USING BOOTSTRAP LIKELIHOOD RATIO IN FINITE MIXTURE MODELS

Z.D. Feng
Division of Public Health Sciences
Fred Hutchinson Cancer Research Center
1124 Columbia Street, MP702
Seattle, WA 98104

C.E. McCulloch
Biometrics Unit and Statistics Center
Cornell University
Ithaca, New York, 14853

BU-1268-M

November 1994

Abstract

Statistical inference on the likelihood ratio statistic for the number of components in a mixture model is complicated when the true number of components is less than that of the proposed model since this represents a non-regular problem: the true parameter is on the boundary of the parameter space and in some cases the true parameter is in a nonidentifiable subset of the parameter space. The maximum likelihood estimator is shown to converge to the subset characterized by the same density function, and connection is made to the bootstrap method proposed by Aitkin *et al.* (1981) and McLachlan (1987) for testing the number of components in a finite mixture and deriving confidence regions in a finite mixture.

Key words: FINITE MIXTURE; LIKELIHOOD RATIO TEST; BOOTSTRAP; IDENTIFIABILITY; BOUNDARY.

1 INTRODUCTION

Finite mixture models have been widely used in biology, medicine and engineering (Everitt and Hand, 1981; Titterington *et al.*, 1985; McLachlan and Basford, 1988). Let $X = (x_1, \cdots x_n)$ be a random sample of size n from a probability distribution with a real-valued density function $f(x_i, \theta_0)$, where x can be univariate or vector-valued, discrete or continuous and θ_0 is the unknown true parameter vector of dimension p. Let $l(\theta; x_i)$ denote the log likelihood. A finite mixture density has the form

$$f(x, \theta) = \sum_{j=1}^{k} \pi_j f_j(x, \psi_j), \tag{1.1}$$

where ψ_j is the m-dimensional parameter vector for component j, π_j is the mixing probability for component j with restriction $\sum_{j=1}^{k} \pi_j = 1$ and $\pi_j \geq 0$, $\theta = (\pi_1, \cdots, \pi_{k-1}, \psi_1, \cdots, \psi_k)$, and $p = k - 1 + km$. The number of components, k, may be known or unknown. When the number of components is known, statistical inferential procedures about the parameters are well developed, mostly via likelihood based inference. Although the inferential problem for the number of components in a mixture has wide applications, it is still an open question without satisfactory treatment (Titterington, 1990). Suppose we want to test

$$H_0 : f(x, \theta) = N(0, 1) \tag{1.2}$$

against

$$H_1 : f(x, \theta) = (1 - \pi)N(0, 1) + \pi N(\mu, 1), \tag{1.3}$$

where $N(\mu, \sigma^2)$ is a Gaussian density with mean μ and variance σ^2. We have $\theta = (\pi, \mu) \in [0, 1]\text{x}(-\infty, \infty)$. The parameter space where H_0 holds is, $\Omega_0 = ([0]\text{x}(-\infty, \infty)) \bigcup ([0, 1]\text{x}[0])$, i.e., the entire μ axis when $\pi = 0$ and the line segment $[0, 1]$ on the π axis when $\mu = 0$. Therefore, the parameter is on the boundary of the parameter space and the null hypothesis corresponds to a nonidentifiable subset of the parameter space. The classic assumptions (Cramér, 1946) about the asymptotic properties of the maximum likelihood estimator and the likelihood ratio statistic are not valid under the null hypothesis.

