A Appendix

To emphasize the underlying parameters of the NN, by some abuse of notation, we introduce

$$\mathcal{G}_k(\Theta) := \left\{ g : \mathbb{R}^d \to \mathbb{R} : g(x) = \sum_{i=1}^{k} \beta_i \phi\left(w_i \cdot x + b_i\right) + b_0,\ \left(\{\beta_i, w_i, b_i\}_{i=1}^k, b_0\right) \in \Theta \right\}, \tag{34a}$$

$$\Theta_k(\mathbf{a}) := \left\{ \left(\{\beta_i, w_i, b_i\}_{i=1}^k, b_0\right) : \begin{array}{l} w_i \in \mathbb{R}^d,\ b_0, b_i, \beta_i \in \mathbb{R},\ \max_{\substack{i=1,\ldots,k \\ j=1,\ldots,d}} \{|w_{i,j}|, |b_i|\} \le a_1 \\ |\beta_i| \le a_2,\quad i = 1, \ldots, k,\ |b_0| \le a_3 \end{array} \right\}. \tag{34b}$$

Also, throughout the Appendix, we denote $g(x) = \sum_{i=1}^{k} \beta_i \phi\left(w_i \cdot x + b_i\right) + b_0$ for $\theta = \left(\{\beta_i, w_i, b_i\}_{i=1}^k, b_0\right)$ by g_θ, whenever the underlying θ needs to be emphasized.

We first state an auxiliary result which will be useful in the proofs that follow. For $b \ge 0$, an integer $l \ge 0$, consider the function class $\mathcal{S}_{l,b}(\mathbb{R}^d)$ defined below:

$$\mathcal{S}_{l,b}(\mathbb{R}^d) := \left\{ f \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d) : \begin{array}{l} |f(0)| \le b,\ D^{\boldsymbol{\alpha}} f \text{ exists Lebesgue a.e. on } \mathbb{R}^d\ \forall \boldsymbol{\alpha} \text{ s.t. } |\boldsymbol{\alpha}| = l, \\ \|D^{\boldsymbol{\alpha}} f\|_{L^j(\mathbb{R}^d)} \le b \text{ for } j = 1, 2,\ |\boldsymbol{\alpha}| \in \{1, l\} \end{array} \right\}. \tag{35}$$

The following lemma states that functions in $\mathcal{S}_{l,b}(\mathbb{R}^d)$ with sufficient smoothness order l belong to the Barron class. Its proof essentially follows using arguments from Barron (1993), where it was mentioned without explicit quantification. Below, we provide a proof for completeness.

Lemma 1 (Smoothness and Barron class)**.** *If $f \in \mathcal{S}_{s,b}\left(\mathbb{R}^d\right)$ for $s := \lfloor \frac{d}{2} \rfloor + 2$, then we have*

$$B(f) \le b\kappa_d \sqrt{d}, \tag{36a}$$

$$\kappa_d^2 := (d + d^s) \int_{\mathbb{R}^d} \left(1 + \|\omega\|^{2(s-1)}\right)^{-1} \mathsf{d}\omega < \infty. \tag{36b}$$

Consequently, $\mathcal{S}_{s,b}\left(\mathbb{R}^d\right) \subseteq \mathcal{B}_{b\kappa_d\sqrt{d} \vee b}$.

Proof. Since $f \in L^1\left(\mathbb{R}^d\right)$, its Fourier transform $\hat{f}$ is well-defined. Also,

$$\begin{aligned} \int_{\mathbb{R}^d} \left|\hat{f}(\omega)\right| d\omega &\overset{(a)}{\le} \left(\int_{\mathbb{R}^d} \frac{d\omega}{1 + \|\omega\|^{2s}}\right)^{\frac{1}{2}} \left(\int_{\mathbb{R}^d} \left(1 + \|\omega\|^{2s}\right) \left|\hat{f}(\omega)\right|^2 d\omega\right)^{\frac{1}{2}} \\ &\overset{(b)}{\le} \left(\int_{\mathbb{R}^d} \frac{d\omega}{1 + \|\omega\|^{2s}}\right)^{\frac{1}{2}} \left(\|f\|_{L^2(\mathbb{R}^d)}^2 + \sum_{\boldsymbol{\alpha}: |\boldsymbol{\alpha}| = s} \|D^{\boldsymbol{\alpha}} f\|_{L^2(\mathbb{R}^d)}^2\right)^{\frac{1}{2}} < \infty, \end{aligned}$$

where

(a) follows from Cauchy-Schwarz inequality;

(b) is by Plancherel's theorem and definition of $\mathcal{S}_{l,b}(\mathbb{R}^d)$.

Hence, $\hat{f} \in L^1\left(\mathbb{R}^d\right)$ and the Fourier inversion formula holds with $\tilde{F}(d\omega) = \hat{f}(\omega)d\omega$. Then, it follows that

$$B(f) = \int_{\mathbb{R}^d} \sup_{x \in \mathcal{X}} |\omega \cdot x| \left|\hat{f}(\omega)\right| d\omega \le \sqrt{d} \int_{\mathbb{R}^d} \|\omega\| \left|\hat{f}(\omega)\right| d\omega, \tag{37}$$

where we used $\sup_{x \in \mathcal{X}} |\omega \cdot x| \le \sqrt{d}\, \|\omega\|$ which holds by Cauchy-Schwarz inequality.

Next, recall that if the partial derivatives $D^{\boldsymbol{\alpha}} f$, $|\boldsymbol{\alpha}| = s$, exists on $\mathbb{R}^d$, then all partial derivatives $D^{\boldsymbol{\alpha}} f$, $0 \le |\boldsymbol{\alpha}| \le s$, also exists. Hence, if $\|D^{\boldsymbol{\alpha}} f\|_{L^2(\mathbb{R}^d)} \le b$ for all α with $|\alpha| \in \{1, s\}$, we have

$$\int_{\mathbb{R}^d} \|\omega\| \left|\hat{f}(\omega)\right| d\omega \overset{(a)}{\le} \left(\int_{\mathbb{R}^d} \frac{d\omega}{1 + \|\omega\|^{2(s-1)}}\right)^{\frac{1}{2}} \left(\int_{\mathbb{R}^d} \left(\|\omega\|^2 + \|\omega\|^{2s}\right) \left|\hat{f}(\omega)\right|^2 d\omega\right)^{\frac{1}{2}} \tag{38}$$

$$\overset{(b)}{\leq} \left(\int_{\mathbb{R}^d} \frac{d\omega}{1 + \|\omega\|^{2(s-1)}} \right)^{\frac{1}{2}} \left(\sum_{\boldsymbol{\alpha}:|\boldsymbol{\alpha}|=1} \|D^{\boldsymbol{\alpha}} f\|^2_{L^2(\mathbb{R}^d)} + \sum_{\boldsymbol{\alpha}:|\boldsymbol{\alpha}|=s} \|D^{\boldsymbol{\alpha}} f\|^2_{L^2(\mathbb{R}^d)} \right)^{\frac{1}{2}}$$

$$\overset{(c)}{\leq} \kappa_d b, \tag{39}$$

where

(a) follows from Cauchy-Schwarz inequality;

(b) is due to Plancherel's theorem;

(c) follows since $|\{\boldsymbol{\alpha} : |\boldsymbol{\alpha}| = s\}| = d^s$ and $\|D^{\boldsymbol{\alpha}} f\|_{L^2(\mathbb{R}^d)} \leq b$.

Combining (37) and (39) leads to (36a). The final claim follows from (5) and (36a) by noting that $|f(0)| \leq b$ by definition. ☐

A.1 Proof of Theorem 2

The proof relies on arguments from Barron (1992) and Barron (1993), along with the uniform central limit theorem for uniformly bounded VC function classes. Fix an arbitrary (small) $\delta > 0$, and let $f : \mathbb{R}^d \to \mathbb{R}$ be such that $\tilde{f} = f|_{\mathcal{X}}$ and $B(f) \vee f(0) \leq c + \delta$. This is possible since $c^\star_B(\tilde{f}) \leq c$. Then, it follows from the proof of Barron (1993, Theorem 2) that

$$f_0(x) := f(x) - f(0) = \int_{\omega \in \mathbb{R}^d \setminus \{0\}} \varrho(x, \omega) \mu(d\omega),$$

where

$$\varrho(x, \omega) = \frac{B(f)}{\sup_{x \in \mathcal{X}} |\omega \cdot x|} \left(\cos(\omega \cdot x + \zeta(\omega)) - \cos(\zeta(\omega)) \right),$$
$$B(f) := \int_{\mathbb{R}^d} \sup_{x \in \mathcal{X}} |\omega \cdot x| \, F(d\omega),$$
$$\mu(d\omega) = \frac{\sup_{x \in \mathcal{X}} |\omega \cdot x| \, F(d\omega)}{B(f)},$$

and $\zeta : \mathbb{R}^d \to \mathbb{R}$. Note that $\mu \in \mathcal{P}(\mathbb{R}^d)$ is a probability measure.

Let $\tilde{\Theta}_1 (k, B(f)) := \Theta_1(\sqrt{k} \log k, 2B(f), 0)$ (see (34b)). Then, it further follows from the proofs[4] of Barron (1993, Lemma 2-Lemma 4,Theorem 3) that there exists a probability measure $\mu_k \in \mathcal{P}\left(\tilde{\Theta}_1 (k, B(f))\right)$ (see Barron (1993, Eqns. (28)-(32))) such that

$$\left\| f_0 - \int_{\theta \in \tilde{\Theta}_1(k, B(f))} g_\theta(\cdot) \; \mu_k \,(d\theta) \right\|_{\infty, P, Q} \leq \frac{2(B(f) + 1)}{\sqrt{k}}, \tag{40}$$

where $g_\theta(x) = \beta \phi \,(w \cdot x + b)$ for $\theta = (\beta, w, b)$. Note that $\int_{\tilde{\Theta}_1(k, B(f))} \mu_k(d\theta) = 1 < \infty$.

