A MODIFIED LIKELIHOOD RATIO TEST FOR HOMOGENEITY IN FINITE MIXTURE MODELS

Hanfeng Chen, Jiahua Chen and John D. Kalbfleisch

Bowling Green State University and University of Waterloo

Abstract

Testing for homogeneity in finite mixture models has been investigated by many authors. The asymptotic null distribution of the likelihood ratio test (LRT) is very complex and difficult to use in practice. In this paper we propose a modified LRT for homogeneity in finite mixture models with a general parametric kernel distribution family. The modified LRT has a χ^2-type null limiting distribution and is asymptotically most powerful under local alternatives. Simulations show that it performs better than competing tests. They also reveal that the limiting distribution with some adjustment can satisfactorily approximate the quantiles of the test statistic, even for moderate sample sizes.

Key words and phrases. Asymptotic distribution, finite mixture, irregular model, strong identifiability.

AMS 1980 *subject classifications. Primary 62F03; secondary 62F05.*

1 Introduction

Finite mixture models are often used to help determine whether data come from a homogeneous or heterogeneous population. Let $\{f(x, \theta) : \theta \in \Theta\}$ be a parametric family of probability density functions (pdf). We observe a random sample $X_1, \ldots, X_n$ from the mixture pdf

$$(1-\gamma)f(x, \theta_1) + \gamma f(x, \theta_2), \tag{1}$$

where $\theta_1 \leq \theta_2 \in \Theta$ and $0 \leq \gamma \leq 1$. We wish to test the hypothesis

$$H_0 : \theta_1 = \theta_2, \text{ (or equivalently } \gamma = 0, \text{ or } \gamma = 1),$$

that is to test whether the observations come from a homogeneous population $f(x, \theta)$. The pdf $f(x, \theta)$ is called the *kernel function.* The mixture pdf (**??**) can also be written as

$\int f(x,\theta)dG(\theta)$, where

$$G(\theta) = (1-\gamma)I(\theta_1 \le \theta) + \gamma I(\theta_2 \le \theta) \tag{2}$$

is called the mixing distribution.

Likelihood-based methods play a central role in parametric testing problems, and among these, the likelihood ratio test (LRT) is often preferred. The LRT has a simple interpretation, is invariant under re-parameterization and, under standard regularity conditions, has a simple and elegant asymptotic theory (Wilks, 1938). However, the regularity conditions required are not satisfied in the mixture problem. Chernoff and Lander (1995), Ghosh and Sen (1985) and some other authors find that the asymptotic distribution of the LRT involves the supremum of a Gaussian process. Chen and Chen (1998a,b) show that when $f(x,\theta_0)$ is the true null distribution, the asymptotic distribution of the LRT for homogeneity is that of $\{\sup_{\theta\in\Theta} W^+(\theta)\}^2$, where $W(\theta)$ is a Gaussian process with mean 0, variance 1 and autocorrelation function

$$\rho(\theta,\theta') = \frac{cov\{Z_i(\theta) - h(\theta)Y_i(\theta_0), Z_i(\theta') - h(\theta')Y_i(\theta_0)\}}{\sqrt{var\{Z_i(\theta) - h(\theta)Y_i(\theta_0)\}var\{Z_i(\theta') - h(\theta')Y_i(\theta_0)\}}}. \tag{3}$$

Here

$$Y_i(\theta) = Y_i(\theta,\theta_0) = \frac{f(X_i,\theta) - f(X_i,\theta_0)}{(\theta-\theta_0)f(X_i,\theta_0)},\ \theta \ne \theta_0;\quad Y_i(\theta_0) = Y_i(\theta_0,\theta_0) = \frac{f'(X_i,\theta_0)}{f(X_i,\theta_0)}.$$

$$Z_i(\theta) = Z_i(\theta,\theta_0) = \frac{Y_i(\theta) - Y_i(\theta_0)}{\theta-\theta_0},\ \theta \ne \theta_0;\quad Z_i(\theta_0) = Z_i(\theta_0,\theta_0) = \frac{dY_i(\theta,\theta_0)}{d\theta}\Big|_{\theta=\theta_0},$$

and $h(\theta) = EY_i(\theta_0)Z_i(\theta)/EY_i^2(\theta_0)$.

