A Appendix: Remaining Proofs

In this section, we present the proofs of Theorems 3.1 and 3.4.

A.1 Proof of Theorem 3.1

In the following, we quote a version of Talagrand's inequality due to Bousquet (2002) from Steinwart and Christmann (2008, Theorem 7.5) and a (slightly simplified) bound on the expected suprema of empirical processes indexed by Vapnik-Červonenkis (VC) classes of functions, from Giné and Guillou (2001, Proposition 2.1). Both will be used to prove Theorem 3.1.

Theorem A.1. *Let (Z,P) be a probability space and $\mathcal{G}$ be a set of measurable functions from Z to $\mathbb{R}$. Furthermore, let $B \geq 0$ and $\sigma \geq 0$ be constants such that $\mathbb{E}_P g = 0$, $\mathbb{E}_P g^2 \leq \sigma^2$, and $\|g\|_\infty \leq B$ for all $g \in \mathcal{G}$. For $n \geq 1$, define $G : Z^n \to \mathbb{R}$ by*

$$G(z) := \sup_{g\in\mathcal{G}} \left| \frac{1}{n} \sum_{j=1}^{n} g(z_j) \right|, \quad z = (z_1, \ldots, z_n) \in Z^n.$$

Then, for all $\tau > 0$, we have

$$P^n\left(\left\{z\in Z^n \,:\, G(z) \geq 4\mathbb{E}_{P^n} G + \sqrt{\frac{2\tau\sigma^2}{n}} + \frac{\tau B}{n}\right\}\right) \leq e^{-\tau}.$$

Theorem A.2. *Let (Z,P) be a probability space and $\mathcal{G}$ be a set of measurable functions from Z to $\mathbb{R}$. Furthermore, let $B \geq 0$ and $0 \leq \sigma \leq B$ be constants such that $\mathbb{E}_P g^2 \leq \sigma^2$, and $\|g\|_\infty \leq B$ for all $g \in \mathcal{G}$. Suppose $\mathcal{G}$ is a uniformly bounded VC-class, i.e., there exist positive numbers A and ν such that, for every probability measure P on Z and every $0 < \epsilon \leq B$, the covering numbers satisfy*

$$\mathcal{N}(\mathcal{G}, L_2(P), \epsilon) \leq \left(\frac{AB}{\epsilon}\right)^{\nu}.$$

Then there exists a universal constant C such that G defined as in Theorem A.1 satisfies

$$\mathbb{E}_{P^n} G \leq C\left(\frac{\nu B}{n}\log\frac{AB}{\sigma} + \sqrt{\frac{\nu\sigma^2}{n}\log\frac{AB}{\sigma}}\right). \quad (15)$$

Given a measurable $g : \mathbb{R}^d \to \mathbb{R}$ and a $\delta > 0$ we define the function $g_\delta : \mathbb{R}^d \to \mathbb{R}$ by $g_\delta(x) := g(x/\delta)$, $x \in \mathbb{R}^d$. The following lemma, which establishes a stability of covering number bounds under this operation, will also be needed in the proof of Theorem 3.1.

Lemma A.3. *Let $\mathcal{G}$ be set of measurable functions $g : \mathbb{R}^d \to \mathbb{R}$ such that there exists a constant $B \geq 0$ with $\|g\|_\infty \leq B$ for all $g \in \mathcal{G}$. For $\delta > 0$, we write $\mathcal{G}_\delta := \{g_\delta : g \in \mathcal{G}\}$. Then, for all $\epsilon \in (0, B]$ and all $\delta > 0$, we have*

$$\sup_P \mathcal{N}(\mathcal{G}, L_2(P), \epsilon) = \sup_P \mathcal{N}(\mathcal{G}_\delta, L_2(P), \epsilon),$$

where the suprema are taken over all probability measures P on $\mathbb{R}^d$.

