Bernoulli **20**(4), 2014, 1802–1818
DOI: 10.3150/13-BEJ542

Minimax bounds for estimation of normal mixtures

ARLENE K.H. KIM

Statistical Laboratory, Center for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge, CB30WB, UK. E-mail: a.kim@statslab.cam.ac.uk

This paper deals with minimax rates of convergence for estimation of density functions on the real line. The densities are assumed to be location mixtures of normals, a global regularity requirement that creates subtle difficulties for the application of standard minimax lower bound methods. Using novel Fourier and Hermite polynomial techniques, we determine the minimax optimal rate – slightly larger than the parametric rate – under squared error loss. For Hellinger loss, we provide a minimax lower bound using ideas modified from the squared error loss case.

Keywords: Assouad's lemma; Hermite polynomials; minimax lower bound; normal location mixture

1. Introduction

This paper establishes the optimal minimax rate of convergence under squared error loss, for densities that are normal mixtures. The analysis reveals a subtle difficulty in the application of Assouad's lemma to parameter spaces defined by indirect regularity conditions, which complicate the usual construction of subsets of the parameter space indexed by "hyper-rectangles."

More precisely, we consider independent observations from probability distributions P_f on the real line whose densities f (with respect to Lebesgue measure on $\mathbb{R}$) belong to the set of convolutions

$$\mathcal{F}=\left\{f\colon\ f(x)=\phi\star\Pi(x)=\int\phi(x-u)\,\mathrm{d}\Pi(u),\,\Pi\in\mathcal{P}(\mathbb{R})\right\},$$

where ϕ denotes the standard normal $N(0,1)$ density and $\mathcal{P}(\mathbb{R})$ denotes the set of all probability measures on the (Borel sigma-field of the) real line. Our main result gives an asymptotic minimax lower bound for the L_2 risk of estimators of $f\in\mathcal{F}$.

Theorem 1.1. *Let $X_1,\ldots,X_n$ be independent and identically distributed with density $f\in\mathcal{F}$. Then there exists a positive constant c such that*

$$\sup_{f\in\mathcal{F}}\mathbb{E}_{n,f}\int_{-\infty}^{\infty}\bigl(\hat{f}_n(x)-f(x)\bigr)^2\,\mathrm{d}x\geq c\cdot\log n\cdot\frac{1}{n\sqrt{\log n}}:=c\ell_n$$

for every estimator $\hat{f}_n=\hat{f}_n(X_1,\ldots,X_n)$.

Let $\mathcal{F}_0$ denote the subset of $\mathcal{F}$ consisting of those normal mixture densities whose mixing measure is absolutely continuous with respect to Lebesgue measure. The proof of Theorem 1.1, which is given in Section 2, involves the construction of a finite subset of $\mathcal{F}_0$, so the lower bound also holds when the supremum is taken over $f \in \mathcal{F}_0$. Perhaps the most interesting feature of this result is that the same rate has been obtained as an upper bound for the minimax risk with respect to squared error loss over much larger classes of functions. For instance, [4] defined the class $\mathcal{F}^*$ consisting of those densities that can be extended to an entire function f^* on $\mathbb{C}$ satisfying $\sup_{y\in\mathbb{R}} \mathrm{e}^{-y^2/2} \sup_{x\in\mathbb{R}} |f^*(x+\mathrm{i}y)| < \infty$. He proved the following theorem.

Theorem 1.2 ([4], Theorem 4.1). *Let $X_1, \ldots, X_n$ be independent and identically distributed with density $f \in \mathcal{F}^*$. Then there exists an estimator $\hat{f}_n = \hat{f}_n(X_1, \ldots, X_n)$ of f such that*

$$\sup_{f\in\mathcal{F}^*} \mathbb{E}_{n,f} \int_{-\infty}^{\infty} \bigl(\hat{f}_n(x) - f(x)\bigr)^2 \,\mathrm{d}x = \mathrm{O}(\ell_n).$$

For the reader's convenience, in Section 3, we show that $\mathcal{F} \subseteq \mathcal{F}^*$ and summarize Ibragimov's proof. Theorems 1.1 and 1.2 together establish that the minimax optimal rate of estimation for squared L_2 loss is ℓ_n for any class of functions containing $\mathcal{F}_0$ and contained in $\mathcal{F}^*$. In particular, this is the case for $\mathcal{F}$.

While a minimax result under the L_2 loss presents the most successful case, this loss function is often criticized for giving too little weight to errors from the tails. As an alternative, we also consider the Hellinger loss. Define a class of probability measures with sub-Gaussian tails,

$$\mathcal{P}_s(\mathbb{R}) := \bigl\{\Pi \in \mathcal{P}(\mathbb{R})\colon\ \exists C > 0 \text{ such that } \Pi\bigl(|u| > t\bigr) \le C\exp\bigl(-t^2/C\bigr) \text{ for all real } t\bigr\}.$$

For the following class of normal location mixtures

$$\mathcal{F}_s := \left\{f\colon\ f(x) = \phi \star \Pi(x) = \int \phi(x-u)\,\mathrm{d}\Pi(u),\ \Pi \in \mathcal{P}_s(\mathbb{R})\right\},$$

[2] provide a sieved maximum likelihood estimator whose convergence rate is $\mathrm{O}((\log n)^2/n)$. However, as they pointed out, the optimal rate for $\mathcal{F}_s$ is still unknown. Our technique gives a lower bound that lies within a logarithmic factor of Ghosal and van der Vaart's upper bound.

Theorem 1.3. *Let $X_1, \ldots, X_n$ be independent and identically distributed with density $f \in \mathcal{F}_s$. Then there exists a positive constant c such that*

$$\sup_{f\in\mathcal{F}_s} \mathbb{E}_{n,f} \int_{-\infty}^{\infty} \Bigl(\sqrt{\hat{f}_n(x)} - \sqrt{f(x)}\Bigr)^2 \,\mathrm{d}x \ge c \cdot \log n \cdot \frac{1}{n}$$

for every estimator $\hat{f}_n = \hat{f}_n(X_1, \ldots, X_n)$.

To prove Theorems 1.1 and 1.3, we use a variation on Assouad's lemma (cf. [11], page 347). When specialized to density estimation, the lemma can be cast into the following form. (Henceforth, we omit the $\pm\infty$ terminals on the integrals when there is no ambiguity.) For completeness, we provide the proof in the Appendix.

Lemma 1.4. *Let* $\{f_\alpha, \alpha \in \{0,1\}^K\} \subseteq \mathcal{F}$ *where* K *is a finite index set of cardinality* m. *Suppose* W *is a nonnegative loss function for which there exists* $\zeta > 0$ *such that, for all* $g_1, g_2 \in \mathcal{F}$,

$$\inf_{f \in \mathcal{F}} W(f, g_1) + W(f, g_2) \geq \zeta W(g_1, g_2). \tag{1}$$

Suppose also that for some constants $c_0 > 0$ *and* $1 > c_1 > 0$,

$$W(f_\alpha, f_\beta) \geq c_0 \varepsilon^2 \|\alpha - \beta\|_0 \qquad \text{for all } \alpha, \beta \in \{0,1\}^K \tag{2}$$

and

$$\int \frac{(f_\alpha - f_\beta)^2}{f_\alpha} \leq \frac{c_1}{n} \qquad \text{if } \|\alpha - \beta\|_0 = 1, \tag{3}$$

where $\|\alpha - \beta\|_0 = \sum_{k \in K} \mathbb{1}\{\alpha_k \neq \beta_k\}$, *the Hamming distance. Then, for every estimator* $\hat{f}_n$ *based on* n *independent observations*,

$$\sup_{f \in \mathcal{F}} \mathbb{E}_{n,f} W(\hat{f}_n, f) \geq \frac{c_0 \zeta}{4} (1 - \sqrt{c_1}) m \varepsilon^2. \tag{4}$$

Remark 1.1. Assumption (3) regarding the χ^2 distance is merely a convenient way to show that the testing affinity, $\|P^n_{f_\alpha} \wedge P^n_{f_\beta}\|_1$, is at least $1 - \sqrt{c_1}$, where P^n_f is a product probability measure under f and $\|P \wedge Q\|_1$ is defined as $\int \min(\mathrm{d}P, \mathrm{d}Q)$.

