SUPPLEMENTARY MATERIAL FOR 'OPTIMAL SUBGROUP SELECTION'

BY HENRY W. J. REEVE[1,a], TIMOTHY I. CANNINGS[2,b] AND RICHARD J. SAMWORTH[3,c]

[1]*School of Mathematics, University of Bristol*
[a]*henry.reeve@bristol.ac.uk*

[2]*School of Mathematics and Maxwell Institute for Mathematical Sciences, The University of Edinburgh*
[b]*timothy.cannings@ed.ac.uk*[*]

[3]*Statistical Laboratory, University of Cambridge*
[c]*r.samworth@statslab.cam.ac.uk*[†]

This is the supplementary material for Reeve, Cannings and Samworth (2023).

S1. Proof of the hardness result.

PROOF OF PROPOSITION 1. Fix $\epsilon \in (0,1)$, and let $(P_x)_{x\in\mathbb{R}^d}$ be a disintegration of P into conditional probability measures on $[0,1]$; see Section S8.1 and Lemma S34. Since $\mathcal{A}$ has finite VC dimension, it follows from the Vapnik–Chervonenkis concentration inequality (Lemma S36) that there exists a finite set $\mathbb{T} \subseteq \mathbb{R}^d$ for which

$$\sup_{A\in\mathcal{A}} \left| \frac{1}{|\mathbb{T}|} \sum_{t\in\mathbb{T}} \mathbb{1}_{\{t\in A\}} - \mu(A) \right| \leq \epsilon. \tag{S1}$$

Since $\mathbb{T}$ is finite and μ has no atoms, we may choose a radius $r > 0$ sufficiently small that $\mu\big(\bigcup_{t\in\mathbb{T}} B_r(t)\big) \leq \epsilon$. Now define a function $\rho : \mathbb{R}^d \to [0,1]$ by $\rho(x) := 1 \wedge \bigwedge_{t\in\mathbb{T}}\{(2/r)\cdot \|x-t\|_\infty\}$, noting that ρ is Lipschitz. Further, define a family of probability distributions $(Q_x)_{x\in\mathbb{R}^d}$ on $[0,1]$ by

$$\int_{[0,1]} h(y)\, dQ_x(y) = \int_{[0,1]} h\big(\rho(x)\cdot y\big)\, dP_x(y),$$

for all Borel functions $h : [0,1] \to [0,1]$, and define a probability distribution Q on $\mathbb{R}^d \times [0,1]$ by $Q(A\times B) = \int_A Q_x(B)\, d\mu(x)$. It follows that $(Q_x)_{x\in\mathbb{R}^d}$ is a disintegration of Q into conditional probability measures on $[0,1]$. In addition, taking a random pair $(X^Q, Y^Q) \sim Q$ we see by (S40) that for μ-almost every $x \in \mathbb{R}^d$,

$$\eta_Q(x) = \mathbb{E}\big(Y^Q \mid X^Q = x\big) = \int_{[0,1]} y\, dQ_x(y) = \rho(x)\cdot \int_{[0,1]} y\, dP_x(y) = \rho(x)\cdot \eta_P(x).$$

Hence, we may extend the definition of η_Q to $\mathbb{R}^d$ in such a way that $\eta_Q(\cdot) = \rho(\cdot)\eta_P(\cdot)$, which is a product of Lipschitz functions, so is itself Lipschitz; thus, $Q \in \mathcal{P}_{\mathrm{Lip}}(\mu)$. Note also that for every $t \in \mathbb{T}$, we have $\eta_Q(t) = \rho(t)\eta_P(t) = 0 < \tau$, so $\mathbb{T} \cap \mathcal{X}_\tau(\eta_Q) = \emptyset$. Moreover, $\eta_Q(x) \leq \eta_P(x)$ for all $x \in \mathbb{R}^d$, so $\mathcal{X}_\tau(\eta_Q) \subseteq \mathcal{X}_\tau(\eta_P)$. Hence, since $\hat{A}$ controls the Type I error at the level α over $\mathcal{P}_{\mathrm{Lip}}(\mu)$, we have

$$\begin{aligned}\mathbb{P}_Q\big\{\hat{A} \subseteq (\mathbb{R}^d\setminus\mathbb{T}) \cap \mathcal{X}_\tau(\eta_P)\big\} &\geq \mathbb{P}_Q\big\{\hat{A} \subseteq (\mathbb{R}^d\setminus\mathbb{T}) \cap \mathcal{X}_\tau(\eta_Q)\big\} = \mathbb{P}_Q\big\{\hat{A} \subseteq \mathcal{X}_\tau(\eta_Q)\big\} \\ &\geq 1-\alpha. \end{aligned} \tag{S2}$$

Now $Q_x = P_x$ for $x \notin \bigcup_{t\in\mathbb{T}} B_r(t)$. Hence

$$\mathrm{H}^2(P,Q) \leq 2\mathrm{TV}(P,Q) \leq 2\mu\biggl(\bigcup_{t\in\mathbb{T}} B_r(t)\biggr) \leq 2\epsilon,$$

so

(S3)

$$\mathrm{TV}^2(P^{\otimes n},Q^{\otimes n}) \leq \mathrm{H}^2(P^{\otimes n},Q^{\otimes n}) = 2\biggl\{1-\prod_{i=1}^{n}\biggl(1-\frac{\mathrm{H}^2(P,Q)}{2}\biggr)\biggr\} \leq 2\bigl(1-(1-\epsilon)^n\bigr).$$

Note that by (S1) if $A \in \mathcal{A}$ satisfies $A \cap \mathbb{T} = \emptyset$, then $\mu(A) \leq \epsilon$. Hence, by (S2) and (S3), we have

$$\begin{aligned}\mathbb{P}_P\bigl(\{\mu(\hat{A}) \leq \epsilon\} \cap \{\hat{A} \subseteq \mathcal{X}_\tau(\eta)\}\bigr) &\geq \mathbb{P}_P\bigl(\hat{A} \subseteq (\mathbb{R}^d \setminus \mathbb{T}) \cap \mathcal{X}_\tau(\eta_P)\bigr)\\ &\geq 1-\alpha-\sqrt{2\bigl(1-(1-\epsilon)^n\bigr)}.\end{aligned}$$

Letting $\epsilon \searrow 0$ gives (3). Thus,

$$R_\tau(\hat{A}) = M_\tau - \mathbb{E}_P\bigl(\mu(\hat{A}) \mid \hat{A} \subseteq \mathcal{X}_\tau(\eta)\bigr) \geq (1-\alpha)\cdot M_\tau.$$

Finally, note that for any $\xi > 0$, we may take $A_\xi \in \mathcal{A}$ with $\mu(A_\xi) > M_\tau - \xi$ and $A_\xi \subseteq \mathcal{X}_\tau(\eta)$. Hence, we may define $\bar{A} \in \bar{\mathcal{A}}$ that takes the value A_ξ with probability α and $\emptyset$ otherwise; it has regret $R_\tau(\hat{A}) < (1-\alpha)\cdot M_\tau + \alpha\cdot\xi$. Letting $\xi \searrow 0$ yields the final equality in (4). □

S2. Proof of the upper bound in Theorem 2. Recall that Theorem 2(i) will follow from Lemma 3, together with Propositions 4 and 5.

PROOF OF LEMMA 3. We begin by showing that $\sup_{x\in B}\eta(x) \leq t := \tau + \lambda\cdot\mathrm{diam}_\infty(B)^\beta$. Indeed, suppose for a contradiction that there exists some $x_0 \in B$ with $\eta(x_0) > t$. Since $B \not\subseteq \mathcal{X}_\tau(\eta)$ there also exists $x_1 \in B$ with $\eta(x_1) \leq \tau$. Since η is continuous there exists x_2 on the line segment between x_0 and x_1 with $\eta(x_2) = \tau$. Thus, since $x_0, x_2 \in \mathcal{X}_\tau(\eta)$ we have,

$$\begin{aligned}\eta(x_0) &\leq \eta(x_2) + |\eta(x_0)-\eta(x_2)| \leq \tau + \lambda\cdot\|x_0-x_2\|_\infty^\beta\\ &\leq \tau + \lambda\cdot\|x_0-x_1\|_\infty^\beta \leq \tau + \lambda\cdot\mathrm{diam}_\infty(B)^\beta = t < \eta(x_0),\end{aligned}$$

a contradiction which proves the claim $\sup_{x\in B}\eta(x) \leq t$. Now let $m := n\cdot\hat{\mu}_n(B) = \sum_{i\in[n]}\mathbb{1}_{\{X_i\in B\}}$. If $m = 0$, then $\hat{p}_n(B) = 1$, so we may assume without loss of generality that $m \geq 1$. Let $(i_j)_{j\in[m]}$ denote a strictly increasing sequence such that $X_{i_j} \in B$ for all $j \in [m]$. For each $j \in [m]$ let $Z_j := Y_{i_j}$ so that

$$\mathbb{E}\bigl(Z_j \mid \mathcal{D}_X\bigr) = \mathbb{E}\bigl(Y_{i_j} \mid \mathcal{D}_X\bigr) = \eta(X_{i_j}) \leq t.$$

Moreover, $(Z_j)_{j\in[m]}$ are conditionally independent given $\mathcal{D}_X$. Writing $\bar{Z} := m^{-1}\sum_{j\in[m]} Z_j$, we have by construction of $\hat{p}_n(B)$ that

$$\begin{aligned}\mathbb{P}\bigl(\hat{p}_n(B) \leq \alpha \mid \mathcal{D}_X\bigr) &= \mathbb{P}\Bigl[\exp\bigl\{-n\cdot\hat{\mu}_n(B)\cdot\mathrm{kl}\bigl(\hat{\eta}_n(B),t\bigr)\bigr\} \leq \alpha \text{ and } \hat{\eta}_n(B) > t \Bigm| \mathcal{D}_X\Bigr]\\ &= \mathbb{P}\bigl\{\mathrm{kl}\bigl(\bar{Z},t\bigr) \geq m^{-1}\cdot\log(1/\alpha) \text{ and } \bar{Z} > t \bigm| \mathcal{D}_X\bigr\} \leq \alpha,\end{aligned}$$

where the final inequality follows from a Chernoff bound, stated for convenience as Lemma S38. □

The proof of Proposition 4 will rely on the following lemma.

LEMMA S1. *Fix* $\beta \in (0,1]$, $\lambda \geq 1$ *and* $P \in \mathcal{P}_{\mathrm{Höl}}(\beta,\lambda,\tau)$ *with* $\eta \in \mathcal{F}_{\mathrm{Höl}}\big(\beta,\lambda,\mathcal{X}_\tau(\eta)\big)$. *Then, with* $(B_{(\ell)})_{\ell\in[\hat{L}]}$, ℓ_α *and* m *as in Algorithm 1 (and setting* $\ell_\alpha := 0$ *when* $\hat{L}\cdot\hat{p}_n(B_{(1)}) > \alpha$*), we have*

$$\mathbb{P}\bigg(\bigcup_{\ell\in[\ell_\alpha]} B_{(\ell)} \not\subseteq \mathcal{X}_\tau(\eta) \,\bigg|\, \mathcal{D}_X\bigg) \leq \alpha.$$

PROOF. Let $\mathcal{N}(\mathcal{D}_X) := \{B \in \mathcal{H}(\mathcal{D}_X) : B \not\subseteq \mathcal{X}_\tau(\eta)\}$ and $K := |\mathcal{N}(\mathcal{D}_X)|$. Note that

$$\bigg\{\bigcup_{\ell\in[\ell_\alpha]} B_{(\ell)} \not\subseteq \mathcal{X}_\tau(\eta)\bigg\}\bigcap\{K=0\} \subseteq \bigg\{\bigcup_{\ell\in[\hat{L}]} B_{(\ell)} \not\subseteq \mathcal{X}_\tau(\eta)\bigg\}\bigcap\{K=0\} = \emptyset.$$

On the other hand, when $K \geq 1$, we may write

$$\tilde{\ell} := \min\big\{\ell \in [\hat{L}] \;:\; B_{(\ell)} \in \mathcal{N}(\mathcal{D}_X)\big\},$$

so that when $\hat{L}\cdot\hat{p}_n(B_{(1)}) \leq \alpha$, we have $\bigcup_{\ell\in[\ell_\alpha]} B_{(\ell)} \not\subseteq \mathcal{X}_\tau(\eta)$ if and only if $\tilde{\ell} \leq \ell_\alpha$. When $K \geq 1$, we have by the minimality of $\tilde{\ell}$ that $\tilde{\ell} \leq \hat{L} + 1 - K$, so

$$\begin{aligned}
&\mathbb{P}\bigg(\bigcup_{\ell\in[\ell_\alpha]} B_{(\ell)} \not\subseteq \mathcal{X}_\tau(\eta) \,\bigg|\, \mathcal{D}_X\bigg) \\
&\qquad= \mathbb{P}\bigg(\bigg\{\bigcup_{\ell\in[\ell_\alpha]} B_{(\ell)} \not\subseteq \mathcal{X}_\tau(\eta)\bigg\}\bigcap\{\hat{L}\cdot\hat{p}_n(B_{(1)}) \leq \alpha\} \,\bigg|\, \mathcal{D}_X\bigg) \\
&\qquad= \mathbb{P}\big(\{\tilde{\ell} \leq \ell_\alpha\} \cap \{\hat{L}\cdot\hat{p}_n(B_{(1)}) \leq \alpha\} \,\big|\, \mathcal{D}_X\big) \\
&\qquad\leq \mathbb{P}\big(\{(\hat{L}+1-\tilde{\ell})\cdot\hat{p}_n(B_{(\tilde{\ell})}) \leq \alpha\} \cap \{\hat{L}\cdot\hat{p}_n(B_{(1)}) \leq \alpha\} \,\big|\, \mathcal{D}_X\big) \\
&\qquad\leq \mathbb{P}\bigg(K\cdot \min_{B\in\mathcal{N}(\mathcal{D}_X)} \hat{p}_n(B) \leq \alpha \,\bigg|\, \mathcal{D}_X\bigg) \\
&\qquad\leq \sum_{B\in\mathcal{N}(\mathcal{D}_X)} \mathbb{P}\bigg(\hat{p}_n(B) \leq \frac{\alpha}{K} \,\bigg|\, \mathcal{D}_X\bigg) \leq \alpha,
\end{aligned}$$

where we applied Lemma 3 for the final inequality. □

PROOF OF PROPOSITION 4. By construction in Algorithm 1, we have $\hat{A}_{\mathrm{OSS}}(\mathcal{D}) \subseteq \bigcup_{\ell\in[\ell_\alpha]} B_{(\ell)}$. Hence the result follows from Lemma S1. □

We now turn to the proof of Proposition 5. A key component of this result is the following proposition, which states that if a set $A \in \mathcal{A}$ may be covered with a finite collection of hyper-cubes $\{B_1,\ldots,B_L\} \subseteq \mathcal{H}$, each with sufficiently large diameter and μ-measure, in such a way that η is well above the level τ on each B_ℓ, then $\hat{A}_{\mathrm{OSS}}$ will return a set of μ-measure comparable with $\mu(A)$.

PROPOSITION S2. *Take* $\alpha \in (0,1)$, $n \in \mathbb{N}$, $\delta \in (0,1)$, $(\beta,\lambda) \in (0,1]\times(0,\infty)$, $P \in \mathcal{P}_{\mathrm{Höl}}(\beta,\lambda)$ *and* $\mathcal{A} \subseteq \mathcal{B}(\mathbb{R}^d)$ *with* $\dim_{\mathrm{VC}}(\mathcal{A}) < \infty$ *and* $\emptyset \in \mathcal{A}$. *Given* $L \in \mathbb{N}$, *suppose that there exist hyper-cubes* $\{B_1,\ldots,B_L\} \subseteq \mathcal{H}$ *such that* $\min_{q\in[L]} \mu(B_q) \geq 8\log(4L/\delta)/n$, $\min_{q\in[L]} \mathrm{diam}_\infty(B_q) \geq 1/n$ *and*

(S4)

$$\min_{q\in[L]}\bigg\{\sup_{x\in B_q}\eta(x) - 2\lambda\cdot\mathrm{diam}_\infty(B_q)^\beta - \sqrt{\frac{2\log\big(2^{2+d}L\cdot n(2+\log_2 n)/(\alpha\cdot\delta)\big)}{n\cdot\mu(B_q)}}\bigg\} \geq \tau.$$

Let $S^\dagger := \bigcup_{q\in[L]} B_q$*, and taking the universal constant* $C_{\mathrm{VC}} > 0$ *from Lemma S36, let*

$$J_{n,\delta}(S^\dagger) := \sup\big\{\mu(A) : A \in \mathcal{A} \cap \mathrm{Pow}(S^\dagger)\big\} - 2C_{\mathrm{VC}}\sqrt{\frac{\dim_{\mathrm{VC}}(\mathcal{A})}{n}} - \sqrt{\frac{2\log(2/\delta)}{n}}.$$

Then

$$\mathbb{P}\big\{\mu\big(\hat{A}_{\mathrm{OSS}}(\mathcal{D})\big) < J_{n,\delta}(S^\dagger)\big\} \leq \delta.$$

Proposition S2 will be proved through a series of lemmas below.

LEMMA S3. *Let* P *be a distribution on* $\mathbb{R}^d \times [0,1]$ *having marginal* μ *on* $\mathbb{R}^d$*, and let* $\delta \in (0,1)$*,* $n \in \mathbb{N}$ *and* $L \in \mathbb{N}$*. Suppose further that* $\{B_1,\ldots,B_L\} \subseteq \mathcal{H}$ *with* $\min_{q\in[L]} \mu(B_q) \geq 8\log(4L/\delta)/n$*, and define the event*

$$\mathcal{E}_{1,\delta} := \bigg\{\min_{q\in[L]}\bigg(\hat{\mu}_n(B_q) - \frac{\mu(B_q)}{2}\bigg) > 0\bigg\}.$$

Then $\mathbb{P}(\mathcal{E}_{1,\delta}^{\mathrm{c}}) \leq \delta/4$*.*

PROOF. By the multiplicative Chernoff bound (Lemma S39), for each $q \in [L]$,

$$\mathbb{P}\bigg(\hat{\mu}_m(B_q) \leq \frac{\mu(B_q)}{2}\bigg) = \mathbb{P}\bigg(\sum_{i=1}^m \mathbb{1}_{\{X_i\in B_q\}} \leq \frac{m}{2}\cdot\mu(B_q)\bigg) \leq \exp\bigg(-\frac{m}{8}\cdot\mu(B_q)\bigg) \leq \frac{\delta}{4L}.$$

The result therefore follows by a union bound. □

LEMMA S4. *Let* P *be a distribution on* $\mathbb{R}^d \times [0,1]$ *having regression function* $\eta : \mathbb{R}^d \to [0,1]$*, and let* $\delta \in (0,1)$*,* $n \in \mathbb{N}$ *and* $L \in \mathbb{N}$*. Suppose that* $\{B_1,\ldots,B_L\} \subseteq \mathcal{H}$ *and define the event*

$$\mathcal{E}_{2,\delta} := \bigg\{\max_{q\in[L]}\bigg(\inf_{x\in B_q}\eta(x) - \hat{\eta}_n(B_q) - \sqrt{\frac{\log(4L/\delta)}{2n\cdot\hat{\mu}_n(B_q)}}\bigg) < 0\bigg\},$$

where the empirical distribution $\hat{\mu}_n$ *and empirical regression function* $\hat{\eta}_n$ *are defined in* (7) *and* (8) *respectively. Then* $\mathbb{P}(\mathcal{E}_{2,\delta}^{\mathrm{c}}) \leq \delta/4$*.*

PROOF. By Hoeffding's inequality (Lemma S38), for every $q \in [L]$, we have

$$\operatorname{ess\,sup}\,\mathbb{P}\bigg(\hat{\eta}_n(B_q) \leq \inf_{x\in B_q}\eta(x) - \sqrt{\frac{\log(4L/\delta)}{2n\cdot\hat{\mu}_n(B_q)}}\;\bigg|\;\mathcal{D}_X\bigg) \leq \frac{\delta}{4L}.$$

The result now follows by the law of total expectation, combined with a union bound. □

LEMMA S5. *Let* $(\beta,\lambda) \in (0,1] \times [1,\infty)$*,* $P \in \mathcal{P}_{\mathrm{H\ddot{o}l}}(\beta,\lambda)$*,* $\delta \in (0,1)$*,* $n \in \mathbb{N}$*,* $L \in \mathbb{N}$ *and* $\xi \in (0,1)$*. Suppose that for some* $\{B_1,\ldots,B_L\} \subseteq \mathcal{H}$ *we have* $\min_{q\in[L]}\mu(B_q) \geq 8\log(8L/\delta)/n$ *and*

$$\min_{q\in[L]}\bigg\{\sup_{x\in B_q}\eta(x) - 2\lambda\cdot\mathrm{diam}_\infty(B_q)^\beta - \sqrt{\frac{2\log\big(4L/(\xi\cdot\delta)\big)}{n\cdot\mu(B_q)}}\bigg\} \geq \tau. \tag{S5}$$

Then, recalling the definition of the p-values $\hat{p}_n$ *from* (9)*, we have*

$$\mathbb{P}\bigg(\max_{q\in[L]}\hat{p}_n(B_q) \geq \xi\bigg) \leq \frac{\delta}{2}.$$

PROOF. By Lemmas S3 and S4, we have $\mathbb{P}(\mathcal{E}_{1,\delta}^{c} \cup \mathcal{E}_{2,\delta}^{c}) \leq \delta/2$. On $\mathcal{E}_{1,\delta} \cap \mathcal{E}_{2,\delta}$, we have for each $q \in [L]$ that

$$\begin{aligned}\hat{\eta}_n(B_q) &> \inf_{x \in B_q} \eta(x) - \sqrt{\frac{\log(4L/\delta)}{n \cdot \mu(B_q)}} \\ &\geq \sup_{x \in B_q} \eta(x) - \lambda \cdot \mathrm{diam}_\infty(B_q)^\beta - \sqrt{\frac{\log(4L/\delta)}{n \cdot \mu(B_q)}} \\ &\geq \tau + \lambda \cdot \mathrm{diam}_\infty(B_q)^\beta + \sqrt{\frac{\log(1/\xi)}{n \cdot \mu(B_q)}},\end{aligned}$$

where we used the fact that $P \in \mathcal{P}_{\text{Höl}}(\beta, \lambda)$, (S5) and the fact that $\sqrt{2(a+b)} \geq \sqrt{a} + \sqrt{b}$ for all $a, b \geq 0$. Thus, on $\mathcal{E}_{1,\delta} \cap \mathcal{E}_{2,\delta}$, we have for every $q \in [L]$ that

$$\hat{p}_n(B_q) \leq \exp\biggl(-\frac{n \cdot \mu(B_q)}{2} \cdot \mathrm{kl}\bigl\{\hat{\eta}_n(B_q), \tau + \lambda \cdot \mathrm{diam}_\infty(B_q)^\beta\bigr\}\biggr) < \xi,$$

as required, where the final bound uses Pinsker's inequality. □

We can now complete the proof of Proposition S2 before returning to complete the proof of Proposition 5.

PROOF OF PROPOSITION S2. We begin by defining events

$$\mathcal{E}_{\mathrm{PV}} := \left\{ \max_{q \in [L]} \hat{p}_n(B_q) < \frac{\alpha}{2^d n(2 + \log_2 n)} \right\},$$
$$\mathcal{E}_{\mathrm{VC}} := \left\{ \sup_{A \in \mathcal{A}} \bigl|\hat{\mu}_n(A) - \mu(A)\bigr| \leq C_{\mathrm{VC}} \sqrt{\frac{\dim_{\mathrm{VC}}(\mathcal{A})}{n}} + \sqrt{\frac{\log(2/\delta)}{2n}} \right\}.$$

By Lemma S5, with $\xi = \alpha / \bigl(2^d n(2 + \log_2 n)\bigr) \in (0,1)$, and Lemma S36, we have $\mathbb{P}(\mathcal{E}_{\mathrm{PV}}^{c} \cup \mathcal{E}_{\mathrm{VC}}^{c}) \leq \delta$. On $\mathcal{E}_{\mathrm{PV}}$ we have each $\hat{p}_n(B_q) < 1$, which implies $\hat{\mu}_n(B_q) > 0$, and since $\mathrm{diam}_\infty(B_1) \geq 1/n$ we deduce that $B_q \in \mathcal{H}(\mathcal{D}_X)$.

Now $\hat{L} = |\mathcal{H}(\mathcal{D}_X)| \leq 2^d n(2 + \log_2 n)$, so on the event $\mathcal{E}_{\mathrm{PV}}$, for each $q \in [L]$, we have $\hat{L} \cdot \hat{p}_n(B_q) \leq \alpha$, and hence $B_q = B_{(\ell(q))}$ for some $\ell(q) \leq \ell_\alpha$. Thus, on the event $\mathcal{E}_{\mathrm{PV}}$ we have $S^\dagger \subseteq \bigcup_{\ell \in [\ell_\alpha]} B_{(\ell)}$. Now take $\zeta > 0$ and choose $A_\zeta^* \in \mathcal{A} \cap \mathrm{Pow}(S^\dagger)$ with $\mu(A_\zeta^*) > \sup\bigl\{\mu(A) : A \in \mathcal{A} \cap \mathrm{Pow}(S^\dagger)\bigr\} - \zeta$. It follows that $A_\zeta^* \in \mathcal{A} \cap \mathrm{Pow}\bigl(\bigcup_{\ell \in [\ell_\alpha]} B_{(\ell)}\bigr)$ and hence on the event $\mathcal{E}_{\mathrm{PV}} \cap \mathcal{E}_{\mathrm{VC}}$ that

$$\begin{aligned}\mu(\hat{A}_{\mathrm{OSS}}) &\geq \hat{\mu}_n(\hat{A}_{\mathrm{OSS}}) - C_{\mathrm{VC}} \sqrt{\frac{\dim_{\mathrm{VC}}(\mathcal{A})}{n}} - \sqrt{\frac{\log(2/\delta)}{2n}} \\ &\geq \hat{\mu}_n(A_\zeta^*) - C_{\mathrm{VC}} \sqrt{\frac{\dim_{\mathrm{VC}}(\mathcal{A})}{n}} - \sqrt{\frac{\log(2/\delta)}{2n}} \\ &\geq \mu(A_\zeta^*) - 2C_{\mathrm{VC}} \sqrt{\frac{\dim_{\mathrm{VC}}(\mathcal{A})}{n}} - \sqrt{\frac{2\log(2/\delta)}{n}} \geq J_{n,\delta}(S^\dagger) - \zeta.\end{aligned}$$

Letting $\zeta \searrow 0$, we conclude that $\mu(\hat{A}_{\mathrm{OSS}}) \geq J_{n,\delta}(S^\dagger)$, on the event $\mathcal{E}_{\mathrm{PV}} \cap \mathcal{E}_{\mathrm{VC}}$, as required. □

PROOF OF PROPOSITION 5. We define $\rho := \kappa(2\beta + d) + \beta\gamma$,

$$\theta := \frac{8\lambda^{d/\beta}}{n} \log_+\biggl(\frac{4 \cdot 3^d \cdot n}{\alpha \wedge \delta}\biggr),$$

$r_* := \lambda^{-1/\beta}\theta^{\kappa/\rho}$, $\xi := \theta^{\beta\gamma/\rho}$ and $\Delta := 2^4\theta^{\beta\kappa/\rho}$. We initially assume that $\Delta \leq 1$, so that $1/n \leq \lambda^{-d/\beta} \cdot \theta \leq r_* \leq 2^{-4}$. Now choose a maximal subset $\{x_1, \ldots, x_L\} \subseteq \mathcal{X}_\xi(\omega) \cap \mathcal{X}_{\tau+\Delta}(\eta)$ with the property that $\|x_q - x_{q'}\|_\infty > r_*$ for distinct $q, q' \in [L]$. Then $\mathcal{X}_\xi(\omega) \cap \mathcal{X}_{\tau+\Delta}(\eta) \subseteq \bigcup_{q\in[L]} \bar{B}_{r_*}(x_q)$ and $\bar{B}_{r_*/3}(x_q) \cap \bar{B}_{r_*/3}(x_{q'}) = \emptyset$ for distinct $q, q' \in [L]$. Now, since $\xi \leq 1$,

$$\begin{aligned} L \leq \sum_{q=1}^{L} \frac{\mu\bigl(\bar{B}_{r_*/3}(x_q)\bigr)}{\xi \cdot (r_*/3)^d} &= \frac{1}{\xi \cdot (r_*/3)^d} \cdot \mu\biggl(\bigcup_{q=1}^{L} \bar{B}_{r_*/3}(x_q)\biggr) \\ &\leq \frac{(3\lambda^{1/\beta})^d}{\theta^{(\beta\gamma+d\kappa)/\rho}} \leq \frac{3^d\lambda^{d/\beta}}{\theta} \leq 3^d n. \end{aligned}$$

For each $q \in [L]$ we can find $B_q \in \mathcal{H}$ such that $\bar{B}_{r_*}(x_q) \subseteq B_q$ and such that $r_* \leq \mathrm{diam}_\infty(B_q) \leq 2^{-\lceil \log_2(\frac{1}{2r_*})\rceil+1} \leq 4r_*$, which is possible since $r_* \leq 1/4$. We then have that for every $q \in [L]$,

$$\mu(B_q) \geq \mu\bigl(\bar{B}_{r_*}(x_q)\bigr) \geq \xi \cdot r_*^d \geq \frac{\theta}{\lambda^{d/\beta}} \geq \frac{8}{n}\log(4L/\delta).$$

Hence

$$\begin{aligned} &\min_{q\in[L]}\biggl\{\sup_{x\in B_q} \eta(x) - 2\lambda \cdot \mathrm{diam}_\infty(B_q)^\beta - \sqrt{\frac{2\log\bigl(2^{2+d}L \cdot n(2+\log_2 n)/(\alpha\cdot\delta)\bigr)}{n\cdot\mu(B_q)}}\biggr\} \\ &\geq \min_{q\in[L]}\biggl\{\eta(x_q) - 2^{1+2\beta}\cdot\lambda\cdot r_*^\beta - \sqrt{\frac{2\log\bigl(2^{3+d}3^d n^2\log_2(n)/(\alpha\cdot\delta)\bigr)}{n\cdot\xi\cdot r_*^d}}\biggr\} \\ &\geq \tau + \Delta - 2^3\cdot\lambda\cdot r_*^\beta - \sqrt{\frac{\theta}{\xi\cdot r_*^d\cdot\lambda^{d/\beta}}} \geq \tau, \end{aligned}$$

so (S4) holds. Thus, taking $S^\dagger := \bigcup_{q\in[L]} B_q \supseteq \mathcal{X}_\xi(\omega)\cap\mathcal{X}_{\tau+\Delta}(\eta)$, when $\Delta \leq 1$ we may apply Proposition S2 to see that with probability at least $1-\delta$, we have

$$\begin{aligned} \mu(\hat{A}_{\mathrm{OSS}}) &\geq \sup\bigl\{\mu(A) : A \in \mathcal{A}\cap\mathrm{Pow}(S^\dagger)\bigr\} - 2C_{\mathrm{VC}}\sqrt{\frac{\dim_{\mathrm{VC}}(\mathcal{A})}{n}} - \sqrt{\frac{2\log(2/\delta)}{n}} \\ &\geq \sup\bigl\{\mu(A) : A \in \mathcal{A}\cap\mathrm{Pow}\bigl(\mathcal{X}_\xi(\omega)\cap\mathcal{X}_{\tau+\Delta}(\eta)\bigr)\bigr\} - 2C_{\mathrm{VC}}\sqrt{\frac{\dim_{\mathrm{VC}}(\mathcal{A})}{n}} - \sqrt{\frac{2\log(2/\delta)}{n}} \end{aligned}$$

(S6)

$$\begin{aligned} &\geq M_\tau - C_{\mathrm{App}}\cdot(\xi^\kappa + \Delta^\gamma) - 2C_{\mathrm{VC}}\sqrt{\frac{\dim_{\mathrm{VC}}(\mathcal{A})}{n}} - \sqrt{\frac{2\log(2/\delta)}{n}} \\ &= M_\tau - C_{\mathrm{App}}\cdot(1+2^{5\gamma})\cdot\theta^{\beta\kappa\gamma/\rho} - 2C_{\mathrm{VC}}\sqrt{\frac{\dim_{\mathrm{VC}}(\mathcal{A})}{n}} - \sqrt{\frac{2\log(2/\delta)}{n}} \end{aligned}$$

(S7)

$$\geq M_\tau - C\biggl\{\biggl(\frac{\lambda^{d/\beta}\cdot\log_+\bigl(n/(\alpha\wedge\delta)\bigr)}{n}\biggr)^{\frac{\beta\kappa\gamma}{\kappa(2\beta+d)+\beta\gamma}} + \biggl(\frac{\log_+(1/\delta)}{n}\biggr)^{1/2}\biggr\},$$

where $C \geq 1$ depends only on d, κ, γ, C_{App} and $\dim_{\mathrm{VC}}(\mathcal{A})$. Finally, if $\Delta > 1$, then (S6) holds because $\mu(\hat{A}_{\mathrm{OSS}}) \geq 0$, $M_\tau \leq 1$ and $C_{\mathrm{App}} \geq 1$, so (S7) holds too. This completes the proof of the first claim of the proposition.

