

# Estimation of linear mixed models with a mixture of distribution for the random-effects

Cécile Proust-Lima, Hélène Jacqmin-Gadda

► **To cite this version:**

Cécile Proust-Lima, Hélène Jacqmin-Gadda. Estimation of linear mixed models with a mixture of distribution for the random-effects. *Computer Methods and Programs in Biomedicine*, Elsevier, 2005, 78 (2), pp.165-73. inserm-00130037

**HAL Id: inserm-00130037**

**<https://www.hal.inserm.fr/inserm-00130037>**

Submitted on 12 Feb 2007

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Estimation of linear mixed models with a mixture of distribution for the random effects

Cécile Proust<sup>1</sup> , Hélène Jacqmin-Gadda<sup>1</sup>

4 janvier 2005

<sup>1</sup> Institut National de la Santé et de la Recherche Médicale, Equipe de biostatistique E0338, Université de Bordeaux2, 146 rue Léo Saignat, 33076 Bordeaux cedex, France.

**Abstract :** The aim of this paper is to propose an algorithm to estimate linear mixed model when random effect distribution is a mixture of Gaussians. This heterogeneous linear mixed model relaxes the classical Gaussian assumption for the random effects and, when used for longitudinal data, can highlight distinct patterns of evolution. The observed likelihood is maximized using a Marquardt algorithm instead of the EM algorithm which is frequently used for mixture models. Indeed, the EM algorithm is computationally expensive and does not provide good convergence criteria nor direct estimates of the variance of the parameters. The proposed method also allows to classify subjects according to the estimated profiles by computing posterior probabilities of belonging to each component. The use of heterogeneous linear mixed model is illustrated through a study of the different patterns of cognitive evolution in the elderly. HETMIXLIN is a free Fortran90 program available on the

web site : <http://www.isped.u-bordeaux2.fr>.

**keywords :** heterogeneous mixed model, mixture model, longitudinal data, Newton-Raphson like algorithm, cognitive ageing

**Corresponding author :** Cécile Proust, INSERM E0338, ISPED, Université de Bordeaux2, 146 rue Léo Saignat, 33076 Bordeaux cedex, France. Tel : (33) 5 57 57 11 36 ; Fax : (33) 5 56 24 00 81 ; e-mail : [cecile.proust@isped.u-bordeaux2.fr](mailto:cecile.proust@isped.u-bordeaux2.fr)

# 1 Introduction

Many longitudinal studies consist in assessing the changes over the time of a marker measured repeatedly on each participant. These analyses are generally performed with mixed models [1] which allow to take into account the within-subject correlation and the variability of the marker course between the subjects. However such a model is based on the strong assumption that the random effects are sampled from a single multivariate Gaussian distribution which means that the marker course is homogeneous among all the subjects.

To assess this assumption, Verbeke and Lesaffre [2] have proposed a mixed model with a mixture of multivariate Gaussians on the random effects. This heterogeneous linear mixed model allows to relax the normality assumption for the random effects and also allows to highlight distinct evolutions of the marker and classify the subjects according to these different patterns of evolution.

In the Verbeke and Lesaffre's work as in more recent papers [3,4], the mixed models with a mixture on the random effects distribution were estimated using an EM algorithm [5]. For instance, Spiessens and Verbeke [3] recently proposed a free SAS-macro (HETNLMIXED) using the EM algorithm and the NLMIXED procedure for the optimization in the M step. This SAS-macro is an extension of the SAS-macro HETMIXED which was developed earlier for estimating heterogeneous linear mixed models using the MIXED procedure [6]. To our knowledge, HETNLMIXED and its first version HETMIXED are the only free available programs developed for estimating heterogeneous mixed models. The first version HETMIXED was proved to be very slow and limited to small samples due to very large matrices handling and prohibitive computation; it will not be expanded in this work. HETNLMIXED was developed to reduce these computational problems and to allow estimation of

both linear and generalized linear models. However, in the linear case, this SAS-macro has the drawback of computing numerically an integral across the random effects while it has a closed form, and thus the macro is limited to a small number of random-effects. We have also observed convergence problems when using the macro with large samples except for very simple models.

Moreover, the EM algorithm, which is used in these macros, has some general drawbacks. In particular it does not have any good convergence criteria; the convergence is only built on a lack of progression of the likelihood or the parameter estimates [7]. Furthermore, the convergence is slow [8] and the EM algorithm does not provide direct estimates of the variance of the parameters. In the particular case of an heterogeneous mixed model, the M step also requires the estimation of an homogeneous mixed model which is computationally expensive.

