Nonparametric estimation of the mixing density using polynomials

Tabea Rebafka, François Roueff

NONPARAMETRIC ESTIMATION OF THE MIXING DENSITY USING POLYNOMIALS

TABEA REBAFKA, FRANÇOIS ROUEFF

ABSTRACT

We consider the problem of estimating the mixing density f from n i.i.d. observations distributed according to a mixture density with unknown mixing distribution. In contrast with finite mixtures models, here the distribution of the hidden variable is not bounded to a finite set but is spread out over a given interval. We propose an approach to construct an orthogonal series estimator of the mixing density f involving Legendre polynomials. The construction of the orthonormal sequence varies from one mixture model to another. Minimax upper and lower bounds of the mean integrated squared error are provided which apply in various contexts. In the specific case of exponential mixtures, it is shown that the estimator is adaptive over a collection of specific smoothness classes, more precisely, there exists a constant $A > 0$ such that, when the order m of the projection estimator verifies $m \sim A\log(n)$, the estimator achieves the minimax rate over this collection. Other cases are investigated such as Gamma shape mixtures and scale mixtures of compactly supported densities including Beta mixtures. Finally, a consistent estimator of the support of the mixing density f is provided.

1. MIXTURE DISTRIBUTIONS

We consider mixture distributions of densities belonging to some parametric collection $\{\pi_t, t \in \Theta\}$ of densities with respect to the dominating measure ζ on the observation space $(\mathrm{X}, \mathcal{X})$. A general representation of a mixture density uses the so-called *mixing distribution* and is of the following form

$$\pi_f(x) = \int_\Theta f(t)\pi_t(x)\mu(\mathrm{d}t) \ , \tag{1}$$

where the *mixing density* f is a density with respect to some measure μ defined on Θ. If μ is a counting measure with a finite number of support points θ_k, then obviously, π_f is a finite mixture distribution of the form $\sum_{k=1}^K p_k \pi_{\theta_k}$. However, if μ denotes the Lebesgue measure on Θ, and if Θ is a given interval, say $\Theta = [a, b]$, then the distribution of the latent variable t is spread out over this interval and π_f represents a *continuous mixture*. In this paper we consider continuous mixtures and the problem of identifying the mixing density f when a sample of the continuous mixture π_f is observed. Note that when t is a location parameter, the problem of estimating f is referred to as a *deconvolution problem*, which has received considerable attention in the nonparametric statistics literature since [9].

Continuous mixtures have been used in very numerous and various fields of application. We just give some recent examples to show that continuous mixtures are still of much interest from an application point of view. The video-on-demand traffic can be modeled by a continuous Poisson mixture for the purpose of efficient cache managing [20]. In time-resolved fluorescence, where photon lifetimes have exponential distribution and parameters depend on the emitting molecules, typically continuous mixtures of exponential distributions are observed [18, 23]. When t is a scale parameter, the distribution π_f is called a *scale mixture*. Scale mixtures of uniforms are also related to multiplicative censoring introduced in Vardi [24] and length-biased data. A recent application in nanoscience of the latter are length measurements of carbon nanotubes, where observations are partly censored [16]. Exponential mixtures

play a significant role in natural sciences phenomena of discharge or disexcitation as e.g. radioactive decays, the electric discharge of a capacitor or the temperature difference between two objects. Several examples of applications of the exponential mixture model can be found in the references of the seminal paper [15].

Not only for applications, as well from a mathematical point of view, scale mixtures are particularly interesting as they define classes of densities that verify some monotonicity constraints. It is well known that any monotone non-increasing density function with support in $(0, +\infty)$ can be written as a mixture of uniform densities $\mathrm{U}[0, t]$ [12, p. 158]. Moreover, a k-*monotone* density is defined as a non-increasing, convex density function h whose derivatives satisfy for all $j = 1, \ldots, k-2$ that $(-1)^j h^{(j)}$ is non-negative, non-increasing and convex. One can show that any k-monotone density can be represented by a scale mixture of Beta distributions $B(1, k)$. Furthermore, densities that are k-monotone for any $k \geq 1$, also called *completely monotone* functions, can be written as a continuous mixture of exponential distributions [3].

The literature provides various approaches for the estimation of the mixing density, as for example the nonparametric maximum likelihood estimate (NPMLE). A characteristic feature of this estimator is that it yields a discrete mixing distribution [17, 19]. This appears to be unsatisfactory if we have reasons to believe that the mixing density is indeed a smooth function. In this case a functional approach is more appropriate, which relies on smoothness assumptions on the mixing density f. In Zhang [26] kernel estimators are constructed for mixing densities of a location parameter. Goutis [13] proposes an iterative estimation procedure also based on kernel methods. Asgharian et al. [2] show strong uniform consistency of kernel estimators in the specific case of multiplicative censoring. In the same setting, Andersen and Hansen [1] consider the linear operator K verifying $\pi_f = Kf$ and estimate f by an SVD reconstruction in the orthonormal basis of eigenfunctions of K. For mixtures of discrete distributions, that is when π_t are densities with respect to a counting measure on a discrete space, orthogonal series estimators have been developed and studied in Hengartner [14] and Roueff and Ryden [22]. For such mixtures, these estimators turn out to enjoy similar or better rates of convergence than the kernel estimator presented in Zhang [27]. Comte and Genon-Catalot [6] present a projection estimator based on Laguerre functions that has the specific feature that the support of the mixing density f is not a compact as usual, but the entire positive real line. Belomestny and Schoenmakers [4] extend the class of scale mixtures and derive estimation methods based on the Mellin transform.

In this paper we show that orthogonal series estimators can be provided in a very general fashion to estimate mixing densities with compact supports. In contrast to Andersen and Hansen [1], who consider only the case of scale mixtures of uniforms, our approach applies to a large variety of continuous mixtures as our numerous examples demonstrate. In the exponential mixture case, in particular, we exhibit an orthogonal series estimator achieving the minimax rate of convergence in a collection of smoothness classes without requiring a prior knowledge of the smoothness index. In other words, we provide an adaptive estimator of the mixing density of an exponential mixture.

The paper is organized as follows. In Section 2 the general construction of an orthogonal series estimator is presented and the estimator is applied in several different mixture settings. In Section 3 we derive upper bounds on the rate of convergence of the mean integrated squared error of the estimator on some specific smoothness classes. In Section 4 the approximation classes used for the convergence rate are related to more meaningful smoothness classes defined by weighted moduli of smoothness. Section 5 is concerned with the investigation of the minimax rate. On the one hand, a general lower bound of the MISE is provided and on the other hand, some specific cases are studied in detail. Section 6 provides a consistent estimator of the support of the mixing density. Finally, the performance of the projection estimator is evaluated by a simulation study in different mixture settings in Section 7. The Appendix provides some technical results.

2. Estimation Method

In this section we develop an orthogonal series estimator and we provide several examples, namely for mixtures of exponential, Gamma, Beta and uniform densities.

2.1. **Orthogonal Series Estimator.** Throughout this paper the following assumption will be used.

Assumption 1. Let ζ be a dominating measure on the observation space $(\mathrm{X}, \mathcal{X})$. Let $\{\pi_t, t \in \Theta\}$ be a parametric collection of densities with respect to ζ. Furthermore, let the parameter space $\Theta = [a, b]$ be a compact interval with known endpoints $a < b$ in $\mathbb{R}$. We denote by $X, X_1, \ldots, X_n$ an i.i.d. sample from the mixture distribution density π_f defined by (1) with μ equal to the Lebesgue measure on $[a, b]$.

For convenience, we also denote by π_t and π_f the probability measures associated to these densities. Moreover we will use the functional analysis notation $\pi_t(h)$ and $\pi_f(h)$, for the integral of h with respect to these probability measures.

The basic assumption of our estimation approach is that the mixing density f in (1) is square integrable, that is $f \in L^2[a, b]$. Then, for any complete orthonormal basis $(\psi_k)_{k\geq 1}$ of the Hilbert space $\mathbb{H} = L^2[a, b]$, the mixing density f can be represented by the orthogonal series $f(t) = \sum_{k\geq 1} c_k \psi_k(t)$, where the coefficients c_k correspond to the inner products of f and ψ_k. If we have estimators $\hat{c}_{n,k}$ of those coefficients, then an estimator of the mixing density f is obtained by $\sum_{k=1}^{m} \hat{c}_{n,k}\psi_k$.

To construct estimators $\hat{c}_{n,k}$, we remark that the following relation holds: Let g be a nonnegative integrable function on $\mathbb{R}$. Define the function φ on $[a, b]$ by the conditional expectations

$$\varphi(t) = \pi_t(g) = \int_{x\in \mathrm{X}} g(x)\pi_t(x)\ \zeta(\mathrm{d}x)\ , \quad t \in [a, b]\ . \tag{2}$$

Suppose that φ belongs to $\mathbb{H}$. The mean $\pi_f(g)$ can be written as the inner product of f and φ. Namely, by the definition of π_f in (1) and Fubini's theorem,

$$\pi_f(g) = \int_{x\in \mathrm{X}} g(x)\pi_f(x)\ \zeta(\mathrm{d}x) = \int_a^b f(t) \int_{x\in \mathrm{X}} g(x)\pi_t(x)\ \zeta(\mathrm{d}x)\,\mathrm{d}t = \langle f, \varphi\rangle_{\mathbb{H}}\ .$$

Consequently, by the strong law of large numbers, $\frac{1}{n}\sum_i g(X_i)$ is a consistent estimator of the inner product $\langle f, \varphi\rangle_{\mathbb{H}}$ based on an i.i.d. sample $(X_1, \ldots, X_n)$ from the mixture density π_f defined in (1).

We make the following assumption under which the orthogonal series estimator makes sense.

Assumption 2. Assumption 1 holds and there exists a sequence $(g_k)_{k\geq 1}$ of $\mathrm{X} \to \mathbb{R}$ functions such that $(\varphi_k)_{k\geq 1}$ is a dense sequence of linearly independent functions in $\mathbb{H}$, where $\varphi_k(t) = \pi_t(g_k)$ as in (2).

We then proceed as follows. Using linear combinations of the φ_k's, a sequence of orthonormal functions $\psi_1, \psi_2, \ldots$ in $\mathbb{H}$ can be constructed, for instance by the Gram-Schmidt procedure. Say that ψ_k writes as $\sum_{j=1}^{k} Q_{k,j}\varphi_j$ with an array $(Q_{k,j})_{1\leq j\leq k}$ of real values that are computed beforehand. Then we define estimators of $c_k = \langle f, \psi_k\rangle_{\mathbb{H}} = \sum_{j=1}^{k} Q_{k,j}\langle f, \varphi_j\rangle_{\mathbb{H}}$ by the empirical means

$$\hat{c}_{n,k} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k} Q_{k,j}g_j(X_i)\ .$$

Finally, for any integer m, an estimator of f is given by

$$\hat{f}_{m,n} = \frac{1}{n}\sum_{k=1}^{m} \hat{c}_{n,k}\psi_k = \frac{1}{n}\sum_{i=1}^{n}\sum_{j,k,l=1}^{m} Q_{k,j}Q_{k,l}g_j(X_i)\varphi_l\ , \tag{3}$$

with the convention $Q_{k,j} = 0$ for all $j > k$. We refer to $\hat{f}_{m,n}$ as the *orthogonal series estimator* or the *projection estimator* of approximation order m.

Define the subspaces

$$V_m = \text{span}(\varphi_1, \dots, \varphi_m) \ , \quad \text{for all } m \geq 1 \ . \tag{4}$$

By Assumption 2, the sequence $(V_m)_m$ is strictly increasing, V_m has dimension m for all m, and $\cup_m V_m$ has closure equal to $\mathbb{H}$. By construction the orthogonal series estimator $\hat{f}_{m,n}$ belongs to V_m. Consequently, the best squared error achievable by $\hat{f}_{m,n}$ is $\|f - P_{V_m}(f)\|_{\mathbb{H}}^2$, where $\|\cdots\|_{\mathbb{H}}$ denotes the norm associated to $\mathbb{H}$ and P_{V_m} the orthogonal projection on the space V_m. Hence once the functions g_k are chosen, the definition of the subspaces V_m follows and the performance of the estimator will naturally depend on how well f can be approximated by functions in V_m. It is thus of interest to choose a sequence $(g_k)_{k\geq 1}$ yielding a meaningful sequence of approximation spaces $(V_m)_m$. In the context of scale family mixtures (but not only, see Roueff and Ryden [22]), polynomial spaces appear naturally. Indeed, for any function g, we have $\pi_t(g) = \pi_1(g(t\cdot))$, so that, provided that π_1 has finite moments, if g is polynomial of degree k, so is $\varphi(t) = \pi_t(g)$. The following assumption slightly extends this choice for the two following reasons. First, a scale family is not always parameterized by its scale parameter but by its inverse (as for the exponential family). Second, it will appear that the choice of $(g_k)_{k\geq 1}$ not only influences the approximation class (and thus the bias) but also the variance. It may thus be convenient to allow the g_k's not to be polynomials, while still remaining in the context of polynomial approximation. This goal is achieved by the following assumption.

Assumption 3. Assumption 2 holds and there exist two real numbers $a' < b'$ and a linear isometry T from $\mathbb{H}$ to $\mathbb{H}' = L^2[a', b']$ such that, for all $k \geq 1$, $T\varphi_k$ is a polynomial of degree $k-1$. We denote by T^{-1} the inverse isometry.

To compute the coefficients $Q_{k,j}$ under Assumption 3, one may rely on the well known Legendre polynomials which form an orthogonal sequence of polynomials in $\mathbb{H}' = L^2[a', b']$. Indeed, by choosing g_k so that $T\varphi_k$ is the polynomial t^{k-1}, as will be illustrated in all the examples below, the constants $Q_{k,j}$ are the coefficients of the normalized Legendre polynomials $\sum_{j=1}^k Q_{k,j} t^{j-1}$. Let us recall the definition of the Legendre polynomials.

Definition 1 (Legendre polynomials)**.** Let $a' < b'$ be two real numbers and denote $\mu = (a' + b')/2$ and $\delta = (b' - a')/2$. The *Legendre polynomials* associated to the interval $[a', b']$ are defined as the polynomials $r_k(t) = \sum_{l=1}^k R_{k,l} t^{l-1}$, where the coefficients $R_{k,l}$ are given by the following recurrence relation

$$R_{k+1,l} = R_{k,l-1} + \mu R_{k,l} - \beta_k R_{k-1,l} \ , \quad \text{for all } k, l \geq 1 \ ,$$

with $R_{1,1} = 1$ and $R_{k,l} = 0$ for all $l > k$, $\beta_1 = 2\delta$ and $\beta_k = \delta^2(k-1)^2/(4(k-1)^2 - 1)$ for $k \geq 2$. The obtained sequence $(r_k)_{k\geq 1}$ is orthogonal in $\mathbb{H}' = L^2([a', b'])$ with norms given by $\|r_k\|_{\mathbb{H}'} = \sqrt{\beta_1 \dots \beta_k}$. Hence, the coefficients of the *normalized Legendre polynomials* are defined by the relation

$$Q_{k,l} = \frac{R_{k,l}}{\sqrt{\beta_1 \dots \beta_k}} \ , \quad \text{for all } k, l \geq 1 \ . \tag{5}$$

2.2. **Examples.** For illustration we exhibit in this section the orthogonal series estimator in some special cases. Some scale mixtures are presented. As an example for a non scale mixture we also consider Gamma shape mixtures.