For the second claim, observe by Proposition 4 that for $\alpha \in (0, 1/2]$,

$$R_\tau(\hat{A}_{\mathrm{OSS}}) \leq \frac{M_\tau - \mathbb{E}\mu(\hat{A}_{\mathrm{OSS}})}{\mathbb{P}\big(\hat{A}_{\mathrm{OSS}} \subseteq \mathcal{X}_\tau(\eta)\big)} \leq \frac{M_\tau - \mathbb{E}\mu(\hat{A}_{\mathrm{OSS}})}{1-\alpha}$$
$$\leq \tilde{C}\biggl\{\biggl(\frac{\lambda^{d/\beta} \cdot \log_+(n/\alpha)}{n}\biggr)^{\frac{\beta\kappa\gamma}{\kappa(2\beta+d)+\beta\gamma}} + \frac{1}{n^{1/2}}\biggr\},$$

where the final bound follows by integrating the tail bound in the first part of the proposition. □

S3. Proofs of claims in Examples 1, 2 and 3 and a related result. **Example 1:** The marginal density of X is convex on $(-\infty, -\nu-2]$ and on $[\nu+2, \infty)$, so writing ϕ for the standard normal density, we have $\omega(x) = \phi(x-\nu) + \phi(x+\nu)$ for $|x| \geq \nu + 2$. Hence there exists $\xi_0 \in \big(0, 1/\sqrt{2\pi}\big]$, depending only on ν, such that for $\xi \in (0, \xi_0]$ we have $\mathcal{X}_\xi(\omega) = [-x_\xi, x_\xi]$, where $x_\xi \in \Big[\nu + \sqrt{2\log\big(\frac{1}{(2\pi)^{1/2}\xi}\big)}, \nu + \sqrt{2\log\big(\frac{2^{1/2}}{\pi^{1/2}\xi}\big)}\Big]$ satisfies $\phi(x_\xi - \nu) + \phi(x_\xi + \nu) = \xi$. In fact, when $\nu \geq 2$, we may take $\xi_0 = 2\phi(\nu)$. Moreover, $\eta(x) = \frac{\phi(x-\nu)}{\phi(x-\nu)+\phi(x+\nu)} = \frac{1}{1+e^{-2x\nu}}$, so $\mathcal{X}_{\tau+\Delta}(\eta) = [x_{\nu,\tau,\Delta}, \infty)$, where $x_{\nu,\tau,\Delta} := \frac{1}{2\nu}\log\big(\frac{\tau+\Delta}{1-(\tau+\Delta)}\big)$. By reducing $\xi_0 > 0$, depending only on ν and τ, if necessary, we may assume that $-x_\xi \leq x_{\nu,\tau,\Delta} \leq x_\xi$ for $\xi \in (0, \xi_0]$ and $\Delta \in \big(0, (1-\tau)/2\big]$. Writing Φ for the standard normal distribution function, we deduce that for $\xi \in (0, \xi_0]$,

$$\sup\big\{\mu(A) : A \in \mathcal{A}_{\mathrm{int}} \cap \mathrm{Pow}\big(\mathcal{X}_\xi(\omega) \cap \mathcal{X}_{\tau+\Delta}(\eta)\big)\big\} = \mu\big([x_{\nu,\tau,\Delta}, x_\xi]\big)$$
$$= \frac{1}{2}\Phi(x_\xi - \nu) - \frac{1}{2}\Phi(x_{\nu,\tau,\Delta} - \nu) + \frac{1}{2}\Phi(x_\xi + \nu) - \frac{1}{2}\Phi(x_{\nu,\tau,\Delta} + \nu).$$

Using the Mills ratio and the mean value inequality, it follows that for $\xi \in (0, \xi_0]$,

$$M_\tau - \sup\big\{\mu(A) : A \in \mathcal{A}_{\mathrm{int}} \cap \mathrm{Pow}\big(\mathcal{X}_\xi(\omega) \cap \mathcal{X}_{\tau+\Delta}(\eta)\big)\big\}$$
$$= 1 - \frac{1}{2}\Phi(x_\xi - \nu) - \frac{1}{2}\Phi(x_\xi + \nu) - \frac{1}{2}\Phi(x_{\nu,\tau,0} - \nu) - \frac{1}{2}\Phi(x_{\nu,\tau,0} + \nu)$$
$$+ \frac{1}{2}\Phi(x_{\nu,\tau,\Delta} - \nu) + \frac{1}{2}\Phi(x_{\nu,\tau,\Delta} + \nu)$$
$$\leq \frac{\phi(x_\xi - \nu)}{2(x_\xi - \nu)} + \frac{\phi(x_\xi + \nu)}{2(x_\xi + \nu)} + \frac{\Delta}{\sqrt{2\pi}\nu\tau(1-\tau)}$$
$$\leq \frac{\xi}{2\sqrt{2\log\big(\frac{1}{(2\pi)^{1/2}\xi_0}\big)}} + \frac{\Delta}{\sqrt{2\pi}\nu\tau(1-\tau)}.$$

On the other hand, when $\xi > \xi_0$, we have

$$M_\tau - \sup\big\{\mu(A) : A \in \mathcal{A}_{\mathrm{int}} \cap \mathrm{Pow}\big(\mathcal{X}_\xi(\omega) \cap \mathcal{X}_{\tau+\Delta}(\eta)\big)\big\} \leq 1 \leq \frac{\xi}{\xi_0}.$$

We conclude that $P \in \mathcal{P}_{\mathrm{App}}(\mathcal{A}_{\mathrm{int}}, \tau, \kappa, \gamma, \tau, C_{\mathrm{App}})$ with $\kappa = \gamma = 1$ when we take

$$C_{\mathrm{App}}^{-1} = \min\biggl\{2\sqrt{2\log\biggl(\frac{1}{(2\pi)^{1/2}\xi_0}\biggr)}, \sqrt{2\pi}\nu\tau(1-\tau), \xi_0\biggr\}.$$

Example 2: Given $\epsilon_0 > 0$, choose $A_0 \in \mathcal{A}_{\mathrm{hpr}} \cap \mathrm{Pow}\bigl(\mathcal{X}_\tau(\eta) \cap [0,1]^d\bigr)$ such that $\mu(A_0) \geq M_\tau - \epsilon_0$. Let ∂A_0 and $r = (r_1, \ldots, r_d) \in [0,1]^d$ denote the boundary and vector of side-lengths of A_0 respectively. Observe that for $\Delta \leq \epsilon \cdot \delta^{1/\gamma}$,

$$\begin{aligned}\mathcal{X}_{\tau+\Delta}(\eta) &\supseteq \bigl\{x \in \mathcal{X}_\tau(\eta) \cap [0,1]^d : \mathrm{dist}_\infty(x, \mathcal{S}_\tau) \geq (\Delta/\epsilon)^\gamma\bigr\} \\ &\supseteq \bigl\{x \in A_0 : \mathrm{dist}_\infty(x, \partial A_0) \geq (\Delta/\epsilon)^\gamma\bigr\}.\end{aligned}$$

Moreover, $\mathcal{X}_\xi(\omega) = [0,1]^d$ for $\xi \leq 1$. For $s > 0$, let $A_0(s) := \bigl\{x \in A_0 : \mathrm{dist}_\infty(x, \partial A_0) \geq s\bigr\}$. Note that for $s \leq \min_j r_j/2$,

$$\mu(A_0) - \mu(A_0(s)) \leq \prod_{j=1}^d r_j - \prod_{j=1}^d (r_j - 2s) \leq 1 - (1-2s)^d \leq 2ds.$$

On the other hand, if $s > \min_{j \in [d]} r_j/2$, then

$$\mu(A_0) - \mu(A_0(s)) \leq \prod_{j=1}^d r_j \leq \min_{j \in [d]} r_j < 2s.$$

Then, for $\xi \in [0,1]$ and any $\Delta \in (0, \epsilon \cdot \delta^{1/\gamma}]$,

$$\begin{aligned}M_\tau - \sup\bigl\{\mu(A) : A \in \mathcal{A}_{\mathrm{hpr}} \cap \mathrm{Pow}\bigl(\mathcal{X}_\xi(\omega) \cap \mathcal{X}_{\tau+\Delta}(\eta)\bigr)\bigr\} &\leq M_\tau - \mu\bigl(A_{0,(\Delta/\epsilon)^\gamma}\bigr) \\ &\leq M_\tau - \mu(A_0) + 2d\Bigl(\frac{\Delta}{\epsilon}\Bigr)^\gamma \\ &\leq \epsilon_0 + 2d\Bigl(\frac{\Delta}{\epsilon}\Bigr)^\gamma.\end{aligned}$$

On the other hand, if $\xi > 1$ or $\Delta > \epsilon \cdot \delta^{1/\gamma}$, then for any $\kappa \in (0, \infty)$ and $C_{\mathrm{App}} \geq 1/(\epsilon^\gamma \delta)$, we have

$$M_\tau - \sup\bigl\{\mu(A) : A \in \mathcal{A}_{\mathrm{hpr}} \cap \mathrm{Pow}\bigl(\mathcal{X}_\xi(\omega) \cap \mathcal{X}_{\tau+\Delta}(\eta)\bigr)\bigr\} \leq 1 \leq C_{\mathrm{App}} \cdot (\xi^\kappa + \Delta^\gamma).$$

Since $\epsilon_0 > 0$ was arbitrary, the conclusion follows.

Example 3: Writing $\omega_\kappa := \omega_{\mu_\kappa, d}$ for the lower-density of μ_κ, we have for $x \in \mathbb{R}^d$ that

$$\begin{aligned}\omega_\kappa(x) &\geq \sup_{t \in [\|x\|_\infty, \infty)} g_\kappa(t) \biggl\{\mathcal{L}_d\bigl(\bar{B}_{1\wedge t}(x) \cap \bar{B}_t(0)\bigr) \wedge \inf_{r \in (0, 1\wedge t)} \frac{\mathcal{L}_d\bigl(\bar{B}_r(x) \cap \bar{B}_t(0)\bigr)}{r^d}\biggr\} \\ &\geq \sup_{t \in [\|x\|_\infty, \infty)} (1 \wedge t^d) g_\kappa(t) \\ &\geq \begin{cases} g_\kappa(\|x\|_\infty) \wedge g_\kappa(1) & \text{if } \kappa \in (0,2) \\ \Bigl(\frac{1}{2^d(\kappa-1)} \cdot g_\kappa(\|x\|_\infty)\Bigr) \wedge g_\kappa\Bigl(\frac{1}{2(\kappa-1)^{1/d}}\Bigr) & \text{if } \kappa \in [2, \infty).\end{cases}\end{aligned}$$

Now, writing $a_{d,\kappa} := \frac{1}{2(\kappa-1)^{1/d}} \cdot \mathbb{1}_{\{\kappa \geq 2\}} + \mathbb{1}_{\{\kappa < 2\}}$ and $\xi_{d,\kappa} := g_\kappa(a_{d,\kappa})$, we have for $\xi \leq \xi_{d,\kappa}$ that $\mathcal{X}_\xi(\omega_\kappa) \supseteq \bigl\{x \in \mathbb{R}^d : \|x\|_\infty \leq R_{\xi,d,\kappa}\bigr\}$, where

$$R_{\xi,d,\kappa} := \begin{cases} \Bigl(\frac{(\kappa/(2^d\xi))^{1-\kappa} - 1}{1-\kappa}\Bigr)^{1/d} & \text{if } \kappa \in (0,1) \\ \log^{1/d}\Bigl(\frac{1}{2^d \xi}\Bigr) & \text{if } \kappa = 1 \\ \Bigl(\frac{1 - \{2^d\xi/(\kappa a_{d,\kappa}^d)\}^{\kappa-1}}{\kappa - 1}\Bigr)^{1/d} & \text{if } \kappa \in (1, \infty).\end{cases}$$

We now calculate that

$$M_\tau \equiv M_\tau(P_{\kappa,\gamma},\mathcal{A}_{\mathrm{hpr}}) = \mu_\kappa\bigl([0,\infty)\times\mathbb{R}^{d-1}\bigr) = 1/2. \tag{S8}$$

Observe that $\mathcal{X}_{\tau+\Delta}(\eta_\gamma) = \{x=(x_1,\ldots,x_d)^\top \in \mathbb{R}^d : x_1 \geq (\Delta/\lambda)^\gamma\}$ for $\Delta\in(0,1-\tau]$. Hence, for $\Delta\in(0,1-\tau]$ and $\xi\leq\xi_{d,\kappa}$, we have

$$\begin{aligned}
&\sup\bigl\{\mu_\kappa(A) : A\in\mathcal{A}_{\mathrm{hpr}}\cap \mathrm{Pow}\bigl(\mathcal{X}_\xi(\omega_\kappa)\cap\mathcal{X}_{\tau+\Delta}(\eta_\gamma)\bigr)\bigr\}\\
&\qquad\geq \sup\bigl\{\mu_\kappa(A) : A\in\mathcal{A}_{\mathrm{hpr}}\cap\mathrm{Pow}\bigl(\bar{B}_{R_{\xi,d,\kappa}}(0)\cap\bigl([(\Delta/\lambda)^\gamma,\infty)\times\mathbb{R}^{d-1}\bigr)\bigr)\bigr\}\\
&\qquad= \mu_\kappa\bigl(\bar{B}_{R_{\xi,d,\kappa}}(0)\cap\bigl([(\Delta/\lambda)^\gamma,\infty)\times\mathbb{R}^{d-1}\bigr)\bigr)\\
&\qquad\geq \frac{1}{2}\mu_\kappa\bigl(\bar{B}_{R_{\xi,d,\kappa}}(0)\bigr) - \mu_\kappa\bigl([0,(\Delta/\lambda)^\gamma]\times\mathbb{R}^{d-1}\bigr). \qquad\qquad \text{(S9)}
\end{aligned}$$

For $x=(x_1,\ldots,x_d)^\top\in\mathbb{R}^d$, let $x_{-1}:=(x_2,\ldots,x_d)^\top\in\mathbb{R}^{d-1}$. Then

$$\begin{aligned}
\mu_\kappa\bigl([0,(\Delta/\lambda)^\gamma]\times\mathbb{R}^{d-1}\bigr) &= \int_{[0,(\Delta/\lambda)^\gamma]\times\mathbb{R}^{d-1}} g_\kappa(\|x\|_\infty)\,dx\\
&\leq \int_{[0,(\Delta/\lambda)^\gamma]\times\mathbb{R}^{d-1}} g_\kappa(\|x_{-1}\|_\infty)\,dx = b_{d,\kappa}\cdot\Bigl(\frac{\Delta}{\lambda}\Bigr)^\gamma, \qquad \text{(S10)}
\end{aligned}$$

where

$$\begin{aligned}
b_{d,\kappa} &:= \int_{\mathbb{R}^{d-1}} g_\kappa(\|x_{-1}\|_\infty)\,dx_{-1} = (d-1)\cdot 2^{d-1}\int_0^\infty y^{d-2}g_\kappa(y)\,dy\\
&= \begin{cases} \frac{(1-\kappa)^{1/d}\Gamma(2-1/d)\Gamma(\frac{1+(d-1)\kappa}{d(1-\kappa)})}{2\Gamma(\kappa/(1-\kappa))} & \text{if } \kappa\in(0,1)\\ \Gamma(2-1/d)/2 & \text{if } \kappa=1\\ \frac{(\kappa-1)^{1/d}\Gamma(2-1/d)\Gamma(2+\frac{1}{\kappa-1})}{2\Gamma(2-\frac{1}{d}+\frac{1}{1-\kappa})} & \text{if } \kappa\in(1,\infty).\end{cases}
\end{aligned}$$

Moreover, for $\xi\leq\xi_{d,\kappa}\leq\kappa/2^d$,

$$1-\mu_\kappa\bigl(\bar{B}_{R_{\xi,d,\kappa}}(0)\bigr) = d\cdot 2^d\int_{R_{\xi,d,\kappa}}^\infty y^{d-1}g_\kappa(y)\,dy = \biggl(\frac{2^d\xi}{a_{d,\kappa}^d\kappa}\biggr)^\kappa. \tag{S11}$$

But for $\Delta>1-\tau$ or $\xi>\xi_{d,\kappa}$, we have

$$\sup\bigl\{\mu_\kappa(A) : A\in\mathcal{A}_{\mathrm{hpr}}\cap\mathrm{Pow}\bigl(\mathcal{X}_\xi(\omega_\kappa)\cap\mathcal{X}_{\tau+\Delta}(\eta_\gamma)\bigr)\bigr\} \geq 0 \geq \frac{1}{2} - \Bigl(\frac{\Delta}{1-\tau}\Bigr)^\gamma - \Bigl(\frac{\xi}{\xi_{d,\kappa}}\Bigr)^\kappa.$$

We deduce from (S8), (S9), (S10) and (S11) that $P_{\kappa,\gamma}\in\mathcal{P}_{\mathrm{App}}(\mathcal{A}_{\mathrm{hpr}},\tau,\kappa,\gamma,\tau,C_{\mathrm{App}})$, with

$$C_{\mathrm{App}} = \biggl(\frac{b_{d,\kappa}^{1/\gamma}}{\lambda}\vee\frac{1}{1-\tau}\biggr)^\gamma \vee \biggl(\frac{2^d}{a_{d,\kappa}^d\kappa}\vee\frac{1}{\xi_{d,\kappa}}\biggr)^\kappa.$$

Given a closed set $S\subseteq\mathbb{R}^d$, we define the projection $\Pi_S:\mathbb{R}^d\to S$ by

$$\Pi_S(x) := \operatorname*{sargmin}_{z\in S}\|x-z\|_2,$$

where $\mathrm{sargmin}$ denotes the smallest element of the argmin in the lexicographic ordering.

PROPOSITION S6. *Let P be a distribution on $\mathbb{R}^d\times[0,1]$ with marginal μ on $\mathbb{R}^d$ and continuous regression function η. Suppose that $\mathcal{A}\subseteq\mathcal{B}(\mathbb{R}^d)$ has the properties that*

- *there exists a bounded set* $A_0 \in \mathcal{A} \cap \mathrm{Pow}\big(\mathcal{X}_\tau(\eta)\big)$ *such that* $\mu(A_0) = M_\tau$*;*
- *writing* ∂A_0 *for the topological boundary of* A_0*, there exist* $s_0 > 0$ *and* $C'_{\mathrm{App}} > 0$ *such that for every* $s \in (0, s_0]$*, we can find* $A_0(s) \in \mathcal{A}$ *with* $A_0(s) \subseteq \{x \in A_0 : \mathrm{dist}_2(x, \partial A_0) > s\}$*, satisfying*

$$\mu(A_0) - \mu\big(A_0(s)\big) \leq C'_{\mathrm{App}} \cdot s.$$

Suppose that there exist $c_*, \kappa > 0$ *such that*

$$\omega(x) \geq c_* \cdot s^{1/\kappa}$$

for all $x \in A_0(s)$ *and* $s \in (0, s_0)$*. Assume further that* $\mathcal{S}_\tau := \{x \in \mathbb{R}^d : \eta(x) = \tau\}$ *is non-empty, and that there exists* $\delta_0 > 0$ *such that* η *is differentiable on* $\mathcal{S}_{\tau,\delta_0} := \mathcal{S}_\tau + \delta_0 B_{2,1}(0)$*. Let* $A_{0,\delta_0} := A_0 + \delta_0 B_{2,1}(0)$ *and assume that* $\epsilon_0 := \inf_{x \in A_{0,\delta_0} \cap \mathcal{S}_{\tau,\delta_0}} \|\nabla\eta(x)\|_2 > 0$*. Then* $\Delta_* := \inf_{x \in A_0 \setminus \mathcal{S}_{\tau,\delta_0}} \eta(x) - \tau > 0$*, and* $P \in \mathcal{P}_{\mathrm{App}}(\mathcal{A}, \tau, \kappa, 1, C_{\mathrm{App}})$ *for*

$$C_{\mathrm{App}} \geq \max\left\{\frac{C'_{\mathrm{App}}}{\max(c_*^\kappa, \epsilon_0)}, \frac{1}{s_0 c_*^\kappa}, \frac{1}{\min(\epsilon_0\delta_0, \Delta_*)}\right\}.$$

PROOF. Since η is continuous on the intersection of the closure of A_0 with the complement of $\mathcal{S}_{\tau,\delta_0}$, and since this intersection is compact but does not contain any point in $\mathcal{S}_\tau$, we have that $\Delta_* > 0$. By Cannings, Berrett and Samworth (2020, Proposition 2), we have

$$\mathcal{S}_{\tau,\delta_0} = \left\{x_0 + \frac{t\nabla\eta(x_0)}{\|\nabla\eta(x_0)\|_2} : x_0 \in \mathcal{S}_\tau, |t| < \delta_0\right\}. \tag{S12}$$

Moreover, from the proof of that result, we see that for any $x \in \mathcal{S}_{\tau,\delta_1}$, we can take $x_0 = \Pi_{\mathcal{S}_\tau}(x)$ in the representation (S12). Now suppose that $x \in A_0 \cap \mathcal{S}_{\tau,\delta_0}$, so that $x = x_0 + t\nabla\eta(x_0)/\|\nabla\eta(x_0)\|_2$ with $x_0 = \Pi_{\mathcal{S}_\tau}(x) \in A_{0,\delta_0} \cap \mathcal{S}_\tau$ and $|t| = \mathrm{dist}_2(x, \mathcal{S}_\tau) < \delta_0$. Since the line segment joining x_0 and x is contained in $A_{0,\delta_0} \cap \mathcal{S}_{\tau,\delta_0}$, we have

$$|\eta(x) - \tau| = \left|\eta\left(x_0 + \frac{t\nabla\eta(x_0)}{\|\nabla\eta(x_0)\|_2}\right) - \eta(x_0)\right| \geq |t|\epsilon_0. \tag{S13}$$

Now observe that $\mathrm{int}(A_0) \cap \mathcal{S}_\tau = \emptyset$ because if $x_0 \in \mathrm{int}(A_0) \cap \mathcal{S}_\tau$, then for sufficiently small $t > 0$, the point $x_0 - t\nabla\eta(x_0)/\|\nabla\eta(x_0)\|_2$ would belong to A_0 and $\mathcal{X}_\tau(\eta)^{\mathrm{c}}$, a contradiction. Hence, for $x \in A_0$ we have $\mathrm{dist}_2(x, \mathcal{S}_\tau) \geq \mathrm{dist}_2(x, \partial A_0)$, so for $\Delta < \min(\epsilon_0\delta_0, \Delta_*)$,

$$\{x \in A_0 : \eta(x) \in [\tau, \tau + \Delta)\} \subseteq A_0 \cap \mathcal{S}_{\tau,\Delta/\epsilon_0} \subseteq \big\{x \in A_0 : \mathrm{dist}_2(x, \partial A_0) \leq \Delta/\epsilon_0\big\}.$$

Then, for any $\xi \in \big(0, c_* s_0^{1/\kappa}\big)$ and $\Delta \in \big(0, \min(\epsilon_0\delta_0, \Delta_*)\big)$,

$$\begin{aligned} M_\tau - \sup\big\{\mu(A) : A \in \mathcal{A} \cap \mathrm{Pow}\big(\mathcal{X}_\xi(\omega) \cap \mathcal{X}_{\tau+\Delta}(\eta)\big)\big\} &\leq \mu(A_0) - \mu\left(A_0\Big(\frac{\xi^\kappa}{c_*^\kappa} \wedge \frac{\Delta}{\epsilon_0}\Big)\right) \\ &\leq C'_{\mathrm{App}} \cdot \left(\frac{\xi^\kappa}{c_*^\kappa} \wedge \frac{\Delta}{\epsilon_0}\right) \leq C_{\mathrm{App}} \cdot (\xi^\kappa + \Delta). \end{aligned}$$

On the other hand, if $\xi \geq c_* s_0^{1/\kappa}$ or $\Delta \geq \min(\epsilon_0\delta_0, \Delta_*)$, then provided we take $C_{\mathrm{App}} \geq \max\big\{1/(s_0 c_*^\kappa), 1/\min(\epsilon_0\delta_0, \Delta_*)\big\}$, we have

$$\begin{aligned} M_\tau - \sup\big\{\mu(A) : A \in \mathcal{A} \cap \mathrm{Pow}\big(\mathcal{X}_\xi(\omega) \cap \mathcal{X}_{\tau+\Delta}(\eta)\big)\big\} \leq 1 &\leq \frac{1}{s_0} \cdot \left(\frac{\xi}{c_*}\right)^\kappa + \frac{\Delta}{\min(\epsilon_0\delta_0, \Delta_*)} \\ &\leq C_{\mathrm{App}} \cdot (\xi^\kappa + \Delta). \end{aligned}$$

The result follows. □

S4. Proofs from Section 3.

S4.1. *Proofs from Section 3.1.* In order to prove Theorem 6, we first establish Proposition S7 below, which will be useful in the sequel. Recall the definition of $\Delta_{n,\beta,\lambda}$ from (10).

PROPOSITION S7. *Given $\delta \in (0,1)$ and $x_0,\, x_1 \in \mathbb{R}^d$, we let*

$$\mathcal{E}^{\mathrm{Höl}}_{n,\delta}(x_0,x_1) := \bigcap \biggl\{ \biggl(\frac{|\eta(x_0)-\eta(x_1)| - \Delta_{n,\beta,\lambda}(x_0) \vee \Delta_{n,\beta,\lambda}(x_1)}{\|x_0-x_1\|_\infty^\beta} \biggr) \leq \hat{\lambda}_{n,\beta,\delta} \leq \lambda \biggr\},$$

where the intersection is over all $(\beta,\lambda) \in (0,1] \times [1,\infty)$ for which $P \in \mathcal{P}_{\mathrm{Höl}}(\beta,\lambda,\tau)$ and $\min_{x\in\{x_0,x_1\}}\{\eta(x) - \Delta_{n,\beta,\lambda}(x)\} \geq \tau$. Here we adopt the convention that an intersection over an empty set is the entire probability space. Then $\mathbb{P}_P\bigl(\mathcal{E}^{\mathrm{Höl}}_{n,\delta}(x_0,x_1)\bigr) \geq 1-\delta$.

The proof of Proposition S7 will make use of the event

$$\mathcal{E}_{\eta,\delta} := \bigcap_{(i,k)\in[n]^2} \Biggl\{ \Biggl| \hat{\eta}_n\bigl(\bar{B}_{r_{i,k}}(X_i)\bigr) - \frac{\sum_{t\in[n]} \eta(X_t)\cdot \mathbb{1}_{\{X_t \in \bar{B}_{r_{i,k}}(X_i)\}}}{\sum_{t\in[n]} \mathbb{1}_{\{X_t\in\bar{B}_{r_{i,k}}(X_i)\}}} \Biggr| \leq \sqrt{\frac{\log(4n^2/\delta)}{2k}} \Biggr\}.$$

PROOF OF PROPOSITION S7. Fix $n \in \mathbb{N}$, $\delta \in (0,1)$, and $x_0, x_1 \in \mathbb{R}^d$. We will assume without loss of generality that $\eta(x_0) \geq \eta(x_1)$ and $\min\{\omega(x_0),\omega(x_1)\} > 0$ since otherwise the index set in the intersection in the definition of $\mathcal{E}^{\mathrm{Höl}}_{n,\delta}(x_0,x_1)$ is empty. By Hoeffding's lemma and a union bound, we have $\mathbb{P}\bigl(\mathcal{E}^{\mathrm{c}}_{\eta,\delta}|\mathcal{D}_X\bigr) \leq \delta/2$. For $\ell \in [n]$ and $x \in \{x_0,x_1\}$ we let $s_\ell(x) := \bigl(n\omega(x)/(2\ell)\bigr)^{-1/d}$ and define the event

$$\mathcal{E}_{\mu,\delta}(x_0,x_1) := \bigcap_{x\in\{x_0,x_1\}} \bigcap_{\ell=\lceil 4\log(4n/\delta)\rceil}^{\lfloor n\omega(x)/2\rfloor} \bigl\{ |\{X_i\}_{i\in[n]} \cap \bar{B}_{s_\ell(x)}(x)| \geq \ell \bigr\}.$$

Observe that for $x \in \{x_0,x_1\}$ and $\ell \in \{\lceil 4\log(4n/\delta)\rceil,\ldots,\lfloor n\omega(x)/2\rfloor\}$ we have $s_\ell(x) \in (0,1]$ and

$$\mu\bigl(\bar{B}_{s_\ell(x)}(x)\bigr) \geq \omega(x)\cdot s_\ell(x)^d = \frac{2\ell}{n}. \tag{S14}$$

Hence, by the multiplicative Chernoff bound (Lemma S39) we have $\mathbb{P}\bigl(\mathcal{E}_{\mu,\delta}(x_0,x_1)^{\mathrm{c}}\bigr) \leq \delta/2$. Consequently, to complete the proof it suffices to show that $\mathcal{E}_{\eta,\delta} \cap \mathcal{E}_{\mu,\delta}(x_0,x_1) \subseteq \mathcal{E}^{\mathrm{Höl}}_{n,\delta}(x_0,x_1)$. To this end, fix $(\beta,\lambda) \in (0,1] \times [1,\infty)$ for which $P \in \mathcal{P}_{\mathrm{Höl}}(\beta,\lambda,\tau)$ and $\eta(x) - \Delta_n(x) \geq \tau$ for $x \in \{x_0,x_1\}$. On the event $\mathcal{E}_{\eta,\delta}$, we have for any $(i,k) \in [n]^2$ with $\hat{\psi}_{n,\beta,\delta}(i,k) > \lambda$ that

$$\frac{\sum_{t\in[n]} \eta(X_t)\cdot \mathbb{1}_{\{X_t \in \bar{B}_{r_{i,k}}(X_i)\}}}{\sum_{t\in[n]} \mathbb{1}_{\{X_t\in\bar{B}_{r_{i,k}}(X_i)\}}} \geq \hat{\eta}_n\bigl(\bar{B}_{r_{i,k}}(X_i)\bigr) - \sqrt{\frac{\log(4n^2/\delta)}{2k}} > \tau + \lambda\cdot(2r_{i,k})^\beta.$$

Hence, $\bar{B}_{r_{i,k}}(X_i) \subseteq \mathcal{X}_\tau(\eta)$, since $\eta \in \mathcal{F}_{\mathrm{Höl}}(\beta,\lambda,\mathcal{X}_\tau(\eta))$. Consequently, on the event $\mathcal{E}_{\eta,\delta}$, given any $(i,j,k,\ell) \in [n]^4$ with $\min\bigl\{\hat{\psi}_{n,\beta,\delta}(i,k),\hat{\psi}_{n,\beta,\delta}(j,\ell)\bigr\} > \lambda$, we have

$$\begin{aligned}
&\hat{\eta}_n\bigl(\bar{B}_{r_{i,k}}(X_i)\bigr) - \hat{\eta}_n\bigl(\bar{B}_{r_{j,\ell}}(X_j)\bigr) - \sqrt{\frac{2\log(4n^2/\delta)}{k\wedge\ell}} \\
&\qquad\leq \frac{\sum_{t\in[n]} \eta(X_t)\cdot \mathbb{1}_{\{X_t \in \bar{B}_{r_{i,k}}(X_i)\}}}{\sum_{t\in[n]} \mathbb{1}_{\{X_t\in\bar{B}_{r_{i,k}}(X_i)\}}} - \frac{\sum_{t\in[n]} \eta(X_t)\cdot \mathbb{1}_{\{X_t \in \bar{B}_{r_{j,\ell}}(X_j)\}}}{\sum_{t\in[n]} \mathbb{1}_{\{X_t\in\bar{B}_{r_{j,\ell}}(X_j)\}}} \\
&\qquad\leq \eta(X_i) + \lambda\cdot r_{i,k}^\beta - \eta(X_j) + \lambda\cdot r_{j,\ell}^\beta \leq \lambda\cdot\bigl(\|X_i - X_j\|_\infty^\beta + r_{i,k}^\beta + r_{j,\ell}^\beta\bigr),
\end{aligned}$$

and hence $\hat{\phi}_{n,\beta,\delta}(i,j,k,\ell) \leq \lambda$. Thus, on the event $\mathcal{E}_{\eta,\delta}$, we have $\hat{\lambda}_{n,\beta,\delta} \leq \lambda$.

