TESTING FOR A FINITE MIXTURE MODEL WITH TWO COMPONENTS

Hanfeng Chen, Jiahua Chen and John D. Kalbfleisch
Bowling Green State University and University of Waterloo

Abstract

We consider a finite mixture model with k components and a kernel distribution from a general parametric family. We consider the problem of testing the hypothesis $k = 2$ against $k \geq 3$. In this problem, the likelihood ratio test has a very complicated large sample theory and is difficult to use in practice. We propose a test based on the likelihood ratio statistic where the estimates of the parameters, (under the null and the alternative) are obtained from a penalized likelihood which guarantees consistent estimation of the support points. The asymptotic null distribution of the corresponding modified likelihood ratio test is derived and found to be relatively simple in nature and easily applied. Simulations based on a mixture model with normal kernel are encouraging that the modified test performs well, and its use is illustrated in an example involving data from a medical study where the hypothesis arises as a consequence of a potential genetic mechanism.

Key words and phrases. Asymptotic distribution, finite mixture models, likelihood ratio tests, penalty terms, non-regular estimation, strong identifiability.

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

1 Introduction

Finite mixture models are often used to study data from a population that is suspected to be composed of a number of homogeneous subpopulations. For example, mixture distributions are used routinely to accommodate the genetic heterogeneity thought to underlie many human diseases. See, for example, Friedlander and Leitersdorf (1995); Heiba, *et al.* 1995; Schork, Allison and Thiel, 1996; and Ott, 1999.

We consider a finite mixture distribution with probability density function (pdf)

$$f(x; G) = \int f(x, \theta) dG(\theta), \tag{1}$$

where $f(x, \theta)$ is a specified pdf (called the kernel function) with parameter $\theta \in \Theta$, and $G(\theta)$ is a discrete cumulative distribution function (called the mixing distribution) with a finite number of support points. Let $\theta_j \in \Theta$, $j = 1, \cdots, k$ be the support points of G and $\pi_1, \ldots, \pi_k$ be the corresponding weights ($\sum \pi_j = 1$). Then

$$G(\theta) = \sum_{j=1}^{k} \pi_j I(\theta_j \leq \theta),$$

where $I(\cdot)$ is the indicator function. The class of all finite mixing distributions with k support points is denoted by $\mathcal{M}_k$, i.e.,

$$\mathcal{M}_k = \{G(\theta) = \sum_{j=1}^{k} \pi_j I(\theta_j \leq \theta) : \theta_1 \leq \cdots \leq \theta_k, \sum_{j=1}^{k} \pi_j = 1\}.$$

The class of all finite mixing distributions is $\mathcal{M} = \cup_{k \geq 1} \mathcal{M}_k$.

We assume that the model (1) is identifiable by the mixing distribution G in the sense that $f(x; G_1) = f(x; G_2)$, for all x, implies $G_1 = G_2$. We also assume that the parameter space Θ of the kernel function is compact. It should be noted that, while the compactness of Θ is a basic technical requirement of the study (see, e.g., Ghosh and Sen, 1985; Dacunha-Castelle and Gassiat, 1999), it may not be so restrictive in applications since the parameter θ is typically known to be bounded.

The books by Titterington, Smith and Makov (1985), McLachlan and Basford (1988) and Lindsay (1995) give extensive discussion of the background of finite mixture models. Some important more recent developments can be found in Cheng and Traylor (1995), Bickel and Chernoff (1993), Chernoff and Lander (1995), Lemdani and Pons (1999), Dacunha-Castelle and Gassiat (1999) and Chen and Chen (2001), Qin (1998), Liang and Rathouz (1999).

As noted above, finite mixture models are useful models for many types of data, but they are non-regular. As a consequence, it is often difficult to apply standard inferential procedures to finite mixture models. For example, it has long been recognized that the likelihood ratio test statistic for the hypothesis $G \in \mathcal{M}_1$ (or $k = 1$) does not have the usual chi-squared limiting distribution (Hartigan, 1985), and there have been extensive discussions on testing this hypothesis. For example, Neyman and Scott (1966) proposed the well known $C(\alpha)$ test. MaLachlan (1987) used a bootstrap method to simulate the quantiles of the null distribution of the likelihood

ratio statistic. Chen, Chen and Kalbfleisch (2001) developed a modified likelihood approach which they showed to be simple to use and have superior performance in many situations.

Constructing a test of the hypothesis $G \in \mathcal{M}_2$ (or $k = 2$) is similar in principle to $k = 1$. Perhaps due to its mathematical complexity, however, there is a less extensive literature. Some approaches can be found in the diagnostic method (Roeder, 1994 and Lindsay and Roeder, 1997), the moment methods (Lindsay, 1989), and the model selection approach (Chen and Kalbfleisch, 1996; Henna, 1985). If the component distributions are known, Chen and Cheng (1997) discussed a bootstrap method. Even though there is relatively little literature on the subject, the problem of testing $k = 2$ is also important in applications. For example, if a quantitative trait is determined by a simple gene with two alleles, a mixture of two normal distributions might be appropriate when the mode of inheritance is dominant, whereas a mixture of three or more normals will be appropriate when the mode of inheritance is additive or more complex in nature. In this and other examples in genetics and elsewhere, determining the size of k can play a crucial role. More generally, we are interested in determining how large k needs to be to adequately describe the data; in the interest of parsimony, we prefer models with less complexity.

In this paper, we consider mixture models in the class $\mathcal{M}_k$ and, with reference to that class, test the null hypothesis that $G \in \mathcal{M}_2$. Our approach is to define a modified likelihood function on the class $\mathcal{M}_k$ and interpret the null hypothesis within this class. If k is large enough, the modified LRT has a relatively simple asymptotic distribution under the null hypothesis. It is argued that this distribution is appropriate for a test of the null hypothesis $G \in \mathcal{M}_2$ against the general alternative $G \in \mathcal{M}$.

In Section 2, we give a list of conditions with some brief discussions. In Section 3, we introduce the modified likelihood ratio statistic, and show the consistency of the modified maximum likelihood estimate of the mixing distribution. The limiting distribution of the test statistic is given in Section 4, and a sketch of the derivation is also included. Section 5 contains an example based on data from Roeder (1994), and Section 6 presents some simulation results. Technical proofs are presented in the Appendix.

2 A Modified Likelihood Ratio Test Procedure

Let $X_1, ..., X_n$ be a random sample of size n from the model (1). One wishes to test the hypothesis $H_0 : G \in \mathcal{M}_2$. As remarked in the introduction, the ordinary LRT or its bootstrap approximation is impractical for use. In this paper, a modified from the LRT is proposed. The new procedure is described in this section. We will proceed with the modified maximum likelihood estimation of G. Throughout the paper, assume that the *true* mixing distribution is

$$G_0(\theta) = \pi_0 I(\theta_{01} \leq \theta) + (1 - \pi_0) I(\theta_{02} \leq \theta), \tag{2}$$

where θ_{01} and θ_{02} are distinct interior points of Θ and $0 < \pi_0 < 1$. All expectations and probabilities are with respect to this null distribution.

