Standard Errors of Fitted Component Means of Normal Mixtures

K.E. Basford[1], D.R. Greenway[1], G.J. McLachlan[2], and D. Peel[2]

[1] Department of Agriculture, The University of Queensland 4072 Australia
[2] Department of Mathematics, The University of Queensland 4072 Australia

Summary

In this paper use consider the problem of providing standard errors of the component means in normal mixture models fitted to univariate or multivariate data by maximum likelihood via the EM algorithm. Two methods of estimation of the standard errors are considered: the standard information-based method and the computationally- intensive bootstrap method. They are compared empirically by their application to three real data sets and by a small-scale Monte Carlo experiment.

Keywords: Normal mixture models, Observed information, Standard errors, Bootstrap approach.

1 Introduction

This paper is concerned with the calculation of standard errors of the estimated component means obtained when fitting a normal mixture model by maximum likelihood via the EM algorithm. The standard errors provide a crude but useful measure of accuracy of the point estimates so obtained. Also, if estimates of the standard errors are available, asymptotic normal theory leads to approximate confidence intervals for the parameters of interest.

Two methods of obtaining standard errors are presented and their applicability to real data sets examined. The first method is based on the information in the sample. Asymptotic variances of the estimated parameters in the mixture model are obtained from the diagonal elements of the inverse of the Fisher information matrix (Chang, 1976, 1979; Hosmer, 1973b; Tan and Chang, 1972; McLachlan and Basford, 1988 p.47). It is common in practice to estimate the expected information matrix by the observed information matrix, which is the negative of the Hessian matrix of the log likelihood function evaluated at the maximum likelihood solution (Efron and Hinkley, 1978).

While the information-based method is asymptotically applicable, it may not provide reliable estimates of the standard errors of the component means unless the sample size is very large (Day, 1969) or the component means are well separated (Hosmer, 1973a; Dick and Bowden, 1973). Tukey (1986) cautions that we ought not to expect conventional asymptotic theory to be widely useful, and that in many practical situations asymptotics are a serious oversimplification.

The second method involves the calculation of bootstrap estimates of the standard errors of the fitted component means by repeated sampling from the fitted normal mixture distribution. This method will provide accurate standard error estimates provided that a sufficient number of bootstrap samples is generated (Efron, 1982). We shall first define these two methods before reporting some empirical results on the performances of these two methods in their application to three real data sets and to some simulated data in a small-scale Monte Carlo experiment undertaken.

2 Theory

Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ denote an observed random sample of size n, where $\boldsymbol{x}_j$ $(j = 1, \ldots, n)$ is a vector of p dimensions. It is proposed to fit the g-component normal mixture model,

$$f(\boldsymbol{x};\, \boldsymbol{\Psi}) = \sum_{i=1}^{g} \pi_i \phi(\boldsymbol{x};\, \boldsymbol{\mu}_i,\, \boldsymbol{\Sigma}_i), \tag{1}$$

where $\phi(\boldsymbol{x}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ denotes the probability density function (p.d.f.) of the multivariate normal distribution with mean $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}_i$, and π_1, ..., π_g denote the proportions in which these g normal component densities occur in the mixture. Here we let $\boldsymbol{\Psi}$ be the vector of all the unknown parameters, partitioned as

$$\boldsymbol{\Psi} = (\boldsymbol{\pi}',\, \boldsymbol{\mu}_1',\, \ldots,\, \boldsymbol{\mu}_g',\, \xi_1',\, \ldots,\, \xi_g')', \tag{2}$$

where $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_{g-1})'$ and ξ_i is a vector containing the $\frac{1}{2}p(p+1)$ distinct elements of the covariance matrix $\boldsymbol{\Sigma}_i$ $(i = 1, \ldots, g)$. Thus $\boldsymbol{\Psi}$ consists of q

elements, where

$$q = (g-1) + \tfrac{1}{2}g(p^2+3p).$$

The log likelihood for $\boldsymbol{\Psi}$ is given by

$$\log L(\boldsymbol{\Psi}) = \sum_{j=1}^{n} \log\{\sum_{i=1}^{g} \pi_i \phi(\boldsymbol{x}_j;\, \boldsymbol{\mu}_i,\, \boldsymbol{\Sigma}_i)\}. \tag{3}$$

Solutions of the likelihood equation

$$\partial \log L(\boldsymbol{\Psi})/\partial \boldsymbol{\Psi} = \mathbf{0}$$

corresponding to local maxima can be found iteratively by application of the EM algorithm of Dempster, Laird, and Rubin (1977); see McLachlan and Basford (1988, Chapter 2). We let $\hat{\boldsymbol{\Psi}}$ be the estimator so obtained.

