MOX–Report No. 22/2011

Nonlinear nonparametric mixed-effects models for unsupervised classification

Azzimonti, L.; Ieva, F.; Paganoni, A.M.

MOX, Dipartimento di Matematica "F. Brioschi"
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox@mate.polimi.it http://mox.polimi.it

Nonlinear nonparametric mixed-effects models for unsupervised classification

Laura Azzimonti[♯], Francesca Ieva[♯], Anna Maria Paganoni[♯]

May 12, 2011

[♯] MOX– Modellistica e Calcolo Scientifico
Dipartimento di Matematica "F. Brioschi"
Politecnico di Milano
via Bonardi 9, 20133 Milano, Italy

laura.azzimonti@mail.polimi.it, francesca.ieva@mail.polimi.it,
anna.paganoni@polimi.it

Keywords: Mixed-Effects Models, Nonparametric estimation, EM algorithm, Nonlinear Models.

Abstract

In this work we propose a novel estimation method for nonlinear nonparametric mixed-effects models, aimed at unsupervised classification. The proposed method is an iterative algorithm that alternates a nonparametric EM step and a nonlinear Maximum Likelihood step. We perform simulation studies in order to evaluate the algorithm performances and we apply this new procedure to a real dataset.

1 Introduction

Nonlinear mixed-effects models (NLME models) are mixed-effects models in which at least one of the fixed or random effects appears nonlinearly in the model function. NLME models are increasingly used in several biomedical applications, especially in population pharmacokinetics, pharmacodynamic, immune cells reconstruction and epidemiological studies (see Sheiner and Beal, 1980; Davidian and Gallant, 1993; De Lalla et.al., 2011; Ieva et.al., 2010).

In these fields, statistical modeling based on NLME models takes advantage of tools that allow to distinguish overall population effects from drugs effects or unit specific influence. Mixed-effects models include parameters associated with the entire population (fixed effects) and subject/group specific parameters (random effects). For this reason,

mixed-effects models are able to describe the dynamics of the phenomenon under investigation, even in presence of high between subjects variability. When the random effects represent a deviation from the common dynamic of the population, mixed-effects models provide both estimates for the entire population's model and for each subject's one. In this work random effects have a different meaning, in fact they describe the common dynamic of different groups of subjects. In this framework, mixed-effects models provide only estimates for each group-specific model. Thanks to this property, it will be possible to consider mixed-effects models as an unsupervised clustering tool for longitudinal data and repeated measures. For this reason we focus our attention on the estimation of the distribution of the random effects $\mathscr{P}^*$.

A wide literature exists for parametric modeling of random effects distribution in linear and non linear mixed-effects models. In this framework, Maximum Likelihood (ML) estimators are generally preferred because of their consistency and efficiency. However, due to the non linearity of the likelihood, we are not always able to provide explicitly the parameter estimators. A general and complete overview of linear multilevel models is given in Hox (1995). An analogous overview for nonlinear case is given in Gallant (1987). Fox (2002) shows how `R` and `S-plus` tools estimate linear and generalized linear mixed-effects models with parametric, in particular Gaussian, random effects. Concerning non linear models, in Goldstein (1991) a ML estimation of Gaussian random effect is provided for peculiar nonlinear forms. A stochastic approximation of traditional EM algorithm (SAEM) for estimating Gaussian random effects is suggested in Kuhn and Lavielle (2005), whereas an exact EM algorithm is described in Walker (1996). Finally, Wolfinger (1993) introduces a Laplace approximation for nonlinear random effects marginal distributions. However, parametric assumption may sometimes result too restrictive to describe very heterogeneous or grouped populations. Moreover, when the number of measurements for unit is small, predictions for random effects are strongly influenced by the parametric assumptions. For these reasons nonparametric (NP) framework, which allow $\mathscr{P}^*$ to live in an infinite dimensional space, is attractive. Moreover, it provides in a very natural way a classification tool, as we will highlight later.

Methods for the estimation of linear nonparametric random effects distribution in linear and generalized linear mixed-effects models have been proposed in Aitkin (1996; 1999), whereas Lai and Shih (2003), Davidian and Gallant (1993), Vermut (2004), Antic et al. (2009), among others, deal with nonparametric nonlinear models.

In this work we propose a novel estimation method for nonlinear nonparametric mixed-effects models, aimed at unsupervised classification. The proposed method is an iterative algorithm that alternates a nonparametric EM step and a nonlinear Maximum Likelihood step. The present algorithm is implemented in `R` program (version 2.13.0, R Development Core Team, 2009) and the `R` source code is available upon request. To the best of our knowledge, this is the first example of free software for the estimation of nonlinear nonparametric mixed-effects models.

In Section 2 the general framework of the work is sketched out and the algorithm for the estimation of nonlinear nonparametric random effect (NLNPEM) is described. In Section 3 some simulation studies are presented, both for the linear and nonlinear

case. We first test the performances of our procedure with already existing one in the linear framework, comparing the Wasserstein distance between the true and the estimated distribution of random effects and the goodness of fit index $-2\log L$, then we test NLNPEM on nonlinear case. Section 4 contains an application to real data. Concluding remarks and further developments of the present work are finally discussed in Section 5.

2 Methods

2.1 Model and framework

We consider the following NLME model for longitudinal data:

$$\begin{array}{ll} \mathbf{y}_i = f(\boldsymbol{\beta}, \mathbf{b}_i, \mathbf{t}) + \boldsymbol{\epsilon}_i & i = 1, \ldots, N \\ \boldsymbol{\epsilon}_i \sim \mathscr{N}(\mathbf{0}, \sigma^2 \mathbb{I}_n) & \text{i.i.d.} \end{array} \tag{1}$$

where $\mathbf{y}_i \in \mathbb{R}^n$ is the response variable evaluated at times $\mathbf{t} \in \mathbb{R}^n$ and f is a general, real-valued and differentiable function with $p+q$ parameters. Each parameter of f is treated either as fixed or as random. Fixed effects are parameters associated with the entire population whereas random effects are subject-specific parameters that allow to identify clusters of subjects. $\boldsymbol{\beta} \in \mathbb{R}^p$ is a vector that contain all fixed effects and $\mathbf{b}_i \in \mathbb{R}^q$ is the vector for the i-th subject random effects. The function f is non linear at least in one component of the fixed or random effects. The errors ε_{ij} are associated with the j-th measurement of the i-th longitudinal data. They are normally distributed, independent between different subjects and independent within the same subject. In general, the proposed method could also take account of a different number of observations, located at different times, for different subjects. In (1) we chose not to consider this case in order to ease the notation, but the generalization is straightforward.

Usually random effects are assumed to be Normal distributed, $\mathbf{b}_i \sim \mathscr{N}_q(\mathbf{0}, \Sigma)$, with unknown parameters that, together with $\boldsymbol{\beta}$ and σ, can be estimated through methods based on the likelihood function (see Pinheiro and Bates, 2000). In this parametric framework the maximum likelihood estimators are generally favored by their statistical properties, i.e. consistency and efficiency. Nevertheless the parametric assumptions could be too restrictive to describe highly heterogeneous or grouped data, so it might be necessary to move to a non parametric approach. In our case, we assume $\mathbf{b}_i$, for $i = 1, ..., N$, independent and identically distributed according to a probability measure $\mathscr{P}^*$. Looking for the ML estimator $\hat{\mathscr{P}}^*$ of $\mathscr{P}^*$ in the space of all probability measures on $\mathbb{R}^q$, the discreteness theorem proved in Lindsay (1983), states that $\hat{\mathscr{P}}^*$ is a discrete measure with at most N support points. Therefore the ML estimator of the random effects distribution can be expressed as a set of points $(\mathbf{c}_1, \ldots, \mathbf{c}_M)$, where $M \leq N$ and $\mathbf{c}_l \in \mathbb{R}^q$, and a set of weights $(\omega_1, \ldots, \omega_M)$, where $\omega_l \geq 0$ and $\sum_{l=1}^{M} \omega_l = 1$.