2.1 Modified likelihood function of G

The main complications of the asymptotic null distribution of the ordinary LRT under the mixture model (1) are due to the fact that the maximum likelihood estimates (MLE's) of the weights π_j can be very close to the boundary point 0. As a consequence, some of the MLE's of the support points θ_j's become inconsistent, and a quadratic approximation to the likelihood function fails. These complications are expected to disappear if we can prevent the estimates of π_j from being too close to 0. This consideration results in the following definition of modified likelihood function. For $G(\theta) = \sum_{i=1}^{k} \pi_i I(\theta_i \leq \theta) \in \mathcal{M}_k$ with $k \geq 2$, the modified likelihood function is defined as

$$\tilde{l}_n(G|k) = l_n(G) + C_k \sum_{j=1}^{k} \log \pi_j, \tag{3}$$

where C_k is some positive constant and

$$l_n(G) = \sum_{i=1}^{n} \log f(X_i; G)$$

is the ordinary likelihood function. The constant C_k determines the penalty size on the proportion parameters π_j of G. Though the asymptotic properties of the statistical procedures based on $\tilde{l}_n(G|k)$ does not depend on choice of C_k, in practice choice of C_k is expected to have some effects on performace of the statistical procedures for

small or moderate sample sizes. One basis for the choice of C_k is to reflect the size of Θ. See Chen, Chen and Kalbfleisch (2001) for further discussion.

The modified MLE of G is then obtained by maximizing $\tilde{l}_n$ over the space $\mathcal{M}_k$. We denote the modified MLE of G by $\hat{G} = \hat{G}^{(k)}$, and the modified MLE's of θ_j and π_j by $\hat{\theta}_j$ and $\hat{\pi}_j$. Thus, the modified MLE of G under the null hypothesis is $\hat{G}_0 = \hat{G}^{(2)}$ that maximizes the modified likelihood (3) when $k = 2$. Let $\hat{\pi}_0$, $\hat{\theta}_{01}$, and $\hat{\theta}_{02}$ be the modified MLE's of π, θ_{01}, and θ_{02}, respectively, i.e.,

$$\hat{G}_0(\theta) = \hat{\pi}_0 I(\hat{\theta}_{01} \leq \theta) + (1 - \hat{\pi}_0) I(\hat{\theta}_{02} \leq \theta).$$

2.2 An adaptive choice of k

It is important to note that the class $\mathcal{M}_k$ implicitly contains the models with fewer than k distinct support points. These models are obtained by allowing the θ_j's to coincide with one another while still maintaining separate weights π_j. In general, there are infinitely many representations of any specific model with fewer than k distinct support points. By convention, we will consider the representation that gives rise to the maximum modified likelihood (3). More specifically, the true null distribution (2) in $\mathcal{M}_2$ can be written as an element of $\mathcal{M}_k$ as

$$G_0^{(k)} = \sum_{j=1}^{k} \pi_j^{(0)} I(\theta_j^{(0)} \leq \theta) \tag{4}$$

where

$$\theta_i^{(0)} = \theta_{01},\ i = 1, ..., r_0,\ \text{and}\ \theta_i^{(0)} = \theta_{02},\ i = r_0 + 1, ..., k,$$

and

$$\pi_i^{(0)} = \frac{\pi_0}{r_0}, i = 1, ..., r_0 \text{ and } \pi_i^{(0)} = \frac{1 - \pi_0}{k - r_0},\ i = r_0 + 1, ..., k.$$

In these expressions, r_0 is chosen to minimize the penalty term

$$-C_k \left[r \log \frac{\pi_0}{r} + (k - r) \log \frac{1 - \pi_0}{k - r} \right]. \tag{5}$$

Note that

$$\tilde{l}_n(G_0^{(k)} | k) = C_k \left[r_0 \log \frac{\pi_0}{r_0} + (k - r_0) \log \frac{1 - \pi_0}{k - r_0} \right]. \tag{6}$$

We denote the class of all such two point distributions in $\mathcal{M}_k$ as $\mathcal{M}_k^{(0)}$.

Note that r_0 as defined in (5) is implicitly a function of the unknown weight π_0 in the true null distribution and the order k of the mixture model. Without loss of generality, we suppose that $\pi_0 < 0.5$ and define

$$k^* = \min\{k : r_0 \geq 2\}, \text{ i.e. } k^* = [1.5/\pi_0], \tag{7}$$

where $[x]$ is the smallest integer greater than or equal to x. The MLE of k^* under the null hypothesis is $\hat{k}^* = [1.5/\hat{\pi}_0]$. The modified LRT is then based on the statistic

$$R_n = 2\{l_n(\hat{G}) - l_n(\hat{G}_0)\},$$

where $\hat{G}$ is the modified MLE of G in $\mathcal{M}_k$ with $k \geq k^*$.

2.3 The testing procedure

The modified LRT procedure proposed above for testing

$$H_0 : \ G \in \mathcal{M}_2 \ \text{ versus } \ H_1 : \ G \in \mathcal{M}$$

is summarized as follows.

Step 1. Obtain the estimate $\hat{G}_0$ which maximizes the modified likelihood function $\tilde{l}_n(G|2)$ over $\mathcal{M}_2$. Let $\hat{\pi}_0$ be the probability mass of $\hat{G}_0$ at the lower support point and $\hat{k}^* = [1.5/\hat{\pi}_0]$.

Step 2. Let $k \geq \hat{k}^*$ be any integer. Obtain the estimate $\hat{G}$ which maximizes $\tilde{l}_n(G|k)$ over $\mathcal{M}_k$.

Step 3. Compute the testing statistic $R_n = 2\{l_n(\hat{G}) - l_n(\hat{G}_0)\}$. Reject the null hypothesis H_0 if R_n is large.

A critical value of R_n can be approximated by its limiting distribution. As presented in Corollory 1, for large sample size, the null distribution of R_n is asymptotically distributed as a mixture of chisquares as follows:

$$(\frac{1}{2} - \frac{\alpha}{2\pi})\chi_0^2 + \frac{1}{2}\chi_1^2 + \frac{\alpha}{2\pi}\chi_2^2,$$

where α is defined in Theorem 2.

3 A Large Sample Theory of the Modified LRT

The exact null distribution of R_n is intractable, but a critical value or the p-value can be approximated by the limiting null distribution of R_n . We will outline a large sample theory of R_n in this section, while rigorous proofs of the asymptotic results, together with a number of regularity conditions (Conditions 1-5), are given in Appendix.

In the asymptotical analysis of R_n , we first treat $k^* = [1.5/\pi_0]$ as given, so $k \geq k^*$ replaces the requirement $k \geq \hat{k}^*$ in the modified LRT procedure. We will then show that the limiting null distribution of R_n remains the same when k^* is estimated by $\hat{k}^*$.

In the modified LRT procedure, the requirement of $k \geq k^*$ ensures the simplicity of the limiting distribution of R_n without loss of statistical consideration of the testing problem. For details, see the remarkes following Theorem 2.

3.1 Consistency of the modified MLE of G

As similar to the ordinary LRT, the large sample behavior of the modified LRT relies on the asymptotic properties of the estimate of G. Let $\hat{G}$ be the modified MLE of G over $\mathcal{M}_k$ and put

$$\hat{G}(\theta) = \sum_{j=1}^{k} \hat{\pi}_j I(\hat{\theta}_j \leq \theta).$$

Lemma 1 *Suppose that Conditions 1-5 listed in Appendix hold and that the true distribution is $f(x; G_0)$. For any given $k > 0$, there exists a positive constant $\epsilon = \epsilon(k)$ such that*

$$\lim_{n\to\infty} P(\hat{\pi}_1 \geq \epsilon, \cdots, \hat{\pi}_k \geq \epsilon) = 1.$$

An immediate and yet important implication of Lemma 1 is that under the null distribution $f(x; G_0)$, all the modified MLE's $\hat{\theta}_j$ are consistent; that is, the support points of $\hat{G}$ converge to those of G_0. For clarification, let $\theta_0 = (\theta_{01} + \theta_{02})/2$ be the average of the support points of G_0. Define

$$\hat{\pi} = \hat{G}(\theta_0)$$

which is the probability assigned to the support points $\hat{\theta}_j \leq \theta_0$ by the mixing distribution estimate $\hat{G}$. Then $\hat{G}$ can be expressed as a mixture as follows:

$$\hat{G}(\theta) = \hat{\pi}\hat{G}_1(\theta) + (1-\hat{\pi})\hat{G}_2(\theta),$$

where $\hat{G}_1(\theta_0) = 1$ and $\hat{G}_2(\theta_0) = 0$. Similarly, we can express the null mixing distribution G_0 as

$$G_0(\theta) = \pi_0 G_1(\theta) + (1-\pi_0)G_2(\theta),$$

where $G_1(\theta_0) = 1$ and $G_2(\theta_0) = 0$.

