SUPPLEMENTARY MATERIAL TO ‘ADAPTATION IN LOG-CONCAVE DENSITY ESTIMATION’

BY ARLENE K. H. KIM*,§, ADITYANAND GUNTUBOYINA†,¶ and RICHARD J. SAMWORTH‡,∥

University of Cambridge and Sungshin Women’s University§
University of California, Berkeley¶
University of Cambridge∥

This is the supplementary material to Kim, Guntuboyina and Samworth (2017), hereafter referred to as the main text.

1. Auxiliary result from Section 2 of the main text.

Lemma 1. Let $g : [a, b] \to (-\infty, \infty]$ be convex with $g(r) = 0$ for some $r \in [a, b]$. For $\alpha, \beta, c \in \mathbb{R}$, define

$$G(x) := c + \int_{a}^{x} \exp(\alpha t + \beta) g(t) \, dt \quad \text{for } x \in [a, b].$$

Assume $\alpha \neq 0$. If $r \in (a, b]$, then

$$\inf_{x \in [a, r]} \frac{G(x) - G(r)}{1 - e^{-\alpha (r-x)} \{\alpha (r-x) + 1\}} = \frac{G(a) - G(r)}{1 - e^{-\alpha (r-a)} \{\alpha (r-a) + 1\}}$$

and if $r \in [a, b)$

$$\sup_{x \in (r, b]} \frac{G(x) - G(r)}{1 + e^{\alpha (x-r)} \{\alpha (x-r) - 1\}} = \frac{G(b) - G(r)}{1 + e^{\alpha (b-r)} \{\alpha (b-r) - 1\}}.$$  

Now assume $\alpha = 0$. If $r \in (a, b]$, then

$$\inf_{x \in [a, r]} \frac{G(x) - G(r)}{(r-x)^2} = \frac{G(a) - G(r)}{(r-a)^2}$$

and if $r \in [a, b)$, then

$$\sup_{x \in [r, b]} \frac{G(x) - G(r)}{(x-r)^2} = \frac{G(b) - G(r)}{(b-r)^2}.$$
Proof. Assume \( \alpha \neq 0 \) and \( r \in (a, b) \) and consider the linear function
\[
\bar{g}(x) := \frac{\alpha^2 \{G(r) - G(a)\}}{e^{\alpha r + \beta} - e^{\alpha a + \beta} \{\alpha (r - a) + 1\}} (r - x).
\]
Note here that the denominator does not vanish, because \( 1 - e^{-y}(1 + y) > 0 \) for \( y \neq 0 \). Thus
\[
(5) \quad \bar{g}(r) = 0 = g(r) \quad \text{and} \quad \int_a^r \exp(\alpha x + \beta) \bar{g}(x) \, dx = \int_a^r \exp(\alpha x + \beta) g(x) \, dx.
\]
Now the function \( x \mapsto g(x) - \bar{g}(x) \), which is convex on \([a, r]\) and 0 at \( x = r \), can change sign at most once in the interval \([a, r]\). But we deduce from the second part of (5) that either this function is zero for all \( x \in (a, r) \) or it changes sign exactly once in \((a, r)\). In particular, there exists \( x_0 \in (a, r) \) such that \( g(x) \geq \bar{g}(x) \) for \( x \in [a, x_0] \) and \( g(x) \leq \bar{g}(x) \) for \( x \in [x_0, r] \). This further implies that
\[
\int_a^x \exp(\alpha t + \beta) \{g(t) - \bar{g}(t)\} \, dt = -\int_x^r \exp(\alpha t + \beta) \{g(t) - \bar{g}(t)\} \, dt \geq 0
\]
for every \( x \in [a, r] \). Consequently, for \( x \in [a, r] \),
\[
G(x) = G(r) - \int_x^r \exp(\alpha t + \beta) g(t) \, dt \\
\geq G(r) - \int_x^r \exp(\alpha t + \beta) \bar{g}(t) \, dt \\
= G(r) - \frac{1 - e^{-\alpha (r-x)}}{1 - e^{-\alpha (r-a)}} \left\{ \frac{\alpha (r-x) + 1}{\alpha (r-a) + 1} \right\} \{G(r) - G(a)\}.
\]
This yields (1), and the proof of (2) is very similar. The proofs of (3) and (4) then follow by taking limits as \( \alpha \to 0 \) and using the fact that
\[
\lim_{\alpha \to 0} \frac{1 - e^{-\alpha y}(\alpha y + 1)}{\alpha^2} = \frac{y^2}{2} \quad \text{for every} \ y \in \mathbb{R}.
\]

