Bernoulli **25**(4B), 2019, 3400–3420
https://doi.org/10.3150/18-BEJ1094

Rate of divergence of the nonparametric likelihood ratio test for Gaussian mixtures

WENHUA JIANG[1] and CUN-HUI ZHANG[2]

[1]*School of Mathematical Sciences, Fudan University and Shanghai Center for Mathematical Sciences, 220 Handan Road, Shanghai 200433, China. E-mail: wenhuajiang@yahoo.com*
[2]*Department of Statistics and Biostatistics, Rutgers University, 110 Frelinghuysen Road, Piscataway, NJ 08854, U.S.A. E-mail: czhang@stat.rutgers.edu*

We study a nonparametric likelihood ratio test (NPLRT) for Gaussian mixtures. It is based on the nonparametric maximum likelihood estimator in the context of demixing. The test concerns if a random sample is from the standard normal distribution. We consider mixing distributions of unbounded support for alternative hypothesis. We prove that the divergence rate of the NPLRT under the null is bounded by $\log n$, provided that the support range of the mixing distribution increases no faster than $(\log n/\log 9)^{1/2}$. We prove that the rate of $\sqrt{\log n}$ is a lower bound for the divergence rate if the support range increases no slower than the order of $\sqrt{\log n}$. Implications of the upper bound for the rate of divergence are discussed.

Keywords: Gaussian mixtures; Hermite polynomials; likelihood ratio test; rate of divergence; two-component mixtures

1. Introduction

Define the standard normal location-mixture density

$$f_G(x) = \int \varphi(x-u)\,dG(u), \tag{1.1}$$

where $\varphi(x) = \exp(-x^2/2)/\sqrt{2\pi}$ is the standard normal density. Let $\mathscr{G}$ be the collection of all distributions in the real line $\mathbb{R}$ and take $\mathscr{F} = \{f_G : G \in \mathscr{G}\}$ as the family of all standard normal location-mixture densities. Let $X_1, \dots, X_n$ be independent and identically distributed observations with probability density function f. We consider testing the null hypothesis that the sample is generated from φ against the general alternative that the sample is from a mixture density $f_G \in \mathscr{F}$ other than φ. For $f_1, f_2 \in \mathscr{F}$, the log-likelihood ratio is defined as

$$\ell_n(f_1, f_2) = \sum_{i=1}^n \log \frac{f_1(X_i)}{f_2(X_i)}.$$

Let $\mathscr{F}_n \subset \mathscr{F}$ be a sequence of density families. For testing $H_0: f = \varphi$ against $H_1: f \in \mathscr{F}_n \setminus \{\varphi\}$, the nonparametric likelihood ratio test (NPLRT) statistic is

$$\sup_{f\in\mathscr{F}_n} \ell_n(f,\varphi) = \ell_n(\hat{f}_n, \varphi), \tag{1.2}$$

where $\hat{f}_n = \arg\max_{f\in\mathcal{F}_n} \prod_{i=1}^n f(X_i)$ is the nonparametric maximum likelihood estimator (NPMLE, Robbins [17]; Kiefer and Wolfowitz [13]) of normal mixture density. We are interested in the case where

$$\mathcal{F}_n = \{f_G : G([-M_n, M_n]) = 1\} \tag{1.3}$$

with $M_n > 0$, especially $M_n \to \infty$.

The MLE in (1.2) is nonparametric in the sense that $\mathcal{F}_n$ is a family of "infinite Gaussian mixture", instead of the well-known finite mixtures. It is widely thought that the analysis of the NPMLE is challenging (e.g., DasGupta [4], Chapter 33). What are the asymptotic properties of the NPLRT? First of all, $\hat{f}_n$ is consistent (Ghosal and van der Vaart [7]; Zhang [19]). Due to loss of identifiability, the asymptotic distribution of the NPLRT is not the usual χ^2 distribution. Hartigan [9] discovered that under the null, the NPLRT diverges to infinity in probability as $M_n \to \infty$. Jiang and Zhang [12] proved that when $\mathcal{F}_n = \mathcal{F}$, the rate of divergence is bounded by $(\log n)^2$. Gu, Koenker and Volgushev [8] provided interesting numerical results for the asymptotic behaviors of the test under different settings. Liu and Shao [15] considered asymptotics of likelihood ratio tests under loss of identifiability. Azaïs, Gassiat and Mercadier [2] considered the asymptotic null distribution of the NPLRT with mixing distributions of bounded support. See Section 4 for some asymptotic properties of the parametric LRT (PLRT).

The notation to be used is listed first for easy reference. We will use the abbreviation $\mathbb{P}f = Ef(X)$ for an integrable function f. Its empirical counterpart is denoted by $\mathbb{P}_n f = \sum_{i=1}^n f(X_i)/n$. In this paper, the expectation $\mathbb{P}$ is taken with respect to the standard normal density φ. For $f \in \mathcal{L}_p(\mathbb{P})$, define the $\mathcal{L}_p(\mathbb{P})$ norm as $\|f\|_p = \{\mathbb{P}(f^p)\}^{1/p} = \{\int |f(x)|^p \varphi(x)\,dx\}^{1/p}$. The Hellinger distance between two densities f and g is defined as $d^2(f,g) = (1/2)\int(\sqrt{f} - \sqrt{g})^2\,dx$. Given a collection $\mathcal{F}$ of $\mathcal{L}_1(\mathbb{P})$ functions, the $\mathcal{F}$-indexed empirical process ν_n is given by $\{\nu_n f = \sqrt{n}(\mathbb{P}_n - \mathbb{P})f, f \in \mathcal{F}\}$. Throughout the paper, $x \wedge y = \min(x,y)$, $x_+ = \max(x,0)$ and $a_n \asymp b_n$ means $0 < a_n/b_n + b_n/a_n = O(1)$.

The following theorem summarizes some results based on Ghosal and van der Vaart [7], Jiang and Zhang [11] and Jiang and Zhang [12].

Theorem 1. *There exists* $\{f_{G_j}, 1 \le j \le N_n\} \subset \mathcal{F}_n$ *and* $\varepsilon_n \asymp n^{-1/2}\log n$ *such that* $\log N_n \le n\varepsilon_n^2$, $\max_{1\le j\le N_n} d(f_{G_j}, \varphi) \le \varepsilon_n$ *and*

$$\sup_{f\in\mathcal{F}_n} \ell_n(f,\varphi) = (1+\eta_n) \sup_{1\le j\le N_n} \ell_n(f_{G_j},\varphi) \le n\varepsilon_n^2$$

with large probability, where $|\eta_n| \asymp 1/n$.

Theorem 1 says that the NPLRT is nearly achieved by a finite collection of Gaussian mixtures $\{f_{G_j}, 1 \le j \le N_n\}$ of manageable size. This collection can be regarded as approximate NPMLEs such that $d(f_{G_j}, \varphi) \le \varepsilon_n$ for all $1 \le j \le N_n$. In this paper, as mentioned in (1.3), we allow the support range of mixing distribution goes to infinity. We prove that the order of the NPLRT in (1.2) is bounded by $\log n$ in probability, provided that the support range of the mixing distribution goes to infinity no faster than $(\log n/\log 9)^{1/2}$. This gives an upper bound. The discretization in Theorem 1 is an element for the analysis of upper bound. We prove that the rate of $\sqrt{\log n}$ is a lower bound for the divergence rate if the support range increases no slower than the order of $\sqrt{\log n}$.

The rest of this paper is organized as follows. The Hermite polynomial expansion of Gaussian mixtures is introduced in Section 2. In Section 3, we study the upper bound for the rate of divergence of the NPLRT. The lower bound is given in Section 4. The implications of the rate of divergence, some simulations and other mixtures are discussed in Section 5. Proofs are given in Section 6.

2. The Hermite polynomial expansion

As in Azaïs, Gassiat and Mercadier [2], our analysis is based on the expansion of Gaussian mixtures by Hermite polynomials. The Hermite polynomials are defined as

$$H_k(x) = (-1)^k e^{x^2/2} \frac{d^k}{dx^k} e^{-x^2/2} \tag{2.1}$$

with $H_0(x) = 1$, $H_1(x) = x$, $H_2(x) = x^2 - 1$, etc. It is well known that $\int H_j(x) H_k(x) \varphi(x)\, dx = k! I\{j = k\}$. The Gaussian mixtures have the expansion

$$f_G(x) = \varphi(x) \sum_{k=0}^{\infty} \frac{\mu_k(G)}{k!} H_k(x), \tag{2.2}$$

where $\mu_k(G) = \int u^k\, dG(u)$ is the kth moment of G. Let

$$h_G(x) = \frac{f_G(x)/\varphi(x) - 1}{\| f_G/\varphi - 1\|_2} \tag{2.3}$$

be the generalized score function (Liu and Shao [15]). Note that the likelihood ratio f_G/φ is square integrable. It follows from the expansion above that for $f_G \neq \varphi$,

$$h_G(x) = \sum_{k=1}^{\infty} c_k(G) \frac{H_k(x)}{\sqrt{k!}}, \tag{2.4}$$

where $c_k(G) = \{\mu_k(G)/\sqrt{k!}\}/\{\sum_{j=1}^{\infty} \mu_j^2(G)/j!\}^{1/2}$.

