Comparison of the mixture and the classification maximum likelihood in cluster analysis

Gilles Celeux, Gérard Govaert

UNITÉ DE RECHERCHE
INRIA-ROCQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P.105
78153 Le Chesnay Cedex
France
Tél.:(1) 39 63 55 11

Rapports de Recherche

N° 1517

Programme 5
Traitement du Signal,
Automatique et Productique

COMPARISON OF THE MIXTURE AND THE CLASSIFICATION MAXIMUN LIKELIHOOD IN CLUSTER ANALYSIS

Gilles CELEUX
Gérard GOVAERT

Septembre 1991

RR-1517

Comparison of the Mixture and the Classification Maximum Likelihood in Cluster Analysis*

Comparaison des approches vraisemblance et vraisemblance classifiante en classification

Gilles Celeux
INRIA
Rocquencourt

Gérard Govaert
UTC-URA CNRS 817
Compiègne

Abstract

Generally, the mixture and the classification approaches via maximum likelihood had been contrasted under different underlying assumptions. In the classification approach, the mixing proportions are assumed to be equal whereas, in the mixture approach, there are supposed to be unknown. In this paper, Monte-Carlo numerical experiments comparing both approaches, mixture and classification, in both assumptions, equal and unknown mixing proportions are reported. These numerical experiments exhibited that assumptions on the mixing proportions is a more sensitive factor than the choice of the clustering approach, especially in the small sample setting. Moreover, the differences between the finite sample and the asymptotic behaviour of both approaches are analyzed through additional simulations.

Keywords: Maximum likelihood estimates; Mixture and classification clustering approaches; Equal or unknown mixing proportions; Simulation comparison.

Résumé

En général, la comparaison des approches maximum de vraisemblance et vraisemblance classifiante pour les méthodes de classification fondées sur un modèle de mélange a été faite sous des hypothèses différentes. La vraisemblance classifiante est souvent restreinte à des mélanges dont les proportions sont égales alors que l'approche par le maximum de vraisemblance est considérée sans restriction. Dans cet article, nous présentons des simulations de Monte-Carlo où les deux approches sont comparées sous les mêmes hypothèses, proportions égales et proportions inconnues. Ces simulations montrent que l'hypothèse sur les proportions du mélange est plus importante que le choix du type de méthode. De plus, les différences entre les comportements à taille finie et asymptotique sont analysées à l'aide de simulations complémentaires.

Mots-clés : maximum de vraisemblance et vraisemblance classifiante en classification ; proportions égales ou inconnues ; comparaison par simulation.

---

* Prepared for an invited communication at the 3rd Conference of the International Federation of Classification Societies, Edinburgh, Scotland. August 1991.

1. Introduction and Motivation

Many authors (see Scott and Symons 1971, Marriott 1975, 1982, Symons 1981, McLachlan 1982, McLachlan and Basford 1988, among others) have considered nonhierarchical clustering methods in which a mixture of multivariate normal distributions is used as a statistical model. In this context, two commonly used maximum likelihood (m.l.) approaches have been proposed: the mixture approach and the classification approach. Loosely speaking, the mixture approach is aimed to maximize the likelihood over the mixture parameters, whereas the classification approach is aimed to maximize the likelihood over the mixture parameters *and* over the identifying labels of the mixture component origin for each point.

It is important to point out that, before the paper of Symons (1981), the classification approach assumed implicitly that the mixing proportions were equal in the mixture model. And, actually, most of the studies which compared both approaches (Marriott 1975, Bryant and Williamson 1978, 1986, McLachlan 1982, Ganesalingam 1989) restricted attention to the "standard" classification approach with "equal mixing proportions" but placed no restriction on the mixture approach. Thus, Ganesalingam (1989) have performed extensive numerical experiments to contrast the behaviour of the standard classification approach and the mixture approach in practical situations. But, as far as we can know, there is no published simulation comparison of both approaches under the same assumptions for the underlying mixture model.

On the other hand, Marriott (1975) and Bryant and Williamson (1978), showed that when estimation of the mixture parameters is of primary interest, then the standard classification maximum likelihood method is inconsistent. These authors have discussed the reasons of the bias of the standard classification approach estimates which can be serious. The sources of the bias are "all-or-nothing" classification, then the number of parameters increasing indefinitely with the sample size since the identifying labels of the mixture components are considered as parameters when using classification approach, and, at last, the mixture component treated equally by the standard classification approach. The first two reasons are inherent in the classification approach, but the third one concerns clearly the implicit restrictive assumption on the mixing proportions. From the studies of Bryant and Williamson (1986) and Ganesalingam (1989), it appears that, in practical situations, the standard classification approach performs well if the mixture components are present in approximately the same proportion in the population, and, that otherwise the mixture approach should be preferred. But there is no study analyzing the respective part of the three above mentioned reasons for the bias of the mixture parameters estimates from the standard classification approach.

Now, Symons (1981) has proposed a general classification maximum likelihood criterion which places no restriction on the mixing proportions. This unrestricted classification approach can be expected to perform better and to reduce dramatically the bias of the standard classification approach. Some numerical experiments reported in Symons (1981) show that actually the unrestricted classification approach did not favor partitions of equal sizes as the standard classification approach. However, Symons noted a tendency of the unrestricted classification approach to overemphasize slightly the size of the larger groups. This tendency has been confirmed in the paper of Bryant (1991) concerning large sample results in nonhierarchical clustering. Bryant showed that, asymptotically, the unrestricted classification approach did not classify at all for ill-separated components or extreme values of the mixing proportions.

Finally, we think that there is a need to compare the practical behaviour of the classification and the mixture approaches to clustering via maximum likelihood by considering both approaches with, firstly, equal mixing proportions, and secondly,

unknown mixing proportions. Note that Ganesalingam (end of Section 4.1 p. 464, 1989) had previously suggested to perform the mixture approach with equal mixing proportions but it did not provided numerical illustrations. Thus, the aim of the present paper is to extend the studies of Bryant and Williamson (1986) and Ganesalingam (1989) for the comparison of both approaches in a finite sample setting by a series of simulations.

In Section 2, we introduce the likelihood criteria to be optimized under both approaches in a general framework and we focus on Gaussian mixtures with equal mixing proportions and with unknown mixing proportions. Section 3 is devoted to the presentation of the Monte-Carlo simulations. First, we present the experiment plan closely related to the simulation scheme of Ganesalingam (1989). The main differences consisted in that we considered more sample sizes and we covered the factor concerning the dependence of the used algorithms over their initial position. Then, we presented the numerical experiments results. The paper is concluded with a discussion Section in which the main points of the study are summarized.

2. The Two Approaches

Clustering methods based on maximum likelihood consider the situation where the data are $\mathbf{R}^d$-valued vectors $\mathbf{x}_1, ..., \mathbf{x}_n$ assumed to be a sample from a mixture of densities

$$\varphi(\mathbf{x}) = \sum_{k=1}^{K} p_k \, f(\mathbf{x}, \mathbf{a}_k) \tag{2.1}$$

where the p_k's are the mixing proportions ($0 < p_k < 1$ for all $k = 1, ..., K$ and $\Sigma_k p_k = 1$) and the $f(\mathbf{x}, \mathbf{a}_k)$ are densities from the same parametric family. In the following, $f(\mathbf{x}, \mathbf{a}_k)$ will denote the d-dimensional normal density with unknown mean μ_k and covariance matrix Γ and $\mathbf{a}_k = (\mu_k, \Gamma)$. This situation, also considered by Ganesalingam (1989), provides a parsimonious and commonly used clustering model (Friedman and Rubin 1967, Scott and Symons 1971, Symons 1981).

