

High-dimensional nonparametric density estimation via symmetry and shape constraints

Min Xu^{*}and Richard J. Samworth[†]

Wharton Business School, University of Pennsylvania and
Statistical Laboratory, University of Cambridge

September 9, 2017

Abstract

We tackle the problem of high-dimensional nonparametric density estimation by taking the class of log-concave densities on \mathbb{R}^p and incorporating within it a symmetry assumption, which both mitigates the curse of dimensionality and admits a scalable estimation algorithm. The case of spherical symmetry is studied in detail through the notion of spherically symmetric log-concave projection. In the setting of independent and identically distributed data in \mathbb{R}^p , we prove that the spherically symmetric log-concave maximum likelihood estimator has a worst case risk bound with respect to, e.g., squared Hellinger loss, of $O(n^{-4/5})$, independent of p . Moreover, we show that the estimator is adaptive in the sense that if the data generating density admits a special form, then a nearly parametric rate may be attained. Our estimation algorithm is fast even when n and p are on the order of hundreds of thousands, and we illustrate its strong performance on simulated data.

1 Introduction

Density estimation emerged as one of the fundamental challenges in Statistics very soon after its inception as a field. Up until halfway through the last century, approaches based on parametric (often Gaussian) assumptions or histograms/contingency tables were dominant (Fisher, 1922, 1925). However, the restrictions of these techniques

^{*}Email address: minx@wharton.upenn.edu

[†]Email address: r.samworth@statslab.cam.ac.uk

has led, since the 1950s, to an enormous research effort devoted to exploring nonparametric methods, primarily based on smoothness assumptions, but also on shape constraints. These include kernel density estimation (Rosenblatt, 1956; Wand and Jones, 1995), wavelets (Donoho et al., 1996) and other orthogonal series methods, splines (Gu and Qiu, 1993), as well as techniques based on monotonicity (Grenander, 1956), log-concavity (Cule et al., 2010) and others. Although highly successful for low-dimensional data, these approaches all encounter two serious difficulties in moderate- or high-dimensional regimes: first, theoretical performance is limited by minimax lower bounds that characterise the ‘curse of dimensionality’ (e.g. Ibragimov and Khasminskii, 1983); and second, computational issues may become a bottleneck, often exacerbated by the need to choose (multiple) smoothing parameters.

In parallel to these developments, modern technology now allows the routine collection of extremely high-dimensional data sets, leading to a great demand for reliable and scalable density estimation algorithms. To emphasise the challenge here, let \mathcal{F}_p denote the class of upper semi-continuous, log-concave densities on \mathbb{R}^p , and let X_1, \dots, X_n be independent and identically distributed random vectors with density $f_0 \in \mathcal{F}_p$. Kim and Samworth (2016) proved that for each $p \in \mathbb{N}$, there exists $c_p > 0$ such that

$$\inf_{\tilde{f}_n} \sup_{f_0 \in \mathcal{F}_p} \mathbb{E} d_H^2(\tilde{f}_n, f_0) \geq \begin{cases} c_1 n^{-4/5} & \text{if } p = 1 \\ c_p n^{-2/(p+1)} & \text{if } p \geq 2, \end{cases}$$

where $d_H^2(f, g) := \int_{\mathbb{R}^p} (f^{1/2} - g^{1/2})^2$ denotes the squared Hellinger distance between densities f and g on \mathbb{R}^p , and where the infimum is taken over all estimators \tilde{f}_n of f_0 based on X_1, \dots, X_n . This suggests that very large sample sizes would be required for an adequate approximation to the true density, even for $p = 5$. In view of these fundamental theoretical limitations, it is natural to consider imposing additional structure on the problem, while simultaneously seeking to retain the desirable flexibility of the nonparametric paradigm.

In this paper, we propose a new method for high-dimensional, nonparametric density estimation by incorporating symmetry constraints into the shape-constrained class. We demonstrate that this approach facilitates an efficient algorithm that can evade the curse of dimensionality in terms of its rate of convergence. In this preliminary work, we focus on perhaps the simplest symmetry constraint, and consider the class of upper semi-continuous, spherically symmetric log-concave densities $\mathcal{F}_p^{\text{SS}}$. In particular, we study the maximum likelihood estimator (MLE) \hat{f}_n of f_0 over the class $\mathcal{F}_p^{\text{SS}}$. An attractive feature of this estimator is that, despite the potentially high-dimensional nature of the problem, it does not require the choice of any tuning parameters.

Writing Φ for the class of upper semi-continuous, concave and decreasing functions on $[0, \infty)$, it turns out (cf. Proposition 1 below) that the density f_0 belongs to $\mathcal{F}_p^{\text{SS}}$

if and only if $f_0(x) = e^{\phi_0(\|x\|)}$ for some $\phi_0 \in \Phi$. We further note that the univariate random variables $Z_i := \|X_i\|$ for $i = 1, \dots, n$ each have (log-concave) density $h_0(r) := c_p r^{p-1} e^{\phi_0(r)}$, where $c_p := 2\pi^{p/2}/\Gamma(p/2)$. These observations enable us to reduce the problem of computing \hat{f}_n to obtaining the MLE $\hat{\phi}_n$ of ϕ_0 based on Z_1, \dots, Z_n , over the subclass of $\phi \in \Phi$ such that $r \mapsto c_p r^{p-1} e^{\phi(r)}$ is a density.

Our results on the theoretical performance of \hat{f}_n are presented in terms of the divergence measure

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

We show in Proposition 3(iv) below that $d_X^2(\hat{f}_n, f_0) \geq \int_{\mathbb{R}^p} \hat{f}_n \log(\hat{f}_n/f_0) =: d_{\text{KL}}^2(\hat{f}_n, f_0)$, so that our upper bounds on $\mathbb{E}d_X^2(\hat{f}_n, f_0)$ immediately yield the same upper bounds on the expected Kullback–Leibler divergence (as well as the risks in the squared total variation and squared Hellinger distances, for instance). One of our main results (Theorem 8) is that there exists a universal constant $C > 0$ such that

$$\sup_{f_0 \in \mathcal{F}_p^{\text{SS}}} \mathbb{E}d_X^2(\hat{f}_n, f_0) \leq Cn^{-4/5}.$$

Thus, there is no dependence on p in this worst case risk bound, and we can expect good performance even in high dimensions. See Section 6 for empirical verification. We also elucidate the adaptation behaviour of \hat{f}_n . More precisely, for $k \in \mathbb{N}$, we let $\Phi^{(k)}$ denote the set of $\phi \in \Phi$ that are piecewise linear on $\text{dom}(\phi) := \{r : \phi(r) > -\infty\}$, with at most k linear pieces. We also let $\mathcal{H}^{(k)}$ denote the class of densities h of form $h(r) = r^{p-1} e^{\phi(r)}$ for $\phi \in \Phi^{(k)}$ and let $h_0(r) := r^{p-1} e^{\phi_0(r)}$. In Theorem 9, we derive an oracle inequality: writing $\nu_k^2 := 2 \wedge \inf_{h \in \mathcal{H}^{(k)}} d_{\text{KL}}^2(h_0, h)$, there exists a universal constant $C > 0$ such that for every $f_0 \in \mathcal{F}_p^{\text{SS}}$,

$$\mathbb{E}d_X^2(\hat{f}_n, f_0) \leq C \min_{k=1, \dots, n} \left\{ \frac{k^{4/5} \nu_k^{2/5}}{n^{4/5}} \log\left(\frac{en}{k\nu_k}\right) + \frac{k}{n} \log^{5/4}\left(\frac{en}{k}\right) \right\}.$$

This result reveals that \hat{f}_n adapts to densities f_0 that are close to being of the form e^ϕ for some $\phi \in \Phi^{(k)}$ with k not too large; in particular, an almost parametric rate of $\frac{k}{n} \log^{5/4}\left(\frac{en}{k}\right)$ can be attained. In fact, this result is stronger than what is typically referred to as a sharp oracle inequality; see the remark after Theorem 9.

We remark that an alternative to our proposal, which relies only on existing methods, would be to compute the estimator \tilde{f}_n given by

$$\tilde{f}_n(x) := \begin{cases} \tilde{h}_n(\|x\|)/(c_p \|x\|^{p-1}) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0, \end{cases} \quad (1)$$

where \tilde{h}_n is the ordinary log-concave MLE of h_0 based on Z_1, \dots, Z_n . This estimator \tilde{f}_n ignores the special structure of h_0 , but is similarly straightforward to compute

and attains the same worst case rate of convergence as \hat{f}_n (Kim and Samworth, 2016, Theorem 5)¹; this is in fact the minimax rate of convergence over the class of log-concave densities (Kim and Samworth, 2016, Theorem 1). We argue, however, that \hat{f}_n has several advantages over \tilde{f}_n in this context, and list these in roughly decreasing order of importance:

1. The estimator \tilde{f}_n is inconsistent at $x = 0$. Indeed $\tilde{h}_n(x) = 0$ whenever $\|x\| < \min_i Z_i$, and the division by $\|x\|^{p-1}$ in (1) means that the estimator behaves poorly for small $\|x\|$; see Figure 1. By contrast, \hat{f}_n is uniformly consistent over closed Euclidean balls contained in the interior of the support of f_0 (Proposition 4);
2. As mentioned in our discussion above, the estimator \hat{f}_n attains faster rates of convergence when the true density has a simple structure;
3. The estimator \hat{f}_n takes values in the relevant class $\mathcal{F}_p^{\text{SS}}$, whereas \tilde{f}_n does not;
4. The estimator \hat{f}_n exists in slightly greater generality than \tilde{f}_n (cf. the remark following Proposition 2(iv)).

The differences between \hat{f}_n and \tilde{f}_n are illustrated in Figure 1.

Central to our approach is the notion of spherically symmetric log-concave projection, whose existence, uniqueness and properties we study in detail in Section 2. Significant additional technical challenges, relative to the study of the ordinary log-concave MLE, are caused by the fact that this projection does not have the same moment preservation properties. In Section 3, we therefore develop new arguments to understand how moments of the underlying distribution change under projection; these are crucial to prove that, with high probability, $\hat{\phi}_n$ belongs to a class with sufficiently small global and local bracketing entropy bounds (Sections 8.2 and 8.3) for our risk bounds to hold. Our algorithm for computing \hat{f}_n , presented in Section 5, is based on an Active Set approach, and is a variant of that proposed by Dümbgen et al. (2007) for computing the ordinary log-concave MLE in the univariate case. Its algorithmic complexity only depends on p through the need to compute Z_1, \dots, Z_n at the outset, and therefore scales extremely well to high-dimensional settings.

In addition to the work mentioned above on estimation over the class \mathcal{F}_p of log-concave densities on \mathbb{R}^p , multivariate shape-constrained density estimation has also been considered over other classes, including the set of block decreasing densities on $[0, 1]^p$ (Polonik, 1995, 1998; Biau and Devroye, 2003; Gao and Wellner, 2007), and the class of s -concave densities on \mathbb{R}^p (Doss and Wellner, 2016; Han and Wellner, 2016). In the former case, for uniformly bounded densities, Biau and Devroye (2003) established

¹Although 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).

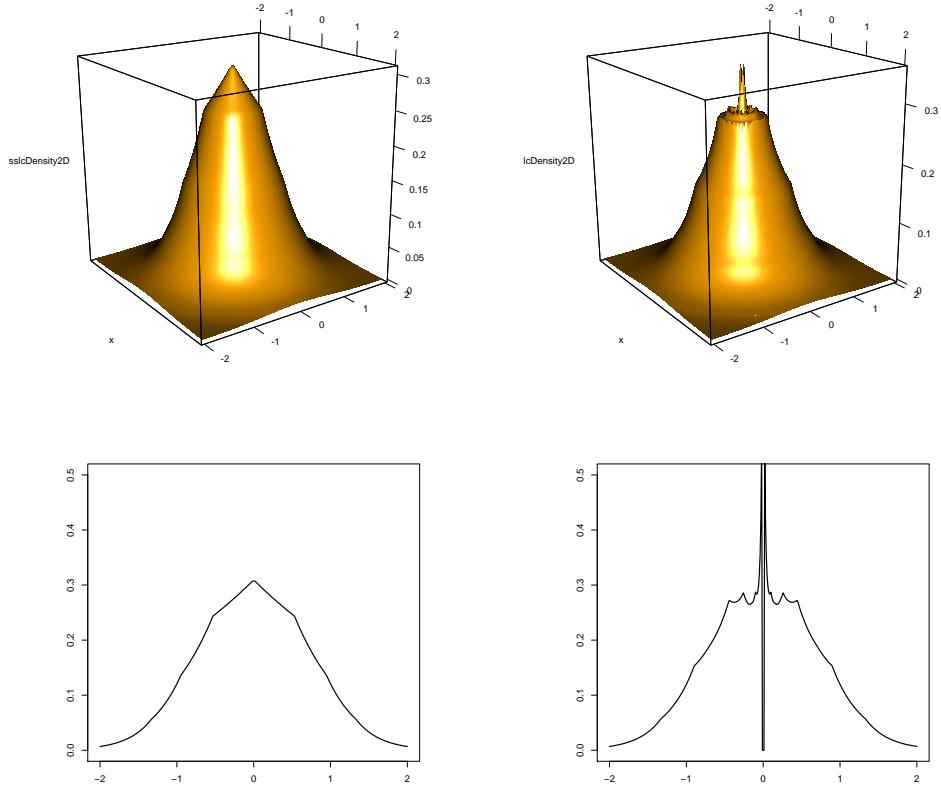


Figure 1: A comparison of the spherically-symmetric log-concave MLE \hat{f}_n (left) and the ordinary log-concave MLE \tilde{f}_n (right) based on a sample of size $n = 1000$ from a standard bivariate normal distribution. The top plots give the bivariate density estimates, while the bottom ones present the corresponding estimates $e^{\hat{\phi}_n}$ and $e^{\tilde{\phi}_n}$ of e^{ϕ_0} .

a minimax lower bound in total variation distance of order $n^{-1/(p+2)}$, revealing a similar curse of dimensionality, while in the latter case, the main interest has been in the classes with $s < 0$, which contain the class \mathcal{F}_p , so the same minimax lower bounds apply as for \mathcal{F}_p . Various other simplifying structures and methods have also been considered for nonparametric high-dimensional density estimation, including kernel approaches for forest density estimation (Liu et al., 2011) and star-shaped density estimation (Liebscher and Richter, 2016), as well as nonparametric maximum likelihood methods for independent component analysis (Samworth and Yuan, 2012). Perhaps most closely related to this work is the approach of Bhattacharya and Bickel (2012), who consider a maximum likelihood approach (as well as spline approximations) to estimating the generator of an elliptically symmetric distribution with decreasing generator.

We conclude the introduction with some notation used throughout the paper. We

use $\|\cdot\|$ to denote the ℓ_2 norm of a vector and $\|\cdot\|_\infty$ for the supremum norm of a function. For $x \in \mathbb{R}$, we write $x_+ := \max(x, 0)$ and $x_- := \max(-x, 0)$. We use $\lambda(\cdot)$ to denote Lebesgue measure on Euclidean space. Given a sample Z_1, \dots, Z_n , we write $Z_{(1)} \leq \dots \leq Z_{(n)}$ for the corresponding order statistics. If h and h' are log-concave functions defined on a subset of \mathbb{R} , we write $h \ll h'$ if $\text{dom}(\log h) \subseteq \text{dom}(\log h')$. Let $S \subseteq \mathbb{R}$, and let \mathcal{G} be a class of non-negative functions whose domains include S . For $\epsilon > 0$, let $N_{[]}(\epsilon, \mathcal{G}, d_H, S)$ 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

$$\int_S (g_{U,j}^{1/2} - g_{L,j}^{1/2})^2 \leq \epsilon^2$$

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

2 Spherically symmetric log-concave densities and projections

Recall from the introduction that $\mathcal{F}_p^{\text{SS}}$ denotes the class of upper semi-continuous, spherically symmetric log-concave densities on \mathbb{R}^p and Φ denotes the class of upper semi-continuous, decreasing and concave functions on $[0, \infty)$. The following basic characterisation of $\mathcal{F}_p^{\text{SS}}$ underpins our methodological and theoretical development.

Proposition 1. *Let f be a density on \mathbb{R}^p . Then $f \in \mathcal{F}_p^{\text{SS}}$ if and only if we can write $f(x) = e^{\phi(\|x\|)}$ for some $\phi \in \Phi$.*

For any fixed $a \geq 0$, let Φ_a be the set of upper semi-continuous functions $\phi : [a, \infty) \rightarrow [-\infty, \infty)$ such that ϕ is decreasing and concave, and set $\mathcal{H}_a := \{r \mapsto r^{p-1}e^{\phi(r)} : \phi \in \Phi_a, \int_a^\infty r^{p-1}e^{\phi(r)} dr = 1\}$. We continue to write $\Phi = \Phi_0$ and also write $\mathcal{H} = \mathcal{H}_0$ as shorthand.

Let P be a probability measure on \mathbb{R}^p . We define

$$\tilde{L}(\phi, P) := \int_{\mathbb{R}^p} \phi(\|x\|) dP(x) - \int_0^\infty r^{p-1}e^{\phi(r)} dr + 1, \quad (2)$$

and write $\tilde{\phi}^*(P) := \text{argmax}_{\phi \in \Phi} \tilde{L}(\phi, P)$. Since $\tilde{L}(\cdot, P)$ is strictly concave, any maximiser of $\tilde{L}(\cdot, P)$ over Φ is unique. If $\tilde{L}(\phi, P) \in \mathbb{R}$, then

$$\frac{\partial}{\partial c} \tilde{L}(\phi + c, P) = 1 - e^c \int_0^\infty r^{p-1}e^{\phi(r)} dr,$$

so $\tilde{L}(\phi + c, P)$ is maximised by choosing $c = -\log(\int_0^\infty r^{p-1} e^{\phi(r)} dr)$. It follows that if $\tilde{\phi}^*(P)$ exists, and $\tilde{L}(\tilde{\phi}^*, P) \in \mathbb{R}$, then we can define the *spherically symmetric log-concave projection* $f^*(P) \in \mathcal{F}_p^{\text{SS}}$ by $f^*(P)(x) := e^{\tilde{\phi}^*(P)(\|x\|)}$.

In fact, in order to study $\tilde{\phi}^*(P)$, it will be convenient also to define a related projection of a one-dimensional probability distribution. To this end, for $a \geq 0$, a probability measure Q on $[a, \infty)$ and $\phi \in \Phi_a$, let

$$L(\phi, Q) := \int_{[a, \infty)} \phi dQ - \int_a^\infty r^{p-1} e^{\phi(r)} dr + 1. \quad (3)$$

Here, we incorporate the greater generality of the translation by a in order to facilitate our analysis of the adaptivity properties of the spherically symmetric log-concave MLE in Section 4. Similarly to before, we let $\phi^*(Q) := \operatorname{argmax}_{\phi \in \Phi_a} L(\phi, Q)$, and set $h^*(Q)(r) := r^{p-1} e^{\phi^*(Q)(r)}$. Again, any maximiser of $L(\cdot, Q)$ over Φ_a is unique, and if $\phi^*(Q)$ exists with $L(\phi^*, Q) \in \mathbb{R}$, then, writing $h^*(r) := r^{p-1} e^{\phi^*(Q)(r)}$, we have that $h^* \in \mathcal{H}_a$, so in particular, h^* is a (log-concave) density.

