Statistica Sinica **20** (2010), 497-512

ASYMPTOTIC RELATIONSHIPS BETWEEN THE D-TEST AND LIKELIHOOD RATIO-TYPE TESTS FOR HOMOGENEITY

Richard Charnigo and Jiayang Sun

University of Kentucky and Case Western Reserve University

Abstract: The D-test for homogeneity in finite mixtures is appealing because the D-test statistic depends on the data solely through parameter estimates, whereas likelihood ratio-type test statistics require both parameter estimates and the full data set. In this paper we establish asymptotic equivalences between the D-test and three likelihood ratio-type tests for homogeneity. The first two equivalences, under maximum likelihood and Bayesian estimation frameworks respectively, apply to mixtures from a one-dimensional exponential family; the second equivalence yields a simple limiting null distribution for the D-test statistic as well as a simple limiting distribution under contiguous local alternatives, revealing that the D-test is asymptotically locally most powerful. The third equivalence, under an empirical Bayesian estimation framework, pertains to mixtures from a normal location family with unknown structural parameter; the third equivalence also yields a simple limiting null distribution for the D-test statistic. Simulation studies are provided to investigate finite-sample accuracy of critical values based on the limiting null distributions and to compare the D-test to its competitors regarding power to detect heterogeneity. We conclude with an application to medical data and a discussion emphasizing computational advantages of the D-test.

Key words and phrases: D-test, homogeneity, L^2 distance, mixture model, structural parameter.

1. Introduction

Mixture models are useful for describing complex, heterogeneous populations. Everitt and Hand (1981), Lemdani and Pons (1997), Ott (1999), and Sun, Morrison, Harding, and Woodroofe (2009) discuss applications of mixture modeling to genetics, astronomy, and other sciences. However, determining the number of distinct components in a finite mixture is challenging (Titterington, Smith and Makov (1985); Lindsay (1995)), exemplified by the non-identifiability of mixture parameters under the null hypothesis of homogeneity (i.e., when the mixture has only one distinct component). In fact, the usual chi-square theory for (negative twice) the (log) likelihood ratio (LR) test statistic is not valid (see, e.g., Ghosh and Sen (1985); Hartigan (1985); Bickel and Chernoff (1993); Chernoff and Lander (1995)).

Under fairly general conditions, the asymptotic null distribution of the LR test statistic is that of $\sup_{\theta\in\Theta}\{\mathcal{W}^+(\theta)\}^2$, where $\mathcal{W}^+$ is the positive part of a Gaussian process $\mathcal{W}$ and Θ is the compact parameter space for the component-specific parameters (see, e.g., Dacunha-Castelle and Gassiat (1999); Chen and Chen (2001); Liu and Shao (2003)). Inferences can be made by computing tail probabilities for $\sup_{\theta\in\Theta}\{\mathcal{W}^+(\theta)\}^2$ using the random field theory in Sun (1993).

Chen, Chen and Kalbfleisch (2001) took a different approach in their seminal work on modified likelihood ratio (MLR) testing. They proposed a test for homogeneity based on the maximum values attained by a penalized log likelihood in restricted (homogeneous) and unrestricted (heterogeneous) models. The estimated weights of the mixture components in the unrestricted model are forced away from zero by the penalty term, which reduces the effects of the non-identifiability. If θ is one-dimensional and regularity conditions are satisfied, then the asymptotic null distribution of the MLR test statistic is an equal-probability mixture of a point mass at zero and a chi-square distribution on one degree of freedom. Chen, Chen and Kalbfleisch (2001) report that the MLR test is competitive with the $C(\alpha)$ test of Neyman and Scott (1966), the bootstrap test by McLachlan (1987), and the methodology of Davies (1987). Thus, we adopt the MLR test as a benchmark in our simulation studies on mixtures from a one-dimensional exponential family (Section 6).

Li, Chen and Marriott (2009) subsequently developed the "EM test" for homogeneity. The EM test is like the MLR test except that: (i) only a few iterations of the expectation maximization (EM) algorithm are run to secure parameter estimates; and, (ii) multiple sets of initial values may be explicitly considered. If θ is one-dimensional, then the asymptotic null distribution of the EM test statistic is the same as that of the MLR test statistic; however, the EM test statistic achieves this with weaker regularity conditions (in particular, no compactness requirement on Θ). Chen and Li (2009) adapted the EM test to normal location mixtures with unknown structural (standard deviation) parameter; the EM test statistic has an analytically tractable asymptotic null distribution in this setting (dependent on what sets of initial values are considered), whereas the MLR test statistic does not. Hence, we view the EM test as a benchmark in our simulation studies on normal location mixtures with unknown structural parameter (Section 6).

Charnigo and Sun (2004) broke away from the LR paradigm and introduced the D-test for homogeneity in finite continuous mixtures. The D-test has a simple geometric motivation: the L^2 distance between fitted homogeneous and heterogeneous models is likely to be large when the null hypothesis is false, so a test based on this L^2 distance is likely to detect multiple distinct components when they are present. The D-test statistic is denoted $d(k, n)$, where k is the number of distinct components under the alternative hypothesis and n is the sample size. While LR-type test statistics cannot be calculated without the full data set, $d(k, n)$ depends on the data only through mixture parameter estimates. Thus, $d(k, n)$ can be computed more quickly than an LR-type test statistic if: (i) parameter estimates are available from a previous review of the data set; or, (ii) updating formulas for parameter estimates can accommodate a new addition to the data set without recalling the old portion of the data set. Point (ii) has become a reality with the partial EM estimator introduced by Sun, Liu, and Chen (2009). We will elaborate on this in Section 7.

In this paper we investigate the asymptotic behavior of the D-test statistic and establish three unanticipated equivalences: one between the D-test and LR test for mixtures from a one-dimensional exponential family, one between the D-test and MLR test, and one between the D-test and EM test for normal location mixtures with unknown structural parameter. Let R_n and M_n denote the LR and MLR test statistics, let $E_n(\alpha_0)$ denote the EM test statistic based on one set of initial values that includes the initial value $\alpha_0 \in (0, 0.5]$ for α (the weight of the second mixture component), and let $d(2, n)$ be abbreviated to d_n.

With a maximum likelihood framework for estimating mixture parameters, we show that $n\, d_n = C^*(\hat{\theta}_2)R_n + o_p(1)$ under the null hypothesis of homogeneity (Theorem 3.1 in Section 3), where $C^*(\theta)$ is a computable deterministic function of its argument and $\hat{\theta}_2$ estimates the component-specific parameter in the second mixture component.