Example 1 (a). Exponential Mixture. We first consider continuous exponential mixtures as they play a meaningful role in physics. That is, we consider $\pi_t(x) = t\mathrm{e}^{-tx}$. For the orthogonal series estimator we

choose the functions $g_k(x) = 1\left\{x > k - \frac{1}{2}\right\}$ for $k \geq 1$. By (2), we obtain

$$\varphi_k(t) = \mathrm{e}^{-(k-\frac{1}{2})t} \ . \tag{6}$$

We claim that the φ_k's can be transformed into polynomials in the space $\mathbb{H}' = L^2[\mathrm{e}^{-b}, \mathrm{e}^{-a}]$. Indeed, define, for all $f \in \mathbb{H} = L^2[a,b]$,

$$Tf(t) = f(-\log t)/\sqrt{t} \ , \quad t \in [\mathrm{e}^{-b}, \mathrm{e}^{-a}] \ . \tag{7}$$

Then one has $\langle Tf, Tg\rangle_{\mathbb{H}'} = \langle f, g\rangle_{\mathbb{H}}$, hence T is an isometry from $\mathbb{H}$ to $\mathbb{H}'$. Moreover $T\varphi_k(t) = t^{k-1}$ are polynomials. Denote by $p_k(t) = \sum_{j=1}^k Q_{k,j} t^{j-1}$ the Legendre polynomials in $\mathbb{H}'$ with coefficients $Q_{k,j}$ defined by (5) with $a' = \mathrm{e}^{-b}$ and $b' = \mathrm{e}^{-a}$. Denote by T^{-1} the inverse operator of T given by $T^{-1}h(t) = \mathrm{e}^{-t/2}h(\mathrm{e}^{-t})$. Since T^{-1} is a linear isometry, we get that the functions $\psi_k \equiv T^{-1}p_k = \sum_{j=1}^k Q_{k,j}\varphi_j$ are orthonormal in $\mathbb{H}$. Consequently, an orthonormal series estimator is given by

$$\hat{f}_{m,n}(t) = \frac{1}{n}\sum_{k,j,l=1}^{m}\sum_{i=1}^{n} 1\left\{X_i > j - \frac{1}{2}\right\} Q_{k,j}Q_{k,l}\mathrm{e}^{-(l-\frac{1}{2})t} \ . \tag{8}$$

Example 1 (b). Exponential Mixture. The choice of the functions g_k is not unique and needs to be done with care. For illustration, consider once again exponential mixtures with $\pi_t(x) = t\mathrm{e}^{-tx}$. This time we take

$$g_k(x) = a_k x^k \quad \text{with } a_k = \left(\int x^k \pi_1(\mathrm{d}x)\right)^{-1} = 1/k!$$

and hence $\varphi_k(t) = t^{-k}$, for $k \geq 1$. To relate φ_k to polynomials, define the isometry $\tilde{T}$ from $\mathbb{H}$ to $\tilde{\mathbb{H}} = L^2[1/b, 1/a]$ by $\tilde{T}f(t) = \frac{1}{t}f(\frac{1}{t})$. We have $\tilde{T}\varphi_k(t) = t^{k-1}$ for all $k \geq 1$. Furthermore, denote by $\tilde{T}^{-1}$ the inverse of $\tilde{T}$ satisfying $\tilde{T}^{-1}h = \frac{1}{t}h(\frac{1}{t})$. Let $\tilde{p}_k(t) = \sum_{j=1}^k Q_{k,j}t^{j-1}$ be the Legendre polynomials in $\tilde{\mathbb{H}}$ defined with $a' = 1/b$ and $b' = 1/a$. Since $\tilde{T}^{-1}$ is an isometry, $\psi_k \equiv \tilde{T}^{-1}p_k = \sum_{j=1}^k Q_{k,j}\varphi_j$ are orthonormal functions in $\mathbb{H}$ and the orthonormal series estimator is given by

$$\hat{f}_{m,n}(t) = \frac{1}{n}\sum_{k,j,l=1}^{m}\sum_{i=1}^{n} \frac{Q_{k,j}Q_{k,l}X_i^j}{j!}t^{-l} \ .$$

Example 2. Gamma Shape Mixture. Polynomial estimators can be used in the context where the mixed parameter is not necessarily a scale parameter. As pointed out earlier, they have first been used for mixtures on a discrete state space X, such as Poisson mixtures, see [14] and [22]. Let us consider the Gamma shape mixture model. Parametric Gamma shape mixtures have been considered in [25]. For this model π_t is the Gamma density with shape parameter t and a fixed scale parameter (here set to 1 for simplicity),

$$\pi_t(x) = \frac{x^{t-1}}{\Gamma(t)}\,\mathrm{e}^{-x}, \quad t \in [a,b] \ ,$$

where Γ denotes the Gamma function. This model has a continuous state space (ζ is the Lebesgue measure on $\mathbb{R}_+$) and is not a scale mixture. We shall construct g_k and $\varphi_k = \pi_\cdot(g_k)$ such that Assumption 3 holds with T being the identity and $\varphi_k(t) = t^{k-1}$. Consider the following sequence of polynomials, $p_1(t) = 1$, $p_2(t) = t$, ..., $p_k(t) = t(t+1)\ldots(t+k-2)$ for all $k \geq 2$. Since $(p_k)_{k\geq 1}$ is a sequence of polynomials with degrees $k-1$, there are coefficients $(\tilde{c}_{k,l})_{1\leq l\leq k}$ such that $t^{k-1} = \sum_l \tilde{c}_{k,l}p_l(t)$ for $k = 1, 2, \ldots$. A simple recursive formula for computing $(\tilde{c}_{k,l})_{1\leq l\leq k}$ is provided in Lemma 6 in the Appendix, see Eq. (48). Observe that, for any $l \geq 1$,

$$\int x^{l-1}\,\pi_t(x)\mathrm{d}x = \frac{\Gamma(t+l-1)}{\Gamma(t)} = p_l(t) \ .$$

Hence, setting $g_k(x) = \sum_l \tilde{c}_{k,l} x^{l-1}$, we obtain

$$\varphi_k(t) = \pi_t(g_k) = \sum_l \tilde{c}_{k,l} p_l(t) = t^{k-1} ,$$

and thus Assumption 3 holds with T being the identity operator and $\varphi_k(t) = t^{k-1}$. Define $(Q_{k,l})_{k,l}$ as the coefficients of Legendre polynomials on $\mathbb{H} = L^2([a,b])$, that is as in (5) with $a' = a$ and $b' = b$. The polynomial estimator defined by (8) reads

$$\hat{f}_{m,n}(t) = \frac{1}{n} \sum_{i=1}^{n} \sum_{k,j,l=1}^{m} Q_{k,j} Q_{k,l} \sum_{h=1}^{j} \tilde{c}_{j,h} X_i^{h-1} t^{l-1} .$$

Example 3. Scale Mixture of Beta Distributions or Uniform Distributions. It is well known that any k-monotone density, for $k \geq 1$, can be represented by a scale mixture of Beta distributions $B(1,k)$ [3] with

$$\pi_t(x) = \frac{k}{t} \left(1 - \frac{x}{t}\right)^{k-1} , \quad \text{for } x \in [0,t] .$$

Note that if $k = 1$, then π_t is the uniform density $U(0,t)$. We take

$$g_p(x) = a_p x^{p-1} \quad \text{with } a_p = \left(\int x^{p-1} \pi_1(\mathrm{d}x) \right)^{-1} = \frac{1}{k\, \beta(p,k)}, \quad p \geq 1 ,$$

where $\beta(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}\mathrm{d}t$ denotes the Beta function. It follows that $\varphi_p(t) = t^{p-1}$. As in the preceding example, if $f \in \mathbb{H}$ then an orthogonal series estimator $\hat{f}_{m,n}$ of f can be constructed by using Legendre polynomials $p_k(t) = \sum_{j=1}^{k} Q_{k,j} t^{j-1}$ where the coefficients $Q_{k,j}$ are defined as in (5) with $a' = a$ and $b' = b$. Then according to (3), the corresponding orthogonal series estimator is given by

$$\hat{f}_{m,n}(t) = \frac{1}{n} \sum_{j,p,l=1}^{m} \sum_{i=1}^{n} Q_{p,j} Q_{p,l} \frac{X_i^{j-1}}{k\beta(j,k)} t^{l-1} .$$

In Example 1 (b) we considered the same functions g_p but here Assumption 3 holds with T equal to the identity operator on $\mathbb{H}$. This difference relies on the parametrization of the exponential family by the inverse of the scale parameter.

Example 4. Mixture of exponential distributions with location parameter. The estimator also applies to the deconvolution setting. As an example, consider $X = Y + \theta$ where Y and θ are independent random variables, Y has exponential distribution with mean 1 and θ has unknown density f supported on $[a,b]$. The density of X is given by $\pi_f(x) = \int_a^b \pi_t(x) f(t)\mathrm{d}t$ with $\pi_t(x) = \mathrm{e}^{-(x-t)} 1\{x > t\}$. Let $g_1(x) = 1$ and

$$g_k(x) = x^{k-1} - (k-1)x^{k-2} , \tag{9}$$

for $k \geq 2$. Then $\varphi_k(t) = t^{k-1}$ for $k \geq 1$. The estimator $\hat{f}_{m,n}$ of f is then given by

$$\hat{f}_{m,n}(t) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j,k,l=1}^{m} Q_{k,l} Q_{k,j} g_j(X_i) t^{l-1} ,$$

where $Q_{k,l}$ are the Legendre coefficients defined by (5) with $a' = a$ and $b' = b$.

3. Analysis of the Orthogonal Series Estimator

In this section the properties of the orthogonal series estimator are analyzed.

3.1. **Bias, Variance and MISE.** It is useful to write the orthogonal series estimator $\hat{f}_{m,n}$ defined in (3) in matrix notation. Therefore, we introduce the $m \times m$–matrix $Q = (Q_{k,j})_{k,j}$, where $Q_{k,j} = 0$ for all $j > k$, and the m–vectors

$$\Phi = [\varphi_1, \ldots, \varphi_m]^T \ , \quad \Psi = [\psi_1, \ldots, \psi_m]^T = Q\Phi \ ,$$
$$\mathbf{g}(x) = [g_1(x), \ldots, g_m(x)]^T \ , \quad \hat{\mathbf{g}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{g}(X_i) \ ,$$
$$\mathbf{c} = [c_1, \ldots, c_m]^T = \langle \Psi, f \rangle_{\mathbb{H}} \ , \quad \hat{\mathbf{c}} = [\hat{c}_{n,1}, \ldots, \hat{c}_{n,m}]^T = Q\hat{\mathbf{g}} \ .$$

It follows that the orthogonal series estimator can be written as

$$\hat{f}_{m,n} = \hat{\mathbf{c}}^T \Psi = \hat{\mathbf{g}}^T Q^T Q \Phi \ .$$

Further, let $\Sigma = \pi_f(\mathbf{g}\mathbf{g}^T) - \pi_f(\mathbf{g})\pi_f(\mathbf{g})^T$ be the covariance matrix of $\mathbf{g}(X_1)$. The MISE is defined by $\mathbb{E}\left\|\hat{f}_{m,n} - f\right\|_{\mathbb{H}}^2$. The orthogonal projection of f on V_m is denoted by

$$P_{V_m} f = \mathbf{c}^T \Psi = \sum_{k=1}^{m} c_{n,k} \psi_k \ .$$

It is clear that the orthogonal series estimator $\hat{f}_{m,n}$ is an unbiased estimator of $P_{V_m} f$. Furthermore, by the usual argument, the MISE is decomposed into two terms representing the integrated variance and integrated squared bias, as summarized in the following result, whose proof is standard and thus omitted.

Proposition 1. *Suppose that Assumption 2 holds. The orthogonal series estimator $\hat{f}_{m,n}$ defined in (3) satisfies*

(i) For every $t \in [a,b]$, $\mathbb{E}[\hat{f}_{m,n}(t)] = P_{V_m} f(t)$.
(ii) For every $t \in [a,b]$, $\mathrm{Var}(\hat{f}_{m,n}(t)) = \frac{1}{n}\Psi^T(t) Q \Sigma Q^T \Psi(t)$.
(iii) $\mathbb{E}\left\|\hat{f}_{m,n} - f\right\|_{\mathbb{H}}^2 = \|P_{V_m} f - f\|_{\mathbb{H}}^2 + \frac{1}{n}\mathrm{tr}\left(Q\Sigma Q^T\right)$.

An important issue for orthogonal series estimators $\hat{f}_{m,n}$ is the choice of the approximation order m. The integrated squared bias $\|P_{V_m} f - f\|_{\mathbb{H}}^2$ only depends on how well $P_{V_m} f$ approximates f, whose rate of convergence depends on the smoothness class to which belongs the density f. To be more precise, define for any approximation rate index α and radius C, the approximation class

$$\mathcal{C}(\alpha, C) = \{f \in \mathbb{H} : \|f\|_{\mathbb{H}} \leq C \text{ and } \|P_{V_m} f - f\|_{\mathbb{H}} \leq C\, m^{-\alpha} \text{ for all } m \geq 1\} \ . \tag{10}$$

So when the mixing density f belongs to $\mathcal{C}(\alpha, C)$, then the bias of the orthogonal series estimator $\hat{f}_{m,n}$ is well controlled, namely it decreases at the rate $m^{-\alpha}$ as m increases. Furthermore, denote the set of densities in $\mathbb{H}$ by $\mathbb{H}_1 = \{f \in \mathbb{H} : f \geq 0, \int_a^b f(t)\mathrm{d}t = 1\}$. We will investigate the rate of convergence of $\hat{f}_{m,n}$ in $\mathbb{H}$ when $f \in \mathcal{C}(\alpha, C) \cap \mathbb{H}_1$. We will obtain the best achievable rate in the case of exponential mixtures and almost the best one in the case of Gamma shape mixtures.

3.2. **Upper Bound of the MISE.** We now provide an upper bound of the MISE for the orthogonal series estimator based on Legendre polynomials, that is, when Assumption 3 holds.

To show an upper bound of the MISE we use the following property [see 22, Lemma A.1]. If $\lambda > \frac{2+a'+b'}{b'-a'} + \sqrt{1 + \frac{2+a'+b'}{b'-a'}}$, then the coefficients of the normalized Legendre polynomials in $L^2[a', b']$

defined by (5) verify

$$\sum_{l=1}^{k} Q_{k,l}^2 = O(\lambda^{2k}) \ , \quad \text{as } k \to \infty \ . \tag{11}$$

By combining Proposition 1 (iii) and the bound given in (11) along with a normalization condition on the g_k's (Condition (12) or Condition (15) below), we obtain the following asymptotic upper bounds of the MISE.

Theorem 1. *Let α be a positive rate index and C be a positive radius. Suppose that Assumption 3 holds with $f \in \mathcal{C}(\alpha, C) \cap \mathbb{H}_1$. Let $\hat{f}_{m,n}$ be defined by (3) with Legendre polynomials coefficients $Q_{k,j}$ given by (5). Then the two following assertions hold.*

(a) If, for some constants $C_0 > 0$ and $B \geq 1$, we have

$$\mathrm{Var}(g_k(X)) < C_0 \, B^{2k} \quad \textit{for all } k \geq 1 \ . \tag{12}$$

Set $m_n = A \log n$ with

$$A < \frac{1}{2} \left\{ \log B + \log \left(\frac{2 + a' + b'}{b' - a'} + \sqrt{1 + \frac{2 + a' + b'}{b' - a'}} \right) \right\}^{-1} \ . \tag{13}$$

Then, as $n \to \infty$,

$$\mathbb{E} \left\| \hat{f}_{m_n,n} - f \right\|_{\mathbb{H}}^2 \leq C^2 m_n^{-2\alpha} (1 + o(1)) \ , \tag{14}$$

where the o-term only depends on the constants α, C, a', b', A and C_0.

