



HAL
open science

Robust density estimation with the L1-loss. Applications to the estimation of a density on the line satisfying a shape constraint

Yannick Baraud, H el ene Halconruy, Guillaume Maillard

► To cite this version:

Yannick Baraud, H el ene Halconruy, Guillaume Maillard. Robust density estimation with the L1-loss. Applications to the estimation of a density on the line satisfying a shape constraint. *Annales de l'Institut Henri Poincar e (B) Probabilit es et Statistiques*, 2025, 61 (3), pp.2098-2148. <10.1214/24-AIHP1485>. <hal-05226651>

HAL Id: hal-05226651

<https://hal.science/hal-05226651v1>

Submitted on 27 Aug 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destin e au d ep ot et  a la diffusion de documents scientifiques de niveau recherche, publi es ou non,  emanant des  tablissements d'enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv es.



HAL Authorization

ROBUST DENSITY ESTIMATION WITH THE L_1 -LOSS. APPLICATIONS TO THE ESTIMATION OF A DENSITY ON THE LINE SATISFYING A SHAPE CONSTRAINT

YANNICK BARAUD, HÉLÈNE HALCONRUY, AND GUILLAUME MAILLARD

ABSTRACT. We tackle the problem of estimating the distribution of presumed i.i.d. observations for the total variation loss. Our approach is based on density models and is versatile enough to cope with many different ones, including some for which the Maximum Likelihood Estimator (MLE for short) does not exist. We mainly illustrate the properties of our estimator on models of densities on the line that satisfy a shape constraint. We show that it possesses some similar optimality properties, with regard to some global rates of convergence, as the MLE does when it exists. It also enjoys some adaptation properties with respect to some specific densities in the model for which our estimator is proven to converge at the parametric rate $1/\sqrt{n}$. More important is the fact that our estimator is robust, not only with respect to model misspecification, but also to contamination, the presence of outliers among the dataset and the equidistribution assumption we started from. This means that the estimator performs almost as well as if the data were i.i.d. with density p in a situation where these data are only independent and most of their marginals are close enough in total variation to a distribution with density p . We also show that our estimator converges to the average density of the data, when this density belongs to the model, even when none of the marginal densities does. Our main result on the risk of the estimator takes the form of an exponential deviation inequality which is nonasymptotic and involves explicit numerical constants. We deduce from it several global rates of convergence, including some bounds for the minimax L_1 -risks over the sets of concave, log-concave or more generally s -concave densities with $s > -1$. These bounds derive from some approximation results of densities which are monotone, convex, concave or more generally s -concave. These results may be of independent interest.

1. INTRODUCTION

Estimating a density under a shape constraint has been addressed by many authors since the pioneering papers by Grenander [16, 17], Rao [26], Groeneboom [18] and Birgé [7] for estimating a nonincreasing density on $(0, +\infty)$. It is well-known that this problem can be elegantly solved by the Grenander estimator – see Grenander [16] – which is the maximum likelihood estimator (MLE for short) over the set of densities that satisfy this monotonicity constraint on $(0, +\infty)$. Practically, the MLE is obtained by taking the (right-hand) derivative of the least concave majorant of the cumulative distribution function of the data. Rao [26] and Grenander [16, 17] established the local asymptotic properties of the Grenander estimator while Groeneboom [18] and Birgé [7] studied its global L_1 -estimation error over some functional classes of interest. Birgé proved that the worst-case risk of the Grenander estimator over the set $\mathcal{F}(H, L)$, that consists of nonincreasing densities bounded by $H > 0$ and supported on $[0, L]$ with $L > 0$, is of order $(\log(1 + HL)/n)^{1/3}$. This rate turns out to be optimal in the minimax sense over this class of densities. The Grenander estimator therefore performs almost as well (apart maybe from numerical constants) as a minimax estimator

2020 *Mathematics Subject Classification.* Primary 62G07, 62G35; Secondary 62C20.

Key words and phrases. Density estimation, robust estimation, shape constraint, total variation loss, minimax theory.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 811017.

over $\mathcal{F}(H, L)$ that would know the values of H and L in advance. Even more surprising is the fact that the Grenander estimator converges at the parametric rate $1/\sqrt{n}$ when the target density is piecewise constant on the elements of partition of $(0, +\infty)$ into a finite number of intervals – see Grenander [17], Groeneboom [18] and Birgé [7]. As a consequence, the Grenander estimator can nicely *adapt* to the specific features of the density to be estimated even though these features are *a priori* unknown.

Because of its nice theoretical properties and the fact that it can easily be calculated in this specific situation, the MLE has widely and almost exclusively been used to solve many other density estimation problems under shape constraints. We refer to Groeneboom *et al* [19] for convex densities and to Balabdaoui and Wellner [1] and Gao and Wellner [15] for k -monotone ones. Although the construction of the MLE is not based on smoothness assumptions in these cases, it still requires that some specific features of the target density, and in particular its support, be known. It was already the case for the monotonicity constraint since the left-endpoint of its support needed to be known to build the Grenander estimator. This kind of prior information might, however, not be available in practice and a more reasonable assumption would be that only an interval containing this support be known. Unfortunately, under this weaker assumption the MLE does not exist and a search for alternative estimators becomes necessary to solve this issue.

To our knowledge, the first attempt to solve it dates back to Wegman [29]. He designed an MLE-type estimator restricted to a class of unimodal densities that attain their modes on an interval of length not smaller than some parameter $\varepsilon > 0$. This parameter needs to be tuned by the statistician and its choice influences the performance of the resulting estimator. Birgé [8] proposed a different approach based on data-driven choice of a Grenander estimator among the collection of those associated to all possible modes. Up to an additional term of order $1/\sqrt{n}$, he proved that the \mathbb{L}_1 -risk of the selected estimator is the same as that of the Grenander estimator that would know the value of the mode in advance.

The situation is different when the density is assumed to be log-concave on \mathbb{R} , or more generally on \mathbb{R}^d with $d > 1$. The construction of the MLE is then free of any assumption on the support of the density. The study of the MLE on the set of log-concave densities has led to intensive work. We refer the reader to Dümbgen and Rufibach [13], Doss and Wellner [12], Cule and Samworth [10], Kim and Samworth [23] and Feng *et al* [14] as well as the references therein. Kim and Samworth [23] described the uniform rates of convergence of the MLE for the squared Hellinger loss over the class of log-concave densities in dimension $d \in \{1, 2, 3\}$ and they proved these rates to be minimax (up to a possible logarithmic factor). Besides, as for the monotonicity constraint in dimension one, the MLE also possesses for log-concave densities some adaptation properties: it converges at the parametric rate $1/\sqrt{n}$ (for the Hellinger loss and up to a possible logarithmic factor) when the logarithm of the target density is piecewise affine on a suitable convex subset of \mathbb{R}^d with $d \in \{1, 2, 3\}$. This result was established by Kim *et al* [22] when $d = 1$ and extended to the dimensions $d \in \{2, 3\}$ by Feng *et al* [14]. In dimension $d \geq 4$, Kur *et al* [25] showed that the MLE converges at a minimax rate (up to a logarithmic factor) for the Hellinger loss.

In the literature, there exist various generalisations of log-concavity. They usually take the following form. Given some function \mathcal{L} on \mathbb{R}_+ , one considers the class of these densities p on \mathbb{R}^d for which $\mathcal{L} \circ p$ is concave. In the present paper, we shall restrict to the case $d = 1$ and call these densities \mathcal{L} -concave. Of particular interest is the case where the function \mathcal{L} is of the form $u \mapsto \text{sign}(s)u^s$, with $s \in \mathbb{R} \setminus \{0\}$. An \mathcal{L} -concave density is then called s -concave and log-concave ones are 0-concave by convention. While log-concave densities possess light tails that decrease at least exponentially fast as we move away from their modes, s -concave densities may exhibit heavier tails that decrease at a polynomial rate. The study of the existence, the consistency and the rates of convergence of the MLE over classes of \mathcal{L} -concave densities can be found in Seregin and Wellner [27] and Doss and Wellner [12]. In particular, Doss and Wellner [12] showed that the MLE does not exist over the class of s -concave densities when $s < -1$ while for $s > -1$, it converges at a rate which is at least $n^{-2/5}$ for the Hellinger loss. The convergence established by Doss and Wellner is, however, not uniform over the

class when $s \in (-1, 0)$. For these values of s , we do not know whether or not the convergence of the MLE is uniform. It seems that the study of the MLE for estimating such densities has received less attention than for log-concave ones. In particular, we are not aware of any adaptation properties of the MLE over classes of s -concave densities when $s \neq 0$. Finally, we mention that some alternative approaches based on Rényi divergences have been proposed by Koenker and Mizera [24] and Han and Wellner [21] to solve the problem of estimating s -concave densities. The results established in these papers mainly address the calculability of these estimators, their connexions with the MLE and their asymptotic performance.

In the one dimensional case, our aim is to design a versatile estimation strategy that can be applied to a wide variety of density models, including some for which the MLE does not exist, and that automatically results in estimators with good estimation properties. In particular, these estimators should keep the nice minimax and adaptation properties of the MLE, when it exists, for estimating a density under a shape constraint. They should also remain stable with respect to a slight departure from the ideal situation where the data are truly i.i.d. and their density satisfies the required shape constraint. In particular, the estimator should still perform well when the equidistribution assumption is slightly violated and the data set contains a small portion of outliers. It should also perform well when the shape of the density is slightly different from what was originally expected, that is, when the true density of the data does not satisfy the shape constraint but is close enough (with respect to the \mathbb{L}_1 -loss) to a density that does satisfy it. In a nutshell, our aim is to build estimators that are *robust*.

By robust, we mean here that the performance of the estimator should remain stable with respect to a small discrepancy (with respect to the Hellinger or the \mathbb{L}_1 distance) between the target density and the statistical model. It turns out that in some cases the MLE provides a robust solution when the support of the density is known or partly known. This is for example the case for estimating a nonincreasing or a convex density on $(a, +\infty)$ when $a \in \mathbb{R}$ is known or a concave density on (a, b) when $a, b \in \mathbb{R}$ are both known. The robustness of the MLE in such models can be viewed as a consequence of Baraud and Birgé [5][Section 6, Proposition 5] which states the following: if the MLE over a convex set of densities exists and takes positive values at the data, it necessarily coincides with a ρ -estimator and therefore inherits its robustness properties. The sets that consist of nonincreasing, convex or concave densities on a given interval are convex and this result therefore applies. An illustration of this is that the Grenander estimator on the set \mathcal{F} of nonincreasing densities on $(0, +\infty)$ remains stable for estimating a density p^* on $(0, +\infty)$ which does not belong to \mathcal{F} but lies close enough to it (with respect to the \mathbb{L}_1 -distance). However, this apparent stability of the Grenander estimator should be interpreted with caution. The following example shows that a small error on the presumed support of p^* may result in a dramatic misspecification of the model. Assume that p^* is the uniform distribution on $(a, 2a)$, where a is a small positive number, and consider an arbitrary (left continuous) density q in the set \mathcal{F} of nonincreasing densities on $(0, +\infty)$. Then,

$$\int_0^{+\infty} |p^*(x) - q(x)| dx \geq \int_0^a q(x) dx + \int_a^{2a} \left| \frac{1}{a} - q(x) \right| dx \geq aq(a) + \left(\frac{1}{a} - q(a) \right) a = 1.$$

In particular, the \mathbb{L}_1 -risk of the Grenander estimator on \mathcal{F} for estimating p^* is at least 1 while it would not exceed $3.6/\sqrt{n}$ if a were known, as proven in Birgé [7] [Theorem 1 and inequality (3.3)]. This example shows that a small error on the presumed support of the target density may result in a large estimation error. Our approach allows one to deal with statistical models that only rely on the presumed shape of the density and not on some information on its support. As a consequence, our estimators do not suffer from such weaknesses.

We are not aware of many robust strategies for estimating a density under a shape constraint. For estimating concave and log-concave densities, Chan *et al* [9] proposed a piecewise linear estimator on a data-driven partition of \mathbb{R} into intervals. Their estimator is minimax optimal on the sets of concave and log-concave densities and it enjoys some robustness properties with respect to a departure (in \mathbb{L}_1 -distance)

of the true density from the model. Their approach is based on the estimation procedure described in Devroye and Lugosi [11] and uses the fact that the Yatracos class associated to the set of the densities that are piecewise linear on a partition of the line into a fixed number of intervals is a Vapnik-Chervonenkis (VC) class. We refer the reader to van der Vaart and Wellner [28] for an introduction to VC classes and their applications to statistics. Despite the desirable properties described above, the estimator proposed by Chan *et al* does not possess some of the nice ones that make the MLE so popular. For estimating a log-concave density, the MLE converges at a global rate of order $n^{-2/5}$ (for the \mathbb{L}_1 and Hellinger distances) but, as already mentioned, it also possesses some adaptation properties with respect to these densities the logarithms of which are piecewise linear. The estimator proposed by Chan *et al* does not possess such a property. Besides, their approach provides competitors to the MLE for some specific density models only. Chan *et al*'s approach cannot deal with the estimation of a monotone density on a half-line for example and therefore cannot be used to provide a surrogate to the Grenander estimator.

In dimension one, Baraud and Birgé [4][Section 7] proposed to solve the problem of robust estimation of a density under a shape constraint by using ρ -estimation. Their results hold for the squared Hellinger loss while ours is for the total variation one (TV-loss for short). The estimator we propose is more specifically designed for this loss and quite surprisingly the risk bounds we get for the TV-loss are slightly different from those obtained by Baraud and Birgé for the (squared) Hellinger one. We do not know if ρ -estimators would satisfy the same \mathbb{L}_1 -risk bounds as those we establish here.

Our procedure shares some similarities with that proposed by Devroye and Lugosi [11]. When the Yatracos class associated with the density model is VC, the risk bound we establish is similar to theirs except for the fact that we provide explicit numerical constants. However, unlike them, we also consider density models for which the Yatracos class is not VC, which is typically the case for models of densities that satisfy a shape constraint. Nevertheless, it is likely that with the same proof techniques, we could establish similar results for Devroye and Lugosi's estimators as those we prove for ours.

The theory of ℓ -estimation introduced in Baraud [3] provides a generic way of building estimators that possess the robustness properties we are looking for. Even though the present paper is in the same vein, we modify Baraud's procedure and establish, for the modified ℓ -estimator, risk bounds with numerical constants that are essentially divided by a factor 2 as compared to his. Another important difference with Baraud's result lies in the following fact. When the data are only independent with marginal densities p_1^*, \dots, p_n^* , we measure the performance of our density estimator \hat{p} in terms of its \mathbb{L}_1 -distance $\|p^* - \hat{p}\|$ between \hat{p} and the average of the marginal densities $p^* = n^{-1} \sum_{i=1}^n p_i^*$. In contrast, Baraud considered, as a loss function, the average of the \mathbb{L}_1 -distances of \hat{p} to the p_i^* , i.e. the quantity $n^{-1} \sum_{i=1}^n \|\hat{p} - p_i^*\|$. As a consequence, unlike Baraud, we can establish the convergence of our estimator to p^* , as soon as it belongs to the model, even in the unfavourable situation where none of the marginals p_i^* belongs to it.

The risk bounds we obtain hold for very general density models but our applications focus on the estimation of a density on the line that satisfies a shape constraint. In a nutshell, we establish the following results which are to our knowledge new in the literature.

- The procedure applies to a large variety of density models including some for which the MLE does not exist, e.g. the set of all monotone densities on an unknown half-line, the set of all unimodal densities on \mathbb{R} , the set of all convex densities on an interval or the set of s -concave densities with $s < -1$.
- The global rates of convergence that we establish for our estimator are optimal in all the models we consider.
- The estimator possesses some adaptation properties: it converges at the parametric rate $1/\sqrt{n}$ when the data are i.i.d. with a density that belongs to the model and satisfies some special properties. In particular, our estimator shares similar adaptation properties as those established for the MLE

under a monotonicity or a log-concavity constraint. We also establish some adaptation properties on density models for which the MLE does not even exist.

- The estimator is robust with respect to model misspecification, contamination, the presence of outliers and is robust with respect to a departure from the equidistribution assumption we started from.

Of particular interest is the class of \mathcal{L} -concave densities. In this special case, we prove that our TV-estimator exists and enjoys some adaptation properties as soon as the function \mathcal{L} is increasing. This result applies thus to all classes of s -concave densities without any restriction on s . If the function \mathcal{L} is additionally convex, the \mathbb{L}_1 -risk of our TV-estimator can be bounded uniformly over the class of \mathcal{L} -concave densities by a quantity that goes to 0 at rate $n^{-2/5}$. A similar bound can also be established when \mathcal{L} is increasing and concave provided that the right and left derivatives of \mathcal{L} satisfy some suitable conditions. We show that these results apply, in particular, to all classes of s -concave densities with parameters $s > -1$.

The paper is organized as follows. The statistical framework is described in Section 2 and the construction of the estimator as well as its properties are presented in Section 3. The more specific properties of our estimator for estimating a mixture of densities that are monotone, convex or concave can be found in Sections 4 and 5 respectively while the cases of log-concave, s -concave and more generally \mathcal{L} -concave densities are tackled in Section 6. These sections also contain some approximation results which may be of independent interest and are central to our approach. Some concluding remarks in Section 7 discuss the specificities and advantages of our approach as well as its limitations. The proofs are postponed to Section 8.

2. THE STATISTICAL FRAMEWORK AND MAIN NOTATIONS

Let X_1, \dots, X_n be n independent random variables and P_1^*, \dots, P_n^* their marginals on a measurable space $(\mathcal{X}, \mathcal{A})$. Our aim is to estimate the n -tuple $\mathbf{P}^* = (P_1^*, \dots, P_n^*)$ from the observation of $\mathbf{X} = (X_1, \dots, X_n)$ on the basis of a suitable *model* for \mathbf{P}^* . More precisely, given a σ -finite measure μ on $(\mathcal{X}, \mathcal{A})$ and a family $\overline{\mathcal{M}}$ of densities with respect to μ , we shall treat the X_i as if they were i.i.d. with a density that belongs to $\overline{\mathcal{M}}$, even though this might not be true, and estimate \mathbf{P}^* by a n -tuple of the form $(\hat{P}, \dots, \hat{P})$ where $\hat{P} = \hat{P}(\mathbf{X}) = \hat{p} \cdot \mu$ is a random element of the set $\overline{\mathcal{M}} = \{P = p \cdot \mu, p \in \overline{\mathcal{M}}\}$. We refer to $\overline{\mathcal{M}}$ and $\overline{\mathcal{M}}$ as our probability and density models respectively. For the sake of simplicity, we abusively identify \mathbf{P}^* with the distribution $\otimes P_i^*$ of the observation \mathbf{X} .

The density models we have in mind are nonparametric and consist of densities that satisfy a given shape constraint: monotonicity on a half line, convexity on an interval, log-concavity on the line, among other examples.

In order to evaluate the accuracy of our estimator, we use the TV-loss d on the set \mathcal{P} of all probability measures on $(\mathcal{X}, \mathcal{A})$. We denote by $\|\cdot\|$ the \mathbb{L}_1 -norm on the set $\mathbb{L}_1(\mathcal{X}, \mathcal{A}, \mu)$ that consists of the equivalence classes of integrable functions on $(\mathcal{X}, \mathcal{A}, \mu)$. We recall that the TV-loss is a distance defined for $P, Q \in \mathcal{P}$ by

$$(1) \quad d(P, Q) = \sup_{A \in \mathcal{A}} [P(A) - Q(A)]$$

and if P and Q are absolutely continuous with respect to our dominating measure μ ,

$$d(P, Q) = \frac{1}{2} \left\| \frac{dP}{d\mu} - \frac{dQ}{d\mu} \right\|.$$

In general, for each \mathbf{P}^* , denote by P^* the uniform mixture of the marginals:

$$P^* = \frac{1}{n} \sum_{i=1}^n P_i^*.$$

In particular, when the data are i.i.d., their common distribution is $P^* \in \mathcal{P}$.

Throughout this paper, we assume the following.

Assumption 1. There exists a countable subset \mathcal{M} of $\overline{\mathcal{M}}$ that is dense in $\overline{\mathcal{M}}$ for the \mathbb{L}_1 -norm.

We recall that a subset of a separable metric space is separable. In particular, when the space $\mathbb{L}_1(\mathcal{X}, \mathcal{A}, \mu)$ is separable for the \mathbb{L}_1 -norm, so is any subset $\overline{\mathcal{M}}$ of densities on $(\mathcal{X}, \mathcal{A}, \mu)$ and Assumption 1 is automatically satisfied. This is in particular the case when $(\mathcal{X}, \mathcal{A}) = (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ with $k \geq 1$ and μ is the Lebesgue measure. If a family $\overline{\mathcal{M}}$ of densities satisfies our Assumption 1, so does any subset $\overline{\mathcal{D}}$ of $\overline{\mathcal{M}}$. The set $\overline{\mathcal{D}}$ may in turn be associated with a subset \mathcal{D} and a probability set $\mathcal{P} = \{P = p \cdot \mu, p \in \mathcal{D}\}$ that are both countable and respectively dense in $(\overline{\mathcal{D}}, \|\cdot\|)$ and $\overline{\mathcal{P}} = \{P = p \cdot \mu, p \in \overline{\mathcal{D}}\}$ for the total variation distance d . We may therefore write

$$\inf_{P \in \overline{\mathcal{P}}} d(P^*, P) = \inf_{P \in \mathcal{P}} d(P^*, P).$$

We shall repeatedly apply this equality to sets $\overline{\mathcal{D}}$ of interest without any further notice. As a consequence, replacing a density model $\overline{\mathcal{D}}$ by a countable and dense subset \mathcal{D} changes nothing from the approximation point of view. Nevertheless, we prefer to work with \mathcal{D} rather than $\overline{\mathcal{D}}$ in order to avoid some measurability issues that may result from the calculation of the supremum of an empirical process indexed by $\overline{\mathcal{D}}$.

Throughout the present paper, we use the same kind of notations as $\overline{\mathcal{D}}, \mathcal{D}, \mathcal{P}, \overline{\mathcal{P}}$ in order to distinguish between the density model, a countable and dense subset of it and their corresponding probability models. Following these notations $\mathcal{M} = \{P = p \cdot \mu, p \in \mathcal{M}\}$. An interval I of \mathbb{R} is said to be *nontrivial* if its interior $\overset{\circ}{I}$ is not empty or equivalently if its length is positive. When we say that p is a density on a (nontrivial) interval I , we mean that p is a density that vanishes outside I . The set of positive integers is denoted \mathbb{N}^* and $|A|$ is the cardinality of a set A . By convention, $\sum_{\emptyset} = 0$. For an integrable function f on $(\mathcal{X}^{\otimes n}, \mathcal{A}^{\otimes n})$, $\mathbb{E}[f(\mathbf{X})]$ is the integral of f with respect to the probability measure $\otimes_{i=1}^n P_i^* = \mathbf{P}^*$ while for f on $(\mathcal{X}, \mathcal{A})$ and $S \in \mathcal{P}$, $\mathbb{E}_S[f(X)]$ is the integral of f with respect to S . We use the same conventions for $\text{Var}(f(\mathbf{X}))$ and $\text{Var}_S(f(X))$. Finally, given a sequence of positive numbers $(v_n)_{n \geq 1}$ tending to 0 as n tends to infinity, we shall (abusively) say that an estimator converges at rate v_n if its rate of convergence is v_n or possibly faster than v_n .

3. AN ℓ -TYPE ESTIMATOR FOR THE TV-LOSS

Let $\overline{\mathcal{M}}$ be a density model that satisfies our Assumption 1 for some $\mathcal{M} \subset \overline{\mathcal{M}}$. Given $P = p \cdot \mu$ and $Q = q \cdot \mu$ in \mathcal{M} , we define

$$(2) \quad t_{(P,Q)} = \mathbb{1}_{q > p} - P(q > p) = P(p \geq q) - \mathbb{1}_{p \geq q}.$$

Given the family $\mathcal{T} = \{t_{(P,Q)}, (P,Q) \in \mathcal{M}^2\}$, we define for $P, Q \in \mathcal{M}$ and $\mathbf{x} \in \mathcal{X}^n$

$$(3) \quad \mathbf{T}(\mathbf{x}, P, Q) = \sum_{i=1}^n t_{(P,Q)}(x_i) = \sum_{i=1}^n [\mathbb{1}_{q > p}(x_i) - P(q > p)]$$

and

$$\mathbf{T}(\mathbf{x}, P) = \sup_{Q \in \mathcal{M}} \mathbf{T}(\mathbf{x}, P, Q).$$

For $\varepsilon > 0$, we finally define our estimator as any (measurable) element $\hat{P} = \hat{p} \cdot \mu$ that belongs to the set

$$(4) \quad \mathcal{E}(\mathbf{X}) = \left\{ P \in \mathcal{M}, \mathbf{T}(\mathbf{X}, P) \leq \inf_{P' \in \mathcal{M}} \mathbf{T}(\mathbf{X}, P') + \varepsilon \right\}.$$

We call \widehat{P} and \widehat{p} a *TV-estimator* on \mathcal{M} and \mathcal{M} respectively. The parameter ε is introduced in case a minimizer of $P \mapsto \mathbf{T}(\mathbf{X}, P)$ does not exist on \mathcal{M} . Any ε -minimizer would do provided that ε is not too large.

The construction of estimators from an appropriate family of test statistics $t_{(P,Q)}$ is described in Baraud [3] and our approach is in a similar vein. In particular, we use the following key property on the family \mathcal{T} (which can be compared to Assumption 1 in Baraud [3]).

Lemma 1. *For all probabilities $P, Q \in \mathcal{M}$ and $S \in \mathcal{P}$,*

$$(5) \quad d(P, Q) - d(S, Q) \leq \mathbb{E}_S [t_{(P,Q)}(X)] \leq d(S, P).$$

In particular,

$$(6) \quad d(P, Q) - d(P^*, Q) \leq \frac{1}{n} \sum_{i=1}^n \mathbb{E} [t_{(P,Q)}(X_i)] \leq d(P^*, P) \quad \text{where} \quad P^* = \frac{1}{n} \sum_{i=1}^n P_i^*.$$

However, our family \mathcal{T} does not satisfy the anti-symmetry assumption, namely $t_{(P,Q)} = -t_{(Q,P)}$, which is required for Baraud's construction. The risk bound that we establish below cannot therefore be deduced from Baraud [3]. In fact, for the specific problem we want to solve here the anti-symmetry assumption can be relaxed which leads to an improvement on the numerical constants that are involved in the risk bounds. Nevertheless, Baraud's approach and ours rely on similar heuristics. Denoting by \overline{P} the best approximation of P^* in \mathcal{M} with respect to the total variation distance, assuming that it exists for the sake of simplicity, we deduce from (6) that the random variable $\mathbf{T}(P) = \sup_{Q \in \mathcal{M}} \mathbb{E} [\mathbf{T}(\mathbf{X}, P, Q)]$ satisfies

$$d(P, \overline{P}) - d(P^*, \overline{P}) \leq \sup_{Q \in \mathcal{M}} [d(P, Q) - d(P^*, Q)] \leq \frac{\mathbf{T}(P)}{n} \leq d(P^*, P).$$

These inequalities imply that

$$\left| \frac{\mathbf{T}(P)}{n} - d(P, \overline{P}) \right| \leq (d(P^*, P) - d(P, \overline{P})) \vee d(P^*, \overline{P}) \leq d(P^*, \overline{P}) \quad \text{for all } P \in \mathcal{M}$$

which means that the mappings $P \mapsto \mathbf{T}(P)/n$ and $P \mapsto d(P, \overline{P})$ are uniformly close over \mathcal{M} when P^* is close enough to the model \mathcal{M} . We therefore expect that the minimizer of the mapping $P \mapsto d(P, \overline{P})$, which is \overline{P} , will be close to that of $P \mapsto \mathbf{T}(P)/n$ over \mathcal{M} . The procedure we propose here is to replace $\mathbf{T}(P)$ by its empirical counterpart $\mathbf{T}(\mathbf{X}, P) = \sup_{Q \in \mathcal{M}} \mathbf{T}(\mathbf{X}, P, Q)$.

Our construction also shares some similarities with that proposed by Devroye and Lugosi [11][Section 6.8 p.55]. However, a careful look at their selection criterion shows that it is slightly different from ours. They replace our function $\mathbf{T}(\cdot, P, Q)$ given by (3) by

$$\mathbf{T}_{\text{DL}}(\cdot, P, Q) : \mathbf{x} \mapsto \left| \sum_{i=1}^n [\mathbb{1}_{q \geq p}(x_i) - P(q \geq p)] \right|.$$

Their approach leads to a set of estimators $\mathcal{E}_{\text{DL}}(\mathbf{X})$ defined in the same way as (4) for \mathbf{T}_{DL} in place of \mathbf{T} and $\varepsilon = 1$.

Proof of Lemma 1. Let $P, Q \in \mathcal{M}$. Using the definition (2) of $t_{(P,Q)}$ and that of the TV-loss given by (1), we obtain that for all $S \in \mathcal{P}$,

$$\mathbb{E}_S [t_{(P,Q)}(X)] = S(q > p) - P(q > p) \leq d(S, P),$$

which is exactly the second inequality in (5). To establish the first one, we use the fact that $d(P, Q) = Q(q > p) - P(q > p)$. This leads to

$$\mathbb{E}_S [t_{(P, Q)}(X)] = S(q > p) - Q(q > p) + [Q(q > p) - P(q > p)] \geq -d(S, Q) + d(P, Q).$$

Finally, (6) results from the observation that

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E} [t_{(P, Q)}(X_i)] = \mathbb{E}_{P^*} [t_{(P, Q)}(X)].$$

□

3.1. Properties of the estimator. As for Baraud and Birgé [4], our main result is based on the key notion of *extremal point* in a model. However, in our situation, this notion is different from theirs. While we require a VC condition on a class of sets, they require a VC condition on a class of functions.

Definition 1. Let \mathcal{F} be a class of real-valued functions on a set \mathcal{X} with values in \mathbb{R} . We say that an element $\bar{f} \in \mathcal{F}$ is *extremal* in \mathcal{F} (or is an *extremal point* of \mathcal{F}) with degree not larger than $\mathfrak{D} \geq 1$ if the classes of subsets

$$\mathcal{E}^>(\mathcal{F}, \bar{f}) = \{\{x \in \mathcal{X} \mid q(x) > \bar{f}(x)\}, q \in \mathcal{F} \setminus \{\bar{f}\}\}$$

and

$$\mathcal{E}^<(\mathcal{F}, \bar{f}) = \{\{x \in \mathcal{X} \mid q(x) < \bar{f}(x)\}, q \in \mathcal{F} \setminus \{\bar{f}\}\}$$

are both VC with dimension not larger than \mathfrak{D} .

Additionally, we say that \bar{P} is an *extremal point* of $\overline{\mathcal{M}}$ with degree not larger than $\mathfrak{D} \geq 1$ if there exists $\bar{p} \in \overline{\mathcal{M}}$ such that $\bar{P} = \bar{p} \cdot \mu$ and \bar{p} is extremal in $\overline{\mathcal{M}}$ with degree not larger than \mathfrak{D} . For each $\mathfrak{D} \geq 1$, we denote by $\overline{\mathcal{O}}(\mathfrak{D})$ the set of extremal points \bar{p} in $\overline{\mathcal{M}}$ with degree not larger than \mathfrak{D} , $\mathcal{O}(\mathfrak{D})$ a countable and dense subset of it, $\mathcal{O}(\mathfrak{D})$ the corresponding set of probability measures and $\overline{\mathcal{O}} = \bigcup_{\mathfrak{D} \geq 1} \overline{\mathcal{O}}(\mathfrak{D})$ the set of all extremal points in $\overline{\mathcal{M}}$. Finally, let \mathcal{M} be a countable and dense subset of $\overline{\mathcal{M}}$ containing $\bigcup_{\mathfrak{D} \geq 1} \mathcal{O}(\mathfrak{D})$.

