The Annals of Statistics
2018, Vol. 46, No. 5, 2279–2306
https://doi.org/10.1214/17-AOS1619

ADAPTATION IN LOG-CONCAVE DENSITY ESTIMATION

BY ARLENE K. H. KIM*,†,[1], ADITYANAND GUNTUBOYINA‡,[2] AND RICHARD J. SAMWORTH*,[3]

University of Cambridge, Sungshin Women's University† and University of California, Berkeley‡*

The log-concave maximum likelihood estimator of a density on the real line based on a sample of size n is known to attain the minimax optimal rate of convergence of $O(n^{-4/5})$ with respect to, for example, squared Hellinger distance. In this paper, we show that it also enjoys attractive adaptation properties, in the sense that it achieves a faster rate of convergence when the logarithm of the true density is k-affine (i.e., made up of k affine pieces), or close to k-affine, provided in each case that k is not too large. Our results use two different techniques: the first relies on a new Marshall's inequality for log-concave density estimation, and reveals that when the true density is close to log-linear on its support, the log-concave maximum likelihood estimator can achieve the parametric rate of convergence in total variation distance. Our second approach depends on local bracketing entropy methods, and allows us to prove a sharp oracle inequality, which implies in particular a risk bound with respect to various global loss functions, including Kullback–Leibler divergence, of $O(\frac{k}{n}\log^{5/4}(en/k))$ when the true density is log-concave and its logarithm is close to k-affine.

1. Introduction. It is well known that nonparametric shape constraints such as monotonicity, convexity or log-concavity have the potential to offer the practitioner the best of both the nonparametric and parametric worlds: on the one hand, the infinite-dimensional classes allow considerable modelling flexibility, while on the other one can often obtain estimation procedures that do not require the choice of tuning parameters. Examples include isotonic regression [Barlow et al. (1972), Van Eeden (1956)], convex regression [Hildreth (1954), Lim and Glynn (2012), Seijo and Sen (2011)], generalised additive models [Chen and Samworth (2016)], the Grenander estimator [Grenander (1956)], convex density estimation [Groeneboom, Jongbloed and Wellner (2001)], independent component analysis [Samworth and Yuan (2012)] and many others. See Groeneboom and Jongbloed (2014) for a recent book-length treatment of the field.

Received September 2016; revised July 2017.

[1]Supported by National Research Foundation of Korea (NRF) Grant 2017R1C1B5017344.

[2]Supported by NSF Grant DMS-1309356.

[3]Supported by an EPSRC Early Career Fellowship and a grant from the Leverhulme Trust.

MSC2010 subject classifications. 62G07, 62G20.

Key words and phrases. Adaptation, bracketing entropy, log-concavity, maximum likelihood estimation, Marshall's inequality.

These attractive properties have led to intensive efforts in recent years, to try to understand the theoretical properties of shape-constrained estimators. In some cases, for instance, it is now known that these estimators can achieve minimax optimal rates of convergence; see, for example, Birgé (1987) for the Grenander estimator, Baraud and Birgé (2016) for ρ-estimators, Kim and Samworth (2016) for the log-concave maximum likelihood estimator and Han and Wellner (2016) for convex regression estimators. However, the fact that these estimators are tuning-free raises the prospect of an additional allure, namely that they might adapt to certain types of data generating mechanisms in the sense of attaining a faster rate of convergence than that predicted by the "worst-case" minimax theory.

The purpose of this paper is to explore this adaptation phenomenon in the context of log-concave density estimation. Recall that a density f on the real line is said to be log-concave if it is of the form $\exp(\phi)$ for some concave function $\phi : \mathbb{R} \to [-\infty, \infty)$. We write $\mathcal{F}$ for the set of all upper semi-continuous log-concave densities. The class $\mathcal{F}$ serves as a particularly attractive nonparametric surrogate for the class of Gaussian densities, because it is closed under linear transformations, marginalisation, conditioning and convolution, and because it contains many commonly encountered parametric families of unimodal densities with exponentially decaying tails. For this reason, the log-concave maximum likelihood estimator of f, first introduced by Walther (2002), has been studied in great detail in recent years; see, for example, Cule and Samworth (2010), Cule, Samworth and Stewart (2010), Dümbgen and Rufibach (2009), Dümbgen, Samworth and Schuhmacher (2011), Pal, Woodroofe and Meyer (2007), Schuhmacher and Dümbgen (2010), Seregin and Wellner (2010).

Very recently, Kim and Samworth (2016) proved that if $X_1, \dots, X_n$ are an independent sample from $f_0 \in \mathcal{F}$, then[4]

$$\inf_{\tilde{f}_n} \sup_{f_0 \in \mathcal{F}} \mathbb{E}_{f_0} d_{\mathrm{H}}^2(\tilde{f}_n, f_0) \asymp n^{-4/5}, \tag{1}$$

and moreover that the log-concave maximum likelihood estimator $\hat{f}_n$ based on $X_1, \dots, X_n$, defined by $\hat{f}_n := \operatorname{argmax}_{f \in \mathcal{F}} \sum_{i=1}^n \log f(X_i)$, attains this minimax optimal rate. Here, the infimum is taken over all estimators $\tilde{f}_n$ of f_0, and $d_{\mathrm{H}}^2(f, g) := \int_{-\infty}^{\infty} (f^{1/2} - g^{1/2})^2$ denotes the squared Hellinger distance. In fact, there are various other choices of global loss function that one can study, including the total variation distance and Kullback–Leibler divergence, defined respectively by

$$d_{\mathrm{TV}}(f, g) := \frac{1}{2} \int_{-\infty}^{\infty} |f - g| \quad \text{and} \quad d_{\mathrm{KL}}^2(f, g) := \int_{-\infty}^{\infty} f \log \frac{f}{g},$$

where we set $d_{\mathrm{KL}}^2(f, g) := \infty$ if the support of f is not contained in the support of g. We recall the standard inequalities $d_{\mathrm{TV}}^2(f, g) \le d_{\mathrm{H}}^2(f, g) \le d_{\mathrm{KL}}^2(f, g)$ that

[4]Here, we write $a_n \asymp b_n$ to mean that $0 < \liminf_{n\to\infty} |a_n/b_n| \le \limsup_{n\to\infty} |a_n/b_n| < \infty$.

relate these loss functions. In fact, in this work, we will also be interested in another notion of divergence: by an application of Remark 2.3 of Dümbgen, Samworth and Schuhmacher (2011) to the function $x \mapsto \log \frac{f_0(x)}{\hat{f}_n(x)}$, we have that

$$d_{\mathrm{KL}}^2(\hat{f}_n, f_0) \le \frac{1}{n}\sum_{i=1}^n \log \frac{\hat{f}_n(X_i)}{f_0(X_i)} =: d_X^2(\hat{f}_n, f_0).$$

Thus, an upper bound on the risk of the log-concave maximum likelihood estimator in the d_X^2 divergence immediately yields bounds in each of the other global loss functions mentioned above.

Writing $X_{(1)} := \min_i X_i$ and $X_{(n)} := \max_i X_i$, the log-concave maximum likelihood estimator can be expressed as

$$\hat{f}_n(x) = \begin{cases} \exp\{\min(b_1 x - \beta_1, \dots, b_m x - \beta_m)\} & \text{if } x \in [X_{(1)}, X_{(n)}], \\ 0 & \text{otherwise,} \end{cases}$$

for some $m \in \mathbb{N}$ and $b_1, \dots, b_m, \beta_1, \dots, \beta_m \in \mathbb{R}$. This motivates the thought that if $\log f_0$ is itself composed of a relatively small number of affine pieces (e.g., the logarithm of a Laplace density comprises two affine pieces), then we might expect $\hat{f}_n$ to converge to f_0 at an especially fast rate.

To this end, for $k \in \mathbb{N}$, we define $\mathcal{F}^k$ to be the class of log-concave densities f for which $\log f$ is k*-affine* in the sense that there exist intervals $I_1, \dots, I_k$ such that f is supported on $I_1 \cup \cdots \cup I_k$, and $\log f$ is affine on each I_j. We then study adaptation in log-concave density estimation via two different approaches. The first, presented in Section 2, establishes risk bounds in total variation distance for true densities that are close to $\mathcal{F}^1$, showing in some cases (such as when the true density is uniform on a compact interval), that the log-concave maximum likelihood estimator achieves the parametric rate of convergence. Our key tool for this approach is an analogue of Marshall's inequality [Marshall (1970)], which we use to relate $\sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F_0(x)|$ to $\sup_{x\in\mathbb{R}} |\mathbb{F}_n(x) - F_0(x)|$, where $\mathbb{F}_n$, F_0 and $\hat{F}_n$ denote the empirical distribution function and the distribution functions corresponding to f_0 and $\hat{f}_n$ respectively. An attraction of this strategy is that the true density need not be assumed to be log-concave.

Our second approach, developed in Section 3, studies more general adaptation of the log-concave maximum likelihood estimator to densities in $\mathcal{F}^k$ via local bracketing entropy methods. More precisely, we provide risk bounds in the d_X^2 divergence when the true density is log-concave and close to $\mathcal{F}^k$, which reveal that a rate of $\frac{k}{n}\log^{5/4}(en/k)$ can be attained. Thus, when k is relatively small, we obtain a significant improvement over the minimax rate.

There has been considerable interest in adaptation in shape-constrained estimation, especially in recent years, on problems including decreasing density estimation [Birgé (1987)], isotonic regression [Chatterjee, Guntuboyina and Sen (2014), Zhang (2002)], matrix estimation under shape constraints [Chatterjee, Guntuboyina and Sen (2017)] and convex regression [Chen and Wellner (2016), Han and

Wellner (2016)]. However, all of these works consider the least squares estimator, which has a more explicit expression as a projection onto a convex set. The class of log-concave densities is not convex, and the maximum likelihood estimator does not have such a simple characterisation, so we have to develop new techniques. We finally mention the work of Baraud and Birgé (2016), who study a procedure called a ρ-estimator in various shape-constrained density estimation problems. We discuss their results in the context of log-concave density estimation in Section 3.

Proofs of our main results are given in Sections 4 and 4.2. These rely on several auxiliary results that are presented in the Appendix, which appears as the Supplementary Material [Kim, Guntuboyina and Samworth (2018)]. We conclude this Introduction with some notation used throughout the paper. Given a function $g : \mathbb{R} \to \mathbb{R}$, we write $\|g\|_\infty := \sup_{x\in\mathbb{R}} |g(x)|$. For $f, g \in \mathcal{F}$, we write $D_f := \{x : f(x) > 0\} = \{x : \log f(x) > -\infty\}$ for the domain of $\log f$, and write $f \ll g$ if $D_f \subseteq D_g$. Also for $f \in \mathcal{F}$, let $\mu_f := \int_{-\infty}^{\infty} x f(x)\,dx$, $\sigma_f^2 := \int_{-\infty}^{\infty} (x-\mu_f)^2 f(x)\,dx$ and $\mathcal{F}^{0,1} := \{f \in \mathcal{F} : \mu_f = 0, \sigma_f^2 = 1\}$. We use C to denote a generic universal positive constant, whose value may be different at different instances, and also write $a \lesssim b$ to mean that there exists a universal constant $C > 0$ such that $a \leq Cb$.

2. Rates for densities that are close to log-affine on their support. This section concerns settings where the true density is close to $\mathcal{F}^1$, the class of densities that are log-affine on their support, but not necessarily log-concave. It will be convenient to have an explicit parameterisation of such densities. Let $\mathcal{T}_0 := \{(s_1, s_2) \in \mathbb{R}^2 : s_1 < s_2\}$ and

$$\mathcal{T} := (\mathbb{R} \times \mathcal{T}_0) \cup \big((0,\infty) \times \{-\infty\} \times \mathbb{R}\big) \cup \big((-\infty, 0) \times \mathbb{R} \times \{\infty\}\big).$$

Now, for $(\alpha, s_1, s_2) \in \mathcal{T}$, let

$$f_{\alpha,s_1,s_2}(x) := \begin{cases} \dfrac{1}{s_2 - s_1} \mathbb{1}_{\{x\in[s_1,s_2]\}} & \text{if } \alpha = 0, \\ \dfrac{\alpha}{e^{\alpha s_2} - e^{\alpha s_1}} e^{\alpha x} \mathbb{1}_{\{x\in[s_1,s_2]\}} & \text{if } \alpha \neq 0. \end{cases}$$

Then we can write

$$\mathcal{F}^1 = \{f_{\alpha,s_1,s_2} : (\alpha, s_1, s_2) \in \mathcal{T}\}.$$

Thus the class $\mathcal{F}^1$ consists of uniform and (possibly truncated) exponential densities. It is also convenient to define a continuous, strictly increasing function $q : \mathbb{R} \to [0, 1]$ by

$$q(x) := \begin{cases} \dfrac{x - 2 + e^{-x}(x+2)}{x\{1 - e^{-x}(x+1)\}} & \text{for } x \neq 0, \\ \dfrac{1}{3} & \text{for } x = 0, \end{cases} \tag{2}$$

and to set $\rho(x) := \frac{1+q(x)}{1-q(x)}$. As a preliminary calculation, we note that for $x \geq 2$,

$$q(x) = 1 - \frac{2}{x} + \frac{x}{e^x - (1+x)} \leq 1 - \frac{1}{x},$$

so that $\rho(x) \leq \max\{\rho(2), \rho(x)\} \leq \max(3, 2x)$ for all $x \in \mathbb{R}$.