Therefore the first aim of this paper is to propose a program for estimating more general heterogeneous linear mixed models suitable for large samples. The proposed program HETMIXLIN is written in Fortran90 and uses a direct maximization of the likelihood via a Marquardt optimization algorithm. The second objective of this paper is to illustrate the use of heterogeneous linear mixed model through a study of the different patterns of evolution in cognitive ageing.

## 2 Computational methods and theory

### 2.1 The heterogeneous linear mixed model

Let  $Y_i = (Y_{i1}, \dots, Y_{in_i})$  be the response vector for the  $n_i$  measurements of the subject  $i$  with  $i = 1, \dots, N$ . The linear mixed model [1] for the response vector  $Y_i$  is defined as :

$$Y_i = X_i\beta + Z_iu_i + \epsilon_i \quad (1)$$

$X_i$  is a  $n_i \times p$  design matrix for the  $p$ -vector of fixed effects  $\beta$ , and  $Z_i$  is a  $n_i \times q$  design matrix associated to the  $q$ -vector of random effects  $u_i$  which represents the subject specific regression coefficients. The errors  $\epsilon_i$  are assumed to be normally distributed with mean zero and covariance matrix  $\sigma^2 I_{n_i}$ , and are assumed to be independent from the vector of random effects  $u_i$ .

In an homogeneous mixed model [1],  $u_i$  is normally distributed with mean  $\mu$  and covariance matrix  $D$  i.e.

$$u_i \sim N(\mu, D) \quad (2)$$

In the heterogeneous mixed model [2-4],  $u_i$  is assumed to follow a mixture of  $G$  multivariate Gaussians with different means  $(\mu_g)_{g=1,G}$  and a common covariance matrix  $D$  i.e.

$$u_i \sim \sum_{g=1}^G \pi_g N(\mu_g, D) \quad (3)$$

Each component  $g$  of the mixture has a probability  $\pi_g$  and the  $(\pi_g)_{g=1,G}$  verify the following conditions :

$$0 \leq \pi_g \leq 1 \quad \forall g = 1, G \quad \text{and} \quad \sum_{g=1}^G \pi_g = 1 \quad (4)$$

In this work, we propose a slightly more general formulation of the model described in (1) in which the effect of some covariates may depend on the components of mixture and some of the random effects may have a common mean whatever the component of mixture. Thus, the  $X_i$  design matrix is split in  $X_{1i}$  associated with the vector  $\beta$  of fixed effects which are common to all the components and  $X_{2i}$  associated with the vectors  $\gamma_g$  of fixed effects which are specific to the components. The  $Z_i$  design matrix is also splitted in  $Z_{1i}$  associated with the vector  $v_i$  of random effects following a single Gaussian distribution and  $Z_{2i}$  associated with the vector  $u_i$  of random effects following a mixture of Gaussian distributions. The model is then

written as :

$$Y_i = X_{1i}\beta + \sum_{g=1}^G \pi_g X_{2i}\delta_g + Z_{1i}v_i + Z_{2i}u_i + \epsilon_i \quad (5)$$

where  $v_i \sim N(0, D_v)$  and  $u_i \sim \sum_{g=1}^G \pi_g N(\mu_g, D_u)$ ; given the component  $g$ , the conditional distribution of the vector  $\begin{pmatrix} v_i \\ u_i \end{pmatrix}$  is  $N\left(\begin{pmatrix} 0 \\ \mu_g \end{pmatrix}, D\right)$  with  $D = \begin{pmatrix} D_v & D_{vu} \\ D_{uv} & D_u \end{pmatrix}$ .

## 2.2 Likelihood

Following the previous works [3,4], we define  $w_{ig}$  the unobserved variable indicating if the subject  $i$  belongs to the component  $g$ . We have  $P(w_{ig} = 1) = \pi_g$ . The density for the vector  $y_i$  can then be written as :

$$f_i(y_i) = \sum_{g=1}^G \pi_g f(y_i | w_{ig} = 1) \quad (6)$$

Given  $w_{ig}$ ,  $y_i$  follows a linear mixed model, and the density  $f(y_i | w_{ig} = 1)$  denoted by  $\phi_{ig}$  is the multivariate Gaussian density with mean  $E_{ig}$  and covariance matrix  $V_i$  given by :

$$E_{ig} = E(Y_i | w_{ig} = 1) = X_{1i}\beta + X_{2i}\delta_g + Z_{2i}\mu_g$$

and (7)

