GLOBAL RATES OF CONVERGENCE OF THE MLES OF LOG-CONCAVE AND S–CONCAVE DENSITIES

By Charles Doss∗,‡ and Jon A. Wellner†,‡

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 \(-1/5 < s < \infty\) where \( s = 0 \) corresponds to the log-concave case.

CONTENTS

1 Introduction and overview ................. 2
2 Basic definitions and notation ............. 4
3 Maximum likelihood estimators: basic properties ...................... 5
  3.1 Log-concave densities: basic properties and consistency ....... 5
  3.2 s-concave densities: basic properties and consistency ......... 6
4 Bracketing entropy bounds and rates of convergence for log-concave
  and s-concave densities ........................................ 7
  4.1 Log-concave and s–concave densities: rates for the MLE ...... 7
5 Bracketing entropy bounds: extending Guntuboyina and Sen [2013] 10
6 Bracketing entropy bounds: the dual induction ...................... 12
7 Bracketing entropy bounds: putting the pieces together ............ 21
  7.1 Bracketing results .............................................. 21
  7.2 Rate results .................................................... 27
8 Appendix, part 1: Technical Lemmas and Inequalities ................. 35
  8.1 Nonexistence of the MLE for \( P_{1,s} \) with \( s < -1 \) .......... 35
  8.2 Some Technical Lemmas ......................................... 36
  8.3 Univariate s-concave and general \( h \)--concave density bounds 39
9 Appendix, part 2: Proofs of Propositions 5.1 and 5.2 ................. 48
10 Appendix, part 3: The Dual Induction Proof ........................ 52
Acknowledgements .................................................. 63
References ........................................................... 64

∗Supported by NSF Grant DMS-1104832
†Supported in part by NSF Grant DMS-1104832 and NI-AID grant 2R01 AI291968-04
AMS 2000 subject classifications: Primary 62G07, 62H12; secondary 62G05, 62G20
Keywords and phrases: bracketing entropy, consistency, empirical processes, global rate, Hellinger metric, induction, log-concave, s-concave

1
1. Introduction and overview.
We study global rates of convergence of nonparametric estimators of log-
concave densities, with focus on maximum likelihood estimation and the
Hellinger metric. A density \( p \) on \( \mathbb{R}^d \) is log-concave if
\[
p = e^{\varphi} \quad \text{where} \quad \varphi : \mathbb{R}^d \rightarrow \mathbb{R} \quad \text{is concave.}
\]
Log-concave densities are always unimodal and have convex level sets. Fur-
thermore, log-concavity is preserved under marginalization and convolution;
see e.g. Dharmadhikari and Joag-Dev \[1988\], chapter 2, pages 61-66. 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 as is nicely explained by Dharmadhikari and
Joag-Dev \[1988\], pages 84-99. These classes are defined by the generalized
means of order \( s \) as follows. Let
\[
M_s(a, b; \theta) \equiv \begin{cases} 
((1 - \theta)a^s + \theta b^s)^{1/s}, & s \neq 0, \ a, b \geq 0, \ \\
\min(a, b), & s = 0, \ \\
\min(a, b), & s = -\infty.
\end{cases}
\]
Then \( p \in \mathcal{P}_{d,s} \), the class of \( s \)-concave densities on \( C \subset \mathbb{R}^d \) if \( p \) satisfies
\[
p((1 - \theta)x_0 + \theta x_1) \geq M_s(p(x_0), p(x_1); \theta)
\]
for all \( x_0, x_1 \in C \) and \( \theta \in (0, 1) \). It is not hard to see that \( \mathcal{P}_{d,0} \) consists of
densities of the form \( p = e^{\varphi} \) where \( \varphi \in (-\infty, \infty) \) is concave, and densities \( p \)
in \( \mathcal{P}_{d,s} \) with \( s < 0 \) have the form \( p = \varphi^{1/s} \) where \( \varphi \in [0, \infty) \) is convex, and
with \( s > 0 \) have the form \( p = \varphi_{+}^{1/s} \) where \( \varphi \) is concave on \( C \) (and then we
write \( \mathcal{P}_{d,s}(C) \)); see for example Dharmadhikari and Joag-Dev \[1988\] page
86. These classes are nested since
\[
\mathcal{P}_{d,s}(C) \subset \mathcal{P}_{d,0} \subset \mathcal{P}_{d,r} \subset \mathcal{P}_{d,-\infty}, \quad \text{if} \quad -\infty < r < 0 < s < \infty.
\]
Here we view the classes \( \mathcal{P}_{1,s} \) defined above for \( d = 1 \) in terms of the gener-
alized 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 we define

\[ \mathcal{P}_{1,0} = \{ p = e^{\varphi} : \varphi \text{ is concave} \}, \]

\[ \mathcal{P}_{1,s} = \{ p = h_s(\varphi) : \varphi \text{ is concave} \}, \quad s < 0, \]

\[ \mathcal{P}_{1,s} = \{ p = h_s(\varphi) : \varphi \text{ is concave} \}, \quad s > 0 \]

where all the concave functions are assumed to be closed (i.e. upper semi-continuous), proper, and are viewed as concave functions on all of \( \mathbb{R} \) rather than on a (possibly) specific set \( C \). Thus we allow \( \text{ran}(\varphi) \subset [-\infty, \infty) \). See (2.1) in Section 2. This view simplifies our treatment in much the same way as the treatment in Seregin and Wellner [2010], but with “increasing” transformations replacing the “decreasing” transformations of Seregin and Wellner, and “concave functions” here replacing the “convex functions” of Seregin and Wellner.

Nonparametric estimation of log-concave and \( s \)-concave densities has developed rapidly in the last decade: For log-concave densities on \( \mathbb{R} \), Pal, Woodroofe and Meyer [2007] established existence of the Maximum Likelihood Estimator (MLE) \( \hat{p}_n \) of \( p_0 \), provided a method to compute it, and showed that it is Hellinger consistent: \( H(\hat{p}_n, p_0) \to_{a.s.} 0 \) where \( H^2(p,q) = (1/2) \int (\sqrt{p} - \sqrt{q})^2 dx \) is the (squared) Hellinger distance. Dümbgen and Rufibach [2009] also discussed algorithms to compute \( \hat{p}_n \) and rates of convergence with respect to supremum metrics on compact subsets of the support of \( p_0 \) under Holder smoothness assumptions on \( p_0 \). Balabdaoui, Rufibach and Wellner [2009] 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. Seregin and Wellner [2010] showed that the MLE exists and is consistent for the classes \( \mathcal{P}_{d,s} \) with \( s \in (-1, 0) \cup (0, \infty) \). Although it has been conjectured that the MLE is Hellinger-consistent at rate \( n^{-2/5} \) in the one-dimensional cases (see e.g. Seregin and Wellner [2010], pages 3378-3379), to the best of our knowledge this has not yet been proved.

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 seek bounds of the form

\[ \log N_{[]} (\epsilon, \mathcal{P}, H) \lesssim K \epsilon^{-1/2}, \quad \epsilon \leq \epsilon_0 \]

where \( N_{[]} (\epsilon, \mathcal{P}, H) \) denotes the minimal number of \( \epsilon \)-brackets with respect to the Hellinger metric \( H \) needed to cover \( \mathcal{P} \). We will establish such bounds in Section 5 using recent results of Dryanov [2009] for convex functions on \( \mathbb{R} \).
and Guntuboyina and Sen [2013] who extended the results of Dryanov from $\mathbb{R}$ to $\mathbb{R}^d$. These recent results build on earlier work by Bronštein [1976] and Dudley [1984]; see also Dudley [1999], pages 269-281. The main difficulty has been that the bounds of Bronštein [1976] 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 $\hat{p}_n$ (at least with arbitrarily high probability for large $n$). It is well-known that log-concave densities on $\mathbb{R}$ can have at most two discontinuities with these occurring at the endpoints of the support; see e.g. Schoenberg [1951], page 339. Since the estimators $\hat{p}_n$ are discontinuous at the upper and lower ends 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 end points of their support. Dryanov [2009] 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, Guntuboyina and Sen [2013] showed how to extend Dryanov’s results from $\mathbb{R}$ to $\mathbb{R}^d$. Here we show how the results of Dryanov [2009] and Guntuboyina and Sen [2013] 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 log-concave densities by an argument which we call dual induction, since it involves keeping track of brackets for concave (or convex) functions, and their transforms by an exponential function, simultaneously.

Once bounds of the form (1.1) are available, then relatively standard tools from empirical process theory going back to Birgé and Massart [1993], van de Geer [1993], Wong and Shen [1995], and developed further in van de Geer [2000] and van der Vaart and Wellner [1996], become available.

Our focus here will be on global convergence rates for the MLE’s $\hat{p}_n$ of log-concave and $s-$concave densities $p_0$ on $\mathbb{R}$. This seems to be a natural first step toward rates of MLEs and regularized versions of MLEs in the more difficult log-concave and $s-$concave problems with $d \geq 2$.

2. Basic definitions and notation. We will restrict attention to the class of concave functions

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

where Rockafellar [1970] 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. We also follow the convention that all concave functions $\varphi$ are defined on all of $\mathbb{R}$ and take the value
continuous with jump discontinuities (down to zero) at both $X$.

From Meyer [2007] and Dümbgen and Rufibach [2009] (Theorem 2.1) we know

over all concave functions $\phi$. Write $h_0(u) = e^{u}$, and, for $s < 0$, the class of $s$-concave densities $\mathcal{P}_{1,s}$ is a subset of the class of non-negative functions $\mathcal{F}_{1,s} = \{h_s \circ \phi : \text{concave}\}$ where $h_1(u) = (-u)^{1/s}$, $u < 0$, $h_s(u) = +\infty$ for $u \geq 0$. We will elaborate on this in Section 6. We will slightly abuse notation by allowing the domain and range operators to apply to such concave-transformed functions. In this case, we let $dom h \circ \phi := \{x : h(\phi(x)) > 0\}$ and $ran h \circ \phi := h(\phi(dom h \circ \phi))$.

3. Maximum likelihood estimators: basic properties. We divide our treatment here according to $s = 0$ and then $s \in (-1, 0) \cup (0, \infty)$.

3.1. Log-concave densities: basic properties and consistency. Let $X_1, \ldots, X_n$ be i.i.d. with density $p_0 = e^{\varphi_0}$ where $\varphi_0 : \mathbb{R} \to [-\infty, \infty)$ is concave. Thus $p_0$ is log-concave. 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 $\hat{\varphi}_n = \exp(\hat{\varphi}_n)$ of $p_0$ maximizes

$$
\Psi_n(\varphi) = \mathbb{P}_n \log p - \int \log p(x) dx = \mathbb{P}_n \varphi - \int e^{\varphi(x)} dx
$$

over all concave functions $\varphi$. From Walther [2002], Pal, Woodroofe and Meyer [2007] and Dümbgen and Rufibach [2009] (Theorem 2.1) we know that $\hat{\varphi}_n$ exists and is unique. It is linear on all intervals $[X_{(j)} , X_{(j+1)}]$, $j = 1, \ldots, n - 1$, where $X_{(1)} \leq \cdots \leq X_{(n)}$ denote the order statistics of the $X_i$’s. Furthermore $\hat{\varphi}_n = -\infty$ on $\mathbb{R} \setminus [X_{(1)} , X_{(n)}]$. Thus $\hat{\varphi}_n$ is upper semi-continuous with jump discontinuities (down to zero) at both $X_{(1)}$ and $X_{(n)}$.

The following lemma is basic.
Lemma 3.1. For any log-concave density $p$ on $\mathbb{R}$, there exist $a > 0$ and $b \in \mathbb{R}$ such that

$$p(x) \leq e^{-a|x|+b} \text{ for all } x \in \mathbb{R}.$$ 

This is a simplified version of Lemma A.1 of Dümbgen and Rufibach [2009]; earlier results with a similar spirit were given by Schoenberg [1951] and Devroye [1984]. An analogous result for log-concave densities on $\mathbb{R}^d$ is given in Lemma 1 of Cule and Samworth [2010].

Theorem 3.1. (Consistency and boundedness of $\hat{p}_n$ for $\mathcal{P}_{1,0}$)

(i) $H(\hat{p}_n, p_0) \to a.s. 0.$

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

$$\sup_{x \in S} |\hat{p}_n(x) - p_0(x)| \to a.s. 0.$$ 

(iii) If $p_0$ is continuous on $\mathbb{R}$, and $p_0(x) \leq e^{-a_0|x|+b_0}$ (by Lemma 3.1), then for any $0 \leq a < a_0$,

$$\sup_{x \in \mathbb{R}} e^{a|x|} |\hat{p}_n(x) - p_0(x)| \to a.s. 0.$$ 

(iv) For any $p_0$ log-concave with $p_0(x) \leq e^{-a_0|x|+b_0}$ (by Lemma 3.1), then for any $0 \leq a < a_0$,

$$\int_{\mathbb{R}} e^{a|x|} |\hat{p}_n(x) - p_0(x)| dx \to a.s. 0.$$ 

(v) $\limsup_{n \to \infty} \sup_x \hat{p}_n(x) \leq M(p_0) < \infty$ almost surely.

The first statement (i) is proved by Pal, Woodroofe and Meyer [2007]; statement (ii) is a corollary of Theorem 4.1 of Dümbgen and Rufibach [2009]; (iii) and (iv) are special cases of Theorem 4.1 of Cule and Samworth [2010], but (iv) with $a = 0$ is also given in Corollary 4.2 by Dümbgen and Rufibach [2009]. (v) This is Theorem 3.2 of Pal, Woodroofe and Meyer [2007] and will be needed to handle cases in which the mode(s) of $p_0$ are in the boundary of the support.

3.2. s-concave densities: basic properties and consistency. Let $X_1, \ldots, X_n$ be i.i.d. with density $p_0 = \varphi_0^{1/s}$ where $\varphi_0 : \mathbb{R} \to [0, \infty)$ is convex. Thus $p_0$ is s-concave. Write $P_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$ for the empirical measure of the $X_i$'s. The maximum likelihood estimator $\hat{p}_n = \varphi_n^{1/s}$ of $p_0$ maximizes

$$\Psi_n(\varphi) = \mathbb{P}_n \varphi^{1/s}$$
over all convex functions $\varphi$ for which $\int (\varphi)^{1/s}(x)dx = 1$. From Seregin and Wellner [2010] (Theorem 2.12, page 3757) we know that $\hat{\varphi}_n$ exists if $n \geq n_1 = 1/(r - 1)$ with $r \equiv -1/s > 1$ the case $s < 0$ and if $n \geq 2$ when $s > 0$. Seregin and Wellner [2010] page 3762 conjectured that $\hat{\varphi}_n$ is unique when it exists. Note that the class $P_{1, -\infty}$ corresponds to the class of all unimodal densities; see e.g. Dharmadhikari and Joag-Dev [1988] page 85, and for this class it is known that the MLE does not exist (see e.g. Birgé [1997]). In fact, it is easily seen that the MLE $\hat{p}_n$ of $p_0 \in P_{1, s}$ does not exist for any $s < -1$; see Section 8, Proposition 8.1.

**Theorem 3.2.** (Consistency and boundedness of $\hat{p}_n$ for $P_{1, s}$, $s \in (-1, 0) \cup (0, \infty)$)

(i) $H(\hat{p}_n, p_0) \to_{a.s.} 0$.

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

$$\sup_{x \in C} |\hat{p}_n(x) - p_0(x)| \to_{a.s.} 0.$$ 

(iii) If $p_0$ is continuous on $\mathbb{R}$, then

$$\sup_{x \in \mathbb{R}} |\hat{p}_n(x) - p_0(x)| \to_{a.s.} 0.$$ 

(iv) $\limsup_{n \to \infty} \sup_{x \in \mathbb{R}} \hat{p}_n(x) \leq M(p_0) < \infty$ almost surely.

The first statement (i) is Theorem 2.17 of Seregin and Wellner [2010]; statements (ii) and (iii) are consequences of Theorem 2.18 of Seregin and Wellner [2010]. (iv) is Lemma 3.17 in Seregin and Wellner [2010] and will be needed to handle cases in which the mode(s) of $p_0$ are in the boundary of the support (of $p_0$).

4. Bracketing entropy bounds and rates of convergence for log-concave and s-concave densities.

4.1. Log-concave and s-concave densities: rates for the MLE. Our main goal is to establish rates of convergence for the Hellinger consistency given in (i) of Theorems 3.1 and 3.2. It suffices, without loss of generality, to suppose that the mode (or smallest mode to be more precise) $m_0$ of $p_0$ is 0. Although the MLE $\hat{p}_n$ of $p_0$ will then have mode $\hat{m}_n \to_{a.s.} m_0 = 0$, $\hat{p}_n$ will usually not have mode exactly at 0. We therefore consider the following subclasses $P_{1, M, s}$ of log-concave and s-concave densities which will contain both $p_0$ and $\hat{p}_n$ with high probability for large $n$. Let $m_p$ denote the (smallest) mode
of $p$. Since all $s$-concave densities are unimodal, this is well-defined. Then, for $0 < M < \infty$, let

$$\mathcal{P}_{1,M,0} \equiv \left\{ p = e^{\varphi} : \varphi \text{ is concave}, \varphi \in [-\infty, \infty), m_p \in [-1, 1], 1/M \leq p(m_p) \leq M, \ p(0) > 0 \right\},$$

$$\mathcal{P}_{1,M,s} \equiv \left\{ p = \varphi_{+}^{1/s} : \varphi \text{ is convex}, m_p \in [-1, 1], \ 1/M \leq p(m_p) \leq M, \ p(0) > 0 \right\}, \ s < 0,$$

and

$$\mathcal{P}_{1,M,s} \equiv \left\{ p = \varphi_{+}^{1/s} : \varphi \text{ is concave}, m_p \in [-1, 1], \ 1/M \leq p(m_p) \leq M, \ p(0) > 0 \right\}, \ s > 0.$$

The following lemma gives upper envelopes for the classes $\mathcal{P}_{1,M,s}$ with $-1 < s \leq 0$.

**Lemma 4.1.** For any $p \in \mathcal{P}_{1,M,0}$ and $0 < M < \infty$,

$$p(x) \leq \begin{cases} 
Me \exp \left( -\frac{1}{M} (x - 1) \right), & x \geq 1, \\
Me \exp \left( -\frac{1}{M} |x + 1| \right), & x \leq -1, \\
Me, & x \in [-1, +1], 
\end{cases} \equiv p_{u,1,0}(x).$$

For any $p \in \mathcal{P}_{1,M,s}$ with $-1 < s < 0$ and $0 < M < \infty$

$$p(x) \leq \begin{cases} 
M \left( 1 + s - \frac{s}{M} (x - 1) \right)^{1/s}, & x \geq 1, \\
M \left( 1 + s - \frac{s}{M} (x - 1) \right)^{1/s}, & x \leq -1, \\
M(1 + s)^{1/s}, & x \in [-1, +1], 
\end{cases} \equiv p_{u,1,s-}(x).$$

For any $p \in \mathcal{P}_{1,M,s}$ with $0 < s \leq 1$ and $0 < M < \infty$

$$p(x) \leq \begin{cases} 
M \left( 1 + s - \frac{s}{M} (x - 1) \right)^{1/s}, & x \geq 1, \\
M \left( 1 + s - \frac{s}{M} (x - 1) \right)^{1/s}, & x \leq -1, \\
M(1 + s)^{1/s}, & x \in [-1, +1], 
\end{cases} \equiv p_{u,1,s+}(x).$$
Proof. Let \( P(x) \equiv \int_{-\infty}^{x} p(y) \, dy \) be the distribution function corresponding to \( p \). Then, by Lemma A.1 of Dümbgen and Rufibach [2009] (see also Proposition 8.2) with \( x_0 = m_p \):

\[
p(x) \leq p(m_p) e^{\exp\left(-\frac{p(m_p)}{1 - P(m_p)} \wedge \frac{p(m_p)}{P(m_p)} |x - m_p|\right)}
\leq M e^{\exp\left(-\frac{1}{M}|x - m_p|\right)} ,
\]

and the right side of the last display is upper bounded by \( p_{u,1,0} \) as claimed since

\[
|x - m_p| \geq (x - 1)1_{[1,\infty)}(x) + |x + 1|1_{(-\infty,-1]}(x) + 0 \cdot 1_{[-1,1]}(x)
\]

for all \( p \in \mathcal{P}_{1,M,0} \). The inequality (4.5) follows from the natural generalization of Lemma A.1 of Dümbgen and Rufibach [2009] with \( x_0 = m_p \) which is given in Proposition 8.2. For example, for \(-1 < s < 0\), with \( P(x) \equiv \int_{-\infty}^{x} p(y) \, dy \) the distribution function corresponding to \( p \),

\[
p(x) \leq p(m_p) \left(1 + s - s \frac{p(m_p)}{1 - P(m_p)} \wedge \frac{p(m_p)}{P(m_p)} |x - m_p|\right)^{1/s}
\leq M \left(1 + s - \frac{s}{M} |x - m_p|\right)^{1/s} ,
\]

and the right side of the last display is upper bounded by \( p_{u,1,s} \) as claimed since

\[
|x - m_p| \geq (x - 1)1_{[1,\infty)}(x) + |x + 1|1_{(-\infty,-1]}(x) + 0 \cdot 1_{[-1,1]}(x)
\]

for all \( p \in \mathcal{P}_{1,M,s} \). \( \square \)

Now let the bracketing entropy of a class of functions \( \mathcal{F} \) with respect to a semi-metric \( d \) on \( \mathcal{F} \) be defined in the usual way; see e.g. Dudley [1999] page 234, van der Vaart and Wellner [1996], page 83, or van de Geer [2000], page 16. With this preparation we can state our main results as follows: 

**Theorem 4.1.** Suppose that \( s \in (-1/5,1] \). Then

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

for all \( \epsilon \leq \epsilon_0 \) where the constant implied by \( \lesssim \) and \( \epsilon_0 \) depends only on \( M \) and \( s \).
Theorem 4.1 is the main tool we need to obtain rates of convergence for the MLEs \( \hat{p}_n \). This is given in our second main theorem:

**Theorem 4.2.** Suppose that \( s \in (-1/5, \infty) \) and \( \hat{p}_n \) is the MLE of the \( s \)-concave density \( p_0 \). Then \( n^{2/5} H(\hat{p}_n, p_0) = O_p(1) \).

Theorem 4.2 is a fairly straightforward consequence of Theorem 4.1 by applying van de Geer [2000], Theorem 7.4, page 99, or van der Vaart and Wellner [1996], Theorem 3.4.4 in conjunction with Theorem 3.4.1, pages 322-323. In addition, we have further consequences since the Hellinger metric dominates the total variation or \( L_1 \)-metric and via van de Geer [2000], Corollary 7.5, page 100:

**Corollary 4.1.** Suppose that \( s \in (-1/5, \infty) \) and \( \hat{p}_n \) is the MLE of the \( s \)-concave density \( p_0 \). Then \( n^{2/5} \int_{\mathbb{R}} |\hat{p}_n(x) - p_0(x)| dx = O_p(1) \).

**Corollary 4.2.** Suppose that \( s \in (-1/5, \infty) \). If \( \hat{p}_n \) is the MLE of the \( s \)-concave density \( p_0 \), then the log-likelihood ratio (divided by \( n \)) \( \mathbb{P}_n \log(\hat{p}_n/p_0) \) satisfies

\[
(4.7) \quad n^{4/5} \mathbb{P}_n \log \left( \frac{\hat{p}_n}{p_0} \right) = O_p(1).
\]

The result (4.7) 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 4.2 and Corollary 4.2 are also true for the constrained maximum likelihood estimator \( \hat{p}_n^0 \) 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 Doss [2013] and Doss and Wellner [2013].

**5. Bracketing entropy bounds: extending Guntuboyina and Sen [2013].** 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 \( L_r \) with \( r \geq 1 \) and related distances instead. We are really interested in \( L_r \) distances of concave-transformed classes of functions. Thus, when we measure distances between concave functions, we will not use the \( L_r \) metrics themselves, but new metrics that will give us information about \( L_r \) distance on the transformed scale. We will define such distances in Section 6. First, we will define the classes of concave and concave-transformed functions which we will be studying.
Definition 5.1. (i) For fixed $0 \leq u \leq \infty$, $-\infty < b_1 < b_2 < \infty$, and $-\infty < B_1 < B_2 < \infty$, we let $C([b_1, b_2], [B_1, B_2], u)$ be the class of all functions $\varphi \in C$ satisfying:

(a) The domain of $\varphi$ ends within $u$ of the endpoints, i.e. letting $\text{dom}(\varphi) := [d_{\varphi,1}, d_{\varphi,2}] \subseteq [b_1, b_2]$, we have

$$d_{\varphi,1} - b_1 \leq u \quad \text{and} \quad b_2 - d_{\varphi,2} \leq u,$$

(b) $\text{ran} \varphi \subseteq [B_1, B_2]$.

