ROBUST DENSITY ESTIMATION WITH THE $\mathbb{L}_1$-LOSS. APPLICATIONS TO THE ESTIMATION OF A DENSITY ON THE LINE SATISFYING A SHAPE CONSTRAINT

YANNICK BARAUD, HÉLÈNE HALCONRUY, AND GUILLAUME MAILLARD

Abstract. We solve 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 density models 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 target densities in the model for which our estimator is proven to converge at parametric rate. 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. 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 belongs to it. Our main result on the risk of the estimator takes the form of an exponential deviation inequality which is non-asymptotic and involves explicit numerical constants. We deduce from it several global rates of convergence, including some bounds for the minimax $\mathbb{L}_1$-risks over the sets of concave and log-concave densities. These bounds derive from some specific results on the approximation of densities which are monotone, convex, concave and log-concave. Such results may be of independent interest.

1. Introduction

Estimating a density under a shape constraint has been addressed by many authors since the pioneer papers by Grenander (1956; 1981), Rao (1969),

---

Date: June 24, 2022.
1991 *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.

Groeneboom (1985) and Birgé (1989) for estimating a nonincreasing density on $(0, +\infty)$. It is well-known that this problem can elegantly be solved by the Grenander estimator – see Grenander (1956)– which is the maximum likelihood estimator (MLE for short) over the set of densities that satisfy this monotonicity constraint on $(0, +\infty)$. Rao (1969) and Grenander (1956; 1981) established the local asymptotic properties of the Grenander estimator while Groeneboom (1985) and Birgé (1989) studied its global estimation errors for the $\mathbb{L}_1$-loss on some functional classes of interest. Birgé proved that the uniform risk of the Grenander estimator over the set $\mathscr{F}(H, L)$, that consists of nonincreasing densities bounded by $H > 0$ and supported on $[0, L]$ with $L > 0$, is of order $(\log(1 + HL)/n)^{1/3}$. Since this rate is optimal, the Grenander estimator performs almost as well (apart maybe from numerical constants) as a minimax estimator over $\mathscr{F}(H, L)$ that would know the values of H and L in advance. Even more surprising is the fact that the Grenander estimator can converge at a parametric rate $(1/\sqrt{n})$ when the density is piecewise constant on the elements of partition of $(0, +\infty)$ into a finite number of intervals – see Grenander (1981), Groeneboom (1985) and Birgé (1989). As a consequence, the Grenander estimator can nicely *adapt* to the specific features of the target density even though these features are *a priori* unknown.

Because of these *adaptation* properties for estimating a monotone density, 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* (2001) for convex densities and to Balabdaoui and Wellner (2007) and Gao and Wellner (2009a) for k-monotone ones. Although the construction of the MLE is not based on any smoothness assumption in these cases, it still needs to have some pieces of information on the support of the target density. It was already the case for the monotonicity constraint since the left-endpoint of the support of the density 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 the search of alternative estimators becomes necessary to solve this issue.

To our knowledge, the first attempt to solve it dates back to Wegman (1970). He designed a 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. Wegman and Grenander estimators both converge at the same rate except on an interval of length ε around the mode. Birgé (1997) proposed a different approach based on data-driven choice of a Grenander estimator among the collection of those associated to all possible modes. 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, up to an additional term of order $1/\sqrt{n}$.

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 the MLE on the set of log-concave densities has led to an intensive work. We refer the reader to Dümbgen and K. Rufibach (2009), Doss and Wellner (2016) , Cule and Samworth (2010), Kim and Samworth (2016) and Feng *et al* (2021) as well as the references therein. Kim and Samworth (2016) 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 parametric rate (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* (2018) when $d = 1$ and extended to the dimensions $d \in \{2, 3\}$ by Feng *et al* (2021). In dimension $d \geqslant 4$, Kur *et al.* Kur et al. (2019) showed that the MLE converges at a minimax rate - up to a logarithm factor - for the Hellinger loss.

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. 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*. Except for the Grenander estimator (which is a particular case of a ρ-estimator — see Baraud and Birgé (2018)[Section 6]—, we are not aware of any result that establishes such robustness properties for the MLE.

Actually, 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* (2014) 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 (2001) 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 VC. Despite the desirable properties described above, this estimator does not possess some of the nice ones that makes the MLE so popular. For estimating a log-concave density, the MLE converges at 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 (2016)[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 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 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 (2001). When the Yatracos class associated to density model is VC, the risk bound we establish is similar to theirs except from 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 these models of densities that satisfy a shape constraint. Nevertheless, it is likely that with the same techniques of proofs, we could establish for Devroye and Lugosi's estimators the similar results as those we establish here for ours.

The theory of ℓ-estimation introduced in Baraud (2021) 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 line, 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 $p_1^\star, \ldots, p_n^\star$, we measure the performance of our density estimator $\widehat{p}$ in terms of its $\mathbb{L}_1$-distance $\|p^\star - \widehat{p}\|$ between $\widehat{p}$ and the average of the marginal densities $p^\star = n^{-1}\sum_{i=1}^n p_i^\star$. In contrast, Baraud considered, as a loss function, the average of the $\mathbb{L}_1$-distances of $\widehat{p}$ to the $p_i^\star$, i.e. the quantity $n^{-1}\sum_{i=1}^n \|\widehat{p} - p_i^\star\|$. As a consequence, unlike Baraud, we can establish the convergence of our estimator to $p^\star$, as soon as its belongs to the model, even in the unfavourable situation where none of the marginals $p_i^\star$ 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 model including some for which the MLE does not exist (the set of all monotone densities on a half-line, the set of all unimodal densities on $\mathbb{R}$, the set of all convex densities on an interval, etc).
- 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 parametric rate when the data are i.i.d. with a density that belongs to 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 on 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.

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 case of a log-concave density is tackled in 6 respectively. These sections also contain some approximation results which may be of independent interest and are central to our approach. The proofs are postponed to Section 7.

2. The statistical framework and main notations

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

The density models we have in mind are nonparametric and gather 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 $\mathscr{P}$ of all probability measures on $(\mathscr{X}, \mathscr{A})$. We denote by $\|\cdot\|$ the $\mathbb{L}_1$-norm on the set $\mathbb{L}_1(\mathscr{X}, \mathscr{A}, \mu)$ that consists of the equivalence classes of integrable functions on $(\mathscr{X}, \mathscr{A}, \mu)$. We recall that the TV-loss is a distance defined for $P, Q \in \mathscr{P}$ by

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

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, whatever $\mathbf{P}^\star$, denote by $P^\star$ the uniform mixture of the marginals:

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

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

The quantity $d(P^\star, P)$ is small as compared to 1 when only a small portion of the marginals $P_1^\star, \ldots, P_n^\star$ are far away from P. Note that this quantity can be small even if none of the $P_i^\star$ equals 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 $\mathbb{L}_1(\mathscr{X}, \mathscr{A}, \mu)$ is separable for the $\mathbb{L}_1$-norm, so is any subset $\overline{\mathcal{M}}$ of densities on $(\mathscr{X}, \mathscr{A}, \mu)$ and Assumption 1 is automatically satisfied. This is in particular the case when $(\mathscr{X}, \mathscr{A}) = (\mathbb{R}^k, \mathscr{B}(\mathbb{R}^k))$, $k \geqslant 1$ 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 to a subset $\mathcal{D}$ and a probability set $\mathscr{D} = \{P = p \cdot \mu,\ p \in \mathcal{D}\}$ that are both countable and respectively dense in $(\overline{\mathcal{D}}, \|\cdot\|)$ and $\overline{\mathscr{D}} = \{P = p \cdot \mu,\ p \in \overline{\mathcal{D}}\}$ for the total variation distance d. We may therefore write

$$\inf_{P \in \mathscr{D}} d(P^\star, P) = \inf_{P \in \overline{\mathscr{D}}} d(P^\star, P).$$

We shall repeatedly apply this equality to sets $\overline{\mathscr{D}}$ 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 $\overline{\mathcal{D}}, \mathcal{D}, \mathscr{D}, \overline{\mathscr{D}}$ in order to distinguish between the density model, a countable and dense subset of it and their corresponding probability models. Following these notations $\mathscr{M} = \{P = p \cdot \mu,\ p \in \mathcal{M}\}$. An interval I of $\mathbb{R}$ is said to be *nontrivial* if its interior $\mathring{I}$ is not empty or equivalently if its length is positive. Given $a \in \mathbb{R} \cup \{-\infty\}$ and $b \in \mathbb{R}$, $(a, b|$ denotes any of the intervals (a, b) and $(a, b]$. 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 $\mathbb{N}^{\star}$ and $|A|$ is the cardinality of a set A. By convention, $\sum_{\varnothing} = 0$. For an integrable function f on $(\mathscr{X}^{\otimes n}, \mathscr{A}^{\otimes n})$, $\mathbb{E}[f(\boldsymbol{X})]$ is the integral of f with respect to the probability measure $\bigotimes_{i=1}^{n} P_i^{\star} = \mathbf{P}^{\star}$ while for f on $(\mathscr{X}, \mathscr{A})$ and $S \in \mathscr{P}$, $\mathbb{E}_S[f(X)]$ is the integral of f with respect to S. We use the same conventions for $\mathrm{Var}\,(f(\boldsymbol{X}))$ and $\mathrm{Var}_S\,(f(X))$.

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 $\mathscr{M}$, we define

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

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

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

and

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

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

$$\mathscr{E}(\boldsymbol{X}) = \left\{P \in \mathscr{M},\ \mathbf{T}(\boldsymbol{X}, P) \leqslant \inf_{P' \in \mathscr{M}} \mathbf{T}(\boldsymbol{X}, P') + \varepsilon\right\}. \tag{4}$$

We call $\widehat{P}$ and $\widehat{p}$ a *TV-estimator* on $\mathscr{M}$ and $\mathcal{M}$ respectively. The parameter ε is introduced in case a minimizer of $P \mapsto \mathbf{T}(\boldsymbol{X}, P)$ does not exist on $\mathscr{M}$. 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 (2021) and our approach is in the same line. In particular, we use the following key property on the family $\mathscr{T}$ (which can be compared to Assumption 1 in Baraud (2021)).

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

$$d(P, Q) - d(S, Q) \leqslant \mathbb{E}_S\left[t_{(P,Q)}(X)\right] \leqslant d(S, P). \tag{5}$$

In particular,

$$d(P,Q) - d(P^\star, Q) \leqslant \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[t_{(P,Q)}(X_i)\right] \leqslant d(P^\star, P)\,, \tag{6}$$

where $P^\star = n^{-1}\sum_{i=1}^{n} P_i^\star$.

However, our family $\mathscr{T}$ 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 (2021). 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.

Our construction also shares some similarities with that proposed by Devroye and Lugosi (2001)[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

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

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

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

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

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

$$\begin{aligned}\mathbb{E}_S\left[t_{(P,Q)}(X)\right] &= S(q>p) - Q(q>p) + \left[Q(q>p) - P(q>p)\right]\\ &\geqslant -d(S,Q) + d(P,Q)\,.\end{aligned}$$

Finally, (6) results from the observation that

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

□

3.1. **Properties of the estimator.** Our main result is based on the key notion of *extremal point* in a model.

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

$$\mathscr{C}^{>}(\mathcal{F}, \overline{f}) = \left\{\{x \in \mathscr{X} \,\middle|\, q(x) > \overline{f}(x)\},\ q \in \mathcal{F} \setminus \{\overline{f}\}\right\}$$

and

$$\mathscr{C}^{<}(\mathcal{F}, \overline{f}) = \left\{ \{x \in \mathscr{X} \mid q(x) < \overline{f}(x)\},\ q \in \mathcal{F} \setminus \{\overline{f}\} \right\}$$

are both VC with dimension not larger than D.

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

Theorem 1. *Let $\overline{\mathcal{M}}$ be a density model satisfying our Assumption 1 for which $\overline{\mathcal{O}}$ is nonempty. Any TV-estimator $\widehat{P}$ on $\mathscr{M}$ satisfies for all $\xi > 0$ and all product distribution $\mathbf{P}^{\star}$, with a probability at least $1 - e^{-\xi}$, for all $D \geqslant 1$ and all $P \in \overline{\mathscr{O}}(D)$,*

$$d\left(P, \widehat{P}\right) \leqslant 2d(P^{\star}, P) + 20\sqrt{\frac{5D}{n}} + \sqrt{\frac{2(\xi + \log 2)}{n}} + \frac{\varepsilon}{n}. \tag{7}$$

In particular,

$$d\left(P^{\star}, \widehat{P}\right) \leqslant \inf_{D \geqslant 1} \left[3 \inf_{P \in \overline{\mathscr{O}}(D)} d(P^{\star}, P) + 20\sqrt{\frac{5D}{n}}\right] + \sqrt{\frac{2(\log 2 + \xi)}{n}} + \frac{\varepsilon}{n}, \tag{8}$$

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

$$\mathbb{E}\left[d\left(P, \widehat{P}\right)\right] \leqslant 2d(P^{\star}, P) + 48\sqrt{\frac{D}{n}} + \frac{\varepsilon}{n} \tag{9}$$

for all $D \geqslant 1$ and all $P \in \overline{\mathscr{O}}(D)$, moreover by (8),

$$\mathbb{E}\left[d\left(P^{\star}, \widehat{P}\right)\right] \leqslant \inf_{D \geqslant 1} \left\{3 \inf_{P \in \overline{\mathscr{O}}(D)} d(P^{\star}, P) + 48\sqrt{\frac{D}{n}}\right\} + \frac{\varepsilon}{n}. \tag{10}$$

Proof. The proof is postponed to Subsection 7.2. □

Let us now comment on this result.