This result illustrates that, as a consequence of the breakdown of the usual regularity conditions, the LRT has unusual asymptotic properties. The main difficulties in applying these asymptotic results are that:

1. The asymptotic distribution under H_0 depends on the true (unknown) value θ_0 of θ.

2. The asymptotic null distribution depends on the parametric family. For instance, the autocorrelation function (**??**) differs for normal and Poisson kernels.

3. From a more practical point of view, the asymptotic LRT loses its appeal since the supremum of a Gaussian process is very complicated, compared to the usual χ^2 distribution. As a result, simulation-based tests (see e.g., McLachlan 1987 and Schork 1992) that circumvent the asymptotic theory are often used.

In this paper we propose a modified LRT which retains the power of the LRT and enjoys an elegant asymptotic theory as well.

The following is a brief review of some key references. The books by Titterington, Smith and Makov (1985), McLachlan and Basford (1988) and Lindsay (1995) provide extensive discussion about the background of finite mixture models. Hartigan (1985) gave an important insight into the irregularity problems of finite mixture models. He proved that the LRT statistic becomes unbounded in probability when Θ is unbounded and the sample size is large. Neyman and Scott (1966) studied the use of $C(\alpha)$ tests against the alternative of any mixture. Lindsay(1989) proposed moment-based testing procedures. Cheng and Traylor (1995) discussed irregular models which include finite mixture models. Ghosh and Sen (1985), Bickel and Chernoff (1993), Chernoff and Lander (1995), Lemdani and Pons (1999), Dacunha-Castelle and Gassiat (1999) and Chen and Chen (1998a) studied the asymptotic distributions of the LRT in mixture models. This paper extends work in Chen (1998) on multinomial mixture models.

The paper is organized as follows. Section 2 describes the modified LRT and presents its asymptotic theory. It is shown that the asymptotic null distribution is a mixture of central χ^2_1 and χ^2_0 with equal weights and that the modified LRT is asymptotically most powerful under local alternatives. To study the testing power for small or moderate sample sizes, three competing methods, the $C(\alpha)$ test, a bootstrap test and the method of Davies (1977, 1987), are considered. Section 3 gives a brief summary of the three competing methods and Section 4 presents simulation results. It is demonstrated that the modified LRT is slightly better than $C(\alpha)$ when an alternative model is close to the null model (i.e., the Kullback-Leibler information is small), but becomes much more powerful than $C(\alpha)$ as the Kullback-Leibler information increases. The modified LRT performs better, both in terms of size and power, than the bootstrap method discussed in McLachlan (1987). Davies' test for the current problem is also shown to be less appealing than the modified LRT. The simulations also reveal that the limiting distribution satisfactorily approximates the quantiles of the test statistic even for moderate sample sizes.

2 The modified LRT

Complications of the asymptotic null distribution of the ordinary LRT have two sources: (a) the null hypothesis lies on the boundary of the parameter space ($\gamma = 0$) and (b) the parameters γ, θ_1 and θ_2 are not identifiable under the null model. For example, the two statements $\gamma = 0$ and $\theta_1 = \theta_2$ are equivalent. The complications are expected to disappear if we can overcome the boundary problem and the non-identifiability. The test based on the following modified log-likelihood function $l_n(\gamma, \theta_1, \theta_2)$ provides a satisfactory solution. For $0 < \gamma < 1$, $\theta_1, \theta_2 \in \Theta$ with $\theta_1 \leq \theta_2$, define

$$l_n(\gamma, \theta_1, \theta_2) = \sum_{i=1}^{n} \log\{(1-\gamma)f(X_i, \theta_1) + \gamma f(X_i, \theta_2)\} + C \log\{4\gamma(1-\gamma)\}, \tag{4}$$

where $C > 0$ is constant and used to control the level of modification. Even though $\gamma = 0$ and 1 are not in the domain of $l_n(\gamma, \theta_1, \theta_2)$, the corresponding distributions are not excluded since they are covered by $\theta_1 = \theta_2$. The modified likelihood function $l_n(\gamma, \theta, \theta)$ is often called a penalized likelihood function, referring to the penalty when γ is close to 0 or 1.

