Statistica Sinica **14**(2004), 1165-1177

LIKELIHOOD RATIO TEST FOR HOMOGENEITY IN NORMAL MIXTURES IN THE PRESENCE OF A STRUCTURAL PARAMETER

Yong Song Qin[1,2] and Bruce Smith[1]

Dalhousie University and Guangxi Normal University

Abstract: This paper investigates the asymptotic properties of the likelihood ratio statistic for testing homogeneity in normal mixture models in the presence of a structural parameter. The asymptotic null distributions of the ordinary likelihood ratio statistic and the modified likelihood ratio statistic are the same, having the probability density function(pdf) $(1/2)g_1(x)+(1/2)g_2(x)$ where $g_1(x)$ and $g_2(x)$ are the probability density functions of $\chi^2_{(1)}$ and $\chi^2_{(2)}$, respectively. For the ordinary likelihood ratio statistic, we employ the assumption that $\min\{\alpha_1, \alpha_2\} \geq \epsilon$ for some $1/2 > \epsilon > 0$, where α_1 and α_2 are the coefficients of the mixtures.

Key words and phrases: Mixture model, likelihood ratio test, modified likelihood ratio test, asymptotic distribution.

1. Introduction

Tests for homogeneity in finite mixture models have been investigated by many researchers. Titterington, Smith and Makov (1985), McLachlan and Basford (1988) and Lindsay (1995) provide extensive background discussions. For further related work, refer to Lindsay (1989), Leroux (1992), Chernoff and Lander (1995), Cheng and Traylor (1995), Chen and Chen (2001, 2003), Chen, Chen and Kalbfleisch (2001) and Chen (1998).

An example of much current interest arises in genetics, in the case that a major gene is thought to underly a phenotypic trait. Until the underlying gene has been identified, the genotype is unobservable. Consider the case of a major gene with two possible alleles (say A and a), and suppose that the genotype aa occurs in a population with frequency α. Suppose that A is dominant over a, that the genotype aa confers an average effect size θ_1, and that one or more copies of A confers an average effect size θ_2. Assuming that the phenotypic trait is normally distributed with mean θ_1 or θ_2 (depending on the genotype) and variance σ^2, the phenotypic distribution for a randomly sampled individual follows a two component mixture distribution $\alpha_1\phi(x;\theta_1,\sigma^2) + \alpha_2\phi(x;\theta_2,\sigma^2)$. A test for the presence of a major gene is a test that the phenotypic distribution is a single normal distribution. What is of primary interest are the mean effect sizes θ_1, θ_2, and so the hypothesis of primary interest is that the effect sizes are equal, in which case the underlying gene contributes nothing to the phenotypic trait.

The asymptotic null distribution of the likelihood ratio test is very complex and difficult to use in practice. Chen et al. (2001) propose a modified likelihood ratio test for homogeneity in the finite mixture models. They prove that the modified likelihood ratio test enjoys the simple χ^2-type null limiting distribution and is asymptotically most powerful under the local alternative models when there are no structural parameters. Chen and Chen (2003) investigate the large sample behavior of the likelihood ratio test for testing homogeneity in the normal mixture models with an unknown structural parameter. The asymptotic null distribution of the likelihood ratio test is even more complex than that in the case that no structural parameter is involved.

In this paper, we continue to study the likelihood ratio test for homogeneity in normal mixtures in the presence of a structural parameter. At first, we show that the ordinary likelihood ratio test has the simple χ^2-type null limiting distribution under the assumption $\min\{\alpha_1, \alpha_2\} \geq \epsilon$ for some $1/2 > \epsilon > 0$, where α_1 and α_2 are the coefficients of the mixtures. Second, without the above assumption, we show that the modified likelihood ratio test has the same null limiting distribution. The interesting thing is that here the value of α_1 or α_2 contributes to the degrees of freedom of the χ^2-type null limiting distribution, while that in Chen et al. (2001) does not when $\min\{\alpha_1, \alpha_2\} \geq \epsilon$ and α_1, α_2 are known . This new fact also complicates the analysis in the present case.

Let $X_1, \cdots, X_n$ be a random sample of size n from a mixture population with the probability density function (pdf)

$$f(x; \alpha_1, \alpha_2, \theta_1, \theta_2, \sigma^2) = \alpha_1 \phi(x; \theta_1, \sigma^2) + \alpha_2 \phi(x; \theta_2, \sigma^2), \tag{1.1}$$

where $\alpha_1, \alpha_2 \geq 0, \alpha_1 + \alpha_2 = 1$, and $\phi(x; \theta, \sigma^2)$ is the pdf of $N(\theta, \sigma^2)$.

We wish to test

$$H_0: \ N(\theta_0, \sigma_0^2), \tag{1.2}$$

versus the full model (1.1).

To avoid nonidentifiability, we take $\alpha_1 \geq \alpha_2$. As noted by Hartigan (1985), a bounded assumption on the mean parameters is also necessary $-$ we take $|\theta_i| \leq M < \infty$ for $i = 1, 2$.

The likelihood function is

$$\ell_n(\alpha_1, \alpha_2, \theta_1, \theta_2, \sigma^2) = \sum_{i=1}^{n} \log\{\alpha_1 \phi(X_i; \theta_1, \sigma^2) + \alpha_2 \phi(X_i; \theta_2, \sigma^2)\}. \tag{1.3}$$

The paper is organized as follows. Section 2 presents the asymptotic theory of the ordinary likelihood ratio test with the assumption that $\min\{\alpha_1, \alpha_2\} \geq \epsilon$

under the null hypothesis. In Section 3, we give the null limiting distribution of the modified likelihood ratio test. Some simulation results are given in Section 4. The proofs of Lemmas used in Sections 2 and 3 are in Appendix.

