

Adaptive density estimation for clustering with Gaussian mixtures

Cathy Maugis-Rabuseau

*Institut de Mathématiques de Toulouse, INSA de Toulouse, Université de Toulouse
INSA de Toulouse,
135, avenue de Rangueil,
31077 Toulouse Cedex 4, France.
e-mail: cathy.maugis@insa-toulouse.fr*

and

Bertrand Michel

*Laboratoire de Statistique Théorique et Appliquée,
Université Pierre et Marie Curie - Paris 6,
4 place Jussieu, 75252 Paris cedex 05 France
e-mail: bertrand.michel@upmc.fr*

Abstract: Gaussian mixture models are widely used to study clustering problems. These model-based clustering methods require an accurate estimation of the unknown data density by Gaussian mixtures. In Maugis and Michel (2009), a penalized maximum likelihood estimator is proposed for automatically selecting the number of mixture components. In the present paper, a collection of univariate densities whose logarithm is locally β -Hölder with moment and tail conditions are considered. We show that this penalized estimator is minimax adaptive to the β regularity of such densities in the Hellinger sense.

AMS 2000 subject classifications: Primary 62G07; secondary 62G20.

Keywords and phrases: Rate adaptive density estimation, Gaussian mixture clustering, Hellinger risk, Non asymptotic model selection.

1. Introduction

Clustering methods consists of discovering clusters among observations. Many cluster analysis methods have been proposed in statistics and learning theory, roughly fall into three categories. The first one is based on similarity or dissimilarity distances, the best-known are partitioned clustering methods as k-means and the hierarchical clustering methods (see for instance Sections 14.3.6 and 14.3.12 in [Hastie et al., 2009](#)). The second category consists of density level set clustering methods which consider clusters as the connected components of high density regions (see [Hartigan, 1975](#)). The third category is composed of model-based clustering methods which define clusters as observations having most likely the same distribution. In this last case, each subpopulation is assumed to be distributed from a parametric density, like a Gaussian one and thus the unknown data density is a mixture of these distributions (see for instance [McLachlan and Peel, 2000](#)). The data clustering is then deduced thanks to the maximum a posteriori (MAP) rule. The clustering problem being based on data density estimation, it is then essential that this density be efficiently estimated.

Because of their wide range flexibility, Gaussian mixture densities are widely used to model the unknown distribution of continuous data for clustering analysis (see for instance [Lindsay, 1995](#);

McLachlan and Peel, 2000). By recasting the clustering problem into a model selection problem, we have proposed in Maugis and Michel (2009) a non asymptotic penalized criterion. We proved that the selected Gaussian mixture estimator fulfills an oracle inequality. The aim of this new paper is to investigate the adaptive properties of this estimator in order to justify the validity of our clustering method. More precisely, adapting a recent approximation result, we show that our estimator is minimax adaptive to the regularity parameter of a particular class of Hölder spaces defined further. As far as we know, such a minimax adaptive result has never been shown for a density estimator used for model-based clustering methods.

We first recall the context of Maugis and Michel (2009) in the unidimensional case. Let us consider n independent identically distributed random variables X_1, \dots, X_n with values in \mathbb{R} . Their common unknown density s belongs to the set \mathcal{S} of all density functions with respect to the Lebesgue measure on \mathbb{R} . The considered unidimensional Gaussian mixtures are characterized by their number of components m and their means and variances parameters are assumed to be bounded. These mixture densities are grouped into a model collection $(\mathcal{S}_m)_{m \in \mathcal{M}_n}$, subsets of \mathcal{S} , defined by

$$\mathcal{S}_m = \left\{ x \in \mathbb{R} \mapsto \sum_{u=1}^m p_u \psi_{\sigma_u}(x - \mu_u); \mu_u \in [-\bar{\mu}, \bar{\mu}], \sigma_u^2 \in [\underline{\lambda}, \bar{\lambda}]; p_u \in [0, 1], \sum_{u=1}^m p_u = 1 \right\} \quad (1)$$

where ψ is the Gaussian kernel defined by $\psi(x) = \pi^{-\frac{1}{2}} \exp(-x^2)$ for all $x \in \mathbb{R}$ and $\psi_{\sigma}(\cdot) = \sigma^{-1} \psi(\frac{\cdot}{\sigma})$ for all $\sigma > 0$. The number of free parameters, common to all the mixture densities of a given model \mathcal{S}_m is called dimension and is denoted $D(m)$. Considering a non asymptotic point of view (see for instance Massart, 2007), the three bounds $\bar{\mu}$, $\underline{\lambda}$ and $\bar{\lambda}$ of each model \mathcal{S}_m and also the maximum number of mixture components in the collection may depend on n . Such mixtures are called sieves according to the terminology introduced by Grenander (1981).

Over each model \mathcal{S}_m , a maximum likelihood estimator (MLE) \hat{s}_m is obtained by minimizing the empirical contrast

$$\gamma_n(t) = -\frac{1}{n} \sum_{i=1}^n \ln \{t(X_i)\}.$$

The loss function associated to the likelihood contrast is the Kullback-Leibler divergence: For two densities s and t in \mathcal{S} , the Kullback-Leibler divergence is defined by

$$\text{KL}(s, t) = \int \ln \left\{ \frac{s(x)}{t(x)} \right\} s(x) dx$$

if $s dx$ is absolutely continuous with respect to $t dx$ and $+\infty$ otherwise. The model m^* in the collection minimizing the Kullback-Leibler risk

$$m^* \in \underset{m \in \mathcal{M}_n}{\text{argmin}} \mathbb{E}_s[\text{KL}(s, \hat{s}_m)]$$

is considered as the “best” model of the collection. Nevertheless this best model m^* and also the associated density \hat{s}_{m^*} (called oracle) are unknown since they depend on the true density s . A model \hat{m} is then chosen by minimizing over \mathcal{M}_n the following penalized criterion

$$\text{crit}(m) = \gamma_n(\hat{s}_m) + \text{pen}(m).$$

The penalty function $\text{pen} : m \in \mathcal{M}_n \mapsto \text{pen}(m) \in \mathbb{R}^+$ has to be chosen such that the Kullback-Leibler risk $\mathbb{E}_s[\text{KL}(s, \hat{s}_{\hat{m}})]$ of $\hat{s}_{\hat{m}}$ is close to the oracle risk $\mathbb{E}_s[\text{KL}(s, \hat{s}_{m^*})]$. The construction of such

penalties is proposed in Theorem 2.2 in [Maugis and Michel \(2009\)](#). This theorem can be stated as follows in the univariate context, where $d_H(g, h) = \frac{1}{\sqrt{2}} \|\sqrt{g} - \sqrt{h}\|_2$ denotes the Hellinger distance between two densities g and h of \mathcal{S} :

Theorem 1. *There exists two absolute constants κ and C such that, if*

$$\text{pen}(m) \geq \kappa \frac{D(m)}{n} \left\{ 1 + 2\mathcal{A}^2 + \ln \left(\frac{1}{1 \wedge \frac{D(m)}{n} \mathcal{A}^2} \right) \right\}$$

where

$$\mathcal{A} = \sqrt{\ln(6\pi e^2)} + \sqrt{\pi} + \sqrt{\ln \left(\bar{\mu} \sqrt{\frac{8}{c_1 \bar{\lambda}}} \right)} + \sqrt{\ln \left(\frac{144 \bar{\lambda}}{\bar{\lambda}} \right)}, \quad (2)$$

then the model \hat{m} minimizing

$$\text{crit}(m) = \gamma_n(\hat{s}_m) + \text{pen}(m)$$

over \mathcal{M}_n exists and

$$\mathbb{E} [d_H^2(s, \hat{s}_{\hat{m}})] \leq C \left[\inf_{m \in \mathcal{M}_n} \{ \text{KL}(s, \mathcal{S}_m) + \text{pen}(m) \} + \frac{1}{n} \right]. \quad (3)$$

Note that a similar result can be found in [Maugis and Michel \(2009\)](#) for multivariate data clustering with variable selection. The method has been successfully implemented and tested in practice (see [Maugis and Michel, 2010](#)).

Minimax adaptive estimation has been intensively studied in nonparametric statistics, see for instance [Tsybakov \(2009\)](#), and [Massart \(2007\)](#) for adaptive minimax methods based on l_0 penalization. A natural optimality criterion is the minimax risk, first introduced by [Wolfowitz \(1950\)](#). Let

$$\mathcal{R}(\tilde{s}_n, \mathcal{H}_\beta) = \sup_{s \in \mathcal{H}_\beta} \mathbb{E}_s [d_H^2(s, \tilde{s}_n)]$$

be the maximal Hellinger risk of an estimator \tilde{s}_n of s . The minimax Hellinger risk on a density class \mathcal{H}_β is then defined by

$$\mathcal{R}_n(\mathcal{H}_\beta) = \inf_{\tilde{s}_n} \mathcal{R}(\tilde{s}_n, \mathcal{H}_\beta)$$

where the infimum is taken over all the possible estimators \tilde{s}_n of s . An estimator is said to be minimax on \mathcal{H}_β if its maximal risk over \mathcal{H}_β reaches the minimax risk on this density class. Let us now consider a collection $(\mathcal{H}_\beta)_{\beta \in \mathcal{B}}$ of density classes indexed by a set \mathcal{B} of regularity parameters β . An estimator is said to be minimax adaptive if it reaches the minimax risk over \mathcal{H}_β for all β of \mathcal{B} , without using the knowledge of β . In order to motivate the clustering method based on Gaussian mixture estimator $\hat{s}_{\hat{m}}$ proposed in [Maugis and Michel \(2009\)](#), we prove in this new paper that this estimator is minimax adaptive over a particular collection of Hölder density classes $(\mathcal{H}_\beta)_{\beta \in \mathcal{B}}$ defined further. Of course, adaptive density estimation in one dimension is now a classical problem and several adaptive estimators have been already proposed such as kernel estimators or thresholding wavelet estimators. Nevertheless, although these alternative methods maybe perform better than our penalized estimator $\hat{s}_{\hat{m}}$ concerning density estimation in general, these have no interest for clustering purposes.

The link between model selection and adaptive estimation is made through approximation theory. Indeed, an adaptive estimation is possible only for functional classes \mathcal{H}_β that can be efficiently approximated by our Gaussian mixture collection. Convolution is widely used in approximation theory

and many results are known on this topic. It is well known that the convolution of a density f with scaled versions ψ_σ of the Gaussian kernel ψ converges to f (see for instance [Cheney and Light, 2009](#), chapter 20). The so-called quasi-interpolation method consists of replacing the functions $\psi_\sigma * f$ by infinite linear combinations of scaled and translated Gaussian kernels (see for instance [Cheney and Light, 2009](#), chapter 36). In a recent paper of [Hangelsbroek and Ron \(2010\)](#), a nonlinear approximation algorithm based on finite combinations of scaled and translated Gaussian kernels is defined to give some approximation results in L^p norm on some particular density classes. Nevertheless, all these results cannot be straightly applied to study the approximation capacities of Gaussian mixtures. Indeed, the coefficients in these linear combinations are not necessary positive and their sum is not constrained to be equal to one. Furthermore, the approximation results provided by all these methods are not given for the Kullback-Leibler divergence as required by our statistical context.

The approximation capacity of Gaussian mixtures has also been studied in non parametric Bayesian works. Lemma 3.1 in [Ghosal and van der Vaart \(2001\)](#) gives a discretization result for Gaussian mixtures: assume that s is a location or location-scale mixture with a mixing distribution compactly supported or with sub-Gaussian tails, s can be approximated by a finite Gaussian mixture with a small number of components, the error being controlled in L_1 and L_∞ norms. In [Ghosal and van der Vaart \(2007\)](#), these authors take advantage of this method for approximating by finite Gaussian mixtures some twice continuously differentiable functions with additional regularity conditions. More recently, [Kruijer et al. \(2010\)](#) prove an approximation result by finite Gaussian mixtures for densities whose logarithm is locally Hölder. Their approximation result is given for the Kullback-Leibler divergence. This last result can be successfully adapted in our context to control the bias term in the right side term of the oracle inequality [\(3\)](#) on these particular density classes. Concerning approximation, the contribution of our work consists of checking that the non explicit constants of the approximation bounds given in [Kruijer et al. \(2010\)](#) are actually uniform over a density class \mathcal{H}_β we define. For easier reading, all the approximation results are given and proved in this preprint version although a large part of them can be found in [Kruijer et al. \(2010\)](#).

The paper is organized as follows: The main results are presented in Section 2. The density classes \mathcal{H}_β are introduced in Section 2.1 and an approximation result, adapted of [Kruijer et al. \(2010\)](#), is given in Section 2.2. Next, a lower bound of the minimax risk is given in Section 2.3 and the adaptive property of our penalized Gaussian mixture estimator on these density classes \mathcal{H}_β is addressed in Section 2.4. The approximation result, the lower bound and the adaptive result are respectively proved in Sections 3, 4 and 5. Finally, some technical results are developed in Appendices A and B.

2. Main results

2.1. The density classes $\mathcal{H}(\beta, \mathcal{P})$

The adaptation result given further requires a slightly modified version of the approximation result by finite Gaussian mixtures proved in [Kruijer et al. \(2010\)](#). This approximation result concerns densities whose logarithm is locally β -Hölder and that fulfills additional tail, moments and monotonicity conditions. More precisely, let $\beta > 0$, $r = \lfloor \beta \rfloor$ be the largest integer less than β and $k \in \mathbb{N}$ such that $\beta \in (2k, 2k + 2]$. Let also \mathcal{P} be the set of parameters $\{\gamma, l^+, L, \varepsilon, C, \alpha, \xi, M\}$ where L is a polynomial function on \mathbb{R} and the other parameters are positive constants. We then define the density class $\mathcal{H}(\beta, \mathcal{P})$ of all densities f satisfying the following conditions:

1. **Smoothness.** $\ln f$ is assumed to be locally β -Hölder: for all x and y such that $|y - x| \leq \gamma$,

$$\left| (\ln f)^{(r)}(x) - (\ln f)^{(r)}(y) \right| \leq r! L(x) |y - x|^{\beta-r}. \quad (4)$$

Furthermore for all $j \in \{0, \dots, r\}$,

$$|(\ln f)^{(j)}(0)| \leq l^+. \quad (5)$$

2. **Moments.** The derivative functions $(\ln f)^{(j)}$ for $j = 1, \dots, r$ and the polynomial function L fulfill

$$\int_{\mathbb{R}} \left| (\ln f)^{(j)}(x) \right|^{\frac{2\beta+\varepsilon}{j}} f(x) dx \leq C, \quad \int_{\mathbb{R}} |L(x)|^{2+\frac{\varepsilon}{\beta}} f(x) dx \leq C. \quad (6)$$

3. **Tail.** For all $x \in \mathbb{R}$,

$$f(x) \leq M\psi(x). \quad (7)$$

4. **Monotonicity.** f is strictly positive, f is nondecreasing on $(-\infty, -\alpha)$ and nonincreasing on (α, ∞) , and $f(x) \geq \xi$ for all $x \in [-\alpha, \alpha]$.

Remarks 1. The monotonicity assumption can be relaxed by assuming that there exist two constants $c > 0$ and $\bar{\sigma} > 0$ such that $\forall 0 < \sigma < \bar{\sigma}, \forall x \in \mathbb{R}$,

$$\frac{K_{\sigma} f(x)}{f(x)} \geq c.$$

This condition corresponds to the first point given in Lemma 13 in Appendix A which is a key point to prove the approximation result. In the following, the strong monotonicity condition is assumed in the definition of the density class $\mathcal{H}(\beta, \mathcal{P})$ to simplify the proofs of the lower bound.

Remarks 2. For easier reading, the monotonicity assumption is stated on a symmetric interval but it is possible to consider this assumption on a general interval $[\alpha_1, \alpha_2]$ with $\alpha_1 < \alpha_2$. This monotonicity assumption allows us to lower bound the convolution $f * \psi_{\sigma}$ by f up to a multiplicative constant according to Remark 3 in Ghosal et al. (1999).

Remarks 3. These density classes are more restrictive than those considered in Kruijer et al. (2010): Indeed the upper bounds in (6) have to be uniform on the density class $\mathcal{H}(\beta, \mathcal{P})$ and we also need the additional Condition (5). These restrictions allow us to control the Kullback-Leibler divergence between a density of $\mathcal{H}(\beta, \mathcal{P})$ and a convenient finite Gaussian mixture, uniformly over $\mathcal{H}(\beta, \mathcal{P})$. Note that Condition (7) is here assumed on \mathbb{R} but it could be assumed only outside an interval as in Kruijer et al. (2010).

Remarks 4. In the sequel, \mathcal{P}' is said to be "larger than" \mathcal{P} if at least one of the following conditions is fulfilled:

- at least one constant among M , C or l^+ of \mathcal{P}' is larger than the corresponding one of \mathcal{P} ,
- the constant γ of \mathcal{P}' is smaller than the corresponding one of \mathcal{P} ,
- for all $x \in \mathbb{R}$, $L(x) \leq L'(x)$ where L (resp. L') belongs to \mathcal{P} (resp. \mathcal{P}')

2.2. Approximation result

For any function f , $K_\sigma f$ denotes the convolution $f * \psi_\sigma$ and $\Delta_\sigma f$ is the error term $K_\sigma f - f$. As explained in Kruijjer et al. (2010), for a β -smooth density f with $\beta \leq 2$ and under reasonable regularity assumptions, it is possible to define a finite location-scale Gaussian mixture \wp_σ such that $\text{KL}(f, \wp_\sigma) = O(\sigma^{2\beta})$. The usual approach consists of discretizing the continuous mixture $K_\sigma f$. But as $\|f - K_\sigma f\|_\infty$ remains of order σ^2 when $\beta > 2$, this approach appears to be inefficient for smoother densities. An alternative strategy is proposed in Kruijjer et al. (2010), based on the following successive convolutions of f : $f_0 = f$ and for all $j \geq 0$, $f_{j+1} = f - \Delta_\sigma f_j$. In their paper, the density is approximated by a discretized version of the continuous mixture $K_\sigma f_k$ where $k \in \mathbb{N}$ is such that $\beta \in (2k, 2k + 2]$.

In our framework, Lemma 4 in Kruijjer et al. (2010) cannot be directly used since the upper bound over the Kullback-Leibler divergence between f and the finite Gaussian mixture is not uniform over $\mathcal{H}(\beta, \mathcal{P})$. Thus some additional work is necessary in order to prove an uniform version of this approximation result. Another reason for revisiting the approximation results given in Kruijjer et al. (2010) is that these ones are stated for $\sigma \leq \bar{\sigma}$ where $\bar{\sigma}$ depends on the approximated density f . Thus we also need to check that it is possible to choose the same $\bar{\sigma}$ for all the densities of $\mathcal{H}(\beta, \mathcal{P})$. The proof of Theorem 2 consists of carefully following the method of Kruijjer et al. (2010) in order to obtain this uniform version. A sketch of the proof is given below and a self-contained proof is detailed in Section 3.

Theorem 2. *There exists a positive constant $\bar{\sigma}(\beta) < 1$ such that for all $f \in \mathcal{H}(\beta, \mathcal{P})$ and for all $\sigma < \bar{\sigma}(\beta)$, there exists a finite Gaussian mixture of density \wp_σ with less than $G_\beta \sigma^{-1} |\ln \sigma|^{\frac{3}{2}}$ support points, with the same variance σ for each component and with means belonging to $[-\mu_\sigma, \mu_\sigma]$ where*

$$\mu_\sigma \leq \tilde{G}_\beta |\ln \sigma|^{\frac{1}{2}}$$

such that

$$\text{KL}(f, \wp_\sigma) = \int_{\mathbb{R}} f(x) \ln \left(\frac{f(x)}{\wp_\sigma(x)} \right) dx \leq c_\beta \sigma^{2\beta} \quad (8)$$

where c_β is uniform on $\mathcal{H}(\beta, \mathcal{P})$ and continuous on β . The constant $\bar{\sigma}(\beta)$ only depends on $\mathcal{H}(\beta, \mathcal{P})$ and is a continuous function of β . Moreover, G_β and \tilde{G}_β are two positive constants that only depend on $\mathcal{H}(\beta, \mathcal{P})$, and are both increasing functions of β .

The two constants G_β and \tilde{G}_β are explicitly defined by Equations (55) and (56) in the proof of Theorem 2 in Section 3.2.

Sketch of the proof. Let f be a density in a given class $\mathcal{H}(\beta, \mathcal{P})$. First, the convolution $K_\sigma f_k$ is shown to be close to f on a subspace of \mathbb{R} where the derivative functions of $\ln f$ and L are efficiently controlled (see Lemma 1). On this subspace, the difference $K_\sigma f_k - f$ is controlled by $f(x)R_f(x)O(\sigma^\beta)$, apart from a term σ^H where H can be arbitrarily large. The term $O(\sigma^\beta)$ is uniform on $\mathcal{H}(\beta, \mathcal{P})$ and R_f is a polynomial function of L and the derivative functions of $\ln f$. Next, since f_k is not necessarily a positive function, a density function h_k is defined from f_k . The previous result is then adapted for controlling $K_\sigma h_k - f$ on a more restrictive subspace of \mathbb{R} (see Lemma 2). Based on this result, a control of the Kullback-Leibler divergence between f and the continuous Gaussian mixture $K_\sigma h_k$ is obtained in Proposition 2: $\text{KL}(f, K_\sigma h_k) \leq c_\beta \sigma^{2\beta}$ where c_β is a multiplicative constant uniform on $\mathcal{H}(\beta, \mathcal{P})$. Finally, using a discretization result, a similar control is obtained for $\text{KL}(f, \wp_\sigma)$ where \wp_σ is a finite Gaussian mixture fulfilling conditions given in Theorem 2. \square

2.3. Lower bound

In order to show that the MLE penalized estimator $\hat{s}_{\hat{m}}$ is adaptive to the smoothness parameter β , a lower bound of the minimax risk $\mathcal{R}_n(\mathcal{H}(\beta, \mathcal{P}))$ is required. For all $0 < \underline{\beta} < \bar{\beta}$, a “large enough” parameter set $\mathcal{P}(\underline{\beta}, \bar{\beta})$ is found such that for all $\beta \in [\underline{\beta}, \bar{\beta}]$, $\mathcal{H}(\beta, \mathcal{P}(\underline{\beta}, \bar{\beta}))$ is well defined and a lower bound is given for the density classes $\mathcal{H}(\beta, \mathcal{P}(\underline{\beta}, \bar{\beta}))$. Note that in Theorem 2, the constants c_β , $\bar{\sigma}(\beta)$, G_β and \tilde{G}_β cannot be bounded uniformly for all $\beta \in \mathbb{R}^+$. Nevertheless, it can be proved that $\hat{s}_{\hat{m}}$ is minimax adaptive on a range of regularity $[\underline{\beta}, \bar{\beta}]$.

First, the parameter set $\mathcal{P}(\underline{\beta}, \bar{\beta})$ has to be defined rigorously. Its definition is rather technical since it depends on the way the lower bound is proved. The proof is based on the construction of some oscillating functions, this standard method is presented for instance in Massart (2007, see Section 7.5). Let us take some infinitely differentiable function $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ with compact support included into $(\frac{1}{4}, \frac{3}{4})$ such that

$$\int_{\mathbb{R}} \varphi(x) dx = 0 \text{ and } \int_{\mathbb{R}} \varphi(x)^2 dx = 1.$$

We set $A = \max_{0 \leq k \leq r+1} \|\varphi^{(k)}\|_\infty > 1$ and let D be some positive even integer. For any positive integer $j \in \{1, \dots, D\}$, we consider the function

$$\begin{aligned} \varphi_j : \mathbb{R} &\rightarrow \mathbb{R} \\ x &\mapsto \frac{\xi D^{-\beta}}{A} \varphi\left(\frac{D}{\alpha}\left(x + \frac{\alpha}{2}\right) - (j-1)\right). \end{aligned}$$

Moreover, let $\mathcal{T}(\alpha, \xi)$ be the space of functions $\omega : \mathbb{R} \rightarrow \mathbb{R}^+$ such that ω is nondecreasing on $(-\infty, -\frac{\alpha}{2})$, nonincreasing on $(\frac{\alpha}{2}, +\infty)$, $\omega(x) = 2\xi$ for all $x \in [-\frac{3\alpha}{4}, \frac{3\alpha}{4}]$, and $\omega(-\alpha) = \omega(\alpha) = \xi$.

