Katholieke Universiteit Leuven
Biostatistisch Centrum

A SAS-Macro for Linear Mixed Models with Finite Normal Mixtures as Random-Effects Distribution, version 1.1

Arnošt Komárek
Geert Verbeke
Geert Molenberghs

e-mail address: arnost.komarek@med.kuleuven.ac.be

April 18, 2002

Preface

A SAS macro `HetMixed,` version 1.1 that is able to fit a linear mixed model with finite normal mixtures as random-effects distribution is described in this text. This version of the macro is an improved original `HetMixed` macro which was released through the URL of the Biostatistical Centre, K.U. Leuven: `www.med.kuleuven.ac.be/biostat/` in September 2001 and was desribed in Komárek (2001), available via the same site. The two analyses were presented in the original text. The analysis of the growth curves of schoolgirls and the analysis of the prostate data set (see Chapters 4 and 5 of both the original and this text). The convergence of the EM algorithm that is used to compute desirable estimates was checked partially "by hand" when using the original `HetMixed` macro (by manual increasing of the `A` option of the macro). The version 1.1 enables the user to increase the A[1] value automatically which is the main difference between the original and version 1.1 `HetMixed` macro.

This text is, in fact, only properly changed original macro `HetMixed` manual that is referred as Komárek (2001). Thus the person who has never worked with the original `HetMixed` macro before, need not read the original work and can proceed directly to the version 1.1 of the macro and its manual which is contained in this work.

I hope you will enjoy the macro. All comments to the macro are highly appreciated through my e-mail address.

Arnošt Komárek
Leuven, November 2001

[1] These who do not know what is meant by A value, are referred to the original text or to Chapter 2 of this text.

Acknowledgments

This text would not exist without a direct but also indirect contribution of several people and institutions. I would like to gratefully acknowledge them.

First of all, it would be very hard to develop a SAS-macro described in this text without the idea how to use the existing `MIXED` procedure. I am grateful to Geert Verbeke (Katholieke Universiteit Leuven) for it. He also commented on earlier versions of the macro and the text which led to their considerable improvement. Geert Molenberghs (Limburgs Universitair Centrum Diepenbeek) and Steffen Fieuws (Katholieke Universiteit Leuven) suggested some innovations of the macro during its first public presentation.

It has been an exciting time to spend a year in Diepenbeek while following the Biostatistics Programme of the Limburgs Universitair Centrum. I am grateful to all teachers of this programme. They enabled me to make the links among many parts of my statistical knowledge, I received during my previous studies.

However, a real understanding of many biostatistical methods would not be possible without a theoretical basis obtained in an excellent way at the Faculty of Mathematics and Physics of the Charles University in Prague.

I am also highly grateful to all my LUC schoolmates, especially to all inhabitants of the student house at Stationsstraat in Diepenbeek, for creating a pleasant atmosphere, full of a fruitful cooperation.

Special thanks to Iva for development of the SAS-dance, to Inge for making a nice *lehm* photograph that characterizes Belgium (or at least its province Limburg), to Lenka for her SMS support during hot days spent under the roof of the Stationsstraat house and to the director of Doughnut & Co. for not closing the Belgium office of this company.

Arnošt
Diepenbeek, August 2001

Contents

CHAPTER 1

Introduction

Linear mixed models became one of the most frequently used techniques among statisticians in the last decades. Especially, since the SAS procedure `PROC MIXED` became available. The popularity of this class of models comes from its flexibility and also from the fact that it can deal with correlated data which the statistician is often confronted with. As an example of a simple study when correlated data can be obtained, assume that a child's height is measured repeatedly over time. The obtained data setting is usually called a longitudinal one due to the time ordering of measurements taken on one child. This can be considered as a part of broader class of repeated measures data settings. Let us use this data setting for describing the main features of the linear mixed model even if its use is much broader than just only in this area. Further, let us focus on the situation when the outcome of interest is continuous and normally distributed in some sense.

In repeated measures settings, measurements taken on one (ith) individual can be represented by a response vector $\boldsymbol{Y}_i$. Let us start with the well known *linear model* which specifies the distribution of the response vectors $\boldsymbol{Y}_i$ using the following relationship.

$$\boldsymbol{Y}_i = \mathbb{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i,$$

where matrix $\mathbb{X}_i$ is given by a subject specific set of predictor variables, $\boldsymbol{\beta}$ is a vector of regression parameters that are common to all subjects in the study. Vector $\boldsymbol{\varepsilon}_i$ of residual components is usually assumed to be normally distributed with zero mean and covariance matrix Σ_i. Residual vectors $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{\varepsilon}_j$ for the ith and jth subjects are considered to be independent which reflects investigator's belief that measurements taken on different subjects are independent.

More flexible *linear mixed model* supplements above mentioned relationship by subject specific random effects $\boldsymbol{b}_i$ yielding the following model.

$$\boldsymbol{Y}_i = \mathbb{X}_i\boldsymbol{\beta} + \mathbb{Z}_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i.$$

Vectors $\boldsymbol{b}_i$ are assumed to be random as indicated above and are linked with the response vector $\boldsymbol{Y}_i$ by a matrix of covariates $\mathbb{Z}_i$. The distribution of the random effects is usually considered to be normal with zero mean and some covariance matrix $\mathbb{D}$ common for all subjects. When taking earlier mentioned example of study where the heights of children are measured over the time and assuming that the height evolves linearly in time, the following model can be considered.

$$\text{Height}_{ij} = \beta_0 + \beta_1\text{Age}_{ij} + b_{0i} + b_{1i}\text{Age}_{ij} + \varepsilon_{ij},$$

where Height_{ij} and Age_{ij} are the height and age of the ith child at the time of the jth measurement. The line with intercept β_0 and slope β_1 then describes the mean evolution of the height over the time in the population of interest. If the residual component ε_{ij} is considered as the error of the jth measurement taken on the ith child, the line with intercept $\beta_0 + b_{0i}$ and slope $\beta_1 + b_{1i}$ then describes the true evolution of the height of the ith child cleaned from the measurement error (assuming that the model is correct). So that, the term b_{0i} represents the deviation of the intercept of the line for the ith child from the mean population intercept and similarly, the term b_{1i} the deviation of the slope for the ith child from the mean population slope. Since the mean of components of vectors $\boldsymbol{b}_i$ is assumed to be zero, all subject specific lines should be randomly scattered around the population line given by intercept β_0 and slope β_1. But let us assume the following situation. One part of a study population comes from China where people are generally smaller than for example in Europe and the second part from the Netherlands. Then the intercepts of height lines for children from China can be expected to be systematically below the population mean intercept and intercepts for children from the Netherlands systematically above the population mean intercepts. The same can be valid for slopes. Let us further suppose that the

information containing the nationality of children is not available so that it is not possible to distinguish between children and consequently it is not possible to include two different mean slopes and intercepts into the model. The assumption of normal distribution of random effects $\boldsymbol{b}_i$ can be then violated and rather a mixture of two normals should be considered as a distribution of random effects $\boldsymbol{b}_i$. This change of distributional assumptions for random effects makes a bridge to the so called *heterogeneity linear mixed model* which generally assumes that random effects $\boldsymbol{b}_i$ are distributed according to a mixture of a prespecified number g of normal distributions with different means and common covariance matrix $\mathbb{D}$. Since the mixture of normal distributions is able to approximate many other different distributions, the heterogeneity linear mixed model is a powerful tool in situations when the normality of random effects is possibly violated.

Unfortunatelly, there was no easily available software that would be able to compute estimates of parameters of the heterogeneity linear mixed model. That is why, the purpose of this project was to develop a SAS macro that would be able to do this job. This objective was fulfilled by creating a SAS macro called `HetMixed` and this text should serve mainly as a user's manual. The strategy of the macro is to compute desirable estimates using the EM algorithm with utilizing the SAS procedure `PROC MIXED` within each iteration of the algorithm. The user of the macro `HetMixed` is assumed to be experienced in working with the SAS procedure `PROC MIXED` since the philosophy of the macro is very similar to the philosophy of this procedure.

Broadly, the structure of the text is as follows. The theoretical basis of the heterogeneity linear mixed models is given in Chapter 2 along with description of the technique used for the computation of the estimates. Chapter 3 gives a detailed description of the syntax and output of the macro while the using of the macro is illustrated in Chapters 4 and 5 on real data. It should be mentioned that the purpose of the illustrating examples is to show various possibilities of the macro. So that presented models should not be taken as the optimal ones for given data.

CHAPTER 2

The Heterogeneity Linear Mixed Model

This chapter introduces the concept of the heterogeneity linear mixed model that was proposed by Verbeke and Lesaffre (1996) and also described by Verbeke and Molenberghs (2000, Chapter 12). The heterogeneity linear mixed model can be seen as an extension of common linear mixed model which will be called as the homogeneity linear mixed model. Theoretical basis for a computation of parameter estimates is also given in this chapter.

2.1 Definition of the Heterogeneity Model

Notation that is used within the whole text and the definition of the heterogeneity model is going to be introduced on this place. Let the random variable Y_{ik} denote the (possibly transformed) response of interest, for the ith individual measured at time t_{ik}, $i = 1, \ldots, N$, $k = 1, \ldots, n_i$, and let $\boldsymbol{Y}_i$ be the n_i-dimensional vector of all repeated measurements for the ith subject, that is, $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in_i})^T$. The heterogeneity linear mixed model starts from similar relationship as the homogeneity model, that is from[1]

$$\boldsymbol{Y}_i = \begin{pmatrix} \mathbb{X}_i & \mathbb{Z}_i \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}^F \\ \boldsymbol{\beta}^R \end{pmatrix} + \mathbb{Z}_i \boldsymbol{b}_i + \boldsymbol{\varepsilon}_i, \tag{2.1}$$

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 ith subject. Further, $\boldsymbol{\beta}^F$ and $\boldsymbol{\beta}^R$ are p-dimensional, respectively q-dimensional vectors of unknown regression parameters. Variables $\boldsymbol{b}_i$ are subject-specific q-dimensional random effects, and $\boldsymbol{\varepsilon}_i$ is n_i-dimensional vector of residual components ε_{ik}, $k = 1, \ldots, n_i$. Matrix $\begin{pmatrix} \mathbb{X}_i & \mathbb{Z}_i \end{pmatrix}$ is assumed to have a rank equal to $p + q$. All $\boldsymbol{\varepsilon}_i$ are assumed to be independent and normally distributed with mean vector zero and covariance matrix Σ_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 $\boldsymbol{b}_i$. They are assumed to be independent by both models. The homogeneity model consideres 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[2] $\mathbb{D}$, i.e.

$$\boldsymbol{b}_i \sim \sum_{j=1}^{g} \pi_j N(\boldsymbol{\mu}_j, \mathbb{D}), \tag{2.2}$$

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

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

[1] Slightly different notation than in the introductory chapter is used now to simplify further derivations.

[2] More general case assumes different covariance matrices $\mathtt{D}_1, \ldots, \mathtt{D}_g$ for each component of the mixture. But this can lead to infinitely large likelihood. In order to avoid numerical problems in the estimating procedure, which will be described later, we will assume $\mathtt{D}_1 = \cdots = \mathtt{D}_g = \mathtt{D}$.

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

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

Expectation of the response is then

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

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

$$\sum_{j=1}^{g}\pi_j\boldsymbol{\mu}_j = \mathbf{0} \tag{2.3}$$

is needed.

The model (2.1) with assumptions (2.2) can be also rewritten as a following hierarchical Bayes model

$$\begin{aligned}
&\boldsymbol{Y}_i|\boldsymbol{b}_i \sim N(\mathbb{X}_i\boldsymbol{\beta}^F + \mathbb{Z}_i\boldsymbol{\beta}^R + \mathbb{Z}_i\boldsymbol{b}_i, \Sigma_i),\\
&\boldsymbol{b}_i|\boldsymbol{\mu} \sim N(\boldsymbol{\mu}, \mathbb{D}),\\
&\boldsymbol{\mu} \in \{\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_g\}, \qquad \text{with } P(\boldsymbol{\mu} = \boldsymbol{\mu}_j) = \pi_j.
\end{aligned} \tag{2.4}$$

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 $\boldsymbol{Y}$ into one of the populations.

2.2 Estimation of the Heterogeneity Model

Estimates of unknown parameters of the heterogeneity model should be based on a marginal distribution of the observations $\boldsymbol{Y}_i$. This distribution under (2.1) and (2.2) can be easily found to be given by

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

Let $\boldsymbol{\pi}$ be the vector of component probabilities (i.e. $\boldsymbol{\pi}^T = (\pi_1, \ldots, \pi_g)$) 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 Σ_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 $\boldsymbol{Y}_i$ is of the form

$$L^*(\boldsymbol{\theta}|\boldsymbol{y}) = \prod_{i=1}^{N}\left\{\sum_{j=1}^{g}\pi_j f_{ij}(\boldsymbol{y}_i|\boldsymbol{\gamma})\right\}. \tag{2.5}$$

where $\boldsymbol{y}^T = (\boldsymbol{y}_1^T, \ldots, \boldsymbol{y}_N^T)$ is the vector containing all observed response values and f_{ij} is a density of 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 (2.5) 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 be easily overcome by imposing some constraint on the parameters. For example, the constraint

$$\pi_1 \geq \pi_2 \geq \cdots \geq \pi_g \tag{2.6}$$

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 (2.6).

The log-likelihood function corresponding to the likelihood (2.5) is then

$$l^*(\boldsymbol{\theta}|\boldsymbol{y}) = \sum_{i=1}^{N} \log \left\{ \sum_{j=1}^{g} \pi_j f_{ij}(\boldsymbol{y}_i|\boldsymbol{\gamma}) \right\}. \tag{2.7}$$

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. Concept of complete and incomplete data is introduced for the purpose of using the EM algorithm. Response vectors $\boldsymbol{Y}_i$ along with (unobserved) population indicators $\boldsymbol{W}_i$ can be seen as complete data whereas vectors $\boldsymbol{Y}_i$ alone can be viewed as incomplete data since information containing population pertinence is missing. The likelihood function (2.5) corresponds then to the incomplete data. The likelihood function that would have been obtained if values $\boldsymbol{w}_i = (w_{i1}, \ldots, w_{ig})^T$ of population indicators $\boldsymbol{W}_i$ had been observed is then

$$L(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{w}) = \prod_{i=1}^{N} \prod_{j=1}^{g} \{\pi_j f_{ij}(\boldsymbol{y}_i|\boldsymbol{\gamma})\}^{w_{ij}} \tag{2.8}$$

where $\boldsymbol{w}^T = (\boldsymbol{w}_1^T, \ldots, \boldsymbol{w}_N^T)$ is the vector containing all hypothetically observed population indicators.

The log-likelihood function corresponding to (2.8) has then the more attractable form

$$l(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{w}) = \sum_{i=1}^{N} \sum_{j=1}^{g} w_{ij} \{\log \pi_j + \log f_{ij}(\boldsymbol{y}_i|\boldsymbol{\gamma})\}. \tag{2.9}$$

Maximizing $l(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{w})$ with respect to $\boldsymbol{\theta}$ yields estimates which depend on the unobserved ("missing") indicators $\boldsymbol{w}$. The EM algorithm offers a solution to this problem by maximizing the expected value of $l(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{w})$, rather than $l(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{w})$ with respect to $\boldsymbol{\theta}$, where the expectation is taken over all unobserved w_{ij}. The conditional expectation of $l(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{w})$, given the observed data vector $\boldsymbol{y}$, 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. The expected log-likelihood function will be called as the objective function and will be denoted as Q within this text.