As mentioned above, in this paper we propose an algorithm for the joint estimation of $\boldsymbol{\beta}$, M, $(\mathbf{c}_1, \ldots, \mathbf{c}_M)$ and $(\omega_1, \ldots, \omega_M)$ in the non linear framework of model (1). The estimation of fixed effects $\boldsymbol{\beta}$ and variance σ^2 is performed through the maximization of

the restricted likelihood:

$$L(\boldsymbol{\beta},\sigma^2|\,\mathbf{y}) = p(\mathbf{y}|\,\boldsymbol{\beta},\sigma^2) = \sum_{l=1}^{M} \omega_l \frac{1}{(2\pi\sigma^2)^{(nN)/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{j=1}^{n}\left(y_{ij}-f(\boldsymbol{\beta},\mathbf{c}_l,t_j)\right)^2}.$$

Notice that the number of support points M is estimated by the algorithm as well and we do not have to fix it a priori. Since we don't have to specify a priori the number of support points and in consequence the number of groups, the nonparametric mixed-effects model could be interpreted as an unsupervised clustering tool for longitudinal data. This tool could be very useful in order to identify the groups of subjects to be used in the analysis.

2.2 NLNPEM algorithm

The algorithm proposed for the estimation of the parameters of model (1) arises from the framework described in Schumitzky (1991), and alternates two steps. The first one is a nonparametric EM step whereas the second one is a non linear maximum-likelihood step. The nonparametric EM step estimates the discrete q-dimensional distribution $(\mathbf{c},\boldsymbol{\omega})$ of the random effects $\mathbf{b}_i$. The non linear maximum likelihood step provides an estimation of the fixed effects β and the variance σ^2, given $\mathbf{b}_i$.

The nonparametric EM step consists in an update of the parameters of the discrete distribution $(\mathbf{c},\boldsymbol{\omega})$ that increases the likelihood function. The property of increasing the likelihood was proved in Schumitzky (1991). The update is the following:

$$\begin{cases} \tilde{\omega}_l = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} W_{il} \\ \tilde{\mathbf{c}}_l = \arg\max\limits_{\mathbf{c}} \left[\displaystyle\sum_{i=1}^{N} W_{il} \ln p(\mathbf{y}_i|\,\boldsymbol{\beta},\sigma^2,\mathbf{c})\right] \end{cases} \tag{2}$$

where

$$W_{il} = \frac{\omega_l p(\mathbf{y}_i|\,\boldsymbol{\beta},\sigma^2,\mathbf{c}_l)}{\sum_{k=1}^{M}\omega_k p(\mathbf{y}_i|\,\boldsymbol{\beta},\sigma^2,\mathbf{c}_k)}$$

and

$$p(\mathbf{y}_i|\,\boldsymbol{\beta},\sigma^2,\mathbf{c}_l) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_{j=1}^{n}\left(y_{ij}-f(\boldsymbol{\beta},\mathbf{c}_l,t_j)\right)^2}.$$

The coefficients W_{il} represent the probability of $\mathbf{b}_i$ being equal to $\mathbf{c}_l$ conditionally to the observation $\mathbf{y}_i$ and given the fixed effects $\boldsymbol{\beta}$ and the variance σ^2, that is

$$W_{il} = p(\mathbf{c}_l|\,\mathbf{y}_i,\boldsymbol{\beta},\sigma^2)$$

in fact,

$$W_{il} = \frac{p(\mathbf{c}_l)p(\mathbf{y}_i|\,\boldsymbol{\beta},\sigma^2,\mathbf{c}_l)}{p(\mathbf{y}_i|\,\boldsymbol{\beta},\sigma^2)} = \frac{p(\mathbf{y}_i,\mathbf{c}_l|\,\boldsymbol{\beta},\sigma^2)}{p(\mathbf{y}_i|\,\boldsymbol{\beta},\sigma^2)} = p(\mathbf{c}_l|\mathbf{y}_i,\boldsymbol{\beta},\sigma^2).$$

In order to estimate $\mathbf{b}_i$ for $i = 1,\ldots,N$ we want to maximize the conditional probability of $\mathbf{b}_i$ conditionally to the observations $\mathbf{y}_i$ and given the fixed effects $\boldsymbol{\beta}$ and the error variance σ^2. For this reason the estimation of the random effects, $\hat{\mathbf{b}}_i$, is obtained maximizing W_{il} over l, that is

$$\hat{\mathbf{b}}_i = \mathbf{c}_{\tilde{l}} \text{ if } \tilde{l} = \arg\max_l W_{il}.$$

During the nonparametric EM step, we could also reduce the support of the discrete distribution. The reduction of the support is performed in order to cluster the support of random effects. This support reduction consists in both making points very close to each other collapse and removing points with very low weight and not associated with any subject. In particular if two points are too close, that is $\|\mathbf{c}_l - \mathbf{c}_k\| < D$, where D is a tuning tolerance parameter, than we replace $\mathbf{c}_l$ and $\mathbf{c}_k$ with a new point $\mathbf{c}_{\min\{l,k\}} = (\mathbf{c}_l + \mathbf{c}_k)/2$ with weight $\omega_{\min\{l,k\}} = \omega_l + \omega_k$. Otherwise, if $\omega_l < \tilde{\omega}$, where $\tilde{\omega}$ is another tuning tolerance parameter, and the subset $\left\{i : \hat{\mathbf{b}}_i = \mathbf{c}_l\right\}$ is empty, we remove the point $\mathbf{c}_l$. The thresholds D and $\tilde{\omega}$ are two complexity parameters that affect the estimation of the nonparametric distribution; the higher D is set, the lower is the number of groups. For this reason the two complexity parameters define a trade off between bias and high number of groups. In this work we prefer setting D low in order to obtain an higher number of groups and, in case, cluster them later.

The non linear maximum likelihood step provides the estimation of the fixed effects $\boldsymbol{\beta}$ and the errors variance σ^2, given $\mathbf{b}_i = \hat{\mathbf{b}}_i$. In this step we maximize the non linear log-likelihood:

$$\ell(\boldsymbol{\beta}, \sigma^2 \big| \mathbf{y}, \hat{\mathbf{b}}) = -\frac{nN}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{j=1}^{n}\left(y_{ij} - f(\boldsymbol{\beta}, \hat{\mathbf{b}}_i, t_j)\right)^2$$

where $\hat{\mathbf{b}}_i$ is the estimation of random effects for the i-th subject provided in the nonparametric EM step.

The algorithm, given a starting discrete distribution with N support points for the random effects and a starting estimate for the fixed effects, alternate the nonparametric EM step and the non linear maximum likelihood step until convergence. Technical details together with the sketch of the algorithm are reported in Appendix A.

3 Simulation studies

In order to validate the proposed estimation algorithm and to compare it with different procedures, we perform two simulation studies. Since we are mainly interested in classifying curves in an unsupervised framework, we focus our attention on the estimation of random effects distribution.

