Statistica Sinica **13**(2003), 351-365

TESTS FOR HOMOGENEITY IN NORMAL MIXTURES IN THE PRESENCE OF A STRUCTURAL PARAMETER

Hanfeng Chen and Jiahua Chen

Bowling Green State University and University of Waterloo

Abstract: Often a question arises as to whether observed data are a sample from a homogeneous population or from a heterogeneous population. If in particular, one wants to test for a single normal distribution versus a mixture of two normal distributions, classic asymptotic results do not apply since the model does not satisfy regularity conditions. This paper investigates the large sample behavior of the likelihood ratio statistic for testing homogeneity in the normal mixture in location parameters with an unknown structural parameter. It is proved that the asymptotic null distribution of the likelihood ratio statistic is the maximum of a χ^2_2-variable and the supremum of the square of a truncated Gaussian process with mean 0 and variance 1. This result exposes the unusual large sample behavior of the likelihood function under the null distribution. The correlation structure of the process involved in the limiting distribution is presented explicitly. From the large sample study, it is also found that even though the structural parameter is not part of the mixing distribution, the convergence rate of its maximum likelihood estimate is $n^{-1/4}$ rather than $n^{-1/2}$, while the mixing distribution has a convergence rate $n^{-1/8}$ rather than $n^{-1/4}$. This is in sharp contrast to ordinary semi-parametric models and to mixture models without a structural parameter.

Key words and phrases: Asymptotic distribution, finite mixture, Gaussian process, genetic analysis, likelihood ratio, non-regular model, semi-parametric model.

1. Introduction

Let $X_1, \ldots, X_n$ be a random sample from a mixture population $(1-\alpha)$ $N(\theta_1, \sigma^2) + \alpha N(\theta_2, \sigma^2)$ with the probability density function (pdf)

$$(1-\alpha)\sigma^{-1}\phi\{(x-\theta_1)/\sigma\} + \alpha\sigma^{-1}\phi\{(x-\theta_2)/\sigma\}, \tag{1}$$

where $\phi(\cdot)$ is the pdf of the standard normal $N(0,1)$. Alternatively, write (1) as $\int \sigma^{-1}\phi\{(x-u)/\sigma\}dG(u)$, with the mixing distribution

$$G(u) = (1-\alpha)I(u \geq \theta_1) + \alpha I(u \geq \theta_2). \tag{2}$$

Suppose we wish to test $H_0 : \alpha(1-\alpha) = 0$ or $\theta_1 = \theta_2$ versus the full model (1), i.e., to test $N(\theta, \sigma^2)$ versus $(1-\alpha)N(\theta_1, \sigma^2) + \alpha N(\theta_2, \sigma^2)$.

There has been increasing interest in finite mixture models in recent years. The large sample behavior of the likelihood ratio test (LRT) for homogeneity in the finite mixture model is a long-standing mystery. Hartigan (1985) showed that the LRT statistic tends to infinity with probability one if the mean parameters are unbounded. The divergence behavior of the LRT is further detailed by Bickel and Chernoff (1993). On the other hand, Ghosh and Sen (1985) gave the first version of the asymptotic distributions of the LRT statistic when the mean parameters are bounded. However, in addition to the boundedness, they imposed a separation condition, i.e., $|\theta_1 - \theta_2| > \epsilon$ for some given $\epsilon > 0$. The separation condition is obviously unsatisfactory. There have been many attempts made to remove the separation condition. Lemdani and Pons (1999) used a reparameterization approach to investigate the testing problem when one of the mean parameters is known and their study showed that there is no obvious way to remove the separation condition. Dacunha-Castelle and Gassiat (1999) developed a general reparameterization method for the testing problem in locally conic models. In the meantime, Chen and Chen (2001a and b) took a different approach, i.e., the so-called sandwich method, to attack the problem without the separation condition.

In this paper, a structural parameter is included in the mixture model to bring it closer to the reality. A test for homogeneity is considered when both the two mean parameters are assumed unknown, and we remove the separation condition.

We start our study in Section 2 with the single mean parameter mixture model: one of the mean parameters θ_1 and θ_2 in (1) is assumed known. While study of the single mean parameter mixtures has its virtue, the purpose of the section is to demonstrate the main ideas behind our approach to the general mixture model (1). The asymptotic distribution of the LRT for homogeneity under the model (1) is investigated in Section 3. It is shown that the asymptotic null distribution of the LRT statistic is the maximum of a χ^2_2-variable and the supremum of the square of a truncated Gaussian process with mean 0 and variance 1.

Throughout the paper, without loss of generality, let the null underlying distribution be $N(0,1)$. For convenience, we write $X_n(t) = O_p(a_n)$ or $= o_p(a_n)$ if $\sup_{t\in T} |X_n(t)/a_n| = O_p(1)$ or $\sup_{t\in T} |X_n(t)/a_n| = o_p(1)$, where T is a suitably specified index set and a_n is a sequence of constants or random variables.

To save space, only the main ideas behind the analyses and results are presented, and all technical details are in Chen and Chen (2001c), available from the web site http://www.stats.uwaterloo.ca/Stats_Dept/techreports/node8.html.

2. Single Mean Parameter Mixtures

In (1), take θ_1 to be 0, $\theta_2 = \theta$ to be unknown. In addition assume that $|\theta| \leq M$. Based on the observations X_i, we wish to use the LRT to test the null hypothesis $H_0 : N(0, \sigma^2)$ versus $H_a : (1-\alpha)N(0,\sigma^2) + \alpha N(\theta, \sigma^2)$. The log-likelihood function of α, θ and σ is

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

Let $\hat{\sigma}_0^2 = n^{-1}\sum X_i^2$ be the MLE of σ^2 under the null hypothesis. Let

$$\begin{aligned} r_n(\alpha,\theta,\sigma) &= 2\{l_n(\alpha,\theta,\sigma) - l_n(0,0,\hat{\sigma}_0)\} \\ &= 2\sum_{i=1}^{n} \log\Big\{1 + \alpha\Big(\exp\Big\{\frac{2X_i\theta - \theta^2}{2\sigma^2}\Big\} - 1\Big)\Big\} \\ &\quad -n\log\sigma^2 - \frac{\sum X_i^2}{\sigma^2} + n\Big(1 + \log\frac{\sum X_i^2}{n}\Big). \end{aligned} \tag{3}$$

Let $\hat{\alpha}$, $\hat{\theta}$ and $\hat{\sigma}^2$ be the MLEs for α, θ and σ^2 under the full model. Then the LRT is to reject the null hypothesis when $R_n = r_n(\hat{\alpha}, \hat{\theta}, \hat{\sigma})$ is large.

2.1. Large sample behavior of the MLE's

We first show that under the null hypothesis $\hat{\sigma}^2$ is bounded away from zero and infinity with probability approaching 1.