For the lower bound, assume without loss of generality that $\Delta_n(x_0) \vee \Delta_n(x_1) \leq 1$ and let $i_0 := \mathrm{sargmin}_{i\in[n]} \|X_i - x_0\|_\infty$ and $i_1 := \mathrm{sargmin}_{i\in[n]} \|X_i - x_1\|_\infty$. For $x \in \{x_0, x_1\}$, let $\rho_n(x) := \bigl\{\Delta_n(x)/(24\lambda)\bigr\}^{1/\beta}$ and $k_n(x) := \lfloor \frac{n}{2} \cdot \omega(x) \cdot \rho_n(x)^d \rfloor$. Since for each $x \in \{x_0, x_1\}$, the fact that $\Delta_n(x) \leq 1$ ensures that $\frac{n}{2} \cdot \omega(x) \cdot \rho_n(x)^d \geq 4\log(4n/\delta) + 2$, we have $k_n(x) \in \{\lceil 4\log(4n/\delta)\rceil, \ldots, n\}$. Hence, on the event $\mathcal{E}_{\mu,\delta}(x_0,x_1)$ we have $|\{X_i\}_{i\in[n]} \cap \bar{B}_{\rho_n(x)}(x)| \geq k_n(x)\}$ for $x \in \{x_0, x_1\}$.

Now fix $j \in \{0,1\}$, so that $\|X_{i_j} - x_j\|_\infty \leq \rho_n(x_j)$ and $r_{i_j,k_n(x_j)} \leq 2 \cdot \rho_n(x_j)$. Then $\eta(x_j) \geq \tau + \Delta_n(x_j) = \tau + 24\lambda\rho_n(x_j)^\beta$ and $\eta \in \mathcal{F}_{\mathrm{H\ddot{o}l}}(\beta,\lambda,\mathcal{X}_\tau(\eta))$, so $\bar{B}_{r_{i_j,k_n(x_j)}}(X_{i_j}) \subseteq \bar{B}_{3\cdot\rho_n(x_j)}(x_j) \subseteq \mathcal{X}_\tau(\eta)$. Since $\Delta_n(x_j) \leq 1$, we also have $\sqrt{2\log(4n^2/\delta)/k_n(x_j)} \leq \Delta_n(x_j)/4$. Hence, on the event $\mathcal{E}_{\eta,\delta} \cap \mathcal{E}_{\mu,\delta}(x_0,x_1)$, we have

$$
\begin{aligned}
\hat{\eta}_n\bigl(\bar{B}_{r_{i_j,k_n(x_j)}}(X_{i_j})\bigr) - \tau - \sqrt{\frac{\log(4n^2/\delta)}{2k_n(x_j)}}
&\geq \frac{\sum_{t\in[n]} \eta(X_t)\cdot \mathbb{1}_{\{X_t \in \bar{B}_{r_{i_j,k_n(x_j)}}(X_{i_j})\}}}{\sum_{t\in[n]} \mathbb{1}_{\{X_t \in \bar{B}_{r_{i_j,k_n(x_j)}}(X_{i_j})\}}} - \tau - \sqrt{\frac{2\log(4n^2/\delta)}{k_n(x_j)}} \\
&\geq \eta(x_j) - \lambda\cdot\{3\cdot\rho_n(x_j)\}^\beta - \tau - \sqrt{\frac{2\log(4n^2/\delta)}{k_n(x_j)}} \\
&\geq \Delta_n(x_j) - \frac{3^\beta}{24}\cdot\Delta_n(x_j) - \frac{1}{4}\cdot\Delta_n(x_j) \geq \frac{5}{8}\cdot\Delta_n(x_j) \\
&= 15\lambda\cdot\rho_n(x_j)^\beta > \lambda\cdot\bigl(2r_{i_j,k_n(x_j)}\bigr)^\beta.
\end{aligned}
$$

Thus, on the event $\mathcal{E}_{\eta,\delta} \cap \mathcal{E}_{\mu,\delta}(x_0,x_1)$, we have $\hat{\psi}_{n,\beta,\delta}\bigl(i_j, k_n(x_j)\bigr) > \lambda$. Hence, whenever $\Delta_n(x_0) \vee \Delta_n(x_1) \leq 1$, we have on the event $\mathcal{E}_{\eta,\delta} \cap \mathcal{E}_{\mu,\delta}(x_0,x_1)$ that

$$
\begin{aligned}
&\hat{\phi}_{n,\beta,\delta}\bigl(i_0, i_1, k_n(x_0), k_n(x_1)\bigr) \\
&\quad= \frac{\hat{\eta}_n\bigl(\bar{B}_{r_{i_0,k_n(x_0)}}(X_{i_0})\bigr) - \hat{\eta}_n\bigl(\bar{B}_{r_{i_1,k_n(x_1)}}(X_{i_1})\bigr) - \sqrt{2\log(4n^2/\delta)/\{k_n(x_0)\wedge k_n(x_1)\}}}{\|X_{i_0} - X_{i_1}\|_\infty^\beta + r_{i_0,k_n(x_0)}^\beta + r_{i_1,k_n(x_1)}^\beta} \\
&\quad\geq \frac{\eta(x_0) - \eta(x_1) - 2\lambda\bigl\{3\bigl(\rho_n(x_0)\vee\rho_n(x_1)\bigr)\bigr\}^\beta - 2\sqrt{2\log(4n^2/\delta)/\{k_n(x_0)\wedge k_1(x_0)\}}}{\|x_0 - x_1\|_\infty^\beta + 3\cdot\bigl\{2\cdot\bigl(\rho_n(x_0)\vee\rho_n(x_1)\bigr)\bigr\}^\beta} \\
&\quad\geq \frac{\eta(x_0) - \eta(x_1) - \frac{3}{4}\bigl(\Delta_n(x_0)\vee\Delta_n(x_1)\bigr)}{\|x_0 - x_1\|_\infty^\beta + \frac{1}{4\lambda}\bigl(\Delta_n(x_0)\vee\Delta_n(x_1)\bigr)} \\
&\quad\geq \frac{\eta(x_0) - \eta(x_1) - \frac{3}{4}\bigl(\Delta_n(x_0)\vee\Delta_n(x_1)\bigr)}{\|x_0 - x_1\|_\infty^\beta} - \frac{\Delta_n(x_0)\vee\Delta_n(x_1)}{4\lambda}\cdot\frac{\eta(x_0)-\eta(x_1)}{\|x_0 - x_1\|_\infty^{2\beta}} \\
&\quad\geq \frac{\eta(x_0) - \eta(x_1) - \Delta_n(x_0)\vee\Delta_n(x_1)}{\|x_0 - x_1\|_\infty^\beta}.
\end{aligned}
$$

Thus, whenever $\Delta_n(x_0) \vee \Delta_n(x_1) \leq 1$, we have on the event $\mathcal{E}_{\eta,\delta} \cap \mathcal{E}_{\mu,\delta}(x_0,x_1)$ that

$$
\begin{aligned}
\hat{\lambda}_{n,\beta,\delta} &\geq \min\bigl\{\hat{\psi}_{n,\beta,\delta}\bigl(i_0,k_n(x_0)\bigr), \hat{\psi}_{n,\beta,\delta}\bigl(i_1,k_n(x_1)\bigr), \hat{\phi}_{n,\beta,\delta}\bigl(i_0,i_1,k_n(x_0),k_n(x_1)\bigr)\bigr\} \\
&\geq \min\biggl\{\frac{\eta(x_0) - \eta(x_1) - \Delta_n(x_0)\vee\Delta_n(x_1)}{\|x_0 - x_1\|_\infty^\beta}, \lambda\biggr\}
\end{aligned}
$$

$$\geq \min\left\{\frac{\eta(x_0)-\eta(x_1)-\Delta_n(x_0)\vee\Delta_n(x_1)}{\|x_0-x_1\|_\infty^\beta}, \frac{\eta(x_0)-\eta(x_1)}{\|x_0-x_1\|_\infty^\beta}\right\}$$
$$= \frac{\eta(x_0)-\eta(x_1)-\Delta_n(x_0)\vee\Delta_n(x_1)}{\|x_0-x_1\|_\infty^\beta},$$

as required. □

PROOF OF THEOREM 6. Theorem 6 is an immediate consequence of Proposition S7 combined with the continuity of probability from below. □

The following four lemmas are used in the proof of Corollary 7.

LEMMA S8. *Let μ be a Borel probability distribution on $\mathbb{R}^d$ with lower density ω. Then* $\mu\big(\{x\in\mathbb{R}^d:\omega(x)=0\}\big)=0$. *Hence* $\mathrm{supp}(\mu)=\mathrm{cl}\big(\{x\in\mathbb{R}^d:\omega(x)>0\}\big)$.

PROOF. Let $Z:=\{x\in\mathbb{R}^d:\omega(x)=0\}$ and fix $\epsilon\in(0,1)$. By the monotone convergence theorem we may choose $R>0$ sufficiently large that $\mu\big(Z\setminus\bar{B}_R(0)\big)<\epsilon/2$ and let $\xi:=\epsilon/\big(2\cdot\{5(R+1)\}^d\big)$. By Reeve, Cannings and Samworth (2021, Lemma S4),

$$\mu\big(Z\cap\bar{B}_R(0)\big)\leq\mu\big(\{x\in\bar{B}_R(0):\omega(x)<\xi\}\big)\leq\{5(R+1)\}^d\cdot\xi=\frac{\epsilon}{2}.$$

Hence $\mu(Z)=0$, so $\mu\big(\{x\in\mathbb{R}^d:\omega(x)>0\}\big)=1$, and consequently $\mathrm{supp}(\mu)\subseteq\mathrm{cl}\big(\{x\in\mathbb{R}^d:\omega(x)>0\}\big)$. Moreover, if $z\in\mathrm{cl}\big(\{x\in\mathbb{R}^d:\omega(x)>0\}\big)$ then every neighbourhood of z has positive μ-measure, so $z\in\mathrm{supp}(\mu)$. The second conclusion therefore follows. □

Recall the definitions of $\underline{\lambda}_\beta(P)$ and $\mathcal{P}_{\mathrm{Reg}}(\tau)$ from Section 3.1.

LEMMA S9. *Given any distribution $P\in\mathcal{P}_{\mathrm{Reg}}(\tau)$, we have $\underline{\lambda}_\beta(P)=\lambda_{\beta,\flat}(P)$ where*

$$\lambda_{\beta,\flat}(P):=\sup\left\{\frac{|\eta(x_0)-\eta(x_1)|}{\|x_0-x_1\|_\infty^\beta}:x_0\neq x_1,\eta(x_0)\wedge\eta(x_1)>\tau,\omega(x_0)\wedge\omega(x_1)>0\right\}\vee 1.$$

PROOF. First note that if $\eta(x)>\tau$ then $x\in\mathcal{X}_\tau(\eta)$ and if $\omega(x)>0$ then $x\in\mathrm{supp}(\mu)$, so $\lambda_{\beta,\flat}(P)\leq\underline{\lambda}_\beta(P)$.

Since $P\in\mathcal{P}_{\mathrm{Reg}}(\tau)$, it has a continuous regression function η such that $\eta^{-1}([\tau,\tau+\varepsilon))\subseteq\mathrm{supp}(\mu)$ for some $\varepsilon>0$. Now take distinct points $x_0,x_1\in\mathrm{supp}(\mu)\cap\mathcal{X}_\tau(\eta)$, and suppose initially that $\eta(x_0)\wedge\eta(x_1)>\tau$. Given $\epsilon'>0$, it follows from Lemma S8 that we may choose $x_0',x_1'\in\mathbb{R}^d$ with $\eta(x_0')\wedge\eta(x_1')>\tau$ and $\omega(x_0')\wedge\omega(x_1')>0$ such that

$$\max\big\{\|x_0-x_0'\|_\infty,\|x_1-x_1'\|_\infty,|\eta(x_0)-\eta(x_0')|,|\eta(x_1)-\eta(x_1')|\big\}\leq\epsilon'.$$

Consequently, we have

$$\begin{aligned}|\eta(x_0)-\eta(x_1)|&\leq|\eta(x_0')-\eta(x_1')|+2\epsilon'\leq\lambda_{\beta,\flat}(P)\cdot\|x_0'-x_1'\|_\infty^\beta+2\epsilon'\\ &\leq\lambda_{\beta,\flat}(P)\cdot\big(\|x_0-x_1\|_\infty+2\epsilon'\big)^\beta+2\epsilon'.\end{aligned}\tag{S15}$$

Since $\epsilon'>0$ was arbitrary, we deduce that $|\eta(x_0)-\eta(x_1)|\leq\lambda_{\beta,\flat}(P)\cdot\|x_0-x_1\|_\infty^\beta$ for all distinct $x_0,x_1\in\mathrm{supp}(\mu)\cap\eta^{-1}\big((\tau,1]\big)$. Now consider $x_0,x_1\in\mathcal{X}_\tau(\eta)\cap\mathrm{supp}(\mu)$ with $\eta(x_0)=\tau$ and $\eta(x_1)>\tau$. Given $\epsilon'\in\big(0,\min\{\varepsilon,\eta(x_1)-\tau\}\big)$, by the intermediate value theorem we may choose x_2 on the line segment between x_0 and x_1 with $\eta(x_2)=\tau+\epsilon'$. It then follows that $x_2 \in \eta^{-1}\big([\tau, \tau+\varepsilon)\big) \subseteq \mathrm{supp}(\mu)$. By the consequence of (S15), we deduce that $|\eta(x_2) - \eta(x_1)| \leq \lambda_{\beta,\flat}(P) \cdot \|x_2 - x_1\|_\infty^\beta$, and hence

$$|\eta(x_0) - \eta(x_1)| \leq |\eta(x_0) - \eta(x_2)| + |\eta(x_2) - \eta(x_1)| \leq \epsilon' + \lambda_{\beta,\flat}(P) \cdot \|x_2 - x_1\|_\infty^\beta \leq \epsilon' + \lambda_{\beta,\flat}(P) \cdot \|x_0 - x_1\|_\infty^\beta.$$

Again, since $\epsilon' \in \big(0, \min\{\varepsilon, \eta(x_1) - \tau\}\big)$ can be taken arbitrarily small, we deduce that $|\eta(x_0) - \eta(x_1)| \leq \lambda_{\beta,\flat}(P) \cdot \|x_0 - x_1\|_\infty^\beta$ for all $x_0, x_1 \in \mathcal{X}_\tau(\eta) \cap \mathrm{supp}(\mu)$ with $\eta(x_1) > \tau$. Finally, when $\eta(x_0) = \eta(x_1) = \tau$, the inequality $|\eta(x_0) - \eta(x_1)| \leq \lambda_{\beta,\flat}(P) \cdot \|x_0 - x_1\|_\infty^\beta$ is immediate. Thus, we have $|\eta(x_0) - \eta(x_1)| \leq \lambda_{\beta,\flat}(P) \cdot \|x_0 - x_1\|_\infty^\beta$ for all $x_0, x_1 \in \mathcal{X}_\tau(\eta) \cap \mathrm{supp}(\mu)$, so $\underline{\lambda}_\beta(P) \leq \lambda_{\beta,\flat}(P)$, as required. □

Our next lemma adapts ideas from McShane (1934).

LEMMA S10. *Given $P \in \mathcal{P}_{\mathrm{Reg}}(\tau)$ with $\underline{\lambda}_\beta(P) < \infty$, we have $P \in \mathcal{P}_{\mathrm{Höl}}\big(\beta, \underline{\lambda}_\beta(P), \tau\big)$.*

PROOF. Since $P \in \mathcal{P}_{\mathrm{Reg}}(\tau)$, it has a continuous regression function η_0 such that $\eta_0^{-1}\big([\tau, \tau+\varepsilon)\big) \subseteq \mathrm{supp}(\mu)$ for some $\varepsilon > 0$. We construct a regression function $\eta_1 : \mathbb{R}^d \to [0,1]$ by

$$\eta_1(x) := \begin{cases} 1 \wedge \inf\big\{\eta_0(z) + \underline{\lambda}_\beta(P) \cdot \|z - x\|_\infty^\beta : z \in \mathrm{supp}(\mu) \cap \mathcal{X}_\tau(\eta_0)\big\} & \text{if } x \in \mathcal{X}_\tau(\eta_0) \\ \eta_0(x) & \text{otherwise.} \end{cases}$$

First note that $\mathcal{X}_\tau(\eta_0) = \mathcal{X}_\tau(\eta_1)$. We claim that $|\eta_1(x) - \eta_1(x')| \leq \underline{\lambda}_\beta(P) \cdot \|x - x'\|_\infty^\beta$ for all $x, x' \in \mathcal{X}_\tau(\eta_0)$. Indeed, if $\mathrm{supp}(\mu) \cap \mathcal{X}_\tau(\eta_0) = \emptyset$, then $\eta_1(x) = \eta_1(x')$, and if $z \in \mathrm{supp}(\mu) \cap \mathcal{X}_\tau(\eta_0)$, then

$$\eta_0(z) + \underline{\lambda}_\beta(P) \cdot \|z - x\|_\infty^\beta \leq \eta_0(z) + \underline{\lambda}_\beta(P) \cdot \|z - x'\|_\infty^\beta + \underline{\lambda}_\beta(P) \cdot \|x - x'\|_\infty^\beta,$$

and taking an infimum over $z \in \mathrm{supp}(\mu) \cap \mathcal{X}_\tau(\eta_0)$ yields $\eta_1(x) - \eta_1(x') \leq \underline{\lambda}_\beta(P) \cdot \|x - x'\|_\infty^\beta$. By interchanging the roles of x and x', the claim follows. Our second claim is that $\eta_1(x) = \eta_0(x)$ for all $x \in \mathrm{supp}(\mu) \cap \mathcal{X}_\tau(\eta_0)$. To see this, first observe that $\eta_1(x) \leq \eta_0(x)$ for all $x \in \mathrm{supp}(\mu) \cap \mathcal{X}_\tau(\eta_0)$ by definition of η_1. On the other hand, by definition of $\underline{\lambda}_\beta(P)$, we have for any $z \in \mathrm{supp}(\mu) \cap \mathcal{X}_\tau(\eta_0)$ that

$$\eta_0(x) \leq \eta_0(z) + \underline{\lambda}_\beta(P) \cdot \|z - x\|_\infty^\beta,$$

so taking an infimum over $z \in \mathrm{supp}(\mu) \cap \mathcal{X}_\tau(\eta_0)$ yields $\eta_0(x) \leq \eta_1(x)$. In particular, it now follows that $\eta_0(x) = \eta_1(x)$ for all $x \in \mathrm{supp}(\mu)$. To show that $\eta_1 \in \mathcal{F}_{\mathrm{Höl}}\big(\beta, \underline{\lambda}_\beta(P), \mathcal{X}_\tau(\eta_1)\big)$, we must also verify that η_1 is continuous. To this end, let $x \in \mathbb{R}^d$. If $\eta_0(x) > \tau$, then since η_0 is continuous, we have for all sufficiently small $\epsilon' > 0$ and $z \in \bar{B}_{\epsilon'}(x)$ that $\eta_0(z) > \tau$. Hence $|\eta_1(x) - \eta_1(z)| \leq \underline{\lambda}_\beta(P) \cdot \|x - z\|_\infty^\beta$ by our first claim, so η_1 is continuous on $\eta_0^{-1}\big((\tau, 1]\big)$. On the other hand, if $\eta_0(x) < \tau + \varepsilon$ then, again, since η_0 is continuous, for all sufficiently small $\epsilon' > 0$ and $z \in \bar{B}_{\epsilon'}(x)$ we also have have $\eta_0(z) < \tau + \varepsilon$, so $x, z \in \mathrm{supp}(\mu) \cap \mathcal{X}_\tau(\eta_0)$. We deduce from our second claim that $\eta_1(x) = \eta_0(x)$ and $\eta_1(z) = \eta_0(z)$, so again, η_1 is continuous at x. We conclude that $\eta_1 \in \mathcal{F}_{\mathrm{Höl}}\big(\beta, \underline{\lambda}_\beta(P), \mathcal{X}_\tau(\eta_1)\big)$. Moreover, since η_0 is a regression function for P, and $\eta_0(x) = \eta_1(x)$ for all $x \in \mathrm{supp}(\mu)$, we have that η_1 is also a regression function for P, and hence $P \in \mathcal{P}_{\mathrm{Höl}}\big(\beta, \underline{\lambda}_\beta(P), \tau\big)$. □

Recall the definition of the classes $\mathcal{P}^+_{\mathrm{Höl}}(\beta, \lambda, \tau, \boldsymbol{\epsilon})$ from Section 3.1.

LEMMA S11. *Fix $\beta \in (0,1]$ and $\lambda \geq 1$. Then*

$$\mathcal{P}_{\mathrm{Reg}}(\tau) \cap \mathcal{P}_{\mathrm{H\ddot{o}l}}(\beta,\lambda,\tau) \subseteq \bigcup_{\underline{\epsilon} \in (0,\infty)^3} \mathcal{P}^{+}_{\mathrm{H\ddot{o}l}}(\beta,\lambda,\tau,\boldsymbol{\epsilon}).$$

PROOF. Let $P \in \mathcal{P}_{\mathrm{Reg}}(\tau) \cap \mathcal{P}_{\mathrm{H\ddot{o}l}}(\beta,\lambda,\tau)$. Then $\underline{\lambda}_\beta(P) \leq \lambda < \infty$. Moreover, since $P \in \mathcal{P}_{\mathrm{Reg}}(\tau)$, we have by Lemma S9 that $P \in \mathcal{P}^{+}_{\mathrm{H\ddot{o}l}}(\beta,\lambda,\tau,\boldsymbol{\epsilon})$ for sufficiently small $\underline{\epsilon} = (\epsilon_0,\epsilon_1,\epsilon_2) \in (0,1]^3$. □

Finally, we are in a position to prove Corollary 7.

PROOF OF COROLLARY 7. First, since $P \in \mathcal{P}^{+}_{\mathrm{H\ddot{o}l}}(\beta,\lambda,\tau,\boldsymbol{\epsilon})$ we may choose $x_0, x_1 \in \mathcal{X}_{\tau+\epsilon_0}(\eta)$ with $\|x_0 - x_1\|_\infty \geq \epsilon_1$, as well as $\min\{\omega(x_0),\omega(x_1)\} \geq \epsilon_2$ and

$$|\eta(x_0) - \eta(x_1)| \geq \frac{3}{4} \cdot \underline{\lambda}_\beta(P) \cdot \|x_0 - x_1\|_\infty^\beta \cdot \mathbb{1}_{\{\underline{\lambda}_\beta(P) > 1\}}.$$

Writing $\lambda = \underline{\lambda}_\beta(P)$, we have by Lemma S10 that $P \in \mathcal{P}_{\mathrm{H\ddot{o}l}}(\beta,\lambda,\tau)$. Define

$$\begin{aligned}\Delta &:= 192 \cdot \lambda^{d/(2\beta+d)} \cdot \left(\frac{\log(2n/\delta)}{n\{\omega(x_0) \wedge \omega(x_1)\}}\right)^{\beta/(2\beta+d)} \\ &\leq 192 \cdot \lambda^{d/(2\beta+d)} \cdot \left(\frac{\log(2n/\delta)}{n \cdot \epsilon_2}\right)^{\beta/(2\beta+d)}.\end{aligned}$$

Observe that when $n \in \mathbb{N}$ and $\delta \in (0,1)$ satisfy (11), we have $\Delta \leq \min\{\epsilon_0, \lambda\epsilon_1^\beta/4\}$. By Theorem 6 we have $\mathbb{P}_P(\mathcal{E}) \geq 1-\delta$, where

$$\mathcal{E} := \left\{\frac{|\eta(x_0) - \eta(x_1)| - \Delta}{\|x_0 - x_1\|_\infty^\beta} \leq \hat{\lambda}_{n,\beta,\delta} \leq \lambda\right\}.$$

Note that on the event $\mathcal{E}$, we have $2\hat{\lambda}_{n,\beta,\delta} \leq 2\lambda$, and if $\lambda = 1$, then $\lambda \leq 2\hat{\lambda}_{n,\beta,\delta}$ is immediate. On the other hand, if $\lambda > 1$ then on the event $\mathcal{E}$, we have

$$\frac{|\eta(x_0) - \eta(x_1)| - \Delta}{\|x_0 - x_1\|_\infty^\beta} \geq \frac{3}{4} \cdot \lambda - \frac{\lambda \cdot \epsilon_1^\beta}{4 \cdot \|x_0 - x_1\|_\infty^\beta} \geq \frac{\lambda}{2},$$

as required. □

PROOF OF THEOREM 8. First, for $\lambda' \in [1,\infty)$, let $S_{\lambda'}$ denote the union $\bigcup_{\ell \in [\ell_{\alpha_n}]} B_{(\ell)}$ appearing in line 6 of Algorithm 1 when it is applied with $\alpha_n = (\alpha/2) \wedge (1/n)$ in place of α and λ' in place of λ, and let $\hat{A}_{\lambda'}$ denote the corresponding output set in $\mathcal{A}$. Observe that if $\lambda'_0 \leq \lambda'_1$ then $S_{\lambda'_1} \subseteq S_{\lambda'_0}$, since the p-values in (9) satisfy $\hat{p}_{n,\beta,\lambda'_0}(\cdot) \leq \hat{p}_{n,\beta,\lambda'_1}(\cdot)$. Consequently, $\hat{\mu}_n(\hat{A}_{\lambda'_1}) \leq \hat{\mu}_n(\hat{A}_{\lambda'_0})$ by line 6 in Algorithm 1.

(i) Let $P \in \mathcal{P}_{\mathrm{Reg}}(\tau) \cap \mathcal{P}^{+}_{\mathrm{H\ddot{o}l}}(\beta,\lambda,\tau,\boldsymbol{\epsilon})$. By Lemma S10 we have $P \in \mathcal{P}_{\mathrm{H\ddot{o}l}}\big(\beta,\underline{\lambda}_\beta(P),\tau\big)$. Hence, by Lemma S1 and Corollary 7 we have

$$\begin{aligned}\mathbb{P}_P\big(\hat{A}'_{\mathrm{OSS}}(\mathcal{D}) \not\subseteq \mathcal{X}_\tau(\eta)\big) &\leq \mathbb{P}_P\big(S_{2\hat{\lambda}_{n,\beta,\alpha_n}} \not\subseteq \mathcal{X}_\tau(\eta)\big) \\ &\leq \mathbb{P}_P\big(S_{\underline{\lambda}_\beta(P)} \not\subseteq \mathcal{X}_\tau(\eta)\big) + \mathbb{P}_P\big(S_{2\hat{\lambda}_{n,\beta,\alpha_n}} \not\subseteq S_{\underline{\lambda}_\beta(P)}\big) \\ &\leq \mathbb{P}_P\big(S_{\underline{\lambda}_\beta(P)} \not\subseteq \mathcal{X}_\tau(\eta)\big) + \mathbb{P}_P\big(2\hat{\lambda}_{n,\beta,\alpha_n} < \underline{\lambda}_\beta(P)\big) \leq 2\alpha_n \leq \alpha. \qquad \text{(S16)}\end{aligned}$$

(ii) Now suppose that $P \in \mathcal{P}_{\mathrm{Reg}}(\tau) \cap \mathcal{P}^{+}_{\mathrm{Höl}}(\beta,\lambda,\tau,\boldsymbol{\epsilon}) \cap \mathcal{P}_{\mathrm{App}}(\mathcal{A},\kappa,\gamma,\tau,C_{\mathrm{App}})$. On the event $\{2\hat{\lambda}_{n,\beta,\alpha_n} \leq 2\underline{\lambda}_\beta(P)\}$, we have $\hat{\mu}_n(\hat{A}'_{\mathrm{OSS}}) = \hat{\mu}_n(\hat{A}_{2\hat{\lambda}_{n,\beta,\alpha_n}}) \geq \hat{\mu}_n(\hat{A}_{2\underline{\lambda}_\beta(P)})$. Hence, by Lemma S36, we have

$$
\begin{aligned}
\mathbb{E}_P\{\mu(\hat{A}_{2\underline{\lambda}_\beta(P)}) - \mu(\hat{A}'_{\mathrm{OSS}})\} &= \mathbb{E}_P\{\mu(\hat{A}_{2\underline{\lambda}_\beta(P)}) - \hat{\mu}_n(\hat{A}_{2\underline{\lambda}_\beta(P)}) \\
&\qquad + \hat{\mu}_n(\hat{A}_{2\underline{\lambda}_\beta(P)}) - \hat{\mu}_n(\hat{A}'_{\mathrm{OSS}}) + \hat{\mu}_n(\hat{A}'_{\mathrm{OSS}}) - \mu(\hat{A}'_{\mathrm{OSS}})\} \\
&\leq 2C_{\mathrm{VC}}\sqrt{\frac{\dim_{\mathrm{VC}}(\mathcal{A})}{n}} =: \epsilon_n^{\mathrm{VC}}. \qquad \text{(S17)}
\end{aligned}
$$

By Lemmas S10 and S1 we have $\mathbb{P}_P\big(\hat{A}_{2\underline{\lambda}_\beta(P)} \nsubseteq \mathcal{X}_\tau(\eta)\big) \leq \mathbb{P}_P\big(S_{2\underline{\lambda}_\beta(P)} \nsubseteq \mathcal{X}_\tau(\eta)\big) \leq \alpha_n$. It follows by (S16) and (S17) that

$$
\begin{aligned}
&\mathbb{E}_P\{(M_\tau - \mu(\hat{A}'_{\mathrm{OSS}})) \cdot \mathbb{1}_{\{\hat{A}'_{\mathrm{OSS}} \subseteq \mathcal{X}_\tau(\eta)\}}\} \leq \mathbb{E}_P\{M_\tau - \mu(\hat{A}'_{\mathrm{OSS}})\} + \mathbb{P}_P\big(\hat{A}'_{\mathrm{OSS}}(\mathcal{D}) \nsubseteq \mathcal{X}_\tau(\eta)\big) \\
&\quad \leq \mathbb{E}_P\{M_\tau - \mu(\hat{A}_{2\underline{\lambda}_\beta(P)})\} + \epsilon_n^{\mathrm{VC}} + 2\alpha_n \\
&\quad \leq \mathbb{E}_P\{(M_\tau - \mu(\hat{A}_{2\underline{\lambda}_\beta(P)})) \cdot \mathbb{1}_{\{\hat{A}_{2\underline{\lambda}_\beta(P)} \subseteq \mathcal{X}_\tau(\eta)\}}\} + \mathbb{P}_P\big(\hat{A}_{2\underline{\lambda}_\beta(P)} \nsubseteq \mathcal{X}_\tau(\eta)\big) + \epsilon_n^{\mathrm{VC}} + 2\alpha_n \\
&\quad \leq \mathbb{E}_P\big(M_\tau - \mu(\hat{A}_{2\underline{\lambda}_\beta(P)}) \,\big|\, \hat{A}_{2\underline{\lambda}_\beta(P)} \subseteq \mathcal{X}_\tau(\eta)\big) + 3\alpha_n + \epsilon_n^{\mathrm{VC}} \\
&\quad = R_\tau(\hat{A}_{2\underline{\lambda}_\beta(P)}) + 3\alpha_n + \epsilon_n^{\mathrm{VC}} \\
&\quad \leq C'\left\{\left(\frac{\big(2\underline{\lambda}_\beta(P)\big)^{d/\beta}}{n} \cdot \log_+\left(\frac{n}{(\alpha/2) \wedge (1/n)}\right)\right)^{\frac{\beta\kappa\gamma}{\kappa(2\beta+d)+\beta\gamma}} + \frac{1}{n^{1/2}}\right\} + 3\alpha_n + \epsilon_n^{\mathrm{VC}},
\end{aligned}
$$

where we have used Proposition 5 with C' in place of C for the final inequality. The regret bound on $\hat{A}'_{\mathrm{OSS}}$ follows by once again using $\mathbb{P}_P\big(\hat{A}'_{\mathrm{OSS}}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta)\big) \geq 1 - \alpha_n \geq 1/2$. □

S4.2. *Proofs from Section 3.2.* Recall the definition of the event $\mathcal{E}_{\eta,\delta}$ from Section S4.1.