The EM algorithm for finding estimates of the heterogeneity model is going to be described more in detail in the following paragraph. Suppose that $\boldsymbol{\theta}^{(t)}$ is 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 $l(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{w})$, given the observed data vector $\boldsymbol{y}$ is given by

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) &= E\big[l(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{w})\big|\boldsymbol{y}, \boldsymbol{\theta}^{(t)}\big] \\ &= \sum_{i=1}^{N} \sum_{j=1}^{g} p_{ij}(\boldsymbol{\theta}^{(t)}) \big\{\log \pi_j + \log f_{ij}(\boldsymbol{y}_i|\boldsymbol{\gamma})\big\}. \end{aligned} \tag{2.10}$$

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

$$\begin{aligned} p_{ij}(\boldsymbol{\theta}^{(t)}) &= E\big[W_{ij}\big|\boldsymbol{y}_i, \boldsymbol{\theta}^{(t)}\big] = P\big(W_{ij} = 1\big|\boldsymbol{y}_i, \boldsymbol{\theta}^{(t)}\big) = \\ &= \frac{\pi_j^{(t)} f_{ij}(\boldsymbol{y}_i|\boldsymbol{\gamma}^{(t)})}{\sum_{k=1}^{g} \pi_k^{(t)} f_{ik}(\boldsymbol{y}_i|\boldsymbol{\gamma}^{(t)})}. \end{aligned} \tag{2.11}$$

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 (2.10) 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

$$Q_1(\boldsymbol{\pi}|\boldsymbol{\theta}^{(t)}) = \sum_{i=1}^{N}\sum_{j=1}^{g} p_{ij}(\boldsymbol{\theta}^{(t)}) \log \pi_j, \tag{2.12}$$

$$Q_2(\boldsymbol{\gamma}|\boldsymbol{\theta}^{(t)}) = \sum_{i=1}^{N}\sum_{j=1}^{g} p_{ij}(\boldsymbol{\theta}^{(t)}) \log f_{ij}(\boldsymbol{y}_i|\boldsymbol{\gamma}). \tag{2.13}$$

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 the Q function.

We first maximize the expression (2.12). This has to be done under the restriction imposed on the prior probabilities, $\sum_{j=1}^{g} \pi_j = 1$. Rewriting (2.12) into the form

$$Q_1(\boldsymbol{\pi}|\boldsymbol{\theta}^{(t)}) = \sum_{i=1}^{N}\sum_{j=1}^{g-1} p_{ij}(\boldsymbol{\theta}^{(t)}) \log \pi_j + \sum_{i=1}^{N} p_{ig}(\boldsymbol{\theta}^{(t)}) \log\Big(1 - \sum_{j=1}^{g-1} \pi_j\Big)$$

and setting all first-order derivatives with respect to $\pi_1, \ldots, \pi_{g-1}$ equal to zero yields that the updated estimates satisfy

$$\frac{\pi_j^{(t+1)}}{\pi_g^{(t+1)}} = \frac{\sum_{i=1}^{N} p_{ij}(\boldsymbol{\theta}^{(t)})}{\sum_{i=1}^{N} p_{ig}(\boldsymbol{\theta}^{(t)})}, \qquad j = 1, \ldots, g-1.$$

Using the condition

$$1 = \sum_{j=1}^{g} \pi_j^{(t+1)} = \frac{N\pi_g^{(t+1)}}{\sum_{i=1}^{N} p_{ig}(\boldsymbol{\theta}((t))}$$

gives us the relationship

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

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

Unfortunatelly, the term (2.13) cannot be maximized analytically as the first one. A numerical maximization procedure such as Newton-Raphson is needed to maximize $Q_2(\boldsymbol{\gamma}|\boldsymbol{\theta}^{(t)})$ with respect to $\boldsymbol{\gamma}$. Derivatives of Q_2 with respect to components of $\boldsymbol{\gamma}$ have to be obtained to be able to use above mentioned algorithm. This is not a trivial task, even for simple cases such as e.g. $\Sigma_i = \sigma^2 I_{n_i}$. 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`.

2.3 How to Maximize Q_2 – the Second Part of the Objective Function

The function

$$Q_2(\boldsymbol{\gamma}|\boldsymbol{\theta}^{(t)}) = \sum_{i=1}^{N}\sum_{j=1}^{g} p_{ij}(\boldsymbol{\theta}^{(t)}) \log f_{ij}(\boldsymbol{y}_i|\boldsymbol{\gamma}) \tag{2.14}$$

is to be maximized with respect to $\boldsymbol{\gamma}$. Let us first explore the log-likelihood function of the homogeneity model. That can be obtained from (2.7) by setting $g = 1$ and $\pi_1 = 1$ which yields

$$l^{HOM}(\boldsymbol{\theta}|\boldsymbol{y}) = \sum_{i=1}^{N} \log f_i(\boldsymbol{y}_i|\boldsymbol{\gamma}),$$

where f_i is a density of n_i-dimensional normal distribution $N(\mathbb{X}_i\boldsymbol{\beta}^F + \mathbb{Z}_i\boldsymbol{\beta}^R, \mathbb{V}_i)$. The constraint (2.3) yields common condition $E(\boldsymbol{b}_i) = \boldsymbol{\mu}_1 = \mathbf{0}$. If posterior probabilities $p_{ij}(\boldsymbol{\theta}^{(t)})$ are integers, the function (2.14) would be a log-likelihood for the homogeneity model based on observations from $\sum_{i=1}^{N}\sum_{j=1}^{g} p_{ij}(\boldsymbol{\theta}^{(t)})$ individuals. Moreover, maximization of (2.14) with respect to $\boldsymbol{\gamma}$ is equivalent to maximization of

$$A \cdot Q_2(\boldsymbol{\gamma}|\boldsymbol{\theta}^{(t)}) = \sum_{i=1}^{N}\sum_{j=1}^{g} A \cdot p_{ij}(\boldsymbol{\theta}^{(t)}) \log f_{ij}(\boldsymbol{y}_i|\boldsymbol{\gamma}) = \sum_{i=1}^{N}\sum_{j=1}^{g} a_{ij}(\boldsymbol{\theta}^{(t)}) \log f_{ij}(\boldsymbol{y}_i|\boldsymbol{\gamma}) \tag{2.15}$$

for an arbitrary positive constant A. Further, numbers

$$a_{ij}(\boldsymbol{\theta}^{(t)}) = A \cdot p_{ij}(\boldsymbol{\theta}^{(t)}) \tag{2.16}$$

can be arbitrarily close to integers by choosing A sufficiently large. In practice, rounded values of $A \cdot p_{ij}(\boldsymbol{\theta}^{(t)})$ can be used to approximate the function $A \cdot Q_2(\boldsymbol{\gamma}|\boldsymbol{\theta}^{(t)})$. Subsequently common software for homogeneity linear mixed models such as the SAS procedure `PROC MIXED` 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 since observations from approximatelly $\sum_{i=1}^{N}\sum_{j=1}^{g} A \cdot p_{ij}(\boldsymbol{\theta}^{(t)})$ individuals are used to find desirable estimates.

When implementing this method, one also has to take into account the constraint (2.3) 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 chapter. Fortunatelly, it is not too difficult to ensure that this constraint is satisfied as one can immediately see. Suppose that all $a_{ij}(\boldsymbol{\theta}^{(t)})$ are integers. Then the function

$$Q_2^A(\boldsymbol{\gamma}|\boldsymbol{\theta}^{(t)}) = \sum_{i=1}^{N}\sum_{j=1}^{g} a_{ij}(\boldsymbol{\theta}^{(t)}) \log f_{ij}(\boldsymbol{y}_i|\boldsymbol{\gamma}) \tag{2.17}$$

that is to be maximized with respect to $\boldsymbol{\gamma}$ can represent the log-likelihood of the homogeneity linear mixed model where originally restricted q-dimensional parameters $\boldsymbol{\beta}^R, \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_g$ are replaced by unrestricted q-dimensional parameters $\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_g$. Original and new parameters are binded together by the relationship.

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

In fact, parameters $\boldsymbol{\delta}_j$ express real population means, whereas parameters $\boldsymbol{\mu}_j$ a mean deviation of each population from the overall mean $\boldsymbol{\beta}^R$. Restriction (2.3) also gives the way to compute $\boldsymbol{\beta}^R$ from $\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_g$, that is

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

Homogeneity linear mixed model corresponding to the log-likelihood (2.17) is then of the form

$$\boldsymbol{Y}_i{}^* = \begin{pmatrix}\mathbb{X}_i^* & \mathbb{Z}_i^*\end{pmatrix} \begin{pmatrix}\boldsymbol{\beta}^F \\ \boldsymbol{\delta}_1 \\ \vdots \\ \boldsymbol{\delta}_g\end{pmatrix} + \mathbb{Z}_i^{**}\boldsymbol{b}_i^* + \boldsymbol{\varepsilon}_i^*, \tag{2.18}$$

where

$\boldsymbol{Y}_i^* = (\boldsymbol{Y}_i^T, \ldots, \boldsymbol{Y}_i^T)^T$ in which the response vector $\boldsymbol{Y}_i$ is repeated $\sum_{j=1}^{g} a_{ij}(\boldsymbol{\theta}^{(t)})$ times ,

$$\mathbb{X}_i^* = \begin{pmatrix} \mathbb{X}_i \\ \vdots \\ \mathbb{X}_i \end{pmatrix} \text{ with } \sum_{j=1}^{g} a_{ij}(\boldsymbol{\theta}^{(t)}) \text{ times repeated matrix } \mathbb{X}_i,$$

$$\mathbb{Z}_i^* = \begin{pmatrix} \mathbb{Z}_i^1 \\ \vdots \\ \mathbb{Z}_i^1 \\ \vdots \\ \mathbb{Z}_i^g \\ \vdots \\ \mathbb{Z}_i^g \end{pmatrix}, \qquad \begin{aligned} \mathbb{Z}_i^1 &= \begin{pmatrix} \mathbb{Z}_i & \mathbb{O} & \ldots & \mathbb{O} \end{pmatrix}, \\ \mathbb{Z}_i^2 &= \begin{pmatrix} \mathbb{O} & \mathbb{Z}_i & \ldots & \mathbb{O} \end{pmatrix}, \\ &\ldots \\ \mathbb{Z}_i^g &= \begin{pmatrix} \mathbb{O} & \mathbb{O} & \ldots & \mathbb{Z}_i \end{pmatrix}, \end{aligned}$$

where $\mathbb{Z}_i^j$ is a $n_i \times qg$ matrix which is repeated $a_{ij}(\boldsymbol{\theta}^{(t)})$ times and $\mathbb{O}$ states for $n_i \times q$ matrix of all zeros,

$$\mathbb{Z}_i^{**} = \begin{pmatrix} \mathbb{Z}_i & \mathbb{O} & \ldots & \mathbb{O} \\ \mathbb{O} & \mathbb{Z}_i & \ldots & \mathbb{O} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{O} & \mathbb{O} & \ldots & \mathbb{Z}_i \end{pmatrix} \text{ with } \sum_{j=1}^{g} a_{ij}(\boldsymbol{\theta}^{(t)}) \text{ times repeated matrix } \mathbb{Z}_i \text{ on the diagonal,}$$

$$\boldsymbol{b}_i^* = \begin{pmatrix} \boldsymbol{b}_{i1}^* \\ \vdots \\ \boldsymbol{b}_{iA_i}^* \end{pmatrix}, \qquad \boldsymbol{\varepsilon}_i^* = \begin{pmatrix} \boldsymbol{\varepsilon}_{i1}^* \\ \vdots \\ \boldsymbol{\varepsilon}_{iA_i}^* \end{pmatrix}, \qquad A_i = \sum_{j=1}^{g} a_{ij}(\boldsymbol{\theta}^{(t)}).$$

Vectors $\boldsymbol{b}_{il}^*,\ i = 1, \ldots, N,\ l = 1, \ldots, A_i$ are assumed to be independent following a q-dimensional normal distribution $N(\boldsymbol{0}, \mathbb{D})$. The residual vectors $\boldsymbol{\varepsilon}_{il}^*,\ i = 1, \ldots, N,\ l = 1, \ldots, A_i$ are assumed to be independent following n_i-dimensional normal distribution $N(\boldsymbol{0}, \Sigma_i)$.

Just described homogeneity linear mixed model can now be used to compute updated approximate estimates of $\boldsymbol{\gamma}$ parameters within the M step of the EM algorithm when computing estimates for the heterogeneity linear mixed model.

2.4 Empirical Bayes Inference

The random effects $\boldsymbol{b}_i$ in model (2.1) 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 $\hat{\boldsymbol{b}}_i$ can be used as a basic tool for the inference for the random effects.

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 $\hat{\boldsymbol{b}}_i$ of the random effects is then given by

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

where the expected value is based on a posterior distribution derived from the model (2.4) 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

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

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

$$\hat{\boldsymbol{b}}_i = \hat{\mathbb{D}}\mathbb{Z}_i^T\hat{\mathbb{V}}_i^{-1}(\boldsymbol{y}_i - \mathbb{X}_i\hat{\boldsymbol{\beta}}^F - \mathbb{Z}_i\hat{\boldsymbol{\beta}}^R) + (I_q - \hat{\mathbb{D}}\mathbb{Z}_i^T\hat{\mathbb{V}}_i^{-1}\mathbb{Z}_i)\sum_{j=1}^{g} p_{ij}(\hat{\boldsymbol{\theta}})\hat{\boldsymbol{\mu}}_j. \tag{2.19}$$

We can further derive from (2.19) the following.

$$\begin{aligned}\hat{\boldsymbol{b}}_i &= \left\{\sum_{j=1}^{g}\left[p_{ij}(\hat{\boldsymbol{\theta}})\hat{\mathbb{D}}\mathbb{Z}_i^T\hat{\mathbb{V}}_i^{-1}(\boldsymbol{y}_i - \mathbb{X}_i\hat{\boldsymbol{\beta}}^F - \mathbb{Z}_i\hat{\boldsymbol{\beta}}^R - \mathbb{Z}_i\hat{\boldsymbol{\mu}}_j)\right]\right\} + \sum_{j=1}^{g} p_{ij}(\hat{\boldsymbol{\theta}})\hat{\boldsymbol{\mu}}_j \\ &= \left\{\sum_{j=1}^{g}\left[p_{ij}(\hat{\boldsymbol{\theta}})\hat{\mathbb{D}}\mathbb{Z}_i^T\hat{\mathbb{V}}_i^{-1}(\boldsymbol{y}_i - \mathbb{X}_i\hat{\boldsymbol{\beta}}^F - \mathbb{Z}_i\hat{\boldsymbol{\delta}}_j)\right]\right\} + \sum_{j=1}^{g} p_{ij}(\hat{\boldsymbol{\theta}})\hat{\boldsymbol{\mu}}_j,\end{aligned} \tag{2.20}$$

where $\hat{\boldsymbol{\delta}}_j = \hat{\boldsymbol{\beta}}^R + \hat{\boldsymbol{\mu}}_j$, $j = 1, \ldots, g$. Let us denote

$$\hat{\boldsymbol{b}}_i^j = \hat{\mathbb{D}}\mathbb{Z}_i^T\hat{\mathbb{V}}_i^{-1}(\boldsymbol{y}_i - \mathbb{X}_i\hat{\boldsymbol{\beta}}^F - \mathbb{Z}_i\hat{\boldsymbol{\delta}}_j).$$

The relationship (2.20) can then be rewritten as

$$\hat{\boldsymbol{b}}_i = \sum_{j=1}^{g} p_{ij}(\hat{\boldsymbol{\theta}})\hat{\boldsymbol{b}}_i^j + \sum_{j=1}^{g} p_{ij}(\hat{\boldsymbol{\theta}})\hat{\boldsymbol{\mu}}_j. \tag{2.21}$$

It can be easily revealed that the quantities $\hat{\boldsymbol{b}}_i^j$ are EB estimates of random effects from the homogeneity linear mixed model (2.18) that was used in the last iteration of the EM algorithm when computing the estimates $\hat{\boldsymbol{\theta}}$ of parameters $\boldsymbol{\theta}$. 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. However, it should be emphasized that the EB estimators $\hat{\boldsymbol{b}}_i$ all have different distributions unless all covariate matrices $\mathbb{X}_i$ and $\mathbb{Z}_i$ are the same so that the inference based on them has to be done with a high care. More information concerning the use of the EB estimates can be found in Verbeke and Molenberghs (2000, Chapter 7).

