
Volume 138 Issue 12 1 December 2008 ISSN 0378-3758
ELSEVIER
journal of
statistical planning
and inference
For Contents see page iii
Available online at
ScienceDirect
www.sciencedirect.com

Journal of Statistical Planning and Inference 138 (2008) 3774 – 3788

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Test for homogeneity in Hardy–Weinberg normal mixture model

Jiahua Chen[a,*], Yongsong Qin[b]

[a]Department of Statistics, University of British Columbia, Vancouver, BC, Canada
[b]Department of Mathematics, Guangxi Normal University, Guilin, Guangxi 541004, China

ARTICLE INFO

Article history:
Received 16 January 2006
Received in revised form
17 June 2007
Accepted 18 December 2007
Available online 28 March 2008

MSC:
primary 62F03
secondary 62F05

Keywords:
Gaussian process
LOD score
Mixture model
Modified likelihood ratio test
Quantitative trait locus
Statistical genetics

ABSTRACT

The phenotype of a quantitative trait locus (QTL) is often modeled by a finite mixture of normal distributions. If the QTL effect depends on the number of copies of a specific allele one carries, then the mixture model has three components. In this case, the mixing proportions have a binomial structure according to the Hardy–Weinberg equilibrium. In the search for QTL, a significance test of homogeneity against the Hardy–Weinberg normal mixture model alternative is an important first step. The LOD score method, a likelihood ratio test used in genetics, is a favored choice. However, there is not yet a general theory for the limiting distribution of the likelihood ratio statistic in the presence of unknown variance. This paper derives the limiting distribution of the likelihood ratio statistic, which can be described by the supremum of a quadratic form of a Gaussian process. Further, the result implies that the distribution of the modified likelihood ratio statistic is well approximated by a chi-squared distribution. Simulation results show that the approximation has satisfactory precision for the cases considered. We also give a real-data example.

1. Introduction

In genetics, the phenotype of a quantitative trait locus (QTL) is often modeled by a finite mixture of normal distributions. Assume that the QTL has two alleles *A* and *a*, then the corresponding genotypes *AA*, *Aa*, and *aa* induce three components in the mixture model. The mixing proportions have a binomial structure according to the Hardy–Weinberg (HW) equilibrium (Ott, 1999). In the absence of quantitative trait loci, the mixture model degenerates to a single-component normal model. Hence, in the search for quantitative trait loci, a significance test of homogeneity against the HW-normal mixture model alternative is an important first step (Jones and McLachlan, 1991; Roeder, 1994). The LOD score method, a likelihood ratio test used in genetics, is a favored choice.

The class of normal mixture models is part of the general class of finite mixture models. Recently, there has been rapid development in this area because of their great importance in a wide range of disciplines. Three recent books by Lindsay (1995), McLachlan and Peel (2003), and Titterington et al. (1985) among others are devoted to the theory and methods of finite mixture models. For genetic applications of finite mixture models, we recommend Schork et al. (1996) and the recent papers of Zhu and Zhang (2004) and Tadesse et al. (2005).

Due to the non-regularity of finite mixture models, many classical asymptotic results do not apply. Many researchers have contributed to the understanding of the asymptotic properties related to the analysis of finite mixture models. Hartigan (1985) pointed out the peculiar behavior of the likelihood ratio statistics. Placing a separation condition, Ghosh and Sen (1985) derived

* Corresponding author. Tel.: +1 604 822 1848; fax: +1 604 822 6960.
E-mail address: jhchen@stat.ubc.ca (J. Chen).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2007.12.013

the limiting distribution of the likelihood ratio statistic for an important class of finite mixture models. More recently, Chernoff and Lander (1995), Dacunha-Castelle and Gassiat (1999), Garel (2001, 2005), Chen and Chen (2001, 2003), Liu and Shao (2003), Charnigo and Sun (2004), and many others have worked on this class of problems and their results have led to much insight into finite mixture models.

A general normal mixture model in the presence of a structure parameter has density function

$$f(x; G, \sigma^2) = \sum_{j=1}^{K} \alpha_j \phi(x; \theta_j, \sigma^2), \tag{1.1}$$

where $\phi(x; \theta, \sigma^2)$ is the univariate density function of a normal random variable with mean θ and variance σ^2; and G is the mixing distribution on θ with K support points such that the probability of observing θ_j is α_j for $j = 1, 2, \ldots, K$.

Assume that a QTL has two alleles A and a and hence three genotypes AA, Aa, and aa. Their proportions are given by $p^2, 2pq$, and q^2 under the HW equilibrium, where $q = 1 - p$ and p is the population prevalence of A. The QTL determined by this gene is modeled by a finite mixture of normal distributions with $K = 3$ and mixing proportions $\alpha_1 = p^2, \alpha_2 = 2pq$, and $\alpha_3 = q^2$. An important genetic problem is to test the null hypothesis of $H_0 : pq = 0$ or $\theta_1 = \theta_2 = \theta_3$. Despite its illusive simplicity, general results on the limiting distribution of the likelihood ratio test are not available. The result by Dacunha-Castelle and Gassiat (1999) is fairly general yet not applicable to this model. Liu and Shao (2003) provided an insightful general principle. Chen and Chen (2003) produced a concrete solution to a related problem, but it lacks generality.

In this paper, we derive the limiting distribution of the likelihood ratio test statistic for the HW mixture. In addition to the fact that the HW mixture model has important genetic applications, it also presents a class of models where the discrete mixing distributions are parameterized. Our result can be combined with the techniques discussed in Davies (1977, 1987) for applications. We also apply the asymptotic result to the modified likelihood method (Chen et al., 2001, 2004). The resulting chi-squared approximation to the modified likelihood ratio test (MLRT) statistic is found to be very satisfactory. Our results can be easily generalized to other HW location-scale mixture models under some regularity conditions. Yet the process of verification is tedious and differs from one model to another. For clarity, we present results only for this most useful model.

The paper is organized as follows. In Section 2, we present some consistency results for the estimation of the common unknown variance. The asymptotic distribution of the likelihood ratio statistic is presented in Section 3.1. Its application to the MLRT is in Section 4. In Section 5, we report some simulation results, which confirm the accuracy of the chi-squared approximation. A real genetic data example is also included. Proofs are given in Section 6 with some technical details deferred to the Appendix.

2. Consistency results

Let $X_1, \ldots, X_n$ be a random sample of size n from the mixture population (1.1). We wish to test $H_0 : K = 1$ versus the HW alternative $H_1 : K = 3, \alpha_1 = p^2, \alpha_2 = 2pq$, and $\alpha_3 = q^2$ such that $0 < p < 1$ and $\theta_1 \neq \theta_2 \neq \theta_3$. For convenience, both α_j and p, q are used. The likelihood function is given by

$$\ell_n(G, \sigma^2) = \sum_{i=1}^{n} \log \left\{ \sum_{j=1}^{3} \alpha_j \phi(X_i; \theta_j, \sigma^2) \right\}. \tag{2.1}$$

Finite normal mixture models are identifiable in terms of mixing distributions. That is, two distinct mixing distributions result in two distinct finite mixture models. A mixing distribution, however, can be parameterized differently due to the exchangeability of its components. To avoid non-identifiability of the HW model in this sense, we adopt a restriction $0 \leqslant q \leqslant \frac{1}{2}$. We further assume that $|\theta_j| \leqslant M < \infty$ for $j = 1, 2, 3$, or the likelihood ratio statistic is stochastically unbounded when the sample size increases.

Let $\hat{G}_0, \hat{\sigma}_0^2$ be the maximum likelihood estimators of G and σ^2 under the null $K = 1$, and $\hat{G}_1, \hat{\sigma}_1^2$ be the corresponding maximum likelihood estimators under the HW alternative. The likelihood ratio statistic is

$$R_n = 2\{\ell_n(\hat{G}_1, \hat{\sigma}_1^2) - \ell_n(\hat{G}_0, \hat{\sigma}_0^2)\}. \tag{2.2}$$

Without loss of generality, we assume that the true null distribution is $N(\theta_0, \sigma_0^2)$ with $\theta_0 = 0$ and $\sigma_0^2 = 1$. Put $\hat{G}_1(\theta) = \sum_{j=1}^{3} \hat{\alpha}_j I(\hat{\theta}_j \leqslant \theta)$, and $\hat{G}_0(\theta) = I(\hat{\theta}_0 \leqslant \theta)$.

Even if the sample is from a null model, the maximum likelihood estimator $\hat{\sigma}_1^2$ under the HW alternative is still consistent. This result is established in the following two lemmas.

Lemma 1. *Under the null hypothesis* $N(0, 1)$, *there exist constants* $0 < \varepsilon < \Delta < \infty$ *such that*

$$\lim_{n \to \infty} P(\varepsilon \leqslant \hat{\sigma}_1^2 \leqslant \Delta) = 1.$$

Proof. It is seen that

$$\begin{aligned}\ell_n(G,\sigma) &= \sum_{i=1}^n \log\left\{\sum_{j=1}^3 \alpha_j \phi(X_i;\theta_j,\sigma^2)\right\}\\ &= -\frac{n}{2}\log\sigma^2 + \sum_{i=1}^n \log\left[\sum_{j=1}^3 \alpha_j \exp\left\{-\frac{(X_i-\theta_j)^2}{2\sigma^2}\right\}\right]\\ &\leqslant -\frac{n}{2}\left(\log\sigma^2 + \sigma^{-2}n^{-1}\sum_{i=1}^n H_i\right),\end{aligned}$$

where $H_i = H_i(\theta_1,\theta_2,\theta_3) = \min\{(X_i-\theta_j)^2, j=1,2,3\}$.

