
Supplement: A Statistical Perspective on Coreset Density Estimation

1 PROOFS FROM SECTION 2

1.1 Proof of Lemma 1

Here we prove Lemma 1, restated below for convenience.

Lemma. *Let $K^{-1} = c(\log n)/n$ for $c > 0$ a sufficiently large absolute constant, and let $A = A_{\beta, L, K}$ denote a sufficiently small constant. Then for all $f \in \mathcal{P}_{\mathcal{H}}(\beta, L)$ and $X_1, \dots, X_n \stackrel{iid}{\sim} \mathbb{P}_f$, the event that for every $j = 1, \dots, K$ there exists some x_i in bin B_j holds with probability at least $1 - O(n^{-2})$.*

Proof. Note that $f_1(x_1) \in \mathcal{P}_{\mathcal{H}}(\beta, L)$ as a univariate density because $f(x) \in \mathcal{P}_{\mathcal{H}}(\beta, L)$. Hence, f_1 satisfies

$$|f_1(x) - f_1(y)| \leq L|x - y|^\alpha$$

for some absolute constants $L > 0$ and $\alpha \in (0, 1)$. If $B_{ik} = B_{jk} + s$ for $s \leq A$, then

$$|\mathbb{P}(B_{ik}) - \mathbb{P}(B_{jk})| \leq \int_{B_{ik}} |f(x_1) - f(x_1 + s)| dx_1 \leq LK^{-1}A^{1+\alpha}. \quad (1)$$

Thus for all i, j ,

$$|\mathbb{P}(B_i) - \mathbb{P}(B_j)| \leq \sum_{k=1}^{1/A} |\mathbb{P}(B_{ik}) - \mathbb{P}(B_{jk})| \leq LK^{-1}A^\alpha. \quad (2)$$

It follows that for all $i = 1, \dots, K$,

$$\lim_{A \rightarrow 0} \mathbb{P}(B_i) = K^{-1}. \quad (3)$$

Let \mathcal{E} denote the event that every bin B_i contains at least one observation x_k . By the union bound,

$$\mathbb{P}(\mathcal{E}^c) \leq \sum_{j=1}^K \mathbb{P}(X_{11} \notin B_j)^n \leq K \max_j (1 - \mathbb{P}(B_j))^n.$$

By (3), choosing A small enough ensures that $\mathbb{P}[B_j] \geq (1/2)K^{-1}$ for all j . In fact, by (1) one may take $A = (\frac{1}{2K^{-2}L})^{1/\alpha}$. Hence, setting $K^{-1} = c(\log n)/n$ for c sufficiently large, we have

$$\mathbb{P}(\mathcal{E}^c) = O(n^{-2}).$$

□

1.2 Proof of the lower bound in Theorem 1

In this section, $X = X_1, \dots, X_n \in \mathbb{R}^d$ denotes the sample. It is convenient to consider a more general family of *decorated coreset-based estimators*. A *decorated coreset* consists of a coreset X_S along with a data-dependent binary string σ of length R . A decorated coreset-based estimator is then given by $\hat{f}[X_S, \sigma]$, where $\hat{f} : \mathbb{R}^{d \times m} \times \{0, 1\}^R \rightarrow L^2([-1/2, 1/2]^d)$ is a measurable function. As with coreset-based estimators, we require that $\hat{f}[x_1, \dots, x_m, \sigma]$ is invariant under permutation of the vectors $x_1, \dots, x_m \in \mathbb{R}^d$. We slightly abuse notation and refer to the channel $S : X \rightarrow Y_S = (X_S, \sigma)$ as a decorated coreset scheme and \hat{f}_S as the decorated coreset-based estimator. The next proposition implies the lower bound in Theorem 1 on setting $R = 0$, in which case a

decorated coreset-based estimator is just a coreset-based estimator. This more general framework allows us to prove Theorem 4 on lower bounds for weighted coreset KDEs.

Proposition 2. *Let \hat{f}_S denote a decorated coreset-based estimator with decorated coreset scheme S such that $\sigma \in \{0, 1\}^R$. Then*

$$\sup_{f \in \mathcal{P}_{\mathcal{H}}(\beta, L)} \mathbb{E}_f \|\hat{f}_S - f\|_2 \geq c_{\beta, d, L} \left((m \log n + R)^{-\frac{\beta}{d}} + n^{-\frac{\beta}{2\beta+d}} \right).$$

1.2.1 Choice of function class

Fix $h \in (0, 1)$ such that $1/h^d$ is integral to be chosen later. Let $z_1, \dots, z_{1/h^d}$ label the points in $\{\frac{1}{2}h \cdot \mathbb{1}_d + h\mathbb{Z}^d\} \cap [-1/2, 1/2]^d$, where $\mathbb{1}_d$ denotes the all-ones vector of \mathbb{R}^d . We consider a class of functions of the form $f_\omega(x) = 1 + \sum_{j=1}^{1/h^d} \omega_j g_j(x)$ indexed by $\omega \in \{0, 1\}^{1/h^d}$. Here, $g_j(x)$ is defined to be

$$g_j(x) = h^\beta \phi\left(\frac{x - z_j}{h}\right)$$

where $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ is L -Hölder smooth of order β , has $\|\phi\|_\infty = 1$, and has $\int \phi(x) dx = 0$.

Informally, f_ω puts a bump on the uniform distribution with amplitude h^β over z_j if and only if $\omega_j = 1$. Using a standard argument (Tsybakov, 2009, Chapter 2) we can construct a packing \mathcal{V} of $\{0, 1\}^{1/h^d}$ which results $\mathcal{G} = \{f_\omega : \omega \in \mathcal{V}\}$ of the function class $\{f_\omega : \omega \in \{0, 1\}^{1/h^d}\}$ such that

- (i) $\|f - g\|_2 \geq c_{\beta, d, L} h^\beta$ for all $f, g \in \mathcal{G}$, $f \neq g$ and,
- (ii) \mathcal{G} is large in the sense that $M := |\mathcal{G}| \geq 2^{c_{\beta, d, L}/h^d}$.

1.2.2 Minimax lower bound

Using standard reductions from estimation to testing, we obtain that

$$\begin{aligned} \inf_{\substack{\hat{f}, |S|=m, \\ \sigma \in \{0, 1\}^R}} \sup_{f \in \mathcal{P}_{\mathcal{H}}(\beta, L)} \mathbb{E}_f \|\hat{f}_S - f\|_2 &\geq \inf_{\substack{\hat{f}, |S|=m, \\ \sigma \in \{0, 1\}^R}} \max_{f \in \mathcal{G}} \mathbb{E}_f \|\hat{f}_S - f\|_2 \\ &\geq c_{\beta, d, L} h^\beta \cdot \inf_{\psi_S} \frac{1}{M} \sum_{\omega \in \mathcal{V}} \mathbb{P}_{f_\omega}[\psi_S(X) \neq \omega]. \end{aligned} \quad (4)$$

where the infimum in the last line is over all tests $\psi_S : \mathbb{R}^{d \times n} \rightarrow [M]$ of the form $\psi_S(X) = \psi(Y_S)$ for a decorated coreset scheme S and a measurable function $\psi : \mathbb{R}^{d \times m} \times \{0, 1\}^R \rightarrow [M]$.

Let V denote a random variable that is distributed uniformly over \mathcal{V} and observe that

$$\frac{1}{M} \sum_{\omega \in \mathcal{V}} \mathbb{P}_{f_\omega}[\psi_S(X) \neq \omega] = \mathbb{P}[\psi_S(X) \neq V]$$

where \mathbb{P} denotes the joint distribution of (X, V) characterized by the conditional distribution $X|V = \omega$ which is assumed to have density f_ω for all $\omega \in \mathcal{V}$.

Next, by Fano's inequality (Cover & Thomas, 2006, Theorem 2.10.1) and the chain rule, we have

$$\mathbb{P}[\psi_S(X) \neq V] \geq 1 - \frac{I(V; \psi_S(X)) + 1}{\log M}, \quad (5)$$

where $I(V; \psi_S(X))$ denotes the mutual information between V and $\psi_S(X)$ and we used the fact that the entropy of V is $\log M$. Therefore, it remains to control $I(V; \psi_S(X))$. To that end, note that it follows from the data processing inequality that

$$I(V; \psi_S(X)) \leq I(V; (X_S, \sigma)) = I(V; Y_S) = \text{KL}(P_{V, Y_S} \| P_V \otimes P_{Y_S}),$$

where P_{V, Y_S}, P_V and P_{Y_S} denote the distributions of (V, Y_S) , V and Y_S respectively and observe that P_{Y_S} is the mixture distribution given by $P_{Y_S}(A, t) = M^{-1} \sum_{\omega \in \mathcal{V}} P_{f_\omega}(X_S \in A, \sigma = t)$ for $A \subset \mathbb{R}^{d \times m}$ and $t \in \{0, 1\}^R$.