2.5 Inference Based on the Heterogeneity Model

In practice, one is often interested in drawing inferences on the paramaters in a model. Since the parameter estimates of the heterogeneity model are computed using the EM algorithm, it is quite difficult to compute their standard errors that are usually used for drawing above mentioned inferences. Also computation of standard errors using the technique suggested for the EM algorithm by Louis (1982) is far from straightforward. To obtain the standard errors according to this paper, one has to compute, among others, the derivatives of the log-likelihood (2.9) and this is exactly the procedure we want to omit and the reason, why the approximate technique to maximize the objective function Q defined by (2.10) is used.

So that, one has to be satisfied with the likelihood ratio test that can be performed by fitting two nested models and subtracting appropriate doubled values of log-likelihoods of these models. It should be mentioned that this procedure can sometimes be quite time consuming.

Moreover, drawing the inference about the parameters of the mixture (i.e. parameters $\pi_1, \ldots, \pi_g$ and $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_g$) and about the number of components g in (2.2) is even more complicated due to boundary problems as discussed by Ghosh and Sen (1985). In order to briefly highlight the main problems, we can

consider testing $\mathrm{H}_0 : g = 1$ versus $\mathrm{H}_A : g = 2$. The null hypothesis can then be expressed as $\mathrm{H}_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$. However, the same hypothesis is obtained by setting $\mathrm{H}_0 : \pi_1 = 0$ or $\mathrm{H}_0 : \pi_2 = 0$, which clearly illustrates that H_0 is on the boundary of the parameter space, and hence also that the usual regularity conditions for application of the classical maximum likelihood theory are violated. In practice however, it may be sufficient to fit several heterogeneity models and to explore how increasing g affects the inference for the parameters of interest.

CHAPTER 3

The Macro HetMixed Description

The objective of this project was to develop a SAS macro in which the previously described methodology for fitting heterogeneity models has been implemented. The used approximate method to maximize the objective function Q is preferred to the derivation of all necessary derivatives for the exact maximization of this function using the Newton-Raphson algorithm. The reasons are as follows. The approximate method uses the SAS procedure `PROC MIXED` for the most difficult part of the estimates computation, which is the M step of the EM algorithm. Consequently, a variety of covariance structures offered by the `PROC MIXED` can be specified by the user. Analytical derivation of all expressions necessary for the Newton-Raphson algorithm would have had to be done separately for each of these covariance structures if the exact maximization of the objective function Q had been to be performed. As already stated in the previous chapter, it is not a trivial task to derive analytically all necessary expressions even in the simplest case which is the so called conditional independence, i.e. $\Sigma_i = \sigma^2 I_{n_i}$. Even if all this tough derivation had been done, the resulting implementation of the maximization algorithm would not have likely been as stable and trustworthy as the used implementation where the maximization is provided by the widely reputable SAS procedure `PROC MIXED`.

The created macro is called `HetMixed` and the main purpose of this chapter is to describe its syntax and its possibilities. The macro was developed using SAS Version 8 but it should work also with Version 7. It will not work in older versions due to the change of `MAKE` statement of `PROC MIXED` into `ODS`. The macro code can be found in Appendix and its electronic form, as well as the SAS script files used for the reported examples, can be found at the URL of the Biostatistical Centre, K.U. Leuven: `www.med.kuleuven.ac.be/biostat/`.

3.1 The Syntax of the Macro

The syntax of the macro is similar to the syntax of the SAS `PROC MIXED`. The options of the macro `HetMixed` with the same label as the statements of `PROC MIXED` has the same meaning. The following options can be used with the `HetMixed` macro.

%MACRO HetMixed(
DATA = *SAS data set*,
SUBJECT = *subject effect*,
REPEATED = *repeated effect*,
RESPONSE = *response effect*,
FIXED = *fixed effects*,
RANDOM = *random effects*,
TYPEREP = *covariance structure*,
TYPERAND = *covariance structure*,
G = *number*,
AMIN= *number*,
AMAX = *number*,
ABY = *number*,
DECISWIT = *number*,
DECISBET = *number*,
STOPWIT = *number*,
STOPBET = *number*,

MAXITER = *number*,
MIXEDMI = *number*,
INITPOST = *SAS data set*,
ENDPOST = *SAS data set*,
EB = *SAS data set*,
PIEPS = *number*
)

Following sections provide a detailed description of each of these options.

3.2 The Macro Strategy

It can be useful to describe more in detail the strategy that is used by the macro `HetMixed` to the estimates computation before a detailed description of all options will be shown.

As already stated, the technique described in Chapter 2 is used by the macro to compute desirable estimates of unknown parameters. The first problem that has to be solved before starting the EM algorithm, is finding suitable initial estimates of the parameter $\boldsymbol{\theta}$ that is to be estimated. Note that the knowledge of posterior probabilities $p_{ij}(\boldsymbol{\theta}^{(0)})$ defined by (2.11) is sufficient to compute the value of the objective function $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(0)})$ given by (2.10) that is to be maximized in the M step of the first iteration of the EM algorithm. This nice property is utilizes by the macro and initial posterior probabilities $p_{ij}(\boldsymbol{\theta}^{(0)})$ for the ith individual to belong to the jth component of the mixture are used to start the EM algorithm. Such initial posterior probabilities can be given by the user as described later in the `INITPOST` option section or can be randomly generated as decribed in the `PIEPS` option section.

The E step of the EM algorithm is quite straightforward alike the computation of updated estimates of component probabilities $\boldsymbol{\pi}$. On the other hand, the computation of updated estimates of $\boldsymbol{\gamma}$ parameters is much more complicated and multiplication technique described in Chapter 2 is used. At each iteration of the EM algorithm, an extended data set corresponding to the homogeneity linear mixed model (2.18) is created and subsequently updated estimates of $\boldsymbol{\gamma}$ parameters are computed using `PROC MIXED`. Multiplication factors $a_{ij}(\boldsymbol{\theta}^{(t)})$ defined by (2.16) are rounded in the way that is described in the `AMIN, AMAX` and `ABY` options section.

Exact description of the stopping rule for the EM algorithm can be found in the `DECISWIT, DECISBET, STOPWIT, STOPBET` and `MAXITER` options sections.

The empirical Bayes estimates of random effects are also computed by the macro and can be saved in a prespecified data set (see `EB` option section). The relationship (2.21) is used to compute them.

3.3 Detailed Options Description

3.3.1 The DATA Option

It specifies the input data set, structured as required by the SAS procedure `PROC MIXED` (1999, Version 8). The default is the most recently created data set.

3.3.2 The SUBJECT Option

It identifies the subjects in your mixed model. This option has the same meaning as the `SUBJECT` option of the `REPEATED` or `RANDOM` statements of `PROC MIXED`. The macro `HetMixed` creates its own data set where only necessary variables are stored and the observations are sorted according to the `SUBJECT` variable. The observations with the same value of the `SUBJECT` variable determine the vector $\boldsymbol{Y}_i$ as defined by (2.1).

3.3.3 The REPEATED Option

It indicates ordering of observations within one subject. It could be also called as a time variable. Be aware of the fact that the specified variable is treated as of the categorical type (`'CLASS'`) inside the macro. Thus if you want to use this time variable in a model in its continuous form, you have to create one replication of it. Number of different values of `REPEATED` variables within one subject determine the number of observations n_i within vector $\boldsymbol{Y}_i$. Make sure that the levels of the repeated effect are different for each observation within a subject.

3.3.4 The RESPONSE Option

It names the response variable which forms vectors $\boldsymbol{Y}_i$.

3.3.5 The FIXED Option

The `FIXED` option names the fixed effects, which determine the $\mathbb{X}_i$ matrices of the mixed model (2.1). Note that, in contrast to most regression procedures in SAS, no intercept is specified by default. Hence, if an intercept is to be included it has to be specified explicitly as one of the effects in this option. Moreover, the variable containing all ones has to be created beforehand using a `DATA` step when an intercept is to be included.

The `FIXED` option **must not** contain any effects that are to be included in the following `RANDOM` effect. This is in agreement with the definition (2.1) where the design matrix for fixed effects is strictly splited into two parts. This will be further clarified in the next section which describes the `RANDOM` option.

3.3.6 The RANDOM Option

This option names the random effects, which determine the $\mathbb{Z}_i$ matrices of the mixed model (2.1). If a random intercept is to be included in the model, a variable containing all ones has to be again created beforehand and then specified in the `RANDOM` option. All effects specified in this option has their representation also in fixed effects design matrix as one can see in (2.1). Be aware of the fact that this is in contrast to `PROC MIXED`.

Warning: The user is encouraged not to use variable names for random effects that end by a number like e.g. `time2`. This can cause errors of the macro since numbers indicating mixture components are added inside the macro to the random effects names and problems can occur if e.g. random effects labeled as `time` and `time2` are specified by the user.

3.3.7 Common Notes Related to the FIXED and RANDOM Options

There is no `'CLASS'` statement in the macro. That is why, dummy variables for all categorical effects have to be created using separate `DATA` step by the user and then specified in `FIXED` or `RANDOM` option. Also the variables for all interaction terms have to be created by the user using the `DATA` step. Literally, all effects named in either `FIXED` or `RANDOM` option of the macro have to have their column representation in the used data set. I have to admit that this is not too handy for the user, nevertheless to implement the same treatment of categorical variables and interaction terms as it is done in `PROC MIXED` would be quite tough for me when programming the macro. Moreover, I think that it is not as big inconvenience for the user and an advantage of this approach is the fact that the user exactly knows the interpretation of the model parameters and does not have to explore whether dummy variables are defined such that all zeros are put to the last level or to the first level of the categorical covariate.

Note also that the matrix $\begin{pmatrix}\mathbb{X}_i & \mathbb{Z}_i\end{pmatrix}$ from (2.1) is assumed to be of the rank equal $p+q$. It also means that all effects named in the `RANDOM` option must not be expressable as linear combinations of arbitrary effects named in the `FIXED` option. As a consequence, only "contrast" parametrization is allowed. See the following example for the illustration.

TABLE 3.1. Hypothetical Illustration Data

Y	age category	*id* (i)	*time* (j)	TIMEC	INT	AGE1	AGE2
5	old	1	1	1	1	1	0
5.5	old	1	2	2	1	1	0
5.6	old	1	3	3	1	1	0
4.1	old	2	1	1	1	1	0
4.8	old	2	2	2	1	1	0
4.9	old	2	3	3	1	1	0
5.8	old	3	1	1	1	1	0
5.9	old	3	2	2	1	1	0
6.6	old	3	3	3	1	1	0
3.4	young	4	1	1	1	0	1
3.6	young	4	2	2	1	0	1
3.9	young	4	3	3	1	0	1
4.4	young	5	1	1	1	0	1
4.6	young	5	2	2	1	0	1
4.9	young	5	3	3	1	0	1

Example. A simple example should help to bring more light into the problem. Let us consider the data set given by Table 3.1 where three repeated measurements on five subjects of two age categories were done. We can assume the following two models.

$$Y_{ij} = \beta_0 \,\mathrm{INT}_i + \beta_1 \,\mathrm{AGE1}_i + \beta_2 \,\mathrm{TIMEC}_{ij} + b_0 \,\mathrm{INT}_i + \varepsilon_{ij}, \qquad i = 1, \ldots, 5, \quad j = 1, 2, 3$$

and

$$Y_{ij} = \beta_1 \,\mathrm{AGE1}_i + \beta_2 \,\mathrm{AGE2}_i + \beta_3 \,\mathrm{TIMEC}_{ij} + b_0 \,\mathrm{INT}_i + \varepsilon_{ij}, \qquad i = 1, \ldots, 5, \quad j = 1, 2, 3.$$

Both models are equivalent as one can easily find. However, the second parametrization is not allowed by the macro `HetMixed` since $\mathrm{INT}_i = \mathrm{AGE1}_i + \mathrm{AGE2}_i$. Only the first parametrization can be used and the macro options have to be filled out in the following way.

```
SUBJECT = id,
REPEATED = time,
RESPONSE = Y,
FIXED = AGE1 TIMEC,
RANDOM = INT.
```

3.3.8 The TYPEREP Option

The `TYPEREP` option names the type of the covariance matrix Σ_i of residual components of the model as in (2.1). It corresponds to the `TYPE` option of the `REPEATED` statement of `PROC MIXED`. All covariance structures available within `PROC MIXED` can be used in the macro `HetMixed.` If no `TYPEREP` statement is specified, Σ_i is assumed to be equal to $\sigma^2 I_{n_i}$, i.e. `'SIMPLE'` covariance structure of residual components.

3.3.9 The TYPERAND Option

This option specifies the type of the covariance matrix of random effects, i.e. $\mathbb{D}$ matrix from (2.2). The `TYPE` option of the `RANDOM` statement of `PROC MIXED` is analogical to this macro option. Like in the case

of the TYPEREP option, all covariance structures available within PROC MIXED are allowed. The default value of TYPERAND is the unstructured matrix $\mathbb{D}$ ('UN').

3.3.10 The G Option

The G option specifies the number of components of the heterogeneity model as in (2.2). So that, G has to be a positive integer. Its default value is one although G=1 does not give any reasons to use the macro HetMixed since PROC MIXED can be used instead.

3.3.11 The Options Controlling the Iterative Process

The options called AMIN, AMAX, ABY, DECISWIT, DECISBET, STOPWIT, STOPBET and MAXITER control the iterative process that is used to compute the estimates.

As indicated earlier, the EM algorithm with a technique of multiplying the data set introduced in Chapter 2 is used to compute requested estimates. Within each iteration of the EM algorithm, a specific value of the multiplication factor A for computing numbers

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

as defined in (2.16) is used. Values of $a_{ij}(\boldsymbol{\theta}^{(t)})$ are obtained by rounding off $A \cdot p_{ij}(\boldsymbol{\theta}^{(t)})$ and changing possible zeros into ones. Be aware of the fact that each consequently created data set that is used to compute updated estimates of $\boldsymbol{\theta}$ has to correspond to the homogeneity linear mixed model (2.18). That is why it has to reflect observations taken on $\sum_{i=1}^{N}\sum_{j=1}^{g} a_{ij}(\boldsymbol{\theta}^{(t)})$ subjects. This number of subjects should never exceed 32767 which is a restriction given by the VI option of the RANDOM statement of PROC MIXED. But with respect to the power of present computers, this is not any restriction since data set containing more than 30000 subjects and only two observations per subjects would have more than 60000 observations. The PROC MIXED used within each iteration of the EM algorithm would either crash or compute several days estimates for a data set with such amount of observations using presently available computers. The value of A for each iteration of the EM algorithm is derived from the values of the options AMIN, AMAX and ABY as described further.

The A value has an influence both on obtaining correct estimates and computational time. The influence on the computational time can be quite considerable. That is why, it is desirable to start with smaller value of A which can be consequently increased. This can be done automatically by the macro and a sequence of sets of the iterations of the EM algorithm can be computed. At the same time, each set of iterations uses different value of A, the later sets use higher values of A than the earlier sets. One set of the EM algorithm finishes according to the values of the DECISWIT and STOPWIT options, the all iteration process finishes according to the values of the DECISBET and STOPBET options or according to the value of the MAXITER and AMAX options.

3.3.12 The AMIN, AMAX and ABY Options

These options control the value of the multiplication factor A. The first set of iterations of the EM algorithm is computed using $A =$ AMIN until convergence given by the DECISWIT and STOPWIT options is obtained or maximum number of iterations given by MAXITER is reached. Since increased value of A provides a better approximation to the objective function Q and hence more correct estimates of unknown parameters, at least one additional iteration of the EM algorithm using higher value of A should be computed and so obtained value of the objective function or so obtained estimates should be compared to the value of the objective function or to the estimates received after the last iteration of the EM algorithm with smaller value of A. Such procedure can protect us from a "false" convergence caused by too rough approximation to the objective function when insufficiently high value of A is used. The higher value of A is obtained by adding the value of ABY to the previous smaller value of A. The value of the objective function or the estimates obtained after one iteration of the EM algorithm with the higher value of A are compared to the value of the objective function or to the estimates received after the last iteration with smaller A and the iterative process either stop or continue according to the stopping rule given by the DECISBET and STOPBET options. If this rule is not satisfied, the next set of iterations of the EM algorithm using the higher value of A is computed until either the stopping rule given by DECISWIT

and STOPWIT is satisfied or maximum number of iterations given by MAXITER is reached. The process of increasing the A value continues until either the stopping rule given by DECISBET and STOPBET for two consequent iterations with two different values of A is met or the maximum number of iterations given by the MAXITER option is reached or the next value of A is higher than AMAX.