In the first simulation study (Section 3.2), we test our algorithm in a linear framework, in order to compare results of our procedure with those obtained with the algorithm introduced in Aitkin (1996) and implemented in the `npmlreg` `R`-package (see Einbeck et al., 2009). In the second one (Section 3.3), we consider two classic non

linear functions f in (1): the exponential and the logistic growth curves. For each case we design a test set of simulated curves (details are provided in Appendix B) and we evaluate the algorithm performance in the estimation of the random effects computing the Wasserstein distance (defined in Section 3.1) between the true and the estimated distribution of the random effects.

3.1 Wasserstein Distance

Evaluating the goodness of the estimation of a discrete distribution is not a straightforward task. Indeed, ways of comparing the true and the estimated probability distribution of the model effects have to take in account both support location and weights, for this reason we adopt a multidimensional version of Wasserstein distance (see Gibbs and Su, 2002). The Wasserstein distance between two probability measures $\mu,\ \nu$ on a subset Ω of the metric space $\mathbb{R}$ is defined as

$$d_W(\mu,\nu) = \int_{-\infty}^{\infty} |F(x) - G(x)| dx, \tag{3}$$

where F and G are the cumulative distribution functions of μ and ν respectively. The generalization to the q-dimensional case is straightforward. When the probability measures μ and ν are discrete, the computation of the integral in (3) is very easy, even in the q-dimensional case. It is known that the Wasserstein metric assumes values in $[0, |\Omega|]$, where $|\Omega|$ is the Euclidean measure of the support space Ω. For this reason, the Wasserstein distance divided by $|\Omega|$ is a good performance index for the evaluation of the estimates in the simulations study.

3.2 Linear cases

In this section, a simulation study for linear models is considered, therefore f in model (1) is linear. The general model, for $i = 1, \ldots, N$, include three different cases, that are:

$$\mathbf{y}_i = \begin{cases} \alpha + d_i\mathbf{t} + \boldsymbol{\epsilon}_i & \text{(random-slope case)} \\ a_i + \boldsymbol{\delta}\mathbf{t} + \boldsymbol{\epsilon}_i & \text{(random-intercept case)} \\ a_i + d_i\mathbf{t} + \boldsymbol{\epsilon}_i & \text{(fully random case)} \end{cases}$$

where $\boldsymbol{\epsilon}_i$ are i.i.d. from $\mathcal{N}(\mathbf{0}, \sigma^2\mathbb{I}_n)$ and $\mathbf{t}$ is the vector of sampling times. Intercept and slope are treated as fixed or random effects according to the different cases. In the fully random case, both slope and intercept parameters are considered random, i.e. $\mathbf{b}_i = (d_i, a_i)$, whereas in the random-slope and random intercept case, $b_i = d_i$ and $b_i = a_i$ respectively. The interest is focused on random effects estimation, because our main goal is to test the performance of our algorithm in identifying the correct number of groups in simulated data and in estimating properly location and weights of different groups. Testing the linear case enables us to compare results of our algorithm with those carried out by the `R` algorithm `npmlreg`, which implements Aitkin (1996) procedure of non parametric random effect estimation. To be noticed is that our method is not efficient in the linear case, since it doesn't take advantage of the linearity of the problem.

However it doesn't need any a priori specification of the number of support points of the random effects. Even if we don't specify the exact number of groups beforehand, the proposed method is able of carrying out a good estimation of the random effects distribution.

We simulated 8 datasets of linear growth curves each grouped in a different number of balanced or unbalanced clusters (from 2 to 10 clusters). Some examples of simulated data are shown in left panels of Figure 1. Parameters specification and details of each set of curves are reported in Appendix B. All these datasets represent typical situations in which fitting a parametric mixed-effects model could be wrong because random effects are not normally distributed. On these datasets, we fitted models with both the NLNPEM method and the nonparametric maximum likelihood approach introduced in Aitkin (1996).

The method introduced in Aitkin (1996) is a method for fitting overdispersed generalized linear models: the idea is to approximate the unknown and unspecified distribution of the random effects by a discrete mixture of densities from exponential family. This approximation leads to a simple expression of the marginal likelihood that can be maximized using a standard EM algorithm. Once specified the model and the number of random effects groups k, the `R` package `npmlreg` fits a linear mixed-effects model using nonparametric maximum likelihood. Since we are testing the proposed method in a simulation setting, when `npmlreg` method is used we provide the correct number of groups, whereas, when NLNPEM is used, we don't have to. The N starting points for random effects distribution are randomly chosen in a proper range and the starting fixed effects are estimated through linear least squares. Finally, the tolerance D is set equal to 0.05 and $\tilde{\omega} = 0.05$. According to the dimension of the random effect ($q = 1$ for random-slope or random-intercept case, $q = 2$ when both effects are random), we properly define the model in `npmlreg` and NLNPEM algorithms. Notice that `npmlreg` does not allow to select one dimensional random effect for slope only but provides a random effects estimation for both intercept and slope parameters. In this case, in order to correctly compare the two methods, we have set also in the NLNPEM method both slope and intercept to be random in the random-slope case. Of course, in the NLNPEM method, random effects only for the slope may be selected by the user, if necessary.

In Tables 4, 5 and 6, of Appendix C, results of `npmlreg` and NLNPEM algorithms for three representative cases are compared, i.e the estimations of random effects in terms of points and weights are reported and compared with the corresponding true distributions. Observing estimated values reported in Tables 4, 5 and 6, it can be argued that both methods estimate well both the discrete random components of the model and the fixed effects when a small number of groups is considered. Increasing the number of groups, the two algorithms show different behaviors. In particular we notice that, for large number of groups, `npmlreg` misses some points of the nonparametric distribution, whereas NLNPEM performs better, even ignoring the true number of groups. The number of groups estimated by the NLNPEM algorithm depends in general on the tuning tolerances D and $\tilde{\omega}$, introduced in Section 2.2. This algorithm tends to overestimate the number of points of the discrete distribution. However, even if the number of points