Next, let $\tilde{\mathcal{P}} = \left\{ \frac{\alpha}{4}, \ln(2\xi), \tilde{L}, \tilde{\varepsilon}, \tilde{C}, \alpha, \xi, \tilde{M} \right\}$ be a parameter set such that $\mathcal{T}(\alpha, \xi) \cap \mathcal{H}(\beta, \tilde{\mathcal{P}})$ is nonempty. Based on a function $\omega \in \mathcal{T}(\alpha, \xi) \cap \mathcal{H}(\beta, \tilde{\mathcal{P}})$ and the functions φ_j , we consider the functional space $\mathcal{J}(\beta, D) = \{f_\theta; \theta \in \{0, 1\}^D\}$ where for all $\theta \in \{0, 1\}^D$ and for all $x \in \mathbb{R}$,

$$f_\theta(x) = \omega(x) + \sum_{j=1}^D (2\theta_j - 1) \varphi_j(x). \quad (9)$$

Proposition 1. *There exists a parameter set $\mathcal{P}(\underline{\beta}, \bar{\beta})$ such that for all $D \in \mathbb{N}^*$ and for all $\beta \in [\underline{\beta}, \bar{\beta}]$,*

$$\mathcal{J}(\beta, D) \subset \mathcal{H}(\beta, \mathcal{P}(\underline{\beta}, \bar{\beta})).$$

Remarks 5. *Note that if such a parameter set exists, Proposition 1 is also true for all the parameter sets larger than it (in the sense given in Remark 4). A key point to prove the lower bound stated in the next theorem is that the parameter set $\mathcal{P}(\underline{\beta}, \bar{\beta})$ does not depend on D .*

Theorem 3. *Suppose that one observes independent random variables X_1, \dots, X_n with common density s with respect the Lebesgue measure on \mathbb{R} . For any $\beta \in [\underline{\beta}, \bar{\beta}]$ and any parameter set $\mathcal{P}(\underline{\beta}, \bar{\beta})$ given by Proposition 1, there exists a positive constant κ_β such that*

$$\mathcal{R}_n(\mathcal{H}(\beta, \mathcal{P}(\underline{\beta}, \bar{\beta}))) := \inf_{\tilde{s}} \sup_s \mathbb{E}[d_H^2(s, \tilde{s})] \geq \kappa_\beta n^{-\frac{2\beta}{2\beta+1}}$$

where the supremum (resp. the infimum) is taken over all densities s in $\mathcal{H}(\beta, \mathcal{P}(\underline{\beta}, \bar{\beta}))$ (resp. over all possible estimators \tilde{s} of s).

Proposition 1 and Theorem 3 are proved in Section 4.1 and Section 4.2 respectively. After establishing Proposition 1, the Hellinger distance and the Kullback-Leibler divergence between two functions of $\mathcal{J}(\beta, D)$ are controlled in Lemma 5 and Lemma 6 respectively. These controls are required to combine a corollary of a Birgé's Lemma (see Birgé, 2005) and the so-called Varshamov-Gilbert's Lemma. These last two results can be found in Massart (2007, see Corollary 2.19 and Lemma 4.7) and are reminded in Appendix B.

2.4. Adaptive density estimation

In a non asymptotic model selection approach, the model collection may increase with the sample size n , leading to an adaptive procedure. As it was already explained, the adaptive properties of $\hat{s}_{\hat{m}}$ are studied on a range of regularity $[\underline{\beta}, \bar{\beta}]$. Preliminary, we fix $0 < \underline{\beta} < \bar{\beta}$ and we also choose $a_{\bar{\beta}} > 1$ large enough such that

$$\frac{G_{\bar{\beta}}}{a_{\bar{\beta}}} \left(\frac{\ln a_{\bar{\beta}}}{\ln 2} + 3 \right)^{3/2} \leq 1, \quad (10)$$

where $G_{\bar{\beta}}$ is defined in Theorem 2. The parameters of the Gaussian mixture models $(\mathcal{S}_m)_{m \in \mathcal{M}_n}$ are now specified in order to apply the approximation results provided by Theorem 2:

$$\mathcal{S}_m = \left\{ x \in \mathbb{R} \mapsto \sum_{u=1}^m p_u \psi_{\sigma_u}(x - \mu_u); \mu_u \in [-\bar{\mu}(m), \bar{\mu}(m)], \sigma_u^2 \in [\underline{\lambda}(m), \bar{\lambda}(m)], p_u \in [0, 1], \sum_{u=1}^m p_u = 1 \right\}$$

where $\sqrt{\underline{\lambda}(m)} := a_{\bar{\beta}} m^{-1} (\ln m)^{3/2}$, $\bar{\mu}(m) = \tilde{G}_{\bar{\beta}} |\ln \sqrt{\underline{\lambda}(m)}|^{1/2}$ and $\bar{\lambda}(m) > \underline{\lambda}(m)$ for all m . Note that the last parameter $\bar{\lambda}(m)$ can be taken the same for all m and is denoted $\bar{\lambda}$ in the sequel. Since a n -sample is observed, it is natural to suppose that the number of mixture components m is less than n and we also assume that the mixtures have at least two components: $\mathcal{M}_n = \{2, \dots, n\}$. Note that when the sample size n increases, mixtures with small component variances and many components m are available in the model collection. This obviously improves the approximation capacity of the Gaussian mixtures.

Theorem 4. *Assume that $n \geq 3$ and let $\hat{s}_{\hat{m}}$ be the penalized maximum likelihood estimator minimizing the penalized criterion defined in Theorem 1. Then there exists a constant $c_{\underline{\beta}, \bar{\beta}}$ such that for all $\beta \in [\underline{\beta}, \bar{\beta}]$ and for all $s \in \mathcal{H}(\beta, \mathcal{P}(\underline{\beta}, \bar{\beta}))$,*

$$\mathbb{E} [d_H^2(s, \hat{s}_{\hat{m}})] \leq c_{\underline{\beta}, \bar{\beta}} (\ln n)^{\frac{5\beta}{2\beta+1}} n^{\frac{-2\beta}{2\beta+1}}.$$

Theorem 4 shows that the penalized estimator $\hat{s}_{\hat{m}}$ is adaptive on the regularity β of the density classes defined in Section 2.1, up to a power of $\ln(n)$. This logarithm term is due to the penalty shape given in Theorem 1. It is not detected in practice as shown in Maugis and Michel (2010) and we suspect that it could be removed from the penalty shape. Note that the non parametric Bayesian estimator defined in Kruijer et al. (2010) has a similar rate of convergence with a greater power of the logarithm term.

3. Proof of the approximation result

In this section, the density functional space $\mathcal{H}(\beta, \mathcal{P})$ is fixed. To make the proofs and the results easier to read, we use the notation c_β (resp. $\bar{\sigma}(\beta)$) for denoting constants (resp. upper bound on σ)

that only depends on β and \mathcal{P} . We also use the notation $c_{\beta,p}$ (resp. $\bar{\sigma}(\beta,p)$) if it also depends on another parameter p . Moreover, we introduce the following notation: For any nonnegative integer h , the h -fold convolution of the Gaussian kernel ψ is denoted ψ^{*h} and for any nonnegative t , the t -th moment of ψ^{*h} is defined by $\nu_{h,t} = \int x^t \psi^{*h}(x) dx$. We also denote as $l_j(\cdot)$ the j th derivative $\frac{d^j}{dx^j} \ln f(x)$ of $\ln f$ and we consider a subset A_σ defined by

$$A_\sigma := \left\{ x \in \mathbb{R}; |l_j(x)| \leq \mathfrak{B}\sigma^{-j} |\ln \sigma|^{-j/2}, j = 1 \dots r, L(x) \leq \mathfrak{B}\sigma^{-\beta} |\ln \sigma|^{-\beta/2} \right\}$$

if $\beta > 1$ and $A_\sigma := \{x \in \mathbb{R}; L(x) \leq \mathfrak{B}\sigma^{-\beta} |\ln \sigma|^{-\beta/2}\}$ otherwise.

3.1. Approximation by a continuous mixture

Lemma 1. *Let $\beta > 0$ and $k \in \mathbb{N}$ such that $\beta \in (2k, 2k+2]$. For all $H > 0$, there exists $\bar{\sigma}(\beta, H) > 0$ such that for all $\sigma < \bar{\sigma}(\beta, H)$, for all $f \in \mathcal{H}(\beta, \mathcal{P})$ and for all $x \in A_\sigma$ we have*

$$(K_\sigma f_k)(x) = f(x) [1 + R_f(x) O_{\beta, H}(\sigma^\beta)] + O_{\beta, H}(\sigma^H)$$

with $R_f(x) = a_{r+1}L(x)$ if $\beta \leq 1$, and

$$R_f(x) = a_{r+1}L(x) + \sum_{j=1}^r a_j |l_j(x)|^{\frac{\beta}{j}}$$

otherwise. In both cases, the a_j 's are nonnegative constants that are uniform on $\mathcal{H}(\beta, \mathcal{P})$. Furthermore, $\bar{\sigma}(\beta, H)$ is a continuous function of β and H .

Proof. Let $H > 0$ and $f \in \mathcal{H}(\beta, \mathcal{P})$. If $\beta > 1$, for all x and y such that $|y - x| \leq \gamma$, there exists ρ such that $|x - \rho| \leq |x - y|$ and

$$\ln f(y) = \ln f(x) + \sum_{j=1}^r \frac{l_j(x)}{j!} (y-x)^j + \frac{l_r(\rho) - l_r(x)}{r!} (y-x)^r.$$

Then, the smoothness condition (4) implies, since $|\rho - x| \leq |y - x| \leq \gamma$, that

$$\begin{aligned} \left| \ln f(y) - \ln f(x) - \sum_{j=1}^r \frac{l_j(x)}{j!} (y-x)^j \right| &\leq L(x) |\rho - x|^{\beta-r} |y-x|^r \\ &\leq L(x) |y-x|^\beta. \end{aligned}$$

Thus we have

$$\ln f(y) \leq \ln f(x) + B(x, y) \tag{11}$$

and

$$\ln f(y) \geq \ln f(x) + \tilde{B}(x, y) \tag{12}$$

with $B(x, y) = \sum_{j=1}^r \frac{l_j(x)}{j!} (y-x)^j + L(x) |y-x|^\beta$ and $\tilde{B}(x, y) = \sum_{j=1}^r \frac{l_j(x)}{j!} (y-x)^j - L(x) |y-x|^\beta$. Note that for $\beta \leq 1$, (11) and (12) are valid with $B(x, y) = -\tilde{B}(x, y) = L(x) |x-y|^\beta$.

Let $x \in A_\sigma$ and $y \in D_x := \{y \in \mathbb{R}; |y - x| \leq k'\sigma |\ln \sigma|^{1/2}\}$ where k' , chosen below, has to be identical for all $f \in \mathcal{H}(\beta, \mathcal{P})$ and we also assume that σ is small enough to satisfy

$$k'\sigma |\ln \sigma|^{1/2} \leq \gamma. \quad (13)$$

Then (11) gives that for all $y \in D_x$, $f(y) \leq f(x) \exp[B(x, y)]$ and thus

$$K_\sigma f(x) \leq f(x) \int_{D_x} e^{B(x, y)} \psi_\sigma(y - x) dy + \int_{D_x^c} f(y) \psi_\sigma(y - x) dy. \quad (14)$$

For the sequel, note that, for $x \in A_\sigma$ and $y \in D_x$, if $\beta \geq 1$,

$$\begin{aligned} |B(x, y)| &\leq \sum_{j=1}^r \frac{k'^j}{j!} \sigma^j |\ln \sigma|^{j/2} \mathfrak{B} \sigma^{-j} |\ln \sigma|^{-j/2} + k'^\beta \sigma^\beta |\ln \sigma|^{\beta/2} \mathfrak{B} \sigma^{-\beta} |\ln \sigma|^{-\beta/2} \\ &\leq \mathfrak{B} \sum_{j=1}^r \frac{k'^j}{j!} + k'^\beta := d_1(\beta, k') \end{aligned}$$

and thus

$$\begin{aligned} e^{B(x, y)} &\leq \sum_{j=0}^r \frac{1}{j!} B^j(x, y) + |B^{r+1}(x, y)| \left| \sum_{j \geq 0} \frac{1}{(j+r+1)!} B^j(x, y) \right| \\ &\leq \sum_{j=0}^r \frac{1}{j!} B^j(x, y) + |B^{r+1}(x, y)| \left| \sum_{j \geq 0} \frac{1}{j!} d_1(\beta, k')^j \right| \\ &\leq \sum_{j=0}^r \frac{1}{j!} B^j(x, y) + d_2(\beta, k') |B^{r+1}(x, y)| \end{aligned} \quad (15)$$

with $d_2(\beta, k') = \exp[d_1(\beta, k')]$.

Case $k=0$: We consider that $\beta \in (1, 2]$ thus $r = 1$. The case $\beta \in (0, 1]$ is discussed hereafter. We have $B(x, y) = l_1(x)(y - x) + L(x)|y - x|^\beta$ and (15) yields

$$\begin{aligned} e^{B(x, y)} &\leq 1 + B(x, y) + d_2(\beta, k') B^2(x, y) \\ &\leq 1 + l_1(x)(y - x) + L(x)|y - x|^\beta \\ &\quad + d_2(\beta, k') [l_1(x)^2(y - x)^2 + 2L(x)l_1(x)|y - x|^\beta(y - x) + L^2(x)|y - x|^{2\beta}] \\ &\leq 1 + l_1(x)(y - x) + L(x)|y - x|^\beta \\ &\quad + d_2(\beta, k') [(\mathfrak{B}k')^{2-\beta} |l_1(x)|^\beta |y - x|^\beta + 2\mathfrak{B}k'^\beta l_1(x)(y - x) + \mathfrak{B}k'^\beta L(x)|y - x|^\beta] \end{aligned}$$

since $|l_1(x)(y - x)| \leq \mathfrak{B}k'$ and $|L(x)(y - x)^\beta| \leq \mathfrak{B}k'^\beta$. Since ψ_σ is symmetric, $\int_{D_x} (y - x) \psi_\sigma(y - x) dy = 0$ and thus,

$$\begin{aligned} \int_{D_x} e^{B(x, y)} \psi_\sigma(y - x) dy &\leq 1 + L(x) \int_{D_x} \psi_\sigma(y - x) |y - x|^\beta dy \\ &\quad + d_2(\beta, k') [(\mathfrak{B}k')^{2-\beta} |l_1(x)|^\beta + \mathfrak{B}k'^\beta L(x)] \int_{D_x} \psi_\sigma(y - x) |y - x|^\beta dy \\ &\leq 1 + \{d_2(\beta, k')(\mathfrak{B}k')^{2-\beta} |l_1(x)|^\beta + [1 + d_2(\beta, k')\mathfrak{B}k'^\beta] L(x)\} \nu_{1, \beta} \sigma^\beta \\ &\leq 1 + d_3(\beta, k') [2L(x) + |l_1(x)|^\beta] \sigma^\beta. \end{aligned} \quad (16)$$

Let $k' = k'(2, 0, 1, H)$ given by Lemma 8, and $\bar{\sigma}(\beta, H)$ such that (13) is satisfied. Note that $\bar{\sigma}(\beta, H)$ can be taken as a continuous function of β and H . Then, for the second integral in the right hand of (14), using (7), it gives

$$\begin{aligned} \int_{D_x^c} f(y)\psi_\sigma(y-x)dy &= \int_{\mathbb{R}} f(y)\psi_\sigma(y-x)\mathbb{1}_{|y-x|\geq k'\sigma|\ln\sigma|^{\frac{1}{2}}}(y)dy \\ &\leq M \int_{|u|\geq k'|\ln\sigma|^{\frac{1}{2}}} \psi(u)du \\ &\leq c_{\beta,H}\sigma^H. \end{aligned} \quad (17)$$

Furthermore, since k' depends on H , $d_3(\beta, k')$ can be also rewritten as a constant $c_{\beta,H}$. Finally, (14), (16) and (17) give that

$$(K_\sigma f)(x) \leq f(x) \{1 + c_{\beta,H}R_f(x)\sigma^\beta\} + c_{\beta,H}\sigma^H$$

with $R_f(x) = 2L(x) + |l_1(x)|^\beta$.

For $\beta \in (0, 1]$, (14) is still valid with $B(x, y) = L(x)|y-x|^\beta$. For $x \in A_\sigma$, the first integral in (14) can be treated as in the case $\beta \in (1, 2]$: it yields

$$\int_{D_x} \exp(L(x)|x-y|^\beta)\psi_\sigma(x-y)dy \leq 1 + c_\beta L(x)\nu_{1,\beta}\sigma^\beta.$$

Using Lemma (8) as before, it gives

$$\int_{D_x^c} f(y)\psi_\sigma(x-y) \leq c_{\beta,H}\sigma^H$$

and finally, for all $x \in A_\sigma$,

$$K_\sigma f(x) \leq f(x) [1 + c_{\beta,H}L(x)\sigma^\beta] + c_{\beta,H}\sigma^\beta.$$

A similar lower bound can be shown in the same way, it is proved in the general case further.

Case $k=1$: We consider $\beta \in (3, 4]$, a similar proof gives the result for $\beta \in (2, 3]$. According to (15), for $x \in A_\sigma$ and $y \in D_x$,

$$e^{B(x,y)} \leq 1 + B(x, y) + \frac{1}{2}B(x, y)^2 + \frac{1}{6}B(x, y)^3 + d_2(\beta, k')B(x, y)^4 \quad (18)$$

with $B(x, y) = l_1(x)(y-x) + \frac{1}{2}l_2(x)(y-x)^2 + \frac{1}{6}l_3(x)(y-x)^3 + L(x)|y-x|^\beta$. Thus $\exp[B(x, y)]$ is upper bounded by a linear combination of terms of the form

$$[L(x)|x-y|^\beta]^{\eta_4} \prod_{j=1}^3 [l_j(x)(y-x)^j]^{\eta_j}$$

with $\sum_{j=1}^4 \eta_j \leq 4$. Let $\mathcal{A}_1(x, y)$ be the sum of such terms for which $\eta_4\beta + \sum_{j=1}^3 j\eta_j < \beta$ and let $\mathcal{A}_2(x, y)$ be the others terms and thus $e^{B(x,y)} \leq \mathcal{A}_1(x, y) + \mathcal{A}_2(x, y)$. Note that the constant $d_2(k', \beta)$

only appears in the terms of $\mathcal{A}_2(x, y)$. By removing inside $\mathcal{A}_1(x, y)$ all the terms for which the power of $(y - x)$ is an odd integer (since $\int u^t \psi_\sigma(u) du = 0$ if t is an odd integer), it yields

$$\begin{aligned} \int_{D_x} \mathcal{A}_1(x, y) \psi_\sigma(y - x) dy &= \int_{D_x} \psi_\sigma(y - x) \left\{ 1 + \frac{1}{2} [l_1^2(x) + l_2(x)] (y - x)^2 \right\} dy \\ &\leq 1 + \frac{\nu_{1,2}}{2} \{l_1^2(x) + l_2(x)\} \sigma^2. \end{aligned} \quad (19)$$

Next, for each term of $\mathcal{A}_2(x, y)$, we have for all $x \in A_\sigma$ and all $y \in D_x$

$$\begin{aligned} [L(x)|x - y|^\beta]^{\eta_4} \prod_{j=1}^3 |l_j(x)(y - x)^j|^{\eta_j} &= \left\{ [L(x)|x - y|^\beta]^{\eta_4} \prod_{j=1}^3 |l_j(x)(x - y)^j|^{\eta_j} \right\}^{\frac{\eta_4 \beta + \sum_{j=1}^3 j \eta_j - \beta}{\eta_4 \beta + \sum_{j=1}^3 j \eta_j}} \\ &\quad \times \left\{ L(x)^{\eta_4} \prod_{j=1}^3 |l_j(x)|^{\eta_j} \right\}^{\frac{\beta}{\eta_4 \beta + \sum_{j=1}^3 j \eta_j}} |x - y|^\beta \end{aligned}$$

and finally

$$[L(x)|x - y|^\beta]^{\eta_4} \prod_{j=1}^3 |l_j(x)(y - x)^j|^{\eta_j} \leq c_\beta \left\{ L(x)^{\eta_4} \prod_{j=1}^3 |l_j(x)|^{\eta_j} \right\}^{\frac{1}{\eta_4 + \sum_{j=1}^3 \frac{j}{\beta} \eta_j}} |x - y|^\beta$$

according to Lemma 11. Note that

$$\begin{aligned} \frac{1}{\eta_4 + \sum_{j=1}^3 \frac{j}{\beta} \eta_j} \ln \left\{ L(x)^{\eta_4} \prod_{j=1}^3 |l_j(x)|^{\eta_j} \right\} &= \frac{1}{\eta_4 + \sum_{j=1}^3 \frac{j}{\beta} \eta_j} \left\{ \eta_4 \ln L(x) + \sum_{j=1}^3 \eta_j \frac{j}{\beta} \ln |l_j(x)|^{\beta/j} \right\} \\ &\leq \ln \left\{ \frac{1}{\eta_4 + \sum_{j=1}^3 \frac{j}{\beta} \eta_j} \left(\eta_4 L(x) + \sum_{j=1}^3 \eta_j \frac{j}{\beta} |l_j(x)|^{\beta/j} \right) \right\}. \end{aligned}$$

since the logarithm function is concave. According to Lemma 11, for each term of $\mathcal{A}_2(x, y)$ we thus have

$$[L(x)|x - y|^\beta]^{\eta_4} \prod_{j=1}^3 |l_j(x)(y - x)^j|^{\eta_j} \leq \frac{c_\beta}{\eta_4 + \sum_{j=1}^3 \frac{j}{\beta} \eta_j} |x - y|^\beta \left\{ \eta_4 L(x) + \sum_{j=1}^3 \eta_j \frac{j}{\beta} |l_j(x)|^{\beta/j} \right\}$$

and then

$$|\mathcal{A}_2(x, y)| \leq c_{\beta, k'} \left\{ a_4 L(x) + \sum_{j=1}^3 a_j |l_j(x)|^{\beta/j} \right\} |x - y|^\beta$$

where $c_{\beta, k'}$ comes from $d_2(\beta, k')$ in (18) and where the a_j 's are positive constants that only depend on $\mathcal{H}(\beta, \mathcal{P})$. It leads to

$$\int_{D_x} |\mathcal{A}_2(x, y)| \psi_\sigma(y - x) dy \leq c_{\beta, k'} R_f(x) \sigma^\beta \quad (20)$$

where $R_f(x) = a_4 L(x) + \sum_{j=1}^3 a_j |l_j(x)|^{\beta/j}$. Finally, (19) and (20) together yield

$$\int_{D_x} e^{B(x,y)} \psi_\sigma(y-x) dy \leq 1 + \frac{\nu_2}{2} \{l_1^2(x) + l_2(x)\} \sigma^2 + c_{\beta,k'} R_f(x) \sigma^\beta.$$

Let $k' = k'(2, 0, 1, H)$ given by Lemma 8, and $\bar{\sigma}(\beta, H)$ such that (13) is satisfied and where $\bar{\sigma}(\beta, H)$ can be taken as a continuous function of β and H . Next, for all $f \in \mathcal{H}(\beta, \mathcal{P})$, and all $\sigma < \bar{\sigma}(\beta, H)$,

$$\int_{D_x^c} f(y) \psi_\sigma(y-x) dy \leq c_{\beta,H} \sigma^H.$$

Finally, for $\sigma < \bar{\sigma}(\beta, H)$,

$$(K_\sigma f)(x) \leq f(x) \left[1 + \frac{\nu_{1,2}}{2} \{l_1^2(x) + l_2(x)\} \sigma^2 + c_{\beta,H} R_f(x) \sigma^\beta \right] + c_{\beta,H} \sigma^H$$

and the similar lower bound is obtained in the same way, see the general case further. Thus,

$$(K_\sigma f)(x) = f(x) \left[1 + \frac{\nu_{1,2}}{2} \{l_1^2(x) + l_2(x)\} \sigma^2 + R_f(x) O_{\beta,H}(\sigma^\beta) \right] + O_{\beta,H}(\sigma^H). \quad (21)$$

Now, we need a similar result for f_1 instead of f . Equation (21) depends on the kernel ψ through the values of $\nu_{1,2}$. In fact, it holds for any symmetric kernel ϕ such that $\int \phi(x) x^t dx = \nu_{1,t} < \infty$ and $\int_{|x| > k' |\ln \sigma|^{1/2}} \phi(x) |x|^t dx = O_{\beta,H}(\sigma^H)$ when k' is large enough. For ψ^{*2} , these properties follow from Lemma 8 : let k' and $\bar{\sigma}(\beta, H)$ such that (13) is satisfied and where $\bar{\sigma}(\beta, H)$ is a continuous function of β and H . Thus, denoting $\nu_{2,u} = \int x^u \psi^{*2}(x) dx$, for all $\sigma < \bar{\sigma}(\beta, H)$,

$$(K_\sigma^2 f)(x) = f(x) \left[1 + \frac{\nu_{2,2}}{2} \{l_1^2(x) + l_2(x)\} \sigma^2 + R_f(x) O_{\beta,H}(\sigma^\beta) \right] + O_{\beta,H}(\sigma^H).$$

Now, since $f_1 = 2f - K_\sigma f$ and $\nu_{2,2} = 2\nu_{1,2}$, it yields for all $\sigma < \bar{\sigma}(\beta, H)$ that

$$(K_\sigma f_1)(x) = f(x) \left[1 + R_f(x) O_{\beta,H}(\sigma^\beta) \right] + O_{\beta,H}(\sigma^H).$$

General case: Let $\beta \in (2k, 2k+2]$. We give the main ideas of the proof in the general case. According to (15), for $x \in A_\sigma$ and $y \in D_x$, $\exp[B(x, y)]$ is upper bounded by a linear combination of terms of the form

$$[L(x) |x-y|^\beta]^{\eta_{r+1}} \prod_{j=1}^r [l_j(x) (y-x)^j]^{\eta_j}$$

with $\sum_{j=0}^r \eta_j \leq r+1$. We then decompose $e^{B(x,y)}$ into $\mathcal{A}_1(x, y)$ and $\mathcal{A}_2(x, y)$ as before. By removing inside $\mathcal{A}_1(x, y)$ all the terms for which the power of $(y-x)$ is an odd integer, it yields

$$\int_{D_x} \mathcal{A}_1(x, y) \psi_\sigma(y-x) dy \leq 1 + \sum_{u=1}^k \nu_{1,2u} Q_u(x) \sigma^{2u} \quad (22)$$

where the Q_u 's are positive functions that can be expressed in function of L and the l_u 's. Following the same method as for $\beta \in (3, 4]$, it yields

$$|\mathcal{A}_2(x, y)| \leq c_{\beta,k'} \left\{ a_{r+1} L(x) + \sum_{j=1}^r a_j |l_j(x)|^{\beta/j} \right\} |x-y|^\beta \quad (23)$$

and

$$\int_{D_x} |\mathcal{A}_2(x, y)| \psi_\sigma(y - x) dy \leq c_{\beta, k'} R_f(x) \sigma^\beta \quad (24)$$

where $R_f(x) = a_{r+1} L(x) + \sum_{j=1}^r a_j |l_j(x)|^{\frac{\beta}{j}}$. Finally, (22) and (24) together yield

$$\int_{D_x} e^{B(x, y)} \psi_\sigma(y - x) dy \leq 1 + \sum_{u=1}^k \nu_{1, 2u} Q_u(x) \sigma^{2u} + c_{\beta, k'} R_f(x) \sigma^\beta.$$