Note, that increasing the A value can, unfortunatelly, run up against the possibilities of the computer. Hence, in practice, it is not possible to increase the value of A up to an arbitrary level. That is why, the value of the AMAX option should be chosen with respect to the size of the data set and to the computer power.

Be also aware that each increase of the A value provides better approximation to the objective function Q. Hence, after performing the M step of the EM algorithm, the new estimates are closer to the real maximizer of this function than these obtained using the smaller value of A. So that, one can expect that after using the higher value of A and computing one iteration of the EM algorithm, the objective function increases more than during several last iterations of the EM algorithm with the smaller value of A. That is why, one has to find a compromise value which the objective function Q can increase by or which the parameter estimates can change by, to be satisfied with when computing an additional iteration of the EM algorithm with increased A constant. This is also the reason why two different stopping rules given by the DECISWIT, STOPWIT or DECISBET, STOPBET options respectively, are used for comparing two iterations with common or different values of A. Do not forget that the estimates provided by the macro are always based on an approximation.

Starting multiplication factor AMIN=10 is a default value for the macro HetMixed. If no AMAX option is specified, it is assumed that AMAX=AMIN and hence only one set of iterations of the EM algorithm with $A =$ AMIN is performed until the stopping rule given by DECISWIT, and STOPWIT is satisfied or the maximum number of iterations given by MAXITER is reached. The default value of the ABY option is equal to 10.

3.3.13 The DECISWIT and DECISBET Options

These two options determines the type of the stopping rule of the EM algorithm that is to be used. The DECISION given by the DECISWIT *("decision within")* option is used for comparing of two iterations with common value of A. The DECISBET *("decision between")* option specifies the DECISION for comparing of two consequtive iterations of the EM algorithm with different A values.

Three possibilities indicated by numbers $1, 2, 3$ are available.

DECISION=1

The EM algorithm (either one of its sets with a specific value of A or the all iteration process) stops if

$$\left|Q\big(\boldsymbol{\theta}^{(t+1)}\big|\boldsymbol{\theta}^{(t)}\big) - Q\big(\boldsymbol{\theta}^{(t)}\big|\boldsymbol{\theta}^{(t-1)}\big)\right| < \varepsilon$$

for two consecutive iterations. DECISWIT=1 and DECISBET=1 are also the default values for these options.

DECISION=2

The EM algorithm (either one of its sets with a specific value of A or the all iteration process) stops if the **average** absolute difference between estimates of all parameters that are to be estimated in its two consecutive iterations is smaller than prespecified ε.

DECISION=3

The EM algorithm (either one of its sets with a specific value of A or the all iteration process) stops if the **maximal** absolute difference between estimates of all parameters that are to be estimated in its two consecutive iterations is smaller than prespecified ε.

The value of ε is specified by the STOPWIT or STOPBET options respectively. The author of the macro recommends (according to his empirical experience) to use, at least for the within comparison, the default value of the decision rule. The reasons are as follows. The convergence of the EM algorithm can be painfully slow and parameter estimates can show negligible changes in some phases of the computational

process. On the other hand, the objective function Q changes usually at least a bit during all iterations of the EM algorithm.

3.3.14 The STOPWIT and STOPBET Options

These options define the value of ε as indicated in the DECISWIT and DECISBET options section. All positive numbers are allowed. If no STOPWIT is specified, ε for the within comparison of two consecutive iterations of the EM algorithm with common A values is assumed to be equal to 10^{-5}. If no STOPBET is specified, ε for the between comparison of two consecutive iterations of the EM algorithm with two different values of A is assumed to be equal to 10^{-4}.

3.3.15 The MAXITER Option

The MAXITER option specifies the maximum number of iterations of the EM algorithm, i.e. the EM algorithm stops when this maximum number of iterations is reached and the stopping rule defined by the DECISWIT and STOPWIT or DECISBET and STOPBET options are not taken into account. The value of MAXITER should be a positive integer. If no MAXITER option is specified, the EM algorithm stops after 100 iterations. The author recommends to choose maximum number of iterations according to a concrete situation. In practice, the value of 100 may not always be sufficient.

3.3.16 The Summary of the Stopping Rules

Let us summarize all possibilities for finishing the iterative process of estimates computation. This will finish if at least one of the following conditions is met.

- the two consecutive iterations of the EM algorithm with different values of A, which differ by ABY, satisfy the rule given by the DECISBET and STOPBET options;
- the two consecutive iterations of the EM algorithm with common value of A satisfy the rule given by the DECISWIT and STOPWIT options and $A +$ ABY is higher than AMAX;
- the number of all performed iterations of the EM algorithm reached its maximum value given by the MAXITER option.

Note that if two consecutive iterations of the EM algorithm with common value of A satisfy the rule given by the DECISWIT and STOPWIT options and $A +$ ABY is at most AMAX, at least one additional iteration of the EM algorithm with $A +$ ABY is computed.

As already mentioned, if the user wants to compute only one set of the iterations of the EM algorithm with one value of A specified by the AMIN option, the AMAX option has to be set to the same value as the AMIN option. This choice for the AMAX option is also its default value.

3.3.17 The MIXEDMI Option

The MIXEDMI option specifies the maximum number of iterations of the Newton-Raphson algorithm used by the procedure PROC MIXED to compute the updated estimates of the parameter $\boldsymbol{\gamma}$ within each step of the EM algorithm. The value of this option is used in the MAXITER statement of PROC MIXED. Smaller values of the MIXEDMI can, sometimes, shorten the computational time. The default value of this option is 50 which is also the default value of the procedure PROC MIXED.

3.3.18 The INITPOST Option

The INITPOST option specifies a SAS data set with initial posterior probabilities if the user wants to use its own initial estimates instead of these generated randomly as described in the PIEPS option section. The data set specified in this option has to contain one variable of the same name as specified in the SUBJECT option and G variables named as POST1,POST2, ... ,POSTG. Values of SUBJECT variable have to be mutually different and have to be the same as the different values of the SUBJECT variable of the data set specified by the DATA option. Observations of the INITPOST data set correspond to the initial values $p_{i1}(\boldsymbol{\theta}^{(0)}), \ldots, p_{ig}(\boldsymbol{\theta}^{(0)})$ of posterior probabilities for the ith subject where i is specified by the value of the variable SUBJECT. So that, the INITPOST data set contains $g + 1$ variables and N observations

when using notation introduced in (2.1) and (2.2). If there is a data set specified by the `INITPOST` option, initial estimates are always based on this data set and they are never generated at random. Be aware of the fact that it is the user's responsibility to ensure that all initial posterior probabilities lie between zero and one and the sum of them is equal to one for each subject.

3.3.19 The ENDPOST Option

The posterior probabilities computed after the last iteration of the EM algorithm are saved in the data set specified by the `ENDPOST` option if the user wishes to save them. The structure of the created data set is exactly the same as the structure requested by the `INITPOST` option.

3.3.20 The EB Option

The empirical Bayes estimates of the random effects computed according to (2.21) can be saved in the data set specified by the `EB` option if the user wishes to know them. Created data set contains one variable that is the same as that one specified by the `SUBJECT` option and q variables corresponding to random effects in the model. Labels of these variables are the same as effects specified by the `RANDOM` option. The data set contains N observations and one observation corresponds to the EB estimate $\hat{\boldsymbol{b}}_i$ of the random effect for the ith subject where i reflects the value of the `SUBJECT` variable for that observation. Numbers q and N are defined as indicated by (2.1).

3.3.21 Common Notes Related to the DATA, INITPOST, ENDPOST and EB Options

The user is encouraged to specify data sets from the SAS library `WORK` in above mentioned options. Specifying data sets from other SAS libraries named as *library.dataset* can cause problems in the procedure `PROC IML` that is used several times inside the macro.

3.3.22 The PIEPS Option

The `PIEPS` option influences the way of random generating initial posterior probabilities. It should be a positive number. Let us denote the number specified by this option as $\varepsilon(\pi)$. Let u_{ij} be random numbers generated from the uniform distribution on interval $(0, 1)$, $i = 1, \ldots, N$, $j = 1, \ldots, g$, where N and g have the same meaning as specified by (2.1) and (2.2). Further, let

$$\nu = \frac{1}{1 + \varepsilon(\pi)} \cdot \frac{1}{g}.$$

It can be easily found that ν can lie in the interval $(0, 1/g)$. In the next step, numbers $\tilde{p}_{ij}$ are computed by

$$\tilde{p}_{ij} = \frac{1}{g} + \nu\,(2u_{ij} - 1).$$

One can find that the closer $\varepsilon(\pi)$ is to zero, the more variable $\tilde{p}_{ij}$ will be. On the other hand, high values of $\varepsilon(\pi)$ force $\tilde{p}_{ij}$ to be close to $1/g$. Neverthelles, all values of $\varepsilon(\pi)$ ensure that $\tilde{p}_{ij}$ lies in the interval $(0, 2/g)$. Initial posterior probabilities are then computed as

$$p_{ij}(\boldsymbol{\theta}^{(0)}) = \frac{\tilde{p}_{ij}}{\sum_{k=1}^{g} \tilde{p}_{ik}}.$$

The default value of `PIEPS` is 0.1.

It should be mentioned that if all initial posterior probabilities are equal to $1/g$, the EM algorithm finishes after the first two iterations since exactly the same estimates are obtained in these two iterations. Similarly, convergence problems can occur if all initial probabilities are only close to $1/g$. That is why, the author of the macro does not recommend to use too high values of `PIEPS`.

3.4 The Log Window

Since the computation of the estimates using the EM algorithm can take some time even for quite small data sets, the information containing few facts about the state of the iteration process appears in the Log Window after each iteration. Reported information looks similarly like the following rows.

****

EM algorithm reached the iteration number 56.
Currently used value of A is equal to 30.

The value of the objective function in the previous iteration
was equal to -182.8121186.

Difference between values of objective function in previous two iterations
was equal to 0.000015.

Average absolute difference between estimates in previous two iterations
was equal to 0.0000123.

Maximal absolute difference between estimates in previous two iterations
was equal to 0.000106.

****

The first message informs about the number of the iteration step that is performed and about currently used value of the multiplication factor A.. The value of the objective function in the previous iteration shows the value of the Q function defined by (2.10) reached after the M step of the previous iteration of the EM algorithm. So that, in this case,

$$Q(\boldsymbol{\theta}^{(55)}|\boldsymbol{\theta}^{(54)}) = -182.8121186.$$

Next three messages show the state of all decision rules defined in the `DECISION` option section. So that, above shown example reflects the following situation.

$$Q(\boldsymbol{\theta}^{(55)}|\boldsymbol{\theta}^{(54)}) - Q(\boldsymbol{\theta}^{(54)}|\boldsymbol{\theta}^{(53)}) = 0.000015,$$
$$\text{the average of the elements in } |\boldsymbol{\theta}^{(55)} - \boldsymbol{\theta}^{(54)}| = 0.0000123,$$
$$\text{the maximum of the elements in } |\boldsymbol{\theta}^{(55)} - \boldsymbol{\theta}^{(54)}| = 0.000106.$$

Be aware of the fact that the quantity $Q(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^{(t-1)}) - Q(\boldsymbol{\theta}^{(t-1)}|\boldsymbol{\theta}^{(t-2)})$ can sometimes be negative even though one of the properties of the EM algorithm ensures that the objective function increases after each iteration. This is given by the fact that the maximization of the objective function is only approximate in our case.

The information about used number of iterations of the EM algorithm appears in the Log Window after the last one provided the convergence was reached, i.e. the iteration process finished due to the fact that the stopping rule defined by options `DECISBET` and `STOPBET` was satisfied. The message is similar to that following.

Convergence reached after 61 iterations.

If the maximum number of iterations of the EM algorithm was reached, the iteration process stops even if the stopping rule is not satisfied and the following message appears in the Log Window.

No convergence. Maximum number of iterations was reached.

If the value of the A constant that is to be used after obtaining "within" convergence given by the `DECISWIT` and `STOPWIT` options is higher than `AMAX,` the iteration process stops and the following message appears in the Log Window.

No convergence. Maximum A was reached.

In the case the user wants to save empirical Bayes estimates of random effects in a data set specified by the `EB` option, these have to be computed. This process lasts approximatelly the same time as one iteration of the EM algorithm. The following message informs the user about the fact that EB estimates are computed.

****

Empirical Bayes estimates of random effects are computed.

****

These were messages provided by the macro `HetMixed` itself. Lots of other messages produced by all procedures used within the macro appear also in the Log Window. The all produced messages should be of blue or black colour if there are no problems. Green or even red messages can appear as a consequence of errors caused by `PROC MIXED` that is used within the M step of each iteration. Among the most common problems that `PROC MIXED` is responsible for, we can assign the following ones. The maximization algorithm used by this procedure does not converge or this algorithm converge but final Hessian of the maximized function is not positive definite. These facts can lead into the situation that not all requested output is created by `PROC MIXED` and the EM algorithm cannot successfully continue in its work. Unfortunatelly, the macro `HetMixed` not always recognizes these problems and continues in the computation until either the stopping rule is satisfied or the maximum number of iterations is reached. But reported estimates (if there are some) are not valid then. The user is encouraged to check the Log Window and not to believe in reported estimates in the case of appearance of green or red messages. On the other hand, above mentioned problems are usually caused by too complicated covariance structures, especially too complicated covariance structure of residual components (i.e. of Σ_i matrices) and most of them can be solved by requesting some simpler covariance structure. Moreover, these problems nearly always appear already in the first iteration of the EM algorithm. So that it is pertinent to run the macro first with small maximal number of iterations to find whether it is possible to fit the requested covariance structure and subsequently, if no problems appear, to run the macro once more, to really compute the estimates. If there are problems, simpler covariance structures can be tried. The combination of the options `TYPEREP = simple, TYPERAND = un` should work nearly always according to the empirical experience of the author of the macro.

3.5 The Output of the Macro

The output described in this section is provided by the macro. I also refer to the illustration examples for better understanding of the output. Each part of the macro corresponds to one created data set that is saved in the SAS library `WORK`. I state also the names of these data sets to enable the user to use them if necessary.

An important property of the macro should be also mentioned. Since I did not want to overfill the Out Window by the output produced by the `PROC MIXED` within each iteration of the EM algorithm, the SAS statement `ods listing close` is used within the macro to inhibit the production of this output. If the macro normally runs till the end, the output production is restored by the statement `ods listing.` However, if the macro work is interruped by the user (e.g. if the user thinks that the iteration process lasts too long), the latter SAS statement is not processed and consequently no output is displayed in the Out Window. This can be remedy, if the user himself/herself proceeds the SAS statement `ods listing` after the violent interruption of the macro.

3.5.1 Initial Posterior Probabilities

Posterior probabilities used for starting the EM algorithm and either created as decribed in the `PIEPS` option section or given by the user as described in the `INITPOST` option section are reported in the first part of the output. They are reported in the form of the data set as required by the `INITPOST` option. Initial posterior probabilities shown in this section of the output are saved in the data set `_postout1`.

3.5.2 Iteration History

Substantial information about all iterations of the EM algorithm is shown in this part of the output. Seven columns labeled as Obs, Q1, Q2, Q, QDiff, AverPDif and MaxPDif are reported. Their meaning is as follows.

