DOI:10.2478/bile-2013-0008

Biometrical Letters
Vol. 49 (2012), No. 2, 103-119

Finite mixture models with fixed weights applied to growth data

Marek Molas[1], **Emmanuel Lesaffre**[1,2]

[1]Department of Biostatistics, Erasmus MC, P.O. Box 2040, 3000 CA Rotterdam, Netherlands, e-mail: e.lesaffre@erasmusmc.nl

[2]Catholic University of Leuven, L-Biostat, U.Z. St. Rafael, Kapucijnenvoer 35, 3000 Leuven, Belgium

Summary

To model cross-sectional growth data the LMS method is widely applied. In this method the distribution is summarized by three parameters: the Box-Cox power that converts outcome to normality (L); the median (M); and the coefficient of variation (S).

Here, we propose an alternative approach based on fitting finite mixture models with several components which may perform better than the LMS method in case the data show an unusual distribution. Further, we explore fixing the weights of the mixture components in contrast to the standard approach where weights are estimated. Having fixed weights improves the speed of computation and the stability of the solution. In addition, fixing the weights provides almost as good a fit as when the weights are estimated. Our methodology combines Gaussian mixture modelling and spline smoothing. The estimation of the parameters is based on the joint modelling of mean and dispersion.

We illustrate the methodology on the Fourth Dutch Growth Study, which is a cross-sectional study that contains information on the growth of 7303 boys as a function of age. This information is used to construct centile curves, so-called growth curves, which describe the distribution of height as a smooth function of age. Further, we analyse simulated data showing a bimodal structure at some time point.

In its full generality, this approach permits the replacement of the Gaussian components by any parametric density. Further, different components of the mixture can have a different probabilistic (multivariate) structure, allowing for censoring and truncation.

Key words: mixture models, growth curves, splines, IWLS algorithm, flexible distributions

1. Introduction

The Fourth Dutch Growth Study is a cross-sectional study which recorded several variables for a sample of boys and girls, conducted in the Netherlands in 1997. Here we consider the height of 7303 boys, and we aimed to estimate centile curves of height as a function of age. The age ranges from 0.032 to 21.7 years, with a mean of 9.29 years. The median height is 145 cm with range 48.5–205.8 cm. Further details of the study can be found in Buuren and Fredriks (2001).

The standard method for estimating growth curves is the LMS method (Cole and Green, 1992). This method transforms the data to normality, and models the median, the coefficient of variation and the skewness as a smooth function of covariates, e.g. age. It performs well in most situations. However, when the data exhibits a special structure, such as bimodal, at some ages, and unimodal at other ages, the LMS method might not be optimal. Such data can arise when the total population divides into subgroups that have different growth patterns.

Another approach useful in the presence of mixtures is available in the R package **gamlss.mx** (Rigby and Stasinopoulos, 2005). This package is based on an extension of generalized additive models for location scale and shape to mixtures of distributions. Mixtures can be fitted such that means and standard deviations depend on age. Further, the weights of each mixture component are estimated.

We explore in this paper a simplified finite mixture modelling approach (McLachlan and Peel, 2000), where weights are fixed and equal. This reduces the computation time, and offers greater stability of the solution. The disadvantage of the simplified computational approach is a less than optimal fit. The observed loss of fit is often small, and can even be avoided by adding some extra mixture components. Further, by the addition of splines we can enable smooth change of density along the covariates, where means and variances of the component densities can be expressed as a non-linear function of covariates. Further, one can tune the approach with respect to the number of components, and the flexibility of splines in each part of the model. The choice of the final model is based on the Akaike Information Criterion (AIC).

In the next section we first review the estimation process of finite mixture models, using the standard case of Gaussian mixtures. Generalization to exponential family densities, multivariate distributions and censoring follows. In Section 3 we present an application of the modelling approach to the Fourth Dutch Growth data. A limited numerical study is performed in Section 4. Finally we give some concluding remarks.

2. Mixture models

2.1. Mixtures of Gaussian Distributions

Throughout this section we will use indices as follows: $k = 1, \ldots, K$ for the number of mixture components, $i = 1, \ldots, N$ for the number of observations, $p = 1, \ldots, P_\beta$ to index both fixed parameters in the mean structure and $p = 1, \ldots, P_\gamma$ for the parameters in the dispersion structure. A mixture model evaluated at data point y_i is given by:

$$g(y_i) = \sum_{k=1}^{K} w_k f_k(y_i), \tag{1}$$

with $f_k(y_i)$, $(k = 1, \ldots, K)$ the mixture components.

