Bootstrap Confidence Regions Using the Likelihood Ratio Statistic in Mixture Models

by

Ziding Feng[1] and Charles E. McCulloch[2]

BU-1114-M February 1991

Summary

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 an 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. Bootstrap confidence regions based on the likelihood ratio statistic are shown by analysis and Monte Carlo simulation to be superior to the traditional likelihood based confidence region in all three cases: regular case, simple boundary case, and nonidentifiable, boundary case.

1. Introduction

Mixture models have been widely used in biology, medicine and engineering (Titterington, Smith and Makov, 1985). When the number of components is known, the statistical inferential procedures about the parameters are well developed, mostly via likelihood based inferences. However, inferential procedures for the number of components in a mixture is still an open question. If there are possibly k components in a mixture, the null hypothesis of k' components where $k' < k$ corresponds to the parameter being on the boundary of the parameter space. Furthermore, if there exists at least one unknown parameter, nuisance or of interest, in addition to the mixing parameters, the above null hypothesis often corresponds to an nonidentifiable subset of the parameter space. The classic

[1] Ziding Feng was a graduate student in the Biometrics Unit and Statistics Center, Cornell University. He is now Assistant Member, Fred Hutchinson Cancer Research Center, Seattle, WA 98104.

[2] Charles E. McCulloch is an Associate Professor, Biometrics Unit and Statistics Center, Cornell University, Ithaca, NY 14853.

assumptions (Cramer, 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 likelihood ratio statistic for the mixture models (Wolfe, 1971; Hartigan, 1979, 1985; McLachlan, 1987; Thode, Finch and Mendell, 1988). Redner (1981) extended Wald's results on consistency of the maximum likelihood estimator to the nonidentifiable case in a quotient space approach. Ghosh and Sen (1985) showed that choosing an identifiable parametrization can create a problem of differentiability of the density. Due to the difficulties discussed above, statisticians have turned to the bootstrap method. Bootstrapping the likelihood ratio to test number of components of a normal mixture was investigated via simulation by McLachlan (1987). Beran (1988) showed that bootstrapping the likelihood ratio test in the regular case automatically accomplishes the Bartlett adjustment and has level error $0(n^{-3/2})$. For general theoretical discussions of the performance of the bootstrap procedure see Bickel and Freedman (1981), Singh (1981), Beran (1984), Hinkley (1988) and Hall (1986, 1988). There does not appear to be any work regarding the theoretical evaluation of the bootstrap method under non-regular conditions as in mixture models.

This paper first extends Beran's result to confidence regions in a more general setting, namely by showing the bootstrap likelihood confidence region has level error $0(n^{-3/2})$ both in the regular case and in cases where the parameter is on the boundary of the parameter space. Some theoretical justification and Monte Carlo simulation are provided for bootstrap, likelihood ratio confidence regions in the cases where the parameter is on the boundary of the parameter space and in a nonidentifiable subset of the parameter space.

2. Level error of bootstrap likelihood ratio confidence region in regular and boundary cases

Let $X=(x_1,...,x_n)$ be an iid sample of size n from a population with probability density $f(x,\theta)$, where θ is an unknown parameter of order p, $\theta \in \Omega$. Let $l(\theta, X)$ denote the log-likelihood. Under the classical regularity condition, the generalized likelihood ratio test of H_0: $\theta = \theta^*$ level α is:

$$\text{Reject } H_0 \text{ if} \quad 2(l(\hat{\theta},x) - l(\theta^*,x)) \leq \chi^2_{p,\ \alpha} \tag{2.1}$$

The 100(1-α)% confidence region based on inverting the likelihood ratio test is:

$$\Re_\alpha = \{\theta\text{: } 2(l(\hat{\theta},x) - l(\theta,x)) \leq \chi^2_{p,\ \alpha}\ \} \tag{2.2}$$

In general, this confidence region has the correct 1-α coverage only asymptotically. To improve the confidence region in finite samples, the bootstrap of likelihood-based confidence regions was introduced by Hall (1987). The statistic which he bootstraps is $n^{1/2}\hat{V}^{-1/2}(\hat{\theta}-\theta)$ where $\hat{\theta}$ and $\hat{V}$ are, respectively, the maximum likelihood estimate of θ and an estimate of the variance matrix of $n^{1/2}(\hat{\theta}-\theta)$. A likelihood-based region is a random region $\Re_\alpha$ which has the property that all parameter values inside $\Re_\alpha$ have higher likelihood than those outside (Cox & Hinkley, 1974, p. 218) and the bootstrap procedure enables us to construct $\Re_\alpha$ based on the empirical distribution F_n.

A drawback of the above approach is that this statistic is of dimension p, the dimension of θ. A contour device needs to be used to get the bootstrap confidence region and this is difficult to express analytically. Therefore, the use of it by practitioners is limited. Hall's procedure can be totally nonparametric if $\hat{\theta}$ and $\hat{V}$ are obtained by some nonparametric procedure and, e.g., a kernel density estimator is used to draw the contours.

For cases in which the likelihood can be specified, it is desirable to reduce the problem to a one dimensional statistic. Two natural candidates are the quadratic form

$$Q(\hat{\theta},\theta) \equiv n(\hat{\theta}-\theta)^T\hat{V}^{-1}(\hat{\theta}-\theta)$$

and the likelihood ratio statistic

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

where $l(\theta,X)$ is the log likelihood evaluated at θ.

Hall (1987) argued that likelihood-based regions should not be approximated by ellipses, if we are to have any hope in capturing first-order departures from normality. This rules out the possibility of using a quadratic form as the statistic to bootstrap.