Define two envelope functions

$$F_{1,n}(x) = \left\{ \sum_{k=1}^{m} \frac{H_k^2(x)}{k!} \right\}^{1/2} \quad \text{and} \quad F_{2,n}(x) = \sum_{k=m+1}^{\infty} \sqrt{C_0/M_n} e^{-(k-M_n^2)^2/(4k)} \frac{|H_k(x)|}{\sqrt{k!}}, \tag{2.5}$$

where C_0 is a suitable constant. The expansion in (2.4) and Lemma 1 below imply that for all integers $m \geq M_n^2$,

$$\sup_{G([-M_n, M_n])=1} h_G(x) \leq F_{1,n}(x) + F_{2,n}(x). \tag{2.6}$$

Lemma 1. *Let $c_k(G)$ be the coefficients as in (2.4). There exists a positive constant C_0 such that for all $G([-M_n, M_n]) = 1$,*

$$c_k^2(G) \le \left(\frac{C_0}{M_n}\right) e^{-(k-M_n^2)^2/(2k)}, \qquad \forall k \ge M_n^2. \tag{2.7}$$

Proof of (2.6). For all $G([-M_n, M_n]) = 1$, (2.4) and (2.7) lead to

$$\begin{aligned} h_G(x) &= \sum_{k=1}^{m} c_k(G) \frac{H_k(x)}{\sqrt{k!}} + \sum_{k=m+1}^{\infty} c_k(G) \frac{H_k(x)}{\sqrt{k!}} \\ &\le \left\{ \sum_{k=1}^{m} H_k^2(x)/k! \right\}^{1/2} + \sum_{k=m+1}^{\infty} \sqrt{C_0/M_n} e^{-(k-M_n^2)^2/(4k)} \frac{|H_k(x)|}{\sqrt{k!}} \\ &= F_{1,n}(x) + F_{2,n}(x). \end{aligned}$$

This completes the proof of (2.6). □

Lemma 2. *Let $F_{1,n}(x)$ and $F_{2,n}(x)$ be the envelope functions defined as in (2.5). Then,*

(i) *For all positive integers m, $\|F_{1,n}\|_2 = m^{1/2}$ and*

$$\|F_{1,n}\|_4^4 \le 9^{m+1}/8. \tag{2.8}$$

(ii) *There exists a constant C_1 such that for all integers m satisfying $(m - M_n^2)/\sqrt{m} \ge 2t_n$,*

$$\|F_{2,n}\|_2 \le \sqrt{\frac{C_0}{M_n}} \sum_{k=m+1}^{\infty} e^{-(k-M_n^2)^2/(4k)} \le C_1 M_n^{1/2} e^{-t_n^2}. \tag{2.9}$$

The upper bound of the fourth moment of the Hermite polynomials in (2.8) will be applied in Lemma 3. Specifically, it is used to prove the uniform square integrability of h_G when f_G is in a neighborhood of φ in $\mathcal{F}_n$. This provides sufficient conditions for the equivalence between the Hellinger distance $d(f, \varphi)$ and the Pearson type $\mathcal{L}_2$ distance $\|f/\varphi - 1\|_2$. The detail of the proofs of (2.8) is in Section 6. There, it is shown that

$$\int H_k^4(x)\varphi(x)\,dx = (k!)^2 \sum_{l=0}^{k} \binom{k}{l}^2 \binom{2l}{l} \le (k!)^2 \sum_{l=0}^{k} \binom{k}{l}^2 2^{2l} = (k!)^2 9^k.$$

Since $\binom{2l}{l} \asymp 2^{2l}/\sqrt{l}$ by Stirling's formula, the base 9 in (2.8) is tight.

3. Upper bound for the rate of divergence

Let $\{f_{G_j}, 1 \le j \le N_n\}$ be the collection of Gaussian mixtures as in Theorem 1. Define $\psi(x) = 2(x - \log(1+x))$. We have $\psi(x) \ge 0$ and

$$2\ell_n(f_{G_j}, \varphi) = 2n\mathbb{P}_n(f_{G_j}/\varphi - 1) - n\mathbb{P}_n\psi(f_{G_j}/\varphi - 1). \tag{3.1}$$

Suppose that we have the following condition,

$$\mathbb{P}_n\psi(f_{G_j}/\varphi - 1) \ge \big(1 + o_{\mathbb{P}}(1)\big)\|f_{G_j}/\varphi - 1\|_2^2, \qquad \forall 1 \le j \le N_n. \tag{3.2}$$

Let $h_j = (f_{G_j}/\varphi - 1)/\|f_{G_j}/\varphi - 1\|_2$. Then, when (3.2) holds,

$$\begin{aligned} &2\max_{1\le j\le N_n} \ell_n(f_{G_j}, \varphi) \\ &\quad \le \max_{1\le j\le N_n} \big\{2n\mathbb{P}_n(f_{G_j}/\varphi - 1) - \big(1 + o_{\mathbb{P}}(1)\big)n\|f_{G_j}/\varphi - 1\|_2^2\big\} \\ &\quad = \max_{1\le j\le N_n} \big\{2\sqrt{n}\|f_{G_j}/\varphi - 1\|_2\nu_n(h_j) - \big(1 + o_{\mathbb{P}}(1)\big)n\|f_{G_j}/\varphi - 1\|_2^2\big\} \end{aligned} \tag{3.3}$$

with large probability. The supremum of $\ell_n(f, \varphi)$ can be bounded by maximizing the quadratic form of $\sqrt{n}\|f_{G_j}/\varphi - 1\|_2$ in (3.3), which can be written as

$$2\max_{1\le j\le N_n} \ell_n(f_{G_j}, \varphi) \le \big(1 + o_{\mathbb{P}}(1)\big)\Big\{\max_{1\le j\le N_n} \nu_n(h_j)\Big\}_+^2. \tag{3.4}$$

This approach was taken in Liu and Shao [15].

Theorem 2. *Let $\mathcal{F}_n$ be as in (1.3). Suppose (3.2) holds. Then,*

$$\sup_{f\in\mathcal{F}_n} \ell_n(f, \varphi) = O_{\mathbb{P}}\big(M_n^2\big). \tag{3.5}$$

Proof of Theorem 2. By (3.4), it suffices to bound $\{\max_{j\le N_n} \nu_n(h_j)\}_+^2$. As in (2.4), we write h_j as $h_j = \sum_{k=1}^\infty c_k(G_j)H_k/\sqrt{k!}$. It follows from (2.7) and (2.9) that for integers m satisfying $(m - M_n^2)/\sqrt{m} \ge 2t_n$, we have

$$\begin{aligned} &\mathbb{P}\max_{1\le j\le N_n}\left|\nu_n\left(\sum_{k=m+1}^\infty c_k(G_j)H_k/\sqrt{k!}\right)\right| \\ &\quad \le \sum_{k=m+1}^\infty \sqrt{C_0/M_n}e^{-(k-M_n^2)^2/(4k)}\mathbb{P}\big|\nu_n(H_k/\sqrt{k!})\big| \\ &\quad \le C_1 M_n^{1/2}e^{-t_n^2}. \end{aligned}$$

In the last inequality we used the fact that $\mathbb{P}|\nu_n(H_k/\sqrt{k!})|^2 = 1$. Taking m such that $t_n = \sqrt{\log M_n}$, by the Markov inequality we have $\max_{1\le j\le N_n} \nu_n(\sum_{k=m+1}^{\infty} c_k(G_j)H_k/\sqrt{k!}) = o_{\mathbb{P}}(1)$. Since $(a+b)^2 \le 2(a^2+b^2)$, by Theorem 1 and (3.4),

$$\begin{aligned} 2\sup_{f\in\mathcal{F}_n} \ell_n(f,\varphi) &\le 2\big(1+o(1)\big)\max_{1\le j\le N_n} \ell_n(f_{G_j},\varphi) \\ &\le 2\big(1+o_{\mathbb{P}}(1)\big)\left\{\max_{1\le j\le N_n}\sum_{k=1}^{m} c_k(G_j)\nu_n(H_k/\sqrt{k!})\right\}_+^2 + o_{\mathbb{P}}(1). \end{aligned} \tag{3.6}$$

Note that

$$\mathbb{P}\left\{\max_{1\le j\le N_n}\sum_{k=1}^{m} c_k(G_j)\nu_n(H_k/\sqrt{k!})\right\}_+^2 \le \mathbb{P}\sum_{k=1}^{m}\big\{\nu_n(H_k/\sqrt{k!})\big\}^2 = m.$$

Hence by the Markov inequality

$$\left\{\max_{1\le j\le N_n}\sum_{k=1}^{m} c_k(G_j)\nu_n(H_k/\sqrt{k!})\right\}_+^2 = O_{\mathbb{P}}(m).$$

This, (3.6) and $(m-M_n^2)/\sqrt{m} \ge 2t_n = 2\sqrt{\log M_n}$ imply (3.5). □

Remark 1. The crucial elements that lead to Theorem 2 are the bounds for the coefficients $c_k(G)$ in the Hermite polynomial expansion (2.4). They are established in Lemmas 1 and 2. The consequence is that for the expansion of $\max_{1\le j\le N_n}\nu_n(h_j)$, the remainder beyond the mth term is negligible. Because the order of m is as large as M_n^2, the upper bound of the rate of divergence is M_n^2.