In the favourable situation where the X_i are i.i.d. with distribution $P^{\star}$ in $\overline{\mathscr{O}}$, $\inf_{D \geqslant 1} \inf_{P \in \overline{\mathscr{O}}(D)} d(P^{\star}, P) = 0$ and we deduce from (10) that the estimator $\widehat{P}$ converges toward $P^{\star}$ at rate $1/\sqrt{n}$ for the total variation distance. More precisely, the risk of the estimator is not larger than $48\sqrt{D/n} + \varepsilon/n$ when $P^{\star}$ belongs to $\overline{\mathscr{O}}(D)$ for some $D \geqslant 1$. Note that the result also holds when the data are independent only, provided that $P^{\star} = n^{-1} \sum_{i=1}^{n} P_i^{\star}$ is extremal. This situation may occur even when none of the marginals $P_i^{\star}$ is extremal or even belongs to the model $\mathscr{M}$.

In the general case where the data are independent only and their marginals write for all $i \in \{1, \ldots, n\}$ as

$$P_i^\star = (1-\alpha_i)\overline{P} + \alpha_i R_i = \overline{P} + \alpha_i \left(R_i - \overline{P}\right) \tag{11}$$

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

$$\begin{aligned}\mathbb{E}\left[d\left(\overline{P}, \widehat{P}\right)\right] &\leqslant 2d\left(P^\star, \overline{P}\right) + 48\sqrt{\frac{D}{n}} + \frac{\varepsilon}{n} \\ &\leqslant \frac{2}{n}\sum_{i=1}^n \alpha_i + 48\sqrt{\frac{D}{n}} + \frac{\varepsilon}{n}.\end{aligned}$$

As compared to the previous situation where $P_i^\star = \overline{P} \in \overline{\mathscr{O}}(D)$, 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 $\sqrt{D/n}$. Note that this situation may occur even when $\alpha_i > 0$ for all i, i.e. when none of the marginals $P_i^\star$ belongs to $\overline{\mathscr{O}}(D)$. 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 $\overline{P} \in \overline{\mathscr{O}}(D)$ 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 $S \subset \{1, \ldots, n\}$ 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 $\sqrt{D/n}$. 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 $\overline{P} \in \overline{\mathscr{O}}(D)$. Then (11) holds with $\alpha_i = \alpha$ and $R_i = R$ for all $i \in \{1, \ldots, n\}$. The risk bound we get remains of the stable under contamination as long as the level $\overline{\alpha} = \alpha$ of contamination remains small as compared to $\sqrt{D/n}$.

A bound similar to (8) has been established in Baraud (2021) for his ℓ-estimators. His inequality (48) can be reformulated in our context as

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n d\left(P_i^\star, \widehat{P}\right) \leqslant \inf_{D\geqslant 1}\left[6\inf_{P\in\overline{\mathscr{O}}(D)}\left[\frac{1}{n}\sum_{i=1}^n d(P_i^\star, P)\right] + 40\sqrt{\frac{5D}{n}}\right] + 2\sqrt{\frac{2\xi}{n}} + \frac{2\varepsilon}{n} \\ - \inf_{P\in\mathscr{M}}\left[\frac{1}{n}\sum_{i=1}^n d(P_i^\star, P)\right].\end{aligned}$$

In comparison, equation (9) and the triangle inequality imply that

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n d\left(P_i^\star, \widehat{P}\right) \leqslant \inf_{D\geqslant 1}\left[3\inf_{P\in\overline{\mathscr{O}}(D)}\left[\frac{1}{n}\sum_{i=1}^n d(P_i^\star, P)\right] + 20\sqrt{\frac{5D}{n}}\right] \\ + \sqrt{\frac{2(\log 2 + \xi)}{n}} + \frac{\varepsilon}{n}\end{aligned}$$

If we omit the term $\inf_{P\in\overline{\mathscr{M}}}\left[\frac{1}{n}\sum_{i=1}^{n} d(P_i^\star, P)\right]$ 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 $\{\{p > q\}, \underline{p, q} \in \overline{\mathcal{M}}\}$ is VC with dimension not larger than $D \geqslant 1$, all the elements of $\overline{\mathcal{M}}$ are extremal with degree not larger than D and (10) becomes

$$\mathbb{E}\left[d\left(P^\star, \widehat{P}\right)\right] \leqslant 3 \inf_{P\in\overline{\mathscr{M}}} d(P^\star, P) + 48\sqrt{\frac{D}{n}} + \frac{\varepsilon}{n}. \tag{12}$$

In the particular case of i.i.d. data with common distribution $P^\star$, an inequality of the same flavour was established by Devroye and Lugosi (2001)[Section 8.2] for their minimum distance estimate. Both inequalities involve a constant 3 in front of the approximation term $\inf_{P\in\overline{\mathscr{M}}} d(P^\star, P)$. In our inequality 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.

In the remaining part of this paper, $(\mathscr{X}, \mathscr{A}) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ and μ is the *Lebesgue measure* on $\mathbb{R}$. Since $\mathbb{L}_1(\mathbb{R}, \mathscr{B}(\mathbb{R}), \mu)$ is separable, Assumption 1 is automatically satisfied. Each of the forthcoming sections is devoted to the problem of estimating densities or a mixture of densities with respect to μ under the assumption that they satisfy one of the following shape constraint: monotonicity, concavity, convexity or log-concavity.

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 $k \geqslant 1$. The elements of A provide a partition of $\mathbb{R}$ into $l = |A| + 1 \leqslant k + 1$ intervals $I_1, \ldots, I_l$ the endpoints of which belong to A. We denote by $\mathbf{I}(A)$ the set $\{\mathring{I}_1, \ldots, \mathring{I}_l\}$ 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 $k \geqslant 2$. 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

$$g: \left| \begin{array}{ccc} \mathbb{R} & \longrightarrow & \mathbb{R}_+ \\ x & \longmapsto & \frac{1}{\sqrt{|x|}} 1\!\!1_{|x|>0} \end{array} \right.$$

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. The sets $\overline{\mathcal{M}}_k$ are obviously increasing with k for the 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 $D \geqslant 1$ 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 bound on the number of bounded intervals in the class, 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 $D \geqslant 1$ and $k \geqslant 2$. The sets $\overline{\mathcal{O}}_{D,k}$ with $D \geqslant 1$ and $k \geqslant 2$ are therefore nonempty. They satisfy the following property which is a consequence of Proposition 3 of Baraud and Birgé (2016).

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

Proof. The proof is postponed to Subsection 7.3. □

For all $D \geqslant 1$ and $k \geqslant 2$, 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 $\bigcup_{D\geqslant 1} \mathcal{O}_{D,k}$. It follows from Proposition 1 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 $D \geqslant 1$. We immediately deduce from Theorem 1 the following result.

Theorem 2. *Let $k \geqslant 2$. Whatever the product distribution $\mathbf{P}^\star$ of the data, any TV-estimator $\widehat{P}$ on $\mathscr{M}_k$ satisfies*

$$\mathbb{E}\left[d\left(P^\star, \widehat{P}\right)\right] \leqslant \inf_{D\geqslant 1}\left[3 \inf_{P\in\overline{\mathscr{O}}_{D,k}} d(P^\star, P) + 83.2\sqrt{\frac{D+k+1}{n}}\right] + \frac{\varepsilon}{n}. \tag{13}$$

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 (2021) (see Proposition 6) for his ℓ-estimator. When the data are i.i.d. with distribution $P^\star$, our result shows better constants in the risk bound (13) as compared to Baraud's. In particular, the constant 3 in front of the approximation term $\inf_{P\in\overline{\mathscr{O}}_{D,k}} d(P^\star, P)$

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 $k \geqslant 2$, the problem of estimating a k-piecewise monotone density was also considered in Baraud and Birgé (2016). The authors used a ρ-estimator and the Hellinger loss in place of the total variation one to evaluate its risk— see their Corollary 2. Their bound is similar to ours except from 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 Hellinger loss while it disappears with the total variation one.

In order to specify further the risk bound given by (13), and see what properties can be established on our estimator, let us consider different assumptions on the distribution of the data. These assumptions can be done *a posteriori*, since the estimator is solely based on the model $\overline{\mathscr{M}}_k$ (or a dense subset of it). Such assumptions enable us to bound the quantity

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

that appear in the right-hand side of (13).

If the X_i are not i.i.d. but only independent and their marginals are close enough to a distribution of the form $\overline{P} = \overline{p} \cdot \mu$, by the triangle inequality we may bound A by $3d\big(P^\star, \overline{P}\big) + \mathbb{B}_{k,n}(\overline{p})$ where $\mathbb{B}_{k,n}(p)$ is defined for a density p on the line by the formulas

$$\begin{aligned} \mathbb{B}_{k,n}(p) &= \inf_{D \geqslant 1} \left[3 \inf_{Q \in \overline{\mathscr{O}}_{D,k}} d(P,Q) + 83.2\sqrt{\frac{D+k+1}{n}} \right] \\ &= \inf_{D \geqslant 1} \left[\frac{3}{2} \inf_{q \in \overline{\mathcal{O}}_{D,k}} \|p-q\| + 83.2\sqrt{\frac{D+k+1}{n}} \right]. \end{aligned} \tag{14}$$

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

If we apply a location-scale transformation to the data, that is, if in place of the original data $X_1, \dots, X_n$ with density $\overline{p}$ 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 density $\overline{p}_{m,\sigma}$ of the new data would satisfy $\mathbb{B}_{k,n}(\overline{p}_{m,\sigma}) = \mathbb{B}_{k,n}(\overline{p})$ since the set $\overline{\mathcal{O}}_{D,k}$ and the total variation distance remain invariant under such a transformation. This means that the performance of the TV-estimator is independent of the unit that is used to measure the data.

In the remaining part this section, we provide upper bounds on the quantity $\mathbb{B}_{k,n}(p)$ for some densities $p \in \overline{\mathcal{M}}_k$ of special interest.

4.2. **Estimation of bounded and compactly supported k-piecewise monotone densities.** For $k \geqslant 3$, 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

$$p = \sum_{i=1}^{k-2} w_i p_i \mathbb{1}_{(x_{i-1}, x_i)} \tag{15}$$

where

(i) $(x_i)_{i \in \{0,\ldots,k-2\}}$ is an increasing sequence of real numbers;
(ii) $w_1, \ldots, w_{k-2}$ are nonnegative numbers such that $\sum_{i=1}^{k-2} w_i = 1$;
(iii) for $i \in \{1, \ldots, k-2\}$, 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

$$\boldsymbol{R}_{k,0}(p) = \inf \left[\sum_{i=1}^{k-2} \sqrt{w_i \log(1 + L_i V_i)} \right]^2 \tag{16}$$

where the infimum runs among all ways of writing p under the form (15) 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 $\boldsymbol{R}_{k,0}(p)$. For $k \geqslant 3$ and $R > 0$, let $\overline{\mathcal{M}}_k^{\infty}(R)$ be the subset of $\overline{\mathcal{M}}_k^{\infty}$ that gathers these densities p for which $\boldsymbol{R}_{k,0}(p) < R$. When a density p belongs to $\overline{\mathcal{M}}_k^{\infty}(R)$, we may therefore write p under the form (15) 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 classes of sets $(\overline{\mathcal{M}}_k^{\infty})_{k \geqslant 3}$ and $(\overline{\mathcal{M}}_k^{\infty}(R))_{k \geqslant 3}$ are both increasing (for the inclusion): if $p \in \overline{\mathcal{M}}_l^{\infty}(R)$ with $l < k$, we may write

$$p = \sum_{i=1}^{l-2} w_i p_i \mathbb{1}_{(x_{i-1}, x_i)} \quad \text{a.e. with} \quad \left[\sum_{i=1}^{l-2} \sqrt{w_i \log(1 + L_i V_i)} \right]^2 < R$$

and alternatively

$$p = \sum_{i=1}^{l-2} w_i p_i \mathbb{1}_{(x_{i-1}, x_i)} + \sum_{j=l-1}^{k-2} 0 \times \mathbb{1}_{(x_i + j - l + 1), x_i + j - l + 2)} \text{ a.e.}$$

hence, $p \in \overline{\mathcal{M}}_k^{\infty}$ and

$$\boldsymbol{R}_{k,0}(p) \leqslant \left[\sum_{i=1}^{l-2} \sqrt{w_i \log\left(1 + L_i V_i\right)} + 0\right]^2 < R.$$

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

$$\mathbb{B}_{k,n}(p) \leqslant 41.3 \left(\frac{R}{n}\right)^{1/3} + 83.2\sqrt{\frac{2k}{n}}. \tag{17}$$

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 $(R/n)^{1/3} \vee kn^{-1/2}$ 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 (2016)[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 $(\sqrt{L\left\|p\right\|_{\infty}}/n)^{1/3}\log n$ up to some numerical constant. With our ℓ-estimator, the bound we get is of order $(\log(1 + L\left\|p\right\|_{\infty})/n)^{1/3}$ and only depends logarithmically on the quantity $L\left\|p\right\|_{\infty}$ (which is not smaller than 1 since p is a density).