2. Ordinary Likelihood Ratio Test

We derive the asymptotic distribution of the ordinary likelihood ratio test when $H_0 : N(\theta_0, \sigma_0^2)$ is the true distribution. Let $\hat{\alpha}_1, \hat{\alpha}_2, \hat{\theta}_1, \hat{\theta}_2$ and $\hat{\sigma}^2$ maximize $\ell_n(\alpha_1, \alpha_2, \theta_1, \theta_2, \sigma^2)$ over the parameter space $0 < \epsilon \le \alpha_2 \le 1/2$, $|\theta_j| \le M$, $j = 1, 2$, $0 < \sigma^2 < \infty$, and let $\hat{\theta}_0$ and $\hat{\sigma}_0^2$ maximize $\ell_n(1, 0, \theta_0, \theta_0, \sigma_0^2)$ over the parameter space $-\infty < \theta_0 < \infty$, $0 < \sigma_0^2 < \infty$. Then the ordinary likelihood ratio test is to reject the null hypothesis H_0 if

$$R_n = 2\{\ell_n(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\theta}_1, \hat{\theta}_2, \hat{\sigma}^2) - \ell_n(\frac{1}{2}, \frac{1}{2}, \hat{\theta}_0, \hat{\theta}_0, \hat{\sigma}_0^2)\} \tag{2.1}$$

is suitably large. The asymptotic null distribution of R_n is used to determine a critical value of the test or a P-value.

We assume that $\theta_0 = 0$, and $\sigma_0^2 = 1$. This presents no loss of generality, as if $Z_i = (X_i - \theta_0)/\sigma_0$, $\hat{\mu}_1 = (\hat{\theta}_1 - \theta_0)/\sigma_0$, $\hat{\mu}_2 = (\hat{\theta}_2 - \theta_0)/\sigma_0$, $\hat{\rho} = \hat{\sigma}/\sigma_0$, $\hat{\mu}_0 = (\hat{\theta}_0 - \theta_0)/\sigma_0$, $\hat{\rho}_0 = \hat{\sigma}_0/\sigma_0$, then

$$\begin{aligned}
R_n &= 2\{\ell_n(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\theta}_1, \hat{\theta}_2, \hat{\sigma}^2) - \ell_n(\frac{1}{2}, \frac{1}{2}, \hat{\theta}_0, \hat{\theta}_0, \hat{\sigma}_0^2)\} \\
&\hat{=} 2\{\ell_n(X_1, \cdots, X_n; \hat{\alpha}_1, \hat{\alpha}_2, \hat{\theta}_1, \hat{\theta}_2, \hat{\sigma}^2) - \ell_n(X_1, \cdots, X_n; \frac{1}{2}, \frac{1}{2}, \hat{\theta}_0, \hat{\theta}_0, \hat{\sigma}_0^2)\} \\
&= 2\{\ell_n(Z_1, \cdots, Z_n; \hat{\alpha}_1, \hat{\alpha}_2, \hat{\mu}_1, \hat{\mu}_2, \hat{\rho}^2) - \ell_n(Z_1, \cdots, Z_n; \frac{1}{2}, \frac{1}{2}, \hat{\mu}_0, \hat{\mu}_0, \hat{\rho}_0^2)\}.
\end{aligned}$$

The following result is given in Chen and Chen (2003).

Lemma 1. *Under the null hypothesis $N(0, 1)$, there exist constants $0 < \epsilon < \Delta < \infty$ such that $\lim_{n\to\infty} P(\epsilon \le \hat{\sigma}^2 \le \Delta) = 1$.*

Referring to Lemma 1, we can restrict σ within a closed interval $[\epsilon_1, M_1]$ for $\epsilon_1 > 0$ and $M_1 > 0$. Using Lemma 1 and $\min\{\alpha_1, \alpha_2\} \ge \epsilon$, the following result can be proved in a similar fashion to Theorem 1 in Chen et al. (2000).

Lemma 2. *Under the null hypothesis $N(0, 1)$, $\hat{\theta}_1 = o_p(1)$, $\hat{\theta}_2 = o_p(1)$, $\hat{\sigma}^2 - 1 = o_p(1)$.*

To find the asymptotic distribution of R_n defined in (2.1), it is convenient to partition it into two parts:

$$\begin{aligned}
R_n &= 2\{\ell_n(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\theta}_1, \hat{\theta}_2, \hat{\sigma}^2) - \ell_n(\frac{1}{2}, \frac{1}{2}, 0, 0, 1)\} \\
&\quad + 2\{\ell_n(\frac{1}{2}, \frac{1}{2}, 0, 0, 1) - \ell_n(\frac{1}{2}, \frac{1}{2}, \hat{\theta}_0, \hat{\theta}_0, \hat{\sigma}_0^2)\} \\
&\hat{=} R_{1n} + R_{2n}.
\end{aligned} \tag{2.2}$$

Note that $-R_{2n}$ is an ordinary likelihood ratio(no mixture involved) and hence an approximation is immediate:

$$-R_{2n} = \frac{(\sum_{i=1}^{n} X_i)^2}{\sum_{i=1}^{n} X_i^2} + \frac{\{\sum_{i=1}^{n}(-1+X_i^2)\}^2}{\sum_{i=1}^{n}(-1+X_i^2)^2} + o_p(1). \tag{2.3}$$

The main task is thus to analyze R_{1n}. Write

$$R_{1n} = 2\sum_{i=1}^{n}\log(1+\delta_i), \tag{2.4}$$

$$\delta_i = \hat{\alpha}_1 \frac{\phi(X_i;\hat{\theta}_1,\hat{\sigma}^2) - \phi(X_i;0,1)}{\phi(X_i;0,1)} + \hat{\alpha}_2 \frac{\phi(X_i;\hat{\theta}_2,\hat{\sigma}^2) - \phi(X_i;0,1)}{\phi(X_i;0,1)}. \tag{2.5}$$

