Statistica Sinica **19** (2009), 1603-1619

MODIFIED LIKELIHOOD RATIO TEST FOR HOMOGENEITY IN A TWO-SAMPLE PROBLEM

Yuejiao Fu, Jiahua Chen and John D. Kalbfleisch

York University, University of British Columbia and University of Michigan

Abstract: We consider testing for homogeneity in a two-sample problem in which one of the samples has a mixture structure. The problem arises naturally in many applications such as case-control studies with contaminated controls, or the test of a treatment effect in the presence of nonresponders in biological experiments or clinical trials. In this paper, we suggest using the modified likelihood ratio test (MLRT), which is devised to restore a degree of regularity in the mixture situation. The asymptotic properties of the MLRT statistic are investigated in mixtures of general one-parameter kernels, and in a situation where the kernels have an additional structural parameter. The MLRT statistic is shown to have a simple χ_1^2 null limiting distribution in both cases and simulations indicate that the MLRT performs better than other tests under a variety of model specifications. The proposed method is also illustrated in an example arising from a trial relating to morphine addiction in rats.

Key words and phrases: Asymptotic distribution, likelihood ratio test, mixture models, normal mixture, structural parameter, two-sample problem.

1. Introduction

Consider two samples, $x_{11}, \ldots, x_{1n_1}$ and $x_{21}, \ldots, x_{2n_2}$, from distributions F and H, respectively. In many applications, two-sample problems lead naturally to the formulation of a null hypothesis that the two samples are drawn from the same distribution, or that $H_0: F = H$. The choice of a test is usually affected by the alternatives that are considered likely to hold if H_0 is false. Typically one describes the alternative hypothesis by some measure of discrepancy between the two distributions, and there are many ways to do this. For instance, the two distributions may be taken as identical except for a location shift, or they may be allowed to differ in both location and scale. In this paper, we consider a mixture alternative as is described more precisely below. The case-control study with contaminated controls (Lancaster and Imbens (1996)) serves as a good preliminary example.

In classical case-control studies, two independent random samples are collected: the first is a sample of individuals who have experienced the event of interest (e.g. diagnosis with a specific disease), referred to as cases; the second is a sample of similar individuals, typically matched for potential confounding factors, who have not experienced the event, referred to as the controls. For each sample, the values of some risk factor (e.g. exposure to an environmental toxin) are observed. In the case-control study with contaminated controls, the second sample contains a mixture of cases and controls. For example, researchers may collect a sample of cases but, for economic reasons, they compare them with a sample arising from a source in which the event status is not observed. A similar example arises in genetic case-control studies with unscreened controls (Moskvina, Holmans, Schmidt and Craddock (2005)).

Suppose $x_{11}, \ldots, x_{1n_1}$ are independent and identically distributed (i.i.d.) with distribution $F(x)$, and $x_{21}, \ldots, x_{2n_2}$ are i.i.d. with distribution $H(x) = (1-\lambda)F(x) + \lambda G(x)$, where λ represents the contamination proportion. A statistical problem of interest is to test the null hypothesis $H_0 : H(x) = F(x)$ or $\lambda = 0$ against the mixture alternative.

The contaminated case-control study is one example of an application that can be formulated in this way. Another arises when testing for a treatment effect when not all experimental subjects respond to the new treatment. This is the so-called nonresponse phenomenon. In this case, F represents the distribution in the control group, G is the distribution of responders to the treatment, and λ is the proportion of subjects that respond. It is of interest to test whether $\lambda \neq 0$ so that there exists a subgroup which responds to the new treatment. See Good (1979) Boos and Brownie (1986, 1991), Conover and Salsburg (1988), Razzaghi and Nanthakumar (1994), and Razzaghi and Kodell (2000).

When the distribution functions F and G have different means, a classical t-test or a simple permutation test could be used. However, these tests are not designed to deal with the specific mixture alternative and specific methods designed for mixture model alternatives would perform better. It is generally accepted that preferred testing procedures are likelihood-based. Due to the non-regularity of the mixture models, however, the likelihood ratio test (LRT) statistic does not have the usual chi-squared limiting distribution. Recently, it has been discovered in many situations that the limiting distribution is that of the squared supremum of a truncated Gaussian process, a result that is not particularly convenient for inference (Liu and Shao (2003)). For a large variety of mixture models, the modified likelihood ratio test (MLRT) has a limiting distribution that is chi-squared or a mixture of chi-squared distributions (Chen (1998), Chen, Chen and Kalbfleisch (2001, 2004), and Zhu and Zhang (2004)) and hence provides a nice alternative to the ordinary LRT.

In this paper, we investigate the use of the MLRT to test homogeneity in the two-sample problem discussed above. Its asymptotic properties are studied both in mixtures of general one-parameter kernels and in a situation with an extra structural parameter. In either case, the limiting null distribution of the MLRT statistic is shown to be χ^2_1. In sharp contrast to the problem discussed in this paper, the distribution of the MLRT statistic lacks a simple analytical form in the one-sample normal mixture model with unknown variance, see Chen and Kalbfleisch (2005). Another interesting result that arises from the MLRT in the two-sample problem is that the component parameter estimators converge to their true values at the rate $n^{-1/2}$ rather than $n^{-1/4}$ (see Chen (1995)).

The remainder of this paper is organized as follows. The main results are given in Section 2. Sections 3 and 4 present simulation studies and an example. Some concluding comments are provided in Section 5.

2. Main Results

This section has three parts. Section 2.1 states the asymptotic properties of the MLRT with general one-parameter kernels, Section 2.2 gives the asymptotic properties in the presence of an unknown structural parameter, and Section 2.3 derives the local asymptotic power.

2.1. Mixtures of a general one-parameter kernel

Consider the two-sample parametric model

$$x_{11},\ldots,x_{1n_1} \overset{\text{i.i.d.}}{\sim} f(x;\theta_1), \qquad x_{21},\ldots,x_{2n_2} \overset{\text{i.i.d.}}{\sim} (1-\lambda)f(x;\theta_1)+\lambda f(x;\theta_2),$$

where $f(x;\theta)$ is a probability density function (pdf) with respect to a σ-finite measure, belonging to a parametric family $\{f(x;\theta):\theta\in\Theta\}$ with the parameter space Θ being a compact subset of the real line. The log-likelihood function is given by $l_n(\lambda,\theta_1,\theta_2)=\sum\log f(x_{1i};\theta_1)+\sum\log\{(1-\lambda)f(x_{2i};\theta_1)+\lambda f(x_{2i};\theta_2)\}$. We are concerned with testing the hypothesis

$$H_0: \quad \theta_1=\theta_2, \quad \text{or} \quad \lambda=0. \tag{2.1}$$

This is a non-regular testing problem since the null hypothesis is on the boundary of the parameter space ($\lambda=0$), and because the parameters (λ and θ_2) are not identifiable under the null hypothesis. As in many similar situations, the LRT statistic for this problem does not have the usual chi-squared limiting distribution.