$$V_i = Var(Y_i | w_{ig} = 1) = Z_i D Z_i' + \sigma^2 I_{n_i}$$

Let now  $\theta$  be the vector of the  $m$  parameters of the model.  $\theta$  contains  $\psi$  with  $\psi' = (\beta', (\delta_g)'_{g=1,G}, (\mu_g)'_{g=1,G}, Vec(D)', \sigma^2)$  and  $\pi$  the vector of the  $G - 1$  first component probabilities  $(\pi_g)_{g=1,G-1}$ . Note that  $\pi_G$  is entirely determined by  $\pi$  as  $1 - \sum_{g=1}^{G-1} \pi_g$ .  $Vec(D)$  represents the vector of the upper triangular elements of  $D$ . The estimates

of  $\theta$  are obtained as the vector  $\hat{\theta}$  that maximizes the observed log-likelihood :

$$\begin{aligned}
L(Y; \theta) &= \sum_{i=1}^N \ln(f_i(y_i)) \\
&= \sum_{i=1}^N \ln\left(\sum_{g=1}^G \pi_g \phi_{ig}(y_i)\right) \\
&= \sum_{i=1}^N -\frac{n_i}{2} \ln(2\pi) - \frac{1}{2} \ln(|V_i|) + \ln\left(\sum_{g=1}^G \pi_g e^{-\frac{1}{2}(Y_i - E_{ig})' V_i^{-1} (Y_i - E_{ig})}\right)
\end{aligned} \tag{8}$$

## 2.3 Estimation procedure

We propose to maximize directly the observed log-likelihood (8) using a modified Marquardt optimization algorithm [9], a Newton-Raphson like algorithm [10]. The diagonal of the Hessian at iteration  $k$ ,  $H^{(k)}$ , is inflated to obtain a positive definite matrix as :  $H^{*(k)} = (H_{ij}^{*(k)})$  with  $H_{ii}^{*(k)} = H_{ii}^{(k)} + \lambda[(1 - \eta)|H_{ii}^{(k)}| + \eta \text{tr}(H)]$  and  $H_{ij}^{*(k)} = H_{ij}^{(k)}$  if  $i \neq j$ . Initial values for  $\lambda$  and  $\eta$  are  $\lambda = 0.01$  and  $\eta = 0.01$ . They are reduced when  $H^*$  is positive definite and increased if not. The estimates  $\theta^{(k)}$  are then updated to  $\theta^{(k+1)}$  using the current modified Hessian  $H^{*(k)}$  and the current gradient of the parameters  $g(\theta^{(k)})$  according to the formula :

$$\theta^{(k+1)} = \theta^{(k)} - \alpha H^{*(k)-1} g(\theta^{(k)}) \tag{9}$$

where, if necessary,  $\alpha$  is modified to ensure that the log-likelihood is improved at each iteration.

To ensure that the covariance matrix  $D$  is positive, we maximize the log-likelihood on the non zero elements of  $U$ , the Cholesky factor of  $D$  (i.e.  $U'U = D$ ) [7]. Furthermore, to deal with the constraints on  $\pi$  (4) we use the transformed parameters  $(\gamma_g)_{g=1, G-1}$  with :



$$\gamma_g = \ln\left(\frac{\pi_g}{\pi_G}\right) \quad (10)$$

Standard errors of the elements of  $D$  and  $(\pi_g)_{g=1, G-1}$  are computed by the  $\Delta$ -method [11] while standard errors of the other parameters are directly computed using the inverse of the observed Hessian matrix.

The convergence is reached when the three following convergence criteria are satisfied :  $\sum_{j=1}^m (\theta_j^{(k)} - \theta_j^{(k-1)})^2 \leq \epsilon_a$ ,  $|L^{(k)} - L^{(k-1)}| \leq \epsilon_b$  and  $g(\theta^{(k)})' H^{(k)-1} g(\theta^{(k)}) \leq \epsilon_d$ . The default values are  $\epsilon_a = 10^{-5}$ ,  $\epsilon_b = 10^{-5}$  and  $\epsilon_d = 10^{-8}$ .

As the log-likelihood of a mixture model may have several maxima [8], we use a grid of initial values to find the global maximum. The multimodality of the log-likelihood in mixture models has been often discussed and some authors proposed different strategies to choose the set of initial values [12]. However, none of them seems to be optimal in a general way. We have observed, in our experience, that the results were mainly sensitive to initial values of  $(\pi_g)_{g=1, G-1}$  and  $(\mu_g)_{g=1, G}$  and less sensitive to the other parameters ( $Vec(U)$ ,  $\beta$  and  $\sigma$ ) for which estimates of the homogeneous mixed models were good initial values.