Using Lemma 8, let k' depending on H and $\bar{\sigma}(\beta, H) > 0$ such that (13) is satisfied and where $\bar{\sigma}(\beta, H)$ is a continuous function of β and H . For all $\sigma < \bar{\sigma}(\beta, H)$,

$$\int_{D_x^c} f(y) \psi_\sigma(y - x) dy \leq c_{\beta, H} \sigma^H.$$

Finally, since k' depends on H ,

$$(K_\sigma f)(x) \leq f(x) \left[1 + \sum_{u=1}^k \nu_{1, 2u} Q_u(x) \sigma^{2u} + c_{\beta, H} R_f(x) \sigma^\beta \right] + c_{\beta, H} \sigma^H$$

and the similar lower bound is obtained in the same way (see further). Thus,

$$(K_\sigma f)(x) = f(x) \left[1 + \sum_{u=1}^k \nu_{1, 2u} Q_u(x) \sigma^{2u} + R_f(x) O_{\beta, H}(\sigma^\beta) \right] + O_{\beta, H}(\sigma^H).$$

Now, we need a similar result for f_k instead of f . According to Lemma 12,

$$f_k = \sum_{i=0}^k (-1)^i \binom{k+1}{i+1} K_\sigma^i f. \quad (25)$$

For all $h \leq k$, the same method can be applied with ψ^{*h} instead of ψ and it yields with the same functions Q_u and R_f : for $\sigma < \bar{\sigma}(\beta, H)$,

$$(K_\sigma^h f)(x) = f(x) \left[1 + \sum_{u=1}^k \nu_{h, 2u} Q_u(x) \sigma^{2u} + R_f(x) O_{\beta, H}(\sigma^\beta) \right] + O_{\beta, H}(\sigma^H).$$

According to (25), for $\sigma < \bar{\sigma}(\beta, H)$

$$\begin{aligned} K_\sigma f_k(x) &= \sum_{i=0}^k (-1)^i \binom{k+1}{i+1} K_\sigma^{i+1} f(x) \\ &= \sum_{j=1}^{k+1} (-1)^{j+1} \binom{k+1}{j} K_\sigma^j f(x) \\ &= \sum_{j=1}^{k+1} (-1)^{j+1} \binom{k+1}{j} f(x) \left[1 + \sum_{u=1}^k \nu_{j, 2u} Q_u(x) \sigma^{2u} + R_f(x) O_{\beta, H}(\sigma^\beta) \right] + O_{\beta, H}(\sigma^H) \\ &= f(x) \left\{ 1 + \sum_{u=1}^{k+1} \left[\sum_{j=1}^{k+1} (-1)^{j+1} \binom{k+1}{j} \nu_{j, 2u} \right] Q_u(x) \sigma^{2u} + R_f(x) O_{\beta, H}(\sigma^\beta) \right\} + O_{\beta, H}(\sigma^H) \end{aligned}$$

and then

$$K_\sigma f_k(x) = f(x) \{ 1 + R_f(x) O_{\beta,H}(\sigma^\beta) \} + O_{\beta,H}(\sigma^H)$$

since $\sum_{i=0}^k (-1)^i \binom{k+1}{i+1} = 1$ and $\sum_{j=1}^{k+1} (-1)^{j+1} \binom{k+1}{j} \nu_{j,2u} = 0$ according to Lemma 9.

To complete this proof, we give the method for obtaining the lower bound in the general case. Using (12) and proceeding in the same way as for the upper bound, it yields

$$\begin{aligned} K_\sigma f(x) &\geq f(x) \int_{D_x} e^{\tilde{B}(x,y)} \psi_\sigma(y-x) dy \\ &\geq f(x) \int_{D_x} \left[1 + \sum_{j=1}^r \frac{1}{j!} \tilde{B}^j(x,y) - d_2(\beta, k') \left| \tilde{B}^{r+1}(x,y) \right| \right] \psi_\sigma(y-x) dy \\ &\geq f(x) \int_{D_x} \mathcal{A}_1(x,y) \psi_\sigma(y-x) dy + f(x) \int_{D_x} \mathcal{A}_2(x,y) \psi_\sigma(y-x) dy \end{aligned} \quad (26)$$

where $\mathcal{A}_1(x,y)$ (resp. $\mathcal{A}_2(x,y)$) contains the terms which powers are less than β (resp. larger than β). For the first integral,

$$\begin{aligned} \int_{D_x} \mathcal{A}_1(x,y) \psi_\sigma(y-x) dy &= \int_{\mathbb{R}} \mathcal{A}_1(x,y) \psi_\sigma(y-x) dy - \int_{D_x^c} \mathcal{A}_1(x,y) \psi_\sigma(y-x) dy \\ &\geq 1 + \sum_{u=1}^k \nu_{1,2u} Q_u(x) \sigma^{2u} - \left| \int_{D_x^c} \mathcal{A}_1(x,y) \psi_\sigma(y-x) dy \right|. \end{aligned}$$

Note that $\mathcal{A}_1(x,y)$ is a linear combination of terms of the form $\prod_{j=1}^r [l_j(x)(y-x)^j]^{\eta_j}$, where $\sum_{j=1}^r j\eta_j$ is even. Since $x \in A_\sigma$, and since $|x-y| \leq 1$, we can find $h > 0$ and a constant c_β that only depends on β such that

$$|\mathcal{A}_1(x,y)| \leq c_\beta \sigma^{-h} (x-y)^2.$$

Finally,

$$\left| \int_{D_x^c} \mathcal{A}_1(x,y) \psi_\sigma(y-x) dy \right| \leq c_\beta \sigma^{-h} \int_{D_x^c} (x-y)^2 \psi_\sigma(y-x) dy.$$

Then we apply Lemma 8 with $H' > h + H$ and it gives that

$$\left| \int_{D_x^c} \mathcal{A}_1(x,y) \psi_\sigma(y-x) dy \right| \leq c_{\beta,H} \sigma^H.$$

To find a lower bound for the second integral in (26), we note that according to (23),

$$|\mathcal{A}_2(x,y)| \geq -c_{\beta,H} \left\{ a_{r+1} L(x) + \sum_{j=1}^r a_j |l_j(x)|^{\beta/j} \right\} |x-y|^\beta$$

and thus

$$\int_{D_x} |\mathcal{A}_2(x,y)| \psi_\sigma(y-x) dy \leq c_{\beta,H} R_f(x) \sigma^\beta$$

where R_f is defined as before. We finally obtain that for $\sigma < \bar{\sigma}(\beta, H)$,

$$K_\sigma f(x) \geq f(x) \left[1 + \sum_{u=1}^k \nu_{1,2u} Q_u(x) \sigma^{2u} + c_{\beta,H} R_f(x) \sigma^\beta \right] + c_{\beta,H} \sigma^H.$$

□

For a density f belonging to $\mathcal{H}(\beta, \mathcal{P})$, Lemma 1 shows that the convolution $K_\sigma f_k$ is close to f on a subspace of \mathbb{R} where the derivative functions of $\ln f$ and L are efficiently controlled. Furthermore, the control on the difference $K_\sigma f_k - f$ is uniform over $\mathcal{H}(\beta, \mathcal{P})$, which is required to upper bound the Kullback-Leibler divergence between f and $K_\sigma f_k$. Thus $K_\sigma f_k$ seems to be a good candidate to approximate the density function f . Nevertheless, the function f_k is not a density function: Its integral over \mathbb{R} is equal to 1 (see Lemma 12) but it can take negative values. To remedy this problem, Kruijer et al. (2010) define a density function h_k as follows: Considering the subspace

$$J_{\sigma,k} = \left\{ x \in \mathbb{R}; f_k(x) > \frac{1}{2} f(x) \right\},$$

the following positive function is defined

$$\forall x \in \mathbb{R}, g_k(x) = f_k(x) \mathbb{1}_{J_{\sigma,k}}(x) + \frac{1}{2} f(x) \mathbb{1}_{J_{\sigma,k}^c}(x)$$

and it is normalized to obtain a density function

$$\forall x \in \mathbb{R}, h_k(x) = \frac{g_k(x)}{\int g_k(u) du}. \quad (27)$$

Note that the constant 1/2 is arbitrary in the definition of $J_{\sigma,k}$, any other number of $(0, 1)$ could be used.

Now, the result of Lemma 1 has to be extended for the convolution $K_\sigma h_k$. For this purpose, the integral of $K_\sigma^t f$ for all nonnegative integers $t \leq k$ is controlled over A_σ^c and E_σ^c where A_σ is defined by (3) and $E_\sigma = \{x \in \mathbb{R}; f(x) \geq \sigma^{H_1}\}$ with $H_1 > 4\beta$.

Lemma 2. *Let $\beta > 0$ and $k \in \mathbb{N}$ such that $\beta \in (2k, 2k + 2]$. There exists $\bar{\sigma}(\beta, H_1) > 0$ such that for all $\sigma \leq \bar{\sigma}(\beta, H_1)$, for all $f \in \mathcal{H}(\beta, \mathcal{P})$ and for all nonnegative integers $t \leq k$,*

$$\int_{A_\sigma^c} (K_\sigma^t f)(x) dx = O_\beta(\sigma^{2\beta}) \quad (28)$$

and

$$\int_{E_\sigma^c} (K_\sigma^t f)(x) dx = O_{\beta, H_1}(\sigma^{2\beta}). \quad (29)$$

Furthermore, for $\sigma \leq \bar{\sigma}(\beta, H_1)$, $A_\sigma \cap E_\sigma \subset J_{\sigma,k}$ and

$$\int_{\mathbb{R}} g_k(x) dx = 1 + O_{\beta, H_1}(\sigma^{2\beta}). \quad (30)$$

Thus, for all $H > 0$, there exists $\bar{\sigma}(\beta, H_1, H) > 0$ such that for all $\sigma \leq \bar{\sigma}(\beta, H_1, H)$ and for all $x \in A_\sigma \cap E_\sigma$,

$$|(K_\sigma h_k)(x) - f(x)| = f(x)R_f(x)O_{\beta, H_1, H}(\sigma^\beta) + O_{\beta, H_1, H}(\sigma^H). \quad (31)$$

Furthermore, $\bar{\sigma}(\beta, H_1)$ and $\bar{\sigma}(\beta, H_1, H)$ are both continuous functions of β , H_1 and H for the last one.

Remarks 6. The left term in (30) does not depend on H_1 whereas the right term does. Indeed, the presence of H_1 here is only technical and by choosing for instance $H_1 = 4\beta + 1$, it gives that there exists a positive constant $\bar{\sigma}(\beta)$, continuous in β such that for all $\sigma \leq \bar{\sigma}(\beta)$,

$$\int_{\mathbb{R}} g_k(x) dx = 1 + O_\beta(\sigma^{2\beta}). \quad (32)$$

Proof. For $\delta \leq 1$ to be chosen further, let

$$A_{\sigma, \delta} := \{x \in \mathbb{R}, |l_j(x)| \leq \delta \mathfrak{B} \sigma^{-j} |\ln \sigma|^{-j/2}, \forall j \in 1 \dots r, L(x) \leq \delta \mathfrak{B} \sigma^{-\beta} |\ln \sigma|^{-\beta/2}\}$$

if $\beta > 1$ and let

$$A_{\sigma, \delta} := \{x \in \mathbb{R}, L(x) \leq \delta \mathfrak{B} \sigma^{-\beta} |\ln \sigma|^{-\beta/2}\}$$

otherwise. Note that for all $\delta \leq 1$, $A_{\sigma, \delta} \subset A_\sigma$. In the sequel we assume that $\beta > 1$, the proof being easily adapted for $\beta < 1$.

Proof of (28):

• Case $t = 0$: If $X \sim f$, then

$$\begin{aligned} \int_{A_\sigma^c} (K_\sigma^0 f)(x) dx &= \int_{A_\sigma^c} f(x) dx \\ &\leq \sum_{j=1}^r P(|l_j(X)| > (\delta \mathfrak{B}) \sigma^{-j} |\ln \sigma|^{-j/2}) + P(|L(X)| > (\delta \mathfrak{B}) \sigma^{-\beta} |\ln \sigma|^{-\beta/2}) \\ &\leq \sum_{j=1}^r P(|l_j(X)|^{\frac{2\beta+\varepsilon}{j}} > (\delta \mathfrak{B})^{\frac{2\beta+\varepsilon}{j}} \sigma^{-2\beta-\varepsilon} |\ln \sigma|^{-\frac{2\beta+\varepsilon}{2}}) \\ &\quad + P(|L(X)|^{\frac{2\beta+\varepsilon}{\beta}} > (\delta \mathfrak{B})^{\frac{2\beta+\varepsilon}{\beta}} \sigma^{-2\beta-\varepsilon} |\ln \sigma|^{-\frac{2\beta+\varepsilon}{2}}) \\ &\leq \sum_{j=1}^r P(|l_j(X)|^{\frac{2\beta+\varepsilon}{j}} > (\delta \mathfrak{B})^{\frac{2\beta+\varepsilon}{j}} \sigma^{-2\beta}) + P(|L(X)|^{\frac{2\beta+\varepsilon}{\beta}} > (\delta \mathfrak{B})^{\frac{2\beta+\varepsilon}{\beta}} \sigma^{-2\beta}) \end{aligned}$$

since $\sigma^{-\varepsilon} |\ln \sigma|^{-\frac{2\beta+\varepsilon}{2}} > 1$ for σ small enough (say $\sigma \leq \bar{\sigma}(\beta)$). Then, Markov Inequality together with (6) gives

$$P(|l_j(X)|^{\frac{2\beta+\varepsilon}{j}} > (\delta \mathfrak{B})^{\frac{2\beta+\varepsilon}{j}} \sigma^{-2\beta}) \leq (\delta \mathfrak{B})^{-\frac{2\beta+\varepsilon}{j}} \sigma^{2\beta} \mathbb{E}[|l_j(X)|^{\frac{2\beta+\varepsilon}{j}}] \leq c_\beta \sigma^{2\beta}$$

and

$$P(|L(X)|^{\frac{2\beta+\varepsilon}{\beta}} > (\delta \mathfrak{B})^{\frac{2\beta+\varepsilon}{\beta}} \sigma^{-2\beta}) \leq c_\beta \sigma^{2\beta}.$$

Finally,

$$\int_{A_\sigma^c} (K_\sigma^0 f)(x) dx \leq c_\beta \sigma^{2\beta}.$$

• Case $t = 1$: Let $X \sim f$ and $U \sim \psi$, then $X + \sigma U \sim K_\sigma f$. By applying Lemma 8 with $H = 2\beta$, let k' depending on β and $\bar{\sigma}(\beta)$ such that for all $\sigma \leq \bar{\sigma}(\beta)$, $k'\sigma |\ln \sigma|^{1/2} \leq \gamma$ and $P(|U| > k' |\ln \sigma|^{1/2}) \leq c_\beta \sigma^{2\beta}$. Then,

$$\begin{aligned} \int_{A_\sigma^c} K_\sigma f(x) dx &= P(X + \sigma U \in A_\sigma^c) \\ &= P(X + \sigma U \in A_\sigma^c \cap |U| \leq k' |\ln \sigma|^{1/2}) + P(X + \sigma U \in A_\sigma^c \cap |U| > k' |\ln \sigma|^{1/2}) \end{aligned} \quad (33)$$

and

$$\begin{aligned} \int_{A_\sigma^c} K_\sigma f(x) dx &\leq P(X + \sigma U \in A_\sigma^c \cap X \in A_{\sigma,\delta} \cap |U| \leq k' |\ln \sigma|^{1/2}) \\ &\quad + P(X + \sigma U \in A_\sigma^c \cap X \in A_{\sigma,\delta}^c \cap |U| \leq k' |\ln \sigma|^{1/2}) + P(|U| > k' |\ln \sigma|^{1/2}) \\ &\leq P(X + \sigma U \in A_\sigma^c \cap X \in A_{\sigma,\delta} \cap |U| \leq k' |\ln \sigma|^{1/2}) + P(X \in A_{\sigma,\delta}^c) + c_\beta \sigma^{2\beta}. \end{aligned} \quad (34)$$

The second term in (34) can be shown to be bounded by a multiple of $\sigma^{2\beta}$ in the same manner as for $t = 0$ for $\sigma \leq \bar{\sigma}(\beta)$. We now show that for σ small enough, the first term in (34) is zero for every function $f \in \mathcal{H}(\beta, \mathcal{P})$. On the one hand, according to (4) there exists $y \in [X, X + \sigma U]$ such that

$$l_j(X + \sigma U) = \sum_{u=0}^{r-j-1} \frac{l_{j+u}(X)}{u!} (\sigma U)^u + \frac{l_r(y)}{(r-j)!} (\sigma U)^{r-j}.$$

If $X \in A_{\sigma,\delta}$ and $|U| \leq k' |\ln \sigma|^{1/2}$ it yields

$$\begin{aligned} |l_j(X + \sigma U)| &\leq \sum_{u=0}^{r-j-1} \left| \frac{l_{j+u}(X)}{u!} \right| |\sigma U|^u + \left| \frac{l_r(y) - l_r(X)}{(r-j)!} \right| |\sigma U|^{r-j} + \left| \frac{l_r(X)}{(r-j)!} \right| |\sigma U|^{r-j} \\ &\leq \sum_{u=0}^{r-j} \left| \frac{l_{j+u}(X)}{u!} \right| |\sigma U|^u + \frac{r!}{(r-j)!} L(X) |y - X|^{\beta-r} |\sigma U|^{r-j} \\ &\leq \sum_{u=0}^{r-j} \left| \frac{l_{j+u}(X)}{u!} \right| |\sigma U|^u + \frac{r!}{(r-j)!} L(X) |\sigma U|^{\beta-j} \\ &\leq \sum_{u=0}^{r-j} \frac{1}{u!} \delta \mathfrak{B} \left(\sigma |\ln \sigma|^{1/2} \right)^{-(u+j)} \left(\sigma k' |\ln \sigma|^{1/2} \right)^u + \frac{r!}{(r-j)!} \delta \mathfrak{B} \left(\sigma |\ln \sigma|^{1/2} \right)^{-\beta} \left(\sigma k' |\ln \sigma|^{1/2} \right)^{\beta-j}. \end{aligned}$$

And thus for δ small enough, $|l_j(X + \sigma U)| \leq \mathfrak{B} \left(\sigma |\ln \sigma|^{1/2} \right)^{-j}$ for all $j \in \{1 \dots r\}$. Since $X + \sigma U \in A_\sigma^c$, this means that

$$L(X + \sigma U) > \mathfrak{B} \left(\sigma |\ln \sigma|^{1/2} \right)^{-\beta}. \quad (35)$$

On the other hand, let $\eta = \max |z_i|$ where the z_i 's are the roots of L . Suppose that $\deg(L) = q$, for $j = 1, \dots, q$, $|L^{(j)}(x)|/|L(x)| \rightarrow 0$ when $|x|$ tends to infinity. Consequently, since L does not vanish out of $[-\eta, \eta]$, there exists $c > 0$ only depending on L such that if $|x| > \eta + 1$, then $|L^{(j)}(x)| \leq c|L(x)|$.

If $|X| > \eta + 1$, then

$$\begin{aligned}
|L(X + \sigma U)| &\leq L(X) + \sum_{j=1}^q \left| \frac{L^{(j)}(X)}{j!} \right| |\sigma U|^j \\
&\leq \delta \mathfrak{B} \left(\sigma |\ln \sigma|^{1/2} \right)^{-\beta} + c |L(X)| \sum_{j=1}^q \frac{1}{j!} \left(\sigma k' |\ln \sigma|^{1/2} \right)^j \\
&\leq \delta \mathfrak{B} \left(\sigma |\ln \sigma|^{1/2} \right)^{-\beta} + c_\beta \delta \mathfrak{B} \left(\sigma |\ln \sigma|^{1/2} \right)^{-\beta+1} \\
&\leq 2\delta \mathfrak{B} \left(\sigma |\ln \sigma|^{1/2} \right)^{-\beta}
\end{aligned} \tag{36}$$

for $\sigma \leq \bar{\sigma}(\beta)$ where $\bar{\sigma}(\beta)$ can be chosen as a continuous function of β . It then leads to a contradiction with (35) for δ chosen small enough and thus $\mathbb{P}(X + \sigma U \in A_\sigma^c \cap X \in A_{\sigma,\delta} \cap |U| \leq k' |\ln \sigma|^{1/2} \cap |X| > \eta + 1) = 0$. Next, let $\bar{L} := \max_{j=0 \dots r} \sup_{|x| \leq \eta+1} |L^{(j)}(x)|$. If $|X| \leq \eta + 1$, (36) implies that

$$\begin{aligned}
|L(X + \sigma U)| &\leq \delta \mathfrak{B} \left(\sigma |\ln \sigma|^{1/2} \right)^{-\beta} + c \bar{L} \sum_{j=1}^q \frac{1}{j!} \left(\sigma k' |\ln \sigma|^{-1/2} \right)^j \\
&\leq \delta \mathfrak{B} \left(\sigma |\ln \sigma|^{1/2} \right)^{-\beta} + c_\beta \sigma |\ln \sigma|^{-1/2} \\
&\leq 2\delta \mathfrak{B} \left(\sigma |\ln \sigma|^{1/2} \right)^{-\beta}
\end{aligned}$$

for $\sigma \leq \bar{\sigma}(\beta)$ where $\bar{\sigma}(\beta)$ can be chosen as a continuous function of β , which also gives a contradiction with (35). Thus $\mathbb{P}(X + \sigma U \in A_\sigma^c \cap X \in A_{\sigma,\delta} \cap |U| \leq k' |\ln \sigma|^{1/2} \cap |X| \leq \eta + 1) = 0$. Finally, for $\sigma \leq \bar{\sigma}(\beta)$, $\mathbb{P}(X + \sigma U \in A_\sigma^c \cap X \in A_{\sigma,\delta} \cap |U| \leq k' |\ln \sigma|^{1/2}) = 0$ and (34) gives that

$$\int_{A_\sigma^c} K_\sigma f(x) dx = O_\beta(\sigma^{2\beta}).$$

- Case $t \geq 2$. The same method as before can be applied by assuming $X \sim K_\sigma^{t-1} f$ and $U \sim \psi$. Similarly, $\int_{A_\sigma^c} K_\sigma^t f(x) dx$ can be decomposed into three terms as in (34). Two of them are $O_\beta(\sigma^{2\beta})$ and the remaining term is zero for δ small enough.

Proof of (29):

- Case $t = 0$: According to Condition (7),

$$\begin{aligned}
\int_{E_\sigma^c} f(x) dx &\leq \sigma^{H_1/2} \int_{E_\sigma^c} \sqrt{f(x)} dx \\
&\leq \sigma^{H_1/2} \int_{\mathbb{R}} \sqrt{M \pi^{-\frac{1}{2}}} \exp(-x^2) dx \\
&\leq \sigma^{2\beta} M^{\frac{1}{2}} \pi^{\frac{1}{4}}
\end{aligned}$$

since $H_1 > 4\beta$.