The Associate Editor has noted that the modified likelihood function can be motivated by a Bayesian procedure or by incorporating a conceptual auxiliary experiment. In the Bayesian motivation, let $(\gamma, \theta_1, \theta_2)$ have prior density proportional to $\{\gamma(1-\gamma)\}^C$ so that $\exp\{l_n(\gamma, \theta_1, \theta_2)\}$ is proportional to the posterior density. Alternatively, one can think of (**??**) as the likelihood arising from the mixture experiment along with an auxiliary experiment. In the auxiliary experiment, an additional $2C$ observations are taken and we observe that exactly C arise from the smaller and C from the larger value of θ. This interpretation enables the construction of a simple EM algorithm for maximizing (**??**). Specifically, the E-step results in the imputed log-likelihood

$$\sum_{i=1}^{n}[w_{1i} \log f(x_i, \theta_1) + w_{2i} \log f(x_i, \theta_2)] + (\sum_{i=1}^{n} w_{1i} + C) \log\{2(1-\gamma)\} + (\sum_{i=1}^{n} w_{2i} + C) \log(2\gamma)$$

where

$$w_{1i} = \frac{(1-\gamma)f(x_i, \theta_1)}{(1-\gamma)f(x_i, \theta_1) + \gamma f(x_i, \theta_2)}$$

and $w_{2i} = 1 - w_{1i}$. For the case of a normal kernel with known variance, this leads to the self-consistency equations

$$\hat{\theta}_j = \sum_{i=1}^{n} \hat{w}_{ji} x_i / \sum_{i=1}^{n} \hat{w}_{ji}, \quad j = 1, 2$$

and $\hat{\gamma} = (\sum_{i=1}^{n} \hat{w}_{2i} + C)/(n + 2C)$. Under the null hypothesis, $\theta_1 = \theta_2 = \theta$ and $\hat{\theta}$ is the maximum likelihood estimator from the kernel model. In this case, $\hat{\gamma} = 1/2$, so the penalty term is zero. The penalty term affects only the maximized likelihood under the alternative.

Let $(\hat{\gamma}, \hat{\theta}_1, \hat{\theta}_2)$ maximize $l_n(\gamma, \theta_1, \theta_2)$ over the full parameter space, and let $\hat{\theta}$ maximize the null modified likelihood function $l_n(1/2, \theta, \theta)$, $\theta \in \Theta$. The modified LRT rejects the null hypothesis H_0 for large values of

$$M_n = 2\{l_n(\hat{\gamma}, \hat{\theta}_1, \hat{\theta}_2) - l_n(1/2, \hat{\theta}, \hat{\theta})\}. \tag{5}$$

Note that the additional term $C \log\{4\gamma(1-\gamma)\}$ in (**??**) is non-positive. Using methods similar to Chen (1998), we can show that the estimate $\hat{\gamma}$ satisfies $0 < \hat{\gamma} < 1$, and $\hat{\theta}_1$ and $\hat{\theta}_2$ are consistent under the null hypothesis. Regularity conditions 1-5, on the kernel distribution $f(x, \theta)$ are given with some discussion in the Appendix. We have the following asymptotic result.

Theorem 1 *If Conditions 1-5 hold, the asymptotic null distribution of the modified LRT statistic M_n is the mixture of χ_1^2 and χ_0^2 with equal weights, that is*

$$\frac{1}{2}\chi_1^2 + \frac{1}{2}\chi_0^2, \tag{6}$$

where χ_0^2 is a degenerate distribution with all its mass at 0.

An outline of the proof of Theorem **??** is given in the Appendix. For a detailed proof, see Chen, Chen and Kalbfleisch (2000).

There are two issues related to the applications of Theorem **??**. The first concerns the choice of C in the modified likelihood function (**??**). The second concerns possible improvements on the approximation $0.5\chi_1^2 + 0.5\chi_0^2$ when the sample size is small or moderate.

Choice of C. Consider the finite normal mixture model and assume that Θ is unbounded. Bickel and Chernoff (1993) show that when the sample size is large, the maximum likelihood estimator of the location parameter does not often take values beyond $\{\log(n)/2\}^{1/2}$ and consequently, the rate of divergence of the LRT is $\log\log n$. This rate would also hold for a wide range of kernel functions. From this result, an appropriate choice of C is $C = \log M$, when the parameter θ in the kernel density is restricted to $[-M, M]$. From a number of simulation studies, we found that the method is not sensitive to the values of C and the choice of $C = \log M$ works well.

