



ELSEVIER

Computational Statistics & Data Analysis 41 (2003) 349–357

COMPUTATIONAL
STATISTICS
& DATA ANALYSIS

www.elsevier.com/locate/csda

Editorial: recent developments in mixture models

Dankmar Böhning^{a,*}, Wilfried Seidel^b

^a*Unit of Biometry and Epidemiology, Institute for International Health, Joint Center for Humanities and Health Sciences, Free University Berlin/Humboldt-University Berlin, Fabeckstr. 60-62, Haus 562, 14195 Berlin, Germany*

^b*School of Economic and Organizational Sciences, University of the Federal Armed Forces Hamburg, D-22039 Hamburg, Germany*

Received 1 March 2002; received in revised form 1 April 2002

Abstract

Recent developments in the area of mixture models are introduced, reviewed and discussed. The paper introduces this special issue on mixture models, which touches upon a diversity of developments which were the topic of a recent conference on mixture models, taken place in Hamburg, July 2001. These developments include issues in nonparametric maximum likelihood theory, the number of components problem, the non-standard distribution of the likelihood ratio for mixture models, computational issues connected with the EM algorithm, several special mixture models and application studies.

© 2002 Elsevier Science B.V. All rights reserved.

Keywords: Nonparametric maximum likelihood; Unobserved heterogeneity; Number of components; Likelihood ratio; Application studies

1. Introduction

Mixture models have experienced increased interest and popularity over last decades. A recent conference in Hamburg (Germany) in 2001 on *Recent developments in mixture models* tried to capture contemporary issues of debate in this area and have them discussed in a broader audience. This special issue of CSDA introduces some of these topics and puts them together in a common context. In the following mixture modeling is introduced in an elementary way and the various contributions to the Special Issue (SI) of CSDA are set into perspective. Note that references to contributions in the

* Corresponding author. Fax: 84451290.

E-mail addresses: boehning@zedat.fu-berlin.de (D. Böhning), wilfried.seidel@unibw-hamburg.de (W. Seidel).

SI are of the form: *author(s)* (SI \backslash year), such as Böhning and Seidel, 2003(SI) which refers to this editorial in the SI. This is in contrast to normal referencing which would be without the prefix SI: *author(s)* (year).

The importance of *mixture distributions*, their enormous developments and their frequent applications over recent years is due to the fact that mixture models offer natural models for *unobserved population heterogeneity*. What does this mean? Suppose that a *one-parametric density* $f(x|\lambda)$ can be assumed for the phenomenon of interest. Here, λ denotes the parameter of the population, whereas x is in the sample space X . We call this the *homogeneous case*. However, often this model is too strict to capture the variation of the parameter over a diversity of subpopulations. In this case, we have that the population consists of various subpopulations, denoted by $\lambda_1, \lambda_2, \dots, \lambda_k$, where k denotes the number (possibly unknown) of subpopulations. We call this situation the *heterogeneous case*. In contrast to the homogenous case, we have the same type of density in each subpopulation j , but a potentially different parameter: $f(x|\lambda_j)$ is the density in subpopulation j .

In the sample x_1, x_2, \dots, x_n it is not observed which subpopulation the observations are coming from. Therefore, we speak of *unobserved heterogeneity*. Let k latent, dummy variables Z_j describe the population membership for $j = 1, \dots, k$, e.g. $Z_j = 1$, means that x derives from subpopulation j . Then, the joint density $f(x, z_j)$ can be written as $f(x, z_j) = f(x|z_j)f(z_j) = f(x|\lambda_j)p_j$, where $f(x|z_j) = f(x|\lambda_j)$ is the density conditionally on membership in subpopulation z_j . Therefore, the unconditional density $f(x)$ is the *marginal density*