2.1 The Mixture Approach

In the mixture maximum likelihood approach, the parameters p_k and $\mathbf{a}_k$ are chosen to maximize the log-likelihood

$$L_M = \sum_{i=1}^{n} \log\{\sum_{k=1}^{K} p_k \, f(\mathbf{x}_i, \mathbf{a}_k)\}. \tag{2.2}$$

In the restricted case where the mixing proportions are assumed to be equal, the criterion to be maximized takes the form

$$L_{MR} = \sum_{i=1}^{n} \log\{\sum_{k=1}^{K} f(\mathbf{x}_i, \mathbf{a}_k)\}. \tag{2.3}$$

It does not seem that there is any interest to detail the expressions (2.2) and (2.3) for the Gaussian mixture under consideration in this paper.

The value of the parameters μ_k, Γ and p_k (resp. μ_k, Γ) are chosen to maximize equation (2.2) (resp. (2.3)) using, generally, the EM algorithm (Dempster, Laird and Rubin 1977). In this approach, a partition $P = (P_1, ..., P_K)$ of the data can directly be derived from the m.l. estimates of the mixture parameters by assigning each $\mathbf{x}_i$ to the component which provides the greatest posterior probability that $\mathbf{x}_i$ arises from it. The estimated posterior probability that $\mathbf{x}_i$ arises from the kth component is given by

$$t_k(\mathbf{x}_i) = \frac{\hat{p}_k \, f(\mathbf{x}_i, \hat{\mu}_k, \hat{\Gamma})}{\sum_{j=1}^{K} \hat{p}_j \, f(\mathbf{x}_i, \hat{\mu}_j, \hat{\Gamma})} \qquad k = 1, ..., K \; ; \; i = 1, ..., n \quad (2.4)$$

for the unrestricted model and by

$$t_k(\mathbf{x}_i) = \frac{f(\mathbf{x}_i, \hat{\mu}_k, \hat{\Gamma})}{\sum_{j=1}^{K} f(\mathbf{x}_i, \hat{\mu}_j, \hat{\Gamma})} \qquad k = 1, ..., K \; ; \; i = 1, ..., n \quad (2.5)$$

for the restricted model.

2.2 *The Classification Approach*

In the classification maximum likelihood (CML) approach, the indicator vectors $\mathbf{z}_i = (z_{ij}, j = 1, ..., K)$ with $z_{ij} = 1$ or 0 according as $\mathbf{x}_i$ $(1 \leq i \leq n)$ has been drawn from the jth component or from another one, identifying the mixture component origin, are treated as unknown parameters. Two different CML criteria have been proposed according to the sampling scheme.

Under the separate sampling scheme, the sample $\mathbf{x}_1, ..., \mathbf{x}_n$ is formed by separately taking n_k observations from the kth component where n_k is fixed before sampling. In this formulation, the proportions p_k's do not appear explicitly and, thus, they are implicitly assumed to be equal. In this situation, the restricted CML criterion takes the form (see, for instance, Scott and Symons 1971)

$$L_{CR} = \sum_{k=1}^{K} \sum_{\mathbf{x}_i \in P_k} \log f(\mathbf{x}_i, \mathbf{a}_k) \quad (2.6)$$

where $P = (P_1, ..., P_K)$ is a partition of $\mathbf{x}_1, ..., \mathbf{x}_n$ associated to the indicator vectors $\mathbf{z}_1, ..., \mathbf{z}_n$: $P_k = \{\mathbf{x}_i \,/\, z_{ik} = 1\}$.

Under the mixture sampling, the sample $\mathbf{x}_1, ..., \mathbf{x}_n$ is taken at random from the mixture density (2.1), so that the number of observations from the components has a multinomial distribution with sample size n and probability parameters $p_1, ..., p_K$. In this situation, the CML criterion takes the form (Symons 1981)

$$L_C = \sum_{k=1}^{K} \sum_{\mathbf{x}_i \in P_k} \log \{p_k \, f(\mathbf{x}_i, \mathbf{a}_k)\} \quad (2.7)$$

which can be written

$$L_C = L_{CR} + \sum_{k=1}^{K} \#P_k \log p_k \ . \tag{2.8}$$

For a Gaussian mixture where the covariance matrices are constant across all clusters, it can be proved (Scott and Symons 1971) that maximizing the restricted CML criterion L_{CR} is equivalent to minimizing the determinant criterion $|W|$ where W is the within-group scatter matrix

$$W = \sum_{k=1}^{K} \sum_{\mathbf{x}_i \in P_k} (\mathbf{x}_i - \bar{\mathbf{x}}_k)(\mathbf{x}_i - \bar{\mathbf{x}}_k)' \tag{2.9}$$

where

$$\bar{\mathbf{x}}_k = \frac{1}{\#P_k} \sum_{\mathbf{x}_i \in P_k} \mathbf{x}_i \ . \tag{2.10}$$

In the same manner, it can be proved (Symons 1981) that maximizing the unrestricted CML criterion L_C is equivalent to minimizing

$$n \log |W| - 2 \sum_{k=1}^{K} \#P_k \log\{\#P_k\} . \tag{2.11}$$

Both criteria can be optimized by standard iterative partitioning algorithms. For our experiments, we ran a classification version of the EM algorithm, the so-called CEM algorithm (Celeux and Govaert 1991) that we described briefly in the unrestricted situation (maximization of L_C):

Starting from an initial partition P^0, the mth iteration of CEM ($m > 0$) is defined as follows:

E-step: Compute for $i = 1, ..., n$ and $k = 1, ..., K$ the current posterior probabilities that $\mathbf{x}_i$ belongs to P_k: $t_k^m(\mathbf{x}_i) = p_k^m f(\mathbf{x}_i, \mathbf{a}_k^m) / \sum_j p_j^m f(\mathbf{x}_i, \mathbf{a}_j^m)$ for the current parameter estimates.

C-step: Assign each $\mathbf{x}_i$ to the cluster which provides the maximum posterior probability $t_k^m(\mathbf{x}_i)$, $1 \le k \le K$, (if the maximum posterior probability is not unique, we choose the cluster with the smallest index). Let P^m denote the resulting partition.

M-step: For $k = 1, ..., K$ compute the m.l. estimates $(p_k^{m+1}, \mathbf{a}_k^{m+1})$ using the sub-samples P_k^m.

3. Simulation Experiments

As we pointed out in the introduction, numerical experiments comparing the mixture approach and the classification approach concerned the restricted classification approach and the unrestricted mixture approach and there is no information in the

literature on the relative performance of both methods for the same underlying model in the finite sample setting. The presented simulation experiments aimed to provide information on which approach is preferable in different circumstances and to extend the Ganesalingam's study.

In the following, the restricted classification approach will be abbreviated CAR and the unrestricted classification approach CA. In the same manner, the restricted mixture approach will be abbreviated MAR and the unrestricted mixture approach MA.