(ii) Similarly, for $u$, $b_1$, and $b_2$ as above, and for $0 \leq B_1 \leq B_2 < \infty$, we define $F_h([b_1, b_2], [B_1, B_2], u)$ to be the class of all concave-transformed functions $f = h \circ \varphi$ satisfying

(a) $\text{ran} f \subseteq [B_1, B_2]$,

(b) The domain of $f$ ends within $u$ of the endpoints, i.e. letting $\text{dom}(f) := [d_{f,1}, d_{f,2}] \subseteq [b_1, b_2]$, we have

$$d_{f,1} - b_1 \leq u \quad \text{and} \quad b_2 - d_{f,2} \leq u.$$

Note that this means

$$F_h([b_1, b_2], [B_1, B_2], u) = h \circ C([b_1, b_2], [h^{-1}(B_1), h^{-1}(B_2)], u).$$

In the above, we take $u = \infty$ to mean that the domains may be any subinterval of $[b_1, b_2]$.

(iii) When $F_h$ is a class of concave-transformed densities $p$, then we denote the class described in (ii) by $P_h([b_1, b_2], [B_1, B_2], u)$.

The above classes are the classes whose bracketing entropy we will control via the methods of Guntuboyina and Sen [2013] and the new methods developed here in Section 6.

Proposition 5.1. Let $r \geq 1$, $-\infty < b_1 < b_2 < \infty$, $-\infty < B_1 < B_2 < \infty$, and $0 < \epsilon \leq \epsilon_0(B_2 - B_1)(b_2 - b_1)^{1/r}$ for absolute constants $c_2 > 0$ and $\epsilon_0 > 0$. Then

$$\log N([\epsilon, C([b_1, b_2], [B_1, B_2], c_2 \epsilon^r(b_2 - b_1)), L_r(\lambda))] \leq c \left(\frac{(B_2 - B_1)(b_2 - b_1)^{1/r}}{\epsilon}\right)^{1/2}.$$ 

(5.1)

We may take all brackets $[l, u]$ such that $l(x) = B_1$ and $u(x) = B_2$ for all $x$ such that $|x - b_i| < c_2 \epsilon^r(b_2 - b_1)$, i.e. for all $x$ in the set where the domains may end.
We also note that we can simply state Theorem 3.1 in Guntuboyina and Sen [2013] in terms of bracketing entropy instead of metric entropy. This yields:

**Proposition 5.2 (Extension of Theorem 3.1 of Guntuboyina and Sen [2013]).** Let $r \geq 1$, $-\infty < b_1 < b_2 < \infty$, $-\infty < B_1 < B_2 < \infty$, and $0 < \epsilon \leq \epsilon_0(B_2 - B_1)(b_2 - b_1)^{1/r}$, where $\epsilon_0 > 0$ is an absolute constant. Then

$$\log N[(\epsilon, C([b_1, b_2], [B_1, B_2], 0), L_r(\lambda)) \leq \frac{c}{\epsilon} \frac{(B_2 - B_1)(b_2 - b_1)^{1/r}}{\epsilon^{1/2}},$$

where $c$ is a constant depending only on $r$.

**Proof.** The proof consists mostly of noticing that Theorem 3.1 in Guntuboyina and Sen [2013] essentially yields the results stated here. See Section 9 for the complete proofs.

Now the above result gives us a bracketing number for a single $\epsilon$ value that governs the size of the window in which the domains of the concave functions can vary, for classes of bounded functions. However, we are really interested in classes who allow domain endpoints to vary throughout the interval $[b_1, b_2]$. In the next section we will use the above result which allows for small windows of varying domain to build up to bracketing entropy control for classes with varying domain over the entire interval.

### 6. Bracketing entropy bounds: the dual induction.

The function classes in which we will be interested in the end are the classes $P_{1,M,0}$ or $P_{1,M,s}$ define in Section 4, or, more generally the classes $P_{1,M,h}$ defined in Section 7, to which the MLEs belong (with high probability as sample size gets large). However, such classes contain functions that are arbitrarily close to or are equal to 0, and these correspond to concave functions that take unboundedly large (negative) values. 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, and we will define a new metric for the concave functions that relates to the $L_r$ distance of the corresponding concave-transformed class.

**Definition 6.1.** A concave-function transformation, $h$, is a nondecreasing function from $[-\infty, \infty]$ to $[0, \infty]$ such that $h(\infty) = \infty$ and $h(-\infty) = 0$. We define its limit points $\tilde{y}_0 \leq \tilde{y}_\infty$ by $\tilde{y}_0 = \inf\{y : h(y) > 0\}$ and $\tilde{y}_\infty = \sup\{y : h(y) < \infty\}$. 
sup\{y : h(y) < \infty\}, we assume that h(\tilde{y}_0) = 0 and h(\tilde{y}_\infty) = \infty, and we define h_0 = \lim_{y \searrow \tilde{y}_0} h(y). We assume h is continuously differentiable on \((\tilde{y}_0, \tilde{y}_\infty)\).

The transformation h is not necessarily continuous at \(\tilde{y}_0\) if \(\tilde{y}_0\) is not \(-\infty\), so \(h_0\) is the minimum value of \(h\) that is not 0.

**Remark 6.1.** These transformations correspond to “decreasing transformations” in the terminology of Seregin and Wellner [2010]. 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 defines a decreasing transformation \(\tilde{h}\) as defined in Seregin and Wellner [2010] via \(\tilde{h}(y) = h(-y)\).

We will sometimes make the following assumptions.

**Assumption 6.1.** Assume that the transformation h satisfies:

T.1 \(h'(y) = o((-y)^{-(\alpha+1)})\) as \(y \searrow \tilde{y}_0\) 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 \(h'(y) \leq M_c\) for all \(y \in (\tilde{y}_0,c]\);

T.3 if \(\tilde{y}_\infty < \infty\) then for some \(0 < c < C, c(y - \tilde{y}_\infty)^{-\beta} \leq h(y) \leq C(y - \tilde{y}_\infty)^{-\beta}\) 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 \(y \to \infty\).

Note that Assumption (T.2) does not preclude \(h\) from being discontinuous at \(\tilde{y}_0\) when \(\tilde{y}_0 < \infty\). Additionally, notice this assumption holds automatically if \(\tilde{y}_0 = -\infty\) when Assumption (T.1) holds.

**Definition 6.2.** For a decreasing vector \(y_k = (y_0, \ldots, y_k) \in \mathbb{R}^{k+1}\), i.e. a vector with \(\tilde{y}_\infty \geq y_0 > y_1 > \ldots > y_k \geq \tilde{y}_0\), for \(-\infty < b_1 < b_2 < \infty\), and letting \(D(b_1) := \{\varphi \in C| \text{dom } \varphi = [b_1, d_\varphi] \text{ or } \text{dom } \varphi = \emptyset\}\) we define

\[
C_k \equiv C([b_1, b_2], [y_k, y_0], \infty) \cap D(b_1),
\]

a class of concave functions whose domains have right endpoint which may vary freely in \([b_1, b_2]\). We also define the concave-transformed functions \(\mathcal{F}_{k,h} = h \circ C_k\).

**Example 6.1.** The class of log-concave densities, as discussed in Section 4.1 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 6.2. The classes of $s$-concave functions with $s \in (-1,0)$, as discussed in Section 4.1 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 $\alpha \in (1, -1/s)$.

Example 6.3. The classes of $s$-concave functions with $0 < s < \infty$, as discussed in Section 4.1 are obtained by taking $h(y) = (y)^{1/s} \equiv h_s(y)$ for $s \in (0,\infty)$. Here $\tilde{y}_0 = 0$ and $\tilde{y}_\infty = \infty$. Assumption (T.1) holds for any $\alpha > 1 > -1/s$, Assumption (T.2) fails if $s > 1$, and Assumption (T.4) holds for any (small) $C, \gamma > 0$. These (small) classes $P_h$ are covered by our Corollary 7.3.

Example 6.4. To connect the preceding two examples, consider $\tilde{h}_s(y) = \log(1 + sy)^{1/s}$ for $y \in (-\infty, -1/s)$ with $-1 < s < 0$. Set $\tilde{h}_s(y) = e^y$ as $s \uparrow 0$. 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 $\alpha \in (1, -1/s)$.

Example 6.5. To illustrate the possibilities further, consider $h(y) = \log(1 + sy)^{1/s}$ 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 this example fails to satisfy Assumption (T.4) when $r < 0$ and $s = 0$ (and then $\tilde{y}_\infty = \infty$).

We are interested in bracketing entropy of $F_{k,h}$ with the $L_r(\lambda)$ norm but we can get control of bracketing entropy for $C_k$ since it is a class of concave functions. Thus we define a metric on the latter space that relates to $L_r(\lambda)$ distance on $F_{k,h}$.

Definition 6.3. Let $h$ be a concave-function transformation and assume we have a decreasing sequence $\tilde{y}_\infty > y_0 > y_1 \cdots > y_k > \infty$, denoted $y_k = (y_0, y_1, \ldots, y_k) \in \mathbb{R}^{k+1}$. If $\tilde{y}_0 > -\infty$ then if $h$ is discontinuous at $\tilde{y}_0$, we also assume $y_k > \tilde{y}_0$, but if $h$ is continuous at $\tilde{y}_0$ (i.e. approaches 0 from the right) we allow $y_k$ to possibly be $\tilde{y}_0$. Set $w_j \equiv w_{j,h} = \sup_{y \in [y_j,y_{j-1}]} h'(y)$ for $j = 1, \ldots, k + 1$. Then for $x \in \mathbb{R}$ define

$$W(x) \equiv W(x; y_k, h) \equiv \sum_{j=1}^k w_j \lambda([y_k, x] \cap [y_j, y_{j-1}])$$

where $\lambda$ denotes Lebesgue measure on $\mathbb{R}$. (Alternatively, for Borel subsets $A$
of \( \mathbb{R} \) define the measure \( W \) by

\[
W(A) \equiv W(A; y_k, h) \equiv \sum_{j=1}^{k} w_j \lambda(A \cap [y_j, y_{j-1}]).
\]

Then for \( a, b \in \mathbb{R} \equiv [-\infty, \infty] \) define the (weighted) distance \( d_{k,h} \equiv d_{y_{k+1},h} \) by

\[
d_{k,h}(a, b) \equiv d_{y_{k+1},h}(a, b) \equiv \int_{a \wedge b}^{a \vee b} dW(x) = W(a \vee b) - W(a \wedge b).
\]

Note that if we make Assumption (T.2) then by the definition of \( y_k \) and the fact that \( h' \) is continuous on \( (\tilde{y}_0, \tilde{y}_\infty) \), then in all scenarios (i.e. if \( \tilde{y}_0 = -\infty \) or if \( \tilde{y}_0 \) is finite and \( h \) is continuous or discontinuous) the weights \( w_j \) are finite.

Next, we define the \( L_r(\lambda) \) generalization of the above metric by integrating.

**Definition 6.4.** Let \( \lambda \) denote Lebesgue measure on \( \mathbb{R} \), let \( h \) be a concave-function transformation, and let \( y_k \) be as in Definition 6.3. For two functions \( \varphi_1 \) and \( \varphi_2 \) defined on \( [b_1, b_2] \) we define

\[
d_{r,k,h}(\varphi_1, \varphi_2) = d_{y_{k+1},h}(\varphi_1, \varphi_2) = \left( \int_{x \in \text{dom}(\varphi_1) \cup \text{dom}(\varphi_2)} d_{y_{k+1},h}(\varphi_1(x), \varphi_2(x))^r \, dx \right)^{1/r} = \left( \int_{b_1}^{b_2} d_{y_{k+1},h}(\varphi_1(x), \varphi_2(x))^r \, dx \right)^{1/r},
\]

where we take \( \varphi_1 \) or \( \varphi_2 \) to be \(-\infty\) outside their respective domains.

Note that since \( dW \) only puts mass on \([y_k, y_0]\), the distance is always finite (since the two domains are bounded). \( d_{r,k,h}(\cdot, \cdot) \) is indeed a metric; i.e. the triangle inequality holds; see Lemma 8.2 for the proof.

We will generally apply \( d_{r,k,h} \) to concave functions, but in some instances it will be useful to take \( h(x) = \text{Id}(x) := x \) and apply \( d_{r,k,\text{Id}} \) to log-concave functions (or brackets thereof), which yields a truncated version of \( L_r \) distance. The next results provide the motivation for using these new metrics \( d_{r,k,h} \).

**Lemma 6.1.** Let \( h \) be a concave-function transformation, let \( y_k \) be as in Definition 6.3, let \( \varphi_i \in \mathcal{C}_k \), and take \( r \geq 1 \). Then

\[
d_{r,(h(y_0), h(y_k)), \text{Id}}(h \circ \varphi_1, h \circ \varphi_2) \leq d_{r,k,h}(\varphi_1, \varphi_2), \tag{6.1}
\]
and

\begin{equation}
\|(h \circ \varphi_1 - h \circ \varphi_2)1_{\text{dom } \varphi_1 \cap \text{dom } \varphi_2}\|_r \leq d_{r,k,h}(\varphi_1, \varphi_2).
\end{equation}

**Proof.** For \(a, b \in (\tilde{y}_0, \tilde{y}_\infty)\), which is the set on which \(h'\) is defined, we have

\[
\int_{a \wedge b}^{a \vee b} h'(z)dz = \sum_{i=1}^{k} \int_{a \wedge b}^{a \wedge b} 1_i(z)h'(z)dz
\]

where

\[
1_i(z) = 1_i(z,a,b) = 1^L_i(z)1^U_i(z) = 1_{L_i \cap U_i}(z),
\]

and where

\[
1^L_i(z) = 1_{[a \wedge b, y_{i-1}]}(z) = 1_{L_i}(z), \quad 1^U_i(z) = 1_{[y_i, a \vee b]}(z) = 1_{U_i}(z).
\]

We take a fixed \(x\) and assume without loss of generality that \(\varphi_1(x) \leq \varphi_2(x)\). While, for \(x\) in the domain of \(\varphi_i\), for \(i = 1\) or \(2\), \(\varphi_i(x) \in [y_k, y_0]\) because \(\varphi_i \in C_k\), we do not assume \(x\) lies in the domain of \(\varphi_i\). Thus for now we just assume that \(\varphi_i\) are concave functions, possibly taking values of \(-\infty\). For symmetry, we also allow \(\varphi_i\) to take values larger than \(y_0\). We then have

\[
0 \leq h(\varphi_2(x) \wedge y_0) - h(\varphi_1(x) \vee y_k) = h(\varphi_2(x) \wedge y_0) - h(\varphi_1(x) \vee y_k) -
\]

\[
= \int_{\varphi_1(x) \vee y_k}^{\varphi_2(x) \wedge y_0} h'(z)dz
\]

\[
= \sum_{\gamma=1}^{k} \int_{\varphi_1(x) \vee y_k}^{\varphi_2(x) \wedge y_0} 1_\gamma(z, \varphi_1(x), \varphi_2(x))h'(z)dz
\]

where the first inequality is because \(h\) is increasing and the next equality is because \(y_k\) is only allowed to be equal to \(\tilde{y}_0\) if \(h\) is continuous at \(\tilde{y}_0\). The next equality is because \(h\) is continuously differentiable on \((\tilde{y}_0, \tilde{y}_\infty)\) (\(h\) may not be differentiable at \(\varphi_1(x) \vee y_k\) if this equals \(\tilde{y}_0\), but this singleton point has Lebesgue measure 0 so we can ignore it). The last equality is because \(1_\gamma(\cdot, \varphi_1(x), \varphi_2(x))\) are indicators of disjoint intervals (excluding their endpoints) which partition \((\varphi_1(x) \vee y_k, \varphi_2(x) \wedge y_0)\). Next, we see that the above display is bounded above by

\[
\sum_{\gamma=1}^{k} \sup_{y \in [y_{\gamma}, y_{\gamma-1}]} h'(y) \int_{y_k}^{y_0} 1_\gamma(z, \varphi_1(x), \varphi_2(x))dz
\]

\[
\leq \sum_{\gamma=1}^{k} \sup_{y \in [y_{\gamma}, y_{\gamma-1}]} h'(y) \int_{y_k}^{y_0} 1_\gamma(z, \varphi_1(x), \varphi_2(x))dz
\]

\[
= d_{y_k, h}(\varphi_1(x), \varphi_2(x)).
\]
Letting $f_i = h \circ \varphi_i$, $i = 1, 2$, we see that
\[
\begin{align*}
  d_{r,(h(y_0), h(y_k))}(f_1, f_2)^r &= \int_{b_1}^{b_2} \left( \int_{h(\varphi_1(x) \vee y_k)}^{h(\varphi_2(x) \wedge y_0)} dz \right)^r dx \\
  &\leq \int_{b_1}^{b_2} d_{y,h}(\varphi_1(x), \varphi_2(x))^r dx,
\end{align*}
\]
which is (6.1). This holds for any concave $\varphi_i$, $i = 1, 2$. Now, assume $\varphi_i \in \mathcal{C}_k$, $i = 1, 2$, and see that the above display immediately yields (6.2) since the $L_r$ distance equals $d_{r,(h(y_0), h(y_k))}$ for functions in $\mathcal{F}_{k,h}$ with the same domain, i.e. if we let $A = \text{dom } f_1 \cap \text{dom } f_2$ then
\[
\| (f_1 - f_2)^1_A \|_r = d_{r,(h(y_0), h(y_k))}(f_1^1_A, f_2^1_A),
\]
since on the intersection of their domains functions in $\mathcal{F}_{k,h}$ have values in $[h(y_k), h(y_0)]$. Clearly
\[
d_{r,(h(y_0), h(y_k))}(f_1^1_A, f_2^1_A) \leq d_{r,(h(y_0), h(y_k))}(f_1, f_2),
\]
so (6.1) implies (6.2).

This means we can control entropy of the classes $\mathcal{F}_{k,h}$ of log-concave functions in terms of entropy with the $d_{y,h}$ metric on concave functions.

**Lemma 6.2.** Fix $1 \leq r < \infty$. Let $-\infty < b_1 < b_2 < \infty$, $-\infty < B_1 < B_2 < \infty$, and let $[B_1, B_2] \subseteq (\bar{y}_0, \bar{y}_\infty)$. Then
\[
N_{[\cdot]}(\epsilon, \mathcal{F}_{k,h}, L_r(\lambda)) \leq N_{[\cdot]}(\epsilon, \mathcal{F}_{k,h}, d_{r,(h(y_0), h(y_k))}, \mathcal{I}_d) \leq N_{[\cdot]}(\epsilon, \mathcal{C}_k, d_{r,k,h}),
\]
(6.3)
for all $\epsilon > 0$. Moreover, setting $w_1 = \sup y \in [B_1, B_2] |h'(y)|$, we have
\[
N_{[\cdot]}(\epsilon, h \circ C([b_1, b_2], [B_1, B_2], 0), L_r(\lambda)) \leq N_{[\cdot]}(\epsilon, C([b_1, b_2], [B_1, B_2], 0), d_{r,(B_2, B_1)}, h) \leq N_{[\cdot]}(\epsilon/w_1, C([b_1, b_2], [B_1, B_2], 0), L_r(\lambda)) \leq \exp \left( c \left( \frac{(B_2 - B_1)(b_2 - b_1)^{1/r}}{\epsilon/w_1} \right)^{1/2} \right),
\]\n(6.4)
for all $\epsilon \leq \epsilon_0 w_1(B_2 - B_1)(b_2 - b_1)^{1/r}$ where $\epsilon_0$ is the constant depending only on $r$ taken from Proposition 5.2.
PROOF. Given an $\epsilon-$bracket $[l, u]$ for $C_k$ with respect to $d_{r,k,h}$, we can immediately construct an $\epsilon-$bracket $[h \circ l, h \circ u]$ for $F_{k,h}$ with respect to $d_{r,(h(y_0),h(y_k))}\triangleq$. Lemma 6.1 says that these new brackets have smaller size than the original, so the second inequality in (6.3) holds. The first inequality is trivial just by constraining all brackets to take values in $[h(y_k), h(y_0)]$.

The same argument applies to show (6.4), where we use the last comment of Lemma 6.1 to apply that lemma to $C([b_1, b_2], [B_1, B_2], 0)$ This gives the first inequality of (6.4). The next inequality follows by noticing that

$$d_{r,(B_2,B_1),h} (\varphi_1, \varphi_2) = w_1 \| \varphi_1 - \varphi_2 \|_r,$$

where $w_1 = \sup_{x \in [b_1, b_2]} |h'(x)|$, so that if for a bracket $[l, u]$ we have $\|u-l\|_r < \epsilon/w_1$, then we have $d_{r,(B_2,B_1),h} (\varphi_1, \varphi_2) < \epsilon$. The final inequality is simply Proposition 5.2. \hfill \Box

Remark 6.2. If we take functions $f_1$ and $f_2$ which take values in $[y_{k-1}, y_0]$, then from the definitions of $d_{r,k,h}$ and $d_{r,k-1,h}$, we can see that

$$d_{r,k,h}(f_1, f_2) = d_{r,k-1,h}(f_1, f_2).$$

This is because it is true pointwise, i.e. $d_{k,h}(p_1, p_2) = d_{k-1,h}(p_1, p_2)$ if $p_1$ and $p_2$ are in $[y_{k-1}, y_0]$, since the corresponding measures $w_{\mathcal{L},h}$ and $w_{\mathcal{L},h}$ are the same on $[y_{k-1}, y_0]$.

Similarly, since the measure $W$ corresponding to $d_{r,k,IA}$ is Lebesgue measure $\lambda$ on $[y_k, y_0]$, for two functions $f_1$ and $f_2$ that take values in $[y_k, y_0]$, we have

$$\|f_1 - f_2\|_r = d_{r,k,IA}(f_1, f_2).$$

Remark 6.3. The space $(F_{k,h}, d_{r,(h(y_0),h(y_k))},\triangleq)$ corresponds to the space $(C_k, d_{\mathcal{L},h})$ (for any decreasing vector $\mathcal{L}$). The metric on the former space is not quite the $L_r(\lambda)$ metric because distance is truncated when functions fall below the cutoff $h(y_k)$. Thus to get from $d_{r,(h(y_0),h(y_k))}\triangleq(f_1, f_2)$ to $\|f_1 - f_2\|_r$, we need to also control the difference in the domains of $f_1$ and $f_2$.

We can now state the main technical proposition needed for the bracketing entropy bound on $F_{k,h}$. For now, we do not make any of (T.1)–(T.4) of Assumption 6.1 on $h$, which may mean that some of the weights $w_{\gamma,h}$ are infinite; in such a case, the conclusion of the following proposition is tautological.
Proposition 6.1. Let $h$ be as in Definition 6.1, let $y_k \in R^{k+1}$ as in Definition 6.3, and let $d_{r,k,h}$ and its corresponding weight sequence $w_{\gamma,h}$ be as in Definition 6.4. Fix $\epsilon > 0$ and assume that for all $1 \leq \gamma \leq k$, $y_\gamma$ satisfy

$$\frac{\epsilon}{w_{\gamma,h}} \leq \epsilon_0 (b_2 - b_1)^{1/r} (y_{\gamma-1} - y_\gamma),$$

for $\epsilon_0 > 0$ a positive constant (not necessarily defined as in Proposition 5.2).

Then we have for any $\zeta \in [0,1]$, $1 \leq r < \infty$,

$$\log N_{[\epsilon]} \left( \epsilon \cdot \left(1 + \sum_{\gamma=1}^k h(y_{\gamma-1})^{r(1-\zeta)} \right)^{\frac{1}{r}} \right)_{F_{k,h}, L_r(\lambda)} \leq \epsilon \sum_{\gamma=1}^k \left( \left( \frac{(y_{\gamma-1} - y_\gamma)(b_2 - b_1)^{1/r}}{\epsilon/w_{\gamma,h}} \right)^{1/2} \right)$$

$$+ 2 \log \left( 1 + \frac{b_2 - b_1}{\epsilon^{r/h}(y_{\gamma-1})^{r\zeta}} \right).$$

Proof. In this proof we will use $L_r$ to denote $L_r(\lambda)$ where $\lambda$ is Lebesgue measure. We will assume that all $w_{\gamma,h}$ are finite, since otherwise the right-hand side of (10.3) is infinite and the conclusion is trivial. Additionally, note that (10.1) forces $w_{\gamma,h} > 0$ for all $\gamma$.