(b) If, for some constants $C_0 > 0$ and $\eta > 0$, we have

$$\mathrm{Var}(g_k(X)) < C_0 \, k^{\eta k} \quad \textit{for all } k \geq 1 \ . \tag{15}$$

Set $m_n = A \log n / \log \log n$ with $A < \eta^{-1}$. Then, as $n \to \infty$,

$$\mathbb{E} \left\| \hat{f}_{m_n,n} - f \right\|_{\mathbb{H}}^2 \leq C^2 m_n^{-2\alpha} (1 + o(1)) \ , \tag{16}$$

where the o-term only depends on the constants α, C, a', b', A and C_0.

Remark 1. The larger A, the lower the upper bound in (14). Hence, since a', b' and B directly depend on the g_k's, the constraint (13) on A indicates how appropriate the choice of the g_k's is.

Remark 2. In the examples treated in this paper, C_0 and B or η can be chosen independently of $f \in \mathcal{C}(\alpha, C) \cap \mathbb{H}_1$. Consequently, the bounds given in (14) and (15) show that $\hat{f}_{m_n,n}$ achieves the MISE rates $(\log n)^{-2\alpha}$ and $(\log(n)/\log\log n)^{-2\alpha}$, respectively, uniformly on $f \in \mathcal{C}(\alpha, C) \cap \mathbb{H}_1$. In the exponential mixture case, we show below that $\hat{f}_{m_n,n}$ of Example 1(a) is minimax rate adaptive in these classes (since m_n does not depend on α). In the Gamma shape mixture case, we could only show that $\hat{f}_{m_n,n}$ of Example 2 is minimax rate adaptive in these classes up to the multiplicative $\log\log n$ term.

Proof. We first consider Case (a). By (13), we may choose a number λ strictly lying between $\frac{2+a'+b'}{b'-a'} + \sqrt{1+\frac{2+a'+b'}{b'-a'}}$ and $\mathrm{e}^{1/(2A)}/B$. Note that from Condition (12), it follows by the Cauchy-Schwarz inequality that $|\Sigma_{k,l}| = |\mathrm{Cov}(g_k(X), g_l(X))| \leq C_0 B^k B^l$ for all k, l. Thus, we obtain

$$\begin{aligned}
\mathrm{tr}\left(Q\Sigma Q^T\right) &\leq C_0 \sum_{k=1}^{m}\sum_{j=1}^{k}\sum_{l=1}^{k} |Q_{k,j}Q_{k,l}|\, B^j B^l \\
&\leq C_0 \sum_{k=1}^{m}\left(\sum_{j=1}^{k} Q_{k,j}^2 \sum_{j=1}^{k} B^{2j}\right) \\
&\leq Km\{B\lambda\}^{2m}\,,
\end{aligned}$$

where the last inequality comes from (11) and K is a positive constant (the multiplicative term m is necessary only for $B = 1$). It follows by the decomposition of the MISE in Proposition 1 (iii) that

$$\begin{aligned}
\mathbb{E}\left\|\hat{f}_{m_n,n} - f\right\|_{\mathbb{H}}^2 &\leq C^2 m_n^{-2\alpha} + Kn^{-1}m_n(B\lambda)^{2m_n} \\
&\leq C^2 m_n^{-2\alpha}\left(1 + \frac{K}{C^2}n^{-1}m_n^{2\alpha+1}(B\lambda)^{2m_n}\right)\,.
\end{aligned}$$

Now we have for $m_n = A\log n$ that

$$n^{-1}m_n^{2\alpha+1}(B\lambda)^{2m_n} = A^{2\alpha+1}(\log n)^{2\alpha+1}n^{2A\log B\lambda - 1} = o(1)\,,$$

since $A < 1/(2\log B\lambda)$.

Let us now consider Case (b). Proceeding as above, for any $\lambda > \frac{2+a'+b'}{b'-a'} + \sqrt{1+\frac{2+a'+b'}{b'-a'}}$, we get $\mathrm{tr}\left(Q\Sigma Q^T\right) \leq K\, C_0 \lambda^{2m} m^{1+\eta m}$, which yields

$$\mathbb{E}\left\|\hat{f}_{m_n,n} - f\right\|_{\mathbb{H}}^2 \leq C^2 m_n^{-2\alpha}\left(1 + \frac{K}{C^2}n^{-1}m_n^{2\alpha+1+\eta m_n}\lambda^{2m_n}\right)\,.$$

To conclude, it suffices to check that the log of the second term between parentheses tends to $-\infty$ as $n \to \infty$ for $m_n = A\log n/\log\log n$ with $A < \eta^{-1}$, which is easily done. □

Let us check the validity of Condition (12) or Condition (15) for the above examples.

Example 1 (a). Exponential Mixture (continued). Condition (12) immediately holds with $B = C_0 = 1$ for the exponential mixture of Example 1(a) since $g_k(x) = 1\left\{x > k - \frac{1}{2}\right\}$.

Example 1 (b). Exponential Mixture (continued). Interestingly, Condition (12) does not hold for Example 1(b), where a different choice of g_k's is proposed. In fact, one finds that $\log \mathrm{Var}(g_k(X))$ is of order $k\log(k)$. Hence, only Condition (15) holds and we fall in case (b) of Theorem 1. Since a slower rate is achieved in this case, this clearly advocates to choose the estimator obtained in Example 1(a) rather than the one in Example 1(b) for the exponential mixture model.

Example 2. Gamma Shape Mixture (continued). We recall that here we set $g_k(t) = \sum_{l=1}^{k}\tilde{c}_{k,l}t^{l-1}$ where the coefficients $(\tilde{c}_{k,l})$ are those defined and computed in Lemma 6 of the Appendix. Using the bound given by (49) in the same lemma, we obtain that $g_k(x) \leq k!(1 \vee |x|^{k-1})$. It follows that $\pi_t(g_k^2) \leq (k!)^2(1 + \Gamma(t+2k-2)/\Gamma(t))$, and, for any $f \in \mathbb{H}_1$, $\pi_f(g_k^2) \leq (k!)^2(1 + \Gamma(b+2k-2)/\Gamma(b))$. Hence, by Stirling's formula, we find that Condition (15) holds for $\eta = 4$ and some C_0 independent of $f \in \mathbb{H}_1$.

Example 3. Scale Mixture of Beta Distributions or Uniform Distributions (continued). We now verify Condition (12) for Beta mixtures and the g_p of Example 3. Note that we can write $X = \theta X_0$ with

independent random variables $\theta \sim f$ and $X_0 \sim B(1,k)$. We have for all $p \geq 1$

$$\mathrm{Var}(g_p(X)) \leq \frac{\mathbb{E}[X^{2p-2}]}{k^2\beta^2(p,k)} = \frac{\mathbb{E}[\theta^{2p-2}]\mathbb{E}[X_0^{2p-2}]}{k^2\beta^2(p,k)} \leq \frac{b^{2p-2}}{k^2\beta^2(p,k)} \leq b^{2p-2}k^{2p-2} \ .$$

Hence Condition (12) holds with $B = k$ if $b < 1$, with $B = bk$ if $b \geq 1$.

A close inspection of Example 3 indicates that it is a particular case of the following more general result concerning mixtures of compactly supported scale families.

Lemma 1. *Suppose that Assumption 1 holds in the context of a scale mixture on $\mathbb{R}_+$, that is, ζ is the Lebesgue measure on $\mathbb{R}_+$ and $\pi_t = t^{-1}\pi_1(t^{-1}\cdot)$ for all $t \in \Theta = [a,b] \subset (0,\infty)$. Assume in addition that π_1 is compactly supported in $\mathbb{R}_+$. Define, for all $k \geq 1$,*

$$g_k(x) = \left(\int x^{k-1}\pi_1(x)\,\mathrm{d}x \right)^{-1} x^{k-1} \ .$$

Then Assumption 2 holds with $\varphi_k(t) = t^{k-1}$, and thus also does Assumption 3 with T being the identity operator on $L^2([a,b])$. Moreover there exists C_0 and B only depending on π_1 and b such that Condition (12) holds.

Proof. Using the assumptions on π_1 and Jensen's inequality, we have

$$B_1^m \leq \int x^m \ \pi_1(x)\,\mathrm{d}x \leq B_2^m \quad \text{for all } m \geq 1 \ ,$$

with $B_1 = \int x \ \pi_1(x)\,\mathrm{d}x$ and $B_2 > 0$ such that the support of π_1 is included in $[0, B_2]$. The result then follows from the same computations as in Examples 3. □

An immediate consequence of Theorem 1 and Lemma 1 is the following.

Corollary 1. *Under the assumptions of Lemma 1, the estimator $\hat{f}_{m,n}$ defined by (3) with Legendre polynomials coefficients $Q_{k,j}$ given by (5) achieves the MISE rate $(\log n)^{-2\alpha}$ uniformly on $f \in \mathcal{C}(\alpha, C) \cap \mathbb{H}_1$ for any $\alpha > 0$ and $C > 0$.*

Example 4. Exponential mixture with location parameter (continued). One can show that Condition (15) of Theorem 1 is satisfied, so that the rate of the MISE of the estimator is $(\log(n)/\log\log n)^{-2\alpha}$. Indeed,

$$\mathbb{E}[X^r] = r! \int_a^b f(t) \sum_{j=0}^{r} \frac{t^j}{j!}\mathrm{d}t \leq r! \sum_{j=0}^{r} \frac{b^j}{j!} \leq r!\mathrm{e}^b \ ,$$

and thus, using the definition of g_k in (9), $\mathrm{Var}(g_k(X)) \leq 2(2k-2)!\mathrm{e}^b \approx 2\sqrt{2\pi}\mathrm{e}^{b-2k+2}(2k-2)^{2k-3/2}$.

4. Approximation Classes

Although the approximation classes $\mathcal{C}(\alpha, C)$ appear naturally when studying the bias of the orthogonal series estimator defined in (3), it is legitimate to ask whether such classes can be interpreted in a more intuitive way, say using a smoothness criterion. This section provides a positive answer to this question.

4.1. **Weighted Moduli of Smoothness.** Let us recall the concept of weighted moduli of smoothness as introduced by Ditzian and Totik [8] for studying the rate of polynomial approximations. For $a < b$ in $\mathbb{R}$, $f : [a, b] \to \mathbb{R}$, $r \in \mathbb{N}^*$ and $h \in \mathbb{R}$ denote by $\Delta_h^r(f, \cdot)$ the *symmetric difference* of f of order r with step h, that is

$$\Delta_h^r(f, x) = \sum_{i=0}^{r} \binom{r}{i} (-1)^i f(x + (i - r/2)h) \ . \tag{17}$$

with the convention that $\Delta_h^r(f, x) = 0$ if $x \pm mh/2 \notin [a, b]$. Define the step-weight function φ on the bounded interval $[a, b]$ as $\varphi(x) = \sqrt{(x - a)(b - x)}$. Then for $f : [a, b] \to \mathbb{R}$ the *weighted modulus of smoothness* of f of order r and with the step-weight function φ in the $L^p([a, b])$ norm is defined as

$$\omega_\varphi^r(f, t)_p = \sup_{0<h\le t} \|\Delta_{h\varphi(\cdot)}^r(f, \cdot)\|_p \ .$$

We recall an equivalence relation of the modulus of smoothness with the so-called K*-functional*, which is defined as

$$K_{r,\varphi}(f, t^r)_p = \inf_h \{\|f - h\|_p + t^r \|\varphi^r h^{(r)}\|_p \ : \ h^{(r-1)} \in A.C._{\text{loc}}\} \ , \tag{18}$$

where $h^{(r-1)} \in A.C._{\text{loc}}$ means that h is $r - 1$ times differentiable and $h^{(r-1)}$ is absolutely continuous on every closed finite interval. If $f \in L^p([a, b])$, then

$$M^{-1}\omega_\varphi^r(f, t)_p \le K_{r,\varphi}(f, t^r)_p \le M\omega_\varphi^r(f, t)_p \ , \quad \text{for } t \le t_0 \ , \tag{19}$$

for some constants M and t_0, see Theorem 6.1.1. in Ditzian and Totik [8].

4.2. **Equivalence Result.** We show that the classes $\mathcal{C}(\alpha, C)$ are equivalent to classes defined using weighted moduli of smoothness. This, in turn, will relate them to Sobolev and Hölder classes. To make this precise, we define for constants $\alpha > 0$ and $C > 0$ the following class of functions in $\mathbb{H} = L^2([a, b])$

$$\tilde{\mathcal{C}}(\alpha, C) = \{f \in \mathbb{H} : \|f\|_{\mathbb{H}} \le C \text{ and } \omega_\varphi^r(f, t)_2 \le Ct^\alpha \text{ for all } t > 0\} \ , \tag{20}$$

where $\varphi(x) = \sqrt{(x - a)(b - x)}$ and $r = [\alpha] + 1$.

The following theorem states the equivalence of the classes $\mathcal{C}(\alpha, C)$ and $\tilde{\mathcal{C}}(\alpha, C)$. This result is an extension of Proposition 7 in Roueff and Ryden [22] to the case where the subspaces V_m correspond to transformed polynomial classes through an isometry T which includes both a multiplication and a composition with smooth functions.

Theorem 2. *Let $\alpha > 0$. Suppose that Assumption 3 holds with a linear isometry $T : \mathbb{H} = L^2([a, b]) \to \mathbb{H}' = L^2([a', b'])$ given by $Tg = \sigma \times g \circ \tau$, where σ is non-negative and $[\alpha] + 1$ times continuously differentiable, and τ is $[\alpha] + 1$ times continuously differentiable with a non-vanishing first derivative. Then for any positive number α, there exist positive constants C_1 and C_2 such that for all $C > 0$*

$$\mathcal{C}(\alpha, C_1 C) \subset \tilde{\mathcal{C}}(\alpha, C) \subset \mathcal{C}(\alpha, C_2 C) \ . \tag{21}$$

where $\tilde{\mathcal{C}}(\alpha, C)$ is defined in (20) and $\mathcal{C}(\alpha, C')$ is defined in (10) with approximation classes (V_m) given by (4).

For short, we write $\mathcal{C}(\alpha, \cdot) \hookrightarrow \tilde{\mathcal{C}}(\alpha, \cdot)$ when there exists $C_1 > 0$ such that the first inclusion in (21) holds for all $C > 0$. The validity of both inclusions is denoted by the equivalence $\mathcal{C}(\alpha, \cdot) \asymp \tilde{\mathcal{C}}(\alpha, \cdot)$.