To circumvent these problems, we consider the modified log-likelihood function $pl_n(\lambda,\theta_1,\theta_2)=l_n(\lambda,\theta_1,\theta_2)+C\log(\lambda)$, for some positive constant C that controls the level of modification. The penalty term $C\log(\lambda)$ penalizes estimates in which the mixing proportion λ is close to zero. A smooth penalty like this avoids the problematic parts of the parameter space and hence partially restores the regularity of the problem. This general idea is appealing and has quite general applicability to testing problems in mixture models. One advantage of this approach is that the resulting test statistic often has a simple null limiting distribution. Let $(\hat{\lambda}, \hat{\theta}_1, \hat{\theta}_2)$ maximize pl_n over $0 < \lambda \leq 1$, $\theta_j \in \Theta$, $j = 1, 2$, and let $\hat{\theta}$ maximize the null modified log-likelihood function $pl_n(1, \theta, \theta)$ over $\theta \in \Theta$. We call $(\hat{\lambda}, \hat{\theta}_1, \hat{\theta}_2)$ the modified maximum likelihood estimates (MMLE's). The MLRT statistic is $M_n = 2\{pl_n(\hat{\lambda}, \hat{\theta}_1, \hat{\theta}_2) - pl_n(1, \hat{\theta}, \hat{\theta})\}$ and the null hypothesis H_0 will be rejected when M_n is large enough.

To study the asymptotic properties of the MLRT under H_0, we assume that the kernel function $f(x; \theta)$ satisfies some regularity conditions. First, however, we introduce some notation. Let θ_0 be the true value of θ under the null model, and for $i = 1, \ldots, n_j$ and $j = 1, 2$, let

$$Y_{ji}(\theta) = Y_{ji}(\theta, \theta_0) = \frac{1}{\theta - \theta_0} \left\{ \frac{f(x_{ji}; \theta)}{f(x_{ji}; \theta_0)} - 1 \right\}, \quad \theta \neq \theta_0,$$
$$Y'_{ji}(\theta) = \frac{\partial Y_{ji}(\theta, \theta_0)}{\partial \theta},$$

where $f'(x; \theta) = \partial f / \partial \theta$. We write the continuous extension of $Y_{ji}(\theta)$ at θ_0 as Y_{ji}. We assume the following regularity conditions on $f(x; \theta)$.

A1. (Wald's integrability conditions for consistency of the MLE.) For each $\theta \in \Theta$, where Θ is a compact subset of the real line, (i) $E(|\log f(x; \theta)|) < \infty$, and (ii) there exists $\rho > 0$ such that $E[\log f(x; \theta, \rho)] < \infty$, where $f(x; \theta, \rho) = 1 + \sup_{|\theta' - \theta| \leq \rho} f(x; \theta')$.

A2. (Smoothness.) The support of $f(x; \theta)$ is the same for all $\theta \in \Theta$ and $f(x; \theta)$ is twice continuously differentiable with respect to θ.

A3. (Identifiability.) $E\{Y_{ji}^2(\theta)\} > 0$ for all $\theta \in \Theta$.

A4. (Condition for uniform strong law of large numbers.) There exists an integrable g such that $|Y_{ji}(\theta)|^2 \leq g(X_{ji})$ and $|Y'_{ji}(\theta)|^3 \leq g(X_{ji})$, for all $\theta \in \Theta$.

A5. (Tightness.) The processes ${n_j}^{-1/2} \sum Y_{ji}(\theta)$ and ${n_j}^{-1/2} \sum Y'_{ji}(\theta)$ are tight for $j = 1, 2$.

Conditions A1 and A2 effectively cover the assumptions of Wald (1949) for consistency of the MLE in the one-sample problem. Conditions A3-A5 guarantee a quadratic expansion of the log-likelihood function so that the remainder terms are statistically small-o. These conditions are satisfied by many mixture models, such as binomial, Poisson, and Normal with known variance. They have been used in Chen (1995), Chen and Chen (2003), Dacunha-Castelle and Gassiat (1999), and Liu and Shao (2003). Interestingly, for the two-sample problem, the identifiability condition needed is much simpler.

Asymptotic null distribution of the MLRT statistic, M_n

Theorem 1. *In the hypothesis testing problem* (2.1), *assume that $f(x;\theta)$ satisfies conditions A1–A5, the null model with parameter $\theta_1 = \theta_2 = \theta_0$ is true, θ_0 is an interior point of the compact set Θ, and $n_j/n \to \rho_j > 0$, $j = 1, 2$, as $n \to \infty$. Then M_n has a χ^2_1 limiting distribution under H_0.*

The key step in the proof is a quadratic approximation to the MLRT statistic when θ_1 and θ_2 are in a neighborhood of θ_0. The following lemma shows that the modified MLEs, $\hat{\theta}_1$ and $\hat{\theta}_2$, are consistent. For the proof of the lemma, we refer the readers to `http://www.stat.sinica.edu.tw/statistica` for the on-line version of the paper containing the supplement.

Lemma 1. *Under the conditions of Theorem 1,* $\log(\hat{\lambda}) = O_p(1)$ *and the modified MLEs, $\hat{\theta}_1$ and $\hat{\theta}_2$, both converge to θ_0 in probability.*

Proof of Theorem 1. Let $R_{1n} = 2\{pl_n(\lambda, \theta_1, \theta_2) - pl_n(1, \theta_0, \theta_0)\}$, $R_{2n} = 2\{pl_n(1, \theta_0, \theta_0) - pl_n(1, \hat{\theta}, \hat{\theta})\}$, and note that $M_n = \sup_{\lambda,\theta_1,\theta_2}(R_{1n} + R_{2n})$. It can be seen that

$$R_{1n} = 2\sum_{i=1}^{n_1} \log\left\{\frac{f(x_{1i};\theta_1)}{f(x_{1i};\theta_0)}\right\} + 2\sum_{i=1}^{n_2} \log(1+\delta_i) + 2C\log(\lambda),$$

where

$$\begin{aligned}
\delta_i &= (1-\lambda)\left\{\frac{f(x_{2i};\theta_1)}{f(x_{2i};\theta_0)} - 1\right\} + \lambda\left\{\frac{f(x_{2i};\theta_2)}{f(x_{2i};\theta_0)} - 1\right\}\\
&= m_1 Y_{2i} + e_i,\\
m_1 &= (1-\lambda)(\theta_1 - \theta_0) + \lambda(\theta_2 - \theta_0), \quad \text{and}\\
e_i &= (1-\lambda)(\theta_1 - \theta_0)\{Y_{2i}(\theta_1) - Y_{2i}\} + \lambda(\theta_2 - \theta_0)\{Y_{2i}(\theta_2) - Y_{2i}\}.
\end{aligned}$$