Theorem 3. *Let $\mathcal{F}_n$ be as in* (1.3). *If $M_n \le \sqrt{a_0 \log n}$ with $a_0 < 1/\log 9$, then* (3.2) *holds. Consequently,*

$$\sup_{f\in\mathcal{F}_n} \ell_n(f,\varphi) = O_{\mathbb{P}}\big(M_n^2 \wedge (\log n)\big). \tag{3.7}$$

It is known that the NPLRT with mixing distributions of unbounded support diverges to infinity in probability (Hartigan [9]). Jiang and Zhang [12] proved that the rate of divergence is of equal or smaller order than $(\log n)^2$. Theorem 3 improves upon the rate to $\log n$, provided that the support range of mixing distribution goes to infinity no faster than $(\log n/\log 9)^{1/2}$. Clearly the divergence rate is slow. Numerical results of the slow divergence of the critical values are demonstrated in Gu, Koenker and Volgushev [8].

To prove Theorem 3, we need the following lemma which provides sufficient conditions for the equivalence between the Hellinger distance $d(f,\varphi)$ and the Pearson type $\mathcal{L}_2$ distance $\|f/\varphi - 1\|_2$.

Lemma 3. *Let ε_n and C_n be sequences of positive constants satisfying $\varepsilon_n C_n \to 0$. Suppose that*

$$\sup_{f\in\mathscr{F}_n, d(f,\varphi)\le\varepsilon_n} \left\| \left(\frac{|f/\varphi - 1|}{\|f/\varphi - 1\|_2} - C_n \right)_+ \right\|_2 \to 0. \tag{3.8}$$

Then,

$$\sup_{f\in\mathscr{F}_n, d(f,\varphi)\le\varepsilon_n} \left| 1 - \frac{2\sqrt{2}d(f,\varphi)}{\|f/\varphi - 1\|_2} \right| \to 0. \tag{3.9}$$

In particular, condition (3.8) holds when $\mathscr{F}_n = \{f_G : G([-M_n, M_n]) = 1\}$ with $M_n \le \sqrt{a_0 \log n}$ for any $a_0 < 1/\log 9$.

For each f, there exists constants C_n such that $\|(|f/\varphi - 1|/\|f/\varphi - 1\|_2 - C_n)_+\|_2 \to 0$. When (3.8) holds, $|f/\varphi - 1|/\|f/\varphi - 1\|_2$ is said to be uniformly square integrable over $\mathscr{F}_n \cap \{f : d(f,\varphi) \le \varepsilon_n\}$. Our proof of (3.8) essentially requires the condition that $\int H_m^4(x)\varphi(x)\,dx \ll n(m!)^2$ for $m = M_n^2(1 + o(1))$ (see proof of Lemma 3). This condition holds iff $M_n < \sqrt{\log n/\log 9}$.

Proof of Theorem 3. Let

$$\psi_-(x) = \begin{cases} 2x - 2\log(1+x) & x > 0, \\ x^2 & x \le 0. \end{cases} \tag{3.10}$$

Let $\{f_{G_j}, 1 \le j \le N_n\}$ be the collection of Gaussian mixtures as in Theorem 1. As $\psi_-(x) \le \psi(x) = 2(x - \log(1+x))$, it suffices to prove that

$$\inf_{1\le j\le N_n} \mathbb{P}_n \frac{\psi_-(f_{G_j}/\varphi - 1)}{\|f_{G_j}/\varphi - 1\|_2^2} \ge 1 + o_{\mathbb{P}}(1). \tag{3.11}$$

We have $0 \le \psi_-(x) \le x^2$ for all x. For $\sigma > 0$ and $x > 0$,

$$\psi_-(\sigma x) = 2\int_0^{\sigma x} \left(1 - \frac{1}{1+t}\right) dt = 2\sigma^2 \int_0^x \frac{t}{1+\sigma t}\,dt \ge \frac{2\sigma^2}{1+\sigma x}\int_0^x t\,dt = \frac{(\sigma x)^2}{1+\sigma x}.$$

This implies that for all $x \in \mathbb{R}$,

$$\max_{\tau>0} \frac{x(x\wedge\tau)}{1+\sigma\tau} = \frac{x^2}{1+\sigma x_+} \le \frac{\psi_-(\sigma x)}{\sigma^2} \le x^2. \tag{3.12}$$

Then, (3.12) yields

$$0 \le \frac{h_j(h_j \wedge C_n)}{1 + \|f_{G_j}/\varphi - 1\|_2 C_n} \le \frac{\psi_-(\|f_{G_j}/\varphi - 1\|_2 h_j)}{\|f_{G_j}/\varphi - 1\|_2^2} = \frac{\psi_-(f_{G_j}/\varphi - 1)}{\|f_{G_j}/\varphi - 1\|_2^2} \le h_j^2. \tag{3.13}$$

Since $M_n^2 \le a_0 \log n$ with $a_0 < 1/\log 9$, it follows from Lemma 3 that $2d(f_{G_j},\varphi)/\|f_{G_j}/\varphi - 1\|_2 = 1 + o(1)$ uniformly. By Theorem 1, we assume without loss of generality that $\|f_{G_j}/\varphi -$

$1\|_2 \le \varepsilon_n$. Let $C_n = 1/(\varepsilon_n \log n)$, so that $\|f_{G_j}/\varphi - 1\|_2 C_n \le \varepsilon_n C_n \to 0$. Then, due to (3.13) and $\mathbb{P}h_j^2 = 1$, (3.11) follows from

$$\max_{1\le j\le N_n} \bigl|(\mathbb{P}_n - \mathbb{P})\bigl(h_j^2 \wedge C_n^2\bigr)\bigr| = o_{\mathbb{P}}(1), \tag{3.14}$$

and

$$\mathbb{P}\Bigl\{\max_{1\le j\le N_n} \mathbb{P}_n\bigl(h_j^2 - C_n^2\bigr)_+\Bigr\} = o(1). \tag{3.15}$$

Since $\mathrm{var}(h_j^2 \wedge C_n^2) \le \mathbb{P}(h_j^2 \wedge C_n^2)^2 \le C_n^2 \mathbb{P}h_j^2 = C_n^2$, the Bernstein's inequality (van der Vaart and Wellner [18], page 102) yields

$$\mathbb{P}\Bigl\{\max_{1\le j\le N_n} \bigl|(\mathbb{P}_n - \mathbb{P})\bigl(h_j^2 \wedge C_n^2\bigr)\bigr| > t\Bigr\} \le 2N_n \exp\biggl\{-\frac{nt^2/2}{C_n^2 + C_n^2 t/3}\biggr\}.$$

Since $\log N_n \le n\varepsilon_n^2$ by Theorem 1 and $n/C_n^2 = n\varepsilon_n^2(\log n)^2 \gg n\varepsilon_n^2$, (3.14) holds. For (3.15), (2.6) and (2.8) give that

$$\begin{aligned}
\mathbb{P}\max_{1\le j\le N_n} \mathbb{P}_n\bigl(h_j^2 - C_n^2\bigr)_+ &\le \mathbb{P}\max_{1\le j\le N_n} \mathbb{P}_n\bigl(2|h_j| - C_n\bigr)_+^2 \\
&\le 4\mathbb{P}\bigl\{(F_{1,n} - C_n/2)_+ + F_{2,n}\bigr\}^2 \\
&\le 8\mathbb{P}(F_{1,n} - C_n/2)_+^2 + 8\mathbb{P}F_{2,n}^2 \\
&\le \bigl(32/C_n^2\bigr)\mathbb{P}F_{1,n}^4 + 8\mathbb{P}F_{2,n}^2 \\
&\le O(1)3^{2m}/C_n^2 + o(1).
\end{aligned}$$

In the last inequality we used (2.9) when $(m - M_n^2)/\sqrt{m} \ge 2t_n = 2\sqrt{\log M_n}$. The right-hand side above is $o(1)$ by (6.9) in the proof of Lemma 3. Hence, (3.15) also holds. This completes the proof of Theorem 3. $\square$