- **Obs.** The number of iteration of the EM algorithm, further in the text denoted as t;
- **Q1.** The value of the first part of the objective function (see (2.12)) after t iterations of the EM algorithm. I.e.

$$\mathsf{Q1} = Q_1(\boldsymbol{\pi}^{(t)}|\boldsymbol{\theta}^{(t-1)});$$

- **Q2.** The value of the second part of the objective function (see (2.13)) after t iterations of the EM algorithm. I.e.

$$\mathsf{Q2} = Q_2(\boldsymbol{\gamma}^{(t)}|\boldsymbol{\theta}^{(t-1)});$$

- **Q.** The value of the objective function (see (2.10)) after t iterations of the EM algorithm. I.e.

$$\mathsf{Q} = Q(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^{(t-1)}) = Q_1(\boldsymbol{\pi}^{(t)}|\boldsymbol{\theta}^{(t-1)}) + Q_2(\boldsymbol{\gamma}^{(t)}|\boldsymbol{\theta}^{(t-1)});$$

- **QDiff.** The difference between the values of the objective function in the current and previous iterations. I.e.

$$\mathsf{QDiff} = Q(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^{(t-1)}) - Q(\boldsymbol{\theta}^{(t-1)}|\boldsymbol{\theta}^{(t-2)}).$$

  As already stated in the previous section, this difference is not necessarily positive despite the properties of the EM algorithm since we only maximize an approximation to the objective function at each M step of the EM algorithm;
- **AverPDif.** The average of absolute values of differencies between parameter estimates in the present and previous iterations. I.e.

$$\mathsf{AverPDif} = \text{the average of the elements in the vector } |\boldsymbol{\theta}^{(t)} - \boldsymbol{\theta}^{(t-1)}|;$$

- **MaxPDif.** The maximum of absolute values of differencies between parameter estimates in the present and previous iterations. I.e.

$$\mathsf{MaxPDif} = \text{the maximum of the elements in the vector } |\boldsymbol{\theta}^{(t)} - \boldsymbol{\theta}^{(t-1)}|.$$

The iteration history is stored in the data set `_history`.

3.5.3 Final Likelihood and Log-Likelihood

The likelihood of the observed data computed as indicated by (2.5), the log-likelihood computed according to (2.7) and evaluated in $\boldsymbol{\theta}$ obtained after the last iteration of the EM algorithm along with its value multiplied by two are reported in this part of the output. Table with columns labeled as _Likelihood_, _Log_Likelihood_ and _2Log_Likelihood_ appears in the Output Window. The data set containing above mentioned values is called `_Lhood`.

3.5.4 Estimates of Component Probabilities

The estimates of component probabilities $\pi_1, \ldots, \pi_g$ defined by (2.2) can be found under labels Pl1, ... ,Plg in this section of the output. The name of the data set where this information is saved, is `_piset`.

3.5.5 Overall Estimates of Component Means

This section reports the estimates of parameters

$$\boldsymbol{\delta}_j = \boldsymbol{\beta}^R + \boldsymbol{\mu}_j, \qquad j = 1, \ldots, g,$$

and parameter $\boldsymbol{\beta}^F$, where $\boldsymbol{\beta}^F$, $\boldsymbol{\beta}^R$ and $\boldsymbol{\mu}_j$ are defined by (2.1) and (2.2) A table labeled by Effect and Estimate is shown in the output. Index j indicating the component of the mixture is reflected by a number added to the name of the effect. The appropriate data set is called `_Muset.`

3.5.6 Estimates of Beta Parameters

Estimates of parameters $\boldsymbol{\beta}^F$ and $\boldsymbol{\beta}^R$ defined by (2.1) are reported in the same form as estimates of $\boldsymbol{\delta}_j$ parameters in this section of the output. The name of the corresponding data set is `_Betaset.`

3.5.7 Estimates of Mu Parameters

Estimates of parameters $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_g$ defined by (2.2) are shown again in the same form as estimates of $\boldsymbol{\delta}_j$ parameters under above mentioned title of the output. The corresponding data set is `_reMuset.`

3.5.8 Estimates of Elements of Sigma Matrices

This section shows estimates of elements of Σ_i matrices defined by (2.1). Two columns labeled as CovParm and Estimate can be found here. The values of the column CovParm denote the elements of the estimated matrix in the same way as used by the output of `PROC MIXED` for the requested type of the covariance structure. The appropriate data set is called `_Sestset.`

3.5.9 Estimates of Elements of D Matrix

Estimates of the elements of the $\mathbb{D}$ matrix defined by (2.2) are reported in the same spirit as the estimates of elements of Σ_i matrices in this part of the output. The name of the corresponding data set is `_Destset.`

3.5.10 Final Posterior Probabilities

Posterior probabilities computed after the last iteration of the EM algorithm are not reported automatically. They can only be saved in the data set specified by the `ENDPOST` option. The section devoted to this option also describes the form of created data set.

3.5.11 Empirical Bayes Estimates of Random Effects

Empirical Bayes estimates of random effects computed by (2.21) are not reported automatically as well. They can be saved in the data set specified by the `EB` option. The form of created data set is described in the `EB` option section.

3.6 `Out of Memory` Appeared in the Log Window

The most common interruption of the computation process is caused by insufficient memory supply. A message `Out of Memory` given by the `PROC MIXED` used within the M step of the EM algorithm appears in the Log Window in such cases. In fact, memory is usually full due to the fact that the SAS stores results from all invocation of the `PROC MIXED` during all iterations of the EM algorithm as the user can find in the filled Results Window. This can be quite memory consuming, especially if higher A value is used or if higher number of iterations is computed since `'VINV'` matrices for all subject of the **extended** data set are, among others, stored. Note that the number of subjects in the extended data set is approximatelly equal to A times the number of subjects in the original data set. I am aware of this shortcomming but was not able to prevent SAS from this useless information storing. If somebody

knows how to prevent the SAS from storing results, I can immediatelly improve the macro in this way. I will be greatful for each information leading to this improvement.

So that, if the message `Out of Memory` appears in the Log Window, the computation can be restored after performing the following steps.

- Submit in SAS the following statement.
  `ods listing;`
  This is necessary to restore printing of the output into the Output Window.

- If you want to see parameter estimates after the last successfully performed iteration, submit in SAS the following.
  `proc print data=_piset; run;`
  – to see estimates of $\pi_1, \ldots, \pi_g$;
  `proc print data=_Muset; run;`
  – to see estimates of the parameter $\boldsymbol{\beta}^F$ and overall means $\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_g$. Ignore columns labeled as `StdErr, DF, tValue, Probt` in the output;
  `proc print data=_CovPset`
  – to see estimates of $\mathbb{D}$ and Σ_i matrices.

- If you want to see the iteration history till the last successful iteration, submit the following.
  `proc print data=_history; run;`

- Store the posterior probabilities after the last successful iteration into your prespecified library, let say `mylib`.
```
data mylib.posprob;
  set _postset;
run;
```

- Close and again open the SAS System.

- Move the posterior probabilities obtained after the last successful iteration into the SAS `Work` library.
```
data posprob;
  set mylib.posprob;
run;
```

- Invoke again the `HetMixed` macro with `INITPOST=posprob` and `AMIN` according to your choice (probably with `AMIN` equal to the last used A before the computation crashed).

CHAPTER 4

Example: The Heights of Schoolgirls

As a first example of the use of heterogeneity models, we consider the growth curves of 20 preadolescent schoolgirls reported by Goldstein (1979, Table 4.3, p. 101). The height of girls was measured on a yearly basis from age 6 to 10. The measurements are given at exact years of age, some having been previously adjusted to these. Measurements reported by Goldstein for one of the girls are 114.5, 112, 126.4, 131.2, 135. The second measurement is obviously incorrect and was replaced by 122. Further, the girls were classified according to the height of their mother into three categories – *short mothers, medium mothers* and *tall mothers.* Data can be found in Table 4.1. The individual profiles are shown in Figure 4.1 and one can observe quite obvious difference among the groups, at least in average intercepts for each group.

The homogeneity linear mixed model, suggested by Verbeke and Molenberghs (2000, Section 12.7) for the data is of the form

$$\begin{aligned}\text{Height}_{ij} &= \beta_0 + \beta_1 \,\text{Small}_i + \beta_2 \,\text{Medium}_i + \\ &+ \{\beta_0 + \beta_1 \,\text{Small}_i + \beta_2 \,\text{Medium}_i\} \,\text{Age}_{ij} + \\ &+ b_{0i} + b_{1i} \,\text{Age}_{ij} + \\ &+ \varepsilon_{ij}, \qquad\qquad i = 1, \ldots, 20, \quad j = 1, \ldots, 5,\end{aligned}$$

where Small_i and Medium_i are dummy variables defined to be 1 if the mother of the ith girl is small, or medium, respectively, and defined to be 0 otherwise. The terms b_{0i} and b_{1i} can be characterized as random intercepts and random slopes, respectively and are assumed to have joint two dimensional normal distribution $N(\mathbf{0}, \mathbb{D})$ for each i.

The next paragraph can be motivated by the endeavour to show how the heterogeneity model can be used for the cluster analysis. The following heterogeneity linear mixed model, loosely connected to the previous homogeneity linear mixed model can be obtained by ignoring the group structure used so far,

FIGURE 4.1. The Heights of Schoolgirls. Growth curves of 20 schoolgirls from age 6 to 10, for girls with small, medium, or tall mothers.

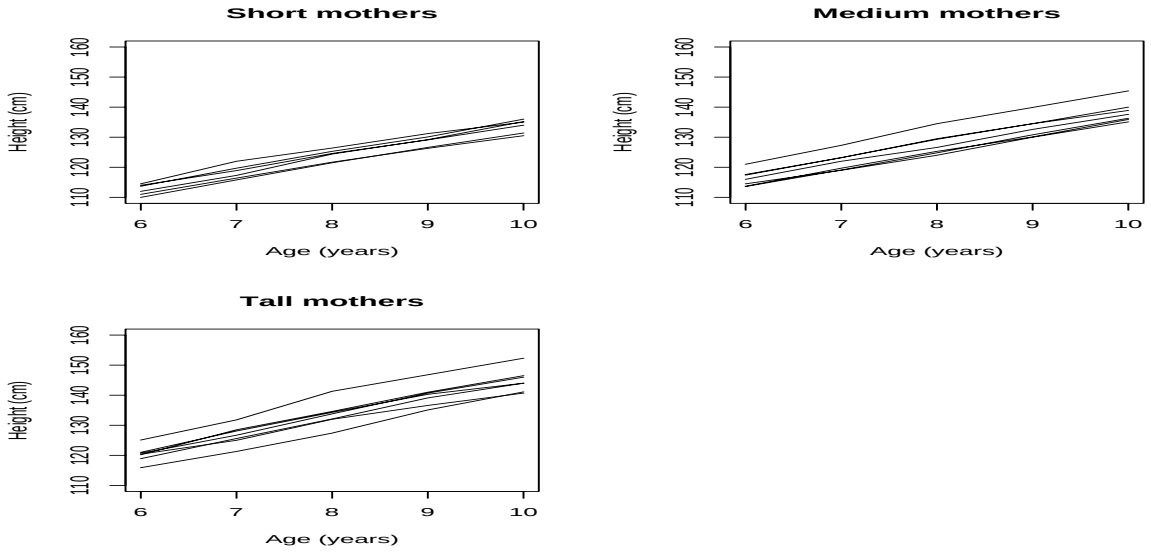


TABLE 4.1. The Heights of Schoolgirls.

Child	Type of mother	Height (cm) at the age of 6	7	8	9	10
1	small	111.0	116.4	121.7	126.3	130.5
2	small	110.0	115.8	121.5	126.6	131.4
3	small	113.7	119.7	125.3	130.1	136.0
4	small	114.0	118.9	124.6	129.1	134.0
5	small	114.5	122.0	126.4	131.2	135.0
6	small	112.0	117.3	124.4	129.2	135.2
7	medium	116.0	122.0	126.6	132.6	137.6
8	medium	117.6	123.2	129.3	134.5	138.9
9	medium	121.0	127.3	134.5	139.9	145.4
10	medium	114.5	119.0	124.0	130.0	135.1
11	medium	117.4	123.2	129.5	134.5	140.0
12	medium	113.7	119.7	125.3	130.1	135.9
13	medium	113.6	119.1	124.8	130.8	136.3
14	tall	120.4	125.0	132.0	136.6	140.7
15	tall	120.2	128.5	134.6	141.0	146.5
16	tall	118.9	125.6	132.1	139.1	144.0
17	tall	120.7	126.7	133.8	140.7	146.0
18	tall	121.0	128.1	134.3	140.3	144.0
19	tall	115.9	121.3	127.4	135.1	141.1
20	tall	125.1	131.8	141.3	146.8	152.3

Source: Goldstein (1979).

i.e. by pretending that the information about the type of the mother is missing. The model can then be specified as

$$\text{Height}_{ij} = \beta_0 + \beta_1 \,\text{Age}_{ij} + b_{0i} + b_{1i} + \varepsilon_{ij}, \tag{4.1}$$

where β_0 and β_1 denote the overall average intercept and linear age effect, respectively. Random intercepts and random slopes $\boldsymbol{b}_i = (b_{0i}, b_{1i})^T$ are now assumed to follow a mixture of two, three, or even more normal distributions with common covariance matrix $\mathbb{D}$ as in (2.2), i.e. $\boldsymbol{b}_i \sim \sum_{j=1}^{g} \pi_j \, N(\boldsymbol{\mu}_j, \mathbb{D})$ with $\boldsymbol{\mu}_j = (\mu_0^j, \mu_1^j)^T$. The structure of the matrix $\mathbb{D}$ will not be more closely specified. Vectors $\boldsymbol{b}_i$, $i = 1, \ldots, 20$ are assumed to be independent. All error components ε_{ij} are considered to be independent and normally distributed with mean zero and common variance σ^2, i.e. matrix Σ_i from (2.1) is equal to $\sigma^2 I_5$. No information about the type of the mother is used now. But after obtaining the estimates, the girls can be classified into one of two, three, or more groups (according to the fitted model). Classification rules can be based on posterior probabilities (2.11), i.e. the ith girl is classified into the jth population if the posterior probability of the ith girl belonging to the jth component of the mixture is maximal among all posterior probabilities for the ith girl.

Two heterogeneous models were fitted: with two and three components for the mixture. Initial estimates for the two-component model were generated randomly as indicated on the place where option `PIEPS` is described. The three-component model was fitted using the initial posterior probabilities derived from the final posterior probabilities of the two-component model. Subjects with one of the final posterior probabilities in the two-component model close to one received high initial posterior probability of belonging to one of the first two populations also in the three-component model. All other subjects

received high initial posterior probability of belonging to the third population in the three-component model.

Estimates for both the two and three-component models were computed with the value of the `AMIN` option equal to 30, `AMAX` option equal to 160 and `ABY` option equal to 10. It means that the approximation to the Q_2 function in the M steps of the EM algorithm started with $A = 30$ and after obtaining the convergence defined by `DECISWIT` and `STOPWIT`, the A value was increased by 10 and at least one iteration of the EM algorithm with this increased A value was computed. The overall convergence was evaluated by the rule defined by `DECISBET` and `STOPBET` options. If the overall convergence is not reached, the computation would stop after the last iteration of the EM algorithm with $A = 160$. Used values of `DECISWIT` and `DECISBET` equal to one reflect the fact that the convergence was always evaluated using the absolute difference of the values of the objective function Q in two consecutive iterations. The `STOPWIT` option equal to 0.000001 informs us that the EM algorithm with one value of A stops if the absolute difference between the values of the objective function Q is less than 10^{-6}. On the other hand, the overall convergence is obtained if the absolute value of the objective function Q in two consecutive iterations of the EM algorithm with different A values is less than 10^{-2} as driven by the `STOPBET` option.

The computation was performed on *Pentium IV, 2000 MHz, 512 MB RAM.* Approximatelly 7 seconds were needed to compute one iteration of the EM algorithm with $A = 30$, one minute for $A = 90$, two minutes for $A = 120$ and three minutes for $A = 150$. According to the empirical experience, the magnitude of the RAM memory is much more crucial factor for the computational time than the quality of the processor.

The SAS data set used for the computation was created in the following way.

```
data schgirls;
   input height child age group @@;
   agecat=age;
   int=1;
   cards;
111.0  1 6 1   116.4  1 7 1   121.7  1 8 1   126.3  1 9 1   130.5  1 10 1
...
125.1 20 6 3   131.8 20 7 3   141.3 20 8 3   146.8 20 9 3   152.3 20 10 3
; run;
```

The Two-Component Model

To compute the estimates for the two-component model, 245 iterations of the EM algorithm with initial $A = 30$ and subsequent 60 iterations with A that was sequentally increased by 10 from 40 to 120 were needed to satisfy the requested stopping rule. So that, totally 305 iterations of the EM algorithm were computed which took approximatelly 70 minutes on the above mentioned computer.