There have been only conjectures and simulation results for the limiting distribution of the likelihood ratio statistic for the mixture models under the null hypothesis (Wolfe, 1971; Hartigan, 1977, 1985; McLachlan, 1987; Thode *et al.*, 1988). Ghosh and Sen (1985) showed that choosing an identifiable parameterization can create a problem of differentiability of the density. Bootstrapping the likelihood ratio to test the number of components of a normal mixture was investigated by Aitkin *et al.* (1981) and McLachlan (1987). McLachlan (1987) used a parametric bootstrap in which the parameter estimate was obtained from the maximum likelihood method under the null hypothesis, bootstrap samples were drawn from the null hypothesis with the estimated parameter and for each bootstrap sample the likelihood ratio was computed to form the reference distribution. Feng and McCulloch (1995) noted that bootstrap is a preferred method for testing the number of components of normal mixture with unequal variances. For regular cases, Beran (1988) and Martin (1990) studied level error and coverage probability of the bootstrap likelihood ratio on the testing and confidence region problems respectively. Since the maximum likelihood estimator under the alternative hypothesis is not a consistent estimator when the null hypothesis is true, there are questions about the validity of the bootstrap likelihood ratio test based on this inconsistent estimator, such as whether the observed rejection rates will match the hoped-for levels (Titterington, 1990).

This work provides some justification for the bootstrap method by showing that the maximum likelihood estimator converges to the nonidentifiable subset to which the true parameter belongs. This property has a natural connection to the bootstrap likelihood ratio tests and confidence regions. The test sizes and coverage probabilities of bootstrap methods match the nominal levels well in simulation studies when the null hypothesis is true.

2 MAIN RESULTS

Along with other classic regularity conditions, if the true parameter is an interior point of the parameter space then standard asymptotic theory applies. We extend the standard results by considering a situation where the true parameter lies in a non-identifiable subset, Ω_0, and this subset may be on the boundary of the parameter space. We first extend the definition of consistency and prove that the unrestricted maxima $\hat{\theta}$, (i.e., maximizing θ without restricting it to Ω) is consistent in the following sense: $\hat{\theta} - \theta_0^*(\hat{\theta}) \to 0$, with probability one for some $\theta_0^*(\hat{\theta}) \in \Omega_0$, where Ω_0 identifies a subset of Ω in which the distributions are not distinguishable. Consistency also holds for the maximum likelihood estimator in the same sense. We first extend the definition of $l(\theta; x)$ to $\Re^p$:

$$l^*(\theta; x) \equiv \sum_{i=1}^{n} log[f^*(\theta; x_i)1(f^*(\theta; x_i) > 0], \tag{2.1}$$

where $1(.)$ is an indicator function and $f^*(\theta, x_i)$ is the extension of $f(\theta, x_i)$ to all $\theta \in \Re^p$.

We need the following assumptions, with θ_0 an arbitrary fixed point in Ω_0 :

(A) the parameter space Ω has finite dimension.

(B) $f(x, \theta_0) = f(x, \theta_0')$ for all $\theta_0, \theta_0' \in \Omega_0$.

(C) there exists an open subset ω_ϵ of $\Re^p$ containing Ω_0, such that for almost all x, $f(\theta; x)$ admits all third derivatives w.r.t. θ for all $\theta \in \omega_\epsilon$, and

$$|\frac{\partial^3}{\partial\theta_j\partial\theta_k\partial\theta_l} logf(x;\theta)| \leq M_{jkl}(x, \theta),$$

and $m_{jkl}(\theta_0) = E_{\theta_0}[M_{jkl}(x, \theta_0)] < \infty$ for all j, k, l, for any fixed $\theta_0 \in \Omega_0$.

Remark: ω_ϵ is not necessarily an open ball and can be expressed as $\omega_\epsilon \equiv \bigcup_{\theta_0 \in \Omega_0} B_{\epsilon(\theta_0)}(\theta_0)$, for each $\epsilon(\theta_0) > 0$, depending on θ_0 and $B_\epsilon(.)$ is an open ball of radius ϵ cented at (.). In the example of (1.3) when (1.2) is true, ω_ϵ is an open stripe with unequal width surrounding Ω_0 with ϵ the maximum of $\epsilon(\theta_0)$s.