With a Bayesian framework corresponding to the penalized likelihood of Chen, Chen and Kalbfleisch (2001), we prove that $n\, d_n = C^*(\theta_0)M_n + o_p(1)$ under the null hypothesis (Theorem 4.1 in Section 4), where θ_0 is the true value of θ in the restricted (homogeneous) model. Thus, $n\, d_n\, C^*(\theta_0)^{-1}$ converges in law to an equal-probability mixture of a point mass at zero and a chi-square random variable on one degree of freedom, yielding an asymptotic critical value of $n^{-1}C^*(\theta_0)\chi^2_{1,2\alpha}$ for d_n (Corollary 4.1). Here $\chi^2_{1,2\alpha}$ denotes the upper 2α quantile of the chi-square distribution on one degree of freedom, where the subscript α is the significance level, not the weight of the second mixture component. Moreover, $n\, d_n\, C^*(\theta_0)^{-1}$ converges in law under contiguous local alternatives (Corollary 4.2), and the limiting distribution is such that the D-test is asymptotically locally most powerful.

With an empirical Bayesian framework corresponding to the penalized likelihood of Chen and Li (2009), we find that $n\, d_n = A^*(\alpha_0)\{E_n(\alpha_0) - 2\log[2\alpha_0]\} + o_p(1)$ under the null hypothesis (Theorem 5.1 in Section 5), where $A^*(\alpha)$ is an estimable deterministic function of its argument. If $\alpha_0 = 0.5$, then $A^*(\alpha_0)^{-1} n\, d_n$ converges in law to an equal-probability mixture of a point mass at zero and a chi-square random variable on one degree of freedom, producing an asymptotic critical value of $n^{-1}A^*(\alpha_0)\chi^2_{1,2\alpha}$ for d_n; if $\alpha_0 < 0.5$, then $A^*(\alpha_0)^{-1}n\, d_n$ converges to a chi-square random variable on one degree of freedom, yielding an asymptotic critical value of $n^{-1}A^*(\alpha_0)\chi^2_{1,\alpha}$ (Corollary 5.1).

In addition to developing large-sample theory, we study the finite-sample accuracy of critical values suggested by the second and third equivalences (Section 6.1). We also compare the D-test to the MLR test and the EM test regarding power to detect heterogeneity (Section 6.2). A practical illustration is provided with a medical data set (Section 6.3). The paper concludes with a discussion (Section 7) to explain further why the D-test is a competitive alternative to LR-type tests, especially for large and online data sets. Proofs and regularity conditions appear in the online Appendix, `http://www.stat.sinica.edu.tw/statistica`.

2. Models and Test Statistics

A general mixture model. Let $X_1, \ldots, X_n$ be a simple random sample from the general two-component mixture model

$$(1-\alpha)f_\sigma(x,\theta_1) + \alpha\, f_\sigma(x,\theta_2), \tag{2.1}$$

where $0 \le \alpha \le 1/2$ and σ denotes, if applicable, an unknown structural parameter common to both components. The σ subscripts are dropped if there is no unknown structural parameter.

Mixture from a one-dimensional exponential family. A specific model of interest to us, in which there is no unknown structural parameter, takes

$$f(x,\theta) := a(x)\exp[-b(\theta) + t(x)\theta] \tag{2.2}$$

in (2.1). Besides assuming that (2.2) belongs to a minimal exponential family (Cf., Brown (1986)), we require the regularity conditions in the Appendix (Conditions A1-A5). We then refer to (2.2) as *regular*, or say that it belongs to a *regular exponential family of distributions.*

Normal location mixture with unknown structural parameter. A second specific model of interest puts

$$f_\sigma(x,\theta) := (2\pi\sigma^2)^{-1/2}\exp[-\frac{(x-\theta)^2}{2\sigma^2}] \tag{2.3}$$

in (2.1). Consequently, both mixture components have the same standard deviation.

Four tests of homogeneity. We now describe four approaches to testing $H_0: \alpha(\theta_2 - \theta_1) = 0$ against $H_1: \alpha(\theta_2 - \theta_1) \neq 0$ in (2.1). The null hypothesis says that (2.1) simplifies to $f_{\sigma_0}(x,\theta_0)$, where the use of σ_0 here (rather than σ) helps to avoid ambiguity in defining the test statistics below.

The D-test. Let $\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\alpha}$, and $\hat{\sigma}$ denote estimators (maximum likelihood, Bayesian, empirical Bayesian, or any other kind) of the corresponding parameters in the unrestricted (heterogeneous) model allowed by the alternative hypothesis. Let $\hat{\theta}_0$ and $\hat{\sigma}_0$ denote estimators of the corresponding parameters in the restricted (homogeneous) model determined by the null hypothesis. The D-test statistic (Charnigo and Sun (2004)) is

$$d_n := \int \Big[(1-\hat{\alpha})f_{\hat{\sigma}}(x,\hat{\theta}_1) + \hat{\alpha}\, f_{\hat{\sigma}}(x,\hat{\theta}_2) - f_{\hat{\sigma}_0}(x,\hat{\theta}_0)\Big]^2\, dx. \tag{2.4}$$

Clearly, d_n depends on $X_1,\ldots,X_n$ only through the estimators $\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\alpha}$, $\hat{\sigma}$, and $\hat{\theta}_0$, $\hat{\sigma}_0$. The integral in (2.4) can usually be evaluated analytically to produce a simple closed-form expression. Although x and θ will be one-dimensional in this paper, we note that (2.4) is well-defined for multi-dimensional x and θ. Expression (6) and Section 4.2 of Charnigo and Sun (2004) consider multi-dimensional x and θ, respectively.

The LR test. The LR test statistic is $R_n := 2l_n(\hat{\alpha},\hat{\theta}_1,\hat{\theta}_2,\hat{\sigma}) - 2l_n(1/2,\hat{\theta}_0,\hat{\theta}_0,\hat{\sigma}_0)$, where

$$l_n(\alpha,\theta_1,\theta_2,\sigma) := \sum_{i=1}^{n} \log\left[(1-\alpha)f_\sigma(X_i,\theta_1) + \alpha\, f_\sigma(X_i,\theta_2)\right] \tag{2.5}$$

is the log likelihood and the estimators $\hat{\alpha},\hat{\theta}_1,\hat{\theta}_2,\hat{\sigma}$, and $\hat{\theta}_0,\hat{\sigma}_0$ maximize (2.5) for the unrestricted and restricted models, respectively.