We now introduce the bootstrap testing and confidence region procedures. The *bootstrap* *likelihood ratio test* procedure for simple H_0: $\theta = \theta_0$ is:

$$\text{Reject } H_0 \text{ if } \quad W(\theta_0, X) \leq W_\alpha(\theta_0, X^*(\theta_0)) \tag{2.3}$$

where $W_\alpha(.,.)$ is the upper α quantile of $W(.,.)$, and $X^*(\theta_0)$ means bootstrap resampling under θ_0, the parameter under H_0 and

$$W(\theta_0, X^*(\theta_0)) = 2(l(\hat{\theta}, X^*(\theta_0)) - l(\theta_0, X^*(\theta_0))) \tag{2.4}$$

where $\hat{\theta}$ is the maximum likelihood estimator under Ω from X^*.

We should emphasize that $\hat{\theta}$ depends on the sample from which it is calculated, and therefore it really is $\hat{\theta}(X)$ or $\hat{\theta}(X^*(.))$, For composite H_0 with $\theta = (\theta_1, \theta_2)$ and an H_0 of the form 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

$$W(\hat{\theta}_0, X) \leq W_\alpha(\hat{\theta}_0, X^*(\hat{\theta}_0)) \tag{2.5}$$

The inversion of (2.3) and (2.5) to get confidence regions $\Re_\alpha$'s are:

$$\Re_\alpha \equiv \{ \theta\text{: } W(\theta, X) \leq W_\alpha(\theta, X^*(\theta) \}. \tag{2.6}$$

Beran (1988) proved that (2.3) and (2.5) have level error $O(n^{-3/2})$. This means that the actual size of the test is $\alpha + O(n^{-3/2})$. When we are inverting (2.3) to form a confidence region, the 100(1-α)% confidence region is (2.6). This is computationally difficult since the bootstrap distribution under each θ has to be obtained, usually by a Monte Carlo simulation, to decide the critical value for that particular θ. On the other hand, if W is a pivot, then the distribution is independent of θ and any θ can be used to generate the bootstrap samples and get the critical value when inverting the likelihood ratio-tests into a confidence region. However, in general, W is a pivot only asymptotically, since its distribution approaches a chi-squared distribution. In finite samples when W is not a pivot, the bootstrap sample generated from each θ has a different distribution and therefore has a different critical value. Below we introduce a procedure which is easier to compute. That is, we show that it is permissible to use $\hat{\theta}$ or $\theta(\hat{F})$ to get the critical value and the level error remains the same as Beran's bootstrap likelihood ratio test.

Define the new bootstrap confidence region $\hat{\mathfrak{R}}_\alpha$:

$$\hat{\mathfrak{R}}_\alpha \equiv \{ \theta\colon \; W(\theta,X) \le W_\alpha(\hat{\theta},X^*(\hat{\theta})) \} \tag{2.7}$$

where $X^*(\hat{\theta})$ is 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,

Theorem 2.1.

Assume that the joint distributions of the components of $\hat{\theta}$, and the first four derivatives of the log likelihood, after standardization by location and scale have multivariate Edgeworth expansions. We further assume that the first five derivatives of the log likelihood exist. Then we have for any real x,

$$\Pr\{W(\hat{\theta},X^*(\hat{\theta})) \le x\} - \Pr\{W(\theta,X) \le x\} = O(n^{-3/2}) \tag{2.8}$$

Furthermore, the boundaries of confidence regions defined by (2.6) and (2.7) differ by the order $O_p(n^{-3/2})$. This means if θ_α denotes any one point in $\mathfrak{R}_\alpha$ constructed by (2.6), then

$$\Pr\{W(\hat{\theta},X^*(\hat{\theta})) \le x\} - \Pr\{W(\theta_\alpha,X^*(\theta_\alpha)) \le x\} = O(n^{-3/2}) \tag{2.9}$$

and

$$\Pr(\theta \in \hat{\mathfrak{R}}_\alpha) = \alpha + O(n^{-3/2}) \tag{2.10}$$

i.e. $\hat{\mathfrak{R}}_\alpha$ has level error $O(n^{-3/2})$

Proof: (in Appendix).

Remark 1:

Barndorff-Nielsen and Hall (1988) applied a Taylor expansion and Edgeworth expansion to $(1+n^{-1}\hat{b}_1)^{-1}W$, the Bartlett adjustment of W, to prove the level error of the Bartlett adjustment on the likelihood ratio statistic. The proof of Theorem 3.1 uses a similar technique based on a Taylor expansion and Edgeworth expansion of $W(\hat{\theta},X^*(\hat{\theta}))$ and $W(\theta,X)$ and then a comparison of the difference. The asymptotic expansion of the likelihood ratio statistic has been studied by Lawley (1956), Hayakawa (1976), Chandra and Ghosh (1979), Chandra (1985) and Bhattacharya (1985). The details of the existence of Edgeworth expansions are in Bhattacharya and Ghosh (1978).

Hall (1986) investigated the level error of the bootstrap confidence interval in the class of "studentized" statistics which includes the maximum likelihood estimator when it can be expressed as a function of the sample mean. Our finding is different from Hall's in two aspects: 1) $-2\log\lambda$ does not have to be a function of the sample mean, i.e. a "studentized statistic"; 2) θ can be p-dimensional where $p \geq 1$.

Remark 2:

The reason the bootstrap likelihood ratio is superior to the bootstrap quadratic form is that the quadratic form uses a confidence region which is a symmetric ellipsoid. This is the correct shape of a confidence region only in the limiting case, i.e. as $n \to \infty$, while the first one is based on the likelihood and able to capture the asymmetry in finite samples.

For the case where the true parameter is on the boundary of the parameter space, Theorem 2.2 below indicates that the bootstrap confidence region or test still has the level error $0(n^{-3/2})$. For the assumptions and proof of the theorem see Feng and McCulloch (1990). The basic idea is to enlarge the parameter space and not force the maximum likelihood estimate of θ to lie in the original parameter space. We call such an estimate the *unrestricted maxima*.