LEMMA S12. *Let P be a distribution with regression function $\eta \in \mathcal{F}_{\mathrm{Höl}}(\beta,\lambda,\mathbb{R}^d)$. On the event $\mathcal{E}_{\eta,\delta}$, we have $v \geq \beta \cdot u - \log\lambda$ for all $(u,v) \in \hat{\Gamma}^{\dagger}_{n,\delta}$.*

PROOF. If $(u,v) \in \hat{\Gamma}^{\dagger}_{n,\delta}$, then we can find $(i,j,k,\ell) \in [n]^4$ with $\hat{\varphi}^{\dagger}_{n,\beta,\delta}(i,j,k,\ell) < \infty$ such that $u = \hat{\varepsilon}^{\dagger}_{n,\beta,\delta}(i,j,k,\ell)$ and $v = \hat{\varphi}^{\dagger}_{n,\beta,\delta}(i,j,k,\ell)$. Thus, on the event $\mathcal{E}_{\eta,\delta}$, we have

$$
\begin{aligned}
v &\geq -\log\left(\left|\frac{\sum_{t=1}^n \eta(X_t) \cdot \mathbb{1}_{\{X_t \in \bar{B}_{r_{i,k}}(X_i)\}}}{\sum_{t=1}^n \mathbb{1}_{\{X_t \in \bar{B}_{r_{i,k}}(X_i)\}}} - \frac{\sum_{t=1}^n \eta(X_t) \cdot \mathbb{1}_{\{X_t \in \bar{B}_{r_{j,\ell}}(X_j)\}}}{\sum_{t=1}^n \mathbb{1}_{\{X_t \in \bar{B}_{r_{j,\ell}}(X_j)\}}}\right|\right) \\
&\geq -\log\big(\lambda(\|X_i - X_j\|_\infty + r_{i,k} + r_{j,\ell})^\beta\big) = \beta \cdot u - \log\lambda,
\end{aligned}
$$

where we have applied $\eta \in \mathcal{F}_{\mathrm{Höl}}(\beta,\lambda,\mathbb{R}^d)$ in the final inequality. □

Now for $x \in \mathbb{R}^d$ and $s > 0$ we define the event $\mathcal{E}_\mu(x,s) := \big\{\hat{\mu}_n\big(\bar{B}_s(x)\big) \geq \frac{1}{2} \cdot \mu\big(\bar{B}_s(x)\big)\big\}$.

LEMMA S13. *Let P be a distribution with regression function $\eta \in \mathcal{F}_{\mathrm{Höl}}(\beta,\lambda,\mathbb{R}^d)$. Suppose that x_0, $x_1 \in \mathcal{R}_\circ(\mu,c_0)$ and $s \in (0,1]$ satisfy*

$$
|\eta(x_0) - \eta(x_1)| \geq 4\sqrt{\frac{\log(4n^2/\delta)}{c_0 n s^d}} + 6\lambda s^\beta.
$$

Then, on the event $\mathcal{E}_{\eta,\delta} \cap \mathcal{E}_\mu(x_0, s) \cap \mathcal{E}_\mu(x_1, s)$*, there exists* $(u, v) \in \hat{\Gamma}^{\dagger}_{n,\delta}$ *with*

$$u \geq -\log\bigl(\|x_0 - x_1\|_\infty + 6s\bigr)$$
$$v \leq -\log\biggl(|\eta(x_0) - \eta(x_1)| - 4\sqrt{\frac{\log(4n^2/\delta)}{c_0 n s^d}} - 6\lambda s^\beta\biggr).$$

PROOF. Suppose that the event $\mathcal{E}_{\eta,\delta} \cap \mathcal{E}_\mu(x_0, s) \cap \mathcal{E}_\mu(x_1, s)$ holds. Then in particular,

$$\min\bigl\{\hat{\mu}_n\bigl(\bar{B}_s(x_0)\bigr), \hat{\mu}_n\bigl(\bar{B}_s(x_1)\bigr)\bigr\} \geq \frac{1}{2} \cdot c_0 \cdot s^d > 0. \tag{S18}$$

As such, we may choose $X_i \in \bar{B}_s(x_0)$ and $X_j \in \bar{B}_s(x_1)$. Moreover, letting $k = \ell = \lceil \frac{n}{2} \cdot c_0 \cdot s^d \rceil$, it follows from (S18) that $\max(r_{i,k}, r_{j,\ell}) \leq 2s$. Now take $u = \hat{\varepsilon}^{\dagger}_{n,\beta,\delta}(i, j, k, \ell)$ and $v = \hat{\varphi}^{\dagger}_{n,\beta,\delta}(i, j, k, \ell)$ so that $(u, v) \in \hat{\Gamma}^{\dagger}_{n,\delta}$. The lower bound on u follows. The upper bound on v then follows from our event combined with the facts that $\bar{B}_{r_{i,k}}(X_i) \subseteq \bar{B}_{3s}(x_0)$ and $\bar{B}_{r_{j,k}}(X_j) \subseteq \bar{B}_{3s}(x_1)$. □

LEMMA S14. *Let* $P \in \mathcal{P}^{\dagger}_{\text{Höl}}(\beta, \lambda, \lambda_0, c_0, r_0)$ *with* $\lambda_0 \leq \lambda$ *and* $0 < r \leq r' \leq r_0 \leq 1$*. Let* $\mathcal{E}^{\dagger}_{\delta}(r, r')$ *denote the event that there exist* $(u, v), (u', v') \in \hat{\Gamma}^{\dagger}_{n,\delta}$ *with* $u \geq -\log(7r)$*,* $u' \geq -\log(7r')$*,* $v \leq -\log\bigl((1/2)\lambda_0 r^\beta\bigr)$ *and* $v' \leq -\log\bigl((1/2)\lambda_0 (r')^\beta\bigr)$*. Then provided that*

$$\frac{n}{\log(4n^2/\delta)} \geq \frac{\bigl(16\lambda/\lambda_0\bigr)^{d/\beta}}{c_0} \cdot \max\biggl\{\frac{2^{10}}{\lambda_0^2 \cdot r^{2\beta+d}}, \frac{8}{r^d}\biggr\}, \tag{S19}$$

we have $\mathbb{P}\bigl(\mathcal{E}_{\eta,\delta} \cap \mathcal{E}^{\dagger}_{\delta}(r, r')\bigr) \geq 1 - \delta$*.*

PROOF. Since $P \in \mathcal{P}^{\dagger}_{\text{Höl}}(\beta, \lambda, \lambda_0, c_0, r_0)$, there exist $x_0, x_1, x'_0, x'_1 \in \mathcal{R}_\circ(\mu, c_0)$ with $\|x_0 - x_1\|_\infty \leq r$, $\|x'_0 - x'_1\|_\infty \leq r'$, $|\eta(x_0) - \eta(x_1)| \geq \lambda_0 \cdot r^\beta$ and $|\eta(x'_0) - \eta(x'_1)| \geq \lambda_0 \cdot r'^\beta$. Now let $s := \{\lambda_0/(16\lambda)\}^{1/\beta} \cdot r \in (0, 1]$ and introduce the event

$$\tilde{\mathcal{E}}_\delta(r, r') := \mathcal{E}_{\eta,\delta} \cap \bigcap_{z \in \{x_0, x_1, x'_0, x'_1\}} \mathcal{E}_\mu(z, s).$$

When (S19) holds, we have

$$4\sqrt{\frac{\log(4n^2/\delta)}{c_0 n s^d}} + 6\lambda s^\beta = 4\sqrt{\frac{(16\lambda/\lambda_0)^{d/\beta} \log(4n^2/\delta)}{c_0 n r^d}} + \frac{3}{8} \cdot \lambda_0 r^\beta \leq \frac{1}{2} \cdot \lambda_0 r^\beta.$$

Hence, by Lemma S13, on the event $\tilde{\mathcal{E}}_\delta(r, r')$ and when (S19) holds, there exist pairs $(u, v), (u', v') \in \hat{\Gamma}^{\dagger}_{n,\delta}$ with $u \geq -\log(7r)$, $u' \geq -\log(7r')$, $v \leq -\log\bigl((1/2)\lambda_0 r^\beta\bigr)$ and $v' \leq -\log\bigl((1/2)\lambda_0 (r')^\beta\bigr)$. Thus, when (S19) holds, we have

$$\mathbb{P}\bigl(\mathcal{E}_{\eta,\delta} \cap \mathcal{E}^{\dagger}_{\delta}(r, r')\bigr) \geq \mathbb{P}\bigl(\tilde{\mathcal{E}}_\delta(r, r')\bigr) \geq 1 - \delta$$

by Hoeffding's inequality (Lemma S38) and the multiplicative Chernoff bound (Lemma S39). □

We are now in a position to prove Theorem 9.

PROOF OF THEOREM 9. First, for $n \in \mathbb{N}$ satisfying (13), let

$$r_n := \left\{ \frac{2^{10}\big(16\lambda/\lambda_0\big)^{d/\beta}}{c_0 \cdot \lambda_0^2} \cdot \frac{\log(4n^2/\delta)}{n} \right\}^{1/(2\beta+d)}.$$

and $r_n' := n^{-1/(7+2d)}$. The sample size condition (13) ensures that (S19) holds and that $r_n \leq r_n' \leq r_0$, so we may apply Lemma S14 to obtain $\mathbb{P}\big(\mathcal{E}_{\eta,\delta} \cap \mathcal{E}_\delta^\dagger(r_n, r_n')\big) \geq 1-\delta$.

Next, we show that on the event $\mathcal{E}_{\eta,\delta} \cap \mathcal{E}_\delta^\dagger(r_n, r_n')$ we have $\hat{\beta}_{n,\delta} \leq \beta$. Indeed, suppose that $\mathcal{E}_{\eta,\delta} \cap \mathcal{E}_\delta^\dagger(r_n, r_n') \cap \{\hat{\beta}_{n,\delta} > 0\}$ holds and let $(u_0, v_0) \in \hat{\Gamma}_{n,\delta}^\dagger$ be such that $u_0 \leq \frac{\log n}{6+2d}$. By Lemma S12 we have $v_0 \geq \beta \cdot u_0 - \log \lambda$. Since $\mathcal{E}_\delta^\dagger(r_n, r_n')$ holds, there exists $(u_1, v_1) \in \hat{\Gamma}_{n,\delta}^\dagger$ with $u_1 \geq -\log(7r_n)$ and $v_1 \leq -\log\big((1/2)\lambda_0 r_n^\beta\big) \leq \beta u_1 + \log(14/\lambda_0)$. Moreover, by (13) we have $u_1 \geq \frac{\log n}{3+d} \geq 2u_0$. Hence,

$$\hat{\beta}_{n,\delta} \leq \frac{v_1 - v_0 - \log f(n)}{u_1 - u_0} \leq \beta + \frac{\log(14\lambda/\lambda_0) - \log f(n)}{u_1 - u_0} \leq \beta,$$

where we have again applied (13) to ensure that $f(n) \geq 14\lambda/\lambda_0$.

Finally, we show that on the event $\mathcal{E}_{\eta,\delta} \cap \mathcal{E}_\delta^\dagger(r_n, r_n')$ we have $\hat{\beta}_{n,\delta} \geq \beta - \frac{2(7+2d)\log f(n)}{\log n}$. Indeed, on $\mathcal{E}_\delta^\dagger(r_n, r_n') \cap \{\hat{\beta}_{n,\delta} < \beta\}$, there exists $(u_0, v_0) \in \hat{\Gamma}_{n,\delta}^\dagger$ with $u_0 \geq -\log(7r_n')$ and $v_0 \leq -\log\big((1/2)\lambda_0 (r_n')^\beta\big) \leq \beta \cdot u_0 + \log\big(14/\lambda_0\big)$. Now fix any $(u_1, v_1) \in \hat{\Gamma}_{n,\delta}^\dagger$ satisfying $u_1 \geq 2u_0$. By Lemma S12 we have $v_1 \geq \beta \cdot u_1 - \log \lambda$ and consequently,

$$\hat{\beta}_{n,\delta} \geq \frac{v_1 - v_0 - \log f(n)}{u_1 - u_0} \geq \beta - \frac{\log(14\lambda/\lambda_0) + \log f(n)}{u_1 - u_0} \geq \beta - \frac{2(7+2d)\log f(n)}{\log n},$$

as required. □

The following lemma is analogous to Corollary 7 but with an estimated Hölder exponent.

LEMMA S15. *Fix* $\beta \in (0,1]$, $d \in \mathbb{N}$, $\lambda \in [1,\infty)$, $\lambda_0 \in (0,\lambda]$, $c_0, r_0 \in (0,1]$ *and* $\boldsymbol{\epsilon} = (\epsilon_0, \epsilon_1, \epsilon_2) \in (0,1]^3$ *and take* $P \in \mathcal{P}_{\mathrm{Reg}}(\tau) \cap \mathcal{P}_{\mathrm{H\ddot{o}l}}^+(\beta,\lambda,\tau,\boldsymbol{\epsilon}) \cap \mathcal{P}_{\mathrm{H\ddot{o}l}}^\dagger(\beta,\lambda,\lambda_0,c_0,r_0)$. *Let* $n \in \mathbb{N}$ *and* $\delta \in (0,1)$ *be such that* $14\lambda/\lambda_0 \leq f(n) \leq n^{\frac{\log(9/7)}{2(7+2d)\log(1/\epsilon_2)}}$ *the sample size condition* (13) *holds and*

$$\text{(S20)} \qquad \frac{n}{\log(2n/\delta)} \geq \frac{1}{\epsilon_2} \cdot \left\{ f(n)^{2(7+2d)} \cdot \left(192 \cdot \max\left\{\frac{\lambda}{\epsilon_0}, \frac{12}{\epsilon_1}\right\}\right)^{2+d} \right\}^{1/\beta}.$$

Then

$$\mathbb{P}_P\left(\left\{ \beta - \frac{2(7+2d)\log f(n)}{\log n} \leq \hat{\beta}_{n,\delta} \leq \beta \right\} \cap \left\{ \frac{\underline{\lambda}_\beta(P)}{2} \leq \hat{\lambda}_{n,\hat{\beta}_{n,\delta},\delta} \leq \underline{\lambda}_\beta(P) \right\} \right) \geq 1 - 2\delta.$$

PROOF OF LEMMA S15. For brevity we write $\underline{\lambda}_\beta = \underline{\lambda}_\beta(P)$. Since $P \in \mathcal{P}_{\mathrm{H\ddot{o}l}}^+(\beta,\lambda,\tau,\boldsymbol{\epsilon})$, we may choose $x_0, x_1 \in \mathcal{X}_{\tau+\epsilon_0}(\eta)$ with $\|x_0 - x_1\|_\infty \geq \epsilon_1$, as well as $\min\{\omega(x_0), \omega(x_1)\} \geq \epsilon_2$ and

$$|\eta(x_0) - \eta(x_1)| \geq \frac{3}{4} \cdot \underline{\lambda}_\beta \cdot \|x_0 - x_1\|_\infty^\beta \cdot \mathbb{1}_{\{\underline{\lambda}_\beta > 1\}}.$$

By Lemma S10, we have $P \in \mathcal{P}_{\mathrm{H\ddot{o}l}}(\beta, \underline{\lambda}_\beta, \tau)$. Hence, on the event $\{\hat{\beta}_{n,\delta} \leq \beta\}$ we have $P \in \mathcal{P}_{\mathrm{H\ddot{o}l}}(\hat{\beta}_{n,\delta}, \underline{\lambda}_\beta, \tau)$. Letting $\beta_n^\flat := \max\{\beta - 2(7+2d)\log f(n)/\log n, 0\}$, it follows from Theorem 9, combined with the sample size condition (13) and the lower bound on $f(n)$, that

$\mathbb{P}\big(\beta_n^\flat \le \hat{\beta}_{n,\delta} \le \beta\big) \ge 1-\delta$. In addition, define $\hat{\Delta}_n := \Delta_{n,\hat{\beta}_{n,\delta},\underline{\lambda}_\beta}(x_0) \vee \Delta_{n,\hat{\beta}_{n,\delta},\underline{\lambda}_\beta}(x_1)$, so that provided $\hat{\beta}_{n,\delta} \ge \beta_n^\flat$ we have

$$\hat{\Delta}_n = 192 \cdot \underline{\lambda}_\beta^{d/(2\hat{\beta}_{n,\delta}+d)} \cdot \biggl(\frac{\log(2n/\delta)}{n\{\omega(x_0)\wedge\omega(x_1)\}}\biggr)^{\hat{\beta}_{n,\delta}/(2\hat{\beta}_{n,\delta}+d)}$$
$$\le 192 \cdot \underline{\lambda}_\beta \cdot \biggl(\frac{\log(2n/\delta)}{n\cdot\epsilon_2}\biggr)^{\beta_n^\flat/(2+d)} \le \min\Bigl\{\epsilon_0, \frac{\underline{\lambda}_\beta\epsilon_1}{12}\Bigr\},$$

where we have applied the sample size condition (S20) in both inequalities. In particular, we have $\min_{x\in\{x_0,x_1\}}\{\eta(x) - \Delta_{n,\beta,\lambda}(x)\} \ge \tau$ on the event $\bigl\{\beta_n^\flat \le \hat{\beta}_{n,\delta}\bigr\}$ and under our sample size conditions. Consequently, if we let $\tilde{\mathcal{E}}_{\beta,\underline{\lambda}_\beta,\delta}$ denote the event

$$\tilde{\mathcal{E}}_{\beta,\underline{\lambda}_\beta,\delta} := \bigl\{\beta_n^\flat \le \hat{\beta}_{n,\delta} \le \beta\bigr\} \cap \biggl\{\frac{|\eta(x_0)-\eta(x_1)| - \hat{\Delta}_n}{\|x_0-x_1\|_\infty^{\hat{\beta}_{n,\delta}}} \le \hat{\lambda}_{n,\hat{\beta}_{n,\delta},\delta} \le \underline{\lambda}_\beta\biggr\},$$

then it follows from Proposition S7 that $\mathbb{P}\bigl(\tilde{\mathcal{E}}_{\beta,\underline{\lambda}_\beta,\delta}\bigr) \ge 1-2\delta$. Thus, to complete the proof it suffices to show that on the event $\tilde{\mathcal{E}}_{\beta,\underline{\lambda}_\beta,\delta}$ we have $\underline{\lambda}_\beta \le 2\hat{\lambda}_{n,\hat{\beta}_{n,\delta},\delta}$. If $\underline{\lambda}_\beta = 1$, then the required bound is immediate. On the other hand, if $\underline{\lambda}_\beta > 1$, then on the event $\tilde{\mathcal{E}}_{\beta,\underline{\lambda}_\beta,\delta}$, we have

$$\hat{\lambda}_{n,\hat{\beta}_{n,\delta},\delta} \ge \frac{|\eta(x_0)-\eta(x_1)| - \hat{\Delta}_n}{\|x_0-x_1\|_\infty^{\hat{\beta}_{n,\delta}}} \ge \frac{3}{4}\cdot\underline{\lambda}_\beta\cdot\epsilon_1^{\beta-\beta_n^\flat} - \frac{1}{12}\cdot\underline{\lambda}_\beta\cdot\epsilon_1^{1-\beta} \ge \frac{\underline{\lambda}_\beta}{2},$$

since $\epsilon_1^{\beta-\beta_n^\flat} \ge 7/9$ by the upper bound on $f(n)$. The result follows. □

We conclude this section by applying Lemma S15 to prove Theorem 10.

PROOF OF THEOREM 10. We proceed similarly to the proof of Theorem 8. For $\beta' \in (0,1]$, $\lambda' \in [1,\infty)$ we let $S^\circ_{\beta',\lambda'}$ denote the union $\bigcup_{\ell\in[\ell_{\tilde{\alpha}_n}]} B_{(\ell)}$ appearing in line 6 of Algorithm 1 when it is applied with $\tilde{\alpha}_n$ in place of α, β' in place of β and λ' in place of λ. Furthermore, we let $\hat{A}^\circ_{\beta',\lambda'}$ denote the corresponding output set in $\mathcal{A}$. If $\beta_0' \le \beta_1'$ and $\lambda_0' \ge \lambda_1'$, then the p-values in (9) satisfy $\hat{p}_{n,\beta_1',\lambda_1'}(\cdot) \le \hat{p}_{n,\beta_0',\lambda_0'}(\cdot)$, and so $S^\circ_{\beta_0',\lambda_0'} \subseteq S^\circ_{\beta_1',\lambda_1'}$. Consequently, $\hat{\mu}_n(\hat{A}^\circ_{\beta_0',\lambda_0'}) \le \hat{\mu}_n(\hat{A}^\circ_{\beta_1',\lambda_1'})$ by line 6 in Algorithm 1.

(i) Suppose that $P \in \mathcal{P}_{\mathrm{Reg}}(\tau) \cap \mathcal{P}^+_{\mathrm{Höl}}(\beta,\lambda,\tau,\boldsymbol{\epsilon}) \cap \mathcal{P}^\dagger_{\mathrm{Höl}}(\beta,\lambda,\lambda_0,c_0,r_0)$. Then by Lemmas S1, S10 and S15, we have

$$\begin{aligned}\mathbb{P}_P\bigl(\hat{A}''_{\mathrm{OSS}}(\mathcal{D}) \not\subseteq \mathcal{X}_\tau(\eta)\bigr) &\le \mathbb{P}_P\bigl(S^\circ_{\hat{\beta}_{n,\tilde{\alpha}_n},2\hat{\lambda}_{n,\hat{\beta}_{n,\tilde{\alpha}_n},\tilde{\alpha}_n}} \not\subseteq \mathcal{X}_\tau(\eta)\bigr)\\ &\le \mathbb{P}_P\bigl(S^\circ_{\beta,\underline{\lambda}_\beta} \not\subseteq \mathcal{X}_\tau(\eta)\bigr) + \mathbb{P}_P\bigl(S^\circ_{\hat{\beta}_{n,\tilde{\alpha}_n},2\hat{\lambda}_{n,\hat{\beta}_{n,\tilde{\alpha}_n},\tilde{\alpha}_n}} \not\subseteq S^\circ_{\beta,\underline{\lambda}_\beta}\bigr)\\ &\le \mathbb{P}_P\bigl(S^\circ_{\beta,\underline{\lambda}_\beta} \not\subseteq \mathcal{X}_\tau(\eta)\bigr) + \mathbb{P}_P\bigl(\hat{\beta}_{n,\tilde{\alpha}_n} > \beta\bigr) + \mathbb{P}_P\bigl(2\hat{\lambda}_{n,\hat{\beta}_{n,\tilde{\alpha}_n},\tilde{\alpha}_n} < \underline{\lambda}_\beta\bigr)\\ &\le 3\tilde{\alpha}_n \le \alpha,\end{aligned}$$

as required.

(ii) Now suppose further that $P \in \mathcal{P}_{\mathrm{Reg}}(\tau) \cap \mathcal{P}^+_{\mathrm{Höl}}(\beta,\lambda,\tau,\boldsymbol{\epsilon}) \cap \mathcal{P}^\dagger_{\mathrm{Höl}}(\beta,\lambda,\lambda_0,c_0,r_0) \cap \mathcal{P}_{\mathrm{App}}(\mathcal{A},\kappa,\gamma,\tau,C_{\mathrm{App}})$ so that by Theorem 9 we have $\mathbb{P}_P\bigl(\{\hat{\beta}_{n,\tilde{\alpha}_n} < \beta_n^\flat\} \cup \{\hat{\lambda}_{n,\hat{\beta}_{n,\tilde{\alpha}_n},\tilde{\alpha}_n} > \underline{\lambda}_\beta\}\bigr) \le 2\tilde{\alpha}_n$, where $\beta_n^\flat$ is defined as in the proof of Lemma S15. Moreover, on the event $\{\hat{\beta}_{n,\tilde{\alpha}_n} \ge \beta_n^\flat\} \cap \{\hat{\lambda}_{n,\hat{\beta}_{n,\tilde{\alpha}_n},\tilde{\alpha}_n} \le \underline{\lambda}_\beta\}$, we have

$$\hat{\mu}_n(\hat{A}''_{\mathrm{OSS}}) = \hat{\mu}_n(\hat{A}^\circ_{\hat{\beta}_{n,\tilde{\alpha}_n},2\hat{\lambda}_{n,\hat{\beta}_{n,\tilde{\alpha}_n},\tilde{\alpha}_n}}) \ge \hat{\mu}_n(\hat{A}^\circ_{\beta_n^\flat,2\underline{\lambda}_\beta}).$$

Hence by Lemma S36,

$$\mathbb{E}_P\{\mu(\hat{A}^{\circ}_{\beta_n^\flat,2\underline{\lambda}_\beta}) - \mu(\hat{A}''_{\mathrm{OSS}})\} \leq 2C_{\mathrm{VC}}\sqrt{\frac{\dim_{\mathrm{VC}}(\mathcal{A})}{n}} =: \epsilon_n^{\mathrm{VC}}.$$

By Lemma S1, we have $\mathbb{P}_P\bigl(\hat{A}^{\circ}_{\beta_n^\flat,2\underline{\lambda}_\beta} \not\subseteq \mathcal{X}_\tau(\eta)\bigr) \leq \mathbb{P}_P\bigl(S^{\circ}_{\beta_n^\flat,2\underline{\lambda}_\beta} \not\subseteq \mathcal{X}_\tau(\eta)\bigr) \leq \tilde{\alpha}_n$. Observe that we may assume without loss of generality that $(2\underline{\lambda}_\beta)^{d/\beta} \leq n$, because otherwise the regret bound is vacuous. Thus, combining with the derivation in **(i)**, we have

$$\begin{aligned}
&\mathbb{E}_P\bigl\{\bigl(M_\tau - \mu(\hat{A}''_{\mathrm{OSS}})\bigr)\cdot \mathbb{1}_{\{\hat{A}''_{\mathrm{OSS}}\subseteq \mathcal{X}_\tau(\eta)\}}\bigr\} \leq \mathbb{E}_P\bigl\{M_\tau - \mu(\hat{A}''_{\mathrm{OSS}})\bigr\} + \mathbb{P}_P\bigl(\hat{A}''_{\mathrm{OSS}}(\mathcal{D}) \not\subseteq \mathcal{X}_\tau(\eta)\bigr)\\
&\quad\leq \mathbb{E}_P\bigl\{M_\tau - \mu\bigl(\hat{A}^{\circ}_{\beta_n^\flat,2\underline{\lambda}_\beta}\bigr)\bigr\} + \epsilon_n^{\mathrm{VC}} + 3\tilde{\alpha}_n\\
&\quad\leq \mathbb{E}_P\bigl\{\bigl(M_\tau - \mu(\hat{A}^{\circ}_{\beta_n^\flat,2\underline{\lambda}_\beta})\bigr)\cdot \mathbb{1}_{\{\hat{A}^{\circ}_{\beta_n^\flat,2\underline{\lambda}_\beta}\subseteq \mathcal{X}_\tau(\eta)\}}\bigr\} + \mathbb{P}_P\bigl(\hat{A}^{\circ}_{\beta_n^\flat,2\underline{\lambda}_\beta} \not\subseteq \mathcal{X}_\tau(\eta)\bigr) + \epsilon_n^{\mathrm{VC}} + 3\tilde{\alpha}_n\\
&\quad\leq \mathbb{E}_P\bigl(M_\tau - \mu(\hat{A}^{\circ}_{\beta_n^\flat,2\underline{\lambda}_\beta}) \bigm| \hat{A}^{\circ}_{\beta_n^\flat,2\underline{\lambda}_\beta} \subseteq \mathcal{X}_\tau(\eta)\bigr) + 4\tilde{\alpha}_n + \epsilon_n^{\mathrm{VC}}\\
&\quad= R_\tau(\hat{A}^{\circ}_{\beta_n^\flat,2\underline{\lambda}_\beta}) + 4\tilde{\alpha}_n + \epsilon_n^{\mathrm{VC}}\\
&\quad\leq C'\biggl\{\biggl(\frac{(2\underline{\lambda}_\beta)^{d/\beta_n^\flat}}{n}\cdot\log_+\Bigl(\frac{n}{\tilde{\alpha}_n}\Bigr)\biggr)^{\frac{\beta_n^\flat\kappa\gamma}{\kappa(2\beta_n^\flat+d)+\beta_n^\flat\gamma}} + \frac{1}{n^{1/2}}\biggr\} + 4\tilde{\alpha}_n + \epsilon_n^{\mathrm{VC}}\\
&\quad\leq C'\biggl\{n^{\frac{2(\beta-\beta_n^\flat)\kappa\gamma}{\kappa(2\beta_n^\flat+d)+\beta_n^\flat\gamma}}\biggl(\frac{(2\underline{\lambda}_\beta)^{d/\beta}}{n}\log_+\Bigl(\frac{3n^2}{\alpha}\Bigr)\biggr)^{\frac{\beta\kappa\gamma}{\kappa(2\beta+d)+\beta\gamma}} + \frac{1}{n^{1/2}}\biggr\} + 4\tilde{\alpha}_n + \epsilon_n^{\mathrm{VC}}\\
&\quad\leq C'\biggl\{f(n)^{\frac{4\gamma(7+2d)}{d}}\biggl(\frac{(2\underline{\lambda}_\beta)^{d/\beta}}{n}\log_+\Bigl(\frac{3n^2}{\alpha}\Bigr)\biggr)^{\frac{\beta\kappa\gamma}{\kappa(2\beta+d)+\beta\gamma}} + \frac{1}{n^{1/2}}\biggr\} + 4\tilde{\alpha}_n + \epsilon_n^{\mathrm{VC}},
\end{aligned}$$

where in the third to last inequality we applied Proposition 5 with C' in place of C by noting that $P \in \mathcal{P}_{\mathrm{Höl}}(\beta,\underline{\lambda}_\beta,\tau) \subseteq \mathcal{P}_{\mathrm{Höl}}(\beta_n^\flat, 2\underline{\lambda}_\beta,\tau)$. The regret bound on $\hat{A}''_{\mathrm{OSS}}$ follows by using **(i)** once again. □

S5. Proof of the upper bound in Theorem 11. Recall that the upper bound in Theorem 11 will follow from Proposition 13. First we state the following consequence of Hoeffding's inequality.

