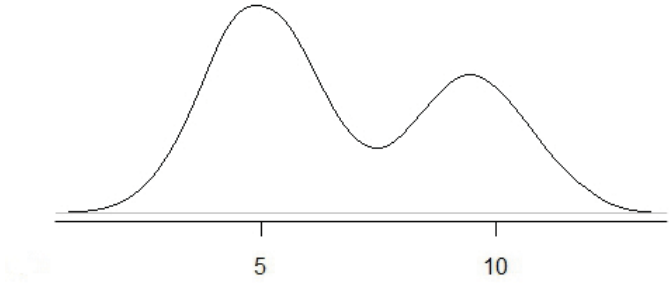
# Parameter Estimation for Mixture Models Given Grouped Data

vorgelegt von
Joanna Wengrzik




**Dissertation**

zur Erlangung des Grades
eines Doktors der Naturwissenschaften
- Dr. rer. nat. -

im April 2012
im Fachbereich 03 (Mathematik und Informatik)

Universität Bremen

Datum des Promotionskolloquiums: 31-05-2012

Gutachter: Prof. Dr. Dr. h.c. Jürgen Timm (Universität Bremen)
Prof. Dr. Werner Brannath (Universität Bremen)

*Für meinen Vater*

# Abstract

Finite mixture models are increasingly used to model heterogeneous data in various important practical situations, where the data can be viewed as arising from two or more subpopulations (components). The decomposition of those components leads to the problem of estimating the parameters of the mixture. The maximum likelihood estimation is a useful tool for obtaining estimates for the parameters. However, since in the case of a finite mixture the corresponding likelihood equations can not be solved analytically, a numerical procedure is necessary. The iterative Expectation Maximization (EM) algorithm provides a convenient way to obtain a solution for a likelihood equation, if a closed-form solution does not exist. Hence, the EM algorithm can be used in the case of mixing distributions to estimate the parameters.

Additionally, this work deals with the problem of observations, which are grouped into intervals. In fact, the basis of this work is observations arising from a mixing distribution, whereby only the number of observations falling into previously specified intervals is known rather than individual observations. To estimate the parameters of a mixture given such grouped observations, several EM algorithm based methods may be applied. For that purpose, two published methods are described in detail. However, both methods tend to be complicated, therefore, four new algorithms, which are much simpler, are proposed. In fact, they are partial modifications of existing algorithms and/or combinations of approaches from other contexts. Furthermore, this work focuses on the comparison of these algorithms. Particularly, a simulation study is presented that compares the different estimation approaches for various two component Gaussian mixtures. Thereby, twelve specified mixture models are chosen, which differ in the choice of the mixing proportions and the component means.

Given specific sample sizes, starting values and interval widths, the proposed algorithms are compared with respect to different mixture models. Thereby, the relative root mean squared error (rRMSE) is used as comparison criterion to measure both the accuracy and the precision for each parameter. Since the likelihood function of finite mixtures may have several maxima, a convergence to a global maximum can not be guaranteed. In fact, convergence can not be guaranteed at all and indeed in several samples and methods divergence occurs. Hence, the convergence failure rate is chosen as second comparison criterion. Moreover, to handle this problem, a variation of the sample size is investigated, since a higher number of samples may improve the performance.

Another way of dealing with the problem of convergence failures is a variation of the starting values. Starting with values that are far away from the global maximum may cause a convergence to a local maximum or even divergence. Accordingly, different initial values may result in different estimations. Therefore, it is indispensable to choose suitable initial values. However, to my knowledge, works on this topic are not published yet. Some techniques for obtaining starting values for the EM algorithm for mixtures with individual observations have been found and it is attempted to modify four of them for the case of grouped observations. Since it turned out that these modifications are not suitable for every mixture model, a new strategy is developed which seems to cover a higher number of mixture models.

In addition, the influence of the interval width is investigated. First, an analysis for selected mixture models with different interval widths is presented. Subsequently, an attempt is made to generalize the findings. More specifically, observations grouped into intervals of, for instance, width $h = 1$ may lead to different estimation results dependent on the underlying mixture model. In mixtures with well separated and widespread components, an interval length of $h = 1$ may be sufficient to achieve good results. In contrast, it might be difficult to obtain results for mixtures with highly overlapped and constricted components when the interval width is 1. For this purpose, a strategy is developed to evaluate a sample whether or not good results can be achieved. Furthermore, recommendations are given with respect to the method that should be taken.

In summary, the simulation study revealed interesting results. First of all, it is discovered that the new simple methods that circumvent the grouping structure achieved much better results than the more complex algorithms. Thereby, it has to be distinguished between mixtures with highly overlapped and well separated components. The former mixtures cause failures in a higher number of samples and the estimation results differ distinctly from the true values. Particularly, algorithms that consist of more than one iteration procedure are more affected. Furthermore, it could be shown that all considered methods are almost comparable in cases where the interval width is small and/or the sample size large. In contrast, situations where the interval width is large and/or the sample size is small can be handled best with the new proposed algorithms.

Finally, to illustrate the recommended proceeding in the estimation of the parameters of a mixture, when the observations are only given in grouped form, this work contains two practical examples. The first example illuminates the Old Faithful Geyser Data to which two well separated Gaussian components could be fitted. The second example contains Sodium measurements of pathological and non-pathological individuals. Here, a decomposition reveals a highly overlapped two component Gaussian mixture.

# Zusammenfassung

Mischverteilungen werden zunehmend eingesetzt um heterogene Daten, die aus zwei oder mehr Subgruppen (Komponenten) bestehen, zu modellieren. Die Zerlegung der einzelnen Komponenten führt zu dem Problem der Parameterschätzung, wobei die Maximum Likelihood Methode ein nützliches Werkzeug darstellt. Da jedoch im Fall einer Mischverteilung die entsprechenden Likelihoodgleichungen nicht analytisch gelöst werden können, ist an dieser Stelle ein numerisches Verfahren zu wählen. Dabei bietet der iterativ arbeitende Expectation-Maximization (EM) Algorithmus eine bequeme Möglichkeit um das Maximum der Likelihoodfunktion zu bestimmen. Folglich können mit dem EM Algorithmus die Parameter einer Mischverteilung geschätzt werden.

Zusätzlich werden dabei Beobachtungen betrachtet, die in Intervalle gruppiert sind. Genauer gesagt, bilden Beobachtungen einer Mischverteilung die Grundlage der Arbeit, wobei anstelle der einzelnen Beobachtungen nur die Anzahl der Beobachtungen bekannt ist, die in zuvor festgelegte Intervalle fallen. Um auch in diesem Fall die Parameter der Mischverteilung zu schätzen, werden in dieser Arbeit verschiedene Methoden vorgestellt, die alle auf der Idee des EM-Algorithmus basieren. Dabei werden zunächst zwei bereits publizierte Methoden präsentiert. Diese sind jedoch aufwändig in der Implementierung und daher werden vier neue, einfachere Algorithmen vorgeschlagen, die teilweise auf Modifizierungen von bereits bestehenden Verfahren, aber auch auf Ansätzen aus anderen Kontexten basieren. Der Schwerpunkt der vorliegenden Arbeit liegt dabei auf dem Vergleich der sechs vorgestellten Algorithmen. Insbesondere wird eine Simulationsstudie präsentiert, die die unterschiedlichen Ansätze für ausgewählte Zwei-Komponenten-Gaußmischverteilungen vergleicht. Im Speziellen werden zwölf verschiedene Mischmodelle untersucht, die in der Wahl ihrer Mischverhältnisse und Komponentenmittelwerte variieren.

Beginnend mit festem Stichprobenumfang, Startwerten und Intervallbreiten, werden die vorgeschlagenen Algorithmen in Hinblick auf die verschiedenen Mischmodelle betrachtet, um einen ersten Eindruck von den einzelnen Verfahren zu erhalten. Dabei wird der relative Root Mean Squared Error (rRMSE) als Vergleichskriterium verwendet um sowohl die Genauigkeit als auch die Präzision der einzelnen Parameter zu messen. Da die Likelihoodfunktion der Mischverteilungen mehrere Maxima aufweisen kann, kann eine Konvergenz zu einem globalen Maximum nicht garantiert werden. Tatsächlich kann sogar Divergenz auftreten, was sich auch in den Untersuchungen bei einigen Methoden und mehreren Stichproben zeigt. Daher wird die Ausfallrate der Methoden als weiteres Vergleichskriterium

herangezogen. Um das Problem der Divergenz einzudämmen, wird in einer weiteren Untersuchung eine Variation des Stichprobenumfangs untersucht, da eine größere Anzahl von Beobachtungen die Ergebnisse einiger Methoden verbessert.

Eine andere Herangehensweise um die Konvergenzprobleme der Algorithmen zu kontrollieren, ist eine genaue Betrachtung der Startwerte. Wählt man Startwerte, die weit von dem globalen Maximum entfernt sind, konvergieren die Algorithmen möglicherweise zu einem lokalen Maximum oder divergieren im Extremfall. Eine Konsequenz daraus ist, dass mit unterschiedlichen Anfangswerten auch unterschiedliche Schätzungen erzielt werden, weshalb es unabdingbar ist geeignete Startwerte zu wählen. Obwohl dieses ein sehr wichtiges Thema ist, konnten in der Literatur keine Untersuchungen zu dem hier untersuchten Szenario gefunden werden. Einige bekannte Techniken zur Bestimmung von Startwerten für den EM-Algorithmus für Mischungen basierend auf Einzelbeobachtungen, wurden daraufhin für den Fall von gruppierten Beobachtungen modifiziert. Da sich herausstellte, dass diese Modifikationen nicht für jedes Mischmodell geeignet sind, wurde eine neue Technik entwickelt, die eine höhere Anzahl von Mischverteilungen abdeckt.

Darüber hinaus werden die Folgen einer Variation der Intervallbreite betrachtet. Zunächst wird eine Analyse für ausgewählte Mischverteilungen mit unterschiedlichen Intervallbreiten vorgestellt. Anschließend wird ein Versuch unternommen die Ergebnisse zu verallgemeinern. Beobachtungen, die beispielsweise in Intervalle mit der Breite $h = 1$ gruppiert sind, können abhängig von dem zugrunde liegenden Mischmodell zu unterschiedlichen Schätzergebnissen führen. Für Mischverteilungen mit deutlich voneinander getrennten Komponenten könnte eine Intervalllänge von 1 ausreichend sein um gute Ergebnisse zu erzielen. Im Gegensatz dazu kann es schwierig sein bei gleicher Intervallbreite Ergebnisse für Mischverteilungen mit stark überlappenden Komponenten zu erhalten. Zu diesem Zweck wurde eine Strategie entwickelt, um zu bewerten, inwiefern bei einer zu untersuchenden Stichprobe gute Ergebnisse erzielt werden können. AnschlieSSend werden Empfehlungen zugunsten einer bestimmten Methode aufgrund der Datenlage ausgesprochen.

Zusammenfassend konnte die präsentierte Simulationsstudie interessante Ergebnisse aufzeigen. Zunächst einmal konnte gezeigt werden, dass die neuen, einfachen Methoden, die die Gruppierung durch eine Datentransformation umgehen, bessere Ergebnisse erreichen als die komplexen Algorithmen. Dabei muss zwischen Mischverteilungen mit stark überlappenden und gut getrennten Komponenten unterschieden werden. Mit den erstgenannten ergeben sich mehr Konvergenzprobleme und die Ergebnisse der Schätzungen unterscheiden sich deutlich von den wahren Werten. Dies trifft insbesondere für Algorithmen zu, die mehr als ein Iterationsverfahren anwenden. Darüber hinaus konnte gezeigt werden, dass in Fällen, wo die Intervallbreite klein ist und / oder der Stichprobenumfang groß, alle Verfahren annähernd gleichwertige Ergebnisse liefern. Für breite Intervalle und /oder kleine Stichproben hingegen bieten die neu vorgeschlagenen Methoden die besten Ergebnisse.

Um die empfohlene Vorgehensweise der Schätzung der Parameter einer Mischverteilung zu illustrieren, bei der die Beobachtungen nur in gruppierter Form vorliegen, werden zwei Beispiele aus der Praxis diskutiert. Das erste Beispiel beleuchtet die Old Faithful Geyser Daten. Hier kann mit Hilfe der vorgestellten Methoden und Startwerten eine Mischverteilung mit zwei gut voneinander getrennten Gaußkomponenten angepasst werden. Das zweite Beispiel beinhaltet Natrium-Messungen von pathologischen und nicht-pathologischen Personen. Eine Zerlegung der beiden Gruppen führt zu der Anpassung einer stark überlappenden Zwei-Komponenten-Gaußmischverteilung.

# Contents

**Abstract** ii

**1 Introduction** 1
- 1.1 Motivation . . . 1
- 1.2 Structure of the Dissertation . . . 3

**2 Mixture Models and Grouped Observations** 5
- 2.1 Finite Mixture Models . . . 5
  - 2.1.1 Identifiability . . . 6
  - 2.1.2 Two Component Gaussian Mixtures . . . 7
  - 2.1.3 Maximum Likelihood Estimation . . . 10
- 2.2 The EM Algorithm . . . 13
  - 2.2.1 General Introduction . . . 13
  - 2.2.2 Formulation of the EM Algorithm . . . 14
  - 2.2.3 Stopping Criterion . . . 15
  - 2.2.4 Convergence Properties . . . 16
- 2.3 The EM Algorithm for Gaussian Mixtures . . . 17
- 2.4 The Problem of Starting Values . . . 20
- 2.5 Grouped Observations . . . 22
  - 2.5.1 Approximated Estimation . . . 23
  - 2.5.2 EM Algorithm for Grouped Data . . . 24
- 2.6 The Newton-Raphson Algorithm . . . 26

**3 Estimation for Gaussian Mixtures given Grouped Data** 27
- 3.1 Likelihood of Mixtures with Grouped Observations . . . 28
- 3.2 Replacing the Grouped Data Structure . . . 29
  - 3.2.1 The RMEM Algorithm . . . 29
  - 3.2.2 The UDEM Algorithm . . . 30
- 3.3 The MGT Algorithm . . . 31
- 3.4 MacDonalds Approach . . . 33
  - 3.4.1 The APEM Algorithm . . . 35
  - 3.4.2 The EMNR Algorithm . . . 36
  - 3.4.3 The MIX Algorithm . . . 37
  - 3.4.4 Similarity of the APEM and the RMEM Algorithm . . . 38
- 3.5 Techniques for Choosing Initial Values . . . 40

**4 Simulation Comparison** **44**
4.1 Simulation Study . . . 44
4.1.1 Comparison Criteria . . . 45
4.1.2 Investigated Mixture Models . . . 46
4.2 Comparison Results for $\pi_1$=0.5 . . . 48
4.2.1 Mixture Model M1 . . . 48
4.2.2 Mixture Model M2 . . . 52
4.2.3 Mixture Model M3 . . . 55
4.2.4 Mixture Model M4 . . . 57
4.3 Comparison Results for $\pi_1$=0.7 . . . 59
4.3.1 Mixture Model M5 . . . 59
4.3.2 Mixture Model M6 . . . 62
4.3.3 Mixture Model M7 . . . 64
4.3.4 Mixture Model M8 . . . 66
4.4 Comparison Results for $\pi_1$=0.9 . . . 68
4.4.1 Mixture Model M9 . . . 68
4.4.2 Mixture Model M10 . . . 71
4.4.3 Mixture Model M11 . . . 73
4.4.4 Mixture Model M12 . . . 75
4.5 Summary of the Results . . . 77
4.6 Variation of the Sample Size . . . 79
4.6.1 Results for Sample Size n=300 . . . 79
4.6.2 Results for Sample Size n=500 . . . 80
4.7 Variation of the Starting Values . . . 82
4.8 Variation of the Interval Width . . . 89
4.8.1 Results for Different Interval Widths . . . 89
4.8.2 Generalization of the Results . . . 93

**5 Practical Application** **96**
5.1 Example 1: Old Faithful Geyser Data . . . 96
5.2 Example 2: Reference Limits for Laboratory Parameters . . . 98

**6 Summary and Discussion** **102**
6.1 Summary . . . 102
6.2 Outlook . . . 104

**Acknowledgment** **107**

**List of Symbols** **109**

**List of Figures** **110**

**List of Tables** **112**

**A Appendix** **113**
A.1 Estimation Results . . . . . 113
A.2 Cumulated rRMSE for Different Interval Widths . . . . . 115

**References** **123**

# 1 Introduction

## 1.1 Motivation

In many research areas it is often not sufficient to fit only one classical distribution to an underlying data set. Particularly, if the collected data can be viewed as arising from two or more subpopulations, it is necessary to fit a composition of distributions. Such compositions are named mixture distributions or mixture models and are specified by the parameters of each component and the mixing proportions in which the components of the mixture occur.

Especially in a medical context it is often the case that data arise from a heterogeneous population, since patients are not alike. For instance, it is known that genetic predispositions influence the speed of metabolism. Some individuals have a high metabolism rate and others may need much more time to metabolize a particular drug or are even unable to do so. Hence, it can be distinguished between slow/poor metabolizer and rapid/extensive metabolizer, see Wehling (2005). Analyzing a population containing individuals with different metabolism rates leads to substantial different results. Accordingly, a mixing distribution seems to be a reasonable fit to such heterogeneous data. Decomposing this mixing distribution into its components is accompanied by a separation of the two underlying groups, namely the slow- and rapid-metabolizer. Thereby, an estimation for the proportions of the mixture may give an indication how many individuals belong to the one and how many to the other group.

Furthermore, the estimated parameters of each component indicate the mean results and deviations for each group. This is an important finding, since individuals with different metabolism rates need different drug timetables. While a rapid-metabolizer should take a particular drug at short time intervals, the time intervals of a slow-metabolizer need to be longer to achieve a similar constant plasma concentration. Of course, this is not the only setting in which mixture models may be used. In fact, mixture models cover a wide range of applications. In Schlattmann (2009), for instance, examples from several disciplines including Pharmacokinetics, Meta-analysis, Disease Mapping, and much more are demonstrated. In brief, mixture models are an important and powerful tool to model heterogeneous data.

Moreover, in practice, values may often not be measured with high precision. For instance, some laboratory parameters can only be quantified as whole number. Reporting those measured values as integer is equivalent to an unintentional grouping of the observa-

tions into intervals of length 1. Additionally, in certain situations the observations might be intended grouped into intervals, if for instance the confidentiality must be preserved and data are first collected with a high precision and afterwards coarsened. Thus, analyzing a data set containing grouped observations is elaborate, since the individual observations are not known and only the number of observations falling into previously specified intervals is reported. To handle this problem investigations have been performed previously and several approaches have been proposed including a strategy to fit a classical distribution to such data, see for instance Heitjan (1989). However, fitting of a mixing distribution to grouped observations has only be rarely investigated. Hence, this work concentrates on this problem.

An intensive literature search reveals only two works that focus on this issue. First, the work of MacDonald and Pitcher (1979) and second that of McLachlan and Jones (1988). Both use an approach based on the Expectation Maximization (EM) idea developed and introduced by Dempster et al. (1977). However, they differ in their methodology and no comparison study seems to have been made. The present work contributes to fill this gap.

Furthermore, it turns out that both algorithms are very complex, i.e. the first requires large storage and a high runtime and the second has an formidable implementation and a complicated construction. Hence, it seems reasonable to develop less complex methods. Therefore, the approach of the first mentioned method has been investigated and based on that, two less complex methods are proposed. Subsequently, a second approach attempting to circumvent the grouping structure by mapping the grouped data in two different ways is chosen. In summary, in this work six methods for the fitting of a mixture model to data that is grouped into intervals are investigated and compared. Thereby, the focus is on mixtures with two Gaussian components.

To cover a wide range of possible situations, several settings are investigated including a variation of the mixture models, the sample size, the interval widths and the starting values. The latter are particularly interesting since all six methods are based on the EM approach. Thus, they involve an iterative procedure and suitable starting values are essential. However, literature does not reveal any recommendations for that kind of data situation. Therefore, for the case of grouped observations, new techniques are proposed to obtain suitable initial values for the parameters of a mixture. For this purpose, techniques used for mixtures with individual observations are modified to be able to apply them in the case of grouped observations. It turns out that none of them is adequate for every mixture model, thus a further technique is developed. This new technique covers a higher number of mixture models.

## 1.2 Structure of the Dissertation

After having described the motivation of this work, chapter 2 begins with a brief introduction into the concept of finite mixture models. Basic definitions and notations are given as well as a historical background. Furthermore, a standard technique for the estimation of the parameters of a mixture model is described, which is the maximum likelihood estimation. Since the corresponding likelihood equations can not be solved analytically, a numerical procedure is necessary. In particular, the EM algorithm is chosen to obtain solutions for the likelihood equations. The introduction and formulation of this algorithm in general is followed by the elaboration in the case of finite mixture models. Related problems, like convergence properties and the choice of suitable starting values are as well discussed briefly. Moreover, the problem of grouped observations is presented. Likewise in the case of mixtures, a solution for the likelihood equation, based on the grouped data, may not be given in a closed form. Hence, the EM algorithm seems to be the first choice to solve the problem. However, in such data situations this algorithm is not able to manipulate the likelihood so that it can be displayed in a convenient way. Thus, a further algorithm is required, which may be the Newton-Raphson (NR) algorithm. This algorithm is introduced briefly at the end of this chapter as a second iteration procedure if one algorithm is not sufficient or as an alternative iterative procedure for obtaining a solution for the likelihood equation.

In chapter 3 the attention is focused on a combination of the two situations presented above, which is the estimation of the parameters of finite mixtures given grouped observations. For this purpose, several methods based on a kind of EM algorithm are considered here. In particular, six algorithms are described, including two simple methods, which circumvent the grouping data structure, a method based on an approach proposed by McLachlan and Jones (1988) and finally three methods based on an approach suggested by MacDonald and Pitcher (1979). The last part of this chapter addresses the problem of suitable starting values in this context. Since no solutions could be found in the literature, four techniques were presented which are modifications of strategies that are proposed originally for the EM algorithm with individual observations. Besides, a new technique is proposed in this work, which incorporates some advantages of the other techniques.

Chapter 4 discusses the comparison of the described methods. Thereby, first, a general description of the simulation requirements and proceedings is given. Then twelve mixture models that are going to be investigated are presented, which differ in the degree of separation as well as balance. Starting with fixed sample size, starting values and interval width the methods are compared against each other with respect to these different models. After that some variations are investigated. First, an increase of the sample size is presented. This is followed by a variation of the starting values. Thereby, on the one hand the proposed techniques for obtaining initial values are compared and on the other hand the results of the methods are investigated for different initial values. Finally, it is examined

how much influence the interval width has. Therefore, a simulation study with different interval widths is presented. This chapter ends with the attempt to generalize the findings for mixture models besides those considered here.

In chapter 5, the practical application of the presented methods and the proposed techniques for obtaining starting values is demonstrated in two examples. In the first example a well separated two-component Gaussian mixture is fitted to the Old Faithful Geyser Data. The second example contains Sodium measurements from pathological and non-pathological persons. Here a highly overlapped two component Gaussian mixture seem to be reasonable.

Finally, the subject of the last chapter is a brief summary of the important findings and conclusions, followed by an outlook. In addition, the appendix contains the numerical results of the presented simulation studies.

# 2 Mixture Models and Grouped Observations

Starting with basic definitions and notations related to finite mixture models and maximum likelihood estimation (section 2.1), this chapter continues with the presentation of the EM algorithm in general as a method for obtaining a solution for a likelihood equation if an analytical solution does not exist (section 2.2). Since this is the case in mixture models, a specification of the EM algorithm for finite mixtures follows in section 2.3. Subsequently, the importance of suitable starting values is investigated in section 2.4. Furthermore, the problem of grouped observations is discussed in section 2.5 and in particular the EM algorithm for grouped observations is highlighted. Finally, section 2.6 presents briefly the NR algorithm as another numerical procedure for obtaining a solution for the likelihood equation.

## 2.1 Finite Mixture Models

Mixture distributions are increasingly used to model heterogenous data in various important practical situations, where the data can be viewed as arising from two or more subpopulations (components). The problem of decomposing a mixture into its components, that is, estimating the parameters of the mixing distribution, has a long history and dates back to Pearson (1894), who dealt with a mixture of two components with equal variances by using the method of moments. Further work using this approach is Charlier (1906), Charlier and Wicksell (1924), Cohen (1967) and Tan and Chang (1972). Attempts to use the maximum likelihood estimation in this context were made by Rao (1948) and by Hasselblad (1966, 1969).

A wide range of practical applications and a detailed statistical analysis of finite mixtures considering different estimation methods were presented by Everitt and Hand (1981) and Titterington et al. (1985). More general descriptions were published by McLachlan and Basford (1988), McLachlan and Jones (1988), McLachlan and Krishnan (1997) and McLachlan and Peel (2000), which are the basis for the descriptions in the present work.

A good historical review can be found in Holgersson and Jorner (1978), Redner and Homer (1984) and Everitt (1996). Some applications in a medical context were presented by Frühwirth-Schnatter (2010) and Schlattmann (2009). The most recent work by Mengerse et al. (2011) shows the relevance of this topic considering the problem in a Bayesian framework.

In brief, a mixture distribution is a compounding of distributions, which arises when a sample from a heterogeneous population with different density functions is taken. According to Titterington et al. (1985), a finite mixture can be defined more formal as follows.

**Definition 2.1** *The distribution of a random variable $X$ with density function of the form*

$$g(x, \Psi) = \sum_{j=1}^{k} \pi_j f(x, \theta_j) \tag{2.1}$$

*is named a* ***(finite) mixture distribution*** *with $k$ components and*

$$\Psi = (\pi_1, \ldots, \pi_k, \theta_1, \ldots, \theta_k). \tag{2.2}$$

*Thereby, $f(x, \theta_j)_{j=1,\ldots,k}$ denote the* ***component densities*** *of the mixture with parameter $\theta_j$ and the* ***mixing proportions*** *or* ***mixing weights*** *$\pi_1, \ldots, \pi_k$ are positive and sum to unity.*

**Remark 2.1** In earlier literature the alternative nomenclature of **compound distributions** can be found, see Everitt and Hand (1981).

It is not absolutely necessary that the component densities $f(x, \theta_j)_{j=1,\ldots,k}$ belong to the same parametric family, but throughout this work, it is assumed that they are Gaussian.

**Definition 2.2** *A random variable $X$ with density function*

$$f(x, \theta) = \phi(x, \mu, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\big( -\frac{1}{2}(x-\mu)^2/\sigma^2 \big), \tag{2.3}$$

*where $\theta = (\mu, \sigma^2)$, is said to be* ***normal*** *or* ***Gaussian*** *distributed with parameters $\mu$ and $\sigma^2$.*

For compositions of other distributions, the reader is referred to the literature mentioned above, or for instance to Simar (1976) and Moharir (1992) for Poisson mixtures, Falls (1970) for Weibull mixtures or Blischke (1962) and Medgyessy (1961) for Binomial mixtures.

Since in many practical situations the number of components $k$ is known, the problem of estimating $k$ is not considered here. In cases where uncertainty about the number of components exists, solutions may be found for example in Aitkin and Rubin (1985) or Lo et al. (2001).

### 2.1.1 Identifiability

For given observations $x_1, \ldots, x_n$ the aim is to fit a mixture distribution to these data by estimating all parameters of this mixture $\Psi$ as defined in (2.2). This is, however, only meaningful if $\Psi$ is identifiable, which refers to the existence of a unique characterization for any mixture model, see McLachlan and Peel (2000).

In general, a parametric family of density functions $f(x, \theta)$ is said to be identifiable if distinct values of $\theta$ determine distinct members of the family. That is

$$f(x, \theta) = f(x, \theta^*) \tag{2.4}$$

if and only if

$$\theta = \theta^*.$$

In the case of finite mixture models, where the density function is $g(x, \Psi)$ from (2.1), the definition of the identifiability is slightly different. To see why this is necessary, suppose that $g(x, \Psi)$ has component densities $f(x, \theta_j)$ and $f(x, \theta_{j^*})$ that belong to the same parametric family. Then (2.4) will still hold when the component labels $j$ and $j^*$ are interchanged in $\Psi$. Although this class of mixtures may be identifiable, $\Psi$ is not. Thus, identifiability can be interpreted in the case of mixtures as follows.

**Definition 2.3** *Let*

$$g(x, \Psi) = \sum_{j=1}^{k} \pi_j f(x, \theta_j) \quad \textit{and} \quad g(x, \Psi^*) = \sum_{j=1}^{k^*} \pi_j^* f(x, \theta_j^*)$$

*be any two members of a parametric family of mixture densities. This class of finite mixtures is said to be* ***identifiable*** *for $\Psi$ if*

$$g(x, \Psi) = g(x, \Psi^*)$$

*if and only if $k = k^*$ and the component labels can be permuted so that*

$$\pi_j = \pi_j^* \quad \textit{and} \quad f(x, \theta_j) = f(x, \theta_j^*).$$

With this definition mixtures of normal distributions are identifiable up to a permutation of the component labels. To handle the interchange of the component labels, restrictions may be imposed on the solution, which may be, for example, $\mu_1 < \mu_2 < \ldots < \mu_k$. Thus, in practice, all parameters can be determined exactly. For a detailed description of the concept of identifiability the reader is referred to Titterington et al. (1985), Teicher (1961, 1963), Frühwirth-Schnatter (2010), and Yakowitz and Spragins (1968).

### 2.1.2 Two Component Gaussian Mixtures

For simplicity, the simulation study contains a consideration of two component Gaussian mixtures. Therefore, in the following, some useful properties and examples are given.

**Notation 2.1** *A distribution having the density function*

$$g(x, \Psi) = \pi_1 \phi(x, \mu_1, \sigma_1^2) + \pi_2 \phi(x, \mu_2, \sigma_2^2) \tag{2.5}$$

*is named a* ***two component Gaussian mixture****, where* $\phi(\cdot)$ *is the density of a Gaussian distribution as given in (2.3) and*

$$\Psi = (\pi_1,\ \pi_2,\ \mu_1,\ \mu_2,\ \sigma_1^2,\ \sigma_2^2). \tag{2.6}$$

To ensure the identifiability of $\Psi$, it is assumed that the component means are in ascending order, i.e.

$$\mu_1 < \mu_2. \tag{2.7}$$

Moreover, for convenience, assumptions are made on the standard deviations. In example, common standard deviations equal 1 are considered.

**Notation 2.2** *The separation of the components of a two component Gaussian mixture with* $\sigma_1 = \sigma_2 = 1$ *can be expressed by the difference between the component means, which is*

$$\Delta = \mu_2 - \mu_1.$$

*Mixtures with* $\Delta \leq 2$ *are referred to as* ***highly overlapped*** *mixture distributions, while mixtures with* $\Delta > 2$ *are denoted by* ***well separated*** *mixtures.*

This classification is approximately equal to the definition of unimodality of a mixture. Eisenberger (1964) gave the sufficient condition that a mixture is unimodal if

$$\Delta^2 < \frac{27\sigma_1^2\sigma_2^2}{4(\sigma_1^2 + \sigma_2^2)}.$$

Accordingly to this condition, a mixture with $\sigma_1 = \sigma_2 = 1$ is unimodal for $\Delta < 1.84$.
Behboodian (1970) considered this problem, too, and derived the following sufficient condition for a mixture of two Gaussian distributions to be unimodal

$$\Delta \leqq 2\min(\sigma_1, \sigma_2).$$

Since $\sigma_1 = \sigma_2 = 1$ is assumed, his classification corresponds to that one chosen in this work, which is $\Delta < 2$. In brief, unimodal mixture distributions may roughly be referred to have highly overlapped components.

Furthermore, a mixture distribution may be characterized by its mean and variance.

**Lemma 2.1** *Suppose* $X$ *is a random variable with a two component Gaussian mixture distribution as defined in (2.5). The mean* $\mu_m$ *and variance* $\sigma_m^2$ *of such a mixture are given by*

$$\mu_m = \pi_1\mu_1 + \pi_2\mu_2, \tag{2.8}$$
$$\sigma_m^2 = \pi_1(\sigma_1^2 + \mu_1^2) + \pi_2(\sigma_2^2 + \mu_2^2) - \mu_m^2. \tag{2.9}$$

**Proof 2.1** Using the operator E as expectation, the mean of the mixture is given by

$$\begin{aligned}\mu_m = E[X] &= \int_{-\infty}^{\infty} x g(x, \Psi) dx \\ &= \int_{-\infty}^{\infty} x\pi_1 f(x, \theta_1) + \int_{-\infty}^{\infty} x\pi_2 f(x, \theta_2) \\ &= \pi_1\mu_1 + \pi_2\mu_2.\end{aligned}$$

The variance can be calculated as follows

$$\begin{aligned}\sigma_m^2 &= E[X^2] - E[X]^2 \\ &= \int_{-\infty}^{\infty} x^2 g(x, \Psi) dx - \mu_m^2 \\ &= \pi_1 \int_{-\infty}^{\infty} x^2 f(x, \theta_1) dx + \pi_2 \int_{-\infty}^{\infty} x^2 f(x, \theta_2) dx - \mu_m^2 \\ &= \pi_1(\sigma_1^2 + \mu_1^2) + \pi_2(\sigma_2^2 + \mu_2^2) - \mu_m^2.\end{aligned}$$

□

Some examples of two component Gaussian mixtures are illustrated in figure 2.1 and 2.2. The former shows mixtures with standard deviations $\sigma_1 = \sigma_2 = 1$, mixing proportions $\pi_1 = \pi_2 = 0.5$ and different component means. Starting with a highly overlapped mixture where $\Delta = 1$, the second component is shifted three times until a well separated mixture with $\Delta = 4$ occurs.

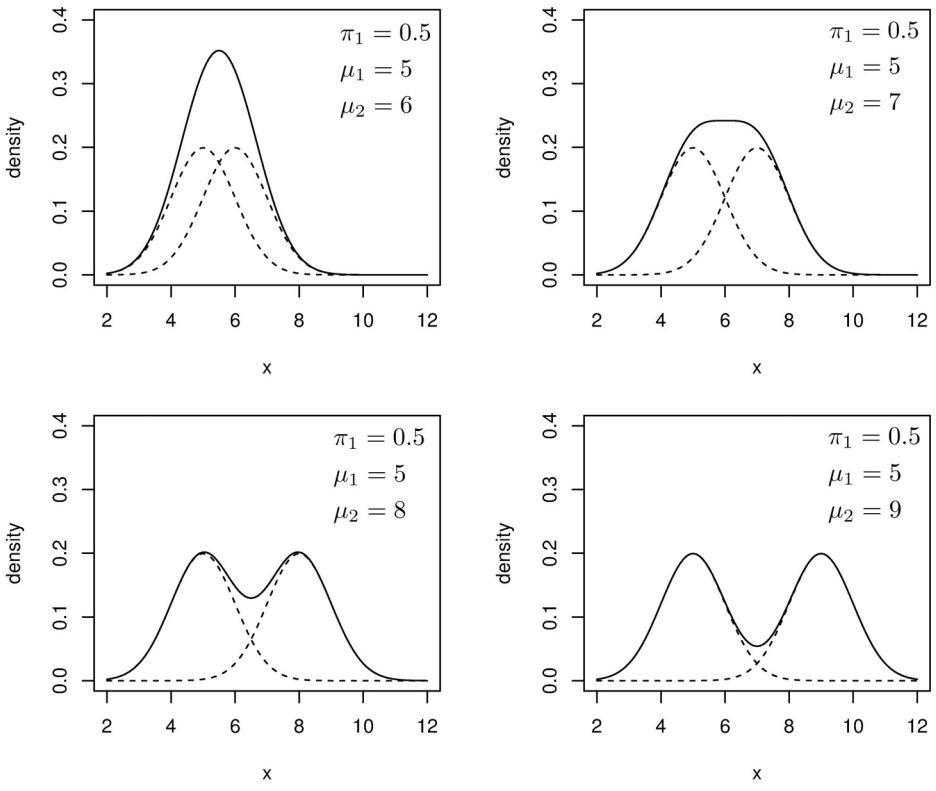



Figure 2.1: Densities of two component Gaussian mixtures with equal variances $\sigma_1^2 = \sigma_2^2 = 1$ and different component means.

Figure 2.2 displays a variation of the mixing proportions. Thereby, the component standard deviations are again chosen to be 1, the component means are $\mu_1 = 5$ and $\mu_2 = 9$, and the mixing proportions are now $\pi_1 = 0.7$ and $\pi_1 = 0.8$, respectively.

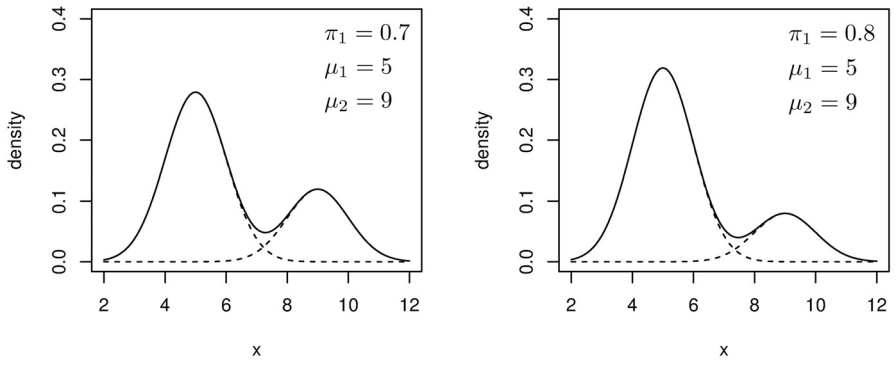



Figure 2.2: Densities of two component Gaussian mixtures with equal variances $\sigma_1^2 = \sigma_2^2 = 1$ and different mixing proportions.

### 2.1.3 Maximum Likelihood Estimation

There is a large amount of estimation methods including the methods of moments, graphical estimation procedures, maximum likelihood estimation, minimum $\chi^2$ and Bayes' estimators. The question arises which estimation method should be taken in estimating parameters of a mixture distribution. A partial answer can be found, for instance, in Tan and Chang (1972). They provided a comparison between the method of moments and the method of maximum likelihood estimation and showed that the maximum likelihood estimation is superior.

Holgersson and Jorner (1978) also compared several estimation techniques and similarly come to the conclusion that the best fit is achieved by the maximum likelihood estimation.

Finally, Day (1969) mentioned the advantages of the maximum likelihood estimation over the minimum $\chi^2$ and the Bayes' estimators. Therefore, the maximum likelihood estimation is used in the present work, which involves the maximization of the likelihood function or equivalently the maximization of the log-likelihood function, which is described, e.g. by Little and Rubin (2002).

**Definition 2.4** *Let $x = (x_1, \ldots, x_n)$ denote independent observations from a random sample of size $n$ of a random variable $X$ with density function $f(x, \theta)$, where $\theta$ is the parameter vector we wish to estimate. Then*

$$L(\theta, x) = \prod_{s=1}^{n} f(x_s, \theta)$$

*denotes the **likelihood function**, abbreviated by likelihood.*

Often it is more convenient to use the logarithm of the likelihood function instead of the likelihood itself, which is the **log-likelihood** given by

$$\log L(\theta, x) = \sum_{s=1}^{n} \log f(x_s, \theta). \tag{2.10}$$

This is meaningful, since log is a monotonic transformation and $L$ will take its maxima at the same parameter values that log $L$ does.

**Definition 2.5** *A **maximum likelihood estimate** $\widehat{\theta}$ of $\theta$ is a value that maximizes the likelihood $L(\theta, x)$, which is*

$$\widehat{\theta} = \arg\max_{\theta} \; L(\theta, x). \tag{2.11}$$

This definition allows the possibility of more than one maximum likelihood estimate. In fact, in several practical applications multiple maxima may occur. However, for many important models "the maximum likelihood estimate is unique and furthermore, the likelihood function itself is differentiable and bounded above", Little and Rubin (2002, p. 81). In such cases, a solution can be found by solving the corresponding score equation.

**Definition 2.6** *The **likelihood equation** or **score equation** is defined by*

$$S(x, \theta) = \frac{\partial \log L(\theta, x)}{\partial \theta} = 0. \tag{2.12}$$

Letting $c$ be the number of components in $\theta$, then the likelihood equation is a set of $c$ simultaneous equations, defined by differentiating $\log L(\theta, x)$ with respect to all $c$ components of $\theta$. In cases where more then one solution exist, the second-order derivatives are examined to verify that it is indeed a maximum which has been achieved and not some other stationary point, see e.g. Everitt and Hand (1981).

**Definition 2.7** *Let*

$$I(\theta, x) = -\frac{\partial^2 \log L(\theta, x)}{\partial \theta \; \partial \theta^T} \tag{2.13}$$

*be the negative of the second-order partial derivatives of the log-likelihood function with respect to $\theta$ and the superscript $T$ denotes the transpose of $\theta$. Then $I(\theta, x)$ is called the **observed information matrix**. The **expected information matrix** is under regularity conditions given by*

$$\boldsymbol{I}(\theta, x) = E\big[S(x, \theta) S^T(x, \theta)\big] = E\big[I(\theta, x)\big]. \tag{2.14}$$

The asymptotic covariance matrix of $\widehat{\theta}$ is equal to the inverse of the expected information matrix $\mathbf{I}(\theta, x)$, which can be approximated by $\mathbf{I}(\widehat{\theta}, x)$. In practice, it is common to estimate

the inverse of the covariance matrix of the maximum solution by the observed information matrix $I(\widehat{\theta}, x)$ rather than the expected information matrix $\mathbf{I}(\theta, x)$, see McLachlan and Peel (2000).

**Likelihood of Mixture Distributions**

In the case of a mixture distribution, things are slightly more complicated. The likelihood may be written in the form

$$L(\Psi, x) = \prod_{s=1}^{n} g(x_s, \Psi) = \prod_{s=1}^{n} \Big( \sum_{j=1}^{k} \pi_j f(x_s, \theta_j) \Big), \tag{2.15}$$

or with logarithm

$$\log L(\Psi, x) = \sum_{s=1}^{n} \log g(x_s, \Psi) = \sum_{s=1}^{n} \log \sum_{j=1}^{k} \pi_j f(x_s, \theta_j). \tag{2.16}$$

The corresponding score equations

$$\frac{\partial}{\partial \Psi} \sum_{s=1}^{n} \log \sum_{j=1}^{k} \pi_j f(x_s, \theta_j) = 0$$

have no analytical solutions, as can be seen in McLachlan and Jones (1988). Therefore, the solving requires a numerical procedure.
Many different iterative methods to solve this problem exist including the NR algorithm, the method of scoring and the EM algorithm. All of them have three essential requirements in common: (1) the choice of reasonable initial values, (2) an iterative algorithm that defines the new estimates and (3) an appropriate stopping criterion. However, the differences between the algorithms are enormous. While the NR algorithm, depending on the starting values, converges to a solution very quickly, the EM algorithm is much slower, but less sensitive to the choice of the initial values.