Theorem 1. *Let $\overline{\mathcal{M}}$ be a density model that satisfies our Assumption 1 and assume that the set $\overline{\mathcal{O}}$ is nonempty. Then any TV-estimator \hat{P} on \mathcal{M} satisfies the following properties. For all product distributions P^* of the data and for all $\xi > 0$, with a probability at least $1 - e^{-\xi}$*

$$(7) \quad d(P, \hat{P}) \leq 2d(P^*, P) + 20\sqrt{\frac{5\mathfrak{D}}{n}} + \sqrt{\frac{2(\xi + \log 2)}{n}} + \frac{\varepsilon}{n} \quad \text{for all } \mathfrak{D} \geq 1 \text{ and } P \in \overline{\mathcal{O}}(\mathfrak{D}).$$

In particular,

$$(8) \quad d(P^*, \hat{P}) \leq \inf_{\mathfrak{D} \geq 1} \left[3 \inf_{P \in \overline{\mathcal{O}}(\mathfrak{D})} d(P^*, P) + 20\sqrt{\frac{5\mathfrak{D}}{n}} \right] + \sqrt{\frac{2(\log 2 + \xi)}{n}} + \frac{\varepsilon}{n},$$

with the convention $\inf_{\emptyset} = +\infty$. As a consequence of (7),

$$(9) \quad \mathbb{E} \left[d(P, \hat{P}) \right] \leq 2d(P^*, P) + 48\sqrt{\frac{\mathfrak{D}}{n}} + \frac{\varepsilon}{n} \quad \text{for all } \mathfrak{D} \geq 1 \text{ and } P \in \overline{\mathcal{O}}(\mathfrak{D}).$$

Moreover by (8),

$$(10) \quad \mathbb{E} \left[d(P^*, \hat{P}) \right] \leq \inf_{\mathfrak{D} \geq 1} \left\{ 3 \inf_{P \in \overline{\mathcal{O}}(\mathfrak{D})} d(P^*, P) + 48\sqrt{\frac{\mathfrak{D}}{n}} \right\} + \frac{\varepsilon}{n}.$$

Proof. The proof is postponed to Subsection 8.2. □

Let us now comment on this result. In the favourable situation where the X_i are i.i.d. with distribution P^* in $\overline{\mathcal{O}}$, $\inf_{\mathfrak{D} \geq 1} \inf_{P \in \overline{\mathcal{O}}(\mathfrak{D})} d(P^*, P) = 0$ and we deduce from (10) that the estimator \widehat{P} converges toward P^* at the rate $1/\sqrt{n}$ for the total variation distance. More precisely, the risk of the estimator is not larger than $48\sqrt{\mathfrak{D}/n} + \varepsilon/n$ when P^* belongs to $\overline{\mathcal{O}}(\mathfrak{D})$ for some $\mathfrak{D} \geq 1$. Note that the result also holds when the data are independent only, provided that $P^* = n^{-1} \sum_{i=1}^n P_i^*$ is extremal. As we shall see, this situation may arise even when none of the marginals P_i^* is extremal or even belongs to the model \mathcal{M} . In the general case where the data are independent only and for all $i \in \{1, \dots, n\}$, their marginals are of the form

$$(11) \quad P_i^* = (1 - \alpha_i)\overline{P} + \alpha_i R_i = \overline{P} + \alpha_i (R_i - \overline{P})$$

for some $\overline{P} \in \overline{\mathcal{O}}(\mathfrak{D})$ with $\mathfrak{D} \geq 1$, $\alpha_1, \dots, \alpha_n$ in $[0, 1]$ and distributions R_1, \dots, R_n in \mathcal{P} , we deduce from (9) that

$$\mathbb{E} \left[d(\overline{P}, \widehat{P}) \right] \leq 2d(P^*, \overline{P}) + 48\sqrt{\frac{\mathfrak{D}}{n}} + \frac{\varepsilon}{n} \leq \frac{2}{n} \sum_{i=1}^n \alpha_i + 48\sqrt{\frac{\mathfrak{D}}{n}} + \frac{\varepsilon}{n}.$$

As compared to the previous situation where $P_i^* = \overline{P} \in \overline{\mathcal{O}}(\mathfrak{D})$, hence $\alpha_i = 0$ for all i , we see that the risk bound we get only inflates by the additional term $2\overline{\alpha} = (2/n) \sum_{i=1}^n \alpha_i$ and it remains thus of the same order when $\overline{\alpha}$ is small enough as compared to $\sqrt{\mathfrak{D}/n}$. Note that this situation may occur even when $\alpha_i > 0$ for all i , i.e. when possibly none of the marginals P_i^* belong to $\overline{\mathcal{O}}(\mathfrak{D})$. In order to be more specific, we may consider the two following situations. In the first one, there exists some subset of the data which are i.i.d. with distribution $\overline{P} \in \overline{\mathcal{O}}(\mathfrak{D})$ while the other part, corresponding to what we shall call *outliers*, are independently drawn according to some arbitrary distributions. In this case, there exists a subset $S \subset \{1, \dots, n\}$ such that $\alpha_i = 1$ for $i \in S$ and $\alpha_i = 0$ otherwise in (11). Our procedure is stable with respect to the presence of such outliers as soon as $\overline{\alpha} = |S|/n$ remains small as compared to $\sqrt{\mathfrak{D}/n}$. In the other situation, which is called the *contamination* case, the data are i.i.d., a portion $\alpha \in (0, 1]$ of them are drawn according to an arbitrary distribution R while the other part follows the distribution $\overline{P} \in \overline{\mathcal{O}}(\mathfrak{D})$. Then (11) holds with $\alpha_i = \alpha$ and $R_i = R$ for all $i \in \{1, \dots, n\}$. The risk bound we get remains stable under contamination as long as the level $\overline{\alpha} = \alpha$ of contamination remains small as compared to $\sqrt{\mathfrak{D}/n}$.

A bound similar to (8) has been established in Baraud [3] for his ℓ -estimators. His inequality (48) can be reformulated in our context as

$$\frac{1}{n} \sum_{i=1}^n d(P_i^*, \widehat{P}) \leq \inf_{\mathfrak{D} \geq 1} \left[6 \inf_{P \in \overline{\mathcal{O}}(\mathfrak{D})} \left[\frac{1}{n} \sum_{i=1}^n d(P_i^*, P) \right] + 40\sqrt{\frac{5\mathfrak{D}}{n}} + 2\sqrt{\frac{2\xi}{n}} + \frac{2\varepsilon}{n} - \inf_{P \in \overline{\mathcal{M}}} \left[\frac{1}{n} \sum_{i=1}^n d(P_i^*, P) \right] \right].$$

In comparison, (7) together with the inequalities

$$\frac{1}{n} \sum_{i=1}^n d(P_i^*, \widehat{P}) \leq \frac{1}{n} \sum_{i=1}^n d(P_i^*, P) + d(P, \widehat{P}) \quad \text{and} \quad d(P^*, P) \leq \frac{1}{n} \sum_{i=1}^n d(P_i^*, P)$$

which hold for all $P \in \overline{\mathcal{O}}$, imply that

$$\frac{1}{n} \sum_{i=1}^n d(P_i^*, \widehat{P}) \leq \inf_{\mathfrak{D} \geq 1} \left[3 \inf_{P \in \overline{\mathcal{O}}(\mathfrak{D})} \left[\frac{1}{n} \sum_{i=1}^n d(P_i^*, P) \right] + 20\sqrt{\frac{5\mathfrak{D}}{n}} + \sqrt{\frac{2(\xi + \log 2)}{n}} + \frac{\varepsilon}{n} \right].$$

If we omit the term $\inf_{P \in \overline{\mathcal{M}}} [(1/n) \sum_{i=1}^n d(P_i^*, P)]$ that appears in his inequality and $\log 2$ that appears in ours, all the constants we get are divided by a factor 2 as compared to his.

When the Yatracos class $\{p > q\}, p, q \in \overline{\mathcal{M}}$ is VC with dimension not larger than $\mathfrak{D} \geq 1$, all the elements of $\overline{\mathcal{M}}$ are extremal with degree not larger than \mathfrak{D} and (10) becomes

$$(12) \quad \mathbb{E} \left[d(P^*, \hat{P}) \right] \leq 3 \inf_{P \in \overline{\mathcal{M}}} d(P^*, P) + 48 \sqrt{\frac{\mathfrak{D}}{n}} + \frac{\varepsilon}{n}.$$

In the special case where the data are i.i.d. with distribution P^* , an inequality of the same flavour was established by Devroye and Lugosi [11][Section 8.2] for their minimum distance estimate. Both inequalities involve a constant 3 in front of the approximation term $\inf_{P \in \overline{\mathcal{M}}} d(P^*, P)$. In our inequality all the numerical constants are explicit.

In the next sections, we take advantage of the stronger inequality (8) to consider density models $\overline{\mathcal{M}}$ for which the Yatracos classes $\{p > q\}, p, q \in \overline{\mathcal{M}}$ are not VC.

3.2. First examples based on geometric properties. The way we have characterized extremal elements in our Definition 1 is combinatorial in nature. We provide here a sufficient condition which is rather geometric. To do so, we start with some notations. Given a positive integer k , an element $\mathbf{x} = (x_1, \dots, x_k) \in \mathcal{X}^k$ and $f \in \mathcal{F}$, we denote by $f_{\mathbf{x}}$ and $\mathcal{F}_{\mathbf{x}}$ the images of f and \mathcal{F} respectively by the mapping $f \mapsto (f(x_1), \dots, f(x_k))$ with values in \mathbb{R}^k . The vectors $f_{\mathbf{x}}$ and the subsets $\mathcal{F}_{\mathbf{x}} \subset \mathbb{R}^k$ can be viewed as finite-dimensional projections of f and \mathcal{F} on \mathbb{R}^k . We recall that the border $\partial\mathcal{F}_{\mathbf{x}}$ of $\mathcal{F}_{\mathbf{x}}$ is the set of elements that belong to the closure of $\mathcal{F}_{\mathbf{x}}$ but not to its interior (with respect to topology induced by the Euclidean distance in \mathbb{R}^k). Given an element $\bar{f} \in \mathcal{F}$, a sufficient condition for the projection $\bar{f}_{\mathbf{x}}$ to belong to $\partial\mathcal{F}_{\mathbf{x}}$ is that there exists a hyperplane through $\bar{f}_{\mathbf{x}}$ that supports $\mathcal{F}_{\mathbf{x}}$. More precisely, this means that we can find a vector $\mathbf{u}(\bar{f}_{\mathbf{x}}) = (u_1(\bar{f}_{\mathbf{x}}), \dots, u_k(\bar{f}_{\mathbf{x}})) \in \mathbb{R}^k \setminus \{\mathbf{0}\}$ which is orthogonal to an hyperplane that contains $\bar{f}_{\mathbf{x}}$ and which satisfies

$$(13) \quad \langle \mathbf{u}(\bar{f}_{\mathbf{x}}), f_{\mathbf{x}} - \bar{f}_{\mathbf{x}} \rangle = \sum_{i=1}^k u_i(\bar{f}_{\mathbf{x}}) (f(x_i) - \bar{f}(x_i)) \leq 0 \quad \text{for all } f \in \mathcal{F}.$$

When \mathcal{F} is convex, so is $\mathcal{F}_{\mathbf{x}} \subset \mathbb{R}^k$, and (13) is actually equivalent to the fact that $\bar{f}_{\mathbf{x}} \in \partial\mathcal{F}_{\mathbf{x}}$. Finally, we say that a point $\mathbf{x} \in \mathcal{X}^k$, with $k \geq 2$, is \mathcal{F} -separated if for all $i, j \in \{1, \dots, k\}$ with $i \neq j$, there exists $f \in \mathcal{F}$ such that $f(x_i) \neq f(x_j)$. The following result holds.

Proposition 1. *Let \mathcal{F} be a class of functions on a set \mathcal{X} , \bar{f} an element of \mathcal{F} and D a positive integer. If for all \mathcal{F} -separated $\mathbf{x} \in \mathcal{X}^{D+1}$, (13) is satisfied for a vector $\mathbf{u}(\bar{f}_{\mathbf{x}}) \in \mathbb{R}^{D+1} \setminus \{\mathbf{0}\}$ whose coordinates are not all of the same sign, the element \bar{f} is extremal in \mathcal{F} with degree not larger than D .*

In the above condition, 0 is assumed to have the same sign as any other number. When \mathcal{F} is convex, Proposition 1 tells us that \bar{f} is extremal in \mathcal{F} with degree not larger than $\mathfrak{D} = D$ if any of its $(D+1)$ -projections $\bar{f}_{\mathbf{x}}$ belong to the border of $\mathcal{F}_{\mathbf{x}}$ and for each $\mathbf{x} \in \mathbb{R}^{D+1}$ one can find a vector $\mathbf{u} = \mathbf{u}(\bar{f}_{\mathbf{x}})$ whose coordinates are not all of the same sign such that the affine hyperplane $\{\bar{f}_{\mathbf{x}} + \mathbf{v} \mid \mathbf{v} \in \mathbb{R}^{D+1}, \langle \mathbf{v}, \mathbf{u} \rangle = 0\}$ supports $\mathcal{F}_{\mathbf{x}}$. It is actually enough to check that the condition holds for these $\mathbf{x} \in \mathbb{R}^{D+1}$ which are \mathcal{F} -separated since for those which are not, i.e. there exist $i \neq j$ with $f(x_i) = f(x_j)$ for all $f \in \mathcal{F}$, the set $\mathcal{F}_{\mathbf{x}}$ is contained in the hyperplane $\{\mathbf{y} \in \mathbb{R}^{D+1}, y_i - y_j = 0\}$ so that $\mathcal{F}_{\mathbf{x}} \subset \partial\mathcal{F}_{\mathbf{x}}$ and $\bar{f}_{\mathbf{x}}$ automatically belongs to the border of $\mathcal{F}_{\mathbf{x}}$.

The assumptions of Proposition 1 implies, in particular, Property (P1): any $(D+1)$ -dimensional projection $\bar{f}_{\mathbf{x}}$ of \bar{f} lies at the border of the projection $\mathcal{F}_{\mathbf{x}}$ of \mathcal{F} . Being extremal is, however, not equivalent to (P1) nor is it equivalent to being an extreme point when \mathcal{F} is convex. For example, if \mathcal{F} is the (convex) set of nonincreasing functions on $\mathcal{X} = [0, 2]$, the function $\bar{f} = (3/4)\mathbb{1}_{[0,1]} + (1/4)\mathbb{1}_{(1,2]}$ is extremal, as we shall see, but not extreme in \mathcal{F} since it can also be written as the convex combination $(1/2) [\mathbb{1}_{[0,1]} + (1/2)\mathbb{1}_{[0,2]}$ where both functions $\mathbb{1}_{[0,1]}$ and $(1/2)\mathbb{1}_{[0,2]}$ belong to \mathcal{F} . To prove in this example that \bar{f} is extremal with

degree not larger than $\mathfrak{D} = 2$, it suffices to argue as follows. Given three distinct points $x_1 < x_2 < x_3$, at least two consecutive ones, say $z_1 < z_2$, satisfies $\bar{f}(z_1) = \bar{f}(z_2)$ so that $(f - \bar{f})(z_2) - (f - \bar{f})(z_1) \leq 0$ for all $f \in \mathcal{F}$. Inequality (13) is therefore satisfied with $\mathbf{u}(\bar{f}_{\mathbf{x}}) = (-1, +1, 0)$ or $\mathbf{u}(\bar{f}_{\mathbf{x}}) = (0, -1, +1)$, depending on $z_1, z_2 \in \{x_1, x_2, x_3\}$. Proposition 1 then applies and we conclude that \bar{f} is extremal with degree not larger than 2.

Conversely, an extreme element of a convex set of functions, or more generally an element $\bar{f} \in \mathcal{F}$ that only satisfies (P1), might not be extremal. If \mathcal{F} is the class of nonnegative functions on \mathbb{R} , the function \bar{f} which is identically equal to 0 is extreme in the convex set \mathcal{F} and clearly satisfies (P1). However, \bar{f} is not extremal in \mathcal{F} since the class $\mathcal{C}^>(\mathcal{F}, \bar{f}) = \{\{f > 0\}, f \in \mathcal{F}\}$, which is that of all the subsets of \mathbb{R} (take $f = \mathbf{1}_B$ with $B \subset \mathbb{R}$), is not VC. In this example, we also note that for such a function \bar{f} inequality (13) is satisfied for all $\mathbf{x} \in \mathcal{X}^2$ with $\mathbf{u}(\bar{f}_{\mathbf{x}}) = (-1, -1)$ although \bar{f} is not extremal with degree not larger than 1. This shows in passing that our assumption on the signs of the coordinates of $\mathbf{u}(\bar{f}_{\mathbf{x}})$ in Proposition 1 is necessary to guarantee the fact \bar{f} is extremal.

Proof. Let us set $k = D + 1$. According to Definition 1, if \bar{f} were not extremal in \mathcal{F} with degree not larger than $\mathfrak{D} = D$, there would exist a subset $\{x_1, \dots, x_k\}$ of \mathcal{X} that can be shattered by at least one of the classes of sets $\mathcal{C}^>(\mathcal{F}, \bar{f})$ and $\mathcal{C}^<(\mathcal{F}, \bar{f})$. We only consider the case where this class is $\mathcal{C}^>(\mathcal{F}, \bar{f})$ since the proof is the similar for $\mathcal{C}^<(\mathcal{F}, \bar{f})$. By definition, $\mathcal{C}^>(\mathcal{F}, \bar{f})$ shatters $\{x_1, \dots, x_k\}$ if for each subset $I \subset \{1, \dots, k\}$, there exists $f_I \in \mathcal{F}$ such that

$$I = \{i \in \{1, \dots, k\}, f_I(x_i) - \bar{f}(x_i) > 0\}.$$

In particular, the point $\mathbf{x} = (x_1, \dots, x_k)$ is necessarily \mathcal{F} -separated. Since (13) is satisfied for this choice of \mathbf{x} and the set $I^* = \{i \in \{1, \dots, k\}, u_i(\bar{f}_{\mathbf{x}}) > 0\}$ is nonempty, we may write that

$$\sum_{i \in I^*} (f_{I^*}(x_i) - \bar{f}(x_i)) u_i(\bar{f}_{\mathbf{x}}) \leq - \sum_{i \notin I^*} (f_{I^*}(x_i) - \bar{f}(x_i)) u_i(\bar{f}_{\mathbf{x}})$$

with the convention $\sum_{\emptyset} = 0$. By definition of I^* and f_{I^*} , the right-hand side of this inequality is nonpositive while the left-hand side is positive. This leads to a contradiction and shows that $\mathcal{C}^>(\mathcal{F}, \bar{f})$ cannot shatter any subset of \mathcal{X} with cardinality larger than D . \square

Let us now take some generic example to illustrate Proposition 1. Let r be a positive integer, \mathcal{X} an open subset of \mathbb{R}^l , $l \geq 1$, g a function on \mathcal{X} with values in a (nontrivial) interval $I \subset \mathbb{R}$, and Φ_r the class of functions ϕ defined on I which satisfy for all $\mathbf{v} = (v_1, \dots, v_{r+1}) \in I^{r+1}$ such $v_1 < \dots < v_{r+1}$,

$$(14) \quad L_{\mathbf{v}}(\phi) = \begin{vmatrix} 1 & v_1 & \dots & v_1^{r-1} & \phi(v_1) \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 1 & v_{r+1} & \dots & v_{r+1}^{r-1} & \phi(v_{r+1}) \end{vmatrix} \geq 0.$$

For example Φ_1 is the class of nondecreasing functions on I while Φ_2 is the class of convex ones. Let us now define $\mathcal{F} = \{f = \phi \circ g, \phi \in \Phi_r\}$ and for $d \geq 1$, let \mathcal{F}_d be the subset of \mathcal{F} which consists of these functions \bar{f} of the form $\bar{\phi} \circ g$ where $\bar{\phi} \in \Phi_r$ satisfies the following property: the restrictions of $\bar{\phi}$ to the elements of a partition of I into d nontrivial intervals coincide with a polynomial of degree at most $r - 1$. We shall see now that a consequence of Proposition 1 is that the elements of \mathcal{F}_d are all extremal in \mathcal{F} with degrees not larger than $\mathfrak{D} = rd \geq 1$. To prove this, it suffices to prove that the requirements of the proposition are satisfied, which we shall now do. Let $k = rd + 1$, $\bar{f} = \bar{\phi} \circ g \in \mathcal{F}_d$ and $\mathbf{x} \in \mathcal{X}^k$ be a \mathcal{F} -separable point. In particular, $g(x_1), \dots, g(x_k)$ are distinct and we may relabel the coordinates of \mathbf{x} , if ever necessary, in such a way that $v_1 = g(x_1) < \dots < v_k = g(x_k)$. Since $\bar{\phi}$ is based on a partition of I into d intervals and $k = rd + 1$, one of these intervals contains at least $r + 1$ (consecutive) numbers v_i , i.e. there exists $j \in \{1, \dots, k - r\}$ for which

$\bar{\phi}$ is a polynomial of degree at most $r - 1$ on the interval $[v_j, v_{j+r}] \subset I$. Setting $\mathbf{v} = (v_j, \dots, v_{j+r}) \in I^{r+1}$, we deduce from (14), the linearity of $L_{\mathbf{v}}$ and the fact that $L_{\mathbf{v}}(\bar{\phi}) = 0$ that for all $f = \phi \circ g \in \mathcal{F}$

$$\begin{aligned} 0 \leq L_{\mathbf{v}}(\phi - \bar{\phi}) &= (-1)^r \begin{vmatrix} 1 & v_{j+1} & \cdots & v_{j+1}^{r-1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & v_{j+r} & \cdots & v_{j+r}^{r-1} \end{vmatrix} (\phi - \bar{\phi})(v_j) + \cdots + \begin{vmatrix} 1 & v_j & \cdots & v_j^{r-1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & v_{j+r-1} & \cdots & v_{j+r-1}^{r-1} \end{vmatrix} (\phi - \bar{\phi})(v_{j+r}) \\ &= (-1)^r \begin{vmatrix} 1 & v_{j+1} & \cdots & v_{j+1}^{r-1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & v_{j+r} & \cdots & v_{j+r}^{r-1} \end{vmatrix} (f - \bar{f})(x_j) + \cdots + \begin{vmatrix} 1 & v_j & \cdots & v_j^{r-1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & v_{j+r-1} & \cdots & v_{j+r-1}^{r-1} \end{vmatrix} (f - \bar{f})(x_{j+r}). \end{aligned}$$

Since $v_j < \dots < v_{j+r}$, these Vandermonde determinants are all positive, $-L_{\mathbf{v}}(\phi - \bar{\phi})$ can be written in the form (13) where the coordinates of $\mathbf{u}(\bar{f}_{\mathbf{x}})$ are all zero except for those with indices $j, \dots, j+r$. Besides, the signs of these nonzero coordinates are not constant. The requirements of Proposition 1 are satisfied and the elements of \mathcal{F}_d are therefore extremal with degree at most rd .

Let us now see some of the consequences of this result and some preliminary applications of our Theorem 1. When μ is the Lebesgue measure on \mathcal{X} , $\mathbb{L}_1(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mu)$ is separable, Assumption 1 is therefore automatically satisfied and our Theorem 1 applies to the subset $\overline{\mathcal{M}}$ of \mathcal{F} that consists of densities. The set $\overline{\mathcal{O}} \subset \overline{\mathcal{M}}$ that consists of those extremal ones therefore contains $\bigcup_{d \geq 1} (\mathcal{F}_d \cap \overline{\mathcal{M}})$. For example, in dimension 1, when $r = 1$, $\mathcal{X} = (0, +\infty)$, $I = (-\infty, 0)$ and $g : x \mapsto -x$, the set $\overline{\mathcal{M}}$ is that of all nonincreasing densities on \mathcal{X} and $\overline{\mathcal{O}}$ contains those which are piecewise constant on a partition of \mathcal{X} into a finite number of (nontrivial) intervals. When $r = 2$, $I = \mathcal{X}$ is a (nontrivial) interval of \mathbb{R} and g the identity function on I , the set $\overline{\mathcal{M}}$ consists of the convex densities on \mathcal{X} and $\overline{\mathcal{O}}$ contains those which are piecewise affine on a partition of \mathcal{X} into a finite number of (nontrivial) intervals. Denoting by $\langle \cdot, \cdot \rangle$ the Euclidean inner product of \mathbb{R}^l and $|\cdot|$ the corresponding norm, our example also includes in higher dimensions, models of densities on a compact domain of \mathbb{R}^l which are of the form $\mathbf{y} \mapsto \phi(\langle \mathbf{y}, \mathbf{t} \rangle)$, for a given \mathbf{t} in the unit sphere of \mathbb{R}^l , or of the form $\mathbf{y} \mapsto \phi(|A\mathbf{y}|)$ for a given $(l \times l)$ -matrix A , taking for g the mappings $\mathbf{y} \mapsto \langle \mathbf{y}, \mathbf{t} \rangle$ and $\mathbf{y} \mapsto A\mathbf{y}$ respectively and functions ϕ that run among a set of monotone, convex or concave functions on the line.

Nevertheless, these examples may appear too restrictive for practical purposes, especially in dimensions larger than 1. These restrictions are due to the fact that the geometric-type assumption required in Proposition 1 is fulfilled for specific convex sets only although we typically wish to consider density models which are not convex. Fortunately, this assumption is sufficient but not necessary and our Definition 1 allows us to find extremal elements in sets that are possibly non-convex. The most interesting ones, nevertheless, arise in dimension $l = 1$. In the remaining part of this paper we shall therefore restrict ourselves to the problem of estimating a density (with respect to the *Lebesgue measure*) on the line. Each of the forthcoming sections is devoted to the problem of estimating densities or a mixture of densities that are monotone, concave, convex, log-concave or more generally s -concave. None of these density models is convex.

4. ESTIMATING A PIECEWISE MONOTONE DENSITY

We denote by $\mathcal{A}(k)$ the class of nonempty subsets $A \subset \mathbb{R}$ with cardinality not larger than $k \geq 1$. The elements of A provide a partition of \mathbb{R} into $l = |A| + 1 \leq k + 1$ intervals I_1, \dots, I_l the endpoints of which belong to A . We denote by $\mathbf{I}(A)$ the set $\{\overset{\circ}{I}_1, \dots, \overset{\circ}{I}_l\}$ of their interiors. Although there exist several ways of partitioning \mathbb{R} into intervals with endpoints in A , the set $\mathbf{I}(A)$ is uniquely defined.

4.1. Piecewise monotone densities.

Definition 2. Let $k \geq 2$. A function g on \mathbb{R} is said to be k -piecewise monotone if there exists $A \in \mathcal{A}(k-1)$ such that the restriction of g to each interval $I \in \mathbf{I}(A)$ is monotone. In particular, there exist at most k monotone functions g_I on $I \in \mathbf{I}(A)$ such that

$$g(x) = \sum_{I \in \mathbf{I}(A)} g_I(x) \mathbb{1}_I(x) \quad \text{for all } x \in \mathbb{R} \setminus A.$$

A k -piecewise monotone function g associated to $A \in \mathcal{A}(k-1)$ may not be monotone on each element of a partition based on A . We only require that g be monotone on the interiors of these elements. The function

$$g : \begin{cases} \mathbb{R} & \rightarrow \mathbb{R}_+ \\ x & \mapsto \frac{1}{\sqrt{|x|}} \mathbb{1}_{|x|>0} \end{cases}$$

is 2-piecewise monotone, associated to $A = \{0\} \in \mathcal{A}(1)$, but g is neither monotone on $(-\infty, 0]$ nor on $[0, +\infty)$.

We denote by $\overline{\mathcal{M}}_k$ the set of k -piecewise monotone densities. These sets obviously satisfy $\overline{\mathcal{M}}_j \subset \overline{\mathcal{M}}_k$ for $j \leq k$ and form thus an increasing sequence with respect to set inclusion. The set $\overline{\mathcal{M}}_2$ contains the unimodal densities on the line and in particular all the densities that are monotone on a half-line and vanish elsewhere.

Of special interest are those densities in $\overline{\mathcal{M}}_k$ which are also piecewise constant on a finite partition of \mathbb{R} into intervals. More precisely, for $D \geq 1$ let $\overline{\mathcal{O}}_{D,k}$ be the subset of $\overline{\mathcal{M}}_k$ that consists of those densities that are constant on each element of a class $\mathbf{I}(A)$ with $A \in \mathcal{A}(D+1)$. The number D is a majorant of the number of bounded intervals in the class $\mathbf{I}(A)$, hence on the number of positive values that a density in $\overline{\mathcal{O}}_{D,k}$ may take. The uniform distribution on a nontrivial interval has a density that belongs to $\overline{\mathcal{O}}_{1,2}$, and also to $\overline{\mathcal{O}}_{D,k}$ for all $D \geq 1$ and $k \geq 2$. The sets $\overline{\mathcal{O}}_{D,k}$ with $D \geq 1$ and $k \geq 2$ are therefore nonempty. They satisfy the following property which is a consequence of Proposition 3 of Baraud and Birgé [4].

Proposition 2. *For all $D \geq 1$ and $k \geq 2$, all the elements of $\overline{\mathcal{O}}_{D,k}$ are extremal in $\overline{\mathcal{M}}_k$ with degree not larger than $3(k+D+1)$.*

For all $D \geq 1$ and $k \geq 2$, let $\mathcal{O}_{D,k}$ be a countable and dense subset of $\overline{\mathcal{O}}_{D,k}$ (for the \mathbb{L}_1 -norm) and \mathcal{M}_k a countable and dense subset of $\overline{\mathcal{M}}_k$ that contains $\bigcup_{D \geq 1} \mathcal{O}_{D,k}$. It follows from Proposition 2 that the elements of $\mathcal{O}_{D,k}$ are also extremal in \mathcal{M}_k with degree not larger than $3(k+D+1)$ for all $D \geq 1$. We immediately deduce from Theorem 1 the following result.

Theorem 2. *Let $k \geq 2$. For every product distribution \mathbf{P}^* of the data, a TV-estimator \hat{P} on \mathcal{M}_k satisfies*

$$(15) \quad \mathbb{E} \left[d(P^*, \hat{P}) \right] \leq \inf_{D \geq 1} \left[3 \inf_{P \in \overline{\mathcal{O}}_{D,k}} d(P^*, P) + 83.2 \sqrt{\frac{D+k+1}{n}} \right] + \frac{\varepsilon}{n}.$$

Our approach solves the problem of estimating a nonincreasing density on a half-line I by taking $k = 2$. For this specific problem, a bound of the same flavour was established by Baraud [3] (see Proposition 6) for his ℓ -estimator. When the data are i.i.d. with distribution P^* , our result shows better constants in the risk bound (15) as compared to Baraud's. In particular, the constant 3 in front of the approximation term $\inf_{P \in \overline{\mathcal{O}}_{D,k}} d(P^*, P)$ improves on his constant 5. Our estimator (as well as Baraud's) also improves on the Grenander estimator since our construction does not require the prior knowledge of the half-line I .

For general values of $k \geq 2$, the problem of estimating a k -piecewise monotone density was also considered in Baraud and Birgé [4]. The authors used a ρ -estimator and the squared Hellinger loss in place of the total variation one to evaluate its risk—see their Corollary 2. Their bound is similar to ours except for the fact that the quantity $\sqrt{(D+k+1)/n}$ appears there multiplied by a logarithmic factor. This logarithmic factor turns out to be necessary when one deals with the squared Hellinger loss while it disappears with the total variation one.

Up to the additional term ε/n , the risk bound that we establish for the TV-estimator is given by the quantity

$$A = \inf_{D \geq 1} \left[3 \inf_{P \in \overline{\mathcal{O}}_{D,k}} d(P^*, P) + 83.2 \sqrt{\frac{D+k+1}{n}} \right].$$

If we apply a location-scale transformation to the data X_1, \dots, X_n , that is, if in place of the original data X_1, \dots, X_n we observe the random variables $Y_i = \sigma X_i + m$ for $i \in \{1, \dots, n\}$ with $(m, \sigma) \in \mathbb{R} \times (0, +\infty)$, the marginal distributions P_i^* of the X_i are then replaced by their images Q_i^* by the mapping $x \mapsto \sigma x + m$. This mapping leaves the sets $\overline{\mathcal{O}}_{D,k}$ and the total variation distance between two probabilities invariant. The quantity A remains thus unchanged under such a transformation and the performance of the TV-estimator is therefore independent of the unit that is used to measure the data.