Theorem 2.2.

Let $X=(x_1,...,x_n)$ be i.i.d. observations with density $f(x,\theta_0)$ with some regularity assumptions but where the unknown θ_0 is on the boundary of Ω. Then with probability tending to 1 as $n \rightarrow \infty$, there exists a $\hat{\theta} \in \mathbb{R}^p$, a local maxima, with the property that

$$\text{(i)} \quad \hat{\theta} \rightarrow \theta_0 \quad \text{w.p. } 1$$

$$\text{(ii)} \quad n^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{D} N(0, I(\theta_0)^{-1}).$$

Remark:

For bootstrap likelihood ratio test or bootstrap confidence region where the unrestricted maxima still identifies a probability distribution, either the parametric or nonparametric bootstrap can be used. For the bootstrap confidence region procedure where the unrestricted maxima does not identify a probability distribution when it is out of the parameter space, only the nonparametric bootstrap can be used to generate the bootstrap sample. In both cases, the maximum likelihood estimator can be used to generate the bootstrap sample, and the level error probably remains the same but is difficult to obtain since the maximum likelihood estimator does not have an asymptotic normal distribution when the true parameter is on the boundary of the parameter space and the assumptions for Theorem 2.1 are not satisfied.

We illustrate the use of the bootstrap likelihood ratio method and compare it to the classic likelihood ratio method by using a Monte Carlo simulation of the level error of the likelihood confidence regions obtained by both methods. We consider two cases: a multinomial mixture and a mixture of two normals. The simulations were programmed in GAUSS (Edlefsen and Jones, 1988).

The mixture multinomial model has been used in genetics by McCulloch (1987) and Roeder, Devlin and Lindsay (1989). We here follow the notation of McCulloch. Suppose we observe $f=(f_1, ..., f_G)$ sampled from multinomial $(N,\mathbf{p})$ with parameters:

$$N = \sum_{j=1}^{G} f_j , \quad \mathbf{p} = (p_1, ..., p_G)^T$$

$$\text{and } p_j = \sum_{i=1}^{s} \pi_i p(j|i)$$

$$\text{with } 0 \le \pi_i \le 1 \text{ and } \sum_{i=1}^{s} \pi_i = 1.$$

When the p(j|i)'s are assumed known and when at least one $\pi_i = 0$ this is the case of a parameter on the boundary of the parameter space. The coverage probability of the bootstrap likelihood ratio method by (2.7) and the likelihood ratio confidence regions for π's based on 500 simulations are given in Table 1 and 2. A bootstrap sample size of 1000 was used to get the cutoff point of the simulated bootstrap distribution for each single replicate of the simulation. Sample sizes of 10, 25, 50 and 100 are used to investigate the convergence rate of the bootstrap likelihood ratio and the likelihood ratio statistic. 1 - α = 0.9 is used. Table 1 represents the boundary problem with:

$$\pi = (1.0\ 0.0),\ P^* \equiv (p(j|i)) = \begin{bmatrix} .5 & .5 \\ 1 & 0 \end{bmatrix} \quad \text{and}$$

Table 2 represents the regular problem with:

$$\pi = (0.6\ 0.35\ 0.05),\ P^* \equiv (p(j|i)) = \begin{bmatrix} .5 & .25 & .25 \\ 0 & .75 & .25 \\ .875 & .125 & 0 \end{bmatrix}.$$

Table 1 indicates that bootstrap likelihood ratio method outperforms the likelihood ratio method since the coverage probability is not significantly different from 0.9 when N $\ge$ 25, while the likelihood ratio method has coverage probability significantly less than 0.9 even at N = 50. Note that when N = 100, the likelihood ratio method has coverage probability 0.884, supporting the fact that it has asymptotic correct coverage probability as from Theorem 2.2. The results of Table 2 indicate that the bootstrap method still outperforms the likelihood ratio method but on the whole both methods performed pretty well in this example.

Table 3 describes the simulation results of the bootstrap and likelihood ratio test procedures based on the unrestricted maxima for mixture normal alternative $(1-\pi)N(1,1) + \pi N(0,1)$ with the true population being standard normal. 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 still outperformed the likelihood ratio method. This difference was clearer for small 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 the nominal level while the bootstrap procedure has the probability only 0.010 to 0.018 away from the nominal level.

Table 4 gives the simulation results for the same model as used in Table 3 except that the coverage probabilities of bootstrap confidence region constructed by (2.7) are compared with the confidence region constructed by the likelihood ratio statistic based on the unrestricted maxima. First of all, when $n = 100$, both procedures performed very well. The bootstrap confidence region still has quite an accurate coverage probability when the sample size is as low as 10 while the likelihood ratio method begins to perform poorly when $n = 30$ and is much worse when $n = 10$.

3. Bootstrap confidence region for nonidentifiable case

We first extend the definition of consistency and prove that the unrestricted maxima $\hat{\theta}$ 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. Simulation results are provided to evaluate the bootstrap, likelihood ratio confidence region procedure for the model:

$$(1-\pi)\ N(0,1) + \pi\ N(\mu,1) \tag{3.1}$$

when the true distribution is N(0,1).

We first extend the definition of $\ell(\theta; \mathbf{x})$ to $\mathbb{R}^p$:

$$\ell^*(\theta; \mathbf{x}) \equiv \sum_{i=1}^{n} \ell og[f^*(\theta; x_i)1(f^*(\theta; x_i) > 0)] \tag{3.2}$$

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

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

(A) the parameter space Ω has finite dimension.