THEOREM 1. *Let f_0 be any density on the real line, let $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} f_0$ for some $n \geq 5$, and let $\hat{f}_n$ denote the corresponding log-concave maximum likelihood estimator. Fix an arbitrary $f_{\alpha,s_1,s_2} \in \mathcal{F}^1$, write $\kappa^* := \alpha(s_2 - s_1)$, let $d_{\mathrm{TV}} := d_{\mathrm{TV}}(f_{\alpha,s_1,s_2}, f_0)$ and let $d_{\mathrm{KS}}^{(n)} := \|F_{\alpha,s_1,s_2}^n - F_0^n\|_\infty + \|(1 - F_{\alpha,s_1,s_2})^n - (1 - F_0)^n\|_\infty$, where F_{α,s_1,s_2} and F_0 are the distribution functions corresponding to f_{α,s_1,s_2} and f_0, respectively. Then, for $t \geq 0$, the following two bounds hold*:

$$\mathbb{P}_{f_0}\bigl[d_{\mathrm{TV}}(\hat{f}_n, f_0) \geq t + \bigl\{1 + 2\rho\bigl(|\kappa^*|\bigr)\bigr\}d_{\mathrm{TV}}\bigr] \leq 2e^{-\frac{nt^2}{2\rho^2(|\kappa^*|)}} + d_{\mathrm{KS}}^{(n)}, \tag{3}$$

$$\mathbb{P}_{f_0}\bigl\{d_{\mathrm{TV}}(\hat{f}_n, f_0) \geq t + (1 + 6\log n)d_{\mathrm{TV}}\bigr\} \leq 2e^{-\frac{nt^2}{18\log^2 n}} + \frac{1}{n^{1/2}} + d_{\mathrm{KS}}^{(n)}, \tag{4}$$

where we interpret (3) as uninformative if $|\kappa^*| = \infty$. Moreover,

$$\mathbb{E}_{f_0} d_{\mathrm{TV}}(\hat{f}_n, f_0) \leq \inf_{f_{\alpha,s_1,s_2} \in \mathcal{F}^1} \biggl\{\frac{c_n}{n^{1/2}} + (1 + c_n)d_{\mathrm{TV}} + 2d_{\mathrm{KS}}^{(n)}\biggr\}, \tag{5}$$

where $c_n = c_n(f_{\alpha,s_1,s_2}) := \min\{2\rho(|\kappa^*|), 6\log n\}$.

To aid with the interpretation of the last part of Theorem 1, first consider the case where $f_0 = f_{\alpha,s_1,s_2} \in \mathcal{F}^1$, so that $d_{\mathrm{TV}} = d_{\mathrm{KS}}^{(n)} = 0$. In that case, provided $|\kappa^*| = |\alpha|(s_2 - s_1)$ is not too large, the first term in the minimum in the definition of c_n guarantees that the log-concave maximum likelihood estimator attains the parametric rate of convergence. In particular, if $f_0 \in \mathcal{F}^1$ is a uniform density on a compact interval, then we may take $\alpha = 0 = \kappa^*$, and find that

$$\mathbb{E}_{f_0} d_{\mathrm{TV}}(\hat{f}_n, f_0) \leq \frac{4}{n^{1/2}}.$$

On the other hand, if $f_0 = f_{\alpha,s_1,s_2} \in \mathcal{F}^1$ where $|\kappa^*|$ is large (e.g., if it is infinite) then the second term in the minimum in the definition of c_n may give a better bound, and guarantees that we attain the parametric rate up to a logarithmic factor. More generally, there exist universal constants $\varepsilon > 0$ and $n_0 \in \mathbb{N}$ such that if $n \geq n_0$ and f_0 is any density with

$$\inf_{f_{\alpha,s_1,s_2} \in \mathcal{F}^1} \bigl(d_{\mathrm{TV}} + d_{\mathrm{KS}}^{(n)}\bigr) \leq \varepsilon n^{-2/5} \log^{-1} n,$$

then the bound provided by (5) is better than that given by the worst-case minimax theory.[5] In fact, there is a special class $\mathcal{F}_* \subseteq \mathcal{F}$ such that when $f_0 \in \mathcal{F}_*$ we can prove an alternative bound on the total variation distance between $\hat{f}_n$ and f_0 that slightly improves and simplifies the bounds provided in Theorem 1. To define this class, for $f \in \mathcal{F}$, let $D_f := \{x : f(x) > 0\}$, and let

$$\mathcal{F}_* := \bigl\{f \in \mathcal{F} : f(x) = e^{\gamma x} h(x) \text{ for all } x \in D_f, \text{ for some } \gamma \in \mathbb{R}, \\ \text{and } h : D_f \to [0, \infty) \text{ concave}\bigr\}. \tag{6}$$

As examples, if $f \in \mathcal{F}$ is concave on its (necessarily bounded) support D_f, then $f \in \mathcal{F}_*$ since we can take $\gamma = 0$ and $h(x) = f(x)$ for $x \in D_f$. Moreover, $\mathcal{F}^1 \subseteq \mathcal{F}_*$, and the family of $\Gamma(\alpha, \beta)$ densities with $\alpha \in [1, 2]$, $\beta > 0$ also belongs to $\mathcal{F}_*$. When $f_0 \in \mathcal{F}_*$, the factors of $1 + 2\rho(|\kappa^*|)$, $1 + 6 \log n$ and $1 + c_n$ in (3), (4) and (5), respectively, can be replaced simply with 3. See Proposition 6 in Section 4.1 for details.

Another setting where we can provide an alternative version of Theorem 1 occurs when f_0 is a (not necessarily log-concave) density that is both close to $\mathcal{F}^1$ and has tails that are lighter than those of an exponential distribution. We refer the reader to Proposition 7 in Section 4.1 for further details and discussion.

The proof of Theorem 1 is crucially based on the following analogue of the classical Marshall's inequality for decreasing density estimation [Marshall (1970)].

LEMMA 2. *Let $n \geq 2$, let $X_1, \ldots, X_n$ be real numbers that are not all equal, with empirical distribution function $\mathbb{F}_n$, and let $\hat{f}_n$ denote the corresponding log-concave maximum likelihood estimator. Let $X_{(1)} := \min_i X_i$ and $X_{(n)} := \max_i X_i$. Let f_0 be a density such that $f_0(x) = e^{\alpha_0 x} h_0(x)$ for $x \in [X_{(1)}, X_{(n)}]$, where $\alpha_0 \in \mathbb{R}$ and $h_0 : [X_{(1)}, X_{(n)}] \to \mathbb{R}$ is concave, and let $\kappa := \alpha_0(X_{(n)} - X_{(1)})$. Writing F_0 and $\hat{F}_n$ for the distribution functions corresponding to f_0 and $\hat{f}_n$ respectively, we have*

$$\|\hat{F}_n - F_0\|_\infty \leq \rho\bigl(|\kappa|\bigr) \|\mathbb{F}_n - F_0\|_\infty. \tag{7}$$

REMARK. Dümbgen and Rufibach (2009) found that in all of their simulations,

$$\|\hat{F}_n - F_0\|_\infty \leq \|\mathbb{F}_n - F_0\|_\infty, \tag{8}$$

provided F_0 has a log-concave density. However, since $\rho(x) \to \infty$ as $x \to \infty$, it is worth noting that Lemma 2 is in line with their observation that "…one can construct counterexamples showing that (8) may be violated, even if the right-hand side is multiplied with any fixed constant $C > 1$".

[5]The fact that we work with the risk in total variation distance rather than squared total variation distance is not significant. However, it is worth recalling that Theorem 1 does not control the (larger) Hellinger risk.

Although Lemma 2 is stated as a deterministic result, the main case of interest is where $X_1, \dots, X_n$ are independent and identically distributed, and we apply the result to some density $f_0 \in \mathcal{F}_*$ (not necessarily the true density). The original Marshall's inequality applies to the integrated Grenander estimator when F_0 is concave; in that case, the multiplicative factor $\rho(|\kappa|)$ can be replaced with 1. Dümbgen, Rufibach and Wellner (2007) proved a similar result for the integrated version of the least squares estimator of a convex density on $[0, \infty)$; there, a multiplicative constant 2 is needed. In the special case where f_0 is concave on the convex hull of the data, we can take $\alpha_0 = 0 = \kappa$, and the multiplicative constant in Lemma 2 can also be taken to be 2. A limitation of Lemma 2 is the fact that not all log-concave densities are of the form required for our result to hold. Indeed, our proof crucially relies on the fact that if a and b are consecutive knots of $\hat{f}_n$ and $x \in [a, b]$, then we can write

$$\hat{F}_n(x) - F_0(x) = c + \int_a^x e^{\alpha_0 t} g(t)\, dt,$$

for some constant $c \in \mathbb{R}$, where g is *convex*; cf. Lemma 1 in the Supplementary Material [Kim, Guntuboyina and Samworth (2018)]. It is an interesting open question as to whether the restriction on f_0 in Lemma 2 can be removed.

3. Rates for densities whose logarithms are close to k-affine. In this section, we extend significantly the class of densities for which we can prove adaptation of the log-concave maximum likelihood estimator. Recall that, for $k \in \mathbb{N}$, the class $\mathcal{F}^k$ denotes the set of log-concave densities $f \in \mathcal{F}$ for which $\log f$ is k-affine. The following is the main theorem of this section.

THEOREM 3. *There exists a universal constant $C > 0$ such that for every $n \geq 2$ and every $f_0 \in \mathcal{F}$, we have*

$$\mathbb{E}_{f_0} d_{\mathrm{X}}^2(\hat{f}_n, f_0) \leq \inf_{k \in \mathbb{N}} \left\{ \frac{Ck}{n} \log^{5/4}\left(\frac{en}{k}\right) + \inf_{f_k \in \mathcal{F}^k} d_{\mathrm{KL}}^2(f_0, f_k) \right\}. \tag{9}$$

One consequence of Theorem 3 is that when $\log f_0$ is k-affine for some k, then the log-concave MLE $\hat{f}_n$ converges to f_0 at nearly the parametric rate when k is small. In particular, there exist universal constants $\varepsilon > 0$ and $n_0 \in \mathbb{N}$ such that if $n \geq n_0$ and $k \leq \varepsilon n^{1/5} \log^{-5/4} n$, then the bound provided by Theorem 3 is better than the minimax bound over all log-concave densities [Kim and Samworth (2016)].[6]

[6] Although Theorem 5 of Kim and Samworth (2016) is stated for the squared Hellinger risk, it can easily be extended to a bound for the d_X^2 risk by appealing to Corollary 7.5 of van de Geer (2000).

A result similar to (9) was recently proved by Baraud and Birgé (2016) for their ρ-estimator. More precisely, they proved that there exists a universal constant $C > 0$ such that

$$\mathbb{E}_{f_0} d_{\mathrm{H}}^2(\hat{f}_\rho, f_0) \le C \inf_{k \in \mathbb{N}} \left\{ \frac{k}{n} \log^3\left(\frac{en}{k}\right) + \inf_{f_k \in \mathcal{F}^k} d_{\mathrm{H}}^2(f_0, f_k) \right\}, \tag{10}$$

where $\hat{f}_\rho$ denotes the ρ-estimator based on a sample of size n, defined in Baraud and Birgé (2016). The differences between Theorem 3 and (10) are as follows:

1. Theorem 3 deals with the log-concave MLE while (10) deals with the ρ-estimator. While the ρ-estimator is very interesting and general, at the moment, we are not aware of algorithms for computing it. On the other hand, the log-concave MLE can be easily computed via active set methods for convex optimisation [Dümbgen and Rufibach (2011)].
2. Theorem 3 is a sharp oracle inequality in the sense that the approximation term $\inf_{f_k \in \mathcal{F}^k} d_{\mathrm{KL}}^2(f_0, f_k)$ in (9) has leading constant 1.
3. Theorem 3 bounds the risk of the log-concave MLE with respect to the loss function d_{X}^2, which is larger than its squared Hellinger risk. On the other hand, the right-hand side of (10) involves $\inf_{f_k \in \mathcal{F}^k} d_{\mathrm{H}}^2(f_0, f_k)$, which may be smaller than the term $\inf_{f_k \in \mathcal{F}^k} d_{\mathrm{KL}}^2(f_0, f_k)$ that appears on the right-hand side of (9).
4. Inequality (9) has a $\log^{5/4}(en/k)$ term on the right-hand side while inequality (10) has a (larger) $\log^3(en/k)$ term.

Our proof of Theorem 3 proceeds by first studying the special case where the infimum in the right-hand side of (9) is replaced by $k = 1$. That case can be handled using empirical process theory techniques [e.g., van de Geer (2000)] together with a local bracketing entropy result for log-concave densities (cf. Theorem 4 below). Before stating this result, we first recall the following definition of bracketing entropy.

