On a Resampling Approach to Choosing the Number of Components in Normal Mixture Models

G.J. McLachlan

Department of Mathematics
The University of Queensland
Brisbane, Queensland 4072

D. Peel

Departments of Mathematics
and Mining & Metallurgical
Engineering
The University of Queensland
Brisbane, Queensland 4072

Abstract

We consider the fitting of a g-component normal mixture to multivariate data. The problem is to test whether g is equal to some specified value versus some specified alternative value. This problem would arise, for example, in the context of a cluster analysis effected by a normal mixture model, where the decision on the number of clusters is undertaken by testing for the smallest value of g compatible with the data. A test statistic can be formed in terms of the likelihood ratio. Unfortunately, regularity conditions do not hold for the likelihood ratio statistic to have its usual asymptotic null distribution of chi-squared. One approach to the assessment of P-values with the use of this statistic is to adopt a resampling approach. An investigation is undertaken of the accuracy of P-values assessed in this manner.

1 Introduction

Often an important consideration in cluster analysis is deciding on the number of clusters in the data. With a mixture model-based method of clustering, the problem can be approached by testing for the smallest number of components g in the mixture model compatible with the data. This can be undertaken using the likelihood ratio test statistic λ to test for g versus $g+1$ components, proceeding sequentially from $g = 1$. Unfortunately, regularity conditions do not hold for $-2\log\lambda$ to have its usual asymptotic null distribution of chi-squared; see McLachlan and Basford (1988, Chapter 10) and the references therein.

McLachlan(1987) proposed a resampling approach to the assessment of the P-value of $-2\log\lambda$ in testing

$$H_0 : g = g_0 \quad \text{v} \quad H_1 : g = g_0 + 1 \tag{1}$$

for a specified value of g_0, whereby bootstrap samples are generated from the normal mixture model fitted under the null hypothesis of g_0 components; see also Aitkin, Anderson, and Hinde (1981). In general, the use of estimates in place of the unknown parameters will affect the accuracy of the P-values assessed on the basis of the bootstrap replications of $-2\log\lambda$. In this paper, we report some simulation results to demonstrate this effect.

2 Finite Mixture Model

Let $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ denote a p-dimensional random sample of size n. The observed values of this sample can be clustered by fitting a normal mixture model with the number of components g equal to the number of clusters to be inferred from the data. This mixture model-based approach to clustering assumes that the probability density function (p.d.f) of each $\boldsymbol{Y}_j$ is given by

$$f(\boldsymbol{y}_j;\boldsymbol{\Psi}) = \sum_{i=1}^{g} \pi_i f_i(\boldsymbol{y}_j;\, \boldsymbol{\theta}) \quad (j = 1, \ldots, n), \tag{2}$$

where $f_i(\boldsymbol{y}_j;\, \boldsymbol{\theta})$ is the p.d.f. corresponding to the ith component and $\boldsymbol{\theta}$ is the vector containing those unknown parameters in the component p.d.f. 's that are known *a priori* to be distinct. Here $\pi_1, \ldots, \pi_g$ denote the mixing proportions, which are nonnegative and sum to one. We let

$$\boldsymbol{\Psi} = (\pi_1, \ldots, \pi_{g-1}, \boldsymbol{\theta}')'$$

be the vector containing all the unknown parameters, where the prime denotes vector transpose. The log likelihood function for $\boldsymbol{\Psi}$ is given by

$$\log L(\boldsymbol{\Psi}) = \sum_{j=1}^{n} \log \sum_{i=1}^{g} \pi_i f_i(\boldsymbol{y}_j;\, \boldsymbol{\theta}). \tag{3}$$

The mixture model (2) can be fitted by maximum likelihood whereby the estimate of $\boldsymbol{\Psi}$ is given by an appropriate root of the likelihood equation

$$\partial \log L(\boldsymbol{\Psi})/\partial \boldsymbol{\Psi} = \mathbf{0}. \tag{4}$$

Solutions of (4) can be found by using the EM algorithm of Dempster, Laird, and Rubin (1977).

3 Assessing the Number of Components

When fitting the finite mixture model (2) in practice, the number of components g is often unknown and must be inferred from the data. An obvious approach to this problem is to use the likelihood ratio statistic λ to test for the smallest value of g compatible with the data. Unfortunately with mixture models, regularity conditions do not hold for $-2\log\lambda$ to have its usual asymptotic distribution of chi-squared with degrees of freedom equal to the difference between the number of parameters under the null and alternative hypotheses. A comprehensive account of this breakdown in regularity conditions has been provided by Ghosh and Sen (1985) and Hartigan (1985).