Next, for each fixed x, let $\upsilon_x : \tilde{\Theta}_1 (k, B(f)) \to \mathbb{R}$ be given by $\upsilon_x(\theta) := g_\theta(x)$, and consider the function class $\mathcal{V}_k \left(\tilde{\Theta}_1 (k, B(f))\right) = \left\{\upsilon_x, \; x \in \mathbb{R}^d\right\}$. Note that every $\upsilon_x \in \mathcal{V}_k \left(\tilde{\Theta}_1 (k, B(f))\right)$ is a composition of an affine function in θ with the bounded monotonic function $\beta\phi(\cdot)$. Hence, noting that $\mathcal{V}_k \left(\tilde{\Theta}_1 (k, B(f))\right)$ is a VC function class

[4] The claims in Barron (1993, Lemma 2- Lemma 4, Theorem 3) are stated for L^2 norm, but it is not hard to see from the proof therein that the same also holds for L^∞ norm, apart from the following subtlety. In the proof of Lemma 3, it is shown that $\varrho(x, \omega)$, $\omega \in \mathbb{R}^d$, lies in the convex closure of a certain class of step functions, whose discontinuity points are adjusted to coincide with the continuity points of the underlying measure μ. Similarly, here, the step discontinuities needs to be adjusted to coincide with the continuity points of both P and Q. Nevertheless, the same arguments hold since the common continuity points of P and Q form a dense set.

(Van Der Vaart and Wellner (1996)), it follows from Van Der Vaart and Wellner (1996, Theorem 2.8.3) that it is a uniform Donsker class (in particular, μ_k-Donsker) for all probability measures $\mu \in \mathcal{P}\left(\tilde{\Theta}_1\left(k, B(f)\right)\right)$. Furthermore, an application of Van Der Vaart and Wellner (1996, Corollary 2.2.8)) yields that there exists k parameter vectors, $\theta_i := (\beta_i, w_i, b_i) \in \tilde{\Theta}_1\left(k, B(f)\right),\ 1 \leq i \leq k$, such that (see also Yukich et al. (1995, Theorem 2.1))

$$\sup_{x \in \mathbb{R}^d} \left| \int_{\theta \in \tilde{\Theta}_1(k, B(f))} g_\theta(x)\ \mu_k(d\theta) - \frac{1}{k}\sum_{i=1}^{k} g_{\theta_i}(x) \right| \leq \hat{c}_d B(f) k^{-\frac{1}{2}}, \tag{41}$$

where $\hat{c}_d$ is a constant which depends only on d. Note that the R.H.S. of (41) is independent of μ_k and depends on f and $\mathcal{X}$ only via $B(f)$.

From (40), (41) and triangle inequality, we obtain

$$\left\| f_0 - \frac{1}{k}\sum_{i=1}^{k} g_{\theta_i} \right\|_{\infty, P, Q} \leq \left(\hat{c}_d B(f) + 2B(f) + 2\right) k^{-\frac{1}{2}}.$$

Setting $\theta = \left\{ \left\{ \left(\frac{\beta_i}{k}, w_i, b_i\right) \right\}_{i=1}^{k}, f(0) \right\}$ and $g_\theta(x) = f(0) + \frac{1}{k}\sum_{i=1}^{k} g_{\theta_i}(x)$, we have

$$\|f - g_\theta\|_{\infty,P,Q} \leq \left((\hat{c}_d + 2)B(f) + 2\right) k^{-\frac{1}{2}} \leq \left((\hat{c}_d + 2)(c + \delta) + 2\right) k^{-\frac{1}{2}}.$$

Next, note that $\left\| \tilde{f} - g_\theta \right\|_{\infty,P,Q} = \|f - g_\theta\|_{\infty,P,Q}$ and $g_\theta \in \mathcal{G}_k^*\left(B(f) \vee f(0)\right) \subseteq \mathcal{G}_k^*\left(c + \delta\right)$. Since $\delta > 0$ is arbitrary, we obtain that there exists $g_\theta \in \mathcal{G}_k^*\left(c\right)$

$$\left\| \tilde{f} - g_\theta \right\|_{\infty,P,Q} \leq \left((\hat{c}_d + 2)c + 2\right) k^{-\frac{1}{2}} =: \tilde{C}_{d,c}\ k^{-\frac{1}{2}}, \tag{42}$$

thus proving the claim in (12).

On the other hand, it follows similar to (38) in Lemma 1 that for a fixed $\epsilon > 0$ and $l(\epsilon) = d/2 + 1 + \epsilon$, the set of functions $f \in \mathbb{R}^d \to \mathbb{R}$ such that $B(f) \leq c$ includes those whose Fourier transform $\hat{f}(\omega)$ satisfies

$$\int_{\mathbb{R}^d} \left(\|\omega\|^2 + \|\omega\|^{2l(\epsilon)} \right) \left| \hat{f}(\omega) \right|^2 d\omega \leq c^2 d^{-1} \left(\int_{\mathbb{R}^d} \frac{d\omega}{1 + \|\omega\|^{2(l(\epsilon)-1)}} \right)^{-1}, \tag{43}$$

since $\int_{\mathbb{R}^d} \frac{d\omega}{1+\|\omega\|^{2(l(\epsilon)-1)}} < \infty$. Then, (13) follows from the proof of Barron (1992)[Theorem 3]. Note from the proof therein that the constant in (13) may in general depend on d and ϵ.

A.2 Proof of Corollary 1

By Theorem 2, it suffices to show that there exists an extension f_{e} of f from $\mathcal{U}$ to $\mathbb{R}^d$ such that $B(f_{\mathsf{e}}) \vee f_{\mathsf{e}}(0) \leq \bar{c}_{b,c,d}$. Let $\boldsymbol{\alpha}_j$ denote a multi-index of order j, and recall that $s := \lfloor \frac{d}{2} \rfloor + 2$. Consider an extension of $D^{\boldsymbol{\alpha}_s} f$ from $\mathcal{U}$ to $\mathbb{R}^d$ for each $\boldsymbol{\alpha}_s$ as follows:

$$D^{\boldsymbol{\alpha}_s} f(x) := \inf_{x' \in \mathcal{U}} D^{\boldsymbol{\alpha}_s} f(x') + c\left\|x - x'\right\|^{\delta},\ x \in \mathbb{R}^d \setminus \mathcal{U}. \tag{44}$$

Note that $D^{\boldsymbol{\alpha}_s} f$ extended this way is Hölder continuous with the same constant c and exponent δ on $\mathbb{R}^d$. Fixing $D^{\boldsymbol{\alpha}_s} f$ on $\mathbb{R}^d$ induces an extension of all lower (and also higher) order derivatives $D^{\boldsymbol{\alpha}_j} f,\ 0 \leq j < s$ to $\mathbb{R}^d$, which can be defined recursively as $D^{\boldsymbol{\alpha}_1} D^{\boldsymbol{\alpha}_{s-j}} f(x) = D^{\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_{s-j}} f(x)$, $x \in \mathbb{R}^d$, for all $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_{s-j}$ and $j = 1, \ldots, s$.

Let $\mathcal{U}' := \{x' \in \mathbb{R}^d : \|x' - x\| < 1 \text{ for some } x \in \mathcal{X}\}$. Suppose $\mathcal{U} \subset \mathcal{U}'$. By the mean value theorem, we have for any $x, x' \in \mathcal{U}'$ and $j = 1, \ldots, s$,

$$|D^{\boldsymbol{\alpha}_{s-j}} f(x')| \leq |D^{\boldsymbol{\alpha}_{s-j}} f(x)| + \max_{\substack{\tilde{x} \in \mathcal{U}', \\ \boldsymbol{\alpha}_1}} \left| D^{\boldsymbol{\alpha}_{s-j} + \boldsymbol{\alpha}_1} f(\tilde{x}) \right| \|x - x'\|_1$$

$$\leq \left|D^{\boldsymbol{\alpha}_{s-j}} f(x)\right| + \max_{\substack{\tilde{x}\in\mathcal{U}',\\ \boldsymbol{\alpha}_1}} \left|D^{\boldsymbol{\alpha}_{s-j}+\boldsymbol{\alpha}_1} f(\tilde{x})\right| \sqrt{d}\left\|x-x'\right\|, \tag{45}$$

where the last step follows from $\|x-x'\|_1 \leq \sqrt{d}\,\|x-x'\|$. Also, note from (44) that $D^{\boldsymbol{\alpha}_s} f(x) < b+c$ for all $x \in \mathcal{U}'$, and recall that since $f \in \mathcal{H}^{s,\delta}_{b,c}(\mathcal{U})$, we have $|D^{\boldsymbol{\alpha}_{s-j}} f(x)| \leq b$ for all $x \in \mathcal{U}$. Then, for any $x' \in \mathcal{U}'$, taking $x \in \mathcal{X}$ satisfying $\|x-x'\| \leq 1$ (such an x exists by definition of $\mathcal{U}'$) in (45) yields

$$|D^{\boldsymbol{\alpha}_{s-1}} f(x')| \leq b + (b+c)\sqrt{d}. \tag{46}$$

Starting from (46) and recursively applying (45), we obtain for $j = 1, \ldots, s$, and $x' \in \mathcal{U}'$,

$$|D^{\boldsymbol{\alpha}_{s-j}} f(x')| \leq b\sum_{i=1}^{j} d^{\frac{i-1}{2}} + (b+c)d^{\frac{j}{2}} \leq b\frac{1-d^{\frac{s}{2}}}{1-\sqrt{d}} + (b+c)d^{\frac{s}{2}} =: \tilde{b}. \tag{47}$$

Thus, the extension f from $\mathcal{U}$ to $\mathbb{R}^d$ satisfies $f|_{\mathcal{U}'} \in \mathcal{H}^{s,\delta}_{\tilde{b},c}(\mathcal{U}')$. If $\mathcal{U}' \subseteq \mathcal{U}$, then $f|_{\mathcal{U}'} \in \mathcal{H}^{s,\delta}_{b,c}(\mathcal{U}')$ by definition, and thus, in either case, $f|_{\mathcal{U}'} \in \mathcal{H}^{s,\delta}_{\tilde{b},c}(\mathcal{U}')$.

