The Annals of Statistics
2016, Vol. 44, No. 3, 954–981
DOI: 10.1214/15-AOS1394

GLOBAL RATES OF CONVERGENCE OF THE MLES OF LOG-CONCAVE AND s-CONCAVE DENSITIES

BY CHARLES R. DOSS AND JON A. WELLNER

University of Minnesota and University of Washington

We establish global rates of convergence for the Maximum Likelihood Estimators (MLEs) of log-concave and s-concave densities on $\mathbb{R}$. The main finding is that the rate of convergence of the MLE in the Hellinger metric is no worse than $n^{-2/5}$ when $-1 < s < \infty$ where $s = 0$ corresponds to the log-concave case. We also show that the MLE does not exist for the classes of s-concave densities with $s < -1$.

1. Introduction and overview.

1.1. *Preliminary definitions and notation.* We study global rates of convergence of nonparametric estimators of log-concave and s-concave densities, with focus on maximum likelihood estimation and the Hellinger metric. A density p on $\mathbb{R}^d$ is *log-concave* if

$$p = e^{\varphi} \qquad \text{where } \varphi : \mathbb{R}^d \mapsto [-\infty, \infty) \text{ is concave.}$$

We denote the class of all such densities p on $\mathbb{R}^d$ by $\mathcal{P}_{d,0}$. Log-concave densities are always unimodal and have convex level sets. Furthermore, log-concavity is preserved under marginalization and convolution. Thus, the classes of log-concave densities can be viewed as natural nonparametric extensions of the class of Gaussian densities.

The classes of log-concave densities on $\mathbb{R}$ and $\mathbb{R}^d$ are special cases of the classes of s-concave densities studied and developed by [5–7] and [30]. Dharmadhikari and Joag-Dev [11], pages 84–99, gives a useful summary. These classes are defined by the generalized means of order s as follows. Let

$$M_s(a, b; \theta) \equiv \begin{cases} \big((1-\theta)a^s + \theta b^s\big)^{1/s}, & s \neq 0, a, b \geq 0, \\ a^{1-\theta} b^{\theta}, & s = 0, \\ \min(a, b), & s = -\infty. \end{cases}$$

Then $p \in \widetilde{\mathcal{P}}_{d,s}$, the class of s-concave densities on $C \subset \mathbb{R}^d$, if p satisfies

$$p\big((1-\theta)x_0 + \theta x_1\big) \geq M_s\big(p(x_0), p(x_1); \theta\big)$$

Received June 2013; revised September 2015.
MSC2010 subject classifications. Primary 62G07; secondary 62G05, 62G20.
Key words and phrases. Bracketing entropy, consistency, empirical processes, global rate, Hellinger metric, log-concave, s-concave.

for all $x_0, x_1 \in C$ and $\theta \in (0,1)$. It is not hard to see that $\widetilde{\mathcal{P}}_{d,0} = \mathcal{P}_{d,0}$ consists of densities of the form $p = e^{\varphi}$ where $\varphi \in [-\infty, \infty)$ is concave; densities p in $\widetilde{\mathcal{P}}_{d,s}$ with $s < 0$ have the form $p = \varphi^{1/s}$ where $\varphi \in [0, \infty)$ is convex; and densities p with $s > 0$ have the form $p = \varphi_+^{1/s}$ where $x_+ = \max(x, 0)$ and φ is concave on C [and then we write $\widetilde{\mathcal{P}}_{d,s}(C)$]; see, for example, [11] page 86. These classes are nested since

$$\widetilde{\mathcal{P}}_{d,s}(C) \subset \widetilde{\mathcal{P}}_{d,0} \subset \widetilde{\mathcal{P}}_{d,r} \subset \widetilde{\mathcal{P}}_{d,-\infty} \qquad \text{if } -\infty < r < 0 < s < \infty. \tag{1.1}$$

Here, we view the classes $\widetilde{\mathcal{P}}_{1,s}$ defined above for $d = 1$ in terms of the generalized means M_s as being obtained as increasing transforms h_s of the class of concave functions on $\mathbb{R}$ with

$$h_s(y) = \begin{cases} e^y, & s = 0, \\ (-y)_+^{1/s}, & s < 0, \\ y_+^{1/s}, & s > 0. \end{cases}$$

Thus, with λ denoting Lebesgue measure on $\mathbb{R}^d$ we define

$$\mathcal{P}_{d,s} = \{p = h_s(\varphi) : \varphi \text{ is concave on } \mathbb{R}^d\} \cap \left\{p : \int p \, d\lambda = 1\right\},$$

where the concave functions φ are assumed to be closed (i.e., upper semicontinuous), proper and are viewed as concave functions on all of $\mathbb{R}^d$ rather than on a (possibly) specific convex set C. Thus, we consider φ as a function from $\mathbb{R}$ into $[-\infty, \infty)$. See (2.1) in Section 2. This view simplifies our treatment in much the same way as the treatment in [33], but with "increasing" transformations replacing the "decreasing" transformations of Seregin and Wellner, and "concave functions" here replacing the "convex functions" of Seregin and Wellner.

1.2. *Motivations and rationale.* There are many reasons to consider the s-concave classes $\mathcal{P}_s$ with $s \neq 0$, and especially those with $s < 0$. In particular, these classes contain the log-concave densities corresponding to $s = 0$, while retaining the desirable feature of being unimodal (or quasi-concave), and allowing many densities with tails heavier than the exponential tails characteristic of the log-concave class. In particular, the classes $\mathcal{P}_{1,s}$ with $s \leq -1/2$ contain all the t_ν-densities with degrees of freedom $\nu \geq 1$. Thus, choice of an s-concave class $\mathcal{P}_s$ may be viewed as a choice of how far to go in including heavy tailed densities. For example, choosing $s = 1/2$ yields a class which includes all the t_ν-densities with $\nu \geq 1$ (and all the classes $\mathcal{P}_s$ with $s > -1/2$ since the classes are nested), but not the t_ν-densities for any $\nu \in (0,1)$. Once a class $\mathcal{P}_s$ is fixed, it is known that the MLE over $\mathcal{P}_s$ exists (for sufficiently large sample size n) without any choice of tuning parameters, and as will be reviewed in Theorem 2.1, below, is consistent in several senses. The choice of s plays a role somewhat analogous to some index of smoothness, α say, in more classical nonparametric estimation based on smoothness assumptions: smaller values of s yield larger classes of densities, much as

smaller values of a smoothness index α yield larger classes of densities. But for the shape constrained families $\mathcal{P}_s$, no bandwidth or other tuning parameter is needed to define the estimator, whereas such tuning parameters are typically needed for estimation in classes defined by smoothness conditions. For further examples and motivations for the classes $\mathcal{P}_s$, see [6] and [28]. Heavy tailed data are quite common in many application areas including data arising from financial instruments (such as stock returns, commodity returns, and currency exchange rates), and measurements that arise from data networks (such as sizes of files being transmitted, file transmission rates and durations of file transmissions) often empirically exhibit heavy tails. Yet another setting where heavy-tailed data arise is in the purchasing of reinsurance: small insurance companies may themselves buy insurance from a larger company to cover possible extreme losses. Assuming such losses to be heavy-tailed is natural since they are by definition extreme. Two references (of many) providing discussion of these examples and of inference in heavy-tailed settings are [1] and [29].

1.3. *Review of progress on the statistical side.* Nonparametric estimation of log-concave and s-concave densities has developed rapidly in the last decade. Here is a brief review of recent progress.

1.3.1. *Log-concave and $d = 1$.* For log-concave densities on $\mathbb{R}$, [27] established existence of the Maximum Likelihood Estimator (MLE) $\widehat{p}_n$ of p_0, provided a method to compute it, and showed that it is Hellinger consistent: $H(\widehat{p}_n, p_0) \to_{\text{a.s.}} 0$ where $H^2(p, q) = (1/2)\int\{\sqrt{p} - \sqrt{q}\}^2\, dx$ is the (squared) Hellinger distance. Dümbgen and Rufibach [19] also discussed algorithms to compute $\widehat{p}_n$ and rates of convergence with respect to supremum metrics on compact subsets of the support of p_0 under Hölder smoothness assumptions on p_0. Balabdaoui et al. [2] established limit distribution theory for the MLE of a log-concave density at fixed points under various differentiability assumptions and investigated the natural mode estimator associated with the MLE.

1.3.2. *Log-concave and $d \geq 2$.* Estimation of log-concave densities on $\mathbb{R}^d$ with $d \geq 2$ was initiated by [10]; they established existence and uniqueness and algorithms for computation. Cule and Samworth [9] proved consistency in weighted L_1 and appropriate supremum metrics, while [20, 32] investigated stability and robustness properties and use of the log-concave MLE in regression problems. Recently, [25] study upper and lower bounds for minimax risks based on Hellinger loss. When specialized to $d = 1$ and $s = 0$, their results are consistent with (and somewhat stronger than) the results we obtain here. (See Section 5 for further discussion.)

1.3.3. *s-concave and $d \geq 1$.* While the log-concave (or 0-concave) case has received the most attention among the s-concave classes, some progress has been made for other s-concave classes. Seregin and Wellner [33] showed that the MLE exists and is Hellinger consistent for the classes $\mathcal{P}_{d,s}$ with $s \in (-1/d, \infty)$. Koenker and Mizera [26] studied estimation over s-concave classes via estimators based on Rényi and other divergence criteria rather than maximum likelihood. Consistency and stability results for these divergence estimator analogous to those established by [20] and [32] for the MLE in the log-concave case have been investigated by [24].

1.4. *What we do here.* In this paper, we will focus on global rates of convergence of MLEs for the case $d = 1$. We make this choice because of additional technical difficulties when $d > 1$. Although it has been conjectured that the s-concave MLE is Hellinger consistent at rate $n^{-2/5}$ in the one-dimensional cases (see, e.g., [33], pages 3778–3779), to the best of our knowledge this has not yet been proved (even though it follows for $s = 0$ and $d = 1$ from the unpublished results of [14] and [25]).

The main difficulty in establishing global rates of convergence with respect to the Hellinger or other metrics has been to derive suitable bounds for the metric entropy with bracketing for appropriately large subclasses $\mathcal{P}$ of log-concave or s-concave densities. We obtain bounds of the form

$$\log N_{[\cdot]}(\epsilon, \mathcal{P}, H) \leq K\epsilon^{-1/2}, \qquad \epsilon > 0, \tag{1.2}$$

where $N_{[\cdot]}(\epsilon, \mathcal{P}, H)$ denotes the minimal number of ϵ-brackets with respect to the Hellinger metric H needed to cover $\mathcal{P}$. We will establish such bounds in Section 3 using recent results of [16] (see also [22]) for convex functions on $\mathbb{R}$. These recent results build on earlier work by [8] and [17]; see also [18], pages 269–281. The main difficulty has been that the bounds of [8] involve restrictions on the Lipschitz behavior of the convex functions involved as well as bounds on the supremum norm of the functions. The classes of log-concave functions to be considered must include the estimators $\widehat{p}_n$ (at least with arbitrarily high probability for large n). Since the estimators $\widehat{p}_n$ are discontinuous at the boundary of their support (which is contained in the support of the true density p_0), the supremum norm does not give control of the Lipschitz behavior of the estimators in neighborhoods of the boundary of their support. Dryanov [16] showed how to get rid of the constraint on Lipschitz behavior when moving from metric entropy with respect to supremum norms to metric entropies with respect to L_r norms. Furthermore, [22] showed how to extend Dryanov's results from $\mathbb{R}$ to $\mathbb{R}^d$ and the particular domains $[0, 1]^d$. Here, we show how the results of [16] and [22] can be strengthened from metric entropy with respect to L_r to bracketing entropy with respect to L_r, and we carry these results over to the class of concave-transformed densities. Once bounds of the form (1.2) are available, then tools from empirical process theory due to [4, 34, 38] and developed further in [35] and [36], become available.