LEMMA S16. *Fix $(\beta,\lambda) \in (0,\infty)\times[1,\infty)$ and let $P \in \mathcal{P}_{\mathrm{Höl}}(\beta,\lambda)$. Suppose that $\mathcal{D} = \bigl((X_i,Y_i)\bigr)_{i\in[n]} \sim P^{\otimes n}$ and let $\mathcal{D}_X = (X_i)_{i\in[n]}$. Given $x \in \mathbb{R}^d$, $h \in [0,1]$, and $\alpha \in (0,1)$ define*

$$\hat{\Delta}_{x,h}(\alpha) := \begin{cases} \sqrt{e_0^\top\bigl(Q^\beta_{x,h}\bigr)^{-1}e_0}\cdot\biggl(\lambda\cdot h^\beta|\mathcal{N}_{x,h}|^{1/2} + \sqrt{\frac{\log(1/\alpha)}{2}}\biggr) & \text{if } Q^\beta_{x,h} \text{ is invertible,}\\ 1 & \text{otherwise.}\end{cases}$$

Then

$$\max\bigl\{\mathbb{P}\bigl(\hat{\eta}(x) - \eta(x) \geq \hat{\Delta}_{x,h}(\alpha) \bigm| \mathcal{D}_X\bigr), \mathbb{P}\bigl(\eta(x) - \hat{\eta}(x) \geq \hat{\Delta}_{x,h}(\alpha) \bigm| \mathcal{D}_X\bigr)\bigr\} \leq \alpha.$$

PROOF. Fix a realisation of $\mathcal{D}_X = (X_i)_{i\in[n]}$. It suffices to restrict our attention to the case where $Q^\beta_{x,h}$ is invertible. Writing $u_i := \bigl\langle e_0, \bigl(Q^\beta_{x,h}\bigr)^{-1}\Phi^\beta_{x,h}(X_i)\bigr\rangle$ for $i \in [n]$, we have

$$\begin{aligned}\sum_{i\in\mathcal{N}_{x,h}} u_i\cdot\langle w^\beta_{x,h}, \Phi^\beta_{x,h}(X_i)\rangle &= e_0^\top\bigl(Q^\beta_{x,h}\bigr)^{-1}\sum_{i\in\mathcal{N}_{x,h}}\Phi^\beta_{x,h}(X_i)(w^\beta_{x,h})^\top\Phi^\beta_{x,h}(X_i) = e_0^\top w^\beta_{x,h}\\ &= \eta(x).\end{aligned}$$

Hence,

$$\langle e_0, \hat{w}^{\beta}_{x,h}\rangle = e_0^{\top}\left(Q^{\beta}_{x,h}\right)^{-1}V^{\beta}_{x,h} = \sum_{i\in\mathcal{N}_{x,h}} Y_i \cdot e_0^{\top}\left(Q^{\beta}_{x,h}\right)^{-1}\Phi^{\beta}_{x,h}(X_i)$$

$$= \sum_{i\in\mathcal{N}_{x,h}} u_i \cdot \Big(\big\{Y_i - \eta(X_i)\big\} + \big\{\eta(X_i) - \mathcal{T}^{\beta}_x[\eta](X_i)\big\} + \big\langle w^{\beta}_{x,h}, \Phi^{\beta}_{x,h}(X_i)\big\rangle\Big)$$

$$= \sum_{i\in\mathcal{N}_{x,h}} u_i \cdot \Big(\big\{Y_i - \eta(X_i)\big\} + \big\{\eta(X_i) - \mathcal{T}^{\beta}_x[\eta](X_i)\big\}\Big) + \eta(x). \tag{S21}$$

Note also that

$$\sum_{i\in\mathcal{N}_{x,h}} u_i^2 = \sum_{i\in\mathcal{N}_{x,h}} e_0^{\top}\left(Q^{\beta}_{x,h}\right)^{-1}\Phi^{\beta}_{x,h}(X_i)\Phi^{\beta}_{x,h}(X_i)^{\top}\left(Q^{\beta}_{x,h}\right)^{-1}e_0 = e_0^{\top}\left(Q^{\beta}_{x,h}\right)^{-1}e_0. \tag{S22}$$

In addition, since $P \in \mathcal{P}_{\mathrm{H\ddot{o}l}}(\beta,\lambda)$, for each $i \in \mathcal{N}_{x,h}$, we have

$$|\eta(X_i) - \mathcal{T}^{\beta}_x[\eta](X_i)| \le \lambda \cdot \|X_i - x\|^{\beta}_{\infty} \le \lambda \cdot h^{\beta}, \tag{S23}$$

and so by (S23), the Cauchy–Schwarz inequality and (S22), we have

$$\Bigg|\sum_{i\in\mathcal{N}_{x,h}} u_i \cdot \big\{\eta(X_i) - \mathcal{T}^{\beta}_x[\eta](X_i)\big\}\Bigg| \le \lambda \cdot h^{\beta} \cdot \sum_{i\in\mathcal{N}_{x,h}} |u_i|$$

$$\le \lambda \cdot h^{\beta} \cdot \sqrt{|\mathcal{N}_{x,h}| \cdot e_0^{\top}\left(Q^{\beta}_{x,h}\right)^{-1}e_0}. \tag{S24}$$

We conclude, by the definition of $\hat{\eta}(x)$, (S21), (S22), (S24) and Hoeffding's inequality, that

$$\mathbb{P}\Big(\hat{\eta}(x) - \eta(x) \ge \hat{\Delta}_{x,h}(\alpha)\Big) = \mathbb{P}\left(\sum_{i\in\mathcal{N}_{x,h}} u_i \cdot \big\{Y_i - \eta(X_i)\big\} \ge \sqrt{\frac{\log(1/\alpha)}{2}\sum_{i\in\mathcal{N}_{x,h}} u_i^2}\right) \le \alpha.$$

The other inequality follows similarly. □

LEMMA S17. *Fix $(\beta,\lambda) \in (0,\infty)\times[1,\infty)$ and let $P \in \mathcal{P}_{\mathrm{H\ddot{o}l}}(\beta,\lambda)$. Suppose that $\mathcal{D} = \big((X_i,Y_i)\big)_{i\in[n]} \sim P^{\otimes n}$ and let $\mathcal{D}_X = (X_i)_{i\in[n]}$. Then for any closed hyper-cube $B \subseteq \mathbb{R}^d$ with $\mathrm{diam}_{\infty}(B) \le 1$ and $\inf_{x'\in B}\eta(x') \le \tau$, and any $\alpha \in (0,1)$, we have*

$$\mathbb{P}\big(\hat{p}^{+}_n(B) \le \alpha \,|\, \mathcal{D}_X\big) \le \alpha.$$

PROOF. Recall that $x \in \mathbb{R}^d$ and $r \in [0,1/2]$ denote the centre and ℓ_∞-radius of B, and that $h = (2r)^{1\wedge\frac{1}{\beta}} \in [0,1]$. Again, it restrict our attention to the case where $Q^{\beta}_{x,h}$ is invertible. Since $\inf_{x'\in B}\eta(x') \le \tau$ and $P \in \mathcal{P}_{\mathrm{H\ddot{o}l}}(\beta,\lambda)$, we have $\eta(x) \le \tau + \lambda \cdot r^{\beta\wedge 1}$, and hence the lemma follows from Lemma S16. □

PROOF OF PROPOSITION 12. This follows from Lemma S17 in the same way as Proposition 4 followed from Lemma 3. □

We now turn to the proof of Proposition S2, which will rely on several lemmas. For $a > 0$, let $\mathcal{K}(a)$ denote the set of measurable sets $K \subseteq \bar{B}_1(0)$ with $\mathcal{L}_d(K) \ge a$.

LEMMA S18. *Given $d \in \mathbb{N}$, $\beta \in (0,\infty)$ and $a \in (0,1)$, we have*

$$c_{\min}(d,\beta,a) := 1 \wedge \inf_{K\in\mathcal{K}(a)}\left\{\lambda_{\min}\left(\int_K \Phi^{\beta}_{0,1}(z)\Phi^{\beta}_{0,1}(z)^{\top}\,d\mathcal{L}_d(z)\right)\right\} > 0.$$

PROOF. Suppose, for a contradiction, that $c_{\min}(d,\beta,a)=0$. Then we can find a sequence $(K^{(\ell)})_{\ell\in\mathbb{N}}$ in $\mathcal{K}(a)$, along with a sequence $(w^{(\ell)})_{\ell\in\mathbb{N}}$ with $w^{(\ell)}\in\mathbb{R}^{\mathcal{V}(\beta)}$, $\|w^{(\ell)}\|_2=1$ and

$$\lim_{\ell\to\infty}\int_{K^{(\ell)}}\langle w^{(\ell)},\Phi^{\beta}_{0,1}(z)\rangle^2\,d\mathcal{L}_d(z)=\lim_{\ell\to\infty}\bigl(w^{(\ell)}\bigr)^{\top}\biggl(\int_{K^{(\ell)}}\Phi^{\beta}_{0,1}(z)\Phi^{\beta}_{0,1}(z)^{\top}\,d\mathcal{L}_d(z)\biggr)w^{(\ell)}$$
$$=0. \tag{S25}$$

By moving to a subsequence if necessary, we may assume that $\lim_{\ell\to\infty}w^{(\ell)}=w^*$ for some $w^*\in\mathbb{R}^{\mathcal{V}(\beta)}$ with $\|w^*\|_2=1$. Now since $z\mapsto\langle w^*,\Phi^{\beta}_{0,1}(z)\rangle$ is a non-zero polynomial, the zero-set $\mathcal{Z}_{w^*}:=\bigl\{z\in\bar{B}_1(0):\langle w^*,\Phi^{\beta}_{0,1}(z)\rangle=0\bigr\}$ satisfies $\mathcal{L}_d(\mathcal{Z}_{w^*})=0$ (e.g. Okamoto, 1973, Lemma 1). In addition, by the continuity of $z\mapsto\langle w^*,\Phi^{\beta}_{0,1}(z)\rangle$, the set $\mathcal{Z}_{w^*}$ is closed. By countable additivity of the finite measure $\mathcal{L}_d|_{\bar{B}_1(0)}$, there exists $\epsilon_a>0$ such that $\mathcal{L}_d\bigl(\mathcal{Z}^{\epsilon_a}_{w^*}\bigr)\le a/2$ where $\mathcal{Z}^{\epsilon_a}_{w^*}:=\bigcup_{z\in\mathcal{Z}^{\epsilon_a}_{w^*}}B_{\epsilon}(z)=\mathcal{Z}_{w^*}+B_{\epsilon_a}(z)$. By continuity again, $\delta_a:=\inf_{z\in\bar{B}_1(0)\setminus\mathcal{Z}^{\epsilon_a}_{w^*}}\bigl|\langle w^*,\Phi^{\beta}_{0,1}(z)\rangle\bigr|>0$. Now choose $\ell_0\in\mathbb{N}$ sufficiently large that

$$\sup_{\ell\ge\ell_0}\bigl\|w^{(\ell)}-w^*\bigr\|_2\le\frac{\delta_a}{2\sqrt{|\mathcal{V}(\beta)|}},$$

so that, by Cauchy–Schwarz,

$$\bigl|\langle w^{(\ell)},\Phi^{\beta}_{0,1}(z)\rangle\bigr|\ge\bigl|\langle w^*,\Phi^{\beta}_{0,1}(z)\rangle\bigr|-\bigl|\langle w^{(\ell)}-w^*,\Phi^{\beta}_{0,1}(z)\rangle\bigr|\ge\delta_a-\frac{\delta_a}{2\sqrt{|\mathcal{V}(\beta)|}}\cdot\bigl\|\Phi^{\beta}_{0,1}(z)\bigr\|_2$$
$$\ge\frac{\delta_a}{2}$$

for all $\ell\ge\ell_0$ and $z\in\bar{B}_1(0)\setminus\mathcal{Z}^{\epsilon_a}_{w^*}$. Hence, for all $\ell\ge\ell_0$,

$$\int_{K^{(\ell)}}\langle w^{(\ell)},\Phi^{\beta}_{0,1}(z)\rangle^2\,d\mathcal{L}_d(z)\ge\int_{K^{(\ell)}\setminus\mathcal{Z}^{\epsilon_a}_{w^*}}\langle w^{(\ell)},\Phi^{\beta}_{0,1}(z)\rangle^2\,d\mathcal{L}_d(z)\ge\frac{a\cdot\delta_a^2}{8}>0,$$

which contradicts (S25), and completes the proof of the lemma. □

LEMMA S19. *Suppose that $\upsilon\in(0,1)$, $\xi\in(0,\infty)$, $\beta\in(0,\infty)$, $x\in\mathbb{R}^d$ and $r\in(0,1/2]$ satisfy $\bar{B}_r(x)\cap\mathcal{R}_{\upsilon}(\mu)\cap\mathcal{X}_{\xi}(f_\mu)\neq\emptyset$. Given any $h\in[2r,1]$, we have $\mu\bigl(\bar{B}_h(x)\bigr)\ge\upsilon\cdot\xi\cdot(h/2)^d$. In addition, if either $\beta\in(0,1]$ or $3r\le\upsilon h$, then*

$$\lambda_{\min}\biggl(\int_{\bar{B}_h(x)}\Phi^{\beta}_{x,h}(z)\Phi^{\beta}_{x,h}(z)^{\top}\,d\mu(z)\biggr)\ge2^{-(3d+1)}\cdot\upsilon\cdot c^0_{\min}\cdot\mu\bigl(\bar{B}_h(x)\bigr), \tag{S26}$$

where $c^0_{\min}\equiv c_{\min}(d,\beta,2^{-(d+1)}\upsilon)\in(0,1]$ is taken from Lemma S18.

PROOF. First choose $x_0\in\bar{B}_r(x)\cap\mathcal{R}_{\upsilon}(\mu)\cap\mathcal{X}_{\xi}(f_\mu)$, noting that $\bar{B}_{h/2}(x_0)\subseteq\bar{B}_{h-r}(x_0)\subseteq\bar{B}_h(x)$. Hence, since $x_0\in\mathcal{R}_{\upsilon}(\mu)\cap\mathcal{X}_{\xi}(f_\mu)$, we deduce

$$\mu\bigl(\bar{B}_h(x)\bigr)\ge\mu\bigl(\bar{B}_{h-r}(x_0)\bigr)\ge\mu\bigl(\bar{B}_{h/2}(x_0)\bigr)\ge\upsilon\cdot\Bigl(\frac{h}{2}\Bigr)^d\cdot\xi.$$

For $\beta\in(0,1]$, we have $\Phi^{\beta}_{x,h}(\cdot)\equiv1$, so (S26) is immediate. Suppose now that $3r\le\upsilon h$, so that $B_{(1+\upsilon)(h-r)}(x_0)\supseteq\bar{B}_{h+r}(x_0)\supseteq\bar{B}_h(x)$. Thus, since $x_0\in\mathcal{R}_{\upsilon}(\mu)$, we infer that with $M_{x,h}:=\sup_{x'\in\bar{B}_h(x)}f_\mu(x')$,

$$\mu\bigl(\bar{B}_h(x)\bigr)\ge\mu\bigl(B_{h-r}(x_0)\bigr)\ge\upsilon(h-r)^d\cdot\sup_{x'\in B_{(1+\upsilon)(h-r)}(x_0)}f_\mu(x')\ge\upsilon\cdot\Bigl(\frac{h}{2}\Bigr)^d\cdot M_{x,h}. \tag{S27}$$

Moreover, if we take $J_{x,h} := \{x' \in \bar{B}_h(x) : f_\mu(x') \geq 2^{-(2d+1)} \cdot \upsilon \cdot M_{x,h}\}$, then

$$\mu(\bar{B}_h(x)) \leq \mathcal{L}_d(J_{x,h}) \cdot M_{x,h} + \mathcal{L}_d(\bar{B}_h(x) \setminus J_{x,h}) \cdot 2^{-(2d+1)} \cdot \upsilon \cdot M_{x,h}$$
$$\leq \mathcal{L}_d(J_{x,h}) \cdot M_{x,h} + \frac{\upsilon}{2} \cdot \left(\frac{h}{2}\right)^d \cdot M_{x,h},$$

so by (S27) we have $\mathcal{L}_d(J_{x,h}) \geq 2^{-(d+1)} \cdot \upsilon \cdot h^d$. Taking $K_{x,h} := h^{-1} \cdot (J_{x,h} - x) \subseteq \bar{B}_1(0)$, we have $\mathcal{L}_d(K_{x,h}) \geq 2^{-(d+1)} \cdot \upsilon$. Given any $w \in \mathbb{R}^{\mathcal{V}(\beta)}$ with $\|w\|_2 = 1$, it follows from Lemma S18 that

$$\int_{\bar{B}_h(x)} \langle w, \Phi^\beta_{x,h}(z)\rangle^2 \, d\mu(z) \geq 2^{-(2d+1)} \cdot \upsilon \cdot M_{x,h} \cdot \int_{J_{x,h}} \langle w, \Phi^\beta_{x,h}(z)\rangle^2 \, d\mathcal{L}_d(z)$$
$$\geq 2^{-(2d+1)} \cdot \upsilon \cdot M_{x,h} \cdot h^d \cdot \int_{K_{x,h}} \langle w, \Phi^\beta_{0,1}(z')\rangle^2 \, d\mathcal{L}_d(z')$$
$$\geq 2^{-(3d+1)} \cdot \upsilon \cdot c^0_{\mathrm{min}} \cdot \mu(\bar{B}_h(x)).$$

The result follows. □

LEMMA S20. *Suppose that* $\upsilon \in (0,1)$, $\xi \in (0,\infty)$, $\beta \in (0,\infty)$, $x \in \mathbb{R}^d$, $r \in (0,1/2]$ *and* $h \in [2r,1]$ *satisfy* $\bar{B}_r(x) \cap \mathcal{R}_\upsilon(\mu) \cap \mathcal{X}_\xi(f_\mu) \neq \emptyset$, *and choose* $\delta \in (0,1)$. *Suppose also that*

$$h \geq \left\{ \frac{2^{4(d+1)} \cdot |\mathcal{V}(\beta)|}{c^0_{\mathrm{min}} \cdot \upsilon^2 \cdot \xi \cdot n} \cdot \log\left(\frac{2|\mathcal{V}(\beta)|}{\delta}\right) \right\}^{1/d}, \tag{S28}$$

and that either $\beta \in (0,1]$ *or* $3r \leq \upsilon h$. *Then*

$$\mathbb{P}\Big(\big\{|\mathcal{N}_{x,h}| \leq 2n \cdot \mu(\bar{B}_h(x))\big\} \cap \big\{\lambda_{\mathrm{min}}(Q^\beta_{x,h}) \geq 2^{-(3d+2)} \cdot n \cdot c^0_{\mathrm{min}} \cdot \upsilon \cdot \mu(\bar{B}_h(x))\big\}\Big)$$
$$\geq 1 - \delta.$$

PROOF. By Lemma S19 and (S28), we have $\mu(\bar{B}_h(x)) \geq \upsilon \cdot \xi \cdot (h/2)^d \geq (8/3)\log(2/\delta)/n$. Hence, by the multiplicative Chernoff bound (Lemma S39),

$$\mathbb{P}\{|\mathcal{N}_{x,h}| > 2n \cdot \mu(\bar{B}_h(x))\} \leq \frac{\delta}{2}. \tag{S29}$$

In addition, if either $\beta \in (0,1]$ or $3r \leq \upsilon h$, then by Lemma S19 again,

$$\lambda_{\mathrm{min}}\left(\int_{\bar{B}_h(x)} \Phi^\beta_{x,h}(z)\Phi^\beta_{x,h}(z)^\top \, d\mu(z)\right) \geq 2^{-(3d+1)} \cdot \upsilon \cdot c^0_{\mathrm{min}} \cdot \mu(\bar{B}_h(x))$$
$$\geq 2^{-(4d+1)} \cdot \upsilon^2 \cdot c^0_{\mathrm{min}} \cdot \xi \cdot h^d$$
$$\geq \frac{8|\mathcal{V}(\beta)|}{n} \cdot \log\left(\frac{2|\mathcal{V}(\beta)|}{\delta}\right).$$

Note also that $\lambda_{\mathrm{max}}\big(\Phi^\beta_{x,h}(X_1)\Phi^\beta_{x,h}(X_1)^\top \cdot \mathbb{1}_{\{X_1 \in \bar{B}_h(x)\}}\big) \leq |\mathcal{V}(\beta)|$. Hence, by a matrix multiplicative Chernoff bound (Lemma S40) applied with $m = n$, $\mathbf{Z}_i = \Phi^\beta_{x,h}(X_i)\Phi^\beta_{x,h}(X_i)^\top \cdot \mathbb{1}_{\{X_i \in \bar{B}_h(x)\}}$ and $q = |\mathcal{V}(\beta)|$, we have

$$\mathbb{P}\{\lambda_{\mathrm{min}}(Q^\beta_{x,h}) < 2^{-(3d+2)} \cdot n \cdot c^0_{\mathrm{min}} \cdot \upsilon \cdot \mu(\bar{B}_h(x))\}$$
$$\leq \mathbb{P}\left\{\lambda_{\mathrm{min}}(Q^\beta_{x,h}) < \frac{n}{2} \cdot \lambda_{\mathrm{min}}\left(\int_{\bar{B}_h(x)} \Phi^\beta_{x,h}(z)\Phi^\beta_{x,h}(z)^\top d\mu(z)\right)\right\} \leq \frac{\delta}{2}. \tag{S30}$$

The result now follows by combining (S29) and (S30) with a union bound. □

LEMMA S21. *Suppose that* $\alpha \in (0,1)$, $\beta > 0$, $\lambda \geq 1$, $\kappa, \gamma > 0$, $\upsilon \in (0,1)$, $C_{\mathrm{App}} \geq 1$, *take* $P \in \mathcal{P}_{\mathrm{Höl}}(\beta,\lambda) \cap \mathcal{P}^+_{\mathrm{App}}(\mathcal{A},\tau,\kappa,\gamma,\upsilon,\tau,C_{\mathrm{App}})$ *and let* $C_{\mathrm{pv}} := 2^{2d+5} \cdot \bigl(\upsilon \cdot \sqrt{c^0_{\min}}\bigr)^{-1}$. *Suppose further that* $\xi, \Delta \in (0,\infty)$, $x \in \mathbb{R}^d$ *and* $r \in (0,1/2]$ *satisfy* $\bar{B}_r(x) \cap \mathcal{R}_\upsilon(\mu) \cap \mathcal{X}_\xi(f_\mu) \cap \mathcal{X}_{\tau+\Delta}(\eta) \neq \emptyset$, *where* μ *is the marginal distribution of* P *on* $\mathbb{R}^d$, *and* $\eta : \mathbb{R}^d \to [0,1]$ *is the regression function. Given any* $\delta \in (0,1)$ *with*

$$r \geq \biggl\{\frac{C^2_{\mathrm{pv}} \cdot |\mathcal{V}(\beta)|}{\xi \cdot n} \cdot \log\biggl(\frac{4|\mathcal{V}(\beta)|}{\delta}\biggr)\biggr\}^{\frac{\beta}{d(\beta\wedge 1)}}, \quad \Delta \geq C_{\mathrm{pv}}\Biggl(\lambda r^{\beta\wedge 1} + \sqrt{\frac{\log\bigl(2/(\alpha\wedge\delta)\bigr)}{\xi \cdot n \cdot r^{d(\beta\wedge 1)/\beta}}}\Biggr),$$

and either $\beta \in (0,1]$ *or* $r \leq \{2(\upsilon/3)^\beta\}^{\frac{1}{\beta-1}}$, *we have* $\mathbb{P}\bigl\{\hat{p}^+_n\bigl(\bar{B}_r(x)\bigr) \leq \alpha\bigr\} \geq 1-\delta$.

PROOF. First recall that in the construction of our p-values $\hat{p}^+_n(\cdot)$ in (19) we take $h = (2r)^{1\wedge\frac{1}{\beta}}$. To prove the lemma we define events

$$\mathcal{E}^\eta_\delta := \Biggl\{\hat{\eta}(x) > \eta(x) - \sqrt{e_0^\top\bigl(Q^\beta_{x,h}\bigr)^+ e_0} \cdot \Biggl(\lambda \cdot h^\beta \cdot |\mathcal{N}_{x,h}|^{1/2} + \sqrt{\frac{\log(2/\delta)}{2}}\Biggr)\Biggr\},$$

$$\mathcal{E}^\mu_\delta := \Biggl\{\lambda_{\min}\bigl(Q^\beta_{x,h}\bigr) \geq 2^{-(3d+2)} \cdot c^0_{\min} \cdot \upsilon^2 \cdot \max\biggl\{\xi \cdot n \cdot \biggl(\frac{h}{2}\biggr)^d, \frac{|\mathcal{N}_{x,h}|}{2\upsilon}\biggr\}\Biggr\}.$$

Note that since $P \in \mathcal{P}_{\mathrm{Höl}}(\beta,\lambda)$ and $\bar{B}_r(x) \cap \mathcal{X}_{\tau+\Delta}(\eta) \neq \emptyset$, we have $\eta(x) \geq \tau + \Delta - \lambda \cdot r^{\beta\wedge 1}$. Hence, on the event $\mathcal{E}^\eta_\delta \cap \mathcal{E}^\mu_\delta$ we have

$$\begin{aligned}
&\hat{\eta}(x) - \tau - \lambda\Bigl(1 + 2\sqrt{e_0^\top\bigl(Q^\beta_{x,h}\bigr)^{-1}e_0 \cdot |\mathcal{N}_{x,h}|}\Bigr)r^{\beta\wedge 1} \\
&\quad > \eta(x) - \tau - \lambda\Bigl(1 + 4\sqrt{e_0^\top\bigl(Q^\beta_{x,h}\bigr)^{-1}e_0 \cdot |\mathcal{N}_{x,h}|}\Bigr)r^{\beta\wedge 1} - \sqrt{\frac{1}{2} \cdot e_0^\top\bigl(Q^\beta_{x,h}\bigr)^{-1}e_0 \cdot \log(2/\delta)} \\
&\quad > \Delta - 2\lambda\Bigl(1 + 2\sqrt{e_0^\top\bigl(Q^\beta_{x,h}\bigr)^{-1}e_0 \cdot |\mathcal{N}_{x,h}|}\Bigr)r^{\beta\wedge 1} - \sqrt{\frac{1}{2} \cdot e_0^\top\bigl(Q^\beta_{x,h}\bigr)^{-1}e_0 \cdot \log(2/\delta)} \\
&\quad \geq \Delta - 2\lambda\biggl(1 + \sqrt{\frac{2^{3d+5}}{c^0_{\min} \cdot \upsilon}}\biggr)r^{\beta\wedge 1} - \sqrt{\frac{2^{4d+1} \cdot \log(2/\delta)}{c^0_{\min} \cdot \upsilon^2 \cdot \xi \cdot n \cdot r^{d(\beta\wedge 1)/\beta}}} \\
&\quad \geq \sqrt{\frac{2^{4d+1} \cdot \log(2/\alpha)}{c^0_{\min} \cdot \upsilon^2 \cdot \xi \cdot n \cdot r^{d(\beta\wedge 1)/\beta}}} \geq \sqrt{\frac{1}{2} \cdot e_0^\top\bigl(Q^\beta_{x,h}\bigr)^{-1}e_0 \cdot \log(1/\alpha)}.
\end{aligned}$$

Hence, on the event $\mathcal{E}^\eta_\delta \cap \mathcal{E}^\mu_\delta$ we have $\hat{p}^+_n\bigl(\bar{B}_r(x)\bigr) \leq \alpha$. Now by Lemma S16 we have $\mathbb{P}\bigl((\mathcal{E}^\eta_\delta)^{\mathrm{c}} \cap E^\mu_\delta\bigr) \leq \delta/2$. Moreover, by Lemma S19 we have $\mu\bigl(\bar{B}_h(x)\bigr) \geq \upsilon \cdot \xi \cdot (h/2)^d$. Hence, by Lemma S20 we have $\mathbb{P}\bigl((\mathcal{E}^\mu_\delta)^{\mathrm{c}}\bigr) \leq \delta/2$. Thus $\mathbb{P}\bigl((\mathcal{E}^\eta_\delta)^{\mathrm{c}} \cup (E^\mu_\delta)^{\mathrm{c}}\bigr) \leq \delta$, and the conclusion follows. □

Given any $\upsilon \in (0,1)$, ξ, $\Delta \in (0,\infty)$, $r \in (0,1/2]$, we let

$$\mathcal{H}'_\upsilon(\xi,\Delta,r) := \bigl\{B \in \mathcal{H}^+ : \mathrm{diam}_\infty(B) = 2r \text{ and } B \cap \mathcal{R}_\upsilon(\mu) \cap \mathcal{X}_\xi(f_\mu) \cap \mathcal{X}_{\tau+\Delta}(\eta) \neq \emptyset\bigr\}.$$

LEMMA S22. *We have* $|\mathcal{H}'_\upsilon(\xi,\Delta,r)| \leq (2/r)^d \cdot (\upsilon \cdot \xi)^{-1}$ *for every* $\upsilon \in (0,1)$, ξ, $\Delta \in (0,\infty)$ *and* $r \in (0,1/2]$.