2. Auxiliary results from Section 3 of the main text.

2.1. Auxiliary results for the proof of Theorem 3 in the main text.

Lemma 2. There exists a universal constant \( C > 0 \) such that for every \( n \geq 2 \), we have
\[
(6) \quad \sup_{f_0 \in F} \mathbb{E}_{f_0} \left\{ \sup_{x \in \mathbb{R}} \log \hat{f}_n(x) + \sup_{x \in [X_{(1)}, X_{(n)}]} \log \frac{1}{f_0(x)} \right\} \leq C \log n.
\]
Proof. For $\mu \in \mathbb{R}$, $\sigma > 0$ and $i = 1, \ldots, n$, let $Y_i := \sigma X_i + \mu$, so $Y_i$ has density $g_0(y) := \sigma^{-1} f_0((y - \mu)/\sigma)$. By affine equivariance, the log-concave maximum likelihood estimator based on $Y_1, \ldots, Y_n$ is $\hat{g}_n(y) := \sigma^{-1} \hat{f}_n((y - \mu)/\sigma)$. Moreover, writing $X_{(1)} := \min_i X_i$ and $X_{(n)} := \max_i X_i$, we have

$$Y_{(1)} := \min_i Y_i = \sigma X_{(1)} + \mu \quad \text{and} \quad Y_{(n)} := \max_i Y_i = \sigma X_{(n)} + \mu.$$ 

Thus

$$\sup_{y \in \mathbb{R}} \log \hat{g}_n(y) = \sup_{x \in \mathbb{R}} \log \hat{f}_n(x) - \log \sigma$$

and

$$\sup_{y \in [Y_{(1)}, Y_{(n)}]} \log \frac{g_0(y)}{\sigma} = \sup_{x \in [X_{(1)}, X_{(n)}]} \log \frac{1}{f_0(x)} + \log \sigma.$$

The left-hand side of (6) is therefore affine invariant, and there is no loss of generality in assuming that $\mu f_0 = 0$ and $\sigma^2 f_0 = 1$. Let $\mathcal{P}$ denote the class of probability distributions $P$ on $\mathbb{R}$ for which $\int_{-\infty}^{\infty} |x| dP(x) < \infty$ and $P$ is not a Dirac point mass. We recall from Dümbgen, Samworth and Schuhmacher (2011, Theorem 2.2) that there is a well-defined projection $\psi^* : \mathcal{P} \to \mathcal{F}$ given by

$$\psi^*(P) := \arg\max_{f \in \mathcal{F}} \int_{-\infty}^{\infty} \log f dP.$$

Now, for $\sigma > 0$, let $\mathcal{P}_{\geq \sigma}$ denote the subset of $\mathcal{P}$ consisting distributions $P$ on the real line with $\int_{-\infty}^{\infty} (x - \mu_P)^2 dP(x) \geq \sigma^2$, where $\mu_P := \int_{-\infty}^{\infty} x dP(x)$. By a very similar argument to that given in the proof of Lemma 6 of Kim and Samworth (2016),

$$\sup_{P \in \mathcal{P}_{\geq \sigma}} \sup_{x \in \mathbb{R}} \psi^*(P)(x) \leq \frac{C}{\sigma}.$$

As $\hat{f}_n = \psi^*(\mathbb{P}_n)$, where $\mathbb{P}_n$ denotes the empirical distribution of $X_1, \ldots, X_n$, we have for $t > 0$

$$\mathbb{P} \left( \sup_{x \in \mathbb{R}} \log \hat{f}_n(x) > \frac{t}{2} \log n \right) \leq \mathbb{P} \left( \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 < \frac{C}{n^{1/2}} \right)$$

$$\leq \mathbb{P} \left( |X_1 - \bar{X}| < \frac{C^{1/2}}{n^{1/4-1/2}} \right),$$

where $\bar{X} := n^{-1} \sum_{i=1}^{n} X_i$. But $X_1 - \bar{X}$ has mean 0, variance $1 - 1/n$ and has a log-concave density (which is therefore bounded by a universal constant). Hence

$$\sup_{x \in \mathbb{R}} \log \hat{f}_n(x) > \frac{t}{2} \log n \leq \frac{C}{n^{1/4-1/2}}.$$