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

Proof. Consider $r_n(\alpha,\theta,\sigma)$ defined by (3). Note that when $2x\theta - \theta^2 \ge 0$,

$$1 + \alpha\Big[\exp\{\frac{2x\theta - \theta^2}{2\sigma^2}\} - 1\Big] \le \exp\Big\{\frac{2x\theta - \theta^2}{2\sigma^2}\Big\}.$$

We thus have the inequality

$$r_n(\alpha,\theta,\sigma) \le \frac{\sum[(2X_i\theta - \theta^2)^+ - X_i^2]}{\sigma^2} - n\log\sigma^2 + n\Big(1 + \log\frac{\sum X_i^2}{n}\Big),$$

where $t^+ = tI(t > 0)$ denotes the positive part of t. Since $(2X_i\theta - \theta^2)^+ - X_i^2$ is equal to either $-(\theta - X_i)^2$ or $-X_i^2$, we see that $r_n(\alpha,\theta,\sigma) \le -n\log\sigma^2 + n\{1 + \log(\sum X_i^2/n)\}$. Since $\log(n^{-1}\sum X_i^2) \to 0$ almost surely, the function $r_n(\alpha,\theta,\sigma) < 0$ for all $\sigma^2 > \Delta$ with probability approaching 1 for some large constant Δ. That is, $\lim P(\hat{\sigma}^2 > \Delta) = 0$ for some constant Δ.

Next we show that $\hat{\sigma}^2$ is also bounded away from zero asymptotically. By the Uniform Strong Law of Large Numbers (see Rubin (1956)), $n^{-1}\sum\{X_i^2 - (2\theta X_i - \theta^2)^+\} \to S(\theta) = E\{X^2 - (2\theta X - \theta^2)^+\}$, almost surely and uniformly in $|\theta| \le M$. Since $S(\theta)$ is continuous and positive, the minimum value of $S(\theta)$ is positive, say equal to $2q$ for some $q > 0$. Then with probability approaching one uniformly in α, θ and σ, $r_n(\alpha,\theta,\sigma) \le -nq\sigma^{-2} - n\log\sigma^2 + n\{1+\log(\sum X_i^2/n)\}$. Let $\epsilon > 0$ be small enough such that $-q/\epsilon - \log\epsilon + 1 < 0$. It follows that with probability approaching 1 uniformly, the function $r_n(\alpha,\theta,\sigma) < 0$ if $\sigma^2 < \epsilon$, implying $P(\hat\sigma^2 \ge \epsilon) \to 1$ as $n \to \infty$. The proof is completed.

By Lemma 1, the parametric space of interest can be reduced to a compact one by restricting σ^2 within the interval $[\epsilon, \Delta]$.

Lemma 2. *Under the null distribution $N(0,1)$, as $n \to \infty$, $\hat\alpha\hat\theta \to 0$ and $\hat\sigma^2 \to 1$, in probability.*

Proof. Here is an outline of the proof. For a detailed proof see Chen and Chen (2001c). As remarked, we only need to consider $\epsilon \le \sigma^2 \le \Delta$ for some constants $0 < \epsilon < 1 < \Delta < \infty$. Let $\mathcal{G} = \{G(\cdot) : G(u) = (1-\alpha)I(u \ge 0) + \alpha I(u \ge \theta), |\theta| \le M, 0 \le \alpha \le 1\}$. The space of the distribution functions metrized by taking the Lévy distance between two distribution functions is compact because $M < \infty$. (Note that the Lévy distance convergence is equivalent to weak convergence of distribution functions.) So the product space $\Omega = \{\omega = (\sigma^2, G) : \sigma^2 \in [\epsilon,\Delta], G \in \mathcal{G}\}$ is compact. Moreover, for $\omega = (\sigma^2, G) \in \Omega$, put $f(x;\omega) = \sigma^{-1}\int \phi\{(x-u)/\sigma\}dG(u)$. Then the parameter $\omega \in \Omega$ is identifiable, i.e., for any $\omega_i \in \Omega$, $i = 1,2$, $f(x;\omega_1) = f(x;\omega_2)$ for all x, implies $\omega_1 = \omega_2$. With compactness and identifiability, Wald (1949)'s argument leads to consistency of the MLE $\hat\omega = (\hat\sigma^2, \hat G)$ for $\omega = (\sigma^2, G)$ under the null model. Furthermore, it is implied that the MLEs of the moments $\int u^k dG(u) = \alpha\theta^k$ are consistent. Since $\int u^k dG(u) = 0$ under $N(0,1)$, the lemma is proved.

By Lemma 2, σ^2 can be technically restricted to any neighborhood of $\sigma^2 = 1$, say $[1-\delta, 1+\delta]$ for a small $\delta > 0$. This restriction will be used to ensure the tightness of some processes later.

Here we would like to point out that Lemma 2 does not imply anything about the rate of convergence. We also like to remark that Lemma 2 does not say that $\hat\alpha$ or $\hat\theta$ is consistent. In fact, $\hat\alpha$ and $\hat\theta$ are inconsistent under the null model. See Chernoff and Lander (1995)'s discussion of the binomial mixture model that is also applicable to the normal mixture model.

2.2. Asymptotic distribution of the LRT

We proceed to study the large sample behavior of the LRT using a sandwich idea to derive the asymptotic null distribution of R_n. We first establish an asymptotic upper bound for R_n. Write $r_n(\alpha,\theta,\sigma) = r_{1n}(\alpha,\theta,\sigma) + r_{2n}$, where $r_{1n}(\alpha,\theta,\sigma) = 2\{l_n(\alpha,\theta,\sigma) - l_n(0,0,1)\}$ and $r_{2n} = 2\{l_n(0,0,1) - l_n(0,0,\hat\sigma_0)\}$.

First, analyze $r_{1n}(\alpha,\theta,\sigma)$. Write $r_{1n}(\alpha,\theta,\sigma) = 2\sum\log(1+\delta_i)$, where $\delta_i = (\sigma^2-1)U_i(\sigma) + \alpha\theta Y_i(\theta,\sigma)$, with

$$U_i(\sigma) = (\sigma^2-1)^{-1}\Big[\frac{1}{\sigma}\exp\Big\{-\frac{X_i^2}{2}(\frac{1}{\sigma^2}-1)\Big\} - 1\Big], \tag{4}$$

$$Y_i(\theta,\sigma) = \frac{1}{\sigma\theta}\Big[\exp\Big\{-\frac{(X_i-\theta)^2}{2\sigma^2}+\frac{X_i^2}{2}\Big\} - \exp\Big\{-\frac{X_i^2}{2\sigma^2}+\frac{X_i^2}{2}\Big\}\Big]. \tag{5}$$

The functions $U_i(\sigma)$ and $Y_i(\theta,\sigma)$ are continuously differentiable with $U_i(1) = (X_i^2-1)/2$ and $Y_i(0,\sigma) = \sigma^{-3}X_i\exp\{-X_i^2(\sigma^{-2}-1)/2\}$. Also, note that under the null distribution $N(0,1)$, $E\{U_i(\sigma)\}=0$ and $E\{Y_i(\theta,\sigma)\}=0$ for any σ and θ.