Proof of Theorem 2. Weighted moduli of smoothness are used to characterize the rate of polynomial approximations. We start by relating $\mathcal{C}(\alpha, C)$ to classes defined by the rate of polynomial approximations, namely

$$\bar{\mathcal{C}}(\alpha, C) = \{g \in \mathbb{H}' : \|g\|_{H'} \leq C \text{ and } \inf_{p \in \mathcal{P}_{m-1}} \|g - p\|_{H'} \leq C m^{-\alpha}, \text{ for all } m \geq 1\} ,$$

where $\mathcal{P}_m$ is the set of polynomials of degree at most m. Indeed, we see that, since T is a linear isometry,

$$\begin{aligned}\mathcal{C}(\alpha, C) &= \{f \in \mathbb{H} : \|f\|_H \leq C \text{ and } \|P_{V_m} f - f\|_H \leq C m^{-\alpha} \text{ for all } m \geq 1\} \\ &= \{T^{-1} g : g \in \mathbb{H}', \|g\|_{H'} \leq C \text{ and } \|P_{TV_m} g - g\|_{H'} \leq C m^{-\alpha} \text{ for all } m \geq 1\} \\ &= T^{-1} \bar{\mathcal{C}}(\alpha, C) .\end{aligned}$$

As stated in Corollary 7.25 in Ditzian and Totik [8], we have the equivalence $\bar{\mathcal{C}}(\alpha, \cdot) \asymp \tilde{\mathcal{C}}'(\alpha, \cdot)$, where $\tilde{\mathcal{C}}'(\alpha, C)$ is defined as $\tilde{\mathcal{C}}(\alpha, C)$ but with a' and b' replacing a and b. Hence, it only remains to show that

$$T^{-1} \tilde{\mathcal{C}}'(\alpha, \cdot) \asymp \tilde{\mathcal{C}}(\alpha, \cdot) . \tag{22}$$

To show this, we use the assumed particular form of T, that is $T(g) = \sigma \times g \circ \tau$. Since T is an isometry from $\mathbb{H} = L^2([a, b])$ to $\mathbb{H}' = L^2([a', b'])$ and σ is non-negative, we necessarily have that τ is a bijection from $[a', b']$ to $[a, b]$ (whose inverse bijection is denoted by τ^{-1}) and $\sigma = 1/\sqrt{\tau' \circ \tau^{-1}}$. Moreover the inverse isometry writes $T^{-1}(g) = (\sigma \circ \tau^{-1})^{-1} \times g \circ \tau^{-1}$. From the assumptions on τ we have that σ, $(\sigma \circ \tau^{-1})^{-1}$, τ and τ^{-1} all are $[\alpha]+1$ times continuously differentiable and the two latter's first derivative do not vanish. The equivalence (22) then follows by Lemma 5 in the appendix. □

Example 1 (a). Exponential Mixture (continued). In Example 1(a) of continuous exponential mixtures, the operator T is given by (7), that is $\sigma(t) = 1/\sqrt{t}$ and $\tau(t) = -\log t$ and further $\mathbb{H}' = L^2(\mathrm{e}^{-b}, \mathrm{e}^{-a})$. Both σ and τ are infinitely continuously differentiable on $[a, b]$ if $a > 0$, and thus the equivalence given in (21) holds.

Example 1 (b). Exponential Mixture (continued). For the estimator exhibited in Example 1(b) for exponential mixtures, the isometry T is such that $\sigma(t) = \tau(t) = 1/t$ with $a' = 1/b$ and $b' = 1/a$. Hence, the conclusion of Theorem 2 holds if $a > 0$.

Example 2, 3 and 4. Gamma Shape Mixture, Scale Mixture of Beta Distributions and Exponential mixture with location parameter (continued). In the cases of Example 2, 3 and 4, the transform T is the identity and hence Theorem 2 applies. However, this result is also obtained by Corollary 7.25 in Ditzian and Totik [8].

5. Lower Bound of the Minimax Risk

Our goal in this section is to find a lower bound of the minimax risk

$$\inf_{\hat{f} \in \mathcal{S}_n} \sup_{f \in \mathcal{C}} \pi_f^{\otimes n} \|\hat{f} - f\|_{\mathbb{H}}^2 ,$$

where $\mathcal{S}_n$ is the set of all Borel functions from $\mathbb{R}^n$ to $\mathbb{H}$, $\mathcal{C}$ denotes a subset of densities in $\mathbb{H}_1$ and $\pi_f^{\otimes n}$ denotes the joint distribution of the sample $(X_1, \ldots, X_n)$ under Assumption 1. We first provide a general lower bound, which is then used to investigate the minimax rate in the specific cases of exponential mixtures, Gamma shape mixtures and mixtures of compactly supported scale families.

5.1. **A General Lower Bound for Mixture Densities.** We now present a new lower bound for the minimax risk of mixture density estimation. As in Proposition 2 in [22], it relies on the mixture structure. However, in contrast with this previous result which only applies for mixtures of discrete distributions, we will use the following lower bound in the case of mixtures of exponential distributions, Gamma shape mixtures and scale mixtures of compactly supported densities.

Theorem 3 (Lower bound)**.** *Let $f_0 \in \mathbb{H}_1$ and $f_* \in \mathbb{H}$ with $\|f_*\|_{\mathbb{H}} \leq 1$ and $f_0 \pm f_* \in \mathbb{H}_1$ the following lower bound holds, for any $c \in (0,1)$,*

$$\inf_{\hat{f} \in \mathcal{S}_n} \sup_{f \in \{f_0, f_0 \pm f_*\}} \pi_f^{\otimes n} \|f - \hat{f}\|_{\mathbb{H}}^2 \geq c\|f_*\|_{\mathbb{H}}^2 - \frac{c}{(1-c)^2} \left(\left(1 + \int |\pi_{f_*}(x)| \, \zeta(\mathrm{d}x) \right)^n - 1 \right) , \tag{23}$$

where $\pi_f^{\otimes n}$ denotes the joint distribution of the sample $(X_1, \ldots, X_n)$ under Assumption 1.

Proof. Let f_* be as in the Theorem. We define for a fixed $\hat{f} \in \mathcal{S}_n$ and any $c \in (0,1)$ the set $A = \{\|f_0 - \hat{f}\|_{\mathbb{H}} \leq \frac{c}{1-c}\}$. Then, for all $\hat{f} \in \mathcal{S}_n$, $\sup_{f \in \{f_0, f_0 \pm f_*\}} \pi_f^{\otimes n} \|f - \hat{f}\|_{\mathbb{H}}^2$ is bounded from below by

$$\begin{aligned}
&\frac{c}{2}\pi_{f_0+f_*}^{\otimes n}\|f_0 + f_* - \hat{f}\|_{\mathbb{H}}^2 + \frac{c}{2}\pi_{f_0-f_*}^{\otimes n}\|f_0 - f_* - \hat{f}\|_{\mathbb{H}}^2 + (1-c)\pi_{f_0}^{\otimes n}\|f_0 - \hat{f}\|_{\mathbb{H}}^2 \\
&\geq \frac{c}{2}\pi_{f_0+f_*}^{\otimes n}\left[1_A\|f_0 + f_* - \hat{f}\|_{\mathbb{H}}^2\right] \\
&\qquad + \frac{c}{2}\pi_{f_0-f_*}^{\otimes n}\left[1_A\|f_0 - f_* - \hat{f}\|_{\mathbb{H}}^2\right] + (1-c)\pi_{f_0}^{\otimes n}\|f_0 - \hat{f}\|_{\mathbb{H}}^2 .
\end{aligned}$$

Note that for a function k defined on $\mathbb{R}^n$ we have

$$\begin{aligned}
\pi_{f_0 \pm f_*}^{\otimes n} k &= \int k(x_1, \ldots, x_n) \prod_{i=1}^n [\pi_{f_0}(x_i) \pm \pi_{f_*}(x_i)] \prod_{i=1}^n \zeta(\mathrm{d}x_i) \\
&= \int k(x_1, \ldots, x_n) \sum_{I,J} \left[(\pm 1)^{\#J} \prod_{j \in J} \pi_{f_*}(x_j) \prod_{i \in I} \pi_{f_0}(x_i) \right] \prod_{i=1}^n \zeta(\mathrm{d}x_i) ,
\end{aligned}$$

where the sum is take over all sets I and J such that $I \cup J = \{1, \ldots, n\}$ and $I \cap J = \emptyset$. Therefore,

$$\begin{aligned}
&\pi_{f_0+f_*}^{\otimes n}\left[1_A\|f_0 + f_* - \hat{f}\|_{\mathbb{H}}^2\right] + \pi_{f_0-f_*}^{\otimes n}\left[1_A\|f_0 - f_* - \hat{f}\|_{\mathbb{H}}^2\right] \\
&= \sum_{I,J} \int \prod_{i \in I} \pi_{f_0}(x_i) \prod_{j \in J} \pi_{f_*}(x_j) 1_A \left[\|f_0 + f_* - \hat{f}\|_{\mathbb{H}}^2 + (-1)^{\#J}\|f_0 - f_* - \hat{f}\|_{\mathbb{H}}^2 \right] \prod_{i=1}^n \zeta(\mathrm{d}x_i) .
\end{aligned}$$

Since $\|f_*\|_{\mathbb{H}} \leq 1$ and, on A, $\|f_0 - \hat{f}\|_{\mathbb{H}} \leq \frac{c}{1-c}$, we obtain that, on A, $\|f_0 \pm f_* - \hat{f}\|_{\mathbb{H}} \leq \|f_0 - \hat{f}\|_{\mathbb{H}} + \|f_*\|_{\mathbb{H}} \leq \frac{1}{1-c}$. This implies that the absolute value of the sum in the last display taken over all sets I and J such that the cardinality of set J is positive, $\#J \geq 1$, is lower than

$$\begin{aligned}
&\frac{2}{(1-c)^2} \sum_{I,J:\#J \geq 1} \int \prod_{i \in I} \pi_{f_0}(x_i) \prod_{j \in J} |\pi_{f_*}(x_j)| \prod_{i=1}^n \zeta(\mathrm{d}x_i) \\
&= \frac{2}{(1-c)^2} \sum_{I,J:\#J \geq 1} \prod_{i \in I} \int \pi_{f_0}(x_i) \zeta(\mathrm{d}x_i) \prod_{j \in J} \int |\pi_{f_*}(x_j)| \zeta(\mathrm{d}x_j) \\
&= \frac{2}{(1-c)^2} \left\{ \left(1 + \int |\pi_{f_*}(x)| \zeta(\mathrm{d}x) \right)^n - 1 \right\}
\end{aligned}$$

Moreover, the term with $\#J = 0$ writes

$$\pi_{f_0}^{\otimes n}\left(1_A(\|f_0 + f_* - \hat{f}\|_{\mathbb{H}}^2 + \|f_0 - f_* - \hat{f}\|_{\mathbb{H}}^2)\right) = 2\pi_{f_0}^{\otimes n}\left(1_A(\|f_0 - \hat{f}\|_{\mathbb{H}}^2 + \|f_*\|_{\mathbb{H}}^2)\right) ,$$

by the Parallelogram law. By combining these results, the minimax risk is bounded from below by

$$(1-c)\pi_{f_0}^{\otimes n}\|f_0-\hat{f}\|_{\mathbb{H}}^2+ \\ c\pi_{f_0}^{\otimes n}\left[1_A(\|f_0-\hat{f}\|_{\mathbb{H}}^2+\|f_*\|_{\mathbb{H}}^2)\right]-\frac{c}{(1-c)^2}\left[\left(1+\int|\pi_{f_*}(x)|\zeta(\mathrm{d}x)\right)^n-1\right] .$$

Finally we see that

$$\begin{aligned}(1-c)\|f_0-\hat{f}\|_{\mathbb{H}}^2+c1_A\left(\|f_0-\hat{f}\|_{\mathbb{H}}^2+\|f_*\|_{\mathbb{H}}^2\right)& \\ =c1_A\|f_*\|_{\mathbb{H}}^2+((1-c)+c1_A)\|f_0-\hat{f}\|_{\mathbb{H}}^2& \\ \geq c1_A\|f_*\|_{\mathbb{H}}^2+c1_{A^c}& \\ \geq c\|f_*\|_{\mathbb{H}}^2 ,&\end{aligned}$$

where we used $1\geq\|f_*\|_{\mathbb{H}}^2$. This yields the lower bound asserted in the theorem. □

5.2. **Application to Polynomial Approximation Classes.** The lower bound given in (23) relies on the choice of a function f_* such that f_0 and $f_0\pm f_*$ are in the smoothness class of interest. In this subsection, we give conditions which provide a tractable choice of $\|f_*\|_{\mathbb{H}}\leq 1$ for the class $\mathcal{C}(\alpha,C)$ defined in (10). Following the same lines as Theorem 1 in [22], the key idea consists in restricting our choice using the space $V_m^\perp$ (the orthogonal set of V_m in $\mathbb{H}$) and to control separately the two terms that appear in the right hand-side of (23) within this space.

An important constraint on f_* is that $f_0\pm f_*\in\mathbb{H}_1$. In particular, for controlling the sign of $f_0\pm f_*$, we use the following semi-norm on $\mathbb{H}$,

$$\|f\|_{\infty,f_0}=\operatorname*{ess\,sup}_{t\in\Theta}\frac{|f(t)|}{f_0(t)} ,$$

with the convention $0/0=0$ and $s/0=\infty$ for $s>0$. Further, for any subspace V of $\mathbb{H}$, we denote

$$K_{\infty,f_0}(V)=\sup\{\|f\|_{\infty,f_0}:f\in V,\|f\|_{\mathbb{H}}=1\} .$$

The following lemma will serve to optimize the term $\|f_*\|_{\mathbb{H}}$ on the right-hand side of (23). It is similar to Lemma 2 in [22], so we omit its proof.

Lemma 2. *Suppose that Assumption 2 holds. Let f_0 be in $\mathbb{H}_1$, $\alpha,C_0>0$, $K\leq 1$ and let $\mathcal{C}(\alpha,C_0)$ be defined by (10) with V_m given by (4). Let moreover $w\in\mathbb{H}$. Then there exists $g\in\mathcal{C}(\alpha,C_0)\cap V_m^\perp\cap w^\perp$ such that $\|g\|_{\infty,f_0}\leq K$ and*

$$\|g\|_{\mathbb{H}}=\min\left(C_0\,(m+1)^{-\alpha},\frac{K}{K_{\infty,f_0}(V_{m+2}\cap V_m^\perp\cap w^\perp)}\right) .$$

Under Assumption 3, where the orthonormal functions ψ_k are related to polynomials in some space $\mathbb{H}'=L^2[a',b']$, the constant $K_{\infty,f_0}(V_{m+2}\cap V_m^\perp\cap w^\perp)$ can be bounded by $K_{\infty,f_0}(V_{m+2})$ and then using the following lemma.

Lemma 3. *Suppose that Assumption 3 holds. Let f_0 be in $\mathbb{H}_1$ and suppose that*

$$\sup\left\{\|f\|_{\infty,f_0}\ :\ f\in\mathbb{H}\text{ such that }\sup_{t\in[a',b']}|Tf(t)|\leq 1\right\}<\infty . \tag{24}$$

Then there exists a constant $C_0>0$ satisfying

$$K_{\infty,f_0}(V_{m+2})\leq C_0m ,\quad \text{for all } m\geq 1 .$$

Proof. Note that $\{Tf : f \in V_m\}$ is the set of polynomials in $\mathbb{H}'$ of degree at most $m-1$, denoted by $\mathcal{P}_{m-1}$. Using $\|f\|_{\mathbb{H}} = \|Tf\|_{\mathbb{H}'}$ and denoting by B the left-hand side of (24), we have

$$\begin{aligned} K_{\infty,f_0}(V_m) &= \sup\{\|f\|_{\infty,f_0} : f \in V_m, \|f\|_{\mathbb{H}} = 1\} \\ &\le B \sup\left\{ \sup_{t\in[a',b']} |Tf(t)| \,:\, f \in V_m, \|f\|_{\mathbb{H}} = 1 \right\} \\ &= B \sup\left\{ \sup_{t\in[a',b']} |p(t)| : p \in \mathcal{P}_{m-1}, \int_{a'}^{b'} p^2(t)\mathrm{d}t = 1 \right\} \end{aligned}$$

By the Nikolskii inequality (see e.g. DeVore and Lorentz [7], Theorem 4.2.6), there exists a constant $C > 0$ such that the latter sup is at most Cm. Hence, there exists $C_0 > 0$ such that $K_{\infty,f_0}(V_m) \le C_0 m$ for all $m \ge 1$. □

Theorem 3 and Lemmas 2 and 3 yield the following result.