Theorem 1 *Suppose that Conditions 1-5 hold and that the true distribution is $f(x; G_0)$. Then*

(a) $\hat{\pi} = \pi_0 + o_p(1)$.

(b) For $i = 1, 2$

$$|\hat{G}_i - G_i| = o_p(1),$$

where $|F_1 - F_2|$ is the supremum distance between two probability distributions F_1 and F_2, namely,

$$|F_1 - F_2| = \sup_x |F_1(x) - F_2(x)|.$$

(c) All support points of $\hat{G}_i$ converge to those of G_i for $i = 1, 2$.

(d) The absolute moment $\int |\theta - \theta_{0i}|^r d\hat{G}_i(\theta) = o_p(1)$ for $i = 1, 2$ and $r > 0$.

3.2 Limiting distribution of R_n.

Recall $R_n = 2\{l_n(\hat{G}) - l_n(\hat{G}_0)\}$, where $\hat{G}$ maximizes $\tilde{l}_n(G)$ over $\mathcal{M}_k$ with $k > k^*$. Let

$$R_n = R_{1n} - R_{0n},$$

where $R_{1n} = 2\{l_n(\hat{G}) - l_n(G_0)\}$ and $R_{0n} = 2\{l_n(\hat{G}_0) - l_n(G_0)\}$. The limiting distribution of R_n will be established by the quadratic-type expansions of R_{1n} and R_{0n}. The following quantities play an important role in our study: for $i = 1, \cdots, n$ and $j = 1, 2$

$$Y_{ij}(\theta) = \frac{f(X_i, \theta) - f(X_i, \theta_{0j})}{f(X_i; G_0)}, \qquad Y_i'(\theta) = \frac{f'(X_i, \theta)}{f(X_i; G_0)},$$

$$Y_i''(\theta) = \frac{f''(X_i,\theta)}{f(X_i;G_0)}, \quad Y_i'''(\theta) = \frac{f'''(X_i,\theta)}{f(X_i;G_0)}. \tag{8}$$

Quadratic expansion of R_{1n}. Put $R_{1n} = 2\sum \log(1+\delta_i)$, with

$$\delta_i = \frac{f(X_i;\hat{G}) - f(X_i;G_0)}{f(X_i;G_0)}. \tag{9}$$

A quadratic-type expansion of R_{1n} can be given by the Taylor expansion of the function $\log(1+\delta)$. As always in a large sample study, the key point and yet the difficult part is the justification of negligibility of the remainder.

Using a typical technique of adding and subtracting an identical expression, we have

$$\delta_i = (\hat{\pi} - \pi_0)\Delta_i + \hat{\pi}\frac{f(X_i;\hat{G}_1) - f(X_i,\theta_{01})}{f(X_i;G_0)} + (1-\hat{\pi})\frac{f(X_i;\hat{G}_2) - f(X_i,\theta_{02})}{f(X_i;G_0)}, \tag{10}$$

where

$$\Delta_i = [f(X_i,\theta_{01}) - f(X_i,\theta_{02})]/f(X_i;G_0).$$

Note the symmetry of the second and third terms on the right side of (10). Note also that the consistency result of Theorem 1 implies that all three terms converge to zero in probability.

When $\theta - \theta_{01} = o_p(1)$, we have

$$\sum_{i=1}^n \frac{f(X_i,\theta) - f(X_i,\theta_{01})}{f(X_i;G_0)} \approx (\theta-\theta_{01})\sum_{i=1}^n Y_i'(\theta_{01}) + \frac{1}{2}(\theta-\theta_{01})^2\sum_{i=1}^n Y_i''(\theta_{01}),$$

where Y_i' and Y_i'' are defined in (8). Put

$$\hat{m}_{ij} = \int (\theta - \theta_{0j})^i d\hat{G}_j(\theta).$$

Then

$$\sum_{i=1}^n \frac{f(X_i,\hat{G}_1) - f(X_i,\theta_{01})}{f(X_i;G_0)} \approx \hat{m}_{11}\sum_{i=1}^n Y_i'(\theta_{01}) + \frac{\hat{m}_{21}}{2}\sum_{i=1}^n Y_i''(\theta_{01}),$$

with a similar expression for $\hat{G}_2$, It follows that

$$\begin{aligned}\sum_{i=1}^n \delta_i \;\approx\; \sum_{i=1}^n \Big[&(\hat{\pi}-\pi_0)\Delta_i + \hat{\pi}\hat{m}_{11}Y_i'(\theta_{01}) + (1-\hat{\pi})\hat{m}_{12}Y_i'(\theta_{02}) \\ &+ \hat{\pi}\frac{\hat{m}_{21}}{2}Y_i''(\theta_{01}) + (1-\hat{\pi})\frac{\hat{m}_{22}}{2}Y_i''(\theta_{02})\Big]\end{aligned}$$

Using the inequality $\log(1+\delta_i) \leq \delta_i - \delta_i^2/2 + \delta_i^3/3$ and neglecting the high order terms (see the Appendix), we get

$$R_{1n} = 2\sum_{i=1}^{n} \log(1+\delta_i) = L_n - Q_n + o_p(1), \tag{11}$$

where $L_n = 2\sum_{i=1}^n \delta_i$ and $Q_n = \sum_{i=1}^n \delta_i^2$

Note that L_n and Q_n are, respectively, linear and quadratic functions of

$$\hat{\mathbf{t}} = (\hat{\pi} - \pi_0, \hat{\pi}_0\hat{m}_{11}, (1-\hat{\pi}_0)\hat{m}_{12}, \hat{\pi}_0\frac{\hat{m}_{21}}{2}, (1-\hat{\pi}_0)\frac{\hat{m}_{22}}{2})^\tau.$$

Let

$$\mathbf{b}_i = \left(\Delta_i, Y_i'(\theta_{01}), Y_i'(\theta_{02}), Y_i''(\theta_{01}), Y_i''(\theta_{02})\right).$$

Finally, let

$$\mathbf{b} = \sum_{i=1}^{n} \mathbf{b}_i \qquad \text{and} \qquad \mathbf{B} = \sum_{i=1}^{n} \mathbf{b}_i\mathbf{b}_i^\tau.$$

It can then be seen that $L_n = 2\mathbf{b}^\tau\hat{\mathbf{t}}$ and $Q_n = \hat{\mathbf{t}}^\tau\mathbf{B}\hat{\mathbf{t}}$ so that, from (11)

$$R_{1n} \approx L_n - Q_n \leq \sup_{\mathbf{t}} \left[2\mathbf{b}^\tau\mathbf{t} - \mathbf{t}^\tau\mathbf{B}\mathbf{t}\right] + o_p(1). \tag{12}$$

where

$$\mathbf{t} = (\pi - \pi_0, \pi_0 m_{11}, (1-\pi_0)m_{12}, \pi_0\frac{m_{21}}{2}, (1-\pi_0)\frac{m_{22}}{2})^\tau$$

and $\pi = \pi(G) = G(\theta_0)$ and $m_{ij} = m_{ij}(G) = \int(\theta - \theta_{0j})^i dG(\theta)$. Thus, to find the supremum in (12), $\mathbf{t}$ ranges over its admissible values generated by π and m_{ij} as G ranges over the region defined by the alternative hypothesis.