Put $\hat{m}_k = \hat{\alpha}_1\hat{\theta}_1^k + \hat{\alpha}_2\hat{\theta}_2^k,\ k \geq 1,$

$$Y_i(\theta,\sigma^2) = \begin{cases} \dfrac{\phi(X_i;\theta,\sigma^2) - \phi(X_i;0,\sigma^2)}{\theta\phi(X_i;0,1)}, & \theta \neq 0 \\ \sigma^{-3}X_i\exp\{-\frac{1}{2}X_i^2(\sigma^{-2}-1)\}, & \theta = 0, \end{cases}$$

$$U_i(\sigma^2) = \begin{cases} \dfrac{\phi(X_i;0,\sigma^2) - \phi(X_i;0,1)}{(\sigma^2-1)\phi(X_i;0,1)}, & \sigma^2 \neq 1 \\ \frac{1}{2}(-1+X_i^2), & \sigma^2 = 1. \end{cases}$$

Then $\delta_i = \hat{\alpha}_1\hat{\theta}_1 Y_i(\hat{\theta}_1,\hat{\sigma}^2) + \hat{\alpha}_2\hat{\theta}_2 Y_i(\hat{\theta}_2,\hat{\sigma}^2) + (\hat{\sigma}^2-1)U_i(\hat{\sigma}^2)$. The method used in Chen and Chen (2003) is used to deal with δ_i in the following . Using the Taylor expansion of $Y_i(\hat{\theta}_j,\hat{\sigma}^2)$, $j=1,2$, and $U_i(\hat{\sigma}^2)$, it follows that

$$\begin{aligned} \delta_i = {} & \hat{m}_1 Y_i(0,1) + (\hat{\sigma}^2 - 1 + \hat{m}_2)Y_i'(0,1) + \frac{1}{2}\hat{m}_3 Y_i''(0,1) \\ & + \frac{1}{6}\{3(\hat{\sigma}^2-1)^2 + \hat{m}_4 + 6(\hat{\sigma}^2-1)\hat{m}_2\}Y_i'''(0,1) + \hat{\epsilon}_{in}, \end{aligned} \tag{2.6}$$

where $Y_i'(0,1)$ is the first partial derivative of $Y_i(\theta,\sigma^2)$ with respect to θ at $\theta=0$ and $\sigma^2=1$, while $Y_i''(0,1)$ and $Y_i'''(0,1)$ are the associated second and third partial derivatives with respect to θ. It can be shown that $Y_i(0,1) = X_i,\ Y_i'(0,1) = U_i(1) = (X_i^2-1)/2,\ Y_i''(0,1) = (X_i^3 - 3X_i)/3$, and $Y_i'''(0,1) = 2U_i'(1) = (X_i^4 - 6X_i^2 + 3)/4$. Put $Y_i = Y_i(0,1),\ Y_i' = Y_i'(0,1),\ Y_i'' = Y_i''(0,1),\ Y_i''' = Y_i'''(0,1)$. As the processes resulting from $Y_i(\theta,\sigma^2),\ U_i(\sigma^2)$ and their derivatives are tight, it can be shown, as in Chen and Chen (2003), that the sum of the remainders $\hat{\epsilon}_n = \sum_{i=1}^{n}\hat{\epsilon}_{in}$ satisfies

$$\begin{aligned} \hat{\epsilon}_n = {} & n^{1/2}(\hat{\sigma}^2-1)^3 O_p(1) + n(\hat{m}_1^2 + \hat{m}_3^2)o_p(1) \\ & + n^{1/2}(|\hat{m}_5| + \hat{m}_6)O_p(1) + o_p(1). \end{aligned} \tag{2.7}$$

Also,

$$\begin{aligned}\delta_i &\hat{=} S_1\cdot X_i+S_2\cdot\frac{1}{2}(-1+X_i^2)+S_3\cdot\frac{1}{6}(-3X_i+X_i^3)+S_5\cdot\frac{1}{24}(3-6X_i^2+X_i^4)+\hat{\epsilon}_{in}\\ &\hat{=} S_1\cdot X_i+S_2\cdot\frac{1}{2}(-1+X_i^2)+S_3\cdot\frac{1}{6}(-3X_i+X_i^3)+S_4\cdot\frac{1}{24}(3-6X_i^2+X_i^4)+\epsilon_{in}, \quad (2.8)\end{aligned}$$

where $S_1=\hat{m}_1$, $S_2=\hat{\sigma}^2-1+\hat{m}_2$, $S_3=\hat{m}_3$, $S_5=3(\hat{\sigma}^2-1)^2+\hat{m}_4+6(\hat{\sigma}^2-1)\hat{m}_2$, $S_4=\hat{m}_4-3\hat{m}_2^2$, $\epsilon_{in}=\hat{\epsilon}_{in}+3\{\hat{m}_2+(\hat{\sigma^2}-1)\}^2\times(1/24)(3-6X_i^2+X_i^4)$. Here we use the fact that $S_5=S_4+3\{\hat{m}_2+(\hat{\sigma^2}-1)\}^2$.

Let $|S|=\sum_{i=1}^4|S_i|$. We have the following result concerning the convergence rates of the MLE's.

Lemma 3. If $\hat{\alpha}_2\ge\epsilon$, then under $H_0, N(0,1)$, $\hat{\theta}_1=O_p(|S|^{1/4})$, $\hat{\theta}_2=O_p(|S|^{1/4})$, $\hat{\sigma}^2-1=O_p(|S|^{1/2})$.

The proof of Lemma 3 is given in Appendix.

By (2.7), Lemma 2, Lemma 3 and $n^{1/2}|S|\le 1+nS^2$,

$$\sum_{i=1}^n\epsilon_{in}=o_p(1)+nS^2\cdot o_p(1). \quad (2.9)$$

Since $nS^2\le 4n\sum_{j=1}^4 S_j^2$, using similar reasoning as in Chen and Chen (2003), $nS^2\cdot o_p(1)$ can be absorbed into the quadratic sum $\sum_{i=1}^n\{S_1^2Y_i^2+S_2^2(Y_i')^2+(1/4)S_3^2(Y_i'')^2+(1/36)S_4^2(Y_i''')^2\}$. That is $nS^2\cdot o_p(1)=o_p(1)\cdot\sum_{i=1}^n\{S_1^2Y_i^2+S_2^2(Y_i')^2+(1/4)S_3^2(Y_i'')^2+(1/36)S_4^2(Y_i''')^2\}$. Similarly, it can be shown that

$$\begin{aligned}R_{1n}&\le 2\sum_{i=1}^n\delta_i-\sum_{i=1}^n\delta_i^2+\frac{2}{3}\sum_{i=1}^n\delta_i^3\\ &=2\sum_{i=1}^n(S_1Y_i+S_2Y_i'+\frac{1}{2}S_3Y_i''+\frac{1}{6}S_4Y_i''')-\sum_{i=1}^n\{S_1^2Y_i^2+S_2^2(Y_i')^2\\ &\quad+\frac{1}{4}S_3^2(Y_i'')^2+\frac{1}{36}S_4^2(Y_i''')^2\}\{1+o_p(1)\}+o_p(1). \quad (2.10)\end{aligned}$$