A mixture model is estimated with a fixed number of components  $G$ , otherwise the number of parameters in the model is unknown. To choose the right number of components, one has to estimate models with different values for  $G$  and select the best model according to a test or a criterion. Some works favor a bootstrap approach to approximate the asymptotic distribution of the likelihood ratio test between models with different number of components [13] but this approach is very heavy in particular for mixture models with random effects. Criteria such as Akaike's Information Criterion (AIC) [14] or Bayesian Information Criterion (BIC) [15] are often preferred. We use these selection criteria to select the optimal number of components.

## 2.4 A posteriori classification

After parameter estimation, mixture models allow to classify subjects according to the  $G$  components. The classification is based on the posterior probabilities  $(\pi_{ig})_{g=1,G}$  that the subject  $i$  follows each of the  $G$  components. Using  $\hat{\theta} = (\hat{\psi}', \hat{\pi}')$ , these probabilities are obtained by the Bayes theorem [2-4] as :

$$\hat{\pi}_{ig} = P(w_{ig} = 1 | Y_i, \hat{\theta}) = \frac{\hat{\pi}_g \phi_{ig}(\hat{\psi}, Y_i)}{\sum_{g=1}^G \hat{\pi}_g \phi_{ig}(\hat{\psi}, Y_i)} \quad (11)$$

We then assign to each subject  $i$  the component to which he has the highest probability  $(\pi_{ig})_{g=1,G}$  to belong.

## 3 Program description

The program requires two distinct input files : the data file described in appendix 1 and the parameter file named HETMIXLIN.inf which contains the information needed for the estimation of the model : the names of the data file and output files, the number of subjects, the description of the model (number of components  $G$ , dimension of the random effects, covariates  $X_1$ ,  $X_2$ ,  $Z_1$  and  $Z_2$  and covariance structure of  $D$ ) and the initial values of the parameters. An example of the parameter file is given in appendix 2.

The main output file gives the final log-likelihood, the AIC, the BIC, the convergence criteria, the number of iterations and the parameter estimates with the standard errors, the Wald statistics and the 95% confidence interval. The number of subjects classified in each component is also given.

Finally, another output file contains the posterior probabilities for each subject to belong to each class and the final class membership.

## 4 Applications

### 4.1 The height of schoolgirls

We consider the sample of 20 preadolescent schoolgirls introduced by Goldstein [16]. Verbeke *et al* [2] and Komárek *et al* [6] modelled the growth curves of their height according to age from 6 to 10. They showed in the homogeneous mixed model that the height course of girls differed significantly according to the category of height of their mother (small, medium and tall). Thus they used the heterogeneous linear mixed model without introducing the height of the mother in the model to try to highlight clusters with distinct growth curves among the girls. In this work, we compare the results obtained using our program to those obtained with the HETNLMIXED SAS-macro which uses the EM algorithm. The model is written as :

$$\text{Height}_{ij} = u_{0i} + u_{1i} \times \text{age}_{ij} + \epsilon_{ij} \quad (12)$$

where  $u_i = (u_{0i}, u_{1i})' \sim \sum_{g=1}^G \pi_g N(\mu_g, D)$  with  $\mu_g = (\mu_{0g}, \mu_{1g})'$  and  $\epsilon_{ij} \sim_{iid} N(0, \sigma^2)$

We fitted the heterogeneous linear mixed model for two and three components. An extract of the data file and the parameter file for the model with two components are presented in appendices 1 and 2. The results for the model with two components of mixture obtained with our program HETMIXLIN and the SAS-macro HETNLMIXED are shown in Table 1. The estimates obtained using the two methods are the same but a difference is observed in the standard error estimates; the standard errors estimates from HETNLMIXED seem bigger than those from HETMIXLIN. This difference in the standard error estimates from the two algorithms was also observed in the homogeneous case comparing HETNLMIXED, the MIXED procedure, the NLMIXED procedure and HETMIXLIN program. In the three latter programs,

standard errors are estimated by the inverse of the Hessian matrix which estimates the Fisher Information matrix [11] and led to the same standard error estimates. By contrast, HETNLMIXED uses an approximation of the Louis's method based on the product of the expectations of the gradient of the complete likelihood [17], the Louis's method [18] being itself an approximation of the observed Hessian matrix. This method appeared to overestimate standard errors in this small sample. However, in our experience, this approximation of the observed Hessian matrix seemed to be improved when the sample size increased. For instance, using a linear mixed model estimated on the 1,392 subjects of the PAQUID sample from section 4.2, the discrepancy was lower.