(D) $E_{\theta_0}[\frac{\partial}{\partial\theta_j} logf(x, \theta)] = 0$ for $j = 1, \cdots, p$ and all $\theta_0 \in \Omega_0$.

$$\begin{aligned} I_{jk}(\theta_0) &= E_{\theta_0}[\frac{\partial}{\partial\theta_j} logf(x, \theta) \frac{\partial}{\partial\theta_k} logf(x, \theta)] \\ &= E_{\theta_0}[\frac{\partial^2}{\partial\theta_j\partial\theta_k} logf(x, \theta)] \end{aligned}$$

for any $\theta_0 \in \Omega_0$. Also, for any $\theta \notin \Omega_0$, the $\theta_0 \in \Omega_0$ which is nearest to θ in Euclidean distance, (i.e. $|\theta - \theta_0| \leq |\theta - \theta_0'|$ for all $\theta_0' \in \Omega_0$), is selected and has the property that $(\theta - \theta_0)^t I(\theta_0)(\theta - \theta_0) > 0$ for all θ in the neighborhood of θ_0.

Remark: Notice that the above properties need not be held by all $\theta_0 \in \Omega_0$. For those θ_0 which will not be selected by the above rule the assumptions can be relaxed, i.e., the quadratic form can be zero. It is not difficult to check that for the MLE under (1.3)when (1.2) is true, (A)-(D) are satisfied.

Theorem 2.1.

Let $X = (x_1, \cdots, x_n)$ be independent, identically distributed observations with density $f(x, \theta)$ satisfying assumptions (A)-(D) above and with the true parameter θ_0 being any point in Ω_0 (The value of θ_0 is not important since all points in Ω_0 identify the same density function). Then with probability tending to 1 as $n \to \infty$, there exists a $\hat{\theta} \in \Re^p$, a local maxima of $l^*(\theta, X)$ as defined in (2.1), which has the property that there exists a $\theta_0^*(\hat{\theta}) \in \Omega_0$ which depends on $\hat{\theta}$ such that $\hat{\theta} - \theta_0^*(\hat{\theta}) \to 0$ with probability 1. Moreover, the maximum likelihood estimator, $\hat{\theta}_{ml} - \theta_0^*(\hat{\theta}_{ml}) \to 0$ with probability 1.

Proof. The proof is similar to that of Lehmann (1983) with modifications to adapt assumptions (B) and (C). We only need to show that for sufficient small $\epsilon(\theta_0) > 0$, $l^*(\theta, X) < l^*(\theta_0, X)$ at all points θ on the boundary of some stripe ω_ϵ surrounding Ω_0, since this means that there exists at least a local maxima within ω_ϵ. We can choose $\epsilon(\theta_0)$ small enough such that $f(x_i, \theta) > 0$ for all x_i's in the sample and Taylor expansion of $l^*(\theta, X)$ about θ_0 is justified in ω_ϵ. For any fixed θ on the boundary of ω_ϵ, we define $\theta_0^*(\theta)$, such that $|\theta - \theta_0^*(\theta)| \leq |\theta - \theta_0|$ for all $\theta_0 \in \Omega_0$, i.e., $\theta_0^*(\theta)$ is the point in Ω_0 closest to θ in Euclidian distance. Taylor expansion of $l^*(\theta, X)$ about $\theta_0^*(\theta)$ leads to:

$$\frac{1}{n}l^*(\theta, X) - \frac{1}{n}l^*(\theta_0^*(\theta, X)) = S_1 + S_2 + S_3,$$

with

$$S_1 = \frac{1}{n}\sum_{j=1}^{p}(\theta_j - \theta_{0j}^*(\theta))[\frac{\partial}{\partial\theta_j}l^*(\theta, X)|_{\theta=\theta_0^*(\theta)}],$$

$$S_2 = \frac{1}{2n}\sum_{j=1}^{p}\sum_{k=1}^{p}(\theta_j - \theta_{0j}^*(\theta))(\theta_k - \theta_{0k}^*(\theta))[\frac{\partial^2}{\partial\theta_j\partial\theta_k}l^*(\theta, X)|_{\theta=\theta_0^*(\theta)}],$$

$$S_3 = \frac{1}{6n}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p}(\theta_j - \theta_{0j}^*(\theta))(\theta_k - \theta_{0k}^*(\theta))(\theta_l - \theta_{0l}^*(\theta))\sum_{i=1}^{n}\gamma_{jkl}(x_i)M_{jkl}(x_i, \theta),$$

where $0 \leq |\gamma_{jkl}(x)| \leq 1$ by assumption (C).