By the law of large numbers, and the easily verified tightness of the process, we have

$$n^{-1}\sum_{i=1}^n H_i \to E\{H_1(\theta_1,\theta_2,\theta_3)\}$$

in probability uniformly in the range of $|\theta_j| \leqslant M$. See Rubin (1956) for the result of uniform convergence.

Since $E\{H_1(\theta_1,\theta_2,\theta_3)\}>0$ for all $|\theta_j|\leqslant M$, and $E\{H_1(\theta_1,\theta_2,\theta_3)\}$ is continuous in the compact region $|\theta_j|\leqslant M$, we get $\inf_{|\theta_j|\leqslant M} E\{H_1(\theta_1,\theta_2,\theta_3)\}>0$. Thus, when n is large (but finite), either very small or very large σ^2 will lead $\ell_n(G,\sigma^2)$ to negative infinity. Hence, these σ^2 values cannot be maximum likelihood estimates. This leads to the conclusion. □

This result does not rely on K or the HW structure. It enables us to restrict σ^2 within a closed interval $[\varepsilon, \Delta]$. Once the relevant parameters θ_j and σ^2 are confined into a compact space, the Wald (1949) type proof can be carried out easily which leads to the following consistency result. The details are omitted.

Lemma 2. *Under the null hypothesis, and with the convention* $q \leqslant \frac{1}{2}$, *we have* $\hat\theta_1 = o_p(1)$, $\hat\alpha_2\hat\theta_2 = o_p(1)$, $\hat\alpha_3\hat\theta_3 = o_p(1)$, *and* $\hat\sigma_1^2 - 1 = o_p(1)$.

Consequently, either all support points of $\hat G_1$ stay close to the true value 0 or those that wander away are associated with diminishing probabilities. Thus, the asymptotic distribution of the likelihood ratio statistic is determined by parameters in a small range as indicated by this lemma.

3. Limiting distribution

We state the main result as follows. The proof will be given in Section 6.

Theorem 1. *Let* $\{\xi(s): -M \leqslant s \leqslant M\}$ *be a Gaussian process with* $E\{\xi(s)\}=0$, $Var\{\xi(s)\}=1$, *and correlation function*

$$\rho(s,t) = \frac{b(st)}{\sqrt{b(s^2)b(t^2)}},$$

where $b(s) = s^{-3}\{\exp(s) - 1 - s - s^2/2\}$. *Let* η *be a random variable with standard normal distribution such that*

$$Cov\{\eta,\xi(s)\} = \frac{s^2}{\sqrt{5!\,b(s^2)}}.$$

Then, under the null hypothesis, the likelihood ratio statistic R_n *given by* (2.2) *satisfies*

$$R_n \to \sup_{|s,t|\leqslant M}\left\{\frac{\xi^2(s)+\xi^2(t)-2\rho^2(s,t)\xi(s)\xi(t)}{1-\rho^2(s,t)} + \eta^2 I(s=t=0)\right\} \tag{3.1}$$

in distribution, as $n\to\infty$.

We can use the techniques of Davies (1977, 1987) to approximate the upper-tail probabilities in applications. Another approach is to use modified likelihood, as will be discussed. It will be seen that at $s=t=0$, the process reduces to $\xi^2(0)+\eta^2+\zeta^2$ which has a chi-squared distribution with three degrees of freedom, where ζ has a standard normal distribution and is independent of $(\xi(0),\eta)'$. Further, the spike of the process at $s=t=0$ implies that the null rejection rate is dominated by the outcomes in this area.

Table 1
The HW-normal mixture models used in the simulation

Model	A1	A2	A3	A4
μ_1	−1.0	−1.0	−2.0	−2.0
μ_2	0.0	0.0	0.0	0.0
μ_3	1.0	1.0	1.0	1.0
p	0.5	0.3	0.5	0.3

4. Application to the modified likelihood method

The modified likelihood approach (Chen, 1998; Chen et al., 2001, 2004) penalizes models with small mixing proportions. It effectively prevents the over-fittings caused by spurious mixing components with very low mixing proportions. As a result, all parts of the mixing distribution are consistently estimated. When used for a test of homogeneity, it automatically focuses on the region where $s = t = 0$ as discussed in the last section. Hence, the modified likelihood method is expected to be highly efficient. The ultimate advantage of this method is the simplicity of its limiting distribution for many commonly used models.

The modified likelihood for the current problem is defined as

$$\tilde{\ell}_n(G,\sigma^2) = \ell_n(G,\sigma^2) + C\sum_{j=1}^{3}\log\alpha_j$$

for some positive constant C. The usual choice is $C = 1$ which has been found satisfactory in many investigations (Chen, 1998; Charnigo and Sun, 2004; Zhu and Zhang, 2004) as well as in our simulation. In applications, we recommend a pilot simulation study being conducted to ensure the null type I error is close to the nominal value. If the simulated null rejection rate is more than 5.5% when the nominal level is 5%, C should be increased. Yet, our experience indicates that such a suitable C can be found quickly. Let $\tilde{G}_1$ and $\tilde{\sigma}_1^2$ be the maximum modified likelihood ratio estimators of G and σ^2. Let $\tilde{G}_0$ and $\tilde{\sigma}_0^2$ be the maximum modified likelihood ratio estimators of G and σ^2 under the null model. The modified likelihood ratio statistic is defined as

$$\tilde{R}_n = 2[\ell_n(\tilde{G}_1,\tilde{\sigma}_1^2) - \ell_n(\tilde{G}_0,\tilde{\sigma}_0^2)].$$

The most important property of the modified likelihood method for finite mixture models is that $\tilde{\theta}_j \to 0$ for all $j = 1,2,3$. This can be verified by using Theorem 1, similarly to the proof of Theorem 1 in Chen et al. (2001). We ignore the technical details here. Asymptotically, this claim reduces to

$$\tilde{R}_n \leqslant 2\sup_{|\theta_j|\leqslant\varepsilon_0, j=1,2,3}\{\ell_n(G,\sigma^2) - \ell_n(\hat{G}_0,\hat{\sigma}_0^2)\} + o_p(1)$$

for any $\varepsilon_0 > 0$. That is, the limiting distribution of $\tilde{R}_n$ is bounded by R_n with $M = \varepsilon_0$. In Section 6.2, a result summarized in the following lemma will be shown.

Lemma 3. *Under the same assumptions as Theorem* 1,

$$2\sup_{|\theta_j|\leqslant\varepsilon_0, j=1,2,3}\{\ell_n(G,\sigma^2) - \ell_n(\hat{G}_0,\hat{\sigma}_0^2)\} \leqslant \sum_{k=3}^{5}\frac{\{\sum_{i=1}^n Y_i^{(k-1)}\}^2}{\sum_{i=1}^n \{Y_i^{(k-1)}\}^2} + o_p(1) \xrightarrow{d} \chi_3^2,$$

where the $Y_i^{(k-1)}, k = 1,\ldots,5$ *are defined in Section* 6.

This lemma implies that $\tilde{R}_n$ has χ_3^2 as an asymptotic upper bound. Simulation shows that $\tilde{R}_n$ is well approximated by χ_3^2. One may be interested in the theoretical aspect of finding the exact limiting distribution of the modified likelihood ratio statistic. This is a technical problem for future research.

5. Simulation and an example

Simulation: We generated 2000 samples of size $n = 200$ and 500 from a standard normal distribution and compared the sample quantiles of $\tilde{R}_n$ with the quantiles of the χ_3^2 distribution. The modification constant C was set to 0.0 and 1.0. When $C = 0.0$, $\tilde{R}_n$ becomes the ordinary likelihood ratio statistic, whereas when $C = 1.0$, $\tilde{R}_n$ is a modified likelihood ratio statistic. We used the EM-algorithm with eight random initial values plus the true value of the parameters to compute the statistics of the MLRT. Fig. 1 contains four Q–Q plots corresponding to the two statistics with respect to the chi-squared distribution with three degrees of freedom for $n = 200$ and 500. The chi-squared approximation is liberal for both tests when $n = 200$. When $n = 500$, the chi-squared approximation improves substantially for both methods, particularly for the MLRT. This and other simulation experiences