By the inequality $2\log(1+x) \le 2x - x^2 + 2x^3/3$, $r_{1n}(\alpha,\theta,\sigma) = 2\sum_{i=1}^n \log(1+\delta_i) \le 2\sum_{i=1}^n \delta_i - \sum_{i=1}^n \delta_i^2 + (2/3)\sum_{i=1}^n \delta_i^3$. Re-write δ_i as

$$\delta_i = (\sigma^2-1)U_i(1) + \alpha\theta Y_i(\theta,1) + \epsilon_{in}, \tag{6}$$

where the remainder $\epsilon_{in} = (\sigma^2-1)\{U_i(\sigma)-U_i(1)\} + \alpha\theta\{Y_i(\theta,\sigma)-Y_i(\theta,1)\}$. Since the processes $U^*(\sigma) = n^{-1/2}\sum\{U_i(\sigma)-U_i(1)\}/(\sigma^2-1)$ and $Y^*(\theta,\sigma) = n^{-1/2}\sum\{Y_i(\theta,\sigma)-Y_i(\theta,1)\}/(\sigma^2-1)$, $\sigma^2\in[1-\delta,1+\delta]$ and $|\theta|\le M$, are tight (see Chen and Chen (2001c), Proposition 1), we see that $U^*(\sigma) = O_p(1)$ and $Y^*(\theta,\sigma) = O_p(1)$, implying

$$\sum_{i=1}^n \epsilon_{in} = n^{1/2}(\sigma^2-1)^2 O_p(1) + n^{1/2}\alpha\theta(\sigma^2-1)O_p(1). \tag{7}$$

Our convention has $\sup_{|\sigma^2-1|\le\delta}|U^*(\sigma)| = O_p(1)$, and $\sup_{|\theta|\le M, |\sigma^2-1|\le\delta}|Y^*(\theta,\sigma)| = O_p(1)$. Put $E_{n1} = (\sigma^2-1)^2O_p(1)$, $E_{n2} = \alpha\theta(\sigma^2-1)O_p(1)$, $U_i = U_i(1)$, and $Y_i(\theta) = Y_i(\theta,1)$. By (6) and (7), we obtain

$$\sum_{i=1}^n \delta_i = \sum_{i=1}^n\{(\sigma^2-1)U_i + \alpha\theta Y_i(\theta)\} + n^{1/2}(E_{n1}+E_{n2}). \tag{8}$$

Similarly, we can replace σ^2 with 1 in the square and cubic terms of δ_i, and arrive at

$$\sum_{i=1}^n \delta_i^2 = \sum_{i=1}^n\{(\sigma^2-1)U_i + \alpha\theta Y_i(\theta)\}^2 + n(E_{n1}^2+E_{n2}^2), \tag{9}$$

$$\left|\sum_{i=1}^n \delta_i^3 - \sum_{i=1}^n\{(\sigma^2-1)U_i + \alpha\theta Y_i(\theta)\}^3\right| = n(|E_{n1}|^3+|E_{n2}|^3). \tag{10}$$

It is important to note that in (10), the remainder terms have a factor of n rather than $n^{3/2}$. To see this, e.g.,

$$\begin{aligned}\sum_{i=1}^n |(\sigma^2-1)\{U_i(\sigma)-U_i(1)\}|^3 &= n(\sigma^2-1)^6(1/n)\sum_{i=1}^n |\{U_i(\sigma)-U_i(1)\}/(\sigma^2-1)|^3\\ &= n(\sigma^2-1)^6 O_p(1) = n|E_{n1}|^3.\end{aligned}$$

Now by (8), (9) and (10),

$$r_{1n}(\alpha,\theta,\sigma) \le 2\sum_{i=1}^{n}\{(\sigma^2-1)U_i+\alpha\theta Y_i(\theta)\} - \sum_{i=1}^{n}\{(\sigma^2-1)U_i+\alpha\theta Y_i(\theta)\}^2$$
$$+(2/3)\sum_{i=1}^{n}\{(\sigma^2-1)U_i+\alpha\theta Y_i(\theta)\}^3 + n^{1/2}(E_{n1}+E_{n2})$$
$$+n\sum_{j=2}^{3}(|E_{n1}|^j+|E_{n2}|^j). \quad (11)$$

Introduce $Z_i(\theta) = Y_i(\theta) - \theta U_i$. Then $(\sigma^2-1)U_i + \alpha\theta Y_i(\theta) = t_1 U_i + t_2 Z_i(\theta)$, where $t_1 = \sigma^2 - 1 + \alpha\theta^2$, and $t_2 = \alpha\theta$. Note that $EU_iZ_i(\theta) = 0$. It can be proved that the cubic and remainder terms in (11) are controlled by the square term, see Chen and Chen (2001c) for details. As for the remainder terms in (11), since $|\theta| \le M$, $n^{1/2}|E_{n1}| = n^{1/2}(\sigma^2-1)^2|O_p(1)| \le n^{1/2}(t_1^2+t_2^2)|O_p(1)| = o_p\{\sum_{i=1}^n[t_1U_i+t_2Z_i(\theta)]^2\}$, and similarly $n^{1/2}|E_{n2}| \le n^{1/2}\{t_2^2+(\sigma^2-1)^2\}|O_p(1)| \le n^{1/2}(t_1^2+t_2^2)|O_p(1)| = o_p\{\sum_{i=1}^n[t_1U_i+t_2Z_i(\theta)]^2\}$. Note that when $t_1 = t_2 = 0$, i.e., $\sigma^2 = 1$ and $\theta = 0$ in the above inequalities, $r_{1n} = 0 = o_p(1)$. Thus, this case can be ignored. The other remainder terms resulting from the square or cubic sum are of the same (or higher) order as that from the linear sum. In fact, $n(E_{n1}^2+E_{n2}^2) \le n(t_1^2+t_2^2)^2 O_p(1) = (t_1^2+t_2^2)O_p\{\sum_{i=1}^n[t_1U_i+t_2Z_i(\theta)]^2\}$ and $n(|E_{n1}|^3+|E_{n2}|^3) \le (|t_1|+|t_2|)O_p(nE_{n1}^2+nE_{n2}^2)$. Then (11) can be expressed as

$$r_{1n}(\alpha,\theta,\sigma)$$
$$\le 2\sum_{i=1}^{n}\{t_1U_i+t_2Z_i(\theta)\} - \sum_{i=1}^{n}\{t_1U_i+t_2Z_i(\theta)\}^2\{1+(|t_1|+|t_2|)O_p(1)+o_p(1)\}. \quad (12)$$

Since U_i and $Z_i(\theta)$ are orthogonal, (12) is further reduced to

$$r_{1n}(\alpha,\theta,\sigma)$$
$$\le 2\sum_{i=1}^{n}\{t_1U_i+t_2Z_i(\theta)\} - \sum_{i=1}^{n}\{t_1^2U_i^2+t_2^2Z_i^2(\theta)\}\{1+(|t_1|+|t_2|)O_p(1)+o_p(1)\}. \quad (13)$$