4. Lower bound for the rate of divergence

To derive a lower bound, we may consider a subfamily $\mathcal{F}_n' \subset \mathcal{F}_n$. Since $\sup_{f\in\mathcal{F}_n} \ell_n(f,\varphi) \ge \sup_{f\in\mathcal{F}_n'} \ell_n(f,\varphi)$, any lower bound for $\sup_{f\in\mathcal{F}_n'} \ell_n(f,\varphi)$ is a lower bound for $\sup_{f\in\mathcal{F}_n} \ell_n(f,\varphi)$. Probably, the family of two-component Gaussian mixture is the most natural and simplest choice. Let

$$\mathcal{H} = \bigl\{(1-p)\varphi(x) + p\varphi(x-\mu)\colon\ p \in [0,1], \mu \in \mathbb{R}\bigr\}. \tag{4.1}$$

The parametric LRT (PLRT) statistic $\sup_{f\in\mathcal{H}} \ell_n(f,\varphi)$ has been investigated in literature. Hartigan [9] discovered the PLRT statistic diverges to infinity in probability, and further conjectured that the rate of divergence is $\log\log n$. Using a result in Bickel and Chernoff [3], Liu and Shao

[16] confirmed Hartigan's conjecture, and further proved the asymptotic null distribution. Although the class of mixtures with mixing distributions of unbounded support is considerably larger than the class of two-component mixtures, the support range of the mixing distribution in Theorem 3 increases no faster than $(\log n/\log 9)^{1/2}$. Hence, the rate of $\log\log n$ cannot serve as a lower bound for the rate of divergence directly. The next theorem shows that the $\sqrt{\log n}$ rate is a lower bound if the support range increases no slower than the order of $\sqrt{\log n}$.

Theorem 4. *Let $\mathcal{F}_n$ be as in (1.3). For $M_n = \sqrt{c_0 \log n}$ with $c_0 > 0$,*

$$\mathbb{P}\Big\{\sup_{f\in\mathcal{F}_n} \ell_n(f,\varphi) \lesssim (\log n)^{1/2}\Big\} \to 0. \tag{4.2}$$

Proof of Theorem 4. Let $0 = u_0 < u_1 \le \cdots \le u_m = \sqrt{c_0 \log n}$ for $c_0 > 0$. Let $\widetilde{G}$ be the mixing distribution putting mass π_j at u_j for $j = 1, \ldots, m$, and the rest of the mass π_0 at 0. Similar to Bickel and Chernoff [3] and Liu and Shao [16], let $\pi_j = n^{-1/2} w_j e^{-u_j^2/2}$ with random w_j, $j = 1, \ldots, m$. For a single observation X_i, the likelihood ratio for $N(u_j, 1)$ against $N(0, 1)$ is $e^{u_j X_i - u_j^2/2}$. The log-likelihood ratio for all data points can be written as

$$\begin{aligned}
\sum_{i=1}^n \log \frac{f_{\widetilde{G}}(X_i)}{\varphi(X_i)} &= \sum_{i=1}^n \log\Bigg(\sum_{j=0}^m \pi_j e^{u_j X_i - u_j^2/2}\Bigg) \\
&= \sum_{i=1}^n \log\Bigg(1 + \sum_{j=1}^m \pi_j \big(e^{u_j X_i - u_j^2/2} - 1\big)\Bigg) \\
&= \sum_{i=1}^n \log\Bigg(1 + \sum_{j=1}^m w_j n^{-1/2} \big(e^{u_j X_i - u_j^2/2} - 1\big) e^{-u_j^2/2}\Bigg).
\end{aligned}$$

Let $Z_{i,j} = n^{-1/2}(e^{u_j X_i - u_j^2/2} - 1)e^{-u_j^2/2}$. Consider the case where $|\sum_{j=1}^m w_j Z_{i,j}| \le 1/2$. Let $f(x) = \log(1+x) - x + x^2$. Since $f(0) = 0$, $f'(x) \le 0$ for $-1/2 \le x \le 0$ and $f'(x) > 0$ for $0 < x \le 1/2$, $\log(1+x) - x \ge -x^2$ for $|x| \le 1/2$. Thus,

$$\begin{aligned}
\sum_{i=1}^n \log \frac{f_{\widetilde{G}}(X_i)}{\varphi(X_i)} &\ge \sum_{i=1}^n \Bigg\{\sum_{j=1}^m w_j Z_{i,j} - \Bigg(\sum_{j=1}^m w_j Z_{i,j}\Bigg)^2\Bigg\} \\
&= \sum_{j=1}^m w_j S_j - \sum_{j=1}^m \sum_{k=1}^m w_j w_k U_{j,k},
\end{aligned} \tag{4.3}$$

where

$$S_j = \sum_{i=1}^n Z_{i,j} = \sum_{i=1}^n n^{-1/2}\big(e^{u_j X_i - u_j^2/2} - 1\big)e^{-u_j^2/2},$$

and

$$U_{j,k} = \sum_{i=1}^{n} Z_{i,j} Z_{i,k} = \frac{1}{n} \sum_{i=1}^{n} \big(e^{u_j X_i - u_j^2/2} - 1\big)\big(e^{u_k X_i - u_k^2/2} - 1\big) e^{-u_j^2/2 - u_k^2/2}.$$

Let $w_j = \tau\{\operatorname{sgn}(S_j) + 1\}/2$ with $\tau > 0$. We have

$$\sum_{j=1}^{m} w_j S_j - \sum_{j=1}^{m}\sum_{k=1}^{m} w_j w_k U_{j,k} \geq \tau \sum_{j=1}^{m} \max(S_j, 0) - \tau^2 \sum_{j=1}^{m}\sum_{k=1}^{m} |U_{j,k}|.$$

Since $\mathbb{E} e^{tX_i} = e^{t^2/2}$, we have $\mathbb{E}|U_{j,k}| \leq (e^{(u_j+u_k)^2/2 - u_j^2/2 - u_k^2/2} + 3) e^{-u_j^2/2 - u_k^2/2} \leq 4e^{-(u_j - u_k)^2/2}$. Let $u_j = j$. Then,

$$\mathbb{E} \sum_{j=1}^{m}\sum_{k=1}^{m} |U_{j,k}| \leq 4 \sum_{j=1}^{m}\sum_{k=1}^{m} e^{-(j-k)^2/2} \leq C_1 m.$$

Because $\mathbb{E}(e^{u_j X_i - u_j^2/2} - 1) e^{-u_j^2/2} = 0$ and $\mathbb{E}\{(e^{u_j X_i - u_j^2/2} - 1) e^{-u_j^2/2}\}^2 = 1 - e^{-u_j^2}$, $S_j \to N(0,1)$ as $j \to \infty$. Then, there exists a small constant τ such that

$$\mathbb{E}\left\{\tau \sum_{j=1}^{m} \max(S_j, 0) - \tau^2 \sum_{j=1}^{m}\sum_{k=1}^{m} |U_{j,k}|\right\} \geq \tau C_2 m \mathbb{E}\big(N(0,1)\big)_+ - \tau^2 C_1 m \geq C_3 m.$$

Because the order of m is $\sqrt{\log n}$, we find that

$$\sum_{i=1}^{n} \log \frac{f_{\widetilde{G}}(X_i)}{\varphi(X_i)} \gtrsim \sqrt{\log n}$$

with large probability.

It remains to verify that $|\sum_{j=1}^{m} w_j Z_{i,j}| \leq 1/2$ with large probability. It suffices to consider $\max |X_i| \leq \sqrt{2\log n}$ as it holds with large probability. We have,

$$\begin{aligned}
\left|\sum_{j=1}^{m} w_j Z_{i,j}\right| &= \left|\sum_{j=1}^{m} w_j n^{-1/2} \big(e^{u_j X_i - u_j^2/2} - 1\big) e^{-u_j^2/2}\right| \\
&\leq 2\tau n^{-1/2} \sum_{j=1}^{m} e^{u_j \sqrt{2\log n} - u_j^2} \\
&= 2\tau \sum_{j=1}^{m} e^{-(u_j - \sqrt{2\log n}/2)^2} \leq C_4 \tau.
\end{aligned}$$

The last inequality holds because for $u_j = j$ the summands decrease faster than geometric rate from center. Taking a constant $\tau \leq C_4/2$ gives that $|\sum_{j=1}^{m} w_j Z_{i,j}| \leq 1/2$. This completes the proof of Theorem 4. $\square$

In the proof, we set $u_j = j$, so the number of support points is of order $\sqrt{\log n}$. It is known that the NPMLE is a mixture of at most n components (Lindsay [14]). However, putting denser equal-spaced support points in support range does not improve the lower bound in our analysis.