Corollary 2. *Let $\alpha \ge 1$ and $C > (b-a)^{-1/2}$. Suppose that Assumption 3 holds with an isometry T satisfying the assumptions of Theorem 2. Let w be an $[\alpha]+1$ times continuously differentiable function defined on $[a,b]$ and set*

$$v_m = \sup_{g\in V_m^{\perp}, \|g\|_{\mathbb{H}}\le 1} \int |\pi_{w\,g}(x)|\, \zeta(\mathrm{d}x)\ . \tag{25}$$

Then there exists a small enough $C_* > 0$ and $C^* > 0$ such that, for any sequence (m_n) of integers increasing to ∞ satisfying $v_{m_n} \le C_*\, n^{-1} m_n^{-\alpha}$, we have

$$\inf_{\hat f\in\mathcal{S}_n} \sup_{f\in\tilde{\mathcal{C}}(\alpha,C)\cap\mathbb{H}_1} \pi_f^{\otimes n}\|\hat f - f\|_{\mathbb{H}}^2 \ge C^*\, m_n^{-2\alpha}(1+o(1))\ , \tag{26}$$

where $\tilde{\mathcal{C}}(\alpha,C)$ is the smoothness class defined by (20).

Remark 3. The assumption $C > (b-a)^{-1/2}$ is necessary, otherwise $\tilde{\mathcal{C}}(\alpha,C)\cap\mathbb{H}_1$ is reduced to one density for $C = (b-a)^{-1/2}$ and is empty for $C < (b-a)^{-1/2}$. To see why, observe that for any $f \in \mathbb{H}_1$, by Jensen's inequality, $\|f\|_{\mathbb{H}}^2 = \int_a^b f^2(t)\mathrm{d}t \ge (b-a)^{-1}$, with equality implying that f is the uniform density on $[a,b]$.

Proof. We apply Theorem 3 with f_0 set as the uniform density on $[a,b]$ and f_* chosen as follows. For some $C_0 > 0$ and an integer m to be determined later, we choose $f_* = wg$ where g is given by Lemma 2 with $K = \min(1, \sup_{t\in[a,b]}|w(t)|)$. Since $g \in w^{\perp}$ and $\|g\|_{\infty,f_0} \le K$, we get that $f_0 \pm f_* \in \mathbb{H}_1$.

Now we show that $\{f_0, f_0 \pm f_*\} \subset \tilde{\mathcal{C}}(\alpha,C)$ for a well chosen C_0. We have $\|f_0\|_{\mathbb{H}} = (b-a)^{-1/2}$ and, since the symmetric differences of all order vanishes on f_0, we get that $f_0 \in \tilde{\mathcal{C}}(\alpha,(b-a)^{-1/2})$. By definition of g in Lemma 2 and Lemma 5 successively, we get that $f_* \in \tilde{\mathcal{C}}(\alpha, C_1' C_0)$ for some $C_1' > 0$ not depending on C_0. Choosing $C_0 = (C-(b-a)^{-1/2})/C_1'$, we finally get that

$$\{f_0, f_0 \pm f_*\} \subset \tilde{\mathcal{C}}(\alpha,C)\cap\mathbb{H}_1\ .$$

By Lemma 2, $\|g\|_{\mathbb{H}} \to 0$ as $m \to \infty$ and, since w is bounded, it implies that $\|f_*\|_{\mathbb{H}} \le 1$ for m large enough. Hence we may apply Theorem 3 and, to conclude the proof, it remains to provide a lower bound of the right-hand side of (23) for the above choice of f_*. Under the assumptions of Theorem 2, Condition (24) clearly holds. So Lemma 3 and the definition of g in Lemma 2 give that

$$\|g\|_H \le C_0' m^{-\alpha}\ ,$$

for some constant $C_0' > 0$. By definition of v_m and since $g \in V_m^\perp$, we have

$$\int |\pi_{f_*}(x)|\, \zeta(\mathrm{d}x) \le \|g\|_H\, v_m \le C_0' m^{-\alpha} v_m \ .$$

We now apply the lower bound given by (23) with $m = m_n$ for (m_n) satisfying $v_{m_n} \le C_* n^{-1} m_n^{-\alpha}$. We thus obtain

$$\begin{aligned}
&\inf_{\hat f \in \mathcal{S}_n} \sup_{f \in \tilde{\mathcal{C}}(\alpha,C) \cap \mathbb{H}_1} \pi_f^{\otimes n} \|\hat f - f\|_{\mathbb{H}}^2 \\
&\qquad \ge c(C_0' m_n^{-\alpha})^2 - \frac{c}{(1-c)^2} C_* C_0' m_n^{-2\alpha}(1+o(1)) \\
&\qquad \ge C^* m_n^{-2\alpha}\,(1+o(1)) \ ,
\end{aligned}$$

where the last inequality holds for some $C^* > 0$ provided that C_* is small enough. □

To apply Corollary 2, one needs to investigate the asymptotic behavior of the sequence (v_m) defined in (25). The following lemma can be used to achieve this goal.

Lemma 4. *Under Assumption 3, if $\pi_\cdot(x) \in \mathbb{H}$ for all $x \in \mathrm{X}$, then v_m defined in (25) satisfies*

$$v_m \le \int \|T[w\pi_\cdot(x)] - P_{\mathcal{P}_{m-1}}(T[w\pi_\cdot(x)])\|_{\mathbb{H}'}\, \zeta(\mathrm{d}x) \ , \tag{27}$$

where $\mathcal{P}_{m-1}$ is the set of polynomials of degree at most $m-1$ in $\mathbb{H}'$ and $P_{\mathcal{P}_{m-1}}$ denotes the orthogonal projection in $\mathbb{H}'$ onto $\mathcal{P}_{m-1}$.

Proof. Let $g \in V_m^\perp$ such that $\|g\|_{\mathbb{H}} \le 1$. Then we have, for all $x \in \mathbb{R}$,

$$\pi_{wg}(x) = \langle wg, \pi_\cdot(x)\rangle_{\mathbb{H}} = \langle g, w\pi_\cdot(x)\rangle_{\mathbb{H}} = \langle Tg, T[w\pi_\cdot(x)]\rangle_{\mathbb{H}'} \ .$$

Recall that $TV_m = \mathcal{P}_{m-1}$ is the set of polynomials of degree at most $m-1$ in $\mathbb{H}'$. Hence, Tg is orthogonal to $\mathcal{P}_{m-1}$, and for any $p \in \mathcal{P}_{m-1}$, we get, for all $x \in \mathbb{R}$,

$$|\pi_{wg}(x)| = |\langle Tw, T[w\pi_\cdot(x)] - p\rangle_{\mathbb{H}'}| \le \|T[w\pi_\cdot(x)] - p\|_{\mathbb{H}'} \ , \tag{28}$$

where we used the Cauchy-Schwarz inequality and $\|Tg\|_{\mathbb{H}'} = \|g\|_{\mathbb{H}} \le 1$. Now the bound given by (27) is obtained by taking p equal to the projection of $[w\pi_\cdot(x)]$ onto $\mathcal{P}_{m-1}$ (observe that the right-hand side of (28) is then minimal). □

5.3. **Minimax Rate for Exponential Mixtures.** In this section, we show that in the case of exponential mixtures the orthogonal series estimator of Example 1(a) achieves the minimax rate.

Theorem 4. *Consider the exponential case, that is, let Assumption 1 hold with ζ defined as the Lebesgue measure on $\mathbb{R}_+$, $\Theta = [a,b] \subset (0,\infty)$ and $\pi_t(x) = t\mathrm{e}^{-tx}$. Let $C > (b-a)^{-1/2}$ and $\alpha > 1$ and define $\tilde{\mathcal{C}}(\alpha, C)$ as in (20). Then there exists $C^* > 0$ such that*

$$\inf_{\hat f \in \mathcal{S}_n} \sup_{f \in \tilde{\mathcal{C}}(\alpha,C) \cap \mathbb{H}_1} \pi_f^{\otimes n} \|\hat f - f\|_{\mathbb{H}}^2 \ge C^* (\log n)^{-2\alpha}(1+o(1)) \ . \tag{29}$$

Proof. Let $g_k(x) = 1\left\{x > k - \frac{1}{2}\right\}$, for $k \ge 1$. Then Assumption 3 holds with φ_k and T defined by (6) and (7), respectively. Since $a > 0$, T satisfies the assumptions of Theorem 2. We may thus apply Corollary 2 with $w = 1_{[a,b]}$. Hence the minimax lower bound given in (29) thus follows from (26), provided that we have for some constant $C' > 0$, setting $m_n = C' \log n$,

$$v_{m_n} = o(n^{-1} m_n^{-\alpha}) \quad \text{as } n \to \infty \ , \tag{30}$$

where v_m is defined by (25). Note that $\pi_t(x) = t\mathrm{e}^{-xt} 1_{\mathbb{R}_+}(x)$. We apply Lemma 4 to bound v_m. Using the definition of T in (7), we have for all $x \ge 0$, $[T\pi_\cdot(x)](t) = -\log t\; t^{x-1/2}$. We write $x \in \mathbb{R}_+$ as the

sum of its entire and decimal parts, $x = [x] + \langle x \rangle$, and observe that, since $< x > -1/2 \in [-1/2, 1/2)$ and $[a', b'] = [\mathrm{e}^{-b}, \mathrm{e}^{-a}] \subset (0, 1)$, the expansion of $t^{<x>-1/2} = \sum_{k \geq 0} \alpha_k(x)(1-t)^k$ as a power series about $t = 1$ satisfies $|\alpha_k(x)| = \prod_{j=1}^k |\langle x \rangle - 1/2) - j|/k! \leq 1$. Extending $-\log t$ about $t = 1$, we thus get $-\log(t) t^{<x>-1/2} = \sum_{k \geq 0} \beta_k(x)(1-t)^k$ with $|\beta_k(x)| = |\sum_{l=1}^k \alpha_{k-l}/l| \leq 1 + \log(k)$. For any $x < m$, we use this expansion to approximate $[T\pi_\cdot(x)](t) = -\log(t) t^{<x>-1/2} \times t^{[x]}$ by a polynomial of degree m. Namely, we obtain

$$\sup_{t \in [a', b']} |[T\pi_\cdot(x)](t) - \sum_{k=0}^{m-[x]} \beta_k(x)\, t^{k+[x]}| \leq \sum_{k > m-[x]} (1 + \log(k))(b')^{k+[x]} \leq C_1 c^m \ ,$$

where we used the bound $1 + \log(k) \leq C_1 (c/b')^k$, valid for some constants $C_1 > 0$ and $c \in (b', 1)$ not depending on x. This bound also applies to $\|T\pi_\cdot(x) - P_{\mathcal{P}_{m-1}}(T\pi_\cdot(x))\|_{\mathbb{H}'}$ by definition of the projection $P_{\mathcal{P}_{m-1}}$. For $x \geq m$, we simply observe that $|[T\pi_\cdot(x)](t)| \leq -\log(a') b'^{x-1/2}$. This also provides an upper bound for $\|T\pi_\cdot(x) - P_{\mathcal{P}_{m-1}}(T\pi_\cdot(x))\|_{\mathbb{H}'}$. Finally, integrating on $x \geq 0$ we get

$$\int_{\mathbb{R}_+} \|T\pi_\cdot(x) - P_{\mathcal{P}_{m-1}}(T\pi_\cdot(x))\|_{\mathbb{H}'} \, \mathrm{d}x \leq C_2\, m\, c^m \ ,$$

with constants $C_2 > 0$ and $c < 1$ not depending on m, and this upper bound applies to v_m by Lemma 4. This shows that (30) holds provided that $C' > 0$ is taken small enough. This completes the proof. □

5.4. **Minimax Rate for Gamma Shape Mixtures.** In this section, we show that in the case of Gamma shape mixtures the orthogonal series estimator of Example 4 achieves the minimax rate up to the $\log \log n$ multiplicative term.

Theorem 5. *Consider the Gamma shape mixture case, that is, let Assumption 1 hold with ζ defined as the Lebesgue measure on $\mathbb{R}_+$, $\Theta = [a, b] \subset (0, \infty)$ and $\pi_t(x) = x^{t-1}\mathrm{e}^{-x}/\Gamma(t)$. Let $C > (b-a)^{-1/2}$ and $\alpha > 1$ and define $\tilde{\mathcal{C}}(\alpha, C)$ as in (20). Then there exists $C^* > 0$ such that*

$$\inf_{\hat{f} \in \mathcal{S}_n} \sup_{f \in \tilde{\mathcal{C}}(\alpha, C) \cap \mathbb{H}_1} \pi_f^{\otimes n} \|\hat{f} - f\|_{\mathbb{H}}^2 \geq C^* (\log n)^{-2\alpha} (1 + o(1)) \ . \tag{31}$$

Proof. We proceed as in the proof of Theorem 4. This time we set $g_k(x) = \sum_{l=1}^k \tilde{c}_{k,l} t^{l-1}$ with coefficients $(\tilde{c}_{k,l})$ defined in Lemma 6. Assumption 3 then holds with $\mathbb{H}' = \mathbb{H}$ and T defined as the identity operator. Applying Corollary 2 with $w(t) = \Gamma(t)$, we obtain the lower bound given in (31) provided that Condition (30) holds with $m_n = C' \log n / \log \log n$ for some $C' > 0$. Again we use Lemma 4 to check this condition in the present case. To this end we must, for each $x > 0$, provide a polynomial approximation of $w(t)\pi_t(x) = x^{t-1}\mathrm{e}^{-x}$ as a function of t. Expanding the exponential function as a power series, we get

$$\sup_{t \in [a,b]} \left| w(t)\pi_t(x) - \mathrm{e}^{-x} \sum_{k=0}^{m-1} \frac{\log^k(x)}{k!} (t-1)^k \right| \leq \mathrm{e}^{-x} \sum_{k \geq m} \frac{|\log(x)|^k}{k!} c^k \ ,$$

where $c = \max(|a-1|, |b-1|)$. Let (x_m) be a sequence of real numbers tending to infinity. The right hand side of the previous display is less than $\mathrm{e}^{c|\log(x)| - x} (c|\log(x)|)^m / m!$. We use this for bounding $\|w\pi_\cdot(x) - P_{\mathcal{P}_{m-1}}(w\pi_\cdot(x))\|_{\mathbb{H}}$ (recall that T is the identity and $\mathbb{H}' = \mathbb{H}$) when $x \in [\mathrm{e}^{-x_m}, x_m]$. When $x \in (0, \mathrm{e}^{-x_m})$ we use that the latter is bounded by $O(1)$ and when $x > x_m$ by $O(\mathrm{e}^{-x/2})$. Hence Lemma 4 gives that

$$v_m = O(\mathrm{e}^{-x_m}) + \frac{c^m}{m!} \int_{\mathrm{e}^{-x_m}}^{x_m} \mathrm{e}^{c|\log(x)| - x} \, |\log(x)|^m \, \mathrm{d}x + O\left(\mathrm{e}^{-x_m/2}\right) \ .$$

Now observe that, as $x_m \to \infty$, separating the integral $\int_{\mathrm{e}^{-x_m}}^{x_m}$ as $\int_{\mathrm{e}^{-x_m}}^{1} + \int_{1}^{x_m}$, we get

$$\int_{\mathrm{e}^{-x_m}}^{x_m} \mathrm{e}^{c|\log(x)|-x} \, |\log(x)|^m \, \mathrm{d}x = O(\mathrm{e}^{cx_m} x_m^m) + O(\log^m(x_m)) \; .$$

Set $x_m = c_0 m$. By Stirling's formula, for $c_0 > 0$ small enough, we get $v_m = O(c_1^m)$ with $c_1 \in (0,1)$. We conclude as in the proof of Theorem 4. □

5.5. **Lower Bound for Compactly Supported Scale Families.** We derived in Corollary 1 an upper bound of the minimax rate for estimating f in $\mathcal{C}(\alpha, C)$. It is thus legitimate to investigate whether, as in the exponential mixture case, this upper bound is sharp for mixtures of compactly supported scale families. A direct application of Corollary 2 provides the following lower bound, which, unfortunately, is far from providing a complete and definite answer.