(B) $f(x,\theta_0) = f(x,\theta_0') \ \forall \ \theta_0, \theta_0' \in \Omega_0$. That is, all parameters in Ω_0 generate the same distribution and therefore the distribution is not identifiable in Ω_0.

(C) there exists an open subset ω_ϵ of $\mathbb{R}^p$ 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

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

and $m_{jkl}(\theta_0) = E_{\theta_0}[M_{jkl}(x,\theta_0)] < \infty \ \forall$ 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. In the above example, ω_ϵ is then an open stripe with unequal width surrounding Ω_0.

(D) $E_{\theta_0}[\frac{\partial}{\partial\theta_j}\log f(X,\theta)] = 0$ for $j = 1,\ldots, p$ and all $\theta_0 \in \Omega_0$.

$$I_{jk}(\theta_0) = E_{\theta_0}[\frac{\partial}{\partial\theta_j}\log f(X,\theta)\ \frac{\partial}{\partial\theta_k}\log f(X,\theta)]$$

$$= E_{\theta_0}[-\frac{\partial^2}{\partial\theta_j\partial\theta_k}\log f(X,\theta)] \text{ for any } \theta_0 \in \Omega_0, \text{ also,}$$

for any $\theta \notin \Omega_0$, the $\theta_0 \in \Omega_0$ which is nearest to θ in Euclidian distance, (i.e. $|\theta - \theta_0| \leq |\theta - \theta_0'| \ \forall \ \theta_0' \in \Omega_0$), 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 the above example satisfies the (A) – (D) with Fisher information evaluated in Ω_0 as follows:

$$I((p,0)^T) = \begin{bmatrix} 0 & 0 \\ 0 & p^2 \end{bmatrix}, \; I((0,\mu)^T) = \begin{bmatrix} e^{\mu^2} - 1 & 0 \\ 0 & 0 \end{bmatrix}$$

and

$$I((0,0)^T) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

For any $\theta = (p, \mu)^T$ not in Ω_0, $(\theta - \theta_0)^T I(\theta_0)(\theta - \theta_0)$ is as follows:

If $|p| < |\mu|$, i.e., it is near the μ-axis, then:

$$((p-0),(\mu-\mu))I((0, \mu)^T)((p-0),(\mu-\mu))^T = p^2(e^{\mu^2} - 1) > 0.$$

On the other hand, if $|p| > |\mu|$, then

$$((p-p),(\mu-0))I((p, 0)^T)((p-p),(\mu-0))^T = p^2\mu^2 > 0.$$

For $|p| = |\mu|$, choose either $(0,\mu)$ or $(p,0)$. Though

$I(\theta) = 0$ at $\theta = (0,0)$, this point is never selected by the selection rule.

Theorem 3.1.

Let $X = (x_1,\ldots,x_n)$ be i.i.d. 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 $\hat{\theta} \in \mathbb{R}^p$, a local maxima of $l^*(\theta, X)$ as defined in (3.2), which has the property that there exists a $\theta_0^*(\hat{\theta}) \in \Omega_0$ which depends on $\hat{\theta}$ such that

$$\hat{\theta} - \theta_0^*(\hat{\theta}) \quad \to \quad 0 \quad \text{with probability 1}$$

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 sufficiently small $\epsilon(\theta_0) > 0$, $l^*(\theta, X) < l^*(\theta_0, X)$ at all points θ on the boundary of some stripe w_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 $f(x_i,\theta) > 0 \; \forall \; x_i$'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 $|\theta - \theta_0^*(\theta)| \le |\theta - \theta_0| \; \forall \; \theta_0 \in \Omega_0$, 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:

$$\tfrac{1}{n}l^*(\theta,X) - \tfrac{1}{n}l^*(\theta_0^*(\theta,X)) \tag{3.3}$$

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

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

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

$$\equiv S_1 + S_2 + S_3$$

and $0 \leq |\gamma_{jkl}(x)| \leq 1$ 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 the classic one (Lehmann, 1983).

Remark:

Redner (1981) proved the strong consistency of the maximum likelihood estimator in the quotient space. The above proof is the parallel result expressed in Euclidean space and is therefore easier to interpret. This approach might also lead to obtaining the asymptotic distribution of -2logλ^*, the likelihood ratio statistics based on the unrestricted maxima. However, we have not been able to derive the asymptotic distribution of -2logλ^* or -2logλ. Hartigan (1985) pointed out that for the example of (3.1), -2logλ is asymptotically not bounded in probability but tends to infinity at a very slow rate ($\frac{1}{2}$log log n). This implies the asymptotic distribution of -2logλ does not exist in this case. However, the distribution of -2logλ or -2logλ^* for any finite sample does exist. This suggests that the bootstrap procedure is a natural candidate for this inference problem since it mimics the underlying finite sample distribution. The significance of Theorem 3.1 is that since the maximum likelihood estimator converges to Ω_0, the bootstrap procedure should asymptotically mimic the true distribution though the theoretical level error of the bootstrap procedure is difficult to obtain. The merit is that every point in Ω_0 corresponds to the same distribution and therefore the bootstrap distributions are asymptotically the same no matter which point in ω_ϵ is used to generate bootstrap samples.

Simulation of the coverage probability of bootstrap confidence region was carried out on Cornell's supercomputer. -2logλ and -2logλ^* are the statistics we bootstrapped. A subroutine, DBCONF, which uses a quasi-Newton and a finite-difference gradient method in IMSL was called in a FORTRAN program to find the maximum likelihood estimators from random samples of size 100, 30 and 10. Then 500 bootstrap samples were generated by the parametric bootstrap method and the 90%, 95% and 99% cut off points for the distribution of -2logλ (or -2logλ^*) were obtained from 500 simulated -2logλs (or -2logλ^*s) each computed from its corresponding bootstrap sample to form the confidence region as (2.7). This process was repeated 500 times to evaluate the coverage probability.