The MLR test. The MLR test statistic (Chen, Chen and Kalbfleisch (2001)) is $M_n := 2l_n^*(\hat{\alpha},\hat{\theta}_1,\hat{\theta}_2,\hat{\sigma}) - 2l_n^*(1/2,\hat{\theta}_0,\hat{\theta}_0,\hat{\sigma}_0)$, where

$$l_n^*(\alpha,\theta_1,\theta_2,\sigma) := \sum_{i=1}^{n} \log\left[(1-\alpha)f_\sigma(X_i,\theta_1) + \alpha\, f_\sigma(X_i,\theta_2)\right] + C\log\left[4\alpha(1-\alpha)\right] \tag{2.6}$$

is a penalized log likelihood and the estimators $\hat{\alpha},\hat{\theta}_1,\hat{\theta}_2,\hat{\sigma}$, and $\hat{\theta}_0,\hat{\sigma}_0$ maximize (2.6) for the unrestricted and restricted models, respectively. The specific value of C does not affect the development of asymptotic theory; for our simulation studies we set $C := \log 10$ based on documented success with that choice (Chen, Chen and Kalbfleisch (2001)).

The EM test. Assuming one set of initial values that includes the initial value $\alpha_0 \in (0, 0.5]$ for α, the EM test statistic (Li, Chen and Marriott (2009); Chen and Li (2009)) is $E_n(\alpha_0) := 2l_n^\dagger(\hat{\alpha},\hat{\theta}_1,\hat{\theta}_2,\hat{\sigma}) - 2l_n^\dagger(1/2,\hat{\theta}_0,\hat{\theta}_0,\hat{\sigma}_0)$, where for (2.3) we have

$$\begin{aligned} l_n^\dagger(\alpha,\theta_1,\theta_2,\sigma) := &\sum_{i=1}^{n} \log\left[(1-\alpha)f_\sigma(X_i,\theta_1) + \alpha\, f_\sigma(X_i,\theta_2)\right] + \log\left\{1 - |1-2\alpha|\right\} \\ &-\left\{\frac{S^2}{\sigma^2} + \log\left(\frac{\sigma^2}{S^2}\right)\right\} \end{aligned} \tag{2.7}$$

with $S^2 := n^{-1}\sum_{i=1}^n (X_i - \bar{X})^2$. The estimators $\hat{\alpha}, \hat{\theta}_1, \hat{\theta}_2, \hat{\sigma}$ are obtained by applying K iterations of the EM algorithm to (2.7), while the estimators $\hat{\theta}_0, \hat{\sigma}_0$ maximize (2.7) for the restricted model. The specific value of K does not affect the development of asymptotic theory; for our simulation studies we set $K := 1$ based on documented success with that choice (Chen and Li (2009)). If one wishes to consider J sets of initial values that include the initial values $\alpha_{0;1}, \ldots, \alpha_{0;J} \in (0, 0.5]$ for α, then the EM test statistic is the maximum of $E_n(\alpha_{0;1}), \ldots, E_n(\alpha_{0;J})$. We assume a single set of initial values when developing the asymptotic equivalence in Theorem 5.1, but we address the possibility of multiple sets of initial values after stating Corollary 5.1.

3. Asymptotic Equivalence: D-Test and LR Test

In this section we consider model (2.2) and suppose that $\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\alpha}$, and $\hat{\theta}_0$ are *maximum likelihood* estimators. Our main result is the following.

Theorem 3.1. *Suppose that $f(x,\theta)$ is regular and that the null hypothesis is true. Then, under a maximum likelihood estimation framework, as $n \to \infty$, $n\, d_n = C^*(\hat{\theta}_2) R_n + o_p(1)$. Here $C^*(\theta)$ is an asymptotic scaling factor defined in* (3.1)−(3.2) *below.*

The proof of Theorem 3.1, along with the proofs of other results in this paper, can be found in the Appendix.

To describe $C^*(\theta)$, we need to define certain *key quantities* that have appeared previously, without the names given here, in Chen and Chen (2001).

1. The quasi-derivative of $\log[f(X_i, \theta)]$ is
$$Y_i(\theta) := \begin{cases} \frac{f(X_i,\theta) - f(X_i,\theta_0)}{(\theta-\theta_0) f(X_i,\theta_0)}, & \text{if } \theta \neq \theta_0 \\ \frac{\frac{\partial}{\partial\theta} f(X_i,\theta)|_{\theta=\theta_0}}{f(X_i,\theta_0)}, & \text{if } \theta = \theta_0 \end{cases}.$$
2. The quasi-derivative of $Y_i(\theta)$ is
$$Z_i(\theta) := \begin{cases} \frac{Y_i(\theta) - Y_i(\theta_0)}{\theta-\theta_0}, & \text{if } \theta \neq \theta_0 \\ Y_i'(\theta_0), & \text{if } \theta = \theta_0 \end{cases}.$$
3. The regression coefficient is $h(\theta) := E[Y_i(\theta_0) Z_i(\theta)]/E[Y_i^2(\theta_0)]$, where the expectations are taken with respect to $f(x, \theta_0)$.
4. The residual from regression of $Z_i(\theta)$ on $Y_i(\theta_0)$ is $W_i(\theta) := Z_i(\theta) - Y_i(\theta_0) h(\theta)$.

The following lemma, stated here rather than in the Appendix because it is useful in evaluating $C^*(\theta)$, shows that these quantities are simple functions of $t(X_i)$ and the derivatives of $b(\theta)$, where $t(x)$ and $b(\theta)$ are as defined in (2.2).

Lemma 3.1. *Suppose that $f(x, \theta)$ is regular. Then*

(i) $Y_i(\theta_0) = t(X_i) - b'(\theta_0)$;

(ii) $Z_i(\theta_0) = (1/2)\left[-b''(\theta_0) + \{t(X_i) - b'(\theta_0)\}^2\right]$;

(iii) $E[Y_i(\theta_0)^2] = b''(\theta_0)$; *and,*

(iv) $h(\theta_0) = (1/2)b'''(\theta_0)/b''(\theta_0)$.