Remark 1.2. To apply Lemma 1.4, we try to maximize $m\varepsilon^2$ for the best possible lower bound. While we construct the finite density class satisfying the loss separation condition (2), we need to restrict the size ε^2 and m so that two nearest densities should be reasonably close as in (3), and so that the constructed densities are truly in the parameter space $\mathcal{F}$.

For the proof in Section 2, we construct f_α's of the form

$$f_\alpha(x) = f_0(x) + \varepsilon \sum_{k \in K} \alpha_k \Delta_k(x), \qquad \alpha \in \{0,1\}^K,$$

where f_0 is the normal density function with a zero mean (and variance specified later), where $K = \{1, 3, \ldots, 2m-1\}$, and where m, $\varepsilon > 0$, and Δ_k could depend on n. The main difficulty lies in choosing the (signed) perturbations Δ_k so that each f_α is a normal location mixture. The natural way around this problem is to construct the Assouad hyper-rectangle in the space of mixing distributions,

$$f_\alpha = \phi \star \Pi_\alpha, \qquad \text{where } \Pi_\alpha(u) = \Pi_0(u) + \varepsilon \sum_{k \in K} \alpha_k V_k(u), \alpha \in \{0,1\}^K,$$

where the signed measures V_k must be chosen so that each Π_α is a probability measure. In contrast to the standard construction, the indirect form of $f_\alpha = \phi \star \Pi_\alpha$ leads to an embedding condition of the form

$$W(\phi \star \Pi_\alpha, \phi \star \Pi_\beta) \geq \tau_n \sum_{k\in K} (\alpha_k - \beta_k)^2 \tag{5}$$

for some τ_n. The right side of (5) is expressed in terms of $\sum_{k\in K}(\alpha_k - \beta_k)^2$ instead of the Hamming distance, in order to emphasize the orthogonality relation. If the convolution with the normal density were not present, such a property could be obtained by choosing the perturbations to be exactly orthogonal to each other, subject to various other regularity properties that define the parameter space. The smoothing effect of the convolution operation, however, makes it difficult to choose the V_k to achieve such near-orthogonality. Nevertheless, we can achieve (5) by choosing the perturbations so that their Fourier transforms are orthogonal as elements in $L_2(\phi^2)$, the space of complex-valued functions g such that $\int \phi(x)^2 |g(x)|^2 \, \mathrm{d}x < \infty$ for the L_2 loss. Similarly, we achieve (5) under the Hellinger loss using the similar ideas under L_2 except that ϕ^2 is replaced by a different weight function.

2. Proofs of the lower bounds

First, we introduce some notation used in this section. We let ϕ_{σ^2} be the normal density with mean zero and variance σ^2. Following, for example, [9], Chapter 9, we define the Fourier transform $\mathcal{T}$ by

$$\mathcal{T} f(t) := \check{f}(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-\mathrm{i}xt) f(x) \, \mathrm{d}x$$

for $f \in \mathcal{L}_1(\lambda)$ where λ is Lebesgue measure, and then extend from $\mathcal{L}_1 \cap \mathcal{L}_2$ to $\mathcal{L}_2$ by extending an isometry of $\mathcal{L}_1 \cap \mathcal{L}_2$ into $\mathcal{L}_2$ to an isometry of $\mathcal{L}_2$ onto $\mathcal{L}_2$.

For both theorems, we construct the signed measures V_k to have (signed) densities v_k with respect to λ:

$$\pi_\alpha(u) = \frac{\mathrm{d}\Pi_\alpha}{\mathrm{d}\lambda}(u) = \pi_0(u) + \varepsilon \sum_{k\in K} \alpha_k v_k(u), \qquad \alpha \in \{0,1\}^K, \tag{6}$$

where π_0 is the normal density with zero mean and each v_k is a function for which $\int v_k = 0$ and

$$\pi_0(u) + \varepsilon \sum_{k\in K} \alpha_k v_k(u) \geq 0 \qquad \text{for all } u.$$

We then need to check the assumptions for Lemma 1.4.

2.1. Ideas in the proof of Theorem 1.1

Here we let $W(f,g) := \|f - g\|_2^2 = \int (f-g)^2$, so (1) is satisfied with $\zeta = 1/2$. The choice of the v_k's is suggested by Fourier methods. By the Plancherel formula (and the fact that $\check{\phi} = \phi$), recalling that $f_\alpha = \phi \star \Pi_\alpha$,

$$\frac{1}{2\pi}\|f_\alpha - f_\beta\|_2^2 = \frac{1}{2\pi}\|\breve{f}_\alpha - \breve{f}_\beta\|_2^2 = \varepsilon^2 \int_{-\infty}^{\infty}\left|\sum_{k\in K}(\alpha_k - \beta_k)\phi(t)\breve{v}_k(t)\right|^2 \mathrm{d}t,$$

which lets us write the desired property (2) of Lemma 1.4 as

$$\int_{-\infty}^{\infty}\left|\sum_{k\in K}(\alpha_k - \beta_k)\phi(t)\breve{v}_k(t)\right|^2 \mathrm{d}t \ge \frac{c_0}{2\pi}\sum_{k\in K}(\alpha_k - \beta_k)^2 \qquad \forall \alpha, \beta \in \{0,1\}^K.$$

We might achieve such an inequality by choosing the v_k's to make the functions $\psi_k(t) := \phi(t)\breve{v}_k(t)$ orthogonal. Ignoring other requirements for the moment, we could even start from an orthonormal set $\{\psi_k\}$ and then try to define v_k as the (inverse) Fourier transform of $\psi_k(t)/\phi(t)$, provided that the ratio is square integrable. This heuristic succeeds if we start from the normalized orthogonal functions (see [5], Chapter 9),

$$\psi_k(t) = C\mathrm{i}^{-k}\phi(t)^2\frac{H_k(2t)}{\sqrt{k!}} = \mathrm{i}^{-k}\sqrt{2\phi(2t)}\frac{H_k(2t)}{\sqrt{k!}} \tag{7}$$

for $k \in K := \{1, 3, \ldots, 2m-1\}$, where $C = \sqrt{2}(2\pi)^{3/4}$ is chosen so that $C\phi(t)^2 = \sqrt{2\phi(2t)}$ and $H_k(t)$ is the Hermite polynomial of order k, the polynomial for which $\phi(t)$ has kth derivative $(-1)^k H_k(t)\phi(t)$.