The major results in this paper are developed for classes of densities, more general than the s-concave classes, which we call *concave-transformed classes*. (They will be rigorously defined later; see Section 4.) These are the classes studied in [33]. The main reason for this generality is that it does not complicate the proofs, and, in fact, actually makes the proofs easier to understand. For instance, when $h(y) = e^y$, $h'(y) = h(y)$, but the proofs are more intuitively understood if one can tell the difference between h' and h. Similarly, this generality allows us to keep track of the tail behavior and the peak behavior of the concave-transformed classes separately (via the parameters α and β, see page 964). The tail behavior turns out to be relevant for global rates of convergence, as we see in this paper.

Here is an outline of the rest of our paper. In Section 2, we define the MLEs for s-concave classes and briefly review known properties of these estimators. We also show that the MLE does not exist for $\mathcal{P}_s$ for any $s < -1$. In Section 3, we state our main rate results for the MLEs over the classes $\mathcal{P}_{1,s}$ with $s > -1$. In Section 4, we state our main general rate results for h-transformed concave classes. Section 5 gives a summary as well as further problems and prospects. The proofs are given in Section 6 and in supplementary material [12].

2. Maximum likelihood estimators: Basic properties. We will restrict attention to the class of concave functions

$$\mathcal{C} := \{\varphi : \mathbb{R} \to [-\infty, \infty) | \varphi \text{ is a closed, proper concave function}\}, \tag{2.1}$$

where [31] defines proper (page 24) and closed (page 52) convex functions. A concave function is proper or closed if its negative is a proper or closed convex function, respectively. Since we are focusing on the case $d = 1$, we write $\mathcal{P}_s$ for $\mathcal{P}_{1,s}$; this can be written as

$$\mathcal{P}_s = \left\{ p : \int p \, d\lambda = 1 \right\} \cap h_s \circ \mathcal{C}. \tag{2.2}$$

We also follow the convention that all concave functions φ are defined on all of $\mathbb{R}$ and take the value $-\infty$ off of their *effective domains*, $\operatorname{dom} \varphi := \{x : \varphi(x) > -\infty\}$. These conventions are motivated in [31] (page 40). For any unimodal function p, we let m_p denote the (smallest) mode of p. For two functions f and g and $r \geq 1$, we let $L_r(f, g) = \|f - g\|_r = (\int |f - g|^r \, d\lambda)^{1/r}$. We will make the following assumption.

ASSUMPTION 2.1. We assume that X_i, $i = 1, \ldots, n$ are i.i.d. random variables having density $p_0 = h_s \circ \varphi_0 \in \mathcal{P}_s$ for $s \in \mathbb{R}$.

Write $\mathbb{P}_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$ for the empirical measure of the X_i's. The maximum likelihood estimator $\widehat{p}_n = h_s(\widehat{\varphi}_n)$ of p_0 maximizes

$$\Psi_n(\varphi) = \mathbb{P}_n \log p = \mathbb{P}_n (\log h_s) \circ \varphi$$

over all functions $\varphi \in \mathcal{C}$ for which $\int h_s(\varphi)\, d\lambda = 1$. When $s > -1$, from [33] (Theorem 2.12, page 3757) we know that $\widehat{\varphi}_n$ exists if $n \geq \gamma/(\gamma - 1)$ with $\gamma \equiv -1/s > 1$ in the case $s < 0$, and if $n \geq 2$ when $s \geq 0$. Seregin and Wellner [33], page 3762, conjectured that $\widehat{\varphi}_n$ is unique when it exists. See also [27, 37] and [19] (Theorem 2.1) for the $s = 0$ case.

The existence of the MLE has been shown only when $s > -1$. One might wonder if this is a deficiency in the proofs or is fundamental. It is well known that the MLE does not exist for the class of unimodal densities, $\mathcal{P}_{-\infty}$; see, for example, [3]. The following proposition shows that in fact the MLE does not exist for $\mathcal{P}_s$ when $s < -1$. The case $s = -1$ is still not resolved.

PROPOSITION 2.1. *A maximum likelihood estimator does not exist for the class $\mathcal{P}_s$ for any $s < -1$.*

Proposition 2.1 gives a negative result about the MLE for an s-concave density when $s < -1$. When $s > -1$, there are many known positive results, some of which are summarized in the next theorem, which gives boundedness and consistency results. In particular, we already know that the MLEs for s-concave densities are Hellinger consistent; our main Theorem 3.2 extends this result to give the rate of convergence, when $s > -1$.

Additionally, from lemmas and corollaries involved in the proof of Hellinger consistency, we know that on compact sets strictly contained in the support of p_0 we have uniform convergence, and we know that the s-concave MLE is uniformly bounded almost surely. We will need these latter two results in the proof of the rate theorem to show we only need to control the bracketing entropy of an appropriate subclass of $\mathcal{P}_s$.

THEOREM 2.1 (Consistency and boundedness of $\widehat{p}_n$ for $\mathcal{P}_s$). *Let Assumption 2.1 hold with $s > -1$ and let $\widehat{p}_n$ be the corresponding MLE. Then*:

(i) $H(\widehat{p}_n, p_0) \to_{a.s.} 0$ *as* $n \to \infty$,

(ii) *If S is a compact set strictly contained in the support of p_0,*

$$\sup_{x \in S} |\widehat{p}_n(x) - p_0(x)| \to_{a.s.} 0 \qquad \text{as } n \to \infty,$$

(iii) $\limsup_{n\to\infty} \sup_x \widehat{p}_n(x) \leq \sup_x p_0(x) \equiv M_0 < \infty$ *almost surely.*

PROOF. The first statement (i) is proved by [27] for $s = 0$, and for $s > -1$ in Theorem 2.17 of [33]. Statement (ii) for $s = 0$ is a corollary of Theorem 4.1 of [19], and for $s > -1$ follows from Theorem 2.18 of [33]. Statement (iii) is Theorem 3.2 of [27] for $s = 0$, and is Lemma 3.17 in [33] for $s > -1$. □

In order to find the Hellinger rate of convergence of the MLEs, we will bound the bracketing entropy of classes of s-concave densities. In general, by using

known consistency results, one does not need to bound the bracketing entropy of the entire function class being considered, but rather of a smaller subclass in which the MLE is known to lie with high probability. This is the approach we will take, by using parts (ii) and (iii) of Theorem 2.1. We therefore consider the following subclasses $\mathcal{P}_{M,s}$ of s-concave densities which (we show in the proof of Theorem 3.2) for some $M < \infty$ will contain both p_0 and $\widehat{p}_n$, after translation and rescaling, with high probability for large n. (Recall the Hellinger distance is invariant under translations and rescalings.) For $0 < M < \infty$, let

$$\mathcal{P}_{M,s} \equiv \Big\{ p \in \mathcal{P}_s : \sup_{x \in \mathbb{R}} p(x) \le M, 1/M \le p(x) \text{ for all } |x| \le 1 \Big\}. \tag{2.3}$$

The next proposition gives an envelope for the class $\mathcal{P}_{M,s}$. This envelope is an important part of the proof of the bracketing entropy of the class $\mathcal{P}_{M,s}$.

PROPOSITION 2.2. *Fix $0 < M < \infty$ and $s > -1$. Then there exists a constant $0 < L < \infty$ depending only on s and M such that for any $p \in \mathcal{P}_{M,s}$*

$$p(x) \le \left\{ \begin{array}{ll} \left(M^s + \dfrac{L}{2M}|x| \right)^{1/s}, & |x| \ge 2M+1 \\ M, & |x| < 2M+1 \end{array} \right\}. \tag{2.4}$$

PROOF. A corresponding statement for the more general h-transformed density classes is given in Proposition 4.2 in the Appendix. However, (2.4) does not immediately follow from the statement of Proposition 4.2 applied to $h \equiv h_s(y) = (-y)_+^{1/s}$, since the requirement $\alpha > -1/s$ disallows the case $\alpha = -1/s$, which is what we need. However, (6.6) from the proof of Proposition 4.2 with $h_s^{-1}(y) = -y^s$ for $y \in (0, \infty)$, yields

$$p(x) \le h_s\left(-M^s - \frac{L}{2M}|x| \right)$$

for $|x| \ge 2M + 1$, which gives us (2.4). $\square$

3. Main results: Log-concave and s-concave classes. Our main goal is to establish rates of convergence for the Hellinger consistency given in (i) of Theorem 2.1 for the s-concave MLE. As mentioned earlier, the key step toward proving rate results of this type is to bound the size, in terms of bracketing entropy, of the function class over which we are estimating. Thus, we have two main results in this section. In the first, we bound the bracketing entropy of certain s-concave classes of functions. This shows that for appropriate values of s, the transformed classes have the same relevant metric structure as (compact) classes of concave functions. Next, using the bracketing bound, our next main result gives the rates of convergence of the s-concave MLEs.

Now let the bracketing entropy of a class of functions $\mathcal{F}$ with respect to a semimetric d on $\mathcal{F}$ be defined in the usual way; see, for example, [18] page 234, [36],

page 83 or [35], page 16. The L_r-size of the brackets depends on the relationship of s and r. In particular, for our results, we need to have light enough tails, which is to say we need $-1/s$ to be large enough. Our main results are as follows.

THEOREM 3.1. *Let $r \geq 1$ and $M > 0$. Assume that either $s \geq 0$ or that $\gamma \equiv -1/s > 2/r$. Then*

$$\log N_{[\cdot]}(\epsilon, \mathcal{P}_{M,s}^{1/2}, L_r) \lesssim \epsilon^{-1/2}, \tag{3.1}$$

where the constants in $\lesssim$ depend only on r, M and s. By taking $r = 2$ and $s > -1$, we have that

$$\log N_{[\cdot]}(\epsilon, \mathcal{P}_{M,s}, H) \lesssim \epsilon^{-1/2}.$$

Theorem 3.1 is the main tool we need to obtain rates of convergence for the MLEs $\widehat{p}_n$. This is given in our second main theorem.

THEOREM 3.2. *Let Assumption 2.1 hold, and let $s > -1$. Suppose that $\widehat{p}_{n,s}$ is the MLE of the s-concave density p_0. Then*

$$H(\widehat{p}_{n,s}, p_0) = O_p(n^{-2/5}).$$

Theorem 3.2 is a fairly straightforward consequence of Theorem 3.1 by applying [35], Theorem 7.4, page 99, or [36], Theorem 3.4.4 in conjunction with Theorem 3.4.1, pages 322–323.

In the case $s = 0$, one can extend our results (an upper bound on the rate of convergence) to an upper bound on the risk $E_{p_0}(H^2(\widehat{p}_{n,0}, p_0))$ over the entire class of log-concave densities p_0; [25] show how this can be done; they use the fact that the log-concave density class is compact in the sense that one can translate and rescale to have, for example, any fixed mean and covariance matrix one would like (since the Hellinger metric is invariant under translation and rescaling), and the class of densities with fixed mean and variance is uniformly bounded above. However, to show the risk bound for $s = 0$, [25] use many convergence results that are available for 0-concave densities but not yet available for s-concave densities with $s < 0$. In particular, their crucial Lemma 11, page 33, relies on results concerning the asymptotic behavior of the MLE beyond the log-concave model $\mathcal{P}_0$ due to Dümbgen et al. [20]. We do not yet know if such a result holds for the MLE in any of the classes $\mathcal{P}_s$ with $s < 0$. Thus, for the moment, we leave our results as rates of convergence rather than risk bounds.

In addition to Theorem 3.2, we have further consequences since the Hellinger metric dominates the total variation or L_1-metric and via [35], Corollary 7.5, page 100.