We now define the *asymptotic scaling factor* $C^*(\theta)$ in Theorem 3.1. Let $g_j(x, \theta) := \frac{\partial^j}{\partial\theta^j} f(x, \theta)$ for $j = 1, 2, 3$, and put

$$C^*(\theta) := \frac{1}{E[W_i(\theta)^2]} \int \left[-h(\theta)g_1(x, \theta_0) + \frac{f(x,\theta) - f(x,\theta_0) - (\theta-\theta_0)g_1(x,\theta_0)}{(\theta-\theta_0)^2}\right]^2 dx \tag{3.1}$$

for $\theta \neq \theta_0$, and

$$C^*(\theta_0) := \frac{1}{E[W_i(\theta_0)^2]} \int \left[-h(\theta_0)g_1(x, \theta_0) + \frac{g_2(x,\theta_0)}{2}\right]^2 dx = \frac{E\left[W_i(\theta_0)^2 f(X_i, \theta_0)\right]}{E[W_i(\theta_0)^2]}. \tag{3.2}$$

For instance, if $f(x, \theta) := (2\pi\tau^2)^{-1/2} \exp[-(x-\theta)^2/(2\tau^2)]$ for known $\tau > 0$, then $C^*(\theta_0) = (3/16)\pi^{-1/2}\tau^{-1}$.

Theorem 3.1 provides an unanticipated connection between the D-test and the LR test. In particular, Theorem 3.1 implies that $n\,d_n$ converges in law to $C^*(\tilde{\theta}) \sup_{\theta\in\Theta}\{\mathcal{W}^+(\theta)\}^2$ when the null hypothesis is true, where $\tilde{\theta} := \arg\sup_{\theta\in\Theta} \{\mathcal{W}^+(\theta)\}^2$ and $\mathcal{W}$ is the Gaussian process used to characterize the limiting distribution of R_n (Dacunha-Castelle and Gassiat (1999); Chen and Chen (2001); Liu and Shao (2003)). Thus, Theorem 3.1 allows application of random field theory (Sun (1993)) to the calculation of critical values for the D-test statistic in model (2.2) with a maximum likelihood estimation framework. However, we do not pursue random field theory here since an easier option for calculating critical values is available with a Bayesian estimation framework in Section 4.

4. Asymptotic Equivalence: D-Test and MLR Test

In this section we consider model (2.2) and suppose that $\hat{\theta}_1$, $\hat{\theta}_2$, and $\hat{\alpha}$ are *Bayesian maximum a posteriori* estimators. The prior distribution is $\pi(\alpha, \theta_1, \theta_2) \propto (4\alpha(1-\alpha))^C$ for some $C > 0$. With this prior, *maximum a posteriori* estimation is equivalent to maximum penalized likelihood estimation with penalty as in (2.6). Note that $\hat{\theta}_0$ is actually a maximum likelihood estimator since α effectively disappears in the restricted (homogeneous) model.

The main result of this section is another unanticipated connection between the D-test and an LR-type test.

Theorem 4.1. *Suppose $f(x, \theta)$ is regular and the null hypothesis is true. Under the indicated Bayesian estimation framework, as $n \to \infty$, $n\, d_n = C^*(\theta_0)\, M_n + o_p(1)$.*

Corollary 4.1. *Under the conditions of Theorem 4.1, the asymptotic distribution of $n\, d_n\, C^*(\theta_0)^{-1}$ is $0.5\chi_0^2 + 0.5\chi_1^2$, an equal-probability mixture of a point mass at zero and a chi-square distribution on one degree of freedom. Hence, the asymptotic level α critical value for d_n is $C^*(\theta_0) n^{-1} \chi_{1,2\alpha}^2$.*

For some parametric families, $C^*(\theta_0)$ can be evaluated without knowledge of θ_0. For other parametric families, we can replace $C^*(\theta_0)$ by $C^*(\hat{\theta}_0)$ since $\hat{\theta}_0$ converges to θ_0 when the null hypothesis is true. In any case, computation of $C^*(\theta_0)$ or $C^*(\hat{\theta}_0)$ is no obstacle to using the D-test. Also, we note that the asymptotic level α critical value does *not* depend on the size of Θ, as long as Θ is compact.

Corollary 4.2. *Suppose $f(x, \theta)$ is regular and the following contiguous local alternatives hold: $\alpha = \alpha_0 \in (0, 1)$, $\theta_1 = \theta_0 - n^{-1/4}\tau_1$, and $\theta_2 = \theta_0 + n^{-1/4}\tau_2$, where $\tau_1 = \tau\sqrt{\alpha_0/(1-\alpha_0)}$ and $\tau_2 = \tau\sqrt{(1-\alpha_0)/\alpha_0}$ for some fixed $\tau > 0$. Then, with the indicated Bayesian estimation framework, $n\ d_n\, C^*(\theta_0)^{-1}$ converges in law to $\left\{\left(Z + \tau^2 E[W_i(\theta_0)^2]^{1/2}\right)^+\right\}^2$, where Z is standard normal and the expectation is with respect to $f(x, \theta_0)$.*

This second corollary, which follows from the definition of contiguity and Theorem 2 in Chen, Chen and Kalbfleisch (2001), assures us that the Bayesian-framework D-test is asymptotically locally most powerful.

5. Asymptotic Equivalence: D-Test and EM Test

In this section we consider model (2.3) and suppose that $\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\alpha}$, and $\hat{\sigma}$ are "approximate" *empirical Bayesian maximum a posteriori* estimators. The prior distribution is $\pi(\alpha, \theta_1, \theta_2, \sigma) \propto \{1 - |1 - 2\alpha|\} \exp\left[-S^2/\sigma^2\right] S^2/\sigma^2$, with $S^2 := n^{-1}\sum_{i=1}^n (X_i - \bar{X})^2$. The reason for the "approximate" designation is that we run only K iterations of the EM algorithm to secure $\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\alpha}$, and $\hat{\sigma}$. Thus, these estimators correspond to those used for the EM test when there is one set of initial values (Chen and Li (2009)). Note that $\hat{\theta}_0$ and $\hat{\sigma}_0$ are actually maximum likelihood estimators. Also, unlike for model (2.2), we now take Θ to be the real line rather than a compact subset. Likewise, we place no restriction on σ except that it be positive.

The main result of this section is a third unanticipated connection between the D-test and an LR-type test.

Theorem 5.1. *Suppose model* (2.3) *applies and the null hypothesis is true. Under the indicated empirical Bayesian estimation framework, as* $n \to \infty$, $n\, d_n = A^*(\alpha_0)\{E_n(\alpha_0) - 2\log[2\alpha_0]\} + o_p(1)$, *where* α_0 *is the initial value for* α *and*

$$A^*(\alpha) := \frac{5}{32\pi^{1/2}\sigma_0} \text{ for } \alpha < 0.5, \quad A^*(0.5) := \frac{35}{256\pi^{1/2}\sigma_0}. \tag{5.1}$$

Note that Slutsky's Theorem permits us to substitute $\hat{\sigma}_0$ for σ_0 without disturbing the conclusions of Theorem 5.1 or of the following corollary.