Remark 2.1. $\{H_k, k = 1, 2, \ldots\}$ is sometimes called the "probabilists' Hermite Polynomials" (denoted as "*He*" in [3]), as opposed to the "physicists' Hermite Polynomials" $\mathbf{H}$. There is one-to-one relation between H and $\mathbf{H}$, given by

$$H_k(t) = 2^{-k/2}\mathbf{H}_k\left(\frac{t}{\sqrt{2}}\right).$$

To calculate the Fourier inverse transform of $\psi_k(t)/\phi(t)$, we provide the following lemma.

Lemma 2.1. *For $b > a > 0$,*

$$\mathcal{T}^{-1}\bigl[\phi(at)H_k(bt)\bigr](u) = Q_k\phi\left(\frac{u}{a}\right)H_k\bigl(b'u\bigr), \tag{8}$$

where $Q_k = (\mathrm{i}c_{a,b})^k/a$ with $c_{a,b} = \sqrt{b^2/a^2 - 1}$ and $b' = b/(a^2 c_{a,b})$.

Remark 2.2. Lemma 2.1 illustrates a general form of the eigenvalue-eigenfunction relation for the Fourier transform of Hermite functions,

$$\mathcal{T}\bigl[\phi(t)H_k(\sqrt{2}t)\bigr](u) = (-\mathrm{i})^k\phi(u)H_k(\sqrt{2}u).$$

(See (7.376) in [3], or for more details, see Section 4.11 in [6]).

We now formulate these arguments into a proof.

Proof of Theorem 1.1. By Lemma 2.1, defining $\{\psi_k, k \in K\}$ as in (7) leads to

$$v_k(u) = C\sqrt{\frac{3^k}{k!}}\phi(u)H_k\left(\frac{2}{\sqrt{3}}u\right) \qquad \text{for } k \in K, \tag{9}$$

because $\mathcal{T}^{-1}[\phi(t)H_k(2t)](u) = \mathrm{i}^k 3^{k/2}\phi(u)H_k(2u/\sqrt{3})$. By restricting to odd values of k, we make the v_k's real-valued and odd, thereby ensuring that $\int v_k \,\mathrm{d}\lambda = 0$ and $\int \pi_\alpha \,\mathrm{d}\lambda = 1$ for each α in $\{0, 1\}^K$.

In summary, the choice of v_k as in (9) gives

$$\frac{1}{2\pi}\|f_\alpha - f_\beta\|_2^2 = \varepsilon^2 \int_{-\infty}^{\infty}\left(\sum_{k\in K}(\alpha_k - \beta_k)\psi_k(t)\right)^2 \mathrm{d}t = \varepsilon^2 \sum_{k\in K}(\alpha_k - \beta_k)^2. \tag{10}$$

That is, the condition (2) of Lemma 1.4 is satisfied with $c_0 = 2\pi$.

We still need to check the condition (3), and also show that ε can be chosen small enough to make all the π_α's nonnegative. Actually, we first show that $\pi_\alpha \geq \pi_0/2 > 0$ by choosing

$$\varepsilon \leq \frac{1}{16}3^{-m+1/2}m^{-3/2}, \tag{11}$$

and by choosing $\pi_0 = \phi_m$. Secondly, we determine the largest size m while the two densities f_α and f_β are close in terms of the χ^2 distance as $O(1/n)$ when there is only one different coordinate between α and β.

To control the denominator in (3), we first show that $|v_k(u)| \leq C_k\sqrt{m}\pi_0(u)$ where $C_k = 8 \cdot 3^{k/2}$. By Cramér's inequality [3], equation (8.954)

$$\left|H_k(u)\right| \leq \kappa\sqrt{k!}\exp\left(u^2/4\right) \qquad \text{with } \kappa \approx 1.086. \tag{12}$$

Applying this inequality to (9),

$$\left|v_k(u)\right| \leq \kappa C 3^{k/2}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{6}u^2\right) \leq C_k\phi(u/\sqrt{3}) \tag{13}$$

$$\leq C_k\phi(u/\sqrt{m}) = C_k\sqrt{m}\pi_0(u). \tag{14}$$

Using (14), we have

$$\begin{aligned}\pi_\alpha(u) &= \pi_0(u) + \varepsilon\sum_{k\in K}\alpha_k v_k(u) \geq \pi_0(u) - \varepsilon\sum_{k\in K}C_k\sqrt{m}\pi_0(u)\\ &\geq \pi_0(u)\left[1 - C_{2m-1}m^{3/2}\varepsilon\right]\\ &= \pi_0(u)\left[1 - 8\cdot 3^{m-1/2}m^{3/2}\varepsilon\right] \geq \frac{\pi_0(u)}{2}\end{aligned}$$

by the choice of ε in (11).

Hence, under the condition (11), $\phi \star \Pi_\alpha := f_\alpha \geq f_0/2 := \phi \star \Pi_0/2$, which implies that the second condition in Lemma 1.4 is rewritten as $\int (f_\alpha - f_\beta)^2/f_0 \leq c_1/2n$ for α and β having only one different coordinate. The denominator $f_0 = \phi \star \Pi_0$ is again normally distributed with mean zero and variance $1+m$ by the choice of $\mathrm{d}\Pi_0/\mathrm{d}\lambda := \pi_0 = \phi_m$ density.

For convenience, we let $\alpha_1 \neq \beta_1$ (all the other cases work the same way). By splitting the integral into two regions $|x| \leq M\sqrt{m}$ and $|x| > M\sqrt{m}$ with a constant $M^2 = 8\log 9$,

$$\int \frac{(f_\alpha - f_\beta)^2}{f_0} = \int_{|x|\leq M\sqrt{m}} \frac{(f_\alpha - f_\beta)^2}{f_0} + \varepsilon^2 \int_{|x|>M\sqrt{m}} \frac{(\int \phi(x-u)v_1(u)\,\mathrm{d}\lambda)^2}{f_0}.$$

For the first integral, the denominator is bounded below on the interval $\{|x| \leq M\sqrt{m}\}$, since

$$f_0(x)\bigl\{|x| \leq M\sqrt{m}\bigr\} > \exp\bigl(-M^2/2\bigr)/(2\sqrt{2\pi}\sqrt{m}) := 1/\bigl(C^*\sqrt{m}\bigr),$$

where $C^* := 2\sqrt{2\pi}\exp(M^2/2)$. Then, using the $\mathcal{L}_2$ loss calculation from (10),

$$\int_{|x|\leq M\sqrt{m}} \frac{(f_\alpha(x) - f_\beta(x))^2}{f_0(x)}\,\mathrm{d}x \leq C^*\sqrt{m}\|f_\alpha - f_\beta\|_2^2 = 2\pi C^*\sqrt{m}\varepsilon^2.$$