Denote by f_{ω, Y_S} the mixed density of $P_{f_\omega}(X_S \in \cdot, \sigma = \cdot)$, where the continuous component is with respect to the Lebesgue measure on $[-1/2, 1/2]^{d \times m}$. Denote by \bar{f}_{Y_S} the mixed density of the uniform mixture of these:

$$\bar{f}_{Y_S} := \frac{1}{M} \sum_{\omega \in \mathcal{V}} f_{\omega, Y_S}.$$

By a standard information-theoretic inequality, for all measures \mathbb{Q} it holds that

$$\text{KL}(P_{V, Y_S} \| P_V \otimes P_{Y_S}) = \frac{1}{M} \sum_{\omega} \text{KL}(P_{Y_S | \omega} \| P_{Y_S}) \leq \frac{1}{M} \sum_{\omega} \text{KL}(P_{Y_S | \omega} \| \mathbb{Q}). \quad (6)$$

In fact, we have equality precisely when $\mathbb{Q} = P_{Y_S}$, and (6) follows immediately from the nonnegativity of the KL-divergence. Setting $\mathbb{Q} = \text{Unif}[-\frac{1}{2}, \frac{1}{2}]^d \otimes \text{Unif}\{0, 1\}^R$, for all ω we have

$$\begin{aligned} \text{KL}(P_{Y_S | \omega}, \mathbb{Q}) &= \sum_{t \in \{0, 1\}^R} \int_{[-\frac{1}{2}, \frac{1}{2}]^d} f_{\omega, Y_S}(x, t) \log \frac{f_{\omega, Y_S}(x, t)}{2^{-R}} dx \\ &\leq \sum_{t \in \{0, 1\}^R} \int_{[-\frac{1}{2}, \frac{1}{2}]^d} f_{\omega, Y_S}(x, t) \log f_{\omega, Y_S}(x, t) dx + R. \end{aligned} \quad (7)$$

Our next goal is to bound the first term on the right-hand-side above.

Lemma 2. *For any $\omega \in \mathcal{V}$, we have*

$$\sum_{t \in \{0, 1\}^R} \int_{[-\frac{1}{2}, \frac{1}{2}]^d} f_{\omega, Y_S}(x, t) \log f_{\omega, Y_S}(x, t) dx \leq 3m \log n.$$

Proof. Let \mathbb{P}_{X_S} denote the distribution of the (undecorated) coresot X_S , and note that the density of this distribution is given by $f_{\omega, X_S}(x) := \sum_{t \in \{0, 1\}^R} f_{\omega, Y_S}(x, t)$. Then because the logarithm is increasing,

$$\begin{aligned} \sum_{t \in \{0, 1\}^R} \int_{[-\frac{1}{2}, \frac{1}{2}]^d} f_{\omega, Y_S}(x, t) \log f_{\omega, Y_S}(x, t) dx &\leq \sum_{t \in \{0, 1\}^R} \int_{[-\frac{1}{2}, \frac{1}{2}]^d} f_{\omega, Y_S}(x, t) \log f_{\omega, X_S}(x) dx \\ &= \int_{[-\frac{1}{2}, \frac{1}{2}]^d} f_{\omega, X_S}(x) \log f_{\omega, X_S}(x) dx. \end{aligned}$$

By the union bound,

$$\mathbb{P}_{X_S}(\cdot) \leq \sum_{s \in \binom{[n]}{m}} \mathbb{P}_{X_s}(\cdot) = \binom{n}{m} \mathbb{P}_{X_{[m]}}(\cdot).$$

It follows readily that $f_{\omega, X_S}(\cdot) \leq \binom{n}{m} f_{\omega, X_{[m]}}(\cdot)$. Next, let $Z \in [-1/2, 1/2]^{d \times m}$ be a random variable with density f_{ω, X_S} and note that

$$\int f_{\omega, X_S} \log f_{\omega, X_S} = \mathbb{E} \log f_{\omega, X_S}(Z) \leq \log \binom{n}{m} + \mathbb{E} \log f_{\omega, X_{[m]}}(Z) \leq m \log \left(\frac{en}{m} \right) + m \log 2,$$

where in the last inequality, we use the fact that $f_{\omega, X_{[m]}} = f_\omega^m \leq 2^m$. The lemma follows. \square

Since $\log M \geq c_{\beta, d, L} h^{-d}$, it follows from (5)–(7) and Lemma 2 that

$$\mathbb{P}[\psi_S(X) \neq V] \geq 1 - \frac{3m \log n + R + 1}{\log M} \geq 0.5$$

on setting $h = c_{\beta, d, L} (m \log n + R)^{-1/d}$. Plugging this value back into (4) yields

$$\inf_{\hat{f}, |S|=m} \sup_{f \in \mathcal{P}_{\mathcal{H}}(\beta, L)} \mathbb{E}_f \|\hat{f}_S - f\|_2 \geq c_{\beta, d, L} (m \log n + R)^{-\beta/d}.$$

Moreover, it follows from standard minimax theory (see e.g. Tsybakov, 2009, Chapter 2) that

$$\inf_{\hat{f}, |S|=m} \sup_{f \in \mathcal{P}_{\mathcal{H}}(\beta, L)} \mathbb{E}_f \|\hat{f}_S - f\|_2 \geq c_{\beta, d, L} n^{-\frac{\beta}{2\beta+d}}.$$

Combined together, the above two displays give the lower bound of Proposition 2.

2 PROOFS FROM SECTION 3

2.1 Proof of Proposition 1

We restate the result below.

Proposition. Let $k(x) = \prod_{i=1}^d \kappa(x_i)$ denote a kernel with $\kappa \in \mathcal{S}(\gamma, L')$ such that $|\kappa(x)| \leq c_{\beta,d} |x|^{-\nu}$ for some $\nu \geq \beta + d$, and the KDE

$$\hat{f}(y) = \frac{1}{n} \sum_{i=1}^n k_h(X_i - y)$$

with bandwidth $h = n^{-\frac{1}{2\beta+d}}$ satisfies

$$\sup_{f \in \mathcal{P}_{\mathcal{H}}(\beta, L)} \mathbb{E} \|f - \hat{f}\|_2 \leq c_{\beta,d,L} n^{-\frac{\beta}{2\beta+d}}.$$

Then the Carathéodory coreset estimator $\hat{g}_S(y)$ constructed from \hat{f} with $T = c_{d,\gamma,L'} n^{\frac{d/2+\beta+\gamma}{\gamma(2\beta+d)}}$ satisfies

$$\sup_{f \in \mathcal{P}_{\mathcal{H}}(\beta, L)} \mathbb{E} \|\hat{g}_S - f\|_2 \leq c_{\beta,d,L} n^{-\frac{\beta}{2\beta+d}}.$$

Let $\varphi : \mathbb{R}^d \rightarrow [0, 1]$ denote a cutoff function that has the following properties: $\varphi \in \mathcal{C}^\infty$, $\varphi|_{[-1,1]^d} \equiv 1$, and φ is compactly supported on $[-2, 2]^d$.

Lemma 3. Let $\tilde{k}_h(x) = k_h(x)\varphi(x)$ where $|\kappa(x)| \leq c_{\beta,d} |x|^{-\nu}$. Then

$$\|\tilde{k}_h - k_h\|_2 \leq c_{\beta,d} h^{-d+\nu}.$$

Proof.

$$\begin{aligned} \|\tilde{k}_h - k_h\|_2 &= \|(1 - \varphi)k_h\|_2 \\ &\leq \|(1 - \mathbb{1}_{[-1,1]^d})k_h\|_2 \\ &= h^{-d/2} \|(1 - \mathbb{1}_{[-\frac{1}{h}, \frac{1}{h}]^d})k\|_2 \\ &\leq dh^{-d/2} \|\mathbb{1}_{|x_1| \geq \frac{1}{h}} k\|_2 \\ &\leq c_{\beta,d} h^{-d/2} \sqrt{\int_{|x_1| \geq \frac{1}{h}} \kappa^2(x_1) dx_1} \\ &\leq c_{\beta,d} h^{-d+\nu}. \end{aligned}$$

□

The triangle inequality and the previous lemma yield the next result.

Lemma 4. Let k denote a kernel such that $|\kappa(x)| \leq c_{\beta,d} |x|_2^{-\nu}$. Recall the definition of \tilde{k}_h from Lemma 3. Let $\{X_j : j \in S\} \subset \mathbb{R}^d$ denote an arbitrary set of points (not necessarily from a sample), and let

$$\hat{g}_S(y) = \sum_{j \in S} \lambda_j k_h(X_j - y)$$

denote a weighted KDE on the points labeled by S where $\lambda_j \geq 0$ and $\mathbb{1}^T \lambda = 1$. Let

$$\tilde{g}_S(y) = \sum_{j \in S} \lambda_j \tilde{k}_h(X_j - y).$$

Then

$$\|\hat{g}_S - \tilde{g}_S\|_2 \leq c_{\beta,d} h^{-\nu+d}.$$

Next we show that \tilde{k}_h is well approximated by its Fourier expansion on $[-2, 2]^d$. Since \tilde{k}_h is a smooth periodic function on $[-2, 2]^d$, it is expressed in L^2 as a Fourier series on $\frac{\pi}{2}\mathbb{Z}^d$. Thus we bound the tail of this expansion. In what follows, $\alpha \in \mathbb{Z}_{\geq 0}^d$ is a multi-index and

$$\bar{\mathcal{F}}[f](\omega) = \frac{1}{4^{2d}} \int f(x) e^{i\langle x, \omega \rangle} dx$$

denotes the (rescaled) Fourier transform on $[-2, 2]^d$, where $\omega \in \frac{\pi}{2}\mathbb{Z}^d$.