Table 5 indicates that the bootstrap confidence procedure is again clearly better than the confidence region based on ${\chi^2}_1$ (${\chi^2}_2$ approximation was even worse in this simulation). We should mention that neither ${\chi^2}_1$ nor ${\chi^2}_2$ 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 size and at all nominal levels, while the chi-squared approximation, though not bad in 0.95 and 0.99 nominal levels, performed poorly at the 0.90 nominal level.

The simulation results of the coverage probabilities of the bootstrap confidence region for the above example are in Table 6, the only difference between Table 6 and Table 5 being that the bootstrap statistic is -2logλ^*, i.e., based on the unrestricted maxima. Table 6 indicates that the coverage probability based on ${\chi^2}_1$ is much too liberal while that based on ${\chi^2}_2$ is a little too conservative. The bootstrap procedure is again a winner at all sample sizes.

4. Discussion

The singular Fisher information is the major difficulty in investigating the level error of the bootstrap procedure for the nonidentifiable case. As we saw in section 2, the proof of the level error of the bootstrap procedure depends on the existence of an Edgeworth expansion of $-2\log\lambda$. Bhattacharya (1985) proved that if $-2\log\lambda = 2n(H(\bar{Z}) - H(\mu))$ where $\bar{Z} = n^{-1}(Z_1 + ... + Z_n)$ with $\mu = EZ_1$ and with some other regularity conditions, the Edgeworth expansion of $-2\log\lambda$ 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 $-2\log\lambda$ 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 an uniquely defined real-valued infinitely differentiable function H on the neighborhood of μ. The difference between $n^{1/2}(\hat{\theta} - \theta_0)$ and its Edgeworth representation is then of the 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 $-2\log\lambda$ 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.

Appendix:

Proof of Theorem 2.1. By the assumption of the continuity of the derivatives of the log likelihood, the Taylor expansion of $W(\theta,X) \equiv 2(l(\hat{\theta},X) - l(\theta,X))$ about θ is:

$$W(\theta,X) = 2\sum_{j=1}^{p} (\hat{\theta}_j - \theta_j)\, l'_j(\theta,X) + \sum_{j=1}^{p}\sum_{k=1}^{p} (\hat{\theta}_j-\theta_j)(\hat{\theta}_k-\theta_k)\, l''_{jk}(\theta,X)$$

$$+ \frac{1}{3}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p} (\hat{\theta}_j-\theta_j)(\hat{\theta}_k-\theta_k)(\hat{\theta}_l-\theta_l)\, l'''_{jkl}(\theta,X)$$

$$+ \frac{1}{12}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p}\sum_{m=1}^{p}(\hat{\theta}_j-\theta_j)(\hat{\theta}_k-\theta_k)(\hat{\theta}_l-\theta_l)(\hat{\theta}_m-\theta_m)\, l''''_{jklm}(\theta,X) + 0_p(n^{-3/2}). \qquad (A.1)$$

Taylor expansion of $l'_j(\hat{\theta},X)$ about θ gives:

$$
\begin{aligned}
0 &= \ell'_j(\hat\theta, X) \\
&= \ell'_j(\theta, X) + \sum_{k=1}^{p} (\hat\theta_k - \theta_k)\, \ell''_{jk}(\theta, X) \\
&+ \tfrac{1}{2} \sum_{k=1}^{p} \sum_{l=1}^{p} (\hat\theta_k - \theta_k)(\hat\theta_l - \theta_l)\, \ell'''_{jkl}(\theta, X) \\
&+ \tfrac{1}{6} \sum_{k=1}^{p} \sum_{l=1}^{p} \sum_{m=1}^{p} (\hat\theta_k - \theta_k)(\hat\theta_l - \theta_l)(\hat\theta_m - \theta_m)\, \ell''''_{jklm}(\theta, X) \\
&+ 0_p(n^{-1}). \qquad \text{(A.2)}
\end{aligned}
$$

Substituting (A.2) into the first term of the right hand side of (A.1), gives

$$
\begin{aligned}
W(\theta, X) &= \sum_{j=1}^{p} \sum_{k=1}^{p} (\hat\theta_j - \theta_j)(\hat\theta_k - \theta_k)(- \ell''_{jk}\ (\theta, X)) \\
&- \tfrac{2}{3} \sum_{j=1}^{p} \sum_{k=1}^{p} \sum_{l=1}^{p} (\hat\theta_j - \theta_j)(\hat\theta_k - \theta_k)(\hat\theta_l - \theta_l)\ell'''_{jkl}\ (\theta, X) \\
&- \tfrac{1}{4} \sum_{j=1}^{p} \sum_{k=1}^{p} \sum_{l=1}^{p} \sum_{m=1}^{p} (\hat\theta_j - \theta_j)(\hat\theta_k - \theta_k)(\hat\theta_l - \theta_l)(\hat\theta_m - \theta_m)\ell''''_{jklm}(\theta, X) \\
&+ 0_p(n^{-3/2}). \qquad \text{(A.3)}
\end{aligned}
$$

$$
\equiv S_1 + S_2 + S_3 + 0_p(n^{-3/2}).
$$

Furthermore,

$$
\begin{aligned}
S_1 &= \sum_{j=1}^{p} \sum_{k=1}^{p} n^{1/2}(\hat\theta_j - \theta_j) n^{1/2}(\hat\theta_k - \theta_k) n^{-1} E[-\ell''_{jk}\ (\theta, X)] \\
&+ \sum_{j=1}^{p} \sum_{k=1}^{p} n^{1/2}(\hat\theta_j - \theta_j) n^{1/2}(\hat\theta_k - \theta_k)\{n^{-1/2}(-\ell''_{jk}(\theta, X) - E[-\ell''_{jk}(\theta, X)])\} n^{-1/2}.
\end{aligned}
$$