The following proposition gives necessary and sufficient conditions for the spherically symmetric log-concave projection to be well-defined. We write \mathcal{P}_p for the set of probability distributions on \mathbb{R}^p with $\int_{\mathbb{R}^p} \|x\| dP(x) < \infty$ and $P(\{0\}) < 1$. We let \mathcal{Q}_a denote the class of probability measures Q on $[a, \infty)$ with $\int_a^\infty r dQ(r) < \infty$ and $Q(\{a\}) < 1$, and let $\mathcal{Q} := \mathcal{Q}_0$.

Proposition 2. *We have*

- (i) if $\int_a^\infty r dQ(r) = \infty$, then $L(\phi, Q) = -\infty$ for all $\phi \in \Phi_a$;
- (ii) if $Q(\{a\}) = 1$, then $\sup_{\phi \in \Phi_a} L(\phi, Q) = \infty$;
- (iii) if $Q \in \mathcal{Q}_a$, then $\sup_{\phi \in \Phi_a} L(\phi, Q) \in \mathbb{R}$ and ϕ_a^* is a well-defined map from \mathcal{Q}_a to Φ_a ;
- (iv) if P is a probability measure on \mathbb{R}^p and we define a probability measure Q on $[0, \infty)$ by $Q([0, r)) := P(\{x : \|x\| < r\})$, then $\tilde{L}(\phi, P) = L(\phi, Q)$ for every $\phi \in \Phi$. In particular, if $P \in \mathcal{P}_p$, then $Q \in \mathcal{Q}$ and $\tilde{\phi}^*(P) = \phi^*(Q)$.

Remark: From Proposition 2(iv), we see that the conditions on P required for the spherically symmetric log-concave projection to exist, namely $\int_{\mathbb{R}^p} \|x\| dP(x) < \infty$ and $P(\{0\}) < 1$, are weaker than the corresponding conditions for the ordinary log-concave projection to exist, namely $\int_{\mathbb{R}^p} \|x\| dP(x) < \infty$ and $P(H) < 1$ for every hyperplane H ; cf. D\"umbgen et al. (2011, Theorem 2.2).

The following constitute basic properties of $\phi_a^*(Q)$.

Proposition 3. *Let $Q \in \mathcal{Q}_a$, let $\phi^* := \phi_a^*(Q)$ and let $h^*(r) := r^{p-1} e^{\phi^*(r)}$.*

- (i) The projection ϕ_a^* is scale equivariant in the sense that if $\alpha > 0$, and $Q_\alpha \in \mathcal{Q}_{\alpha a}$ is defined by $Q_\alpha([\alpha a, r)) := Q([a, r/\alpha))$ for all $r \in [\alpha a, \infty)$, then $\phi_{\alpha a}^*(Q_\alpha)(r) = \phi_a^*(Q)(r/\alpha) - p \log \alpha$.
- (ii) Let $\Delta : [a, \infty) \rightarrow [-\infty, \infty)$ be a function satisfying the property that there exists $t > 0$ such that $r \mapsto \phi^*(r) + t\Delta(r) \in \Phi_a$. Then

$$\int_{[a, \infty)} \Delta dQ \leq \int_a^\infty \Delta(r) h^*(r) dr.$$

- (iii) $\int_a^\infty r h^*(r) dr \leq \int_{[a, \infty)} r dQ(r)$.
- (iv) For any $h_0 \in \mathcal{H}_a$, we have $\int_a^\infty h^* \log(h^*/h_0) \leq \int_a^\infty \log(h^*/h_0) dQ$.

Remark: Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} P \in \mathcal{P}_p$ with empirical distribution \mathbb{P}_n , and for a measurable set $A \subseteq \mathbb{R}$, let $Q(A) := P(\{x : \|x\| \in A\})$. Let $Z_i := \|X_i\|$ for $i = 1, \dots, n$ and let \mathbb{Q}_n denote the empirical distribution of \mathbb{Q}_n . Writing $\hat{f}_n := f^*(\mathbb{P}_n)$, $f^* := f^*(P)$, $\hat{h}_n := h^*(\mathbb{Q}_n)$ and $h^* := h^*(Q)$, we have by Proposition 3(iv) that

$$\begin{aligned} d_X^2(\hat{f}_n, f^*) &= \int_{\mathbb{R}^p} \log \frac{\hat{f}_n}{f^*} d\mathbb{P}_n = \int_0^\infty \log \frac{\hat{h}_n}{h^*} d\mathbb{Q}_n \geq \int_0^\infty \hat{h}_n \log \frac{\hat{h}_n}{h^*} = \int_0^\infty \hat{f}_n \log \frac{\hat{f}_n}{f^*} \\ &= d_{\text{KL}}^2(\hat{f}_n, f^*). \end{aligned}$$

As a final basic property of our projections, we establish continuity with respect to the Wasserstein distance. Recall that if P, P' are probability measures on a Euclidean space with finite first moments, then the Wasserstein distance between P and P' is defined as

$$d_W(P, P') := \inf_{(X, X') \sim (P, P')} \mathbb{E} \|X - X'\|,$$

where the infimum is taken over all pairs of random vectors X, X' , defined on the same probability space, with $X \sim P$ and $X' \sim P'$. We also recall that if P has a finite first moment, then $d_W(P_n, P) \rightarrow 0$ if and only if both $P_n \xrightarrow{d} P$ and $\int \|x\| dP_n(x) \rightarrow \int \|x\| dP(x)$.

Proposition 4. Suppose that $P \in \mathcal{P}_p$ and $d_W(P_n, P) \rightarrow 0$. Then $\sup_{\phi \in \Phi} L(\phi, P_n) \rightarrow \sup_{\phi \in \Phi} L(\phi, P)$, $f^*(P_n)$ is well-defined for large n and

$$\int_{\mathbb{R}^p} |f^*(P_n)(x) - f^*(P)(x)| dx \rightarrow 0. \quad (4)$$

Moreover, given any compact set K contained in the interior of the support of P , we have $\sup_{x \in K} |f^*(P_n)(x) - f^*(P)(x)| \rightarrow 0$.

Remark: Proposition 4 immediately yields a consistency (and robustness to misspecification) result for the spherically symmetric log-concave MLE. In particular, suppose that $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} P \in \mathcal{P}_p$ with empirical distribution \mathbb{P}_n , and let $\hat{f}_n := f^*(\mathbb{P}_n)$, $f^* := f^*(P)$. Then, by the strong law of large numbers and Varadarajan's theorem (e.g. Dudley, 2002, Theorem 11.4.1), we have $d_W(\mathbb{P}_n, P) \xrightarrow{\text{a.s.}} 0$, so

$$\int_{\mathbb{R}^p} |\hat{f}_n - f^*| \xrightarrow{\text{a.s.}} 0.$$

Remark: In fact, the conclusion of Proposition 4 also holds for stronger norms than the total variation norm. In particular, taking $a_0 > 0$ and $b_0 \in \mathbb{R}$ such that $f^*(P)(x) \leq e^{-a_0\|x\|+b_0}$ for all $x \in \mathbb{R}^p$, we have by, e.g., Cule and Samworth (2010, Proposition 2) that for every $a < a_0$,

$$\int_{\mathbb{R}^p} e^{a\|x\|} |f^*(P_n)(x) - f^*(P)(x)| dx \rightarrow 0.$$

3 Moment preservation properties

The aim of this section is to elucidate the way in which the first two moments of the empirical distribution \mathbb{Q}_n of a set of n data points in $[0, \infty)$ change under the projection h_a^* . These results enable us to show that if the data are drawn independently from a common distribution on $[0, \infty)$, then with high probability, the first two moments of $\hat{h}_n := h_a^*(\mathbb{Q}_n)$ are close to their population analogues.

For $f \in \mathcal{F}_1$, define $\mu_f := \int_{-\infty}^{\infty} rf(r) dr$ and

$$\sigma_f^2 := \int_{-\infty}^{\infty} r^2 f(r) dr - \mu_f^2.$$

Our first proposition concerns bounds on $\mu_{\hat{h}_n}$:

Proposition 5. *Fix $a \geq 0$, and suppose that Z_1, \dots, Z_n are real numbers in the interval $[a, \infty)$ that are not all equal to a . Let \mathbb{Q}_n be the empirical distribution corresponding to Z_1, \dots, Z_n . Let $\hat{h}_n := h_a^*(\mathbb{Q}_n)$. Then, writing $\bar{Z} := n^{-1} \sum_{i=1}^n Z_i$, there exists $s \in (0, \infty)$ such that*

$$\bar{Z} - \min\left\{\frac{\bar{Z} - a}{\rho(s)}, \frac{s}{\rho(s)}\right\} \leq \mu_{\hat{h}_n} \leq \bar{Z}.$$

We now study bounds for $\sigma_{\hat{h}_n}$, and their consequences for $\sup_{r \geq a} \log \hat{h}_n(r)$.

Proposition 6. *Let $a \geq 0$ and let \mathbb{Q}_n be an empirical distribution of n points in $[a, \infty)$ that are not all equal to a . Let $\delta_{\mathbb{Q}_n} > 0$ be such that $\mathbb{Q}_n(A) \leq 1/2$ for every interval A of length at most $\delta_{\mathbb{Q}_n}$. Writing $\hat{h}_n := h_a^*(\mathbb{Q}_n)$, let $\ell_{\hat{h}_n} \leq \int_a^{\infty} \log \hat{h}_n d\mathbb{Q}_n$. Then there exists a universal constant $C_{\sigma} > 0$ such that*

$$\frac{1}{C_{\sigma}} \min(\delta_{\mathbb{Q}_n}^2, e^{\ell_{\hat{h}_n}}, 1) \leq \sigma_{\hat{h}_n} \leq C_{\sigma} e^{-\ell_{\hat{h}_n}}.$$

Moreover,

$$\sup_{r \geq a} \log \hat{h}_n(r) \leq \max \left\{ -2 \log \left(\frac{\delta_{\mathbb{Q}_n}}{6} \right), -2\ell_{\hat{h}_n}, 3 \right\}. \quad (5)$$

We are now in a position to argue that, with high probability, \hat{h}_n belongs to a subclass of \mathcal{H}_a with restricted first two moments. These moment restrictions are important for enabling us to obtain the bracketing entropy bounds that drive the rates of convergence of the spherically symmetric log-concave MLE. For $C_\mu, C_\sigma > 0$, $a_0 \geq 0$ and $h_0 \in \mathcal{H}_{a_0}$ with $\sigma_{h_0} = 1$, let

$$\mathcal{H}(h_0, C_\mu, C_\sigma) := \left\{ h \in \mathcal{H}_{a_0} : |\mu_h - \mu_{h_0}| \leq C_\mu, \frac{1}{C_\sigma} \leq \sigma_h \leq C_\sigma \right\}.$$

Proposition 7. *Let $a_0 \geq 0$, fix a density $h_0 \in \mathcal{H}_{a_0}$ with $\sigma_{h_0} = 1$, and suppose that $Z_1, \dots, Z_n \stackrel{\text{iid}}{\sim} h_0$, with empirical distribution \mathbb{Q}_n . Writing $\hat{h}_n := h_{a_0}^*(\mathbb{Q}_n)$, there exist universal constants $C_\mu, C_\sigma, C > 0$ such that*

$$\mathbb{P}(\hat{h}_n \notin \mathcal{H}(h_0, C_\mu, C_\sigma)) \leq \frac{C}{n}.$$

4 Worst case and adaptive risk bounds

Let $f_0 \in \mathcal{F}_p^{\text{SS}}$, and suppose that $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} f_0$ with empirical distribution \mathbb{P}_n . Let $\hat{f}_n := f^*(\mathbb{P}_n)$ be the spherically symmetric log-concave MLE. Our first main result below provides a worst case risk bound for \hat{f}_n as an estimator of f_0 in terms of the d_X^2 measure defined in the introduction.

Theorem 8. *Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} f_0 \in \mathcal{F}_p^{\text{SS}}$ with empirical distribution \mathbb{P}_n . Let $\hat{f}_n := f^*(\mathbb{P}_n)$ be the spherically symmetric log-concave MLE. There exists a universal constant $C > 0$ such that*

$$\sup_{f_0 \in \mathcal{F}_p^{\text{SS}}} \mathbb{E} d_X^2(\hat{f}_n, f_0) \leq \frac{C}{n^{4/5}}.$$

As mentioned in the introduction, the attractive aspect of this bound is that it does not depend on p . The proof relies heavily on the special moment preservation properties of the spherically symmetric log-concave MLE.

We now turn to the adaptation properties of \hat{f}_n . For $k \in \mathbb{N}$ and $a > 0$, we say $\phi \in \Phi_a$ is k -affine, and write $\phi \in \Phi_a^{(k)}$, if there exist $r_0 \in (a, \infty]$ and a partition I_1, \dots, I_k of $[a, r_0)$ into intervals such that ϕ is affine on each I_j for $j = 1, \dots, k$, and $\phi(r) = -\infty$ for $r > r_0$. Define $\mathcal{H}_a^{(k)} := \{h \in \mathcal{H}_a : h(r) = r^{p-1} e^{\phi(r)} \text{ for some } \phi \in \Phi_a^{(k)}\}$. Again, we write $\Phi^{(k)} := \Phi_0^{(k)}$ and $\mathcal{H}^{(k)} := \mathcal{H}_0^{(k)}$.

Theorem 9. Let $f_0 \in \mathcal{F}_p^{\text{SS}}$ be given by $f_0(x) = e^{\phi_0(\|x\|)}$ and let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} f_0$ with empirical distribution \mathbb{P}_n . Let $\hat{f}_n := f^*(\mathbb{P}_n)$ be the spherically symmetric log-concave MLE. Define $h_0 \in \mathcal{H}$ by $h_0(r) := r^{p-1} e^{\phi_0(r)}$. Then, writing $\nu_k^2 := 2 \wedge \inf_{h \in \mathcal{H}^{(k)}} d_{\text{KL}}^2(h_0, h)$, there exists a universal constant $C > 0$ such that

$$\mathbb{E} d_X^2(\hat{f}_n, f_0) \leq C \min_{k=1, \dots, n} \left(\frac{k^{4/5} \nu_k^{2/5}}{n^{4/5}} \log \frac{en}{k \nu_k} + \frac{k}{n} \log^{5/4} \frac{en}{k} \right).$$

Remark: Taking the universal constant $C > 0$ from the conclusion of Theorem 9 and setting $C_* := \max\{(3C/2)^{5/4}, 1\}$, we see that if $\nu_k^2 \geq C_* \frac{k}{n} \log^{5/4} \left(\frac{en}{k} \right)$ and $k \in \{1, \dots, n\}$, then

$$C \frac{k^{4/5} \nu_k^{2/5}}{n^{4/5}} \log \frac{en}{k \nu_k} \leq C \nu_k^2 \left\{ \frac{1}{C_*^{4/5} \log(en/k)} \log \left(\frac{en^{3/2}}{C_*^{1/2} k^{3/2} \log^{5/8}(en/k)} \right) \right\} \leq \nu_k^2.$$

On the other hand, if $\nu_k^2 \leq C_* \frac{k}{n} \log^{5/4} \left(\frac{en}{k} \right)$ and $k \in \{1, \dots, n\}$, then

$$\frac{k^{4/5} \nu_k^{2/5}}{n^{4/5}} \log \frac{en}{k \nu_k} \lesssim \frac{k}{n} \log^{1/4} \left(\frac{en}{k} \right) \log \left(\frac{en^{3/2}}{k^{3/2} \log^{5/8}(en/k)} \right) \lesssim \frac{k}{n} \log^{5/4} \frac{en}{k}.$$

It follows that Theorem 9 implies the following sharp oracle inequality: there exists a universal constant $C > 0$ such that

$$\mathbb{E}_{h_0} d_X^2(\hat{f}_n, f_0) \leq \min_{k=1, \dots, n} \left(\nu_k^2 + C \frac{k}{n} \log^{5/4} \frac{en}{k} \right).$$

The proof of Theorem 9 proceeds by first considering the case $k = 1$, described in Proposition 10 below, for which we obtain a slightly different approximation error term.

Proposition 10. Let $a \in [0, \infty)$ and suppose that $Z_1, \dots, Z_n \stackrel{\text{iid}}{\sim} h_0$ for some $h_0 \in \mathcal{H}_a$ with empirical distribution function \mathbb{Q}_n , and let $\hat{h}_n := h_a^*(\mathbb{Q}_n)$. Set $\nu := \inf\{d_{\text{H}}(h_0, h) : h \in \mathcal{H}_a^{(1)}, h_0 \ll h\}$. Then there exists a universal constant $C > 0$ such that

$$\mathbb{E}_{h_0} d_X^2(\hat{h}_n, h_0) \leq C \left(\frac{\nu^{2/5}}{n^{4/5}} \log \frac{en}{\nu} + \frac{1}{n} \log^{5/4}(en) \right).$$

Remark: Since $2^{1/2} \leq ne^{-3/2}$ for $n \geq 8$ and the function $x \mapsto x^{2/5} \log(en/x)$ is increasing for $x \leq ne^{-3/2}$, Proposition 10 remains true if we redefine $\nu^2 := 2 \wedge \inf_{h \in \mathcal{H}_a^{(1)}} d_{\text{KL}}^2(h_0, h)$. Hence, the conclusion of Proposition 10 is stronger than that obtained by specialising Theorem 9 to the case $k = 1$.

5 Algorithm

In this section, we assume we are given data X_1, \dots, X_n in \mathbb{R}^p and let $Z_i := \|X_i\|$ for $i = 1, \dots, n$. We let \mathbb{P}_n and \mathbb{Q}_n denote the empirical distributions of X_1, \dots, X_n and Z_1, \dots, Z_n respectively. Proposition 2 above shows that, provided at least one of Z_1, \dots, Z_n is non-zero, the function $\hat{\phi}_n := \phi^*(\mathbb{Q}_n)$ is well-defined, and we can then set $\hat{f}_n(x) := e^{\hat{\phi}_n(\|x\|)}$. Our aim is therefore to provide an algorithm for computing $\hat{\phi}_n$.