For the second integral, recall that for any $k = 1, 3, \ldots, 2m-1$, we have

$$\bigl|v_k(u)\bigr| \leq C_{2m-1}\phi(u/\sqrt{3}) := C_{2m-1}\sigma_0\phi_{\sigma_0^2}$$

with $\sigma_0 = \sqrt{3}$ as in (13). Using $C_{2m-1} := 8 \cdot 3^{m-1/2}$ and $\phi_{1+\sigma_0^2}(x) \leq \sqrt{m}\phi_{1+m}(x)$, with a notation $R(x) := \{|x| > M\sqrt{m}\}$, we bound the second integral:

$$\begin{aligned}
\varepsilon^2 \int_{R(x)} \frac{(\int \phi(x-u)v_1(u)\,\mathrm{d}\lambda)^2}{f_0(x)}\,\mathrm{d}x &\leq \varepsilon^2 C_{2m-1}^2\sigma_0^2 \int_{R(x)} \frac{(\int \phi(x-u)\phi_{\sigma_0^2}(u)\,\mathrm{d}\lambda)^2}{\phi_{1+m}(x)}\,\mathrm{d}x \\
&\leq \sqrt{m}\varepsilon^2 C_{2m-1}^2\sigma_0^2 \int_{R(x)} \phi_{1+\sigma_0^2}(x)\,\mathrm{d}x \\
&= \left(\frac{64}{3}\sqrt{m}\varepsilon^2\right)\left(3^{2m}\int_{R(x)} \phi_{1+\sigma_0^2}(x)\,\mathrm{d}x\right) \\
&\leq \frac{64}{3}\sqrt{m}\varepsilon^2.
\end{aligned}$$

Here the last inequality is obtained by a Gaussian tail property with $\sqrt{m} \gg \sigma_0 := \sqrt{3}$, namely

$$\int_{|x|>M\sqrt{m}} \phi_{1+\sigma_0^2}(x)\,\mathrm{d}x \leq \exp\left(-\frac{1}{8}M^2 m\right) = 3^{-2m}$$

since $M^2 = 8\log 9$.

Combining these two upper bounds for the integral, we obtain

$$\int \frac{(f_\alpha - f_\beta)^2}{f_0} \le \sqrt{m}\varepsilon^2\bigl(2\pi C^* + 64/3\bigr) = \frac{c_1}{2n}$$

as long as

$$\sqrt{m}\varepsilon^2 \le \frac{1}{n}\frac{c_1}{2(2\pi C^* + 64/3)}. \tag{15}$$

As a consequence, the constructed mixing densities fulfil the two requirements in Assouad's lemma under conditions (11) and (15), with

$$\varepsilon^2 \le \min\biggl(\frac{1}{16^2}3^{-2m+1}m^{-3}, \frac{1}{n\sqrt{m}}\frac{c_1}{2(2\pi C^* + 64/3)}\biggr).$$

From Lemma 1.4, the lower bound is obtained as $c\varepsilon^2 m$, which is at most

$$\min\bigl(3^{-2m}m^{-2}, \sqrt{m}/n\bigr)$$

up to a constant. To find the largest $m\varepsilon^2$, by equating $3^{-2m}m^{-2} = \sqrt{m}/n$, we obtain m and ε^2 as $\log n$ and $1/(n\sqrt{\log n})$, respectively, up to a constant, and hence the lower bound is obtained as $\sqrt{\log n}/n$ up to a constant. □

2.2. Ideas in the proof of Theorem 1.3

Here we let $W(f,g) := \|\sqrt{f} - \sqrt{g}\|_2^2 = \int(\sqrt{f} - \sqrt{g})^2$, so (1) is satisfied with $\zeta = 1$. First, we relate the Hellinger distance and the χ^2 distance. That is, suppose we can show $(1/2)\pi_0(u) \le \pi_\varsigma(u) \le (3/2)\pi_0(u)$ so that $(1/2)f_0(x) \le f_\varsigma(x) \le (3/2)f_0(x)$ for both $\varsigma = \alpha$ and $\varsigma = \beta$, by convolving with the standard normal density. Then, using the upper bound for f_α and f_β,

$$\int (\sqrt{f_\alpha} - \sqrt{f_\beta})^2 = \int \frac{(f_\alpha - f_\beta)^2}{(\sqrt{f_\alpha} + \sqrt{f_\beta})^2} \ge \frac{1}{6}\int \frac{(f_\alpha - f_\beta)^2}{f_0}. \tag{16}$$

Similarly, the lower bound for f_α would give an upper bound for the testing condition

$$\int \frac{(f_\alpha - f_\beta)^2}{f_\alpha} \le 2\int \frac{(f_\alpha - f_\beta)^2}{f_0}. \tag{17}$$

Thus it would be enough to work with the following quantity

$$\int \frac{(f_\alpha - f_\beta)^2}{f_0} = \int \biggl(\frac{f_\alpha}{\sqrt{f_0}} - \frac{f_\beta}{\sqrt{f_0}}\biggr)^2 = \varepsilon^2\int \biggl(\sum_{k\in K}(\alpha_k - \beta_k)\frac{\phi \star v_k}{\sqrt{f_0}}\biggr)^2,$$

where the second equality is given by (6).

At first glance, $\int (f_\alpha - f_\beta)^2/f_0$ does not look amenable to Fourier techniques. However, as Lemma 2.2 below shows, $\phi \star v_k/\sqrt{f_0}$ is expressed as convolution of a normal density (with a variance larger than 1) with a certain choice of the perturbation function v_k and base function $\pi_0 = \phi_{\sigma^2}$.

Lemma 2.2. *Consider the perturbation functions*

$$v_k(u) = \frac{C_k}{\sqrt{k!}}\phi(\rho u)H_k(\gamma u), \qquad \rho^2 \geq \frac{1}{\sigma^2} + \frac{\gamma^2}{2},$$

where C_k is a constant depending on k and $\gamma > 0$. Then

$$\frac{[\phi \star v_k](x)}{\sqrt{\phi \star \phi_{\sigma^2}(x)}} = \phi_{\tilde{\sigma}^2} \star \tilde{v}_k, \tag{18}$$

where

$$\tilde{v}_k(u) := \frac{\tilde{C}_k}{\sqrt{k!}}\phi(\tilde{\rho} u)H_k(\tilde{\gamma} u), \tag{19}$$

with

$$\tilde{\sigma}^2 = 1 + \frac{1}{2\sigma^2 + 1}, \qquad \tilde{C}_k = C_k\frac{(4\pi)^{1/4}}{\tilde{\sigma}}, \qquad \tilde{\rho} = \frac{\sqrt{\rho^2 + 1 - \tilde{\sigma}^2}}{\tilde{\sigma}^2}, \qquad \tilde{\gamma} = \frac{\gamma}{\tilde{\sigma}^2}. \tag{20}$$

By Lemma 2.2, the denominator effect can be incorporated into the normal convolution. Then we follow similar ideas used in the proof of Theorem 1.1.

Proof of Theorem 1.3. Again, the choice of v_k's is suggested by Fourier methods. For convenience, we let $\pi_0 = \phi$, so $f_0 = \phi_2$ and $\sqrt{f_0} = 2\pi^{1/4}\phi_4$. Assuming $\tilde{v}_k$ in (19) are in $\mathcal{L}_2$,

$$\begin{aligned}\mathcal{T}\left[\frac{f_\alpha}{\sqrt{f_0}}\right](t) &= \mathcal{T}[\sqrt{f_0}](t) + \varepsilon\sum_{k\in K}\alpha_k\mathcal{T}\left[\frac{\phi \star v_k}{\sqrt{f_0}}\right](t)\\ &= \mathcal{T}[\sqrt{f_0}](t) + \varepsilon\sum_{k\in K}\alpha_k\mathcal{T}\phi_{4/3}(t)\mathcal{T}\tilde{v}_k(t)\end{aligned}$$

by Lemma 2.2.