We will first reduce to the case $b_1 = 0$ and $b_2 = 1$ by translating and scaling the domain. We will not scale the $y_\gamma$, however. For any function $f \in F_{k,h}$ with $f \subseteq [b_1, b_2]$, we can define a function $\tilde{f}(x) = f(b_1 + (b_2 - b_1)x)$ with $\tilde{f} \subseteq [0,1]$. $\tilde{f}$ is an element of $F_{k,h}$ with domain specified as $[0,1]$. If the proposition holds when $b_1 = 0$ and $b_2 = 1$, then we can find a collection of brackets $\tilde{l}$ and $\tilde{u}$ such that $\tilde{l} \leq \tilde{f} \leq \tilde{u}$. We can invert the above scaling to arrive at $l$ and $u$ (e.g. $l(x) = \tilde{l}((x - b_1)/(b_2 - b_1))$) such that $l \leq f \leq u$. For any two functions $f_1$ and $f_2$ defined on $[0,1]$, we have

$$d_{r,k,h}(f_1, f_2) = \left( \int_0^1 d_{k,h}(f_1(x), f_2(x))^{r} dx \right)^{1/r}$$

$$= \left( \int_{b_1}^{b_2} d_{k,h}(\tilde{f}_1(x), \tilde{f}_2(x))^{r} \frac{dx}{b_2 - b_1} \right)^{1/r}$$

$$= \frac{1}{(b_2 - b_1)^{1/r}} d_{r,k,h}(\tilde{f}_1, \tilde{f}_2).$$

Applying this to $[\tilde{l}, \tilde{u}]$ and the corresponding $[l, u]$, we see that if we are given a collection of $\epsilon$-brackets when $b_1 = 0$ and $b_2 = 1$ then we immediately have
a collection of \((b_2 - b_1)^{1/r} \epsilon\)-brackets of the same cardinality for general \(b_1\) and \(b_2\), i.e. if the proposition holds when \(b_1 = 0\) and \(b_2 = 1\) then it holds for general \(b_1\) and \(b_2\).

We thus now proceed with \(b_1 = 0\) and \(b_2 = 1\). We set \(\eta_{\gamma,h} = \epsilon/h(y_{\gamma-1})^{c}\) and \(\epsilon_{\gamma,h} = \epsilon/w_{\gamma,h}\). Note that \(0 < \epsilon_{\gamma,h} < \infty\) by the comments at the beginning of the proof. The proof now proceeds by induction. The induction hypothesis is as follows.

**Induction Hypothesis:**
1. For \(k^* \in \mathbb{N}\), we assume that:
   (a) (Bracketing) There is a collection of \(d_{r,k^*,h}\) brackets \([l_\beta, u_\beta]\) for the class \(C_{k^*}\) (defined in Definition 6.2, using \(b_1 = 0\) and \(b_2 = 1\));
   (b) (Cardinality) and that they have cardinality

\[
\exp \left( \epsilon \sum_{\gamma=1}^{k^*} \left( \frac{(y_{\gamma-1} - y_{\gamma})}{\epsilon/w_{\gamma,h}} \right)^{1/2} \right) + (1 + 1_{1\leq \gamma \leq k^*-1}) \log \left( 1 + \frac{1}{\epsilon^r/h(y_{\gamma-1})^{c}} \right).
\]

(6.9)

2. Denoting \(\text{dom}(\varphi)\) by \([0, d_{\varphi}]\) we now define the classes

\[
C_{\gamma,j} := \{ \varphi \in C_\gamma | d_{\varphi} \in ((j-1)\eta_{\gamma}, j\eta_{\gamma}] \} \subseteq C ((0, j\eta_{\gamma}], [y_{\gamma}, y_0], \eta_{\gamma})
\]

for \(j = 1, \ldots, N_\gamma := \lceil 1/\eta_{\gamma,h} \rceil\), which disjointly partition \(C_\gamma, \gamma = 1, \ldots, k^*\). We assume that each bracket \([l_\beta, u_\beta]\) has a corresponding index \(j_{\gamma,\beta}\), and:

(a) (Partition) We assume that all the brackets \([l_\beta, u_\beta]\) that share an index \(j = j_{k^*,\beta}\) form a collection of brackets for the class \(C_{k^*,j}\);

(b) (Extension) we assume that if \(\varphi \in C_{\gamma,j}\), for some \(1 \leq j \leq N_\gamma\) and \(1 \leq \gamma \leq k^*\), and \(\varphi\) is bracketed by \([l_\beta, u_\beta]\) (which thus has index \(j_{\gamma,\beta} = j\)), then any \((j, \gamma)\)-concave extension \(\phi\) of \(\varphi\) is bracketed by \([l_\beta, u_\beta]\), where a \((j, \gamma)\)-concave extension \(\phi\) is a function such that

\[
\phi(x) = \varphi(x) \text{ for } x \in \text{dom}(\varphi), \quad \text{dom}(\phi) \subseteq [0, j\eta_{\gamma,h}],
\]

\[
\phi(x) \leq y_0 \text{ for } x \in \mathbb{R}, \quad \text{and } \phi \in C.
\]

3. (Size) We assume each bracket \([l_\beta, u_\beta]\) has a corresponding set of “jump-down” intervals \(J_{\gamma,\beta} := ((j_{\gamma,\beta}-1)\eta_{\gamma,h}, j_{\gamma,\beta}\eta_{\gamma,h}] [0, j_{\gamma,\beta}\eta_{\gamma,h}], \gamma = 1, \ldots, k^*\), where we take \(j_{0,\beta}\eta_{0,h}\) to be 0 for all \(\beta\). For \(\gamma = 1, \ldots, k^*\), we assume for all \(\beta\) that the brackets \([l_\beta, u_\beta]\) satisfy:

(a) for all \(k \geq k^*\)

\[
d_{r,k,h}(u_\beta \cdot 1_{(j_{\gamma,\beta}-1)\eta_{\gamma,h}]}, l_\beta \cdot 1_{(j_{\gamma,\beta}\eta_{\gamma,h})}) \leq \epsilon(j_{k^*,\beta}\eta_{k^*,h})^{1/r}.
\]

(6.10)
(b) and

\[(6.11) \quad \|(h \circ u_\beta - h \circ l_\beta)_{\gamma=1}^{k^*} f_{\gamma,\beta}\|_r^p \leq \epsilon^r \sum_{\gamma=1}^{k^*} h(y_{\gamma-1})^{r(1-\zeta)}.
\]

4. (Conclusion) We assume that this transformed collection of brackets 
\([h \circ l_\beta, h \circ u_\beta]\) for \(F_{k^*,h}\) has size not larger than 
\((1 + \sum_{\gamma=1}^{k^*} h(y_{\gamma-1})^{r(1-\zeta)})^{1/r} \epsilon\)

in the \(L_r\) metric, which yields

\[
\log N_{[\epsilon]} \left( \epsilon \left( 1 + \sum_{\gamma=1}^{k^*} h(y_{\gamma-1})^{r(1-\zeta)} \right)^{1/r}, F_{k^*,h}, L_r \right) 
\leq c \sum_{\gamma=1}^{k^*} \left( \frac{(y_{\gamma-1} - y_{\gamma})}{\epsilon/w_{\gamma,h}} \right)^{1/2} 
+ (1 + 1_{[\gamma \leq k^* - 1]}) \log \left( 1 + \frac{1}{\epsilon^r/h(y_{\gamma-1})^{r\zeta}} \right) .
\]

This ends the induction hypothesis.

The remainder of the proof of Proposition 10.1 is given in Section 10.

Here we give a brief outline of how it works. For our decreasing sequence 
\(y_0 > y_1 > \ldots\), we can consider the domain on which a concave function 
\(\varphi\) takes values in \([y_{k-1}, y_0]\) and inductively form brackets there, and we can consider the domain on which \(\varphi\) takes values in \([y_k, y_{k-1}]\) and form brackets there via Proposition 5.2. Then we merge the two brackets together. See Figure 1 for a diagram of these three steps and Section 10 for all the details. \(\square\)

The condition (10.1) is not fundamental. It is essentially keeping \(\epsilon\) from being too large, which is an unimportant constraint. The proposition could be phrased without this condition, but we phrase it with the condition and then pick \(y_k\) sequences later that satisfy it.

7. Bracketing entropy bounds: putting the pieces together.

7.1. Bracketing results. We now use the above result to prove an actual bracketing entropy bound of the type in which we are interested. Recall, for \(b_1 < b_2\) and \(B > 0\), Definition 5.1 of \(\mathcal{F}([b_1, b_2], [0, B], \infty)\), and recall the definition of \(\mathcal{D}(b_1)\) in Definition 6.2.
Fig 1. Diagram of the inductive bracketing construction for $k = 3$. The uppermost figure diagrams the inductive bracketing on $C_{k-1,i}$. The middle figure diagrams the application of Proposition 5.2 to get brackets for the class (10.23). The final figure diagrams merging the two previous types of brackets to form brackets for $C_k$. 
THEOREM 7.1. Let $r \geq 1$. Assume $h$ is a concave-function transformation and that Assumptions (T.1) and (T.2) hold, and let 

$$
\mathcal{G} \equiv \mathcal{F}_h([b_1, b_2], [0, B], \infty) \cap h \circ \mathcal{D}(b_1).
$$

Assume $h$ is continuous. For some $\epsilon_0 > 0$ and all $\epsilon \leq \epsilon_0 B(b_2 - b_1)^{1/r}$, we have 

$$
\log N_{\|}(\epsilon, \mathcal{G}, L_r(\lambda)) \lesssim \left( \frac{B(b_2 - b_1)^{1/r}}{\epsilon} \right)^{1/2},
$$

where $\lesssim$ means $\leq$ up to a constant depending only on $r$ and $h$. Similarly, $\epsilon_0$ is a constant depending on $r$ and $h$.

PROOF. We will first reduce to the case where $b_1 = 0, b_2 = 1,$ and $B = 1$. For any function $f \in \mathcal{F}([b_1, b_2], [0, B], \infty) = h \circ \mathcal{C}([b_1, b_2], [h^{-1}(0), h^{-1}(B)], \infty)$, we can define a function $\tilde{f}(x) = f(b_1 + (b_2 - b_1)x)/B$. Then $\tilde{f}$ is an element of $\tilde{h} \circ \mathcal{C}([0, 1], [\tilde{h}^{-1}(0), \tilde{h}^{-1}(1)], \infty)$ where $\tilde{h}(y) = h(y)/B$ (i.e. it is a function of the type $\tilde{h}(\varphi(x))$, where $\varphi \in \mathcal{C}$ and dom $\tilde{h} \subseteq [0, 1]$ and ran $\tilde{h} \subseteq [0, 1]$). Note that of course if $h$ is a concave-function transformation then $\tilde{h}$ is a concave-function transformation. Thus if the proposition holds when $b_1 = 0, b_2 = 1,$ and $B = 1$, then we can find a collection of brackets for the class $\tilde{h}^{-1} \circ \mathcal{C}([0, 1], [\tilde{h}^{-1}(0), \tilde{h}^{-1}(1)], \infty)$. If one of these brackets $[l, U]$ satisfies $\tilde{l} \leq \tilde{f} \leq \tilde{u}$ then by inverting the above scaling relationship we can arrive at $l$ and $u$ (e.g. $l(x) = \tilde{l}((x - b_1)/(b_2 - b_1))$) such that $l \leq f \leq u$. We see that the size of the bracket $[l, U]$ is related to the size of $[\tilde{l}, \tilde{u}]$ by

$$
B^r \int_0^1 \left| \tilde{u}(x) - \tilde{l}(x) \right|^r dx = \frac{1}{b_2 - b_1} \int_{b_1}^{b_2} |u(x) - l(x)|^r dx.
$$

Thus, if we have a collection of $\epsilon$-brackets for $\tilde{h} \circ \mathcal{C}([0, 1], [\tilde{h}^{-1}(0), \tilde{h}^{-1}(1)], \infty)$ then we immediately get a collection of $B(b_2 - b_1)^{1/r} \epsilon$-brackets of the same cardinality for $h \circ \mathcal{C}([b_1, b_2], [0, B], \infty)$. That is, if the theorem holds for any concave-function transformation $h$ and $b_1 = 0, b_2 = 1$, $B = 1$, then it holds for any concave-function transformation and general $b_1, b_2$, and $B$.

Thus, we now consider $b_1 = 0, b_2 = 1$, and $B = 1$, and fix $0 < \epsilon \leq \epsilon_0$. By replacing $h$ by a translation of $h$, and using the fact that $\text{ran} \ h = (0, \infty)$, we assume that $h^{-1}(1) < 0$. (Translating $h$ has no effect on the class $\mathcal{G}$, since concave functions plus a constant are still concave.) We apply Proposition 10.1. To do so, we need to pick the sequence $y_\gamma, \gamma = 1, \ldots, k$, such that
S.1 $\sum_{\gamma=1}^{k} h(y_{\gamma-1})^{r(1-\zeta)} < C < \infty,$
S.2 $\sum_{\gamma=1}^{k} (w_{\gamma,h}(y_{\gamma-1} - y_{\gamma}))^{1/2} < C < \infty,$
S.3 $\sum_{\gamma=1}^{k} \log \left( 1 + \frac{h(y_{\gamma-1})^{r}}{e^r} \right) \leq C e^{-1/2},$

where the constant $C$ cannot depend on $\epsilon$, upon which $k$ may depend.

We will consider the cases $\tilde{y}_0 = -\infty$ and $\tilde{y}_0 > -\infty$ separately. First let us assume $\tilde{y}_0 > -\infty$. We take $k = 1$ (regardless of $\epsilon$), set $y_0 = h^{-1}(1)$ and $y_1 = h^{-1}(0) = \tilde{y}_0$. We can take any $\zeta \in [0,1]$. Then with $\mathcal{F}_{k,h}$ as defined in Definition 6.2, $\mathcal{F}_{1,h} = \mathcal{G}$. Thus since $k = 1$ the sum in (S.1) is automatically finite and independent of $\epsilon$. Similarly, by (T.2) we see that $w_{1,h} < \infty$, so that (S.2) also holds. Since $\log(1+x) \leq C \beta x^\beta$ for all $x > 0$, $0 < \beta \leq 1$, we can also conclude that (S.3) holds by taking $\beta = 1/(2r)$. Thus, for $\epsilon$ small enough (10.1) holds (since $y_0$, $y_1$, and $w_{1,h}$ are fixed constants depending only on $h$), and we can apply Proposition 10.1. Since $\mathcal{F}_{1,h} = \mathcal{G}$, we have shown (7.1) and are done with the case $\tilde{y}_0 > -\infty$.

Now let us assume that $\tilde{y}_0 = -\infty$. We take some $\alpha > 1$ satisfying Assumption (T.1) and set $c = 2^{1/\alpha}$. Take any $\zeta \in [0,1)$. We set $y_0 \in h^{-1}(1)$, so $y_0 < 0$, and define $y_\gamma$ recursively, for $\gamma = 1, \ldots, \infty$, by

\begin{equation}
y_\gamma = \sup \left\{ y \mid y \leq (y_{\gamma-1} - y_0(1-c)c^{\gamma-1}) \wedge \left( y_{\gamma-1} - \frac{1}{\epsilon_0 \sup_{z \in [y_{\gamma-1}] h'(z)}} \right) \right\}, \tag{7.2}
\end{equation}

or, equivalently, define $y_\gamma$ by

\begin{equation}
y_{\gamma-1} - y_\gamma = \inf \left\{ d \mid d \geq y_0(1-c)c^{\gamma-1} \vee \frac{1}{\epsilon_0 \sup_{z \in [y_{\gamma-1}-d,y_{\gamma-1}] h'(z)}} \right\}. \tag{7.3}
\end{equation}

Note that all quantities are negative in (7.2). The idea behind this definition is simple. We want to set $y_\gamma$ to be $y_{\gamma-1} - y_0(1-c)c^{\gamma-1}$ which makes $|y_\gamma|$ an exponentially increasing sequence, since we would then have, by induction, that $y_\gamma$ equals

$$y_{\gamma-1} - y_0(1-c)c^{\gamma-1} = y_0c^{\gamma-1} - y_0(1-c)c^{\gamma-1} = y_0c^{\gamma-1}(1 - (1-c)) = y_0c^{\gamma}.$$ 

However, to apply Proposition 10.1, we need (10.1) to hold, which may not happen if $h$ is very flat for long periods and we use the simple definition just mentioned. Thus, (7.2) defines $y_\gamma$ to be the first point such that the sequence is increasing at least exponentially fast and (10.1) holds. Note that because of the supremum over $h'(z)$ in the definitions, we cannot simply set $y_\gamma$ to be the minimum of two quantities, rather we need this more complicated definition.
Next we define the truncation index, $k$, by

$$k - 1 := \min \left\{ k^* - 1 | h(y_{k^*} - 1) \leq \frac{\epsilon}{\epsilon_0} \right\}.$$  

Now (10.1) holds by definition, and by Lemma 8.4, $k$ also satisfies

$$k \leq 2 \log_c \left( \frac{1}{y_0} h^{-1} \left( \frac{\epsilon}{\epsilon_0} \right) \right).$$

Now we need to show that S.1-S.3 hold. We start with S.3. We first see that

$$\sum_{\gamma=1}^{k} \log \left( 1 + \frac{h(y_{\gamma-1})^{r \zeta}}{\epsilon^r} \right) \leq k \log \left( 1 + \frac{h(y_0)^{r \zeta}}{\epsilon^r} \right).$$

Since $h$ is nondecreasing and $h(y_0) = 1$ by the definition of $y_0$, and by Lemma 8.3 and (7.5) the above display, ignoring a factor of 2, is bounded above by

$$\log_c \left( \frac{h^{-1}(\epsilon/\epsilon_0)}{y_0} \right) \log \left( 1 + \frac{1}{\epsilon^r} \right) \leq \log_c \left( \frac{M}{-y_0} \left( \frac{\epsilon}{\epsilon_0} \right)^{-1/\alpha} \right) \log \left( 1 + \frac{1}{\epsilon^r} \right)$$

for some $M > 0$. Now for $x > 0$ and $\beta \leq 1$, there exists a $C_{\beta} > 0$ such that $\log_c(1 + x) \vee \log(1 + x) \leq C_{\beta} x^\beta$. Thus, taking $\beta \leq 1/(2(r+1))$, for $\epsilon \leq \epsilon_0$, the above display is bounded above by

$$\frac{1}{\alpha} C_{\beta} \left( \left( \frac{M}{-y_0} \right)^{\alpha} - 1 \right)^{1/(2(r+1))} C_{\beta} \left( \frac{1}{\epsilon^r} \right)^{1/(2(r+1))},$$

where we take $M$ large enough that $(-M/y_0)^\alpha \epsilon_0/\epsilon - 1 > 0$. The above display is then bounded by

$$\tilde{C} \epsilon^{-1/(2(r+1))} \epsilon^{-r/(2(r+1))} = \tilde{C} \epsilon^{-1/2},$$

where we collect all the constants into $\tilde{C} > 0$. Thus we have shown S.3.

Then S.1 and S.2 follow from arguments similar to each other. We first show S.1. Recall by (8.11) that $y_\gamma \leq y_0 c^\gamma$, so that for all $\gamma \geq 1$, $h(y_\gamma) \leq M(-y_0 c^\gamma)^{-\alpha}$ for some $M > 0$ by the fact that $h$ is nondecreasing and by Assumption (T.1). Then

$$\sum_{\gamma=1}^{k} \frac{h(y_{\gamma-1})^{r(1-\zeta)}}{\gamma^{r(1-\zeta)}} \leq \sum_{\gamma=1}^{k} (M(-y_0 c^{-\alpha})^{r(1-\zeta)},$$
which, since $c = 2^{1/\alpha}$, equals $C \sum_{\gamma=1}^{k} (2^{-r(1-\zeta)})^{\gamma-1}$ for a constant $C > 0$. Since $r \geq 1$, $0 < 2^{-r(1-\zeta)} < 1$ as long as $\zeta \in [0,1)$, and then the sum is less than $C \sum_{\gamma=1}^{\infty} (2^{-r(1-\zeta)})^{\gamma-1} < \infty$, as desired.

Next we show S.2. We have, for $\gamma = 1, \ldots, k$,

\begin{equation}
(7.7) \quad (w_{\gamma,h}(y_{\gamma-1} - y_{\gamma}))^{1/2} \leq (w_{\gamma,h}y_{0}(1-c)c^{\gamma-1})^{1/2} \vee \left( \frac{\epsilon}{\epsilon_{0}} \right)^{1/2};
\end{equation}

this is basically (7.3), which says that if $y_{\gamma-1} - y_{\gamma}$ is strictly greater than $y_{0}(1-c)c^{\gamma-1}$ then it equals $\epsilon/(\epsilon_{0}w_{\gamma,h})$. Then, by (7.7) and the bound $a \vee b \leq a + b$, we can see

\begin{equation}
(7.8) \quad \sum_{\gamma=1}^{k}(w_{\gamma,h}(y_{\gamma-1} - y_{\gamma}))^{1/2} \leq \sum_{\gamma=1}^{k}(w_{\gamma,h}y_{0}(1-c)c^{\gamma-1})^{1/2} + \sum_{\gamma=1}^{k} \left( \frac{\epsilon}{\epsilon_{0}} \right)^{1/2}.
\end{equation}

Now, for $\gamma \geq 1$,

\[ w_{\gamma,h} \leq M(-y_{\gamma-1})^{-(\alpha+1)} \leq M(-y_{0}c^{\gamma-1})^{-(\alpha+1)}, \]

by (8.11) and Assumption (T.1). Thus, the first sum on the right side of (7.8) is bounded above by

\[ \sum_{\gamma=1}^{k} \left( M(-y_{0})^{-(\alpha+1)}c^{-\alpha(\gamma-1)-(\gamma-1)}y_{0}(1-c)c^{\gamma-1} \right)^{1/2}, \]

which is equal to

\[ C \sum_{\gamma=1}^{k} (2^{1/\alpha})^{-\alpha(\gamma-1)/2} \leq C \sum_{\gamma=1}^{\infty} 2^{-(\gamma-1)/2} \]

for a constant $C > 0$, and this sum is finite and does not depend on $\epsilon$, as desired. Next, we consider the second sum on the right side of (7.8), which is bounded by $k(\epsilon/\epsilon_{0})^{1/2}$. Now, by (7.5) and Lemma 8.3, $k(\epsilon/\epsilon_{0})^{1/2}$ is bounded by

\[ 2 \log_{e} \left( \frac{1}{y_{0}}h^{-1} \left( \frac{\epsilon}{\epsilon_{0}} \right) \right) \left( \frac{\epsilon}{\epsilon_{0}} \right)^{1/2} \leq 2 \log_{e} \left( \frac{M}{-y_{0}} \left( \frac{\epsilon}{\epsilon_{0}} \right)^{-1/\alpha} \right) \left( \frac{\epsilon}{\epsilon_{0}} \right)^{1/2}, \]

and this is uniformly bounded for all $\epsilon \in (0,\epsilon_{0})$. Thus we have shown S.2.