PROOF. Given $B = 2r\prod_{j\in[d]}[a_j, a_j+1] \in \mathcal{H}'_\upsilon(\xi,\Delta,r)$, for some $(a_j)_{j\in[d]} \in \mathbb{Z}^d$, we write $\phi(B) := (a_j \mod 2)_{j\in[d]} \in \{0,1\}^d$ and $\psi(B) := r\prod_{j\in[d]}[2a_j-1, 2a_j+3]$. Note that if

$\phi(B_0) = \phi(B_1)$ for distinct $B_0, B_1 \in \mathcal{H}'_\upsilon(\xi,\Delta,r)$ then $\mu\bigl(\psi(B_0)\cap\psi(B_1)\bigr) = 0$, since μ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^d$. Moreover, by Lemma S19 we have $\mu\bigl(\psi(B)\bigr) \geq \upsilon\cdot\xi\cdot r^d$, so

$$|\mathcal{H}'_\upsilon(\xi,\Delta,r)|\cdot\upsilon\xi r^d \leq \sum_{B\in\mathcal{H}'_\upsilon(\xi,\Delta,r)} \mu\bigl(\psi(B)\bigr) = \sum_{z\in\{0,1\}^d}\ \sum_{B\in\mathcal{H}'_\upsilon(\xi,\Delta,r)\cap\phi^{-1}\{z\}} \mu\bigl(\psi(B)\bigr) \leq 2^d,$$

as required. □

PROOF OF PROPOSITION 13. Let

$$\rho := \kappa(2\beta+d)+\beta\gamma, \quad \theta := \frac{\lambda^{d/\beta}}{n}\log_+\biggl(\frac{8n^{\frac{\beta}{\beta\wedge1}}|\mathcal{V}(\beta)|\log_2 n}{\alpha\wedge\delta}\biggr),$$

$$\xi := \theta^{\beta\gamma/\rho}, \quad r_* := 2^{-\bigl\lfloor\frac{1}{\beta\wedge1}\{\frac{\beta\kappa}{\rho}\log_2(1/\theta)+\log_2\lambda\}\bigr\rfloor},$$

and, recalling the definition of $C_{\mathrm{pv}} := 2^{2d+5}\cdot\bigl(\upsilon\cdot\sqrt{c^0_{\min}}\bigr)^{-1}$ from Lemma S21, define

$$A_0 := \begin{cases} 2\vee C_{\mathrm{pv}}|\mathcal{V}(\beta)|^{1/2}\vee 2^{\frac{\beta-2}{\beta-1}}(3/\upsilon)^{\frac{\beta}{\beta-1}} & \text{if } \beta>1 \\ 2^\beta\vee C_{\mathrm{pv}}|\mathcal{V}(\beta)|^{1/2} & \text{if } \beta\leq 1. \end{cases}$$

Then for $\theta^\beta \leq A_0^{-\frac{\rho}{\kappa}}$ we have

$$\biggl(\frac{C^2_{\mathrm{pv}}\cdot|\mathcal{V}(\beta)|}{\xi}\cdot\theta\lambda^{-d/\beta}\biggr)^{\frac{\beta}{d(\beta\wedge1)}} \leq r_* \leq \frac{1}{2},$$

and $r_* \leq \{2(\upsilon/3)^\beta\}^{\frac{1}{\beta-1}}$ if $\beta>1$. Now let

$$\Delta := C_{\mathrm{pv}}\Biggl(\lambda r_*^{\beta\wedge1} + \sqrt{\frac{\log\bigl(2/(\alpha\wedge\delta)\bigr)}{\xi\cdot n\cdot r_*^{d(\beta\wedge1)/\beta}}}\Biggr),$$

so that $\Delta \leq 3\cdot C_{\mathrm{pv}}\theta^{\frac{\beta\kappa}{\rho}}$ when $\theta\leq A_0^{-\frac{\rho}{\beta\kappa}}$. By Lemma S22, we have $|\mathcal{H}'_\upsilon(\xi,\Delta,r_*)| \leq (2/r_*)^d\cdot(\upsilon\cdot\xi)^{-1} \leq (\lambda^d\theta^{-\beta})^{\frac{1}{\beta\wedge1}} \leq n^{\frac{\beta}{\beta\wedge1}}$ when $\theta\leq A_0^{-\frac{\rho}{\beta\kappa}}$. Hence we may apply a union bound and Lemma S21 with $\delta/(2n^{\frac{\beta}{\beta\wedge1}})$ in place of δ and $\alpha/(n\log_2 n)$ in place of α to deduce that whenever $\theta\leq A_0^{-\frac{\rho}{\beta\kappa}}$, we have

$$\mathbb{P}\biggl(\bigcup_{B\in\mathcal{H}'_\upsilon(\xi,\Delta,r_*)}\Bigl\{\hat{p}^+_n(B) > \frac{\alpha}{|\mathcal{H}^+(\mathcal{D}_X)|}\Bigr\}\biggr) \leq \sum_{B\in\mathcal{H}'_\upsilon(\xi,\Delta,r_*)}\mathbb{P}\Bigl(\hat{p}^+_n(B) > \frac{\alpha}{n\log_2 n}\Bigr) \leq \frac{\delta}{2}.$$

Hence, whenever $\theta\leq A_0^{-\frac{\rho}{\beta\kappa}}$, we have

$$\mathcal{R}_\upsilon(\mu)\cap\mathcal{X}_\xi(f_\mu)\cap\mathcal{X}_{\tau+\Delta}(\eta) \subseteq \bigcup_{B\in\mathcal{H}'_\upsilon(\xi,\Delta,r_*)} B \subseteq \bigcup_{\ell\in[\ell_\alpha]} B_{(\ell)},$$

with probability at least $1-\delta/2$. Thus, with probability at least $1-\delta/2$,

$$\begin{aligned} &M_\tau - \sup\biggl\{\mu(A) : A\in\mathcal{A}\cap\mathrm{Pow}\biggl(\bigcup_{\ell\in[\ell_\alpha]}B_{(\ell)}\biggr)\biggr\} \\ &\quad\leq M_\tau - \sup\Bigl\{\mu(A) : A\in\mathcal{A}\cap\mathrm{Pow}\bigl(\mathcal{R}_\upsilon(\mu)\cap\mathcal{X}_\xi(f_\mu)\cap\mathcal{X}_{\tau+\Delta}(\eta)\bigr)\Bigr\} \\ &\quad\leq C_{\mathrm{App}}\cdot(\xi^\kappa+\Delta^\gamma)\mathbb{1}_{\{\theta\leq A_0^{-\frac{\rho}{\beta\kappa}}\}} + \mathbb{1}_{\{\theta> A_0^{-\frac{\rho}{\beta\kappa}}\}} \\ &\quad\leq C_{\mathrm{App}}\cdot\bigl\{1+(3C_{\mathrm{pv}})^\gamma + A_0^\gamma\bigr\}\cdot\theta^{\beta\kappa\gamma/\rho}. \end{aligned} \tag{S31}$$

Finally, since $\hat{A}^+_{\mathrm{OSS}}$ is chosen from $\mathcal{A} \cap \mathrm{Pow}\big(\bigcup_{\ell \in [\ell_\alpha]} B_{(\ell)}\big)$ with maximal empirical measure, it follows from Lemma S36 as in the proof of Proposition S2 that with probability at least $1-\delta$,

$$M_\tau - \mu(\hat{A}^+_{\mathrm{OSS}}) \leq C_{\mathrm{App}}\big\{1 + (3C_{\mathrm{pv}})^\gamma + A_0^\gamma\big\} \cdot \theta^{\beta\kappa\gamma/\rho} + 2C_{\mathrm{VC}}\sqrt{\frac{\dim_{\mathrm{VC}}(\mathcal{A})}{n}} + \sqrt{\frac{2\log(2/\delta)}{n}}.$$

The second part of Proposition 13 follows integrating the tail bound and applying Proposition 12, as at the end of the proof of Theorem 5. □

S6. Proofs of the lower bounds in Theorems 2 and 11. Recall the construction of the probability distributions $\{P^\ell_{L,r,w,s,\theta} : \ell \in [L]\}$ on $\mathbb{R}^d \times \{0,1\}$ from Section 5, with corresponding regression functions $\{\eta^\ell_{L,r,w,s,\theta} : \ell \in [L]\}$ and common marginal distribution $\mu_{L,r,w}$ on $\mathbb{R}^d$. Recall also the definition of $\mathcal{R}_\upsilon(\cdot)$ from (15). Our initial goal is to prove that $\{P^\ell_{L,r,w,s,\theta} : \ell \in [L]\}$ is a subset of $\mathcal{P}_{\mathrm{H\ddot{o}l}}(\beta,\lambda)$ (see Lemma S28) and $\mathcal{P}_{\mathrm{App}}(\mathcal{A},\kappa,\gamma,\tau,C_{\mathrm{App}}) \cap \mathcal{P}^+_{\mathrm{App}}(\mathcal{A},\tau,\kappa,\gamma,\upsilon,\tau,C_{\mathrm{App}})$ (see Lemma S30) for suitable L, r, w, s and θ. The first of these lemmas will rely on several auxiliary results,

Given two multi-indices $\nu = (\nu_1,\ldots,\nu_d)^\top, \nu' = (\nu'_1,\ldots,\nu'_d)^\top \in \mathbb{N}_0^d$, we write $\nu \prec \nu'$ if either $\|\nu\|_1 < \|\nu'\|_1$ or both $\|\nu\|_1 = \|\nu'\|_1$ and there exists $j \in \{0,1,\ldots,d-1\}$ such that $\nu_1 = \nu'_1,\ldots,\nu_j = \nu'_j$ and $\nu_{j+1} < \nu'_{j+1}$. Now, given $m \in \mathbb{N}$ and $j \in [m]$, we write

$$\mathcal{Q}_j(\nu,m) := \biggl\{(k_1,\ldots,k_j,\ell_1,\ldots,\ell_j) \in \mathbb{N}^j \times (\mathbb{N}_0^d)^j : 0 \prec \ell_1 \prec \ldots \prec \ell_j, \sum_{q=1}^j k_q = m, \sum_{q=1}^j k_q \ell_q = \nu\biggr\}.$$

In addition, for multi-indices $\nu = (\nu_1,\ldots,\nu_d)^\top \in \mathbb{N}_0^d$, we let $\nu! := \prod_{m=1}^{\|\nu\|_1} \nu_j!$. The following lemma is a version of the Faà di Bruno formula.

LEMMA S23 (Corollary 2.10 of Constantine and Savits (1996)). *Let $x \in \mathbb{R}^d$ and $\nu = (\nu_1,\ldots,\nu_d)^\top \in \mathbb{N}_0^d$. Suppose that all partial derivative of order $\|\nu\|_1$ of $f : \mathbb{R}^d \to \mathbb{R}$ exist and are continuous in a neighbourhood of x, and that $g : \mathbb{R} \to \mathbb{R}$ is $\|\nu\|_1$-times continuously differentiable in a neighbourhood of $f(x)$. Then*

$$\partial_x^\nu(g \circ f) = \nu! \cdot \sum_{m=1}^{\|\nu\|_1} g^{(m)}\big(f(x)\big) \sum_{j=1}^{\|\nu\|_1} \sum_{\substack{(k_1,\ldots,k_j,\ell_1,\ldots,\ell_j) \\ \in \mathcal{Q}_j(\nu,m)}} \prod_{q=1}^j \frac{\{\partial_x^{\ell_q}(f)\}^{k_q}}{k_q! \cdot (\ell_q!)^{k_q}}.$$

LEMMA S24. *Given $\nu = (\nu_1,\ldots,\nu_d)^\top \in \mathbb{N}_0^d$, we have $\sup_{\|x\|_2 \geq 1} |\partial_x^\nu(\|\cdot\|_2)| < \infty$.*

PROOF. For $t \in [d]$, write $e_t = (0,\ldots,0,1,0,\ldots,0)^\top \in \mathbb{R}^d$ for the t^{th} standard basis vector in $\mathbb{R}^d$. By Lemma S23 with $f(x) = \|x\|_2^2$ and $g(z) = \sqrt{z}$, we have for $x = (x_1,\ldots,x_d)^\top \in \mathbb{R}^d$ that

$$\partial_x^\nu(\|\cdot\|_2) = \nu! \sum_{m=1}^{\|\nu\|_1} \frac{(-1)^{m+1}(2m-3)!!}{2^m\|x\|_2^{2m-1}} \sum_{j=1}^{\|\nu\|_1} \sum_{\substack{(k_1,\ldots,k_j,\ell_1,\ldots,\ell_j) \\ \in \mathcal{Q}_j(\nu,m)}} \prod_{q=1}^j \frac{2^{k_q}\big\{\sum_{t=1}^d (x_t \mathbb{1}_{\{\ell_q=e_t\}} + \mathbb{1}_{\{\ell_q=2e_t\}})\big\}^{k_q}}{k_q! \cdot (\ell_q!)^{k_q}}.$$

It follows that for all $x \in \mathbb{R}^d$ with $\|x\|_2 \geq 1$ we have

$$|\partial_x^\nu(\|\cdot\|_2)| \leq \nu! \sum_{m=1}^{\|\nu\|_1} \frac{(2m-3)!!}{\|x\|_2^{2m-1}} \sum_{j=1}^{\|\nu\|_1} \sum_{\substack{(k_1,\ldots,k_j,\ell_1,\ldots,\ell_j) \\ \in \mathcal{Q}_j(\nu,m)}} \prod_{q=1}^{j} \frac{\|x\|_2^{k_q}}{k_q! \cdot (\ell_q!)^{k_q}}$$

$$\leq \nu! \sum_{m=1}^{\|\nu\|_1} (2m-3)!! \cdot \sum_{j=1}^{\|\nu\|_1} \sum_{\substack{(k_1,\ldots,k_j,\ell_1,\ldots,\ell_j) \\ \in \mathcal{Q}_j(\nu,m)}} \prod_{q=1}^{j} \frac{1}{k_q! \cdot (\ell_q!)^{k_q}} < \infty,$$

as required. □

LEMMA S25. *For each $m, d \in \mathbb{N}$, there exists $C_{m,d} > 0$, depending only on m and d, such that for any infinitely differentiable function $g : [0,\infty) \to [0,\infty)$ with $g'(z) = 0$ for all $z \in [0,1]$, and any $\nu = (\nu_1,\ldots,\nu_d)^\top \in \mathbb{N}_0^d$ with $\|\nu\|_1 = m$, we have*

$$\left|\partial_x^\nu(g \circ \|\cdot\|_2)\right| \leq C_{m,d} \cdot \max_{k\in[m]} \sup_{z\in[0,\infty)} |g^{(k)}(z)|$$

for all $x \in \mathbb{R}^d$.

PROOF. The lemma follows from combining Lemmas S23 and S24, and by considering the cases $\|x\|_2 < 1$ and $\|x\|_2 \geq 1$ separately. □

LEMMA S26. *Let $L, d \in \mathbb{N}$, $r \in (0,\infty)$, $s \in (0, 1 \wedge (r/2)]$, $w \in \big(0, (2r)^{-d} \wedge 1\big)$ and $\theta \in (0, \epsilon_0/2]$. Then for $\ell \in [L]$, $\nu = (\nu_1,\ldots,\nu_d)^\top \in \mathbb{N}_0^d$ with $\|\nu\|_1 = m$ and $x \in \mathbb{R}^d$, we have*

$$\left|\partial_x^\nu(\eta^\ell_{L,r,w,s,\theta})\right| \leq 2A_m C_{m,d} \cdot \frac{\theta}{s^m}, \tag{S32}$$

where A_m is taken from (21) and $C_{m,d}$ is taken from Lemma S25. Hence, given any $\xi \in [0,1]$ and x, $x' \in \mathbb{R}^d$, we have

$$\left|\partial_x^\nu(\eta^\ell_{L,r,w,s,\theta}) - \partial_{x'}^\nu(\eta^\ell_{L,r,w,s,\theta})\right| \leq 2A_{m+1}(2C_{m,d} \vee dC_{m+1,d}) \cdot \frac{\theta}{s^{m+\xi}} \cdot \|x-x'\|_\infty^\xi. \tag{S33}$$

PROOF. To prove (S32), we construct an open cover of $\mathbb{R}^d$ by $\{U_1,\ldots,U_{L+1}\}$ where $U_{\ell'} := B_{r_\sharp(w)}(z_{\ell'})$ for $\ell' \in [L]$ and $U_{L+1} := \mathbb{R}^d \setminus \bigcup_{\ell'\in[L]} \bar{B}_{2d^{1/2}r}(z_{\ell'})$. First suppose that $\ell' \in [L]\setminus\{\ell\}$ and consider the function $g_0 : [0,\infty) \to [0,\infty)$ defined by

$$g_0(t) := \begin{cases} \tau - \theta & \text{if } t \leq 1 \\ \tau + \theta - 2\theta \cdot h\big(t-1\big) & \text{if } 1 < t \leq 2 \\ \tau + \theta & \text{if } 2 < t \leq \frac{d^{1/2}r}{s} \\ \tau - \theta + 2\theta \cdot h\big(\frac{s\cdot t}{d^{1/2}r} - 1\big) & \text{if } \frac{d^{1/2}r}{s} < t < \frac{2d^{1/2}r}{s} \\ \tau - \theta & \text{otherwise.} \end{cases}$$

By Lemma S25, together with $s \leq r/2 \leq d^{1/2}r$, we have

$$\sup_{x\in\mathbb{R}^d} \left|\partial_x^\nu(g_0 \circ \|\cdot\|_2)\right| \leq C_{m,d} \cdot \max_{k\in[m]} \sup_{z\in[0,\infty)} |g_0^{(k)}(z)| \leq 2A_m C_{m,d}\theta.$$

Moreover, for all $x \in U_{\ell'}$ we have $\eta^\ell_{L,r,w,s,\theta}(x) = g_0(\|s^{-1}\cdot(x - z_{\ell'})\|_2)$. Hence, for all $x \in U_{\ell'}$ we have $\left|\partial_x^\nu(\eta^\ell_{L,r,w,s,\theta})\right| \leq 2A_m C_{m,d}\theta s^{-m}$ since $\|\nu\|_1 = m$. Next, consider the open set U_ℓ

and define a function $g_1 : [0,\infty) \to [0,\infty)$ by

$$g_1(t) := \begin{cases} \tau + \theta & \text{if } t \leq 1 \\ \tau - \theta + 2\theta \cdot h\big(t-1\big) & \text{if } 1 < t < 2 \\ \tau - \theta & \text{otherwise.} \end{cases}$$

By applying Lemma S25 again, we have $\big|\partial_x^\nu(g_1 \circ \|\cdot\|_2)\big| \leq 2A_m C_{m,d}\theta$ for all $x \in \mathbb{R}^d$. Moreover, for $x \in U_\ell$ we have $\eta^\ell_{L,r,w,s,\theta}(x) = g_1(\|(d^{1/2}r)^{-1} \cdot (x - z_\ell)\|_2)$. Hence, for all $x \in U_\ell$, we have $\big|\partial_x^\nu(\eta^\ell_{L,r,w,s,\theta})\big| \leq 2A_m C_{m,d}\theta(d^{1/2}r)^{-m} \leq 2A_m C_{m,d}\theta s^{-m}$. Finally we note that $\eta^\ell_{L,r,w,s,\theta}\big|_{U_{L+1}} \equiv \tau - \theta$, so $\sup_{x \in U_{L+1}}\big|\partial_x^\nu(\eta^\ell_{L,r,w,s,\theta})\big| = 0 \leq 2A_m C_{m,d}\theta s^{-m}$. The claim (S32) follows.

To prove (S33), we first consider the case where $\|x - x'\|_\infty \leq s$, in which case, we may apply the mean value theorem combined with (S32) and Hölder's inequality to obtain

$$\begin{aligned} \big|\partial_x^\nu(\eta_{L,r,w,s,\theta}) - \partial_{x'}^\nu(\eta_{L,r,w,s,\theta})\big| &\leq dA_{m+1}C_{m+1,d} \cdot \frac{\theta}{s^{m+1}} \cdot \|x - x'\|_\infty \\ &\leq dA_{m+1}C_{m+1,d} \cdot \frac{\theta}{s^{m+\xi}} \cdot \|x - x'\|_\infty^\xi. \end{aligned}$$

Moreover, when $\|x - x'\|_\infty > s$, (S33) follows immediately from (S32) and the triangle inequality. □

LEMMA S27. *Take $\beta > 0$, $C_f > 0$ and let $f : \mathbb{R}^d \to \mathbb{R}$ be a $\lceil\beta\rceil$-times differentiable function such that for all $\nu = (\nu_1, \ldots, \nu_d)^\top \in \mathbb{N}_0^d$ with $\|\nu\|_1 = \lceil\beta\rceil - 1 =: m$, and x, $x' \in \mathbb{R}^d$, we have*

$$\big|\partial_{x'}^\nu(f) - \partial_x^\nu(f)\big| \leq C_f \cdot \|x' - x\|_\infty^{\beta - m}.$$

Then for all $x, x' \in \mathbb{R}^d$ we have

$$\big|f(x') - \mathcal{T}_x^\beta[f](x')\big| \leq C_f \cdot \binom{m+d-1}{d-1} \cdot \|x' - x\|_\infty^\beta.$$

PROOF. By Taylor's theorem, there exists $t \in (0,1)$ such that

$$f(x') = \sum_{\nu \in \mathbb{N}_0^d : \|\nu\|_1 < m} \frac{(x'-x)^\nu}{\nu!} \cdot \partial_x^\nu(f) + \sum_{\nu \in \mathbb{N}_0^d : \|\nu\|_1 = m} \frac{(x'-x)^\nu}{\nu!} \cdot \partial_{x + t\cdot(x'-x)}^\nu(f).$$

Hence,

$$\begin{aligned} \big|f(x') - \mathcal{T}_x^\beta[f](x')\big| &= \left|\sum_{\nu \in \mathbb{N}_0^d : \|\nu\|_1 = m} \frac{(x'-x)^\nu}{\nu!} \cdot \big(\partial_{x+t\cdot(x'-x)}^\nu(f) - \partial_x^\nu(f)\big)\right| \\ &\leq C_f \cdot \binom{m+d-1}{d-1} \cdot \|x' - x\|_\infty^\beta, \end{aligned}$$

as required. □

LEMMA S28. *Let $\beta > 0$, $\lambda > 1$, $L, d \in \mathbb{N}$, $r \in (0,\infty)$, $s \in (0, 1 \wedge (r/2)]$, $w \in \big(0, (2r)^{-d} \wedge 1\big)$ and $\theta \in (0, \epsilon_0/2]$. There exists $c^\flat_{\beta,d} > 0$, depending only on β and d, such that whenever $\theta \leq c^\flat_{\beta,d} \cdot \lambda \cdot s^\beta$, we have that for each $\ell \in [L]$, the function $\eta^\ell_{L,r,w,s,\theta}$ is (β,λ)-Hölder on $\mathbb{R}^d$; i.e. $P^\ell_{L,r,w,s,\theta} \in \mathcal{P}_{\text{Höl}}(\beta,\lambda)$.*

PROOF. By taking

$$c^{\flat}_{\beta,d} := \min_{q\in\mathbb{N}_0:q\leq\lceil\beta\rceil-1}\left\{2A_{q+1}(2C_{q,d}\vee dC_{q+1,d})\cdot\binom{q+d-1}{d-1}\right\}^{-1},$$

the result follows from Lemmas S26 and S27. □

Lemma S30 also requires one auxiliary lemma.

LEMMA S29. *Given $L,d\in\mathbb{N}$, $r>0$ and $w\in\bigl(0,(2r)^{-d}\wedge 1\bigr)$, we have $\omega_{\mu_{L,r,w},d}(x)\geq\frac{w}{L\cdot(4d^{1/2})^d}$ for all $x\in\bigcup_{\ell\in[L]}K_r^0(\ell)$. Moreover, $\bigcup_{\ell\in[L]}K_r^0(\ell)\subseteq\mathcal{R}_\upsilon(\mu_{L,r,w})$ for every $\upsilon\leq(4d^{1/2})^{-d}$.*

PROOF. Let $\ell\in[L]$, let $x\in K_r^0(\ell)=\bar{B}_r(z_\ell)$ and let $\tilde{r}\in(0,1)$. If $\tilde{r}\in(0,2r]$, then $\bar{B}_{\tilde{r}}(x)\cap K_r^0(\ell)$ contains a hyper-cube of radius $\tilde{r}/2$, so $\mu_{L,r,w}\bigl(\bar{B}_{\tilde{r}}(x)\bigr)\geq w\cdot L^{-1}\cdot\tilde{r}^d$. Consequently, when $\tilde{r}\in\bigl(0,8d^{1/2}r\bigr]$, we have

$$\mu_{L,r,w}\bigl(\bar{B}_{\tilde{r}}(x)\bigr)\geq\mu_{L,r,w}\bigl(\bar{B}_{\tilde{r}/(4d^{1/2})}(x)\bigr)\geq\frac{w}{L\cdot(4d^{1/2})^d}\cdot\tilde{r}^d.$$

Note also that since $x\in K_r^0(\ell)\subseteq\bar{B}_{2d^{1/2}r}(z_\ell)$ there exists $\sigma_x\in\{-1,1\}^d$ and $\tilde{x}=z_\ell+\sigma_x\cdot 2d^{1/2}r\in\mathbb{R}^d$ with $\|\tilde{x}-x\|_\infty\leq 2d^{1/2}r$. Hence, if $\tilde{r}\in(4d^{1/2}r,r_\sharp(w)]$, then $\bar{B}_{\tilde{r}/4}(z_\ell+\sigma_x\cdot(2d^{1/2}r+\tilde{r}/4))\subseteq\bar{B}_{\tilde{r}}(x)\cap K_r^1(\ell)$. Thus, we have $\mu_{L,r,w}\bigl(\bar{B}_{\tilde{r}}(x)\bigr)\geq(w/L)\cdot(\tilde{r}/2)^d$ for $\tilde{r}\in(4d^{1/2}r,r_\sharp(w)]$, and consequently, $\mu_{L,r,w}\bigl(\bar{B}_{\tilde{r}}(x)\bigr)\geq(w/L)\cdot(\tilde{r}/4)^d$ for $\tilde{r}\in(8d^{1/2}r,2r_\sharp(w)]$. Finally, if $\tilde{r}\in(2r_\sharp(w),1)$, then $K_r^0(\ell)\cup K_r^1(\ell)\subseteq\bar{B}_{r_\sharp(w)}(z_\ell)\subseteq\bar{B}_{\tilde{r}}(x)$ so $\mu_{L,r,w}\bigl(\bar{B}_{\tilde{r}}(x)\bigr)\geq 1/L>(w/L)\cdot\tilde{r}^d$. The first conclusion of the lemma therefore follows. The second part then follows from the fact that the Lebesgue density of $\mu_{L,r,w}$ is at most w/L on $\mathbb{R}^d$. □

LEMMA S30. *Let $\beta>0$, $\lambda>1$, $L,d\in\mathbb{N}$, $r\in(0,\infty)$, $s\in(0,1\wedge(r/2)]$, $w\in\bigl(0,(2r)^{-d}\wedge 1\bigr)$ and $\theta\in(0,\epsilon_0/2]$, $\ell\in[L]$, $\upsilon\leq(4d^{1/2})^{-d}$ and let $\eta=\eta^\ell_{L,r,w,s,\theta}$, $\mu=\mu_{L,r,w}$ and $P=P^\ell_{L,r,w,s,\theta}$. Suppose also that $\mathcal{A}_{\mathrm{hpr}}\subseteq\mathcal{A}\subseteq\mathcal{A}_{\mathrm{conv}}$. Given any $A\in\mathcal{A}\cap\mathrm{Pow}\bigl(\mathcal{X}_\tau(\eta)\bigr)$ and $\ell'\in[L]$ with $A\cap K_r^0(\ell')\neq\emptyset$ and $z_{\ell'}\notin A$ for some $\ell'\in[L]$, we have $\mu(A)\leq(w/L)\cdot(2r)^d/2$. In particular, $\mu(A)\leq(w/L)\cdot(2r)^d/2$ whenever $A\cap K_r^0(\ell')\neq\emptyset$ for some $A\in\mathcal{A}\cap\mathrm{Pow}\bigl(\mathcal{X}_\tau(\eta)\bigr)$ and $\ell'\in[L]\setminus\{\ell\}$. Moreover, $M_\tau(P,\mathcal{A})=\mu\bigl(K_r^0(\ell)\bigr)=(w/L)\cdot(2r)^d$. Finally, if $(w/L)\cdot(2r)^d\leq C_{\mathrm{App}}\cdot\min\bigl\{(w/\{(4d^{1/2})^d\cdot L\})^\kappa,\theta^\gamma\bigr\}$, then $P\in\mathcal{P}_{\mathrm{App}}(\mathcal{A},\kappa,\gamma,\tau,C_{\mathrm{App}})\cap\mathcal{P}^+_{\mathrm{App}}(\mathcal{A},\tau,\kappa,\gamma,\upsilon,\tau,C_{\mathrm{App}})$.*

PROOF. First take $A\in\mathcal{A}\cap\mathrm{Pow}\bigl(\mathcal{X}_\tau(\eta)\bigr)$ and $\ell'\in[L]$ with $A\cap K_r^0(\ell')\neq\emptyset$ and $z_{\ell'}\notin A$. Since $\eta(x)=\tau-\theta$ for all $x\in K_r^1(\ell')$, we must have $A\cap\bigl(K_r^1(\ell')\cup\{z_{\ell'}\}\bigr)=\emptyset$. Moreover, A is convex with $A\cap K_r^0(\ell')\neq\emptyset$, and it follows that $A\subseteq\{x\in\mathbb{R}^d:\|x-z_{\ell'}\|_\infty<2d^{1/2}r\}$. Thus $A\cap\mathrm{supp}(\mu)=A\cap\{x\in\mathbb{R}^d:\|x-z_{\ell'}\|_\infty<2d^{1/2}r\}\cap\mathrm{supp}(\mu)=A\cap K_r^0(\ell')$ is the intersection of two axis-aligned hyper-rectangles, so is itself an axis-aligned hyper-rectangle. Since $A\cap\mathrm{supp}(\mu)\subseteq K_r^0(\ell')\setminus\{z_{\ell'}\}=\bar{B}_r(z_{\ell'})\setminus\{z_{\ell'}\}$, we deduce that $\mu(A)\leq(w/L)\cdot(2r)^d/2$. In particular, if $\ell'\in[L]\setminus\{\ell\}$, then $\eta(z_{\ell'})=\tau-\theta$, so $z_{\ell'}\notin\mathcal{X}_\tau(\eta)$ and the conclusion $\mu(A)\leq(w/L)\cdot(2r)^d/2$ holds.

For the next part, note that $K_r^0(\ell)=\bar{B}_r(z_\ell)\in\mathcal{A}_{\mathrm{hpr}}\cap\mathrm{Pow}\bigl(\mathcal{X}_\tau(\eta)\bigr)\subseteq\mathcal{A}\cap\mathrm{Pow}\bigl(\mathcal{X}_\tau(\eta)\bigr)$ since $\eta(x)=\tau+\theta$ for all $x\in K_r^0(\ell)$. Hence, $M_\tau(P,\mathcal{A})\geq\mu\bigl(K_r^0(\ell)\bigr)=(w/L)\cdot(2r)^d$. On the other hand, given $A\in\mathcal{A}\cap\mathrm{Pow}\bigl(\mathcal{X}_\tau(\eta)\bigr)$, we have either $A\cap\mathrm{supp}(\mu)\subseteq K_r^0(\ell)$, in which case $\mu(A)\leq\mu\bigl(K_r^0(\ell)\bigr)=(w/L)\cdot(2r)^d$, or $A\cap\mathrm{supp}(\mu)\cap K_r^0(\ell')\neq\emptyset$ for some $\ell'\in[L]\setminus$

$\{\ell\}$, since $\mathrm{supp}(\mu) \cap \mathcal{X}_\tau(\eta) \subseteq \bigcup_{\ell \in [L]} K_r^0(\ell)$, in which case $\mu(A) \leq (w/L) \cdot (2r)^d/2$. Hence $M_\tau(P, \mathcal{A}) = (w/L) \cdot (2r)^d$.

For the final part, assume that $(w/L) \cdot (2r)^d \leq C_{\mathrm{App}} \cdot \min\bigl\{\{w/(4^d d^{1/2} \cdot L)\}^\kappa, \theta^\gamma\bigr\}$, and fix $(\xi, \Delta) \in (0,\infty)^2$. We consider two cases: first suppose that $\xi \leq w/\{L \cdot (4d^{1/2})^d\}$ and $\Delta \leq \theta$. By Lemma S29, we have $K_r^0(\ell) \subseteq \mathcal{X}_\xi(\omega_{\mu,d})$. Moreover, it follows from the construction of η that $K_r^0(\ell) \subseteq \mathcal{X}_{\tau+\theta}(\eta) \subseteq \mathcal{X}_{\tau+\Delta}(\eta)$. Thus, with $A_{\xi,\Delta} = K_r^0(\ell) \in \mathcal{A} \cap \mathrm{Pow}\bigl(\mathcal{X}_\xi(\omega_{\mu,d}) \cap \mathcal{X}_{\tau+\Delta}(\eta)\bigr)$, we have

$$\mu(A_{\xi,\Delta}) = \frac{w}{L} \cdot (2r)^d \geq \frac{w}{L} \cdot (2r)^d - C_{\mathrm{App}} \cdot (\xi^\kappa + \Delta^\gamma) = M_\tau - C_{\mathrm{App}} \cdot (\xi^\kappa + \Delta^\gamma).$$

On the other hand, if $\xi > w/\{L \cdot (4d^{1/2})^d\}$ or $\Delta > \theta$, then with $A_{\xi,\Delta} = \emptyset \in \mathcal{A} \cap \mathrm{Pow}\bigl(\mathcal{X}_\xi(\omega_{\mu,d}) \cap \mathcal{X}_{\tau+\Delta}(\eta)\bigr)$, we have

$$\mu(A_{\xi,\Delta}) = 0 \geq \frac{w}{L} \cdot (2r)^d - C_{\mathrm{App}} \cdot \min\biggl\{\biggl(\frac{w}{(4d^{1/2})^d \cdot L}\biggr)^\kappa, \theta^\gamma\biggr\} \geq M_\tau - C_{\mathrm{App}} \cdot (\xi^\kappa + \Delta^\gamma).$$

We conclude that $P \in \mathcal{P}_{\mathrm{App}}(\mathcal{A}, \kappa, \gamma, \tau, C_{\mathrm{App}})$. To prove $P \in \mathcal{P}^+_{\mathrm{App}}(\mathcal{A}, \tau, \kappa, \gamma, \upsilon, \tau, C_{\mathrm{App}})$, we proceed similarly using the facts that $K_r^0(\ell) \subseteq \mathcal{R}_\upsilon(\mu)$ by Lemma S29 and that μ has Lebesgue density at least w/L on $K_r^0(\ell)$. □

Lemma S31 below bounds the χ^2-divergence between pairs of distributions in our class $\{P^\ell_{L,r,w,s,\theta} : \ell \in [L]\}$.