The desired final extension is $f_{\mathsf{e}} : \mathbb{R}^d \to \mathbb{R}$ given by $f_{\mathsf{e}}(x) := f(x)\cdot f_{\mathsf{C}}(x)$, where

$$f_{\mathsf{C}}(x) := \mathbb{1}_{\mathcal{X}'} * \psi_{\frac{1}{2}}(x) := \int_{\mathbb{R}^d} \mathbb{1}_{\mathcal{X}'}(y)\psi_{\frac{1}{2}}(x-y)dy,\ x \in \mathbb{R}^d, \tag{48}$$

$$\mathcal{X}' := \left\{x' \in \mathbb{R}^d : \|x'-x\| \leq 0.5 \text{ for some } x \in \mathcal{X}\right\},$$

$$\psi(x) := \begin{cases} u^{-1}e^{-\frac{1}{\frac{1}{2}-\|x\|^2}}, & \|x\| < \frac{1}{2},\\ 0, & \text{otherwise}, \end{cases} \tag{49}$$

and u is the normalization constant such that $\int_{\mathbb{R}^d}\psi(x)dx = 1$. Note that $\psi \in \mathsf{C}^\infty(\mathbb{R}^d)$, and consequently, $f_{\mathsf{C}} \in \mathsf{C}^\infty(\mathbb{R}^d)$ from (48) by dominated convergence theorem. Also, observe that $f_{\mathsf{C}}(x) = 1$ for $x \in \mathcal{X}$, $f_{\mathsf{C}}(x) = 0$ for $x \in \mathbb{R}^d \setminus \mathcal{U}'$ and $f_{\mathsf{C}}(x) \in (0,1)$ for $x \in \mathcal{U}' \setminus \mathcal{X}$. Hence, $f_{\mathsf{e}}(x) = f(x)$ for $x \in \mathcal{X}$, $f_{\mathsf{e}}(x) = 0$ for $x \in \mathbb{R}^d \setminus \mathcal{U}'$ and $|f_{\mathsf{e}}(x)| \leq |f(x)|$ for $x \in \mathcal{U}' \setminus \mathcal{X}$, thus satisfying $f_{\mathsf{e}}|_{\mathcal{X}} = f|_{\mathcal{X}} = \tilde{f}$ as required. Moroever, for all $j = 0, \ldots, s$,

$$|D^{\boldsymbol{\alpha}_j} f_{\mathsf{e}}(x)| \overset{(a)}{\leq} 2^j \tilde{b} \max_{\substack{x\in\mathcal{U}',\\ \boldsymbol{\alpha}:|\boldsymbol{\alpha}|\leq j}} |D^{\boldsymbol{\alpha}} f_{\mathsf{C}}(x)| \overset{(b)}{\leq} 2^s\tilde{b} \max_{\substack{x:\|x\|\leq 0.5,\\ \boldsymbol{\alpha}:|\boldsymbol{\alpha}|\leq s}} |D^{\boldsymbol{\alpha}}\psi(x)| =: \hat{b},\ x \in \mathcal{U}', \tag{50a}$$

$$D^{\boldsymbol{\alpha}_j} f_{\mathsf{e}}(x) = 0,\ x \notin \mathcal{U}', \tag{50b}$$

where

(a) follows using chain rule for differentiation and (47);

(b) follows from the definition in (48).

Then, we have for $j = 0, \ldots, s$ and $i = 1, 2$,

$$\begin{aligned}
\|D^{\boldsymbol{\alpha}_j} f_{\mathsf{e}}\|^i_{L^i(\mathbb{R}^d)} &= \int_{\mathbb{R}^d} (D^{\boldsymbol{\alpha}_j} f_{\mathsf{e}})^i(x)dx\\
&= \int_{\mathcal{U}'} (D^{\boldsymbol{\alpha}_j} f_{\mathsf{e}})^i(x)dx \leq \hat{b}^i\ \mathsf{Vol}_d(0.5\sqrt{d}+1)\\
&= \hat{b}^i\frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2}+1)}(0.5\sqrt{d}+1)^d,
\end{aligned} \tag{51}$$

where $\mathsf{Vol}_d(r)$ denotes the volume of a Euclidean ball in $\mathbb{R}^d$ with radius r and Γ denotes the gamma function. Defining $b' := \hat{b}\pi^{\frac{d}{2}}\Gamma^{-1}(\frac{d}{2}+1)(0.5\sqrt{d}+1)^d$ and noting that $b' \geq \hat{b}$, we have from (50) and (51) that $f_{\mathsf{e}}(x) \in \tilde{\mathcal{S}}_{s,b'}(\mathbb{R}^d)$, where

$$\tilde{\mathcal{S}}_{s,b'}(\mathbb{R}^d) := \left\{ f \in L^1(\mathbb{R}^d)\cap L^2(\mathbb{R}^d) : \begin{array}{l} |f(0)| \leq b',\ D^{\boldsymbol{\alpha}} f \text{ exists Lebesgue a.e. on } \mathbb{R}^d\ \forall \boldsymbol{\alpha} \text{ s.t. } |\boldsymbol{\alpha}| = s,\\ \|D^{\boldsymbol{\alpha}} f\|_{L^i(\mathbb{R}^d)} \leq b' \text{ for } i = 1,2,\ \ |\boldsymbol{\alpha}| = 1,\ldots,s \end{array} \right\}, \tag{52}$$

Observe that $\tilde{\mathcal{S}}_{s,b'}\left(\mathbb{R}^d\right) \subseteq \mathcal{S}_{s,b'}\left(\mathbb{R}^d\right)$ (see (35)). This implies via Lemma 1 that $B(f_{\mathsf{e}}) \leq c' := \kappa_d \sqrt{d} b'$ and

$$f_{\mathsf{e}} \in \mathcal{B}_{b' \vee c'} \cap \tilde{\mathcal{S}}_{s,b'}\left(\mathbb{R}^d\right) \subseteq \mathcal{B}_{b' \vee c'} \cap \mathcal{S}_{s,b'}\left(\mathbb{R}^d\right). \tag{53}$$

Then, by defining

$$\bar{c}_{b,c,d} := b' \vee c', \tag{54}$$

where

$$b' = \pi^{\frac{d}{2}} \Gamma^{-1}(0.5d+1)(0.5\sqrt{d}+1)^d 2^s \left(b \frac{1-d^{\frac{s}{2}}}{1-\sqrt{d}} + (b+c)d^{\frac{s}{2}}\right) \max_{\substack{x:\|x\| \leq 0.5, \\ \boldsymbol{\alpha}:|\boldsymbol{\alpha}| \leq s}} \psi^{(\boldsymbol{\alpha})}(x), \tag{55}$$

$$c' = \sqrt{d} \kappa_d b', \tag{56}$$
$$\kappa_d^2 = (d+d^s) \int_{\mathbb{R}^d} \left(1 + \|\omega\|^{2(s-1)}\right)^{-1} d\omega,$$

it follows from Theorem 2 (see (42)) that there exists $g \in \mathcal{G}_k^*\left(\bar{c}_{b,c,d}\right)$ such that

$$\left\|\tilde{f} - g\right\|_{\infty,P,Q} \leq \tilde{C}_{d,\bar{c}_{b,c,d}}\ k^{-\frac{1}{2}}. \tag{57}$$

This completes the proof.

A.3 Proof of Theorem 3

We will show that Theorem 3 holds with

$$V_{k,\mathbf{a},\gamma} := 4Ca_2^2 k R_{k,\mathbf{a},\gamma}^2, \tag{58}$$
$$E_{k,\mathbf{a},n,\gamma} := 2\sqrt{2} n^{-\frac{1}{2}} k a_2 R_{k,\mathbf{a},\gamma} = 4\sqrt{2} n^{-\frac{1}{2}} k^{3/2} a_2 \left(\bar{\gamma}'_{\mathcal{G}_k(\mathbf{a})} + 1\right), \tag{59}$$

where

$$R_{k,\mathbf{a},\gamma} := 2\left(\bar{\gamma}'_{\mathcal{G}_k(\mathbf{a})} + 1\right)\sqrt{k}, \tag{60}$$

and $\bar{\gamma}'_{\mathcal{G}_k(\mathbf{a})}$ is defined in (16). We have

$$\begin{aligned}
&\hat{\mathsf{H}}_{\gamma,\mathcal{G}_k(\mathbf{a})}(x^n, y^n) - \mathsf{H}_{\gamma,\mathcal{G}_k(\mathbf{a})}(P,Q) \\
&= \sup_{g_\theta \in \mathcal{G}_k(\mathbf{a})} \frac{1}{n}\sum_{i=1}^n g_\theta(x_i) - \frac{1}{n}\sum_{i=1}^n \gamma(g_\theta(y_i)) - \left(\sup_{g_\theta \in \mathcal{G}_k(\mathbf{a})} \mathbb{E}_P[g_\theta(X)] - \mathbb{E}_Q\left[\gamma(g_\theta(Y))\right]\right) \\
&\leq \sup_{g_\theta \in \mathcal{G}_k(\mathbf{a})} \frac{1}{n}\sum_{i=1}^n g_\theta(x_i) - \frac{1}{n}\sum_{i=1}^n \gamma(g_\theta(y_i)) - \mathbb{E}_P[g_\theta(X)] + \mathbb{E}_Q\left[\gamma(g_\theta(Y))\right].
\end{aligned} \tag{61}$$

Let

$$Z_\theta := \frac{1}{n}\sum_{i=1}^n g_\theta(X_i) - \frac{1}{n}\sum_{i=1}^n \gamma\left(g_\theta(Y_i)\right) - \mathbb{E}_P[g_\theta(X)] + \mathbb{E}_Q\left[\gamma(g_\theta(Y))\right]. \tag{62}$$

We have

$$\begin{aligned}
|Z_\theta - Z_{\theta'}| \leq \sum_{i=1}^n &\frac{1}{n}\left|g_\theta(X_i) - g_{\theta'}(X_i) - \mathbb{E}_P[g_\theta(X) - g_{\theta'}(X)]\right| \\
&+ \frac{1}{n}\left|\gamma(g_\theta(Y_i)) - \gamma(g_{\theta'}(Y_i)) - \mathbb{E}_Q\left[\gamma(g_\theta(Y)) - \gamma(g_{\theta'}(Y))\right]\right|.
\end{aligned} \tag{63}$$

