

## ON A FITTING OF A LINEAR MIXED MODEL WITH A FINITE NORMAL MIXTURE AS RANDOM-EFFECTS DISTRIBUTION

ARNOŠT KOMÁREK, GEERT VERBEKE

**ABSTRACT.** This paper describes an approximate method to compute maximum likelihood estimates of the parameters in the linear mixed model with a finite normal mixture as random-effects distribution. The proposed method uses an EM algorithm with an approximate M step which can be performed using procedures designed to fit a common linear mixed model. This approach enables, among others, to include easily various covariance structures of the residuals and random effects in the model. The suggested method has been implemented as a SAS macro which is publicly available on the Internet.

**Abstrakt.** Článek se zabývá přibližným postupem pro výpočet maximálně věrohodných odhadů v lineárním smíšeném modelu s normální směsí v rozdělení náhodných efektů. Navržená metoda využívá EM algoritmu, jehož M část je založena na přibližné maximalizaci optimalizované funkce a lze k ní využít procedur určených k výpočtu odhadů běžného lineárního smíšeného modelu. Výhodou tohoto přístupu je mj. možnost snadného zahrnutí rozmanitých struktur pro varianční matice reziduí a náhodných efektů. Popsaná metoda byla následně implementována jako makro programu SAS, které lze volně získat na síti Internet.

### 1. INTRODUCTION

A linear mixed model is a frequently used tool for describing longitudinal continuous data. Its random effects are usually assumed to be normally distributed. Unfortunately, this basic assumption can very often be violated. This will occur, for example, if an important categorical covariate is omitted from the fixed part of the model. That is why Verbeke and Lesaffre (1996) proposed to assume that random effects are distributed according to a finite normal mixture. The advantage of this approach is, among others, the fact that many continuous distributions can be well approximated by a finite normal mixture illustrating that the proposed model is generally applicable. On the other hand, a big disadvantage is a lack of available computational tools to fit such models in practice.

In this article, we concentrate on describing an approximate method to compute maximum likelihood estimates of the linear mixed model with a finite normal mixture as random effects distribution that can be quite easily implemented using common software for the linear mixed models.

After defining the model in Section 2, we show in Section 3 how the estimates can be computed using the EM algorithm and how the most difficult part of it – the M step can be approximately performed by tools designed for the classical linear mixed model.

---

1991 *Mathematics Subject Classification.* Primary 62J99; Secondary 62H30.

*Key words and phrases.* Finite normal mixture, linear mixed model, maximum likelihood estimates, random-effects distribution, repeated measurements.

The research has been partially supported by the project CEZ: MSM 113200008.

## 2. MODEL FORMULATION

This section introduces briefly the concept of the linear mixed model with a finite normal mixture as random-effects distribution that was proposed by Verbeke and Lesaffre (1996) and also described by Verbeke and Molenberghs (2000, Chapter 12). In accordance to the terminology used by them, the linear mixed model with a finite normal mixture as random effects distribution will be further called as the heterogeneity linear mixed model. It can be seen as an extension of the classical linear mixed model which will be called the homogeneity linear mixed model.

Let the random variable  $Y_{ik}$  denote the (possibly transformed) response of interest, for the  $i$ th individual measured at time  $t_{ik}$ ,  $i = 1, \dots, N$ ,  $k = 1, \dots, n_i$ , and let  $\mathbf{Y}_i$  be the  $n_i$ -dimensional vector of all repeated measurements for the  $i$ th subject, that is,  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^T$ . The heterogeneity linear mixed model starts from similar relationship as the homogeneity model, that is from

$$(1) \quad \mathbf{Y}_i = (\mathbb{X}_i \quad \mathbb{Z}_i) \begin{pmatrix} \boldsymbol{\beta}^F \\ \boldsymbol{\beta}^R \end{pmatrix} + \mathbb{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i,$$