(7)
Now let $F_0$ denote the distribution function corresponding to $f_0$ and for $t \geq 2$ let

$$
\Omega_t := \{X_{(1)} \geq F_0^{-1}(n^{-t/2}/\alpha)\} \cap \{X_{(n)} \leq F_0^{-1}(1 - n^{-t/2}/\alpha)\},
$$

where $\alpha > 0$ is taken from Lemma 3 below. Then by a union bound,

$$
\sup_{f_0 \in F_0} P_{f_0}(\Omega_t) \leq \frac{2}{\alpha n^{t/2-1}}.
$$

Moreover, on $\Omega_t$,

$$
\sup_{x \in [X_{(1)}, X_{(n)}]} \log \frac{1}{f_0(x)} \leq \max\left\{ \log \frac{1}{f_0(F_0^{-1}(n^{-t/2}/\alpha))}, \log \frac{1}{f_0(F_0^{-1}(1 - n^{-t/2}/\alpha))} \right\}
= \frac{t}{2} \log n
$$

(9)

where the equality holds because the minimum of a concave function on a compact interval is attained at one of the endpoints of the interval, and the second inequality holds due to Lemma 3 below. It follows from (7), (8) and (9) that for $t \geq 2$,

$$
P\left( \sup_{x \in \mathbb{R}} \log \hat{f}_n(x) + \sup_{x \in [X_{(1)}, X_{(n)}]} \log \frac{1}{f_0(x)} > t \log n \right)
\leq P\left( \sup_{x \in \mathbb{R}} \log \hat{f}_n(x) > \frac{t}{2} \log n \right) + P\left( \sup_{x \in [X_{(1)}, X_{(n)}]} \log \frac{1}{f_0(x)} > \frac{t}{2} \log n \right)
\leq \frac{C}{n^{t/4-1/2}} + \frac{2}{\alpha n^{t/2-1}},
$$

(10)

and the result follows.

The following result is a small generalisation of Proposition A.1(c) of Bobkov (1996).

**Lemma 3.** There exists $\alpha > 0$ such that for all $p \in (0, 1)$ and all $f_0 \in \mathcal{F}_{0,1}$ with corresponding distribution function $F_0$,

$$
f_0(F_0^{-1}(p)) \geq \alpha \min(p, 1 - p).
$$
Proof. Fix \( f_0 \in \mathcal{F}^{0,1} \) with associated distribution function \( F_0 \). Proposition A.1(c) of Bobkov (1996) gives that \( p \mapsto f_0(F_0^{-1}(p)) \) is positive and concave on \((0,1)\). But, by Theorem 2(b) of Kim and Samworth (2016), there exists \( \alpha > 0 \) (not depending on \( f_0 \)) such that
\[
f_0(0) \geq \alpha.
\]
Noting that \( F_0(0) \in (0,1) \), we deduce by concavity that for \( p \in (0,F_0(0)) \),
\[
f_0(F_0^{-1}(p)) \geq \frac{p}{F_0(0)} \alpha \geq \alpha p \geq \alpha \min(p,1-p).
\]
A very similar argument handles the case \( p \in (F_0(0),1) \), and this concludes the proof.

2.2. Auxiliary results for the proof of Theorem 4 in the main text. Recall that we can write \( \mathcal{F}^1 = \{f_{\alpha,s_1,s_2} : (\alpha,s_1,s_2) \in \mathcal{T} \} \).

Lemma 4. If \( X \sim f_{\alpha,s_1,s_2} \in \mathcal{F}^1 \), then there exist \( \alpha \neq 0 \) and \( b \in \mathbb{R} \) such that \( aX + b \) has a density \( f_0 \in \mathcal{F}^1 \) of one of the following three forms:

1. \( f_0 = f_{0,0,1} \);
2. \( f_0 = f_{-\alpha_0,0,1} \) for some \( \alpha_0 \in (0,18) \);
3. \( f_0 = f_{-1,0,s_0} \) for some \( s_0 \in [18,\infty] \).