DEFINITION 1. Let $S \subseteq \mathbb{R}$, and let $\mathcal{G}$ be a class of nonnegative functions whose domains include S. For $\varepsilon > 0$, let $N_{[]}(\varepsilon, \mathcal{G}, d_{\mathrm{H}}, S)$ denote the smallest $M \in \mathbb{N}$ for which there exist pairs of functions $\{[g_{L,j}, g_{U,j}] : j = 1, \dots, M\}$ such that

$$\int_S (g_{U,j}^{1/2} - g_{L,j}^{1/2})^2 \le \varepsilon^2 \qquad \text{for every } j = 1, \dots, M$$

and such that for every $g \in \mathcal{G}$, there exists $j^* \in \{1, \dots, M\}$ with $g_{L,j^*}(x) \le g(x) \le g_{U,j^*}(x)$ for every $x \in S$. We also define the ε-bracketing entropy of $\mathcal{G}$ over S with respect to the Hellinger distance by $H_{[]}(\varepsilon, \mathcal{G}, d_{\mathrm{H}}, S) := \log N_{[]}(\varepsilon, \mathcal{G}, d_{\mathrm{H}}, S)$ and write $H_{[]}(\varepsilon, \mathcal{G}, d_{\mathrm{H}}) := H_{[]}(\varepsilon, \mathcal{G}, d_{\mathrm{H}}, \mathbb{R})$ when $S = \mathbb{R}$.

For $f_0 \in \mathcal{F}$ and $\delta > 0$, we also define $\mathcal{F}(f_0, \delta) := \{f \in \mathcal{F} : f \ll f_0, d_{\mathrm{H}}(f, f_0) \le \delta\}$. We are now in a position to state our main local bracketing entropy bound for log-concave densities.

THEOREM 4. *There exist universal constants $C, \kappa > 0$ such that for every $f_0 \in \mathcal{F}$ with $\upsilon := \inf\{d_{\mathrm{H}}(f_0, f_1) : f_1 \in \mathcal{F}^1, f_0 \ll f_1\}$, and every $\varepsilon > 0$,*

$$H_{[]}\big(2^{1/2}\varepsilon, \mathcal{F}(f_0, \delta), d_{\mathrm{H}}\big) \le C \log^{5/4}\Big(\frac{1}{\delta}\Big)\Big(\frac{\delta + \upsilon}{\varepsilon}\Big)^{1/2} \tag{11}$$

provided $\delta + \upsilon < \kappa$.

It is instructive to compare Theorem 4 with other recent global bracketing entropy results for log-concave densities on the real line. The class $\mathcal{F}$ is not totally bounded with respect to Hellinger distance, but since this metric is invariant to affine transformations, one can consider subclasses of $\mathcal{F}$ with mean and variance restrictions. More precisely, for $\xi \ge 0$ and $\eta \in (0, 1)$, let

$$\tilde{\mathcal{F}}^{\xi,\eta} := \{f \in \mathcal{F} : |\mu_f| \le \xi, |\sigma_f^2 - 1| \le \eta\}.$$

Kim and Samworth [(2016), Theorem 4] proved that

$$H_{[]}\big(\varepsilon, \tilde{\mathcal{F}}^{1,1/2}, d_{\mathrm{H}}\big) \le C\varepsilon^{-1/2}; \tag{12}$$

see also Doss and Wellner [(2016), Theorem 3.1] for a closely related result with different but similar restrictions on the class $\mathcal{F}$. Thus, Theorem 4 reveals that when $f_0 \in \mathcal{F}$ is close to some $f_1 \in \mathcal{F}^1$ with $f_0 \ll f_1$, and when $\delta > 0$ is small, the local bracketing entropy is much smaller than the global bracketing entropy described by (12).

The proof of Theorem 4 is lengthy, but the main ideas are as follows. By a triangle inequality, one can show that it suffices to prove the result for $f_0 \in \mathcal{F}^1$. In fact, by an affine transformation, it is enough to consider f_0 belonging to one of three canonical forms within the class $\mathcal{F}^1$. When $f_0 \in \mathcal{F}^1$, we have $\upsilon = 0$, and we can exploit natural boundedness properties enjoyed by $f \in \mathcal{F}(f_0, \delta)$ when $f \ll f_0$ and $\delta > 0$ is sufficiently small. For example, when f_0 is the uniform density on $[0, 1]$, it is possible to show (see Lemma 5 in the Supplementary Material) that such f satisfy $\log f(x) \le C\delta$ for all $x \in [0, 1]$ and $\log f(x) \ge -C\delta \max\{x^{-1/2}, (1-x)^{-1/2}\}$ whenever $\min(x, 1-x) \ge 4\delta^2$. These boundedness properties allow us to apply bracketing entropy bounds for bounded log-concave functions developed in Propositions 7 and 8 in the Supplementary Material to deduce the result.

Theorem 4 enables us to prove the following risk bound for the log-concave maximum likelihood estimator when the true density is close to $\mathcal{F}^1$, a key step in proving Theorem 3.

THEOREM 5. *There exists a universal constant $C > 0$ such that for every $n \ge 2$ and $f_0 \in \mathcal{F}$, we have*

$$\mathbb{E}_{f_0} d_{\mathrm{X}}^2(\hat{f}_n, f_0) \le \frac{C}{n} \log^{5/4} n + \inf_{f_1 \in \mathcal{F}^1 : f_0 \ll f_1} d_{\mathrm{H}}^2(f_0, f_1). \tag{13}$$

Since $d_{\mathrm{H}}^2(f_0, f_1) \le d_{\mathrm{KL}}^2(f_0, f_1)$ and since $d_{\mathrm{KL}}^2(f_0, f_1) = \infty$ unless $f_0 \ll f_1$, the inequality given in (13) is stronger than the inequality obtained by replacing the infimum on the right-hand side of (9) by $k = 1$.

4. Proofs of main results.

4.1. *Proofs from Section 2 and alternative total variation bounds.* We first present the proof of Theorem 1, and then give the proof of Lemma 2, on which it relies.

PROOF OF THEOREM 1. Let $\hat{F}_n$ and F_0 denote the distribution functions of $\hat{f}_n$ and f_0, respectively, and let $\mathbb{F}_n$ denote the empirical distribution function of $X_1, \ldots, X_n$. Fix $f_{\alpha,s_1,s_2} \in \mathcal{F}^1$ with $(\alpha, s_1, s_2) \in \mathcal{T}$, and let F_{α,s_1,s_2} denote its corresponding distribution function. Then $\{x : \hat{f}_n(x) \geq f_{\alpha,s_1,s_2}(x)\} = \{x : \log \hat{f}_n(x) \geq \log f_{\alpha,s_1,s_2}(x)\}$ is an interval. It follows that

$$
\begin{aligned}
d_{\mathrm{TV}}(\hat{f}_n, f_{\alpha,s_1,s_2}) &= \int_{x:\hat{f}_n(x)\geq f_{\alpha,s_1,s_2}(x)} \{\hat{f}_n(x) - f_{\alpha,s_1,s_2}(x)\}\,dx \\
&= \sup_{s\leq t} \int_s^t \{\hat{f}_n(x) - f_{\alpha,s_1,s_2}(x)\}\,dx \\
&= \sup_{s\leq t} \bigl[\hat{F}_n(t) - F_{\alpha,s_1,s_2}(t) - \{\hat{F}_n(s) - F_{\alpha,s_1,s_2}(s)\}\bigr] \\
&\leq 2\|\hat{F}_n - F_{\alpha,s_1,s_2}\|_\infty.
\end{aligned} \tag{14}
$$

Hence, writing $d_{\mathrm{TV}} := d_{\mathrm{TV}}(f_{\alpha,s_1,s_2}, f_0)$,

$$
\begin{aligned}
d_{\mathrm{TV}}(\hat{f}_n, f_0) &\leq d_{\mathrm{TV}}(\hat{f}_n, f_{\alpha,s_1,s_2}) + d_{\mathrm{TV}} \\
&\leq 2\|\hat{F}_n - F_{\alpha,s_1,s_2}\|_\infty + d_{\mathrm{TV}} \\
&\leq 2\rho(|\kappa|)\|\mathbb{F}_n - F_{\alpha,s_1,s_2}\|_\infty + d_{\mathrm{TV}} \\
&\leq 2\rho(|\kappa|)\|\mathbb{F}_n - F_0\|_\infty + \{1 + 2\rho(|\kappa|)\}d_{\mathrm{TV}},
\end{aligned} \tag{15}
$$

where $\kappa := \alpha(X_{(n)} - X_{(1)})$. Here, the penultimate inequality follows from Lemma 2, and the final one follows by the triangle inequality and the fact that $\|F - G\|_\infty \leq d_{\mathrm{TV}}(f, g)$ for any densities f and g with corresponding distribution functions F and G, respectively. It is now convenient to introduce $Y_1, \ldots, Y_n \overset{\text{i.i.d.}}{\sim} f_{\alpha,s_1,s_2}$, with $Y_{(1)} := \min_i Y_i$ and $Y_{(n)} := \max_i Y_i$. Then, writing $\kappa^* := \alpha(s_2 - s_1)$ and $d_{\mathrm{KS}}^{(n)} := \|F_{\alpha,s_1,s_2}^n - F_0^n\|_\infty + \|(1 - F_{\alpha,s_1,s_2})^n - (1 - F_0)^n\|_\infty$,

$$
\begin{aligned}
\mathbb{P}_{f_0}\bigl(|\kappa| > |\kappa^*|\bigr) &\leq \mathbb{P}_{f_0}(X_{(n)} > s_2) + \mathbb{P}_{f_0}(X_{(1)} < s_1) \\
&\leq \mathbb{P}_{f_{\alpha,s_1,s_2}}(Y_{(n)} > s_2) + \mathbb{P}_{f_{\alpha,s_1,s_2}}(Y_{(1)} < s_1) + d_{\mathrm{KS}}^{(n)} = d_{\mathrm{KS}}^{(n)}.
\end{aligned}
$$

Since ρ is strictly increasing, we can therefore apply the Dvoretzky–Kiefer–Wolfowitz inequality [Dvoretzky, Kiefer and Wolfowitz (1956)] with the sharp constant of Massart (1990) to conclude that for any $t \geq 0$,

$$\begin{aligned}
&\mathbb{P}_{f_0}\big[d_{\mathrm{TV}}(\hat{f}_n, f_0) \geq t + \{1 + 2\rho(|\kappa^*|)\}d_{\mathrm{TV}}\big] \\
&\quad \leq \mathbb{P}_{f_0}\{2\rho(|\kappa^*|)\|\mathbb{F}_n - F_0\|_\infty \geq t\} + \mathbb{P}_{f_0}(|\kappa| > |\kappa^*|) \\
&\quad \leq 2\exp\biggl(-\frac{nt^2}{2\rho^2(|\kappa^*|)}\biggr) + d_{\mathrm{KS}}^{(n)}.
\end{aligned}$$

This proves (3). For the other bound (4), note first that if $B \geq 2$ and $\alpha < 0$, then $s_1 > -\infty$, so

$$\begin{aligned}
&\mathbb{P}_{f_0}\biggl(|\kappa| > \frac{B}{2}\log n\biggr) \\
&\quad \leq \mathbb{P}_{f_0}\biggl(X_{(n)} > s_1 - \frac{B\log n}{2\alpha}\biggr) + \mathbb{P}_{f_0}(X_{(1)} < s_1) \\
&\quad \leq \mathbb{P}_{f_{\alpha,s_1,s_2}}\biggl(Y_{(n)} > s_1 - \frac{B\log n}{2\alpha}\biggr) + \mathbb{P}_{f_{\alpha,s_1,s_2}}(Y_{(1)} < s_1) + d_{\mathrm{KS}}^{(n)} \\
&\quad = 1 - \biggl(\frac{1 - n^{-B/2}}{1 - e^{\alpha(s_2 - s_1)}}\biggr)^n + d_{\mathrm{KS}}^{(n)} \\
&\quad \leq 1 - (1 - n^{-B/2})^n + d_{\mathrm{KS}}^{(n)} \leq n^{-(B/2-1)} + d_{\mathrm{KS}}^{(n)},
\end{aligned}$$

where the final inequality follows because $1 - x \leq (1 - x/n)^n$ for $x \in [0, 1)$ (this can be proved by taking logarithms and examining the Taylor series). A very similar calculation yields the same bound when $\alpha > 0$. Recalling that $\rho(x) \leq \max(3, 2x)$, it follows that if $t \geq 0$ and $(\alpha, s_1, s_2) \in \mathcal{T}$ with $\alpha \neq 0$, then provided $B \geq 2$ and $B\log n \geq 3$,

$$\begin{aligned}
&\mathbb{P}_{f_0}\{d_{\mathrm{TV}}(\hat{f}_n, f_0) \geq t + (1 + 2B\log n)d_{\mathrm{TV}}\} \\
&\quad \leq \mathbb{P}_{f_0}\{\rho(|\kappa|) > B\log n\} + \mathbb{P}_{f_0}\{2B\log n\|\mathbb{F}_n - F_0\|_\infty \geq t\} \\
&\quad \leq n^{-(B/2-1)} + d_{\mathrm{KS}}^{(n)} + 2\exp\biggl(-\frac{nt^2}{2B^2\log^2 n}\biggr),
\end{aligned}$$

where the final inequality follows by another application of the Dvoretzky–Kiefer–Wolfowitz inequality. Taking $B = 3$ and $n \geq 3$ therefore yields (4).