An often used approximation to the null distribution of $-2\log\lambda$ given by Wolfe (1971) is

$$-2\log\lambda \sim \chi^2_{2d},$$

where d is the difference between the number of parameters under the null and alternative hypotheses, excluding the mixing proportions. Wolfe suggested this approximation on the basis of a small scale simulation study performed for $g = 1$ versus $g = 2$ in the case of multivariate normal component densities with a common covariance matrix. McLachlan (1987) has demonstrated that Wolfe's approximation can be quite misleading for heteroscedastic normal component distributions. It leads to too small an estimate of the P-value and thus to too many clusters if the outcome of the test is rigidly interpreted from the assessed P-value. In some other related work, Thode, Finch, and Mendell (1988), Mendell, Thode, and Finch (1992), and Mendell, Finch, and Thode, (1993) have conducted some simulations including power and sample size analyses on the use of $-2\log\lambda$ in the case of a single normal versus a mixture of two univariate normal components with a common variance.

An alternative approach to approximating the null distribution of $-2\log\lambda$ in testing the hypothesis (1) is to adopt a resampling method, which can be viewed as a particular application of the general bootstrap approach of Efron (1979, 1982); see also Efron and Tibshirani (1994). More specifically, the statistic λ for the test (1) can be bootstrapped as follows. Proceeding under the null hypothesis H_0, a so-called bootstrap sample is generated from the mixture density $\hat{\boldsymbol{\Psi}}_{g_0}$, where $\hat{\boldsymbol{\Psi}}_{g_0}$ is taken to be the maximum likelihood estimate of $\boldsymbol{\Psi}_{g_0}$ formed under H_0 from the original observed sample. We are now writing $\boldsymbol{\Psi}$ as $\boldsymbol{\Psi}_g$, where g is the number of components, to explicitly denote that it is the parameter vector for a g-component normal mixture. The value of $-2\log\lambda$ is calculated for the bootstrap sample after fitting a mixture model for $g = g_0$ and $g = g_0 + 1$ in turn to it. This process is repeated independently a number of times K and the replicated values of $-2\log\lambda$ evaluated from the successive bootstrap samples can be used to assess the bootstrap, and hence the true, null distribution of $-2\log\lambda$. In particular , it enables an approximation to be made to the achieved level of significance P corresponding to the values of $-2\log\lambda$ evaluated from the original sample. The value of the jth-order statistic of the K replicates can be taken as an estimate of the quantile of order $j/(K+1)$, and the P-value can be assessed by reference with respect to the ordered bootstrap replications of $-2\log\lambda$. If a very accurate estimate of the P-value were required, then K may have to be very large. Usually, however, there is no interest in estimating a P-value with high precision. Even with a limited replication number K, the amount of computation is still considerable. McLachlan et al. (1995) have developed an algorithm for implementing this resampling approach in the case of normal mixture models.

In some special situations, for example, with $g_0 = 1$ in the case of normal mixture models , the null distribution of λ depends only on the sample size n, and so there is no difference between the bootstrap and true null distributions of $-2\log\lambda$. In these situations where it is the actual statistic $-2\log\lambda$ and not its bootstrap analogue that is being resampled, the resampling may be viewed as an application of the Monte Carlo approach to the construction of a hypothesis test having an exact level of significance. This approach was proposed originally by Barnard (1963); see Hope (1968) and Hall and Titterington (1989).

Other approaches to the problem of testing for the number of components in a mixture model have been considered recently by Bozdogan (1992), Goffinet, Loisel, and Laurent (1992), Windham and Cutler (1992), Banfield and Raftery (1993), Mangin, Goffinet, and Elsen (1993), Furman and Lindsay (1994), Celeux and Soromenho (1993), and Soromenho (1994), among others.