3.1 Experiment Conditions

Our simulation experiments were carried out along almost the same program used by Ganesalingam (1989). We generated a two-component Gaussian mixture with a common covariance matrix $\Gamma = I$ and mean vectors $\mu_1 = -\mu_2 = (1/2\,\Delta, 0, ..., 0)'$ where Δ is the Mahalanobis distance between the two subpopulations with means μ_1 and μ_2 and common covariance matrix Γ and is given by $\Delta = \{(\mu_1 - \mu_2)'\Gamma^{-1}(\mu_1 - \mu_2)\}^{1/2}$. 30 simulated trials were performed over 216 different combinations of the parameters Δ = 1, 2, 3; p_1 = 0.25, 0.35, 0.50; d = 1, 2, 4; n = 20, 40, 100, 200; two sampling schemes: mixture sampling and separate sampling. It is worth to remark that we considered more sample sizes than Ganesaligam: he considered only small sample sizes (20 and 40) and we also considered moderate sample sizes (100 and 200).

Moreover, we considered a factor generally neglected by authors who have performed simulation studies in cluster analysis. This factor, which is of considerable practical interest, is the dependence of commonly used optimization algorithms over their initial position for the different approaches. Before describing the way we proceeded to deal with this factor, it is worth noting that it is always possible to run algorithms which aim to reduce the initial-position dependence such as simulated annealing and the stochastic versions of EM (Celeux and Diebolt 1985) and CEM (Celeux and Govaert 1991). But, in the present paper, we are concerned with the statistical behaviour of two clustering approaches; thus, for the sake of simplicity, we only considered algorithms available in current software libraries. The reader is referred to the two above mentioned references for a discussion on the optimization problem. We proceeded as follows to take the initial-position dependence factor in consideration. We performed two series of 216 numerical experiments. In the first series, we initiated the EM and the CEM algorithms with the theoretical parameter values of the involved mixtures. Acting in such a way, we drastically attenuated the initial-position dependence of the solutions. In the second series, for each generated sample, we ran each algorithm 20 times from random initial position and selected the solution out of the 20 runs which provided the best value of the optimized criterion. In such a classical scheme, the initial-position dependence of the solutions can be expected to be important in some situations.

3.2 Assessing the Clustering Performances

Since we are concerned with the mixture approach and the classification approach as clustering procedures, the performances have been assessed on the basis of the overall expected error rate. Let $\hat{\mu}_1$, $\hat{p}_2$, $\hat{\Gamma}$, $\hat{p}_1$ and $\hat{p}_2$ denote the mixture parameter estimates from one of the discussed methods. (Remark: if the concerned method is CAR or MAR, the estimated of p_1 and p_2 are $\hat{p}_1 = \#P_1/n$ and $\hat{p}_2 = \#P_2/n$, where P_1 and P_2 are the clusters obtained from CAR or MAR). The overall expected error rate associated to these estimates is

$$R = p_1\ \Phi\left[\frac{-\{\mu_1 - 1/2\,(\hat{\mu}_1+\hat{\mu}_2)\}'\ \hat{\Gamma}^{-1}\,(\hat{\mu}_1-\hat{\mu}_2) + s}{\{(\hat{\mu}_1-\hat{\mu}_2)'\ \hat{\Gamma}^{-1}\,\Gamma\,\hat{\Gamma}^{-1}\,(\hat{\mu}_1-\hat{\mu}_2)\}^{1/2}}\right]$$
$$+p_2\ \Phi\left[\frac{-\{\mu_2 - 1/2\,(\hat{\mu}_1+\hat{\mu}_2)\}'\ \hat{\Gamma}^{-1}\,(\hat{\mu}_2-\hat{\mu}_1) + s}{\{(\hat{\mu}_1-\hat{\mu}_2)'\ \hat{\Gamma}^{-1}\,\Gamma\,\hat{\Gamma}^{-1}\,(\hat{\mu}_1-\hat{\mu}_2)\}^{1/2}}\right] \qquad (3.1)$$

with $s = \log(\hat{p}_1/\hat{p}_2)$ and where Φ denotes the cumulative distribution function of a standard Gaussian random variable with mean 0 and standard deviation 1. If the mixture parameter estimates were exactly the true values, the overall expected error rate would simplify in

$$R^* = p_1\,\Phi(-\frac{\Delta}{2}+\frac{s^*}{\Delta}) + p_2\,\Phi(-\frac{\Delta}{2}-\frac{s^*}{\Delta}) \qquad (3.2)$$

with $s^* = \log(p_1/p_2)$.

The overall error rates obtained over the simulations have to be compared with the associated 'ideal' error rate provided by (3.2). Table 1 gives 'ideal' misclassification risks for the situations considered in our numerical experiments.

p_1 \ Δ	1	2	3
0.25	0.222	0.127	0.055
0.35	0.277	0.148	0.063
0.50	0.309	0.159	0.067

Table 1. Ideal overall error rates

As Ganesaligam (1989), we also assessed the performances of the classification rules on the basis of their overall apparent error rate $\hat{R}$ which is simply the proportion of members in the original sample misallocated by the classification rule.

3.3 The Simulation Results

The simulation results obtained are displayed in Tables 2 and 3. In order to present tables with reasonable sizes, we do not display the results for every combination of factors: Results for the separate sampling scheme, for the sample sizes $n = 20$ and $n = 200$, and for the dimension $d = 1$ are not reported here. The behaviour under the separate sampling scheme and under the mixture sampling were similar. In the same manner, the results for the small sample sizes $n = 20$ and $n = 40$ were quite analogous as the results for sample sizes $n = 100$ and $n = 200$. At last, the dimension $d = 1$ (which is not of particular interest in cluster analysis) did not exhibit a marked difference with other dimensions.

Both tables displayed the sample mean and standard error of the overall expected error rate over the 30 trials for each combination of the parameters and for each approach in, respectively, the columns 'mre' and 'ere'. In the same manner, the sample mean and standard error of the overall apparent error rate over the 30 trials are displayed in the columns 'mra' and 'era'. Then, in each row, the number n_1, n_2 and n_3 are, respectively, the number of times over the 30 trials that the apparent error rate for the considered approach in the row was less than the apparent error rate for the approach considered, respectively, in the next row, the second next row, and the third next row... And this scheme is circular. (Example: in Table 2, for the sample size $n = 40$, the dimension

$d = 2$, the Mahalanobis distance $\Delta = 1$ and the mixing proportion $p_1 = 0.25$, CA outperformed 11 times MA, 17 times CAR and 21 times MAR out of the 30 trials.) Table 2 reported the results when each approach has been initiated with the true mixture parameter values. Table 3 reported the results obtained by the solution which provided the best value for the optimized criterion out of 20 randomly initiated runs.

Tables 4 and 5 are carrying out a qualitative evaluation of the partitions derived from the numerical experiments. Table 4 is related to the numerical experiments reported in Table 2 (starting from the true parameter values) and Table 5 is related to the numerical experiments reported in Table 3 (the best solution out of 20 runs). Each table contained 4 sub-tables which contrasted the performances of 4 couples of approach (CAR vs. MAR, CAR vs. CA, CA vs. MA and MAR vs. MA). The keys to interpret these sub-tables are the following: a star (resp. a blank) in a cell means that for this configuration of the parameters, the first (resp. second) listed approach had outperformed the second (resp. first) listed approach. An hyphen in a cell means that no approach had outperformed the other one. We considered that an approach had outperformed the other one if the number of times it produced a smaller apparent error rate out of the 30 trials is, at least, greater than the number of times it produced a greater apparent error rate *plus* one.