In a Gaussian mixture model each $f_k(y_i)$ is assumed to be a normal density with mean μ_k and variance $\phi_k \equiv \sigma_k^2$. Further, each mixture component $f_k(y_i)$ contributes to the total density with weight w_k. Suppose that the density described in (1) changes over a range of known factors (which may be continuous, such as age, or discrete, such as treatment). Denote by $\mathbf{z}_i$ a vector of covariates pertaining to observation y_i, where the location parameters μ_k are a function of $\mathbf{x}_i$, and the variances ϕ_k are a function of $\mathbf{r}_i$, with $\mathbf{x}_i$ and $\mathbf{r}_i$ subsets of $\mathbf{z}_i$, then we obtain the following model:

$$g(y_i|\mathbf{x}, \mathbf{r}) = \sum_{k=1}^{K} w_k f_k(y_i|\mathbf{x}_i, \mathbf{r}_i), \tag{2}$$

where $g(y_i|\mathbf{x}_i, \mathbf{r}_i)$ is the distribution of the response y_i. We assume that the parameters in each of the K components are allowed to be distinct. This is of interest for the Fourth Dutch Growth Study, as we wish to allow for a change in the distributional shape of the mixture along with covariates. The log-likelihood of all N subjects is

$$\begin{aligned} &\log L(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K, \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_K, w_1, \ldots, w_k; \mathbf{y}) = \\ &= \textstyle\sum_{i=1}^{N} \log \left[\sum_{k=1}^{K} w_k \cdot f_k(y_i|\mu_k(\mathbf{x}_i), \phi_k(\mathbf{r}_i)) \right], \end{aligned} \tag{3}$$

We distinguish three types of the parameters in the above likelihood: (1) location parameters $\mu_k(\mathbf{x}_i) = \mathbf{x}_i^T \boldsymbol{\beta}_k$, (2) scale parameters $\phi_k(\mathbf{r}_i) = exp(\mathbf{r}_i^T \gamma_k)$, and (3) weights w_k. The score equations with respect to $\boldsymbol{\beta}$ are expressed as follows:

$$\frac{\partial \log L}{\partial \beta_{kp}} = \sum_{i=1}^{N} \left(\frac{w_k \frac{\partial f_k(y_i|\mathbf{x}_i, \mathbf{r}_i)}{\partial \beta_{kp}}}{\sum_{k=1}^{K} w_k \cdot f_k(y_i|\mathbf{x}_i, \mathbf{r}_i)} \right) = \sum_{i=1}^{N} c_{ki} \frac{\partial \log(f_k(y_i|\mathbf{x}_i, \mathbf{r}_i))}{\partial \beta_{kp}}, \quad (4)$$

with $c_{ki} = \frac{w_k \cdot f_k(y_i|\mathbf{x}_i, \mathbf{r}_i)}{\sum_{k=1}^{K} w_k \cdot f_k(y_i|\mathbf{x}_i, \mathbf{r}_i)}$, and $p = 1, \ldots, P_\beta$. The same score equations are obtained for the parameters of the P_γ dimensional $\boldsymbol{\gamma}_k$ vectors with entries γ_{kp}, $p = 1, \ldots, P_\gamma$.

Similarly we can derive the score equations for weights w_k:

$$\frac{\partial \log L}{\partial w_k} = \sum_{i=1}^{N} c_{ki} \frac{\partial \log(w_k)}{\partial w_k} - \sum_{i=1}^{N} c_{Ki} \frac{\partial \log(w_k)}{\partial w_k}, \quad (5)$$

which are the score equations of a multinomial model where the parameters are w_k given the estimated posterior weights c_{ki}. When the weights do not depend on covariates the solution for w_k has a closed form, i.e.

$$\hat{w}_k = \frac{\sum_{i=1}^{N} c_{ki}}{N}. \quad (6)$$

However, when the weights are functions of covariates, iterative procedures are necessary.

We propose here to keep weights fixed and equal $w_k = w$. Then equations (5) do not need to be solved, but we estimate only location and scale parameters. This speeds up the convergence, as no iteration for weights is needed and fixing weights avoides problems such as infinite likelihood (McLachlan and Peel, 2000 pp.94-97). Thus fixing equal weights enables speeding up of the estimation and improvement of the numerical stability of the solution. We refer to this approach as the “FMIX” approach.

For $f_k(y_i|\mathbf{x}_i, \mathbf{r}_i)$ a Gaussian probability density function:

$$f_k(y_i|\mathbf{x}_i, \mathbf{r}_i) = \mathcal{N}_k(\mu_k(\mathbf{x}_i), \phi_k(\mathbf{r}_i)), \quad (7)$$

the score equations become

$$\frac{\partial \log L}{\partial \beta_{kp}} = \sum_{i=1}^{N} c_{ki} \left(\frac{y_i - \mu_{ki}}{\phi_{ki}} \right) x_{ip} = \sum_{i=1}^{N} \tilde{c}_{ki}(y_i - \mu_{ki}) x_{ip}, \quad (8)$$

$p = 1, \ldots, P_\beta$, resembling score equations of a normal distribution with unequal variances. The solution is provided by the weighted least squares procedure:

$$\hat{\boldsymbol{\beta}}_k = (\mathbf{X}^T \widetilde{\mathbf{C}}_k \mathbf{X})^{-1} \mathbf{X}^T \widetilde{\mathbf{C}}_k \mathbf{y}, \tag{9}$$

$k = 1, \ldots, K$, with $\widetilde{\mathbf{C}}_k$ a diagonal matrix with entries $\tilde{c}_{ki} = \frac{c_{ki}}{\phi_{ki}}$.