Theorem 6. *Consider the case of scale mixtures of a compactly supported density on $\mathbb{R}_+$, that is, suppose that the assumptions of Lemma 1 hold. Suppose moreover that π_1 has a k-th derivative bounded on $\mathbb{R}_+$. Let $C > (b-a)^{-1/2}$ and $\alpha \geq 1$, and define $\tilde{\mathcal{C}}(\alpha, C)$ as in (20). Then if $k > \alpha$,*

$$\inf_{\hat{f} \in \mathcal{S}_n} \sup_{f \in \tilde{\mathcal{C}}(\alpha,C) \cap \mathbb{H}_1} \pi_f^{\otimes n} \|\hat{f} - f\|_{\mathbb{H}}^2 \geq n^{-2\alpha/(k-\alpha)}(1 + o(1)) \; . \tag{32}$$

Proof. We proceed as in the proof of Theorem 4, that is, we observe that Assumption 3 holds with the same choice of (g_k) as in Lemma 1 and apply Corollary 2 with $w = 1_{[a,b]}$. Here, the lower bound given in (32) is obtained by showing that

$$v_{m_n} = O(n^{-1} m_n^{-\alpha}) \quad \text{as } n \to \infty \; , \tag{33}$$

holds with $m_n = n^{1/(k-\alpha)}$ and with (v_m) defined by (25). Again we use Lemma 4 to bound v_m. Here T is the identity operator on $\mathbb{H} = \mathbb{H}'$ and $\pi_t(x) = t^{-1}\pi_1(x/t)$. Let $M > 0$ such that the support of π_1 is included in $[0, M]$. Then for $t \in [a, b]$ and $x > Mb$, $\pi_t(x) = 0$. Hence

$$\|\pi_\cdot(x) - P_{\mathcal{P}_{m-1}}(\pi_\cdot(x))\|_{\mathbb{H}} = 0 \quad \text{for all } x > Mb \; . \tag{34}$$

We now consider the case $x \leq Mb$. By the assumption on π_1 and a, we have that $t \mapsto \pi_t(x) = t^{-1}\pi_1(x/t)$ is k-times differentiable on $[a, b]$. Moreover its k-th derivative is bounded by $C_k x^k$ on $[a, b]$, where $C_k > 0$ does not depend on x. It follows that, for any $h > 0$ and $t \in [a, b]$,

$$\left| \Delta_h^k(\pi_\cdot(x), t) \right| \leq C_k \, k! \, (xh)^k \; ,$$

where Δ_h^k is the k-th order symmetric difference operator defined by (17). Observing moreover that

$$\|\pi_\cdot(x)\|_{\mathbb{H}}^2 = \int_a^b t^{-2} \pi_1^2(x/t) \, \mathrm{d}t \leq C' \, ,$$

for some $C' > 0$ not depending on x, we get that $\pi_\cdot(x) \in \tilde{\mathcal{C}}(k, C' \vee C_k \, k! \, x^k)$. Using Corollary 7.25 in Ditzian and Totik [8], we thus have for a constant $C'' > 0$ not depending on x,

$$\|\pi_\cdot(x) - P_{\mathcal{P}_{m-1}}(\pi_\cdot(x))\|_{\mathbb{H}} \leq C''(1 + x^k) m^{-k} \quad \text{for all } x \leq Mb \; . \tag{35}$$

Applying Lemma 4 with (34) and (35), we obtain $v_m = O(m^{-k})$. We conclude that (33) holds with $m_n = n^{1/(k-\alpha)}$, which completes the proof. □

Theorem 6 provides polynomial lower bounds of the minimax MISE rate whereas Corollary 1 gives logarithmic upper bounds in the same smoothness spaces. Hence the question of the minimax rate is left completely open in this case. Moreover the lower bound relies on smoothness conditions on π_1 which rule out Example 3 (for which π_1 is discontinuous). On the other hand, the case of scale families can be

related with the deconvolution problem that has received a considerable attention in a series of papers of the 1990's (see e.g. [26, 9, 10, 11, 21]). The following section sheds a light on this relationship.

5.6. **Scale Families and Deconvolution.** The following lower bound is obtained from classical lower bounds in the deconvolution problem, derived in [11].

Theorem 7. *Consider the case of scale mixtures on $\mathbb{R}_+$, that is, suppose that Assumption 1 with ζ equal to the Lebesgue measure on $\mathbb{R}_+$, $\Theta = [a, b] \subset (0, \infty)$ and $\pi_t(x) = t^{-1}\pi_1(x/t)$. Denote by ϕ the characteristic function of the density $\mathrm{e}^t\pi_1(\mathrm{e}^t)$ on $\mathbb{R}$,*

$$\phi(\xi) = \int \mathrm{e}^{t+\mathrm{i}\xi t}\pi_1(\mathrm{e}^t)\,\mathrm{d}t\ .$$

Define $\tilde{T}$ as the operator $\tilde{T}(g) = \tilde{g}$, where $g : \mathbb{R} \to \mathbb{R}$ and $\tilde{g}(t) = t^{-1}g(\log(t))$ for all $t \in (0, \infty)$. Let $C > 0$ and $\alpha > 0$, and define $\mathcal{L}(\alpha, C)$ as the set containing all densities g on $\mathbb{R}$ such that

$$\left|g^{(r)}(t) - g^{(r)}(u)\right| \le C|t - u|^{\alpha - r} \text{ for all } t, u \in \mathbb{R}\ ,$$

where $r = [\alpha]$.

(a) Assume that $\phi^{(j)}(t) = O(|t|^{-\beta-j})$ as $|t| \to \infty$ for $j = 0, 1, 2$, where $\phi^{(j)}$ is the j-th derivative of ϕ. Then there exists $C^* > 0$ such that

$$\inf_{\hat{f}\in\mathcal{S}_n} \sup_{f\in\tilde{T}(\mathcal{L}(\alpha,C))} \pi_f^{\otimes n}\|\hat{f} - f\|_{\mathbb{H}}^2 \ge C^*\ n^{-2\alpha/(2(\alpha+\beta)+1)}(1 + o(1))\ . \tag{36}$$

(b) Assume that $\phi(t) = O(|t|^{\beta_1}\mathrm{e}^{-|t|^\beta/\gamma})$ as $|t| \to \infty$ for some $\beta, \gamma > 0$ and β_1, and that $\pi_1(u) = o(u^{-1}|\log(u)|^{-a})$ as $u \to 0, \infty$ for some $a > 1$. Then there exists $C^* > 0$ such that

$$\inf_{\hat{f}\in\mathcal{S}_n} \sup_{f\in\tilde{T}(\mathcal{L}(\alpha,C))} \pi_f^{\otimes n}\|\hat{f} - f\|_{\mathbb{H}}^2 \ge C^*\ \log(n)^{-2\alpha/\beta}(1 + o(1))\ . \tag{37}$$

Proof. In the scale mixture case the observation X can be represented as $X = \theta Y$, where Y and θ are independent variables having density π_1 and (unknown) density f, respectively. By taking the log of the observations, the problem of estimating the density of $\log(\theta)$, that is $f^*(t) = \mathrm{e}^t f(\mathrm{e}^t)$, is a deconvolution problem. Hence we may apply Theorem 2 in [11] to obtain lower bounds on the nonparametric estimation of f^* from $\log(X_1), \ldots, \log(X_n)$ under appropriate assumptions on ϕ, which is the characteristic function of $\log(Y)$. Let $a' = \log(a)$ and $b' = \log(b)$. The lower bounds in (a) and (b) above are those appearing in (a) and (b) in Theorem 2 of [11] of the minimax quadratic risk in $\mathbb{H}' = L^2([a', b'])$ for estimating f^* in the Lipschitz smoothness class $\mathcal{L}(\alpha, C)$. Observe that $\tilde{T}$ is defined for all function $g : [a', b'] \to \mathbb{R}$ by $\tilde{T}(g) = \tilde{g}$ with $\tilde{g}$ defined on $[a, b]$ by $\tilde{g}(t) = t^{-1}g(\log(t))$, so that $\tilde{T}(f^*) = f$. Observing that $\tilde{T}$ is a linear operator and that for any $g \in \mathbb{H}'$, $\|\tilde{T}(g)\|_{\mathbb{H}} \asymp \|g\|_{\mathbb{H}'}$, we obtain the lower bounds given in (36) and (37). □

As in Theorem 6, the smoother π_1 is assumed, the slower the lower bound of the minimax rate. However the lower bounds obtained in Theorem 7 hold for a much larger class of scale families. Indeed, if π_1 is compactly supported, the condition induced on π_1 in case (a) are much weaker than in Theorem 6. For instance, it holds with $\beta = k$ for Example 3. For an infinitely differentiable π_1 both theorems say that the minimax rate is slower than any polynomial rate. However, in this case, case (b) in Theorem 7 may provide a more precise logarithmic lower bound. It is interesting to note that, as a consequence of [5], the MISE rate $(\log n)^{-2\alpha}$, which is the rate obtained in Corollary 1 by the polynomial estimator for any compactly supported π_1, is the slowest possible minimax rate obtained in Theorem 7(b) for a compactly supported π_1. Such a comparison should be regarded with care since the smoothness class

in the latter theorem is different and cannot be compared to the smoothness classes considered in the previous results, as we explain hereafter.

The arguments for adapting the lower bounds of Theorem 7 also apply for minimax upper bounds. More precisely, using the kernel estimators for the deconvolution problem from the observations $\log(X_1)$, ..., $\log(X_n)$ and mapping the estimator through $\tilde{T}$, one obtains an estimator of f achieving the same integrated quadratic risk. The obtained rates depend on similar assumptions on ϕ as those in (a) and (b), see [9, 10, 11]. Although the scale mixture and the convolution model are related to one another by taking the exponential (or the logarithm in the reverse sense) of the observations, it is important to note that, except for Theorem 7, our results are of different nature. Indeed, the upper and lower bounds in the deconvolution problem cannot be compared with those obtained previously in the paper because there are no possible inclusions between the smoothness classes considered in the deconvolution problem and those defined by polynomial approximations.

Let us examine more closely the smoothness class $\tilde{T}(\mathcal{L}(\alpha, C))$ that appears in the lower bounds of Theorem 7, inherited from the results on the deconvolution problem. This class contains densities with non-compact supports whereas $\tilde{\mathcal{C}}(\alpha, C) \cap \mathbb{H}_1$ only contains densities with supports in $[a, b]$. Hence neither (36) nor (37) can be used for deriving minimax rates in $\tilde{\mathcal{C}}(\alpha, C) \cap \mathbb{H}_1$. In fact the densities exhibited in [11] to prove the lower bound have infinite support by construction and the argument does not at all seem to be adaptable for a class of compactly supported densities. As for upper bounds in the deconvolution problem, they are based on Lipschitz or Sobolev type of smoothness conditions which are not compatible with compactly supported densities on $[a, b]$ except for those that are smoothly decreasing close to the end points. This follows from the fact that, in the deconvolution problem, standard estimators (kernel or wavelet) highly rely on the Fourier behavior both of the mixing density and of the additive noise density. In contrast, such boundary constraints are not necessary for densities in $\tilde{\mathcal{C}}(\alpha, C)$. For instance the uniform density on $[a, b]$ belongs to $\tilde{\mathcal{C}}(\alpha, C)$ for all $\alpha > 0$ and $C > (b - a)^{-1/2}$, but has a Fourier transform decreasing very slowly. A natural conclusion of this observation is that polynomial estimators should be used preferably to standard deconvolution estimators when the mixing density has a known compact support $[a, b] \subset (0, \infty)$. Of course this conclusion holds for both deconvolution and scale mixture problems.

6. Support Estimation

A basic assumption of our estimation approach is that the mixing density f belongs to $\mathbb{H} = L^2[a, b]$. However, in practice the exact interval $[a, b]$ is generally unknown. To compass this problem, we propose an estimator of the support of the mixing density f, or more precisely of the support of Tf. It can be shown that the support estimator is consistent when it is based on an estimator $T\hat{f}_{n,m_n}$, which is a polynomial, and Tf behaves as follows on the bounds of the support interval.

Denote by $[a_0, b_0]$ the smallest interval such that $Tf(u) = 0$ for all $u \in [a', b']\backslash[a_0, b_0]$. In other words, $a_0 = \inf\{u \in [a', b'], Tf(u) > 0\}$ and $b_0 = \sup\{u \in [a', b'], Tf(u) > 0\}$. Furthermore, we suppose that there exist constants $D > a_0, E < b_0, D', E', \alpha' > 0$ such that

$$Tf(u) \geq ((u - a_0)/D')^{\alpha'} \text{ , for all } u \in [a_0, D] \text{ ,} \tag{38}$$

$$Tf(u) \geq ((b_0 - u)/E')^{\alpha'} \text{ , for all } u \in [E, b_0] \text{ .} \tag{39}$$

For fixed $\varepsilon_n, \eta_n > 0$, we define the estimators $\hat{a}_n$ and $\hat{b}_n$ of the interval bounds a_0 and b_0 by

$$\hat{a}_n = \inf\left\{u \in [a', b'] : T\hat{f}_{n,m_n}(v) > \frac{\varepsilon_n}{2} \text{ for all } v \in [u, u+\eta_n]\right\} \tag{40}$$

$$\hat{b}_n = \sup\left\{u \in [a', b'] : T\hat{f}_{n,m_n}(v) > \frac{\varepsilon_n}{2} \text{ for all } v \in [u-\eta_n, u]\right\} \ . \tag{41}$$

Roughly, these estimators take the smallest and largest value where the estimator $T\hat{f}_{n,m_n}$ exceeds $\varepsilon_n/2$, by disregarding side-effects of size η_n. For a convenient choice of the sequences $(\varepsilon_n)_n$ and $(\eta_n)_n$ these estimators are consistent.

