Submitted to the Annals of Statistics

HYPOTHESIS TEST FOR NORMAL MIXTURE MODELS: THE EM APPROACH WITH TECHNICAL DETAILS

BY JIAHUA CHEN*† AND PENGFEI LI

University of British Columbia

Normal mixture distributions are arguably the most important mixture models, and also the most challenging technically. The likelihood function of the normal mixture model is unbounded based on a set of random samples unless an artificial bound is placed on its component variance parameter. Moreover, the model is not strongly identifiable so it is hard to differentiate between over-dispersion caused by the presence of a mixture and that caused by a large variance; and it has infinite Fisher information with respect to mixing proportions. There has been extensive research on finite normal mixture models, but much of it addresses merely consistency point estimation or useful practical procedures, and many results require undesirable restrictions on the parameter space. We show that an EM-test for homogeneity is effective at overcoming many challenges in the context of finite normal mixtures. We find that the limiting distribution of the EM-test is a simple function of the $0.5\chi_0^2 + 0.5\chi_1^2$ and χ_1^2 distributions when the mixing variances are equal but unknown, and the χ_2^2 when variances are unequal and unknown. Simulations show that the limiting distributions approximate the finite sample distribution satisfactorily. Two genetic examples are used to illustrate the application of the EM-test.

1. Introduction. The class of finite normal mixture models has many applications. More than a hundred years ago, Pearson (1894) modeled a set of crab observations with a two-component normal mixture distribution. In genetics, such models are often used for quantitative traits influenced

*Research was partially supported by the Natural Sciences and Engineering Research Council of Canada and by the Mathematics of Information Technology and Complex Systems.

† *Corresponding author*: Jiahua Chen, Department of Statistics, University of British Columbia, Vancouver, BC, V6T 1Z2, Canada; jhchen@stat.ubc.ca; (604) 822-1848.

AMS 2000 subject classifications: Primary 62F03; secondary 62F05

Keywords and phrases: Chi-square limiting distribution, compactness, normal mixture models, homogeneity test, likelihood ratio test, statistic genetics

by major genes. Roeder (1994) discusses an example in which the blood chemical concentration of interest is influenced by a major gene with additive effects; see Schork et al. (1996) for additional examples in human genetics. Normal mixture models are also used to account for heterogeneity in the age of onset for male and female schizophrenia patients (Everitt, 1996), and used in hematology studies (McLaren, 1996). They play a fundamental role in cluster analysis (Tadesse et al., 2005; Raftery and Dean, 2006), and in the study of the false discovery rate (Efron, 2004; McLachlan et al., 2006; Sun and Cai, 2007; Cai et al. 2007). In financial economics they are used for daily stock returns (Kon, 1984).

Contrary to intuition, of all the finite mixture models, the normal mixture models have the most undesirable mathematical properties. Their likelihood functions are unbounded unless the component variances are assumed equal or constrained, the Fisher information can be infinite, and the strong identifiability condition is not satisfied. We demonstrate these points in the following example.

EXAMPLE 1. *Let $X_1, \ldots, X_n$ be a random sample from the following normal mixture model:*

$$(1 - \alpha)N(\theta_1, \sigma_1^2) + \alpha N(\theta_2, \sigma_2^2). \tag{1.1}$$

Let $f(x, \theta, \sigma)$ be the density function of a normal distribution with mean θ and variance σ^2. The likelihood function is given by

$$l_n(\alpha, \theta_1, \theta_2, \sigma_1, \sigma_2) = \sum_{i=1}^{n} \log\{(1 - \alpha)f(X_i; \theta_1, \sigma_1) + \alpha f(X_i; \theta_2, \sigma_2)\}. \tag{1.2}$$

1. (Unbounded likelihood function). *The log-likelihood function is unbounded for any given n because when $\theta_1 = X_1$, $0 < \alpha < 1$, it goes to infinity when $\sigma_1 \to 0$.*
2. (Infinite Fisher information). *For each given θ_1, θ_2, σ_1^2, and σ_2^2, we have*

$$S_n = \frac{\partial l_n(\alpha, \theta_1, \theta_2, \sigma_1, \sigma_2)}{\partial \alpha}\Big|_{\alpha=0} = \sum_{i=1}^{n} \Big\{ \frac{f(X_i; \theta_2, \sigma_2)}{f(X_i; \theta_1, \sigma_1)} - 1 \Big\}.$$

Under the homogeneous model in which $\theta_1 = 0$, $\sigma_1 = 1$, and $\alpha = 0$, i.e., N(0,1), the Fisher information

$$E\{S_n^2\} = \infty, \text{ whenever } \sigma_2^2 > 2.$$

3. (Loss of strong identifiability). *It can be seen that*

$$\frac{\partial^2 f(x;\theta,\sigma)}{\partial\theta^2}|_{(\theta,\sigma^2)=(0,1)} = 2\frac{\partial f(x;\theta,\sigma)}{\partial(\sigma^2)}|_{(\theta,\sigma^2)=(0,1)}.$$

This is in violation of the strong identifiability condition introduced in Chen (1995).

The above properties of finite normal mixture models are in addition to other undesirable properties of general finite mixture models. In Hartigan (1985), Liu et al. (2003), and Liu and Shao (2004), the likelihood ratio statistic is shown to diverge to infinity as the sample size increases, which forces most researchers to restrict the mixing parameter (θ) into some compact space. Without which, Hall and Stewart (2005) find the likelihood ratio test can only consistently detect alternative models at distance $(n^{-1}\log\log n)^{1/2}$ rather than at the usual distance $n^{-1/2}$. The partial loss of identifiability, when $\theta_1 = \theta_2$, once forced in a technical separate condition, $|\theta_1 - \theta_2| \geq \epsilon > 0$ (Ghosh and Sen, 1985). This condition has recently been shown to be unnecessary by many authors, for instance, Garel (2005).

The unbounded likelihood prevents straightforward use of maximum likelihood estimation. Placing an additional constraint on the parameter space (e.g., Hathaway, 1985) or adding a penalty function (Chen at al., 2008) to the log-likelihood regains the consistency and efficiency of the maximum constrained/penalized likelihood estimators.

The loss of strong identifiability results in a lower best possible rate of convergence (Chen, 1995; Chen and Chen, 2003). Furthermore it invalidates many elegant asymptotic results such as those in Dacunha-Castelle and Gassiat (1999), Chen et al. (2001), and Charnigo and Sun (2004). Finite Fisher information is a common hidden condition of these papers, but it did not gain much attention until the paper of Li et al. (2008).

Due to the indisputable importance of finite normal mixture models, developing valid and useful statistical procedures is an urgent task, particularly for the test of homogeneity. Yet the task is challenging for the reasons presented. Many existing results used simulated quantiles of the corresponding statistics, see Wolfe (1971), McLachlan (1987), and Feng and McCulloch (1994). Without rigorous theory, however, it is difficult to reconcile their varying recommendations.

In this paper, we investigate the application of the EM-test (Li et al., 2008) to finite normal mixture models, and show that this test provides a most satisfactory and general solution to the problem. Interestingly, our asymptotic results do not require any constraints on the mean and variance parameters, or compactness of the parameter space.

In Section 2, we present the result for the normal mixture model (1.1) when $\sigma_1^2 = \sigma_2^2 = \sigma^2$. The limiting distribution of the EM-test is shown to be a simple function of the $0.5\chi_0^2 + 0.5\chi_1^2$ and the χ_1^2 distributions. In Section 3, we present the result for the general normal mixture model (1.1). The limiting distribution of the EM-test is found to be the χ_2^2. Both results are stunningly simple and convenient to apply. In both cases, we conduct simulation studies and the outcomes are in good agreement with the asymptotic results. In Section 4, we give two genetic examples. For convenience of the presentation, the proofs are deferred to the Appendix.

2. Normal mixture models in the presence of the structural parameter. When $\sigma_1 = \sigma_2 = \sigma$ and σ is unknown in model (1.1), we call σ the structural parameter. We are interested in the test of the homogeneity null hypothesis

$$H_0 : \alpha(1 - \alpha)(\theta_1 - \theta_2) = 0$$

under this assumption. Without loss of generality, we assume $0 \leq \alpha \leq 0.5$.

Because the population variance $\mathrm{Var}(X_1)$ is the sum of the component variance σ^2 and the variance of the mixing distribution $\alpha(1-\alpha)(\theta_1 - \theta_2)^2$, σ^2 is often underestimated by straight likelihood methods. Furthermore, most asymptotic results are obtained by approximating the likelihood function with some form of quadratic function (Liu and Shao, 2003; Marriott, 2007). The approximation is most precise when the fitted α value is away from 0 and 1. Based on these considerations, we recommend using the modified log-likelihood

$$pl_n(\alpha, \theta_1, \theta_2, \sigma) = l_n(\alpha, \theta_1, \theta_2, \sigma, \sigma) + p_n(\sigma) + p(\alpha)$$

with $l_n(\cdot)$ given in (1.2). We usually select $p_n(\sigma)$ such that it is bounded when σ is large, but goes to negative infinity as σ goes to 0, and $p(\alpha)$ such that it is maximized at $\alpha = 0.5$ and goes to negative infinity as α goes to 0 or 1. Concrete recommendations will be given later.

To construct the EM-test, we first choose a set of $\alpha_j \in (0, 0.5]$, $j = 1, 2, \ldots, J$, and a positive integer K. For each $j = 1, 2, \ldots, J$, let $\alpha_j^{(1)} = \alpha_j$

and compute

$$(\theta_{j1}^{(1)}, \theta_{j2}^{(1)}, \sigma_j^{(1)}) = \arg\max_{\theta_1,\ \theta_2,\ \sigma} pl_n(\alpha_j^{(1)}, \theta_1,\ \theta_2,\ \sigma).$$

For $i = 1, 2, \ldots, n$ and the current k, we use an E-step to compute

$$w_{ij}^{(k)} = \frac{\alpha_j^{(k)} f(X_i; \theta_{j2}^{(k)}, \sigma_j^{(k)})}{(1-\alpha_j^{(k)}) f(X_i; \theta_{j1}^{(k)}, \sigma_j^{(k)}) + \alpha_j^{(k)} f(X_i; \theta_{j2}^{(k)}, \sigma_j^{(k)})}$$

and then update α and other parameters by an M-step such that

$$\alpha_j^{(k+1)} = \arg\max_{\alpha}\{(n - \sum_{i=1}^{n} w_{ij}^{(k)}) \log(1-\alpha) + \sum_{i=1}^{n} w_{ij}^{(k)} \log(\alpha) + p(\alpha)\}$$

and

$$(\theta_{j1}^{(k+1)}, \theta_{j2}^{(k+1)}, \sigma_j^{(k+1)}) = \arg\left[\max_{\theta_1,\ \theta_2,\ \sigma} \sum_{h=1}^{2}\sum_{i=1}^{n} w_{ij}^{(k)} \log\{f(X_i; \theta_h, \sigma)\} + p_n(\sigma)\right].$$

The E-step and the M-step are iterated $K-1$ times.

For each k and j, we define

$$M_n^{(k)}(\alpha_j) = 2\{pl_n(\alpha_j^{(k)}, \theta_{j1}^{(k)}, \theta_{j2}^{(k)}, \sigma_j^{(k)}) - pl_n(1/2, \hat\theta_0, \hat\theta_0, \hat\sigma_0)\}$$

where $(\hat\theta_0, \hat\sigma_0) = \arg\max_{\theta,\sigma} pl_n(1/2, \theta, \theta, \sigma)$.

The EM-test statistic is then defined as

$$EM_n^{(K)} = \max\{M_n^{(K)}(\alpha_j) : j = 1, \ldots, J\}.$$

We reject the null hypothesis when $EM_n^{(K)}$ exceeds some critical value to be determined.

Consider the simplest case where $J = K = 1$ and $\alpha_1 = 0.5$. In this case, the EM-test is the likelihood ratio test against the alternative models with known $\alpha = 0.5$. The removal of one unknown parameter in the model simplifies the asymptotic property of the (modified) likelihood ratio test, and the limiting distribution becomes the $0.5\chi_0^2 + 0.5\chi_1^2$ which does not require the parameter space of θ to be compact. The price of this simplicity is a loss of efficiency when the data are from an alternative model with $\alpha \neq 0.5$. Choosing $J > 1$ initial values of α reduces the efficiency loss because the true α value can be close to one of the initial values. The EM-iteration updates the value of α_j and moves it toward the true α-value while retaining the nice asymptotic property.

Specific choice of initial set of α values is not crucial in general. This is another benefit of the EM-iteration. The updated α-values from either $\alpha = 0.3$ or $\alpha = 0.4$ are likely very close after two iterations. Hence, we recommend $\{0.1, 0.3, 0.5\}$. If some prior information indicates that the potential α value under the alternative model is low, then choosing $\{0.01, 0.025, 0.05, 0.1\}$ can improve the power of the test. We leave refinement considerations into a future research project at this stage.

The idea of the EM-test was introduced by Li et al. (2008) for mixture models with a single mixing parameter. Yet finite normal mixture models do not fit into the general theory and pose specific technical challenges. The asymptotic properties of the EM-test will be presented in the next subsection. The recommendation for penalty functions will be given in subsection 2.2.

2.1. *Asymptotic properties.* We study the asymptotic properties of the EM-test under the following conditions on the penalty functions $p(\alpha)$ and $p_n(\sigma)$:

C0. $p(\alpha)$ is a continuous function such that it is maximized at $\alpha = 0.5$ and goes to negative infinity as α goes to 0 or 1.

C1. $\sup\{|p_n(\sigma)| : \sigma > 0\} = o(n)$.

C2. The derivative $p'_n(\sigma) = o_p(n^{1/4})$ at any $\sigma > 0$.

We allow p_n to be dependent on the data. To ensure that the EM-test has the invariant property, we recommend choosing a p_n that also satisfies

C3. $p_n(a\sigma; aX_1 + b, \ldots, aX_n + b) = p_n(\sigma; X_1, \ldots, X_n)$.

The following intermediate results reveal some curious properties of the finite normal mixture model:

THEOREM 1. *Suppose Conditions C0, C1, and C2 hold. Under the null distribution $N(\theta_0, \sigma_0^2)$ we have, for $j = 1, \ldots, J$ and any $k \le K$,*

(a) if $\alpha_j = 0.5$ then

$$\theta_{j1}^{(k)} - \theta_0 = O_p(n^{-1/8}), \qquad \theta_{j2}^{(k)} - \theta_0 = O_p(n^{-1/8}),$$
$$\alpha_j^{(k)} - \alpha_j = O_p(n^{-1/4}), \qquad \sigma_j^{(k)} - \sigma_0 = O_p(n^{-1/4});$$

(b) if $0 < \alpha_j < 0.5$ *then*

$$\theta_{j1}^{(k)} - \theta_0 = O_p(n^{-1/6}), \qquad \theta_{j2}^{(k)} - \theta_0 = O_p(n^{-1/6}),$$
$$\alpha_j^{(k)} - \alpha_j = O_p(n^{-1/4}), \qquad \sigma_j^{(k)} - \sigma_0 = O_p(n^{-1/3}).$$

Note that the convergence rates of $(\theta_{j1}^{(k)}, \theta_{j2}^{(k)}, \sigma_j^{(k)})$ depend on the choice of initial α value, and it singles out $\alpha = 0.5$. Even when $\alpha_1 = 0.5$, $\alpha_1^{(k)} \neq 0.5$ when $k > 1$. However, this does not reduce Case (a) to Case (b) because $\alpha_1^{(k)} = 0.5 + o_p(1)$ rather than equaling a non-random constant $\alpha_1 \neq 0.5$.

