Supplement

The supplement contains the detailed proofs of our results (Section A), a few technical lemmas used during these arguments (Section B), the McDiarmid inequality for self-containedness (Section C), and the pseudocode of the two-sample test performed in Experiment-2 (Section D).

A. Proofs of Theorem 1 and Theorem 2

This section contains the detailed proofs of Theorem 1 (Section A.1) and Theorem 2 (Section A.2).

A.1. Proof of Theorem 1

The structure of the proof is as follows:

1. We show that $\|\hat{\mu}_{\mathbb{P},Q} - \mu_{\mathbb{P}}\|_K \le (1+\sqrt{2}) r_{Q,N}$, where $r_{Q,N} = \sup_{f\in B_K} \text{MON}_Q\Big[\underbrace{\langle f, K(\cdot,x)-\mu_{\mathbb{P}}\rangle_K}_{f(x)-\mathbb{P}f}\Big]$, i.e. the analysis can be reduced to B_K.
2. Then $r_{Q,N}$ is bounded using empirical processes.

Step-1: Since $\mathcal{H}_K$ is an inner product space, for any $f \in \mathcal{H}_K$

$$\begin{aligned}&\|f-K(\cdot,x)\|_K^2 - \|\mu_{\mathbb{P}} - K(\cdot,x)\|_K^2 \\ &= \|f-\mu_{\mathbb{P}}\|_K^2 - 2\langle f-\mu_{\mathbb{P}}, K(\cdot,x)-\mu_{\mathbb{P}}\rangle_K. \qquad (14)\end{aligned}$$

Hence, by denoting $e = \hat{\mu}_{\mathbb{P},Q} - \mu_{\mathbb{P}}$, $\tilde{g} = g - \mu_{\mathbb{P}}$ we get

$$\begin{aligned}
&\|e\|_K^2 - 2r_{Q,N}\|e\|_K \\
&\overset{(a)}{\le} \|e\|_K^2 - 2\text{MON}_Q\left[\left\langle \frac{e}{\|e\|_K}, K(\cdot,x)-\mu_{\mathbb{P}}\right\rangle\right]_K \|e\|_K \\
&\overset{(b)}{\le} \text{MON}_Q\left[\|e\|_K^2 - 2\left\langle \frac{e}{\|e\|_K}, K(\cdot,x)-\mu_{\mathbb{P}}\right\rangle_K \|e\|_K\right] \\
&\overset{(c)}{\le} \text{MON}_Q\left[\|\hat{\mu}_{\mathbb{P},Q} - K(\cdot,x)\|_K^2 - \|\mu_{\mathbb{P}} - K(\cdot,x)\|_K^2\right] \\
&\overset{(d)}{\le} \sup_{g\in\mathcal{H}_K} \text{MON}_Q\left[\|\hat{\mu}_{\mathbb{P},Q} - K(\cdot,x)\|_K^2 - \|g - K(\cdot,x)\|_K^2\right] \\
&\overset{(e)}{\le} \sup_{g\in\mathcal{H}_K} \text{MON}_Q\left[\|\mu_{\mathbb{P}} - K(\cdot,x)\|_K^2 - \|g - K(\cdot,x)\|_K^2\right] \\
&\overset{(f)}{=} \sup_{g\in\mathcal{H}_K}\Big\{2\text{MON}_Q\Big[\underbrace{\langle \tilde{g}, K(\cdot,x)-\mu_{\mathbb{P}}\rangle_K}_{\|\tilde{g}\|_K\left\langle \frac{\tilde{g}}{\|\tilde{g}\|_K}, K(\cdot,x)-\mu_{\mathbb{P}}\right\rangle_K}\Big] - \|\tilde{g}\|_K^2\Big\} \\
&\overset{(g)}{=} \sup_{g\in\mathcal{H}_K}\left\{2\|\tilde{g}\|_K r_{Q,N} - \|\tilde{g}\|_K^2\right\} \overset{(h)}{\le} r_{Q,N}^2, \qquad (15)
\end{aligned}$$

where we used in (a) the definition of $r_{Q,N}$, (b) the linearity[7] of $\text{MON}_Q[\cdot]$, (c) Eq. (14), (d) $\sup_g$, (e) the definition of $\hat{\mu}_{\mathbb{P},Q}$, (f) Eq. (14) and the linearity of $\text{MON}_Q[\cdot]$, (g) the definition of $r_{Q,N}$. In step (h), by denoting $a = \|\tilde{g}\|_K$, $r = r_{Q,N}$, the argument of the sup takes the form $2ar - a^2$; $2ar - a^2 \le r^2 \Leftrightarrow 0 \le r^2 - 2ar + a^2 = (r-a)^2$.

In Eq. (15), we obtained an equation $a^2 - 2ra \le r^2$ where $a := \|e\|_K \ge 0$. Hence $r^2 + 2ra - a^2 \ge 0$, $r_{1,2} = \left[-2a \pm \sqrt{4a^2+4a^2}\right]/2 = \left(-1\pm\sqrt{2}\right)a$, thus by the non-negativity of a, $r \ge (-1+\sqrt{2})a$, i.e., $a \le \frac{r}{\sqrt{2}-1} = (\sqrt{2}+1)r$. In other words, we arrived at

$$\|\hat{\mu}_{\mathbb{P},Q} - \mu_{\mathbb{P}}\|_K \le \left(1+\sqrt{2}\right) r_{Q,N}. \qquad (16)$$

It remains to upper bound $r_{Q,N}$.