For the variance structure parameters $\boldsymbol{\gamma}_k$, the general framework was described by Nelder and Pregibon (1987). The score equations have the following forms:

$$\frac{\partial \log L}{\partial \gamma_{kp}} = \sum_{i=1}^{N} c_{ki} \left[-\frac{1}{2\phi_{ki}} + \frac{d_{ki}}{2\phi_{ki}^2} \right] \frac{\partial \phi_{ki}}{\partial \gamma_{kp}} = \sum_{i=1}^{N} \left[\frac{c_{ki}}{2} \frac{d_{ki} - \phi_{ki}}{\phi_{ki}^2} \right] \frac{\partial \phi_{ki}}{\partial \gamma_{kp}}, \tag{10}$$

$p = 1, \ldots, P_\gamma$, with d_{ki} the deviance residual, which for a normal distribution is equal to the squared residual $(y_i - \mu_{ki})^2$. Equation (10) is a score equation of a gamma generalized linear model (GLM), with response d_{ki}, prior weight $c_{ki}/2$, and mean ϕ_{ki} linked to the covariates. The parameters can be estimated by Iterative Weighted Least Squares (IWLS); see also Nelder and Wedderburn (1972).

The above two (I)WLS procedures can be combined into an interchangeable IWLS algorithm to find estimates of $\boldsymbol{\beta}_k$ and $\boldsymbol{\gamma}_k$ for each k^{th} component of the mixture. In total the following estimation procedure at iteration t is proposed:

1. Weights c_{ki} are computed given (starting) values of $\boldsymbol{\beta}_k(t-1)$ and $\boldsymbol{\gamma}_k(t-1)$
2. Each of K components is estimated by the above procedure yielding $\boldsymbol{\beta}_k(t)$ and $\boldsymbol{\gamma}_k(t)$
3. Iterate [1] and [2] until convergence

This approach can be proven to be equivalent to the EM-algorithm. Briefly, the E-step of the algorithm corresponds to finding the weights c_{ki}, while the M-step is solving equations (4) and (5) given the posterior weights c_{ki}. More details on the EM-algorithm in this setting can be found in McLachlan and Peel (2000) (pp. 48–51).

Upon convergence, proper standard errors of each component location and scale parameters can be computed numerically using the Hessian matrix of the likelihood evaluated at the maximum.

2.2. Mixtures of exponential family distributions

We can easily extend the previous approach to mixtures of other distributions from the exponential family:

$$f(y_i) = \exp\left(\frac{y_i\theta_i - b(\theta_i)}{\phi_i} + c(y_i, \phi_i)\right). \tag{11}$$

In general, the score equations (4) become:

$$\frac{\partial \log L}{\partial \beta_{kp}} = \sum_{i=1}^{N} c_{ki}\left(\frac{y_i - \mu_{ki}}{\phi_{ki}V(\mu_{ki})}\right)\frac{\partial \mu_{ki}}{\partial \eta_{ki}} x_{ip}, \tag{12}$$

where $p = 1, \ldots, P_\beta$. These are the score equations of a standard GLM with a modified prior weight equal to $\tilde{c}_{ki} = c_{ki}/\phi_{ki}$. Therefore a standard IWLS algorithm can be used for the estimation; see e.g. Lee *et al.* (2006) (pp. 85–87). The estimation of the dispersion parameters $\boldsymbol{\gamma}_k$ requires the use of deviance residuals, corresponding to the distribution used as a component of the mixture (e.g. Poisson, binomial, ...). To find the estimates of dispersion parameters γ_{kp} with $p = 1, \ldots, P_\gamma$ the GLM assuming a gamma distribution is fitted as in Section 2.1.

3. Applications

3.1. Fourth Dutch Growth Study

Here we present analyses of the cross-sectional Fourth Dutch Growth Study. The objective is to model the centile curves of height as a function of age among the 7303 boys in the study.

3.1.1. FMIX approach

We first illustrate the FMIX approach on a subset of boys aged 14 to 16. We used 5 Gaussian mixture components. Every component has its own mean and variance, and the weights are 0.2. The assumed density for each height is then:

$$f(y_i|\mu_1 \ldots \mu_5, \sigma_1 \ldots \sigma_5) = \frac{1}{5}\sum_{i=1}^{5}\frac{1}{\sqrt{2\pi}\sigma_i}\exp\left\{-\frac{1}{2}\left(\frac{y_i - \mu_i}{\sigma_i}\right)^2\right\}.$$

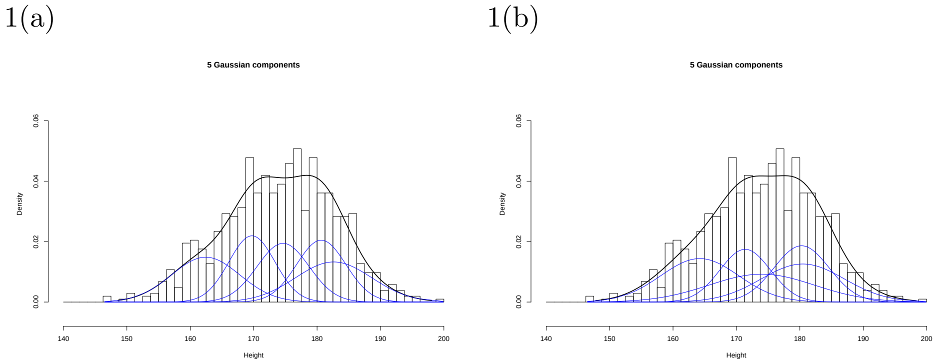

Figure 1. Fourth Dutch Growth Study: estimation of density for boys of age 14 to 16 with 1(a) fixed weights 1(b) estimated weights

The result of the fit is shown in Figure 1(a). For comparison we present the fitted density when weights are estimated. The fitted density is presented in Figure 1(b). Comparison of Figures 1(a) and 1(b) demonstrates the flexibility with which mixtures are fitted.