Note that Tables 4 and 5 provided also the evaluation of the four approaches for sample size $n = 20$ and $n = 200$ and for the dimension $d = 1$.

In its simulation study, Ganesalingam (1989) compared CAR and MA. He concluded that, whenever the mixing proportions are close to equality, CAR is generally preferable to MA and that, whenever the mixing proportions are far from equality, MA is preferable to CAR. Moreover, he recommended to use MA in preference of CAR as a clustering procedure, unless we can be sure that the observations are present in approximately the same proportion from each subpopulation. Our experiments confirmed these conclusions.

Moreover, the results in Tables 2-5 showed that, whenever the mixing proportions are close to equality, CAR is preferable to CA and MAR is preferable to MA. Otherwise, CA should be preferred to CAR and MA should be preferred to MAR, and the superiority of the unrestricted approaches is more marked as the mixing proportions are extreme. On the other hand, Tables 2-5 did not suggest a general superiority of MA over CA or of MAR over CAR. These results show that, before the choice of the classification or the mixture approach, the choice of the underlying model is of primary importance.

Now, it can be seen that, when the algorithms are initiated with the true parameter values, CA outperformed MA. But, this superiority can be thought of as somewhat artificial as, in general, we have no ideas on the mixture parameter values and since it is apparent from Tables 3 and 5 that CA is more dependent on the initial position than MA. Moreover, CA exhibited no better results than MA when the algorithms are randomly initiated. The same remarks on the comparison of CAR and MAR could be expressed.

It is difficult to recommend one of the two approaches in a general setting. From an attentive examination of Table 3, it seems that CA could be preferred for small sample sizes and that MA could be preferred for moderate or large sample sizes.

				n=40							n=100						
d	Δ	p_1		*mre*	*ere*	*mra*	*era*	$n1$	$n2$	$n3$		*mre*	*ere*	*mra*	*era*	$n1$	$n2$	$n3$
2	1	.25	CAR	0.362	0.077	0.332	0.085	19	10	9	CAR	0.354	0.031	0.357	0.058	18	7	5
			MAR	0.396	0.062	0.380	0.074	5	4	3	MAR	0.382	0.033	0.380	0.056	4	6	9
			CA	0.305	0.103	0.299	0.105	11	17	21	CA	0.298	0.088	0.290	0.080	13	22	25
			MA	0.295	0.069	0.290	0.082	16	20	16	MA	0.305	0.071	0.291	0.074	22	24	13
		.35	CAR	0.321	0.024	0.301	0.082	13	12	19	CAR	0.321	0.019	0.309	0.046	19	6	11
			MAR	0.359	0.066	0.316	0.073	9	15	7	MAR	0.358	0.048	0.338	0.066	3	7	6
			CA	0.321	0.087	0.288	0.068	14	14	19	CA	0.292	0.023	0.275	0.048	18	21	24
			MA	0.339	0.056	0.315	0.072	7	11	4	MA	0.325	0.036	0.307	0.057	18	19	5
		.50	CAR	0.320	0.017	0.307	0.067	16	6	13	CAR	0.322	0.019	0.336	0.047	18	9	18
			MAR	0.342	0.035	0.327	0.071	5	10	5	MAR	0.347	0.040	0.369	0.043	7	11	6
			CA	0.324	0.019	0.308	0.067	14	5	15	CA	0.325	0.019	0.337	0.047	17	13	22
			MA	0.363	0.049	0.325	0.079	8	13	7	MA	0.367	0.050	0.368	0.045	9	12	8
	2	.25	CAR	0.193	0.044	0.188	0.072	17	7	8	CAR	0.208	0.034	0.200	0.063	28	2	6
			MAR	0.254	0.072	0.240	0.086	4	3	5	MAR	0.271	0.052	0.276	0.071	1	1	1
			CA	0.152	0.033	0.149	0.062	14	16	22	CA	0.160	0.065	0.142	0.058	20	24	29
			MA	0.186	0.063	0.178	0.081	17	20	5	MA	0.174	0.047	0.175	0.061	21	28	5
		.35	CAR	0.199	0.040	0.182	0.052	14	7	13	CAR	0.179	0.026	0.158	0.041	25	8	17
			MAR	0.245	0.073	0.222	0.083	7	10	5	MAR	0.225	0.059	0.209	0.064	5	14	2
			CA	0.181	0.030	0.156	0.052	19	15	19	CA	0.159	0.008	0.142	0.032	20	18	23
			MA	0.234	0.088	0.208	0.100	9	11	3	MA	0.201	0.057	0.193	0.073	11	16	3
		.50	CAR	0.170	0.013	0.164	0.065	15	5	16	CAR	0.166	0.007	0.168	0.036	20	12	23