p	Δ	$\bar{\hat{\Delta}}$	Simulated Percentile			Average Bootstrap Estimate of Percentile (Actual Percentage)			RMSE of Bootstrap Estimate of Percentile		
			90%	95%	99%	90%	95%	99%	90%	95%	99%
1	2	3.08	5.0	6.2	9.6	5.1 (91.0%)	6.7 (96.0%)	10.7 (99.4%)	0.7430	1.2616	2.9252
	3	3.12	4.9	6.5	9.8	4.8 (89.6%)	6.3 (94.5%)	10.6 (99.6%)	0.6946	0.9424	2.5659
2	2	2.35	9.3	11.0	15.5	9.8 (91.7%)	11.8 (96.3%)	17.0 (99.4%)	1.0389	1.4596	3.3759
	3	3.18	9.1	10.8	14.9	9.0 (89.7%)	11.0 (95.2%)	15.8 (99.3%)	0.8967	1.5942	3.0977
3	2	2.61	15.7	18.0	22.0	14.0 (84.1%)	16.7 (92.6%)	23.1 (99.2%)	2.1701	2.1429	3.9362
	3	3.18	13.2	15.5	20.1	13.1 (89.8%)	15.2 (94.9%)	21.4 (99.4%)	1.0119	1.5142	3.7432
4	2	2.89	20.9	23.0	28.1	17.8 (79.1%)	20.3 (89.1%)	26.8 (98.6%)	3.3432	3.0930	4.2796
	3	3.34	17.9	20.2	26.1	16.6 (87.1%)	18.8 (92.4%)	24.6 (98.7%)	1.9057	2.2301	4.1480
5	2	2.94	25.3	27.8	35.0	22.3 (79.4%)	24.9 (89.1%)	32.2 (97.7%)	3.7073	3.7000	4.6107
	3	3.44	21.2	24.1	30.1	20.1 (86.9%)	22.6 (93.2%)	28.2 (97.9%)	1.7402	2.4174	4.1180
6	2	3.11	30.7	33.2	38.6	25.7 (74.1%)	29.0 (86.7%)	36.1 (97.7%)	5.3458	4.8659	5.1005
	3	3.53	26.5	29.3	35.4	23.5 (80.7%)	26.7 (90.4%)	34.1 (98.9%)	3.3717	3.5130	5.3061
7	2	3.27	35.3	38.3	45.2	29.5 (70.0%)	32.7 (82.3%)	40.6 (97.0%)	6.2204	6.2375	6.3222
	3	3.60	30.8	34.1	41.6	27.7 (82.0%)	30.9 (90.2%)	37.7 (97.7%)	3.6922	4.1464	5.2560
8	2	3.38	40.2	44.3	48.9	32.7 (67.2%)	36.3 (81.2%)	45.1 (95.8%)	7.8245	8.6065	6.1437
	3	3.67	36.1	39.0	47.1	30.6 (72.1%)	33.7 (82.7%)	41.4 (96.7%)	5.9259	5.9942	7.5482

Table 1: Summary of Simulation Results for the Null Distribution of $-2\log\lambda$

4 Simulation Experiment

A simulation experiment was performed to investigate the effect of assessing the P-value of the likelihood ratio test on the basis of the bootstrap rather than actual replicates of the test statistic $-2\log\lambda$. Attention was confined to mixtures of normal components with equal covariance matrices. For simplicity, the focus was on the case of $g = g_0$ versus $g = g_0 + 1$, where $g_0 = 2$. In this special case of two normal components with means $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ and common covariance matrix $\boldsymbol{\Sigma}$, it can be assumed without loss of generality that

$$\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{I}_p,$$

$$\boldsymbol{\mu}_1 = \mathbf{o}_{1,p} \textit{ and } \boldsymbol{\mu}_2 = [\Delta, \mathbf{o}_{1,p-1}]^T,$$

where $\boldsymbol{I}_p$ is the $p \times p$ identity matrix, $\mathbf{o}_{q,p}$ is the $q \times p$ zero matrix, and Δ

$$\Delta = \{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\}^{1/2}$$

is the Mahalanobis distance between the two components.

In this case, the null distribution of $-2\log\lambda$ will depend only on n, π_1, and Δ. For the purposes of our simulations, we took the mixing proportions to be equal ($\pi_1 = \pi_2 = 0.5$), and the sample size n to be 100. There were two levels of $\Delta\,(\Delta = 2$ and $3)$, corresponding to moderately and widely separated components, respectively. The dimension of the observation vector ranged from $p = 1$ to $p = 8$.

The percentiles of the null distribution of $-2\log\lambda$ were based on 999 simulation trials. Concerning the effectiveness of the bootstrapping of $-2\log\lambda$, fifty simulation trials were performed for each combination of the parameters Δ and p with $\pi_1 = 0.5$ and $n = 100$. On each trial, $K = 99$ bootstrap replications were used to assess the percentiles of the null distribution of $-2\log\lambda$.