Remark 2. It is unclear whether the exact rate of divergence of the NPLRT is $\sqrt{\log n}$ or $\log n$, or a rate between them. Our analyses for the upper and lower bounds are very different. The proof of the upper bound is based on the expansion of the standardized likelihood ratio in the Hermite polynomial basis in (2.4), which leads to the use of the fourth moment condition on the envelope function $F_{1,n}$ as we remarked below the statement of Lemma 3. This seems to be cruder than the lower bound analysis. However, we are unable to use the more natural expansion in the lower bound calculation to derive an upper bound as the magnitude of w_j is hard to control. In any case, little is known about the NPMLE, in spite of Lindsay [14], Genovese and Wasserman [6], Ghosal and van der Vaart [7], Zhang [19] and other studies in the literature.

5. Discussion

5.1. Implications of the upper bound

The divergence rate of the NPLRT was proved to be bounded by $(\log n)^2$ in Jiang and Zhang [12] by a large deviation inequality:

$$P_{H_0}\left\{\sum_{i=1}^{n} \log \frac{\hat{f}_n(X_i)}{\varphi(X_i)} \geq n\varepsilon_n^2\right\} \to 0. \tag{5.1}$$

This has an implication of the two-component testing problem:

$$H_0\colon\ f(x) = \varphi(x) \quad \text{against} \quad H_1^{(n)}\colon\ f(x) = (1-p_n)\varphi(x) + p_n\varphi(x-\mu_n). \tag{5.2}$$

The alternative says only a small fraction of normal means is nonzero, and they have the same value. In the sparse case, namely, $1/2 < \beta < 1$, let $p_n = n^{-\beta}$ and $\mu_n = \sqrt{2r\log n}$ be calibration. Let $\rho^*(\beta) = \beta - 1/2$ if $1/2 < \beta \leq 3/4$, and $\rho^*(\beta) = (1-\sqrt{1-\beta})^2$ otherwise. The detection limit of (5.2) is given by $\rho^*(\beta)$ (Ingster [10]; Donoho and Jin [5]). The NPLRT separates the null and the alternative asymptotically in the "detectable region", that is, $r > \rho^*(\beta)$ (Jiang and Zhang [12]).

The PLRT for contiguous alternative hypothesis was studied in Azaïs, Gassiat and Mercadier [1]. This is the dense case of (5.2), where $0 < \beta \leq 1/2$ under $p_n = n^{-\beta}$. In the dense case, μ_n is calibrated by $\mu_n = n^{-r}$. The detectable region is $r < \rho^*(\beta)$ where $\rho^*(\beta) = 1/2 - \beta$. When $n^{1/2}p_n\mu_n \to \gamma \in \mathbb{R}$ and $\mu_n \to \mu_0 \in \mathbb{R}$, the asymptotic power of the PLRT is equal to the asymptotic level (Azaïs, Gassiat and Mercadier [1]). This implies that in the case where $r = (1+o(1))\rho^*(\beta)$, the PLRT cannot distinguish the null from the alternative asymptotically. We provide a result which says the NPLRT is consistent in the interior of the detectable region.

Theorem 5. *Consider the testing problem (5.2) where $0 < \beta \le 1/2$ under $p_n = n^{-\beta}$. Let $\mu_n = n^{-r}$. Let $q(n,\alpha)$ be the critical value such that $\mathbb{P}_{H_0}\{\ell_n(\hat{f}_n, \varphi) > q(n,\alpha)\} = \alpha$. The rejection region is $\ell_n(\hat{f}_n, \varphi) > q(n,\alpha)$. Then*

$$\mathbb{P}_{H_1^{(n)}}\big\{\textit{NPLRT in (1.2) rejects } H_0\big\} \to 1, \qquad n \to \infty$$

if and only if $r < \rho^*(\beta)$ where $\rho^*(\beta) = 1/2 - \beta$.

If we let $\alpha \to 0$, then the sum of Type I and Type II errors tends to zero. So the NPLRT separates the null and the alternative asymptotically in the detectable region.

5.2. Simulations

We provide some simulation results to compare the NPLRT and the PLRT. Because the $\log\log n$ rate of divergence is very slow, the asymptotic distribution is not directly applicable in computing the critical values of the PLRT under the null. In our simulations, both the critical values of the PLRT and the NPLRT under the null are simulated.

We first considered (5.2). We set $(n, p_n) = (1000, 0.005)$. The corresponding sparse parameter under $p_n = n^{-\beta}$ is $\beta = 0.767$. We let the amplitude parameter μ_n range from 1.5 to 4 with an increment of 0.25. We set the significance level $\alpha = 0.05$. Table 1 displays the powers of the NPLRT and the PLRT based on 1000 replications. The PLRT is slightly better than the NPLRT. This is not surprising since the PLRT is the benchmark for (5.2).

We next considered testing H_0 against the Gaussian hierarchical model. Under the alternative, the non-standard normal observations are from $N(\mu_i, 1)$ where $\mu_i \sim N(0, \tau^2)$. We set $(n, p_n) = (1000, 0.005)$ and $(1000, 0.01)$. We let τ range from 1 to 4.5 with an increment of 0.5. The results are reported in Table 2. In this simulation, the NPLRT yields much stronger performance.

5.3. Other mixtures

The methods to derive the upper bound is applicable to multivariate Gaussian location mixtures. For independent bivariate normal distribution with unit variances, the Gaussian mixtures have the expansion

$$f_G(x, y) = \varphi(x, y) \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \frac{\mu_{j,k}(G)}{j!k!} H_j(x) H_k(y), \tag{5.3}$$

Table 1. Powers for (5.2) at level $\alpha = 0.05$. $(n, p_n) = (1000, 0.005)$

	μ_n										
Method	1.5	1.75	2	2.25	2.5	2.75	3	3.25	3.5	3.75	4
NPLRT	0.097	0.126	0.171	0.237	0.334	0.481	0.623	0.771	0.884	0.952	0.982
PLRT	0.096	0.127	0.189	0.279	0.398	0.535	0.681	0.817	0.919	0.968	0.990

Table 2. The nonnull means are sampled from $N(0, \tau^2)$. $n = 1000$, $\alpha = 0.05$

		τ							
p_n	Method	1	1.5	2	2.5	3	3.5	4	4.5
0.005	NPLRT	0.060	0.150	0.309	0.482	0.649	0.758	0.855	0.894
	PLRT	0.057	0.115	0.206	0.328	0.438	0.557	0.631	0.702
0.01	NPLRT	0.102	0.237	0.512	0.737	0.873	0.947	0.983	0.994
	PLRT	0.080	0.173	0.363	0.539	0.686	0.778	0.860	0.898

where $\varphi(x, y) = \exp(-x^2/2 - y^2/2)/(2\pi)$ and $\mu_{j,k}(G) = \int u^j v^k \, dG(u, v)$ is the (j, k)th moment of G. The generalized score function $h_G(x, y) \equiv (f_G(x, y)/\varphi(x, y) - 1)/\|f_G/\varphi - 1\|_2$ is expanded by

$$h_G(x, y) = \sum_{(j,k)\neq(0,0)} c_{j,k}(G) \frac{H_j(x)}{\sqrt{j!}} \frac{H_k(y)}{\sqrt{k!}}, \tag{5.4}$$

where $c_{j,k}(G) = \{\mu_{j,k}(G)/\sqrt{j!k!}\}/\{\sum_{(j,k)\neq(0,0)} \mu_{j,k}^2(G)/(j!k!)\}^{1/2}$. Let

$$F_{1,n}(x, y) = \left\{ \sum_{j=1}^{m_1} \sum_{k=1}^{m_2} \frac{1}{j!} \frac{1}{k!} H_j^2(x) H_k^2(y) \right\}^{1/2}. \tag{5.5}$$

By analysis parallel to Lemmas 1 and 2, it can be shown that the cutoffs are $m_1 \approx M_n^2$ and $m_2 \approx M_n^2$. Similar analysis can be carried out in general fixed dimension.

The methods is potentially applicable to Poisson mixtures $p_G(x) = \int z^x e^{-z}/x! \, dG(z)$. We test $H_0\colon z = z_0$. The generalized score function can be expanded by the Charlier polynomials (Azaïs, Gassiat and Mercadier [2]),

$$q_G(x) = \frac{p_G(x)/p_0(x) - 1}{\|p_G/p_0 - 1\|_2} = \sum_{k=1}^{\infty} b_k(G) C_k(x), \tag{5.6}$$

where $b_k(G) = \{m_k(G)/(k! z_0^k)\}/\{\sum_{j=1}^{\infty} m_j^2(G)/(j! z_0^j)\}^{1/2}$, $m_k(G) = E_G(z - z_0)^k$ and

$$C_k(x) = z_0^k \frac{d^k}{dz^k} \left(\frac{z}{z_0} \right)^x e^{-z+z_0} \bigg|_{z=z_0}.$$

6. Proofs

Proof of Theorem 1. First of all, $\sup_{f\in\mathcal{F}_n} \ell_n(f, \varphi) \le n\varepsilon_n^2$ follows directly from Theorem 1 of Jiang and Zhang [12].