Proof. Let \( X \sim f_{\alpha,s_1,s_2} \in \mathcal{F}^1 \) for some \( (\alpha,s_1,s_2) \in \mathcal{T} \), and let \( \alpha \neq 0 \) and \( b \in \mathbb{R} \). Then
\[
aX + b \sim \begin{cases} f_{\alpha/a,as_1+b,as_2+b} & \text{if } a > 0 \\ f_{\alpha/a,as_2+b,as_1+b} & \text{if } a < 0. \end{cases}
\]
Thus, if \( \alpha = 0 \), we can set \( a = (s_2-s_1)^{-1}, b = -s_1(s_2-s_1)^{-1} \) so that \( aX + b \sim f_{0,0,1} \). If \( \alpha > 0 \) and \( \alpha(s_2-s_1) < 18 \), then we can set \( a = -(s_2-s_1)^{-1}, b = s_2(s_2-s_1)^{-1} \) while if \( \alpha < 0 \) and \( |\alpha|(s_2-s_1) < 18 \) then we can set \( a = (s_2-s_1)^{-1}, b = -s_1(s_2-s_1)^{-1} \); in either situation, \( aX + b \sim f_{-\alpha_0,0,1} \), with \( \alpha_0 := |\alpha|(s_2-s_1) \in (0,18) \). Finally, if \( \alpha > 0 \) and \( \alpha(s_2-s_1) \in [18,\infty] \), then we can set \( a = -\alpha, b = \alpha s_2 \) while if \( \alpha < 0 \) and \( |\alpha|(s_2-s_1) \in [18,\infty] \) then we can set \( a = -\alpha, b = \alpha s_1 \); in either situation, \( aX + b \sim f_{-1,0,s_0} \) with \( s_0 := |\alpha|(s_2-s_1) \).

Lemma 5. Let \( \phi : \mathbb{R} \to [-\infty,\infty) \) be a concave function whose domain is contained in \([0,1]\) and which satisfies
\[
(11) \quad \int_0^1 (e^{\phi(u)/2} - 1)^2 \, du \leq \delta^2
\]
for some \( \delta \in (0, 2^{-5/2}] \). Then

\[
\phi(x) \leq 2^{13/2}\delta \quad \text{for every } x \in [0, 1].
\]

Moreover,

\[
\phi(x) \geq \frac{-4\delta}{\{\min(x, 1-x)\}^{1/2}} \quad \text{when } \min(x, 1-x) \geq 4\delta^2.
\]

**Proof.** We first prove inequality (12). By symmetry, it suffices to prove that \( \phi(x) \leq 2^{13/2}\delta \) for all \( x \in [0, 1/2] \). Fix \( x \in [0, 1/2] \) and assume that \( \phi(x) > 0 \), for otherwise there is nothing to prove. Let \( x_\ast \in (x, 1] \) be such that \( \phi(x_\ast) = 0 \) if such an \( x_\ast \) exists; otherwise, set \( x_\ast = 1 \).

We first consider the case \( x_\ast \geq \frac{3}{4} \). Since \( e^x \geq 1 + x \) and \( \phi \) is a concave function with \( \phi(x_\ast) \geq 0 \),

\[
\delta^2 \geq \int_x^{x_\ast} \left( e^{\phi(u)/2} - 1 \right)^2 \, du \geq \frac{1}{4} \int_x^{x_\ast} \phi^2(u) \, du \geq \frac{\phi^2(x)}{4} \int_x^{x_\ast} \left( \frac{x_\ast - u}{x_\ast - x} \right)^2 \, du
\]

\[
= \frac{x_\ast - x}{12} \phi^2(x) \geq \frac{\phi^2(x)}{48},
\]

so \( \phi(x) \leq 4\sqrt{3}\delta \).

Now suppose instead that \( x_\ast < \frac{3}{4} \), so that \( \phi(x_\ast) = 0 \). Then for \( u \in [\frac{7}{8}, 1] \),

\[
\phi(u) \leq -\frac{u - x_\ast}{x_\ast - x} \phi(x) \leq -\frac{\phi(x)}{8}.
\]

We deduce that

\[
\delta^2 \geq \int_{\frac{7}{8}}^1 (1 - e^{\phi(u)/2})^2 \, du \geq \frac{1}{8} (1 - e^{-\phi(x)/16})^2,
\]

so

\[
\phi(x) \leq 16 \log \left( \frac{1}{1 - 2^{3/2}\delta} \right) \leq \frac{2^{11/2}\delta}{1 - 2^{3/2}\delta} \leq 2^{13/2}\delta,
\]

since \( \delta \in (0, 2^{-5/2}] \). This completes the proof of (12).