Let $\mathbf{b}^\tau = (\mathbf{b}_1^\tau, \mathbf{b}_2^\tau)$, $\mathbf{t}^\tau = (\mathbf{t}_1^\tau, \mathbf{t}_2^\tau)$ and

$$\mathbf{B} = \left(\begin{array}{c|c} B_{11} & B_{12} \\ \hline B_{21} & B_{22} \end{array}\right)$$

where $\mathbf{b}_1$ and $\mathbf{t}_1$ are 3×1 vectors and B_{11} is a 3×3 matrix. Some algebra shows that

$$2\mathbf{b}^\tau\mathbf{t} - \mathbf{t}^\tau\mathbf{B}\mathbf{t} = 2\mathbf{b}_1^\tau\tilde{\mathbf{t}}_1 - \tilde{\mathbf{t}}_1^\tau B_{11}\tilde{\mathbf{t}}_1 + 2\tilde{\mathbf{b}}_2^\tau\mathbf{t}_2 - \mathbf{t}_2^\tau\tilde{B}_{22}\mathbf{t}_2, \tag{13}$$

where $\tilde{\mathbf{t}}_1 = \mathbf{t}_1 - B_{11}^{-1}B_{12}\mathbf{t}_2$, $\tilde{\mathbf{b}}_2^\tau = \mathbf{b}_2^\tau - \mathbf{b}_1^\tau B_{11}^{-1}B_{12}$, and $\tilde{B}_{22} = B_{22} - B_{21}B_{11}^{-1}B_{12}$. When $\tilde{\mathbf{t}}_1$ and $\mathbf{t}_2$ are regarded as free variables, we find that

$$R_{1n} \approx L_n - Q_n \leq \mathbf{b}_1^\tau B_{11}^{-1}\mathbf{b}_1 + \sup_{\mathbf{t}_2}\{2\tilde{\mathbf{b}}_2^\tau\mathbf{t}_2 - \mathbf{t}_2^\tau\tilde{B}_{22}\mathbf{t}_2\} + o_p(1). \tag{14}$$

Since $\mathbf{t}_2^\tau = (\pi_0 \frac{m_{21}}{2}, (1-\pi_0)\frac{m_{22}}{2})$, the supremum over $\mathbf{t}_2$ in (14) is taken within the first quadrant of the R^2 plane.

The upper bound in (14) can be attained and a useful quadratic-type expansion of R_{1n} is thus obtained. The result is summarized in the following lemma.

Lemma 2 *Suppose that Conditions 1-5 hold and that the true distribution is $f(x; G_0)$ with $0 < \pi_0 < 1$ and $\theta_{01} \neq \theta_{02}$. Then as $n \to \infty$*

$$R_{1n} = \mathbf{b}_1^\tau B_{11}^{-1} \mathbf{b}_1 + \sup_{\mathbf{t}_2} \{2\tilde{\mathbf{b}}_2^\tau \mathbf{t}_2 - \mathbf{t}_2^\tau \tilde{B}_{22} \mathbf{t}_2\} + o_p(1),$$

where the supremum over $\mathbf{t}_2$ is taken within the first quadrant of the R^2 plane.

It is important to note that all components of $\mathbf{b}$ have mean zero and that the matrix $n^{-1}\mathbf{B}$ converges to the covariance matrix of $n^{-1/2}\mathbf{b}$. This is the key ingredient of obtaining the classical chi-squared limiting distribution.

Quadratic expansion of R_{0n}. The analysis of $R_{0n} = l_n(\hat{G}_0) - l_n(G_0)$ is similar to that above for R_{1n} except for the following difference: When $k = 2$ (the null hypothesis), each of $\hat{G}_1$ and $\hat{G}_2$ has a single support point so that $\hat{m}_{2j} = \hat{m}_{1j}^2$. The terms of Y_i'' are thus controlled by those of Y_i', implying that in (13),

$$2\tilde{\mathbf{b}}_2^\tau \mathbf{t}_2 - \mathbf{t}_2^\tau \tilde{B}_{22} \mathbf{t}_2 = o_p\{2\mathbf{b}_1^\tau \tilde{\mathbf{t}}_1 - \tilde{\mathbf{t}}_1^\tau B_{11} \tilde{\mathbf{t}}_1\}.$$

Thus, it follows that

$$R_{0n} \leq \sup_{\tilde{\mathbf{t}}_1} 2\mathbf{b}_1^\tau \tilde{\mathbf{t}}_1 - \tilde{\mathbf{t}}_1^\tau B_{11} \tilde{\mathbf{t}}_1.$$

This gives the following lemma.

Lemma 3 *Suppose that Conditions 1-5 hold and that the true distribution is $f(x; G_0)$ with $0 < \pi_0 < 1$ and $\theta_{01} \neq \theta_{02}$. Then as $n \to \infty$*

$$R_{0n} = \mathbf{b}_1^\tau B_{11}^{-1} \mathbf{b}_1 + o_p(1).$$

From Lemmas 2 and 3, it follows that

$$R_n = \sup_{\mathbf{t}_2} \{2\tilde{\mathbf{b}}_2^\tau \mathbf{t}_2 - \mathbf{t}_2^\tau \tilde{B}_{22} \mathbf{t}_2\} + o_p(1), \tag{15}$$

where the supremum over $\mathbf{t}_2$ is taken within the first quadrant of the R^2 plane. As remarked before, the two components of $\tilde{\mathbf{b}}_2$ have mean zero and the matrix $n^{-1}\tilde{B}_{22}$ converges to the covariance matrix of $n^{-1/2}\tilde{\mathbf{b}}_2$. Depending on the nature of the alternative hypothesis, which can place various restrictions on the range $\mathbf{t}_2$, the limiting distribution of R_n may have different forms. In the following, we state the simplest and the most important case.

Theorem 2 *Suppose that Conditions 1-5 hold and that the true distribution is $f(x; G_0)$ with $0 < \pi_0 < 1$ and $\theta_{01} \neq \theta_{02}$. Suppose that k satisfies $k \geq k^*$. Then the asymptotic distribution of the modified LRT statistic R_n is that of the mixture*

$$(\frac{1}{2} - \frac{\alpha}{2\pi})\chi_0^2 + \frac{1}{2}\chi_1^2 + \frac{\alpha}{2\pi}\chi_2^2,$$

where $\alpha = \arccos(\rho)$ and ρ is the correlation coefficient between the two components of $\tilde{\mathbf{b}}_2$.

At last, it is pointed out that the result of Theorem 2 remains true even if $k^* = [1.5/\pi_0]$ actually has to be estimated by $\hat{k}^* = [1.5/\hat{\pi}_0]$ as proposed in the modified LRT procedure.

Corollary 1 *Suppose that Conditions 1-5 hold and that the true distribution is $f(x; G_0)$ with $0 < \pi_0 < 1$ and $\theta_{01} \neq \theta_{02}$. Suppose that $k \geq \hat{k}^*$. The result of Theorem 2 remains true, i.e., the asymptotic distribution of the modified LRT statistic R_n is that of the mixture*

$$(\frac{1}{2} - \frac{\alpha}{2\pi})\chi_0^2 + \frac{1}{2}\chi_1^2 + \frac{\alpha}{2\pi}\chi_2^2,$$

where α is given in Theorem 2.