Let $\mathscr{F}_n = \{f_G\colon G([-M_n, M_n]) = 1\}$. Since $\varphi \in \mathscr{F}_n$, we have $\ell_n(\hat{f}_n, \varphi) \geq 0 \geq -n\varepsilon_n^2$. By the proof of Theorem 1 of Zhang [19], the nonparametric MLE over $\mathscr{F}_n$ converges: $P\{d(\hat{f}_n, \varphi) \geq \varepsilon_n\} \to 0$, where $\varepsilon_n \asymp (\log n)/\sqrt{n}$. For any semi-distance d_0, let $\mathrm{Ball}(h_0, \eta, d_0) \equiv \{h\colon d_0(h, h_0) < \eta\}$ be the ball of radius η around h_0. We consider the η-cover of $\mathscr{F}_n \cap \{f : d(f, \varphi) \leq \varepsilon_n\}$ for $\eta = 1/n^3$. It follows from Lemma 2 of Zhang [19] that there exists $\{f_{G_j}, 1 \leq j \leq N_n\} \subset \mathscr{F}_n \cap \{f : d(f, \varphi) \leq \varepsilon_n\}$ with $\log N_n \leq n\varepsilon_n^2$ such that for $M_n \asymp \sqrt{\log n}$,

$$\mathscr{F}_n \cap \bigl\{f : d(f, \varphi) \leq \varepsilon_n\bigr\} \subset \bigcup_{j=1}^{N_n} \mathrm{Ball}\bigl(f_{G_j}, 1/n^3, \|\cdot\|_{\infty, M_n}\bigr),$$

where $\|h\|_{\infty,M} = \sup_{|x|\leq M} |h(x)|$ is the supreme norm in bounded intervals.

Proposition 2 of Jiang and Zhang [11] asserts that for all $i = 1, \dots, n$, $\hat{f}_n(X_i) \geq 1/(\sqrt{2\pi} en)$. Then, for f_{G_j} such that $\|\hat{f}_n - f_{G_j}\|_{\infty, M_n} \leq 1/n^3$, we have

$$\begin{aligned}
&\Bigl|\sup_{f\in\mathscr{F}_n} \ell_n(f, \varphi) - \ell_n(f_{G_j}, \varphi)\Bigr| \\
&\quad= \Biggl|\sum_{i=1}^n \bigl(\log f_{G_j}(X_i) - \log \varphi(X_i)\bigr) - \sum_{i=1}^n \bigl(\log \hat{f}_n(X_i) - \log \varphi(X_i)\bigr)\Biggr| \\
&\quad\leq \Biggl|\sum_{i=1}^n \log\biggl(1 + \frac{\|f_{G_j} - \hat{f}_n\|_{\infty, M_n}}{\hat{f}_n(X_i)}\biggr)\Biggr| \\
&\quad\leq \sqrt{2\pi} e/n.
\end{aligned}$$

This and $\sup_{f\in\mathscr{F}_n} \ell_n(f, \varphi) \geq \chi_1^2$ imply that $\sup_{f\in\mathscr{F}_n} \ell_n(f, \varphi) = (1 + \eta_n) \sup_{1\leq j\leq N_n} \ell_n(f_{G_j}, \varphi)$ with large probability, where $|\eta_n| \asymp 1/n$. □

Proof of Lemma 1. We use C_0 to denote a universal constant which may take different values from one appearance to another.

Let m_0 be an integer satisfying $2(m_0 - 1) < M_n^2 \leq 2m_0$. For $j \leq m_0$, $\mu_{2m_0}^2(G) \leq (M_n^2)^{2m_0 - 2j} \mu_{2j}^2(G)$. So

$$\sum_{j=1}^\infty \frac{\mu_j^2(G)}{j!} \geq \sum_{j=1}^{m_0} \frac{\mu_{2j}^2(G)}{(2j)!} \geq \frac{\mu_{2m_0}^2(G)}{(2m_0)!} \sum_{j=1}^{m_0} \frac{(2m_0)!}{(2j)!(M_n^2)^{2m_0-2j}}. \tag{6.1}$$

By Stirling's formula,

$$\begin{aligned}
\sum_{j=1}^{m_0} \frac{(2m_0)!}{(2j)!(M_n^2)^{2m_0-2j}} &= \frac{(2m_0)!}{M_n^{4m_0}} \sum_{j=1}^{m_0} \frac{(M_n^2)^{2j}}{(2j)!} \\
&\geq \frac{\sqrt{2\pi M_n^2}}{M_n^{4m_0}} \biggl(\frac{M_n^2}{e}\biggr)^{2m_0} \bigl(C_0 e^{M_n^2}\bigr) \geq \frac{M_n}{C_0}.
\end{aligned} \tag{6.2}$$

Since $\mu_k^2(G) \le (2m_0)^{k-2m_0}\mu_{2m_0}^2(G)$ for $k \ge 2m_0$, it follows from (6.1) and (6.2) that

$$c_k^2(G) \le \frac{\mu_k^2(G)/k!}{(M_n/C_0)\mu_{2m_0}^2(G)/(2m_0)!} \le \left(\frac{C_0}{M_n}\right)(2m_0)^{k-2m_0}(2m_0)!/k!.$$

Since $(2m_0)^{k-2m_0}(2m_0)!/k! = \prod_{j=1}^{k-2m_0}(1 - j/(2m_0+j)) \le e^{-(k-2m_0)^2/(2k)}$, (2.7) follows. □

Proof of Lemma 2. (i) First of all, $\|F_{1,n}\|_2^2 = m$ follows directly from $\int H_j(x)H_k(x)\varphi(x)\,dx = k!I\{j=k\}$. We include a proof since it leads to the proof of (2.8). Due to

$$\begin{aligned} H_k(x) &= (-1)^k e^{x^2/2}\frac{d^k}{dt^k}e^{-t^2/2}\bigg|_{t=x} \\ &= e^{x^2/2}\frac{d^k}{du^k}e^{-(x-u)^2/2}\bigg|_{u=0} \\ &= \frac{d^k}{du^k}e^{xu-u^2/2}\bigg|_{u=0}, \end{aligned}$$

the second moment of the Hermite polynomials can be computed via

$$\begin{aligned} \int H_j(x)H_k(x)\varphi(x)\,dx &= \frac{\partial^j}{\partial u^j}\frac{\partial^k}{\partial v^k}\int e^{xu-u^2/2+xv-v^2/2}\varphi(x)\,dx\bigg|_{u=v=0} \\ &= \frac{\partial^j}{\partial u^j}\frac{\partial^k}{\partial v^k}e^{uv}\bigg|_{u=v=0} \\ &= k!I\{j=k\}. \end{aligned}$$

This gives the orthogonality. Similarly, the fourth moment of the Hermite polynomials can be computed via

$$\begin{aligned} &\int \prod_{j=1}^4 H_{k_j}(x)\varphi(x)\,dx \\ &\quad = \left(\prod_{j=1}^4 \frac{\partial^{k_j}}{\partial u_j^{k_j}}\right)\int \exp\left\{\sum_{1\le j\le 4}\left(xu_j - u_j^2/2\right)\right\}\varphi(x)\,dx\bigg|_{u_1=u_2=u_3=u_4=0} \\ &\quad = \left(\prod_{j=1}^4 \frac{\partial^{k_j}}{\partial u_j^{k_j}}\right)\exp\{u_1u_2+u_1u_3+u_1u_4+u_2u_3+u_2u_4+u_3u_4\}\bigg|_{u_1=u_2=u_3=u_4=0}. \end{aligned} \tag{6.3}$$

Let $m = (k_1+k_2+k_3+k_4)/2$. Assume that m is an integer as the right-hand side above is zero otherwise. Let $k_1 \ge k_2 \ge k_3 \ge k_4$. We have $k_4 + k_3 \le m$ and

$$\int \prod_{j=1}^4 H_{k_j}(x)\varphi(x)\,dx$$

$$
\begin{aligned}
&= \frac{1}{m!}\left(\prod_{j=1}^{4}\frac{\partial^{k_j}}{\partial u_j^{k_j}}\right)\left(\sum_{1\le j_1<j_2\le 4} u_{j_1}u_{j_2}\right)^m\Bigg|_{u_1=u_2=u_3=u_4=0} \\
&= \frac{k_4!}{m!}\binom{m}{k_4}\left(\prod_{j=1}^{3}\frac{\partial^{k_j}}{\partial u_j^{k_j}}\right)\left(\sum_{1\le j\le 3} u_j\right)^{k_4}\left(\sum_{1\le j_1<j_2\le 3} u_{j_1}u_{j_2}\right)^{m-k_4}\Bigg|_{u_1=u_2=u_3=0} \\
&= \sum_{l=0}^{k_4}\frac{k_3!}{(m-k_4)!}\binom{k_4}{l}\binom{m-k_4}{k_3-l}\left(\prod_{j=1}^{2}\frac{\partial^{k_j}}{\partial u_j^{k_j}}\right)(u_1+u_2)^{k_4-l+(k_3-l)}(u_1u_2)^{m-k_4-(k_3-l)}\Bigg|_{u_1=u_2=0} \\
&= \sum_{l=0}^{k_4}\frac{k_3!}{(m-k_4)!}\binom{k_4}{l}\binom{m-k_4}{k_3-l}\binom{k_4+k_3-2l}{k_2-(m-k_4-k_3+l)}k_2!k_1! \\
&= \sum_{l=0}^{k_4}\frac{k_1!k_2!k_3!k_4!(k_4+k_3-2l)!}{l!(k_4-l)!(k_3-l)!(m-k_3-k_4+l)!(m-k_2-l)!(m-k_1-l)!}. \qquad (6.4)
\end{aligned}
$$