			MAR	0.216	0.085	0.205	0.089	7	11	7	MAR	0.195	0.061	0.196	0.067	8	18	6
			CA	0.175	0.017	0.163	0.070	17	5	16	CA	0.176	0.024	0.173	0.034	22	6	16
			MA	0.244	0.095	0.218	0.110	6	9	4	MA	0.230	0.073	0.223	0.071	2	6	3
	3	.25	CAR	0.091	0.031	0.094	0.054	18	4	6	CAR	0.100	0.027	0.100	0.043	26	3	5
			MAR	0.191	0.103	0.209	0.117	1	1	0	MAR	0.186	0.084	0.194	0.091	1	2	2
			CA	0.073	0.037	0.074	0.052	9	16	23	CA	0.064	0.011	0.062	0.021	10	22	29
			MA	0.080	0.024	0.081	0.047	13	21	2	MA	0.072	0.030	0.077	0.041	22	28	4
		.35	CAR	0.077	0.015	0.076	0.048	13	3	9	CAR	0.077	0.014	0.072	0.033	18	8	12
			MAR	0.113	0.065	0.106	0.090	2	5	3	MAR	0.106	0.056	0.103	0.061	4	7	1
			CA	0.080	0.024	0.064	0.048	13	10	17	CA	0.073	0.015	0.068	0.025	11	14	21
			MA	0.093	0.037	0.088	0.064	6	11	0	MA	0.081	0.028	0.080	0.042	11	18	5
		.50	CAR	0.078	0.015	0.070	0.054	10	5	12	CAR	0.073	0.005	0.074	0.025	9	1	8
			MAR	0.092	0.036	0.082	0.056	6	7	4	MAR	0.075	0.006	0.078	0.027	3	4	3
			CA	0.080	0.016	0.076	0.054	10	0	9	CA	0.073	0.007	0.073	0.025	8	3	11
			MA	0.096	0.035	0.087	0.058	2	4	3	MA	0.076	0.008	0.075	0.027	7	10	8
4	1	.25	CAR	0.372	0.075	0.346	0.080	15	15	9	CAR	0.355	0.027	0.331	0.052	25	6	6
			MAR	0.397	0.048	0.373	0.075	8	9	9	MAR	0.405	0.043	0.381	0.055	4	3	4
			CA	0.361	0.108	0.345	0.101	9	14	17	CA	0.270	0.071	0.262	0.081	19	24	26
			MA	0.325	0.059	0.307	0.092	18	19	17	MA	0.309	0.062	0.292	0.075	22	26	9
		.35	CAR	0.328	0.022	0.332	0.089	18	11	19	CAR	0.330	0.019	0.323	0.059	20	5	17
			MAR	0.376	0.038	0.375	0.072	8	11	6	MAR	0.381	0.062	0.367	0.062	1	7	6
			CA	0.313	0.040	0.318	0.067	20	16	20	CA	0.300	0.027	0.289	0.054	25	22	28
			MA	0.355	0.044	0.359	0.078	10	12	7	MA	0.341	0.055	0.332	0.071	9	22	2
		.50	CAR	0.332	0.018	0.333	0.071	12	5	19	CAR	0.327	0.019	0.325	0.047	20	11	23
			MAR	0.368	0.060	0.352	0.070	8	16	7	MAR	0.361	0.041	0.353	0.059	7	19	9
			CA	0.335	0.019	0.332	0.071	19	5	11	CA	0.333	0.023	0.330	0.054	22	10	20
			MA	0.360	0.031	0.367	0.075	3	6	4	MA	0.382	0.045	0.376	0.061	3	7	5
	2	.25	CAR	0.203	0.037	0.167	0.075	22	5	13	CAR	0.210	0.031	0.199	0.050	28	2	11
			MAR	0.310	0.089	0.301	0.123	1	3	2	MAR	0.293	0.064	0.290	0.077	1	4	0
			CA	0.170	0.054	0.138	0.071	17	16	25	CA	0.147	0.017	0.135	0.032	23	26	29
			MA	0.236	0.093	0.207	0.105	7	20	1	MA	0.194	0.067	0.185	0.070	18	25	2
		.35	CAR	0.197	0.036	0.197	0.060	16	6	13	CAR	0.182	0.017	0.174	0.039	28	7	21
			MAR	0.262	0.081	0.247	0.089	2	6	6	MAR	0.279	0.079	0.260	0.077	1	4	0
			CA	0.182	0.029	0.165	0.054	17	19	22	CA	0.168	0.016	0.156	0.031	24	19	29
			MA	0.237	0.055	0.216	0.077	11	15	3	MA	0.219	0.048	0.206	0.055	6	23	3
		.50	CAR	0.181	0.021	0.152	0.059	24	5	28	CAR	0.175	0.018	0.167	0.038	19	7	24
			MAR	0.234	0.052	0.213	0.090	1	8	2	MAR	0.208	0.043	0.193	0.050	5	17	5
			CA	0.184	0.020	0.150	0.063	26	7	23	CA	0.180	0.023	0.171	0.040	21	4	18
			MA	0.252	0.056	0.219	0.084	1	8	0	MA	0.235	0.075	0.215	0.070	2	7	4
	3	.25	CAR	0.097	0.041	0.094	0.063	21	4	14	CAR	0.097	0.028	0.085	0.043	25	1	7
			MAR	0.196	0.117	0.188	0.125	2	2	3	MAR	0.204	0.094	0.190	0.104	0	1	0
			CA	0.078	0.032	0.073	0.047	18	13	24	CA	0.064	0.009	0.056	0.023	21	24	28
			MA	0.134	0.105	0.118	0.105	8	19	4	MA	0.080	0.027	0.072	0.041	16	25	4
		.35	CAR	0.085	0.014	0.078	0.048	15	8	12	CAR	0.081	0.015	0.066	0.025	22	7	18
			MAR	0.129	0.078	0.114	0.086	4	4	4	MAR	0.114	0.061	0.108	0.071	2	7	1
			CA	0.080	0.016	0.073	0.042	15	10	17	CA	0.073	0.012	0.059	0.025	18	18	27
			MA	0.116	0.067	0.100	0.071	6	10	1	MA	0.085	0.024	0.079	0.039	10	18	3
		.50	CAR	0.085	0.015	0.070	0.043	9	4	13	CAR	0.076	0.008	0.073	0.021	16	4	15
			MAR	0.110	0.043	0.088	0.062	5	10	6	MAR	0.089	0.023	0.089	0.036	2	10	2
			CA	0.087	0.017	0.074	0.048	11	1	7	CA	0.078	0.011	0.072	0.025	17	5	14
			MA	0.126	0.065	0.102	0.083	4	2	4	MA	0.090	0.020	0.086	0.031	4	8	3