Now, writing $s^* := (2\log 2)^{1/2}\rho(|\kappa^*|)/n^{1/2}$ and using the fact that the total variation distance is bounded above by 1, we have

$$\begin{aligned}
&\mathbb{E}_{f_0} d_{\mathrm{TV}}(\hat{f}_n, f_0) \\
&\quad = \mathbb{E}_{f_0}\{d_{\mathrm{TV}}(\hat{f}_n, f_0)\mathbb{1}_{\{|\kappa| \leq |\kappa^*|\}}\} + \mathbb{E}_{f_0}\{d_{\mathrm{TV}}(\hat{f}_n, f_0)\mathbb{1}_{\{|\kappa| > |\kappa^*|\}}\} \\
&\quad \leq \mathbb{E}_{f_0}\bigl([d_{\mathrm{TV}}(\hat{f}_n, f_0) - \{1 + 2\rho(|\kappa^*|)\}d_{\mathrm{TV}}]\mathbb{1}_{\{|\kappa| \leq |\kappa^*|\}}\bigr) \\
&\qquad + \{1 + 2\rho(|\kappa^*|)\}d_{\mathrm{TV}} + d_{\mathrm{KS}}^{(n)}
\end{aligned}$$

$$\le s^* + 2\int_{s^*}^{\infty} \exp\biggl(-\frac{ns^2}{2\rho^2(|\kappa^*|)}\biggr)\,ds + \bigl\{1 + 2\rho\bigl(|\kappa^*|\bigr)\bigr\}d_{\mathrm{TV}} + 2d_{\mathrm{KS}}^{(n)}$$
$$\le \frac{2\rho(|\kappa^*|)}{n^{1/2}} + \bigl\{1 + 2\rho\bigl(|\kappa^*|\bigr)\bigr\}d_{\mathrm{TV}} + 2d_{\mathrm{KS}}^{(n)}.$$

On the other hand, writing $s' := 3(2\log 2)^{1/2}n^{-1/2}\log n$, we also have

$$\mathbb{E}_{f_0} d_{\mathrm{TV}}(\hat{f}_n, f_0) \le \frac{1}{n^{1/2}} + s' + 2\int_{s'}^{\infty} e^{-\frac{ns^2}{18\log^2 n}}\,ds + (1 + 6\log n)d_{\mathrm{TV}} + 2d_{\mathrm{KS}}^{(n)}$$
$$\le \frac{6\log n}{n^{1/2}} + (1 + 6\log n)d_{\mathrm{TV}} + 2d_{\mathrm{KS}}^{(n)},$$

for $n \ge 5$. Since these inequalities hold for any $f_{\alpha,s_1,s_2} \in \mathcal{F}^1$, the conclusion follows. $\square$

PROOF OF LEMMA 2. This proof has some similarities with the proof of Dümbgen, Rufibach and Wellner (2007), Lemma 1. We define the set of *knots* of $\hat{f}_n$ by

$$\mathcal{S} := \bigl\{t \in (X_{(1)}, X_{(n)}) : \hat{f}_n'(t-) \ne \hat{f}_n'(t+)\bigr\} \cup \{X_{(1)}, X_{(n)}\},$$

where $X_{(1)}$ and $X_{(n)}$ denote the smallest and largest order statistics of the data $X_1, \dots, X_n$. By, for example, Dümbgen and Rufibach [(2009), Theorem 2.1], $\mathcal{S} \subseteq \{X_1, \dots, X_n\}$, and we therefore write $\mathcal{S} = \{t_0, \dots, t_k\}$ for some $k \in \{1, \dots, n-1\}$ where $X_{(1)} = t_0 < \cdots < t_k = X_{(n)}$. We first write the left-hand side of (7) as

$$\max\Bigl\{\sup_{x<t_0}\bigl|\hat{F}_n(x) - F_0(x)\bigr|, \max_{i\in\{0,1,\dots,k-1\}} \sup_{x\in[t_i,t_{i+1})}\bigl|\hat{F}_n(x) - F_0(x)\bigr|,$$
$$\sup_{x\ge t_k}\bigl|\hat{F}_n(x) - F_0(x)\bigr|\Bigr\}.$$

Observe now that $\hat{F}_n(x) = 0 = \mathbb{F}_n(x)$ for $x < t_0$ and $\hat{F}_n(x) = 1 = \mathbb{F}_n(x)$ for $x \ge t_k$. It therefore follows that in order to establish (7), we need only establish the two statements

$$\sup_{x\in[t_i,t_{i+1})}\bigl\{\hat{F}_n(x) - F_0(x)\bigr\} \le \rho(\kappa_+)\|\mathbb{F}_n - F_0\|_\infty \quad \text{for } i = 0, 1, \dots, k-1, \tag{16}$$

$$\inf_{x\in[t_i,t_{i+1})}\bigl\{\hat{F}_n(x) - F_0(x)\bigr\} \ge -\rho(\kappa_-)\|\mathbb{F}_n - F_0\|_\infty \quad \text{for } i = 0, 1, \dots, k-1, \tag{17}$$

where $\kappa_+ := \max(\kappa, 0)$ and $\kappa_- := \max(-\kappa, 0)$. It is convenient to prove the second statement first. We may assume that the infimum of $\hat{F}_n(x) - F_0(x)$ over

$x \in [X_{(1)}, X_{(n)})$ is attained at $r \in [t_i, t_{i+1})$, say, for some $i \in \{0, 1, \dots, k-1\}$ and let $a := t_i$ and $b := t_{i+1}$. By hypothesis, there exist $\alpha_0 \in \mathbb{R}$ and a concave function $h_0 : [a, b] \to [0, \infty)$ such that $f_0(x) = e^{\alpha_0 x} h_0(x)$ for $x \in [a, b]$. Moreover, there exist $\alpha, \beta \in \mathbb{R}$ such that $\hat{f}_n(x) = \exp(\alpha x + \beta)$ for $x \in [a, b]$. It follows that if we define

$$g(x) := e^{(\alpha-\alpha_0)x+\beta} - h_0(x) = e^{-\alpha_0 x}\{\hat{f}_n(x) - f_0(x)\},$$

then g is convex on $[a, b]$ and $g(r) = 0$. Moreover, defining $G(x) := \hat{F}_n(x) - F_0(x)$, we have

$$G(x) = c + \int_a^x e^{\alpha_0 t} g(t)\, dt \qquad \text{for } x \in [a, b],$$

where $c := \int_{-\infty}^a \hat{f}_n(t) - f_0(t)\, dt$. We may therefore apply either inequality (4) or inequality (2) in Lemma 1 in the Supplementary Material (depending on whether or not $\alpha_0 = 0$) to obtain that for every $x \in [r, b]$,

$$G(x) \le \begin{cases} G(r) + \dfrac{(x-r)^2}{(b-r)^2}\{G(b) - G(r)\} & \text{if } \alpha_0 = 0, \\ G(r) + \dfrac{1 + e^{\alpha_0(x-r)}\{\alpha_0(x-r) - 1\}}{1 + e^{\alpha_0(b-r)}\{\alpha_0(b-r) - 1\}}\{G(b) - G(r)\} & \text{if } \alpha_0 \ne 0. \end{cases}$$

Integrating from $x = r$ to $x = b$, writing $A := \alpha_0(b - r)$ and recalling the definition of the function q in (2), we deduce that

$$G(r) \ge \frac{1}{b-r}\frac{1}{1-q(-A)}\int_r^b G(x)\, dx - \frac{q(-A)}{1-q(-A)}G(b). \tag{18}$$

Now Dümbgen and Rufibach [(2009), Theorem 2.4] yields that

$$\int_{-\infty}^t \hat{F}_n(x)\, dx \le \int_{-\infty}^t \mathbb{F}_n(x)\, dx \quad \text{and} \quad \int_{-\infty}^s \hat{F}_n(x)\, dx = \int_{-\infty}^s \mathbb{F}_n(x)\, dx \tag{19}$$

for every $t \in \mathbb{R}$ and $s \in \mathcal{S}$. Moreover, Dümbgen, Samworth and Schuhmacher [(2011), Remark 2.8] gives that

$$\mathbb{F}_n(x) - \frac{1}{n} \le \hat{F}_n(x) \le \mathbb{F}_n(x) \qquad \text{for every } x \in \mathcal{S}. \tag{20}$$

It follows from (18), (19) and (20) that

$$\begin{aligned} G(r) &\ge \frac{1}{b-r}\frac{1}{1-q(-A)}\left\{\int_{-\infty}^b \hat{F}_n(x)\, dx - \int_{-\infty}^r \hat{F}_n(x)\, dx - \int_r^b F_0(x)\, dx\right\} \\ &\quad - \frac{q(-A)G(b)}{1-q(-A)} \\ &\ge \frac{1}{b-r}\frac{1}{1-q(-A)}\int_r^b \{\mathbb{F}_n(x) - F_0(x)\}\, dx - \frac{q(-A)\{\mathbb{F}_n(b) - F_0(b)\}}{1-q(-A)} \\ &\ge -\rho(-A)\|\mathbb{F}_n - F_0\|_\infty \ge -\rho(\kappa_-)\|\mathbb{F}_n - F_0\|_\infty. \end{aligned}$$

This establishes (17). For (16), let $Y_i := -X_i$, let $\hat{h}_n$ denote the log-concave maximum likelihood estimator based on $Y_1, \ldots, Y_n$, and let $\hat{H}_n$ denote its corresponding distribution function, so that by affine equivariance of the log-concave maximum likelihood estimator [Dümbgen, Samworth and Schuhmacher (2011), Remark 2.4], we have $\hat{h}_n(y) = \hat{f}_n(-y)$ and $\hat{H}_n(y) = 1 - \hat{F}_n(-y)$. Similarly, let $h_0(y) := f_0(-y)$ (so h_0 is concave on the convex hull of $Y_1, \ldots, Y_n$), and let H_0 denote the distribution function corresponding to the density h_0, so that $H_0(y) = 1 - F_0(-y)$. Finally, let $\mathbb{H}_n$ denote the empirical distribution function corresponding to $Y_1, \ldots, Y_n$, so $\mathbb{H}_n(y) = n^{-1}\sum_{i=1}^n \mathbb{1}_{\{Y_i \le y\}} = 1 - \lim_{z \searrow y} \mathbb{F}_n(-z)$. Then for any two consecutive knots a and b of $\hat{f}_n$,

$$\begin{aligned}\sup_{x\in(a,b]}\{\hat{F}_n(x) - F_0(x)\} &= \sup_{x\in(a,b]} -\{\hat{H}_n(-x) - H_0(-x)\}\\ &= -\inf_{y\in[-b,-a)}\{\hat{H}_n(y) - H_0(y)\}\\ &\le \rho(\kappa_+)\|\mathbb{H}_n - H_0\|_\infty\\ &= \rho(\kappa_+)\|\mathbb{F}_n - F_0\|_\infty,\end{aligned}$$

as required, where the inequality follows from an application of (17) to the transformed data $Y_1, \ldots, Y_n$, noting that $-\alpha_0(Y_{(n)} - Y_{(1)}) = -\alpha_0(X_{(n)} - X_{(1)}) = -\kappa$. □

Recall the definition of $\mathcal{F}_*$ in (6). We now provide a result which improves the bounds given in Theorem 1 in the special case where the true density belongs to the class $\mathcal{F}_*$.

PROPOSITION 6. *Let $n \ge 5$, let $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} f_0 \in \mathcal{F}_*$, and let $\hat{f}_n$ denote the corresponding log-concave maximum likelihood estimator. Fix an arbitrary $f_{\alpha,s_1,s_2} \in \mathcal{F}^1$, write $\kappa^* := \alpha(s_2 - s_1)$, let $d_{\mathrm{TV}} := d_{\mathrm{TV}}(f_{\alpha,s_1,s_2}, f_0)$ and let $d_{\mathrm{KS}}^{(n)} := \|F_{\alpha,s_1,s_2}^n - F_0^n\|_\infty + \|(1 - F_{\alpha,s_1,s_2})^n - (1 - F_0)^n\|_\infty$, where F_{α,s_1,s_2} and F_0 are the distribution functions corresponding to f_{α,s_1,s_2} and f_0 respectively. Then, for $t \ge 0$,*

$$\begin{aligned}&\mathbb{P}_{f_0}\{d_{\mathrm{TV}}(\hat{f}_n, f_0) \ge t + 3d_{\mathrm{TV}}\}\\ &\qquad\le \min\left\{2e^{-\frac{nt^2}{2\rho^2(|\kappa^*|)}}, \frac{1}{n^{1/2}} + 2e^{-\frac{nt^2}{18\log^2 n}}\right\} + d_{\mathrm{KS}}^{(n)}.\end{aligned}$$

Moreover,

$$\mathbb{E}_{f_0} d_{\mathrm{TV}}(\hat{f}_n, f_0) \le \inf_{f_{\alpha,s_1,s_2}\in\mathcal{F}^1}\left\{\frac{\min\{2\rho(|\kappa^*|), 6\log n\}}{n^{1/2}} + 3d_{\mathrm{TV}} + 2d_{\mathrm{KS}}^{(n)}\right\}.$$

PROOF. For any $f_{\alpha,s_1,s_2} \in \mathcal{F}^1$ with $(\alpha, s_1, s_2) \in \mathcal{T}$ and corresponding distribution function F_{α,s_1,s_2}, we have by (14) that

$$\begin{aligned} d_{\mathrm{TV}}(\hat{f}_n, f_0) &\le d_{\mathrm{TV}}(\hat{f}_n, f_{\alpha,s_1,s_2}) + d_{\mathrm{TV}} \\ &\le 2\|\hat{F}_n - F_{\alpha,s_1,s_2}\|_\infty + d_{\mathrm{TV}} \\ &\le 2\|\hat{F}_n - F_0\|_\infty + 2\|F_0 - F_{\alpha,s_1,s_2}\|_\infty + d_{\mathrm{TV}} \\ &\le 2\rho\big(|\kappa|\big)\|\mathbb{F}_n - F_0\|_\infty + 3d_{\mathrm{TV}}, \end{aligned}$$

where $\kappa := \alpha(X_{(n)} - X_{(1)})$, and where the last line follows again from Lemma 2. The proof now follows that of Theorem 1, mutatis mutandis, so we omit the details for brevity. □

As mentioned in Section 2, we can also give an alternative version of Theorem 1 that is particularly applicable in settings where f_0 is both close to $\mathcal{F}^1$ and has light tails.