COROLLARY 3.1. *Let Assumption 2.1 hold and let $s > -1$. Suppose that $\widehat{p}_{n,s}$ is the MLE of the s-concave density p_0. Then*

$$\int_{\mathbb{R}} |\widehat{p}_{n,s}(x) - p_0(x)|\, dx = O_p(n^{-2/5}).$$

COROLLARY 3.2. *Let Assumption 2.1 hold and let $s > -1$. Suppose that $\widehat{p}_{n,s}$ is the MLE of the s-concave density p_0. Then the log-likelihood ratio (divided by n) $\mathbb{P}_n \log(\widehat{p}_{n,s}/p_0)$ satisfies*

$$\mathbb{P}_n \log\left(\frac{\widehat{p}_{n,s}}{p_0}\right) = O_p(n^{-4/5}). \tag{3.2}$$

The result (3.2) is of interest in connection with the study of likelihood ratio statistics for tests (and resulting confidence intervals) for the mode m_0 of p_0 which are being developed by the first author. In fact, the conclusions of Theorem 3.2 and Corollary 3.2 are also true for the *constrained maximum likelihood estimator* $\widehat{p}_n^0$ of p_0 constrained to having (known) mode at 0. We will not treat this here, but details will be provided along with the development of these tests in [13] and [15].

The rates we have given are for the Hellinger distance (as well as any distance smaller than the Hellinger distance) and also for the log-likelihood ratio. The Hellinger metric is very natural for maximum likelihood estimation given i.i.d. observations, and thus many results are stated in terms of Hellinger distance (e.g., [35] focuses much attention on Hellinger distance). Use of the Hellinger metric is not imperative, for example, Theorem 3.4.1 of [36] is stated for a general metric, but getting rates for other metrics (e.g., L_r for $r > 1$) would require additional work since using Theorem 3.4.1 of [36] requires verification of additional conditions which are not immediate.

Estimators based on shape constraints have been shown to have a wide range of adaptivity properties. For instance, [19] study the sup-norm on compacta (which we expect to behave differently than Hellinger distance) and show that the log-concave MLE is rate-adaptive to Hölder smoothness β when $\beta \in [1, 2]$. In the case of univariate convex regression, [23] were able to show that the least-squares estimator achieves a parametric rate (up to log factors) at piecewise linear functions φ_0. They do this by computing entropy bounds for local classes of convex functions within a distance δ of the true function. We have not yet succeeded in extending the bracketing entropy bound of our Theorem 3.1 to analogous local classes, because the proof method used for our theorem does not keep tight enough control of concave-function classes that do *not* drop to 0 except near a prespecified boundary (where one expects the entropies to be smaller). It seems that techniques more similar to those used by [16] or [22] may be applicable.

4. Main results: General h-transformed classes. Here, we state and prove the main results of the paper in their most general form, via arbitrary concave-function transformations, h. Similar to our definition of $\mathcal{P}_s$, we define

$$\mathcal{P}_h := \{h \circ \mathcal{C}\} \cap \left\{p : \int p \, d\lambda = 1\right\}, \tag{4.1}$$

the class of h-concave-transformed densities, and we study the MLE over $\mathcal{P}_h$. These will be described in more detail in Definition 4.1 and Assumption 4.1. In order to study rates of convergence, we need to bound bracketing entropies of relevant function classes. Control of the entropies of classes of concave (or convex) functions with respect to supremum metrics requires control of Lipschitz constants, which we do not have. Thus, we will use L_r metrics with $r \geq 1$. First, we will define the classes of concave and concave-transformed functions which we will be studying.

While we consider $\varphi \in \mathcal{C}$ to be defined on $\mathbb{R}$, we will still sometimes consider a function ψ which is the "restriction of φ to I" for an interval $I \subset \mathbb{R}$. By this, in keeping with the above-mentioned convention, we still mean that ψ is defined on $\mathbb{R}$, where if $x \notin I$ then $\psi(x) = -\infty$, and otherwise $\psi(x) = \varphi(x)$. We will let $\varphi|_I$ denote such restricted functions ψ. When we want to speak about the range of any function f (not necessarily concave) we will use set notation, for example, for $S \subseteq \mathbb{R}$, $f(S) := \{y : f(x) = y \text{ for some } x \in S\}$. We will sometimes want to restrict not the domain of φ but, rather, the range of φ. We will thus let $\varphi|^I$ denote $\varphi|_{D_{\varphi,I}}$ for any interval $I \subset \mathbb{R}$, where $D_{\varphi,I} = \{x : \varphi(x) \in I\}$. Thus, for instance, for all intervals I containing $\varphi(\operatorname{dom}\varphi)$ we have $\varphi|^I \equiv \varphi$.

We will be considering classes of nonnegative concave-transformed functions of the type $h \circ \mathcal{C}$ for some transformation h where $h(-\infty) = 0$ and $h(\infty) = \infty$. We will elaborate on these transformations shortly, in Definition 4.1 and Assumption 4.1. We will slightly abuse notation by allowing the dom operator to apply to such concave-transformed functions, by letting $\operatorname{dom} h \circ \varphi := \{x : h(\varphi(x)) > 0\}$ be the support of $h \circ \varphi$.

The function classes in which we will be interested in the end are the classes $\mathcal{P}_{M,s}$ defined in (2.3), or, more generally $\mathcal{P}_{M,h}$ defined in (4.3), to which the MLEs (of translated and rescaled data) belong, for some $M < \infty$, with high probability as sample size gets large. However, such classes contain functions that are arbitrarily close to or equal to 0 on the support of the true density p_0, and these correspond to concave functions that take unboundedly large (negative) values on the support of p_0. Thus, the corresponding concave classes do not have finite bracketing entropy for the L_r distance. To get around this difficulty, we will consider classes of truncated concave functions and the corresponding concave-transformed classes.

DEFINITION 4.1. *A* concave-function transformation, h, *is a continuously differentiable increasing function from* $[-\infty, \infty]$ *to* $[0, \infty]$ *such that* $h(\infty) = \infty$ *and* $h(-\infty) = 0$. *We define its limit points* $\tilde{y}_0 < \tilde{y}_\infty$ *by* $\tilde{y}_0 = \inf\{y : h(y) > 0\}$ *and* $\tilde{y}_\infty = \sup\{y : h(y) < \infty\}$, *we assume that* $h(\tilde{y}_0) = 0$ *and* $h(\tilde{y}_\infty) = \infty$.

REMARK 4.1. These transformations correspond to "decreasing transformations" in the terminology of [33]. In that paper, the transformations are applied to convex functions whereas here we apply our transformations to concave ones. Since negatives of convex functions are concave, and vice versa, each of our transformations h defines a decreasing transformation $\tilde{h}$ as defined in [33] via $\tilde{h}(y) = h(-y)$.

We will sometimes make the following assumptions.

ASSUMPTION 4.1 (Consistency assumptions on h). Assume that the transformation h satisfies:

T.1 $h'(y) = o(|y|^{-(\alpha+1)})$ as $y \searrow -\infty$ for some $\alpha > 1$;

T.2 If $\tilde{y}_0 > -\infty$, then for all $\tilde{y}_0 < c < \tilde{y}_\infty$, there is an $0 < M_c < \infty$ such that $h'(y) \le M_c$ for all $y \in (\tilde{y}_0, c]$;

T.3 If $\tilde{y}_\infty < \infty$, then for some $0 < c < C$, $c(\tilde{y}_\infty - y)^{-\beta} \le h(y) \le C(\tilde{y}_\infty - y)^{-\beta}$ for some $\beta > 1$ and y in a neighborhood of $\tilde{y}_\infty$;

T.4 If $\tilde{y}_\infty = \infty$, then $h(y)^\gamma h(-Cy) = o(1)$ for some $\gamma, C > 0$, as $y \to \infty$.

EXAMPLE 4.1. The class of log-concave densities, as discussed in Section 3 is obtained by taking $h(y) = e^y \equiv h_0(y)$ for $y \in \mathbb{R}$. Then $\tilde{y}_0 = -\infty$ and $\tilde{y}_\infty = \infty$. Assumption T.4 holds with any $\gamma > C > 0$, and assumption T.1 holds for any $\alpha > 1$.

EXAMPLE 4.2. The classes $\mathcal{P}_s$ of s-concave densities with $s \in (-1, 0)$, as discussed in Section 3, are obtained by taking $h(y) = (-y)_+^{1/s} \equiv h_s(y)$ for $s \in (-1, 0)$ and for $y < 0$. Here, $\tilde{y}_0 = -\infty$ and $\tilde{y}_\infty = 0$. Assumption T.3 holds for $\beta = -1/s$, and assumption T.1 holds for any $\alpha \in (1, -1/s)$.

Note that the same classes of densities $\mathcal{P}_s$ result from the transforms $\tilde{h}_s(y) = (1 + sy)_+^{1/s}$ for $y \in (-\infty, -1/s) = (\tilde{y}_0, \tilde{y}_\infty)$: if $p = h_s(\varphi) \in \mathcal{P}_s$, then also $p = \tilde{h}_s(\tilde{\varphi}_s) \in \mathcal{P}_s$ where $\tilde{\varphi}_s \equiv -(\varphi + 1)/s$ is also concave. With this form of the transformation, we clearly have $\tilde{h}_s(y) \to e^y$ as $s \nearrow 0$, connecting this example with Example 4.1.

EXAMPLE 4.3. The classes of s-concave functions with $0 < s < \infty$, as discussed in Section 3 are obtained by taking $h(y) = (y)_+^{1/s} \equiv h_s(y)$. Here $\tilde{y}_0 = 0$ and $\tilde{y}_\infty = \infty$. Assumption T.1 holds for any $\alpha > 1$, assumption T.2 fails if $s > 1$, and assumption T.4 holds for any (small) $C, \gamma > 0$. These (small) classes $\mathcal{P}_h$ are covered by our Corollary 4.3.

EXAMPLE 4.4. To illustrate the possibilities further, consider $h(y) = \tilde{h}_s(y) = (1 + sy)^{1/s}$ for $y \in [0, -1/s)$ with $-1 < s < 0$, and $h(y) = \tilde{h}_r(y)$ for $y \in (-\infty, 0)$ and $r \in (-1, 0]$. Here, $\tilde{y}_0 = -\infty$ and $\tilde{y}_\infty = -1/s$. Assumption T.3 holds for $\beta =$

$-1/s$, and assumption T.1 holds for any $\alpha \in (1, -1/r)$. Note that $s = 0$ is not allowed in this example, since then if $r < 0$, assumption T.4 fails.

The following lemma shows that concave-transformed classes yield nested families $\mathcal{P}_h$ much as the s-concave classes are nested, as was noticed in Section 1.

LEMMA 4.1. *Let h_1 and h_2 be concave-function transformations. If Ψ is a concave function such that $h_1 = h_2 \circ \Psi$, then $\mathcal{P}_{h_1} \subseteq \mathcal{P}_{h_2}$.*

PROOF. Lemma 2.5, page 6, of [33] gives this result, in the notation of "decreasing (convex) transformations." □

Now, for an interval $I \subset \mathbb{R}$, let

$$\mathcal{C}\big(I, [-B, B]\big) = \big\{\varphi \in \mathcal{C} : -B \le \varphi(x) \le B \text{ if } x \in \operatorname{dom}\varphi = I\big\}.$$

Despite making no restrictions on the Lipschitz behavior of the function class, we can still bound the entropy, as long as our metric is L_r with $1 \le r < \infty$ rather than L_∞.

PROPOSITION 4.1 (Extension of Theorem 3.1 of [22]). *Let $b_1 < b_2$. Then there exists $C < \infty$ such that*

$$\log N_{[\cdot]}\big(\epsilon, \mathcal{C}\big([b_1, b_2], [-B, B]\big), L_r\big) \le C\left(\frac{B(b_2 - b_1)^{1/r}}{\epsilon}\right)^{1/2} \tag{4.2}$$

for all $\epsilon > 0$.