As the convergence of the two algorithms depends on the choice of the initial values, we fitted the model by the two approaches with the same grid of 32 sets of initial values. The sets differed on all kinds of parameters. HETNLMIXED provided the global maximum 11 times out of the 32 tries and our program found the global maximum 23 times. HETMIXLIN was also faster than the HETNLMIXED SAS-macro (several seconds compared with at least several minutes using a *Bi-Xeon 3,06 GHz 1024 MB RAM*).

The results for the model with three components of mixture are also shown in Table 1, but we cannot compare our results with those given by HETNLMIXED since it converges toward a non-positive definite  $\hat{D}$  matrix. Indeed, HETNLMIXED uses the NLMIXED procedure which does not constrain  $D$  to be positive definite.

## 4.2 Cognitive decline in the elderly

The second example illustrates the use of heterogeneous linear mixed models with our estimation method on a large data set as it can be encountered in epidemiological studies. The aim of this analysis is to describe, in a cohort of elderly subjects, the

heterogeneity of the evolution of the Mini Mental State Examination (MMSE), the most important psychometric test to evaluate dementia and cognitive impairment, and to compare the classification of subjects stemmed from the mixture model with the dementia diagnosis. The MMSE score ranges from 0 to a maximum of 30 points.

Data come from the French prospective cohort study PAQUID initiated in 1988 to study normal and pathological ageing [19]. The cohort includes 3,777 subjects of 65 years and older who lived at home in southwestern France at baseline. Subjects were interviewed at baseline and were seen again 1 (T1), 3 (T3), 5 (T5), 8 (T8) and 10 (T10) years after the baseline visit (T0). At each visit, a battery of psychometric tests was completed and a diagnosis of dementia was carried out. In this analysis, we excluded data from T0 because of a learning effect previously described for the cognitive tests between T0 and T1 [20]. We studied the evolution of the MMSE between T1 and T8 for subjects free of dementia till T5 and compared the estimated classification with the dementia diagnosis at T8 and then with the health status at T10. We excluded subjects not seen at T8 to ensure that we had their diagnosis at this visit. This leads to a sample of 1,392 subjects having between 1 and 4 measures of the MMSE between T1 and T8.

The model is a quadratic function of time adjusted on covariates associated with cognitive evolution in order to exclude heterogeneity introduced by known factors. The time ( $t_{ij}$  for subject  $i$  at visit  $j$ ) is the negative time between the measurement and the visit at T8 (time is zero for diagnosis time at T8). We model the square root of the number of errors to satisfy the normality assumption of the error terms. The model is written as :

$$Y_{ij} = \sqrt{30 - \text{MMSE}_{ij}} = \beta X_{ij} + u_{0i} + u_{1i}t_{ij} + u_{2i}t_{ij}^2 + \epsilon_{ij} \quad (13)$$

where  $X_{ij}$  is the vector of covariates for subject  $i$  at visit  $j$  including age, occupation,

educational level, living place and interactions with time for age and educational level;  $u_i = (u_{0i}, u_{1i}, u_{2i})' \sim \sum_{g=1}^G \pi_g N(\mu_g, D)$  with  $\mu_g = (\mu_{0g}, \mu_{1g}, \mu_{2g})'$  and  $\epsilon_{ij} \sim_{iid} N(0, \sigma^2)$ .

We fitted the heterogeneous linear mixed model with two components of mixture. The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) were largely improved compared with the homogeneous mixed model ( $\Delta AIC = 98.7$ ;  $\Delta BIC = 77.8$ ). Table 2 displays the parameter estimates for the homogeneous and the heterogeneous linear mixed models obtained with HETMIXLIN. The heterogeneous linear mixed model distinguished two different MMSE courses (Figure 1). First, a large class including 98% of the sample follows a linear evolution with a slight decline of 0.42 points ( $\sigma = 0.15$ ) per year. The second class including 2% of the sample follows a nonlinear evolution which speeds to fell down from the second follow-up to the end.

Then, we tried to evaluate if the cognitive profiles highlighted by the model were associated with dementia diagnosis. Among the 1,392 subjects, 26 (1.9%) are classified in the second component with the nonlinear decline (Table 3). Among them, 21 have a positive dementia diagnosis at T8. The predictive positive value (81%) and the specificity (99.6%) of this classification are high but the sensitivity is poor (31%) : 47 subjects among the 68 subjects diagnosed as demented at T8 are not detected by the model. These subjects are significantly less disabled ( $p < 0.0001$ ) than the other demented subjects (23% of disabled people vs. 76%) and have a significantly lower educational level ( $p = 0.001$ ) : 51% have no education or no diploma from primary school vs. 9.5% in the demented group detected in the model. All of the 5 subjects without dementia but classified in the declining group are disabled, 2 die in the two years and 1 has a positive dementia diagnosis at T10.