The asterisk on the log-likelihood can be dropped in each term of the Taylor expansion since $\theta_0^*(\theta) \in \Omega$. Therefore, the rest of the proof of $S_1 + S_2 + S_3 < 0$ as $n \to \infty$ follows Lehmann (1983). The proof of convergence of $\hat{\theta}_{ml}$ is the same except Taylor expansion of $l^*(\theta, X)$ is about any fixed θ on the boundary of the intersection of ω_ϵ and Ω.

Remark: Redner (1981) proved the strong consistency of the maximum likelihood estimator in the quotient topological space. The above proof is the parallel result expressed in Euclidean space and is therefore easier for practitioners to apply. It is also valid for the true parameter being on the boundary of the parameter space and in a nonidentifiable subset while Redner's result dealt with the nonidentifiable situation only. If the parameter is identifiable but on the boundary of the parameter space, Self and Liang (1987) proved the consistency and gave the characteristic of the asymptotical distribution of the maximum likelihood estimator. Feng and McCulloch (1992) proposed an unrestricted maximum likelihood estimator which is consistent and has asymptotic normality. So, Theorem 2.1 is the extension of results of Redner (1981) and Self and Liang (1987). When Ω_0 is a single boundary point, Theorem 2.1 reduces to Self and Liang's consistency result of the maximum likelihood estimator. It reduces to Redner's result when Ω_0 is in the interior of

Ω, and to the classic consistency of the maximum likelihood estimator when Ω_0 is a single point in the interior of Ω. It also extends Feng and McCulloch's results to nonidentifiable cases.

3 CONNECTION TO BOOTSTRAP LIKELIHOOD RATIO

Following Beran (1988), Martin (1990), and McLachlan (1987), we denote

$$W(\theta, X) \equiv 2(l(\hat{\theta}, X) - l(\theta, X))$$

and

$$W(\theta_0, X^*(\theta_0)) \equiv 2(l(\hat{\theta}^*, X^*(\theta_0)) - l(\theta_0, X^*(\theta_0))),$$

where $\hat{\theta}$ and $\hat{\theta}^*$ are the maximum likelihood estimators under Ω from X and X^* respectively, and $X^*(\theta_0)$ means a bootstrap sample under θ_0, where θ_0 is contained in H_0. The size α bootstrap likelihood ratio test procedure for simple $H_0 : \theta = \theta_0$ is: Reject H_0 if

$$W(\theta_0, X) > W_\alpha(\theta_0, X^*(\theta_0))$$

where $W_\alpha(.,.)$ is the upper α quantile of $W(.,.)$.

For $\theta = (\theta_1, \theta_2)$ and a composite $H_0 : \theta_1 = \theta_{01}$, θ_2 unspecified, denote the maximum likelihood estimator under H_0 as $\hat{\theta}_0 = (\theta_{01}, \hat{\theta}_{02})$, The bootstrap test is then: Reject $H_0 : \theta_1 = \theta_{01}$ if

$$W(\hat{\theta}_0, X) > W_\alpha(\hat{\theta}_0, X^*(\hat{\theta}_0)).$$

By Theorem 2.1, when H_0 is true, the likelihood computed under H_1 is based on the maximum likelihood estimate converging to Ω_0 with probability 1. If this convergence does not hold, it is clear that the bootstrap likelihood ratio test is not valid. For example (1.2)-(1.3), Hartigan (1985) pointed out that $W(\theta_0, X)$ is asymptotically unbounded above in probability at a very slow rate ($\frac{1}{2} log\ log\ n$) when the null hypothesis is true. However, the distribution of W or W^* for any finite sample size does exist. Since the likelihood is identifiable while the parameters are not, the bootstrap likelihood ratio is a natural candidate for this inference problem as compared to other parameter-based tests (e.g. Wald's test).