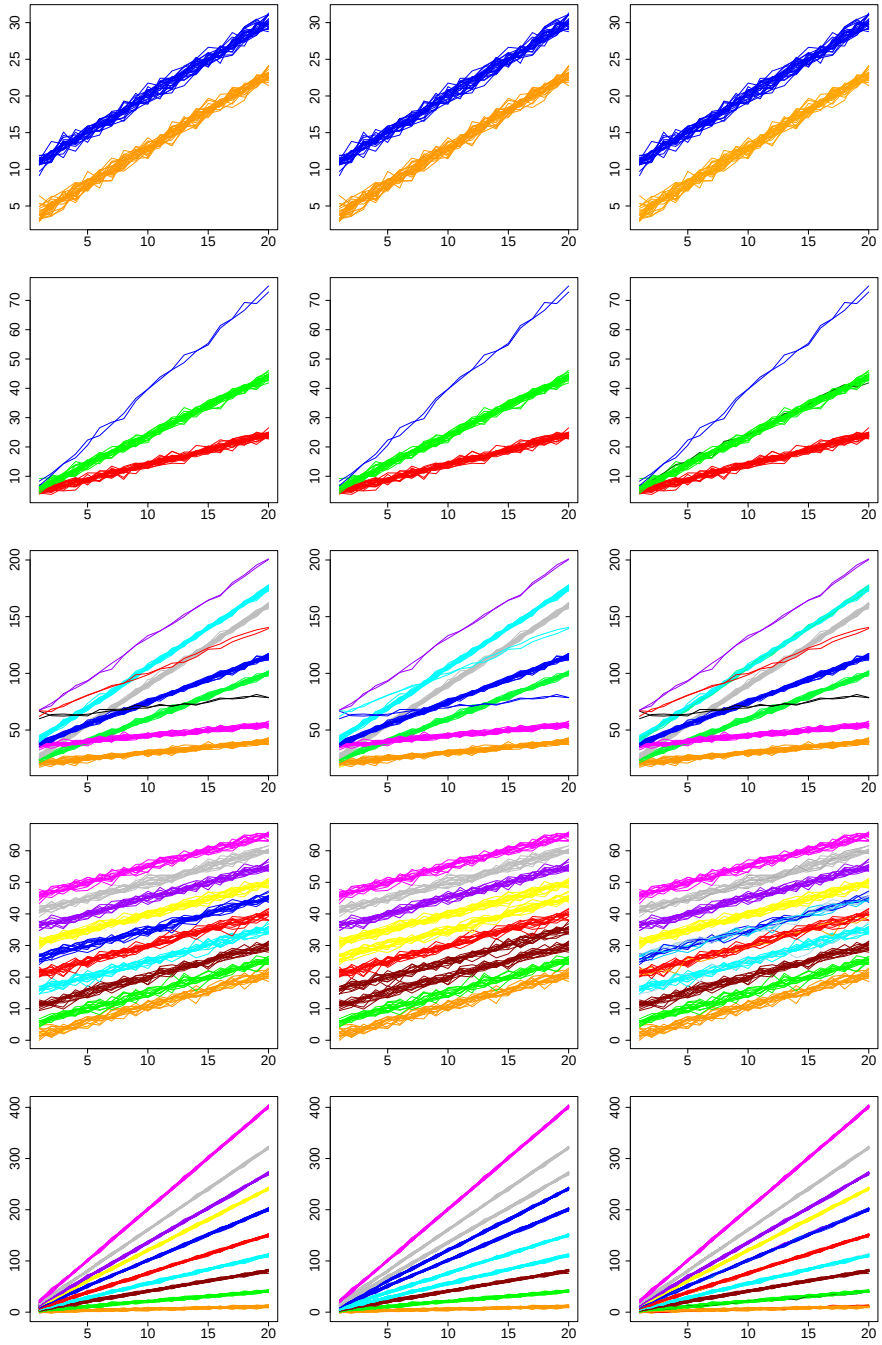

Figure 1: Simulated data (left panels), `npmlreg` (central panels) and NLNPEM classification (right panels) in lin2I, lin3S, lin9SI, lin10I and lin10S datasets respectively. Different colors are used to represent real groups (left panels), groups identified by `npmlreg` and NLNPEM methods (central and right panels respectively).

is greater than the real number, the points tend to cluster near the true ones. Moreover, summing the weights of the points in each cluster, we obtain results similar to the exact weights. The hints concerning the number of groups provided by NLNPEM algorithm make this method a powerful tool in explorative analyses within an unsupervised framework. The NLNPEM method is also capable of detecting outlier groups, whereas the `npmlreg` method is able to detect them only in presence of small number of groups. In general, we notice that sometimes `npmlreg` method performs poorly in estimation or even misses convergence, whereas NLNPEM doesn't. These situations happen in particular when there are 9 different groups both for intercept and slope ("lin9SI" dataset) and when there are 10 groups for slope or intercept("lin10S" and "lin10I" dataset respectively).

Model	Wasserstein distance		$-2\log L$	
	`npmlreg`	NLNPEM	`npmlreg`	NLNPEM
lin2S	0.013572	0.013724	2861.2	942.0
lin2I	0.004538	0.005187	2097.7	190.5
lin4SI	0.008121	0.006298	5974.4	2017.7
lin3S	0.003041	0.004651	2839.8	912.7
lin3I	0.003454	0.003454	2938.3	1017.2
lin9SI	0.017756	0.001565	16127.0	5376.7
lin10S	0.033632	0.000410	76716.1	18025.9
lin10I	0.023045	0.001649	12795.8	2947.3

Table 1: Normalized Wasserstein distances and $-2\log L$ index for `npmlreg` and NLNPEM algorithm respectively in the simulated linear cases.

In order to resume the goodness of fit of NLNPEM method and the `npmlreg` one, we finally compare the normalized Wasserstein distances between the true discrete random effects distribution and the estimated one through the two methods, for each simulated set of linear curves. Results are reported in Table 1, together with the goodness of fit index $-2\log L$.
To be noticed is that, in the case of Wasserstein distance, results are similar for all datasets where both algorithm perform well. On the other hand, significant differences exist in cases with large number of groups, where NLNPEM performances are much better both in terms of Wasserstein distance and $-2\log L$.

3.3 Non linear cases

In this section we describe two nonlinear case studies: the exponential and the logistic growth model. These two models are among the most used in nonlinear mixed-effects framework because they find application in several areas like pharmacokinetics and epidemiological studies.
Since other nonlinear nonparametric methods are not available for free software, we are not able to compare the NLNPEM results with those obtained with other methods; for this reason we will only test NLNPEM performances, providing the normalized

Wasserstein distance between the true distribution and the estimated one.

3.3.1 Exponential growth model

We first describe the exponential case, in which we consider the following nonlinear function in model (1):

$$f(t) = \alpha\left(1 - e^{-\lambda t}\right)$$

which is nonlinear in λ. The two parameters α and λ represent respectively the asymptote and the growth rate.
In this case study we consider only random effects for the asymptote, that means that the mixed-effects model becomes

$$\mathbf{y}_i = a_i\left(1 - e^{-\lambda \mathbf{t}}\right) + \boldsymbol{\epsilon}_i$$

where $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbb{I}_n)$ are i.i.d. errors, a_i are the random effects for the asymptote ($b_i = a_i$) and λ is the fixed effect for the growth rate ($\beta = \lambda$).

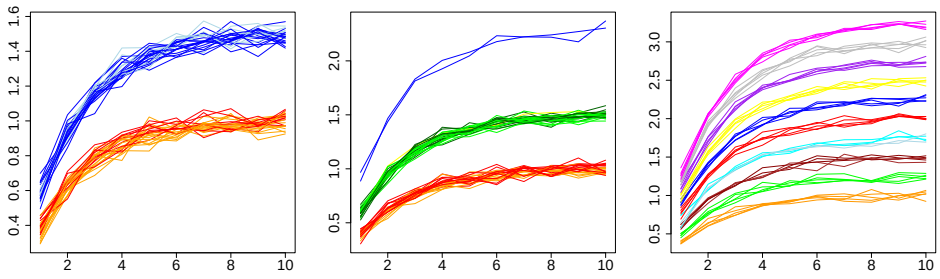

Figure 2: NLNPEM classification in exp2A, exp3A and exp10A datasets respectively with exponential model.

We simulated 3 datasets of exponential growth curves, described in Appendix B, in which only asymptote varies. The starting random effects distribution has N support points, randomly chosen in a proper range, and the starting fixed effects are estimated through nonlinear least squares. The tuning tolerance parameter D is set equal to 0.01 and $\tilde{\omega} = 0.05$. Figure 2 shows original datasets, where each curve is colored according to the group estimated by NLNPEM method.
The estimated number of groups is larger than the real one in all the three cases; however the estimated random effects create the right number of clusters located close to the correct points. In the exp3A case the NLNPEM method is also able to identify the *outlier* group estimating well the location and the weight of the random effects.
The performance of NLNPEM method is evaluated in this case only in terms of normalized Wasserstein distance, shown in Table 2.