Table 2. Means and standard errors of error rates for solutions obtained when initiating with the true parameter values

				n=40							n=100						
d	Δ	p_1		mre	ere	mra	era	n1	n2	n3		mre	ere	mra	era	n1	n2	n3
2	1	.25	CAR	0.415	0.087	0.398	0.079	4	7	5	CAR	0.405	0.053	0.390	0.054	15	4	11
			MAR	0.405	0.071	0.378	0.091	10	11	10	MAR	0.412	0.051	0.402	0.050	4	9	6
			CA	0.340	0.086	0.339	0.092	9	16	14	CA	0.301	0.061	0.319	0.066	18	23	23
			MA	0.358	0.094	0.348	0.101	14	10	7	MA	0.357	0.087	0.361	0.085	14	18	4
		.35	CAR	0.399	0.079	0.375	0.087	10	10	11	CAR	0.382	0.054	0.369	0.061	12	12	14
			MAR	0.379	0.052	0.371	0.074	14	13	10	MAR	0.380	0.057	0.369	0.061	13	17	14
			CA	0.390	0.066	0.378	0.081	5	11	13	CA	0.358	0.038	0.360	0.051	14	15	15
			MA	0.395	0.073	0.371	0.085	12	14	8	MA	0.386	0.056	0.378	0.058	11	7	10
		.50	CAR	0.397	0.075	0.369	0.080	9	17	16	CAR	0.375	0.050	0.360	0.083	10	18	12
			MAR	0.400	0.077	0.383	0.083	15	15	5	MAR	0.368	0.060	0.350	0.077	17	14	12
			CA	0.425	0.072	0.404	0.071	5	1	5	CA	0.458	0.048	0.414	0.056	9	7	12
			MA	0.423	0.071	0.404	0.082	3	5	8	MA	0.422	0.072	0.386	0.084	10	12	14
	2	.25	CAR	0.308	0.108	0.286	0.106	11	12	10	CAR	0.329	0.089	0.319	0.090	17	7	5
			MAR	0.313	0.103	0.300	0.096	9	7	4	MAR	0.337	0.076	0.341	0.074	4	6	10
			CA	0.284	0.127	0.277	0.139	6	10	17	CA	0.242	0.104	0.229	0.096	9	20	26
			MA	0.286	0.127	0.274	0.134	11	16	10	MA	0.237	0.112	0.223	0.104	21	23	17
		.35	CAR	0.296	0.109	0.280	0.121	5	8	12	CAR	0.279	0.111	0.268	0.119	15	17	17
			MAR	0.291	0.101	0.273	0.119	10	14	6	MAR	0.296	0.115	0.282	0.114	14	10	8
			CA	0.295	0.115	0.280	0.112	7	9	8	CA	0.282	0.091	0.268	0.103	9	11	15
			MA	0.312	0.117	0.294	0.121	6	4	4	MA	0.279	0.110	0.259	0.108	13	17	13
		.50	CAR	0.239	0.097	0.215	0.109	9	13	17	CAR	0.215	0.070	0.212	0.076	7	20	18
			MAR	0.255	0.108	0.233	0.116	12	15	1	MAR	0.205	0.069	0.198	0.072	26	20	13
			CA	0.332	0.128	0.296	0.133	9	5	7	CA	0.347	0.143	0.328	0.127	6	2	3
			MA	0.340	0.128	0.305	0.134	7	7	9	MA	0.280	0.118	0.269	0.126	9	6	20
	3	.25	CAR	0.261	0.168	0.260	0.163	7	3	4	CAR	0.140	0.110	0.139	0.112	16	4	4
			MAR	0.286	0.167	0.290	0.153	3	3	5	MAR	0.199	0.124	0.202	0.126	1	1	7
			CA	0.153	0.148	0.150	0.154	2	16	20	CA	0.073	0.024	0.072	0.039	6	20	27
			MA	0.149	0.149	0.145	0.156	15	20	5	MA	0.069	0.017	0.072	0.033	21	28	4
		.35	CAR	0.144	0.129	0.147	0.132	7	10	9	CAR	0.082	0.019	0.078	0.039	9	9	9
			MAR	0.146	0.122	0.152	0.123	12	11	6	MAR	0.089	0.030	0.084	0.042	13	11	8
			CA	0.172	0.157	0.172	0.157	6	7	9	CA	0.082	0.032	0.082	0.048	6	8	12
			MA	0.144	0.133	0.147	0.142	8	11	6	MA	0.078	0.030	0.080	0.045	11	14	8
		.50	CAR	0.122	0.130	0.108	0.114	5	6	7	CAR	0.076	0.010	0.076	0.035	2	5	10
			MAR	0.130	0.143	0.117	0.126	8	7	4	MAR	0.072	0.005	0.073	0.032	11	11	10
			CA	0.139	0.133	0.120	0.120	4	2	4	CA	0.085	0.028	0.080	0.043	10	6	5
			MA	0.139	0.147	0.122	0.128	1	2	3	MA	0.082	0.028	0.083	0.043	9	6	5
4	1	.25	CAR	0.463	0.080	0.415	0.075	8	6	7	CAR	0.444	0.060	0.425	0.053	12	5	8
			MAR	0.459	0.073	0.407	0.072	6	6	14	MAR	0.447	0.062	0.424	0.067	5	8	11
			CA	0.402	0.069	0.364	0.091	12	18	15	CA	0.345	0.087	0.357	0.080	14	22	21
			MA	0.430	0.069	0.383	0.086	17	13	9	MA	0.382	0.080	0.376	0.076	17	15	10
		.35	CAR	0.436	0.064	0.385	0.061	11	11	11	CAR	0.425	0.041	0.421	0.055	10	4	7
			MAR	0.442	0.060	0.398	0.058	7	7	5	MAR	0.425	0.057	0.403	0.066	7	7	13
			CA	0.436	0.074	0.406	0.052	10	7	12	CA	0.380	0.051	0.378	0.058	17	22	19
			MA	0.438	0.064	0.400	0.055	5	11	10	MA	0.400	0.049	0.384	0.057	20	16	7
		.50	CAR	0.418	0.056	0.388	0.086	7	11	13	CAR	0.408	0.057	0.386	0.059	13	19	17
			MAR	0.437	0.074	0.391	0.087	11	10	8	MAR	0.415	0.059	0.388	0.055	17	17	11
			CA	0.438	0.067	0.393	0.079	14	10	11	CA	0.435	0.050	0.412	0.060	11	5	7
			MA	0.454	0.075	0.410	0.073	7	7	5	MA	0.425	0.057	0.406	0.057	6	10	14
	2	.25	CAR	0.404	0.101	0.379	0.100	11	8	10	CAR	0.373	0.112	0.351	0.110	11	8	10
			MAR	0.415	0.092	0.392	0.081	6	5	6	MAR	0.396	0.100	0.376	0.100	5	7	7
			CA	0.351	0.113	0.331	0.122	14	17	17	CA	0.285	0.104	0.279	0.105	16	20	22
			MA	0.405	0.108	0.364	0.108	12	13	5	MA	0.320	0.124	0.311	0.124	14	15	6
		.35	CAR	0.365	0.115	0.324	0.118	8	7	12	CAR	0.341	0.080	0.339	0.091	12	17	19
			MAR	0.382	0.104	0.337	0.115	7	10	6	MAR	0.350	0.085	0.345	0.088	14	17	14
			CA	0.325	0.107	0.289	0.119	12	14	15	CA	0.345	0.097	0.344	0.111	14	11	14
			MA	0.363	0.123	0.319	0.128	9	9	5	MA	0.372	0.100	0.373	0.114	9	6	8
		.50	CAR	0.322	0.115	0.283	0.115	11	14	13	CAR	0.318	0.106	0.303	0.105	16	24	20
			MAR	0.360	0.116	0.320	0.104	11	8	4	MAR	0.331	0.108	0.320	0.112	24	21	10
			CA	0.382	0.113	0.342	0.118	9	9	12	CA	0.413	0.106	0.389	0.107	8	4	5
			MA	0.384	0.117	0.339	0.112	6	10	6	MA	0.370	0.115	0.353	0.116	5	4	18
	3	.25	CAR	0.344	0.120	0.343	0.127	9	6	5	CAR	0.275	0.158	0.258	0.148	19	1	2
			MAR	0.374	0.135	0.337	0.106	6	4	14	MAR	0.343	0.157	0.323	0.152	2	0	6
			CA	0.190	0.124	0.176	0.141	8	22	22	CA	0.124	0.108	0.119	0.104	8	24	27
			MA	0.231	0.157	0.196	0.142	20	20	7	MA	0.124	0.116	0.122	0.113	25	26	9
		.35	CAR	0.188	0.131	0.177	0.144	6	10	10	CAR	0.135	0.101	0.133	0.100	13	9	7
			MAR	0.211	0.129	0.200	0.147	9	8	5	MAR	0.190	0.131	0.183	0.131	7	6	6
			CA	0.221	0.155	0.198	0.157	9	6	9	CA	0.126	0.092	0.120	0.100	9	17	22
			MA	0.243	0.152	0.223	0.165	5	5	5	MA	0.118	0.095	0.114	0.097	18	22	10
		.50	CAR	0.209	0.140	0.178	0.133	12	9	10	CAR	0.110	0.072	0.106	0.074	11	19	13
			MAR	0.256	0.158	0.222	0.146	6	5	4	MAR	0.120	0.079	0.118	0.081	20	16	9
			CA	0.210	0.128	0.179	0.126	10	10	12	CA	0.191	0.147	0.183	0.144	6	3	3
			MA	0.239	0.151	0.201	0.142	8	7	6	MA	0.141	0.115	0.138	0.109	5	8	15

Table 3. Means and standard errors of error rates for the best solution obtained out of 20 randomly initiated runs

CAR-MAR

n	d	Δ=1 pl=.25	Δ=1 pl=.35	Δ=1 pl=.50	Δ=2 pl=.25	Δ=2 pl=.35	Δ=2 pl=.50	Δ=3 pl=.25	Δ=3 pl=.35	Δ=3 pl=.50
20	1	*	*	–	*	*	–	*	*	–
	2	*	*	*	*	*	*	*	*	*
	4	*		*	*	*	*	*	*	*
40	1	*	–	*	*	*	*	*	*	–
	2	*	*	*	*	*	*	*	*	*
	4	*	*	*	*	*	*	*	*	*
100	1	*	*	–	*	*	*	*	*	–
	2	*	*	*	*	*	*	*	*	*
	4	*	*	*	*	*	*	*	*	*
200	1	*	*	*	*	*	–	*	*	–
	2	*	*	*	*	*	*	*	*	*
	4	*	*	*	*	*	*	*	*	*