Since $0 \leq \phi(x) \leq 1$ for all $x \in \mathbb{R}^d$, for any $x, x' \in \mathcal{X}$ and $\theta = \left(\{\beta_i, w_i, b_i\}_{i=1}^k, b_0\right), \theta' = \left(\{\beta_i', w_i', b_i'\}_{i=1}^k, b_0'\right) \in \Theta_k(\mathbf{a})$,

$$|g_\theta(x) - g_{\theta'}(x')| \leq \sum_{i=1}^k |\beta_i - \beta_i'| \leq \|\boldsymbol{\beta}(\theta) - \boldsymbol{\beta}(\theta')\|_1, \tag{64}$$

where $\boldsymbol{\beta}(\theta) := (\beta_1, \ldots, \beta_k)$. Moreover, an application of the mean value theorem yields that for all $\theta, \theta' \in \Theta_k(\mathbf{a})$,

$$|\gamma(g_\theta(x)) - \gamma(g_{\theta'}(x'))| \le \bar{\gamma}'_{\mathcal{G}_k(\mathbf{a})} |g_\theta(x) - g_{\theta'}(x')| \le \bar{\gamma}'_{\mathcal{G}_k(\mathbf{a})} \|\boldsymbol{\beta}(\theta) - \boldsymbol{\beta}(\theta')\|_1 , \tag{65}$$

where $\bar{\gamma}'_{\mathcal{G}_k(\mathbf{a})}$ is defined in (16). Hence, with probability one

$$\begin{aligned} &\frac{1}{n} |g_\theta(X_i) - g_{\theta'}(X_i) - \mathbb{E}_P[g_\theta(X_i) - g_{\theta'}(X_i)]| + \frac{1}{n}\left|\gamma(g_\theta(Y_i)) - \gamma(g_{\theta'}(Y_i)) - \mathbb{E}_Q\left[\gamma(g_\theta(Y_i)) - \gamma(g_{\theta'}(Y_i))\right]\right| \\ &\le \frac{1}{n}\left[|g_\theta(X_i) - g_{\theta'}(X_i)| + |\mathbb{E}_P[g_\theta(X_i) - g_{\theta'}(X_i)]| + |\gamma(g_\theta(Y_i)) - \gamma(g_{\theta'}(Y_i))| + |\mathbb{E}_Q\left[\gamma(g_\theta(Y_i)) - \gamma(g_{\theta'}(Y_i))\right]|\right] \\ &\le \frac{1}{n} s_{k,\mathbf{a},\gamma} \|\boldsymbol{\beta}(\theta) - \boldsymbol{\beta}(\theta')\|_1 , \end{aligned} \tag{66}$$

where $s_{k,\mathbf{a},\gamma} := 2\left(\bar{\gamma}'_{\mathcal{G}_k(\mathbf{a})} + 1\right)$. Note that $\mathbb{E}[Z_\theta] = 0$ for all $\theta \in \Theta_k(\mathbf{a})$. Then, using the fact that $\|\boldsymbol{\beta}(\theta) - \boldsymbol{\beta}(\theta')\|_1 \le \sqrt{k}\, \|\boldsymbol{\beta}(\theta) - \boldsymbol{\beta}(\theta')\|$, it follows from (63) and (66) via Hoeffding's lemma that

$$\mathbb{E}\left[e^{t(Z_\theta - Z_{\theta'})}\right] \le e^{\frac{1}{2}t^2 \mathsf{d}_{k,\mathbf{a},n,\gamma}(\theta,\theta')^2}, \tag{67}$$

where

$$\mathsf{d}_{k,\mathbf{a},n,\gamma}(\theta, \theta') := \frac{s_{k,\mathbf{a},\gamma}\sqrt{k}\, \|\boldsymbol{\beta}(\theta) - \boldsymbol{\beta}(\theta')\|}{\sqrt{n}} := \frac{R_{k,\mathbf{a},\gamma}}{\sqrt{n}} \|\boldsymbol{\beta}(\theta) - \boldsymbol{\beta}(\theta')\| . \tag{68}$$

It follows that $\{Z_\theta\}_{\theta \in \Theta_k(\mathbf{a})}$ is a separable subgaussian process on the metric space $(\Theta_k(\mathbf{a}), \mathsf{d}_{k,\mathbf{a},n,\gamma}(\theta,\theta'))$. Next, note that $N(\Theta_k(\mathbf{a}), \mathsf{d}_{k,\mathbf{a},n,\gamma}(\cdot,\cdot), \epsilon) = N\left([-a_2, a_2]^k, n^{-\frac{1}{2}} R_{k,\mathbf{a},\gamma} \|\cdot\|, \epsilon\right)$. Also, $[-a_2, a_2]^k \subseteq B^k\left(\sqrt{k}\; a_2\right)$. Hence, we have

$$\begin{aligned} N(\Theta_k(\mathbf{a}), \mathsf{d}_{k,\mathbf{a},n,\gamma}(\cdot,\cdot), \epsilon) &\le N\left(B^k\left(\sqrt{k}\; a_2\right), n^{-\frac{1}{2}} R_{k,\mathbf{a},\gamma} \|\cdot\|, \epsilon\right) \\ &= N\left(B^k\left(\sqrt{k}\; a_2\right), \|\cdot\|, \sqrt{n} R^{-1}_{k,\mathbf{a},\gamma}\epsilon\right) \\ &\le \frac{\left(\sqrt{k}\; a_2 + \sqrt{n} R^{-1}_{k,\mathbf{a},\gamma}\epsilon\right)^k}{\left(\sqrt{n} R^{-1}_{k,\mathbf{a},\gamma}\epsilon\right)^k} \\ &= \left(1 + \frac{\sqrt{k}\; a_2\; R_{k,\mathbf{a},\gamma}}{\sqrt{n}\epsilon}\right)^k , \end{aligned} \tag{69}$$

where, in (69), we used that the covering number of Euclidean ball $B^d(r)$ w.r.t. Euclidean norm satisfies

$$N\left(B^d(r), \|\cdot\|, \epsilon\right) \le \left(\frac{r + \epsilon}{\epsilon}\right)^d . \tag{70}$$

Also, for $\epsilon \ge \mathsf{diam}(\Theta_k(\mathbf{a}), \mathsf{d}_{k,\mathbf{a},n,\gamma}) := \max_{\theta,\theta' \in \Theta_k(\mathbf{a})} \mathsf{d}_{k,\mathbf{a},n,\gamma}(\theta, \theta') = 2\sqrt{k} a_2 R_{k,\mathbf{a},\gamma} n^{-\frac{1}{2}}$, we have that $N(\Theta_k(\mathbf{a}), \mathsf{d}_{k,\mathbf{a},n,\gamma}(\cdot,\cdot), \epsilon) = 1$. Then,

$$\begin{aligned} E_{k,\mathbf{a},n,\gamma} &:= \int_0^\infty \sqrt{\log N(\Theta_k(\mathbf{a}), \mathsf{d}_{k,\mathbf{a},n,\gamma}(\cdot,\cdot), \epsilon)}\, d\epsilon \\ &= \int_0^{\mathsf{diam}(\Theta_k(\mathbf{a}), \mathsf{d}_{k,\mathbf{a},n,\gamma})} \sqrt{\log N(\Theta_k(\mathbf{a}), \mathsf{d}_{k,\mathbf{a},n,\gamma}(\cdot,\cdot), \epsilon)}\, d\epsilon \\ &\le \sqrt{k} \int_0^{\mathsf{diam}(\Theta_k(\mathbf{a}), \mathsf{d}_{k,\mathbf{a},n,\gamma})} \sqrt{\log\left(1 + \frac{a_2\sqrt{k} R_{k,\mathbf{a},\gamma}}{\sqrt{n}\epsilon}\right)} d\epsilon \\ &\le n^{-\frac{1}{4}} k^{\frac{3}{4}} \sqrt{a_2 R_{k,\mathbf{a},\gamma}} \int_0^{\mathsf{diam}(\Theta_k(\mathbf{a}), \mathsf{d}_{k,\mathbf{a},n,\gamma})} \epsilon^{-\frac{1}{2}} d\epsilon \end{aligned} \tag{71}$$

$$= 2k^{\frac{3}{4}} n^{-\frac{1}{4}} \sqrt{a_2 R_{k,\mathbf{a},\gamma} \; \mathsf{diam}\left(\Theta_k(\mathbf{a}), \mathsf{d}_{k,\mathbf{a},n,\gamma}\right)}, \tag{72}$$

where, we used the inequality $\log(1+x) \leq x$ (for $x \geq -1$) in (71). It follows from Theorem 1 that there exists a constant C such that for $\delta > 0$,

$$\mathbb{P}\left(\sup_{g_\theta \in \mathcal{G}_k(\mathbf{a})} Z_\theta \geq C E_{k,\mathbf{a},n,\gamma} + \delta\right) = \mathbb{P}\left(\sup_{g_\theta \in \mathcal{G}_k(\mathbf{a})} Z_\theta - Z_{\mathbf{0}} \geq C E_{k,\mathbf{a},n,\gamma} + \delta\right)$$

$$\leq C e^{-\frac{\delta^2}{C \mathsf{diam}\left(\Theta_k(\mathbf{a}), \mathsf{d}_{k,\mathbf{a},n,\gamma}\right)^2}} = C e^{-\frac{n\delta^2}{4Ca_2^2 R_{k,\mathbf{a},\gamma}^2 k}}, \tag{73}$$

where $Z_{\mathbf{0}} = 0$. It follows similarly that for $\delta > 0$,

$$\mathbb{P}\left(\sup_{g_\theta \in \mathcal{G}_k(\mathbf{a})} -Z_\theta \geq \delta + C E_{k,\mathbf{a},n,\gamma}\right) \leq C e^{-\frac{n\delta^2}{4Ca_2^2 R_{k,\mathbf{a},\gamma}^2 k}}. \tag{74}$$

Combining (73) and (74) yields

$$\mathbb{P}\left(\sup_{g_\theta \in \mathcal{G}_k(\mathbf{a})} |Z_\theta| \geq \delta + C E_{k,\mathbf{a},n,\gamma}\right) \leq 2C e^{-\frac{n\delta^2}{4Ca_2^2 R_{k,\mathbf{a},\gamma}^2 k}}. \tag{75}$$

From (61), (62) and (75), we obtain that for $\delta > 0$,

$$\mathbb{P}\left(\left|\mathsf{H}_{\gamma,\mathcal{G}_k(\mathbf{a})}(P,Q) - \hat{\mathsf{H}}_{\gamma,\mathcal{G}_k(\mathbf{a})}(X^n,Y^n)\right| \geq \delta + C E_{k,\mathbf{a},n,\gamma}\right)$$

$$\leq \mathbb{P}\left(\sup_{g_\theta \in \mathcal{G}_k(\mathbf{a})} |Z_\theta| \geq \delta + C E_{k,\mathbf{a},n,\gamma}\right) \leq 2C e^{-\frac{n\delta^2}{4Ca_2^2 R_{k,\mathbf{a},\gamma}^2 k}}. \tag{76}$$

B Appendix: KL divergence

B.1 Proof of Theorem 4

Let $D_{\mathcal{G}_k(\mathbf{a}_k)}(P,Q) := \mathsf{H}_{\gamma_{\mathsf{KL}},\mathcal{G}_k(\mathbf{a}_k)}(P,Q)$. The proof of Theorem 4 relies on the following lemma, whose proof is given in Appendix B.1.1.