- Case $t = 1$: We have

$$\int_{E_\sigma^c} K_\sigma f(x) dx = \int_{E_\sigma^c \cap A_\sigma} K_\sigma f(x) dx + \int_{E_\sigma^c \cap A_\sigma^c} K_\sigma f(x) dx$$

where the second integral is less than $\int_{A_\sigma^c} K_\sigma f(x) dx$ which is $O_\beta(\sigma^{2\beta})$ for $\sigma \leq \bar{\sigma}(\beta)$ according to (28). For $\delta \leq 1$ to be chosen further uniformly on $\mathcal{H}(\beta, \mathcal{P})$, consider the set $E_{\sigma, \delta} = \{x \in \mathbb{R}; f(x) \geq \sigma^{\delta H_1}\}$. Let $X \sim f$ and $U \sim \psi$. By applying Lemma 8 as before with $H = 2\beta$, let k' depending on β and $\bar{\sigma}(\beta)$ such that for all $\sigma \leq \bar{\sigma}(\beta)$, $k'\sigma |\ln \sigma|^{1/2} \leq \gamma$ and $P(|U| > k' |\ln \sigma|^{1/2}) \leq c_\beta \sigma^{2\beta}$. Then,

$$\begin{aligned} \int_{E_\sigma^c \cap A_\sigma} K_\sigma f(x) dx &= \mathbb{P}(X + \sigma U \in E_\sigma^c \cap A_\sigma) \\ &\leq \mathbb{P}(X + \sigma U \in E_\sigma^c \cap A_\sigma; |U| \leq k' |\ln \sigma|^{1/2}) + \mathbb{P}(|U| > k' |\ln \sigma|^{1/2}) \end{aligned} \quad (37)$$

$$\begin{aligned} \int_{E_\sigma^c \cap A_\sigma} K_\sigma f(x) dx &\leq \mathbb{P}(X + \sigma U \in E_\sigma^c \cap A_\sigma; |U| \leq k' |\ln \sigma|^{1/2}; X \in A_\sigma) + \mathbb{P}(X \in A_\sigma^c) \\ &\quad + \mathbb{P}(|U| > k' |\ln \sigma|^{1/2}) \\ &\leq \mathbb{P}(X + \sigma U \in E_\sigma^c; |U| \leq k' |\ln \sigma|^{1/2}; X \in A_\sigma \cap E_{\sigma, \delta}) \\ &\quad + \mathbb{P}(X \in E_{\sigma, \delta}^c) + \mathbb{P}(X \in A_\sigma^c) + c_\beta \sigma^{2\beta}. \end{aligned} \quad (38)$$

$$(39)$$

According to (28), the second term in (39) is $O_\beta(\sigma^{2\beta})$ for $\sigma \leq \bar{\sigma}(\beta)$. The first term in (39) can be bounded as previously for $t = 0$, leading to the condition

$$\delta H_1 \geq 4\beta$$

and thus we choose $\delta \in (0, 1)$ to satisfy this last condition. It remains to control the probability given in (38). On the one hand, since $X + \sigma U \in E_\sigma^c$ and $X \in E_{\sigma, \delta}$,

$$|\ln f(X + \sigma U) - \ln f(X)| > (1 - \delta) H_1 |\ln \sigma|. \quad (40)$$

On the other hand, since $X \in A_\sigma$ and $|U| \leq k' |\ln \sigma|^{1/2}$,

$$\begin{aligned} |\ln f(X + \sigma U) - \ln f(X)| &\leq \sum_{j=1}^r \left| \frac{l_j(X)}{j!} \sigma^j U^j \right| + L(X) |\sigma U|^\beta \\ &\leq \mathfrak{B} e^{k'} + \mathfrak{B} k'^\beta := d_1(\beta, k'). \end{aligned}$$

This is in contradiction with (40) for $\sigma \leq \exp\left(-\frac{d_1(\beta, k')}{(1-\delta)H_1}\right)$ and then (38) is zero for $\sigma \leq \bar{\sigma}(\beta, H_1)$ where $\bar{\sigma}(\beta, H_1)$ can be chosen continuous.

- Case $t \geq 2$. We follow the same proof as before:

$$\int_{E_\sigma^c} K_\sigma^t f(x) dx = \int_{E_\sigma^c \cap A_\sigma} K_\sigma^t f(x) dx + \int_{E_\sigma^c \cap A_\sigma^c} K_\sigma^t f(x) dx$$

where the second integral is less than $\int_{A_\sigma^c} K_\sigma^t f(x) dx$ which is $O_\beta(\sigma^{2\beta})$ according to (28). Let $X \sim f$ and $U_1, \dots, U_t \sim \psi$. By applying Lemma 8 as before with $H = 2\beta$, let k' depending on β and $\bar{\sigma}(\beta)$ such that for all $\sigma \leq \bar{\sigma}(\beta)$, $k'\sigma |\ln \sigma|^{1/2} \leq \gamma$ and $P(|U| > k' |\ln \sigma|^{1/2}) \leq c_\beta \sigma^{2\beta}$. Then,

$$\begin{aligned} \int_{E_\sigma^c \cap A_\sigma} K_\sigma^t f(x) dx &= \mathbb{P}(X + \sigma U_1 + \dots + \sigma U_t \in E_\sigma^c \cap A_\sigma) \\ &\leq \mathbb{P}\left(X + \sigma U_1 + \dots + \sigma U_t \in E_\sigma^c \cap A_\sigma; \forall j |U_j| \leq k' |\ln \sigma|^{1/2}\right) + \sum_{j=1}^t \mathbb{P}\left(|U_j| > k' |\ln \sigma|^{1/2}\right) \\ &\leq \mathbb{P}\left(X + \sigma U_1 + \dots + \sigma U_t \in E_\sigma^c \cap A_\sigma; \forall j |U_j| \leq k' |\ln \sigma|^{1/2}; X \in A_\sigma \cap E_{\sigma, \delta}\right) \\ &\quad + \mathbb{P}(X \in E_{\sigma, \delta}^c) + \mathbb{P}(X \in A_\sigma^c) + c_\beta. \end{aligned} \quad (41)$$

According to (28), the second term in (42) is $O_\beta(\sigma^{2\beta})$ for $\sigma \leq \bar{\sigma}(\beta)$, as well as the first term if $\delta H_1 \geq 4\beta$. We thus choose $\delta \in (0, 1)$ to satisfy this last condition. As before, we check that the probability given in (41) is 0. On the one hand, since $X + \sigma U_1 + \dots + \sigma U_t \in E_\sigma^c$ and $X \in E_{\sigma, \delta}$,

$$|\ln f(X + \sigma U_1 + \dots + \sigma U_t) - \ln f(X)| > (1 - \delta) H_1 |\ln \sigma|. \quad (43)$$

On the other hand, since $X \in A_\sigma$ and for all j , $|U_j| \leq k' |\ln \sigma|^{1/2}$, then

$$\begin{aligned} |\ln f(X + \sigma U_1 + \dots + \sigma U_t) - \ln f(X)| &\leq \sum_{j=1}^r \left| \frac{l_j(X)}{j!} \sigma^j \left(\sum_{i=1}^t U_i \right)^j \right| + L(X) \left| \sigma \sum_{i=1}^t U_i \right|^\beta \\ &\leq \sum_{j=1}^r \left| \frac{l_j(X)}{j!} \left(\sigma k' t |\ln \sigma|^{1/2} \right)^j \right| + L(X) \left(\sigma t k' |\ln \sigma|^{1/2} \right)^\beta \\ &\leq \mathfrak{B} e^{tk'} + \mathfrak{B} (tk')^\beta. \end{aligned}$$

This in contradiction with (43) for $\sigma \leq \bar{\sigma}(\beta, H_1)$ and finally (41) is zero.

Proof of $E_\sigma \cap A_\sigma \subset \mathbf{J}_{\sigma, k}$:

For $\beta \leq 2$, the inclusion is obvious since $f_0 = f$ and thus $J_{\sigma, k} = \mathbb{R}$. To prove the case $\beta > 2$, we show by induction on $u \in \mathbb{N}$, $1 \leq u \leq k$ that for every $h \in (0, 1)$, there exists a continuous function $\bar{\sigma}(\beta, H_1, h)$ such that for all $\sigma \leq \bar{\sigma}(\beta, H_1, h)$, for all $f \in \mathcal{H}(\beta, \mathcal{P})$ and all $x \in E_\sigma \cap A_\sigma$,

$$f_u(x) \geq \left(1 - \frac{h}{2}\right) f(x). \quad (44)$$

- Let $u = 1$ and $L^{(1)}$ defined by

$$L^{(1)}(x) = \sum_{j=1}^r \frac{|l_j(x)|}{j!} \left(k' \sigma |\ln \sigma|^{1/2} \right)^{j-1} + L(x) \left(k' \sigma |\ln \sigma|^{1/2} \right)^{\beta-1}.$$

Then, for all x and all $y \in D_x$, we have

$$\ln f(x) - L^{(1)}(x) |y - x| \leq \ln f(y) \leq \ln f(x) + L^{(1)}(x) |y - x|. \quad (45)$$

Note that for all x in A_σ , $L^{(1)}(x) \leq \mathfrak{B}e^{k'}\sigma^{-1}|\ln \sigma|^{-1/2}$. Starting from (45) and following the proof of Lemma 1 for the case $\beta = 1$, it yields for all x in A_σ and for all $H' > 0$ and all $\sigma < \bar{\sigma}(\beta, H')$

$$(K_\sigma f)(x) = f(x) \left[1 + L^{(1)}(x)\sigma O_\beta(1) \right] + O_{\beta, H'}(\sigma^{H'}). \quad (46)$$

For every $x \in E_\sigma \cap A_\sigma$, taking $H' = H_1 + 1$ in (46) it yields

$$\begin{aligned} \frac{f_1(x)}{f(x)} &= 2 - \frac{K_\sigma f(x)}{f(x)} \\ &= 1 - L^{(1)}(x)\sigma O_\beta(1) - \sigma \frac{\sigma^{H_1}}{f(x)} O_{\beta, H_1+1}(1). \end{aligned}$$

Next, $\left| \frac{\sigma^{H_1}}{f(x)} \right| \leq 1$ since $x \in E_\sigma$. Thus, for every $h \in (0, 1)$, there exists $\bar{\sigma}(\beta, H_1, h)$ such that for all $\sigma \leq \bar{\sigma}(\beta, H_1, h)$, for all $f \in \mathcal{H}(\beta, \mathcal{P})$ and every $x \in E_\sigma \cap A_\sigma$, $f_1(x) > (1 - h)f(x)$.

• The previous point is sufficient for $\beta \leq 4$ since $k = 1$ in this case. We now also suppose that $\beta > 4$ and thus that $k \geq 2$. Suppose that the integer $2 \leq u \leq k$ is such that (44) is true for the integer $u - 1$. Let $h \in (0, 1)$, there exists $\bar{\sigma}(\beta, H_1, h)$ such that for all $\sigma < \bar{\sigma}(\beta, H_1, h)$, for all $f \in \mathcal{H}(\beta, \mathcal{P})$ and every $x \in E_\sigma \cap A_\sigma$, $f_{u-1}(x) > (1 - \frac{h}{2})f(x)$. Note that since $2u \leq \beta$ we find that for all x and all $y \in D_x$,

$$\ln f(x) + \sum_{j=1}^{2u-1} \frac{l_j(x)}{j!} (y-x)^j - (y-x)^{2u} L^{(u)}(x) \leq \ln f(y) \leq \ln f(x) + \sum_{j=1}^{2u-1} \frac{l_j(x)}{j!} (y-x)^j + (y-x)^{2u} L^{(u)}(x)$$

with $L^{(u)}(x) = \sum_{j=2u}^r \frac{|l_j(x)|}{j!} (k'\sigma|\ln \sigma|^{1/2})^{j-2u} + L(x) (k'\sigma|\ln \sigma|^{1/2})^{\beta-2u}$. Thus for all x in A_σ , $|L^{(u)}(x)| \leq \mathfrak{B}e^{k'}\sigma^{-2u}|\ln \sigma|^{-u}$. Following the proof of Lemma 1 for the case $\beta = 2u$, it yields for all $x \in A_\sigma$ and for all $H' > 0$ and all $\sigma < \bar{\sigma}(\beta, H')$,

$$(K_\sigma f_{u-1})(x) = f(x) \left[1 + R^{(u)}(x)O_\beta(\sigma^{2u}) \right] + O_{\beta, H'}(\sigma^{H'}) \quad (47)$$

with $R^{(u)} = a_{2u+2}L^{(u)}(x) + \sum_{j=1}^{2u} a_j |l_j(x)| \frac{\sigma^{2u}}{j}$ and we have $\sup_{f \in \mathcal{H}(\beta, \mathcal{P})} \sup_{x \in E_\sigma \cap A_\sigma} |\sigma^{2u} R^{(u)}(x)| \leq \frac{c_\beta}{|\ln \sigma|^u}$. Then, using (47) with $H' = H_1 + 1$, it yields for all $x \in A_\sigma \cap E_\sigma$

$$\begin{aligned} \frac{f_u(x)}{f(x)} &= 1 - \frac{K_\sigma f_{u-1}(x) - f_{u-1}}{f(x)} \\ &= - \left[R^{(u)}(x)O_\beta(\sigma^{2u}) + \frac{\sigma^{H_1+1}}{f(x)} c_{\beta, H_1+1} \right] + \frac{f_{u-1}(x)}{f(x)}. \end{aligned}$$

There exists $\bar{\sigma}(\beta, H_1, h)$ such that for all $\sigma \leq \bar{\sigma}(\beta, H_1, h)$, $\left| R^{(u)}(x)O_\beta(\sigma^{2u}) + \frac{\sigma^{H_1+1}}{f(x)} c_{\beta, H_1+1} \right| \leq h/2$ and the induction is complete. By choosing $u = k$ and $h = 1/2$, it finally gives that $E_\sigma \cap A_\sigma \subset J_{\sigma, k}$ for $\sigma \leq \bar{\sigma}(\beta, H_1)$.

Proof of (30): We have

$$\begin{aligned} \int_{\mathbb{R}} g_k(x) dx &= \int_{J_{\sigma, k}} f_k(x) dx + \int_{J_{\sigma, k}^c} \frac{1}{2} f(x) dx \\ &= 1 + \int_{J_{\sigma, k}^c} \left[\frac{1}{2} f(x) - f_k(x) \right] dx \end{aligned}$$

since $\int_{\mathbb{R}} f_k(x) dx = 1$ (see Lemma 12). Moreover, f_k is a linear combination of $K_{\sigma}^t f$, $t = 0, \dots, k$, according to Lemma 12. Thus it yields $\int_{\mathbb{R}} g_k(x) dx = 1 + O_{\beta, H_1}(\sigma^{2\beta})$ for $\sigma \leq \bar{\sigma}(\beta, H_1)$ thanks to (28), (29) and that $J_{\sigma, k}^c \subset A_{\sigma}^c \cup E_{\sigma}^c$ for $\sigma \leq \bar{\sigma}(\beta, H_1)$.

Proof of (31): Let $H > 0$. According to Lemma 1 and (30), for all $\sigma \leq \bar{\sigma}(\beta, H, H_1)$, for all $x \in E_{\sigma} \cap A_{\sigma} \subset J_{\sigma, k}$, we have

$$\begin{aligned} |K_{\sigma} h_k(x) - f(x)| &\leq \left(\int_{\mathbb{R}} g_k(y) dy \right)^{-1} |K_{\sigma} f_k(x) - f(x)| + \left| \left(\int_{\mathbb{R}} g_k(y) dy \right)^{-1} - 1 \right| f(x) \\ &\quad + \left(\int_{\mathbb{R}} g_k(y) dy \right)^{-1} \int_{J_{\sigma, k}^c} \left\{ \frac{1}{2} f(u) - f_k(u) \right\} \psi_{\sigma}(x-u) du \\ &\leq c_{\beta} f(x) R_f(x) \sigma^{\beta} + c_{\beta} \sigma^H + \left(\int_{\mathbb{R}} g_k(y) dy \right)^{-1} (\diamond) \end{aligned}$$

where $(\diamond) = \int_{J_{\sigma, k}^c} \left\{ \frac{1}{2} f(u) - f_k(u) \right\} \psi_{\sigma}(x-u) du$. Let $D_x = \{u \in \mathbb{R}; |x-u| \leq k' \sigma |\ln \sigma|^{\frac{1}{2}}\}$ such that $k' \sigma |\ln \sigma|^{\frac{1}{2}} \leq \gamma$. According to the third result of Lemma 12 and Lemma 8,

$$0 \leq \int_{J_{\sigma, k}^c \cap D_x^c} \left\{ \frac{1}{2} f(u) - f_k(u) \right\} \psi_{\sigma}(x-u) du \leq 2^{k+2} \frac{M}{\sqrt{\pi}} \int_{D_x^c} \psi_{\sigma}(x-u) du \leq c_{\beta} \sigma^H.$$

Next, if $x \in A_{\sigma} \cap E_{\sigma}$, there exists $t > 1$ such that for all $u \in D_x$, $u \in A_{\sigma, t} \cap E_{\sigma, t}$. This result can be proved by adapting some parts of the proof of (28) and (29). Moreover, by changing \mathfrak{B} into $t\mathfrak{B}$, it can be shown that there exists $\bar{\sigma}(\beta)$ such that for all $\sigma < \bar{\sigma}(\beta)$, $A_{\sigma, t} \cap E_{\sigma, t} \subset J_{\sigma, k}$. Thus for σ small enough, $\int_{J_{\sigma, k}^c \cap D_x^c} \left\{ \frac{1}{2} f(u) - f_k(u) \right\} \psi_{\sigma}(x-u) du = 0$. Finally, $(\diamond) \leq c_{\beta} \sigma^H$ and $|K_{\sigma} h_k(x) - f(x)| \leq c_{\beta} f(x) R_f(x) \sigma^{\beta} + c'_{\beta} \sigma^H$. \square

Proposition 2.

Let $\beta > 0$ and $k \in \mathbb{N}$ such that $\beta \in (2k, 2k+2]$. There exists a positive constant $\bar{\sigma}(\beta)$ such that for all $f \in \mathcal{H}(\beta, \mathcal{P})$ and all $\sigma < \bar{\sigma}(\beta)$,

$$KL(f, K_{\sigma} h_k) = \int_{\mathbb{R}} f(x) \ln \left(\frac{f(x)}{K_{\sigma} h_k(x)} \right) dx = O_{\beta}(\sigma^{2\beta})$$

where h_k is defined by (27) and where $\bar{\sigma}(\beta)$ can be chosen as a continuous function of β .

Proof. Preliminary, we remark that if p and q are two densities and S is a set, then

$$\int_S p \ln \left(\frac{p}{q} \right) \leq \int_S p \frac{p-q}{q} = \int_S \frac{(p-q)^2}{q} + \frac{q(p-q)}{q} = \int_S \frac{(p-q)^2}{q} + \int_{S^c} (q-p)$$

since $\int_S p = 1 - \int_{S^c} p$, $\int_S q = 1 - \int_{S^c} q$ and $\int_S (p-q) = \int_{S^c} (q-p)$. We use this inequality with the densities f and $K_{\sigma} h_k$, and the sets A_{σ} and E_{σ} , where E_{σ} is defined with $H_1 = 4\beta + 1$, to obtain

the following control of $\text{KL}(f, K_\sigma h_k)$:

$$\begin{aligned} \int_{\mathbb{R}} f(x) \ln \left(\frac{f(x)}{K_\sigma h_k(x)} \right) dx &= \int_{A_\sigma \cap E_\sigma} f(x) \ln \left(\frac{f(x)}{K_\sigma h_k(x)} \right) dx + \int_{A_\sigma^c \cup E_\sigma^c} f(x) \ln \left(\frac{f(x)}{K_\sigma h_k(x)} \right) dx \\ &\leq \int_{A_\sigma \cap E_\sigma} \frac{[f(x) - K_\sigma h_k(x)]^2}{K_\sigma h_k(x)} dx \end{aligned} \quad (48)$$

$$+ \int_{A_\sigma^c \cup E_\sigma^c} [K_\sigma h_k(x) - f(x)] dx \quad (49)$$

$$+ \int_{A_\sigma^c \cup E_\sigma^c} f(x) \ln \left(\frac{f(x)}{K_\sigma h_k(x)} \right) dx. \quad (50)$$

• **Control of (48):**

Let $H > 0$. According to Lemma 2 with $H_1 = 4\beta + 1$, there exists $\bar{\sigma}(\beta, H) > 0$ such that for all $x \in A_\sigma \cap E_\sigma$ and for all $\sigma < \bar{\sigma}(\beta, H)$, $[K_\sigma h_k(x) - f(x)]^2 \leq [\Lambda_{\beta, H} f(x) R_f(x) \sigma^\beta + \Omega_{\beta, H} \sigma^H]^2$ where $\Lambda_{\beta, H}$ and $\Omega_{\beta, H}$ are two constants. Moreover, according to Lemma 13, there exists $\bar{\sigma}(\beta) > 0$ such that for all $\sigma < \bar{\sigma}(\beta)$,

$$K_\sigma h_k(x) \geq \frac{D}{1 + A_\beta \sigma^{2\beta}} f(x)$$

with $D = \frac{\xi\sqrt{\pi}}{6M}$. Thus for all $\sigma < \bar{\sigma}(\beta, H) \wedge \bar{\sigma}(\beta)$,

$$\begin{aligned} \frac{[f(x) - K_\sigma h_k(x)]^2}{K_\sigma h_k(x)} &\leq \frac{\Lambda_{\beta, H}^2}{D} (1 + A_\beta \sigma^{2\beta}) \sigma^{2\beta} R_f(x)^2 f(x) + \frac{\Omega_{\beta, H}^2}{D} (1 + A_\beta \sigma^{2\beta}) \sigma^{2H} \frac{1}{f(x)} \\ &+ \frac{2\Lambda_{\beta, H} \Omega_{\beta, H}}{D} (1 + A_\beta \sigma^{2\beta}) \sigma^{\beta+H} R_f(x). \end{aligned}$$

Then,

$$\begin{aligned} \int_{A_\sigma \cap E_\sigma} \frac{[f(x) - K_\sigma h_k(x)]^2}{K_\sigma h_k(x)} dx &\leq \frac{\Lambda_{\beta, H}^2}{D} (1 + A_\beta \sigma^{2\beta}) \sigma^{2\beta} \int_{A_\sigma \cap E_\sigma} R_f(x)^2 f(x) dx \\ &+ \frac{\Omega_{\beta, H}^2}{D} (1 + A_\beta \sigma^{2\beta}) \sigma^{2H-2(4\beta+1)} \int_{A_\sigma \cap E_\sigma} f(x) dx \\ &+ \frac{2\Lambda_{\beta, H} \Omega_{\beta, H}}{D} (1 + A_\beta \sigma^{2\beta}) \sigma^{\beta+H-4\beta-1} \int_{A_\sigma \cap E_\sigma} R_f(x) f(x) dx. \end{aligned} \quad (51)$$

Thus the two integrals $\int_{A_\sigma \cap E_\sigma} R_f(x)^2 f(x) dx$ and $\int_{A_\sigma \cap E_\sigma} R_f(x) f(x) dx$ have to be controlled.

The first integral can be decomposed into

$$\begin{aligned}
\int_{A_\sigma \cap E_\sigma} R_f(x)^2 f(x) dx &= \int_{A_\sigma \cap E_\sigma} \left[a_{r+1} L(x) + \sum_{j=1}^r a_j |l_j(x)|^{\frac{\beta}{j}} \right]^2 f(x) dx \\
&= a_{r+1}^2 \int_{A_\sigma \cap E_\sigma} L(x)^2 f(x) dx + \sum_{j=1}^r a_j^2 \int_{A_\sigma \cap E_\sigma} |l_j(x)|^{\frac{2\beta}{j}} f(x) dx \\
&\quad + 2 \sum_{j=1}^r a_{r+1} a_j \int_{A_\sigma \cap E_\sigma} |l_j(x)|^{\frac{\beta}{j}} L(x) f(x) dx \\
&\quad + \sum_{\substack{j, j'=1 \\ j \neq j'}}^r a_j a_{j'} \int_{A_\sigma \cap E_\sigma} |l_j(x)|^{\frac{\beta}{j}} |l_{j'}(x)|^{\frac{\beta}{j'}} f(x) dx.
\end{aligned}$$

Using the Hölder inequality and Condition (6), for all $j = 1, \dots, r$,

$$\int_{A_\sigma \cap E_\sigma} |l_j(x)|^{\frac{2\beta}{j}} f(x) dx \leq \left[\int_{\mathbb{R}} |l_j(x)|^{\frac{2\beta+\varepsilon}{j}} f(x) dx \right]^{\frac{2\beta}{2\beta+\varepsilon}} \left[\int_{\mathbb{R}} f(x) dx \right]^{\frac{\varepsilon}{2\beta+\varepsilon}} \leq C^{\frac{2\beta}{2\beta+\varepsilon}} \quad (52)$$

and $\int_{A_\sigma \cap E_\sigma} L(x)^2 f(x) dx \leq \left[\int_{\mathbb{R}} L(x)^{2+\frac{\varepsilon}{\beta}} f(x) dx \right]^{\frac{2\beta}{2\beta+\varepsilon}} \left[\int_{\mathbb{R}} f(x) dx \right]^{\frac{\varepsilon}{2\beta+\varepsilon}} \leq C^{\frac{2\beta}{2\beta+\varepsilon}}$. Next, using the Cauchy-Schwarz inequality and (52), for all $j, j' \in \{1, \dots, r\}, j \neq j'$,

$$\int_{A_\sigma \cap E_\sigma} |l_j(x)|^{\frac{\beta}{j}} |l_{j'}(x)|^{\frac{\beta}{j'}} f(x) dx \leq \left[\int_{\mathbb{R}} |l_j(x)|^{\frac{2\beta}{j}} f(x) dx \right]^{\frac{1}{2}} \left[\int_{\mathbb{R}} |l_{j'}(x)|^{\frac{2\beta}{j'}} f(x) dx \right]^{\frac{1}{2}} \leq C^{\frac{2\beta}{2\beta+\varepsilon}}$$

and for all $j \in \{1, \dots, r\}$,

$$\int_{A_\sigma \cap E_\sigma} |l_j(x)|^{\frac{\beta}{j}} L(x) f(x) dx \leq \left[\int_{\mathbb{R}} |l_j(x)|^{\frac{2\beta}{j}} f(x) dx \right]^{\frac{1}{2}} \left[\int_{\mathbb{R}} L(x)^2 f(x) dx \right]^{\frac{1}{2}} \leq C^{\frac{2\beta}{2\beta+\varepsilon}}.$$

Finally, $\int_{A_\sigma \cap E_\sigma} R_f(x)^2 f(x) dx \leq \left(\sum_{j=1}^{r+1} a_j \right)^2 C^{\frac{2\beta}{2\beta+\varepsilon}}$.