Proof. The proof of Theorem 3 is based on (14) and the following approximation result. The complete proof is postponed to subsection 7.3. □

Proposition 2. *Let $V \geqslant 0$ and I be a bounded interval of length $L > 0$. For all $D \geqslant 1$, there exists a partition $\mathcal{J} = \mathcal{J}(D, L, V)$ of I into $D \geqslant 1$ nontrivial intervals with the following properties. For any monotone density p on I for which*

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

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

$$\overline{p} = \sum_{J \in \mathcal{J}} \overline{p}_J \mathbb{1}_J \quad \textit{with} \quad \overline{p}_J = \frac{1}{\mu(J)} \int_J p \, d\mu,$$

is a monotone density on I that satisfies

$$\int_I |p - \overline{p}| \, d\mu \leqslant \left[(1 + VL)^{1/D} - 1\right] \wedge 2 \leqslant \frac{2\log\left(1 + VL\right)}{D}. \tag{18}$$

Proof. The proof is postponed to subsection 7.3. □

Although the result is hidden in his calculations, Birgé (1987) has established a bound of the same flavour except from the fact that the variation V is replaced by a uniform bound on p. Unlike his, our bound (18) 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:

Corollary 1. *Let $k \geqslant 3$. If $X_1, \ldots, X_n$ is a n sample which density $p \in \overline{\mathcal{M}}_k^{\infty}(R)$, then the TV-estimator $\widehat{p}$ on $\mathcal{M}_k$ satisfies*

$$\mathbb{E}\left[\|p - \widehat{p}\|\right] \leqslant 82.6\left(\frac{R}{n}\right)^{1/3} + 166.4\sqrt{\frac{2k}{n}} + \frac{2\varepsilon}{n}. \tag{19}$$

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

$$\mathbb{B}_{k,n}(p) \leqslant \inf_{\overline{p} \in \overline{\mathcal{M}}_k^{\infty}} \left[\frac{3}{2}\|p - \overline{p}\| + \mathbb{B}_{k,n}(\overline{p})\right], \tag{20}$$

an upper bound on $\mathbb{B}_{k,n}(p)$ can be obtained by combining Theorem 3 with an approximationt 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 $\widetilde{p}$ as the mapping on $(0, +\infty)$ given by*

$$\widetilde{p}(y) = \inf\left\{x > 0,\ p(a + x) < y\right\} \geqslant 0. \tag{21}$$

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

$$\tau_x(p,\cdot) : \left|\begin{array}{ccc} [0, +\infty) & \longrightarrow & \mathbb{R}_+ \\ t & \longmapsto & \displaystyle\int_t^{+\infty} p(a + x)d\mu(x) \end{array}\right.,$$

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

$$\tau_y(p,\cdot) : \left|\begin{array}{ccc} [0, +\infty) & \longrightarrow & \mathbb{R}_+ \\ t & \longmapsto & \displaystyle\int_t^{+\infty} \widetilde{p}(y)d\mu(y) \end{array}\right.$$

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

$$\tau(p, t) = \inf_{s > 0}\left[\tau_x(p, st) + \tau_y(p, p(a + s))\right] \quad \text{for all } t \geqslant 1. \tag{22}$$

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 these definitions. When p is a continuous decreasing density from $(0,+\infty)$ onto $(0,+\infty)$, $\widetilde{p}$ is the reciprocal function p^{-1}. By taking the symmetric of the graph $x \mapsto p(x)$ with respect to the first diagonal, we easily see that $\widetilde{p} = p^{-1}$ 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 $\widetilde{p}$ while $\tau_x(p,\cdot)$ is that of $p(a+\cdot)$.

Lemma 2. *Let p be a nonincreasing density on $(a,+\infty)$ with $a \in \mathbb{R}$ and $\widetilde{p}$ the mapping defined by (21). For all $B \geqslant 0$,*

$$\int_0^{+\infty} [p(a+x)-B]_+ \, d\mu(x) = \int_B^{+\infty} \widetilde{p}(y) d\mu(y) = \tau_y(p,B). \tag{23}$$

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

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

Proof. Let $y > 0$. Since p is 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 $\widetilde{p}(y) < +\infty$ by definition of $\widetilde{p}(y)$. In particular,

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

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

$$\mathbb{1}_{x<\widetilde{p}(y)} \leqslant \mathbb{1}_{p(a+x)\geqslant y} \leqslant \mathbb{1}_{x\leqslant\widetilde{p}(y)}.$$

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

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

which proves (23). □

It follows from (23) that if p is a nonincreasing density on $(a,+\infty)$, $\tau(p,t)$ also writes for all $t \geqslant 1$ 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 $\ell\geqslant 2$. Changing the values of p, if ever necessary, on a negligible set, which will not change the way it can be approximated in $\mathbb{L}_1$-norm, we may assume with no loss of generality that it writes as

$$p=w_1p_1\mathbb{1}_{(-\infty,x_1)}+\sum_{i=2}^{\ell-1}w_ip_i\mathbb{1}_{(x_{i-1},x_i)}+w_\ell p_\ell\mathbb{1}_{(x_{\ell-1},+\infty)} \tag{24}$$

where p_1 and p_ℓ are monotone densities on $(-\infty,x_1)$ and $(x_{\ell-1},+\infty)$ respectively, $w_1=\int_{-\infty}^{x_1}pd\mu$, $w_\ell=\int_{x_\ell}^{+\infty}pd\mu$ and when $\ell>2$, $x_1<x_2<\ldots<x_{\ell-1}$ is an increasing sequence of real numbers, the p_i are monotone densities on (x_{i-1},x_i) and $w_i=\int_{x_{i-1}}^{x_i}pd\mu$ for all $i\in\{2,\ldots,\ell-1\}$. For p written under the form (24), we set

$$\tau_\infty(p,t)=\max_{i\in\{1,\ldots,\ell\}}\tau\left(p_i,t\right)\quad\text{for all } t\geqslant 1. \tag{25}$$

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

Theorem 4. *Let $\ell\geqslant 2$, $k\geqslant 2\ell$ and $R\geqslant\ell\log 2$. If p is a density of the form (24) a.e.,*

$$\inf_{\overline{p}\in\overline{\mathcal{M}}_k^\infty(R)}\|p-\overline{p}\|\leqslant 2\tau_\infty\left(p,\exp\left(\frac{R}{\ell}\right)-1\right). \tag{26}$$

Proof. The proof is postponed to Subsection 7.3. □

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

Corollary 2. *Let $\ell\geqslant 2$ and $k\geqslant 2\ell$. If p is a density of the form (24) a.e.,*

$$\mathbb{B}_{k,n}(p)\leqslant 44.3\left(\frac{\ell\log(1+r_n)}{n}\right)^{1/3}+83.2\sqrt{\frac{2k}{n}} \tag{27}$$

where

$$r_n=\inf\left\{t\geqslant 1,\ \tau_\infty(p,t)\leqslant\left(\frac{\ell\log(1+t)}{n}\right)^{1/3}\right\} \tag{28}$$

and $\tau_\infty(p,\cdot)$ is defined by (25). Then, the TV-estimator $\widehat{p}$ on $\mathcal{M}_k$ satisfies

$$\mathbb{E}\left[\|p-\widehat{p}\|\right]\leqslant 88.6\left(\frac{\ell\log(1+r_n)}{n}\right)^{1/3}+166.4\sqrt{\frac{2k}{n}}+\frac{2\varepsilon}{n}. \tag{29}$$

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

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

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

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

Using (20), Theorems 3 and 4 with $R = \ell \log(1+t) > \ell \log 2$ we obtain that

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

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

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

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

When $(\alpha, \beta) \in (0, +\infty) \times [-1, +\infty)$ and when $\alpha = 0$ and $\beta > 0$, q is 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 is under the form (24) with $\ell = 2$, $w_1 = 0$, $w_2 = 1$, $a = x_1 = 0$ 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 that when $\alpha > 0$, for all $t \geqslant 2$

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

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

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

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

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

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

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

Taking

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

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

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

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

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

for some constant $C' > 0$ large enough, we obtain that $\tau_{\infty}(p,t_n) \leqslant (2\log(1+t_n)/n)^{1/3}$ and consequently, r_n defined by (28) satisfies $r_n \leqslant t_n$. Applying Corollary 2, we conclude that for all $k \geqslant 4$

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

If $\alpha = 0$ and $\beta > 1$, 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

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

hence $r_n \leqslant t_n = C' n^{1/(1+3\beta)}$ and we get that for all $k \geqslant 4$

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

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 $k \geqslant 2$, 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 (nontrivial) 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 $D \geqslant 1$, 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 3. *Let $k \geqslant 2$, $D \geqslant 1$. 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 7.4. □

For all $D \geqslant 1$, $k \geqslant 2$, 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 $\bigcup_{D \geqslant 1} \mathcal{O}_{D,k}^1$. By proposition 1, 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 $D \geqslant 1$. We deduce from Theorem 1 the following result.

Theorem 5. *Let $k \geqslant 2$. Whatever the product distribution $\mathbf{P}^\star$ of the data, any TV-estimator $\widehat{P}$ on $\mathscr{M}_k^1$ satisfies*

$$\mathbb{E}\left[d\left(P^\star, \hat{P}\right)\right] \leqslant \inf_{D \geqslant 1} \left\{ 3 \inf_{P \in \overline{\mathscr{O}}_{D,k}^1} d(P^\star, P) + 68\sqrt{\frac{D + 2k - 1}{n}} \right\} + \frac{\varepsilon}{n}. \tag{30}$$

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 (30) writes as $\mathbb{B}_{k,n}^1(p) + \varepsilon/n$ with

$$\mathbb{B}_{k,n}^1(p) = \inf_{D \geqslant 1} \left[\frac{3}{2} \inf_{q \in \overline{\mathcal{O}}_{D,k}^1} \|p - q\| + 68\sqrt{\frac{D + 2k - 1}{n}} \right]. \tag{31}$$

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 $D \geqslant 1$ 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 $a = x_0 < \ldots < x_D = b$ such that $\overline{f}(x_i) = f(x_i)$ and $\overline{f}$ is affine on $[x_{i-1}, x_i]$ for all $i \in \{1, \ldots, D\}$.*

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 $i \in \{1, \ldots, D\}$. 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 $\overline{f}$. If f is convex (respectively concave), $\overline{f} \geqslant f$ (respectively $\overline{f} \leqslant f$).

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

$$|p_r'(a)| \wedge |p_l'(b)| \leqslant |\Delta| \leqslant |p_r'(a)| \vee |p_l'(b)| \quad \text{and}$$

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 $D \geqslant 1$, there exists a $2D$-linear interpolation $\overline{p}$ of p such that*

$$\int_a^b |p - \overline{p}|\, d\mu \leqslant \frac{4}{3}\left[\left(1 + \sqrt{2LV\Gamma}\right)^{1/D} - 1\right]^2. \tag{32}$$

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

$$\int_a^b |p - q|\, d\mu \leqslant 5.14 \frac{\log^2\left(1 + \sqrt{2LV\Gamma}\right)}{D^2}. \tag{33}$$

Since $\Gamma \in [0,1]$, (33) implies that

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

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

The fact that a bounded convex (or concave) function on a compact interval can be approximated in $\mathbb{L}_1$ by piecewise affine functions at rate $\mathcal{O}(1/D^2)$ had already been established by Guérin et al. (2006). What is novel in Theorem 6 is 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 7.4. □

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 $k \geqslant 3$, that is of the form

$$p = \sum_{i=1}^{k-2} w_i p_i \mathbb{1}_{(x_{i-1}, x_i)} \tag{34}$$

where

(i) $(x_i)_{i \in \{0,\ldots,k-2\}}$ is an increasing sequence of real numbers;
(ii) $w_1, \ldots, w_{k-2}$ are nonnegative numbers such that $\sum_{i=1}^{k-2} w_i = 1$;
(iii) for $i \in \{1, \ldots, k-2\}$, 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 $\Gamma_i \in [0,1]$.