Our first main result has a statement analogous to that of the previous proposition, but it is not about concave or convex classes of functions but rather about concave-transformed classes, defined as follows. Let h be a concave-function transformation. Let $\mathcal{I}[b_1, b_2]$ be all intervals I contained in $[b_1, b_2]$, and let

$$\mathcal{F}\big(\mathcal{I}[b_1, b_2], [0, B]\big) = \big\{f : f = h \circ \varphi, \varphi \in \mathcal{C}, \operatorname{dom}\varphi \subset [b_1, b_2], 0 \le f \le B\big\}.$$

THEOREM 4.1. *Let $r \ge 1$. Assume h is a concave-function transformation. If $\tilde{y}_0 = -\infty$ then assume $h'(y) = o(|y|^{-(\alpha+1)})$ for some $\alpha > 0$ as $y \to -\infty$. Otherwise assume assumption* T.2 *holds. Then for all $\epsilon > 0$*

$$\frac{\log N_{[\cdot]}(\epsilon, \mathcal{F}(\mathcal{I}[b_1, b_2], [0, B]), L_r)}{(B(b_2 - b_1)^{1/r})^{1/2}} \lesssim \epsilon^{-1/2},$$

where $\lesssim$ means $\le$ up to a constant. The constant implied by $\lesssim$ depends only on r and h.

Thus, a bounded class of transformed-functions for any reasonable transformation behaves like a compact class of concave functions.

We extend the definition (2.3) to an arbitrary concave-function transformation h as follows:

$$\mathcal{P}_{M,h} \equiv \Big\{ p \in \mathcal{P}_h : \sup_{x\in\mathbb{R}} p(x) \le M,\ 1/M \le p(x) \text{ for all } |x| \le 1 \Big\}. \tag{4.3}$$

As with the analogous classes of log-concave and s-concave densities, the class $\mathcal{P}_{M,h}$ is important because it has an upper envelope, which is given in the following proposition.

PROPOSITION 4.2. *Let h be a concave-function transformation such that assumption* T.1 *holds with exponent $\alpha_h > 1$. Then for any $p^{1/2} \in \mathcal{P}_{M,h}^{1/2}$ with $0 < M < \infty$,*

$$\begin{aligned} p^{1/2}(x) &\le \begin{cases} D^{1/2}\Big(1 + \dfrac{L}{2M}|x|\Big)^{-\alpha_h/2}, & |x| \ge 2M+1, \\ M^{1/2}, & |x| < 2M+1 \end{cases} \\ &\equiv p_{u,h}^{1/2}(x), \end{aligned} \tag{4.4}$$

where $0 < D, L < \infty$ are constants depending only on h and M.

We would like to bound the bracketing entropy of the classes $\mathcal{P}_{M,h}$. This requires allowing possibly unbounded support. To do this, we will apply the envelope from the previous proposition and then apply Theorem 4.1. Because the size or cardinality of the brackets depends on the height of the function class, the upper bound on the heights given by the envelope allows us to take brackets of correspondingly decreasing size and cardinality out towards infinity. Combining all the brackets from the partition of $\mathbb{R}$ yields the result. Before we state the theorem, we need the following assumption, which is the more general version of Assumption 2.1.

ASSUMPTION 4.2. We assume that X_i, $i = 1, \dots, n$ are i.i.d. random variables having density $p_0 = h \circ \varphi_0 \in \mathcal{P}_h$ where h is a concave-function transformation.

THEOREM 4.2. *Let $r \ge 1$, $M > 0$, and $\epsilon > 0$. Let h be a concave-function transformation such that for $g \equiv h^{1/2}$, Assumption* 4.1, T.1–T.4 *hold, with $\alpha \equiv \alpha_g > 1/r \vee 1/2$. Then*

$$\log N_{[\cdot]}\big(\epsilon, \mathcal{P}_{M,h}^{1/2}, L_r\big) \le K_{r,M,h}\epsilon^{-1/2}, \tag{4.5}$$

where $K_{r,M,h}$ is a constant depending only on r, M, and h.

For the proof of this theorem (given with the other proofs, in Section 6), we will pick a sequence y_γ, for $\gamma = 1, \ldots, k_\epsilon$ to discretize the range of values that a concave function φ may take, where k_ϵ defines the index of truncation which necessarily depends on ϵ in order to control the fineness of the approximation. This allows us to approximate a concave function φ more coarsely as y_γ decreases, corresponding to approximating the corresponding concave-*transformed* function $h \circ \varphi$ at the *same* level of fineness at all y_γ levels.

REMARK 4.2. We require that $h^{1/2}$, rather than h itself, is a concave-function transformation here because to control Hellinger distance for the class $\mathcal{P}_{M,h}$, we need to control L_2 distance for the class $\mathcal{P}_{M,h}^{1/2}$. Note that when h is h_s for any $s \in \mathbb{R}$, $h^{1/2}$ is also a concave-function transformation.

We can now state our main rate result theorem, which is the general form of Theorem 3.2. It is proved by using Theorem 4.2, specifying to the case $r = 2$. There is seemingly a factor of two different in the assumptions for the s-concave rate theorem (requiring $-1/s > 1$) and the assumption in the h-concave rate theorem, requiring $\alpha > 1/2$ (where, intuitively, we might think α corresponds to $-1/s$). The reason for this discrepancy is that α in the h-concave theorem is α_g corresponding to $g \equiv h^{1/2}$, rather than corresponding to h itself; thus α_g corresponds not to $(-1/s)$ but to $(-1/s)/2$.

THEOREM 4.3. *Let Assumption 4.2 hold and let $\widehat{p}_n$ be the h-transformed MLE of p_0. Suppose that Assumption 4.1, T.1–T.4 holds for $g \equiv h^{1/2}$. Assume that $\alpha \equiv \alpha_g > 1/2$. Then*

$$H(\widehat{p}_n, p_0) = O_p\big(n^{-2/5}\big). \tag{4.6}$$

The following corollaries connect the general Theorem 4.3 with Theorem 3.2 via Examples 4.1, 4.2 and 4.3.

COROLLARY 4.1. *Suppose that p_0 in Assumption 4.2 is log-concave; that is, $p_0 = h_0 \circ \varphi_0$ with $h_0(y) = e^y$ as in Example 4.1 and φ_0 concave. Let $\widehat{p}_n$ be the MLE of p_0. Then $H(\widehat{p}_n, p_0) = O_p(n^{-2/5})$.*

COROLLARY 4.2. *Suppose that p_0 in Assumption 4.2 is s-concave with $-1 < s < 0$; that is, $p_0 = h_s \circ \varphi_0$ with $h_s(y) = (-y)^{1/s}$ for $y < 0$ as in Example 4.2 with $-1 < s < 0$ and φ_0 concave. Let $\widehat{p}_n$ be the MLE of p_0. Then $H(\widehat{p}_n, p_0) = O_p(n^{-2/5})$.*

COROLLARY 4.3. *Suppose that p_0 in Assumption 4.2 is h-concave where h is a concave transformation satisfying Assumption 4.1. Suppose that h satisfies $h =$*

$h_2 \circ \Psi$ where Ψ is a concave function and h_2 is a concave-function transformation such that $g \equiv h_2^{1/2}$ also satisfies Assumption 4.1, and such that $\alpha \equiv \alpha_g > 1/2$. Let $\widehat{p}_n$ be the h-concave MLE of p_0. Then

$$H(\widehat{p}_n, p_0) = O_p(n^{-2/5}). \tag{4.7}$$

In particular, the conclusion holds for $h = h_s$ given by $h_s(y) = y_+^{1/s}$ with $s > 0$.

Corollaries 4.1 and 4.2 follow immediately from Theorem 4.3 (see Examples 4.1 and 4.2). However, Corollary 4.3 requires an additional argument (given in the proofs section). Together, these three corollaries yield Theorem 3.2 in the main document.

Theorem 4.3 has further corollaries, for example, via Example 4.4.

5. Summary, further problems, and prospects. In this paper, we have shown that the MLEs of s-concave densities on $\mathbb{R}$ have Hellinger convergence rates of $n^{-2/5}$ for all $s > -1$ and that the MLE does not exist for $s < -1$. Our bracketing entropy bounds explicitly quantify the growth of these classes as $s \searrow -1$ and are of independent interest in the study of convergence rates for other possible estimation methods. In the rest of this section, we briefly discuss some further problems.

5.1. *Behavior of the constants in our bounds.* It can be seen from the proof of Theorem 4.2 that the constants in our entropy bounds diverge to $+\infty$ as $\alpha = \alpha_g \searrow 1/r$. When translated to Theorem 3.1 and $r = 2$, this occurs as $(-1/(2s)) \searrow 1/2$. It would be of interest to establish lower bounds for these entropy numbers with the same property. On the other hand, when $r = 2$ and $s = -1/2$, the constant $K_{r,\alpha}$ in the proof of Theorem 4.2 becomes very reasonable: $K_{2,1} = M^{1/5}(4M+2)^{1/5} + 16(2D^{1/2}M/L)^{2/5}$ where M, D, and L are the constants in the envelope function $p_{u,h}$ of Proposition 4.2. Note that the constant $\tilde{K}_{r,\alpha}$ from Theorem 4.1 arises as a factor in the constant for Theorem 4.2, but from the proof of Theorem 4.1 it can be seen that unless $\alpha \searrow 0$, $\tilde{K}_{r,\alpha}$ stays bounded.

5.2. *Alternatives to maximum likelihood.* As noted by [26], page 2999, there are great algorithmic advantages in adapting the method of estimation to the particular class of shape constraints involved, thereby achieving a convex optimization problem with a tractable computational strategy. In particular, [26] showed how Rényi divergence methods are well-adapted to the s-concave classes in this regard. As has become clear through the work of [24], there are further advantages in terms of robustness and stability properties of the alternative estimation procedures obtained in this way.

5.3. *Rates of convergence for nonparametric estimators, $d \geq 2$.* Here, we have provided global rate results for the MLEs over $\mathcal{P}_{1,s}$ with respect to the Hellinger metric. Global rate results are still lacking for the classes $\mathcal{P}_{d,s}$ on $\mathbb{R}^d$ with $d \geq 2$. Kim and Samworth [25] provides interesting and important minimax lower bounds for squared Hellinger risks for the classes $\mathcal{P}_{d,0}$ with $d \geq 1$, and their lower bounds apply to the classes $\mathcal{P}_{d,s}$ as well in view of the nesting properties in (1.1) and Lemma 4.1. Establishment of comparable upper bounds for $d \geq 2$ remains an active area of research.

5.4. *Rates of convergence for the Rényi divergence estimators.* Although global rates of convergence of the Rényi divergence estimators of [26] have not yet been established even for $d = 1$, we believe that the bracketing entropy bounds obtained here will be useful in establishing such rates. The results of [24] provide some useful starting points in this regard.

5.5. *Global rates of convergence for density estimation in L_1.* Rates of convergence with respect to the L_1 metric for MLEs for the classes $\mathcal{P}_{d,0}$ and $\mathcal{P}_{d,s}$ with $d \geq 2$ and $s < 0$ are not yet available. At present, further tools seem to be needed.

5.6. *Rate efficient estimators when $d \geq 3$.* It has become increasingly clear that nonparametric estimators based on minimum contrast methods (either MLE or minimum Rényi divergence) for the classes $\mathcal{P}_{d,s}$ with $d \geq 3$ will be rate inefficient. This modified form of the conjecture of [33], Section 2.6, page 3762, accounts for the fact pointed out by [25] that the classes $\mathcal{P}_{d,s}$ with $-1/d < s \leq 0$ contain all uniform densities on compact convex subsets of $\mathbb{R}^d$, and these densities have Hellinger entropies of order $\epsilon^{-(d-1)}$. Hence, alternative procedures based on sieves or penalization will be required to achieve optimal rates of convergence. Although these problems have not yet been pursued in the context of log-concave and s-concave densities, there is related work by [21], in a closely related problem involving estimation of the support functions of convex sets.