3 Information-Based Method

Under regularity conditions, the inverse of the covariance matrix of $\hat{\boldsymbol{\Psi}}$ is given approximately by the expected information

$$\mathcal{I}(\boldsymbol{\Psi}) = E\{\boldsymbol{I}(\boldsymbol{\Psi})\}, \tag{4}$$

where

$$\boldsymbol{I}(\boldsymbol{\Psi}) = -\partial^2 \log L(\boldsymbol{\Psi})/\partial \boldsymbol{\Psi} \partial \boldsymbol{\Psi}'. \tag{5}$$

Hence the covariance matrix of $\hat{\boldsymbol{\Psi}}$ can be approximated by $\mathcal{I}^{-1}(\hat{\boldsymbol{\Psi}})$. Although $\mathcal{I}(\hat{\boldsymbol{\Psi}})$ can be calculated for mixtures of any number of variables (Chang, 1979), it can involve substantial computation for large multivariate data sets.

Another way to proceed is to approximate the covariance matrix of $\hat{\boldsymbol{\Psi}}$ by the inverse of the observed information matrix, $\boldsymbol{I}^{-1}(\hat{\boldsymbol{\Psi}})$, which is common practice with estimation via maximum likelihood. Recently, Louis (1982), Meilijson (1989), Jones and McLachlan (1992), and Meng and Rubin (1993) have considered the calculation and approximation of the observed information matrix $\boldsymbol{I}(\hat{\boldsymbol{\Psi}})$ within the EM framework. As noted by McLachlan and Basford (1988, page 48), among others, it is computationally convenient in the case of independent data to approximate $\boldsymbol{I}(\hat{\boldsymbol{\Psi}})$ in terms of just the gradient vector of the log likelihood function. This approximation is given by

$$\boldsymbol{I}(\hat{\boldsymbol{\Psi}}) \approx \sum_{j=1}^{n} \hat{\boldsymbol{h}}_j \hat{\boldsymbol{h}}_j', \tag{6}$$

where

$$\hat{\boldsymbol{h}}_j = \partial \log L_j(\hat{\boldsymbol{\Psi}})/\partial \boldsymbol{\Psi} \tag{7}$$

is the gradient vector at $\boldsymbol{\Psi} = \hat{\boldsymbol{\Psi}}$ of the log likelihood based on the single observation vector $\boldsymbol{x}_j$.

If $\hat{\boldsymbol{\Psi}}$ is partitioned as in (2), then the corresponding partition of $\hat{\boldsymbol{h}}_j$ ($j = 1, \ldots, n$) is

$$\hat{\boldsymbol{h}}_j = (\hat{\boldsymbol{h}}'_{\pi_j}, \hat{\boldsymbol{h}}'_{\boldsymbol{\mu}_{1j}}, \ldots, \hat{\boldsymbol{h}}'_{\boldsymbol{\mu}_{gj}}, \hat{\boldsymbol{h}}'_{\xi_{1j}}, \ldots, \hat{\boldsymbol{h}}'_{\xi_{gj}})'. \tag{8}$$

The dth element of $\hat{\boldsymbol{h}}_{\pi_j}$ is given by

$$(\hat{\boldsymbol{h}}_{\pi_j})_d = \hat{\tau}_{dj}/\hat{\pi}_d - \hat{\tau}_{gj}/\hat{\pi}_g \qquad (d = 1, \ldots, g-1) \tag{9}$$

and

$$\hat{\boldsymbol{h}}_{\boldsymbol{\mu}_{ij}} = \hat{\tau}_{ij}\hat{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{x}_j - \hat{\boldsymbol{\mu}}_i) \qquad (i = 1, \ldots, g) \tag{10}$$

for $j = 1, \ldots, n$, where

$$\hat{\tau}_{ij} = \hat{\pi}_i \phi(\boldsymbol{x}_j;\, \hat{\boldsymbol{\mu}}_i,\, \hat{\boldsymbol{\Sigma}}_i)/f(\boldsymbol{x}_j;\, \hat{\boldsymbol{\Psi}}) \tag{11}$$

is the estimated posterior probability that the jth entity with observation vector $\boldsymbol{x}_j$ belongs to the ith component of the mixture.

To calculate $\hat{\boldsymbol{h}}_{\xi_{ij}}$, let $\boldsymbol{\Sigma}_i^{-1} = (\boldsymbol{\sigma}_i^{(1)}, \ldots, \boldsymbol{\sigma}_i^{(p)})$ for $i = 1, \ldots, g$, where $\boldsymbol{\sigma}_i^{(r)}$ denotes the rth column of $\boldsymbol{\Sigma}_i^{-1}$. If the mth element of $\hat{\boldsymbol{h}}_{\xi_{ij}}$ corresponds to differentiation with respect to $(\boldsymbol{\Sigma}_i)_{rs}$, the (r, s)th element of $\boldsymbol{\Sigma}_i (r \leq s)$, then

$$\begin{aligned}(\hat{\boldsymbol{h}}_{\xi_{ij}})_m &= \tfrac{1}{2}\hat{\tau}_{ij}(2 - \delta_{rs})[-(\hat{\boldsymbol{\Sigma}}_i^{-1})_{rs} \\ &\qquad + \{(\boldsymbol{x}_j - \hat{\boldsymbol{\mu}}_i)'\hat{\boldsymbol{\sigma}}_i^{(r)}\}\{(\boldsymbol{x}_j - \hat{\boldsymbol{\mu}}_i)'\hat{\boldsymbol{\sigma}}_i^{(s)}\}],\end{aligned} \tag{12}$$

where δ_{rs} is the Kronecker delta so that

$$\delta_{rs} = \begin{cases} 1, & (r = s) \\ 0, & (r \neq s). \end{cases} \tag{13}$$

Once the observed information matrix $\boldsymbol{I}(\hat{\boldsymbol{\Psi}})$ has been computed or approximated as above, the variance of each of the estimates in $\hat{\boldsymbol{\Psi}}$, partitioned according to (2), can be obtained from the corresponding diagonal elements of $\boldsymbol{I}^{-1}(\hat{\boldsymbol{\Psi}})$.