CAR-CA

n	d	Δ=1 pl=.25	Δ=1 pl=.35	Δ=1 pl=.50	Δ=2 pl=.25	Δ=2 pl=.35	Δ=2 pl=.50	Δ=3 pl=.25	Δ=3 pl=.35	Δ=3 pl=.50
20	1			–			–			
	2	*		*	*					–
	4	*	–		*		*	–	*	–
40	1			*					–	
	2			–			–			*
	4	–		–						*
100	1			*			–			–
	2						*			
	4			–			*			–
200	1			–			*			–
	2						*			*
	4			*						–

CA-MA

n	d	Δ=1 pl=.25	Δ=1 pl=.35	Δ=1 pl=.50	Δ=2 pl=.25	Δ=2 pl=.35	Δ=2 pl=.50	Δ=3 pl=.25	Δ=3 pl=.35	Δ=3 pl=.50
20	1		*	–	*	*	*	–	*	*
	2		–	*	*	*	*	*	*	*
	4					*	*	–	–	*
40	1	*	*	*	*	*	*	*	*	*
	2		*	*	*	*	*	*	*	*
	4		*	*	*	*	*	*	*	*
100	1		*	*		*		*	*	*
	2	–	*	*	*	*	*	*	*	–
	4	*	*	*	*	*	*	*	*	*
200	1		–	*	–	*	*	*	–	–
	2		–	*	*	*	*	–	*	*
	4	*	*	*	*	*	*	*	*	*

MAR-MA

n	d	Δ=1 pl=.25	Δ=1 pl=.35	Δ=1 pl=.50	Δ=2 pl=.25	Δ=2 pl=.35	Δ=2 pl=.50	Δ=3 pl=.25	Δ=3 pl=.35	Δ=3 pl=.50
20	1		–	*			*			*
	2			*		*	*			
	4									
40	1		–	*			*			*
	2		*			–	*			*
	4		–	*			–			*
100	1			*			*			*
	2			–			*			
	4			*			*			*
200	1			*			*			–
	2			*			*			*
	4			*			*			*

Table 4. Qualitative comparisons of the four methods when initiating with the true parameter values

CAR-MAR

n	d	Δ=1 pl=.25	Δ=1 pl=.35	Δ=1 pl=.50	Δ=2 pl=.25	Δ=2 pl=.35	Δ=2 pl=.50	Δ=3 pl=.25	Δ=3 pl=.35	Δ=3 pl=.50
20	1	–	–			*		*	–	*
	2	*	–	–	*	–	–	*	–	–
	4	–	*	*	*	*	–	*	*	*
40	1	*	*	–	–	–		*	*	*
	2		–	*	*	–	*	*	–	–
	4		*	–	*	*	*		–	*
100	1		*	–	*	*	–	*	*	–
	2	*			*	*		*	–	
	4	–		*	*		*	*	*	*
200	1	*		*	*	*		*	*	*
	2	*		*	*	*		*	*	–
	4			*	*	*	–	*	*	

CAR-CA

n	d	Δ=1 pl=.25	Δ=1 pl=.35	Δ=1 pl=.50	Δ=2 pl=.25	Δ=2 pl=.35	Δ=2 pl=.50	Δ=3 pl=.25	Δ=3 pl=.35	Δ=3 pl=.50
20	1		*	–			*	–	–	*
	2	*	–	–	*	*	*		*	
	4	–	–	–		*	*		–	
40	1		–	*		*	*		–	*
	2		–	*	*	–	*		*	*
	4		*	–			*		*	–
100	1		*	*		*	*		–	*
	2			*		*	*		–	–
	4			*		*	*			*
200	1		*	*			*			*
	2		*	*		*	*			*
	4			*		*	*			–

CA-MA

p	d	Δ=1 pl=.25	Δ=1 pl=.35	Δ=1 pl=.50	Δ=2 pl=.25	Δ=2 pl=.35	Δ=2 pl=.50	Δ=3 pl=.25	Δ=3 pl=.35	Δ=3 pl=.50
20	1		–	*		–	*	*		
	2	–	*	*	–	*	–	*		*
	4		–	*	*	*	*	*	*	*
40	1	*	*							–
	2	*				*	–		–	–
	4	*	–	*	*	*	*	–	*	*
100	1	*						–		
	2	*	*					*		*
	4	*	*		*	*		–	–	
200	1			*	–				*	*
	2	*	–							
	4	*	*		–			–		*

MAR-MA

n	d	Δ=1 pl=.25	Δ=1 pl=.35	Δ=1 pl=.50	Δ=2 pl=.25	Δ=2 pl=.35	Δ=2 pl=.50	Δ=3 pl=.25	Δ=3 pl=.35	Δ=3 pl=.50
20	1			*			*		–	–
	2	–		–		*	*		–	–
	4			*		–	*	–	*	*
40	1			*			*			*
	2	–	–	*		*	*		–	–
	4			*		–	–		*	
100	1		*	*			*			–
	2		*	*			*			*
	4			*		*	*			*
200	1		–	*			*			*
	2		*	*			*			*
	4			–			*			–

Table 5. Qualitative comparisons of the four methods when initiating with 20 randomly initiated runs

In order to precise the influence of the sample sizes, we have performed additional experiments. We selected the following parameter configurations: $d = 2$, $\Delta = 1$, $p_1 = 0.25$ and $n = 40$, 100, 200, 500, 800 and 1000. For each configuration, we generated 30 samples, then CA and MA have been experimented from 20 random initial positions and the solution which provided the best value of the optimized criterion out of the 20 runs has been selected. Proceeding to these additional experiments, we also aimed to evaluate how useful the asymptotic figures of Bryant (1991) may be in practical situations. The main figures are, firstly, that MA is expected to provide consistent parameter estimates as the sample size grows to infinity (but not CA) and, secondly, that the optimal solution for CA turns out to occur on the boundary of the mixture parameter space, if the mixture components are poorly separated.

In Table 6, the sample mean of the overall expected error rate over the 30 trials and, into parentheses, its sample standard error are displayed for the different sample sizes. In Table 7 are reported the number of times the CEM algorithm (for CA) and the EM algorithm (for MA) hit the boundary of the parameter space and had to be started afresh.

n	40	100	200	500	800	1000
CA	0.280 (0.076)	0.282 (0.086)	0.304 (0.080)	0.324 (0.085)	0.323 (0.090)	0.317 (0.089)
MA	0.305 (0.089)	0.264 (0.065)	0.271 (0.061)	0.261 (0.047)	0.244 (0.031)	0.242 (0.026)

Table 6. Variations of the mean and standard error of the overall expected error rate as a function of the sample size n

n	40	100	200	500	800	1000
CA	66	136	328	951	1165	1361
MA	476	0	0	7	0	0

Table 7. Number of times the algorithms had to be started afresh to produce a classification as a function of the sample size n

Both tables highlight the asymptotic analyze of Bryant (1991). It seems that the overall error rate of MA tends to the ideal overall expected error rate (see Table 1) as n grows with a decreasing standard error. On the contrary, the expected error rate of CA seems go away from the ideal value with a constant standard error as n grows.

On the other hand, the CA asymptotic tendency to not classify at all for ill-separated components or extreme values of the mixing proportions (Bryant 1991) appeared strongly even for moderate sample size from Table 7. This tendency is certainly a drawback of CA.

