

McLachlan G, Peel, D A 2000: *Finite mixture models*. New York: Wiley 419pp, \$94.50 (HB). ISBN 0 471 00626 2.

Finite mixture distributions are certainly not new – Karl Pearson’s classic paper on estimating the five parameters in a mixture of two univariate normal densities is now over 100 years old – but it is only in the past decade that the potential of such distributions for modelling a range of random phenomena has been fully appreciated, and they have been applied in astronomy, biology, genetics, medicine, psychiatry, economics, engineering and marketing. The book by McLachlan and Peel is a masterly overview of the area, which manages the difficult task of integrating the technical and the practical almost seamlessly, so that material included should appeal to a variety of statisticians ranging from those interested in the minutia of the convergence properties of the EM algorithm to those more concerned with deciding how many components to include when modelling a particular data set.

The book opens with a general introduction in which mixture distributions are defined and the shapes of univariate mixtures considered. Identifiability is also covered in this chapter, and a number of topics to be considered in more detail in later chapters are introduced.

Chapter 2 gets down to the details of the maximum likelihood fitting of mixture models. The EM algorithm is comprehensively covered in this chapter; choice of starting values, rate of convergence of the algorithm, and variants of the basic algorithm are all given a sub-section, and it is extremely useful to have the material from a large variety of sources collected together and integrated as it is here. Some of this chapter is quite technically demanding but it provides an unbeatable source of information for anybody developing algorithms for fitting mixture distributions.

Chapter 3 concentrates on mixture models with normal components, both univariate and multivariate. Examples of the former include screening for hemochromatosis and diagnostic criteria for diabetes, and of the latter the well-known crab data and data on haemophilia. These applications are well described, as are some of the problems that can be encountered when using the EM algorithm in practice. My only disappointment with this chapter is that Fisher’s iris data manages to make an appearance!

With the advent of MCMC methods, Bayesian estimation is now a viable proposition for fitting mixture models as it is in many other areas of statistics. Chapter 4 describes the Bayesian approach but sadly no numerical examples are included. In Chapter 5 mixtures with non-normal components are discussed. Perhaps the most practically interesting model here is that for data sets where some of the variables are continuous and some are categorical. The approach proposed is based on the location model used originally in discriminant analysis; an interesting application involving prostate cancer data is used to illustrate the possibilities of the model.

In Chapter 6 McLachlan and Peel deal with one of the most difficult problems faced by users of finite mixture distributions, namely estimating the number of components to be used. Well-known problems with the likelihood ratio test in this context are discussed, and both theoretical and simulation results are discussed. Bootstrapping and a variety of information criteria are also considered. There is a lot of material in this chapter and I would have liked to see a summary of results and recommendations at the end, but no

such summary section is given, an omission that applies to most of the other chapters in the book.

Chapters 7, 8 and 9 deal respectively with multivariate t mixtures, mixtures of factor analysers and fitting mixture models to binned data, topics which will probably be of interest only to a minority of readers. But Chapter 10, dealing with mixture models for failure-time data, contains much material that will be of particular interest to medical statisticians dealing with survival time data.

The last two chapters of the book deal with directional data, and variants of the EM algorithm for large databases. The latter contains some nice examples, including one on segmentation of MR images. An appendix on mixture software and a list of 800 references, 40% of which have appeared since 1995, complete the book.

This is a scholarly and authoritative account of finite mixture distributions, well written and containing many interesting examples. It is difficult to ask for more and there is no doubt that McLachlan and Peel's book will be the standard reference on mixture models for many years to come.

Reviewed by Brian Everitt, Biostatistics and Computing Department, Institute of Psychiatry, London, UK.