$$f(x|P) = \sum_{j=1}^k f(x, z_j) = \sum_{j=1}^k f(x|\lambda_j)p_j, \quad (1)$$

where the margin is taken over the latent variable Z_j . Note that p_j is the probability of belonging to the j th subpopulation having parameter λ_j . Therefore, the p_j have to meet the constraints $p_j \geq 0$, $p_1 + \dots + p_k = 1$. Note that (1) is a mixture distribution with *kernel* $f(x, \lambda)$ and *mixing distribution* P , in which weights p_1, \dots, p_k are given to parameters $\lambda_1, \dots, \lambda_k$. Estimation is done conventionally by maximum likelihood, that is we have to find \hat{P} which maximizes the log-likelihood $l(P) = \sum_{i=1}^n \log f(x_i, P)$. \hat{P} is called the *nonparametric maximum likelihood estimator* (Laird, 1978), if it exists. Note that also the number of subpopulations k is treated as unknown and is estimated. Note, in addition, that it is essential for the existence of the NPMLE that the likelihood is bounded. For example, in the case of the Poisson, Binomial, or Exponential the likelihood is bounded and the NPMLE exists. For the normal with common variance parameter one has to fix the number of components to bound the likelihood, for the normal with component-specific variances in the mixture any $k > 1$ will make the likelihood unbounded (though a version of the EM algorithm exist as well for this situation and is frequently used in practice; see Nityasuddhi and Böhning, 2003(SI)).

Many applications are of the following type: Under standard assumptions the population is homogeneous, leading to a simple, one-parametric and *natural* density. Examples include the binomial, the Poisson, the geometric, the exponential, the normal distribution (with additional variance parameter). If these standard assumptions are violated

because of population heterogeneity, mixture models can easily capture these additional complexities. Therefore, mixture models are considered for conventional densities such as *normal* (common unknown, known and unknown different variances), Poisson, Poisson for standardized mortality ratio data, Binomial, Binomial for rate data, Geometric, Exponential among others. We speak of *direct applications* of mixture models. Incorporating covariates, these direct applications lead in a very natural way to an extension of the generalized linear model: incorporating unobserved heterogeneity we are lead to the generalized linear mixed model.

The development of mixture models is underlined by the appearance of a number of recent books on mixtures including Lindsay (1995), Böhning (2000), McLachlan and Peel (2000) which update previous books by Everitt and Hand (1981), Titterington et al. (1985) and McLachlan and Basford (1988).

2. Some basic results on the NPMLE

The strong results of nonparametric mixture distributions are based on the fact that the log-likelihood l is a *concave* functional on the set of *all* discrete probability distributions Ω . The NPMLE (Laird, 1978) is defined as an element in Ω which maximizes the log-likelihood (if it exists). Van de Geer, 2003(SI) reviews some asymptotic properties of the NPMLE. It is very important to distinguish between the set of all discrete distributions and the set Ω_k of all distributions with a fixed number of k support points (subpopulations). The latter set is *not* convex. For details, see Lindsay (1995) or Böhning (2000). The major tool for achieving characterizations and algorithms is the *gradient function*, defined as

$$d(\lambda, P) = \frac{1}{n} \sum_{i=1}^n \frac{f(x_i|\lambda)}{f(x_i|P)}.$$

We have the general *mixture maximum likelihood theorem* (Lindsay, 1983; Böhning, 1982):

- (a) \hat{P} is NPMLE if and only if $D_{\hat{P}}(\lambda) \leq 0$ for all λ or, if and only if $d(\lambda, \hat{P}) \leq 1$ for all λ
- (b) $d(\lambda, \hat{P}) = 0$ for all support points λ of \hat{P} .

Secondly, the concept of the gradient function is important in developing *reliable, globally converging algorithms* (for a review see Böhning, 1995; Lesperance and Kalbfleisch, 1992). Schlattmann, 2003(SI) uses an improved version of one of these algorithms to develop an inferential Bootstrap approach for estimating the number of components in a mixture.