THEOREM 2. *Suppose Conditions C0, C1, and C2 hold and* $\alpha_1 = 0.5$. *Then under the null distribution* $N(\theta_0, \sigma_0^2)$ *and for any finite* K, *as* $n \to \infty$,

$$\Pr(EM_n^{(K)} \leq x) \to F(x - \Delta)\{0.5 + 0.5F(x)\},$$

where $F(x)$ *is the cumulative density function* (*cdf*) *of the* χ_1^2 *and*

$$\Delta = 2 \max_{\alpha_j \neq 0.5} \{p(\alpha_j) - p(0.5)\}.$$

To shed some light on the non-conventional results, we reveal some helpful momental relationships. Without loss of generality, assume that under the null model, $\theta_1 = \theta_2 = 0$ and $\sigma^2 = 1$. The EM-test or other likelihood-based methods fit the data from the null model with an alternative model $(1-\alpha)N(\theta_1, \sigma^2) + \alpha N(\theta_2, \sigma^2)$. Asymptotically, the fit matches the first few sample moments. When $\alpha = 0.5$ is presumed, the first three moments of a homogeneous model and an alternative model can be made identical with proper choice of the values of the remaining parameters. Which model fits the data better is revealed through the fourth moment,

$$E(X_1^4) = 3 - (\theta_1^4 + \theta_2^4) \leq 3.$$

Thus, for local alternatives, we may as well test

$$H_0 : E(X_1^4) = 3 \text{ versus } H_a : E(X_1^4) < 3.$$

The parameter of this null hypothesis is on the boundary so that the null limiting distribution of $M_n(0.5)$ is the $0.5\chi_0^2 + 0.5\chi_1^2$.

When $\alpha = \alpha_0 \in (0, 0.5)$, the first two moments of the null and alternative models can be made identical, but their third moments differ because

$$E(X_1^3) = (1-\alpha_0)\theta_1^3 + \alpha_0\theta_2^3,$$

which can take any value in a neighborhood of 0. Thus, for local alternatives, we may as well test

$$H_0 : E(X_1^3) = 0 \text{ versus } H_a : E(X_1^3) \neq 0.$$

Because the null hypothesis is an interior point, $M_n(\alpha_0)$ has the asymptotic distribution $\chi_1^2 + 2\{p(\alpha_0) - p(0.5)\}$ in which $2\{p(\alpha_0) - p(0.5)\}$ is due to the penalty.

Since the sample third and fourth moments are asymptotically orthogonal, the limiting distribution of the EM-test involves the maximum of two independent distributions, the χ_1^2 and the $0.5\chi_0^2 + 0.5\chi_1^2$, and a term caused by the penalty $p(\alpha)$. This is the result as in the above theorem.

The order assessment results in Theorem 1 can be similarly explained. If $\alpha = 0.5$ is presumed, the fitted fourth moment of the mixing distribution will be $O_p(n^{-1/2})$ and hence both fitted θ_1 and θ_2 are $O_p(n^{-1/8})$. For other α values, the fitted third moment is $O_p(n^{-1/2})$, which implies that the fitted θ_1 and θ_2 are $O(n^{-1/6})$.

2.2. *Simulation results.* We demonstrate the precision of the limiting distribution via simulation and explore the power properties. Among several existing results, the modified likelihood ratio test (MLRT) in Chen and Kalbfleisch (2005) is known to have an accurate asymptotic upper bound. Thus we also include this method in our simulation.

The key idea of the MLRT is to define the modified likelihood function as

$$\tilde{l}_n(\alpha, \theta_1, \theta_2, \sigma) = l_n(\alpha, \theta_1, \theta_2, \sigma) + p(\alpha)$$

and the recommended penalty function is $\log\{4\alpha(1-\alpha)\}$. The corresponding statistic is defined as

$$M_n = 2\{l_n(\tilde{\alpha}, \tilde{\theta}_1, \tilde{\theta}_2, \tilde{\sigma}) - l_n(0.5, \tilde{\theta}_0, \tilde{\theta}_0, \tilde{\sigma}_0)\}$$

where $(\tilde{\alpha}, \tilde{\theta}_1, \tilde{\theta}_2, \tilde{\sigma})$ and $(0.5, \tilde{\theta}_0, \tilde{\theta}_0, \tilde{\sigma}_0)$ maximize $\tilde{l}_n$ under the alternative and null models respectively. Unlike that for the EM-test, the limiting distribution of M_n is unknown, but is shown to have an upper bound χ_2^2 when θ is confined in a compact space. Chen and Kalbfleisch (2005) show that the type I errors of the MLRT with critical values determined by the χ_2^2 distribution are close to the nominal values.

For the EM-test statistics, we choose the penalty function

$$p_n(\sigma) = -\Big\{s_n^2/\sigma^2 + \log(\sigma^2/s_n^2)\Big\},$$

where $s_n^2 = n^{-1}\sum_{i=1}^n (X_i - \bar{X})^2$ with $\bar{X} = n^{-1}\sum_{i=1}^n X_i$.

It can be seen that (a) $p_n(\sigma)$ satisfies Conditions C1-C3; (b) it effectively places an inverse gamma prior on σ^2; (c) it allows a closed form expression for $\sigma_j^{(k)}$; and (d) it is maximized at $\sigma^2 = s_n^2$. In fact, even a constant function $p_n(\sigma)$ satisfies C1-C2. This choice of $p_n(\sigma)$ prevents under-estimation of σ^2 and plays a role of higher order adjustment.

For the penalty function $p(\alpha)$, we choose $p(\alpha) = \log(1 - |1 - 2\alpha|)$. We refer to Li et al. (2008) for reasons of this choice. The combination of $p_n(\sigma)$ and $p(\alpha)$ results in accurate type I errors for the EM-test.

We conducted the simulation with two groups of initial values for α: (0.1, 0.2, 0.3, 0.4, 0.5) and (0.1, 0.3, 0.5). We generated 20,000 random samples from $N(0,1)$ with sample size n (n=100, 200). The simulated null rejection rates are summarized in Table 1. The EM-test and the MLRT both have accurate type I errors, especially $EM_n^{(2)}$ with the three initial values (0.1, 0.3, 0.5) for α.

TABLE 1
Type I errors (%) of the EM-test and the MLRT.

Level	$EM_n^{(1)}$	$EM_n^{(2)}$	$EM_n^{(3)}$	$EM_n^{(1)}$	$EM_n^{(2)}$	$EM_n^{(3)}$	MLRT
	$n = 100$						
10%	8.9	9.1	9.2	9.2	9.9	10.2	10.9
5%	4.6	4.8	4.8	4.6	5.1	5.3	5.7
1%	0.9	1.0	1.0	0.9	1.0	1.1	1.2
	$n = 200$						
10%	9.3	9.4	9.5	9.7	10.0	10.3	9.8
5%	4.6	4.8	4.8	4.7	5.0	5.1	5.0
1%	1.0	1.1	1.1	0.9	1.1	1.1	1.1

Results in columns (2, 3, 4) used $\alpha = (0.1, 0.2, 0.3, 0.4, 0.5)$.
Results in columns (5, 6, 7) used $\alpha = (0.1, 0.3, 0.5)$.

We selected four models for power assessment. The parameter settings are shown in Rows 2 to 5 of Table 2. The powers of the EM-test and the MLRT are estimated based on 5,000 repetitions and are presented in Table 3. We used the simulated critical values to ensure fairness of the comparison. The results show that the EM-test statistics based on three initial values have almost the same power as those from five initial values. Combining the type

I error results and the power comparison results, we recommend the use of $EM_n^{(2)}$ with three initial values (0.1, 0.3, 0.5) for α.

The EM-test has higher power when the mixing proportion α is close to 0.5, while the MLRT statistic performs better when α is close to 0. However, the limiting distribution of the EM-test is obtained without any restrictions on the model, while the limiting distribution of the MLRT is unknown, and the upper bound result is obtained under some restrictions. When α is small and some prior information on α value is known, the lower efficiency problem of the EM-test can be easily fixed. We conducted additional simulation by choosing the set of initial α-values $\{0.1, 0.05, 0.025, 0.01\}$. In this case, the limiting distribution of the EM-test becomes $\chi_1^2 + 2\{p(0.1) - p(0.5)\}$. When $n = 100$, the power comparison between the EM-test and the MLRT becomes 71.5% versus 71.0% for Model III, and 73.1% versus 75% for Model IV. Therefore the EM-test can be refined to attain higher efficiency and in many ways. Naturally, a systematic way is preferential and is best left to a future research project.

The other eight models in Table 2 have unequal variances, which are mainly selected for power comparisons in Section 3.3. To examine the importance of the equal variance assumption, we applied the current EM-test designed for finite normal mixture models in the presence of a structural parameter to the data from models V and IX. In some sense, Model V is a null model because its two component means are equal; while Model IX is an alternative model because its two component means are unequal. It can be seen in Table 3 that the current EM-test has a rightfully low rejection rate against Model V. This property is not shared by the MLRT. At the same time, the current EM-test has good power for detecting model IX. In fact, the power is comparable to that of the EM-test designed for finite normal mixture models with unequal variances, to be introduced in the next section. We conclude that when σ_1/σ_2 is close to 1, the power of the current EM-test is not sensitive to the $\sigma_1 = \sigma_2$ assumption.

To explore what happens when σ_1/σ_2 is large, we generated data from Model IX with σ_1 re-set to 2.4. The current EM-test rejected the null hypothesis 84% of the time, compared to a 96% rejection rate for the EM-test designed for finite mixture models without an equal variance assumption when $n = 100$. We conclude that when the two component variances are rather different, the current EM-test should not be used. An EM-test designed for finite mixture models without an equal variance assumption is preferred.

TABLE 2
Parameter values of normal mixture models for power assessment.

	$1-\alpha$	θ_1	θ_2	σ_1	σ_2
Model I	0.50	-1.15	1.20	1.00	1.00
Model II	0.25	-1.15	1.15	1.00	1.00
Model III	0.10	-1.30	1.30	1.00	1.00
Model IV	0.05	-1.55	1.55	1.00	1.00
Model V	0.50	0.00	0.00	1.20	0.50
Model VI	0.25	0.00	0.00	1.15	0.50
Model VII	0.10	0.00	0.00	1.40	0.50
Model VIII	0.05	0.00	0.00	1.85	0.50
Model IX	0.50	0.75	-0.75	1.20	0.80
Model X	0.25	0.65	-0.65	1.20	0.80
Model XI	0.10	0.85	-0.85	1.20	0.80
Model XII	0.05	1.15	-1.15	1.20	0.80

TABLE 3
Powers (%) of the EM-test and the MLRT at 5% level.

Model	$EM_n^{(1)}$	$EM_n^{(2)}$	$EM_n^{(3)}$	$EM_n^{(1)}$	$EM_n^{(2)}$	$EM_n^{(3)}$	MLRT
	$n=100$						
I	53.4	53.2	52.8	53.8	53.4	53.4	45.2
II	51.8	51.7	51.6	50.3	50.5	50.7	50.7
III	51.9	52.2	52.2	50.7	51.3	51.7	59.2
IV	49.5	51.2	51.5	50.7	51.6	52.0	63.1
V	15.2	17.0	17.6	16.0	17.8	18.1	33.4
IX	49.4	49.3	49.1	48.1	48.7	48.6	48.3
	$n=200$						
I	85.2	85.2	85.1	85.3	85.4	85.3	80.1
II	85.0	84.9	84.9	84.7	84.8	84.7	84.3
III	86.0	86.1	86.1	85.7	85.8	85.9	90.9
IV	81.4	82.3	82.5	82.5	83.1	83.2	91.1
V	23.0	25.0	25.9	24.4	26.0	26.9	52.9
IX	82.3	82.2	82.2	81.8	82.0	82.0	82.1

Results in columns (2, 3, 4) used $\alpha = (0.1, 0.2, 0.3, 0.4, 0.5)$.
Results in columns (5, 6, 7) used $\alpha = (0.1, 0.3, 0.5)$.

3. Normal Mixture Models in Both Mean and Variance Parameters.

3.1. *The EM-test Procedure.* In this section, we apply the EM-test to the test of homogeneity in the general normal mixture model (1.1) where both θ and σ are mixing parameters. We wish to test

$$H_0 : \alpha(1-\alpha) = 0 \text{ or } (\theta_1, \sigma_1^2) = (\theta_2, \sigma_2^2).$$

Compared to the case where σ is a structural parameter, the asymptotic properties of likelihood-based methods become much more challenging because of the unbounded log-likelihood and infinite Fisher information. Especially because of the latter, there exist few asymptotic results for general finite normal mixture models. Interestingly, we find that the EM-test can be directly applied and the asymptotic distribution is particularly simple. However, its derivation is complex.

To avoid the problem of unbounded likelihood, adding a penalty becomes essential in our approach. We define

$$pl_n(\alpha, \theta_1, \theta_2, \sigma_1, \sigma_2) = l_n(\alpha, \theta_1, \theta_2, \sigma_1, \sigma_2) + p_n(\sigma_1) + p_n(\sigma_2) + p(\alpha),$$

where $p_n(\sigma)$, $p(\alpha)$ are the same as before.

The EM-test statistic is constructed similarly. We first choose a set of $\alpha_j \in (0, 0.5]$, $j = 1, 2, \ldots, J$, and a positive integer K. For each $j = 1, 2, \ldots, J$, let $\alpha_j^{(1)} = \alpha_j$ and compute

$$(\theta_{j1}^{(1)}, \theta_{j2}^{(1)}, \sigma_{j1}^{(1)}, \sigma_{j2}^{(1)}) = \arg \max_{\theta_1,\ \theta_2,\ \sigma_1,\ \sigma_2} pl_n(\alpha_j^{(1)}, \theta_1,\ \theta_2,\ \sigma_1,\ \sigma_2).$$

For $i = 1, 2, \ldots, n$, and the current k, we use the E-step to compute

$$w_{ij}^{(k)} = \frac{\alpha_j^{(k)} f(X_i; \theta_{j2}^{(k)}, \sigma_j^{(k)})}{(1-\alpha_j^{(k)}) f(X_i; \theta_{j1}^{(k)}, \sigma_j^{(k)}) + \alpha_j^{(k)} f(X_i; \theta_{j2}^{(k)}, \sigma_j^{(k)})}$$

and then we use the M-step to update α and other parameters such that

$$\alpha_j^{(k+1)} = \arg\max_{\alpha} \{(n - \sum_{i=1}^{n} w_{ij}^{(k)}) \log(1-\alpha) + \sum_{i=1}^{n} w_{ij}^{(k)} \log(\alpha) + p(\alpha)\}$$

and

$$\begin{aligned}
&(\theta_{j1}^{(k+1)}, \theta_{j2}^{(k+1)}, \sigma_{j1}^{(k+1)}, \sigma_{j2}^{(k+1)}) \\
&\quad = \arg \max_{\theta_1,\ \theta_2,\ \sigma_1,\ \sigma_2} \sum_{h=1}^{2} \Big[\sum_{i=1}^{n} w_{ij}^{(k)} \log\{f(X_i; \theta_h, \sigma_h)\} + p_n(\sigma_h) \Big].
\end{aligned}$$

The E-step and the M-step are iterated $K-1$ times.