Model	Wasserstein distance
exp2A	0.030048
exp3A	0.015025
exp10A	0.011524

Table 2: Normalized Wasserstein distances for NLNPEM algorithm in the simulated exponential cases.

3.3.2 Logistic growth model

The second nonlinear model tested is the logistic growth model. In this case, the nonlinear function is:

$$f(t) = \frac{\alpha}{1+e^{-\frac{t-\delta}{\gamma}}}$$

where α represent the asymptote, δ is the inflection point, which correspond to the time at which the growth curve reaches the half of the asymptote, and γ is the time elapsed between δ and the time at which the growth curve reaches 3/4 of the asymptote level. The parameter γ will always be treated as a fixed effect while the asymptote and the inflection point will be treated either as fixed or as random effect according to different cases. The general model, which is nonlinear in λ and γ, include three different cases:

$$\mathbf{y}_i = \begin{cases} \dfrac{a_i}{1+e^{-\frac{\mathbf{t}-\delta}{\gamma}}} + \boldsymbol{\epsilon}_i & \text{(random-asymptote case)} \\ \dfrac{\alpha}{1+e^{-\frac{\mathbf{t}-d_i}{\gamma}}} + \boldsymbol{\epsilon}_i & \text{(random-inflection case)} \\ \dfrac{a_i}{1+e^{-\frac{\mathbf{t}-d_i}{\gamma}}} + \boldsymbol{\epsilon}_i & \text{(random-asymptote and inflection case)} \end{cases} \tag{4}$$

where $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbb{I}_n)$ are i.i.d. errors, a_i and d_i represent the random effects for the asymptote and the inflection point, while α, δ and γ represent the fixed effects. In particular in the varying asymptote case $b_i = a_i$ and $\beta = (\delta, \gamma)$, in the varying inflection case $b_i = d_i$ and $\beta = (\alpha, \gamma)$ and in the varying asymptote and inflection case $\mathbf{b}_i = (a_i, d_i)$ and $\beta = \gamma$.
We simulated 8 datasets of logistic growth curves that include all the cases resumed in (4). Each dataset is composed by a different number of balanced or unbalanced groups (from 2 to 10 clusters) similar to those presented in the linear framework. Details are provided in Appendix B.

Since the NLNPEM method is able to fit all three models resumed in (4), we fit the right model for each dataset. The starting random effects distribution has N support points, randomly chosen in a proper range, and the starting fixed effects are estimated through nonlinear least squares. We set the tolerance D equal to 0.05 and $\tilde{\omega} = 0.05$. Figure 3 shows original datasets, where each curve is colored according to the group estimated by NLNPEM method. We notice in Figure 3 that, even if we don't specify a priori the correct number of groups, we are able to cluster correctly the subjects both when there are few groups and when there are many. The method is also able to capture correctly *outliers* groups; in all the unbalanced cases the proposed method recognize the

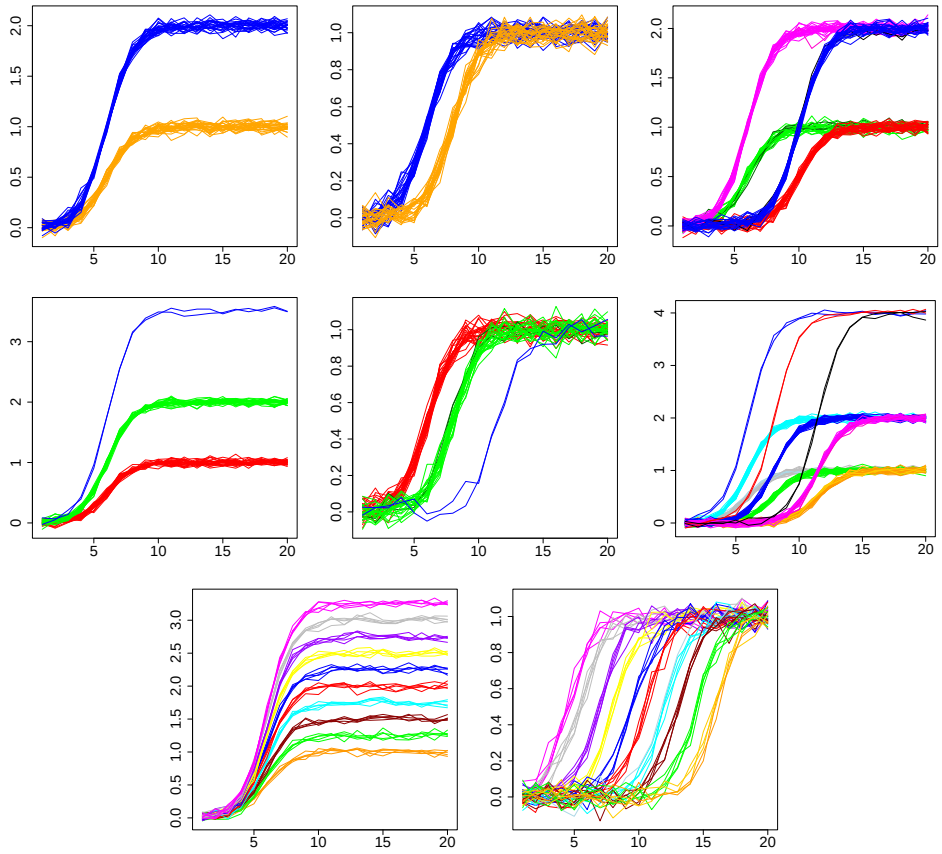

Figure 3: NLNPEM classification in logis2A, logis2I, logis4AI, logis3A, logis3I, logis9AI, logis10A and logis10I datasets respectively with logistic growth model.

outliers groups and estimate well both the location and the weight of random effects. In order to test the NLNPEM method we can compare these results with those obtained considering always both asymptote and inflection point as random effects. For the two varying asymptote and inflection cases we have obviously fitted only the model with two random effects. The normalized Wasserstein distances are shown in Table 3; the left column represents the normalized Wasserstein distance for a model with one random effect while the right one represent the same distance for models with two random effects. We first notice that the normalized Wasserstein distances are always very low, that means that the NLNPEM method is able to estimate well both random and fixed effects even in presence of a high number of groups. We also notice, comparing results for the same case study with one or two random effects, that the normalized Wasserstein distances are always very close together. Both observing the Wasserstein distances and the fitted curves obtained with the model with two random effects, we notice that in the NLNPEM method we are allowed to consider more parameters as random effects than needed, without damaging the parameter estimation. In particular this approach could

Model	Wasserstein distance	
	$q=1$	$q=2$
logis2A	0.000150	0.000450
logis2I	0.003202	0.012171
logis4AI	–	0.004869
logis3A	0.000396	0.000629
logis3I	0.007243	0.010250
logis9AI	–	0.006477
logis10A	0.001286	0.0.0015
logis10I	0.004664	0.005207

Table 3: Normalized Wasserstein distances for NLNPEM algorithm in the simulated logistic cases.

be useful when we don't know which are the parameters to be considered random. For this purpose we could perform a first analysis considering all parameters as random effects and then fit a second model fixing the parameters that show a very low variability. This approach could be performed with the NLNPEM method because it can handle both random and fixed effects whereas other previous methods cannot.