3. Mixture models and the EM algorithm

When the number of components k is fixed in advance, the set of all probability distributions with k support points is *non-convex* which makes the maximization of the log-likelihood a lot more difficult, and multiple local maxima might exist. However,

the advantage of this restricted situation lies in the fact that parameter estimation can be treated with the EM algorithm (Dempster et al., 1977). Here, the complete data likelihood is simply

$$\prod_j [f(x|\lambda_j) p_j]^{Z_j} \quad (2)$$

and the complete data log-likelihood separates the estimation of p_1, \dots, p_k and estimation of the parameters involved in $f(x|\lambda_j)$. The *E*-step is can be executed easily as well using Bayes theorem:

$$\begin{aligned} e_j &= E(Z_j|x, P) = Pr(Z_j = 1|x, P) = f(x|Z_j = 1, P) \\ &= Pr(Z_j = 1|P) / \sum_{j'} f(x|Z_{j'} = 1, P) Pr(Z_{j'} = 1|P) \\ &= f(x|\lambda_j) p_j / \sum_{j'} f(x|\lambda_{j'}) p_{j'} \end{aligned} \quad (3)$$

The expected complete data log-likelihood becomes

$$\sum_j e_j \log p_j + \sum_j e_j \log f(x|\lambda_j)$$

and based upon the full sample x_1, \dots, x_n

$$\sum_i \sum_j e_{ij} \log p_j + \sum_i \sum_j e_{ij} \log f(x_i|\lambda_j), \quad (4)$$

where $e_{ij} = E(Z_{ij}|x_i, P)$. From (4) new estimates for p_1, \dots, p_k are easily determined as $\sum_i e_{ij}/n$, whereas the estimates for the parameters involved in $f(x|\lambda_j)$ will depend on the form of f . The EM algorithm is one of the most frequently used algorithms for determining maximum likelihood estimates in mixture models because of its numerical stability and monotonicity property. However, recently more attention is focussed on its dependence on initial values (in converging to local maxima). See Karlis and Xekalaki, 2003(SI), Biernacki et al., 2003(SI), Seidel et al. (2000), or Böhning (2002). In the latter contribution, the concept of the gradient function is utilized to improve the convergence behavior of the EM algorithm.

The idea of incomplete data can be extended to incorporate *missing data* (Hunt and Jorgensen, 2003(SI) or *truncated data* like data stemming from capture-recapture experiments in which the counting distributions of the number of catches is modeled. Here, the frequencies of zero-catches remains unobserved and can be imputed with an appropriate modification of the EM algorithm. A contribution to this area is provided by Mao and Lindsay, 2003(SI).

4. Likelihood ratio test and number of components

Although the nonparametric estimation of the heterogeneity distribution provides an estimate of the number of components itself, it is sometimes requested to use the likelihood ratio test for testing whether a reduced number of components is likewise sufficient. It is well known (Titterton et al., 1985; McLachlan and Basford, 1988) that conventional asymptotic results for the null distribution of the likelihood ratio statistic do not hold, since the null hypothesis lies on the boundary of the alternative hypothesis. In some cases theoretical results are available (Böhning et al., 1994), but in other cases simulation results must be used. In general, a parametric Bootstrap procedure can

be used (McLachlan, 1992), and it was pointed out recently that this approach leads to valid statistical inference (Feng and McCulloch, 1996). Recent papers focussing on the likelihood ratio include Schlattmann, 2003(SI) or Pons and Lemdani, 2003(SI); Pavlic and van der Laan, 2003(SI) consider a distance estimation approach. Other approaches are based upon the Akaike information criterion, Bayesian information criterion and alike; for details, see McLachlan and Peel (2000).

5. Mixture models and covariates

In many application areas the analysis of additional covariates is often desirable. Mixture modeling with covariates leads to the area of *mixed generalized linear models*. Consider again the basic mixture model (1): $\sum_{j=1}^k f(x, \lambda_j) p_j$. Suppose that a number of covariates u_1, \dots, u_p is given. These are then related to the component mean by means of a linear model $\lambda_j = \mathbf{u}^t \beta_j = \beta_{0j} + u_1 \beta_{1j} + \dots + u_p \beta_{pj}$ or by means of generalized linear model $\lambda_j = h^{-1}(\mathbf{u}^t \beta_j)$, where h is the link function like $h(\cdot) = \log(\cdot)$ for a Poisson model. Mixing can be done on all or only on part of the covariates, or solely on the intercept. Also, the multi-level structure of the data might be taken into account. Various authors have contributed developments in this area including Aitkin (1996, 1999) or Dietz (1992), Dietz and Böhning (1996). Sometimes, also the weights $p_j = p_j(\mathbf{u})$ in the mixture (2) are modeled in dependency of the covariates. In an application study, models of this kind are used for modeling *length of stay in a hospital* (Yau et al., 2003(SI)).