We now proceed to prove inequality (13), and by symmetry it suffices to consider a fixed \( x \in [4\delta^2, 1/2] \). We assume that \( \phi(x) < 0 \), because otherwise there is nothing to prove. By concavity of \( \phi \), we have either \( \phi(u) \leq \phi(x) \) for all \( u \in [0, x] \) or \( \phi(u) \leq \phi(x) \) for all \( u \in [x, 1] \). In the former case,

\[
\delta^2 \geq \int_0^x (1 - e^{\phi(u)/2})^2 \, du \geq x (1 - e^{\phi(x)/2})^2.
\]
Thus
\[
\phi(x) \geq 2 \log \left( 1 - \frac{\delta}{x^{1/2}} \right) \geq \frac{-4\delta}{x^{1/2}}.
\]

In the latter case, where \( \phi(u) \leq \phi(x) \) for all \( u \in [x, 1] \), we find
\[
\delta^2 \geq \int_x^1 (1 - e^{\phi(u)/2})^2 \, du \geq (1 - x)(1 - e^{\phi(x)/2})^2 \geq x(1 - e^{\phi(x)/2})^2,
\]
and the conclusion follows as before.

**Lemma 6.** Let \( f_0 = f_{-1,0,a} \in \mathcal{F}^1 \) for some \( a \in [18, \infty) \), and let \( \phi : \mathbb{R} \to (-\infty, \infty) \) be a concave function whose domain is contained in \([0, a]\) and which satisfies
\[
\int_0^a \{ e^{\phi(u)/2} - f_0^{1/2}(u) \}^2 \, du \leq \delta^2
\]
for some \( \delta \in (0, e^{-9/8}) \). Let
\[
x_0 := \min \left\{ \log \frac{1}{2^6 e^{52}(1 - e^{-a})}, a - 1 \right\} \geq 17.
\]
Then with \( \tilde{\phi}_a \) defined as in (31) in the main text, we have
\[
-4 e^{x/2}(1 - e^{-a})^{1/2} \delta \leq \tilde{\phi}_a(x) \leq 2^{13/2} e^{x/2}(1 - e^{-a})^{1/2} \delta
\]
for every \( x \in [1, x_0] \), and
\[
\tilde{\phi}_a(x) \leq 8 \frac{x - x_0}{x_0 - 1} + 7
\]
for every \( x \in [x_0, a] \).

**Proof.** Fix \( f_0 = f_{-1,0,a} \) for some \( a \in [18, \infty) \), and fix \( \delta \in (0, e^{-9/8}) \) and \( \phi \) satisfying the conditions of the lemma. For ease of notation, let us denote \( \tilde{\phi}_a \) by \( \psi \). We first prove the lower bound for \( \psi \) in (16). Fix \( x \in [1, x_0] \) and assume that \( \psi(x) < 0 \) because otherwise there is nothing to prove. By concavity of \( \psi \), the inequality \( \psi(u) \leq \psi(x) \) is true either for all \( u \in [0, x] \) or for all \( u \in [x, a] \). In the former case,
\[
\delta^2 \geq \int_0^x \{ e^{\phi(u)/2} - f_0^{1/2}(u) \}^2 \, du = \int_0^x (1 - e^{\psi(u)/2})^2 \frac{e^{-u}}{1 - e^{-a}} \, du
\]
\[
\geq (1 - e^{\psi(x)/2})^2 \frac{e^{-x}}{1 - e^{-a}} \geq (1 - e^{\psi(x)/2})^2 \frac{e^{-x}(e - 1)}{1 - e^{-a}},
\]
where we used the fact that $x \geq 1$ in the final inequality. Similarly in the latter case, we can consider the integral from $x$ to $a$ instead to obtain

$$
\delta^2 \geq (1 - e^{\psi(x)/2})^2 \frac{e^{-x} - e^{-a}}{1 - e^{-a}} \geq (1 - e^{\psi(x)/2})^2 \frac{e^{-x}(1 - e^{-1})}{1 - e^{-a}},
$$