Let $\bar{\Phi}$ denote the set of $\phi \in \Phi$ with the property that ϕ is constant on the interval $[0, Z_{(1)}]$ and affine on the intervals $[Z_{(i-1)}, Z_{(i)}]$ for $i = 2, \dots, n$, with $\phi(r) = -\infty$ for $r > Z_{(n)}$. Observe that if we fix $\phi \in \Phi$, and $\bar{\phi} \in \bar{\Phi}$ be such that $\bar{\phi}(0) = \phi(0)$ and $\bar{\phi}(Z_i) = \phi(Z_i)$ for all $i = 1, \dots, n$. Then by concavity of ϕ , we have $\phi(r) \geq \bar{\phi}(r)$ for all $r \in [0, \infty)$. Hence

$$\begin{aligned} L(\phi, \mathbb{Q}_n) &= \frac{1}{n} \sum_{i=1}^n \phi(Z_i) - \int_0^\infty r^{p-1} e^{\phi(r)} dr + 1 \leq \frac{1}{n} \sum_{i=1}^n \bar{\phi}(Z_i) - \int_0^\infty r^{p-1} e^{\bar{\phi}(r)} dr + 1 \\ &= L(\bar{\phi}, \mathbb{Q}_n). \end{aligned} \quad (6)$$

We now assume for simplicity of exposition that Z_1, \dots, Z_n are distinct. The more general case can be treated similarly by assigning appropriate weights to duplicated points. Any $\phi \in \bar{\Phi}$ can be identified with $\phi = (\phi_1, \dots, \phi_n)^\top \in \mathbb{R}^n$ given by $\phi_i := \phi(Z_i)$ for $i = 1, \dots, n$. For $i = 1, \dots, n-1$, let $\delta_i := Z_{(i+1)} - Z_{(i)}$. Define $v_1 = (v_{1,j})_{j=1}^n \in \mathbb{R}^n$ to have two non-zero entries, namely $v_{1,1} := -1$, $v_{1,2} := 1$. Further, for $i = 2, \dots, n-1$, let $v_i = (v_{i,j})_{j=1}^n \in \mathbb{R}^n$ have three non-zero entries, namely

$$v_{i,i-1} := \frac{1}{\delta_{i-1}}, \quad v_{i,i} := -\frac{1}{\delta_i} - \frac{1}{\delta_{i-1}} \quad \text{and} \quad v_{i,i+1} := \frac{1}{\delta_i}.$$

Finally, let $\bar{\Phi}_n := \{\phi \in \mathbb{R}^n : v_i^\top \phi \leq 0 \text{ for } i = 1, \dots, n-1\}$. By (6), we see that it suffices to compute $\phi^* = (\phi_1^*, \dots, \phi_n^*)^\top \in \operatorname{argmax}_{\phi \in \bar{\Phi}_n} F(\phi)$, where

$$F(\phi) := \frac{1}{n} \sum_{i=1}^n \phi_i - \frac{1}{p} Z_{(1)}^p - \sum_{i=1}^{n-1} \int_{Z_{(i)}}^{Z_{(i+1)}} r^{p-1} \exp\left(\frac{Z_{(i+1)} - r}{\delta_i} \phi_i + \frac{r - Z_{(i)}}{\delta_i} \phi_{i+1}\right) dr + 1. \quad (7)$$

This is a finite-dimensional convex optimisation problem with linear inequality constraints. We propose an active set algorithm for the optimisation of (7), a variant of the algorithm used in Dümbgen et al. (2007) to compute the ordinary univariate log-concave MLE. For $\phi \in \bar{\Phi}_n$, we define $A(\phi) := \{i \in \{1, \dots, n-1\} : v_i^\top \phi = 0\}$ to be the set of ‘active’ constraints. Note that this is the complement in $\{1, \dots, n\}$ of the set of ‘knots’ of ϕ . Given a set $A \subseteq \{1, \dots, n-1\}$, we define $V(A) := \{\phi \in \mathbb{R}^n : v_i^\top \phi = 0, \forall i \in A\}$, and

$$V^*(A) := \operatorname{argmax}_{\phi \in V(A)} F(\phi). \quad (8)$$

Here, the maximiser is unique because $F(\cdot)$ is strictly concave on \mathbb{R}^n with $F(\phi) \rightarrow -\infty$ as $\|\phi\| \rightarrow \infty$. It is convenient to define, for $i = 1, \dots, n-1$, vectors $b_i = (b_{i,j})_{j=1}^n \in \mathbb{R}^n$ by

$$b_{i,j} := - \sum_{k=i}^{j-1} \delta_k,$$

where, as usual, we interpret an empty sum as 0, and also define $b_n := \mathbf{1}_n \in \mathbb{R}^n$, the all-one vector. It follows from this definition that $b_i^\top v_i = -1$ for $i \in \{1, \dots, n-1\}$ and $b_i^\top v_j = 0$ for all $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, n-1\}$ with $i \neq j$. Finally, given $\phi \in \bar{\Phi}_n$ and $\phi' \in \mathbb{R}^n$, we define

$$t(\phi, \phi') := \max \left\{ \frac{v_i^\top \phi'}{v_i^\top (\phi' - \phi)} : i \in \{1, \dots, n-1\} \setminus A(\phi), v_i^\top \phi' > 0 \right\}.$$

We are now in a position to present the full algorithm; see Algorithm 1. It is guaranteed to terminate in finitely many steps with the exact solution.

Algorithm 1 Computing the spherically symmetric log-concave MLE

Input: $X_1, \dots, X_n \in \mathbb{R}^p$. **Output:** $\phi \in \bar{\Phi}_n$.

```

1:  $Z_i \leftarrow \|X_i\|$  for all  $i = 1, \dots, n$ .
2:  $A \leftarrow \{1, \dots, n-1\}$ .
3:  $\phi \leftarrow V^*(A)$ .
4: while  $\max_{i=1, \dots, n} b_i^\top \nabla F(\phi) \geq 0$  do
5:    $i^* \leftarrow \operatorname{argmax}_{i=1, \dots, n} b_i^\top \nabla F(\phi)$ 
6:    $\phi' \leftarrow V^*(A \setminus \{i^*\})$ 
7:   while  $\phi' \notin \bar{\Phi}_n$  do
8:      $\phi \leftarrow t(\phi, \phi')\phi + \{1 - t(\phi, \phi')\}\phi'$ 
9:      $A \leftarrow A(\phi)$ 
10:     $\phi' \leftarrow V^*(A)$ 
11:   end while
12:    $\phi \leftarrow \phi'$ 
13:    $A \leftarrow A(\phi)$ 
14: end while

```

We complete this section by providing further detail on how to solve the optimisation problem in (8). Given the active set $A \subseteq \{1, \dots, n-1\}$, let us define $I := \{1, \dots, n\} \setminus A$. We index the elements of I by $i_1 < \dots < i_T$ where $T := |I|$. Given $v \in \mathbb{R}^{(n-1) \times n}$, we also write v_A for the matrix in $\mathbb{R}^{|A| \times n}$ obtained by extracting the rows of v with indices in A . Observe that the set $\{\phi = (\phi_1, \dots, \phi_n)^\top \in \mathbb{R}^n : v_A \phi = 0\}$ is the

subspace of \mathbb{R}^n where for $j < i_1$, we have $\phi_j = \phi_{i_1}$, and for $j \in \{i_t + 1, \dots, i_{t+1} - 1\}$, we have

$$\phi_j = \frac{Z_{(i_{t+1})} - Z_{(j)}}{Z_{(i_{t+1})} - Z_{(i_t)}} \phi_{i_t} + \frac{Z_{(j)} - Z_{(i_t)}}{Z_{(i_{t+1})} - Z_{(i_t)}} \phi_{i_{t+1}}.$$

It follows we can solve the optimisation problem (8) by solving instead an unconstrained optimisation over T variables, i.e. by computing

$$\begin{aligned} \max_{\phi_I \in \mathbb{R}^T} \frac{1}{n} & \left(i_1 \phi_{i_1} + \sum_{t=1}^{T-1} \sum_{j=i_t+1}^{i_{t+1}} \frac{Z_{(i_{t+1})} - Z_{(j)}}{Z_{(i_{t+1})} - Z_{(i_t)}} \phi_{i_t} + \frac{Z_{(j)} - Z_{(i_t)}}{Z_{(i_{t+1})} - Z_{(i_t)}} \phi_{i_{t+1}} \right) \\ & - \frac{1}{p} \exp(\phi_{i_1}) Z_{(i_1)}^p - \sum_{t=1}^{T-1} \int_{Z_{(i_t)}}^{Z_{(i_{t+1})}} r^{p-1} \exp\left(\frac{Z_{(i_{t+1})} - r}{Z_{(i_{t+1})} - Z_{(i_t)}} \phi_i + \frac{r - Z_{(i_t)}}{Z_{(i_{t+1})} - Z_{(i_t)}} \phi_{i_{t+1}}\right) dr. \end{aligned}$$

We solve this latter problem via Newton's method.

6 Empirical performance

In this section, we compare the performance of the spherically symmetric log-concave MLE against two alternative approaches. The first competitor is the spherically symmetric estimator \tilde{f}_n defined in the introduction, based on computing the univariate log-concave MLE of the density of the norms of the data. The second competitor is the spherically symmetric estimator \tilde{f}_n^{ker} , which is similar but estimates the density of the norms of the data via kernel smoothing.

In the first experiment, we generate data from an isotropic Gaussian distribution f_0 and measure the error in Kullback–Leibler divergence $d_{\text{KL}}(\cdot, f_0)$. For each setting of n and p , we repeat the simulation 100 times and report the average error. The results are shown in Figure 2a and 2b. In Figure 2b, we plot $n^{-4/5}$ on the x -axis for various sample sizes to demonstrate the scaling proved in Theorem 8.

In the second experiment, we generate data whose true density is uniform on a Euclidean ball, i.e., $f_0(x) = p^{-p} c_p^{-1} \mathbb{1}_{\{\|x\| \leq p\}}$. We measure the error in Hellinger distance because f_0 has bounded support and there exist $x \in \mathbb{R}^p$ where $\tilde{f}_n^{\text{ker}}(x) > 0$ but $f_0(x) = 0$. We plot the results in Figures 2c and 2d. For the simulation study described in Figure 2d, the error for \tilde{f}_n^{ker} is always above 0.01 and hence does not appear in the figure. In Figure 2d, we plot n^{-1} on the x-axis to demonstrate the scaling proved in Theorem 9.

In the third experiment, the true density is $f_0(x) = \Gamma(p)^{-1} c_p^{-1} e^{-\|x\|}$ and, as with the first experiment, we measure the error in Kullback–Leibler divergence. We plot the results in Figure 2e and 2f. In Figure 2e, we plot n^{-1} on the x -axis to demonstrate the scaling proved in Theorem 9.

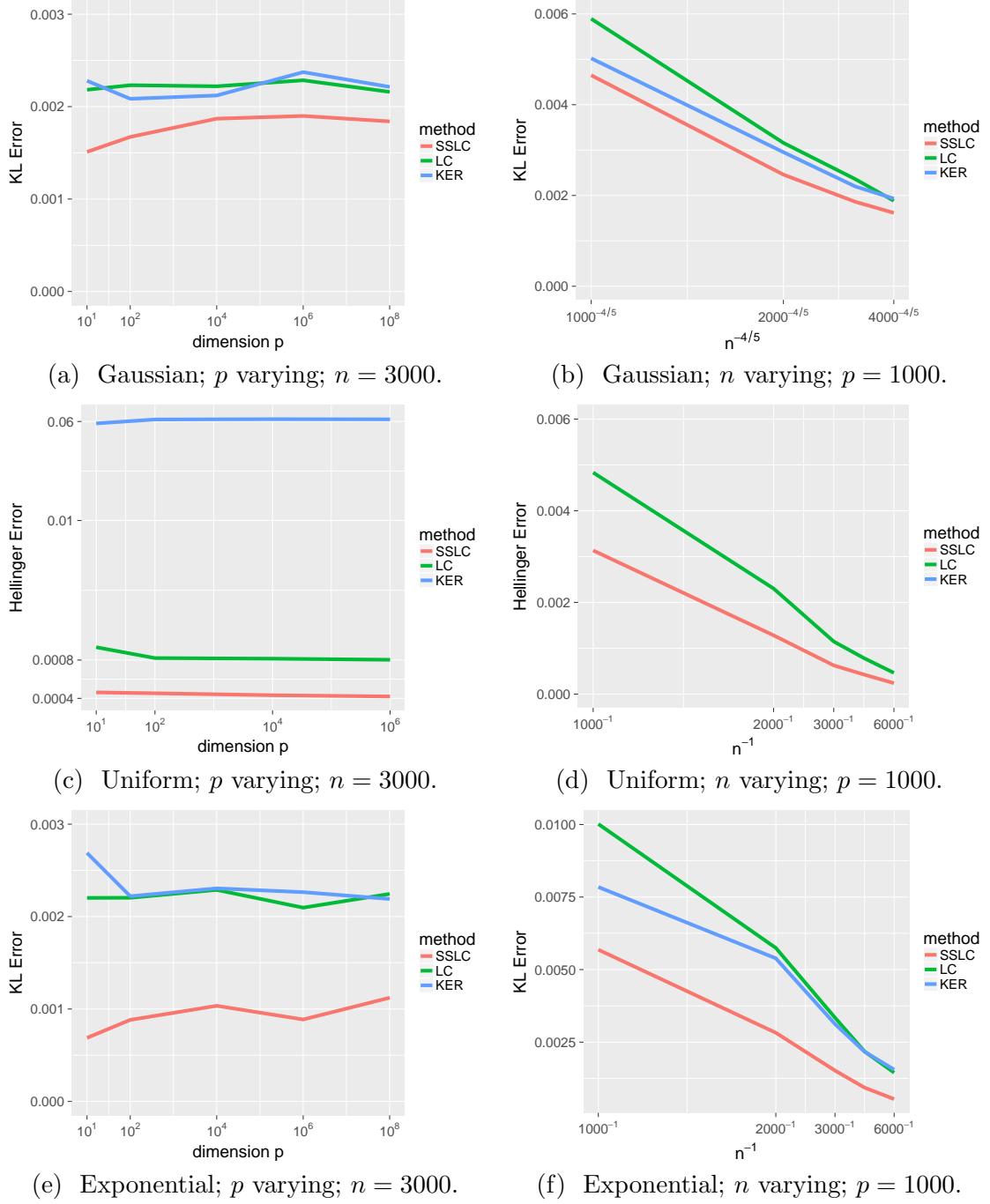


Figure 2: Simulation studies.

These simulations confirm the theoretical findings in Section 8 and illustrate the strong finite-sample performance of the spherically symmetric log-concave MLE.

7 Proofs of main results

7.1 Proofs from Section 2

Proof of Proposition 1. Suppose that f is a density on \mathbb{R}^p that can be written as $f(x) = e^{\phi(\|x\|)}$ for some $\phi \in \Phi$. Then f is upper semi-continuous and spherically symmetric. Moreover, for any $t \in (0, 1)$, and $x, y \in \mathbb{R}^p$, and using the facts that ϕ is decreasing and concave,

$$\begin{aligned} \log f(tx + (1-t)y) &= \phi(\|tx + (1-t)y\|) \geq \phi(t\|x\| + (1-t)\|y\|) \\ &\geq t\phi(\|x\|) + (1-t)\phi(\|y\|) = t \log f(x) + (1-t) \log f(y). \end{aligned}$$

Thus, f is log-concave, as required.

We now turn to the converse. Suppose that $f \in \mathcal{F}_p^{\text{SS}}$. Then since f is upper semi-continuous and spherically symmetric, we can write $f(x) = g(\|x\|)$ for some upper semi-continuous $g : [0, \infty) \rightarrow [0, \infty)$. It remains to show that $\log g$ is decreasing and concave. To this end, suppose for a contradiction that there exists $r > r' \geq 0$ such that $\log g(r) > \log g(r')$, and let $x \in \mathbb{R}^p$ be such that $\|x\| = r$. Then

$$\log f\left(\frac{r'x}{r}\right) = \log g(r') < \log g(r) = \frac{r+r'}{2r} \log f(x) + \frac{r-r'}{2r} \log f(-x),$$

contradicting the log-concavity of f .

On the other hand, if g is not log-concave, then there exist $r, r' \in [0, \infty)$ and $t \in (0, 1)$ such that $\log g(tr + (1-t)r') < t \log g(r) + (1-t) \log g(r')$. But then, for any $x \in \mathbb{R}^p$ with $\|x\| = 1$, we have

$$\begin{aligned} \log f(trx + (1-t)r'x) &= \log g(tr + (1-t)r') \\ &< t \log g(r) + (1-t) \log g(r') = t \log f(rx) + (1-t) \log f(r'x), \end{aligned}$$

again contradicting the log-concavity of f . □

Proof of Proposition 2. (i) Fix $\phi \in \Phi_a$. Observe that if $\lim_{r \rightarrow \infty} \phi(r) = c > -\infty$, then $L(\phi, Q) \leq \phi(a) - e^c \int_a^\infty r^{p-1} dr = -\infty$. Otherwise $\lim_{r \rightarrow \infty} \phi(r) = -\infty$, and then there exist $\alpha > 0, \beta \in \mathbb{R}$ such that $\phi(r) \leq -\alpha r + \beta$ (Cule and Samworth, 2010, Lemma 1). Hence, $L(\phi, Q) \leq \int_a^\infty \phi dQ \leq -\alpha \int_a^\infty r dQ(r) + \beta = -\infty$.

(ii) Now suppose that $Q(\{a\}) = 1$ and let $e^{\phi_n(r)} := n \mathbb{1}_{\{r \in [a, a+n^{-1}]\}}$. Then,

$$L(\phi_n, Q) = \log n - n \int_a^{a+n^{-1}} r^{p-1} dr + 1 \geq \log n - (a + n^{-1})^{p-1} + 1 \rightarrow \infty$$

as $n \rightarrow \infty$.

(iii) Finally, suppose that $Q \in \mathcal{Q}_a$. For $\phi(r) := -r$, we have

$$L(\phi, Q) = - \int_a^\infty r dQ(r) - \int_a^\infty r^{p-1} e^{-r} dr + 1 \geq - \int_a^\infty r dQ(r) - \Gamma(p) + 1 > -\infty,$$

so $\sup_{\phi \in \Phi_a} L(\phi, Q) > -\infty$.

For $\delta, \epsilon > 0$, let $\mathcal{Q}_a(\delta, \epsilon) := \{Q \in \mathcal{Q}_a : Q((a + \delta, \infty)) \geq \epsilon\}$. Then, since $Q(\{a\}) < 1$, we have $Q \in \mathcal{Q}_a(\delta, \epsilon)$ for some $\delta, \epsilon > 0$. We also write $M := \phi(a)$ and $M' := \phi(a + \delta)$. Then by the concavity of ϕ ,

$$L(\phi, Q) \leq M(1 - \epsilon) + M'\epsilon - \int_a^{a+\delta} r^{p-1} \exp\left(M - \frac{r-a}{\delta}(M-M')\right) dr + 1. \quad (9)$$

If $M > 0$ and $(M - M')\epsilon > 2M$, then

$$\begin{aligned} L(\phi, Q) &\leq M - e^M \int_a^{a+\delta} r^{p-1} \exp\left(-\frac{2M(r-a)}{\epsilon\delta}\right) dr + 1 \\ &\leq M - e^M \left(\frac{\epsilon\delta}{2M}\right)^p \int_0^{2M/\epsilon} \left(\frac{2Ma}{\epsilon\delta} + s\right)^{p-1} e^{-s} ds + 1 \\ &\leq M - e^M \left(\frac{\epsilon\delta}{2M}\right)^p \int_0^{2M/\epsilon} s^{p-1} e^{-s} ds + 1. \end{aligned}$$

On the other hand, if $M > 0$ and $(M - M')\epsilon > 2M$, then from (9) we see that $L(\phi, Q) \leq -M + 1$. We deduce that there exists $M^* > 0$, depending only on δ and ϵ , such that

$$\sup_{\phi \in \Phi_a} L(\phi, Q) = \sup_{\phi \in \Phi_a : \phi(a) \leq M^*} L(\phi, Q) < \infty.$$

The existence of ϕ^* then follows from the proof of Theorem 2.2 in [Dümbgen et al. \(2011\)](#).