6. Special mixture models

One class of special mixture models arises when certain component parameters are fixed to specific values. An example is the *zero inflated Poisson* which is a two component mixture with the first component fixed at 0 which makes this component distribution singular (one-point distribution). The zero-inflated Poisson model is: $(1 - p)\text{Po}(0|\lambda) + p\text{Po}(x|\lambda)$ where $\text{Po}(x|\lambda) = \exp(-\lambda)\lambda^x/x!$. Note that $\text{Po}(0|\lambda) = 0$, unless $\lambda = 0$ in which case $\text{Po}(0|\lambda) = 1$. Models of this kind are considered in Dalrymple et al., 2003(SI), Lambert (1992), Böhning (1998); see also Mooney et al., 2003(SI) for modeling count data in their time dependency. Similarly in survival analysis, when a certain part of the (treated) population remains disease free, we are led to *cure models*. Cure models are also special mixture models, namely (two-component) mixtures of survival time distributions in which one component is a singular one-point distribution (survival guaranteed). Contributions to this area are from Pons and Lemdani, 2003(SI), Peng, 2003(SI) and Morbiducci et al., 2003(SI).

7. Application of mixture models and connection to Bayesian methods

Mixture models do have many application areas of their own dimension, depth and value—to mention just two: *disease mapping and meta analysis*. In these areas interest

is often in a stabilized estimate of the quantity of interest in the spatial unit (disease mapping) or the study (meta analysis). These can be provided by the *empirical Bayes estimators* which are developed from the posterior distribution (3) like a mean measure such as the expected value with respect to the posterior. The key idea is to think of the heterogeneity distribution P as a prior distribution on the unknown parameter λ . Then (3) establishes the posterior distribution. The mixture model comes in by providing a tool to replace the unknown prior (mixing) distribution by an appropriate estimate. A classical paper on this is from Clayton and Kaldor (1987). For applications in meta-analysis see also Böhning (2000). Full Bayesian procedures arise when further prior distributions (hyperpriors) are assumed for modeling all parameters involved in the prior. Aitkin (2002) has reviewed full Bayesian approaches to cope with the number of components problem. He could demonstrate a rather strong dependence on the posterior from the assumed hyperprior for the number of components. This is not surprising since mixture likelihoods can be extremely flat, bringing out the influence of the hyperprior model. A contribution to these areas is from Biggeri et al., 2003(SI). Other applications include texture modeling by mixtures (Grim and Haindl, 2003(SI)) and distance-based modeling of ranking data (Murphy and Martin, 2003(SI)).

8. Multivariate mixtures

Developments in the area of multivariate mixtures are few and mostly concentrated on mixtures of multivariate normals. Contributions are given here by McLachlan et al., 2003(SI) as well as Biernacki et al., 2003(SI). A further special multivariate mixture model is provided by the latent class approach and associated developments in which a multidimensional contingency table is analyzed with respect to a latent, discrete variable. The ideas go back originally to Lazarsfeld and Henry (1968) and has been further developed by several authors including Clogg (1979) and Formann (1982), to mention only a few. Contributions here are from Croft and Smith, 2003(SI), Formann, 2003(SI), and Vermunt, 2003(SI). Finally, Berchtold, 2003(SI) considers mixture modeling of heteroscedastic times series.