where  $\mathbb{X}_i$  and  $\mathbb{Z}_i$  are  $(n_i \times p)$ , respectively  $(n_i \times q)$  matrices of known covariates, modeling how the response evolves over time for the  $i$ th subject. Further,  $\boldsymbol{\beta}^F$  and  $\boldsymbol{\beta}^R$  are  $p$ -dimensional, respectively  $q$ -dimensional vectors of unknown regression parameters. Variables  $\mathbf{b}_i$  are subject-specific  $q$ -dimensional random effects, and  $\boldsymbol{\varepsilon}_i$  is  $n_i$ -dimensional vector of residual components  $\varepsilon_{ik}$ ,  $k = 1, \dots, n_i$ . All  $\boldsymbol{\varepsilon}_i$  are assumed to be independent and normally distributed with mean vector zero and covariance matrix  $\Sigma_i$ .

We have just described the part of the heterogeneity model that is the same as for the homogeneity model. The former one differs from the latter one in assumptions on subject-specific effects  $\mathbf{b}_i$ . They are assumed to be independent by both models. The homogeneity model considers them as normally distributed with mean vector zero and covariance matrix  $\mathbb{D}$ . The heterogeneity model is obtained by replacing this distributional assumption by a mixture of a prespecified number  $g$  of  $q$ -dimensional normal distributions with mean vectors  $\boldsymbol{\mu}_j$  and covariance matrices  $\mathbb{D}$ , i.e.

$$(2) \quad \mathbf{b}_i \sim \sum_{j=1}^g \pi_j N(\boldsymbol{\mu}_j, \mathbb{D}),$$

with  $\sum_{j=1}^g \pi_j = 1$ . More general case assumes different covariance matrices  $\mathbb{D}_1, \dots, \mathbb{D}_g$  for each component of the mixture. But this can lead to infinitely large likelihood. In order to avoid numerical problems in the estimating procedure, we will assume  $\mathbb{D}_1 = \dots = \mathbb{D}_g = \mathbb{D}$ .

Vectors  $\mathbf{W}_i = (W_{i1}, \dots, W_{ig})^T$  can now be defined as follows. The term  $W_{ij} = 1$  if  $\mathbf{b}_i$  is sampled from the  $j$ th component of the mixture and 0 otherwise,  $j = 1, \dots, g$ . The distribution of  $\mathbf{W}_i$  is then described by

$$P(W_{ij} = 1) = E(W_{ij}) = \pi_j,$$

which is called *the prior probability* to be sampled from component  $j$ . Expected values of  $\mathbf{b}_i$  can then be easily obtained as

$$E(\mathbf{b}_i) = E(E[\mathbf{b}_i | \mathbf{W}_i]) = E \left( \sum_{j=1}^g \boldsymbol{\mu}_j W_{ij} \right) = \sum_{j=1}^g \pi_j \boldsymbol{\mu}_j.$$

Expectation of the response is then

$$E(\mathbf{Y}_i) = E(\mathbb{X}_i\boldsymbol{\beta}^F + \mathbb{Z}_i\boldsymbol{\beta}^R + \mathbb{Z}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i) = \mathbb{X}_i\boldsymbol{\beta}^F + \mathbb{Z}_i\boldsymbol{\beta}^R + \mathbb{Z}_i \sum_{j=1}^g \pi_j \boldsymbol{\mu}_j.$$

The homogeneity model usually assumes  $E(\mathbf{Y}_i) = \mathbb{X}_i\boldsymbol{\beta}^F + \mathbb{Z}_i\boldsymbol{\beta}^R$ . It is quite desirable to keep this property even for the heterogeneity model and therefore the additional constraint

$$(3) \quad \sum_{j=1}^g \pi_j \boldsymbol{\mu}_j = \mathbf{0}$$

is needed.