In above expression, the cubic and mixed terms are absorbed into the quadratic terms, again reasoning as in Chen and Chen (2003). Let $x^+=xI_{[x\ge0]}$ where I is the indicator function. At the end of Appendix, we prove that $-S_4\ge0$, $a.s.$ It follows that, when

$$S_1=\frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n Y_i^2},\quad S_2=\frac{\sum_{i=1}^n Y_i'}{\sum_{i=1}^n (Y_i')^2},\quad S_3=2\frac{\sum_{i=1}^n Y_i''}{\sum_{i=1}^n (Y_i'')^2},\quad S_4=-6\frac{(-\sum_{i=1}^n Y_i''')^+}{\sum_{i=1}^n (Y_i''')^2}, \quad (2.11)$$

$$R_{1n} \leq \frac{(\sum_{i=1}^n Y_i)^2}{\sum_{i=1}^n Y_i^2} + \frac{(\sum_{i=1}^n Y_i')^2}{\sum_{i=1}^n (Y_i')^2} + \frac{(\sum_{i=1}^n Y_i'')^2}{\sum_{i=1}^n (Y_i'')^2} + \frac{\{(-\sum_{i=1}^n Y_i''')^+\}^2}{\sum_{i=1}^n (Y_i''')^2} + o_p(1). \tag{2.12}$$

From (2.3) and (2.12),

$$R_n \leq \frac{(\sum_{i=1}^n Y_i'')^2}{\sum_{i=1}^n (Y_i'')^2} + \frac{\{(-\sum_{i=1}^n Y_i''')^+\}^2}{\sum_{i=1}^n (Y_i''')^2} + o_p(1). \tag{2.13}$$

We are going to get a lower bound for R_n. Let S_1, S_2, S_3 and S_4 be defined by (2.11).

Lemma 4. *There exist $\tilde\alpha_1, \tilde\alpha_2, \tilde\theta_1, \tilde\theta_2$ and $\tilde\sigma^2$ such that $\tilde\alpha_1 - \tilde\alpha_2 = o(1), a.s.,$ and that*

$$\tilde\alpha_1\tilde\theta_1 + \tilde\alpha_2\tilde\theta_2 = S_1, \tag{2.14}$$
$$\tilde\alpha_1\tilde\theta_1^2 + \tilde\alpha_2\tilde\theta_2^2 + (\tilde{\sigma^2} - 1) = S_2, \tag{2.15}$$
$$\tilde\alpha_1\tilde\theta_1^3 + \tilde\alpha_2\tilde\theta_2^3 = S_3, \tag{2.16}$$
$$\tilde\alpha_1\tilde\theta_1^4 + \tilde\alpha_2\tilde\theta_2^4 - 3(\tilde\alpha_1\tilde\theta_1^2 + \tilde\alpha_2\tilde\theta_2^2)^2 = S_4. \tag{2.17}$$

The proof of Lemma 4 is given in Appendix.

From Lemma 3, under H_0, $\tilde\theta_1 = O_p(|S|^{1/4})$, $\tilde\theta_2 = O_p(|S|^{1/4})$, $\tilde\sigma^2 - 1 = O_p(|S|^{1/2})$. Consider the Taylor expansion

$$\begin{aligned}\tilde R_{1n} &\hat= 2\{\ell_n(\tilde\alpha_1, \tilde\alpha_2, \tilde\theta_1, \tilde\theta_2, \tilde\sigma^2) - \ell_n(\frac{1}{2}, \frac{1}{2}, 0, 0, 1)\} \\ &= 2\sum_{i=1}^n \tilde\delta_i - \sum_{i=1}^n \tilde\delta_i^2 (1 + \tilde\eta_i)^{-2},\end{aligned}$$

where $|\tilde\eta_i| < |\tilde\delta_i|$, and

$$\tilde\delta_i = \tilde\alpha_1 \frac{\phi(X_i; \tilde\theta_1, \tilde\sigma^2) - \phi(X_i; 0, 1)}{\phi(X_i; 0, 1)} + \tilde\alpha_2 \frac{\phi(X_i; \tilde\theta_2, \tilde\sigma^2) - \phi(X_i; 0, 1)}{\phi(X_i; 0, 1)}.$$

Similar to Chen et al. (2000), it can be shown that $\tilde\delta_i = o_p(1)$. So $\tilde R_{1n} = 2\sum_{i=1}^n \tilde\delta_i - \sum_{i=1}^n \tilde\delta_i^2\{1 + o_p(1)\}$. Thus, with the similar derivation as (2.10),

$$\tilde R_{1n} = \frac{(\sum_{i=1}^n Y_i)^2}{\sum_{i=1}^n Y_i^2} + \frac{(\sum_{i=1}^n Y_i')^2}{\sum_{i=1}^n (Y_i')^2} + \frac{(\sum_{i=1}^n Y_i'')^2}{\sum_{i=1}^n (Y_i'')^2} + \frac{\{(-\sum_{i=1}^n Y_i''')^+\}^2}{\sum_{i=1}^n (Y_i''')^2} + o_p(1), \tag{2.18}$$

$$R_n \geq \frac{(\sum_{i=1}^{n} Y_i'')^2}{\sum_{i=1}^{n}(Y_i'')^2} + \frac{\{(-\sum_{i=1}^{n} Y_i''')^+\}^2}{\sum_{i=1}^{n}(Y_i''')^2} + o_p(1). \tag{2.19}$$