Since $\log(1+x) \le x - (1/2)x^2 + (1/3)x^3$, it follows that

$$\begin{aligned}
&2\sum_{i=1}^{n_2}\log(1+\delta_i)\\
&\quad \le 2\sum_{i=1}^{n_2}(m_1 Y_{2i} + e_i) - \sum_{i=1}^{n_2}(m_1 Y_{2i} + e_i)^2 + \frac{2}{3}\sum_{i=1}^{n_2}(m_1 Y_{2i} + e_i)^3. \qquad (2.2)
\end{aligned}$$

From Lemma 1 we can restrict attention to θ_1 and θ_2 values in a small neighborhood of θ_0, and assume that there is a de facto positive lower bound for λ. It is easy to see that $m_1^2 + (\theta_1 - \theta_0)^2$ is a strictly positive definite quadratic form in $\theta_1 - \theta_0$ and $\theta_2 - \theta_0$. Further, we may regard m_1 as if it is a small-o term. Under the null hypothesis, the Law of Large Numbers implies that

$n_2^{-1}\sum_{i=1}^{n_2} Y_{2i}^2 \to \mathrm{E}(Y_{21}^2)$ almost surely and $n_2^{-1}\sum_{i=1}^{n_2} Y_{2i}^3 \to \mathrm{E}(Y_{21}^3)$ almost surely. Hence,

$$\frac{m_1^3\sum_{i=1}^{n_2} Y_{2i}^3}{m_1^2\sum_{i=1}^{n_2} Y_{2i}^2} = m_1\frac{\sum_{i=1}^{n_2} Y_{2i}^3/n_2}{\sum_{i=1}^{n_2} Y_{2i}^2/n_2} = o_p(1).$$

Straightforward derivation under Condition A4 yields

$$\sum_{i=1}^{n_2}|e_i|^3 = \{(1-\lambda)^3(\theta_1-\theta_0)^6+\lambda^3(\theta_2-\theta_0)^6\}O_p(n_2) = \{m_1^2+(\theta_1-\theta_0)^2\}o(1)O_p(n_2).$$

The term $o(1)$ arises from the fact that both $\theta_1-\theta_0$ and $\theta_2-\theta_0$ can be made arbitrarily small, as mentioned earlier. These arguments establish that the cubic term of e_i in the inequality (2.2) is bounded by a quantity of order $\{m_1^2+(\theta_1-\theta_0)^2\}o(1)O_p(n_2)$.

Noting that $\sum\{Y_{2i}(\theta)-Y_{2i}\} = (\theta-\theta_0)\sum Y_{2i}'(\theta')$ for some θ' and the tightness of $Y_{2i}'(\theta)$, we have $\sum_{i=1}^{n_2} e_i = m_2 O_p(n_2^{1/2}) = \{m_1^2+(\theta_1-\theta_0)^2\}O_p(n_2^{1/2})$. Similarly, because we only need consider small $\theta_1-\theta_0$ and $\theta_2-\theta_0$,

$$\begin{aligned}\sum_{i=1}^{n_2} e_i^2 &= \{(1-\lambda)^2(\theta_1-\theta_0)^4+\lambda^2(\theta_2-\theta_0)^4\}O_p(n_2)\\ &= \{m_1^2+(\theta_1-\theta_0)^2\}o(1)O_p(n_2).\end{aligned}$$

With these results we obtain

$$2\sum_{i=1}^{n_2}\log(1+\delta_i) \le 2m_1\sum_{i=1}^{n_2} Y_{2i} - m_1^2\sum_{i=1}^{n_2} Y_{2i}^2 + \{m_1^2+(\theta_1-\theta_0)^2\}o(1)O_p(n_2).$$

Since $f(x;\theta)$ is regular, we have

$$2\sum_{i=1}^{n_1}\log\left\{\frac{f(x_{1i},\theta_1)}{f(x_{1i},\theta_0)}\right\} \le 2(\theta_1-\theta_0)\sum_{i=1}^{n_1} Y_{1i} - (\theta_1-\theta_0)^2\left\{\sum_{i=1}^{n_1} Y_{1i}^2\right\}\{1+o(1)\}$$

in probability. Adding the two bounds together, taking note that both n_1 and n_2 are of the same order, and that $n_j^{-1}\sum_i Y_{ji}^2$, $j=1,2$ converge to positive constants, we have in probability

$$\begin{aligned}R_{1n}(\lambda,\theta_1,\theta_2) \le\; & 2(\theta_1-\theta_0)\sum_{i=1}^{n_1} Y_{1i} + 2m_1\sum_{i=1}^{n_2} Y_{2i}\\ & -\left\{(\theta_1-\theta_0)^2\sum_{i=1}^{n_1} Y_{1i}^2 + m_1^2\sum_{i=1}^{n_2} Y_{2i}^2\right\}\{1+o(1)\}.\end{aligned}$$

It is easily seen that the above upper bound is maximized at

$$\tilde{m}_1 = \frac{\sum_{i=1}^{n_2} Y_{2i}}{\sum_{i=1}^{n_2} Y_{2i}^2} \quad\text{and}\quad \tilde{\theta}_1-\theta_0 = \frac{\sum_{i=1}^{n_1} Y_{1i}}{\sum_{i=1}^{n_1} Y_{1i}^2},$$

where $\tilde{m}_1$ may be positive or negative. By evaluating R_{1n} at parameter values for which $m_1 = \tilde{m}_1$, $\theta_1 = \tilde{\theta}_1$ and $\lambda = 1$, we find

$$R_{1n}(1, \tilde{m}_1, \tilde{\theta}_1) = \frac{\{\sum_{i=1}^{n_2} Y_{2i}\}^2}{\sum_{i=1}^{n_2} Y_{2i}^2} + \frac{\{\sum_{i=1}^{n_1} Y_{1i}\}^2}{\sum_{i=1}^{n_1} Y_{1i}^2} + o_p(1),$$

so that the upper bound is achievable.