Proof. We only prove "$\leq$", the converse inequality can be shown analogously. Let us fix $\epsilon, \delta > 0$ and a distribution P on $\mathbb{R}^d$. We define a new distribution P' on $\mathbb{R}^d$ by $P'(A) := P(\frac{1}{\delta}A)$ for all measurable $A \subset \mathbb{R}^d$. Furthermore, let $\mathcal{C}'$ be an ϵ-net of $\mathcal{G}_\delta$ with respect to $L_2(P')$. For $\mathcal{C} := \mathcal{C}'_{1/\delta}$, we then have $|\mathcal{C}| = |\mathcal{C}'|$, and hence it suffices to show that $\mathcal{C}$ is an ϵ-net of $\mathcal{G}$ with respect to $L_2(P)$. To this end, we fix a $g \in \mathcal{G}$. Then $g_\delta \in \mathcal{G}_\delta$, and hence there exists an $h' \in \mathcal{C}'$ with $\|g_\delta - h'\|_{L_2(P')} \leq \epsilon$. Moreover, we have $h := h'_{1/\delta} \in \mathcal{C}$, and since the definition of P' ensures $\mathbb{E}_{P'} f_\delta = \mathbb{E}_P f$ for all measurable $f : \mathbb{R}^d \to [0, \infty)$, we obtain

$$\|g - h\|_{L_2(P)} = \|g_\delta - h_\delta\|_{L_2(P')} = \|g_\delta - h'\|_{L_2(P')} \leq \epsilon,$$

i.e. $\mathcal{C}$ is an ϵ-net of $\mathcal{G}$ with respect to $L_2(P)$. □

We further need the following result, which is a reformulation of van der Vaart and Wellner (1996, Theorem 2.6.4).

Theorem A.4. *Let $\mathcal{A}$ be a set of subsets of Z that has finite VC-dimension V. Then the corresponding set of indicator functions $\mathcal{G} := \{\mathbf{1}_A : A \in \mathcal{A}\}$ is a uniformly bounded VC-class and the corresponding VC-characteristics A and ν only depend on V.*

With these preparation we are now able to establish the following generalization of Theorem 3.1. Applying this generalization to K of the form (3) immediately proves Theorem 3.1.

Proposition A.5. *Let P be a probability measure on $\mathbb{R}^d$ with a bounded Lebesgue density h and K be a real-valued function on X such that $K \in L_\infty(\mathbb{R}^d) \cap L_2(\mathbb{R}^d)$. Suppose that*

$$\mathcal{F} := \{K(x - \cdot) \,:\, x \in X\}$$

is a uniformly bounded VC-class. Then, there exists a positive constant C only depending on K, h and VC-characteristics A and ν of $\mathcal{F}$ such that, for all $n \geq 1$, $\delta > 0$, and $\tau > 0$ we have

$$P^n\left(\left\{D \in X^n \,:\, \|\bar{h}_{D,\delta} - \bar{h}_{P,\delta}\|_\infty < \frac{C}{n\delta^d}\log\frac{C}{\delta} + \sqrt{\frac{C}{n\delta^d}\log\frac{C}{\delta}} + \frac{\tau C}{n\delta^d} + \frac{C\sqrt{\tau}}{\sqrt{n\delta^d}}\right\}\right) \geq 1 - e^{-\tau}.$$

Proof. Let us assume without loss of generality that $\|K\|_\infty \leq 1$. We define $k_{x,\delta} := \delta^{-d} K\left(\frac{x-\cdot}{\delta}\right)$ and $f_{x,\delta} := k_{x,\delta} - \mathbb{E}_P k_{x,\delta}$. Then it is easy to check that $\mathbb{E}_P f_{x,\delta} = 0$ and $\|f_{x,\delta}\|_\infty \leq 2\delta^{-d}$ for all $x \in X$ and $\delta > 0$. Moreover, we have

$$\begin{aligned}\mathbb{E}_P f_{x,\delta}^2 &\leq \mathbb{E}_P k_{x,\delta}^2 = \delta^{-2d} \int_{\mathbb{R}^d} K^2\left(\frac{x-y}{\delta}\right) h(y)\, dy \\ &\leq \delta^{-d} \|h\|_\infty \|K\|_2^2\end{aligned}$$

for all $x \in X$ and $\delta > 0$, where the norm $\|K\|_2$ is with respect to the Lebesgue measure on $\mathbb{R}^d$. In addition, we have

$$\frac{1}{n}\sum_{i=1}^n f_{x,\delta}(x_i) = \bar{h}_{D,\delta}(x) - \bar{h}_{P,\delta}(x),$$