4 Bootstrap Method

The bootstrap method used to obtain estimates of the standard errors of the elements of the component-mean vectors $\hat{\boldsymbol{\mu}}_i$, has been developed by Efron (1979, 1981a, 1981b, 1982); see also Efron and Tibshirani (1993). Under a normal mixture model with g components, the parametric estimate of the distribution F of the observation vector is

$$\hat{F} = \sum_{i=1}^{g} \hat{\pi}_i \hat{F}_i, \tag{14}$$

where $\hat{F}_i$ is the distribution function for the p-dimensional normal distribution with mean $\hat{\boldsymbol{\mu}}_i$ and covariance matrix $\hat{\boldsymbol{\Sigma}}_i$; $\hat{\boldsymbol{\mu}}_i$, $\hat{\boldsymbol{\Sigma}}_i$, and $\hat{\pi}_i$ are the parameter estimates obtained from fitting a normal mixture model to the original data. A Monte Carlo approximation to the bootstrap estimate of the covariance matrix of the component-means can be obtained by

1) drawing a "bootstrap sample" $\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, \ldots, \boldsymbol{x}_n^*$ from the fitted mixture distribution $\hat{F}$ and fitting the mixture model to this sample to obtain the maximum likelihood estimate $\hat{\boldsymbol{\mu}}_i^*$ of the bootstrap component mean $\hat{\boldsymbol{\mu}}_i$ (i=1,...,g), and

2) repeating step 1 a number of times, say K, thereby obtaining bootstrap replications $\hat{\boldsymbol{\mu}}_i^{*(1)}, \hat{\boldsymbol{\mu}}_i^{*(2)}, \ldots, \hat{\boldsymbol{\mu}}_i^{*(K)}$ of the bootstrap component mean $\hat{\boldsymbol{\mu}}_i^* (i = 1, \ldots, g)$.

The sample covariance matrix of the bootstrap replicates of the fitted means is then calculated as

$$\mathbf{S}_i^{(B)} = \sum_{k=1}^{K} (\hat{\boldsymbol{\mu}}_i^{*(k)} - \bar{\hat{\boldsymbol{\mu}}}_i^*)(\hat{\boldsymbol{\mu}}_i^{*(k)} - \bar{\hat{\boldsymbol{\mu}}}_i^*)'/(K-1) \tag{15}$$

where

$$\bar{\hat{\boldsymbol{\mu}}}_i^* = \sum_{k=1}^{K} \hat{\boldsymbol{\mu}}_i^{*(k)}/K \qquad (i = 1, \ldots, g). \tag{16}$$

Thus $\mathbf{S}_i^{(B)}$ is an approximation to the bootstrap covariance matrix of $\hat{\boldsymbol{\mu}}_i^*$, and hence, to the covariance matrix of $\hat{\boldsymbol{\mu}}_i$.

5 Application to Real Data Sets

To demonstrate the calculation of the standard errors of the estimates $\hat{\mu}_{iv} = (\hat{\boldsymbol{\mu}}_i)_v, (i = 1, \ldots, g;\, v = 1, \ldots, p)$, the two methods were applied to three sets of real data, two of which were multivariate. The normal mixture model was fitted assuming both equal and arbitrary covariance structure for each component. However, as the results were similar, only those obtained for the case of arbitrary component-covariance matrices are presented here.

All calculations for the information-based method were carried out using amended versions of the FORTRAN program KMM (listed in the Appendix of McLachlan and Basford, 1988). The amendments provide an assessment of the covariance matrix of $\hat{\boldsymbol{\mu}}_i$ from the observed information matrix approximated according to (6).

As noted by Efron (1979, 1981a, 1981b),the number of bootstrap replications K does not seem to be critical past 50 or 100 in the task of providing estimates of the standard errors. Here $K = 100$ was used in all examples. The 100 randomly generated sets of data were obtained, each set representing a bootstrap sample arising from a mixture of normal distributions, with parameters equal to the parameter estimates obtained from the original data. The mixture model was fitted to these data sets in the same way as to the original data, and the bootstrap estimates of the standard errors of the $\hat{\mu}_{iv}$ calculated via (15).

The standard errors of the estimated component means $\hat{\mu}_{iv}$ obtained by each of the two methods for each data set are presented separately.