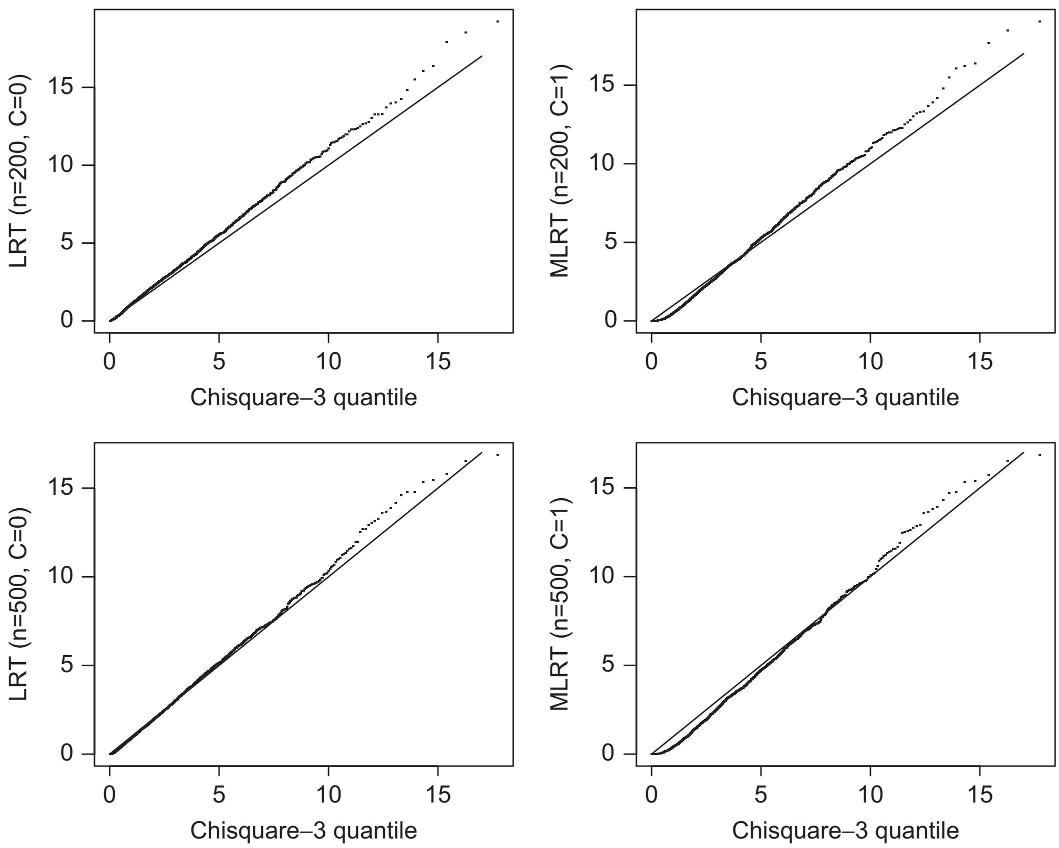

Fig. 1. Q–Q plots of the modified likelihood ratio statistics.

Table 2
Rejection rates for $K = 1$ versus the HW mixture model

	$C = 0$				$C = 1$			
Nominal	0.100	0.050	0.025	0.010	0.100	0.050	0.025	0.010
Null ($n = 200$)	0.134	0.074	0.041	0.019	0.119	0.069	0.039	0.017
Null ($n = 500$)	0.109	0.052	0.030	0.013	0.095	0.049	0.026	0.013
A1 ($n = 200$)	0.153	0.094	0.054	0.023	0.136	0.086	0.049	0.022
A2 ($n = 200$)	0.153	0.086	0.040	0.021	0.137	0.078	0.037	0.020
A3 ($n = 200$)	0.639	0.521	0.412	0.291	0.606	0.498	0.397	0.281
A4 ($n = 200$)	0.642	0.512	0.409	0.283	0.630	0.508	0.405	0.280

indicate that $C = 1$ is suitable in a wide range of applications. Only if controlling type I error is very important should one select a larger value suggested by additional simulations. See Chen and Kalbfleisch (2004) for further discussion.

Table 2 contains simulated null rejection rates, powers of both ordinary and MLRTs with the χ^2_3 approximation. The alternative models are specified in Table 1. The null rejection rates are very close to the nominal values. The agreement between the simulated quantiles and those of the χ^2_3 approximation is very good. On the other hand, the powers of the two methods are comparable among the models considered.

The MLRT against simple heterogeneity alternatives, as discussed in Chen and Chen (2003) and Chen and Kalbfleisch (2004), is also effective. In fact, some unreported simulation results show that it has a slightly higher power. However, its rejection does not indicate whether the departure is in the direction of the HW mixture, which is the goal of the current method.

Real-data example: We re-analyze the data presented in Roeder (1994) to illustrate the MLRT. The data set consists of 190 observations of red blood cell sodium–lithium countertransport (SLC). As discussed by Roeder, geneticists are interested in SLC because it is correlated with blood pressure and hence may be an important cause of hypertension. The condition is also easier to study than blood pressure because the latter is a complex trait that is highly variable and affected by environmental and perhaps many genetic factors.

One possibility is that SLC is determined by a simple mode of inheritance compatible with the action of a single gene with two alleles, A_1 and A_2, that occur with probabilities p and q. In this case, we might suppose that each observation is composed of the sum of a genetic component and a normally distributed measurement error. This would lead to a finite normal mixture model with common variance. A single dominance model for the gene yields a finite mixture model with $K=2$ components whereas an additive model yields a finite mixture model with $K=3$. A mixture model with more components is also possible if the mode of inheritance is complex. Roeder (1994) gives more background and references as well as some graphical methods of analysis.

With $C=1$ for a test against the HW-normal mixture alternative, we found the value of $\tilde{R}_n$ to be 36.5 with fitted normal means equaling 0.58, 0.37, 0.22 and corresponding $\hat{q}=0.144$. Thus, there is strong evidence for rejecting the homogeneous model in favor of the HW-normal mixture alternative model.

Since the outcome may simply reflect the need for a more complex model than a single normal, but not necessarily the HW-normal mixture model, we also fitted a 2-component normal mixture model. The outcome is summarized as $\tilde{R}_n=27.65$ with fitted normal means equaling 0.442, 0.236 and corresponding proportions 0.136, 0.864. This result indicates that the HW-mixture is more suitable than a simple 2-component mixture without a formal test.

If we fit a 3-component model without the HW structure or mixture models with more components, the log-likelihood is increased by less than 0.1 (over the maximum log-likelihood of HW). Hence, when regarded as a null model, the HW-mixture will not be rejected against any of these possible alternatives. Thus, there is no need to fit models more complex than the HW-normal mixture.

6. Proofs

We divide the proof of Theorem 1 into several major steps, and leave some details to the Appendix. Write

$$R_n=2\{\ell_n(\hat{G}_1,\hat{\sigma}_1^2)-\ell_n(G_0,1)\}-2\{\ell_n(\hat{G}_0,\hat{\sigma}_0^2)-\ell_n(G_0,1)\}=R_{1n}-R_{2n}, \tag{6.1}$$

where the last equality defines R_{1n} and R_{2n}. Under the null model, R_{2n} is an ordinary likelihood ratio statistic. Hence the following expansion is immediate and convenient for asymptotic derivation:

$$R_{2n}=\frac{(\sum_{i=1}^n X_i)^2}{\sum_{i=1}^n X_i^2}+\frac{\{\sum_{i=1}^n (X_i^2-1)\}^2}{\sum_{i=1}^n (X_i^2-1)^2}+o_p(1). \tag{6.2}$$

Consequently, finding the asymptotic distribution of R_n relies on a careful analysis of R_{1n}. Letting $R_{1n}(G,\sigma^2)=2\{\ell_n(G,\sigma^2)-\ell_n(G_0,1)\}$, we have

$$R_{1n}=\sup\{R_{1n}(G,\sigma^2): G(t)=p^2I(\theta_1\leqslant t)+2pqI(\theta_2\leqslant t)+q^2I(\theta_3\leqslant t), \sigma^2>0\}.$$

Let $\varepsilon_0>0$ be arbitrarily small. We partition the parameter space in the above expression into four regions in terms of the relative sizes of $\theta_j, j=1,2,3$ as follows:

(I) $|\theta_2|\leqslant\varepsilon_0, |\theta_3|\leqslant\varepsilon_0$;
(II) $|\theta_2|\leqslant\varepsilon_0, |\theta_3|>\varepsilon_0$;
(III) $|\theta_2|>\varepsilon_0, |\theta_3|\leqslant\varepsilon_0$;
(IV) $|\theta_2|>\varepsilon_0, |\theta_3|>\varepsilon_0$.

Because $\hat{\theta}_1=o_p(1)$ and $\hat{\sigma}^2-1=o_p(1)$ by Lemma 1, we may further ignore parameters not satisfying $|\theta_1|\leqslant\varepsilon_0$ and $|\sigma^2-1|\leqslant\varepsilon_0$.

Our asymptotic analysis of R_{1n} builds on region-specific quadratic expansions of $R_{1n}(G,\sigma^2)$. Let R_{1n}(I), R_{1n}(II), R_{1n}(III), and R_{1n}(IV) be its suprema over corresponding regions. It turns out that R_{1n}(IV) asymptotically dominates R_{1n}(II) and R_{1n}(III). When the parameters are confined to Region I, the model is ”almost” regular so that

$$R_{1n}(\mathrm{I})-R_{2n}\approx\chi_3^2.$$

Two free-range parameters θ_2 and θ_3 in Region IV makes $R_{1n}(\mathrm{IV})-R_{2n}$ behave like the supremum of a two-parameter stochastic process. Hence, we get an overall picture of the asymptotic results.

Before presenting the technical details, we introduce some simplifying notation. Write

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

with

$$\delta_i=\delta_i(G,\sigma^2)=\delta(X_i;G,\sigma^2)=\frac{f(X_i;G,\sigma^2)}{\phi(X_i;0,1)}-1.$$

Introduce

$$Y_i(\theta,\sigma^2)=\frac{\phi(X_i;\theta,\sigma^2)-\phi(X_i;0,\sigma^2)}{\theta\phi(X_i;0,1)}$$

and

$$U_i(\sigma^2)=\frac{\phi(X_i;0,\sigma^2)-\phi(X_i;0,1)}{(\sigma^2-1)\phi(X_i;0,1)}$$

with $Y_i(0,\sigma^2)$ and $U_i(1)$ being the corresponding continuity limits. Then a formal expansion of δ_i is as follows:

$$\delta_i(G,\sigma^2)=\sum_{j=1}^{3}\alpha_j\theta_j Y_i(\theta_j,\sigma^2)+(\sigma^2-1)U_i(\sigma^2). \tag{6.3}$$

The expansion of $Y_i(\theta,\sigma^2)$, $U_i(\sigma^2)$, and hence $R_{1n}(G,\sigma^2)$ is undertaken in the following subsections.