Lemma 5. *Suppose that $\kappa \in \mathcal{S}(\beta, L')$. Let $A = \{\omega \in \frac{\pi}{2}\mathbb{Z}^d : |\omega|_1 \leq T\}$, and define*

$$\tilde{k}_h^T(y) = \sum_{\omega \in A} \bar{\mathcal{F}}[\tilde{k}_h](\omega) e^{i\langle y, \omega \rangle}.$$

Then

$$\|(\tilde{k}_h - \tilde{k}_h^T) \mathbb{1}_{[-2, 2]^d}\|_2 \leq c_{\gamma, d, L'} T^{-\gamma} h^{-d/2-\gamma}$$

Proof. Observe that for $\omega \notin A$, it holds that

$$\sum_{|\alpha|_1 = \gamma} \frac{\gamma!}{\alpha!} |\omega|^\alpha = (|\omega_1| + \dots + |\omega_d|)^\gamma \geq T^\gamma.$$

Therefore,

$$\begin{aligned} \|\bar{\mathcal{F}}[\tilde{k}_h](\omega) \mathbb{1}_{\omega \notin A}\|_{\ell_2} &\leq T^{-\gamma} \left\| \sum_{|\alpha|_1 = \gamma} \frac{\gamma!}{\alpha!} |\omega|^\alpha \bar{\mathcal{F}}[\tilde{k}_h](\omega) \mathbb{1}_{\omega \notin A} \right\|_{\ell_2} \\ &\leq T^{-\gamma} \sum_{|\alpha|_1 = \gamma} \frac{\gamma!}{\alpha!} \|\omega^\alpha \bar{\mathcal{F}}[\tilde{k}_h](\omega)\|_{\ell_2} \\ &= c_d T^{-\gamma} \sum_{|\alpha|_1 = \gamma} \frac{\gamma!}{\alpha!} \left\| \frac{\partial^\alpha}{\partial x^\alpha} \tilde{k}_h(x) \right\|_2, \end{aligned} \tag{8}$$

where in the last line we used Parseval's identity. For any multi-index α with $|\alpha|_1 = \gamma$,

$$\begin{aligned} \left\| \frac{\partial^\alpha}{\partial x^\alpha} \tilde{k}_h(x) \right\|_2 &= \left\| \sum_{\eta \leq \alpha} \frac{\partial^\eta}{\partial x^\eta} k_h(x) \frac{\partial^{\alpha-\eta}}{\partial x^{\alpha-\eta}} \varphi(x) \right\|_2 \\ &\leq h^{-\frac{d}{2}-\gamma} \sum_{\eta \leq \alpha} c_{d, \gamma} \left\| \frac{\partial^\eta}{\partial x^\eta} k(x) \right\|_2, \end{aligned} \tag{9}$$

where we used that the derivatives of φ are bounded. Next by Parseval's identity,

$$\left\| \frac{\partial^\eta}{\partial x^\eta} k(x) \right\|_2^2 = \prod_{i=1}^d \|\omega_i^{\eta_i} \mathcal{F}[\kappa](\omega_i)\|_2^2. \tag{10}$$

For $0 \leq a \leq \gamma$, we have

$$\int |\omega^a \mathcal{F}[\kappa](\omega)|^2 d\omega \leq 2\|\kappa\|_1^2 + \int_{|\omega| \geq 1} |\omega^\gamma \mathcal{F}[\kappa](\omega)|^2 d\omega \leq 2\|\kappa\|_1^2 + L'. \tag{11}$$

By (8)–(11),

$$\|\bar{\mathcal{F}}[\tilde{k}_h](\omega) \mathbb{1}_{\omega \notin A}\|_{\ell_2} \leq c_{d, \gamma, L'} T^{-\gamma} h^{-\frac{d}{2}-\gamma},$$

as desired. \square

Applying the previous lemma and linearity of the Fourier transform, we have the next corollary that gives an expansion for a general KDE on the smaller domain $[-\frac{1}{2}, \frac{1}{2}]^d$.

Corollary 2. *Let \tilde{g}_S denote the weighted KDE built from \tilde{k}_h from Lemma 4 where $\{X_j : j \in S\} \subset [-\frac{1}{2}, \frac{1}{2}]^d$ is an arbitrary set of points (not necessarily from a sample) and moreover $\kappa \in \mathcal{S}(\beta, L')$. Let $A = \{\omega \in \frac{\pi}{2}\mathbb{Z}^d : |\omega|_1 \leq T\}$, and define*

$$\tilde{g}_S^T(y) = \sum_{\omega \in A} \bar{\mathcal{F}}[\tilde{g}_S](\omega) e^{i\langle y, \omega \rangle}.$$

Then

$$\|(\tilde{g}_S - \tilde{g}_S^T)\mathbf{1}_{[-\frac{1}{2}, \frac{1}{2}]^d}\|_2 \leq c_{d,\gamma,L'} T^{-\gamma} h^{-d/2-\gamma} L.$$

Now we have all the ingredients needed to prove Proposition 1.

Proof of Proposition 1 . Let

$$\tilde{f}(y) = \frac{1}{n} \sum_{j=1}^n \tilde{k}_h(X_j - y),$$

and

$$\tilde{g}_S(y) = \sum_{j \in S} \lambda_j \tilde{k}_h(X_j - y)$$

where the coreset $\{X_j : j \in S\}$ is constructed by Carathéodory's theorem as in Section 3 of the main text and \tilde{k}_h is defined as in Lemma 3. Also consider their Fourier expansions \tilde{f}^T and \tilde{g}_S^T as defined in Corollary 2. Observe that, by construction of the Carathéodory coreset,

$$\tilde{f}^T(y) = \tilde{g}_S^T(y) \quad \forall y \in [-\frac{1}{2}, \frac{1}{2}]^d.$$

In what follows, $\|\cdot\|_2$ is computed on $[-\frac{1}{2}, \frac{1}{2}]^d$. By the triangle inequality,

$$\begin{aligned} \|\hat{g}_S - \hat{f}\|_2 &\leq \|\hat{g}_S - \tilde{g}_S\|_2 + \|\tilde{g}_S - \tilde{g}_S^T\|_2 + \|\tilde{g}_S^T - \tilde{f}^T\|_2 \\ &\quad + \|\tilde{f}^T - \tilde{f}\|_2 + \|\tilde{f} - \hat{f}\|_2 \\ &\leq c_{\beta,d} h^{-d+\nu} + c_{d,\gamma,L'} T^{-\gamma} h^{-d/2-\gamma} + 0 \\ &\quad + c_{d,\gamma,L'} T^{-\gamma} h^{-d/2-\gamma} + c_{\beta,d} h^{-d+\nu} \end{aligned} \tag{12}$$

On the right-hand-side of the first line, the first and last terms are bounded via Lemma 4. The second and fourth terms are bounded via Lemma 5, and the third term is 0 by Carathéodory. By our choice of T and the decay properties of k , we have

$$\|\hat{g}_S - \hat{f}\|_2 \leq c_{\beta,d,L} h^\beta \leq c_{\beta,d,L} n^{-\beta/(2\beta+d)}.$$

The conclusion follows by the hypothesis on k , the previous display, and the triangle inequality. \square

2.2 Proof of Theorem 2

We restate Theorem 2 here for convenience.

Theorem. *Let $\varepsilon > 0$. The Carathéodory coreset estimator $\hat{g}_S(y)$ built using the kernel k_s and setting $T = c_{d,\beta,\varepsilon} n^{\frac{\varepsilon}{d} + \frac{1}{2\beta+d}}$ satisfies*

$$\sup_{f \in \mathcal{P}_{\mathcal{H}}(\beta,L)} \mathbb{E}_f \|\hat{g}_S - f\|_2 \leq c_{\beta,d,L} n^{-\frac{\beta}{2\beta+d}}.$$

The corresponding coreset has cardinality

$$m = c_{d,\beta,\varepsilon} n^{\frac{d}{2\beta+d} + \varepsilon}.$$

Proof. Our goal is to apply Proposition 1 to k_s . First we show that the standard KDE built from k_s attains the minimax rate on $\mathcal{P}_{\mathcal{H}}(\beta, L)$. The Fourier condition

$$\text{ess sup}_{\omega \neq 0} \frac{1 - \mathcal{F}[k_s](\omega)}{|\omega|^\alpha} \leq 1, \quad \forall \alpha \leq \beta,$$

implies that k_s is a kernel of order β (Tsybakov, 2009, Definition 1.3). Since $\mathcal{F}[k_s](0) = 1 = \int k_s(x) dx$, it remains to show that the ‘moments’ of order at most β of k_s vanish. In fact all of the moments vanish. We have, expanding the exponential and using the multinomial formula,

$$\begin{aligned} \psi(\omega) &= \mathcal{F}^{-1}[k_s](\omega) \\ &= \int k_s(x) e^{i\langle x, \omega \rangle} dx \\ &= \sum_{t=0}^{\infty} \int k_s(x) \frac{(i\langle x, \omega \rangle)^t}{t!} dx \\ &= \sum_{t=0}^{\infty} \sum_{|\alpha|_1=t} \frac{i^t}{\alpha!} \omega^\alpha \left\{ \int k_s(x) x^\alpha dx \right\}. \end{aligned}$$

Since $\psi(\omega) \equiv 1$ in a neighborhood near the origin, it follows that all of the terms $\int k_s(x) x^\alpha dx = 0$. Thus k_s is a kernel of order β for all $\beta \in \mathbb{Z}_{\geq 0}$, and the standard KDE on all of the dataset with bandwidth $h = n^{-1/(2\beta+d)}$ attains the rate of estimation $n^{-\beta/(2\beta+d)}$ over $\mathcal{P}_{\mathcal{H}}(\beta, L)$ (see e.g. Tsybakov, 2009, Theorem 1.2).