PROPOSITION 7. *Let f_0 be any density on the real line, let $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} f_0$ for some $n \ge 2$, and let $\hat{f}_n$ denote the log-concave maximum likelihood estimator. Fix an arbitrary $f_{\alpha,s_1,s_2} \in \mathcal{F}^1$, let $d_{\mathrm{TV}} := d_{\mathrm{TV}}(f_{\alpha,s_1,s_2}, f_0)$ and let $\tilde{\kappa} := \alpha\{F_0^{-1}(1 - n^{-3/2}) - F_0^{-1}(n^{-3/2})\}$, where F_0^{-1} denotes the quantile function corresponding to f_0. Then, for $t \ge 0$,*

$$\mathbb{P}_{f_0}\big[d_{\mathrm{TV}}(\hat{f}_n, f_0) \ge t + \big\{1 + 2\rho\big(|\tilde{\kappa}|\big)\big\}d_{\mathrm{TV}}\big] \le 2e^{-\frac{nt^2}{2\rho^2(|\tilde{\kappa}|)}} + \frac{2}{n^{1/2}}.$$

Moreover,

$$\mathbb{E}_{f_0} d_{\mathrm{TV}}(\hat{f}_n, f_0) \le \inf_{f_{\alpha,s_1,s_2} \in \mathcal{F}^1} \left\{ \frac{\tilde{c}_n}{n^{1/2}} + (1 + \tilde{c}_n)d_{\mathrm{TV}} \right\} + \frac{4}{n^{1/2}},$$

where $\tilde{c}_n = \tilde{c}_n(f_{\alpha,s_1,s_2}) := 2\rho(|\tilde{\kappa}|)$.

PROOF. We follow the proof of Theorem 1 up to (15). Now, writing $x_0 := F_0^{-1}(n^{-3/2})$ and $x_1 := F_0^{-1}(1 - n^{-3/2})$, we have

$$\begin{aligned} \mathbb{P}_{f_0}\big(|\kappa| > |\tilde{\kappa}|\big) &\le \mathbb{P}_{f_0}(X_{(n)} > x_1) + \mathbb{P}_{f_0}(X_{(1)} < x_0) \\ &= 1 - F_0^n(x_1) + 1 - \big\{1 - F_0(x_0)\big\}^n \le \frac{2}{n^{1/2}}. \end{aligned}$$

The proof now again follows that of Theorem 1, mutatis mutandis, so we omit the details for brevity. □

A nice aspect of Proposition 7 is that the $d_{\mathrm{KS}}^{(n)}$ term in Theorem 1 (and Proposition 6) no longer appears. To illustrate Proposition 7, for $\beta \ge 1$, let $x_1 :=$

$(3/2)^{1/\beta} \log^{1/\beta} n$, let $x_* = -\log(1 + e^{-x_1} - n^{-3/2})$ and consider the density

$$f_0(x) = f_0^{(n)}(x) = e^{-x} \mathbb{1}_{\{x \in [x_*, x_1]\}} + \beta x^{\beta-1} e^{-x^\beta} \mathbb{1}_{\{x > x_1\}}.$$

Then $F_0^{-1}(n^{-3/2}) = -\log(e^{-x_*} - n^{-3/2}) = -\log(1 + e^{-x_1} - 2n^{-3/2})$ and $F_0^{-1}(1 - n^{-3/2}) = x_1$. Defining $f_{\alpha,s_1,s_2} \in \mathcal{F}^1$ by

$$f_{\alpha,s_1,s_2}(x) = \frac{e^{-x}}{e^{-x_*} - e^{-x_1}} \mathbb{1}_{\{x \in [x_*, x_1]\}},$$

we deduce that $|\tilde{\kappa}| = x_1 + \log(1 + e^{-x_1} - 2n^{-3/2}) \le 2x_1$ and $d_{\mathrm{TV}} = n^{-3/2}$, so from Proposition 7,

$$\mathbb{E}_{f_0} d_{\mathrm{TV}}(\hat{f}_n, f_0) \le \frac{8x_1}{n^{1/2}} + \frac{1 + 8x_1}{n^{3/2}} + \frac{4}{n^{1/2}} \le 17\left(\frac{3}{2}\right)^{1/\beta} \frac{\log^{1/\beta} n}{n^{1/2}}.$$

In other words, a continuum of logarithmic powers in the rates is achievable.

4.2. *Proofs from Section 3.*

PROOF OF THEOREM 3. Fix a density $f_0 \in \mathcal{F}$. Also fix $k \in \mathbb{N}$ and an arbitrary density $f \in \mathcal{F}^k$ such that $d_{\mathrm{KL}}^2(f_0, f) < \infty$. Note that this implies that $f_0 \ll f$. Suppose that $I_1, \dots, I_k$ is a partition of the support of f into maximal intervals such that $\log f$ is affine on each I_j. Since f_0 is absolutely continuous with respect to f, it follows that $\sum_{j=1}^k p_j = 1$, where $p_j := \int_{I_j} f_0$. For $j = 1, \dots, k$, we also let $N_j := \sum_{i=1}^n \mathbb{1}_{\{X_i \in I_j\}}$, $J_1 := \{j : N_j \ge 2\}$ and $J_2 := \{j : N_j \le 1\}$. Observe that the sets J_1 and J_2 as well as the integers $N_1, \dots, N_k$ are random. We initially assume that $k \le \min(n^{1/5}, n/e^{5/4})$. We write

$$\begin{aligned} d_{\mathrm{X}}^2(\hat{f}_n, f_0) &= \frac{1}{n} \sum_{i=1}^n \log \frac{\hat{f}_n(X_i)}{f_0(X_i)} \\ &= \frac{1}{n} \sum_{j \in J_1} \sum_{i: X_i \in I_j} \log \frac{\hat{f}_n(X_i)}{f_0(X_i)} + \frac{1}{n} \sum_{j \in J_2} \sum_{i: X_i \in I_j} \log \frac{\hat{f}_n(X_i)}{f_0(X_i)} \qquad (21) \\ &\le \frac{1}{n} \sum_{j \in J_1} \sum_{i: X_i \in I_j} \log \frac{\hat{f}_n(X_i)}{f_0(X_i)} + \frac{k}{n} \max_{1 \le i \le n} \log \frac{\hat{f}_n(X_i)}{f_0(X_i)}, \end{aligned}$$

where the final inequality follows because $|J_2| \le k$ and $|N_j| \le 1$ whenever $j \in J_2$. To handle the first term, let $\tilde{f}_n$ denote the maximum likelihood estimator based on the data $\{X_i : X_i \in \bigcup_{j \in J_1} I_j\}$ over the class of all densities f for which $\log f$ is concave on each of the intervals $\{I_j : j \in J_1\}$. Since $\log \hat{f}_n$ is concave on each I_j and since $\int_{-\infty}^{\infty} \hat{f}_n(x) \mathbb{1}_{\{x \in \bigcup_{j \in J_1} I_j\}} \, dx \le 1$, it follows that

$$\sum_{j \in J_1} \sum_{i: X_i \in I_j} \log \hat{f}_n(X_i) \le \sum_{j \in J_1} \sum_{i: X_i \in I_j} \log \tilde{f}_n(X_i). \qquad (22)$$

Writing $M_1 := \sum_{j \in J_1} N_j$, we claim that

$$\tilde{f}_n(x) = \frac{N_j}{M_1} \hat{f}^{(j)}(x) \qquad \text{for } x \in I_j \text{ with } j \in J_1, \tag{23}$$

where $\hat{f}^{(j)}$ denotes the log-concave maximum likelihood estimator based on $\{X_i : X_i \in I_j\}$. To see this, let $\bar{\Phi}$ denote the class of functions $\phi : \mathbb{R} \to [-\infty, \infty)$ that are concave on each I_j for $j \in J_1$ and that satisfy $\phi(x) \to -\infty$ as $|x| \to \infty$. Now, $\log \tilde{f}_n$ maximises

$$\begin{aligned} L(\phi) &:= \frac{1}{M_1} \sum_{j \in J_1} \sum_{i : X_i \in I_j} \phi(X_i) - \sum_{j \in J_1} \int_{I_j} e^{\phi} \\ &= \sum_{j \in J_1} \frac{N_j}{M_1} \biggl\{ \frac{1}{N_j} \sum_{i : X_i \in I_j} \phi(X_i) - \int_{I_j} e^{\phi + \log(M_1/N_j)} \biggr\} \end{aligned}$$

over $\phi \in \bar{\Phi}$. For $j \in J_1$, let $\bar{\Phi}_j$ denote the set of functions $\phi : I_j \to [-\infty, \infty)$ that are restrictions of functions in $\bar{\Phi}$ to I_j. Then, on each interval I_j with $j \in J_1$, we have

$$\begin{aligned} \log \tilde{f}_n &= \operatorname*{argmax}_{\phi \in \bar{\Phi}_j} \biggl\{ \frac{1}{N_j} \sum_{i : X_i \in I_j} \phi(X_i) - \int_{I_j} e^{\phi + \log(M_1/N_j)} \biggr\} \\ &= \operatorname*{argmax}_{\tilde{\phi} \in \bar{\Phi}_j} \biggl\{ \frac{1}{N_j} \sum_{i : X_i \in I_j} \tilde{\phi}(X_i) - \int_{I_j} e^{\tilde{\phi}} \biggr\} - \log \frac{M_1}{N_j} \\ &= \log \hat{f}^{(j)} - \log \frac{M_1}{N_j}, \end{aligned}$$

which establishes the claim (23). Let $f_0^{(j)}(x) := \frac{1}{p_j} f_0(x) \mathbb{1}_{\{x \in I_j\}}$. We deduce from (22) and (23) that

$$\begin{aligned} &\frac{1}{n} \mathbb{E}_{f_0} \biggl\{ \sum_{j \in J_1} \sum_{i : X_i \in I_j} \log \frac{\hat{f}_n(X_i)}{f_0(X_i)} \biggr\} \\ &\quad \le \frac{1}{n} \mathbb{E}_{f_0} \biggl\{ \sum_{j \in J_1} \sum_{i : X_i \in I_j} \log \frac{N_j \hat{f}^{(j)}(X_i)/M_1}{p_j f_0^{(j)}(X_i)} \biggr\} \\ &\quad = \frac{1}{n} \mathbb{E}_{f_0} \biggl\{ \sum_{j \in J_1} \sum_{i : X_i \in I_j} \log \frac{\hat{f}^{(j)}(X_i)}{f_0^{(j)}(X_i)} \biggr\} + \mathbb{E}_{f_0} \biggl(\sum_{j \in J_1} \frac{N_j}{n} \log \frac{N_j}{n p_j} \biggr) \\ &\qquad + \mathbb{E}_{f_0} \biggl(\frac{M_1}{n} \log \frac{n}{M_1} \biggr). \end{aligned} \tag{24}$$

Now let $f^{(j)}(x) := \frac{1}{q_j} f(x)\mathbb{1}_{\{x\in I_j\}}$, where $q_j := \int_{I_j} f$, and note both that $f^{(j)} \in \mathcal{F}^1$ and $f_0^{(j)} \ll f^{(j)}$. To evaluate the first expectation on the right-hand side of (24), we condition on the set of random variables $\{N_j : j = 1, \dots, k\}$. After this conditioning, and since $N_j \geq 2$ for every $j \in J_1$, we can apply the risk bound in Theorem 5 for each $f_0^{(j)}$ to deduce that

$$\begin{aligned}
&\frac{1}{n}\mathbb{E}_{f_0}\biggl\{\sum_{j\in J_1}\sum_{i:X_i\in I_j}\log\frac{\hat{f}^{(j)}(X_i)}{f_0^{(j)}(X_i)}\biggr\} \\
&\quad\leq \frac{1}{n}\mathbb{E}_{f_0}\sum_{j\in J_1} N_j\biggl\{\frac{C}{N_j}\log^{5/4} N_j + \inf_{f_1\in\mathcal{F}^1: f_0^{(j)}\ll f_1} d_{\mathrm{H}}^2\bigl(f_0^{(j)}, f_1\bigr)\biggr\}.
\end{aligned}$$