We can define a bootstrap confidence region, $\hat{\Re}_\alpha$ as:

$$\hat{\Re}_\alpha \equiv \{\theta : W(\theta, X) \leq W_\alpha(\hat{\theta}, X^*(\hat{\theta}))\},$$

where $X^*(\hat{\theta})$ is the bootstrap sample from $F(\hat{\theta})$, i.e., the parametric bootstrap, $\hat{\theta}$ and $\hat{\theta}^*$ are the maximum likelihood estimates of θ from X and $X^*(\hat{\theta})$ respectively.

The validity of $\hat{\Re}_\alpha$ relies heavily on Theorem 2.1, since the resampling is from $F(\hat{\theta})$, the maximum likelihood estimate under H_1, while in the hypotheis testing situation the resampling is from the null hypothesis.

For the hypothesis testing in (1.2)-(1.3), bootstrap samples are generated from the standard normal distribution. The maximum likelihood estimate $(\hat{\pi}, \hat{\mu})$ by Theorem 2.1 will converge to $\Omega_0 : ([0]\text{x}(-\infty, \infty)) \bigcup([0,1]\text{x}[0])$ which identifies the standard normal density. Therefore, we expect the bootstrap test to have a high probability of accepting H_0 when the sample size is sufficiently large and H_0 holds. For the confidence region based on the model of (1.3) when the unknown true distribution is a standard normal, $X^*(\hat{\theta})$ is sampled from (1.3) with $\hat{\theta}$ converging to Ω_0. Therefore, we expect the bootstrap confidence region $\hat{\Re}_\alpha$ to have a high probability of intersecting Ω_0 for a large sample size. Since the asymptotic normality of the unrestricted maximum likelihood estimator by Feng and McCulloch (1992) does not hold under nonidentifiable cases and the asymptotical distribution for the maximum likelihood estimator is unknown, the critical values for tests and confidence regions should come from bootstrapping.

4 SIMULATION STUDIES

We conducted two simulation studies to examine the acceptance probabilities of the bootstrap likelihood tests and the coverage probabilities of the bootstrap likelihood confidence regions under null hypotheses. We focus on the null hypothesis since this is where the classic asymptotic distribution theory for the likelihood ratio fails. Table 1 describes the simulation results of bootstrap likelihood ratio test for a mixture normal alternative: $(1-\pi)N(1,1)+\pi N(0,1)$ with the true density standard normal. This is the case where the parameter is on the boundary of Ω. We also included the likelihood ratio test based on the unrestricted maximum by Feng and McCulloch (1992) which has a limiting chi-square distribution with one degree of freedom. The probabilities of correctly accepting H_0 for both procedures are near the nominal level when the sample size N is 100, but the bootstrap procedure outperforms the likelihood ratio method. This difference is clear for smaller sample sizes. The likelihood ratio method performed poorly when n=10 with the probability of making a correct conclusion from 0.128 to 0.180 less than that of the nominal level while the bootstrap procedure has the probability only 0.010 to 0.018 away from the nominal level. Table 2 summarizes the coverage probabilities of bootstrap confidence regions for a mixture normal model: $(1-\pi)N(0,1)+\pi N(\mu,1)$ when the true distribution is a standard normal. This is the case where the parameter is on the boundary of Ω and in an nonidentifiable subset Ω_0. The bootstrap confidence procedure is again clearly better than the confidence region based on χ_1^2 (a χ_2^2 approximation was even worse in this simulation). We should mention that neither χ_1^2 nor χ_2^2 has a theoretical basis for use. The bootstrap confidence region has good coverage probabilities that are very close to the nominal levels in all sample sizes at all nominal levels, while the chi-squared approximation, although not bad in 0.95 and 0.99 nominal levels, performed poorly at the 0.90 nominal level.