The model (1) with assumptions (2) can also be rewritten as a following hierarchical Bayes model

$$(4) \quad \begin{aligned} \mathbf{Y}_i | \mathbf{b}_i &\sim N(\mathbb{X}_i\boldsymbol{\beta}^F + \mathbb{Z}_i\boldsymbol{\beta}^R + \mathbb{Z}_i\mathbf{b}_i, \Sigma_i), \\ \mathbf{b}_i | \boldsymbol{\mu} &\sim N(\boldsymbol{\mu}, \mathbb{D}), \\ \boldsymbol{\mu} &\in \{\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_g\}, \quad \text{with } P(\boldsymbol{\mu} = \boldsymbol{\mu}_j) = \pi_j. \end{aligned}$$

This expression might be useful when the heterogeneity model is going to be used for classification of individual profiles into one of  $g$  populations. The underlying data generating mechanism can be viewed as a two step process. First, the population is chosen and second, response is generated according to the chosen population. In practice, one can wish to reveal the first step of this mechanism and to try to classify an individual with observed response vector  $\mathbf{Y}$  into one of the populations.

### 3. ESTIMATION OF THE HETEROGENEITY MODEL

**3.1. The General Concept of the Estimation Procedure.** Estimation of unknown parameters of the heterogeneity model is based on a marginal distribution of the observations  $\mathbf{Y}_i$ . Under (1) and (2) This distribution can easily be found to be given by

$$\mathbf{Y}_i \sim \sum_{j=1}^g \pi_j N(\mathbb{X}_i\boldsymbol{\beta}^F + \mathbb{Z}_i\boldsymbol{\beta}^R + \mathbb{Z}_i\boldsymbol{\mu}_j, \mathbb{V}_i), \quad \text{with } \mathbb{V}_i = \mathbb{Z}_i\mathbb{D}\mathbb{Z}_i^T + \Sigma_i.$$

Let  $\boldsymbol{\pi}$  be the vector of component probabilities (i.e.  $\boldsymbol{\pi}^T = (\pi_1, \dots, \pi_g)$ ) and let  $\boldsymbol{\gamma}$  be the vector of all other unknown parameters (i.e.  $\boldsymbol{\beta}^F, \boldsymbol{\beta}^R$ , components of matrices  $\mathbb{D}$  and  $\Sigma_i$ ). Further, let  $\boldsymbol{\theta}^T = (\boldsymbol{\pi}^T, \boldsymbol{\gamma}^T)$  denote the vector of all unknown parameters that are to be estimated. Method of maximum likelihood can be used to find requested estimates. The likelihood function corresponding to the marginal distribution of the observations  $\mathbf{Y}_i$  is of the form

$$(5) \quad L^*(\boldsymbol{\theta} | \mathbf{y}) = \prod_{i=1}^N \left\{ \sum_{j=1}^g \pi_j f_{ij}(\mathbf{y}_i | \boldsymbol{\gamma}) \right\}.$$

where  $\mathbf{y}^T = (\mathbf{y}_1^T, \dots, \mathbf{y}_N^T)$  is the vector containing all observed response values and  $f_{ij}$  is the density of an  $n_i$ -dimensional normal distribution  $N(\mathbb{X}_i\boldsymbol{\beta}^F + \mathbb{Z}_i\boldsymbol{\beta}^R + \mathbb{Z}_i\boldsymbol{\mu}_j, \mathbb{V}_i)$ .

Note that the likelihood function (5) is invariant under the  $g!$  possible permutations of the mean vectors and corresponding probabilities of the components of

the mixture. However, this lack of identifiability can easily be overcome by imposing some constraint on the parameters. For example, the constraint

$$(6) \quad \pi_1 \geq \pi_2 \geq \dots \geq \pi_g$$

suggested by Aitkin and Rubin (1985) can be used. The likelihood is then maximized without the restriction, and the component labels are permuted afterward to achieve (6).