For a density p satisfying (34) we define

$$\boldsymbol{R}_{k,1}(p) = \inf \left[\sum_{i=1}^{k-2} \left(w_i \log^2 \left(1 + \sqrt{2L_i V_i \Gamma_i} \right) \right)^{1/3} \right]^{3/2}, \tag{35}$$

where the infimum runs among all ways of writing p under the form (34). We denote by $\overline{\mathcal{M}}_{k,1}^{\infty}$ the class of all densities of the form (34) a.e. and for $R > 0$, $\overline{\mathcal{M}}_{k,1}^{\infty}(R)$ the subset of those which satisfy $\boldsymbol{R}_{k,1}(p) < R$.

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

$$\mathbb{B}_{k,n}^1(p) \leqslant 7.71 \left[15.06 \left(\frac{R}{n} \right)^{2/5} + 8.82 \sqrt{\frac{4k-5}{n}} \right], \tag{36}$$

where $\mathbb{B}_{k,n}^1(p)$ is defined by (31).

Theorem 7 implies, together with Theorem 5, that the TV-estimator converges at rate $n^{-2/5}$ in total variation distance, whenever the underlying density $p^\star$ 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 (2001)[Section 15.5] for bounded convex densities. This rate is therefore optimal.

To the best of our knowledge, Theorem 8 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 (2009b) 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 is 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 $L|p(a) - p(b)|/2 \leqslant 1$, hence $\boldsymbol{R}_{3,1}(p) \leqslant \log(1 + \sqrt{2L|p(a) - p(b)|}) \leqslant \log 3$. 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 $[c, a]$ 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 $z^{1/3} + (1-z)^{1/3} \leqslant 2^{2/3}$ which holds for all $z \in [0, 1]$, we obtain that

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

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

Corollary 3. *If $X_1,\ldots,X_n$ is a n-sample which density is concave on an interval of $\mathbb{R}$, then the TV-estimator $\widehat{p}$ on $\mathcal{M}_4^1$ satisfies*

$$\mathbb{E}\left[\|p-\widehat{p}\|\right] \leqslant \frac{320}{n^{2/5}} + \frac{451}{\sqrt{n}} + \frac{2\varepsilon}{n}.$$

In particular,

$$\inf_{\widetilde{p}} \sup_{p} \mathbb{E}\left[\|p-\widetilde{p}\|\right] \leqslant \frac{320}{n^{2/5}} + \frac{451}{\sqrt{n}},$$

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

Proof of Theorem 7. Let $p \in \overline{\mathcal{M}}_{k,1}^{\infty}(R)$. With no loss of generality we may assume that p is of the form (34) everywhere and choose a subdivision $(x_i)_{i\in\{0,\ldots,k-2\}}$ in such a way that

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

Let $D \geqslant k-2$ and $D_1,\ldots,D_{k-2}$ be some positive integers to be chosen later on that satisfy the constraint $\sum_{i=1}^{k-2} D_i \leqslant D$. By Theorem 6, we may find for all $i \in \{1,\ldots,k-2\}$ 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

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

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 $I \in \mathbf{I}(\{x_0,\ldots,x_{k-2}\})$ and affine on each interval of a partition $(x_0,x_{k-2}] = \bigcup_{i=1}^{k-2}(x_{i-1},x_i]$ into $\sum_{i=1}^{k-2} 2D_i \leqslant 2D$ intervals. It therefore belongs to $\overline{\mathcal{O}}_{2D,k}$. Besides,

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

and it follows from (31) that

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

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

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

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

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

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

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

which gives (36). □

6. Log-concave densities

In this section, we consider the set $\overline{\mathcal{M}}^{\mathrm{LC}}$ of log-concave densities. With the convention $\exp(-\infty)=0$, $\overline{\mathcal{M}}^{\mathrm{LC}}$ is the set of densities of the form $p=\exp\phi$ where the set $J=J(p)=\{x\in\mathbb{R},\ p(x)>0\}$ is an open interval and the mapping $\phi:\mathbb{R}\to\mathbb{R}\cup\{-\infty\}$ is a continuous and concave function on J. Given some $D\geqslant 1$, a subset of $\overline{\mathcal{M}}^{\mathrm{LC}}$ of special interest is the set $\overline{\mathcal{O}}^{\mathrm{LC}}_D$ of those densities the logarithm of which is either affine or takes the value $-\infty$ on the elements of a set $\mathbf{I}(A)$ with $A\in\mathcal{A}(D)$. Since a density $p\in\overline{\mathcal{O}}^{\mathrm{LC}}_D$ is log-concave, the logarithm of p may take the value $-\infty$ on the two unbounded elements of $\mathbf{I}(A)$ only. For example, the densities $x\mapsto \mathbb{1}_{x>0}\exp(-x)$ and $x\mapsto(1/2)\exp(-|x|)$ of the exponential and Laplace distributions belong to $\overline{\mathcal{O}}^{\mathrm{LC}}_1$ while the standard Gaussian density belongs to $\overline{\mathcal{M}}^{\mathrm{LC}}$.

Proposition 4. *Let $D \geqslant 1$. The elements of $\overline{\mathcal{O}}_D^{LC}$ are extremal in $\overline{\mathcal{M}}^{LC}$ with degree not larger than $2(D+2)$.*

Proof. The proof is postponed to Subsection 7.5. □

For all $D \geqslant 1$, let $\mathcal{O}_D^{\mathrm{LC}}$ be a countable and dense subset of $\overline{\mathcal{O}}_D^{\mathrm{LC}}$ (for the $\mathbb{L}^1$-norm) and $\mathcal{M}^{\mathrm{LC}}$ a countable and dense subset of $\overline{\mathcal{M}}^{\mathrm{LC}}$ that contains $\bigcup_{D\geqslant 1}\mathcal{O}_D^{\mathrm{LC}}$. We immediately deduce from Theorem 1 together with Proposition 4 the following result.

Theorem 8. *Whatever the product distribution $\mathbf{P}^\star$ of the data, any TV-estimator $\widehat{P}$ on $\mathscr{M}^{LC}$ satisfies*

$$\mathbb{E}\left[d\left(P^\star, \widehat{P}\right)\right] \leqslant \inf_{D\geqslant 1}\left[3\inf_{P\in\overline{\mathscr{O}}_D^{LC}} d(P^\star, P) + 48\sqrt{2}\sqrt{\frac{D+2}{n}}\right] + \frac{\varepsilon}{n}. \tag{37}$$

In the remaining part of this section we assume that the X_i are i.i.d. with a density $p \in \overline{\mathcal{M}}_{\mathrm{LC}}$, in which case, the right-hand side of (37) writes as $\mathbb{B}_n(p) + \varepsilon/n$ with

$$\mathbb{B}_n(p) = \inf_{D\geqslant 1}\left[\frac{3}{2}\inf_{q\in\overline{\mathcal{O}}_D^{\mathrm{LC}}}\|p-q\| + 48\sqrt{2}\sqrt{\frac{D+2}{n}}\right] \tag{38}$$

In particular, if p belongs to $\overline{\mathcal{O}}_D^{\mathrm{LC}}$ for some $D \geqslant 1$, then $\mathbb{B}_n(p) \leqslant \sqrt{D/n}$, so that $\hat{p}$ converges at the parametric rate. Thus, the TV-estimator shares the *adaptivity* property established for the MLE by Kim et al (2018) (and Feng et al. (2021)), and by Baraud and Birgé (2016) for their $\rho-$estimator. As compared to theirs, our upper bound does not involve logarithmic factors. This difference is due to the fact that Baraud and Birgé's results are established for the Hellinger distance and these logarithmic factors are sometimes necessary for such a loss.

As we did in Sections 4 and 5 , our aim is now to bound the quantity $\mathbb{B}_n(p)$. This can be done by using the following approximation result.

Proposition 5. *Let $D \geqslant 6$. For all log-concave densities p,*

$$\inf_{q\in\overline{\mathcal{O}}_{6D}^{LC}}\|p-q\| \leqslant \frac{2}{D^2}.$$

Proof. The proof is postponed to Subsection 7.5. □

Combining (38) with Proposition 5, we deduce that for all $D \geqslant 6$

$$\begin{aligned}\mathbb{B}_n(p) &= \inf_{D\geqslant 6}\left[\frac{3}{2}\inf_{q\in\overline{\mathcal{O}}_{6D}^{\mathrm{LC}}}\|p-q\| + 48\sqrt{2}\sqrt{\frac{6D+2}{n}}\right]\\ &\leqslant \inf_{D\geqslant 6}\left[\frac{3}{D^2} + 48\sqrt{2}\sqrt{\frac{6D+2}{n}}\right].\end{aligned}$$

By choosing

$$D = \left\lceil \left(\frac{3n}{96^2}\right)^{1/5} \right\rceil \vee 6 \leqslant \left(\frac{6n}{2 \times 96^2}\right)^{1/5} + 6,$$

we obtain the following estimation result for log-concave densities.

Corollary 4. *If $X_1, \ldots, X_n$ is a n-sample which density p is log-concave on $\mathbb{R}$, then the TV-estimator $\widehat{p}$ on $\mathcal{M}^{LC}$ satisfies*

$$\mathbb{E}\left[\|p - \widehat{p}\|\right] \leqslant \frac{300}{n^{2/5}} + \frac{837}{\sqrt{n}} + \frac{2\varepsilon}{n}.$$

The shape-constrained TV-estimator attains the global convergence rate of $n^{-2/5}$ on the class of log-concave densities, i.e. the same global rate as the MLE — see Kim and Samworth (2016).

Acknowledgement. One of the authors thanks Lutz Dümbgen for asking if we could establish a result with $d(P^\star, \widehat{P})$ in place of $n^{-1}\sum_{i=1}^n d(P_i, \widehat{P})$ 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.

7. Proofs

7.1. **Technical result.** First, we need a way to approximate p not just by nonnegative functions, but by densities. The following lemma shows that this can be achieved with a simple renormalization.

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

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

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

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

comes from the triangle inequality. Let us now prove that

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

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

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

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

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

□

7.2. **Proofs of Section 3.**

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

$$\mathbf{Z}_+(\boldsymbol{X},P) = \sup_{Q\in\mathscr{M}} \left[\mathbf{T}(\boldsymbol{X},P,Q)-\mathbb{E}\left[\mathbf{T}(\boldsymbol{X},P,Q)\right]\right]-\zeta$$
$$\mathbf{Z}_-(\boldsymbol{X},P) = \sup_{Q\in\mathscr{M}} \left[\mathbb{E}\left[\mathbf{T}(\boldsymbol{X},Q,P)\right]-\mathbf{T}(\boldsymbol{X},Q,P)\right]-\zeta$$

and

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

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

$$\begin{aligned} nd\big(\overline{P},Q\big) &\leqslant nd\big(P^\star,\overline{P}\big)+\mathbb{E}\left[\mathbf{T}(\boldsymbol{X},Q,\overline{P})\right] \\ &= nd\big(P^\star,\overline{P}\big)+\mathbb{E}\left[\mathbf{T}(\boldsymbol{X},Q,\overline{P})\right]-\mathbf{T}(\boldsymbol{X},Q,\overline{P})+\mathbf{T}(\boldsymbol{X},Q,\overline{P}) \\ &\leqslant nd\big(P^\star,\overline{P}\big)+\mathbf{Z}(\boldsymbol{X},\overline{P})+\mathbf{T}(\boldsymbol{X},Q,\overline{P})+\zeta \\ &\leqslant nd\big(P^\star,\overline{P}\big)+\mathbf{Z}(\boldsymbol{X},\overline{P})+\mathbf{T}(\boldsymbol{X},Q)+\zeta. \end{aligned}$$

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

$$\mathbf{T}(\boldsymbol{X},\widehat{P})\leqslant \inf_{P\in\mathscr{M}}\mathbf{T}(\boldsymbol{X},P)+\varepsilon\leqslant\mathbf{T}(\boldsymbol{X},\overline{P})+\varepsilon,$$

we deduce that

$$nd\big(\overline{P},\widehat{P}\big)\leqslant nd\big(P^\star,\overline{P}\big)+\mathbf{Z}(\boldsymbol{X},\overline{P})+\mathbf{T}(\boldsymbol{X},\overline{P})+\zeta+\varepsilon. \tag{39}$$