The parameter estimation was performed using the following syntax.

```
%HetMixed(DATA = schgirls,
          SUBJECT = child, REPEATED = agecat,
          RESPONSE = height, FIXED = , RANDOM = int age,
          TYPEREP = simple, TYPERAND = un,
          G = 2,
          AMIN = 30, AMAX = 160, ABY = 10,
          DECISWIT = 1, DECISBET = 1, STOPWIT = 0.000001, STOPBET = 0.01,
          MAXITER = 1000,
          PIEPS = 0.1, ENDPOST = dvapost);
```

The part of the output obtained by the macro `HetMixed` for the two-component model equals.

The HetMixed Macro
Initial posterior probabilities

Obs	child	POST1	POST2
1	1	0.83074	0.16926
2	2	0.71487	0.28513
3	3	0.67121	0.32879
4	4	0.69833	0.30167
5	5	0.59085	0.40915
6	6	0.58066	0.41934
7	7	0.77926	0.22074
8	8	0.10985	0.89015
9	9	0.57326	0.42674
10	10	0.71687	0.28313
11	11	0.59866	0.40134
12	12	0.51405	0.48595
13	13	0.65439	0.34561
14	14	0.53268	0.46732
15	15	0.60447	0.39553
16	16	0.82301	0.17699
17	17	0.57976	0.42024
18	18	0.11008	0.88992
19	19	0.41062	0.58938
20	20	0.48017	0.51983

The HetMixed Macro
Iteration History

Obs	A	Q1	Q2	Q	QDiff	AverPDif	MaxPDif
1	30	-13.6142	-169.197	-182.811	.	.	.
2	30	-13.6144	-169.184	-182.798	0.012355	0.000439022	0.004412
3	30	-13.6141	-169.167	-182.782	0.016884	0.000424211	0.002992
4	30	-13.6137	-169.153	-182.767	0.014986	0.000938052	0.009564
...	...	...	...	...	...	...	...
243	30	-12.5955	-157.592	-170.187	-0.00002	1.1102E-17	0.000002
244	30	-12.5956	-157.592	-170.187	-0.00000	2.2204E-17	0.000000
245	30	-12.5956	-157.592	-170.187	-0.00000	3.8858E-17	0.000000
246	40	-12.5956	-157.456	-170.051	0.13606	0.002576448	0.056155
...	...	...	...	...	...	...	...
302	110	-12.5053	-155.864	-168.369	-0.000027	5.5511E-18	0.000002439
303	110	-12.5054	-155.864	-168.369	-0.000003	5.5511E-18	0.000000248
304	110	-12.5054	-155.864	-168.369	-0.000000	5.5511E-18	0.000000025
305	120	-12.5054	-155.856	-168.361	0.008344	0.00013437	0.002517402

The HetMixed Macro
Final Likelihood and Log-Likelihood

Obs	Likelihood	Log Likelihood	2Log Likelihood
1	3.8688E-73	-166.736	-333.472

The HetMixed Macro
Estimates of Component Probabilities

Obs	PI1	PI2
1	0.68211	0.31789

The HetMixed Macro
Overall Estimates of Component Means

Obs	Effect	Estimate
1	INT1	82.8042
2	AGE1	5.3866
3	INT2	81.9404
4	AGE2	6.4139

The HetMixed Macro
Estimates of Beta Parameters

Obs	Effect	Estimate
1	INT	82.5296
2	AGE	5.7132

The HetMixed Macro
Estimates of Mu Parameter

Obs	Effect	Estimate
1	INT1	0.27461
2	AGE1	-0.32658
3	INT2	-0.58925
4	AGE2	0.70076

The HetMixed Macro
Estimates of Elements of Sigma Matrices

Obs	CovParm	Estimate
1	agecat	0.47595

The HetMixed Macro
Estimates of Elements of D Matrix

Obs	CovParm	Estimate
1	UN(1,1)	6.46005
2	UN(2,1)	0.12575
3	UN(2,2)	0.04332

The first part of the output shows the user the initial posterior probabilities generated at random as described in the `PIEPS` option section. Iteration history informs us, among others, that when using previously reported initial posterior probabilities, 245 iterations with $A = 30$ and a total of 305 iterations was needed to satisfy the requested stopping rule. If all iteration history for the EM algorithm with $A = 30$ had been shown, one would have found that the objective function increased several times only by less than 0.001 in not few consequtive iterations but then succeeded to increase more considerably. As an example, we can state

$$Q(\boldsymbol{\theta}^{(1)}|\boldsymbol{\theta}^{(0)}) = -182.811, \quad Q(\boldsymbol{\theta}^{(219)}|\boldsymbol{\theta}^{(218)}) = -181.908, \quad Q(\boldsymbol{\theta}^{(245)}|\boldsymbol{\theta}^{(244)}) = -170.187$$

which shows that 219 iterations were needed to increase the value of the objective function by less than one but then approximatelly 30 iterations were sufficient to increase this value by more than ten. This phenomenon was caused by initial estimates that were not good enough. Especially, the initial posterior probabilities for children number 8 and 16 reflect completely different component pertinence than the final posterior probabilities that are shown further in the text.

The next part of the output reports the values of the likelihood and log-likelihood of the fitted model defined by (2.5) and (2.7). So that

$$L^*(\hat{\boldsymbol{\theta}}|\boldsymbol{y}) = 3.8688 \cdot 10^{-73}, \quad l^*(\hat{\boldsymbol{\theta}}|\boldsymbol{y}) = -166.736, \quad 2\,l^*(\hat{\boldsymbol{\theta}}|\boldsymbol{y}) = -333.472.$$

Last sections of the output are devoted to the parameter estimates where parameters are defined by (2.1), (2.2) and (4.1). Hence, the estimates take the following values.

$$\hat{\pi}_1 = 0.6821, \qquad \hat{\pi}_2 = 0.3179,$$

$$\hat{\boldsymbol{\delta}}_1 = \begin{pmatrix} \hat{\beta}_0 + \hat{\mu}_0^1 \\ \hat{\beta}_1 + \hat{\mu}_1^1 \end{pmatrix} = \begin{pmatrix} 82.8042 \\ 5.3866 \end{pmatrix}, \qquad \hat{\boldsymbol{\delta}}_2 = \begin{pmatrix} \hat{\beta}_0 + \hat{\mu}_0^2 \\ \hat{\beta}_1 + \hat{\mu}_1^2 \end{pmatrix} = \begin{pmatrix} 81.9404 \\ 6.4139 \end{pmatrix},$$

$$\hat{\boldsymbol{\beta}}^R = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} 82.5296 \\ 5.7132 \end{pmatrix},$$

$$\hat{\boldsymbol{\mu}}_1 = \begin{pmatrix} \hat{\mu}_0^1 \\ \hat{\mu}_1^1 \end{pmatrix} = \begin{pmatrix} 0.27461 \\ -0.32658 \end{pmatrix}, \qquad \hat{\boldsymbol{\mu}}_2 = \begin{pmatrix} \hat{\mu}_0^2 \\ \hat{\mu}_1^2 \end{pmatrix} = \begin{pmatrix} -0.58925 \\ 0.70076 \end{pmatrix},$$

$$\hat{\sigma}^2 = 0.4760, \qquad \widehat{\mathbb{D}} = \begin{pmatrix} 6.4601 & 0.1258 \\ 0.1258 & 0.0433 \end{pmatrix}.$$

The posterior probabilities evaluated in $\hat{\boldsymbol{\theta}}$ can be obtained by using the procedure `PROC PRINT` with the data set `dvapost.` The following values are then shown.

child (i)	$p_{i1}(\hat{\boldsymbol{\theta}})$	$p_{i2}(\hat{\boldsymbol{\theta}})$	child (i)	$p_{i1}(\hat{\boldsymbol{\theta}})$	$p_{i2}(\hat{\boldsymbol{\theta}})$	child (i)	$p_{i1}(\hat{\boldsymbol{\theta}})$	$p_{i2}(\hat{\boldsymbol{\theta}})$
1	1.00000	0.00000	8	0.99782	0.00218	15	0.00158	0.99842
2	0.99972	0.00028	9	0.07034	0.92966	16	0.01069	0.98931
3	0.99716	0.00284	10	0.99985	0.00015	17	0.00299	0.99701
4	0.99998	0.00002	11	0.96908	0.03092	18	0.66270	0.33730
5	0.99997	0.00003	12	0.99769	0.00231	19	0.01514	0.98486
6	0.94410	0.05590	13	0.97693	0.02307	20	0.00001	0.99999
7	0.99867	0.00133	14	0.99935	0.00065			

The first component of the mixture can be characterized as girls who are taller at the age of 6 but the rate of their growth is not as rapid as the rate of growth of girls belonging to the second component of the mixture where the girls are smaller at the age of 6. The first group consists of girls number 1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 18, the second one of girls number 9, 15, 16, 17, 19, 20 if the classification rule is based on posterior probabilities obtained after the last iteration of the EM algorithm.

The Three-Component Model

Since the initial posterior probabilities for the three-component model were derived from the final posterior probabilities from the two-component model, they provided much better starting point for the EM algorithm than the random initial probabilities used in the previous case. Hence, only 55 iterations with initial $A = 30$ were needed to satisfy the requested "within" stopping rule. Nevertheless, next 198 iterations with higher A values were needed to obtain also the desired overall convergence. Moreover, alike the estimation process for the two-component model, the value $A = 160$ had to be used now. Since much more iterations with higher A values were used for the three-component model, the whole computation took about 5 hours compared to 70 minutes needed for the two-component model.

The following data set was used as the initial posterior probabilities for the three-component model.

```
data initdrie;
   input child post1 post2 post3 @@;
   cards;
1  0.80 0.10 0.10   2  0.75 0.10 0.15   3  0.77 0.12 0.11   4  0.76 0.09 0.15
5  0.78 0.06 0.16   6  0.30 0.05 0.65   7  0.79 0.11 0.10   8  0.74 0.10 0.16
9  0.05 0.75 0.20   10 0.78 0.03 0.19   11 0.65 0.05 0.30   12 0.80 0.05 0.15
13 0.75 0.04 0.21   14 0.84 0.01 0.15   15 0.05 0.75 0.20   16 0.02 0.74 0.24
17 0.08 0.70 0.22   18 0.10 0.10 0.80   19 0.06 0.50 0.44   20 0.05 0.60 0.35
; run;
```

The computation was performed using the following syntax.

```
%HetMixed(DATA = schgirls,
```

```
SUBJECT = child, REPEATED = agecat,
RESPONSE = height, FIXED = , RANDOM = int age,
TYPEREP = simple, TYPERAND = un,
G = 3,
AMIN = 30, AMAX = 160, ABY = 10,
DECISWIT = 1, DECISBET = 1, STOPWIT = 0.000001, STOPBET = 0.01,
MAXITER = 1000,
INITPOST = initdrie, ENDPOST = tripost);
```

The estimates obtained after the last iteration of the EM algorithm follow.

$$\hat{\pi}_1 = 0.49634, \qquad \hat{\pi}_2 = 0.29650, \qquad \hat{\pi}_3 = 0.20713,$$

$$\hat{\boldsymbol{\delta}}_1 = \begin{pmatrix} \hat{\beta}_0 + \hat{\mu}_0^1 \\ \hat{\beta}_1 + \hat{\mu}_1^1 \end{pmatrix} = \begin{pmatrix} 84.3682 \\ 5.3196 \end{pmatrix}, \qquad \hat{\boldsymbol{\delta}}_2 = \begin{pmatrix} \hat{\beta}_0 + \hat{\mu}_0^2 \\ \hat{\beta}_1 + \hat{\mu}_1^2 \end{pmatrix} = \begin{pmatrix} 81.6782 \\ 6.4558 \end{pmatrix},$$
$$\hat{\boldsymbol{\delta}}_3 = \begin{pmatrix} \hat{\beta}_0 + \hat{\mu}_0^3 \\ \hat{\beta}_1 + \hat{\mu}_1^3 \end{pmatrix} = \begin{pmatrix} 79.3721 \\ 5.6041 \end{pmatrix},$$

$$\hat{\boldsymbol{\beta}}^R = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} 82.5340 \\ 5.7154 \end{pmatrix},$$

$$\hat{\boldsymbol{\mu}}_1 = \begin{pmatrix} \hat{\mu}_0^1 \\ \hat{\mu}_1^1 \end{pmatrix} = \begin{pmatrix} 1.83417 \\ -0.39584 \end{pmatrix}, \qquad \hat{\boldsymbol{\mu}}_2 = \begin{pmatrix} \hat{\mu}_0^2 \\ \hat{\mu}_1^2 \end{pmatrix} = \begin{pmatrix} -0.86117 \\ 0.74041 \end{pmatrix},$$
$$\hat{\boldsymbol{\mu}}_3 = \begin{pmatrix} \hat{\mu}_0^3 \\ \hat{\mu}_1^3 \end{pmatrix} = \begin{pmatrix} -3.16193 \\ -0.11130 \end{pmatrix},$$

$$\hat{\sigma}^2 = 0.4771, \qquad \widehat{\mathbb{D}} = \begin{pmatrix} 2.6700 & 0.4079 \\ 0.4079 & 0.0303 \end{pmatrix}.$$

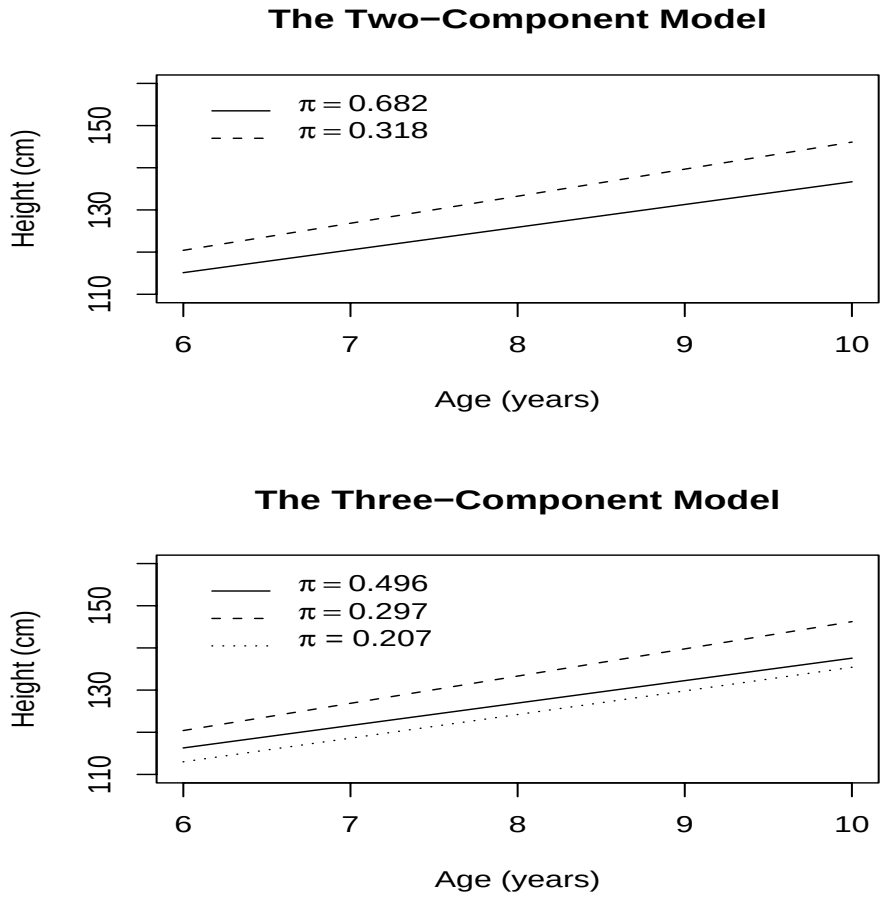
The likelihood (2.5) and the log-likelihood (2.7) of the fitted three-component model are reported below.

$$L^*(\hat{\boldsymbol{\theta}}|\boldsymbol{y}) = 9.68 \cdot 10^{-73}, \qquad l^*(\hat{\boldsymbol{\theta}}|\boldsymbol{y}) = -165.819.$$

The posterior probabilities obtained after the last iteration can be found in the following table.

child (i)	$p_{i1}(\hat{\boldsymbol{\theta}})$	$p_{i2}(\hat{\boldsymbol{\theta}})$	$p_{i3}(\hat{\boldsymbol{\theta}})$	child (i)	$p_{i1}(\hat{\boldsymbol{\theta}})$	$p_{i2}(\hat{\boldsymbol{\theta}})$	$p_{i3}(\hat{\boldsymbol{\theta}})$
1	0.96422	0.00000	0.03578	11	0.84347	0.01047	0.14606
2	0.12352	0.00040	0.87608	12	0.53061	0.00215	0.46723
3	0.48007	0.00282	0.51711	13	0.08032	0.02591	0.89377
4	0.98842	0.00000	0.01158	14	0.99973	0.00000	0.00027
5	0.99737	0.00000	0.00263	15	0.00011	0.99761	0.00228
6	0.00732	0.04581	0.94687	16	0.00046	0.98461	0.01494
7	0.95206	0.00020	0.04774	17	0.00020	0.99565	0.00415
8	0.98787	0.00010	0.01203	18	0.95559	0.02584	0.01857
9	0.07786	0.87519	0.04694	19	0.00002	0.96368	0.03629
10	0.93126	0.00003	0.06870	20	0.00000	0.99999	0.00001

FIGURE 4.2. The Heights of Schoolgirls. Estimated component means based on the two and three-component models



If the girls are to be classified according to these probabilities, one can find that the second component of the three-population model consists of exactly the same girls as the second component of the two-population model. The first component of this model is splited into two parts now, consisting of girls with numbers 1, 3, 4, 5, 7, 8, 10, 11, 12, 14, 18 and girls with numbers 2, 6, 13, respectively. One can find that currently created three groups of girls do not coincide completely with prior group structure given by the mother type. But there is no a priori reason why the mixture classification should exactly correspond to some predefined group structure. Estimated component mean profiles based on the two and three-component models can be found in Figure 4.2.