The log-likelihood function corresponding to the likelihood (5) is then

$$l^*(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^N \log \left\{ \sum_{j=1}^g \pi_j f_{ij}(\mathbf{y}_i|\boldsymbol{\gamma}) \right\}.$$

It is quite difficult to maximize this function and the EM algorithm introduced by Dempster, Laird and Rubin (1977) can be used to compute the desired estimates. The response vectors  $\mathbf{Y}_i$  along with the (unobserved) population indicators  $\mathbf{W}_i$  can be seen as complete data whereas the vectors  $\mathbf{Y}_i$  alone can be viewed as incomplete data since information containing population pertinence is missing. The likelihood function (5) corresponds then to the incomplete data. The likelihood function that would have been obtained if values  $\mathbf{w}_i = (w_{i1}, \dots, w_{ig})^T$  of population indicators  $\mathbf{W}_i$  had been observed equals

$$(7) \quad L(\boldsymbol{\theta}|\mathbf{y}, \mathbf{w}) = \prod_{i=1}^N \prod_{j=1}^g \{\pi_j f_{ij}(\mathbf{y}_i|\boldsymbol{\gamma})\}^{w_{ij}}$$

where  $\mathbf{w}^T = (\mathbf{w}_1^T, \dots, \mathbf{w}_N^T)$  is the vector containing all hypothetically observed population indicators. The log-likelihood function corresponding to (7) has then the more attractable form

$$l(\boldsymbol{\theta}|\mathbf{y}, \mathbf{w}) = \sum_{i=1}^N \sum_{j=1}^g w_{ij} \{\log \pi_j + \log f_{ij}(\mathbf{y}_i|\boldsymbol{\gamma})\}.$$

Maximizing  $l(\boldsymbol{\theta}|\mathbf{y}, \mathbf{w})$  with respect to  $\boldsymbol{\theta}$  yields estimates which depend on the unobserved ("missing") indicators  $\mathbf{w}$ . The EM algorithm offers a solution to this problem by maximizing the expected value of  $l(\boldsymbol{\theta}|\mathbf{y}, \mathbf{w})$ , rather than  $l(\boldsymbol{\theta}|\mathbf{y}, \mathbf{w})$  with respect to  $\boldsymbol{\theta}$ , where the expectation is taken over all unobserved  $w_{ij}$ . The conditional expectation of  $l(\boldsymbol{\theta}|\mathbf{y}, \mathbf{w})$ , given the observed data vector  $\mathbf{y}$ , is calculated within the E step (expectation step) of each iteration of the EM algorithm. The obtained expected log-likelihood function is then maximized within the M step (maximization step) of the algorithm.

Let  $\boldsymbol{\theta}^{(t)}$  be the current estimate for  $\boldsymbol{\theta}$ , and  $\boldsymbol{\theta}^{(t+1)}$  stands for the updated estimate, obtained from one further iteration of the EM algorithm. The following E and M steps have to be followed to compute the updated estimate.

**The E step.** The conditional expectation

The conditional expectation of  $l(\boldsymbol{\theta}|\mathbf{y}, \mathbf{w})$ , given the observed data vector  $\mathbf{y}$  is given by

$$(8) \quad \begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) &= E[l(\boldsymbol{\theta}|\mathbf{y}, \mathbf{w})|\mathbf{y}, \boldsymbol{\theta}^{(t)}] \\ &= \sum_{i=1}^N \sum_{j=1}^g p_{ij}(\boldsymbol{\theta}^{(t)}) \{\log \pi_j + \log f_{ij}(\mathbf{y}_i|\boldsymbol{\gamma})\}. \end{aligned}$$

The terms  $p_{ij}(\boldsymbol{\theta}^{(t)})$  are called *the posterior probabilities* for the  $i$ th individual to belong to the  $j$ th component of the mixture and can easily be computed using Bayes' theorem as

$$\begin{aligned} p_{ij}(\boldsymbol{\theta}^{(t)}) &= E[W_{ij} | \mathbf{y}_i, \boldsymbol{\theta}^{(t)}] = P(W_{ij} = 1 | \mathbf{y}_i, \boldsymbol{\theta}^{(t)}) = \\ &= \frac{\pi_j^{(t)} f_{ij}(\mathbf{y}_i | \boldsymbol{\gamma}^{(t)})}{\sum_{k=1}^g \pi_k^{(t)} f_{ik}(\mathbf{y}_i | \boldsymbol{\gamma}^{(t)})}. \end{aligned}$$