By the Plancherel formula,

$$\frac{1}{2\pi}\left\|\frac{f_\alpha}{\sqrt{f_0}} - \frac{f_\beta}{\sqrt{f_0}}\right\|_2 = \varepsilon^2\int\left|\sum_{k\in K}(\alpha_k - \beta_k)\mathcal{T}\phi_{4/3}(t)\mathcal{T}[\tilde{v}_k](t)\right|^2 \mathrm{d}t,$$

which lets us write the condition (2) in Assouad's Lemma 1.4 as

$$\int\left|\sum_{k\in K}(\alpha_k - \beta_k)\mathcal{T}[\phi_{4/3}](t)\mathcal{T}[\tilde{v}_k](t)\right|^2 \mathrm{d}t \geq \frac{3c_0}{\pi}\sum_{k\in K}(\alpha_k - \beta_k)^2 \qquad \forall\alpha, \beta \in \{0,1\}^K,$$

with $\delta = \pi c_0 \varepsilon^2/3$ by (16).

Similar to the case for squared error loss, we might achieve even an equality with $c_0 = \pi/3$ by choosing $\tilde{v}_k$'s to make the functions $\psi_k(t) := \mathcal{T}[\phi_{4/3}](t)\mathcal{T}[\tilde{v}_k](t)$ orthonormal. Ignoring other requirements, we also start from the same orthonormal set (7), and then try to define $\tilde{v}_k$ as the inverse Fourier transform.

From the fact that

$$\mathcal{T}[\phi_{4/3}](t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{2}{3}t^2\right)$$

and by definition of $\tilde{v}_k$ in (19), the requirement is that

$$\psi_k(t) := \frac{\tilde{C}_k}{\sqrt{2\pi k!}} \exp\left(-\frac{2t^2}{3}\right)\mathcal{T}\big[\phi(\tilde{\rho}u)H_k(\tilde{\gamma}u)\big](t) = \mathrm{i}^{-k}\sqrt{2\phi(2t)}\frac{H_k(2t)}{\sqrt{k!}}. \tag{21}$$

If we determine all the parameters to make (21) true, we have the desired property for the loss separation condition (2), that is, we have

$$\int \frac{(f_\alpha - f_\beta)^2}{f_0} = 2\pi\varepsilon^2 \sum_{k\in K}(\alpha_k - \beta_k)^2. \tag{22}$$

We have to find $\tilde{\rho}$, $\tilde{\gamma}$ and $\tilde{C}_k$ so that (21) is satisfied. The solutions are derived below and given in (23).

After some calculations,

$$\mathcal{T}\big[\phi(\tilde{\rho}u)H_k(\tilde{\gamma}u)\big](t) = \mathrm{i}^{-k}\frac{\sqrt{2}(2\pi)^{3/4}}{\tilde{C}_k}\phi\left(t\sqrt{\frac{2}{3}}\right)H_k(2t),$$

which leads to

$$\mathcal{T}^{-1}\left[\phi\left(t\sqrt{\frac{2}{3}}\right)H_k(2t)\right](u) = \mathrm{i}^k\frac{\tilde{C}_k}{\sqrt{2}(2\pi)^{3/4}}\phi(\tilde{\rho}u)H_k(\tilde{\gamma}u).$$

Substituting $a = \sqrt{2/3}$ and $b = 2$ into Lemma 2.1, we have the following solutions,

$$\tilde{C}_k = (2\pi)^{3/4}\sqrt{3}\sqrt{5^k}, \qquad \tilde{\rho} = \sqrt{\frac{3}{2}}, \qquad \tilde{\gamma} = \frac{3}{\sqrt{5}}. \tag{23}$$

We need to ensure that the choice of $\sigma^2 = 1$ satisfies the inequality $\rho^2 \geq \frac{1}{\sigma^2} + \frac{\gamma^2}{2}$ needed for the Lemma 2.2. Comparing (20) and (23), we obtain $\rho^2 = 3$ and $\gamma = \frac{4}{\sqrt{5}}$, which satisfy the condition. Also, C_k is obtained as $C_k = (2^{5/4}\sqrt{\pi})\sqrt{5}^k$.

Therefore, this choice for the ψ_k's leads to

$$v_k(u) = 2^{5/4}\sqrt{\pi}\sqrt{\frac{5^k}{k!}}\phi(\sqrt{3}u)H_k\left(\frac{4}{\sqrt{5}}u\right) \qquad \text{for } k \in K. \tag{24}$$

By restricting to odd values of k, we make the v_k's real-valued and odd, thereby ensuring that $\int v_k \,\mathrm{d}\lambda = 0$.

Using exactly the same idea as in the previous section, if

$$\varepsilon \le \frac{1}{2\kappa m C_{2m-1}} \qquad \text{with } \kappa \simeq 1.086, \tag{25}$$

then

$$\tfrac{1}{2}\pi_0(u) \le \pi_\alpha(u) \le \tfrac{3}{2}\pi_0(u) \qquad \text{for all } u \in \mathbb{R}, \alpha \in \{0, 1\}^K.$$

Now the second testing condition can be treated straightforwardly. Indeed, once we choose orthonormal functions $\{\psi_k, k \in K\}$, we obtain

$$\int \frac{(f_\alpha - f_\beta)^2}{f_\alpha} \le \int 2\frac{(f_\alpha - f_\beta)^2}{f_0} = 4\pi\varepsilon^2 \qquad \text{for } \|\alpha - \beta\|_0 = 1 \text{ by (17) and (22).}$$

Thus it is enough to choose $\varepsilon^2 < 1/(4\pi n)$. With our choice $\varepsilon = 1/(4\sqrt{n})$, the testing condition is satisfied.

From the lower bound $m\varepsilon^2$, we want to choose m as large as possible. The condition in (25) restricts the size of m,

$$2\kappa m C_{2m-1} < 6m5^m \le 4\sqrt{n}.$$

Thus, we have the upper bound for m,

$$m \lesssim \bigl(1/(2\log 5)\bigr) \log n \simeq (0.31) \log n.$$

Finally, we check these constructed π_α's are inside of the parameter space $\mathcal{P}_s(\mathbb{R})$. From the fact that $\pi_\alpha(u) \le (3/2)\pi_0(u)$ for all $u \in \mathbb{R}$ and $\alpha \in \{0, 1\}^K$, it is clear that π_0 is in the space $\mathcal{P}_s(\mathbb{R})$ from the tail property of normal density.

Consequently, the lower bound is obtained as $\log n/n$ up to a constant. □

3. Proof of the upper bound

For the reader's convenience, we summarize the arguments for Theorem 1.2, following pages 365–369 by [4]. Before turning to that result, we first show that if $f = \phi \star \Pi \in \mathcal{F}$, then f can be extended to an entire function f^*. To see this, we let $f^*(x + \mathrm{i}y) = \frac{1}{\sqrt{2\pi}} \int \exp(-(x + \mathrm{i}y - u)^2)/2) \,\mathrm{d}\Pi(u)$. Defining $a(x, y, u) \equiv y(u - x)$, $b(x, y, u) \equiv -\frac{1}{2}\{(x - u)^2 - y^2\}$, we write

$$\begin{aligned}\sqrt{2\pi} f^*(x + \mathrm{i}y) &= \int \bigl[\bigl\{\cos a(x, y, u) + \mathrm{i} \sin a(x, y, u)\bigr\} \mathrm{e}^{b(x,y,u)}\bigr] \,\mathrm{d}\Pi(u) \\ &= \int \bigl\{\cos a(x, y, u) \mathrm{e}^{b(x,y,u)}\bigr\} \,\mathrm{d}\Pi(u) + \mathrm{i} \int \bigl\{\sin a(x, y, u) \mathrm{e}^{b(x,y,u)}\bigr\} \,\mathrm{d}\Pi(u) \\ &:= v(x, y) + \mathrm{i}w(x, y).\end{aligned}$$