The first term of S_1 is a quadratic form in $(\hat\theta - \theta)$ and can be written as $\sum_{j=1}^{p} (\alpha_j)^2$, where α is a p-dimensional vector with limiting normal distribution $N(0, I_p)$, where I_p is the identity matrix of dimension p. The variances are all unity because the first term is standardized by the Fisher information. The second term is a function of limiting normals times $n^{-1/2}$ which also can be treated as a linear transformation of α_j's and β_{jk}'s with limiting joint standard normal distribution, times $n^{-1/2}$. β_{jk} comes from the linear mapping of $\{n^{-1/2}(-\ell''_{jk}(\theta,X) - E[-\ell''_{jk}(\theta,X)])\}$ to the limiting standard normal.

We apply the same method to S_2 and S_3 and get,

$$S_2 = -\tfrac{2}{3}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p} n^{3/2}(\hat{\theta}_j - \theta_j)(\hat{\theta}_k - \theta_k)(\hat{\theta}_l - \theta_l)\{n^{-1}E\,\ell'''_{jkl}(\theta,X)\}\; n^{-1/2}$$

$$-\tfrac{2}{3}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p} n^{3/2}(\hat{\theta}_j-\theta_j)(\hat{\theta}_k-\theta_k)(\hat{\theta}_l-\theta_l)\{n^{-1/2}(-\ell'''_{jkl}(\theta,X)-E[-\ell'''_{jkl}(\theta,X)])\}n^{-1}.$$

$$S_3 = -\tfrac{1}{4}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p}\sum_{m=1}^{p} n^{2}(\hat{\theta}_j - \theta_j)(\hat{\theta}_k - \theta_k)(\hat{\theta}_l - \theta_l)(\hat{\theta}_m - \theta_m)\{n^{-1}E\,\ell''''_{jklm}(\theta,X)\}\; n^{-1}$$

$$-\tfrac{1}{4}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p}\sum_{m=1}^{p} n^{2}(\hat{\theta}_j-\theta_j)(\hat{\theta}_k-\theta_k)(\hat{\theta}_l-\theta_l)(\hat{\theta}_m-\theta_m)$$

$$\{n^{-1/2}(-\ell''''_{jklm}(\theta,X) - E[-\ell''''_{jklm}(\theta,X)])\}n^{-3/2} \qquad (A.4)$$

The second term of S_3 is of the order $0_p(n^{-3/2})$ and therefore can be collapsed into the remainder which is $0_p(n^{-3/2})$. We can express the terms in S_1, S_2 and S_3 as a polynomial in the α's and β's. We get:

$$W(\theta, X) = \sum_{j=1}^{p}(\alpha_j)^2$$

$$+ n^{-1/2}\Big(\sum_{j=1}^{p}\sum_{k=1}^{p}\alpha_j\alpha_k\beta_{jk}C_{1jk} - \tfrac{2}{3}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p}\alpha_j\alpha_k\alpha_l C_{1jkl}\Big)$$

$$+ n^{-1}\Big(-\tfrac{2}{3}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p}\alpha_j\alpha_k\alpha_l\beta_{jkl}C_{2jkl}$$

$$- \tfrac{1}{4}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p}\sum_{m=1}^{p}\alpha_j\alpha_k\alpha_l\alpha_m C_{2jklm})$$

$$+ 0_p(n^{-3/2}). \tag{A.5}$$

The C_{1jk}, C_{2jkl}, C_{1jkl} and C_{2jklm} are constants representing the linear transformation from α's and β's to the limiting normals in S_1, S_2 and S_3. Notice that they do not depend on n.

Now assume α's and β's have a joint density function f which admits an Edgeworth expansion of the form:

$$f(z) = \phi(z) + \sum_{j=1}^{2} n^{-\frac{1}{2}j} p_j(z)\phi(z) + O(n^{-3/2}) \tag{A.6}$$

where $\phi(z)$ is the N(0, I) density.

By the results of Bhattacharya(1985), the p_j's with j odd vanish. Then

$$\Pr(\, W(\theta,X) \le x \,) = \int_{\Re(x;n)} \{\, \phi(z) + n^{-1}p_2(z)\phi(z) \,\}\, dz \quad + O(n^{-3/2}) \tag{A.7}$$

where

$$\Re(x;n) \equiv \quad \{\, z: \sum_{j=1}^{p}(\alpha_j)^2$$

$$+ n^{-1/2}(\sum_{j=1}^{p}\sum_{k=1}^{p}\alpha_j\alpha_k\beta_{jk}C_{1jkl} - \tfrac{2}{3}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p}\alpha_j\alpha_k\alpha_l C_{1jkl})$$

$$+ n^{-1}(- \tfrac{2}{3}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p}\alpha_j\alpha_k\alpha_l\beta_{jkl}C_{2jklm}$$

$$- \tfrac{1}{4}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p}\sum_{m=1}^{p}\alpha_j\alpha_k\alpha_l\alpha_m C_{2jklm}) \;\le\; x \;\} \tag{A.8}$$

The remainder term, after introducing the integral, is still $O(n^{-3/2})$ since $\Re(x;n)$ is a region with finite area and the integrand is also finite and thus the order of the error is preserved.