We also considered the construction of confidence intervals for the μ_{iv}. Asymptotic normal theory leads to approximate confidence intervals of the form

$$\mu_{iv} = \hat{\mu}_{iv} \pm z_{1-\alpha} s_{iv}, \tag{17}$$

where $z_{1-\alpha}$ denotes the quantile of order $(1-\alpha)$ of the standard normal distribution and s_{iv} denotes the estimate adopted for the standard error of $\hat{\mu}_{iv}$. As noted above, the reliance on asymptotic theory is of particular concern for mixture models. Another way of providing confidence intervals is to adopt a nonparametric approach by using the bootstrap.

The distribution function of $\hat{\mu}^*_{iv}$ can be approximated by

$$\widehat{CDF}_{iv}(u) = \#\{\hat{\mu}^{*(k)}_{iv} \leq u\}/K \quad (i = 1, \ldots, g;\, v = 1, \ldots, p), \tag{18}$$

the empirical distribution function formed from the K bootstrap replications $\hat{\mu}^{*(1)}_{iv}, \ldots, \hat{\mu}^{*(K)}_{iv}$ of $\hat{\mu}^*_{iv}$.

With the percentile bootstrap method of Efron (1981b, 1982), a nominal $100(1-\alpha)\%$ confidence interval for μ_{iv} is given by

$$[\widehat{CDF}^{-1}_{iv}(\tfrac{1}{2}\alpha),\ \widehat{CDF}^{-1}_{iv}(1-\tfrac{1}{2}\alpha)], \tag{19}$$

where

$$\widehat{CDF}^{-1}_{iv}(\alpha) = \sup\{u :\ \widehat{CDF}_{iv}(u) \leq \alpha\}. \tag{20}$$

That is, the percentile method interval consists of the central $1-\alpha$ proportion of the (simulated) bootstrap distribution of the $\hat{\mu}^*_{iv}$. As explained by Efron (1982), setting confidence intervals is a harder problem than standard error estimation. Hence a larger number of bootstrap replications is required; see also Hall (1986a, 1986b, 1988) and Efron (1987), who consider ways of improving the percentile bootstrap confidence intervals.

In the examples below, the bootstrap confidence intervals (19) are to be contrasted with the approximate confidence intervals of the form (17), using the information-based and bootstrap estimates of the standard errors. The nominal value of these confidence intervals was taken to be 90%.

Example 1

This data set contains 485 observations on stamp thickness. The stamps were believed to come from seven separate, normally distributed populations (Izenman & Sommer, 1988). Accordingly, a mixture of $g = 7$ components was fitted to the data and standard errors and corresponding confidence intervals of the component means calculated using the two methods. The results are presented in Table 1. Here, the observed information matrix contained 20×20 elements.

Table 1: Standard errors of estimated component means obtained from fitting a normal mixture model to the stamp data of Izenman & Sommer (1988).

Component	Number assigned to component	Estimated component mean	Standard error of mean	
			Method	
			Information	Bootstrap
1	150	0.07242	.00078	.00038
2	156	0.07970	.00032	.00022
3	45	0.09048	.00053	.00053
4	61	0.10025	.00048	.00046
5	52	0.10949	.00066	.00066
6	15	0.12078	.00092	.00092
7	6	0.12934	.00129	.00051

As this example has a large n and reasonably well separated means, it was considered likely that the information-based method would provide reliable estimates of the standard errors of the fitted parameters. However, comparison with those obtained from the bootstrap method shows that it provides substantially larger estimates of the standard error for both Component 1 and Component 7. As Component 7 contained only 6 individuals, it could be expected that this method based on asymptotic theory may not provide a reliable result. However, the reason for the large estimate obtained for Group 1 is not clear.

Plots showing 90% confidence intervals based on the standard errors obtained from each method are presented (Figure 1) and show a clear separation between components regardless of the method used.

Example 2

The haemophilia data of Habbema, Hermans, and van den Broek (1974) consist of 75 bivariate observations. The observations are partitioned into 2 groups, representing normal women and carriers of haemophilia A, by the

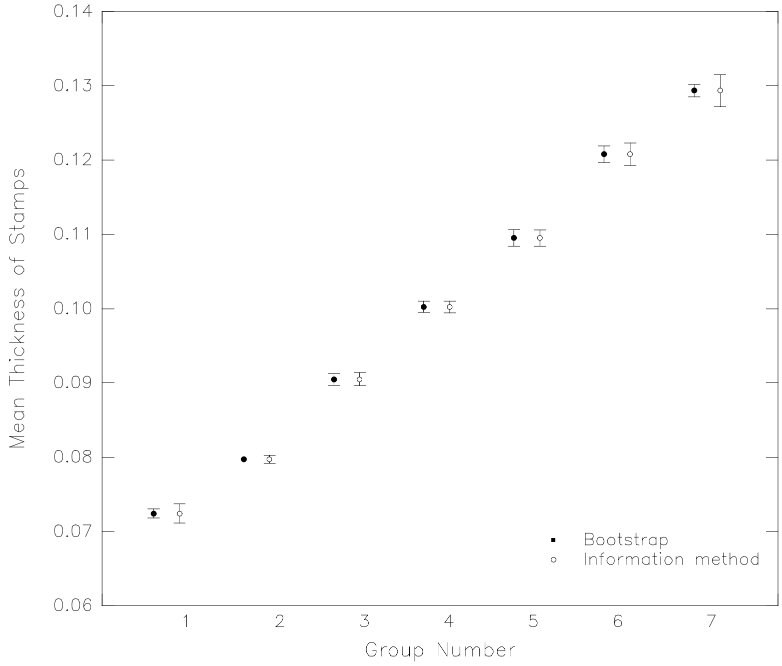

Figure 1. Plot of the estimated group means together with the 90% confidence intervals obtained by each method for the Stamp data in Example 1.

fitting of a two-component bivariate normal mixture. Calculated standard errors are presented below (Table 2).