Lemma 2. *Let $P, Q \in \mathcal{P}_{\mathsf{KL}}(\mathcal{X})$. Then, for $X^n \sim P^{\otimes n}$ and $Y^n \sim Q^{\otimes n}$, the following holds for any $\alpha > 0$:*

(i) For $n, k_n, \mathbf{a}_{k_n} = (a_{1,k_n}, a_{2,k_n}, a_{3,k_n})$ such that $k_n^{\frac{3}{2}} a_{2,k_n} e^{k_n a_{2,k_n} + a_{3,k_n}} = O\left(n^{\frac{1-\alpha}{2}}\right)$,

$$\hat{D}_{\mathcal{G}_{k_n}(\mathbf{a}_{k_n})}(X^n,Y^n) \xrightarrow[n\to\infty]{} D_{\mathcal{G}_{k_n}(\mathbf{a}_{k_n})}(P,Q), \quad \mathbb{P}-\text{a.s.}. \tag{77}$$

(ii) For $n, k, \mathbf{a}_k = (a_{1,k}, a_{2,k}, a_{3,k})$ such that $k^{\frac{3}{2}} a_{2,k} e^{k a_{2,k} + a_{3,k}} = O\left(n^{\frac{1-\alpha}{2}}\right)$

$$\mathbb{E}\left[\left|\hat{D}_{\mathcal{G}_k(\mathbf{a}_k)}(X^n,Y^n) - D_{\mathcal{G}_k(\mathbf{a}_k)}(P,Q)\right|\right] = O\left(n^{-\frac{1}{2}} k^{\frac{3}{2}} a_{2,k} e^{k a_{2,k} + a_{3,k}}\right). \tag{78}$$

We proceed to prove (20). Since $f_{\mathsf{KL}} \in \mathsf{C}(\mathcal{X})$ for a compact set $\mathcal{X}$, it follows from Stinchcombe and White (1990, Theorem 2.8) that for any $\epsilon > 0$ and $k \geq k_0(\epsilon)$, there exists a $g_{\tilde{\theta}} \in \mathcal{G}_k(\mathbf{1})$ such that

$$\sup_{x\in\mathcal{X}} \left|f_{\mathsf{KL}}(x) - g_{\tilde{\theta}}(x)\right| \leq \epsilon. \tag{79}$$

This implies that

$$\lim_{k\to\infty} D_{\mathcal{G}_k(\mathbf{1})}(P,Q) = \mathsf{D}_{\mathsf{KL}}\left(P\|Q\right). \tag{80}$$

To see this, note that

$$D_{\mathcal{G}_k(\mathbf{1})}(P,Q) \le \mathsf{D}_{\mathsf{KL}}\left(P\|Q\right),\ \forall k \in \mathbb{N}, \tag{81}$$

by (18) since g_θ is continuous and bounded ($|g_\theta| \le k+1$). Moreover, the left hand side (L.H.S.) of (81) is monotonically increasing in k, and being bounded, has a limit point. Then, (80) will follow if we show that the limit point is $\mathsf{D}_{\mathsf{KL}}\left(P\|Q\right)$. Assume otherwise that $\lim_{k\to\infty} D_{\mathcal{G}_k(\mathbf{1})}(P,Q) < \mathsf{D}_{\mathsf{KL}}\left(P\|Q\right)$. Note that $\mathcal{G}_k(\mathbf{1})$ is a closed set and hence the supremum in the variational form of the L.H.S. of (81) is a maximum. Then, defining

$$D(g) := 1 + \mathbb{E}_P[g(X)] - \mathbb{E}_Q\left[e^{g(Y)}\right], \tag{82}$$

this implies that there exists $\delta > 0$ and

$$g_{\theta_k^*} := \underset{g_\theta \in \mathcal{G}_k(\mathbf{1})}{\arg\max}\, D(g_\theta), \tag{83}$$

such that for all k,

$$\mathsf{D}_{\mathsf{KL}}\left(P\|Q\right) - D(g_{\theta_k^*}) \ge \delta. \tag{84}$$

However, it follows from (79) that for all $k \ge k_0(\epsilon)$,

$$\begin{aligned}
\mathsf{D}_{\mathsf{KL}}\left(P\|Q\right) - D(g_{\theta_k^*}) &\le \mathsf{D}_{\mathsf{KL}}\left(P\|Q\right) - D(g_{\tilde{\theta}}) \\
&\le \mathbb{E}_P\left[\left|f_{\mathsf{KL}}(X) - g_{\tilde{\theta}}(X)\right|\right] + \mathbb{E}_Q\left[\left|e^{f_{\mathsf{KL}}(Y)} - e^{g_{\tilde{\theta}}(Y)}\right|\right] \\
&\le \mathbb{E}_P\left[\left|f_{\mathsf{KL}}(X) - g_{\tilde{\theta}}(X)\right|\right] + L_{P,Q}\ \mathbb{E}_Q\left[\left|1 - e^{g_{\tilde{\theta}}(Y) - f_{\mathsf{KL}}(Y)}\right|\right] && (85)\\
&\le \epsilon + L_{P,Q}(e^{\epsilon} - 1), && (86)
\end{aligned}$$

where (86) follows from (79). Note that

$$0 \le L_{P,Q} := \left\|\frac{\mathsf{d}P}{\mathsf{d}Q}\right\|_\infty < \infty, \tag{87}$$

since $e^{f_{\mathsf{KL}}}$ is a continuous function and hence bounded over a compact support $\mathcal{X}$. Taking ϵ sufficiently small in (86) contradicts (84), thus proving (80). Next, for $a_{3,k} = a_{2,k} = a_{1,k} = 1$ and any $\eta > 0$, $k^{\frac{3}{2}} a_{2,k} e^{k a_{2,k} + a_{3,k}} < e^{k(1+\eta)}$ provided k is sufficiently large. Then, (20) follows from (77) and (80) by letting $k = k_n \to \infty$ (subject to constraint in Lemma 2(i)), and noting that $\eta > 0$ is arbitrary.

Next, we prove (21). Note that since $f_{\mathsf{KL}} \in \mathcal{I}(M)$, we have from (42) that for k such that $m_k \ge M$, there exists $g_\theta \in \mathcal{G}_k^*(m_k)$ satisfying

$$\|f_{\mathsf{KL}} - g_\theta\|_{\infty,P,Q} \le \tilde{C}_{d,M} k^{-\frac{1}{2}} = \left((\hat{c}_d + 2)M + 2\right) k^{-\frac{1}{2}}.$$

On the other hand, for k such that $m_k < M$, taking $g_{\mathbf{0}} = 0$ yields $\|f_{\mathsf{KL}} - g_{\mathbf{0}}\|_{\infty,P,Q} \le M$. Hence, for all k, there exists $g_{\theta_k^*} \in \mathcal{G}_k^*(m_k)$ such that

$$\left\|f_{\mathsf{KL}} - g_{\theta_k^*}\right\|_{\infty,P,Q} \le D_{d,M,\mathbf{m}} k^{-\frac{1}{2}}, \tag{88}$$

where $\mathbf{m} = \{m_k\}_{k\in\mathbb{N}}$,

$$D_{d,M,\mathbf{m}} := \tilde{C}_{d,M} \vee \sqrt{\bar{m}(M,\mathbf{m})}M, \tag{89}$$

$$\bar{m}(M,\mathbf{m}) := \min\left\{k \in \mathbb{N} : m_k \ge M\right\}. \tag{90}$$

Also, observe that $\mathsf{D}_{\mathsf{KL}}\left(P\|Q\right) \ge D_{\mathcal{G}_k^*(m_k)}(P,Q)$ since $g_{\theta_k^*} \in \mathcal{G}_k^*(m_k)$ is bounded. Then, the following chain of inequalities hold:

$$\left|\mathsf{D}_{\mathsf{KL}}\left(P\|Q\right) - D_{\mathcal{G}_k^*(m_k)}(P,Q)\right|$$

$$= \mathsf{D}_{\mathsf{KL}}\left(P\|Q\right) - D_{\mathcal{G}_k^*(m_k)}(P,Q)$$
$$\stackrel{(a)}{\leq} \mathbb{E}_P\left[\left|f_{\mathsf{KL}}(X) - g_{\theta_k^*}(X)\right|\right] + L_{P,Q}\,\mathbb{E}_Q\left[\left|1 - e^{g_{\theta_k^*}(Y) - f_{\mathsf{KL}}(Y)}\right|\right]$$
$$\stackrel{(b)}{\leq} D_{d,M,\mathbf{m}}k^{-\frac{1}{2}} + e^M\left(e^{D_{d,M,\mathbf{m}}k^{-\frac{1}{2}}} - 1\right), \tag{91}$$

where

(a) follows similar to (85);

(b) is due to (88) and $L_{P,Q} \leq e^M$ since $f_{\mathsf{KL}} \in \mathcal{I}(M)$.

On the other hand, taking $a_{1,k} = \sqrt{k}\log k$, $ka_{2,k} = a_{3,k} = m_k$, and k satisfying $\sqrt{k}e^{2m_k} = O\left(n^{\frac{1-\alpha}{2}}\right)$ for some $\alpha > 0$, we have

$$\mathbb{E}\left[\left|\hat{D}_{\mathcal{G}_k^*(m_k)}(X^n,Y^n) - \mathsf{D}_{\mathsf{KL}}\left(P\|Q\right)\right|\right]$$
$$\stackrel{(a)}{\leq} \left|D_{\mathcal{G}_k^*(m_k)}(P,Q) - \mathsf{D}_{\mathsf{KL}}\left(P\|Q\right)\right| + \mathbb{E}\left[\left|D_{\mathcal{G}_k^*(m_k)}(P,Q) - \hat{D}_{\mathcal{G}_k^*(M)}(X^n,Y^n)\right|\right]$$
$$\stackrel{(b)}{\leq} D_{d,M,\mathbf{m}}k^{-\frac{1}{2}} + e^M\left(e^{D_{d,M,\mathbf{m}}k^{-\frac{1}{2}}} - 1\right) + O\left(e^{2m_k}\sqrt{k}\;n^{-\frac{1}{2}}\right) \tag{92}$$
$$\stackrel{(c)}{=} O_M\left(e^{D_{d,M,\mathbf{m}}k^{-\frac{1}{2}}} - 1\right) + O\left(e^{2m_k}\sqrt{k}n^{-\frac{1}{2}}\right), \tag{93}$$

where

(a) is due to triangle inequality;

(b) follows from (78) and (91).

Choosing $m_k = 0.5\log k$ in (93) yields

$$\mathbb{E}\left[\left|\hat{D}_{\mathcal{G}_k^*(0.5\log k)}(X^n,Y^n) - \mathsf{D}_{\mathsf{KL}}\left(P\|Q\right)\right|\right] = O\left(k^{-\frac{1}{2}}\right) + O\left(k^{\frac{3}{2}}n^{-\frac{1}{2}}\right), \tag{94}$$

since for k sufficiently large,

$$e^{D_{d,M,\mathbf{m}}k^{-\frac{1}{2}}} - 1 = \sum_{j=1}^{\infty}\frac{\left(D_{d,M,\mathbf{m}}k^{-\frac{1}{2}}\right)^j}{j!} \leq \sum_{j=1}^{\infty}\left(D_{d,M,\mathbf{m}}k^{-\frac{1}{2}}\right)^j = O\left(k^{-\frac{1}{2}}\right).$$

This completes the proof.