From (2.13) and (2.19), it follows that

$$\frac{(\sum_{i=1}^{n} Y_i'')^2}{\sum_{i=1}^{n}(Y_i'')^2} + \frac{\{(-\sum_{i=1}^{n} Y_i''')^+\}^2}{\sum_{i=1}^{n}(Y_i''')^2} + o_p(1) \leq R_n \leq \frac{(\sum_{i=1}^{n} Y_i'')^2}{\sum_{i=1}^{n}(Y_i'')^2} + \frac{\{(-\sum_{i=1}^{n} Y_i''')^+\}^2}{\sum_{i=1}^{n}(Y_i''')^2} + o_p(1).$$

By the Central Limit Theorem, $n^{-1/2}\sum_{i=1}^{n}(Y_i, Y_i', Y_i'', Y_i''')^T \xrightarrow{d} N(0, \Sigma)$, where

$$\Sigma = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{2}{3} & 0 \\ 0 & 0 & 0 & \frac{3}{2} \end{pmatrix},$$

and $E(Y_i^2, (Y_i')^2, (Y_i'')^2, (Y_i''')^2) = (1, 1/2, 2/3, 3/2)$.

Theorem 1. *Assume that* $\min\{\alpha_1, \alpha_2\} \geq \epsilon$ *for some* $0 < \epsilon < 1/2$. *Then under the null hypothesis* H_0, $R_n \xrightarrow{d} R$ *as* $n \to \infty$, *where* R *has the probability density function* $g(x) = (1/2)g_1(x) + (1/2)g_2(x)$, *where* $g_1(x)$ *and* $g_2(x)$ *are the probability density functions of* $\chi^2_{(1)}$ *and* $\chi^2_{(2)}$, *respectively.*

Remark 1. If α_1 and α_2 are known and $\alpha_1 \neq \alpha_2$, then under null hypothesis, $R_n \xrightarrow{d} \chi^2_{(1)}$ as $n \to \infty$. If α_1 and α_2 are known and $\alpha_1 = \alpha_2$, then under null hypothesis, $R_n \xrightarrow{d} (1/2)\chi^2_{(0)} + (1/2)\chi^2_{(1)}$ as $n \to \infty$.

Remark 1 can be verified, with a similar derivations as the proof of Theorem 1, by letting $S_1 \cdot X_i + S_2 \cdot (1/2)(-1 + X_i^2) + S_3 \cdot (1/6)(-3X_i + X_i^3)$ as the main order term of δ_i if $\alpha_1 \neq \alpha_2$, and $S_1 \cdot X_i + S_2 \cdot (1/2)(-1 + X_i^2) + S_4 \cdot (1/24)(3 - 6X_i^2 + X_i^4)$ as the main order term of δ_i if $\alpha_1 = \alpha_2$, respectively.

3. The Modified Likelihood Ratio Test

To avoid nonidentifiability of this model, we adopt a restriction $\alpha_1 \geq \alpha_2$. As noted by Hartigan(1985), a bounded assumption on the mean parameters is necessary. We therefore assume that $|\theta_i| \leq M < \infty$ for $i = 1, 2$.

Similar to Chen et al. (2001), define the modified likelihood function

$$p\ell_n(\alpha_1, \alpha_2, \theta_1, \theta_2, \sigma^2) = \sum_{i=1}^{n} \log\{\alpha_1\phi(X_i; \theta_1, \sigma^2) + \alpha_2\phi(X_i; \theta_2, \sigma^2)\} + C\log(2\alpha_2), \tag{3.1}$$

where $C > 0$ is a constant used to control the level of modification. The modified likelihood function is often called a penalized likelihood function, referring to the penalty when α_2 is close to 0.

We derive the asymptotic distribution of the modified likelihood ratio test when $H_0 : N(\theta_0, \sigma_0^2)$ is the true distribution. Let $\hat{\alpha}_1, \hat{\alpha}_2, \hat{\theta}_1, \hat{\theta}_2$ and $\hat{\sigma}^2$ maximize $p\ell_n(\alpha_1, \alpha_2, \theta_1, \theta_2, \sigma^2)$ over the parameter space Ω: $|\theta_j| \le M, j = 1, 2,\ 0 < \sigma^2 < \infty$, and let $\hat{\theta}_0$ and $\hat{\sigma}_0^2$ maximize $\ell_n(1, 0, \theta_0, \theta_0, \sigma_0^2)$ over the parameter space Ω_0: $0 < \sigma_0^2 < \infty$. Then the modified likelihood ratio test is to reject the null hypothesis H_0 if

$$M_n = 2\{p\ell_n(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\theta}_1, \hat{\theta}_2, \hat{\sigma}^2) - p\ell_n(\frac{1}{2}, \frac{1}{2}, \hat{\theta}_0, \hat{\theta}_0, \hat{\sigma}_0^2)\} \tag{3.2}$$

is large. The asymptotic null distribution of M_n is used to determine a critical value of the test or a P-value.

Using the fact that the ordinary likelihood ratio test over the parameter spaces Ω and Ω_0 is bounded in probability (refer to Theorem 2 in Chen and Chen (2003)), the following result can be shown (refer to the proof of Lemma 1 in Chen et al. (2000)).

Lemma 5. *Under the null hypothesis,* $\log(2\hat{\alpha}_2) = O_p(1)$.

From this result, we can restrict $\hat{\alpha}_2$ to the closed interval $[\delta_1, 1/2]$ for a positive constant δ_1. Thus, similar to the proof of Theorem 1, We have the following result.

Theorem 2. *Under the null hypothesis H_0, $M_n \xrightarrow{d} R$ as $n \to \infty$, where R has the probability density function $g(x) = g_1(x)/2 + g_2(x)/2$, where $g_1(x)$ and $g_2(x)$ are the probability density functions of $\chi^2_{(1)}$ and $\chi^2_{(2)}$, respectively.*

Choice of C. According to Chen et al. (2001), an appropriate choice of C is $C = \log(M)$ where M is the value such that $|\theta_i| \le M$ for $i = 1, 2$.