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

$$\begin{aligned} \mathbf{T}(\boldsymbol{X},\overline{P}) &= \sup_{Q\in\mathscr{M}}\mathbf{T}(\boldsymbol{X},\overline{P},Q) \\ &\leqslant \sup_{Q\in\mathscr{M}}\left[\mathbf{T}(\boldsymbol{X},\overline{P},Q)-\mathbb{E}\left[\mathbf{T}(\boldsymbol{X},\overline{P},Q)\right]-\zeta\right]+\sup_{Q\in\mathscr{M}}\mathbb{E}\left[\mathbf{T}(\boldsymbol{X},\overline{P},Q)\right]+\zeta \\ &\leqslant \mathbf{Z}(\boldsymbol{X},\overline{P})+nd\big(P^\star,\overline{P}\big)+\zeta, \end{aligned}$$

which with (39) lead to

$$nd\left(\overline{P},\widehat{P}\right)\leqslant 2nd\left(P^{\star},\overline{P}\right)+2\zeta+\varepsilon+2\mathbf{Z}(\boldsymbol{X},\overline{P}). \tag{40}$$

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

$$\begin{aligned}\mathbf{w}(P)=\mathbb{E}&\left[\sup_{Q\in\mathscr{M}}\left[\mathbf{T}(\boldsymbol{X},P,Q)-\mathbb{E}\left[\mathbf{T}(\boldsymbol{X},P,Q)\right]\right]\right]\\&\vee\mathbb{E}\left[\sup_{Q\in\mathscr{M}}\left[\mathbb{E}\left[\mathbf{T}(\boldsymbol{X},Q,P)\right]-\mathbf{T}(\boldsymbol{X},Q,P)\right]\right].\end{aligned}$$

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

$$\left|\mathbf{Z}_{+}((x_1,\ldots,x_i,\ldots,x_n),\overline{P})-\mathbf{Z}_{+}((x_1,\ldots,x_i',\ldots,x_n),\overline{P})\right|\leqslant 1$$

for all $\mathbf{x}\in\mathbf{E}$, $x_i'\in E$ and $i\in\{1,\ldots,n\}$. Following the same lines as in the proof of Lemma 2 of Baraud (2021) (with $\xi+\log 2$ in place of ξ), we deduce that with a probability at least $1-(1/2)e^{-\xi}$,

(41)

$$\begin{aligned}\mathbf{Z}_{+}(\boldsymbol{X},\overline{P})&\leqslant\mathbb{E}\left[\mathbf{Z}_{+}(\boldsymbol{X},\overline{P})\right]+\sqrt{\frac{n(\xi+\log 2)}{2}}\\&=\mathbb{E}\left[\sup_{Q\in\mathscr{M}}\left[\mathbf{T}(\boldsymbol{X},\overline{P},Q)-\mathbb{E}\left[\mathbf{T}(\boldsymbol{X},\overline{P},Q)\right]\right]\right]+\sqrt{\frac{n(\xi+\log 2)}{2}}-\zeta\\&\leqslant\mathbf{w}(\overline{P})+\sqrt{\frac{n(\xi+\log 2)}{2}}-\zeta. \qquad (42)\end{aligned}$$

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

$$\mathbf{Z}_{-}(\boldsymbol{X},\overline{P})\leqslant\mathbf{w}(\overline{P})+\sqrt{\frac{n(\xi+\log 2)}{2}}-\zeta. \tag{43}$$

Putting (42) and (43) together and choosing $\zeta=\mathbf{w}(\overline{P})+\sqrt{n(\xi+\log 2)/2}$, we obtain that with a probability at least $1-e^{-\xi}$,

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

which with (40) lead to the bound

$$d\left(\overline{P},\widehat{P}\right)\leqslant 2d\left(P^{\star},\overline{P}\right)+\frac{2\mathbf{w}(\overline{P})}{n}+\sqrt{\frac{2(\xi+\log 2)}{n}}+\frac{\varepsilon}{n}. \tag{44}$$

It remains now to control $\mathbf{w}(\overline{P})$. Since $\overline{p}\in\mathcal{O}(D)\subset\overline{\mathcal{O}}(D)$, it is extremal in $\overline{\mathcal{M}}\supset\mathcal{M}$ with degree not larger than D, the classes $\{\{q<\overline{p}\},\ q\in\mathcal{M}\setminus\{\overline{p}\}\}$ and $\{\{q>\overline{p}\},\ q\in\mathcal{M}\setminus\{\overline{p}\}\}$ are both VC with dimensions not larger than

D. We may therefore apply Proposition 3.1 in Baraud (2016) with $\sigma = 1$ and get

$$\mathbb{E}\left[\sup_{q\in\mathcal{M}\setminus\{\overline{p}\}}\left|\sum_{i=1}^{n}\left(\mathbb{1}_{\overline{p}>q}(X_i) - P_i^\star(\overline{p}>q)\right)\right|\right] \leqslant 10\sqrt{5nD}, \tag{45}$$

and

$$\mathbb{E}\left[\sup_{q\in\mathcal{M}\setminus\{\overline{p}\}}\left|\sum_{i=1}^{n}\left(\mathbb{1}_{\overline{p}<q}(X_i) - P_i^\star(\overline{p}<q)\right)\right|\right] \leqslant 10\sqrt{5nD}. \tag{46}$$

These inequalities entail $\mathbf{w}(\overline{P}) \leqslant 10\sqrt{5nD}$, and we infer from (44) that

$$d\left(\overline{P},\widehat{P}\right) \leqslant 2d\left(P^\star,\overline{P}\right) + 20\sqrt{\frac{5D}{n}} + \sqrt{\frac{2(\xi+\log 2)}{n}} + \frac{\varepsilon}{n}. \tag{47}$$

Since $\overline{P}$ is arbitrary in the set $\mathscr{O}(D)$ which is dense on $\overline{\mathscr{O}}(D)$, we infer that equation (47) holds for all $\overline{P} \in \overline{\mathscr{O}}(D)$, which yields (7). Hence, by the triangle inequality,

$$\begin{aligned} d\left(P^\star,\widehat{P}\right) &\leqslant \inf_{P\in\overline{\mathscr{O}}(D)}\left\{d(P^\star,P) + d\left(P,\widehat{P}\right)\right\} \\ &\leqslant 3\inf_{P\in\overline{\mathscr{O}}(D)} d(P^\star,P) + 20\sqrt{\frac{5D}{n}} + \sqrt{\frac{2(\xi+\log 2)}{n}} + \frac{\varepsilon}{n}. \end{aligned}$$

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

7.3. **Proofs of Section 4.**

Proof of Proposition 2. 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 $\mathring{I} = (a,b)$ 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 $V_I(p) = q(a) - q(b) = V_I(q)$. With no loss of generality, we may therefore assume that $I = [a,b]$ and that $V_I(p) = p(a) - p(b)$, what we shall do now.

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

$$\int_J |p - \overline{p}_J|\, d\mu \leqslant \frac{(v-u)(p(u)-p(v))}{2}. \tag{48}$$

In particular, when $D = 1$ it suffices to take $\mathcal{J} = \{I\}$ and the result follows from (48) with $u = a$ and $v = b$ and the trivial inequality $\int_I |p - \overline{p}_J| d\mu \leqslant 2$. It remains to prove the result for $D \geqslant 2$ and since (18) 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 $a = x_0 < x_1 < \ldots < x_D = b = a + L$ and a partition $\mathcal{J}$ of I into D intervals based on $\{x_0, \ldots, x_D\}$. Using (48) and the facts that $x_{j+1} - x_j = (1+\eta)(x_j - x_{j-1}) > x_j - x_{j-1}$ for $j \in \{1, \ldots, D-1\}$, we obtain

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

Together with the trivial bound $\int_I |p - \overline{p}|\, d\mu \leqslant 2$, this last inequality leads to (18). The second inequality derives from the fact that $(e^x - 1) \wedge 2 \leqslant 2x/\log 3 \leqslant 2x$ for all $x \geqslant 0$. □

Proof of Theorem 3. Let $D, D_1, \ldots, D_{k-2}$ be positive integers and p a density in $\overline{\mathcal{M}}_k^{\infty}(R)$. We may therefore write p under the form (15) with

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

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

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

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 $i \in \{1, \ldots, k-2\}$. Since

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

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

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

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

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

Using the sub-additivity property of the square root, we deduce from (14) that

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

which is (17). □

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 4. *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 $p_{|I}^{\wedge B} = (p \wedge B)\mathbb{1}_I / \int_I (p\wedge B) d\mu$ satisfies,*

$$\left\| p - p_{|I}^{\wedge B} \right\| \leqslant 2\left[\int_I (p-B)_+ \, d\mu + \int_{I^c} p d\mu\right]. \tag{49}$$

If p is a monotone density on a half-line

$$\inf_{\overline{p} \in \overline{\mathcal{M}}_3^{\infty}(R)} \|p - \overline{p}\| \leqslant 2\tau\left(p, \exp(R) - 1\right) \quad \text{for all } R \geqslant \log 2. \tag{50}$$

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 $p_{|I}^{\wedge B}$ is therefore a well-defined density on I. By Lemma 3,

$$\begin{aligned}\int_{\mathbb{R}} \left| p - p_{|I}^{\wedge B} \right| d\mu &\leqslant 2\int_{\mathbb{R}} \left| p - (p\wedge B)\mathbb{1}_I \right| d\mu \\ &= 2\int_{I^c} p d\mu + 2\int_I (p-B)_+ \, d\mu.\end{aligned}$$

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 with no 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

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

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 $(a,+\infty)$, $p(x)\geqslant p(a+s)=B$ for all $x\in(a,a+s)\subset I$ and consequently

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

The density $\overline{p}_s=p_{|I}^{\wedge B}$ 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 that

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

Hence, it follows from (16) that

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

and consequently, $\overline{p}_s\in\overline{\mathcal{M}}_3^\infty(R)$. Applying (49) and Lemma 2, we obtain that for all $s>0$

$$\inf_{\overline{p}\in\overline{\mathcal{M}}_3^\infty(R)}\|p-\overline{p}\|\leqslant\|p-\overline{p}_s\|\leqslant 2\left[\tau_x(p,st)+\tau_y\left(p,p(a+s)\right)\right]\tag{51}$$

and we derive (50) from (22). Since for all $s>0$, $\overline{p}_s$ 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 $s_0=l/t\leqslant l$. 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

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

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

$$\tau_y\left(p,p(a+s)\right)\leqslant\tau_y\left(p,p(a+s_0)\right)\leqslant\tau_y\left(p,p(a+s')\right),$$

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

$$\begin{aligned}&\inf_{s\in(0,s_0)}\left[\tau_x(p,st)+\tau_y\left(p,p(a+s)\right)\right]\\&\leqslant\inf_{s\in(0,s_0)}\tau_x(p,st)+\tau_y\left(p,p(a+s')\right)=\tau_x(p,s_0t)+\tau_y\left(p,p(a+s')\right)\\&=0+\tau_y\left(p,p(a+s')\right)=\tau_x(p,s't)+\tau_y\left(p,p(a+s')\right).\end{aligned}$$

Consequently, (51) 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 $\overline{p}_s=p_{|I}^{\wedge B}$ vanishes outside $(a,a+s)\subset(a,a+l)$, we may restrict the infimum in (50) to those densities in $\overline{\mathcal{M}}_3^{\infty}(R)$ that vanish outside $(a, a+l)$. □

Let us set $\eta = \tau_\infty(p, \exp(R/\ell) - 1)$. Since $R/\ell \geqslant \log 2$, by applying Lemma 4 to the densities p_i, we may find for all $i \in \{1, \dots, \ell\}$ a density $\overline{p}_i \in \overline{\mathcal{M}}_3^{\infty}(R/\ell)$ such that $\|p_i - \overline{p}_i\| \leqslant 2\tau(p_i, \exp(R/\ell) - 1)$. In particular, the density $\overline{p} = \sum_{i=1}^{\ell} w_i \overline{p}_i$ satisfies

$$\|p - \overline{p}\| \leqslant \sum_{i=1}^{\ell} w_i \|p_i - \overline{p}_i\| \leqslant \max_{i \in \{1,\dots,\ell\}} \|p_i - \overline{p}_i\| \leqslant 2\eta. \tag{52}$$

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 4 that we may choose $\overline{p}_i$ 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

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

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 $\overline{p}_1$ 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

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

Similarly, $\overline{p}_\ell$ 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

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

The density $\overline{p} = \sum_{i=1}^{\ell} w_i \overline{p}_i$ also writes as

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

and may therefore be written under the form (15) when $k \geqslant 2\ell$. Moreover, by Cauchy-Schwarz inequality

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

and consequently, $\overline{p} \in \overline{\mathcal{M}}_k^{\infty}(R)$. We deduce from (52) that

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

which is (26). □

7.4. **Proofs of Section 5.**

Proof of Proposition 3. The proof relies on the following lemma the proof of which is a direct consequence of convexity and is therefore omitted.