Corollary 5.1. *Under the conditions of Theorem* 5.1, *with* $\alpha_0 = 0.5$ *the asymptotic distribution of* $n\, d_n\, \{256/35\}\pi^{1/2}\,\sigma_0$ *is* $0.5\chi^2_0 + 0.5\chi^2_1$, *an equal-probability mixture of a point mass at zero and a chi-square distribution on one degree of freedom, and with* $\alpha_0 < 0.5$ *the asymptotic distribution of* $n\, d_n\, \{32/5\}\pi^{1/2}\ \sigma_0$ *is* χ^2_1, *a chi-square distribution on one degree of freedom. Hence, the asymptotic level* α *critical value for* d_n *is* $\{35/256\}\pi^{-1/2}\,\sigma_0^{-1} n^{-1}\chi^2_{1,2\alpha}$ *or* $\{5/32\}\pi^{-1/2}\,\sigma_0^{-1} n^{-1}\chi^2_{1,\alpha}$, *according as* $\alpha_0 = 0.5$ *or* $\alpha_0 < 0.5$.

Chen and Li (2009) suggest performing the EM test with $J = 3$ sets of initial values that include initial values of $\alpha_{0;1} = 0.5$, $\alpha_{0;2} = 0.3$, and $\alpha_{0;3} = 0.1$ for α. The reason for this suggestion is that the EM test, if performed with a single set of initial values, has difficulty detecting heterogeneity for specific alternatives in which α is far away from α_0. This motivates us to ask whether the D-test can also be performed with multiple sets of initial values and, if so, how.

Let $d_n(\alpha_{0;1}), \ldots, d_n(\alpha_{0;J})$ denote the D-test statistics corresponding to the initial values $\alpha_{0;1} = 0.5, \alpha_{0;2} < 0.5, \ldots, \alpha_{0;J} < 0.5$. Let

$$\widetilde{d}_n := \max\{\{\frac{256}{35}\}d_n(\alpha_{0;1}), \{\frac{32}{5}\}d_n(\alpha_{0;2}), \ldots, \{\frac{32}{5}\}d_n(\alpha_{0;J})\}. \tag{5.2}$$

Under the null hypothesis, $E_n(\alpha_{0;2})$ through $E_n(\alpha_{0;J})$ differ asymptotically only by additive constants of the form $2\log[2\alpha_{0;j}]$ (Chen and Li (2009)). Moreover, $E_n(\alpha_{0;2})$ through $E_n(\alpha_{0;J})$ are asymptotically independent of $E_n(\alpha_{0;1})$ (Chen and Li (2009)). Hence, Corollary 5.1 implies that $n\,\widetilde{d}_n\,\pi^{1/2}\,\hat{\sigma}_0$ converges in law to a random variable with distribution function $\mathbb{F}_1(x)\{0.5 + 0.5\mathbb{F}_1(x)\}$, where $\mathbb{F}_1(x)$ denotes the distribution function of a chi-square random variable on one degree of freedom. So, for example, rejection of the null hypothesis when $\widetilde{d}_n > 4.509\, n^{-1}\pi^{-1/2}\hat{\sigma}_0^{-1}$ provides a test with asymptotic significance level 0.05.

Before leaving this section, we comment on the discontinuity in asymptotic behavior as α_0 reaches 0.5. This occurs because, as shown in Chen and Li (2009), the third-order term in the Taylor expansion for $2l^\dagger_n(\hat{\alpha}, \hat{\theta}_1, \hat{\theta}_2, \hat{\sigma}) - 2l^\dagger_n(1/2, 0, 0, 1)$ includes a factor of $(1 - 2\hat{\alpha})$. When $\alpha_0 = 0.5$, the third-order term disappears asymptotically and the fourth-order term becomes relevant; when $\alpha_0 < 0.5$, the third-order term does not vanish and the fourth-order term does not enter. Moreover, since Chen and Li (2009) find that $\hat{\alpha} - \alpha_0 = O_p(n^{-1/4})$, we conjecture that $n\, d_n = A^*(0.5)E_n(0.5) + o_p(1)$ under the conditions of Theorem 5.1 if the initial value for α is $0.5 - \xi n^{-1/4}$, where $\xi > 0$.

6. Simulation Studies and Data Application

Previous simulation experiments (Charnigo and Sun (2004)) have shown that the *maximum likelihood-framework D-test* (Section 3) is competitive with the MLR test. Now we describe new simulation studies examining the performances of the *Bayesian-framework D-test* (Section 4) and the *empirical Bayesian-framework D-test* (Section 5). We also estimate the Type I error rates associated with the asymptotic level 0.05 critical values of Corollary 4.1 and Corollary 5.1. This section concludes with application of the empirical Bayesian-framework D-test to the SLC data set (Roeder (1994)).

6.1. Assessing the Bayesian-framework D-test

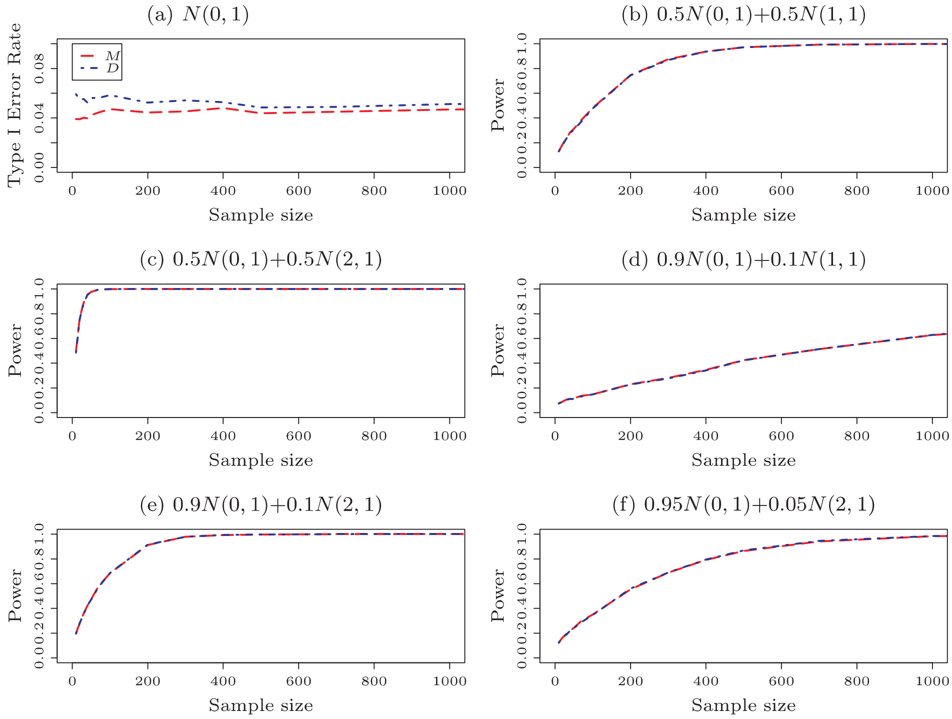
First we generated $10,000$ random samples of size n from the standard normal distribution $N(0,1)$ for each of various n. From each sample we calculated d_n and M_n to test $H_0 : N(\theta_0, 1)$ against $H_1 : (1-\alpha)N(\theta_1, 1) + \alpha N(\theta_2, 1)$. Panel a of Figure 1 shows the frequencies with which d_n and M_n exceeded their asymptotic level 0.05 critical values. The estimated Type I error rates were close to 0.05 for both d_n and M_n, typically slightly less than 0.05 for M_n, and slightly greater than 0.05 for d_n. For a normal location mixture with known component standard deviation, we can comfortably invoke Corollary 4.1 when $n \geq 10$.