Let $\hat{t}_1 = \hat{\sigma}^2 - 1 + \hat{\alpha}\hat{\theta}^2$ and $\hat{t}_2 = \hat{\alpha}\hat{\theta}$ be the MLE's. By Lemma 2, $\hat{t}_1 = o_p(1)$ and $\hat{t}_2 = o_p(1)$. Consequently, replacement of the MLE's in (13) gives

$$r_{1n}(\hat{\alpha},\hat{\theta},\hat{\sigma}) \le 2\sum_{i=1}^{n}\{\hat{t}_1U_i+\hat{t}_2Z_i(\hat{\theta})\} - \sum_{i=1}^{n}\{\hat{t}_1^2U_i^2+\hat{t}_2^2Z_i^2(\hat{\theta})\}\{1+o_p(1)\}$$
$$\le \frac{(\sum U_i)^2}{\sum U_i^2}\{1+o_p(1)\} + \sup_{|\theta|\le M}\frac{[\{\mathrm{sgn}(\theta)\sum Z_i(\theta)\}^+]^2}{\sum Z_i^2(\theta)}\{1+o_p(1)\}$$
$$\le \frac{(\sum U_i)^2}{\sum U_i^2} + \sup_{|\theta|\le M}\frac{[\{\mathrm{sgn}(\theta)\sum Z_i(\theta)\}^+]^2}{\sum Z_i^2(\theta)} + o_p(1), \quad (14)$$

where sgn(θ) is the sign function. For detailed analysis leading to (14), see Chen and Chen (2001c).

Recall that $R_n = r_n(\hat{\alpha}, \hat{\theta}, \hat{\sigma}) = r_{1n}(\hat{\alpha}, \hat{\theta}, \hat{\sigma}) + r_{2n}$, and note that r_{2n} renders an ordinary quadratic approximation, i.e., $r_{2n} = -(\sum U_i)^2/\sum U_i^2 + o_p(1)$. An upper bound for R_n is obtained as

$$R_n \leq \sup_{|\theta|\leq M} \frac{[\{\text{sgn}(\theta)\sum_{i=1}^n Z_i(\theta)\}^+]^2}{nEZ_1^2(\theta)} + o_p(1). \tag{15}$$

To obtain a lower bound for R_n, let $\epsilon > 0$ be any fixed small number and let $R_n(\epsilon)$ be the supremum of $r_n(\alpha, \theta, \sigma)$ under the restriction $\epsilon \leq |\theta| \leq M$. Chen and Chen (2001c) show that, for any fixed $\epsilon \leq |\theta| \leq M$, if $\alpha = \tilde{\alpha}(\theta)$ and $\sigma = \tilde{\sigma}(\theta)$ assume the values determined by $\sigma^2 - 1 + \alpha\theta^2 = \sum U_i/\sum U_i^2$ and $\alpha\theta = [\text{sgn}(\theta)\sum Z_i(\theta)]^+/\sum Z_i^2(\theta)$, then

$$r_{1n}(\tilde{\alpha}(\theta), \theta, \tilde{\sigma}(\theta)) = \frac{\{\sum U_i\}^2}{\sum U_i^2} + \frac{[\{\text{sgn}(\theta)\sum Z_i(\theta)\}^+]^2}{nEZ_1^2(\theta)} + o_p(1).$$

It thus follows that

$$R_n(\epsilon) \geq \sup_{\epsilon\leq|\theta|\leq M} r_n(\tilde{\alpha}(\theta), \theta, \tilde{\sigma}(\theta)) = \sup_{\epsilon\leq|\theta|\leq M} \frac{[\{\text{sgn}(\theta)\sum Z_i(\theta)\}^+]^2}{nEZ_1^2(\theta)} + o_p(1). \tag{16}$$

Theorem 1. *Let $X_1, \ldots, X_n$ be a random sample from the mixture distribution $(1-\alpha)N(0, \sigma^2) + \alpha N(\theta, \sigma^2)$, where $0 \leq \alpha \leq 1$, $|\theta| \leq M$ and $\sigma > 0$, otherwise unknown. Let R_n be (twice) the log-likelihood ratio test statistic for testing $H_0 : \alpha = 0$. Then under the null distribution, $R_n \to \sup_{|\theta|\leq M}\{\zeta^+(\theta)\}^2$, as $n \to \infty$, where $\zeta(0) = 0$ and, for $0 < |\theta| \leq M$, $\zeta(\theta)$ is a Gaussian process with mean 0 and variance 1 and the autocorrelation given by $\rho(s,t) = sgn(st)a(st)/\{a(s^2)a(t^2)\}^{1/2}$ for $s, t \neq 0$, and $\rho(0,t) = |t|/(b^2(t))^{1/2}$, where $a(x) = e^x - 1 - x^2/2$.*

Proof. In light of (15) and (16), the result follows by first letting $n \to \infty$ and then $\epsilon \to 0$. For details, see Chen and Chen (2001c).

3. Two-Mean Parameter Mixtures

In this section we study the testing problem when both mean parameters θ_1 and θ_2 are unknown. Assume that $|\theta_i| \leq M$, $i = 1, 2$, and that $0 \leq \alpha \leq 1/2$, so that θ_1 and θ_2 are distinguishable. We wish to test $H_0 : \alpha = 0$ versus the full model (1).

Let $r_n(\alpha, \theta_1, \theta_2, \sigma) = 2\{l_n(\alpha, \theta_1, \theta_2, \sigma) - l_n(0, \hat{\theta}, \hat{\theta}, \hat{\sigma}_0)\}$, where l_n is the log-likelihood function, and $\hat{\theta} = \bar{X}$ and $\hat{\sigma}_0^2 = n^{-1}\sum(X_i - \bar{X})^2$ are the MLE's of $\theta_1 = \theta_2 = \theta$ and σ^2 under the null model. Let $\hat{\alpha}$, $\hat{\theta}_1$, $\hat{\theta}_2$ and $\hat{\sigma}^2$ be the MLEs for α, θ_1, θ_2 and σ^2 under the full model (1). The LRT is to reject H_0 if $R_n = r_n(\hat{\alpha}, \hat{\theta}_1, \hat{\theta}_2, \hat{\sigma})$ is large.

3.1. The MLE's

The statement of Lemma 1 remains true, i.e., under the null distribution, there are constants $0 < \epsilon < \Delta < \infty$ such that $P(\epsilon \le \hat{\sigma}^2 \le \Delta) \to 1$ as $n \to \infty$. The proof is also similar to that of Lemma 1 (see Chen and Chen (2001c)). Lemma 2 can be re-written as follows.

Lemma 3. *Under the null distribution,* $\hat{\theta_1} \to 0$, $\hat{\alpha}\hat{\theta_2} + (1-\hat{\alpha})\hat{\theta_1} \to 0$, $\hat{\alpha}\hat{\theta_2}^2 \to 0$ *and* $\hat{\sigma}^2 \to 1$ *in probability, as* $n \to \infty$.

Proof. The proof is similar to that of Lemma 2. See Chen and Chen (2001c) for details.

In light of Lemma 3, without loss of generality, σ^2 can be restricted to a neighborhood of $\sigma^2 = 1$, say $[1-\delta, 1+\delta]$ for a small number $\delta > 0$.