Since $-R_{2n}$ is an ordinary LRT statistic under a regular model, it has the asymptotic approximation

$$R_{2n} = -\frac{\{\sum_{i=1}^{n_1} Y_{1i} + \sum_{i=1}^{n_2} Y_{2i}\}^2}{\sum_{i=1}^{n_1} Y_{1i}^2 + \sum_{i=1}^{n_2} Y_{2i}^2} + o_p(1).$$

Therefore,

$$M_n = \frac{\{\sum_{i=1}^{n_2} Y_{2i}\}^2}{\sum_{i=1}^{n_2} Y_{2i}^2} + \frac{\{\sum_{i=1}^{n_1} Y_{1i}\}^2}{\sum_{i=1}^{n_1} Y_{1i}^2} - \frac{\{\sum_{i=1}^{n_1} Y_{1i} + \sum_{i=1}^{n_2} Y_{2i}\}^2}{\sum_{i=1}^{n_1} Y_{1i}^2 + \sum_{i=1}^{n_2} Y_{2i}^2} + o_p(1).$$

Replacing $(\sum_{i=1}^{n_1} Y_{1i}^2 + \sum_{i=1}^{n_2} Y_{2i}^2)/n$ by $E(Y_{11}^2)$ yields

$$M_n = \frac{n_2 n_1}{n} \frac{1}{E(Y_{11}^2)} \Big\{ \frac{1}{n_2} \sum_{i=1}^{n_2} Y_{2i} - \frac{1}{n_1} \sum_{i=1}^{n_1} Y_{1i} \Big\}^2 + o_p(1), \tag{2.3}$$

and Theorem 1 follows from the Central Limit Theorem.

2.2. Mixture model with a structural parameter

We now consider the MLRT when the kernel density has a structural parameter. An example is a normal mixture with a common unknown variance. We show that the limiting distribution of the MLRT statistic is again χ_1^2.

Suppose that we have independent samples

$$x_{11}, \ldots, x_{1n_1} \overset{i.i.d.}{\sim} f(x; \theta_1, \xi), \quad x_{21}, \ldots, x_{2n_2} \overset{i.i.d.}{\sim} (1-\lambda) f(x; \theta_1, \xi) + \lambda f(x; \theta_2, \xi),$$

where $0 \le \lambda \le 1$, $\theta_j \in \Theta$ for $j = 1, 2$, and $\xi \in \Xi$. As before, we consider a test of

$$H_0: \quad \theta_1 = \theta_2, \quad \text{or} \quad \lambda = 0. \tag{2.4}$$

The modified log-likelihood function is

$$pl_n(\lambda, \theta_1, \theta_2, \xi) = \sum_{i=1}^{n_1} \log f(x_{1i}; \theta_1, \xi) + \sum_{i=1}^{n_2} \log\{(1-\lambda) f(x_{2i}; \theta_1, \xi) + \lambda f(x_{2i}; \theta_2, \xi)\} + C \log(\lambda),$$

and the modified MLEs, $\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\lambda}$ and $\hat{\xi}$, maximize $pl_n(\lambda, \theta_1, \theta_2, \xi)$ over the region $\{0 < \lambda \leq 1,\ \theta_j \in \Theta \text{ for } j = 1, 2,\quad \xi \in \Xi\}$. Let $\hat{\theta}$ and $\hat{\xi}_0$ be the values that maximize $pl_n(1, \theta, \theta, \xi)$. The MLRT statistic is defined as $M_n = 2\{pl_n(\hat{\lambda}, \hat{\theta}_1, \hat{\theta}_2, \hat{\xi}) - pl_n(1, \hat{\theta}, \hat{\theta}, \hat{\xi}_0)\}$.

To give regularity conditions, we set

$$Y_{ji}(\theta, \xi) = \frac{f(x_{ji}; \theta, \xi) - f(x_{ji}; \theta_0, \xi)}{(\theta - \theta_0) f(x_{ji}; \theta_0, \xi_0)}, \quad U_{ji}(\xi) = \frac{f(x_{ji}; \theta_0, \xi) - f(x_{ji}; \theta_0, \xi_0)}{(\xi - \xi_0) f(x_{ji}; \theta_0, \xi_0)},$$

for $i = 1, \ldots, n_j$; $j = 1, 2$. For convenience, we may write Y_{ji} for $Y_{ji}(\theta_0, \xi_0)$ and U_{ji} for $U_{ji}(\xi_0)$ as the continuous extensions of these expressions. The regularity conditions on $f(x; \theta, \xi)$ are similar to A1−A5.

B1. (Wald's integrability conditions.) Let G denote the mixing distribution. The parameter space of (G, ξ) can be compactified into $\{(\mathcal{G}, \Xi), d\}$ such that, for each $(G, \xi) \in (\mathcal{G}, \Xi)$, (i) $E(|\log f(x; G, \xi)|) < \infty$, and (ii) there exists $\rho > 0$ such that $E[\log f(x; G, \xi, \rho)] < \infty$, where

$$f(x; G, \xi, \rho) = 1 + \sup_{d((G', \xi'), (G, \xi)) \leq \rho} f(x; G', \xi').$$

B2. (Smoothness.) The kernel density $f(x; \theta, \xi)$ has common support and is three times continuously differentiable with respect to θ and ξ.

B3. (Identifiability.) Fisher information of $f(x; \theta, \xi)$ is of full rank at any (θ, ξ).

B4. (Condition for the uniform strong law of large numbers.) Let $Y'_\theta(\theta, \xi)$ and $Y'_\xi(\theta, \xi)$ be partial derivatives of $Y_{ji}(\theta, \xi)$ with ji omitted, and similarly for U'_ξ. There exists an integrable $g(x)$ such that $|Y_{ji}(\theta, \xi)|^2 \leq g(X_{ji})$, $|U_{ji}(\theta, \xi)|^2 \leq g(X_{ji})$, $|Y'_\theta(\theta, \xi)|^3 \leq g(X_{ji})$, $|Y'_\xi(\theta, \xi)|^3 \leq g(X_{ji})$, and $|U'_\xi(\xi)|^3 \leq g(X_{ji})$ for all $(\theta, \xi) \in \Theta \times \Xi$.

B5. (Tightness.) The processes $n_j^{-1/2} \sum Y_{ji}(\theta, \xi)$, $n_j^{-1/2} \sum Y'_\theta(\theta, \xi)$, $n_j^{-1/2} \sum Y'_\xi(\theta, \xi)$, and $n_j^{-1/2} \sum U_\xi(\xi)$ are all tight.

Conditions B1−B5 are satisfied by many mixture models used in practice, such as mixture models of most location-scale families including the normal, Cauchy and logistic. One can easily examine these conditions on the joint density function of a group of observations; see Kiefer and Wolfowitz (1956) for details of this technique.

Theorem 2. *In the hypothesis testing problem* (2.4), *assume that* $f(x; \theta, \xi)$ *satisfies conditions B1–B5, the null model with parameters* $\theta_1 = \theta_2 = \theta_0$ *and* $\xi = \xi_0$ *is true,* θ_0 *is an interior point of the compact set* Θ, ξ_0 *is an interior point of* Ξ, *and* $n_j/n \to \rho_j > 0$, $j = 1, 2$, *as* $n \to \infty$. *The MLRT statistic* M_n *has a* χ^2_1 *limiting distribution under* H_0.