Next, $|\kappa_s(x)| \leq c_{\beta,d} |x|^\nu$ for $\nu = \lceil \beta + d \rceil$. This is because

$$x^\nu \kappa_s(x) = x^\nu \mathcal{F}[\psi](x) = \mathcal{F} \left[\frac{d^\nu}{dx^\nu} \psi \right] (x) \leq \left\| \frac{d^\nu}{dx^\nu} \psi \right\|_1 \leq c_{\beta,d}.$$

Moreover for all $\gamma \in \mathbb{Z}_{>0}$, $\kappa_s \in \mathcal{S}(\gamma, c_\gamma)$. By Parseval’s identity,

$$\left\| \frac{d^\gamma}{dx^\gamma} \kappa_s \right\|_2 = \left\| \mathcal{F} \left[\frac{d^\gamma}{dx^\gamma} \kappa_s \right] \right\|_2 = \left\| \omega^\gamma \psi(\omega) \right\|_2 \leq c_\gamma$$

because ψ has compact support (see e.g. Katznelson, 2004, Chapter VI).

All of the hypotheses of Proposition 1 are satisfied, so we apply the result with

$$\gamma = \frac{d}{2\varepsilon}$$

to derive Theorem 2. □

2.3 Proof of Corollary 1

Corollary. Let $\varepsilon > 0$ and $m \leq c_{\beta,d,\varepsilon} n^{\frac{d}{2\beta+d} + \varepsilon}$. The Carathéodory coreset estimator $\hat{g}_S(y)$ built using the kernel k_s , setting $h = m^{-\frac{1}{d} + \frac{\varepsilon}{\beta}}$ and $T = c_d m^{1/d}$, satisfies

$$\sup_{f \in \mathcal{P}_{\mathcal{H}}(\beta, L)} \mathbb{E} \|\hat{g}_S - f\|_2 \leq c_{\beta,d,\varepsilon,L} \left(m^{-\frac{\beta}{d} + \varepsilon} + n^{-\frac{\beta}{2\beta+d} + \varepsilon} \right),$$

and the corresponding coreset has cardinality m .

Proof. Recall from the proof of Theorem 2 that k_s is a kernel of all orders. By a standard bias-variance trade-off (see e.g. Tsybakov, 2009, Section 1.2), it holds for the KDE \hat{f} with bandwidth h built on the entire dataset that

$$\mathbb{E}_f \|\hat{f} - f\|_2 \leq c_{\beta,d,L} \left(h^\beta + \frac{1}{\sqrt{nh^d}} \right). \quad (13)$$

Moreover, from (12) applied to k_s , setting $T = c_d m^{1/d}$, we get

$$\|\hat{g}_S - \hat{f}\|_2 \leq c_{\beta,d} h^\beta + c_{d,\gamma} m^{-\gamma/d} h^{-d/2-\gamma}. \quad (14)$$

Choosing

$$\gamma = (\beta + \frac{d}{2})(\frac{\beta}{d\varepsilon} - 1), \quad h = m^{-\frac{1}{d} + \frac{\varepsilon}{\beta}}$$

(assuming without loss of generality that $\varepsilon > 0$ is sufficiently small so that $\gamma > 0$), then the triangle inequality, (13), (14), and the upper bound on m yield the conclusion of Corollary 1. \square

2.4 Proof of Theorem 4

For convenience, we restate Theorem 4 here.

Theorem. Let $A, B \geq 1$. Let k denote a kernel with $\|k\|_2 \leq n$. Let \hat{g}_S denote a weighted coreset KDE with bandwidth $h \geq n^{-A}$ built from k with weights $\{\lambda_j\}_{j \in S}$ satisfying $\max_{j \in S} |\lambda_j| \leq n^B$. Then

$$\sup_{f \in \mathcal{P}_{\mathcal{H}}(\beta, L)} \mathbb{E}_f \|\hat{g}_S - f\|_2 \geq c_{\beta,d,L} \left[(A+B)^{-\frac{\beta}{d}} (m \log n)^{-\frac{\beta}{d}} + n^{-\frac{\beta}{2\beta+d}} \right].$$

Proof. Let $\lambda = \lambda_1, \dots, \lambda_m$ and let $\tilde{\lambda} = \tilde{\lambda}_1, \dots, \tilde{\lambda}_m$. Observe that

$$\begin{aligned} \left\| \sum_{j \in S} \lambda_j k_h(X_j - y) - \sum_{j \in S} \tilde{\lambda}_j k_h(X_j - y) \right\|_2 &\leq \sum_{j \in S} |\lambda_j - \tilde{\lambda}_j| \|k_h(X_j - y)\|_2 \\ &\leq \|\lambda - \tilde{\lambda}\|_\infty n^2 h^{-d/2}. \end{aligned} \quad (15)$$

Using this we develop a decorated coreset-based estimator \hat{f}_S (see Section 1.2 of this Supplement) that approximates \hat{g}_S well. Set $\delta = c_{\beta,d,L} n^{-4} h^{d/2}$ for $c_{\beta,d,L}$ sufficiently small and to be chosen later. Order the points of the coreset X_S according to their first coordinate. This gives rise to an ordering \preceq so that

$$X'_1 \preceq X'_2 \preceq \dots \preceq X'_m$$

denote the elements of X_S . Let $\lambda \in \mathbb{R}^m$ denote the correspondingly reordered collection of weights so that

$$\hat{g}_S(y) = \sum_{j=1}^m \lambda_j k_h(X'_j - y).$$

Construct a δ -net \mathcal{N}_δ with respect to the sup-norm $|\cdot|_\infty$ on the set $\{\nu \in \mathbb{R}^m : |\nu|_\infty \leq n^B\}$. Observe that

$$\log |\mathcal{N}_\delta| = \log(n^B \delta^{-1})^m = c_{\beta,d,L} (B+A)m \log n \quad (16)$$

Define R to be the smallest integer larger than the right-hand-side above. Then we can construct a surjection $\phi : \{0, 1\}^R \rightarrow \mathcal{N}_\delta$. Note that ϕ is constructed before observing any data: it simply labels the elements of the δ -net \mathcal{N}_δ by strings of length R .

Given $\hat{g}_S(y) = \sum_{j \in S} \lambda_j k_h(X_j - y)$, define \hat{f}_S as follows:

1. Let $\tilde{\lambda} \in \mathbb{R}^m$ denote the closest element in \mathcal{N}_δ to $\lambda \in \mathbb{R}^m$.
2. Choose $\sigma \in \{0, 1\}^R$ such that $\phi(\sigma) = \tilde{\lambda}$.
3. Define the decorated coreset $Y_S = (X_S, \sigma)$.
4. Order the points of X_S by their first coordinate. Pair the i -th element of $\tilde{\lambda}$ with the i -th element X'_i of X_S , and define

$$\hat{f}_S(y) = \sum_{j=1}^m \tilde{\lambda}_j k_h(X'_j - y)$$

We see that \hat{f}_S is a decorated-coreset based estimator because in step 4 this estimator is constructed only by looking at the coreset X_S and the bit string σ . Moreover, by (15) and the setting of δ ,

$$\|\hat{f}_S - \hat{g}_S\|_2 \leq c_{\beta,d,L} n^{-2}. \quad (17)$$

By Proposition 2 and our choice of R ,

$$\sup_{f \in \mathcal{P}_{\mathcal{H}}(\beta,L)} \mathbb{E}_f \|\hat{f}_S - f\|_2 \geq c_{\beta,d,L} \left((A+B)^{-\frac{\beta}{d}} (m \log n)^{-\frac{\beta}{d}} + n^{-\frac{\beta}{2\beta+d}} \right).$$

Applying the triangle inequality and (17) yields Theorem 4. \square

3 PROOFS FROM SECTION 4

Notation: Given a set of points $X = x_1, \dots, x_m \in [-1/2, 1/2]$ (not necessarily a sample), we let

$$\hat{f}_X(y) = \frac{1}{m} \sum_{i=1}^m k_h(X_i - y)$$

denote the uniformly weighted KDE on X .