(iv) By the change of variable formula (e.g. [Billingsley, 1995](#), Theorem 16.13), we have $\int_{\mathbb{R}^p} \phi(\|x\|) dP(x) = \int_{[0, \infty)} \phi(r) dQ(r)$ for all $\phi \in \Phi$. The result then follows from (iii), specialised to the case $a = 0$. \square

Proof of Proposition 3. (i) For any $\phi \in \Phi_a$, we may define $\phi_\alpha \in \Phi_{\alpha a}$ by $\phi_\alpha(r) := \phi(r/\alpha) - p \log \alpha$. The map $\phi \mapsto \phi_\alpha$ is a bijection from Φ_a to $\Phi_{\alpha a}$. Let $\phi^* = \phi_a^*(Q)$. Then, for any $\phi \in \Phi_a$, we have

$$\begin{aligned} \int_{[\alpha a, \infty)} \phi_\alpha^* dQ_\alpha - \int_{\alpha a}^\infty r^{p-1} e^{\phi_\alpha^*(r)} dr &= \int_{[a, \infty)} \phi^* dQ - p \log \alpha - \int_a^\infty s^{p-1} e^{\phi^*(s)} ds \\ &\geq \int_{[a, \infty)} \phi dQ - p \log \alpha - \int_a^\infty s^{p-1} e^{\phi(s)} ds \\ &\geq \int_{[\alpha a, \infty)} \phi_\alpha dQ_\alpha - \int_{\alpha a}^\infty r^{p-1} e^{\phi_\alpha(r)} dr. \end{aligned}$$

This establishes that $\phi_\alpha^* = \phi_a^*(Q_\alpha)$ and thus proves scale equivariance.

(ii) For any $t > 0$, we have

$$0 \geq \frac{1}{t} \{L(\phi^* + t\Delta, Q) - L(\phi^*, Q)\} = \int_{[a, \infty)} \Delta dQ - \frac{1}{t} \int_a^\infty (e^{t\Delta(r)} - 1) h^*(r) dr. \quad (10)$$

Choose $t_0 > 0$ small enough that $\phi^* + t_0\Delta \in \Phi_a$. Since $\int_a^\infty r^{p-1} e^{\phi^*(r)} dr < \infty$, we must have $\phi^*(r) \rightarrow -\infty$ as $r \rightarrow \infty$, and hence, by reducing $t_0 > 0$ if necessary, we may assume that $\int_a^\infty \Delta(r) r^{p-1} e^{\phi^*(r)+t_0\Delta(r)} dr < \infty$. Now, for $t \in (0, t_0]$,

$$\frac{1}{t} (e^{t\Delta(r)} - 1) \leq \frac{1}{t_0} (e^{t_0\Delta(r)} - 1) \mathbb{1}_{\{\Delta(r) \geq 0\}} + \Delta(r) \mathbb{1}_{\{\Delta(r) < 0\}}.$$

Hence, if $\int_a^\infty \Delta(r) h^*(r) dr > -\infty$, then we may apply the dominated convergence theorem to (10) to take the limit as $t \searrow 0$ and reach the desired conclusion. On the other hand, if $\int_a^\infty \Delta(r) h^*(r) dr = -\infty$, then for every $t \in (0, t_0]$,

$$\begin{aligned} & \frac{1}{t} \int_a^\infty (e^{t\Delta(r)} - 1) h^*(r) dr \\ & \leq \int_a^\infty \left(\frac{1}{t_0} (e^{t_0\Delta(r)} - 1) \mathbb{1}_{\{\Delta(r) \geq 0\}} + \Delta(r) \mathbb{1}_{\{\Delta(r) < 0\}} \right) h^*(r) dr = -\infty. \end{aligned}$$

The result follows. \square

(iii) Letting $\Delta(r) = -r$, this is a consequence of (ii).

(iv) Letting $\Delta(r) = \log(h_0(r)/h^*(r))$, this also follows from an application of (ii).

\square

Proof of Proposition 4. The proof is very similar to (in fact, somewhat more straightforward than) the proof of Dümbgen et al. (2011, Theorem 4.5), so we focus on the main differences. We first observe that if $X_n \sim P_n$ and $X \sim P$ are defined on the same probability space, then

$$|\mathbb{E}\|X_n\| - \mathbb{E}\|X\|\| \leq \mathbb{E}\|X_n - X\| \leq d_W(P_n, P).$$

Hence, writing Q_n and Q for the distributions of $\|X_n\|$ and $\|X\|$ respectively, we deduce that $d_W(Q_n, Q) \leq d_W(P_n, P) \rightarrow 0$. It follows that $\int_0^\infty r dQ_n(r) \rightarrow \int_0^\infty r dQ(r) < \infty$, and $\limsup_{n \rightarrow \infty} Q_n(\{0\}) \leq Q(\{0\}) < 1$, so $Q_n \in \mathcal{Q}$ for $n \geq n_0$, say. For such n , we write $\phi_n^* := \phi^*(Q_n)$ and $\phi^* := \phi^*(Q)$. Let $n_0 \leq n_1 < n_2 < \dots$ be an arbitrary, strictly increasing sequence of positive integers. By extracting a further subsequence if necessary, we may assume that $L(\phi_{n_k}^*, Q_{n_k}) \rightarrow \alpha \in [-\infty, \infty]$. First note that, by considering the function $\phi(r) = -r$,

$$\alpha \geq \lim_{k \rightarrow \infty} \int_0^\infty -r dQ_{n_k}(r) - (p-1)! + 1 = \int_0^\infty -r dQ(r) - (p-1)! + 1 > -\infty.$$

Our next claim is that $\limsup_{k \in \mathbb{N}} \sup_{r \in [0, \infty)} \phi_{n_k}^*(r) < \infty$. To see this, recall the definition of the classes $\mathcal{Q}(\delta, \epsilon)$ from the proof of Proposition 2, and let $\delta_0, \epsilon_0 > 0$ be such

that $Q \in \mathcal{Q}(\delta_0, \epsilon_0)$. By examining the proof of Proposition 2, it suffices to prove that $Q_{n_k} \in \mathcal{Q}(\delta_0, \epsilon_0)$ for large k (with $a = 0$). But $\liminf_{k \rightarrow \infty} Q_{n_k}((\delta_0, \infty)) \geq Q((\delta_0, \infty)) \geq \epsilon_0$, which establishes the claim.

Let $r_0 := \sup\{r \in [0, \infty) : Q(r) < 1\}$. Our next claim is that $\liminf_{k \rightarrow \infty} \phi_{n_k}^*(r) > -\infty$ for all $r \in [0, r_0)$. To see this, note by our first claim that we may assume without loss of generality that there exists $M^* \geq \max(\alpha, 0)$ such that $\sup_{k \in \mathbb{N}} \sup_{r \in [0, \infty)} \phi_{n_k}^*(r) \leq M^*$. Then, for any $r \in [0, r_0)$,

$$L(\phi_{n_k}^*, Q_{n_k}) \leq \phi_{n_k}^*(0) \{1 - Q_{n_k}((r, \infty))\} + \phi_{n_k}^*(r) Q_{n_k}((r, \infty)).$$

Since $Q((r, \infty)) > 0$, we deduce that

$$\liminf_{k \rightarrow \infty} \phi_{n_k}^*(r) \geq \liminf_{k \rightarrow \infty} \frac{L(\phi_{n_k}^*, Q_{n_k}) - \phi_{n_k}^*(0) \{1 - Q_{n_k}((r, \infty))\}}{Q_{n_k}((r, \infty))} \geq -\frac{M^* - \alpha}{Q((r, \infty))} > -\infty,$$

as required.

These two claims allow us to extract a further subsequence $(\phi_{n_{k(\ell)}}^*)$ that converges in an appropriate sense to a limit $\phi^* \in \Phi$ (in particular, this convergence occurs Lebesgue almost everywhere). It turns out that $\phi^* = \phi^*(Q)$, that $L(\phi_{n_{k(\ell)}}^*, Q) \rightarrow L(\phi^*, Q)$, and, writing $f_\ell^* := f^*(P_{n_{k(\ell)}})$, we have $\int_{\mathbb{R}^p} |f_\ell^* - f^*| \rightarrow 0$. The desired total variation convergence (4) follows. See the proof of Theorem 4.5 of [Dümbgen et al. \(2011\)](#) for details.

For the final claim, note that our previous argument allows us to conclude that $f^*(P_n)$ converges to $f^*(P)$ Lebesgue almost everywhere. The conclusion therefore follows from [Rockafellar \(1997, Theorem 10.8\)](#). \square

7.2 Proofs from Section 3

Proof of Proposition 5. Let $r_0 := \sup\{r \in [a, \infty) : \hat{\phi}_n(r) = \hat{\phi}_n(a)\}$ and define $s_0 := r_0 - a$. By (6), we have that $r_0 = Z_i$ for some i because \mathbb{Q}_n is an empirical distribution. We also know that the right derivative of $\hat{\phi}_n$ at r_0 is strictly negative. Hence, by Proposition 3(ii), applied to the functions $\Delta(r) := \pm(r - r_0)_+$, we have

$$\begin{aligned} \int_a^\infty (r - r_0)_+ \hat{h}_n(r) dr + r_0 &= \frac{1}{n} \sum_{i=1}^n (Z_i - r_0)_+ + r_0 \\ &= \frac{1}{n} \sum_{i=1}^n (Z_i - r_0) + \frac{1}{n} \sum_{i=1}^n (Z_i - r_0)_- + r_0 \\ &= \bar{Z} + \frac{1}{n} \sum_{i=1}^n (r_0 - Z_i)_+. \end{aligned} \tag{11}$$

Now, since $\hat{\phi}_n(r) = \hat{\phi}_n(a)$ for all $r \in [a, r_0]$, we have

$$1 \geq e^{\hat{\phi}_n(a)} \int_a^{r_0} r^{p-1} dr = e^{\hat{\phi}_n(a)} \frac{r_0^p - a^p}{p}.$$

We deduce that

$$\begin{aligned} \int_a^\infty (r - r_0)_- \hat{h}_n(r) dr &= e^{\hat{\phi}_n(a)} \int_a^{r_0} (r_0 - r) r^{p-1} dr \\ &\leq \frac{p}{r_0^p - a^p} \left(\frac{r_0^{p+1}}{p} - \frac{r_0 a^p}{p} - \frac{r_0^{p+1}}{p+1} + \frac{a^{p+1}}{p+1} \right) \\ &= r_0 - \frac{p}{p+1} \frac{r_0^{p+1} - a^{p+1}}{r_0^p - a^p} \leq \frac{s_0}{\rho(s_0)}, \end{aligned} \quad (12)$$

where we used Lemma 14 to obtain the final bound. From (11) and (12), we find that

$$\mu_{\hat{h}_n} = \int_a^\infty (r - r_0)_+ \hat{h}_n(r) dr - \int_a^\infty (r - r_0)_- \hat{h}_n(r) dr + r_0 \geq \bar{Z} + \frac{1}{n} \sum_{i=1}^n (r_0 - Z_i)_+ - \frac{s_0}{\rho(s_0)}. \quad (13)$$

In particular, $\mu_{\hat{h}_n} \geq \bar{Z} - \frac{s_0}{\rho(s_0)}$.

For $i = 1, \dots, n$, let $\tilde{Z}_i := Z_i$ if $Z_i \leq r_0$ and $\tilde{Z}_i := s_0 + a$ if $Z_i > r_0$. Then $n^{-1} \sum_{i=1}^n (r_0 - Z_i)_+ = s_0 - n^{-1} \sum_{i=1}^n (\tilde{Z}_i - a) \geq 0$ and $n^{-1} \sum_{i=1}^n \tilde{Z}_i \leq \bar{Z}$. Hence

$$\begin{aligned} \frac{s_0}{\rho(s_0)} - \frac{1}{n} \sum_{i=1}^n (r_0 - Z_i)_+ \\ &= \frac{s_0 - n^{-1} \sum_{i=1}^n (\tilde{Z}_i - a)}{\rho(s_0)} + \frac{n^{-1} \sum_{i=1}^n (\tilde{Z}_i - a)}{\rho(s_0)} - \left(s_0 - \frac{1}{n} \sum_{i=1}^n (\tilde{Z}_i - a) \right) \\ &\leq \left(s_0 - \frac{1}{n} \sum_{i=1}^n (\tilde{Z}_i - a) \right) \left(\frac{1}{\rho(s_0)} - 1 \right) + \frac{\bar{Z} - a}{\rho(s_0)} \leq \frac{\bar{Z} - a}{\rho(s_0)}. \end{aligned} \quad (14)$$

The second lower bound for $\mu_{\hat{h}_n}$ follows from (13) and (14). The upper bound on $\mu_{\hat{h}_n}$ follows from Proposition 3. \square

Proof of Proposition 6. We have

$$\sup_{r \geq a} \log \hat{h}_n(r) \geq \int_a^\infty \log \hat{h}_n d\mathbb{Q}_n \geq \ell_{\hat{h}_n},$$

so since \hat{h}_n is upper semi-continuous, there exists $r_0 \geq a$ such that $\log \hat{h}_n(r_0) \geq \ell_{\hat{h}_n}$. By Lemma 15, we then have that $\sigma_{\hat{h}_n} \leq C' e^{-\ell_{\hat{h}_n}}$ for some universal constant $C' > 0$. Now, let $M := 3 \vee \sup_{r \geq a} \log \hat{h}_n(r)$ and let $D_t := \{t : \log \hat{h}_n(r) \geq t\}$. By (Dümbgen et al., 2011, Lemma 4.1), we have that $\lambda(D_{-2M}) \leq 6M e^{-M}$. Suppose now for contradiction that $M > \max\{-2 \log(\delta_{\mathbb{Q}_n}/6), -2\ell_{\hat{h}_n}\}$. Then

$$\lambda(D_{-2M}) < 6e^{-M/2} < \delta_{\mathbb{Q}_n},$$

so $\mathbb{Q}_n(D_{-2M}) \leq 1/2$. Hence

$$\int_a^\infty \log \hat{h}_n d\mathbb{Q}_n \leq -2M\{1 - \mathbb{Q}_n(D_{-2M})\} + M\mathbb{Q}_n(D_{-2M}) \leq -M/2 < \ell_{\hat{h}_n},$$

a contradiction. We deduce that (5) holds, so by Lemma 15, there exists a universal constant $c' > 0$ such that $\sigma_{\hat{h}_n} \geq c' \min(\delta_{\mathbb{Q}_n}^2, e^{2\ell_{\hat{h}_n}}, 1)$. The lower bound for $\sigma_{\hat{h}_n}$ follows, with $C_\sigma := \max(C', 1/c')$. \square

Proof of Proposition 7. We may assume that $n \geq 500$. Let $E := \{|\bar{Z} - \mu_{h_0}| \leq 1\}$, so that $\mathbb{P}(E^c) \leq 1/n$, by Chebychev's inequality. On the event E , we have $\mu_{\hat{h}_n} \leq \bar{Z} \leq \mu_{h_0} + 1$ by Proposition 5. If $s_0 \leq 1$, then $\mu_{\hat{h}_n} \geq \bar{Z} - 1 \geq \mu_{h_0} - 2$ because $\rho(s_0) \geq 1$ (Lemma 13(iii)). If $s_0 > 1$, then by Proposition 5 and Lemma 13(iv),

$$\begin{aligned} \mu_{\hat{h}_n} &\geq \bar{Z} - \frac{\bar{Z} - a_0}{\rho(s_0)} \geq \mu_{h_0} - 1 - \frac{20(\mu_{h_0} + 1 - a_0)}{\rho(s_0 + 1)} \\ &\geq \mu_{h_0} - 21 - \frac{20(\mu_{h_0} - a_0)}{\rho(s_0 + 1)}. \end{aligned}$$

Hence, if $\rho(s_0)/s_0 < 2^{-7}$, then by Lemma 11, we have $\mu_{\hat{h}_n} - \mu_{h_0} \geq -21 - 20 \times 2^{12} \geq -2^{17}$. On the other hand, if $\rho(s_0)/s_0 \geq 2^{-7}$, then by Proposition 5,

$$\mu_{\hat{h}_n} - \mu_{h_0} \geq \bar{Z} - \mu_{h_0} - \frac{s_0}{\rho(s_0)} \geq -1 - 2^7 \geq -2^8.$$

It follows that there exists a universal constant $C_\mu > 0$ such that

$$\mathbb{P}(|\mu_{\hat{h}_n} - \mu_{h_0}| > C_\mu) \leq 1/n.$$

To bound $\sigma_{\hat{h}_n}$, we show that, taking $\delta_{\mathbb{Q}_n} = 2^{-11}$ and $\ell_{\hat{h}_n} = -3$, there exists a universal constant $C > 0$ such that the hypotheses of Proposition 6 hold with probability at least $1 - C/n$. To verify the first condition, note that $\|h_0\|_\infty \leq 2^9$ (Lovász and Vempala, 2007, Theorem 5.14(b) and (d)). Let H_0 and \mathbb{H}_n denote the distribution functions corresponding to h_0 and \mathbb{Q}_n respectively. Let $(a, b]$ be an interval of length at most 2^{-11} . Then

$$\begin{aligned} \mathbb{Q}_n((a, b]) &= \mathbb{H}_n(b) - \mathbb{H}_n(a) = \mathbb{H}_n(b) - H_0(b) + H_0(b) - H_0(a) + \{H_0(a) - \mathbb{H}_n(a)\} \\ &\leq 2\|\mathbb{H}_n - H_0\|_\infty + \|h_0\|_\infty(b - a) \leq 2\|\mathbb{H}_n - H_0\|_\infty + \frac{1}{4}. \end{aligned}$$

Now, by the Dvoretzky–Kiefer–Wolfowitz inequality,

$$\mathbb{P}\left(\|\mathbb{H}_n - H_0\|_\infty > \sqrt{\frac{1}{2n} \log(4n)}\right) \leq \frac{1}{2n}.$$

Hence, for any $n \geq 20$, we have $\mathbb{P}(\|\mathbb{H}_n - H_0\|_\infty > 1/8) \leq 1/(2n)$. Thus, when $\delta_{\mathbb{Q}_n} = 2^{-11}$,

$$\mathbb{P}\left(\sup_{\substack{a_0 \leq a < b < \infty \\ |b-a| \leq \delta_{\mathbb{Q}_n}}} \mathbb{Q}_n((a, b]) > 1/2\right) \leq \frac{1}{2n}.$$

Now, by Lemma 16 and Bobkov and Madiman (2011, Theorem 1.1), for $n \geq 500$ and $\ell_{\hat{h}_n} = -3$,

$$\mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n \log h_0(Z_i) \leq \ell_{\hat{h}_n}\right) \leq \mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n \log h_0(Z_i) - \int_{a_0}^{\infty} h_0 \log h_0\right| > 1\right) \leq 2e^{-n^{1/2}/16} \leq \frac{250}{n}.$$