Let $Y^{(k)}=Y^{(k)}(0,1)$ be the kth partial derivative of $Y(\theta,\sigma^2)$ with respect to θ at $\theta=0$ and $\sigma^2=1$. Let $U^{(k)}=U^{(k)}(1)$ be the kth derivative of $U(\sigma^2)$ with respect to σ^2 at $\sigma^2=1$. Simple calculations reveal that

$$\begin{aligned}
&Y_i(0,1)=X_i,\\
&Y_i'(0,1)=U_i(1)=\tfrac{1}{2}(X_i^2-1),\\
&Y_i''(0,1)=\tfrac{1}{3}(X_i^3-3X_i),\\
&Y_i^{(3)}(0,1)=2U_i'(1)=\tfrac{1}{4}(X_i^4-6X_i^2+3),\\
&Y_i^{(4)}(0,1)=\tfrac{1}{5}(X_i^5-10X_i^3+15X_i).
\end{aligned}$$

Note that these have zero mean and are uncorrelated. We also denote $m_k=\sum_{j=1}^{3}\alpha_j\theta_j^k, k\geqslant 1$ as the kth moment of G.

We now outline the stepping stones of the proof further. By (6.3), δ_i is a linear combination of $Y_i(\theta,\sigma^2)$ and $U_i(\sigma^2)$. We first expand these at $\sigma^2=1$ so that after ignoring high-order terms ("$\approx$"),

$$\delta_i\approx\sum_{j=1}^{3}\alpha_j\theta_j\{Y_i(\theta_j,1)+(\sigma^2-1)A_i(\theta_j,1)\}+(\sigma^2-1)U_i(1)+(\sigma^2-1)^2U_i'(1)$$

with $A_i(\theta,1)$ to be introduced in (6.6).

Section 6.2 expands $Y_i(\theta,1)$ and $A_i(\theta,1)$ in the above expression further at $\theta=0$ for θ in Region I. Ultimately, we arrive at

$$R_{1n}(\mathrm{I})=\sup_{\sigma^2,G\in I}R_{1n}(\sigma^2,G)\approx\sum_{k=1}^{5}\frac{\{\sum_{i=1}^n Y_i^{(k-1)}\}^2}{\sum_{i=1}^n\{Y_i^{(k-1)}\}^2}.$$

Expansions over the other three regions in the other subsections are obtained in the same way. In particular,

$$R_{1n}(\mathrm{IV})\approx\frac{(\sum_{i=1}^n Y_i)^2}{\sum_{i=1}^n Y_i^2}+\frac{\{\sum_{i=1}^n Y_i'\}^2}{\sum_{i=1}^n Y_i'^2}+\frac{\{\sum_{i=1}^n V_i(\theta_2)\}^2}{\sum_{i=1}^n V_i^2(\theta_2)}+\frac{\{\sum_{i=1}^n Z_i(\theta_2,\theta_3)\}^2}{\sum_{i=1}^n Z_i^2(\theta_2,\theta_3)}.$$

The exact forms of $R_{1n}(\mathrm{II})$ and $R_{1n}(\mathrm{III})$ are not important because they are bounded by $R_{1n}(\mathrm{IV})$. The two expansions are combined in the final subsection to give the conclusion of Theorem 1.

6.1. Expanding δ_i with respect to σ^2

In this subsection, δ_i in (6.3) is expanded at $\sigma^2=1$ through $U_i(\sigma^2)$ and $Y_i(\theta,\sigma^2)$.

We first expand $U_i(\sigma^2)$ as

$$(\sigma^2-1)U_i(\sigma^2)=(\sigma^2-1)U_i(1)+(\sigma^2-1)^2U_i'(1)+\varepsilon_{1i}(\sigma^2). \tag{6.4}$$

Note that $(\sigma^2-1)^{-3}\varepsilon_{1i}(\sigma^2)$ has mean zero (including the case of $\sigma^2=1$). As a stochastic process, it is easily shown that $n^{-1/2}$ $(\sigma^2-1)^{-3}\sum_{i=1}^n\varepsilon_{1i}(\sigma^2)$ is tight in a neighborhood of $\sigma^2=1$. Thus, uniformly in a small neighborhood of $\sigma^2=1$,

$$\sum_{i=1}^n\varepsilon_{1i}(\sigma^2)=(\sigma^2-1)^3\mathrm{O_p}(\sqrt{n}). \tag{6.5}$$

Next, we expand $Y_i(\theta,\sigma^2)$. Put

$$A_i(\theta,\sigma^2)=(\sigma^2-1)^{-1}\{Y_i(\theta,\sigma^2)-Y_i(\theta,1)\} \tag{6.6}$$

with $A_i(\theta, 1)$ being the corresponding continuity limit. We have

$$Y_i(\theta, \sigma^2) = Y_i(\theta, 1) + (\sigma^2 - 1)A_i(\theta, 1) + \varepsilon_{2i}(\theta, \sigma^2),$$

where

$$\varepsilon_{2i}(\theta, \sigma^2) = (\sigma^2 - 1)^2 \left\{ \frac{A_i(\theta, \sigma^2) - A_i(\theta, 1)}{\sigma^2 - 1} \right\}.$$

Similarly to the case for $\varepsilon_{1i}(\sigma^2)$, it is seen that

$$\sum_{i=1}^{n} \sum_{j=1}^{3} \alpha_j \theta_j \varepsilon_{2i}(\theta_j, \sigma^2) = (|m_1| + m_2)(\sigma^2 - 1)^2 \mathrm{O_p}(\sqrt{n}). \tag{6.7}$$

In summary, we have

$$\delta_i = \sum_{j=1}^{3} \alpha_j \theta_j \{Y_i(\theta_j, 1) + (\sigma^2 - 1)A_i(\theta_j, 1)\} + (\sigma^2 - 1)U_i(1) + (\sigma^2 - 1)^2 U_i'(1) + \{\varepsilon_{1i}(\sigma^2) + \varepsilon_{2i}(\theta, \sigma^2)\} \tag{6.8}$$

with the order of $\sum_i \{\varepsilon_{1i}(\sigma^2) + \varepsilon_{2i}(\theta, \sigma^2)\}$ assessed by (6.5) and (6.7). Expanding the terms in the above expression at $\theta = 0$ in the four regions, and therefore obtaining corresponding expansions of $R_{1n}(G, \sigma^2)$, are the tasks of the next three subsections.

6.2. Expansion of $R_{1n}(G, \sigma^2)$ in Region I

In this subsection, $Y_i(\theta, 1)$, $A_i(\theta, 1)$, and $U_i(\theta)$ in (6.8) are expanded at $\theta = 0$, and hence that of R_{1n} over Region I.

When θ is in a small neighborhood of $\theta = 0$, we write

$$Y_i(\theta, 1) - Y_i(0, 1) = \sum_{k=1}^{4} \frac{\theta^k}{k!} Y_i^{(k)}(0, 1) + \varepsilon_{3i}(\theta). \tag{6.9}$$

Note that $\theta^{-5}\varepsilon_{3i}(\theta)$ has zero mean and finite variance. As a stochastic process in θ, $n^{-1/2}\theta^{-5}\sum_{i=1}^{n}\varepsilon_{3i}(\theta)$ is tight in $|\theta| \leqslant M$. Thus $\sup_{|\theta| \leqslant M} n^{-1/2}\theta^{-5}\sum_{i=1}^{n}\varepsilon_{3i}(\theta) = \mathrm{O_p}(1)$. Hence,

$$\sum_{j=1}^{3} \alpha_j \theta_j \sum_{i=1}^{n} \varepsilon_{3i}(\theta_j) = m_6 \mathrm{O_p}(\sqrt{n}) \tag{6.10}$$

uniformly over $|\theta_j| \leqslant M$.

We also need to expand $A_i(\theta, 1)$ defined in (6.6). Direct calculation shows that

$$A_i(\theta, 1) = \frac{\{(X_i - \theta)^2 - 1\}\phi(X_i; \theta, 1) - (X_i^2 - 1)\phi(X_i; 0, 1)}{2\theta\phi(X_i; 0, 1)}.$$

By expanding $\phi(X_i; \theta, 1)/\phi(X_i; 0, 1) = \exp\{\theta X_i - \theta^2/2\}$, we easily find

$$A_i(0, 1) = \tfrac{1}{2}(X_i^3 - 3X_i) = \tfrac{3}{2}Y_i'',$$

and

$$\left.\frac{\mathrm{d}A_i(\theta, 1)}{\mathrm{d}\theta}\right|_{\theta=0} = A_i'(0, 1) = \frac{1}{4}(X_i^4 - 6X_i^2 + 3) = Y_i^{(3)}.$$

Hence, we may write

$$A_i(\theta, 1) = A_i(0, 1) + \theta Y_i^{(3)} + \tfrac{1}{2}\theta^2 A_i''(0, 1) + \varepsilon_{4i}(\theta). \tag{6.11}$$

Similarly to (6.10), we note that

$$\sum_{j=1}^{3} \alpha_j \theta_j \sum_{i=1}^{n} \varepsilon_{4i}(\theta_j) = m_4 \mathrm{O_p}(\sqrt{n}). \tag{6.12}$$

Combining these expansions, we get

$$\delta_i = m_1 Y_i + (\sigma^2 - 1 + m_2)Y_i' + \tfrac{1}{2}m_3 Y_i'' + \tfrac{1}{6}\{3(\sigma^2-1)^2 + m_4 + 6(\sigma^2-1)m_2\}Y_i^{(3)} + \tfrac{1}{24}m_5 Y_i^{(4)} + \varepsilon_{5i} \tag{6.13}$$

with ε_{5i} having the order assessment

$$\sum_{i=1}^n \varepsilon_{5i} = \{m_6 + (|m_1| + |m_3| + m_4)|\sigma^2 - 1| + (|m_1| + m_2)(\sigma^2-1)^2 + |\sigma^2 - 1|^3\}O_p(\sqrt{n}).$$

The leading terms in (6.13) are from a number of places: $\sum_{j=0}^4 (1/j!)m_{j+1}Y_i^{(j)}$ is from (6.9), $(\sigma^2-1)Y_i' + \frac{1}{2}(\sigma^2-1)^2 Y_i^{(3)}$ is from (6.4), $(\sigma^2-1)m_2$ is from $\theta Y_i^{(3)}$ in (6.11).