9. Diagnostics, testing, and adjusting for heterogeneity

Before one enters into the cumbersome task of fitting a mixture model it might be of value to know about the presence of heterogeneity. If there is none, there is also no need to go into mixture modeling. This is the area of diagnostics and testing for heterogeneity (mixing). Susko, 2003(SI) contributes to this area as well as Mao and Lindsay, 2003(SI), the latter in combination with a very interesting application of truncated Poisson distributions arising from capture-recapture data which are considered more and more in many areas including biology/ecology, but also epidemiology. Recently, a major contribution in the American Journal of Epidemiology (IWGDMF, 1995a, b) pointed out the importance of these techniques for determining the completeness of disease incidence registries. Sometimes one considers the occurrence of heterogeneity as a

nuisance that needs to be adjusted for. Viwatwongkasem and Böhning, 2003(SI) consider such a situation, namely the risk difference in multi-center studies under baseline heterogeneity, and study a number of estimators adjusting for heterogeneity.

Acknowledgements

This research is under support of the *German Research Foundation*. In addition, the guest editors are most grateful to the *German Research Foundation* for funding a conference on mixture models which took place in Hamburg, Germany, 2001. This Special Issue of CSDA was initiated from contributions to this conference. The guest editors would also acknowledge the help of the Associate Editors H. Friedl (Graz, Austria), B. Lindsay (Penn State, USA), G. McLachlan (Brisbane, Australia), A. Raftery (Seattle, USA) dealing with the quite numerous submissions for this Special Issue. Special thanks go to several referees. Without their continuous efforts the reviewing could not have been done. They are (in alphabetical order):

M. Aitkin	A. Formann	M. Markatou	E. Stampfer
F. Bartolucci	H. Grossmann	R. Pilla	E. Susko
H. Bensmail	P. Hoff	V. Patilea	S. van de Geer
A. Berchtold	D. Karlis	K. Roeder	G. Verbeke
A. Biggeri	G. Kauermann	B. Rossi	J. Vermunt
G. Celeux	L. Knorr-Held	T. Rudas	Chukiat Viwatwongkasem
E. Dietz	E. Lessaffre	J. Rost	K. Yau
E. Dreassi	M. Lesperance	J. Sarol	X. Zhou
L. Fahrmeir	J. Marden	P. Schlattmann	

Finally, the guest editors would express their sincerest thanks to the co-editors Peter Naeve and Erricos John Kontoghiorghes.

References

- Aitkin, M.A., 1996. NPML estimation of the mixing distribution in general statistical models with unobserved random variation. In: Seeber, G.U.H., Francis, B.J., Hatzinger, R., Steckel-Berger, G. (Eds.), *Statistical Modeling*. Springer, Berlin, pp. 1–9.
- Aitkin, M.A., 1999. A general maximum likelihood analysis of variance. *Biometrics* 55, 117–128.
- Aitkin, M.A., 2002. Likelihood and Bayes approaches to the number of mixture components. Volume of Abstracts, Conference on Recent Developments in Mixture Models, Hamburg, 2001.
- Berchtold, A., 2003. Mixture transition distribution (MTD) modeling of heteroscedastic time series. Special issue on mixtures. *Comput. Statist. Data Anal.*, 41 (this Vol.).
- Biernacki, C., Celeux, G., Govaert, G., 2003. Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. Special issue on mixtures. *Comput. Statist. Data Anal.*, 41 (this Vol.).
- Biggeri, A., Dreassi, E., Lagazio, C., Böhning, D., 2003. A transitional non-parametric maximum pseudo-likelihood estimator for disease mapping. Special issue on mixtures. *Comput. Statist. Data Anal.*, 41 (this Vol.).
- Böhning, D., 1982. Convergence of Simar's algorithm for finding the maximum likelihood estimate of a compound Poisson process. *Ann. Statist.* 10, 1006–1008.