The result follows by Proposition 6. \square

7.3 Proofs from Section 4

Proof of Theorem 8. Let h_0 denote the density of $\|X_1\|$, so that $h_0(r) = r^{p-1}e^{\phi(r)}$ for some $\phi \in \Phi$. For $i = 1, \dots, n$, let $Z_i := \|X_i\|$, and write \mathbb{Q}_n for the empirical distribution of Z_1, \dots, Z_n . Now let $\hat{h}_n := h^*(\mathbb{Q}_n)$, so that $\hat{h}_n(r) = r^{p-1}e^{\hat{\phi}_n(r)}$, where $\hat{\phi}_n := \phi^*(\mathbb{Q}_n) = \tilde{\phi}^*(\mathbb{P}_n)$, by Proposition 2(iv). Then $\hat{h}_n(Z_i)/h_0(Z_i) = \hat{f}_n(X_i)/f_0(X_i)$ for $i = 1, \dots, n$, so $d_X^2(\hat{f}_n, f_0) = d_X^2(\hat{h}_n, h_0)$. By scale equivariance (Proposition 3(i)), we may assume without loss of generality that $\sigma_{h_0} = 1$. By Proposition 7, there exist universal constants $C_\mu, C_\sigma, C > 0$ such that $\mathbb{P}(\hat{h}_n \notin \mathcal{H}(h_0, C_\mu, C_\sigma)) \leq C/n$. By Lemma 17, there exists a universal constant $K > 0$ such that for every $\delta > 0$,

$$\int_0^\delta H_{\square}^{1/2}(\epsilon, \mathcal{H}(h_0, C_\mu, C_\sigma), d_H) d\epsilon \leq K\delta^{3/4}.$$

Define $\Psi(\delta) := \max(K\delta^{3/4}, \delta)$, so that $\delta \mapsto \Psi(\delta)/\delta^2$ is decreasing. By choosing $\delta_* := K_0 n^{-2/5}$ for a suitably large universal constant $K_0 > 0$, we may apply Kim et al. (2017b, Theorem 10) (a minor restatement of van de Geer (2000, Corollary 7.5)), to deduce that there exists a universal constant $K_* > 0$ such that

$$\begin{aligned} \mathbb{E}d_X^2(\hat{h}_n, h_0) &\leq \int_0^{16 \log n} \mathbb{P}(\{d_X^2(\hat{h}_n, h_0) \geq t\} \cap \{\hat{h}_n \in \mathcal{H}(h_0, C_\mu, C_\sigma)\}) dt \\ &\quad + 16 \log n \mathbb{P}(\hat{h}_n \notin \mathcal{H}(h_0, C_\mu, C_\sigma)) + \int_{16 \log n}^{\infty} \mathbb{P}(d_X^2(\hat{h}_n, h_0) \geq t) dt \\ &\leq \delta_*^2 + K_* \int_{\delta_*^2}^{16 \log n} \exp\left(-\frac{nt}{K_*^2}\right) dt + \frac{16C \log n}{n} + \int_{16 \log n}^{\infty} \mathbb{P}\left(\max_{i=1, \dots, n} \log \frac{\hat{h}_n(Z_i)}{h_0(Z_i)} \geq t\right) dt \\ &\lesssim n^{-4/5}, \end{aligned}$$

where, to obtain the final inequality, we have applied Lemmas 18 and 19. \square

Proof of Theorem 9. Fix $h_* \in \mathcal{H}^{(k)}$ where $h_*(r) = r^{p-1}e^{\phi_*(r)}$ and $\phi_* \in \Phi^{(k)}$, let I_1, \dots, I_k be the k intervals on which ϕ_* is affine, and let $r_0 := \sup\{r \in [a, \infty) : \phi_*(r) > -\infty\}$. Define $\mathcal{Z}_j := \{i : Z_i \in I_j\}$ and $n_j := |\mathcal{Z}_j|$. Let $J := \{j : n_j \geq 1\}$ be the set of indices of intervals with at least one data point. Define $\tilde{\Phi}$ to be the set of upper semi-continuous functions $\phi : [0, \infty) \rightarrow [-\infty, \infty)$ such that $\phi|_{I_j}$ is decreasing

and concave for each $j = 1, \dots, k$, and such that $\phi(r) = -\infty$ for $r > r_0$. Note that a function $\phi \in \tilde{\Phi}$ need not be globally decreasing and, in fact, need not be continuous on $[0, r_0]$. Given any $\phi \in \tilde{\Phi}$ and $j \in J$, let $\phi^{(j)}(r) := \phi(r) + \log(n/n_j)$ for $r \in I_j$, and let $\Phi_{I_j} := \{\phi|_{I_j} : \phi \in \Phi\}$. Now define

$$\tilde{\phi}_n^{(j)} := \operatorname{argmax}_{\phi_j \in \Phi_{I_j}} \left\{ \frac{1}{n_j} \sum_{i \in \mathcal{Z}_j} \phi_j(Z_i) - \int_{I_j} r^{p-1} e^{\phi_j(r)} dr \right\},$$

and let $\tilde{\phi}_n(r) := \tilde{\phi}_n^{(j)}(r) - \log(n/n_j)$ for $r \in I_j$, and $\tilde{\phi}_n(r) := -\infty$ for $r > r_0$. Then, for any $\phi \in \tilde{\Phi}$,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \tilde{\phi}_n(Z_i) - \int_0^\infty r^{p-1} e^{\tilde{\phi}_n(r)} dr &= \sum_{j \in J} \frac{n_j}{n} \left(\frac{1}{n_j} \sum_{i \in \mathcal{Z}_j} \tilde{\phi}_n(Z_i) - \frac{n}{n_j} \int_{I_j} r^{p-1} e^{\tilde{\phi}_n(r)} dr \right) \\ &= \sum_{j \in J} \frac{n_j}{n} \left(\frac{1}{n_j} \sum_{i \in \mathcal{Z}_j} \tilde{\phi}_n^{(j)}(Z_i) - \int_{I_j} r^{p-1} e^{\tilde{\phi}_n^{(j)}(r)} dr \right) - \sum_{j \in J} \frac{n_j}{n} \log \frac{n}{n_j} \\ &\geq \sum_{j \in J} \frac{n_j}{n} \left(\frac{1}{n_j} \sum_{i \in \mathcal{Z}_j} \phi^{(j)}(Z_i) - \int_{I_j} r^{p-1} e^{\phi^{(j)}(r)} dr \right) - \sum_{j \in J} \frac{n_j}{n} \log \frac{n}{n_j} \\ &\geq \frac{1}{n} \sum_{i=1}^n \phi(Z_i) - \int_0^\infty r^{p-1} e^{\phi(r)} dr. \end{aligned} \tag{15}$$

It follows that the function \tilde{h}_n defined by $\tilde{h}_n(r) := r^{p-1} e^{\tilde{\phi}_n(r)}$ is a density. Moreover, for $j \in J$, the function $\tilde{h}_n^{(j)} := \frac{n_j}{n} \tilde{h}_n|_{I_j}$ is a density. Writing $p_j := \int_{I_j} h_0$, and $h_0^{(j)} := \frac{1}{p_j} h_0|_{I_j}$, we deduce from (15) that

$$\begin{aligned} \mathbb{E}_{h_0} d_X^2(\hat{h}_n, h_0) &\leq \mathbb{E}_{h_0} \frac{1}{n} \sum_{i=1}^n \log \frac{\tilde{h}_n(Z_i)}{h_0(Z_i)} \\ &= \mathbb{E}_{h_0} \sum_{j \in J} \frac{n_j}{n} \left(\frac{1}{n_j} \sum_{i \in \mathcal{Z}_j} \log \frac{\tilde{h}_n^{(j)}(Z_i)}{h_0^{(j)}(Z_i)} \right) + \mathbb{E}_{h_0} \sum_{j \in J} \frac{n_j}{n} \log \frac{n_j}{np_j}. \end{aligned} \tag{16}$$

Now, since $n_j \sim \text{Bin}(n, p_j)$ and $\log x \leq x - 1$ for $x > 0$, we have

$$\begin{aligned} \mathbb{E}_{h_0} \sum_{j \in J} \frac{n_j}{n} \log \frac{n_j}{np_j} &\leq \mathbb{E}_{h_0} \sum_{j \in J} \frac{n_j}{n} \left(\frac{n_j}{np_j} - 1 \right) = \sum_{j \in J} \left(\frac{n^2 p_j^2 + np_j(1-p_j)}{n^2 p_j} - p_j \right) \\ &= \frac{1}{n} \sum_{j \in J} (1-p_j) \leq \frac{k}{n}. \end{aligned} \tag{17}$$

To bound the first term in (16), for $j = 1, \dots, k$, let us first define $q_j := \int_{I_j} h_*(r) dr$, $h_*^{(j)} := h_*/q_j$, $\nu_{*,j}^2 := d_H^2(h_0^{(j)}, h_*^{(j)})$, $\nu_*^2 := \sum_{j=1}^k p_j \nu_{*,j}^2$, $J_0 := \{j : n_j \nu_{*,j}^2 \geq 1\}$ and temporarily assume that $k \leq e^{-1/4}n$. By Proposition 10 (and the subsequent remark) below, applied conditionally on n_1, \dots, n_k , a simple extension of Jensen's inequality

using the fact that $x \mapsto \log^{5/4} x$ is concave on $[e^{1/4}, \infty)$ (e.g. [Han et al., 2017](#), Lemma 2) and the fact that $k \mapsto k \log^{5/4}(en/k)$ is increasing for $k \in [1, e^{-1/4}n]$,

$$\begin{aligned} \mathbb{E}_{h_0} \sum_{j \in J} \frac{n_j}{n} \left(\frac{1}{n_j} \sum_{i \in \mathcal{Z}_j} \log \frac{\tilde{h}_n^{(j)}(Z_i)}{h_0^{(j)}(Z_i)} \right) &\lesssim \mathbb{E}_{h_0} \sum_{j \in J} \frac{n_j}{n} \left(\frac{1}{n_j} \log^{5/4}(en_j) + \frac{\nu_{*,j}^{2/5}}{n_j^{4/5}} \log \frac{en_j}{\nu_{*,j}} \right) \\ &\lesssim \frac{k}{n} \log^{5/4} \left(\frac{en}{k} \right) + \frac{1}{n} \mathbb{E}_{h_0} \sum_{j \in J} n_j^{1/5} \nu_{*,j}^{2/5} \{ \log(n_j \nu_{*,j}^2) + \log(e/\nu_{*,j}^3) \}. \end{aligned} \tag{18}$$

To bound the second term of (18), observe that by two applications of Jensen's inequality,

$$\begin{aligned} \frac{1}{n} \mathbb{E}_{h_0} \sum_{j \in J} n_j^{1/5} \nu_{*,j}^{2/5} \log(n_j \nu_{*,j}^2) &\leq \frac{1}{n} \mathbb{E}_{h_0} \left\{ |J_0|^{4/5} \left(\sum_{j \in J_0} n_j \nu_{*,j}^2 \right)^{1/5} \log \left(\frac{\sum_{j \in J_0} n_j \nu_{*,j}^2}{|J_0|} \right) \right\} \\ &\leq \frac{1}{n} \mathbb{E}_{h_0} \left\{ |J_0|^{4/5} \left(\sum_{j=1}^k n_j \nu_{*,j}^2 \right)^{1/5} \log \left(\frac{\sum_{j=1}^k n_j \nu_{*,j}^2}{|J_0|} \right) \right\} \\ &\leq \frac{k^{4/5}}{n} \log \left(\frac{en}{k} \right) \mathbb{E}_{h_0} \left\{ \left(\sum_{j=1}^k n_j \nu_{*,j}^2 \right)^{1/5} \right\} \\ &\leq \frac{k^{4/5}}{n^{4/5}} \nu_*^{2/5} \log \frac{en}{k}. \end{aligned} \tag{19}$$

But

$$\begin{aligned} \nu_*^2 &= \sum_{j=1}^k p_j \nu_{*,j}^2 \leq 2 \wedge \sum_{j=1}^k p_j d_{\text{KL}}^2(h_0^{(j)}, h_*^{(j)}) \leq 2 \wedge \left\{ \sum_{j=1}^k p_j d_{\text{KL}}^2(h_0^{(j)}, h_*) + \sum_{j=1}^k p_j \log \frac{p_j}{q_j} \right\} \\ &= 2 \wedge d_{\text{KL}}^2(h_0, h_*). \end{aligned} \tag{20}$$

Moreover, by three further applications of Jensen's inequality, and using the fact that $x \mapsto \log^5 x$ is concave for $x \geq e^5$, we have

$$\begin{aligned} \frac{1}{n} \mathbb{E}_{h_0} \sum_{j \in J} n_j^{1/5} \nu_{*,j}^{2/5} \log(e/\nu_{*,j}^3) &\leq \frac{3}{n} \mathbb{E}_{h_0} \sum_{j=1}^k n_j^{1/5} \nu_{*,j}^{2/5} \log \frac{2^{1/2} e^5}{\nu_{*,j}} \\ &\leq \frac{3k^{4/5}}{n} \mathbb{E}_{h_0} \left\{ \left(\sum_{j=1}^k n_j \nu_{*,j}^2 \log^5 \frac{2^{1/2} e^5}{\nu_{*,j}} \right)^{1/5} \right\} \leq \frac{3k^{4/5}}{n^{4/5}} \left(\sum_{j=1}^k p_j \nu_{*,j}^2 \log^5 \frac{2^{1/2} e^5}{\nu_{*,j}} \right)^{1/5} \\ &\leq \frac{3k^{4/5}}{n^{4/5}} \nu_*^{2/5} \log \left(2^{1/2} e^5 \sum_{j=1}^k \frac{p_j \nu_{*,j}}{\nu_k^2} \right) \leq \frac{3k^{4/5}}{n^{4/5}} \nu_*^{2/5} \log \frac{2e^5}{\nu_*^2} \\ &\leq \frac{3k^{4/5}}{n^{4/5}} \{2 \wedge d_{\text{KL}}^2(h_0, h_*)\}^{1/5} \log \frac{2e^5}{2 \wedge d_{\text{KL}}^2(h_0, h_*)}, \end{aligned} \tag{21}$$

where the last step follows from the fact that $x \mapsto x^{1/5} \log \frac{2e^5}{x}$ is increasing for $x \leq 2$. Combining (16), (17), (18), (19), (20) and (21), the result follows in the case $k \leq e^{-1/4}n$.

Now suppose $k > e^{-1/4}n$. Then, by Theorem 8,

$$\mathbb{E}_{h_0} d_X^2(\hat{h}_n, h_0) \lesssim n^{-4/5} \lesssim \frac{k}{n} \log^{5/4} \frac{en}{k},$$

which completes the proof. \square

Proof of Proposition 10. By the scale equivariance described in Proposition 3(i), we may assume without loss of generality that $\sigma_{h_0} = 1$. Define $\nu := \inf\{d_H(h_0, h) : h \in \mathcal{H}_a^{(1)}, h_0 \ll h\} \in [0, 2^{1/2}]$. By Lemma 20, if $\delta \in (0, 2^{-9} - \nu)$, then for every $\epsilon > 0$, it holds that

$$H_{\mathbb{H}}(\epsilon, \mathcal{H}(h_0, C_\mu, C_\sigma) \cap \mathcal{H}(h_0, \delta), d_H) \leq H_{\mathbb{H}}(\epsilon, \mathcal{H}(h_0, \delta), d_H) \lesssim \frac{(\delta + \nu)^{1/2}}{\epsilon^{1/2}} \log^{5/4} \frac{1}{\delta}.$$

On the other hand, if $\delta \geq 2^{-9} - \nu$, then by Lemma 17, for every $\epsilon > 0$, we have that

$$H_{\mathbb{H}}(\epsilon, \mathcal{H}(h_0, C_\mu, C_\sigma) \cap \mathcal{H}(h_0, \delta), d_H) \leq H_{\mathbb{H}}(\epsilon, \mathcal{H}(h_0, C_\mu, C_\sigma), d_H) \lesssim \frac{1}{\epsilon^{1/2}} \lesssim \frac{(\delta + \nu)^{1/2}}{\epsilon^{1/2}}.$$

It follows that

$$\begin{aligned} \int_0^\delta H_{\mathbb{H}}^{1/2}(\epsilon, \mathcal{H}(h_0, \delta) \cap \mathcal{H}(h_0, C_\mu, C_\sigma), d_H) d\epsilon &\lesssim (\delta + \nu)^{1/4} \{\log^{5/8}(1/\delta) \vee 1\} \int_0^\delta \epsilon^{-1/4} d\epsilon \\ &\lesssim \delta^{3/4} (\delta + \nu)^{1/4} \{\log^{5/8}(1/\delta) \vee 1\}. \end{aligned}$$

Define $\Psi(\delta) := C\delta^{3/4}(\delta + \nu)^{1/4} \{\log^{5/8}(1/\delta) \vee 1\}$, where $C > 0$ is chosen such that

$$\Psi(\delta) \geq \max \left\{ \int_0^\delta H_{\mathbb{H}}^{1/2}(\epsilon, \mathcal{H}(h_0, \delta) \cap \mathcal{H}(h_0, C_\mu, C_\sigma), d_H) d\epsilon, \delta \right\},$$

Set $\delta_n := K \left\{ \frac{\nu^{2/5}}{n^{4/5}} \log \frac{en}{\nu} + \frac{1}{n} \log^{5/4}(en) \right\}^{1/2}$ for a universal constant $K > 0$ to be chosen later. Then, because $\Psi(\delta)/\delta^2$ is non-increasing, we have

$$\inf_{\delta \geq \delta_n} \frac{n^{1/2}\delta^2}{\Psi(\delta)} = \frac{n^{1/2}\delta_n^2}{\Psi(\delta_n)} = \frac{n^{1/2}\delta_n^{5/4}}{C(\delta_n + \nu)^{1/4} \{\log^{5/8}(1/\delta_n) \vee 1\}}.$$

By choosing the universal constant $K > 0$ sufficiently large, we can ensure that this ratio is larger than the universal constant required to apply Theorem 10 in the online supplement of [Kim et al. \(2017b\)](#) (a minor restatement of [van de Geer \(2000, Corollary 7.5\)](#)). We deduce from this result that there exists a universal constant $C > 0$ such that for $\delta \geq \delta_n$,

$$\mathbb{P}(d_X^2(\hat{h}_n, h_0) \geq \delta^2) \leq C \exp\left(-\frac{n\delta^2}{C}\right). \quad (22)$$

Moreover, by Lemmas 18 and 19, for $n \geq 4$,

$$\begin{aligned} \int_{16 \log n}^\infty \mathbb{P}\left(\max_{i=1,\dots,n} \log \frac{\hat{h}_n(Z_i)}{h_0(Z_i)} \geq t\right) dt &= \log n \int_{16}^\infty \mathbb{P}\left(\max_{i=1,\dots,n} \log \frac{\hat{h}_n(Z_i)}{h_0(Z_i)} \geq s \log n\right) ds \\ &\lesssim \log n \int_{16}^\infty n^{-(s/2-1)} + \left(\frac{2^8 e}{n^{s/16}}\right)^n + n^{-sn^{1/2}/256} ds \lesssim \frac{\log n}{n}. \end{aligned} \quad (23)$$

It follows from (22), (23) and Proposition 7 that

$$\begin{aligned}
\mathbb{E}d_X^2(\hat{h}_n, h_0) &\leq \int_0^{16\log n} \mathbb{P}\left(\{d_X^2(\hat{h}_n, h_0) \geq t\} \cap \{\hat{h}_n \in \mathcal{H}(h_0, C_\mu, C_\sigma)\}\right) dt \\
&\quad + 16\log n \mathbb{P}(\hat{h}_n \notin \mathcal{H}(h_0, C_\mu, C_\sigma)) + \int_{16\log n}^\infty \mathbb{P}(d_X^2(\hat{h}_n, h_0) \geq t) dt \\
&\leq \delta_n^2 + C \int_{\delta_n^2}^\infty \exp\left(-\frac{nt}{C}\right) dt + \frac{C\log n}{n} + \int_{16\log n}^\infty \mathbb{P}\left(\max_{i=1,\dots,n} \log \frac{\hat{h}_n(Z_i)}{h_0(Z_i)} \geq t\right) dt \\
&\lesssim \frac{\nu^{2/5}}{n^{4/5}} \log \frac{en}{\nu} + \frac{\log^{5/4}(en)}{n},
\end{aligned}$$

as required. \square

8 Auxiliary results

8.1 Lemmas used in results in Section 3

The mean μ_h of any $h \in \mathcal{H}_a$ is constrained because $h(r) = r^{p-1}e^{\phi(r)}$ and ϕ is decreasing. The next lemma formalises this notion.