Everitt (1984) compared six algorithms for estimating the parameters of a mixture of two Gaussian distributions and came to the conclusion that the NR and the EM algorithm lead to the most satisfying results. In the present work, the investigation of the EM algorithm is chosen, because, among other good properties, this algorithm has an easy implementation, a more convenient calculation, since it is not required to calculate the second derivatives, and in particular the most reliable convergence. Moreover, the advantage of the fast convergence of the NR algorithm vanishes if the separation between components is poor, as has been shown by Dick and Bowden (1973).

In the following, the EM algorithm is described first in general and then for finite mixture models. In addition, the NR algorithm is presented briefly at the end of this chapter, since a combination of both methods is investigated in the next chapter.

## 2.2 The EM Algorithm

### 2.2.1 General Introduction

The EM algorithm was described and introduced by Dempster et al. (1977) as an algorithm for obtaining a maximum of the likelihood if an analytical calculation is not feasible. This algorithm can be used in a wide variety of situations where the data can be seen as being *incomplete*. Beside the obvious incomplete-data problems like missing-data or censored observations, this algorithm can also be used in cases without apparent missing values like in the case of finite mixtures. In such situations, the challenge is to formulate the problem as an incomplete-data problem to apply the EM algorithm. The broad scope of the application of this algorithm in so many fields led to its popularity. In 1992, Meng and Pedlow published a bibliography with more than 1000 papers related to the EM algorithm. McLachlan and Krishnan (1997) estimated at least 1700 publications.

The basic idea of the algorithm has been used much earlier than Dempster et al. (1977). The earliest reference to literature on a EM-type algorithm is Newcomb (1886), who considered the estimation of parameters of a two component Gaussian mixture. This work was followed by many others like McKendrick (1926), who presented an application in a medical context, and Healy and Westmacott (1956), who proposed an example of the EM algorithm in a randomized block design. Baum et al. (1970) used this algorithm in connection with Markov models and Orchard and Woodbury (1972) worked on a similar algorithm called missing information principle. A well-arranged summary of the history of the EM algorithm can be found in McLachlan and Krishnan (1997) or Redner and Homer (1984). In addition, McLachlan and Jones (1988) and Little and Rubin (2002) provide a detailed overview.

The EM algorithm has many appealing advantages compared to other iterative algorithms such as the NR algorithm and Fisher's scoring method. These are, for instance, the economy of storage, the easy implementing, and the numerical stability. Furthermore, in most practical situations this algorithm converges, under fairly mild regularity conditions, to a local maximum. A detailed description of the convergence properties is presented at the end of this section.

Nevertheless, the attempt to use this algorithm in many statistical situations reveals its limitations. Hence, numerous modifications and extensions were developed. Especially the fact that in certain situations the algorithm converges very slowly, resulted in the development of modifications and speed-up mechanisms. For instance, Redner and Homer (1984) recommended combining the EM algorithm with an algorithm of the Newton-type, where the good convergence properties of the EM algorithm are used together with the rapid local convergence of Newton's method. Another hybrid algorithm, called EM/GN algorithm, is introduced by Aitkin and Aitkin (1996), where the EM algorithm is combined with the Gauss-Newton (GN) algorithm.

Furthermore, a combination of algorithms was proposed by Du (2002), who suggested to combine the EM algorithm with the NR algorithm. As this method is already implemented in an R package and contains the estimation of mixtures given grouped observations, this algorithm is one of the investigated methods presented in the next chapter. For more details about modified EM algorithms and hybrid algorithms, the reader is referred to McLachlan and Krishnan (1997) and Little and Rubin (2002).

### 2.2.2 Formulation of the EM Algorithm

In the following, the formulation of the EM algorithm is presented as described in McLachlan and Krishnan (1997). Each iteration of the EM algorithm consists of two steps, the *expectation* and *maximization step*, which are called briefly the E- and the M-step. These names were given by Dempster et al. (1977). The basic idea is to associate with a given *incomplete-data* problem a *complete-data* problem for which the maximum likelihood estimation is more tractable. For instance, the solution of the maximum likelihood estimation may result in a closed-form or there might exist standard computer packages with which the likelihood estimation can be calculated more easily.

In brief, the E-step contains the preparation of the complete-data, which includes the calculation of the log-likelihood for the complete-data set. As this log-likelihood is based partly on unobserved data, it is replaced by its conditional expectation given the observed data using the current fit for the unknown parameters. Finally, the M-step maximizes this obtained complete log-likelihood over $\theta$. Starting from suitable initial parameter values, the E- and the M-step are then repeated until convergence.

**Missing-Data**
Let $x = (x_1, \ldots, x_n)$ denote an observed sample of size $n$ taken from a random variable $X$ with probability density function $f(x, \theta)$, where $\theta$ is the parameter vector which is to be estimated. The parameter space is denoted by $\Omega$.

By introducing $y$ as the *unobserved-data* or *missing-data* vector with corresponding random variable $Y$, a complete-data vector can be written as $w = (x, y)$, where $W$ denotes the corresponding random variable. The log-likelihood for the complete-data set, the *complete log-likelihood*, is denoted by $\log L_c(\theta, w)$.

**Expectation-step**
Let $\theta^{(0)}$ be some initial value for $\theta$. Then on the first iteration, the E-step requires the calculation of the conditional expectation of the complete log-likelihood given the observed data $x$ using the initial value $\theta^{(0)}$, which is

$$E_{\theta^{(0)}}\left[\log L_c(\theta, W) \mid X = x\right] =: Q(\theta, \theta^{(0)}). \tag{2.17}$$

Here and elsewhere in this work, the operator $E_\theta$ denotes expectation using the parameter $\theta$.

**Maximization-step**
On the M-step, the $Q$-function (2.17) is maximized with respect to $\theta$ over the parameter space $\Omega$, which results in a new estimation of $\theta$, denoted by $\theta^{(1)}$:

$$\theta^{(1)} = \arg\max_{\theta}\big(Q(\theta, \theta^{(0)})\big).$$

The E- and M-step are then repeated, but this time with $\theta^{(1)}$ instead of $\theta^{(0)}$.

On the $(t+1)$th iteration, the E- and M-step are defined as

$$\textbf{E-step} \quad : \qquad Q(\theta, \theta^{(t)}) = E_{\theta^{(t)}}\big[\log L_c(\theta, W) \mid X = x\big] \tag{2.18}$$

$$\textbf{M-step} \quad : \qquad \theta^{(t+1)} = \arg\max_{\theta}\big(Q(\theta, \theta^{(t)})\big) \tag{2.19}$$

In summary, the EM algorithm may be formulated in the following way:

1. Determine a missing-data and define a complete-data
2. Calculate the conditional expectation of the complete log-likelihood given the observed data using some initial estimation
3. Maximize the corresponding Q-function to obtain a new estimation
4. Replace the initial estimation with the new estimation
5. Repeat step 2 and 3 until a stopping criterion is reached

### 2.2.3 Stopping Criterion

The E- and M-step are alternated repeatedly until a suitable stopping criterion is satisfied. For instance, this process delivers a sequence of values of the observed log-likelihood as defined in (2.10). To stop the iteration, the absolute difference

$$|\log L(\theta^{(t+1)}, x) - \log L(\theta^{(t)}, x)|$$

or the relative difference

$$\frac{|\log L(\theta^{(t+1)}, x) - \log L(\theta^{(t)}, x)|}{|\log L(\theta^{(t)}, x)|}$$

in this sequence can be considered. If the chosen difference is smaller than a small chosen value $\varepsilon$, the algorithm is terminated. If this occurs at the $(t+1)$th iteration, the estimate of $\theta$ is $\widehat{\theta} = \theta^{(t+1)}$.

Besides the change in the likelihood, one may as well consider the change in the parameters after each iteration. In fact, Seidel et al. (2000) have shown, that the estimation results strongly depend on the implementation. "Different starting strategies and stopping rules yield completely different estimators of the parameter" (p. 481). Though, it is not finally discussed in the literature which convergence criterion should be taken. However, the most frequently used stopping criterion is one based on the likelihood, hence it is used in the present study. Moreover, when comparing different algorithms it is more appropriate to apply the relative change of the log-likelihood, since it is dimensionless.

### 2.2.4 Convergence Properties

Having chosen an appropriate stopping criterion, the convergence properties are of interest. These are briefly outlined in this subsection. A detailed description can be found in Wu (1983) or Dempster et al. (1977). The latter have shown the monotonicity of the log-likelihood sequence.

**Lemma 2.2** *The log-likelihood is not decreased after an EM iteration, that is*

$$\log L(\theta^{(t+1)}) \geq \log L(\theta^{(t)}).$$

**Proof 2.2** See Dempster et al. (1977). □

According to this finding, and if the log-likelihood values are bounded above, the log-likelihood sequence nearly always converges to some $L^* = L(\theta^*)$, see Wu (1983).

Furthermore, Wu (1983) formulated regularity conditions, among others the compactness of the parameter space, under which any likelihood sequence is bounded. Furthermore, under mild conditions, namely if $Q(\theta, \theta^{(t)})$ is continuous in both $\theta$ and $\theta^{(t)}$, $L^*$ is a stationary point.

As the likelihood may have several stationary points, the convergence to a maximum depends on the initial values. However, a convergence to a local or even to a global maximum can not be guaranteed. Only in the case that the likelihood is unimodal (and a differentiation condition is satisfied), any EM sequence converges to the unique maximum likelihood estimate, irrespective of the starting point, see Wu (1983).

In addition, Wu (1983) proved that if the sequence of the likelihood is not trapped at some saddle point, the stationary point is a local maximum. However, in practice, this condition is limited and typically hard to verify. Hence, he recommended to try several EM iterations with different starting points, since a small perturbation away from the saddle point will cause the EM algorithm to diverge away from the saddle point, see McLachlan and Basford (1988).

Nevertheless, examples can be found, where the EM algorithm converges to a saddle point

and not to a local maximum, as presented by McLachlan and Krishnan (1997), where an example adapted from Murray (1977) is discussed. Consequently, it can not be guaranteed that the EM algorithm converges to the global maximum, hence the recommendation of Wu (1983) to try different starting values seems to be meaningful.

## 2.3 The EM Algorithm for Gaussian Mixtures

As mentioned above, the likelihood equation of a mixture distribution, given in (2.16), has no explicit solution. However, to obtain a solution an iterative procedure may be applied. With an appropriate choice of the unobserved-data, the EM algorithm can be used as described by McLachlan and Peel (2000).

**Theorem 2.1** *Suppose $X$ is a random variable with density function $g(x, \Psi)$ as defined in (2.1) with Gaussian components and $\Psi = (\pi_1, \ldots, \pi_k, \theta_1, \ldots, \theta_k)$ contains the unknown parameters. A sample $x = (x_1, \ldots, x_n)$ is made on $X$. To estimate the parameters of the mixture an EM algorithm with the following E- and M-step may be applied*

$$\textbf{E-step:} \quad Q(\Psi, \Psi^{(t)}) = \sum_{j=1}^{k}\sum_{s=1}^{n} e_{js}^{(t)} \log \pi_j + \sum_{j=1}^{k}\sum_{s=1}^{n} e_{js}^{(t)} \log f(x_s, \theta_j) \tag{2.20}$$

$$\textbf{M-step:} \quad \pi_j^{(t+1)} = \frac{1}{n}\sum_{s=1}^{n} e_{js}^{(t)} \tag{2.21}$$

$$\mu_j^{(t+1)} = \sum_{s=1}^{n} e_{js}^{(t)} x_s \;/\; \sum_{s=1}^{n} e_{js}^{(t)} \tag{2.22}$$

$$\sigma_j^{2(t+1)} = \sum_{s=1}^{n} e_{js}^{(t)} (x_s - \mu_j^{(t+1)})^2 \;/\; \sum_{s=1}^{n} e_{js}^{(t)}, \tag{2.23}$$

*where*

$$e_{js}^{(t)} = \frac{\pi_j^{(t)} f(x_s, \theta_j^{(t)})}{g(x_s, \Psi^{(t)})}. \tag{2.24}$$

**Proof 2.3** First, an unobservable or missing-data vector

$$z = (z_1, \ldots, z_n)$$

is introduced, where each $(z_s)_{s=1,\ldots,n}$ is a $k$-dimensional vector of zero-one indicator variables that are one or zero according to whether the $s$th observation did or did not arise from the $j$th component of the mixture, i.e.

$$z_{js} = \begin{cases} 1, & \text{if } x_s \text{ belongs to the } j\text{th component,} \\ 0, & \text{if } x_s \text{ does not belong to the } j\text{th component,} \end{cases}$$

for $s = 1, \ldots, n$ and $j = 1, \ldots, k$. The corresponding random variable is denoted by $Z$ and its density function by $\widetilde{g}(z, \Psi)$.

By defining the complete-data vector as $w = (x, z)$ and denoting $h(w) = h(x, z, \Psi)$ as joint density of the random variable $W$, the complete log-likelihood may be written as

$$\begin{aligned}
\log\ L_c(\Psi, w) &= \log \prod_{s=1}^{n} h(w_s, \Psi) = \sum_{s=1}^{n} \log h(x_s, z_s, \Psi) \\
&= \sum_{s=1}^{n} \log \left(g(x_s, \Psi | z_s) \cdot \widetilde{g}(z_s, \Psi)\right) = \sum_{s=1}^{n} \log \Big( \prod_{j=1}^{k} f(x_s, \theta_j)^{z_{js}} \ \prod_{j=1}^{k} \pi_j^{z_{js}} \Big) \\
&= \sum_{j=1}^{k} \sum_{s=1}^{n} z_{js} \log \left(\pi_j f(x_s, \theta_j)\right). \qquad (2.25)
\end{aligned}$$

**E-step**

The E-step of the EM algorithm requires the calculation of the conditional expectation of the complete log-likelihood given the observed data $x$ using $\Psi^{(t)}$, which is

$$\begin{aligned}
Q(\Psi, \Psi^{(t)}) &= E_{\Psi^{(t)}}\big[\log L_c(\Psi, W) \mid X = x\big] \\
&= E_{\Psi^{(t)}}\Big[\sum_{j=1}^{k} \sum_{s=1}^{n} Z_{js} \log \left(\pi_j f(x_s, \theta_j)\right) \mid X = x\Big] \\
&= \sum_{j=1}^{k} \sum_{s=1}^{n} \Big(\log \left(\pi_j\ f(x_s, \theta_j)\right)\Big)\ E_{\Psi^{(t)}}[Z_{js} \mid X_s = x_s]. \qquad (2.26)
\end{aligned}$$

With the application of Bayes' theorem the conditional expectation of $Z_{js}$ is at the $t$th iteration

$$\begin{aligned}
E_{\Psi^{(t)}}[Z_{js} \mid X_s = x_s] &= P(Z_{js} = 1 \mid X_s =\ x_s) \\
&= \frac{P(X_s = x_s \mid Z_{js} = 1)\ P(Z_{js} = 1)}{\sum_{r=1}^{k} P(X_s = x_s \mid Z_{rs} = 1)\ P(Z_{rs} = 1)} \\
&= \frac{\pi_j^{(t)} f(x_s, \theta_j^{(t)})}{g(x_s, \Psi^{(t)})} = e_{js}^{(t)}, \qquad (2.27)
\end{aligned}$$

for $s = 1, \ldots, n$ and $j = 1, \ldots, k$. Hence, $e_{js}^{(t)}$ represents the probability under $\Psi^{(t)}$ that the $s$th observation belongs to the $j$th component of the mixture given that $X_s$ equals $x_s$. Hence, (2.26) can be written as

$$\begin{aligned}
Q(\Psi, \Psi^{(t)}) &= \sum_{j=1}^{k} \sum_{s=1}^{n} \log \left(\pi_j f(x_s, \theta_j)\right) e_{js}^{(t)} \\
&= \sum_{j=1}^{k} \sum_{s=1}^{n} e_{js}^{(t)} \log \pi_j + \sum_{j=1}^{k} \sum_{s=1}^{n} e_{js}^{(t)} \log f(x_s, \theta_j). \qquad (2.28)
\end{aligned}$$

**M-step**

The M-step now requires the maximization of the Q-function with respect to $\Psi$. Since $\pi_j$ appears only in the first term and $\theta_j$ only in the second term of the righthand side of (2.28), the maximization can be done separately. Starting with the maximization of the first term,

it is necessary to solve

$$\frac{\partial}{\partial \pi_j}\Big( \sum_{j=1}^{k}\sum_{s=1}^{n} e_{js}^{(t)} \log \pi_j + \lambda[\sum_{j=1}^{k} \pi_j - 1]\Big) = 0, \tag{2.29}$$

$$\frac{\partial}{\partial \lambda}\Big( \sum_{j=1}^{k}\sum_{s=1}^{n} e_{js}^{(t)} \log \pi_j + \lambda[\sum_{j=1}^{k} \pi_j - 1]\Big) = 0, \tag{2.30}$$

where $\lambda$ denotes a Lagrange multiplier, since the constraint $\sum_{j=1}^{k} \pi_j = 1$ needs to hold. This yields

$$\lambda = -n \tag{2.31}$$

and an iterative estimator for $\pi_j$, which is

$$\pi_j^{(t+1)} = \frac{1}{n}\sum_{s=1}^{n} e_{js}^{(t)}. \tag{2.32}$$

The maximization of (2.28) with respect to $\theta_j$ depends on the density function $f(x, \theta_j)$. As here Gaussian mixtures are considered, the $f(x, \theta_j)$ are given as in (2.3) where $\theta_j = (\mu_j, \sigma_j^2)_{j=1,\ldots,k}$. Starting with $\mu_j$, the maximization includes calculating

$$\frac{\partial}{\partial \mu_j}\sum_{j=1}^{k}\sum_{s=1}^{n} e_{js}^{(t)}\Big( -\frac{1}{2}\log(2\pi) - \log(\sigma_j) - (x_s - \mu_j)^2/2\sigma_j^2\Big), \tag{2.33}$$

and setting this derivative to zero, which is

$$\sum_{s=1}^{n} e_{js}^{(t)} \sigma_j^{-2}(x_s - \mu_j) = 0. \tag{2.34}$$

Solving (2.34) provides an iterative estimator for $\mu_j$, i.e.

$$\mu_j^{(t+1)} \;=\; \sum_{s=1}^{n} e_{js}^{(t)} x_s \;/\; \sum_{s=1}^{n} e_{js}^{(t)}.$$

An iterative estimator for $\sigma_j^2$ can be obtained by differentiating the second term of (2.28) with respect to $\sigma_j$. Solving the corresponding score equation

$$-\sum_{s=1}^{n} e_{js}^{(t)} \sigma_j^{-1} + \sum_{s=1}^{n} e_{js}^{(t)} (x_s - \mu_j)^2 \sigma_j^{-3} = 0$$

yields an iterative estimator for $\sigma_j^2$, namely

$$\sigma_j^{2(t+1)} \;=\; \sum_{s=1}^{n} e_{js}^{(t)} (x_s - \mu_j^{(t+1)})^2 \;/\; \sum_{s=1}^{n} e_{js}^{(t)}.$$

□

As mentioned in the previous section, the log-likelihood monotonically converges if the sequence is bounded above. Unfortunately, examples of finite mixture models can be found, where the likelihood is unbounded, so that the maximum likelihood estimate does not exist. Redner and Homer (1984) mentioned that "if one of the mixture means coincides with a sample observation and if the corresponding variance tends to zero [...] then the log-likelihood function increases without bound" (p.208). In particular, Kiefer and Wolfowitz (1956) offered an example with a mixture of two normal densities with unknown different means and unknown different variances, where no maximum likelihood estimate exists.

Nevertheless, as suggested in McLachlan and Krishnan (1997), this problem can be circumvented by adding constraints on the parameters, e.g. $\sigma_j^2 \geq \varepsilon$, $j = 1, \ldots, k$ which ensure the compactness of the parameter space and hence the boundedness of the likelihood function. Other constraints are proposed by Day (1969), namely the equality of the variances, or by Hathaway (1983, 1985), who chose for $k = 2$ the constraint $\min(\sigma_1/\sigma_2) \geq \varepsilon > 0$.

In most practical situations these constraints are naturally satisfied, hence useful maximum likelihood solutions can be obtained, see Everitt (1984). However, since several maxima of the likelihood may occur, a global solution depends on the starting values, which is briefly discussed in the next section.

## 2.4 The Problem of Starting Values

As Böhning (1999) stated, the problem of finding the right starting values is scarcely discussed in literature or not mentioned at all. Furthermore, Böhning et al. (1992) remarked that "there is no straightforward method to find starting values for the algorithm." (p. 296). Some authors, for instance Hosmer (1973b), claimed that the initial values do not have much effect on the estimates. In contrast, others emphasize the need for a careful consideration of the starting values, since the likelihood equation often has multiple roots in finite mixtures, which may lead to different solutions.

Everitt (1984) compared several algorithms for obtaining a maximum likelihood estimate in three types of mixtures of two normal distributions. Thereby, he used in each case three different starting values to investigate the estimation results. In most data sets each algorithm converges to the same local maxima of the likelihood function from a variety of different starting values. However, he noted as well that in some situations different results or actually convergence to a singularity may occur. Hence, he suggested to make a choice between solutions on the basis of the value of $L$. In practice, one may choose different starting points, investigate the behavior of the likelihood and decide to accept a maximum or to move on.

The problem of how to find reliable initial values was considered by McLachlan (1988)

for the case of multivariate data. He proposed to use a two dimensional scatter plot to explore the data initially for the presence of clusters. "Any visual clustering of the data so obtained can be reflected in the initial values [..]" (p. 418). A similar procedure can by applied in the univariate case. With a kernel density estimation, one can obtain a first impression of the data. Based on such graphical considerations an initial guess may be verbalized, see example 2.1 below.

Strategies for the univariate case can be found in Karlis and Xekalaki (2003), where as well a good review of previous works on this research field is given. The authors compared several methods for choosing initial values for the EM algorithm. Among others, they analyzed an approach of Finch et al. (1989), who proposed that for a two component Gaussian mixture only the mixture proportions need to be given as initial values. Their idea was that with given mixing proportions, the sample can be separated into two parts, one containing the smallest $n \cdot \pi_1$ observations assumed to belong to the first component of the mixture and one containing the remaining observations assumed to belong to the second component. The mean of the first part of the sample is then used as initial value for the first component mean while the mean obtained by the second part officiates as starting value for the second component mean. Indeed, this method provides good results. In contrast, using random initialization, where the starting points are generated randomly under adequate conditions, the EM algorithm failed in a much higher number of samples. However, the performance of the EM algorithm depends as well on the underlying mixture model. Mixtures with well separated components have been decomposed correctly in their study with almost all investigated initial values strategies. Similar results were reproduced in the present work and will be shown later.

To illustrate the importance of the choice of the right initial values in finite mixture models and their influence on the convergence of an algorithm and so on the estimation results, the following example is given.

**Example 2.1** A mixture with parameter $\Psi = (0.9,\ 0.1,\ 5,\ \mu_2,\ 1,\ 1)$ is chosen. The true value of $\mu_2$ is 9, however, this value is desired to be estimated. A density kernel estimation of a sample of size $n = 100$ is shown in figure 2.3. The corresponding likelihood function for $\mu_2$ is displayed in figure 2.4. As can be seen, this likelihood has two maxima.

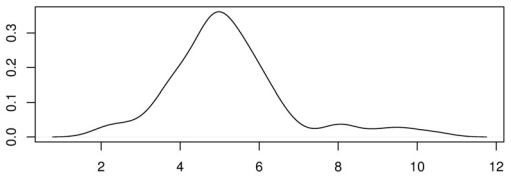



Figure 2.3: Sample of a mixture.

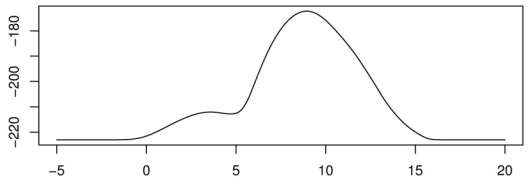



Figure 2.4: Likelihood function for $\mu_2$.

Performing the EM algorithm with the starting point $\mu_2^{(0)} = 10$, which is close to the true value, one obtains the solution $\widehat{\mu}_2 = 9.46$. If the EM algorithm is started with the initial value $\mu_2^{(0)} = 3$, the EM algorithm converges to the local maximum, which results in $\widehat{\mu}_2 = 3.16$. Hence, different starting points cause different estimation results. However, comparing the values of the likelihood, it is obvious which result should be associated with the global maximum.
A similar result can be obtained considering other iterative procedures like the NR algorithm, which is explained in section 2.6. Here, the initial value $\mu_2^{(0)} = 10$ results in $\widehat{\mu}_2 = 8.94$, while a starting value $\mu_2^{(0)} = 3$ yields $\widehat{\mu}_2 = 3.56$.

□

Hence, the choice of the initial values is important and affects the estimation results. In the present work, different algorithms are compared for the estimation of the parameters of mixture distributions when the observations are grouped. Since in that case multiple maxima may occur in the likelihood, the present simulation study also deals with different starting values. In fact, the influence of a variation of the starting values on the results is investigated. However, since the strategies proposed and compared by Karlis and Xekalaki (2003) are designed for obtaining starting values for mixtures with individual data and since no other elaborations on this topic have been found in literature, new techniques for the considered case are proposed in section 3.5.

## 2.5 Grouped Observations

Having introduced mixing distributions and estimation techniques in the previous sections, this section turns to the problem of grouped observations. A well-known example is the histogram, where $n$ observations $x_1, \ldots, x_n$, taken from a random variable $X$, are grouped into $m$ intervals, like demonstrated in figure 2.5. A good review of this topic is given by Heitjan (1989), who studied the grouping and its ramifications on data analysis.

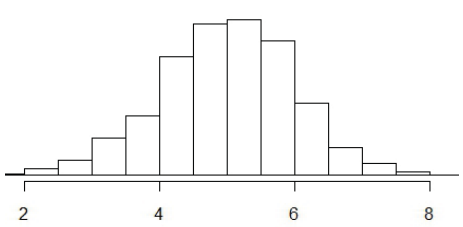



Figure 2.5: Grouped observations.

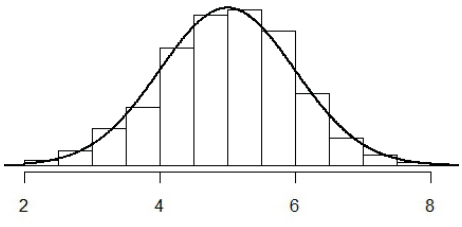



Figure 2.6: A fitted Gaussian distribution to grouped data.

In many practical situations, it is desired to fit a distribution to such data. Unfortunately, a straightforward fitting is not possible and again a numerical procedure is essential. Figure 2.6 illustrates a fitting of a Gaussian distribution to the previous example, which is achieved by the application of the EM algorithm described in section 2.5.2.

**Definition 2.8** *Let $X$ be a random variable with density function $f(x, \theta)$. Suppose that the sample space is partitioned into $m$ intervals. The boundaries of the intervals are denoted by $a_0, a_1, \ldots, a_m$. Observations $x_1, \ldots, x_n$ are made on $X$, but only the number of observations $n_i$ falling into the interval $[a_{i-1}, a_i)_{i=1,\ldots,m}$, is reported and $n = \sum_{i=1}^{m} n_i$. Data of the form*

$$y_i = (a_i, n_i), \quad i = 1, \ldots, m \tag{2.35}$$

*is referred to as* ***grouped data****.*

**Notation 2.3** *The midpoint of the* i*th interval is denoted by $\bar{a}_i = \frac{1}{2}(a_i + a_{i-1})$ and the length by $h_i = a_i - a_{i-1}$. Throughout this work it is assumed that the intervals are of equal length, which is $h = h_i$ for $i = 1, \ldots, m$.*

**Remark 2.2** The observed data vector $(n_1, \ldots, n_m)$ has a multinomial distribution consisting of $n$ draws on $m$ intervals with probabilities $P_1(\theta), \ldots, P_m(\theta)$, where

$$P_i(\theta) := \int_{a_{i-1}}^{a_i} f(x, \theta) dx, \quad i = 1, \ldots, m \tag{2.36}$$

gives the probability that an individual observation belongs to the $i$th interval.

Assuming the individual observations are normally distributed and disregarding the terms that are not functions of $\mu$ and $\sigma^2$, the likelihood and the log-likelihood are

$$L(\theta, y) \propto \prod_{i=1}^{m} P_i(\theta)^{n_i} \quad \text{and} \quad \log L(\theta, y) \propto \sum_{i=1}^{m} n_i \log P_i(\theta), \quad \text{respectively.} \tag{2.37}$$

As the score equation of the log-likelihood in (2.37) can not be solved analytically, again a numerical procedure is reasonable. Several iteration techniques including the scoring-method, the EM algorithm and the NR algorithm are elaborated and compared in detail in this context by Schader and Schmidt (1984, 1988). They pointed out that the EM algorithm is slower in comparison to the other two algorithms, though it has the most numerical stability and is less sensitive to the choice of starting values. How to apply the EM algorithm in the case of grouped observations will be demonstrated in section 2.5.2. Before to that, a simple approximation method for obtaining estimations for the parameters of a Gaussian distribution is presented.

### 2.5.1 Approximated Estimation

The basic idea of this approach, described by Gjeddebaek (1959), Hasselblad (1966), Titterington et al. (1985), and Hald (2001), is to approximate the probability (2.36) that an individual observation falls into the $j$th interval by

$$P_i(\theta) \approx h \cdot f(\bar{a}_i, \theta), \tag{2.38}$$

where $h$ is the interval width and $f(\bar{a}_i, \theta)$ the density function evaluated at the midpoint of the $i$th interval. Supposing $n_1, \ldots, n_m$ give the number of observations in each interval and $a_0, a_1, \ldots, a_m$ are the interval borders, then with (2.38) the corresponding approximated log-likelihood is given by

$$\log L(\theta, y) \approx \sum_{i=1}^{m} n_i \ \log\big(h \cdot f(\bar{a}_i, \theta)\big). \tag{2.39}$$

Assuming $f$ to be the Gaussian distribution, maximization of (2.39) with respect to $\theta = (\mu, \sigma^2)$ is now solvable analytically and yields the following maximum likelihood estimates

$$\begin{aligned} \widehat{\mu} &= \frac{1}{n} \sum_{i=1}^{m} n_i \bar{a}_i \\ \widehat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^{m} n_i (\bar{a}_i - \widehat{\mu})^2. \end{aligned}$$

Gjeddebaek (1959) however, considered $\bar{a}_i$ as a *reference point*, which is referred to some point lying between $a_i$ and $a_{i-1}$ and which does not need to be the midpoint.

This described approximation method provides an efficient maximum likelihood estimate for an estimation of $\mu$ and $\sigma^2$ for intervals less then twice the standard deviation and for sample sizes less then 100, see Gjeddebaek (1959). Since the application of this simple method is limited to small intervals, the maximum likelihood estimation obtained by the EM algorithm is presented in the next section.

### 2.5.2 EM Algorithm for Grouped Data

**Theorem 2.2** *Suppose $X$ is a random variable with a Gaussian density function $f(x, \theta)$. Observations $x_1, \ldots, x_n$ are made on $X$. Instead of the individual observations, only the number of observations $n_i$ of each interval $[a_{i-1}, a_i)$ is reported, so that the observed data can be written as $y = (y_1, \ldots, y_m)$, where $(y_i)_{i=1,\ldots,m}$ are defined in (2.35) and $Y$ denotes the corresponding random variable.*
*To estimate the parameters of the Gaussian distribution an EM algorithm may be applied with the following E- step*

$$Q(\theta, \theta^{(t)}) = \sum_{i=1}^{m} n_i \frac{1}{P_i(\theta^{(t)})} \int_{a_{i-1}}^{a_i} \log f(x, \theta) \ f(x, \theta^{(t)}) dx, \tag{2.40}$$

*where*

$$P_i(\theta) := \int_{a_{i-1}}^{a_i} f(x, \theta) dx, \quad i = 1, \ldots, m. \tag{2.41}$$

*The M-step can not be expressed in a closed-form.*

**Proof 2.4** According to Dempster et al. (1977) and McLachlan and Krishnan (1997), the missing-data vector can be defined for this purpose by

$$x^* = (x_1^*, \ldots, x_m^*),$$

where $x_i^* = (x_{i1}^*, \ldots, x_{in_i}^*)_{i=1,\ldots,m}$, denotes $n_i$ observations from the density

$$f(x,\theta)\mathbf{I}_{x\in[a_{i-1},a_i)}/P_i(\theta), \tag{2.42}$$

with $P_i(\theta)$ as defined in (2.41). The likelihood of the complete-data $w = (x^*, y)$ can then be written as

$$L_c(\theta, w) = \prod_{i=1}^{m} P_i(\theta)^{n_i} \prod_{l=1}^{n_i} \frac{f(x_{il}^*, \theta)}{P_i(\theta)}$$

and the log-likelihood is given by

$$\log L_c(\theta, w) = \sum_{i=1}^{m}\sum_{l=1}^{n_i} \log f(x_{il}, \theta).$$

**E-step**

The E-step of the EM algorithm contains the calculation of the conditional expectation of the complete log-likelihood given the observed data $y$, which is

$$\begin{aligned} Q(\theta, \theta^{(t)}) &= E_{\theta^{(t)}}(\log L_c(\theta, W) \mid Y = y) \\ &= \sum_{i=1}^{m} E_{\theta^{(t)}}\Big(\sum_{l=1}^{n_i} \log f(X_{il}^*, \theta) \mid Y = y)\Big) \\ &= \sum_{i=1}^{m} n_i \int_{a_{i-1}}^{a_i} \Big(\log f(x,\theta)\frac{f(x,\theta^{(t)})}{P_i(\theta^{(t)})}\Big)\, dx. \\ &= \sum_{i=1}^{m} n_i \frac{1}{P_i(\theta^{(t)})} \int_{a_{i-1}}^{a_i} \log f(x,\theta)\ f(x,\theta^{(t)})dx. \end{aligned} \tag{2.43}$$

**M-step**

The M-step contains the maximization of the Q-function from (2.43), which is the solution of

$$\frac{\partial}{\partial\theta} \sum_{i=1}^{m} n_i \frac{1}{P_i(\theta^{(t)})} \int_{a_{i-1}}^{a_i} \log f(x,\theta)\ f(x,\theta^{(t)})dx = 0 \tag{2.44}$$

In the case of a Gaussian distribution, the solution of the equation (2.44) is not expressible in a closed-form. Therefore, the M-step of the EM algorithm itself requires a further iterative procedure. □

Since the numerical complexity is enormous using a second iterations procedure, Schader and Schmidt (1988) discourage from using the EM algorithm for Gaussian distributions.

## 2.6 The Newton-Raphson Algorithm

The NR algorithm is an iterative procedure for solving a score equation as defined in (2.12) if a closed-form solution can not be found. This is done by an approximation of the gradient vector $S(x, \theta)$ by a linear Taylor series expansion about the current fit $\theta^{(t)}$, see e.g. Little and Rubin (2002). This approach gives

$$S(x, \theta) \approx S(x, \theta^{(t)}) - I(\theta^{(t)}, x)(\theta - \theta^{(t)}), \tag{2.45}$$

where $I(\theta, x)$ is the observed information matrix as defined in (2.13). A new fit $\theta^{(t+1)}$ is obtained by solving the approximated score equation

$$0 \quad = \quad S(x, \theta^{(t)}) - I(\theta^{(t)}, x)(\theta^{(t+1)} - \theta^{(t)}),$$

which gives

$$\theta^{(t+1)} \quad = \quad \theta^{(t)} + I^{-1}(\theta^{(t)}, x)S(x, \theta^{(t)}). \tag{2.46}$$

If the log-likelihood is concave and unimodal, then the sequence of the $\theta^{(t)}$ converges to the maximum likelihood estimate of $\theta$. If the log-likelihood is a quadratic function, the convergence is obtained within one step. This fast convergence is the greatest advantage of the algorithm. However, if the log-likelihood is not concave, the NR algorithm is not guaranteed to converge from an arbitrary initial guess. Only sufficiently accurate starting values cause the desired convergence, see example 2.1.

A combination of the EM algorithm, with its stability according to the choice of the initial values and the NR algorithm may improve the convergence. Hence, it will be one of the considered methods for estimating the parameters of a Gaussian mixture given grouped observations in the next chapter.

# 3 Estimation for Gaussian Mixtures given Grouped Data

In this chapter both afore described data situations are now combined. On the one hand there are observations taken from a mixture distribution and on the other hand these observations are grouped. Starting with an introduction into the problem in the first section, several estimation techniques for the unknown parameters of a mixture distribution given grouped observations are evaluated.

First, two simple methods are presented in section 3.2. Both use a data mapping, where the grouped data are transformed into continuous data to apply the EM algorithm for finite mixtures as presented above. In the first one, a well known approach is investigated, where a new data set is generated by repeating each interval midpoint the number of observations falling into this interval. In the second one, a new approach is investigated. Here, it is assumed that the observations are uniformly distributed on each interval and a corresponding data set is generated to circumvent the grouping data structure.

Second, in section 3.3, an approach of McLachlan and Jones (1988) is investigated. Here the problem is considered within the EM framework, where it is proposed to introduce two missing-data vectors, which results in a closed-form solution for each parameter.

Third, in section 3.4, an approach based on advisements of MacDonald and Pitcher (1979) is considered, where with the choice of the missing-data a complete likelihood function occurs that can not be maximized in a straightforward way. Thus, to obtain a solution three methods are investigated. (1) The idea of the approximated estimation, as presented in the case of grouped observations in section 2.5.1, is combined here with the EM approach. As a results, a new algorithm is received, with which the iterative estimators are given in a closed-form. (2) The problem of not being able to obtain a solution of the score equation analytically, is solved here by a second numerical procedure similar to the EM algorithm for grouped observations, explained in section 2.5.2. (3) An extension of the previous approach is considered, where the EM algorithm is in addition combined with the NR algorithm as proposed by Du (2002).

This chapter ends with the consideration of the problem of suitable initial values in section 3.5. Since each iteration procedure requires starting values and no adequate suggestions seem to be given in literature, some modified techniques are presented.

## 3.1 Likelihood of Mixtures with Grouped Observations

Now, observations coming from different components are considered, where it is not known to which component a single observation belongs to and additionally, the single observations are not observed directly. Only the number of observations $n_i$ falling into an interval with boundaries $[a_{i-1}, a_i)_{i=1,\ldots,m}$ are reported.

Throughout this chapter, suppose X is a random variable with probability density function $g(x, \Psi)$ as defined in (2.1), where $\Psi = (\pi_1, \ldots \pi_k, \theta_1, \ldots \theta_k)$ contains the parameters which are desired to be estimated. The parameter space is denoted, as before, by $\Omega$. A sample $x = (x_1, \ldots, x_n)$ taken from X is grouped into $m$ intervals, with boundaries $a_0, a_1, \ldots, a_m$. The number of observations $n_1, \ldots, n_m$ falling into the intervals $[a_0, a_1)$, $[a_1, a_2)$,..., $[a_{m-1}, a_m)$ are given. The observed data can be written as $y_i = (a_i, n_i)_{i=1,\ldots,m}$ as defined in (2.35).

Estimating the parameters of a mixture model given grouped data with the maximum likelihood method requires an appropriate likelihood function. According to (2.37) such log-likelihood can be written as

$$\log L(\Psi, y) \propto \sum_{i=1}^{m} n_i \ \log P_i(\Psi), \tag{3.1}$$

where

$$\begin{aligned} P_i(\Psi) &:= \int_{a_{i-1}}^{a_i} g(x, \Psi) dx && (3.2) \\ &= \int_{a_{i-1}}^{a_i} \Big( \sum_{j=1}^{k} \pi_j f(x, \theta_j) \Big) dx \\ &= \sum_{j=1}^{k} \pi_j \underbrace{\int_{a_{i-1}}^{a_i} f(x, \theta_j^{(t)}) dx}_{=: P_i(\theta_j)} = \sum_{j=1}^{k} \pi_j P_i(\theta_j). && (3.3) \end{aligned}$$

is a composition of the weighted individual component probabilities. Hence, the log-likelihood for a mixture distribution given grouped data can be written as

$$\log L(\Psi, w) = \sum_{i=1}^{m} n_i \ \log \sum_{j=1}^{k} \pi_j P_i(\theta_j). \tag{3.4}$$

Since no explicit solutions of the corresponding likelihood equations in the case of Gaussian mixtures are available, an iterative procedure is essential.

In the following, several techniques to compute the roots of the Score equations of (3.4) based on the EM approach are presented.

## 3.2 Replacing the Grouped Data Structure

The first two proposed methods are simple, since they generate new data sets to circumvent the grouping data structure. With this new data set, the EM algorithm for mixtures from section 2.3 can be applied in a straightforward way.

### 3.2.1 The RMEM Algorithm

The simplest approach being examined, using a mapping of the grouped data into individual data, is based on an extension of a method of Gjeddebaek (1959). As he used a method to estimate the parameters of a single Gaussian distribution, his approach is adapted here for mixture distributions. However, applying his method directly provides estimators that strongly depend on the underlying mixture distribution. Only a good initial guess of this mixture results in good estimations. To avoid this effect, this approach is combined in this work with the EM framework to improve the estimates, similar to the proceeding in Everitt and Hand (1981).

For a given grouped data set $y$, the boundaries $a_0, a_1, \ldots, a_m$ and frequencies $n_1, \ldots, n_m$ are taken to generate a new data set, where for each interval $[a_{i-1}, a_i)$ the midpoint $\bar{a}_i$ is repeated $n_i$ times. Therefore, this algorithm is denoted by the Repeated Midpoint EM algorithm or briefly the RMEM algorithm.
Each interval contains $n_i$ individual observations $x_i = (x_{i1}, \ldots, x_{in_i})_{i=1,\ldots,m}$. As these individual observations are unknown, they are replaced by the interval midpoint $\bar{a}_i$. These new generated observations are denoted by

$$x^* = (x_1^*, \ldots, x_n^*) := (\underbrace{\bar{a}_1, \ldots, \bar{a}_1}_{n_1-times}, \underbrace{\bar{a}_2, \ldots, \bar{a}_2}_{n_2-times}, \ \ldots \ , \underbrace{\bar{a}_m, \ldots, \bar{a}_m}_{n_m-times}).$$

The new observations are considered to be continuous and the EM algorithm for finite mixture models, introduced in section 2.3, can now be applied.

**Missing-Data**
The missing-data vector

$$z = (z_1, \ldots, z_n) \tag{3.5}$$

contains the k-dimensional indicator variable $(z_s)_{s=1,\ldots,n}$, where $z_{js}$ indicates whether an individual observation $x_s^*$ did or did not arise from the $j$th component.