Note that the posterior probabilities  $p_{ij}(\hat{\boldsymbol{\theta}})$  evaluated in the estimate  $\hat{\boldsymbol{\theta}}$  of the vector  $\boldsymbol{\theta}$  can be used for classification or clustering of individual profiles in the sense that the  $i$ th individual is classified into the  $k$ th component for which  $\max_{j=1, \dots, g} p_{ij}(\hat{\boldsymbol{\theta}}) = p_{ik}(\hat{\boldsymbol{\theta}})$ .

**The M step.** The maximization

The objective function  $Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(t)})$  has to be maximized with respect to  $\boldsymbol{\theta}$  to get the updated estimate  $\boldsymbol{\theta}^{(t+1)}$ . Expression (8) is the sum of two terms as indicated below.

$$Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(t)}) = Q_1(\boldsymbol{\pi} | \boldsymbol{\theta}^{(t)}) + Q_2(\boldsymbol{\gamma} | \boldsymbol{\theta}^{(t)}),$$

where

$$\begin{aligned} Q_1(\boldsymbol{\pi} | \boldsymbol{\theta}^{(t)}) &= \sum_{i=1}^N \sum_{j=1}^g p_{ij}(\boldsymbol{\theta}^{(t)}) \log \pi_j, \\ Q_2(\boldsymbol{\gamma} | \boldsymbol{\theta}^{(t)}) &= \sum_{i=1}^N \sum_{j=1}^g p_{ij}(\boldsymbol{\theta}^{(t)}) \log f_{ij}(\mathbf{y}_i | \boldsymbol{\gamma}). \end{aligned}$$

The first term depends only on the parameter  $\boldsymbol{\pi}$ , the second one only on the parameter  $\boldsymbol{\gamma}$ . Hence, it is possible to maximize each of these terms separately to find a maximum of  $Q$ . The function  $Q_1$  can be easily found to be maximized for

$$\pi_j^{(t+1)} = \frac{1}{N} \sum_{i=1}^N p_{ij}(\boldsymbol{\theta}^{(t)}).$$

In fact, these estimates are equal to an average of posterior probabilities for all subjects belonging to a given population.

Unfortunately, the function  $Q_2$  cannot be maximized analytically as the first one. In the next section, it will be derived how an approximate optimization of  $Q_2$  can be obtained using the common software for fitting the homogeneity linear mixed models, such as the SAS procedure PROC MIXED or the R/SPLUS function lme.

### 3.2. How to Maximize $Q_2$ – the Second Part of the Objective Function.

The function

$$(9) \quad Q_2(\boldsymbol{\gamma} | \boldsymbol{\theta}^{(t)}) = \sum_{i=1}^N \sum_{j=1}^g p_{ij}(\boldsymbol{\theta}^{(t)}) \log f_{ij}(\mathbf{y}_i | \boldsymbol{\gamma})$$

is to be maximized with respect to  $\boldsymbol{\gamma}$ . Maximization of (9) with respect to  $\boldsymbol{\gamma}$  is equivalent to maximization of

$$A \cdot Q_2(\boldsymbol{\gamma} | \boldsymbol{\theta}^{(t)}) = \sum_{i=1}^N \sum_{j=1}^g A \cdot p_{ij}(\boldsymbol{\theta}^{(t)}) \log f_{ij}(\mathbf{y}_i | \boldsymbol{\gamma})$$