5 Simulation Results

Corresponding to the (simulated) 90th, 95th, and 99th percentiles of the null distribution of $-2\log\lambda$, we list in Table 1 the averages of the bootstrap-based estimates of these percentiles. The percentage areas to the left of these average percentile estimates under the actual (simulated) null distribution of $-2\log\lambda$ (that is, the actual order of these percentiles) are given in parentheses alongside them. In Table 1, $\bar{\hat{\Delta}}$ is the mean of $\hat{\Delta}$ over the 50 simulation trials. We also list in Table 1, the root mean squared error (RMSE) of these bootstrap-based estimates of the percentiles. In Table 2, we give the empirical results obtained in each series of 50 simulation trials for the Type I error of the resampling approach when applied at a nominal level of $\alpha = 10\%, 5\%$, and 1%, respectively.

The results in Tables 1 and 2 provide some guide as to the effect of using the bootstrap replications in place of actual replicates of the test statistic $-2\log\lambda$ under the null hypothesis. As to be expected, the effect is more marked for $\Delta = 2$ than for $\Delta = 3$, and increases as the dimension p of the observation vector increases for a given level of Δ.

As Table 1 indicates, there does seem to be a tendency for the resampling approach using bootstrap replications to underestimate the upper percentiles of the null distribution of $-2\log\lambda$, and hence overestimate the P-value of tests based on this statistic. The associated bias is in the direction that we would anticipate. For it is well known that the plug-in estimate of Δ that effectively is being used here in the generation of the bootstrap samples tends to overestimate the true value of Δ. Hence the bootstrap data are from a normal mixture with components more widely separated than in the original mixture, and so the null hypothesis of $g_0 = 2$ would be favoured by this wider separation. This leads to the bootstrap replicates of $-2\log\lambda$ tending to be smaller than what the actual replicates would be. This tendency is confirmed in the the simulation results in Table 1.

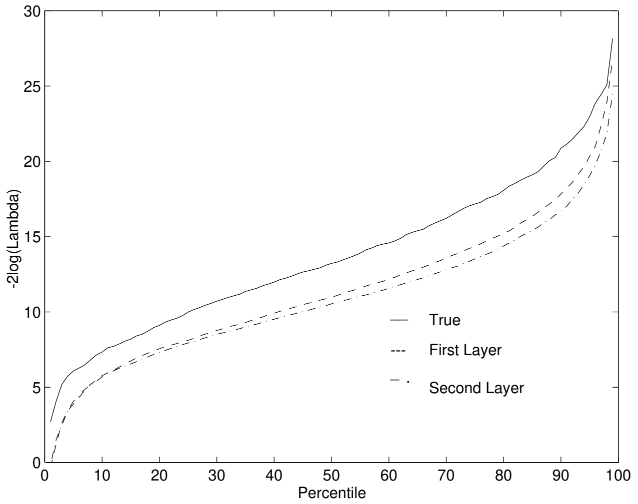

Figure 1: Plot of the Inverse of the Bootstrap-Based Estimate of the Null Distribution of $-2\log\lambda$ for $p = 4$ and $\Delta = 2$

p	Δ	Percentage of Observed Type I Errors		
		10%	5%	1%
1	2	14%	8%	4%
	3	14%	0%	0%
2	2	18%	10%	0%
	3	16%	6%	4%
3	2	14%	4%	0%
	3	8%	6%	0%
4	2	18%	6%	2%
	3	4%	4%	0%
5	2	22%	16%	0%
	3	10%	4%	0%
6	2	28%	16%	6%
	3	24%	8%	2%
7	2	20%	10%	2%
	3	24%	12%	0%
8	2	42%	18%	4%
	3	32%	14%	4%

Table 2: Percentage of the 50 Simulations Trials for which a Type I Error Occurred

6 Double Bootstrap

In fitting a mixture density $f(\boldsymbol{y}; \boldsymbol{\Psi}_g)$, some guide to the sensitivity of the assessed P-value to the use of $\hat{\boldsymbol{\Psi}}_{g_0}$ for $\boldsymbol{\Psi}_{g_0}$ in $f(\boldsymbol{y}\,; \boldsymbol{\Psi}_{g_0})$ in the generation of the bootstrap replications can be obtained by performing another layer of bootstrapping as follows. Replications of $-2\log\lambda$ can be formed from bootstrap samples drawn from $f(\boldsymbol{y}\,; \hat{\boldsymbol{\Psi}}^*_{g_0})$, where $\hat{\boldsymbol{\Psi}}^*_{g_0}$ denotes the maximum likelihood estimate calculated from a bootstrap sample drawn from $f(\boldsymbol{y}\,; \hat{\boldsymbol{\Psi}}_{g_0})$. These latter replications of $-2\log\lambda$ and the implied P-value can be contrasted with the original bootstrap replications of $-2\log\lambda$ and the implied P-value.