To derive the asymptotic distribution of the LRT in the present case, we face the challenge of a loss of positive-definiteness of the quadratic term in (11). To overcome the difficulty, the parameter space is partitioned into two parts: $|\theta_2| > \epsilon$ and $|\theta_2| \le \epsilon$, for an arbitrarily small $\epsilon > 0$. The LRT will be analyzed within each part by using the sandwich approach. Let $R_n(\epsilon; I)$ denote the supremum of the likelihood function within the part $|\theta_2| \ge \epsilon$, and $R_n(\epsilon; II)$ the supremum within $|\theta_2| \le \epsilon$. Then $R_n = \max\{R_n(\epsilon; I), R_n(\epsilon; II)\}$. The number ϵ will remain fixed as n approaches infinity. It is easily seen that Lemma 3 remains true under either restriction $|\theta_2| \ge \epsilon$ or $|\theta_2| \le \epsilon$. Dependence on ϵ will be suppressed notationally for the MLE's of the parameters. Thus $\hat{\alpha}$, $\hat{\theta}_1$, $\hat{\theta}_2$ and $\hat{\sigma}$ will denote the constrained MLE's of α, θ_1, θ_2 and σ with restriction $|\theta_2| \ge \epsilon$ in the analysis of $R_n(\epsilon; I)$, but stand for the constrained MLE's with restriction $|\theta_2| \le \epsilon$ in the analysis of $R_n(\epsilon; II)$.

3.2. Analysis of $R_n(\epsilon; I)$

We first establish an asymptotic upper bound for $R_n(\epsilon; I)$. As in Section 2.2, write $r_n(\alpha, \theta_1, \theta_2, \sigma) = r_{1n}(\alpha, \theta_1, \theta_2, \sigma) + r_{2n}$, where $r_{2n} = 2\{l_n(0,0,0,1) - l_n(0, \hat{\theta}, \hat{\theta}, \hat{\sigma}_0)\}$. To analyze $r_{1n}(\alpha, \theta_1, \theta_2, \sigma)$, express $r_{1n}(\alpha, \theta_1, \theta_2, \sigma) = 2\sum \log(1 + \delta_i)$ with

$$\delta_i = (1-\alpha)\theta_1 Y_i(\theta_1, \sigma) + \alpha\theta_2 Y_i(\theta_2, \sigma) + (\sigma^2 - 1)U_i(\sigma), \tag{17}$$

where $Y_i(\theta, \sigma)$ and $U_i(\sigma)$ are defined in (5) and (4). Re-write $\delta_i = m_1 Y_i(0,1) + (\sigma^2 - 1 + m_2)U_i(1) + m_3 V_i(\theta_2) + \epsilon_{in}$, where ϵ_{in} is the remainder, $m_1 = (1-\alpha)\theta_1 + \alpha\theta_2$, $m_2 = (1-\alpha)\theta_1^2 + \alpha\theta_2^2$, $m_3 = \alpha\theta_2^3$, and $V_i(\theta_2) = \{Y_i(\theta_2, 1) - Y_i(0,1) -$

$\theta_2 U_i(1)\}/\theta_2^2$. Define $V_i(0) = -(X_i/2) + (X_i^3/6)$ so that the function $V_i(\theta)$ is continuously differentiable. By an analysis similar to the single mean parameter case, the sum of the remainder ϵ_{in}'s satisfies

$$\epsilon_n = \sum_{i=1}^n \epsilon_{in} = O_p\{\sqrt{n}|\sigma^2 - 1|[|m_1| + \theta_1^2 + \alpha\theta_2^2 + |\sigma^2 - 1|] + \sqrt{n}|\theta_1^3|\}. \qquad (18)$$

Note that $U_i = U_i(1) = (X_i^2 - 1)/2$ and $Y_i(0,1) = X_i$. We have $\sum_{i=1}^n \delta_i = m_1 \sum_{i=1}^n X_i + (\sigma^2 - 1 + m_2)\sum_{i=1}^n U_i + m_3 \sum_{i=1}^n V_i(\theta_2) + \epsilon_n$. Since the remainders resulting from the square and cubic sums are of the same (or higher) order as the remainder from the linear sum (see the similar analysis in the case of single mean parameter mixtures), we have

$$\begin{aligned} r_{1n}(\alpha, \theta_1, \theta_2, \sigma) &\le 2\sum_{i=1}^n \delta_i - \sum_{i=1}^n \delta_i^2 + (2/3)\sum_{i=1}^n \delta_i^3 \\ &= 2\sum_{i=1}^n \{m_1 X_i + (\sigma^2 - 1 + m_2)U_i + m_3 V_i(\theta_2)\} \\ &\quad - \sum_{i=1}^n \{m_1 X_i + (\sigma^2 - 1 + m_2)U_i + m_3 V_i(\theta_2)\}^2 \\ &\quad + (2/3)\sum_{i=1}^n \{m_1 X_i + (\sigma^2 - 1 + m_2)U_i + m_3 V_i(\theta_2)\}^3 \\ &\quad + O_p\{\sqrt{n}|\sigma^2 - 1|[|m_1| + \theta_1^2 + \alpha\theta_2^2 + |\sigma^2 - 1|] + \sqrt{n}|\theta_1^3|\}. \end{aligned}$$

Furthermore, the cubic sum is negligible when compared to the square sum. This can be justified by using the idea leading to (14). First, the square sum times n^{-1} approaches $E\{m_1 X_1 + (\sigma^2 - 1 + m_2)U_1 + m_3 V_1(\theta_2)\}^2$ uniformly. The limit is a positive definitive quadratic form in variables m_1, $\sigma^2 - 1 + m_2$ and m_3. Next, noting that X_i, U_i and $V_i(\theta_2)$ are mutually orthogonal, we have that

$$\begin{aligned} &r_{1n}(\hat\alpha, \hat\theta_1, \hat\theta_2, \hat\sigma) \\ &\le 2\Big\{\hat m_1 \sum_{i=1}^n X_i + (\hat\sigma^2 - 1 + \hat m_2)\sum_{i=1}^n U_i + \hat m_3 \sum_{i=1}^n V_i(\hat\theta_2)\Big\} \\ &\quad - \Big\{\hat m_1^2 \sum_{i=1}^n X_i^2 + (\hat\sigma^2 - 1 + \hat m_2)^2 \sum_{i=1}^n U_i^2 + \hat m_3^2 \sum_{i=1}^n V_i^2(\hat\theta_2)\Big\}\{1 + o_p(1)\} + \hat\epsilon_n. \end{aligned}$$

Here the terms with a hat are their (constrained) MLE's with restriction $|\theta_2| \ge \epsilon$ as remarked at the end of Section 3.1. In particular, from (18), $\hat\epsilon_n = O_p\{\sqrt{n}|\hat\sigma^2 - 1|[|\hat m_1| + \hat\theta_1^2 + \hat\alpha\hat\theta_2^2 + |\hat\sigma^2 - 1|] + \sqrt{n}|\hat\theta_1^3|\}$. By the Cauchy inequality (e.g., $\sqrt{n}|\hat m_1| \le$