Now the function $x \mapsto \log^{5/4}(x)$ is concave on $[e^{1/4}, \infty)$, so by Jensen's inequality, together with the fact that the function $x \mapsto x\log^{5/4}(n/x)$ is increasing for $x \in (0, n/e^{5/4}]$,

$$\begin{aligned}
&\frac{1}{n}\mathbb{E}_{f_0}\biggl\{\sum_{j\in J_1}\sum_{i:X_i\in I_j}\log\frac{\hat{f}^{(j)}(X_i)}{f_0^{(j)}(X_i)}\biggr\} \\
&\quad\leq \frac{C}{n}\mathbb{E}\biggl\{|J_1|\log^{5/4}\biggl(\frac{n}{|J_1|}\biggr)\biggr\} + \sum_{j=1}^k p_j d_{\mathrm{H}}^2\bigl(f_0^{(j)}, f^{(j)}\bigr) \qquad (25)\\
&\quad\leq \frac{Ck}{n}\log^{5/4}(n/k) + d_{\mathrm{KL}}^2(f_0, f),
\end{aligned}$$

where the second inequality follows because

$$\begin{aligned}
\sum_{j=1}^k p_j d_{\mathrm{H}}^2\bigl(f_0^{(j)}, f^{(j)}\bigr) &\leq \sum_{j=1}^k p_j d_{\mathrm{KL}}^2\bigl(f_0^{(j)}, f^{(j)}\bigr) \\
&\leq \sum_{j=1}^k p_j d_{\mathrm{KL}}^2\bigl(f_0^{(j)}, f^{(j)}\bigr) + \sum_{j=1}^k p_j \log\frac{p_j}{q_j} = d_{\mathrm{KL}}^2(f_0, f).
\end{aligned}$$

To handle the second term on the right-hand side of (24), we use the facts that $\log x \leq x - 1$ for $x > 0$, $N_j \sim \mathrm{Bin}(n, p_j)$ and $N_j \log N_j = 0$ for $j \in J_2$, to obtain

$$\begin{aligned}
\mathbb{E}_{f_0}\biggl(\sum_{j\in J_1}\frac{N_j}{n}\log\frac{N_j}{np_j}\biggr) &\leq \sum_{j=1}^k \mathbb{E}_{f_0}\biggl\{\frac{N_j}{n}\biggl(\frac{N_j}{np_j} - 1\biggr)\biggr\} - \mathbb{E}_{f_0}\biggl(\sum_{j\in J_2}\frac{N_j}{n}\log\frac{N_j}{np_j}\biggr) \\
&= \frac{1}{n}\sum_{j=1}^k (1 - p_j) + \frac{1}{n}\sum_{j=1}^k p_j \log(np_j) \qquad (26)\\
&\leq \frac{2k}{n} + \frac{\log(n/k)}{n}.
\end{aligned}$$

Finally, for the third term on the right-hand side of (24), we note that $M_1 = n - \sum_{j \in J_2} N_j \geq n - k > n/2$, so in particular $\mathbb{E}(M_1/n) \geq 1 - k/n$. Using the fact that $x \mapsto -x \log x$ is decreasing on $[1/2, 1]$ and the fact that $-\log(1-x) \leq x + x^2$ for $x \in (0, 1/2]$, we deduce that

$$\begin{aligned}\mathbb{E}_{f_0}\left(\frac{M_1}{n}\log\frac{n}{M_1}\right) &\leq -\left(1-\frac{k}{n}\right)\log\left(1-\frac{k}{n}\right) \leq -\log\left(1-\frac{k}{n}\right) \\ &\leq \frac{k}{n} + \frac{k^2}{n^2} \leq \frac{2k}{n}.\end{aligned} \tag{27}$$

To handle the second term on the right-hand side of (21), we apply Lemma 2 in the Supplementary Material and use the fact that $k \leq n^{1/5}$ to find that

$$\mathbb{E}_{f_0}\left\{\max_{1 \leq i \leq n} \log\frac{\hat{f}_n(X_i)}{f_0(X_i)}\right\} \lesssim \log n \lesssim \log^{5/4}\left(\frac{en}{k}\right). \tag{28}$$

Thus, in the case $k \leq \min(n^{1/5}, n/e^{5/4})$, the desired conclusion follows from (21), (24), (25), (26), (27) and (28). When $k > \min(n^{1/5}, n/e^{5/4})$, we can argue as in the final paragraph of the proof of Theorem 5 below that $\mathbb{E}_{f_0} d_X^2(\hat{f}_n, f_0) \lesssim n^{-4/5} \lesssim (k/n)\log^{5/4}(en/k)$. The result follows. $\square$

PROOF OF THEOREM 4. We consider first the case where $\upsilon = 0$, so that $f_0 \in \mathcal{F}^1$. In that case, recalling the parameterisation $\mathcal{F}^1 = \{f_{\alpha,s_1,s_2} : (\alpha, s_1, s_2) \in \mathcal{T}\}$ used in Section 2, the affine invariance of the Hellinger distance, together with Lemma 4 in the Supplementary Material, shows that we may assume without loss of generality that f_0 is of one of the following three forms:

1. $f_0 = f_{0,0,1}$;
2. $f_0 = f_{-\alpha,0,1}$ for some $\alpha \in (0, 18)$;
3. $f_0 = f_{-1,0,a}$ for some $a \in [18, \infty]$.

We refer to these three forms as "uniform", "exponential conditioned on [0, 1]" and "truncated exponential", respectively, and treat the three cases separately. The choice of splitting these cases at the constant 18 is convenient rather than essential; for instance, it guarantees that the infinite sum in (35) below converges.

The case where f_0 is uniform: Fix $\delta \in (0, 2^{-5/2}]$. Observe first that for every $\varepsilon > 0$, we have

$$\begin{aligned}&H_{[]}(2^{1/2}\varepsilon, \mathcal{F}(f_0,\delta), d_{\mathrm{H}}, [0,1]) \\ &\quad \leq H_{[]}(\varepsilon, \mathcal{F}(f_0,\delta), d_{\mathrm{H}}, [0,1/2]) + H_{[]}(\varepsilon, \mathcal{F}(f_0,\delta), d_{\mathrm{H}}, [1/2,1]) \\ &\quad = 2H_{[]}(\varepsilon, \mathcal{F}(f_0,\delta), d_{\mathrm{H}}, [0,1/2]) \\ &\quad \leq 2H_{[]}(\varepsilon/2^{1/2}, \mathcal{F}(f_0,\delta), d_{\mathrm{H}}, [0,4\delta^2]) \\ &\qquad + 2H_{[]}(\varepsilon/2^{1/2}, \mathcal{F}(f_0,\delta), d_{\mathrm{H}}, [4\delta^2, 1/2]).\end{aligned} \tag{29}$$

We bound the two terms on the right-hand side of (29) separately. For the first term, we use inequality (12) in Lemma 5 in the Supplementary Material, which gives that $\sup_{x\in[0,1]} \log f(x) \le 2^{13/2}\delta \le 16$ for every $f \in \mathcal{F}(f_0,\delta)$. From this, Proposition 8 in the Supplementary Material and the fact that $\delta \in (0, 2^{-5/2}]$, we therefore obtain

$$H_{[]}\bigl(\varepsilon/2^{1/2}, \mathcal{F}(f_0,\delta), d_{\mathrm{H}}, [0, 4\delta^2]\bigr) \lesssim \frac{e^4(4\delta^2)^{1/4}}{\varepsilon^{1/2}} \lesssim \frac{\delta^{1/2}}{\varepsilon^{1/2}}, \tag{30}$$

which takes care of the first term in (29). For the second term, let $\eta_j := 4\delta^2 2^j$ for $j = 0, 1, \ldots, l$ where l is the largest integer for which $4\delta^2 2^l < 1/2$. Also let $\eta_{l+1} := 1/2$. By Lemma 5 in the Supplementary Material, for every $f \in \mathcal{F}(f_0,\delta)$ and $j = 0, 1, \ldots, l$, we have that

$$-\frac{4\delta}{\eta_j^{1/2}} \le \log f(x) \le 2^{13/2}\delta \le 16 \qquad \text{for every } x \in [\eta_j, \eta_{j+1}].$$

Set $\varepsilon_j := \varepsilon/(2l+2)^{1/2}$. Then by Proposition 7 in the Supplementary Material,

$$\begin{aligned}
H_{[]}\bigl(\varepsilon/2^{1/2}, \mathcal{F}(f_0,\delta), d_{\mathrm{H}}, [4\delta^2, 1/2]\bigr) &\le \sum_{j=0}^{l} H_{[]}\bigl(\varepsilon_j, \mathcal{F}(f_0,\delta), d_{\mathrm{H}}, [\eta_j, \eta_{j+1}]\bigr) \\
&\lesssim \sum_{j=0}^{l} \biggl(2^{13/2}\delta + \frac{4\delta}{\eta_j^{1/2}}\biggr)^{1/2} \frac{e^4(\eta_{j+1}-\eta_j)^{1/4}}{\varepsilon_j^{1/2}} \\
&\lesssim \delta^{1/2} \sum_{j=0}^{l} \frac{1}{\varepsilon_j^{1/2}} \biggl(\frac{\eta_{j+1}-\eta_j}{\eta_j}\biggr)^{1/4},
\end{aligned}$$

where we have used the fact that $\eta_j \le 1$ in the final inequality. Observe now that by our choice of $\eta_j = 4\delta^2 2^j$ for $j = 0, 1, \ldots, l$ and $\eta_{l+1} = 1/2 \le 4\delta^2 2^{l+1}$, it follows that $\eta_{j+1} - \eta_j \le \eta_j$ for every $j = 0, 1, \ldots, l$. We therefore obtain

$$\begin{aligned}
H_{[]}\bigl(\varepsilon/2^{1/2}, \mathcal{F}(f_0,\delta), d_{\mathrm{H}}, [4\delta^2, 1/2]\bigr) &\lesssim \delta^{1/2} \sum_{j=0}^{l} \frac{1}{\varepsilon_j^{1/2}} \lesssim \frac{\delta^{1/2}}{\varepsilon^{1/2}} (l+1)^{5/4} \\
&\lesssim \frac{\delta^{1/2}}{\varepsilon^{1/2}} \log^{5/4}\biggl(\frac{1}{\delta}\biggr),
\end{aligned}$$

as required, where the final inequality follows because $4\delta^2 2^l < 1/2$, so

$$l + 1 < \frac{-\log(4\delta^2)}{\log 2} \lesssim \log\biggl(\frac{1}{\delta}\biggr).$$

The exponential conditioned on [0, 1] *case*: Now suppose $f_0 = f_{-\alpha,0,1}$ for some $\alpha \in (0, 18)$, let $C_\alpha := \alpha(1 - e^{-\alpha})^{-1}$ and again fix $\delta \in (0, 2^{-5/2}]$. For every $f =$

$e^{\phi} \in \mathcal{F}(f_0, \delta)$, we have

$$\begin{aligned}\delta^2 &\geq \int_0^1 (e^{\phi(x)/2} - C_\alpha^{1/2} e^{-\alpha x/2})^2 \, dx \\ &= C_\alpha \int_0^1 e^{-\alpha x} \biggl(\frac{1}{C_\alpha^{1/2}} e^{\{\phi(x)+\alpha x\}/2} - 1 \biggr)^2 \, dx \\ &\geq C_\alpha e^{-\alpha} \int_0^1 \biggl(\frac{1}{C_\alpha^{1/2}} e^{\{\phi(x)+\alpha x\}/2} - 1 \biggr)^2 \, dx.\end{aligned}$$

Write $\tilde{\delta} := \delta e^{\alpha/2}/C_\alpha^{1/2}$, so that

$$\delta \leq \tilde{\delta} \leq \biggl(\frac{e^{18} - 1}{18} \biggr)^{1/2} \delta.$$

Thus, arguing as for the uniform case, given $\delta \in (0, (\frac{18}{e^{18}-1})^{1/2} 2^{-5/2}]$ and $\varepsilon > 0$, we can find an $\varepsilon/C_\alpha^{1/2}$-Hellinger bracketing set $\{[g_{L,j}, g_{U,j}], j = 1, \ldots, N\}$ for the class $\{x \mapsto C_\alpha^{-1} f(x) e^{\alpha x} : f \in \mathcal{F}(f_0, \delta)\}$ with

$$\log N \lesssim \log^{5/4}\biggl(\frac{1}{\tilde{\delta}} \biggr) \frac{\tilde{\delta}^{1/2} C_\alpha^{1/4}}{\varepsilon^{1/2}} \lesssim \log^{5/4}\biggl(\frac{1}{\delta} \biggr) \frac{\delta^{1/2}}{\varepsilon^{1/2}}.$$

Now let $f_{L,j}(x) := C_\alpha g_{L,j}(x) e^{-\alpha x}$ and $f_{U,j}(x) := C_\alpha g_{U,j}(x) e^{-\alpha x}$ for $j = 1, \ldots, N$. Then

$$\begin{aligned}\int_0^1 (f_{U,j}^{1/2} - f_{L,j}^{1/2})^2 &= C_\alpha \int_0^1 e^{-\alpha x} \{ g_{U,j}^{1/2}(x) - g_{L,j}^{1/2}(x) \}^2 \, dx \\ &\leq C_\alpha \int_0^1 (g_{U,j}^{1/2} - g_{L,j}^{1/2})^2 \leq \varepsilon^2,\end{aligned}$$

so $\{[f_{L,j}, f_{U,j}], j = 1, \ldots, N\}$ form an ε-Hellinger bracketing set for $\mathcal{F}(f_0, \delta)$, as required.