The above quantity is counted as zero when $m < k_1 + l$. In particular, for $k_4 = k_3 = k_2 = k_1$,

$$
\begin{aligned}
\int H_k^4(x)\varphi(x)\,dx &= (k!)^2\sum_{l=0}^{k}\binom{k}{l}^2\binom{2k-2l}{k-l} \\
&= (k!)^2\sum_{l=0}^{k}\binom{k}{k-l}^2\binom{2k-2l}{k-l} \\
&= (k!)^2\sum_{l=0}^{k}\binom{k}{l}^2\binom{2l}{l} \\
&\le (k!)^2\sum_{l=0}^{k}\binom{k}{l}^2 2^{2l} \le (k!)^2\left\{\sum_{l=0}^{k}\binom{k}{l}2^l\right\}^2 = (k!)^2 9^k. \qquad (6.5)
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\|F_{1,n}\|_4^4 &= \int\left\{\sum_{k=1}^{m}\left(\frac{H_k(x)}{\sqrt{k!}}\right)^2\right\}^2\varphi(x)\,dx \\
&= \int\sum_{k=1}^{m}\left(\frac{H_k(x)}{\sqrt{k!}}\right)^4\varphi(x)\,dx = \sum_{k=1}^{m}\left\|\frac{H_k}{\sqrt{k!}}\right\|_4^4 \le \sum_{k=1}^{m}9^k \le 9^{m+1}/8.
\end{aligned}
$$

Since $\binom{2l}{l} \asymp 2^{2l}/\sqrt{l}$, the base 9 in (6.5) is tight.

(ii) Let $y = (x - M_n^2)/\sqrt{x}$. For $x > m$, we have

$$
\begin{aligned}
dx &= \frac{2\,dy}{x^{-1/2} + M_n^2 x^{-3/2}} = \frac{(2y + 2M_n^2/x^{1/2})\,dy}{1 + M_n^2/x} \\
&\le \left(2y + \frac{2M_n^2/x^{1/2}}{1 + M_n^2/x}\right) dy \le (2y + M_n)\,dy.
\end{aligned}
\tag{6.6}
$$

When $x \ge m$, $y = (x - M_n^2)/\sqrt{x} \ge (m - M_n^2)/\sqrt{m} \ge 2t_n$. By the triangle inequality and (6.6),

$$
\begin{aligned}
\|F_{2,n}\|_2 &\le \sqrt{\frac{C_0}{M_n}} \sum_{k=m+1}^{\infty} e^{-(k-M_n^2)^2/(4k)} \\
&\le \sqrt{C_0/M_n} \int_m^\infty e^{-(x-M_n^2)^2/(4x)}\,dx \\
&\le \sqrt{C_0/M_n} \int_{2t_n}^\infty e^{-y^2/4}(2y + M_n)\,dy \\
&\le \sqrt{C_0/M_n}\left(4e^{-t_n^2} + M_n e^{-t_n^2} \int_0^\infty e^{-y^2/4}\,dy\right) = C_1 M_n^{1/2} e^{-t_n^2}.
\end{aligned}
$$

This completes the proof of (2.9). □

Proof of Lemma 3. Since

$$
\|\sqrt{f/\varphi} - 1\|_2^2 = \int (\sqrt{f/\varphi} - 1)^2 \varphi\,dx = \int (\sqrt{f} - \sqrt{\varphi})^2\,dx = 2d^2(f,\varphi),
\tag{6.7}
$$

we have

$$
\sup_{f\in\mathcal{F}_n, d(f,\varphi)\le\varepsilon_n} \left| 1 - \frac{2\sqrt{2}d(f,\varphi)}{\|f/\varphi - 1\|_2} \right| \le \sup_{f\in\mathcal{F}_n, d(f,\varphi)\le\varepsilon_n} \left\| \frac{f/\varphi - 1}{\|f/\varphi - 1\|_2} - \frac{2(\sqrt{f/\varphi} - 1)}{\|f/\varphi - 1\|_2} \right\|_2
\tag{6.8}
$$

by the triangle inequality. Let $h = (f/\varphi - 1)/\|f/\varphi - 1\|_2$. Then, (3.8), (6.7) and $\varepsilon_n C_n \to 0$ imply that

$$
\begin{aligned}
&\sup_{f\in\mathcal{F}_n, d(f,\varphi)\le\varepsilon_n} \left\| \frac{f/\varphi - 1}{\|f/\varphi - 1\|_2} - \frac{2(\sqrt{f/\varphi} - 1)}{\|f/\varphi - 1\|_2} \right\|_2 \\
&= \sup_{f\in\mathcal{F}_n, d(f,\varphi)\le\varepsilon_n} \left\| \frac{\sqrt{f/\varphi} - 1}{\sqrt{f/\varphi} + 1} h \right\|_2 \\
&= \sup_{f\in\mathcal{F}_n, d(f,\varphi)\le\varepsilon_n} \left\| \frac{\sqrt{f/\varphi} - 1}{\sqrt{f/\varphi} + 1} \big(|h| - C_n + C_n\big) \right\|_2 \\
&\le \sup_{f\in\mathcal{F}_n, d(f,\varphi)\le\varepsilon_n} \left\{ \big\| \big(|h| - C_n\big)_+ \big\|_2 + \sqrt{2}\varepsilon_n C_n \right\} \to 0.
\end{aligned}
$$

It remains to verify (3.8) for the specific M_n. For bounded M_n, (3.8) follows from the fact that $F_{1,n} + F_{2,n}$ for some fixed m is an L_2 envelope function. Assume $M_n \to \infty$. Let $C_n = 1/(\varepsilon_n \log n)$. Since $(F_{1,n} - C_n)_+ \le F_{1,n}^2/C_n$, by Lemma 2, when $(m - M_n^2)/\sqrt{m} \ge 2t_n = 2\sqrt{\log M_n}$,

$$\begin{aligned}
\sup_{f\in\mathscr{F}_n, d(f,\varphi)\le\varepsilon_n} \left\| \left(\frac{|f/\varphi - 1|}{\|f/\varphi - 1\|_2} - C_n \right)_+ \right\|_2 &\le \left\|(F_{1,n} - C_n)_+\right\|_2 + \|F_{2,n}\|_2 \\
&\le \|F_{1,n}\|_4^2/C_n + o(1) \\
&\le O(1)3^m/C_n + o(1).
\end{aligned}$$

Thus, when $M_n^2 \le a_0 \log n$ with $a_0 < 1/\log 9$,

$$\begin{aligned}
\log\bigl(3^{2m}/C_n^2\bigr) &= \log\bigl((\log n)^2 \varepsilon_n^2 9^m\bigr) \\
&\le \log\bigl((\log n)^4 9^m/n\bigr) + O(1) \\
&\le (\log n)\bigl(\bigl(1 + o(1)\bigr)a_0 \log 9 - 1\bigr) + O(\log\log n) \\
&\to -\infty.
\end{aligned} \tag{6.9}$$

This completes the proof of Lemma 3. □

Proof of Theorem 5. Let $G^{(n)} = (1 - p_n)\delta_0 + p_n\delta_{\mu_n}$ where δ_μ is the probability distribution giving its entire mass to μ. Using Theorem 2 of Jiang and Zhang [12], it remains to prove that $nd^2(\varphi, f_{G^{(n)}})/(\log n)^2 \to \infty$ if and only if $r < 1/2 - \beta$. Note that $f_{G^{(n)}}(x) = (1 - p_n)\varphi(x) + p_n\varphi(x - \mu_n)$. We divide the Hellinger distance into two parts:

$$\begin{aligned}
2d^2(\varphi, f_{G^{(n)}}) &= \int \Bigl(\sqrt{\varphi(x)} - \sqrt{(1 - p_n)\varphi(x) + p_n\varphi(x - \mu_n)}\Bigr)^2 dx \\
&= \int \Bigl(1 - \sqrt{1 - n^{-\beta} + n^{-\beta}\exp\bigl(x\mu_n - \mu_n^2/2\bigr)}\Bigr)^2 \varphi(x)\, dx \\
&= \int_{-\infty}^{\frac{\beta}{\mu_n}\log n + \frac{\mu_n}{2}} + \int_{\frac{\beta}{\mu_n}\log n + \frac{\mu_n}{2}}^{+\infty} \\
&\triangleq I_1 + I_2.
\end{aligned} \tag{6.10}$$