For each k and j, we define

$$M_n^{(k)}(\alpha_j) = 2\{pl_n(\alpha_j^{(k)}, \theta_{j1}^{(k)}, \theta_{j2}^{(k)}, \sigma_{j1}^{(k)}, \sigma_{j2}^{(k)}) - pl_n(1/2, \hat{\theta}_0, \hat{\theta}_0, \hat{\sigma}_0, \hat{\sigma}_0)\}$$

where $(\hat{\theta}_0, \hat{\sigma}_0) = \arg\max_{\theta,\sigma} pl_n(1/2, \theta, \theta, \sigma, \sigma)$. The EM-test statistic is then defined as

$$EM_n^{(K)} = \max\{M_n^{(K)}(\alpha_j) : j = 1, \ldots, J\}.$$

We reject the null hypothesis when $EM_n^{(K)}$ exceeds some critical value to be determined.

In terms of statistical procedure, the EM-test for the case of $\sigma_1^2 = \sigma_2^2$ is a special case of $\sigma_1^2 \neq \sigma_2^2$. However, the asymptotic distributions and their derivations are different.

3.2. *Asymptotic properties.* We further require that $p_n(\sigma)$ satisfies C1 and

C4. $p_n'(\sigma) = o_p(n^{1/6})$ for all $\sigma > 0$.
C5. $p_n(\sigma) \leq 4(\log n)^2 \log(\sigma)$, when $\sigma \leq n^{-1}$ and n is large.

The following theorems consider the consistency of $(\alpha_j^{(k)}, \theta_{j1}^{(k)}, \theta_{j2}^{(k)}, \sigma_{j1}^{(k)}, \sigma_{j2}^{(k)})$ and give the major result. The proofs are given in the Appendix.

THEOREM 3. *Suppose Conditions C0, C1, and C4-C5 hold. Under the null distribution $N(\theta_0, \sigma_0^2)$ we have, for $j = 1, \ldots, J$, $h = 1, 2$, and any $k \leq K$,*

$$\alpha_j^{(k)} - \alpha_j = o_p(1), \;\; \theta_{jh}^{(k)} - \theta_0 = o_p(1), \;\; \text{and } \sigma_{jh}^{(k)} - \sigma_0 = o_p(1).$$

THEOREM 4. *Suppose Conditions C0, C1, and C4-C5 hold. When $\alpha_1 = 0.5$, under the null distribution $N(\theta_0, \sigma_0^2)$ and for any finite K, as $n \to \infty$,*

$$EM_n^{(K)} \xrightarrow{d} \chi_2^2.$$

It is a surprise that the EM-test has a simpler limiting distribution when applied to a more complex model. We again shed some light on this via some moment consideration.

The test of homogeneity is to compare the fit of the null $N(0, 1)$ and the fit of the full model. The limiting distribution amounts to considering this problem when the data are from the null model. By matching the first two moments of the full model to the first two sample moments, we roughly select a full model such that

$$(1-\alpha)\theta_1 + \alpha\theta_2 = 0 \text{ and } (1-\alpha)(\theta_1^2+\sigma_1^2) + \alpha(\theta_2^2+\sigma_2^2) = 1.$$

Let $\beta_1 = \theta_1^2 + \sigma_1^2 - 1$. When the value of $\alpha = \alpha_0 \in (0, 0.5]$ (say $\alpha_0 = 0.5$), the third moment and the fourth moment of the full model are

$$\begin{aligned} E(X_1^3) &= 3\theta_1\beta_1, \\ E(X_1^4) &= 3\beta_1^2 - 2\theta_1^4 + 3. \end{aligned}$$

It is easy to verify that $\{E(X_1^3), E(X_1^4)\} = \{0, 3\}$ if and only if the mixture model is the homogeneous model. Therefore, we may as well test

$$H_0 : \{E(X_1^3), E(X_1^4)\} = \{0, 3\} \text{ versus } H_a : \{E(X_1^3), E(X_1^4)\} \neq \{0, 3\}.$$

As shown in Figure 1, $\{0, 3\}$ is an interior point of the parameter space of $\{E(X_1^3), E(X_1^4)\}$. Therefore the null limiting distribution of the EM-test is the χ_2^2. We note that when the observations are from an alternative model, the situation is totally different. A test on moments is not equivalent to the EM-test.

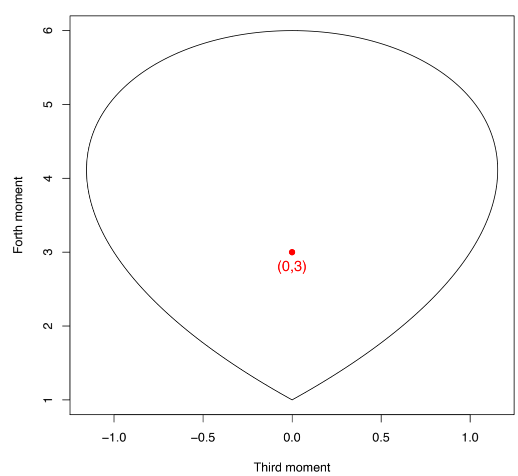

FIG 1. *The range (area inside the solid line) of* $\{E(X_1^3), E(X_1^4)\}$.

3.3. *Simulation Studies.* We demonstrate the precision of the limiting distribution and explore the power properties via simulations. In contrast to the case where $\sigma_1^2 = \sigma_2^2$, the EM-test does not have many competitors. Thus we set up an MLRT method with

$$M_n = 2\left\{ \sup_{\alpha,\theta_1,\theta_2,\sigma_1,\sigma_2} pl_n(\alpha,\theta_1,\theta_2,\sigma_1,\sigma_2) - pl_n(0.5,\hat{\theta}_0,\hat{\theta}_0,\hat{\sigma}_0,\hat{\sigma}_0)\right\}.$$

Although the limiting distribution of M_n is not available, we simulate the critical values and use the MLRT as an efficiency barometer.

We suggest using the penalty function $p_n(\sigma) = -0.25\left\{s_n/\sigma^2 + \log(\sigma^2/s_n)\right\}$ which is almost the same as before, except for the coefficient because we have two penalty terms in this problem. Our simulation shows that this choice works well in terms of providing accurate type I errors. We use $p(\alpha) = \log(1-|1-2\alpha|)$ according to the recommendation of Li et al. (2008).

In the simulations, the type I errors were calculated based on 20,000 samples from N(0, 1). As in Section 2.2, we used two groups of initial values (0.1, 0.2, 0.3, 0.4, 0.5) and (0.1, 0.3, 0.5) to calculate $EM_n^{(K)}$. The simulation results are summarized in Table 4. The EM-test statistics based on (0.1, 0.3, 0.5) give accurate type I errors.

TABLE 4
Type I errors (%) of the EM-test.

Level	$EM_n^{(1)}$	$EM_n^{(2)}$	$EM_n^{(3)}$	$EM_n^{(1)}$	$EM_n^{(2)}$	$EM_n^{(3)}$
	$n=100$					
10%	10.8	10.9	10.9	10.5	10.6	10.6
5%	5.5	5.5	5.6	5.3	5.4	5.4
1%	1.2	1.2	1.2	1.1	1.2	1.2
	$n=200$					
10%	10.7	10.7	10.7	10.4	10.5	10.5
5%	5.4	5.4	5.4	5.1	5.2	5.2
1%	1.1	1.1	1.1	1.0	1.0	1.0

Results in columns (2, 3, 4) used $\alpha = (0.1, 0.2, 0.3, 0.4, 0.5)$
Results in columns (5, 6, 7) used $\alpha = (0.1, 0.3, 0.5)$

The powers of the EM-test and the MLRT for the models in Table 2 are calculated based on 5,000 repetitions and presented in Table 5. Since the limiting distribution of the MLRT is unavailable and hence is not a viable method, the simulated critical values were used for power calculation. The simulation results show that the $EM_n^{(2)}$ and $EM_n^{(3)}$ based on three initial values (0.1,0.3,0.5) for α have almost the same power as the MLRT. Further increasing the number of iterations or the number of initial values for α does not increase the power of the EM-test statistics. We hence recommend the use of $EM_n^{(2)}$ or $EM_n^{(3)}$ based on three initial values (0.1, 0.3, 0.5) for α.

We note that when $\sigma_1 = \sigma_2$, the current EM-test loses some power compared to the EM-test designed for finite mixture models in the presence of a structural parameter if the mixing parameter α is close to 0.5, but it has higher power when α is near 0 or 1. Nevertheless, we recommend the use of the current EM-test if the equal variance assumption is likely violated.

TABLE 5
Powers (%) of the EM-test and the MLRT at the 5% level.

Model	$EM_n^{(1)}$	$EM_n^{(2)}$	$EM_n^{(3)}$	$EM_n^{(1)}$	$EM_n^{(2)}$	$EM_n^{(3)}$	MLRT
	$n = 100$						
I	44.0	44.0	43.9	44.1	43.8	43.8	44.0
II	47.7	47.9	47.8	47.5	47.5	47.4	47.9
III	55.5	55.5	55.4	55.6	55.6	55.5	55.5
IV	56.9	56.9	56.8	57.4	56.9	56.8	56.8
V	58.6	58.4	58.4	58.8	58.8	58.7	58.2
VI	63.5	63.3	63.3	63.7	63.6	63.6	63.2
VII	66.8	66.6	66.6	66.9	66.8	66.8	66.6
VIII	67.3	67.2	67.2	67.4	67.2	67.1	67.4
IX	48.9	48.8	48.7	49.1	48.8	48.7	48.7
X	54.6	54.6	54.5	55.0	54.8	54.6	54.3
XI	56.5	56.5	56.5	57.0	56.6	56.6	56.3
XII	57.1	57.1	57.0	57.4	57.1	57.0	57.0
	$n = 200$						
I	78.3	78.2	78.2	78.3	78.2	78.2	78.2
II	82.0	81.9	81.9	82.2	82.1	82.1	81.9
III	88.6	88.6	88.6	88.8	88.7	88.7	88.5
IV	88.7	88.6	88.6	88.9	88.8	88.8	88.5
V	90.0	89.9	89.9	90.1	90.0	90.0	89.8
VI	91.6	91.5	91.5	91.7	91.6	91.6	91.5
VII	91.4	91.3	91.3	91.5	91.4	91.4	91.3
VIII	89.6	89.6	89.6	89.6	89.5	89.5	89.7
IX	81.7	81.5	81.5	81.9	81.8	81.7	81.4
X	88.1	88.0	88.0	88.3	88.2	88.1	87.9
XI	86.4	86.3	86.3	86.5	86.4	86.4	86.2
XII	87.7	87.6	87.6	87.9	87.9	87.8	87.5

Results in columns (2, 3, 4) used $\alpha = (0.1, 0.2, 0.3, 0.4, 0.5)$.
Results in columns (5, 6, 7) used $\alpha = (0.1, 0.3, 0.5)$.

4. Genetic Applications.

EXAMPLE 2. We apply the EM-test to the example discussed in Loisel et al. (1994). Due to the potential use for hybrid production, cytoplasmic male sterility in plant species is a trait of much scientific and economic interest. To efficiently use this character, it is important to find nuclear genes—preferably dominant ones—that induce fertility restoration (MacKenzie and Bassett, 1987). Loisel et al. (1994) carried out an experiment for detecting a major restoration gene. In this experiment, 150 F2 bean plants were obtained. The number of pods with one up to a maximum of ten grains were then counted on each F2 plant. Loisel et al. (1994) suggested analyzing the square root of the total number of grains for each plant. If a major restoration gene exists, the normal mixture model will provide a more suitable fit; otherwise the single normal distribution best fits the data. The histogram of the transformed counts is given in Figure 2. It indicates the existence of two modes, and an unequal variance normal mixture model is a good choice.

Based on some genetic background, Loisel et al. (1994) postulated a three-component normal mixture model:

$$\frac{1}{4}N(\theta_1,\sigma^2)+\frac{1}{2}N(\theta_2,\sigma^2)+\frac{1}{4}N(\theta_3,\sigma^2) \tag{4.1}$$

and tested the null hypothesis that $\theta_1 = \theta_2 = \theta_3$. They found that the limiting distribution of the LRT statistic is a 50-50 mixture of the χ^2_1 and χ^2_2, and the resulting p-value is 0.002%. We investigated the null rejection rates of the LRT under model (4.1) when $n = 150$ and the critical values were determined by a 50-50 mixture of the χ^2_1 and χ^2_2 limiting distributions. Based on 40,000 repetitions, the simulated null rejection rates were 15.6%, 8.8%, and 2.2% for nominal values of 10%, 5%, and 1%. The above p-value may be biased toward the liberal side.

For illustration purposes, we re-analyzed the data with the EM-test under model (1.1) with $\sigma_1^2 = \sigma_2^2$. The p-value of the MLRT calibrated with the χ^2_2 distribution was found to be 1.4%. We found $EM_n^{(2)} = 6.827$ with three initial values $(0.1, 0.3, 0.5)$ for α, corresponding to the p-value 1.0%. It can be seen that the EM-test provides stronger evidence against the null model than the MLRT test.

It appears that the equal variance assumption is not suitable. We consider the EM-test for a finite normal mixture with unequal variance. We found that $EM_n^{(1)} = 15.966$ and $EM_n^{(2)} = 20.590$ with three initial values $(0.1, 0.3, 0.5)$ for α, resulting in the p-values 0.03% and 0.003%, respectively. Further iteration does not change the p-value much. This result is in line with the outcome of Loisel et al. (1994). The modified MLES of $(\alpha, \theta_1, \theta_2, \sigma_1, \sigma_2)$ are (0.175, 10.663, 2.535, 3.203, 1.080), confirming that $\sigma_1 \neq \sigma_2$ and explaining why the EM-test under the general model gives much stronger evidence against the null model.

Figure 2 shows the fitted density functions of models (1.1) and (4.1). Our analysis indicates that a two-component mixture model can fit the data just as well as the model suggested by Loisel et al. (1994). The question of which model is more appropriate is not the focus of this paper.