for an arbitrary positive constant  $A$ . Further, numbers  $A \cdot p_{ij}(\boldsymbol{\theta}^{(t)})$  can be arbitrarily close to integers by choosing  $A$  sufficiently large. In practice, their rounded values can be used to approximate the function  $A \cdot Q_2(\boldsymbol{\gamma}|\boldsymbol{\theta}^{(t)})$ . Let  $a_{ij}(\boldsymbol{\theta}^{(t)})$  denote integers such that

$$a_{ij}(\boldsymbol{\theta}^{(t)}) \doteq A \cdot p_{ij}(\boldsymbol{\theta}^{(t)})$$

and let

$$(10) \quad Q_2^A(\boldsymbol{\gamma}|\boldsymbol{\theta}^{(t)}) = \sum_{i=1}^N \sum_{j=1}^g a_{ij}(\boldsymbol{\theta}^{(t)}) \log f_{ij}(\mathbf{y}_i|\boldsymbol{\gamma}).$$

The function  $Q_2^A(\boldsymbol{\gamma}|\boldsymbol{\theta}^{(t)})$  can be interpreted as the log-likelihood function for the homogeneity linear mixed model based on observations from  $\sum_{i=1}^N \sum_{j=1}^g a_{ij}(\boldsymbol{\theta}^{(t)})$  mutually independent individuals. Note that the  $i$ th response vector  $\mathbf{Y}_i$  from the original data set appears  $\sum_{j=1}^g a_{ij}(\boldsymbol{\theta}^{(t)})$  times in a data set which corresponds to the desirable homogeneity model. At the same time, the marginal distribution of  $a_{ij}(\boldsymbol{\theta}^{(t)})$  response vectors  $\mathbf{Y}_i$  out of their  $\sum_{j=1}^g a_{ij}(\boldsymbol{\theta}^{(t)})$  replications follows the  $n_i$ -dimensional normal distribution  $N(\mathbb{X}_i\boldsymbol{\beta}^F + \mathbb{Z}_i\boldsymbol{\beta}^R + \mathbb{Z}_i\boldsymbol{\mu}_j, \mathbb{V}_i)$ , with  $\mathbb{V}_i = \mathbb{Z}_i\mathbb{D}\mathbb{Z}_i^T + \Sigma_i$ . At this moment, common software for homogeneity linear mixed models is able to compute updated approximate estimates of  $\boldsymbol{\gamma}$ . The higher the value of  $A$  is used, the better the approximation is obtained. One has to take into account only present computational possibilities.

When implementing this method, one also has to consider the constraint (3) of the form  $\sum_{j=1}^g \pi_j \boldsymbol{\mu}_j = \mathbf{0}$  that was exposed to the population means at the beginning of this section. Fortunately, it is not too difficult to ensure that this constraint is satisfied since the originally restricted  $q$ -dimensional parameters  $\boldsymbol{\beta}^R, \boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_g$  can be replaced by unrestricted  $q$ -dimensional parameters  $\boldsymbol{\delta}_1, \dots, \boldsymbol{\delta}_g$  using the relationship

$$\boldsymbol{\delta}_j = \boldsymbol{\beta}^R + \boldsymbol{\mu}_j, \quad j = 1, \dots, g.$$

In fact, parameters  $\boldsymbol{\delta}_j$  express real population means, whereas parameters  $\boldsymbol{\mu}_j$  represent the contrasts between a population mean and the overall mean  $\boldsymbol{\beta}^R$ . Restriction (3) also gives the way to compute  $\boldsymbol{\beta}^R$  from  $\boldsymbol{\delta}_1, \dots, \boldsymbol{\delta}_g$ , that is

$$\boldsymbol{\beta}^R = \sum_{j=1}^g \pi_j \boldsymbol{\delta}_j.$$

**3.3. Empirical Bayes Inference.** The random effects  $\mathbf{b}_i$  in model (1) are assumed to be random variables and that is why they cannot be estimated in a standard way. Bayesian techniques can perfectly help in such situation and so called *Empirical Bayes* (EB) estimates  $\hat{\mathbf{b}}_i$  can be used as a basic tool for the inference for the random effects. It will be immediately shown that they can be also quite easily obtained using the common software for fitting the homogeneity linear mixed models. Let us denote the estimate of  $\boldsymbol{\theta}$  parameters obtained using the EM algorithm described in the previous section as  $\hat{\boldsymbol{\theta}}$ . The EB estimate  $\hat{\mathbf{b}}_i$  of the random effects is then given by