Step-2: Our goal is to provide a probabilistic bound on

$$\begin{aligned}r_{Q,N} &= \sup_{f\in B_K} \text{MON}_Q\left[x \mapsto \langle f, K(\cdot,x)-\mu_{\mathbb{P}}\rangle_K\right] \\ &= \sup_{f\in B_K} \operatorname*{med}_{q\in[Q]}\Big\{\underbrace{\left\langle f, \mu_{S_q} - \mu_{\mathbb{P}}\right\rangle_K}_{=:r(f,q)}\Big\}.\end{aligned}$$

The N_c corrupted samples can affect (at most) N_c of the $(S_q)_{q\in[Q]}$ blocks. Let $U := [Q]\backslash C$ stand for the indices of the uncorrupted sets, where $C := \{q\in[Q] : \exists n_j \text{ s.t. } n_j \in S_q,\ j\in[N_c]\}$ contains the indices of the corrupted sets. If

$$\forall f\in B_K : \underbrace{|\{q\in U : r(f,q)\ge\epsilon\}|}_{\sum_{q\in U}\mathbb{I}_{r(f,q)\ge\epsilon}} + N_c \le \frac{Q}{2}, \qquad (17)$$

then for $\forall f \in B_K$, $\text{med}_{q\in[Q]}\{r(f,q)\} \le \epsilon$, i.e. $\sup_{f\in B_K}\text{med}_{q\in[Q]}\{r(f,q)\} \le \epsilon$. Thus, our task boils down to controlling the event in (17) by appropriately choosing ϵ.

- **Controlling** $r(f,q)$: For any $f\in B_K$ the random variables $\langle f, k(\cdot,x_i)-\mu_{\mathbb{P}}\rangle_{\mathcal{H}_K} = f(x_i) - \mathbb{P}f$ are independent, have zero mean, and

$$\begin{aligned}&\mathbb{E}_{x_i\sim\mathbb{P}}\langle f, k(\cdot,x_i)-\mu_{\mathbb{P}}\rangle_K^2 = \langle f, \Sigma_{\mathbb{P}} f\rangle_K \\ &\le \|f\|_K\|\Sigma_{\mathbb{P}} f\|_K \le \|f\|_K^2\|\Sigma_{\mathbb{P}}\| = \|\Sigma_{\mathbb{P}}\| \qquad (18)\end{aligned}$$

  using the reproducing property of the kernel and the covariance operator, the Cauchy-Schwarz (CBS) inequality and $\|f\|_{\mathcal{H}_K} = 1$.
  For a zero-mean random variable z by the Chebyshev's inequality $\mathbb{P}(z>a) \le \mathbb{P}(|z|>a) \le \mathbb{E}\left(z^2\right)/a^2$, which implies $\mathbb{P}\left(z > \sqrt{\mathbb{E}\left(z^2\right)/\alpha}\right) \le \alpha$ by a $\alpha = \mathbb{E}\left(z^2\right)/a^2$ substitution. With $z := r(f,q)$ $(q\in U)$, using $\mathbb{E}\left[z^2\right] = \mathbb{E}\left\langle f, \mu_{S_q} - \mu_{\mathbb{P}}\right\rangle_K^2 = \frac{Q}{N}\mathbb{E}_{x_i\sim\mathbb{P}}\langle f, k(\cdot,x_i)-\mu_{\mathbb{P}}\rangle_K^2$ and Eq. (18) one gets that for all $f\in B_K$, $\alpha\in(0,1)$ and $q\in U$: $\mathbb{P}\left(r(f,q) > \sqrt{\frac{\|\Sigma_{\mathbb{P}}\|Q}{\alpha N}}\right) \le \alpha$. This means $\mathbb{P}\left(r(f,q) > \frac{\epsilon}{2}\right) \le \alpha$ with $\epsilon \ge 2\sqrt{\frac{\|\Sigma_{\mathbb{P}}\|Q}{\alpha N}}$.

[7] $\text{MON}_Q[c_1 + c_2 f] = c_1 + c_2\text{MON}_Q[f]$ for any $c_1, c_2\in\mathbb{R}$.