where $\bar{h}_{P,\delta}$ and $\bar{h}_{D,\delta}$ are defined in (1) and (2) respectively. Applying Theorem A.1 to $\mathcal{G} := \{f_{x,\delta} : x \in \mathbb{R}^d\}$, we hence obtain, for all $\delta > 0$, $\tau > 0$, and $n \geq 1$, that

$$\begin{aligned}\|\bar{h}_{D,\delta} - \bar{h}_{P,\delta}\|_\infty &< 4\mathbb{E}_{P^n} \|\bar{h}_{D,\delta} - \bar{h}_{P,\delta}\|_\infty + \frac{2\tau}{n\delta^d} \\ &\quad + \sqrt{\frac{2\tau \|h\|_\infty \|K\|_2^2}{n\delta^d}} \qquad (16)\end{aligned}$$

holds with probability P^n not smaller than $1 - e^{-\tau}$. It thus remains to bound the term $\mathbb{E}_{P^n} \|\bar{h}_{D,\delta} - \bar{h}_{P,\delta}\|_\infty$. Note that since $\mathcal{F}$ is a uniformly bounded VC-class, so is $\tilde{\mathcal{F}} := \{f - a : f \in \mathcal{F}, a \in [-1,1]\}$, i.e. there exist positive numbers A and ν such that

$$\sup_P \mathcal{N}\big(\tilde{\mathcal{F}}, L_2(P), \epsilon\big) \leq \left(\frac{2A}{\epsilon}\right)^\nu$$

for all $0 < \epsilon \leq 2$. For $\delta > 0$, we further have $\delta^d \mathcal{G} \subset \tilde{\mathcal{F}}_\delta$, and hence Lemma A.3 implies

$$\mathcal{N}\left(\delta^d \mathcal{G}, L_2(P), \epsilon\right) \leq \mathcal{N}(\tilde{\mathcal{F}}_\delta, L_2(P), \epsilon) \leq \left(\frac{2A}{\epsilon}\right)^\nu,$$

for all probability measures P on $\mathbb{R}^d$ and all $0 < \epsilon \leq 2$. Now, our very first estimates show that every $g \in \tilde{\mathcal{G}} := \delta^d \mathcal{G}$ satisfies $\|g\|_\infty \leq 2$ and $\mathbb{E}_P g^2 \leq \delta^d \|h\|_\infty \|K\|_2^2$, and hence Theorem A.2 yields

$$\begin{aligned}\mathbb{E}_{P^n} \sup_{g \in \tilde{\mathcal{G}}} \left| \frac{1}{n} \sum_{j=1}^n g(X_j) \right| &\leq C\left(\frac{2\nu}{n} \log \frac{2A}{\sqrt{\delta^d \|h\|_\infty \|K\|_2^2}} \right. \\ &\quad \left. + \sqrt{\frac{\nu \delta^d \|h\|_\infty \|K\|_2^2}{n} \log \frac{2A}{\sqrt{\delta^d \|h\|_\infty \|K\|_2^2}}} \right).\end{aligned}$$

Multiplying both sides by δ^{-d}, we obtain

$$\begin{aligned}\mathbb{E}_{P^n} \|\bar{h}_{D,\delta} - \bar{h}_{P,\delta}\|_\infty &\leq C\left(\frac{2\nu}{n\delta^d} \log \frac{2A}{\sqrt{\delta^d \|h\|_\infty \|K\|_2^2}} \right. \\ &\quad \left. + \sqrt{\frac{\nu \|h\|_\infty \|K\|_2^2}{n\delta^d} \log \frac{2A}{\sqrt{\delta^d \|h\|_\infty \|K\|_2^2}}} \right),\end{aligned}$$

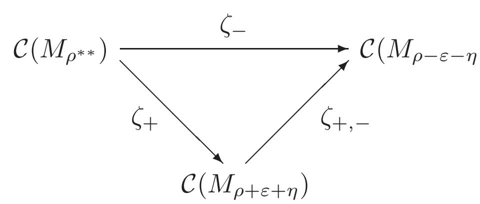
which, when used in (16), yields the result. □