Another consequence of Theorem 2 is the following one. If the X_i are not i.i.d. but only independent and their distributions are close enough to a probability \overline{P} which admits a density \overline{p} with respect to μ , by using the triangle inequality we may bound A by $3d(P^*, \overline{P}) + \mathbb{B}_{k,n}(\overline{p})$ where $\mathbb{B}_{k,n}(p)$ is defined for a density p on the line by

$$\begin{aligned} \mathbb{B}_{k,n}(p) &= \inf_{D \geq 1} \left[3 \inf_{Q \in \overline{\mathcal{O}}_{D,k}} d(P, Q) + 83.2 \sqrt{\frac{D+k+1}{n}} \right] \\ (16) \quad &= \inf_{D \geq 1} \left[\frac{3}{2} \inf_{q \in \overline{\mathcal{O}}_{D,k}} \|p - q\| + 83.2 \sqrt{\frac{D+k+1}{n}} \right]. \end{aligned}$$

This means that as long as $3d(P^*, \overline{P})$ remains small as compared to $\mathbb{B}_{k,n}(\overline{p})$, the bound on $\mathbb{E}[d(P^*, \widehat{P})]$ would be almost the same as if the X_i were truly i.i.d. with distribution \overline{P} . This property accounts for the robustness of our approach.

In the remaining part of this section, our aim is to specify the order of magnitude of the quantity $\mathbb{B}_{k,n}(p)$ under different *a posteriori* assumptions on the density p . Note that the quantity $\mathbb{B}_{k,n}(p)$ also remains invariant under a location-scale transformation of the density p .

4.2. Estimation of bounded and compactly supported k -piecewise monotone densities. For $k \geq 3$, let $\overline{\mathcal{M}}_k^\infty$ be the subset of $\overline{\mathcal{M}}_k$ that consists of the densities on \mathbb{R} which coincide almost everywhere with a density of the form

$$(17) \quad p = \sum_{i=1}^{k-2} w_i p_i \mathbb{1}_{(x_{i-1}, x_i)}$$

where

- (i) $(x_i)_{i \in \{0, \dots, k-2\}}$ is an increasing sequence of real numbers;
- (ii) w_1, \dots, w_{k-2} are nonnegative numbers such that $\sum_{i=1}^{k-2} w_i = 1$;
- (iii) for $i \in \{1, \dots, k-2\}$, p_i is a monotone density on the interval $I_i = (x_{i-1}, x_i)$ of length $L_i > 0$ with variation

$$V_i = \sup_{x \in I_i} p_i(x) - \inf_{x \in I_i} p_i(x) < +\infty.$$

A density p in $\overline{\mathcal{M}}_k^\infty$ is necessarily bounded and compactly supported. A monotone density on \mathbb{R}_+ which is bounded and compactly supported belongs to $\overline{\mathcal{M}}_3^\infty$. A bounded unimodal density supported on a compact interval belongs to $\overline{\mathcal{M}}_4^\infty$.

For $p \in \overline{\mathcal{M}}_k^\infty$, we set

$$(18) \quad \mathbf{R}_{k,0}(p) = \inf \left[\sum_{i=1}^{k-2} \sqrt{w_i \log(1 + L_i V_i)} \right]^2$$

where the infimum runs among all ways of writing p in the form (17) a.e. Note that we allow some of the w_i to be zero in which case the corresponding densities p_i may be chosen arbitrarily and their choices do not contribute to the value of $\mathbf{R}_{k,0}(p)$. For $k \geq 3$ and $R > 0$, let $\overline{\mathcal{M}}_k^\infty(R)$ be the subset of $\overline{\mathcal{M}}_k^\infty$ that consists of these densities p for which $\mathbf{R}_{k,0}(p) < R$. When a density p belongs to $\overline{\mathcal{M}}_k^\infty(R)$, we may therefore write p in the form (17) a.e. with L_i and V_i such that $\left[\sum_{i=1}^{k-2} \sqrt{w_i \log(1 + L_i V_i)} \right]^2 < R$. The sequences of sets $(\overline{\mathcal{M}}_k^\infty)_{k \geq 3}$ and $(\overline{\mathcal{M}}_k^\infty(R))_{k \geq 3}$ are both increasing with respect to set inclusion. To see that a density $p \in \overline{\mathcal{M}}_l^\infty(R)$ belongs to $p \in \overline{\mathcal{M}}_k^\infty(R)$ when $l < k$, it suffices to introduce $k - l$ extra numbers $x_{l-1} < \dots < x_{k-2}$ which are larger than x_{l-2} and for $i \in \{l-2, \dots, k-1\}$, consider the uniform density p_i on (x_i, x_{i+1}) with the weight $w_i = 0$. It is not difficult to check that the set $\overline{\mathcal{M}}_k^\infty(R)$ is invariant under a location-scale transformation.

Theorem 3. *Let $k \geq 3$ and $R > 0$. If $p \in \overline{\mathcal{M}}_k^\infty(R)$,*

$$(19) \quad \mathbb{B}_{k,n}(p) \leq 41.3 \left(\frac{R}{n} \right)^{1/3} + 83.2 \sqrt{\frac{2k}{n}}.$$

This result is to our knowledge new in the literature. We deduce that the minimax risk for the \mathbb{L}_1 -norm over $\overline{\mathcal{M}}_k^\infty(R)$ is not larger than $(R/n)^{1/3} \vee (k/n)^{1/2}$ up to a positive multiplicative constant. For large enough values of n , the bound is of order $(R/n)^{1/3}$ while for moderate ones and values of R which are close enough to 0, which means that the densities in $\overline{\mathcal{M}}_k^\infty(R)$ are close to a mixture of $k - 2$ uniform distributions, the bound is of order $\sqrt{k/n}$.

It is interesting to compare this result to that established in Baraud and Birgé [4][page 3900] for their ρ -estimators. For the problem of estimating a bounded unimodal density p supported on an interval of length L , which is an element of $\overline{\mathcal{M}}_4$, Baraud and Birgé show that the Hellinger risk of the ρ -estimator is not larger than $(\sqrt{L} \|p\|_\infty / n)^{1/3} \log n$ up to some numerical constant. With our TV-estimator, the bound we get is of order $(\log(1 + L \|p\|_\infty) / n)^{1/3}$ and therefore only depends logarithmically on the quantity $L \|p\|_\infty$ (which is not smaller than 1 since p is a density).

Proof. The proof of Theorem 3 is based on (16) and the following approximation result. The complete proof is postponed to subsection 8.3. \square

Proposition 3. *Let $V \geq 0$ and I be a bounded interval of length $L > 0$. For all $D \geq 1$, there exists a partition $\mathcal{J} = \mathcal{J}(D, L, V)$ of I into $D \geq 1$ nontrivial intervals with the following properties. For any monotone density p on I for which*

$$V_I(p) = \sup_{x \in \tilde{I}} p(x) - \inf_{x \in \tilde{I}} p(x) \leq V,$$

the function $\bar{p} = \bar{p}(\mathcal{J})$ defined by

$$\bar{p} = \sum_{J \in \mathcal{J}} \bar{p}_J \mathbf{1}_J \quad \text{with} \quad \bar{p}_J = \frac{1}{\mu(J)} \int_J p \, d\mu,$$

is a monotone density on I that satisfies

$$(20) \quad \int_I |p - \bar{p}| \, d\mu \leq \left[(1 + VL)^{1/D} - 1 \right] \wedge 2 \leq \frac{2 \log(1 + VL)}{D}.$$

Proof. The proof is postponed to subsection 8.3. \square

Although the result is hidden in his calculations, Birgé [6] has established a bound of the same flavour where the variation V is replaced by a uniform bound on p . Unlike his, our bound (20) allows to recover the fact that when $V = 0$, i.e. when p is constant on I , the left-hand side equals 0 as expected. The combination of Theorem 2 and Theorem 3 immediately leads to the following

Corollary 1. *Let $k \geq 3$. If X_1, \dots, X_n are i.i.d. with a density $p \in \overline{\mathcal{M}}_k^\infty(R)$, then the TV-estimator \hat{p} on \mathcal{M}_k satisfies*

$$(21) \quad \mathbb{E} [\|p - \hat{p}\|] \leq 82.6 \left(\frac{R}{n}\right)^{1/3} + 166.4 \sqrt{\frac{2k}{n} + \frac{2\varepsilon}{n}}.$$

4.3. Estimation of other k -piecewise monotone densities. Corollary 1 provides an upper bound on the \mathbb{L}_1 -risk of the TV-estimator for estimating a density $p \in \overline{\mathcal{M}}_k^\infty$. A natural question is how the estimator performs when the density p is neither bounded nor supported on a compact interval. Since for such densities we may write

$$(22) \quad \mathbb{B}_{k,n}(p) \leq \inf_{\overline{p} \in \overline{\mathcal{M}}_k^\infty} \left[\frac{3}{2} \|p - \overline{p}\| + \mathbb{B}_{k,n}(\overline{p}) \right],$$

an upper bound on $\mathbb{B}_{k,n}(p)$ can be obtained by combining Theorem 3 with an approximation result showing how general densities in $\overline{\mathcal{M}}_k$ can be approximated by elements of $\overline{\mathcal{M}}_k^\infty$. In this section, we therefore study the approximation properties of the set $\overline{\mathcal{M}}_k^\infty$ with respect to possibly unbounded and non-compactly supported densities. We start with the case of a monotone density on a half-line and introduce the following definitions.

Definition 3. Given a nonincreasing density p on $(a, +\infty)$ with $a \in \mathbb{R}$, we define \tilde{p} as the mapping on $(0, +\infty)$ given by

$$(23) \quad \tilde{p}(y) = \inf \{x > 0, p(a+x) < y\} \geq 0.$$

We define the x -tail function $\tau_x(p, \cdot)$ associated to p as

$$\tau_x(p, \cdot) : \begin{cases} [0, +\infty) & \longrightarrow \mathbb{R}_+ \\ t & \longmapsto \int_t^{+\infty} p(a+x) d\mu(x) \end{cases}$$

the y -tail function $\tau_y(p, \cdot)$ as

$$\tau_y(p, \cdot) : \begin{cases} [0, +\infty) & \longrightarrow \mathbb{R}_+ \\ t & \longmapsto \int_t^{+\infty} \tilde{p}(y) d\mu(y) \end{cases}$$

and the tail function $\tau(p, \cdot)$ as

$$(24) \quad \tau(p, t) = \inf_{s>0} [\tau_x(p, st) + \tau_y(p, p(a+s))] \quad \text{for all } t \geq 1.$$

When p is a nondecreasing density on $(-\infty, -a)$ with $a \in \mathbb{R}$, we define $\tau_x(p, \cdot)$, $\tau_y(p, \cdot)$ and $\tau(p, \cdot)$ as respectively the x -tail, y -tail and tail functions of the nonincreasing density $x \mapsto p(-x)$.

Let us comment on these definitions. When p is a continuous decreasing density from $(0, +\infty)$ onto $(0, +\infty)$, \tilde{p} is the reciprocal function p^{-1} . By reflecting the graph of $x \mapsto p(x)$ in the line $y = x$, we easily see that $\tilde{p} = p^{-1}$ is a nonincreasing density on $(0, +\infty)$. This property remains true in the general case as shown by the lemma below with the special value $B = 0$. As a consequence, $\tau_y(p, \cdot)$ can be interpreted as the tail of the distribution function associated to the density \tilde{p} while $\tau_x(p, \cdot)$ is that of $p(a + \cdot)$. Figure 1 displays a graphical representations of the quantities $\tau_x(p, st)$ and $\tau_y(p, p(a+s))$ for $s > 0$ and $t > 1$.

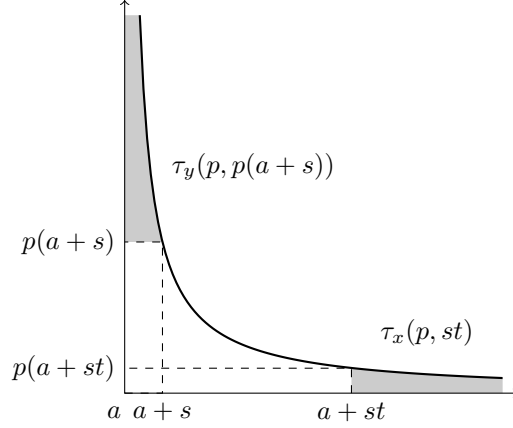


FIGURE 1. The quantities $\tau_x(p, st)$ and $\tau_y(p, p(a+s))$ (dark areas) for $s > 0$ and $t > 1$.

Lemma 2. Let p be a nonincreasing density on $(a, +\infty)$ with $a \in \mathbb{R}$ and \tilde{p} the mapping defined by (23). For all $B \geq 0$,

$$(25) \quad \int_0^{+\infty} [p(a+x) - B]_+ d\mu(x) = \int_B^{+\infty} \tilde{p}(y) d\mu(y) = \tau_y(p, B).$$

By changing p into $x \mapsto p(-x)$, we also obtain that when p is nondecreasing on $(-\infty, -a)$,

$$\int_{-\infty}^0 [p(-a+x) - B]_+ d\mu(x) = \int_B^{+\infty} \tilde{p}(y) d\mu(y) = \tau_y(p, B) \quad \text{for all } B \geq 0.$$

Proof. Let $y > 0$. Since p is a nonincreasing density on $(a, +\infty)$, it necessarily tends to 0 at $+\infty$. The set $I(y) = \{x > 0, p(a+x) < y\}$ is therefore a nonempty unbounded interval with endpoint $\tilde{p}(y) < +\infty$ by definition of $\tilde{p}(y)$. In particular,

$$(\tilde{p}(y), +\infty) \subset I(y) \subset [\tilde{p}(y), +\infty),$$

and by taking the complements of those sets we obtain that for all $(x, y) \in (0, +\infty) \times (0, +\infty)$

$$\mathbb{1}_{x < \tilde{p}(y)} \leq \mathbb{1}_{p(a+x) \geq y} \leq \mathbb{1}_{x \leq \tilde{p}(y)}.$$

Integrating these inequalities on $(0, +\infty) \times (B, +\infty)$ with respect to $\mu \otimes \mu$ and using Fubini's theorem, we obtain that

$$\begin{aligned} & \int_B^{+\infty} \tilde{p}(y) d\mu(y) \\ &= \int_B^{+\infty} \left[\int_0^{+\infty} \mathbb{1}_{x < \tilde{p}(y)} d\mu(x) \right] d\mu(y) \leq \int_B^{+\infty} \left[\int_0^{+\infty} \mathbb{1}_{p(a+x) \geq y} d\mu(x) \right] d\mu(y) \\ &= \int_0^{+\infty} \left[\int_B^{+\infty} \mathbb{1}_{p(a+x) \geq y} d\mu(y) \right] d\mu(x) = \int_a^{+\infty} [p(a+x) - B]_+ d\mu(x) \\ &\leq \int_B^{+\infty} \left[\int_0^{+\infty} \mathbb{1}_{x \leq \tilde{p}(y)} d\mu(x) \right] d\mu(y) = \int_B^{+\infty} \tilde{p}(y) d\mu(y), \end{aligned}$$

which proves (25). \square

It follows from (25) that if p is a nonincreasing density on $(a, +\infty)$, $\tau(p, t)$ also writes for all $t \geq 1$ as

$$\tau(p, t) = \inf_{s>0} \left[\int_{st}^{+\infty} p(a+x) d\mu(x) + \int_0^s [p(a+x) - p(a+s)] d\mu(x) \right].$$

It is not difficult to check that the mapping $\tau(p, \cdot)$ is nonincreasing on $[1, +\infty)$, tends to 0 at $+\infty$ and is invariant under a location-scale transformation, i.e. by changing p into the density $\sigma^{-1}p[\sigma^{-1}(\cdot - m)]$ on $(\sigma a + m, +\infty)$ with $m \in \mathbb{R}$ and $\sigma > 0$.

We consider the general situation where p is an arbitrary element of $\overline{\mathcal{M}}_\ell$ with $\ell \geq 2$. Changing the values of p on a negligible set, if ever necessary, which will not change the way it can be approximated in \mathbb{L}_1 -norm, we may assume without loss of generality that it can be written in the form:

$$(26) \quad p = w_1 p_1 \mathbb{1}_{(-\infty, x_1)} + \sum_{i=2}^{\ell-1} w_i p_i \mathbb{1}_{(x_{i-1}, x_i)} + w_\ell p_\ell \mathbb{1}_{(x_{\ell-1}, +\infty)}$$

where $x_1 < \dots < x_{\ell-1}$ is an increasing sequence of real numbers and $p_1, \dots, p_\ell, w_1, \dots, w_\ell$ are defined as follows. The functions p_1 and p_ℓ denote monotone densities on $(-\infty, x_1)$ and $(x_{\ell-1}, +\infty)$ respectively, $w_1 = \int_{-\infty}^{x_1} p d\mu$, $w_\ell = \int_{x_{\ell-1}}^{+\infty} p d\mu$ and when $\ell > 2$, the p_i are monotone densities on (x_{i-1}, x_i) and $w_i = \int_{x_{i-1}}^{x_i} p d\mu$ for all $i \in \{2, \dots, \ell-1\}$. For p written in the form (26), we set

$$(27) \quad \tau_\infty(p, t) = \max_{i \in \{1, \dots, \ell\}} \tau(p_i, t) \quad \text{for all } t \geq 1.$$

The mapping $t \mapsto \tau(p, t)$ is nonincreasing on \mathbb{R}_+ and tends to 0 at $+\infty$.

Theorem 4. *Let $\ell \geq 2$, $k \geq 2\ell$ and $R \geq \ell \log 2$. If p is a density of the form (26) a.e.,*

$$(28) \quad \inf_{\bar{p} \in \overline{\mathcal{M}}_k(R)} \|p - \bar{p}\| \leq 2\tau_\infty \left(p, \exp \left(\frac{R}{\ell} \right) - 1 \right).$$

Proof. The proof is postponed to Subsection 8.3. □

By combining Theorem 3 and Theorem 4 we obtain the following corollary.

Corollary 2. *Let $\ell \geq 2$ and $k \geq 2\ell$. If p is a density of the form (26) a.e.,*

$$(29) \quad \mathbb{B}_{k,n}(p) \leq 44.3 \left(\frac{\ell \log(1 + r_n)}{n} \right)^{1/3} + 83.2 \sqrt{\frac{2k}{n}}$$

where

$$(30) \quad r_n = \inf \left\{ t \geq 1, \tau_\infty(p, t) \leq \left(\frac{\ell \log(1 + t)}{n} \right)^{1/3} \right\}$$

and $\tau_\infty(p, \cdot)$ is defined by (27). Then, the TV-estimator \hat{p} on \mathcal{M}_k satisfies

$$(31) \quad \mathbb{E} [\|p - \hat{p}\|] \leq 88.6 \left(\frac{\ell \log(1 + r_n)}{n} \right)^{1/3} + 166.4 \sqrt{\frac{2k}{n}} + \frac{2\varepsilon}{n}.$$

Proof. Since $\tau_\infty(p, \cdot)$ tends to 0 at $+\infty$, the set

$$\mathcal{R} = \left\{ t \geq 1, \tau_\infty(p, t) \leq \left(\frac{\ell \log(1 + t)}{n} \right)^{1/3} \right\}$$

is nonempty, r_n is well-defined and for all $t > r_n \geq 1$,

$$\tau_\infty(p, t) \leq \left(\frac{\ell \log(1+t)}{n} \right)^{1/3}.$$

Using (22), Theorems 3 and 4 with $R = \ell \log(1+t) > \ell \log 2$ we obtain that

$$\begin{aligned} \mathbb{B}_{k,n}(p) &\leq \frac{3}{2} \inf_{\bar{p} \in \overline{\mathcal{M}}_k^\infty(R)} \|p - \bar{p}\| + \sup_{\bar{p} \in \overline{\mathcal{M}}_k^\infty(R)} \mathbb{B}_{k,n}(\bar{p}) \\ &\leq 3\tau_\infty(p, t) + 41.3 \left(\frac{\ell \log(1+t)}{n} \right)^{1/3} + 83.2 \sqrt{\frac{2k}{n}} \\ &\leq 44.3 \left(\frac{\ell \log(1+t)}{n} \right)^{1/3} + 83.2 \sqrt{\frac{2k}{n}}, \end{aligned}$$

and the result follows from the fact that t is arbitrary in $(r_n, +\infty)$. \square

Example 1. Let $n \geq 2$, $\alpha \geq 0$, $\beta \geq -1$, $\gamma \in (0, 1)$ and q be the mapping defined by

$$q(x) = \frac{2^{1-\gamma}}{x^{1-\gamma}} \mathbf{1}_{(0,2)}(x) + \frac{2^{1+\alpha} (\log 2)^{1+\beta}}{x^{1+\alpha} (\log x)^{1+\beta}} \mathbf{1}_{[2,+\infty)}(x).$$

When $(\alpha, \beta) \in (0, +\infty) \times [-1, +\infty)$ and when $\alpha = 0$ and $\beta > 0$, q is a positive, integrable, nonincreasing function on $(0, +\infty)$ and we may denote by p the corresponding density, i.e. $p = cq$ for some $c > 0$ depending on α, β and γ . The density p may be written in the form (26) with $\ell = 2$, $w_1 = 0$, $w_2 = 1$, $x_1 = 0$ and $p_2 = p$. Throughout this example, C denotes a positive number depending on α, β and γ that may vary from line to line.

It follows from Definition 3 with $a = 0$ that when $\alpha > 0$, for all $t \geq 2$

$$\frac{\tau_x(p, t)}{c2^{1+\alpha} (\log 2)^{1+\beta}} = \int_t^{+\infty} \frac{dx}{x^{1+\alpha} (\log x)^{1+\beta}} = \int_{\log t}^{+\infty} \frac{e^{-\alpha s}}{s^{1+\beta}} ds \leq \frac{1}{\alpha t^\alpha (\log t)^{1+\beta}},$$

and when $\alpha = 0$ and $\beta > 0$

$$\frac{\tau_x(p, t)}{c2^{1+\alpha} (\log 2)^{1+\beta}} = \int_{\log t}^{+\infty} \frac{1}{s^{1+\beta}} ds = \frac{1}{\beta (\log t)^\beta}.$$

For $y > c$, $\tilde{p} : y \mapsto 2(c/y)^{1/(1-\gamma)}$, hence

$$\tau_y(p, t) = \int_t^{+\infty} \tilde{p}(y) d\mu(y) = \frac{2(1-\gamma)c^{1/(1-\gamma)}}{\gamma t^{\gamma/(1-\gamma)}} \quad \text{for all } t \geq c.$$

We deduce that for all $t \geq 1$ and $s \in [2/t, 2]$, $p(s) \geq p(2) = c$ and

$$C^{-1} \tau_\infty(p, t) \leq \begin{cases} \left[(st)^\alpha (\log(st))^{1+\beta} \right]^{-1} + s^\gamma & \text{when } \alpha > 0 \text{ and } \beta \geq -1 \\ (\log(st))^{-\beta} + s^\gamma & \text{when } \alpha = 0 \text{ and } \beta > 0. \end{cases}$$

Taking

$$s = \begin{cases} 2 \left[\left(t^\alpha \left(\frac{\log(1+t)}{\log 2} \right)^{1+\beta} \right)^{-\frac{1}{\alpha+\gamma}} \vee t^{-1} \right] & \text{when } \alpha > 0 \text{ and } \beta \geq -1 \\ 2 \left[\left(\frac{\log(1+t)}{\log 2} \right)^{-\frac{\beta}{\gamma}} \vee t^{-1} \right] & \text{when } \alpha = 0 \text{ and } \beta > 0 \end{cases}$$

the value of which belongs to $[2/t, 2]$, we obtain that for all $t \geq 1$

$$C^{-1}\tau_\infty(p, t) \leq \begin{cases} \left[t^\alpha (\log(1+t))^{1+\beta} \right]^{-\frac{\gamma}{\alpha+\gamma}} \vee t^{-\gamma} & \text{when } \alpha > 0 \text{ and } \beta \geq -1 \\ (\log(1+t))^{-\beta} \vee t^{-\gamma} & \text{when } \alpha = 0 \text{ and } \beta > 0. \end{cases}$$

If $\alpha > 0$ and $\beta \geq -1$, by taking

$$t = t_n = C'n^{\frac{1}{3}(\frac{1}{\alpha} + \frac{1}{\gamma})} \log^{-\kappa} n \quad \text{with} \quad \kappa = \frac{1}{3} \left(\frac{1}{\alpha} + \frac{1}{\gamma} \right) + \frac{1+\beta}{\alpha}$$

for some constant $C' > 0$ large enough, we obtain that $\tau_\infty(p, t_n) \leq (2 \log(1+t_n)/n)^{1/3}$ and consequently, r_n defined by (30) satisfies $r_n \leq t_n$. Applying Corollary 2, we conclude that for all $k \geq 4$

$$\mathbb{B}_{k,n}(p) \leq C \left[\left(\frac{\log n}{n} \right)^{1/3} + \sqrt{\frac{2k}{n}} \right].$$

If $\alpha = 0$ and $\beta > 0$, we take $t = t_n$ such that $\log(1+t_n) = C'n^{1/(1+3\beta)}$ for some constant $C' > 0$ large enough, we obtain that

$$\tau_\infty(p, t_n) \leq Cn^{\frac{-\beta}{1+3\beta}} \leq \left(\frac{2 \log(1+t_n)}{n} \right)^{1/3},$$

hence $r_n \leq t_n$ and we get that for all $k \geq 4$

$$\mathbb{B}_{k,n}(p) \leq C \left[n^{\frac{-\beta}{1+3\beta}} + \sqrt{\frac{2k}{n}} \right].$$

We are not aware of any \mathbb{L}_1 -risk bound for the Grenander estimator when the target density has an infinite support and heavy tails.

5. CONVEX-CONCAVE DENSITIES

5.1. Piecewise monotone convex-concave densities. In this section, our aim is to estimate a density on the line which is piecewise monotone convex-concave in the sense defined below.

Definition 4. A function g is said to be convex-concave on an interval I if it is either convex or concave on I . For $k \geq 2$, a function g on \mathbb{R} is said to be k -piecewise monotone convex-concave if there exists $A \in \mathcal{A}(k-1)$ such that the restriction of g to the each interval $I \in \mathbf{I}(A)$ is monotone and convex-concave. In particular, there exist at most k functions $\{g_I, I \in \mathbf{I}(A)\}$, where g_I is monotone and convex-concave on I such that

$$g(x) = \sum_{I \in \mathbf{I}(A)} g_I(x) \mathbb{1}_I(x) \quad \text{for all } x \in \mathbb{R} \setminus A.$$

We denote by $\overline{\mathcal{M}}_k^1$ the set of k -piecewise monotone convex-concave densities. The Laplace density $x \mapsto (1/2)e^{-|x|}$ belongs to $\overline{\mathcal{M}}_2^1$, the uniform density on a bounded interval belongs to $\overline{\mathcal{M}}_3^1$, all convex-concave densities on an interval belong to $\overline{\mathcal{M}}_4^1$. A function $g \in \overline{\mathcal{M}}_k^1$ associated to $A \in \mathcal{A}(k-1)$ admits left and right derivatives at any point $x \in \mathbb{R} \setminus A$. These derivatives are denoted by g'_l, g'_r respectively. More generally, when a function f is continuous and convex-concave on a nontrivial bounded interval $[a, b]$, we define

$$f'_r(z) = \lim_{x \downarrow z} \frac{f(x) - f(z)}{x - z} \quad \text{for all } z \in [a, b)$$

and

$$f'_l(z) = \lim_{x \uparrow z} \frac{f(x) - f(z)}{x - z} \quad \text{for all } z \in (a, b].$$

These quantities are finite for all $z \in (a, b)$ and belong to $[-\infty, +\infty]$ when $z \in \{a, b\}$. We say that f admits a right derivative at a and a left derivative at b when $f'_r(a)$ and $f'_l(b)$ are finite respectively.

The role played by piecewise constant functions in the previous section is here played by piecewise linear functions. For $D \geq 1$, let $\overline{\mathcal{O}}_{D,k}^1$ be the subset of $\overline{\mathcal{M}}_k^1$ that consists of those densities that are left-continuous and affine on each interval of a class $\mathbf{I}(A)$ with $A \in \mathcal{A}(D+1)$. For example, the left-continuous version of the density of a uniform distribution on a nontrivial interval belongs to $\overline{\mathcal{O}}_{1,3}^1$. The proposition below shows that the elements of $\overline{\mathcal{O}}_{D,k}^1$ are extremal in $\overline{\mathcal{M}}_k^1$.

Proposition 4. *Let $k \geq 2$, $D \geq 1$. If $p \in \overline{\mathcal{M}}_k^1$ and $q \in \overline{\mathcal{O}}_{D,k}^1$, the sets $\{x \in \mathbb{R}, p(x) - q(x) > 0\}$ and $\{x \in \mathbb{R}, p(x) - q(x) < 0\}$ are unions of at most $D + 2k - 1$ intervals. In particular, the elements of $\overline{\mathcal{O}}_{D,k}^1$ are extremal in $\overline{\mathcal{M}}_k^1$ with degree not larger than $2(D + 2k - 1)$.*

Proof. The proof is postponed to Subsection 8.4. □

For all $D \geq 1$, $k \geq 2$, let $\mathcal{O}_{D,k}^1$ be a countable and dense subset of $\overline{\mathcal{O}}_{D,k}^1$ (for the \mathbb{L}_1 -norm) and \mathcal{M}_k^1 a countable and dense subset of $\overline{\mathcal{M}}_k^1$ that contains $\bigcup_{D \geq 1} \mathcal{O}_{D,k}^1$. By proposition 2, the elements of $\mathcal{O}_{D,k}^1$ are also extremal in \mathcal{M}_k^1 with degree no larger than $2(D + 2k - 1)$ for all $D \geq 1$. We deduce from Theorem 1 the following result.

Theorem 5. *Let $k \geq 2$. For every product distribution \mathbf{P}^* of the data, a TV-estimator \hat{P} on \mathcal{M}_k^1 satisfies*

$$(32) \quad \mathbb{E} \left[d(P^*, \hat{P}) \right] \leq \inf_{D \geq 1} \left\{ 3 \inf_{P \in \overline{\mathcal{O}}_{D,k}^1} d(P^*, P) + 48 \sqrt{\frac{2(D + 2k - 1)}{n}} \right\} + \frac{\varepsilon}{n}.$$

In the remaining part of this section we assume that the X_i are i.i.d. with a density $p \in \overline{\mathcal{M}}_k^1$, in which case, the right-hand side of (32) can be written as $\mathbb{B}_{k,n}^1(p) + \varepsilon/n$ with

$$(33) \quad \mathbb{B}_{k,n}^1(p) = \inf_{D \geq 1} \left[\frac{3}{2} \inf_{q \in \overline{\mathcal{O}}_{D,k}^1} \|p - q\| + 48 \sqrt{\frac{2(D + 2k - 1)}{n}} \right].$$

As we did in Section 4, our aim is to bound the quantity $\mathbb{B}_{k,n}^1(p)$ under some suitable additional assumptions on the density p .

5.2. Approximation of a monotone convex-concave density by a piecewise linear function. Let us now turn to the approximation of a monotone convex-concave density by a convex-concave piecewise linear function. The approximation result that we establish is actually true for a *sub-density* on an interval $[a, b]$, i.e. a nonnegative function on $[a, b]$ the integral of which is not larger than 1. In the remaining part of this chapter, we use the following convenient definition.

Definition 5. Let $D \geq 1$ and f be a continuous function on a compact nontrivial interval $[a, b]$. We say that \overline{f} is a D -linear interpolation of f on $[a, b]$ if there exists a subdivision $a = x_0 < \dots < x_D = b$ such that $\overline{f}(x_i) = f(x_i)$ and \overline{f} is affine on $[x_{i-1}, x_i]$ for all $i \in \{1, \dots, D\}$.

This definition automatically determines the values of \overline{f} on $[a, b]$ since \overline{f} corresponds on $[x_{i-1}, x_i]$ to the chord that connects $(x_{i-1}, f(x_{i-1}))$ to $(x_i, f(x_i))$ for all $i \in \{1, \dots, D\}$. The function \overline{f} is therefore continuous and piecewise linear on a partition of $[a, b]$ into D intervals and it inherits of some of the features

of the function f . For example, if f is nonnegative, increasing, decreasing, convex or concave, so is \bar{f} . If f is convex (respectively concave), $\bar{f} \geq f$ (respectively $\bar{f} \leq f$).