Thus, by Proposition 10.1, we have, for $\epsilon$ small enough,

\[ \log N_{\parallel}(\epsilon C, \mathcal{F}_{k,h}, \mathcal{L}_{r}) \leq 2C \epsilon^{-1/2}, \]

for a constant $C < \infty$ that depends on $r$ and $h$ but not $\epsilon$ or $k$. We now take the brackets we have for $\mathcal{F}_{k,h}$ and extend each bracket $[l, u]$ to a new bracket $[\tilde{l}, \tilde{u}]$ for the space of interest, $\mathcal{G}$, by

\begin{equation}
(7.9) \quad \tilde{l}(x) = \begin{cases} l(x), & l(x) > y_{k} \\ 0, & l(x) \leq y_{k} \end{cases}, \quad \tilde{u}(x) = \begin{cases} u(x), & u(x) \geq y_{k} \\ \frac{x}{\epsilon_{0}}, & u(x) < y_{k}. \end{cases}
\end{equation}
This defines a collection of brackets for $G$ since for $g$ in this space, $g^{[h(y_k), \infty)} \in \mathcal{F}_{k,h}$, so $l \leq g^{[h(y_k), \infty)} \leq u$. Then by the definition of $k$, $h(y_k) \leq \epsilon/\epsilon_0$, so that $0 \leq g^{[0,h(y_k))] \leq \epsilon/\epsilon_0$. Next, using that $[l, u]$ is a bracket of size $C\epsilon$, the computation

$$\|\tilde{u} - \tilde{l}\|_r \leq \epsilon^r C^r + \int_0^1 \left( \frac{\epsilon}{\epsilon_0} \right)^r dx = \left( \frac{C^r + 1}{\epsilon_0^r} \right)^{1/r},$$

shows that size of the new collection of brackets is bounded by $\epsilon (C^r + 1/\epsilon_0^r)^{1/r}$. Thus we have shown

$$\log N([\epsilon, G, L_r]) \lesssim \left( \frac{(C^r + 1/\epsilon_0^r)^{1/r}}{\epsilon} \right)^{1/2},$$

which completes the proof of the theorem for the case $\gamma_0 = -\infty$. 

7.2. Rate results. We now use Theorem 7.1 to control the bracketing entropy for the log-concave classes in which we are really interested, i.e. those to which the Maximum Likelihood Estimator (MLE) $\hat{p}_n$ belongs with high probability, and to establish Hellinger rates of convergence.

Similarly to our previous definitions, we define

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

the class of $h$-concave-transformed densities, and we extend the definitions (4.1) and (7.10) to an arbitrary concave-function transformation $h$ as follows.

$$\mathcal{P}_{1,M,h} \equiv \left\{ p \in \mathcal{P}_h : \mbox{m}_p \in [-1, 1], \ 0 < p(0), \ \frac{1}{M} \leq p(\mbox{m}_p) \leq M \right\}. \quad (7.10)$$

As with the analogous classes of log-concave and the $s$-concave densities, the class $\mathcal{P}_{1,M,h}$ has an upper envelope, given in the following proposition.

**Proposition 7.1.** Let $h$ be a concave-function transformation such that Assumption (T.1) holds with exponent $\alpha = -1/t$ where $-1 < t \leq 0$. Then for any $p \in \mathcal{P}_{1,M,h}$ with $0 < M < \infty$,

$$p(x) \leq \begin{cases} D \left( -h^{-1}(M) + \frac{L}{2M} (x - 1) \right)^{1/t}, & x \geq 2M + 1, \\ D \left( -h^{-1}(M) + \frac{L}{2M} |x + 1| \right)^{1/t}, & x \leq -(2M + 1), \\ M, & \text{otherwise}, \end{cases} \quad (7.11)$$

$$\equiv p_{u,1,h}(x),$$

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


Proof. First, since \( p(x) \leq M \) for all \( x \) by the definition of \( \mathcal{P}_{1,M,h} \), the envelope for \( x \in [-2M+1,2M+1] \) holds automatically.

We will now take \( x \geq 2M + 1 \). The argument for \( x \leq -(2M + 1) \) is symmetric. We consider two cases for \( p \). First, the “nearly uniform” case, wherein \( p = h \circ \varphi \in \mathcal{P}_{1,M,h} \) satisfies

\[
\inf_{x \in R} p(x) \geq 1/(2M),
\]

where \( R \equiv \text{dom} \varphi \cap (1, \infty) \). (Not all densities considered in this case are necessarily nearly uniform, but this case does include some densities whose maximum is \( 1/M \) and whose minimum is larger than \( 1/(2M) \), and which we thus think of as “nearly uniform”). When (7.12) holds, since \( p \) integrates to 1 and has mode in \([-1,1]\), we can conclude that \( p(x) = 0 \) for \( x > 2M + 1 \), and the bound given by the proposition holds trivially.

Now we consider the case wherein

\[
\inf_{x \in R} p(x) \leq 1/(2M).
\]

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

\[
\varphi(m_p) - \varphi(x_1) \geq h^{-1}(M^{-1}) - h^{-1}(M^{-1}/2) \equiv L > 0.
\]

This is possible since \( \varphi(m_p) \geq h^{-1}(M^{-1}) \) by the definition of \( \mathcal{P}_{1,M,h} \) and by our choice of \( x_1 \) (and by the fact that \( \text{dom} \varphi \) is closed, so that we attain equality in (7.13)).

We will use (8.14) with \( x_1 \) as just defined and with \( x_0 = m_p \), which satisfy \( m_p < x_1 \), as needed for (8.14) to hold. Take \( x > 2M + 1 \geq x_1 \), which (by concavity of \( \varphi \)) also means \( \varphi(m_p) > \varphi(x_1) > \varphi(x) \). Also, assume \( \varphi(x) > -\infty \), since otherwise the needed envelope bound holds trivially.

Then, we can apply (8.14) to see

\[
p(x) \leq h\left(\varphi(m_p) - h(\varphi(x_1))\varphi(m_p) - \varphi(x_1)\right) F(x) - F(m_p)(x - m_p).
\]

For any \( x \in \mathbb{R} \), \( (F(x) - F(m_p))^{-1} \geq 1 \). Thus (7.15) is bounded above by

\[
h \left( h^{-1}(M) - \frac{L}{2M} (x - m_p) \right).
\]

Note that \( h^{-1}(M) - (L/(2M))(x - m_p) \leq h^{-1}(M) \). Now, \( h(y) = o(y)^{1/t} \), which implies that \( h(y) \leq D(-y)^{1/t} \) on \((-\infty, h^{-1}(M)]\) for a constant \( D \) that depends only on \( h \) and on \( M \). Thus, (7.16) is bounded above by

\[
D \left( -h^{-1}(M) + \frac{L}{2M} (x - m_p) \right)^{1/t}.
\]
We have thus shown the proposition for the case wherein (7.13) holds and when \( x \geq 2M + 1 \). The case \( x \leq -(2M + 1) \) is symmetric.

For our asymptotic results, we make the following assumption:

**Assumption 7.1.** We assume that \( X_i, i = 1, \ldots, n \) are i.i.d. random variables with distribution \( P_0 \) having density \( p_0 = h \circ \varphi_0 \in \mathcal{P}_h \) with respect to Lebesgue measure, where \( \varphi_0 \) is concave.

We can now state and prove our main theorem.

**Theorem 7.2.** Assume that \( h^{1/2} \) is a concave-function transformation and that Assumption 6.1, (T.1)–(T.4), hold for \( \sqrt{h} \), where the exponent \( \alpha \) in (T.1) satisfies \( \alpha > 5/2 \). Assume \( h \) is continuous. Suppose Assumption 7.1 holds and suppose that \( \hat{p}_n \) is the concave-transformed MLE of \( p_0 \). Then

\[
H(\hat{p}_n, p_0) = O_p(n^{-2/5}).
\]

The following corollaries connect the general Theorem 7.2 with Theorem 4.2 and Examples 6.1, 6.2, and 6.3.

**Corollary 7.1.** Suppose that \( p_0 \) in Assumption 7.1 is log-concave; that is, \( p_0 = h_0 \circ \varphi_0 \) with \( h_0(y) = e^y \) as in Example 6.1 and \( \varphi_0 \) concave. Then \( H(\hat{p}_n, p_0) = O_p(n^{-2/5}) \).

**Corollary 7.2.** Suppose that \( p_0 \) in Assumption 7.1 is \( s \)-concave with \(-1/5 < s < 0\); that is, \( p_0 = h_s \circ \varphi_0 \) with \( h_s(y) = (-y)^{1/s} \) for \( y < 0 \) as in Example 6.2 with \(-1/5 < s < 0\) and \( \varphi_0 \) concave. Then \( H(\hat{p}_n, p_0) = O_p(n^{-2/5}) \).

**Corollary 7.3.** Suppose that \( p_0 \) in Assumption 7.1 is \( h- \)concave where \( h \) is a concave transformation satisfying (T.1) and (T.4) of Assumption 6.1, and that \( h \) satisfies \( h = h_2 \circ \Psi \) where \( \Psi \) is concave and \( h_2 \) is a concave transformation for which \( \sqrt{h_2} \) satisfies Assumption 7.1, (T.1)–(T.4), where the exponent \( \alpha \) in (T.1) satisfies \( \alpha > 5/2 \). Then, if \( \hat{p}_n \) is the concave transformed MLE of \( p_0 \), \( H(\hat{p}_n, p_0) = O_p(n^{-2/5}) \). In particular the conclusion holds for \( h = h_s \) given by \( h_s(y) = y^{1/s} \) with \( s > 0 \).

Theorem 7.2 has further corollaries, for example via Examples 6.4 and 6.5. We do not yet know if the hypothesis \( s > -1/5 \) in Corollary 7.2 can be improved to (say) \( s > -1/2 \) or \( s > -1 \).
Step 1: Reduction from \( P_h \) to \( P_{1,M,h} \). We first show that we may assume, without loss of generality, that \( p_0 \in P_{1,M,h} \) for some \( M > 0 \).

To see this, consider translating and rescaling the data: we let \( \tilde{X}_i \equiv cX_i + b \) with \( c > 0 \) and \( b \in \mathbb{R} \) so that each \( \tilde{X}_i \) has density \( \tilde{p}_0(x) = p_0((x-b)/c)/c \). To choose \( b \) and \( c > 0 \) so that \( \tilde{p}_0 \in P_{1,M,h} \) we argue as follows. For definiteness, let \( m_0 \equiv \inf\{m : m \text{ is a mode of } p_0\} \). Now \( m_0 = m_{p_0} \) is in the support of \( p_0 \). (Note that \( m_0 \) in the boundary of the support of \( p_0 \) is possible.) Furthermore, there exists a point \( a_0 \neq m_0 \) in the interior of the support of \( p_0 \) (since otherwise the support of \( p_0 \) is degenerate and any such \( p_0 \) is not a density). Thus there exists a closed interval \([c_0, d_0]\) with \( m_0, a_0 \in (c_0, d_0) \).

Without loss of generality we may assume that \( m_0 < a_0 \) and we may take \( d_0 = a_0 + 2(a_0 - m_0) = 3a_0 - 2m_0 \), \( c_0 = m_0 - (a_0 - m_0) = 2m_0 - a_0 \). Thus \( a_0 = (c_0 + d_0)/2 \) and \( m_0 = (3c_0 + d_0)/4 \). Then it is easily seen that if \( c = 1/(d_0 - a_0) \) and \( b = -ca_0 \) we have

\[
\tilde{p}_0(0) = p_0(a_0)/c > 0, \quad \tilde{p}_0(-1/2) = p_0(m_0)/c, \quad \text{and} \quad \tilde{p}_0(1) = p_0(c_0)/c > 0, \quad \tilde{p}_0(-1) = p_0(0)/c,
\]

so that \( \tilde{m}_0 \equiv m_{\tilde{p}_0} = -1/2 \) is the (smallest) mode of \( \tilde{p}_0 \).

For the rescaled data the MLE satisfies \( \hat{p}_n((x-b)/c; \tilde{X})/c = \hat{p}_n(x; \tilde{X}) \), and since the Hellinger metric is invariant under affine transformations, it follows that

\[
H(\hat{p}_n(\cdot; \tilde{X}), p_0) = H(\hat{p}_n(\cdot; \tilde{X}), \tilde{p}_0).
\]

Hence if (7.18) holds for \( \tilde{p}_0 \) and the transformed data, it also holds for \( p_0 \) and the original data. Thus we can henceforth assume that \( p_0 \in P_{1,M,h} \) for any \( M > p_0(m_{p_0}) \).

Now by the consistency results in Theorems 3.1 and 3.2 (and the general version of the latter in Theorem 2.17 of Seregin and Wellner [2010]) which holds under their assumptions (D.1)–(D.4) and which are in turn implied by our (T.1)–(T.4)) for \( g \equiv h^{1/2} \), it follows that \( H(\hat{p}_n, p_0) \rightarrow a.s. \ 0 \), and we also have uniform convergence of \( \hat{p}_n \) to \( p_0 \) on compact subsets strictly contained in the support of \( p_0 \). When the mode \( m_{p_0} \) is not in the interior of the support we indeed do not have consistency at the mode, but we do have almost sure uniform boundedness of the MLEs by Theorem 3.2, (iv). The scaling given above gives 0 as an interior point of the support of \( p_0 \), and hence consistency at 0 holds. Then uniform boundedness lets us conclude that all possible MLEs \( p \) satisfy \( p(m_p) < M \), and by consistency at \( x = 0 \) we can conclude that \( p(m_p) \geq p(0) > 1/M \). Since \( p_0 \in P_{1,M,h} \) for any \( M > p_0(m_{p_0}) \), it follows that

\[
P_0(\hat{p}_n \in P_{1,M,h}) \rightarrow 1 \quad \text{as} \quad n \rightarrow \infty
\]
for all $M > M(p_0) \geq p_0(m_{p_0})$ where $M(p_0)$ is as defined in Theorem 3.1 part (v) when $s = 0$ and Theorem 3.2 part (iv). This completes step 1.

**Step 2. Control of Hellinger bracketing entropy for $\mathcal{P}_{1,M,h}$ suffices.**

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

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

Suppose that we can show that

$$\log N_{[1]}(\epsilon, \overline{P}_h(\delta), H) \preceq \epsilon^{-1/2}$$  

for all $0 < \delta \leq \delta_0$ for some $\delta_0 > 0$. Then it follows from van der Vaart and Wellner [1996], Theorems 3.4.1 and 3.4.4 (with $p_n = p_0$ in Theorem 3.4.4) or, alternatively, from van de Geer [2000], Theorem 7.4 and an inspection of her proofs, that any $r_n$ satisfying

$$r_n^2 \Psi(1/r_n) \leq \sqrt{n}$$  

where

$$\Psi(\delta) \equiv J_{[1]}(\delta, \overline{P}_h(\delta), H) \left(1 + \frac{J_{[1]}(\delta, \overline{P}_h(\delta), H)}{\delta^2 \sqrt{n}}\right)$$  

and

$$J_{[1]}(\delta, \overline{P}_h(\delta), H) \equiv \int_0^\delta \sqrt{\log N_{[1]}(\epsilon, \overline{P}_h(\delta), H)} \, d\epsilon$$

gives a rate of convergence for $H(\hat{p}_n, p_0)$. But it is easily seen that if (7.20) holds, then $r_n = n^{2/5}$ satisfies (7.21), and hence (7.18) holds.

**Step 2b.** Thus we want to show that (7.20) holds if we have an appropriate bracketing entropy bound for $\mathcal{P}_{1,M,h}$. First note that

$$N_{[1]}(\epsilon, \overline{P}_h(\delta), H) \leq N_{[1]}(\epsilon, \mathcal{P}_h(4\delta), H)$$

in view of van der Vaart and Wellner [1996], exercise 3.4.4 (or van de Geer [2000]), Lemma 4.2, page 48). Furthermore,

$$N_{[1]}(\epsilon, \mathcal{P}_h(4\delta), H) \leq N_{[1]}(\epsilon, \mathcal{P}_{1,M,h}, H)$$

since $\mathcal{P}_h(4\delta) \subset \mathcal{P}_{1,M,h}$ for all $0 < \delta \leq \delta_0$ 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 Rockafellar [1970], Theorem 10.8. See Lemma 8.1 for details of the proofs.

Finally, note that
\[
N_1(\epsilon, P_{1,M,h}, H) = N_1(\epsilon, P_{1,M,h}^{1/2}, L_2(\lambda/2))
= N_1(\epsilon, P_{1,M,h}^{1/2}, L(\lambda)/\sqrt{2}) = N_1(\epsilon/\sqrt{2}, P_{1,M,h}^{1/2}, L_2(\lambda))
\]
by the definition of \( H \) and \( L_2(\lambda) \). Thus it suffices to show that
\[
\log N_1(\epsilon, P_{1,M,h}^{1/2}, L_2(\lambda)) \lesssim \frac{1}{\epsilon^{1/2}}
\]
where the constant involved depends only on \( M \) and \( h \). This completes the proof of Step 2.

**Step 3: Proof of the bracketing bound (7.22).**

In fact we will show that
\[
\log N_1(\epsilon, P_{1,M,h}^{1/2}, L_r(Q)) \lesssim \frac{1}{\epsilon^{1/2}}
\]
for any measure \( Q \) with Lebesgue-density \( q \) where \( q \) is bounded above by 1 and where the constants in \( \lesssim \) depend only on \( r, M \) and \( h \). (Note that \( q \equiv 1 \) works.) Now (7.23) holds if it holds when we replace \( P_{1,M,h}^{1/2} \) by the two classes \( P_{1}^{1/2}(\cdot,0,\lambda) \) and \( P_{1}^{1/2}([0,\infty]) \). This is because if we have \( \exp(c_1/\epsilon^{1/2}) \) brackets \( [l_{n_1}, u_{n_1}] \) for \( P_{1}^{1/2}(\cdot,0,\lambda) \) and \( \exp(c_2/\epsilon^{1/2}) \) brackets \( [l_{n_2}, u_{n_2}] \) for \( P_{1}^{1/2}([0,\infty]) \), where \( c_i > 0 \) are constants depending only on \( p_0 \) and on \( M \), and where we let \( n_i \) range from 1 to the appropriate index, for \( i = 1, 2 \), then we can define functions
\[
u_{n_1,n_2}(x) = u_{n_1}(x)1_{(-\infty,0]}(x) + u_{n_2}(x)1_{[0,\infty)}(x)
\]
and similarly define \( l_{n_1,n_2} \) via \( l_{n_1} \) and \( l_{n_2} \). These form brackets for \( P_{1}^{1/2} \); there are no more than \( \exp((c_1 + c_2)/\epsilon^{1/2}) \) of them; and their size is no larger than \( 2^{1/r}\epsilon \).

These two intervals are symmetric to each other, so we will only consider the restriction to \([0,\infty)\). We will use the method of Theorem 2.7.4 of van der Vaart and Wellner [1996] with Theorem 7.1. We first partition \([0,\infty)\) into intervals \( I_j = [j,j+1] \) of length 1, \( j \in \mathbb{N} \) and consider \( P_{1}^{1/2} \), the restriction of \( P_{1,M,h}^{1/2} \) to those intervals.
Thus by the envelope $p_{u,1,\sqrt{h}}$ for $\mathcal{P}_{1,M,\sqrt{h}}$, defined in (7.11), $\mathcal{F}_j^{1/2} \subset h^{1/2} \circ \mathcal{C}(I_j, [0, B_j], \infty)$ where $B_j := K(j - 1)^{-\alpha}$ (with $\alpha = -1/t$) for $j \geq 2$ and $B_0, B_1$, and $K$ are given by a constant depending on $M$ and $h$. By definition, all $p \in \mathcal{P}_{1,M,h}$ satisfy $p(0) > 0$, which means $p|_{I_j} \in h \circ \mathcal{D}(j)$, so we can apply Theorem 7.1.

Now, we let $a_j$ be any sequence of numbers in $(0, \infty]$. We fix $\epsilon > 0$ and for each $j \in \mathbb{N}$ take an $\epsilon a_j$-bracket $[l_{j,1}, u_{j,1}], \ldots, [l_{j,p_j}, u_{j,p_j}]$ for $\sqrt{h} \circ \mathcal{C}(I_j, [0, (\sqrt{h})^{-1}(B_j)], \infty)$ for the $L_r(\lambda)$ norm on $I_j$. By Theorem 7.1, the $p_j$’s satisfy

\begin{equation}
\log p_j \lesssim \left( \frac{B_j \cdot 1}{\epsilon a_j} \right)^{1/2}.
\end{equation}

If $\epsilon a_j \geq B_j$ then we can take $p_j = 1$. We form brackets

\begin{equation}
\left[ \sum_j l_{j,i_j} 1_{I_j}, \sum_j u_{j,i_j} 1_{I_j} \right]
\end{equation}

where $i_j$ range over all possible values $1, \ldots, p_j$. The number of brackets is bounded by $\prod_j p_j$. Now, the $L_r(Q)$ size of a bracket $[l, u]$ defined above is $\|u - l\|_{r,Q}$ where

\begin{align*}
\|u - l\|_{r,Q} &= \int_{x_0}^\infty |u - l|^r q d\lambda \\
&\leq \sum_j \int_{I_j} \|q\|_{\infty,j} |u(x) - l(x)|^r dx \\
&= \sum_j \|q\|_{\infty,j} \|u - l\|_{r,\lambda} \leq \sum_j \|q\|_{\infty,j} a_j^r \epsilon^r \\
&\leq \epsilon^r \sum_j \|q\|_{\infty,j} a_j^r.
\end{align*}

Thus,

\begin{equation}
\log N[\epsilon \left( \sum_j \|q\|_{\infty,j} a_j^r \right)^{1/r}, \mathcal{P}_{1,M,h}^{1/2}([0, \infty), L_r(Q)) \lesssim \sum_j \left( \frac{B_j}{\epsilon a_j} \right)^{1/2}.
\end{equation}

The choice

\begin{equation}
a_j^{r+1/2} = \frac{B_j^{1/2}}{\|q\|_{\infty,j}} =: \frac{B_j^{1/2}}{q_j},
\end{equation}
or \( a_j = B_j^{1/(1+2r)}q_j^{2/(1+2r)} \), reduces both sums in (7.26) to \( \sum B_j^{r/(1+2r)}q_j^{1/(1+2r)} \), by the computations

\[
\begin{align*}
a_j^r q_j &= B_j^{r/(1+2r)}q_j^{1-2r/(1+2r)} = B_j^{r/(1+2r)}q_j^{1/(1+2r)} \\
\left( \frac{B_j}{a_j} \right)^{1/2} &= \left( \frac{B_j}{B_j^{1/(1+2r)}q_j^{2/(1+2r)}} \right)^{1/2} = \left( B_j^{2r/(1+2r)}q_j^{2/(1+2r)} \right)^{1/2} = B_j^{r/(1+2r)}q_j^{1/(1+2r)}.
\end{align*}
\]

Thus, for the moment denoting \( \sum B_j^{r/(1+2r)}q_j^{1/(1+2r)} \) by the symbol \( S \), (7.26) says

\[
(7.28) \quad \log N[\epsilon S^{1/r}, P_{1,M,h}\|_{[0,\infty)}, L_r(Q)] \lesssim \frac{1}{\epsilon^{1/2}} S,
\]

so letting \( v = \epsilon S^{1/r} \), we have

\[
(7.29) \quad \log N[\epsilon, P_{1/2,M,h}\|_{[0,\infty)}, L_r(Q)] \lesssim \frac{1}{\epsilon^{1/2}} S^{(1+2r)/(2r)}.
\]

So we just need to show that \( S = \sum B_j^{r/(1+2r)}q_j^{1/(1+2r)} < \infty \). Since we assumed \( q_j \leq 1 \) and \( B_j = K(j - 1)^{-\alpha} \), the inequality \( S < \infty \) holds where \( S \) depends only on \( r \), \( M \), and \( h \), as long as \( \alpha > (1 + 2r)/r \). This completes the proof of (7.23). Taking \( r = 2 \) and \( Q = \lambda \) yields (7.22) for \( \alpha > 5/2 \) and completes the proof of the theorem. \( \square \)

**Proof.** Corollary 7.3: The proof is based on the proof of Theorem 7.2. In Step 1 of that proof, the only requirement on \( h \) is that we can conclude that \( \hat{p}_n \) is almost surely Hellinger consistent. Almost sure Hellinger consistency is given by Theorem 2.18 of Seregin and Wellner [2010] 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 (7.20) for \( P_h(\delta) \). This argument does not reply on specific properties of \( h \). Thus we now only need to show that (7.20) holds for our current \( h \).

Step 3 of the proof shows that the bound (7.20) holds for transforms \( h \) which satisfy the assumptions of Theorem 7.2, which are the same assumptions as in the Corollary for \( h_2 \). By Lemma 7.1 below we have

\[
\log N[\epsilon, P_h(\delta), H] \leq \log N[\epsilon, P_{h_2}(\delta), H],
\]

and we are then done, since Step 3 in the original proof bounded the right side of the above display already. \( \square \)
Lemma 7.1. Let $h_1$ and $h_2$ be concave-function transformations. If $\Psi$ 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 Seregin and Wellner [2010] gives this result, in the notation of “decreasing (convex) transformations.” Thus, let $\tilde{h}_i(y) = h_i(-y)$, $i = 1, 2$, and recall that $\tilde{h}_i(y)$ is a decreasing (convex) transformation. Then,

$$\tilde{h}_1(y) = h_1(-y) = h_2(\Psi(-y)) = \tilde{h}_2(-\Psi(-y)),$$

where the middle equality is by assumption. Now $-\Psi(-y)$ is convex if and only if $-\Psi(y)$ is convex if and only if $\Psi(y)$ is concave. Thus we conclude by Lemma 2.5, page 6, of Seregin and Wellner [2010],

$$\mathcal{P}_{h_1} \subseteq \mathcal{P}_{h_2},$$

since our classes $\mathcal{P}_{h_i}$ are the classes “$\mathcal{P}(\tilde{h}_i)$” in Seregin and Wellner [2010].