By the same Taylor expansion on $W(\hat{\theta},X^*(\hat{\theta}))$ and denoting $\hat{\theta}(X^*(\hat{\theta}))$ as $\hat{\theta}^*$ for simplicity, we get

$$W(\hat{\theta},X^*(\hat{\theta})) = \sum_{j=1}^{p}\sum_{k=1}^{p}(\hat{\theta}_j^* - \hat{\theta}_j)(\hat{\theta}_k^* - \hat{\theta}_k)(- \ell''_{jk}\ (\hat{\theta},X))$$

$$- \tfrac{2}{3}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p}(\hat{\theta}_j^*-\hat{\theta}_j)(\hat{\theta}_k^*-\hat{\theta}_k)(\hat{\theta}_l^*-\hat{\theta}_l)\ell'''_{jkl}(\hat{\theta},X)$$

$$- \tfrac{1}{4}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p}\sum_{m=1}^{p}(\hat{\theta}_j^*-\hat{\theta}_j)(\hat{\theta}_k^*-\hat{\theta}_k)(\hat{\theta}_l^*-\hat{\theta}_l)(\hat{\theta}_m^*-\hat{\theta}_m)\cdot l''''_{jklm}(\hat{\theta},X)$$

$$+ 0_p(n^{-3/2}). \tag{A.9}$$

$$= \sum_{j=1}^{p}(\alpha_j^*)^2$$

$$+ n^{-1/2}\left(\sum_{j=1}^{p}\sum_{k=1}^{p}\alpha_j^*\alpha_k^*\beta_{jk}^*C_{1jk} - \tfrac{2}{3}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p}\alpha_j^*\alpha_k^*\alpha_l^*C_{1jkl}\right)$$

$$+ n^{-1}\Big(- \tfrac{2}{3}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p}\alpha_j^*\alpha_k^*\alpha_l^*\beta_{jkl}^*C_{2jkl}$$

$$- \tfrac{1}{4}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p}\sum_{m=1}^{p}\alpha_j^*\alpha_k^*\alpha_l^*\alpha_m^*C_{2jklm}\Big)$$

$$+ 0_p(n^{-3/2}). \tag{A.10}$$

The constants C_{1jk}, C_{2jkl}, C_{1jkl} and C_{2jklm} are the same as in (A.5) since the same linear transformation can be used from α^* to the limiting normals corresponding to the terms in (A.4) except with $\hat{\theta}$ replacing θ and $\hat{\theta}^*$ replacing $\hat{\theta}$.

Similarly,

$$\Pr(\ W(\hat{\theta},X^*(\hat{\theta})) \ \le\ x\ ) \ =\ \int_{\Re(x;\ n)}\ \{\ \phi(z)\ +\ n^{-1}p_2^*(z)\phi(z)\ \}\ dz\ +\ O(n^{-3/2})\ , \tag{A.11}$$

where $\Re(x;\ n)$ is the same as in (A.8) since the same mapping was applied. $p_2(z)$ is a polynomial whose coefficients are the first and second cumulants and Hermite polynomials $H_4(z)$ and $H_6(z)$ which are defined by

$$\partial^r(e^{-\frac{1}{2}z^2}) \,/\, \partial x^r = (-1)^r \, H_r(z) \, e^{-\frac{1}{2}z^2}$$

$p_2^*(z)$ is the same polynomial except with the cumulants replaced by their bootstrap estimates. $p_2(z) - p_2^*(z) = O_p(n^{-1/2})$ since they are smooth functions of sample moments and population moments which have differences of the order $O_p(n^{-1/2})$ as long as the fourth moment is finite. (Serfling, 1981, p.68). Therefore, again by noticing that $\Re(x;\, n)$ and the integrand are finite, we have

$$\Pr(W(\hat\theta, X^*(\hat\theta)) \le x) - \Pr(W(\theta, X) \le x)$$

$$= \int_{\Re(x;\, n)} n^{-1} \left(p_2^*(z) - p_2(z) \right) dz + 0(n^{-3/2})$$

$$= O(n^{-3/2}). \tag{A.12}$$

This proves (2.8).

To prove (2.9) and (2.10), since direct inversion of bootstrap likelihood ratio confidence region $\Re(\theta)$ has level error $O(n^{-3/2})$, it is sufficient to prove (2.9) and then (2.10) follows. To show that $\Pr(W(\hat\theta, X^*(\hat\theta)) \le x) - \Pr(W(\theta_\alpha, X^*(\theta_\alpha)) \le x) = O(n^{-3/2})$, where θ_α is any $\theta \in \Re_\alpha$ constructed from (2.6), similar expansions of $W(\hat\theta, X^*(\hat\theta))$ and $W(\theta_\alpha, X^*(\theta_\alpha))$ as before give:

$$\Pr(W(\hat\theta, X^*(\hat\theta)) \le x) - \Pr(W(\theta_\alpha, X^*(\theta_\alpha)) \le x)$$

$$= \int_{\Re(x;\, n)} n^{-1} \left(p_2^*(z, \hat\theta) - p_2^*(z, \theta_\alpha) \right) dz + O(n^{-3/2}). \tag{A.13}$$

When the Edgeworth expansion exists, the p_2^*'s are bounded functions of order $O_p(1)$ uniformly in θ. Since they are polynomials, they are continuous in θ. When θ's are restricted in $\Re_\alpha$, notice that $\Re_\alpha$ shrinks in each of its linear coordinates as $O_p(n^{-1/2})$, therefore, $|\theta_\alpha - \hat\theta| = O_p(n^{-1/2})$ uniformly in $\Re_\alpha$ and $p_2^*(z, \hat\theta) - p_2^*(z, \theta_\alpha) = O_p(n^{-1/2})$ uniformly in $\Re_\alpha$. This indicates that (A.13) is of order $0(n^{-3/2})$□

Bibliography

Barndorff-Nielsen, O.E. and Hall, P (1988). On the level-error after Bartlett adjustment of the likelihood ratio statistic. *Biometrika,* 75: 374-378.

Beran, R. (1988). Prepivoting test statistics: a bootstrap view of asymptotic refinements. *J. Amer. Statist. Assoc.* 83: 687-697.