**E-step**
The Q-function is similar to (2.26) given by

$$Q(\Psi, \Psi^{(t)}) = \sum_{j=1}^{k} \sum_{s=1}^{n} \Big( \log \left( \pi_j \cdot f(x_s, \theta_j) \right) \Big) \, E_{\Psi^{(t)}}[Z_{js} \mid X_s^* = x_s^*]. \tag{3.6}$$

The conditional expectation of $Z_{js}$ is again

$$E_{\Psi^{(t)}}[Z_{js} \mid X_s^* = x_s^*] = \frac{\pi_j^{(t)} f(x_s^*, \theta_j^{(t)})}{g(x_s^*, \Psi^{(t)})} := e_{js}^{(t)}. \tag{3.7}$$

The calculation of the conditional expectation of the log-likelihood for the complete data $w = (y, x^*)$ results equivalently to (2.28) in the following E-step

$$Q(\Psi, \Psi^{(t)}) = \sum_{j=1}^{k} \sum_{s=1}^{n} e_{js}^{(t)} \log \pi_j + \sum_{j=1}^{k} \sum_{s=1}^{n} e_{js}^{(t)} \log f(x_s^*, \theta_j). \tag{3.8}$$

**M-step**

The maximization of this Q-function yields the following iterative estimates, which are analogous to those from section 2.3

$$\pi_j^{(t+1)} = \frac{1}{n} \sum_{s=1}^{n} e_{js}^{(t)} \tag{3.9}$$

$$\mu_j^{(t+1)} = \sum_{s=1}^{n} e_{js}^{(t)} x_s^* \; / \; \sum_{s=1}^{n} e_{js}^{(t)} \tag{3.10}$$

$$\sigma_j^{2(t+1)} = \sum_{s=1}^{n} e_{js}^{(t)} (x_s^* - \mu_j^{(t+1)})^2 \; / \; \sum_{s=1}^{n} e_{js}^{(t)}. \tag{3.11}$$

With appropriate starting values and an adequate stopping criterion, this algorithm can provide excellent results as shown in the simulation study in the next chapter.

### 3.2.2 The UDEM Algorithm

In the previous section it was supposed that all observations of an interval lie in the middle of those. Since this does not reflect the real data situation, a new approach is investigated in this work with the aim to better reflect the spreading of the individual observation in each interval. Hence, it is now assumed that the individual observations are uniformly distributed on each interval. Therefore, this algorithm is named the Uniformly Distributed EM algorithm or briefly the UDEM algorithm.

The proceeding is similar to that of the previous described RMEM algorithm. Again grouped observations in the form (2.35) are considered. As the individual observations of each interval $x_i = (x_{i1}, \ldots, x_{in_i})_{i=1,\ldots,m}$ are unknown, they are replaced by observations $x_i^* = (x_{i1}^*, \ldots, x_{in_i}^*)$, where $(x_{il}^*)_{l=1,\ldots,n_i}$ is randomly drawn from a uniform distribution, i.e.

$$f(x_{il}) = \begin{cases} \frac{1}{a_{i-1} - a_i}, & \text{for } a_{i-1} \leq x_{il} \leq a_i, \\ 0, & \text{else.} \end{cases}$$

This new data set $x^*$ may now be applied by the EM algorithm for finite mixture models as presented in section 2.3. The proceeding may be summarized in the following way.

**Missing-Data**
The missing-data vector from (3.5) is introduced, where $z_{js}$ indicates whether an individual observation $x_s^*$ did or did not arise from the $j$th component.

**E-step**
The E-step is given by (3.8) with $e_{js}^{(t)}$ as defined in (3.7) with the difference that $x^*$ is now replaced by the new generated individual observations.

**M-step**
The maximization in the M-step provides the iterative estimators (3.9), (3.10), and (3.11).

Note that since the generation of the individual observations itself is a random process, the likelihood and, thus, the results vary.

## 3.3 The MGT Algorithm

McLachlan and Jones (1988, 1990) also dealt with the fitting of finite mixture models via the EM algorithm for data, which are available only in grouped form and are also truncated. Furthermore, this problem is considered in McLachlan and Peel (2000), too, where the data is called to be binned. Since the truncation aspect is not of interest in the present work, their considerations were modified for the case that the observations are completely reported in a grouped form.

As the authors mentioned, fitting a mixture distribution to grouped or binned data via the EM algorithm is difficult since the solution of the M-step of the EM algorithm generally does not exist in closed-form. Therefore, they proposed to include a second missing-data vector.

**Missing-Data**
Suppose again X is a random variable with density function $g(x, \Psi)$. The observations $n_1, \ldots, n_m$ have a multinomial distribution consisting of $n$ draws and $m$ categories with probabilities $P_i(\Psi)$ as defined in (3.2). To apply the EM algorithm, a missing-data vector

$$x_i = (x_{i1}, \ldots, x_{in_i})$$

is introduced, which contains the $n_i$ individual observations $x_{il}$ for each interval, where $i = 1, \ldots, m$ and $l = 1, \ldots, n_i$. Using this missing-data the log-likelihood of the complete-data vector $w^* = (y, x_i)$ is given by

$$\log L_c(\Psi, w^*) = \sum_{i=1}^{m} \sum_{l=1}^{n_i} \log g(x_{il}, \Psi). \tag{3.12}$$

Working with (3.12) the M-step of the EM algorithm will itself require again an iterative procedure. Therefore, McLachlan and Jones (1988) additionally proposed an indicator variable denoting the membership of the data with respect to the components of the mixture. This further extension of the complete-data vector includes a zero-one indicator variable

$$z_{il} = (z_{i1l}, \ldots, z_{ikl}) \quad i = 1, \ldots, m, \quad l = 1, \ldots, n_i,$$

where $\sum_{j=1}^{k} z_{ijl} = 1$, the $z_{il}$ gives the membership of an individual observation $x_{il}$ to one of the components and

$$P(z_{ijl} = 1 \mid X_{il} = x_{il}) \quad = \quad \frac{\pi_j f(x_{il}, \theta_j)}{g(x_{il}, \Psi)} =: \tau_j(x_{il}, \Psi).$$

Denoting the new complete data vector by $w = (y, x_i, z_{il})$, the complete log-likelihood becomes

$$\log L_c(\Psi, w) = \sum_{i=1}^{m} \sum_{j=1}^{k} \sum_{l=1}^{n_i} z_{ijl} \big( \log f(x_{il}, \theta_j) + \log \pi_j \big). \tag{3.13}$$

**E-step**

On the $(t+1)$th iteration of the EM algorithm, the E-step requires the calculation of $Q(\Psi, \Psi^{(t)})$, the conditional expectation of the complete log-likelihood, $\log L_c(\Psi, w)$, given the observed data $y$. On taking the conditional expectation of (3.13), it follows that

$$\begin{aligned} Q(\Psi, \Psi^{(t)}) \quad &= \quad E_{\Psi^{(t)}} \Bigg[ \sum_{i=1}^{m} \sum_{j=1}^{k} \sum_{l=1}^{n_i} Z_{ijl} \big( \log f(X, \theta_j) + \log \pi_j \big) \mid Y = y \Bigg] \\ &= \quad \sum_{i=1}^{m} \sum_{j=1}^{k} n_i \; E_{\Psi^{(t)}} \Big[ \tau_j(X, \Psi^{(t)}) \big( \log f(X, \theta_j) + \log \pi_j \big) \mid Y = y \Big]. \end{aligned} \tag{3.14}$$

**M-step**

On the M-step at the $(t+1)$th iteration, $Q(\Psi, \Psi^{(t)})$ has to be maximized. Here, it is assumed that the components are Gaussian. For general descriptions the reader is referred to McLachlan and Peel (2000).

Solving

$$\frac{\partial}{\partial \Psi} \sum_{i=1}^{m} \sum_{j=1}^{k} n_i \; E_{\Psi^{(t)}} \Big[ \tau_j(X, \Psi^{(t)}) \big( \log f(X, \theta_j) + \log \pi_j \big) \mid Y = y \Big] = 0$$

yields the iterative estimators, which are

$$\pi_j^{(t+1)} = \frac{1}{n} \sum_{i=1}^{m} n_i \; E_{\Psi^{(t)}} \big[ \tau_j(X, \Psi^{(t)}) \mid Y = y \big], \tag{3.15}$$

$$\begin{aligned}
\mu_j^{(t+1)} &= \sum_{i=1}^{m} n_i \; E_{\Psi^{(t)}}\big[\tau_j(X,\Psi^{(t)})X \mid Y=y\big] \; / \\
&\quad \sum_{i=1}^{m} n_i E_{\Psi^{(t)}}\big[\tau_j(X,\Psi^{(t)}) \mid Y=y\big], \qquad (3.16)\\
\sigma_j^{2(t+1)} &= \sum_{i=1}^{m} n_i \; E_{\Psi^{(t)}}\big[\tau_j(X,\Psi^{(t)})(X-\mu_i^{(t+1)})^2 \mid Y=y\big] \; / \\
&\quad \sum_{i=1}^{m} n_i E_{\Psi^{(t)}}\big[\tau_j(X,\Psi^{(t)}) \mid Y=y\big]. \qquad (3.17)
\end{aligned}$$

Equations (3.15), (3.16), and (3.17) may be expressed in several ways. Defining $P_i(\Psi)$ and $P_i(\theta_j)$ as in (3.2), the iterative estimators can be written as

$$\begin{aligned}
\pi_j^{(t+1)} &= \frac{1}{n}\sum_{i=1}^{m}\pi_j^{(t)}\frac{n_i P_i(\Psi^{(t)})}{P_i(\theta_j^{(t)})},\\
\mu_j^{(t+1)} &= \mu_j^{(t)} - \sigma_j^{2(t)}\Big\{\big(\sum_{i=1}^{m}\frac{n_i}{P_i(\theta_j^{(t)})}(f(a_i,\theta_j^{(t)}) - f(a_{i-1},\theta_j^{(t)}))\big)/\sum_{i=1}^{m}\frac{n_i P_i(\Psi^{(t)})}{P_i(\theta_j^{(t)})}\Big\}\\
\sigma_j^{2(t+1)} &= \sigma_j^{2(t)}\Bigg\{1+\Big(2\mu_j^{(t+1)}-\mu_j^{(t)}\Big)\Big(\sum_{i=1}^{m}\frac{n_i}{P_i(\theta_j^{(t)})}(f(a_i,\theta_j^{(t)}) - f(a_{i-1},\theta_j^{(t)}))\Big)/\\
&\quad \Big(\sum_{i=1}^{m}\frac{n_i P_i(\Psi^{(t)})}{P_i(\theta_j^{(t)})}\Big) - \Big(\sum_{i=1}^{m}\frac{n_i}{P_i(\theta_j^{(t)})}(a_i f(a_i,\theta_j^{(t)}) - a_{i-1}f(a_{i-1},\theta_j^{(t)}))\Big)/\\
&\quad \Big(\sum_{i=1}^{m}\frac{n_i P_i(\Psi^{(t)})}{P_i(\theta_j^{(t)})}\Big)\Bigg\} + \Big(\mu_i^{(t+1)}-\mu_i^{(t)}\Big)^2.
\end{aligned}$$

This approach was implemented by McLachlan and Jones (1990) as algorithm AS 254 in Fortran77. The Subroutine was called `MGT` and uses several other Fortran routines like `ALNORM` (= ASb66), `CHOL` (= ASb6) and `SYMINV` (= ASb7). Since Fortran77 is obsolete, this algorithm was rewritten in R for the present simulation study.

## 3.4 MacDonalds Approach

The following three methods have the choice of the missing-data in common. Instead of regarding the individual observations as missing, as have been done by McLachlan and Peel (2000), this approach considers the number of observations of each component to be unknown.

This approach was first described by MacDonald and Pitcher (1979) for the analysis of fisheries length-frequency data. The main idea of this approach is to consider the log-likelihood obtained if the number of observations of each component were known.

**Missing-Data**

The following missing-data vector is introduced

$$n_i^* = (n_{i1}^*, \ldots, n_{ik}^*), \qquad (3.18)$$

where $(n^*_{ij})_{i=1,...,m,j=1,...,k}$ denotes the number of observations from the $j$th component falling into the $i$th interval and $N^*_{ij}$ is the corresponding random variable. The log-likelihood of the complete-data $w = (w_1, ..., w_m)$ with $w_i = (y_i, n^*_i)_{i=1,...,m}$ can therefore be written as

$$\log L_c(\Psi, w) = \sum_{i=1}^{m}\sum_{j=1}^{k} n^*_{ij} \log\big(\pi_j \cdot P_i(\theta_j)\big), \tag{3.19}$$

where as before $P_i(\theta_j)$ denotes the probability that an individual observation from the $j$th component falls into the $i$th interval, as defined in (3.3).

**E-step**

The E-step of the EM algorithm is now given by

$$\begin{aligned} Q(\Psi, \Psi^{(t)}) &= E_{\Psi^{(t)}}\big[\log L_c(\Psi, W) \mid Y = y\big] \\ &= \sum_{j=1}^{k}\sum_{i=1}^{m} \log\big(\pi_j \cdot P_i(\theta_j)\big)\; E_{\Psi^{(t)}}\big[N^*_{ij} \mid Y = y\big]. \end{aligned} \tag{3.20}$$

The conditional expectation of $N^*_{ij}$ can again be calculated with Bayes' theorem and yields

$$E_{\Psi^{(t)}}\big[N^*_{ij} \mid Y = y\big] = \frac{n_i \pi_j^{(t)} P_i(\theta_j^{(t)})}{P_i(\Psi^{(t)})} =: e^{*(t)}_{ij}, \tag{3.21}$$

where $i = 1, \ldots, m$, $j = 1, \ldots, k$ and $P(\Psi)$ as defined in (3.2).
Thus, (3.20) can be written as

$$\begin{aligned} Q(\Psi, \Psi^{(t)}) &= \sum_{j=1}^{k}\sum_{i=1}^{m} \log\big(\pi_j \cdot P_i(\theta_j)\big) \cdot e^{*(t)}_{ij} \\ &= \sum_{j=1}^{k}\sum_{i=1}^{m} e^{*(t)}_{ij} \log \pi_j + \sum_{j=1}^{k}\sum_{i=1}^{m} e^{*(t)}_{ij} \log P_i(\theta_j). \end{aligned} \tag{3.22}$$

**M-step**

The M-step requires the calculation of

$$\frac{\partial}{\partial \Psi} \sum_{j=1}^{k}\sum_{i=1}^{m} e^{*(t)}_{ij} \log \pi_j + \sum_{j=1}^{k}\sum_{i=1}^{m} e^{*(t)}_{ij} \log P_i(\theta_j) = 0$$

Since $\pi_j$ again appears only in the first term of the right hand side of (3.20), the required maximization of the M-step can again be done separately. Applying again Bayes' theorem and regarding the condition $\sum_{j=1}^{k} \pi_j = 1$, the maximization results similarly to (2.21) in

$$\pi_j^{(t+1)} = \frac{1}{m} \sum_{i=1}^{m} e^{*(t)}_{ij}, \tag{3.23}$$

which gives the iterative estimator of the mixing proportions $(\pi_j)_{j=1,...,k}$.

The maximization of the right term of the right hand side of (3.20) provides the iterative estimators for $\theta_j$, which apparently depend on the distribution components $f(x, \theta_j)$. In the case of a Gaussian mixture distribution, there exists no analytical solution for

$$\frac{\partial}{\partial \mu_j} \sum_{j=1}^{k} \sum_{i=1}^{m} e_{ij}^* \log P_i(\theta_j) = 0 \quad \text{and} \quad \frac{\partial}{\partial \sigma_j} \sum_{j=1}^{k} \sum_{i=1}^{m} e_{ij}^* \log P_i(\theta_j) = 0, \quad \text{respectively.}$$

The following three methods provide techniques to solve this problem.

### 3.4.1 The APEM Algorithm

The basic idea of this approach is that the probability that an observation from the $j$th component falls into the $i$th interval, which is $P_i(\theta_j)$, is approximated by the integrand evaluated at the midpoint of the interval, similar to the suggested approximation of Gjeddebaek (1959) in section 2.5. This idea is combined in the present work with the approach of the method described above and the new obtained method is called the approximated EM algorithm or briefly APEM.

**Missing-Data**
As before, the observations are grouped into $m$ intervals with boundaries $a_0, a_1, \ldots, a_m$ and frequencies $n_1, \ldots, n_m$. The missing-data vector again is defined by

$$n_i^* = (n_{i1}^*, \ldots, n_{ik}^*)_{i=1,\ldots,m},$$

which gives the number of observations of each component in each interval. The probability that an individual observation from the $j$th component falls into the $i$th interval is now approximated by

$$P_i(\theta_j) \approx h \cdot f(\bar{a}_i, \theta_j), \tag{3.24}$$

where $i = 1, \ldots, m$, $j = 1, \ldots, k$, and where $h$ denotes the interval width and $\bar{a}_i$ the midpoint of the $i$th interval. Thus, the following complete log-likelihood occurs

$$\log L_c(\Psi, w) = \sum_{i=1}^{m} \sum_{j=1}^{k} n_{ij}^* \log \left(\pi_j \cdot h \cdot f(\bar{a}_i, \theta_j)\right). \tag{3.25}$$

**E-step**
Using this approximation in the E-step of the EM algorithm gives

$$\begin{aligned} Q(\Psi, \Psi^{(t)}) &= E_{\Psi^{(t)}}\left[\log L_c(\Psi, W) \mid Y = y\right] \\ &= \sum_{i=1}^{m} \sum_{j=1}^{k} \log \left(\pi_j \cdot h \cdot f(\bar{a}_i, \theta_j)\right) \; E_{\Psi^{(t)}}\left[N_{ij}^* \mid Y = y\right]. \end{aligned}$$

The conditional expectation of $N_{ij}^*$ is now

$$E_{\Psi^{(t)}}\left[N_{ij}^* \mid n_i\right] = \frac{n_i \pi_j^{(t)} h f(\bar{a}_i, \theta_j^{(t)})}{\sum_{r=1}^{k} \pi_r^{(t)} h f(\bar{a}_i, \theta_r^{(t)})} = \frac{n_i \pi_j^{(t)} f(\bar{a}_i, \theta_j^{(t)})}{g(\bar{a}_i, \Psi^{(t)})} =: e_{ij}^{**(t)}. \tag{3.26}$$

Hence, the Q-function can be written as

$$Q(\Psi, \Psi^{(t)}) = \sum_{i=1}^{m}\sum_{j=1}^{k} e_{ij}^{**(t)} \log \pi_j + \sum_{i=1}^{m}\sum_{j=1}^{k} e_{ij}^{**(t)} \log\left(h \cdot f(\bar{a}_i, \theta_j)\right). \tag{3.27}$$

**M-step**

Maximization of the left term of the right hand side of (3.27) provides analogously to (3.23) the iterative estimator for $\pi_j$, which is

$$\pi_j^{(t+1)} \quad = \quad \frac{1}{m}\sum_{i=1}^{m} e_{ij}^{**(t)}. \tag{3.28}$$

The score equation of the right term of the right hand side of (3.27) can now be calculated analytically and yields the following iterative estimators for $\mu_j$ and $\sigma_j^2$

$$\mu_j^{(t+1)} \quad = \quad \sum_{i=1}^{m} e_{ij}^{**(t)} \bar{a}_i / \sum_{i=1}^{m} e_{ij}^{**(t)} \tag{3.29}$$

$$\sigma_j^{2^{(t+1)}} \quad = \quad \sum_{i=1}^{m} e_{ij}^{**(t)} (\bar{a}_i - \mu_j^{(t+1)})^2 / \sum_{i=1}^{m} e_{ij}^{**(t)}. \tag{3.30}$$

### 3.4.2 The EMNR Algorithm

This method to obtain a maximum of the likelihood function in (3.4) is based on a modification of an approach developed by MacDonald and Green (1988). They implemented and published this approach as a computer program called MIX, which analyzes histograms by finding a set of distributions that gives the best fit to the histograms.
Recall that the aim is still to maximize the Q-function from (3.4). As has been shown, a solution for that problem can not be obtained in a closed-form, which therefore results in the application of the EM algorithm.

**Missing-Data**

As before, the missing-data $n^*$ from (3.18) is introduced.

**E-step**

The E-step is

$$Q(\Psi, \Psi^{(t)}) = \sum_{i=1}^{m}\sum_{j=1}^{k} e_{ij}^{*(t)} \log \pi_j + \sum_{i=1}^{m}\sum_{j=1}^{k} e_{ij}^{*(t)} \log P_i(\theta_j). \tag{3.31}$$

**M-step**

The maximization with respect to $\pi_i$ can be done straight forward as described before. The maximization with respect to $\mu_i$ and $\sigma_i$ can not be calculated analytically. Hence, it is proposed to use another iteration procedure to obtain a solution. In fact, it is proposed to use the NR algorithm as described in section 2.6.

Starting with suitable initial values $\pi_j^{(0)}, \mu_j^{(0)}$, and $\sigma_j^{(0)}$, the first iteration of the EM-algorithm, including the NR algorithm in the M-step, yields new estimated parameters $\pi_j^{(1)}, \mu_j^{(1)}$ and $\sigma_j^{(1)}$. Replacing the previous starting values with the new obtained parameters leads to $\pi_j^{(2)}, \mu_j^{(2)}$ and $\sigma_j^{(2)}$. Both steps are then repeated until convergence is achieved. In brief, this procedure contains two iterations procedures - the EM algorithm and in the M-step the NR algorithm. Therefore, this algorithm is named EMNR algorithm.

As mentioned before, the NR algorithm has good convergence properties and, consequently, the iterative estimators are obtained very fast. However, as the main iterative procedure is an EM algorithm, this method may converge slowly and may not satisfy the chosen convergence properties at all, which is demonstrated in the simulation study in chapter 4. To speed up this algorithm, an expansion is proposed by Du (2002) and explained in the next section.

### 3.4.3 The MIX Algorithm

The previously described method may converge very slowly. Hence, Du (2002) proposed a combination of the previous EMNR algorithm with another NR algorithm to speed up this algorithm. As mentioned previously, "the Newton-type algorithm converges more quickly than the EM algorithm, but, in general, only when starting from an initial value very close to the minimum" (Du (2002), p. 46). He therefore took advantage of the good properties of each method and suggested to perform one step of the EM algorithm, as presented in section 3.4.2, to improve the initial values and to apply subsequently the NR algorithm, as introduced in section 2.6, to estimate the final parameters. This obtained algorithm is implemented into the function `mix`, which is a part of the R package `mixdist`.
To remind, the aim is to maximize

$$\log L(\Psi, w) = \sum_{i=1}^{m} n_i \ \log \sum_{j=1}^{k} \pi_j P_i(\theta_j) \tag{3.32}$$

with respect to $\Psi$. Introducing the missing-data vector $n^*$ as before and choosing some starting values $\pi_j^{(0)}, \mu_j^{(0)}$, and $\sigma_j^{(0)}$ leads to the following E-step of the EM algorithm

$$Q(\Psi, \Psi^{(0)}) = \sum_{i=1}^{m} \sum_{j=1}^{k} e_{ij}^{*(0)} \log \pi_j + \sum_{i=1}^{m} \sum_{j=1}^{k} e_{ij}^{*(0)} \log P_i(\theta_j). \tag{3.33}$$

The maximization with respect to $\pi_j^{(0)}, \mu_j^{(0)}$, and $\sigma_j^{(0)}$ is performed as presented in the previous section with the help of the NR algorithm, which provides new estimates $\pi_j^{(1)}, \mu_j^{(1)}$, and $\sigma_j^{(1)}$. However, instead of repeating the EM iteration with these new estimates as before, the new parameters are now used as initial values for the NR algorithm in order to solve the score equation with the likelihood of (3.32). Hence, the EM algorithm is only applied to improve the starting values while the main iteration is performed by the NR algorithm.

This procedure has the advantage that the solution is obtained much faster than by using the previous method, since the NR algorithm converges faster than the EM algorithm. Nevertheless, this method contains a combination of three iteration procedures, since the EM algorithm itself requires in the M-step a NR algorithm, too. This increases the time needed for the performance compared to methods that are using only one iteration procedure.

A huge benefit of using the `mix` function is, that it is already implemented. Besides the estimation of the parameters of a mixture composed of two normal distributions, this package offers the decomposition of mixtures of Weibull, Poisson, Negative Binomial, Binomial, Gamma, and Log-Normal distributions with up to 15 components. Previous knowledge about the components, for example, equal variances or fixed mixing proportions can also be added as constraints. To my knowledge, this package is the only one written in R, which deals with mixture models given grouped observations. Hence, it is assumed that this approach is the most known and applied.

Nevertheless, Du (2002) mentioned that the estimation of the parameters of a mixture distribution is difficult when the components are heavily overlapped. The overlapping obscures information about the individual components and, indeed, this effect is reproduced by the simulation study shown in chapter 4.

### 3.4.4 Similarity of the APEM and the RMEM Algorithm

**Lemma 3.1** *The APEM and RMEM algorithms provide the same iterative estimators, which are*

$$\pi_j^{(t+1)} = \frac{1}{m}\sum_{i=1}^{m} e_{ij}^{**(t)}, \tag{3.34}$$

$$\mu_j^{(t+1)} = \sum_{i=1}^{m} e_{ij}^{**(t)}\bar{a}_i / \sum_{i=1}^{m} e_{ij}^{**(t)}, \tag{3.35}$$

$$\sigma_j^{2(t+1)} = \sum_{i=1}^{m} e_{ij}^{**(t)}(\bar{a}_i - \mu_j^{(t+1)})^2 / \sum_{i=1}^{m} e_{ij}^{**(t)}, \tag{3.36}$$

*where*

$$e_{ij}^{**(t)} = \frac{n_i \pi_j^{(t)} f(\bar{a}_i, \theta_j^{(t)})}{g(\bar{a}_i, \Psi^{(t)})}.$$

**Proof 3.1** Recall, the Q-function from the APEM algorithm is given by

$$\begin{aligned} Q(\Psi, \Psi^{(t)}) &= \sum_{i=1}^{m}\sum_{j=1}^{k} e_{ij}^{**(t)} \log \pi_j + \sum_{i=1}^{m}\sum_{j=1}^{k} e_{ij}^{**(t)} \log\left(h \cdot f(\bar{a}_i, \theta_j)\right). \\ &= \sum_{i=1}^{m}\sum_{j=1}^{k} e_{ij}^{**(t)} \log \pi_j + \sum_{i=1}^{m}\sum_{j=1}^{k} e_{ij}^{**(t)} \Big(\log\left(f(\bar{a}_i, \theta_j)\right) + \log(h)\Big) \end{aligned} \tag{3.37}$$

with $e_{ij}^{**(t)}$ as defined in (3.37). The Q-function of the RMEM algorithm is given by

$$Q(\Psi, \Psi^{(t)}) = \sum_{j=1}^{k}\sum_{s=1}^{n} e_{js}^{(t)} \log \pi_j + \sum_{j=1}^{k}\sum_{s=1}^{n} e_{js}^{(t)} \log f(x_s^*, \theta_j), \tag{3.38}$$

with

$$e_{js}^{(t)} = \frac{\pi_j^{(t)} f(x_s^*, \theta_j^{(t)})}{g(x_s^*, \Psi^{(t)})}. \tag{3.39}$$

Since the generated individual observations $(x_s^*)_{s=1,\ldots,n}$ of the RMEM algorithm consist of the midpoints $(\bar{a}_i)_{i=1,\ldots,m}$ of each interval, the $e_{js}^{(t)}$ of (3.39) may be reparameterized as follows

$$\sum_{s=1}^{n} e_{js}^{(t)} = \sum_{s=1}^{n} \frac{\pi_j^{(t)} f(x_s^*, \theta_j^{(t)})}{g(x_s^*, \Psi^{(t)})} = \sum_{i=1}^{m} n_i \frac{\pi_j^{(t)} f(\bar{a}_i, \theta_j^{(t)})}{g(\bar{a}_i, \Psi^{(t)})} = \sum_{i=1}^{m} e_{ij}^{**(t)}. \tag{3.40}$$

Hence, the Q-function of the RMEM algorithm (3.38) can be rewritten as

$$Q(\Psi, \Psi^{(t)}) = \sum_{i=1}^{m}\sum_{j=1}^{k} e_{ij}^{**(t)} \log \pi_j + \sum_{i=1}^{m}\sum_{j=1}^{k} e_{ij}^{**(t)} \log f(\bar{a}_i, \theta_j), \tag{3.41}$$

which differs from the Q-function of the APEM algorithm (3.37) only by an additive term containing $h$. A maximization of both Q-functions leads now to the same iterative estimators (3.34), (3.35), and (3.36).

□

**Remark 3.1** Especially when the interval width $h$ is 1, the Q-functions of both methods are equal and, therefore, the estimates are the same.

For interval widths unequal to 1, the final estimations may differ because both Q-functions are not equal and each algorithm stops at a different iterative step. Nevertheless, the equality of the estimators of both algorithms is an interesting finding as both methods use different approaches.

The difference in the approaches involve different implementations, hence, the complexity of the calculation varies. The implementation of the RMEM algorithm simply requires a data mapping and the use of the EM algorithm for finite mixture models, which is already

implemented in several statistical software, among others in the R package `mixtools`. The implementation of the APEM algorithm and the calculation of the estimators are more complicated and require more time. A detailed comparison is provided by the simulation study in chapter 4. Before that, some techniques for obtaining suitable starting values are presented.

## 3.5 Techniques for Choosing Initial Values

Several investigations on the right choice of initial values have been made previously. For instance Karlis and Xekalaki (2003) compared some techniques for obtaining starting values for the EM algorithm for finite mixtures. Since these techniques assume individual observations, none of them can be used directly in the context presented here, where the observations are only available in a grouped form. An intensive literature search could not reveal previous works on this topic, thus, new strategies for obtaining initial values for mixtures given grouped observations are proposed. Most of them are based on modifications of techniques presented in Karlis and Xekalaki (2003). In particular, these techniques are presented for mixtures with two Gaussian components.

**Starting Value Technique S1**
First, the technique of Finch et al. (1989), briefly introduced in section 2.4, is adapted and modified to determine initial values for mixtures given grouped observations. Finch et al. (1989) claimed that "[...] the optimal estimates of the means are closely approximated by the arithmetic means of the split sample and the optimal estimated variance is close to the weighted average of the variance of the split sample" (p. 1021). They obtained the splitting by randomly generating a value from a uniform distribution for $\pi_1$. Then, the smallest $n \cdot \pi_1$ observations are taken to be the first part and the remaining observations are the second part. Since this proceeding turns out to be too vague, Karlis and Xekalaki (2003) proposed to use equal mixing proportions. To achieve the splitting, the number of observations from the first part of the sample is denoted by $n_1$, where

$$n_1 := \begin{cases} \frac{n}{2}, & \text{if n even,} \\ \frac{n+1}{2}, & \text{if n uneven.} \end{cases}$$

Then, the starting values for the component means and variances are

$$\mu_1^{(0)} = \frac{1}{n_1} \sum_{s=1}^{n_1} x_s \tag{3.42}$$

$$\mu_2^{(0)} = \frac{1}{n - n_1} \sum_{s=n_1+1}^{n} x_s \tag{3.43}$$

$$\sigma_{1,2}^{2(0)} = \Big( \sum_{s=1}^{n_1} (x_s - \mu_1^{(0)})^2 + \sum_{s=n_1+1}^{n} (x_s - \mu_2^{(0)})^2 \Big) / (n-2). \tag{3.44}$$

In cases when $n$ is uneven, the sum of (3.43) and the second sum in (3.44) start from $n_1$ instead of $n_1 + 1$.

Since the individual observations are not available in our context, a data mapping like presented in section 3.2.1 is proposed, where the midpoint of each interval is repeated dependent on the number of observations in each interval.

A similar proceeding was proposed by Schader and Schmidt (1988) for a single Gaussian distribution. They also recommended to include the middle of each interval and suggested to use $\mu^{(0)} = 1/n \sum_{i=1}^{m} n_i \bar{a}_i$ as an initial value for the mean of the Gaussian distribution. Their idea was adapted here and with the resulting data set, the starting values for $\mu_j$ and $\sigma_j^2, j = 1, 2$ can be calculated from (3.42), (3.43), and (3.44). This technique is denoted by S1.

**Starting Value Technique S2**

This approach is adapted from Böhning et al. (1994) and denoted by S2. They used equal mixing proportions, too, and made the following suggestions for the component means

$$\mu_1^{(0)} = x_1 + 1/2 \tag{3.45}$$
$$\mu_2^{(0)} = x_n - 1/2, \tag{3.46}$$

"[..] since well separated values have often turned out to be a good strategy for avoiding local maxima which are not global ones" (p. 381). Although they did not make further specifications on the choice of the initial values for the variance, Karlis and Xekalaki (2003) proposed to use the following values

$$\sigma_{1,2}^{2(0)} = \bar{\sigma}^2 - \Big(((\mu_1^{(0)} - \bar{x})^2 + (\mu_2^{(0)} - \bar{x})^2)/2\Big), \tag{3.47}$$

where $\bar{x}$ and $\bar{\sigma}^2$ are the mean and the variance of the sample with individual observations. To obtain individual observations, the afore mentioned mapping is suggested. Note that if the variance estimate $\sigma_{1,2}^{2(0)}$ came out to be negative, $\bar{\sigma}^2/2$ is used instead. The proceeding is equivalent to the one used by Karlis and Xekalaki (2003).

**Starting Value Technique S3**

The following technique based on the range of the sample likewise considered by Karlis and Xekalaki (2003) is indexed by S3. Starting again with equal mixing proportions, the initial values for the component means are now

$$\mu_1^{(0)} = \min_s(x_s) + \frac{d}{3} \tag{3.48}$$
$$\mu_2^{(0)} = \min_s(x_s) + \frac{2d}{3}, \tag{3.49}$$

where $s = 1, \dots n$ and $d$ is the range of the sample. Once again, this approach requires individual observations. Hence, the previously suggested mapping is applied here, too.
The starting values for the component standard deviation are chosen to be

$$\sigma_1 = \sigma_2 = 1/2 \cdot \bar{\sigma}, \tag{3.50}$$

where $\bar{\sigma}$ is the standard deviation of the transformed sample.

**Starting Value Technique S4**
The fourth considered approach was proposed by Seidel et al. (2000). This technique also involves the assumption that the mixing proportions are equal. The initial values for the component means are now

$$\mu_1 = \bar{x} - 1/2 \cdot \bar{\sigma} \tag{3.51}$$

$$\mu_2 = \bar{x} + 1/2 \cdot \bar{\sigma}, \tag{3.52}$$

where $\bar{x}$ and $\bar{\sigma}$ are the mean and the standard deviation of the sample. Since the individual observations are as well required here, a data mapping is suggested as describes earlier. The starting values for the component variances equal the variance of the whole transformed sample. This technique is denoted by S4.

**Starting Value Technique S5**
Finally, a new technique indexed by S5 is proposed. Once again, a data mapping is essential. This technique is the only one that uses estimated values for the mixing proportions. Therefore, at first, the middle of the sample range $d$ is determined, which is

$$\bar{d} := \min_s(x_s) + d/2. \tag{3.53}$$

Then, the number of observations from the first part of the sample, denoted by $N_1$, divided by the sample size $n$ is chosen to be the starting value for the first mixing proportion. The second part, denoted by $N_2$, divided by $n$, yields equivalently a first guess for the second mixing proportion. Both terms are summarized

$$\pi_1 = N_1/n \tag{3.54}$$

$$\pi_2 = N_2/n. \tag{3.55}$$

The starting values for the component means are calculated in a similar way, as suggested by Böhning et al. (1994). The authors used the smallest observation and added 0.5 to obtain a starting value for the first component mean. Since the minimum of the samples may be an outlier, which would be inappropriate, the proposed technique instead considers the 1% percentile of the sample. In the same manner, the calculation of the second component

mean is performed, but now, the 99% percentile is chosen. The component means are

$$\mu_1 = q_{0.01} \quad \text{and} \quad \mu_2 = q_{0.99}. \tag{3.56}$$

The starting values for the common component variance are determined by the approach of Finch et al. (1989) given in equation (3.44).

In the following the results of the simulation studies are presented. One investigation (subsection 4.7) addresses the comparison of these five techniques against each other. A recommendation will be given for different mixture models. Additionally, the previously presented methods for the estimation of the parameters of a Gaussian mixture will be compared. First, the true values will be used as starting values and subsequently starting values obtained by technique S1 - S5 will by applied.

# 4 Simulation Comparison

In this chapter the procedure and the results of the simulation study are outlined. Starting in section 4.1 with the description of the implementation, this chapter turns to the specification of the comparison criteria, which is followed by the presentation of the investigated situations and mixture models. For convenience, the simulation study is presented for Gaussian mixtures with two components. Since the $(\pi_j)_{j=1,2}$ sum to one, one of them is redundant. Hence $\pi_2$ will not be specified and $\Psi = (\pi_1,\ \mu_1,\ \mu_2,\ \sigma_1,\ \sigma_2)$.
The main simulation results are shown in the sections 4.2, 4.3, and 4.4, where the afore mentioned mixture models are considered. A variation of the sample size can be found in section 4.6. Section 4.7 addresses the problem of different initial values by presenting results of different starting values obtained by the techniques proposed in the previous section. Finally, the influence of the interval width on the estimation is investigated and presented in section 4.8.

## 4.1 Simulation Study

All simulations were performed with the program R using additional packages.

**Package mixdist[1]**
The `mixdist` package was implemented by MacDonald (2011) with contributions from Du. This package contains the function `mix`, which is a realization of the method described in section 3.4.3, where only one step of the EM algorithm is performed to improve the initial values and then the main iteration is obtained by the Newton algorithm. A modification of this function is implemented in the new function `emnr`, where the main iteration is performed by the EM algorithm and the M-step contains the NR algorithm as explained in section 3.4.2. Furthermore, a modification of the function `mixgroup` from the `mixdist` package was implemented here to achieve the grouping of the individual observations, where the borders of the intervals may be specified.

**Package mixtools[2]**
To generate observations coming from a mixture distribution, the function `rnormmix` from the package `mixtools` was used. It was implemented by Young et al. (2010). This package is designed for analyzing finite mixture models with continuous data. Especially for the estimation of the parameters of a two component Gaussian mixture, the mixtools package

[1] available at http://cran.r-project.org/web/packages/mixdist/index.html
[2] available at http://cran.r-project.org/web/packages/mixtools/index.html

offers a function named `normalmixEM2comp`. The authors state that the implementation of the EM algorithm bases on the specifications from McLachlan and Peel (2000) as described in section 2.3. Actually, it could be revealed that this is not exactly the case. Their function differs from the original descriptions by the estimations of the component variances $\sigma_1^2$ and $\sigma_2^2$. Their algorithm estimates $\sigma_j^2$ on the $(t+1)$th iteration by using the estimate of $\mu_j$ from the $t$-th iteration. Since it is more common to apply $\mu_j$ from the $(t+1)$-th iteration, a modification of this function is given in this work, denoted by `gmix2`. This new function contains now the adequate realization of the EM algorithm for finite mixtures as shown in McLachlan and Peel (2000).

### 4.1.1 Comparison Criteria

The following results are based on simulations with 1000 samples. For comparison reasons, the relative change in the log-likelihood was chosen as stopping criterion and the tolerance is set to $10^{-4}$. To avoid problems that occur if an algorithm is trapped, if the convergence is very slow, or in the case of divergence, the algorithms stop if the number of iterations exceeds 500. In such a case, the corresponding sample is excluded from the analysis, which is a similar handling likewise described by Everitt (1984). Since each algorithm fails in a different number of samples, the convergence failure rate is an important criterion in the comparison of the algorithms.

Furthermore, the main focus is on the quality of the estimation of $\Psi$, measured by the discrepancy between the estimated and the true parameters. Due to the fact that 1000 samples are considered, the repeated estimations provide a distribution of each parameter estimator. Accordingly, the mean and the standard deviation of these distributions are estimated so that the mean squared error (MSE) may be calculated, which is defined as

$$MSE(\widehat{\theta}) = \mathrm{Var}(\widehat{\theta}) + (E(\widehat{\theta}) - \theta)^2,$$

where $\theta$ is the parameter, which is desired to be estimated and $\widehat{\theta}$ is an estimation of $\theta$. The MSE incorporates both the accuracy and the precision of an estimator and was used for comparison reasons by several authors, e.g. Hosmer (1973b). However, since the MSE is not scale independent, this measurement seems not to be adequate for the comparison between the parameters and the methods. Instead, it is proposed to use the relative root mean squared error, which is abbreviated as rRMSE. This measurement is obtained by taking the root of the MSE, which gives the root mean squared error (RMSE), also known as root mean squared deviation (RMSD). Deviating the RMSE by the true value of the parameter results in the rRMSE, which is

$$rRMSE(\widehat{\theta}) = \frac{\sqrt{MSE(\widehat{\theta})}}{\theta}.$$

To compare the methods the cumulated rRMSE is considered, which is the sum of the rRMSE of each parameter. With this measurement each parameter has the same influence on the quality of the results and a comparison between the methods with regards to the parameters is possible.

In addition, a graphical comparison is included. Since the average parameter estimations provide an average distribution for each method, the corresponding density functions are displayed to demonstrate the differences between the methods. To get an impression of the dimension of the discrepancy between the true values and the estimated ones, the density function with the true parameters is as well depicted. For a clear illustration, the densities obtained by the six methods and the reference density are abbreviated and color-coded as shown in table 4.1.

Besides the graphical highlighting and the consideration of the cumulated rRMSE, the methods are also compared by the average number of iterations and the average time needed for their performance. Since the EMNR and MIX algorithms consist of two and three iterative procedures, respectively, the average number of iterations of the main iteration procedure is taken, which is the EM algorithm in the EMNR method and the NR algorithm in the MIX method. The other iterations are neglected.

Table 4.1: Abbreviations and color coding of the methods.

| Color | Abbr. | Method | Section |
|---|---|---|---|
| ■ | TRUE | True parameter | |
| ■ | RMEM | Repeated midpoint EM algorithm | 3.2.1 |
| ■ | UDEM | Uniformly distributed EM algorithm | 3.2.2 |
| ■ | MGT | EM algorithm proposed by McLachlan and Jones (1988) | 3.3 |
| ■ | APEM | Approximated EM algorithm | 3.4.1 |
| ■ | EMNR | EM algorithm with NR algorithm | 3.4.2 |
| ■ | MIX | Combined algorithms proposed by Du (2002) | 3.4.3 |

### 4.1.2 Investigated Mixture Models

In order to examine the performance of the methods in various situations, several parameter constellations are considered. Authors like Hosmer (1973a), Day (1969) and Hasselblad (1969), pointed out that problems with divergence arise, when the components are not well separated. Therefore, in the present work, several mixture models with different degrees of separation are investigated. Throughout all simulations, the parameters of the first component are fix with mean equal to five and standard deviation equal to one. The standard deviation of the second Gaussian distribution is chosen to be also 1, while the mean varies. Thereby, according to Everitt (1996), an increasing separation between the two components is considered. The difference $\Delta$ between the means is one, two, three and four.