*Adjustment to (**??**).* Since the proportion $1/2$ of χ_0^2 in (**??**) is the limit of the probability $p_n = P_{H_0}(M_n \leq 0)$, $(1-p_n)\chi_1^2 + p_n\chi_0^2$ should provide a more accurate approximation. In most circumstances, p_n can be easily evaluated. For example, if $f(x,\theta)$ is the normal density, $M_n \leq 0$ if and only if the sample variance is smaller than $n\sigma_0^2/(n-1)$, implying that $p_n = 0.547$ when $n = 100$ and $p_n = 0.533$ when $n = 200$.

We next consider the distribution of M_n under a local alternative: for any $0 < \gamma_0 < 1$ and $\tau > 0$, let

$$H_a^n: \quad \gamma = \gamma_0, \quad \theta_1 = \theta_0 - n^{-1/4}\tau_1 \quad \theta_2 = \theta_0 + n^{-1/4}\tau_2,$$

where $\tau_1 = \tau\{\gamma_0/(1-\gamma_0)\}^{1/2}$ and $\tau_2 = \tau\{(1-\gamma_0)/\gamma_0\}^{1/2}$. The local alternative is contiguous to the null distribution $f(x,\theta_0)$; see LeCam and Yang (1990) or Bickel, et al. (1993, page 17). From LeCam's contiguity theory, the limiting distribution of M_n under H_a^n can be determined. A sketch of the proof of the following theorem is given in the Appendix.

Theorem 2 *Under the alternatives H_a^n, the limiting distribution of M_n is that of $\{(Z+\sigma_{12})^+\}^2$, where Z is a standard normal random variable and $\sigma_{12} = \tau^2\{EW_1^2\}^{1/2}$ where $W_1 = Z_1(\theta_0) - h(\theta_0)Y_1(\theta_0)$ and the expectation is with respect to the pdf $f(x,\theta_0)$.*

It can be shown that $\sigma_{12} = \tau^2\{EW_1^2\}^{1/2} > 0$ and that the modified LRT is locally unbiased. Furthermore, it is clear that the ordinary LRT for testing $f(x,\theta)$ versus H_a^n has the same limiting distribution as the modified LRT. Thus, the modified LRT is locally (asymptotically) optimal for this class of alternatives.

To conclude this section, we remark that the modified LRT approach can be extended to the finite mixture model with k mixture components: $\gamma_1 f(x,\theta_1) + \cdots + \gamma_k f(x,\theta_k)$ where $\theta_1 \leq \cdots \leq \theta_k$ and $\sum \gamma_i = 1$. In this case, the penalty term is $C\sum_{i=1}^k \log(2\gamma_i)$. The asymptotic null distribution of this new modified LRT would be (**??**), the same as before. The power of the test would depend mainly on the Kullback-Leibler information associated with the alternative hypothesis. If the true mixture model has three components say, the modified LRT has lower power than that for a two-component alternative (with the same Kullback-Leibler information). This difference in power is due to the larger penalty in the three-component case, but this effect is negligible, compared to the effect of the Kullback-Leibler information.

3 Competing methods

To study the power of the modified LRT in samples of moderate size, consider three competing methods: Neyman and Scott's $C(\alpha)$ test; a bootstrap test; and Davies' test.

Neyman and Scott's $C(\alpha)$ test is designed to test homogeneity against general mixture alternatives (Neyman and Scott, 1966, and Lindsay, 1995). The $C(\alpha)$ test reduces to a test for over-dispersion when the kernel distribution belongs to a regular exponential family. If the kernel distribution is normal with variance σ^2, the $C(\alpha)$ test is based on the ratio of the sample variance to σ^2. If the kernel distribution is Poisson, the test is based on the ratio of the sample variance to the sample mean. It is known that the $C(\alpha)$ test is also locally most powerful.

McLachlan (1987) suggested a bootstrap method to evaluate the p-value of the LRT for normal mixture models. This method involves generating a parametric bootstrap sample from the null model and comparing the observed LRT statistics with the 95th percentile of the bootstrap distribution.