Given a continuous monotone convex-concave function f with increment $\Delta = (f(b) - f(a))/(b - a)$ on a bounded nontrivial interval $[a, b]$, we define its linear index $\Gamma = \Gamma(f)$ as

$$\Gamma = 1 - \frac{1}{2} \left(\frac{|p'_r(a)| \wedge |p'_l(b)|}{|\Delta|} + \frac{|\Delta|}{|p'_r(a)| \vee |p'_l(b)|} \right),$$

with the conventions $0/0 = 1$ and $1/(+\infty) = 0$. Since f is convex-concave, monotone and continuous

$$|p'_r(a)| \wedge |p'_l(b)| \leq |\Delta| \leq |p'_r(a)| \vee |p'_l(b)| \quad \text{and}$$

and

$$\Delta = 0 \implies |p'_r(a)| = |p'_l(b)| = 0.$$

With our conventions, Γ is well-defined and belongs to $[0, 1]$. When f is affine, $\Delta = p'_r(a) = p'_l(b)$ and its linear index is 0. In the opposite direction when f is far from being affine, say when for some $c \in (a, b)$ and $v > 0$

$$f(x) = \frac{v}{b-c}(x-c)_+ \quad \text{for all } x \in [a, b]$$

its linear index $\Gamma = 1 - (b-c)/[2(b-a)]$ increases to 1 as c approaches b .

Theorem 6. *Let p be a monotone, continuous, convex-concave density on a bounded interval $[a, b]$ of length $L > 0$ with variation $V = |p(b) - p(a)|$ and linear index $\Gamma \in [0, 1]$. For all $D \geq 1$, there exists a $2D$ -linear interpolation \bar{p} of p such that*

$$(34) \quad \int_a^b |p - \bar{p}| d\mu \leq \frac{4}{3} \left[\left(1 + \sqrt{2LVT\Gamma} \right)^{1/D} - 1 \right]^2.$$

In particular, there exists a continuous convex-concave piecewise linear density q based on a partition of $[a, b]$ into $2D$ intervals that satisfies

$$(35) \quad \int_a^b |p - q| d\mu \leq 5.14 \frac{\log^2 \left(1 + \sqrt{2LVT\Gamma} \right)}{D^2}.$$

Since $\Gamma \in [0, 1]$, (35) implies that

$$\int_a^b |p - q| d\mu \leq 5.14 \frac{\log^2 \left(1 + \sqrt{2LV} \right)}{D^2}.$$

Nevertheless, when p is affine on $[a, b]$, $\Gamma = 0$ and we recover the fact that we may choose \bar{p} and q on $[a, b]$ such that $\int_a^b |p - \bar{p}| d\mu = \int_a^b |p - q| d\mu = 0$.

Guérin *et al* [20] already established that any bounded convex (or concave) function on a compact interval can be approximated by a piecewise affine function on a partition of this interval into D -pieces with an \mathbb{L}_1 -error not larger than C/D^2 , for some number $C > 0$ which is independent of D . The novelty in Theorem 6 lies in the fact that for probability densities the approximation error depends logarithmically on the product LV .

Proof. The proof is based on several preliminary approximation results whose statements and proofs are postponed to Subsection 8.4. \square

5.3. Estimation of k -piecewise monotone convex-concave bounded and compactly supported densities. In this section, we consider a density $p \in \overline{\mathcal{M}}_k^1$, with $k \geq 3$, that is of the form (17) except for the fact that (iii) is here replaced by

(iii') for $i \in \{1, \dots, k-2\}$, p_i is a monotone continuous convex-concave density on the interval $[x_{i-1}, x_i]$ of length $L_i > 0$, with variation $V_i = |p_i(x_{i-1}) - p(x_i)| < +\infty$ and linear index $\Gamma_i \in [0, 1]$.

Throughout this section, when we refer to the form (17), it means with (iii) replaced by (iii'). For such a density p , we define

$$(36) \quad \mathbf{R}_{k,1}(p) = \inf \left[\sum_{i=1}^{k-2} \left(w_i \log^2 \left(1 + \sqrt{2L_i V_i \Gamma_i} \right) \right)^{1/3} \right]^{3/2},$$

where the infimum runs among all ways of writing p in the form (17). We denote by $\overline{\mathcal{M}}_{k,1}^\infty$ the class of all densities of the form (17) a.e. and for $R > 0$, $\overline{\mathcal{M}}_{k,1}^\infty(R)$ the subset of those which satisfy $\mathbf{R}_{k,1}(p) < R$.

Note that a concave (or convex) density on a (necessarily) compact interval belongs to $\overline{\mathcal{M}}_4^1$. The following result holds.

Theorem 7. *Let $k \geq 3$ and $R > 0$. For all $p \in \overline{\mathcal{M}}_{k,1}^\infty(R)$*

$$(37) \quad \mathbb{B}_{k,n}^1(p) \leq 7.71 \left[15.06 \left(\frac{R}{n} \right)^{2/5} + 8.82 \sqrt{\frac{4k-5}{n}} \right],$$

where $\mathbb{B}_{k,n}^1(p)$ is defined by (33).

Theorem 7 together with Theorem 5 imply that the TV-estimator converges at rate $n^{-2/5}$ in total variation distance, whenever the underlying density p^* is bounded, compactly supported and belongs to the class $\overline{\mathcal{M}}_{k,1}^\infty$ of k -piecewise monotone convex-concave densities. The rate $n^{-2/5}$ matches the minimax lower bound established in Devroye and Lugosi [11][Section 15.5] for bounded convex densities. This rate is therefore optimal.

To the best of our knowledge, Theorem 7 provides the sharpest known minimax upper bound in this setting, including the case of a monotone, convex or concave density on a compact interval. In comparison, Gao and Wellner [15] proved that the MLE on the set of convex non-increasing densities on a given interval achieves the rate $n^{-2/5}$ (for the Hellinger distance). Note that the construction of the MLE requires that the support of the target density be known while our TV-estimator assumes nothing.

Consider now the special case of a continuous, concave density on an interval $[a, b]$ (which is necessarily bounded). A monotone continuous concave density p on a bounded interval $[a, b]$ with length $L > 0$ belongs to $\overline{\mathcal{M}}_3^1$. It necessarily satisfies $L|p(a) - p(b)|/2 \leq 1$, hence $\mathbf{R}_{3,1}(p) \leq \log(1 + \sqrt{2L|p(a) - p(b)|}) \leq \log 3$. If p is not monotone but only continuous and concave on $[a, b]$, we may write p as $w_1 p_1 + w_2 p_2$ where p_1 and p_2 are monotone and concave densities on the intervals $[a, c]$ and $[c, b]$ respectively where c is a maximizer of p in (a, b) . The density p belongs to $\overline{\mathcal{M}}_4^1$ and by applying the previous inequality to the densities p_1 and p_2 successively and the inequality $z^{1/3} + (1-z)^{1/3} \leq 2^{2/3}$ which holds for all $z \in [0, 1]$, we obtain that

$$\begin{aligned} \mathbf{R}_{4,1}^{2/3}(p) &\leq \left(w_1 \log^2(1 + \sqrt{2(c-a)(p_1(c) - p_1(a))}) \right)^{1/3} + \left(w_2 \log^2(1 + \sqrt{2(b-c)(p_2(c) - p_2(b))}) \right)^{1/3} \\ &\leq \left(w_1^{1/3} + w_2^{1/3} \right) (\log 3)^{2/3} \leq (2 \log 3)^{2/3}. \end{aligned}$$

We immediately deduce from Theorems 5 and 7 the following corollary.

Corollary 3. *If X_1, \dots, X_n are i.i.d. with a density that is concave on an interval of \mathbb{R} , a TV-estimator \hat{p} on \mathcal{M}_4^1 satisfies*

$$\mathbb{E}[\|p - \hat{p}\|] \leq \frac{320}{n^{2/5}} + \frac{451}{\sqrt{n}} + \frac{2\varepsilon}{n}.$$

In particular,

$$\inf_{\tilde{p}} \sup_p \mathbb{E}[\|p - \tilde{p}\|] \leq \frac{320}{n^{2/5}} + \frac{451}{\sqrt{n}},$$

where the supremum runs among all concave densities p on an interval of \mathbb{R} and the infimum over all density estimators \tilde{p} based on a n -sample with density p .

Proof. Let $p \in \overline{\mathcal{M}}_{k,1}^\infty(R)$. Without loss of generality we may assume that p is of the form (17) everywhere and choose a subdivision $(x_i)_{i \in \{0, \dots, k-2\}}$ in such a way that

$$\left[\sum_{i=1}^{k-2} \left(w_i \log^2 \left(1 + \sqrt{2L_i V_i \Gamma_i} \right) \right)^{1/3} \right]^{3/2} \leq R.$$

Let $D \geq k-2$ and D_1, \dots, D_{k-2} be some positive integers to be chosen later on that satisfy the constraint $\sum_{i=1}^{k-2} D_i \leq D$. By Theorem 6, we may find for all $i \in \{1, \dots, k-2\}$ a density q_i that is continuous and supported on $[x_{i-1}, x_i]$, piecewise linear on a partition of $[x_{i-1}, x_i]$ into $2D_i$ intervals, that satisfies

$$\int_{x_{i-1}}^{x_i} |p_i - q_i| d\mu \leq 5.14 \frac{\log^2 \left(1 + \sqrt{2L_i V_i \Gamma_i} \right)}{D_i^2}.$$

The function $q = \sum_{i=1}^{k-2} w_i q_i \mathbb{1}_{(x_{i-1}, x_i]}$ is a density, that is left-continuous, convex-concave on each interval $I \in \mathbf{I}(\{x_0, \dots, x_{k-2}\})$ and affine on each interval of a partition $(x_0, x_{k-2}] = \bigcup_{i=1}^{k-2} (x_{i-1}, x_i]$ into $\sum_{i=1}^{k-2} 2D_i \leq 2D$ intervals. It therefore belongs to $\overline{\mathcal{O}}_{2D,k}$. Besides,

$$\|p - q\| \leq \sum_{i=1}^{k-2} w_i \int_{x_{i-1}}^{x_i} |p_i - q_i| d\mu \leq 5.14 \sum_{i=1}^{k-2} \frac{w_i \log^2 \left(1 + \sqrt{2L_i V_i \Gamma_i} \right)}{D_i^2}$$

and it follows from (33) that

$$\begin{aligned} \mathbb{B}_{k,n}^1(p) &\leq \frac{3}{2} \inf_{q \in \overline{\mathcal{O}}_{2D,k}^1} \|p - q\| + 68 \sqrt{\frac{2D + 2k - 1}{n}} \\ &\leq 7.71 \sum_{i=1}^{k-2} \frac{w_i \log^2 \left(1 + \sqrt{2L_i V_i \Gamma_i} \right)}{D_i^2} + 68 \sqrt{\frac{2D + 2k - 1}{n}}. \end{aligned}$$

Let us set $c = 68/7.71$, $s_0 = [nR^4/(2c^2)]^{1/5}$,

$$s_i = \left[w_i \log^2 \left(1 + \sqrt{2L_i V_i \Gamma_i} \right) \right]^{1/3} \quad \text{for all } i \in \{1, \dots, k-2\}$$

and choose $D = s_0 + k - 2$ and

$$D_i = \left\lceil \frac{s_0 s_i}{\sum_{j=1}^{k-2} s_j} \right\rceil \geq \frac{s_0 s_i}{\sum_{j=1}^{k-2} s_j} \vee 1 \quad \text{for all } i \in \{1, \dots, k-2\},$$

so that $\sum_{i=1}^{k-2} D_i \leq D$. Then,

$$\begin{aligned} \mathbb{B}_{k,n}^1(p) &\leq 7.71 \left[\sum_{i=1}^{k-2} \frac{s_i^3}{D_i^2} + c\sqrt{\frac{2D+2k-1}{n}} \right] \leq 7.71 \left[\frac{\left(\sum_{j=1}^{k-2} s_j\right)^3}{s_0^2} + c\sqrt{\frac{2s_0+4k-5}{n}} \right] \\ &\leq 7.71 \left[\frac{R^2}{s_0^2} + c\sqrt{\frac{2s_0}{n}} + c\frac{\sqrt{4k-5}}{n} \right] = 7.71 \left[2(2c^2)^{2/5} \frac{R^{2/5}}{n^{2/5}} + c\frac{\sqrt{4k-5}}{n} \right] \end{aligned}$$

which gives (37). \square

6. THE CASE OF s -CONCAVE DENSITIES AND THEIR GENERALISATIONS

Given an increasing function \mathcal{L} on $(0, +\infty)$ and a density p for which $\{p > 0\}$ is an open subinterval of \mathbb{R} , we define

$$\mathcal{L}p : x \mapsto \begin{cases} \mathcal{L} \circ p(x) & \text{when } x \in \{p > 0\} \\ -\infty & \text{otherwise.} \end{cases}$$

A density p is said to be \mathcal{L} -concave if $\mathcal{L}p$ is concave on $\{p > 0\}$. We denote by $\overline{\mathcal{M}}(\mathcal{L})$ the class of \mathcal{L} -concave densities and for $D \geq 1$, $\overline{\mathcal{O}}_D(\mathcal{L})$ the subset of $\overline{\mathcal{M}}(\mathcal{L})$ which consists of these densities \overline{p} for which $\mathcal{L}\overline{p}$ is either affine or takes the value $-\infty$ on the elements of a partition $\mathbf{I}(A)$, $A \in \mathcal{A}(D)$, containing thus at most $D+1$ intervals. Note that the uniform distributions on an interval belong to $\overline{\mathcal{O}}_D(\mathcal{L})$ for all $D \geq 2$, which shows in passing that the sets $\overline{\mathcal{M}}(\mathcal{L})$ and $\overline{\mathcal{O}}_D(\mathcal{L})$ are nonempty for every choices of \mathcal{L} and $D \geq 2$. As before, we denote by $\mathcal{O}_D(\mathcal{L})$ a countable and dense subset of $\overline{\mathcal{O}}_D(\mathcal{L})$ for $D \geq 1$ and $\mathcal{M}(\mathcal{L})$ a countable and dense subset of $\overline{\mathcal{M}}(\mathcal{L})$ which contains $\bigcup_{D \geq 1} \mathcal{O}_D(\mathcal{L})$. The sets $\mathcal{O}_D(\mathcal{L})$, $\overline{\mathcal{O}}_D(\mathcal{L})$, $\mathcal{M}(\mathcal{L})$ and $\overline{\mathcal{M}}(\mathcal{L})$ are the classes of probability distributions on \mathbb{R} that are associated with the sets of densities $\mathcal{O}_D(\mathcal{L})$, $\overline{\mathcal{O}}_D(\mathcal{L})$, $\mathcal{M}(\mathcal{L})$ and $\overline{\mathcal{M}}(\mathcal{L})$ respectively.

The functions \mathcal{L} we have in mind are typically of the form $\mathcal{L} = \mathcal{L}_s$ with $s \in \mathbb{R}$ and defined for $u > 0$ by

$$\mathcal{L}_s(u) = \begin{cases} -u^s & \text{when } s < 0 \\ \log u & \text{when } s = 0 \\ u^s & \text{when } s > 0. \end{cases}$$

In the literature, a \mathcal{L}_s -concave density is called s -concave and log-concave in the special case $s = 0$. For $s < s'$, the reader can check that $\overline{\mathcal{M}}(\mathcal{L}_{s'}) \subset \overline{\mathcal{M}}(\mathcal{L}_s)$, hence that the family $(\overline{\mathcal{M}}(\mathcal{L}_s))_{s \in \mathbb{R}}$ is nonincreasing with respect to set inclusion.

Seregin and Wellner [27] proved that the MLE on $\overline{\mathcal{M}}(\mathcal{L}_s)$ exists for those $s > -1$ (at least for n large enough). The following result shows that it is unstable at least for those $s \geq 1$.

Proposition 5. *Given $n \geq 5$ and $c \geq 5$, consider the distribution $P^* = (1 - 1/n)P_0 + (1/n)\delta_c$ where P_0 denotes the uniform distribution on $(0, 1)$. Let \widehat{L}_n be the likelihood function on $\mathcal{M}(\mathcal{L}_s)$, with $s \geq 1$, based on an n -sample with distribution P^* . With a probability at least 63%, for any density $p_s \in \mathcal{M}(\mathcal{L}_s)$ such that $\widehat{L}_n(p_s) > 0$, the associated distribution $P_s = p_s \cdot \mu \in \mathcal{M}(\mathcal{L}_s)$ satisfies*

$$d(P_0, P_s) \geq \frac{1}{2} \left(1 - \frac{2}{c-1} \right) \geq \frac{1}{4} \quad \text{while} \quad P_0 \in \mathcal{M}(\mathcal{L}_s) \quad \text{and} \quad d(P_0, P^*) = \frac{1}{n}.$$

Proof. Let $X_{(1)}, \dots, X_{(n)}$ the order statistics associated to the n -sample with distribution P^* . Since $c > 1$,

$$\mathbb{P}[X_{(1)} = c] = \mathbb{P}(X_1 = \dots = X_n = c) = n^{-n} \quad \text{and} \quad \mathbb{P}[X_{(n)} \neq c] = \left(1 - \frac{1}{n} \right)^n < e^{-1}.$$

We deduce that the probability of the set $\Omega = \{0 \leq X_{(1)} < X_{(n)} = c\}$ is at least $1 - e^{-1} - n^{-n} > 63\%$ for $n \geq 5$. Since $s \geq 1$, a density p_s in $\mathcal{M}(\mathcal{L}_s)$ is continuous and concave on $\{p_s > 0\}$, $\{p_s > 0\}$ is a bounded interval of the form (a, b) with $a < b$ and p_s can be extended continuously at a and b . In particular, there exists $z \in [a, b]$ such that $p_s(z) = \sup_{x \in [a, b]} p_s(x)$. Besides, the condition $\widehat{L}_n(p_s) > 0$ implies that on Ω , $a \leq X_{(1)} \leq 1 \leq X_{(n)} = c \leq b$ and since the graph of the density p_s on $[a, b]$ lies above its chord,

$$1 = \int_a^b p_s(x) dx \geq \frac{p_s(z)(b-a)}{2} \geq \frac{p_s(z)(c-1)}{2}.$$

Hence, $p_s(z) \leq 2/(c-1) < 1$ for $c > 3$ and we conclude that

$$d(P_0, P_s) \geq \frac{1}{2} \int_0^1 |1 - p_s(x)| dx \geq \frac{1}{2} \int_0^1 (1 - p_s(x)) dx \geq \frac{1}{2} \left(1 - \frac{2}{c-1}\right).$$

□

6.1. Extremal points and risk bounds over $\overline{\mathcal{M}}(\mathcal{L})$. The following result shows that the elements of $\overline{\mathcal{O}}_D(\mathcal{L})$ are extremal.

Proposition 6. *For all $D \geq 1$, the elements of $\overline{\mathcal{O}}_D(\mathcal{L})$ are extremal in $\overline{\mathcal{M}}(\mathcal{L})$ with degree not larger than $2(D+2)$.*

Proof. The proof is postponed to Section 8.5. □

We immediately deduce from Theorem 1 the following risk bound for the TV-estimator $\widehat{P} = \widehat{p} \cdot \mu$ on $\overline{\mathcal{M}}(\mathcal{L})$.

Theorem 8. *For every product distribution \mathbf{P}^* of the data, a TV-estimator \widehat{P} on $\mathcal{M}(\mathcal{L})$ satisfies*

$$(38) \quad \mathbb{E} \left[d(P^*, \widehat{P}) \right] \leq \inf_{D \geq 1} \left[3 \inf_{P \in \overline{\mathcal{O}}_D(\mathcal{L})} d(P^*, P) + 48\sqrt{2} \sqrt{\frac{D+2}{n}} \right] + \frac{\varepsilon}{n}.$$

As a consequence, if the data are i.i.d. with a density p^* that belongs to $\overline{\mathcal{O}}(\mathcal{L}) = \bigcup_{D \geq 1} \overline{\mathcal{O}}_D(\mathcal{L})$, the density $\widehat{p} \in \overline{\mathcal{M}}(\mathcal{L})$ of the TV-estimator converges in \mathbb{L}_1 to p^* at rate $1/\sqrt{n}$. When \mathcal{L} is continuous, hence one-to-one from $(0, +\infty)$ onto its range, this result applies to these densities p^* of the form $\mathcal{L}^{-1}(-a|x-c|+b)\mathbb{1}_I(x)$ where $a \in [0, +\infty)$, $b, c \in \mathbb{R}$ and I is an open interval. This includes all uniform densities on \mathbb{R} (even when \mathcal{L} is not continuous). In the special case where $\mathcal{L} = \mathcal{L}_s$ with $s \in (-1, 0)$, densities of the form

$$p(x) = \left(\frac{1}{a|x|+b} \right)^{1/|s|} \quad \text{or} \quad p(x) = \left(\frac{1}{ax+b} \right)^{1/|s|} \mathbb{1}_{x>0} \quad \text{with } a, b > 0$$

and their translated versions belong to $\overline{\mathcal{O}}(\mathcal{L}_s)$.

Of special interest is the case $s = 0$, which corresponds to log-concave densities. As a result, we obtain that our TV-estimator on the set of log-concave densities estimates at rate $1/\sqrt{n}$ densities of the form $x \mapsto 2a \exp[-a|x-b|]$ or $x \mapsto a \exp[-a(x-b)] \mathbb{1}_{x>b}$ with $a > 0$ and $b \in \mathbb{R}$. Our estimator shares thus some adaptation property which is of the same flavour as those established for the MLE by Kim *et al* [22] (and Feng *et al.* [14]), and by Baraud and Birgé [4] for their ρ -estimator. As compared to theirs, our upper bound does not involve logarithmic factors. When considering the Hellinger loss, these logarithmic factors are sometimes necessary while they disappear for the total variation one. Nevertheless, the paper by Kim *et al* [22] additionally contains a risk bound for the MLE with respect to total variation distance (see their Theorem 1). For the exponential distribution (and their translated versions), they obtain a rate of convergence which is $(\log n)/\sqrt{n}$ for the MLE while our estimator would converge at the faster rate $1/\sqrt{n}$. We do not know whether or not this extra logarithmic factor for the MLE is technical or necessary. In the latter

case, this would mean that the MLE is slightly sub-optimal for estimating exponential distributions among the set of log-concave densities. Nevertheless, the authors also exhibit examples of log-concave densities for which this extra logarithmic factor can be removed. For these cases, both the MLE and our TV estimator converge at the rate $1/\sqrt{n}$.

In order to specify further the risk bound (38), we need to investigate the approximation properties of $\overline{\mathcal{O}}_D(\mathcal{L})$ with respect to P^* . In the remaining part of this section, we shall restrict ourselves to the case where \mathcal{L} is either convex or concave on $(0, +\infty)$ and the data i.i.d. with a density $p^* \in \overline{\mathcal{M}}(\mathcal{L})$. Within this framework, our aim is to study the approximation properties of the sets of densities $\overline{\mathcal{O}}_D(\mathcal{L})$, $D \geq 1$, with respect to the target density p^* for the L_1 -norm.

6.2. Approximation properties of $\overline{\mathcal{O}}_D(\mathcal{L})$ when \mathcal{L} is convex. In this section, we study the case where the function \mathcal{L} is convex which includes functions of the form $\mathcal{L} = \mathcal{L}_s$ with $s \geq 1$. When the function \mathcal{L} is not only increasing but also convex, it is continuous on $(0, +\infty)$ and its admits an inverse \mathcal{L}^{-1} , defined on the range of \mathcal{L} , which is increasing, continuous and concave. Since $p = \mathcal{L}^{-1}(\mathcal{L}p)$ on $\{p > 0\}$, the density p is necessarily concave on the interval $\{p > 0\}$ and can therefore be well approximated by piecewise affine functions. However, such functions may not belong to the sets $\overline{\mathcal{O}}_D(\mathcal{L})$. Nevertheless, we show that the approximation properties of the elements of $\overline{\mathcal{O}}_D(\mathcal{L})$ with respect to p are, up to a factor 2, as good as those we would get by using piecewise affine functions.

Proposition 7. *Let \mathcal{L} be a convex increasing function on $(0, +\infty)$ and p an element of $\overline{\mathcal{M}}(\mathcal{L})$. The set $\{p > 0\}$ is a bounded open interval (a, b) with $a < b$ and p can be extended continuously on $[a, b]$. For all $D \geq 1$, there exist a $2D$ -linear interpolation \overline{p} of p on $[a, b]$ as well as a density $s \in \overline{\mathcal{O}}_{2D+1}(\mathcal{L})$ such that*

$$(39) \quad \int_{\mathbb{R}} |p - s| d\mu \leq 2 \int_a^b |p - \overline{p}| d\mu \leq \frac{2}{D^2}.$$

In particular, this result applies to $\mathcal{L} = \mathcal{L}_s$ with $s \geq 1$.

Proof. The proof of the proposition is postponed to Section 8.5. □

6.3. Approximation properties of $\overline{\mathcal{O}}_D(\mathcal{L})$ when \mathcal{L} is concave. In this section we investigate the situation where the function \mathcal{L} is concave.

Theorem 9. *Assume that the function \mathcal{L} is increasing and concave on $(0, +\infty)$ and that there exist $\lambda, \kappa > 1$ such that*

$$(40) \quad \lambda^2 \mathcal{L}'_l(\lambda y) \geq \kappa \mathcal{L}'_r(y) \quad \text{for all } y > 0.$$

Then, set

$$(41) \quad \gamma = \gamma(\lambda, \kappa) = \left[\frac{2}{\log \kappa} \left(\frac{\kappa}{\kappa - 1} \right)^{1/2} + \frac{\lambda \sqrt{\lambda - 1}}{\sqrt{2\kappa}} \left(1 + \frac{1}{2(\sqrt{\kappa} - 1)} \right) \right]^2.$$

For any $p \in \overline{\mathcal{M}}(\mathcal{L})$ and any integer $D \geq 1$, there exists an extremal point $\overline{p} \in \overline{\mathcal{O}}_{2D+3}(\mathcal{L})$ such that

$$(42) \quad \int_{\mathbb{R}} |p(x) - \overline{p}(x)| dx \leq \frac{2\gamma}{D^2}.$$

Proof. The proof is postponed to Section 8.5. □

Let us comment on condition (40). If we wish to interpret the fact that a density p belongs to the set $\overline{\mathcal{M}}(\mathcal{L})$ as a condition on its shape, it is natural to require that this shape remain unchanged under a location-scale transformation: the density $p_\sigma : x \mapsto \sigma p(\sigma x)$ with $\sigma > 0$ should therefore belong to $\overline{\mathcal{M}}(\mathcal{L})$ if p does. Let

$p \in \overline{\mathcal{M}}(\mathcal{L})$ and assume for the sake of simplicity that both \mathcal{L} and p are differentiable so that the derivatives of $\mathcal{L}p$ and $\mathcal{L}p_\sigma$ are respectively given by

$$(\mathcal{L}p)'(x) = p'(x)\mathcal{L}'[p(x)] \quad \text{and} \quad (\mathcal{L}p_\sigma)'(x) = p'(\sigma x)\sigma^2\mathcal{L}'[\sigma p(\sigma x)] \quad \text{for all } x \in \mathbb{R}.$$

Under the assumption

$$(43) \quad \forall \sigma > 0, \exists c > 0 \quad \text{such that} \quad \sigma^2\mathcal{L}'(\sigma y) = c\mathcal{L}'(y) \quad \text{for all } y > 0,$$

the following equality holds

$$(\mathcal{L}p_\sigma)'(x) = cp'(\sigma x)\mathcal{L}'[p(\sigma x)] = c(\mathcal{L}p)'(\sigma x) \quad \text{for all } x \in \mathbb{R}.$$

Since $(\mathcal{L}p)'$ is nonincreasing because $\mathcal{L}p$ is concave, so is $(\mathcal{L}p_\sigma)'$ and we conclude that under (43) the densities p_σ with $\sigma > 0$ also belong $\overline{\mathcal{M}}(\mathcal{L})$. The set $\overline{\mathcal{M}}(\mathcal{L})$ is stable under a scale transformation. Condition (43) is satisfied for the functions \mathcal{L}_s with $s \in \mathbb{R}$ that are associated to the so-called s -concavity property. In our framework, we do not require that the set $\overline{\mathcal{M}}(\mathcal{L})$ remain invariant under rescaling. From this point of view, the fact that p belongs to $\overline{\mathcal{M}}(\mathcal{L})$ may not be interpreted in general as a genuine condition on the shape of p . Nevertheless, (40) and (43) share some similarities. On the one hand, our condition (40) only requires that the inequality hold in (43) for given values of $\sigma = \lambda$ and $c = \kappa$, on the other hand we impose the constraint that these values be both larger than 1.

The functions $\mathcal{L} = \mathcal{L}_s$ are differentiable on $(0, +\infty)$ and they satisfy the assumptions of Theorem 9 when $s \in (-1, 1]$. More precisely, since $\mathcal{L}'_s(y) = |s|y^{s-1}$ for $s \in (-1, 1] \setminus \{0\}$ and $\mathcal{L}'_0(y) = y^{-1}$ for all $y > 0$, $\lambda^2\mathcal{L}'_s(\lambda y)/\mathcal{L}'_s(y) = \lambda^{s+1}$ for all $\lambda > 1$ and we may then choose $\kappa = \lambda^{s+1}$. Consequently, for a given choice of $\lambda > 1$, the value of γ involved in (42) can be chosen as

$$\gamma = \gamma(\lambda, \lambda^{s+1}) = \left[\frac{2}{(s+1)\log\lambda} \left(\frac{\lambda^{s+1}}{\lambda^{s+1}-1} \right)^{1/2} + \frac{\lambda^{(1-s)/2}\sqrt{\lambda^{s+1}-1}}{\sqrt{2}} \left(1 + \frac{1}{2(\sqrt{\lambda^{s+1}-1})} \right) \right]^2.$$

For values of $s \in \{-1/2, 0, 1/2, 1\}$ and the choice $\lambda = 3.3$, which approximately gives the minimal value of $\lambda \mapsto \gamma(\lambda, \lambda^{s+1})$, we obtain the following upper bounds for $\gamma = \gamma(s)$

$$(44) \quad \gamma(-1/2) < 130, \quad \gamma(0) < 26.5, \quad \gamma(1/2) = 10.1 \quad \text{and} \quad \gamma(1) < 4.8.$$

6.4. Risk bounds of the TV-estimator over sets of \mathcal{L} -concave densities. Combining the risk bound established for the TV-estimator $\hat{P} = \hat{p} \cdot \mu$ in Theorem 8 with the approximation properties of the sets $\overline{\mathcal{O}}_D(\mathcal{L})$ provided by Proposition 7 and Theorem 9, we derive the following result.

Corollary 4. *Let X_1, \dots, X_n be an n -sample with density $p^* \in \overline{\mathcal{M}}(\mathcal{L})$ and $\hat{p} \in \overline{\mathcal{M}}(\mathcal{L})$ the density of the TV-estimator. The following results hold.*

(i) *If \mathcal{L} is convex and increasing on $(0, +\infty)$,*

$$(45) \quad \mathbb{E} [\|p^* - \hat{p}\|] \leq \frac{192}{n^{2/5}} + \frac{96\sqrt{10}}{\sqrt{n}} + \frac{2\varepsilon}{n}.$$

In particular, this result applies to the family of s -concave densities $\overline{\mathcal{M}}(\mathcal{L}_s)$ with $s \geq 1$.

(ii) *if \mathcal{L} is concave, increasing on $(0, +\infty)$ and satisfies (40) for some constants $\lambda, \kappa > 1$,*

$$(46) \quad \mathbb{E} [\|p^* - \hat{p}\|] \leq \frac{192\gamma^{1/5}}{n^{2/5}} + \frac{96\sqrt{14}}{\sqrt{n}} + \frac{2\varepsilon}{n},$$

where the constant γ is given by (41). In particular, this result applies to the family of s -concave densities $\overline{\mathcal{M}}(\mathcal{L}_s)$ with $s \in (-1, 1)$. In the special case $s = 0$, $\overline{\mathcal{M}}(\mathcal{L}_s)$ is the family of log-concave densities and one may take $\gamma = 26.5$.