$1+n\hat m_1^2$) and the restriction $|\theta_2|\ge\epsilon$ (hence $|\hat\theta_2|\ge\epsilon$), we have

$$\begin{aligned}&\sqrt{n}|\hat\sigma^2-1|[|\hat m_1|+\hat\theta_1^2+\hat\alpha\hat\theta_2^2+|\hat\sigma^2-1|]+\sqrt{n}|\hat\theta_1^3|\\ &\le|\hat\sigma^2-1|[4+n\{\hat m_1^2+\hat\theta_1^4+(\hat\alpha\hat\theta_2^2)^2+(\hat\sigma^2-1)^2\}]+|\hat\theta_1|(1+n\hat\theta_1^4)\\ &=o_p(1)+no_p\{\hat m_1^2+\hat\theta_1^4+(\hat\alpha\hat\theta_2^2)^2+(\hat\sigma^2-1)^2\}\\ &=o_p(1)+no_p\{\hat m_1^2+(\hat\sigma^2-1+\hat m_2)^2+\hat m_3^2\}.\end{aligned}$$

Thus the remainder term $\hat\epsilon_n$ can also be absorbed into the quadratic sum, i.e.,

$$\begin{aligned}&r_{1n}(\hat\alpha,\hat\theta_1,\hat\theta_2,\hat\sigma)\\ &\le 2\Big\{\hat m_1\sum_{i=1}^n X_i+(\hat\sigma^2-1+\hat m_2)\sum_{i=1}^n U_i+\hat m_3\sum_{i=1}^n V_i(\hat\theta_2)\Big\}\\ &\quad-\Big\{\hat m_1^2\sum_{i=1}^n X_i^2+(\hat\sigma^2-1+\hat m_2)^2\sum_{i=1}^n U_i^2+\hat m_3^2\sum_{i=1}^n V_i^2(\hat\theta_2)\Big\}\{1+o_p(1)\}+o_p(1).\end{aligned}$$

Similar to (14), the right-hand side of the above inequality becomes even greater when $\hat m_1$, $\hat\sigma^2-1+\hat m_2$ and $\hat m_3$ are replaced with

$$\tilde m_1=\frac{\sum X_i}{\sum X_i^2},\quad \tilde\sigma^2-1+\tilde m_2=\frac{\sum U_i}{\sum U_i^2},\quad \tilde m_3=\frac{\{\mathrm{sgn}(\theta_2)\sum V_i(\theta_2)\}^+}{\sum V_i^2(\theta_2)},\tag{19}$$

for any $\epsilon\le|\theta_2|\le M$, so that

$$r_{1n}(\hat\alpha,\hat\theta_1,\hat\theta_2,\hat\sigma)\le\frac{\{\sum X_i\}^2}{n}+2\frac{\{\sum U_i\}^2}{n}+\sup_{\epsilon\le|\theta|\le M}\frac{[\{\mathrm{sgn}(\theta)\sum V_i(\theta)\}^+]^2}{nEV_1^2(\theta)}+o_p(1).\tag{20}$$

On the other hand, classic analysis gives

$$r_{2n}=2\{l_n(0,0,0,1)-l_n(0,\hat\theta,\hat\theta,\hat\sigma_0)\}=-n\bar X^2-(2/n)\Big\{\sum_{i=1}^n U_i\Big\}^2+o_p(1).\tag{21}$$

Combining (20) and (21) yields

$$R_n(\epsilon;I)\le\sup_{\epsilon\le|\theta|\le M}\frac{[\{\mathrm{sgn}(\theta)\sum V_i(\theta)\}^+]^2}{nEV_1^2(\theta)}+o_p(1).\tag{22}$$

We have thus established an asymptotic upper bound for $R_n(\epsilon;I)$. The upper bound is also achievable. To see this, for $|\theta_2|\ge\epsilon$ fixed, let $\tilde\alpha$, $\tilde\theta_1$ and $\tilde\sigma$ be the solutions for α, θ_1 and σ of (19). Then $\tilde\alpha=O_p(n^{-1/2})$, $\tilde\theta_1=O_p(n^{-1/2})$ and $\tilde\sigma^2-1=O_p(n^{-1/2})$ uniformly in θ_2. Chen and Chen (2001c) show that $r_{1n}(\tilde\alpha,\tilde\theta_1,\theta_2,\tilde\sigma)=2\sum\tilde\delta_i-\sum\tilde\delta_i^2(1+o_p(1))$. By (19), $\tilde\alpha$, $\tilde\theta_1$ and $\tilde\sigma$ are such that

$$\sup_{\epsilon\le|\theta|\le M}r_n(\tilde\alpha,\tilde\theta_1,\theta,\tilde\sigma)=\sup_{\epsilon\le|\theta|\le M}\frac{[\{\mathrm{sgn}(\theta)\sum V_i(\theta)\}^+]^2}{nEV_1^2(\theta)}+o_p(1).$$

It is thus shown that the upper bound in (22) is achievable and hence

$$R_n(\epsilon; I) = \sup_{\epsilon \le |\theta| \le M} \frac{[\{\text{sgn}(\theta) \sum V_i(\theta)\}^+]^2}{nEV_1^2(\theta)} + o_p(1). \tag{23}$$

3.3. Analysis of $R_n(\epsilon; II)$

Now consider the restriction $|\theta_2| \le \epsilon$. In this case, θ_1 and θ_2 can be treated equally. In fact, since the MLE of θ_1 is consistent, we can take $|\theta_1| \le \epsilon$ as well.

As before, we know that $r_{1n}(\alpha, \theta_1, \theta_2, \sigma) = 2\sum_{i=1}^n \log(1+\delta_i) \le 2\sum_{i=1}^n \delta_i - \sum_{i=1}^n \delta_i^2 + (2/3)\sum_{i=1}^n \delta_i^3$. Let $\hat{m}_k = (1-\hat{\alpha})\hat{\theta}_1^k + \hat{\alpha}\hat{\theta}_2^k$. Expanding $Y_i(\theta, \sigma)$ and $U_i(\sigma)$ at $\theta = 0$ and $\sigma = 1$ in (17), we have

$$\begin{aligned}\hat{\delta}_i &= \hat{m}_1 Y_i(0,1) + (\hat{\sigma}^2 - 1 + \hat{m}_2) Y_i'(0,1) + \frac{1}{2}\hat{m}_3 Y_i''(0,1) \\ &\quad + \frac{1}{6}\{3(\hat{\sigma}^2-1)^2 + \hat{m}_4 + 6(\hat{\sigma}^2-1)\hat{m}_2\} Y_i'''(0,1) + \hat{\epsilon}_{in},\end{aligned} \tag{24}$$

where $Y_i'(0,1)$ is the first partial derivative of $Y_i(\theta, \sigma)$ with respect to θ at $\theta = 0$ and $\sigma^2 = 1$, and similarly for $Y_i''(0,1)$ and $Y_i'''(0,1)$. As before, put $Y_i = Y_i(0,1)$, $Y_i' = Y_i'(0,1)$, $Y_i'' = Y_i''(0,1)$ and $Y_i''' = Y_i'''(0,1)$. By calculation, $Y_i' = U_i(1) = (X_i^2 - 1)/2$, $Y_i'' = (X_i^3 - 3X_i)/3$, and $Y_i''' = 2U_i'(1) = (X_i^4 - 6X_i^2 + 3)/4$. The sum of the remainders, $\hat{\epsilon}_n = \sum \hat{\epsilon}_{in}$ satisfies

$$\hat{\epsilon}_n = n^{1/2}(\hat{\sigma}^2 - 1)^3 O_p(1) + n(\hat{m}_1^2 + \hat{m}_3^2) o_p(1) + n^{1/2}(|\hat{m}_5| + \hat{m}_6) O_p(1) + o_p(1). \tag{25}$$