- **Reduction to** ϕ: As a result
$$\sum_{q\in U}\mathbb{P}\left(r(f,q)\geq\frac{\epsilon}{2}\right)\leq |U|\alpha$$
happens if and only if
$$\sum_{q\in U}\mathbb{I}_{r(f,q)\geq\epsilon} \leq |U|\alpha+\sum_{q\in U}\Big[\mathbb{I}_{r(f,q)\geq\epsilon}-\underbrace{\mathbb{P}\left(r(f,q)\geq\frac{\epsilon}{2}\right)}_{\mathbb{E}\left[\mathbb{I}_{r(f,q)\geq\frac{\epsilon}{2}}\right]}\Big]=:A.$$
Let us introduce $\phi: t\in\mathbb{R}\to(t-1)\mathbb{I}_{1\leq t\leq 2}+\mathbb{I}_{t\geq 2}$. ϕ is 1-Lipschitz and satisfies $\mathbb{I}_{2\leq t}\leq\phi(t)\leq\mathbb{I}_{1\leq t}$ for any $t\in\mathbb{R}$. Hence, we can upper bound A as
$$A\leq |U|\alpha+\sum_{q\in U}\left[\phi\left(\frac{2r(f,q)}{\epsilon}\right)-\mathbb{E}\phi\left(\frac{2r(f,q)}{\epsilon}\right)\right]$$
by noticing that $\epsilon\leq r(f,q)\Leftrightarrow 2\leq 2r(f,q)/\epsilon$ and $\epsilon/2\leq r(f,q)\Leftrightarrow 1\leq 2r(f,q)/\epsilon$, and by using the $\mathbb{I}_{2\leq t}\leq\phi(t)$ and the $\phi(t)\leq\mathbb{I}_{1\leq t}$ bound, respectively. Taking supremum over B_K we arrive at
$$\sup_{f\in B_K}\sum_{q\in U}\mathbb{I}_{r(f,q)\geq\epsilon} \leq |U|\alpha+\underbrace{\sup_{f\in B_K}\sum_{q\in U}\left[\phi\left(\frac{2r(f,q)}{\epsilon}\right)-\mathbb{E}\phi\left(\frac{2r(f,q)}{\epsilon}\right)\right]}_{=:Z}.$$

- **Concentration of** Z **around its mean:** Notice that Z is a function of x_V, the samples in the uncorrupted blocks; $V=\cup_{q\in U}S_q$. By the bounded difference property of Z (Lemma 4) for any $\beta>0$, the McDiarmid inequality (Lemma 6; we choose $\tau:=Q\beta^2/8$ to get linear scaling in Q on the r.h.s.) implies that
$$\mathbb{P}\left(Z<\mathbb{E}_{x_V}[Z]+Q\beta\right)\geq 1-e^{-\frac{Q\beta^2}{8}}.$$

- **Bounding** $\mathbb{E}_{x_V}[Z]$: Let $M=N/Q$ denote the number of elements in S_q-s. The $\mathcal{G}=\{g_f: f\in B_K\}$ class with $g_f:\mathcal{X}^M\to\mathbb{R}$ and $\mathbb{P}_M:=\frac{1}{M}\sum_{m=1}^M\delta_{u_m}$ defined as
$$g_f(u_{1:M})=\phi\left(\frac{\langle f,\mu_{\mathbb{P}_M}-\mu_{\mathbb{P}}\rangle_K}{\epsilon}\right)$$
is uniformly bounded separable Carathéodory (Lemma 5), hence the symmetrization technique (Steinwart & Christmann, 2008, Prop. 7.10), (Ledoux & Talagrand, 1991) gives
$$\mathbb{E}_{x_V}[Z]\leq 2\mathbb{E}_{x_V}\mathbb{E}_{\mathbf{e}}\sup_{f\in B_K}\left|\sum_{q\in U}e_q\phi\left(\frac{2r(f,q)}{\epsilon}\right)\right|,$$
where $\mathbf{e}=(e_q)_{q\in U}\in\mathbb{R}^{|U|}$ with i.i.d. Rademacher entries $[\mathbb{P}(e_q=\pm 1)=\frac{1}{2}\ (\forall q)]$.

- **Discarding** ϕ: Since $\phi(0)=0$ and ϕ is 1-Lipschitz, by Talagrand's contraction principle of Rademacher processes (Ledoux & Talagrand, 1991), (Koltchinskii, 2011, Theorem 2.3) one gets
$$\mathbb{E}_{x_V}\mathbb{E}_{\mathbf{e}}\sup_{f\in B_K}\left|\sum_{q\in U}e_q\phi\left(\frac{2r(f,q)}{\epsilon}\right)\right| \leq 2\mathbb{E}_{x_V}\mathbb{E}_{\mathbf{e}}\sup_{f\in B_K}\left|\sum_{q\in U}e_q\frac{2r(f,q)}{\epsilon}\right|.$$