Here the dimensions of the observed information matrix are 11×11. In this instance n is not particularly large and the estimated component means for Attribute 2 are not well separated. Despite this, the information method appears to have provided sensible results. Examination of the plots of the confidence intervals obtained from the two methods (Figure 2) show that the means of Group 1 and Group 2 are well-separated for Attribute 1 but not Attribute 2.

Example 3

This example uses the well-known data set considered by Fisher (1936), which consist of measurements of the length and width of both sepals and petals of 50 plants from each of 3 species of *Iris*, namely *virginica, versicolor* and *setosa*. Calculated standard errors of estimated component means are presented below (Table 3).

Unlike the previous two examples, the standard errors obtained from both

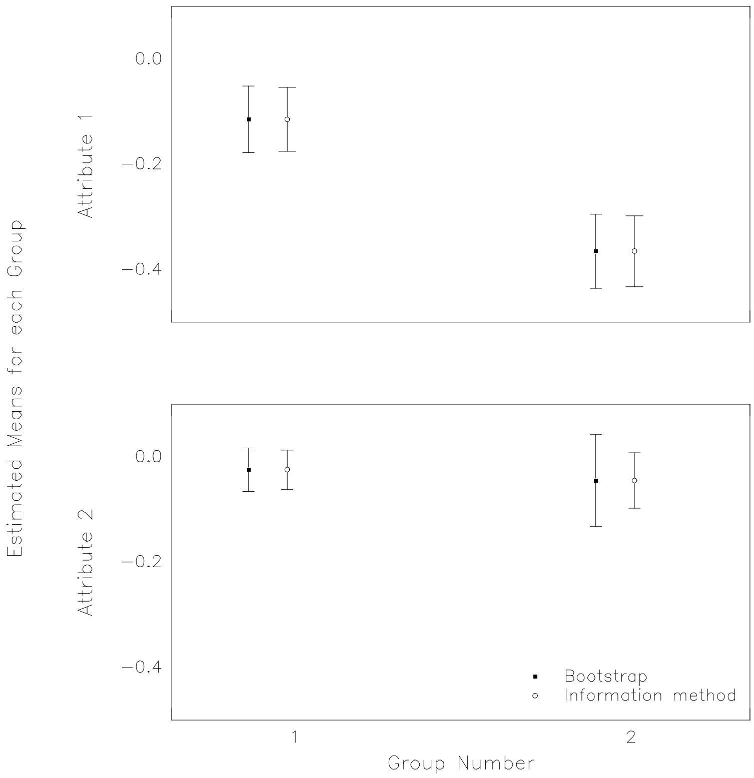

Figure 2. Plot of the estimated group/attribute means together with the 90% confidence intervals obtained by each method for the Haemophilia data in Example 2.

Table 2: Standard errors of estimated component means obtained from fitting a normal mixture model to the Haemophilia data of Habbema, Hermans and van den Broek (1974).

Component	Attribute	Number assigned to component	Estimated component mean	Standard error of mean — Method: Information	Bootstrap
1	1	39	-0.114	0.037	0.042
	2		-0.024	0.023	0.025
2	1	36	-0.365	0.041	0.042
	2		-0.045	0.032	0.058

Table 3: Standard errors of estimated component means obtained from fitting a normal mixture model to the *Iris* data of Fisher (1936)

Component	Attribute	Number assigned to component	Estimated component mean	Standard error — Method: Information	Bootstrap
1	1	50	5.006	0.057	0.040
	2		3.428	0.059	0.049
	3		1.462	0.030	0.023
	4		0.246	0.020	0.013
2	1	45	5.914	0.103	0.069
	2		2.777	0.056	0.040
	3		4.201	0.097	0.076
	4		1.296	0.033	0.034
3	1	55	6.544	0.108	0.070
	2		2.948	0.049	0.041
	3		5.479	0.103	0.072
	4		1.984	0.043	0.045

methods are of comparable magnitude. The plots of the confidence intervals reveal that the data can be separated into 3 distinct groups for each of the four attributes regardless of which estimate of standard error was used (Figure 3).