6. Main results: Proofs. This section contains the proofs of the main results.

PROOF OF PROPOSITION 2.1. Let $s < -1$ and set $r \equiv -1/s < 1$. Consider the family of convex functions $\{\varphi_a\}$ given by

$$\varphi_a(x) = a^{-1/r}(b_r - ax)1_{[0,b_r/a]}(x),$$

where $b_r \equiv (1-r)^{1/(1-r)}$ and $a > 0$. Then φ_a is convex and

$$p_a(x) \equiv \varphi_a(x)^{1/s} = \varphi_a(x)^{-r} = \frac{a}{(b_r - ax)^r} 1_{[0,b_r/a)}(x)$$

is a density. The log-likelihood is given by

$$\ell_n(a) = \log L_n(a) = \log \prod_{i=1}^n p_a(X_i) = \sum_{i=1}^n \{\log a - r \log(b_r - aX_i)\}$$

on the set $X_i < b_r/a$ for all $i \le n$, and hence for $a < b_r/X_{(n)}$ where $X_{(n)} \equiv \max_{1\le i\le n} X_i$. Note that $\ell_n(a) \nearrow \infty$ as $a \nearrow b_r/X_{(n)}$. Hence, the MLE does note exist for $\{p_a : a > 0\}$, and a fortiori the MLE does not exist for $\{p : p \in \mathcal{P}_{1,s}\}$ with $s < -1$. □

PROOF OF PROPOSITION 4.1. The proof consists mostly of noticing that Theorem 3.1 in [22] essentially yields the result stated here; the difference in the statements is that we use L_r bracketing entropy whereas they use L_r metric entropy. For the details of the proof, see supplementary material [12]. □

To prove Theorem 4.1, we discretize the domains and the range of the concave-transformed functions. We define a sequence of values y_γ that discretize the range of the concave functions. As $|y_\gamma|$ get large, $h(y_\gamma)$ get small, so we can define brackets of increasing size. The increasing size of the brackets will be governed by the values of ϵ_γ^B in the proof. We also have to discretize the domain of the functions to allow for regions where the concave-transformed functions can become 0 (which corresponds to concave functions becoming infinite, and which thus cannot be bracketed at the concave level). The sizes of the discretization of the domain corresponding to each level y_γ is governed by the values of ϵ_γ^S in the proof.

PROOF OF THEOREM 4.1. First note that the L_r bracketing numbers scale in the following fashion. For a function f supported on a subset of $[b_1, b_2]$ and with $|f|$ bounded by B, we can define a scaled and translated version of f,

$$\tilde{f}(x) := \frac{f(b_1 + (b_2 - b_1)x)}{B},$$

which is supported on a subset of $[0, 1]$ and bounded by 1. Then

$$B^r \int_{[0,1]} \left|\tilde{f}(x) - \tilde{g}(x)\right|^r dx = \frac{1}{(b_2 - b_1)} \int_{[b_1,b_2]} \left|f(x) - g(x)\right|^r dx.$$

Thus, a class of ϵ-sized L_r brackets when $b_1 = 0$, $b_2 = 1$ and $B = 1$ scales to be a class of $\epsilon(b_2 - b_1)^{1/r} B$ brackets for general b_1, b_2, and B. Thus, for the remainder of the proof we take $b_1 = 0$, $b_2 = 1$ and $B = 1$. By replacing h by a translation of h (since concave functions plus a constant are still concave), and using the fact that the range of h is $(0, \infty)$, we assume that $h^{-1}(1) < 0$.

We will shortly define a sequence of epsilons, ϵ_γ^B and ϵ_γ^S, depending on ϵ. We will need $\epsilon_\gamma^S \le 1$ for all γ. Thus, we will later specify a constant ϵ^* such that $\epsilon \le \epsilon^*$ guarantees $\epsilon_\gamma^S \le 1$.

We will consider the cases $\tilde{y}_0 = -\infty$ and $\tilde{y}_0 > -\infty$ separately; the former case is more difficult, so let us begin by assuming that $\tilde{y}_0 = -\infty$. Let $y_\gamma = -2^\gamma$ for $\gamma = 1, \ldots, k_\epsilon \equiv \lfloor \log_2 h^{-1}(\epsilon) \rfloor$. The y_γ's discretize the range of possible values a concave function takes. We let $\epsilon_\gamma^B = \epsilon(-y_{\gamma-1})^{(\alpha+1)\zeta}$ and $\epsilon_\gamma^S = \epsilon^r(-y_{\gamma-1})^{r\alpha\zeta}$, where we choose ζ to satisfy $1 > \zeta > 1/(\alpha + 1)$.

We start by discretizing the support $[0,1]$. At each level $\gamma = 1, \dots, k_\epsilon$, we use ϵ_γ^S to discretize the support into intervals on which a concave function can cross below y_γ.

We place $\lceil 2/\epsilon_\gamma^S \rceil$ points a_l in $[0,1]$, $l = 1, \dots, \lceil 2/\epsilon_\gamma^S \rceil$, such that $0 < a_{l+1} - a_l < \epsilon_\gamma^S/2$, $l = 0, \dots, \lceil 2/\epsilon_\gamma^S \rceil$ taking $a_{l_0} = 0$ and $a_{l_{\lceil 2/\epsilon_\gamma^S \rceil + 1}} = 1$. There are $N_\gamma^S \equiv \binom{\lceil 2/\epsilon_\gamma^S \rceil}{2}$ pairs of the points, and for each pair (l_1, l_2) we define a pair of intervals, $I_{i,\gamma}^L$ and $I_{i,\gamma}^U$ by

$$I_{i,\gamma}^L = [a_{l_1}, a_{l_2}] \quad \text{and} \quad I_{i,\gamma}^U = [a_{l_1 - 1}, a_{l_2 + 1}],$$

for $i = 1, \dots, N_\gamma^S$. We see that $\log N_\gamma^S \le 4 \log(1/\epsilon_\gamma^S)$, that $\lambda(I_{i,\gamma}^U \setminus I_{i,\gamma}^L) \le \epsilon_\gamma^S$ and that for each γ, for all intervals $I \subset [0,1]$ [i.e., for all possible domains I of a concave function $\varphi \in \mathcal{C}([0,1], [-1,1])$], there exists $1 \le i \le N_\gamma^S$ such that $I_{i,\gamma}^L \subseteq I \subseteq I_{i,\gamma}^U$.

Now, we can apply Proposition 4.1 so for each $\gamma = 1, \dots, k_\epsilon$ we can pick brackets $[l_{\alpha,i,\gamma}(x), u_{\alpha,i,\gamma}(x)]$ for $\mathcal{C}(I_{i,\gamma}^L, [y_\gamma, y_0])$ with $\alpha = 1, \dots, N_\gamma = \lfloor \exp(C(|y_\gamma|/\epsilon_\gamma^B)^{1/2}) \rfloor$ (since $y_0 \le |y_\gamma|$) and $L_r(l_{\alpha,i,\gamma}, u_{\alpha,i,\gamma}) \le \epsilon_\gamma^B$. Note that by Lemma A.2 $k_\epsilon \le \log_2 M \epsilon^{-1/\alpha}$ for some $M \ge 1$, so we see that

$$\epsilon_\gamma^S \le \epsilon^{(1-\zeta)r} \left(\frac{M}{2} \right)^{r\alpha\zeta},$$

and thus taking $\epsilon^* \equiv (2/M)^{\alpha\zeta/(1-\zeta)}$ the above display is bounded above by 1 for all $\epsilon \le \epsilon^*$, as needed.

Now we can define the brackets for $\mathcal{F}(\mathcal{I}[0,1], [0,1])$. For multi-indices $\mathbf{i} = (i_1, \dots, i_{k_\epsilon})$ and $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{k_\epsilon})$, we define brackets $[f_{\mathbf{i},\boldsymbol{\alpha}}^U, f_{\mathbf{i},\boldsymbol{\alpha}}^L]$ by

$$\begin{aligned} f_{\mathbf{i},\boldsymbol{\alpha}}^U(x) = \sum_{\gamma=1}^{k_\epsilon} \big(& h\big(u_{\alpha_\gamma, i_\gamma, \gamma}(x)\big) \mathbb{1}_{\{x \in I_{i_\gamma,\gamma}^L \setminus \bigcup_{j=1}^{\gamma-1} I_{i_j,j}^U\}} \\ & + h(y_{\gamma-1}) \mathbb{1}_{\{x \in I_{i_\gamma,\gamma}^U \setminus (\bigcup_{j=1}^{\gamma} I_{i_j,j}^L \cup \bigcup_{j=1}^{\gamma-1} I_{i_j,j}^U)\}} \big) + \epsilon \mathbb{1}_{\{x \in [0,1] \setminus \bigcup_{j=1}^{\gamma} I_{i_j,j}^U\}}, \end{aligned}$$

$$f_{\mathbf{i},\boldsymbol{\alpha}}^L(x) = \sum_{\gamma=1}^{k_\epsilon} h\big(l_{\alpha_\gamma, i_\gamma, \gamma}(x)\big) \mathbb{1}_{\{x \in I_{i_\gamma,\gamma}^L \setminus \bigcup_{j=1}^{\gamma-1} I_{i_j,j}^U\}}.$$

Figure A in the supplementary material [12] gives a plot of $[f_{\mathbf{i},\boldsymbol{\alpha}}^L, f_{\mathbf{i},\boldsymbol{\alpha}}^U]$. For $x \in I_{i_\gamma,\gamma}^L \setminus \bigcup_{j=1}^{\gamma-1} I_{i_j,j}^U$, we can assume that $y_\gamma \le u_{i_\gamma,\alpha_\gamma,\gamma}(x) \le y_{\gamma-1}$ by replacing $u_{i_\gamma,\alpha_\gamma,\gamma}(x)$ by $(u_{i_\gamma,\alpha_\gamma,\gamma}(x) \wedge y_{\gamma-1}) \vee y_\gamma$. We do the same for $l_{i_\gamma,\alpha_\gamma,\gamma}(x)$.

We will check that these do indeed define a set of bracketing functions for $\mathcal{F}(\mathcal{I}[0,1], [0,1])$ by considering separately the different domains on which $f_{\mathbf{i},\boldsymbol{\alpha}}^U$ and $f_{\mathbf{i},\boldsymbol{\alpha}}^L$ are defined. We take any $h(\varphi) \in \mathcal{F}(\mathcal{I}[0,1], [0,1])$, and then for $\gamma =$

$1, \ldots, k_\epsilon$, we can find $I^L_{i_\gamma,\gamma} \subseteq \mathrm{dom}(\varphi^{[y_\gamma,\infty)}) \subseteq I^U_{i_\gamma,\gamma}$ for some $i_\gamma \le N^S_\gamma$. So, in particular,

$$\varphi(x) < y_\gamma \qquad \text{for } x \notin I^U_{i_\gamma,\gamma} \quad \text{and} \quad y_\gamma \le \varphi(x) \qquad \text{for } x \in I^L_{i_\gamma,\gamma}. \tag{6.1}$$

Thus, there is an α_γ such that $l_{\alpha_\gamma,i_\gamma,\gamma}$ and $u_{\alpha_\gamma,i_\gamma,\gamma}$ have the bracketing property for φ on $I^L_{i_\gamma,\gamma}$, by which we mean that for $x \in I^L_{i_\gamma,\gamma}$, $l_{\alpha_\gamma,i_\gamma,\gamma}(x) \le \varphi(x) \le u_{\alpha_\gamma,i_\gamma,\gamma}(x)$. Thus, on the sets $I^L_{i_\gamma,\gamma} \setminus \bigcup_{j=1}^{\gamma-1} I^U_{i_j,j}$, the functions $f^U_{\mathbf{i},\boldsymbol{\alpha}}$ and $f^L_{\mathbf{i},\boldsymbol{\alpha}}$ have the bracketing property for $h(\varphi)$. Now, $f^L_{\mathbf{i},\boldsymbol{\alpha}}$ is 0 everywhere else and so is everywhere below $h(\varphi)$. $f^U_{\mathbf{i},\boldsymbol{\alpha}}$ is everywhere above $h(\varphi)$ because for $x \in (\bigcup_{j=1}^{\gamma-1} I^U_{i_j,j})^c$, we know $h(\varphi(x)) \le h(y_{\gamma-1})$ by (6.1). It just remains to check that $f^U_{\mathbf{i},\boldsymbol{\alpha}}(x) \ge h(\varphi(x))$ for $x \in [0,1] \setminus \bigcup_{j=1}^{\gamma} I^U_{i_j,j}$, and this follows by the definition of k_ϵ which ensures that $h(y_{k_\epsilon}) \le \epsilon$ and from (6.1). Thus, $[f^L_{\mathbf{i},\boldsymbol{\alpha}}, f^U_{\mathbf{i},\boldsymbol{\alpha}}]$ are indeed brackets for $\mathcal{F}(\mathcal{I}[0,1],[0,1])$.