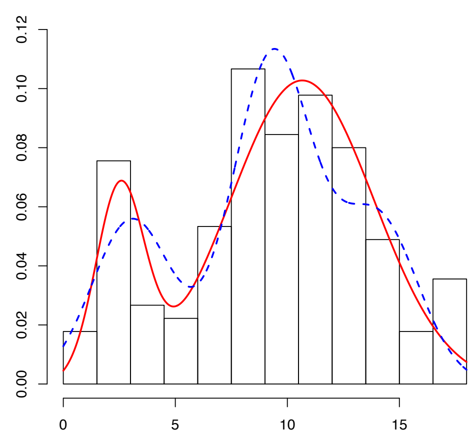

Fig 2. *The histogram of the square root of the total number of grains per plant, the fitted densities of the normal mixtures in (1.1) (solid line) and in (4.1) (dashed line).*

Example 3. The second example considers the data presented in Everitt et al. (2001); see part (b) of Table 6.2. This data set is from a schizophrenia study reported by Levine (1981), who collated the results of seven studies on the age of onset of schizophrenia including 99 females and 152 males. We use the male data to illustrate the use of the EM-test. As suggested by Levine (1981), there are two types of schizophrenia in males. The first type is diagnosed at a younger age and is generally more severe; the second type is diagnosed later in life. We wish to test the existence of the two types of schizophrenia.

Everitt et al. (2001) fitted the 152 observations using a two-component normal mixture model, and used the LRT to test the homogeneity. Using the χ^2_3 distribution for calibration, they found the p-value was less than 0.01%. Following Everitt (1996), our analysis is based on logarithmic transformed data. Assuming model (1.1) with $\sigma_1^2 = \sigma_2^2$, the p-value of the MLRT calibrated with the χ^2_2 distribution is 1.8%, but $EM_n^{(2)} = 0$ with three initial values $(0.1, 0.3, 0.5)$ for α.

Removing the $\sigma_1 = \sigma_2$ assumption, we find that $EM_n^{(1)} = 13.301$ and $EM_n^{(2)} = 13.323$ with three α initial values $(0.1, 0.3, 0.5)$ and both p-values are 0.1%. The modified MLEs of $(\alpha, \mu_1, \mu_2, \sigma_1, \sigma_2)$ are (0.448, 1.379, 1.319, 0.192, 0.071). Our analysis indicates that there are two subpopulations in the population with close mean-ages of onset but different variances. This also explains why the EM-test designed for finite mixture models in the presence of a structural parameter is insignificant.

Figure 3 contains the histogram and the fitted densities. It can be seen that the mixture model with unequal variances fits better. We also computed the LRT statistic which equals 15.27 under the unequal variance assumption. If it is calibrated with the χ^2_4 distribution, as suggested by Wolfe (1971), the p-value is 0.4%, and if calibrated with the χ^2_6 distribution, as suggested by McLachlan (1987), the p-value is 1.8% Without a solid theory, it would be hard to reconcile these inconsistent outcomes.

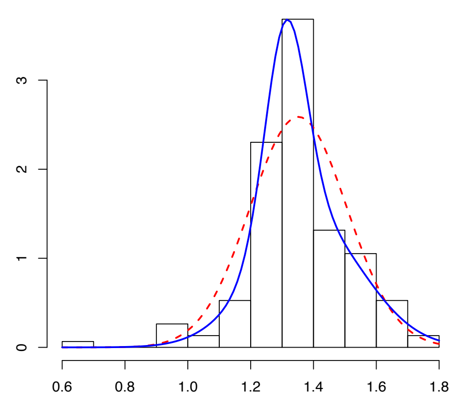

FIG 3. *The histogram of log age of onset for male schizophrenics, the fitted densities of the single normal model (dashed line) and normal mixture model in (1.1) (solid line).*

APPENDIX A: PROOFS

A.1. Proofs of Theorems 1 and 2. Without loss of generality, we assume that the null distribution is $N(0,1)$. A brief roadmap is as follows.

Lemma 1 shows that any estimator with α bounded away from 0 and 1, and with a large likelihood value, is consistent for θ_1, θ_2, and σ under the null model. Lemma 2 provides technical preparation for Lemma 3. Lemma 3 strengthens Lemma 1 by providing specific convergence rates. Lemma 4 makes Lemmas 1 and 3 applicable to $(\alpha_j^{(k)}, \theta_{1j}^{(k)}, \theta_{2j}^{(k)}, \sigma_j^{(k)})$, by showing that the EM-iteration keeps $\alpha_j^{(k)}$ away from 0 and 1. Theorems 1 and 2 then follow easily.

Two results, one from Chen et al. (2008) and another from Chen and Chen (2003), are stated as Lemmas A and B as follows. We include them here without proof for easy reference.

LEMMA A. *Except for a zero-probability event not depending on σ, and under the null model $N(0,1)$, we have for all large enough n,*

$$\sup_{\theta} \sum_{i=1}^{n} I(|X_i - \theta| \le |\sigma \log \sigma|) \le \begin{cases} 8n\{|\sigma \log \sigma| + n^{-1}\} & n^{-1} \le \sigma \le e^{-2} \\ 4(\log n)^2 & 0 < \sigma < n^{-1} \end{cases}.$$

That is, the number of observations in a small neighborhood of θ has the above upper bound, uniformly in θ.

The next lemma concerns the expansion of the likelihood function when θ and σ are in a small neighborhood of the true values. For $i = 1, \ldots, n$, let $Z_i = (X_i^2 - 1)/2$, $U_i = (X_i^3 - 3X_i)/6$, and $V_i = (X_i^4 - 6X_i^2 + 3)/24$.

LEMMA B. *Under the null model N(0, 1), and when $(\bar\theta_1, \bar\theta_2, \bar\sigma) = (0,0,1) + o_p(1)$, we have*

$$\begin{aligned} & l_n(\bar\alpha, \bar\theta_1, \bar\theta_2, \bar\sigma) - l_n(0.5, 0, 0, 1) \\ & \quad \le \bar s_1 \sum_{i=1}^n X_i + \bar s_2 \sum_{i=1}^n Z_i + \bar s_3 \sum_{i=1}^n U_i + \bar s_4 \sum_{i=1}^n V_i \\ & \qquad - \frac{1}{2}\{\bar s_1^2 \sum_{i=1}^n X_i^2 + \bar s_2^2 \sum_{i=1}^n Z_i^2 + \bar s_3^2 \sum_{i=1}^n U_i^2 + \bar s_4^2 \sum_{i=1}^n V_i^2\}\{1 + o_p(1)\} + o_p(1). \end{aligned}$$

LEMMA 1. *Let $(\bar\alpha, \bar\theta_1, \bar\theta_2, \bar\sigma)$ be any estimators of $(\alpha, \theta_1, \theta_2, \sigma)$ such that*

$$pl_n(\bar\alpha, \bar\theta_1, \bar\theta_2, \bar\sigma) - pl_n(0.5, 0, 0, 1) \ge c > -\infty$$

and $\bar\alpha \in [\delta, 1-\delta]$ *for some* $\delta \in (0, 0.5]$. *Under the null model* $N(0,1)$, $\bar\theta_1 = o_p(1)$, $\bar\theta_2 = o_p(1)$, *and* $\bar\sigma - 1 = o_p(1)$.

PROOF. We first show that, with probability approaching 1, $\bar\sigma$ has a non-zero lower bound. Then we apply the result in Kiefer and Wolfowitz (1956) to show the consistency of $(\bar\theta_1, \bar\theta_2, \bar\sigma)$.

Let $A = \{i : \min(|X_i - \theta_1|, |X_i - \theta_2|) < |\sigma \log \sigma|\}$, and let $n(A)$ be the number of indices in set A. For any index set, for instance A, we define

$$l_n(\alpha, \theta_1, \theta_2, \sigma; A) = \sum_{i \in A} \log\{(1-\alpha) f(X_i; \theta_1, \sigma) + \alpha f(X_i; \theta_2, \sigma)\}.$$

Since for any $i \in A$, the mixture density function $(1-\alpha) f(X_i; \theta_1, \sigma) + \alpha f(X_i; \theta_2, \sigma) \le (\sqrt{2\pi}\sigma)^{-1}$, we have

$$l_n(\alpha, \theta_1, \theta_2, \sigma; A) \le -n(A) \log(\sqrt{2\pi}\sigma). \tag{A.1}$$

For any $i \in A^c$, the complement of A,

$$(1-\alpha) f(X_i; \theta_1, \sigma) + \alpha f(X_i; \theta_2, \sigma) \le (\sqrt{2\pi}\sigma)^{-1} \exp\{-\frac{1}{2} \log^2 \sigma\}.$$

Hence,

$$l_n(\alpha, \theta_1, \theta_2, \sigma; A^c) \le -n(A^c)\{\log(\sqrt{2\pi}\sigma) + \frac{1}{2} \log^2 \sigma\}. \tag{A.2}$$

Combining (A.1) and (A.2), we have

$$l_n(\alpha, \theta_1, \theta_2, \sigma) \le -n \log(\sqrt{2\pi}\sigma) - \frac{1}{2}\{n - n(A)\} \log^2 \sigma.$$

This further implies that

$$l_n(\alpha, \theta_1, \theta_2, \sigma) - l_n(0.5, 0, 0, 1) \le \frac{1}{2} \sum_{i=1}^n X_i^2 - n \log(\sigma) - \frac{1}{2}\{n - n(A)\} \log^2 \sigma.$$

According to Lemma A, it is easy to check that there exists a non-random $\epsilon > 0$ not depending on n such that for large enough n and $\sigma < \epsilon$, we have $n(A) \le n/2$. Consequently, as $n \to \infty$,

$$l_n(\alpha, \theta_1, \theta_2, \sigma) - l_n(0.5, 0, 0, 1) \le \frac{1}{2} \sum_{i=1}^n X_i^2 - n\{\log(\sigma) + \frac{1}{4} \log^2 \sigma\}.$$

Let ϵ be small enough such that for all $\sigma < \epsilon$,

$$\log(\sigma) + \frac{1}{4}\log^2 \sigma \geq 1.$$

Hence, we must have

$$l_n(\alpha, \theta_1, \theta_2, \sigma) - l_n(0.5, 0, 0, 1) \leq -\frac{n}{2} + o(n).$$

Since the penalty is $o(n)$ by Conditions C0 and C1, this furthermore implies that uniformly over the parameter space such that $\sigma < \epsilon$,

$$pl_n(\alpha, \theta_1, \theta_2, \sigma) - pl_n(0.5, 0, 0, 1) \leq -\frac{n}{2} + o(n).$$

Thus for any estimator $(\bar{\alpha}, \bar{\theta}_1, \bar{\theta}_2, \bar{\sigma})$ such that $pl_n(\bar{\alpha}, \bar{\theta}_1, \bar{\theta}_2, \bar{\sigma}) - pl_n(0.5, 0, 0, 1) \geq c > -\infty$ for all n, we must have

$$\lim_{n\to\infty} P(\epsilon \leq \bar{\sigma}) = 1.$$

This result is equivalent to placing a positive constant lower bound for the variance parameter for searching the maximal value of $pl_n(\alpha, \theta_1, \theta_2, \sigma)$. Thus, the consistency of $(\bar{\theta}_1, \bar{\theta}_2, \bar{\sigma})$ is covered by the result in Kiefer and Wolfowitz (1956). Note that their proof can be modified to accommodate a penalty of size $o(n)$. □

Let $\bar{m}_j = (1-\bar{\alpha})\bar{\theta}_1^j + \bar{\alpha}\bar{\theta}_2^j$, $j = 1, 2, 3, 4$ be the first four moments of the mixing distribution, and let

$$\bar{s}_1 = \bar{m}_1, \bar{s}_2 = \bar{m}_2 + \bar{\sigma}^2 - 1, \bar{s}_3 = \bar{m}_3, \text{ and } \bar{s}_4 = \bar{m}_4 - 3\bar{m}_2^2. \tag{A.3}$$

The following lemma provides technical preparation for Lemma 3.

LEMMA 2. *Under the same conditions as in Lemma 1 we have*

$$\bar{\theta}_1^4 = O_p\Big(\sum_{j=1}^4 |\bar{s}_j|\Big),\ \bar{\theta}_2^4 = O_p\Big(\sum_{j=1}^4 |\bar{s}_j|\Big),\ \textit{and } (\bar{\sigma}^2 - 1)^2 = O_p\Big(\sum_{j=1}^4 |\bar{s}_j|\Big).$$

PROOF. It is obvious that $\bar{s}_j = o_p(1)$ for all $j = 1, 2, 3, 4$. Write

$$\bar{\theta}_1 = \frac{\bar{s}_1 - \alpha\bar{\theta}_2}{1-\bar{\alpha}} \tag{A.4}$$

Substituting (A.4) into the expression for $\bar{s}_3$, we obtain

$$\bar{s}_3 = \frac{\bar{\alpha}(1-2\bar{\alpha})}{(1-\bar{\alpha})^2}\bar{\theta}_2^3 + o_p(\bar{s}_1).$$

Furthermore, this implies that

$$\begin{aligned}
\bar{s}_4 &= \bar{m}_4 - 3\bar{m}_2^2 \\
&= \frac{\bar{\alpha}\{\alpha^3 + (1-\alpha)^3\}}{(1-\bar{\alpha})^3}\bar{\theta}_2^4 - \frac{3\bar{\alpha}^2}{(1-\bar{\alpha})^2}\bar{\theta}_2^4 + o_p(\bar{s}_1) \\
&= \frac{\bar{\alpha}(1-6\bar{\alpha}+6\bar{\alpha}^2)}{(1-\bar{\alpha})^3}\theta_2^4 + o_p(\bar{s}_1) \\
&= -\frac{\bar{\alpha}}{2(1-\bar{\alpha})^3}\bar{\theta}_2^4 + \frac{3(1-2\bar{\alpha})}{2(1-\bar{\alpha})}\bar{\theta}_2 s_3 + o_p(\bar{s}_1) \\
&= -\frac{\bar{\alpha}}{2(1-\bar{\alpha})^3}\bar{\theta}_2^4 + o_p(\bar{s}_1) + o_p(\bar{s}_3). \qquad \text{(A.5)}
\end{aligned}$$

The coefficient of $\bar{\theta}_2^4$ is bounded away from 0 because $\delta < \bar{\alpha} < 1-\delta$ for some positive δ, and this implies that

$$\bar{\theta}_2^4 = O_p(\bar{s}_4) + o_p(\bar{s}_1) + o_p(\bar{s}_3) = O_p\Big(\sum_{j=1}^4 |\bar{s}_j|\Big)$$

and by symmetry, $\bar{\theta}_1^4 = O_p\Big(\sum_{j=1}^4 |\bar{s}_j|\Big)$. Furthermore,

$$(\bar{\sigma}^2-1)^2 = (\bar{s}_2 - \bar{m}_2)^2 \le 2\bar{s}_2^2 + 2\bar{m}_2^2 = O_p\Big(\sum_{j=1}^4 |\bar{s}_j|\Big).$$

□

The next lemma strengthens the results of Lemma 1 and more.