3.1 Proof of Theorem 5

Theorem. *Let k denote a nonnegative kernel satisfying*

$$k(t) = O(|t|^{-(k+1)}), \quad \text{and} \quad \mathcal{F}[k](\omega) = O(|\omega|^{-\ell})$$

for some $\ell > 0$, $k > 1$. Suppose that $0 < \alpha < 1/3$. If

$$m \leq \frac{n^{\frac{2}{3}-2(\alpha(1-\frac{2}{\ell})+\frac{2}{3\ell})}}{\log n},$$

then

$$\inf_{h,S:|S| \leq m} \sup_{f \in \mathcal{P}_{\mathcal{H}}(1,L)} \mathbb{E} \|\hat{f}_S^{\text{unif}} - f\|_2 = \Omega_k \left(\frac{n^{-\frac{1}{3}+\alpha}}{\log n} \right).$$

The infimum above is over all possible choices of bandwidth h and all coreset schemes S of cardinality at most m .

The proof of Theorem 5 follows directly from Propositions 3 and 4, which are presented in Sections 3.1.1 and 3.1.2, respectively, of this Supplement.

3.1.1 Small bandwidth

First we show that uniformly weighted coreset KDEs on m points poorly approximate densities that are very close to 0 everywhere.

Lemma 6. *Let \hat{f}_X denote a uniformly weighted coreset KDE built from an even kernel $k : \mathbb{R} \rightarrow \mathbb{R}$ with bandwidth h on m points $X = x_1, \dots, x_m \in \mathbb{R}$. Suppose that quantiles $0 \leq q_1 \leq q_2$ satisfy*

$$\int_{-q_1}^{q_1} k(t) dt \geq 0.9, \quad \text{and} \quad (18)$$

$$\int_{-q_2}^{q_2} k(t) dt \geq 1 - \gamma. \quad (19)$$

Let U denote an interval $[0, u]$ where

$$u \geq 8q_2 h, \quad (20)$$

and suppose that $f : U \rightarrow \mathbb{R}$ satisfies

$$\frac{1}{100q_1mh} \leq f(x) \leq \frac{45}{44} \cdot \frac{1}{100q_1mh} \quad (21)$$

for all $x \in U$.

Then

$$\inf_{X:|X|=m} \|(\hat{f}_X - f)\mathbf{1}_U\|_1 \geq \frac{u}{440q_1mh} - \gamma.$$

Proof. Let N denote the number of $x_i \in X$ such that $[x_i - q_1h, x_i + q_1h] \subset [0, u]$. The argument proceeds in two cases. With foresight, we set $\alpha = 1/(44q_1)$. Also let $C_1 = 1/(100q_1)$ and $C_2 = 45/(4400q_1)$.

Case 1: $N \geq \frac{\alpha u}{h}$. Then by (18) and the nonnegativity of k ,

$$\|\hat{f}_X \mathbf{1}_U\|_1 \geq \frac{0.9N}{m} \geq \frac{0.9\alpha u}{mh}.$$

By (21),

$$\|f\|_1 \leq \frac{C_2 u}{mh}.$$

Hence,

$$\|(\hat{f}_X - f)\mathbf{1}_U\|_1 \geq \frac{u}{mh}(0.9\alpha - C_2) = C_2 \frac{u}{mh} = \frac{45}{4400} \cdot \frac{u}{q_1mh}.$$

Thus Lemma 6 holds in Case 1 where $N \geq \alpha u/h$.

Case 2: $N \leq \frac{\alpha u}{h}$. Let

$$V = [2hq_2, u - 2hq_2] \setminus \bigcup_{j \in T} [x_j - q_1h, x_j + q_1h]$$

where T is the set of indices j so that $[x_j - q_1h, x_j + q_1h] \subset U$. Observe that if $j \notin T$, then by (19),

$$\int_V \frac{1}{h} k\left(\frac{x_j - t}{h}\right) dt \leq \gamma.$$

If $j \in T$, then by (18),

$$\int_V \frac{1}{h} k\left(\frac{x_j - t}{h}\right) dt \leq 0.1.$$

Thus,

$$\|\hat{f}_X \mathbf{1}_V\|_1 \leq \frac{0.1N}{m} + \gamma \leq \frac{\alpha 0.1u}{mh} + \gamma.$$

By the union bound, observe that the Lebesgue measure of V is at least

$$u - 4hq_2 - 2Nq_1h \geq \frac{u}{2} - 2Nq_1h \geq u\left(\frac{1}{2} - 2\alpha q_1\right).$$

Next, by (21),

$$\|f \mathbf{1}_V\|_1 \geq C_1 \frac{u}{mh} \left(\frac{1}{2} - 2\alpha q_1\right).$$

Therefore,

$$\|(\hat{f}_X - f)\mathbf{1}_U\|_1 \geq \frac{u}{mh} (C_1(1/2 - 2\alpha q_1) - 0.1\alpha) - \gamma = \frac{u}{440q_1mh} - \gamma. \quad (22)$$

□

Proposition 3. Let $L > 2$. Let $0 < \delta < 1/3$ denote an absolute constant. Let \hat{f}_X denote a uniformly weighted coresets KDE with bandwidth h built from a kernel k on $X = x_1, \dots, x_m$. Suppose that $k(t) \leq \Delta|t|^{-(k+1)}$ for some absolute constants $\Delta > 0, k \geq 1$. If $h \leq n^{-1/3+\delta}$, then for

$$m \leq \frac{n^{2/3-2\delta}}{\log n}$$

it holds that

$$\sup_{f \in \mathcal{P}_{\mathcal{H}}(1, L)} \inf_{X: |X|=m} \|\hat{f}_X - f\|_2 = \Omega\left(\frac{n^{-1/3+\delta}}{\log n}\right). \quad (23)$$

Proof. Let

$$f(t) = \lambda \left(e^{-1/t} \mathbf{1}(t \in [-1/2, 0]) + e^{-1/(1-t)} \mathbf{1}(t \in [0, 1/2]) \right),$$

where λ is a normalizing constant so that $\int f = 1$. Observe that $f \in \mathcal{P}_{\mathcal{H}}(1, L)$. Our first goal is to show that

$$\|\hat{f}_X - f\|_1 = \Omega\left(\frac{1}{mh \log^2(mh)}\right)$$

holds for all $\tau/h \leq m \leq h^{-2}$ and for all $h \leq n^{-1/3+\delta}$, where τ is an absolute constant to be determined.

We apply Lemma 6 to the density f . Let q_1 be defined as in Lemma 6, and set $C_1 = 1/(100q_1)$ and $C_2 = 45/(4400q_1)$. Set $\tau = 10C_2/\lambda$. Let

$$U = [t_1, t_2] := \left[\frac{1}{\log(\lambda mh/C_1)}, \frac{1}{\log(\lambda mh/C_2)} \right].$$

The function $f|_U$ satisfies the bounds (21) from Lemma 6. Observe that the length of U is

$$u := t_2 - t_1 = \Omega\left(\frac{1}{\log^2(mh)}\right).$$

We set the parameter γ in Lemma 6 to be

$$\gamma = \frac{1}{800q_1 mh \log^2(mh)}.$$

By the decay assumption on k , we may set

$$q_2 := \left(\frac{2\Delta}{k\gamma}\right)^{1/k}.$$

Therefore,

$$u - 8q_2h = \Omega\left(\frac{1}{\log^2(mh)}\right) - 8h \left(\frac{2\Delta}{k\gamma}\right)^{1/k} \quad (24)$$

$$= \Omega\left(\frac{1}{\log^2(mh)}\right) - O(h(mh \log^2(mh))^{1/k}) \quad (25)$$

$$= \Omega\left(\frac{1}{\log^2(h^{-1})}\right) - O(h^{1-1/k} \log^2(h^{-1})) > 0 \quad (26)$$

for n sufficiently large, because we assume $\tau/h \leq m \leq h^{-2}$, $h \leq n^{-1/3+\delta}$, and $k > 1$. Hence, condition (20) is satisfied for m, h in the specified range, so we apply Cauchy-Schwarz and Lemma 6 to conclude that for all $\tau/h \leq m \leq h^{-2}$ and $h \leq n^{-1/3+\delta}$,

$$\|\hat{f}_X - f\|_2 \geq \|\hat{f}_X - f\|_1 = \Omega\left(\frac{1}{mh \log^2(mh)}\right) = \Omega\left(\frac{1}{mh \log^2(h^{-1})}\right). \quad (27)$$

Suppose first that $\log^2(1/h) \geq n^{1/3-\delta}$. Then clearly the right-hand side of (27) is $\Omega(1)$ for $m \leq n$. Otherwise, we have for all $h \leq n^{-1/3+\delta}$ that if m is in the range

$$\frac{\tau}{h} \leq m \leq \min \left(\frac{n^{1/3-\delta} \log n}{h \log^2(1/h)}, h^{-2} \right) =: N_h,$$

then (27) implies

$$\|\hat{f}_X - f\|_2 = \Omega \left(\frac{n^{-1/3+\delta}}{\log n} \right). \quad (28)$$

Moreover, a uniformly weighted coreset KDE on $m = O(1/h)$ points can be expressed as a uniformly weighted coreset KDE on $\Omega(1/h)$ points by setting some of the x_i 's to be duplicates. Hence (28) holds for all $1 \leq m \leq N_h$. Since N_h is a decreasing function of h , it follows that (28) holds for all $m \leq n^{2/3-2\delta}/\log n$ and $h \leq n^{-1/3+\delta}$, as desired. \square