8. Appendix, part 1: Technical Lemmas and Inequalities.

8.1. Nonexistence of the MLE for $\mathcal{P}_{1,s}$ with $s < -1$. Consider $s$–concave densities $p$ with $s < -1$; i.e.

$$p(x) = \varphi(x)^{1/s}$$

with $s < -1$ and $\varphi$ convex. We denote the class of all such densities by $\mathcal{P}_{1,s}$. It follows from Seregin and Wellner [2010] that the MLE exists for $\mathcal{P}_{1,s}$ when $-1 < s < 0$ and existence of the MLE for $s = 0$ (the log-concave case) follows from Pal, Woodroofe and Meyer [2007] and Dümbgen and Rufibach [2009] as discussed in more detail in Section 3. It is well-known that the MLE does not exist for the class of unimodal densities $\mathcal{P}_{1,-\infty}$; see for example Birgé [1997]. The following proposition shows that the MLE does not exist for any of the classes $\mathcal{P}_{1,s}$ with $s < -1$.

Proposition 8.1. A maximum likelihood estimator does not exist for the class $\mathcal{P}_{1,s}$ for any $s < -1$.

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

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

with $s < -1$ and $\varphi$ convex. We denote the class of all such densities by $\mathcal{P}_{1,s}$. It follows from Seregin and Wellner [2010] that the MLE exists for $\mathcal{P}_{1,s}$ when $-1 < s < 0$ and existence of the MLE for $s = 0$ (the log-concave case) follows from Pal, Woodroofe and Meyer [2007] and Dümbgen and Rufibach [2009] as discussed in more detail in Section 3. It is well-known that the MLE does not exist for the class of unimodal densities $\mathcal{P}_{1,-\infty}$; see for example Birgé [1997]. The following proposition shows that the MLE does not exist for any of the classes $\mathcal{P}_{1,s}$ with $s < -1$.

Proposition 8.1. A maximum likelihood estimator does not exist for the class $\mathcal{P}_{1,s}$ for any $s < -1$.

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

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

with $s < -1$ and $\varphi$ convex. We denote the class of all such densities by $\mathcal{P}_{1,s}$. It follows from Seregin and Wellner [2010] that the MLE exists for $\mathcal{P}_{1,s}$ when $-1 < s < 0$ and existence of the MLE for $s = 0$ (the log-concave case) follows from Pal, Woodroofe and Meyer [2007] and Dümbgen and Rufibach [2009] as discussed in more detail in Section 3. It is well-known that the MLE does not exist for the class of unimodal densities $\mathcal{P}_{1,-\infty}$; see for example Birgé [1997]. The following proposition shows that the MLE does not exist for any of the classes $\mathcal{P}_{1,s}$ with $s < -1$.

Proposition 8.1. A maximum likelihood estimator does not exist for the class $\mathcal{P}_{1,s}$ for any $s < -1$.

Proof. 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 \( \varphi_a \) is convex and

\[
p_a(x) \equiv \varphi_a(x)^{1/s} = \varphi_a(x)^{-r} = \frac{a}{(b_r - ax)^r} \mathbf{1}_{[0,b_r/a]}(x)
\]
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 \( i \leq n \) and hence for \( a < b_r/X(n) \) where \( X(n) \equiv \max_{1 \leq i \leq n} X_i \). Note that \( \ell_n(a) \to \infty \) as \( a \to b_r/X(n) \). Hence the MLE does not exist for \( \{ p_a : a > 0 \} \), and a fortiori the MLE does not exist for \( \{ p : p \in \mathcal{P}_{1,s} \} \) with \( s < -1 \).

8.2. Some Technical Lemmas. For \( \delta > 0 \) and \( \mathcal{P} \) consisting of all log-concave densities on \( \mathbb{R} \), let

\[
\mathcal{P}(\delta) \equiv \{ p \in \mathcal{P} : H(p,p_0) < \delta \},
\]
\[
\overline{\mathcal{P}}(\delta) \equiv \{ (p + p_0)/2 : p \in \mathcal{P}, H((p + p_0)/2, p_0) < \delta \},
\]
and let \( \mathcal{P}_{1,M} \) be as defined in (4.1).

**Lemma 8.1.** Let \( \delta > 0 \) and \( 0 < \epsilon \leq \delta \). With the definitions in the previous display

\[
N[1](\epsilon, \overline{\mathcal{P}}^{1/2}(\delta), L_2(\lambda)) \lesssim N[1](\epsilon, \mathcal{P}^{1/2}(4\delta), L_2(\lambda)) \lesssim N[1](\epsilon, \mathcal{P}_{1,M}^{1/2}, L_2(\lambda)).
\]

By the definition of Hellinger distance (8.1) is identical to

\[
N[1](\epsilon, \mathcal{P}(\delta), H) \lesssim N[1](\epsilon, \mathcal{P}(4\delta), H) \lesssim N[1](\epsilon, \mathcal{P}_{1,M}, H).
\]

Note that the bracketing number in (8.1) is the key quantity in the integrand of (7.8) of van de Geer [2000].

**Proof.** To see the inequalities in (8.1), we first show (8.2). To do this we follow the notation in van de Geer [2000] (see e.g. chapter 4) and set \( \overline{p} = (p + p_0)/2 \) for any function \( p \). Then if \( \overline{p}_1 \in \overline{\mathcal{P}}(\delta) \), by (4.6) on page 48 of van de Geer [2000], we have \( H(p_1, p_0) < 4H(\overline{p}_1, p_0) < 4\delta \), so that \( p_1 \in \mathcal{P}(4\delta) \). Then given \( \epsilon \)-brackets \([l_\alpha, u_\alpha]\), of \( \mathcal{P}(4\delta) \), with \( 1 \leq \alpha \leq N[1](\epsilon, \mathcal{F}(4\delta), H) \), we
can construct brackets of $\mathcal{P}(\delta)$ since for any $p_1 \in \mathcal{P}(4\delta)$ which is bracketed by $[l_\alpha, u_\alpha]$ for some $\alpha$, $p_1$ is bracketed by $[l_\alpha, u_\alpha]$, so that $[l_\alpha, u_\alpha]$ form a collection of brackets for $\mathcal{P}(\delta)$ with size bounded by $H(l_\alpha, u_\alpha) \leq \frac{1}{\sqrt{2}} H(l_\alpha, u_\alpha) < \frac{1}{\sqrt{2}} \epsilon$, where we used (4.5) on page 48 of van de Geer [2000]. Thus we have a collection of brackets of Hellinger size $\epsilon/\sqrt{2} < \epsilon$ with cardinality bounded by $N_\epsilon(\epsilon, \mathcal{P}(4\delta), H)$ and (8.2) holds.

Next we show (8.3), which will follow from showing $\mathcal{P}(4\delta) \subset \mathcal{P}_{1,M}$. Now if $0 < M^{-1} < \inf_{x \in [-1,1]} p_0(x)$ then for any $p$ that has its mode in $[-1,1]$ and satisfies

$$\sup_{x \in [-1,1]} |p(x) - p_0(x)| \leq \min \left( \inf_{x \in [-1,1]} p_0(x) - M^{-1}, M - \sup_{x \in [-1,1]} p_0(x) \right),$$

we can conclude that $p \in \mathcal{P}_{1,M}$.

The proof of Lemma 3.14 of Seregin and Wellner [2010] shows that for any sequence of log-concave densities $p_i$,

$$H(p_i, p_0) \to 0 \quad \text{implies} \quad \sup_{x \in [-1,1]} |p_i(x) - p_0(x)| \to 0.$$  

This says that the topology defined by the Hellinger metric has more open sets than that defined by the supremum distance on $[-1,1]$, which implies that open supremum balls are nested within open Hellinger balls, i.e. for $\epsilon > 0$

$$(8.6) \quad B_\epsilon(p_0, \sup_{[-1,1]}) \subseteq B_{4\delta}(p_0, H)$$

for some $\delta > 0$, where $B_\epsilon(p_0, d)$ denotes an open ball about $p_0$ of size $\epsilon$ in the metric $d$.

Alternatively, we can argue directly by contradiction that (8.6) holds given (8.5). The latter says that for a fixed $\epsilon > 0$ that any sequence of functions $p_i$ that converges to $p_0$ in Hellinger distance has a $\delta > 0$, which potentially depends on the sequence $p_i$, such that once $H(p_i, p_0) < \delta$, $p_i$ and $p_0$ are uniformly within $\epsilon$ on $[-1,1]$. We need to show that the $\delta$ does not depend on the sequence $p_i$. We thus consider a set of sequences, $p_{i,j}$, $i, j = 1, 2, \ldots$, such that as $j \to \infty$, $H(p_{i,j}, p_0) \to 0$. For fixed $\epsilon$ each sequence has a minimum $\delta_i > 0$ such that for fixed $i$ and all $j$ large enough such that $H(p_{i,j}, p_0) < \delta_i$, $p_{i,j}$ is within $\epsilon$ of $p_0$ on $[-1,1]$. Assume for contradiction
that $\delta_i \to 0$. Then we consider the diagonal sequence $p_{i,i}$ as $i \to \infty$, and it follows that $p(p_{i,i}, p_0) \to 0$ but for no $\delta > 0 = \lim_{i \to \infty} \delta_i$ is $p_{i,i}$ uniformly within $\epsilon$ of $p_0$ on $[-1, 1]$, which contradicts (8.5), and so we have shown (8.6).

Now, if $p$ is uniformly within $\epsilon$ of $p_0$ on $[-1, 1]$, then for $\epsilon$ small enough we know that the mode of $p$ is in $[-1, 1]$. Thus for $0 < M^{-1} < \inf_{x \in [-1,1]} p_0(x)$ and $\delta$ small enough, any $p \in \mathcal{P}(4\delta)$ is also in $\mathcal{P}_{1,M}$ as desired, and so (8.3) has been shown.

**Lemma 8.2.** $d_{r,k,h}(\cdot, \cdot)$ is a metric; i.e. the triangle inequality holds.

**Proof.** For the purposes of this proof, for $a < b$, $a, b \in [-\infty, \infty]$, let $[b, a)$ be defined to be $[a, b)$. Now, letting $a, b, c$ be any values in $[-\infty, \infty]$ we know that the triangle inequality holds for set containment of intervals, i.e. $[a, c) \subseteq [a, b) \cup [b, c)$. Thus

$$W([a, c)) \leq W([a, b) \cup [b, c)) \leq W([a, b)) + W([b, c)),$$

where $W$ is the measure on which $d_{k,h}$ depends, the first inequality above follows by the triangle inequality for set containment and the second follows because $W$ is a positive measure. Thus, the above display says

$$d_{k,h}(a, c) \leq d_{k,h}(a, b) + d_{k,h}(b, c).$$

So we have the triangle inequality for $d_{k,h}$. Then the triangle inequality for $d_{r,k,h}$ follows from Minkowski’s inequality, i.e. for any functions $\varphi_1, \varphi_2,$ and $\varphi_3$ defined on $[0, b]$,

$$d_{r,k,h}(\varphi_1, \varphi_3) = \left( \int_0^b d_{k,h}(\varphi_1(x), \varphi_3(x))^r \, dx \right)^{1/r} \leq \left( \int_0^b (d_{k,h}(\varphi_1(x), \varphi_2(x)) + d_{k,h}(\varphi_2(x), \varphi_3(x)))^r \, dx \right)^{1/r} \leq \left( \int_0^b d_{k,h}(\varphi_1(x), \varphi_2(x))^r \, dx \right)^{1/r} + \left( \int_0^b d_{k,h}(\varphi_2(x), \varphi_3(x))^r \, dx \right)^{1/r} = d_{r,k,h}(\varphi_1, \varphi_2) + d_{r,k,h}(\varphi_2, \varphi_3),$$

where the first inequality is by (8.7) and the next is Minkowski’s triangle inequality for $L_r$ distance.

**Lemma 8.3.** For a concave-function transformation $h$ that satisfies Assumption T.1, we can have that $h^{-1}$ is nondecreasing and as $f \searrow 0$,

$$h^{-1}(f) = o(f^{-1/\alpha}).$$

In particular, for $f \in (0, L]$, $h^{-1}(f) \leq M_L f^{-1/\alpha}$. 

---

imsart-aos ver. 2011/12/06 file: LCgr-v23-all.tex date: May 7, 2014
Proof. For two increasing functions $h \leq g$ defined on $(-\infty, \infty)$ taking values in $[-\infty, \infty]$, where ran $h$ and ran $g$ are both intervals, we will show that $g^{-1}(f) \leq h^{-1}(f)$ for any $f \in \text{ran} h \cap \text{ran} g$. By definition, for such $f$, we can find a $z \in (-\infty, \infty)$ such that $f = g(z)$. That is, $g(z) = h(h^{-1}(f)) \leq g(h^{-1}(f))$ since $h \leq g$. Applying $g^{-1}$, we see $z = g^{-1}(f) \leq h^{-1}(f)$, as desired.

Then (8.8) follows by letting $g(y) = \delta(-y)^{-\alpha}$, which has $g^{-1}(f) = -(1/y)^{-\alpha}$. The statement that $h^{-1}(f) \leq M_L f^{-1/\alpha}$ follows since on neighborhoods away from 0, $h^{-1}$ is bounded above and $f \mapsto f^{-1/\alpha}$ is bounded below.

To see that $h^{-1}$ is nondecreasing, we differentiate to see $(h^{-1})'(f) = 1/h'(h^{-1}(f))$. Since $h' \geq 0$ so is $(h^{-1})'$.

Lemma 8.4. Let $h$ be as in the statement of Theorem 7.1, taking $y_\gamma$ defined for $\gamma = 1, \ldots, k$ by (7.2) and $k$ by (7.4). Then for $\epsilon$ small enough,

$$k \leq 2 \log_e \left( \frac{1}{y_0} h^{-1}\left( \frac{\epsilon}{\epsilon_0} \right) \right).$$

Proof. Set

$$k^* - 1 := \log_e \left( \frac{1}{y_0} h^{-1}\left( \frac{\epsilon}{\epsilon_0} \right) \right).$$

The value on the right side of (8.10) is what $k - 1$ would be if $y_\gamma$ were always equal to $y_{\gamma-1} - y_0(1-c)\gamma^{-1}$. That is, if we define $z_\gamma$ by $z_0 = y_0$ and $z_\gamma - z_\gamma = y_0(1-c)\gamma^{-1}$ for all $\gamma \geq 1$, then $z_\gamma = y_0 \epsilon_0^{-1}$, so that $z_{k^*-1} = y_0 \epsilon_0^{k^*-1} = y_0 \frac{1}{y_0} h^{-1}(\epsilon/\epsilon_0)$, i.e. $h(z_{k^*-1}) = \epsilon/\epsilon_0$. (Induction shows that $z_\gamma = y_0 \epsilon_0^{\gamma}$; the base case, $\gamma = 0$, is trivial, and assuming $z_{\gamma-1} = y_0 \epsilon_0^{\gamma-1}$, we have $z_\gamma = y_{\gamma-1} - y_0(1-c)\gamma^{-1}$ which equals $y_0 \epsilon_0^{\gamma-1} - y_0(1-c)\gamma^{-1} = y_0 \epsilon_0^{\gamma-1}(1-(1-c))$ as desired.) Now, by (7.2) for all $\gamma \geq 1$,

$$y_0 \epsilon_0^{\gamma} = z_\gamma \geq y_\gamma$$

(i.e. $|z_\gamma| \leq |y_\gamma|$) so since $h$ is nondecreasing, $h(y_{k^*}) \leq h(z_{k^*}) = \epsilon/\epsilon_0$. Thus, by the definition of $k$, (7.4), $k \leq k^*$, i.e. $k - 1 \leq k^* - 1$. To remove the summand of 1 in (8.9), we multiply by the factor 2 and on the right and take $\epsilon$ small enough.

8.3. Univariate $s$-concave and general $h-$concave density bounds. Recall that a non-negative function $f$ defined on an open, convex subset $C \subset \mathbb{R}^d$ and taking values in $\mathbb{R}^+$ is $s-$concave on $C$ if

$$f((1-\theta)x_0 + \theta x_1) \geq M_s(f(x_0), f(x_1); \theta) = \{(1-\theta)f(x_0)^s + \theta f(x_1)^s\}^{1/s}$$
for all $x_0, x_1 \in C$ and $0 \leq \theta \leq 1$. If this holds, write $f \in \mathcal{F}_s(C)$. If this holds with $C = \mathbb{R}^d$, write $f \in \mathcal{F}_{s,d}$. Then:

(a) If $s < 0$, then $f \in \mathcal{F}_{s,d}$ if and only if $f(x) = \varphi(x)^{1/s}$ with $\varphi$ convex into $[0, \infty)$.
(b) If $s = 0$, then $f \in \mathcal{F}_{0,d}$ if and only if $f(x) = \exp(\varphi(x))$ with $\varphi$ concave into $[-\infty, \infty)$.
(c) If $s > 0$, then $f \in \mathcal{F}_{s,d}$ if and only if $p(x) = \varphi(x)^{1/s}$ where either $\varphi \equiv 0$ on $C$ or $\varphi$ is concave on $C$ into $(0, \infty)$.

See e.g. Dharmadhikari and Joag-Dev [1988], page 86.

In the following we take $d = 1$ and let $F(x) \equiv \int_{-\infty}^{x} f(y) dy$ be the (generalized) distribution function corresponding to the non-negative function $f$.

**Proposition 8.2.** Suppose that $f \in \mathcal{F}_{s,1}$ with $s \leq 1$. Then for arbitrary points $x_1 < x_2$

$$
(8.12) \quad \sqrt{f(x_1)f(x_2)} \leq \frac{F(x_2) - F(x_1)}{x_2 - x_1}.
$$

Furthermore, for $x_0 \in \{f > 0\}$ and any real $x \neq x_0$, we have:

(i) If $-1 \leq s < 0$, then

$$
f(x) \leq \begin{cases}
    f(x_0) \cdot \left\{1 + s + (-s) \left(\frac{f(x_0)|x-x_0|}{h(x_0,x)}\right)\right\}^{1/s}, & \text{if } f(x_0)|x-x_0| \geq h(x_0, x), \\
    f(x_0) \left(\frac{h(x_0,x)}{f(x_0)|x-x_0|}\right)^2, & \text{otherwise}
\end{cases}
$$

where

$$
(8.13) \quad h(x_0, x) \equiv F(x \vee x_0) - F(x \wedge x_0) \leq \begin{cases}
    F(x_0), & \text{if } x < x_0, \\
    1 - F(x_0), & \text{if } x > x_0.
\end{cases}
$$

(ii) If $s = 0$, then

$$
f(x) \leq \begin{cases}
    f(x_0) \exp\left(1 - \frac{f(x_0)|x-x_0|}{h(x_0,x)}\right), & \text{if } f(x_0)|x-x_0| \geq h(x_0, x) \\
    f(x_0) \left(\frac{h(x_0,x)}{f(x_0)|x-x_0|}\right)^2, & \text{otherwise}
\end{cases}
$$

and where $h(x_0, x)$ is as in (8.13).

(iii) If $0 < s \leq 1$, then, with $w_+ \equiv \max\{0, w\}$ for $w \in \mathbb{R}$,

$$
f(x) \leq \begin{cases}
    f(x_0) \left\{1 + s \left(1 - \frac{f(x_0)|x-x_0|}{h(x_0,x)}\right)\right\}^{1/s} + \left(\frac{h(x_0,x)}{f(x_0)|x-x_0|}\right)^2, & \text{if } f(x_0)|x-x_0| \geq h(x_0, x), \\
    f(x_0) \left(\frac{h(x_0,x)}{f(x_0)|x-x_0|}\right)^2, & \text{otherwise}
\end{cases}
$$
**Proposition 8.3.** Let $h$ be a concave-function transformation and $f = h \circ \varphi$ for $\varphi \in C$. Then for $x_0 < x_1 < x$ or $x < x_1 < x_0$, all such that $-\infty < \varphi(x) < \varphi(x_1) < \varphi(x_0) < \infty$, we have

\begin{equation}
\begin{aligned}
f(x) &\leq h \left( \varphi(x_0) - h(\varphi(x_1)) \frac{\varphi(x_0) - \varphi(x_1)}{F(x) - F(x_0)} (x - x_0) \right).
\end{aligned}
\end{equation}

**Remark 1.** Part (ii) of the statement above is Lemma A.1 of Dümbgen and Rufibach [2009]. For an early version of the bounds in (ii) see Devroye [1984].

**Remark 2.** Note that the bound in (i) converges to the bound in (iii) as $s \nearrow 0$, and that the bound in (iii) converges to the bound in (ii) as $s \searrow 0$.

**Proof.** (Proposition 8.2). We first prove that (8.12) holds. This was shown by Dümbgen and Rufibach [2009] when $s = 0$; see their Lemma A.1. Now suppose that $f \in \mathcal{F}_{s,1}$ with $s < 0$ or $0 < s \leq 1$ and let $x_1, x_2 \in \{f > 0\}$. 


Then

\[ F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(y) dy = \int_{x_1}^{x_2} \varphi(y)^{1/s} dy \]

\[ = \int_{x_1}^{x_2} \varphi\left(\frac{x_2 - y}{x_2 - x_1} + \frac{y - x_1}{x_2 - x_1} x_2\right)^{1/s} dy \]

\[ \geq \int_{x_1}^{x_2} \left\{ \frac{x_2 - y}{x_2 - x_1} \varphi(x_1) + \frac{y - x_1}{x_2 - x_1} \varphi(x_2) \right\}^{1/s} dy \]

when \( s < 0 \) since \( \varphi \) is convex and

\[ g(v) = v^{1/s} \] is decreasing

when \( 0 < s \leq 1 \) since \( \varphi \) is concave and

\[ g(v) = v^{1/s} \] is increasing

\[ (8.15) = (x_2 - x_1) \int_0^1 \left\{ (1 - u)\varphi(x_1) + u\varphi(x_2) \right\}^{1/s} du \]

\[ \geq \int_0^1 (1 - u)\varphi(x_1) + u\varphi(x_2) du \]

when \( s \neq 0, \ s \leq 1 \) by Jensen’s inequality since

\[ g(v) = v^{1/s} \] is convex

\[ = (x_2 - x_1) \left\{ \frac{1}{2} (\varphi(x_2) + \varphi(x_2)) \right\}^{1/s} \]

\[ = (x_2 - x_1) \left\{ \frac{1}{2} (f(x_1)^s + f(x_2)^s) \right\}^{1/s} \]

\[ \geq (x_2 - x_1) \left( \sqrt{f(x_1)^s \cdot f(x_2)^s} \right)^{1/s} \]

since \( u^\alpha v^{1-\alpha} \leq \alpha u + (1 - \alpha)v \)

\[ = (x_2 - x_1) \sqrt{f(x_1) f(x_2)}. \]

This is very similar to the proof of the first part of Lemma A.1 of Dümbgen and Rufibach [2009].

Part (ii) of the Proposition is the second part of Lemma A.1 of Dümbgen and Rufibach [2009]. It remains to prove (i) and (iii).