Next we generated $10,000$ random samples of size n from the two-component mixture $0.5N(0,1) + 0.5N(1,1)$. From each sample we calculated d_n and M_n. Panel b of Figure 1 shows the frequencies with which d_n and M_n exceeded their actual level 0.05 critical values (i.e., their upper 0.05 quantiles under $N(0,1)$, as estimated during the simulation for panel a). These steps were repeated for $0.5N(0,1) + 0.5N(2,1)$, $0.9N(0,1) + 0.1N(1,1)$, $0.9N(0,1) + 0.1N(2,1)$, and $0.95N(0,1) + 0.05N(2,1)$; the corresponding results appear in panels c through f. In all cases, the power curves of the Bayesian-framework D-test are nearly indistinguishable from those of the MLR test.

6.2. Assessing the empirical Bayesian-framework D-test

We proceed much as in Section 6.1 except that now we are testing $H_0 : N(\theta_0, \sigma_0^2)$ against $H_1 : (1-\alpha)N(\theta_1, \sigma^2) + \alpha N(\theta_2, \sigma^2)$ and are using the EM test as the benchmark. Let $E_n(0.5)$, $E_n(0.3)$, and $E_n(0.1)$ denote the EM test statistics based on initial values of 0.5, 0.3, and 0.1 for α; let $d_n(0.5)$, $d_n(0.3)$,

Figure 1. Type I error and power comparisons of MLR test and Bayesian-framework D-test

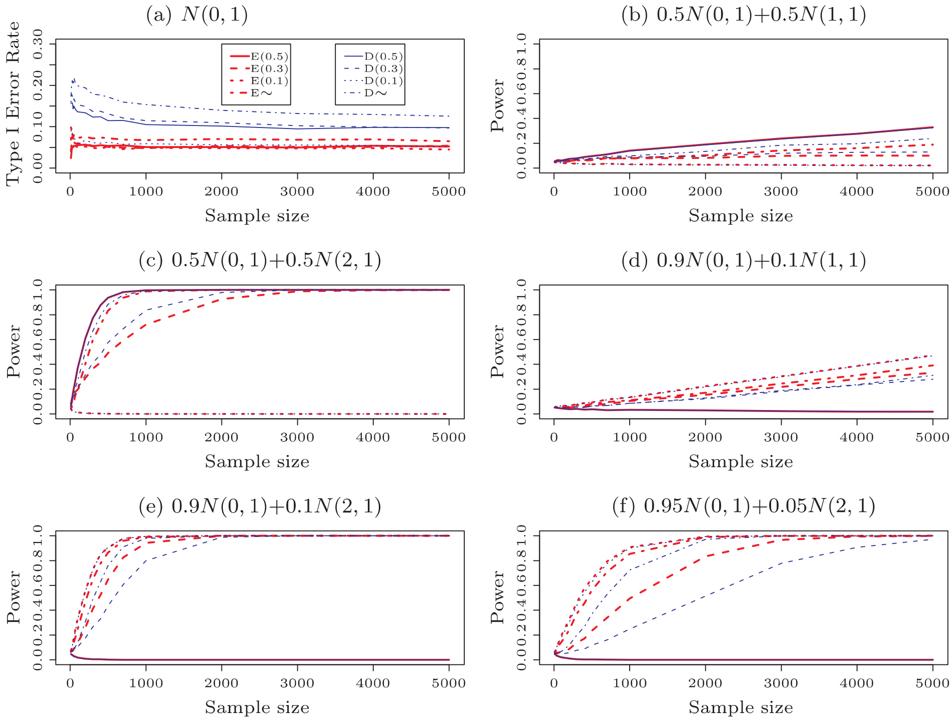


Panel a: $10,000$ random samples of size n are generated from $N(0,1)$ for each of various n from 10 to $1,000$. For each random sample we test $H_0 : N(\theta_0, 1)$ against $H_1 : (1-\alpha)N(\theta_1, 1) + \alpha N(\theta_2, 1)$. Fractions of MLR test and Bayesian-framework D-test statistics exceeding their asymptotic level 0.05 critical values are displayed, as a function of n. Panels b through f: $10,000$ random samples of size n are generated from the indicated mixture distributions for each of various n. Fractions of MLR test and Bayesian-framework D-test statistics exceeding their actual level 0.05 critical values are displayed.

and $d_n(0.1)$ denote the analogous empirical Bayesian-framework D-test statistics. Put $\widetilde{E}_n := \max\{E_n(0.5), E_n(0.3) - 2\log[0.6], E_n(0.1) - 2\log[0.2]\}$, with reference distribution function $\mathbb{F}_1(x)\{0.5+0.5\mathbb{F}_1(x)\}$, and let $\widetilde{d}_n$ be as defined in (5.2) with $J := 3$, $\alpha_{0;1} := 0.5$, $\alpha_{0;2} := 0.3$, and $\alpha_{0;3} := 0.1$.