The terms in ε_{5i} are from the following places: $(|m_1| + m_2)(\sigma^2-1)^2$ is from (6.7), $(\sigma^2-1)^3$ is from (6.5), m_6 is from (6.10), $(|m_1| + |m_3|)(\sigma^2-1)$ is from $A_i(0,1) + \frac{1}{2}\theta^2 A_i''(0,1)$ in (6.11), and $(\sigma^2-1)m_4$ is from (6.12). Note that $m_1(\sigma^2-1)^2$ has higher order than $m_1(\sigma^2-1)$ but is included for completeness.

For notational simplicity, write $s_1 = m_1, s_2 = \sigma^2 - 1 + m_2, s_3 = m_3/2, s_4 = (m_4 - 3m_2^2)/6, s_5 = m_5/24$. Hence, by keeping only linear terms in $s_j, j = 1, \ldots, 5$, we may write

$$\delta_i = \sum_{k=1}^5 s_k Y_i^{(k-1)}(0,1) + \tilde{\varepsilon}_{5i}, \tag{6.14}$$

where $Y_i^{(0)}(0,1) = Y_i(0,1)$. The spillover of a quadratic term in s_2 is accommodated by letting $\tilde{\varepsilon}_{5i} = \varepsilon_{5i} - \frac{1}{2}s_2^2 Y_i^{(3)}$. Note that according to Lemma 2, we need consider only small values of s_j^2.

The order of the remainder can be presented in terms of s_j^2. The following lemma will be proved in the Appendix.

Lemma 4. (i) *When* $\theta_j \to 0$ *for* $j = 1, 2, 3$, $m_6^2 = o(\sum_{j=1}^5 s_j^2)$.

(ii) *When in addition* $\sigma^2 \to 1$,

$$\{(|m_1| + |m_3| + m_4)|\sigma^2 - 1| + m_2(\sigma^2-1)^2 + |\sigma^2-1|^3\}O(\sqrt{n}) = o(1) + \left(\sum_{j=1}^5 s_j^2\right) o(n).$$

Region I contains parameter values satisfying the conditions in Lemma 4. Hence

$$\sum_{i=1}^n \tilde{\varepsilon}_{5i} = o_p(1) + \left(\sum_{j=1}^5 s_j^2\right) o_p(n).$$

Using the inequality $\log(1+x) \leqslant x - \frac{1}{2}x^2 + \frac{1}{3}x^3$, and similarly to Chen and Chen (2003), we have

$$\begin{aligned} R_{1n}(G, \sigma^2) &= 2\sum_{i=1}^n \log(1+\delta_i) \leqslant 2\sum_{i=1}^n \delta_i - \sum_{i=1}^n \delta_i^2 + \frac{2}{3}\sum_{i=1}^n \delta_i^3 \\ &= \sum_{k=1}^5 \left\{ s_k \sum_{i=1}^n Y_i^{(k-1)} \right\} - \left[\sum_{k=1}^5 s_k^2 \sum_{i=1}^n \{Y_i^{(k-1)}\}^2 \right] \{1 + o_p(1)\} + o_p(1). \end{aligned}$$

Thus, over Region I,

$$R_{1n}(\mathrm{I}) = \sup_{\sigma^2, G \in I} R_{1n}(\sigma^2, G) \leqslant \sum_{k=1}^5 \frac{\{\sum_{i=1}^n Y_i^{(k-1)}\}^2}{\sum_{i=1}^n \{Y_i^{(k-1)}\}^2} + o_p(1).$$

On the other hand, this upper bound is attained when

$$s_k = \frac{\sum_{i=1}^n Y_i^{(k-1)}}{\sum_{i=1}^n \{Y_i^{(k-1)}\}^2}$$

for $k = 1, \ldots, 5$ with appropriate choice of $p, \theta_1, \theta_2, \theta_3$, and σ^2.

That is, focusing on a very small region where all $\theta_1, \theta_2, \theta_3$ are close to zero, the likelihood ratio statistic can be approximated by the quadratic form shown. The conclusion of Lemma 3 also follows. For regular models, this kind of expansion provides the justification of the usual chi-squared limiting distribution. In the context of the finite mixture model, the likelihood may attain its maximum when θ_2 and θ_3 are not close to zero. That is why the proof does not stop here.

6.3. *Expansion of $R_{1n}(G, \sigma^2)$ in Regions II and III*

In this subsection, we expand $Y_i(\theta, 1)$, $A_i(\theta, 1)$, and $U_i(\theta)$ in (6.8) at $\theta = 0$ to obtain expansions of R_{1n} over Regions II and III. In Region II, θ_3 is outside of the small neighborhood of $\theta_0 = 0$ while θ_1 and θ_2 are close to 0. The expansion of δ_i given by (6.14) is still valid in principle. However, the order assessment of the residual does not work because θ_3 can be large. For example, m_6 could be much larger than m_2 (rather than being a higher order quantity).

To accommodate this fact, we use a shortened expansion

$$Y_i(\theta, 1) - Y_i(0, 1) = \sum_{k=1}^{2} \frac{\theta^k}{k!} Y_i^{(k)}(0, 1) + \varepsilon_{3i}(\theta), \tag{6.15}$$

where $\varepsilon_{3i}(\theta)$ satisfies

$$\sum_{j=1}^{2} \alpha_j \theta_j \sum_{i=1}^{n} \varepsilon_{3i}(\theta_j) = (\alpha_1 \theta_1^4 + \alpha_2 \theta_2^4) O_p(\sqrt{n}).$$

Note that the $\varepsilon_{3i}(\theta)$ here is different from the $\varepsilon_{3i}(\theta)$ in the previous subsection.

Expanding $Y_i(\theta_3, 1)$ at $\theta_3 = 0$ is not useful because in this case θ_3^4 is not a higher order infinitely small term. Instead, we define

$$W_i(\theta_3) = \theta_3^{-3} \left\{ Y_i(\theta_3, 1) - Y_i(0, 1) - \theta_3 Y_i'(0, 1) - \frac{\theta_3^2}{2} Y_i''(0, 1) \right\},$$

so that

$$\delta_i = m_1 Y_i(0, 1) + (\sigma^2 - 1 + m_2) Y'(0, 1) + \tfrac{1}{2} m_3 Y''(0, 1) + \alpha_3 \theta_3^4 W_i(\theta_3) + \varepsilon_{4i}$$

with

$$\sum_{i=1}^{n} \varepsilon_{4i} = \{(\alpha_1 \theta_1^4 + \alpha_2 \theta_2^4) + (|m_1| + m_2 + |m_3| + m_4)|\sigma^2 - 1| + |\sigma^2 - 1|^2\} O_p(\sqrt{n}).$$

The sources of all the leading terms of $\delta_i(G, \sigma^2)$ can be easily identified as before. The contributions of $A_i(\theta, 1)$ are all included in the residuals in this case.

Similarly to, but different from, the previous subsection we define $s_1 = m_1$, $s_2 = \sigma^2 - 1 + m_2$, $s_3 = m_3$, and $s_4 = \alpha_3 \theta_3^4$. The order assessment is given by a lemma here with its proof in the Appendix.