The case where f_0 is truncated exponential:

Now suppose that $f_0 = f_{-1,0,a}$ for some $a \in [18, \infty]$. Given a function $\phi : \mathbb{R} \to [-\infty, \infty)$, we define $\tilde{\phi}_a : \mathbb{R} \to [-\infty, \infty)$ by

$$\tilde{\phi}_a(x) := \phi(x) + x + \log\bigl(1 - e^{-a}\bigr). \tag{31}$$

Let x_0 be defined as in the statement of Lemma 6 in the Supplementary Material and assume that $\delta \leq \kappa := e^{-9}/8$, so that $x_0 \geq 17$. Also let $l = \lfloor x_0 \rfloor$ and $J := \sup\{j \in \mathbb{N} : x_0 + j - l - 1 \leq a\}$. We define subintervals of $[0, a]$ (or $[0, a)$ when $a = \infty$) by

$$\begin{aligned}&S_1 := [0, 1], \\ &S_j := \bigl[j - 1, \min(j, x_0)\bigr] \qquad \text{for } j = 2, \ldots, l + 1, \\ &S_j := \bigl[x_0 + j - l - 2, \min(x_0 + j - l - 1, a)\bigr] \qquad \text{for } j = l + 2, \ldots, J + 1.\end{aligned}$$

Also let

$$\varepsilon_j^2 := \begin{cases} \frac{2(1-e^{-18})}{3}\varepsilon^2 & \text{for } j = 1, \\ \frac{2(1-e^{-18})}{3}\frac{e^{j-1}\varepsilon^2}{l} & \text{for } j = 2, \dots, l+1, \\ \frac{2(1-e^{-18})}{3}e^{x_0+j-l-2}\varepsilon^2 u_j^2 & \text{for } j = l+2, \dots, J+1, \end{cases} \tag{32}$$

where (u_j) is a sequence with $\sum_{j=l+2}^{J+1} u_j^2 \leq 1$ to be specified later. Applying Lemma 9 in the Supplementary Material with $\mathcal{G} := \{\exp(\tilde{\phi}_a) : \exp(\phi) \in \mathcal{F}(f_0, \delta)\}$, we obtain

$$H_{[]}\big(2^{1/2}\varepsilon, \mathcal{F}(f_0, \delta), d_{\mathrm{H}}, [0, a)\big) \leq \sum_{j=1}^{J+1} H_{[]}(\varepsilon_j, \mathcal{G}, d_{\mathrm{H}}, S_j). \tag{33}$$

We now break the right-hand side of (33) into the three parts:

$$\begin{aligned} H_1 &:= H_{[]}(\varepsilon_1, \mathcal{G}, d_{\mathrm{H}}, S_1), \\ H_2 &:= \sum_{j=2}^{l+1} H_{[]}(\varepsilon_j, \mathcal{G}, d_{\mathrm{H}}, S_j), \\ H_3 &:= \sum_{j=l+2}^{J+1} H_{[]}(\varepsilon_j, \mathcal{G}, d_{\mathrm{H}}, S_j), \end{aligned}$$

and bound each of them below separately. For H_1, we have for every $f = e^{\phi} \in \mathcal{F}(f_0, \delta)$ that

$$\begin{aligned} \delta^2 &\geq \int_0^1 (f^{1/2} - f_0^{1/2})^2 = \int_0^1 (e^{\tilde{\phi}_a(x)/2} - 1)^2 \frac{e^{-x}}{1-e^{-a}}\, dx \\ &\geq \frac{e^{-1}}{1-e^{-a}} \int_0^1 (e^{\tilde{\phi}_a(x)/2} - 1)^2\, dx. \end{aligned}$$

Thus, arguing as for the uniform case, since $\delta e^{1/2}(1-e^{-a})^{1/2} \leq \kappa e^{1/2} \leq 2^{-5/2}$,

$$H_1 \lesssim \log^{5/4}\left(\frac{1}{\delta e^{1/2}(1-e^{-a})^{1/2}}\right)\frac{\delta^{1/2}}{\varepsilon^{1/2}} \lesssim \log^{5/4}\left(\frac{1}{\delta}\right)\frac{\delta^{1/2}}{\varepsilon^{1/2}}.$$

We next bound H_2. Note that $\bigcup_{j=2}^{l+1} S_j \subseteq [1, x_0]$. We can therefore apply Lemma 6 in the Supplementary Material to deduce that whenever $e^{\phi} \in \mathcal{F}(f_0, \delta)$ and $x \in S_j$,

$$\big|\tilde{\phi}_a(x)\big| \leq C e^{x/2}(1-e^{-a})^{1/2}\delta \leq C e^{j/2}\delta.$$

An application of Proposition 7 in the Supplementary Material therefore gives, for $j = 2, \dots, l+1$, that

$$H_{[]}(\varepsilon_j, \mathcal{G}, d_{\mathrm{H}}, S_j) \le \frac{Ce^{j/4}\delta^{1/2}}{\varepsilon_j^{1/2}} \exp(Ce^{j/2}\delta) \le Cl^{1/4}\frac{\delta^{1/2}}{\varepsilon^{1/2}} \exp(Ce^{j/2}\delta)$$
$$\lesssim l^{1/4}\frac{\delta^{1/2}}{\varepsilon^{1/2}},$$

where the final inequality follows because

$$e^{j/2} \le e^{1/2}e^{x_0/2} \le \left\{\frac{1}{2^6(1-e^{-18})}\right\}^{1/2}\delta^{-1}.$$

We therefore obtain

$$H_2 = \sum_{j=2}^{l+1} H_{[]}(\varepsilon_j, \mathcal{G}, d_{\mathrm{H}}, S_j) \le Cl^{5/4}\frac{\delta^{1/2}}{\varepsilon^{1/2}}$$
$$\le Cx_0^{5/4}\frac{\delta^{1/2}}{\varepsilon^{1/2}} \le C\log^{5/4}\left(\frac{1}{2^6e\delta^2(1-e^{-18})}\right)\frac{\delta^{1/2}}{\varepsilon^{1/2}} \lesssim \log^{5/4}\left(\frac{1}{\delta}\right)\frac{\delta^{1/2}}{\varepsilon^{1/2}}.$$

We next turn to H_3, where we consider two cases. First, suppose that $x_0 = a - 1$ so that $J = l + 2$ and $S_{l+2} = [a-1, a]$. This means that $H_3 = N_{[]}(\varepsilon_{l+2}, \mathcal{G}, d_{\mathrm{H}}, [a-1, a])$. We take $u_{l+2} = 1$ in the definition of ε_{l+2} in (32). From the definition of x_0 in (15) in the Supplementary Material, we find that

$$a = 1 + x_0 \le 1 + \log\frac{1}{2^6e\delta^2(1-e^{-a})} = \log\frac{1}{2^6\delta^2(1-e^{-a})}. \tag{34}$$

For every $f = e^{\phi} \in \mathcal{F}(f_0, \delta)$, we can write

$$\delta^2 \ge \int_{a-1}^{a}(f^{1/2} - f_0^{1/2})^2 = \int_{a-1}^{a}(e^{\tilde{\phi}_a(x)/2} - 1)^2\frac{e^{-x}}{1-e^{-a}}\,dx$$
$$\ge \frac{e^{-a}}{1-e^{-a}}\int_{a-1}^{a}(e^{\tilde{\phi}_a/2} - 1)^2.$$

Now $\delta e^{a/2}(1-e^{-a})^{1/2} \le 2^{-5/2}$ from (34), and it follows again by the same argument as in the uniform case that

$$H_3 \lesssim \log^{5/4}\left(\frac{1}{\delta e^{a/2}(1-e^{-a})^{1/2}}\right)\frac{\delta^{1/2}e^{a/4}(1-e^{-a})^{1/4}}{\varepsilon_{l+2}^{1/2}}$$
$$\lesssim \log^{5/4}\left(\frac{1}{\delta}\right)\frac{\delta^{1/2}}{\varepsilon^{1/2}}.$$

Now suppose that $x_0 < a - 1$, so that

$$x_0 = \log\frac{1}{2^6e\delta^2(1-e^{-a})}.$$

For every $j \in \{l+2, \dots, J+1\}$, every $x \in S_j$ and every $f = e^{\phi} \in \mathcal{F}(f_0, \delta)$, it follows from Lemma 6 in the Supplementary Material that

$$\tilde{\phi}_a(x) \le \frac{8(x - x_0)}{x_0 - 1} + 7 \le \frac{8(j - l - 1)}{x_0 - 1} + 7 \le \frac{8(j - x_0)}{x_0 - 1} + 7.$$

Let $u_j := ce^{-(j-x_0)/8}$ in (32), where the universal constant $c > 0$ is chosen such that $\sum_{j=l+2}^{\infty} u_j^2 \le 1$. Then by Proposition 8 in the Supplementary Material, for $j = l+2, \dots, J+1$,

$$\begin{aligned} H_{[]}(\varepsilon_j, \mathcal{G}, d_{\mathrm{H}}, S_j) &\le C \frac{e^{2(j-x_0)/(x_0-1)}}{\varepsilon_j^{1/2}} \\ &\lesssim \frac{1}{\varepsilon^{1/2}} \exp\left\{ \frac{2(j - x_0)}{x_0 - 1} - \frac{x_0 + j - l - 2}{4} + \frac{j - x_0}{16} \right\} \\ &\lesssim \frac{1}{\varepsilon^{1/2}} \exp\left\{ \frac{2(j - x_0)}{x_0 - 1} - \frac{j - 2}{4} + \frac{j - x_0}{16} \right\}. \end{aligned}$$

Hence,

$$\begin{aligned} H_3 &\lesssim \frac{1}{\varepsilon^{1/2}} \sum_{j=l+2}^{J+1} \exp\left\{ \frac{2(j - x_0)}{x_0 - 1} - \frac{j - x_0 + x_0 - 2}{4} + \frac{j - x_0}{16} \right\} \\ &\lesssim \frac{e^{-x_0/4}}{\varepsilon^{1/2}} \sum_{r=1}^{\infty} \exp\left\{ -r\left(\frac{3}{16} - \frac{2}{x_0 - 1} \right) \right\} \lesssim \frac{e^{-x_0/4}}{\varepsilon^{1/2}} \lesssim \frac{\delta^{1/2}}{\varepsilon^{1/2}}. \end{aligned} \tag{35}$$

This completes the proof of Theorem 4 in the case where $\upsilon = 0$. We can now treat the case of general $\upsilon \in [0, 2^{1/2}]$ as follows. Fix $f_0 \in \mathcal{F}$, $\varepsilon > 0$, let κ be as above and let $\delta \in (0, \kappa - \upsilon)$. Also let $\eta \in (0, \kappa - \upsilon - \delta)$ and $f_1 \in \mathcal{F}^1$ be such that $f_0 \ll f_1$ and

$$d_{\mathrm{H}}(f_0, f_1) \le \upsilon + \eta < \kappa - \delta. \tag{36}$$

Then by the triangle inequality, $\mathcal{F}(f_0, \delta) \subseteq \mathcal{F}(f_1, d_{\mathrm{H}}(f_0, f_1) + \delta)$, so that the result in the case $\upsilon = 0$ and (36) give

$$\begin{aligned} H_{[]}\big(2^{1/2}\varepsilon, \mathcal{F}(f_0, \delta), d_{\mathrm{H}}\big) &\le C \log^{5/4}\left(\frac{1}{\delta}\right) \frac{\{\delta + d_{\mathrm{H}}(f_0, f_1)\}^{1/2}}{\varepsilon^{1/2}} \\ &\le C \log^{5/4}\left(\frac{1}{\delta}\right) \frac{(\delta + \upsilon + \eta)^{1/2}}{\varepsilon^{1/2}}. \end{aligned}$$

Since $\eta \in (0, \kappa - \upsilon - \delta)$ was arbitrary, the result follows. □

PROOF OF THEOREM 5. For $\xi \ge 0$ and $\eta \in (0, 1)$, let

$$\tilde{\mathcal{F}}^{\xi,\eta} := \{ f \in \mathcal{F} : |\mu_f| \le \xi, |\sigma_f^2 - 1| \le \eta \}.$$

Since the log-concave maximum likelihood estimator is affine equivariant [Dümbgen, Samworth and Schuhmacher (2011), Remark 2.4] and since the Hellinger distance between densities is affine invariant, we may assume without loss of generality that $f_0 \in \mathcal{F}^{0,1}$. By Kim and Samworth [(2016), Lemma 6] there exist universal constants $\eta \in (0,1)$, $C'' > 0$ such that

$$\sup_{f_0 \in \mathcal{F}^{0,1}} \mathbb{P}_{f_0}\bigl(\hat{f}_n \notin \tilde{\mathcal{F}}^{1,\eta}\bigr) \leq \frac{C''}{n}. \tag{37}$$