- Böhning, D., 1995. A review of reliable maximum likelihood algorithms for the semi-parametric mixture maximum likelihood estimator. *J. Statist. Plann. Inference* 47, 5–28.
- Böhning, D., 1998. Zero-inflated poisson models and C.A.MAN: a tutorial collection of evidence *Biometrical J.* 7, 833–843.
- Böhning, D., 2000. *Computer-Assisted Analysis of Mixtures and Applications. Meta-Analysis, Disease Mapping and Others.* Chapman & Hall /CRC, Boca Raton.
- Böhning, D., 2002. A globally-convergent adjustment of the EM algorithm for mixture models with fixed number of components. submitted for publication.
- Böhning, D., Dietz, E., Schaub, R., Schlattmann, P., Lindsay, B.G., 1994. The distribution of the likelihood ratio for mixtures of densities from the one-parametric exponential family. *Ann. Inst. Statist. Math.* 46, 373–388.
- Clayton, D., Kaldor, J., 1987. Empirical Bayes estimates for age-standardized relative risks. *Biometrics* 43, 671–681.
- Clogg, C.C., 1979. Some latent structure models for the analysis of Likert-type data. *Soc. Sci. Res.* 8, 287–301.
- Croft, J., Smith, J.Q., 2003. Discrete mixtures in Bayesian networks with hidden variables: a latent time budget example. Special issue on mixtures. *Comput. Statist. Data Anal.*, 41 (this Vol.).
- Dalrymple, M.L., Hudson, I.L., Ford, R.P.K., 2003. Finite Mixture, Zero-inflated Poisson and Hurdle Models with Application to SIDS. Special issue on mixtures. *Comput. Statist. Data Anal.*, 41 (this Vol.).
- Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. B* 39, 1–38.
- Dietz, E., 1992. Estimation of heterogeneity—a GLM-approach. In: Fahrmeir, L., Francis, F., Gilchrist, R., Tutz, G. (Eds.), *Advances in GLIM and Statistical Modeling. Lecture Notes in Statistics*, Springer, Berlin, pp. 66–72.
- Dietz, E., Böhning, D., 1996. Statistical inference based on a general model of unobserved heterogeneity. In: Fahrmeir, L., Francis, F., Gilchrist, R., Tutz, G. (Eds.), *Advances in GLIM and Statistical Modeling. Lecture Notes in Statistics*, Springer, Berlin, pp. 75–82.
- Feng, Z.D., McCulloch, C.E., 1996. Using Bootstrap likelihood ratios in finite mixture models. *J. Roy. Statist. Soc. B* 58, 609–617.
- Formann, A.K., 1982. Linear logistic latent class analysis. *Biometrical J.* 24, 171–190.
- Formann, A.K., 2003. Latent class model diagnostics—a review and some proposals. Special issue on mixtures. *Comput. Statist. Data Anal.*, 41 (this Vol.).
- Grim, J., Haindl, M., 2003. Texture modelling by discrete distribution mixtures. Special issue on mixtures. *Comput. Statist. Data Anal.*, 41 (this Vol.).
- Hunt, L., Jorgensen, M., 2003. Mixture model clustering for mixed data with missing information. Special issue on mixtures. *Comput. Statist. Data Anal.*, 41 (this Vol.).
- International Working Group for Disease Monitoring and Forecasting, 1995a. Capture–recapture and multiple record systems estimation I: history and theoretical development *Amer. J. Epidemiol.* 142, 1047–1058.
- International Working Group for Disease Monitoring and Forecasting, 1995b. Capture–recapture and multiple record systems estimation II: Applications in human diseases *Amer. J. Epidemiol.* 142, 1059–1068.
- Karlis, D., Xekalaki, E., 2003. Choosing initial values for the EM algorithm for finite mixtures. Special issue on mixtures. *Comput. Statist. Data Anal.*, 41 (this Vol.).
- Laird, N.M., 1978. Nonparametric maximum likelihood estimation of a mixing distribution. *J. Amer. Statist. Assoc.* 73, 805–811.
- Lazarsfeld, P.F., Henry, N.W., 1968. *Latent Structure Analysis.* Houghton Mifflin, Boston.
- Lesperance, M., Kalbfleisch, J.D., 1992. An algorithm for computing the nonparametric MLE of a mixing distribution. *J. Amer. Statist. Assoc.* 87, 120–126.
- Lindsay, B.G., 1983. The geometry of mixture likelihoods, Part I: a general theory *Ann. Statist.* 11, 783–792.
- Lindsay, B.G., 1995. *Mixture Models: Theory, Geometry, and Applications.* . NSF-CBMS Regional Conference Series in Probability and Statistics, Vol. 5. Institute of Mathematical Statistics, Hayward.
- Mao, C.X., Lindsay, B.G., 2003. Tests and diagnostics for heterogeneity in the species problem. Special issue on mixtures. *Comput. Statist. Data Anal.*, 41 (this Vol.).
- McLachlan, G.J., 1992. Cluster analysis and related techniques in medical research. *Statist. Meth. Med. Res.* 1, 27–49.