Next we analysed the complete dataset to see the evolution of height with age. The relationship is clearly non-linear, and therefore we used splines to capture this behaviour.

We started with restricted cubic splines (Harrell, 2001 pp. 20–21) and cubic B-splines (Eilers and Marx, 1996), but the B-splines converge quicker and gave a better fit. In what follows we denote a B-spline of age with n degrees of freedom as $bs(age, n)$, with degrees of freedom (df) equal to the spline order plus the number of interior breakpoints. The following structure was used in the final model:

$$\mu_k = bs(age, 10)\boldsymbol{\beta}_k,$$

$$\sigma_k^2 \equiv \phi_k = \exp(bs(age, 3)\boldsymbol{\gamma}_k),$$

with the optimal df determined by minimizing the AIC (smaller is better). However no claim is made that we determined the optimal df. In Table 1 we present the AIC of various FMIX models. We considered models with a different number of components and a different df in the mean and variance structure. The optimal model contains four mixture components with df=10 in the mean structure and df=3 in the variance structure. The result of the estimation (fitted centile curves) is presented in Figure 2.

Figure 3 shows fitted densities at different pre-selected ages of boys. To check the fit of the model worm plots (Buuren and Fredriks, 2001) are

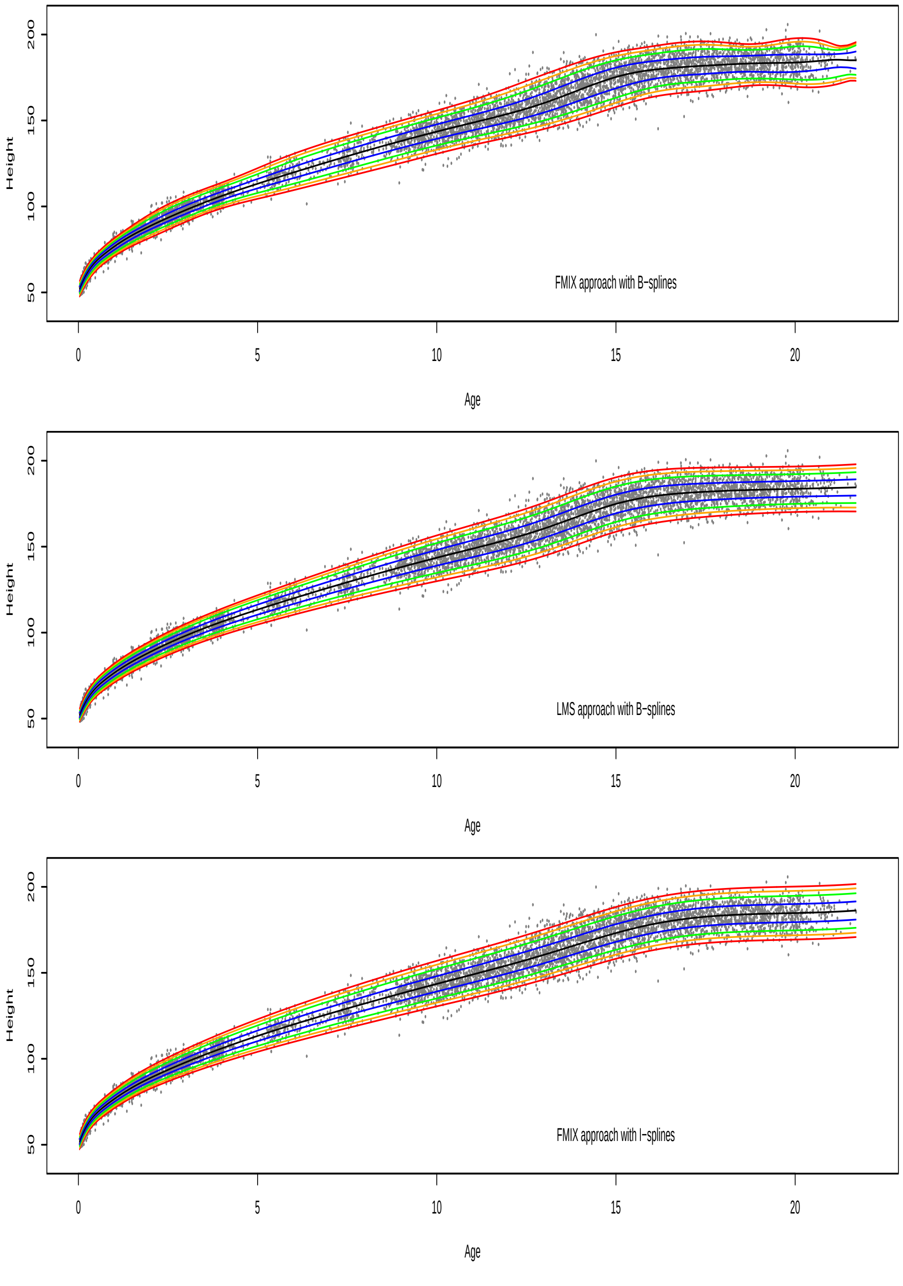

Figure 2. Fourth Dutch Growth Study: fitted centile curves with FMIX and LMS approaches