At first, the mixing proportions are chosen to be equal, i.e. $\pi_1 = \pi_2 = 0.5$ and only the mean of the second distribution varies. Starting with highly overlapped components, where the means are $\mu_1 = 5$ and $\mu_2 = 6$ , the second component is shifted three steps until two well separated components occur. This yields a bimodal distribution, where $\mu_2 = 9$. The corresponding four mixture models are denoted by M1, M2, M3, and M4 and are displayed in the first row of figure 4.1.

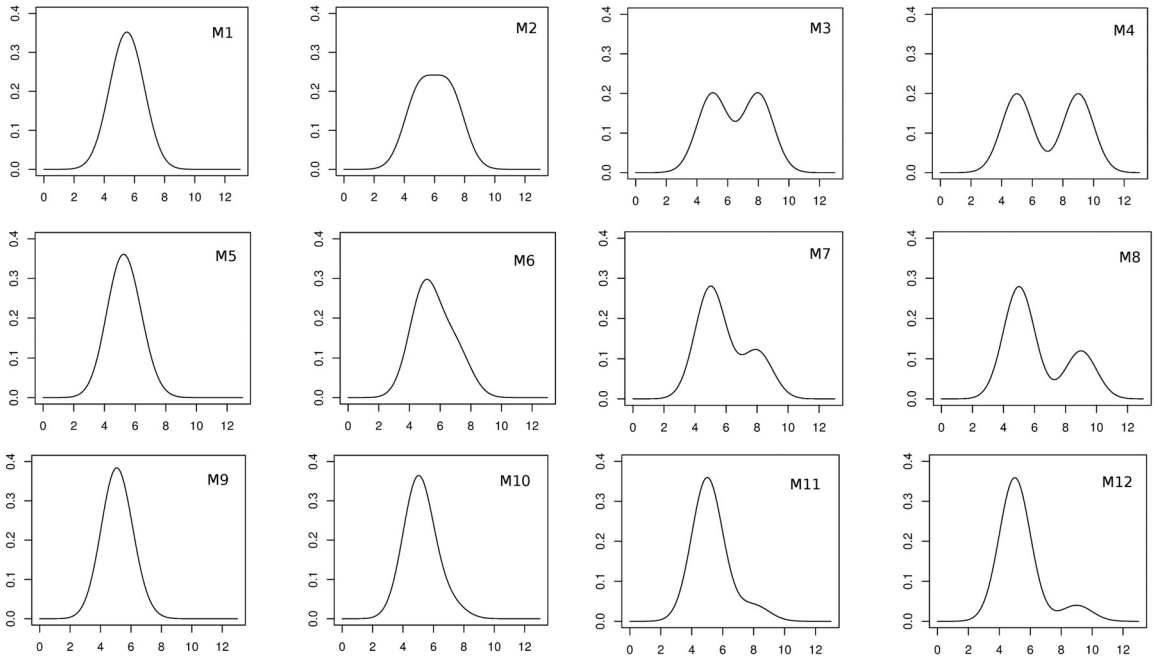



Figure 4.1: Investigated mixture models. First row: $\Psi = (0.5,\ 0.5,\ 5,\ \mu_2,\ 1,\ 1)$, second row: $\Psi = (0.7,\ 0.3,\ 5,\ \mu_2,\ 1,\ 1)$, third row: $\Psi = (0.9,\ 0.1,\ 5,\ \mu_2,\ 1,\ 1)$, where $\mu_2 = 6,\ 7,\ 8$ and 9, respectively.

Besides the well balanced components, two other cases are investigated. In the first case, the first mixing weight is chosen to be 0.7 and the second 0.3. The four different separated mixture models described above are also investigated with these unbalanced proportions. The corresponding mixture models are denoted by M5, M6, M7, and M8 and are shown in the second row of figure 4.1. The second case includes a further unbalance of the components. Here, one main component with $\pi_1 = 0.9$ and a small second component with $\pi_2 = 0.1$ are chosen. The last row of figure 4.1 illustrates these mixture models, which are denoted by M9, M10, M11, and M12.

As the component means and the mixing proportions are in most practical situations of greatest interest, attention has been focused on these values. A variation of the standard deviation has not been investigated explicitly. However, an attempt to generalize the results for a wide range of mixture models is given in section 4.8.2.

Furthermore, the sample size plays an important role on the convergence of an algorithm. Day (1969) and Hasselblad (1969) noted the need for large samples to achieve adequate results. Additionally, Hosmer (1973a) gave at least empirical evidence that sample sizes smaller than 300 may be unreliable. However, in a medical context, particularly in clinical trials when infrequent diseases are investigated, it is not uncommon that the population consists only of a few hundred patients. Thus, it is important to know the accuracy of the method in such sample size dimensions. Therefore, three different situations are considered: a) $n = 100$, b) $n = 300$, and c) $n = 500$.

Finally, as mentioned in section 2.2, the convergence of the algorithms can not be guaranteed. Even if an algorithm converges, the likelihood may have several local maxima and the result depends on the initial values. To minimize the error caused by a wrong choice of the starting values, at first, the true values are taken as initial values. However, since the true values are seldom known in practice, a comparison between several initial values obtained by the previously presented techniques is discussed in section 4.7.

In addition, the interval width of the grouped observations has also an influence on the convergence and the results. In the first investigations the interval width $h$ is set to be 1. Nevertheless, in some practical situations, the interval width may be wider or thinner. Therefore, several interval widths are discussed in section 4.8.

## 4.2 Comparison Results for $\pi_1 = 0.5$

This section contains the comparison of the methods for the mixture models M1 - M4, where the mixing proportions are equal. This case is the simplest one, since half of the observations belongs to the first component and the other half to the second.
All simulations in this and the following two sections are based on samples with 100 observations. The interval width $h$ is chosen to be 1 and the true parameters are taken as initial values.

### 4.2.1 Mixture Model M1

The first considered mixture model M1 is well balanced and highly overlapped with the following true parameters

$$\Psi = (0.5,\ 0.5,\ 5,\ 6,\ 1,\ 1).$$

**Convergence Investigation**
In the comparison of the methods, the focus is at first on the convergence properties of each algorithm. Since divergence and the exceeding of the maximum number of iterations occurred in several cases, the samples without convergence were excluded and the analysis was done with the remaining samples. It turned out that the MIX algorithm did not satisfy

the chosen convergence criteria in 21% of the samples. With the EMNR algorithm no results were obtained in 8%. The APEM and RMEM algorithms did not provide a solution in 1.1% of the samples. Both the UDEM and MGT algorithms did not have any convergence problems. These results are summarized in table 4.2.

**Graphical Consideration**

As each simulated sample provides parameter estimations, a distribution for each parameter can be obtained. Taking the mean of these distributions leads to average parameter estimations. These values can be used to illustrate the corresponding average densities, which are displayed in figure 4.2. For this and the following mixture models, the underlying numerical results, i.e. the average means and, additionally, the average standard deviations of each parameter can be found in table A.1 in the appendix.
Besides the densities with the average estimated parameters obtained by the six methods, figure 4.2 also shows the density with the true parameters, which officiates as reference.

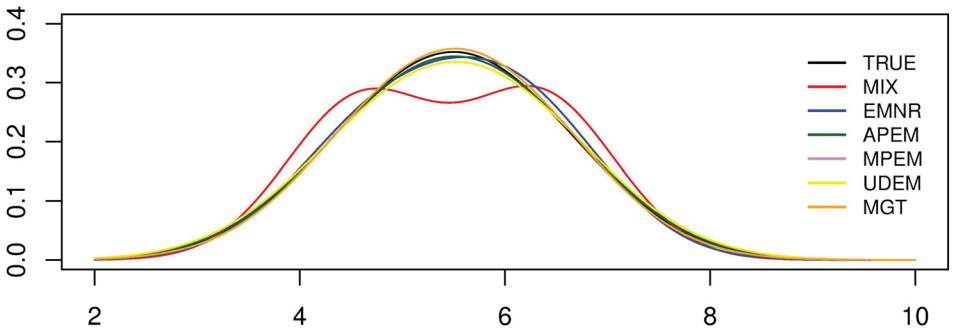



Figure 4.2: Mixture densities with average estimated parameters for mixture model M1.

For this parameter constellation, the first that can be noticed, is that all curves are close to each other except the red one, which gives the distribution obtained by the MIX function. With two obvious peaks, this curve has a totally different appearance than the others including the reference one. Hence, the discrepancy between the true and the parameters estimated by MIX is very large. This corresponds to the results of Du (2002), who explained this demeanor by the high overlapping and the obscuring of each component. To circumvent this problem, he proposed to include additional knowledge about the parameters to handle the difficulties that arise with heavily overlapped components. However, in practise often further information about the parameters does not exist. Thus, the MIX method seems to be less appropriate for this data situation.

Furthermore, a closer examination of figure 4.2 indicates that the curve obtained by the estimations of the EMNR algorithm, is shifted to the righthand side compared to the curve with the true values. However, this discrepancy is obviously not as large as that of the curve provided by MIX. Hence, it can be deduced that the EMNR method provides on

average a better fit than the MIX method.

**Numerical Comparison**

The same conclusion can be drawn considering table 4.2, where, besides the average runtime and the average number of iterations, the rRMSE for each parameter and the cumulated rRMSE for each method are presented.

Table 4.2: The rRMSEs, cumulated rRMSEs, average runtimes in seconds, average number of iterations, and convergence failure rates in % for mixture model M1.

| Method | $\pi_1$ | $\mu_1$ | $\mu_2$ | $\sigma_1$ | $\sigma_2$ | Cum | Time | Iter | Failures |
|---|---|---|---|---|---|---|---|---|---|
| MIX | 0.648 | 0.208 | 0.157 | 0.475 | 0.457 | 1.945 | 0.676 | 35.1 | 21.0 |
| EMNR | 0.443 | 0.129 | 0.103 | 0.318 | 0.356 | 1.349 | 16.364 | 161.0 | 8.0 |
| APEM | 0.075 | 0.041 | 0.034 | 0.169 | 0.165 | 0.484 | 0.010 | 9.2 | 1.1 |
| RMEM | 0.075 | 0.041 | 0.034 | 0.169 | 0.165 | 0.484 | 0.004 | 9.2 | 1.1 |
| UDEM | 0.104 | 0.045 | 0.038 | 0.184 | 0.179 | 0.550 | 0.003 | 9.9 | 0 |
| MGT | 0.080 | 0.040 | 0.035 | 0.189 | 0.192 | 0.536 | 0.030 | 11.6 | 0 |

Although the MIX and EMNR algorithms provide the same average estimations for the mixing proportion, which is 0.492 as presented in table A.1, the standard deviation of the estimation obtained by MIX is much higher, 0.324 compared to 0.221. This higher standard deviation results in a higher rRMSE as can be seen in table 4.2. Furthermore, the discrepancies between the average estimations of the component means and standard deviations and the corresponding true values are much larger using the MIX method than applying the EMNR algorithm, which accordingly has an effect on the rRMSE.

In summary, with a cumulated rRMSE of rounded 1.95, MIX has a large uncertainty about all parameter estimations. In comparison, the cumulated rRMSE of the EMNR method is 1.35, which reinforces the superiority of this method over the MIX method. This finding is very interesting since the aim of Du (2002) was to speed up the EM algorithm by creating a hybrid algorithm by combining the EM algorithm and the NR algorithm. Indeed, MIX is much faster than EMNR - the average time for one performance is 0.68 in comparison to 16.36 seconds - but this is at the expense of the accuracy.

Turning to the estimation results of the APEM algorithm, it can be seen that the results are remarkably better in comparison to the MIX and EMNR algorithms. With an obviously smaller cumulated rRMSE (0.49), the APEM methods is superior. Since those three methods use the same approach and a direct comparison is apposite. Moreover, one may come to the conclusion that the approximation of the probability, that a single observation falls into a chosen interval, causes a smaller error than the application of a second or a third numerical procedure. In addition, avoiding a second iteration also has a positive effect on the runtime. With an average time of 0.01 seconds per run, the APEM algorithm

is clearly faster than the MIX and EMNR method.
The advantage of a simplification can also be seen considering the RMEM and UDEM algorithms, which both circumvent the grouping structure through a data mapping. Not only the implementation is much easier, but also the runtime benefits from the simple approach. With an average time of 0.004 and 0.003 seconds per run, respectively, both methods are faster than the others. As the runtime is correlated with the number of iterations, similar results are achieved considering the number of iterations. In addition, the accuracy of these simple methods is surprisingly good. The cumulated rRMSE of the RMEM algorithm is 0.49 and that of the UDEM algorithm is 0.55.
The value of the RMEM algorithm is the same as that of the APEM algorithm, since we have seen in section 3.4.1, that for interval width $h = 1$ the APEM and RMEM algorithms provide the same results although they use different approaches.

Furthermore, comparing the average estimation results of RMEM and UDEM algorithms directly, table A.1 shows that both methods estimated the same mixing proportions. However, the standard deviation is a bit higher using the UDEM algorithm. The same effect can be seen in the component means and standard deviations. Hence, even if the estimation results of UDEM are on average close to the true values, the deviations are slightly larger. This causes higher rRMSEs in the parameters and, therewith, a higher cumulated rRMSE. In conclusion, the estimates can not be improved by taking the spreading of the individual observations within an interval into consideration. Only with respect to convergence an improvement is achieved, since the UDEM algorithm provides solutions in samples where the RMEM algorithm fails.

Finally, figure 4.2 shows that the curve obtained by the estimation results of the MGT method, is a bit higher than the reference curve. This can be explained by the underestimation of the component standard deviations, which are 0.941 and 0.931, respectively, see table A.1. Nevertheless, since the other estimations are close to the true values, the cumulated rRMSE is 0.54, which indicates acceptable estimation results.

**Summary of Mixture Model M1**
In summary, fitting highly overlapped mixture models to grouped observations is not naturally accompanied with inferior estimations. Moreover, simple and convenient tractable methods like the RMEM algorithm provide estimation results that are on average close to the true parameters. In contrast, methods with more than one iteration procedure like the MIX algorithm fail to construct a right decomposition of the two components.

### 4.2.2 Mixture Model M2

The second well balanced mixture model M2 leads to slightly better separated components with true values

$$\Psi = (0.5,\ 0.5,\ 5,\ 7,\ 1,\ 1).$$

**Convergence Investigation**

Compared to the previous mixture model, in this mixture model less convergence problems occurred. In fact, the MIX algorithm stopped iteration due to convergence problems only in 11.5% of the samples, which is a distinct improvement compared to the failure rate in the previous mixture model. The EMNR algorithm did not provide a solution in 2.1% of the samples and the APEM and RMEM methods failed in 0.6%. The UDEM and the MGT algorithms were again free of convergence problems, as can be seen in table 4.3.

**Graphical Consideration**

The graphical presentation of the average parameter estimations and the corresponding densities can be found in figure 4.3. The numerical results are available in table A.1 in the appendix.

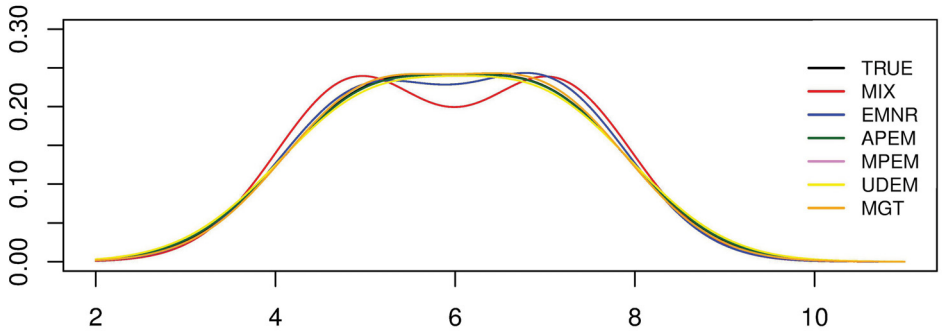



Figure 4.3: Mixture densities with average estimated parameters for mixture model M2.

Results similar to the previously discussed mixture model occur. The MIX and EMNR algorithms result in curves clearly different from the other one. This indicates again a less accurate fit compared to methods using only one iteration procedure. Comparing separately the curves of the MIX and EMNR algorithms to the reference one, a higher difference of the curve associated with the results of MIX is observed. This higher discrepancy becomes even more apparent considering the numerical results of table A.1.

Except of the MIX and EMNR algorithms, the other methods provide on average estimation results, which are close to the true values, since their curves can hardly be distinguished from the reference one.

**Numerical Comparison**

Considering the numerical results given in table A.1, the MIX algorithm, provides for the first and the second standard deviation on average a value of about 0.73, which is a large discrepancy compared to the true values of 1. In contrast, the average estimations obtained by the EMNR algorithm for the standard deviations are 0.86 and 0.82, respectively, which is closer to the true values.

Moreover, the other parameters are also estimated less accurate with a higher variability using the MIX and EMNR algorithms compared to the other methods, which use one iteration procedure. This also is reflected in higher rRMSEs as presented in table 4.3. The cumulated rRMSE of the MIX algorithm is 1.53, while that of the EMNR algorithm is 1.20. Nevertheless, both are slightly better in comparison to the previous mixture model M1. There, the cumulated rRMSE was 1.95 and 1.35, respectively.

Additionally, having a look at the runtime of these two methods, a superiority of the MIX algorithm can be seen - 0.5 compared to 11.5 seconds. However, as before, this benefit is achieved at the cost of accuracy.

The results obtained by the APEM algorithm, which uses the same approach as MIX and EMNR, are again remarkably better. It is hard to distinguish in figure 4.3 the corresponding curve from the reference one, which indicates on average a good fit. Indeed, all parameter estimations obtained by APEM are closer to the true values than by the MIX or EMNR algorithms. In fact, the cumulated rRMSE (0.55) is much smaller, which underlines the superiority of the APEM algorithm over the MIX and EMNR algorithms.

Table 4.3: The rRMSEs, cumulated rRMSEs, average runtimes in seconds, average number of iterations, and convergence failure rates in % for mixture model M2.

| Method | $\pi_1$ | $\mu_1$ | $\mu_2$ | $\sigma_1$ | $\sigma_2$ | Cum | Time | Iter | Failures |
|---|---|---|---|---|---|---|---|---|---|
| MIX | 0.526 | 0.158 | 0.104 | 0.379 | 0.367 | 1.534 | 0.519 | 23.3 | 11.5 |
| EMNR | 0.414 | 0.112 | 0.079 | 0.273 | 0.323 | 1.201 | 11.538 | 127.2 | 2.1 |
| APEM | 0.146 | 0.046 | 0.033 | 0.169 | 0.156 | 0.550 | 0.010 | 9.0 | 0.6 |
| RMEM | 0.146 | 0.046 | 0.033 | 0.169 | 0.156 | 0.550 | 0.004 | 9.0 | 0.6 |
| UDEM | 0.148 | 0.047 | 0.035 | 0.176 | 0.169 | 0.575 | 0.003 | 8.8 | 0 |
| MGT | 0.145 | 0.046 | 0.033 | 0.181 | 0.168 | 0.573 | 0.031 | 10.6 | 0 |

As already mentioned in the previous model, the APEM and RMEM algorithms provide comparable estimation results since the interval width is still 1. Therefore, the RMEM algorithm provides also estimations that are in average close to the true values. Since implementation is easier, the RMEM method is even faster - 0.003 compared to 0.004 seconds.

Furthermore, comparing the results of the RMEM and UDEM algorithms with those obtained by the MGT algorithm, it can be seen that the results of the simple methods are approximately as good as those of the MGT algorithm, which is more complex. In particular, the cumulated rRMSE of the RMEM algorithm is even smaller - 0.550 compared to 0.573. This is a surprising finding, since the RMEM and UDEM algorithms assume that the real observations lie somewhere between the interval borders. Hence, one may conclude that it is not relevant where within the interval the observations are located, as long as the number of observations in each interval is specified.

Moreover, having a look at the runtime and the iterations that are required, it can be seen that the RMEM and UDEM algorithms need less time and iterations for a performance than the MGT algorithm. In fact, the UDEM method is about ten times faster than MGT. A direct comparison between the RMEM and UDEM algorithms shows no substantial differences regarding the average runtime and number of iterations. Though, the estimation results are on average closer to the true values, if they are obtained by the RMEM algorithm. Therefore, this method is to be preferred.
However, if, due to convergence problems, a solution can not be found with the RMEM algorithm, the UDEM algorithm may be applied, since this algorithm provided a solution for each considered sample.

Although the MGT algorithm has a higher cumulated rRMSE than the RMEM algorithm and its performance is slower, this algorithm may also be applied in data situation like the one considered here. The mixing proportions and the component means are estimated as accurate as by the RMEM algorithm. Only the component standard deviations are on average slightly underestimated, which leads to clearly better results than by the MIX and EMNR algorithms.

**Summary of Mixture Model M2**
For the mixture model M2, algorithms with one iteration procedure, APEM, RMEM, UDEM, and MGT, on average provide parameter estimations that are closer to the true values than algorithms with more iteration procedures like EMNR and MIX. Moreover, the estimates of the former are obtained faster and have a smaller deviation. Therefor, they should be preferred in similar data situations.

### 4.2.3 Mixture Model M3

In this mixture model, the difference between the component means is three. Hence, the true values are

$$\Psi = (0.5,\ 0.5,\ 5,\ 8,\ 1,\ 1).$$

**Convergence Investigation**

This mixture model is accompanied by two clearly identifiable components, which results in more satisfying convergence properties in each method. The MIX method did not provide a solution in only 0.5% of the samples. The EMNR method had no convergence problems at all. The APEM and RMEM algorithms stopped iterations in 0.2% of the samples due to convergence problems. As before, the UDEM and the MGT method provided a solution for each generated sample as it is shown in table 4.4.

**Graphical Consideration**

A graphical fitting of the distributions obtained by the average parameter estimations can be found in the following figure.

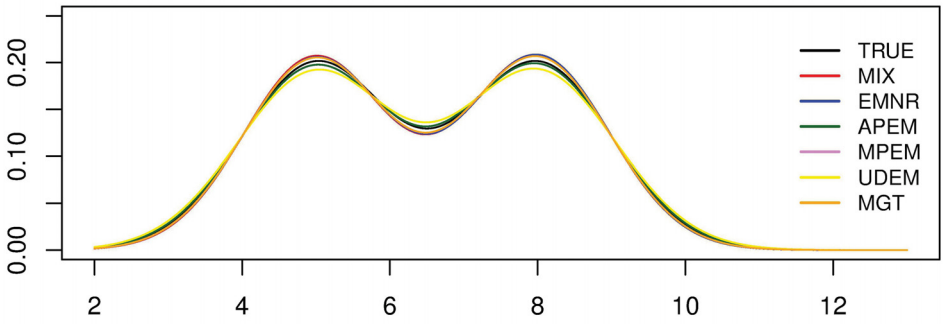



Figure 4.4: Mixture densities with average estimated parameters for mixture model M3.

It is clearly visible that the more the two components are separated the better are the results. There is only a hardly difference between the reference curve, representing the true density, and the other curves, standing for the densities obtained by the average estimations of each method.

Especially the MIX and EMNR algorithms provide much better results than in the previous highly overlapped mixture models M1 and M2. The corresponding curves are much closer to that one of the true density than before. Nevertheless, both densities are slightly higher in the two peaks. This can be explained by the underestimation of the standard deviation, see table 4.4.

**Numerical Comparison**

Using the MIX and EMNR algorithms leads to larger deviances in the parameters and

accordingly higher rRMSEs compared to the other methods. Even though, the results are a distinct improvement in comparison to those obtained by the highly overlapped mixture models M1 and M2, which emphasizes the ability of both methods to decompose the components of bimodal mixtures. Moreover, besides the accuracy, the runtime also benefits from the better separation. With an average time of 0.32 and 3.94 seconds, respectively, the MIX and EMNR algorithms are much faster than before.

A closer examination of figure 4.4 reveals that the curves, representing the results obtained by the the APEM and UDEM algorithms, slightly deviate from the reference one. The reason is that both methods underestimate the component means and overestimate the standard deviations, see table A.1. However, the deviates of these average estimations and so the corresponding rRMSEs are small. In fact, the cumulated rRMSEs are 0.55 for the APEM algorithm and 0.59 for the UDEM algorithm. Hence, also for this mixture model, these methods provide estimates that are close to the true values.
Since the estimators of the APEM and RMEM algorithms are identical for interval width $h = 1$, the same holds for the RMEM algorithm, with the difference that this algorithm is approximately twice as fast as the APEM algorithm.

Table 4.4: The rRMSEs, cumulated rRMSEs, average runtimes in seconds, average number of iterations, and convergence failure rates in % for mixture model M3.

| Method | $\pi_1$ | $\mu_1$ | $\mu_2$ | $\sigma_1$ | $\sigma_2$ | Cum | Time | Iter | Failures |
|---|---|---|---|---|---|---|---|---|---|
| MIX | 0.252 | 0.077 | 0.048 | 0.243 | 0.244 | 0.864 | 0.324 | 12.4 | 0.5 |
| EMNR | 0.228 | 0.069 | 0.043 | 0.214 | 0.224 | 0.778 | 3.936 | 48.0 | 0 |
| APEM | 0.152 | 0.048 | 0.030 | 0.159 | 0.161 | 0.550 | 0.009 | 8.6 | 0.2 |
| RMEM | 0.152 | 0.048 | 0.030 | 0.159 | 0.161 | 0.550 | 0.004 | 8.6 | 0.2 |
| UDEM | 0.157 | 0.051 | 0.031 | 0.173 | 0.179 | 0.591 | 0.003 | 8.7 | 0 |
| MGT | 0.153 | 0.048 | 0.030 | 0.170 | 0.169 | 0.570 | 0.032 | 10.2 | 0 |

Finally, likewise the MIX and EMNR algorithms, the MGT algorithm underestimates the parameters, but this discrepancy is lower as by MIX or EMNR. In fact, the deviates and so the rRMSEs are much smaller. Hence, the cumulated rRMSE (0.57) is comparable with those of the APEM and UDEM algorithms.

**Summary of Mixture Model M3**

In conclusion, the MIX and EMNR algorithms benefit most from the better separation of the components. In particular, the more separate the components are the better are the results obtained by these methods. This will be seen even more clearly in the following mixture model, where the two component are much more separated. In contrast, this huge improvement can not be confirmed for the other methods. They provide estimates that are similarly accurate as those obtained for the highly overlapped mixture models.

### 4.2.4 Mixture Model M4

In this mixture model the difference between the component means is four. Hence, the mixture density shows two well separated peaks. The true parameters are

$$\Psi = (0.5,\ 0.5,\ 5,\ 9,\ 1,\ 1).$$

**Convergence Investigation**

Since the two components can be identified clearly, none of the methods failed at the convergence. Thus, all simulated samples are incorporated into the analysis.

**Graphical Consideration**

The obtained average densities and the reference density are shown in figure 4.5.

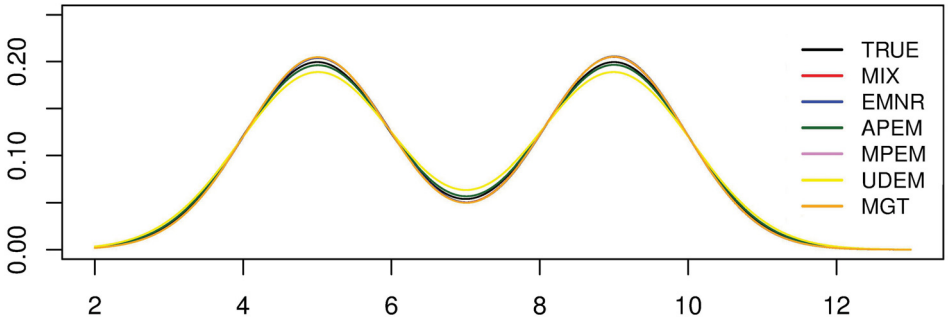



Figure 4.5: Mixture densities with average estimated parameters for mixture model M4.

For this well separated mixture model, all methods result on average estimations that are close to the true parameters. This can be seen by the small difference between the curves obtained by the methods and the reference one. However, the curves of the MIX, EMNR and MGT algorithms are higher in the peaks than the reference curve, which indicates an underestimation of the component standard deviations. Even though, in contrast to the previous considered mixture models, particularly considering MIX and EMNR, an improvement is observable.

**Numerical Comparison**

Table 4.5 shows that the rRMSEs of MIX and EMNR are much smaller than before and almost comparable to the results obtained by the other methods. In particular, the EMNR method provides estimates that are almost as good as those of the UDEM algorithm and which are slightly inferior to those of the MGT algorithm. As before, in comparison to the MIX algorithm, the results of the EMNR algorithm are superior. Moreover, having a look at the average number of iterations and time required for a performance, also a large improvement of the EMNR algorithm due to the more separated components is visible.

A direct comparison between the MIX and EMNR algorithms to the simplified APEM algorithm again demonstrates the superiority of the latter with respect to accuracy and runtime. Additionally, the cumulated rRMSE of 0.44 indicates the best fit of all considered methods. As before, the same is valid for the RMEM algorithm, which is even faster.

Table 4.5: The rRMSEs, cumulated rRMSEs, average runtimes in seconds, average number of iterations, and convergence failure rates in % for mixture model M4.

| Method | $\pi_1$ | $\mu_1$ | $\mu_2$ | $\sigma_1$ | $\sigma_2$ | Cum | Time | Iter | Failures |
|---|---|---|---|---|---|---|---|---|---|
| MIX | 0.120 | 0.040 | 0.023 | 0.164 | 0.163 | 0.510 | 0.278 | 8.7 | 0 |
| EMNR | 0.118 | 0.039 | 0.022 | 0.155 | 0.156 | 0.490 | 1.163 | 14.5 | 0 |
| APEM | 0.111 | 0.035 | 0.020 | 0.138 | 0.134 | 0.438 | 0.006 | 4.7 | 0 |
| RMEM | 0.111 | 0.035 | 0.020 | 0.138 | 0.134 | 0.438 | 0.003 | 4.7 | 0 |
| UDEM | 0.115 | 0.037 | 0.022 | 0.159 | 0.153 | 0.486 | 0.002 | 5.3 | 0 |
| MGT | 0.111 | 0.036 | 0.020 | 0.146 | 0.144 | 0.457 | 0.021 | 6.3 | 0 |

Furthermore, figure 4.5 shows that the curve representing the average estimation results of the UDEM algorithm slightly deviates from the reference curve. This can be explained by the overestimation of the standard deviation, see table A.1. Nevertheless, the discrepancies of this average estimations are small and, consequently, the cumulated rRMSE is with a value of 0.48 almost as good as that of the APEM or MGT algorithm. Moreover, with an average number of iterations of 5.3 and an average time of 0.002 seconds per run, the UDEM algorithm represents one of the fastest algorithms.

For the remaining MGT algorithm, similar results as with the previous presented mixture models are obtained. Particularly, the component standard deviations are underestimated, which was already assumed by the consideration of the corresponding density. However, since on average the deviance in each parameter is small, the cumulated rRMSE is 0.46, which is slightly higher than that one obtained by the best algorithm APEM. In fact, considering the mixing proportions and the component means, the estimations of RMEM, UDEM and MGT are equivalent.

**Summary of Mixture Model M4**

In conclusion, mixture distributions with well separated components are decomposed more easily. In particular, in such cases, methods with two or more iteration procedures like MIX and EMNR provide estimates that are much closer to the true values than with highly overlapped mixtures. This effect can not be observed that clearly with the other methods. The APEM algorithm again provides the best results in such well separated mixture models.

## 4.3 Comparison Results for $\pi_1 = 0.7$

In this section, results for the mixture models M5 - M8 are discussed, where the mixing proportion of the first component is 0.7 and that one of the second 0.3. This choice of the mixing proportions leads to the existence of a main population and a subpopulation. This is an interesting case, because in many practical applications, the existence of a subpopulation changes further analysis significantly. In particular, when components are highly overlapped, one might misleadingly suppose that the underlying distribution is a single Gaussian distribution. This can cause that Gaussian based procedures are applied, although the Gaussian assumption is not valid. Hence, it is important to know how reliable an algorithm is with regard to the decomposition of a mixture with unbalanced mixing proportions. Therefore, the simulation results and the analysis of the four mixture models M5 - M8 are presented in the following.

### 4.3.1 Mixture Model M5

The mixing components of this mixture model are highly overlapped with a difference between the component means equal to one. The mixing proportions are $\pi_1 = 0.7$ and $\pi_2 = 0.3$, which yields the following true parameters

$$\Psi = (0.7,\ 0.3,\ 5,\ 6,\ 1,\ 1).$$

**Convergence Investigation**

As shown before, highly overlapped mixtures are accompanied with high convergence failure rates, which can also be observed in this mixture model. Similar to mixture model M1, the MIX algorithm failed in 22% of the samples. The EMNR algorithm did not provide a solution in 7.2%, and the APEM and RMEM algorithms determined iteration due to convergence problems in 2.1%. As before, the UDEM and the MGT were free of convergence problems. These results are displayed in table 4.14.

**Graphical Consideration**

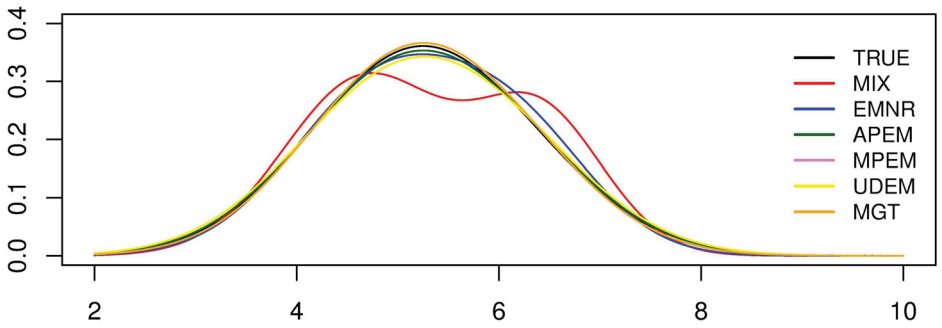



Figure 4.6: Mixture densities with average estimated parameters for mixture model M5.

Figure 4.6 demonstrates the average estimation results by showing the mixing densities obtained by the average parameter estimations for each method. The corresponding numerical results are presented in table A.1 in the appendix.

Similar results as in mixture model M1 were achieved. The density of the mixture distribution with the average estimation results obtained by the MIX algorithm differs distinctly from the curve with the true values. As before, this can be explained by the large underestimation of the mixing proportions as shown in table A.1. In contrast, the discrepancy between the curve of the EMNR algorithm and the reference one is smaller, which indicates that this method achieves on average estimates that are closer to the true values compared to MIX.

The other methods provide on average estimations which give densities close to the reference one. In particular, the curve obtained by the MGT algorithm can hardly be distinguished from the density with the true parameters.

**Numerical Comparison**

Considering the cumulated rRMSEs from table 4.6, it can be seen that almost all methods provide results worse than those of the highly overlapped distribution with balanced mixing proportions. It is more difficult to decompose a mixture containing a main distribution and a subdistribution than a mixture with balanced components. This effect increases the more the mixing proportions are unbalanced, as will be seen in the next section.

Table 4.6: The rRMSEs, cumulated rRMSEs, average runtimes in seconds, average number of iterations, and convergence failure rates in % for mixture model M5.

| Method | $\pi_1$ | $\mu_1$ | $\mu_2$ | $\sigma_1$ | $\sigma_2$ | Cum | Time | Iter | Failures |
|---|---|---|---|---|---|---|---|---|---|
| MIX | 0.469 | 0.163 | 0.18 | 0.396 | 0.560 | 1.768 | 0.713 | 36.3 | 22 |
| EMNR | 0.306 | 0.096 | 0.125 | 0.261 | 0.454 | 1.242 | 17.196 | 163 | 7.2 |
| APEM | 0.070 | 0.034 | 0.043 | 0.137 | 0.220 | 0.504 | 0.010 | 9.5 | 2.1 |
| RMEM | 0.070 | 0.034 | 0.043 | 0.137 | 0.220 | 0.504 | 0.004 | 9.5 | 2.1 |
| UDEM | 0.086 | 0.041 | 0.050 | 0.149 | 0.249 | 0.575 | 0.003 | 11.1 | 0 |
| MGT | 0.072 | 0.035 | 0.043 | 0.152 | 0.247 | 0.549 | 0.030 | 11.6 | 0 |

Furthermore, almost all findings from mixture model M1 are also valid in this mixture model. For example, the APEM algorithm again provides much better results than the MIX or EMNR algorithm does. With a cumulated rRMSE of 0.50, this method calculates estimates that are closest to the true values and, in addition, have the smallest deviances. Considering the runtime and the number of iterations, no distinct difference compared to mixture model M1 can be observed. Due to the equality of the estimators of the APEM and RMEM algorithms, the same results are obtained by the RMEM algorithm. However,

in this mixture model, the RMEM algorithm is about twice as fast as the APEM algorithm.

Looking at the estimations of the two components separately, it can be seen, that the estimations of the mean and the standard deviation of the first component are closer to the true values and the divergences are smaller than those of the second component. Table 4.7 shows this effect for each method. In the first row, the sum of the rRMSE of $\mu_1$ and $\sigma_1$ is displayed and the second row shows the cumulated rRMSE of $\mu_2$ and $\sigma_2$.

Table 4.7: Cumulated rRMSEs separately listed for the components of mixture model M5.

| Component | MIX | EMNR | APEM | RMEM | UDEM | MGT |
|---|---|---|---|---|---|---|
| 1th | 0.559 | 0.357 | 0.171 | 0.171 | 0.190 | 0.187 |
| 2nd | 0.740 | 0.579 | 0.263 | 0.263 | 0.299 | 0.290 |

It becomes clear, that all methods provide better results for the first component than for the second, which can be explained by the unbalance. The first component consists of 70 observations and the estimations of the second component are based on only 30 observations. The lower the number of observations incorporated into the analysis, the higher are the discrepancies. This effect increases, the more unbalanced the components are.

**Summary of Mixture Model M5**

In brief, all methods using only one iteration procedure are again superior in accuracy, time and number of failures compared to those containing more numerical procedures. As before, the APEM and RMEM algorithms provide estimations that are closest to the true values with the fewest deviation. They are followed by the UDEM and MGT algorithms.

Moreover, the parameters of the first component are estimated better than of the second one. The former are even better than those obtained by the well balanced mixture model M1.

### 4.3.2 Mixture Model M6

In this mixture model, the difference between the component means is two and the true parameters are

$$\Psi = (0.7,\ 0.3,\ 5,\ 7,\ 1,\ 1).$$

**Convergence Investigation**
As before, this mixture model causes less convergence problems than the highly overlapped mixture model M5. To be specific, the MIX algorithm did not provide a solution in 15.9% of the samples. The EMNR algorithm terminated iterations in 2.5% and the APEM and RMEM algorithms in 0.9% of the samples. As before, the UDEM and MGT algorithms were free of convergence problems. These results are summarized in table 4.8.

Compared to the balanced mixture model M2, these convergence failure rates are higher, except for the UDEM and MGT algorithms, which are still zero.

**Graphical Consideration**
The average densities in figure 4.7 emphasize once again that the MIX and EMNR algorithms distinctly and visibly deviate from the reference curve with the true parameters. In contrast, the other curves can barely be distinguished from the reference one. They provide on average estimations close to the true values.

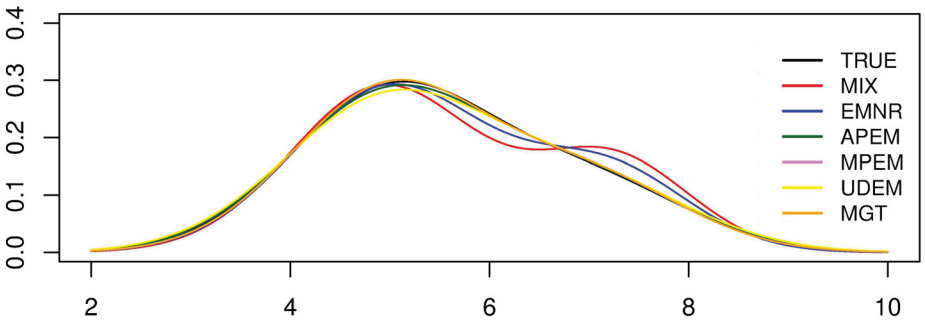



Figure 4.7: Mixture densities with average estimated parameters for mixture model M6.

**Numerical Comparison**
Having a look at the rRMSEs, which are displayed in table 4.8, it is obvious that the APEM and so the RMEM algorithms convince once again with the most accurate estimations in each parameter. Moreover, these methods and the UDEM algorithm are again the fastest ones.

Comparing these results with those obtained by mixture model M5, a slight improvement of the estimates in almost each method is visible. Particularly, the MIX and EMNR al-

gorithms benefit from the better separation and provide superior estimations compared to mixture model M5. However, this improvement is minor and both methods are inferior compared to the other algorithms.

Table 4.8: The rRMSEs, cumulated rRMSEs, average runtimes in seconds, average number of iterations, and convergence failure rates in % for mixture model M6.

| Method | $\pi_1$ | $\mu_1$ | $\mu_2$ | $\sigma_1$ | $\sigma_2$ | Cum | Time | Iter | Failures |
|---|---|---|---|---|---|---|---|---|---|
| MIX | 0.339 | 0.096 | 0.131 | 0.269 | 0.460 | 1.295 | 0.553 | 23.9 | 15.9 |
| EMNR | 0.266 | 0.071 | 0.108 | 0.214 | 0.430 | 1.089 | 12.032 | 121.1 | 2.5 |
| APEM | 0.091 | 0.033 | 0.039 | 0.131 | 0.213 | 0.507 | 0.009 | 8.5 | 0.9 |
| RMEM | 0.091 | 0.033 | 0.039 | 0.131 | 0.213 | 0.507 | 0.003 | 8.5 | 0.9 |
| UDEM | 0.095 | 0.035 | 0.044 | 0.146 | 0.238 | 0.558 | 0.003 | 8.9 | 0 |
| MGT | 0.090 | 0.032 | 0.039 | 0.138 | 0.232 | 0.531 | 0.029 | 10.3 | 0 |

As before, the parameters of the first component are estimated more accurately than those of the second one. This is shown in detail in table 4.9, where the cumulated rRMSEs of the parameters of the first and second component are displayed separately. All in all, similar results compared to the previously discussed mixture models are observed.

Table 4.9: Cumulated rRMSEs separately listed for the components of mixture model M6.

| Component | MIX | EMNR | APEM | RMEM | UDEM | MGT |
|---|---|---|---|---|---|---|
| 1th | 0.365 | 0.285 | 0.164 | 0.164 | 0.181 | 0.170 |
| 2nd | 0.591 | 0.567 | 0.252 | 0.252 | 0.282 | 0.271 |

**Summary of Mixture Model M6**

In conclusion, in mixture models like the one considered here, the MIX algorithm has to be examine with caution. This is due to a high convergence failure rate and predominantly large discrepancies to the true parameters that cause a high inferiority compared to the other methods. In contrast, the less complex APEM and RMEM methods once again provide the best results, followed by the MGT and UDEM algorithms.