Panel a of Figure 2 shows that $d_n(0.1)$, $E_n(0.5)$, $E_n(0.3)$, and $E_n(0.1)$ had estimated Type I error rates close to 0.05, while those of $\widetilde{E}_n$ were only modestly above 0.05. The estimated Type I error rates for $d_n(0.5)$, $d_n(0.3)$, and $\widetilde{d}_n$ were very slow to approach 0.05, remaining well above 0.05 even when $n = 5,000$. For a normal location mixture with unknown structural parameter and with $\alpha_0 = 0.1$,

Figure 2. Type I error and power comparisons of EM test and empirical Bayesian-framework D-test.



Panel a: $10,000$ random samples of size n are generated from $N(0,1)$ for each of various n from 10 to $5,000$. For each random sample we test $H_0 : N(\theta_0, \sigma_0^2)$ against $H_1 : (1-\alpha)N(\theta_1, \sigma^2) + \alpha N(\theta_2, \sigma^2)$. Fractions of EM test and empirical Bayesian-framework D-test statistics exceeding their asymptotic level 0.05 critical values are displayed. Panels b through f: $10,000$ random samples of size n are generated from the indicated mixture distributions for each of various n. Fractions of EM test and empirical Bayesian-framework D-test statistics exceeding their actual level 0.05 critical values are displayed.

we can comfortably invoke Corollary 5.1 when $n \geq 70$. With $\alpha_0 \in \{0.3, 0.5\}$, we recommend employing actual critical values; these are available from the corresponding author.

Panels b through f of Figure 2 reveal several interesting phenomena; we emphasize that actual critical values have been employed to generate all power curves in these panels. For equal-probability mixtures (panels b and c), $d_n(0.5)$ and $E_n(0.5)$ appear competitive with each other and better than the other six tests; in fact, $d_n(0.1)$ and $E_n(0.1)$ are apparently not consistent for equal-probability mixtures. For "high-low"-probability mixtures (panels d through f),

$d_n(0.1)$ and $E_n(0.1)$ look to be competitive and better than the other six tests; now $d_n(0.5)$ and $E_n(0.5)$ are apparently not consistent. Thus, we favor taking $\alpha_0 := 0.5$ if a near-equal-probability mixture is suspected, and taking $\alpha_0 := 0.1$ if a high-low-probability mixture is suspected. For a user who does not know what kind of mixture to anticipate, we recommend employing $\widetilde{d}_n$ or $\widetilde{E}_n$ rather than $d_n(0.3)$ or $E_n(0.3)$. In this case, $\widetilde{d}_n$ performs slightly better than $\widetilde{E}_n$ if the mixture happens to be near-equal-probability, while $\widetilde{E}_n$ performs somewhat better than $\widetilde{d}_n$ if the mixture happens to be high-low-probability.

6.3. Application to the SLC data set

To illustrate the empirical Bayesian-framework D-test in a practical setting, we apply it to the SLC data set (Roeder (1994)). The data set contains red blood cell sodium-lithium countertransport (SLC) measurements for 190 individuals. As Roeder (1994) notes, such measurements are of interest since SLC is correlated with blood pressure and may be a cause of hypertension.

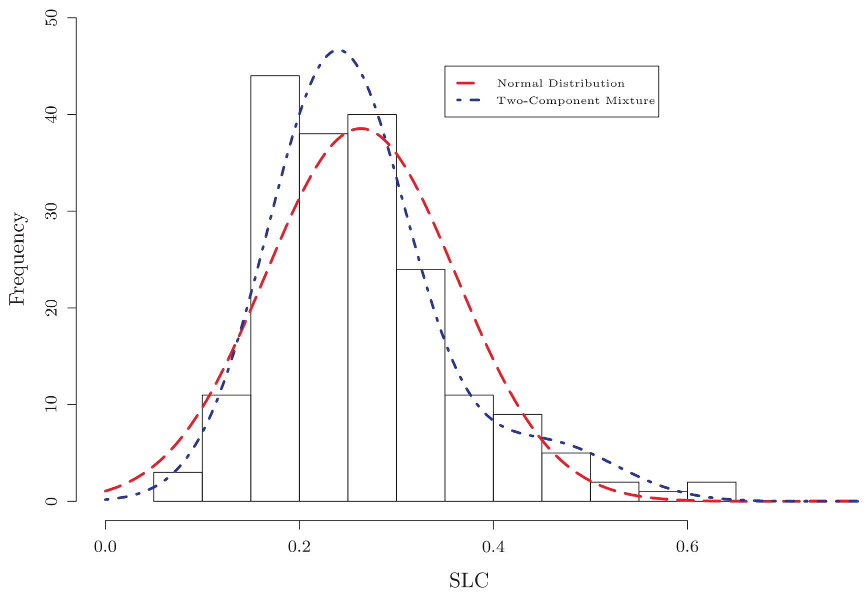
We find that $n\, d_n(0.5)\, \{256/35\}\pi^{1/2}\, \hat{\sigma}_0 \approx 0$ and $E_n(0.5) \approx 0$. On the other hand, $n\, d_n(0.1)\, \{32/5\}\pi^{1/2}\, \hat{\sigma}_0 = n\, \widetilde{d}_n\, \pi^{1/2}\, \hat{\sigma}_0 = 35.731$ and $E_n(0.1) - 2\log[0.2] = \widetilde{E}_n = 27.155$. The actual level 0.05 critical values for $n\, d_n(0.1)\, \{32/5\}\pi^{1/2}\, \hat{\sigma}_0$, $n\, \widetilde{d}_n\, \pi^{1/2}\, \hat{\sigma}_0$, $E_n(0.1) - 2\log[0.2]$, and $\widetilde{E}_n$ at $n = 190$ are 4.359, 9.643, 3.810, and 5.229, respectively. Both the EM test and the empirical Bayesian-framework D-test suggest heterogeneity in the form of a high-low-probability mixture.

Figure 3 shows the SLC data along with the fitted homogeneous model $N(0.263, 0.098^2)$ and the fitted heterogeneous model $0.887N(0.240, 0.072^2) + 0.113N(0.456, 0.072^2)$. The parameter estimates for the latter model are based on an initial value of 0.1 for α. The former model exhibits clear lack of fit but cannot be effectively supplanted by a near-equal-probability normal location mixture.

7. Discussion

The three asymptotic equivalences developed herein show that, under a null hypothesis of homogeneity, the D-test is closely related to three LR-type tests. Useful by-products of the equivalences found include asymptotic critical values for the Bayesian-framework D-test that can be obtained from a chi-square table, a theoretical assurance that the Bayesian-framework D-test is competitive with the MLR test under contiguous local alternatives, and asymptotic critical values for the empirical Bayesian-framework D-test that can be obtained from a chi-square table. The simulations in Section 6.1 suggest that the asymptotic critical values of Corollary 4.1 can be used with small n, while those in Section 6.2 indicate that the asymptotic critical values of Corollary 5.1 can be used with moderate n when α_0 is small.