- **Switching from** $|U|$ **to** N **terms**: Applying an other symmetrization [(a)], the CBS inequality, $f\in B_K$, and the Jensen inequality
$$\begin{aligned}
&\mathbb{E}_{x_V}\mathbb{E}_{\mathbf{e}}\sup_{f\in B_K}\left|\sum_{q=1}^{Q}e_q\frac{r(f,q)}{\epsilon}\right| \\
&\overset{(a)}{\leq}\frac{2Q}{\epsilon N}\mathbb{E}_{x_V}\mathbb{E}_{\mathbf{e}'}\Bigg[\sup_{f\in B_K}\Big|\underbrace{\sum_{n\in V}e'_n\langle f,K(\cdot,x_n)-\mu_{\mathbb{P}}\rangle_K}_{=\left\langle f,\sum_{n\in V}e'_n[K(\cdot,x_n)-\mu_{\mathbb{P}}]\right\rangle_K}\Big|\Bigg] \\
&\leq\frac{2Q}{\epsilon N}\mathbb{E}_{x_V}\mathbb{E}_{\mathbf{e}'}\Bigg[\sup_{f\in B_K}\underbrace{\|f\|_K}_{=1}\left\|\sum_{n\in V}e'_n[K(\cdot,x_n)-\mu_{\mathbb{P}}]\right\|_K\Bigg] \\
&=\frac{2Q}{\epsilon N}\mathbb{E}_{x_V}\mathbb{E}_{\mathbf{e}'}\left\|\sum_{n\in V}e'_n[K(\cdot,x_n)-\mu_{\mathbb{P}}]\right\|_K \\
&\leq\frac{2Q}{\epsilon N}\sqrt{\mathbb{E}_{x_V}\mathbb{E}_{\mathbf{e}'}\left\|\sum_{n\in V}e'_n[K(\cdot,x_n)-\mu_{\mathbb{P}}]\right\|_K^2} \\
&\overset{(b)}{=}\frac{2Q\sqrt{|V|\operatorname{Tr}(\Sigma_{\mathbb{P}})}}{\epsilon N}.
\end{aligned}$$
In (a), we proceed as follows:
$$\begin{aligned}
&\mathbb{E}_{x_V}\mathbb{E}_{\mathbf{e}}\sup_{f\in B_K}\left|\sum_{q\in U}e_q\frac{r(f,q)}{\epsilon}\right| \\
&=\mathbb{E}_{x_V}\mathbb{E}_{\mathbf{e}}\sup_{f\in B_K}\left|\sum_{q\in U}e_q\frac{\langle f,\mu_{S_q}-\mu_{\mathbb{P}}\rangle_K}{\epsilon}\right| \\
&\overset{(c)}{\leq}\frac{2Q}{N\epsilon}\mathbb{E}_{x_V}\mathbb{E}_{\mathbf{e}}\mathbb{E}_{\mathbf{e}'}\sup_{f\in B_K}\left|\sum_{n\in V}e'_ne''_n\langle f,K(\cdot,x_n)-\mu_{\mathbb{P}}\rangle_K\right| \\
&=\frac{2Q}{N\epsilon}\mathbb{E}_{x_V}\mathbb{E}_{\mathbf{e}'}\sup_{f\in B_K}\left|\sum_{n\in V}e'_n\langle f,K(\cdot,x_n)-\mu_{\mathbb{P}}\rangle_K\right|,
\end{aligned}$$
where in (c) we applied symmetrization, $\mathbf{e}'=(e'_n)_{n\in V}\in\mathbb{R}^{|V|}$ with i.i.d. Rademacher entries, $e''_n=e_q$ if $n\in S_q$ $(q\in U)$, and

we used that $\left(e'_n e''_n \left\langle f, K(\cdot,x_n)-\mu_{\mathbb{P}}\right\rangle_K\right)_{n\in V} \overset{\text{distr}}{=} \left(e'_n \left\langle f, K(\cdot,x_n)-\mu_{\mathbb{P}}\right\rangle_K\right)_{n\in V}$.
In step (b), we had

$$\begin{aligned}
&\mathbb{E}_{x_V}\mathbb{E}_{\mathbf{e}'}\left\|\sum_{n\in V} e'_n\left[K(\cdot,x_n)-\mu_{\mathbb{P}}\right]\right\|_K^2\\
&=\mathbb{E}_{x_V}\mathbb{E}_{\mathbf{e}'}\sum_{n\in V}\left[e'_n\right]^2\left\langle K(\cdot,x_n)-\mu_{\mathbb{P}},K(\cdot,x_n)-\mu_{\mathbb{P}}\right\rangle_K\\
&=|V|\mathbb{E}_{x\sim\mathbb{P}}\left\langle K(\cdot,x)-\mu_{\mathbb{P}},K(\cdot,x)-\mu_{\mathbb{P}}\right\rangle_K\\
&=|V|\mathbb{E}_{x\sim\mathbb{P}}\operatorname{Tr}\left(\left[K(\cdot,x)-\mu_{\mathbb{P}}\right]\otimes\left[K(\cdot,x)-\mu_{\mathbb{P}}\right]\right)\\
&=|V|\operatorname{Tr}(\Sigma_{\mathbb{P}})
\end{aligned}$$

exploiting the independence of e'_n-s and $[e'_n]^2=1$.

Until this point we showed that for all $\alpha\in(0,1)$, $\beta>0$, if $\epsilon\geq 2\sqrt{\frac{\|\Sigma_{\mathbb{P}}\|Q}{\alpha N}}$ then

$$\sup_{f\in B_K}\sum_{q=1}^{Q}\mathbb{I}_{r(f,q)\geq\epsilon}\leq|U|\alpha+Q\beta+\frac{8Q\sqrt{|V|\operatorname{Tr}(\Sigma_{\mathbb{P}})}}{\epsilon N}$$

with probability at least $1-e^{-\frac{Q\beta^2}{8}}$. Thus, to ensure that $\sup_{f\in B_K}\sum_{q=1}^{Q}\mathbb{I}_{r(f,q)\geq\epsilon}+N_c\leq Q/2$ it is sufficient to choose (α,β,ϵ) such that $|U|\alpha+Q\beta+\frac{8Q\sqrt{|V|\operatorname{Tr}(\Sigma_{\mathbb{P}})}}{\epsilon N}+N_c\leq\frac{Q}{2}$, and in this case $\left\|\hat{\mu}_{\mathbb{P},Q}-\mu_{\mathbb{P}}\right\|_K\leq(1+\sqrt{2})\epsilon$. Applying the $|U|\leq Q$ and $|V|\leq N$ bounds, we want to have

$$Q\alpha+Q\beta+\frac{8Q\sqrt{\operatorname{Tr}(\Sigma_{\mathbb{P}})}}{\epsilon\sqrt{N}}+N_c\leq\frac{Q}{2}. \tag{19}$$