The association between cognitive ageing and educational level is an important

issue [21] : educational level could have a different effect on normal cognitive ageing and on the decline before dementia. As HETMIXLIN allows to specify distinct parameters for the covariates per component, we fitted the model where the interactions between educational level and time (educational level  $\times t$  and educational level  $\times t^2$ ) were different according to the components. The log-likelihood was not improved enough for significance ( $\Delta(-2L) = 3.4$ ;  $p = 0.17$ ). Thus the association between educational level and cognitive evolution appears to be similar in the two subpopulations. Moreover, the discrimination of demented people was not improved with this model.

Due to limitations of HETNLMIXED explained in section 1, we did not compare our approach with this program on the PAQUID data set. However, as Newton-Raphson algorithms have been criticized on their global convergence behavior compared with the EM algorithm [8], we compared the convergence performances of our program with those of an EM algorithm we developed in Fortran90. Thus the comparison was free from the limitations due to SAS environment and to the use of the NLMIXED procedure. The EM algorithm we developed uses a Marquardt optimization in the M-step (convergence criteria :  $\epsilon_a = 10^{-2}$ ,  $\epsilon_b = 10^{-2}$  and  $\epsilon_d = 10^{-3}$ ) and the global convergence is reached when two successive calculations of the likelihood differ less than  $10^{-8}$ . This algorithm was tested using the schoolgirls data. It was faster and converged more often than HETNLMIXED : the global maximum was reached 22 times out of the 32 tries with a mean computational time around 30 seconds (versus 11 times out of the 32 tries in at least several minutes for HETNLMIXED). On the PAQUID data set, HETMIXLIN and the EM algorithm we implemented led to the same parameter estimates. Among the 15 sets of initial values, the two programs provided the global maximum an equivalent number of times (9 times for HETMIXLIN vs. 10 times for the EM algorithm) but HETMIXLIN was much

faster : the CPU time was less than 10 minutes for HETMIXLIN and more than 2 hours for the EM algorithm.

## 5 Availability of the program and hardware specification

The program HETMIXLIN is written in Fortran90 and all the subroutines needed in the program are provided. The Fortran source code HETMIXLIN.f, an example of HETMIXLIN.inf, a documentation HETMIXLIN.pdf and the example data file for the schoolgirls are available at no charge on the web site : <http://www.isped.u-bordeaux2.fr>.

Two versions are provided on the web site : one for Unix and one for Windows. The version for Windows includes an executable file (a DOS application) and does not need any Fortran90 compiler whereas the version for Unix needs to be compiled. The Unix version has been tested using an Intel Fortran Compiler for Linux version 7 or 8, a Fortran90 Compaq compiler for Alpha and a Forte Developer 6 update 2 on Solaris SPARC. Examples of the compilation command are given in the documentation HETMIXLIN.pdf.

## 6 Conclusion

We proposed in this paper a Newton-Raphson like algorithm to estimate heterogeneous linear mixed models. The main advantages of Newton-Raphson like algorithms are the speed of convergence, the availability of good convergence criteria based on the derivatives of the likelihood and direct estimates of the variance of the parameters via the inverse of the Hessian matrix. Moreover, using a simple mo-



dification of the Marquardt algorithm, we ensure the monotonicity of the algorithm which is considered as a main advantage of the EM algorithm [8].

We compared our program HETMIXLIN with a SAS-macro developed by Spiessens *et al* using an EM algorithm. This SAS-macro allows to estimate heterogeneous generalized linear mixed models, but when the model is linear, this macro has the drawback of computing numerically an integral across the random effects while it has a closed form. Our algorithm HETMIXLIN allows to estimate more complex linear models (models with a larger number of mixture components, a larger number of random effects and more covariate effects depending on the mixture components) and is suitable for much larger samples. Moreover, it converges faster.

This paper also illustrates the usefulness of heterogeneous linear mixed models on a study about cognitive ageing. These models allow to highlight various evolution profiles taking covariates into account. The cross-classification of the groups defined by the model and clinical events in the next years enables to evaluate whether the cognitive profiles are associated with different clinical evolutions.

As a conclusion, we hope this work will improve the availability and the use of heterogeneous linear mixed models.