The proof of Theorem 2 parallels that of Theorem 1. The extra complexity lies in the proof of the following lemma, which is given in the on-line supplement. Hence, the proof of Theorem 2 itself will be brief. Note that the structural parameter ξ does not have to be a scalar; similar results hold when it is a vector parameter.

Lemma 2. *Under the same assumptions as in Theorem* 2, $\log(\hat{\lambda}) = O_p(1)$, $\hat{\theta}_j \to \theta_0$ *for* $j = 1, 2$, *and* $\hat{\xi} \to \xi_0$, *in probability.*

Proof of Theorem 2. Write $R_{1n}(\lambda, \theta_1, \theta_2, \xi) = 2\{pl_n(\lambda, \theta_1, \theta_2, \xi) - pl_n(1, \theta_0, \theta_0, \xi_0)\}$ and $R_{2n} = 2\{pl_n(1, \theta_0, \theta_0, \xi_0) - pl_n(1, \hat{\theta}, \hat{\theta}, \hat{\xi}_0)\}$. Also write

$$
\begin{aligned}
R_{1n}(\lambda, \theta_1, \theta_2, \xi) &= 2\sum_{i=1}^{n_1} \log\left\{\frac{f(x_{1i}; \theta_1, \xi)}{f(x_{1i}; \theta_0, \xi_0)}\right\} \\
&\quad +2\sum_{i=1}^{n_2} \log\left\{(1-\lambda)\frac{f(x_{2i}; \theta_1, \xi)}{f(x_{2i}; \theta_0, \xi_0)} + \lambda\frac{f(x_{2i}; \theta_2, \xi)}{f(x_{2i}; \theta_0, \xi_0)}\right\} + 2C\log\lambda \\
&= 2\sum_{i=1}^{n_1} \log\left\{\frac{f(x_{1i}; \theta_1, \xi)}{f(x_{1i}; \theta_0, \xi_0)}\right\} + 2\sum_{i=1}^{n_2}\log(1+\delta_i) + 2C\log\lambda,
\end{aligned}
$$

with

$$
\begin{aligned}
\delta_i &= (1-\lambda)(\theta_1-\theta_0)Y_{2i}(\theta_1, \xi) + \lambda(\theta_2-\theta_0)Y_{2i}(\theta_2, \xi) + (\xi-\xi_0)U_{2i}(\xi) \\
&= m_1 Y_{2i} + (\xi-\xi_0)U_{2i} + e_i.
\end{aligned}
$$

By Lemma 2, we need only consider values of θ_1 and θ_2 near θ_0, of ξ near ξ_0, and of λ not near 0. In this case, we have

$$
\begin{aligned}
\sum_{i=1}^{n_2} e_i &= \{m_1^2 + (\theta_1-\theta_0)^2 + (\xi-\xi_0)^2\}O_p(n_2^{1/2}), \\
\sum_{i=1}^{n_2} e_i^2 &= \{m_1^2 + (\theta_1-\theta_0)^2 + (\xi-\xi_0)^2\}o(1)O_p(n_2), \\
\sum_{i=1}^{n_2} |e_i|^3 &= \{m_1^2 + (\theta_1-\theta_0)^2 + (\xi-\xi_0)^2\}o(1)O_p(n_2).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
2\sum_{i=1}^{n_2}\log(1+\delta_i) \le 2\sum_{i=1}^{n_2}\{m_1Y_{2i} + (\xi-\xi_0)U_{2i}\} - \sum_{i=1}^{n_2}\{m_1Y_{2i} + (\xi-\xi_0)U_{2i}\}^2 \\
+\{m_1^2 + (\theta_1-\theta_0)^2 + (\xi-\xi_0)^2\}o(1)O_p(n_2).
\end{aligned}
$$

For the first sample, due to the regularity of $f(x;\theta,\xi)$, we have

$$2\sum_{i=1}^{n_1}\log\left\{\frac{f(x_{1i};\theta_1,\xi)}{f(x_{1i};\theta_0,\xi_0)}\right\}$$
$$\le 2\sum_{i=1}^{n_1}\{(\theta_1-\theta_0)Y_{1i}+(\xi-\xi_0)U_{1i}\}-\sum_{i=1}^{n_1}\{(\theta_1-\theta_0)Y_{1i}+(\xi-\xi_0)U_{1i}\}^2\{1+o(1)\}$$

in probability. Combining the above two inequalities, we obtain

$$\begin{aligned}&R_{1n}(\lambda,\theta_1,\theta_2,\xi)\\ &\le 2(\theta_1-\theta_0)\sum_{i=1}^{n_1}Y_{1i}+2m_1\sum_{i=1}^{n_2}Y_{2i}+2(\xi-\xi_0)\sum_{j=1}^{2}\sum_{i=1}^{n_j}U_{ji}\\ &\quad-\left[\sum_{i=1}^{n_1}\{(\theta_1-\theta_0)Y_{1i}+(\xi-\xi_0)U_{1i}\}^2+\sum_{i=1}^{n_2}\{m_1Y_{2i}+(\xi-\xi_0)U_{2i}\}^2\right]\{1+o(1)\}\end{aligned}$$

in probability. Let $\sigma_Y^2=Var(Y_{11})$, $\sigma_U^2=Var(U_{11})$, and $\sigma_{YU}=Cov(Y_{11},U_{11})$. By the Law of Large Numbers, $n_1^{-1}\sum Y_{1i}^2=\sigma_Y^2+o_p(1)$, and similarly for other quantities. From approximation, we find

$$\begin{aligned}R_{1n}(\lambda,\theta_1,\theta_2,\xi)&\le 2(\theta_1-\theta_0)\sum_{i=1}^{n_1}Y_{1i}+2m_1\sum_{i=1}^{n_2}Y_{2i}+2(\xi-\xi_0)\sum_{j=1}^{2}\sum_{i=1}^{n_j}U_{ji}\\ &\quad-\big[n_1\{(\theta_1-\theta_0)^2\sigma_Y^2+2(\theta_1-\theta_0)(\xi-\xi_0)\sigma_{YU}+(\xi-\xi_0)^2\sigma_U^2\}\\ &\quad+n_2\{m_1^2\sigma_Y^2+2m_1(\xi-\xi_0)\sigma_{YU}+(\xi-\xi_0)^2\sigma_U^2\}\big]\{1+o(1)\}\end{aligned}$$

in probability. That is, $R_{1n}(\lambda,\theta_1,\theta_2,\xi)$ is bounded by a quadratic function in $(\theta_1-\theta_0,m_1,\xi-\xi_0)$. The quadratic function is maximized when $(\theta_1-\theta_0,m_1,\xi-\xi_0)^\tau=I_n^{-1}W_n+o_p(n^{-1/2})$, where

$$I_n=\begin{bmatrix}n_1\sigma_Y^2 & 0 & n_1\sigma_{YU}\\ 0 & n_2\sigma_Y^2 & n_2\sigma_{YU}\\ n_1\sigma_{YU} & n_2\sigma_{YU} & (n_1+n_2)\sigma_U^2\end{bmatrix}\quad\text{and}\quad W_n=\begin{bmatrix}\sum_{i=1}^{n_1}Y_{1i}\\ \sum_{i=1}^{n_2}Y_{2i}\\ \sum_{j=1}^{2}\sum_{i=1}^{n_j}U_{ji}\end{bmatrix}.$$