Proposition 2. *Let $\hat{f}_{n,m_n}$ be the density estimator defined in (3) under Assumption 3 with $\alpha > 1/2$. Suppose that f verifies (38-39) for appropriate constants $D > a_0, E < b_0, D', E', \alpha' > 0$. Assume that there are sequences $m_n \to \infty$, $\varepsilon_n \to 0$ and $\eta_n \to 0$ such that*

$$\mathbb{E}\left\|\hat{f}_{n,m_n} - f\right\|_{\mathbb{H}}^2 = O\left(m_n^{-2\alpha}\right) \ , \qquad \varepsilon_n^{-1} = o\left(m_n^{(2\alpha-1)/(2+1/\alpha')}\right) \ , \qquad \eta_n = O\left(\varepsilon_n^{1/\alpha'} m_n^{-1}\right) \ .$$

Then the estimators $\hat{a}_n$ and $\hat{b}_n$ defined by (40) and (41) are consistent for the support bounds a_0 and b_0. More precisely, as $n \to \infty$,

$$(\hat{a}_n - a_0)_+ = O_P\left(\varepsilon_n^{1/\alpha'}\right) \quad \textit{and} \quad (\hat{a}_n - a_0)_- = O_P\left(\varepsilon_n^{1/\alpha'} m_n^{-1}\right) \ ,$$
$$(\hat{b}_n - b_0)_+ = O_P\left(\varepsilon_n^{1/\alpha'} m_n^{-1}\right) \quad \textit{and} \quad (\hat{b}_n - b_0)_- = O_P\left(\varepsilon_n^{1/\alpha'}\right) \ .$$

Proof. First we consider $(\hat{a}_n - a_0)_+$. We set $\delta_n = MD'\varepsilon_n^{1/\alpha'}$ for some $M > 1$ and denote

$$\begin{aligned} A_n &= \{(\hat{a}_n - a_0)_+ > \delta_n\} = \{\hat{a}_n > a_0 + \delta_n\} \\ &= \left\{\forall u \in [a', a_0 + \delta_n] \ \exists v \in [u, u+\eta_n] \text{ such that } T\hat{f}_{n,m_n}(v) \le \frac{\varepsilon_n}{2}\right\} \ . \end{aligned}$$

As $T\hat{f}_{n,m_n}$ is a polynomial of degree m_n, $T\hat{f}_{n,m_n}$ has at most m_n intersections with any constant function. Hence, the number of subintervals of $[a', b']$ where $T\hat{f}_{n,m_n}$ exceeds $\varepsilon/2$ for any fixed $\varepsilon > 0$ is bounded by m_n. On A_n, all such intervals included in $[a', a_0 + \delta_n]$ are at most of size η_n. Thus, on A_n,

$$\int_{a'}^{a_0+\delta_n} 1\left\{T\hat{f}_{n,m_n}(u) > \frac{\varepsilon_n}{2}\right\} \mathrm{d}u \le m_n\eta_n \ .$$

It follows, that on A_n,

$$\int_{a'}^{a_0+\delta_n} 1\left\{T\hat{f}_{n,m_n}(u) \le \frac{\varepsilon_n}{2}\right\} \mathrm{d}u \ge a_0 + \delta_n - a' - m_n\eta_n \ ,$$

and thus

$$\int_{a_0+D'\varepsilon_n^{1/\alpha'}}^{a_0+\delta_n} 1\left\{T\hat{f}_{n,m_n}(u) \le \frac{\varepsilon_n}{2}\right\} \mathrm{d}u \ge \delta_n - m_n\eta_n - D'\varepsilon_n^{1/\alpha'} \ .$$

For large n such that $\delta_n^{1/\alpha'} < D$ and since $Tf > \varepsilon_n$ on $[a_0 + D'\varepsilon_n^{1/\alpha'}, D]$ by (38), we obtain on A_n,

$$\begin{aligned} &\int_{a_0+D'\varepsilon_n^{1/\alpha'}}^{a_0+\delta_n} 1\left\{T\hat{f}_{n,m_n}(u) \le \frac{\varepsilon_n}{2}\right\} \mathrm{d}u \le \int_{a_0+D'\varepsilon_n^{1/\alpha'}}^{a_0+\delta_n} 1\left\{|T\hat{f}_{n,m_n}(u) - Tf(u)| > \frac{\varepsilon_n}{2}\right\} \mathrm{d}u \\ &\quad \le \frac{4}{\varepsilon_n^2} \int_{a_0+D'\varepsilon_n^{1/\alpha'}}^{a_0+\delta_n} |T\hat{f}_{n,m_n}(u) - Tf(u)|^2 \mathrm{d}u \le \frac{4}{\varepsilon_n^2}\|T\hat{f}_{n,m_n} - Tf\|_{\mathbb{H}'}^2 = \frac{4}{\varepsilon_n^2}\|\hat{f}_{n,m_n} - f\|_{\mathbb{H}}^2 \ . \end{aligned}$$

For sufficiently large M we have $m_n\eta_n < \delta_n - D'\varepsilon_n^{1/\alpha'}$. Then, it follows by Markov's inequality that

$$\begin{aligned}\mathbb{P}((\hat{a}_n - a_0)_+ > \delta_n) &\le \mathbb{P}\left(\frac{4}{\varepsilon_n^2}\|\hat{f}_{n,m_n} - f\|_{\mathbb{H}}^2 \ge \delta_n - m_n\eta_n - D'\varepsilon_n^{1/\alpha'}\right)\\ &\le \frac{4\mathbb{E}[\|\hat{f}_{n,m_n} - f\|_{\mathbb{H}}^2]}{\varepsilon_n^2(\delta_n - m_n\eta_n - D'\varepsilon_n^{1/\alpha'})} \longrightarrow 0\ , \quad n\to\infty\ ,\end{aligned}$$

by the assumptions on $(\varepsilon_n)_n$ and $\mathbb{E}\left\|\hat{f}_{n,m_n} - f\right\|_{\mathbb{H}}^2$ and as $\delta_n = MD'\varepsilon_n^{1/\alpha'}$. Thus $(\hat{a}_n - a_0)_+ = O_P(\delta_n) = O_P\left(\varepsilon_n^{1/\alpha'}\right)$.

To investigate $(\hat{a}_n - a_0)_-$ put $\delta_n = M'\eta_n$ for some $M' > 1$. By using that $Tf = 0$ on $[a, a_0]$, we have

$$\begin{aligned}\mathbb{P}((\hat{a}_n - a_0)_- > \delta_n) &= \mathbb{P}(\hat{a}_n < a_0 - \delta_n)\\ &= \mathbb{P}\left(\exists x \in [a', a_0 - \delta_n[\colon \int_x^{x+\eta_n} 1\left\{T\hat{f}_{n,m_n}(u) > \frac{\varepsilon_n}{2}\right\}\mathrm{d}u = \eta_n\right)\\ &\le \mathbb{P}\left(\int_{a'}^{a_0} 1\left\{T\hat{f}_{n,m_n}(u) > \frac{\varepsilon_n}{2}\right\}\mathrm{d}u \ge \eta_n\right)\\ &= \mathbb{P}\left(\int_{a'}^{a_0} 1\left\{|T\hat{f}_{n,m_n}(u) - Tf(u)|^2 > \frac{\varepsilon_n^2}{4}\right\}\mathrm{d}u > \eta_n\right)\\ &\le \mathbb{P}\left(\frac{4}{\varepsilon_n^2}\int_a^{a_0} |T\hat{f}_{n,m_n}(u) - Tf(u)|^2\mathrm{d}u > \eta_n\right)\\ &\le \frac{4\mathbb{E}[\|\hat{f}_{n,m_n} - f\|_{\mathbb{H}}^2]}{\eta_n\varepsilon_n^2}\\ &\longrightarrow 0\ ,\end{aligned}$$

where again we applied Markov's inequality. Consequently, $(\hat{a}_n - a_0)_- = O_P(\eta_n) = O_P\left(\varepsilon_n^{1/\alpha'} m_n^{-1}\right)$.

By symmetry, the properties on $\hat{b}_n$ stated in the proposition hold as well. □

By Theorem 1 the proposition applies to Example 1 (a) and 3 with $m_n = A\log n$ and to Example 1 (b) and 2 with $m_n = A\log n/\log\log n$.

7. Numerical results

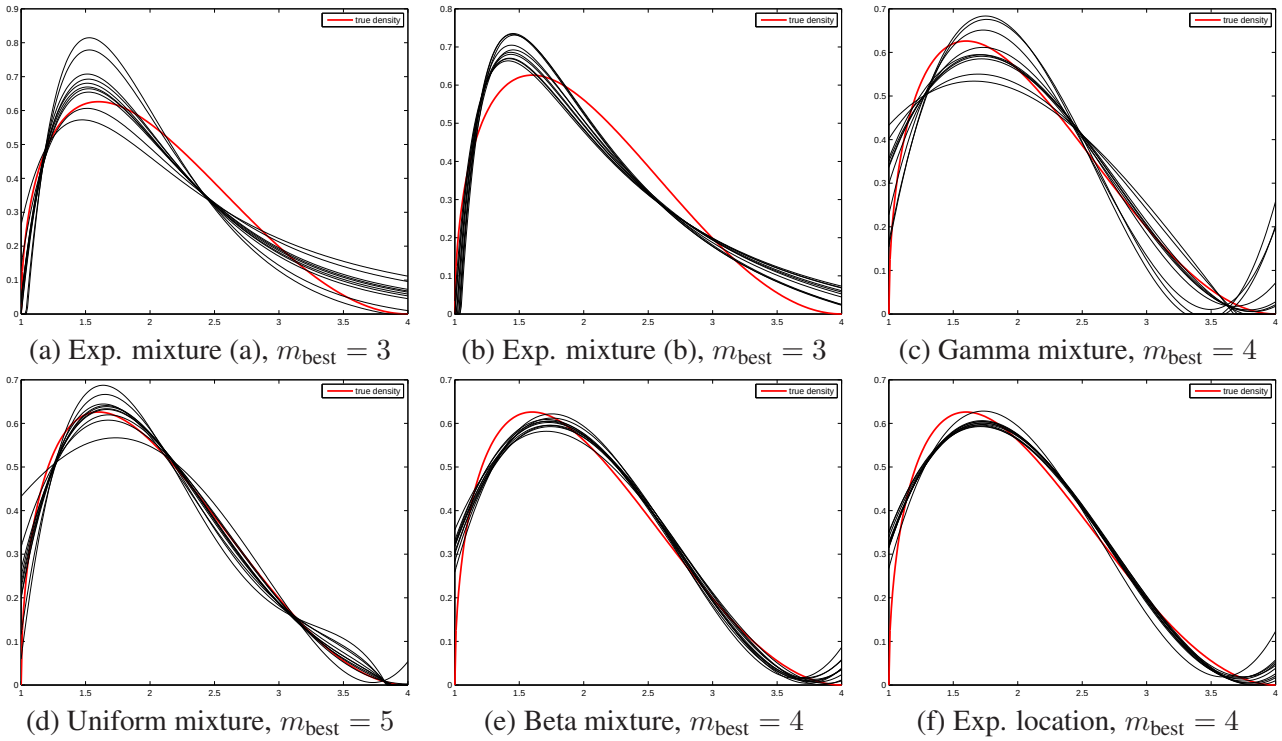
A simulation study is conducted to evaluate the performance of the estimator on finite datasets. Six different mixture settings are considered, namely the exponential mixture from Example 1 (a) and 1 (b), the Gamma shape mixture from Example 2, the uniform mixture and the Beta mixture with $k = 4$ from Example 3 and the exponential mixture with a location parameter from Example 4.

We consider the case where the mixing density f is the Beta distribution on the interval $[1, 4]$ with parameters $\alpha = 3/2$ and $\beta = 3$. Remark that for the exponential mixture setting of Example 1 (b) we cannot take a mixing distribution with support $[a, b]$ with $a = 0$, since $b' = 1/a$ must be finite.

For every mixture setting, the estimator $\hat{f}_{m,n}$ with $m = 5$ is computed on a large number of datasets (for sample sizes n varying from 100 to 10^9) and the corresponding MISE is evaluated. Table 1 gives the mean values of the different MISE and the associated standard deviations. Obviously, in all six settings the MISE decreases when n increases. Note that in the last four settings, where the mixing density f

TABLE 1. Estimated MISE (and standard deviation) of estimator $\hat{f}_{m,n}$ with $m = 5$ in six different mixture settings when the mixing density f is a Beta distribution.

		n							
		10^2	10^3	10^4	10^5	10^6	10^7	10^8	10^9
Exp. (a)	MISE	0.72	0.69	0.62	0.48	0.26	0.058	8.4e-03	1.1e-03
	sd	(0.16)	(0.17)	(0.16)	(0.19)	(0.18)	(0.085)	(0.10)	(1.4e-03)
Exp. (b)	MISE	0.61	0.52	0.35	0.21	0.084	0.015	2.0e-03	4.4e-04
	sd	(0.21)	(0.25)	(0.25)	(0.21)	(0.12)	(0.027)	(2.9e-03)	(2.7e-04)
Gamma	MISE	0.58	0.47	0.31	0.12	0.020	3.4e-03	1.5e-03	1.3e-03
	sd	(0.20)	(0.22)	(0.21)	(0.13)	(0.024)	(3.0e-03)	(4.5e-04)	(4.9e-05)
Uniform	MISE	0.32	0.10	0.015	2.9e-03	1.5e-03	1.3e-03	1.3e-03	1.3e-03
	sd	(0.25)	(0.12)	(0.018)	(2.1e-03)	(2.6e-04)	(4.4e-05)	(1.2e-05)	(3.6e-06)
Beta	MISE	0.46	0.19	0.035	5.4e-03	1.7e-03	1.4e-03	1.3e-03	1.3e-03
	sd	(0.27)	(0.17)	(0.040)	(5.3e-03)	(6.5e-04)	(8.6e-05)	(2.2e-05)	(5.8e-06)
Exp. loc.	MISE	0.55	0.47	0.29	0.11	0.015	2.9e-03	1.5e-03	1.3e-03
	sd	(0.20)	(0.23)	(0.21)	(0.11)	(0.018)	(1.9e-03)	(2.6e-04)	(5.3e-05)



(a) Exp. mixture (a), $m_{\text{best}} = 3$ (b) Exp. mixture (b), $m_{\text{best}} = 3$ (c) Gamma mixture, $m_{\text{best}} = 4$

(d) Uniform mixture, $m_{\text{best}} = 5$ (e) Beta mixture, $m_{\text{best}} = 4$ (f) Exp. location, $m_{\text{best}} = 4$

FIGURE 1. 10 estimators $\hat{f}_{m_{\text{best}},n}$ (black) in six different settings with $n = 10^5$ when the mixing density f is a Beta distribution (red).

is approximated in the same polynomial basis, the MISE tends to the same value, which is obviously the squared bias of the estimator when $m = 5$. In the exponential mixture settings, different values are obtained because different bases are used to approach f. The exponential mixture setting from Example 1 (a) always has the largest mean MISE value, while the uniform and the Beta mixtures are doing best.

Figure 1 illustrates the estimator $\hat{f}_{m,n}$ when $n = 10^5$ and where the order m is the value minimizing the MISE when $n = 10^5$, say m_{best}. The values of m_{best} have been obtained by extra simulations. We see that in the first two settings, we only have $m_{\text{best}} = 3$ and the estimator seems slightly biased. On the contrary, the uniform mixture setting allows for the best approximation with $m_{\text{best}} = 5$.