CHAPTER 5

Example: The Prostate Data

The second illustration of the heterogeneity linear mixed model is devoted to the so called *Prostate Data Set.* The same data set was also analyzed by Verbeke and Molenberghs (2000, Sections 12.1 and 12.6). *The Prostate Data Set* is an excellent example of the situation where classification of subjects based on longitudinal profiles is clearly of interest as will immediately be explained. This data set was also chosen to show some shortcomings of the macro `HetMixed` caused by insufficient computational power of available computers.

The prostate cancer has become one of the most common medical problems in the United States. As a consequence of this fact, lots of studies were performed to find markers which can detect the disease at an early stage. The prostate specific antigen (PSA) is assumed to be such a marker. PSA is an enzyme produced by both normal and cancerous prostate cells, and its level is related to the volume of prostate tissue. Still, an elevated PSA level is not necessarily an indicator of prostate cancer because patients with bening prostatic hyperplasia (BPH) also have an enlarged volume of prostate tissue and therefore also an increased PSA level. Based on clinical practice, researchers have hypothesized that the rate of change in PSA level might be a more accurate method of detecting prostate cancer in early stages of the disease. This has been extensively investigated by Pearson *et al.* (1994), who analyzed repeated PSA measures from the Baltimore Longitudinal Study of Aging (BLSA), using linear mixed models. The BLSA is a multidisciplinary observational study, which started in 1958, and with the study of normal human aging as primary objective. More information concerning this study can be found in Shock *et al.* (1984). Some of above mentioned repeated PSA measures are refered as *the Prostate Data Set* in this text. A description of the data, differentiating between healthy subjects (controls), BPH cases, local/regional (L/R) cancer cases and metastatic cancer cases is given in Table 5.1. The number of repeated PSA measurements per individual varies between 4 and 15, and the follow-up period ranges from 6.9 to 25.3 years. Measurements were taken approximatelly every second year. Since it was anticipated that PSA values would increase exponentially in prostate cancer cases, the responses were transformed to $\ln(PSA+1)$. Individual profiles can be found in Figure 5.1.

As previously stated, the rate of change of PSA can serve as a diagnostic tool for the prostate cancer. For such aims, the information containing the diagnostic group cannot be used and the goal is to classify subjects into one of these diagnostic groups and to base the classification rule on a subject's individual profile.

Verbeke and Molenberghs (2000, Section 12.1) suggest the following heterogeneity linear mixed model for the above mentioned purposes.

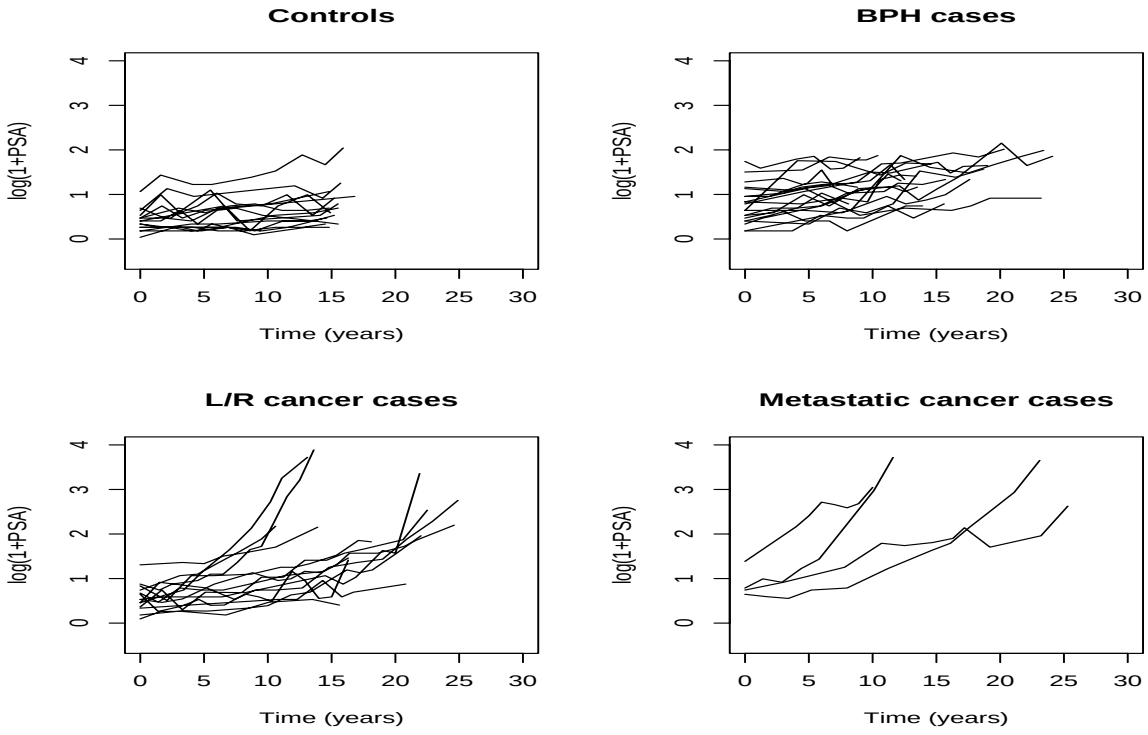
$$\ln(1+\text{PSA}_{ij}) = \beta_1 \text{Age}_i + (\beta_2 + b_{1i}) + (\beta_3 + b_{2i})t_{ij} + (\beta_4 + b_{3i})t_{ij}^2 + \varepsilon_{ij},$$

where Age_i is defined as the age of the ith subject at entry in the study (or at the time the first measurement was taken) and where the time points t_{ij} are expressed as time since entry in years. Other assumptions are derived from (2.1) and (2.2). I.e. vectors $\boldsymbol{b}_i = (b_{1i}, b_{2i}, b_{3i})^T$, $i = 1, \ldots, N$ are assumed to be independent with $\boldsymbol{b}_i \sim \sum_{j=1}^{g} \pi_j N(\boldsymbol{\mu}_j, \mathbb{D})$, $\boldsymbol{\mu}_j = (\mu_1^j, \mu_2^j, \mu_3^j)^T$. Residual components $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{in_i})^T$, $i = 1, \ldots, N$ are also assumed to be independent with $\boldsymbol{\varepsilon}_i \sim N(\boldsymbol{0}, \Sigma_i)$. The matrix Σ_i is considered to be equal to $\sigma^2 I_{n_i}$ (conditional independence).

TABLE 5.1. Prostate Data. Description of subjects included in the prostate data set, by diagnostic group. The cancer cases are subdivided into local/regional (L/R) and metastatic (M) cancer cases.

				Cancer Cases	
	Overall	Controls	BPH cases	L/R	M
Number of participants	54	16	20	14	4
Age at the entry					
Median	53.3	52.0	54.9	53.8	49.7
Range	38.9-75.7	42.2-65.9	47.1-75.7	38.9-67.6	44.4-69.9
Years of follow-up					
Median	15.4	15.1	14.3	17.2	17.4
Range	6.9-25.3	9.4-16.8	6.9-24.1	10.6-24.9	10.0-25.3
Number of measurements per individual					
Median	8	8	8	11	9.5
Range	4-15	4-10	5-11	7-15	7-12

FIGURE 5.1. Prostate Data. Individual profiles for men with prostate cancer, benign prostatic hyperplasia, or no evidence of prostate disease.



In order to enable the reader to compare the obtained results with these reported by Verbeke and Molenberghs (2000), I excluded the benign prostatic hyperplasia patients (BPH) from my analysis, yielding a total of 34 remaining patients. The two and three-component models were fitted. The initial estimates of the posterior probabilities were based on the knowledge of the real subject's status. Note that this is the information that is not generally available but can be helpful for finding the good initial posterior probabilities.

Estimates for both the two and three-component models were computed with the value of the `AMIN` option equal to 30, `AMAX` option equal to 110 and `ABY` option equal to 10. Again *Pentium IV, 2000 MHz, 512 MB RAM* was used for the computation. The value of 110 for `AMAX` was used due to the fact that one iteration of the EM algorithm lasted 30 minutes for the two-component model and 55 minutes for the three-component model and when trying to use $A = 120$, the first iteration was still under computation after 17 hours. The same decision rules as for the previous example were used. I.e. the absolute difference of the values of the objective function Q small enough led to the end of the iteration process (`DECISWIT=1, DECISBET=1)`. The meaning of the words small enough is precised by the values of the following two options: `STOPWIT=0.000001, STOPBET=0.01.`

It should be reported that for neither the two nor the three-component model, the overall convergence was obtained since maximal value of A was used and it was not possible, due to practical reasons caused by insufficient computer power, to increase its value and continue in the iteration process to obtain the convergence. That is why, the reported estimates have to be considered as not fully correct.

The SAS data set used for the computation can be created in the following way.

```
data prostate;
   input id age lnpsa time;
   int=1;
   timesq=time**2;
   timecat=time;
   cards;
88   47.3   0.470   0.0
88   47.3   0.693   3.0
88   47.3   0.588   5.1
...
; run;
```

The SAS data sets with the initial posterior probabilities for the two and three-component models were created as follows.

```
data inittwee;
   input id POST1 POST2;
   cards;
88    0.8881597   0.1118403
146   0.7971899   0.2028101
299   0.7985212   0.2014788
...
; run;

data initdrie;
   input id POST1 POST2 POST3;
   cards;
88    0.88146317   0.05689768   0.06163915
146   0.79479922   0.09849637   0.10670440
```

```
299   0.74352354   0.12310870   0.13336776
...
; run;
```

The two-component model can be fitted using the following syntax.

```
%HetMixed(DATA = prostate,
          SUBJECT = id, REPEATED = timecat,
          RESPONSE = lnpsa, FIXED = age, RANDOM = int time timesq,
          TYPEREP = simple, TYPERAND = un,
           G = 2,
           AMIN = 30, AMAX = 110, ABY = 10,
           DECISWIT = 1, DECISBET = 1, STOPWIT = 0.000001, STOPBET = 0.01,
           MAXITER = 1000,
           INITPOST = inittwee, ENDPOST = dvapost);
```

The estimates for **the three-component model** can be computed analogically, with proper changes of the `G, INITPOST` and `ENDPOST` options. To show how to compute empirical Bayes estimates of the random effects, the full syntax of the macro `HetMixed` for this model is reported below. Note that the empirical Bayes estimates are stored in a SAS data set specified by the `EB` option.

```
%HetMixed(DATA = prostate,
          SUBJECT = id, REPEATED = timecat,
          RESPONSE = lnpsa, FIXED = age, RANDOM = int time timesq,
          TYPEREP = simple, TYPERAND = un,
           G = 3,
           AMIN = 30, AMAX = 110, ABY = 10,
           DECISWIT = 1, DECISBET = 1, STOPWIT = 0.000001, STOPBET = 0.01,
           MAXITER = 1000,
          INITPOST = initdrie, ENDPOST = tripost, EB = trieb);
```

The Two-Component Model

The part of the output concerning the estimates of the unknown parameters obtained after the last iteration of the EM algorithm with $A = 110$ for the two-component model equals.

The HetMixed Macro
Final Likelihood and Log-Likelihood

Obs	_Likelihood_	_Log_ Likelihood_	_2Log_ Likelihood_
1	4.7999E-8	-16.8521	-33.7042

The HetMixed Macro
Estimates of Component Probabilities

Obs	PI1	PI2
1	0.77726	0.22274

The HetMixed Macro
Overall Estimates of Component Means

Obs	Effect	Estimate
1	AGE	0.01209
2	INT1	-0.1738
3	TIME1	0.02613
4	TIMESQ1	0.001211
5	INT2	0.2493
6	TIME2	-0.03894
7	TIMESQ1	0.009839

The HetMixed Macro
Estimates of Beta Parameters

Obs	Effect	Estimate
1	AGE	0.012088
2	INT	-0.079532
3	TIME	0.011634
4	TIMESQ	0.003133

The HetMixed Macro
Estimates of Mu Parameter

Obs	Effect	Estimate
1	INT1	-0.09424
2	TIME1	0.01449
3	TIMESQ1	-0.00192
4	INT2	0.32885
5	TIME2	-0.05057
6	TIMESQ1	0.00671

The HetMixed Macro
Estimates of Elements of Sigma Matrices

Obs	CovParm	Estimate
1	timecat	0.026791

The HetMixed Macro
Estimates of Elements of D Matrix

Obs	CovParm	Estimate
1	UN(1,1)	0.040924
2	UN(2,1)	0.010452
3	UN(2,2)	0.001477
4	UN(3,1)	-0.000507
5	UN(3,2)	-0.000027
6	UN(3,3)	0.000020

These sections of the output inform us about the parameter estimates obtained after the last iteration of the EM algorithm with $A = 110$. Their values equal.

$$\hat{\pi}_1 = 0.7773, \qquad \hat{\pi}_2 = 0.2227,$$

$$\hat{\boldsymbol{\delta}}_1 = \begin{pmatrix} \hat{\beta}_2 + \hat{\mu}_1^1 \\ \hat{\beta}_3 + \hat{\mu}_2^1 \\ \hat{\beta}_4 + \hat{\mu}_3^1 \end{pmatrix} = \begin{pmatrix} -0.1738 \\ 0.02613 \\ 0.001211 \end{pmatrix}, \qquad \hat{\boldsymbol{\delta}}_2 = \begin{pmatrix} \hat{\beta}_0 + \hat{\mu}_1^2 \\ \hat{\beta}_1 + \hat{\mu}_2^2 \\ \hat{\beta}_4 + \hat{\mu}_3^2 \end{pmatrix} = \begin{pmatrix} 0.2493 \\ -0.03894 \\ 0.009839 \end{pmatrix},$$

$$\hat{\beta}^F = \hat{\beta}_1 = 0.012088, \qquad \hat{\boldsymbol{\beta}}^R = \begin{pmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \end{pmatrix} = \begin{pmatrix} -0.079532 \\ 0.011634 \\ 0.003133 \end{pmatrix},$$

$$\hat{\boldsymbol{\mu}}_1 = \begin{pmatrix} \hat{\mu}_1^1 \\ \hat{\mu}_2^1 \\ \hat{\mu}_3^1 \end{pmatrix} = \begin{pmatrix} -0.09424 \\ 0.01449 \\ -0.00192 \end{pmatrix}, \qquad \hat{\boldsymbol{\mu}}_2 = \begin{pmatrix} \hat{\mu}_1^2 \\ \hat{\mu}_2^2 \\ \hat{\mu}_3^1 \end{pmatrix} = \begin{pmatrix} 0.32885 \\ -0.05057 \\ 0.00671 \end{pmatrix},$$

$$\hat{\sigma}^2 = 0.026791, \qquad \widehat{\mathbb{D}} = \begin{pmatrix} 0.040924 & 0.010452 & -0.000507 \\ 0.010452 & 0.001477 & -0.000027 \\ -0.000507 & -0.000027 & 0.000020 \end{pmatrix}.$$

The likelihood (2.5) and the log-likelihood (2.7) of the two-component model after the last iteration of the EM algorithm with $A = 110$ take the following values.

$$L^*(\hat{\boldsymbol{\theta}}|\boldsymbol{y}) = 4.80 \cdot 10^{-8}, \qquad l^*(\hat{\boldsymbol{\theta}}|\boldsymbol{y}) = -16.8521.$$

At the end of this paragraph, few facts concerning the iteration history should be mentioned. These are summarized in the Table 5.2 where the value of the objective function Q after the first and last iteration with each used A is reported along with the difference between the values of Q after the last iteration with smaller A and and the first iteration with higher A. The table was obtained from the `Iteration History` part of the output. Totally, 121 iterations of the EM algorithm were performed as one can find in this table.