Proof. For $a, b, c > 0$, define on the set of positive integers D ,

$$F(D) = \frac{a}{D^2} + b\sqrt{\frac{D-1}{n}} + c.$$

Then, for the choice

$$D^* = \left\lceil \left(\frac{a}{b}\right)^{2/5} n^{1/5} \right\rceil \quad \text{which satisfies} \quad 1 \vee \left\lceil \left(\frac{a}{b}\right)^{2/5} n^{1/5} \right\rceil \leq D^* \leq \left\lceil \left(\frac{a}{b}\right)^{2/5} n^{1/5} \right\rceil + 1$$

we obtain that

$$(47) \quad \inf_{D \geq 1} F(D) \leq F(D^*) \leq 2a^{1/5}b^{4/5}n^{-2/5} + c.$$

When \mathcal{L} is convex, it follows from (38), (39) and the subadditivity of the square root that for all $D \geq 1$,

$$\begin{aligned} \mathbb{E} [\|p^* - \hat{p}\|] &= 2\mathbb{E} \left[d(P^*, \hat{P}) \right] \leq \inf_{D \geq 1} \left[3 \inf_{s \in \mathcal{O}_{2D+1}(\mathcal{L})} \|p^* - s\| + 96\sqrt{2}\sqrt{\frac{2D+3}{n}} \right] + \frac{2\varepsilon}{n} \\ &\leq \inf_{D \geq 1} \left[\frac{6}{D^2} + 96\sqrt{2}\sqrt{\frac{2(D-1)+5}{n}} \right] + \frac{2\varepsilon}{n} \leq \inf_{D \geq 1} \left[\frac{6}{D^2} + (96 \times 2)\sqrt{\frac{D-1}{n}} \right] + \frac{96\sqrt{10}}{\sqrt{n}} + \frac{2\varepsilon}{n}. \end{aligned}$$

We obtain (45) by applying (47) with $a = 6$, $b = 96 \times 2$ and $c = 96\sqrt{10}/\sqrt{n} + 2\varepsilon/n$.

When \mathcal{L} is concave and fulfils the requirements of Theorem 9, we may argue similarly by applying (42) in place of (39) and by using the fact that for all $D \geq 1$,

$$\mathbb{E} [\|p^* - \hat{p}\|] \leq \inf_{D \geq 1} \left[\frac{6\gamma}{D^2} + 96\sqrt{2}\sqrt{\frac{2(D-1)+7}{n}} \right] + \frac{2\varepsilon}{n} \leq \inf_{D \geq 1} \left[\frac{6\gamma}{D^2} + 96 \times 2\sqrt{\frac{D-1}{n}} \right] + \frac{96\sqrt{14}}{\sqrt{n}} + \frac{2\varepsilon}{n}.$$

□

7. CONCLUDING REMARKS

7.1. About the computational aspects of the TV-estimators. The estimation procedure that we propose here requires very few assumptions on the statistical model and the density to be estimated. From this point of view, it provides a competitor to ρ -estimators. As for the latter, it is unlikely that an algorithm for calculating TV-estimators can be designed at this level of generality. A more reasonable approach is to see how the estimator can be calculated in a specific situation that is, for a given statistical model, as we would do for the MLE. It is likely that the complexity of the algorithm will be related not only to the size of the model the statistician considers but also to the assumptions he or she wants to make on the target density. For illustration, let us consider the case of the Grenander estimator. In this situation, the statistician has a nice expression of the optimizer of the likelihood function because he or she has beforehand restricted the optimization to the set of densities that are supported on a given half-line (restriction to a specific model) and assumed that all the data do belong to this half-line (assumption on the target density). The MLE would not exist otherwise. Under the same assumptions, it is worth noticing that the calculation of the ρ -estimator becomes tractable as well (we merely recovers the MLE). The drawback of these somewhat restrictive assumptions, which make the calculability of the MLE possible and easy (as well as that of the ρ -estimator), lie in the fact that a very small error on the support of the density may result in a large distance between the true density and the model and, in turn, to a poor performance of these estimators. This is what we have shown in our Introduction. By considering the whole set of nonincreasing densities on a half-line, the TV- (and ρ -) estimator would not suffer from these weaknesses but may be more difficult to compute. This example shows that it seems difficult to disentangle the computational issues from the properties we wish the resulting estimator to achieve. We believe that these computational aspects, of both the TV- and

ρ -estimators, are an area of research on their own and that solving these computational issues may require different sets of assumptions from those which are needed to establish their mathematical properties.

7.2. About the limitation of the approach for estimating a density under a shape constraint.

Our approach offers the advantage that it solves in a uniform framework various problems of estimation of a density under a shape constraint. It provides an estimator that is well-defined, even for sets of densities on which the MLE does not exist, and it leads to an estimator that enjoys some desirable adaptation properties. When the data are i.i.d. with a distribution P_0 whose density has the expected shape as well as some additional features, the risk bound that we establish for our TV-estimator is of order $1/\sqrt{n}$. Interestingly, the estimator is robust to a departure from this ideal situation. When the data are only independent but the average $P^\star = n^{-1} \sum_{i=1}^n P_i^\star$ of their marginals lies close enough to P_0 , say in a ball (with respect to the total variation distance) centered at P_0 with radius $r > 0$, the risk bound of the TV-estimator does not deteriorate by more than the additional term $3r$. This stability of the risk is the key ingredient that allows us to establish not only the robustness of our estimator but also, by means of approximation theory, some uniform \mathbb{L}_1 -risk bounds over classes of densities of interest. However, this versatile approach suffers from some restrictions: it mainly applies to densities on the line. This is due to the fact that the level sets of unimodal, convex, concave or log-concave functions on the line are quite simple. They take the form of an interval or a union of two intervals. The situation dramatically changes in higher dimensions and the notion of extremal points, that we have introduced here, becomes much less interesting. In dimension 2, the class of all uniform densities on polygons, for example, does not possess any extremal points. It is unclear to us whether or not this limitation to the line is due to our approach or is rather inherent to the problem of robustness (or stability) which we want to solve.

8. PROOFS

8.1. Some useful lemmas for approximating densities and convex functions with the \mathbb{L}_1 -norm.

The following result shows that normalizing to one a nonnegative approximating function of a density cannot worsen its \mathbb{L}_1 -error by more than a factor 2.

Lemma 3. *Let p be a density on a measured space (E, \mathcal{E}, ν) and f a nonnegative integrable function on (E, \mathcal{E}, ν) which is not ν -a.e. equal to 0 on E . Then*

$$\int_E \left| p - \frac{f}{\int_E f d\nu} \right| d\nu \leq 2 \left[1 \wedge \int_E |p - f| d\nu \right].$$

This inequality cannot be improved in general since equality holds when $f = p\mathbb{1}_I$ and I is a measurable subset of E on which p is not ν -a.e. equal to 0.

Proof. The fact that

$$\int_E \left| p - \frac{f}{\int_E f d\nu} \right| d\nu \leq 2$$

comes from the triangle inequality. Let us now prove that

$$\int_E \left| p - \frac{f}{\int_E f d\nu} \right| d\nu \leq 2 \int_E |p - f| d\nu.$$

We first assume that $c = \int_E f d\nu \in (0, 1]$. Since f/c and p are two densities,

$$\int_E \left| \frac{f}{c} - p \right| d\nu = 2 \int_I \left[p - \frac{f}{c} \right] \mathbb{1}_{\{cp \geq f\}} d\nu \leq 2 \int_E [p - f] \mathbb{1}_{\{cp \geq f\}} d\nu \leq 2 \int_E |p - f| d\nu.$$

This proves the lemma when $c \in (0, 1]$. let us now assume that $c > 1$. The previous case of the lemma applies to the density f/c and the nonnegative function p/c the integral of which is not larger than 1. This yields

$$\int_E \left| \frac{f}{c} - p \right| d\nu \leq 2 \int_E \left| \frac{f}{c} - \frac{p}{c} \right| d\nu = \frac{2}{c} \int_I |f - p| d\nu \leq 2 \int_E |f - p| d\nu.$$

□

The following lemma applies to approximation functions that are unimodal.

Lemma 4. *Let f be an integrable function which is nondecreasing on $(-\infty, c)$ and nonincreasing on $(c, +\infty)$ for some $c \in \mathbb{R}$, and such that $\kappa = \int_{-\infty}^{+\infty} f(x) dx \in (0, 1]$. Then, for any probability density p on the line,*

$$\int_{-\infty}^{+\infty} |p(x) - f(c + \kappa(x - c))| dx \leq 2 \int_{-\infty}^{+\infty} |p(x) - f(x)| dx.$$

The constant 2 is optimal as shown by the following example. Given $\varepsilon \in (0, 1)$, let $f = \varepsilon \mathbb{1}_{[-1/2, 1/2]}$ and $p = \mathbb{1}_{[1/\varepsilon, 1+1/\varepsilon]}$ so that $\kappa = \varepsilon$ and one may take $c = 0$. The function $x \mapsto f(\kappa x)$ is then the uniform density on $[-1/(2\varepsilon), 1/(2\varepsilon)]$ and its support, as that of f , is disjoint from that of p . We conclude that

$$\int_{-\infty}^{+\infty} |p(x) - f(\kappa x)| dx = 2 \quad \text{and} \quad \int_{-\infty}^{+\infty} |p(x) - f(x)| dx = 1 + \varepsilon.$$

Proof. Since the Lebesgue measure is translation invariant, we may assume without loss of generality that $c = 0$. Let $x \in \mathbb{R}$. If $x < 0$, $\kappa x > x$ and $f(\kappa x) \geq f(x)$ since f is nondecreasing on $(-\infty, 0)$ and if $x > 0$, $\kappa x < x$ and $f(\kappa x) \geq f(x)$ since f is nonincreasing on $(0, +\infty)$. We deduce that for all $x \in \mathbb{R}$, $f(\kappa x) \geq f(x)$ and consequently,

$$\begin{aligned} \int_{-\infty}^{+\infty} |f(\kappa x) - f(x)| dx &= \int_{-\infty}^{+\infty} f(\kappa x) dx - \int_{-\infty}^{+\infty} f(x) dx = \frac{1}{\kappa} \int_{-\infty}^{+\infty} f(x) dx - \int_{-\infty}^{+\infty} f(x) dx \\ &= 1 - \int_{-\infty}^{+\infty} f(x) dx = \int_{-\infty}^{+\infty} p(x) dx - \int_{-\infty}^{+\infty} f(x) dx \leq \int_{-\infty}^{+\infty} |p(x) - f(x)| dx. \end{aligned}$$

Using the triangle inequality, we conclude that

$$\int_{-\infty}^{+\infty} |f(\kappa x) - p(x)| dx \leq \int_{-\infty}^{+\infty} |f(\kappa x) - f(x)| dx + \int_{-\infty}^{+\infty} |p(x) - f(x)| dx \leq 2 \int_{-\infty}^{+\infty} |p(x) - f(x)| dx.$$

□

The following result shows how well a convex (or concave) function can be approximated by its chord, that is, its 1-linear interpolation on a compact interval.

Proposition 8. *Let f be a convex-concave continuous function on a bounded nontrivial interval $[a, b]$ and ℓ_f be the linear function*

$$(48) \quad \ell_f : x \mapsto f(a) + \frac{f(b) - f(a)}{b - a} (x - a).$$

The following results hold.

(i) *If f admits a right derivative $f'_r(a)$ at a and a left derivative $f'_l(b)$ at b ,*

$$(49) \quad \int_a^b |f - \ell_f| d\mu \leq \frac{(b - a)^2}{8} |f'_r(a) - f'_l(b)|.$$

(ii) If f is strictly monotone on $[a, b]$ with $f'_r(a) \neq 0$ and $f'_l(b) \neq 0$

$$(50) \quad \int_a^b |f - \ell_f| d\mu \leq \frac{(f(b) - f(a))^2}{8} \left| \frac{1}{f'_r(a)} - \frac{1}{f'_l(b)} \right|,$$

with the convention $1/(\pm\infty) = 0$.

Proof. Let us first assume that f is concave on $[0, 1]$, admits a right derivative at 0, a left derivative at 1 and satisfies $f(0) = 0$ and $f(1) = 1$. Let us show that

$$(51) \quad \int_0^1 |f - \ell_f| d\mu \leq \frac{1}{2} \frac{(f'_r(0) - 1)(1 - f'_l(1))}{f'_r(0) - f'_l(1)},$$

with the convention $0/0 = 1$. If f is linear on $[0, 1]$, then $f'_r(0) = f'_l(1) = 1$ and the inequality is satisfied with our convention. Otherwise, $f'_l(1) < 1 = f(1) - f(0) < f'_r(0)$ and since f lies above its chord and under its tangents at 0 and 1,

$$\ell_f(x) \leq f(x) \leq \min \{f'_r(0)x, 1 + f'_l(1)(x - 1)\} \quad \text{for all } x \in [0, 1].$$

Since $\ell_f(x) = x$, we deduce that for all $c \in [0, 1]$

$$\begin{aligned} \int_0^1 |f - \ell_f| d\mu &\leq \int_0^c (f'_r(0) - 1)x d\mu + \int_c^1 (1 - f'_l(1))(1 - x) d\mu \\ &= \frac{1}{2} [(f'_r(0) - 1)c^2 + (1 - f'_l(1))(1 - c)^2]. \end{aligned}$$

The result follows by minimizing with respect to c , i.e. by taking $c = (1 - f'_l(1))/(f'_r(0) - f'_l(1)) \in (0, 1)$. In particular, we deduce from (52) that

$$\int_0^1 |f - \ell_f| d\mu \leq \frac{c(1 - c)}{2} (f'_r(0) - f'_l(1))$$

and since $c(1 - c) \leq 1/4$, we obtain that

$$(52) \quad \int_0^1 |f - \ell_f| d\mu \leq \frac{1}{8} (f'_r(0) - f'_l(1)).$$

Note that the inequality also holds when $f'_r(0) = f'_l(1) = 1$.

When f is increasing on $[0, 1]$ and satisfies $0 < 1 - f'_l(1) < 1$, i.e. $f'_l(1) \neq 0$, we also deduce from (52) and the convexity of $z \mapsto 1/z$ on $(0, +\infty)$ that

$$\begin{aligned} \int_0^1 |f - \ell_f| d\mu &\leq \frac{1}{4} \frac{1}{\frac{1}{2} \left[\left(\frac{1}{1 - f'_l(1)} - 1 \right) + \left(1 + \frac{1}{f'_r(0) - 1} \right) \right]} \leq \frac{1}{4} \left[\frac{1}{2} \left(\frac{1}{\frac{1}{1 - f'_l(1)} - 1} + \frac{1}{1 + \frac{1}{f'_r(0) - 1}} \right) \right] \\ &= \frac{1}{8} \left[\frac{1 - f'_l(1)}{f'_l(1)} + \frac{f'_r(0) - 1}{f'_r(0)} \right] \end{aligned}$$

which leads to

$$(53) \quad \int_0^1 |f - \ell_f| d\mu \leq \frac{1}{8} \left(\frac{1}{f'_l(1)} - \frac{1}{f'_r(0)} \right).$$

Note that the inequality is still satisfied when $f'_r(0) = +\infty$ with the convention $1/(+\infty) = 0$.

Let us now turn to the proofs of (49) and (50). Note that (49) is clearly true when f is constant on $[a, b]$ and we may therefore assume that $f(a) \neq f(b)$. We obtain (49) and (50) by applying (52) and (53) respectively to the function

$$g(x) = \frac{f(a + x(b - a)) - f(a)}{f(b) - f(a)}$$

when f is concave and satisfies $f(b) > f(a)$ or when f is convex and satisfies $f(a) > f(b)$. In the other cases, one may use the function

$$g(x) = \frac{f(b - x(b - a)) - f(b)}{f(a) - f(b)}.$$

□

The following result is an extension of Proposition 8. It shows how well a convex (or concave) function can be approximated by a suitable D -linear interpolation on a compact interval.

Proposition 9. *Let D be some positive integer and f a convex-concave function on a non-trivial interval $[a, b]$ such that*

$$\Delta = \frac{f(b) - f(a)}{b - a} \neq 0.$$

(i) *If f admits a right derivative at a and a left derivative at b , there exists a D -linear interpolation ℓ_1 of f such that*

$$(54) \quad \int_a^b |f - \ell_1| d\mu \leq \frac{(b - a)^2}{2D^2} \frac{|f'_r(a) - \Delta| |\Delta - f'_l(b)|}{|f'_r(a) - f'_l(b)|} \leq \frac{(b - a)^2}{8D^2} |f'_r(a) - f'_l(b)|,$$

with the convention $0/0 = 0$ in the right-hand side of (54).

(ii) *If f is strictly monotone on $[a, b]$ with $f'_r(a) \neq 0$ and $f'_r(b) \neq 0$, there exists a D -linear interpolation ℓ_2 of f*

$$(55) \quad \int_a^b |f - \ell_2| d\mu \leq \frac{(f(a) - f(b))^2}{8D^2} \left| \frac{1}{f'_r(a)} - \frac{1}{f'_l(b)} \right|$$

with the convention $1/(\pm\infty) = 0$.

(iii) *If f is monotone and convex or concave on $[a, b]$, there exists a D -linear interpolation ℓ_3 of f such that*

$$(56) \quad \int_a^b |f - \ell_3| d\mu \leq \frac{(b - a) |f(a) - f(b)|}{2D^2}.$$

Proof. If f is a concave and monotone function on $[0, 1]$ satisfying $f(0) = 0$ and $f(1) = 1$, Guérin *et al* [20] proved that there exists a D -linear interpolation ℓ of f on $[0, 1]$ such that

$$\int_0^1 |f - \ell| d\mu \leq \frac{1}{2D^2}.$$

If f admits a right derivative at 0 and a left derivative at 1, they proved that there exists a D -linear interpolation ℓ of f on $[0, 1]$ such that

$$(57) \quad \int_0^1 |f - \ell| d\mu \leq \frac{1}{2D^2} \frac{(f'_r(0) - 1)(1 - f'_l(1))}{f'_r(0) - f'_l(1)}.$$

The results established in Proposition 9 are deduced from (57) by arguing as in the proof of Proposition 8. □

8.2. Proofs of Section 3.

Proof of Theorem 1. Let $\mathfrak{D} \geq 1$ such that $\overline{\mathcal{O}}(\mathfrak{D})$ is not empty. Such an integer \mathfrak{D} exists since $\overline{\mathcal{O}}$ is nonempty. Let $\overline{P} = \overline{p} \cdot \mu$ be an arbitrary point in $\mathcal{O}(\mathfrak{D})$ with $\overline{p} \in \mathcal{O}(\mathfrak{D})$. For $P, Q \in \mathcal{M}$ and $\zeta \geq 0$, we set

$$\begin{aligned} \mathbf{Z}_+(\mathbf{X}, P) &= \sup_{Q \in \mathcal{M}} [\mathbf{T}(\mathbf{X}, P, Q) - \mathbb{E}[\mathbf{T}(\mathbf{X}, P, Q)]] - \zeta \\ \mathbf{Z}_-(\mathbf{X}, P) &= \sup_{Q \in \mathcal{M}} [\mathbb{E}[\mathbf{T}(\mathbf{X}, Q, P)] - \mathbf{T}(\mathbf{X}, Q, P)] - \zeta \end{aligned}$$

and

$$\mathbf{Z}(\mathbf{X}, P) = \mathbf{Z}_+(\mathbf{X}, P) \vee \mathbf{Z}_-(\mathbf{X}, P).$$

Applying the first inequality of (6) with $P = Q$ and $Q = \overline{P}$, we infer that for all $Q \in \mathcal{M}$,

$$\begin{aligned} nd(\overline{P}, Q) &\leq nd(P^*, \overline{P}) + \mathbb{E}[\mathbf{T}(\mathbf{X}, Q, \overline{P})] \\ &= nd(P^*, \overline{P}) + \mathbb{E}[\mathbf{T}(\mathbf{X}, Q, \overline{P})] - \mathbf{T}(\mathbf{X}, Q, \overline{P}) + \mathbf{T}(\mathbf{X}, Q, \overline{P}) \\ &\leq nd(P^*, \overline{P}) + \mathbf{Z}(\mathbf{X}, \overline{P}) + \mathbf{T}(\mathbf{X}, Q, \overline{P}) + \zeta \\ &\leq nd(P^*, \overline{P}) + \mathbf{Z}(\mathbf{X}, \overline{P}) + \mathbf{T}(\mathbf{X}, Q) + \zeta. \end{aligned}$$

In particular, the inequality applies to $Q = \widehat{P} \in \mathcal{E}(\mathbf{X})$ and using the fact that

$$\mathbf{T}(\mathbf{X}, \widehat{P}) \leq \inf_{P \in \mathcal{M}} \mathbf{T}(\mathbf{X}, P) + \varepsilon \leq \mathbf{T}(\mathbf{X}, \overline{P}) + \varepsilon,$$

we deduce that

$$(58) \quad nd(\overline{P}, \widehat{P}) \leq nd(P^*, \overline{P}) + \mathbf{Z}(\mathbf{X}, \overline{P}) + \mathbf{T}(\mathbf{X}, \overline{P}) + \zeta + \varepsilon.$$

Using now the second inequality of (6) with $P = \overline{P}$, we obtain that

$$\begin{aligned} \mathbf{T}(\mathbf{X}, \overline{P}) &= \sup_{Q \in \mathcal{M}} \mathbf{T}(\mathbf{X}, \overline{P}, Q) \\ &\leq \sup_{Q \in \mathcal{M}} [\mathbf{T}(\mathbf{X}, \overline{P}, Q) - \mathbb{E}[\mathbf{T}(\mathbf{X}, \overline{P}, Q)] - \zeta] + \sup_{Q \in \mathcal{M}} \mathbb{E}[\mathbf{T}(\mathbf{X}, \overline{P}, Q)] + \zeta \\ &\leq \mathbf{Z}(\mathbf{X}, \overline{P}) + nd(P^*, \overline{P}) + \zeta, \end{aligned}$$

which with (58) lead to

$$(59) \quad nd(\overline{P}, \widehat{P}) \leq 2nd(P^*, \overline{P}) + 2\zeta + \varepsilon + 2\mathbf{Z}(\mathbf{X}, \overline{P}).$$

Let us now bound from above $\mathbf{Z}(\mathbf{X}, \overline{P})$. We set for $P \in \mathcal{M}$

$$\mathbf{w}(P) = \mathbb{E} \left[\sup_{Q \in \mathcal{M}} [\mathbf{T}(\mathbf{X}, P, Q) - \mathbb{E}[\mathbf{T}(\mathbf{X}, P, Q)]] \vee \mathbb{E} \left[\sup_{Q \in \mathcal{M}} [\mathbb{E}[\mathbf{T}(\mathbf{X}, Q, P)] - \mathbf{T}(\mathbf{X}, Q, P)] \right] \right].$$

The functions $t_{(\overline{P}, Q)}$ satisfy $|t_{(\overline{P}, Q)}(x) - t_{(\overline{P}, Q)}(x')| \leq 1$ for all $Q \in \mathcal{M}$ and $x, x' \in E$, hence

$$|\mathbf{Z}_+((x_1, \dots, x_i, \dots, x_n), \overline{P}) - \mathbf{Z}_+((x_1, \dots, x'_i, \dots, x_n), \overline{P})| \leq 1$$

for all $\mathbf{x} \in \mathbf{E}$, $x'_i \in E$ and $i \in \{1, \dots, n\}$. Arguing as in the proof of Lemma 2 of Baraud [3] (with $\xi + \log 2$ in place of ξ), we deduce that with a probability at least $1 - (1/2)e^{-\xi}$,

$$(60) \quad \begin{aligned} \mathbf{Z}_+(\mathbf{X}, \bar{P}) &\leq \mathbb{E} [\mathbf{Z}_+(\mathbf{X}, \bar{P})] + \sqrt{\frac{n(\xi + \log 2)}{2}} \\ &= \mathbb{E} \left[\sup_{Q \in \mathcal{M}} [\mathbf{T}(\mathbf{X}, \bar{P}, Q) - \mathbb{E} [\mathbf{T}(\mathbf{X}, \bar{P}, Q)]] \right] + \sqrt{\frac{n(\xi + \log 2)}{2}} - \zeta \end{aligned}$$

$$(61) \quad \leq \mathbf{w}(\bar{P}) + \sqrt{\frac{n(\xi + \log 2)}{2}} - \zeta.$$

Arguing similarly, with a probability at least $1 - (1/2)e^{-\xi}$,

$$(62) \quad \mathbf{Z}_-(\mathbf{X}, \bar{P}) \leq \mathbf{w}(\bar{P}) + \sqrt{\frac{n(\xi + \log 2)}{2}} - \zeta.$$

Putting (61) and (62) together and choosing $\zeta = \mathbf{w}(\bar{P}) + \sqrt{n(\xi + \log 2)/2}$, we obtain that with a probability at least $1 - e^{-\xi}$,

$$\mathbf{Z}(\mathbf{X}, \bar{P}) = \mathbf{Z}_+(\mathbf{X}, \bar{P}) \vee \mathbf{Z}_-(\mathbf{X}, \bar{P}) \leq \mathbf{w}(\bar{P}) + \sqrt{\frac{n(\xi + \log 2)}{2}} - \zeta \leq 0$$

which with (59) lead to the bound

$$(63) \quad d(\bar{P}, \hat{P}) \leq 2d(P^*, \bar{P}) + \frac{2\mathbf{w}(\bar{P})}{n} + \sqrt{\frac{2(\xi + \log 2)}{n}} + \frac{\varepsilon}{n}.$$

It remains now to control $\mathbf{w}(\bar{P})$. Since $\bar{p} \in \mathcal{O}(\mathfrak{D}) \subset \overline{\mathcal{O}}(\mathfrak{D})$, it is extremal in $\overline{\mathcal{M}} \supset \mathcal{M}$ with degree not larger than \mathfrak{D} , the classes $\{q < \bar{p}\}$, $q \in \mathcal{M} \setminus \{\bar{p}\}$ and $\{q > \bar{p}\}$, $q \in \mathcal{M} \setminus \{\bar{p}\}$ are both VC with dimensions not larger than \mathfrak{D} . We may therefore apply Proposition 3.1 in Baraud [2] with $\sigma = 1$ and get

$$(64) \quad \mathbb{E} \left[\sup_{q \in \mathcal{M} \setminus \{\bar{p}\}} \left| \sum_{i=1}^n (\mathbb{1}_{\bar{p} > q}(X_i) - P_i^*(\bar{p} > q)) \right| \right] \leq 10\sqrt{5n\mathfrak{D}},$$

and

$$(65) \quad \mathbb{E} \left[\sup_{q \in \mathcal{M} \setminus \{\bar{p}\}} \left| \sum_{i=1}^n (\mathbb{1}_{\bar{p} < q}(X_i) - P_i^*(\bar{p} < q)) \right| \right] \leq 10\sqrt{5n\mathfrak{D}}.$$

These inequalities entail $\mathbf{w}(\bar{P}) \leq 10\sqrt{5n\mathfrak{D}}$, and we infer from (63) that

$$(66) \quad d(\bar{P}, \hat{P}) \leq 2d(P^*, \bar{P}) + 20\sqrt{\frac{5\mathfrak{D}}{n}} + \sqrt{\frac{2(\xi + \log 2)}{n}} + \frac{\varepsilon}{n}.$$

Since \bar{P} is arbitrary in the set $\mathcal{O}(\mathfrak{D})$ which is dense on $\overline{\mathcal{O}}(\mathfrak{D})$, we infer that equation (66) holds for all $\bar{P} \in \overline{\mathcal{O}}(\mathfrak{D})$, which yields (7). Hence, by the triangle inequality,

$$\begin{aligned} d(P^*, \hat{P}) &\leq \inf_{P \in \overline{\mathcal{O}}(\mathfrak{D})} \left\{ d(P^*, P) + d(P, \hat{P}) \right\} \\ &\leq 3 \inf_{P \in \overline{\mathcal{O}}(\mathfrak{D})} d(P^*, P) + 20\sqrt{\frac{5\mathfrak{D}}{n}} + \sqrt{\frac{2(\xi + \log 2)}{n}} + \frac{\varepsilon}{n}. \end{aligned}$$

With our convention that $\inf_{\emptyset} = +\infty$, the inequality is also true when $\overline{\mathcal{O}}(\mathfrak{D}) = \emptyset$, hence for all values of \mathfrak{D} , which leads to (8). Inequality (10) follows by integrating this deviation bound with respect to ξ . \square

8.3. Proofs of Section 4.

Proof of Proposition 3. We restrict ourselves to the case where p is nonincreasing on I , the proof in the other case is similar. Let q be the function that coincides with p on $\overset{\circ}{I} = (a, b)$ and satisfies $q(a) = \sup_{x \in (a, b)} p(x)$ and $q(b) = \inf_{x \in (a, b)} p(x)$. Clearly, $p = q$ a.e. and satisfies $V_I(p) = q(a) - q(b) = V_I(q)$. Without loss of generality, we may therefore assume that $I = [a, b]$ and that $V_I(p) = p(a) - p(b)$.

Since p is nonincreasing in I , for all intervals $J \subset I$ with endpoints $u < v$,

$$(67) \quad \int_J |p - \bar{p}_J| d\mu \leq \frac{(v - u)(p(u) - p(v))}{2}.$$

In particular, when $D = 1$ it suffices to take $\mathcal{J} = \{I\}$ and the result follows from (67) with $u = a$ and $v = b$ and the trivial inequality $\int_I |p - \bar{p}_J| d\mu \leq 2$. It remains to prove the result for $D \geq 2$ and since (20) is trivially true for $V = 0$ we may also assume that $V > 0$.