4 Application to NON STEMI data

In this section we study a dataset concerning Hospital Discharges of patients affected by Acute Myocardial Infarction (AMI) without ST-segment elevation (NON-STEMI). These data have already been studied in Ieva et al., 2010. Figure 4 represents the normalized number of NON-STEMI diagnoses along the time period 2000-2007 grouped by hospital and relative to the 30 largest clinical institutions of Regione Lombardia. For each hospital the yearly number of diagnoses has been standardized by the hospital total number of diagnoses in the time period 2000-2007.

As pointed out in Ieva et al., 2010, the random-inflection case in model (4) seems to capture the common "S-shaped" growing pattern. The NLNPEM algorithm clusters the hospitals in $M=2$ different groups, according to the estimated discrete distribution of the random effect for the inflection point (see Figure 4). The estimated fixed effects are $\hat{\alpha}=0.16$ and $\hat{\gamma}=1.31$, the estimated discrete measure $\hat{\mathscr{P}}^*$ is concentrated on $(\hat{\mathbf{c}}_1,\hat{\mathbf{c}}_2)=(-3.76,-2.43)$ with weights $(\hat{\omega}_1,\hat{\omega}_2)=(0.2,0.8)$ and the estimated variance is $\hat{\sigma}^2=7.7\cdot 10^{-4}$. This analysis, performed with $D=0.05$ and $\tilde{\omega}=0.05$, backs up the presence of two groups of hospitals according to different inflection points and automatically detects an unsupervised cluster structure. Even if clinical best practice maintains that there is no evidence for a greater incidence of NON-STEMI in this period it is known that since the early 2000s a new diagnostic procedure - the *troponin* exam - has been introduced and this could have produced an increased number of positive diagnoses, by easing NON-STEMI detection. Hence, the presence of 2 clusters could be a consequence of the different hospital timings in the introduction and adoption of this practice. This hypothesis cannot be validated directly since the timings of

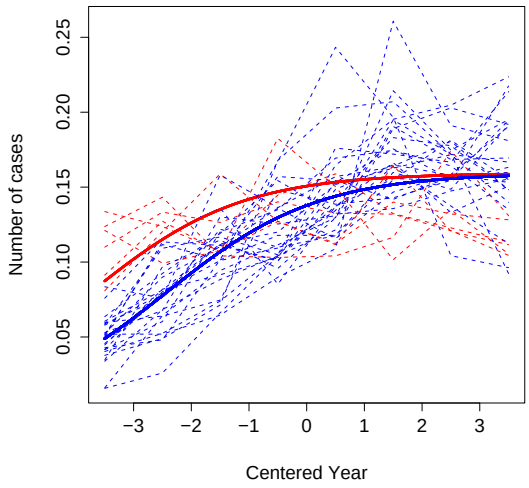

Figure 4: Standardized number of AMI without ST-segment elevation diagnoses in the period 2000 - 2007 in the 30 largest clinical institutions of Lombardia Region. The year has been centered and normalization has been carried out standardizing the yearly number of diagnoses for each hospital by total number of diagnoses in the time window $2000-2007$. Real data are colored according to the NLNPEM clusters and NLNPEM fitted models are superimposed.

adoption of the troponin exam by the 30 different hospitals included in the analysis are not available.
The good agreement with previous results detailed in Ieva et al., 2010 together with the great advantage of a non-parametric approach advocates the real profit in using this new estimation algorithm.

5 Conclusions

In this work, we proposed a new estimation method for nonlinear nonparametric mixed effect models, aimed at unsupervised classification.
The proposed method is based on an iterative algorithm (named NLNPEM) that alternates a nonparametric EM step and a ML step for the maximization of a nonlinear likelihood function. We first tested this procedure in a linear framework against the already existing tool for nonparametric random effects estimation (the `npmlreg` `R` package), in order to compare the performances of the new method in terms of random effects distribution. Results show that our method performs well both in terms of Wasserstein distance and $-2\log L$, even ignoring the real number of groups, and that it always converges, even in those cases where several groups are present. Then we tested NLNPEM algorithm also in simulated test set within nonlinear frameworks of exponential and logistic growth. In both these cases, the number of groups and distribution of random effect are correctly and effectively identified. Finally, an application to real

data of NON-STEMI is presented, where the potential of our method in unsupervised clustering analysis is highlighted.

A Details on NLNPEM Algorithm

The NLNPEM is the following:

1. Define a starting discrete distribution for random effects with support on N points $(\mathbf{c}^{(0)}, \boldsymbol{\omega}^{(0)})$, a starting estimate for the fixed effects $\boldsymbol{\beta}^{(0)}$ and for $\sigma^{2(0)}$ and the tolerance parameters D and $\tilde{\omega}$;

2. given $(\mathbf{c}^{(k-1)}, \boldsymbol{\omega}^{(k-1)})$, $\boldsymbol{\beta}^{(k-1)}$ and $\sigma^{2(k-1)}$, perform the EM step (without the support reduction) in order to update the support points $\mathbf{c}^{(k)}$ and the weights $\boldsymbol{\omega}^{(k)}$ of the random effect distribution, according to equation (2);

3. given $(\mathbf{c}^{(k)}, \boldsymbol{\omega}^{(k)})$, perform the nonlinear maximum likelihood step in order to estimate the fixed effects $\boldsymbol{\beta}^{(k)}$ and the error variance $\sigma^{2(k)}$;

4. iterate steps 2 and 3 until convergence;

5. reduce the support of the discrete distribution, according with the tuning parameters D and $\tilde{\omega}$;

6. given $(\mathbf{c}^{(k-1)}, \boldsymbol{\omega}^{(k-1)})$, $\boldsymbol{\beta}^{(k-1)}$, $\sigma^{2(k-1)}$, D and $\tilde{\omega}$, perform the EM step with the support reduction in order to update the support points $\mathbf{c}^{(k)}$ and the weights $\boldsymbol{\omega}^{(k)}$ of the random effect distribution, according to equation (2);

7. given $(\mathbf{c}^{(k)}, \boldsymbol{\omega}^{(k)})$, perform the nonlinear maximum likelihood step in order to estimate the fixed effects $\boldsymbol{\beta}^{(k)}$ and the error variance $\sigma^{2(k)}$;

8. iterate steps 6 and 7 until convergence.

The algorithm reaches convergence when parameters and discrete distribution stop changing or when there is no variation in the log-likelihood function.