Davies (1977, 1987) proposed a method which can be applied to the test of mixture models when one of the mixture components is completely known. For example, the method is applicable to $H_0 : \theta_1 = \theta_2 = 0$ versus $H_1 : \theta_1 = 0, \theta_2 \neq 0$ and $\gamma > 0$. In this case, the likelihood function (with θ replacing θ_2) is given by

$$l_n(\theta, \gamma) = \sum_{i=1}^{n} \log[1 + \gamma\{\exp(\theta X_i - \frac{1}{2}\theta^2) - 1\}].$$

Davies suggested using the test statistic $\sup_{\theta \in \Theta} Z_n(\theta)$ where $Z_n(\theta)$ is the standardized score at $\gamma = 0$ with θ treated as known and defined as follows:

$$Z_n(\theta) = [n\{\exp(\theta^2) - 1\}]^{-1/2} \sum_{i=1}^{n} \{\exp(\theta X_i - \frac{1}{2}\theta^2) - 1\}.$$

It can be seen that $Z_n(\theta)$ converges to a Gaussian process, say $S(\theta)$. Since the quantiles of $\sup_{\theta \in \Theta} S(\theta)$ are intractable, Davies (1987) suggested approximating the p-value with

$$\Phi(-Q) + V \exp(-\frac{1}{2}Q^2)/\sqrt{8\pi},$$

where $Q = \sup_{\theta \in \Theta} Z_n(\theta)$ and $V = \int_{\Theta} |S'(\theta)| d\theta$. The value of V can be approximated with the total variation of $Z_n(\theta)$.

The modified LRT to the models with one component completely known can be applied by defining $M_n = 2[\sup_{\Theta} l_n(\theta, \gamma) + C \log(2\gamma)]$. It can be seen that $M_n \to \chi_1^2$ in distribution.

4 Simulation

A simulation experiment was conducted with normal and Poisson kernels. The null distribution in the normal case is $N(0,1)$ and in the Poisson case, the null distribution has a mean of 5. Eight alternative models were selected as follows: Four from a normal mixture, each of which has a mixing distribution with mean $(1-\gamma)\theta_1 + \gamma\theta_2 = 0$ and variance $(1-\gamma)\theta_1^2 + \gamma\theta_2^2 = 1/4$; and four from a Poisson mixture, each of which has a mixing distribution with mean $(1-\gamma)\theta_1 + \gamma\theta_2 = 5$ and variance $(1-\gamma)(\theta_1 - 5)^2 + \gamma(\theta_2 - 5)^2 = 1$. The choice of the null mean has no impact on the power, but the variance does. The four normal and four Poisson mixtures are identified by $\gamma = 0.5, 0.75, 0.9$ and 0.95, respectively. These alternative models together with their Kullback-Leibler information are listed in Table 1.

For each simulation, except those involving Davies' method, three significance levels $10\%, 5\%$ and 1%, and two sample sizes, $n = 64$ and $n = 100$ were considered. As mentioned in Section 2, we chose $C = \log M$ in the modified LRT. In the normal mixture, with the pdf $(1-\gamma)\phi(x-\theta_1) + \gamma\phi(x-\theta_2)$, we assumed that $\theta_i \in [-10, 10]$, $i = 1, 2$ and hence $C = \log(10) = 2.303$. In the Poisson mixture, $(1-\gamma)\theta_1^x \exp\{-\theta_1\}/x! + \gamma\theta_2^x \exp\{-\theta_2\}/x!$, we assumed that $\theta_i \in [0, 50]$, $i = 1, 2$ and $C = \log(50) = 3.912$.

For the two mixture models considered, the $C(\alpha)$ test measures over-dispersion. Therefore, the test statistic is the sample variance S_n^2 for the normal mixture, and the ratio $S_n^2/\bar{X}_n$ of the sample variance to the sample mean for the Poisson mixture, respectively. The bootstrap method was applied to normal mixture models with 200 bootstrap samples and 2,000 repetitions.

Davies' method was applied to normal mixture models (with θ_1 assumed known) in a separate simulation.

To conduct a fair and meaningful power comparison, we should ensure comparable finite sample size significance levels between various methods. For each nominal significance level, the critical value and the null rejection level were obtained by using 10,000 Monte Carlo trials for the modified LRT and the $C(\alpha)$ test, and 2,000 Monte Carlo trials for the bootstrap method. The simulated results are quite close to the values given by the asymptotic theory. To save space, only null rejection rates are reported in Tables 2 and 3.

The simulation results for power comparison under normal mixture models are reported in Table 2. We see that when the Kullback-Leibler information is small, so that the alternatives are near the null, the modified LRT and the $C(\alpha)$ test have comparable power, but the bootstrap method does relatively poorly. When the Kullback-Leibler information is large, however, the modified LRT is clearly preferable to the $C(\alpha)$ test and comparable with the bootstrap test.