### 4.3.3 Mixture Model M7

In this mixture model, the difference of the component means is three. Therefore, the density consists of two obvious peaks. The true values are

$$\Psi = (0.7,\ 0.3,\ 5,\ 8,\ 1,\ 1).$$

**Convergence Investigation**

Since the mixture components are now better separated, they can be decomposed more easily, which leads to less convergence failures. In fact, the MIX algorithm only failed in 1.5% of the samples. The EMNR algorithm did not provide a solution in 0.1% of the samples and the APEM and RMEM algorithms in 0.5%. The UDEM and MGT methods again offered a solution for each sample. These results can also be found in table 4.10.

Compared to the highly overlapped mixture models M5 and M6, these rates are smaller. However, in comparison to the balanced mixture model M3, more terminations due to convergence problems occur.

**Graphical Consideration**

The easier decomposition results in more accurate estimates as shown by the average densities in figure 4.8. All curves are close to each other and only the shapes of the peaks differ slightly. In fact, the higher a peak is, the more the standard deviation is underestimated, see table A.1.

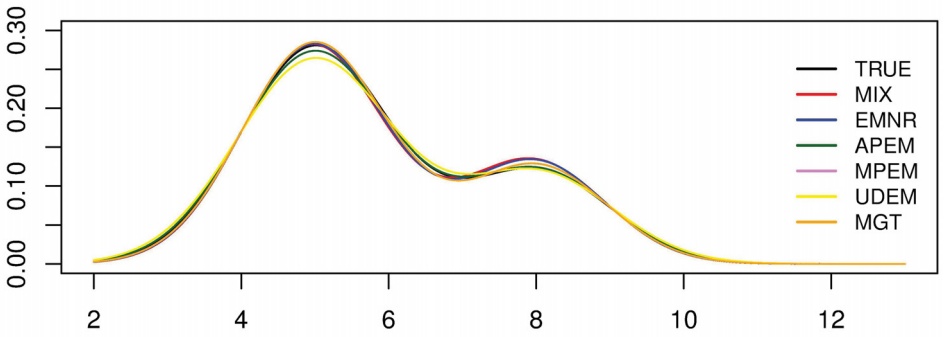



Figure 4.8: Mixture densities with average estimated parameters for mixture model M7.

**Numerical Comparison**

The results of this mixture model correspond to those of mixture model M3. The MIX and EMNR algorithms estimate on average the parameters of this well separated mixture much better than in the case where the components are highly overlapped. This becomes clearer when considering the cumulated rRMSE in table 4.10. Nevertheless, compared to the other four algorithms, the MIX and EMNR perform worse since their rRMSEs are higher in each parameter.

A direct comparison of the MIX and EMNR algorithms reveals a superiority of the MIX with respect to the runtime. However, considering the accuracy, results estimated by the EMNR algorithm are on average closer to the true values.

In contrast, a benefit of the better separation is not observable in the other methods. They are almost as accurate as before. Since there is no significant difference to the previous considered model, it can be deduced that the underlying mixture model has no large influence on the estimation results of these methods. In brief, for both highly overlapped and well separated mixture models, these methods provide similarly good results.

Table 4.10: The rRMSEs, cumulated rRMSEs, average runtimes in seconds, average number of iterations, and convergence failure rates in % for mixture model M7.

| Method | $\pi_1$ | $\mu_1$ | $\mu_2$ | $\sigma_1$ | $\sigma_2$ | Cum | Time | Iter | Failures |
|---|---|---|---|---|---|---|---|---|---|
| MIX | 0.177 | 0.053 | 0.066 | 0.185 | 0.305 | 0.786 | 0.366 | 14.3 | 1.5 |
| EMNR | 0.161 | 0.050 | 0.061 | 0.175 | 0.299 | 0.746 | 4.670 | 50.9 | 0.1 |
| APEM | 0.099 | 0.035 | 0.038 | 0.129 | 0.204 | 0.505 | 0.009 | 8.4 | 0.5 |
| RMEM | 0.099 | 0.035 | 0.038 | 0.129 | 0.204 | 0.505 | 0.004 | 8.4 | 0.5 |
| UDEM | 0.103 | 0.038 | 0.041 | 0.147 | 0.225 | 0.554 | 0.003 | 8.6 | 0 |
| MGT | 0.100 | 0.036 | 0.038 | 0.135 | 0.223 | 0.532 | 0.031 | 10.1 | 0 |

Comparing separately the rRMSEs of the mixture components, similar results to the previous mixture models M6 and M7 occur. All considered methods estimate the parameters of the first component more accurately than those of the second component, as shown in table 4.11.

Table 4.11: Cumulated rRMSEs separately listed for the components of mixture model M7.

| Component | MIX | EMNR | APEM | RMEM | UDEM | MGT |
|---|---|---|---|---|---|---|
| 1th | 0.238 | 0.225 | 0.164 | 0.164 | 0.185 | 0.171 |
| 2nd | 0.371 | 0.360 | 0.242 | 0.242 | 0.266 | 0.261 |

**Summary of Mixture Model M7**

To conclude, the APEM and RMEM algorithms are again the most reasonable methods. They are followed by the MGT and UDEM algorithms. In contrast, the MIX and EMNR algorithms provide worse estimations than the other four methods. As before, in comparison to the other four methods, they provide worse estimations.

### 4.3.4 Mixture Model M8

The last considered mixture model with $\pi_1 = 0.7$ is model M8, where the two components are well separated and true parameters are

$$\Psi = (0.7,\ 0.3,\ 5,\ 9,\ 1,\ 1).$$

**Convergence Investigation**

The convergence failure rates are similar to those of mixture model M4. The good separation of the components is accompanied with less convergence failures. Expect of the MIX algorithm, all methods provided a solution for each considered sample. The MIX algorithm failed in 0.3% of the samples.

**Graphical Consideration**

Figure 4.9 gives a first impression of the average estimation results by showing the densities obtained by the average estimated parameters. All shapes are similar to that one obtained by the true parameters. As before, only a difference in the peaks arises.

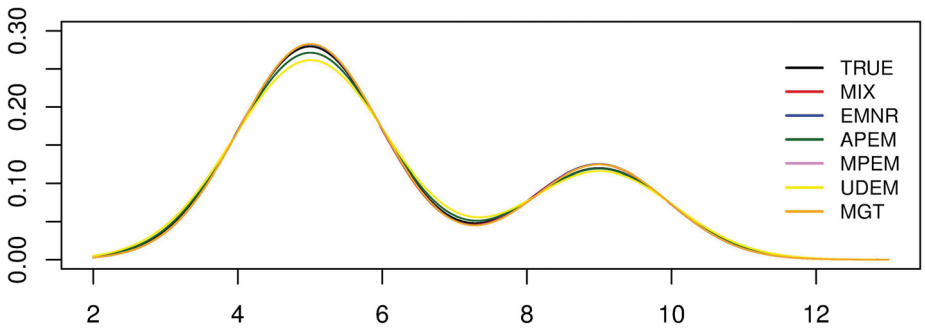



Figure 4.9: Mixture densities with average estimated parameters for mixture model M8.

**Numerical Comparison**

The results are comparable with those obtained by the corresponding, well separated mixture model M4. As before, the MIX and EMNR algorithms estimate the parameters of this well separated mixture much more accurate than of mixtures with highly overlapped components. A comparison of both algorithms against each other reveals that although the EMNR method needs almost five times more time per run, the estimates are closer to the true parameter, see table 4.9. Thus, as well in this mixture model, the EMNR algorithm is superior in the accuracy of the estimations.

The APEM algorithm provide again estimations that are more accurate than those obtained by MIX and EMNR. In particular, this method gives the most accurate parameters estimations. As before, the same is valid for the RMEM algorithm, which is also in this mixture model twice as fast as the APEM methods.

The MGT method performs almost as good as the APEM and RMEM algorithms. Only a slight difference in the component standard deviations is observable, particulary in the second component. Hence, the corresponding rRMSEs are worse than with the APEM and RMEM algorithms. However, this difference is small and all three methods are comparable at least in the parameter estimation. A comparison with respect to the runtime reveals an inferiority of MGT since this algorithm requires almost four times more time for a performance than APEM.

Table 4.12: The rRMSEs, cumulated rRMSEs, average runtimes in seconds, average number of iterations, and convergence failure rates in % for mixture model M8.

| Method | $\pi_1$ | $\mu_1$ | $\mu_2$ | $\sigma_1$ | $\sigma_2$ | Cum | Time | Iter | Failures |
|---|---|---|---|---|---|---|---|---|---|
| MIX | 0.087 | 0.034 | 0.034 | 0.134 | 0.234 | 0.523 | 0.329 | 11.5 | 0.3 |
| EMNR | 0.082 | 0.033 | 0.031 | 0.128 | 0.226 | 0.500 | 1.480 | 15.8 | 0 |
| APEM | 0.075 | 0.031 | 0.027 | 0.116 | 0.185 | 0.434 | 0.006 | 5.1 | 0 |
| RMEM | 0.075 | 0.031 | 0.027 | 0.116 | 0.185 | 0.434 | 0.003 | 5.1 | 0 |
| UDEM | 0.077 | 0.032 | 0.030 | 0.135 | 0.202 | 0.476 | 0.002 | 5.4 | 0 |
| MGT | 0.076 | 0.031 | 0.027 | 0.119 | 0.199 | 0.452 | 0.023 | 6.7 | 0 |

Similar to the other mixture models with unbalance components, the parameters of the first component are estimated more accurate than those of the second component. This effect is observable in each method, see table 4.13.

Table 4.13: Cumulated rRMSEs separately listed for the components of mixture model M8.

| Component | MIX | EMNR | APEM | RMEM | UDEM | MGT |
|---|---|---|---|---|---|---|
| 1th | 0.168 | 0.161 | 0.147 | 0.147 | 0.167 | 0.150 |
| 2nd | 0.268 | 0.257 | 0.212 | 0.212 | 0.232 | 0.226 |

**Summary of Mixture Model M8**

In conclusion, comparing these results to those obtained by mixture model M4, which has the same difference in the component means and differs only in the mixing proportions, no significant differences in the rRMSE are observable. Hence, it can be concluded that the unbalance of the components has not much influence on the estimation results if the components are sufficiently separated.

## 4.4 Comparison Results for $\pi_1 = 0.5$

This section addresses mixture models with the most unbalance components. This case is the most interesting and difficult one, because now 90% of the observations belong to the first and only 10% belong to the second component. Thus, it is difficult to identify the two components, especially when they are highly overlapped. Nevertheless, such data situations arise in several research fields like for instance in clinical trials, if rare genetic polymorphism occur as preliminary described. Hence, this parameter constellation is an important case and, therefore, considered here for different component separations in the mixture models M9 - M12.

### 4.4.1 Mixture Model M9

Given that the first mixing proportion is 0.9 and the difference between the component means is $\Delta = 1$, true values are

$$\Psi = (0.9,\ 0.1,\ 5,\ 6,\ 1,\ 1).$$

**Convergence Investigation**
This mixture model describes a distribution with highly overlapped components likewise the mixture models M1 and M5. As we have seen there, a decomposition of components of such highly overlapped mixtures is hard to perform and if additionally the component weights are highly unbalanced, it is even more difficult to estimate the parameters of such mixture. In example, this is reflected in higher convergence failure rates. The MIX algorithm did not provide a solution in 23.8% of the samples. The EMNR algorithm failed in 11.8% and the APEM and RMEM algorithms in 13.6% of the samples. Both, the UDEM and MGT algorithms had again no convergence problems and a solution is obtained in each considered sample. These results are summarized in table 4.14.

Compared to the failure rates in mixture models M1 and M5, which differ only in the choice of the mixing proportions, higher failure rates occur with all algorithms expect the UDEM and MGT algorithms. Especially, the APEM and, so the RMEM algorithms have a more than five times higher convergence failure rate compared to the results obtained by mixture model M5. However, the remaining samples provide on average appropriate estimates as shown next.

**Graphical Consideration**
The densities obtained by the estimated parameters are shown in figure 4.10. The underlying parameter estimations and deviations are presented in table A.1 in the appendix.

With the MIX algorithm almost one quarter of the samples has to be excluded due to convergence problems. Even though, the average parameter estimations obtained by this algorithm are distinct worse than those of the other methods. In fact, the MIX algorithm

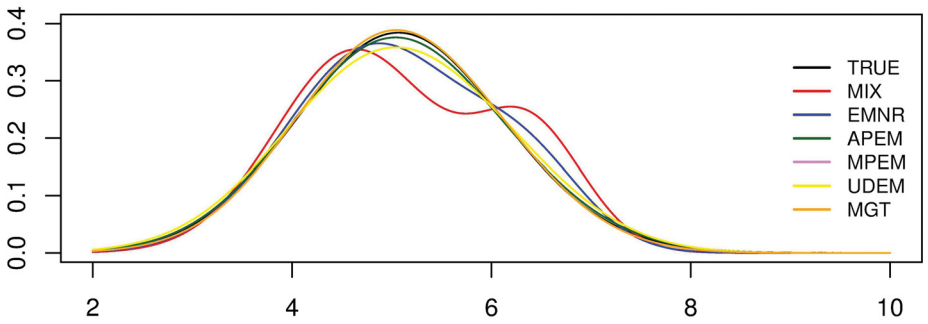



Figure 4.10: Mixture densities with average estimated parameters for mixture model M9.

underestimates almost all parameters, which causes the large difference between this curve and the reference one, see figure 4.10.

Furthermore, the curve of the EMNR algorithm indicates that this method provides parameter estimations, which are closer to the true values compared to the MIX algorithm. Comparing the UDEM curve with the reference one, a distinct smoother shape is observable. This implies an overestimation of at least the first standard deviation. Finally, the average results of the APEM, RMEM, and MGT algorithms are, as before, close to the true parameters.

**Numerical Comparison**

The weak decomposition results of the MIX algorithm are also reflected in the cumulated rRMSE as shown in table 4.14. With a value of 1.69, this method has a large uncertainty in the decomposition of the components compared to the other algorithms. Hence, for highly unbalanced mixture models this algorithm has to be applied with caution, when the components are highly overlapped.

Table 4.14: The rRMSEs, cumulated rRMSEs, average runtimes in seconds, average number of iterations, and convergence failure rates in % for mixture model M9.

| Method | $\pi_1$ | $\mu_1$ | $\mu_2$ | $\sigma_1$ | $\sigma_2$ | Cum | Time | Iter | Failures |
|---|---|---|---|---|---|---|---|---|---|
| MIX | 0.403 | 0.138 | 0.183 | 0.351 | 0.610 | 1.685 | 0.876 | 38.2 | 23.8 |
| EMNR | 0.230 | 0.083 | 0.149 | 0.216 | 0.584 | 1.262 | 19.519 | 132.9 | 11.8 |
| APEM | 0.069 | 0.032 | 0.068 | 0.107 | 0.320 | 0.596 | 0.012 | 10.4 | 13.6 |
| RMEM | 0.069 | 0.032 | 0.068 | 0.107 | 0.320 | 0.596 | 0.004 | 10.5 | 13.6 |
| UDEM | 0.085 | 0.035 | 0.080 | 0.116 | 0.389 | 0.705 | 0.004 | 13 | 0 |
| MGT | 0.071 | 0.031 | 0.067 | 0.120 | 0.382 | 0.671 | 0.035 | 13.8 | 0 |

Having a look at the cumulated rRMSE of the EMNR algorithm reinforce the superiority to MIX. Especially the estimation of the mixing proportion $\pi_1$ is closer to the true value using the EMNR algorithm instead of MIX. Indeed, the EMNR algorithm is better in terms of accuracy compared to the MIX algorithm, however the runtime is inferior. In particular, compared to the other algorithms, the EMNR algorithm needs much time for a performance. The same is true for the average number of iterations. This value is about 13 times higher than the one of the best algorithm, which is APEM. Comparing the cumulated rRMSE of the EMNR algorithm with the other methods an twice time higher value can be observed. This underline the inferiority to methods with only one iteration procedure.

The comparison of these results to those obtained by the mixture model M1 or M5, reveals that expect of the MIX and EMNR algorithms the difference between the true and the estimated parameters is larger. This is due to the fact that the estimation of the parameters of the second component is based on a small number of observation. This naturally causes a higher rRMSE in the second component, as can be seen in table 4.15, where the cumulated rRMSEs of the components is separated listed for each method.

Table 4.15: Cumulated rRMSEs separately listed for the components of mixture model M9.

| Component | MIX | EMNR | APEM | RMEM | UDEM | MGT |
|---|---|---|---|---|---|---|
| 1th | 0.489 | 0.299 | 0.139 | 0.139 | 0.151 | 0.151 |
| 2nd | 0.793 | 0.733 | 0.388 | 0.388 | 0.469 | 0.449 |

**Summary of Mixture Model M9**

Summarized, it is difficult to decompose mixtures with highly overlapped and highly unbalanced components. Especially when the sample size is small like in our case, only a few observations are available for the estimation of the second component. As will be shown in section 4.6, an increase of the sample size causes a strong improvement of the estimation results. However, with samples of about 100 observations, the best results were achieved using the APEM and RMEM algorithms. With both algorithms estimates are obtained which provide the best fit. The worst results are obtained by the MIX and EMNR algorithms.

### 4.4.2 Mixture Model M10

This mixture model is the counterpart of M2 and M6 with a difference between the component means of two. The true values are

$$\Psi = (0.9,\ 0.1,\ 5,\ 7,\ 1,\ 1).$$

**Convergence Investigation**

In general, this mixture model causes less convergence problems in comparison to mixture model M9. Only the MIX algorithm failed in more samples compared to model M9 (24.1%). The EMNR algorithm did not provide a solution in 5.6% of the samples and the APEM and RMEM algorithms in 6.7%. As before, a solution for each sample was obtained by the UDEM and MGT algorithms, which is displayed in table 4.16.

**Graphical Consideration**

The inferiority of the MIX and EMNR algorithms can not only be notified in the convergence failure rates, but also by the average parameter estimations. These are displayed in figure 4.11 as average densities.

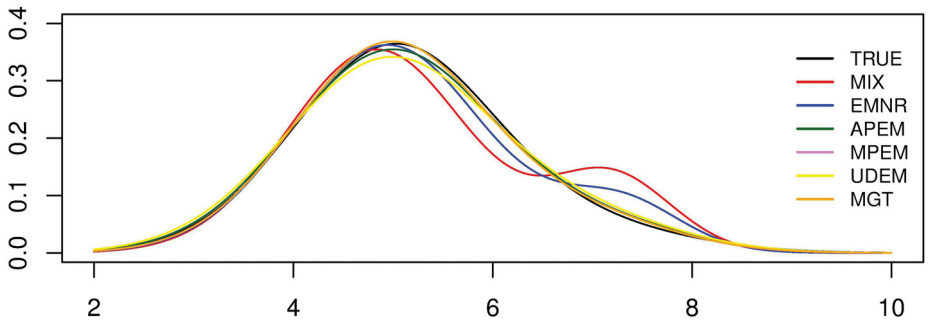



Figure 4.11: Mixture densities with average estimated parameters for mixture model M10.

The curves obtained by the MIX and EMNR algorithms can clearly be distinguished from the curve representing the true parameters. This can be explained by the underestimation of the mixing proportions, as shown in table A.1. Furthermore, all densities differ more from the reference curve than in mixture models with more balanced components.

**Numerical Comparison**

In general, all methods underestimate the second variance, which is the main reason for the difference of the curves obtained by the methods and the reference one in figure 4.11. Moreover, a large deviation in this parameter leads to large rRMSEs, see table 4.16. This uncertainty in the parameter estimation can be explained by the unbalance of the components. The second component is estimated on the basis of only 10 observations. Thus, all methods have difficulties to separate the second component, which can be seen in table 4.17, where the cumulated rRMSE of each component is separately displayed.

However, some methods provide better results than others. For instance, the APEM algorithm provide the smallest cumulated rRMSE, which indicates that this algorithm on average lead to the best estimates. Moreover, this method is fast, requires only a few iterations and can only be surpassed by the RMEM algorithm, which computes the same estimates, but is at the same time twice as fast.

Some methods provide less accurate estimation results. In fact, the MIX and EMNR algorithms have cumulated rRMSEs that are almost two times higher than those of the best method, which is the APEM algorithm. Moreover, both methods require a higher number of iterations, more time for a run, and failed in a distinct number of samples. Thus, they are distinctly inferior compared to the other methods.

Table 4.16: The rRMSEs, cumulated rRMSEs, average runtimes in seconds, average number of iterations, and convergence failure rates in % for mixture model M10.

| Method | $\pi_1$ | $\mu_1$ | $\mu_2$ | $\sigma_1$ | $\sigma_2$ | Cum | Time | Iter | Failures |
|---|---|---|---|---|---|---|---|---|---|
| MIX | 0.293 | 0.084 | 0.155 | 0.245 | 0.603 | 1.380 | 0.856 | 34.7 | 24.1 |
| EMNR | 0.172 | 0.054 | 0.139 | 0.162 | 0.595 | 1.122 | 13.818 | 100 | 5.6 |
| APEM | 0.053 | 0.028 | 0.055 | 0.097 | 0.325 | 0.558 | 0.010 | 7.9 | 6.7 |
| RMEM | 0.053 | 0.028 | 0.055 | 0.097 | 0.325 | 0.558 | 0.004 | 7.9 | 6.7 |
| UDEM | 0.065 | 0.032 | 0.061 | 0.118 | 0.363 | 0.639 | 0.003 | 8.8 | 0 |
| MGT | 0.058 | 0.029 | 0.060 | 0.110 | 0.377 | 0.634 | 0.029 | 10.6 | 0 |

Table 4.17: Cumulated rRMSEs separately listed for the components of mixture model M10.

| Component | MIX | EMNR | APEM | RMEM | UDEM | MGT |
|---|---|---|---|---|---|---|
| 1th | 0.329 | 0.216 | 0.125 | 0.125 | 0.150 | 0.139 |
| 2nd | 0.758 | 0.734 | 0.380 | 0.380 | 0.424 | 0.437 |

**Summary of Mixture Model M10**

In conclusion, the simple algorithms again dominate for this mixture model. Particulary, the APEM and RMEM algorithms provide the most reliable results. The MIX and EMNR algorithms are distinctly inferior.

### 4.4.3 Mixture Model M11

This mixture model consists of two components that can clearly be identified. The underlying true parameters are

$$\Psi = (0.9,\ 0.1,\ 5,\ 8,\ 1,\ 1).$$

**Convergence Investigation**

The convergence failure rates of this mixture model are lower compared to those obtained by the highly overlapped models M9 and M10. Nevertheless, they are higher compared to the more balanced models M3 or M7. In example, the MIX algorithm failed in 17.3% of the samples, while the EMNR algorithm did not find a solution in only 0.2% of the samples. The APEM and RMEM algorithms did not provide a solution in 4.8% and the UDEM and MGT algorithms had again no convergence problems. These results are summarized in table 4.18.

**Graphical Consideration**

For a first impression of the average estimation results, the average densities are displayed in figure 4.12.

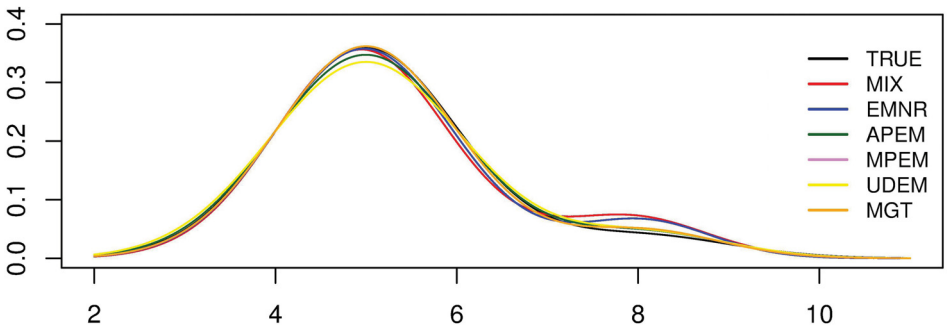



Figure 4.12: Mixture densities with average estimated parameters for mixture model M11.

All methods provide parameter estimations that are closer to the true values than in the previous mixture model M10. This can be seen at the densities, which are closer to the reference one. However, the MIX and EMNR algorithms have more difficulties in the estimation of the parameters of the second component than the other algorithms. Both curves have a higher second peak, which indicates an underestimation of the second standard deviation.

**Numerical Comparison**

Considering the cumulated rRMSEs, displayed in table 4.18, the inferiority of the MIX and EMNR algorithms becomes more clear. With a cumulated rRMSE of 0.98 and 0.92, respectively, these methods are worse than the others. In particular, a high uncertainty about the mixing proportions is noticeable. The rRMSE of $\pi_1$ is more than two times higher than for instance those obtained by the APEM algorithm.

Table 4.18: The rRMSEs, cumulated rRMSEs, average runtimes in seconds, average number of iterations, and convergence failure rates in % for mixture model M11.

| Method | $\pi_1$ | $\mu_1$ | $\mu_2$ | $\sigma_1$ | $\sigma_2$ | Cum | Time | Iter | Failures |
|---|---|---|---|---|---|---|---|---|---|
| MIX | 0.158 | 0.044 | 0.115 | 0.165 | 0.497 | 0.979 | 0.682 | 27.0 | 17.3 |
| EMNR | 0.121 | 0.038 | 0.101 | 0.138 | 0.517 | 0.915 | 7.141 | 54.8 | 0.2 |
| APEM | 0.059 | 0.028 | 0.061 | 0.107 | 0.326 | 0.581 | 0.009 | 7.8 | 4.8 |
| RMEM | 0.059 | 0.028 | 0.061 | 0.107 | 0.326 | 0.581 | 0.003 | 7.8 | 4.8 |
| UDEM | 0.061 | 0.029 | 0.068 | 0.128 | 0.385 | 0.671 | 0.003 | 8.5 | 0 |
| MGT | 0.059 | 0.028 | 0.061 | 0.108 | 0.381 | 0.637 | 0.030 | 9.9 | 0 |

As before, considering the cumulated rRMSE for each component separately, it is revealed that the parameters of the first component are estimated better than those of the second. The corresponding results are shown in table 4.19.

Table 4.19: Cumulated rRMSEs separately listed for the components of mixture model M11.

| Component | MIX | EMNR | APEM | RMEM | UDEM | MGT |
|---|---|---|---|---|---|---|
| 1th | 0.209 | 0.176 | 0.135 | 0.135 | 0.157 | 0.136 |
| 2nd | 0.612 | 0.618 | 0.387 | 0.387 | 0.453 | 0.442 |

Furthermore, comparing the average runtime of the EMNR algorithm for this mixture model (7.1 seconds) to those required for the mixture models M3 or M8, which are 3.9 and 4.7 seconds, it can be seen that this model needs on average distinctly more time for a performance. The same is valid for the MIX algorithm. In this mixture model, the MIX algorithm needs 0.7 seconds, while for the other two mixture models 0.3 and 0.4 seconds, respectively, are necessary.

In contrast, there is no significant difference in the runtime observable for the other methods. Independently of the underlying mixture model, the UDEM algorithm does not need more than 0.004 seconds for one run. On average, the RMEM algorithm is slightly slower. However, irrespective of the mixture model, this method never needs more than 0.004 seconds. The APEM method is slightly slower and requires the most time for model M9, which is the most unbalanced and highest overlapped model. Nevertheless, the APEM algorithm never exceeded 0.012 seconds, which is an exceptinal improvement to the MIX and EMNR algorithms.

**Summary of Mixture Model M11**

In brief, the APEM and RMEM algorithms provide the best results, followed by the MGT and UDEM algorithms. The results with the highest difference to the true values are offered by the MIX and EMNR algorithms.

### 4.4.4 Mixture Model M12

The last mixture model has a difference between the component means of four. The true parameter are

$$\Psi = (0.9, 0.1, 5, 9, 1, 1).$$

**Convergence Investigation**

Although the two components are well separated, some methods failed in the right decomposition in a higher number of samples compared to the well separated mixture models M4 or M8. The MIX algorithm did not provide a solution in 7.6% of the samples. The APEM and RMEM algorithms failed in 1.0% of the samples. In contrast, the EMNR, UDEM and MGT algorithms had no convergence problems. Hence, it seems that the balance of the components has an influence on the estimation results, at least in some methods. However, if the samples without convergence are excluded, satisfying results are obtained.

**Graphical Consideration**

Figure 4.13 shows the average densities obtained by the six methods and the true parameters.

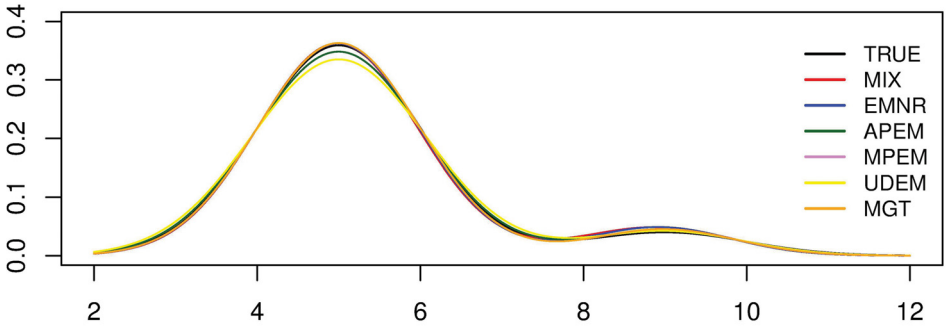



Figure 4.13: Mixture densities with average estimated parameters for mixture model M12.

All method provide on average densities which are close to the density with the true parameters. Similar to mixture model M4 and M8, the only difference that occurs is observable in the two peaks. Especially the curves obtained by the APEM and UDEM algorithms results in a lower peak, which indicates on average an overestimation of the first standard deviation, which is shown in table A.1.

**Numerical Comparison**

All findings from the previous mixture models M11 are valid in this mixture model, too. The MIX and EMNR algorithms once again provide the highest rRMSEs, which indicates the largest discrepancies to the true values. Thereby, as before, the MIX algorithm is much faster, but this is at the expense of the accuracy.

Moreover, the APEM method is superior to the other methods, particularly to the MIX and EMNR algorithms. The UDEM algorithm is the fastest method, but the estimation results are not as good as those obtained by the RMEM algorithm. Finally, the results obtained by the MGT algorithm are slightly worse than those of the APEM algorithm.

Table 4.20: The rRMSEs, cumulated rRMSEs, average runtimes in seconds, average number of iterations, and convergence failure rates in % for mixture model M12.

| Method | $\pi_1$ | $\mu_1$ | $\mu_2$ | $\sigma_1$ | $\sigma_2$ | Cum | Time | Iter | Failures |
|---|---|---|---|---|---|---|---|---|---|
| MIX | 0.066 | 0.027 | 0.079 | 0.112 | 0.431 | 0.715 | 0.591 | 22.7 | 7.6 |
| EMNR | 0.057 | 0.026 | 0.070 | 0.106 | 0.443 | 0.702 | 3.337 | 22.7 | 0 |
| APEM | 0.040 | 0.024 | 0.051 | 0.095 | 0.313 | 0.523 | 0.007 | 5.4 | 1 |
| RMEM | 0.040 | 0.024 | 0.051 | 0.095 | 0.313 | 0.523 | 0.003 | 5.4 | 1 |
| UDEM | 0.044 | 0.026 | 0.056 | 0.119 | 0.338 | 0.583 | 0.002 | 5.7 | 0 |
| MGT | 0.040 | 0.024 | 0.051 | 0.096 | 0.353 | 0.564 | 0.025 | 7.3 | 0 |

It can clearly be seen from table 4.21 that, as before, the estimations of the parameters of the first component are calculated more accurately with less deviations than those of the second component. Compared to the models with more balanced components, the results for the first component are even better. A similar behavior can be revealed, having a look at the mixing proportions. Throughout all methods, the rRMSE of $\pi_1$ is much smaller than in the mixture models with more balanced components. A possible explanation is that more observations are available for the estimation of the first component.

Table 4.21: Cumulated rRMSEs separately listed for the components of mixture model M12.

| Component | MIX | EMNR | APEM | RMEM | UDEM | MGT |
|---|---|---|---|---|---|---|
| 1th | 0.139 | 0.132 | 0.119 | 0.119 | 0.145 | 0.120 |
| 2nd | 0.510 | 0.513 | 0.364 | 0.364 | 0.394 | 0.404 |

**Summary of Mixture Model M12**

The APEM algorithm, once again, provide the best estimation results. The same results are achieved by the RMEM algorithm, which is even faster. Similar estimation are obtained by the MGT and UDEM algorithm. The MIX and EMNR algorithm are much better in this model than in mixtures with highly overlapped components, however, they still remain inferior.

In the next section, the results for all twelve mixture models are summarized.

## 4.5 Summary of the Results

The previous sections dealt with the estimation of the parameters of the mixture models M1 - M12. Throughout these investigations, the sample size $n$ was chosen to be 100, the interval width $h$ was 1 and the true values were taken as starting values. The latter choice does not reflect the situation in practical applications since the true values are seldom known. However, this assumption illustrates the power of each method starting with values close to the true ones. Moreover, a variation of the initial values is investigated in section 4.7.

First, four mixture models with equal underlying mixing proportions were considered (M1 - M4). This is the easiest case, since half of the observations belong to the first component, and the other half to the second. Methods containing simplifications, like APEM, RMEM and UDEM, had approved themselves for cases with highly overlapped components, where it is not obvious that the distribution consists of two subpopulations, and well separated components, where two peaks can clearly be identified. They were fast and, at the same time, they only failed in a small number of samples. Furthermore, they provided estimations that were close to the true values. Hence all of them can be recommended.

An interesting finding concerns the UDEM algorithm. This method assumes the observations to lie somewhere between the interval borders and, indeed, it seems to be irrelevant where exactly the observations are located, as long as the number of observations of each interval is specified. This effect can be observed even more clearly in the RMEM method, where only the number of observations of each interval is of interest irrespective of their exact location. This method has the simplest construction and, therefore, the easiest implementation. It is fast and provides, additionally, the best fit for most of the mixture models. Therefore, this method is to be preferred.

The more complex algorithms, MIX and EMNR, which contain two and tree iterations procedures, respectively, seem to be less suitable. In particular, for highly overlapped mixture models, they provided less accurate fits and with respect to the convergence properties, both methods failed in distinctively higher number of samples. However, the results of these two methods were strongly dependent on the separation of the components. The more the components were separated the better were the estimations. Furthermore, the same is valid for the runtime. The larger the distance between the component means was, the faster the results were achieved. Finally, the separation also had an influence on the convergence, since in well separated mixture distributions less convergence problems occurred. In brief, the MIX and EMNR algorithms can not be recommended for highly overlapped mixture models. However, for well separated mixtures, they provided estimates that were close to the true values. Compared to the simple methods, the MIX and EMNR algorithms are inferior.

Comparing the MIX and EMNR algorithms directly leads to the conclusion that the EMNR algorithm provides estimates that are closer to the true values. The improvement of the

runtime of the MIX algorithm caused less accuracy in the estimates, which is a surprising result, since the aim of the combination of the EM and NR algorithms was to improve the method. In fact, an improvement of the runtime was achieved, but at the same time a deterioration of the accuracy occurred.

In contrast, considering the APEM algorithm, which is a simplification of the EMNR and, hence, the MIX method, one can see that the estimates of the parameters obtained by this method were much closer to the true values than those obtained by the MIX or EMNR algorithm. This finding led us to the conclusion that the approximation of the probability of an individual observation to fall into an interval causes a smaller deviation than an additional iteration.

The largest advantage of the MGT algorithm is the reliability of this method in terms of the convergence properties. This method provided a solution for each considered mixture model, independent of the degree of overlapping or balance. Besides its excellent convergence properties, this method as well estimated parameters, which were close to the true values. Hence, this method can be recommended, too.

Furthermore, mixture models M5 - M8 were investigated, where the first mixing proportion is chosen to be 0.7. With such unbalanced mixing proportions, the decomposition of a mixture distribution into its components is more difficult compared to models with well balanced mixing weights. In fact, higher convergence failure rates, worse estimations and larger deviations occurred in almost all applied methods. This effect increases the more unbalanced the mixing weights were, which was shown for the mixture models M9 - M12, where $\pi_1 = 0.9$ and $\pi_2 = 0.1$. Though, it has to be kept in mind, that all these estimates resulted on sample of size $n = 100$. Accordingly, a second mixing proportion of 0.1 lead to 10 observations in the second component. From those the second means and standard deviations had to be calculated. Due to this fact, it is natural that the second standard deviation was underestimated by all methods, independent of the component separation.

To conclude, the results show that grouped observations do not automatically cause a less capable fit. Depending on the underlying mixture model and the choice of an appropriate method, an adequate fitting can be achieved. Nevertheless, it should be noted that these results are based on samples that lead to a solution within 500 iterations. Samples with a higher number of iterations and those where the algorithms converge to a singularity or divergence, are excluded. This might be the reason for the good results comparing those computed by other authors, e.g. Hosmer (1973a). He investigated the accuracy of maximum likelihood estimates in two component mixtures with highly overlapped components and small samples sizes. His simulation study reinforces the need for well separated components with $\Delta \geq 3$ and large sample sizes $n > 300$. For the other cases, he stated, that“[...] the ML estimates may not to be accurate enough to provide useful estimates”

(p. 223). To investigate the improvement of higher number of observations, the following section focuses on this topic.

## 4.6 Variation of the Sample Size

The previously presented simulations were based on $n = 100$ observations. Naturally, there is a large uncertainty about the parameter estimation in such sample dimensions, especially when the components are highly unbalanced and only a few observations from the second component are incorporated in the estimation. Nevertheless, the presented results showed that on average, even with a small amount of observations, useful estimates can be obtained for most mixture models. Though, to achieve better results, larger samples are necessary. As mentioned before, Hosmer (1973b) refers to his experience and recommended sample sizes of at least 300. This sample dimension is investigated in the following study, followed by a simulation with 500 observations.

### 4.6.1 Results for Sample Size n=300

Starting with the consideration of the convergence failure rates, which are displayed in table 4.22, it can be seen that for well balanced components (mixture model M1 - M4), almost all methods provided a solution for each sample. The only convergence problems occurred with the MIX and EMNR algorithms in highly overlapped mixtures. Hence, even with larger sample sizes both methods could not decompose the mixture of each sample.

Likewise, failure rates are obtained for model M5 - M8, where the mixing proportion is $\pi_1 = 0.7$. Again, the MIX and EMNR algorithms were the only algorithms that could not provide a solution for each sample. For more unbalanced mixtures, like models M9 - M12, where $\pi_1 = 0.9$, more convergence problems arose in the other methods. However, compared to the failure rates in the previous section, where the sample size was $n = 100$, less iteration terminations occurred. This is natural, because more information about each component are available.

Table 4.22: Convergence failure rates in % for $n = 300$.

| Method | M1 | M2 | M3 | M4 | M5 | M6 | M7 | M8 | M9 | M10 | M11 | M12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MIX | 16.1 | 3.5 | 0 | 0 | 16.7 | 3.7 | 0 | 0 | 17.5 | 14.9 | 2.7 | 0.5 |
| EMNR | 21.6 | 7.3 | 0 | 0 | 19.8 | 6.0 | 0 | 0 | 14.9 | 6.4 | 0.4 | 0 |
| APEM | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3.6 | 0.1 | 0.3 | 0 |
| RMEM | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3.6 | 0.1 | 0.3 | 0 |
| UDEM | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| MGT | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Turning to the accuracy and the precision of the methods in terms of the cumulated rRMSE, which is presented in table 4.23, the first that can be identified is that the values are smaller than in the previously considered case. This indicates that for all mixture models all methods provide solutions that are closer to the true values and/or the deviations of the estimates are smaller.

Thereby, as before, the cumulated rRMSE of the MIX and EMNR algorithms decrease the better separated the components are. In particular, mixtures with well separated components are decomposed by MIX and EMNR almost as good as by the other methods. In fact, compared to the UDEM algorithm, these methods are even slightly better.

Table 4.23: Cumulated rRMSEs for sample size $n = 300$.

| Method | M1 | M2 | M3 | M4 | M5 | M6 | M7 | M8 | M9 | M10 | M11 | M12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MIX | 1.76 | 1.16 | 0.47 | 0.28 | 1.54 | 1.01 | 0.46 | 0.28 | 1.44 | 1.08 | 0.61 | 0.40 |
| EMNR | 0.79 | 0.89 | 0.45 | 0.27 | 0.76 | 0.79 | 0.43 | 0.27 | 0.81 | 0.77 | 0.57 | 0.39 |
| APEM | 0.24 | 0.25 | 0.28 | 0.25 | 0.25 | 0.24 | 0.26 | 0.24 | 0.32 | 0.29 | 0.32 | 0.30 |
| RMEM | 0.24 | 0.25 | 0.28 | 0.25 | 0.25 | 0.24 | 0.26 | 0.24 | 0.32 | 0.29 | 0.32 | 0.30 |
| UDEM | 0.27 | 0.28 | 0.31 | 0.31 | 0.27 | 0.28 | 0.30 | 0.30 | 0.36 | 0.33 | 0.36 | 0.34 |
| MGT | 0.24 | 0.25 | 0.27 | 0.24 | 0.26 | 0.24 | 0.25 | 0.24 | 0.34 | 0.30 | 0.33 | 0.31 |

In contrast, the improvement due to better separation of the components can not be observed within the other methods. Indeed, the estimations obtained from mixture models with highly unbalanced components are slightly worse. However, this effect is not as much as by MIX and EMNR. For example, the cumulated rRMSEs of the APEM and RMEM algorithms are within the range 0.24 and 0.32. The values of the MGT algorithm are never higher than 0.34 and those of the UDEM algorithm are between 0.27 and 0.36. Hence, it can be deduced that these methods provide good estimation results irrespective of the underlying mixture model.

In summary, larger sample sizes cause better estimation results. However, the results of the MIX and EMNR algorithms could not be improved much for highly overlapped mixtures by increasing the sample size to 300. Consequently it is not recommended to use these methods in such cases. In contrast, the APEM, RMEM, UDEM, and MGT algorithms can be recommended without restrictions for mixture models with highly overlapped and highly unbalanced components.

### 4.6.2 Results for Sample Size n=500

In this simulation study, each sample consists of 500 observations. First, the convergence failure rate of each method is considered for each model. The results are summarized in table 4.24. It can be deduced that the increase of the sample size does not improve the

failure rate of all methods. In fact, the convergence failure rate of the EMNR algorithm is even higher in some mixture models compared to the results obtained by 300 observations. Especially for the mixture models M5 and M6 an increase is noticeable.

However, this is the only algorithm that causes more iteration terminations with higher sample sizes. As before, the UDEM and MGT algorithms had no convergence problems at all. The APEM and RMEM algorithms provided solutions for each sample except for mixture models M9 and M10. Even the MIX method benefits from the increase of the sample size and provided convergence failure rates that are smaller than before.