4. Discussion

In approaching cluster analysis via a mixture of Gaussian distributions, we have isolated two important factors to tackle this problem. These factors are the type of approach (mixture or classification) and the underlying assumption on the mixing proportions (equal proportions or unknown proportions). From our numerical experiments, the classification approach can be preferred to deal with small samples and

the mixture approach can be preferred to deal with moderate or large samples. The choice of the underlying mixture model is more difficult since, in general, there is no prior information available on the mixing proportions. And, moreover, the assumptions concerning the mixing proportions can lead to quite different partitions. In particular, unrestricted approaches assuming unknown mixing proportions are not guaranteed to provide the more relevant results when the true mixing proportions are close to equality. In effect, approaches assuming equal mixing proportions are more parsimonious and less initial-position dependent. For this very reason, these approaches can be thought of as more reliable and more robust than unrestricted approaches in practical situations. The aforementioned unrestricted approaches CA and MA are jeopardized by the occurrence of sub-optimal solutions or singularities (especially CA). To attenuate the initial-position dependence of these approaches, the following strategy can be proposed:

- Start with some runs of CAR and run CA or MA from the best position obtained with CAR.

This hybrid approach has been tested on 30 samples from the configurations n = 100, Δ = 1, 2, 3, d = 4, p_1 = 0.25, 0.35, 0.50 for a two-component Gaussian mixture. For each sample, we started the procedure with 20 runs of CAR. The results, presented in the same way than in Tables 2 and 3, are displayed in Table 8. In this experiment, we also ran MAR from the best solution provided by CAR.

			n=100							
d	Δ	p_1		*mre*	*ere*	*mra*	*era*	*n1*	*n2*	*n3*
4	1	.25	CAR	0.447	0.066	0.434	0.053	13	5	8
			MAR	0.453	0.062	0.435	0.044	7	9	12
			CA	0.446	0.072	0.432	0.054	11	12	12
			MA	0.440	0.074	0.425	0.056	13	13	13
		.35	CAR	0.402	0.062	0.386	0.067	12	3	12
			MAR	0.414	0.064	0.396	0.062	5	7	7
			CA	0.400	0.064	0.382	0.068	14	10	16
			MA	0.404	0.070	0.389	0.072	9	13	6
		.50	CAR	0.429	0.066	0.410	0.053	15	5	10
			MAR	0.432	0.065	0.419	0.059	8	10	8
			CA	0.425	0.065	0.406	0.055	13	12	14
			MA	0.435	0.062	0.414	0.052	10	14	9
	2	.25	CAR	0.380	0.105	0.360	0.113	14	6	6
			MAR	0.385	0.098	0.366	0.098	8	4	10
			CA	0.377	0.121	0.353	0.118	8	9	13
			MA	0.353	0.121	0.342	0.120	16	18	12
		.35	CAR	0.319	0.092	0.309	0.095	14	6	12
			MAR	0.323	0.093	0.317	0.091	5	8	8
			CA	0.301	0.093	0.295	0.093	14	15	16
			MA	0.315	0.098	0.304	0.100	12	16	7
		.50	CAR	0.337	0.101	0.324	0.109	7	3	11
			MAR	0.335	0.103	0.318	0.108	10	14	16
			CA	0.325	0.102	0.310	0.109	16	18	12
			MA	0.351	0.112	0.324	0.108	9	8	6
	3	.25	CAR	0.284	0.155	0.272	0.161	6	3	8
			MAR	0.288	0.150	0.268	0.156	6	11	8
			CA	0.250	0.162	0.246	0.163	10	19	18
			MA	0.247	0.169	0.247	0.175	17	16	10
		.35	CAR	0.134	0.082	0.130	0.089	10	3	9
			MAR	0.133	0.083	0.127	0.090	7	9	8
			CA	0.110	0.082	0.105	0.088	15	19	21
			MA	0.123	0.091	0.117	0.093	16	18	4
		.50	CAR	0.121	0.078	0.109	0.075	1	1	7
			MAR	0.104	0.059	0.095	0.061	4	7	10
			CA	0.095	0.050	0.083	0.055	16	18	16
			MA	0.104	0.060	0.096	0.063	13	6	6

Table 8. Means and standard errors of error rates for solutions obtained by an hybrid approach

It is noteworthy that, using the hybrid approach, CA outperformed CAR even for p_1 = 0.50. However, it appears from the comparison of Table 3 and Table 8, that the hybrid approach did not improve the results of CA and MA randomly initiated, except

for p_1 =0.50. Finally, this hybrid approach seems to be useful but can not be regarded as a general answer to the initial-position dependence of CA and MA.

As a conclusion, we agree with Symons (1981) when he said "There seems to be no simple recommendation to guide the users of these criteria... the general impression from the above sketch of some empirical results is that the choice of the most appropriate approach depend upon a knowledge of data... The only reasonable *a priori* suggestion is to compare results obtained from several approaches".

References

P. Bryant (1991), Large-Sample Results for Optimization Based Clustering Methods, *Journal of Classification*, **8**, 31-44.

P. Bryant and J.A. Williamson (1978), Asymptotic Behaviour of Classification Maximum Likelihood Estimates, *Biometrika*, **65**, 273-281.

P. Bryant and J.A. Williamson (1986), Maximum Likelihood and Classification: A Comparison of three Approaches, in: W. Gaul and M. Schader (Eds.), *Classification as a Tool of Research* (North-Holland, Amsterdam) 33-45.

G. Celeux and J. Diebolt (1985), The SEM Algorithm: A Probabilistic Teacher Algorithm derived from the EM Algorithm for the Mixture Problem, *Computational Statistics Quarterly*, **2**, 73-82.

G. Celeux and G. Govaert (1991), A Classification EM Algorithm for Clustering and Two Stochastic Versions, *Computational Statistics and Data Analysis* (to appear).

A.P. Dempster, N.M. Laird and D.B. Rubin (1977), Maximum Likelihood from Incomplete Data via the EM Algorithm (with discussion), *Journal of the Royal Statistical Society*, **B 39**, 1-38.

H.P. Friedman and J. Rubin (1967), On Some Invariant Criterion for Grouping Data, *Journal of the American Statistical Association*, **62**, 1159-1178.

S. Ganesalingam (1989), Classification and Mixture Approach to Clustering via Maximum Likelihood, *Applied Statistics*, **38**, 455-466.

F.H.C. Marriott (1975), Separating Mixtures of Normal Distributions, *Biometrics*, **31**, 767-769.

F.H.C. Marriott (1982), Optimization Methods of Cluster Analysis, *Biometrika*, **69**, 417-421.

G.J. McLachlan (1982), The Classification and Maximum Likelihood Approach to Cluster Analysis, in: P.R. Krishnaiah and L.N. Kanal (Eds), *Handbook of Statistics* (North-Holland, Amsterdam), vol. 2, 199-208.

G.J. McLachlan and K.E. Basford (1988), *Mixture Models* (Marcel Dekker, New York).

A.J. Scott and M.J. Symons (1971), Clustering Methods based on Likelihood Ratio Criteria, *Biometrics*, **27**, 387-397.

M.J. Symons (1981), Clustering Criteria and Multivariate Normal Mixture, *Biometrics*, **37**, 35-43.

ISSN 0249 - 6399