Choosing $\alpha=\beta=\frac{\delta}{3}$ in Eq. (19), the sum of the first two terms is $Q\frac{2\delta}{3}$; $\epsilon\geq\max\left(2\sqrt{\frac{3\|\Sigma_{\mathbb{P}}\|Q}{\delta N}},\frac{24}{\delta}\sqrt{\frac{\operatorname{Tr}(\Sigma_{\mathbb{P}})}{N}}\right)$ gives $\leq Q\frac{\delta}{3}$ for the third term. Since $N_c\leq Q(\frac{1}{2}-\delta)$, we got

$$\left\|\hat{\mu}_{\mathbb{P},Q}-\mu_{\mathbb{P}}\right\|_K\leq c_1\max\left(\sqrt{\frac{3\|\Sigma_{\mathbb{P}}\|Q}{\delta N}},\frac{12}{\delta}\sqrt{\frac{\operatorname{Tr}(\Sigma_{\mathbb{P}})}{N}}\right)$$

with probability at least $1-\mathrm{e}^{-\frac{Q\delta^2}{72}}$. With an $\eta=\mathrm{e}^{-\frac{Q\delta^2}{72}}$, and hence $Q=\frac{72\ln\left(\frac{1}{\eta}\right)}{\delta^2}$ reparameterization Theorem 1 follows.

A.2. Proof of Theorem 2

The reasoning is similar to Theorem 1; we detail the differences below. The high-level structure of the proof is as follows:

- First we prove that $\left|\widehat{\mathrm{MMD}}_Q(\mathbb{P},\mathbb{Q})-\mathrm{MMD}(\mathbb{P},\mathbb{Q})\right|\leq r_{Q,N}$, where $r_{Q,N}=\sup_{f\in B_K}\left|\operatorname{med}_{q\in[Q]}\left\{\left\langle f,\left(\mu_{S_q,\mathbb{P}}-\mu_{S_q,\mathbb{Q}}\right)-\left(\mu_{\mathbb{P}}-\mu_{\mathbb{Q}}\right)\right\rangle_K\right\}\right|$.
- Then $r_{Q,N}$ is bounded.

Step-1:

- $\widehat{\mathrm{MMD}}_Q(\mathbb{P},\mathbb{Q})-\mathrm{MMD}(\mathbb{P},\mathbb{Q})\leq r_{Q,N}$: By the subadditivity of supremum $[\sup_f(a_f+b_f)\leq\sup_f a_f+\sup_f b_f]$ one gets

$$\begin{aligned}
&\widehat{\mathrm{MMD}}_Q(\mathbb{P},\mathbb{Q})\\
&=\sup_{f\in B_K}\operatorname{med}_{q\in[Q]}\Big\{\Big\langle f,\left(\mu_{S_q,\mathbb{P}}-\mu_{S_q,\mathbb{Q}}\right)-\left(\mu_{\mathbb{P}}-\mu_{\mathbb{Q}}\right)\\
&\qquad+\left(\mu_{\mathbb{P}}-\mu_{\mathbb{Q}}\right)\Big\rangle_K\Big\}\\
&\leq\sup_{f\in B_K}\operatorname{med}_{q\in[Q]}\left\{\left\langle f,\left(\mu_{S_q,\mathbb{P}}-\mu_{S_q,\mathbb{Q}}\right)-\left(\mu_{\mathbb{P}}-\mu_{\mathbb{Q}}\right)\right\rangle_K\right\}\\
&\qquad+\sup_{f\in B_K}\left\langle f,\mu_{\mathbb{P}}-\mu_{\mathbb{Q}}\right\rangle_K\\
&\leq\underbrace{\sup_{f\in B_K}\left|\operatorname{med}_{q\in[Q]}\left\{\left\langle f,\left(\mu_{S_q,\mathbb{P}}-\mu_{S_q,\mathbb{Q}}\right)-\left(\mu_{\mathbb{P}}-\mu_{\mathbb{Q}}\right)\right\rangle_K\right\}\right|}_{=r_{Q,N}}\\
&\qquad+\mathrm{MMD}(\mathbb{P},\mathbb{Q}).
\end{aligned}$$

- $\mathrm{MMD}_Q(\mathbb{P},\mathbb{Q})-\widehat{\mathrm{MMD}}_Q(\mathbb{P},\mathbb{Q})\leq r_{Q,N}$: Let $a_f:=\left\langle f,\mu_{\mathbb{P}}-\mu_{\mathbb{Q}}\right\rangle_K$ and $b_f:=\operatorname{med}_{q\in[Q]}\left\{\left\langle f,\left(\mu_{\mathbb{P}}-\mu_{\mathbb{Q}}\right)-\left(\mu_{S_q,\mathbb{P}}-\mu_{S_q,\mathbb{Q}}\right)\right\rangle_K\right\}$. Then

$$\begin{aligned}
&a_f-b_f\\
&=\left\langle f,\mu_{\mathbb{P}}-\mu_{\mathbb{Q}}\right\rangle_K\\
&\quad+\operatorname{med}_{q\in[Q]}\left\{\left\langle f,\left(\mu_{S_q,\mathbb{P}}-\mu_{S_q,\mathbb{Q}}\right)-\left(\mu_{\mathbb{P}}-\mu_{\mathbb{Q}}\right)\right\rangle_K\right\}\\
&=\operatorname{med}_{q\in[Q]}\left\{\left\langle f,\mu_{S_q,\mathbb{P}}-\mu_{S_q,\mathbb{Q}}\right\rangle_K\right\}
\end{aligned}$$