Remark 1. As seen in the analysis, using $k \geq k^*$ in the modified LRT procedure enables the fitted second moments $\hat{m}_{j2}$ to achieve the upper bound. The resulting limiting distribution becomes simple. On the other hand, the procedure itself is suitable for any $k \geq 3$. However, we would like to point out that the limiting distribution with $k < k^*$ is stochastically less than the limiting distribution with $k \geq k^*$. Thus, the asymptotic result given by Theorem 2 could be used for any $k \geq 3$ and in the case of $k < k^*$, the asymptotic critical value would be conservative with some loss in power.

Remark 2. Note that the limiting distribution of R_n does not depend on k as long as $k \geq k^*$. This gives the advantage and flexibility to fit the finite mixture models without bound on the number of mixture components.

4 Examples and Simulation Studies

4.1 Example 1.

The data set considered in Roeder(1994) was analyzed to illustrate the modified LRT approach. The data set consists of 190 observations of red blood cell sodium-lithium countertransport(SLC). As discussed by Roeder, geneticists are interested in SLC because it is correlated with blood pressure and hence may be an important cause of hypertension. The condition is also easier to study than blood pressure because the latter is a complex trait that is highly variable and affected by environmental and perhaps many genetic factors.

One possibility is that SLC is determined by a simple mode of inheritance compatible with the action of a single gene with two alleles, A_1 and A_2, which occur with probabilities p and $1-p$. In this case, we might suppose that each observation is composed of the sum of a genetic component and a normally distributed measurement error. This would lead to a finite normal mixture model with common variance. A single dominance model for the gene yields a finite mixture model with $k = 2$ components whereas an additive model yields a finite mixture model with $k = 3$. A mixture model with more components is also possible if the mode of inheritance is complex. Roeder (1994) gives more background and references as well as illustrating some graphical methods of analysis.

Using the modified likelihood with $C_2 = 1$, the best fit with $k = 2$ is a model with $\theta_1 = 0.236$, $\theta_2 = 0.443$, $\pi_1 = 0.866$ and $\pi_2 = 0.134$. The common standard deviation is $\sigma = 0.070$. For $k \geq 3$, we fit a model with $k = 4$ and $C_4 = 0.5$, and the maximum modified likelihood estimates of the four location parameters $0.223, 0.223, 0.377, 0.571$ with corresponding probabilities $0.382, 0.382, 0.205, 0.030$. The common standard deviation is $\sigma = 0.058$. Note that the first and second location parameters are identical, the outcome in fact chooses a model with $k = 3$. This fit is almost identical to that given by Roeder (1994) except that the penalty term in the modified likelihood results

in fitted probabilities that are shifted slightly toward the center of interval [0, 1].

Using a graphical diagnostic method, Roeder (1994) suggested that $k = 3$ is the most appropriate model, which would correspond to the additive model in genetics. The diagnostic curve for $k = 2$ also falls within the 90% confidence bands but only just. The graphical analysis is suggestive that $k = 3$ is more appropriate, but it does not give a conclusive answer.

The modified likelihood ratio statistic for testing $k = 2$ against $k \geq 3$ is found to be 9.47. (The choices of C_2 and C_4 have very limited influence). The mixing parameter $\frac{\alpha}{2\pi}$ in the limiting distribution is estimated as 0.15. Accordingly, the asymptotic p-value is 0.0025. Therefore, we can conclusively reject the null model $k = 2$. In addition, models with $k \geq 4$ do not outperform the model with $k = 3$ as is evident from the fact that the overall modified MLE is concentrated on three points. We further claim that there is no evident to suggest that $k \geq 4$.

It should be noted that our theoretical result is based on the assumption that the variance is known, whereas we have estimated the variance above. An important generalization of this work would include models with unknown variance. We plan to report on this extension in a future publication.

4.2 Similation studies

We have conducted simulations under finite normal mixture models. For each model considered, we generated $n = 100$ and 200 samples to evaluate the modified LRT of the hypothesis $k = 2$ against $k \geq 3$. A slightly adjusted E-M algorithm (see Chen, Chen and Kalbfleisch, 2001) was used to compute the maximum modified likelihood estimate of the mixing distribution. Since the data sets were simulated, the choice of initial values was simplified. We selected two or k initial support points which were more or less consistent with the generating distribution. For example, if the true distribution had 2 support points, we took the true values as the starting values for fitting $k = 2$. For fitting $k = 6$, we spread 6 support points over the range of data, and set all the initial weights equal. Since we were able to use the true values in specifying initial values, convergence of the algorithm was not a problem. We did some detailed investigations of special cases which indicated that we usually obtained a global maximum or, in a relatively small percentage of cases, a local maximum that

gave a likelihood value near the global maximum. (In the case of analyzing real data, one should always try different initial values to be reasonably sure that the algorithm converges to the global maximum. This was our approach in the example discussed in the last section.)

For the simulations, we chose $C_2 = C_6 = 1$. We explored other choices of C_k, but found little difference. We chose $k = 6$ since this was large enough to exceed the critical k^* for the cases considered. The results were not affected by selecting other larger values of k. Once the estimate $\hat{G}_0$ is obtained, the the angle α in the asymptotic distribution is easily estimated. The nominal p-value or significance level can then be determined for each simulated sample.

We selected six distributions under the normal null model (See Table 1) to represent a variety of situations. When the difference between the two support points is less than or equal to two standard deviations, the corresponding mixture density has only one mode. One might expect difficulties in identifying the need for even as many as two components for these model. We have done some preliminary investigations for such mixtures, but in the present study we concentrated attention on true null mixture distributions that are bimodal, but with some variation in the degree separation and sizes of the modes.

The rejection rates for normal models in Table 2 are based on 2,000 repetitions. The significance levels are chosen as 10%, 5%, and 2.5%. According to the simulation results, the null rejection rates are close to the nominal values for both $n = 100$ and 200.

Our simulation indicates that the null limiting distribution gives satisfactory approximations in all cases. These results are very encouraging and suggest that the asymptotic approximations work reasonably well. More numerical work is needed, however, to investigate how best to select the penalty constant and to identify the applicability of the asymptotic results.

The rejection rates under the alternative models are usually substantially larger than the significance levels. When $n = 200$, and the more extreme alternatives are considered, the power of the test approaches 1 suggesting the consistency of the method. The alternative models 1A, 2A and 3A give mixture densities that are similar to those achievable under the null hypothesis. While the method still appears to be consistent in these cases, there is less power as should be expected.

Appendix: Regularity Conditions and Proofs

A. Regularity conditions

Although the mixture model (1) is non-regular, it is reasonable and necessary to require that the kernel function $f(x, \theta)$ be regular. We list these regular conditions followed by brief discussion.

Condition 1. *Wald's integrability conditions.* The kernel function $f(x, \theta)$ is such that the mixture distribution $f(x; G)$ satisfies Wald's integrability conditions for consistency of the maximum likelihood estimate (see Leroux, 1992). It is sufficient to assume that

(i) $E|\log f(X; G_0)| < \infty$, and (ii) there exists $\rho > 0$ such that for each G, $f(x; G, \rho)$ is measurable and $E \log f(X; G, \rho) < \infty$, where

$$f(x; G, \rho) = 1 + \sup_{|Q-G| \leq \rho} \{f(x; Q)\}.$$

Condition 2. *Smoothness.* The support of $f(x, \theta)$ is independent of θ and $f(x, \theta)$ is three times differentiable with respect to θ in Θ. Further, $f(x, \theta)$ and its derivatives with respect to θ, $f'(x, \theta)$, $f''(x, \theta)$ and $f'''(x, \theta)$, are jointly continuous in x and θ.