TABLE 5.2. The Two-Component Model. A brief summary of the iteration history for the two-component model.

Iteration Number	A	Q	Q Difference
1	30	-46.1676	–
60	30	-22.2723	–
61	40	-21.8102	0.4621
70	40	-21.3362	–
71	50	-21.0230	0.3132
77	50	-20.7555	–
78	60	-20.5517	0.2039
85	60	-20.3329	–
86	70	-20.2007	0.1321
92	70	-20.1105	–
93	80	-20.0011	0.1094
100	80	-19.8986	–
101	90	-19.8169	0.0818
107	90	-19.7479	–
108	100	-19.6888	0.0591
114	100	-19.6389	–
115	110	-19.5919	0.0470
121	110	-19.5514	–

The Three-Component Model

The parameter estimates for the three-component model were received from the similar output as these for the two-component model. Even now, the real convergence was not obtained and below reported estimates cannot be considered as fully correct. They reflect the state of the iteration process after the last iteration of the EM algorithm with $A = 110$.

$$\hat{\pi}_1 = 0.8007, \qquad \hat{\pi}_2 = 0.0919, \qquad \hat{\pi}_3 = 0.1074$$

$$\hat{\boldsymbol{\delta}}_1 = \begin{pmatrix} \hat{\beta}_2 + \hat{\mu}_1^1 \\ \hat{\beta}_3 + \hat{\mu}_2^1 \\ \hat{\beta}_4 + \hat{\mu}_3^1 \end{pmatrix} = \begin{pmatrix} 0.4368 \\ 0.002200 \\ 0.001778 \end{pmatrix}, \qquad \hat{\boldsymbol{\delta}}_2 = \begin{pmatrix} \hat{\beta}_0 + \hat{\mu}_1^2 \\ \hat{\beta}_1 + \hat{\mu}_2^2 \\ \hat{\beta}_4 + \hat{\mu}_3^2 \end{pmatrix} = \begin{pmatrix} 1.1732 \\ 0.06549 \\ 0.000844 \end{pmatrix},$$

$$\hat{\boldsymbol{\delta}}_3 = \begin{pmatrix} \hat{\beta}_0 + \hat{\mu}_1^3 \\ \hat{\beta}_1 + \hat{\mu}_2^3 \\ \hat{\beta}_4 + \hat{\mu}_3^3 \end{pmatrix} = \begin{pmatrix} 0.6808 \\ -0.01390 \\ 0.01769 \end{pmatrix},$$

$$\hat{\beta}^F = \hat{\beta}_1 = 0.00078, \qquad \hat{\boldsymbol{\beta}}^R = \begin{pmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \end{pmatrix} = \begin{pmatrix} 0.53073 \\ 0.00629 \\ 0.00340 \end{pmatrix},$$

$$\hat{\boldsymbol{\mu}}_1 = \begin{pmatrix} \hat{\mu}_1^1 \\ \hat{\mu}_2^1 \\ \hat{\mu}_3^1 \end{pmatrix} = \begin{pmatrix} -0.09390 \\ -0.00409 \\ -0.00162 \end{pmatrix}, \qquad \hat{\boldsymbol{\mu}}_2 = \begin{pmatrix} \hat{\mu}_1^2 \\ \hat{\mu}_2^2 \\ \hat{\mu}_3^1 \end{pmatrix} = \begin{pmatrix} 0.64252 \\ 0.05920 \\ -0.00256 \end{pmatrix},$$

$$\hat{\boldsymbol{\mu}}_3 = \begin{pmatrix} \hat{\mu}_1^3 \\ \hat{\mu}_2^3 \\ \hat{\mu}_3^3 \end{pmatrix} = \begin{pmatrix} 0.15012 \\ -0.02019 \\ 0.01429 \end{pmatrix},$$

$$\hat{\sigma}^2 = 0.02703, \qquad \widehat{\mathbb{D}} = \begin{pmatrix} 0.032007 & 0.002102 & 0.000068 \\ 0.002102 & 0.002202 & -0.000115 \\ 0.000068 & -0.000115 & 0.000011 \end{pmatrix}.$$

The likelihood (2.5) and the log-likelihood (2.7) of the three-component model after the last iteration with $A = 110$ are reported below.

$$L^*(\hat{\boldsymbol{\theta}}|\boldsymbol{y}) = 36.2310, \qquad l^*(\hat{\boldsymbol{\theta}}|\boldsymbol{y}) = 3.5899.$$

As in the two-component model section, few facts concerning the iteration history can be mentioned. They are summarized in Table 5.3. The entries of this table are analogical to the entries of the Table 5.2. As one can find, by comparing Tables 5.2 and 5.3, we are quite close to the desired convergence (absolute difference between the values of the Q function in two consecutive iterations with smaller and higher A less than 0.01) in the case of the two-component model but still quite far in the case of the three-component model.

TABLE 5.3. The Three-Component Model. A brief summary of the iteration history for the three-component model.

Iteration Number	A	Q	Q Difference
1	30	-55.9586	–
48	30	-11.2952	–
48	40	-8.6048	2.6904
61	40	-7.5418	–
62	50	-5.6663	1.8755
74	50	-4.9217	–
75	60	-3.5105	1.4112
85	60	-3.0432	–
86	70	-1.9247	1.1185
95	70	-1.6306	–
96	80	-0.7587	0.8719
106	80	-0.4665	–
107	90	0.2190	0.6854
115	90	0.4304	–
116	100	0.9991	0.5686
123	100	1.1205	–
124	110	1.5961	0.4756
121	110	1.7115	–

TABLE 5.4. Prostate Data. Cross-classification of 34 patients according to the three-component mixture model and according to their true disease status.

	Mixture classification				
Disease	The two-component model		The three-component model		
status	1	2	1	2	3
Control	15	1	15	1	0
L/R cancer	10	4	10	1	3
Metast. cancer	2	2	2	1	1

The estimated component means and probabilities based on the reported estimates for the two and three-component heterogeneity models, not taking into account the effect of age are shown in Figure 5.2.

Since the true disease status of all subjects in the study is known, a cross-classification of the subjects according to the two or three-component model and according to this status can be done using posterior probabilities obtained after the last performed iteration of the EM algorithm. These are stored in the SAS data sets specified by the `ENDPOST` option of the macro. The result of this cross-classification is shown in Table 5.4. Except for one patient, all controls were classified in the first component for both the two and three-component models. However, quite huge ammount of cancer cases was also classified in the first component. The fact that the used estimates are not fully correct is not responsible for this phenomenon since the initial posterior probabilities moreless reflected the three components of the mixture created according to the disease status but were changed, during the iteration process, into the state described in Table 5.4. Some reasons why this is observed, can be found in Verbeke and Molenberghs (2000, Section 12.6). In any case, this example has shown that the mixture approach does not necessarily model what one might hope.

To illustrate the using of the `EB` option of the macro, histograms of the empirical Bayes estimates of the random effects for the three-component model obtained after the last iteration of the EM algorithm with $A = 110$ are reported in Figure 5.3. The plots were obtained by further work with the SAS data set `trieb`.

Some shortcomings of the macro were highlighted by this example. First of all, quite long computational time is needed to find the desirable estimates. It should be stated that most of this time is consumed by the procedure `PROC MIXED` within each iteration of the EM algorithm. Second, the satisfactory convergence cannot be, sometimes, obtained due to the insufficient computational possibilities, as the *Prostate Data Set* clearly illustrates.

FIGURE 5.2. Prostate Data. Estimated component means and probabilities based on the two and three-component heterogeneity models, not taking into account the effect of age.

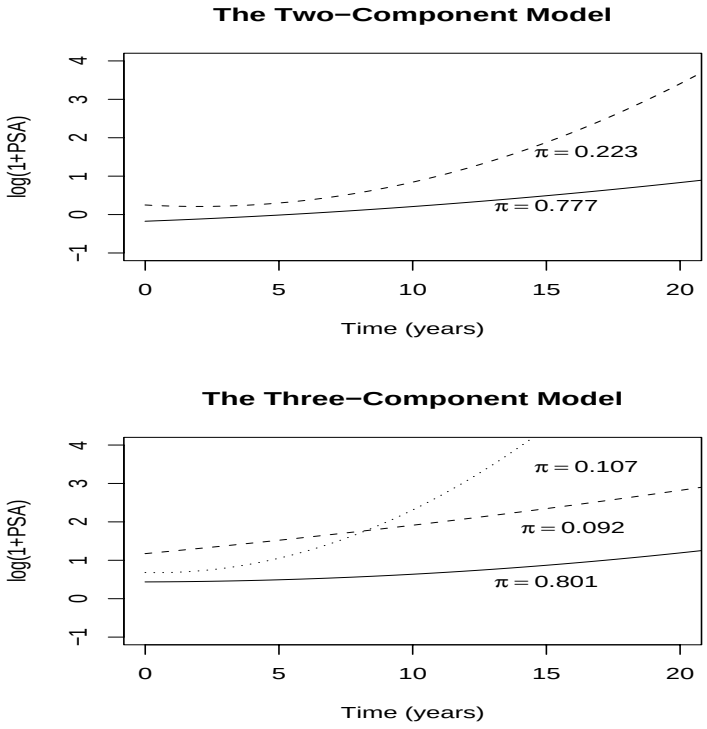
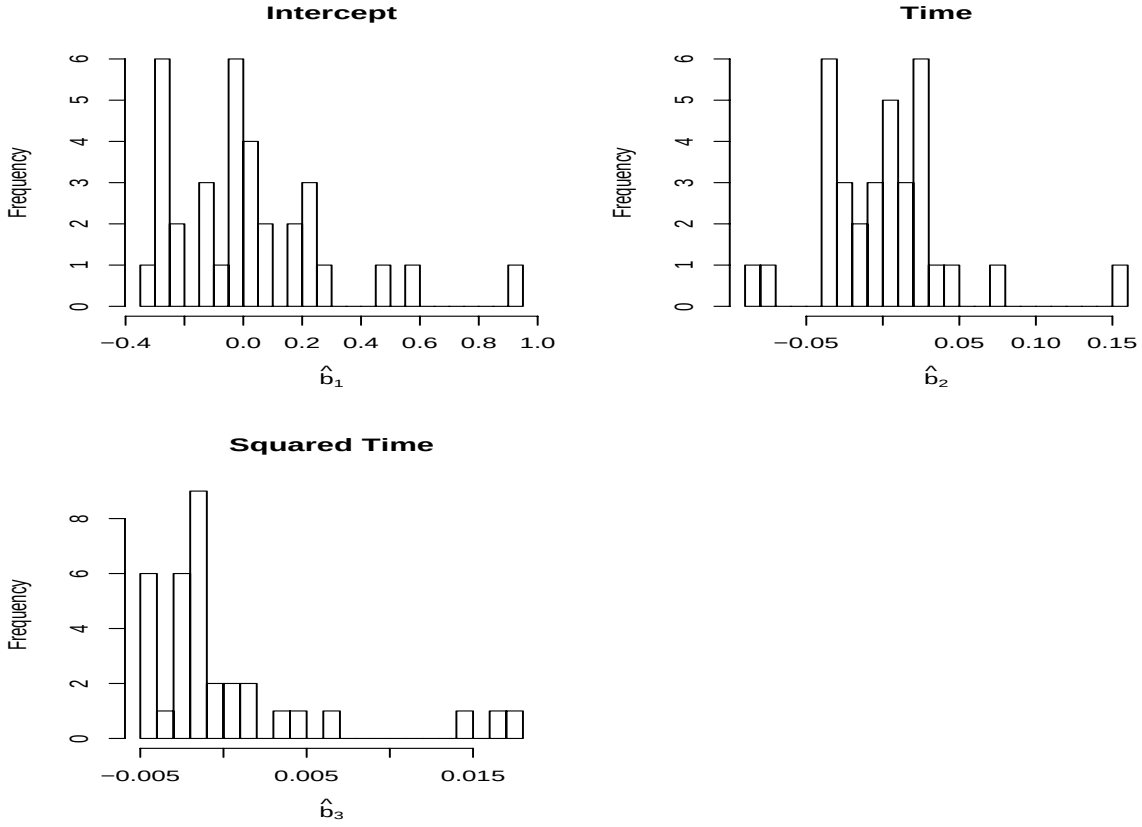


FIGURE 5.3. Prostate Data. Histograms of empirical Bayes estimates of random effects from the three-component heterogeneity model.



CHAPTER 6

Discussion

Modelling repeated measures by the homogeneity linear mixed model is not always the best thing one can do since the assumed normal distribution of random effects can 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 form models with many other distributions for random effects was therefore introduced. Desirable properties of the heterogeneity linear mixed model are given by the fact that the distribution of random effects is assumed to be a mixture of normal distributions which can approximate lots of other, commonly used continuous distributions. Note that the normality assumption for the random effects is violated, whenever a 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 is inhibited by insufficient software support. The SAS macro `HetMixed` created as the main objective of this project and described in this text, tries to fill in, at least a bit, this gap. As one could see, especially on the example with *the Prostate Data Set,* the macro still has some shortcomings, caused mainly by insufficient computational power of present computers. But this should continuously improve during following years.

References

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.

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.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (1995) *Bayesian Data Analysis,* Texts in Statistical Science. London: Chapman & Hall.

Ghosh, J.K., and Sen, P.K. (1985) On the asymptotic performance of the log likelihood ratio statistic for the mixture model and related results. In: *Proceedings of the Berekely Conference in Honor of Jerzy Neyman and Jack Kiefer,* Vol. 2, L.M. Le Cam and R.A. Olshen (Eds.). Monterey: Wadsworth, Inc., pp. 789-806.

Goldstein, H. (1979) *The Design and Analysis of Longitudinal Studies.* London: Academic Press.

Komárek, A. (2001) *A SAS Macro for Linear Mixed Models with Finite Normal Mixtures as Random-Effects Distribution.* M.S. thesis, Diepenbeek: Limburgs Universitair Centrum, Faculty of Science, Center for Statistics.

Louis, T.A. (1982) Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B,* **44,** 226-233.

Pearson, J.D., Morrell, C.H., Landis, P.K., Carter, H.B., and Brant, L.J. (1994) Mixed-effects regression models for studying the natural history of prostate disease. *Statistics in Medicine,* **13,** 587-601.

Shock, N.W., Greullich, R.C., Andres, R., Arenberg, D., Costa, P.T., Lakatta, E.G., and Tobin, J.D. (1984) *Normal Human Aging: The Baltimore Longitudinal Study of Aging.* National Institutes of Health Publication 84-2450. Washington, DC: National Institutes of Health.

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.

Verbeke, G., and Molenberghs, G. (2000) *Linear Mixed Models for Longitudinal Data.* New York: Springer-Verlag.