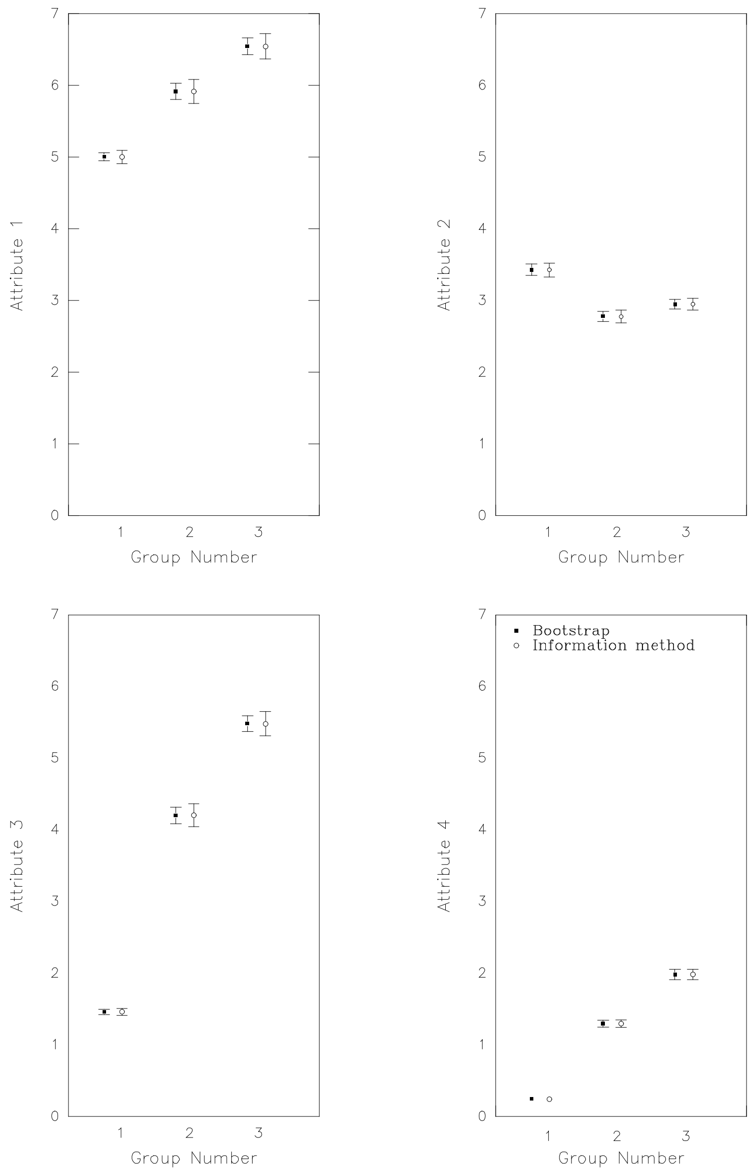

Figure 3. Plot of the estimated group/attribute means together with the 90% confidence intervals obtained by each method for the Iris data in Example 3.

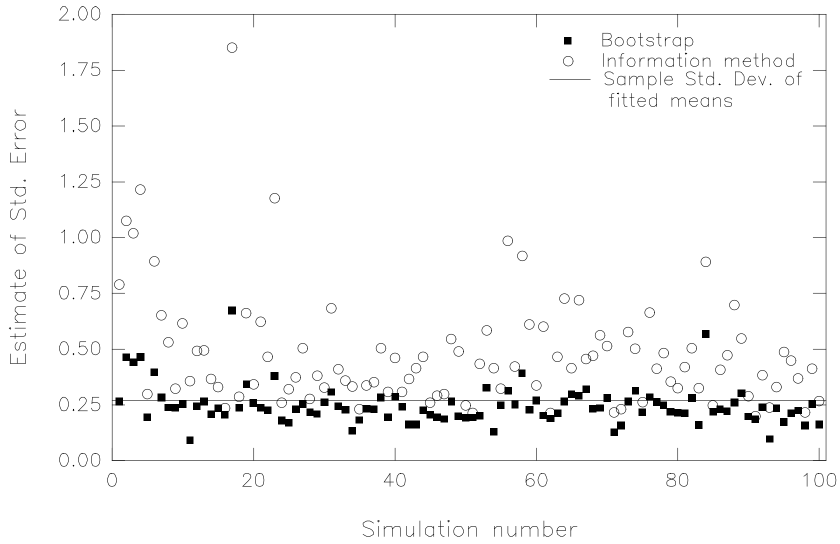

Figure 4a. Estimates of standard errors of Group 1 Attribute 1 means for the Monte Carlo experiment.