Lemma 5. *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: $\varnothing, J,\ J \cap (-\infty, c_0)$, $J \cap (c_1, +\infty)$, $[J \cap (-\infty, c_0)] \cup [J \cap (c_1, +\infty)]$ 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 $A = \{a_1, \ldots, a_l\}$ with $l \in \{1, \ldots, k-1\}$ 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 $l \geqslant 2$, $J_i = (a_{i-1}, a_i]$ for all $i \in \{2, \ldots, l\}$. Besides, we set $I_i = \mathring{J}_i$ for all $i \in \{1, \ldots, l+1\}$. We shall repeatedly use Lemma 5 with $g = p - q$ throughout the proof. Given $\epsilon \in \{\pm 1\}$, we set

$$C_\epsilon = \{x \in \mathbb{R},\ \epsilon g(x) > 0\} = \bigcup_{i=1}^{l+1} \{x \in J_i,\ \epsilon g(x) > 0\}\,. \tag{53}$$

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 (2016).

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, $\{x \in J_1,\ g(x) < 0\} = \varnothing$ and $\{x \in J_1,\ g(x) > 0\}$ is an interval.

If $m_1 = |B \cap I_1| \geqslant 1$, we may partition J_1 into $s = m_1 + 1 \geqslant 2$ consecutive intervals $K_1, \ldots, K_s$ 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 $K_1, \ldots, K_{s-1}$ and on $\mathring{K}_s$. On K_1, $\{x \in K_1,\ g(x) < 0\} = \varnothing$ 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 $s \geqslant 2$, we may apply Lemma 5 to g and the intervals K_i with $i \in \{2, \ldots, s-1\}$. 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 $\{x \in K_s,\ g(x) < 0\}$ writes 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 $l \geqslant 2$, let us now consider an interval of the form $J_i = (a_{i-1}, a_i]$ with $i \in \{2, \ldots, l\}$ 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 whatever $\epsilon \in \{\pm 1\}$. If $m_i \geqslant 1$, 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 5 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 (53) and the facts that $\sum_{i=1}^{l+1} m_i \leqslant |B| \leqslant D + 1$ and $l \leqslant k - 1$, we conclude that the sets C_ε are unions of at most

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

intervals. □

Proof Theorem 6. The proof is based on Lemma 6 and three preliminary approximation results given in Proposition 6, Proposition 7 and Proposition 8, whose proofs follow that one.

Proposition 6. *Let f be a convex-concave continuous function on a bounded nontrivial interval $[a, b]$ and ℓ_f be the linear function*

$$\ell_f : x \mapsto f(a) + \frac{f(b) - f(a)}{b - a}(x - a). \tag{54}$$

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,

$$\int_a^b |f - \ell_f|\, d\mu \leqslant \frac{(b-a)^2}{8} \left|f'_r(a) - f'_l(b)\right|. \tag{55}$$

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

$$\int_a^b |f - \ell_f|\, d\mu \leqslant \frac{(f(b)-f(a))^2}{8} \left| \frac{1}{f_r'(a)} - \frac{1}{f_l'(b)} \right|, \tag{56}$$

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

Since ℓ_f is the equation of the chord connecting $(a, f(a))$ to $(b, f(b))$, it is clear that $\ell_f(x) = f(x)$ for all $x \in \{a, b\}$, $\ell_f \geqslant f$ on $[a,b]$ when f is convex on $[a,b]$ and $\ell_f \leqslant f$ when f is concave.

Proposition 7. *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 $D \geqslant 1$, there exists a D-linear interpolation $\overline{p}$ of p such that*

$$\int_a^b |p - \overline{p}|\, d\mu \leqslant \frac{4}{3} \left[\left(1 + L\sqrt{|p_l'(b) - p_r'(a)|}\right)^{1/D} - 1 \right]^2. \tag{57}$$

Proposition 8. *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 $D \geqslant 1$ there exists a D-linear interpolation $\overline{p}$ of p on $[a,b]$ such that*

$$\int_a^b |p - \overline{p}|\, d\mu \leqslant \frac{4}{3} \left[\left(1 + V\sqrt{\left| \frac{1}{p_r'(a)} - \frac{1}{p_l'(b)} \right|}\right)^{1/D} - 1 \right]^2. \tag{58}$$

Lemma 6. *If F is a nondecreasing, differentiable, concave function on $\mathbb{R}_+$, the mapping*

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

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

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

is concave on $\mathbb{R}_+$.

Proof. For all $u > 0$,

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

is the product of $u \mapsto [F\left(\sqrt{u}\right) - F(0)]/\sqrt{u}$ 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. □

Let's turn to the proof of Theorem 6. We first introduce some mappings of interest we will use in the proofs of of Theorem 6 and Proposition 7. Let $\mathcal{V}$ be the linear space of continuous functions f on $[a,b]$ that admits 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

$$\mathcal{T}_1 : f \mapsto [x \mapsto f(a+b-x)], \quad \mathcal{T}_2 : f \mapsto [x \mapsto m - f(x)]. \tag{59}$$

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 $f \in \mathcal{V}$

$$|f(a) - f(b)| = |\mathcal{T}_j(f)(a) - \mathcal{T}_j(f)(b)| \tag{60}$$

$$\left|f'_r(a) - f'_l(b)\right| = \left|(\mathcal{T}_j(f))'_r(a) - (\mathcal{T}_j(f))'_l(b)\right| \tag{61}$$

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 $p'_r(a) \geqslant 0$. If p is constant, the result is clear by taking $\overline{p} = p$. 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

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

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 7 to p_1/w_1 and find a D-linear interpolation $\overline{p}_1$ of p on $[a,c]$ that satisfies

$$\int_a^c |p - \overline{p}_1|\, d\mu \leqslant \frac{4w_1}{3}\left[\left(1 + (c-a)\sqrt{\frac{p'_l(c) - p'_r(a)}{w_1}}\right)^{1/D} - 1\right]^2. \tag{62}$$

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 8 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 $\overline{p}_2$ of p on $[c,b]$ that satisfies

$$\int_c^b |p - \overline{p}_2|\, d\mu \leqslant \frac{4w_2}{3}\left[\left(1 + \frac{p(b) - p(c)}{\sqrt{w_2}}\sqrt{\frac{1}{p'_r(c)} - \frac{1}{p'_l(b)}}\right)^{1/D} - 1\right]^2. \tag{63}$$

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

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

Then

$$
\begin{aligned}
A &= (c-a)^2\left(p_l'(c)-p_r'(a)\right)+(p(b)-p(c))^2\left(\frac{1}{p_r'(c)}-\frac{1}{p_l'(b)}\right)\\
&\leqslant (b-a)^2\left(\Delta-p_r'(a)\right)+(p(b)-p(a))^2\left(\frac{1}{\Delta}-\frac{1}{p_l'(b)}\right)\\
&= 2(b-a)(p(b)-p(a))\left[1-\frac{1}{2}\left(\frac{p_r'(a)}{\Delta}+\frac{\Delta}{p_l'(b)}\right)\right]=2LV\Gamma. \qquad (64)
\end{aligned}
$$

The function $\overline{p}=\overline{p}_1 1\!\!1_{[a,c)}+\overline{p}_2 1\!\!1_{[c,b]}$ is a $2D$-linear interpolation of p on $[a,b]$. Since by Lemma 6 the function F_D is concave and increasing, we deduce from (64), (62) and (63) that

$$
\begin{aligned}
\frac{3}{4}\left\|p-\overline{p}\right\| &\leqslant w_1\int_a^c\left|p-\overline{p}_1\right|d\mu+w_2\int_c^b\left|p-\overline{p}_2\right|d\mu\\
&\leqslant w_1F_D\left(\frac{(c-a)^2\left(p_l'(c)-p_r(a)\right)}{w_1}\right)\\
&\quad+w_2F_D\left(\frac{(p(b)-p(c))^2}{w_2}\left(\frac{1}{p_r'(c)}-\frac{1}{p_l'(b)}\right)\right)\\
&\leqslant F_D(A)\leqslant F_D(2LV\Gamma),
\end{aligned}
$$

which is (32).

The density $q=\overline{p}/\int_a^b\overline{p}d\mu$ 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

$$
\int_a^b|p-q|\,d\mu\leqslant 2\left\{1\wedge\left[\frac{4}{3}\left(\left(1+\sqrt{2LV\Gamma}\right)^{1/D}-1\right)^2\right]\right\}. \qquad (65)
$$

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

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

Applying this inequality with $z=D^{-1}\log\left(1+\sqrt{2LV\Gamma}\right)$, we deduce (33) from (65).

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 (59). 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 $j\in\{1,2\}$, $\Gamma(f)=\Gamma\left(\mathcal{T}_j(f)\right)$. Applying $\mathcal{T}_1$ to nonincreasing continuous convex densities we establish (33) and we extend it to all monotone concave continuous densities by applying $\mathcal{T}_2$. □

Proof of Proposition 6. 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

$$\int_0^1 |f - \ell_f|\, d\mu \leqslant \frac{1}{2} \frac{(f_r'(0) - 1)\,(1 - f_l'(1))}{f_r'(0) - f_l'(1)}, \tag{66}$$

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,

$$\ell_f(x) \leqslant f(x) \leqslant \min\left\{f_r'(0)x, 1 + f_l'(1)(x-1)\right\} \quad \text{for all } x \in [0,1].$$

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

$$\begin{aligned}\int_0^1 |f - \ell_f|\, d\mu &\leqslant \int_0^c (f_r'(0) - 1)x d\mu + \int_c^1 (1 - f_l'(1))(1-x) d\mu \\ &= \frac{1}{2}\left[(f_r'(0) - 1)c^2 + (1 - f_l'(1))(1-c)^2\right].\end{aligned}$$

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 (67) that

$$\int_0^1 |f - \ell_f|\, d\mu \leqslant \frac{c(1-c)}{2}\left(f_r'(0) - f_l'(1)\right)$$

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

$$\int_0^1 |f - \ell_f|\, d\mu \leqslant \frac{1}{8}\left(f_r'(0) - f_l'(1)\right). \tag{67}$$

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 (67) and the convexity of $z \mapsto 1/z$ on $(0, +\infty)$ that

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

which leads to

$$\int_0^1 |f - \ell_f|\, d\mu \leqslant \frac{1}{8}\left(\frac{1}{f_l'(1)} - \frac{1}{f_r'(0)}\right). \tag{68}$$

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 (55) and (56). Note that (55) is clearly true when f is constant on $[a,b]$ and we may therefore assume that $f(a) \neq$

$f(b)$. We obtain (55) and (56) by applying (67) and (68) 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

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

□

Proof of Proposition 7. 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 $0 < \int_a^b g d\mu \leqslant 1$. 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,

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

Let $q > 1$ and

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

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 $i \in \{1, \ldots, D-1\}$ 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 $g'_l(x) \leqslant g'(x) \leqslant g'_r(x)$ for all $x \in (a,b)$. Since g is convex, we may write

$$g(x) \geqslant g(x_i) + g'(x_i)(x - x_i) \quad \text{for all } i \in \{1, \ldots, D\} \text{ and } x \in [a,b]. \tag{69}$$

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

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

hence,

$$g'(x_i)\Delta_{i+1}^2 \leqslant 2\int_{x_i}^{x_i+1} (g - g(x_i)) d\mu. \tag{70}$$

Moreover, applying (69) with $x_i = x_D = b$ and using the facts that g is nonnegative and nonincreasing, we get

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

consequently

$$g(b) \leqslant \sqrt{2g'(b)}. \tag{71}$$

Let g_i be the restriction of g on the interval $[x_{i-1}, x_i]$ and $\overline{g}$ the function on $[a,b]$ that coincides on $[x_{i-1}, x_i]$ with ℓ_{g_i} defined by (54) with $f = g_i$ for all $i \in \{1, \dots, D\}$. The function $\overline{g}$ is a D-linear interpolation of g on $[a,b]$ that satisfies

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

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

$$\int_a^b |g - \overline{g}|\, d\mu \leqslant \frac{1}{8}\left[\Delta_D^2 g'(b) + (q^2 - 1)\sum_{i=1}^{D-1} g'(x_i)\Delta_{i+1}^2\right]. \tag{72}$$

Using (70), (71) and the fact that g is nondecreasing,

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

It follows from (72) that

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

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

$$\begin{aligned}
\int_a^b |g-\overline{g}|\,d\mu &\leqslant \frac{1+4(1+\sqrt{2})}{8}\left[\left(1+L\sqrt{g'(b)}\right)^{1/D}-1\right]^2 \\
&\leqslant \frac{4}{3}\left[\left(1+L\sqrt{g'(b)}\right)^{1/D}-1\right]^2. \qquad (73)
\end{aligned}$$

Let us now prove Proposition 7 in the case where p is convex and nondecreasing. Then $p_r'(a)\geqslant 0$ 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 $\overline{p}=p$. Otherwise $0<\int_a^b g d\mu\leqslant 1$, since ℓ is nonnegative on $[a,b]$, and we may apply our previous result to g. This leads to a D-linear interpolation $\overline{g}$ of g from which we may define $\overline{p}=\overline{g}-\ell(x)$ which is a D-linear interpolation of p on $[a,b]$. Inequality (57) follows from (73) and the facts that $|p-\overline{p}|=|g-\overline{g}|$ and $g_l'(b)=p_l'(b)-p_r'(a)$.