Lemma 5. *When $\theta_1, \theta_2 \to 0$, $p \to 1$, and $\sigma^2 \to 1$,*

$$(\alpha_1 \theta_1^4 + \alpha_2 \theta_2^4) O(\sqrt{n}) = o(1) + \left(\sum_{j=1}^{4} s_j^2 \right) o(n)$$

and

$$\{(|m_1| + m_2 + |m_3|)|\sigma^2 - 1| + |\sigma^2 - 1|^2\} O(\sqrt{n}) = o(1) + \left(\sum_{j=1}^{4} s_j^2 \right) o(n).$$

At last, we have arrived at

$$\sum_{i=1}^{n} \delta_i = s_1 \sum_{i=1}^{n} Y_i + s_2 \sum_{i=1}^{n} Y_i' + \frac{1}{2} s_3 \sum_{i=1}^{n} Y_i'' + s_4 \sum_{i=1}^{n} W_i(\theta_3) + o_p(1) + \left(\sum_{j=1}^{4} s_j^2 \right) o_p(n).$$

Hence, over Region II,

$$\begin{aligned}R_{1n} &\leqslant \sum_{i=1}^{n}\left(2\delta_i - \delta_i^2 + \frac{2}{3}\delta_i^3\right)\\ &= 2\left\{s_1\sum_{i=1}^{n} Y_i + s_2\sum_{i=1}^{n} Y_i' + \frac{1}{2}s_3\sum_{i=1}^{n} Y_i'' + s_4\sum_{i=1}^{n} W_i(\theta_3)\right\}\\ &\quad - \left[s_1^2\sum_{i=1}^{n} Y_i^2 + s_2^2\sum_{i=1}^{n}(Y_i')^2 + \frac{1}{4}s_3^2\sum_{i=1}^{n}(Y_i'')^2 + s_4^2\sum_{i=1}^{n}\{W_i(\theta_3)\}^2\right]\{1+o_p(1)\} + o_p(1).\end{aligned}$$

The right-hand side would attain a maximum if we could make

$$s_1 = \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n Y_i^2},\quad s_2 = \frac{\sum_{i=1}^n Y_i'}{\sum_{i=1}^n (Y_i')^2},\quad s_3 = \frac{2\sum_{i=1}^n Y_i''}{\sum_{i=1}^n (Y_i'')^2},\quad s_4 = \frac{\sum_{i=1}^n W_i(\theta_3)}{\sum_{i=1}^n W_i^2(\theta_3)}.$$

Therefore,

$$R_{1n}(\mathrm{II}) \leqslant \frac{(\sum_{i=1}^n Y_i)^2}{\sum_{i=1}^n Y_i^2} + \frac{(\sum_{i=1}^n Y_i')^2}{\sum_{i=1}^n (Y_i')^2} + \frac{(\sum_{i=1}^n Y_i'')^2}{\sum_{i=1}^n (Y_i'')^2} + \frac{\{\sum_{i=1}^n W_i(\theta_3)\}^2}{\sum_{i=1}^n W_i^2(\theta_3)} + o_p(1).$$

Since $s_4 \geqslant 0$, the above upper bound may not be attained. However, it will be seen that $R_{1n}(\mathrm{IV})$ dominates $R_{1n}(\mathrm{II})$. Hence, an exact expansion is not needed. Region III is a mirror version of Region II. Thus, we turn our attention to Region IV immediately.

6.4. Expansion of $R_{1n}(G, \sigma^2)$ in Region IV

In this subsection, we expand $Y_i(\theta, 1)$, $A_i(\theta, 1)$, and $U_i(\theta)$ in (6.8) at $\theta = 0$ to obtain an expansion of R_{1n} over Region IV.

When both θ_2 and θ_3 are outside of the neighborhood of 0, only expanding $Y_i(\theta_1, 1)$ at $\theta_1 = 0$ is meaningful. We write

$$Y_i(\theta_1, 1) = Y_i(0, 1) + \theta_1 Y_i'(0, 1) + \varepsilon_{3i}(\theta_1)$$

where $\sum_{i=1}^n \varepsilon_{3i}(\theta_1) = \theta_1^2 O_p(\sqrt{n})$. Define $V_i(\theta) = \theta^{-2}\{Y_i(\theta, 1) - Y_i(0, 1) - Y_i'(0, 1)\theta\}$, which has one fewer term than $W_i(\theta)$. We have

$$\delta_i(G, \sigma^2) = m_1 Y_i + (\sigma^2 - 1 + m_2)Y_i' + \alpha_2\theta_2^3 V_i(\theta_2) + \alpha_3\theta_3^3 V_i(\theta_3) + \varepsilon_{4i}. \tag{6.16}$$

The order of ε_{4i} will be assessed using two cases as follows.

Case I: $|\theta_2 - \theta_3| \geqslant \varepsilon_0$. Let $\gamma(\theta_2, \theta_3) = Cov\{V_i(\theta_2), V_i(\theta_3)\}/Var\{V_i(\theta_2)\}$ and define $Z_i(\theta_2, \theta_3) = V_i(\theta_3) - \gamma(\theta_2, \theta_3)V_i(\theta_2)$. Due to identifiability, $Var(Z_i) > 0$ uniformly over $|\theta_2 - \theta_3| \geqslant \varepsilon_0$. It is also seen that $V_i(\theta_2)$ and $Z_i(\theta_2, \theta_3)$ are uncorrelated for all θ_2 and θ_3.

Let $s_1 = m_1$, $s_2 = \sigma^2 - 1 + m_2$, $s_3 = \alpha_2\theta_2^3 + \alpha_3\gamma(\theta_2, \theta_3)\theta_3^3$, and $s_4 = \alpha_3\theta_3^3$. We have

$$\delta_i(G, \sigma^2) = s_1 Y_i + s_2 Y_i' + s_3 V_i(\theta_2) + s_4 Z_i(\theta_2, \theta_3) + \varepsilon_{5i}$$

with

$$\sum_{i=1}^{n}\varepsilon_{5i} = \{\alpha_1|\theta_1^3| + (|m_1| + m_2 + \alpha_1|\theta_1^3| + \alpha_2|\theta_2^3| + \alpha_3|\theta_3^3|)|\sigma^2 - 1| + (\sigma^2 - 1)^2\}O_p(\sqrt{n}).$$

We further conclude that

$$\sum_{i=1}^{n}\varepsilon_{5i} = o_p(1) + \left(\sum_{j=1}^{4} s_j^2\right) o_p(n)$$

based on the following lemma to be proved in the Appendix.

Lemma 6. *When* $\theta_1 \to 0$, $p \to 1$, *and* $\sigma^2 \to 1$,

$$\{\alpha_1|\theta_1^3| + (|m_1| + m_2 + \alpha_2|\theta_2^3| + \alpha_3|\theta_3^3|)|\sigma^2 - 1| + (\sigma^2 - 1)^2\}O(\sqrt{n}) = o(1) + \left(\sum_{j=1}^{4} s_j^2\right) o(n).$$

When $|\theta_2-\theta_3|\geqslant\varepsilon_0$, $Var(s_1Y_i+s_2Y_i'+s_3V_i(\theta_2)+s_4Z_i(\theta_2,\theta_3))$ is a positive definite quadratic form of s_1, s_2, s_3, and s_4 uniformly. With the positive definiteness, we find

$$R_{1n}\leqslant 2\sum_{i=1}^n\delta_i-\sum_{i=1}^n\delta_i^2+\frac{2}{3}\sum_{i=1}^n\delta_i^3$$
$$=2\left\{s_1\sum_{i=1}^nY_i+s_2\sum_{i=1}^nY_i'+s_3\sum_{i=1}^nV_i(\theta_2)+s_4\sum_{i=1}^nZ_i(\theta_2,\theta_3)\right\}$$
$$-\left\{s_1^2\sum_{i=1}^nY_i^2+s_2^2\sum_{i=1}^n(Y_i')^2+s_3^2\sum_{i=1}^nV_i^2(\theta_2)+s_4^2\sum_{i=1}^nZ_i^2(\theta_2,\theta_3)\right\}\{1+o_p(1)\}+o_p(1).$$

For each given θ_2 and θ_3 in the region, it attains a maximum when

$$s_1=\frac{\sum_{i=1}^nX_i}{\sum_{i=1}^nX_i^2},\quad s_2=\frac{\sum_{i=1}^nU_i(1)}{\sum_{i=1}^nU_i^2(1)},\quad s_3=\frac{\sum_{i=1}^nV_i(\theta_2)}{\sum_{i=1}^nV_i^2(\theta_2)},\quad s_4=\frac{\sum_{i=1}^nZ_i(\theta_2,\theta_3)}{\sum_{i=1}^nZ_i^2(\theta_2,\theta_3)}.$$

Therefore,

$$R_{1n}=\frac{(\sum_{i=1}^nX_i)^2}{\sum_{i=1}^nX_i^2}+\frac{\{\sum_{i=1}^n(X_i^2-1)\}^2}{\sum_{i=1}^n(X_i^2-1)^2}+\frac{\{\sum_{i=1}^nV_i(\theta_2)\}^2}{\sum_{i=1}^nV_i^2(\theta_2)}+\frac{\{\sum_{i=1}^nZ_i(\theta_2,\theta_3)\}^2}{\sum_{i=1}^nZ_i^2(\theta_2,\theta_3)}+o_p(1).$$

Case II: $|\theta_2-\theta_3|\leqslant\varepsilon_0$. When θ_2 and θ_3 are close, $V_i(\theta_2)$ and $V_i(\theta_3)$ are almost equal. Hence, we write

$$\delta_i=m_1Y_i+(\sigma^2-1+m_2)Y_i'+(\alpha_2\theta_2^3+\alpha_3\theta_3^3)V_i(\theta_2)+\alpha_3\theta_3^3\{V_i(\theta_3)-V_i(\theta_2)\}+\varepsilon_{6i}. \tag{6.17}$$

We now show that $\alpha_3\theta_3^3\{V_i(\theta_3)-V_i(\theta_2)\}$ is negligible in the final expansion.