3.1.2 Large bandwidth

Lemma 7. *Let $\varepsilon = \varepsilon(n) > 0$, and let \hat{f}_X denote the uniformly weighted coreset KDE on X with bandwidth h . Suppose that $\phi : \mathbb{R} \rightarrow \mathbb{R}$ is an odd C^∞ function supported on $[-1/4, 1/4]$. Let $f(t) : [-1/2, 1/2] \rightarrow \mathbb{R}_{\geq 0}$ denote the density*

$$f(t) = \frac{12}{11}(1 - t^2) + \varepsilon \phi(t) \cos \left(\frac{t}{\varepsilon} \right).$$

Then

$$\begin{aligned} \|\hat{f}_X - f\|_2^2 &\geq \frac{1}{2} \varepsilon^2 (\|\phi\|_2^2 - |\mathcal{F}[\phi^2](2\varepsilon^{-1})|) \\ &\quad - \|\phi\|_1 \sup_{|\omega| \geq h\varepsilon^{-1}/2} |\mathcal{F}[k](\omega)| - 2\varepsilon \int_{|\omega| \geq \varepsilon^{-1}/2} |\mathcal{F}[\phi](\omega)| d\omega. \end{aligned} \quad (29)$$

Proof. Let $g(t) = (12/11)(1 - t^2)$ and $\psi(t) = \varepsilon \phi(t) \cos(t/\varepsilon)$. Observe that

$$\begin{aligned} \|\hat{f}_X - f\|_2^2 &\geq \|g - f\|_2^2 - 2\langle \hat{f}_X, g - f \rangle + 2\langle g, \psi(t) \rangle \\ &= \|g - f\|_2^2 - 2\langle \hat{f}_X, g - f \rangle \end{aligned} \quad (30)$$

because $g(t)\psi(t)$ is an odd function. Next, using $\cos^2(\theta) = (1/2)(\cos(2\theta) + 1)$,

$$\begin{aligned} \|g - f\|_2^2 &= \varepsilon^2 \int_{-1/2}^{1/2} \cos^2(t/\varepsilon) \phi^2(t) dt \\ &\geq \frac{\varepsilon^2}{2} \|\phi\|_2^2 - \frac{\varepsilon^2}{2} |\mathcal{F}[\phi^2](2\varepsilon^{-1})|. \end{aligned} \quad (31)$$

By the triangle inequality and Parseval's formula,

$$\frac{|\langle \hat{f}_X, g - f \rangle|}{\varepsilon} \leq \left(\underbrace{\int_{|\omega| \leq h\varepsilon^{-1}/2}}_{=:A} + \underbrace{\int_{|\omega| \geq h\varepsilon^{-1}/2}}_{=:B} \right) \left| \mathcal{F}[k] \left(-\frac{h}{\varepsilon} - \omega \right) \frac{1}{h} \mathcal{F}[\phi] \left(-\frac{\omega}{h} \right) \right| d\omega.$$

Moreover,

$$A \leq \frac{1}{2\varepsilon} \|\phi\|_1 \cdot \sup_{|\omega| \geq h\varepsilon^{-1}/2} |\mathcal{F}[k](\omega)|, \quad (32)$$

$$B \leq \|k\|_1 \cdot \int_{|\omega| > \varepsilon^{-1}/2} |\mathcal{F}[\phi](\omega)| d\omega. \quad (33)$$

Then (29) follows from $\|k\|_1 = 1$ and equations (30), (31), (32), and (33). \square

Proposition 4. Let $\varepsilon = n^{-1/3+\gamma}$ for some absolute constant $\gamma > 0$. Let \hat{f}_X denote a uniformly weighted coreset KDE with bandwidth h built from a kernel k on $X = x_1, \dots, x_m$. Suppose that $|\mathcal{F}[k](\omega)| \leq |\omega|^{-\ell}$. If $h \geq c\varepsilon^{1-2/\ell} = cn^{(-1/3+\gamma)(1-2/\ell)}$ for c sufficiently large, then for all m it holds that

$$\sup_{f \in \mathcal{P}_{\mathcal{H}}(\beta, L)} \inf_{X: |X|=m} \|\hat{f}_X - f\|_2 = \Omega(\varepsilon) = \Omega\left(n^{-1/3+\gamma}\right) \quad (34)$$

Proof. The proof is a direct application of Lemma 7. Let $f(t) = g(t) + \varepsilon\phi(t) \cos(t/\varepsilon)$, where we set

$$\phi(t) = -e^{\frac{1}{x(x+1/4)}} \mathbf{1}(x \in [-1/4, 0]) + e^{-\frac{1}{x(x-1/4)}} \mathbf{1}(x \in [0, 1/4]).$$

Observe that ϕ is odd and $\phi \in \mathcal{C}^\infty$. Thus, $\phi^2 \in \mathcal{C}^\infty$, so by the Riemann–Lebesgue lemma (see e.g. Katznelson, 2004, Chapter VI), $\mathcal{F}[\phi^2](\varepsilon^{-1}) \leq 10\varepsilon$. Using a similar argument and noting that $\mathcal{F}[\phi](\omega) = \omega^{-2}\mathcal{F}[\phi''](\omega) \leq 10\omega^{-3}$, we obtain

$$\int_{|\omega| \geq 2\varepsilon^{-1}} |\mathcal{F}[\phi](\omega)| d\omega \leq 100\varepsilon^2.$$

Also $\|\phi\|_2 \geq c'$ for a small absolute constant, and $\|\phi\|_1 \leq 2$.

Thus Lemma 7, the hypothesis on k , and $h \geq c'\varepsilon^{1-2/\ell}$ imply that

$$\|\hat{f}_X - f\|_2^2 \geq \frac{c'^2}{2}\varepsilon^2 - 2\left(\frac{\varepsilon}{h}\right)^\ell - 200\varepsilon^3 = \Omega(\varepsilon^2).$$

Since $f \in \mathcal{P}_{\mathcal{H}}(1, L)$, the statement of the lemma follows. \square

3.2 Proof of Theorem 6

Theorem. Fix $\beta > 0$ and a nonnegative kernel k on \mathbb{R} satisfying the following fast decay and smoothness conditions:

$$\lim_{s \rightarrow +\infty} \frac{1}{s} \log \frac{1}{\int_{|t|>s} k(t) dt} > 0, \quad (35)$$

$$\lim_{\omega \rightarrow \infty} \frac{1}{|\omega|} \log \frac{1}{|\mathcal{F}[k](\omega)|} > 0, \quad (36)$$

where we recall that $\mathcal{F}[k]$ denotes the Fourier transform. Let \hat{f}_S^{unif} be the uniformly weighted coreset KDE. Then there exists $L_\beta > 0$ such that for $L \geq L_\beta$ and any m and $h > 0$, we have

$$\inf_{h, S: |S| \leq m} \sup_{f \in \mathcal{P}_{\mathcal{H}}(\beta, L)} \mathbb{E} \|\hat{f}_S^{\text{unif}} - f\|_2 = \Omega_{\beta, k} \left(\frac{m^{-\frac{\beta}{1+\beta}}}{\log^{\beta+\frac{1}{2}} m} \right).$$

Proof. We follow a similar strategy to the proof of Theorem 5 by handling the cases of small and large bandwidth separately.

Let $q_1 = q_1(k) > 0$ be the minimum number such that $\int_{|t|>q_1} k(t) dt \leq 0.1$. By the assumption in the theorem, there exists $a > 0$ such that

$$\int_{|t|>s} k(t) dt \leq \frac{1}{a} \exp(-as), \quad \forall s \geq 0.$$

Note that we can set $L_\beta^{(1)}$ large such that for any $\delta \in [0, 1]$, there exists $f \in \mathcal{P}_{\mathcal{H}}(\beta, L_\beta^{(1)})$ such that $f(x) = \delta$ for $x \in [0, 1/2]$. We first show that for any given m and h , we have

$$\inf_{S: |S| \leq m} \sup_{f \in \mathcal{P}_{\mathcal{H}}(\beta, L_\beta^{(1)})} \mathbb{E} \|\hat{f}_S^{\text{unif}} - f\|_1 \geq 0.2 \left(1 \wedge \frac{1}{100q_1mh} \right) \mathbf{1} \left\{ h \leq \frac{0.02a}{\log \left(\frac{mq_1}{0.001a} \vee \frac{10}{a} \right)} \wedge 1 \right\}. \quad (37)$$

Let f be an arbitrary function in $f \in \mathcal{P}_{\mathcal{H}}(\beta, L_{\beta}^{(1)})$ such that

$$f(x) = 1 \wedge \frac{1}{100q_1mh}, \quad \forall x \in [0, 1/2].$$

Let T be the set of $i \in S$ for which $x_i \in [q_1h, 1/2 - q_1h]$.