Next, we compute the size of these brackets. We have that $L^r_r(f^U_{\mathbf{i},\boldsymbol{\alpha}}, f^L_{\mathbf{i},\boldsymbol{\alpha}})$ is

$$\begin{aligned}
\int (f^U_{\mathbf{i},\boldsymbol{\alpha}} - f^L_{\mathbf{i},\boldsymbol{\alpha}})^r \, d\lambda &\le \sum_{\gamma=1}^{k_\epsilon} \int_{I^L_{i_\gamma,\gamma} \setminus I^U_{i_{\gamma-1},\gamma-1}} \left(h(u_{\alpha_\gamma,i_\gamma,\gamma}) - h(l_{\alpha_\gamma,i_\gamma,\gamma})\right)^r d\lambda \\
&\quad + \int_{I^U_{i_\gamma,\gamma} \setminus I^L_{i_\gamma,\gamma}} h(y_{\gamma-1})^r \, d\lambda + \epsilon^r \\
&\le \sum_{\gamma=1}^{k_\epsilon} \sup_{y \in [y_\gamma, y_{\gamma-1}]} h'(y)^r \int_{I^L_{i_\gamma,\gamma} \setminus I^U_{i_{\gamma-1},\gamma-1}} (u_{\alpha_\gamma,i_\gamma,\gamma} - l_{\alpha_\gamma,i_\gamma,\gamma})^r \, d\lambda \\
&\quad + \sum_{\gamma=1}^{k_\epsilon} h(y_{\gamma-1})^r \epsilon^S_\gamma + \epsilon^r,
\end{aligned}$$

since we specified the brackets to take values in $[y_\gamma, y_{\gamma-1}]$ on $I^L_{i_\gamma,\gamma} \setminus I^U_{i_{\gamma-1},\gamma-1}$. By our assumption that $h'(y) = o(|y|^{-(\alpha+1)})$ [so, additionally, $h(y) = o(|y|^{-\alpha})$] as $y \to -\infty$, and the definition of ϵ^B_γ, the above display is bounded above by

$$\epsilon^r + \sum_{\gamma=1}^{k_\epsilon} (-y_{\gamma-1})^{-(\alpha+1)r} \epsilon^r (-y_{\gamma-1})^{(\alpha+1)\zeta r} + \epsilon^r (-y_{\gamma-1})^{-\alpha r(1-\zeta)} \le \tilde{C}_1 \epsilon^r$$

since $\alpha r(1-\zeta)$ and $(\alpha+1)r(1-\zeta)$ are both positive, where $\tilde{C}_1 = (1 + 2/(1 - 2^{-\alpha r(1-\zeta)}))$.

Finally, we can see that the log-cardinality of our set of bracketing functions, $\log \prod_{\gamma=1}^{k_\epsilon} N_\gamma N^S_\gamma$, is

$$\sum_{\gamma=1}^{k_\epsilon} C \left(\frac{|y_\gamma|}{\epsilon^B_\gamma}\right)^{1/2} + 4 \log\left(\frac{1}{\epsilon^S_\gamma}\right), \tag{6.2}$$

with C from Proposition 4.1. The above display is bounded above by

$$C\sum_{\gamma=1}^{k_\epsilon}\frac{2^{\gamma/2}}{\epsilon^{1/2}}2^{-(\gamma-1)(\alpha+1)\zeta/2}+4\log\bigl(\epsilon^{-r}(-y_{\gamma-1})^{-r\alpha\zeta}\bigr)$$

$$\le (C\vee 4)\biggl(\sum_{\gamma=0}^{\infty}\frac{2^{-((\alpha+1)\zeta-1)\gamma/2+1/2}}{\epsilon^{1/2}}+\sum_{\gamma=0}^{\infty}\frac{(-y_\gamma)^{-\alpha\zeta/2}}{\epsilon^{1/2}}\biggr).$$

Since $(\alpha+1)\zeta-1>0$, the above display is finite and can be bounded by $\tilde{C}_2\epsilon^{-1/2}$ where $\tilde{C}_2=(C\vee 4)(\frac{2^{3/2}}{1-2^{-((\alpha+1)/4-1/2)}}\vee\frac{2}{1-2^{-\alpha/4}})$. We have now shown, for $\tilde{y}_0=-\infty$ and $\epsilon\le\epsilon^*$ that

$$\log N_{[\cdot]}\bigl(\epsilon\tilde{C}_1^{1/r},\mathcal{F}\bigl(\mathcal{I}[0,1],[0,1]\bigr),L_r\bigr)\le\tilde{C}_2\epsilon^{-1/2}$$

or for $\epsilon\le\tilde{C}_1^{1/r}\epsilon^*$,

$$\log N_{[\cdot]}\bigl(\epsilon,\mathcal{F}\bigl(\mathcal{I}[0,1],[0,1]\bigr),L_r\bigr)\le\tilde{K}_{r,h}\epsilon^{-1/2},$$

with $\tilde{K}_{r,h}\equiv\tilde{C}_1^{1/(2r)}\tilde{C}_2$. We mention how to extend to all $\epsilon>0$ at the end.

Now let us consider the simpler case, $\tilde{y}_0>-\infty$. Here, we take $k_\epsilon=1$, $y_0=h^{-1}(1)<0$, and $y_1=h^{-1}(0)=\tilde{y}_0$. Then we define $\epsilon^B=\epsilon$, take $\epsilon^*\le 1$, and $\epsilon^S=\epsilon^r\le\epsilon^*$ and we define $I_{i,\gamma}^U$, $I_{i,\gamma}^L$, N_γ^S, $[l_{\alpha,i,\gamma},u_{\alpha,i,\gamma}]$, and N_γ^B as before, except we will subsequently drop the γ subscript since it only takes one value. We can define brackets $[f_{i,\alpha}^L,f_{i,\alpha}^U]$ by

$$f_{i,\alpha}^U(x)=h\bigl(u_{\alpha,i,}(x)\bigr)\mathbb{1}_{A_i^L}(x)+h(y_0)\mathbb{1}_{A_i^U\setminus A_i^L}(x),$$

$$f_{i,\alpha}^L(x)=h\bigl(l_{\alpha,i,}(x)\bigr)\mathbb{1}_{A_i^L}(x).$$

Their size, $L_r^r(f_{i,\alpha}^U,f_{i,\alpha}^L)$ is bounded above by

$$\sup_{y\in[y_1,y_0]}h'(y)^r\int_{A_i^L}(u_{\alpha,i,}-l_{\alpha,i,})^r\,d\lambda+h(y_0)^r\int_{A_i^U\setminus A_i^L}d\lambda\le M^r\epsilon^r+h(y_0)^r\epsilon^r$$

for some $0<M<\infty$ by assumption T.2. Thus, the bracket size is of order ϵ, as desired. The log cardinality $\log N^BN^S$ is

$$C\biggl(\frac{|y_1|}{\epsilon}\biggr)^{1/2}+4\log\bigl(\epsilon^{-r}\bigr).$$

Thus, we get the same conclusion as in the case $\tilde{y}_0=-\infty$, and we have completed the proof for $\epsilon<\epsilon^*$.

When either $\tilde{y}_0=-\infty$ or $\tilde{y}_0>-\infty$, we have proved the theorem when $0<\epsilon\le\epsilon^*$. The result can be extended to apply to any $\epsilon>0$ in a manner identical to the extension at the end of the proof of Proposition 4.1. □

PROOF OF PROPOSITION 4.2. First, we find an envelope for the class $\mathcal{P}_{M,h}$ with $\alpha_h > 1$. For $x \in [-(2M+1), 2M+1]$, the envelope is trivial. Thus, let $x \geq 2M+1$. The argument for $x \leq -(2M+1)$ is symmetric. We show the envelope holds by considering two cases for $p = h \circ \varphi \in \mathcal{P}_{M,h}$. Let $R \equiv \operatorname{dom}\varphi \cap [1, \infty)$. First, consider the case

$$\inf_{x \in R} p(x) \leq 1/(2M). \tag{6.3}$$

We pick $x_1 \in R$ such that $p(x_1) = h(\varphi(x_1)) = 1/(2M)$ and such that

$$\varphi(0) - \varphi(x_1) \geq h^{-1}\big(M^{-1}\big) - h^{-1}\big(M^{-1}/2\big) \equiv L > 0. \tag{6.4}$$

This is possible since $\varphi(0) \geq h^{-1}(M^{-1})$ by the definition of $\mathcal{P}_{1,M,h}$ and by our choice of x_1 [and by the fact that $\operatorname{dom}\varphi$ is closed, so that we attain equality in (6.3)].

If $p(z) \geq 1/(2M)$, then concavity of φ means $p \geq 1/(2M)$ on $[0, z]$ and since p integrates to 1, we have $z \leq 2M$. Thus, $x_1 \leq 2M$. Fix $x > 2M+1 \geq x_1 > 0$, which (by concavity of φ) means $\varphi(0) > \varphi(x_1) > \varphi(x)$. We will use Proposition A.1 with $x_0 = 0$ and x_1 and x as just defined. Also, assume $\varphi(x) > -\infty$, since otherwise any $0 < D, L < \infty$ suffice for our bound. Then we can apply (A.16) to see

$$p(x) \leq h\bigg(\varphi(0) - h\big(\varphi(x_1)\big)\frac{\varphi(0) - \varphi(x_1)}{F(x) - F(0)}x\bigg). \tag{6.5}$$

Since $(F(x) - F(0))^{-1} \geq 1$ (since $\alpha > 1$), (6.5) is bounded above by

$$h\bigg(h^{-1}(M) - \frac{L}{2M}x\bigg) < \infty. \tag{6.6}$$

We can assume $h^{-1}(M) = -1$ without loss of generality. This is because, given an arbitrary h, we let $h_M(y) = h(y + 1 + h^{-1}(M))$ which satisfies $h_M^{-1}(M) = -1$. Note that $\mathcal{P}_{M,h} = \mathcal{P}_{M,h_M}$ since translating h does not change the class $\mathcal{P}_h$ or $\mathcal{P}_{M,h}$. Thus, if (4.4) holds for all $p \in \mathcal{P}_{M,h_M}$ then it holds for all $p \in \mathcal{P}_{M,h}$. So without loss of generality, we assume $h^{-1}(M) = -1$. Then (6.6) is equal to

$$h\bigg(-1 - \frac{L}{2M}x\bigg) < \infty. \tag{6.7}$$

Now, $h(y) = o(|y|^{-\alpha})$ as $y \to -\infty$, which implies that $h(y) \leq D(-y)^{-\alpha}$ on $(-\infty, -1]$ for a constant D that depends only on h and on M, since $-1 - (L/(2M))x \leq -1$. Thus, (6.6) is bounded above by

$$D\bigg(1 + \frac{L}{2M}x\bigg)^{-\alpha}. \tag{6.8}$$

We have thus found an envelope for the case wherein (6.3) holds and when $x \geq 2M+1$. The case $x \leq -(2M+1)$ is symmetric.