Remark 10. *Setting $m_k = M$ in* (93) *and via steps leading to* (94)*, we obtain* (22).

B.1.1 Proof of Lemma 2

Note that for $\gamma_{\mathsf{KL}}(x) = e^x - 1$,

$$\bar{\gamma}'_{\mathcal{G}_k(\mathbf{a}_k)} = \sup_{\substack{x\in\mathcal{X},\\ g_\theta\in\mathcal{G}_k(\mathbf{a}_k)}} \gamma'_{\mathsf{KL}}(g_\theta(x)) \leq e^{ka_{2,k}+a_{3,k}},$$
$$R_{k,\mathbf{a}_k,\gamma} \leq 2\sqrt{k}\left(e^{ka_{2,k}+a_{3,k}} + 1\right),$$

where γ'_{KL} denotes the derivative of γ_{KL}. Since

$$E_{k,\mathbf{a}_k,n,\gamma} \leq 4\sqrt{2}n^{-\frac{1}{2}}k^{\frac{3}{2}}a_{2,k}\left(e^{ka_{2,k}+a_{3,k}} + 1\right) \xrightarrow[n\to\infty]{} 0, \tag{95}$$

for $k, \mathbf{a}_k$ such that $k^{\frac{3}{2}} a_{2,k} e^{k a_{2,k} + a_{3,k}} = O\left(n^{\frac{1-\alpha}{2}}\right)$ for $\alpha > 0$, it follows from (17) that for any $k \in \mathbb{N}$, $\delta > 0$, and n sufficiently large,

$$\mathbb{P}\left(\left|D_{\mathcal{G}_k(\mathbf{a}_k)}(P,Q) - \hat{D}_{\mathcal{G}_k(\mathbf{a}_k)}(X^n,Y^n)\right| \geq \delta\right) \leq 2Ce^{-\frac{n(\delta - CE_{k,\mathbf{a}_k,n,\gamma})^2}{16Ca_{2,k}^2 k^2 \left(e^{ka_{2,k}+a_{3,k}}+1\right)^2}}. \tag{96}$$

Hence, for $k_n, \mathbf{a}_{k_n}$ such that $k_n^{\frac{3}{2}} a_{2,k_n} e^{k_n a_{2,k_n} + a_{1,k_n}} = O\left(n^{\frac{1-\alpha}{2}}\right)$,

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\left|D_{\mathcal{G}_{k_n}(\mathbf{a}_{k_n})}(P,Q) - \hat{D}_{\mathcal{G}_{k_n}(\mathbf{a}_{k_n})}(X^n,Y^n)\right| \geq \delta\right) \leq 2C \sum_{n=1}^{\infty} e^{-\frac{n(\delta - CE_{k,\mathbf{a}_k,n,\gamma})^2}{16Ca_{2,k_n}^2 k_n^2 \left(e^{k_n a_{2,k_n}+a_{1,k_n}}+1\right)^2}} < \infty, \tag{97}$$

where the final inequality in (97) can be established via integral test for sum of series. This implies (77) via the first Borel-Cantelli lemma. To prove (78), note that

$$\begin{aligned}
&\mathbb{E}\left[\left|D_{\mathcal{G}_k(\mathbf{a}_k)}(P,Q) - \hat{D}_{\mathcal{G}_k(\mathbf{a}_k)}(X^n,Y^n)\right|\right] \\
&= \int_0^{\infty} \mathbb{P}\left(\left|D_{\mathcal{G}_k(\mathbf{a}_k)}(P,Q) - \hat{D}_{\mathcal{G}_k(\mathbf{a}_k)}(X^n,Y^n)\right| \geq \delta\right) d\delta \\
&\leq CE_{k,\mathbf{a}_k,n,\gamma} + \int_{CE_{k,\mathbf{a}_k,n,\gamma}}^{\infty} 2Ce^{-\frac{n(\delta - CE_{k,\mathbf{a}_k,n,\gamma})^2}{16Ca_{2,k}^2 k^2 \left(e^{ka_{2,k}+a_{3,k}}+1\right)^2}} d\delta \\
&= O\left(n^{-\frac{1}{2}} k^{\frac{3}{2}} a_{2,k} e^{ka_{2,k}+a_{3,k}}\right).
\end{aligned} \tag{98}$$

B.2 Proof of Proposition 1

From proof of Corollary 1 (see (53)), there exists extensions $f_p^{(\mathsf{e})}, f_q^{(\mathsf{e})} \in \mathcal{B}_{b' \vee c'} \cap \mathcal{S}_{s,b'}\left(\mathbb{R}^d\right)$ of $f, \bar{f}$, respectively (see (55) and (56) for definitions of b' and c'). Define $f_{\mathsf{KL}}^{(\mathsf{e})} := f_p^{(\mathsf{e})} - f_q^{(\mathsf{e})}$. Since $f_p^{(\mathsf{e})}, f_q^{(\mathsf{e})} \in \mathcal{S}_{s,b'}\left(\mathbb{R}^d\right)$, their Fourier transforms exists such that corresponding Fourier inversion formulas hold. Also, we have

$$B\left(f_{\mathsf{KL}}^{(\mathsf{e})}\right) \overset{(a)}{\leq} B\left(f_p^{(\mathsf{e})}\right) + B\left(f_q^{(\mathsf{e})}\right) \overset{(b)}{\leq} 2(b' \vee c'), \tag{99}$$

$$\max_{x \in \mathcal{X}} \left|f_{\mathsf{KL}}^{(\mathsf{e})}(x)\right| \leq \max_{x \in \mathcal{X}} \left|f_p^{(\mathsf{e})}(x)\right| + \max_{x \in \mathcal{X}} \left|f_q^{(\mathsf{e})}(x)\right| \overset{(d)}{\leq} 2b, \tag{100}$$

where

(a) follows from the definition in (4) and linearity of the Fourier transform;

(b) (c) is since $f_p^{(\mathsf{e})}, f_q^{(\mathsf{e})} \in \mathcal{B}_{b' \vee c'}$;

(d) is due to $(P,Q) \in \mathcal{L}_{\mathsf{KL}}(b,c)$.

Hence, it follows from (99)-(100) that $f_{\mathsf{KL}}^{(\mathsf{e})}|_{\mathcal{X}} \in \mathcal{I}(M)$ with $M = 2\bar{c}_{b,c,d}$ (since $b \leq b'$), where $\bar{c}_{b,c,d}$ is given in (54). The claim then follows from Theorem 4 since $f_{\mathsf{KL}} = f_{\mathsf{KL}}^{(\mathsf{e})}|_{\mathcal{X}}$.

C Appendix: χ^2 divergence

C.1 Proof of Theorem 5

Let $\chi^2_{\mathcal{G}_k(\mathbf{a}_k)}(P,Q) := \mathsf{H}_{\gamma_{\chi^2}, \mathcal{G}_k(\mathbf{a}_k)}(P,Q)$. The proof of Theorem 5 is based on the lemma below (see Appendix C.1.1 for proof).

Lemma 3. *Let $P, Q \in \mathcal{P}_{\chi^2}(\mathcal{X})$. For $X^n \sim P^{\otimes n}$ and $Y^n \sim Q^{\otimes n}$, the following holds for any $\alpha > 0$:*

(i) For $n, k_n, \mathbf{a}_{k_n}$ such that $k_n^{\frac{5}{2}} a_{2,k_n}^2 + k_n^{\frac{3}{2}} a_{2,k_n} a_{3,k_n} = O\left(n^{\frac{1-\alpha}{2}}\right)$,

$$\hat{\chi^2}_{\mathcal{G}_k(\mathbf{a}_{k_n})}(X^n, Y^n) \xrightarrow[n\to\infty]{} \chi^2_{\mathcal{G}_{k_n}(\mathbf{a}_{k_n})}(P, Q), \quad \mathbb{P} - a.s. \tag{101}$$

(ii) For $n, k, \mathbf{a}_k$ such that $k^{\frac{5}{2}} a_{2,k}^2 + k^{\frac{3}{2}} a_{2,k} a_{3,k} = O\left(n^{\frac{1-\alpha}{2}}\right)$,

$$\mathbb{E}\left[\left|\hat{\chi^2}_{\mathcal{G}_k(\mathbf{a}_k)}(X^n, Y^n) - \chi^2_{\mathcal{G}_k(\mathbf{a}_k)}(P, Q)\right|\right] = O\left(n^{-\frac{1}{2}}\left(k^{\frac{5}{2}} a_{2,k}^2 + k^{\frac{3}{2}} a_{2,k} a_{3,k}\right)\right). \tag{102}$$

The proof of (25) follows from (101), using similar arguments used to establish (20) and steps leading to (104) below. The details are omitted.

We proceed to prove (26). Since $f_{\chi^2} \in \mathcal{I}(M)$, we have similar to (88) that there exists $g_{\theta_k^*} \in \mathcal{G}_k^*(m_k)$

$$\left\|f_{\chi^2} - g_{\theta_k^*}\right\|_{\infty,P,Q} = D_{d,M,\mathbf{m}} k^{-\frac{1}{2}}, \tag{103}$$

where $D_{d,M,\mathbf{m}}$ is defined in (89). Also, $\chi^2(P\|Q) \geq \chi^2_{\mathcal{G}_k^*(m_k)}(P, Q)$ since $g_\theta \in \mathcal{G}_k^*(m_k)$ is bounded. Then, we have