- McLachlan, G.J., Basford, K.E., 1988. *Mixture Models. Inference and Applications to Clustering*. Marcel Dekker, New York.
- McLachlan, G.J., Peel, D., 2000. *Finite Mixture Models*. Wiley, New York.
- McLachlan, G.J., Peel, D., Bean, R.W., 2003. Modelling high-dimensional data by mixtures of factor analyzers. Special issue on mixtures. *Comput. Statist. Data Anal.*, 41 (this Vol.).
- Mooney, J.A., Helms, P.J., Jolliffe, I.T., 2003. Fitting mixtures of von Mises distributions: a case study involving Sudden Infant Death Syndrome. Special issue on mixtures. *Comput. Statist. Data Anal.*, 41 (this Vol.).
- Morbiducci, M., Nardi, A., Rossi, C., 2003. Classification of cured individuals in survival analysis: the mixture approach to the diagnostic-prognostic problem. Special issue on mixtures. *Comput. Statist. Data Anal.*, 41 (this Vol.).
- Murphy, T.B., Martin, D., 2003. Mixtures of distance-based models for ranking data. Special issue on mixtures. *Comput. Statist. Data Anal.*, 41 (this Vol.).
- Nityasuddhi, D., Böhning, D., 2003. Asymptotic properties of the EM algorithm estimate for normal mixture models with component specific variances. Special issue on mixtures. *Comput. Statist. Data Anal.*, 41 (this Vol.).
- Pavlic, M., van der Laan, M.J., 2003. Fitting of mixtures with unspecified number of components using cross validation distance estimate. Special issue on mixtures. *Comput. Statist. Data Anal.*, 41 (this Vol.).
- Peng, Y., 2003. Fitting semiparametric cure models. Special issue on mixtures. *Comput. Statist. Data Anal.*, 41 (this Vol.).
- Pons, O., Lemdani, M., 2003. Estimation and test in long-term survival mixture models. Special issue on mixtures. *Comput. Statist. Data Anal.*, 41 (this Vol.).
- Schlattmann, P., 2003. Estimating the number of components in a finite mixture model: The special case of homogeneity. Special issue on mixtures. *Comput. Statist. Data Anal.*, 41 (this Vol.).
- Seidel, W., Mosler, K., Alker, M., 2000. A cautionary note on likelihood ratio tests in mixture models. *Ann. Instit. Statist. Math.* 52, 481–487.
- Susko, E., 2003. Weighted tests of homogeneity for testing the number of components in a mixture. Special issue on mixtures. *Comput. Statist. Data Anal.*, 41 (this Vol.).
- Titterton, D.M., Smith, A.F.M., Makov, U.E., 1985. *Statistical Analysis of Finite Mixture Distributions*. Wiley, New York.
- Van de Geer, S., 2003. Maximum likelihood for nonparametric mixture models. Special issue on mixtures. *Comput. Statist. Data Anal.*, 41 (this Vol.).
- Vermunt, J.K., 2003. Latent class models for classification. Special issue on mixtures. *Comput. Statist. Data Anal.*, 41 (this Vol.).
- Viwatwongkasem, C., Böhning, W., 2003. A comparison of risk difference estimators in multi-center studies under baseline-risk heterogeneity. Special issue on mixtures. *Comput. Statist. Data Anal.*, 41 (this Vol.).
- Yau, K.K.W., Lee, A.H., Ng, A.S.K., 2003. Finite mixture regression model with random effects: Application to neonatal hospital length of stay. Special issue on mixtures. *Comput. Statist. Data Anal.*, 41 (this Vol.).