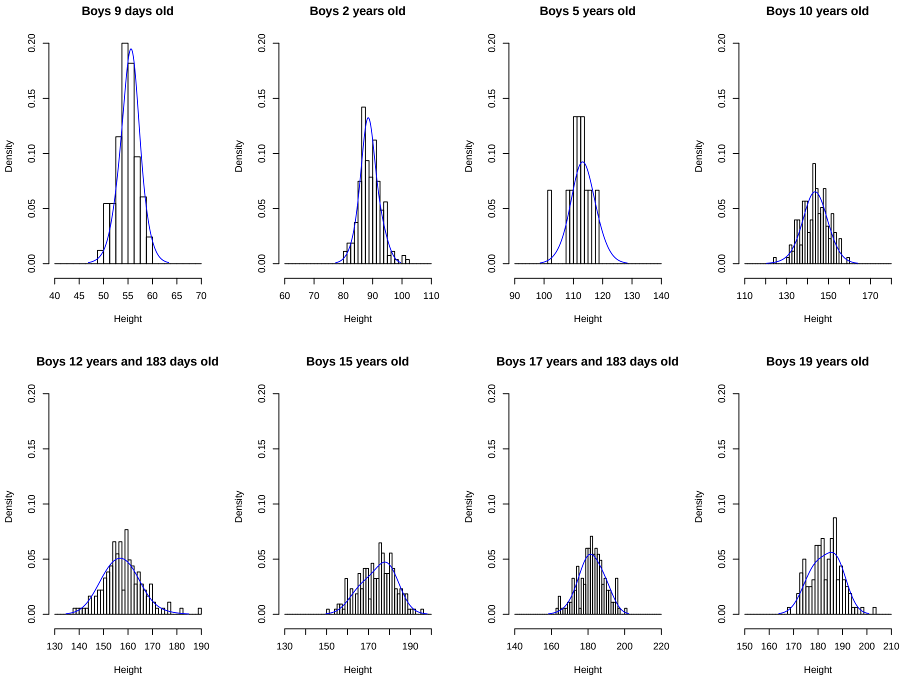

Figure 3. Fourth Dutch Growth Study: fitted densities at pre-selected boys ages

created; see Figure 4. A worm plot presents the expected quantiles and observed quantiles of a model for a range of covariate values, augmented with confidence bands of the estimated quantiles.

The fit of the model seems appropriate, although there is still some misfit for the height of boys younger than 35 days. Modelling height at this young age is quite difficult since in this period of life boys exhibit growth spurts. Further we computed the percentage of observations falling below every centile curve. This computed percentage is very close to the nominal centile i.e. below the 95%-ile lie about 95 percent of observations for the whole span of age. Fitting the Gaussian mixture model with 10 bases in the mean, 3 basis functions in the variance structure and 4 Gaussian components with fixed weights took 2.8 minutes on a Pentium 2.33 GHz core duo 2GB RAM.

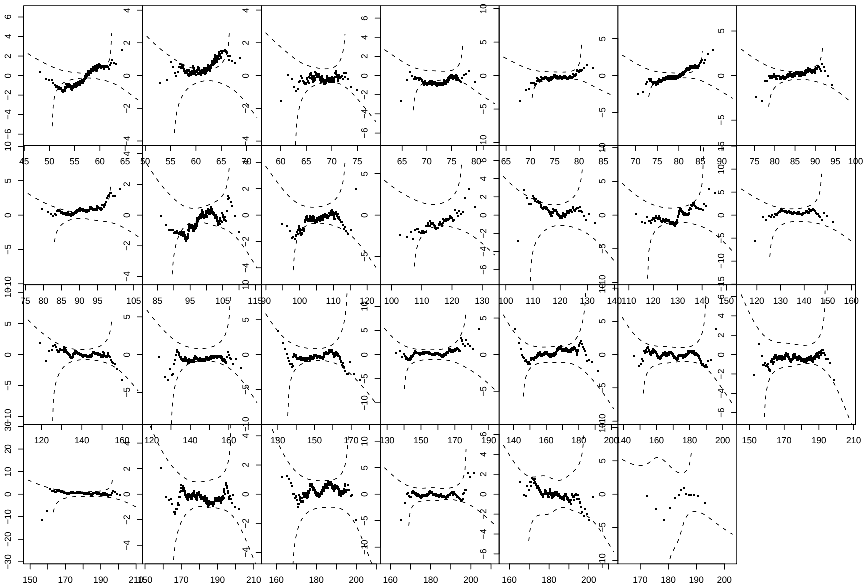

Figure 4. Fourth Dutch Growth Study: worm plots

3.1.2. LMS approach

A competitor to the FMIX approach, popular in growth curve modelling, is the LMS method of Cole and Green (1992). The LMS models were fitted using the gamlss function with distribution 'BCCG' and an adequate number of dfs in each of the structures (mean,variance,skewness). We compared the AIC of LMS models in Table 1. The AIC is worse than that of the FMIX models considered. The fitted centile curves with the LMS approach are shown in Figure 2.