A.2 Proof of Theorem 3.4

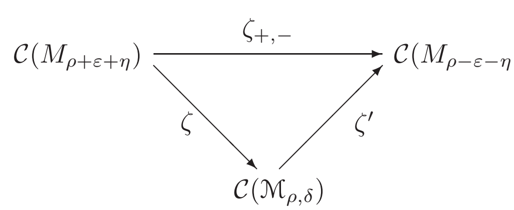
Proof of Theorem 3.4. i). Let $D \in X^n$ be a dataset such that $\|\bar{h}_{D,\delta} - \bar{h}_{P,\delta}\|_\infty < \varepsilon$. Moreover, let $\rho \geq 0$ be the current level that is considered by Algorithm 1. Then, Theorem 3.3 shows that, for $\rho \in [0, \rho^{**} - 3\varepsilon - 3\eta]$, Algorithm 1 identifies exactly the topologically connected components of $\mathcal{M}_{\rho,\delta}$ in its loop that belong to the set $\zeta(\mathcal{C}(M_{\rho+\varepsilon+\eta}))$, where $\zeta : \mathcal{C}(M_{\rho+\varepsilon+\eta}) \to \mathcal{C}(\mathcal{M}_{\rho,\delta})$ is the top-CCRM. In the following, we thus consider the set $\zeta(\mathcal{C}(M_{\rho+\varepsilon+\eta}))$ for $\rho \in [0, \rho^{**} - 3\varepsilon - 3\eta]$.

Let us first consider the case $\rho \in [0, \rho^* - \varepsilon - \eta)$. Then, (c_1) and (c_3) together with the assumed $\rho + \varepsilon + \eta < \rho^*$ show $|\mathcal{C}(M_{\rho+\varepsilon+\eta})| = 1$. This yields $|\zeta(\mathcal{C}(M_{\rho+\varepsilon+\eta}))| = 1$, and hence Algorithm 1 does not stop. Consequently, we have $\rho^*(D) \geq \rho^* - \varepsilon - \eta$.

Let us now consider the case $\rho \in [\rho^* + \varepsilon^* + \eta^* + \varepsilon + \eta, \rho^* + \varepsilon^* + \eta^* + 2\varepsilon + 2\eta]$. Then we first note that Algorithm 1 actually inspects such an ρ, since it iteratively inspects all $\rho = i\varepsilon + i\eta$, $i = 0, 1, \ldots$, and the width of the interval above is $\varepsilon + \eta$. Moreover, our assumptions on ε^*, η^*, ε and η guarantee $\rho^* + \varepsilon^* + \eta^* + 2\varepsilon + 2\eta \leq \rho^{**} - 3\varepsilon - 3\eta$ and hence we have $\rho \in [\rho^* + \varepsilon^* + \eta^* + \varepsilon + \eta, \rho^{**} - 3\varepsilon - 3\eta]$. Let us write $\zeta_+ : \mathcal{C}(M_{\rho^{**}}) \to \mathcal{C}(M_{\rho+\varepsilon+\eta})$, $\zeta_- : \mathcal{C}(M_{\rho^{**}}) \to \mathcal{C}(M_{\rho-\varepsilon-\eta})$, and $\zeta_{+,-} : \mathcal{C}(M_{\rho+\varepsilon+\eta}) \to \mathcal{C}(M_{\rho-\varepsilon-\eta})$ for the top-CCRMs between the involved sets. Using the composition property of top-CCRMs in (b_2), we then obtain the following diagram:



Moreover, we have $\rho - \varepsilon - \eta \geq \rho^* + \varepsilon^* + \eta^* > \rho^*$ and $\rho + \varepsilon + \eta > \rho^*$, and hence (c_1) and (c_2) show that $|\mathcal{C}(M_{\rho-\varepsilon-\eta})| = 2$ and $|\mathcal{C}(M_{\rho+\varepsilon+\eta})| = 2$. Consequently, (c_3) ensures that the maps ζ_+ and ζ_- are bijective. Consequently, $\zeta_{+,-}$ is bijective. Let us further consider the top-CCRM $\zeta' : \mathcal{C}(\mathcal{M}_{\rho,\delta}) \to \mathcal{C}(M_{\rho-\varepsilon-\eta})$. Then the composition property of top-CCRMS in (b_2)—yields another diagram:



Since $\zeta_{+,-}$ is bijective, we then find that ζ is injective, and since we have already seen that $M_{\rho+\varepsilon+\eta}$ has two top-connected components, we conclude that $\zeta(\mathcal{C}(M_{\rho+\varepsilon+\eta}))$ contains two elements. Consequently, the stopping criterion of Algorithm 1 is satisfied, that is, $\rho^*(D) \le \rho^* + \varepsilon^* + \eta^* + 2\varepsilon + 2\eta$.

ii). Theorem 3.3 shows that in its last loop Algorithm 1 identifies exactly the topologically connected components of $\mathcal{M}_{\rho^*(D),\delta}$ that belong to the set $\zeta_\varepsilon(\mathcal{C}(M_{\rho^*(D)+\varepsilon+\eta}))$, where $\zeta_\varepsilon : \mathcal{C}(M_{\rho^*(D)+\varepsilon+\eta}) \to \mathcal{C}(\mathcal{M}_{\rho^*(D),\delta})$ is the top-CCRM. Moreover, since Algorithm 1 stops at $\rho^*(D)$, we have $|\zeta_\varepsilon(\mathcal{C}(M_{\rho^*(D)+\varepsilon+\eta}))| \ne 1$ and thus $|\mathcal{C}(M_{\rho^*(D)+\varepsilon+\eta})| \ne 1$. From $\rho^*(D) + \varepsilon + \eta \le \rho^{**}$ and (c_1) we thus conclude that $|\mathcal{C}(M_{\rho^*(D)+\varepsilon+\eta})| = 2$. For later purposes, note that the latter implies the injectivity of ζ_ε. In addition, since $|\mathcal{C}(M_{\rho^*(D)+\varepsilon+\eta})| = 2$, (c_3) yields $\zeta_{-,\rho^*(D)+\varepsilon+\eta} : \mathcal{C}(M_{\rho^{**}}) \to \mathcal{C}(M_{\rho^*(D)+\varepsilon+\eta})$ is bijective. Since $\rho^*(D) + 3\varepsilon + 3\eta > \rho^*$, it follows from (c_1)–(c_3) that we have $\zeta_{+,\rho^*(D)+\varepsilon+\eta} : \mathcal{C}(M_{\rho^{**}}) \to \mathcal{C}(M_{\rho^*(D)+3\varepsilon+3\eta})$ is bijective. Using the composition property of top-CCRMS in (b_2), we obtain that $\zeta_{+,-,\rho^*(D)+\varepsilon+\eta} : \mathcal{C}(M_{\rho^*(D)+3\varepsilon+3\eta}) \to \mathcal{C}(M_{\rho^*(D)+\varepsilon+\eta})$ is bijective, and hence $|\mathcal{C}(M_{\rho^*(D)+3\varepsilon+3\eta})| = 2$. Let us now consider the following commutative diagram:

$$\begin{array}{ccc}
\mathcal{C}(M_{\rho^*(D)+3\varepsilon+3\eta}) & \xrightarrow{\zeta_{+,-,\rho^*(D)+\varepsilon+\eta}} & \mathcal{C}(M_{\rho^*(D)+\varepsilon+\eta}) \\
\downarrow{\scriptstyle \zeta_{3\varepsilon}} & & \downarrow{\scriptstyle \zeta_{\varepsilon}} \\
\mathcal{C}(\mathcal{M}_{\rho^*(D)+2\varepsilon+2\eta,\delta}) & \xrightarrow[\zeta_f]{} & \mathcal{C}(\mathcal{M}_{\rho^*(D),\delta})
\end{array}$$

where again, all occurring maps are the top-CCRMs between the respective sets. Now we have already shown that ζ_ε is injective and that $\zeta_{+,-,\rho^*(D)+\varepsilon+\eta}$ is bijective. Consequently, $\zeta_{3\varepsilon}$ is injective.

iii). Follows from Theorem 3.3 and $\rho^*(D) + 2\varepsilon + 2\eta \le \rho^{**} - 3\varepsilon - 3\eta$.

iv). Since $\rho^*(D) + 3\varepsilon + 3\eta > \rho^*(D) + \varepsilon + \eta > \rho^*$, by (c_1)–(c_3), we see that the maps $\zeta_{-,\rho^{**}}$ and $\zeta_{+,\rho^{**}}$ are bijective. Therefore $\zeta_{+,-,\rho^*(D)+\varepsilon+\eta}$ is bijective and the diagram follows. □