where we used the fact that $x \leq a - 1$ for the final inequality. Now

$$
\frac{e^{x/2}(1 - e^{-a})^{1/2}}{(1 - e^{-1})^{1/2}} \delta \leq \frac{e^{x_0/2}(1 - e^{-a})^{1/2}}{(1 - e^{-1})^{1/2}} \delta \leq \frac{1}{2},
$$

and we deduce from (18) and (19) that

$$
\psi(x) \geq 2 \log \left(1 - \frac{e^{x/2}(1 - e^{-a})^{1/2}}{(1 - e^{-1})^{1/2}} \delta \right) \geq -\frac{4e^{x/2}(1 - e^{-a})^{1/2}}{(1 - e^{-1})^{1/2}} \delta,
$$

as required.

We next prove the upper bound in (16). To this end, again fix $x \in [1, x_0]$ and note by very similar arguments to those above that

$$
\delta^2 \geq \int_{x-1}^{x} \left( e^{\psi(u)/2} - 1 \right)^2 \frac{e^{-u}}{1 - e^{-a}} \, du \geq \frac{e^{-x}}{1 - e^{-a}} \int_{0}^{1} \left( e^{\psi(u+x-1)/2} - 1 \right)^2 \, du.
$$

Now

$$
e^{x/2}(1 - e^{-a})^{1/2} \delta \leq e^{x_0/2}(1 - e^{-a})^{1/2} \delta \leq \frac{1}{8e^{1/2}} \leq 2^{-5/2},$$

so the result follows by (12) in Lemma 5.

Finally, we prove (17). Fix $x \in [x_0, a]$. Inequality (16) gives

$$
\psi(x_0) \leq 2^{13/2} e^{x_0/2}(1 - e^{-a})^{1/2} \delta \leq 2^{7/2} e^{-1/2}
$$

and also that

$$
\psi(1) \geq -\frac{4e^{1/2}(1 - e^{-a})^{1/2}}{e^{1/2}} \delta \geq -\frac{e^{1/2}}{2e^{9}(1 - e^{-1})^{1/2}} \geq -\frac{1}{2}.
$$

It therefore follows by concavity of $\psi$ that

$$
\psi(x) \leq \frac{x - x_0}{x_0 - 1} \psi(x_0) + \psi(1) \leq 8 \frac{x - x_0}{x_0 - 1} + 7,
$$

as required. 

In order to prove Theorem 4 for these three cases, we need to prove two results on the bracketing numbers of log-concave functions on bounded subintervals of $\mathbb{R}$. For $a < b$ and $-\infty \leq B_1 \leq B_2 < \infty$, let $\mathcal{F}([a, b], B_1, B_2)$ denote the class of all non-negative functions $f$ on $[a, b]$ such that $\log f$ is concave and such that $B_1 \leq \log f(x) \leq B_2$ for every $x \in [a, b]$. 

Proposition 7. There exists a universal constant $C > 0$ such that

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

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

Proof. Fix $\epsilon > 0$, $a < b$ and $B_1 \leq B_2$, and let $\delta := 2\epsilon e^{-B_2/2}$. By Kim and Samworth (2016b, Proposition 4) (see also Guntuboyina and Sen (2015); Doss and Wellner (2016)), there exists a bracketing set $\{[\phi_{L,j}, \phi_{U,j}] : j = 1, \ldots, M\}$ for the set of concave functions on $[a, b]$ that are bounded below by $B_1$ and above by $B_2$ with $\int_a^b (\phi_{U,j} - \phi_{L,j})^2 dx \leq \delta^2$ and

$$\log M \leq C \left\{ \frac{(b - a)^{1/2}(B_2 - B_1)}{\delta} \right\}^{1/2}.
$$

Now take $f_{L,j} := e^{\phi_{L,j}}$ and $f_{U,j} := e^{\phi_{U,j}}$ for $j = 1, \ldots, M$. Since there is no loss of generality in assuming $\phi_{U,j}(x) \leq B_2$ for every $j \in \{1, \ldots, M\}$ and $x \in [a, b]$, we have

$$\int_a^b (f_{U,j}^{1/2} - f_{L,j}^{1/2})^2 = \int_a^b e^{\phi_{U,j}} \left\{ 1 - e^{-(\phi_{U,j} - \phi_{L,j})/2} \right\}^2 \leq \frac{e^{B_2/4}}{4} \int_a^b (\phi_{U,j} - \phi_{L,j})^2
$$

$$\leq \frac{\delta^2}{4} e^{B_2} = \epsilon^2.
$$

The result follows.