We first calculate I_1. When $x < (\beta/\mu_n)\log n + \mu_n/2$, the Taylor series gives that

$$\sqrt{1 - n^{-\beta} + n^{-\beta}\exp\bigl(x\mu_n - \mu_n^2/2\bigr)} = 1 - \frac{1}{2}\bigl(1 + o(1)\bigr)n^{-\beta}\left\{1 - \exp\left(x\mu_n - \frac{\mu_n^2}{2}\right)\right\}.$$

Then,

$$I_1 = \bigl(1 + o(1)\bigr)\int_{-\infty}^{\frac{\beta}{\mu_n}\log n + \frac{\mu_n}{2}} \frac{1}{4}n^{-2\beta}\left\{1 - \exp\left(x\mu_n - \frac{\mu_n^2}{2}\right)\right\}^2 \varphi(x)\, dx$$

$$= \big(1+o(1)\big)\frac{1}{4}n^{-2\beta}\int_{-\infty}^{\frac{\beta}{\mu_n}\log n+\frac{\mu_n}{2}}\bigg\{1-2\exp\bigg(x\mu_n-\frac{\mu_n^2}{2}\bigg)+\exp\big(2x\mu_n-\mu_n^2\big)\bigg\}\varphi(x)\,dx$$
$$\triangleq \big(1+o(1)\big)(I_3+I_4+I_5). \tag{6.11}$$

Here

$$I_3=\frac{1}{4}n^{-2\beta}\Phi\bigg(\frac{\beta}{\mu_n}\log n+\frac{\mu_n}{2}\bigg).$$

The cross term is

$$\begin{aligned} I_4 &\equiv -\frac{1}{2}n^{-2\beta}\int_{-\infty}^{\frac{\beta}{\mu_n}\log n+\frac{\mu_n}{2}}\exp\bigg(x\mu_n-\frac{\mu_n^2}{2}\bigg)\varphi(x)\,dx \\ &= -\frac{1}{2}n^{-2\beta}\Phi\bigg(\frac{\beta}{\mu_n}\log n-\frac{\mu_n}{2}\bigg). \end{aligned}$$

Some calculations show that

$$\begin{aligned} I_5 &\equiv \frac{1}{4}n^{-2\beta}\int_{-\infty}^{\frac{\beta}{\mu_n}\log n+\frac{\mu_n}{2}}\exp\big(2x\mu_n-\mu_n^2\big)\varphi(x)\,dx \\ &= \frac{1}{4}n^{-2\beta}\exp\big(n^{-2r}\big)\Phi\bigg(\frac{\beta}{\mu_n}\log n-\frac{3}{2}\mu_n\bigg). \end{aligned}$$

Putting I_3, I_4 and I_5 together, we obtain

$$\begin{aligned} &I_3+I_4+I_5 \\ &\quad = \frac{1}{4}n^{-2\beta}\bigg\{\Phi\bigg(\frac{\beta}{\mu_n}\log n+\frac{\mu_n}{2}\bigg)-2\Phi\bigg(\frac{\beta}{\mu_n}\log n-\frac{\mu_n}{2}\bigg)+\exp\big(n^{-2r}\big)\Phi\bigg(\frac{\beta}{\mu_n}\log n-\frac{3}{2}\mu_n\bigg)\bigg\} \\ &\quad = \frac{1}{4}\big(1+o(1)\big)n^{-2\beta}\big(\exp\big(n^{-2r}\big)-1\big) \\ &\quad = \frac{1}{4}\big(1+o(1)\big)n^{-2r-2\beta}. \end{aligned}$$

This and (6.11) give that

$$I_1 \gg (\log n)^2/n \quad \text{iff} \quad r<\frac{1}{2}-\beta. \tag{6.12}$$

The analysis I_2 is easier. When $x \ge (\beta/\mu_n)\log n+\mu_n/2$, the main term in the square root is $n^{-\beta}\exp(x\mu_n-\mu_n^2/2)$. So

$$I_2 \equiv \int_{\frac{\beta}{\mu_n}\log n+\frac{\mu_n}{2}}^{+\infty}\Big(1-\sqrt{1-n^{-\beta}+n^{-\beta}\exp\big(x\mu_n-\mu_n^2/2\big)}\Big)^2\varphi(x)\,dx$$

$$
\begin{aligned}
&= O(1)n^{-\beta} \int_{\frac{\beta}{\mu_n}\log n + \frac{\mu_n}{2}}^{+\infty} \exp\left(x\mu_n - \frac{\mu_n^2}{2}\right)\varphi(x)\,dx \\
&= O(1)n^{-\beta}\Phi\left(n^{-r}/2 - \beta n^r \log n\right) \\
&\ll (\log n)^2/n. \qquad (6.13)
\end{aligned}
$$

The last step is due to $\Phi(-x) = (1 + o(1))\varphi(x)/x$.

Combining $2d^2(\varphi, f_{G^{(n)}}) = I_1 + I_2$, (6.12) and (6.13), we have that $d^2(\varphi, f_{G^{(n)}}) \gg (\log n)^2/n$ if and only if $r < 1/2 - \beta$. □

Acknowledgements

The authors are grateful to the Associate Editor and a referee for helpful comments and suggestions, which helped to improve an earlier version of the paper.

References

[1] Azaïs, J.-M., Gassiat, É. and Mercadier, C. (2006). Asymptotic distribution and local power of the log-likelihood ratio test for mixtures: Bounded and unbounded cases. *Bernoulli* **12** 775–799. MR2265342

[2] Azaïs, J.-M., Gassiat, É. and Mercadier, C. (2009). The likelihood ratio test for general mixture models with or without structural parameter. *ESAIM Probab. Stat.* **13** 301–327. MR2528086

[3] 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* (J.K. Ghosh, S.K. Mitra, K.R. Parthasarathy and B.L.S. Prakasa Rao, eds.) 83–96. New Delhi: Wiley Eastern.

[4] DasGupta, A. (2008). *Asymptotic Theory of Statistics and Probability. Springer Texts in Statistics*. New York: Springer. MR2664452

[5] Donoho, D. and Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994. MR2065195

[6] Genovese, C.R. and Wasserman, L. (2000). Rates of convergence for the Gaussian mixture sieve. *Ann. Statist.* **28** 1105–1127. MR1810921

[7] Ghosal, S. and van der Vaart, A.W. (2001). Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *Ann. Statist.* **29** 1233–1263. MR1873329

[8] Gu, J., Koenker, R. and Volgushev, S. (2018). Testing for homogeneity in mixture models. *Econometric Theory* **34** 850–895. MR3815469

[9] Hartigan, J.A. (1985). A failure of likelihood asymptotics for normal mixtures. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, Vol. II* (*Berkeley, Calif.*, 1983). *Wadsworth Statist./Probab. Ser.* 807–810. Belmont, CA: Wadsworth. MR0822066

[10] Ingster, Yu.I. (1999). Minimax detection of a signal for l^n-balls. *Math. Methods Statist.* **7** 401–428. MR1680087

[11] Jiang, W. and Zhang, C.-H. (2009). General maximum likelihood empirical Bayes estimation of normal means. *Ann. Statist.* **37** 1647–1684. MR2533467

[12] Jiang, W. and Zhang, C.-H. (2016). Generalized likelihood ratio test for normal mixtures. *Statist. Sinica* **26** 955–978. MR3559938

[13] Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Stat.* **27** 887–906. MR0086464
[14] Lindsay, B.G. (1995). *Mixture Models: Theory, Geometry, and Applications*. Hayward, CA: IMS.
[15] Liu, X. and Shao, Y. (2003). Asymptotics for likelihood ratio tests under loss of identifiability. *Ann. Statist.* **31** 807–832. MR1994731
[16] Liu, X. and Shao, Y. (2004). Asymptotics for the likelihood ratio test in a two-component normal mixture model. *J. Statist. Plann. Inference* **123** 61–81. MR2058122
[17] Robbins, H. (1950). A generalization of the method of maximum likelihood: Estimating a mixing distribution (abstract). *Ann. Math. Stat.* **21** 314–315.
[18] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics. Springer Series in Statistics*. New York: Springer. MR1385671
[19] Zhang, C.-H. (2009). Generalized maximum likelihood estimation of normal mixture densities. *Statist. Sinica* **19** 1297–1318. MR2536157

Received July 2017 and revised November 2018