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

Bhattacharya, R.N. (1985). Some recent results on Cramer-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

Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics.* Chapman and Hall, London. 511 pp.

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

Feng, Z. and McCulloch, C.E. (1990). Statistical inference using maximum likelihood estimation and the generalized likelihood ratio when the true parameter is on the boundary of the parameter space. BU-1109-M. December, 1990. Biometrics Unit, Cornell University.

Hall, P. (1986). On the bootstrap and confidence intervals. *Ann. Statist. 86: 1431-1452.*

Hall, P. (1987). On the bootstrap and likelihood-based confidence regions. *Biometrika.* 74: 481-493.

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 the mixture model. *Proceedings of the Berkeley Symposium in honor of Jerzy Neyman and Jack Kiefer,* Vol. II, L. LeCam and R.A. Olshen (eds.), Wadsworth, New York, pp. 807-810.

Lehmann, E.L. (1983). *Theory of Point Estimation.* John Wiley and Sons, New York. 506 pp.

McCulloch, C.E. (1987). Maximum likelihood estimation in a multinomial mixture model. No. BU-934-MA in the Biometrics Unit, Technical Report Series, Cornell University.

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.

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

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.

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. San Diego: U.S. Naval Personnel and Training Research Laboratory.

Table 1 Coverage probabilities of the bootstrap likelihood ratio and the likelihood ratio based confidence regions from 500 simulations on mixture multinomials with two components and with the parameter on the boundary of the parameter space

(nominal coverage probability = 0.9, 500 replications)

Estimated coverage probability (standard error)

Sample Size	Boot-LR	LR
10	0.968 (0.008)	0.884 (0.014)
25	0.902 (0.013)	0.890 (0.014)
50	0.908 (0.013)	0.866 (0.015)
100	0.886 (0.014)	0.884 (0.014)
$1 - \alpha$	0.900	0.900

Table 2. Coverage probabilities of the bootstrap likelihood ratio and the likelihood ratio based confidence regions from 500 simulations on mixture multinomials with three components and the parameter is near the boundary of the parameter space

(nominal coverage probability = 0.9, 500 replications)

Estimated coverage probability (standard error)

Sample Size	Boot-LR	LR
10	0.884 (0.014)	0.802 (0.018)
25	0.928 (0.012)	0.894 (0.014)
50	0.906 (0.013)	0.906 (0.013)
100	0.910 (0.013)	0.906 (0.013)
$1 - \alpha$	0.900	0.900

Table 3. Probabilities of acceptance of H_0 by the bootstrap and the likelihood ratio test procedures in a test of H_0: N(0,1) vs. H_1: π N(0,1) + (1-π) N(1,1) when H_0 is true. The likelihood ratio statistic is based on the unrestricted maxima (500 replications).

Estimated Coverage Probability (standard error)

Nominal $1-\alpha$	Sample Size	Boot-LR	LR
	10	0.882 (0.014)	0.720 (0.020)
.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)
.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)
.99	30	0.998 (0.002)	0.956 (0.009)
	100	0.990 (0.004)	0.982 (0.006)

Table 4. Coverage probabilities of the bootstrap and the likelihood ratio based confidence intervals for a mixture normal model: π N(0,1) + (1-π) N(1,1) when the true distribution is N(0,1). The likelihood ratio statistic is based on the unrestricted maxima (500 replications.)

Estimated coverage probability (standard error)

Nominal $1-\alpha$	Sample Size	Boot-LR	LR
	10	0.912 (0.013)	0.728 (0.020)
.90	30	0.924 (0.012)	0.840 (0.016)
	100	0.898 (0.014)	0.874 (0.015)
	10	0.944 (0.013)	0.796 (0.018)
.95	30	0.968 (0.008)	0.910 (0.013)
	100	0.946 (0.010)	0.946 (0.010)
	10	0.952 (0.010)	0.858 (0.016)
.99	30	0.992 (0.002)	0.952 (0.010)
	100	0.984 (0.006)	0.976 (0.007)

Table 5. Coverage probabilities of the bootstrap and the likelihood ratio based confidence regions for a mixture normal model: (1-π) N(0,1) + π N(μ,1) when the true distribution is N(0,1). The likelihood ratio statistic is based on the maximum likelihood estimator (500 replications).

Simulated coverage probability (standard error)

Nominal $1-\alpha$	Sample Size	Boot-LR	LR
	10	0.916 (0.012)	0.878 (0.015)
.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)
.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)
.99	30	0.984 (0.006)	0.970 (0.008)
	100	0.996 (0.003)	0.992 (0.004)

Table 6. Coverage probabilities of the bootstrap and the likelihood ratio based confidence regions for a mixture normal model: (1-π) N(0,1) + π N(μ,1) when the true distribution is N(0,1). The likelihood ratio statistic is based on the unrestricted maxima (500 replications).

Simulated coverage probability (standard error)

Nominal $1-\alpha$	Sample size	Boot-LR	LR(χ^2_1)	LR(χ^2_2)
	10	0.902(0.013)	0.788(0.018)	0.896(0.014)
.90	30	0.932(0.011)	0.828(0.017)	0.942(0.010)
	100	0.896(0.014)	0.830(0.017)	0.932(0.011)
	10	0.952(0.010)	0.866(0.015)	0.940(0.011)
.95	30	0.968(0.008)	0.918(0.012)	0.974(0.007)
	100	0.942(0.010)	0.894(0.014)	0.964(0.008)
	10	0.998(0.002)	0.948(0.010)	0.980(0.006)
.99	30	0.992(0.004)	0.978(0.007)	0.994(0.003)
	100	0.992(0.004)	0.978(0.007)	0.996(0.003)