Let us set $\eta = (1 + VL)^{1/D} - 1 > 0$, $x_0 = a$ and for all $j \in \{1, \dots, D\}$,

$$x_j = x_{j-1} + L \frac{(1 + \eta)^j}{\sum_{k=1}^D (1 + \eta)^k} = x_0 + L \frac{(1 + \eta)^j - 1}{(1 + \eta)^D - 1} = x_0 + \frac{(1 + \eta)^j - 1}{V}.$$

Then, we obtain an increasing sequence of points $a = x_0 < x_1 < \dots < x_D = b = a + L$ and a partition \mathcal{J} of I into D intervals based on $\{x_0, \dots, x_D\}$. Using (67) and the facts that $x_{j+1} - x_j = (1 + \eta)(x_j - x_{j-1}) > x_j - x_{j-1}$ for $j \in \{1, \dots, D - 1\}$, we obtain

$$\begin{aligned} \int_I |p - \bar{p}| d\mu &= \sum_{J \in \mathcal{J}} \int_J |p - \bar{p}_J| d\mu \leq \frac{1}{2} \sum_{j=1}^D (x_j - x_{j-1}) [p(x_{j-1}) - p(x_j)] \\ &= \frac{1}{2} \left[p(x_0)(x_1 - x_0) + \sum_{j=1}^{D-1} p(x_j) [(x_{j+1} - x_j) - (x_j - x_{j-1})] \right] - \frac{p(x_D)(x_D - x_{D-1})}{2} \\ &\leq \frac{1}{2} \left[V_I(p)(x_1 - x_0) + \eta \sum_{j=1}^{D-1} p(x_j)(x_j - x_{j-1}) \right] + \frac{p(x_D) [(x_1 - x_0) - (x_D - x_{D-1})]}{2} \\ &\leq \frac{1}{2} \left[V(x_1 - x_0) + \eta \sum_{j=1}^{D-1} \int_{x_{j-1}}^{x_j} p d\mu \right] \leq \frac{1}{2} [V(x_1 - x_0) + \eta] = \eta. \end{aligned}$$

Together with the trivial bound $\int_I |p - \bar{p}| d\mu \leq 2$, this last inequality leads to (20). The second inequality derives from the fact that $(e^x - 1) \wedge 2 \leq 2x / \log 3 \leq 2x$ for all $x \geq 0$. \square

Proof of Theorem 3. Let D, D_1, \dots, D_{k-2} be positive integers and p a density in $\overline{\mathcal{M}}_k^\infty(R)$. We may therefore write p in the form (17) with

$$\left[\sum_{i=1}^{k-2} \sqrt{w_i \log(1 + L_i V_i)} \right]^2 \leq R.$$

Applying Proposition 3 to the density p_i , with $I = I_i = (x_{i-1}, x_i)$, $L = L_i = (x_i - x_{i-1})$ and $V = V_i$ for each $i \in \{1, \dots, k - 2\}$, we build a monotone density \bar{p}_i on I_i which is piecewise constant on a partition of I_i into $D_i \geq 1$ nontrivial intervals and that satisfies

$$\int_{I_i} |p_i - \bar{p}_i| d\mu \leq \frac{2S_i}{D_i} \quad \text{with} \quad S_i = \log(1 + V_i L_i).$$

Let us now take $D_i = \left\lceil D\sqrt{w_i S_i} / (\sum_{i=1}^{k-2} \sqrt{w_i S_i}) \right\rceil \vee 1$ for all $i \in \{1, \dots, k-2\}$. Since

$$\frac{D\sqrt{w_i S_i}}{\sum_{i=1}^{k-2} \sqrt{w_i S_i}} \vee 1 \leq D_i \leq \frac{D\sqrt{w_i S_i}}{\sum_{i=1}^{k-2} \sqrt{w_i S_i}} + 1,$$

the density $\bar{p} = \sum_{i=1}^{k-2} w_i \bar{p}_i$ satisfies

$$\|p - \bar{p}\| \leq \sum_{i=1}^{k-2} w_i \int_{I_i} |p_i - \bar{p}_i| d\mu \leq \sum_{i=1}^{k-2} \frac{2w_i S_i}{D_i} \leq \frac{2}{D} \left(\sum_{i=1}^{k-2} \sqrt{w_i S_i} \right)^2 \leq \frac{2R}{D}.$$

Besides, the density \bar{p} is k -piecewise monotone, supported on $[x_1, x_{k-2}]$ and piecewise constant on a partition of \mathbb{R} consisting of at most $\sum_{i=1}^{k-2} D_i \leq D + k - 2$ bounded intervals. It therefore belongs to $\overline{\mathcal{O}}_{D+k-2, k}$. Finally, let us choose

$$D = \left\lceil \left(\frac{9R^2 n}{83.2^2} \right)^{1/3} \right\rceil \leq \left(\frac{9R^2 n}{83.2^2} \right)^{1/3} + 1.$$

Using the sub-additivity property of the square root, we deduce from (16) that

$$\begin{aligned} \mathbb{B}_{k,n}(p) &\leq \frac{3R}{2D} + 83.2 \sqrt{\frac{D-1+2k}{n}} \\ &\leq \frac{3^{1/3} \times 83.2^{2/3}}{2} \left(\frac{R}{n} \right)^{1/3} + \frac{83.2}{\sqrt{n}} \sqrt{\left(\frac{9nR^2}{83.2^2} \right)^{1/3} + 2k} \\ &\leq \left[\frac{3^{1/3} \times 83.2^{2/3}}{2} + 3^{1/3} \times 83.2^{2/3} \right] \left(\frac{R}{n} \right)^{1/3} + 83.2 \sqrt{\frac{2k}{n}}, \end{aligned}$$

which is (19). \square

Proof of Theorem 4. Let us start with the following lemma where we show that the mapping $\tau(p, \cdot)$ controls the \mathbb{L}_1 -approximation error of a monotone density p by the elements of the class $\overline{\mathcal{M}}_3^\infty(R)$.

Lemma 5. *Let p be a density on \mathbb{R} , B some positive number and I a subset of \mathbb{R} on which the density p is not a.e. equal to 0. The density $p \hat{p}_I^B = (p \wedge B) \mathbf{1}_I / \int_I (p \wedge B) d\mu$ satisfies,*

$$(68) \quad \left\| p - p \hat{p}_I^B \right\| \leq 2 \left[\int_I (p - B)_+ d\mu + \int_{I^c} p d\mu \right].$$

If p is a monotone density on a half-line

$$(69) \quad \inf_{\bar{p} \in \overline{\mathcal{M}}_3^\infty(R)} \|p - \bar{p}\| \leq 2\tau(p, \exp(R) - 1) \quad \text{for all } R \geq \log 2.$$

Besides, if p is a nonincreasing density on $(a, a+l)$ (respectively a nondecreasing density on $(a-l, a)$) with $a \in \mathbb{R}$ and $l \in (0, +\infty]$, we may restrict the infimum to the nonincreasing densities on $(a, a+l)$ (respectively the nondecreasing densities on $(a-l, a)$) that belong to $\overline{\mathcal{M}}_3^\infty(R)$.

Proof. Since p is not equal to 0 a.e. on I , $\int_I (p \wedge B) d\mu > 0$ and $p \hat{p}_I^B$ is therefore a well-defined density on I . By Lemma 3,

$$\int_{\mathbb{R}} |p - p \hat{p}_I^B| d\mu \leq 2 \int_{\mathbb{R}} |p - (p \wedge B) \mathbf{1}_I| d\mu = 2 \int_{I^c} p d\mu + 2 \int_I (p - B)_+ d\mu.$$

Changing p into $x \mapsto p(-x)$ if necessary and possibly changing the value of p at the endpoint of the half-line, we may assume without loss of generality that p is a nonincreasing density on a half-line of the form $(a, +\infty)$ with $a \in \mathbb{R}$. Let us now set

$$l = \sup\{z > 0, p(a+z) > 0\} \in (0, +\infty] \quad \text{and} \quad t = \exp(R) - 1 \geq 1.$$

We first consider the case where $l = +\infty$. Given $s > 0$, let us take $B = p(a+s) > 0$ and $I = (a, a+st)$. Since p is nonincreasing on $(a, +\infty)$, $p(x) \geq p(a+s) = B$ for all $x \in (a, a+s) \subset I$ and consequently

$$\int_I (p \wedge B) d\mu \geq \int_a^{a+s} (p \wedge B) d\mu = sB.$$

The density $\bar{p}_s = p \wedge_I^B$ belongs to $\overline{\mathcal{M}}_3^\infty$, is supported on an interval of length not larger than st and its variation on I is not larger than

$$\frac{B}{\int_I (p \wedge B) d\mu} - p(a+st) < \frac{B}{\int_I (p \wedge B) d\mu}.$$

Hence, it follows from (18) that

$$\mathbf{R}_{k,0}(\bar{p}_s) < \log \left(1 + \frac{tsB}{\int_I (p \wedge B) d\mu} \right) \leq \log(1+t) = R$$

and consequently, $\bar{p}_s \in \overline{\mathcal{M}}_3^\infty(R)$. Applying (68) and Lemma 2, we obtain that for all $s > 0$

$$(70) \quad \inf_{\bar{p} \in \overline{\mathcal{M}}_3^\infty(R)} \|p - \bar{p}\| \leq \|p - \bar{p}_s\| \leq 2[\tau_x(p, st) + \tau_y(p, p(a+s))]$$

and we derive (69) from (24). Since for all $s > 0$, \bar{p}_s is a density on $(a, +\infty)$, we may restrict the infimum to these densities in $\overline{\mathcal{M}}_3^\infty(R)$ that satisfy this property.

Let us now turn to the case where $l < +\infty$ and define $s_0 = l/t \leq l$. Given $s \in (0, s_0)$, we take $B = p(a+s) > 0$ and $I = (a, a+st)$. Since $st < l$, $p(a+st) > 0$ and by arguing as before, we obtain that

$$\inf_{\bar{p} \in \overline{\mathcal{M}}_3^\infty(R)} \|p - \bar{p}\| \leq 2[\tau_x(p, st) + \tau_y(p, p(a+s))] \quad \text{for all } s \in (0, s_0).$$

It follows from the monotonicity of $\tau_y(p, \cdot)$ that for all $0 < s < s_0 \leq s'$,

$$\tau_y(p, p(a+s)) \leq \tau_y(p, p(a+s_0)) \leq \tau_y(p, p(a+s')),$$

and since the mapping $u \mapsto \tau_x(p, u)$ is continuous and nonincreasing, for all $s' \geq s_0$

$$\begin{aligned} & \inf_{s \in (0, s_0)} [\tau_x(p, st) + \tau_y(p, p(a+s))] \\ & \leq \inf_{s \in (0, s_0)} \tau_x(p, st) + \tau_y(p, p(a+s')) = \tau_x(p, s_0 t) + \tau_y(p, p(a+s')) \\ & = 0 + \tau_y(p, p(a+s')) = \tau_x(p, s' t) + \tau_y(p, p(a+s')). \end{aligned}$$

Consequently, (70) remains satisfied for all $s > 0$. Since it is actually enough to restrict the infimum to those $s \in (0, s_0)$ and since for such values of s the density $\bar{p}_s = p \wedge_I^B$ vanishes outside $(a, a+s) \subset (a, a+l)$, we may restrict the infimum in (69) to those densities in $\overline{\mathcal{M}}_3^\infty(R)$ that vanish outside $(a, a+l)$. \square

Let us set $\eta = \tau_\infty(p, \exp(R/\ell) - 1)$. Since $R/\ell \geq \log 2$, by applying Lemma 5 to the densities p_i , we may find for all $i \in \{1, \dots, \ell\}$ a density $\bar{p}_i \in \overline{\mathcal{M}}_3^\infty(R/\ell)$ such that $\|p_i - \bar{p}_i\| \leq 2\tau(p_i, \exp(R/\ell) - 1)$. In particular,

the density $\bar{p} = \sum_{i=1}^{\ell} w_i \bar{p}_i$ satisfies

$$(71) \quad \|p - \bar{p}\| \leq \sum_{i=1}^{\ell} w_i \|p_i - \bar{p}_i\| \leq \max_{i \in \{1, \dots, \ell\}} \|p_i - \bar{p}_i\| \leq 2\eta.$$

When $\ell > 2$ and $i \in \{2, \dots, \ell - 1\}$, it follows from the definition of $\overline{\mathcal{M}}_3^{\infty}(R/\ell)$ and Lemma 5 that we may choose \bar{p}_i in such a way that it vanishes outside an interval $I_i = (x_{i,0}, x_{i,1}) \subset (x_{i-1}, x_i)$ with

$$\log \left[1 + (x_{i,0} - x_{i,1}) \left(\sup_{x \in I_i} \bar{p}_i(x) - \inf_{x \in I_i} \bar{p}_i(x) \right) \right] < \frac{R}{\ell}$$

and $x_{i,0} = x_{i-1}$ when p_i is nonincreasing and $x_{i,1} = x_i$ when p_i is nondecreasing. The mapping \bar{p}_1 is a nondecreasing density on an interval of the form $I_1 = (x_{1,0}, x_{1,1})$ with $x_{1,0} < x_{1,1} = x_1$ and

$$\log \left[1 + (x_{1,1} - x_{1,0}) \left(\sup_{x \in I_1} \bar{p}_1(x) - \inf_{x \in I_1} \bar{p}_1(x) \right) \right] < \frac{R}{\ell}.$$

Similarly, \bar{p}_{ℓ} is a nonincreasing density on an interval of the form $I_{\ell} = (x_{\ell,0}, x_{\ell,1})$ with $x_{\ell,0} = x_{\ell-1} < x_{\ell,1}$ and

$$\log \left[1 + (x_{\ell,1} - x_{\ell,0}) \left(\sup_{x \in I_{\ell}} \bar{p}_{\ell}(x) - \inf_{x \in I_{\ell}} \bar{p}_{\ell}(x) \right) \right] < \frac{R}{\ell}.$$

The density $\bar{p} = \sum_{i=1}^{\ell} w_i \bar{p}_i$ can also be written as

$$w_1 \bar{p}_1 \mathbb{1}_{(x_{1,0}, x_1)} + \sum_{i=2}^{\ell-1} w_i [\bar{p}_i \mathbb{1}_{I_i} + 0 \mathbb{1}_{(x_{i-1}, x_i) \setminus I_i}] + w_{\ell} \bar{p}_{\ell} \mathbb{1}_{(x_{\ell-1}, x_{\ell,1})}$$

and may therefore be written in the form (17) when $k \geq 2\ell$. Moreover, by the Cauchy-Schwarz inequality

$$\mathbf{R}_{k,0}(\bar{p}) < \left[\sum_{i=1}^{\ell} \sqrt{w_i \left(\frac{R}{\ell} \right) + 0} \right]^2 = \frac{R}{\ell} \left[\sum_{i=1}^{\ell} \sqrt{w_i} \right]^2 \leq R$$

and consequently, $\bar{p} \in \overline{\mathcal{M}}_k^{\infty}(R)$. We deduce from (71) that

$$\inf_{q \in \overline{\mathcal{M}}_k^{\infty}(R)} \|p - q\| \leq \|p - \bar{p}\| \leq 2\eta$$

which is (28). □

8.4. Proofs of Section 5.

Proof of Proposition 4. The proof relies on the following lemma the proof of which is a direct consequence of convexity and is therefore omitted.

Lemma 6. *Let g be a convex and continuous function on a nontrivial interval J . The set $\{x \in J, g(x) < 0\}$ is an interval (possibly empty). The set $\{x \in J, g(x) > 0\}$ has one of the following forms: $\emptyset, J, J \cap (-\infty, c_0), J \cap (c_1, +\infty), [J \cap (-\infty, c_0)] \cup [J \cap (c_1, +\infty)]$ with $c_0, c_1 \in J$ and $c_0 < c_1$. In particular $\{x \in J, g(x) > 0\}$ is the union of at most two intervals. When g is continuous and concave on J , the same conclusion holds with $\{x \in J, g(x) > 0\}$ in place of $\{x \in J, g(x) < 0\}$ and vice-versa.*

Since p belongs to $\overline{\mathcal{M}}_k^1$ and q belongs to $\overline{\mathcal{O}}_{D,k}^1$, there exists $A = \{a_1, \dots, a_l\}$ with $l \in \{1, \dots, k-1\}$ such that p is convex-concave on each element of $\mathbf{I}(A)$ and there exists a subset $B \subset \mathbb{R}$ with cardinality not larger than $D+2$ such that q is left-continuous on \mathbb{R} and affine on each element of $\mathbf{I}(B)$. Since p and q are densities, p is necessarily convex on the two unbounded intervals of $\mathbf{I}(A)$ and q vanishes on the two unbounded intervals of $\mathbf{I}(B)$.

We define $J_1 = (-\infty, a_1]$, $J_{l+1} = (a_l, +\infty)$ and when $l \geq 2$, $J_i = (a_{i-1}, a_i]$ for all $i \in \{2, \dots, l\}$. Besides, we set $I_i = J_i$ for all $i \in \{1, \dots, l+1\}$. We shall repeatedly use Lemma 6 with $g = p - q$ throughout the proof. Given $\epsilon \in \{\pm 1\}$, we set

$$(72) \quad C_\epsilon = \{x \in \mathbb{R}, \epsilon g(x) > 0\} = \bigcup_{i=1}^{l+1} \{x \in J_i, \epsilon g(x) > 0\}.$$

Our aim is to show that C_ϵ is the union of at most $D + 2k - 1$ intervals. The second part of the proposition is a consequence of Lemma 1 of Baraud and Birgé [4].

If $m_1 = |B \cap I_1| = 0$ then $q = 0$ on J_1 (since it is left-continuous), $g = p - 0$ is continuous, monotone (nondecreasing) and convex on I_1 , $\{x \in J_1, g(x) < 0\} = \emptyset$ and $\{x \in J_1, g(x) > 0\}$ is an interval.

If $m_1 = |B \cap I_1| \geq 1$, we may partition J_1 into $s = m_1 + 1 \geq 2$ consecutive intervals K_1, \dots, K_s that we may choose to be of the form $(a, b]$, $a < b$, $a \in \mathbb{R} \cup \{-\infty\}$, $b \in \mathbb{R}$. Since p is continuous on I_1 and q is left-continuous, g is continuous on K_1, \dots, K_{s-1} and on $\overset{\circ}{K}_s$. On K_1 , $\{x \in K_1, g(x) < 0\} = \emptyset$ and the set $\Lambda_{1,] }^+ = \{x \in K_1, g(x) > 0\}$ is an interval which is either empty or contains the right endpoint of K_1 . When $s \geq 2$, we may apply Lemma 6 to g and the intervals K_i with $i \in \{2, \dots, s-1\}$. We obtain that $\Lambda_i^- = \{x \in K_i, g(x) < 0\}$ is an interval and $\{x \in K_i, g(x) > 0\}$ is of the form $\Lambda_{i,(}^+ \cup \Lambda_{i,] }^+$ where $\Lambda_{i,(}^+, \Lambda_{i,] }^+$ are two (possibly empty) intervals and when they are not, $\inf \Lambda_{i,(}^+ = \inf K_i$ and the right endpoint of K_i belongs to $\Lambda_{i,] }^+$. The set $\{x \in K_s, g(x) < 0\}$ can also be written as $\Lambda_s^- \cup \Lambda_{s,] }^-$ where $\Lambda_s^-, \Lambda_{s,] }^-$ are two possibly empty intervals and when $\Lambda_{s,] }^-$ is not empty, it reduces to $\{a_1\}$. The set $\{x \in K_s, g(x) > 0\}$ writes $\Lambda_{s,(}^+ \cup \Lambda_s^+$ where $\Lambda_{s,(}^+, \Lambda_s^+$ are two possibly empty intervals and when they are not $\inf \Lambda_{s,(}^+ = \inf K_s$ and $\sup \Lambda_s^+ = \sup K_s$. We conclude that

$$\{x \in J_1, g(x) < 0\} = \left[\bigcup_{i=2}^{s-1} \Lambda_i^- \right] \cup \left[\Lambda_s^- \cup \Lambda_{s,] }^- \right]$$

and

$$\{x \in J_1, g(x) > 0\} = \Lambda_{1,] }^+ \cup \left[\bigcup_{i=2}^{s-1} \left(\Lambda_{i,(}^+ \cup \Lambda_{i,] }^+ \right) \right] \cup \left[\Lambda_{s,(}^+ \cup \Lambda_s^+ \right]$$

are both the unions of at most $s = m_1 + 1$ intervals.

By arguing similarly, we obtain that on the interval J_{l+1} : the sets $\{x \in J_{l+1}, \epsilon g(x) > 0\}$ with $\epsilon \in \{\pm 1\}$ are the unions of at most $m_{l+1} + 1$ intervals where $m_{l+1} = |B \cap J_{l+1}|$.

When $l \geq 2$, let us now consider an interval of the form $J_i = (a_{i-1}, a_i]$ with $i \in \{2, \dots, l\}$ and set $m_i = |B \cap I_i|$. If $m_i = 0$, g is continuous and convex-concave on I_i and by arguing as for K_s , we obtain that $\{x \in J_i, \epsilon g(x) > 0\}$ is the union of at most 2 intervals for every $\epsilon \in \{\pm 1\}$. If $m_i \geq 1$, we may partition I_i with $m_i + 1$ intervals of the form $(a, b]$ with $a < b$, $a, b \in \mathbb{R}$. On each of these intervals, g is continuous and convex-concave and by applying Lemma 6 and arguing as previously, we obtain that $\{x \in J_i, \epsilon g(x) > 0\}$ is a union of at most $m_i + 2$ intervals.

Using (72) and the facts that $\sum_{i=1}^{l+1} m_i \leq |B| \leq D + 1$ and $l \leq k - 1$, we conclude that the sets C_ϵ are unions of at most

$$m_1 + 1 + m_{l+1} + 1 + \sum_{i=2}^l (m_i + 2) \leq |B| + 2l \leq D + 2k - 1$$

intervals. □

Proof of Theorem 6. The proof is based on Proposition 8 as well as two other approximation results, namely Proposition 10 and Proposition 11 below. We shall additionally use a elementary lemma to simplify some calculations. The proofs the two propositions and the lemma follow that of Theorem 6.

Proposition 10. *Let p be a monotone, continuous and convex-concave sub-density on a bounded interval $[a, b]$ of length $L > 0$ with a right derivative $p'_r(a)$ at a and a left derivative $p'_l(b)$ at b . For all $D \geq 1$, there exists a D -linear interpolation \bar{p} of p such that*

$$(73) \quad \int_a^b |p - \bar{p}| d\mu \leq \frac{4}{3} \left[\left(1 + L \sqrt{|p'_l(b) - p'_r(a)|} \right)^{1/D} - 1 \right]^2.$$

Proposition 11. *Let p be a strictly monotone, continuous, convex-concave sub-density with variation $V = |p(a) - p(b)|$ on a nontrivial bounded interval $[a, b]$. Assume furthermore that $p'_r(a)$ and $p'_l(b)$ are nonzero. Then, for all $D \geq 1$ there exists a D -linear interpolation \bar{p} of p on $[a, b]$ such that*

$$(74) \quad \int_a^b |p - \bar{p}| d\mu \leq \frac{4}{3} \left[\left(1 + V \sqrt{\left| \frac{1}{p'_r(a)} - \frac{1}{p'_l(b)} \right|} \right)^{1/D} - 1 \right]^2.$$

Lemma 7. *If F is a nondecreasing, differentiable, concave function on \mathbb{R}_+ , the mapping*

$$\phi : u \mapsto [F(\sqrt{u}) - F(0)]^2$$

is concave on \mathbb{R}_+ . In particular for all $D \geq 1$,

$$F_D : u \mapsto [(1 + \sqrt{u})^{1/D} - 1]^2$$

is concave on \mathbb{R}_+ .

Let's turn to the proof of Theorem 6. We first introduce some mappings of interest we will use in the proofs of Theorem 6 and Proposition 10. Let \mathcal{V} be the linear space of continuous functions f on $[a, b]$ that admit a right derivative $f'_r(a)$ at a and a left derivative $f'_l(b)$ at b . Given $m \in \mathbb{R}$, we define \mathcal{T}_1 and \mathcal{T}_2 as the mappings defined on \mathcal{V} by

$$(75) \quad \mathcal{T}_1 : f \mapsto [x \mapsto f(a + b - x)], \quad \mathcal{T}_2 : f \mapsto [x \mapsto m - f(x)].$$

Although \mathcal{T}_2 depends on the choice of m , we drop this dependency in the notation for the sake of convenience. The mappings \mathcal{T}_j are one-to-one from \mathcal{V} onto itself, isometric with respect to the \mathbb{L}_1 -norm on \mathcal{V} and they satisfy for all $f \in \mathcal{V}$

$$(76) \quad |f(a) - f(b)| = |\mathcal{T}_j(f)(a) - \mathcal{T}_j(f)(b)|$$

$$(77) \quad |f'_r(a) - f'_l(b)| = |(\mathcal{T}_j(f))'_r(a) - (\mathcal{T}_j(f))'_l(b)|$$

and $\mathcal{T}_j^{-1} = \mathcal{T}_j$ for all $j \in \{1, 2\}$.

Let us now turn to the proof of Theorem 6 and assume first that p is nondecreasing, continuous and convex on $[a, b]$ so that $p'_r(a) \geq 0$. If p is constant, the result is clear by taking $\bar{p} = p$. We may therefore assume that $p(a) < p(b)$ and choose a point $c \in (a, b)$ such that $p(c) > p(a)$. In particular, $w_1 = \int_a^c p d\mu > 0$ and

$$0 < \frac{p(c) - p(a)}{c - a} \leq p'_l(c) \leq \frac{p(b) - p(c)}{b - c} < +\infty.$$

The restriction p_1 of p on the interval $[a, c]$ is nondecreasing, continuous and convex and so is the density p_1/w_1 . We may therefore apply Proposition 10 to p_1/w_1 and find a D -linear interpolation \bar{p}_1 of p on $[a, c]$

that satisfies

$$(78) \quad \int_a^c |p - \bar{p}_1| d\mu \leq \frac{4w_1}{3} \left[\left(1 + (c-a) \sqrt{\frac{p'_l(c) - p'_r(a)}{w_1}} \right)^{1/D} - 1 \right]^2.$$

The restriction p_2 of p to $[c, b]$ is increasing, continuous and convex with nonzero right and left derivatives at c and b respectively. We may therefore apply Proposition 11 to the density p_2/w_2 with $w_2 = \int_c^b p d\mu = 1 - w_1 > 0$ and find a D -linear interpolation \bar{p}_2 of p on $[c, b]$ that satisfies

$$(79) \quad \int_c^b |p - \bar{p}_2| d\mu \leq \frac{4w_2}{3} \left[\left(1 + \frac{p(b) - p(c)}{\sqrt{w_2}} \sqrt{\frac{1}{p'_r(c)} - \frac{1}{p'_l(b)}} \right)^{1/D} - 1 \right]^2.$$

We may choose $c \in (a, b)$ such that

$$p'_l(c) \leq \Delta \quad \text{and} \quad p'_r(c) \geq \Delta \quad \text{with} \quad \Delta = \frac{p(b) - p(a)}{b - a}.$$

Then

$$(80) \quad \begin{aligned} A &= (c-a)^2 (p'_l(c) - p'_r(a)) + (p(b) - p(c))^2 \left(\frac{1}{p'_r(c)} - \frac{1}{p'_l(b)} \right) \\ &\leq (b-a)^2 (\Delta - p'_r(a)) + (p(b) - p(a))^2 \left(\frac{1}{\Delta} - \frac{1}{p'_l(b)} \right) \\ &= 2(b-a)(p(b) - p(a)) \left[1 - \frac{1}{2} \left(\frac{p'_r(a)}{\Delta} + \frac{\Delta}{p'_l(b)} \right) \right] = 2LVT. \end{aligned}$$

The function $\bar{p} = \bar{p}_1 \mathbb{1}_{[a,c]} + \bar{p}_2 \mathbb{1}_{[c,b]}$ is a $2D$ -linear interpolation of p on $[a, b]$. Since by Lemma 7 the function F_D is concave and increasing, we deduce from (80), (78) and (79) that

$$\begin{aligned} \frac{3}{4} \|p - \bar{p}\| &\leq w_1 \int_a^c |p - \bar{p}_1| d\mu + w_2 \int_c^b |p - \bar{p}_2| d\mu \\ &\leq w_1 F_D \left(\frac{(c-a)^2 (p'_l(c) - p_r(a))}{w_1} \right) \\ &\quad + w_2 F_D \left(\frac{(p(b) - p(c))^2}{w_2} \left(\frac{1}{p'_r(c)} - \frac{1}{p'_l(b)} \right) \right) \\ &\leq F_D(A) \leq F_D(2LVT), \end{aligned}$$

which is (34).

The density $q = \bar{p} / \int_a^b \bar{p} d\mu$ is continuous convex-concave on $[a, b]$ and piecewise linear on a partition of $[a, b]$ into $2D$ intervals and it follows from Lemma 3 that

$$(81) \quad \int_a^b |p - q| d\mu \leq 2 \left\{ 1 \wedge \left[\frac{4}{3} \left(\left(1 + \sqrt{2LVT} \right)^{1/D} - 1 \right)^2 \right] \right\}.$$

The mapping $z \mapsto (4/3)(e^z - 1)^2$ is not larger than 1 if and only if $z \leq z_0 = \log(1 + \sqrt{3/4})$ and for all $z \in [0, z_0]$, $e^z - 1 \leq (e^{z_0} - 1)z/z_0$. Consequently, for all $z \geq 0$

$$2 \left\{ 1 \wedge \left[\frac{4}{3} (e^z - 1)^2 \right] \right\} = \frac{8}{3} (e^{z_0 \wedge z} - 1)^2 \leq \frac{8}{3} \left(\frac{e^{z_0} - 1}{z_0} z \right)^2 \leq 5.14z^2.$$

Applying this inequality with $z = D^{-1} \log \left(1 + \sqrt{2LV\Gamma} \right)$, we deduce (35) from (81).

Theorem 6 is therefore proven for a nondecreasing, continuous convex density p . In order to prove the result in the other cases, we use the transformations \mathcal{T}_1 and \mathcal{T}_2 introduced above and defined by (75). These transformations are isometric with respect to the \mathbb{L}_1 -norm and they preserve the variation of a monotone function. Note that they also preserve the linear index, i.e. for all continuous convex-concave function f on $[a, b]$ and $j \in \{1, 2\}$, $\Gamma(f) = \Gamma(\mathcal{T}_j(f))$. Applying \mathcal{T}_1 to nonincreasing continuous convex densities we establish (35) and we extend it to all monotone concave continuous densities by applying \mathcal{T}_2 . \square

Proof of Proposition 10. Let us first consider a function g that is monotone, continuous and convex on $[a, b]$ and that satisfies $g(a) = g'_r(a) = 0$ and $0 < \int_a^b g d\mu \leq 1$. Then g is nonnegative and nondecreasing on $[a, b]$. Since $\int_a^b g d\mu > 0$, g is not identically equal to 0 on $[a, b]$ and consequently,

$$g'_l(b) \geq \frac{g(b) - g(a)}{b - a} > 0.$$

Let $q > 1$ and

$$x_0 = a \quad \text{and} \quad x_i = x_{i-1} + L \frac{q^{-i}}{\sum_{j=1}^D q^{-j}} \quad \text{for all } i \in \{1, \dots, D\},$$

so that $x_D = b$ and $\Delta_{i+1} = x_{i+1} - x_i = q^{-1}(x_i - x_{i-1}) = q^{-1}\Delta_i$ for all $i \in \{1, \dots, D-1\}$ and

$$\Delta_D = L \frac{q^{-D}(1 - q^{-1})}{q^{-1} - q^{-(D+1)}} = L \frac{q - 1}{q^D - 1}.$$

Let g' be any nondecreasing function on $[a, b]$ satisfying $0 = g'_r(a) = g'(a)$, $g'_l(b) = g'(b)$ and $g'_l(x) \leq g'(x) \leq g'_r(x)$ for all $x \in (a, b)$. Since g is convex, we may write

$$(82) \quad g(x) \geq g(x_i) + g'(x_i)(x - x_i) \quad \text{for all } i \in \{1, \dots, D\} \text{ and } x \in [a, b].$$

In particular, for all $i \in \{1, \dots, D-1\}$

$$\int_{x_i}^{x_{i+1}} (g(x_i) + g'(x_i)(x - x_i)) d\mu(x) = \left[g(x_i) + \frac{g'(x_i)\Delta_{i+1}}{2} \right] \Delta_{i+1} \leq \int_{x_i}^{x_{i+1}} g d\mu,$$

hence,

$$(83) \quad g'(x_i)\Delta_{i+1}^2 \leq 2 \int_{x_i}^{x_{i+1}} (g - g(x_i)) d\mu.$$

Moreover, applying (82) with $x_i = x_D = b$ and using the facts that g is nonnegative and nonincreasing, we get

$$1 \geq \int_a^b g d\mu \geq \int_a^b (g(b) + g'(b)(x - b))_+ d\mu(x) = \int_{b-g(b)/g'(b)}^b (g(b) + g'(b)(x - b)) d\mu(x) = \frac{g^2(b)}{2g'(b)},$$

consequently

$$(84) \quad g(b) \leq \sqrt{2g'(b)}.$$

Let g_i be the restriction of g on the interval $[x_{i-1}, x_i]$ and \bar{g} the function on $[a, b]$ that coincides on $[x_{i-1}, x_i]$ with ℓ_{g_i} defined by (48) with $f = g_i$ for all $i \in \{1, \dots, D\}$. The function \bar{g} is a D -linear interpolation of g on

$[a, b]$ that satisfies

$$\begin{aligned} \int_a^b |g - \bar{g}| d\mu &= \sum_{i=1}^D \int_{x_{i-1}}^{x_i} (\bar{g} - g) d\mu \leq \frac{1}{8} \sum_{i=1}^D \Delta_i^2 (g'_i(x_i) - g'_r(x_{i-1})) \leq \frac{1}{8} \sum_{i=1}^D \Delta_i^2 (g'(x_i) - g'(x_{i-1})) \\ &= \frac{1}{8} \left[-\Delta_1^2 g'(a) + \Delta_D^2 g'(b) + \sum_{i=1}^{D-1} g'(x_i) (\Delta_i^2 - \Delta_{i+1}^2) \right]. \end{aligned}$$

Since $g'(a) = 0$ and $\Delta_i = q\Delta_{i+1}$ for all $i \in \{1, \dots, D-1\}$ we deduce that

$$(85) \quad \int_a^b |g - \bar{g}| d\mu \leq \frac{1}{8} \left[\Delta_D^2 g'(b) + (q^2 - 1) \sum_{i=1}^{D-1} g'(x_i) \Delta_{i+1}^2 \right].$$

Using (83), (84) and the fact that g is nondecreasing,

$$\begin{aligned} \frac{1}{2} \sum_{i=1}^{D-1} g'(x_i) \Delta_{i+1}^2 &\leq \sum_{i=1}^{D-1} \int_{x_i}^{x_{i+1}} (g - g(x_i)) d\mu \\ &= \int_{x_1}^{x_D} g d\mu - \sum_{i=1}^{D-1} \Delta_{i+1} g(x_i) = \int_{x_1}^{x_D} g d\mu - q^{-1} \sum_{i=1}^{D-1} \Delta_i g(x_i) \\ &\leq \int_{x_1}^{x_D} g d\mu - q^{-1} \sum_{i=1}^{D-1} \int_{x_{i-1}}^{x_i} g d\mu = \int_{x_1}^{x_D} g d\mu - q^{-1} \int_{x_0}^{x_{D-1}} g d\mu \\ &= (1 - q^{-1}) \int_{x_1}^{x_D} g d\mu + q^{-1} \int_{x_{D-1}}^{x_D} g d\mu - q^{-1} \int_{x_0}^{x_1} g d\mu \\ &\leq (1 - q^{-1}) \times 1 + q^{-1} \Delta_D g(b) \leq 1 - q^{-1} + q^{-1} \sqrt{2\Delta_D^2 g'(b)}. \end{aligned}$$

It follows from (85) that

$$\begin{aligned} \int_a^b |g - \bar{g}| d\mu &\leq \frac{1}{8} \left[\Delta_D^2 g'(b) + 2(q^2 - 1) \left(1 - q^{-1} + q^{-1} \sqrt{2\Delta_D^2 g'(b)} \right) \right] \\ &\leq \frac{(q-1)^2}{8} \left[\frac{L^2 g'(b)}{(q^D - 1)^2} + 2 \left(1 + \frac{1}{q} \right) \left(1 + \frac{L\sqrt{2g'(b)}}{q^D - 1} \right) \right]. \end{aligned}$$

Finally, choosing q such that

$$q^D - 1 = L\sqrt{g'(b)} \quad \text{i.e.} \quad q = \left(1 + L\sqrt{g'(b)} \right)^{1/D} > 1,$$

we get that

$$(86) \quad \begin{aligned} \int_a^b |g - \bar{g}| d\mu &\leq \frac{1 + 4(1 + \sqrt{2})}{8} \left[\left(1 + L\sqrt{g'(b)} \right)^{1/D} - 1 \right]^2 \\ &\leq \frac{4}{3} \left[\left(1 + L\sqrt{g'(b)} \right)^{1/D} - 1 \right]^2. \end{aligned}$$

Let us now prove Proposition 10 in the case where p is convex and nondecreasing. Then $p'_r(a) \geq 0$ and we may set $\ell : x \mapsto p(a) + p'_r(a)(x - a)$ and $g : x \mapsto p(x) - \ell(x)$. The function g is nonnegative, nondecreasing, continuous and convex on $[a, b]$ and it satisfies $g(a) = g'_r(a) = 0$. If $\int_a^b g d\mu = 0$, then $p = \ell$ and we may choose $\bar{p} = p$. Otherwise $0 < \int_a^b g d\mu \leq 1$, since ℓ is nonnegative on $[a, b]$, and we may apply our previous

result to g . This leads to a D -linear interpolation \bar{g} of g from which we may define $\bar{p} = \bar{g} - \ell(x)$ which is a D -linear interpolation of p on $[a, b]$. Inequality (73) follows from (86) and the facts that $|p - \bar{p}| = |g - \bar{g}|$ and $g'_l(b) = p'_l(b) - p'_r(a)$.