To prove (i), suppose that \( f \in F_{s,1} \) with \(-1 \leq s < 0\). Then (8.12) yields, with \( x_2 \equiv x, \ x_1 = x_0, \ x > x_0, \)

\[ \frac{F(x) - F(x_0)}{f(x_0)(x - x_0)} \geq \sqrt{\frac{f(x)}{f(x_0)}}, \]

or

\[ \frac{f(x)}{f(x_0)} \leq \left(\frac{F(x) - F(x_0)}{f(x_0)(x - x_0)}\right)^2 < 1 \]
if \( f(x_0)(x - x_0) > F(x) - F(x_0) \). But recall that

\[
F(x) - F(x_0) \geq (x - x_0) \int_0^1 \{(1 - u)\varphi(x_0) + u\varphi(x)\}^{1/s} \, du
\]

\[
= (x - x_0) \int_0^1 \varphi(x_0)^{1/s} \left\{ 1 + u \left( \frac{\varphi(x)}{\varphi(x_0)} - 1 \right) \right\}^{1/s} \, du
\]

\[
= (x - x_0) f(x_0) J_{1,s} \left( \frac{-1}{s} \cdot \left( \frac{\varphi(x)}{\varphi(x_0)} - 1 \right) \right)
\]

with \( \varphi(x)/\varphi(x_0) - 1 \geq 0 \) where

\[
J_{1,s}(y) \equiv \int_0^1 (1 - sy)^{1/s} \, du \rightarrow \int_0^1 e^{-yu} \, du \equiv J_0(-y)
\]

as \( s \searrow 0 \). But

\[
J_{1,s}(y) = \int_0^1 (1 - syu)^{1/s} \, du = \frac{s}{1 + s} \cdot \frac{(-1 + (1 - sy)^{1+1/s})}{-sy}
\]

\[
= \frac{1}{1 + s} \cdot \frac{1 - (1 - sy)^{(1+1/s)}}{y} \rightarrow \frac{1 - e^{-y}}{y}
\]

as \( s \searrow 0 \), and satisfies (see Lemma 8.5 below)

\[
J_{1,s}(y) \geq \frac{1}{1 + y}, \quad \text{for all } y \geq 0, \quad -1 < s < 0.
\]

This yields

\[
F(x) - F(x_0) \geq (x - x_0) f(x_0) J_{1,s} \left( \frac{-1}{s} \cdot \left( \frac{\varphi(x)}{\varphi(x_0)} - 1 \right) \right)
\]

\[
\geq (x - x_0) f(x_0) \frac{1}{1 + \left( \frac{-1}{s} \right) \left( \frac{\varphi(x)}{\varphi(x_0)} - 1 \right)},
\]

\[
= (x - x_0) f(x_0) \frac{1}{1 + s^{-1} + \left( \frac{-1}{s} \right) \left( \frac{\varphi(x)}{\varphi(x_0)} \right)}
\]

or

\[
1 + s^{-1} + \left( \frac{-1}{s} \right) \left( \frac{f(x)}{f(x_0)} \right)^s \geq \frac{f(x_0)(x - x_0)}{F(x) - F(x_0)}
\]

or, equivalently,

\[
\left( \frac{f(x)}{f(x_0)} \right)^s \geq (-s) \left\{ \frac{f(x_0)(x - x_0)}{F(x) - F(x_0)} - (1 + s^{-1}) \right\},
\]
and hence that the claimed inequality holds:

\[
f(x) \leq f(x_0) \cdot \left\{ 1 + s + (-s) \left( \frac{f(x_0)(x - x_0)}{F(x) - F(x_0)} \right) \right\}^{1/s}.
\]

To prove (iii), suppose that \( f \in F_{s,1} \) with \( 0 < s \leq 1 \) and let \( x_0, x \in \{ f > 0 \} \) with \( x > x_0 \). Then (8.12) yields, with \( x_2 \equiv x, x_1 = x_0, x > x_0 \),

\[
\frac{F(x) - F(x_0)}{f(x_0)(x - x_0)} \geq \sqrt{\frac{f(x)}{f(x_0)}},
\]

or

\[
\frac{f(x)}{f(x_0)} \leq \left( \frac{F(x) - F(x_0)}{f(x_0)(x - x_0)} \right)^2 < 1
\]

if \( f(x_0)(x - x_0) > F(x) - F(x_0) \). But recall that (8.15) yields

\[
F(x) - F(x_0) \geq (x-x_0) \left\{ (1-u)\varphi(x_0) + u\varphi(x) \right\}^{1/s} du
\]

\[
= (x-x_0) \int_0^1 \varphi(x_0)^{1/s} \left\{ 1 + u \left( \frac{\varphi(x)}{\varphi(x_0)} - 1 \right) \right\}^{1/s} du
\]

\[
= (x-x_0)f(x_0)J_{3,s} \left( -\frac{1}{s} \left( \frac{\varphi(x)}{\varphi(x_0)} - 1 \right) \right)
\]

where now \( \varphi(x)/\varphi(x_0) - 1 \leq 0 \) and where \( J_{3,s} \) is defined for \( 0 \leq y \leq 1/s \) by

\[
(8.17) \quad J_{3,s}(y) = \int_0^1 (1-sy)^{1/s} du = \frac{1}{1+s} \cdot \frac{1-(1-sy)^{1+1/s}}{y}
\]

satisfies \( J_{3,s}(y) \geq 1/(1+y) \) for all \( 0 \leq y \leq 1/s \); see Lemma 8.5 below. This yields

\[
F(x) - F(x_0) \geq (x-x_0)f(x_0)J_{3,s} \left( -\frac{1}{s} \left( \frac{\varphi(x)}{\varphi(x_0)} - 1 \right) \right)
\]

\[
\geq (x-x_0)f(x_0) \frac{1}{\left( 1 - \frac{1}{s} \left( \frac{\varphi(x)}{\varphi(x_0)} - 1 \right) \right)},
\]

or

\[
\left( 1 - \frac{1}{s} \left( \frac{\varphi(x)}{\varphi(x_0)} - 1 \right) \right) \geq \frac{f(x_0)(x - x_0)}{h(x_0, x)},
\]
or
\[
1 + \frac{1}{s} - \frac{f(x_0)(x - x_0)}{h(x_0, x)} \geq \frac{1}{s} \frac{\varphi(x)}{\varphi(x_0)} = \frac{1}{s} \left( \frac{f(x)}{f(x_0)} \right)^s,
\]
or
\[
\frac{f(x)}{f(x_0)} \leq \left\{ 1 + s \left( 1 - \frac{f(x_0)(x - x_0)}{h(x_0, x)} \right) \right\}^{1/s},
\]
and this yields the claimed inequality in (iii).

The bounds for the functions \( J_1 \) and \( J_2 \) needed in this proof are established in the following lemma:

**Lemma 8.5.** (a) For \(-1 \leq s < 0\)
\[
J_{1,s}(y) = \begin{cases} \frac{1}{1+\frac{1}{s}} \cdot y^{-1}(1 - (1 - sy)^{1+1/s}), & -1 < s < 0, \\ y^{-1} \log(1 + y), & s = -1, \end{cases}
\]
satisfies
\[
J_{1,s}(y) \geq \frac{1}{1+y} \text{ for all } y \geq 0.
\]
(b) For \(0 < s \leq 1\)
\[
J_{3,s}(y) = \frac{1}{1+s} \frac{1 - (1 - sy)^{1+1/s}}{y}
\]
satisfies
\[
J_{3,s}(y) \geq \frac{1}{1+y} \text{ for all } 0 \leq y \leq 1/s.
\]

**Proof.** (a) First, when \( s = -1 \), the claimed inequality holds if and only if
\[
(1 + y) \log(1 + y) - y \geq 0
\]
for all \( y \geq 0 \). But the left side of the last display is \( h(1 + y) \) where \( h(x) \equiv x \log(x - 1) + 1 \geq 0 \) for all \( x \geq -1 \) with equality if and only if \( x = 1 \); see e.g. Shorack and Wellner [1986], page 416.

For \(-1 < s < 0\), the claimed inequality holds if and only if
\[
1 - (1 - sy)^{1+1/s} \geq (1 + s) \frac{y}{1+y},
\]
or, equivalently, if and only if $1 - (1 + s)y/(1 + y) \geq (1 - sy)^{1+1/s}$, or, equivalently, if and only if $1 - sy \geq (1+y)(1-sy)^{1+1/s}$, or, equivalently, if and only if $1 \geq (1 + y)/(1 + ry)^{1/r}$, or, equivalently, if and only if $(1 + ry) \geq (1 + y)^r$, where $r = -s \in (0, 1)$ for $s \in (-1, 0)$. But $(1 + ry) \geq (1 + y)^r$ holds for $y \geq 0$ and $0 < r \leq 1$, so the claimed inequality holds.

(b) Much as in the proof of (a), for $-1 < s < 0$, the claimed inequality in (a) holds if and only if $1 - (1 - sy)^{1+1/s} \geq (1 + s)y/(1 + y)$, or, equivalently, if and only if $1 - (1 + sy)/(1 + y) \geq (1 - sy)^{1+1/s}$, or, equivalently, if and only if $1 \geq (1+y)(1-sy)^{1+1/s}$, or, equivalently, if and only if $(1 + ry)/(1 - sy)^{1/s}$, or, equivalently, if and only if $1 \geq (1 + y)^s(1 - sy)$ for all $0 \leq y \leq 1/s$. But this inequality holds for $0 < s \leq 1$ so the inequality of part (c) holds.

PROOF. Proposition 8.3: Let $F(x) = \int_{-\infty}^{x} f(y) \, dy$. Take $x_1, x_2 \in \mathbb{R}$ with $x_1 < x_2$. Then

$$F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x) \, dx = \int_{x_1}^{x_2} h(\varphi(x)) \, dx = \int_{x_1}^{x_2} h \left( \varphi \left( \frac{x_2 - x}{x_2 - x_1} + \frac{x - x_1}{x_2 - x_1} \right) \right) \, dx,$$

and since $h$ is nondecreasing and $\varphi$ is concave, the above is not smaller than

$$\int_{x_1}^{x_2} h \left( \varphi \left( \frac{x_2 - x}{x_2 - x_1} \varphi(x_1) + \frac{x - x_1}{x_2 - x_1} \varphi(x_2) \right) \right) \, dx,$$

which, by the change of variables $u = (x - x_1)/(x_2 - x_1)$, can be written as

$$\int_{0}^{1} h \left( (1 - u)\varphi(x_1) + u\varphi(x_2) \right) (x_2 - x_1) \, du. \tag{8.18}$$

Now we let $x_1 = x_0$ and $x_2 = x$ with $x_0 < x_1 < x$ as in the statement. Since $x_0$ and $x_1$ are in dom $\varphi$,

$$C \equiv \int_{0}^{1} h((1 - u)\varphi(x_0) + u\varphi(x_1)) \, du \tag{8.19}$$

satisfies

$$0 < h(\varphi(x_1)) \leq C \leq h(\varphi(x_0)) \tag{8.20}.$$
Now, let \( \eta = (\varphi(x_0) - \varphi(x_1))/(\varphi(x_0) - \varphi(x)) \), so that \( \eta \in (0, 1) \) by the assumption of the proposition. Then

\[
\int_0^1 h((1 - u)\varphi(x_0) + u\varphi(x)) \, du \\
= \left( \int_0^\eta + \int_{\eta}^1 \right) h((1 - u)\varphi(x_0) + u\varphi(x)) \, du \\
\geq \int_0^\eta h((1 - u)\varphi(x_0) + u\varphi(x)) \, du.
\]

Then by the substitution \( v = u/\eta \), this is equal to

\[
\int_0^1 h((1 - \eta v)\varphi(x_0) + \eta v\varphi(x)) \eta \, dv.
\]

which is

\[
\int_0^1 h ((1 - v)\varphi(x_0) + v\varphi(x_1)) \frac{\varphi(x_0) - \varphi(x_1)}{\varphi(x_0) - \varphi(x)} \, dv,
\]

by the construction of \( \eta \), i.e. because

\[
(1 - \eta v)\varphi(x_0) + \eta v\varphi(x) = \left( 1 - \frac{\varphi(x_0) - \varphi(x_1)}{\varphi(x_0) - \varphi(x)} v \right) \varphi(x_0) + \frac{\varphi(x_0) - \varphi(x_1)}{\varphi(x_0) - \varphi(x)} v\varphi(x) \\
= v\varphi(x_0) + \frac{\varphi(x_0) - \varphi(x_1)}{\varphi(x_0) - \varphi(x)} v(\varphi(x) - \varphi(x_0))) \\
= v\varphi(x_0) + v(\varphi(x_0) - \varphi(x_1)) \\
= (1 - v)\varphi(x_0) + v\varphi(x_1).
\]

And, by definition of \( C \), (8.22) equals \( C(\varphi(x_0) - \varphi(x_1))/(\varphi(x_0) - \varphi(x)) \). This gives, by applying (8.18), that

\[
F(x) - F(x_0) \geq (x - x_0) \int_0^1 h((1 - u)\varphi(x_0) + u\varphi(x)) \, du \\
\geq (x - x_0)C\frac{\varphi(x_0) - \varphi(x_1)}{\varphi(x_0) - \varphi(x)}.
\]

Now we rearrange the above display to get an inequality for \( \varphi(x) \). From (8.23), we have

\[
\varphi(x) \leq \varphi(x_0) - C\frac{\varphi(x_0) - \varphi(x_1)}{F(x) - F(x_0)} (x - x_0),
\]
and, since \( h \) is nondecreasing,
\[
\begin{align*}
    h(\varphi(x)) &\leq h \left( \varphi(x_0) - C \varphi(x_0) - \varphi(x_1) \right) \\
    &\leq h \left( \varphi(x_0) - h(\varphi(x_1)) \right) \\
    &\leq \frac{\varphi(x_0) - \varphi(x_1)}{F(x) - F(x_0)} (x - x_0),
\end{align*}
\]
by (8.20). This proves the claim for \( x_0 < x_1 < x \). The proof for \( x < x_1 < x_0 \) is similar.

\[ \square \]

9. Appendix, part 2: Proofs of Propositions 5.1 and 5.2.

Note that by Rockafellar [1970], Theorem 23.1, page 213, right- and left-derivatives are well defined at all points in \( \text{dom}(\varphi) \) if we allow the derivatives to take the values \( \pm \infty \). The classes \( \mathcal{C}^* \) to be defined below have fixed domains and bounded Lipschitz constants, which allows us to use Theorem 3.2 of Guntuboyina and Sen [2013] (which is based on earlier work by Bronštejn [1976]; see also Dudley [1984]) to control the bracketing entropy of the class. In Guntuboyina and Sen [2013], Theorem 3.2 is the building block for the proof of their Theorem 3.1. Our Proposition 5.1 (and Proposition 9.1 here) extends their Theorem 3.1, via a similar proof method. We will extend the bracketing entropy bound to allow for classes with varying domains. We now define these classes more precisely.

**Definition 9.1.** Let \( -\infty < b_1 < b_2 < \infty \), \( -\infty < B_1 < B_2 < \infty \), \( c_2 > 0 \), and \( 0 < \epsilon \leq \epsilon_0 B(b_2 - b_1)^{1/r} \) for absolute constants \( c_2 > 0 \) and \( \epsilon_0 > 0 \). Then
\[
\log N_{\frac{1}{2}} \left( \epsilon, \mathcal{C}\left([b_1, b_2], [B_1, B_2], c_2 \epsilon^r (b_2 - b_1) \right), L_r(\lambda) \right) \leq c \left( \frac{(B_2 - B_1)(b_2 - b_1)^{1/r}}{\epsilon} \right)^{1/2},
\]
(9.1)
We may take all brackets \([l, u]\) such that \(l(x) = B_1\) and \(u(x) = B_2\) for all \(x\) such that \(|x - b_i| < c_2\epsilon^r(b_2 - b_1)\), i.e. for all \(x\) in the set where the domains may end.

We also note that we can simply state Theorem 3.1 in Guntuboyina and Sen [2013] in terms of bracketing entropy instead of metric entropy. This yields:

**Proposition 9.2** (Extension of Theorem 3.1 of Guntuboyina and Sen [2013]). Let \(r \geq 1\), \(-\infty < b_1 < b_2 < \infty\), \(-\infty < B_1 < B_2 < \infty\), and \(0 < \epsilon \leq \epsilon_0(B_2 - B_1)(b_2 - b_1)^{1/r}\), where \(\epsilon_0 > 0\) is an absolute constant. Then

\[
\log N_{[l]}(\epsilon, C([b_1, b_2], [B_1, B_2], 0), L_r(\lambda)) \leq c \left( \frac{(B_2 - B_1)(b_2 - b_1)^{1/r}}{\epsilon} \right)^{1/2},
\]

where \(c\) is a constant depending only on \(r\).

**Proof.** The proof consists mostly of noticing that Theorem 3.1 in Guntuboyina and Sen [2013] essentially yields the result stated here. In our proof here we will attempt to maintain similar notation to the original proof of their result. We will prove the result for domains ending in the interval \([b_1, b_1 + c_2\epsilon^r]\), since the arguments are totally symmetric and this allows us to line up our argument with that of Guntuboyina and Sen [2013]. The proof is entirely symmetric if the domains end in the interval \([b_2 - c_2\epsilon^r, b_2]\).

First, notice that the \(L_r\) bracketing numbers scale in the following fashion. For a function \(f \in C([b_1, b_2], [B_1, B_2], c_2\epsilon^r(b_2 - b_1))\) we can define

\[
\tilde{f}(x) := \frac{f(b_1 + (b_2 - b_1)x) - (B_1 + B_2)/2}{(B_2 - B_1)/2},
\]

a scaled and translated version of \(f\) that satisfies \(\tilde{f} \in C([0, 1], [-1, 1], c_2\epsilon^r)\). Thus, if \([l, u]\) is a bracket for \(C([b_1, b_2], [B_1, B_2], c_2\epsilon^r(b_2 - b_1))\), then we have

\[
\left( \frac{B_2 - B_1}{2} \right)^{r} \int_{0}^{1} \left| \tilde{u}(x) - \tilde{l}(x) \right|^r dx = \frac{1}{b_2 - b_1} \int_{b_1}^{b_2} |u(x) - l(x)|^r dx.
\]

Thus an \(\epsilon\)-size \(L_r\) bracket for \(C([0, 1], [-1, 1], c_2\epsilon^r)\) immediately scales to be an \(\epsilon(b_2 - b_1)^{1/r}(B_2 - B_1)/2\) bracket for \(C([b_1, b_2], [B_1, B_2], c_2\epsilon^r(b_2 - b_1))\). Thus for the remainder of the proof we set \(b_1 = 0\), \(b_2 = 1\), \(B_1 = -1\), \(B_2 = 1\), and \(u = c_2\epsilon^r\).
We take the domain to be fixed for these classes so that we can apply the result Theorem 3.2 of Guntuboyina and Sen [2013] which is the building block of the proof. Now we fix

\[(9.3) \quad \mu := \exp(-2(r + 1)^2(r + 2) \log 2) \quad \text{and} \quad \nu := 1 - \mu.\]

(Note that $\mu$ and $\nu$ are $u$ and $v$, respectively, in Guntuboyina and Sen [2013].)

We will consider the intervals $[0, \mu]$, $[\mu, \nu]$, and $[\nu, 1]$ separately, and will show the bound (5.2) separately for the restriction of $C([0, 1], [-1, 1], c_2 \epsilon^r)$ to each of these sub-intervals. This will imply (5.2). We fix $\epsilon > 0$, let $\eta = (3/17)^{1/r} \epsilon$, choose an integer $A$ and $\delta_0, \ldots, \delta_{A+1}$ such that

\[(9.4) \quad 0 = \delta_0 < \delta_1 < \eta = \delta_2 < \cdots < \delta_A < \mu \leq \delta_{A+1}.\]

We set $\delta_1 := \delta_2/2$, and define $c_2 = 3/34$. Thus with these definitions, $\delta_1 = c_2 \epsilon^r$, so that by definition the class $C([0, 1], [-1, 1], c_2 \epsilon^r)$ has its jumps down in the interval $[0, \delta_1]$. For two functions $f$ and $g$ on $[0, 1]$, we can decompose the integral $\int_0^1 |f - g|^r d\lambda$ as

\[(9.5) \quad \int_0^1 |f - g|^r d\lambda = \int_0^\mu |f - g|^r d\lambda + \int_\mu^\nu |f - g|^r d\lambda + \int_\nu^1 |f - g|^r d\lambda.\]

The first term can be bounded by

\[(9.6) \quad \int_0^\mu |f - g|^r d\lambda \leq \sum_{m=0}^A \int_{\delta_m}^{\delta_{m+1}} |f - g|^r d\lambda,\]

since $\delta_{A+1} \geq \mu$. Now for a fixed $m \in \{2, \ldots, A\}$, we consider the problem of covering the functions in $C([0, 1], [-1, 1], c_2 \epsilon^r)$ on the interval $[\delta_m, \delta_{m+1}]$. Defining $f(x) = f(\delta_m + (\delta_{m+1} - \delta_m) x)$ and $g(x) = g(\delta_m + (\delta_{m+1} - \delta_m) x)$, we have

\[(9.7) \quad \int_{\delta_m}^{\delta_{m+1}} |f - g|^r d\lambda = (\delta_{m+1} - \delta_m) \int_0^1 |\tilde{f} - \tilde{g}|^r d\lambda.\]

Since concavity is certainly preserved by restriction of a function, the restriction of any function $f$ in $C([0, 1], [-1, 1], c_2 \epsilon^r)$ to $[\delta_m, \delta_{m+1}]$ belongs to the Lipschitz class with fixed domain $C^r([\delta_m, \delta_{m+1}], [-1, 1], 2/|\delta_{m+1} - \delta_m|)$ (since the domain of $f$ could end at $\delta_1$, so the largest “rise” is of size $1 - (-1) = 2$ over a smallest “run” of $\delta_{m+1} - \delta_1$). Thus the corresponding $\tilde{f}$ belongs to $C^r([0, 1], [-1, 1], 2(\delta_{m+1} - \delta_m)/(\delta_m - \delta_1))$. We can now use Theorem 3.2 of Guntuboyina and Sen [2013] to assert the existence of positive constants $\epsilon_0$ and $c$ that depend only on $r$ such that for all $\alpha_m \leq \epsilon_0$ there exists an
MLE LOG-CONCAVE GLOBAL RATES

\( \alpha_m \)-net for \( C^*([0, 1], [-1, 1], 2(\delta_{m+1} - \delta_m)/(\delta_m - \delta_1)) \) in the supremum norm of cardinality smaller than

\[
\exp \left( c \alpha_m^{-1/2} \left( 2 + \frac{2(\delta_{m+1} - \delta_m)}{\delta_m - \delta_1} \right)^{1/2} \right) \leq \exp \left( c \left( \frac{\delta_{m+1}}{(\delta_m - \delta_1)\alpha_m} \right)^{1/2} \right).
\]

Denote the net by \( \{f_{m,n_m}: n_m = 1, \ldots, N_m\} \) where \( N_m \) is bounded by (9.8) and \( m = 2, \ldots, A \). Now, define the brackets \( [l_{n_m}, u_{n_m}] \) by

\[
l_{n_m}(x) \equiv -1_{[0, \delta_2]}(x) + \sum_{m=2}^{A} 1_{[\delta_m, \delta_{m+1}]}(x) (f_{m,n_m}(x) - \alpha_m),
\]

\[
u_{n_m}(x) \equiv 1_{[0, \delta_2]}(x) + \sum_{m=2}^{A} 1_{[\delta_m, \delta_{m+1}]}(x) (f_{m,n_m}(x) + \alpha_m)
\]

for the restrictions of the functions in \( C([0, 1], [-1, 1], c_2\epsilon^r) \) to the set \([0, \mu]\), where the tuple \( (n_2, \ldots, n_A) \) defining the bracket varies over all possible tuples with components \( n_m \leq N_m, m = 2, \ldots, A \). The brackets were chosen in the supremum norm, so we can compute their \( L_r(\lambda) \) size as \( S_1^{1/r} \) where

\[
S_1 = 2^r \delta_2 + \sum_{m=2}^{A} (2\alpha_m)^r (\delta_{m+1} - \delta_m),
\]

and the cardinality is \( \exp(S_2) \) where

\[
S_2 = c \sum_{m=2}^{A} \left( \frac{\delta_{m+1}}{(\delta_m - \delta_1)\alpha_m} \right)^{1/2}.
\]

Since we picked \( \delta_1 = \delta_2/2 \), we have \( 2(\delta_2 - \delta_1) = \delta_2 \). Then for \( m > 2 \),

\[
2(\delta_m - \delta_1) = 2(\delta_m - \delta_2) + 2(\delta_2 - \delta_1) \geq \delta_m - \delta_2 + \delta_m - \delta_2 + \delta_2 = \delta_m - \delta_2 + \delta_m > \delta_m.
\]

Thus \( 1/\delta_m > 1/(2(\delta_m - \delta_1)) \) or \( 2/\delta_m > 1/(\delta_m - \delta_1) \), so

\[
S_2 = c \sum_{m=2}^{A} \left( \frac{\delta_{m+1}}{(\delta_m - \delta_1)\alpha_m} \right)^{1/2} \leq c \sum_{m=2}^{A} \left( \frac{2\delta_{m+1}}{\delta_m\alpha_m} \right)^{1/2} \leq \sqrt{2}c \sum_{m=2}^{A} \left( \frac{\delta_{m+1}}{\delta_m\alpha_m} \right)^{1/2}.
\]

Thus our \( S_1 \) and \( S_2 \) are identical to those in (7) in Guntuboyina and Sen [2013], except that the absolute constants have changed in \( S_1 \) and \( S_2 \) and
our index $m$ is off by one (our $\delta_2$ is their $\delta_1$). Thus replacing their $m$ by $m - 1$, the remaining argument is now identical to the remaining argument there, in which $\delta_m$ and $\alpha_m$ are chosen and the size of $S_1$ and $S_2$ are computed. Our values are identical except $S_2$ is off by a factor of $\sqrt{2}$. Thus Proposition 5.1 holds.