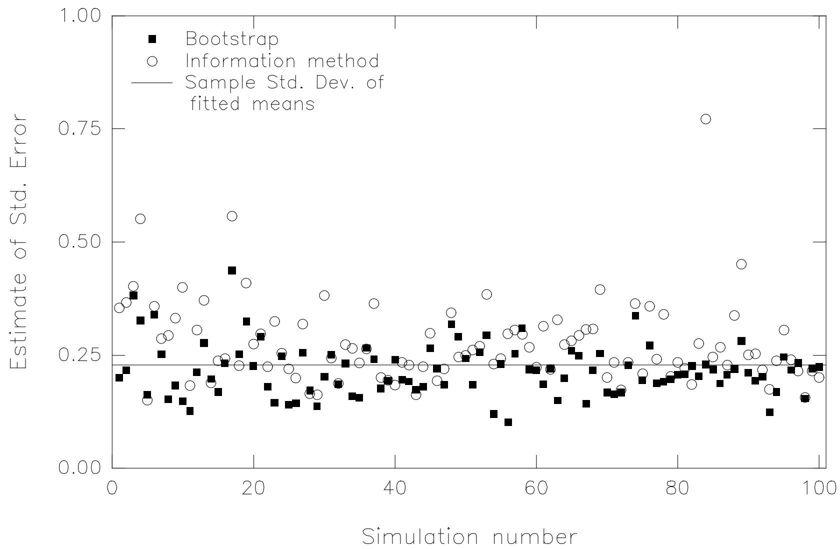

Figure 4b. Estimates of standard errors of Group 1 Attribute 2 means for the Monte Carlo experiment.

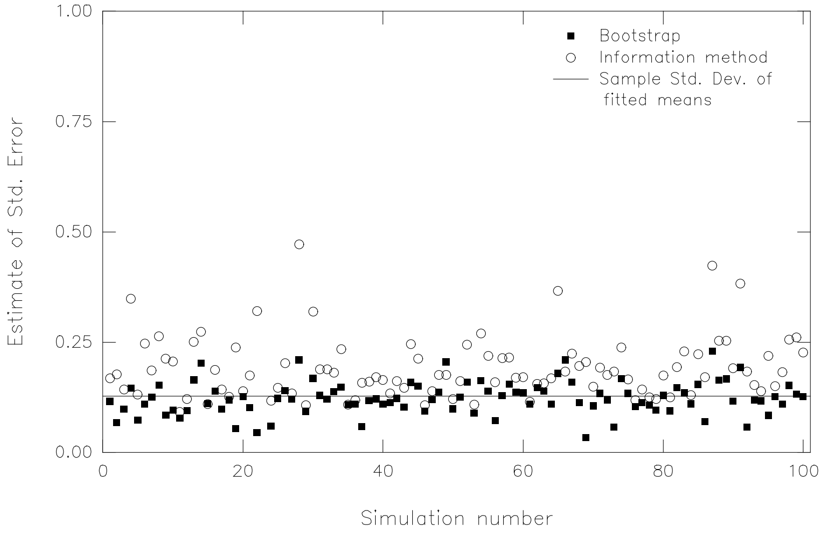

Figure 4c. Estimates of standard errors of Group 2 Attribute 1 means for the Monte Carlo experiment.

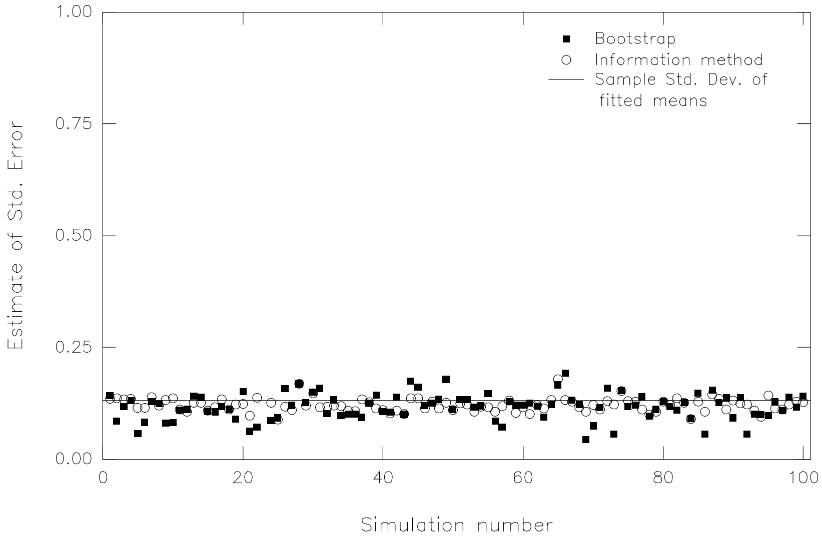

Figure 4d. Estimates of standard errors of Group 2 Attribute 2 means for the Monte Carlo experiment.

6 Monte Carlo Experiment

In addition to the above, a small-scale Monte Carlo experiment was carried out. In these simulations, 100 bivariate samples each consisting of $n = 100$ observations were generated. These observations were generated from a mixture of two bivariate normal component distributions with means $(0, 0)'$ and $(3, 0)'$ in the proportions 0.25 and 0.75, respectively. The component distributions were taken to have a common covariance matrix set equal to the identity matrix.

In applying the mixture model to the generated samples the same program KMM was used. However, the covariance matrices for the components were assumed to be equal.

Standard errors of the $\hat{\mu}_{iv}$ were calculated for each of the 100 bivariate samples using the two methods previously described. To provide a basis for comparing the methods, the sample standard deviation of each $\hat{\mu}_{iv}$ over the 100 simulation trials was also calculated. These results are plotted in Figure 4. As expected, the bootstrap method provided the better estimates of the standard errors of the $\hat{\mu}_{iv}$, although the method tended to underestimate their true size in some instances. Variable results were obtained from the information method which frequently overestimated the magnitude of the standard error, in some instances to such a degree that the results cause concern for the practical use of this procedure.