For $B_1 = -\infty$, Proposition 7 unfortunately gives the trivial upper bound $H(\epsilon, \mathcal{F}([a, b], -\infty, B_2), d_H, [a, b]) \leq \infty$. It turns out however that this quantity is actually finite, as shown by the following result, essentially due to Doss and Wellner (2016, Theorem 4.1).

Proposition 8. There exists a universal constant $C > 0$ such that

$$H(\epsilon, \mathcal{F}([a, b], -\infty, B), d_H, [a, b]) \leq Ce^{B/4}(b - a)^{1/4}/\epsilon^{1/2}
$$

for every $\epsilon > 0$, $a < b$ and $B \in \mathbb{R}$.

*In fact, formally, only the case $B_1 = -B_2$ is covered by Kim and Samworth (2016b, Proposition 4), but the proof proceeds by first considering the case $B_1 = -1, B_2 = 1$, so a simple scaling argument can be used to obtain the claimed result.
Proof. First note that
\[ \{ f^{1/2} : f \in \mathcal{F}([a, b], -\infty, B) \} \subseteq \mathcal{F}(\mathbb{R}, -\infty, B/2). \]
Thus
\begin{align*}
H_{[]} (\epsilon, \mathcal{F}([a, b], -\infty, B), d_{H}, [a, b]) &= H_{[]} (\epsilon, \{ f^{1/2} : f \in \mathcal{F}([a, b], -\infty, B) \}, L_2, [a, b]) \\
& \leq C \frac{e^{B/4} (b - a)^{1/4}}{\epsilon^{1/2}},
\end{align*}
where the final inequality follows from Theorem 4.1 of Doss and Wellner (2016).

The following lemma is also used in the proof of Theorem 4 in the main text.

**Lemma 9.** Let \( S, S_1, S_2, \ldots S_k \) denote measurable subsets of \( \mathbb{R} \) such that \( S \subseteq \bigcup_{j=1}^k S_j \). Let \( \mathcal{F}_0 \) denote an arbitrary class of non-negative functions on \( \bigcup_{j=1}^k S_j \) and let \( \mathcal{G} := \{ e^{\tilde{\phi}_a} : \tilde{\phi} \in \mathcal{F}_0 \} \), where \( \tilde{\phi}_a \) is defined in (31) of the main text. Let \( \alpha_j := \inf\{ x : x \in S_j \} \) and suppose that \( \epsilon, \epsilon_1, \ldots, \epsilon_k > 0 \) satisfy
\[ \sum_{j=1}^k e^{-\alpha_j} \epsilon_j^2 \leq (1 - e^{-a}) \epsilon^2. \]
Then
\[ H_{[]} (\epsilon, \mathcal{F}_0, d_{H}, S) \leq \sum_{j=1}^k H_{[]} (\epsilon_j, \mathcal{G}, d_{H}, S_j). \]