B Details on simulation study

B.1 The Linear case

We simulated 8 datasets of linear curves grouped in a number of clusters that vary form 2 to 10. Different values of the error variance σ^2 have been chosen for each test set, in order to obtain noisy observations for each curve. Some examples of simulated data are shown in left panels of Figure 1. Datasets addressed with the name "S" contain groups in which only slopes is random, "I" datasets contain groups where only intercept is random and "SI" datasets contain curves where both slope and intercept are random. The simulated datasets are then:

- lin2S: 2 balanced groups, each one composed by 25 curves, with the same intercept (equal to 4), 2 different slopes ($\mathbf{c} = (c_1, c_2) = (1, 2)$) and $\sigma = 1$;
- lin2I: 2 balanced groups, each one composed by 25 curves with the same slope (equal to 1), 2 different intercept ($\mathbf{c} = (c_1, c_2) = (3, 10)$) and $\sigma = 0.65$;
- lin4SI: 4 balanced groups, each one composed by 25 curves, where location points $\mathbf{c} = (\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3, \mathbf{c}_4)$ are obtained from all possible combinations of 2 different slopes (equal to 1 and 3) and 2 different intercepts (equal to 40 and 60), i.e. $\mathbf{c}_1 = (1, 40)$, $\mathbf{c}_2 = (1, 60)$, $\mathbf{c}_3 = (3, 40)$ and $\mathbf{c}_4 = (3, 60)$ with $\sigma = 1$;
- lin3S: 3 unbalanced groups, composed by 24, 24 and 2 curves respectively, with the same intercept (equal to 4), 3 different slopes ($\mathbf{c} = (c_1, c_2, c_3) = (1, 2, 3.5)$) and $\sigma = 1$;
- lin3I: 3 unbalanced groups, composed by 24, 24 and 2 curves respectively, with the same slope (equal to 1), 3 different intercepts ($\mathbf{c} = (c_1, c_2, c_3) = (2, 7, 14)$) and $\sigma = 1$;
- lin9SI: 9 unbalanced groups, 6 of whom containing 24 curves and 3 containing 2 curves, where location points $\mathbf{c} = (c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9)$ are obtained from all possible combinations of 3 different slopes (equal to 1, 4 and 7) and 3 different intercept (equal to 20, 35 and 60) with $\sigma = 1.5$;
- lin10S: 10 balanced groups, each one composed by 50 curves with the same intercept (equal to 1), 10 different slopes ($\mathbf{c} = (c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9, c_{10}) = (0.5, 2, 4, 5.5, 7.5, 10, 12, 13.5, 16, 20)$) and $\sigma = 1.5$;
- lin10I: 10 balanced groups, each one composed by 15 curves with the same slope (equal to 1), 10 different intercepts ($\mathbf{c} = (c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9, c_{10}) = (1, 5, 10, 15, 20, 25, 30, 35, 40, 45)$) and $\sigma = 1$.

B.2 The Exponential case

We simulated 3 datasets of exponential growth curves where only asymptote varies and is considered as random. All datasets are then addressed with the name "A". They are:

- exp2A: 2 balanced groups, each one composed by 25 curves, with the same growth rate ($\lambda = 0.5$), 2 different asymptotes ($\mathbf{c} = (c_1, c_2) = (1, 1.5)$) and $\sigma = 0.04$;
- exp3A: 3 unbalanced groups of 24, 24 and 2 curves respectively, with the same growth rate ($\lambda = 0.5$), 3 different asymptotes ($\mathbf{c} = (c_1, c_2, c_3) = (1, 1.5, 2.3)$) and $\sigma = 0.04$;
- exp10A: 10 balanced groups, each one composed by 5 curves, with the same growth rate ($\lambda = 0.5$), 10 different asymptotes ($\mathbf{c} = (c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9, c_{10}) = (1, 1.25, 1.5, 1.75, 2, 2.25, 2.5, 2.75, 3, 3.25)$ and $\sigma = 0.04$.

B.3 The Logistic case

We simulated 8 datasets of logistic growth curves. Datasets addressed with the name "A" represent random asymptote cases, "I" datasets contain groups where only inflection point is random and "AI" ones contain curves where both asymptote and inflection point are random. We then have:

- logis2A: 2 balanced groups, each one composed by 25 curves, with $\delta = 6$, $\gamma = 1$, 2 different asymptotes ($\mathbf{c} = (c_1, c_2) = (1, 2)$) and $\sigma = 0.04$;
- logis2I: 2 balanced groups, each one composed by 25 curves, with $\alpha = 1$, $\gamma = 1$, 2 different inflection points ($\mathbf{c} = (c_1, c_2) = (6, 8)$) and $\sigma = 0.04$;
- logis4AI: 4 balanced groups, each one composed by 25 curves, where location points $\mathbf{c} = (\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3, \mathbf{c}_4)$ are obtained from all possible combinations of 2 different asymptotes (equal to 1 and 2) and 2 different inflection points (equal to 6 and 10), i.e. $\mathbf{c}_1 = (1, 6)$, $\mathbf{c}_1 = (1, 10)$, $\mathbf{c}_1 = (2, 6)$ and $\mathbf{c}_4 = (2, 10)$ with $\gamma = 1$ and $\sigma = 0.04$;
- logis3A: 3 unbalanced groups of 24, 24 and 2 curves respectively, with $\delta = 6$, $\gamma = 1$, 3 different asymptotes ($\mathbf{c} = (c_1, c_2, c_3) = (1, 2, 3.5)$) and $\sigma = 0.04$;
- logis3I: 3 unbalanced groups of 24, 24 and 2 curves respectively, with $\alpha = 1$, $\gamma = 1$, 3 different inflection points ($\mathbf{c} = (c_1, c_2, c_3) = (6, 8, 11.5)$) and $\sigma = 0.04$;
- logis9AI: 9 unbalanced groups of curves (6 of whom containing 24 curves and 3 containing 2 curves), where location points $\mathbf{c} = (c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9)$ are obtained from all possible combinations of 3 different asymptotes (equal to 1, 2 and 4) and 3 different inflection points (equal to 6, 8 and 11.5) with $\gamma = 1$ and $\sigma = 0.04$;
- logis10A: 10 balanced groups, each one composed by 5 curves, with $\delta = 6$, $\gamma = 1$, 10 different asymptotes ($\mathbf{c} = (c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9, c_{10}) = (1, 1.25, 1.5, 1.75, 2, 2.25, 2.5, 2.75, 3, 3.25)$ and $\sigma = 0.04$;
- logis10I: 10 balanced groups, each one composed by 5 curves, with $\alpha = 1$, $\gamma = 1$, 10 different inflection points ($\mathbf{c} = (c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9, c_{10}) = (4.5, 5.5, 7, 8, 9.5, 10.5, 12, 13, 14.5, 16)$ and $\sigma = 0.04$.

C Comparison of results

Comparison of estimates carried out by `npmlreg` and NLNPEM method are reported here, for some cases of interest mentioned in the paper.

- Linear case - Random-intercept case (lin2I)
- Linear case - Random-slope case (lin3S)
- Linear case - Random-intercept case (lin10I)

effects		True	npmlreg	NLNPEM
fixed	slope	1	1.0021	1.0022
random	intercept 1	3	2.9382	2.9368
	(weight 1)	(0.5)	(0.5)	(0.5)
	intercept 2	10	10.0150	10.0136
	(weight 2)	(0.5)	(0.5)	(0.5)

Table 4: Estimates carried out by npmlreg and NLNPEM method on lin2I dataset, where intercept is considered as random, with 2 balanced groups.

effects		True	npmlreg	NLNPEM
random	slope 1	1	1.0107	1.0107
	(weight 1)	(0.48)	(0.48)	(0.48)
	slope 2	2	1.9982	2.0030 1.9637
	(weight 2)	(0.48)	(0.48)	(0.4214) (0.0585)
	slope 3	3.5	3.5250	3.5250
	(weight 3)	(0.04)	(0.04)	(0.04)
random	intercept	4	3.9326	3.9326
	intercept	4	4.0751	3.9954 4.6502
	intercept	4	3.3717	3.7174

Table 5: Estimates carried out by npmlreg and NLNPEM method on lin3S dataset, where slope is considered as random, with 3 unbalanced groups.