At the same time, the resulting upper bound is attained by $R_{1n}(\lambda,\theta_1,\theta_2,\xi)$ at exactly the same parameter values as above with $\lambda=1$. Thus, we have $\sup_{\lambda,\theta_1,\theta_2,\xi}R_{1n}=W_n^\tau I_n^{-1}W_n+o_p(1)$. It is seen that this quantity has a χ_3^2 limiting distribution.

On the other hand, R_{2n} is defined on a regular parameter space. Classical theory implies

$$-R_{2n}=\begin{pmatrix}\sum_{j=1}^{2}\sum_{i=1}^{n_j}Y_{ji}\\ \sum_{j=1}^{2}\sum_{i=1}^{n_j}U_{ji}\end{pmatrix}^\tau\begin{bmatrix}n\sigma_Y^2 & n\sigma_{YU}\\ n\sigma_{YU} & n\sigma_U^2\end{bmatrix}^{-1}\begin{pmatrix}\sum_{j=1}^{2}\sum_{i=1}^{n_j}Y_{ji}\\ \sum_{j=1}^{2}\sum_{i=1}^{n_j}U_{ji}\end{pmatrix}+o_p(1)$$

has a χ_2^2 limiting distribution. It follows that $M_n = \sup_{\lambda,\theta_1,\theta_2,\xi}(R_{1n} + R_{2n})$ is asymptotically positive definite and has a chi-squared limiting distribution with $3 - 2 = 1$ degree of freedom. This completes the proof.

2.3. Asymptotic power

To examine the asymptotic power of the MLRT without structural parameters, we consider the local alternative

$$H_a^n: \quad \lambda = \lambda_0, \quad \theta_1 = \theta_0, \quad \theta_2 = \theta_0 + n_2^{-1/2}\tau, \tag{2.5}$$

where $0 < \lambda_0 \le 1$ and $\tau \neq 0$. Note that θ_2 gets closer to θ_0 as n increases, and tends to θ_0 at rate of $n^{-1/2}$. This rate is based on the fact that the convergence rates of $\hat{\theta}_1$ and $\hat{\theta}_2$ are $n^{-1/2}$ as shown in the previous sections. Let $\chi_1^2(c)$ denote the noncentral chi-squared distribution with one degree of freedom and noncentrality parameter c.

Theorem 3. *Suppose that $f(x;\theta)$ satisfies the regularity conditions* A1$-$A5 *and $n_j/n \to \rho_j > 0$, $j = 1, 2$, as $n \to \infty$. Under the local alternative* (2.5), *the limiting distribution of M_n is $\chi_1^2(\lambda_0\tau\sqrt{\rho_1 E(Y_{11}^2)})$.*

Proof of Theorem 3. The local alternative is contiguous to the null distribution (see Le Cam and Yang (1990)). By Le Cam's contiguity theory, the limiting distribution of M_n under H_a^n is determined by the null limiting distribution of (M_n, Λ_n) where

$$\Lambda_n = \sum_{i=1}^{n_2} \log \frac{(1-\lambda_0)f(x_{2i};\theta_0) + \lambda_0 f(x_{2i};\theta_0 + n_2^{-1/2}\tau)}{f(x_{2i};\theta_0)}.$$

Equation (2.3) gives a quadratic approximation to M_n, and a quadratic approximation to Λ_n can be obtained in a similar manner to give

$$\Lambda_n = \lambda_0\tau n_2^{-1/2}\sum_{i=1}^{n_2} Y_{2i} - \frac{1}{2}\lambda_0^2\tau^2 n_2^{-1}\sum_{i=1}^{n_2} Y_{2i}^2 + o_p(1).$$

Let

$$V_n = \frac{{n_2}^{-1}\sum_{i=1}^{n_2} Y_{2i} - {n_1}^{-1}\sum_{i=1}^{n_1} Y_{1i}}{\{({n_1}^{-1} + {n_2}^{-1})E(Y_{11}^2)\}^{1/2}}.$$

It follows that the null limiting joint distribution of (V_n, Λ_n) is bivariate normal

$$L((V_n,\Lambda_n)^\tau | H_0) \to \mathcal{N}_2\left(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \begin{pmatrix}1 & \sigma_2\sqrt{\rho_1}\\ \sigma_2\sqrt{\rho_1} & \sigma_2^2\end{pmatrix}\right),$$

where $(\mu_1, \mu_2) = (0, -(1/2)\sigma_2^2)$ and $\sigma_2^2 = \lambda_0^2\tau^2 E(Y_{11}^2)$. By Le Cam's third lemma (Hájek and Šidák (1967)), we have $L(V_n|H_a^n) \to N(\sigma_2\sqrt{\rho_1}, 1)$. Since M_n is asymptotically equivalent to V_n^2, the limiting distribution of M_n under the local alternative H_a^n is that of $\chi_1^2(\lambda_0\tau\sqrt{\rho_1 E(Y_{11}^2)})$.

The key step in the above proof is the contiguity of the null model and the alternative model. Under the null model, the MLRT statistic has asymptotic expansion $M_n = V_n^2 + o_p(1)$. The joint distribution of V_n and Λ_n under the null model determines the distribution of V_n under the contiguous alternative H_a^n. To investigate the asymptotic power of the MLRT in the presence of a structural parameter, the same approach is useful. The key step is to find the explicit expansion of M_n under the null model. Under the assumptions of Theorem 2, we proved that M_n has a χ_1^2 limiting distribution without giving an explicit expansion. In fact, with some straightforward linear algebra, it can be shown that the expansion $M_n = V_n^2 + o_p(1)$ remains true for the same V_n as in the above proof. Further, Λ_n at $\xi = \xi_0$ can be expanded in the same way. Thus, the conclusion of Theorem 3 holds when a structural parameter is present.