Condition 3. *Strong identifiability.* 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 $a_j = b_j = c_j = 0$, $j = 1, 2$.

Note that Condition 3 is stronger than an ordinary identifiability condition. In addition to $f(x, \theta)$ itself, $f'(x, \theta)$ and $f''(x, \theta)$ are also identifiable. The strong identifiability condition is first proposed by Chen (1995). It is related to the non-singularity of the Fisher Information in regular models, and to the Bhattacharyya inequality (Lehmann, 1983, pp. 129). Chen (1995) proves that location and scale kernels are strongly identifiable if $f(\pm\infty, \theta) = f'(\pm\infty, \theta) = 0$. Using the same argument, we can show that all regular exponential families are strongly identifiable.

Condition 4. *Uniform boundedness.* There exists an integrable function g and some $\delta > 0$ such that $|Y_i(\theta)|^{4+\delta} \leq g(X_i)$, $|Y_i'(\theta)|^3 \leq g(X_i)$, $|Y_i''(\theta)|^3 \leq g(X_i)$ and $|Y_i'''(\theta)|^3 \leq g(X_i)$ for all θ.

Condition 5. *Tightness.* For $j = 1, 2$, the processes

$$n^{-1/2}\sum_{i=1}^{n} Y_{ij}(\theta), \quad n^{-1/2}\sum_{i=1}^{n} Y_i'(\theta), \quad n^{-1/2}\sum_{i=1}^{n} Y_i''(\theta), \quad n^{-1/2}\sum_{i=1}^{n} Y_i'''(\theta)$$

are tight.

The tightness condition ensures the weak convergence of the processes. It is noted that the tightness of $n^{-1/2}\sum Y_{ij}(\theta)$, $n^{-1/2}\sum_{i=1}^{n} Y_i'(\theta)$, and $n^{-1/2}\sum_{i=1}^{n} Y_i''(\theta)$ are implied by Condition 4. To see this, consider

$$E\{n^{-1/2}\sum Y_{ij}(\theta_1) - n^{-1/2}\sum Y_{ij}(\theta_2)\}^2 = E\{Y_{1j}(\theta_1) - Y_{1j}(\theta_2)\}^2 \le Eg^{2/3}(X_1)|\theta_1 - \theta_2|^2.$$

The tightness then follows from Theorem 12.3 of Billingsley (1968, p95). The same argument also applies to Y_i', Y_i'' and Y_i'''.

B. Proofs

Proof of Lemma 1. From (6), it follows that $\tilde{l}_n^{(k)}(\hat{G}) \ge \tilde{l}_n^{(k)}(G_0^{(k)}) = A > -\infty$. Also, from Dacunha-Castelle and Gassiate (1999), $l_n(\bar{G}) - l_n(G_0) = O_p(1)$ where $\bar{G}$ is the ordinary MLE of G. If follows that

$$A \le \tilde{l}_n^{(k)}(\hat{G}) - C_k\sum \log\hat{\pi}_j = l_n(\hat{G}) - l_n(G_0) \le l_n(\bar{G}) - l_n(G_0) = O_p(1).$$

Thus $\sum \log\hat{\pi}_j = O_p(1)$ and the lemma follows. □

Proof of Theorem 1. In light of Lemma 1, we can assume that the weight on each support point of $\hat{G}$ is at least ϵ. Thus it is immediate that conclusions (a) and (b) imply (c) and (d). The claims (a) and (b) are equivalent to the consistency of $\hat{G}$. Since the space of the mixing distributions G is compact, the consistency of the modified MLE of G follows the classical proof given by Wald (1945). □

Proof of Lemma 2 We start with the inequality

$$R_{1n} \le 2\sum_{i=1}^{n}\delta_i - \sum_{i=1}^{n}\delta_i^2 + (2/3)\sum_{i=1}^{n}\delta_i^3, \tag{16}$$

where δ_i is defined in (9). Thus, from (10), we have

$$\sum_{i=1}^{n}\delta_i = (\hat{\pi} - \pi_0)\sum_{i=1}^{n}\Delta_i + \hat{\pi}\sum_{i=1}^{n}\frac{f(X_i;\hat{G}_1) - f(X_i,\theta_{01})}{f(X_i;G_0)} + (1-\hat{\pi})\sum_{i=1}^{n}\frac{f(X_i;\hat{G}_2) - f(X_i,\theta_{02})}{f(X_i;G_0)}. \tag{17}$$

Note that

$$\sum_{i=1}^{n} \frac{f(X_i;\hat{G}_1) - f(X_i,\theta_{01})}{f(X_i;G_0)} = \int \left\{ \sum_{i=1}^{n} \frac{f(X_i,\theta) - f(X_i,\theta_{01})}{f(X_i;G_0)} \right\} d\hat{G}_1(\theta).$$

and a Taylor expansion of the integrand of this expression gives

$$(\theta - \theta_{01}) \sum_{i=1}^{n} Y_i'(\theta_{01}) + \frac{(\theta-\theta_{01})^2}{2} \sum_{i=1}^{n} Y_i''(\theta_{01}) + \frac{(\theta-\theta_{01})^3}{6} \sum_{i=1}^{n} Y_i'''(\eta_1),$$

where η_1 is between θ and θ_{01}. As a consequence, we have

$$\begin{aligned} \sum_{i=1}^{n} \frac{f(X_i,\hat{G}_1) - f(X_i,\theta_{01})}{f(X_i;G_0)} &= m_{11} \sum_{i=1}^{n} Y_i'(\theta_{01}) + \frac{m_{21}}{2} \sum_{i=1}^{n} Y_i''(\theta_{01}) \\ &\quad + \frac{1}{6} \int (\theta - \theta_{01})^3 \sum_{i=1}^{n} Y_i'''(\eta_1) d\hat{G}_1(\theta). \end{aligned}$$

A similar expression holds for the term with $\hat{G}_2$. Equation (17) becomes

$$\begin{aligned} \sum_{i=1}^{n} \delta_i &= \sum_{i=1}^{n} [(\hat{\pi} - \pi_0)\Delta_i + \hat{\pi} m_{11} Y_i'(\theta_{01}) + (1-\hat{\pi}) m_{21} Y_i'(\theta_{02}) \\ &\quad + \hat{\pi} \frac{m_{12}}{2} Y_i''(\theta_{01}) + (1-\hat{\pi}) \frac{m_{22}}{2} Y_i''(\theta_{02}) \Big] + \epsilon_n, \end{aligned}$$

where

$$\epsilon_n = \frac{1}{6} \sum_{i=1}^{n} \left[\int \hat{\pi} (\theta - \theta_{01})^3 Y_i'''(\eta_1) d\hat{G}_1(\theta) + \int (1-\hat{\pi})(\theta - \theta_{02})^3 Y_i'''(\eta_2) d\hat{G}_2(\theta) \right].$$

The inequality (16) can now be written as

$$R_{1n} \leq L_n - Q_n + C_n + \epsilon_n,$$

where L_n and Q_n are defined in (11), and

$$\begin{aligned} C_n &= \frac{2}{3} \sum_{i=1}^{n} \Big[(\hat{\pi} - \pi_0)\Delta_i + \hat{\pi} \{ m_{11} Y_i'(\theta_{01}) + m_{21} Y_i''(\theta_{01}) \} \\ &\quad + \frac{1-\hat{\pi}}{2} \{ m_{12} Y_i'(\theta_{02}) \} + m_{22} Y_i''(\theta_{02}) \} \Big]^3. \end{aligned}$$

Note that only the leading terms in Q_n and C_n are included since the remainders are negligible and result in no higher order than the remainder in the linear term L_n.