Table 4.24: Convergence failure rates in % for $n = 500$.

| Method | M1 | M2 | M3 | M4 | M5 | M6 | M7 | M8 | M9 | M10 | M11 | M12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MIX | 13.5 | 1.9 | 0 | 0 | 13.8 | 10 | 0.1 | 0 | 14.8 | 7.9 | 0.8 | 0.1 |
| EMNR | 21.0 | 11.5 | 0 | 0 | 23.6 | 8.4 | 0 | 0 | 13.3 | 5.2 | 0.1 | 0 |
| APEM | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.7 | 0.1 | 0 | 0 |
| RMEM | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.7 | 0.1 | 0 | 0 |
| UDEM | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| MGT | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Moving on to the cumulated rRMSE, displayed in table 4.25, an improvement in the accuracy and/or the deviation of the estimates can be seen.

Table 4.25: Cumulated rRMSEs for sample size $n = 500$.

| Method | M1 | M2 | M3 | M4 | M5 | M6 | M7 | M8 | M9 | M10 | M11 | M12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MIX | 1.67 | 0.96 | 0.36 | 0.21 | 1.51 | 0.82 | 0.36 | 0.21 | 1.37 | 0.90 | 0.50 | 0.30 |
| EMNR | 0.55 | 0.71 | 0.34 | 0.21 | 0.53 | 0.62 | 0.34 | 0.21 | 0.66 | 0.65 | 0.45 | 0.30 |
| APEM | 0.17 | 0.18 | 0.21 | 0.20 | 0.17 | 0.18 | 0.20 | 0.19 | 0.22 | 0.23 | 0.24 | 0.24 |
| RMEM | 0.17 | 0.18 | 0.21 | 0.20 | 0.17 | 0.18 | 0.20 | 0.19 | 0.22 | 0.23 | 0.24 | 0.24 |
| UDEM | 0.20 | 0.22 | 0.26 | 0.26 | 0.21 | 0.22 | 0.25 | 0.26 | 0.26 | 0.26 | 0.30 | 0.28 |
| MGT | 0.16 | 0.18 | 0.19 | 0.18 | 0.16 | 0.17 | 0.18 | 0.18 | 0.22 | 0.23 | 0.23 | 0.23 |

However, the MIX and ENMR algorithms remain distinctively inferior with respect to the estimation of the parameters of mixtures with highly overlapped components. Thereby, the estimates obtained by the EMNR algorithm are more accurate than those obtained by MIX. Though, this is only valid for highly overlapped components. For well separated mixture models like M4, M8, and M12, estimates are achieved, which are almost as good as those obtained by the other methods. In fact, the cumulated rRMSE of the UDEM algorithm is even higher.

Furthermore, an interesting finding is that the MGT method is in this sample dimension at least as good as the APEM algorithm. Moreover, in most mixture models the MGT method provides even better estimation results. Since this method is free of convergence problems and in most cases provides the best results, a recommendation for this method is given.

In addition, if one prefers an easier algorithm, the APEM or the RMEM algorithm is a good choice since both algorithms provide reasonable estimations. In contrast, even for larger sample sizes, methods with more than one iteration procedure are reliable only for mixture models with well separated components.

In conclusion, an increase of the sample size may improve the estimation results. With a higher number of observations, it is more likely that a solution will be found, particularly with the APEM and RMEM algorithms. Finally, the highest benefit arises in the more complex methods. They provide better results the more information is given.

## 4.7 Variation of the Starting Values

As previously mentioned, the starting values have an important influence on the convergence of the likelihood sequence and, hence, on the estimation results. To study the behavior of the estimations and to examine the convergence failures due to inappropriate starting values, this section turns to a variation of the initial values. In the previous simulation studies, the true values were taken as initial values to demonstrate the power of the algorithms. But now, it is assumed that the true values are not known and different starting values are considered.

In section 3.5, five techniques for obtaining starting values were proposed and will now be compared with each other. Recall that the modification of the technique of Finch et al. (1989) to obtain starting values is denoted by S1, the modified technique based on Böhning et al. (1994) by S2, the Range strategy mentioned by Karlis and Xekalaki (2003) by S3, and the adapted technique proposed by Seidel et al. (2000) by S4. Finally, the new technique is denoted by S5. For comparison reasons, the results obtained by the true values are also shown and denoted by S0.

**Convergence Failures**

Starting with the consideration of the convergence failure rates, the results obtained by the different initial values are summarized in table 4.26. Surprisingly, for some mixture models some initial values lead to smaller failure rates than those obtained by applying the true values as initial values. Considering for example mixture model M5, it can be seen that taking the true parameters as initial values the APEM algorithm failed in 21 of 1000 samples. In contrast, using the starting values obtained by technique S5, the APEM algorithm did

not provide a solution for only 8 samples, which is a distinct improvement. In particular, all techniques provided starting values with which the APEM algorithm achieved smaller convergence failure rates than with the true values. In fact, in some models the true values caused an up to 17 times higher failure rate compared to the results obtained by technique S5. This effect can be seen in most of the other methods, too.

The reason for this behavior might be, that if a particular sample is unfavorably generated, the true values might be poor as initial values. In other words, the data might be an untypical realization of the underlying distribution and the true values might be inappropriate for this sample to achieve any convergence. Since the proposed techniques generate individual initial values for each sample, they are in general more suitable and lead to a higher number of successfully decomposed samples.

However, it needs to be noted that a higher number of successfully decomposed samples not necessarily lead to better results. Having in mind that the estimation results depend on the convergence to a local or global maximum, it might be that for an unfavorable sample the generated starting values lead to convergence to a local maximum instead of the global one. Indeed, the estimation results are worse than those obtained by the true values, although nearly all starting value techniques results in less convergence failures. This will be discussed in the second part of the section. Though, this problem will not arise in practice since the true values are generally not known and the generation of initial values is essential.

Furthermore, as has been shown in section 4.6.2, if the sample size is higher, less convergence problems occur. In fact, convergence problems decrease with increasing sample size independent of the choice of the starting values. Hence, all methods are less sensitive to the choice of initial values if the sample size is sufficiently high.

Table 4.26: Number of failures in the same 1000 samples for starting values S0 - S5.

| Model | Method | S0 | S1 | S2 | S3 | S4 | S5 |
|---|---|---|---|---|---|---|---|
| M1 | MIX | 211 | 208 | 212 | 213 | 205 | 228 |
| | EMNR | 80 | 73 | 70 | 74 | 80 | 72 |
| | APEM/RMEM | 11 | 8 | 8 | 7 | 11 | 8 |
| | UDEM | 0 | 0 | 0 | 0 | 0 | 0 |
| | MGT | 0 | 0 | 0 | 0 | 0 | 0 |
| M2 | MIX | 111 | 126 | 117 | 113 | 129 | 113 |
| | EMNR | 21 | 16 | 25 | 23 | 23 | 18 |
| | APEM/RMEM | 6 | 6 | 5 | 5 | 5 | 5 |
| | UDEM | 0 | 0 | 0 | 0 | 0 | 0 |
| | MGT | 0 | 0 | 0 | 0 | 0 | 0 |

Table 4.26: (continued)

| Model | Method | S0 | S1 | S2 | S3 | S4 | S5 |
|---|---|---|---|---|---|---|---|
| M3 | MIX | 5 | 8 | 5 | 7 | 8 | 6 |
| | EMNR | 0 | 0 | 0 | 0 | 0 | 0 |
| | APEM/RMEM | 2 | 2 | 2 | 2 | 3 | 2 |
| | UDEM | 0 | 0 | 0 | 0 | 0 | 0 |
| | MGT | 0 | 0 | 0 | 0 | 0 | 0 |
| M4 | MIX | 0 | 0 | 0 | 0 | 0 | 0 |
| | EMNR | 0 | 0 | 0 | 0 | 0 | 0 |
| | APEM/RMEM | 0 | 0 | 0 | 0 | 0 | 0 |
| | UDEM | 0 | 0 | 0 | 0 | 0 | 0 |
| | MGT | 0 | 0 | 0 | 0 | 0 | 0 |
| M5 | MIX | 220 | 248 | 239 | 252 | 243 | 241 |
| | EMNR | 72 | 78 | 75 | 73 | 83 | 72 |
| | APEM/RMEM | 21 | 12 | 10 | 9 | 14 | 8 |
| | UDEM | 0 | 0 | 1 | 0 | 0 | 0 |
| | MGT | 0 | 0 | 0 | 0 | 0 | 0 |
| M6 | MIX | 159 | 154 | 139 | 144 | 150 | 159 |
| | EMNR | 25 | 34 | 25 | 24 | 40 | 23 |
| | APEM/RMEM | 9 | 6 | 3 | 3 | 7 | 3 |
| | UDEM | 0 | 0 | 0 | 0 | 0 | 0 |
| | MGT | 0 | 0 | 0 | 0 | 0 | 0 |
| M7 | MIX | 15 | 12 | 14 | 17 | 26 | 16 |
| | EMNR | 1 | 2 | 1 | 1 | 4 | 2 |
| | APEM/RMEM | 5 | 2 | 2 | 1 | 3 | 3 |
| | UDEM | 0 | 0 | 0 | 0 | 0 | 0 |
| | MGT | 0 | 0 | 0 | 0 | 0 | 0 |
| M8 | MIX | 3 | 4 | 5 | 4 | 12 | 5 |
| | EMNR | 0 | 0 | 0 | 0 | 0 | 0 |
| | APEM/RMEM | 0 | 0 | 0 | 0 | 0 | 0 |
| | UDEM | 0 | 0 | 0 | 0 | 0 | 0 |
| | MGT | 0 | 0 | 0 | 0 | 0 | 0 |
| M9 | MIX | 244 | 274 | 245 | 247 | 239 | 251 |
| | EMNR | 118 | 141 | 138 | 139 | 153 | 127 |
| | APEM/RMEM | 136 | 13 | 10 | 8 | 11 | 8 |
| | UDEM | 0 | 0 | 0 | 0 | 0 | 0 |
| | MGT | 0 | 0 | 0 | 0 | 0 | 0 |

Table 4.26: (continued)

| Model | Method | S0 | S1 | S2 | S3 | S4 | S5 |
|---|---|---|---|---|---|---|---|
| M10 | MIX | 232 | 225 | 225 | 223 | 215 | 231 |
| | EMNR | 56 | 76 | 56 | 55 | 78 | 52 |
| | APEM/RMEM | 67 | 5 | 6 | 6 | 11 | 5 |
| | UDEM | 1 | 0 | 1 | 0 | 1 | 0 |
| | MGT | 0 | 0 | 0 | 0 | 0 | 0 |
| M11 | MIX | 173 | 128 | 148 | 132 | 137 | 151 |
| | EMNR | 2 | 10 | 1 | 1 | 12 | 2 |
| | APEM/RMEM | 48 | 3 | 3 | 3 | 5 | 4 |
| | UDEM | 1 | 0 | 1 | 0 | 0 | 0 |
| | MGT | 0 | 0 | 0 | 0 | 0 | 0 |
| M12 | MIX | 76 | 67 | 69 | 74 | 77 | 81 |
| | EMNR | 0 | 0 | 0 | 0 | 0 | 0 |
| | APEM/RMEM | 10 | 1 | 1 | 1 | 1 | 4 |
| | UDEM | 0 | 0 | 0 | 0 | 0 | 0 |
| | MGT | 0 | 0 | 0 | 0 | 0 | 0 |

**Cumulated rRMSEs**

Turning to the estimation results in terms of the cumulated rRMSE reveals interesting results. Those are displayed in table 4.27. For a better illustration, the lowest values for each model and each method are color-coded. Beside the best values, results which are close to the best values are highlighted in a lighter color. Those values are restricted to values with a ratio smaller or equal 1.1. The results obtained by the true values (S0) are shown for a better understanding of the dimensions. Thus, at a glance one can see which technique provides suitable starting values and which does not.

Table 4.27: Cumulated rRMSEs for starting values S0 - S5.

| Model | Method | S0 | S1 | S2 | S3 | S4 | S5 |
|---|---|---|---|---|---|---|---|
| M1 | MIX | 1.95 | 1.94 | 2.01 | 2.00 | 1.96 | 2.02 |
| | EMNR | 1.35 | 1.36 | 1.44 | 1.42 | 1.35 | 1.43 |
| | APEM/RMEM | 0.48 | 0.56 | 0.85 | 0.82 | 0.48 | 0.79 |
| | UDEM | 0.56 | 0.60 | 0.85 | 0.84 | 0.56 | 0.82 |
| | MGT | 0.54 | 0.60 | 0.85 | 0.84 | 0.53 | 0.87 |
| M2 | MIX | 1.54 | 1.52 | 1.57 | 1.58 | 1.56 | 1.56 |
| | EMNR | 1.20 | 1.21 | 1.23 | 1.22 | 1.20 | 1.22 |
| | APEM/RMEM | 0.55 | 0.57 | 0.65 | 0.64 | 0.71 | 0.65 |

Table 4.27: (continued)

| Model | Method | S0 | S1 | S2 | S3 | S4 | S5 |
|---|---|---|---|---|---|---|---|
| | UDEM | 0.56 | 0.60 | 0.85 | 0.84 | 0.56 | 0.70 |
| | MGT | 0.57 | 0.59 | 0.67 | 0.65 | 0.63 | 0.68 |
| M3 | MIX | 0.86 | 0.85 | 0.87 | 0.86 | 0.88 | 0.86 |
| | EMNR | 0.78 | 0.78 | 0.78 | 0.78 | 0.79 | 0.78 |
| | APEM/RMEM | 0.55 | 0.55 | 0.56 | 0.56 | 0.62 | 0.56 |
| | UDEM | 0.61 | 0.60 | 0.62 | 0.64 | 0.72 | 0.62 |
| | MGT | 0.57 | 0.57 | 0.58 | 0.58 | 0.60 | 0.58 |
| M4 | MIX | 0.51 | 0.51 | 0.51 | 0.51 | 0.68 | 0.51 |
| | EMNR | 0.49 | 0.49 | 0.49 | 0.49 | 0.51 | 0.49 |
| | APEM/RMEM | 0.44 | 0.44 | 0.44 | 0.44 | 0.47 | 0.44 |
| | UDEM | 0.49 | 0.49 | 0.49 | 0.49 | 0.52 | 0.50 |
| | MGT | 0.46 | 0.46 | 0.46 | 0.46 | 0.47 | 0.46 |
| M5 | MIX | 1.77 | 1.81 | 1.85 | 1.86 | 1.83 | 1.92 |
| | EMNR | 1.24 | 1.33 | 1.39 | 1.37 | 1.32 | 1.39 |
| | APEM/RMEM | 0.50 | 0.80 | 0.89 | 0.87 | 0.75 | 0.87 |
| | UDEM | 0.57 | 0.82 | 0.90 | 0.86 | 0.79 | 0.89 |
| | MGT | 0.55 | 0.85 | 0.94 | 0.90 | 0.79 | 0.94 |
| M6 | MIX | 1.29 | 1.33 | 1.39 | 1.39 | 1.44 | 1.33 |
| | EMNR | 1.09 | 1.15 | 1.15 | 1.15 | 1.19 | 1.12 |
| | APEM/RMEM | 0.51 | 0.77 | 0.70 | 0.71 | 0.94 | 0.66 |
| | UDEM | 0.54 | 0.81 | 0.75 | 0.74 | 0.99 | 0.71 |
| | MGT | 0.53 | 0.79 | 0.73 | 0.73 | 0.88 | 0.68 |
| M7 | MIX | 0.79 | 0.82 | 0.81 | 0.81 | 1.09 | 0.79 |
| | EMNR | 0.75 | 0.80 | 0.77 | 0.78 | 0.91 | 0.75 |
| | APEM/RMEM | 0.51 | 0.79 | 0.60 | 0.60 | 0.99 | 0.56 |
| | UDEM | 0.55 | 0.83 | 0.65 | 0.65 | 1.07 | 0.61 |
| | MGT | 0.53 | 0.79 | 0.60 | 0.63 | 0.93 | 0.58 |
| M8 | MIX | 0.52 | 0.59 | 0.53 | 0.53 | 1.16 | 0.52 |
| | EMNR | 0.50 | 0.62 | 0.51 | 0.51 | 0.85 | 0.50 |
| | APEM/RMEM | 0.43 | 0.61 | 0.47 | 0.46 | 0.78 | 0.44 |
| | UDEM | 0.48 | 0.68 | 0.53 | 0.53 | 0.89 | 0.49 |
| | MGT | 0.45 | 0.59 | 0.47 | 0.47 | 0.68 | 0.46 |
| M9 | MIX | 1.70 | 1.86 | 1.95 | 1.94 | 1.91 | 1.94 |
| | EMNR | 1.26 | 1.48 | 1.53 | 1.52 | 1.48 | 1.50 |
| | APEM/RMEM | 0.60 | 1.07 | 1.08 | 1.07 | 0.98 | 1.05 |

Table 4.27: (continued)

| Model | Method | S0 | S1 | S2 | S3 | S4 | S5 |
|---|---|---|---|---|---|---|---|
| | UDEM | 0.68 | 1.05 | 1.07 | 1.06 | 0.98 | 1.05 |
| | MGT | 0.67 | 1.12 | 1.16 | 1.12 | 1.04 | 1.13 |
| M10 | MIX | 1.37 | 1.58 | 1.60 | 1.60 | 1.65 | 1.64 |
| | EMNR | 1.12 | 1.40 | 1.36 | 1.36 | 1.42 | 1.36 |
| | APEM/RMEM | 0.56 | 1.17 | 1.03 | 1.04 | 1.19 | 1.02 |
| | UDEM | 0.64 | 1.19 | 1.05 | 1.04 | 1.22 | 1.03 |
| | MGT | 0.63 | 1.19 | 1.07 | 1.07 | 1.20 | 1.05 |
| M11 | MIX | 0.98 | 1.27 | 1.13 | 1.13 | 1.32 | 1.09 |
| | EMNR | 0.91 | 1.29 | 1.11 | 1.12 | 1.31 | 1.09 |
| | APEM/RMEM | 0.58 | 1.40 | 0.99 | 1.00 | 1.45 | 0.95 |
| | UDEM | 0.65 | 1.44 | 1.02 | 1.04 | 1.48 | 0.99 |
| | MGT | 0.64 | 1.42 | 1.00 | 1.06 | 1.45 | 0.96 |
| M12 | MIX | 0.71 | 1.06 | 0.89 | 0.90 | 1.06 | 0.75 |
| | EMNR | 0.70 | 1.11 | 0.91 | 0.91 | 1.12 | 0.78 |
| | APEM/RMEM | 0.52 | 1.36 | 0.90 | 0.90 | 1.37 | 0.75 |
| | UDEM | 0.59 | 1.44 | 0.95 | 0.95 | 1.48 | 0.82 |
| | MGT | 0.56 | 1.38 | 0.89 | 0.95 | 1.39 | 0.77 |

best close to best less appropriate

Considering for instance model M1. The estimation results obtained using the starting values generated by technique S4 are better in almost each algorithm compared to the results obtained by other techniques. The same is valid for the other highly overlapped models M5 and M9. Consequently, it is deduced that for highly overlapped components, technique S4 provides the most appropriate initial values.

Having a look at mixtures with well balanced components like models M2 - M4, it seems that technique S1 provides the most suitable starting values. Utilizing these initial values leads to cumulated rRMSEs that in general are smaller than those achieved with other techniques. However, for well balanced and well separated mixture models, the choice of starting values seems to have minor influence on the results as shown by mixture model M3 and more clearly by mixture model M4. In the latter almost all techniques provide suitable initial values which yield results close to the true values.

As noted before, the correct decomposition of models with unbalanced components is much more difficult. All values of the cumulated rRMSE are higher compared to the values obtained by the true values. Especially when the components are highly unbalanced, like

M9 - M12, the resulting values are much higher. However, one of the techniques provides starting values that lead to the smallest cumulated rRMSEs throughout all models that are not highly overlapped. This is the new developed technique S5.

In addition, for unbalanced and well separated mixture models it should be avoided to use techniques S1 or S4, since both techniques generate initial values with which all considered methods provide estimates that differ substantially from the true parameters.

**Recommendations for Appropriate Initial Values**

In conclusion, not all starting value techniques proposed for the EM algorithm for mixtures can be suitably modified in a way that they can be applied for all situations with data sets containing grouped observations. In fact, if it is assumed that a data set consists of two components, it is recommended to first consider the corresponding histogram to get a first impression of the data. If the histogram reveals only one mode, although it is assumed that the data consists of two subpopulations, technique S4 is recommended for obtaining appropriate starting values.

In contrast, if there are two obvious peaks, the new technique S5 should be applied to generate initial values since with this technique the best results were achieved. However, if the histogram reveals two potentially well balanced components or if previous knowledge about the components exists, e.g. if it is known that a data set consists of measurements coming from a population with women and men whose measurements differ in a particular value, then technique S1 is to be preferred. In well balanced mixtures, this technique provided the most suitable starting values.

In brief, the following table demonstrates, which starting value technique should be taken in a certain data situation. An illustration of the whole proceeding is shown in chapter 5, where two examples are given.

Table 4.28: Recommended techniques for obtaining starting values dependent on the shape of the histogram.

| separated | | overlapped |
|---|---|---|
| balanced | unbalanced | |
| S1 | S5 | S4 |

Finally, since none of the algorithms and none of the starting value techniques can assure convergence to the global maximum, it is recommended to use more than one set of starting values to ensure that a solution is the global one. In many mixture models, the techniques S2 and S3 provide initial values with which results are obtained that are close to the best ones. Hence, it is recommended to use these techniques additionally to confirm an obtained result.

## 4.8 Variation of the Interval Width

This section is divided into two parts. In the first part, the results of the variation of the interval width for selected models and the corresponding recommendations are presented. In the second part, it is attempted to generalize these results for arbitrary models and a practical proceeding is proposed.

### 4.8.1 Results for Different Interval Widths

In the previous simulation studies it is assumed that the interval width is 1. In practice, this is often the case when measured values can only be collected as integers, which in other words is a natural grouping with interval width $h = 1$. However, in some research fields the measured data may be grouped into intervals of other widths. If, for instance, the measurements are quantified with one decimal place, a grouping with interval width $h = 0.1$ occurs. In contrast, in other research areas, the measurements may be summarized in intervals of width $h = 5$ or more. Hence, this section addresses the problem of estimating the parameters of a two component Gaussian mixture with observations grouped into intervals with different widths.

First, some small interval widths like $h = 0.1$, $h = 0.2$, and $h = 0.5$ are investigated. Then, from $h = 1$ on the length of the intervals is increased in steps of 0.1 until $h = 2.5$ is reached. On the one hand, as before, the cumulated rRMSE based on 1000 samples is chosen as comparison criterion. Thereby, again only those samples are evaluated, which satisfy the convergence criteria. On the other hand, the number of the remaining samples, i.e. the number of convergence failures, is compared.

**Results for Mixture Model M4**
In the following the results for the well separated and balanced mixture model M4 are presented. A graphical illustration of the results can be found in figure 4.14, where the black curve displays the cumulated rRMSE and the red one the convergence failures. The numerical results for that model and all others are shown in table A.2 in the appendix.

The cumulated rRMSE increases with increasing interval width. However, there are differences between the methods. For example, the increase of the cumulated rRMSEs of the APEM and RMEM algorithms is less pronounced than those of the other methods. Even for large interval widths like $h = 2.5$, their cumulated rRMSE is 0.78 and 0.79, respectively. As previously mentioned, for interval width $h = 1$, both methods provide the same estimates. For other interval widths, the estimators are still the same but the likelihood functions differ. Accordingly, the iteration procedures terminate at different stages and the results differ. However, since the difference is small, both methods provide similar results, which are in comparison very good.

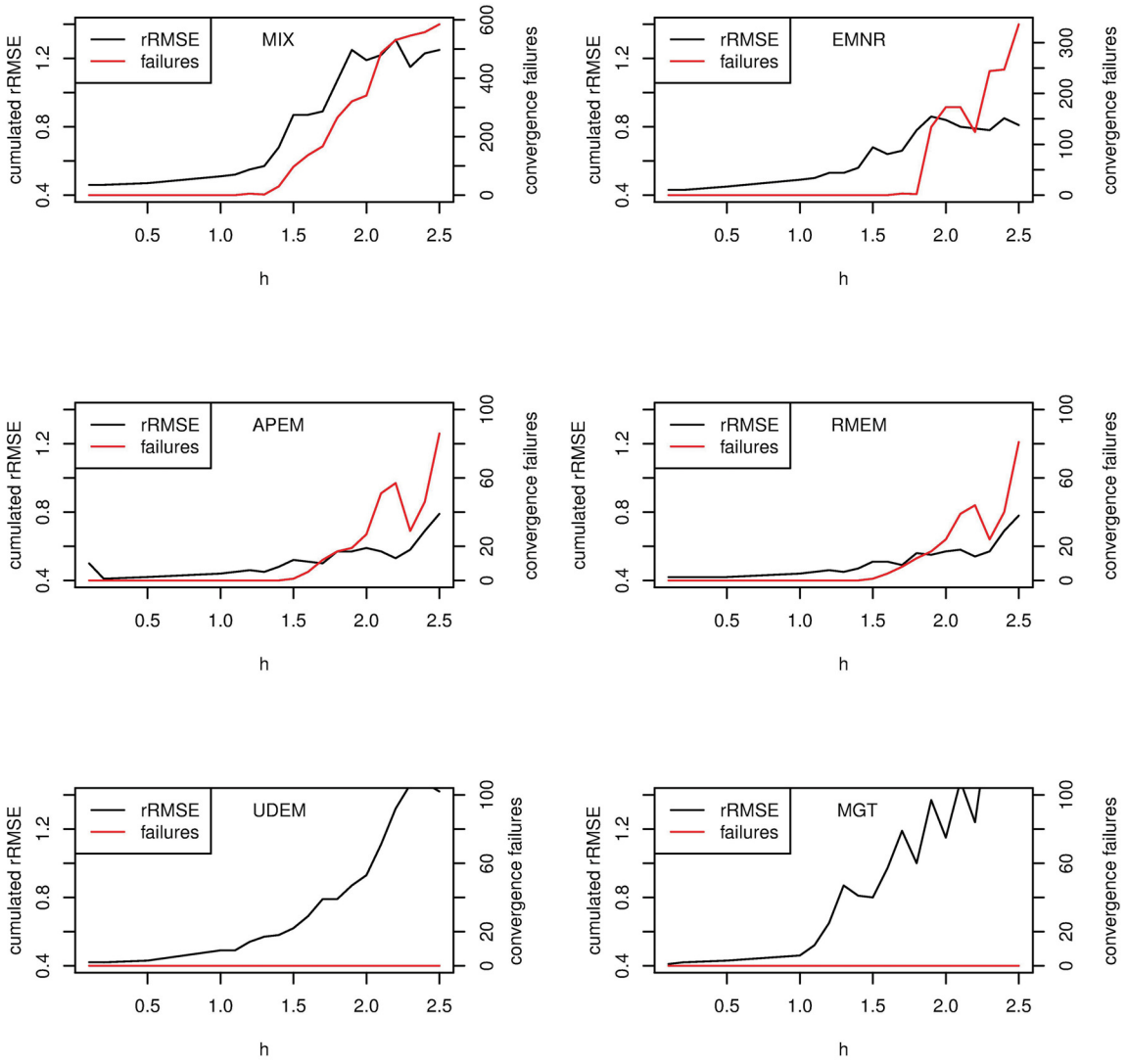



Figure 4.14: Cumulated rRMSE (black line) and convergence failures (red line) for mixture model M4 with interval widths $h$ from 0.1 to 2.5.

Considering the EMNR algorithm, the increase of the cumulated rRMSE is similar to that of the APEM and RMEM algorithms. Indeed, the slope of the black line is slightly higher in larger interval widths. However, the final value of the cumulated rRMSE at an interval width of $h = 2.5$ is 0.81, which is similar to those obtained by the APEM and RMEM algorithms.

In contrast, the MIX algorithm provides worse results even for small interval widths. Actually, for interval widths larger than about $h = 1.4$, a high slope of the black curve can be observed, which indicates a higher deviation of the estimates from the true parameters. For example, for the largest considered interval width, $h = 2.5$, a cumulated rRSME of 1.25 was calculated.

Furthermore, considering the UDEM and MGT algorithms, a similar development of values like with the MIX method occurs. In fact, the slope of both methods is even higher and ends at an interval width of $h = 2.5$ with a cumulated rRMSE of 1.42 and 1.99, respectively. Particularly, the MGT algorithm seems to provide appropriate estimation results only for interval widths smaller than 1.

Turning to the convergence failures, one can see that both the UDEM and MGT algorithms are nearly free of convergence problems. In contrast, the number of failures of the APEM and RMEM algorithms increases with increasing interval width up to a rate of 8.6% and 8.1%, respectively. The increasing number of convergence failures in consequence of larger interval widths can be seen even more clearly considering the methods with more than one iteration procedure. If the interval width is $h = 2.5$, nearly a third of the samples is not satisfying the convergence criteria if using the EMNR algorithm. Applying the MIX algorithm, almost 60% of the samples have to be omitted due to convergence problems.

In brief, for this well separated and balanced model and interval widths less than about $h = 1.1$, a solution for a particular sample is found with a high probability independent of the method applied. Larger interval widths have to be evaluated with caution. In this simulation study, the best results for interval widths larger than $h = 1.1$ were obtained using the APEM and RMEM algorithms. However, since the number of failures increases with increasing interval width, a solution may not be achieved at all. So if these algorithms provide a solution, it is probably the best one. In contrast, for nearly every sample a solution is obtained when using the UDEM and MGT algorithms. However, this solution might differ strongly from the true values.

As has been described before, it is important to keep the underlying mixture model in mind. In particular, the degree of separation and the balance are crucial. The preceding results are only valid for well separated and balanced models. For other models different recommendations arise. For a better illustration, the results of the highly overlapped and unbalanced model M9 are presented in the following.

**Results for Mixture Model M9**

For this mixture model, previous findings concerning the quality of the algorithms could be confirmed. For instance, particularly for small interval widths, the MIX and EMNR algorithms provide much higher cumulated rRMSEs in this mixture models than those of the other methods, see figure 4.15. In contrast, the APEM and RMEM algorithms provide small cumulated rRMSEs. Nevertheless, compared to the previous well separated model, the results are worse even for small widths. This is natural, since the decomposition of mixtures with well separated components is easier than that of mixtures with highly overlapped components, as was shown in section 4.2.

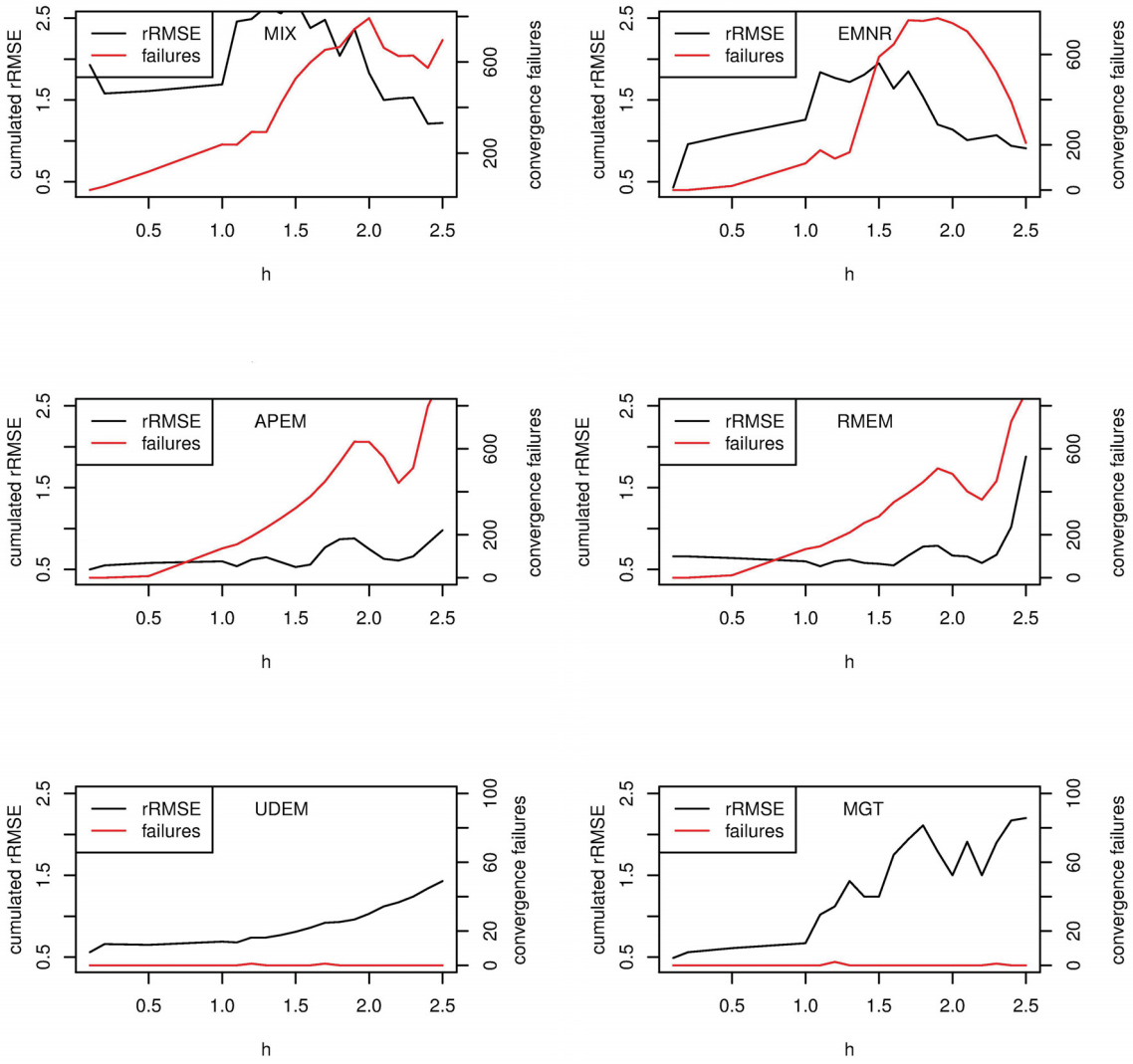



Figure 4.15: Cumulated rRMSE (black line) and convergence failures (red line) for mixture model M9 with interval widths $h$ from 0.1 to 2.5.

Furthermore, considering the accuracy of the UDEM and MGT algorithms in terms of the cumulated rRMSE, similar results to those of the APEM and RMEM algorithms are observed for small interval widths. Larger interval widths cause a high slope of the black curve of the MGT method, which indicates that this method is less appropriate in such cases. This finding is similar to that in the previously presented mixture model.

Turning to the convergence failures one can see again that the UDEM and MGT algorithms are still nearly free of convergence failures, which is the greatest advantage of these methods. In contrast, the convergence failure rates of the APEM and RMEM algorithms increase more compared to the previous model. The latter method reaches values up to 941.

However, the convergence failure rates have to be evaluated with caution. As has been

shown in the previous section, the failure rates are dependent on the choice of the starting values. In this investigation, the true values are taken as initial values. As has been seen before, the true values may cause more convergence failures, since a particular sample might have been unfavorably generated. Hence, these rates should only be taken as orientation.

To conclude, depending on the interval width, some methods are more appropriate than others to estimate the parameters of a two component Gaussian mixture. Using an interval width of $h = 1.5$, the APEM and RMEM algorithms, for example, provide on average estimates with the smallest cumulated rRMSEs. Therefore, one of those algorithms should be preferred. If, however, a solution can not be found due to a convergence failure, the UDEM algorithm is recommended. The estimates might be not as accurate as if they were obtained by the APEM algorithm, but with a high probability a solution can be obtained.

### 4.8.2 Generalization of the Results

The investigated models do not cover each possible parameter constellation of the six parameters of the two component Gaussian mixture and each possible interval width. In practice, however, many different situations occur. Hence, it is desired to transfer the obtained results to other data situations. Therefore, a kind of standardization is necessary. Here, a standardization of the interval width with respect to the mixing standard deviation is suggested. As seen in (4.1), the variance of a two component mixture is given by

$$\sigma_m^2 = \pi_1(\sigma_1^2 + \mu_1^2) + \pi_2(\sigma_2^2 + \mu_2^2) - \mu_m^2.$$

The recommendations given in the previous section are only valid for particular mixture models and interval widths. To apply these results in other mixture models and for a variety of interval widths, it is suggested to consider the relative interval width $H$.

**Definition 4.1** *The **relative interval width H** is defined as the ratio of a given interval width $\widetilde{h}$ and the standard deviation of the mixture $\sigma_m$, i.e.*

$$H = \widetilde{h}/\sigma_m. \tag{4.1}$$

Table 4.29 shows the relative interval width $H$ for the twelve mixture models and the corresponding interval width $h$ used in this work.

Table 4.29: Relative interval width $H$ depending on the underlying mixture model.

| M1 | M2 | M3 | M4 | M5 | M6 | M7 | M8 | M9 | M10 | M11 | M12 | $h$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.09 | 0.07 | 0.06 | 0.04 | 0.09 | 0.07 | 0.06 | 0.05 | 0.10 | 0.09 | 0.07 | 0.06 | 0.1 |
| 0.18 | 0.14 | 0.11 | 0.09 | 0.18 | 0.15 | 0.12 | 0.10 | 0.19 | 0.17 | 0.15 | 0.13 | 0.2 |
| 0.45 | 0.35 | 0.28 | 0.22 | 0.45 | 0.37 | 0.29 | 0.24 | 0.48 | 0.43 | 0.37 | 0.32 | 0.5 |
| 0.89 | 0.71 | 0.55 | 0.45 | 0.91 | 0.74 | 0.59 | 0.48 | 0.96 | 0.86 | 0.74 | 0.64 | 1.0 |

Table 4.29: (continued)

| M1 | M2 | M3 | M4 | M5 | M6 | M7 | M8 | M9 | M10 | M11 | M12 | $h$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.98 | 0.78 | 0.61 | 0.49 | 1.00 | 0.81 | 0.65 | 0.53 | 1.05 | 0.94 | 0.82 | 0.70 | 1.1 |
| 1.07 | 0.85 | 0.67 | 0.54 | 1.09 | 0.88 | 0.71 | 0.57 | 1.15 | 1.03 | 0.89 | 0.77 | 1.2 |
| 1.16 | 0.92 | 0.72 | 0.58 | 1.18 | 0.96 | 0.76 | 0.62 | 1.25 | 1.11 | 0.97 | 0.83 | 1.3 |
| 1.25 | 0.99 | 0.78 | 0.63 | 1.27 | 1.03 | 0.82 | 0.67 | 1.34 | 1.20 | 1.04 | 0.90 | 1.4 |
| 1.34 | 1.06 | 0.83 | 0.67 | 1.36 | 1.11 | 0.88 | 0.72 | 1.44 | 1.29 | 1.11 | 0.96 | 1.5 |
| 1.43 | 1.13 | 0.89 | 0.72 | 1.45 | 1.18 | 0.94 | 0.77 | 1.53 | 1.37 | 1.19 | 1.02 | 1.6 |
| 1.52 | 1.20 | 0.94 | 0.76 | 1.55 | 1.25 | 1.00 | 0.81 | 1.63 | 1.46 | 1.26 | 1.09 | 1.7 |
| 1.61 | 1.27 | 1.00 | 0.80 | 1.64 | 1.33 | 1.06 | 0.86 | 1.72 | 1.54 | 1.34 | 1.15 | 1.8 |
| 1.70 | 1.34 | 1.05 | 0.85 | 1.73 | 1.40 | 1.12 | 0.91 | 1.82 | 1.63 | 1.41 | 1.22 | 1.9 |
| 1.79 | 1.41 | 1.11 | 0.89 | 1.82 | 1.47 | 1.18 | 0.96 | 1.92 | 1.71 | 1.49 | 1.28 | 2.0 |
| 1.88 | 1.48 | 1.16 | 0.94 | 1.91 | 1.55 | 1.24 | 1.01 | 2.01 | 1.80 | 1.56 | 1.34 | 2.1 |
| 1.97 | 1.56 | 1.22 | 0.98 | 2.00 | 1.62 | 1.29 | 1.05 | 2.11 | 1.89 | 1.64 | 1.41 | 2.2 |
| 2.06 | 1.63 | 1.28 | 1.03 | 2.09 | 1.70 | 1.35 | 1.10 | 2.20 | 1.97 | 1.71 | 1.47 | 2.3 |
| 2.15 | 1.70 | 1.33 | 1.07 | 2.18 | 1.77 | 1.41 | 1.15 | 2.30 | 2.06 | 1.78 | 1.54 | 2.4 |
| 2.24 | 1.77 | 1.39 | 1.12 | 2.27 | 1.84 | 1.47 | 1.20 | 2.39 | 2.14 | 1.86 | 1.60 | 2.5 |

A practical proceeding may be the following. First, as before, it is recommended to consider the histogram of the grouped data to get a first impression of the data. To decide which mixture model is to be fitted, table 4.30 is given.

Table 4.30: Mixture Models listed dependent on the distribution shape.

| $\pi_1 = 0.5$ | | | | $\pi_1 = 0.7$ | | | | $\pi_1 = 0.9$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| overlapped | | separated | | overlapped | | separated | | overlapped | | separated | |
| M1 | M2 | M3 | M4 | M5 | M6 | M7 | M8 | M9 | M10 | M11 | M12 |

Second, the standard deviation of the mixture is required. Of course, in practical situations the standard deviation of the mixture can not be calculated since the component parameters are not known. However, the empirical standard deviation is a good estimation for the mixed standard deviation. This calculation may be achieved by using the data mapping as proposed in section 3.2.1.

Third, having estimated the standard deviation of the mixture, the relative interval width $H$ may be calculated by division of the given interval width $\widetilde{h}$ by the empirical standard deviation $\widehat{\sigma}_m$. With the relative interval width $H$, one can have a look at table 4.29, where the relative interval widths for the different mixture models are presented.

Fourth, having found the calculated value in the corresponding column, the interval width

$h$ used in this work can be obtained.

Finally, depending on the value of $h$, one may have a clue whether adequate estimations can be obtained by a particular data set. It can be seen, which method is best for this particular data set, by having a look at table A.2 in the appendix, where the best results and values close to the best are displayed color-coded.

To demonstrate the whole procedure, two examples, which are investigated in the next chapter, are given. In the first example, the histogram of a sample reveals two well separated components that are unbalanced. Accordingly, a mixture model like M8 may be fitted. This example is presented in detail in section 5.1. The estimated standard deviation of the mixture is $\widehat{\sigma}_m = 13.6$. The interval width is given to be $\widetilde{h} = 3$. Hence, the relative interval width is $H = 3/13.6 = 0.22$. Looking for this value in table 4.29 (column of M8), a corresponding value for the interval width $h$ of slightly less than 0.5 is obtained. Since all presented methods provide appropriate estimation results with such a grouping, see table A.2, all of them may be used for the fitting of a mixture to this particular data set.