## References

- [1] N.M Laird, J.H. Ware, Random-effects models for longitudinal data, *Biometrics* 38 (1982) 963-74.
- [2] G. Verbeke, E. Lesaffre, A linear mixed-effects model with heterogeneity in the random-effects population, *JASA* 91 (1996) 217-21.
- [3] B. Spiessens, G. Verbeke, A. Komárek, A SAS-macro for the classification of longitudinal profiles using mixtures of normal distributions in nonlinear and generalized linear models, <http://www.med.kuleuven.ac.be/biostat/research/software.htm> (2002).
- [4] B. Muthén, K. Shedden, Finite mixture modeling with mixture outcomes using a EM algorithm, *Biometrics* 55 (1999) 463-9.
- [5] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via EM algorithm (with discussion), *J. Roy. Statist. Soc. Ser. B* 39 (1977) 1-38.
- [6] A. Komárek, A SAS-macro for linear mixed models with a finite normal mixture as random-effects distribution, <http://www.med.kuleuven.ac.be/biostat/research/software.htm> (2001).
- [7] M.J. Linstrom, D.M. Bates, Newton-Raphson and EM algorithms for linear mixed models for repeated-measures data, *JASA* 83 (1988) 1014-22.
- [8] R.A. Redner, H.F. Walker, Mixture densities, maximum likelihood and the EM algorithm, *SIAM Review* 26 (1984) 195-239.
- [9] D. Marquardt, An algorithm for least-squares estimation of nonlinear parameters, *SIAM J. Appl. Math.* 11 (1963) 431-41.
- [10] R. Fletcher, *Practical methods of optimization*, second edition, chap. 3 (John Wiley & Sons 2000).

- [11] K. Knight, *Mathematical Statistics*, chapt. 3;5 (Chapman & Hall/CRC 2000).
- [12] D. Karlis, E. Xekalaki, Choosing initial values for the EM algorithm for finite mixtures, *Comput. Statist. Data Anal.* 41 (2003) 577-90.
- [13] P. Schlattmann, Estimating the number of components in a finite mixture model : the special case of homogeneity, *Comput. Statist. Data Anal.* 41 (2003) 441-51.
- [14] H. Akaike, A new look at the statistical model identification, *IEEE trans. automat. control* 19 (1974) 716-23.
- [15] G. Schwarz, Estimating the dimension of a model, *The Annals of Statistics* 6 (1978) 461-64.
- [16] H. Goldstein, *The design and analysis of longitudinal studies*, London : Academic press (1979).
- [17] G.J. McLachlan, T. Krishnan, *The EM algorithm and extensions*, John Wiley & Sons (1997).
- [18] T.A. Louis, Finding the Observed Information Matrix when Using the EM algorithm, *J. Roy. Statis. Soc. Ser. B* 44 (1982) 226-33.
- [19] L. Letenneur, D. Commenges, J.F. Dartigues, P. Barberger-Gateau, Incidence of dementia and Alzheimer's disease in elderly community residents of southwestern France, *Int. J. Epidemiol.* 23 (1994) 1256-61.
- [20] H. Jacqmin-Gadda, C. Fabrigoule, D. Commenges, J.F. Dartigues, A 5-year longitudinal study of the Mini Mental State Examination in normal aging, *Am. J. Epidemiol.* 145 (1997) 498-506.
- [21] L. Letenneur, V. Gilleron, D. Commenges, C. Helmer, J.M. Orgogozo, J.F. Dartigues, Are sex and educational level independent predictors of dementia and Alzheimer's disease? Incidence data from the PAQUID project, *J. Neurol. Neuro-*

surg. Psychiatry 66 (1999) 177-83.

## Appendix 1 : Extract of the Schoolgirl data file

(two first subjects)

1	← identification number of the unit (subject)
5	← number of measures
111 116.4 121.7 126.3 130.5	← raw vector of the $n_i$ responses
1 1 1 1 1	← raw vector of the first covariate
6 7 8 9 10	← raw vector of the second covariate
2	← identification number of the next unit (subject)
5	← number of measures
110 115.8 121.5 126.6 131.4	← raw vector of the $n_i$ responses
1 1 1 1 1	← raw vector of the first covariate
6 7 8 9 10	← raw vector of the second covariate
3	← identification number of the next unit (subject)
...	

## Appendix 2 : Example of the parameter file

An example of HETMIXLIN.inf used in the application about the height of schoolgirls is given below. The user should notice that each asked piece of information is preceded by a line summing it up.

```
-> Filename for the data
schoolgirls.txt
-> Filename for the output
girls.out
-> Title of the procedure (in inverted commas)
```

'G=2 : school girls '

-> Number of units (subjects)

20

-> Number of mixture components (G) and, if and only if  $G > 1$ , the initial values for the  $G-1$  first component probabilities below and the filename for the posterior probabilities below again

2

0.5

p.out

-> Number of explanatory variables (including the intercept) in the data file

2

-> Indicator that the explanatory variable is in the model (1 if present 0 if not)