Lemma 11. *Let $a \geq 0$ and let $h \in \mathcal{H}_a$ with $\sigma_h = 1$. For $s \in (0, \infty)$, let $\rho(s) := \frac{(a+s)^{p-1}ps}{(a+s)^p - a^p}$. Then, writing $s^* := \sup\{s \in (0, \infty) : \frac{\rho(s)}{s} \geq 2^{-7}\}$, we have*

$$a \leq \mu_h \leq a + 2^{12}\rho(s^* + 1).$$

Remark: Even though we cannot obtain an analytic expression for $\rho(s^* + 1)$, we can apply the bounds developed in Lemma 13 to control μ_h . For example, since $\rho(s) \leq p$ for any $a \geq 0$ by Lemma 13(iii), we have that $\mu_h - a \lesssim p$. When $a = 0$, this bound is sharp up to the universal constant, because taking $e^{\phi(r)} = \frac{p}{a^p} \mathbb{1}_{\{r \in [0, a]\}}$, where $a = (p+1)(p+2)^{1/2}/p^{1/2}$, yields $\mu_h = p^{1/2}(p+2)^{1/2}$.

Proof. If $h \in \mathcal{H}_a$ with $\sigma_h = 1$, then $h(r) = 0$ for $r \in (-\infty, a)$ and thus, $\mu_h \geq a$. For the upper bound on μ_h , we first observe that $h(\mu_h) \geq 2^{-7}$ by Lovász and Vempala (2007, Theorem 5.14(d)). Since $\rho(s)/s$ is decreasing by Lemma 13(ii) and $1 \leq \rho(s) \leq p$ by Lemma 13(iii), we have that $s^* \in (0, \infty)$. Suppose for a contradiction that $\mu_h > a + 2^{12}\rho(s^* + 1)$. Then, since ρ is increasing by Lemma 13(i), we have $\frac{\rho(s+1)}{\mu_h - a} < 2^{-12}$ for all $s \leq s^*$. Moreover, by definition of s^* , we have $\rho(s)/s < 2^{-7}$ for all $s > s^*$. Hence $\sup_{s \in (0, \infty)} \min\left(24\frac{\rho(s)}{\mu_h - a}, \frac{\rho(s)}{s}\right) < 2^{-7}$. But then Lemma 12 establishes a contradiction, so $\mu_h \leq a + 2^{12}\rho(s^* + 1)$, as desired. \square

The next lemma is used in the proof of Lemma 11.

Lemma 12. For any $a \geq 0$, $h \in \mathcal{H}_a$ and any $s \in (0, \infty)$,

$$h(a+s) \leq \min \left\{ 16 \frac{\rho(s+1)}{\mu_h - a}, \frac{\rho(s)}{s} \right\}.$$

Proof. Let us fix $h \in \mathcal{H}_a$ and define $\tilde{h}(s) := h(a+s)$. Observe that \tilde{h} is a density of the form $\tilde{h}(s) = e^{\phi(s)}(a+s)^{p-1}$ for some $\phi \in \Phi_0$. We will show that $\tilde{h}(s) \leq \min \left\{ 16 \frac{\rho(s+1)}{\mu_{\tilde{h}}}, \frac{\rho(s)}{s} \right\}$ for all $s \in [0, \infty)$. The lemma then follows because $\mu_{\tilde{h}} = \mu - a$.

To this end, fix $s_0 \in (0, \infty)$ and define $\tilde{h}_{s_0}(s) := \alpha(s+a)^{p-1} \mathbb{1}_{\{s \in [0, s_0]\}}$ where $\alpha^{-1} := \int_0^{s_0} (a+s)^{p-1} ds = \frac{(a+s_0)^p}{p} - \frac{a^p}{p}$. Then

$$1 = \alpha \int_0^{s_0} (a+s)^{p-1} ds \geq \int_0^{s_0} e^{\phi(s)}(a+s)^{p-1} ds \geq e^{\phi(s_0)} \int_0^{s_0} (a+s)^{p-1} ds.$$

Thus, we see that $e^{\phi(s_0)} \leq \alpha$ and therefore $\tilde{h}(s_0) \leq \tilde{h}_{s_0}(s_0) = \alpha(s_0+a)^{p-1} \leq \rho(s_0)/s_0$, as desired.

To prove the second part of the bound, observe that if $s_0 \geq \mu_{\tilde{h}}/8$, then $\frac{\rho(s_0)}{s_0} \leq 8 \frac{\rho(s_0)}{\mu_{\tilde{h}}}$ and lemma follows. We therefore fix $s_0 \in (0, \mu_{\tilde{h}}/8)$, and also define $M := \log 4 \vee \sup_{s \in [0, \infty)} \log \tilde{h}(s)$ and fix $m \in (-\infty, M-2]$. For $t \in [m, M]$, let $D_t := \{s \in [0, \infty) : \log \tilde{h}(s) \geq t\}$. Since \tilde{h} is itself a log-concave density, we have that, any $t \in [m, M]$ and $s \in D_m$,

$$\log \tilde{h} \left(\frac{t-m}{M-m} s_0 + \frac{M-t}{M-m} s \right) \geq \frac{(t-m)M}{M-m} + \frac{(M-t)m}{M-m} = t.$$

Hence, writing λ for Lebesgue measure on \mathbb{R} ,

$$\lambda(D_t) \geq \lambda \left(\frac{t-m}{M-m} s_0 + \frac{M-t}{M-m} D_m \right) = \frac{M-t}{M-m} \lambda(D_m).$$

Using Fubini's theorem as in [Dümbgen et al. \(2011\)](#), we can now compute

$$\begin{aligned} 1 &\geq \int_{D_m} \tilde{h}(s) - e^m ds \geq \int_{D_m} \int_m^M e^t \mathbb{1}_{\{\log \tilde{h}(s) \geq t\}} dt ds \\ &= \int_m^M e^t \lambda(D_t) dt \geq \frac{\lambda(D_m)}{M-m} \int_m^M (M-t) e^t dt = \frac{\lambda(D_m) e^M}{M-m} \int_0^{M-m} x e^{-x} dx \\ &\geq \frac{\lambda(D_m) e^M}{2(M-m)}. \end{aligned}$$

Since D_m is an interval containing s_0 , we conclude that $\log \tilde{h}(s) \leq m$ whenever $|s-s_0| > 2(M-m)e^{-M}$. Thus, for $|s-s_0| \geq 4e^{-M}$, we have

$$\log \tilde{h}(s) \leq \inf \{m \in (-\infty, M-2] : 2(M-m)e^{-M} < |s-s_0|\} = M - \frac{|s-s_0|e^M}{2}.$$

We now consider two cases. If $s_0 - 4e^{-M} > 0$, then using the bound $\tilde{h}(s) \leq \rho(s)/s$, Lemma 13(i) and the fact that $e^M \geq 4$, we have

$$\begin{aligned}\mu_{\tilde{h}} &\leq \int_0^\infty s\tilde{h}(s) ds \leq \int_0^{s_0-4e^{-M}} s \exp\left\{M - \frac{(s_0-s)e^M}{2}\right\} ds + \int_{s_0-4e^{-M}}^{s_0+4e^{-M}} \rho(s) ds \\ &\quad + \int_{s_0+4e^{-M}}^\infty s \exp\left\{M - \frac{(s-s_0)e^M}{2}\right\} ds \\ &\leq 2 \int_2^\infty \left(s_0 - \frac{2t}{e^M}\right) e^{-t} dt + 8e^{-M} \rho(s_0+1) + 2 \int_2^\infty \left(s_0 + \frac{2t}{e^M}\right) e^{-t} dt \\ &\leq 4s_0 + 8e^{-M} \rho(s_0+1).\end{aligned}$$

Since $4s_0 < \mu_{\tilde{h}}/2$, we conclude that $e^M \leq 16 \frac{\rho(s_0+1)}{\mu_{\tilde{h}}}$.

Now, suppose $s_0 - 4e^{-M} \leq 0$. Then, similarly,

$$\begin{aligned}\mu_{\tilde{h}} &\leq \int_0^\infty s\tilde{h}(s) ds \leq \int_0^{s_0+4e^{-M}} \rho(s) ds + \int_{s_0+4e^{-M}}^\infty s \exp\left\{M - \frac{(s-s_0)e^M}{2}\right\} ds \\ &\leq 8e^{-M} \rho(s_0+1) + 2 \int_2^\infty \left(s_0 + \frac{2t}{e^M}\right) e^{-t} dt \\ &\leq 2s_0 + 4e^{-M} \{2\rho(s_0+1) + 1\}.\end{aligned}$$

Using the facts that $s_0 \leq \mu_{\tilde{h}}/8$ and $\rho(s_0+1) \geq 1$ yields the desired upper bound on e^M . \square

The next lemma provides basic properties of the function ρ defined in Lemma 11.

Lemma 13. *For any $p \in \mathbb{N}$ and $a \geq 0$, we have*

- (i) $\rho(s)$ is increasing;
- (ii) $\frac{\rho(s)}{s}$ is decreasing;
- (iii) $1 \leq \rho(s) \leq p$ for all $s \in (0, \infty)$;
- (iv) $\frac{\rho(s+1)}{\rho(s)} \leq 20$ for $s \geq 1$.

Proof. (i) Define $\alpha := (a+s)/a \in (1, \infty)$. Then

$$\rho(s) = \frac{(a+s)^{p-1}ps}{(a+s)^p - a^p} = \frac{\alpha^{p-1}p(\alpha-1)}{\alpha^p - 1} = p\left(1 - \frac{\alpha^{p-1} - 1}{\alpha^p - 1}\right),$$

which is increasing in α , as required.

(ii) If $a > 0$, then

$$\frac{\rho(s)}{s} = \frac{p}{a} \frac{\alpha^{p-1}}{\alpha^p - 1},$$

which is decreasing in α . If $a = 0$, then $\rho(s)/s = p/s$ and the claim also follows.

(iii) The lower bound follows from (i) and the fact that $\lim_{s \searrow 0} \rho(s) = 1$, while the upper bound follows from (i) and the fact that $\lim_{s \rightarrow \infty} \rho(s) = p$.

(iv) We have

$$\begin{aligned} \frac{\rho(s+1)}{\rho(s)} &= \frac{(a+s+1)^{p-1}}{(a+s)^{p-1}} \frac{s+1}{s} \frac{(a+s)^p - a^p}{(a+s+1)^p - a^p} \\ &\leq 2 \left(1 + \frac{1}{a+s}\right)^{p-1} \left\{ \frac{1 - (\frac{a}{a+s})^p}{(1 + \frac{1}{a+s})^p - (\frac{a}{a+s})^p} \right\}. \end{aligned}$$

Hence, if $a+s \geq p/2$, then $\rho(s+1)/\rho(s) \leq 2e^2 \leq 20$. On the other hand, if $a+s < p/2$, then $(1 + \frac{1}{a+s})^p - (\frac{a}{a+s})^p \geq \frac{1}{2}(1 + \frac{1}{a+s})^p$ and the claim follows again. \square

Lemma 14. *For any $p \geq 1$ and $r_0 > a \geq 0$, we have that*

$$r_0 \leq \frac{p}{p+1} \frac{r_0^{p+1} - a^{p+1}}{r_0^p - a^p} + \frac{r_0^p - a^p}{r_0^{p-1} p}$$

Proof. Writing $x := 1 - (a/r_0)^p$, we are required to prove that for $x \in (0, 1]$,

$$1 \leq \frac{p}{p+1} \frac{1 - (1-x)^{\frac{p+1}{p}}}{x} + \frac{x}{p}. \quad (24)$$

The inequality holds when $x \searrow 0$ and at $x = 1$. To finish the proof, it suffices to show that $t_p(x) := \frac{1 - (1-x)^{(p+1)/p}}{x}$ is concave on $(0, 1)$. But

$$\begin{aligned} p^2(1-x)x^3t_p''(x) &= 2p^2\{1 - (1-x)^{1/p}\}(1-x)^{1-1/p} - p(2-x)x - x^2 \\ &\leq 2px\{1 - (1-1/p)x\} - p(2-x)x - x^2 = -(p-1)x^2 \leq 0, \end{aligned}$$

as required. \square

Lemma 15. *Let $h \in \mathcal{F}_1$.*

- (i) *There exists a universal constant $c > 0$ such that $\sigma_h \geq c/h(\mu_h)$;*
- (ii) *There exists a universal constant $C > 0$ such that $\sigma_h \leq C/\sup_{r \in \mathbb{R}} h(r)$.*

Proof. (i) Let $f(r) := \sigma_h h(\sigma_h r + \mu_h)$, so f is a log-concave density with $\mu_f = 0$ and $\sigma_f = 1$. Then

$$\sigma_h = \frac{f(0)}{h(\mu_h)} \geq \frac{2^{-7}}{h(\mu_h)},$$

where the final inequality follows from Lovász and Vempala (2007, Theorem 5.14(d)).

(ii) Since h is upper semi-continuous, there exists $r_0 \in \mathbb{R}$ such that $h(r_0) = \sup_{r \in \mathbb{R}} h(r)$. Thus

$$\sigma_h \leq \frac{f(r_0/\sigma_h)}{h(r_0)} \leq \frac{\|f\|_\infty}{\sup_{r \in \mathbb{R}} h(r)} \leq \frac{2^7}{\sup_{r \in \mathbb{R}} h(r)},$$

by Lovász and Vempala (2007, Theorem 5.14(b) and (d)). \square

Lemma 16. For any $f \in \mathcal{F}_1$ with $\sigma_f = 1$, we have that

$$-\frac{1}{2} - \frac{1}{2} \log(2\pi) \leq \int_{-\infty}^{\infty} f \log f \leq 9 \log 2.$$

Proof. By the location invariance of the entropy functional, we may assume without loss of generality that $\int_{-\infty}^{\infty} xf(x) dx = 0$. Since $\|f\|_{\infty} \leq 2^9$ (Lovász and Vempala, 2007, Theorem 5.14(b) and (d)), we have

$$\int_{-\infty}^{\infty} f \log f \leq 9 \log 2.$$

Now let $g(x) := (2\pi)^{-1/2} e^{-x^2/2}$, so by the non-negativity of Kullback–Leibler divergence,

$$\int_{-\infty}^{\infty} f \log f \geq \int_{-\infty}^{\infty} f \log g = -\frac{1}{2} \int_{-\infty}^{\infty} x^2 f(x) dx - \frac{1}{2} \log(2\pi) = -\frac{1}{2} - \frac{1}{2} \log(2\pi),$$

as required. \square

8.2 Lemmas used in both Theorem 8 and Proposition 10

The next lemma is a very slight generalisation of (Kim and Samworth, 2016, Theorem 4) and can be proved in the same manner, with minor modifications to handle the general mean and variance perturbation.

Lemma 17. *Kim and Samworth (2016, Theorem 4) Fix $C_{\mu}, C_{\sigma} > 0$ and $h_0 \in \mathcal{H}_a$. There exists $C > 0$, depending only on C_{μ}, C_{σ} , such that for every $\epsilon > 0$,*

$$H_{\|}(\epsilon, \mathcal{H}(h_0, C_{\mu}, C_{\sigma}), d_H) \leq \frac{C}{\epsilon^{1/2}}.$$

Lemma 18. *Let $a \geq 0$ and let $Z_1, \dots, Z_n \stackrel{\text{iid}}{\sim} h_0 \in \mathcal{H}_a$ with $\sigma_{h_0} = 1$. Then for $n \geq 4$ and $t \geq 8$,*

$$\mathbb{P}\left(\sup_{r \geq a} \log \hat{h}_n(r) > t \log n\right) \leq \left(\frac{2^8 e}{n^{t/8}}\right)^n + 2n^{-tn^{1/2}/128}.$$

Proof. Let \mathbb{Q}_n denote the empirical distribution of Z_1, \dots, Z_n . Let E_1 be the event that $\mathbb{Q}_n(A) \leq 1/2$ for every interval A of length at most $6n^{-t/2}$. Since $\|h_0\|_{\infty} \leq 2^9$ (Lovász and Vempala, 2007, Theorem 5.14(b) and (d)), we have that

$$\begin{aligned} \mathbb{P}(E_1^c) &= \mathbb{P}\left(\bigcup_{i=1}^n \bigcup_{\substack{A \subset \{1, \dots, n\} \setminus \{i\} \\ |A|=\lfloor n/2 \rfloor}} \bigcap_{j \in A} \{Z_j \in [Z_i, Z_i + 6n^{-t/2}]\}\right) \\ &\leq n \binom{n-1}{\lfloor n/2 \rfloor} \mathbb{P}\left(\bigcap_{j=2}^{\lfloor n/2 \rfloor + 1} \{Z_j \in [Z_1, Z_1 + 6n^{-t/2}]\}\right) \\ &\leq n \binom{n-1}{\lfloor n/2 \rfloor} (2^9 6n^{-t/2})^{\lfloor n/2 \rfloor} \leq \left(\frac{2^8 e}{n^{t/8}}\right)^n. \end{aligned}$$

Now let $E_2 := \{\int_a^\infty \log h_0 d\mathbb{Q}_n > -(t/2) \log n\}$, so by Lemma 16 and Bobkov and Madiman (2011, Theorem 1.1),

$$\begin{aligned}\mathbb{P}(E_2^c) &\leq \mathbb{P}\left(\left|\int h_0(r) \log h_0(r) dr - \int \log h_0 d\mathbb{Q}_n\right| \geq -\frac{1}{2} - \frac{1}{2} \log(2\pi) + \frac{t}{2} \log n\right) \\ &\leq \mathbb{P}\left(\left|\int h_0(r) \log h_0(r) dr - \int \log h_0 d\mathbb{Q}_n\right| \geq \frac{t}{8} \log n\right) \leq 2n^{-tn^{1/2}/128}.\end{aligned}$$

But on $E_1 \cap E_2$, and applying Proposition 6 with $\delta_{\mathbb{Q}_n} = 6n^{-t/2}$ and $\ell_{\hat{h}_n} = -(t/2) \log n$, we find that

$$\sup_{r \geq a} \log \hat{h}_n(r) \leq \max\left\{-2 \log\left(\frac{\delta_{\mathbb{Q}_n}}{6}\right), -2\ell_{\hat{h}_n}, 3\right\} \leq t \log n.$$