In the analysis of the Fourth Dutch Growth Study height of boys, the visual fit of the LMS model and the mixture approach with 4 mixture components with fixed weights did not differ much; see Figure 2. The LMS method required less than 5 seconds to perform the fit.

3.1.3. General mixture modelling

The R package **gamlss.mx** enables for fitting of mixtures of distributions with estimated weights, which also can depend on covariates. However

Table 1. Fourth Dutch Growth Study: AIC of different models

Model	AIC (lower is better)
M10V3K4E	44564.9
M10V5K7	44601.9
M10V5K5	44602.3
M10V3K5	44582.3
M10V3K4	44574.1
M10V3K3	44579.1
LMS-M10V3S1	44633.7
LMS-M10V5S1	44619.3

Abbreviations used: M10V5K7E - mixture model with 7 mixture components (K), 5 df of a B-spline in variance structure (V) and 10 df B-spline in the mean structure (M); E denotes that the weights of a mixture are estimated; LMS-M10V3S1 - an LMS model with 10 degrees of freedom in the mean structure, 3 degrees of freedom in the variance structure and one parameter in the skewness part of the distribution

when trying to fit the model with splines in the mean and variance structure the program failed to converge (10 and 3 dfs subsequently). We developed our own codes for fitting mixture models with estimated weights to analyze the Fourth Dutch Growth Study. We allowed the weights to be estimated, but not to depend on covariates. The fit of the model measured by AIC was better (see Table 1); however the visual fit of the centile curves to the data was not improved (results not shown). Fitting the model took 8.9.

3.1.4. Penalized Gaussian Mixture approach

In Ghidey *et al.* (2004) the penalized Gaussian mixture model (PGMM) is introduced. This approach proposes to fix the means and variances of the components of the mixture, allowing the weights of the individual distributions to be estimated. Additionally a penalty is imposed on the weights, reducing the difference between the weight estimates of neighbouring distributions. An implementation of this approach can be found in the R package **smoothSurv**. Applications in survival analysis can be found in Komarek *et al.* (2005).

This model can be applied to positive continuous data, with the exponent of the response as dependent variable. The shape of the distribution (modelled by weights) remains the same over the range of covariates, while the mean and scale are estimated from the data and can vary with independent variables. An extension of the approach could allow the weights to depend on the covariates, thereby allowing the shape of distribution to change with a factor.

The PGMM as described by Komarek *et al.* (2005) was applied to the Fourth Dutch Growth Study; it took 8 hrs (on a 2.33GHz 2GB RAM Intel duo core PC) to converge (results not presented). Due to the long estimation time in the case of constant weights, we did not extend the PGMM system to weights depending on covariates. Further, it is unclear how to create the penalty term of the likelihood when weights change with a covariate.

3.2. Simulated example

In this section we present an artificially simulated dataset, however the example is motivated by the situation described by Muthen and Brown (2009). They describe a 4-class drug trial model, where the patients are assigned to either a drug group or a placebo group. Further, in each drug group there are respondents to the treatment and non-respondents. Therefore, while information on the drug assignment is available in covariates, the information whether a patient is a respondent is a latent factor and cannot be observed. Here we will consider one drug only, so that we have respondents to the drug and non-respondents. For each group of respondents and non-respondents we simulate a different trajectory. Further we assume a cross-sectional situation. We assume that our response variable is a hypothetical performance index measuring the treatment performance, while the covariate is the age of the patient. We simulated a dataset for 5000 individuals, with a uniform distribution of age between 0 and 40. The response originates from the following model:

$$\mu_1 = (2(40 - age)^3 + 3000)/(60 - age)^2 - 1.5,$$

for respondents and

$$\mu_2 = (2(40 - age)^3 - 3000)/(60 - age)^2,$$

for non-respondents. The dispersion parameters were set to one in both respondents and non-respondents. To this dataset we fitted LMS, FMIX and **gamlss.mx** models. Figure 5 shows the fitted centile curves obtained from LMS, FMIX with 2 components, and **gamlss.mx** with standard starting values. We plot there 2.5, 5, 10, 25, 35, 40, 45, 50, 55, 60, 65, 75, 90, 95, 97.5 percentile curves. We also computed the proportion of observations falling below the fitted centile curves. For the LMS method 43.3% of observations are below the 40% centile curve, while in the FMIX approach it

Table 2. TSimulated data: AIC of different models

Model	AIC (lower is better)
LMS	20265
FMIX	17972
gamlssMX	17972

is 39.7%. 56% of observations are below the 60% centile curve for the LMS method, and 59.7% in the FMIX model.

The FMIX method improves the fit of the LMS method by detecting the mixture of respondents and non-respondents. The fits of FMIX and **gamlss.mx** are comparable. This is due to the assumption that half the patient population are respondents and half do not respond to the drug. Therefore the correct weights are assumed in the FMIX approach. Table 2 presents the AIC values of the three approaches shown in this section. We conclude that FMIX and gamlssMX gave very similar results.