1 1

-> Indicator of random effect for each variable in the model (variables included in Z1 or Z2)

1 1

-> Indicator of mixture for each variable in the model (variables included in X2 or Z2)

1 1

-> Initial values for fixed effects. First, initial values for common fixed effects (without mixture) in the same order as in the datafile, then initial values for the covariates with a mixture (G values per covariate).

ex : b1 b3 b21 b22 b23 b41 b42 b43 for a mixture on the second and the fourth covariate and  $G=3$

86 80 5 7

-> Indicator of the random-effect covariance matrix structure (0 if unstructured matrix / 1 if diagonal matrix)

0

-> Initial values for the variance-covariance parameters of the random effects (1/2 superior matrix column by column)

3 1 1

-> Initial value for the variance of the independent Gaussian errors

1

Table 1 : Estimates and standard-errors of the heterogeneous linear mixed model with two components of mixture for the height of schoolgirls using HETMIXLIN (the proposed direct maximisation using a Marquardt algorithm) and HETNLMIXED (Spiessens *et al* SAS-macro using an EM algorithm) and estimates of the heterogeneous linear mixed model with three components of mixture using HETMIXLIN



Parameter	$G = 2$		$G = 2$		$G = 3$	
	HETNLMIXED		HETMIXLIN		HETMIXLIN	
	Estimate	SE*	Estimate	SE**	Estimate	SE**
$\pi_1$	0.68	0.14	0.68	0.12	0.50	0.18
$\pi_2$	0.32		0.32		0.30	0.11
$\pi_3$					0.20	
$\mu_{01}$	82.8	1.12	82.8	0.91	84.2	1.18
$\mu_{11}$	5.38	0.091	5.38	0.086	5.32	0.10
$\mu_{02}$	81.9	2.01	81.9	1.52	81.7	1.12
$\mu_{12}$	6.44	0.18	6.44	0.15	6.47	0.12
$\mu_{03}$					79.4	2.19
$\mu_{13}$					5.60	0.20
$var(u_{0i})$	6.47	4.94	6.47	3.13	3.50	2.39
$cov(u_{0i}, u_{1i})$	0.13	0.40	0.13	0.35	0.32	0.13
$var(u_{1i})$	0.034	0.056	0.034	0.030	0.030	0.024
$\sigma$	0.69	0.10	0.69	0.063	0.68	0.06
-2L				166.67		165.94
AIC				351.35		355.87
BIC				360.32		367.82

\* Standard Errors obtained using Louis's method

\*\* Standard Errors obtained using the inverse of the Hessian matrix and the  $\Delta$ -method for the component probabilities and the variance parameters

Table 2 : Estimates of the homogeneous linear mixed model and the heterogeneous linear mixed model with two components of mixture obtained with HETMIXLIN program for the MMSE evolution adjusted on age, occupation, educational level, living place and interaction with time for age and educational level.

Parameter	Homogeneous model ( $G = 1$ )		Heterogeneous model ( $G = 2$ )	
	Estimate	SE	Estimate	SE
$\pi_1$	1		0.98	0.005
$\pi_2$			0.02	
$\mu_{01}$	-1.92	0.32	-1.51	0.31
$\mu_{11}$	-0.57	0.16	-0.42	0.16
$\mu_{21}$	-0.042	0.018	-0.028	0.024
$\mu_{02}$			0.89	0.36
$\mu_{12}$			0.46	0.19
$\mu_{22}$			0.055	0.028
$var(u_{0i})$	0.43	0.025	0.32	0.024
$cov(u_{0i}, u_{1i})$	0.074	0.010	0.034	0.0095
$var(u_{1i})$	0.066	0.014	0.029	0.013
$cov(u_{0i}, u_{2i})$	0.042	0.0073	0.028	0.0069
$cov(u_{1i}, u_{2i})$	0.052	0.010	0.038	0.010
$var(u_{2i})$	0.073	0.016	0.060	0.016
$\sigma$	0.45	0.009	0.45	0.009
-L		4662.9		4609.5
AIC		9367.7		9269.0
BIC		9477.8		9400.0

Table 3 : Relationship between classification stemmed from the heterogeneous linear mixed model with two components and dementia diagnosis at T8.

Dementia diagnosis at T8	Classification		Total
	Linear class	Nonlinear class	
Positive	47	21	68
Negative	1319	5	1324
Total	1366	26	1392

Figure 1 : Mean curves of the MMSE between T1 and T8 for the heterogeneous linear mixed model with two components of mixture for a 70-year-old worker subject with no education and living in Dordogne.

Solid line : first component with a probability of 98%

Dashed line : second component with a probability of 2%