by $\operatorname{med}_{q\in[Q]}\{-z_q\}=-\operatorname{med}_{q\in[Q]}\{z_q\}$. Applying the $\sup_f(a_f-b_f)\geq\sup_f a_f-\sup_f b_f$ inequality (it follows from the subadditivity of sup):

$$\begin{aligned}
&\widehat{\mathrm{MMD}}_Q(\mathbb{P},\mathbb{Q})\\
&\geq\mathrm{MMD}(\mathbb{P},\mathbb{Q})\\
&\quad-\sup_{f\in B_K}\underbrace{\operatorname{med}_{q\in[Q]}\left\{\left\langle f,\left(\mu_{\mathbb{P}}-\mu_{\mathbb{Q}}\right)-\left(\mu_{S_q,\mathbb{P}}-\mu_{S_q,\mathbb{Q}}\right)\right\rangle_K\right\}}_{-\operatorname{med}_{q\in[Q]}\left\{\left\langle f,\left(\mu_{S_q,\mathbb{P}}-\mu_{S_q,\mathbb{Q}}\right)-\left(\mu_{\mathbb{P}}-\mu_{\mathbb{Q}}\right)\right\rangle_K\right\}}\\
&\geq\mathrm{MMD}(\mathbb{P},\mathbb{Q})\\
&\quad-\underbrace{\sup_{f\in B_K}\left|\operatorname{med}_{q\in[Q]}\left\{\left\langle f,\left(\mu_{S_q,\mathbb{P}}-\mu_{S_q,\mathbb{Q}}\right)-\left(\mu_{\mathbb{P}}-\mu_{\mathbb{Q}}\right)\right\rangle_K\right\}\right|}_{r_{Q,N}}.
\end{aligned}$$

Step-2: Our goal is to control

$$r_{Q,N}=\sup_{f\in B_K}\left|\operatorname{med}_{q\in[Q]}\left\{r(f,q)\right\}\right|,\text{ where}$$
$$r(f,q):=\left\langle f,\left(\mu_{S_q,\mathbb{P}}-\mu_{S_q,\mathbb{Q}}\right)-\left(\mu_{\mathbb{P}}-\mu_{\mathbb{Q}}\right)\right\rangle_K.$$

The relevant quantities which change compared to the proof of Theorem 1 are as follows.

- **Median rephrasing**:

$$\sup_{f\in B_K}\left|\operatorname{med}_{q\in[Q]}\{r(f,q)\}\right|\le\epsilon$$
$$\Leftrightarrow \forall f\in B_K: -\epsilon\le \operatorname{med}_{q\in[Q]}\{r(f,q)\}\le\epsilon$$
$$\Leftarrow \forall f\in B_K: |\{q: r(f,q)\le -\epsilon\}|\le Q/2 \text{ and } |\{q: r(f,q)\ge\epsilon\}|\le Q/2$$
$$\Leftarrow \forall f\in B_K: |\{q: |r(f,q)|\ge\epsilon\}|\le Q/2.$$

Thus, $\forall f\in B_K: |\{q\in U: |r(f,q)|\ge\epsilon\}|+N_c\le\frac{Q}{2}$, implies $\sup_{f\in B_K}\left|\operatorname{med}_{q\in[Q]}\{r(f,q)\}\right|\le\epsilon$.

- **Controlling** $|r(f,q)|$: For any $f\in B_K$ the random variables $[f(x_i)-f(y_i)]-[\mathbb{P}f-\mathbb{Q}f]$ are independent, zero-mean and

$$\mathbb{E}_{(x,y)\sim\mathbb{P}\otimes\mathbb{Q}}([f(x)-\mathbb{P}f]-[f(y)-\mathbb{Q}f])^2$$
$$=\mathbb{E}_{x\sim\mathbb{P}}[f(x)-\mathbb{P}f]^2+\mathbb{E}_{y\sim\mathbb{Q}}[f(y)-\mathbb{Q}f]^2$$
$$\le\|\Sigma_{\mathbb{P}}\|+\|\Sigma_{\mathbb{Q}}\|,$$

where $\mathbb{P}\otimes\mathbb{Q}$ is the product measure. The Chebyshev argument with $z=|r(f,q)|$ implies that $\forall\alpha\in(0,1)$

$$(\mathbb{P}\otimes\mathbb{Q})\left(|r(f,q)|>\sqrt{\frac{(\|\Sigma_{\mathbb{P}}\|+\|\Sigma_{\mathbb{Q}}\|)Q}{\alpha N}}\right)\le\alpha.$$

This means $(\mathbb{P}\otimes\mathbb{Q})(|r(f,q)|>\epsilon/2)\le\alpha$ with $\epsilon\ge 2\sqrt{\frac{(\|\Sigma_{\mathbb{P}}\|+\|\Sigma_{\mathbb{Q}}\|)Q}{\alpha N}}$.