**Proof.** We may assume that \( S_1, \ldots, S_k \) are pairwise disjoint, because otherwise we can work with the sets \( S'_1 := S_1 \) and \( S'_j := S_j \setminus \bigcup_{\ell=1}^{j-1} S_\ell \) for \( j = 2, \ldots, k \). For each \( j = 1, \ldots, k \), let \( \{ [f_{L,\ell}^{(j)}, f_{U,\ell}^{(j)}] : \ell = 1, \ldots, N_{[]} (\epsilon_j, \mathcal{G}, d_{H}, S_j) \} \) denote an \( \epsilon_j \)-Hellinger bracketing set for the class \( \mathcal{G} \) over \( S_j \). Now, for \( x \in S_j \) and
\[ \ell = (\ell_1, \ldots, \ell_k) \in \{ 1, \ldots, N_{[]} (\epsilon_1, \mathcal{G}, d_{H}, S_1) \} \times \ldots \times \{ 1, \ldots, N_{[]} (\epsilon_k, \mathcal{G}, d_{H}, S_k) \}, \]
set
\[ f_{L,\ell}(x) := \frac{e^{-x} f_{L,\ell}^{(j)}(x)}{1 - e^{-a}} \quad \text{and} \quad f_{U,\ell}(x) := \frac{e^{-x} f_{U,\ell}^{(j)}(x)}{1 - e^{-a}}. \]
Then for every $f \in \mathcal{F}_0$, there exists $\ell = (\ell_1, \ldots, \ell_k)$ such that $f_{L,\ell} \leq f \leq f_{U,\ell}$. Moreover,
\[
\int_S (f_{U,\ell}^{1/2} - f_{L,\ell}^{1/2})^2 \leq \sum_{j=1}^k \int_{S_j} \frac{e^{-x}}{1 - e^{-a}} \{f_{U,\ell_j}(x)^{1/2} - f_{L,\ell_j}(x)^{1/2}\}^2 \, dx \\
\leq \sum_{j=1}^k \frac{e^{-\alpha_j}}{1 - e^{-\alpha}} \alpha_j^2 \leq \epsilon^2,
\]
as required.

2.3. Auxiliary result for the proof of Theorem 5 in the main text. The following is the key empirical processes result used in the proof of Theorem 5.

**Theorem 10** (van de Geer (2000), Corollary 7.5). Let $f_0 \in \mathcal{F}$ and let $\mathcal{F}(f_0, \delta) := \{f \in \mathcal{F} : f \ll f_0, d_H(f, f_0) \leq \delta\}$. Suppose $\Psi : (0, \infty) \rightarrow (0, \infty)$ is a function such that
\[
\Psi(\delta) \geq \max \left\{ \delta, \int_0^\delta H^{1/2}(2^{1/2} \epsilon, \mathcal{F}(f_0, 4\delta), d_H) \, d\epsilon \right\}
\]
for every $\delta > 0$ and such that $\delta \mapsto \delta^{-2}\Psi(\delta)$ is decreasing on $(0, \infty)$. Let $\hat{f}_n$ denote the maximum likelihood estimator over $\mathcal{F}$ based on $X_1, \ldots, X_n \iid f_0$. There exists a universal constant $C > 0$ such that if $\delta_* > 0$ is such that $n^{1/2}\delta_*^2 \geq C\Psi(\delta_*)$, then for every $\delta \geq \delta_*$,
\[
\mathbb{P}\{d_X^2(\hat{f}_n, f_0) > \delta^2\} \leq C \exp \left( \frac{-n\delta^2}{C^2} \right).
\]

In fact, van de Geer (2000, Corollary 7.5) relies on a bracketing entropy upper bound in Hellinger distance for $\bar{\mathcal{F}}(f_0, \delta) := \{\frac{f + f_0}{2} : f \in \mathcal{F}, f \ll f_0, d_H(\frac{f + f_0}{2}, f_0) \leq \delta\}$, where the restriction $f \ll f_0$ can be included because the support of $\hat{f}_n$ is contained in the support of $f_0$. But for any non-negative functions $f_0$, $f_L$ and $f_U$ with $f_L \leq f_U$, we have
\[
\left( \frac{f_U + f_0}{2} \right)^{1/2} - \left( \frac{f_L + f_0}{2} \right)^{1/2} \leq \frac{1}{2^{1/2}} \left( f_U^{1/2} - f_L^{1/2} \right).
\]
Moreover, by the triangle inequality and the fact that the squared Hellinger distance is jointly convex in its arguments, if $d_H(\frac{f + f_0}{2}, f_0) \leq \delta$, then
\[
d_H(f, f_0) \leq \frac{2^{1/2}}{2^{1/2} - 1} \left\{ d_H\left( f, \frac{f + f_0}{2} \right) + d_H\left( \frac{f + f_0}{2}, f_0 \right) \right\} - \frac{d_H(f, f_0)}{2^{1/2} - 1} \leq 2^{1/2} \frac{d_H\left( \frac{f + f_0}{2}, f_0 \right)}{2^{1/2} - 1} \leq 4\delta,
\]
so $H\left(2^{1/2}\epsilon, \tilde{F}(f_0, \delta), d_H\right) \leq H\left(2^{1/2}\epsilon, F(f_0, 4\delta), d_H\right)$.

References.