4. Simulation Results

A simulation experiment was conducted with the null distribution $N(0, 1)$. For each simulation, eleven significance levels(denoted as SL, and indicated in the tables), and two sample sizes, $n = 100$ and $n = 160$, were considered. We took $\epsilon = 0.01$, $M = 100$ and $C = \log(M)$ in the simulations.

For the specified SL, the theoretical critical values (denoted as TCV) are for the random varible with pdf $g_1(x)/2 + g_2(x)/2$, where $g_1(x)$ and $g_2(x)$ are the probability density functions of $\chi^2_{(1)}$ and $\chi^2_{(2)}$, respectively. The inverse cumulative distribution was approximated using the Splus procedure “uniroot”.

Simulated critical values (denoted as SCV) were simulated using 5,000 Monte Carlo trials for the ordinary likelihood ratio test R_n. A second set of simulated critical values (denoted as MSCV) were simulated using $5,000$ Monte Carlo trials for the modified likelihood ratio test M_n. These are reported in Table 1. We can see that, even at these relatively small sample sizes, the simulated critical values are quite close to the theoretical values.

Table 1. Comparsion of theoretical and simulated critical values.

SL	90%	80%	70%	60%	50%
TCV	0.0492855	0.1689514	0.345352	0.5729519	0.8670388
SCV($n = 100$)	0.04666001	0.1737475	0.380945	0.6809468	1.128781
SCV ($n = 160$)	0.04470917	0.1686727	0.3692717	0.6697447	1.071691
MSCV ($n = 100$)	0.03417222	0.1133761	0.2375973	0.4028322	0.6151661
MSCV($n = 160$)	0.02714179	0.1113902	0.244251	0.4241742	0.6580295

SL	40%	30%	20%	10%	5%	1%
TCV	1.2474160	1.7582148	2.5016278	3.8078255	5.1383808	8.2732342
SCV ($n = 100$)	1.651546	2.342976	3.256938	5.106185	6.63853	10.73876
SCV ($n = 160$)	1.594131	2.260242	3.170333	4.963923	6.434895	9.917313
MSCV ($n = 100$)	0.9063974	1.343508	2.046623	3.348158	4.981956	8.885286
MSCV ($n = 160$)	0.9826923	1.426866	2.056794	3.274991	4.775348	8.484485

Table 2 reports simulated rejection rates (denoted as RR when using R_n, and MRR when using M_n) under H_0, using $5,000$ Monte Carlo trials for R_n and M_n. The results suggest that in general, the level of the ordinary likelihood ratio test is somewhat elevated over the nominal level, while for the modified LRT, the level is slightly reduced at nominal levels of 0.05 and 0.1.

Table 2. Rejection rates under H_0.

SL	90%	80%	70%	60%	50%
RR(%,$n = 100$)	89.64	80.28	71.62	63.5	55.64
RR(%, $n = 160$)	89.5	79.98	71.26	63.22	55.06
MRR(%,$n = 100$)	87.8	75.14	63.46	51.8	41.08
MRR(%, $n = 160$)	87.08	75.3	64.2	53.18	43.34

SL	40%	30%	20%	10%	5%	1%
RR(%, $n = 100$)	47.56	38.24	28	16.18	9.86	2.48
RR(%, $n = 160$)	46.34	37.3	27.14	15.48	9.26	2.2
MRR(%, $n = 100$)	31.74	23.66	15.52	8.36	4.72	1.36
MRR(%, $n = 160$)	33.52	23.96	15.52	7.62	4.34	1.08

Additional simulations were carried out to investigate the conclusions of Remark 1. For each of $\alpha_2 = 0.5, 0.25$ and 0.1, 5,000 simulation batches of size 1,000 were generated, and the associated values of R_n calculated. Quantile-quantile plots were created for each of the target distributions $\chi^2_{(1)}$ and $0.5\chi^2_{(0)} + 0.5\chi^2_{(1)}$, and are displayed in Figure 1. Observed quantiles are on the ordinate, and theoretical quantiles are on the abscissa. The plots are supportive of Remark 1 in that when $\alpha_1 = \alpha_2 = 0.5$, the fit to the $0.5\chi^2_{(0)}+0.5\chi^2_{(1)}$ mixture is quite good, apart from a slightly long right tail, while when $\alpha_2 \neq 0.5$, the $\chi^2_{(1)}$ distribution is appropriate.

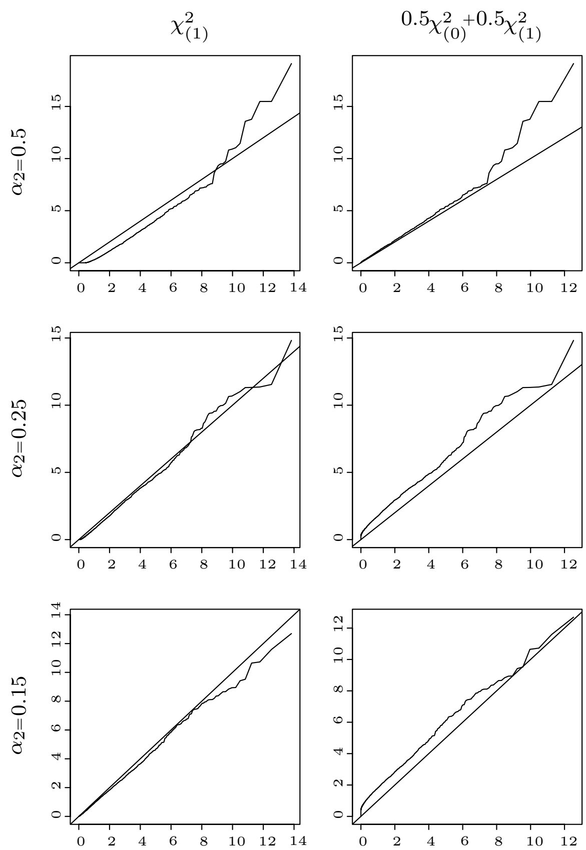

Figure 1. $\chi^2_{(1)}$ (left panels) and $0.5\chi^2_{(0)} + 0.5\chi^2_{(1)}$ (right panels) QQ plots for the distribution of R_n when $\alpha_2 = 0.5$ (top row), $\alpha_2 = 0.25$ (middle row), and $\alpha_2 = 0.15$ (bottom row).