Case 1: $|T| \geq m \left(1 \wedge \frac{1}{100q_1mh}\right)$. Since $k \geq 0$, we have

$$\|\hat{f}_X 1_{[0,1/2]}\|_1 \geq \frac{0.9|T|}{m} \geq 0.9 \left(1 \wedge \frac{1}{100q_1mh}\right).$$

On the other hand,

$$\|f 1_{[0,1/2]}\|_1 \leq \frac{1}{2} \left(1 \wedge \frac{1}{100q_1mh}\right),$$

therefore,

$$\|(\hat{f}_X - f) 1_{[0,1/2]}\|_1 \geq 0.4 \left(1 \wedge \frac{1}{100q_1mh}\right).$$

Case 2: $|T| < m \left(1 \wedge \frac{1}{100q_1mh}\right)$. Define

$$\gamma := 0.1 \left(1 \wedge \frac{1}{100q_1mh}\right)$$

and

$$q_2 := \frac{0.02}{h}.$$

Note that to verify (37) we only need to consider the event of $h \leq \frac{0.02a}{\log(\frac{mq_1}{0.001a} \vee \frac{10}{a})} \wedge 1$, in which case

$$\begin{aligned} \int_{|t| > q_2} k(t) dt &\leq \frac{1}{a} \exp(-aq_2) \\ &\leq \frac{1}{a} \cdot \left(\frac{0.001a}{mq_1} \wedge 0.1a \right) \\ &\leq \frac{1}{a} \cdot \left(\frac{0.001a}{q_1mh} \wedge 0.1a \right) \\ &= 0.1 \left(1 \wedge \frac{1}{100q_1mh}\right) \\ &= \gamma. \end{aligned}$$

Moreover since $\gamma \leq 0.1$ we see that $q_2 \geq q_1$. Now define

$$V := [2hq_2, 1/2 - 2hq_2] \setminus \bigcup_{j \in T} [x_j - q_1h, x_j + q_1h].$$

Then for $j \notin T$, we have

$$\int_V \frac{1}{h} k\left(\frac{x_j - t}{h}\right) dt \leq \gamma$$

while for $j \in T$ we have

$$\int_V \frac{1}{h} k\left(\frac{x_j - t}{h}\right) dt \leq 0.1.$$

Thus,

$$\|\hat{f}_X 1_V\|_1 \leq \frac{0.1|T|}{m} + \gamma \leq 0.2 \left(1 \wedge \frac{1}{100q_1mh} \right).$$

On the other hand, by the union bound we see that the Lebesgue measure of V is at least

$$\frac{1}{2} - 4q_2h - 2q_1h|T| \geq 0.5 - 4q_2h - 0.02 \geq 0.4$$

where we used the fact that $q_2h = 0.02$. Then

$$\|f 1_V\|_1 \geq 0.4 \left(1 \wedge \frac{1}{100q_1mh} \right)$$

and hence

$$\|(\hat{f}_X - f) 1_{[0,1/2]}\|_1 \geq \|(\hat{f}_X - f) 1_V\|_1 \geq 0.2 \left(1 \wedge \frac{1}{100q_1mh} \right).$$

This concludes the proof of (37).

The second step is to show that for given m and h , we have

$$\inf_{S: |S| \leq m} \sup_{f \in \mathcal{P}_{\mathcal{H}}(\beta, L)} \mathbb{E} \|\hat{f}_S^{\text{unif}} - f\|_1 \geq \frac{1}{4} \left(\frac{b(h \wedge 1)}{\log m} \right)^\beta - \frac{1}{bm^2} \quad (38)$$

sufficiently large m and L to be determined later, and $0 < b < \infty$ is such that

$$\mathcal{F}[k](\omega) \leq \frac{1}{b} \exp(-b\omega), \quad \forall \omega \in \mathbb{R}$$

whose existence is guaranteed by the assumption of the theorem. Let ϕ be a smooth, even, nonnegative function supported on $[-1/2, 1/2]$ satisfying $\int_{[-1/2, 1/2]} \phi = 1$. Define

$$f_\epsilon(t) := \phi(t) \left(c_\epsilon + \epsilon^\beta \sin \frac{t}{\epsilon} \right)$$

where $c_\epsilon > 0$ is chosen so that $\int_{[-1/2, 1/2]} f_\epsilon = 1$. Then $\lim_{\epsilon \rightarrow 0} c_\epsilon = 1$, and in particular $f_\epsilon \geq 0$ when $\epsilon < \epsilon(\phi, \beta)$ for some $\epsilon(\phi, \beta)$. Moreover we can find $L_\beta^{(2)} < \infty$ such that $f_\epsilon \in \mathcal{P}_{\mathcal{H}}(\beta, L_\beta^{(2)})$ for all $\epsilon < \epsilon(\phi, \beta)$. Now

$$\begin{aligned} \|f_\epsilon - \hat{f}_X\|_1 &\geq |\mathcal{F}[f_\epsilon](1/\epsilon) - \mathcal{F}[\hat{f}_X](1/\epsilon)| \\ &\geq \left| \int_{[-1/2, 1/2]} f_\epsilon(t) e^{-it/\epsilon} dt \right| - \left| \mathcal{F}[k]\left(\frac{h}{\epsilon}\right) \right| \\ &\geq \left| \int_{[-1/2, 1/2]} f_\epsilon(t) \sin \frac{t}{\epsilon} dt \right| - \left| \mathcal{F}[k]\left(\frac{h}{\epsilon}\right) \right| \\ &= \epsilon^\beta \left| \int_{[-1/2, 1/2]} \phi(t) \sin^2 \frac{t}{\epsilon} dt \right| - \left| \mathcal{F}[k]\left(\frac{h}{\epsilon}\right) \right| \end{aligned} \quad (39)$$

where (39) used the fact that ϕ is even. Since $\lim_{\epsilon \rightarrow 0} \int_{[-1/2, 1/2]} \phi(t) \sin^2 \frac{t}{\epsilon} dt = \frac{1}{2}$, there exists $\epsilon'(\phi)$ such that

$$\int_{[-1/2, 1/2]} \phi(t) \sin^2 \frac{t}{\epsilon} dt \geq \frac{1}{4}$$

for any $\epsilon \leq \epsilon'(\phi)$. Now define

$$\epsilon''(h, m) = \frac{b(h \wedge 1)}{2 \log m}.$$

There exists $m(\phi, \beta, b) < \infty$ such that $\sup_{h>0} \epsilon''(h, m) < \epsilon(\phi, \beta) \wedge \epsilon'(\phi)$ whenever $m \geq m(\phi, \beta, b)$. With the choice of $\epsilon = \epsilon''(h, m)$, we can continue lower bounding (39) as (for $m \geq m(\phi, \beta, b)$):

$$\frac{1}{4} \left(\frac{b(h \wedge 1)}{\log m} \right)^\beta - \frac{1}{bm^2}.$$

Finally, we collect the results for step 1 and step 2. First observe that the main term in the risk in step 1 can be simplified as

$$\begin{aligned} & \left(1 \wedge \frac{1}{100q_1mh} \right) 1 \left\{ h \leq \frac{0.02a}{\log \left(\frac{mq_1}{0.001a} \vee \frac{10}{a} \right)} \wedge 1 \right\} \\ &= \frac{1}{100q_1mh} \wedge 1 \{\mathcal{A}\} \end{aligned} \quad (40)$$

where \mathcal{A} denotes the event in the left side of (40).

Thus up to multiplicative constant depending on k, β , we can lower bound the risk by taking the max of the risks in the two steps:

$$\left(\frac{1}{mh} \wedge 1\{\mathcal{A}\} \right) \vee \left(\left(\frac{b(h \wedge 1)}{\log m} \right)^\beta - \frac{1}{bm^2} \right) \quad (41)$$

whenever $L \geq L_\beta := L_\beta^{(1)} \vee L_\beta^{(2)}$. We can use the distributive law to open up the parentheses in (41). By checking the $h > m^{-\frac{1}{\beta}}$ and $h \leq m^{-\frac{1}{\beta}}$ cases respectively, it is easy to verify that

$$\frac{1}{mh} \vee \left(\left(\frac{b(h \wedge 1)}{\log m} \right)^\beta - \frac{1}{bm^2} \right) = \Omega \left(\frac{m^{-\frac{\beta}{\beta+1}}}{\log^\beta m} \right).$$

Next, if \mathcal{A} is true, we evidently have

$$1\{\mathcal{A}\} \vee \left(\left(\frac{b(h \wedge 1)}{\log m} \right)^\beta - \frac{1}{bm^2} \right) = 1 = \Omega \left(\frac{m^{-\frac{\beta}{\beta+1}}}{\log^\beta m} \right).$$

If \mathcal{A} is not true, then $h > \frac{0.02a}{\log \left(\frac{mq_1}{0.001a} \vee \frac{10}{a} \right)} \wedge 1$, and we have

$$\begin{aligned} 1\{\mathcal{A}\} \vee \left(\left(\frac{b(h \wedge 1)}{\log m} \right)^\beta - \frac{1}{bm^2} \right) &= \left(\left(\frac{b(h \wedge 1)}{\log m} \right)^\beta - \frac{1}{bm^2} \right) \\ &= \Omega \left(\log^{-2\beta} m \right) \\ &= \Omega \left(\frac{m^{-\frac{\beta}{\beta+1}}}{\log^\beta m} \right). \end{aligned}$$

In either case the risk with respect to L_1 is $\Omega \left(\frac{m^{-\frac{\beta}{\beta+1}}}{\log^\beta m} \right)$. It remains to convert this to a lower bound in L^2 .