Furthermore, we shall show that the cubic term as well as the remainder ϵ_n can be controlled by Q_n.

First, consider ϵ_n. By Condition 5,

$$\sup_{\theta\in\Theta} |n^{-1/2}\sum_{i=1}^{n} Y_i'''(\theta)| = O_p(1).$$

Therefore $|\epsilon_n| \le n^{1/2}(\|m_{31}\| + \|m_{32}\|)O_p(1)$, where

$$\|m_{ij}\| = \int |\theta - \theta_{0j}|^i d\hat{G}_j(\theta).$$

Hence $\|m_{2j}\| = m_{2j}$ and $|m_{ij}| \le \|m_{ij}\|$ in general. Since $\|m_{ij}\| = o_p(1)$ which is implied by Theorem 1,

$$\|m_{3j}\| = m_{2j}o_p(1),$$

we obtain

$$|\epsilon_n| = n^{1/2}(m_{21} + m_{22})o_p(1) \le \{1 + n(m_{21}^2 + m_{22}^2)\}o_p(1).$$

Under strong identifiability condition, Q_n is a positive-definite quadratic form in m_{11}, m_{12}, m_{21} and m_{22}. Hence,

$$\epsilon_n = o_p(Q_n) + o_p(1).$$

Second, we consider the cubic term. Since,

$$\frac{C_n}{Q_n} \le O_p(1)\sum_{i,j}\|m_{ij}\| = o_p(1).$$

it follows that $C_n = o_p(Q_n)$.

To sum up, we have established an upper bound of R_{1n} as follows:

$$R_{1n} \le L_n - Q_n(1 + o_p(1)) + o_p(1).$$

Following the analysis and notation in Section 3.2, we have

$$R_{1n} \le \mathbf{b}_1^\tau B_{11}^{-1}\mathbf{b}_1 + \sup_{\mathbf{t}_2}\{2\tilde{\mathbf{b}}_2^\tau\mathbf{t}_2 - \mathbf{t}_2^\tau\tilde{B}_{22}\mathbf{t}_2\} + o_p(1). \tag{18}$$

The proof will be complete if it is shown that the upper bound in (18) is achievable. To see this, let $\mathbf{t}_1^* = B_{11}^{-1}\mathbf{b}_1$ and $\mathbf{t}_2^*$ be such that

$$\sup_{\mathbf{t}_2}\{2\tilde{\mathbf{b}}_2^\tau\mathbf{t}_2 - \mathbf{t}_2^\tau\tilde{B}_{22}\mathbf{t}_2\} = 2\tilde{\mathbf{b}}_2^\tau\mathbf{t}_2^* - \mathbf{t}_2^{*\tau}\tilde{B}_{22}\mathbf{t}_2^*.$$

We suppose that $k \geq k^*$ and consider the mixing distribution $G^* \in \mathcal{M}_k$ whose support points and weights θ_j^*, and π_j^*, $j = 1, ..., k$, are chosen so that $\pi_j^* = \pi_j^{(0)}$, $\mathbf{t}_1(G^*) = \mathbf{t}_1^*$ and $\mathbf{t}_2(G^*) = \mathbf{t}_2^*$. Such a solution exists since, with $k \geq k^*$ there is sufficient latitude to fit both second order moments $\hat{m}_{j2}$ since, for n large enough we can simultaneously partition the mass near θ_{01} and θ_{02} while maintaining weights that allow the penalty term to converge to the penalty under the null.

¿From the fact that both $\mathbf{t}_1^*$ and $\mathbf{t}_2^*$ have an order of $n^{-1/2}$, we also conclude $|G^* - G_0| = O_p(n^{-1/4})$, as Θ is assumed compact. We show next that R_{1n} reaches the upper bound with this choice of G.

Let $R_{1n}^* = 2\{l_n(G^*) - l_n(G_0)\} = 2\sum \log(1 + \delta_i^*)$, where the definition of δ_i^* is similar to that of δ_i. Consider the Taylor expansion

$$2\sum \log(1 + \delta_i^*) = 2\sum_{i=1}^{n} \delta_i^* - \sum_{i=1}^{n} \delta_i^{*2}(1 + \gamma_i)^{-2},$$

where $|\gamma_i| < |\delta_i^*|$. Note that for a constant c

$$|\delta_i^*| \leq c|G^* - G_0| \max_{1 \leq i \leq n} \{\sup_{\theta \in \Theta} |Y_i(\theta)|\}.$$

By Conditional 4, $|Y_i(\theta)|^{4+\delta} \leq g(X_i)$ and $E\{g(X_i)\}$ is finite, implying that

$$\max_{1 \leq i \leq n} \{\sup_{\theta \in \Theta} |Y_i(\theta)|\} = o_p(n^{1/4}).$$

It follows that $\max\{|\gamma_i|\} = o_p(1)$ uniformly in θ. Therefore,

$$R_{1n}^* = 2\sum_{i=1}^{n} \delta_i^* - \sum_{i=1}^{n} \delta_i^{*2}\{1 + o_p(1)\}.$$

Applying the argument that led to (18) yields the require result

$$R_{1n}^* = \mathbf{b}_1^\tau B_{11}^{-1} \mathbf{b}_1 + \sup_{\mathbf{t}_2} \{2\tilde{\mathbf{b}}_2^\tau \mathbf{t}_2 - \mathbf{t}_2^\tau \tilde{B}_{22} \mathbf{t}_2\} + o_p(1).$$

Finally we note that

$$R_{1n} - R_{1n}^* = \tilde{l}_n^{(k)}(\hat{G}) - \tilde{l}_n^{(k)}(G^*) - C_k\{\sum_{j=1}^{k} \log \hat{\pi}_j - \sum_{j=1}^{k} \log \pi_j^*\}.$$

It follows immediately that $\tilde{l}_n^{(k)}(\hat{G}) - \tilde{l}_n^{(k)}(G^*) \geq 0$. At last, note that when $k \geq k^*$, $\hat{G}$ is consistent under the null hypothesis so that $\sum \log \hat{\pi}_j - \sum \log \pi_j^* + o_p(1) \leq 0$. Thus, $R_{1n} \geq R_{1n}^* + o_p(1)$ and the proof is complete. □

Proof of Theorem 2. The proof starts with the equation (15). Without loss of generality, we assume that the covariance matrix of $n^{-1/2}\tilde{\mathbf{b}}_2$ has the following standard form:

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

Let $(Z_1, Z_2)^\tau$ be bivariate normal with mean zero and covariance matrix Σ, and let $W_1 = Z_1$ and $W_2 = (1-\rho^2)^{-1/2}(Z_2 - \rho Z_1)$, so that W_1 and W_2 are independent $N(0,1)$ variates.

As $n \to \infty$, it can be seen that

$$\begin{aligned} R_n &\to \sup_{\xi_1 \geq 0, \xi_2 \geq 0} \{2Z_1\xi_1 + 2Z_2\xi_2 - \xi_1^2 - 2\rho\xi_1\xi_2 - \xi_2^2\} \\ &= W_1^2 + W_2^2 - \inf_{\xi_1 \geq 0, \xi_2 \geq 0} \{[W_1 - (\xi_1 + \rho\xi_2)]^2 + [W_2 - (1-\rho^2)^{1/2}\xi_2]^2\} \\ &= W_1^2 + W_2^2 - \inf_{(\eta_1,\eta_2) \in \mathcal{S}} \{(W_1 - \eta_1)^2 + (W_2 - \eta_2)^2\}. \end{aligned}$$

The restrictions $\xi_1 \geq 0$ and $\xi_2 \geq 0$ are transformed into $\mathcal{S} = \{(\eta_1, \eta_2) : \eta_2 \geq 0, \eta_1 \geq \rho\eta_2/\sqrt{1-\rho^2}\}$, as illustrated in Figure 1.