By differentiating under the integral (see Theorem 16.8 in [1]),

$$\frac{\partial v}{\partial x} = \int \big[\big\{y \sin a(x,y,u) + (u-x)\cos a(x,y,u)\big\} e^{b(x,y,u)}\big] \,\mathrm{d}\Pi(u) = \frac{\partial w}{\partial y},$$

$$\frac{\partial v}{\partial y} = \int \big[\big\{y \cos a(x,y,u) - (u-x)\sin a(x,y,u)\big\} e^{b(x,y,u)}\big] \,\mathrm{d}\Pi(u) = -\frac{\partial w}{\partial x}.$$

Also note that $\partial v/\partial x$, $\partial v/\partial y$, $\partial w/\partial x$, and $\partial w/\partial y$ are continuous. Then by Cauchy–Riemann theorem (see Theorem 1.5.8 in [8]), f^* is analytic.

Now it suffices to show that f^* satisfies the growth condition. Indeed,

$$\begin{aligned}
\sup_x \big|f^*(x+\mathrm{i}y)\big| &= \sup_x \left|\frac{1}{\sqrt{2\pi}} \int \exp\left(-\frac{(x+\mathrm{i}y-u)^2}{2}\right) \mathrm{d}\Pi(u)\right| \\
&\le \frac{1}{\sqrt{2\pi}} \sup_x \int \left|\exp\left(-\frac{(x+\mathrm{i}y-u)^2}{2}\right)\right| \mathrm{d}\Pi(u) \\
&\le \frac{1}{\sqrt{2\pi}} \exp\left(\frac{y^2}{2}\right) \sup_x \int \mathrm{d}\Pi(u) \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(\frac{y^2}{2}\right).
\end{aligned} \tag{26}$$

Thus, $\mathcal{F} \subseteq \mathcal{F}^*$, which ensures that Ibragimov's estimation also gives the upper bound to match Theorem 1.1.

Proof of Theorem 1.2. Ibragimov used a *sinc* kernel estimator,

$$\hat{f}_n(x) = \frac{1}{nh} \sum_{j=1}^{n} \mathcal{K}\left(\frac{X_j - x}{h}\right), \qquad \mathcal{K}(u) = \frac{\sin(u)}{\pi u},$$

with $h = 1/\sqrt{\log n}$. It is important for his method that the Fourier transform of $\mathcal{K}$ is $\breve{\mathcal{K}}(t) = \frac{1}{\sqrt{2\pi}} \mathbb{1}\{|t| \le 1\}$ and also $\breve{\mathcal{K}}^2(t) = \frac{1}{\sqrt{2\pi}} \frac{1}{\pi}(1 - \frac{|t|}{2})_+$ where $x_+ = \max(x, 0)$.

The expected mean integrated squared error (MISE) has the usual squared bias and variance decomposition. As usual, the variance term is bounded by $(nh)^{-1} \int \mathcal{K}^2(u)\,\mathrm{d}u$. For the bias term, note that $\mathbb{E}_{n,f} \hat{f}$ has the Fourier transform $\sqrt{2\pi} \breve{f}(t) \breve{\mathcal{K}}(ht)$, so that

$$\begin{aligned}
\mathrm{bias}^2 := \int (\mathbb{E}_{n,f} \hat{f}_n - f)^2 &= \int \big|\mathcal{T}[\mathbb{E}_{n,f} \hat{f}_n](t) - \mathcal{T}[f](t)\big|^2 \,\mathrm{d}t \qquad \text{by Plancherel} \\
&= \int \big|\breve{f}(t)\big|^2 \big|\sqrt{2\pi}\breve{\mathcal{K}}(ht) - 1\big|^2 \,\mathrm{d}t \\
&= \int_{|t| \ge 1/h} \big|\breve{f}(t)\big|^2 \,\mathrm{d}t \qquad \text{by the form of } \breve{\mathcal{K}}
\end{aligned}$$

$$\leq 2\mathrm{e}^{-y/h}\int \mathrm{e}^{-yt}\left|\breve{f}(t)\right|^2 \mathrm{d}t \qquad \text{for } y>0$$
$$= 2\mathrm{e}^{-y/h}\lim_{M\to\infty}\int \mathrm{e}^{-yt}\left|\breve{f}(t)\right|^2\left(1-\frac{|t|}{M}\right)_+ \mathrm{d}t.$$

Write the last integral as

$$\int \mathrm{e}^{-yt}\left|\breve{f}(t)\right|^2\left(1-\frac{|t|}{M}\right)_+ \mathrm{d}t = \int \breve{f}(t)\mathrm{e}^{-yt}\overline{\breve{\vartheta}(t)}\,\mathrm{d}t = \int \breve{f}(t)\mathrm{e}^{-yt}\breve{\vartheta}(-t)\,\mathrm{d}t,$$

where $\breve{\vartheta}(t) = \breve{f}(t)(1-\frac{|t|}{M})_+$ is the Fourier transform of the nonnegative function $\vartheta(x) = \frac{M\pi^2}{\sqrt{2\pi}}\int f(u)\mathcal{K}^2(\frac{M}{2}(x-u))\,\mathrm{d}u$. Using $\vartheta(x+\mathrm{i}y) = \frac{1}{\sqrt{2\pi}}\int \mathrm{e}^{\mathrm{i}tx}\mathrm{e}^{-yt}\breve{\vartheta}(t)\,\mathrm{d}t$, we have $\int f(x)\vartheta(x+\mathrm{i}y)\,\mathrm{d}x = \int \breve{f}(t)\mathrm{e}^{yt}\breve{\vartheta}(-t)\,\mathrm{d}t$ by Parseval's theorem. By changing the contour of the integration, $\int f(x+\mathrm{i}y)\vartheta(x)\,\mathrm{d}x = \int \breve{f}(t)\mathrm{e}^{-yt}\breve{\vartheta}(-t)\,\mathrm{d}t$. Combining these ideas,

$$\int \breve{f}(t)\mathrm{e}^{-yt}\vartheta(-t)\,\mathrm{d}t = \int f(x+\mathrm{i}y)\vartheta(x)\,\mathrm{d}x \leq \int \sup_x\left|f(x+\mathrm{i}y)\right|\vartheta(x)\,\mathrm{d}x \leq \frac{\exp(y^2/2)}{\sqrt{2\pi}},$$

where the last inequality follows by (26) together with $\int \vartheta(x)\,\mathrm{d}x = \sqrt{2\pi}\breve{\vartheta}(0) = 1$. By taking $y = 1/h$, we obtain the upper bound as $\sqrt{\log n}/n := \ell_n$ up to a constant. □

4. Discussion

It has been claimed that the Fano's method is more general in a sense (see [13], page 428). Indeed, using Varshamov–Gilbert's lemma (e.g., Lemma 2.9 in [10]), it is not very difficult to prove the same rate result for $\mathcal{F}$ with a similar type of sub-parameter space using Fano's method.

However, Assouad's method seems more convenient in some cases. For instance, before knowing how to construct the subspace, it would be extremely difficult to determine the right family of densities when there are only indirect regularity conditions as in this example. Assouad's hyper-rectangle method indicates that the problem can be solved if we can show the orthogonality relations between the constructed densities. These added regularity conditions can cause different difficulties, but we at least have some clues to handle these problems.