7 Discussion

Two methods of calculating estimates of the standard errors of the elements of the fitted component means $\hat{\boldsymbol{\mu}}_i$ have been considered empirically on the basis of some real and simulated data sets. The potential of the bootstrap in providing useful estimates of the standard errors of the elements of the $\hat{\boldsymbol{\mu}}_i$ is demonstrated. Concerning the less demanding (computational-wise) information-based method, the limitations of asymptotic theory were evidenced by the apparent failure of this method to always provide realistic results both in the real examples and in the Monte Carlo experiment.

In the special case of mixture models with components that are very widely separated, the fitted posterior probabilities $\hat{\tau}_{ij}$ of component membership of the mixture model are either close to zero or one. Hence then the covariance matrix of the fitted component mean $\hat{\boldsymbol{\mu}}_i$ can be approximated simply by

$$\hat{\boldsymbol{\Sigma}}_i / \sum_{j=1}^{n} \hat{\tau}_{ij}.$$

Acknowledgement

The authors would like to thank Pieter Kroonenberg for discussions on earlier versions of this paper.

References

Chang, W.C. (1976). The effects of adding a variable in dissecting a mixture of two normal populations with a common covariance matrix. *Biometrika*, **63**, 676–678.

Chang, W.C. (1979). Confidence interval estimation and transformation of data in a mixture of two multivariate normal distributions with any given large dimension. *Technometrics*, **21**, 351–355.

Day, N.E. (1969). Estimating the components of a mixture of two normal distributions. *Biometrika*, **56**, 463–474.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. R. Statist. Soc. B.*, **39**, 1–38.

Dick, N.P. and Bowden, D.C. (1973). Maximum likelihood estimation for mixtures of two normal distributions. *Biometrics*, **29**, 781–790.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.*, **7**, 1–26.

Efron, B. (1981a). Nonparametric estimates of standard error, the jackknife, the bootstrap and other methods. *Biometrika*, **68**, 589–599.

Efron, B. (1981b). Nonparametric standard errors and confidence intervals (with discussion). *Canad. J. Statist.*, **9**, 139–172

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans.* Society for Industrial and Applied Mathematics, Philadelphia.

Efron, B. (1987). Better bootstrap confidence intervals (with discussion). *J. Amer. Statist. Assoc.*, **82**, 171–200.

Efron, B. and Hinkley, D.V. (1978). Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher information. *Biometrika*, **65**, 457–487.

Efron, B. and Tibsharini, R.J. (1993). *An Introduction to the Bootstrap.* Chapman & Hall, London.

Fisher, R.A. (1936). The use of multiple measurements in taxonomic problems. *Ann. Eug.*, **8**, 179–188.

Habbema, J.D.F., Hermans, J., and van den Broek, K. (1974). A stepwise discriminant analysis program using density estimation. *Compstat 1974, Proc. Computational Statistics.* Vienna: Physica-Verlag, pp. 101–110.

Hall, P. (1986a). On the bootstrap and confidence intervals. *Ann. Statist.* , **14**, 1431–1452.

Hall, P. (1986b). On the number of bootstrap simulations required to construct a confidence interval. *Ann. Statist.*, **14**, 1453–1462.

Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals (with discussion). *Ann. Statist.*, **16**, 927–985.

Hosmer, D.W. (1973a). On MLE of the parameters of a mixture of two normal distributions when the sample size is small. *Commun. Statist.*, **1**, 217–227.

Hosmer, D.W. (1973b). 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**, 761–770.

Izenman, A.J. and Sommer C.J. (1988). Philatelic mixtures and multimodal densities. *J. Amer. Statist. Assoc.*, **83**, 941–953.

Jones, P.N. and McLachlan, G.J. (1992). Improving the convergence rate of the EM algorithm for a mixture model fitted to grouped truncated data. *J. Statist. Comput. Simul.*, **43**, 31–44.

Louis, T.A. (1982). Finding the observed information matrix when using the EM algorithm. *J. R. Statist. Soc. B.*, **44**, 226–233.

McLachlan, G.J. and Basford, K.E. (1988). *Mixture Models Inference and Applications to Clustering.* Marcel Dekker, Inc., New York.

Meilijson I. (1989). A fast improvement to the EM algorithm on its own terms. *J. R. Statist. Soc. B*, **51**, 127–138.

Meng, X.L. and Rubin, D.B. (1991). Using EM to obtain asymptotic variance-covariance matrices: the SEM algorithm. *J. Amer. Statist. Assoc.*, **86**, 899–909.

Tan, W.Y. and Chang, W.C. (1972). Some comparisons of the method of moments and the method of maximum likelihood in estimating parameters of a mixture to two normal densities. *J. Amer. Statist. Assoc.* , **67**, 702–708.

Tukey, J.W. (1986). Sunset Salvo. *Amer. Statist.*, **40**, 72–76.