LEMMA 3. *Under the same conditions as in Lemma 1, and if* $\bar{\alpha} - \alpha_0 = o_p(1)$ *for some* $\alpha_0 \in (0, 0.5]$, *then* $\bar{s}_j = O_p(n^{-1/2})$ *for* $j = 1, 2, 3, 4$, *and*

$$\bar{\theta}_1 = O_p(n^{-1/8}), \quad \bar{\theta}_2 = O_p(n^{-1/8}), \quad \textit{and} \;\; \bar{\sigma}^2 - 1 = O_p(n^{-1/4}).$$

Furthermore, when $0 < \alpha_0 < 0.5$, *the orders are refined to*

$$\bar{\theta}_1 = O_p(n^{-1/6}), \quad \bar{\theta}_2 = O_p(n^{-1/6}), \quad \textit{and} \;\; \bar{\sigma}^2 - 1 = O_p(n^{-1/3}).$$

PROOF. Expanding $l_n(\bar\alpha, \bar\theta_1, \bar\theta_2, \bar\sigma)$ at $(\bar\theta_1, \bar\theta_2, \bar\sigma) = (0, 0, 1)$, we find

$$\begin{aligned}
-\infty < c \;&\le\; pl_n(\bar\alpha, \bar\theta_1, \bar\theta_2, \bar\sigma) - pl_n(0.5, 0, 0, 1)\\
&= l_n(\bar\alpha, \bar\theta_1, \bar\theta_2, \bar\sigma) - l_n(0.5, 0, 0, 1)\\
&\quad + \{p_n(\bar\sigma) - p_n(1)\} + \{p(\bar\alpha) - p(0.5)\}\\
&\le \bar s_1 \sum_{i=1}^n X_i + \bar s_2 \sum_{i=1}^n Z_i + \bar s_3 \sum_{i=1}^n U_i + \bar s_4 \sum_{i=1}^n V_i\\
&\quad - \frac{1}{2}\{\bar s_1^2 \sum_{i=1}^n X_i^2 + \bar s_2^2 \sum_{i=1}^n Z_i^2 + \bar s_3^2 \sum_{i=1}^n U_i^2 + \bar s_4^2 \sum_{i=1}^n V_i^2\}\{1 + o_p(1)\}\\
&\quad + \{p(\alpha_0) - p(0.5)\} + o_p(1). \qquad \text{(A.6)}
\end{aligned}$$

In the above, we used Lemma B and the fact that the penalty term

$$p(\bar\alpha) - p(0.5) = p(\alpha_0) - p(0.5) + o_p(1),$$

and that

$$p_n(\bar\sigma) - p_n(1) = (\bar\sigma^2 - 1)o_p(n^{1/4}) \le o_p\{1 + n(\bar\sigma^2 - 1)^4\}$$

which is a higher order term compared to the quadratic part of the expansion. The reason for expanding up to the fourth term is the loss of strong identifiability. Since

$$\bar s_1 \sum_{i=1}^n X_i - \frac{1}{2}\bar s_1^2 \sum_{i=1}^n X_i^2\{1 + o_p(1)\} \le \frac{(\sum_{i=1}^n X_i)^2}{2\sum_{i=1}^n X_i^2}\{1 + o_p(1)\} = O_p(1),$$

and similarly for the other terms, we must have

$$\bar s_1 \sum_{i=1}^n X_i - \frac{1}{2}\bar s_1^2\{\sum_{i=1}^n X_i^2\} = O_p(1).$$

This is possible only if $\bar s_1 = O_p(n^{-1/2})$ and similarly $\bar s_j = O_p(n^{-1/2})$, for $j = 2, 3, 4$.

Next, by Lemma 2, the order results on $\bar s_j$ imply that

$$\bar\theta_1 = O_p(n^{-1/8}),\ \bar\theta_2 = O_p(n^{-1/8}),\ \text{and}\ \bar\sigma^2 - 1 = O_p(n^{-1/4}).$$

Recall that

$$\bar s_3 = \frac{\bar\alpha(1 - 2\bar\alpha)}{(1 - \bar\alpha)^2}\bar\theta_2^3 + o_p(\bar s_1).$$

Hence when $0 < \alpha_0 < 0.5$, the order assessment is refined to $\bar\theta_1 = O_p(n^{-1/6})$, $\bar\theta_2 = O_p(n^{-1/6})$, and $\bar\sigma^2 - 1 = O_p(n^{-1/3})$. □

Now we show that under the null model, the EM-iteration changes the fitted value of α by no more than an $O_p(n^{-1/4})$ quantity. Let $(\bar{\alpha}, \bar{\theta}_1, \bar{\theta}_2, \bar{\sigma})$ be some estimators of $(\alpha, \theta_1, \theta_2, \sigma)$ as before, and let

$$\bar{w}_i = \frac{\bar{\alpha} f(X_i; \bar{\theta}_2, \bar{\sigma})}{(1-\bar{\alpha}) f(X_i; \bar{\theta}_1, \bar{\sigma}) + \bar{\alpha} f(X_i; \bar{\theta}_2, \bar{\sigma})}.$$

We further define

$$R_n(\alpha) = (n - \sum_{i=1}^{n} \bar{w}_i) \log(1-\alpha) + \sum_{i=1}^{n} \bar{w}_i \log(\alpha)$$

and $H_n(\alpha) = R_n(\alpha) + p(\alpha)$. The EM-test updates α by searching for $\bar{\alpha}^* = \arg\max_\alpha H_n(\alpha)$.

LEMMA 4. *Under the same conditions as in Lemma 1, and if $\bar{\alpha} - \alpha_0 = O_p(n^{-1/4})$ for some $\alpha_0 \in (0,1)$, then $\bar{\alpha}^* - \alpha_0 = O_p(n^{-1/4})$.*

PROOF. Putting $\hat{\alpha} = n^{-1} \sum_{i=1}^{n} \bar{w}_i$ which is the maximum point of $R_n(\alpha)$, we find

$$|\hat{\alpha} - \bar{\alpha}| = \frac{\bar{\alpha}(1-\bar{\alpha})}{n} \Big| \sum_{i=1}^{n} \frac{f(X_i; \bar{\theta}_2, \bar{\sigma}) - f(X_i; \bar{\theta}_1, \bar{\sigma})}{(1-\bar{\alpha}) f(X_i; \bar{\theta}_1, \bar{\sigma}) + \bar{\alpha} f(X_i; \bar{\theta}_2, \bar{\sigma})} \Big|. \quad \text{(A.7)}$$

By Lemma 1, $(\bar{\theta}_1, \bar{\theta}_2, \bar{\sigma})$ are in a small neighborhood of $(0,0,1)$ in probability. Therefore, expanding (A.7) at $(0,0,1)$, we get

$$|\hat{\alpha} - \bar{\alpha}| = n^{-1} \bar{\alpha}(1-\bar{\alpha}) \Big| (\bar{\theta}_2 - \bar{\theta}_1) \sum_{i=1}^{n} X_i + O_p(n) \{\bar{\theta}_1^2 + \bar{\theta}_2^2 + (\bar{\sigma}^2 - 1)^2\} \Big| = O_p(n^{-1/4})$$

using the order results of Lemma 3. We hence obtain $\hat{\alpha} - \alpha_0 = O_p(n^{-1/4})$. Thus, the lemma is true if $\bar{\alpha}^* - \hat{\alpha} = O_p(n^{-1/4})$.

First, note that $R_n(\alpha)$ is a binomial log-likelihood. It attains its maximum at and decreases from $\hat{\alpha}$ in both directions. For any $\epsilon > 0$ and $\alpha \geq \hat{\alpha} + 2\epsilon$, by the mean value theorem,

$$R_n(\alpha) - R_n(\hat{\alpha}) \leq R_n(\hat{\alpha} + 2\epsilon) - R_n(\hat{\alpha} + \epsilon) = \epsilon R_n'(\xi)$$

for some $\xi \in [\hat{\alpha} + \epsilon, \hat{\alpha} + 2\epsilon]$. It is easy to verify that $R_n'(\xi) \to -\infty$ in probability as $n \to \infty$ uniformly for ξ in this range. On the other hand, we have

$$p(\alpha) - p(\hat{\alpha}) \leq p(0.5) - p(\alpha_0) + o_p(1) = O_p(1).$$

Hence, with probability approaching 1,

$$H_n(\alpha) - H_n(\hat{\alpha}) = R_n(\alpha) - R_n(\hat{\alpha}) + \{p(\alpha) - p(\hat{\alpha})\} \to -\infty$$

uniformly for any $\alpha > \hat{\alpha}+2\epsilon$. Hence, we must have $\bar{\alpha}^* < \hat{\alpha}+2\epsilon$ in probability. Similarly, we can show that $\bar{\alpha}^* > \hat{\alpha} - 2\epsilon$ in probability. Therefore, we have $\bar{\alpha}^* = \hat{\alpha} + o_p(1)$.

Next, noting that $R_n'(\hat{\alpha}) = 0$, $p(\bar{\alpha}^*) - p(\hat{\alpha}) = o_p(1)$, we find

$$0 < H_n(\bar{\alpha}^*) - H_n(\hat{\alpha}) = R_n(\bar{\alpha}^*) - R_n(\hat{\alpha}) + o_p(1) = \frac{1}{2} R_n''(\eta)(\bar{\alpha}^* - \hat{\alpha})^2 + o_p(1)$$

by Taylor's expansion for some $\eta \in [\hat{\alpha}, \bar{\alpha}^*]$. Since

$$-R_n''(\eta) = \frac{n}{(1-\eta)^2}(1-\hat{\alpha}) + \frac{n}{\eta^2}\hat{\alpha} = \frac{n}{\alpha_0(1-\alpha_0)}\{1 + o_p(1)\}$$

is of order n, we must have

$$(\bar{\alpha}^* - \hat{\alpha})^2 = o_p(n^{-1}).$$

This clearly suffices. □

LEMMA 5. *Suppose the same conditions as in Lemma 1 hold.*

(*a*) *If* $\bar{\alpha} - 0.5 = O_p(n^{-1/4})$, *then*

$$\begin{aligned} &2\{pl_n(\bar{\alpha}, \bar{\theta}_1, \bar{\theta}_2, \bar{\sigma}) - pl_n(0.5, 0, 0, 1)\} \\ &\quad \le \frac{(\sum_{i=1}^n X_i)^2}{\sum_{i=1}^n X_i^2} + \frac{(\sum_{i=1}^n Z_i)^2}{\sum_{i=1}^n Z_i^2} + \frac{\{(\sum_{i=1}^n V_i)^-\}^2}{\sum_{i=1}^n V_i^2} + o_p(1), \end{aligned}$$

where $(\sum_{i=1}^n V_i)^-$ *means the negative part of* $\sum_{i=1}^n V_i$.

(*b*) *If* $\bar{\alpha} - \alpha_0 = o_p(1)$ *for some* $\alpha_0 \in (0, 0.5)$, *then*

$$\begin{aligned} &2\{pl_n(\bar{\alpha}, \bar{\theta}_1, \bar{\theta}_2, \bar{\sigma}) - pl_n(0.5, 0, 0, 1)\} \\ &\quad \le \frac{(\sum_{i=1}^n X_i)^2}{\sum_{i=1}^n X_i^2} + \frac{(\sum_{i=1}^n Z_i)^2}{\sum_{i=1}^n Z_i^2} + \frac{(\sum_{i=1}^n U_i)^2}{\sum_{i=1}^n U_i^2} + 2\{p(\alpha_0) - p(0.5)\} + o_p(1). \end{aligned}$$

PROOF. (a). When $\bar{\alpha} - 0.5 = O_p(n^{-1/4})$, we have $\bar{\theta}_2 = O_p(n^{-1/8})$ and hence

$$\bar{s}_3 = \frac{\bar{\alpha}(1-2\bar{\alpha})}{(1-\bar{\alpha})^2}\bar{\theta}_2^3 + o_p(s_1) = o_p(n^{-1/2}).$$

Consequently, the terms in (A.6) containing $\bar{s}_3$ are $o_p(1)$. Furthermore this value of $\bar{\alpha}$ makes $\bar{s}_4 < 0$. Hence, the upper bound (A.6) simplifies to

$$\begin{aligned} & 2\{pl_n(\bar{\alpha}, \bar{\theta}_1, \bar{\theta}_2, \bar{\sigma}) - pl_n(0.5, 0, 0, 1)\} \\ \leq & \ 2\{\bar{s}_1 \sum_{i=1}^n X_i + \bar{s}_2 \sum_{i=1}^n Z_i + \bar{s}_4 \sum_{i=1}^n V_i\} \\ & -\{\bar{s}_1^2 \sum_{i=1}^n X_i^2 + \bar{s}_2^2 \sum_{i=1}^n Z_i^2 + \bar{s}_4^2 \sum_{i=1}^n V_i^2\}\{1 + o_p(1)\} + o_p(1) \\ \leq & \ \frac{(\sum_{i=1}^n X_i)^2}{\sum_{i=1}^n X_i^2} + \frac{(\sum_{i=1}^n Z_i)^2}{\sum_{i=1}^n Z_i^2} + \frac{\{(\sum_{i=1}^n V_i)^-\}^2}{\sum_{i=1}^n V_i^2} + o_p(1). \end{aligned}$$

(b) When $\bar{\alpha} - \alpha_0 = o_p(1)$ for some $\alpha_0 \in (0, 0.5)$, we have $\bar{\theta}_1 = O_p(n^{-1/6})$, $\bar{\theta}_2 = O_p(n^{-1/6})$, and $\bar{\sigma}^2 - 1 = O_p(n^{-1/3})$ by Lemma 3. These imply that $\bar{s}_4 = o_p(n^{-1/2})$. Hence, the s_4 terms in (A.6) are $o_p(1)$. The upper bound in (A.6) simplifies to

$$\begin{aligned} & 2\{pl_n(\bar{\alpha}, \bar{\theta}_1, \bar{\theta}_2, \bar{\sigma}) - pl_n(0.5, 0, 0, 1)\} \\ \leq & \ 2\{\bar{s}_1 \sum_{i=1}^n X_i + \bar{s}_2 \sum_{i=1}^n Z_i + \bar{s}_3 \sum_{i=1}^n U_i\} \\ & -\{\bar{s}_1^2 \sum_{i=1}^n X_i^2 + \bar{s}_2^2 \sum_{i=1}^n Z_i^2 + \bar{s}_3^2 \sum_{i=1}^n U_i^2\}\{1 + o_p(1)\} + 2\{p(\alpha_0) - p(0.5)\} + o_p(1) \\ \leq & \ \frac{(\sum_{i=1}^n X_i)^2}{\sum_{i=1}^n X_i^2} + \frac{(\sum_{i=1}^n Z_i)^2}{\sum_{i=1}^n Z_i^2} + \frac{(\sum_{i=1}^n U_i)^2}{\sum_{i=1}^n U_i^2} + 2\{p(\alpha_0) - p(0.5)\} + o_p(1). \end{aligned}$$

□

We now prove Theorems 1 and 2 by showing that the slightly more general results in the previous lemmas are applicable.