3. Simulation Studies

We have conducted simulation studies in order to evaluate the finite sample performance of the MLRT in the two-sample problem. We generated two independent random samples with $n_1 = 150$ and $n_2 = 70$ or $n_1 = n_2 = 200$, from $N(\theta_1, 1)$ and $(1-\lambda)N(\theta_1, 1) + \lambda N(\theta_2, 1)$ with $\theta_1 = 0$. When $\lambda = 0$, the data were generated under the null hypothesis. When $\lambda = 0.1$, 0.2, or 0.3 and $\theta_2 = 0.5$, 1, or 2, the data were generated under alternatives. The MLRT with normal kernel and common unknown variance was used to test for homogeneity. Let $MR(C)$ denote the MLRT with the level of modification being C. Let T denote the two-sample t-test, and $v(0.67)$ denote the Fisher's type randomization test proposed by Good (1979). For the MLRT, the EM algorithm was used for finding the maximum of the modified log-likelihood function. For $v(0.67)$, random selection of 500 partitions were used to calculate the statistic. The simulated levels and power of the three tests are given in Table 1. The simulated levels were based on 10,000 repetitions. For the simulated power, the results were based on 2,000 repetitions. For the MLRT and two-sample t-test, we used the critical values of the observed test statistics under the null to determine the simulated power in order to make the power comparison fair. As seen in Table 1, for moderate or large samples, the simulated null rejection rates of the three tests are all close to nominal levels. Under local alternatives, the three tests had comparable power while under distant alternatives, $MR(1)$ had better power than T and Good's randomization test $v(0.67)$. For example, under alternative model with $\lambda = 0.1$, $\theta_2 = 2$ and sample sizes $n_0 = n_1 = 200$, $MR(1)$ outperformed T by 32.4% and outperformed $v(0.67)$ by 3.4% at 5% nominal level, and the improvements were even more remarkable at 1% nominal level.

Table 1. Simulated null rejection rates and power for $MR(1)$, T and $v(0.67)$. Here F is from $N(0,1)$ and H is from $(1-\lambda)N(0,1)+\lambda N(\theta_2,1)$.

			$n_1 = 150$, $n_2 = 70$			$n_1 = n_2 = 200$		
λ	θ_2	Nominal levels	$MR(1)$	T	$v(0.67)$	$MR(1)$	T	$v(0.67)$
0.0		0.10	0.109	0.103	0.101	0.104	0.100	0.104
0.0		0.05	0.056	0.051	0.054	0.053	0.050	0.051
0.0		0.01	0.013	0.011	0.011	0.011	0.010	0.011
0.1	0.5	0.10	0.140	0.137	0.128	0.145	0.146	0.141
0.1	0.5	0.05	0.074	0.074	0.071	0.079	0.082	0.074
0.1	0.5	0.01	0.015	0.016	0.017	0.027	0.026	0.018
0.1	1.0	0.10	0.217	0.205	0.221	0.262	0.246	0.289
0.1	1.0	0.05	0.140	0.128	0.140	0.176	0.167	0.181
0.1	1.0	0.01	0.039	0.037	0.048	0.069	0.061	0.068
0.1	2.0	0.10	0.536	0.383	0.612	0.818	0.561	0.824
0.1	2.0	0.05	0.441	0.283	0.487	0.762	0.438	0.728
0.1	2.0	0.01	0.297	0.126	0.257	0.619	0.225	0.482
0.2	0.5	0.10	0.202	0.201	0.169	0.248	0.247	0.204
0.2	0.5	0.05	0.128	0.129	0.102	0.163	0.164	0.125
0.2	0.5	0.01	0.032	0.035	0.027	0.062	0.064	0.037
0.2	1.0	0.10	0.425	0.400	0.375	0.629	0.600	0.549
0.2	1.0	0.05	0.313	0.284	0.264	0.510	0.470	0.404
0.2	1.0	0.01	0.155	0.130	0.116	0.290	0.246	0.205
0.2	2.0	0.10	0.910	0.789	0.927	0.997	0.967	0.997
0.2	2.0	0.05	0.865	0.699	0.880	0.994	0.939	0.991
0.2	2.0	0.01	0.738	0.486	0.732	0.985	0.817	0.962
0.3	1.0	0.10	0.659	0.642	0.591	0.894	0.881	0.786
0.3	1.0	0.05	0.545	0.522	0.465	0.831	0.816	0.684
0.3	1.0	0.01	0.331	0.296	0.276	0.653	0.604	0.449

We have also conducted simulations to compare the performance of $MR(C)$ and $v(0.67)$ in the situation with small samples. The results are reported in Table 2. As expected, $MR(2)$ had better simulated levels and lower power than $MR(1)$. We also note that when the sample was quite small, there was little to choose among the various methods as far as local alternatives go. When the alternatives were more distant, however, both $MR(1)$ and $MR(2)$ did much better than Good's randomization test.

4. An Example

In this section, we apply our method to the drug abuse data first presented in Weeks and Collins (1971) and analyzed subsequently by Good (1979)

Table 2. Simulated null rejection rates and power for MR and $v(0.67)$. F is from $N(0,1)$ and H is from $(1-\lambda)N(0,1)+\lambda N(\theta_2,1)$.

			$n_1 = 30$, $n_2 = 30$		
λ	θ_2	Nominal levels	$MR(1)$	$MR(2)$	$v(0.67)$
0.0		0.10	0.120	0.109	0.101
0.0		0.05	0.064	0.056	0.050
0.0		0.01	0.019	0.015	0.012
0.2	1.0	0.10	0.203	0.197	0.229
0.2	1.0	0.05	0.122	0.115	0.140
0.2	1.0	0.01	0.030	0.030	0.046
0.2	2.0	0.10	0.539	0.466	0.577
0.2	2.0	0.05	0.434	0.361	0.449
0.2	2.0	0.01	0.212	0.164	0.208
0.2	3.0	0.10	0.878	0.813	0.834
0.2	3.0	0.05	0.828	0.744	0.717
0.2	3.0	0.01	0.680	0.581	0.427
0.2	4.0	0.10	0.980	0.956	0.926
0.2	4.0	0.05	0.968	0.940	0.850
0.2	4.0	0.01	0.934	0.893	0.580
0.3	1.0	0.10	0.319	0.311	0.304
0.3	1.0	0.05	0.213	0.205	0.206
0.3	1.0	0.01	0.068	0.067	0.082
0.3	2.0	0.10	0.759	0.691	0.805
0.3	2.0	0.05	0.662	0.593	0.690
0.3	2.0	0.01	0.419	0.364	0.442
0.3	3.0	0.10	0.974	0.948	0.964
0.3	3.0	0.05	0.955	0.925	0.930
0.3	3.0	0.01	0.885	0.838	0.779
0.3	4.0	0.10	0.999	0.994	0.992
0.3	4.0	0.05	0.997	0.993	0.979
0.3	4.0	0.01	0.993	0.984	0.883