Simulations comparing powers of the modified LRT and the $C(\alpha)$ test for Poisson mixture models are reported in Table 3. The outcomes are similar to the corresponding results in Table 2. The modified LRT is superior to the $C(\alpha)$ test.

Comparisons between the modified LRT and Davies' method for the four normal mixture alternatives are given in Table 4. The nominal significance level is 5% and $n = 100$. The four sets of parameters used are $\theta = \gamma = 0, 0.1, 0.2, 0.3$. Davies' approximation to the p-value is found to be very precise. The simulation results show, however, that Davies' method is decidedly less powerful than the modified LRT.

APPENDIX

In the Appendix, we list the regularity conditions on the kernel function for the validity of the asymptotic results given in Section 2, and we outline the main ideas in the proofs of Theorems 1 and 2. For detailed proofs, see Chen, Chen and Kalbfleisch (2000).

A.1 Regularity conditions on the kernel function

Condition 1. *Wald's integrability conditions.* For each $\theta \in \Theta$, (i) $E|\log f(X,\theta)| < \infty$, and (ii) there exists $\rho > 0$ such that $E \log f(X,\theta,\rho) < \infty$, where $f(x,\theta,\rho) = 1 + \sup_{|\theta'-\theta|\le\rho}\{f(x,\theta')\}$.

Condition 2. *Smoothness.* The kernel function $f(x,\theta)$ has support independent of θ and is twice continuously differentiable with respect to θ. The first two derivatives are denoted by $f'(x,\theta)$ and $f''(x,\theta)$.

Condition 3. *Strong identifiability.* The kernel function $f(x,\theta)$ is strongly identifiable. That is,

(i) For any G_1 and G_2 such that

$$\int f(x,\theta)dG_1(\theta) = \int f(x,\theta)dG_2(\theta), \quad \text{for all} \;\; x,$$

we must have $G_1 = G_2$.

(ii) For any $\theta_1 \neq \theta_2$ in Θ,

$$\sum_{j=1}^{2}\{a_j f(x,\theta_j) + b_j f'(x,\theta_j) + c_j f''(x,\theta_j)\} = 0, \text{ for all } x,$$

implies that $a_j = b_j = c_j = 0$, $j = 1, 2$.

The condition (ii) above was first proposed by Chen (1995), who proved that location and scale kernels satisfy (ii) if $f(\pm\infty,\theta) = f'(\pm\infty,\theta) = 0$. Using the same argument, it can be shown that all regular exponential families are strongly identifiable.

Condition 4. *Condition for uniform strong law of large numbers.* There exists an integrable g and $\delta > 0$ such that $|Y_i(\theta)|^{4+\delta} \le g(X_i)$ and $|Y_i'(\theta)|^3 \le g(X_i)$ for all $\theta \in \Theta$.

Condition 5. *Tightness.* The processes $n^{1/2}\sum Y_i(\theta)$, $n^{1/2}\sum Y_i'(\theta)$ and $n^{1/2}\sum Y_i''(\theta)$ are tight.

A.2 Sketch of the proof of Theorem 1

A key step in the proof of Theorem 1 is to note that since $l_n(\gamma,\theta_1,\theta_2)$ is $O_p(1)$ uniformly, it follows that $\log\{4\hat\gamma(1-\hat\gamma)\} = O_p(1)$. This implies that $\hat\gamma$ is bounded away from 0 and 1. It then follows that both $\hat\theta_1$ and $\hat\theta_2$ must converge to the true value θ_0 in probability under the null model. This justifies the approximation of the modified likelihood ratio statistic by a quadratic form. The typical technique of quadratic approximation to the modified likelihood function is then applicable and yields the result desired.

A.3 Sketch of the proof of Theorem 2

Let

$$\Lambda_n = \sum_{i=1}^{n} \log \frac{(1-\gamma_0) f(X_i, \theta_0 - n^{-1/4}\tau_1) + \gamma_0 f(X_i, \theta_0 + n^{-1/4}\tau_2)}{f(X_i,\theta_0)}.$$