Proof of Theorem 1

For any $k \leq K$, due to the property of the EM-algorithm that the likelihood increases after each iteration (Dempster et al., 1977; Wu, 1981; McLachlan and Krishnan, 1997), we have

$$pl_n(\alpha_j^{(k)}, \theta_{j1}^{(k)}, \theta_{j2}^{(k)}, \sigma_j^{(k)}) \geq pl_n(\alpha_j^{(1)}, \theta_{j1}^{(1)}, \theta_{j2}^{(1)}, \sigma_j^{(1)}) \geq pl_n(\alpha_j, 0, 0, 1).$$

That is,

$$pl_n(\alpha_j^{(k)}, \theta_{j1}^{(k)}, \theta_{j2}^{(k)}, \sigma_j^{(k)}) - pl_n(0.5, 0, 0, 1) \geq p(\alpha_j) - p(0.5) > -\infty.$$

Furthermore, by Lemma 4 and applying mathematical induction in k, we find that

$$\alpha_j^{(k)} - \alpha_j = O_p(n^{-1/4}).$$

Hence the rate conclusions in Lemma 3 are applicable to $(\alpha_j^{(k)}, \theta_{j1}^{(k)}, \theta_{j2}^{(k)}, \sigma_j^{(k)})$ for any finite k and hence we get the conclusion of this theorem. □

Proof of Theorem 2

Expanding the log-likelihood function, we have that

$$2\{\sup_{\theta,\sigma} l_n(0.5, \theta, \theta, \sigma) - l_n(0.5, 0, 0, 1)\} = \frac{(\sum_{i=1}^n X_i)^2}{\sum_{i=1}^n X_i^2} + \frac{(\sum_{i=1}^n Z_i)^2}{\sum_{i=1}^n Z_i^2} + o_p(1).$$

By Theorem 1, $\sigma_j^{(k)} - 1 = O_p(n^{-1/4})$ for any $k \le K$. Hence the penalty $p_n(\sigma_j^{(k)}) - p_n(1) = o_p(1)$ by Condition C2, and the above expansion remains unchanged when $l_n(\alpha, \theta_1, \theta_2, \sigma)$ is replaced by $pl_n(\alpha, \theta_1, \theta_2, \sigma)$.

Due to the properties established in Theorem 1, $(\alpha_j^{(K)}, \theta_{j1}^{(K)}, \theta_{j2}^{(K)}, \sigma^{(K)})$ satisfies the conditions of Lemma 5 and hence for $\alpha_j = 0.5$,

$$\begin{aligned} M_n^{(K)}(0.5) &= 2\{pl_n(\alpha_j^{(K)}, \theta_{j1}^{(K)}, \theta_{j2}^{(K)}, \sigma^{(K)}) - pl_n(0.5, 0, 0, 1)\} \\ &\quad -2\{\sup_{\theta,\sigma} pl_n(0.5, \theta, \theta, \sigma) - pl_n(0.5, 0, 0, 1)\} \\ &\le \frac{\{(\sum_{i=1}^n V_i)^-\}^2}{\sum_{i=1}^n V_i^2} + o_p(1). \end{aligned}$$

Now we show that the $M_n^{(K)}(0.5)$ asymptotically equals this upper bound. Since the EM-iteration increases the penalized likelihood, we need only show that for given $\alpha = 0.5$, there exists a set of feasible $\theta_1, \theta_2, \sigma$ values at which the upper bound in Part (a) of Lemma 5 becomes equality. Consequently, the inequality for $M_n^{(K)}(0.5)$ will also become equality. This is equivalent to finding a set of values such that $s_j = \hat{s}_j + o_p(n^{-1/2})$, for $j = 1, 2, 4$, where s_j are defined in (A.3) and

$$\hat{s}_1 = \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n X_i^2}, \ \hat{s}_2 = \frac{\sum_{i=1}^n Z_i}{\sum_{i=1}^n Z_i^2}, \ \text{ and } \hat{s}_4 = -\frac{(\sum_{i=1}^n V_i)^-}{\sum_{i=1}^n V_i^2}.$$

Ignoring $o_p(n^{-1/2})$ terms, these equations are

$$\text{(A.8)} \qquad \left\{ \begin{array}{rcl} \frac{1}{2}(\theta_1 + \theta_2) &=& \hat{s}_1 \\ \frac{1}{2}(\theta_1^2 + \theta_2^2) + \sigma^2 - 1 &=& \hat{s}_2 \\ \frac{1}{2}(\theta_1^4 + \theta_2^4) - \frac{3}{4}(\theta_1^2 + \theta_2^2)^2 &=& \hat{s}_4 \end{array} \right. ,$$

Furthermore, we may replace the third equation by (A.5): $-2\theta_2^4 = \hat{s}_4$. Let $\tilde{\theta}_2$ be the non-negative solution. We then solve (A.8) to get $\tilde{\theta}_1$ and $\tilde{\sigma}$.

Additional simple algebra shows that the above solutions satisfy $\tilde{s}_j = \hat{s}_j + o_p(n^{-1/2})$, for $j = 1, 2, 4$, $\tilde{\theta}_1 = O_p(n^{-1/8})$, $\tilde{\theta}_2 = O_p(n^{-1/8})$, and $\tilde{\sigma}^2 - 1 = O_p(n^{-1/4})$. With this order information, we can easily expand and obtain

$$
\begin{aligned}
&2\{pl_n(0.5, \tilde{\theta}_1, \tilde{\theta}_2, \tilde{\sigma}) - pl_n(0.5, 0, 0, 1)\} \\
&\quad = \frac{(\sum_{i=1}^n X_i)^2}{\sum_{i=1}^n X_i^2} + \frac{(\sum_{i=1}^n Z_i)^2}{\sum_{i=1}^n Z_i^2} + \frac{\{(\sum_{i=1}^n V_i)^-\}^2}{\sum_{i=1}^n V_i^2} + o_p(1)
\end{aligned}
$$

and hence

$$
2\{pl_n(0.5, \tilde{\theta}_1, \tilde{\theta}_2, \tilde{\sigma}_1, \tilde{\sigma}_2) - pl_n(0.5, \hat{\theta}_0, \hat{\theta}_0, \hat{\sigma}_0, \hat{\sigma}_0)\} = \frac{\{(\sum_{i=1}^n V_i)^-\}^2}{\sum_{i=1}^n V_i^2} + o_p(1).
$$

Because the EM-iteration always increases the penalized likelihood, we must have

$$
\begin{aligned}
M_n^{(K)}(0.5) &\geq 2\{\sup_{(\theta_1,\theta_2,\sigma_1,\sigma_2)} pl_n(0.5, \theta_1, \theta_2, \sigma_1, \sigma_2) - pl_n(0.5, \hat{\theta}_0, \hat{\theta}_0, \hat{\sigma}_0, \hat{\sigma}_0)\} \\
&\geq 2\{pl_n(\tilde{\alpha}, \tilde{\theta}_1, \tilde{\theta}_2, \tilde{\sigma}_1, \tilde{\sigma}_2) - pl_n(0.5, \hat{\theta}_0, \hat{\theta}_0, \hat{\sigma}_0, \hat{\sigma}_0)\} \\
&= \frac{\{(\sum_{i=1}^n V_i)^-\}^2}{\sum_{i=1}^n V_i^2} + o_p(1).
\end{aligned}
$$

That is, the asymptotic upper bound of $M_n^{(K)}(0.5)$ is also a lower bound. Hence

$$
M_n^{(K)}(0.5) = \frac{\{(\sum_{i=1}^n V_i)^-\}^2}{\sum_{i=1}^n V_i^2} + o_p(1).
$$

Similarly for $\alpha_j \neq 0.5$

$$
M_n^{(K)}(\alpha_j) = \frac{(\sum_{i=1}^n U_i)^2}{\sum_{i=1}^n U_i^2} + 2\{p(\alpha_j) - p(0.5)\} + o_p(1).
$$

Therefore

$$
EM_n^{(K)} = \max\left[\frac{(\sum_{i=1}^n U_i)^2}{\sum_{i=1}^n U_i^2} + \Delta, \frac{\{(\sum_{i=1}^n V_i)^-\}^2}{\sum_{i=1}^n V_i^2}\right] + o_p(1).
$$

It is easy to verify that $\sum_{i=1}^n U_i/\sqrt{n}$ and $\sum_{i=1}^n V_i/\sqrt{n}$ are jointly asymptotical bivariate normal and independent. Consequently, the limiting distribution is given by $F(x - \Delta)\{0.5 + 0.5F(x)\}$ with $F(x)$ being the cdf of the χ_1^2 distribution. □

A.2. Proof of Theorems 3 and 4. We first prove some general results and then apply these results to show Theorems 3 and 4. Without loss of generality, we assume the null distribution is $N(0,1)$.

LEMMA 6. *Let $(\bar{\alpha}, \bar{\theta}_1, \bar{\theta}_2, \bar{\sigma}_1, \bar{\sigma}_2)$ be any estimators of $(\alpha, \theta_1, \theta_2, \sigma_1, \sigma_2)$ such that $\delta \le \bar{\alpha} \le 1-\delta$ for some $\delta \in (0, 0.5]$. Assume that*

$$pl_n(\bar{\alpha}, \bar{\theta}_1, \bar{\theta}_2, \bar{\sigma}_1, \bar{\sigma}_2) - pl_n(0.5, 0, 0, 1, 1) \ge c > -\infty.$$

Then under null distribution $N(0,1)$, $\bar{\theta}_h = o_p(1)$ and $\bar{\sigma}_h - 1 = o_p(1)$ for $h = 1, 2$.

The next lemma states that the EM-iteration changes the fitted value of α by no more than an $o_p(1)$ quantity. First we define $R_n(\alpha)$, $H_n(\alpha)$, $\bar{w}_i$, and $\bar{\alpha}^* = \arg\max_\alpha H_n(\alpha)$ as in the case where $\sigma_1 = \sigma_2$.

LEMMA 7. *Assume the same conditions as in Lemma 6. If $|\bar{\alpha} - \alpha_0| = o_p(1)$ for some $\alpha_0 \in (0,1)$, then under the null distribution $N(0,1)$ we have $|\bar{\alpha}^* - \alpha_0| = o_p(1)$.*

The proofs for the above two lemmas are the same as those for Lemmas 1 and 4, respectively, and are therefore omitted.

Proof of Theorem 3

For any $k \le K$, due to the property of the EM-algorithm that the likelihood increases after each iteration (Dempster et al., 1977; Wu, 1981; McLachlan and Krishnan, 1997), we have

$$pl_n(\alpha_j^{(k)}, \theta_{j1}^{(k)}, \theta_{j2}^{(k)}, \sigma_{j1}^{(k)}, \sigma_{j2}^{(k)}) \ge pl_n(\alpha_j^{(1)}, \theta_{j1}^{(1)}, \theta_{j2}^{(1)}, \sigma_{j1}^{(1)}, \sigma_{j2}^{(1)}) \ge pl_n(\alpha_j, 0, 0, 1, 1).$$

Therefore

$$pl_n(\alpha_j^{(k)}, \theta_{j1}^{(k)}, \theta_{j2}^{(k)}, \sigma_{j1}^{(k)}, \sigma_{j2}^{(k)}) - pl_n(0.5, 0, 0, 1, 1) \quad \ge \quad p(\alpha_j) - p(0.5) > -\infty.$$

According to Lemma 6, this property implies the consistency of $\theta_{jh}^{(k)}$ and $\sigma_{jh}^{(k)}$, $h = 1, 2$ whenever $\alpha_j^{(k)} - \alpha_j = o_p(1)$ is firmly established.

This key requirement, $|\alpha_j^{(k)} - \alpha_0| = o_p(1)$, is the result of applying Lemma 7. Hence, the conclusions of the Theorem are proved. □

Theorem 4 derives the asymptotic distribution of the EM-test. Its proof is long. We give some preparatory derivations as well as some preparatory lemmas.

Suppose $(\bar\alpha, \bar\theta_1, \bar\theta_2, \bar\sigma_1, \bar\sigma_2)$, such that $\delta \le \bar\alpha \le 1-\delta$ for some $\delta \in (0, 0.5]$, are the consistent estimators of $(\theta_1, \theta_2, \sigma_1, \sigma_2)$ under the null model $N(0,1)$. That is, $\bar\theta_h = o_p(1)$ and $\bar\sigma_h = 1 + o_p(1)$, $h = 1, 2$.

Denote

$$\begin{aligned} r_{1n}(\bar\alpha, \bar\theta_1, \bar\theta_2, \bar\sigma_1, \bar\sigma_2) &= 2\{l_n(\bar\alpha, \bar\theta_1, \bar\theta_2, \bar\sigma_1, \bar\sigma_2) - l_n(0.5, 0, 0, 1, 1)\}, \\ r_{2n} &= 2\{l_n(0.5, 0, 0, 1, 1) - l_n(0.5, \hat\theta_0, \hat\theta_0, \hat\sigma_0, \hat\sigma_0)\}, \\ r_{3n}(\bar\sigma_1, \bar\sigma_2) &= 2\{p_n(\bar\sigma_1) - p_n(1) + p_n(\bar\sigma_2) - p_n(1)\} + 2\{p(\bar\alpha) - p(0.5)\}. \end{aligned}$$

Their sum resembles the usual likelihood ratio statistic with penalties.

Put $r_{1n} = r_{1n}(\bar\alpha, \bar\theta_1, \bar\theta_2, \bar\sigma_1, \bar\sigma_2) = 2\sum_{i=1}^n \log(1+\bar\delta_i)$ with

$$\bar\delta_i = (1-\bar\alpha)\Big\{\frac{f(X_i, \bar\theta_1, \bar\sigma_1)}{f(X_i; 0, 1)} - 1\Big\} + \bar\alpha\Big\{\frac{f(X_i, \bar\theta_2, \bar\sigma_2)}{f(X_i; 0, 1)} - 1\Big\}.$$

As an aside we note that, in contrast to the EM-test, most other likelihood approaches do not naturally confine the fitted values $\bar\sigma_1$ and $\bar\sigma_2$ to a small neighborhood of the true variance. Consequently, their derivations are valid only if

$$E\Big\{\frac{f(X_i, \bar\theta_1, \bar\sigma_1)}{f(X_i; 0, 1)} - 1\Big\}^2 < \infty$$

for any nonrandom values of $(\bar\theta_1, \bar\sigma_1)$. As observed in the introduction, this condition is not satisfied by finite normal mixture models, which explains why these results do not apply to such models.

By $2\log(1+x) \le 2x - x^2 + (2/3)x^3$, we have

$$r_{1n} \le 2\sum_{i=1}^n \delta_i - \sum_{i=1}^n \delta_i^2 + 2/3\sum_{i=1}^n \delta_i^3.$$

For $l = 0, 1, 2, 3, 4$ and $s = 0, 1, 2, 3, 4$, we define

$$\bar m_{l,s} = (1-\bar\alpha)\bar\theta_1^l(\bar\sigma_1^2 - 1)^s + \bar\alpha\bar\theta_2^l(\bar\sigma_2^2 - 1)^s.$$

Denoting

$$f^{(l,s)}(x; \theta, \sigma) = \frac{\partial^{l+s} f(x; \theta, \sigma)}{\partial\theta^l \partial(\sigma^2)^s},$$

and expanding $f(X_i; \bar{\theta}_h, \bar{\sigma}_h)$ to order 4, we find that

$$\delta_i = \sum_{l+s=1}^{4} \binom{l+s}{s} \frac{\bar{m}_{l,s}}{(l+s)!} \frac{f^{(l,s)}(X_i; 0, 1)}{f(X_i; 0, 1)} + \epsilon_{in}^{(1)}$$

and the remainder term $\epsilon_n^{(1)} = \sum_{i=1}^{n} \epsilon_{in}^{(1)}$ satisfies

$$\epsilon_n^{(1)} = O_p(n^{1/2}) \Big\{ \sum_{h=1}^{2} \sum_{k=0}^{5} |\bar{\theta}_h|^k |\bar{\sigma}_h^2 - 1|^{5-k} \Big\}. \tag{A.9}$$

The reason for expanding to the fourth term is the loss of strong identifiability. Having to go through these extra terms makes the proof complex.