For notational convenience, we write

$$\upsilon := \inf_{f_1 \in \mathcal{F}^1 : f_0 \ll f_1} d_{\mathrm{H}}(f_0, f_1),$$

and initially consider the case $\upsilon \leq \kappa/2$, where κ is taken from Theorem 4. From Theorem 4, we find that

$$\int_0^\delta H_{[]}^{1/2}\bigl(\varepsilon, \mathcal{F}(f_0,\delta) \cap \tilde{\mathcal{F}}^{1,\eta}, d_{\mathrm{H}}\bigr)\, d\varepsilon \leq C\delta^{3/4}(\delta+\upsilon)^{1/4}\log^{5/8}(1/\delta),$$

provided $\delta \leq \kappa - \upsilon$. For $\delta > \kappa - \upsilon$, we have

$$\begin{aligned} H_{[]}\bigl(\varepsilon, \mathcal{F}(f_0,\delta) \cap \tilde{\mathcal{F}}^{1,\eta}, d_{\mathrm{H}}\bigr) \leq H_{[]}\bigl(\varepsilon, \tilde{\mathcal{F}}^{1,\eta}, d_{\mathrm{H}}\bigr) \leq C\varepsilon^{-1/2} \leq C\biggl(\frac{\delta}{\kappa-\upsilon}\biggr)^{1/2}\varepsilon^{-1/2} \\ \leq \frac{2^{1/2}C}{\kappa^{1/2}}\biggl(\frac{\delta}{\varepsilon}\biggr)^{1/2}, \end{aligned}$$

where the second inequality follows by Kim and Samworth (2016), Theorem 4. Thus, in this case,

$$\int_0^\delta H_{[]}^{1/2}\bigl(\varepsilon, \mathcal{F}(f_0,\delta) \cap \tilde{\mathcal{F}}^{1,\eta}, d_{\mathrm{H}}\bigr)\, d\varepsilon \leq C\kappa^{-1/4}\delta.$$

We can therefore define

$$\Psi(\delta) := \begin{cases} C\delta^{3/4}(\delta+\upsilon)^{1/4}\log^{5/8}(1/\delta) & \text{if } \delta \leq \kappa - \upsilon, \\ C'\kappa^{-1/4}\delta & \text{if } \delta > \kappa - \upsilon, \end{cases}$$

where the universal constants $C, C' > 0$ are chosen such that

$$\Psi(\delta) \geq \max\biggl\{\int_0^\delta H_{[]}^{1/2}\bigl(2^{1/2}\varepsilon, \mathcal{F}(f_0, 4\delta) \cap \tilde{\mathcal{F}}^{1,\eta}, d_{\mathrm{H}}\bigr)\, d\varepsilon, \delta\biggr\},$$

and such that $\delta \mapsto \delta^{-2}\Psi(\delta)$ is decreasing on $(0,\infty)$. Moreover, we can define $\delta_* := (cn^{-1}\log^{5/4} n + \upsilon^2)^{1/2}$ for some universal constant $c > 0$, so that

$$\inf_{\delta \geq \delta_*} \frac{n^{1/2}\delta^2}{\Psi(\delta)} \geq \frac{n^{1/2}\delta_*^2}{\Psi(\delta_*)} \geq \frac{n^{1/2}\delta_*}{\max(2^{1/4}C, C'\kappa^{-1/4})\max\{1, \log^{5/8}(1/\delta_*)\}}.$$

By choosing the universal constant $c > 0$ sufficiently large, we can ensure that this ratio exceeds the universal constant required to apply the empirical process bound of van de Geer [(2000), Corollary 7.5] (restated as Theorem 10 in the Supplementary Material for convenience). We deduce from this and (37) that

$$
\begin{aligned}
\mathbb{E}_{f_0} d_{\mathrm{X}}^2(\hat{f}_n, f_0) &\leq \int_0^{10\log n} \mathbb{P}\bigl[\{d_{\mathrm{X}}^2(\hat{f}_n, f_0) \geq t\} \cap \{\hat{f}_n \in \tilde{\mathcal{F}}^{1,\eta}\}\bigr]\,dt \\
&\quad + 10 \log n \mathbb{P}(\hat{f}_n \notin \tilde{\mathcal{F}}^{1,\eta}) + \int_{10\log n}^{\infty} \mathbb{P}\{d_{\mathrm{X}}^2(\hat{f}_n, f_0) \geq t\}\,dt \\
&\leq \delta_*^2 + C \int_{\delta_*^2}^{\infty} \exp\biggl(-\frac{nt}{C^2}\biggr)\,dt + \frac{10C'' \log n}{n} \qquad (38) \\
&\quad + \int_{10\log n}^{\infty} \mathbb{P}_{f_0}\biggl\{\max_{1\leq i\leq n} \log \frac{\hat{f}_n(X_i)}{f_0(X_i)} \geq t\biggr\}\,dt \\
&\leq \frac{C''' \log^{5/4} n}{n} + \upsilon^2,
\end{aligned}
$$

for some universal constant $C''' > 0$, where the final inequality follows from (10) in the proof of Lemma 2 in the Supplementary Material.

Now suppose that $\upsilon > \kappa/2$. In that case, a slightly simpler version of the calculation above, which relies only on the global entropy bound $H_{[]}(\varepsilon, \tilde{\mathcal{F}}^{1,\eta}, d_{\mathrm{H}}) \lesssim \varepsilon^{-1/2}$, yields that $\sup_{f_0 \in \mathcal{F}} \mathbb{E}_{f_0} d_{\mathrm{X}}^2(\hat{f}_n, f_0) \leq Cn^{-4/5} \leq \kappa^2/4 < \upsilon^2$ for large n; see also Kim and Samworth (2016), Theorem 5. By increasing the universal constant to deal with smaller values of n if necessary, the result follows. □

Acknowledgements. The authors are very grateful to the anonymous reviewers for their constructive feedback, which helped to improve the paper.

SUPPLEMENTARY MATERIAL

Supplement to "Adaptation in log-concave density estimation" (DOI: 10.1214/17-AOS1619SUPP; .pdf). Auxiliary results

REFERENCES

BARAUD, Y. and BIRGÉ, L. (2016). RHO-estimators for shape restricted estimation. *Stochastic Process. Appl.* **126** 3888–3912. MR3565484

BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical Inference Under Order Restrictions*. Wiley, London.

BIRGÉ, L. (1987). Estimating a density under order restrictions: Nonasymptotic minimax risk. *Ann. Statist.* **15** 995–1012. MR0902241

CHATTERJEE, S., GUNTUBOYINA, A. and SEN, B. (2014). On risk bounds in isotonic and other shape restricted regression problems. *Ann. Statist.* **43** 1774–1800. MR3357878

CHATTERJEE, S., GUNTUBOYINA, A. and SEN, B. (2018). On matrix estimation under monotonicity constraints. *Bernoulli* **24** 1072–1100. MR3706788

CHEN, Y. and SAMWORTH, R. J. (2016). Generalised additive and index models with shape constraints. *J. Roy. Statist. Soc. Ser. B* **78** 729–754.

CHEN, Y. and WELLNER, J. A. (2016). On convex least squares estimation when the truth is linear. *Electron. J. Stat.* **10** 171–209.

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. MR2645484

CULE, M., SAMWORTH, R. and STEWART, M. (2010). Maximum likelihood estimation of a multidimensional log-concave density. *J. Roy. Statist. Soc. Ser. B* **72** 545–607. MR2758237

DOSS, C. and WELLNER, J. A. (2016). Global rates of convergence of the MLEs of log-concave and s-concave densities. *Ann. Statist.* **44** 954–981. MR3485950

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** 40–68. MR2546798

DÜMBGEN, L. and RUFIBACH, K. (2011). logcondens: Computations related to univariate log-concave density estimation. *J. Stat. Softw.* **39** 1–28.

DÜMBGEN, L., RUFIBACH, K. and WELLNER, J. A. (2007). Marshall's lemma for convex density estimation. In *Asymptotics: Particles, Processes and Inverse Problems. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **55** 101–107. IMS, Beachwood, OH. MR2459933

DÜMBGEN, L., SAMWORTH, R. and SCHUHMACHER, D. (2011). Approximation by log-concave distributions, with applications to regression. *Ann. Statist.* **39** 702–730. MR2816336

DVORETZKY, A., KIEFER, J. and WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Stat.* **33** 642–669. MR0083864

GRENANDER, U. (1956). On the theory of mortality measurement. II. *Skand. Aktuarietidskr.* **39** 125–153. MR0093415

GROENEBOOM, P. and JONGBLOED, G. (2014). *Nonparametric Estimation Under Shape Constraints*. Cambridge Univ. Press, Cambridge.

GROENEBOOM, P., JONGBLOED, G. and WELLNER, J. A. (2001). Estimation of a convex function: Characterizations and asymptotic theory. *Ann. Statist.* **29** 1653–1698. MR1891742

HAN, Q. and WELLNER, J. A. (2016). Multivariate convex regression: Global risk bounds and adaptation. Available at http://arxiv.org/abs/1601.06844.

HILDRETH, C. (1954). Point estimates of ordinates of concave functions. *J. Amer. Statist. Assoc.* **49** 598–619. MR0065093

KIM, A. K. H., GUNTUBOYINA, A. and SAMWORTH, R. J. (2018). Supplement to "Adaptation in log-concave density estimation." DOI:10.1214/17-AOS1619SUPP.

KIM, A. K. H. and SAMWORTH, R. J. (2016). Global rates of convergence in log-concave density estimation. *Ann. Statist.* **44** 2756–2779. MR3576560

LIM, E. and GLYNN, P. W. (2012). Consistency of multidimensional convex regression. *Oper. Res.* **60** 196–208.

MARSHALL, A. W. (1970). Discussion of Barlow and van Zwet's paper. In *Nonparametric Techniques in Statistical Inference Proceedings of the First International Symposium on Nonparametric Techniques Held at Indiana University* (M. L. Puri, ed.) 174–176. Cambridge Univ. Press, Cambridge.

MASSART, P. (1990). The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality. *Ann. Probab.* **18** 1269–1283. MR1062069

PAL, J., WOODROOFE, M. and MEYER, M. (2007). Estimating a Polya frequency function. In *Complex Datasets and Inverse Problems: Tomography, Networks and Beyond* (R. Liu, W. Strawderman and C.-H. Zhang, eds.). *IMS Lecture Notes—Monograph Series* **54** 239–249.

SAMWORTH, R. J. and YUAN, M. (2012). Independent component analysis via nonparametric maximum likelihood estimation. *Ann. Statist.* **40** 2973–3002. MR3097966

SCHUHMACHER, D. and DÜMBGEN, L. (2010). Consistency of multivariate log-concave density estimators. *Statist. Probab. Lett.* **80** 376–380.

SEIJO, E. and SEN, B. (2011). Nonparametric least squares estimation of a multivariate convex regression. *Ann. Statist.* **39** 1633–1657. MR2850215

SEREGIN, A. and WELLNER, J. A. (2010). Nonparametric estimation of multivariate convex-transformed densities. *Ann. Statist.* **38** 3751–3781. MR2766867

VAN DE GEER, S. (2000). *Applications of Empirical Process Theory*. Cambridge Univ. Press, Cambridge.

VAN EEDEN, C. (1956). Maximum likelihood estimation of ordered probabilities. In *Indagationes Mathematicae* (*Proceedings*) **59** 444–455. North-Holland, Amsterdam.

WALTHER, G. (2002). Detecting the presence of mixing with multiscale maximum likelihood. *J. Amer. Statist. Assoc.* **97** 508–513. MR1941467

ZHANG, C.-H. (2002). Risk bounds in isotonic regression. *Ann. Statist.* **30** 528–555. MR1902898

A. K. H. KIM
STATISTICAL LABORATORY
UNIVERSITY OF CAMBRIDGE
WILBERFORCE ROAD
CAMBRIDGE
CB3 0WB
UNITED KINGDOM
AND
DEPARTMENT OF STATISTICS
SUNGSHIN WOMEN'S UNIVERSITY
34DA-GIL 2
BOMUN-RO
SEONGBUK-GU
SEOUL 02844
SOUTH KOREA
E-MAIL: a.kim@statslab.cam.ac.uk
arlenekim@sungshin.ac.kr
URL: http://sites.google.com/site/kyoungheearlene/home

A. GUNTUBOYINA
DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA, BERKELEY
423 EVANS HALL
BERKELEY, CALIFORNIA 94720
USA
E-MAIL: aditya@stat.berkeley.edu
URL: http://www.stat.berkeley.edu/~aditya/

R. J. SAMWORTH
STATISTICAL LABORATORY
UNIVERSITY OF CAMBRIDGE
WILBERFORCE ROAD
CAMBRIDGE
CB3 0WB
UNITED KINGDOM
E-MAIL: r.samworth@statslab.cam.ac.uk
URL: http://www.statslab.cam.ac.uk/~rjs57