For the second integral,

$$\begin{aligned}
\int_{A_\sigma \cap E_\sigma} R_f(x) f(x) dx &= \int_{A_\sigma \cap E_\sigma} \left[a_{r+1} L(x) + \sum_{j=1}^r a_j |l_j(x)|^{\frac{\beta}{j}} \right] f(x) dx \\
&\leq a_{r+1} \sqrt{\int_{\mathbb{R}} L(x)^2 f(x) dx} \sqrt{\int_{\mathbb{R}} f(x) dx} + \sum_{j=1}^r a_j \sqrt{\int_{\mathbb{R}} |l_j(x)|^{\frac{2\beta}{j}} f(x) dx} \sqrt{\int_{\mathbb{R}} f(x) dx} \\
&\leq \sum_{j=1}^{r+1} a_j C^{\frac{\beta}{2\beta+\varepsilon}}.
\end{aligned}$$

Finally, (51) becomes

$$\begin{aligned} \int_{A_\sigma \cap E_\sigma} \frac{[f(x) - K_\sigma h_k(x)]^2}{K_\sigma h_k(x)} dx &\leq \frac{\Lambda_{\beta,H}^2}{D} (1 + A_\beta \sigma^{2\beta}) \sigma^{2\beta} \left(\sum_{j=1}^{r+1} a_j \right)^2 C^{\frac{2\beta}{2\beta+\varepsilon}} \\ &+ \frac{\Omega_{\beta,H}^2}{D} (1 + A_\beta \sigma^{2\beta}) \sigma^{2H-8\beta-2} \\ &+ \frac{2\Lambda_{\beta,H} \Omega_{\beta,H}}{D} (1 + A_\beta \sigma^{2\beta}) \sigma^{H-3\beta-1} \left(\sum_{j=1}^{r+1} a_j \right) C^{\frac{\beta}{2\beta+\varepsilon}}. \end{aligned}$$

By taking $H = 5\beta + 1$, it gives that there exists $\bar{\sigma}(\beta) > 0$ such that for all $\sigma < \bar{\sigma}(\beta)$,

$$\int_{A_\sigma \cap E_\sigma} \frac{[f(x) - K_\sigma h_k(x)]^2}{K_\sigma h_k(x)} dx = O_\beta(\sigma^{2\beta}).$$

• **Control of (49):**

According to Lemma 12,

$$\begin{aligned} h_k(x) &= \left(\int_{\mathbb{R}} g_k(x) dx \right)^{-1} \left\{ f_k(x) \mathbb{1}_{J_{\sigma,k}}(x) + \frac{1}{2} f(x) \mathbb{1}_{J_{\sigma,k}^c}(x) \right\} \\ &\leq \left\{ 2 \sum_{i=0}^k (-1)^i \binom{k+1}{i+1} K_\sigma^i f(x) \right\} \mathbb{1}_{J_{\sigma,k}}(x) + f(x) \mathbb{1}_{J_{\sigma,k}^c}(x) \end{aligned}$$

thus

$$K_\sigma h_k(x) \leq 2 \sum_{j=1}^{k+1} \binom{k+1}{j} K_\sigma^j f(x) + K_\sigma f(x).$$

According to (28) and (29) in Lemma 2 with $H_1 = 4\beta + 1$, there exists $\bar{\sigma}(\beta) > 0$ such that for all $\sigma < \bar{\sigma}(\beta)$,

$$\begin{aligned} \int_{A_\sigma^c \cup E_\sigma^c} [K_\sigma h_k(x) - f(x)] dx &\leq \int_{A_\sigma^c \cup E_\sigma^c} K_\sigma h_k(x) dx + \int_{A_\sigma^c \cup E_\sigma^c} f(x) dx \\ &\leq 2 \sum_{j=1}^{k+1} \binom{k+1}{j} \int_{A_\sigma^c \cup E_\sigma^c} K_\sigma^j f(x) dx + \int_{A_\sigma^c \cup E_\sigma^c} K_\sigma f(x) dx + \int_{A_\sigma^c \cup E_\sigma^c} f(x) dx \\ &\leq 2 \sum_{j=2}^{k+1} \binom{k+1}{j} \int_{A_\sigma^c} K_\sigma^j f(x) dx + [2(k+1) + 1] \int_{A_\sigma^c} K_\sigma f(x) dx + \int_{A_\sigma^c} K_\sigma^0 f(x) dx \\ &\quad + 2 \sum_{j=2}^{k+1} \binom{k+1}{j} \int_{E_\sigma^c} K_\sigma^j f(x) dx + [2(k+1) + 1] \int_{E_\sigma^c} K_\sigma f(x) dx + \int_{E_\sigma^c} K_\sigma^0 f(x) dx \\ &\leq 2 \left[2 \sum_{j=2}^{k+1} \binom{k+1}{j} + 2(k+2) \right] c_\beta \sigma^{2\beta}. \end{aligned}$$

• **Control of (50):**

According to Lemma 13, for all $\sigma < \bar{\sigma}(\beta)$, $K_\sigma h_k(x) \geq \frac{D}{1+A_\beta\sigma^{2\beta}} f(x)$ then

$$\begin{aligned} \int_{A_\sigma^c \cup E_\sigma^c} f(x) \ln \left(\frac{f(x)}{K_\sigma h_k(x)} \right) dx &\leq \ln \left(\frac{1+A_\beta\sigma^{2\beta}}{D} \right) \int_{A_\sigma^c \cup E_\sigma^c} f(x) dx \\ &\leq \ln \left(\frac{1+A_\beta\sigma^{2\beta}}{D} \right) \left\{ \int_{A_\sigma^c} K_\sigma^0 f(x) dx + \int_{E_\sigma^c} K_\sigma^0 f(x) dx \right\} \\ &\leq \ln \left(\frac{1+A_\beta\sigma^{2\beta}}{D} \right) 2c_\beta \sigma^{2\beta}. \end{aligned}$$

In conclusion, there exists $\bar{\sigma}(\beta) > 0$ such that for all $\sigma < \bar{\sigma}(\beta)$, $\text{KL}(f, K_\sigma h_k) = O_\beta(\sigma^{2\beta})$. □

3.2. Proof of Theorem 2

Proof. For the definition of E_σ , we choose $H_1 = 4(\beta + 1)$. Let \tilde{h}_k be the restriction of h_k on an interval $[-\mu_\sigma, \mu_\sigma]$, normalized in order to have a density function:

$$\tilde{h}_k : x \in \mathbb{R} \mapsto \left(\int_{[-\mu_\sigma, \mu_\sigma]} h_k(y) dy \right)^{-1} h_k(x) \mathbb{1}_{[-\mu_\sigma, \mu_\sigma]}(x)$$

where μ_σ depends on σ and will be chosen further such that

$$\mu_\sigma \geq \sigma. \tag{53}$$

Let $\varepsilon \in (0, \pi^{-1/2})$. According to Proposition 3 in Appendix A.2, there exists a discrete distribution \tilde{F} on $[-\mu_\sigma, \mu_\sigma]$ with at most $54\mu_\sigma\sigma^{-1}e^2 [-\ln(\sqrt{\pi\varepsilon}) \vee 1]$ support points such that

$$\|\tilde{h}_k * \psi_\sigma - \tilde{F} * \psi_\sigma\|_\infty \leq \frac{2\varepsilon}{\sigma}. \tag{54}$$

Denoting $\tilde{\varphi}(x)dx = \left(\int_{[-\mu_\sigma, \mu_\sigma]} h_k(x) dx \right) \tilde{F} * \psi_\sigma(x)$, it gives for all $x \in \mathbb{R}$,

$$\begin{aligned} |K_\sigma h_k(x) - \tilde{\varphi}(x)| &= \left(\int_{[-\mu_\sigma, \mu_\sigma]} h_k(x) dx \right) \left| \frac{h_k}{\int_{[-\mu_\sigma, \mu_\sigma]} h_k(y) dy} * \psi_\sigma(x) - \tilde{F} * \psi_\sigma(x) \right| \\ &\leq \left| \frac{h_k \mathbb{1}_{[-\mu_\sigma, \mu_\sigma]}}{\int_{[-\mu_\sigma, \mu_\sigma]} h_k(y) dy} * \psi_\sigma(x) - \tilde{F} * \psi_\sigma(x) \right| + (h_k \mathbb{1}_{[-\mu_\sigma, \mu_\sigma]^c}) * \psi_\sigma(x). \end{aligned}$$

By applying Lemma 14 with $p = \frac{1}{2}$, it gives that for all $\sigma \leq 1 - 2^{-1/k}$ and for all $x \in \mathbb{R}$,

$$h_k(x) \leq 4M \left(\frac{4}{\sqrt{3}} \right)^k \psi \left(\frac{x}{2} \right)$$

and thus $(h_k \mathbb{1}_{[-\mu_\sigma, \mu_\sigma]^c}) * \psi_\sigma(x) \leq 4M \left(\frac{4}{\sqrt{3}}\right)^k \psi\left(\frac{\mu_\sigma}{2}\right)$. Now, we choose $\mu_\sigma := 2\sqrt{\ln\left(\frac{4M}{\sqrt{\pi}} \left(\frac{4}{\sqrt{3}}\right)^k \frac{\sigma}{\varepsilon}\right)}$ in order to obtain that $\|(h_k \mathbb{1}_{[-\mu_\sigma, \mu_\sigma]^c}) * \psi_\sigma\|_\infty \leq \frac{\varepsilon}{\sigma}$. This last inequality together with (54) yields

$$\|K_\sigma h_k - \tilde{\varphi}\|_\infty \leq \frac{3\varepsilon}{\sigma}.$$

We also define the function $t := \tilde{\varphi} + \sigma^{6\beta+5}\psi_\sigma$ and the finite Gaussian mixture with density

$$\varphi(x) := \frac{t(x)}{\int_{\mathbb{R}} t(y) dy} = \frac{\tilde{\varphi}(x) + \sigma^{6\beta+5}\psi_\sigma(x)}{\int_{[-\mu_\sigma, \mu_\sigma]} h_k(y) dy + \sigma^{6\beta+5}}.$$

Then we want to upper bound

$$\begin{aligned} \text{KL}(f, \varphi) &= \int_{\mathbb{R}} f(x) \ln\left(\frac{f(x)}{\varphi(x)}\right) dx \\ &= \int_{\mathbb{R}} f(x) \ln\left(\frac{f(x)}{K_\sigma h_k(x)}\right) dx + \int_{\mathbb{R}} f(x) \ln\left(\frac{K_\sigma h_k(x)}{t(x)}\right) dx + \int_{\mathbb{R}} f(x) \ln\left(\frac{t(x)}{\varphi(x)}\right) dx \\ &= \int_{\mathbb{R}} f(x) \ln\left(\frac{f(x)}{K_\sigma h_k(x)}\right) dx \\ &\quad + \int_{E_\sigma^c} f(x) \ln\left(\frac{K_\sigma h_k(x)}{t(x)}\right) dx + \int_{E_\sigma} f(x) \ln\left(\frac{K_\sigma h_k(x)}{t(x)}\right) dx + \int_{\mathbb{R}} f(x) \ln\left(\frac{t(x)}{\varphi(x)}\right) dx \\ &= \boxed{\text{I1}} + \boxed{\text{I2}} + \boxed{\text{I3}} + \boxed{\text{I4}}. \end{aligned}$$

- **Control of $\boxed{\text{I1}}$:** According to Proposition 2, for all $\sigma < \bar{\sigma}(\beta)$,

$$\int_{\mathbb{R}} f(x) \ln\left(\frac{f(x)}{K_\sigma h_k(x)}\right) dx = O_\beta(\sigma^{2\beta}).$$

- **Control of $\boxed{\text{I2}}$:** According to Lemma 14, $K_\sigma h_k(x) \leq 4M \left(\frac{4}{\sqrt{3}}\right)^k$ for σ small enough and since $s(x) \geq \sigma^{6\beta+5}\psi_\sigma(x)$,

$$\begin{aligned} \boxed{\text{I2}} &\leq \int_{E_\sigma^c} f(x) \ln\left(\frac{4M \left(\frac{4}{\sqrt{3}}\right)^k}{\sigma^{6\beta+5}\psi_\sigma(x)}\right) dx \\ &\leq \left(\int_{E_\sigma^c} f(x) dx\right) \left[(6\beta+4)|\ln \sigma| + \ln\left(4M \left(\frac{4}{\sqrt{3}}\right)^k\right) \right] + \int_{E_\sigma^c} f(x) \frac{x^2}{\sigma^2} dx. \end{aligned}$$

For the second integral,

$$\begin{aligned} \int_{E_\sigma^c} \frac{x^2}{\sigma^2} f(x) dx &\leq \sigma^{\frac{H_1}{2}-2} \int_{E_\sigma^c} x^2 \sqrt{f(x)} dx \\ &\leq \sigma^{2\beta} \int_{\mathbb{R}} x^2 \sqrt{M\psi(x)} dx \\ &\leq \sigma^{2\beta} \sqrt{2M} \pi^{-\frac{1}{4}} 2\nu_{1,2} = 4\pi\sqrt{M}\sigma^{2\beta}. \end{aligned}$$

Similarly, $\int_{E_\sigma^c} f(x)dx \leq \sigma^{2\beta+2}\sqrt{2M}$ and finally

$$\boxed{\text{I2}} \leq \left\{ \ln \left(4M \left(\frac{4}{\sqrt{3}} \right)^k \right) + (6\beta + 4)|\ln \sigma| \right\} \sqrt{2M}\sigma^{2\beta+2} + 4\pi\sqrt{M}\sigma^{2\beta}.$$

Thus $\boxed{\text{I2}} = O_\beta(\sigma^{2\beta})$.

• **Control of $\boxed{\text{I3}}$:** On the one hand,

$$\begin{aligned} |K_\sigma h_k(x) - t(x)| &\leq |K_\sigma h_k(x) - \tilde{\varphi}(x)| + |\tilde{\varphi}(x) - t(x)| \\ &\leq 3\varepsilon\sigma^{-1} + \sigma^{6\beta+5}\psi_\sigma(x) \\ &\leq 3\varepsilon\sigma^{-1} + \sigma^{6\beta+4}\pi^{-1/2}. \end{aligned}$$

On the other hand, according to Lemma 13, for all $x \in \mathbb{R}$ and for all $\sigma < \bar{\sigma}(\beta)$, $K_\sigma h_k(x) \geq \frac{\xi\sqrt{\pi}}{6M(1+A_\beta\sigma^{2\beta})}f(x)$. Since $x \in E_\sigma$ then $K_\sigma h_k(x) \geq \frac{\xi\sqrt{\pi}}{6M(1+A_\beta\sigma^{2\beta})}\sigma^{4(\beta+1)}$. Thus, $t(x) \geq \tilde{\varphi}(x) \geq K_\sigma h_k(x) - 3\varepsilon\sigma^{-1} \geq \frac{\sigma^{4(\beta+1)}\sqrt{\pi}}{6M(1+A_\beta\sigma^{2\beta})} - 3\varepsilon\sigma^{-1}$. Finally,

$$\begin{aligned} \boxed{\text{I3}} &\leq \int_{E_\sigma} f(x) \frac{K_\sigma h_k(x) - t(x)}{t(x)} dx \\ &\leq \frac{3\varepsilon\sigma^{-1} + \sigma^{6\beta+4}\pi^{-1/2}}{\frac{\sigma^{4(\beta+1)}}{2(1+A_\beta\sigma^{2\beta})} - 3\varepsilon\sigma^{-1}} \int_{E_\sigma} f(x) dx \\ &\leq \frac{3\varepsilon\sigma^{-1} + \sigma^{6\beta+4}\pi^{-1/2}}{\frac{\sigma^{4(\beta+1)}}{2(1+A_\beta\sigma^{2\beta})} - 3\varepsilon\sigma^{-1}}. \end{aligned}$$

Let $\delta' := 1 + \frac{\beta}{2(\beta+1)}$ and we set $\varepsilon := \sigma^{\delta'4(\beta+1)+1}$. It yields

$$\boxed{\text{I3}} \leq \frac{(\pi^{-1/2} + 3)\sigma^{6\beta+4}}{\frac{\sigma^{4(\beta+1)}}{2(1+A_\beta\sigma^{2\beta})} - 3\sigma^{6\beta+4}} = O_\beta(\sigma^{2\beta}).$$

• **Control of $\boxed{\text{I4}}$:** Note that $\frac{t(x)}{\varphi(x)} = \int_{[-\mu_\sigma, \mu_\sigma]} h_k(y)dy + \sigma^{6\beta+5} \leq 1 + \sigma^{6\beta+5}$ and thus

$$\begin{aligned} \boxed{\text{I4}} &\leq \int_{\mathbb{R}} f(x) \ln(1 + \sigma^{6\beta+5}) dx \\ &\leq \sigma^{6\beta+5} \leq \sigma^{2\beta}. \end{aligned}$$

Finally, we obtain that $\text{KL}(f, \varphi) = O_\beta(\sigma^{2\beta})$. Moreover, according to the choice of ε , we have that

$$\begin{aligned} \mu_\sigma &= 2\sqrt{\ln \left(\frac{4M}{\sqrt{\pi}} \left(\frac{4}{\sqrt{3}} \right)^k \frac{\sigma}{\varepsilon} \right)} \\ &= 2\sqrt{\ln \left(\frac{4M}{\sqrt{\pi}} \left(\frac{4}{\sqrt{3}} \right)^k \sigma^{-(6\beta+4)} \right)} \\ &= \tilde{G}_\beta |\ln \sigma|^{\frac{1}{2}} \end{aligned}$$

where

$$\tilde{G}_\beta = 2\sqrt{\ln\left(\frac{4M}{\sqrt{\pi}}\right) + k \ln\left(\frac{4}{\sqrt{3}}\right) + (6\beta + 4)}. \quad (55)$$

Thus there exists $\bar{\sigma}(\beta)$ continuous in β such that (53) is fulfilled for $\sigma < \bar{\sigma}(\beta)$. Furthermore, the mixture \wp has k_σ components such that

$$\begin{aligned} k_\sigma &\leq 54\mu_\sigma\sigma^{-1}e^2 \left[1 \vee \ln\left(\frac{1}{\sqrt{\pi}\varepsilon}\right) \right] + 1 \\ &\leq \tilde{G}_\beta |\ln \sigma|^{\frac{1}{2}} 54\sigma^{-1}e^2 \left[1 \vee \ln\left(\frac{1}{\sqrt{\pi}\sigma^{6\beta+5}}\right) \right] + 1 \\ &= G_\beta \sigma^{-1} |\ln \sigma|^{\frac{3}{2}}. \end{aligned} \quad (56)$$

□

4. Proof of the lower bound

4.1. Proof of Proposition 1

Note that for every j , φ_j is supported by

$$J_j := \left[-\frac{\alpha}{2} + \frac{\alpha}{D}(j-1) + \frac{\alpha}{4D}, -\frac{\alpha}{2} + \frac{\alpha}{D}j - \frac{\alpha}{4D} \right] \subsetneq I_j = \left[-\frac{\alpha}{2} + \frac{\alpha}{D}(j-1), -\frac{\alpha}{2} + \frac{\alpha}{D}j \right]$$

and thus the supports of the φ_j , $1 \leq j \leq D$ are disjoint. We also note that for all $x \in [-\frac{\alpha}{2}, \frac{\alpha}{2}]^c$, $f_\theta(x) = \omega(x)$ and for all $x \in [-\frac{\alpha}{2}, \frac{\alpha}{2}]$, there exists a unique $j \in \{1, \dots, D\}$ such that $f_\theta(x) = 2\xi + (2\theta_j - 1)\varphi_j(x)$ where $\varphi_j(x) = 0$ if $x \in I_j \setminus J_j$. The proof of Proposition 1 is decomposed into two lemmas.

Lemma 3. *Density function and monotonicity conditions.*

For all $D \in \mathbb{N}^*$ and all $\theta \in \{0, 1\}^D$, the function f_θ defined by (9) is a positive density function such that for all $x \in [-\frac{\alpha}{2}, \frac{\alpha}{2}]$, $f_\theta(x) \in [\xi, 3\xi]$. This function fulfills also the following monotonicity conditions:

1. $\forall x \in [-\alpha, \alpha]$, $f_\theta(x) \geq \xi$ and $\forall x \in [-\alpha, \alpha]^c$, $f_\theta(x) \leq \xi$.
2. f_θ is nondecreasing on $(-\infty, -\alpha)$ and nonincreasing on (α, ∞) .
3. $\forall x \in \mathbb{R}$, $f_\theta(x) \leq M\psi(x)$ with $M = \bar{M} \vee 3\sqrt{\pi}\xi \exp(\alpha^2/4)$.

Proof. For all $x \in [-\frac{\alpha}{2}, \frac{\alpha}{2}]^c$, $f_\theta(x) = \omega(x) > 0$ since ω is positive. Moreover, for all $x \in [-\frac{\alpha}{2}, \frac{\alpha}{2}]$, $\exists! j \in \{1, \dots, D\}$ such that $x \in I_j$. Then,

$$f_\theta(x) = \omega(x) + (2\theta_j - 1)\varphi_j(x) = 2\xi + (2\theta_j - 1)\varphi_j(x).$$

Thus

$$\begin{aligned} |f_\theta(x) - 2\xi| &= |(2\theta_j - 1)| |\varphi_j(x)| \\ &= \left| \frac{\xi D^{-\beta}}{A} \varphi\left(\frac{D}{\alpha}\left(x + \frac{\alpha}{2}\right) - (j-1)\right) \right| \\ &\leq \xi D^{-\beta} \\ &\leq \xi \end{aligned}$$

since $D^{-\beta} \leq 1$. Thus for all $x \in [-\frac{\alpha}{2}, \frac{\alpha}{2}]$, $f_\theta(x) \in [\xi, 3\xi]$. Finally, f_θ is a positive function on \mathbb{R} . Moreover,

$$\begin{aligned} \int_{\mathbb{R}} f_\theta(x) dx &= \int_{\mathbb{R}} \omega(x) dx + \sum_{j=1}^D (2\theta_j - 1) \int_{I_j} \varphi_j(x) dx \\ &= \int_{\mathbb{R}} \omega(x) dx + \sum_{j=1}^D (2\theta_j - 1) \frac{\xi D^{-\beta} \alpha}{A} \frac{1}{D} \int_{\mathbb{R}} \varphi(y) dy \\ &= 1 \end{aligned}$$

because $\int_{\mathbb{R}} \omega(x) dx = 1$ and $\int_{\mathbb{R}} \varphi(y) dy = 0$. Thus, f_θ is a density function.

On $(-\infty, -\alpha)$, since $f_\theta(x) = \omega(x)$ and ω is a nondecreasing function on $(-\infty, -\alpha)$, the function f_θ is a nondecreasing function on $(-\infty, -\alpha)$. Moreover,

$$\forall x < -\alpha, f_\theta(x) \leq f_\theta(-\alpha) = \omega(-\alpha) = \xi.$$

In the same way, the function f_θ is a nonincreasing function on (α, ∞) and

$$\forall x > \alpha, f_\theta(x) \leq f_\theta(\alpha) = \omega(\alpha) = \xi.$$

For all $x \in [-\alpha, \alpha]$,

- if $x \in [-\alpha, -\frac{\alpha}{2}]$, $f_\theta(x) = \omega(x) \geq \omega(-\alpha) = \xi$ because ω non-decreases and $\omega(-\alpha) = \xi$;
- if $x \in (\frac{\alpha}{2}, \alpha]$, $f_\theta(x) = \omega(x) \geq \omega(\alpha) = \xi$ because ω non-increases and $\omega(\alpha) = \xi$;
- if $x \in [-\frac{\alpha}{2}, \frac{\alpha}{2}]$, $f_\theta(x) \in [\xi, 3\xi]$ thus $f_\theta(x) \geq \xi$.

For the last point, we have that for all $x \in [-\frac{\alpha}{2}, \frac{\alpha}{2}]^c$, $f_\theta(x) = \omega(x) \leq \tilde{M}\psi(x)$. Moreover, for all $x \in [-\frac{\alpha}{2}, \frac{\alpha}{2}]$, $f_\theta(x) \leq 3\xi \leq 3\xi\sqrt{\pi} \exp(\alpha^2/4)\psi(x)$. Finally, for all $x \in \mathbb{R}$, $f_\theta(x) \leq M(\xi, \alpha, \tilde{M})\psi(x)$ with $M(\xi, \alpha, \tilde{M}) := \tilde{M} \vee 3\sqrt{\pi}\xi \exp(\alpha^2/4)$. \square

Lemma 4. Let $\beta \in [\underline{\beta}, \bar{\beta}]$. For all $\theta \in \{0, 1\}^D$, the function $\ln f_\theta$ is locally β -Hölder: for all x, y such that $|x - y| \leq \frac{\alpha}{4}$,

$$|(\ln f_\theta)^{(r)}(x) - (\ln f_\theta)^{(r)}(y)| \leq L(\underline{\beta}, \bar{\beta}, \tilde{L}, \alpha) r! |x - y|^{\beta-r}$$

where $L(\underline{\beta}, \bar{\beta}, \tilde{L}, \alpha)$ does not depend on D . Moreover, there exists a constant $C(\underline{\beta}, \bar{\beta}, \tilde{C}, \alpha)$, which can be taken identical for every D , such that for any integer $j = 1, \dots, r$ and for all $D \in \mathbb{N}^*$,

$$\int_{\mathbb{R}} |(\ln f_\theta)^{(j)}(x)|^{\frac{2\beta+\varepsilon}{j}} f_\theta(x) dx \leq C(\underline{\beta}, \bar{\beta}, \tilde{C}, \alpha),$$

and

$$\int_{\mathbb{R}} |L(\underline{\beta}, \bar{\beta}, \tilde{L}, \alpha)|^{\frac{2\beta+\varepsilon}{\beta}} f_\theta(x) dx \leq C(\underline{\beta}, \bar{\beta}, \tilde{C}, \alpha).$$

If D is a positive even integer, for any integer $j = 0, \dots, r$, $|(\ln f_\theta)^{(j)}(0)| \leq \ln(2\xi)$.