Note that when $k = 0, 1, 2$,

$$\sum_{h=1}^{2} |\bar{\theta}_h|^k |\bar{\sigma}_h^2 - 1|^{5-k} \leq \sum_{h=1}^{2} |\bar{\sigma}_h^2 - 1|^3$$

and when $k = 3, 4$,

$$\sum_{h=1}^{2} |\bar{\theta}_h|^k |\bar{\sigma}_h^2 - 1|^{5-k} \leq \sum_{h=1}^{2} |\bar{\theta}_h|^3 |\bar{\sigma}_h^2 - 1|.$$

Therefore, (A.9) is simplified to

$$\epsilon_n^{(1)} = O_p(n^{1/2}) \sum_{h=1}^{2} \{ |\bar{\theta}_h|^5 + |\bar{\theta}_h|^3 |\bar{\sigma}_h^2 - 1| + |\bar{\sigma}_h^2 - 1|^3 \}.$$

Absorbing the terms $\bar{m}_{l,s}$ with $l + 2s \geq 5$ into the remainder term, we have

$$\delta_i = \sum_{l+2s=1}^{4} \binom{l+s}{s} \bar{m}_{l,s} \frac{f^{(l,s)}(X_i; 0, 1)}{(l+s)! f(X_i; 0, 1)} + \epsilon_{in} \tag{A.10}$$

with

$$\begin{aligned} \epsilon_n &= \sum_{i=1}^{n} \epsilon_{in} \\ &= O_p(n^{1/2}) \sum_{h=1}^{2} \{ |\bar{\theta}_h|^5 + |\bar{\theta}_h|^3 |\bar{\sigma}_h^2 - 1| + |\bar{\theta}_h| (\bar{\sigma}_h^2 - 1)^2 + |\bar{\sigma}_h^2 - 1|^3 \}. \end{aligned} \tag{A.11}$$

Note that all the other terms in δ_i have been looked after in the above absorption. For example,

$$\bar{m}_{2,2} = \sum_{h=1}^{2} \bar{\theta}_h^2 (\sigma_h^2 - 1)^2 \le \sum_{h=1}^{2} |\bar{\theta}_h| (\sigma_h^2 - 1)^2$$

and

$$\bar{m}_{0,4} = \sum_{h=1}^{2} (\sigma_h^2 - 1)^4 \le \sum_{h=1}^{2} |\sigma_h^2 - 1|^3.$$

It is seen that $2|\bar{\theta}_h|^3 |\bar{\sigma}_h^2 - 1| \le \{|\bar{\theta}_h|^5 + |\bar{\theta}_h|(\bar{\sigma}_h^2 - 1)^2\}$. Consequently, (A.11) is reduced to

$$\text{(A.12)} \quad \epsilon_n = \sum_{i=1}^{n} \epsilon_{in} = O_p(n^{1/2}) \sum_{h=1}^{2} \{|\bar{\theta}_h|^5 + |\bar{\theta}_h|(\bar{\sigma}_h^2 - 1)^2 + |\bar{\sigma}_h^2 - 1|^3\}.$$

Next we simplify the dominant term of δ_i. By simple algebra, (A.10) is further reduced to

$$\delta_i = \bar{t}_1 X_i + \bar{t}_2 Z_i + \bar{t}_3 U_i + \bar{t}_4 V_i + \epsilon_{in},$$

where Z_i, U_i, and V_i are defined as before, and
(A.13)
$\bar{t}_1 = \bar{m}_{1,0}$, $\bar{t}_2 = \bar{m}_{2,0} + \bar{m}_{0,1}$, $\bar{t}_3 = \bar{m}_{3,0} + 3\bar{m}_{1,1}$, and $\bar{t}_4 = \bar{m}_{4,0} + 6\bar{m}_{2,1} + 3\bar{m}_{0,2}$.

After some straightforward algebra, we get

$$\begin{aligned} r_{1n}(\bar{\alpha}, \bar{\theta}_1, \bar{\theta}_2, \bar{\sigma}_1, \bar{\sigma}_2) \le\ & 2\{\bar{t}_1 \sum_{i=1}^{n} X_i + \bar{t}_2 \sum_{i=1}^{n} Z_i + \bar{t}_3 \sum_{i=1}^{n} U_i + \bar{t}_4 \sum_{i=1}^{n} V_i\} \\ & -\{\bar{t}_1^2 \sum_{i=1}^{n} X_i^2 + \bar{t}_2^2 \sum_{i=1}^{n} Z_i^2 + \bar{t}_3^2 \sum_{i=1}^{n} U_i^2 + \bar{t}_4^2 \sum_{i=1}^{n} V_i^2\}\{1 + o_p(1)\} \\ & +2/3\{\bar{t}_1 \sum_{i=1}^{n} X_i + \bar{t}_2 \sum_{i=1}^{n} Z_i + \bar{t}_3 \sum_{i=1}^{n} U_i + \bar{t}_4 \sum_{i=1}^{n} V_i\}^3 + O_p(\epsilon_n). \end{aligned}$$

Because (X_i, Z_i, U_i, V_i) are uncorrelated, the potential cross terms are of high orders and have been absorbed into the leading terms. In addition, we note that

$$\{\bar{t}_1 \sum_{i=1}^{n} X_i + \bar{t}_2 \sum_{i=1}^{n} Z_i + \bar{t}_3 \sum_{i=1}^{n} U_i + \bar{t}_4 \sum_{i=1}^{n} V_i\}^3 = o_p(n)\Big\{\sum_{l=1}^{4} \bar{t}_l^2\Big\}.$$

Hence,

$$\begin{aligned} r_{1n}(\bar\alpha,\bar\theta_1,\bar\theta_2,\bar\sigma_1,\bar\sigma_2) \quad \le \quad & 2\{\bar t_1\sum_{i=1}^n X_i + \bar t_2\sum_{i=1}^n Z_i + \bar t_3\sum_{i=1}^n U_i + \bar t_4\sum_{i=1}^n V_i\} \\ & -\{\bar t_1^2\sum_{i=1}^n X_i^2 + \bar t_2^2\sum_{i=1}^n Z_i^2 + \bar t_3^2\sum_{i=1}^n U_i^2 + \bar t_4^2\sum_{i=1}^n V_i^2\}\{1+o_p(1)\} \\ & +O_p(\epsilon_n). \end{aligned} \tag{A.14}$$

Our next step shows that

$$\epsilon_n = o_p(n)\Big\{\sum_{l=1}^4 \bar t_l^2\Big\},$$

which is the immediate consequence of the following lemma.

LEMMA 8. *Assume the same conditions as in Lemma 6. Under null distribution $N(0,1)$ and for $h=1,2$,*

$$\bar\theta_h^5 = o_p\Big\{\sum_{l=1}^4 |\bar t_l|\Big\},\ \bar\theta_h(\bar\sigma_h^2-1)^2 = o_p\Big\{\sum_{l=1}^4 |\bar t_l|\Big\},\ and\ (\bar\sigma_h^2-1)^3 = o_p\Big\{\sum_{l=1}^4 |\bar t_l|\Big\}.$$

PROOF. Clearly, the consistency result given in Lemma 6 implies that $\bar t_l = o_p(1)$, $l=1,2,3,4$. Let $\bar\beta_h = \bar\theta_h^2 + \bar\sigma_h^2 - 1$ for $h=1,2$. By the definitions of $\bar t_1$ and $\bar t_2$, we obtain

$$\bar\theta_2 = \{\bar t_1 - (1-\bar\alpha)\bar\theta_1\}/\bar\alpha, \tag{A.15}$$

$$\bar\beta_2 = \{\bar t_2 - (1-\bar\alpha)\bar\beta_1\}/\bar\alpha. \tag{A.16}$$

Plugging (A.15) and (A.16) into the definitions of $\bar t_3$ and $\bar t_4$ in (A.13), and because $\bar\alpha$ is bounded away from 0 and 1, we get

$$\bar t_3 = 3\frac{1-\bar\alpha}{\bar\alpha}\Big\{\bar\theta_1\bar\beta_1 - \frac{2(2\bar\alpha-1)}{3\bar\alpha}\bar\theta_1^3\Big\} + o_p(\bar t_1) + o_p(\bar t_2), \tag{A.17}$$

$$\bar t_4 = 3\frac{1-\bar\alpha}{\bar\alpha}\Big\{\bar\beta_1^2 - \frac{2(1-3\bar\alpha+3\bar\alpha^2)}{3\bar\alpha^2}\bar\theta_1^4\Big\} + o_p(\bar t_1) + o_p(\bar t_2). \tag{A.18}$$

Hence,

$$\Big\{\bar\beta_1 + \frac{2(2\bar\alpha-1)}{3\bar\alpha}\bar\theta_1^2\Big\}\bar t_3 - \bar\theta_1\bar t_4 = \frac{2(1-\bar\alpha)(1-\bar\alpha+\bar\alpha^2)}{3\bar\alpha^3}\bar\theta_1^5 + o_p(\bar t_1) + o_p(\bar t_2).$$

Since $\bar\theta_h$ and $\bar\beta_h$ are all $o_p(1)$, and the coefficient of $\bar\theta_1^5$ is bounded away from 0, we conclude from this equation that

$$\bar\theta_1^5 = o_p\Big(\sum_{l=1}^4 |\bar t_l|\Big). \tag{A.19}$$

Applying this result to $\bar\theta_1 \bar t_4$, we also get $\bar\theta_1\bar\beta_1^2 = o_p\Big(\sum_{l=1}^4 |\bar t_l|\Big)$. By the definition of $\bar\beta_1$, it is further seen that

$$|\bar\theta_1(\bar\sigma_1^2-1)^2| \le 2|\bar\theta_1|(\bar\beta_1^2+\bar\theta_1^4) = o_p\Big(\sum_{l=1}^4 |\bar t_l|\Big).$$

Applying (A.17) and (A.18) to $\bar\beta_1 t_4 + \{2(1-3\bar\alpha+3\bar\alpha^2)/(3\bar\alpha^2)\}\bar\theta_1^3 t_4$ and from the order assessment given by (A.19), we get

$$|\bar\beta_1|^3 = o_p\Big(\sum_{l=1}^4 |\bar t_l|\Big).$$

Finally, we have

$$|(\bar\sigma_1^2-1)^3| = |\bar\beta_1+\theta_1^2|^3 = O_p(|\bar\beta_1|^3+|\bar\theta_1|^6) = o_p\Big(\sum_{l=1}^4 |\bar t_l|\Big).$$

The same conclusions for $\bar\theta_2$ and $\bar\sigma_2$ are true by symmetry. □

The order result about ϵ_n simplifies (A.14) into

$$\begin{aligned} r_{1n}(\bar\alpha,\bar\theta_1,\bar\theta_2,\bar\sigma_1,\bar\sigma_2) \;\le\; & 2\{\bar t_1\sum_{i=1}^n X_i + \bar t_2\sum_{i=1}^n Z_i + \bar t_3\sum_{i=1}^n U_i + \bar t_4\sum_{i=1}^n V_i\} \\ & -\{\bar t_1^2\sum_{i=1}^n X_i^2 + \bar t_2^2\sum_{i=1}^n Z_i^2 + \bar t_3^2\sum_{i=1}^n U_i^2 + \bar t_4^2\sum_{i=1}^n V_i^2\}\{1+o_p(1)\}. \end{aligned} \tag{A.20}$$

Applying some of the classic results about regular models, we have

$$r_{2n} = -\frac{(\sum_{i=1}^n X_i)^2}{\sum_{i=1}^n X_i^2} - \frac{(\sum_{i=1}^n Z_i)^2}{\sum_{i=1}^n Z_i^2} + o_p(1). \tag{A.21}$$

Next we work on r_{3n}. By the mean value theorem on $p_n(\bar\sigma_h) - p_n(1)$,

$h = 1, 2$, and Conditions C0 and C4, we get

$$
\begin{aligned}
r_{3n}(\bar\sigma_1, \bar\sigma_2) &\le o_p(n^{1/6})(|\bar\sigma_1^2 - 1| + |\bar\sigma_2^2 - 1|) \\
&\le o_p(1) + o_p(n^{1/2})\{(\bar\sigma_1^2 - 1)^3 + (\bar\sigma_2^2 - 1)^3\} \\
&= o_p(1) + o_p(n^{1/2})\Big\{\sum_{l=1}^4 |\bar t_l|\Big\} \\
&\le o_p(1) + o_p(n)\Big\{\sum_{l=1}^4 \bar t_l^2\Big\}. \qquad \text{(A.22)}
\end{aligned}
$$

In the second step above, we used the fact that $|x| \le 1 + |x|^3$, and in the third step, we applied the results of Lemma 8. Combining (A.20) and (A.22), we have

$$
\begin{aligned}
& r_{1n}(\bar\alpha, \bar\theta_1, \bar\theta_2, \bar\sigma_1, \bar\sigma_2) + r_{3n}(\bar\sigma_1, \bar\sigma_2) \\
\le\ & 2\{\bar t_1 \sum_{i=1}^n X_i + \bar t_2 \sum_{i=1}^n Z_i + \bar t_3 \sum_{i=1}^n U_i + \bar t_4 \sum_{i=1}^n V_i\} \\
& -\{\bar t_1^2 \sum_{i=1}^n X_i^2 + \bar t_2^2 \sum_{i=1}^n Z_i^2 + \bar t_3^2 \sum_{i=1}^n U_i^2 + \bar t_4^2 \sum_{i=1}^n V_i^2\}\{1 + o_p(1)\} + o_p(1). \qquad \text{(A.23)}
\end{aligned}
$$

Inequality (A.23) implies that $r_{1n}(\bar\alpha, \bar\theta_1, \bar\theta_2, \bar\sigma_1, \bar\sigma_2) + r_{3n}(\bar\sigma_1, \bar\sigma_2)$ is stochastically bounded by the maximum of the following quadratic function:

$$
\begin{aligned}
Q(t_1, t_2, t_3, t_4) &= 2\{t_1 \sum_{i=1}^n X_i + t_2 \sum_{i=1}^n Z_i + t_3 \sum_{i=1}^n U_i + t_4 \sum_{i=1}^n V_i\} \\
&\quad -\{t_1^2 \sum_{i=1}^n X_i^2 + t_2^2 \sum_{i=1}^n Z_i^2 + t_3^2 \sum_{i=1}^n U_i^2 + t_4^2 \sum_{i=1}^n V_i^2\}.
\end{aligned}
$$

We see that $Q(t_1, t_2, t_3, t_4)$ is maximized at $t_l = \hat t_l$, $l = 1, 2, 3, 4$, with

$$
\hat t_1 = \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n X_i^2},\ \hat t_2 = \frac{\sum_{i=1}^n Z_i}{\sum_{i=1}^n Z_i^2},\ \hat t_3 = \frac{\sum_{i=1}^n U_i}{\sum_{i=1}^n U_i^2},\ \text{and } \hat t_4 = \frac{\sum_{i=1}^n V_i}{\sum_{i=1}^n V_i^2} \qquad \text{(A.24)}
$$

and

$$
Q(\hat t_1, \hat t_2, \hat t_3, \hat t_4) = \frac{(\sum_{i=1}^n X_i)^2}{\sum_{i=1}^n X_i^2} + \frac{(\sum_{i=1}^n Z_i)^2}{\sum_{i=1}^n Z_i^2} + \frac{(\sum_{i=1}^n U_i)^2}{\sum_{i=1}^n U_i^2} + \frac{(\sum_{i=1}^n V_i)^2}{\sum_{i=1}^n V_i^2}.
$$

This implies that

$$
\begin{aligned}
& r_{1n}(\bar\alpha, \bar\theta_1, \bar\theta_2, \bar\sigma_1, \bar\sigma_2) + r_{3n}(\bar\sigma_1, \bar\sigma_2) \\
\le\ & \frac{(\sum_{i=1}^n X_i)^2}{\sum_{i=1}^n X_i^2} + \frac{(\sum_{i=1}^n Z_i)^2}{\sum_{i=1}^n Z_i^2} + \frac{(\sum_{i=1}^n U_i)^2}{\sum_{i=1}^n U_i^2} + \frac{(\sum_{i=1}^n V_i)^2}{\sum_{i=1}^n V_i^2} + o_p(1). \qquad \text{(A.25)}
\end{aligned}
$$

We now summarize (A.21) and (A.25) into the following lemma.