In order to prove Proposition 10 in the other cases, we use transformations defined by (75). We can note that if ℓ is a D -linear interpolation of f , $\mathcal{T}_j(\ell)$ is D -linear interpolation of $\mathcal{T}_j(f)$ (based on the same subdivision).

If p is convex and nonincreasing, we apply the transformation \mathcal{T}_1 . Then $g = \mathcal{T}_1(p)$ is a convex, nondecreasing sub-density on $[a, b]$ and our previous result applies. We may find a D -linear interpolation \bar{g} on $[a, b]$ such that

$$(87) \quad \int_a^b |g - \bar{g}| d\mu \leq \frac{4}{3} \left[\left(1 + L \sqrt{|g'_l(b) - g'_r(a)|} \right)^{1/D} - 1 \right]^2,$$

and the function $\bar{p} = \mathcal{T}_1(\bar{g})$ is a D -linear interpolation of $p = \mathcal{T}_1(g)$ that satisfies (73). The result is therefore proven for all convex continuous monotone sub-density on $[a, b]$.

If p is a concave continuous sub-density, we apply the transformation \mathcal{T}_2 with $m = 2S/(b - a)$ and $S = \int_a^b p d\mu \leq 1$. Since p is concave and monotone, $(p(a) \vee p(b))(b - a)/2 \leq S$, hence $p(x) \leq m$ for all $x \in [a, b]$, and $g = \mathcal{T}_2(p) = m - p$ is a monotone convex sub-density since

$$\int_a^b (m - p) d\mu = \frac{2S}{b - a} \times (b - a) - S = S \leq 1.$$

Applying the previous result, we may find a D -interpolation \bar{g} of g that satisfies (87) and $\bar{p} = \mathcal{T}_2(\bar{g})$ satisfies (73). This completes the proof of Proposition 10. \square

Proof of Proposition 11. We start with the following lemma.

Lemma 8. *If f and g are two continuous increasing or decreasing functions from on interval I onto an interval J ,*

$$(88) \quad \int_I |f - g| d\mu = \int_J |f^{-1} - g^{-1}| d\mu.$$

Moreover, if $\inf I = 0$ and f is nonnegative, continuous and decreasing

$$\int_J f^{-1} d\mu \leq \int_I f d\mu.$$

Proof. Let us first assume that f, g are both increasing. For all $(t, y) \in I \times J$

$$\mathbb{1}_{g(t) \leq y \leq f(t)} + \mathbb{1}_{f(t) \leq y \leq g(t)} = \mathbb{1}_{f^{-1}(y) \leq t \leq g^{-1}(y)} + \mathbb{1}_{g^{-1}(y) \leq t \leq f^{-1}(y)}.$$

By integrating this equality on $I \times J$ and using Fubini's theorem, we obtain that

$$\begin{aligned} \int_I |f(t) - g(t)| d\mu(t) &= \int_I \left[\int_J [\mathbb{1}_{g(t) \leq y \leq f(t)} + \mathbb{1}_{f(t) \leq y \leq g(t)}] d\mu(y) \right] d\mu(t) \\ &= \int_J \left[\int_I [\mathbb{1}_{f^{-1}(y) \leq t \leq g^{-1}(y)} + \mathbb{1}_{g^{-1}(y) \leq t \leq f^{-1}(y)}] d\mu(t) \right] d\mu(y) \\ &= \int_J |f^{-1}(y) - g^{-1}(y)| d\mu(y). \end{aligned}$$

When f and g are both decreasing, we may apply the above equality to $\bar{f} = f(\cdot)$ and $\bar{g} = g(\cdot)$ from $K = \{-x, x \in I\}$ to J then $\bar{f}^{-1} = -f^{-1}$, $\bar{g}^{-1} = -g^{-1}$ and the result follows from the facts

$$\int_I |f - g| d\mu = \int_K |\bar{f} - \bar{g}| d\mu \text{ and } \int_J |\bar{f}^{-1} - \bar{g}^{-1}| d\mu = \int_J |f^{-1} - g^{-1}| d\mu.$$

Let us now turn to the proof of the second part of the lemma. Since $\inf I = 0$ and f is decreasing and continuous from I onto J , for all $t \in J$

$$\mu(\{x \in I, f(x) \geq t\}) = \mu(\{x \in I, x \leq f^{-1}(t)\}) = f^{-1}(t).$$

Besides, f being nonnegative, the interval $J \subset \mathbb{R}_+$ and it follows from Fubini's theorem that

$$\begin{aligned} \int_J f^{-1}(t) d\mu(t) &= \int_J \mu(\{x \in I, f(x) \geq t\}) d\mu(t) \\ &= \int_J \left[\int_I \mathbf{1}_{t \leq f(x)} d\mu(x) \right] d\mu(t) = \int_I \left[\int_J \mathbf{1}_{t \leq f(x)} d\mu(t) \right] d\mu(x) \\ &\leq \int_I \left[\int_0^{+\infty} \mathbf{1}_{t \leq f(x)} d\mu(t) \right] d\mu(x) = \int_I f(x) d\mu(x). \end{aligned}$$

□

Let us now turn to the proof of (74). We first claim that it is sufficient to establish (74) when p is decreasing. If p were increasing, we could apply the result to $q = \mathcal{T}_1(p)$ defined by (75), which is then decreasing, and find a D -linear interpolation of q on $[a, b]$ that satisfies

$$\int_a^b |q - \bar{q}| d\mu \leq \frac{4}{3} \left[\left(1 + V \sqrt{\left| \frac{1}{q'_r(a)} - \frac{1}{q'_l(b)} \right|} \right)^{1/D} - 1 \right]^2.$$

We then conclude by using the facts that \mathcal{T}_1 is an \mathbb{L}_1 -isometry, the function $\bar{p} = \mathcal{T}_1(\bar{q})$ which is a D -linear interpolation of p , and $q'_r(a) = -p'_l(b)$ and $q'_l(b) = -p'_r(a)$.

We may therefore assume that p is decreasing and change p into $p(\cdot - a)$, which amounts to translating the sub-density p , we may also assume without loss of generality that $a = 0$. By Lemma 8, $s = p^{-1}$ is then a sub-density on $[p(b), p(0)]$ which is furthermore decreasing, continuous and convex-concave. Since the right and left derivatives of p at 0 and b respectively are not zero, s admits a right derivative at $p(b)$ and a left derivative at $p(0)$ given by $1/p'_l(b)$ and $1/p'_r(0)$ (with our convention $1/(+\infty) = 0$). We may therefore apply Proposition 10 and find a D -linear interpolation \bar{s} of s on $[p(b), p(0)]$ that satisfies

$$\int_{p(b)}^{p(0)} |s - \bar{s}| d\mu \leq \frac{4}{3} \left[\left(1 + V \sqrt{\left| \frac{1}{p'_r(0)} - \frac{1}{p'_l(b)} \right|} \right)^{1/D} - 1 \right]^2.$$

Since s is continuous and decreasing from $[p(b), p(0)]$ onto $[0, b]$, so is \bar{s} , and we may set $\bar{p} = \bar{s}^{-1} : [0, b] \rightarrow [p(b), p(0)]$. The function \bar{p} is a D -linear interpolation of p on $[0, b]$ and we conclude by using the equality $\int_{p(b)}^{p(0)} |s - \bar{s}| d\mu = \int_0^b |p - \bar{p}| d\mu$ which is a consequence of Lemma 8. □

Proof of Lemma 7. For all $u > 0$,

$$\phi'(u) = \frac{F(\sqrt{u}) - F(0)}{\sqrt{u}} F'(\sqrt{u})$$

is the product of $u \mapsto [F(\sqrt{u}) - F(0)]/\sqrt{u}$ and $u \mapsto F'(\sqrt{u})$ which are both nonnegative and nonincreasing since F is nondecreasing and concave. The function ϕ' is therefore nonincreasing and ϕ concave. □

8.5. Proofs of Section 6.

Proof of Proposition 6. Let $p \in \overline{\mathcal{M}}(\mathcal{L})$, $J = \{p > 0\}$ and $\bar{p} \in \overline{\mathcal{O}}_D(\mathcal{L})$. There exists a partition \mathcal{I} of \mathbb{R} into at most $D + 1$ intervals such that the restriction of $\mathcal{L}\bar{p}$ to an element $I \in \mathcal{I}$ is either affine or equal to $-\infty$. We denote by ϕ and $\bar{\phi}$ the functions $\mathcal{L}p$ and $\mathcal{L}\bar{p}$ respectively which map \mathbb{R} into $\mathbb{R} \cup \{-\infty\}$. Since \mathcal{L} is increasing

$$\{p > \bar{p}\} = \{\phi > \bar{\phi}\} \quad \text{and} \quad \{p < \bar{p}\} = \{\phi < \bar{\phi}\}.$$

To prove that \bar{p} is extremal in $\overline{\mathcal{M}}(\mathcal{L})$ with degree not larger than $2(D + 2)$, it suffices to prove that $\{\phi > \bar{\phi}\}$ and $\{\phi < \bar{\phi}\}$ are both unions of at most $D + 2$ intervals and then to use Lemma 1 of Baraud and Birgé [4]. This is what we shall do now.

Let us start with the set $\{\phi > \bar{\phi}\}$. Either $\bar{\phi}$ is finite on $I \in \mathcal{I}$ and $I \cap J \cap \{\phi > \bar{\phi}\}$ is an interval since $\phi - \bar{\phi}$ is a concave function on the interval $I \cap J$ when it is not empty. Otherwise, $\bar{\phi}$ takes the value $-\infty$ on I and $I \cap J \cap \{\phi > \bar{\phi}\} = I \cap J$ remains an interval. The set $I \cap J^c \cap \{\phi > \bar{\phi}\}$ is empty since ϕ takes the value $-\infty$ on J^c . Hence

$$I \cap \{\phi > \bar{\phi}\} = (I \cap J \cap \{\phi > \bar{\phi}\}) \cup (I \cap J^c \cap \{\phi > \bar{\phi}\})$$

is an interval and

$$\{p > \bar{p}\} = \{\phi > \bar{\phi}\} = \bigcup_{I \in \mathcal{I}} (I \cap \{\phi > \bar{\phi}\})$$

the union of at most $D + 1$ intervals.

Let us now turn to the set $\{\phi < \bar{\phi}\}$ and define $K = (a, b)$ as the open interval $\{\phi > -\infty\} \cap \{\bar{\phi} > -\infty\}$ with $a \in \mathbb{R} \cup \{-\infty\}$, $b \in \mathbb{R} \cup \{+\infty\}$, $a \leq b$. If $K = \emptyset$, i.e. $a = b$, $\{\phi < \bar{\phi}\} = \{\bar{\phi} > -\infty\} = \{\bar{p} > 0\}$ is an interval. Otherwise K is not empty, $\psi = \phi - \bar{\phi}$ is a continuous function on K which is either concave on the whole interval K or only piecewise concave on each element of $I \cap K$ with $I \in \mathbf{I}(A \cap K)$ when $A \cap K \neq \emptyset$. In any case,

$$\{x \in K, \phi(x) < \bar{\phi}(x)\} = \{x \in K, \phi(x) - \bar{\phi}(x) < 0\}$$

is a union of at most $|A \cap K|$ intervals plus, possibly, an additional one with endpoint a and another one with endpoint b . If $b < +\infty$, either $\bar{\phi}(b) = -\infty$ or $\phi(b) = -\infty$ and the set $\{x \geq b, \phi(x) < \bar{\phi}(x)\}$ is either empty or is an interval that contains b . We may argue similarly to establish that $\{x \leq a, \phi(x) < \bar{\phi}(x)\}$ is either empty or an interval that contains a . This means that the set

$$\{\phi < \bar{\phi}\} = (K \cap \{\phi < \bar{\phi}\}) \cup (K^c \cap \{\phi < \bar{\phi}\})$$

is a union of at most $|A \cap K| + 2 \leq D + 2$ intervals. □

Proof of Proposition 7. Since \mathcal{L} is increasing and convex on $(0, +\infty)$, it grows to $+\infty$ at $+\infty$ and it can be extended continuously at 0. Using the same notation \mathcal{L} for this extension, we obtain that $\mathcal{L} : [0, +\infty) \rightarrow [u, +\infty)$ for some $u \in \mathbb{R}$. The inverse $\mathcal{L}^{-1} : [u, +\infty) \rightarrow [0, +\infty)$ is continuously increasing and concave. Since $\mathcal{L}p$ is by definition concave on $\{p > 0\} = (a, b)$, the density $p = \mathcal{L}^{-1}(\mathcal{L}p)$ is also concave on this interval and (a, b) is necessarily bounded. Furthermore, p can be continuously extended on the compact interval $[a, b]$. This extension \tilde{p} is also concave on $[a, b]$ and thus admits a maximum at some point $c \in [a, b]$. The function \tilde{p} is nondecreasing on $[a, c]$ and nonincreasing on $[c, b]$. Applying (iii) in Proposition 9 we know that there exist a D -linear interpolation \bar{p}_1 of \tilde{p} on $[a, c]$ and a D -linear interpolation \bar{p}_2 of \tilde{p} on $[c, b]$ such that

$$\int_a^c |\tilde{p} - \bar{p}_1| d\mu \leq \frac{(c-a)(\tilde{p}(c) - \tilde{p}(a))}{2D^2} \quad \text{and} \quad \int_c^b |\tilde{p} - \bar{p}_2| d\mu \leq \frac{(b-c)(\tilde{p}(c) - \tilde{p}(b))}{2D^2}.$$

Since $\bar{p}_1(c) = \bar{p}_2(c) = \tilde{p}(c)$, the function $\bar{p} = \bar{p}_1 \mathbb{1}_{[a,c]} + \bar{p}_2 \mathbb{1}_{(c,b]}$ is a $(2D)$ -linear interpolation of \tilde{p} on $[a, b]$ which satisfies

$$\int_a^b |p - \bar{p}| d\mu = \int_a^b |\tilde{p} - \bar{p}| d\mu \leq \frac{1}{D^2} \left[\frac{(c-a)(\tilde{p}(c) - \tilde{p}(a))}{2} + \frac{(b-c)(\tilde{p}(c) - \tilde{p}(b))}{2} \right].$$

The quantities $[(c-a)(\tilde{p}(c) - \tilde{p}(a))]/2$ and $[(b-c)(\tilde{p}(c) - \tilde{p}(b))]/2$ are the areas of two disjoint triangles that lie under the graph of the density \tilde{p} on $[a, b]$ and their sum is therefore not larger than 1. This leads to the second inequality in (39). Note that since p is positive on (a, b) , so is its $(2D)$ -linear interpolation \bar{p} : $\bar{p}(x) > 0$ for all $x \in (a, b)$.

Since \mathcal{L} extends continuously at 0, the function $\mathcal{L}\tilde{p}$ is well-defined on $[a, b]$ and corresponds to the continuous extension of $\mathcal{L}p$ on this interval. In particular $\mathcal{L}\tilde{p}$ is concave on the compact interval $[a, b]$. Let $a = x_0 < \dots < x_{2D} = b$ be the subdivision associated with the $2D$ -linear interpolation \bar{p} of \tilde{p} on $[a, b]$. The function \bar{q} defined by

$$\bar{q}(x) = \sum_{i=1}^{2D} \left(\frac{x - x_{i-1}}{x_i - x_{i-1}} \mathcal{L}\tilde{p}(x_i) + \frac{x_i - x}{x_i - x_{i-1}} \mathcal{L}\tilde{p}(x_{i-1}) \right) \mathbb{1}_{[x_{i-1}, x_i)}(x) \quad \text{for all } x \in [a, b)$$

and $\bar{q}(b) = \mathcal{L}\tilde{p}(b)$, is a $2D$ -linear interpolation of $\mathcal{L}\tilde{p}$ based on the same subdivision as \bar{p} . In particular, since $\mathcal{L}\tilde{p}$ is concave, \bar{q} is also concave and it satisfies $\bar{q} \leq \mathcal{L}\tilde{p}$ on $[a, b]$. Since \mathcal{L}^{-1} is increasing and concave, the function $\check{p} = \mathcal{L}^{-1} \circ \bar{q}$ satisfies $0 \leq \check{p} \leq \mathcal{L}^{-1}(\mathcal{L}\tilde{p}) = \tilde{p}$ on $[a, b]$ and for all $x \in [x_{i-1}, x_i)$ with $i \in \{1, \dots, 2D\}$

$$\begin{aligned} \check{p}(x) &= \mathcal{L}^{-1}(\bar{q}(x)) = \mathcal{L}^{-1} \left(\frac{x - x_{i-1}}{x_i - x_{i-1}} \mathcal{L}\tilde{p}(x_i) + \frac{x_i - x}{x_i - x_{i-1}} \mathcal{L}\tilde{p}(x_{i-1}) \right) \\ &\geq \frac{x - x_{i-1}}{x_i - x_{i-1}} \mathcal{L}^{-1}(\mathcal{L}\tilde{p}(x_i)) + \frac{x_i - x}{x_i - x_{i-1}} \mathcal{L}^{-1}(\mathcal{L}\tilde{p}(x_{i-1})) \\ &= \frac{x - x_{i-1}}{x_i - x_{i-1}} \tilde{p}(x_i) + \frac{x_i - x}{x_i - x_{i-1}} \tilde{p}(x_{i-1}) = \bar{p}(x) \end{aligned}$$

and $\check{p}(b) = \mathcal{L}^{-1}(\mathcal{L}\tilde{p}(b)) = \tilde{p}(b) = \bar{p}(b)$. We deduce that $0 < \bar{p} \leq \check{p} \leq \tilde{p}$ on (a, b) and by using the second inequality in (39) we obtain that

$$(89) \quad \int_a^b |p - \check{p}| d\mu = \int_a^b (\tilde{p} - \check{p}) d\mu \leq \int_a^b (\tilde{p} - \bar{p}) d\mu = \int_a^b |p - \bar{p}| d\mu \leq \frac{1}{D^2}.$$

In a nutshell, the function \check{p} possesses the following properties: for all $x \in (a, b)$, $0 < \check{p}(x) \leq p(x)$, hence $\kappa = \int_a^b \check{p} d\mu \in (0, 1]$, and $\mathcal{L} \circ \check{p} = \bar{q}$ is concave and piecewise affine on each element of a partition of $[a, b]$ into $2D$ intervals. The function $\check{p} \mathbb{1}_{(a,b)}$ satisfies thus $\{\check{p} \mathbb{1}_{(a,b)} > 0\} = (a, b)$ as well as all the requirements for belonging to $\overline{\mathcal{O}}_{2D+1}(\mathcal{L})$, except from the fact that $\check{p} \mathbb{1}_{(a,b)}$ might not be a density. Let us define

$$s(x) = (\check{p} \mathbb{1}_{(a,b)})(\kappa x + (1 - \kappa)c) \quad \text{for } x \in \mathbb{R}.$$

Then s is a density, $\{s > 0\}$ is the open interval $(c + \kappa^{-1}(a - c), c + \kappa^{-1}(b - c))$ and for all $x \in \{s > 0\}$ the function

$$\mathcal{L}s(x) = (\mathcal{L} \circ \check{p})(\kappa x + (1 - \kappa)c) = \bar{q}(\kappa x + (1 - \kappa)c)$$

is concave and piecewise affine on each element of a partition of $\{s > 0\}$ into the $2D$ -intervals. By definition, $\mathcal{L}s$ takes the value $-\infty$ on $\{s = 0\}$ which is a union of two intervals. Consequently, s belongs to $\overline{\mathcal{O}}_{2D+1}(\mathcal{L})$ and we conclude by using Lemma 4. \square

Proof of Theorem 9. Since the function \mathcal{L} is concave and increasing on $(0, +\infty)$, it is continuous on this interval and its inverse \mathcal{L}^{-1} is convex, increasing and continuous on the range (u, v) of \mathcal{L} , with $-\infty \leq u < v \leq +\infty$. Throughout the proof, we shall repeatedly use the following equalities: for all $s, t > 0$ with $s < t$

$$(90) \quad \int_s^t z \mathcal{L}'_r(z) dz = \int_s^t z \mathcal{L}'_l(z) dz = \int_{\mathcal{L}(s)}^{\mathcal{L}(t)} \mathcal{L}^{-1}(u) du.$$

In particular, by letting s decrease to 0 and t increase to infinity, we respectively obtain that for all $t > 0$

$$(91) \quad \int_0^t z \mathcal{L}'_r(z) dz = \int_0^t z \mathcal{L}'_l(z) dz = \int_u^{\mathcal{L}(t)} \mathcal{L}^{-1}(u) du$$

and for all $s > 0$

$$(92) \quad \int_s^{+\infty} z \mathcal{L}'_r(z) dz = \int_s^{+\infty} z \mathcal{L}'_l(z) dz = \int_{\mathcal{L}(s)}^v \mathcal{L}^{-1}(u) du.$$

We obtain (90) by using Fubini's theorem and the fact that a concave function is absolutely continuous on any compact interval $[s, t] \subset (0, +\infty)$. More precisely,

$$\begin{aligned} \int_{\mathcal{L}(s)}^{\mathcal{L}(t)} \mathcal{L}^{-1}(u) du &= \int_{\mathcal{L}(s)}^{\mathcal{L}(t)} \left[\int_0^{+\infty} \mathbb{1}_{x \leq \mathcal{L}^{-1}(u)} dx \right] du = \int_{\mathcal{L}(s)}^{\mathcal{L}(t)} \left[\int_0^{+\infty} \mathbb{1}_{\mathcal{L}(x) \leq u} dx \right] du \\ &= \int_0^t \left[\int_{\mathcal{L}(s)}^{\mathcal{L}(t)} \mathbb{1}_{\mathcal{L}(x) \leq u} du \right] dx = \int_0^t (\mathcal{L}(t) - \mathcal{L}(u) \vee \mathcal{L}(s)) du \\ &= \int_0^t (\mathcal{L}(t) - \mathcal{L}(u \vee s)) du = \int_0^t \left(\int_{s \vee u}^t \mathcal{L}'_l(z) dz \right) du = \int_0^t \left(\int_{s \vee u}^t \mathcal{L}'_r(z) dz \right) du \\ &= \int_s^t \mathcal{L}'_r(z) \left(\int_0^z du \right) dz = \int_s^t z \mathcal{L}'_r(z) dz = \int_s^t z \mathcal{L}'_l(z) dz. \end{aligned}$$

Let $a, b \in \mathbb{R} \cup \{\pm\infty\}$, $a < b$, such that $\{p > 0\} = (a, b)$. Since p is \mathcal{L} -concave, $\mathcal{L}p$ is concave, hence continuous on (a, b) and there exists $c \in \mathbb{R} \cup \{\pm\infty\}$, $a \leq c \leq b$, such that $\mathcal{L}p$ is nondecreasing on (a, c) and nonincreasing on (c, b) (with the convention that a function is always monotone on the empty set). In particular, if c belongs to (a, b) , $\mathcal{L}p$ is continuously nondecreasing on $(a, c]$, continuously nonincreasing on $[c, b)$ and the same holds for $p = \mathcal{L}^{-1} \circ \mathcal{L}p$ since \mathcal{L}^{-1} is increasing. Let us now consider the case where $c \in \{a, b\}$. Our aim is to show that c is necessarily a finite number and that p can be continuously extended at c . We may then conclude that in any case, $\mathcal{L}p$ and p are both continuous and monotone on the (possibly empty) intervals $(a, c]$ and $[c, b)$ for some real number $c \in [a, b]$.

Let us start with the case $c = a$. Then $\mathcal{L}p$ and p are both continuously nonincreasing on (c, b) and since p is a density, c is necessarily finite. Let us assume that $p(x)$ tends to $+\infty$ when x tends to c and get a contradiction. Since $\mathcal{L}p$ is concave and nonincreasing on (c, b) , it has a finite limit at c and since under our assumption p grows to infinity at c ,

$$v = \lim_{y \uparrow +\infty} \mathcal{L}(y) = \lim_{x \downarrow c} \mathcal{L}p(x) < +\infty.$$

We may therefore extend $\mathcal{L}p$ continuously at $x = c$ by setting $\mathcal{L}p(c) = v \in \mathbb{R}$. Let $x_0 \in (c, b)$ and set $s = p(x_0) > 0$. It follows from the monotonicity of \mathcal{L} that $\mathcal{L}p(x_0) = \mathcal{L}(s) < \mathcal{L}p(c) = v$. Besides, $\mathcal{L}p$ lies above its chord on the interval $[c, x_0]$ since it is concave and we obtain that

$$v > \mathcal{L}p(x) \geq \theta(x - c) + v > \mathcal{L}p(x_0) > u \quad \text{for all } x \in (c, x_0) \text{ with } \theta = \frac{\mathcal{L}(s) - v}{x_0 - c} < 0.$$

Consequently, using that \mathcal{L}^{-1} is increasing, we get

$$1 \geq \int_c^{x_0} p(x) dx = \int_c^{x_0} \mathcal{L}^{-1}(\mathcal{L}p(x)) dx \geq \int_c^{x_0} \mathcal{L}^{-1}(\theta(x-c) + v) dx \geq \frac{1}{|\theta|} \int_{\mathcal{L}(s)}^v \mathcal{L}^{-1}(u) du.$$

In particular, \mathcal{L}^{-1} is integrable on $[\mathcal{L}(s), v)$. This is in fact impossible under our assumption (40). If \mathcal{L}^{-1} were integrable, we would deduce from (92) and the fact that since \mathcal{L} is increasing that

$$\begin{aligned} \int_{\mathcal{L}(s)}^v \mathcal{L}^{-1}(u) du &\geq \int_{\mathcal{L}(\lambda s)}^v \mathcal{L}^{-1}(u) du = \int_{\lambda s}^{+\infty} z \mathcal{L}'_l(z) dz = \int_s^{+\infty} \lambda^2 u \mathcal{L}'_l(\lambda u) dz \\ &\geq \kappa \int_s^{+\infty} u \mathcal{L}'_r(u) dz = \kappa \int_{\mathcal{L}(s)}^v \mathcal{L}^{-1}(u) du > \int_{\mathcal{L}(s)}^v \mathcal{L}^{-1}(u) du, \end{aligned}$$

which is the contradiction we are looking for. The case $c = b$ is obtained by applying this result to the density $x \mapsto p(-x)$ which is nonincreasing on $(-b, -a)$ and also belongs to $\overline{\mathcal{M}}(\mathcal{L})$.

The density p may therefore be continuously extended at c so that it is nondecreasing on $(a, c]$ and nonincreasing on $[c, b)$. Let us now find an element $\bar{p} \in \overline{\mathcal{O}}_{2D+3}(\mathcal{L})$ that satisfies (42). It is actually sufficient to restrict ourselves to the case $c = 0$ for the following reasons. If p belongs to $\overline{\mathcal{M}}(\mathcal{L})$, so does the density $p_0 : x \mapsto p(x+c)$ which is nondecreasing on $(a-c, 0]$ and nonincreasing on $[0, b-c)$. If we can find $\bar{p}_0 \in \overline{\mathcal{O}}_{2D+3}(\mathcal{L})$ that satisfies (42) with p_0 and \bar{p}_0 in place of p and \bar{p} , then by the invariance property of the Lebesgue measure under translation, the same inequality holds for p and $\bar{p} : x \mapsto p_0(x-c)$. It remains to notice that the so-defined function \bar{p} is also an element of $\overline{\mathcal{O}}_{2D+3}(\mathcal{L})$. From now on, we shall therefore assume that $c = 0$.

Let us assume that $b \neq 0$. We set $b_1 = b$, $y_j = \lambda^{-j} p(0)$ for $j \in \mathbb{N}$, $x_0 = 0$ and for $j \in \mathbb{N} \setminus \{0\}$, x_j a positive solution of the equation $p(x) = y_j$ when such a solution exists and $x_j = b_1 > 0$ otherwise. Since the density p is continuous and monotone on $[0, b_1)$, a positive solution x_j always exists for each $j \geq 1$ when $b_1 = +\infty$. The situation $x_j = b_1$ for some $j \geq 1$ can therefore only occur when b_1 is finite. In this case, we extend p continuously at b_1 so that p is well-defined and continuous on $[0, b_1]$. Note that in any case, $p(x_j) \geq y_j > 0$ for all $j \in \mathbb{N}$.