To illustrate the use of the double bootstrap applied to this problem, we consider two of the aforementioned combinations of the parameters Δ and p ($\Delta = 2$ with $p = 4$ and 8). We plot in Figures 1 and 2, corresponding to these two combinations, the inverse of the simulated null distribution of $-2\log\lambda$ versus the two bootstrap analogues obtained by resampling from the normal mixture model with parameter vector equal to $\hat{\boldsymbol{\Psi}}_{g_0}$ and $\hat{\boldsymbol{\Psi}}^*_{g_0}$, respectively.

The second layer of bootstrapping can be used to adjust the estimates of the percentiles based on the first layer of bootstrap replicates. These adjusted estimates of the percentiles are reported in Table 3, along with the corresponding unadjusted values form Table 1, where $\bar{\hat{\Delta}}^*$ denotes the average value of $\hat{\Delta}^*$ for each series of 50 simulation trials and $\hat{\Delta}^*$ is the maximum likelihood estimate of Δ obtained from a bootstrap sample drawn from $f(\boldsymbol{y}\,; \hat{\boldsymbol{\Psi}}_{g_0})$. It can be seen that by performing a second layer of bootstrapping a more accurate assessment of the percentiles of the null distribution of $-2\log\lambda$ can be obtained. However, the double bootstrap only partially removes the bias in the bootstrap estimates obtained from the first layer. This can be explained by an examination

p	Δ	$\bar{\hat{\Delta}}$	$\bar{\hat{\Delta}}^*$	Average Value of Unadjusted Percentile (Actual Percentage)			Average Value of Adjusted Percentile (Actual Percentage)		
				90%	95%	99%	90%	95%	99%
4	2	2.89	3.19	17.8 (79.1%)	20.3 (89.1%)	26.8 (98.6%)	19.0 (84.1%)	21.6 (92.2%)	29.0 (99.3%)
8	2	3.48	3.89	32.8 (67.2%)	18.8 (81.2%)	4.2 (95.8%)	35.8 (79.7%)	40.2 (90.1%)	49.9 (99.1%)

Table 3: Summary of Simulation Results for the Adjusted Percentile Estimates

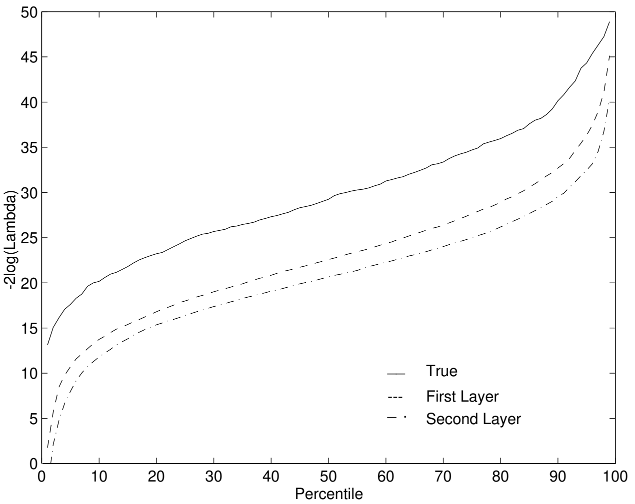

Figure 2: Plot of the Inverse of the Bootstrap-Based Estimate of the Null Distribution of $-2\log\lambda$ for $p = 8$ and $\Delta = 2$

of Figures 1 and 2, where it can be seen that the difference between the true and first layered-bootstrap assessment of the percentiles is more marked than between the first and second layered bootstrap assessments.

In this particular case of $g_0 = 2$, where the null distribution of $-2\log\lambda$ depends on the component parameters through just the Mahalanobis distance Δ, it would be straightforward to produce a suitable adjustment by working in terms of the canonical form of the problem in terms of Δ and using a less biased estimate of Δ (McLachlan (1992, Chapter 10)) in the production of the first layer of bootstrap replications.

In this paper, we have used the jth-order statistic of the K bootstrap replicates to estimate the the quantile of order $j/(K+1)$. A preferable alternative would be to use the jth-order statistic as an estimate of the quantile of order $(3j-1)/(3K+1)$; see Hoaglin (1985).