The histogram of the second example that will be discussed in section 5.2 reveals a highly overlapped mixture with probably highly unbalanced components. The fitting of a mixture model like M9 might be adequate. With a given interval width of $\widetilde{h} = 1$ and an estimated mixture standard deviation of $\widehat{\sigma}_m = 3.9$, a relative interval width of $H = 1/3.9 = 0.26$ results. Looking for this value in table 4.29 (column of M9), it can be seen that this value corresponds to an interval width $h$ of clearly less than 0.5. Hence, as well in this example all presented methods are adequate to fit a mixture to the given data set, see table A.2.

Alternatively to the proposed proceeding, for example, if one does not want to use different algorithms for different models, it is recommended to use the APEM or the RMEM algorithm. As previously mentioned, these algorithms provide the best results or results that are close to the best for almost all models and almost all interval widths. This can be confirmed by having a look at table A.2, where all of the results are summarized. Nevertheless, it has to be kept in mind that the convergence failures increase with increasing interval widths and that for a particular sample, a solution might not be found at all, particularly, if the components are highly overlapped and the interval widths are large. If a solution can not be found, it is suggested to use the UDEM or MGT algorithm, since with these algorithm a solution is obtained by a high probability.

# 5 Practical Application

In this chapter, two examples are presented to demonstrate on the one hand the practical application of the proposed methods and on the other hand to illustrate the proposed techniques for obtaining a suitable set of starting parameters. In the first example the data are grouped into intervals of length $\widetilde{h} = 3$. Due to the shape of the histogram, a fitting of a mixture with two well separated components seems reasonable. The second example contains data that are grouped into intervals of length $\widetilde{h} = 1$. The histogram indicates that a mixture with highly overlapped components is an appropriate fit.

## 5.1 Example 1: Old Faithful Geyser Data

Geysers are natural fountains that erupt heated water and steam into the air, at more or less regular intervals. Old Faithful is such a geyser and is the most popular attraction of the Yellowstone National Park. From August 1st to August 15th, 1985, Old Faithful was observed and the waiting time between eruptions noted. For a more detailed description, the reader is referred to Everitt and Hothorn (2006).

This *Old Faithful Geyser data set* is included in the data set package in R and was analyzed by several authors, for instance Hunter et al. (2007) and Benaglia et al. (2009). As mentioned previously, it is important to get a first impression of the data, hence figure 5.1 shows the corresponding histogram consisting of 272 observations that are grouped into intervals of width $\widetilde{h} = 3$.

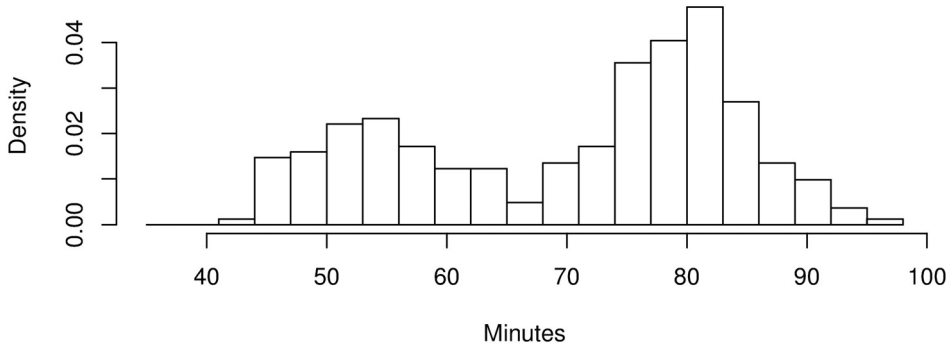



Figure 5.1: The Old Faithful Geyser data

Because of the bimodality in the histogram, a two component Gaussian mixture seems to be a reasonable model. First, a suitable estimation method must be chosen. Therefore,

as described in the previous section, the relative interval width $H$ is calculated. This indicates here that all six presented methods are adequate to fit a mixture model to these data.

The next step is to find suitable starting values. As shown in section 4.7, starting values for such a well separated and unbalanced model can be obtained by the new technique S5. Applying this strategy to the underlying sample yields the following initial values

$$\begin{aligned} \pi_1^{(0)} &= 0.38 \quad \mu_1^{(0)} = 44.5 \quad \sigma_1^{(0)} = 7.48 \\ \pi_2^{(0)} &= 0.62 \quad \mu_2^{(0)} = 92.5 \quad \sigma_2^{(0)} = 7.48. \end{aligned}$$

Using these starting values, the six algorithms provide estimation results, which are summarized in table 5.1.

Table 5.1: Estimation results for the Old Faithful data set

| Method | $\widehat{\pi}_1$ | $\widehat{\pi}_2$ | $\widehat{\mu}_1$ | $\widehat{\mu}_2$ | $\widehat{\sigma}_1$ | $\widehat{\sigma}_2$ |
|---|---|---|---|---|---|---|
| MIX | 0.36 | 0.64 | 54.26 | 79.64 | 6.11 | 5.84 |
| EMNR | 0.36 | 0.64 | 54.22 | 79.63 | 6.10 | 5.85 |
| APEM | 0.37 | 0.63 | 54.39 | 79.71 | 6.26 | 5.81 |
| RMEM | 0.37 | 0.63 | 54.39 | 79.71 | 6.26 | 5.81 |
| UDEM | 0.37 | 0.63 | 54.74 | 79.89 | 6.65 | 5.86 |
| MGT | 0.37 | 0.63 | 54.18 | 79.78 | 6.38 | 5.97 |

For a better illustration, these results are illustrated in figure 5.2, where the corresponding densities are displayed.

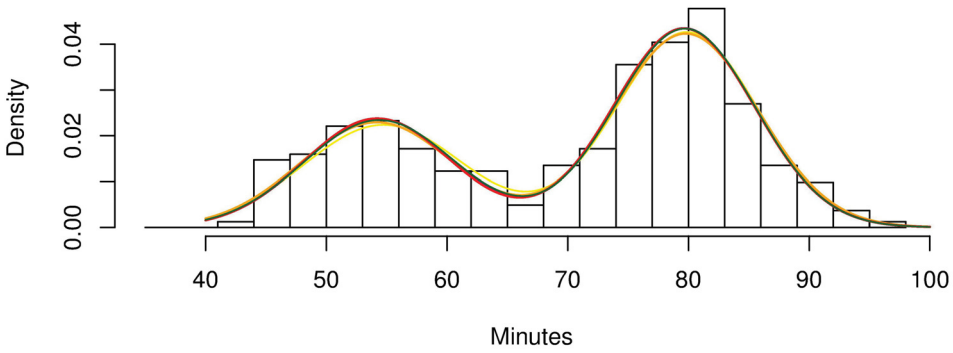



Figure 5.2: Fitted densities to the Old Faithful Geyser data

The first thing that can be observed is, that all methods provide similar estimation results. In fact, the corresponding densities can hardly be distinguished from each other. Accordingly, each algorithm seems to be adequate for the fitting of a two-component Gaussian mixture to these data. In particular, the APEM and RMEM algorithms have once again proven themselves. Furthermore, the starting values appear to be suitably chosen, which

reinforces the recommendation of the proposed strategy in this setting.

Benaglia et al. (2009) fitted also a two-component gaussian mixture to the corresponding continuous data by using $\pi_1^{(0)} = \pi_2^{(0)} = 0.5, \mu_1^{(0)} = 55, \mu_2^{(0)} = 80, \sigma_1^{(0)} = \sigma_2^{(0)} = 6$ as starting values. They obtained by their function `normalmixEM` the following estimation results:

$$\begin{aligned} \widehat{\pi}_1 &= 0.36 \quad \widehat{\mu}_1 = 54.61 \quad \widehat{\sigma}_1 = 5.87 \\ \widehat{\pi}_2 &= 0.64 \quad \widehat{\mu}_2 = 80.09 \quad \widehat{\sigma}_2 = 5.87. \end{aligned}$$

Comparing their results to those obtained by grouped observations, it can be seen that all presented methods only differ slightly from the estimation without grouping. Hence, it can be deduced that all of them provide results that are approximately as good as those obtained by estimation techniques without grouping.

## 5.2 Example 2: Reference Limits for Laboratory Parameters

In health-related fields, a reference range or reference limit usually describes the variation of a measurement or value in healthy individuals. It is a basis for the interpretation of a set of results of a laboratory parameter for a particular patient. The standard definition of a reference range basically originates in what is most prevalent in a reference group taken from the population. In fact, the reference range for a particular measurement is defined as the prediction interval within 95% of values of a reference group fall into. A detailed description of the concept of reference limits can be found in Gräsbeck (2004).
To demonstrate the proceeding of finding a reference limit for a particular laboratory parameter, a data set from the central laboratory *Städtisches Krankenhaus, Kiel, Germany*, is adapted. Specifically, a data set containing 10183 adults is investigated to obtain a reasonable reference range for sodium serum concentration. To get a first impression of the data, a histogram of the data set is displayed in figure 5.3.

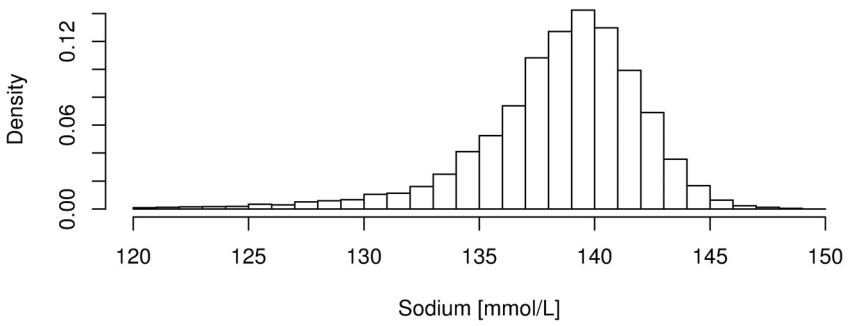



Figure 5.3: Histogram of the sodium data set

This data set consists of non-diseased and diseased subjects, which obviously demands the fitting of a two component mixture. A decomposition of the two components is accompanied by the decomposition of the non-pathological and pathological subpopulations. It is assumed that the central part of the distribution represents the non-diseased, "healthy" population, see Arzideh et al. (2009). Fitting a Gaussian distribution to the non-pathogenic values facilitates an estimation of the reference range, which is circumscribed by the 2.5% percentile in the lower limit and the 97.5% percentile in the upper limit.

The grouping of the observations occurs, because the values are rounded off whole numbers. This, in fact, represents a grouping into intervals with width $\widetilde{h} = 1$. Since it is known that this data set contains values of pathogenic and non-pathogenic subjects, the challenge is to decompose the pathogenic ones to predict a confidence interval for the non-pathogenic subjects and, hence, obtain a reference range for the "healthy" individuals.

As only one peak is observable, a highly overlapped mixture seems to be reasonable. Accordingly to section 4.8.2, the relative interval width $H$ for this data set reveals that all six methods are appropriate to be applied.

Furthermore, the question arises which technique should be taken to obtain suitable initial values. The answer can be found in section 4.7, where it was shown that the starting values for this type of mixtures should be calculated by technique S4, which is a modification of the method introduced by Seidel et al. (2000) as presented in section 3.5. This technique gives the following starting values

$$\begin{aligned} \pi_1^{(0)} &= 0.5 \quad \mu_1^{(0)} = 136.48 \quad \sigma_1^{(0)} = 3.91 \\ \pi_2^{(0)} &= 0.5 \quad \mu_2^{(0)} = 140.38 \quad \sigma_2^{(0)} = 3.91. \end{aligned}$$

Using these initial values, the estimation results of the six methods are given in table 5.2.

Table 5.2: Parameter estimations for the sodium data set

| Method | $\widehat{\pi}_1$ | $\widehat{\pi}_2$ | $\widehat{\mu}_1$ | $\widehat{\mu}_2$ | $\widehat{\sigma}_1$ | $\widehat{\sigma}_2$ |
|---|---|---|---|---|---|---|
| MIX | 0.14 | 0.86 | 133.52 | 139.24 | 6.24 | 2.57 |
| EMNR | 0.15 | 0.85 | 133.78 | 139.25 | 6.23 | 2.56 |
| APEM | 0.19 | 0.81 | 134.68 | 139.34 | 5.99 | 2.46 |
| RMEM | 0.19 | 0.81 | 134.68 | 139.34 | 5.99 | 2.46 |
| UDEM | 0.20 | 0.80 | 134.69 | 139.33 | 5.99 | 2.48 |
| MGT | 0.19 | 0.81 | 134.50 | 139.32 | 6.09 | 2.47 |

As before, the MIX and EMNR algorithms provide similar results, which differ from those of the other methods. Particularly, the estimations of the mixing proportions and the component means are smaller. In contrast, the standard deviations obtained by MIX and EMNR are higher compared to those provided by the other methods.
The results obtained by the APEM and RMEM algorithms are the same since the interval width is $\widetilde{h} = 1$. They only differ slightly to the results achieved by the UDEM and MGT algorithms.

Recall that the aim was to estimate a reference limit. Hence, the corresponding 2.5% and 97.5% percentiles of the distributions of the "healthy" population are of interest and shown in figure 5.4. The densities of the two component Gaussian mixture with the parameter estimations from table 5.2 are displayed and as well the respective reference limits for each method.

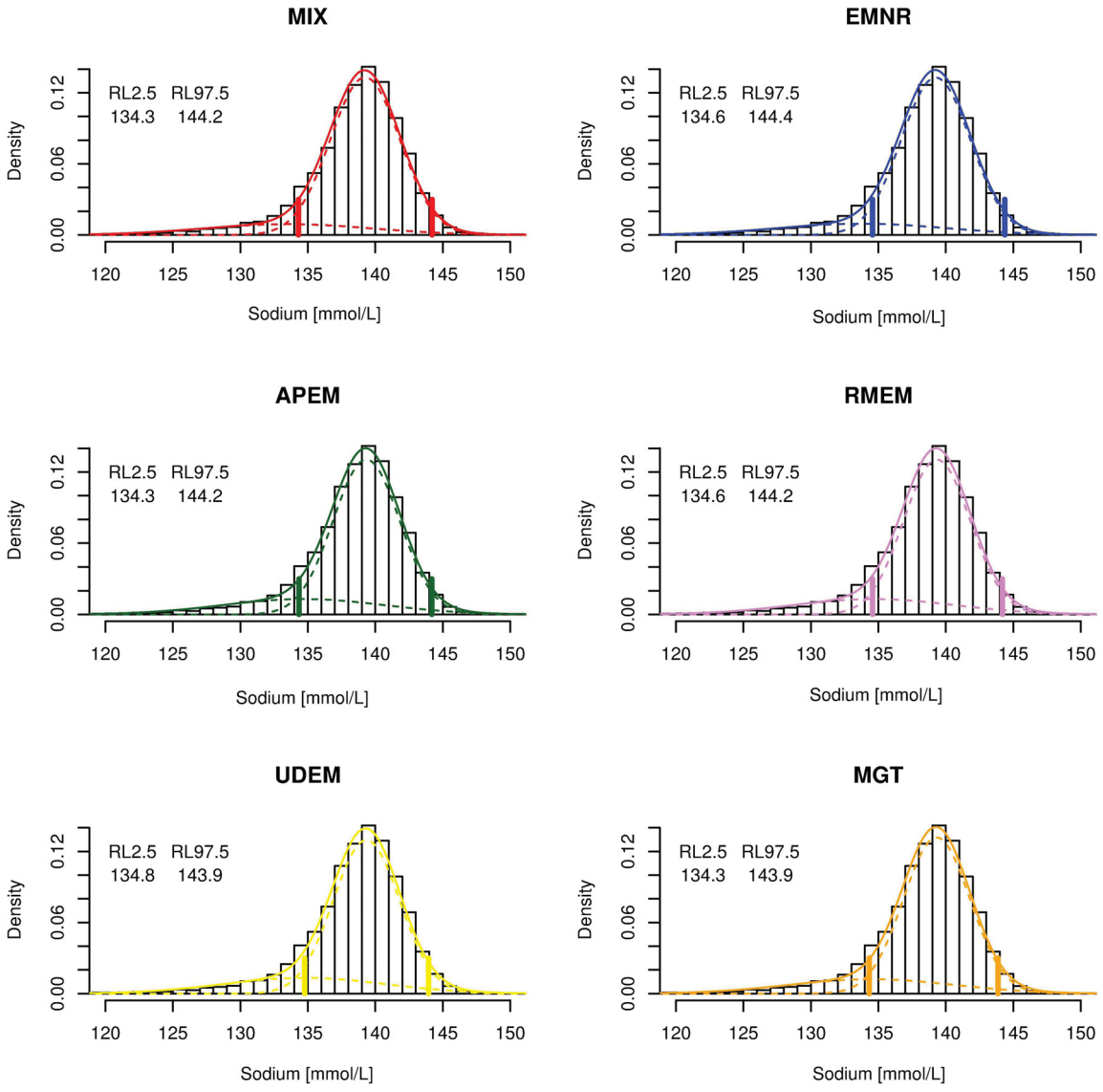



Figure 5.4: Fitted densities and estimated reference limits for the sodium data set

Although the estimation results for the parameters of the MIX and EMNR algorithms are much smaller, the estimations of the reference limits equals those of the others. In fact, all estimated 95% reference limits are close to each other and when comparing them to that of Arzideh et al. (2009), who calculated a reference range of $134.5 - 144.5$, only a minor difference is visible.

In brief, evaluating the estimations obtained using grouped observations with small relative interval widths like in the presented example, leads to the conclusion that the estimations are almost as good as those without grouping. Particularly, simple algorithms like APEM and RMEM are to be preferred due to their fast calculation and their easy implementation.

# 6 Summary and Discussion

## 6.1 Summary

This work dealt with the problem of fitting a mixing distribution to data that are given in a grouped form with intervals of equal length. A fitting can be achieved by estimating the parameters of the mixing distribution, which are the parameters of each component and the mixing weights in which the components occur. Especially in the case of a two component Gaussian mixture, there are five parameters to determine. Generally, as has been shown by many authors, the maximum likelihood estimation is a meaningful strategy to find adequate estimations. In this case, however, no analytical solutions for the corresponding likelihood equations can be found and a numerical procedure is essential. The EM algorithm is such a numerical procedure with which solutions may conveniently be obtained. Hence, the present work focuses on methods based on the EM approach to estimate the parameters of a two component Gaussian mixture.

In fact, six algorithms were considered. One of them is based on an approach proposed by MacDonald and Pitcher (1979). This approach was improved by Du (2002) to an algorithm called MIX, which contains a combination of the EM and NR algorithms and is available for the R program. Based on this idea, two further algorithms were proposed here. The first, EMNR, also contains two iteration procedures. The second one, abbreviated as APEM, uses an approximation to avoid a second iteration procedure. Furthermore, two algorithms, which circumvent the grouping by a data mapping, were proposed. The first one, denoted by RMEM, repeats the midpoints of each interval the number of observations fall into. The second one, called UDEM, generates individual observations from a uniform distribution for each interval. In both cases, the results are individual observations for which the EM algorithm for finite mixtures may be easily applied. Finally, the last considered method proposed by McLachlan and Jones (1988) and called MGT, uses an approach, which affords a closed-form solution for each likelihood equation. Since this algorithm was only available for the obsolete Fortran77, it was implemented here in a function for the R program.

To examine which algorithm is best for which data situation, several simulation studies were performed. To cover a wide range of possible data situations, twelve mixture models were investigated, which varied in the separation and balance of the components as can be seen in figure 4.1. It turned out from the first simulation study that for fixed sample size (n=100), interval width (h=1), and true parameters as initial values, the new simple methods that circumvent the grouping provided surprisingly good results. They failed only

in a small number of samples due to convergence problems and provided throughout all mixture models parameter estimations that hardly varied from the true parameters. Especially the RMEM algorithm seemed to be a good choice and an adequate alternative to the available algorithms MIX and MGT, which were both inferior. Furthermore, it was found that for interval width $h = 1$, the estimators of the APEM and RMEM algorithms are equal and, therefore, all results for the RMEM algorithm are as well valid for the APEM algorithm. In contrast, algorithms that contain more than one iteration procedure, like MIX and EMNR, performed worse compared to the others. Especially highly overlapped mixtures could predominately not be decomposed. Moreover, these methods failed in a distinctly higher number of samples and clearly demanded more time for a performance.

The second simulation study addressed the problem of the sample size. An increase of the sample size revealed that higher numbers of observations led to less convergence problems and more accurate estimations, especially in algorithms with only one iteration procedure. In particular, the number of observations suggested by Hosmer (1973b) could be confirmed. He stated that samples with less than 300 observations may yield infeasible results. Indeed, it was more difficult to decompose the components of samples with only 100 observations than those with 300. Moreover, the results were more reliable. This effect is even more pronounced in samples of size $n = 500$. In this case, the convergence failure rates decreased and the results were more accurate. In addition, it has to be kept in mind that even for a high number of observations, the MIX algorithm performed distinctly worse in mixture models with highly overlapped components. Hence, in such mixture models the results have to be examined with caution.

The third simulation study dealt with the variation of the starting values. Each iteration procedure requires suitable initial values. The choice of those decides whether a global or a local maximum is achieved or may even cause divergence. Hence, finding a suitable set of starting values is of great importance and since no advises could be found in literature for this case, several new techniques were considered. In fact, four techniques, which are modifications of techniques for the EM algorithm for mixture models with available individual observations, were examined. Furthermore, one new technique was developed to compensate the disadvantages of the previous strategies. The present study revealed that with almost all techniques starting values were determined with which the methods achieved in almost all mixture models smaller convergence failure rates compared to the results obtained by the true values. This is natural, since for each sample individual starting values were generated. Even though, there were great differences between the accuracy of the estimation results. As the accuracy of the estimation results substantially depended on the underlying mixture model, recommendations were made on the basis of the shape of the histogram of the data. These recommendations are summarized in table 4.28. In brief, in many cases the new technique provided initial values that were superior to the other techniques. Comparing the methods against each other led to the result that the

algorithms with more than one iteration procedure tended to be less sensitive to the choice of starting values.

The fourth and last simulation study focused on the influence of a variation of the interval widths. As expected, in almost all cases the estimation results became worse with larger interval width. Especially interval widths larger than 1 partially caused inaccurate results dependent on the underlying mixture model. As an exception, the MIX and EMNR algorithms had problems in decomposing highly overlapped mixtures irrespective of the interval width. Even for small interval widths no adequate results could be obtained. In contrast, the other methods provided good results with interval widths less than 1. Particularly, the results of the APEM or RMEM algorithms were close to each other and moreover in most considered interval widths and mixture models the best estimations.

According to all those findings, it is attempted to generalize the results and to formulate a proceeding how to handle grouped observations while estimating the parameters of a mixture. First the histogram has to be examined. Correspondingly to the shape of the histogram, it is proposed to use a technique for obtaining starting values in accordance with table 4.28. Having found suitable initial values, an adequate method to estimate the parameters has to be chosen. Either, one follows the proceeding proposed in section 4.8.2 and calculates the relative interval width to decide whether and with which method for a particular sample accurate estimations can be obtained. Or one always chooses the APEM or the RMEM algorithm, since with both methods good results were achieved in a high number of samples in different data situation. The only disadvantage is that for unfavorable data records, both algorithms may not provide a solution due to divergence of the corresponding likelihood. In such cases, the UDEM or for small interval widths the MGT algorithm is recommended, since both algorithms provide with high probability a solution, which, however, might be less accurate than that one obtained by APEM or RMEM.

## 6.2 Outlook

In the present work, the focus was on two component Gaussian mixtures. In practice, however, a composition of arbitrary components is supposable. If two obviously well separated peaks can be observed or if previous knowledge about two components exists, then the suggested procedure will lead to adequate estimation results. If more than two peaks appear, the presented methods may easily be adapted. Nevertheless, sometimes the histogram of a data set containing grouped observations will not reveal identifiable peaks or it is not clear how many components should be fitted. The problem of finding the correct number of components has not been investigated here. However, an algorithm dealing with this problem was proposed by Schlattmann (2009), who named it *Algorithm for Flexible Support Size*. His proceeding tests how many components are reasonable before those components are fitted. Further work on that topic where pushed forward by Furman and Lindsay (1994b),

who also dealt with the testing of the number of components. Nevertheless, those approaches are only available for individual observations. Hence, to extend them to grouped observations would be an interesting continuative project.

Furthermore, this work is limited to mixtures of Gaussian distributions. As noted earlier, the components of a mixture may consist of different distributions. For this purpose, Du (2002) extended the MIX algorithm to mixtures of Weibull, Poisson, Negative Binomial, Binomial, Gamma and Log-Normal distributions. An expansion of the here proposed methods to those distributions might enlarge their area of application. Moreover, a comparison between MIX and the other algorithms might help to decide, which method is the most suitable for a particular mixture model with non-Gaussian components.

A further restriction in this work is the choice of the interval widths. In some cases, particulary in censored or truncated data, different classifications may occur. For instance, the upper and the lower boundaries might be chosen as infinity or values may fall below the range of a measurement instrument, which is accompanied by one or two sided censoring. This should be treated in a different way as proposed here. An interested reader may find a solution in Efron (1967), who proposed an EM algorithm for single censored data or Turnbull (1976), who extended Efron's approach to arbitrary censored, grouped, and truncated data.

Finally, the problem of finding suitable values does not seem to be discussed to complete satisfaction. In fact, a technique was developed, which can be used for many mixture models, however, no technique is suitable for every data situation. Hence, the attempt to find a more convenient and applicable strategy should be pursued. In particular, the work of Furman and Lindsay (1994a) seems to be promising. They state that moment estimates are highly effective as starting values in maximizing mixture likelihoods. An attempt to extend this approach to grouped observations might lead to more accurate initial values.

# Acknowledgment

Während der Entstehung und Fertigstellung dieser Arbeit standen viele wichtige Personen an meiner Seite ohne die diese Arbeit nicht möglich gewesen wäre. Für ihre Unterstützung, Hilfe und Kritik möchte ich mich ganz herzlich bedanken.

Allen voran möchte ich meinem Betreuer Herrn Prof. Dr. Jürgen Timm dafür danken, dass er mir die Möglichkeit gegeben hat dieses Thema zu bearbeiten. Seine Ratschläge waren immer sehr hilfreich und haben mich während der gesamten Zeit unterstützt. Auch meinem Zweitgutachter Herrn Prof. Dr. Werner Brannath möchte ich dafür danken, dass er ohne zu Zögern sich bereit erklärt hat diese Arbeit zu begutachten und mir viele wichtige Anregungen und Hinweise gegeben hat.

Des Weiteren möchte ich allen Kollegen des KKSB für ihre Ideen und Ratschläge danken. Insbesondere möchte ich Sylvia Schmidt Danke sagen für ihre zahlreichen Vorschläge, die diese Arbeit bereichert haben und für die spannenden Go Partien, die für erholsame Mittagspausen gesorgt haben. Außerdem möchte ich mich bei Werner Wosniok bedanken, der mir eine große Hilfe beim Erstellen der Programme war.

Für das Korrekturlesen möchte ich mich ganz besonders bei Susann Henkel bedanken, sowie bei Elvira Pratsch und Paul Grathwohl. Sie alle haben mir viele wertvolle Tipps und Ratschläge gegeben.

Ein besondere Dank gilt meinem Verlobten Daniel Pratsch, der mir nicht nur bei technischen Fragen geholfen, mich unermüdlich motiviert und bedingungslos unterstützt hat, sondern mir auch in schwierigen Zeiten immer zur Seite stand. Insbesondere in der Endphase habe ich seine Geduld oft auf die Probe gestellt, doch er hat mir immer das Gefühl gegeben, dass wir gemeinsam alles schaffen.

Der größte Dank geht an meine Eltern, insbesondere meinen Vater, dem ich diese Arbeit gewidmet habe. Ohne ihn hätte ich nicht an mich geglaubt und daran, dass ich so eine Arbeit bewerkstelligen könnte. Ich möchte ihm dafür danken, dass er immer für mich da war und mich ermutigt hat an mich zu glauben, auch wenn er leider die Fertigstellung dieser Arbeit nicht miterleben durfte.

# List of Symbols

| | |
|---|---|
| $\Delta$ | Difference of the component means |
| E[X] | Expectation of X |
| $X, Y, Z$ | Random variables |
| $i$ | Index for the intervals |
| $j$ | Index for the mixture components |
| $s$ | Index for the individual observations |
| $t$ | Index for the iterations |
| $n$ | Number of observations |
| $k$ | Number of mixture components |
| $m$ | Number of intervals |
| $x_s$ | Individual observations |
| $n_i$ | Number of observations of the $i$th interval |
| $y_i$ | Grouped data |
| $a_0, \ldots, a_m$ | Interval borders |
| $\bar{a_i}$ | Midpoint of the $i$th interval |
| $\theta$ | Parameter of the density function f |
| $c$ | Number of components in $\theta$ |
| $\Omega$ | Parameter space for $\theta$ |
| $f(x, \theta)$ | Density function of a distribution with parameter $\theta$ |
| $g(x, \Psi)$ | Density function of a mixture distribution with parameter $\Psi$ |
| $\phi(x, \mu_j, \sigma_j^2)$ | Gaussian density function |
| $\Psi$ | Parameter of the mixture distribution |
| $\pi_j$ | Mixing proportions of the mixture |
| $\mu_j$ | Component means of the mixture |
| $\sigma_j$ | Component standard deviation of the mixture |
| $\mu_m$ | Mean of the mixture |
| $\sigma_m$ | Standard deviation of the mixture |
| $L(x, \theta)$ | Likelihood function |
| $S(\theta, x)$ | Score equation: $\partial \log L(\theta, x)/\partial\theta = 0$ |
| $I(x, \theta)$ | Information matrix |
| $\mathbf{I}(x, \theta)$ | Expected information matrix |
| $P_i(\theta)$ | Probability of an observation falling into the $i$ interval |
| $A^T$ | Transpose of a matrix A |
| $\lambda$ | Lagrange multiplier |

# List of Figures


2.1 Densities of two component Gaussian mixtures with equal variances $\sigma_1^2 = \sigma_2^2 = 1$ and different component means. . . . . 9
2.2 Densities of two component Gaussian mixtures with equal variances $\sigma_1^2 = \sigma_2^2 = 1$ and different mixing proportions. . . . . 10
2.3 Sample of a mixture. . . . . 21
2.4 Likelihood function for $\mu_2$. . . . . 21
2.5 Grouped observations. . . . . 22
2.6 A fitted Gaussian distribution to grouped data. . . . . 22

4.1 Investigated mixture models. First row: $\Psi = (0.5,\ 0.5,\ 5,\ \mu_2,\ 1,\ 1)$, second row: $\Psi = (0.7,\ 0.3,\ 5,\ \mu_2,\ 1,\ 1)$, third row: $\Psi = (0.9,\ 0.1,\ 5,\ \mu_2,\ 1,\ 1)$, where $\mu_2 = 6,\ 7,\ 8$ and 9, respectively. . . . . 47
4.2 Mixture densities with average estimated parameters for mixture model M1. 49
4.3 Mixture densities with average estimated parameters for mixture model M2. 52
4.4 Mixture densities with average estimated parameters for mixture model M3. 55
4.5 Mixture densities with average estimated parameters for mixture model M4. 57
4.6 Mixture densities with average estimated parameters for mixture model M5. 59
4.7 Mixture densities with average estimated parameters for mixture model M6. 62
4.8 Mixture densities with average estimated parameters for mixture model M7. 64
4.9 Mixture densities with average estimated parameters for mixture model M8. 66
4.10 Mixture densities with average estimated parameters for mixture model M9. 69
4.11 Mixture densities with average estimated parameters for mixture model M10. 71
4.12 Mixture densities with average estimated parameters for mixture model M11. 73
4.13 Mixture densities with average estimated parameters for mixture model M12. 75
4.14 Cumulated rRMSE (black line) and convergence failures (red line) for mixture model M4 with interval widths $h$ from 0.1 to 2.5. . . . . 90
4.15 Cumulated rRMSE (black line) and convergence failures (red line) for mixture model M9 with interval widths $h$ from 0.1 to 2.5. . . . . 92

5.1 The Old Faithful Geyser data . . . . . 96
5.2 Fitted densities to the Old Faithful Geyser data . . . . . 97
5.3 Histogram of the sodium data set . . . . . 98
5.4 Fitted densities and estimated reference limits for the sodium data set . . . 100

# List of Tables


4.1 Abbreviations and color coding of the methods. . . . . 46
4.2 The rRMSEs, cumulated rRMSEs, average runtimes in seconds, average number of iterations, and convergence failure rates in % for mixture model M1. . . . . 50
4.3 The rRMSEs, cumulated rRMSEs, average runtimes in seconds, average number of iterations, and convergence failure rates in % for mixture model M2. . . . . 53
4.4 The rRMSEs, cumulated rRMSEs, average runtimes in seconds, average number of iterations, and convergence failure rates in % for mixture model M3. . . . . 56
4.5 The rRMSEs, cumulated rRMSEs, average runtimes in seconds, average number of iterations, and convergence failure rates in % for mixture model M4. . . . . 58
4.6 The rRMSEs, cumulated rRMSEs, average runtimes in seconds, average number of iterations, and convergence failure rates in % for mixture model M5. . . . . 60
4.7 Cumulated rRMSEs separately listed for the components of mixture model M5. . . . . 61
4.8 The rRMSEs, cumulated rRMSEs, average runtimes in seconds, average number of iterations, and convergence failure rates in % for mixture model M6. . . . . 63
4.9 Cumulated rRMSEs separately listed for the components of mixture model M6. . . . . 63
4.10 The rRMSEs, cumulated rRMSEs, average runtimes in seconds, average number of iterations, and convergence failure rates in % for mixture model M7. . . . . 65
4.11 Cumulated rRMSEs separately listed for the components of mixture model M7. . . . . 65
4.12 The rRMSEs, cumulated rRMSEs, average runtimes in seconds, average number of iterations, and convergence failure rates in % for mixture model M8. . . . . 67
4.13 Cumulated rRMSEs separately listed for the components of mixture model M8. . . . . 67

4.14 The rRMSEs, cumulated rRMSEs, average runtimes in seconds, average number of iterations, and convergence failure rates in % for mixture model M9. . . . . 69
4.15 Cumulated rRMSEs separately listed for the components of mixture model M9. . . . . 70
4.16 The rRMSEs, cumulated rRMSEs, average runtimes in seconds, average number of iterations, and convergence failure rates in % for mixture model M10. . . . . 72
4.17 Cumulated rRMSEs separately listed for the components of mixture model M10. . . . . 72
4.18 The rRMSEs, cumulated rRMSEs, average runtimes in seconds, average number of iterations, and convergence failure rates in % for mixture model M11. . . . . 74
4.19 Cumulated rRMSEs separately listed for the components of mixture model M11. . . . . 74
4.20 The rRMSEs, cumulated rRMSEs, average runtimes in seconds, average number of iterations, and convergence failure rates in % for mixture model M12. . . . . 76
4.21 Cumulated rRMSEs separately listed for the components of mixture model M12. . . . . 76
4.22 Convergence failure rates in % for $n = 300$. . . . . 79
4.23 Cumulated rRMSEs for sample size $n = 300$. . . . . 80
4.24 Convergence failure rates in % for $n = 500$. . . . . 81
4.25 Cumulated rRMSEs for sample size $n = 500$. . . . . 81
4.26 Number of failures in the same 1000 samples for starting values S0 - S5. . . 83
4.27 Cumulated rRMSEs for starting values S0 - S5. . . . . 85
4.28 Recommended techniques for obtaining starting values dependent on the shape of the histogram. . . . . 88
4.29 Relative interval width $H$ depending on the underlying mixture model. . . . 93
4.30 Mixture Models listed dependent on the distribution shape. . . . . 94

5.1 Estimation results for the Old Faithful data set . . . . . 97
5.2 Parameter estimations for the sodium data set . . . . . 99

A.1 Average mean and standard deviation for mixture models M1-M12. . . . . 113
A.2 Cumulated rRMSE for different interval width. . . . . 115

# A Appendix

This appendix contains all numerical results. The first table contains the mean estimations and mean standard deviation for each parameter obtained by the six methods for each considered mixture model. These results were the base of the graphical illustrations in section 4.2, 4.3, and 4.4.

The second table displays the cumulated rRMSE for each mixture model and each method depended on the different interval widths. The best values and those close to the best are color-coded.

## A.1 Estimation Results

Table A.1: Average mean and standard deviation for mixture models M1-M12.

| | Method | $\pi_1$ | $\mu_1$ | $\mu_2$ | $\sigma_1$ | $\sigma_2$ |
|---|---|---|---|---|---|---|
| | MIX | 0.492 (0.324) | 4.594 (0.957) | 6.346 (0.878) | 0.726 (0.388) | 0.733 (0.371) |
| | EMNR | 0.492 (0.221) | 4.807 (0.615) | 6.167 (0.595) | 0.857 (0.283) | 0.824 (0.310) |
| M1 | APEM | 0.498 (0.037) | 4.927 (0.190) | 6.075 (0.193) | 0.976 (0.167) | 0.970 (0.162) |
| | RMEM | 0.498 (0.037) | 4.927 (0.190) | 6.075 (0.193) | 0.976 (0.167) | 0.970 (0.162) |
| | UDEM | 0.498 (0.052) | 4.912 (0.206) | 6.094 (0.209) | 1.004 (0.184) | 0.995 (0.179) |
| | MGT | 0.499 (0.040) | 4.947 (0.193) | 6.057 (0.202) | 0.941 (0.180) | 0.931 (0.179) |
| | MIX | 0.492 (0.263) | 4.594 (0.781) | 6.346 (0.717) | 0.726 (0.355) | 0.733 (0.343) |
| | EMNR | 0.492 (0.207) | 4.807 (0.561) | 6.167 (0.553) | 0.857 (0.267) | 0.824 (0.308) |
| M2 | APEM | 0.498 (0.073) | 4.927 (0.229) | 6.075 (0.232) | 0.976 (0.169) | 0.970 (0.155) |
| | RMEM | 0.498 (0.073) | 4.927 (0.229) | 6.075 (0.232) | 0.976 (0.169) | 0.970 (0.155) |
| | UDEM | 0.498 (0.074) | 4.912 (0.233) | 6.094 (0.241) | 1.004 (0.171) | 0.995 (0.165) |
| | MGT | 0.499 (0.072) | 4.947 (0.228) | 6.057 (0.229) | 0.941 (0.179) | 0.931 (0.165) |
| | MIX | 0.495 (0.126) | 4.987 (0.384) | 7.992 (0.388) | 0.961 (0.240) | 0.973 (0.243) |
| | EMNR | 0.497 (0.114) | 4.994 (0.347) | 8.000 (0.346) | 0.971 (0.212) | 0.969 (0.222) |
| M3 | APEM | 0.496 (0.076) | 4.994 (0.239) | 7.991 (0.238) | 1.020 (0.158) | 1.027 (0.159) |
| | RMEM | 0.495 (0.076) | 4.987 (0.239) | 7.993 (0.238) | 1.013 (0.158) | 1.023 (0.159) |
| | UDEM | 0.497 (0.078) | 4.988 (0.255) | 7.999 (0.245) | 1.049 (0.166) | 1.054 (0.171) |
| | MGT | 0.495 (0.076) | 4.989 (0.240) | 7.992 (0.239) | 0.971 (0.167) | 0.981 (0.168) |
| | MIX | 0.501 (0.060) | 5.009 (0.201) | 9.000 (0.206) | 0.978 (0.162) | 0.971 (0.160) |

Table A.1: (continued)

| | Method | $\pi_1$ | $\mu_1$ | $\mu_2$ | $\sigma_1$ | $\sigma_2$ |
|---|---|---|---|---|---|---|
| | EMNR | 0.502 (0.059) | 5.012 (0.195) | 9.004 (0.198) | 0.982 (0.154) | 0.967 (0.153) |
| M4 | APEM | 0.501 (0.055) | 5.008 (0.177) | 9.000 (0.180) | 1.021 (0.137) | 1.014 (0.133) |
| | RMEM | 0.501 (0.055) | 5.007 (0.177) | 9.002 (0.180) | 1.020 (0.137) | 1.012 (0.133) |
| | UDEM | 0.501 (0.059) | 5.007 (0.187) | 8.997 (0.198) | 1.058 (0.148) | 1.055 (0.143) |
| | MGT | 0.501 (0.056) | 5.007 (0.178) | 9.002 (0.182) | 0.977 (0.144) | 0.970 (0.140) |
| | MIX | 0.609 (0.315) | 4.673 (0.747) | 6.393 (1.008) | 0.792 (0.337) | 0.637 (0.427) |
| | EMNR | 0.667 (0.211) | 4.863 (0.460) | 6.233 (0.711) | 0.888 (0.236) | 0.742 (0.373) |
| M5 | APEM | 0.687 (0.047) | 4.943 (0.161) | 6.078 (0.249) | 0.992 (0.137) | 0.949 (0.214) |
| | RMEM | 0.687 (0.047) | 4.943 (0.161) | 6.078 (0.249) | 0.992 (0.137) | 0.949 (0.214) |
| | UDEM | 0.687 (0.059) | 4.929 (0.191) | 6.108 (0.282) | 1.022 (0.147) | 0.956 (0.245) |
| | MGT | 0.688 (0.049) | 4.954 (0.168) | 6.054 (0.252) | 0.956 (0.145) | 0.909 (0.230) |
| | MIX | 0.674 (0.236) | 4.937 (0.476) | 7.219 (0.891) | 0.928 (0.259) | 0.784 (0.406) |
| | EMNR | 0.702 (0.186) | 5.000 (0.357) | 7.172 (0.733) | 0.970 (0.212) | 0.820 (0.391) |
| M6 | APEM | 0.699 (0.063) | 4.988 (0.164) | 7.013 (0.274) | 1.016 (0.130) | 0.993 (0.212) |
| | RMEM | 0.698 (0.063) | 4.988 (0.164) | 7.013 (0.273) | 1.016 (0.130) | 0.994 (0.210) |
| | UDEM | 0.701 (0.066) | 4.987 (0.172) | 7.038 (0.303) | 1.050 (0.137) | 1.010 (0.238) |
| | MGT | 0.700 (0.063) | 4.994 (0.162) | 7.011 (0.276) | 0.977 (0.137) | 0.951 (0.227) |
| | MIX | 0.678 (0.122) | 4.966 (0.262) | 7.947 (0.524) | 0.962 (0.181) | 0.965 (0.303) |
| | EMNR | 0.686 (0.112) | 4.978 (0.251) | 7.975 (0.486) | 0.971 (0.172) | 0.949 (0.295) |
| M7 | APEM | 0.699 (0.070) | 4.993 (0.175) | 8.007 (0.304) | 1.023 (0.127) | 0.995 (0.204) |
| | RMEM | 0.699 (0.070) | 4.993 (0.175) | 8.007 (0.304) | 1.023 (0.127) | 0.995 (0.204) |
| | UMEM | 0.699 (0.072) | 4.994 (0.188) | 8.014 (0.328) | 1.061 (0.134) | 1.023 (0.224) |
| | MGT | 0.700 (0.070) | 4.996 (0.178) | 8.008 (0.304) | 0.983 (0.134) | 0.949 (0.217) |
| | MIX | 0.697 (0.061) | 5.001 (0.172) | 8.988 (0.303) | 0.985 (0.133) | 0.969 (0.232) |
| | EMNR | 0.699 (0.058) | 5.004 (0.167) | 8.996 (0.281) | 0.988 (0.128) | 0.962 (0.223) |
| M8 | APEM | 0.700 (0.053) | 5.004 (0.154) | 9.002 (0.241) | 1.030 (0.113) | 1.006 (0.185) |
| | RMEM | 0.700 (0.053) | 5.004 (0.154) | 9.002 (0.241) | 1.030 (0.113) | 1.006 (0.185) |
| | UDEM | 0.701 (0.054) | 5.007 (0.160) | 9.009 (0.266) | 1.069 (0.116) | 1.033 (0.200) |
| | MGT | 0.700 (0.053) | 5.005 (0.155) | 9.002 (0.243) | 0.988 (0.118) | 0.962 (0.196) |
| | MIX | 0.704 (0.305) | 4.630 (0.583) | 6.377 (1.031) | 0.794 (0.285) | 0.547 (0.409) |
| | EMNR | 0.810 (0.186) | 4.827 (0.378) | 6.368 (0.812) | 0.894 (0.189) | 0.582 (0.408) |
| M9 | APEM | 0.873 (0.056) | 4.941 (0.149) | 6.189 (0.364) | 0.988 (0.106) | 0.898 (0.304) |
| | RMEM | 0.873 (0.055) | 4.940 (0.149) | 6.193 (0.366) | 0.987 (0.106) | 0.895 (0.308) |
| | UDEM | 0.867 (0.069) | 4.923 (0.156) | 6.246 (0.413) | 1.022 (0.114) | 0.835 (0.352) |
| | MGT | 0.868 (0.055) | 4.941 (0.146) | 6.135 (0.380) | 0.951 (0.110) | 0.814 (0.334) |