In order to prove Proposition 7 in the other cases, we use transformations defined by (59). 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 $\overline{g}$ on $[a,b]$ such that

$$\int_a^b |g-\overline{g}|\,d\mu \leqslant \frac{4}{3}\left[\left(1+L\sqrt{|g_l'(b)-g_r'(a)|}\right)^{1/D}-1\right]^2, \qquad (74)$$

and the function $\overline{p}=\mathcal{T}_1(\overline{g})$ is a D-linear interpolation of $p=\mathcal{T}_1(g)$ that satisfies (57). 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 $S=\int_a^b p d\mu\leqslant 1$. Since p is concave and monotone, $(p(a)\vee p(b))(b-a)/2\leqslant S$, hence $p(x)\leqslant m$ for all $x\in[a,b]$, and

$g = \mathcal{T}_2(p) = m - p$ is a monotone convex sub-density since

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

Applying the previous result, we may find a D-interpolation $\overline{g}$ of g that satisfies (74) and $\overline{p} = \mathcal{T}_2(\overline{g})$ satisfies (57). This completes the proof of Proposition 7. □

Proof of Proposition 8. We start with the following lemma.

Lemma 7. *If f and g are two continuous increasing or decreasing functions from on interval I onto an interval J,*

$$\int_I |f-g| d\mu = \int_J |f^{-1} - g^{-1}| d\mu. \tag{75}$$

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

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

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

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

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

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

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

$$\int_I |f-g|\, d\mu = \int_K \left|\overline{f} - \overline{g}\right| d\mu \text{ and } \int_J \left|\overline{f}^{-1} - \overline{g}^{-1}\right| d\mu = \int_J \left|f^{-1} - g^{-1}\right| d\mu.$$

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 $t \in J$

$$\mu\left(\{x \in I,\ f(x) \geqslant t\}\right) = \mu\left(\left\{x \in I,\ x \leqslant f^{-1}(t)\right\}\right) = f^{-1}(t).$$

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

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

□

Let us now turn to the proof of (58). We first claim that it is sufficient to establish (58) when p is decreasing. If p were increasing, we could apply the result to $q = \mathcal{T}_1(p)$ defined by (59), which is then decreasing, and find a D-linear interpolation of q on $[a, b]$ that satisfies

$$\int_a^b |q - \overline{q}|\, d\mu \leqslant \frac{4}{3}\left[\left(1 + V\sqrt{\left|\frac{1}{q_r'(a)} - \frac{1}{q_l'(b)}\right|}\right)^{1/D} - 1\right]^2.$$

We then conclude by using the facts that $\mathcal{T}_1$ is an $\mathbb{L}_1$-isometry, the function $\overline{p} = \mathcal{T}_1(\overline{q})$ 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 changing p into $p(\cdot - a)$, which amounts to translate the sub-density p, we may also assume with no loss of generality that $a = 0$. By Lemma 7, $s = p^{-1}$ 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 7 and find a D-linear interpolation $\overline{s}$ of s on $[p(b), p(0)]$ that satisfies

$$\int_{p(b)}^{p(0)} |s - \overline{s}|\, d\mu \leqslant \frac{4}{3}\left[\left(1 + V\sqrt{\left|\frac{1}{p_r'(0)} - \frac{1}{p_l'(b)}\right|}\right)^{1/D} - 1\right]^2.$$

Since s is continuous and decreasing from $[p(b), p(0)]$ onto $[0, b]$, so is $\overline{s}$, and we may set $\overline{p} = \overline{s}^{-1} : [0, b] \to [p(b), p(0)]$. The function $\overline{p}$ is a D-linear interpolation of p on $[0, b]$ and we conclude by using the equality $\int_{p(b)}^{p(0)} |s - \overline{s}|\, d\mu = \int_0^b |p - \overline{p}|\, d\mu$ which is a consequence of Lemma 7. □

7.5. **Proofs of Section 6.**

Proof of Proposition 4. Let $p = \exp\phi \in \overline{\mathcal{M}}^{\mathrm{LC}}$, $\overline{p} = \exp\overline{\phi} \in \overline{\mathcal{O}}_D^{\mathrm{LC}}$ and $A \in \mathcal{A}(D)$ associated to $\overline{p}$. There exists a partition $\mathcal{I}$ of $\mathbb{R}$ into at most $D + 1$ intervals such that the restriction of $\overline{\phi}$ to I with $I \in \mathcal{I}$ is either affine or takes the value $-\infty$. Let J be the interval $\{\phi > -\infty\}$. Let $I \in \mathcal{I}$, either

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

$$I \cap \{\phi > \overline{\phi}\} = \left(I \cap J \cap \{\phi > \overline{\phi}\}\right) \cup \left(I \cap J^c \cap \{\phi > \overline{\phi}\}\right)$$

is an interval and

$$\{p > \overline{p}\} = \{\phi > \overline{\phi}\} = \bigcup_{I \in \mathcal{I}} \left(I \cap \{\phi > \overline{\phi}\}\right)$$

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

Let us now turn to the set $\{\phi < \overline{\phi}\}$ and define $K = (a, b)$ as the open interval $\{\phi > -\infty\} \cap \{\overline{\phi} > -\infty\}$ with $a \in \mathbb{R} \cup \{-\infty\}$, $b \in \mathbb{R} \cup \{+\infty\}$, $a \leqslant b$. If $K = \varnothing$, i.e. $a = b$, $\mathbb{R} = \{\phi = -\infty\} \cup \{\overline{\phi} = -\infty\}$ are disjoint and

$$\{\phi < \overline{\phi}\} = \{\overline{\phi} > -\infty\}$$

is an interval. Otherwise K is not empty, $\psi = \phi - \overline{\phi}$ 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 $A \cap K \neq \varnothing$. In any case,

$$\left\{x \in K,\ \phi(x) < \overline{\phi}(x)\right\} = \left\{x \in K,\ \phi(x) - \overline{\phi}(x) < 0\right\}$$

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 $\overline{\phi}(b) = -\infty$ or $\phi(b) = -\infty$ and the set $\{x \geqslant b,\ \phi(x) < \overline{\phi}(x)\}$ is either empty or is an interval that contains b. We may argue similarly to establish that $\{x \leqslant a,\ \phi(x) < \overline{\phi}(x)\}$ is either empty or an interval that contains a. This means that the set

$$\{\phi < \overline{\phi}\} = \left(K \cap \{\phi < \overline{\phi}\}\right) \cup \left(K^c \cap \{\phi < \overline{\phi}\}\right)$$

is a union of at most $|A \cap K| + 2 \leqslant D + 2$ intervals. We conclude by using Lemma 1 of Baraud and Birgé (2016). □

Proof of Proposition 5. Let us first assume that $p = \exp \phi > 0$ on $\mathbb{R}$. Then ϕ is continuous, concave on $\mathbb{R}$ and since p is integrable, ϕ tends to $-\infty$ at $\pm\infty$ and reaches its maximum at a point $m \in \mathbb{R}$. The sets $\overline{\mathcal{M}}^{\mathrm{LC}}$ and $\overline{\mathcal{O}}_D^{\mathrm{LC}}$ are both location and scale invariant and the $\mathbb{L}_1$-distance between two densities remains unchanged under such transformations. By changing p into

$$x \mapsto \sigma p\,(m + \sigma x) = \exp\left[\phi\,(m + \sigma x) + \log \sigma\right]$$

with $\sigma = 1/p(m)$, if ever necessary, we may assume with no loss of generality that $m = 0$ and $0 < \phi(x) \leqslant \phi(0) = 0$ for all $x \in \mathbb{R}$, what we shall do. We use the following approximation result.

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

$$\begin{aligned}\int_a^b |f-\ell_1|\,d\mu &\leqslant \frac{(b-a)^2}{2D^2}\frac{|f'_r(a)-\Delta|\,|\Delta-f'_l(b)|}{\left|f'_r(a)-f'_l(b)\right|} \\ &\leqslant \frac{(b-a)^2}{8D^2}\left|f'_r(a)-f'_l(b)\right|.\end{aligned} \tag{76}$$

(ii) If ϕ 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

$$\int_a^b |f-\ell_2|\,d\mu \leqslant \frac{(f(a)-f(b))^2}{8D^2}\left|\frac{1}{f'_r(a)}-\frac{1}{f'_l(b)}\right| \tag{77}$$

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

Proof. If f is a concave function on $[0,1]$ satisfying $f(0)=0$ and $f(1)=1$, with a right derivative at 0 and a left derivative at 1, Guérin *et al* (2006) proved that there exist a D-linear interpolation ℓ of f on $[0,1]$ such that

$$\int_0^1 |f-\ell|\,d\mu \leqslant \frac{1}{2D^2}\frac{(f'_r(0)-1)\,(1-f'_l(1))}{f'_r(0)-f'_l(1)}. \tag{78}$$

The results established in Proposition 9 are deduced from (78) by arguing as in the proof of Proposition 6. □

Let $t=2.3$, $a_0=b_0=0$ and a_1,b_1 be numbers such that $a_1<0<b_1$ and $|\phi(a_1)|=|\phi(b_1)|=t>0$. Since ϕ is concave,

$$\phi(x)\geqslant t\left(\frac{x}{|a_1|}\wedge\frac{-x}{b_1}\right) \quad \text{for } x\in[a_1,b_1].$$

In particular,

$$1=\int_{\mathbb{R}} e^{\phi}d\mu > \int_{a_1}^{0} e^{tx/|a_1|}d\mu + \int_0^{b_1} e^{-tx/b_1}d\mu,$$

and consequently,

$$1 > \frac{(|a_1|+b_1)\left(1-e^{-t}\right)}{t}, \text{i.e.} \quad |a_1|+b_1 \leqslant \frac{t}{1-e^{-t}}. \tag{79}$$

Let us now define for all $j\geqslant 2$,

$$a_j=\inf\{x\leqslant a_1, \phi(x)\geqslant -jt\} \quad \text{and} \quad b_j=\sup\{x\geqslant b_1,\phi(x)\geqslant -jt\}.$$

Since ϕ is continuous, concave and tends to $-\infty$ at $\pm\infty$, a_j and b_j exist, satisfy $\phi(a_j)=\phi(b_j)=-jt$ and for all $j\geqslant 1$.

Let $J \geqslant 2$ and $D_1, \dots, D_J$ be positive integers. Applying Proposition 9 to ϕ, we can find two D_1-linear interpolations ℓ_1^- and ℓ_1^+ of ϕ on $[a_1, 0]$ and $[0, b_1]$ respectively that satisfy (76) and for all $j \in \{2, \dots, J\}$, we can find two D_j-linear interpolations ℓ_j^- and ℓ_j^+ of ϕ on $[a_j, a_{j-1}]$ and $[b_{j-1}, b_j]$ respectively that satisfy (77). The function

$$\overline{\phi} = \sum_{i=1}^{J} \left[\ell_j^- \mathbb{1}_{[a_j, a_{j-1})} + \ell_j^+ \mathbb{1}_{[b_{j-1}, b_j)} \right] + (-\infty) \mathbb{1}_{\mathbb{R} \setminus [a_J, b_J]}$$

is $(2 \sum_{j=1}^J D_j)$-linear interpolation of ϕ on $[a_J, b_J]$ and the function $\overline{p} = \exp \overline{\phi}$ is integrable and satisfies

$$1 \geqslant p(x) = e^{\phi(x)} \geqslant \overline{p} = e^{\overline{\phi}(x)} \quad \text{for all } x \in \mathbb{R}.$$

Since ϕ is concave and nondecreasing on $[a_1, 0]$, $\phi_l'(0) \geqslant 0$ and

$$\begin{aligned} 0 &\leqslant \frac{a_1^2}{\phi_r'(a_1) - \phi_l'(0)} \left(\phi_r'(a_1) - \frac{\phi(0) - \phi(a_1)}{|a_1|} \right) \left(\frac{\phi(0) - \phi(a_1)}{|a_1|} - \phi_l'(0) \right) \\ &\leqslant \frac{a_1^2}{\phi_r'(a_1) - \phi_l'(0)} \left(\phi_r'(a_1) - \frac{t}{|a_1|} \right) \left(\frac{t}{|a_1|} - \phi_l'(0) \right) \\ &\leqslant \frac{a_1^2}{\phi_r'(a_1)} \left(\phi_r'(a_1) - \frac{t}{|a_1|} \right) \frac{t}{|a_1|} = t \left(|a_1| - \frac{t}{\phi_r'(a_1)} \right). \end{aligned}$$