Note that the cross-product terms in the Taylor expansion of (24) have been taken into account in the remainder, e.g., $n^{1/2}(\hat{\sigma}^2 - 1)\hat{m}_1 = o_p(n^{1/2}\hat{m}_1) = o_p(1 + n\hat{m}_1^2)$. Also note that $3(\hat{\sigma}^2-1)^2 + \hat{m}_4 + 6(\hat{\sigma}^2-1)\hat{m}_2 = 3(\hat{\sigma}^2 - 1 + \hat{m}_2)^2 + \hat{m}_4 - 3\hat{m}_2^2$. Hence (24) can be written as $\hat{\delta}_i = \hat{s}_1 Y_i + \hat{s}_2 Y_i' + \hat{s}_3 Y_i'' + \hat{s}_4 Y_i''' + \hat{\epsilon}_{in}'$, where

$$\hat{s}_1 = \hat{m}_1, \hat{s}_2 = \hat{\sigma}^2 - 1 + \hat{m}_2, \hat{s}_3 = (1/2)\hat{m}_3, \hat{s}_4 = (1/6)(\hat{m}_4 - 3\hat{m}_2^2). \tag{26}$$

Combining with (25), the sum of the remainders, $\hat{\epsilon}_n' = \sum \hat{\epsilon}_{in}'$, becomes

$$\begin{aligned}\hat{\epsilon}_n' &= n^{1/2}\hat{s}_2^2 O_p(1) + n^{1/2}(\hat{\sigma}^2-1)^3 O_p(1) + n(\hat{m}_1^2 + \hat{m}_3^2) o_p(1) \\ &\quad + n^{1/2}(|\hat{m}_5| + \hat{m}_6) O_p(1) + o_p(1).\end{aligned} \tag{27}$$

Therefore, an upper bound for $r_{1n}(\hat{\alpha}, \hat{\theta}_1, \hat{\theta}_2, \hat{\sigma})$ is

$$\begin{aligned}&r_{1n}(\hat{\alpha}, \hat{\theta}_1, \hat{\theta}_2, \hat{\sigma}) \\ &\le 2\sum_{i=1}^n \{\hat{s}_1 Y_i + \hat{s}_2 Y_i' + \hat{s}_3 Y_i'' + \hat{s}_4 Y_i'''\} - \sum_{i=1}^n \{\hat{s}_1 Y_i + \hat{s}_2 Y_i' + \hat{s}_3 Y_i'' + \hat{s}_4 Y_i'''\}^2 \\ &\quad + \frac{2}{3}\sum_{i=1}^n \{\hat{s}_1 Y_i + \hat{s}_2 Y_i' + \hat{s}_3 Y_i'' + \hat{s}_4 Y_i'''\}^3 + \hat{\epsilon}_n'.\end{aligned}$$

Similarly to the analysis in Section 2, the cubic sum is controlled by the square sum (see Chen and Chen (2001c)). Moreover, Y_i, Y_i', Y_i'' and Y_i''' are mutually orthogonal and hence the quadratic sum is positive-definite. We arrive at

$$r_{1n}(\hat{\alpha}, \hat{\theta}_1, \hat{\theta}_2, \hat{\sigma}) \leq 2L_n - Q_n\{1 + \epsilon O_p(1)\} + \hat{\epsilon}_n', \tag{28}$$

where $L_n = \hat{s}_1 \sum Y_i + \hat{s}_2 \sum Y_i' + \hat{s}_3 \sum Y_i'' + \hat{s}_4 \sum Y_i'''$, and $Q_n = \hat{s}_1^2 \sum_{i=1}^n Y_i^2 + \hat{s}_2^2 \sum_{i=1}^n (Y_i')^2 + \hat{s}_3^2 \sum_{i=1}^n (Y_i'')^2 + \hat{s}_4^2 \sum_{i=1}^n (Y_i''')^2$. Since $\hat{\sigma}^2 - 1 = o_p(1)$, $\hat{m}_2^2 \leq \epsilon^2$ and $\hat{m}_2^3 \leq \hat{m}_6$, we have $n^{1/2}|(\hat{\sigma}^2 - 1)^3| \leq 8n^{1/2}\{|\hat{s}_2|^3 + \hat{m}_2^3\} \leq \epsilon n^{1/2}\hat{s}_2^2 O_p(1) + n^{1/2}\hat{m}_6 O_p(1)$. Thus (27) can be expressed as

$$\hat{\epsilon}_n' = \epsilon n^{1/2}\hat{s}_2^2 O_p(1) + n(\hat{m}_1^2 + \hat{m}_3^2)o_p(1) + n^{1/2}(|\hat{m}_5| + \hat{m}_6)O_p(1) + o_p(1). \tag{29}$$

Now the key point is to show that

$$\hat{\epsilon}_n' = o_p(1) + \epsilon n\{\hat{s}_1^2 + \hat{s}_2^2 + \hat{s}_3^2 + \hat{s}_4^2\}O_p(1). \tag{30}$$

This result implies that the remainder is also negligible when compared to the square sum in (28). Put $\hat{\tau} = (1 - \hat{\alpha})|\hat{\theta}_1|^5 + \hat{\alpha}|\hat{\theta}_2|^5$. Then $|\hat{m}_5| + \hat{m}_6 = O_p(\hat{\tau})$. Therefore, (30) follows immediately from (29) and the following lemma.

Lemma 4. $\hat{\tau} = o_p(1) + \epsilon\{|\hat{s}_1| + |\hat{s}_2| + |\hat{s}_3| + |\hat{s}_4|\}O_p(1)$.

Proof. Let $\gamma > 1$ be a constant. Partition the sample space into several parts: $(1 - \hat{\alpha})|\hat{\theta}_1|^k \geq \gamma\hat{\alpha}|\hat{\theta}_2|^k$ and $\gamma^{-1} \leq (1 - \hat{\alpha})|\hat{\theta}_1|^k/(\hat{\alpha}|\hat{\theta}_2|^k) \leq \gamma$, $k = 1$ and 3. The proof is accomplished by showing that in each part, one of $\hat{s}_i$, $i = 1, \ldots, 4$, controls $\hat{\tau}$. For details, see Chen and Chen (2001c).