Table A.1: (continued)

| | Method | $\pi_1$ | $\mu_1$ | $\mu_2$ | $\sigma_1$ | $\sigma_2$ |
|---|---|---|---|---|---|---|
| | MIX | 0.794 (0.241) | 4.836 (0.387) | 7.182 (1.072) | 0.894 (0.221) | 0.606 (0.456) |
| | EMNR | 0.865 (0.151) | 4.946 (0.265) | 7.278 (0.936) | 0.953 (0.155) | 0.602 (0.442) |
| M10 | APEM | 0.889 (0.046) | 4.970 (0.139) | 7.038 (0.382) | 1.011 (0.097) | 0.923 (0.316) |
| | RMEM | 0.889 (0.046) | 4.970 (0.139) | 7.038 (0.382) | 1.011 (0.097) | 0.922 (0.318) |
| | UDEM | 0.887 (0.057) | 4.962 (0.154) | 7.053 (0.421) | 1.047 (0.109) | 0.907 (0.350) |
| | MGT | 0.889 (0.051) | 4.970 (0.141) | 7.024 (0.418) | 0.970 (0.106) | 0.855 (0.347) |
| | MIX | 0.850 (0.134) | 4.946 (0.212) | 7.908 (0.917) | 0.952 (0.158) | 0.837 (0.469) |
| | EMNR | 0.877 (0.106) | 4.979 (0.190) | 8.040 (0.805) | 0.978 (0.136) | 0.755 (0.455) |
| M11 | APEM | 0.895 (0.053) | 4.994 (0.139) | 8.030 (0.487) | 1.028 (0.103) | 0.926 (0.318) |
| | RMEM | 0.895 (0.053) | 4.994 (0.139) | 8.030 (0.487) | 1.028 (0.103) | 0.926 (0.318) |
| | UDEM | 0.897 (0.055) | 4.996 (0.146) | 8.056 (0.540) | 1.068 (0.108) | 0.916 (0.376) |
| | MGT | 0.897 (0.053) | 4.997 (0.139) | 8.040 (0.485) | 0.989 (0.107) | 0.849 (0.350) |
| | MIX | 0.886 (0.057) | 4.984 (0.133) | 8.880 (0.697) | 0.975 (0.109) | 0.932 (0.426) |
| | EMNR | 0.893 (0.051) | 4.992 (0.128) | 8.923 (0.627) | 0.983 (0.105) | 0.875 (0.425) |
| M12 | APEM | 0.899 (0.036) | 4.998 (0.119) | 8.979 (0.455) | 1.029 (0.090) | 0.946 (0.308) |
| | RMEM | 0.899 (0.036) | 4.998 (0.119) | 8.979 (0.455) | 1.029 (0.090) | 0.946 (0.308) |
| | UDEM | 0.899 (0.039) | 5.000 (0.128) | 8.990 (0.505) | 1.070 (0.096) | 0.957 (0.335) |
| | MGT | 0.899 (0.036) | 4.998 (0.119) | 8.980 (0.459) | 0.988 (0.095) | 0.885 (0.334) |

## A.2 Cumulated rRMSE for Different Interval Widths

Table A.2: Cumulated rRMSE for different interval width.

| Model | Width | MIX | EMNR | APEM | RMEM | UDEM | MGT |
|---|---|---|---|---|---|---|---|
| M1 | 0.1 | 1.93 | 0.82 | 0.50 | 0.55 | 0.56 | 0.45 |
| | 0.2 | 1.90 | 0.93 | 0.43 | 0.56 | 0.55 | 0.47 |
| | 0.5 | 1.87 | 1.13 | 0.45 | 0.53 | 0.56 | 0.50 |
| | 1.0 | 1.95 | 1.35 | 0.45 | 0.48 | 0.51 | 0.54 |
| | 1.1 | 3.48 | 1.56 | 0.46 | 0.49 | 0.56 | 0.99 |
| | 1.2 | 3.62 | 1.89 | 0.52 | 0.55 | 0.58 | 1.25 |
| | 1.3 | 3.58 | 1.71 | 0.54 | 0.56 | 0.58 | 1.46 |
| | 1.4 | 3.00 | 1.65 | 0.58 | 0.58 | 0.60 | 1.42 |
| | 1.5 | 2.55 | 1.62 | 0.53 | 0.53 | 0.62 | 1.56 |
| | 1.6 | 2.56 | 1.43 | 0.67 | 0.64 | 0.68 | 1.76 |

Table A.2: (continued)

| Model | Width | MIX | EMNR | APEM | RMEM | UDEM | MGT |
|---|---|---|---|---|---|---|---|
| | 1.7 | 3.25 | 1.36 | 0.84 | 0.77 | 0.67 | 1.84 |
| | 1.8 | 2.81 | 1.48 | 0.91 | 0.88 | 0.73 | 2.62 |
| | 1.9 | 2.34 | 1.34 | 1.01 | 0.98 | 0.94 | 2.72 |
| | 2.0 | 1.79 | 1.35 | 0.98 | 0.92 | 1.07 | 2.53 |
| | 2.1 | 1.68 | 1.56 | 0.78 | 0.87 | 1.13 | 2.64 |
| | 2.2 | 1.74 | 1.16 | 0.74 | 0.90 | 1.16 | 2.37 |
| | 2.3 | 1.76 | 1.17 | 0.80 | 0.86 | 1.26 | 2.30 |
| | 2.4 | 1.50 | 1.45 | 0.90 | 1.16 | 1.29 | 2.75 |
| | 2.5 | 1.59 | 1.39 | 1.03 | 1.29 | 1.29 | 2.79 |
| M2 | 0.1 | 1.50 | 0.77 | 0.50 | 0.54 | 0.59 | 0.45 |
| | 0.2 | 1.49 | 0.88 | 0.45 | 0.54 | 0.43 | 0.47 |
| | 0.5 | 1.46 | 1.02 | 0.49 | 0.55 | 0.43 | 0.51 |
| | 1.0 | 1.54 | 1.20 | 0.55 | 0.55 | 0.50 | 0.57 |
| | 1.1 | 1.88 | 1.29 | 0.53 | 0.53 | 0.50 | 0.80 |
| | 1.2 | 2.20 | 1.35 | 0.51 | 0.50 | 0.55 | 1.10 |
| | 1.3 | 2.27 | 1.30 | 0.51 | 0.50 | 0.57 | 0.96 |
| | 1.4 | 2.16 | 1.36 | 0.53 | 0.50 | 0.58 | 1.16 |
| | 1.5 | 2.61 | 1.35 | 0.51 | 0.49 | 0.63 | 1.36 |
| | 1.6 | 2.96 | 1.35 | 0.53 | 0.52 | 0.70 | 1.46 |
| | 1.7 | 2.38 | 1.28 | 0.64 | 0.60 | 0.82 | 1.71 |
| | 1.8 | 2.31 | 1.14 | 0.72 | 0.65 | 0.82 | 2.00 |
| | 1.9 | 2.30 | 1.17 | 0.68 | 0.64 | 0.86 | 1.91 |
| | 2.0 | 1.73 | 1.22 | 0.64 | 0.64 | 0.95 | 1.81 |
| | 2.1 | 1.86 | 1.12 | 0.74 | 0.74 | 0.87 | 3.13 |
| | 2.2 | 1.38 | 1.09 | 0.94 | 0.98 | 0.99 | 3.20 |
| | 2.3 | 1.31 | 1.08 | 1.09 | 1.17 | 1.19 | 2.95 |
| | 2.4 | 1.23 | 1.00 | 1.38 | 1.12 | 1.22 | 2.69 |
| | 2.5 | 1.36 | 0.99 | 0.94 | 0.96 | 1.32 | 2.23 |
| M3 | 0.1 | 0.73 | 0.63 | 0.50 | 0.54 | 0.59 | 0.49 |
| | 0.2 | 0.72 | 0.66 | 0.49 | 0.55 | 0.42 | 0.50 |
| | 0.5 | 0.75 | 0.72 | 0.52 | 0.55 | 0.43 | 0.53 |
| | 1.0 | 0.79 | 0.78 | 0.55 | 0.55 | 0.49 | 0.57 |
| | 1.1 | 0.92 | 0.86 | 0.55 | 0.55 | 0.49 | 0.70 |
| | 1.2 | 0.84 | 0.83 | 0.55 | 0.54 | 0.53 | 0.94 |
| | 1.3 | 0.92 | 0.90 | 0.56 | 0.54 | 0.58 | 0.98 |
| | 1.4 | 1.00 | 0.94 | 0.57 | 0.57 | 0.58 | 1.03 |
| | 1.5 | 0.99 | 0.93 | 0.55 | 0.53 | 0.61 | 0.96 |

Table A.2: (continued)

| Model | Width | MIX | EMNR | APEM | RMEM | UDEM | MGT |
|---|---|---|---|---|---|---|---|
| | 1.6 | 1.13 | 1.02 | 0.55 | 0.54 | 0.68 | 1.09 |
| | 1.7 | 1.34 | 1.11 | 0.55 | 0.53 | 0.82 | 1.32 |
| | 1.8 | 1.28 | 1.08 | 0.51 | 0.49 | 0.79 | 1.22 |
| | 1.9 | 1.34 | 1.04 | 0.58 | 0.56 | 0.87 | 1.40 |
| | 2.0 | 1.43 | 1.00 | 0.64 | 0.62 | 0.95 | 1.17 |
| | 2.1 | 1.32 | 0.98 | 0.64 | 0.64 | 1.13 | 1.72 |
| | 2.2 | 2.36 | 1.06 | 0.56 | 0.54 | 1.13 | 1.83 |
| | 2.3 | 1.11 | 0.98 | 0.52 | 0.52 | 1.20 | 3.31 |
| | 2.4 | 1.04 | 0.88 | 0.72 | 0.65 | 1.24 | 3.62 |
| | 2.5 | 1.12 | 0.92 | 0.77 | 0.84 | 1.27 | 2.95 |
| M4 | 0.1 | 0.46 | 0.43 | 0.50 | 0.42 | 0.42 | 0.41 |
| | 0.2 | 0.46 | 0.43 | 0.41 | 0.42 | 0.42 | 0.42 |
| | 0.5 | 0.47 | 0.45 | 0.42 | 0.42 | 0.43 | 0.43 |
| | 1.0 | 0.51 | 0.49 | 0.44 | 0.44 | 0.49 | 0.46 |
| | 1.1 | 0.52 | 0.50 | 0.45 | 0.45 | 0.49 | 0.52 |
| | 1.2 | 0.55 | 0.53 | 0.46 | 0.46 | 0.54 | 0.65 |
| | 1.3 | 0.57 | 0.53 | 0.45 | 0.45 | 0.57 | 0.87 |
| | 1.4 | 0.68 | 0.56 | 0.48 | 0.47 | 0.58 | 0.81 |
| | 1.5 | 0.87 | 0.68 | 0.52 | 0.51 | 0.62 | 0.80 |
| | 1.6 | 0.87 | 0.64 | 0.51 | 0.51 | 0.69 | 0.97 |
| | 1.7 | 0.89 | 0.66 | 0.50 | 0.49 | 0.79 | 1.19 |
| | 1.8 | 1.07 | 0.78 | 0.57 | 0.56 | 0.79 | 1.00 |
| | 1.9 | 1.25 | 0.86 | 0.57 | 0.55 | 0.87 | 1.37 |
| | 2.0 | 1.19 | 0.84 | 0.59 | 0.57 | 0.93 | 1.15 |
| | 2.1 | 1.22 | 0.80 | 0.57 | 0.58 | 1.11 | 1.48 |
| | 2.2 | 1.31 | 0.79 | 0.53 | 0.54 | 1.32 | 1.24 |
| | 2.3 | 1.15 | 0.78 | 0.58 | 0.57 | 1.46 | 1.84 |
| | 2.4 | 1.23 | 0.85 | 0.69 | 0.69 | 1.47 | 1.89 |
| | 2.5 | 1.25 | 0.81 | 0.79 | 0.78 | 1.42 | 1.99 |
| M5 | 0.1 | 1.69 | 0.78 | 0.50 | 0.54 | 0.59 | 0.46 |
| | 0.2 | 1.69 | 0.89 | 0.46 | 0.55 | 0.56 | 0.49 |
| | 0.5 | 1.66 | 1.05 | 0.50 | 0.55 | 0.56 | 0.54 |
| | 1.0 | 1.77 | 1.24 | 0.50 | 0.50 | 0.57 | 0.57 |
| | 1.1 | 2.86 | 1.71 | 0.50 | 0.50 | 0.57 | 0.95 |
| | 1.2 | 3.76 | 1.70 | 0.55 | 0.54 | 0.58 | 1.09 |
| | 1.3 | 3.88 | 1.55 | 0.59 | 0.57 | 0.57 | 1.39 |
| | 1.4 | 2.64 | 1.78 | 0.56 | 0.54 | 0.61 | 1.23 |

Table A.2: (continued)

| Model | Width | MIX | EMNR | APEM | RMEM | UDEM | MGT |
|---|---|---|---|---|---|---|---|
| | 1.5 | 3.82 | 1.40 | 0.54 | 0.50 | 0.64 | 1.37 |
| | 1.6 | 2.86 | 1.57 | 0.77 | 0.65 | 0.69 | 1.69 |
| | 1.7 | 3.68 | 1.62 | 0.95 | 0.80 | 0.72 | 1.86 |
| | 1.8 | 3.88 | 1.63 | 1.02 | 0.93 | 0.78 | 2.45 |
| | 1.9 | 1.98 | 1.36 | 1.01 | 0.94 | 0.90 | 2.39 |
| | 2.0 | 1.82 | 1.16 | 0.96 | 0.88 | 0.99 | 2.09 |
| | 2.1 | 1.44 | 0.97 | 0.74 | 0.73 | 1.09 | 2.13 |
| | 2.2 | 1.48 | 0.99 | 0.67 | 0.67 | 1.08 | 1.97 |
| | 2.3 | 1.55 | 1.01 | 0.78 | 0.84 | 1.19 | 1.93 |
| | 2.4 | 1.22 | 0.82 | 1.23 | 1.17 | 1.26 | 2.30 |
| | 2.5 | 1.31 | 0.84 | 1.14 | 1.59 | 1.29 | 2.35 |
| M6 | 0.1 | 0.43 | 0.73 | 0.50 | 0.52 | 0.59 | 0.49 |
| | 0.2 | 1.69 | 0.82 | 0.44 | 0.52 | 0.52 | 0.45 |
| | 0.5 | 1.66 | 0.92 | 0.48 | 0.52 | 0.53 | 0.49 |
| | 1.0 | 1.74 | 1.09 | 0.51 | 0.50 | 0.55 | 0.53 |
| | 1.1 | 1.82 | 1.27 | 0.50 | 0.49 | 0.56 | 0.78 |
| | 1.2 | 3.28 | 1.43 | 0.48 | 0.46 | 0.57 | 1.06 |
| | 1.3 | 2.51 | 1.65 | 0.50 | 0.49 | 0.58 | 1.00 |
| | 1.4 | 2.11 | 1.32 | 0.49 | 0.48 | 0.61 | 1.07 |
| | 1.5 | 2.30 | 1.28 | 0.49 | 0.48 | 0.59 | 1.29 |
| | 1.6 | 1.92 | 1.54 | 0.49 | 0.48 | 0.66 | 1.24 |
| | 1.7 | 1.99 | 1.36 | 0.57 | 0.53 | 0.69 | 1.38 |
| | 1.8 | 2.58 | 1.24 | 0.80 | 0.72 | 0.68 | 1.60 |
| | 1.9 | 1.78 | 1.30 | 0.78 | 0.76 | 0.74 | 1.75 |
| | 2.0 | 1.68 | 1.34 | 0.71 | 0.70 | 0.79 | 1.84 |
| | 2.1 | 1.45 | 1.23 | 0.69 | 0.69 | 0.98 | 2.60 |
| | 2.2 | 1.32 | 0.97 | 0.85 | 0.84 | 1.06 | 2.59 |
| | 2.3 | 1.30 | 0.97 | 1.00 | 1.03 | 1.14 | 2.48 |
| | 2.4 | 1.08 | 0.82 | 1.29 | 0.96 | 1.21 | 2.30 |
| | 2.5 | 1.45 | 0.85 | 0.91 | 0.87 | 1.24 | 2.06 |
| M7 | 0.1 | 0.72 | 0.58 | 0.50 | 0.50 | 0.59 | 0.45 |
| | 0.2 | 0.73 | 0.60 | 0.46 | 0.50 | 0.50 | 0.47 |
| | 0.5 | 0.74 | 0.65 | 0.47 | 0.50 | 0.51 | 0.49 |
| | 1.0 | 0.79 | 0.75 | 0.51 | 0.51 | 0.56 | 0.53 |
| | 1.1 | 0.83 | 0.79 | 0.51 | 0.51 | 0.57 | 0.65 |
| | 1.2 | 0.90 | 0.83 | 0.50 | 0.50 | 0.60 | 0.89 |
| | 1.3 | 1.03 | 0.92 | 0.52 | 0.51 | 0.61 | 0.93 |

Table A.2: (continued)

| Model | Width | MIX | EMNR | APEM | RMEM | UDEM | MGT |
|---|---|---|---|---|---|---|---|
| | 1.4 | 1.00 | 0.92 | 0.53 | 0.51 | 0.62 | 0.97 |
| | 1.5 | 1.05 | 0.99 | 0.52 | 0.50 | 0.72 | 0.91 |
| | 1.6 | 1.13 | 1.00 | 0.51 | 0.49 | 0.74 | 0.99 |
| | 1.7 | 1.30 | 1.07 | 0.50 | 0.48 | 0.77 | 1.25 |
| | 1.8 | 1.30 | 1.05 | 0.49 | 0.49 | 0.77 | 1.13 |
| | 1.9 | 1.36 | 1.09 | 0.58 | 0.55 | 0.85 | 1.20 |
| | 2.0 | 1.44 | 1.02 | 0.64 | 0.62 | 0.94 | 1.03 |
| | 2.1 | 1.32 | 1.01 | 0.56 | 0.56 | 1.07 | 1.62 |
| | 2.2 | 1.38 | 1.03 | 0.50 | 0.49 | 1.10 | 1.77 |
| | 2.3 | 1.03 | 0.88 | 0.52 | 0.53 | 1.23 | 2.61 |
| | 2.4 | 1.01 | 0.82 | 0.70 | 0.63 | 1.29 | 2.85 |
| | 2.5 | 1.04 | 0.84 | 0.69 | 0.74 | 1.36 | 2.43 |
| M8 | 0.1 | 0.46 | 0.42 | 0.50 | 0.41 | 0.59 | 0.40 |
| | 0.2 | 0.46 | 0.43 | 0.40 | 0.41 | 0.41 | 0.41 |
| | 0.5 | 0.47 | 0.45 | 0.40 | 0.41 | 0.41 | 0.42 |
| | 1.0 | 0.52 | 0.50 | 0.43 | 0.43 | 0.49 | 0.45 |
| | 1.1 | 0.52 | 0.50 | 0.44 | 0.44 | 0.50 | 0.52 |
| | 1.2 | 0.56 | 0.56 | 0.44 | 0.44 | 0.50 | 0.68 |
| | 1.3 | 0.61 | 0.58 | 0.44 | 0.44 | 0.56 | 0.88 |
| | 1.4 | 0.65 | 0.62 | 0.47 | 0.46 | 0.58 | 0.84 |
| | 1.5 | 0.77 | 0.70 | 0.49 | 0.48 | 0.58 | 0.83 |
| | 1.6 | 0.77 | 0.67 | 0.51 | 0.50 | 0.67 | 0.99 |
| | 1.7 | 0.81 | 0.73 | 0.49 | 0.48 | 0.80 | 1.20 |
| | 1.8 | 0.97 | 0.81 | 0.51 | 0.50 | 0.83 | 1.01 |
| | 1.9 | 1.22 | 0.92 | 0.55 | 0.54 | 0.82 | 1.42 |
| | 2.0 | 1.20 | 0.94 | 0.57 | 0.56 | 0.86 | 1.22 |
| | 2.1 | 1.23 | 0.87 | 0.59 | 0.57 | 0.96 | 1.47 |
| | 2.2 | 1.28 | 0.82 | 0.55 | 0.53 | 1.14 | 1.23 |
| | 2.3 | 1.18 | 0.83 | 0.56 | 0.56 | 1.28 | 1.70 |
| | 2.4 | 1.14 | 0.85 | 0.63 | 0.63 | 1.38 | 1.81 |
| | 2.5 | 1.18 | 0.82 | 0.74 | 0.73 | 1.48 | 1.92 |
| M9 | 0.1 | 1.93 | 0.43 | 0.50 | 0.66 | 0.56 | 0.49 |
| | 0.2 | 1.58 | 0.96 | 0.55 | 0.66 | 0.66 | 0.56 |
| | 0.5 | 1.61 | 1.08 | 0.58 | 0.64 | 0.65 | 0.61 |
| | 1.0 | 1.69 | 1.26 | 0.60 | 0.60 | 0.69 | 0.67 |
| | 1.1 | 2.46 | 1.84 | 0.54 | 0.54 | 0.68 | 1.02 |
| | 1.2 | 2.49 | 1.77 | 0.62 | 0.60 | 0.74 | 1.12 |

Table A.2: (continued)

| Model | Width | MIX | EMNR | APEM | RMEM | UDEM | MGT |
|---|---|---|---|---|---|---|---|
| | 1.3 | 2.64 | 1.72 | 0.65 | 0.62 | 0.74 | 1.43 |
| | 1.4 | 2.56 | 1.81 | 0.59 | 0.58 | 0.77 | 1.24 |
| | 1.5 | 2.72 | 1.95 | 0.53 | 0.57 | 0.81 | 1.24 |
| | 1.6 | 2.38 | 1.64 | 0.56 | 0.55 | 0.86 | 1.75 |
| | 1.7 | 2.48 | 1.85 | 0.77 | 0.67 | 0.92 | 1.94 |
| | 1.8 | 2.04 | 1.54 | 0.87 | 0.78 | 0.93 | 2.11 |
| | 1.9 | 2.37 | 1.20 | 0.88 | 0.79 | 0.96 | 1.79 |
| | 2.0 | 1.83 | 1.14 | 0.75 | 0.67 | 1.03 | 1.50 |
| | 2.1 | 1.50 | 1.01 | 0.63 | 0.66 | 1.12 | 1.91 |
| | 2.2 | 1.52 | 1.04 | 0.61 | 0.58 | 1.17 | 1.50 |
| | 2.3 | 1.53 | 1.07 | 0.66 | 0.68 | 1.24 | 1.90 |
| | 2.4 | 1.21 | 0.94 | 0.82 | 1.02 | 1.34 | 2.17 |
| | 2.5 | 1.22 | 0.91 | 0.98 | 1.88 | 1.43 | 2.20 |
| M10 | 0.1 | 1.34 | 0.81 | 0.50 | 0.60 | 0.59 | 0.53 |
| | 0.2 | 1.34 | 0.90 | 0.53 | 0.60 | 0.60 | 0.54 |
| | 0.5 | 1.36 | 1.00 | 0.56 | 0.60 | 0.61 | 0.59 |
| | 1.0 | 1.44 | 1.12 | 0.56 | 0.56 | 0.64 | 0.63 |
| | 1.1 | 1.92 | 1.58 | 0.53 | 0.54 | 0.64 | 0.89 |
| | 1.2 | 3.63 | 1.73 | 0.55 | 0.55 | 0.67 | 1.11 |
| | 1.3 | 1.87 | 1.47 | 0.56 | 0.56 | 0.65 | 1.05 |
| | 1.4 | 1.88 | 1.73 | 0.60 | 0.58 | 0.71 | 1.14 |
| | 1.5 | 1.80 | 1.46 | 0.57 | 0.58 | 0.72 | 1.24 |
| | 1.6 | 1.66 | 1.37 | 0.51 | 0.52 | 0.77 | 1.16 |
| | 1.7 | 1.85 | 1.57 | 0.57 | 0.54 | 0.79 | 1.32 |
| | 1.8 | 1.87 | 1.86 | 0.88 | 0.79 | 0.78 | 1.70 |
| | 1.9 | 1.88 | 1.88 | 0.94 | 0.92 | 0.83 | 1.73 |
| | 2.0 | 1.77 | 1.56 | 0.80 | 0.80 | 0.90 | 1.68 |
| | 2.1 | 1.35 | 1.28 | 0.72 | 0.73 | 1.07 | 1.89 |
| | 2.2 | 1.33 | 1.05 | 0.85 | 0.86 | 1.16 | 1.76 |
| | 2.3 | 1.43 | 1.05 | 0.97 | 1.00 | 1.20 | 1.61 |
| | 2.4 | 1.68 | 0.84 | 1.18 | 0.95 | 1.27 | 1.80 |
| | 2.5 | 1.03 | 0.83 | 0.93 | 0.93 | 1.32 | 1.79 |
| M11 | 0.1 | 0.88 | 0.67 | 0.50 | 0.60 | 0.60 | 0.56 |
| | 0.2 | 0.89 | 0.71 | 0.56 | 0.60 | 0.60 | 0.58 |
| | 0.5 | 0.91 | 0.78 | 0.57 | 0.58 | 0.60 | 0.60 |
| | 1.0 | 1.00 | 0.90 | 0.59 | 0.58 | 0.66 | 0.65 |
| | 1.1 | 1.17 | 1.01 | 0.57 | 0.56 | 0.66 | 0.79 |

Table A.2: (continued)

| Model | Width | MIX | EMNR | APEM | RMEM | UDEM | MGT |
|---|---|---|---|---|---|---|---|
| | 1.2 | 1.68 | 1.39 | 0.58 | 0.58 | 0.70 | 0.98 |
| | 1.3 | 1.51 | 1.27 | 0.57 | 0.55 | 0.70 | 0.96 |
| | 1.4 | 1.49 | 1.28 | 0.58 | 0.56 | 0.71 | 1.03 |
| | 1.5 | 1.35 | 1.23 | 0.57 | 0.56 | 0.74 | 1.03 |
| | 1.6 | 1.39 | 1.23 | 0.56 | 0.54 | 0.79 | 1.07 |
| | 1.7 | 1.45 | 1.39 | 0.56 | 0.54 | 0.84 | 1.27 |
| | 1.8 | 1.40 | 1.36 | 0.58 | 0.56 | 0.85 | 1.18 |
| | 1.9 | 1.68 | 1.41 | 0.67 | 0.61 | 0.82 | 1.30 |
| | 2.0 | 1.48 | 1.43 | 0.75 | 0.72 | 0.88 | 1.21 |
| | 2.1 | 1.20 | 1.12 | 0.65 | 0.63 | 0.95 | 1.64 |
| | 2.2 | 1.34 | 1.46 | 0.54 | 0.53 | 1.13 | 1.73 |
| | 2.3 | 1.22 | 1.18 | 0.63 | 0.63 | 1.24 | 2.02 |
| | 2.4 | 0.91 | 0.83 | 0.88 | 0.72 | 1.40 | 2.04 |
| | 2.5 | 1.01 | 0.90 | 0.78 | 0.81 | 1.48 | 1.88 |
| M12 | 0.1 | 0.63 | 0.56 | 0.50 | 0.52 | 0.59 | 0.50 |
| | 0.2 | 0.64 | 0.57 | 0.49 | 0.52 | 0.52 | 0.51 |
| | 0.5 | 0.66 | 0.61 | 0.51 | 0.52 | 0.52 | 0.53 |
| | 1.0 | 0.71 | 0.70 | 0.52 | 0.52 | 0.59 | 0.56 |
| | 1.1 | 0.77 | 0.78 | 0.52 | 0.52 | 0.61 | 0.68 |
| | 1.2 | 0.88 | 0.85 | 0.54 | 0.54 | 0.62 | 0.84 |
| | 1.3 | 0.93 | 0.89 | 0.54 | 0.53 | 0.66 | 0.98 |
| | 1.4 | 0.96 | 0.92 | 0.53 | 0.53 | 0.69 | 0.95 |
| | 1.5 | 0.99 | 0.93 | 0.55 | 0.54 | 0.69 | 0.92 |
| | 1.6 | 0.96 | 0.90 | 0.55 | 0.55 | 0.73 | 1.05 |
| | 1.7 | 1.07 | 1.06 | 0.56 | 0.55 | 0.84 | 1.22 |
| | 1.8 | 1.09 | 1.03 | 0.55 | 0.54 | 0.90 | 1.11 |
| | 1.9 | 1.21 | 1.23 | 0.53 | 0.54 | 0.93 | 1.39 |
| | 2.0 | 1.28 | 1.13 | 0.56 | 0.59 | 0.90 | 1.28 |
| | 2.1 | 1.32 | 1.14 | 0.65 | 0.64 | 1.03 | 1.42 |
| | 2.2 | 1.29 | 1.11 | 0.63 | 0.63 | 1.13 | 1.31 |
| | 2.3 | 1.21 | 1.06 | 0.62 | 0.61 | 1.23 | 1.66 |
| | 2.4 | 1.06 | 0.91 | 0.67 | 0.68 | 1.31 | 1.81 |
| | 2.5 | 1.04 | 0.88 | 0.76 | 0.75 | 1.39 | 1.90 |

best close to best less appropriate

# References


Aitkin, M. and Aitkin, I. A hybrid EM / Gauss-Newton algorithm for maximum likelihood in mixture distributions. *Statistics and Computing*, 6, pp. 127–130, 1996.

Aitkin, M. and Rubin, D. Estimation and hypothesis testing in finite mixture models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 47(1), pp. 67–75, 1985.

Arzideh, F., Brandhorst, G., Gurr, E. Hinsch, W., Hoff, T., Roggenbuck, L., Rothe, G., Schumann, G., Wolters, B., Wosniok, W., and Haeckel, R. An improved indirect approach for determining reference limits from intra-laboratory data bases exemplified by concentrations of electrolytes. *Laboratoriums Medizin*, 33, pp. 52–66, 2009.

Baum, L., Petrie, T., Soules, G., and Weiss, N. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, 41, pp. 164–171, 1970.

Behboodian, J. On the modes of a mixture of two normal distributions. *Technometrics*, 12(1), pp. 131–139, 1970.

Benaglia, T., Chauveau, D., Hunter, D., and Young, D. Mixtools: An R package for analyzing finite mixture models. *Journal of Statistical Software*, 32(6), pp. 1–29, 2009.

Böhning, D. *Computer-Assisted Analysis of Mixtures and Applications - Meta-analysis, disease mapping and others*. Chapmann & Hall, New York, 1999.

Böhning, D., Dietz, E., Schaub, R., Schlattmann, P., and Lindsay, B. The distribution of the likelihood ratio for mixtures of densities from the one-parameter exponential family. *Annals of the Institute of Statistical Mathematics*, 46, pp. 373–388, 1994.

Böhning, D., Schlattmann, P., and Lindsay, B. Computer-assisted analysis of mixtures (C.A.MAN): Statistical Algorithms. *Biometrics*, 48(1), pp. 283–303, 1992.

Blischke, W. R. Moment estimators for the parameters of a mixture of two binomial distributions. *The Annals of Mathematical Statistics*, 33(2), pp. 444–454, 1962.

Charlier, C. *Researches into the theory of probability*. Lund, 1906.

Charlier, C. and Wicksell, S. *On the dissection of frequency functions*. Arkiv for matematik, astronomi och fysik, 1924.

Cohen, A. Estimation in mixtures of two normal distributions. *Technometrics*, 9(1), pp. 15–28, 1967.

Day, N. Estimating the components of a mixture of normal distributions. *Biometrika*, 56, pp. 463, 1969.

Dempster, A., Laird, N., and Rubin, D. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1), pp. 1–38, 1977.

Dick, N. and Bowden, D. Maximum likelihood estimation for mixtures of two normal distributions. *Biometrics*, 29(4), pp. 781–790, 1973.

Du, J. Combined algorithms for fitting finite mixture distributions. Master's thesis, McMaster University Hamilton, Ontaria, 2002.

Efron, B. The two sample problem with censored data. *Proc. Fifth Berkeley Symp. on Math. Statist. and Prob.*, 4, pp. 831–853, 1967.

Eisenberger, I. Genesis of bimodal distributions. *Technometrics*, 6(4), pp. 357–363, 1964.

Everitt, B. An introduction to finite mixture distributions. *Stat Methods Med Res*, 5, pp. 107–127, 1996.

Everitt, B. and Hand, D. *Finite Mixture Distributions*. Chapman and Hall, London, 1981.

Everitt, B. and Hothorn, T. *A Handbook of Statistical Analysis Using R*. Chapman and Hall, London, 2006.

Everitt, B. S. Maximum likelihood estimation of the parameters in a mixture of two univariate normal distributions; a comparison of different algorithms. *The Statistician*, 33(2), pp. 205–215, 1984.

Falls, L. Estimation of parameters in compound Weibull distributions. *Technometrics*, 12(2), pp. 399–407, 1970.

Finch, S., Mendell, N., and Thode, H. Probabilistic measures of adequacy of a numerical search for a global maximum. *Journal of the American Statistical Association*, 84(408), pp. 1020–1023, 1989.

Frühwirth-Schnatter, S. *Finite Mixture and Markov Switching Models*. Springer, New York, 2010.

Furman, W. and Lindsay, B. Measuring the relative effectiveness of moment estimators as starting values in maximizing likelihoods. *Computational Statistics and Data Analysis*, 17(5), pp. 493 – 507, 1994a.

Furman, W. and Lindsay, B. Testing for the number of components in a mixture of normal distributions using moment estimators. *Computational Statistics and Data Analysis*, 17(5), pp. 473 – 492, 1994b.

Gjeddebaek, N. Contribution to the study of grouped observations. iv. some comments on simple estimates. *Biometrics*, 15(3), pp. 433–439, 1959.

Gräsbeck, R. The evolution of the reference value concept. *Clinical Chemistry and Laboratory Medicine*, 42(7), pp. 692–7, 2004.

Hald, A. On the history of the correction for grouping, 1873-1922. *Scandinavian Journal of Statistics*, 28(3), pp. 417–428, 2001.

Hasselblad, V. Estimation of parameters for a mixture of normal distributions. *Technometrics*, 8(3), pp. 431–444, 1966.

Hasselblad, V. Estimation of finite mixtures of distributions from the exponential family. *Journal of the American Statistical Association*, 64(328), pp. 1459–1471, 1969.

Hathaway, R. *Constrained maximum-likelihood estimation for a mixture of m univariate normal distributions*. PhD thesis, Rice University, 1983.

Hathaway, R. A constrained formulation of maximum-likelihood estimation for normal mixture distributions. *The Annals of Statistics*, 13(2), pp. 795–800, 1985.

Healy, M. and Westmacott, M. Missing values in experiments analysed on automatic computers. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 5(3), pp. 203–206, 1956.

Heitjan, D. Inference from grouped continuous data: A review. *Statistical Science*, 4(2), pp. 164–179, 1989.

Holgersson, M. and Jorner, U. Decomposition of a mixture into normal components: A review. *International Journal of Bio-Medical Computing*, 9(5), pp. 367 – 392, 1978.

Hosmer, D. A comparison of iterative maximum likelihood estimates of the parameters of a mixture of two normal distributions under three different types of sample. *Biometrics*, 29(4), pp. 761–770, 1973a.

Hosmer, D. On MLE of the parameters of a mixture of two normal distributions when the sample size is small. *Communications in Statistics*, 1(3), pp. 217–227, 1973b.

Hunter, D., Wang, S., and Hettmansperger, T. Inference for mixtures of symmetric distributions. *The Annals of Statistics*, 35(1), pp. 224–251, 2007.

Karlis, D. and Xekalaki, E. Choosing initial values for the EM algorithm for finite mixtures. *Computational Statistics & Data Analysis*, 41(3-4), pp. 577 – 590, 2003.

Kiefer, J. and Wolfowitz, J. Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *The Annals of Mathematical Statistics*, 27(4), pp. 887–906, 1956.

Little, R. and Rubin, D. *Statistical Analysis with Missing Data*. Wiley Series in probability and mathematical statistics, Hoboken, 2002.

Lo, Y., Mendell, N., and Rubin, D. Testing the number of components in a normal mixture. *Biometrika*, 88(3), pp. 767–778, 2001.

MacDonald, P. *Finite Mixture Distribution Models*, 2011.

MacDonald, P. and Green, P. *User's Guide to Program MIX: An Interactive Program for Fitting Mixtures of Distributions, release 2.3*. ICHTHUS DATA SYSTEMS, Ontario, Canada, 1988.

MacDonald, P. and Pitcher, T. Age-groups from size-frequency data: a versatile and efficient method of analysing distribution mixtures. *Journal of the Fisheries Research Board of Canada*, 36, pp. 987–1001, 1979.

McKendrick, A. Applications of mathematics to medical problems. *Proceedings of the Edinburgh Mathematical Society*, 44, pp. 98–130, 1926.

McLachlan, G. and Basford, K. *Mixture models: inference and applications to clustering*. Dekker, New York, 1988.

McLachlan, G. and Krishnan, T. *The EM Algorithm and Extensions*. Wiley Series in probability and statistics, New York, 1997.

McLachlan, G. and Peel, D. *Finite Mixture Models*. Wiley Series in probability and statistics, New York, 2000.

McLachlan, G. J. On the choice of starting values for the EM algorithm in fitting mixture models. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 37(4/5), pp. 417–425, 1988.

McLachlan, G. J. and Jones, P. N. Fitting mixture models to grouped and truncated data via the EM algorithm. *Biometrics*, 44(2), pp. 571–578, 1988.

McLachlan, G. J. and Jones, P. N. Algorithm AS 254: Maximum likelihood estimation from grouped and truncated data with finite normal mixture models. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 39(2), pp. 273–282, 1990.

Medgyessy, P. *Decomposition of superpositions of distribution functions.* Pub. House of the Hungarian Academy of Sciences, Budapest, 1961.

Mengerse, K., Robert, C., and Titterington, D. *Mixtures: Estimation and Applications.* Wiley Series in probability and mathematical statistics, 2011.

Moharir, P. Estimation of the compounding distribution in the compound Poisson process model for earthquakes. *Journal of Earth System Science*, 101, pp. 347–359, 1992.

Murray, G. Contribution to discussion of paper A.P.Dempster, N.M. Laird, D.B. Rubin. *J. Roy. Statist. Soc. Ser. B*, 39, pp. 805–8011, 1977.

Newcomb, S. A generalized theory of the combination of observations so as to obtain the best result. *American Journal of Mathematics*, 8(4), pp. 343–366, 1886.

Orchard, T. and Woodbury, M. A missing information principle: theory and applications. *Proc. Sixth Berkeley Symp. on Math. Statist. and Prob*, 1, pp. 697–715, 1972.

Rao, C. The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society. Series B (Methodological)*, 10(2), pp. 159–203, 1948.

Redner, R. and Homer, W. F. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*, 26(2), pp. 195–239, 1984.

Schader, M. and Schmidt, F. Computation of maximum likelihood estimates for $\mu$ and $\sigma$ from a grouped sample of a normal population. a comparison of algorithms. *Statistical Papers*, 25, pp. 245–258, 1984.

Schader, M. and Schmidt, F. Maximum-Likelihood-Schätzung aus gruppierten Daten - eine Übersicht. *OR Spektrum*, 10, pp. 1–12, 1988.

Schlattmann, P. *Medical Applications of Finite Mixture Models.* Springer, Berlin Heidelberg, 2009.

Seidel, W., Mosler, K., and Alker, M. A cautionary note on likelihood ratio tests in mixture models. *Annals of the Institute of Statistical Mathematics*, 52, pp. 481–487, 2000.

Simar, L. Maximum likelihood estimation of a compound Poisson process. *The Annals of Statistics*, 4(6), pp. 1200–1209, 1976.

Tan, W. Y. and Chang, W. C. Some comparisons of the method of moments and the method of maximum likelihood in estimating parameters of a mixture of two normal densities. *Journal of the American Statistical Association*, 67(339), pp. 702–708, 1972.

Teicher, H. Identifiability of mixtures. *The Annals of Mathematical Statistics*, 32(1), pp. 244–248, 1961.

Teicher, H. Identifiability of finite mixtures. *The Annals of Mathematical Statistics*, 34(4), pp. 1265–1269, 1963.

Titterington, D., Smith, A., and Makov, U. *Statistical Analysis of finite mixture distributions*. John Wiley & Sons, Chichester, 1985.

Turnbull. The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 3, pp. 290–295, 1976.

Wehling, M. *Klinische Pharmakologie*. Thieme Georg Verlag, Stuttgart, 2005.

Wu, C. On the convergence properties of the EM algorithm. *The Annals of Statistics*, 11(1), pp. 95–103, 1983.

Yakowitz, S. and Spragins, J. On the identifiability of finite mixtures. *The Annals of Mathematical Statistics*, 39(1), pp. 209–2014, 1968.

Young, D., Benaglia, T., Chauveau, D., Elmore, R., Hettmansperger, T., Hunter, D., Thomas, H., and Xuan, F. *Tools for analyzing finite mixture models*, 2010.