For the values of θ_1 and θ_2 in a small neighbourhood of θ_0, the following quadratic approximation can be obtained:

$$\Lambda_n = \tau^2 n^{-1/2}\sum_{i=1}^{n} Z_i - \frac{1}{2}\tau^4 n^{-1}\sum_{i=1}^{n} Z_i^2 + o_p(1),$$

and $M_n = \{(\sum W_i)^+\}^2/nEW_1^2 + o_p(1)$, where $W_i = Z_i(\theta_0) - h(\theta_0)Y_i(\theta_0)$. Let $V_n = \sum W_i/\sqrt{nEW_1^2}$. It is clear that the joint limiting distribution of (V_n, Λ_n) is bivariate normal with the mean vector $(0, -\tau^4 EZ_1^2/2)$ and the covariance matrix

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} 1 & \tau^2 E(W_1 Z_1)/\sqrt{EW_1^2} \\ \tau^2 E(W_1 Z_1)/\sqrt{EW_1^2} & \tau^4 EZ_1^2 \end{pmatrix}.$$

Since $-\tau^4 E Z_1^2/2 + \sigma_{22}/2 = 0$, the limiting distribution of V_n under the alternatives H_a^n is normal with mean $\sigma_{12} = \tau^2 E(W_1 Z_1)/\sqrt{EW_1^2}$ and variance 1. Since M_n is asymptotically equivalent to $\{V_n^+\}^2$, the theorem follows.

Acknowledgment We want to express our thanks to the Editor, the Associate Editor and the Referee for comments which have lead to a substantial improvement in the article. The research was partially supported by the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

Bickel, P. J. and Chernoff, H. (1993) Asymptotic distribution of the likelihood ratio statistic in a prototypical non regular problem. In *Statistics and Probability: a Raghu Raj Bahadur Festschrift* (eds: J.K. Ghosh, S.K. Mitra, K.R. Parthasarathy, and B.L.S. Prakasa Rao), pp. 83-96. Wiley Eastern.

Bickel, P. J., Klaassen, C. A. J., Ritov Y. and Wellner, J. A. (1993) *Efficient and Adaptive Estimation for Semi-parametric Models.* Baltimore: The Johns Hopkins University Press.

Chen, H. and Chen, J. (1998a) The likelihood ratio test for homogeneity in the finite mixture models. *Technical Report STAT 98-08.* Department of Statistics and Actuarial Science, University of Waterloo.

Chen, H. and Chen, J. (1998b) Tests for homogeneity in normal mixtures with presence of a structural parameter. *Technical Report STAT 98-09.* Department of Statistics and Actuarial Science, University of Waterloo.

Chen, H., Chen, J. and Kalbfleisch, J.D. (2000) A modified likelihood ratio test for homogeneity in the finite mixture models. *Technical Report STAT 2000-01.* Department of Statistics and Actuarial Science, University of Waterloo.

Chen, J. (1995) Optimal rate of convergence in finite mixture models. *Ann. Statist.*, **23**, 221-234.

Chen, J. (1998) Penalized likelihood-ratio test for finite mixture models with multinomial observations. *Can. J. Statist.*, **26**, 583-599.

Chen, J. and Kalbfleisch, J. D. (1996) Penalized minimum-distance estimates in finite mixture models, *Can. J. Statist.*, **24**, 167-175.

Cheng, R.C.H. and Traylor, L. (1995) Non-regular maximum likelihood problems (with discussion). *J. Roy. Statist. Soc.*, B, **57**, 3-44.

Chernoff, H. and Lander, E. (1995) Asymptotic distribution of the likelihood ratio test that a mixture of two binomials is a single binomial. *J. Statist. Plan. Inf.*, **43**, 19-40.

Dacunha-Castelle, D. and Gassiat, E. (1999) Testing the order of a model using locally conic parameterization: Population mixtures and stationary ARMA processes. *Ann. Statist.*, **27**, 1178-1209.

Davies, R. B. (1977) Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, **64**, 247-254.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, **74**, 33-43.

Ghosh, J.K. and Sen, P.K. (1985) On the asymptotic performance of the log likelihood ratio statistic for the mixture model and related results. In *Proc. Berk. Conf. in Honor of J. Neyman and J. Kiefer*, **2**, (eds L. LeCam and R.A. Olshen), pp. 789-806.

Hartigan, J.A. (1985) A Failure of Likelihood Asymptotics for Normal Mixtures. In *Proc. Berk. Conf. in Honor of J. Neyman and J. Kiefer*, **2**, (eds L. LeCam and R.A. Olshen), pp. 807-810.

LeCam, L. and Yang, G. L. (1990) *Asymptotics in Statistics; Some Basic Concepts.* New York: Springer-Verlag.