$$\hat{\mathbf{b}}_i = \hat{\mathbf{b}}_i(\hat{\boldsymbol{\theta}}) = E[\mathbf{b}_i | \mathbf{Y}_i = \mathbf{y}_i, \boldsymbol{\theta} = \hat{\boldsymbol{\theta}}],$$

where the expected value is based on a posterior distribution derived from the model (4) using Bayesian techniques. See, for example, Gelman *et al.* (1995). It follows from Verbeke and Molenberghs (2000, Section 7.2) that for the homogeneity linear mixed model, the EB estimates are equal to

$$\hat{\mathbf{b}}_i = \hat{\mathbb{D}}\mathbb{Z}_i^T \hat{\mathbb{V}}_i^{-1}(\mathbf{y}_i - \mathbb{X}_i \hat{\boldsymbol{\beta}}^F - \mathbb{Z}_i \hat{\boldsymbol{\beta}}^R),$$

where all ‘hat’ expressions are obtained by replacing their components by the estimates  $\hat{\boldsymbol{\theta}}$ . The EB estimates of the random effects for the heterogeneity linear mixed model are according to Verbeke and Molenberghs (2000, Section 12.3) given by

$$(11) \quad \hat{\boldsymbol{b}}_i = \hat{\mathbb{D}}\mathbb{Z}_i^T \hat{\mathbb{V}}_i^{-1}(\mathbf{y}_i - \mathbb{X}_i \hat{\boldsymbol{\beta}}^F - \mathbb{Z}_i \hat{\boldsymbol{\beta}}^R) + (I_q - \hat{\mathbb{D}}\mathbb{Z}_i^T \hat{\mathbb{V}}_i^{-1} \mathbb{Z}_i) \sum_{j=1}^g p_{ij}(\hat{\boldsymbol{\theta}}) \hat{\boldsymbol{\mu}}_j.$$

We can further derive from (11) that

$$(12) \quad \hat{\boldsymbol{b}}_i = \left\{ \sum_{j=1}^g \left[ p_{ij}(\hat{\boldsymbol{\theta}}) \hat{\mathbb{D}}\mathbb{Z}_i^T \hat{\mathbb{V}}_i^{-1}(\mathbf{y}_i - \mathbb{X}_i \hat{\boldsymbol{\beta}}^F - \mathbb{Z}_i \hat{\boldsymbol{\delta}}_j) \right] \right\} + \sum_{j=1}^g p_{ij}(\hat{\boldsymbol{\theta}}) \hat{\boldsymbol{\mu}}_j,$$

where  $\hat{\boldsymbol{\delta}}_j = \hat{\boldsymbol{\beta}}^R + \hat{\boldsymbol{\mu}}_j$ ,  $j = 1, \dots, g$ . Let us denote  $\hat{\boldsymbol{b}}_i^j = \hat{\mathbb{D}}\mathbb{Z}_i^T \hat{\mathbb{V}}_i^{-1}(\mathbf{y}_i - \mathbb{X}_i \hat{\boldsymbol{\beta}}^F - \mathbb{Z}_i \hat{\boldsymbol{\delta}}_j)$ . The relationship (12) can then be rewritten as

$$\hat{\boldsymbol{b}}_i = \sum_{j=1}^g p_{ij}(\hat{\boldsymbol{\theta}}) \hat{\boldsymbol{b}}_i^j + \sum_{j=1}^g p_{ij}(\hat{\boldsymbol{\theta}}) \hat{\boldsymbol{\mu}}_j.$$

It can be easily revealed that the quantities  $\hat{\boldsymbol{b}}_i^j$  are common EB estimates of random effects for  $a_{ij}(\hat{\boldsymbol{\theta}})$  individuals with common response vector  $\mathbf{Y}_i$  from the homogeneity linear mixed model that was used in the last iteration of the EM algorithm when maximizing  $Q_2^A$  function (10). This property can be advantageously used when computing EB estimates for the heterogeneity linear mixed model.