Now consider the case where p satisfies

$$\inf_{x\in R} p(x) \geq 1/(2M). \tag{6.9}$$

As argued earlier, if $p(z) \geq 1/(2M)$, then concavity of φ means $p \geq 1/(2M)$ on $[0, z]$ and since p integrates to 1, we have $z \leq 2M$. So, when (6.9) holds, it follows that $p(z) = 0$ for $z > 2M$. We have thus shown $p \leq p_{u,h}$ [with $p_{u,h}$ defined in (4.4)]. For $q \equiv p^{1/2} \in \mathcal{P}_{M,h}^{1/2}$, it is now immediate that $q \leq p_{u,h}^{1/2}$. $\square$

To prove Theorem 4.2, we partition $\mathbb{R}$ into intervals, and on each interval we apply Theorem 4.1. The envelope from Proposition 4.2 gives a uniform bound on the heights of the functions in $\mathcal{P}_{M,h}^{1/2}$, which allows us to control the cardinality of the brackets given by Theorem 4.1.

PROOF OF THEOREM 4.2. We will use the method of Corollary 2.7.4 of [36] for combining brackets on a partition of $\mathbb{R}$, together with Theorem 4.1. Let $I_0 = [-(2M + 1), 2M + 1]$; for $i > 0$ let $I_i = [i^\gamma, (i + 1)^\gamma] \setminus I_0$, and for $i < 0$ let $I_i = [-|i - 1|^\gamma, -|i|^\gamma] \setminus I_0$. Let $A_0 = M^{1/2}(4M + 2)^{1/r}$ and $A_i = D^{1/2}(1 + |i|^\gamma L/(2M))^{-\alpha}((i + 1)^\gamma - i^\gamma)^{1/r}$ where $\alpha \equiv \alpha_{h^{1/2}}$ (so by Lemma A.3 $\alpha_h = 2\alpha_{h^{1/2}} > 1$) for $|i| > 0$, and with D, L as defined in Proposition 4.2, which will correspond to $B(b_2 - b_1)^{1/r}$ in Theorem 4.1 for $\mathcal{P}_{M,h}^{1/2}$ restricted to I_i. For $i \in \mathbb{Z}$, let $a_i = A_i^\beta$ where we will pick $\beta \in (0, 1)$ later. Fix $\epsilon > 0$. We will apply Theorem 4.1 to yield L_r brackets of size ϵa_i for the restriction of $\mathcal{P}_{M,h}^{1/2}$ to each interval I_i. For $i \in \mathbb{Z}$, we apply Theorem 4.1 and form ϵa_i brackets, which we denote by $[f_{i,j}^L, f_{i,j}^U]$ for $j = 1, \ldots, N_i$, for the restriction of $\mathcal{P}_{M,h}^{1/2}$ to I_i. We will bound N_i later. We have thus formed a collection of brackets for $\mathcal{P}_{M,h}^{1/2}$ by

$$\left\{\left[\sum_{i\in\mathbb{Z}} f_{i,j_i}^L \mathbb{1}_{I_i}, \sum_{i\in\mathbb{Z}} f_{i,j_i}^U \mathbb{1}_{I_i}\right] : j_i \in \{1, \ldots, N_i\}, i \in \mathbb{Z}\right\}.$$

The cardinality of this bracketing set is $\prod_{i\in\mathbb{Z}} N_i$. The L_r^r size of a bracket $[f^L, f^U]$ in the above defined collection is

$$\int_{\mathbb{R}} |f^U - f^L|^r \, d\lambda \leq \sum_{i\in\mathbb{Z}} \epsilon^r a_i^r.$$

By Theorem 4.1, $\log N_i \leq \tilde{K}_{r,h}(A_i/(\epsilon a_i))^{1/2}$ for $i \in \mathbb{Z}$ where $\tilde{K}_{r,h}$ is the constant from that theorem. Thus,

$$\log N_{[\cdot]}\left(\epsilon\left(\sum_{i\in\mathbb{Z}} a_i^r\right)^{1/r}, \mathcal{P}_{M,h}^{1/2}, L_r\right) \leq \tilde{K}_{r,h} \sum_{i\in\mathbb{Z}} \left(\frac{A_i}{\epsilon a_i}\right)^{1/2}.$$

We now set $\beta = 1/(2r+1)$, so that $a_i^r = (A_i/a_i)^{1/2} = A_i^{r/(2r+1)}$ and need only to compute $\sum_{i\in\mathbb{Z}} a_i^r = \sum_{i\in\mathbb{Z}} (A_i/a_i)^{1/2}$. Let $\tilde{A}_i = A_i/D^{1/2}$, and we then see that

$$\begin{aligned}
\sum_{|i|\geq 1} \tilde{A}_i^{r/(2r+1)} &= 2\sum_{i\geq 1}\left(1+\frac{L}{2M}i^\gamma\right)^{-\alpha r/(2r+1)}\left((i+1)^\gamma - i^\gamma\right)^{1/(2r+1)} \\
&\leq 2\sum_{i\geq 1}\left(1+\frac{L}{2M}i^\gamma\right)^{-\alpha r/(2r+1)} i^{\gamma/(2r+1)}\left(\left(\frac{i+1}{i}\right)^\gamma - 1\right)^{1/(2r+1)} \\
&= 2^{1+\gamma/(2r+1)}\sum_{i\geq 1}\left(1+\frac{L}{2M}i^\gamma\right)^{-\alpha r/(2r+1)} i^{\gamma/(2r+1)} \\
&\leq 2^{1+\gamma/(2r+1)}\sum_{i\geq 1}\left(\frac{L}{2M}i^\gamma\right)^{-\alpha r/(2r+1)} i^{\gamma/(2r+1)},
\end{aligned}$$

which equals

$$\begin{aligned}
&2^{1+\gamma/(2r+1)}\left(\frac{L}{2M}\right)^{-\alpha r/(2r+1)}\sum_{i\geq 1} i^{-\gamma\alpha r/(2r+1)+\gamma/(2r+1)} \\
&\quad\leq 2^{1+\gamma/(2r+1)}\left(\frac{L}{2M}\right)^{-\alpha r/(2r+1)}\left(1+\int_1^\infty x^{-\alpha\gamma r/(2r+1)+\gamma/(2r+1)}\right)dx,
\end{aligned}$$

which equals

$$2^{1+\gamma/(2r+1)}\left(\frac{L}{2M}\right)^{-\alpha r/(2r+1)}\left(1+\frac{1}{\alpha\gamma r/(2r+1)-\gamma/(2r+1)-1}\right) \tag{6.10}$$

as long as

$$\frac{\alpha\gamma r}{2r+1} - \frac{\gamma}{2r+1} > 1,$$

which is equivalent to requiring

$$\alpha > \frac{1}{r} + \frac{2r+1}{r}\frac{1}{\gamma}. \tag{6.11}$$

Since $\gamma \geq 1$ is arbitrary, for any $\alpha > 1/r$, we can pick $\gamma = ((2r+1)/r)2/(\alpha - 1/r)$. Then the right-hand side of (6.11) becomes $(1/r)(1-1/(2r)) + \alpha/(2r)$, and thus (6.11) becomes

$$\alpha > \frac{\alpha + 1/r}{2},$$

which is satisfied for any $r \geq 1$ and $\alpha > 1/r$. Then (6.10) equals

$$2^{2+(2/(\alpha-1/r))(1/r)}\left(\frac{L}{2M}\right)^{-\alpha r/(2r+1)}.$$

Thus, defining $K_{r,\alpha} \equiv \sum_{i\in\mathbb{Z}} A_i^{r/(2r+1)}$, we have

$$K_{r,\alpha} = M^{r/(2(2r+1))}(4M+2)^{1/(2r+1)} + D^{r/(2(2r+1))}2^{2+(2/(\alpha-1/r))(1/r)}\left(\frac{L}{2M}\right)^{-\alpha r/(2r+1)}.$$

Then we have shown that

$$\log N_{[\cdot]}\big(\epsilon K_{r,\alpha}^{1/r}, \mathcal{P}_{M,h}^{1/2}, L_r\big) \le \tilde{K}_{r,h} K_{r,\alpha}\epsilon^{-1/2},$$

or

$$\log N_{[\cdot]}\big(\epsilon, \mathcal{P}_{M,h}^{1/2}, L_r\big) \le \tilde{K}_{r,h} K_{r,\alpha}^{1+1/(2r)}\epsilon^{-1/2},$$

and the proof is complete. □

PROOF OF THEOREM 4.3. *Step* 1: *Reduction from* $\mathcal{P}_h$ to $\mathcal{P}_{M,h}$. We first show that we may assume, without loss of generality, for some $M > 0$ that $p_0 \in \mathcal{P}_{M,h}$ and, furthermore, $\widehat{p}_n \in \mathcal{P}_{M,h}$ with probability approaching 1 as $n \to \infty$. To see this, consider translating and rescaling the data: we let $\tilde{X}_i = (X_i - b)/a$ for $b \in \mathbb{R}$ and $a > 0$, so that the $\tilde{X}_i$ are i.i.d. with density $\tilde{p}_0(x) = a p_0(ax + b)$. Now the MLE of the rescaled data, $\widehat{p}_n(\tilde{x}; \underline{\tilde{X}})$ satisfies $\widehat{p}_n(\tilde{x}; \underline{\tilde{X}}) = a\widehat{p}_n(a\tilde{x} + b); \underline{X})$ and, since the Hellinger metric is invariant under affine transformations, it follows that

$$H\big(\widehat{p}_n(\cdot; \underline{X}), p_0\big) = H\big(\widehat{p}_n(\cdot; \underline{\tilde{X}}), \tilde{p}_0\big).$$

Hence, if (4.6) holds for $\tilde{p}_0$ and the transformed data, it also holds for p_0 and the original data. Thus, we can pick b and a as we wish. First, we note that there is some interval $B(x_0, \delta) \equiv \{z : |z - x_0| \le \delta\}$ contained in the interior of the support of $p_0 \in \mathcal{P}_h$ since p_0 has integral 1. We take b and a to be x_0 and δ, and thus assume without loss of generality that $B(0, 1)$ is in the interior of the support of p_0. Now, by Theorem 2.17 of [33] which holds under their assumptions (D.1)–(D.4) it follows that we have uniform convergence of $\widehat{p}_n$ to p_0 on compact subsets strictly contained in the support of p_0, such as $B(0, 1)$. Additionally, by Lemma 3.17 of [33], we know that $\limsup_{n\to\infty} \sup_x \widehat{p}_n(x) \le \sup_x p_0(x) \equiv M_0$ almost surely. Assumptions (D.1)–(D.4) of [33] for h are implied by our T.1–T.4 for $g \equiv h^{1/2}$ [with $\beta_h = 2\beta_g$ and $\alpha_h = 2\alpha_g$, since $h'(y) = 2\sqrt{h(y)}(h^{1/2})'(y)$ and if $g'(y) = o(|y|)^{-(\alpha+1)}$ then $g(y) = o(|y|)^{-\alpha}$ as $y \to -\infty$]. Thus, we let $M = (1 + M_0) \vee 2/(\min_{|x|\le 1} p_0(x)) < \infty$. Then we can henceforth assume that $p_0 \in \mathcal{P}_{M,h}$ and, furthermore, with probability approaching 1 as $n \to \infty$, that $\widehat{p}_n \in \mathcal{P}_{M,h}$. This completes step 1.

Step 2. *Control of Hellinger bracketing entropy for* $\mathcal{P}_{M,h}$ *suffices*.