LEMMA 9. *Assume the same conditions as in Lemma 6. Then under null distribution $N(0,1)$,*

$$2\{pl_n(\bar{\alpha}, \bar{\theta}_1, \bar{\theta}_2, \bar{\sigma}_1, \bar{\sigma}_2) - pl_n(0.5, \hat{\theta}_0, \hat{\theta}_0, \hat{\sigma}_0, \hat{\sigma}_0)\} \leq \frac{(\sum_{i=1}^n U_i)^2}{\sum_{i=1}^n U_i^2} + \frac{(\sum_{i=1}^n V_i)^2}{\sum_{i=1}^n V_i^2} + o_p(1). \tag{A.26}$$

Proof of Theorem 4

The conclusion of Theorem 3 implies that the upper bound given by (A.26) is applicable to $EM_n^{(K)}$. Now we show that the $EM_n^{(K)}$ asymptotically equals this upper bound. Since the EM-iteration increases the penalized likelihood, we need only show that for a given α value not equaling 0 or 1, there exists a set of feasible $\theta_1, \theta_2, \sigma_1$, and σ_2 values at which (A.25) and hence (A.26) become equalities. This is equivalent to finding a set of values such that $t_j = \hat{t}_j + o_p(n^{-1/2})$, for $j = 1, 2, 3, 4$, where t_j and $\hat{t}_j$ are defined in (A.13) and (A.24). Considering the case where $\tilde{\alpha} = 0.5$ and ignoring the $o_p(n^{-1/2})$ terms, these equations are

$$\begin{cases} 1/2(\theta_1+\theta_2) &= \hat{t}_1 \\ 1/2(\beta_1+\beta_2) &= \hat{t}_2 \\ 3/2(\theta_1\beta_1+\theta_2\beta_2) &= \hat{t}_3 \\ 3/2(\beta_1^2+\beta_2^2)-(\theta_1^4+\theta_2^4) &= \hat{t}_4 \end{cases}. \tag{A.27}$$

Furthermore, we may replace the third and fourth equations by (A.17) and (A.18):

$$\begin{cases} 3\theta_1\beta_1 &= \hat{t}_3 \\ 3\beta_1^2 - 2\theta_1^4 &= \hat{t}_4 \end{cases}. \tag{A.28}$$

The equations in (A.28) imply that

$$g(\theta_1) = 6\theta_1^6 + 3\hat{t}_4\theta_1^2 - \hat{t}_3^2 = 0.$$

Note that $g(0) < 0$ and $g(\theta_1) \to \infty$ as $\theta_1 \to \infty$, therefore there exists a positive solution for θ_1. Let $\tilde{\theta}_1$ be the smallest positive solution and $\tilde{\beta}_1$ be the corresponding solution for β_1. We then solve (A.27) to get $\tilde{\theta}_2$ and $\tilde{\beta}_2$.

Note that there are practically no restrictions on the range of θ and β. Thus we have shown that a feasible solution to (A.27) exists.

Additional simple algebra shows that the above solutions satisfy $\tilde{t}_j = \hat{t}_j + o_p(n^{-1/2})$, $\tilde{\theta}_h = O_p(n^{-1/8})$, $\tilde{\beta}_h = O_p(n^{-1/4})$, and $\tilde{\sigma}_h^2 - 1 = O_p(n^{-1/4})$, for $h = 1, 2$. With this order information, we can easily expand and obtain

$$r_{1n}(\tilde{\alpha}, \tilde{\theta}_1, \tilde{\theta}_2, \tilde{\sigma}_1, \tilde{\sigma}_2) = \frac{(\sum_{i=1}^n X_i)^2}{\sum_{i=1}^n X_i^2} + \frac{(\sum_{i=1}^n Z_i)^2}{\sum_{i=1}^n Z_i^2} + \frac{(\sum_{i=1}^n U_i)^2}{\sum_{i=1}^n U_i^2} + \frac{(\sum_{i=1}^n V_i)^2}{\sum_{i=1}^n V_i^2} + o_p(1)$$

and hence

$$\begin{aligned} & 2\{pl_n(\tilde{\alpha}, \tilde{\theta}_1, \tilde{\theta}_2, \tilde{\sigma}_1, \tilde{\sigma}_2) - pl_n(0.5, \hat{\theta}_0, \hat{\theta}_0, \hat{\sigma}_0, \hat{\sigma}_0)\} \\ &= r_{1n}(\tilde{\alpha}, \tilde{\theta}_1, \tilde{\theta}_2, \tilde{\sigma}_1, \tilde{\sigma}_2) + r_{2n} + r_{3n}(\tilde{\sigma}_1, \tilde{\sigma}_2) \\ &= \frac{(\sum_{i=1}^n U_i)^2}{\sum_{i=1}^n U_i^2} + \frac{(\sum_{i=1}^n V_i)^2}{\sum_{i=1}^n V_i^2} + o_p(1). \end{aligned}$$

Because the EM-iteration always increases the penalized likelihood, we must have

$$\begin{aligned} EM_n^{(K)} &\geq 2\{\sup_{(\theta_1,\theta_2,\sigma_1,\sigma_2)} pl_n(0.5, \theta_1, \theta_2, \sigma_1, \sigma_2) - pl_n(0.5, \hat{\theta}_0, \hat{\theta}_0, \hat{\sigma}_0, \hat{\sigma}_0)\} \\ &\geq 2\{pl_n(\tilde{\alpha}, \tilde{\theta}_1, \tilde{\theta}_2, \tilde{\sigma}_1, \tilde{\sigma}_2) - pl_n(0.5, \hat{\theta}_0, \hat{\theta}_0, \hat{\sigma}_0, \hat{\sigma}_0)\} \\ &= \frac{(\sum_{i=1}^n U_i)^2}{\sum_{i=1}^n U_i^2} + \frac{(\sum_{i=1}^n V_i)^2}{\sum_{i=1}^n V_i^2} + o_p(1). \end{aligned}$$

That is, the asymptotic upper bound of $EM_n^{(K)}$ is also a lower bound. Hence

$$EM_n^{(K)} = \frac{(\sum_{i=1}^n U_i)^2}{\sum_{i=1}^n U_i^2} + \frac{(\sum_{i=1}^n V_i)^2}{\sum_{i=1}^n V_i^2} + o_p(1).$$

Consequently, the limiting distribution of $EM_n^{(K)}$ is given by the χ_2^2. □

REFERENCES

[1] Cai, T., Jin, J., and Low, M. (2007). Estimation and confidence sets for sparse normal mixtures. *The Annals of Statistics*, **35**, 2421-2449.

[2] Charnigo, R. and Sun J. (2004). Testing homogeneity in a mixture distribution via the L^2-distance between competing models. *Journal of the American Statistical Association*, **99**, 488-498.

[3] Chen, H. and Chen, J. (2003). Tests for homogeneity in normal mixtures with presence of a structural parameter. *Statistica Sinica*, **13**, 351-365.

[4] Chen, H., Chen, J., and Kalbfleisch, J. D. (2001). A modified likelihood ratio for homogeneity in finite mixture models. *Journal of the Royal Statistical Society, Series B*, **63**, 19-29.

[5] Chen, J. (1995). Optimal rate of convergence in finite mixture models. *The Annals of Statistics*, **23**, 221-234.

[6] Chen, J. and Kalbfleisch, J. D. (2005). Modified likelihood ratio test in finite mixture models with a structural parameter. *Journal of Statistical Planning and Inference*, **129**, 93-107.

[7] Chen, J., Tan, X., and Zhang, R. (2008). Inference for normal mixtures in mean and variance. *Statistica Sinica*. To appear.

[8] Dacunha-Castelle, D. and Gassiat, E. (1999). Testing the order of a model using locally conic parametrization: Population mixtures and stationary ARMA processes. *The Annals of Statistics*, **27**, 1178-1209.

[9] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, **39**, 1-38.

[10] Efron, B. (2004). Large-scale simulation hypothesis testing: The choice of a null hypothesis. *Journal of the American Statistical Association*, **99**, 96-104.

[11] Everitt, B. S. (1996). An introduction to finite mixture distributions. *Statistical Methods in Medical Research*, **5**, 107-127.

[12] Everitt, B. S., Landau, S., and Leese, M. (2001). *Cluster Analysis* (4th ed.). New York, NY: Oxford University Press, Inc.

[13] Feng, Z. D. and McCulloch, C. E. (1994). On the likelihood ratio test statistic for the number of components in a normal mixture with unequal variances. *Biometrics*, **50**, 1158-1162.

[14] Garel, B. (2005). Asymptotic theory of the likelihood ratio test for the identification of a mixture. *Journal of Statistical Planning and Inference*, **131**, 271-296.

[15] 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. Berkeley Conf. in Honor of J. Neyman and Kiefer, Volume 2*, eds L. LeCam and R. A. Olshen, 789-806.

[15] Hall, P. and Stewart, M. (2005). Theoretical analysis of power in a two-component normal mixture model. *Journal of Statistical Planning and Inference*, **134**, 158-179.

[16] Hartigan, J. A. (1985). A failure of likelihood asymptotics for normal mixtures, in *Proc. Berkeley Conf. in Honor of J. Neyman and Kiefer, Volume 2*, eds L. LeCam and R. A. Olshen, 807-810.

[17] Hathaway, R. J. (1985). A constrained formulation of maximum-likelihood estimation for normal mixture distributions. *The Annals of Statistics*, **13**, 795-800.

[18] Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimates in the presence of infinitely many incidental parameters. *Annals of Mathematical Statistics*, **27**, 887-906.

[19] Kon, S, (1984). Models of stock returns–A comparison. *The Journal of Finance*, **39**, 147-165.

[20] Levine, R. (1981). Sex differences in schizophrenia: Timing or subtypes? *Psychological Bulletin*, **90**, 432-444.

[21] Li, P., Chen, J., and Marriott, P. (2008). Non-finite Fisher information and homogeneity: The EM approach. *Biometrika*. In revision.

[22] Liu, X., Pasarica, C., and Shao, Y. (2003). Testing homogeneity in gamma mixture models. *Scandinavian Journal of Statistics*, **30**, 227-239.

[23] Liu, X. and Shao, Y. Z. (2003). Asymptotics for likelihood ratio tests under loss of identifiability. *The Annals of Statistics*, **31**, 807-832.

[24] Liu, X. and Shao, Y. Z. (2004). Asymptotics for the likelihood ratio test in a two-component normal mixture model. *Journal of Statistical Planning and Inference*, **123**, 61-81.

[25] Loisel, P., Goffinet, B., Monod, H., and Montes De Oca, G. (1994). Detecting a major gene in an F2 population. *Biometrics*, **50**, 512-516.

[26] MacKenzie, S. A. and Bassett, M. J. (1987). Genetics of fertility restoration in cytoplasmic sterile *Phaseolus vulgaris* L. I. Cytoplasmic alteration by a nuclear restorer gene. *Theoretical and Applied Genetics*, **74**, 642-645.

[27] Marriott, P. (2007). Extending local mixture models. *Annals of the Institute of Statistical Mathematics*, **59**, 95-110.

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

[29] McLachlan, G. J., Bean, R. W., and Ben-Tovim Jones, L. (2006). A simple implementation of a normal mixture approach to differential gene expression in multiclass microarrays. *Bioinformatics*, **22**, 1608-1615.

[30] McLachlan, G. J. and Krishnan, T. (1997). *The EM algorithm and Extensions*. New York: Wiley.

[31] McLaren C. E. (1996). Mixture models in haematology: A series of case studies. *Statistical Methods in Medical Research*, **5**, 129-153.

[32] Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society Of London A*, **185**, 71-110.

[33] Raftery, A. E. and Dean, N. (2006). Variable selection for model-based clustering. *Journal of the American Statistical Assocation*, **101**, 168-178.

[34] Roeder, K. (1994). A graphical technique for determining the number of components in a mixture of normals. *Journal of the American Statistical Association*, **89**, 487-500.

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

[36] Sun, W. and Cai, T. (2007). Oracle and adaptive compound decision rules for false discovery rate control. *Journal of the American Statistical Association*, **102**, 901-912.

[37] Tadesse, M., Sha, N., and Vannucci, M. (2005). Bayesian variable selection in clustering high-dimensional data. *Journal of the American Statistical Association*, **100**, 602-617.

[38] Wolfe, J. H. (1971). A Monte Carlo study of the sampling distribution of the likelihood ratio for mixtures of multinormal distributions. *Technical Bulletin* **STB 72-2**. San Diego: U.S. Naval Personnel and Training Research Laboratory.

[39] Wu, C. F. J. (1981). On the convergence properties of the EM algorithm. *The Annals of Statistics*, **11**, 95-103.

Department of Statistics
University of British Columbia
Vancouver, BC, V6T 1Z2, Canada
E-mail: jhchen@stat.ubc.ca

Department of Mathematical and Statistical Sciences
University of Alberta,
Edmonton, T6G 2G1, Canada
E-mail: pengfei@stat.ualberta.ca