- **Switching from** $|U|$ **to** N **terms**: With $(xy)_V=\{(x_i,y_i): i\in V\}$, in '(b)' with $\tilde{x}_n:=K(\cdot,x_n)-\mu_{\mathbb{P}}$, $\tilde{y}_n:=K(\cdot,y_n)-\mu_{\mathbb{Q}}$ we arrive at

$$\mathbb{E}_{(xy)_V}\mathbb{E}_{\mathbf{e}'}\left\|\sum_{n\in V}e'_n(\tilde{x}_n-\tilde{y}_n)\right\|_K^2$$
$$=\mathbb{E}_{(xy)_V}\mathbb{E}_{\mathbf{e}'}\sum_{n\in V}[e'_n]^2\langle\tilde{x}_n-\tilde{y}_n,\tilde{x}_n-\tilde{y}_n\rangle_K$$
$$=|V|\mathbb{E}_{(xy)\sim\mathbb{P}}\|[K(\cdot,x)-\mu_{\mathbb{P}}]-[K(\cdot,y)-\mu_{\mathbb{Q}}]\|_K$$
$$=|V|\left[\operatorname{Tr}(\Sigma_{\mathbb{P}})+\operatorname{Tr}(\Sigma_{\mathbb{Q}})\right].$$

- These results imply

$$Q\alpha+Q\beta+\frac{8Q\sqrt{\operatorname{Tr}(\Sigma_{\mathbb{P}})+\operatorname{Tr}(\Sigma_{\mathbb{Q}})}}{\epsilon\sqrt{N}}+N_c\le Q/2.$$

$\epsilon\ge\max\left(2\sqrt{\frac{3(\|\Sigma_{\mathbb{P}}\|+\|\Sigma_{\mathbb{Q}}\|)Q}{\delta N}},\frac{24}{\delta}\sqrt{\frac{\operatorname{Tr}(\Sigma_{\mathbb{P}})+\operatorname{Tr}(\Sigma_{\mathbb{Q}})}{N}}\right)$, $\alpha=\beta=\frac{\delta}{3}$ choice gives that

$$\left|\widehat{\mathrm{MMD}}_Q(\mathbb{P},\mathbb{Q})-\mathrm{MMD}(\mathbb{P},\mathbb{Q})\right|$$
$$\le 2\max\left(\sqrt{\frac{3(\|\Sigma_{\mathbb{P}}\|+\|\Sigma_{\mathbb{Q}}\|)Q}{\delta N}},\frac{12}{\delta}\sqrt{\frac{\operatorname{Tr}(\Sigma_{\mathbb{P}})+\operatorname{Tr}(\Sigma_{\mathbb{Q}})}{N}}\right)$$

with probability at least $1-e^{-\frac{Q\delta^2}{72}}$. $\eta=e^{-\frac{Q\delta^2}{72}}$, i.e. $Q=\frac{72\ln\left(\frac{1}{\eta}\right)}{\delta^2}$ reparameterization finishes the proof of Theorem 2.

B. Technical Lemmas

Lemma 3 (Supremum)**.**

$$\left|\sup_f a_f-\sup_f b_f\right|\le\sup_f|a_f-b_f|.$$

Lemma 4 (Bounded difference property of Z)**.** *Let $N\in\mathbb{Z}^+$, $(S_q)_{q\in[Q]}$ be a partition of $[N]$, $K:\mathcal{X}\times\mathcal{X}\to\mathbb{R}$ be a kernel, μ be the mean embedding associated to K, $x_{1:N}$ be i.i.d. random variables on $\mathcal{X}$, $Z(x_V)=\sup_{f\in B_K}\sum_{q\in U}\left[\phi\left(\frac{2\langle f,\mu_{S_q}-\mu_{\mathbb{P}}\rangle_K}{\epsilon}\right)-\mathbb{E}\phi\left(\frac{2\langle f,\mu_{S_q}-\mu_{\mathbb{P}}\rangle_K}{\epsilon}\right)\right]$, where $U\subseteq[Q]$, $V=\cup_{q\in U}S_q$. Let x'_{V_i} be x_V except for the $i\in V$-th coordinate; x_i is changed to x'_i. Then*

$$\sup_{x_V\in\mathcal{X}^{|V|},x'_i\in\mathcal{X}}\left|Z(x_V)-Z\left(x'_{V_i}\right)\right|\le 4,\ \forall i\in V.$$

Proof. Since $(S_q)_{q\in[Q]}$ is a partition of $[Q]$, $(S_q)_{q\in U}$ forms a partition of V and there exists a unique $r\in U$ such that $i\in S_r$. Let

$$Y_q:=Y_q(f,x_V),$$
$$q\in U=\phi\left(\frac{2\langle f,\mu_{S_q}-\mu_{\mathbb{P}}\rangle_K}{\epsilon}\right)-\mathbb{E}\phi\left(\frac{2\langle f,\mu_{S_q}-\mu_{\mathbb{P}}\rangle_K}{\epsilon}\right),$$
$$Y'_r:=Y_r(f,x'_{V_i}).$$

In this case

$$\left|Z(x_V)-Z\left(x'_{V_i}\right)\right|$$
$$=\left|\sup_{f\in B_K}\sum_{q\in U}Y_q-\sup_{f\in B_K}\left(\sum_{q\in U\setminus\{r\}}Y_q+Y'_r\right)\right|$$
$$\overset{(a)}{\le}\sup_{f\in B_K}|Y_r-Y'_r|\overset{(b)}{\le}\sup_{f\in B_K}\left(\underbrace{|Y_r|}_{\le 2}+\underbrace{|Y'_r|}_{\le 2}\right)\le 4,$$

where in (a) we used Lemma 3, (b) the triangle inequality and the boundedness of ϕ [$|\phi(t)|\le 1$ for all t].

