Extending code from the saemix package to fit parametric joint models in R

Alexandra Lavalley-Morelle, France Mentré, Jimmy Mullaert* Emmanuelle Comets†

Résumé

The saemix package ([Comets et al., 2017]) was developed to implement the Stochastic Approximation Expectation-Maximisation (SAEM) algorithm to model longitudinal data in R. Version 3 extended the types of outcome to include non-Gaussian data including count, categorical or time-to-event data, with the current version of the package limited to handling one response at at time. In this work, we showcase how to perform parameter estimation, build models and evaluate them using the functions provided in the CRAN version of the package. We also extended the code to joint models of longitudinal and time-to-event data, and implemented the stochastic approximation in [Delattre and Kuhn, 2023] to provide standard errors of estimation. A simulation study was performed to evaluate the performance on the algorithm with increasing model complexity. We present an application to data from hospitalised Covid19 patients ([Lavalley-Morelle et al., 2022]), with a joint model including three biomarkers and a competing risk framework for two events (discharge and death). The extended code, available on the github for saemix development, shows the flexibility and scope of saemix to fit longitudinal data.

Mots-clefs : Non-linear mixed effect models – Stochastic Approximation EM algorithm – Package – Model evaluation – Non Gaussian outcomes

Introduction : Non-linear mixed effect models are used in many fields, including agronomy, animal breeding, imagery and PKPD analyses. The saemix package in R computes the maximum likelihood estimator of the population parameters without any approximation of the model ([Comets et al., 2017]). Version 3, available on CRAN, extended the types of outcomes handled by saemix to non-Gaussian outcomes ([Karimi et al., 2020]). Recently, joint models have become increasingly popular to link repeated measures of biomarkers to the occurrence of terminal events. In R however, with many packages the function used to describe the biomarker dynamic is mainly linear in the parameters, and the survival submodel relies on pre-implemented functions. The objective of this work is to (i) present the saemix package through two examples with different outcome types and (ii) extend the code to fit parametric joint models where longitudinal submodels are not necessary linear in their parameters, evaluating the extension through a simulation study and a real data application.

Methods : saemix implements the SAEM algorithm for parameter estimation in (non)linear mixed effects models. It provides standard errors (SE) from the Fisher Information Matrix (FIM) and via bootstrap methods ([Comets et al., 2021]). The conditional distributions of the individual parameters are estimated using the Hastings-Metropolis algorithm. Inference and automated model building ([M Delattre, 2014]) use the log-likelihood which can be estimated by different approaches. Model diagnostics for continuous outcomes are available through the npde package ([Comets et al., 2008]) and some simulation-based diagnostics have been implemented for non-Gaussian outcomes. In this work, we will show how to use the package to model continuous dose-response data and repeated binary data.

In the second part of this work, we extend the main functions of saemix to joint model estimation (`https://github.com/saemixdevelopment/saemixextension`), keeping the flexibility for users

---

[*] Université Paris Cité, INSERM, IAME, F-75018 Paris, France

[†] Université Paris Cité, INSERM, IAME, F-75018 Paris, France; Université de Rennes, Inserm, EHESP, Irset - UMRS 1085, F-35000 Rennes, France `emmanuelle.comets@inserm.fr` (presenting author)

to define any parametric model and link function between longitudinal and survival parts. To compute standard errors (SE) of parameter estimates, we implemented a recently developed stochastic algorithm requiring only first derivatives ([Delattre and Kuhn, 2023]). We assessed in a simulation study (i) the relative bias and relative root mean square errors of the estimated parameters, (ii) the accuracy of the estimated SE and (iii) the adequacy of the type I error when testing independence between the two submodels. Four joint models were considered in the simulation study, combining a linear or nonlinear mixed-effects model for the longitudinal submodel, with a time-to-event or a competing risk model. We considered a natural link setting where the (predicted) longitudinal values are directly related to the survival process. For each joint model, we simulated 200 datasets of 100 patients. We assumed a rich design and parameters were chosen to obtain about 50% of events in single event models, and about 45% for each of the two events in competing risks models. We finally apply the saemix extension to fit a multivariate joint model describing biomarker dynamics (neutrophils, C-reactive protein (CRP) and arterial pH) via three linear and nonlinear mixed-effects models, with a competing risk model to describe the risk of in-hospital death and discharge from hospital, in a real case study in patients hospitalised for SARS-COV-2 infection ([Lavalley-Morelle et al., 2022]).

Results : saemix can be used to fit many different types of longitudinal data, giving users complete control on the model by defining the log-likelihood function. In the joint model simulation, parameters were precisely and accurately estimated with low bias and uncertainty in all scenarios. For complex joint models (with NLMEM), increasing the number of chains of the algorithm was necessary to reduce bias, but earlier censoring in the competing risk scenario still challenged the estimation. The empirical SE of parameters were very close to those computed with the stochastic algorithm. For more complex joint models (involving NLMEM), some estimates of random effects variances had higher uncertainty and their SE were moderately under-estimated. Finally, type I error was controlled in all joint models.

Conclusion : The saemix package, available on CRAN, uses the efficient SAEM algorithm to perform parameter estimation, build models including covariates and interindividual variability, and offers diagnostics to evaluate these models. Through our extension to fit complex parametric joint models, up till now mainly available in specialised pharmacometrics software such as Monolix or NONMEM, saemix is the only R package supporting estimation of nonlinear joint models with competing risks.

Références

E Comets, K Brendel, and F Mentré. Computing normalised prediction distribution errors to evaluate nonlinear mixed-effect models : the npde add-on package for R. *Comput Meth Prog Biomed*, 90 :154–66, 2008. doi : 10.1016/j.cmpb.2007.12.002.

E Comets, A Lavenu, and M Lavielle. Parameter estimation in nonlinear mixed effect models using saemix, an R implementation of the SAEM algorithm. *J Stat Softw*, 80 :1–41, 2017.

E Comets, C Rodrigues, V Jullien, and M Ursino. Conditional non-parametric bootstrap for non-linear mixed effect models. *Pharm Res*, 38 :1057–66, 2021.

M Delattre and E Kuhn. Estimating Fisher information matrix in latent variable models based on the score function. *HAL sciences*, 2023. URL `https://hal.science/hal-02285712`.

B Karimi, M Lavielle, and E Moulines. f-SAEM : A fast stochastic approximation of the em algorithm for nonlinear mixed effects models. *Comput Stat Data Anal*, 141 :123–38, 2020.

A Lavalley-Morelle, JF Timsit, F Mentré, J Mullaert, and The OUTCOMEREA Network. Joint modeling under competing risks : Application to survival prediction in patients admitted in intensive care unit for sepsis with daily sequential organ failure assessment score assessments. *CPT :PSP*, 11 :1472–84, 2022. doi : 10.1002/psp4.12856. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/psp4.12856`.

MA Poursat M Delattre, M Lavielle. A note on BIC in mixed effects models. *Electr J Stat*, 8 :456–75, 2014.