On the other hand, if we know metric entropy (good packing and covering number bounds) results beforehand, the optimal minimax rates can be obtained almost automatically with the predictive Bayes density estimator using the main theorems in [12]. It will be interesting to see if we can calculate a sharper metric entropy for $\mathcal{F}$ or $\mathcal{F}_s$ than the one that appeared in [2].

Appendix

Proof of Lemma 1.4. Most of the proof is based on ideas borrowed from [7,10], and some unpublished notes by David Pollard. Denote $A = \{0,1\}^K$ and for convenience denote $\mathbb{E}_\alpha$ for

$\mathbb{E}_{f_\alpha}$ and $\mathbb{P}_\alpha$ for $\mathbb{P}_{f_\alpha}$ where $\mathbb{P}_{f_\alpha} = P^n_{f_\alpha}$. For any density estimator $\hat{f}_n$ based on the observation $X_1, \dots, X_n$, define an estimator

$$\hat{\alpha} = \arg\min_{\alpha \in A} W(\hat{f}_n, f_\alpha).$$

By restricting the parameter space and by the definition of $\hat{\alpha}$,

$$\begin{aligned}\sup_{f \in \mathcal{F}} \mathbb{E}_f W(\hat{f}_n, f) &\geq \max_{\alpha \in A} \mathbb{E}_\alpha W(\hat{f}_n, f_\alpha) \\ &\geq \frac{1}{2} \max_{\alpha \in A} \mathbb{E}_\alpha \big(W(\hat{f}_n, f_\alpha) + W(\hat{f}_n, f_{\hat{\alpha}}) \big) \\ &\geq \frac{\zeta}{2} \max_{\alpha \in A} \mathbb{E}_\alpha W(f_\alpha, f_{\hat{\alpha}})\end{aligned}$$

using the pseudo-distance property (1). Now, using the condition (2) in the lemma followed by the simple fact that the supremum is bounded by the average, the last equation can be lower bounded by

$$\frac{c_0 \varepsilon^2 \zeta}{2} \max_{\alpha \in A} \sum_{k=1}^m \mathbb{E}_\alpha \mathbb{1}\{\alpha_k \neq \hat{\alpha}_k\} \geq \frac{c_0 \varepsilon^2 \zeta}{2} \frac{1}{2^m} \sum_{\alpha \in A} \sum_{k=1}^m \mathbb{E}_\alpha \mathbb{1}\{\alpha_k \neq \hat{\alpha}_k\}.$$

Define

$$\bar{\mathbb{P}}_{0,k} = \frac{1}{2^{m-1}} \sum_{\alpha \in A_{0,k}} \mathbb{P}_\alpha, \qquad \bar{\mathbb{P}}_{1,k} = \frac{1}{2^{m-1}} \sum_{\alpha \in A_{1,k}} \mathbb{P}_\alpha, \qquad k = 1, \dots, m,$$

where $A_{i,k} = \{\alpha \in A\colon \alpha_k = i\}$ for $i = 0, 1$.

Since $\alpha_k, \hat{\alpha}_k \in \{0, 1\}$, we have

$$\begin{aligned}\frac{1}{2^m} \sum_{\alpha \in A} \sum_{k=1}^m \mathbb{E}_\alpha \mathbb{1}\{\alpha_k \neq \hat{\alpha}_k\} &= \frac{1}{2^m} \sum_{k=1}^m \Bigg(\sum_{\alpha \in A_{0,k}} \mathbb{P}_\alpha \mathbb{1}\{\hat{\alpha}_k \neq 0\} + \sum_{\alpha \in A_{1,k}} \mathbb{P}_\alpha \mathbb{1}\{\hat{\alpha}_k \neq 1\} \Bigg) \\ &= \frac{1}{2} \sum_{k=1}^m \big(\bar{\mathbb{P}}_{0,k} \mathbb{1}\{\hat{\alpha}_k \neq 0\} + \bar{\mathbb{P}}_{1,k} \mathbb{1}\{\hat{\alpha}_k \neq 1\} \big),\end{aligned}$$

which gives us the following lower bound

$$\sup_{f \in \mathcal{F}} \mathbb{E}_f W(\hat{f}_n, f) \geq \frac{c_0 \varepsilon^2 \zeta}{4} \sum_{k=1}^m \|\bar{\mathbb{P}}_{0,k} \wedge \bar{\mathbb{P}}_{1,k}\|_1$$

by $Ph + Q(1-h) \geq \|P \wedge Q\|_1$ for $h \geq 0$ with $h = \mathbb{1}\{\hat{\alpha} \neq 0\}$.

For $k=m$, each α in $A_{0,m}$ is of the form $(\gamma,0)$ with $\gamma\in D:=\{0,1\}^{m-1}$. Similarly, each α in $A_{1,m}$ is of the form $(\gamma,1)$ with $\gamma\in D$. Now

$$\|\bar{\mathbb{P}}_{0,m}\wedge\bar{\mathbb{P}}_{1,m}\|_1=\int\left(\frac{1}{2^{m-1}}\sum_{\gamma\in D}p_{\gamma,0}\right)\wedge\left(\frac{1}{2^{m-1}}\sum_{\gamma\in D}p_{\gamma,1}\right)\geq\int\frac{1}{2^{m-1}}\sum_{\gamma\in D}(p_{\gamma,0}\wedge p_{\gamma,1}).$$

Note that $(\gamma,0)$ and $(\gamma,1)$ have only one different coordinate. By similar calculations for other k's, we obtain

$$\sup_{f\in\mathcal{F}}\mathbb{E}_f W(\hat{f}_n,f)\geq\frac{c_0\varepsilon^2\zeta}{4}m\min_{d(\alpha,\beta)=1}\|\mathbb{P}_\alpha\wedge\mathbb{P}_\beta\|_1.$$

In general, it is difficult to calculate the testing affinity exactly. Fortunately, a convenient lower bound in terms of distances between marginals is available when $\mathbb{P}_\alpha$ and $\mathbb{P}_\beta$ are both product measures. For instance, when $\mathbb{P}_\alpha=P_\alpha^n$ for i.i.d. case, we can bound this using the chi-squared distance χ^2 by the following relation.

$$\bigl(1-\|\mathbb{P}_\alpha\wedge\mathbb{P}_\beta\|_1\bigr)^2\leq n\chi^2(P_\alpha,P_\beta):=n\int\frac{(\theta_\alpha-\theta_\beta)^2}{\theta_\alpha}.$$

Thus, the condition (3) in the lemma yields a lower bound for the maximum risk

$$\sup_{f\in\mathcal{F}}\mathbb{E}_f W(\hat{f}_n,f)\geq\frac{c_0\varepsilon^2\zeta}{4}m(1-\sqrt{c_1}).$$

□

See [10], Lemma 2.7 on page 90, or [7], Lemma 1 on page 40, for the derivation of facts about relations between distances.