Now for the proof of Proposition 5.2. If we replace $B_1$ by $-B$ and $B_2$ by $B$ for $B > 0$, then

\begin{equation}
\log N[| (\epsilon, \mathcal{C}([b_1, b_2], [-B, B], 0), L_p) \leq c \left( \frac{B(b_2 - b_1)^{1/p}}{\epsilon} \right)^{1/2},
\end{equation}

follows from Theorem 5.1 since we consider a subclass of the class considered there (i.e. we take the endpoints of the domains of all functions to be equal to $b_1$ and $b_2$ rather than allowing them some room to vary). Note this class is similar to the $C^*$ classes but we have no constraint on the Lipschitz constant. The only difference between this statement and Theorem 3.1 of Guntuboyina and Sen [2013] is that we consider bracketing numbers here instead of covering numbers.

We thus need only to allow for general $-\infty < B_1 < B_2 < \infty$, which we do by translating. Note $\mathcal{C}([b_1, b_2], [B_1, B_2], 0) - B_1 = \mathcal{C}([b_1, b_2], [0, B_2 - B_1], 0)$, i.e. if $\varphi(x)$ takes values in $[B_1, B_2]$ then $\varphi(x) - B_1$ lies in $[0, B_2 - B_1]$, and concavity is preserved by this translation. Since the bracketing numbers of $\mathcal{C}([b_1, b_2], [0, B_2 - B_1], 0)$ and of $\mathcal{C}([b_1, b_2], [B_1, B_2], 0) - B_1$ are the same, and the bracketing number of $\mathcal{C}([b_1, b_2], [0, B_2 - B_1], 0)$ is given by (9.12) (since it is smaller than the bracketing number of $\mathcal{C}([b_1, b_2], [-(B_2 - B_1), B_2 - B_1], 0)$), Proposition 5.2 holds.

10. Appendix, part 3: The Dual Induction Proof.

**Proposition 10.1.** Let $h$ be as in Definition 6.1, let $y_k \in \mathbb{R}^{k+1}$ as in Definition 6.3, and let $d_{r,h}$ and its corresponding weight sequence $w_{\gamma,h}$ be as in Definition 6.4. Fix $\epsilon > 0$ and assume that for all $1 \leq \gamma \leq k$, $y_{\gamma}$ satisfy

\begin{equation}
\epsilon / w_{\gamma,h} \leq \epsilon_0 (b_2 - b_1)^{1/r}(y_{\gamma-1} - y_{\gamma}),
\end{equation}

for $\epsilon_0 > 0$ a positive constant (not necessarily defined as in Theorem 5.1).
Then we have for any \( \zeta \in [0, 1] \), \( 1 \leq r < \infty \),

\[
\log N[\epsilon \cdot \left( 1 + \sum_{\gamma=1}^{k} h(y_{\gamma-1})^{r(1-\zeta)} \right)^{1/r}, \mathcal{F}_{k,h}, L_r(\lambda)] \leq c \sum_{\gamma=1}^{k} \left\{ \left( \frac{(y_{\gamma-1} - y_{\gamma})(b_2 - b_1)^{1/r}}{\epsilon/w_{\gamma,h}} \right)^{1/2} \right.
\]
\[
+ 2 \log \left( 1 + \frac{b_2 - b_1}{\epsilon^r/h(y_{\gamma-1})^{r\zeta}} \right) \right\}.
\]

(10.2)

**Proof.** In this proof we will use \( L_r \) to denote \( L_r(\lambda) \) where \( \lambda \) is Lebesgue measure. We will assume that all \( w_{\gamma,h} \) are finite, since otherwise the right-hand side of (10.3) is infinite and the conclusion is trivial. Additionally, note that (10.1) forces \( w_{\gamma,h} > 0 \) for all \( \gamma \). We will first reduce to the case \( b_1 = 0 \) and \( b_2 = 1 \) by translating and scaling the domain. We will not scale the \( y_{\gamma} \), however. For any function \( f \in \mathcal{F}_{k,h} \) with \( \text{dom } f \subseteq [b_1, b_2] \), we can define a function \( \tilde{f}(x) = f(b_1 + (b_2 - b_1)x) \) with \( \text{dom } \tilde{f} \subseteq [0, 1] \). \( \tilde{f} \) is an element of \( \mathcal{F}_{k,h} \) with domain specified as \([0, 1]\). If the proposition holds when \( b_1 = 0 \) and \( b_2 = 1 \), then we can find a collection of brackets \( \tilde{l} \) and \( \tilde{u} \) such that \( \tilde{l} \leq \tilde{f} \leq \tilde{u} \). We can invert the above scaling to arrive at \( l \) and \( u \) (e.g. \( l(x) = \tilde{l}((x - b_1)/(b_2 - b_1)) \)) such that \( l \leq f \leq u \). For any two functions \( f_1 \) and \( f_2 \) defined on \([0, 1]\), we have

\[
d_{r,k,h}(f_1, f_2) = \left( \int_{b_1}^{b_2} d_{k,h}(f_1(x), f_2(x))^r dx \right)^{1/r} = \frac{1}{(b_2 - b_1)^{1/r}} d_{r,k,h}(\tilde{f}_1, \tilde{f}_2).
\]

Applying this to \([\tilde{l}, \tilde{u}]\) and the corresponding \([l, u]\), we see that if we are given a collection of \( \epsilon \)-brackets when \( b_1 = 0 \) and \( b_2 = 1 \) then we immediately have a collection of \((b_2 - b_1)^{1/r} \epsilon\)-brackets of the same cardinality for general \( b_1 \) and \( b_2 \), i.e. if the proposition holds when \( b_1 = 0 \) and \( b_2 = 1 \) then it holds for general \( b_1 \) and \( b_2 \).

We thus now proceed with \( b_1 = 0 \) and \( b_2 = 1 \). We set \( \eta_{\gamma,h} = \epsilon/h(y_{\gamma-1})^{\zeta} \) and \( \epsilon_{\gamma,h} = \epsilon/w_{\gamma,h} \). Note that \( 0 < \epsilon_{\gamma,h} < \infty \) by the comments at the beginning of the proof. The proof now proceeds by induction. The induction
hypothesis is as follows.

**Induction Hypothesis:**

1. For \( k^* \in \mathbb{N} \), we assume that:
   
   (a) (Bracketing) There is a collection of \( d_{r,k^*,h} \) brackets \([l_\beta, u_\beta]\) for the class \( C_{k^*} \) (defined in Definition 6.2, using \( b_1 = 0 \) and \( b_2 = 1 \));
   
   (b) (Cardinality) and that they have cardinality
   
   \[
   \exp \left( c \sum_{\gamma=1}^{k^*} \left( \frac{\left( y_{\gamma-1} - y_{\gamma} \right)}{\epsilon/w_{\gamma,h}} \right)^{1/2} \right) + (1 + 1[1 \leq \gamma \leq k^*-1]) \log \left( 1 + \frac{1}{\epsilon^f/h(y_{\gamma-1})^{1/2}} \right). \tag{10.3}
   \]

2. Denoting \( \text{dom}(\varphi) \) by \([0, d_{\varphi}]\) we now define the classes
   
   \( C_{\gamma,j} := \{ \varphi \in C_{\gamma} | d_{\varphi} \in ((j-1)\eta_{\gamma,h}, j\eta_{\gamma,h}] \} \subseteq C ([0, j\eta_{\gamma,h}, [y_{\gamma}, y_0], \eta_{\gamma}') \),
   
   for \( j = 1, \ldots, N_{\gamma} := \lfloor 1/\eta_{\gamma,h} \rfloor \), which disjointly partition \( C_{\gamma}, \gamma = 1, \ldots, k^* \). We assume that each bracket \([l_\beta, u_\beta]\) has a corresponding index \( j_{\gamma,\beta} \), and:
   
   (a) (Partition) We assume that all the brackets \([l_\beta, u_\beta]\) that share an index \( j = jk^*,\beta \) form a collection of brackets for the class \( C_{k^*,j} \);
   
   (b) (Extension) we assume that if \( \varphi \in C_{\gamma,j}, \) for some \( 1 \leq j \leq N_{\gamma} \) and \( 1 \leq \gamma \leq k^* \), and \( \varphi \) is bracketed by \([l_\beta, u_\beta]\) (which thus has index \( j_{\gamma,\beta} = j \)), then any \((j, \gamma)\)-concave extension \( \phi \) of \( \varphi \) is bracketed by \([l_\beta, u_\beta]\), where a \((j, \gamma)\)-concave extension \( \phi \) is a function such that
   
   \[
   \phi(x) = \varphi(x) \text{ for } x \in \text{dom}(\varphi), \quad \text{dom}(\phi) \subseteq [0, j\eta_{\gamma,h}], \quad \phi(x) \leq y_0 \text{ for } x \in \mathbb{R}, \quad \text{and } \phi \in C.
   \]

3. (Size) We assume each bracket \([l_\beta, u_\beta]\) has a corresponding set of “jump-down” intervals \( J_{\gamma,\beta} := ((j_{\gamma,\beta}-1)\eta_{\gamma,h}, j_{\gamma,\beta}\eta_{\gamma,h}] \) \([0, j_{\gamma-1,\beta}\eta_{\gamma-1,h}], \gamma = 1, \ldots, k^* \), where we take \( j_{0,\beta}\eta_{0,h} \) to be 0 for all \( \beta \). For \( \gamma = 1, \ldots, k^* \), we assume for all \( \beta \) that the brackets \([l_\beta, u_\beta]\) satisfy:
   
   (a) for all \( \tilde{k} \geq k^* \)
   
   \[
   d_{r,\tilde{k},h}(u_\beta \cdot 1_{(\cup_{\gamma=1}^{k^*} J_{\gamma,\beta})^c}, l_\beta \cdot 1_{(\cup_{\gamma=1}^{k^*} J_{\gamma,\beta})^c}) \leq \epsilon (j_{k^*,\beta}\eta_{k^*,h})^{1/r}; \tag{10.4}
   \]
   
   (b) and
   
   \[
   \|(h \circ u_\beta - h \circ l_\beta)1_{(\cup_{\gamma=1}^{k^*} J_{\gamma,\beta})^c} \|^r \leq \epsilon \sum_{\gamma=1}^{k^*} h(y_{\gamma-1})^{r(1-\zeta)}. \tag{10.5}
   \]
4. (Conclusion) We assume that this transformed collection of brackets
\([h \circ l_\beta, h \circ u_\beta]\) for \(F_{k^*, h}\) has size not larger than 
\((1 + \sum_{\gamma=1}^{k^*} h(y_{\gamma-1})^{r(1-\zeta)})^{1/r} \epsilon\) in the \(L_r\) metric, which yields

\[
\log N[\epsilon] \left( \epsilon \left(1 + \sum_{\gamma=1}^{k^*} h(y_{\gamma-1})^{r(1-\zeta)} \right)^{1/r}, F_{k^*, h}, L_r \right)
\]

\begin{equation}
\leq \epsilon \sum_{\gamma=1}^{k^*} \left( \frac{(y_{\gamma-1} - y_{\gamma})}{\epsilon/w_{\gamma,h}} \right)^{1/2}
\end{equation}

\begin{equation}
+ (1 + 1_{[\gamma \leq k^*-1]}) \log \left( 1 + \frac{1}{\epsilon^r/h(y_{\gamma-1})^{r\zeta}} \right).
\end{equation}

This ends the induction hypothesis.

Here we give a brief outline of how the proof works. For our decreasing
sequence \(y_0 > y_1 > \ldots\), we can consider the domain on which a concave
function \(\varphi\) takes values in \([y_{k-1}, y_0]\) and inductively form brackets there,
and we can consider the domain on which \(\varphi\) takes values in \([y_k, y_{k-1}]\) and
form brackets there via Proposition 5.2. Then we merge the two brackets
together. See Figure 1 for a diagram of these three steps.

**Base Case.**

We start by discretizing \([0, 1]\) into intervals of length \(\eta_{1,h}^r\), i.e. we will
consider intervals \((c_{1,j-1}, c_{1,j}\) where \(c_{1,j} := j\eta_{1,h}^r\), for \(j = 1, \ldots, N_1\) and
\(N_1 := \lceil 1/\eta_{1,h}^r \rceil\). For \(\varphi \in C_k\), we let

\begin{equation}
d_{\gamma, \varphi} = \sup \{x \in \mathbb{R} | \varphi(x) \geq y_{\gamma}\}
\end{equation}

for \(\gamma = 1, \ldots, \infty\). Note that \(\infty > d_{\gamma, \varphi} > -\infty\) always, and that for \(\varphi \in C_{i,j,k}\),
\(\varphi(d_{\gamma, \varphi}) = y_{\gamma}\), and for \(d_{1, \varphi} \leq x\), \(\varphi(x) < y_1\). That is, \(\varphi\) is nonincreasing on
\([d_{1, \varphi}, \infty)\). We now consider the classes

\[
C_{1,j} := \{\varphi \in C_1 | d_{1, \varphi} \in (c_{1,j-1}, c_{1,j})\} \subseteq C([0, c_{1,j}], [y_1, y_0], \eta_{1,h}^r),
\]

for \(j = 1, \ldots, N_1\), which disjointly partition \(C_1\). We now take a fixed \(j\) and
will form a collection of brackets for \(C_{1,j}\). To do so, we first note that for all
\(\varphi \in C_{1,j}\),

\begin{equation}
\varphi|_{[0,c_{1,j-1}]} \in C([0, c_{1,j-1}], [y_1, y_0], 0).
\end{equation}

By Proposition 5.2 we can pick a collection of brackets \([l_{\beta,j}^*, u_{\beta,j}^*]\) for the
class (10.8) that have \(L_r\) size given by

\begin{equation}
\int_0^{c_{1,j-1}} |u_{\beta,j}^*(x) - l_{\beta,j}^*(x)|^r dx \leq c_{1,j-1} \epsilon_{1,h}^r,
\end{equation}

\begin{equation}
\int_0^{c_{1,j-1}} |u_{\beta,j}^*(x) - l_{\beta,j}^*(x)|^r dx \leq c_{1,j-1} \epsilon_{1,h}^r,
\end{equation}
and with cardinality bounded by

\begin{equation}
1 \leq \beta \leq \exp \left( c \left( \frac{(y_0 - y_1)}{\epsilon_{1,h}} \right)^{1/2} \right),
\end{equation}

since $c_{1,j-1}^{1/r} \epsilon_{1,h} = c_{1,j-1}^{1/r} \epsilon/w_{1,h} \leq \epsilon_0 (y_0 - y_1) c_{1,j-1}^{1/r}$, by assumption (10.1).

Now we extend these brackets to form brackets for $C_1$, defined by

\begin{equation}
[l_{\beta,j}(x), u_{\beta,j}(x)] = \begin{cases}
[l_{\beta,j}^*(x), u_{\beta,j}^*(x)], & 0 \leq x \leq c_{1,j-1} \\
[\infty, y_0], & c_{1,j-1} < x \leq c_{1,j} \\
[-\infty, -\infty], & c_{1,j} < x \leq 1
\end{cases}
\end{equation}

and we will define $J_{1,\beta,j} := (c_{1,j-1}, c_{1,j})$.

We will start by showing parts 1(a), 1(b), and 2(a) of the induction hypothesis. For part 1(a) we need to argue for any given $\varphi \in C_1$ that we have bracketed $\varphi$. Since $\varphi$ lies in $C_{1,j}$ for some $j$, it suffices to show that we have bracketed $C_{1,j}$, which is part 2(a). We will consider the three domains $[0, c_{1,j-1}]$, $(c_{1,j-1}, c_{1,j})$, and $(c_{1,j}, 1]$ separately. For $\varphi \in C_{1,j}$, the restriction $\varphi|_{[0,c_{1,j-1}]}$ lies in the class (10.8), and thus for some $\beta$,

\[ l_{\beta,j}(x) = l_{\beta,j}^*(x) \leq \varphi(x) \leq u_{\beta,j}^*(x) = u_{\beta,j}(x) \]

for all $x \in [0, c_{1,j-1}]$. Now for any $\varphi \in C_{1,j}$ and $x \in (c_{1,j-1}, c_{1,j})$

\[ l_{\beta,j}(x) = -\infty \leq \varphi(x) \leq y_0 = u_{\beta,j}(x). \]

Finally, if $x \in (c_{1,j}, 1]$, any $\varphi \in C_{1,j}$ satisfies $\varphi(x) = -\infty$, and then both $u_{\beta,j}(x)$ and $l_{\beta,j}(x)$ are $-\infty$. Thus we have shown that $[l_{\beta,j}, u_{\beta,j}]$ indeed is a collection of brackets for $C_1$ as $j$ ranges from 1 to $N_1$ and $\beta$ ranges over all its values, as in (10.10). We need next to bound the cardinality of our collection. Since $j$ ranges up to $N_1 = \lceil 1/\eta_{1,h}^r \rceil$, we have no more than $1/\eta_{1,h}^r + 1$ times (10.10) total brackets, or

\begin{equation}
\left( \frac{1}{\epsilon^r/h(y_0)^r} + 1 \right) \exp \left( c \left( \frac{(y_0 - y_1)}{\epsilon/w_{r,h}} \right)^{1/2} \right)
\end{equation}

total brackets. We have now shown parts 1(a), 1(b), and 2(a) of the induction hypothesis.

It is also clear that part 2(b) of the induction hypothesis holds. For $\varphi \in C_{1,j}$, any $(j, 1)$-extension $\phi$ of $\varphi$ is in $C_{\tilde{k},j}$ for some $\tilde{k} \geq 1$ by definition, and is equal to $\varphi$ on $[0, c_{1,j-1}]$ and on $(c_{1,j}, 1]$, and thus bracketed by $[l_{\beta,j}, u_{\beta,j}]$.
there. For \( x \in (c_{1,j-1}, c_{1,j}] \), we assumed \( \phi(x) \) is in \([\pm \infty, y_0]\) and so \( \phi \) is bracketed by \([l_{\beta,j}, u_{\beta,j}]\) for such \( x \)'s as well, so we have shown that part \( \text{(b)} \) of the induction hypothesis holds.

Next we show part \( \text{(a)} \) of the induction hypothesis. We only need to consider \( u_{\beta,j} \) and \( l_{\beta,j} \) on \([0, c_{1,j-1}]\) since \( J_{1,\beta,j} = (c_{1,j-1}, c_{1,j}] \), and since \( u_{\beta,j}(x) = l_{\beta,j}(x) = -\infty \) for \( x \in (c_{1,j}, 1] \), i.e. since

\[
d_{r, k, h}(u_{\beta,j}^{1} J_{1,\beta,j}^{1}, l_{\beta,j}^{1} l_{\beta,j}^{1}) = d_{r, 1, h}(u_{\beta,j}^{1} [0, c_{1,j-1}], l_{\beta,j}^{1} [0, c_{1,j-1}]),
\]

for \( k \geq 1 \). We can use any \( k \geq 1 \) because on \([0, c_{1,j-1}]\), the brackets can be taken to have values in \([y_1, y_0]\) since the functions \( \varphi \in C_{1,j} \) that they bracket also have values in \([y_1, y_0]\) (i.e. since ran \( \varphi \) \( \cap [0, c_{1,j-1}] \subset [y_1, y_0] \)), and on \([y_1, y_0]\) the measures \( W(\cdot, y) \) are identical for all \( k \geq 1 \) (they all assign mass \( w_{1,h} \)). Now, on the set \([0, c_{1,j-1}]\) the brackets are defined to be equal to \( u_{\beta,j}^{1} \) and \( l_{\beta,j}^{1} \), which have \( L_r \) size given by \( \epsilon_{1,h}^{1/c_{1,j-1}} \) by \((10.9)\). Now \( \epsilon_{1,h} \) was defined to be \( \epsilon_{1,h} = \epsilon/w_{1,h} \) precisely to offset \( w_{1,h} \). Thus, we can bound the quantity of interest as follows:

\[
d_{r, 1, h}(u_{\beta,j}^{1} [0, c_{1,j-1}], l_{\beta,j}^{1} [0, c_{1,j-1}]) = w_{1,h}^{1/r} \int_{0}^{c_{1,j-1}} |u_{\beta,j}^{1}(x) - l_{\beta,j}^{1}(x)|^r dx
\]

\[
\leq w_{1,h}^{1/r} \epsilon_{1,h}^{1/c_{1,j-1}}
\]

\( (10.13) \)

as desired. We have now shown part \( \text{(a)} \) of the induction hypothesis.

We will now show part \( \text{(b)} \) of the induction hypothesis. We have

\[
\| (h \circ u_{\beta,j} - h \circ l_{\beta,j}) J_{1,\beta,j} \|^r \leq \| h \circ u_{\beta,j} \|_r^{c_{1,j}}
\]

\[= \int_{c_{1,j-1}}^{c_{1,j}} h(u_{\beta,j}(x))^r dx
\]

\[\leq \int_{c_{1,j-1}}^{c_{1,j}} h(y_0)^r dx
\]

\[
\leq \eta_{1,h}^r \epsilon_{1,h}^r
\]

\[= \epsilon^r \eta_{1,h}^r (1-\zeta), \quad (10.18)
\]

where the inequality \((10.16)\) is because (as we have already mentioned) we can take \( u_{\beta,j} \) to be below \( y_0 \) (and \( h \) is nondecreasing), and then \((10.17)\) and \((10.18)\) are because the length of \( J_{1,\beta,j} \) is bounded by \( \eta_{1,h}^r = \epsilon^r / h(y_0)^{\zeta r} \). We have now shown part \( \text{(b)} \) of the induction hypothesis.
It remains to show part 4 of the induction hypothesis. We have
\[
\| (h \circ u_{\beta,j} - h \circ l_{\beta,j})(1_{J_{1,\beta,j}} + 1_{\bar{J}_{1,\beta,j}}) \|_{r}^r
\leq e^r h(y_0) r^{(1-C)} + d_{r,1,h}^r (u_{\beta,j} 1_{J_{1,\beta,j}} , l_{\beta,j} 1_{\bar{J}_{1,\beta,j}})
\leq e^r h(y_0) r^{(1-C)} + e^r (1 + h(y_0) r^{(1-C)}).
\]
We used (6.2) and the fact that \( u_{\beta,j} 1_{J_{1,\beta,j}} \) and \( l_{\beta,j} 1_{\bar{J}_{1,\beta,j}} \) have the shared domain of \([0,c_{j-1}]\), as well as (10.18), for the first inequality, and the second is (10.13) (and the fact that \( c_{j-1} \leq 1 \)). Since we have already found the cardinality of the collection of brackets for \( \mathcal{F}_{1,h} \) (since it is identical to the cardinality for the collection of brackets for \( \mathcal{C}_1 \)), we have shown (10.6) and so we have now shown part 4 of the induction hypothesis, and we are now done showing the base case of the induction proof.