LEMMA S31. *Suppose that* $\epsilon_0 \in (0, 1/2)$, $\tau \in [\epsilon_0, 1-\epsilon_0]$, $L, d \in \mathbb{N}$, $r \in (0,\infty)$, $s \in (0, 1 \wedge (r/2)]$, $w \in \bigl(0, (2r)^{-d} \wedge 1\bigr)$ *and* $\theta \in (0, \epsilon_0/2]$. *Then*

$$\chi^2\bigl(P^\ell_{L,r,w,s,\theta}, P^{\ell'}_{L,r,w,s,\theta}\bigr) \leq \frac{2^{5+2d} \cdot w \cdot \theta^2 \cdot s^d}{\epsilon_0 \cdot L}$$

for all $\ell, \ell' \in [L]$.

PROOF. Let $Q_{L,r,w} := \mu_{L,r,w} \times m_{\mathrm{ct}}$ where m_{ct} denotes the counting measure on $\{0,1\}$. Note that $P^\ell_{L,r,w,s,\theta}$ is absolutely continuous with respect to $Q_{L,r,w}$, for all $\ell \in [L]$. Given $\ell \in [L]$, define $p_\ell : \mathbb{R}^d \times \{0,1\} \to \mathbb{R}$ by

$$p_\ell(x,y) := \frac{dP^\ell_{L,r,w,s,\theta}}{dQ_{L,r,w}}(x,y) = (1-y) \cdot \bigl(1 - \eta^\ell_{L,r,\theta}(x)\bigr) + y \cdot \eta^\ell_{L,r,\theta}(x).$$

Take ℓ, $\ell' \in [L]$ with $\ell \neq \ell'$ and observe that $\eta^\ell_{L,r,w,s,\theta}(x) = \eta^{\ell'}_{L,r,w,s,\theta}(x)$ for all $x \in J_{L,r,w} \setminus \bigl(\bar{B}_{2s}(z_\ell) \cup \bar{B}_{2s}(z_{\ell'})\bigr)$. Note also that $\mu_{L,r,w}\bigl(\bar{B}_{2s}(z_\ell) \cup \bar{B}_{2s}(z_{\ell'})\bigr) = (2w/L) \cdot (4s)^d$; moreover, $\eta^\ell_{L,r,\theta}(x) \in [\tau - \theta, \tau + \theta] \subseteq [\epsilon_0/2, 1 - \epsilon_0/2]$ for all $x \in J_{L,r,w}$ and $\ell \in [L]$. Hence,

$$\begin{aligned}
&\chi^2(P^\ell_{L,r,w,s,\theta}, P^{\ell'}_{L,r,w,s,\theta}) \\
&\quad = \int_{\mathbb{R}^d \times \{0,1\}} \biggl(\frac{dP^\ell_{L,r,w,s,\theta}}{dP^{\ell'}_{L,r,\theta}} - 1\biggr)^2 dP^{\ell'}_{L,r,\theta} \\
&\quad = \int_{\mathbb{R}^d \times \{0,1\}} \biggl(\frac{p_\ell(x,y)}{p_{\ell'}(x,y)} - 1\biggr)^2 p_{\ell'}(x,y)\, dQ_{L,r,w}(x,y) \\
&\quad = \int_{\mathbb{R}^d \times \{0,1\}} \frac{\{p_\ell(x,y) - p_{\ell'}(x,y)\}^2}{p_{\ell'}(x,y)}\, dQ_{L,r,w}(x,y)
\end{aligned}$$

$$= \int_{\mathbb{R}^d} \biggl(\frac{\{\eta^{\ell}_{L,r,w,s,\theta}(x) - \eta^{\ell'}_{L,r,w,s,\theta}(x)\}^2}{1-\eta^{\ell'}_{L,r,w,s,\theta}(x)} + \frac{\{\eta^{\ell}_{L,r,w,s,\theta}(x) - \eta^{\ell'}_{L,r,w,s,\theta}(x)\}^2}{\eta^{\ell'}_{L,r,w,s,\theta}(x)} \biggr) \, d\mu_{L,r,w}(x)$$

$$\le \frac{4}{\epsilon_0} \int_{\bar{B}_{2s}(z_\ell) \cup \bar{B}_{2s}(z_{\ell'})} \bigl\{\eta^{\ell}_{L,r,w,s,\theta}(x) - \eta^{\ell'}_{L,r,w,s,\theta}(x)\bigr\}^2 \, d\mu_{L,r,w}(x)$$

$$\le \frac{4}{\epsilon_0} \cdot (2\theta)^2 \cdot \mu_{L,r,w}\bigl(\bar{B}_{2s}(z_\ell) \cup \bar{B}_{2s}(z_{\ell'})\bigr) = \frac{2^{5+2d} \cdot w \cdot \theta^2 \cdot s^d}{\epsilon_0 \cdot L},$$

as required. □

We are now in a position to state the crucial proposition for the proof of Proposition 14.

PROPOSITION S32. *Assume that $\mathcal{A}_{\mathrm{hpr}} \subseteq \mathcal{A} \subseteq \mathcal{A}_{\mathrm{conv}}$. Fix $(\beta,\gamma,\kappa,\lambda,C_{\mathrm{App}}) \in (0,\infty)^3 \times [1,\infty)^2$ with $\beta\gamma(\kappa-1) < d\kappa$, $\epsilon_0 \in (0,1/2)$, $\tau \in [\epsilon_0, 1-\epsilon_0]$, $\upsilon \le (4d^{1/2})^{-d}$ and $\zeta > 0$. There exist $C_0 \ge 1$, $c_0, c_1 > 0$, depending only on d, β, γ, κ, C_{App} and ϵ_0, such that for any $n \ge C_0 \lambda^{d/\beta} \log(1+\zeta)$ there exists a family of $L \ge \bigl\{c_0\bigl(n/\{\lambda^{d/\beta}\log(1+\zeta)\}\bigr)^{\frac{\beta\gamma(\kappa\wedge 1)}{\kappa(2\beta+d)+\beta\gamma}}\bigr\} \vee 4$ distributions*

$$\{P_1,\ldots,P_L\} \subseteq \mathcal{P}^\dagger(\mathcal{A},\tau,\beta,\kappa,\gamma,\upsilon,\lambda,C_{\mathrm{App}})$$

with regression functions $\eta_1,\ldots,\eta_L$ and common marginal distribution μ on $\mathbb{R}^d$, such that

(a) *$\chi^2\bigl(P_\ell^{\otimes n}, P_{\ell'}^{\otimes n}\bigr) \le \zeta$ for all ℓ, $\ell' \in [L]$;*
(b) *if $A \in \mathcal{A} \cap \mathrm{Pow}\bigl(\mathcal{X}_\tau(\eta_\ell) \cap \mathcal{X}_\tau(\eta_{\ell'})\bigr)$ for some $\ell,\ell' \in [L]$ with $\ell \neq \ell'$, then*

$$M_\tau\bigl(P_\ell,\mathcal{A}\bigr) - \mu(A) \ge c_1 \cdot \left(\frac{\lambda^{d/\beta}\cdot\log(1+\zeta)}{n}\right)^{\frac{\beta\gamma\kappa}{\kappa(2\beta+d)+\beta\gamma}}. \tag{S34}$$

PROOF. We first define some quantities for our construction. Let $\rho := \kappa(2\beta+d)+\beta\gamma$,

$$\xi_{n,\zeta,\lambda} := \frac{\epsilon_0 \cdot (c^\flat_{\beta,d})^{d/\beta}}{2^{5+4d} d^{d/2}} \cdot \frac{\lambda^{d/\beta}\cdot\log(1+\zeta)}{n}, \quad \theta := \xi_{n,\zeta,\lambda}^{\kappa\beta/\rho}, \quad r := C_{\mathrm{App}}^{1/d} \cdot (8d^{1/2})^{-1} \cdot \theta^{\frac{\gamma(\kappa-1)}{d\kappa}},$$

$$L := \bigl\lfloor (8d^{1/2})^{-1}\theta^{-\gamma/\kappa}\bigl\{(2r)^{-d} \wedge 1\bigr\}\bigr\rfloor, \quad w := (4d^{1/2})^d \cdot L \cdot \theta^{\gamma/\kappa} \quad \text{and} \quad s := \left(\frac{\theta}{c^\flat_{\beta,d}\lambda}\right)^{1/\beta}.$$

Finally, let $C_0 := C_0^0 \cdot (C_0^1 \vee C_0^2 \vee C_0^3 \vee C_0^4 \vee C_0^5)$, where

$$C_0^0 := \frac{\epsilon_0 \cdot (c^\flat_{\beta,d})^{d/\beta}}{2^{5+4d}\cdot d^{d/2}}, \quad C_0^1 := \left(\frac{16d^{1/2}}{(c^\flat_{\beta,d})^{1/\beta}\cdot C_{\mathrm{App}}^{1/d}}\right)^{\frac{d\rho}{d\kappa-\beta\gamma(\kappa-1)}}, \quad C_0^2 := \frac{1}{(c^\flat_{\beta,d})^{\rho/(\kappa\beta)}},$$

$$C_0^3 := \left(\frac{2}{\epsilon_0}\right)^{\frac{\rho}{\kappa\beta}}, \quad C_0^4 := \left(\frac{C_{\mathrm{App}}}{2^{2d-5}d^{(d-1)/2}}\right)^{\frac{\rho}{\kappa\beta\gamma}} \quad \text{and} \quad C_0^5 := (32d^{1/2})^{\frac{\rho}{\beta\gamma}}.$$

Observe that when $n \ge C_0 \cdot \lambda^{d/\beta}\cdot\log(1+\zeta)$, we have $\xi_{n,\zeta,\lambda} \le 1/(C_0^1 \vee C_0^2 \vee C_0^3 \vee C_0^4 \vee C_0^5)$. Hence, the choice of C_0^1 ensures that $s \le r/2$, the choice of C_0^2 guarantees that $s \le 1$, the choice of C_0^3 ensures that $\theta \le \epsilon_0/2$, and C_0^4 and C_0^5 are chosen to guarantee that

$$L = \left\lfloor 4 \cdot \min\left\{ \left(\frac{n}{C_0^0 C_0^4 \cdot \lambda^{d/\beta}\cdot\log(1+\zeta)}\right)^{\frac{\beta\gamma\kappa}{\rho}}, \left(\frac{n}{C_0^0 C_0^5 \cdot \lambda^{d/\beta}\cdot\log(1+\zeta)}\right)^{\frac{\beta\gamma}{\rho}} \right\}\right\rfloor$$

$$\ge \left\{ c_0 \cdot \left(\frac{n}{\lambda^{d/\beta}\cdot\log(1+\zeta)}\right)^{\frac{\beta\gamma(\kappa\wedge 1)}{\kappa(2\beta+d)+\beta\gamma}} \right\} \vee 4,$$

where $c_0 := 2\min\{(C_0^0 C_0^4)^{-\frac{\beta\gamma\kappa}{\rho}}, (C_0^0 C_0^5)^{-\frac{\beta\gamma}{\rho}}\}$. Note also that $w \leq \frac{1}{2}\{(2r)^{-d} \wedge 1\}$. We may therefore apply the construction following (22) to define distributions $P_\ell := P_{L,r,w,s,\theta}^\ell$ for $\ell \in [L]$ when $n \geq C_0 \cdot \lambda^{d/\beta} \cdot \log(1+\zeta)$. We write $\mu = \mu_{L,r,w}$ and $\eta_\ell = \eta_{L,r,w,s,\theta}^\ell$ in this construction.

Our choice of s ensures that $\theta = c_{\beta,d}^\flat \cdot \lambda \cdot s^\beta$, so we may apply Lemma S28 to deduce that $P_\ell \in \mathcal{P}_{\mathrm{H\ddot{o}l}}(\beta,\lambda)$ for all $\ell \in [L]$. Moreover, our choice of w and r guarantee that $(w/L)\cdot(2r)^d \leq C_{\mathrm{App}} \cdot \min\{(w/\{(4d^{1/2})^d \cdot L\})^\kappa, \theta^\gamma\}$, so we may apply Lemma S30 to conclude that $P_\ell \in \mathcal{P}_{\mathrm{App}}(\mathcal{A},\kappa,\gamma,\tau,C_{\mathrm{App}}) \cap \mathcal{P}^+_{\mathrm{App}}(\mathcal{A},\tau,\kappa,\gamma,\upsilon,\tau,C_{\mathrm{App}})$ for all $\ell \in [L]$. Next, by Lemma S31, for each ℓ, $\ell' \in [L]$,

$$\chi^2(P_\ell, P_{\ell'}) \leq \frac{2^{5+2d} \cdot w \cdot \theta^2 \cdot s^d}{\epsilon_0 \cdot L} = \frac{\log(1+\zeta)}{n},$$

by our choice of w, θ and s, so

$$\chi^2(P_\ell^{\otimes n}, P_{\ell'}^{\otimes n}) \leq \{1 + \chi^2(P_\ell, P_{\ell'})\}^n - 1 \leq \left(1 + \frac{\log(1+\zeta)}{n}\right)^n - 1 \leq \zeta.$$

This proves *(a)*. To prove *(b)*, we take $\ell,\ell' \in [L]$ with $\ell \neq \ell'$ and $A \in \mathcal{A} \cap \mathrm{Pow}(\mathcal{X}_\tau(\eta_\ell) \cap \mathcal{X}_\tau(\eta_{\ell'}))$. By Lemma S30, we have $M_\tau(P_\ell, \mathcal{A}) = \mu(K_r^0(\ell)) = (w/L)\cdot(2r)^d$. On the other hand, if $\mu(A) > 0$, then since $\mathrm{supp}(\mu) = \bigcup_{(\ell'',j)\in[L]\times\{0,1\}} K_r^j(\ell'')$ and $\bigcup_{\ell''\in[L]} K_r^1(\ell'') \subseteq \mathbb{R}^d \setminus \mathcal{X}_\tau(\eta_{\ell''})$, we must have $A \cap K_r^0(\tilde{\ell}) \neq \emptyset$ for some $\tilde{\ell} \in [L]$. Since at least one of $\tilde{\ell} \neq \ell$ or $\tilde{\ell} \neq \ell'$ must hold, it follows from Lemma S30 that $\mu(A) \leq (w/L)\cdot(2r)^d/2$. Hence

$$M_\tau(P_\ell, \mathcal{A}) - \mu(A) \geq \frac{w}{L} \cdot 2^{d-1} \cdot r^d = \frac{C_{\mathrm{App}}}{2} \cdot \theta^\gamma = \frac{C_{\mathrm{App}}}{2}\left(\frac{C_0^0 \cdot \lambda^{d/\beta} \cdot \log(1+\zeta)}{n}\right)^{\frac{\beta\gamma\kappa}{\kappa(2\beta+d)+\beta\gamma}},$$

so (S34) holds with $c_1 := \frac{C_{\mathrm{App}}}{2} \cdot (C_0^0)^{\frac{\beta\kappa\gamma}{\rho}}$. □

PROOF OF PROPOSITION 14. Part (i): We initially assume that $n \geq C_0 \lambda^{d/\beta} \log(1/(4\alpha))$. By Proposition S32, with $\zeta = 1/(4\alpha) - 1 > 0$, there exists a pair of distributions P_1, $P_2 \in \mathcal{P}^\dagger(\mathcal{A},\tau,\beta,\kappa,\gamma,\upsilon,\lambda,C_{\mathrm{App}})$ with common marginal distribution μ on $\mathbb{R}^d$ and corresponding regression functions η_1,η_2 such that $\chi^2(P_2^{\otimes n}, P_1^{\otimes n}) + 1 \leq 1/(4\alpha)$ and if $A \in \mathcal{A} \cap \mathrm{Pow}(\mathcal{X}_\tau(\eta_1) \cap \mathcal{X}_\tau(\eta_2))$, then

$$M_\tau(P_2, \mathcal{A}) - \mu(A) \geq c_1 \cdot \left(\frac{\lambda^{d/\beta} \cdot \log(1/(4\alpha))}{n}\right)^{\frac{\beta\gamma\kappa}{\kappa(2\beta+d)+\beta\gamma}}. \tag{S35}$$

We now define a test $\varphi : (\mathbb{R}^d \times [0,1])^n \to \{1,2\}$ by

$$\varphi(D) := \begin{cases} 1 & \text{if } \hat{A}(D) \subseteq \mathcal{X}_\tau(\eta_1) \\ 2 & \text{otherwise.} \end{cases}$$

Since $\hat{A}$ controls the Type I error at level α over $\mathcal{P}^\dagger(\mathcal{A},\tau,\beta,\kappa,\gamma,\upsilon,\lambda,C_{\mathrm{App}})$, we have

$$P_1^{\otimes n}(\{D \in (\mathbb{R}^d \times [0,1])^n : \varphi(D) = 2\}) = \mathbb{P}_{P_1}(\hat{A}(\mathcal{D}) \not\subseteq \mathcal{X}_\tau(\eta_1)) \leq \alpha.$$

Hence, by an immediate consequence of Brown and Low (1996, Theorem 1), which we restate as Lemma S41 for convenience, with $\epsilon = \sqrt{\alpha}$ we have

$$\begin{aligned}\mathbb{P}_{P_2}(\hat{A}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta_1)) &= P_2^{\otimes n}(\{D \in (\mathbb{R}^d \times [0,1])^n : \varphi(D) = 1\}) \\ &\geq \left\{1 - \epsilon\sqrt{\chi^2(P_2^{\otimes n}, P_1^{\otimes n}) + 1}\right\}^2 \geq \frac{1}{4}.\end{aligned}$$

Thus, by (S35) we have

$$\mathbb{E}_{P_2}\big[\big\{M_\tau(P,\mathcal{A})-\mu(\hat{A})\big\}\cdot\mathbb{1}_{\{\hat{A}\subseteq\mathcal{X}_\tau(\eta_2)\}}\big]$$

$$\geq \mathbb{P}_{P_2}\big(\{\hat{A}(\mathcal{D})\subseteq\mathcal{X}_\tau(\eta_1)\}\cap\{\hat{A}(\mathcal{D})\subseteq\mathcal{X}_\tau(\eta_2)\}\big)\cdot c_1\bigg(\frac{\lambda^{d/\beta}\log\big(1/(4\alpha)\big)}{n}\bigg)^{\frac{\beta\gamma\kappa}{\kappa(2\beta+d)+\beta\gamma}}$$

$$\geq \big\{\mathbb{P}_{P_2}\big(\hat{A}(\mathcal{D})\subseteq\mathcal{X}_\tau(\eta_1)\big)-\mathbb{P}_{P_2}\big(\hat{A}(\mathcal{D})\not\subseteq\mathcal{X}_\tau(\eta_2)\big)\big\}\cdot c_1\bigg(\frac{\lambda^{d/\beta}\log\big(1/(4\alpha)\big)}{n}\bigg)^{\frac{\beta\gamma\kappa}{\kappa(2\beta+d)+\beta\gamma}}$$

(S36)

$$\geq\bigg(\frac{1}{4}-\alpha\bigg)\cdot c_1\bigg(\frac{\log\big(1/(4\alpha)\big)}{n}\bigg)^{\frac{\beta\gamma\kappa}{\kappa(2\beta+d)+\beta\gamma}}\geq\frac{c_1}{8}\cdot\bigg(\frac{\lambda^{d/\beta}\cdot\log\big(1/(4\alpha)\big)}{n}\bigg)^{\frac{\beta\gamma\kappa}{\kappa(2\beta+d)+\beta\gamma}},$$

as required. Moreover, if $1\leq n<C_0\cdot\lambda^{d/\beta}\cdot\log\big(1/(4\alpha)\big)$, then by (S36) with $n=\lceil C_0\cdot\lambda^{d/\beta}\cdot\log\big(1/(4\alpha)\big)\rceil$, we have

$$\mathbb{E}_{P_2}\big[\big\{M_\tau(P,\mathcal{A})-\mu(\hat{A})\big\}\cdot\mathbb{1}_{\{\hat{A}\subseteq\mathcal{X}_\tau(\eta_2)\}}\big]\geq\frac{c_1}{8}\cdot C_0^{-\frac{\beta\gamma\kappa}{\kappa(2\beta+d)+\beta\gamma}},$$

again. This completes the proof of Part *(i)*.

Part (ii): Fix $\rho:=\kappa(2\beta+d)+\beta\gamma$ and let $C_0\geq1$ and $c_0,c_1>0$ be as in Proposition S32. Let $C_1\equiv C_1(d,\beta,\gamma,\kappa,C_{\mathrm{App}},\epsilon_0)\geq e^{\frac{2\rho}{\beta\gamma}}-1$ be large enough that for all $n/\lambda^{d/\beta}\geq C_1$, we have both $n/\lambda^{d/\beta}\geq C_0\log\big(1+(n/\lambda^{d/\beta})^{\frac{\beta\gamma(\kappa\wedge1)}{2\rho}}\big)$ and

$$c_0\epsilon_0^2\Bigg(\frac{(n/\lambda^{d/\beta})}{\log\big(1+(n/\lambda^{d/\beta})^{\frac{\beta\gamma(\kappa\wedge1)}{2\rho}}\big)}\Bigg)^{\frac{\beta\gamma(\kappa\wedge1)}{\rho}}\geq2^5(n/\lambda^{d/\beta})^{\frac{\beta\gamma(\kappa\wedge1)}{2\rho}}.$$

Next, for $n\geq C_1$, we apply Proposition S32 with $\zeta=(n/\lambda^{d/\beta})^{\frac{\beta\gamma(\kappa\wedge1)}{2\rho}}$ to obtain a family of $L\geq c_0\big\{(n/\lambda^{d/\beta})/\log\big(1+(n/\lambda^{d/\beta})^{\frac{\beta\gamma(\kappa\wedge1)}{2\rho}}\big)\big\}^{\frac{\beta\gamma(\kappa\wedge1)}{\rho}}\geq2^5(n/\lambda^{d/\beta})^{\frac{\beta\gamma(\kappa\wedge1)}{2\rho}}/\epsilon_0^2$ distributions $\{P_1,\ldots,P_L\}\subseteq\mathcal{P}_{\mathrm{H\ddot{o}l}}(\beta,\lambda)\cap\mathcal{P}_{\mathrm{App}}(\mathcal{A},\kappa,\gamma,\tau,C_{\mathrm{App}})\cap\mathcal{P}^+_{\mathrm{App}}(\mathcal{A},\tau,\kappa,\gamma,\upsilon,\tau,C_{\mathrm{App}})$ with common marginal distribution μ on $\mathbb{R}^d$ and corresponding regression functions $\eta_1,\ldots,\eta_L$, such that

(a) $\chi^2\big(P_\ell^{\otimes n},P_{\ell'}^{\otimes n}\big)\leq(n/\lambda^{d/\beta})^{\frac{\beta\gamma(\kappa\wedge1)}{2\rho}}\leq(\epsilon_0/4)^2\cdot(L-1)$ for all $\ell,\ell'\in[L]$;
(b) if $A\in\mathcal{A}\cap\mathrm{Pow}\big(\mathcal{X}_\tau(\eta_\ell)\cap\mathcal{X}_\tau(\eta_{\ell'})\big)$ for some $\ell,\ell'\in[L]$ with $\ell\neq\ell'$, then

$$M_\tau\big(P_\ell,\mathcal{A}\big)-\mu(A)\geq c_1\cdot\Bigg(\frac{\lambda^{d/\beta}\log\{1+(n/\lambda^{d/\beta})^{\frac{\beta\gamma(\kappa\wedge1)}{2\rho}}\}}{n}\Bigg)^{\beta\gamma\kappa/\rho}. \tag{S37}$$

Now define a test function $\varphi:(\mathbb{R}^d\times[0,1])^n\to[L]$ by

$$\varphi(D):=\begin{cases}\min\{\ell\in[L]:\hat{A}(D)\subseteq\mathcal{X}_\tau(\eta_\ell)\} & \text{if } \hat{A}(D)\subseteq\mathcal{X}_\tau(\eta_\ell)\text{ for some }\ell\in[L]\\ L & \text{otherwise.}\end{cases}$$

By Lemma S42, we have

$$\max_{\ell\in[L]}P_\ell^{\otimes n}\big(\{D\in(\mathbb{R}^d\times[0,1])^n:\varphi(D)\neq\ell\}\big)$$

$$\geq\frac{1}{L-1}\sum_{\ell=2}^{L}P_\ell^{\otimes n}\big(\{D\in(\mathbb{R}^d\times[0,1])^n:\varphi(D)\neq\ell\}\big)$$

$$\geq 1 - \frac{1}{L-1} - \sqrt{\frac{1}{L-1}\sum_{\ell=2}^{L}\chi^2(P_\ell^{\otimes n}, P_1^{\otimes n}) \cdot \frac{1}{L-1}\biggl(1 - \frac{1}{L-1}\biggr)}$$

$$\geq 1 - \biggl(\frac{\epsilon_0}{4}\biggr)^2 - \frac{\epsilon_0}{4} \geq 1 - \frac{\epsilon_0}{2}. \tag{S38}$$

Now choose $\ell_0 \in [L]$ with $P_{\ell_0}^{\otimes n}\bigl(\{D \in (\mathbb{R}^d \times [0,1])^n : \varphi(D) \neq \ell_0\}\bigr) \geq 1 - \epsilon_0/2$, and observe that if $\varphi(D) \neq \ell_0$ and $\hat{A}(D) \subseteq \mathcal{X}_\tau(\eta_{\ell_0})$ then we must also have $\hat{A}(D) \subseteq \mathcal{X}_\tau(\eta_{\ell_1})$ for some $\ell_1 \in [\ell_0 - 1]$. It follows from this and (S38) that

$$\begin{aligned}\mathbb{P}_{P_{\ell_0}}&\biggl(\bigl\{\hat{A}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta_{\ell_0})\bigr\} \cap \bigcup_{\ell_1=1}^{\ell_0-1}\bigl\{\hat{A}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta_{\ell_1})\bigr\}\biggr) \\ &\geq \mathbb{P}_{P_{\ell_0}}\bigl(\bigl\{\hat{A}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta_{\ell_0})\bigr\} \cap \{\varphi(\mathcal{D}) \neq \ell_0\}\bigr) \\ &\geq P_{\ell_0}^{\otimes n}\bigl(\{D \in (\mathbb{R}^d \times [0,1])^n : \varphi(D) \neq \ell_0\}\bigr) - \alpha \geq \frac{\epsilon_0}{2}.\end{aligned}$$

Thus, by (S37), we have for all $n \geq C_1 \cdot \lambda^{d/\beta} \geq e^{\frac{2\rho}{\beta\gamma}} - 1$ that

$$\begin{aligned}&\mathbb{E}_{P_{\ell_0}}\bigl[\bigl\{M_\tau(P_{\ell_0}, \mathcal{A}) - \mu(\hat{A})\bigr\} \cdot \mathbb{1}_{\{\hat{A} \subseteq \mathcal{X}_\tau(\eta_{\ell_0})\}}\bigr] \\ &\geq \mathbb{P}_{P_{\ell_0}}\biggl(\bigl\{\hat{A}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta_{\ell_0})\bigr\} \cap \bigcup_{\ell_1=1}^{\ell_0-1}\bigl\{\hat{A}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta_{\ell_1})\bigr\}\biggr) \\ &\qquad \cdot c_1 \cdot \biggl(\frac{\lambda^{d/\beta}\log\{1 + (n/\lambda^{d/\beta})^{\frac{\beta\gamma(\kappa\wedge 1)}{2\rho}}\}}{n}\biggr)^{\beta\gamma\kappa/\rho} \\ &\geq \frac{c_1\epsilon_0}{2} \cdot \biggl(\frac{\beta\gamma(\kappa\wedge 1) \cdot \lambda^{d/\beta} \cdot \log_+(n/\lambda^{d/\beta})}{2\rho n}\biggr)^{\beta\gamma\kappa/\rho}.\end{aligned}$$

We extend the bound to $n < C_1 \cdot \lambda^{d/\beta}$ by monotonicity as at the end of the proof of Proposition 14*(i)*, with

$$c_2 := \frac{c_1\epsilon_0}{2} \cdot \biggl(\frac{\beta\gamma(\kappa\wedge 1)\log_+\lceil C_1\rceil}{2\rho\lceil C_1\rceil}\biggr)^{\beta\gamma\kappa/\rho},$$

which completes the proof. □

Finally, we prove the parametric lower bounds in Theorems 2 and 11. Some care is required here to show that our constructed distributions belong to the relevant distributional classes.

PROOF OF PROPOSITION 15. Observe that there exists $c_{\mathrm{H}} \equiv c_{\mathrm{H}}(\beta) \in (0,1]$ such that when $\theta \leq c_{\mathrm{H}} \cdot \lambda \cdot s^\beta$, we have that η is (β,λ)-Hölder, so $\{P_\zeta^\ell\}_{\ell\in\{-1,1\}} \subseteq \mathcal{P}_{\mathrm{Höl}}(\beta,\lambda)$. In addition, $\mathrm{supp}(\mu_\zeta^\ell) \cap \mathcal{X}_\tau(\eta) = A_{-1} \cup A_1$ for $\ell \in \{-1,1\}$. Since $\mathcal{A} \subseteq \mathcal{A}_{\mathrm{conv}}$ and $A_0 \subseteq \mathbb{R}^d \setminus \mathcal{X}_\tau(\eta)$, it follows that for $\ell \in \{-1,1\}$,

$$M_\tau(P_\zeta^\ell, \mathcal{A}) = \max_{j\in\{-1,1\}} \mu_\zeta^\ell(A_j) = \frac{s^d}{(2t)^d + 2s^d} + \zeta.$$

Observe also that for any $\ell \in \{-1,1\}$, $x \in A_\ell$ and $r \in (0, 4s]$, we have

$$\mu_\zeta^\ell(\bar{B}_r(x)) \geq \biggl(\frac{1}{(2t)^d + 2s^d} + \frac{\zeta}{s^d}\biggr) \cdot \mathcal{L}_d\bigl(\bar{B}_{r/4}(x) \cap A_\ell\bigr) \geq \biggl(\frac{1}{(2t)^d + 2s^d} + \frac{\zeta}{s^d}\biggr) \cdot \biggl(\frac{r}{4}\biggr)^d.$$

Moreover, for any $\ell \in \{-1,1\}$, $x \in A_\ell$ and $r \in (4s,1]$, we have

$$\mu_\zeta^\ell\bigl(\bar{B}_r(x)\bigr) \geq \frac{\mathcal{L}_d\bigl(\bar{B}_r(x)\cap A_0\bigr)}{(2t)^d+2s^d} \geq \frac{r^{d-1}\cdot(r-2s)}{(2t)^d+2s^d}$$
$$\geq \frac{r^d}{2\{(2t)^d+2s^d\}} \geq \biggl(\frac{1}{(2t)^d+2s^d}+\frac{\zeta}{s^d}\biggr)\cdot\frac{r^d}{3}.$$

Hence, $\omega_{\mu_\zeta^\ell,d}(x) \geq \{(8t)^d+2(4s)^d\}^{-1}$ for all $x \in A_\ell$, and moreover $A_\ell \subseteq \mathcal{R}_\upsilon(\mu_\zeta^\ell)$. Thus, for any $(\xi,\Delta) \in \bigl(0,\{(8t)^d+2(4s)^d\}^{-1}\bigr] \times (0,\theta]$, we have

$$\sup\bigl\{\mu_\zeta^\ell(A) : A \in \mathcal{A}\cap\mathrm{Pow}\bigl(\mathcal{X}_\xi(\omega_{\mu_\zeta^\ell,d})\cap\mathcal{X}_{\tau+\Delta}(\eta)\bigr)\bigr\} = \mu_\zeta^\ell(A_\ell) = M_\tau(P_\zeta^\ell,\mathcal{A}),$$

and similarly,

$$\sup\bigl\{\mu_\zeta^\ell(A) : A \in \mathcal{A}\cap\mathrm{Pow}\bigl(\mathcal{R}_\upsilon(\mu_\zeta^\ell)\cap\mathcal{X}_\xi(f_{\mu_\zeta^\ell})\cap\mathcal{X}_{\tau+\Delta}(\eta)\bigr)\bigr\} = \mu_\zeta^\ell(A_\ell) = M_\tau(P_\zeta^\ell,\mathcal{A}).$$

On the other hand, if either $\xi > \{(8t)^d+2(4s)^d\}^{-1}$ or $\Delta > \theta$, then provided that $\frac{3s^d}{2\{(2t)^d+2s^d\}} \leq C_{\mathrm{App}}\cdot\bigl[\{(8t)^d+2(4s)^d\}^{-\kappa}\wedge\theta^\gamma\bigr]$, we have

$$\begin{aligned}&\sup\bigl\{\mu_\zeta^\ell(A) : A \in \mathcal{A}\cap\mathrm{Pow}\bigl(\mathcal{R}_\upsilon(\mu_\zeta^\ell)\cap\mathcal{X}_\xi(f_{\mu_\zeta^\ell})\cap\mathcal{X}_{\tau+\Delta}(\eta)\bigr)\bigr\}\\ &\quad\wedge\sup\bigl\{\mu_\zeta^\ell(A) : A \in \mathcal{A}\cap\mathrm{Pow}\bigl(\mathcal{X}_\xi(\omega_{\mu_\zeta^\ell,d})\cap\mathcal{X}_{\tau+\Delta}(\eta)\bigr)\bigr\}\\ &\quad\geq 0 \geq M_\tau(P_\zeta^\ell,\mathcal{A}) - C_{\mathrm{App}}\cdot(\xi^\kappa+\Delta^\gamma).\end{aligned}$$

It follows that $\{P_\zeta^\ell\}_{\ell\in\{-1,1\}} \subseteq \mathcal{P}^\dagger(\mathcal{A},\tau,\beta,\kappa,\gamma,\upsilon,\lambda,C_{\mathrm{App}})$ whenever $\frac{3s^d}{2\{(2t)^d+2s^d\}} \leq C_{\mathrm{App}}\cdot\bigl[\{(8t)^d+2(4s)^d\}^{-\kappa}\wedge\theta^\gamma\bigr]$.