The EB estimates  $\hat{\boldsymbol{b}}_i$  of the random effects are usually used for diagnostic purposes, such as the detection of outliers etc. More information concerning the use of the EB estimates can be found in Verbeke and Molenberghs (2000, Chapter 7).

**3.4. A SAS Macro.** The just described methodology for the computation of the maximum likelihood estimates of the heterogeneity linear mixed model can be quite easily implemented using existing procedures and functions for fitting homogeneity linear mixed models such as the SAS procedure PROC MIXED or the R/SPLUS function `lme`. The main advantage of this approach is the fact that all covariance structures for matrices  $\Sigma_i$  and  $\mathbb{D}$  that are offered by these functions and procedures can be used.

A SAS macro called `HetMixed`, for fitting heterogeneity models can be downloaded along with its manual from the URL of the Biostatistical Centre, K.U. Leuven:

<http://www.kuleuven.ac.be/biostat/research/software.htm>.

The macro was developed using the SAS version 8. A lot of efforts were spent in making its syntax as similar as possible to the SAS PROC MIXED. So that the user familiar with this procedure should not have any problems when using this macro. More information can be found in Komárek et al. (2002).

#### 4. DISCUSSION

Modelling repeated measures by the homogeneity linear mixed model is not always satisfactory since the assumed normal distribution of random effects might be violated. The homogeneity linear mixed model is also not useful for classification purposes. The so-called heterogeneity linear mixed model that allows us both to classify individual profiles and to create models with many other underlying distributions for random effects than just only Gaussian one was therefore introduced. Appealing properties of the heterogeneity linear mixed model are mostly given by the fact that the distribution of random effects is assumed to be a mixture of normals which can

well approximate lots of other, commonly used continuous distributions. Note that the normality assumption for the random effects is violated, whenever an important categorical covariate has been omitted as a fixed effect in a linear mixed model. Random effects then follow a mixture of  $g$ , possibly normal distributions, where  $g$  is the number of categories of the missing covariate.

Unfortunately, wider using of the heterogeneity linear mixed models was inhibited by insufficient software support. That is why, we have proposed an approximate method how to compute the maximum likelihood estimates of unknown parameters using available procedures and functions for the homogeneity linear mixed model. The proposed method has been implemented as a macro in standard, commercially available software (SAS). The core procedure (`PROC MIXED`) that comes with the software is used to perform the most involved part of the estimation procedure.

#### REFERENCE

- [1] Aitkin, M., and Rubin, D. B. (1985): Estimation and hypothesis testing in finite mixture models. *Journal of the Royal Statistical Society, Series B*, **47**, 67-75.
- [2] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977): Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, **39**, 1-38.
- [3] Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995): *Bayesian Data Analysis*. London: Chapman & Hall.
- [4] Komárek, A., Verbeke, G., and Molenberghs, G. (2002): An approximate approach to fit a linear mixed model with a finite normal mixture as random-effects distribution and its SAS implementation. *Submitted*.
- [5] Verbeke, G., and Lesaffre, E. (1996): A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association*, **91**, 217-221.
- [6] Verbeke, G., and Molenberghs, G. (2000): *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag.

BIostatistical Centre K.U.L., Kapucijnenvoer 35, B-3000 Leuven, Belgium  
E-MAIL: [arnost.komarek@med.kuleuven.ac.be](mailto:arnost.komarek@med.kuleuven.ac.be), [geert.verbeke@med.kuleuven.ac.be](mailto:geert.verbeke@med.kuleuven.ac.be)