Step 2a: For $\delta > 0$, let

$$\overline{\mathcal{P}}_h(\delta) \equiv \big\{(p + p_0)/2 : p \in \mathcal{P}_h, H\big((p + p_0)/2, p_0\big) < \delta\big\}.$$

Suppose that we can show that

$$\log N_{[\cdot]}\bigl(\epsilon, \overline{\mathcal{P}}_h(\delta), H\bigr) \lesssim \epsilon^{-1/2} \tag{6.12}$$

for all $0 < \delta \le \delta_0$ for some $\delta_0 > 0$. Then it follows from [36], Theorems 3.4.1 and 3.4.4 (with $p_n = p_0$ in Theorem 3.4.4) or, alternatively, from [35], Theorem 7.4 and an inspection of her proofs, that any r_n satisfying

$$r_n^2 \Psi(1/r_n) \le \sqrt{n}, \tag{6.13}$$

where

$$\Psi(\delta) \equiv J_{[\cdot]}\bigl(\delta, \overline{\mathcal{P}}_h(\delta), H\bigr)\biggl(1 + \frac{J_{[\cdot]}(\delta, \overline{\mathcal{P}}_h(\delta), H)}{\delta^2\sqrt{n}}\biggr)$$

and

$$J_{[\cdot]}\bigl(\delta, \overline{\mathcal{P}}_h(\delta), H\bigr) \equiv \int_0^\delta \sqrt{\log N_{[\cdot]}\bigl(\epsilon, \overline{\mathcal{P}}_h(\delta), H\bigr)}\, d\epsilon$$

gives a rate of convergence for $H(\widehat{p}_n, p_0)$. It is easily seen that if (6.12) holds then $r_n = n^{-2/5}$ satisfies (6.13). Thus, (4.6) follows from (6.12).

Step 2b. Thus, we want to show that (6.12) holds if we have an appropriate bracketing entropy bound for $\mathcal{P}_{M,h}^{1/2}$. First, note that

$$N_{[\cdot]}\bigl(\epsilon, \overline{\mathcal{P}}_h(\delta), H\bigr) \le N_{[\cdot]}\bigl(\epsilon, \mathcal{P}_h(4\delta), H\bigr)$$

in view of [36], exercise 3.4.4 (or [35], Lemma 4.2, page 48). Furthermore,

$$N_{[\cdot]}\bigl(\epsilon, \mathcal{P}_h(4\delta), H\bigr) \le N_{[\cdot]}(\epsilon, \mathcal{P}_{M,h}, H)$$

since $\mathcal{P}_h(4\delta) \subset \mathcal{P}_{M,h}$ for all $0 < \delta \le \delta_0$ with $\delta_0 > 0$ sufficiently small. This holds since Hellinger convergence implies pointwise convergence for concave transformed functions which in turn implies uniform convergence on compact subsets of the domain of p_0 via [31], Theorem 10.8. See Lemma A.1 for details of the proofs.

Finally, note that

$$\begin{aligned} N_{[\cdot]}(\epsilon, \mathcal{P}_{M,h}, H) &= N_{[\cdot]}\bigl(\epsilon, \mathcal{P}_{M,h}^{1/2}, L_2(\lambda/2)\bigr) \\ &= N_{[\cdot]}\bigl(\epsilon, \mathcal{P}_{M,h}^{1/2}, L_2(\lambda)/\sqrt{2}\bigr) = N_{[\cdot]}\bigl(\epsilon/\sqrt{2}, \mathcal{P}_{M,h}^{1/2}, L_2(\lambda)\bigr) \end{aligned}$$

by the definition of H and $L_2(\lambda)$. Thus, it suffices to show that

$$\log N_{[\cdot]}\bigl(\epsilon, \mathcal{P}_{M,h}^{1/2}, L_2(\lambda)\bigr) \lesssim \frac{1}{\epsilon^{1/2}}, \tag{6.14}$$

where the constant involved depends only on M and h. This completes the proof of step 2, and completes the proof, since (6.14) is exactly what we can conclude by Theorem 4.2 since we assumed Assumption 4.1 holds and that $\alpha \equiv \alpha_g$ satisfies $\alpha_g > 1/2$. □

PROOF OF COROLLARY 4.3. The proof is based on the proof of Theorem 4.3. In step 1 of that proof, the only requirement on h is that we can conclude that $\widehat{p}_n$ is almost surely Hellinger consistent. Almost sure Hellinger consistency is given by Theorem 2.18 of [33] which holds under their assumptions (D.1)–(D.4), which are in turn implied by our T.1, T.3 and T.4 [recalling that all of our h's are continuously differentiable on $(\tilde{y}_0, \tilde{y}_\infty)$].

Then step 2a of the proof shows that it suffices to show the bracketing bound (6.12) for $\overline{\mathcal{P}}_h(\delta)$. Now, by Lemma 4.1 below we have

$$\log N_{[\cdot]}\bigl(\epsilon, \overline{\mathcal{P}}_h(\delta), H\bigr) \le \log N_{[\cdot]}\bigl(\epsilon, \overline{\mathcal{P}}_{h_2}(\delta), H\bigr).$$

Step 2b of the proof shows that (6.12) holds for transforms h when $g \equiv h^{1/2}$ satisfies $\alpha \equiv \alpha_g > 1/2$, as we have assumed. Thus, we are done. □

Acknowledgments. We owe thanks to Arseni Seregin, Bodhi Sen and Fadoua Balabdaoui as well as two referees for helpful comments, suggestions, and corrections. Thanks also to Tilmann Gneiting for hospitality during our visits to Heidelberg.

SUPPLEMENTARY MATERIAL

Technical proofs (DOI: 10.1214/15-AOS1394SUPP; .pdf). In the supplement, we provide additional proofs and technical details that were omitted from the main paper.

REFERENCES

[1] ADLER, R. J., FELDMAN, R. E. and TAQQU, M. S., eds. (1998). *A Practical Guide to Heavy Tails*. Birkhäuser, Boston, MA. MR1652283

[2] BALABDAOUI, F., RUFIBACH, K. and WELLNER, J. A. (2009). Limit distribution theory for maximum likelihood estimation of a log-concave density. *Ann. Statist.* **37** 1299–1331. MR2509075

[3] BIRGÉ, L. (1997). Estimation of unimodal densities without smoothness assumptions. *Ann. Statist.* **25** 970–981. MR1447736

[4] BIRGÉ, L. and MASSART, P. (1993). Rates of convergence for minimum contrast estimators. *Probab. Theory Related Fields* **97** 113–150. MR1240719

[5] BORELL, C. (1974). Convex measures on locally convex spaces. *Ark. Mat.* **12** 239–252. MR0388475

[6] BORELL, C. (1975). Convex set functions in d-space. *Period. Math. Hungar.* **6** 111–136. MR0404559

[7] BRASCAMP, H. J. and LIEB, E. H. (1976). On extensions of the Brunn–Minkowski and Prékopa–Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *J. Funct. Anal.* **22** 366–389. MR0450480

[8] BRONŠTEĬN, E. M. (1976). ϵ-entropy of convex sets and functions. *Sibirsk. Mat. Zh.* **17** 508–514, 715. MR0415155

[9] CULE, M. and SAMWORTH, R. (2010). Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density. *Electron. J. Stat.* **4** 254–270. MR2645484

[10] CULE, M., SAMWORTH, R. and STEWART, M. (2010). Maximum likelihood estimation of a multi-dimensional log-concave density. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 545–607. MR2758237

[11] DHARMADHIKARI, S. and JOAG-DEV, K. (1988). *Unimodality, Convexity, and Applications*. Academic Press, Boston, MA. MR0954608

[12] DOSS, C. R. and WELLNER, J. A. (2015). Supplement to "Global rates of convergence of the MLEs of log-concave and s-concave densities." DOI:10.1214/15-AOS1394SUPP.

[13] DOSS, C. R. (2013). Shape-constrained inference for concave-transformed densities and their modes. Ph.D. thesis, Dept. Statistics, Univ. Washington, Seattle, WA.

[14] DOSS, C. R. and WELLNER, J. A. (2013). Global rates of convergence of the MLEs of log-concave and s-concave densities. Available at arXiv:1306.1438.

[15] DOSS, C. R. and WELLNER, J. A. (2015). Inference for the mode of a log-concave density. Technical report, Univ. Washington, Seattle, WA.

[16] DRYANOV, D. (2009). Kolmogorov entropy for classes of convex functions. *Constr. Approx.* **30** 137–153. MR2519658

[17] DUDLEY, R. M. (1984). A course on empirical processes. In *École D'été de Probabilités de Saint-Flour, XII*—1982. *Lecture Notes in Math.* **1097** 1–142. Springer, Berlin. MR0876079

[18] DUDLEY, R. M. (1999). *Uniform Central Limit Theorems*. *Cambridge Studies in Advanced Mathematics* **63**. Cambridge Univ. Press, Cambridge. MR1720712

[19] DÜMBGEN, L. and RUFIBACH, K. (2009). Maximum likelihood estimation of a log-concave density and its distribution function: Basic properties and uniform consistency. *Bernoulli* **15** 40–68. MR2546798

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

[21] GUNTUBOYINA, A. (2012). Optimal rates of convergence for convex set estimation from support functions. *Ann. Statist.* **40** 385–411. MR3014311

[22] GUNTUBOYINA, A. and SEN, B. (2013). Covering numbers for convex functions. *IEEE Trans. Inform. Theory* **59** 1957–1965. MR3043776

[23] GUNTUBOYINA, A. and SEN, B. (2015). Global risk bounds and adaptation in univariate convex regression. *Probab. Theory Related Fields* **163** 379–411. MR3405621

[24] HAN, Q. and WELLNER, J. A. (2016). Approximation and estimation of s-concave densities via Rényi divergences. *Ann. Statist.* **44** 1332–1359.

[25] KIM, A. K. H. and SAMWORTH, R. J. (2015). Global rates of convergence in log-concave density estimation. Available at arXiv:1404.2298v2.

[26] KOENKER, R. and MIZERA, I. (2010). Quasi-concave density estimation. *Ann. Statist.* **38** 2998–3027. MR2722462

[27] PAL, J. K., WOODROOFE, M. and MEYER, M. (2007). Estimating a Polya frequency function$_2$. In *Complex Datasets and Inverse Problems. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **54** 239–249. IMS, Beachwood, OH. MR2459192

[28] PRÉKOPA, A. (1995). *Stochastic Programming*. *Mathematics and Its Applications* **324**. Kluwer Academic, Dordrecht. MR1375234

[29] RESNICK, S. I. (2007). *Heavy-Tail Phenomena*: *Probabilistic and Statistical Modeling*. Springer, New York. MR2271424

[30] RINOTT, Y. (1976). On convexity of measures. *Ann. Probab.* **4** 1020–1026. MR0428540

[31] ROCKAFELLAR, R. T. (1970). *Convex Analysis*. *Princeton Mathematical Series* **28**. Princeton Univ. Press, Princeton, NJ. MR0274683

[32] SCHUHMACHER, D., HÜSLER, A. and DÜMBGEN, L. (2011). Multivariate log-concave distributions as a nearly parametric model. *Stat. Risk Model.* **28** 277–295. MR2838319

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

[34] VAN DE GEER, S. (1993). Hellinger-consistency of certain nonparametric maximum likelihood estimators. *Ann. Statist.* **21** 14–44. MR1212164

[35] VAN DE GEER, S. A. (2000). *Applications of Empirical Process Theory. Cambridge Series in Statistical and Probabilistic Mathematics* **6**. Cambridge Univ. Press, Cambridge. MR1739079

[36] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York. MR1385671

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

[38] WONG, W. H. and SHEN, X. (1995). Probability inequalities for likelihood ratios and convergence rates of sieve MLEs. *Ann. Statist.* **23** 339–362. MR1332570

SCHOOL OF STATISTICS
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MINNESOTA 55455
USA
E-MAIL: cdoss@stat.umn.edu
URL: http://users.stat.umn.edu/~cdoss/

DEPARTMENT OF STATISTICS
UNIVERSITY OF WASHINGTON
BOX 354322
SEATTLE, WASHINGTON 98195-4322
USA
E-MAIL: jaw@stat.washington.edu
URL: http://www.stat.washington.edu/jaw/