As before, let $s_1=m_1$, $s_2=\sigma^2-1+m_2$, and $s_3=\alpha_2\theta_2^3+\alpha_3\theta_3^3$. Recall that θ_2 and θ_3 are close to each other, and both are larger than ε_0 in absolute value. Hence, $|s_3|=\alpha_2|\theta_2^3|+\alpha_3|\theta_3^3|$. Similarly, since $|\theta_3|\leqslant M$ and $|\theta_3-\theta_2|\leqslant\varepsilon_0$,

$$\sum_{i=1}^n\alpha_3\theta_3^3\{V_i(\theta_3)-V_i(\theta_2)\}=\alpha_3\theta_3^3|\theta_3-\theta_2|O_p(\sqrt{n})=s_3o_p(\sqrt{n})=o_p(1)+s_3^2o_p(n).$$

Then, we have

$$\delta_i=s_1Y_i+s_2Y_i'+s_3V_i(\theta_2)+\varepsilon_{7i}$$

such that

$$\sum_{i=1}^n\varepsilon_{7i}=o_p(1)+\left(\sum_{k=1}^3s_k^2\right)o_p(n).$$

It is therefore concluded that over this region,

$$R_{1n}\leqslant 2\sum_{i=1}^n\delta_i-\sum_{i=1}^n\delta_i^2+\frac{2}{3}\sum_{i=1}^n\delta_i^3$$
$$=2\left\{s_1\sum_{i=1}^nY_i+s_2\sum_{i=1}^nY_i'+s_3\sum_{i=1}^nV_i(\theta_2)\right\}$$
$$-\left\{s_1^2\sum_{i=1}^nY_i^2+s_2^2\sum_{i=1}^n(Y_i')^2+s_3^2\sum_{i=1}^nV_i^2(\theta_2)\right\}\{1+o_p(1)\}+o_p(1).$$

Clearly, R_{1n} here is bounded by the supremum in Case I. Therefore, this case can be simply ignored in the final expression of R_{1n}(IV).

6.5. Combining expansions

Recall that $R_n = \max\{R_{1n}(\mathrm{I}), R_{1n}(\mathrm{II}), R_{1n}(\mathrm{III}), R_{1n}(\mathrm{IV})\} - R_{2n}$. By examining the expansions of $R_{1n}(\mathrm{IV})$ and $R_{1n}(\mathrm{II})$, we note that $R_{1n}(\mathrm{IV}) \geqslant R_{1n}(\mathrm{II}) + o_p(1)$. This can be seen as follows. As $\theta_2 \to 0$, we have $V_i(\theta_2) \to Y_i''$, and $Z_i(\theta_2, \theta_3) \to \theta_3 W_i(\theta_3)$. Hence,

$$\sup_{\theta_2,\theta_3 \in \mathrm{IV}} \frac{\{\sum_{i=1}^n Z_i(\theta_2,\theta_3)\}^2}{\sum_{i=1}^n Z_i^2(\theta_2,\theta_3)} \geqslant \sup_{\theta_3} \frac{\{\sum_{i=1}^n W_i(\theta_3)\}^2}{\sum_{i=1}^n W_i^2(\theta_3)}.$$

For the same reason, $R_{1n}(\mathrm{IV}) \geqslant R_{1n}(\mathrm{III}) + o_p(1)$. Consequently, Regions II and III do not contribute to the maximum. We have

$$R_n = \max\{R_{1n}(\mathrm{I}), R_{1n}(\mathrm{IV})\} - R_{2n} + o_p(1).$$

We now provide a simple description of the limiting distribution. It can be easily (but somewhat tediously) verified that $\{\sum_{i=1}^n V_i(\theta_2)\}/\{\sum_{i=1}^n V_i^2(\theta_2)\}^{1/2}$ converges to the stochastic process $\xi(s)$ as given in Theorem 1.

Since $W_i(\theta_2, \theta_3) = V_i(\theta_3) - \gamma(\theta_2, \theta_3) V_i(\theta_2)$, we have

$$\begin{aligned} R_{1n}(\mathrm{IV}) - R_{2n} &= \sup_{\theta_2,\theta_3 \in \mathrm{IV}} \left\{ \frac{\{\sum_{i=1}^n V_i(\theta_2)\}^2}{\sum_{i=1}^n V_i^2(\theta_2)} + \frac{\{\sum_{i=1}^n V_i(\theta_3) - \gamma(\theta_2,\theta_3)V_i(\theta_2)\}^2}{\sum_{i=1}^n \{V_i(\theta_3) - \gamma(\theta_2,\theta_3)V_i(\theta_2)\}^2} \right\} \\ &\to \sup_{\theta_2,\theta_3 \in \mathrm{IV}} \left\{ \frac{\xi^2(\theta_2) + \xi^2(\theta_3) - 2\rho(\theta_2,\theta_3)\xi(\theta_2)\xi(\theta_3)}{1 - \rho^2(\theta_2,\theta_3)} \right\} \end{aligned} \tag{6.18}$$

in distribution, where $\rho(\theta_2, \theta_3) = Cov\{\xi(\theta_2), \xi(\theta_3)\}$.

We now link $R_{1n}(\mathrm{I}) - R_{2n}$ into this picture. Recall that the first two terms in the expansion of $R_{1n}(\mathrm{I}) - R_{2n}$ are standardized $\sum_{i=1}^n Y_i''$ and $\sum_{i=1}^n Y_i^{(3)}$. At the same time, as θ_2 and θ_3 go to 0, we have $V_i(\theta_2) \to Y_i''$ and $V(\theta_3) - r(\theta_2, \theta_3)V(\theta_2) \to Y^{(3)}$. Hence,

$$\lim_{s,t,\to 0} \left\{ \frac{\xi^2(\theta_2) + \xi^2(\theta_3) - 2\rho(\theta_2,\theta_3)\xi(\theta_2)\xi(\theta_3)}{1 - \rho^2(\theta_2,\theta_3)} \right\} = \xi^2(0) + \zeta^2,$$

where ζ is the weak limit of the standardized $\sum_{i=1}^n Y_i^{(3)}$.

Let η be the limit of the standardized $\sum_{i=1}^n Y_i^{(4)}$. It is seen that $\xi(0)$, ζ, and η are independent standard normal random variables. Further, for $k = 2, 3$, and 4,

$$Cov\{Y^{(k)}, \xi(s)\} = \frac{s^{k-2}}{\sqrt{(k+1)!\, b(s^2)}}.$$

Therefore, we have the expression

$$R_n \to \sup_{|s,t| \leqslant M} \left\{ \frac{\xi^2(s) + \xi^2(t) - 2\rho^2(s,t)\xi(s)\xi(t)}{1 - \rho^2(s,t)} + \eta^2 I(s = t = 0) \right\}$$

in distribution. When $s = t = 0$, the process reduces to $\xi^2(0) + \eta^2 + \zeta^2$ which has a chi-squared distribution with three degrees of freedom.

Acknowledgments

This research was partially supported by a research grant from the Natural Sciences and Engineering Research Council of Canada and a grant from the NNSF of China (10661003) to Y. Qin. The authors would like to thank the Associate Editor and the referees for constructive comments that substantially improved the presentation of the paper.

Appendix A.

Proof of Lemma 4. Without loss of generality, assume that $|\theta_1| \leqslant |\theta_2| \leqslant |\theta_3| \neq 0$. Let $\tilde{m}_k = m_k/\theta_3^k$, $k = 1, \ldots, 6$. We first show that

$$\tilde{m}_1^2 + \tilde{m}_3^2 + (\tilde{m}_4 - 3\tilde{m}_2^2)^2 + \tilde{m}_5^2 \geqslant C_0 > 0 \tag{A.1}$$

for some constant C_0 uniformly in a neighborhood of $\theta = 0$.

For this purpose, define $\mu_j = \theta_j/\theta_3$ for $j = 1, 2$. Consider any (p, μ_1, μ_2) such that $\tilde{m}_1 = 0, \tilde{m}_3 = 0$, and $\tilde{m}_5 = 0$. In matrix representation

$$(p^2, 2pq, q^2)\begin{bmatrix} \mu_1 & \mu_1^3 & \mu_1^5 \\ \mu_2 & \mu_2^3 & \mu_2^5 \\ 1 & 1 & 1 \end{bmatrix} = 0.$$

Thus, the solution exists only if the matrix is singular:

$$\mu_1\mu_2(\mu_1^2 - 1)(\mu_2^2 - 1)(\mu_2^2 - \mu_1^2) = 0.$$

There are few parameter configurations that solve the above equation. For each solution, it is easy to verify that $|\tilde{m}_4 - 3\tilde{m}_2^2| > 0$. Hence,

$$\tilde{m}_1^2 + \tilde{m}_3^2 + \tilde{m}_5^2 + (\tilde{m}_4 - 3\tilde{m}_2^2)^2$$

is a continuous, nowhere zero function over the range of (p, μ_1, μ_2). Since the range of (p, μ_1, μ_2) is compact, we have shown (A.1).

It follows that $\tilde{m}_6^2 \leqslant 1 \leqslant \{\tilde{m}_1^2 + \tilde{m}_3^2 + \tilde{m}_5^2 + (\tilde{m}_4 - 3\tilde{m}_2^2)\}/C_0$. Relating back to the s_j values, we have $m_6^2 \leqslant \varepsilon_0^2 \tilde{m}_6^2 = o(s_1^2 + s_3^2 + s_4^2 + s_5^2)$ as $\theta_j \to 0$ for $j = 1, 2, 3$.

We now prove (ii). Since $\sqrt{n}s \leqslant 1 + ns^2$,

$$\sqrt{n}(|m_1| + |m_3|)(\sigma^2 - 1) = \sqrt{n}(|s_1| + |s_3|)(\sigma^2 - 1) \leqslant \{2 + n(s_1^2 + s_3^2)\}(\sigma^2 - 1).$$

Hence, as $\sigma^2 \to 1$ and $n \to \infty$, we have the required order assessment for these terms.

For the other terms, we have

$$m_2(\sigma^2 - 1)^2 = m_2\{(\sigma^2 - 1 + m_2) - m_2\}^2 = m_2(s_2 - m_2)^2 \leqslant m_2(s_2^2 + m_2^2).$$

Since $m_2 < 1$ when ε_0 is small, we get $m_2(s_2^2 + m_2^2) \leqslant s_2^2 + m_6 = o(\{\sum_{j=1}^5 s_j^2\}^{1/2})$ as $\theta_j \to 0$. This gives the required order assessment.