We consider two cases. First note that by the fast decay condition on the Fourier transform, $k \in \mathcal{C}^1$. Let $B = B_k$ denote a constant such that

$$\sup_{x \in [-1/2, 1/2]} |k'(x)| \leq B. \quad (42)$$

Set $\Delta = B^{1/2} \vee k(0) \vee 1$.

Case 1: $h \leq \Delta$.

Let $U = \{|y| \geq \frac{1}{2} + c_{\beta, \Delta, a} \log m\}$, and let $U^c = \mathbb{R} \setminus U$. If $h \leq \Delta$, then because $X_i \in [-1/2, 1/2]$ and by the exponential decay of k ,

$$\|\hat{f}_X(y) \mathbf{1}_U\|_1 \leq m^{-2}$$

for $c_{\beta,\Delta,a}$ sufficiently large. Thus by Cauchy–Schwarz,

$$\begin{aligned}
\|(\hat{f}_X - f)\mathbf{1}_{U^c}\|_2 &\geq c'_{\beta,\Delta,a}(\log m)^{-1/2}\|(\hat{f}_X - f)\mathbf{1}_{U^c}\|_2 \\
&= c'_{\beta,\Delta,a}(\log m)^{-1/2}\left(\|(\hat{f}_X - f)\|_1 - \|(\hat{f}_X - f)\mathbf{1}_U\|_1\right) \\
&\geq c'_{\beta,\Delta,a}(\log m)^{-1/2}\left(c_{\beta,k}\left(\frac{m^{-\frac{\beta}{\beta+1}}}{\log^\beta m}\right) - m^{-2}\right) \\
&= \Omega\left(\frac{m^{-\frac{\beta}{\beta+1}}}{\log^{\beta+\frac{1}{2}} m}\right)
\end{aligned}$$

Case 2: $h \geq \Delta$

In this case, $k(X_i - y)$ is nearly constant for all i . By (42) and Taylor’s theorem,

$$\left|k(0) - k\left(\frac{X_i - y}{h}\right)\right| \leq 2B$$

for all $y \in [-1/2, 1/2]$ and for all i . Hence, for all $y \in [-1/2, 1/2]$, using $h \geq \Delta$,

$$\hat{f}_X(y) = \frac{1}{mh} \sum_{i=1}^m k\left(\frac{X_i - y}{h}\right) \leq \frac{1}{h}(k(0) + 2B) \leq 3.$$

For L_β large enough, we see that for the function $f \in \mathcal{P}_\mathcal{H}(\beta, L_\beta)$ with $f|_{[0, \frac{1}{100}]} \equiv 4$,

$$\|\hat{f}_X - f\|_2 \geq \|(\hat{f}_X - f)\mathbf{1}_{[0, \frac{1}{100}]}\|_1 = \Omega(1).$$

□

4 PROOFS FROM SECTION 5

4.1 Proof of Theorem 7

The result is restated below.

Theorem. *Let k_s denote the kernel from Section 3 of the main text. The algorithm of Phillips & Tai (2018) yields in polynomial time a subset S with $|S| = m = \tilde{O}(n^{\frac{\beta+d}{2\beta+d}})$ such that the uniformly weighted coresnet KDE \hat{g}_S satisfies*

$$\sup_{f \in \mathcal{P}_\mathcal{H}(\beta, L)} \mathbb{E}\|f - \hat{g}_S\|_2 \leq c_{\beta,d,L} n^{-\frac{\beta}{2\beta+d}}.$$

Proof. Here we adapt the results in Section 2 of Phillips & Tai (2018) to our setting where the bandwidth $h = n^{-1/(2\beta+d)}$ is shrinking. Using their notation, we define $K_s(x, y) = k_s\left(\frac{x-y}{h}\right)$ and study the kernel discrepancy of the kernel K_s . First we verify the assumptions on the kernel (bounded influence, Lipschitz, and positive semidefiniteness) needed to apply their results.

First, the kernel K_s is *bounded influence* (see Phillips & Tai, 2018, Section 2) with constant $c_K = 2$ and $\delta = n^{-1}$, which means that

$$|K_s(x, y)| \leq \frac{1}{n}$$

if $|x - y|_\infty \geq n^2$. This follows from the fast decay of κ_s .

Note that if x and y differ on a single coordinate i , then

$$|k_s(x) - k_s(y)| \leq \left|c(x_i - y_i) \prod_{j \neq i} \kappa_s(x_j)\right| \leq c|x_i - y_i|$$

because $|\kappa_s(x)| \leq \|\psi\|_1$ for all x and the function κ_s is c -Lipschitz for some constant c . Hence by the triangle and Cauchy–Schwarz inequalities, the function k_s is Lipschitz:

$$|k_s(x) - k_s(y)| \leq dc_k |x - y|_1 \leq d^{3/2} c_k |x - y|_2.$$

Therefore the kernel $K_s(x, y)$ is *Lipschitz* (see Phillips & Tai, 2018) with constant $C_K = d^{3/2} c_k h^{-1}$. Moreover, the kernel K_s is *positive semidefinite* because the Fourier transform of κ_s is nonnegative.

Given the shrinking bandwidth $h = n^{-1/(2\beta+d)}$, we slightly modify the lattice used in Phillips & Tai (2018, Lemma 1). Define the lattice

$$\mathcal{L} = \{(i_1\delta, i_2\delta, \dots, i_d\delta) \mid i_j \in \mathbb{Z}\},$$

where

$$\delta = \frac{1}{c_k d^2 n h^{-1}}.$$

The calculation at the top of page 6 of Phillips & Tai (2018, Lemma 1) yields

$$\begin{aligned} \text{disc}(X, \chi, y) &:= \left| \sum_{i=1}^n \chi(X_i) K_s(X_i, y) \right| \\ &\leq \left| \sum_{i=1}^n \chi(X_i) K_s(X_i, y_0) \right| + 1 \end{aligned}$$

where y_0 is the closest point to y in the lattice \mathcal{L} , and χ assigns either $+1$ or -1 to each element of $X = X_1, \dots, X_n$. Moreover, with the bounded influence of K_s , if

$$\min_i |y - X_i|_\infty \geq n^2,$$

then

$$\text{disc}(X, \chi, y) = \left| \sum_{i=1}^n \chi(X_i) K_s(X_i, y) \right| \leq 1.$$

On defining

$$\mathcal{L}_X = \mathcal{L} \cap \{y : \min_i |y - X_i|_\infty \leq n^2\},$$

we see that

$$\max_{y \in \mathbb{R}^d} \text{disc}(X, \chi, y) \leq \max_{y \in \mathcal{L}_X} \text{disc}(X, \chi, y) + 1$$

for all signings $\chi : X \rightarrow \{-1, +1\}$. This is precisely the conclusion of Phillips & Tai (2018, Lemma 1).

This established, the positive definiteness and bounded diagonal entries of K_s and Phillips & Tai (2018, Lemmas 2 and 3) imply that

$$\text{disc}_{K_s} = O(\sqrt{d \log n}).$$

Given $\varepsilon > 0$, the halving algorithm can be applied to K_s as in Phillips & Tai (2018, Corollary 5) to yield a coreset X_S of size $m = O(\varepsilon^{-1} \sqrt{d \log \varepsilon^{-1}})$ such that

$$\left\| \frac{1}{n} \sum_{j=1}^n K_s(X_j, y) - \frac{1}{m} \sum_{j \in S} K_s(X_j, y) \right\|_\infty \leq \varepsilon.$$

Rescaling by h^{-d} , we have

$$\|\hat{f} - \hat{f}_S^{\text{unif}}\|_\infty = \left\| \frac{1}{n} \sum_{j=1}^n k_s(X_j, y) - \frac{1}{m} \sum_{j \in S} k_s(X_j, y) \right\|_\infty \leq \varepsilon h^{-d}.$$

Recall from Section 2.2 of this Supplement that \hat{f} attains the minimax rate of estimation on $\mathcal{P}_{\mathcal{H}}(\beta, L)$. Thus setting $\varepsilon = h^d n^{-\beta/(2\beta+d)}$ we get a coreset of size $\tilde{O}_d(n^{\frac{\beta+d}{2\beta+d}})$ that attains the minimax rate $c_{\beta,d,L} n^{-\beta/(2\beta+d)}$, as desired. Moreover, by the results of Phillips & Tai (2018), this coreset can be constructed in polynomial time. \square

References

- Cover, Thomas M., & Thomas, Joy A. 2006. *Elements of information theory*. Second edn. Hoboken, NJ: Wiley-Interscience [John Wiley & Sons].
- Katznelson, Yitzhak. 2004. *An Introduction to Harmonic Analysis*. 3 edn. Cambridge Mathematical Library. Cambridge University Press.
- Phillips, Jeff M., & Tai, Wai Ming. 2018. Near-Optimal Coresets of Kernel Density Estimates. *Pages 66:1–66:13 of: 34th International Symposium on Computational Geometry, SoCG 2018, June 11-14, 2018, Budapest, Hungary*.
- Tsybakov, Alexandre B. 2009. *Introduction to Nonparametric Estimation*. Springer series in statistics. Springer.