Proof. Let $j \in \{1, \dots, D\}$ and $1 \leq t \leq r + 1$. We start by upper bounding $\sup_{x \in I_j} |(\ln f_\theta)^{(t)}(x)|$. According to Lemma 17, for all $x \in I_j$,

$$(\ln f_\theta)^{(t)}(x) = f_\theta(x)^{-2^{t-1}} \sum_{(\eta_0, \dots, \eta_t) \in \Xi_t} \rho(\eta_0, \dots, \eta_t) \prod_{u=0}^t \left(f_\theta^{(u)}(x) \right)^{\eta_u}$$

with

$$\Xi_t = \left\{ (\eta_0, \dots, \eta_t) \in \mathbb{N}^{t+1}; \sum_{u=0}^t u \eta_u = t, \sum_{u=0}^t \eta_u = 2^{t-1} \right\}.$$

For all $u \in \{1, \dots, t\}$,

$$|f_\theta^{(u)}| \leq \frac{\xi D^{-\beta}}{A} \left(\frac{D}{\alpha} \right)^u \|\varphi^{(u)}\|_\infty \leq \frac{\xi D^{u-\beta}}{\alpha^u}.$$

Then, for all $(\eta_0, \dots, \eta_t) \in \Xi_t$,

$$\begin{aligned} \left| \prod_{u=0}^t (f_\theta^{(u)})^{\eta_u} \right| &\leq D^{\sum_{u=1}^t u \eta_u - \beta \sum_{u=1}^t \eta_u} \xi^{\sum_{u=1}^t \eta_u} \alpha^{-\sum_{u=1}^t u \eta_u} \times |f_\theta|^{\eta_0} \\ &\leq \xi^{2^{t-1} - \eta_0} D^{t - \beta(2^{t-1} - \eta_0)} \alpha^{-t} \times |f_\theta(x)|^{\eta_0} \end{aligned}$$

since $\sum_{u=1}^t u \eta_u = t$ and $\sum_{u=1}^t \eta_u = 2^{t-1} - \eta_0$. Since $f_\theta(x) \in [\xi, 3\xi]$ and $2^{t-1} - \eta_0 \geq 1$,

$$\begin{aligned} |(\ln f_\theta)^{(t)}(x)| &\leq \sum_{(\eta_0, \dots, \eta_t) \in \Xi_t} |\rho(\eta_0, \dots, \eta_t)| \left| \prod_{u=1}^t (f_\theta^{(u)}(x))^{\eta_u} \right| |f_\theta(x)|^{\eta_0 - 2^{t-1}} \\ &\leq \sum_{(\eta_0, \dots, \eta_t) \in \Xi_t} |\rho(\eta_0, \dots, \eta_t)| \xi^{2^{t-1} - \eta_0} D^{t - \beta(2^{t-1} - \eta_0)} \alpha^{-t} \xi^{\eta_0 - 2^{t-1}} \\ &\leq \sum_{(\eta_0, \dots, \eta_t) \in \Xi_t} |\rho(\eta_0, \dots, \eta_t)| D^{t - \beta(2^{t-1} - \eta_0)} \alpha^{-t}. \end{aligned}$$

Denoting $\mathcal{B}(t) := \text{card}(\Xi_t)$ and $B(t) := \max_{(\eta_0, \dots, \eta_t) \in \Xi_t} |\rho(\eta_0, \dots, \eta_t)|$, it leads to

$$\sup_{x \in I_j} |(\ln f_\theta)^{(t)}(x)| \leq \mathcal{B}(t) B(t) D^{t-\beta} \alpha^{-t}. \quad (57)$$

We now use this preliminary result to prove that $\ln f_\theta$ is locally β -Holder. Let $(x, y) \in \mathbb{R}^2$ such that $|x - y| \leq \frac{\alpha}{4}$.

- If $x, y \in [-\frac{\alpha}{2}, \frac{\alpha}{2}]^c$,

$$\begin{aligned} |(\ln f_\theta)^{(r)}(x) - (\ln f_\theta)^{(r)}(y)| &= |(\ln \omega)^{(r)}(x) - (\ln \omega)^{(r)}(y)| \\ &\leq \tilde{L} r! |x - y|^{\beta-r}. \end{aligned}$$

since $\ln \omega$ is locally β -Holder with $\gamma_\omega = \frac{\alpha}{4}$ and a constant \tilde{L} .

- If $y \in [-\frac{\alpha}{2}, \frac{\alpha}{2}]^c$ and $x \in I_j$:

– If $|x - y| < \frac{\alpha}{4D}$ then $x \in I_j \setminus J_j$. Thus, $\ln f_\theta(x) = \ln \omega(x)$ and

$$\begin{aligned} |(\ln f_\theta)^{(r)}(x) - (\ln f_\theta)^{(r)}(y)| &= |(\ln \omega)^{(r)}(x) - (\ln \omega)^{(r)}(y)| \\ &\leq \tilde{L} r! |x - y|^{\beta-r}. \end{aligned}$$

– If $\frac{\alpha}{4D} \leq |x-y| < \frac{\alpha}{4}$, $\ln \omega(y) = \ln(2\xi)$ since $x \in [-3\alpha/4, -\alpha/2] \cup [\alpha/2, 3\alpha/4]$ thus if $r \geq 1$,

$$\begin{aligned} |(\ln f_\theta)^{(r)}(x) - (\ln f_\theta)^{(r)}(y)| &\leq \|(\ln f_\theta)^{(r)}\|_{\infty, [-\alpha/2, \alpha/2]} + \|(\ln \omega)^{(r)}\|_{\infty, [\alpha/2, 3\alpha/4]} \\ &\leq \mathcal{B}(r)B(r)D^{r-\beta}\alpha^{-r} \left(\frac{4D}{\alpha}\right)^{\beta-r} |x-y|^{\beta-r} + 0 \\ &\leq \frac{\mathcal{B}(r)B(r)}{r!} 4^{\beta-r}\alpha^{-\beta} r! |x-y|^{\beta-r} \end{aligned}$$

and if $r = 0$,

$$\begin{aligned} |(\ln f_\theta)(x) - (\ln f_\theta)(y)| &\leq |\ln(2\xi) - \ln(2\xi + (2\theta_j - 1)\varphi_j(y))| \\ &\leq |-\ln(1 + (2\xi)^{-1}(2\theta_j - 1)\varphi_j(y))| \\ &\leq |(2\xi)^{-1}(2\theta_j - 1)\varphi_j(y)| \\ &\leq (2\xi)^{-1}\xi D^{-\beta}(4D)^\beta \alpha^{-\beta} |x-y|^\beta \\ &\leq 4^\beta \alpha^{-\beta} |x-y|^\beta = \frac{\mathcal{B}(1)B(1)}{0!} 4^\beta \alpha^{-\beta} 0! |x-y|^{\beta-r}. \end{aligned}$$

• For all $x, y \in [-\alpha/2, \alpha/2]$, $\exists!(j, j') \in \{1, \dots, D\}^2$ such that $x \in I_j$ and $y \in I_{j'}$.

– If $|x-y| \leq \frac{\alpha}{4D}$,

* if $j' \neq j$, $x \in I_j \setminus J_j$ and $y \in I_{j'} \setminus J_{j'}$, thus

$$|(\ln f_\theta)^{(r)}(x) - (\ln f_\theta)^{(r)}(y)| = 0.$$

* if $j' = j$,

$$\begin{aligned} |(\ln f_\theta)^{(r)}(x) - (\ln f_\theta)^{(r)}(y)| &\leq |x-y|^{\beta-r} |x-y|^{r+1-\beta} \|\ln f_\theta^{(r+1)}\|_{\infty, [-\alpha/2, \alpha/2]} \\ &\leq \alpha^{-\beta+r+1} (4D)^{\beta-r-1} \frac{\mathcal{B}(r+1)B(r+1)}{r!} \frac{D^{r+1-\beta}}{\alpha^{r+1}} r! |x-y|^{\beta-r} \\ &\leq \frac{\mathcal{B}(r+1)B(r+1)}{r!} 4^{\beta-r-1} \alpha^{-\beta} r! |x-y|^{\beta-r} \end{aligned}$$

– If $\frac{\alpha}{4D} < |x-y| < \frac{\alpha}{4}$: if $r = 0$,

$$\begin{aligned} |(\ln f_\theta)(x) - (\ln f_\theta)(y)| &= \left| \ln \left(\frac{1 + (2\xi)^{-1}(2\theta_j - 1)\varphi_j(x)}{1 + (2\xi)^{-1}(2\theta_j - 1)\varphi_j(y)} \right) \right| \\ &\leq \left| \frac{(2\xi)^{-1}(2\theta_j - 1)[\varphi_j(x) - \varphi_j(y)]}{1 + (2\xi)^{-1}(2\theta_j - 1)\varphi_j(y)} \right| \\ &\leq \frac{2\|\varphi_j\|_\infty}{\xi} \\ &\leq 2D^{-\beta}(4D)^\beta \alpha^{-\beta} |x-y|^\beta \\ &\leq 24^\beta \alpha^{-\beta} |x-y|^\beta = 24^\beta \alpha^{-\beta} \frac{\mathcal{B}(1)B(1)}{0!} 0! |x-y|^\beta \end{aligned}$$

and if $r \geq 1$

$$\begin{aligned} |(\ln f_\theta)^{(r)}(x) - (\ln f_\theta)^{(r)}(y)| &\leq 2\|(\ln f_\theta)^{(r)}\|_{\infty, [-\alpha/2, \alpha/2]} \\ &\leq 2\mathcal{B}(r)B(r)\frac{D^{r-\beta}}{\alpha^r}\left(\frac{4D}{\alpha}\right)^{\beta-r}|x-y|^{\beta-r} \\ &\leq 2\frac{\mathcal{B}(r)B(r)}{r!}4^\beta\alpha^{-\beta}r!|x-y|^{\beta-r}. \end{aligned}$$

Finally, for all $\beta \in [\underline{\beta}, \bar{\beta}]$, for all $(x, y) \in \mathbb{R}^2$ such that $|x - y| < \frac{\alpha}{4}$,

$$|(\ln f_\theta)^{(r)}(x) - (\ln f_\theta)^{(r)}(y)| \leq L(\underline{\beta}, \bar{\beta}, \alpha)r!|x - y|^{\beta-r}$$

with

$$L(\underline{\beta}, \bar{\beta}, \tilde{L}, \alpha) := \tilde{L} \vee \max_{\beta \in [\underline{\beta}, \bar{\beta}]} \left(2 \frac{\mathcal{B}(\lceil \beta \rceil)B(\lceil \beta \rceil)}{\lfloor \beta \rfloor!} \left(\frac{4}{\alpha} \right)^\beta \right).$$

According to (57), for any integer $j \in \{1, \dots, r\}$, $\|(\ln f_\theta)^{(j)}\|_{\infty, [-\alpha/2, \alpha/2]} \leq B(j)\mathcal{B}(j)\alpha^{-j}$ thus it yields

$$\begin{aligned} \int_{\mathbb{R}} |(\ln f_\theta)^{(j)}(x)|^{\frac{2\beta+\varepsilon}{j}} f_\theta(x) dx &\leq \int_{[-\alpha/2, \alpha/2]^c} |(\ln \omega)^{(j)}(x)|^{\frac{2\beta+\varepsilon}{j}} \omega(x) dx + [B(j)\mathcal{B}(j)\alpha^{-j}]^{\frac{2\beta+\varepsilon}{j}} \int_{[-\alpha/2, \alpha/2]} f_\theta(x) dx \\ &\leq \tilde{C} + [B(j)\mathcal{B}(j)\alpha^{-j}]^{\frac{2\beta+\varepsilon}{j}}. \end{aligned}$$

Thus there exists a constant $C(\underline{\beta}, \bar{\beta}, \tilde{C}, \tilde{\varepsilon}, \alpha)$ such that for any integer $j \in \{1, \dots, r\}$,

$$\int_{\mathbb{R}} |(\ln f_\theta)^{(j)}(x)|^{\frac{2\beta+\varepsilon}{j}} f_\theta(x) dx \leq \tilde{C} + \max_{1 \leq j \leq r+1} [B(j)\mathcal{B}(j)]^{\frac{2\beta+\varepsilon}{j}} \leq C(\underline{\beta}, \bar{\beta}, \tilde{C}, \tilde{\varepsilon}, \alpha)$$

and

$$\int_{\mathbb{R}} |L(\underline{\beta}, \bar{\beta}, \tilde{L}, \alpha)|^{2+\frac{\varepsilon}{\beta}} f_\theta(x) dx = |L(\underline{\beta}, \bar{\beta}, \tilde{L}, \alpha)|^{2+\frac{\varepsilon}{\beta}} \leq C(\underline{\beta}, \bar{\beta}, \tilde{C}, \tilde{\varepsilon}, \alpha).$$

The last point assumes that D is even, thus $0 \in I_{D/2} \setminus J_{D/2}$. Then, $\ln f_\theta$ is equal to $\ln(2\xi)$ in a neighborhood of 0 and for all $j \in \{1, \dots, r\}$, $|(\ln f_\theta)^{(j)}(0)| = 0$. \square

Lemmas 3 and 4 show that for any positive even integer D and for all $\beta \in [\underline{\beta}, \bar{\beta}]$, $\mathcal{J}(\beta, D) \subset \mathcal{H}(\beta, \mathcal{P}(\underline{\beta}, \bar{\beta}))$ where

$$\mathcal{P}(\underline{\beta}, \bar{\beta}) = \left\{ \frac{\alpha}{4}, \ln(2\xi), L(\underline{\beta}, \bar{\beta}, \tilde{L}, \alpha), \tilde{\varepsilon}, C(\underline{\beta}, \bar{\beta}, \tilde{C}, \tilde{\varepsilon}, \alpha), \alpha, \xi, M(\xi, \alpha, \tilde{M}) \right\}.$$

4.2. Proof of Theorem 3

Lemma 5. Let $\theta, \theta' \in \{0, 1\}^D$. The Hellinger distance between two functions f_θ and $f_{\theta'}$ of $\mathcal{J}(\beta, D)$ fulfills

$$1. \quad d_H^2(f_\theta, f_{\theta'}) \leq \frac{\xi\alpha}{8A^2} D^{-2\beta},$$

2. $\forall \theta \neq \theta'$, $d_H^2(f_\theta, f_{\theta'}) \geq \xi \alpha (2A)^{-2} \delta(\theta, \theta') D^{-(2\beta+1)}$ where $\delta(\theta, \theta') = \sum_{j=1}^D \mathbf{1}_{\theta_j \neq \theta'_j}$ is the Hamming distance between θ and θ' .

Proof.

The Hellinger distance between f_θ and $f_{\theta'}$ can be decomposed as follows:

$$\begin{aligned} d_H^2(f_\theta, f_{\theta'}) &= \frac{1}{2} \int_{[-\alpha/2, \alpha/2]} \left[\sqrt{f_\theta(x)} - \sqrt{f_{\theta'}(x)} \right]^2 dx + \frac{1}{2} \int_{[-\alpha/2, \alpha/2]^c} \left[\sqrt{\omega(x)} - \sqrt{\omega(x)} \right]^2 dx \\ &= \frac{1}{2} \sum_{j=1}^D \int_{I_j} \left[\sqrt{2\xi + (2\theta_j - 1)\varphi_j(x)} - \sqrt{2\xi + (2\theta'_j - 1)\varphi_j(x)} \right]^2 dx. \end{aligned}$$

Since the quantity under the brackets is equal to zero if $\theta_j = \theta'_j$, it gives

$$\begin{aligned} d_H^2(f_\theta, f_{\theta'}) &= \frac{1}{2} \sum_{j=1}^D \int_{I_j} \left[\sqrt{2\xi + \varphi_j(x)} - \sqrt{2\xi - \varphi_j(x)} \right]^2 dx \mathbf{1}_{\theta_j \neq \theta'_j} \\ &= \frac{1}{2} \sum_{j=1}^D \int_{I_j} \left[4\xi - 2\sqrt{(2\xi)^2 - \varphi_j(x)^2} \right] dx \mathbf{1}_{\theta_j \neq \theta'_j}. \end{aligned}$$

Note that $\left(\frac{\varphi_j(x)}{2\xi}\right)^2 \leq 1$ for all $x \in I_j$ and $\|\varphi_j\|_\infty = \frac{\xi D^{-\beta}}{A} \|\varphi\|_\infty \leq \xi$. Then,

$$\sqrt{(2\xi)^2 - \varphi_j(x)^2} = 2\xi \sqrt{1 - \left(\frac{\varphi_j(x)}{2\xi}\right)^2} \geq \frac{1}{4} \left[1 - \left(\frac{\varphi_j(x)}{2\xi}\right)^2 \right]$$

since $\sqrt{1-y} \geq 1-y$ for all $y \in [0, 1]$. Thus,

$$\begin{aligned} \int_{I_j} \left[4\xi - 2\sqrt{(2\xi)^2 - \varphi_j(x)^2} \right] dx &\leq \int_{I_j} \left[4\xi - 4\xi + \frac{\varphi_j^2(x)}{4\xi} \right] dx \\ &\leq (4\xi)^{-1} \int_{I_j} \left[\left(\frac{\xi D^{-\beta}}{A}\right)^2 \varphi^2 \left(\frac{D}{\alpha}(x+1) - (j-1)\right) \right] dx \\ &\leq (4\xi)^{-1} \left(\frac{\xi D^{-\beta}}{A}\right)^2 \frac{\alpha}{D} \end{aligned}$$

since $\int_{\mathbb{R}} \varphi^2(y) dy = 1$. Finally,

$$\begin{aligned} d_H^2(f_\theta, f_{\theta'}) &\leq (4\xi)^{-1} \left(\frac{\xi D^{-\beta}}{A}\right)^2 \frac{\alpha}{D} \frac{1}{2} \delta(\theta, \theta') \\ &\leq \frac{\xi \alpha}{8A^2} D^{-2\beta} \end{aligned}$$

since $\delta(\theta, \theta') \leq D$.

For the lower bound, we have

$$\sqrt{(2\xi)^2 - \varphi_j(x)^2} = 2\xi \sqrt{1 - \left(\frac{\varphi_j(x)}{2\xi}\right)^2} \leq 2\xi \left[1 - \frac{1}{2} \left(\frac{\varphi_j(x)}{2\xi}\right)^2\right]$$

since $\sqrt{1-y} \leq 1 - \frac{1}{2}y$ for all $y \in [0, 1]$. Thus,

$$\begin{aligned} \int_{I_j} \left[4\xi - 2\sqrt{(2\xi)^2 - \varphi_j(x)^2}\right] dx &\geq \int_{I_j} \left[4\xi - 4\xi + \frac{\varphi_j^2(x)}{2\xi}\right] dx \\ &\geq (2\xi)^{-1} \left(\frac{\xi D^{-\beta}}{A}\right)^2 \frac{\alpha}{D} \int_{\mathbb{R}} \varphi^2(y) dy \\ &\geq (2\xi)^{-1} \left(\frac{\xi D^{-\beta}}{A}\right)^2 \frac{\alpha}{D} \end{aligned}$$

and finally

$$\begin{aligned} d_{H}^2(f_{\theta}, f_{\theta'}) &\geq (2\xi)^{-1} \left(\xi \frac{D^{-\beta}}{A}\right)^2 \frac{\alpha}{D} \frac{1}{2} \sum_{j=1}^D \mathbb{1}_{\theta_j \neq \theta'_j} \\ &\geq \xi \alpha (2A)^{-2} D^{-(2\beta+1)} \delta(\theta, \theta'). \end{aligned}$$

□

Lemma 6. Let $\theta, \theta' \in \{0, 1\}^D$. The Kullback-Leibler divergence between two functions f_{θ} and $f_{\theta'}$ of $\mathcal{J}(\beta, D)$ fulfills

$$\text{KL}(f_{\theta}, f_{\theta'}) \leq \frac{5\xi\alpha}{4A^2} D^{-2\beta}.$$

Proof. The Kullback-Leibler divergence between f_{θ} and $f_{\theta'}$ is given by

$$\begin{aligned} \text{KL}(f_{\theta}, f_{\theta'}) &= \int_{\mathbb{R}} f_{\theta}(x) \ln \left(\frac{f_{\theta}(x)}{f_{\theta'}(x)}\right) dx \\ &= \int_{[-\alpha/2, \alpha/2]} f_{\theta}(x) \ln \left(\frac{f_{\theta}(x)}{f_{\theta'}(x)}\right) dx + \int_{[-\alpha/2, \alpha/2]^c} \omega(x) \ln \left(\frac{\omega(x)}{\omega(x)}\right) dx \\ &= \int_{[-\alpha/2, \alpha/2]} f_{\theta}(x) \ln \left(\frac{f_{\theta}(x)}{f_{\theta'}(x)}\right) dx. \end{aligned}$$

Then for all $x \in [-\alpha/2, \alpha/2]$ and for all $\theta \in \{0, 1\}^D$, $f_{\theta}(x) \in [\xi, 3\xi]$ according to Lemma 3 thus $\left\|\frac{f_{\theta}}{f_{\theta'}}\right\|_{\infty, [-1, 1]} \leq 3$. According to Lemma 7.23 in Massart (2007),

$$\text{KL}(f_{\theta}, f_{\theta'}) \leq 2 \left[2 + \ln \left(\left\|\frac{f_{\theta}}{f_{\theta'}}\right\|_{\infty}\right)\right] d_H^2(f_{\theta}, f_{\theta'}).$$

Lemma 3 gives that for all $x \in [-\alpha/2, \alpha/2]$, $f_\theta(x) \in [\xi, 3\xi]$ and furthermore, $f_\theta = f_{\theta'}$ on $[-\alpha/2, \alpha/2]^c$. Thus,

$$\begin{aligned} \text{KL}(f_\theta, f_{\theta'}) &\leq 2 \left[2 + \ln \left(\sup_{[-\alpha/2, \alpha/2]} \left| \frac{f_\theta(x)}{f_{\theta'}(x)} \right| \right) \right] d_H^2(f_\theta, f_{\theta'}) \\ &\leq 10 d_H^2(f_\theta, f_{\theta'}) \\ &\leq \frac{5\xi\alpha}{4A^2} D^{-2\beta} \end{aligned}$$

according to Lemma 5. □

Proof of Theorem 3. The proof consists of applying Corollary 1 given in Appendix B with the space $\mathcal{J}(\beta, D)$, the Hellinger distance d_H , $p = 2$ and the finite subset $\mathcal{C} = \{f_\theta, \theta \in \Theta\}$ where Θ is the subset of $\{0, 1\}^D$ provided by Lemma 16. Then, it has to be checked that

$$n \max_{\theta, \theta' \in \Theta} \text{KL}(f_\theta, f_{\theta'}) \leq \kappa \ln |\Theta|.$$

According to Lemma 16, $\ln |\Theta| > \frac{D}{8}$ and $\kappa \geq \frac{1}{2}$. Moreover, $\text{KL}(f_\theta, f_{\theta'}) \leq \frac{5\xi\alpha}{4A^2} D^{-2\beta}$ and thus D is chosen such that

$$n \frac{5\xi\alpha}{4A^2} D^{-2\beta} \leq \frac{D}{16} \Leftrightarrow \frac{20\xi\alpha n}{A^2} \leq D^{2\beta+1}.$$

Since $3\xi\alpha \leq 1$ then $20\xi\alpha n A^{-2} \leq \frac{20}{3}n \leq 7n$ and we finally choose $D = \min\{2k; k \in \mathbb{N}^*, (2k)^{2\beta+1} \geq 7n\}$. It gives that for any estimator \tilde{s} ,

$$\begin{aligned} \sup_{\theta \in \Theta} \mathbb{E}_s[d_H^2(f_\theta, \tilde{s})] &\geq 2^{-2}(1 - \kappa) \left[\min_{\theta, \theta' \in \Theta, \theta \neq \theta'} d_H(f_\theta, f_{\theta'}) \right]^2 \\ &\geq 2^{-2}(1 - \kappa) \xi \alpha (2A)^{-2} D^{-(2\beta+1)} \min_{\theta, \theta' \in \Theta, \theta \neq \theta'} \delta(\theta, \theta') \\ &\geq 2^{-2}(1 - \kappa) \xi \alpha (2A)^{-2} D^{-(2\beta+1)} \frac{D}{4} \\ &\geq \frac{(1 - \kappa) \xi \alpha}{A^2} 2^{-6-2\beta} (7n)^{-\frac{2\beta}{2\beta+1}} \end{aligned}$$

according to Lemma 16. □

5. Proof of Theorem 4

Under the hypotheses of Section 2.4, let $\mathcal{P}(\underline{\beta}, \bar{\beta})$ be the parameter set given in Proposition 1. In order to prove Theorem 4, we start with the following lemma that makes the connection between the models \mathcal{S}_m and the approximation result given in Theorem 2.