5 CONCLUSIONS AND OPEN QUESTIONS

This paper provides some justification for the bootstrapping likelihood ratio when the true parameter is on the boundary of the parameter space and in a nonidentifiable set. It shows that the maximum likelihood estimate is consistent to the set identifying the true density function. Therefore, a procedure based on the likelihood is justified although its asymptotic distribution may be difficult to obtain and one needs to use the bootstrap method to construct tests or confidence regions. The parametric bootstrap confidence region works very naturally under this setting since bootstrap resamples from a consistently estimated density. Therefore, to a large extent we overcome the difficulties of the fact that the density is characterized by a set of the parameters and this set is on the boundary of the parameter space. The simulations indicate that the bootstrap tests and confidence region seem to hold the nominal levels and the expected coverage probabilities.

To fully justify the bootstrap likelihood ratio method one needs to obtain the exact convergence rate of the bootstrap approximation of the distribution of sample quantiles. The singular Fisher information is the major difficulty in investigating the error rate of the bootstrap procedure for the nonidentifiable case. As in Singh (1981), the proof of the error rate of the bootstrap procedure depends on the existence of an Edgeworth expansion of W. Bhattacharya (1985) proved that if $W = 2n(H(\bar{Z}) - H(\mu))$ where $\bar{Z} = n^{-1}(Z_1 + \cdots + Z_n)$ with $\mu = EZ_1$ and with some other regularity conditions, the Edgeworth expansion of W is valid. Chandra and Ghosh (1979, p.42) pointed out that if the assumptions (A_1) to (A_4) and (A_6) of Theorem 3 of Bhattacharya and Ghosh (1978) are satisfied, the Edgeworth expansion of the cdf of W agrees with the true distribution up to $o_p(n^{-1})$. Unfortunately, (A_4) is that the Fisher information is nonsingular. The importance of (A_4) is that it enables us to apply the implicit function theorem to ensure that there exists a uniquely defined real-valued infinitely differentiable function H on the neighborhood of μ. The difference between $n^{\frac{1}{2}}(\hat{\theta} - \theta_0)$ and its Edgeworth representation is then of order $o(n^{(s-2)/2})$, where s is the positive integer such that the *ith* derivative of the density function with respect to every θ is continuously differentiable for $1 \leq i \leq s$. The representation of W as a sum of iid random variables is necessary in all three of the above papers' proof. Therefore, a possible approach is to develop some non-Edgeworth expansion method or other criterion to justify the Edgeworth expansion.

ACKNOWLEDGEMENT

The authors would like to thank Dr. R.A. Redner for the helpful communications.

REFERENCES

Aitkin, M., Anderson, D. and Hinde, J. (1981). Statistical modeling of data on teaching styles (with discussion). *Journal of the Royal Statistical Society, Series A*, 144, 419-461.

Beran, R. (1987). Prepivoting test statistics: a bootstrap view of asymptotic refinements. *Journal of the American Statistical Association*, 83, 687-697.

Bhattacharya, R.N. and Ghosh, J.K. (1978). On the validity of the formal Edgeworth expansion. *Annals of Statistics*, 6, 434-451.

Bhattacharya, R.N. (1985). Some recent results on Cramér-Edgeworth expansions with applications. In *Multivariate Analysis - VI.* Krishnaiah, P.R. (Ed.). Elsevier Science Publisher B.V. 57-75.

Chandra, T.K. and Ghosh, J.K. (1979). Valid asymptotic expansions for the likelihood ratio statistic and other perturbed chi-square variables. *Sankhya.* 41, 22-47.

Cramér, H. (1946). Mathematical Methods of Statistics, Princeton, NJ: Princeton University Press.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39, 1-38.