Acknowledgements

This work was partially supported by MITACS via the project Statistical Modeling and Analysis of Complex Traits in Human Populations, and by an

NSERC grant to Bruce Smith. The authors are very grateful to the referees and an associate editor for their valuable comments and constructive suggestions

Appendix. Proof of Lemmas

Proof of Lemma 3. By definition,

$$\hat{\alpha}_1\hat{\theta}_1 + \hat{\alpha}_2\hat{\theta}_2 = S_1, \tag{A.1}$$
$$\hat{\alpha}_1\hat{\theta}_1^2 + \hat{\alpha}_2\hat{\theta}_2^2 + (\hat{\sigma}^2 - 1) = S_2, \tag{A.2}$$
$$\hat{\alpha}_1\hat{\theta}_1^3 + \hat{\alpha}_2\hat{\theta}_2^3 = S_3, \tag{A.3}$$
$$\hat{\alpha}_1\hat{\theta}_1^4 + \hat{\alpha}_2\hat{\theta}_2^4 - 3(\hat{\alpha}_1\hat{\theta}_1^2 + \hat{\alpha}_2\hat{\theta}_2^2)^2 = S_4. \tag{A.4}$$

From (A.1) and (A.2), $\hat{\alpha}_1\hat{\alpha}_2(\hat{\theta}_1 - \hat{\theta}_2)^2 = S_2 - S_1^2 - (\hat{\sigma}^2 - 1)$. It follows that

$$|\hat{\theta}_1 - \hat{\theta}_2| = \frac{1}{\sqrt{\hat{\alpha}_1\hat{\alpha}_2}}\sqrt{S_2 - S_1^2 - (\hat{\sigma}^2 - 1)}.$$

Without loss of generality, assume that

$$\hat{\theta}_1 - \hat{\theta}_2 = \frac{1}{\sqrt{\hat{\alpha}_1\hat{\alpha}_2}}\sqrt{S_2 - S_1^2 - (\hat{\sigma}^2 - 1)}. \tag{A.5}$$

By (A.1) and (A.5),

$$\begin{cases} \hat{\theta}_1 = S_1 + \sqrt{\frac{\hat{\alpha}_2}{\hat{\alpha}_1}}\sqrt{S_2 - S_1^2 - (\hat{\sigma}^2 - 1)}, \\ \hat{\theta}_2 = S_1 - \sqrt{\frac{\hat{\alpha}_1}{\hat{\alpha}_2}}\sqrt{S_2 - S_1^2 - (\hat{\sigma}^2 - 1)}. \end{cases} \tag{A.6}$$

From (A.3) and (A.6),

$$-\hat{\alpha}_2\sqrt{\frac{\hat{\alpha}_2}{\hat{\alpha}_1}}(\hat{\sigma}^2 - 1)\sqrt{S_2 - S_1^2 - (\hat{\sigma}^2 - 1)} + \hat{\alpha}_1\sqrt{\frac{\hat{\alpha}_1}{\hat{\alpha}_2}}(\hat{\sigma}^2 - 1)\sqrt{S_2 - S_1^2 - (\hat{\sigma}^2 - 1)} = O_p(|S|),$$

i.e.,

$$(\hat{\alpha}_1 - \hat{\alpha}_2)(\hat{\sigma}^2 - 1)\sqrt{S_2 - S_1^2 - (\hat{\sigma}^2 - 1)} = O_p(|S|).$$

Therefore if $\hat{\alpha}_1 - \hat{\alpha}_2 > \epsilon_0$ for some small ϵ_0, then by Lemma 2, $(\hat{\sigma}^2 - 1)\{S_2 - S_1^2 - (\hat{\sigma}^2 - 1)\} = o_p(|S|)$. It follows that

$$(\hat{\sigma}^2 - 1)^2 = o_p(|S|). \tag{A.7}$$

From (A.2) and (A.7),

$$\hat{\theta}_1^2 = o_p(|S|^{1/2}),\ \hat{\theta}_2^2 = o_p(|S|^{1/2}). \tag{A.8}$$

By (A.7) and (A.8), we have Lemma 3. Similarly, if $\hat{\alpha}_1 - \hat{\alpha}_2 < \epsilon_0$ for some small ϵ_0, by (A.1),(A.2) and (A.4) (or refer to the (A.16) in the proof of Lemma

4), we can show that $(\hat{\sigma}^2-1)\{S_2-S_1^2-(\hat{\sigma}^2-1)\}=O_p(|S|)$. It follows that $(\hat{\sigma}^2-1)^2=O_p(|S|)$. Thus, it can be shown that $\hat{\theta}_1^2=O_p(|S|^{1/2})$, $\hat{\theta}_2^2=O_p(|S|^{1/2})$. Then Lemma 3 follows.