The result follows. \square

Lemma 19. *Let $a \geq 0$, let $h_0 \in \mathcal{H}_a$ with $\sigma_{h_0} = 1$, and suppose that $Z_1, \dots, Z_n \stackrel{\text{iid}}{\sim} h_0$. Then there exists a universal constant $C > 0$ such that for any $n \geq 2$ and $t \geq 4$,*

$$\mathbb{P}\left(\sup_{r \in [Z_{(1)}, Z_{(n)}]} \log \frac{1}{h_0(r)} \geq t \log n\right) \leq C n^{-(t-1)}.$$

Proof. This result follows from the proof of Kim et al. (2017b, Lemma 2). \square

8.3 Lemmas used in Proposition 10

For $h_0 \in \mathcal{H}_a$ and $\delta > 0$, let

$$\mathcal{H}(h_0, \delta) := \{h \in \mathcal{H}_a : h \ll h_0, d_H(h, h_0) \leq \delta\}.$$

Lemma 20. *Fix $a \geq 0$ and let $h_0 \in \mathcal{H}_a$. Assume that $\nu := \inf\{d_H(h_0, h) : h \in \mathcal{H}_a^{(1)}, h_0 \ll h\} < 2^{-9}$. Then there exists a universal constant $C > 0$ such that, for all $\delta \in (0, 2^{-9} - \nu)$ and all $\epsilon > 0$,*

$$H_{\mathbb{H}}(\epsilon, \mathcal{H}(h_0, \delta), d_H) \leq C \left(\frac{\delta + \nu}{\epsilon}\right)^{1/2} \log^{5/4} \frac{1}{\delta}.$$

Proof. Fix $h_1 \in \mathcal{H}^{(1)}$ with $h_0 \ll h_1$ and let $\nu_1 := d_H(h_0, h_1)$. Then, by the triangle inequality,

$$\mathcal{H}(h_0, \delta) \subseteq \mathcal{H}(h_1, \delta + \nu_1).$$

Since $h_1(r) = r^{p-1} e^{\phi_1(r)}$ where ϕ_1 is affine, $\log(h/h_1)$ is concave for any $h \in \mathcal{H}(h_1, \delta + \nu_1)$. We therefore have

$$\mathcal{H}(h_1, \delta + \nu_1) \subseteq \tilde{\mathcal{F}}(h_1, \delta + \nu_1)$$

where the right-hand side is defined in (25). The result then follows from Lemma 22. \square

Lemma 21. *Let $f, g \in \mathcal{F}_1$. Then there exist universal constants C'_μ, C'_σ such that for all $\delta \in (0, 2^{-8}]$, if $d_H(f, g) \leq \delta$, then*

$$\frac{1}{C'_\sigma} \leq \frac{\sigma_g}{\sigma_f} \leq C'_\sigma, \quad |\mu_g - \mu_f| \leq C'_\mu.$$

Proof. Since the Hellinger distance is affine invariant, we may assume without loss of generality that $\mu_f = 0$ and $\sigma_f = 1$. By Lovász and Vempala (2007, Theorem 5.14(d)), $f(x) \geq 2^{-10}$ for $x \in [-1/9, 1/9]$. We claim that $g(x) \geq 2^{-12}$ for some $x \in [-1/9, 1/9]$. To see this, suppose for a contradiction that $g(x) \leq 2^{-12}$ for all $x \in [-1/9, 1/9]$. Then,

$$\int_{-\infty}^{\infty} (f^{1/2} - g^{1/2})^2 \geq \int_{-1/9}^{1/9} (2^{-5} - 2^{-6})^2 dx \geq (2/9)2^{-12} > 2^{-16} \geq \delta^2,$$

a contradiction. By Lemma 15, it follows that $\sigma_g \leq C'_\sigma$ for some universal constant $C'_\sigma > 0$. The lower bound on σ_g follows by symmetry.

Now assume without loss of generality that $\mu_g \geq 0$. By the first part and Fresen (2013, Lemma 13), there exist universal constants $\alpha > 0$ and $\beta \in \mathbb{R}$ such that $g(x) \leq e^{-\alpha\mu_g + \beta}$ for all $x \leq 0$. It follows that if $\mu_g \geq (\beta + 12 \log 2)/\alpha$, then

$$\int_{-\infty}^{\infty} (f^{1/2} - g^{1/2})^2 \geq \int_{-1/9}^0 (2^{-5} - 2^{-6})^2 dx \geq (1/9)2^{-12} > 2^{-16},$$

a contradiction. The result follows. \square

We prove a general result on the local bracketing entropy of log-concave densities. Recall from the introduction that \mathcal{F}_p denotes the class of all upper semi-continuous, log-concave densities on \mathbb{R}^p . Let $f_0 \in \mathcal{F}_1$. We make no assumption on the support of f_0 . We define

$$\tilde{\mathcal{F}}(f_0, \delta) := \{f \in \mathcal{F}_1 : \log(f/f_0) \text{ concave}, f \ll f_0, d_H(f, f_0) \leq \delta\}, \quad (25)$$

where we adopt the convention that $0/0 := 0$.

Lemma 22. *Suppose that $\delta \in (0, 2^{-9}]$ and that $f_0 \in \mathcal{F}_1$ with $\sigma_{f_0} = 1$. Then, for every $\epsilon > 0$,*

$$H_{\mathbb{D}}(\epsilon, \tilde{\mathcal{F}}(f_0, \delta), d_H) \lesssim \frac{\delta^{1/2}}{\epsilon^{1/2}} \log^{5/4} \frac{1}{\delta}.$$

Proof. In this proof, we let C be a generic universal constant whose value may vary from instance to instance. Also, without loss of generality, let us suppose that $\mu_{f_0} = 0$. Since $f_0(0) \geq 2^{-7}$ (Lovász and Vempala, 2007, Theorem 5.14(d)), we may define $a_L := \inf\{r \in \mathbb{R} : f_0(r) \geq \delta^2\}$ and $a_R := \sup\{r \in \mathbb{R} : f_0(r) \geq \delta^2\}$. By Lemma 21 and Fresen (2013, Lemma 13), there exist $\alpha > 0$ and $\beta \in \mathbb{R}$ such that for any $f \in \tilde{\mathcal{F}}(f_0, \delta)$

and $x \in \mathbb{R}$, we have $f(x) \leq e^{-\alpha|x|+\beta}$. Let $b_L := -\frac{1}{\alpha} \log \frac{e^\beta}{\delta^4}$ and $b_R := \frac{1}{\alpha} \log \frac{e^\beta}{\delta^4}$. Then, for any $f \in \tilde{\mathcal{F}}(f_0, \delta)$, $f(r) < \delta^4$ for $r \in (-\infty, b_L) \cup (b_R, \infty)$, and $[a_L, a_R] \subseteq [b_L, b_R]$.

First, we will bracket the region $[b_L, a_L] \cup (a_R, b_R]$. To this end, fix $\epsilon > 0$, and let $K_L := \min\{k \in \mathbb{N} : e^{-\alpha(k-a_L)+\beta} \leq \delta^4\}$ and $K_R := \min\{k \in \mathbb{N} : e^{-\alpha(k+a_R)+\beta} \leq \delta^4\}$, so that $\max(K_L, K_R) \lesssim \log(1/\delta)$. By these definitions, $a_L - K_L \leq b_L$ and $a_R + K_R \geq b_R$. We segment $[b_L, a_L]$ into subintervals S_k for $k = 1, \dots, K_L$, where

$$S_k := [a_L - k, a_L - (k-1)], \quad k = 1, \dots, K_L - 1, \\ S_{K_L} := [b_L, a_L - (K_L - 1)].$$

For $k = 1, \dots, K$, define $\epsilon_k^2 := \frac{\epsilon^2}{16K_L}$. For any $r \in [b_L, a_L]$, we have that $f_0(r) \leq \delta^2$ because $r < a_L$ and, moreover, $e^{-\alpha|r|+\beta} \geq \delta^4$ because $r \geq b_L$. Hence, by Lemma 24, $f(r) \leq f_0(r)e^{C\delta \log \frac{1}{\delta}} \lesssim f_0(r) \leq \delta^2$ for any $f \in \tilde{\mathcal{F}}(f_0, \delta)$ and $r \in [b_L, a_L]$. Hence, by Lemma 27,

$$H_{\mathbb{I}}\left(\frac{\epsilon}{4}, \tilde{\mathcal{F}}(f_0, \delta), d_{\mathbb{H}}, [b_L, a_L]\right) \leq \sum_{k=1}^{K_L} H_{\mathbb{I}}(\epsilon_k, \tilde{\mathcal{F}}(f_0, \delta), d_{\mathbb{H}}, S_k) \\ \lesssim \sum_{k=1}^{K_L} \frac{\delta^{1/2}}{\epsilon_k^{1/2}} \lesssim \frac{\delta^{1/2}}{\epsilon^{1/2}} \log^{5/4} \frac{1}{\delta}. \quad (26)$$

By symmetry, we obtain the same bound for $H_{\mathbb{I}}(\epsilon/4, \tilde{\mathcal{F}}(f_0, \delta), d_{\mathbb{H}}, [a_R, b_R])$.

Now we bracket the region $(-\infty, b_L) \cup (b_R, \infty)$. For $k \in \mathbb{N}$, define $S_k := [b_L - k, b_L - (k-1)]$ and set $\epsilon_k^2 := C\epsilon^2 e^{-\alpha(k-b_L)/2}$ where C is a constant chosen such that $\sum_{k=1}^{\infty} \epsilon_k^2 = \epsilon^2/16$. Then, by Lemma 27 again,

$$H_{\mathbb{I}}(\epsilon/4, \tilde{\mathcal{F}}(f_0, \delta), d_{\mathbb{H}}, (-\infty, b_L)) \leq \sum_{k=1}^{\infty} H_{\mathbb{I}}(\epsilon_k, \tilde{\mathcal{F}}(f_0, \delta), d_{\mathbb{H}}, S_k) \lesssim \sum_{k=1}^{\infty} \frac{e^{-\alpha(k-b_L-1)/4}}{\epsilon_k^{1/2}} \\ \lesssim \frac{1}{\epsilon^{1/2}} \sum_{k=1}^{\infty} e^{-\alpha(k-b_L-1)/8} \lesssim \frac{\delta^{1/2}}{\epsilon^{1/2}}. \quad (27)$$

The same bound holds for $H_{\mathbb{I}}(\epsilon/4, \tilde{\mathcal{F}}(f_0, \delta), d_{\mathbb{H}}, [b_R, \infty))$.

Next, we bracket the region $[a_L, -1/16]$, and recall that by Lemma 25, $a_L \leq -1/9$. To this end, we write $s_0 := a_L$ and partition $[s_0, -1/16]$ into segments $[s_0, s_1], [s_1, s_2], \dots, [s_{J-1}, s_J]$ (where $s_J = -1/16$) as follows:

1. Choose $s_1 > s_0$ such that $\int_{s_0}^{s_1} f_0(t) dt = 4\delta^2$.
2. For each $j \geq 2$, if there exists $t_0 < -1/16$ such that $\int_{s_{j-1}}^{t_0} f_0(t) dt \geq 2 \int_{-\infty}^{s_{j-1}} f_0(t) dt$, then choose s_j such that $\int_{s_{j-1}}^{s_j} f_0(t) dt = 2 \int_{-\infty}^{s_{j-1}} f_0(t) dt$. Otherwise, set $J := j$ and choose $s_J = -1/16$.

Let $\phi_0(r) := \log f(r)$ and write $\text{Range}_j(\phi_0) := \sup_{t \in [s_{j-1}, s_j]} \phi_0(t) - \inf_{t \in [s_{j-1}, s_j]} \phi_0(t)$. We make the following six claims:

- (1) $s_1 < -1/16$;
- (2) $(s_1 - s_0) \sup_{t \in [s_0, s_1]} f_0(t) \lesssim \delta^2 \log \frac{1}{\delta}$;
- (3) $\int_{s_J}^{\infty} f_0(t) dt > 2^{-11}$;
- (4) $(s_j - s_{j-1}) \sup_{t \in [s_{j-1}, s_j]} f_0(t) \lesssim \text{Range}_j(\phi_0) \int_{-\infty}^{s_{j-1}} f_0(t) dt$ for $j = 2, \dots, J-1$;
- (5) $\int_{-\infty}^{s_J} f_0(t) dt \geq 2^{-13}$;
- (6) $J \lesssim \log(1/\delta)$.

To verify claim (1), observe by Lovász and Vempala (2007, Theorem 5.14(a)) that $f_0(t) \geq 2^{-8}$ for all $t \in [-1/9, 1/9]$. Hence $\int_{s_0}^{-1/16} f_0(t) dt \geq (1/9 - 1/16)2^{-8} > 4\delta^2$, so $s_1 < -1/16$.

For claim (2), note that $-2 \log(1/\delta) \leq \phi_0(t) \leq \beta$ for $t \in [s_0, s_1]$. Thus by the second part of Lemma 26 and the definition of s_1 , we have $(s_1 - s_0) \sup_{t \in [s_0, s_1]} f_0(t) \lesssim \log(1/\delta) \int_{s_0}^{s_1} f_0(t) dt \lesssim \delta^2 \log(1/\delta)$.

For claim (3), we have $\int_{s_J}^{\infty} f_0(t) dt \geq \int_{-1/16}^{1/9} f_0(t) dt \geq 2^{-8}(1/9 + 1/16) > 2^{-11}$.

For claim (4), observe that $2 \int_{-\infty}^{s_{j-1}} f_0(t) dt = \int_{s_{j-1}}^{s_j} f_0(t) dt$ for $j = 2, \dots, J-1$.

Hence

$$\frac{(s_j - s_{j-1}) \sup_{t \in [s_{j-1}, s_j]} f_0(t)}{\int_{-\infty}^{s_{j-1}} f_0(t) dt} = \frac{(s_j - s_{j-1}) \sup_{t \in [s_{j-1}, s_j]} f_0(t)}{2 \int_{s_{j-1}}^{s_j} f_0(t) dt} \lesssim \text{Range}_j(\phi_0),$$

where the final bound follows from Lemma 26.

For claim (5), we have

$$\int_{-\infty}^{s_J} f_0(t) dt \geq \int_{-1/9}^{-1/16} f_0(t) dt \geq 2^{-8}(1/9 - 1/16) \geq 2^{-13}.$$

Finally, for claim (6), we have

$$\int_{-\infty}^{s_J} f_0(t) dt = \int_{-\infty}^{s_{j-1}} f_0(t) dt + \int_{s_{j-1}}^{s_J} f_0(t) dt = 3 \int_{-\infty}^{s_{j-1}} f_0(t) dt$$

for $j = 2, \dots, J-1$. Since $\int_{-\infty}^{s_1} f_0(t) dt \geq \int_{s_0}^{s_1} f_0(t) dt = 2\delta^2$ by definition of s_1 and since $\int_{-\infty}^{\infty} f_0(t) dt = 1$, we conclude that $J \lesssim \log(1/\delta)$.

Now set $\tilde{\epsilon} := \epsilon/(2J^{1/2})$. Then, by Lemma 27 and claim (2),

$$\begin{aligned} H_{\mathbb{H}}(\tilde{\epsilon}, \tilde{\mathcal{F}}(f_0, \delta), d_{\mathbb{H}}, [s_0, s_1]) &\lesssim \frac{(s_1 - s_0)^{1/4}}{\tilde{\epsilon}^{1/2}} \sup_{e^{\phi} \in \tilde{\mathcal{F}}(f_0, \delta)} \sup_{t \in [s_0, s_1]} e^{\phi(t)/4} \\ &\lesssim \frac{\delta^{1/2}}{\tilde{\epsilon}^{1/2}} \log^{1/4}(1/\delta) \sup_{e^{\phi} \in \tilde{\mathcal{F}}(f_0, \delta)} \sup_{t \in [s_0, s_1]} e^{(\phi(t) - \phi_0(t))/4} \\ &\lesssim \frac{\delta^{1/2}}{\epsilon^{1/2}} \log^{1/2} \frac{1}{\delta}, \end{aligned} \tag{28}$$

where the final bound follows from Lemma 24.

Now let $j \in \{2, \dots, J\}$, and observe by claim (1) that s_1, \dots, s_J are strictly increasing. Let $\tilde{\mathcal{F}}(f_0, \delta) := \{e^{\phi-\phi_0} : e^\phi \in \tilde{\mathcal{F}}(f_0, \delta)\}$. If $\{(\check{\psi}_j^L, \check{\psi}_j^U) : j = 1, \dots, N\}$ is an $\tilde{\epsilon}$ -Hellinger bracketing set for $\tilde{\mathcal{F}}(f_0, \delta)$ with $\log N = H_{\mathbb{I}}(\epsilon, \tilde{\mathcal{F}}(f_0, \delta), d_H, [s_{j-1}, s_j])$, then we can define $\{(\tilde{\psi}_j^L, \tilde{\psi}_j^U) : j = 1, \dots, N\}$ by $\tilde{\psi}_j^L := f_0 \check{\psi}_j^L$ and $\tilde{\psi}_j^U := f_0 \check{\psi}_j^U$. Then

$$d_H(\tilde{\psi}_j^L, \tilde{\psi}_j^U) \leq \sup_{t \in [s_{j-1}, s_j]} f_0(t)^{1/2} d_H(\check{\psi}_j^L, \check{\psi}_j^U) \leq \tilde{\epsilon} \sup_{t \in [s_{j-1}, s_j]} f_0(t)^{1/2}.$$

Moreover, on $[s_{j-1}, s_j]$, the conditions of Lemma 23 are fulfilled because $\int_{-\infty}^{s_{j-1}} f_0(t) dt \geq \int_{s_0}^{s_1} f_0(t) dt = 4\delta^2$ and $\int_{s_j}^{\infty} f_0(t) dt > 2^{-11} \geq 4\delta^2$ by claim (3). Thus, we may combine Lemmas 23 and 24 with claim (4) and Lemma 28 to obtain

$$\begin{aligned} & H_{\mathbb{I}}(\tilde{\epsilon}, \tilde{\mathcal{F}}(f_0, \delta), d_H, [s_{j-1}, s_j]) \\ & \leq H_{\mathbb{I}}\left(\frac{\tilde{\epsilon}}{\sup_{t \in [s_{j-1}, s_j]} e^{\phi_0(t)/2}}, \tilde{\mathcal{F}}(f_0, \delta), d_H, [s_{j-1}, s_j]\right) \\ & \lesssim \frac{\sup_{t \in [s_{j-1}, s_j]} e^{\phi_0(t)/4}}{\tilde{\epsilon}^{1/2}} \left[|s_{j-1}| \delta + \frac{\delta}{\left\{ \int_{-\infty}^{s_{j-1}} f_0(t) dt \right\}^{1/2}} \right]^{1/2} (s_j - s_{j-1})^{1/4} \sup_{t \in [s_{j-1}, s_j]} e^{(\phi(t) - \phi_0(t))/4} \\ & \lesssim \frac{\delta^{1/2}}{\tilde{\epsilon}^{1/2}} \text{Range}_j^{1/4}(\phi_0) \left[|s_{j-1}|^{1/2} \left\{ \int_{-\infty}^{s_{j-1}} f_0(t) dt \right\}^{1/4} + 1 \right] \lesssim \frac{\delta^{1/2}}{\tilde{\epsilon}^{1/2}} \text{Range}_j^{1/4}(\phi_0), \end{aligned} \quad (29)$$

where the final bound follows because $s_{j-1}^2 \int_{-\infty}^{s_{j-1}} f_0(t) dt \leq 1$ by Markov's inequality.