From (28) and (30), it follows that $r_{1n}(\hat{\alpha}, \hat{\theta}_1, \hat{\theta}_2, \hat{\sigma}) \leq 2L_n - Q_n\{1 + \epsilon O_p(1)\} + o_p(1)$. We thus see that $R_n(\epsilon; II) \leq (\sum Y_i'')^2/\{nE(Y_1'')^2\} + (\sum Y_i''')^2/\{nE(Y_1''')^2\} + \epsilon O_p(1)$. The upper bound in this inequality is attained when the parameters α, θ_1, θ_2 and σ assume the values determined by $s_1 = \sum Y_i / \sum Y_i^2$, $s_2 = \sum Y_i' / \sum (Y_i')^2$, $s_3 = \sum Y_i'' / \sum (Y_i'')^2$, $s_4 = \sum Y_i''' / \sum (Y_i''')^2$, where s_1, s_2, s_3 and s_4 are defined by (26). It then follows that

$$R_n(\epsilon; II) = \frac{(\sum Y_i'')^2}{nE(Y_1'')^2} + \frac{(\sum Y_i''')^2}{nE(Y_1''')^2} + \epsilon O_p(1). \tag{31}$$

Remark. A by-product of the above analysis shows that the MLE of σ^2 has a convergence rate of at most $n^{-1/4}$. To see this, consider the submodel where $\theta_1 = -\theta_2 = \theta$, $\alpha = 1/2$ and $\sigma^2 - 1 = -\theta^2$. The maximum of the likelihood function is achieved when $m_4 - 3m_2^2 = -2\theta^4 = 6\sum Y_i''' / \sum (Y_i''')^2 = O_p(n^{-1/2})$. This implies that $\hat{\theta} = O_p(n^{-1/8})$ and $\hat{\sigma}^2 - 1 = n^{-1/4}$. This is in contrast to ordinary semi-parametric models, where one may still have the usual rate of

$n^{-1/2}$ for parametric components, see Van der Vaart (1996). Moreover, the result suggests that the best possible rate for estimating the mixing distribution when a structural parameter is present is $n^{-1/8}$, rather than $n^{-1/4}$ as found by Chen (1995) for mixture models without a structural parameter.

3.4. Asymptotic distribution of the LRT

Theorem 2. *Let $X_1, \ldots, X_n$ be a random sample from the mixture distribution $(1-\alpha)N(\theta_1, \sigma^2) + \alpha N(\theta_2, \sigma^2)$, where $0 \le \alpha \le 1/2$, $|\theta_i| \le M$, $i = 1, 2$ and $\sigma > 0$. Let R_n be (twice) the log-likelihood ratio test statistic for testing $H_0 : \alpha = 0$. Then under the null distribution, $R_n \to \sup_{|\theta| \le M}[\{\varsigma^+(\theta)\}^2 I(\theta \ne 0) + \{\varsigma(0)^2 + Z^2\} I(\theta = 0)]$ as $n \to \infty$, where the process involved in the limiting distribution is defined as follows:* (1) *$\varsigma(\theta)$, $|\theta| \le M$, is a Gaussian process with mean* 0, *variance* 1 *and the autocorrelation function $\rho(s,t) = sgn(st)\{b(st)\}/\{b(s^2)b(t^2)\}^{1/2}$ for $s,\ t \ne 0$ and $\rho(0,t) = |t|^3/\{6b(t^2)\}^{1/2}$, where $b(x) = e^x - 1 - x - x^2/2$, and* (2) *$\varsigma(0)$ and $Z \sim N(0,1)$ are independent and for $s \ne 0$, $\text{Cov}\,\{\varsigma(s), Z\} = s^4/2\{6a(s^2)\}^{1/2}$.*

Proof. For any fixed ϵ, $R_n = \max\{R_n(\epsilon; I),\ R_n(\epsilon; II)\}$. By (23) and (31), the results follow by first letting $n \to \infty$ and then $\epsilon \to 0$. For details, see Chen and Chen (2001c).

4. Conclusion Remark

The asymptotic null distribution of the LRT for homogeneity in finite normal mixture models in the presence of a structural parameter has been derived without separation conditions on the mean parameters. It is proved that the asymptotic null distribution of the LRT is the maximum of a χ^2-variable and the supremum of the square of a truncated Gaussian process.

If the structural parameter were removed from the model, the peculiar large sample behavior of the LRT would disappear and the limiting null distribution would be simply the supremum of the square of the truncated Gaussian process and reduce to the one discovered by Chen and Chen (2001a). If in addition $M \to \infty$, the supremum is distributed approximately as $(2\log M)^{1/2} + \{X - \log(2\pi)\}/(2\log M)^{1/2}$, where $P(X \le x) = \exp\{-e^{-x}\}$, the type-I extreme value distribution, see Chernoff and Lander (1995), Appendix D and Adler (1990). The result in Bickel and Chernoff (1993) can be obtained in a heuristic way by letting $M = (\log n/2)^{1/2}$. It is interesting to see that the results from different model set-ups agree formally. Bickel and Chernoff actually dealt with a modified LRT by replacing a random element in the LRT statistic with its mean in order to simplify the analysis. It seems that their modification might not have changed the asymptotic behavior of the LRT substantially.

Computing the quantiles of the supremum of a Gaussian process over a region is a difficult problem. See the comments by Dacunha-Castelle and Gassiat (1999), and Chen and Chen (2001b). Some approximations in special cases can be found in Adler (1990) and Sun (1993).

Owing to the large sample study it is found that, even though the structural parameter is not part of the mixing distribution, the convergence rate of the MLE is $n^{-1/4}$ rather than $n^{-1/2}$. This is in sharp contrast to the ordinary semi-parametric models. Moreover, the estimated mixing distribution has a convergence rate $n^{-1/8}$ rather than $n^{-1/4}$, as discovered by Chen (1995) for finite mixture models without a structural parameter.

Acknowledgement

This work was supported in part by a grant from NSERC of Canada, and an FIL grant from Bowling Green State University. The authors thank the Editor, the associate editor, and two referees for their comments which helped improve the paper.

References

Adler, R. J. (1990). *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*. IMS Lecture Notes-Monograph Serious, Vol. 12. Institute of Mathematical Statistics, Hayward, California.

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* (Edited by J. K. Ghosh, S. K. Mitra, K. R. Parthasarathy and B. L. S. Prakasa Rao), 83-96. Wiley Eastern.

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

Chen, H. and Chen, J. (2001b). The likelihood ratio test for homogeneity in the finite mixture models. *Canad. J. Statist.* **29**, 201-215.

Chen, H. and Chen, J. (2001c). Tests for homogeneity in normal mixtures with presence of a structural parameter: technical details. Working Paper 2001-06 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.

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.

Dacunha-Castelle, D. and Gassiat, É. (1999). Testing in locally conic models, and application to mixture models. *Ann. Statist.* **27**, 1178-1209.

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** (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** (Edited by L. LeCam and R. A. Olshen).

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

Rubin H. (1956). Uniform convergence of random functions with applications to statistics. *Ann. Math. Statist.* **27**, 200-203.

Sun, J. (1993). Tail probabilities of the maxima of Gaussian random fields. *Ann. Probab.* **21**, 34-71.

Van der Vaart, A. W. (1996). Efficient maximum likelihood estimation in semiparametric mixture models. *Ann. Statist.* **24**, 862-878.

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* **20**, 595-601.

Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43403, U.S.A.

E-mail: hchen@math.bgsu.edu

Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ont N2L 3G1, Canada.

E-mail: jhchen@uwaterloo.ca

(Received December 2000; accepted December 2002)