4. Simulation study

We performed a limited numerical comparison of the performance of the three methods: (1) FMIX, (2) mixture models with estimated weights using our code, and (3) mixture models with estimated weights using the **gamlss.mx** package. Approaches (2) and (3) are theoretically the same; however in practice they might perform differently because of for example, a different convergence criterion or numerical algorithm. We sampled the data from a mixture distribution. We considered three scenarios. In the following text we specify mixture weights as integers, e.g. 75-25 means a mixture of two components with contributions of 75% of the first component and 25% of the second component.

In Scenario 1, we sampled data from a 75-25 mixture of normal distributions, with means -20 and 20, and standard deviations both 15. In Scenario 2 data are obtained by sampling from the 6-44-18-32 mixture of 4 Gaussian distributions with respective means of -20,10,10,20 and standard deviations all equal to 7. Finally in Scenario 3 we used the 10-17.5-22.5-22.5-17.5-10 mixture of 6 Gaussian distributions with means -30,-20,-10,10,20,30 and standard deviations 3,5,7,7,5,3 respectively. In each scenario 6000 observations were sampled. No covariates were used in this simulation.

In each scenario models with a varying number of components were fitted with the methods (1)-(3). In Scenario 1 we used 1-5,10,20 mixture components in fitted models, while in Scenario 2: 1-4,6,10 components

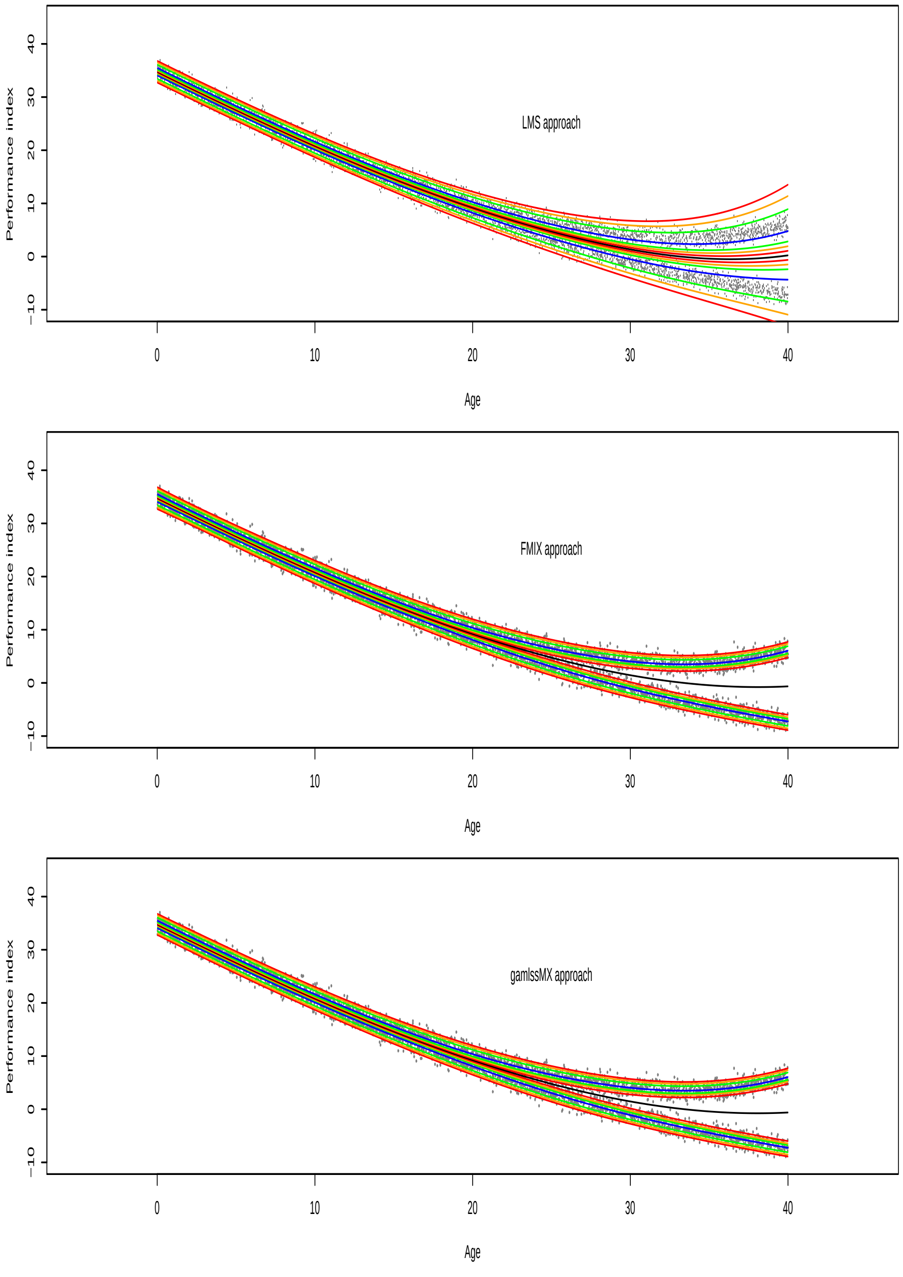

Figure 5. Simulated data fitted centile curves: LMS, FMIX, and gamlssMX approach

were considered. Finally in Scenario 3: 1-6,9,12,18,24 components were considered. We computed the Kullback-Leibler (KL) divergence of the fit of the models against the true distribution. We report ratios of the KL distance of a given model to the best performing model in each scenario. Furthermore we computed the AIC of each model. These two measures were used for comparison of the appropriateness of the model. Note that AIC weights simplicity of the model against the goodness of fit, while KL distance is a measure showing how close to the truth the model is, but without taking into account the complexity of the assumed model.