Table 3. Self-injection rates for rats after six days' treatment with morphine.

control	0	0	3	4	4	6	6	9	9	10
	11	11	13	14	18	18	20	23	26	28
	30	41	50	61	80	94				
treatment	0	0	2	3	5	5	6	10	13	27
	33	40	40	46	51	57	78	87	87	96
	103	113	123	146	159	261	261	281	450	500

and Boos and Brownie (1991). In order to study the addiction to morphine in rats, an experiment was conducted in which rats could get morphine by pressing a lever. The response variable is the frequency of lever presses (self-injection rates) after six days' treatment with morphine. Figure 1 of Good (1979) displays the self-injection rates for five groups of rats corresponding to four different dose levels and one saline control. Boos and Brownie (1991) analyzed the transformed data, $\log_{10}(R+1)$ with R being the number of lever presses by rats, and proposed a mixture model for the dose-response studies.

Here we are interested in comparing the self-injection rates of the treatment (at dose level 0.1) and control groups. The number of lever presses on the sixth day for the two groups of rats are listed in Table 3. We used the same transformation as in Boos and Brownie (1991). A Q-Q plot suggests that a normal model fits the transformed data in the control group very well. We applied the MLRT with normal kernel and common unknown variance to these data with penalty size C chosen to be 2. The MLRT statistic is 10.24, which suggests strong evidence of a treatment effect. The same result can be obtained using Good's randomization test $v(0.67)$. In comparison to the MLRT and Good's test, the usual two-sample t-test suggest weaker evidence.

The presence of nonresponders reduces the power of standard parametric tests. We suggest that the MLRT be employed to detect a treatment effect when there is good empirical or biological evidence that some of the subjects may not be responsive.

5. Concluding Remarks

In this paper, the MLRT is applied to a two-sample problem in which one of the samples has a mixture structure. It is found that the limiting distribution of the MLRT statistic is χ_1^2 under the null hypothesis, and this is very simple and convenient to use in applications. It is remarkable that this same limit is obtained whether or not the kernel contains an unknown structural parameter. This is in contrast to the situation in a single sample problem where the asymptotic distribution of the MLRT statistic is complicated when there is a structural parameter. We have also discussed the asymptotic distribution of the MLRT statistic under a sequence of local alternatives which allows for calculation of local powers. These results may be of assistance in planning experiments through providing guidance on appropriate sample size.

Finally, we provide some guideline on the choice of the modification constant C. Although the limiting distribution of the MLRT does not depend on the specific choice of C, the type I error based on the chi-square limit does in any application. We recommend choosing a value of C so that the type I error is between 4% and 6% when the nominal value is 5%. This can be achieved by reading simulation results in the literature, or by a trial simulation at the sample size of the actual application. For example, in Table 1 with 10,000 replications, $n_1 = 150$ and $n_2 = 70$, the standard error of the estimated size is about 0.2%. So with the simulated type I error rate 5.6% for $MR(1)$, we could say that the level is between 5.2% and 6% when $C = 1$. The choice of $C = 1$ has also been found to be satisfactory for data with multinomial component distributions; see Chen (1998) and Fu, Chen and Kalbfleisch (2006).

Acknowledgement

The research was partially supported by grants from the Natural Sciences and Engineering Research Council of Canada. The authors would like to thank the editors, an associate editor and two referees for their valuable suggestions and comments.

References

Boos, D. D. and Brownie, C. (1986). Testing for a treatment effect in the presence of nonresponders. *Biometrics* **42**, 191-197.

Boos, D. D. and Brownie, C. (1991). Mixture models for continuous data in dose-response studies when some animals are unaffected by treatment. *Biometrics* **47**, 1489-1504.

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

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

Chen, H. and Chen, J. (2003). Tests for homogeneity in normal mixtures with presence of a structural parameter. *Statist. Sinica* **13**, 327-350.

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

Chen, H., Chen, J. and Kalbfleisch, J. D. (2004). Testing for a finite mixture model with two components. *J. Roy. Statist. Soc. Ser. B* **66**, no. 1, 95-115.

Chen, J. and Kalbfleisch, J. D. (2005). Modified likelihood ratio test in finite mixture models with a structural parameter. *J. Statist. Plann. Inference* **129**, 93-107.

Conover, W. J. and Salsburg, D. S. (1988). Locally most powerful tests for treatment effects when only a subset of patients can be expected to respond to treatment. *Biometrics* **44**, 189-196.

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

Fu, Y., Chen, J. and Kalbfleisch, J. D. (2006). Testing for homogeneity in genetic linkage analysis. *Statist. Sinica* **16**, 805-823.

Good, P. I. (1979). Detection of a treatment effect when not all experimental subjects will respond to treatment. *Biometrics* **35**, 483-489.

Hájek, J. and Šidák, Z. (1967). *Theory of Rank Tests.* Academic Press, New York.

Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist* **27**, 887-906.

Lancaster, T. and Imbens, G. (1996). Case-control studies with contaminated controls. *J. Econometrics* **71**, 145-160.

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

Liu, X. and Shao, Y. Z. (2003). Asymptotics for likelihood ratio tests under loss of identifiability. *Ann. Statist.* **31**, 807-832.

Moskvina, V., Holmans, P., Schmidt, K. M. and Craddock, N. (2005). Design of case-controls studies with unscreened controls. *Ann. Human Genetics* **69**, 566-576.

Razzaghi, M. and Kodell, R. L. (2000). Risk assessment for quantitative responses using a mixture model. *Biometrics* **56**, 519-527.

Razzaghi, M. and Nanthakumar, A. (1994). A distribution free test for detecting a treatment effect in the presence of nonresponders. *Biometrical J.* **36**, 373-384.

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

Weeks, J. R. and Collins, R. J. (1971). Primary addiction to morphine in rats. *Federation Proceedings* **30**, 277 ABS.

Zhu, H. T. and Zhang, H. P. (2004). Hypothesis testing in mixture regression models. *J. Roy. Statist. Soc. Ser. B* **66**, 3-16.

Department of Mathematics and Statistics, York University, Toronto, ON, M3J 1P3, Canada.

E-mail: yuejiao@mathstat.yorku.ca

Department of Statistics, University of British Columbia, Vancouver, BC, V6T 1Z2, Canada.

E-mail: jhchen@stat.ubc.ca

Department of Biostatistics, School of Public Health, University of Michigan, 1420 Washington Heights, Ann Arbor, MI 48109-2029, U.S.A.

E-mail: jdkalbfl@umich.edu

(Received October 2007; accepted July 2008)