For $m_4(\sigma^2 - 1)$ and $(\sigma^2 - 1)^3$, we have $m_4(\sigma^2 - 1) = (s_4 + 3m_2^2)(s_2 - m_2)$, and $|\sigma^2 - 1|^3 = |s_2 - m_2|^3 \leqslant 8(|s_2|^3 + m_2^3)$. Both are dominated by $o(\{\sum_{j=1}^5 s_j^2\}^{1/2})$ as $\theta_j \to 0$. Hence, we have the required result and the final conclusion. □

Proof of Lemma 5. It can be verified that

$$(\alpha_1^2 - \alpha_2^2)(\alpha_2\theta_2^3) = \alpha_1^2(m_3 - \alpha_3\theta_3^3) + (\alpha_3\theta_3 - m_1)^3 + 3\alpha_2\theta_2(\alpha_3\theta_3 - m_1)^2 + 3(\alpha_2\theta_2)^2(\alpha_3\theta_3 - m_1). \tag{A.2}$$

Recall that $\alpha_2 = 2pq$ and $\alpha_3 = q^3$. The fact that $|\theta_3| \geqslant \varepsilon_0$ in Region II implies that the estimator $\hat{q} \to 0$ when $n \to \infty$ in probability. Thus, we need consider only very small q and hence the situation where $\alpha_1^2 - \alpha_2^2 \geqslant \frac{1}{5}$. That is, it suffices to show that every term on the right-hand side of (A.2) tends to 0 as fast as $|s_1| + |s_3| + s_4$.

Since $|\alpha_3\theta_3^3| \leqslant s_4/\varepsilon_0$ and $m_3 = s_3$, the first term is controlled by $|s_3| + s_4$. The remaining terms are controlled by $|s_1| + s_4$ for a similar reason. Hence, we have

$$\alpha_2\theta_2^4 = o(\alpha_2\theta_2^3) = o(|s_1| + |s_3| + s_4).$$

Further,

$$\alpha_1\theta_1^4 = \theta_1(m_3 - \alpha_2\theta_2^3 - \alpha_3\theta_3^3) = o(|s_1| + |s_3| + s_4).$$

This completes the proof for the first part of the lemma.

For the second part, we note that $(|m_1| + |m_3|)|\sigma^2 - 1| = o(\{s_1^2 + s_3^2\}^{1/2})$. The crucial step is to show that the terms $m_2(\sigma^2 - 1)$ and $(\sigma^2 - 1)^2$ are under control.

As $\theta_1, \theta_2 \to 0$, $\alpha_3 \to 0$, and $\sigma^2 - 1 \to 0$, we have

$$m_2|\sigma^2 - 1| = m_2|s_2 - m_2| \leqslant m_2|s_2| + m_2^2 = o(|s_2|) + m_2^2$$

and

$$(\sigma^2 - 1)^2 = (s_2 - m_2)^2 \leqslant 2s_2^2 + 2m_2^2 = o(|s_2|) + m_2^2.$$

Thus, it suffices to show that m_2^2 is under control. Since

$$\begin{aligned}m_2^2 &\leqslant 2(\alpha_1\theta_1^2+\alpha_2\theta_2^2)^2+2(\alpha_3\theta_3^2)^2\\ &=2(\alpha_1\theta_1^2+\alpha_2\theta_2^2)^2+2\alpha_3 s_4\\ &\leqslant 2(\alpha_1+\alpha_2)(\alpha_1\theta_1^4+\alpha_2\theta_2^4)+\mathrm{o}(s_4),\end{aligned}$$

and given the first result of this lemma, we get $m_2^2=\mathrm{o}(\{\sum_{j=1}^4 s_j^2\}^{1/2})$ and the conclusion. □

Proof of Lemma 6. As $\sigma^2-1\to 0$, we have $m_1(\sigma^2-1)=\mathrm{o}(s_1)$. Similarly, $\alpha_3\theta_3^3((\sigma^2-1)=\mathrm{o}(s_4)$ and $|\alpha_2\theta_2^3|\leqslant|s_3|+|\gamma(\theta_2,\theta_3)s_4|$. Because $\gamma(\theta_2,\theta_3)$ is continuous over the compact space $|\theta_2|\leqslant M$ and $|\theta_3|\leqslant M$, it is bounded, and therefore $\alpha_2\theta_2^3(\sigma^2-1)=\mathrm{o}(|s_3|+|s_4|)$.

For $\alpha_1\theta_1^3$, as $\alpha_2,\alpha_3,\theta_1\to 0$,

$$\begin{aligned}\alpha_1|\theta_1|^3 &= \frac{1}{\alpha_1^2}|m_1-\alpha_2\theta_2-\alpha_3\theta_3|^3\\ &\leqslant \frac{27}{\alpha_1^2}\{|s_1|^3+\alpha_2^2(\alpha_2|\theta_2|^3)+\alpha_3^2(\alpha_3|\theta_3|^3)\}\\ &=\mathrm{o}(|s_1|+|s_3|+|s_4|).\end{aligned}$$

For $m_2(\sigma^2-1)$ and $(\sigma^2-1)^2$, we need work only on m_2^2 similarly to the proof of Lemma 5. We have

$$m_2^2\leqslant 2\alpha_1^2\theta_1^4+2(\alpha_2\theta_2^2+\alpha_3\theta_3^2)^2\leqslant 2\alpha_1^2\theta_1^4+2(\alpha_2+\alpha_3)^{4/3}(\alpha_2|\theta_2|^3+\alpha_3|\theta_3|^3)^{4/3}.$$

Hence we have Lemma 6. □

References

Charnigo, R., Sun, J., 2004. Testing homogeneity in a mixture distribution via the L^2-distance between competing models. J. Amer. Statist. Assoc. 99, 488–498.
Chen, H., Chen, J., 2001. Large sample distribution of the likelihood ratio test for normal mixtures. Statist. Probab. Lett. 52, 125–133.
Chen, H., Chen, J., 2003. Tests for homogeneity in normal mixtures in the presence of a structural parameter. Statist. Sinica 13, 351–365.
Chen, H., Chen, J., 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., Kalbfleisch, J.D., 2004. Testing for a finite mixture model with two components. J. Roy. Statist. Soc. Ser. B 66, 95–115.
Chen, J., 1998. Penalized likelihood ratio test for finite mixture models with multinomial observations. Canad. J. Statist. 26, 583–599.
Chen, J., Kalbfleisch, J.D., 2004. Modified likelihood ratio test in finite mixture models with a structural parameter. J. Statist. Plann. Inference 129, 93–107.
Chernoff, H., 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., Gassiat, E., 1999. Testing the order of a model using locally conic parameterization: population mixtures and stationary ARMA processes. Ann. Statist. 27, 1178–1209.
Davies, R.B., 1977. Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 64, 247–254.
Davies, R.B., 1987. Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 74, 33–43.
Garel, B., 2001. Likelihood ratio test for univariate Gaussian mixture. J. Statist. Plann. Inference 96, 325–350.
Garel, B., 2005. Asymptotic theory of the likelihood ratio test for the identification of a mixture. J. Statist. Plann. Inference 131, 271–296.
Ghosh, J.K., Sen, P.K., 1985. On the asymptotic performance of the log likelihood ratio statistic for the mixture model and related results. In: LeCam, L., Olshen, R.A. (Eds.), Proceedings of the Berkeley Conference in Honor of J. Neyman and J. Kiefer, vol. 2. Wadsworth, Belmont, pp. 789–806.
Hartigan, J.A., 1985. A failure of likelihood asymptotics for normal mixtures. In: LeCam, L., Olshen, R.A. (Eds.), Proceedings of the Berkeley Conference in Honor of J. Neyman and J. Kiefer, vol. 2. Wadsworth, Belmont, pp. 807–810.
Jones, P.N., McLachlan, G.J., 1991. Fitting mixture distributions to phenylthiocarbamide (PTC) sensitivity. Amer. J. Hum. Gene. 48, 117–120.
Lindsay, B.G., 1995. Mixture Models: Theory, Geometry and Applications. Institute of Mathematical Statistics, Hayward.
Liu, X., Shao, Y., 2003. Asymptotics for likelihood ratio tests under loss of identifiability. Ann. Statist. 31, 807–832.
McLachlan, G.J., Peel, D., 2003. Finite Mixture Models. Wiley Series in Probability and Statistics: Applied Probability and Statistics. Wiley-Interscience, New York.
Ott, J., 1999. Analysis of Human Genetic Linkage. third ed. 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, 487–500.
Rubin, H., 1956. Uniform convergence of random functions with applications to statistics. Ann. Math. Statist. 27, 200–203.
Schork, N.J., Allison, D.B., Thiel, B., 1996. Mixture distributions in human genetics research. Statist. Meth. Med. Res. 5, 155–178.
Tadesse, M.G., Sha, N., Vannucci, M., 2005. Bayesian variable selection in clustering high-dimensional data. J. Amer. Statist. Assoc. 100, 602–617.
Titterington, D.M., Smith, A.F.M., Makov, U.E., 1985. Statistical Analysis of Finite Mixture Distributions. Wiley, New York.
Wald, A., 1949. Note on the consistency of the maximum likelihood estimate. Ann. Math. Statist. 30, 185–191.
Zhu, H.T., Zhang, H.P., 2004. Hypothesis testing in mixture regression models. J. Roy. Statist. Soc. B 66, 3–16.