Under Scenario 1, the lowest KL distance was obtained for the mixture model with 2 components and estimated weights (method 2). This was the optimal model. However, there were still models close to the optimal. The model with 4 components in method 1 attained a KL ratio of 1.11, indicating that the Kullback-Leibler distance of this model was 1.11 times the KL distance of the optimal model. The lowest AIC was obtained by a 2 components mixture model obtained with methods 2 and 3.

Under Scenario 2, the optimal KL distance was obtained for the mixture model of 4 components with fixed weights (method 1). Further, all mixture models with 6 components performed well (methods 1-3), as well as method 2 and 3 with 10 components. These had a ratio of KL distance below 1.16. Note that with method 2 there were computational, difficulties i.e. the boundary of parameter space when 3 or 4 components were used. Note that the two-component model estimated with **gamlss.mx** converged with a maximum log-likelihood value much lower (approximately 500 points) than with other methods.

Under Scenario 3, the lowest KL distance was obtained for the model with 9 components estimated with method 1. Method 2 with 6 components performed equivalently well. Models estimated by method 2 with 9, 12 or 18 components achieved a ratio of their KL distance to the best model of less than 1.11. In this scenario the AICs of **gamlss.mx** models were lower than the AICs of FMIX or estimated weights models with the same number of mixture components.

In summary, by using fixed weights (FMIX) we face fewer less computational problems and we avoid obtaining infinite likelihoods. It is seen that FMIX models often perform as well as general mixture models when the number of mixture components in the FMIX model is slightly larger than the number of components used to generate the data.

5. Conclusions

In this paper we propose the use of fixed weights in finite mixture models. We assume each component of a finite mixture model is parameterized by a separate set of parameters. Therefore, given prior weights (computed in the E-step of the EM algorithm), every mixture component can be separately maximized. Mixture models of this type might be fitted using existing software for generalized linear models or generalized linear mixed models, which allow the inclusion of appropriate weights. The described estimation process is essentially the EM-algorithm of Dempster *et al.* (1977).

The assumption of separate maximization of the components can be relaxed, and the estimation can allow joint parameters over the mixture components. This could be of interest when one wishes to keep the same shape of distribution over the range of the covariates, and vary its mean only.

We used B-splines to model the non-linear distributions; however one could be interested in a monotone centile curves. This can be achieved by using the I-splines of Ramsay (1988) together with some reformulation of the likelihood. Monotonic centile curves fitted to the Fourth Dutch Growth study data are shown in Figure 2. However, the reformulation of the maximization problem required the general Newton-Raphson algorithm to be used instead of the interchangeable (I)WLS computational approach.

In comparison with the estimated weights approach, the speed of computation and stability of the estimation are increased by fixing the weights of the mixture, without much deteriorationing the fit of the model, as can be seen in the Fourth Dutch Growth Study analysis.

References

Cole T.J., Green P.J. (1992): Smoothing reference centile curves: The LMS method and penalized likelihood. Statistics in Medicine 11: 1305–1319.

Dempster A.P., Laird N.M., 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.

Eilers P., Marx B. (1996): Flexible smoothing with b-splines and penalties. Statistical Science 11: 89–121.

Ghidey W., Lesaffre E., Eilers P. (2004): Smooth random effects distribution in a linear mixed model. Biometrics 60: 945–953.

Harrell F.E. (2001): Regression Modelling Strategies. Springer-Verlag, New York.

Komarek A., Lesaffre E., Hilton J.F. (2005): Accelerated failure time model for arbitrarily censored data with smoothed error distribution. Journal of Computational and Graphical Statistics 14: 726–745.

Lee Y., Nelder J.A., Pawitan Y. (2006): Generalized Linear Models with Random Effects. Chapman & Hall / CRC: Boca Raton.

McLachlan G.J., Peel D. (2000): Finite Mixture Models. John Wiley and Sons, New York.

Muthen B., Brown H.C. (2009): Estimating drug effects in the presence of placebo response: Casual inference using growth mixture modelling. Statistics in Medicine 28: 3363–3385.

Nelder J.A., Pregibon D. (1987): An extended quasi-likelihood function. Biometrika 74: 221–232.

Nelder J.A., Wedderburn R.W.M. (1972): Generalized linear models. Journal of Royal Statistical Society A 135: 370–384.

Ramsay J.O. (1988): Monotone regression splines in action. Statistical Science 3: 425–461.

van Buuren S., Fredriks M. (2001): Worm plot: a simple diagnostic device for modelling growth reference curves. Statistics in Medicine 20: 1259–1277.