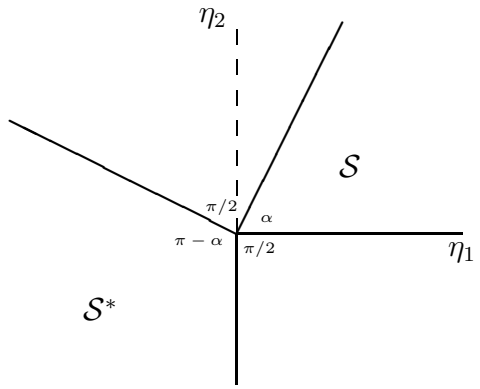

Figure 1: The cone $\mathcal{S} = \{(\eta_1, \eta_2) : \eta_2 \geq 0, \eta_1 \geq \rho\eta_2/\sqrt{1-\rho^2}\}$ and the dual cone S^*.

Since the norm of (W_1, W_2) and its direction vector are statistically independent, it follows that:

a) Given that $(W_1, W_2) \in \mathcal{S}$, R_n has a χ^2_2 distribution;

b) Given that (W_1, W_2) is in the dual cone ($\mathcal{S}^*$ in Figure 1), $R_n = 0$;

c) Given that (W_1, W_2) is in the remaining region, R_n has a χ^2_1 distribution.

This completes the proof. □

REFERENCES

Bickel, P. and Chernoff, H. (1993). Asymptotic distribution of the likelihood ratio statistic in 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. New York: Wiley.

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

Chen, H. and Chen, J. (2001). Large sample distribution of the likelihood ratio test for normal mixtures. *Can. J. Statist.* **29** 201-216.

Chen, H., Chen, J. and Kalbfleisch J.D. (2001). A modified likelihood ratio test for homogeneity in finite mixture models. *J. R. Statist. Soc.* B, **63**, 19-29.

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

Chen, J. and Cheng, P. (1997). A new approach to test the number of components in finite mixture models. *Can. J. Statist.*, **25**, 389-400.

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. R. 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. Plann. 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.

Friedlander Y, Leitersdorf E. (1995). Segregation analysis of plasma lipoprotein(a) levels in pedigrees with molecularly defined familial hypercholesterolemia. *Genetic Epidemiology.* **12** 129-143.

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. *Proc. Berk. Conf. in Honor of J. Neyman and J. Kiefer.* **2** 789-806. Edited by L. LeCam and R.A. Olshen.

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** 807-810. Edited by L. LeCam and R.A. Olshen.

Heiba, I.M., Elston, R.C. and Klein, B.D. (1995). Evidence for a major gene for cortical cataract. *Investigative Ophthalmology and Visual Science*, **36**, 227-235.

Henna J. (1985). On estimating of the number of constituents of a finite mixture of continuous distributions. *Ann. Inst. Statist. Math.*, **37**, 235-240.

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

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

Leroux, B. (1992). Consistent estimation of a mixture distributions. *Ann. Statist.*, **20**, 1350-1360.

Liang, K. Y. and Rathouz, P.J. (1999). Hypothesis testing under mixture models: application to genetic linkage analysis. *Biometrics*, **55**, 65-74.

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

Lindsay, B.G. (1995). *Mixture Models: Theory, Geometry and Applications*. NSF-CBMS Regional Conference Series in Probability and Statistics, Hayward: Institute for Mathematical Statistics.

Lindsay, B.G. and Roeder, K. (1997). Moment-based oscillation properties of mixture models. *Ann. Statist.*, **25**, 378-386.

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: 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.

Ott, J. (1999). *Analysis of Human Genetic Linkage, Third Edition*. Baltimore: The Johns Hopkins University Press.

Qin, J. (1998). Inferences for case-control and semiparametric two-sample density ratio models. *Biometrika*, **85**, 619-630.

Roeder, K. (1994). A graphical technique for determining the number of components in a mixture of normals. *J. Amer. Statist. Assoc.*, **89**, 487-500.

Schork, N.J., Allison, D.B. and Thiel, B. (1996). Mixture distributions in human genetics research. *Statistical Methods in Medical Research*, **5**, 155-178.

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

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.*, **30**, 185-191.

Table 1: Normal Mixture Model Specifications

Model	support			probability			Model	support			probability		
N 1	-1.5	1.5		0.50	0.50		N 4	-2.5	2.5		0.50	0.50	
N 2	-1.5	1.5		0.375	0.625		N 5	-2.5	2.5		0.375	0.625	
N 3	-1.5	1.5		0.25	0.75		N 6	-2.5	2.5		0.25	0.75	
A 1A	0	-1.5	1.5	0.20	0.4	0.4	A 4A	0	-2.0	2.0	0.20	0.4	0.4
A 2A	0	-1.5	1.5	0.30	0.35	0.35	A 5A	0	-2.0	2.0	0.30	0.35	0.35
A 3A	0	-1.5	1.5	0.20	0.30	0.50	A 6A	0	-2.0	2.0	0.20	0.30	0.50
A 1B	-3	-1.5	1.5	0.10	0.45	0.45	A 4B	-4	-2.0	2.0	0.20	0.40	0.40
A 2B	-3	-1.5	1.5	0.20	0.40	0.40	A 5B	-4	-2.0	2.0	0.30	0.35	0.35
A 3B	-3	-1.5	1.5	0.10	0.30	0.60	A 6B	-4	-2.0	2.0	0.20	0.30	0.50

Table 2: The nominal and power of the test

Model	$n = 100$ Rejection rates			$n = 200$ Rejection rates		
N 1	0.0890	0.0490	0.0265	0.1000	0.0550	0.0320
N 2	0.0890	0.0530	0.0295	0.1060	0.0515	0.0305
N 3	0.0930	0.0435	0.0215	0.0975	0.0575	0.0290
N 4	0.1085	0.0540	0.0275	0.1190	0.0685	0.0350
N 5	0.1100	0.0600	0.0335	0.1225	0.0655	0.0345
N 6	0.1000	0.0510	0.0295	0.1080	0.0640	0.0345
A 1A	0.1240	0.0790	0.0465	0.1715	0.1075	0.0670
A 2A	0.1855	0.1085	0.0645	0.2760	0.1800	0.1125
A 3A	0.1270	0.0760	0.0410	0.1840	0.1210	0.0730
A 4A	0.3825	0.2750	0.1925	0.5675	0.4650	0.3735
A 5A	0.5505	0.4395	0.3415	0.8095	0.7390	0.6410
A 6A	0.3690	0.2720	0.1765	0.5675	0.4590	0.3680
A 1B	0.3215	0.2455	0.1145	0.4740	0.3835	0.3005
A 2B	0.5005	0.3950	0.2970	0.7290	0.6420	0.5565
A 3B	0.2960	0.2090	0.1510	0.4100	0.3200	0.2541
A 4B	0.9280	0.8845	0.8375	0.9950	0.9895	0.9830
A 5B	0.9680	0.9470	0.9110	0.9995	0.9990	0.9985
A 6B	0.9315	0.9005	0.8550	0.9965	0.9935	0.9915