

**McLachlan, Geoffrey; Peel, David**

**Finite mixture models.** (English) Zbl 0963.62061

*Wiley Series in Probability and Mathematical Statistics*. Applied Probability and Statistics. Chichester: Wiley. xxii, 419 p. (2000).

Research on finite mixture modeling has increased in recent years. In addition, mixture models are used in a wide range of disciplines. Therefore, a variety of results remain unknown from other disciplines though they could serve as a basis for further developments. The present authors tried hard to gather together the developments made over the last years on this hot topic of scientific research. The book provides an extensive review on finite mixture modeling focusing on the computational aspect of using such models in practice. The impact of computer machines allowed using complicated mixture models in a wide range of applications from several scientific disciplines. The book provides a vast amount of existing work on finite mixture modeling and in particular on estimation via the EM algorithm and related material.

The book is divided into 13 chapters treating several different issues of finite mixture models. The first chapter describes the general finite mixture setting, discussing improvements made over the last years and problems addressed in this area. Chapter 2 reviews maximum likelihood estimation, with special focus on the EM algorithm, its properties and improvements proposed in the literature. Chapter 3 treats the multivariate normal mixture model, which is the basis for clustering algorithms. Problems encountered with this model are discussed in depth, like spurious local maximizers. Chapter 4 describes the Bayesian approach in finite mixture models that has been developed via Gibbs sampling schemes in the last decade. In chapter 5, non-normal models are treated, including mixtures of other distribution families, mixtures of generalized linear models like Poisson regression and logistic regression models, etc.

Chapter 6 is devoted to the problem of assessing the number of components comprising the mixtures. Methodologies developed for this problem are reviewed. Chapter 7 contains material on multivariate  $t$  mixtures as robust alternatives to the multivariate normal mixture model. Mixtures of factor analyzers are described in chapter 8. The material of this chapter is quite novel. Binned data are treated in chapter 9, while mixture models for failure time data are proposed in chapter 10. Recent developments on mixtures of directional data are discussed in chapter 11. Chapter 12 describes variants of the EM algorithm suitable for large data sets. Such variants are quite useful for those working with large databases in disciplines related to statistics, like data miners. The last chapter describes hidden Markov models.

Throughout the book there are several examples and case studies that enable the reader to follow the theoretical material discussed. The main element, in all the cases, is the use of EM algorithms and its variants that facilitate the estimating procedures. Theoretical results about mixtures are sparse in this book, which focuses mainly on an applications basis. The bibliography provided by the authors is considerable and hence one can find a lot of interesting papers that appeared in different research areas. Thus, the literature provided is a very useful guide for those working in this area.

The book will become popular to many researchers as it provides a complete, up to date coverage of the topic and also discusses problems that may be further pursued by researchers in the forthcoming years. Students not familiar with computational topics may have some difficulty to follow the book at once, but they will love the fact that the material covered is so wide that it will make this book a standard reference for the forthcoming years.

Reviewer: [Dimitris Karlis \(Athens\)](#)

**MSC:**

- 62J12 Generalized linear models (logistic models)
- 62-01 Introductory exposition (textbooks, tutorial papers, etc.) pertaining to statistics
- 65C60 Computational problems in statistics (MSC2010)
- 62H30 Classification and discrimination; cluster analysis (statistical aspects)

Cited in **3** Reviews  
Cited in **1107** Documents

**Keywords:**

EM algorithm; clustering algorithms; hidden Markov models; bibliography