Lemma 7. *There exists a positive constant $c_{\underline{\beta}, \bar{\beta}}$ such that for all $\beta \in [\underline{\beta}, \bar{\beta}]$ and for all $s \in \mathcal{H}(\beta, \mathcal{P}(\underline{\beta}, \bar{\beta}))$,*

$$\text{KL}(s, \mathcal{S}_m) \leq c_{\underline{\beta}, \bar{\beta}} \underline{\lambda}(m)^\beta.$$

Proof. According to Theorem 2, the level $\bar{\sigma}(\beta)$ under which the approximation (8) is valid is a continuous function of β . Thus we can define the positive constant $\bar{\sigma}(\underline{\beta}, \bar{\beta}) := \inf_{\beta \in [\underline{\beta}, \bar{\beta}]} \bar{\sigma}(\beta)$. Next, let

$$m_0(\underline{\beta}, \bar{\beta}) := \inf \left\{ m \geq 2; \sqrt{\underline{\lambda}(m)} < \bar{\sigma}(\underline{\beta}, \bar{\beta}) \right\}$$

and consider $m \geq m_0(\underline{\beta}, \bar{\beta})$. Then Theorem 2 can be applied for $\sigma = \sqrt{\underline{\lambda}(m)}$: for all $\beta \in [\underline{\beta}, \bar{\beta}]$ and for all $s \in \mathcal{H}(\beta, \mathcal{P}(\underline{\beta}, \bar{\beta}))$, there exists a mixture φ with less than $G_{\bar{\beta}} \underline{\lambda}(m)^{-\frac{1}{2}} \left| \ln \sqrt{\underline{\lambda}(m)} \right|^{\frac{3}{2}}$ components, with means belonging to $[-\bar{\mu}(m), \bar{\mu}(m)]$ and with the same variance $\underline{\lambda}(m)$ for each component such that

$$\text{KL}(s, \varphi) \leq c_{\beta} \underline{\lambda}(m)^{\beta}.$$

Since $G_{\bar{\beta}}$ is a non decreasing function of β , the number of components is less than

$$\begin{aligned} G_{\bar{\beta}} \left(\sqrt{\underline{\lambda}(m)} \right)^{-1} \left| \ln \sqrt{\underline{\lambda}(m)} \right|^{\frac{3}{2}} &\leq G_{\bar{\beta}} \left[\frac{a_{\bar{\beta}}}{m} (\ln m)^{\frac{3}{2}} \right]^{-1} \left| \ln \left\{ \frac{a_{\bar{\beta}}}{m} (\ln m)^{\frac{3}{2}} \right\} \right|^{\frac{3}{2}} \\ &\leq m \frac{G_{\bar{\beta}}}{a_{\bar{\beta}}} \left[\frac{\ln a_{\bar{\beta}}}{\ln m} + 1 + \frac{3}{2} \frac{|\ln \ln m|}{\ln m} \right]^{\frac{3}{2}} \\ &\leq m \end{aligned}$$

according to the definition of $\sqrt{\underline{\lambda}(m)}$ and Condition (10). This shows that $\varphi \in \mathcal{S}_m$ and thus $\text{KL}(s, \mathcal{S}_m) \leq c_{\beta} \underline{\lambda}(m)^{\beta}$ for all $m \geq m_0(\underline{\beta}, \bar{\beta})$. Since c_{β} is continuous in β , there exists $c_{\underline{\beta}, \bar{\beta}} > 0$ such that for all $\beta \in [\underline{\beta}, \bar{\beta}]$, for all $s \in \mathcal{H}(\beta, \mathcal{P}(\underline{\beta}, \bar{\beta}))$, and for all $m \geq m_0(\underline{\beta}, \bar{\beta})$,

$$\text{KL}(s, \mathcal{S}_m) \leq c_{\underline{\beta}, \bar{\beta}} [\underline{\lambda}(m)]^{\beta}. \quad (58)$$

It remains to show the same result for $m \leq m_0(\underline{\beta}, \bar{\beta})$: let t_m be a mixture of \mathcal{S}_m , for all $\beta \in [\underline{\beta}, \bar{\beta}]$ and for all $s \in \mathcal{H}(\beta, \mathcal{P}(\underline{\beta}, \bar{\beta}))$,

$$\begin{aligned} \text{KL}(s, \mathcal{S}_m) &\leq \text{KL}(s, t_m) \\ &\leq \int M\psi(x) \ln \left(\frac{M\psi(x)}{t_m(x)} \right) < +\infty. \end{aligned}$$

Then it can be easily shown that (58) is valid for all $m \geq 1$ by changing the constant $c_{\underline{\beta}, \bar{\beta}}$. \square

Proof of Theorem 4. In order to upper bound the right-hand side of the oracle inequality (3), we first control the constant \mathcal{A} defined by (2) that depends on the parameters of the Gaussian mixture model \mathcal{S}_m :

$$\mathcal{A}^2 \leq 4 \left\{ \ln(6\pi e^2) + \pi + \ln \left(\bar{\mu}(m) \sqrt{\frac{8}{c_1 \underline{\lambda}(m)}} \right) + \ln \left(\frac{144 \bar{\lambda}(m)}{\underline{\lambda}(m)} \right) \right\}.$$

For the third term, we note that

$$\begin{aligned} \ln \left(\bar{\mu}(m) \sqrt{\frac{8}{c_1 \underline{\lambda}(m)}} \right) &= \frac{1}{2} \ln \left(\frac{4(\tilde{G}_{\bar{\beta}})^2 |\ln \underline{\lambda}(m)|}{c_1 \underline{\lambda}(m)} \right) \\ &\leq c_{\bar{\beta}} \ln(m) \end{aligned}$$

since $\sqrt{\underline{\lambda}(m)} := \frac{a_{\underline{\beta}}}{m}(\ln m)^{\frac{3}{2}}$. For the last term,

$$\ln\left(\frac{144\bar{\lambda}(m)}{\underline{\lambda}(m)}\right) = \ln\left(\frac{144m^2\bar{\lambda}}{(a_{\bar{\beta}})^2(\ln m)^3}\right) \leq c_{\bar{\beta}}\ln(m)$$

and thus \mathcal{A}^2 is upper bounded by $c_{\bar{\beta}}\ln(m)$. For the observation of a n sample, the model collection is indexed by $\mathcal{M}_n = \{2, \dots, n\}$ and then $m \leq n$. Thus for all $m \in \mathcal{M}_n$,

$$\begin{aligned} \text{pen}(m) &= \kappa \frac{3m-1}{n} \left\{ 1 + 2\mathcal{A}^2 + \ln\left(\frac{1}{1 \wedge \frac{D(m)}{n}\mathcal{A}^2}\right) \right\} \\ &\leq c_{\bar{\beta}} \frac{m}{n} [\ln n + \ln m] \\ &\leq c_{\bar{\beta}} \frac{m}{n} \ln(n). \end{aligned}$$

According to Lemma 7 and the definition of $\underline{\lambda}(m)$, the oracle inequality is upper bounded by

$$\begin{aligned} \mathbb{E} [d_H^2(s, \hat{s}_{\hat{m}})] &\leq \mathcal{C} \inf_{m \in \mathcal{M}_n} \left[\text{KL}(s, \mathcal{S}_m) + \text{pen}(m) + \frac{1}{n} \right] \\ &\leq c_{\underline{\beta}, \bar{\beta}} \inf_{m \in \mathcal{M}_n} \left[\frac{(\ln m)^{3\beta}}{m^{2\beta}} + m \frac{\ln n}{n} \right]. \end{aligned}$$

Let $m_n := \inf \left\{ m \geq 2; m \in \mathbb{N}; \frac{(\ln m)^{3\beta}}{m^{2\beta}} \leq m \frac{\ln n}{n} \right\}$. Note that if $m_n = 2$, then $\mathbb{E} [d_H^2(s, \hat{s}_{\hat{m}})] \leq 4c_{\underline{\beta}, \bar{\beta}} \frac{\ln n}{n}$ and this case is completed. Assuming now that $m_n > 2$, we want to check that $m_n \leq n$. According to the definition of m_n ,

$$\frac{(m_n - 1)^{2\beta+1}}{[\ln(m_n - 1)]^{3\beta}} < \frac{n}{\ln n}$$

thus

$$\left[\frac{(m_n - 1)}{\ln(m_n - 1)} \right]^{\square} < \frac{n}{\ln n}$$

where $\square = 3\beta$ if $\beta > 1$ and $\square = 2\beta + 1$ otherwise. Next, since $\frac{(m_n - 1)}{\ln(m_n - 1)} > 1$,

$$\frac{(m_n - 1)}{\ln(m_n - 1)} < \frac{n}{\ln n}$$

in all cases. Assuming that $n \geq 3$, it leads that $m_n \leq n$. Since $m_n \in \mathcal{M}_n$,

$$\begin{aligned} \mathbb{E} [d_H^2(s, \hat{s}_{\hat{m}})] &\leq 2c_{\underline{\beta}, \bar{\beta}} \frac{[\ln(m_n - 1)]^{3\beta}}{(m_n - 1)^{2\beta}} \\ &\leq 2c_{\underline{\beta}, \bar{\beta}} 2^{2\beta} \frac{[\ln m_n]^{3\beta}}{m_n^{2\beta}} \\ &\leq 2^{2\beta+1} c_{\underline{\beta}, \bar{\beta}} [\ln m_n]^{3\beta} \left[\frac{\ln n}{n} (\ln m)^{-3\beta} \right]^{\frac{2\beta}{2\beta+1}} \\ &\leq \tilde{c}_{\underline{\beta}, \bar{\beta}} n^{-\frac{2\beta}{2\beta+1}} (\ln n)^{\frac{5\beta}{2\beta+1}}. \end{aligned}$$

□

6. Conclusion

In this paper, the penalized estimator $\hat{s}_{\hat{m}}$ defined in [Maugis and Michel \(2009\)](#) is shown to be adaptive to the regularity on some density classes \mathcal{H}_β which elements are univariate densities whose logarithm is locally β -Hölder. To prove this result, the approximation result given in [Kruijer et al. \(2010\)](#) has been adapted to control the bias term between our Gaussian mixture models and the density classes \mathcal{H}_β . A lower bound for the minimax risk on the density classes \mathcal{H}_β has also been stated to finally prove that our estimator reaches the minimax rate.

In [Maugis and Michel \(2009\)](#), a Gaussian mixture estimator, fulfilling an oracle inequality as (3), is proposed in the context of multivariate data clustering. In a future work, it would be interesting to extend our adaptive result to this multivariate case. This requires to state an approximation result as Theorem 2 on multivariate density classes which have to be determined, that is obviously a technical task.

Acknowledgements

The authors are grateful to Willem Kruijer and Judith Rousseau for helpful discussions.

Appendix A: Appendices for the approximation result

A.1. Gaussian kernel properties

Lemma 8. *Let $\psi_{(p)}(x) = C_p e^{-|x|^p}$ for all reals x where C_p denotes the normalizing constant $\left[2\Gamma(1 + \frac{1}{p})\right]^{-1}$. Given a positive integer u , let $\varphi_{u,p}$ be the u -fold convolution of $\psi_{(p)}$. Then, for any $t \geq 0$ and for all $H > 0$, there exists a number $k' = k'(p, t, u, H)$ such that for all $\sigma < 1$,*

$$\int_{|x| > k' |\ln \sigma|^{1/p}} \varphi_{u,p}(x) |x|^t dx = O_{\beta, H}(\sigma^H).$$

Furthermore, k' is a continuous function of H .

The reader is referred to Lemma 10 in [Kruijer et al. \(2010\)](#) for the proof of Lemma 8. Next lemma is a technical result used in Lemma 1 to prove the general case $\beta \geq 2$.

Lemma 9. *For all positive integer u and for all integer $k \geq u$,*

$$\sum_{j=1}^{k+1} (-1)^j \binom{k+1}{j} \nu_{j,2u} = 0$$

where $\nu_{j,h}$ is the h -th moment of the j -fold convolution of the Gaussian kernel ψ .

Proof. Let $u = 1$ and $k \in \mathbb{N}^*$. For all $j \in \{1, \dots, k+1\}$, let (X_1, \dots, X_j) be a sample with density ψ . Then

$$\nu_{j,2} = \mathbb{E}[(X_1 + \dots, X_j)^2] = \sum_{q_1 + \dots + q_j = 2} \frac{2!}{q_1! \dots q_j!} \mathbb{E}[X_1^{q_1}] \dots \mathbb{E}[X_j^{q_j}] = j \mathbb{E}[X_1^2] = j \nu_{1,2}$$

since the odd moments of ψ are equal to zero. Thus,

$$\sum_{j=1}^{k+1} (-1)^j \binom{k+1}{j} \nu_{j,2} = \sum_{j=1}^{k+1} (-1)^j \binom{k+1}{j} j \nu_{1,2} = (k+1) \nu_{1,2} \sum_{j=0}^k (-1)^j \binom{k}{j} = 0.$$

We assume now that the result is true until rank $u-1$. Let $k \geq u$ and note that

$$\sum_{j=1}^{k+1} (-1)^j \binom{k+1}{j} \nu_{j,2u} = \left[\sum_{t=1}^{k+1} (-1)^t \binom{k+1}{t} \right] \nu_{1,2u} + \sum_{j=2}^{k+1} \left[\sum_{t=j}^{k+1} (-1)^t \binom{k+1}{t} \right] (\nu_{j,2u} - \nu_{j-1,2u}).$$

Moreover, $\nu_{j,2u} = \sum_{h=0}^{2u} \binom{2u}{h} \mathbb{E}[(X_1 + \dots + X_{j-1})^{2u-h}] \mathbb{E}[X_j^h] = \sum_{p=0}^u \binom{2u}{2p} \nu_{j-1,2(u-p)} \nu_{1,2p}$ with the convention that $\nu_{h,0} = 1$. Thus,

$$\begin{aligned} \sum_{j=1}^{k+1} (-1)^j \binom{k+1}{j} \nu_{j,2u} &= \left[\sum_{t=1}^{k+1} (-1)^t \binom{k+1}{t} \right] \nu_{1,2u} + \sum_{j=2}^{k+1} \left[\sum_{t=j}^{k+1} (-1)^t \binom{k+1}{t} \right] \left(\sum_{p=1}^u \binom{2u}{2p} \nu_{j-1,2(u-p)} \nu_{1,2p} \right) \\ &= \left[\sum_{t=1}^{k+1} (-1)^t \binom{k+1}{t} + \sum_{j=2}^{k+1} \sum_{t=j}^{k+1} (-1)^t \binom{k+1}{t} \right] \nu_{1,2u} \end{aligned} \quad (59)$$

$$+ \sum_{p=1}^{u-1} \binom{2u}{2p} \nu_{1,2p} \left\{ \sum_{j=2}^{k+1} \left[\sum_{t=j}^{k+1} (-1)^t \binom{k+1}{t} \right] \nu_{j-1,2(u-p)} \right\}. \quad (60)$$

It can be checked that the term inside the brackets of (59) is null. For (60), noting that

$$\begin{aligned} \sum_{t=j+1}^{k+1} (-1)^t \binom{k+1}{t} &= \sum_{t=j+1}^k (-1)^t \left[\binom{k}{t-1} + \binom{k}{t} \right] + (-1)^{k+1} \\ &= (-1)^k - (-1)^j \binom{k}{j} + (-1)^{k+1} = -(-1)^j \binom{k}{j}, \end{aligned}$$

we have that for all $1 \leq p \leq u-1$,

$$\sum_{j=2}^{k+1} \left[\sum_{t=j}^{k+1} (-1)^t \binom{k+1}{t} \right] \nu_{j-1,2(u-p)} = - \sum_{j=1}^k (-1)^j \binom{k}{j} \nu_{j,2(u-p)} = 0$$

according to the induction assumption. Finally, $\sum_{j=1}^{k+1} (-1)^j \binom{k+1}{j} \nu_{j,2u} = 0$. \square

A.2. Measure discretization

The following result is adapted from Lemma 2 in Ghosal and van der Vaart (2007). It allows us to approximate a general Gaussian mixture by a finite Gaussian mixture with a limited number of components.

Proposition 3.

Let F be a probability measure on $[-a, a]$ and $\sigma > 0$ such that $\sigma < a$. Let $\varepsilon \in (0, \pi^{-\frac{1}{2}})$. Then there exists a discrete distribution F' on $[-a, a]$ with at most $54a\sigma^{-1}e^2 \left[1 \vee \ln \left(\frac{1}{\sqrt{\pi\varepsilon}}\right)\right]$ support points such that

$$\|F * \psi_\sigma - F' * \psi_\sigma\|_\infty \leq \frac{2\varepsilon}{\sigma}.$$

Proof. The interval $[-a, a]$ can be partitioned into $k = \lfloor \frac{2a}{\sigma} \rfloor$ disjoint consecutive subintervals I_1, \dots, I_k of length σ and a final subinterval I_{k+1} of length $l \leq \sigma$: $I_i = [a_i, a_i + \sigma[$, $i = 1, \dots, k$ and $I_{k+1} = [a_{k+1}, a_{k+1} + l]$.

We decompose F on this partition $F = \sum_{i=1}^{k+1} F(I_i)F_i$ where each F_i is a probability measure concentrated on I_i . Then, $F * \psi_\sigma(x) = \sum_{i=1}^{k+1} F(I_i)(F_i * \psi_\sigma)(x)$. Let Z_i be a random variable distributed according to F_i , and let G_i be the law of $W_i = (Z_i - a_i)/\sigma$. Thus G_i is a probability measure on $[0, 1]$ for $i = 1, \dots, k$ and on $[0, l/\sigma] \subset [0, 1]$ for $i = k+1$. Lemma 10 is applied for each measure G_i

and with $D = \ln \left(\frac{1}{\sqrt{\pi\varepsilon}}\right)^{-\frac{1}{2}}$. We obtain discrete distributions G'_i such that $\|G_i * \psi - G'_i * \psi\|_\infty \leq 2\varepsilon$. Let F'_i be the law of $a_i + \sigma W'_i$ if W'_i has law G'_i and set $F' = \sum_{i=1}^{k+1} F(I_i)F'_i$. We have

$$F_i * \psi_\sigma(x) = \mathbb{E}[\psi_\sigma(x - Z_i)] = \mathbb{E}\left[\frac{1}{\sigma}\psi\left(\frac{x - Z_i}{\sigma}\right)\right] = \mathbb{E}\left[\frac{1}{\sigma}\psi\left(\frac{x - a_i}{\sigma} - W_i\right)\right] = \frac{1}{\sigma}G_i * \psi\left(\frac{x - a_i}{\sigma}\right)$$

and $F'_i * \psi_\sigma(x) = \frac{1}{\sigma}G'_i * \psi\left(\frac{x - a_i}{\sigma}\right)$. Thus

$$|F_i * \psi_\sigma(x) - F'_i * \psi_\sigma(x)| = \frac{1}{\sigma} \left| G_i * \psi\left(\frac{x - a_i}{\sigma}\right) - G'_i * \psi\left(\frac{x - a_i}{\sigma}\right) \right| \leq \frac{1}{\sigma} \|G_i * \psi - G'_i * \psi\|_\infty \leq \frac{2\varepsilon}{\sigma}.$$

Then

$$\begin{aligned} |F * \psi_\sigma(x) - F' * \psi_\sigma(x)| &= \left| \sum_{i=1}^{k+1} F(I_i) [F_i * \psi_\sigma(x) - F'_i * \psi_\sigma(x)] \right| \\ &\leq \frac{2\varepsilon}{\sigma} \sum_{i=1}^{k+1} F(I_i). \end{aligned}$$

Thus $\|F * \psi_\sigma - F' * \psi_\sigma\|_\infty \leq \frac{2\varepsilon}{\sigma}$ and the number of support points of the discrete distribution F' is upper bounded by

$$\begin{aligned} \sum_{i=1}^{k+1} 18 \left[1 \vee \ln \left(\frac{1}{\sqrt{\pi\varepsilon}} \right)^{-1/2} \right]^2 e^2 \ln \left(\frac{1}{\sqrt{\pi\varepsilon}} \right) &= (k+1) 18 \left[1 \vee \ln \left(\frac{1}{\sqrt{\pi\varepsilon}} \right)^{-1} \right] e^2 \ln \left(\frac{1}{\sqrt{\pi\varepsilon}} \right) \\ &\leq 54a\sigma^{-1}e^2 \left[1 \vee \ln \left(\frac{1}{\sqrt{\pi\varepsilon}} \right) \right]. \end{aligned}$$

□

The following lemma is an adaptation of Lemma 3.1 in Ghosal and van der Vaart (2001). For this lemma, one introduces the inverse function of $\psi_\sigma(\cdot)$ defined by $\psi_\sigma^{-1}(y) = \sigma\sqrt{-\ln(\sqrt{\pi}y)}$ on $(0, \pi^{-\frac{1}{2}}]$.

Lemma 10. *Let F be a probability measure on $[0, B]$. Let $\varepsilon \in (0, \pi^{-\frac{1}{2}})$ and let D be a positive constant such that $B \leq D\psi^{-1}(\varepsilon)$. Then there exists a discrete distribution F' on $[0, B]$ with at most $18(1 \vee D)^2 e^2 \ln\left(\frac{1}{\sqrt{\pi\varepsilon}}\right)$ support points such that*

$$\|F * \psi - F' * \psi\|_\infty \leq 2\varepsilon.$$

Proof. Let x_0 be a positive constant which can be calibrated.

- Case 1: Suppose that $|x| \geq x_0$. Then,

$$\begin{aligned} |F * \psi(x) - F' * \psi(x)| &\leq \int_0^B \psi(x-u) d(|F - F'|)(u) \\ &\leq \frac{1}{\sqrt{\pi}} \int_0^B \exp[-|x-u|^2] d(|F - F'|)(u). \end{aligned}$$

If $x_0 \geq 2B$ then $x_0 - B \geq \frac{x_0}{2}$. Thus, for all $|x| > x_0$ and $|u| \leq B$, $|x-u| \geq (x_0 - B)^2 \geq \frac{x_0^2}{4}$ and

$$|F * \psi(x) - F' * \psi(x)| \leq \frac{1}{\sqrt{\pi}} \int_0^B \exp[-|x-u|^2] d(|F - F'|)(u) \leq \frac{2}{\sqrt{\pi}} \exp\left[-\frac{x_0^2}{4}\right].$$

If $\varepsilon \leq \pi^{-\frac{1}{2}}$, we choose x_0 such that $\exp\left[-\frac{x_0^2}{4}\right] \leq \varepsilon\sqrt{\pi} \Leftrightarrow x_0 \geq 2\sqrt{-\ln(\sqrt{\pi\varepsilon})} = 2\psi^{-1}(\varepsilon)$.

Finally if $x_0 = 2\max(B, \psi^{-1}(\varepsilon))$, then $\|F * \psi - F' * \psi\|_{\infty, [-x_0, x_0]^c} \leq 2\varepsilon$.

- Case 2: Suppose that $|x| \leq x_0$. By Taylor's expansion of e^y and $k! \geq k^k e^{-k}$, we have for any $y < 0$, $k > 1$,

$$\left| e^y - \sum_{j=0}^{k-1} \frac{y^j}{j!} \right| \leq \left| \frac{y^k}{k!} \right| \leq \left(\frac{e|y|}{k} \right)^k.$$

We use this inequality with $y = -x^2$ thus

$$\left| \psi(x) - \frac{1}{\sqrt{\pi}} \sum_{j=0}^{k-1} \frac{x^{2j} (-1)^j}{j!} \right| \leq \frac{1}{\sqrt{\pi}} \left(\frac{ex^2}{k} \right)^k.$$

Then, it leads to

$$|F * \psi(x) - F' * \psi(x)| \leq \left| \int_0^B \psi(x-u) - \frac{1}{\sqrt{\pi}} \sum_{j=0}^{k-1} \frac{(x-u)^{2j} (-1)^j}{j!} d(F - F')(u) \right| \quad (61)$$

$$+ \left| \int_0^B \frac{1}{\sqrt{\pi}} \sum_{j=0}^{k-1} \frac{(x-u)^{2j} (-1)^j}{j!} d(F - F')(u) \right|. \quad (62)$$

The term (62) can be written

$$\left| \int_0^B \frac{1}{\sqrt{\pi}} \sum_{j=0}^{k-1} \frac{(x-u)^{2j} (-1)^j}{j!} d(F - F')(u) \right| = \left| \sum_{j=0}^{k-1} \sum_{t=0}^{2j} \frac{1}{\sqrt{\pi} j!} \binom{2j}{t} (-1)^j (-x)^{2j-t} \int_0^B u^t d(F - F')(u) \right|.$$

According to Lemma A.1 in Ghosal and van der Vaart (2001), there exists a discrete distribution F' with at most $2k - 1$ support points such that $\int_0^B u^t dF'(u) = \int_0^B u^t dF(u)$ for all $1 \leq t \leq 2k - 2$. Finally, considering this discrete distribution F' , we obtain that (62) is null. For the term (61),

$$\left| \int_0^B \psi(x-u) - \frac{1}{\sqrt{\pi}} \sum_{j=0}^{k-1} \frac{(x-u)^{2j} (-1)^j}{j!} d(F-F')(u) \right| \leq \int_0^B \frac{1}{\sqrt{\pi}} \left(\frac{e(x-u)^2}{k} \right)^k d(F+F')(u).$$

Since $|x| < x_0$ and $0 \leq u \leq B$, $|x-u| \leq |x| + |u| \leq x_0 + B \leq \frac{3x_0}{2}$, we obtain that

$$\int_0^B \frac{1}{\sqrt{\pi}} \left(\frac{e(x-u)^2}{k} \right)^k d(F+F')(u) \leq \frac{2}{\sqrt{\pi}} \left(\frac{9ex_0^2}{4k} \right)^k.$$

Moreover, since $x_0 = 2 \max(B, \psi^{-1}(\varepsilon))$ and $B \leq D\psi^{-1}(\varepsilon)$, $x_0 \leq 2(1 \vee D)\psi^{-1}(\varepsilon) = 2(1 \vee D)\sqrt{\ln\left(\frac{1}{\sqrt{\pi}\varepsilon}\right)}$. Thus

$$\begin{aligned} \frac{2}{\sqrt{\pi}} \left(\frac{9ex_0^2}{4k} \right)^k &\leq \frac{2}{\sqrt{\pi}} \left[\frac{9e(1 \vee D)^2}{k} \ln\left(\frac{1}{\sqrt{\pi}\varepsilon}\right) \right]^k \\ &\leq \frac{2}{\sqrt{\pi}} \exp \left[-k \left\{ \ln\left(\frac{9e(1 \vee D)^2}{k}\right) - \ln\left[\ln\left(\frac{1}{\sqrt{\pi}\varepsilon}\right)\right] \right\} \right]. \end{aligned}$$

We have that $\varepsilon \leq \pi^{-\frac{1}{2}} \leq 1$ and we choose k such that $k \geq \ln\left(\frac{1}{\varepsilon}\right)$ and $\ln\left(\frac{9e(1 \vee D)^2}{k}\right) - \ln\left[\ln\left(\frac{1}{\sqrt{\pi}\varepsilon}\right)\right] \geq 1$. This is the case if $k = 9(1 \vee D)^2 e^2 \ln\left(\frac{1}{\sqrt{\pi}\varepsilon}\right)$. Finally, the term (61) is upper bounded by $2\pi^{-\frac{1}{2}}\varepsilon \leq 2\varepsilon$ and $\|F * \psi - F' * \psi\|_{\infty, [-x_0, x_0]} \leq 2\varepsilon$.