$$\begin{aligned}
&\left|\chi^2(P\|Q) - \chi^2_{\mathcal{G}_k^*(m_k)}(P, Q)\right| \\
&= \chi^2(P\|Q) - \chi^2_{\mathcal{G}_k^*(m_k)}(P, Q) \\
&\leq \chi^2(P\|Q) - \mathbb{E}_P[g_{\theta_k^*}(X)] - \mathbb{E}_Q\left[g_{\theta_k^*}(Y) + \frac{g_{\theta_k^*}^2(Y)}{4}\right] \\
&\leq \mathbb{E}_P\left[\left|f_{\chi^2}(X) - g_{\theta_k^*}(X)\right|\right] + \mathbb{E}_Q\left[\left|f_{\chi^2}(Y) - g_{\theta_k^*}(Y)\right| + \frac{1}{4}\left|f_{\chi^2}^2(Y) - g_{\theta_k^*}^2(Y)\right|\right] \\
&\leq 2D_{d,M,\mathbf{m}} k^{-\frac{1}{2}} + \mathbb{E}_Q\left[\frac{1}{4}\left|f_{\chi^2}(Y) - g_{\theta_k^*}(Y)\right|\left|f_{\chi^2}(Y) + g_{\theta_k^*}(Y)\right|\right] \\
&\leq 2D_{d,M,\mathbf{m}} k^{-\frac{1}{2}} + \mathbb{E}_Q\Big[\frac{1}{4}\left|f_{\chi^2}(Y) - g_{\theta_k^*}(Y)\right|\left|g_{\theta_k^*}(Y) - f_{\chi^2}(Y)\right| + \frac{1}{2}\left|f_{\chi^2}(Y) - g_{\theta_k^*}(Y)\right|\left|f_{\chi^2}(Y)\right|\Big] \\
&\leq 2D_{d,M,\mathbf{m}} k^{-\frac{1}{2}} + \frac{D_{d,M,\mathbf{m}}^2}{4k} + \frac{D_{d,M,\mathbf{m}} M}{2\sqrt{k}},
\end{aligned} \tag{104}$$

where (104) is due to $f_{\chi^2} \in \mathcal{I}(M)$. Taking $a_{1,k} = \sqrt{k}\log k$, $k a_{2,k} = a_{3,k} = m_k$, and k, m_k satisfying $m_k^2\sqrt{k} = O\left(n^{(1-\alpha)/2}\right)$, we have

$$\begin{aligned}
&\mathbb{E}\left[\left|\hat{\chi^2}_{\mathcal{G}_k^*(m_k)}(X^n, Y^n) - \chi^2(P\|Q)\right|\right] \\
&\overset{(a)}{\leq} \left|\chi^2_{\mathcal{G}_k^*(m_k)}(P, Q) - \chi^2(P\|Q)\right| + \mathbb{E}\left[\left|\chi^2_{\mathcal{G}_k^*(m_k)}(P, Q) - \hat{\chi^2}_{\mathcal{G}_k^*(m_k)}(X^n, Y^n)\right|\right] \\
&\overset{(b)}{\leq} 2D_{d,M,\mathbf{m}} k^{-\frac{1}{2}} + \frac{D_{d,M,\mathbf{m}}^2}{4k} + \frac{D_{d,M,\mathbf{m}} M}{2\sqrt{k}} + O\left(m_k^2\sqrt{k}\, n^{-\frac{1}{2}}\right), \\
&\overset{(c)}{=} O_{d,M}\left(\bar{m}(M, \mathbf{m}) k^{-\frac{1}{2}}\right) + O\left(m_k^2\sqrt{k}\, n^{-\frac{1}{2}}\right),
\end{aligned} \tag{105}$$

where

(a) is due to triangle inequality;

(b) follows from (102) and (104);

(c) is by the definition of $D_{d,M,\mathbf{m}}$ in (89) and since $\bar{m}(M, \mathbf{m}) \geq 1$.

Setting $\mathbf{m} = \{0.5\log k\}_{k\in\mathbb{N}}$ in (105) yields (26), thus completing the proof.

C.1.1 Proof of Lemma 3

For $\gamma_{\chi^2}(x) = x + \frac{x^2}{4}$, we have

$$\bar{\gamma}'_{\mathcal{G}_k(\mathbf{a}_k)} = \sup_{\substack{x\in\mathcal{X},\\ g_\theta\in\mathcal{G}_k(\mathbf{a}_k)}} \gamma'_{\chi^2}(g_\theta(x)) \leq 0.5(ka_{2,k}+a_{3,k})+1,$$
$$R_{k,\mathbf{a}_k,\gamma} \leq 2\sqrt{k}\left(0.5(ka_{2,k}+a_{3,k})+2\right), \tag{106}$$

where $\gamma'_{\chi^2}(\cdot)$ denotes the derivative of γ_{χ^2}. Since

$$0 \leq E_{k,\mathbf{a}_k,n,\gamma} \leq 4\sqrt{2}n^{-\frac{1}{2}}k^{\frac{3}{2}}a_{2,k}\left(0.5(ka_{2,k}+a_{3,k})+2\right) \xrightarrow[n\to\infty]{} 0, \tag{107}$$

for $k, \mathbf{a}_k$ such that $k^{\frac{5}{2}}a_{2,k}^2 + k^{\frac{3}{2}}a_{2,k}a_{3,k} = O\left(n^{\frac{1-\alpha}{2}}\right)$, it follows from (17) that for any $k\in\mathbb{N}$, $\delta>0$, and n sufficiently large,

$$\mathbb{P}\left(\left|\hat{\chi}^2{}_{\mathcal{G}_k(\mathbf{a}_k)}(X^n,Y^n) - \chi^2_{\mathcal{G}_k(\mathbf{a}_k)}(P,Q)\right| \geq \delta\right) \leq 2Ce^{-\frac{n(\delta - CE_{k,\mathbf{a}_k,n,\gamma})^2}{16Ca_{2,k}^2k^2\left(0.5(ka_{2,k}+a_{3,k})+2\right)^2}}. \tag{108}$$

Then, (101) and (102) follows using similar steps used to prove (77) (see (97)) and (78) (see (98)) in Theorem 4, respectively. This completes the proof.

C.2 Proof of Proposition 2

It follows from (53) that there exists extensions $f_p^{(\mathsf{e})}, f_q^{(\mathsf{e})} \in \mathcal{B}_{b'\vee c'} \cap \tilde{\mathcal{S}}_{s,b'}\left(\mathbb{R}^d\right)$ of $f, \bar{f} \in \mathcal{H}^{s,\delta}_{b,c}(\mathcal{U})$, respectively, where $\tilde{\mathcal{S}}_{s,b'}\left(\mathbb{R}^d\right)$ is defined in (52). Let $f_{\chi^2}^{(\mathsf{e})} = 2\left(f_p^{(\mathsf{e})}\cdot f_q^{(\mathsf{e})} - 1\right)$. Recall the notation $\boldsymbol{\alpha}_j$ for a multi-index of order j. We have from the chain rule for differentiation that $D^{\boldsymbol{\alpha}_j} f_{\chi^2}^{(\mathsf{e})}(x)$ is the sum of 2^j terms of the form $D^{\boldsymbol{\alpha}_{j_1}} f_p^{(\mathsf{e})}(x)\cdot D^{\boldsymbol{\alpha}_{j_2}} f_q^{(\mathsf{e})}(x)$, where $\boldsymbol{\alpha}_{j_1}+\boldsymbol{\alpha}_{j_2}=\boldsymbol{\alpha}_j$. Also, note from (50) and (51) that for $j=0,\ldots,s$, $f_p^{(\mathsf{e})}$, $f_q^{(\mathsf{e})}$ satisfies

$$\left|D^{\boldsymbol{\alpha_j}} f_p^{(\mathsf{e})}(x)\right| \vee \left|D^{\boldsymbol{\alpha_j}} f_q^{(\mathsf{e})}(x)\right| \leq \hat{b} \leq b',\ \forall\, x\in\mathbb{R}^d, \tag{109a}$$

$$\left\|D^{\boldsymbol{\alpha}_j} f_p^{(\mathsf{e})}\right\|_{L^i(\mathbb{R}^d)} \vee \left\|D^{\boldsymbol{\alpha}_j} f_q^{(\mathsf{e})}\right\|_{L^i(\mathbb{R}^d)} \leq b',\ i=1,2. \tag{109b}$$

Then, it follows that for $j=0,\ldots,s$ and $i=1,2$,

$$\begin{aligned}\left\|D^{\boldsymbol{\alpha}_j} f_{\chi^2}^{(\mathsf{e})}\right\|_{L^i(\mathbb{R}^d)} &\leq 2 + 2\left\|\sum_{\substack{\boldsymbol{\alpha}_{j_1},\boldsymbol{\alpha}_{j_2}:\\ \boldsymbol{\alpha}_{j_1}+\boldsymbol{\alpha}_{j_2}=\boldsymbol{\alpha}_j}} D^{\boldsymbol{\alpha}_{j_1}} f_p^{(\mathsf{e})}\cdot D^{\boldsymbol{\alpha}_{j_2}} f_q^{(\mathsf{e})}\right\|_{L^i(\mathbb{R}^d)}\\ &\leq 2 + 2^{j+1}b'\max_{\boldsymbol{\alpha}_{j_2}}\left\|D^{\boldsymbol{\alpha}_{j_2}} f_q^{(\mathsf{e})}\right\|_{L^i(\mathbb{R}^d)}\\ &\leq 2 + 2^{j+1}b'^2.\end{aligned} \tag{110}$$

Hence, $f_{\chi^2}^{(\mathsf{e})} \in \tilde{\mathcal{S}}_{s,2+2^{s+1}b'^2}\left(\mathbb{R}^d\right)$. From Lemma 1, it follows that $B\left(f_{\chi^2}^{(\mathsf{e})}\right) \leq (2+2^{s+1}b'^2)\kappa_d\sqrt{d}$. Moreover, we have

$$\sup_{x\in\mathcal{X}}\left|f_{\chi^2}^{(\mathsf{e})}\right| \leq 2 + 2\sup_{x\in\mathcal{X}}\frac{p(x)}{q(x)} \leq 2+2b^2. \tag{111}$$

This implies that $f_{\chi^2}^{(\mathsf{e})}|_{\mathcal{X}} \in \mathcal{I}\left((2+2^{s+1}b'^2)\left(\kappa_d\sqrt{d}\vee 1\right)\right)$ since $b'\geq b$. The claim then follows from Theorem 5 by noting that $f_{\chi^2} = f_{\chi^2}^{(\mathsf{e})}|_{\mathcal{X}}$ and $b'^2 \leq \bar{c}^2_{b,c,d}$.

D Appendix: Squared Hellinger distance

D.1 Proof of Theorem 6

Let $H^2_{\tilde{\mathcal{G}}_k(\mathbf{a}_k,t)}(P,Q) := \mathsf{H}_{\gamma_{H^2},\tilde{\mathcal{G}}_k(\mathbf{a}_k,t)}(P,Q)$. The proof of Theorem 6 hinges on the following lemma, whose proof is given in Appendix D.1.1.