Figure 3. Histogram of SLC data with fitted models.



The SLC data are shown with the fitted homogeneous model $N(0.263, 0.098^2)$ and the fitted heterogeneous model $0.887N(0.240, 0.072^2) + 0.113N(0.456, 0.072^2)$. Parameter estimates for the heterogeneous model were obtained as indicated in Section 5 with an initial value of 0.1 for α.

Yet, the worth of the D-test does not lie solely in its theoretical properties. Rather, the D-test has a computational advantage over LR-type tests in applications involving large or online data that cannot be processed on a single occasion, either because the data set is too massive or because it arrives in batches that are processed individually and then discarded.

More explicitly, suppose that a data set is partitioned into batches $\mathbf{X}_1, \ldots,$ $\mathbf{X}_B$. First, we can estimate the mixture parameters using $\mathbf{X}_1$. Then, for each $j \in \{2, \ldots, B\}$ we can update the parameter estimates using only $\mathbf{X}_j$ and the most recent of the previous estimates (i.e., we need not recall $\mathbf{X}_1$ through $\mathbf{X}_{j-1}$). Once the parameter estimates have been updated from $\mathbf{X}_B$, we can calculate a D-test statistic; all we need are the parameter estimates updated from $\mathbf{X}_B$. On the other hand, to calculate an LR-type test statistic would require us to recall $\mathbf{X}_1$ through $\mathbf{X}_{B-1}$. Such a scenario often occurs in Internet traffic applications, where data come in streams. Of course, an assumption for this updating strategy is that the estimates so acquired are asymptotically indistinguishable from the estimates that would have been calculated based on the full data set if it had been immediately available. Methodology for updating mixture parameter estimates without recalling previous data batches has already been proposed by Sun et al.

(2009). Their "partial" EM estimates have computational advantages over "full" (i.e., the familiar) EM estimates, even if recalling $\mathbf{X}_1$ through $\mathbf{X}_{B-1}$ is feasible; partial EM estimates are also more efficient when later batches of data contain more mixture components than earlier batches.

The present work begins a new line of research. We have recently studied finite mixtures of discrete distributions from an L^2 perspective (Charnigo and Sun (2008)). Moreover, Dai and Charnigo (2008) have examined the D-test and MLR test for contaminated density and regression models. They discovered that, for contaminated density models with a one-dimensional parameter, the limiting null distributions of $n\,d_n$ and M_n coincide up to a multiplicative constant. They also found that $n\,d_n$ and M_n share the limiting null distribution of χ^2_2 for a contaminated beta model, which is exciting because of that model's applicability to microarray data analysis.

Acknowledgement

The authors gratefully acknowledge grant support from the National Science Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Bickel, P. and Chernoff, H. (1993). Asymptotic distribution of the likelihood ratio statistic in a prototypical non regular problem. In *Statistics and Probability: A Raghu Raj Bahadur Festschrift* (Edited by J. K. Ghosh, S. K. Mitra, K. R. Parthasarathy and B. L. S. Prakasa Rao), 83-96. Wiley, New York.

Brown, L. D. (1986). *Fundamentals of Statistical Exponential Families.* Institute of Mathematical Statistics, Hayward, CA.

Charnigo, R. and Sun, J. (2004). Testing homogeneity in a mixture distribution via the L^2 distance between competing models. *J. Amer. Statist. Assoc.* **99**, 488-498.

Charnigo, R. and Sun, J. (2008). Testing homogeneity in discrete mixtures. *J. Statist. Plann. Inference* **138**, 1368-1388.

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

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

Chen, J. and Li, P. (2009). Hypothesis test for normal mixture models: the EM approach. *Ann. Statist.* **37**, 2523-2542.

Chernoff, H. and Lander, E. (1995). Asymptotic distribution of the likelihood ratio test that a mixture of two binomials is a single binomial. *J. Statist. Plann. Inference* **43**, 19-40.

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

Dai, H. and Charnigo, R. (2008). Inferences in contaminated regression and density models. *Sankhyā* **69**, 842-869.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present under the alternative. *Biometrika* **74**, 33-43.

Everitt, B. S. and Hand, D. J. (1981). *Finite Mixture Distributions.* Chapman and Hall, London.

Ghosh, J. K. and Sen, P. K. (1985). On the asymptotic performance of the log likelihood ratio statistics for the mixture model and related results. In *Proceedings of the Berkeley Conference in Honor of J. Neyman and J. Kiefer* **2** (Edieted by Lucien M. Le Cam and Richard A. Olshen), 789-806. Wadsworth.

Hartigan, J. (1985). A failure of likelihood asymptotics for normal mixtures. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (Edited by Lucien M. Le Cam and Richard A. Olshen) **2**, 807-810. Wadsworth.

Lemdani, M. and Pons, O. (1997). Likelihood ratio tests for genetic linkage. *Statist. Probab. Lett.* **33**, 15-22.

Li, P., Chen, J. and Marriott, P. (2009). Non-finite Fisher information and homogeneity: the EM approach. *Biometrika* **96**, 411-426.

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

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

McLachlan, G. (1987). On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *Appl. Statist.* **36**, 318-324.

Neyman, J. and Scott, E. L. (1966). On the use of $C(\alpha)$ optimal tests of composite hypotheses. *Bull. Inst. Statist.* **41**, 447-497.

Ott, J. (1999). *Analysis of Human Genetic Linkage*. 3rd Edition. The Johns Hopkins University Press, Baltimore, MD.

Roeder, K. (1994). A graphical technique for determining the number of components in a mixture of normals. *J. Amer. Statist. Assoc.* **89**, 487-495.

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

Sun, J., Morrison, H., Harding, P. and Woodroofe, M. (2009). Mixtures and bumps: Errors in measurement. *J. Amer. Statist. Assoc.*, submitted.

Sun, J., Liu, P. and Chen, J. (2009). Adaptive estimation for mixture distributions. Manuscript in preparation.

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

851 Patterson Office Tower, Department of Statistics, University of Kentucky, Lexington KY 40506-0027, U.S.A.

E-mail: RJCharn2@aol.com

Department of Statistics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland OH 44106-7054, U.S.A.

E-mail: jiayang@sun.STAT.cwru.edu

(Received February 2008; accepted February 2009)