effects		True	npmlreg	NLNPEM
fixed	slope	1	1.001857	1.001232
random	intercept 1	1	0.9114 0.9114	0.9185
	(weight 1)	(0.1)	(0.00050) (0.09949)	(0.1)
	intercept 2	5	5.0257	5.0328
	(weight 2)	(0.1)	(0.1)	(0.1)
	intercept 3	10	-	10.048
	(weight 3)	(0.1)	-	(0.1)
	intercept 4	15	12.5442	14.8397 15.1058
	(weight 4)	(0.1)	(0.2)	(0.0192) (0.0807)
	intercept 5	20	19.9818	19.9312 20.0026
	(weight 5)	(0.1)	(0.1)	(0.0368) (0.0631)
	intercept 6	25	27.4750	24.9215 25.1181 25.1975
	(weight 6)	(0.1)	(0.2)	(0.0325) (0.0371) (0.0302)
	intercept 7	30	-	29.886
	(weight 7)	(0.1)	-	(0.1)
	intercept 8	35	35.0050	34.9582 35.2459
	(weight 8)	(0.1)	(0.1)	(0.0813) (0.0186)
	intercept 9	40	39.9516	39.6837 39.9624 40.4505
	(weight 9)	(0.1)	(0.1)	(0.0186) (0.0714) (0.0098)
	intercept 10	45	45.0017 45.0017	45.008
	(weight 10)	(0.1)	(0.09949) (0.000507)	(0.1)

Table 6: Estimates carried out by npmlreg and NLNPEM method on lin10I dataset, where intercept is considered as random, with 10 balanced groups.

Acknowledgement The case study in Section 4 is within the Strategic Program "Exploitation, integration and study of current and future health databases in Lombardia for Acute Myocardial Infarction" supported by "Ministero del Lavoro, della Salute e delle Politiche Sociali" and by "Direzione Generale Sanità - Regione Lombardia".

References

[1] Aitkin, M. (1996). A general maximum likelihood analysis of overdispersion in generalized linear models. *Statistics and Computing* **6**, 251-262.

[2] Aitkin, M. (1999) A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics* **55**, 117-128.

[3] Antic, J., Laffont, C.M., Chafaï, D., Concordet, D. (2009) Comparison of Nonparametric Methods in Nonlinear Mixed Effect Models. *Computational Statistics and Data Analysis* **53**, 3, 642-656.

[4] Davidian, M., Gallant, A.R. (1993) The Nonlinear Mixed Effects Model with a Smooth Random Effects Density. *Biometrika* **80**, 3, 475-488.

[5] De Lalla, C., Rinaldi, A., Montagna, D., Azzimonti, L., Bernardo, M.E., Sangalli, L.M., Paganoni, A.M., Maccario, R., Di Cesare Merlone, A., Zecca, M., Locatelli, F., Dellabona, P., Casorati, G. (2011) Invariant Natural Killer T-cell reconstitution in pediatric leukemia patients given HLA-haploidentical stem cell transplantation defines distinct CD4+ and CD4- subset dynamics and correlates with remission state. *The Journal of Immunology* **186**, 7, 4490-4499.

[6] Einbeck, J., Darnell, R., Hinde, J. (2009) npmlreg: Nonparametric maximum likelihood estimation for random effect models. [Online] http://CRAN.R-project.org/package=npmlreg

[7] Fox, J. (2002) Linear Mixed Models, *Appendix to An* `R` *and* `S-PLUS` *Companion to Applied Regression*

[8] Gallant, A.R. (1987) Nonlinear Statistical Models, Wiley, New York

[9] Gibbs, A.L. and Su, F.E. (2002) On choosing and bounding probability metrics. *International Statistical Review* **70**, 3, 419-435.

[10] Goldstein, H. (1991) Nonlinear Multilevel Models, with an Application to Discrete Response Data, *Biometrika* **78**, 1, 45-51

[11] Hox, J.J. (1995) Applied Multilevel Analysis, TT-Publikaties, Amsterdam

[12] Ieva, F., Paganoni, A.M., Secchi, P. (2010) Mining Administrative Health Databases for epidemiological purposes: a case study on Acute Myocardial Infarctions diagnoses, *Mox Report n.*

45/2010, Dipartimento di Matematica, Politecnico di Milano. [Online] http://mox.polimi.it/it/progetti/pubblicazioni/quaderni/45-2010.pdf

[13] Kuhn, E. and Lavielle, M. (2005) Maximum Likelihood estimation in nonlinear mixed effect models, *Computational Statistics and Data Analysis* **49**, 4, 1020-1038

[14] Lai, T.L. and Shih, M.C. (2003) Nonparametric estimation in nonlinear mixed-effects models, *Biometrika*. **90**, 1, 1-13

[15] Lindsay, B.G. (1983) The geometry of mixture likelihoods: a general theory, *The Annals of Statistics*. **11**, 1, 86-94

[16] Pinheiro, J.C., Bates, D.M. (2000). Mixed-Effects Models in S and S-Plus, Springer.

[17] R Development Core Team (2009) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. [Online] http://www.R-project.org

[18] Sheiner, L.B., Beal, S.L. (1980) Evaluation of methods for estimating population pharmacokinetic parameters. III. Monoexponential model: Routine clinical pharmacokinetic data. *Journal of Pharmacokinetics and Pharmacodynamics*. **11**, 3, 303-319.

[19] Schumitzky, A. (1991) Nonparametric EM Algorithms for estimating prior distributions, *Applied Mathematics and Computation* **45**, 2, 143-157.

[20] Vermunt, J.K. (2004) An EM Algorithm for the Estimation of Parametric and Nonparametric Hierarchical Models. *Statistica Neerlandica* **58**, 2: 220-233

[21] Walker, S. (1996) An EM Algorithm for Nonlinear Random Effects Models. *Biometrics* **52**, 3, 934-944

[22] Wolfinger, R. (1993) Laplace's approximation for nonlinear mixed models, *Biometrika* **80**, 4: 791-795

MOX Technical Reports, last issues

Dipartimento di Matematica "F. Brioschi",
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

22/2011 Azzimonti, L.; Ieva, F.; Paganoni, A.M.
Nonlinear nonparametric mixed-effects models for unsupervised classification

21/2011 Ambrosi, D.; Pezzuto, S.
Active stress vs. active strain in mechanobiology: constitutive issues

20/2011 Antonietti, P.F.; Houston, P.
Preconditioning high–order Discontinuous Galerkin discretizations of elliptic problems

19/2011 Passerini, T.; Sangalli, L.; Vantini, S.; Piccinelli, M.; Bacigaluppi, S.; Antiga, L.; Boccardi, E.; Secchi, P.; Veneziani, A.
An Integrated Statistical Investigation of the Internal Carotid Arteries hosting Cerebral Aneurysms

18/2011 Blanco, P.; Gervasio, P.; Quarteroni, A.
Extended variational formulation for heterogeneous partial differential equations

17/2011 Quarteroni, A.; Rozza, G.; Manzoni, A.
Certified Reduced Basis Approximation for Parametrized Partial Differential Equations and Applications

16/2011 Mesin, L; Ambrosi, D.
Spiral waves on a contractile tissue

15/2011 Argiento, R.; Guglielmi, A.; Soriano J.
A semiparametric Bayesian generalized linear mixed model for the reliability of Kevlar fibres

14/2011 Antonietti, P.F.; Mazzieri, I.; Quarteroni, A.; Rapetti, F.
Non-Conforming High Order Approximations for the Elastic Wave Equation

13/2011 Lombardi, M.; Parolini, N.; Quarteroni, A.; Rozza, G.
Numerical simulation of sailing boats: dynamics, FSI, and shape optimization