**Induction Case:** Assume the induction hypothesis for \( k^* = k - 1 \) and show it for \( k^* = k \).

The induction case proceeds similarly to the base case, only we discretize based on both the point at which functions cross below \( y_{k-1} \) (to apply the induction hypothesis) and the point at which they cross below \( y_k \). We start by discretizing \([0,1]\) into intervals in which our concave functions \( \varphi \) may cross below \( y_k \), \((c_{k,j-1}, c_{k,j})\) where \( c_{j,\gamma} := j \eta_{k,h}^r \) for \( j = 1, \ldots, N_k \) where \( N_k = \lceil 1/\eta_{k,h}^r \rceil \). Next, we discretize into intervals in which our concave functions may cross below \( y_{k-1} \), \((c_{k-1,i-1}, c_{k-1,i}) := ((i-1) \eta_{k-1,h}^r, i \eta_{k-1,h}^r)\) for \( i = 1, \ldots, N_{j,k} \), where \( N_{j,k} = \lceil j \eta_{k,h}^r / \eta_{k-1,h}^r \rceil \). That is, the intervals \((c_{k-1,i-1}, c_{k-1,i})\) discretize \([0, j \eta_{k,h}^r]\) into intervals that are of equal size \( \eta_{k-1,h}^r \). Note that for any \( 1 \leq j \leq N_k \),
\[
N_{j,k} = \left\lceil \frac{j \eta_{k,h}^r}{\eta_{k-1,h}^r} \right\rceil \leq \left\lceil \frac{1}{\eta_{k-1,h}^r} \right\rceil = N_{k-1},
\]
so that there are never more than \( N_k N_{k-1} \) pairs \((i,j)\) of jump-down intervals. We consider the class
\[
(10.19) \quad \mathcal{C}_{i,j,k} := \{ \varphi \in \mathcal{C}_k | d_{k-1,\varphi} \in (c_{k-1,i-1}, c_{k-1,i}], \text{ and } d_{k,\varphi} \in (c_{k,j-1}, c_{k,j}) \},
\]
where \( d_{\gamma,\varphi} \) is as in (10.7). These classes, \( \mathcal{C}_{i,j,k} \), for all \( 1 \leq j \leq N_k \) and \( 1 \leq i \leq N_{j,k} \), partition \( \mathcal{C}_k \). For now we will fix a \( j \) and an \( i \) with \( 1 \leq j \leq N_k \) and \( 1 \leq i \leq N_{j,k} \). We will consider the restrictions of \( \varphi \in \mathcal{C}_{i,j,k} \) to \([0,c_{k-1,i})\), to \((c_{k-1,i}, c_{k,j-1})\), and to \([c_{k,j-1}, c_{k,j})\). On the latter two sets
we will proceed as in the base case, using Proposition 5.2 and a by-hand construction, respectively, and then on $[0, c_{k-1,i}]$ we would like to apply the induction hypothesis. However, the restriction of $\varphi \in C_{i,j,k}$ to $[0, c_{k-1,i}]$ does not force $\varphi$ to take values in $[y_{k-1}, y_0]$. Thus, rather than simply considering the restriction of $C_{i,j,k}$ to $[0, c_{k-1,i}]$ we consider the slightly more restricted class,

$$C_{k-1,i} \subseteq C([0, c_{k-1,i}], [y_{k-1}, y_0], \eta_{k-1,i,h}),$$

whose elements take values in $[y_{k-1}, y_0]$. We can thus apply the induction hypothesis to form brackets $[l_{\alpha,i}, u_{\alpha,i}]$ for the class (10.20). We inductively constructed brackets that allowed for domains ending anywhere in $[0, c_{k-1,i}]$ but we actually are currently only bracketing functions that have domains ending in $[c_{k-1,i-1}, c_{k-1,i}]$, and ignoring the other brackets. That is, we keep only brackets $[l_{\alpha,i}, u_{\alpha,i}]$ such that the index $j_{k-1,i}$ defined in part 2 of the induction hypothesis is equal to the $i$ we fixed above, and these form a collection of brackets for $C_{k-1,i}$ by part 2(a) of the induction hypothesis.

These brackets have size

$$d_{r,k,h}(u_{\alpha,i}1{\cup}_{\gamma=1}^{c}(y_{\gamma-1} - y_\gamma)1/2 \leq \epsilon c_{k-1,i}^{1/r},$$

for $\tilde{k} \geq k - 1$, by part 3(a) of the induction hypothesis. The cardinality of this collection is given by

$$1 \leq \alpha \leq \exp \left( c \sum_{\gamma=1}^{\tilde{k}-1} \left( \frac{y_{\gamma}-y_{\gamma-1}}{\epsilon/w_{\gamma,h}} \right)^{1/2} \right.$$

$$(10.22) \left. + (1 + 1_{[\gamma \leq \tilde{k}-2]}) \log \left( 1 + \frac{1}{\epsilon^{r}/h(y_{\gamma-1})^{r}} \right) \right),$$

by (10.3).

Next, we consider the interval $(c_{k-1,i}, c_{k,j-1})$, which may be an empty interval. The following will still hold if the interval is empty. We now construct brackets on this interval, and if the interval is empty these brackets are undefined functions. The restriction of any $\varphi \in C_{i,j,k}$ to $(c_{k-1,i}, c_{k,j-1})$ lies in the class

$$C((c_{k-1,i}, c_{k,j-1}), [y_k, y_{k-1}], 0)).$$

We now apply Proposition 5.2 to get a collection of brackets, $[l_{\beta,i,j}, u_{\beta,i,j}]$ for the class (10.23). The collection has cardinality given by

$$1 \leq \beta \leq \exp \left( c \left( \frac{(y_{k-1} - y_k)}{\epsilon_{k,h}} \right)^{1/2} \right),$$

by (10.3).
and $L_r$ size given by

\begin{equation}
(10.25) \int_{c_{k-1,i}}^{c_{k,j-1}} (u_{\beta,i,j}^*(x) - l_{\beta,i,j}^*(x))^r \, dx < \epsilon_{k,h}^r (c_{k,j-1} - c_{k-1,i})^+, \end{equation}

where $(a)_+ = \max(a, 0)$ is 0 if $c_{k-1,i} \geq c_{k,j-1}$. Of course if $c_{k-1,i} \geq c_{k,j-1}$, the cardinality is also 0, but then the bound given above still (trivially) holds.

We assume the values of all the brackets lie in $[y_k, y_{k-1}]$. Note we move the factor $(c_{k,j-1} - c_{k-1,i})^{1/r}$ from the cardinality to the size of the collection.

We now use the above functions to define the brackets of interest for our functions in $C_{i,j,k}$ by

\begin{equation}
(10.26) [l_{\alpha,i,j}^*(x), u_{\alpha,i,j}^*(x)] = \begin{cases} [l_{\alpha,i}(x), u_{\alpha,i}(x)], & 0 \leq x \leq c_{k-1,i} \\ [l_{\beta,i,j}^*(x), u_{\beta,i,j}^*(x)], & c_{k-1,i} < x < c_{k,j-1} \\ [-\infty, y_{k-1}], & c_{k-1,i} \vee c_{k,j-1} < x \leq c_{k,j} \\ [-\infty, -\infty], & c_{k,j} < x \leq 1 \end{cases},
\end{equation}

for $\alpha$ and $\beta$ as in (10.22) and (10.24), respectively. Note that if $c_{k-1,i}, c_{k,j-1}$ is empty then we have not defined, nor used, $[l_{\beta,i,j}^*, u_{\beta,i,j}^*]$.

We define

$$J_{k,\alpha,\beta,i,j} := (c_{k-1,i} \vee c_{k,j-1}, c_{k,j}) = (c_{k,j-1}, c_{k,j}) \setminus [0, c_{k-1,i}],$$

which satisfies the definition given in part 3 of the induction hypothesis. We define the index, $j_{k,\alpha,\beta,i,j}$, from part (2) of the induction hypothesis to be $j$ for all of these functions, which we will now show form a collection of brackets for $C_{i,j,k}$, and thus indeed form a collection of brackets for $C_{k,j}$ once we let $i$ vary over its range. Now, given $\varphi \in C_{i,j,k}$, there exists an $\alpha$ such that $[l_{\alpha,i}, u_{\alpha,i}]$ bracket $\varphi|_{[y_k, y_{k-1}]}$ (recall the definition of $\varphi|_{I}$ given in Section 2) by the induction hypothesis part 1 since $\varphi|_{[y_k, y_{k-1}]} \in C_{k-1}$. Now, part 2(b) of the induction hypothesis guarantees that the brackets we constructed for the restriction of $\varphi$ actually bracket $\varphi$ for all $x \in [0, c_{k-1,i}]$, i.e.

$$l_{\alpha,\beta,i,j}(x) = l_{\alpha,i}(x) \leq \varphi(x) \leq u_{\alpha,i}(x) = u_{\alpha,\beta,i,j}(x).$$

Next, if $(c_{k-1,i}, c_{k,j})$ is not null then we have constructed brackets $[l_{\beta,i,j}^*, u_{\beta,i,j}^*]$ for $\varphi|_{(c_{k-1,i}, c_{k,j-1})}$, and so for $x \in (c_{k-1,i}, c_{k,j})$,

$$l_{\alpha,\beta,i,j}(x) = l_{\beta,i,j}^*(x) \leq \varphi(x) \leq u_{\beta,i,j}^*(x) = u_{\alpha,\beta,i,j}(x).$$

Now since $\varphi \in C_{i,j,k}$, for $x \in [c_{k-1,i} \vee c_{k,j-1}, c_{k,j}]$ we know $\varphi(x) \leq y_{k-1}$ so that

$$l_{\alpha,\beta,i,j}(x) = -\infty \leq \varphi(x) \leq y_{k-1} = u_{\alpha,\beta,i,j}(x).$$
Finally, since \( \varphi \in \mathcal{C}_{i,j,k} \), for \( x > c_{k,j} \) we know \( \varphi(x) = -\infty \) so that
\[
l_{\alpha,\beta,i,j}(x) = -\infty = \varphi(x) = u_{\alpha,\beta,i,j}(x).
\]

Thus we have shown that the constructed collection of functions \([l_{\alpha,\beta,i,j}, u_{\alpha,\beta,i,j}]\) for all \( \alpha \) and \( \beta \) form a collection of brackets for \( \mathcal{C}_{i,j,k} \), and considering all \( i \), these functions form a collection for \( \mathcal{C}_{k,j} \), as we needed to show for part \( 2(a) \) of the induction hypothesis. We have also thus shown part \( 1(a) \) of the induction hypothesis. We now show that part \( 2(b) \) of the induction hypothesis holds. We need to show that any \( (j,k) \)-concave extension of \( \varphi \) is bracketed by \([l_{\alpha,\beta,i,j}, u_{\alpha,\beta,i,j}]\). Fix such an extension, \( \phi \). For the brackets we constructed inductively, part \( 2(b) \) of the induction hypothesis tells us that any \((i, k - 1)\)-concave extension \( \tilde{\phi} \) of \( \varphi|_{[0,c_{k-1,i}]} \) is bracketed by \([l_{\alpha,i}, u_{\alpha,i}]\), i.e.
\[
l_{\alpha,i,j}(x) \leq \tilde{\phi}(x) \leq u_{\alpha,i,j}(x) \quad \text{for any } x \leq c_{k-1,i}.
\]
We take \( \tilde{\phi}(x) = \phi(x) \) for \( x \leq c_{k-1,i} \) (noting it may be the case that \( \phi(x) = \varphi(x) \) for \( x \leq c_{k-1,i} \), which is fine) to show that \( \phi \) is bracketed on \([0, c_{k-1,i}]\). Next, for \( c_{k-1,i} < x < c_{k,j-1} \), \( \varphi(x) = \phi(x) \) since \( \varphi(x) \geq y_{k} \), so on the interval \((c_{k-1,i}, c_{k,j-1})\) the brackets \([l_{\beta,i,j}, u_{\beta,i,j}]\) bracket \( \varphi \) (and are equal to \([l_{\alpha,i,j}, u_{\alpha,i,j}]\) on this interval). Finally, for \( x \in (c_{k-1,i} \lor c_{k,j-1}, c_{k,j}] \), since \( x > c_{k-1,i} \), \( \varphi(x) \leq y_{k-1} \) and so \( \phi(x) \leq y_{k-1} \) (since \( \varphi \), and thus \( \phi \), are nonincreasing on \([d_{1,\varphi}, \infty)\), and \( d_{1,\varphi} \leq d_{k-1,\varphi} \leq c_{k-1,i} \), so \( \phi \) is nonincreasing on \([c_{k-1,i}, \infty)\). Thus for such \( x \),
\[
l_{\alpha,\beta,i,j}(x) = -\infty \leq \phi(x) \leq y_{k-1} = u_{\alpha,\beta,i,j}(x).
\]

For \( x > c_{k,j} \), both brackets and \( \phi \), which is in \( \mathcal{C}_{k,j} \), are \(-\infty\), and so we have shown part \( 2(b) \) of the induction hypothesis.

Next we will show part \( 1(b) \) of the induction hypothesis. By \((10.22)\) and \((10.24)\), for each \((i, j)\) pair we have no more than
\[
\exp\left(c \left( \frac{y_{k-1} - y_{k}}{\epsilon/w_{k,h}} \right)^{1/2} \right) \cdot \exp\left( c \sum_{\gamma=1}^{k-1} \left( \frac{y_{\gamma-1} - y_{\gamma}}{\epsilon/w_{\gamma,h}} \right)^{1/2} \right) \]
\[
+ (1 + 1_{[\gamma \leq k-2]}) \log \left( 1 + \frac{1}{\epsilon/h(y_{\gamma-1})^{1/2}} \right) \right) \text{ \quad (10.27)}
\]
brackets. There are no more than \((1/\eta_{k-1,h}^r) + 1 \) \( i \)'s for each \( j \), and no more than \((1/\eta_{k,h}^r) + 1 \) \( j \)'s. Thus the total number of brackets we have formed is
\[
\left( \frac{1}{\eta_{k-1,h}^r} + 1 \right) \left( \frac{1}{\eta_{k,h}^r} + 1 \right) = \left( \frac{1}{\epsilon/h(y_{k-2})^{1/2}} + 1 \right) \left( \frac{1}{\epsilon/h(y_{k-1})^{1/2}} + 1 \right)
\]
times (10.27), or (10.28)
\[
\exp\left(c \sum_{\gamma=1}^{k} \left\{ \left( \frac{(y_{\gamma-1} - y_{\gamma})}{\epsilon e^{-y_{\gamma-1}}} \right)^{1/2} + \left( 1 + \frac{1}{\gamma_{\gamma-1}^{1/2}} \right) \log \left( 1 + \frac{1}{\epsilon r/h(y_{\gamma-1})^{\gamma/2}} \right) \right\} \right),
\]
as desired.

We now show part 3 of the induction hypothesis. For brevity in the next display, we will refer to \(J_{\gamma;\alpha,\beta,i,j}\) just as \(J_{\gamma}\). For \(k \geq k\), we have
\[
d^r_{r,k,h} \left( u_{\alpha,\beta,i,j} 1_{(i_{\gamma=1}^{k} J_{\gamma})^c}, l_{\alpha,\beta,i,j} 1_{(i_{\gamma=1}^{k} J_{\gamma})^c} \right)
= \int_{0}^{1} 1_{(i_{\gamma=1}^{k} J_{\gamma})^c}(x) \left( \int_{k_{\alpha,\beta,i,j}(x)}^{u_{\alpha,\beta,i,j}(x)} dW_{k,h}(z) \right)^r dx
= \int_{0}^{f_{k-1,i}\cdots f_{k,j-1}} 1_{(i_{\gamma=1}^{k} J_{\gamma})^c}(x) \left( \int_{k_{\alpha,\beta,i,j}(x)}^{u_{\alpha,\beta,i,j}(x)} dW_{k,h}(z) \right)^r dx
+ 1_{1_{k-1,i} < c_{k,j-1}} \int_{1_{k-1,i} < c_{k,j-1}}^{f_{k,j-1}} 1_{(i_{\gamma=1}^{k} J_{\gamma})^c}(x) \left( \int_{k_{\alpha,\beta,i,j}(x)}^{u_{\alpha,\beta,i,j}(x)} dW_{k,h}(z) \right)^r dx
= d^r_{r,k,h} \left( u_{\alpha,i} 1_{(i_{\gamma=1}^{k} J_{\gamma})^c}, l_{\alpha,i} 1_{(i_{\gamma=1}^{k} J_{\gamma})^c} \right)
+ 1_{1_{k-1,i} < c_{k,j-1}} w^r_{k,h} \int_{1_{k-1,i} < c_{k,j-1}}^{f_{k,j-1}} \left( \int_{k_{\alpha,\beta,i,j}(x)}^{u_{\alpha,\beta,i,j}(x)} dW_{k,h}(z) \right)^r dx,
\]
where the second equality follows from the definition of \(J_{k} = (c_{k-1,i} \cup c_{k,j-1}, c_{k,j})\) and the third is because \(W_{k,h}\) assigns mass \(w_{k,h}\) to the interval \([y_k, y_{k-1}]\), which is where \(l_{\alpha,\beta,i,j}(x)\) and \(u_{\alpha,\beta,i,j}(x)\) lie for \(x \in (c_{k-1,i}, c_{k,j-1})\). Then, by (10.21) (and by (10.25) and the definitions of \(w_{k,h}\) and \(\epsilon_{r,k,h}\)) the above display is bounded above by
\[
(10.29) \quad \epsilon^r c_{k-1,i} + w^r_{k,h} \frac{\epsilon^r}{w^r_{k,h}} (c_{k,j-1} - c_{k-1,i} )_+ = \epsilon^r (c_{k-1,i} \cup c_{k,j-1}) \leq \epsilon^r c_{k,j},
\]
as desired. We can also immediately conclude, letting \(A_k := (\bigcup_{\gamma=1}^{k} J_{\gamma})^c\), that
\[
(10.30) \quad \| (h \circ u_{\alpha,\beta,i,j} - h \circ l_{\alpha,\beta,i,j} 1_{A_k} ) \|_r \leq d_{r,k,h}(u_{\alpha,\beta,i,j} 1_{A_k}, l_{\alpha,\beta,i,j} 1_{A_k}) \leq \epsilon
\]
by (10.29) (and the fact that \(c_{k,j} \leq 1\)), where the first inequality follows from Lemma 6.1 and the fact that \(u_{\alpha,\beta,i,j}(x) 1_{A_k}(x)\) and \(l_{\alpha,\beta,i,j}(x) 1_{A_k}(x)\) lie in \([y_k, y_0]\) when they are not equal to each other.
We next control the size of the bracket on the complementary set, i.e. we will show (10.5). We again use $J_\gamma$ to signify $J_{\gamma,\alpha,\beta,i,j}$. We have

\begin{equation}
\| (h \circ u_{\alpha,\beta,i,j} - h \circ l_{\alpha,\beta,i,j}) 1_{\bigcup_{\gamma=1}^{J_\gamma}} \|_r^r \\
\leq \int_{\bigcup_{\gamma=1}^{J_\gamma}} (h(u_{\alpha,\beta,i,j}(x)) - h(l_{\alpha,\beta,i,j}(x)))^r \, dx \\
+ \int_{J_k} h(u_{\alpha,\beta,i,j}(x))^r \, dx \\
\leq \epsilon^r \sum_{\gamma=1}^{k-1} h(y_{\gamma-1})^{r(1-\zeta)} + \int_{J_k} h(y_{k-1})^r \, dx \\
\leq \epsilon^r \sum_{\gamma=1}^{k-1} h(y_{\gamma-1})^{r(1-\zeta)} + \epsilon^r h(y_{k-1})^{r(1-\zeta)} \\
= \epsilon^r \sum_{\gamma=1}^{k} h(y_{\gamma-1})^{r(1-\zeta)},
\end{equation}

where the inequality (10.32) is by the induction hypothesis (10.5) and because we chose $u_{\alpha,\beta,i,j}$ on $J_k$ to be below $y_{k-1}$. The last inequality, (10.33), is because the length of $J_k$ is less than $\eta_{k,h}^r = \epsilon^r / h(y_{k-1})^{r\zeta}$.

We have thus shown (10.5) for $k^* = k$, and so have shown part 3 of the induction hypothesis.

Now we move on to bounding $\| h \circ u_{\alpha,\beta,i,j} - h \circ l_{\alpha,\beta,i,j} \|_r^r$ and thus showing (10.6). By (10.30) and by (10.34), we have

\begin{equation}
\| (h \circ u_{\alpha,\beta,i,j} - h \circ l_{\alpha,\beta,i,j}) (1_{\bigcup_{\gamma=1}^{J_\gamma}}) \|_r^r \\
\leq \epsilon^r + \epsilon^r \sum_{\gamma=1}^{k} h(y_{\gamma-1})^{r(1-\zeta)},
\end{equation}

as desired. Since we have already found the cardinality of this bracketing of $F_{k^*,h}$, we have shown (10.6) for $k^* = k$, and are thus done with our induction and the proof of Proposition 10.1.

The condition (10.1) is not fundamental. It is essentially keeping $\epsilon$ from being too large, which is an unimportant constraint. The proposition could be phrased without this condition, but we phrase it with the condition and then pick $y_k$ sequences later that satisfy it.

**Acknowledgements.** Thanks go to Arseni Seregin for several helpful suggestions and to Bodhi Sen and Fadoua Balabdaoui for comments and minor corrections. We also owe thanks to Tilmann Gneiting for supporting
the visits of both authors to the Applied Mathematics Institute at the University of Heidelberg during the Autumn of 2011 and Autumn - Spring of 2011-2012 respectively.

References.

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

Birgé, L. (1997). Estimation of unimodal densities without smoothness assumptions. *Ann. Statist.* **25** 970–981. MR1447736 (98g:62056)

Birgé, L. and Massart, P. (1993). Rates of convergence for minimum contrast estimators. *Probab. Theory Related Fields* **97** 113–150. MR1240719 (94m:62095)

Bronšteĭn, E. M. (1976). ε-entropy of convex sets and functions. *Sibirsk. Mat. Ž.* **17** 508–514, 715. MR0415155 (54 ##3246)

Cule, M. and Samworth, R. (2010). Theoretical properties of the log-concave maximum likelihood estimator of a multidimensionsional density. *Electron. J. Stat.* **4** 254–270. MR2645484 (2012f:62071)

Devroye, L. (1984). A simple algorithm for generating random variates with a log-concave density. *Computing* **33** 247–257. MR773927 (86d:65019)

Dharmadhikari, S. and Joag-Dev, K. (1988). *Unimodality, convexity, and applications. Probability and Mathematical Statistics*. Academic Press Inc., Boston, MA. MR954608 (89k:60020)

Doss, C. R. (2013). Shape-Constrained Inference for Concave-Transformed Densities and their Modes PhD thesis, Department of Statistics, University of Washington, in preparation.

Doss, C. R. and Wellner, J. A. (2013). Inference for the mode of a log-concave density Technical Report, University of Washington. in preparation.

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. MR876079 (88e:60029)

Dudley, R. M. (1999). *Uniform central limit theorems. Cambridge Studies in Advanced Mathematics* **63**. Cambridge University Press, Cambridge. MR1720712 (2000k:60050)

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 (2011b:62096)

Guntuboyina, A. and Sen, B. (2013). Covering numbers for convex functions. *IEEE Trans. Inform. Theor.* **59** 1957–1965.

Pal, J. K., Woodroofe, M. and Meyer, M. (2007). Estimating a Polya frequency function2. In *Complex datasets and inverse problems. IMS Lecture Notes Monogr. Ser.* **54** 239–249. Inst. Math. Statist., Beachwood, OH. MR2459192 (2010d:62087)

Rockafellar, R. T. (1970). *Convex analysis. Princeton Mathematical Series, No. 28*. Princeton University Press, Princeton, N.J. MR0274683 (43 #1445)

Schoenberg, I. J. (1951). On Pólya frequency functions. I. The totally positive functions and their Laplace transforms. *J. Analyse Math.* **1** 331–374. MR0047732 (13,923c)

Seregin, A. and Wellner, J. A. (2010). Nonparametric estimation of multivariate convex-transformed densities. *Ann. Statist.* **38** 3751–3781. With supplementary material available online. MR2766867 (2012b:62126)
Shorack, G. R. and Wellner, J. A. (1986). *Empirical processes with applications to statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York. MR838963 (88e:60002)

van de Geer, S. (1993). Hellinger-consistency of certain nonparametric maximum likelihood estimators. *Ann. Statist.* 21 14–44. MR1212164 (94c:62062)

van de Geer, S. A. (2000). *Applications of Empirical Process Theory*. Cambridge Series in Statistical and Probabilistic Mathematics 6. Cambridge University Press, Cambridge. MR1739079 (2001h:62002)

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York. With applications to statistics. MR1385671 (97g:60035)

Walther, G. (2002). Detecting the presence of mixing with multiscale maximum likelihood. *J. Amer. Statist. Assoc.* 97 508–513. MR1941467 (2003k:62152)

Wong, W. H. and Shen, X. (1995). Probability inequalities for likelihood ratios and convergence rates of sieve MLEs. *Ann. Statist.* 23 339–362. MR1332570 (97a:62096)