Proof of Lemma 2.1. For $b>a>0$, we have

$$\phi(at)\exp\left(btx-\frac{1}{2}x^2\right)=\phi(at)\sum_{k=0}^{\infty}\frac{H_k(bt)}{k!}x^k.$$

Thus,

$$\begin{aligned}\mathcal{T}^{-1}\left[\phi(at)\exp\left(btx-\frac{1}{2}x^2\right)\right](u)&=\int_{-\infty}^{\infty}\frac{\exp(\mathrm{i}tu)}{2\pi}\exp\left(-\frac{a^2t^2}{2}+bxt-\frac{1}{2}x^2\right)\mathrm{d}t\\&=\frac{1}{a\sqrt{2\pi}}\exp\left(\frac{(bx+\mathrm{i}u)^2}{2a^2}-\frac{1}{2}x^2\right)\\&=\frac{1}{a}\phi\left(\frac{u}{a}\right)\exp\left(\frac{bxu\mathrm{i}}{a^2}-\frac{1}{2}(\mathrm{i}xc_{a,b})^2\right)\\&=\frac{1}{a}\phi\left(\frac{u}{a}\right)\sum_{k=0}^{\infty}\frac{H_k(b/(a^2c_{a,b})u)}{k!}(\mathrm{i}c_{a,b})^kx^k.\end{aligned}$$

The inverse Fourier transform of the right side is

$$\sum_{k=0}^{\infty} \mathcal{T}^{-1}\left[\phi(at)\frac{H_k(bt)}{k!}\right](u)x^k.$$

By matching the coefficient for the kth power of x,

$$\mathcal{T}^{-1}\big[\phi(at)H_k(bt)\big](u) = (\mathrm{i}c_{a,b})^k \frac{1}{a}\phi\left(\frac{u}{a}\right)H_k\left(\frac{b}{a^2 c_{a,b}}u\right),$$

which proves the claim. □

Proof of Lemma 2.2. First, note that $\phi \star \phi_{\sigma^2} = \phi_{1+\sigma^2}$. We define $[\phi \star v_k(u)](x) = \int \phi(x - u)v_k(u)\,\mathrm{d}u$ and similarly $[\phi \star \phi(\rho u)H_k(\gamma u)](x) = \int \phi(x-u)\phi(\rho u)H_k(ru)\,\mathrm{d}u$. By definition of v_k, we have

$$\begin{aligned}\frac{[\phi \star v_k(u)](x)}{\sqrt{\phi_{1+\sigma^2}(x)}} &= \frac{C_k}{\sqrt{k!}}\frac{[\phi \star \phi(\rho u)H_k(\gamma u)](x)}{\sqrt{\phi_{1+\sigma^2}(x)}} \\ &= \frac{C_k}{\sqrt{k!}}\int \frac{(1/\sqrt{2\pi})\exp(-1/2(x-u)^2)(1/\sqrt{2\pi})\exp(-(1/2)\rho^2u^2)H_k(\gamma u)}{(2\pi(1+\sigma^2))^{-1/4}\exp(-(1/4)(x^2/(1+\sigma^2)))}\,\mathrm{d}u.\end{aligned}$$

Now, by completing the square,

$$\begin{aligned}&\exp\left(-\frac{1}{2}(x-u)^2\right)\exp\left(-\frac{1}{2}\rho^2u^2\right)\exp\left(\frac{1}{4}\frac{x^2}{1+\sigma^2}\right) \\ &\quad= \exp\left(\left(-\frac{1}{2}+\frac{1}{4(1+\sigma^2)}\right)x^2 + xu - \left(\frac{1}{2}+\frac{1}{2}\rho^2\right)u^2\right) \\ &\quad= \exp\left(-\frac{1}{2\tilde{\sigma}^2}x^2 + xu - \left(\frac{1}{2}+\frac{1}{2}\rho^2\right)u^2\right) \qquad \text{by definition of } \tilde{\sigma}^2 \text{ in (20)} \\ &\quad= \exp\left(-\frac{1}{2\tilde{\sigma}^2}(x-\tilde{u})^2\right)\exp\left(-\frac{1}{2}\left(1+\rho^2-\tilde{\sigma}^2\right)\frac{\tilde{u}^2}{\tilde{\sigma}^4}\right) \qquad \text{by } \tilde{u} := \tilde{\sigma}^2 u \\ &\quad= (2\pi\tilde{\sigma})\phi_{\tilde{\sigma}^2}(x-\tilde{u})\phi\left(\frac{\sqrt{1+\rho^2-\tilde{\sigma}^2}}{\tilde{\sigma}^2}\tilde{u}\right),\end{aligned}$$

where the positive value for $(1+\rho^2-\tilde{\sigma}^2)$ is guaranteed by the condition $\rho^2 \ge 1/\sigma^2 + \gamma^2/2 > 1/(1+2\sigma^2) := 1-\tilde{\sigma}^2$. By change of variables,

$$\frac{[\phi \star v_k(u)](x)}{\sqrt{\phi_{1+\sigma^2}(x)}} = \left(\frac{C_k}{\sqrt{k!}}\frac{[2\pi(1+\sigma^2)]^{1/4}}{\tilde{\sigma}}\right)\phi_{\tilde{\sigma}^2} \star \phi\left(\frac{\sqrt{1+\rho^2-\tilde{\sigma}^2}}{\tilde{\sigma}^2}\tilde{u}\right)H_k\left(\frac{\gamma}{\tilde{\sigma}^2}\tilde{u}\right).$$

Using the definitions of each transformed variables (20), the proof is complete. □

Acknowledgements

This work is part of the author's Ph.D. dissertation, written at Yale University. The author is grateful to David Pollard, Harrison Zhou, and Richard J. Samworth for their comments and advice. The research was supported in part by NSF Career Award DMS-06-45676 and NSF FRG Grant DMS-08-54975.

References

[1] Billingsley, P. (1995). *Probability and Measure*, 3rd ed. *Wiley Series in Probability and Mathematical Statistics*. New York: Wiley. MR1324786
[2] 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
[3] Gradshteyn, I.S. and Ryzhik, I.M. (2007). *Table of Integrals, Series, and Products*, 7th ed. Amsterdam: Elsevier/Academic Press. MR2360010
[4] Ibragimov, I. (2001). Estimation of analytic functions. In *State of the Art in Probability and Statistics* (*Leiden*, 1999) (C. Klaasen, M. de Gunst and A.W. van der Vaart, eds.). *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **36** 359–383. Beachwood, OH: IMS. MR1836570
[5] Jackson, D. (2004). *Fourier Series and Orthogonal Polynomials*. Mineola, NY: Dover. MR2098657
[6] Kawata, T. (1972). *Fourier Analysis in Probability Theory*. New York: Academic Press. MR0464353
[7] Le Cam, L. (1973). Convergence of estimates under dimensionality restrictions. *Ann. Statist.* **1** 38–53. MR0334381
[8] Marsden, J.E. and Hoffman, M.J. (1987). *Basic Complex Analysis*, 2nd ed. New York: Freeman. MR0913736
[9] Rudin, W. (1987). *Real and Complex Analysis*, 3rd ed. New York: McGraw-Hill. MR0924157
[10] Tsybakov, A.B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. New York: Springer. MR2724359
[11] Van der Vaart, A.W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge: Cambridge Univ. Press. MR1652247
[12] Yang, Y. and Barron, A. (1999). Information-theoretic determination of minimax rates of convergence. *Ann. Statist.* **27** 1564–1599. MR1742500
[13] Yu, B. (1997). Assouad, Fano, and Le Cam. In *Festschrift for Lucien Le Cam* (D. Pollard, E. Torgersen and G.L. Yang, eds.) 423–435. New York: Springer. MR1462963

Received October 2012 and revised April 2013