Lemma 4. *Let $P,Q \in \mathcal{P}_{H^2}(\mathcal{X})$. For $X^n \sim P^{\otimes n}$ and $Y^n \sim Q^{\otimes n}$, the following holds for any $\alpha > 0$:*

(i) For $n, k_n, \mathbf{a}_{k_n}$ such that $k_n^{\frac{3}{2}} a_{2,k_n} t_{k_n}^{-2} = O\left(n^{\frac{1-\alpha}{2}}\right)$,

$$\hat{H}^2_{\tilde{\mathcal{G}}_{k_n}(\mathbf{a}_{k_n},t_{k_n})}(X^n,Y^n) \xrightarrow[n\to\infty]{} H^2_{\tilde{\mathcal{G}}_{k_n}(\mathbf{a}_{k_n},t_{k_n})}(P,Q), \quad \mathbb{P}-a.s. \tag{112}$$

(ii) For $n, k, \mathbf{a}_k$ such that $k^{\frac{3}{2}} a_{2,k} t_k^{-2} = O\left(n^{\frac{1-\alpha}{2}}\right)$,

$$\mathbb{E}\left[\left|\hat{H}^2_{\tilde{\mathcal{G}}_k(\mathbf{a}_k,t_k)}(X^n,Y^n) - H^2_{\tilde{\mathcal{G}}_k(\mathbf{a}_k,t_k)}(P,Q)\right|\right] = O\left(n^{-\frac{1}{2}} k^{\frac{3}{2}} a_{2,k} t_k^{-2}\right). \tag{113}$$

We first prove (31). Since $f_{H^2} \in \mathsf{C}(\mathcal{X})$ for a compact set $\mathcal{X}$, its supremum is achieved at some $x^* \in \mathcal{X}$. Also, since $\left\|\frac{dP}{dQ}\right\|_\infty < \infty$ by definition of the Radon-Nikodym derivative, we have $\sup_{x\in\mathcal{X}} f_{H^2}(x) = f_{H^2}(x^*) < 1$. Moreover, $t_k \le 1 - f_{H^2}(x^*)$ for sufficiently large k since $t_k \to 0$. Then, it follows from Stinchcombe and White (1990, Theorem 2.8) that for any $\epsilon > 0$ and $k \ge k_0(\epsilon)$ (some integer), there exists a $g_{\theta^*} \in \tilde{\mathcal{G}}^{(1)}_{k,t_k}$ such that

$$\sup_{x\in\mathcal{X}} |f_{H^2}(x) - g_{\theta^*}(x)| \le \epsilon. \tag{114}$$

This implies similar to (80) in Theorem 4 that

$$\lim_{k\to\infty} H^2_{\tilde{\mathcal{G}}^{(1)}_{k,t_k}}(P,Q) = H^2(P,Q). \tag{115}$$

Then, (31) follows from (112) and (115).

Next, we prove (32). Since $f_{H^2} \in \mathcal{I}_{H^2}(M)$, $1 - f_{H^2}(x) \ge \frac{1}{M}$ for all $x \in \mathcal{X}$. Using $t_k \to 0$, we have from (12) that for k such that $t_k \le \frac{1}{M}$ and $m_k \ge M$, there exists $g_\theta \in \tilde{\mathcal{G}}^{(2)}_{k,m_k,t_k}$ such that

$$\|f_{H^2} - g_\theta\|_{\infty,P,Q} \le \tilde{C}_{d,M} k^{-\frac{1}{2}}. \tag{116}$$

On the other hand, for k such that $t_k > \frac{1}{M}$ or $m_k < M$, taking $g_{\mathbf{0}} = 0$ yields $\|f_{H^2} - g_{\mathbf{0}}\|_{\infty,P,Q} \le M$ as $f_{H^2} \in \mathcal{I}(M)$. Then, denoting $\mathbf{t} = \{t_k\}_{k\in\mathbb{N}}$, it follows similar to (88) that for all k, there exists $g_{\theta_k^*} \in \tilde{\mathcal{G}}^{(2)}_{k,m_k,t_k}$ such that

$$\left\|f_{H^2} - g_{\theta_k^*}\right\|_{\infty,P,Q} \le \tilde{C}_{d,M} k^{-\frac{1}{2}} \vee \left(\sqrt{\bar{t}\,(M^{-1},\mathbf{t})} \vee \sqrt{\bar{m}(M,\mathbf{m})}\right) M k^{-\frac{1}{2}} =: \bar{D}_{d,M,\mathbf{t},\mathbf{m}} k^{-\frac{1}{2}}, \tag{117}$$

where $\bar{t}\left(M^{-1},\mathbf{t}\right) := \inf\{k : t_k \le M^{-1}\}$. Moreover, note that by definition, $H^2(P,Q) \ge H^2_{\tilde{\mathcal{G}}^{(2)}_{k,m_k,t_k}}(P,Q)$. Then, we have

$$\begin{aligned}
&\left|H^2(P,Q) - H^2_{\tilde{\mathcal{G}}^{(2)}_{k,m_k,t_k}}(P,Q)\right| \\
&= H^2(P,Q) - H^2_{\tilde{\mathcal{G}}^{(2)}_{k,m_k,t_k}}(P,Q) \\
&\le \mathbb{E}_P\left[f_{H^2}(X)\right] - \mathbb{E}_Q\left[\frac{f_{H^2}(Y)}{1 - f_{H^2}(Y)}\right] - \mathbb{E}_P\left[g_{\theta_k^*}(X)\right] + \mathbb{E}_Q\left[\frac{g_{\theta_k^*}(Y)}{1 - g_{\theta_k^*}(Y)}\right] \\
&\le \mathbb{E}_P\left[\left|f_{H^2}(X) - g_{\theta_k^*}(X)\right|\right] + \mathbb{E}_Q\left[\left|\frac{f_{H^2}(Y)}{1 - f_{H^2}(Y)} - \frac{g_{\theta_k^*}(Y)}{1 - g_{\theta_k^*}(Y)}\right|\right]
\end{aligned}$$

$$\leq \bar{D}_{d,M,\mathbf{t},\mathbf{m}} k^{-\frac{1}{2}} + \mathbb{E}_Q\left[\left|\frac{f_{H^2}(Y) - g_{\theta_k^*}(Y)}{(1 - f_{H^2}(Y))(1 - g_{\theta_k^*}(Y))}\right|\right]$$

$$\leq \bar{D}_{d,M,\mathbf{t},\mathbf{m}} k^{-\frac{1}{2}} + M\ t_k^{-1} \bar{D}_{d,M,\mathbf{t},\mathbf{m}} k^{-\frac{1}{2}}, \tag{118}$$

where (118) is due to $1 - g_{\theta^*}(x) \geq t_k$, $(1 - f_{H^2}(x))^{-1} \leq M$ for all $x \in \mathcal{X}$, and (117).

Then, it follows from (113) and (118) that by taking $a_{1,k} = \sqrt{k}\log k$, $ka_{2,k} = a_{3,k} = m_k$, and $\sqrt{k} m_k t_k^{-2} = O\left(n^{(1-\alpha)/2}\right)$ for some $\alpha > 0$, we have

$$\mathbb{E}\left[\left|\hat{H}^2_{\tilde{\mathcal{G}}^{(2)}_{k,m_k,t_k}}(X^n, Y^n) - H^2(P,Q)\right|\right]$$

$$\leq \left|H^2(P,Q) - H^2_{\tilde{\mathcal{G}}^{(2)}_{k,m_k,t_k}}(P,Q)\right| + \mathbb{E}\left[\left|\hat{H}^2_{\tilde{\mathcal{G}}^{(2)}_{k,m_k,t_k}}(X^n, Y^n) - H^2_{\tilde{\mathcal{G}}^{(2)}_{k,m_k,t_k}}(P,Q)\right|\right]$$

$$\leq \bar{D}_{d,M,\mathbf{t},\mathbf{m}} k^{-\frac{1}{2}} + M\ t_k^{-1} \bar{D}_{d,M,\mathbf{t},\mathbf{m}} k^{-\frac{1}{2}} + O\Big(m_k \sqrt{k} t_k^{-2} n^{-\frac{1}{2}}\Big) \tag{119}$$

$$= O_{d,M}\Big(\sqrt{\bar{t}\,(M^{-1}, \mathbf{t})} \vee \sqrt{\bar{m}(M, \mathbf{m})}\ t_k^{-1} k^{-\frac{1}{2}}\Big) + O\Big(m_k \sqrt{k} t_k^{-2} n^{-\frac{1}{2}}\Big).$$

Setting $m_k = 0.5\log k$ and $t_k = \log^{-1} k$ in (119) yields (32), thus completing the proof.

D.1.1 Proof of Lemma 4

Note that Theorem 3 continues to hold with $\mathcal{G}_k(\mathbf{a})$ in (16) and (17) replaced with $\tilde{\mathcal{G}}_k(\mathbf{a}, t)$, since for $\gamma_{H^2}(x) = \frac{x}{1-x}$,

$$\bar{\gamma}'_{\tilde{\mathcal{G}}_k(\mathbf{a}_k, t_k)} = \sup_{\substack{x \in \mathcal{X},\\ g_\theta \in \tilde{\mathcal{G}}_k(\mathbf{a}_k, t_k)}} \gamma'_{H^2}(g_\theta(x)) = \sup_{\substack{x \in \mathcal{X},\\ g_\theta \in \tilde{\mathcal{G}}_k(\mathbf{a}_k, t_k)}} \frac{1}{(1 - g_\theta)^2} \leq \frac{1}{t_k^2},$$

where $\gamma'_{H^2}(\cdot)$ denotes the derivative of γ_{H^2}. This implies that $R_{k,\mathbf{a}_k,\gamma} \leq 2\sqrt{k}\left(t_k^{-2} + 1\right)$, and

$$0 \leq E_{k,\mathbf{a}_k,n,\gamma} \leq 4\sqrt{2} n^{-\frac{1}{2}} k^{\frac{3}{2}} a_{2,k}\left(t_k^{-2} + 1\right) \xrightarrow[n \to \infty]{} 0,$$

for $k, \mathbf{a}_k$, t_k such that $k^{\frac{3}{2}} a_{2,k} t_k^{-2} = O\left(n^{\frac{1-\alpha}{2}}\right)$. It then follows from (17) that for any $k \in \mathbb{N}$, $\delta > 0$, and n sufficiently large,

$$\mathbb{P}\left(\left|\hat{H}^2{}_{\mathcal{G}_k(\mathbf{a}_k)}(X^n, Y^n) - H^2_{\mathcal{G}_k(\mathbf{a}_k)}(P,Q)\right| \geq \delta\right) \leq 2Ce^{-\frac{n(\delta - CE_{k,\mathbf{a}_k,n,\gamma})^2}{16Ca_{2,k}^2 k^2 \left(t_k^{-2}+1\right)^2}}.$$

Then, (112) and (113) follows using similar steps used to prove (77) (see (97)) and (78) (see (98)) in Theorem 4, respectively. This completes the proof.