Proof of Lemma 4. Consider the following four equations of $\alpha_1(\alpha_2=1-\alpha_1),\theta_1,\theta_2$ and σ^2-1 :

$$\alpha_1\theta_1+\alpha_2\theta_2=S_1, \tag{A.9}$$

$$\alpha_1\theta_1^2+\alpha_2\theta_2^2+(\sigma^2-1)=S_2, \tag{A.10}$$

$$\alpha_1\theta_1^3+\alpha_2\theta_2^3=S_3, \tag{A.11}$$

$$\alpha_1\theta_1^4+\alpha_2\theta_2^4-3(\alpha_1\theta_1^2+\alpha_2\theta_2^2)^2=S_4. \tag{A.12}$$

Define $y=S_2-S_1^2-(\sigma^2-1)$. Then (A.9) and (A.10) imply one of

$$\theta_1=S_1+\sqrt{\alpha_2/\alpha_1}\sqrt{y},\ \theta_2=S_1-\sqrt{\alpha_1/\alpha_2}\sqrt{y}, \tag{A.13}$$

$$\theta_1=S_1-\sqrt{\alpha_2/\alpha_1}\sqrt{y},\ \theta_2=S_1+\sqrt{\alpha_1/\alpha_2}\sqrt{y}, \tag{A.14}$$

where $y=\alpha_1\alpha_2(\theta_1-\theta_2)^2$. (A.11) and (A.13) imply

$$S_1^3+3S_1y-\frac{\alpha_1-\alpha_2}{\sqrt{\alpha_1\alpha_2}}y^{3/2}=S_3. \tag{A.15}$$

(A.12) and (A.13) imply

$$(\frac{1}{\alpha_1\alpha_2}-6)y^2-4S_1\frac{\alpha_1-\alpha_2}{\sqrt{\alpha_1\alpha_2}}y^{3/2}-2S_1^4=S_4. \tag{A.16}$$

From (A.15) and (A.16), using the fact that $(\alpha_1\alpha_2)^{-1}=((\alpha_1-\alpha_2)/\sqrt{\alpha_1\alpha_2})^2+4$ and letting $(\alpha_1-\alpha_2)/\sqrt{\alpha_1\alpha_2}=\gamma y^{1/2}$ by introducing a new parameter γ, it can be shown that

$$2y^3+3S_1^2y^2+(2S_1S_3+S_4)y-(S_3-S_1^3)^2=0. \tag{A.17}$$

Notice that $\sqrt{n}S_j=O_p(1)$ for $j=1,2,3,4$. We can find a positive root $\tilde{y}$ which satisfies (A.17). It is also easy to see that $\tilde{y}=o_p(1)$. Using $\tilde{y}$ to replace y in (A.15), we can get a solution for $(\alpha_1-\alpha_2)/\sqrt{\alpha_1\alpha_2}$, and thus a solution for α_1 or α_2. Here we assume that $3S_1\tilde{y}-(S_3-S_1^3)\ge 0$ to ensure $\alpha_1\ge\alpha_2$. If this is not true, then we need to use (A.14) instead of (A.13) to get a similar expression as (A.15). Let $\beta=(\alpha_1-\alpha_2)/\sqrt{\alpha_1\alpha_2}$. Then combining with (A.13) and (A.14), we can show that there are solutions $\tilde{\alpha}_1,\tilde{\alpha}_2,\tilde{\theta}_1,\tilde{\theta}_2$ and $\tilde{\sigma}^2$ for equations (A.9) through (A.12). We are going to prove that $\tilde{\beta}=(\tilde{\alpha}_1-\tilde{\alpha}_2)/\sqrt{\tilde{\alpha}_1\tilde{\alpha}_2}=\beta_n=o(1)$, a.s. If $\beta_n=o(1)$ is not true, there exists a constant $\delta>0$ and a subsequence n_k of n such that $\beta_{n_k}>\delta$. Without loss of generality, we assume that $\beta_n>\delta$

for all n. Then from (A.15), $\tilde{y} = O_p(n^{-1/3})$. Together with (A.17), one has $n^{5/6}S_4\tilde{y} = o_p(1)$. Let $z = |n^{1/2}S_4|n^{1/3}\tilde{y}$. Then $z = o_p(1)$, $\tilde{y} = n^{-1/3}|n^{1/2}S_4|^{-1}z$ From (A.15) one has $|n^{1/2}S_4|^{3/2}S_3 = o_p(n^{-1/2})$, which is a contradiction. So the probability of the event that $\tilde{\beta} = o(1)$ is not true is 0. Therefore, $\tilde{\beta} = o(1), a.s.$, which is equivalent to $\tilde{\alpha}_1 - \tilde{\alpha}_2 = o(1), a.s.$ Thus we have Lemma 4.

With the same reasoning, we can see that $\hat{\alpha}_1 - \hat{\alpha}_2 = o(1), a.s.$ Therefore, $\hat{\alpha}_2 \geq 1/3$ when n is large enough. Since $\hat{\alpha}_1\hat{\theta}_1^4 + \hat{\alpha}_2\hat{\theta}_2^4 - 3(\hat{\alpha}_1\hat{\theta}_1^2 + \hat{\alpha}_2\hat{\theta}_2^2)^2 = \hat{\alpha}_1\hat{\alpha}_2(\hat{\theta}_1^2 - \hat{\theta}_2^2)^2 - 2(\hat{\alpha}_1\hat{\theta}_1^2 + \hat{\alpha}_2\hat{\theta}_2^2)^2$, it is less than or equal to 0 as $\hat{\alpha}_2 \geq 1/3$. Hence $S_4 \leq 0, a.s.$

References

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

Chen, H. and Chen, J. (2001). Large sample distribution of the likelihood ratio test for normal mixtures. *Statist. Probab. Lett.* **52**, 125-133.

Chen, H. and Chen, J. (2003). Tests for homogeneity in normal mixture in the presence of a structural parameter. *Statist. Sinica.* **13**, 351-365.

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

Cheng, R. C. H. and Traylor, L. (1995). Non-regular maximum likelihood problems(with discussion). *J. Roy. Statist. Soc. Ser. 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. Inference* **43**, 19-40.

Hartigan, J. A. (1985). A failure of likelihood asymptotics for normal mixtures. In *Proc. Berkeley Conf. In Honor of J. Neyman and J. Kiefer* **2** (Edited by L. Lecam and R. A. Olshen), 807-810. Wadsworth, Belmont.

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

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

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

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

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

Department of Mathematics and Statistics, Dalhousie University, Halifax, NS B3H 3J5, Canada.

E-mail: ysqin@mailbox.gxnu.edu.cn

Department of Mathematics, Guangxi Normal University, Guilin, Guangxi 541004, China.

E-mail: bsmith@mathstat.dal.ca

(Received October 2002; accepted January 2004)