Lemdani, M. and Pons, O. (1999) Likelihood ratio tests in contamination models. *Bernoulli*, **5**, 705-719.

Lindsay, B.G. (1989) Moment matrices: applications in mixtures. *Ann. Statist.*, **17**, 722-740.

Lindsay, B.G. (1995) *Mixture Models: Theory, Geometry and Applications.* Hayward: Institute for Mathematical Statistics.

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

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

Neyman, J. and Scott, E.L. (1966) On the use of C(α) optimal tests of composite hypotheses (with discussion). *Bull. Inst. Internat. Statist.*, **41**, 477-497.

Schork, N. (1992) Bootstrapping likelihood ratios in quantitative genetics. In: *Exploring the Limits of the Bootstrap*, (eds Lepage, R. and Billard, L.), pp. 389-393. New York: Wiley.

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

Wilks, S.S. (1938). The large sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.*, **9**, 60-62.

Table 1: Alternative models with their Kullback-Leibler information.

Normal mixture: $(1-\gamma)N(\theta_1,1)+\gamma N(\theta_2,1)$				
γ	0.5	0.75	0.9	0.95
θ_1	-0.500	-0.866	-1.500	-2.179
θ_2	0.500	0.289	0.167	0.115
$KL \times 100$	1.358	1.444	1.842	2.583
Poisson mixture: $(1-\gamma)P(\theta_1)+\gamma P(\theta_2)$				
γ	0.5	0.75	0.9	0.95
θ_1	4.000	3.268	2.000	0.641
θ_2	6.000	5.577	5.333	5.229
$KL \times 100$	0.889	0.996	1.45	2.91

Table 2: Rejection rates (%) for the modified LRT, $C(\alpha)$ test and bootstrap tests. Nominal significance levels are 10%, 5% and 1%.

	Modified LRT			C(α)			Bootstrap		
	Normal Mixture $n = 64$								
H_0	9.4	4.8	1.2	10.8	5.4	1.3	10.5	5.6	1.0
$\gamma = 0.5$	53.1	38.8	16.4	53.3	39.1	16.7	46.8	33.4	12.5
$\gamma = 0.75$	52.5	38.9	16.9	52.7	39.0	16.6	46.6	33.0	13.2
$\gamma = 0.90$	53.2	40.1	19.4	52.4	39.2	17.9	49.8	36.8	17.2
$\gamma = 0.95$	53.1	42.4	23.7	51.1	39.6	19.3	54.5	43.2	23.6
	Normal Mixture $n = 100$								
H_0	9.5	4.9	1.1	9.6	4.75	0.9	9.3	5.1	0.7
$\gamma = 0.5$	63.7	49.6	25.9	63.9	50.3	26.9	57.2	41.1	18.3
$\gamma = 0.75$	63.5	50.1	26.8	63.6	50.5	27.4	56.9	42.5	19.2
$\gamma = 0.9$	63.5	51.3	29.5	62.5	50.1	28.1	63.7	51.0	27.4
$\gamma = 0.95$	64.1	52.9	34.0	61.1	48.9	28.7	67.4	56.1	36.7

Table 3: Rejection rates (%) for the modified LRT and $C(\alpha)$ tests. Nominal significance levels are 10%, 5% and 1%.

Poisson Mixture												
	$n = 64$						$n = 100$					
	Modified LRT			C(α)			Modified LRT			C(α)		
H_0	9.7	4.7	1.0	8.6	4.5	1.1	9.6	4.8	1.2	9.0	4.7	1.4
γ=0.5	38.7	27.8	10.3	38.6	27.7	12.0	51.6	38.6	15.4	51.6	38.4	16.0
γ=0.75	40.2	28.9	11.5	38.5	27.8	12.3	53.6	40.6	16.4	52.2	38.5	14.8
γ=0.9	44.5	33.3	14.8	39.8	28.6	12.7	56.4	44.0	20.8	52.7	38.7	15.5
γ=0.95	48.7	39.2	22.4	39.9	29.1	13.3	61.9	52.2	31.0	52.5	39.3	16.0

Table 4: Rejection rates (%) of the modified LRT and of Davies' method. The nominal significance level is 5% and $n = 100$.

(θ, γ)	(0, 0)	(.1, .1)	(.2, .2)	(.3, .3)
Davies' method	4.9	4.9	5.9	9.7
Modified LRT	5.2	5.3	6.9	15.5