□

Lemma 5 (Uniformly bounded separable Carathéodory family)**.** *Let $\epsilon>0$, $N\in\mathbb{Z}^+$, $Q\in\mathbb{Z}^+$, $M=N/Q\in\mathbb{Z}^+$, $\phi(t)=(t-1)\mathbb{I}_{1\le t\le 2}+\mathbb{I}_{t\ge 2}$, $K:\mathcal{X}\times\mathcal{X}\to\mathbb{R}$ is a continuous kernel on the separable topological domain $\mathcal{X}$, μ is the mean embedding associated to K, $\mathbb{P}_M:=\frac{1}{M}\sum_{m=1}^M\delta_{u_m}$, $\mathcal{G}=\{g_f: f\in B_K\}$, where $g_f:\mathcal{X}^M\to\mathbb{R}$ is defined as*

$$g_f(u_{1:M})=\phi\left(\frac{2\langle f,\mu_{\mathbb{P}_M}-\mu_{\mathbb{P}}\rangle_K}{\epsilon}\right).$$

Then $\mathcal{G}$ is a uniformly bounded separable Carathéodory family: (i) $\sup_{f\in B_K}\|g_f\|_\infty < \infty$ where $\|g\|_\infty = \sup_{u_{1:M}\in\mathcal{X}^M}|g(u_{1:M})|$, (ii) $u_{1:M}\mapsto g_f(u_{1:M})$ is measurable for all $f\in B_K$, (iii) $f\mapsto g_f(u_{1:M})$ is continuous for all $u_{1:M}\in\mathcal{X}^M$, (iv) B_K is separable.

Proof.

(i) $|\phi(t)|\le 1$ for any t, hence $\|g_f\|_\infty \le 1$ for all $f\in B_K$.
(ii) Any $f\in B_K$ is continuous since $\mathcal{H}_K\subset C(\mathcal{X})=\{h:\mathcal{X}\to\mathbb{R}\text{ continuous}\}$, so $u_{1:M}\mapsto (f(u_1),\dots,f(u_M))$ is continuous. ϕ is Lipschitz, specifically continuous. The continuity of these two maps imply that of $u_{1:M}\mapsto g_f(u_{1:M})$, specifically it is Borel-measurable.
(ii) The statement follows by the continuity of $f\mapsto \langle f,h\rangle_K$ ($h=\mu_{\mathbb{P}_M}-\mu_{\mathbb{P}}$) and that of ϕ.
(iv) B_K is separable since $\mathcal{H}_K$ is so by assumption.

□

C. External Lemma

Below we state the McDiarmid inequality for self-containedness.

Lemma 6 (McDiarmid inequality)**.** *Let $x_{1:N}$ be $\mathcal{X}$-valued independent random variables. Assume that $f:\mathcal{X}^N\to\mathbb{R}$ satisfies the bounded difference property*

$$\sup_{u_1,\dots,u_N,u_r'\in\mathcal{X}}|f(u_{1:N})-f(u_{1:N}')|\le c,\quad \forall n\in[N],$$

where $u_{1:N}'=(u_1,\dots,u_{n-1},u_n',u_{n+1},\dots,u_N)$. Then for any $\tau>0$

$$\mathbb{P}\left(f(x_{1:N})<\mathbb{E}_{x_{1:N}}\left[f(x_{1:N})\right]+c\sqrt{\frac{\tau N}{2}}\right)\ge 1-e^{-\tau}.$$

D. Pseudocode of Experiment-2

The pseudocode of the two-sample test conducted in Experiment-2 is summarized in Algorithm 3.

Algorithm 3 Two-sample test (Experiment-2)

Input: Two samples: $(X_n)_{n\in[N]}$, $(Y_n)_{n\in[N]}$. Number of bootstrap permutations: $B\in\mathbb{Z}^+$. Level of the test: $\alpha\in(0,1)$. Kernel function with hyperparameter $\theta\in\Theta$: K_θ.

Split the dataset randomly into 3 equal parts:

$$[N]=\dot{\bigcup_{i=1}^{3}} I_i,\quad |I_1|=|I_2|=|I_3|.$$

Tune the hyperparameters using the 1st part of the dataset:

$$\hat{\theta}=\text{argmax}_{\theta\in\Theta} J_\theta := \widehat{\text{MMD}}_\theta\left((X_n)_{n\in I_1},(Y_n)_{n\in I_1}\right).$$

Estimate the $(1-\alpha)$-quantile of $\widehat{\text{MMD}}_{\hat{\theta}}$ under the null, using B bootstrap permutations from $(X_n)_{n\in I_2}\cup(Y_n)_{n\in I_2}$: $\hat{q}_{1-\alpha}$.

Compute the test statistic on the third part of the dataset:

$$T_{\hat{\theta}}=\widehat{\text{MMD}}_{\hat{\theta}}\left((X_n)_{n\in I_3},(Y_n)_{n\in I_3}\right).$$

Output: $T_{\hat{\theta}}-\hat{q}_{1-\alpha}$.