In addition, recalling that m_{ct} denotes the counting measure, we have

$$\begin{aligned}\mathrm{KL}\bigl(P_\zeta^{-1},P_\zeta^1\bigr) &\leq \chi^2\bigl(P_\zeta^{-1},P_\zeta^1\bigr)\\ &= \int_{\mathbb{R}^d\times\{0,1\}} \frac{\bigl\{f_{\mu_\zeta^{-1}}(x)-f_{\mu_\zeta^1}(x)\bigr\}^2\eta(x)^y\{1-\eta(x)\}^{1-y}}{f_{\mu_\zeta^1}(x)}\,d(\mathcal{L}_d\times m_{\mathrm{ct}})(x,y)\\ &= \frac{(2\zeta/s^d)^2}{\frac{1}{(2t)^d+2s^d}-\frac{\zeta}{s^d}} + \frac{(2\zeta/s^d)^2}{\frac{1}{(2t)^d+2s^d}+\frac{\zeta}{s^d}} \leq \frac{32}{3}\cdot\frac{(2t)^d+2s^d}{s^{2d}}\cdot\zeta^2.\end{aligned}$$

We deduce by Pinsker's inequality that

$$\begin{aligned}\mathrm{TV}\bigl((P_\zeta^{-1})^{\otimes n},(P_\zeta^1)^{\otimes n}\bigr) \leq \sqrt{\mathrm{KL}\bigl((P_\zeta^{-1})^{\otimes n},(P_\zeta^{+1})^{\otimes n}\bigr)/2} &\leq 4\zeta\sqrt{n}\cdot\sqrt{\frac{(2t)^d+2s^d}{3s^{2d}}}\\ &= 2a\cdot\zeta\sqrt{n},\end{aligned}$$

where $a \equiv a_{d,s,t} := 2\cdot\sqrt{\frac{(2t)^d+2s^d}{3s^{2d}}}$. Thus,

$$\begin{aligned}1 &= (P_\zeta^1)^{\otimes n}\bigl(\{D \in (\mathbb{R}^d\times\{0,1\})^n : A_1\cap\hat{A}(D)=\emptyset\}\bigr)\\ &\qquad + (P_\zeta^1)^{\otimes n}\bigl(\{D \in (\mathbb{R}^d\times\{0,1\})^n : A_1\cap\hat{A}(D)\neq\emptyset\}\bigr)\\ &\leq (P_\zeta^1)^{\otimes n}\bigl(\{D : A_1\cap\hat{A}(D)=\emptyset\}\bigr) + (P_\zeta^{-1})^{\otimes n}\bigl(\{D : A_1\cap\hat{A}(D)\neq\emptyset\}\bigr) + 2a\cdot\zeta\sqrt{n}.\end{aligned}$$

Hence, since at most one of A_{-1} and A_1 can have non-empty intersection with $\hat{A}$ whenever $\hat{A} \subseteq \mathcal{X}_\tau(\eta)$, we have

$$\max_{\ell \in \{-1,1\}} (P_\zeta^\ell)^{\otimes n}\bigl(\{D : A_\ell \cap \hat{A}(D) = \emptyset\} \cap \bigl\{D : \hat{A}(D) \subseteq \mathcal{X}_\tau(\eta)\bigr\}\bigr) \geq \frac{1}{2} - a \cdot \zeta\sqrt{n} - \alpha$$
$$\geq \epsilon_0 - a \cdot \zeta\sqrt{n}.$$

We conclude that

$$\max_{\ell \in \{-1,1\}} \mathbb{E}_{P_\zeta^\ell}\bigl[\bigl\{M_\tau\bigl(P_\zeta^\ell, \mathcal{A}\bigr) - \mu_\zeta^\ell(\hat{A})\bigr\} \cdot \mathbb{1}_{\{\hat{A} \subseteq \mathcal{X}_\tau(\eta)\}}\bigr] \geq 2\zeta\bigl(\epsilon_0 - a \cdot \zeta\sqrt{n}\bigr).$$

Taking $t := \bigl\{\bigl(2^{\kappa(3d+1)} \vee (c_{\mathrm{H}} \cdot \lambda)^\gamma\bigr)/C_{\mathrm{App}}\bigr\}^{\frac{\beta\gamma}{d(d\kappa - \beta\gamma(\kappa-1))}} \vee 1$, $s := t^{-\frac{d\kappa}{\beta\gamma}}$ and $\theta := c_{\mathrm{H}} \cdot \lambda \cdot s^\beta$ ensures that the conditions $\frac{3s^d}{2\{(2t)^d + 2s^d\}} \leq C_{\mathrm{App}} \cdot \bigl[\{(8t)^d + 2(4s)^d\}^{-\kappa} \wedge \bigl(c_{\mathrm{H}} \cdot \lambda \cdot s^\beta\bigr)^\gamma\bigr]$ and $\theta \leq c_{\mathrm{H}} \cdot \lambda \cdot s^\beta$ hold. Hence, taking $\zeta := \epsilon_0/(2a\sqrt{n})$ yields the required lower bound. □

PROOF OF THEOREMS 2*(ii)* AND 11*(ii)*. In light of the remarks following Proposition 14, these results follow from Propositions 14 and 15. □

S7. Parameter constraints. The following lemma reveals natural constraints satisfied by the parameters β, κ and γ.

LEMMA S33. *Take $\tau \in (0,1)$, $(\beta, \gamma, \kappa, \upsilon, \lambda, C_{\mathrm{App}}) \in (0,1] \times (0,\infty)^2 \times (0,1) \times [1,\infty)^2$. Let $P \in \mathcal{P}_{\mathrm{H\ddot{o}l}}(\beta, \lambda) \cap \mathcal{P}_{\mathrm{App}}(\mathcal{A}_{\mathrm{hpr}}, \tau, \kappa, \gamma, \tau, C_{\mathrm{App}})$ be a distribution on $\mathbb{R}^d \times [0,1]$ with regression function $\eta : \mathbb{R}^d \to [0,1]$ and with a Lebesgue absolutely continuous marginal μ on $\mathbb{R}^d$ with continuous density f_μ. Suppose that $\mathcal{X}_\tau(\eta) \subseteq \mathcal{R}_\upsilon(\mu)$, that $\mu\bigl(\eta^{-1}((\tau,1])\bigr) > 0$ and that $\eta^{-1}(\{\tau\}) \neq \emptyset$. Then $\beta\gamma(\kappa - 1) \leq d\kappa$.*

PROOF. Note that since $\mu\bigl(\eta^{-1}((\tau,1])\bigr) > 0$ and η is continuous, we must have $M_\tau > 0$. Take $\Delta \in \bigl(0, \{M_\tau/(2C_{\mathrm{App}})\}^{1/\gamma}\bigr)$, and write $\omega := \omega_{\mu,d}$ for the lower density of μ. Since $P \in \mathcal{P}_{\mathrm{App}}(\mathcal{A}_{\mathrm{hpr}}, \tau, \kappa, \gamma, \tau, C_{\mathrm{App}})$, we may take $A_\Delta \in \mathcal{A}_{\mathrm{hpr}} \cap \mathrm{Pow}\bigl(\mathcal{X}_{\Delta^{\gamma/\kappa}}(\omega) \cap \mathcal{X}_{\tau+\Delta}(\eta)\bigr)$ with $\mu(A_\Delta) \geq M_\tau - 2C_{\mathrm{App}} \cdot \Delta^\gamma > 0$. Now A_Δ is a non-empty, compact subset of $\mathbb{R}^d$ and $\eta^{-1}(\{\tau\})$ is a non-empty closed subset of $\mathbb{R}^d$, so we may choose $x_0 \in \eta^{-1}(\{\tau\})$ and $z_0 \in A_\Delta$ such that

$$\|x_0 - z_0\|_\infty = \inf_{x \in \eta^{-1}(\{\tau\})} \inf_{z \in A_\Delta} \|x - z\|_\infty.$$

Let $A_\Delta^\sharp := \{x \in \mathbb{R}^d : \|x - z\|_\infty \leq \|x_0 - z_0\|_\infty$ for some $z \in A_\Delta\}$, and note that $A_\Delta^\sharp \in \mathcal{A}_{\mathrm{hpr}} \cap \mathrm{Pow}\bigl(\mathcal{X}_\tau(\eta)\bigr)$, so $\mu(A_\Delta^\sharp) \leq M_\tau$ and $\mu(A_\Delta^\sharp \setminus A_\Delta) \leq 2C_{\mathrm{App}} \cdot \Delta^\gamma$. In addition, since $P \in \mathcal{P}_{\mathrm{H\ddot{o}l}}(\beta, \lambda)$, we have $\|x_0 - z_0\|_\infty \geq (\Delta/\lambda)^{1/\beta} =: 3r_\Delta$. Hence, if we take

$$w_0 := z_0 + \left(1 + \frac{\upsilon}{2}\right) \cdot r_\Delta \cdot \frac{x_0 - z_0}{\|x_0 - z_0\|_\infty},$$

we have $\bar{B}_{r_\Delta}(w_0) \subseteq \mathcal{X}_\tau(\eta) \cap (A_\Delta^\sharp \setminus A_\Delta)$ and $z_0 \in B_{(1+\upsilon)\cdot r_\Delta}(w_0)$. Thus, as $f_\mu(z_0) \geq 2^{-d} \cdot \omega(z_0) \geq 2^{-d} \cdot \Delta^{\gamma/\kappa}$ and $w_0 \in \mathcal{X}_\tau(\eta) \subseteq \mathcal{R}_\upsilon(\mu)$ and $r_\Delta \in (0,1)$, we have

$$2C_{\mathrm{App}} \cdot \Delta^\gamma \geq \mu(A_\Delta^\sharp \setminus A_\Delta) \geq \mu\bigl(\bar{B}_{r_\Delta}(w_0)\bigr) \geq \upsilon \cdot r_\Delta^d \cdot \sup_{x' \in B_{(1+\upsilon)r_\Delta}(w_0)} f_\mu(x')$$
$$\geq \upsilon \cdot r_\Delta^d \cdot f_\mu(z_0) \geq \frac{\upsilon}{6^d} \cdot \left(\frac{\Delta}{\lambda}\right)^{d/\beta} \cdot \Delta^{\gamma/\kappa}.$$

Letting $\Delta \searrow 0$ we deduce that $\beta\gamma(\kappa - 1) \leq d\kappa$. □

S8. Auxiliary results.

S8.1. *Disintegration and measure-theoretic preliminaries.* Suppose we have a pair of measurable spaces $(\mathcal{X}, \mathcal{G}_X)$ and $(\mathcal{Y}, \mathcal{G}_Y)$ along with a probability distribution P on the product space $(\mathcal{X} \times \mathcal{Y}, \mathcal{G}_X \otimes \mathcal{G}_Y)$. Let μ denote the marginal distribution of P on $(\mathcal{X}, \mathcal{G}_X)$. We say that $(P_x)_{x\in\mathcal{X}}$ is a *disintegration* of P into conditional distributions on $\mathcal{Y}$ if

(a) P_x is a probability measure on $(\mathcal{Y}, \mathcal{G}_Y)$, for each $x \in \mathcal{X}$;
(b) $x \mapsto P_x(B)$ is a $\mathcal{G}_X$-measurable function, for every $B \in \mathcal{G}_Y$;
(c) $P(A \times B) = \int_A P_x(B)\, d\mu(x)$ for all $A \in \mathcal{G}_X$ and $B \in \mathcal{G}_Y$.

We will make use of the following existence result: recall that a topological space $(\mathcal{X}, \mathcal{T}_X)$ is said to be *Polish* if there exists a metric d_X on $\mathcal{X}$ that induces the topology $\mathcal{T}_X$ and for which $(\mathcal{X}, d_X)$ is a complete, separable metric space.

LEMMA S34. *Suppose that $(\mathcal{X}, \mathcal{G}_X)$ and $(\mathcal{Y}, \mathcal{G}_Y)$ are Polish spaces with their corresponding Borel σ-algebras. Let P be a probability distribution on $(\mathcal{X} \times \mathcal{Y}, \mathcal{G}_X \otimes \mathcal{G}_Y)$, with μ denoting the marginal distribution of P on $(\mathcal{X}, \mathcal{G}_X)$. Then there exists a disintegration $(P_x)_{x\in\mathcal{X}}$ of P into conditional distributions on $\mathcal{Y}$ with the property that*

$$\int_{\mathcal{X}\times\mathcal{Y}} g(x,y)\, dP(x,y) = \int_{\mathcal{X}} \biggl(\int_{\mathcal{Y}} g(x,y)\, dP_x(y) \biggr) d\mu(x), \tag{S39}$$

for every P-integrable function $g : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$. Moreover, the disintegration $(P_x)_{x\in\mathcal{X}}$ of P is unique in the sense that if there exists another disintegration $(\tilde{P}_x)_{x\in\mathcal{X}}$ of P into conditional distributions on $\mathcal{Y}$, then $\tilde{P}_x = P_x$ for μ-almost every $x \in \mathcal{X}$.

PROOF. This follows by combining Theorems 10.2.1 and 10.2.2 of Dudley (2018). □

A disintegration has the following useful interpretation. Suppose we have a pair of random variables (X, Y) taking values in $\mathcal{X} \times \mathcal{Y}$ with joint distribution P on $\mathcal{G}_X \times \mathcal{G}_Y$, where $(\mathcal{X}, \mathcal{G}_X)$ and $(\mathcal{Y}, \mathcal{G}_Y)$ are Polish spaces with their corresponding Borel σ-algebras. Let μ be the marginal on $\mathcal{X}$ and $(P_x)_{x\in\mathcal{X}}$ be a disintegration of P into conditional distributions. Then for all P-integrable functions $g : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ we have

$$\mathbb{E}\bigl(g(X,Y) \mid X = x\bigr) = \int_{\mathcal{Y}} g(x,y)\, dP_x(y), \tag{S40}$$

for μ almost every $x \in \mathcal{X}$. Indeed, by Lemma S34 we see that $x \mapsto \int_{\mathcal{Y}} g(x,y)\, dP_x(y)$ is a μ-integrable function, and hence $\mathcal{G}_X$-measurable. Moreover, given any $A \in \mathcal{G}_X$, we have

$$\begin{aligned}
\int_A \biggl(\int_{\mathcal{Y}} g(x,y)\, dP_x(y) \biggr) d\mu(x) &= \int_{\mathcal{X}} \biggl(\int_{\mathcal{Y}} \mathbb{1}_A(x) \cdot g(x,y)\, dP_x(y) \biggr) d\mu(x) \\
&= \int_{\mathcal{X}\times\mathcal{Y}} \mathbb{1}_A(x) \cdot g(x,y)\, dP(x,y) = \int_{A\times\mathcal{Y}} g(x,y)\, dP(x,y),
\end{aligned}$$

where the second equality follows from (S39) with $\mathbb{1}_A(x) \cdot g(x,y)$ in place of $g(x,y)$.

Recall that for Borel subsets $B_0, B_1 \subseteq \mathbb{R}^d$ and a Borel measure μ on $\mathbb{R}^d$, we write $B_0 \subseteq B_1$ if $\mu(B_0 \setminus B_1) = 0$ and $B_0 \nsubseteq B_1$ if $\mu(B_0 \setminus B_1) > 0$.

LEMMA S35. *Suppose that η_0 and $\eta_1 : \mathbb{R}^d \to [0,1]$ are regression functions for a Borel probability distribution P on $\mathbb{R}^d \times [0,1]$ with marginal probability distribution μ. Then, $\mu(\{x \in \mathbb{R}^d : \eta_0(x) \neq \eta_1(x)\}) = 0$. Hence, given $A \in \mathcal{B}(\mathbb{R}^d)$, we have $A \subseteq \mathcal{X}_\tau(\eta_0)$ if and only if $A \subseteq \mathcal{X}_\tau(\eta_1)$.*

PROOF. Given $\epsilon > 0$, let $B_\epsilon := \{x \in \mathbb{R}^d : \eta_0(x) \geq \eta_1(x) + \epsilon\}$. Then,

$$\epsilon \cdot \mu(B_\epsilon) \leq \int_{B_\epsilon} \{\eta_0(x) - \eta_1(x)\}\, d\mu = \int_{B_\epsilon \times [0,1]} y\, dP(x,y) - \int_{B_\epsilon \times [0,1]} y\, dP(x,y) = 0,$$

so $\mu(B_\epsilon) = 0$. By taking a countable union we see that $\mu\bigl(\{x \in \mathbb{R}^d : \eta_0(x) > \eta_1(x)\}\bigr) = 0$, so by symmetry we have $\mu(\{x \in \mathbb{R}^d : \eta_0(x) \neq \eta_1(x)\}) = 0$. Thus, given $A \in \mathcal{B}(\mathbb{R}^d)$ we have $\mu(A \setminus \mathcal{X}_\tau(\eta_0)) = \mu(A \setminus \mathcal{X}_\tau(\eta_1))$, so $A \subseteq \mathcal{X}_\tau(\eta_0)$ if and only if $A \subseteq \mathcal{X}_\tau(\eta_1)$. □

S8.2. *Concentration results.* We will require the following classic result that gives a uniform concentration inequality over classes of finite Vapnik–Chervonenkis dimension; we state it for distributions on $\mathbb{R}^d$ for simplicity.

LEMMA S36 (Vapnik–Chervonenkis concentration). *Let μ be a probability distribution on $\mathbb{R}^d$, and let $X_1, \ldots, X_n \overset{\mathrm{iid}}{\sim} \mu$, with corresponding empirical distribution $\hat{\mu}_n$. There exists a universal constant $C_{\mathrm{VC}} > 0$ such that for any collection of sets $\mathcal{S} \subseteq \mathcal{B}(\mathbb{R}^d)$ with $1 \leq \dim_{\mathrm{VC}}(\mathcal{S}) < \infty$, we have*

$$\mathbb{E}\biggl(\sup_{S \in \mathcal{S}} \bigl|\hat{\mu}_n(S) - \mu(S)\bigr|\biggr) \leq C_{\mathrm{VC}} \sqrt{\frac{\dim_{\mathrm{VC}}(\mathcal{S})}{n}}.$$

Moreover, for all $\delta \in (0,1)$ we have

$$\mathbb{P}\biggl(\sup_{S \in \mathcal{S}} \bigl|\hat{\mu}_n(S) - \mu(S)\bigr| > C_{\mathrm{VC}} \sqrt{\frac{\dim_{\mathrm{VC}}(\mathcal{S})}{n}} + \sqrt{\frac{\log(1/\delta)}{2n}}\biggr) \leq \delta.$$

PROOF. For the expectation bound, see Vershynin (2018, Theorem 8.3.23). The high-probability bound follows by McDiarmid's inequality (Vershynin, 2018, Theorem 2.9.1). □

The following lemma is used in the proof of Lemma S38.

LEMMA S37 (Garivier and Cappé (2011)). *Let $(Z_j)_{j \in [m]}$ be independent random variables taking values in $[0,1]$ with $\max_{j \in [m]} \mathbb{E}[Z_j] \leq t$ for some $t \in (0,1)$. Writing $\bar{Z} := m^{-1} \sum_{j \in [m]} Z_j$, we have for $\kappa \in (t,1)$ that*

$$\mathbb{P}(\bar{Z} \geq \kappa) \leq e^{-m \cdot \mathrm{kl}(\kappa, t)}.$$

PROOF. By Jensen's inequality, for $\theta > 0$ and $j \in [m]$,

$$\mathbb{E}(e^{\theta \cdot Z_j}) \leq 1 - \mathbb{E}(Z_j) + e^\theta \cdot \mathbb{E}(Z_j) \leq 1 + t(e^\theta - 1).$$

Hence, by Markov's inequality,

$$\mathbb{P}(\bar{Z} \geq \kappa) \leq e^{-m\theta\kappa} \prod_{j=1}^m \mathbb{E}(e^{\theta \cdot Z_j}) \leq \bigl[e^{-\theta\kappa}\bigl\{1 + t(e^\theta - 1)\bigr\}\bigr]^m.$$

The lemma follows on taking $\theta = \log\bigl(\frac{\kappa(1-t)}{t(1-\kappa)}\bigr) > 0$. □

LEMMA S38. *Let $(Z_j)_{j \in [m]}$ be independent random variables taking values in $[0,1]$ with $\max_{j \in [m]} \mathbb{E}(Z_j) \leq t$ for some $t \in (0,1)$. Let $\bar{Z} := m^{-1} \sum_{j \in [m]} Z_j$. Then for every $\alpha \in (0,1)$, we have*

$$\mathbb{P}\biggl(\bar{Z} \geq t + \sqrt{\frac{\log(1/\alpha)}{2m}}\biggr) \leq \mathbb{P}\biggl(\biggl\{\mathrm{kl}(\bar{Z}, t) \geq \frac{\log(1/\alpha)}{m}\biggr\} \bigcap \{\bar{Z} > t\}\biggr) \leq \alpha.$$

PROOF. The first inequality follows from the fact that

$$2(\bar{Z}-t)^2 = 2\mathrm{TV}^2\big(\mathrm{Bern}(\bar{Z}),\mathrm{Bern}(t)\big) \le \mathrm{kl}(\bar{Z},t),$$

by Pinsker's inequality. To prove the second inequality we begin by noting that $w \mapsto \mathrm{kl}(w,t)$ is continuous and strictly increasing on the interval $[t,1]$, and consider two cases. If $\alpha \in \big(0, e^{-m\cdot \mathrm{kl}(1,t)}\big)$ and $\bar{Z} > t$, then

$$\mathrm{kl}(\bar{Z},t) \le \mathrm{kl}(1,t) < \frac{\log(1/\alpha)}{m}.$$

On the other hand, if $\alpha \in \big[e^{-m\cdot\mathrm{kl}(1,t)},1\big)$, then by the intermediate value theorem we can find $\kappa_\alpha \in [t,1]$ such that $\mathrm{kl}(\kappa_\alpha,t) = m^{-1}\cdot\log(1/\alpha)$. Then by Lemma S37,

$$\mathbb{P}\bigg(\bigg\{\mathrm{kl}(\bar{Z},t) \ge \frac{\log(1/\alpha)}{m}\bigg\}\bigcap\{\bar{Z}>t\}\bigg) = \mathbb{P}\big(\bar{Z}\ge\kappa_\alpha\big) \le e^{-m\cdot\mathrm{kl}(\kappa_\alpha,t)} = \alpha,$$

as required. □

In addition we shall make use of the following Chernoff bounds.

LEMMA S39 (Multiplicative Chernoff — Theorem 2.3(b,c) of McDiarmid (1998)). *Let $(Z_j)_{j\in[m]}$ be a sequence of independent random variables taking values in $[0,1]$. Then given any $\theta > 0$,*

$$\mathbb{P}\bigg(\sum_{j=1}^m Z_j \le (1-\theta)\cdot\sum_{j=1}^m \mathbb{E}(Z_j)\bigg) \le \exp\bigg(-\frac{\theta^2}{2}\cdot\sum_{j=1}^m\mathbb{E}(Z_j)\bigg)$$

$$\mathbb{P}\bigg(\sum_{j=1}^m Z_j \ge (1+\theta)\cdot\sum_{j=1}^m \mathbb{E}(Z_j)\bigg) \le \exp\bigg(-\frac{\theta^2}{2(1+\theta/3)}\cdot\sum_{j=1}^m\mathbb{E}(Z_j)\bigg).$$

LEMMA S40 (Multiplicative matrix Chernoff – Theorem 1.1 of Tropp (2012)). *Let $(\mathbf{Z}_j)_{j\in[m]}$ be independent, non-negative definite $q\times q$ matrices with $\lambda_{\max}(\mathbf{Z}_j) \le a_{\max}$ almost surely, for every $j\in[m]$. Then, writing $a_{\min} := m^{-1}\cdot\lambda_{\min}\big(\sum_{j\in[m]}\mathbb{E}\mathbf{Z}_j\big)$, we have for every $\theta\in[0,1]$ that*

$$\mathbb{P}\bigg\{\lambda_{\min}\bigg(\sum_{j=1}^m \mathbf{Z}_j\bigg) \le m(1-\theta)a_{\min}\bigg\} \le q\cdot\bigg(\frac{e^{-\theta}}{(1-\theta)^{1-\theta}}\bigg)^{ma_{\min}/a_{\max}} \le q\cdot e^{-\frac{\theta^2 m a_{\min}}{2a_{\max}}}.$$

S8.3. *Useful lemmas for the lower bounds.* We shall make use of the following result from Brown and Low (1996).

LEMMA S41 (Brown–Low constrained risk inequality). *Let $\mathbb{Q}_1$, $\mathbb{Q}_2$ be probability measures on a measurable space $(\Omega,\mathcal{F})$ such that $\mathbb{Q}_2$ is absolutely continuous with respect to $\mathbb{Q}_1$, and assume that*

$$I := \chi^2(\mathbb{Q}_2,\mathbb{Q}_1) + 1 < \infty.$$

Let $\epsilon\in\big(0,I^{-1/2}\big)$ and let $Z:\Omega\to\{1,2\}$ be a $\mathcal{F}$-measurable random variable with $\mathbb{Q}_1(Z=2)\le\epsilon^2$. Then $\mathbb{Q}_2(Z=1)\ge\big(1-\epsilon\sqrt{I}\big)^2$.

The following version of Fano's lemma is a minor variant of Gerchinovitz, Ménard and Stoltz (2020, Lemma 4.3).

LEMMA S42 (Fano's lemma for χ^2 divergences). *Let $\mathbb{P}_1, \ldots, \mathbb{P}_M, \mathbb{Q}_1, \ldots, \mathbb{Q}_M$ denote probability measures on $(\Omega, \mathcal{F})$, and let $A_1, \ldots, A_M \in \mathcal{A}$. Write $\bar{p} := M^{-1} \sum_{j=1}^M \mathbb{P}_j(A_j)$ and $\bar{q} := M^{-1} \sum_{j=1}^M \mathbb{Q}_j(A_j)$. If $\bar{q} \in (0,1)$, then*

$$\bar{p} \leq \bar{q} + \sqrt{\frac{1}{M} \sum_{j=1}^M \chi^2(\mathbb{P}_j, \mathbb{Q}_j) \cdot \bar{q}(1-\bar{q})}.$$

In particular, if $M \geq 2$ and $A_1, \ldots, A_M$ form a partition of Ω, then

$$\frac{1}{M} \sum_{j=1}^M \mathbb{P}_j(A_j) \leq \frac{1}{M} + \sqrt{\inf_{\mathbb{Q} \in \mathcal{Q}} \frac{1}{M} \sum_{j=1}^M \chi^2(\mathbb{P}_j, \mathbb{Q}) \cdot \frac{1}{M}\left(1 - \frac{1}{M}\right)},$$

where $\mathcal{Q}$ denotes the set of all probability distributions on Ω.

PROOF. By the joint convexity of χ^2 divergence, together with the data processing inequality (e.g. Gerchinovitz, Ménard and Stoltz, 2020, Lemma 2.1), we have

$$\frac{(\bar{p}-\bar{q})^2}{\bar{q}(1-\bar{q})} = \chi^2\bigl(\mathrm{Bern}(\bar{p}), \mathrm{Bern}(\bar{q})\bigr) \leq \frac{1}{M} \sum_{j=1}^M \chi^2\bigl(\mathbb{P}_j(A_j), \mathbb{Q}_j(A_j)\bigr) \leq \frac{1}{M} \sum_{j=1}^M \chi^2(\mathbb{P}_j, \mathbb{Q}_j).$$

The first result follows on rearranging this inequality, and the second follows by taking $\mathbb{Q}_1 = \cdots = \mathbb{Q}_M = \mathbb{Q}$ and then taking an infimum over $\mathbb{Q} \in \mathcal{Q}$. □

REFERENCES

BROWN, L. D. and LOW, M. G. (1996). A constrained risk inequality with applications to nonparametric functional estimation. *Annals of Statistics* **24** 2524–2535.

CANNINGS, T. I., BERRETT, T. B. and SAMWORTH, R. J. (2020). Local nearest neighbour classification with applications to semi-supervised learning. *Annals of Statistics* **48** 1789–1814.

CONSTANTINE, G. and SAVITS, T. (1996). A multivariate Faa di Bruno formula with applications. *Transactions of the American Mathematical Society* **348** 503–520.

DUDLEY, R. M. (2018). *Real Analysis and Probability*. CRC Press, Cambridge.

GARIVIER, A. and CAPPÉ, O. (2011). The KL-UCB algorithm for bounded stochastic bandits and beyond. In *Proceedings of the 24th annual Conference On Learning Theory* 359–376.

GERCHINOVITZ, S., MÉNARD, P. and STOLTZ, G. (2020). Fano's inequality for random variables. *Statistical Science* **35** 178–201.

MCDIARMID, C. (1998). Concentration. In *Probabilistic Methods for Algorithmic Discrete Mathematics* 195–248. Springer.

MCSHANE, E. J. (1934). Extension of range of functions. *Bulletin of the American Mathematical Society* **40** 837–842.

OKAMOTO, M. (1973). Distinctness of the Eigenvalues of a Quadratic form in a Multivariate Sample. *Annals of Statistics* **1** 763–765. https://doi.org/10.1214/aos/1176342472

REEVE, H. W. J., CANNINGS, T. I. and SAMWORTH, R. J. (2021). Adaptive Transfer Learning. *Annals of Statistics* **49** 3618-3649.

REEVE, H. W. J., CANNINGS, T. I. and SAMWORTH, R. J. (2023). Optimal subgroup selection. *Submitted*.

TROPP, J. A. (2012). User-Friendly Tail Bounds for Sums of Random Matrices. *Foundations of Computational Mathematics* **12** 389–434. https://doi.org/10.1007/s10208-011-9099-z

VERSHYNIN, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science* **47**. Cambridge University Press, Cambridge.