By arguing similarly on the interval $[0, b_1]$ and using the fact that $\phi_r'(0) \leqslant 0$, we obtain that

$$\begin{aligned} 0 &\leqslant \frac{b_1^2}{\phi_r'(0) - \phi_l'(b)} \left(\phi_r'(0) - \frac{\phi(b_1) - \phi(0)}{b_1} \right) \left(\frac{\phi(b_1) - \phi(0)}{b_1} - \phi_l'(b_1) \right) \\ &\leqslant t \left(b_1 - \frac{t}{|\phi_l'(b_1)|} \right). \end{aligned}$$

Applying Proposition 9-(i), we obtain that

$$\int_{a_1}^{b_1} \left| \phi - \overline{\phi} \right| d\mu \leqslant \frac{t}{2D_1^2} \left[\left(|a_1| - \frac{t}{\phi_r'(a_1)} \right) + \left(b_1 - \frac{t}{|\phi_l'(b_1)|} \right) \right]. \tag{80}$$

Applying Proposition 9-(ii) on the intervals $[a_j, a_{j-1}]$ and $[b_{j-1}, b_j]$ respectively with $j \in \{2, \dots, J\}$, we obtain that

$$\int_{a_j}^{a_{j-1}} \left| \phi - \overline{\phi} \right| d\mu \leqslant \frac{t^2}{8D_j^2} \left(\frac{1}{\phi_l'(a_{j-1})} - \frac{1}{\phi_r'(a_j)} \right) \tag{81}$$

$$\int_{b_{j-1}}^{b_j} \left| \phi - \overline{\phi} \right| d\mu \leqslant \frac{t^2}{8D_j^2} \left(\frac{1}{\phi_l'(b_j)} - \frac{1}{\phi_r'(b_{j-1})} \right). \tag{82}$$

Choosing $K = \sqrt{t/(1 - e^{-t})}$,

$$D_1 = \left\lceil KD\sqrt{t/2} \right\rceil \quad \text{and for all } j \geqslant 2 \quad D_j = \left\lceil KDe^{-(j-1)t/2} \sqrt{t/8} \right\rceil,$$

we deduce from (80) that

$$
\begin{aligned}
\int_{a_1}^{b_1} |p-\overline{p}|\,d\mu &= \int_{a_1}^{b_1} e^{\phi}\left(1-e^{-(\phi-\overline{\phi})}\right)d\mu \leqslant e^{\phi(0)}\int_{a_1}^{b_1}\left(\phi-\overline{\phi}\right)d\mu\\
&\leqslant \frac{t}{2D_1^2}\left[\left(|a_1|-\frac{t}{\phi_r'(a_1)}\right)+\left(b_1-\frac{t}{|\phi_l'(b_1)|}\right)\right]\\
&\leqslant \frac{1}{K^2D^2}\left[\left(|a_1|-\frac{t}{\phi_r'(a_1)}\right)+\left(b_1-\frac{t}{|\phi_l'(b_1)|}\right)\right] \qquad (83)
\end{aligned}
$$

and from (81) and (82) that for all $j\in\{2,\ldots,J\}$,

$$
\begin{aligned}
&\int_{a_j}^{a_{j-1}} |p-\overline{p}|\,d\mu+\int_{b_{j-1}}^{b_j} |p-\overline{p}|\,d\mu\\
&\leqslant e^{\phi(a_{j-1})}\int_{a_j}^{a_{j-1}}\left(\phi-\overline{\phi}\right)d\mu+e^{\phi(b_{j-1})}\int_{b_{j-1}}^{b_j}\left(\phi-\overline{\phi}\right)d\mu\\
&\leqslant \frac{e^{-(j-1)t}t^2}{8D_j^2}\left[\frac{1}{\phi_l'(a_{j-1})}-\frac{1}{\phi_r'(a_j)}+\frac{1}{\phi_l'(b_j)}-\frac{1}{\phi_r'(b_{j-1})}\right]\\
&\leqslant \frac{1}{K^2D^2}\left[\frac{t}{\phi_l'(a_{j-1})}-\frac{t}{\phi_r'(a_j)}+\frac{t}{\phi_l'(b_j)}-\frac{t}{\phi_r'(b_{j-1})}\right]. \qquad (84)
\end{aligned}
$$

Summing up the inequalities (83) and (84) for $j\in\{1,\ldots,J\}$ and using the fact that $\phi_r'(x)\leqslant\phi_l'(x)$ for all $x\in\mathbb{R}$, we obtain

$$
\int_{a_J}^{b_J} |p-\overline{p}|\,d\mu \leqslant \frac{1}{K^2D^2}\left[|a_1|+b_1-\frac{1}{\phi_r'(a_J)}-\frac{1}{\phi_l'(b_J)}\right]. \qquad (85)
$$

For $j\geqslant J+1$ and $x\in[a_j,a_{j-1}]$, $\overline{p}(x)=0\leqslant p(x)\leqslant \phi(a_{j-1})+\phi_l'(a_{j-1})(x-a_{j-1})$ and consequently,

$$
\begin{aligned}
\int_{a_j}^{a_{j-1}} |p-\overline{p}|\,d\mu &\leqslant \int_{a_j}^{a_{j-1}} \exp\left(\phi(a_{j-1})+\phi_l'(a_{j-1})(x-a_{j-1})\right)d\mu\\
&\leqslant \frac{e^{-t(j-1)}}{\phi_l'(a_{j-1})}\left(1-e^{-\phi_l'(a_{j-1})(a_{j-1}-a_j)}\right)\leqslant \frac{e^{-t(j-1)}}{\phi_l'(a_{j-1})}\leqslant \frac{e^{-t(j-1)}}{\phi_l'(a_J)}.
\end{aligned}
$$

Similarly, using that $\overline{p}(x)=0\leqslant p(x)\leqslant \phi(b_{j-1})+\phi_r'(b_{j-1})(x-b_{j-1})$ for all $x\in[b_{j-1},b_j]$ with $j\geqslant J+1$ we obtain that

$$
\begin{aligned}
\int_{b_{j-1}}^{b_j} |p-\overline{p}|\,d\mu &\leqslant \int_{b_{j-1}}^{b_j} \exp\left(\phi(b_{j-1})+\phi_r'(b_{j-1})(x-b_{j-1})\right)d\mu\\
&\leqslant \frac{e^{-t(j-1)}}{\phi_r'(b_{j-1})}\leqslant \frac{e^{-t(j-1)}}{\phi_r'(b_J)}.
\end{aligned}
$$

We deduce that

$$\begin{aligned}\int_{\mathbb{R}\setminus[a_J,b_J]}|p-\overline{p}|\,d\mu &= \sum_{j\geqslant J+1}\left[\int_{a_j}^{a_{j-1}}|p-\overline{p}|\,d\mu+\int_{b_{j-1}}^{b_j}|p-\overline{p}|\,d\mu\right]\\ &\leqslant \left[\frac{1}{\phi_l'(a_J)}+\frac{1}{\phi_r'(b_J)}\right]\sum_{j\geqslant J+1}e^{-t(j-1)}\\ &= \left[\frac{1}{\phi_l'(a_J)}+\frac{1}{\phi_r'(b_J)}\right]\frac{e^{-tJ}}{1-e^{-t}}.\end{aligned}$$

Let us choose J such that

$$J=\left\lceil\frac{1}{t}\log\left(\frac{K^2D^2}{1-e^{-t}}\right)\right\rceil\geqslant\left(\frac{1}{t}\log\left(\frac{K^2D^2}{1-e^{-t}}\right)\right).$$

Note that $J\geqslant 2$ since $D\geqslant 6>D_0=e^t\sqrt{1-e^{-t}}/K$. With this choice of J

$$\int_{\mathbb{R}\setminus[a_J,b_J]}|p-\overline{p}|\,d\mu\leqslant\left[\frac{1}{\phi_l'(a_J)}+\frac{1}{\phi_r'(b_J)}\right]\frac{1}{K^2D^2}. \tag{86}$$

Adding inequalities (85) and (86), using (79) and Lemma 3 we obtain that the density $q=\overline{p}/\int_{\mathbb{R}}\overline{p}d\mu$ satisfies

$$\int_{\mathbb{R}}|p-q|\,d\mu\leqslant 2\int_{\mathbb{R}}|p-\overline{p}|\,d\mu\leqslant\frac{2\left(|a_1|+b_1\right)}{K^2D^2}\leqslant\frac{2t}{(1-e^{-t})K^2D^2}=\frac{2}{D^2}$$

and belongs to $\overline{\mathscr{O}}_{D'}^{\mathrm{LC}}$ with

$$\begin{aligned}D' &\leqslant 2\sum_{j=1}^{J}D_j+1\leqslant KD\sqrt{2t}\left[1+\frac{1}{2}\sum_{j=2}^{J}e^{-(j-1)t/2}\right]+2J+1\\ &\leqslant D\left[K\sqrt{2t}\left(1+\frac{e^{-t/2}}{2(1-e^{-t/2})}\right)+\frac{2}{tD_0}\log\left(\frac{K^2D_0^2}{1-e^{-t}}\right)+\frac{3}{D_0}\right]\\ &\leqslant 6D.\end{aligned}$$

Consequently,

$$\inf_{q\in\overline{\mathscr{O}}_{6D}^{\mathrm{LC}}}\int_{\mathbb{R}}|p-q|\,d\mu\leqslant\frac{2}{D^2}.$$

This inequality, which holds for all positive log-concave densities, extend to nonnegative ones by using the fact that the set gathering of the former are dense in the set gathering the latter for the $\mathbb{L}_1$-norm. □

References

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.

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.

Baraud, Y. (2021). Tests and estimation strategies associated to some loss functions. *Probab. Theory Relat. Fields*, 180(3):799–846.

Baraud, Y. and Birgé, L. (2016). Rho-estimators for shape restricted density estimation. *Stochastic Process. Appl.*, 126(12):3888–3912.

Baraud, Y. and Birgé, L. (2018). Rho-estimators revisited: General theory and applications. *Ann. Statist.*, 46(6B):3767–3804.

Birgé, L. (1987). On the risk of histograms for estimating decreasing densities. *Ann. Statist.*, 15(3):1013–1022.

Birgé, L. (1989). The Grenander estimator: a nonasymptotic approach. *Ann. Statist.*, 17(4):1532–1549.

Birgé, L. (1997). Estimation of unimodal densities without smoothness assumptions. *The Annals of Statistics*, pages 970–981.

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.*

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.

Devroye, L. and Lugosi, G. (2001). *Combinatorial methods in density estimation.* Springer Series in Statistics. Springer-Verlag, New York.

Diakonikolas, I., Kane, D. M., and Stewart, A. (2016). Efficient robust proper learning of log-concave distributions. *ArXiv*, abs/1606.03077.

Diakonikolas, I., Kane, D. M., and Stewart, A. (2017). Learning multivariate log-concave distributions. In Kale, S. and Shamir, O., editors, *Proceedings of the 2017 Conference on Learning Theory*, volume 65 of *Proceedings of Machine Learning Research*, pages 711–727. PMLR.

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.

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.

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.

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.

Gao, F. and Wellner, J. A. (2009b). On the rate of convergence of the maximum likelihood estimator of a k-monotone density. *Science in China Series A: Mathematics*, 52(7):1525–1538.

Grenander, U. (1956). On the theory of mortality measurement. II. *Skand. Aktuarietidskr.*, 39:125–153 (1957).

Grenander, U. (1981). *Abstract inference*. John Wiley & Sons, Inc., New York. Wiley Series in Probability and Mathematical Statistics.

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.

Groeneboom, P., Jongbloed, G., and Wellner, J. A. (2001). A canonical process for estimation of convex functions: the "invelope" of integrated Brownian motion $+t^4$. *Ann. Statist.*, 29(6):1620–1652.

Guérin, J., Marcotte, P., and Savard, G. (2006). An optimal adaptive algorithm for the approximation of concave functions. *Mathematical Programming*, 107(3):357–366.

Kim, A. K. H., Guntuboyina, A., and Samworth, R. J. (2018). Adaptation in log-concave density estimation. *Ann. Statist.*, 46(5):2279–2306.

Kim, A. K. H. and Samworth, R. J. (2016). Global rates of convergence in log-concave density estimation. *Ann. Statist.*, 44(6):2756 – 2779.

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*.

Rao, B. P. (1969). Estimation of a unimodal density. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 23–36.

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

Email address: helene.halconruy@uni.lu

Email address: guillaume.maillard@uni.lu