References

Aitkin, M., Anderson, D., and Hinde, J. (1981). Statistical modelling of data on teaching styles (with discussion), *J. R. Statist. Soc. A* **144**, 419–461.

Banfield, J.D. and Raftery, A. (1993). Model-based Gaussian and non-Gaussian clustering. *Biometrics* **49**, 803–821.

Barnard, G.A. (1963). Contribution to the discussion of paper by M.S. Bartlett. *J. R. Statist. Soc. B* **25**, 294.

Bozdogan, H. (1992). Choosing the number of component clusters in the mixture-model using a new informational complexity criterion of the inverse-Fisher information matrix. In *Information and Classification*, O.Opitz, B. Lausen, and R. Klar (Eds.). Heidelberg: Springer-Verlag, 40–54.

Celeux, G. and Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *J. Classification*. To appear.

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.

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

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*. Philadelphia: SIAM.

Efron, B. and Tibshirani, R.J. (1994). *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Furman, D. and Lindsay, B.G. (1994). Testing for the number of components in a mixture of normal distributions using moment estimators. *Comput. Statist. & Data Anal.*. **17**, 473–492.

Ghosh, J.H. and Sen, P.K. (1985). On the asymptotic performance of the log likelihood ratio statistic for the mixture model and related results. *Proc. Berkeley Conf. in Honor of Jerzy Neyman and Jack Kiefer* (Vol. II), L.M. Le Cam and R.A. Olshen (Eds.). Monterey: Wadsworth, 789–806.

Goffinet, B., Loisel, P., and Laurent, B. (1992). Testing in normal mixture models when the proportions are known. *Biometrika* **79**, 842–846.

Hall, P. and Titterington, D.M. (1989). The effect of simulation order on level accuracy and power of Monte Carlo tests. *J. R. Statist. Soc. B* **51**, 459–467.

Hartigan, J.A. (1985). A failure of likelihood asymptotics for normal mixtures. Proc. Berkeley Conf. in Honor of Jerzy Neyman and Jack Kiefer (Vol. II), L.M. Le Cam and R.A. Olshen (Eds.). Monterey: Wadsworth, 807–810.

Hoaglin, D.C. (1985). Using quantiles to study shape. In *Explaining Data Tables, Trends, and Shapes* D.C. Hoaglin, F. Mosteller, and J.W. Tukey (Eds.). New York: Wiley 417–460.

Hope, A.C.A. (1968). A simplified monte carlo significance test procedure. *J. R. Statist. Soc. B* **30**, 582-598.

Mangin, B., Goffinet, B., and Elsen, J.M. (1993). Testing in normal mixture models with some information on the parameters. *Biom. J.* **35**, 771–783.

McLachlan, G.J. (1987). On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *Appl. Statist.* **36**, 318–324

McLachlan, G.J. (1992). Discriminant Analysis and Statistical Pattern Recognition. New York: Wiley.

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

McLachlan, G.J., Peel D., Adams, P., and Basford, K.E. (1995). An algorithm for assessing by resampling the P-value of the likelihood ratio test on the number of components in normal mixture models. *Research Report No. 31*. Brisbane: Centre for Statistics, The University of Queensland.

Mendell, N.R., Finch, S.J., and Thode, H.C. (1993). Where is the likelihood ratio test powerful for detecting two component normal mixtures? *Biometrics* **49**, 907–915.

Mendell, N.R., Thode, H.C., and Finch, S.J. (1992). The likelihood ratio test for the two-component normal mixture problem: power and sample size analysis. *Biometrics* **47**, 1143–1148. Correction (1992). *Biometrics* **48**, 661.

Soromenho, G. (1994). Comparing approaches for testing the number of components in a finite mixture model. *Comput. Statist.* **9**, 65–78.

Thode, H.C., Finch, S.J., and Mendell, N.R. (1988). Simulated percentage points for the null distribution of the likelihood ratio test for a mixture of two normals. *Biometrics* **44**, 1195–1201.

Windham, M.P. and Cutler, A. (1992). Information ratios for validating mixture analyses. *J. Amer. Statist. Assoc.* **87**, 1188–1192.

Wolfe, J.H. (1971). A Monte Carlo study of sampling distribution of the likelihood ratio for mixtures of multinormal distributions. *Technical Bulletin STB 72-2.* San Diego: U.S. Naval Personnel and Training Research Laboratory