APPENDIX A. TECHNICAL RESULTS

Lemma 5. *Let $\alpha > 0$, $a < b$ and $a' < b'$. Define $\tilde{\mathcal{C}}(\alpha, C)_{\mathbb{H}}$ as in (20) and $\tilde{\mathcal{C}}'(\alpha, C)_{\mathbb{H}}$ similarly with a' and b' replacing a and b. Let σ be $[\alpha]+1$ differentiable on $[a, b]$ and $\tau : [a', b'] \to [a, b]$ be $[\alpha]+1$ differentiable on $[a', b']$ with a non-vanishing first derivative. Then*

$$\left\{\sigma f \,:\, f \in \tilde{\mathcal{C}}(\alpha, \cdot)_{\mathbb{H}}\right\} \hookrightarrow \tilde{\mathcal{C}}(\alpha, \cdot)_{\mathbb{H}} \quad \textit{and} \quad \left\{f \circ \tau \,:\, f \in \tilde{\mathcal{C}}(\alpha, \cdot)_{\mathbb{H}}\right\} \hookrightarrow \tilde{\mathcal{C}}'(\alpha, \cdot)_{\mathbb{H}} \,.$$

Proof. As the first embedding is the inclusion (40) in Roueff and Ryden [22], we only show the second embedding. Let $f \in \tilde{\mathcal{C}}(\alpha, C)$ and denote $r = [\alpha] + 1$. Let $t \in (0, 1]$. By the equivalence (19) with the K-functional given in (18) there exists a function h such that $h^{(r-1)} \in A.C._{\text{loc}}$ and

$$\|f - h\|_H + t^r \|\varphi^r h^{(r)}\|_{\mathbb{H}} \le 2M\omega^r_\varphi(f, t)_{\mathbb{H}} \le 2MCt^\alpha \,, \tag{42}$$

where $\varphi(x) = \sqrt{(x-a)(b-x)}$. Let us set $\tilde{h} = h \circ \tau$ and show that, for some constant $K > 0$ neither depending on t nor C,

$$\|f \circ \tau - \tilde{h}\|_{\mathbb{H}'} + t^r \|\tilde{\varphi}^r \, \tilde{h}^{(r)}\|_{\mathbb{H}'} \le K\, C\, t^\alpha \,, \tag{43}$$

where we defined $\tilde{\varphi}(x) = \sqrt{(x-a')(b'-x)}$, that is the same definition as φ with a' and b' replacing a and b. Using again equivalence (19), the bound given in (43) will achieve the proof of the lemma.

Note that since τ' does not vanish, denoting $C_1 = (\inf |\tau'|)^{-1}$, for all $g \in \mathbb{H}$, we have

$$\|g \circ \tau\|_{H'} \le C_1 \|g\|_H \,. \tag{44}$$

In particular we have that

$$\|f \circ \tau - \tilde{h}\|_{\mathbb{H}'} = \|(f - h) \circ \tau\|_{\mathbb{H}'} \le C_1 \|f\|_H \,. \tag{45}$$

Since τ is r times continuously differentiable and $h^{(r-1)} \in A.C._{\text{loc}}$, we note that $\tilde{h}^{(r-1)} \in A.C._{\text{loc}}$ with $\tilde{h}^{(r)} = \sum_{j=1}^r \tau_j \times h^{(j)} \circ \tau$, where the τ_j's are continuous functions only depending on τ. Hence there is a constant $C_2 > 0$ only depending on τ and r such that

$$\|\tilde{\varphi}^r \, \tilde{h}^{(r)}\|_{\mathbb{H}'} \le C_2 \max_{j=1,\dots,r} \|\tilde{\varphi}^r \, h^{(j)} \circ \tau\|_{\mathbb{H}'} \,.$$

Another simple consequence of τ' not vanishing on $[a', b']$ is that there exists a constant $C_3 > 0$ such that $\tilde{\varphi}(x) \le C_3 \varphi \circ \tau(x)$ for all $x \in [a', b']$. Using this with (44) in the previous display, we get

$$\|\tilde{\varphi}^r \, \tilde{h}^{(r)}\|_{\mathbb{H}'} \le C_1\, C_2\, C_3 \max_{j=1,\dots,r} \|\varphi^r \, h^{(j)}\|_{\mathbb{H}} \,. \tag{46}$$

We shall prove that $\|\varphi^r \, h^{(j)}\|_{\mathbb{H}}$ appearing in the right-hand side of the previous inequality is in fact maximized, up to multiplicative and additive constants, at $j = r$. For $j = 1, \dots, r-1$, we proceed recursively as follows. For any $u \in (a, b)$, we have

$$|h^{(j)}(x)| \le \left| \int_u^x h^{(j+1)}(s) \mathrm{d}s \right| + |h^{(j)}(u)| \,.$$

Then, by Jensen's inequality,

$$\|\varphi^r\, h^{(j)}\|_{\mathbb{H}}$$
$$\leq \left\{ \int_{x=a}^{b} \varphi^{2r}(x) \left(|x-u| \int_{s\in[u,x]} \{h^{(j+1)}(s)\}^2 \mathrm{d}s \right) \mathrm{d}x \right\}^{1/2} + \|\varphi^r\|_{\mathbb{H}}\, |h^{(j)}(u)| \,,$$

where we used the convention that $[c,d]$ denotes the same segment whether $c \leq d$ or not. By Fubini's theorem, the term between braces reads

$$\int_{s=a}^{b} \{h^{(j+1)}(s)\}^2 \psi(s;u) \mathrm{d}s \quad \text{with} \quad \psi(s;u) = \int 1_{[u,x]}(s)\, (x-a)^r (b-x)^r |x-u| \,\mathrm{d}x \,.$$

Let $\tilde{a} < \tilde{b}$ be two fixed numbers in (a,b). It is straightforward to show that, for some constant $C_4 > 0$ only depending on $a, b, \tilde{a}, \tilde{b}$, we have

$$\psi(s;u) \leq C_4^2\, \varphi^{2r}(s) \quad \text{for all } u \in (\tilde{a}, \tilde{b}) \,.$$

The last 3 displays thus give that

$$\|\varphi^r\, h^{(j)}\|_{\mathbb{H}} \leq C_4 \|\varphi^r\, h^{(j+1)}\|_{\mathbb{H}} + \|\varphi^r\|_{\mathbb{H}} \inf_{u\in[\tilde{a},\tilde{b}]} |h^{(j)}(u)| \,.$$

By induction on j, we thus get with (46) that there is a constant C_5 such that

$$\|\tilde{\varphi}^r\, \tilde{h}^{(r)}\|_{\mathbb{H}'} \leq C_5 \left(\|\varphi^r\, h^{(r)}\|_{\mathbb{H}} + \sum_{j=1,\ldots,r-1} \inf_{u\in[\tilde{a},\tilde{b}]} |h^{(j)}(u)| \right) \,. \tag{47}$$

The final step of the proof consists in bounding $\inf_{u\in[\tilde{a},\tilde{b}]} |h^{(j)}(u)|$ for $j = 1, \ldots, r-1$. Let $\delta_j = \inf_{u\in[\tilde{a},\tilde{b}]} |h^{(j)}(u)|$. Then for any $v, v' \in [\tilde{a},\tilde{b}]$, we have $|h^{(j-1)}(v') - h^{(j-1)}(v)| \geq \delta_j\, |v'-v|$. Suppose that v is in the first third part of the segment $[\tilde{a},\tilde{b}]$ and v' in the last third so that $|v-v'| \geq (\tilde{b}-\tilde{a})/3$. On the other hand $|h^{(j-1)}(v') - h^{(j-1)}(v)| \leq |h^{(j-1)}(v')| + |h^{(j-1)}(v)|$. It follows that $|h^{(j-1)}(v')|$ and $|h^{(j-1)}(v)|$ cannot be both less than $\delta_j\, (\tilde{b}-\tilde{a})/3$, which provides a lower bound of $|h^{(j-1)}|$ on at least one sub-interval of $[\tilde{a},\tilde{b}]$ of length $(\tilde{b}-\tilde{a})/3$. Proceeding recursively we get that there exists a sub-interval of $[\tilde{a},\tilde{b}]$ on which h is lower bounded by δ_j multiplied by some constant. This in turns gives that

$$\sum_{j=1,\ldots,r-1} \inf_{u\in[\tilde{a},\tilde{b}]} |h^{(j)}(u)| \leq C_6 \|h\|_{\mathbb{H}} \,.$$

where C_6 is a constant only depending on $\tilde{a}, \tilde{b}$ and r. Observe that, since $f \in \tilde{C}(\alpha, C)$, we have $\|f\|_{\mathbb{H}} \leq C$. Using (42), $t \in (0,1]$ and $\|h\|_{\mathbb{H}} \leq \|f-h\|_{\mathbb{H}} + \|f\|_{\mathbb{H}}$ in the last display we thus get

$$\sum_{j=1,\ldots,r-1} \inf_{u\in[\tilde{a},\tilde{b}]} |h^{(j)}(u)| \leq C_6(2M+1)C \,.$$

Finally, this bound, (47), (45) and (42) yields (43) and the proof is achieved. □

Lemma 6. *Let (p_k) be the sequence of polynomials defined by $p_1(t) = 1$, $p_2(t) = t$, ..., $p_k(t) = t(t+1)\ldots(t+k-2)$ for all $k \geq 2$. Define the coefficients $(\tilde{c}_{k,l})_{1\leq l\leq k}$ by the expansion formula $t^{k-1} = \sum_{l=1}^{k} \tilde{c}_{k,l} p_l(t)$, valid for $k = 1, 2, \ldots$. Then $\tilde{c}_{1,1} = 1$, and for all $k \geq 2$,*

$$\tilde{c}_{k,1} = 0\,, \tilde{c}_{k,k} = 1 \quad \textit{and} \quad \tilde{c}_{k,l} = \tilde{c}_{k-1,l-1} - (l-1)\tilde{c}_{k-1,l} \quad \textit{for all} \quad l = 2, \ldots, k-1 \,. \tag{48}$$

Moreover, we have, for all $k \geq 1$,

$$\sum_{l=1}^{k} |\tilde{c}_{k,l}| \leq k! \,. \tag{49}$$

Proof. By definition of p_l, we have $tp_l(t) = p_{l+1}(t) - (l-1)p_l(t)$ for any $l \geq 1$. Hence, for any $k \geq 2$, writing $t^{k-1} = tt^{k-2} = \sum_l \tilde{c}_{k-1,l} t p_l(t)$, we obtain (48).

We now prove (49). It is obviously true for $k = 1$. From (48), it follows that, for all $k \geq 2$,

$$\sum_{l=1}^{k} |\tilde{c}_{k,l}| \leq \sum_{l=1}^{k-1} l\, |\tilde{c}_{k-1,l}| + 1 \ .$$

Bounding l inside the last sum by $(k-1)$ yields (49). □

REFERENCES

[1] Kim E. Andersen and Martin B. Hansen. Multiplicative censoring: density estimation by a series expansion approach. *Journal of Statistical Planning and Inference*, 98:137–155, 2001.

[2] Masoud Asgharian, Marco Carone, and Vahid Fakoor. Large-sample study of the kernel density estimators under multiplicative censoring. *Annals of Statistics*, 40(1):159–187, 2012.

[3] Fadoua Balabdaoui and Jon A. Wellner. Estimation of a k-monotone density: limit distribution theory and the spline connection. *Annals of Statistics*, 35(6):2536–2564, 2007.

[4] Denis Belomestny and John Schoenmakers. Statistical Skorohod embedding problem and its generalizations. Technical report, Arxiv, 2014. URL `http://arxiv.org/abs/1407.0873`.

[5] A. Beurling and P. Malliavin. On Fourier transforms of measures with compact support. *Acta Math.*, 107:291–309, 1962. ISSN 0001-5962.

[6] Fabienne Comte and Valentine Genon-Catalot. Adaptive laguerre density estimation for mixed poisson models. available at `https://hal.archives-ouvertes.fr/hal-00848158`, 2014.

[7] R. A. DeVore and G. G. Lorentz. *Constructive Approximation*. Springer, 1993.

[8] Z. Ditzian and V. Totik. *Moduli of Smoothness*. Springer Series in Computational Mathematics. Springer-Verlag, 1987.

[9] Jianqing Fan. On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.*, 19(3):1257–1272, 1991. ISSN 0090-5364. doi: 10.1214/aos/1176348248. URL `http://dx.doi.org/10.1214/aos/1176348248`.

[10] Jianqing Fan. Global behavior of deconvolution kernel estimates. *Statist. Sinica*, 1(2):541–551, 1991. ISSN 1017-0405.

[11] Jianqing Fan. Adaptively local one-dimensional subproblems with application to a deconvolution problem. *Ann. Statist.*, 21(2):600–610, 1993. ISSN 0090-5364. doi: 10.1214/aos/1176349139. URL `http://dx.doi.org/10.1214/aos/1176349139`.

[12] Feller. *An Introduction to Probability Theory and Its Applications*, volume 2. Wiley, New York, 2nd edition, 1971.

[13] Constantinos Goutis. Nonparametric estimation of a mixing density via the kernel method. *Journal of the American Statistical Association*, 92(440):1445–1450, 1997.

[14] Nicolas W. Hengartner. Adaptive demixing in poisson mixture models. *Annals of Statistics*, 25(3): 917–928, 1997.

[15] N. P. Jewell. Mixtures of exponential distributions. *Annals of Statistics*, 10(2):479–484, 1982.

[16] Paul Kvam. Length bias in the measurements of carbon nanotubes. *Technometrics,*, 50(4):462–467, 2008.

[17] Nan Laird. Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association*, 73(364):805–811, 1978.

[18] J. R. Lakowicz. *Principles of Fluorescence Spectroscopy*. Academic/Plenum, New York, 1999.

[19] Bruce G. Lindsay. The geometry of mixture likelihoods: A general theory. *Annals of Statistics*, 11 (1):86–94, 1983.

[20] Felipe Olmos, Bruno Kauffmann, Alain Simonian, and Yannick Carlinet. Catalog dynamics: Impact of content publishing and perishing on the performance of a LRU cache. available at http://arxiv.org/abs/1403.5479, 2014.

[21] M. Pensky and B. Vidakovic. Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.*, 27(6):2033–2053, 1999. ISSN 0090-5364. doi: 10.1214/aos/1017939249. URL http://dx.doi.org/10.1214/aos/1017939249.

[22] Francois Roueff and Tobias Ryden. Nonparametric estimation of mixing densities for discrete distributions. *Annals of Statistics*, 33:2066–2108, 2005.

[23] Bernard Valeur. *Molecular Fluorescence*. Wiley-VCH, Weinheim, 2002.

[24] Y. Vardi. Multiplicative censoring, renewal processes, deconvolution and decreasing density: Nonparametric estimation. *Biometrika*, 76:751–761, 1989.

[25] Sergio Venturini, Francesca Dominici, and Giovanni Parmigiani. Gamma shape mixtures for heavy-tailed distributions. *Ann. Appl. Stat.*, 2(2):756–776, 2008. ISSN 1932-6157. doi: 10.1214/07-AOAS156. URL http://dx.doi.org/10.1214/07-AOAS156.

[26] Cun-Hui Zhang. Fourier methods for estimating mixing densities and distributions. *Annals of Statistics*, 18:806–831, 1990.

[27] Cun-Hui Zhang. On estimating mixing densities in discrete exponential family models. *Annals of Statistics*, 23:929–945, 1995.