□

A.3. Technical results for f , f_k , g_k , h_k and their convolutions

The following lemma allows to bound the derivative functions of $\ln f$. It is based on the smoothness assumptions (4) and (5) and is used in the proof of Lemma 1.

Lemma 11.

For all $j \in \{0, \dots, r\}$ and for all $n \in \mathbb{Z}$, there exists a constant $0 < l_{j,n}^+ < \infty$ such that for all $f \in \mathcal{H}(\beta, \mathcal{P})$,

$$\sup_{y \in [n\gamma, (n+1)\gamma]} \left| (\ln f)^{(j)}(y) \right| \leq l_{j,n}^+.$$

Proof. We first prove Lemma 11 on $[-\gamma, \gamma]$. For all $j \in \{1, \dots, r\}$, all $f \in \mathcal{H}(\beta, \mathcal{P})$ and all $y \in [-\gamma, \gamma]$, there exists $\tilde{y} \in [-|y|, |y|]$ such that

$$(\ln f)^{(j)}(y) = \sum_{u=0}^{r-j} \frac{(\ln f)^{(j+u)}(0)}{u!} y^u + \frac{y^{r-j}}{(r-j)!} \left[(\ln f)^{(r)}(\tilde{y}) - (\ln f)^{(r)}(0) \right].$$

Thus,

$$\begin{aligned}
|(\ln f)^{(j)}(y)| &\leq \sum_{u=0}^{r-j} \frac{|(\ln f)^{(j+u)}(0)|}{u!} |y|^u + \frac{r!}{(r-j)!} |y|^{r-j} L(0) |\tilde{y}|^{\beta-r} \\
&\leq l^+ \sum_{u=0}^{r-j} \frac{\gamma^u}{u!} + \frac{r!}{(r-j)!} L(0) \gamma^{\beta-j} \\
&\leq l^+ \exp(\gamma) + \frac{r!}{(r-j)!} L(0) \gamma^{\beta-j} := l_{j,-1}^+ = l_{j,0}^+
\end{aligned}$$

and Lemma 11 is proved for $n = -1$ and $n = 0$. Now, assume that Lemma 11 is valid for $n - 1 \geq 0$. Then, proceeding as before, for all $j \in \{1, \dots, r\}$ and all $y \in [n\gamma, (n+1)\gamma]$, there exists $\tilde{y} \in [n\gamma, (n+1)\gamma]$ such that

$$(\ln f)^{(j)}(y) = \sum_{u=0}^{r-j-1} \frac{(\ln f)^{(j+u)}(n\gamma)}{u!} (y - n\gamma)^u + \frac{(y - n\gamma)^{r-j}}{(r-j)!} [(\ln f)^{(r)}(\tilde{y}) - (\ln f)^{(r)}(n\gamma)]$$

and thus

$$\begin{aligned}
|(\ln f)^{(j)}(y)| &\leq \sum_{u=0}^{r-j} \frac{|(\ln f)^{(j+u)}(n\gamma)|}{u!} \gamma^u + \frac{r!}{(r-j)!} \gamma^{\beta-j} L(n\gamma) \\
&\leq \sum_{u=0}^{r-j} l_{j+u, n-1}^+ \frac{\gamma^u}{u!} + \frac{r!}{(r-j)!} L(n\gamma) \gamma^{\beta-j} := l_{j,n}^+.
\end{aligned}$$

Finally, Lemma 11 is proved for all $n \in \mathbb{N}$ and a similar proof gives this result for all $n \in \mathbb{Z} \setminus \mathbb{N}$. \square

Lemma 12. Let $f_0 = f$ and $\forall k \in \mathbb{N}^*$, $f_{k+1} = f - \Delta_\sigma f_k$ with $\Delta_\sigma f_k = K_\sigma f_k - f_k$.

1. For all $x \in \mathbb{R}$, $f_k(x) = \sum_{i=0}^k \binom{k+1}{i+1} (-1)^i K_\sigma^i f(x)$.
2. For all $k \in \mathbb{N}$, $\int_{\mathbb{R}} f_k(x) dx = 1$.
3. For all $i \in \mathbb{N}$ and for all $x \in \mathbb{R}$, $K_\sigma^i f(x) \leq \frac{M}{\sqrt{\pi}}$ and thus $|f_k(x)| \leq (2^{k+1} - 1) \frac{M}{\sqrt{\pi}}$.

Proof. The first result is trivial for $k = 0$. For $k = 1$, we remark that $f_1(x) = f(x) - \Delta_\sigma f(x) = 2f(x) - K_\sigma f(x)$. Then recursively, we have

$$\begin{aligned}
f_{k+1}(x) &= f(x) - K_\sigma f_k(x) + f_k(x) \\
&= f(x) - \sum_{i=0}^k \binom{k+1}{i+1} (-1)^i K_\sigma^{i+1} f(x) + \sum_{i=0}^k \binom{k+1}{i+1} (-1)^i K_\sigma^i f(x) \\
&= f(x) + \sum_{j=1}^{k+1} \binom{k+1}{j} (-1)^j K_\sigma^j f(x) + \sum_{i=0}^k \binom{k+1}{i+1} (-1)^i K_\sigma^i f(x) \\
&= \binom{k+1}{k+1} (-1)^{k+1} K_\sigma^{k+1} f(x) + \sum_{j=0}^k [\binom{k+1}{j} + \binom{k+1}{j+1}] (-1)^j K_\sigma^j f(x).
\end{aligned}$$

Since $\binom{k+1}{j} + \binom{k+1}{j+1} = \binom{k+2}{j+1}$ and $\binom{k+1}{k+1} = \binom{k+2}{k+2}$, we have

$$f_{k+1}(x) = \sum_{i=0}^{k+1} \binom{k+2}{i+1} (-1)^i K_\sigma^i f(x).$$

Consequently, for all $k \in \mathbb{N}$, $\int_{\mathbb{R}} f_k(x) dx = \sum_{i=0}^k \binom{k+1}{i+1} (-1)^i \int_{\mathbb{R}} K_\sigma^i f(x) dx$. Moreover, it can be easily proved by induction that for all nonnegative integer i , $\int_{\mathbb{R}} K_\sigma^i f(x) dx = 1$. Thus,

$$\int_{\mathbb{R}} f_k(x) dx = \sum_{i=0}^k \binom{k+1}{i+1} (-1)^i = - \sum_{i=0}^{k+1} \binom{k+1}{i} (-1)^i + \binom{k+1}{0} = (1-1)^{k+1} + 1 = 1.$$

For the third result, according to Condition (7), $f(x) \leq M\psi(x) \leq \frac{M}{\sqrt{\pi}}$. And by induction,

$$K_\sigma^i f(x) = \int_{\mathbb{R}} K_\sigma^{i-1} f(u) \psi_\sigma(x-u) du \leq \frac{M}{\sqrt{\pi}} \int_{\mathbb{R}} \psi_\sigma(x-u) du \leq \frac{M}{\sqrt{\pi}}.$$

Finally, $|f_k(x)| \leq \sum_{i=0}^k \binom{k+1}{i+1} K_\sigma^i f(x) \leq (2^{k+1} - 1) \frac{M}{\sqrt{\pi}}$. □

Lemma 13. *Let $\beta > 0$ and $k \in \mathbb{N}$ such that $\beta \in (2k, 2k+2]$. Let f be a density function belonging to $\mathcal{H}(\beta, \mathcal{P})$ where $\mathcal{P} = \{\gamma, l^+, L, \varepsilon, C, \alpha, \xi, M\}$.*

1. *Let $\bar{\sigma} > 0$ such that if Y is distributed from a centered Gaussian density with variance $\bar{\sigma}^2$, then $P(0 < Y < 2\alpha) = \frac{1}{3}$. For all $\sigma < \bar{\sigma}$,*

$$K_\sigma f(x) \geq \frac{\xi\sqrt{\pi}}{3M} f(x). \tag{63}$$

2. *There exists $\bar{\sigma}(\beta) > 0$ and $A_\beta > 0$ such that for all $\sigma < \bar{\sigma}(\beta)$,*

$$K_\sigma h_k(x) \geq \frac{\xi\sqrt{\pi}}{6M(1 + A_\beta\sigma^{2\beta})} f(x).$$

Furthermore, $\bar{\sigma}(\beta)$ can be chosen as a continuous function of β .

Remarks 7. *The first result of Lemma 13 is based on the monotonicity assumption on f . It comes from Remark 3 of Ghosal et al. (1999). In the second result, the constants $\bar{\sigma}(\beta)$ and A_β are due to the result (30) in Lemma 2.*

Proof. For the first point, let $\sigma < \bar{\sigma}$ and Z be a standard centered Gaussian random variable.

- If $x \in [-\alpha, \alpha]$,

$$\begin{aligned}
K_\sigma f(x) &\geq \int_{-\alpha}^{\alpha} f(u) \psi_\sigma(x-u) du \\
&\geq \xi \int_{\frac{x-\alpha}{\sigma}}^{\frac{x+\alpha}{\sigma}} \psi(z) dz \\
&\geq \xi \left\{ P\left(\frac{x-\alpha}{\sigma} \leq Z \leq 0\right) + P\left(0 \leq Z \leq \frac{2\alpha}{\sigma}\right) - P\left(\frac{x+\alpha}{\sigma} \leq Z \leq \frac{2\alpha}{\sigma}\right) \right\} \\
&\geq \xi \left\{ \frac{1}{3} + P\left(0 \leq Z \leq \frac{\alpha-x}{\sigma}\right) - P\left(\frac{x+\alpha}{\sigma} \leq Z \leq \frac{2\alpha}{\sigma}\right) \right\} \\
&\geq \frac{\xi}{3}.
\end{aligned}$$

According to Condition (7), $f(x) \leq M\psi(x) \leq \pi^{-1/2}M$ for all $x \in \mathbb{R}$. Then for all $x \in [-\alpha, \alpha]$,

$$K_\sigma f(x) \geq \frac{\xi\sqrt{\pi}}{3M} f(x).$$

- If $x > \alpha$,

$$\begin{aligned}
K_\sigma f(x) &\geq f(x) \int_0^{\frac{x+\alpha}{\sigma}} \psi(u) du \\
&\geq f(x) \left\{ P\left(0 \leq Z \leq \frac{2\alpha}{\sigma}\right) + P\left(\frac{2\alpha}{\sigma} \leq Z \leq \frac{x+\alpha}{\sigma}\right) \right\} \\
&\geq f(x) P\left(0 \leq Z \leq \frac{2\alpha}{\sigma}\right) \\
&\geq \frac{1}{3} f(x).
\end{aligned}$$

In the same way, for all $x < -\alpha$, $K_\sigma f(x) \geq \frac{1}{3} f(x)$.

Finally, since $M\pi^{-1/2} \geq \xi$, $K_\sigma f(x) \geq \frac{\xi\sqrt{\pi}}{3M} f(x)$ for all $x \in \mathbb{R}$.

For the second point, we take $H_1 = 4\beta + 1$ as in Remark 6; let $\sigma < \bar{\sigma}(\beta)$ in order to have (32) and (63). Then, for all $\sigma < \bar{\sigma}(\beta)$, $\int_{\mathbb{R}} g_k(u) du \leq 1 + A_\beta \sigma^{2\beta}$ and since $g_k(x) \geq \frac{1}{2} f(x)$ for all $x \in \mathbb{R}$,

$$\begin{aligned}
K_\sigma h_k(x) &= \int_{\mathbb{R}} h_k(u) \psi_\sigma(x-u) du \\
&\geq \int_{\mathbb{R}} \frac{f(u)}{2 \int g_k} \psi_\sigma(x-u) du \\
&\geq \frac{\xi\sqrt{\pi}}{6M(1 + A_\beta \sigma^{2\beta})} f(x).
\end{aligned}$$

□

Lemma 14. *Let $p \in (0, 1)$. For all $x \in \mathbb{R}$, we have that*

- for all $i \in \mathbb{N}$ and for all $\sigma < 1 - p^{1/i}$, $K_\sigma^i f(x) \leq M \left(\frac{2}{\sqrt{3}}\right)^i \psi(px)$.
- for all $\sigma < 1 - p^{1/k}$,

$$\max \left(f_k(x), g_k(x), \frac{1}{2} h_k(x) \right) \leq 2M \left(\frac{4}{\sqrt{3}} \right)^k \psi(px).$$

Proof. The control of $K_\sigma^i f$ can be proved by applying successively Lemma 15 to f , $K_\sigma f$, ..., $K_\sigma^{i-1} f$ with $q_1 = p^{k/i}$ and $q_2 = p^{1/i}$ for each step k . It finally gives that $K_\sigma^i f(x) \leq M \left(\frac{4}{\sqrt{3}}\right)^i \psi(px)$ for all x in \mathbb{R} and for all $\sigma < 1 - p^{1/i}$. This control on $K_\sigma^i f$ together with Lemma 12 give the control on f_k . According to the definition of g_k and previous results,

$$g_k(x) \leq 2M \left(\frac{4}{\sqrt{3}} \right)^k \psi(px) \mathbb{1}_{J_{\sigma,k}}(x) + M\psi(x) \mathbb{1}_{J_{\sigma,k}^c}(x) \leq 2M \left(\frac{4}{\sqrt{3}} \right)^k \psi(px).$$

Finally, $h_k(x) = g_k(x) / \int_{\mathbb{R}} g_k(y) dy \leq 2g_k(x) \leq 4M \left(\frac{4}{\sqrt{3}}\right)^k \psi(px)$. \square

Lemma 15. *Let f be a positive application on \mathbb{R} such that there exists $M > 0$ and $q_1 \in (0, 1]$ such that for all $x \in \mathbb{R}$, $f(x) \leq M\psi(q_1 x)$. Then for all $q_2 \in (0, 1)$ and all $\sigma \in (0, 1 - q_2^2)$,*

$$\forall x \in \mathbb{R}, K_\sigma f(x) \leq \frac{2}{\sqrt{3}} M \psi(q_1 q_2 x).$$

Proof. Let $\sigma \in (0, 1 - q_2^2)$. For all $x \in \mathbb{R}$,

$$\begin{aligned} K_\sigma f(x) &= \int_{\mathbb{R}} f(u) \psi_\sigma(x - u) du \\ &\leq \frac{M}{\pi} \int_{\mathbb{R}} \exp[-q_1^2(\sigma y - x)^2] \exp(-y^2) dy \\ &\leq \frac{M}{\pi} \exp[-q_1^2 x^2 (1 - \sigma)] \int_{\mathbb{R}} \exp(-y^2 + q_1^2 y^2 \sigma (1 - \sigma)) dy \\ &\leq \frac{M}{\pi} \exp[-q_1^2 q_2^2 x^2] \int_{\mathbb{R}} \exp\left(-y^2 + \frac{1}{4} q_1^2 y^2\right) dy \\ &\leq \frac{2M}{\sqrt{3}} \psi(q_1 q_2 x). \end{aligned}$$

\square

Appendix B: Appendices for the lower bound result

The two following results are crucial for establishing the lower bound: The first one is the so-called Varshamov-Gilbert's lemma and the second one is a corollary of a lemma given in Birgé (2005). They correspond to Lemma 4.7 and Corollary 2.19 in Massart (2007) respectively.

Lemma 16. Let $\{0, 1\}^D$ be equipped with Hamming distance δ . Given $\alpha \in (0, 1)$, there exists some subset Θ of $\{0, 1\}^D$ with the following properties

$$\begin{cases} \delta(\theta, \theta') > \frac{(1-\alpha)D}{2} \text{ for every } (\theta, \theta') \in \Theta, \theta \neq \theta' \\ \ln |\Theta| > \frac{\rho D}{2} \end{cases}$$

where $\rho = (1 + \alpha) \ln(1 + \alpha) + (1 - \alpha) \ln(1 - \alpha)$. In particular $\rho > \frac{1}{4}$ when $\alpha = \frac{1}{2}$.

Corollary 1. Let (S, d) be some pseudo-metric space, $\{\mathbb{P}_s, s \in S\}$ be some statistical model. Let κ denote an absolute constant (given in Corollary 2.18 of [Massart, 2007](#)). Then for any estimator \tilde{s} and any finite subset \mathcal{C} of S such that $\max_{s, t \in \mathcal{C}} \text{KL}(\mathbb{P}_s, \mathbb{P}_t) \leq \kappa \ln |\mathcal{C}|$, the following lower bound holds for every $p > 1$

$$\sup_{s \in \mathcal{C}} \mathbb{E}_s [d^p(s, \tilde{s})] \geq 2^{-p} (1 - \kappa) \left[\min_{s, t \in \mathcal{C}, s \neq t} d(s, t) \right]^p.$$

The following lemma, used to prove Proposition 1, gives an expression of the derivatives of the logarithm of a function.

Lemma 17. Let $i \in \mathbb{N}^*$ and let t be a strictly positive function, $t \in \mathcal{C}^i$. Then

$$(\ln t)^{(i)}(x) = \frac{P_i(x)}{t(x)^{2^{i-1}}}$$

where

$$P_i(x) = \sum_{(\eta_0, \dots, \eta_i) \in \Xi_i} \rho(\eta_0, \dots, \eta_i) \prod_{j=0}^i [t^{(j)}(x)]^{\eta_j}$$

with

$$\Xi_i = \left\{ (\eta_0, \dots, \eta_i) \in \mathbb{N}^{i+1}; \sum_{j=0}^i \eta_j = 2^{i-1}, \sum_{j=0}^i j \eta_j = i \right\}$$

and $\rho(\eta_0, \dots, \eta_i)$'s are the polynomial coefficients.

Proof. The result is trivial for $i = 1$. Assume that t is \mathcal{C}^{i+1} and that the result is valid for the i -th derivative. Then

$$(\ln t)^{(i+1)}(x) = \frac{t(x)^{2^{i-1}} P_i(x)' - 2^{i-1} t(x)' t(x)^{2^{i-1}-1} P_i(x)}{t(x)^{2^i}} = \frac{\diamond}{t(x)^{2^i}}$$

with

$$\begin{aligned} \diamond = & t(x)^{2^{i-1}} \left\{ \sum_{(\eta_0, \dots, \eta_i) \in \Xi_i} \rho(\eta_0, \dots, \eta_i) \left[\sum_{j=0}^i \eta_j \frac{t(x)^{(j+1)}}{t(x)^{(j)}} \prod_{u=0}^i (t(x)^{(u)})^{\eta_u} \right] \right\} \\ & - 2^{i-1} t(x)' t(x)^{2^{i-1}-1} \sum_{(\eta_0, \dots, \eta_i) \in \Xi_i} \rho(\eta_0, \dots, \eta_i) \prod_{j=0}^i (t(x)^{(j)})^{\eta_j} \end{aligned}$$

$$\begin{aligned} \diamond &= \sum_{(\eta_0, \dots, \eta_i) \in \Xi_i} \rho(\eta_0, \dots, \eta_i) \left[\sum_{j=0}^i \eta_j t(x)^{2^{i-1}} \frac{t(x)^{(j+1)}}{t(x)^{(j)}} \prod_{u=0}^i \left(t(x)^{(u)} \right)^{\eta_u} \right] \\ &\quad - \sum_{(\eta_0, \dots, \eta_i) \in \Xi_i} \rho(\eta_0, \dots, \eta_i) 2^{i-1} t(x)' t(x)^{2^{i-1}-1} \prod_{j=0}^i \left(t(x)^{(j)} \right)^{\eta_j}. \end{aligned}$$

Let $\tilde{\eta}_j$ denotes the new power of the j -th derivative for $j = 0 \dots, i + 1$. In the second sum, we have that $\tilde{\eta}_0 = 2^{i-1} - 1 + \eta_0$, $\tilde{\eta}_1 = \eta_1 + 1$, $\tilde{\eta}_j = \eta_j$ for all $j = 2, \dots, i$ and $\tilde{\eta}_{i+1} = 0$ thus $\sum_{j=0}^{i+1} \tilde{\eta}_j = \sum_{j=0}^i \eta_j + 1 + 2^{i-1} - 1 = 2^i$ and $\sum_{j=0}^{i+1} j \tilde{\eta}_j = \sum_{j=0}^i j \eta_j + 1 = i + 1$. In the first sum,

- if $j < i$: $\tilde{\eta}_0 = 2^{i-1} + \eta_0$, $\tilde{\eta}_j = \eta_j - 1$, $\tilde{\eta}_{j+1} = \eta_{j+1} + 1$, $\forall u \in \{1, \dots, i\} \setminus \{j, j+1\}$, $\tilde{\eta}_u = \eta_u$ and $\tilde{\eta}_{i+1} = 0$ thus $\sum_{u=0}^{i+1} \tilde{\eta}_u = \sum_{u=0}^i \eta_u + 1 + 2^{i-1} - 1 = 2^i$ and $\sum_{u=0}^{i+1} u \tilde{\eta}_u = \sum_{u=0}^i u \eta_u + j + 1 - j = i + 1$.
- if $j = i$: $\tilde{\eta}_0 = 2^{i-1} + \eta_0$, $\tilde{\eta}_i = \eta_i - 1$, $\tilde{\eta}_{i+1} = 1$, $\forall u \in \{1, \dots, i-1\}$, $\tilde{\eta}_u = \eta_u$ thus $\sum_{u=0}^{i+1} \tilde{\eta}_u = \sum_{u=0}^i \eta_u + 1 + 2^{i-1} - 1 = 2^i$ and $\sum_{u=0}^{i+1} u \tilde{\eta}_u = \sum_{u=0}^i u \eta_u + i + 1 - i = i + 1$.

□

References

- Birgé, L. (2005). A new lower bound for multiple hypothesis testing. *IEEE Trans. Inform. Theory.*, 51:1611–1615.
- Cheney, W. and Light, W. (2009). *A course in approximation theory*, volume 101 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI.
- Ghosal, S., Ghosh, J. K., and Ramamoorthi, R. V. (1999). Posterior consistency of Dirichlet mixtures in density estimation. *Annals of Statistics*, 27:143–158.
- Ghosal, S. and van der Vaart, A. (2001). Entropy and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *Annals of Statistics*, 29:1233–1263.
- Ghosal, S. and van der Vaart, A. (2007). Posterior convergence rates of Dirichlet mixtures at smooth densities. *Annals of Statistics*, 35:697–723.
- Grenander, U. (1981). *Abstract inference*. John Wiley and Sons Inc., New York.
- Hanggelbroek, T. and Ron, A. (2010). Nonlinear approximation using Gaussian kernels. *Journal of Functional Analysis*, 259(1):203–219.
- Hartigan, J. A. (1975). *Clustering algorithms*. John Wiley & Sons, New York-London-Sydney. Wiley Series in Probability and Mathematical Statistics.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning*. Springer Series in Statistics. Springer, New York, second edition. Data mining, inference, and prediction.
- Kruijer, W., Rousseau, J., and van der Vaart, A. (2010). Adaptive Bayesian Density Estimation with Location-Scale Mixtures. *Electronic Journal of Statistics*, 4:1225–1257.
- Lindsay, B. (1995). *Mixtures Models: Theory, Geometry and Applications*. IMS, Hayward, CA.
- Massart, P. (2007). *Concentration Inequalities and Model Selection*. *École d'été de Probabilités de Saint-Flour 2003. Lecture Notes in Mathematics*. Springer.

- Maugis, C. and Michel, B. (2009). A non asymptotic penalized criterion for Gaussian mixture model selection. *ESAIM: Probability and Statistics*. doi:10.1051/ps/2009004.
- Maugis, C. and Michel, B. (2010). Data-driven penalty calibration: a case study for Gaussian mixture model selection. *ESAIM: Probability and Statistics*. doi:10.1051/ps/2010002.
- McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*. Wiley.
- Tsybakov, A. B. (2009). *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York.
- Wolfowitz, J. (1950). Minimax estimation of the mean of a normal distribution with known variance. *Annals of Mathematical Statistics*, 21:218–230.