Edlefsen, L.E., and Jones, S.D. (1988). *GAUSS version 2.0.* Aptech System, Inc. Kent, WA.

Everitt, B.S. and Hand, D.J. (1981). Finite Mixture Distributions. Chapman and Hall, London.

Feng, Z. and McCulloch, C.E. (1992). Statistical inference using maximum likelihood estimation and the generalized likelihood ratio when the true parameter is on the boundary of the parameter space. *Statistics & Probability Letters*, 13, 325-332.

Feng, Z. and McCulloch, C.E. (1995). On the likelihood ratio test statistic for the number of components in a normal mixture with unequal variance. *Biometrics*, (In Press).

Ghosh, J.M. 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. Neyman & Kiefer*, II, 789-806. Wadsworth, Monterey.

Hartigan, J.A. (1977). Distribution problems in clustering. In *Classification and Clustering*, J. Van Ryzin (Ed.) New York: Academic Press, pp. 45-71.

Hartigan, J.A. (1985). A failure of likelihood asymptotics for normal mixtures. *Proc. Berkeley Conf. Neyman & Kiefer*, II, 807-810. Wadsworth, Monterey.

Lehmann, E.L. (1983). *Theory of Point Estimation.* John Wiley and Sons, New York.

Martin, M. (1990). On bootstrap iteration for coverage correction in confidence intervals. *Journal of the American Statistical Association*, 85, 1105-1118.

McLachlan, G.J. (1987). On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. A class of statistics with asymptotically normal distribution. *Applied Statistics*, 36, 318-324.

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

Redner, R.A.(1981). Note on the consistency of the maximum likelihood estimate for nonidentifiable distributions. *Annals of Statistics*, 9, 225-228.

Self, S.G. and Liang, K.Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82, 605-610.

Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Annals of Statistics*, 9, 1187-1195.

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

Titterington, D.M., Smith, A.F.M. and Makov, U.E. (1985). *Statistical Analysis of Finite Mixture Distributions.* Wiley, London.

Titterington, D.M. (1990). Some Recent Research in the Analysis of Mixture Distributions. *Statistics,* 21, 619-641.

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

Nominal 1-α	Sample Size	Bootstrap Likelihood Ratio(s.e)	Feng and McCulloch's Likelihood Ratio(s.e)
	10	0.882 (0.014)	0.720 (0.020)
0.90	30	0.892 (0.014)	0.824 (0.017)
	100	0.894 (0.014)	0.864 (0.015)
	10	0.938 (0.011)	0.796 (0.018)
0.95	30	0.950 (0.010)	0.884 (0.014)
	100	0.948 (0.010)	0.924 (0.012)
	10	0.980 (0.006)	0.862 (0.015)
0.99	30	0.998 (0.002)	0.956 (0.009)
	100	0.990 (0.004)	0.982 (0.006)

Table 1: Probabilities of acceptance of $H_0 : N(0,1)$ vs. $H_1 : \pi N(0,1) + (1-\pi)N(1,1)$ when H_0 is true in 500 simulations.

Nominal 1-α	Sample Size	Bootstrap Likelihood Ratio(s.e)	Likelihood Ratio(s.e) (χ_1^2)
	10	0.916 (0.012)	0.878 (0.015)
0.90	30	0.918 (0.012)	0.856 (0.016)
	100	0.916 (0.012)	0.862 (0.015)
	10	0.946 (0.010)	0.926 (0.012)
0.95	30	0.954 (0.009)	0.924 (0.012)
	100	0.966 (0.008)	0.926 (0.012)
	10	0.988 (0.005)	0.986 (0.005)
0.99	30	0.984 (0.006)	0.970 (0.008)
	100	0.996 (0.003)	0.992 (0.004)

Table 2: Coverage Probabilities of bootstrap and the likelihood ratio based confidence regions for a mixture normal model:$(1-\pi)N(0,1) + \pi N(\mu, 1)$ when the true distribution is $N(0,1)$ in 500 simulations.