In what follows, our aim is to find a nonnegative function \bar{q}_1 on $[0, +\infty)$ with the following properties. There exists some $J \geq 1$ such that for all $x \in [0, x_J]$, $0 < \bar{q}_1(x) \leq p(x)$, $\bar{q}_1(x) = 0$ for all $x \notin [0, x_J]$, $\mathcal{L} \circ \bar{q}_1$ is a $(D+1)$ -linear interpolation of $\mathcal{L}p$ on $[0, x_J]$ and

$$(93) \quad \int_0^{b_1} |p - \bar{q}_1| d\mu \leq \frac{\gamma \zeta_1}{D^2} \quad \text{with} \quad \zeta_1 = \int_0^{x_1} p(u) du > 0.$$

In what follows, we set

$$D_1 = \left\lceil \frac{D\lambda\sqrt{(\lambda-1)}}{\sqrt{2\gamma\kappa}} \right\rceil \geq \left(\frac{D\lambda\sqrt{(\lambda-1)\zeta_1}}{\sqrt{2\gamma\kappa}} \right) \vee 1.$$

Since $\mathcal{L}p$ is nonincreasing and concave on $[0, b_1)$, it admits a right and left derivative at all point $x \in (0, b_1)$ as well as a right derivative $(\mathcal{L}p)'_r(0) \in [0, +\infty)$ at 0. If b_1 is finite, the limit $\lim_{x \uparrow b_1} (\mathcal{L}p)'_l(x) \in [-\infty, 0]$ exists and we shall denote it $(\mathcal{L}p)'_l(b_1)$. Throughout the proof, we consider a function ϕ such that for all $x \in (0, b_1)$, $(\mathcal{L}p)'_r(x) \leq \phi(x) \leq (\mathcal{L}p)'_l(x) \leq 0$, hence $|(\mathcal{L}p)'_l(x)| \leq |\phi(x)| \leq |(\mathcal{L}p)'_r(x)|$, and $\phi(b_1) = (\mathcal{L}p)'_l(b_1) \in [-\infty, 0]$ when $b_1 < +\infty$. Since $\mathcal{L}p$ is concave

$$0 \geq (\mathcal{L}p)'_r(0) \geq \Delta_1 = \frac{\mathcal{L}p(x_1) - \mathcal{L}p(0)}{x_1} \geq (\mathcal{L}p)'_l(x_1) \geq \phi(x_1).$$

If $|(\mathcal{L}p)'_i(x_1)|$ is finite, which is always the case when $x_1 \neq b_1$, by using the convention $0/0 = 0$ and applying Proposition 9-(i) to $\mathcal{L}p$, we may find a D_1 -linear interpolation ℓ_1 of $\mathcal{L}p$ on the interval $[0, x_1]$ such that

$$\begin{aligned} \int_0^{x_1} |\mathcal{L}p - \ell_1| d\mu &\leq \frac{x_1^2}{2D_1^2} \frac{((\mathcal{L}p)'_r(0) - \Delta_1)(\Delta_1 - (\mathcal{L}p)'_i(x_1))}{(\mathcal{L}p)'_r(0) - (\mathcal{L}p)'_i(x_1)} \\ &= \frac{x_1^2}{2D_1^2} \left[1 - \frac{\Delta_1 - (\mathcal{L}p)'_i(x_1)}{(\mathcal{L}p)'_r(0) - (\mathcal{L}p)'_i(x_1)} \right] (\Delta_1 - (\mathcal{L}p)'_i(x_1)) \\ &\leq \frac{x_1^2 |\Delta_1|}{2D_1^2} \left(1 - \frac{|\Delta_1|}{|(\mathcal{L}p)'_i(x_1)|} \right) \leq \frac{x_1 (\mathcal{L}p(0) - \mathcal{L}p(x_1))}{2D_1^2} \left(1 - \frac{|\Delta_1|}{|\phi(x_1)|} \right). \end{aligned}$$

It follows from Proposition 9-(iii) that the last inequality remains actually true when $x_1 = b_1 < +\infty$ and $\phi(b_1) = (\mathcal{L}p)'_i(b_1) = -\infty$ with the convention $1/(+\infty) = 0$. Since \mathcal{L} is concave on $(0, +\infty)$, it is Lipschitz on $[y_1, y_0] \subset (0, +\infty)$ with Lipschitz constant not larger than $\mathcal{L}'_r(y_1)$. Using that $p(0) = y_0$ and $p(x_1) \in [y_1, y_0]$, we deduce

$$(94) \quad \int_0^{x_1} |\mathcal{L}p - \ell_1| d\mu \leq \frac{x_1 \mathcal{L}'_r(y_1)(y_0 - y_1)}{2D_1^2} \left(1 - \frac{|\Delta_1|}{|\phi(x_1)|} \right) = \frac{(\lambda - 1)x_1 \mathcal{L}'_r(y_1)y_1}{2D_1^2} \left(1 - \frac{|\Delta_1|}{|\phi(x_1)|} \right).$$

Since $\mathcal{L}p$ is concave, nonincreasing and $p(x_1) \geq y_1 > 0$,

$$u < \mathcal{L}(y_1) \leq \mathcal{L}p(x_1) \leq \ell_1(x) \leq \mathcal{L}p(x) \leq \mathcal{L}p(0) = \mathcal{L}(y_0) < v \quad \text{for all } x \in [0, x_1].$$

Since $\ell_1(x) \in (u, v)$ for $x \in [0, x_1]$, we may define the function $q_1 = \mathcal{L}^{-1} \circ \ell_1$ on $[0, x_1]$. Since \mathcal{L}^{-1} is increasing, the function q_1 satisfies $0 < q_1(x) \leq p(x)$ for all $x \in [0, x_1]$. Furthermore, since \mathcal{L}^{-1} is convex on (u, v) , it is Lipschitz on $[\mathcal{L}(y_1), \mathcal{L}(y_0)]$ with a Lipschitz constant not larger than $(\mathcal{L}^{-1})'_i(\mathcal{L}(y_0)) = (\mathcal{L}'_i(y_0))^{-1} = (\mathcal{L}'_i(\lambda y_1))^{-1}$. Using (40) and (94), we get

$$\begin{aligned} \int_0^{x_1} |p(x) - q_1(x)| dx &= \int_0^{x_1} |\mathcal{L}^{-1}(\mathcal{L}p(x)) - \mathcal{L}^{-1}(\ell_1(x))| dx \leq \frac{1}{\mathcal{L}'_i(\lambda y_1)} \int_0^{x_1} |\mathcal{L}p(x) - \ell_1(x)| dx \\ &\leq \frac{(\lambda - 1)x_1 \mathcal{L}'_r(y_1)y_1}{2\mathcal{L}'_i(\lambda y_1)D_1^2} \left(1 - \frac{|\Delta_1|}{|\phi(x_1)|} \right) \leq \frac{\lambda^2(\lambda - 1)x_1 y_1}{2\kappa D_1^2} \left(1 - \frac{|\Delta_1|}{|\phi(x_1)|} \right). \end{aligned}$$

Since $p(x) \geq y_1$ for all $x \in [0, x_1]$, $x_1 y_1 \leq \int_0^{x_1} p(u) du = \zeta_1 \leq 1$ and we obtain that for our choice of D_1

$$(95) \quad \int_0^{x_1} |p - q_1| d\mu \leq \frac{\lambda^2(\lambda - 1)\zeta_1}{2\kappa D_1^2} \left(1 - \frac{|\Delta_1|}{|\phi(x_1)|} \right) \leq \frac{\gamma \zeta_1}{D^2} \left(1 - \frac{|\Delta_1|}{|\phi(x_1)|} \right) \leq \frac{\gamma \zeta_1}{D^2}.$$

If $x_1 = b_1$ the function $\bar{q}_1 = q_1 \mathbb{1}_{[0, b_1]}$ satisfies our requirements: inequality (93) follows from (95) and it follows from our definition (41) of γ that $\mathcal{L} \circ \bar{q}_1$ is a D_1 -linear interpolation on $[0, x_1]$ with

$$D_1 \leq \frac{D\lambda\sqrt{(\lambda - 1)}}{\sqrt{2\kappa\gamma}} + 1 \leq D + 1.$$

Otherwise $x_1 \neq b_1$, $p(x_1) = y_1$ and

$$\Delta_1 = \frac{\mathcal{L}p(x_1) - \mathcal{L}p(0)}{x_1} = \frac{\mathcal{L}(y_1) - \mathcal{L}(y_0)}{x_1} < 0.$$

In such a situation, we set for $j \geq 1$

$$v_j = \sqrt{(y_{j-1} - y_j) [\mathcal{L}(y_{j-1}) - \mathcal{L}(y_j)]}, \quad w_j = \int_{y_j}^{y_{j-1}} z \mathcal{L}'_i(z) dz$$

for $j \geq 2$,

$$D_j = \left\lceil \frac{\lambda v_j D}{2\sqrt{2\zeta_1 \kappa \gamma} |\Delta_1|} \right\rceil \geq \left(\frac{\lambda v_j D}{2\sqrt{2\zeta_1 \kappa \gamma} |\Delta_1|} \right) \vee 1,$$

and

$$J^* = \left\lceil \frac{2}{\log \kappa} \left(\frac{\kappa}{\kappa - 1} \right)^{1/2} \frac{D}{\sqrt{\gamma}} \right\rceil \leq \left(\frac{2}{\log \kappa} \left(\frac{\kappa}{\kappa - 1} \right)^{1/2} \frac{D}{\sqrt{\gamma}} \right) + 1.$$

Since \mathcal{L} is concave, we deduce from (40) that for all $j \geq 1$,

$$\begin{aligned} v_{j+1}^2 &= (y_j - y_{j+1}) [\mathcal{L}(y_j) - \mathcal{L}(y_{j+1})] = (y_j - y_{j+1}) \int_{y_{j+1}}^{y_j} \mathcal{L}'_r(t) dt = \frac{y_j - y_{j+1}}{\lambda} \int_{y_j}^{y_{j-1}} \mathcal{L}'_r(u/\lambda) du \\ &\leq \frac{\lambda(y_j - y_{j+1})}{\kappa} \int_{y_j}^{y_{j-1}} \mathcal{L}'_i(u) du = \frac{1}{\kappa} (y_{j-1} - y_j) [\mathcal{L}(y_{j-1}) - \mathcal{L}(y_j)] = \frac{v_j^2}{\kappa}. \end{aligned}$$

By using (90), we obtain similarly that

$$w_{j+1} = \int_{y_{j+1}}^{y_j} z \mathcal{L}'_r(z) dz = \frac{1}{\lambda^2} \int_{y_j}^{y_{j-1}} u \mathcal{L}'_r(u/\lambda) du \leq \frac{1}{\kappa} \int_{y_j}^{y_{j-1}} u \mathcal{L}'_i(u) du = \frac{w_j}{\kappa}.$$

Besides, since $\mathcal{L}p$ is concave, it lies above its chord on the interval $[0, x_1]$ which means that

$$\mathcal{L}p(x) \geq \frac{\mathcal{L}(p(x_1)) - \mathcal{L}(p(0))}{x_1} x + \mathcal{L}(p(0)) = \Delta_1 x + \mathcal{L}(y_0) \quad \text{for all } x \in [0, x_1]$$

and consequently,

$$\begin{aligned} w_1 &= \int_{y_1}^{y_0} z \mathcal{L}'_i(z) dz = \int_{\mathcal{L}(y_1)}^{\mathcal{L}(y_0)} \mathcal{L}^{-1}(u) du = |\Delta_1| \int_0^{x_1} \mathcal{L}^{-1}(\Delta_1 x + \mathcal{L}(y_0)) dx \\ &\leq |\Delta_1| \int_0^{x_1} \mathcal{L}^{-1}(\mathcal{L}p(x)) dx = |\Delta_1| \int_0^{x_1} p(x) dx = |\Delta_1| \zeta_1. \end{aligned}$$

We deduce from these inequalities that

$$(96) \quad \sum_{j \geq 2} v_j \leq \frac{v_1}{\sqrt{\kappa} - 1} \quad \text{and for all } J \geq 1 \quad \sum_{j \geq J+1} w_j \leq \frac{\kappa^{1-J} w_1}{\kappa - 1} \leq \frac{\kappa^{1-J} |\Delta_1| \zeta_1}{\kappa - 1}.$$

Since

$$\left(\frac{v_1}{\sqrt{|\Delta_1|}} \right)^2 = \frac{x_1(y_0 - y_1) [\mathcal{L}(y_0) - \mathcal{L}(y_1)]}{\mathcal{L}(y_0) - \mathcal{L}(y_1)} = (\lambda - 1) x_1 y_1 \leq (\lambda - 1) \zeta_1,$$

we derive from (96) that

$$\begin{aligned} D_1 + \sum_{j=2}^{J^*} D_j &\leq \frac{D\lambda\sqrt{(\lambda-1)}}{\sqrt{2\kappa\gamma}} + 1 + \sum_{j=2}^{J^*} \left[1 + \frac{\lambda v_j D}{2\sqrt{2\zeta_1 \kappa \gamma} |\Delta_1|} \right] \leq J^* + \frac{\lambda D}{\sqrt{2\kappa\gamma}} \left(\sqrt{\lambda-1} + \frac{1}{2\sqrt{\zeta_1 |\Delta_1|}} \sum_{j \geq 2} v_j \right) \\ &\leq J^* + \frac{\lambda D}{\sqrt{2\kappa\gamma}} \left(\sqrt{\lambda-1} + \frac{v_1}{2(\sqrt{\kappa}-1)\sqrt{\zeta_1 |\Delta_1|}} \right) \leq J^* + \frac{\lambda\sqrt{\lambda-1} D}{\sqrt{2\kappa\gamma}} \left(1 + \frac{1}{2(\sqrt{\kappa}-1)} \right) \\ &\leq \left[\frac{2}{\log \kappa} \left(\frac{\kappa}{\kappa-1} \right)^{1/2} + \frac{\lambda\sqrt{\lambda-1}}{\sqrt{2\kappa}} \left(1 + \frac{1}{2(\sqrt{\kappa}-1)} \right) \right] \frac{D}{\sqrt{\gamma}} + 1. \end{aligned}$$

which, with the value of γ given by (41) leads to

$$(97) \quad D_1 + \sum_{j=1}^{J^*} D_j \leq D + 1.$$

As long as $x_{j-1} \neq b_1$ with $j \geq 2$, the interval $[x_{j-1}, x_j]$ does not reduce to a singleton and by Proposition 9-(ii), we may find a D_j -linear interpolation ℓ_j of $\mathcal{L}p$ on this interval that satisfies

$$(98) \quad \begin{aligned} \int_{x_{j-1}}^{x_j} |\mathcal{L}p - \ell_j| d\mu &\leq \frac{(\mathcal{L}(y_{j-1}) - \mathcal{L}(y_j))^2}{8D_j^2} \left| \frac{1}{(\mathcal{L}p)'_r(x_{j-1})} - \frac{1}{(\mathcal{L}p)'_l(x_j)} \right| \\ &= \frac{(\mathcal{L}(y_{j-1}) - \mathcal{L}(y_j))^2}{8D_j^2} \left(\frac{1}{|(\mathcal{L}p)'_r(x_{j-1})|} - \frac{1}{|(\mathcal{L}p)'_l(x_j)|} \right) \\ &\leq \frac{(\mathcal{L}(y_{j-1}) - \mathcal{L}(y_j))^2}{8|\Delta_1|D_j^2} \left(\frac{|\Delta_1|}{|\phi(x_{j-1})|} - \frac{|\Delta_1|}{|\phi(x_j)|} \right) \end{aligned}$$

Since $p(x_j) \geq y_j > 0$,

$$u < \mathcal{L}(y_j) \leq \mathcal{L}p(x_j) \leq \ell_1(x) \leq \mathcal{L}p(x) \leq \mathcal{L}p(x_{j-1}) < v \quad \text{for all } x \in [x_{j-1}, x_j]$$

and as before, we may define the function $q_j = \mathcal{L}^{-1} \circ \ell_j$ on $[x_{j-1}, x_j]$. It satisfies $0 < q_j(x) \leq p(x)$ for all $x \in [x_{j-1}, x_j]$ and since \mathcal{L}^{-1} is Lipschitz on $[\mathcal{L}(y_j), \mathcal{L}(y_{j-1})]$ with a Lipschitz constant not larger than $(\mathcal{L}^{-1})'_l(\mathcal{L}(y_{j-1})) = (\mathcal{L}'_l(y_{j-1}))^{-1} = (\mathcal{L}'_l(\lambda y_j))^{-1}$, we deduce from (98) that

$$\int_{x_{j-1}}^{x_j} |p - q_j| d\mu = \int_{x_{j-1}}^{x_j} |\mathcal{L}^{-1}(\mathcal{L}p(x)) - \mathcal{L}^{-1}(\ell_j(x))| dx \leq \frac{(\mathcal{L}(y_{j-1}) - \mathcal{L}(y_j))^2}{8\mathcal{L}'_l(\lambda y_j)|\Delta_1|D_j^2} \left(\frac{|\Delta_1|}{|\phi(x_{j-1})|} - \frac{|\Delta_1|}{|\phi(x_j)|} \right).$$

Using (40) and the fact that \mathcal{L} is concave,

$$\mathcal{L}'_l(\lambda y_j) \geq \frac{\kappa}{\lambda^2} \mathcal{L}'_r(y_j) \geq \frac{\kappa}{\lambda^2} \frac{\mathcal{L}(y_{j-1}) - \mathcal{L}(y_j)}{y_{j-1} - y_j} = \frac{\kappa}{\lambda^2} \frac{(\mathcal{L}(y_{j-1}) - \mathcal{L}(y_j))^2}{v_j^2}$$

which, with our choice of D_j , leads to

$$(99) \quad \int_{x_{j-1}}^{x_j} |p - q_j| d\mu \leq \frac{\lambda^2 v_j^2}{8\kappa|\Delta_1|D_j^2} \left(\frac{|\Delta_1|}{|\phi(x_{j-1})|} - \frac{|\Delta_1|}{|\phi(x_j)|} \right) = \frac{\gamma \zeta_1}{D^2} \left(\frac{|\Delta_1|}{|\phi(x_{j-1})|} - \frac{|\Delta_1|}{|\phi(x_j)|} \right).$$

If $J \geq 1$ and $x_J \neq b_1$, $p(x_J) = y_J$ and since $\mathcal{L}p$ is concave, it lies under its tangent at x_J (which is nonpositive) so that for all $x \in [x_J, b_1)$

$$u < \mathcal{L}p(x) \leq \mathcal{L}p(x_J) + (\mathcal{L}p)'_r(x_J)(x - x_J) \leq \mathcal{L}p(x_J) = \mathcal{L}(y_J) < v.$$

Doing the change of variables $t = \mathcal{L}p(x_J) + (\mathcal{L}p)'_r(x_J)(x - x_J) \in (u, v)$ and using (91), (96) and the monotonicity of \mathcal{L}^{-1} we obtain that

$$\begin{aligned} \int_{x_J}^{b_1} p(x) dx &= \int_{x_J}^{b_1} \mathcal{L}^{-1}(\mathcal{L}p(x)) dx \leq \int_{x_J}^{b_1} \mathcal{L}^{-1}(\mathcal{L}p(x_J) + (\mathcal{L}p)'_r(x_J)(x - x_J)) dx \\ &\leq \frac{1}{|(\mathcal{L}p)'_r(x_J)|} \int_u^{\mathcal{L}(y_J)} \mathcal{L}^{-1}(t) dt = \frac{1}{|(\mathcal{L}p)'_r(x_J)|} \int_0^{y_J} z \mathcal{L}'_l(z) dz \\ &= \frac{1}{|(\mathcal{L}p)'_r(x_J)|} \sum_{j \geq J+1} w_j \leq \frac{\kappa^{1-J} |\Delta_1| \zeta_1}{\kappa - 1 |(\mathcal{L}p)'_r(x_J)|}. \end{aligned}$$

Note that this inequality is clearly satisfied when $x_J = b_1$, hence for all $J \geq 1$. It is in particular true for $J = J^* \geq 1$ and since by definition of J^* ,

$$\kappa^{J^*} = \exp(J^* \log \kappa) \geq \left(1 + \frac{1}{2} J^* \log \kappa\right)^2 \geq \left(\frac{1}{2} J^* \log \kappa\right)^2 \geq \frac{\kappa}{\kappa - 1} \frac{D^2}{\gamma} \quad \text{hence} \quad \frac{\kappa^{1-J^*}}{\kappa - 1} \leq \frac{\gamma}{D^2},$$

we obtain that for $J = J^*$

$$(100) \quad \int_{x_J}^{b_1} p(x) dx \leq \frac{\gamma \zeta_1}{D^2} \frac{|\Delta_1|}{|(\mathcal{L}p)'_r(x_J)|} \leq \frac{\gamma \zeta_1}{D^2} \frac{|\Delta_1|}{|\phi(x_J)|}.$$

Let $\bar{J} = \inf\{j \geq 1, x_j = b_1\}$, with the convention $\inf \emptyset = +\infty$, and $J = J^* \wedge \bar{J} \geq 1$. The function

$$\bar{q}_1 = q_1 \mathbb{1}_{[0, x_1]} + \sum_{j=2}^J q_j \mathbb{1}_{(x_{j-1}, x_j]}$$

satisfies $0 < \bar{q}_1(x) \leq p(x)$ for all $x \in [0, x_J]$, $\bar{q}_1(x) = 0$ outside the interval $[0, x_J]$. Furthermore, it follows from (97) that $\mathcal{L} \circ \bar{q}_1$ is a \bar{D}_1 -linear interpolation of $\mathcal{L}p$ on $[0, x_J]$ with

$$\bar{D}_1 \leq D_1 + \sum_{j=2}^J D_j \leq D_1 + \sum_{j=2}^{J^*} D_j \leq D + 1$$

and by putting the inequalities (95), (99) and (100) together, we obtain that

$$\begin{aligned} \int_0^{b_1} |p - \bar{q}_1| d\mu &\leq \sum_{j=1}^J \int_{x_{j-1}}^{x_j} |p - \bar{q}_j| d\mu + \int_{x_J}^{b_1} |p - \bar{q}_j| d\mu \\ &\leq \frac{\gamma \zeta_1}{D^2} \left(1 - \frac{|\Delta_1|}{|\phi(x_1)|} + \sum_{j=2}^J \left(\frac{|\Delta_1|}{|\phi(x_{j-1})|} - \frac{|\Delta_1|}{|\phi(x_j)|}\right) + \frac{|\Delta_1|}{|\phi(x_J)|}\right) = \frac{\gamma \zeta_1}{D^2}. \end{aligned}$$

The function \bar{q}_1 satisfies thus our requirements.

If $a \neq 0$, by applying the previous result to the density $x \mapsto p(-x)$, which also belongs to $\overline{\mathcal{M}}(\mathcal{L})$, we obtain that there exists numbers $x_{-J} \leq x_{-1} < 0$ and a function \bar{q}_{-1} which vanishes outside $[x_{-J}, 0]$ that satisfies the following properties: $0 < \bar{q}_{-1}(x) \leq p(x)$ for all $x \in [x_{-J}, 0]$, $\mathcal{L} \circ \bar{q}_{-1}$ is a \bar{D}_{-1} -linear interpolation of $\mathcal{L}p$ on $[x_{-J}, 0]$ with $\bar{D}_{-1} \leq D + 1$ and

$$\int_a^0 |p - \bar{q}_{-1}| d\mu \leq \frac{\gamma \zeta_{-1}}{D^2} \quad \text{with} \quad \zeta_{-1} = \int_{x_{-1}}^0 p(x) dx.$$

By taking $\bar{q} = \bar{q}_1$ and $x_{-1} = x_{-J} = 0$ when $a = 0$, $\bar{q} = \bar{q}_{-1}$ and $x_1 = x_J = 0$ when $b = 0$ and $\bar{q} = q_{-1} \mathbb{1}_{[x_{-J}, 0]} + \bar{q}_1 \mathbb{1}_{(0, x_J]}$ otherwise, we obtain a function with the following properties. For all $x \in [x_{-J}, x_J]$, $0 < \bar{q}(x) \leq p(x)$ and $\bar{q}(x) = 0$ for all $x \notin [x_{-J}, x_J]$. The function $\mathcal{L} \circ \bar{q}$ is a \bar{D} -linear interpolation of $\mathcal{L}p$ on the interval $[x_{-J}, x_J]$ with $\bar{D} \leq 2(D + 1)$, it is in particular concave on this interval, and

$$\int_{\mathbb{R}} |p - \bar{q}| d\mu = \int_a^b |p - \bar{q}| d\mu \leq \frac{\gamma}{D^2} (\zeta_{-1} + \zeta_1) \leq \frac{\gamma}{D^2}.$$

By construction, the linear interpolation $\mathcal{L} \circ \bar{q}$ of $\mathcal{L}p$ on $[x_{-J}, x_J]$ is unimodal on this interval with a mode at 0. Since \mathcal{L}^{-1} is increasing, the same is true for \bar{q} . We finally define

$$\bar{p}(x) = \bar{q}(\eta x) \mathbb{1}_{(-x_J, x_J)}(\eta x) \quad \text{with} \quad \eta = \int_{-x_J}^{x_J} \bar{q}(u) du = \int_{\mathbb{R}} \bar{q}(u) du.$$

Since $0 < \bar{q}(x) \leq p(x)$ for all $x \in [x_{-J}, x_J]$, $\eta \in (0, 1]$. By construction, \bar{p} is a density for which the set $\{\bar{p} > 0\}$ is the open interval $(x_{-J}/\eta, x_J/\eta)$. Since $\mathcal{L} \circ \bar{q}$ is a \bar{D} -linear interpolation of $\mathcal{L}p$ on $[x_{-J}, x_J]$, it is concave on this interval and $\mathcal{L}\bar{p} = \mathcal{L} \circ \bar{q}(\eta \cdot)$ is therefore concave on $\{\bar{p} > 0\} = (x_{-J}/\eta, x_J/\eta)$ and affine on each element of a partition of this interval into at most \bar{D} intervals. The function $\mathcal{L}\bar{p}$ takes the value $-\infty$ on $(-\infty, x_{-J}/\eta]$ and $[x_J/\eta, +\infty)$ and is therefore affine or takes the value $-\infty$ on a partition of \mathbb{R} into at most $\bar{D} + 2$ intervals. The density \bar{p} therefore belongs to $\overline{\mathcal{O}}_{\bar{D}+1}(\mathcal{L}) \subset \overline{\mathcal{O}}_{2\bar{D}+3}(\mathcal{L})$. Since \bar{q} is nondecreasing on $(x_{-J}, 0)$ and nonincreasing on $(0, x_J)$, Lemma 4 applies with $f = \bar{q}$, $c = 0$ and $\kappa = \eta$, which leads to

$$\int_{\mathbb{R}} |p - \bar{p}| d\mu \leq 2 \int_{\mathbb{R}} |p - \bar{q}| d\mu \leq \frac{2\gamma}{D^2}.$$

This completes the proof of Theorem 9. □

ACKNOWLEDGMENTS

One of the authors thanks Lutz Dümbgen for asking if we could establish a result with $d(P^*, \hat{P})$ in place of $n^{-1} \sum_{i=1}^n d(P_i, \hat{P})$ at a conference where a preliminary version of this work was presented. His question spurred the authors to improve their result and obtain this new version of Theorem 1.

REFERENCES

- [1] Balabdaoui, F. and Wellner, J. A. (2007). A Kiefer-Wolfowitz theorem for convex densities. In *Asymptotics: particles, processes and inverse problems*, volume 55 of *IMS Lecture Notes Monogr. Ser.*, pages 1–31. Inst. Math. Statist., Beachwood, OH.
- [2] Baraud, Y. (2016). Bounding the expectation of the supremum of an empirical process over a (weak) vc-major class. *Electron. J. Statist.*, 10(2):1709–1728.
- [3] Baraud, Y. (2021). Tests and estimation strategies associated to some loss functions. *Probab. Theory Relat. Fields*, 180(3):799–846.
- [4] Baraud, Y. and Birgé, L. (2016). Rho-estimators for shape restricted density estimation. *Stochastic Process. Appl.*, 126(12):3888–3912.
- [5] Baraud, Y. and Birgé, L. (2018). Rho-estimators revisited: General theory and applications. *Ann. Statist.*, 46(6B):3767–3804.
- [6] Birgé, L. (1987). On the risk of histograms for estimating decreasing densities. *Ann. Statist.*, 15(3):1013–1022.
- [7] Birgé, L. (1989). The Grenander estimator: a nonasymptotic approach. *Ann. Statist.*, 17(4):1532–1549.
- [8] Birgé, L. (1997). Estimation of unimodal densities without smoothness assumptions. *The Annals of Statistics*, pages 970–981.
- [9] Chan, S. O., Diakonikolas, I., Servedio, R. A., and Sun, X. (2014). Efficient density estimation via piecewise polynomial approximation. *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*.
- [10] Cule, M. and Samworth, R. (2010). Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density. *Electron. J. Stat.*, 4:254–270.
- [11] Devroye, L. and Lugosi, G. (2001). *Combinatorial methods in density estimation*. Springer Series in Statistics. Springer-Verlag, New York.
- [12] Doss, C. R. and Wellner, J. A. (2016). Global rates of convergence of the MLEs of log-concave and s -concave densities. *Ann. Statist.*, 44(3):954–981.
- [13] Dümbgen, L. and Rufibach, K. (2009). Maximum likelihood estimation of a log-concave density and its distribution function: basic properties and uniform consistency. *Bernoulli*, 15(1):40–68.

- [14] Feng, O. Y., Guntuboyina, A., Kim, A. K. H., and Samworth, R. J. (2021). Adaptation in multivariate log-concave density estimation. *Ann. Statist.*, 49(1):129 – 153.
- [15] Gao, F. and Wellner, J. A. (2009a). On the rate of convergence of the maximum likelihood estimator of a k -monotone density. *Sci. China Ser. A*, 52(7):1525–1538.
- [16] Grenander, U. (1956). On the theory of mortality measurement. II. *Skand. Aktuarietidskr.*, 39:125–153 (1957).
- [17] Grenander, U. (1981). *Abstract inference*. John Wiley & Sons, Inc., New York. Wiley Series in Probability and Mathematical Statistics.
- [18] Groeneboom, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer, Vol. II (Berkeley, Calif., 1983)*, Wadsworth Statist./Probab. Ser., pages 539–555. Wadsworth, Belmont, CA.
- [19] Groeneboom, P., Jongbloed, G., and Wellner, J. A. (2001). Estimation of a Convex Function: Characterizations and Asymptotic Theory. *Ann. Statist.*, 29(6):1653–1698.
- [20] Guérin, J., Marcotte, P., and Savard, G. (2006). An optimal adaptive algorithm for the approximation of concave functions. *Mathematical Programming*, 107(3):357–366.
- [21] Han, Q. and Wellner, J. A. (2016). Approximation and estimation of s -concave densities via Rényi divergences. *Ann. Statist.*, 44(3):1332–1359.
- [22] Kim, A. K. H., Guntuboyina, A., and Samworth, R. J. (2018). Adaptation in log-concave density estimation. *Ann. Statist.*, 46(5):2279–2306.
- [23] Kim, A. K. H. and Samworth, R. J. (2016). Global rates of convergence in log-concave density estimation. *Ann. Statist.*, 44(6):2756 – 2779.
- [24] Koenker, R. and Mizera, I. (2010). Quasi-concave density estimation. *Ann. Statist.*, 38(5):2998–3027.
- [25] Kur, G., Dagan, Y., and Rakhlin, A. (2019). Optimality of maximum likelihood for log-concave density estimation and bounded convex regression. *arXiv preprint arXiv:1903.05315*.
- [26] Rao, B. P. (1969). Estimation of a unimodal density. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 23–36.
- [27] Seregin, A. and Wellner, J. (2010). Nonparametric estimation of convex-transformed densities. *Ann. Statist.*, 38(6):3751–3781.
- [28] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes. With Applications to Statistics*. Springer Series in Statistics. Springer-Verlag, New York.
- [29] Wegman, E. J. (1970). Maximum likelihood estimation of a unimodal density, ii. *The Annals of Mathematical Statistics*, pages 2169–2174.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF LUXEMBOURG, MAISON DU NOMBRE, 6 AVENUE DE LA FONTE, L-4364 ESCH-SUR-ALZETTE, GRAND DUCHY OF LUXEMBOURG

Email address: `yannick.baraud@uni.lu`

SAMOVAR, TELECOM SUDPARIS, INSTITUT POLYTECHNIQUE DE PARIS, FRANCE

Email address: `helene.halconruy@telecom-sudparis.eu`

Email address: `guillaume.maillard@ensai.fr`

ENSAI CAMPUS DE KER LANN 51 RUE BLAISE PASCAL BP 37203 35172 BRUZ CEDEX, FRANCE