By symmetry, we obtain the same bracketing entropy bound over $[1/16, a_R]$. For the region $[-1/16, 1/16]$, since $\int_{-\infty}^{-1/16} f_0(t) dt \geq 2^{-13}$ and $\int_{1/16}^{\infty} f_0(t) dt \geq 2^{-13}$ by claim (5), so arguing as in (29), we obtain

$$H_{\mathbb{I}}\left(\frac{\epsilon}{2}, \tilde{\mathcal{F}}(f_0, \delta), d_H, [-1/16, 1/16]\right) \lesssim \frac{\delta^{1/2}}{\epsilon^{1/2}}. \quad (30)$$

Now, by Jensen's inequality and the fact that ϕ_0 is unimodal,

$$\sum_{j=2}^J \text{Range}_j^{1/4}(\phi_0) \lesssim J \left(\frac{\log(1/\delta)}{J} \right)^{1/4} \lesssim \log(1/\delta). \quad (31)$$

We conclude from (31), (30), (29), (28), (27), (26) and claim (6) that

$$\begin{aligned} & H_{\mathbb{I}}(\epsilon, \tilde{\mathcal{F}}(f_0, \delta), d_H) \\ & \leq H_{\mathbb{I}}\left(\frac{\epsilon}{2}, \tilde{\mathcal{F}}(f_0, \delta), d_H, [-1/16, 1/16]\right) + 2 \sum_{j=1}^J H_{\mathbb{I}}\left(\frac{\epsilon}{2J^{1/2}}, \tilde{\mathcal{F}}(f_0, \delta), d_H, [s_{j-1}, s_j]\right) \\ & \quad + 2H_{\mathbb{I}}\left(\frac{\epsilon}{4}, \tilde{\mathcal{F}}(f_0, \delta), d_H, [b_L, a_L]\right) + 2H_{\mathbb{I}}\left(\frac{\epsilon}{4}, \tilde{\mathcal{F}}(f_0, \delta), d_H, (-\infty, b_L)\right) \\ & \lesssim \frac{\delta^{1/2}}{\epsilon^{1/2}} \log^{5/4} \frac{1}{\delta}, \end{aligned}$$

as required. \square

Lemma 23. Let $e^\phi \in \tilde{\mathcal{F}}(e^{\phi_0}, \delta)$ for some concave $\phi_0 : \mathbb{R} \rightarrow [-\infty, \infty)$ and $\delta \in (0, \infty)$. Let $r \in \mathbb{R}$ be such that $\int_{-\infty}^r e^{\phi_0(t)} dt \wedge \int_r^\infty e^{\phi_0(t)} dt \geq 4\delta^2$. Then

$$\frac{\phi(r) - \phi_0(r)}{2} \geq -\frac{2\delta}{\{\int_{-\infty}^r e^{\phi_0(t)} dt \wedge \int_r^\infty e^{\phi_0(t)} dt\}^{1/2}}.$$

Proof. As a shorthand, let us write $\psi := (\phi - \phi_0)/2$. Now fix $r \in \mathbb{R}$ and assume without loss of generality that $\psi(r) < 0$ (because otherwise the result is immediate). Since ψ is concave, $\psi(t) \leq \psi(r)$ either for $t \in (-\infty, r)$ or for $t \in (r, \infty)$. In the former case,

$$\delta^2 \geq \int_{-\infty}^\infty e^{\phi_0(t)} (1 - e^{\psi(t)})^2 dt \geq (1 - e^{\psi(r)})^2 \int_{-\infty}^r e^{\phi_0(t)} dt.$$

Hence

$$\psi(r) \geq \log\left(1 - \frac{\delta}{\{\int_{-\infty}^r e^{\phi_0(t)} dt\}^{1/2}}\right) \geq -\frac{2\delta}{\{\int_{-\infty}^r e^{\phi_0(t)} dt\}^{1/2}}.$$

On the other hand, if $\psi(t) < \psi(r)$ for $t \in (r, \infty)$, we can apply an almost identical argument to see that

$$\psi(r) \geq -\frac{2\delta}{\{\int_r^\infty e^{\phi_0(t)} dt\}^{1/2}},$$

as required. \square

Lemma 24. Let $f_0 = e^{\phi_0} \in \mathcal{F}_1$ with $\mu_{f_0} = 0$ and $\sigma_{f_0} = 1$, and let $e^\phi \in \tilde{\mathcal{F}}(f_0, \delta)$ for some $\delta \in (0, 2^{-9})$. Then there exists a universal constant $C > 0$ such that for any $r \in \mathbb{R}$,

$$\frac{\phi(r) - \phi_0(r)}{2} \leq C(|r| + 1)\delta.$$

Proof. Again, we write $\psi := (\phi - \phi_0)/2$. Since we seek an upper bound for ψ , we may assume without loss of generality that ψ is upper semi-continuous, and by symmetry, it suffices to prove the bound at a fixed $r \geq 0$. Further, we assume without loss of generality that $\psi(r) > 0$ (because otherwise the result is immediate).

Let $r_0 := r + 1$. Define $S := \{t \in [-r_0, r_0] : \psi(t) \geq \frac{\psi(r)}{2^{18}c_1r_0}\}$, $s_L := \inf\{t : t \in S\}$, and $s_R := \sup\{t : t \in S\}$. We note that $r \in S$ since $2^{18}c_1r_0 \geq 2$. Then, since $e^x - 1 \geq x$ for any $x \geq 0$, we have

$$\delta^2 \geq \int_{s_L}^{s_R} e^{\phi_0(t)} (e^{\psi(t)} - 1)^2 dt \geq \int_{s_L}^{s_R} e^{\phi_0(t)} \psi(t)^2 dt \geq \left(\frac{\psi(r)}{2^{18}c_1r_0}\right)^2 \int_{s_L}^{s_R} e^{\phi_0(t)} dt. \quad (32)$$

Now, define $S' := \{t \in [-r_0, r_0] : \psi(t) \geq -\frac{\psi(r)}{2^{18}c_1r_0}\}$, $s'_L := \inf\{t : t \in S'\}$, and $s'_R := \sup\{t : t \in S'\}$. Then

$$\delta^2 \geq \int_{[-r_0, r_0] \setminus [s'_L, s'_R]} e^{\phi_0(t)} (1 - e^{\psi(t)})^2 dt \geq \left\{1 - e^{-\frac{\psi(r)}{2^{18}c_1r_0}}\right\}^2 \int_{[-r_0, r_0] \setminus [s'_L, s'_R]} e^{\phi_0(t)} dt.$$

If $\int_{[-r_0, r_0] \setminus [s'_L, s'_R]} e^{\phi_0(t)} dt > 0$, then, by rearranging terms, we have

$$\psi(r) \leq -2^{18} c_1 r_0 \log \left(1 - \frac{\delta}{\left\{ \int_{[-r_0, r_0] \setminus [s'_L, s'_R]} e^{\phi_0(t)} dt \right\}^{1/2}} \right). \quad (33)$$

As a shorthand, let us define

$$\begin{aligned} T_1 &:= \int_{s_L}^{s_R} e^{\phi_0(t)} dt, \quad T_2 := \int_{s'_L}^{s_L} e^{\phi_0(t)} dt + \int_{s_R}^{s'_R} e^{\phi_0(t)} dt, \\ T_3 &:= \int_{-r_0}^{s'_L} e^{\phi_0(t)} dt + \int_{s'_R}^{r_0} e^{\phi_0(t)} dt. \end{aligned}$$

Inequalities (32) and (33) yield

$$\psi(r) \leq 2^{18} c_1 r_0 \frac{\delta}{T_1^{1/2}} \quad (34)$$

$$\psi(r) \leq -2^{18} c_1 r_0 \log \left(1 - \frac{\delta}{T_3^{1/2}} \right) \quad \text{if } T_3 > 0. \quad (35)$$

Since $T_1 + T_2 + T_3 = \int_{-r_0}^{r_0} e^{\phi_0(t)} dt$, we have that $T_1 + T_2 + T_3 \geq 2^{-14}$ from the proof of Lemma 25.

We claim that $T_2 \leq 2^{-15}$. To see this, note that by concavity of ψ (cf. the proof of Theorem 1 of [Cule et al. \(2010\)](#)),

$$s'_R - s'_L \leq (s_R - s_L) \frac{1 + \frac{1}{2^{18} c_1 r_0}}{1 - \frac{1}{2^{18} c_1 r_0}} \leq (s_R - s_L) \left(1 + \frac{4}{2^{18} c_1 r_0} \right).$$

Hence

$$T_2 = \int_{[s'_L, s_L] \cup [s_R, s'_R]} e^{\phi_0(t)} dt \leq c_1 (s'_R - s_R + s_L - s'_L) \leq (s_R - s_L) \frac{4}{2^{18} r_0} \leq 2^{-15}.$$

It follows that either $T_1 \geq 2^{-16}$ or $T_3 \geq 2^{-16}$. If $T_1 \geq 2^{-16}$, then the result follows from (34). On the other hand, if $T_3 \geq 2^{-16}$, then from (35), we have $\psi(r) \leq 2^{19} r_0 \delta$, which again proves the claim. \square

Lemma 25. *Let $f \in \mathcal{F}_1$ with $\mu_f = 0$ and $\sigma_f = 1$. Let $\delta \in (0, 2^{-4}]$. Then $a_L := \inf\{r : f(r) \geq \delta^2\} \leq -1/9$, $a_R := \inf\{r : f(r) \geq \delta^2\} \geq 1/9$, and*

$$\int_{a_L}^{a_R} f(r) dr \geq 2^{-14}. \quad (36)$$

Proof. By [Lovász and Vempala \(2007, Theorem 5.14\(d\)\)](#), $2^{-7} \leq f(0) \leq 2^4$ and $f(r) \geq 2^{-8}$ for all $r \in [-1/9, 1/9]$. Since log-concave functions are unimodal, we conclude that $a_L \leq -1/9$ and $a_R \geq 1/9$. Now define $\alpha := \log f(0)$ and $\beta := \log f(a_L)$. Then

$$\begin{aligned} \int_{a_L}^0 f(r) dr &= |a_L| \int_0^1 f((1-s)a_L) ds \geq |a_L| \int_0^1 e^{s\alpha + (1-s)\beta} ds \\ &= |a_L| e^\beta \frac{1}{\alpha - \beta} (e^{\alpha - \beta} - 1) \geq (1/9) 2^{-8} \frac{1}{\alpha - \beta} (e^{\alpha - \beta} - 1). \end{aligned}$$

Now, by Lovász and Vempala (2007, Theorems 5.14(b) and 5.14(d)) $\alpha - \beta \geq -7 \log 2 - 9 \log 2 = -16 \log 2$, so we deduce that

$$\int_{a_L}^0 f(r) dr \geq (1/9)2^{-8} \frac{1}{16 \log 2} (1 - 2^{-16}) \geq 2^{-15}.$$

By symmetry, the same bound holds for $\int_0^{a_R} f(r) dr$. The lemma follows. \square

Lemma 26. *Let $\phi_0 : \mathbb{R} \rightarrow [-\infty, \infty)$ be concave. Let $a < b$ and fix an arbitrary $t^* \in [a, b]$. Let*

$$q(s) := \begin{cases} (1 - e^{-s})/s & \text{if } s \neq 0 \\ 1 & \text{if } s = 0. \end{cases}$$

Then

$$\int_a^b e^{\phi_0(t)} dt \geq (t^* - a)e^{\phi_0(t^*)} q(\phi_0(t^*) - \phi_0(a)) + (b - t^*)e^{\phi_0(t^*)} q(\phi_0(t^*) - \phi_0(b)).$$

Moreover, if $\max\{\phi_0(a), \phi_0(b)\} \leq \phi_0(t^*) \leq \max\{\phi_0(a), \phi_0(b)\} + \tau$ for some $\tau \geq \log 2$, then

$$\int_a^b e^{\phi_0(t)} dt \geq (b - a)e^{\phi_0(t^*)} \frac{1}{2\tau}.$$

Proof. Let us first suppose that $t^* > a$. We have $\phi_0(t) \geq s\phi_0(a) + (1 - s)\phi_0(t^*)$ for $t \in [a, t^*]$, where $s := \frac{t^* - t}{t^* - a}$. Hence

$$\int_a^{t^*} e^{\phi_0(t)} dt \geq (t^* - a) \int_0^1 e^{s\phi_0(a) + (1 - s)\phi_0(t^*)} ds \geq (t^* - a)e^{\phi_0(t^*)} q(\phi_0(t^*) - \phi_0(a)).$$

We can bound $\int_{t^*}^b e^{\phi_0(t)} dt$ when $t^* < b$ by a similar argument to yield the first conclusion.

For the second part, observe that q is strictly decreasing, so from the first part,

$$\int_a^{t^*} e^{\phi_0(t)} dt \geq (b - a)e^{\phi_0(t^*)} q(\tau) \geq (b - a)e^{\phi_0(t^*)} \frac{1}{2\tau},$$

for $\tau \geq \log 2$. \square

The following two lemmas are from Kim et al. (2017a), though the first is only a minor restatement of Doss and Wellner (2016, Theorem 4.1). For $a < b$ and $-\infty \leq B_1 < B_2 < \infty$, we define $\tilde{\mathcal{F}}([a, b], B_1, B_2)$ to be the set of log-concave functions $f : [a, b] \rightarrow [e^{B_1}, e^{B_2}]$.

Lemma 27. *There exists a universal constant $C > 0$ such that for every $a < b$ and $B, \epsilon > 0$,*

$$H_{[]}(\epsilon, \tilde{\mathcal{F}}([a, b], -\infty, B), d_H, [a, b]) \leq C \frac{e^{B/4}(b - a)^{1/4}}{\epsilon^{1/2}}.$$

Lemma 28. *There exists a universal constant $C > 0$ such that*

$$H_{\mathbb{H}}(\epsilon, \tilde{\mathcal{F}}([a, b], B_1, B_2), d_{\mathbb{H}}, [a, b]) \leq C(B_2 - B_1)^{1/2} \frac{e^{B_2/4}(b - a)^{1/4}}{\epsilon^{1/2}}.$$

for every $\epsilon > 0$, $a < b$ and $-\infty \leq B_1 \leq B_2 < \infty$.

Acknowledgements: The second author would like to thank Sharmodeep Bhattacharya and Richard Lockhart for helpful conversations. The research of the second author is supported by EPSRC fellowship EP/J017213/1 and a grant from the Leverhulme Trust RG81761.

References

- Bhattacharya, S. and Bickel, P. J. (2012). Adaptive estimation in elliptical distributions with extensions to high-dimensions, *Preprint. Available at <http://www.science.oregonstate.edu/~bhattachash/Site/Research.html>*.
- Biau, G. and Devroye, L. (2003). On the risk of estimates for block decreasing densities, *J. Mult. Anal.* **86**: 143–165.
- Billingsley, P. (1995). *Probability and Measure*, Wiley, New York.
- Bobkov, S. and Madiman, M. (2011). Concentration of the information in data with log-concave distributions, *Ann. Probab.* **39**(4): 1528–1543.
- Cule, M. and Samworth, R. (2010). Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density, *Elect. J. of Statist.* **4**: 254–270.
- Cule, M., Samworth, R. and Stewart, M. (2010). Maximum likelihood estimation of a multi-dimensional log-concave density, *J. Roy. Statist. Soc., Ser. B (with discussion)* **72**: 545–607.
- Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding, *Ann. Statist.* **24**: 508–539.
- 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**: 954–981.
- Dudley, R. M. (2002). *Real Analysis and Probability*, Cambridge University Press, Cambridge.

- Dümbgen, L., Hüsler, A. and Rufibach, K. (2007). Active set and EM algorithms for log-concave densities based on complete and censored data, *Preprint arXiv:0707.4643* .
- Dümbgen, L., Samworth, R. and Schuhmacher, D. (2011). Approximation by log-concave distributions, with applications to regression, *Ann. Statist.* **39**: 702–730.
- Fisher, R. A. (1922). On the interpretation of χ^2 from contingency tables, and the calculation of p , *J. Roy. Statist. Soc.* **85**(1): 87–94.
- Fisher, R. A. (1925). Theory of statistical estimation, *Mathematical Proceedings of the Cambridge Philosophical Society* **22**(5): 700–725.
- Fresen, D. (2013). A multivariate Gnedenko law of large numbers, *Ann. Probab.* **41**: 3051–3080.
- Gao, F. and Wellner, J. A. (2007). Entropy estimate for high-dimensional monotonic functions, *J. Mult. Anal.* **98**: 1751–1764.
- Grenander, U. (1956). On the theory of mortality measurement: part ii, *Scand. Actuar. J.* **1956**(2): 125–153.
- Gu, C. and Qiu, C. (1993). Smoothing spline density estimation: Theory, *Ann. Statist.* pp. 217–234.
- Han, Q., Wang, T., Chatterjee, S. and Samworth, R. J. (2017). Isotonic regression in general dimensions, *Preprint arXiv:1708.09468* .
- Han, Q. and Wellner, J. A. (2016). Approximation and estimation of s -concave densities via Rényi divergences, *Ann. Statist.* **44**: 1332–1359.
- Ibragimov, I. A. and Khasminskii, R. (1983). Estimation of distribution density, *Journal of Mathematical Sciences* **21**(1): 40–57.
- Kim, A. K. H., Guntuboyina, A. and Samworth, R. J. (2017a). Adaptation in log-concave density estimation, *Ann. Statist.*, to appear .
- Kim, A. K. H., Guntuboyina, A. and Samworth, R. J. (2017b). Online supplement to ‘Adaptation in log-concave density estimation’, *Ann. Statist.*, to appear .
- Kim, A. K. H. and Samworth, R. J. (2016). Global rates of convergence in log-concave density estimation, *Ann. Statist.* **44**(6): 2756–2779.

- Liebscher, E. and Richter, W.-D. (2016). Estimation of star-shaped distributions, *Risks* **4**: 44.
- Liu, H., Xu, M., Gu, H., Gupta, A., Lafferty, J. and Wasserman, L. (2011). Forest density estimation, *J. Mach. Learn. Res.* **12**: 907–951.
- Lovász, L. and Vempala, S. (2007). The geometry of logconcave functions and sampling algorithms, *Random Struct. & Algor.* **30**(3): 307–358.
- Polonik, W. (1995). Density estimation under qualitative assumptions in higher dimensions, *J. Mult. Anal.* **55**: 61–81.
- Polonik, W. (1998). The silhouette, concentration functions, and ML-density estimation under order restrictions, *Ann. Statist.* **26**: 1857–1877.
- Rockafellar, R. T. (1997). *Convex Analysis*, Princeton University Press, Princeton, NJ.
- Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function, *Ann. Math. Stat.* **27**(3): 832–837.
- Samworth, R. J. and Yuan, M. (2012). Independent component analysis via nonparametric maximum likelihood, *Ann. Statist.* **40**: 2973–3002.
- van de Geer, S. (2000). *Empirical Processes in M-Estimation*, Cambridge University Press, Cambridge.
- Wand, M. and Jones, M. (1995). *Kernel Smoothing*, Chapman and Hall/CRC Monographs on Statistics and Applied Probability, Boca-Raton, Florida.