Supplementary material for Reducing statistical time-series problems to binary classification

Theorem 1. *Let $\mathbf{H} = (\mathcal{H}_1, \mathcal{H}_2, \dots)$, $\mathcal{H}_k \subset \mathcal{X}^k$, $k \in \mathbb{N}$ be a sequence of separable sets of indicator functions of finite VC dimension such that $\mathcal{H}_k$ generates $\mathcal{F}_{\mathcal{X}^k}$. Then, for every stationary ergodic time series distributions ρ_X and ρ_Y generating samples $X_{1..n}$ and $Y_{1..m}$ we have*

$$\lim_{n,m\to\infty} \hat{D}_{\mathbf{H}}(X_{1..n}, Y_{1..m}) = D_{\mathbf{H}}(\rho_X, \rho_Y) \tag{1}$$

Proof. As is established in [2],under the conditions of the theorem we have

$$\lim_{n\to\infty} \sup_{h\in\mathcal{H}_k} \frac{1}{n-k+1} \sum_{i=1}^{n-k+1} h(X_{i..i+k-1}) = \sup_{h\in\mathcal{H}_k} \mathbf{E}_{\rho_X} h(X_1, \dots, X_k) \ \rho_X\text{-a.s.} \tag{2}$$

for all $k \in \mathbb{N}$, and likewise for ρ_Y. Fix an $\varepsilon > 0$. We can find a $T \in \mathbb{N}$ such that

$$\sum_{k>T} w_k \le \varepsilon. \tag{3}$$

Note that T depends only on ε. Moreover, as follows from (2), for each $k = 1..T$ we can find an N_k such that

$$\Big| \sup_{h\in\mathcal{H}_k} \frac{1}{n-k+1} \sum_{i=1}^{n-k+1} h(X_{i..i+k-1}) - \sup_{h\in\mathcal{H}_k} \mathbf{E}_{\rho_X} h(X_{1..k}) \Big| \le \varepsilon/T \tag{4}$$

Let $N_k := \max_{i=1..T} N_i$ and define analogously M for ρ_Y. Thus, for $n \ge N$, $m \ge M$ we have

$$\begin{aligned}
&\hat{D}_{\mathbf{H}}(X_{1..n}, Y_{1..m}) \\
&\le \sum_{k=1}^{T} w_k \sup_{h\in\mathcal{H}_k} \left| \frac{1}{n-k+1} \sum_{i=1}^{n-k+1} h(X_{i..i+k-1}) - \frac{1}{m-k+1} \sum_{i=1}^{m-k+1} h(Y_{i..i+k-1}) \right| + \varepsilon \\
&\qquad \le \sum_{k=1}^{T} w_k \sup_{h\in\mathcal{H}_k} \left\{ \left| \frac{1}{n-k+1} \sum_{i=1}^{n-k+1} h(X_{i..i+k-1}) - \mathbf{E}_{\rho_1} h(X_{1..k}) \right| \right. \\
&\qquad\qquad + |\mathbf{E}_{\rho_1} h(X_{1..k}) - \mathbf{E}_{\rho_2} h(Y_{1..k})| \\
&\qquad\qquad \left. + \left| \mathbf{E}_{\rho_2} h(Y_{1..k}) - \frac{1}{m-k+1} \sum_{i=1}^{m-k+1} h(Y_{i..i+k-1}) \right| \right\} + \varepsilon \\
&\qquad\qquad\qquad \le 3\varepsilon + D_{\mathbf{H}}(\rho_X, \rho_Y),
\end{aligned}$$

where the first inequality follows from the definition of $\hat{D}_{\mathbf{H}}$ (2) and from (3) and the last inequality follows from (4). Since ε was chosen arbitrary the statement follows. $\square$

Lemma 1.1. *Let two samples $X_{1..n}$ and $Y_{1..m}$ be generated by stationary distributions ρ_X and ρ_Y whose β-mixing coefficients satisfy $\beta(\rho_{\cdot}, t) \le \gamma^t$ for some $\gamma < 1$. Let H_k, $k \in \mathbb{N}$ be some sets of indicator functions on $\mathcal{X}^k$ whose VC dimension d_k is finite and non-decreasing with k. Then*

$$P(|\hat{D}_{\mathbf{H}}(X_{1..n}, Y_{1..m}) - D_{\mathbf{H}}(\rho_X, \rho_Y)| > \varepsilon) \le 2\Delta(\varepsilon/4, n') \tag{5}$$

where $n' := \min\{n_1, n_2\}$*, the probability is with respect to* $\rho_X \times \rho_Y$ *and*

$$\Delta(\varepsilon, n) := -\log\varepsilon(n\gamma^{\sqrt{n}+\log(\varepsilon)} + 8n^{(d_{-\log\varepsilon}+1)/2}e^{-\sqrt{n}\varepsilon^2/8}). \tag{6}$$

Proof. Note that $\sum_{k=-\log\varepsilon/2}^{\infty} w_k < \varepsilon/2$. Using this and the definitions (1) and (2) of $D_{\mathbf{H}}$ and $\hat{D}_{\mathbf{H}}$ we obtain

$$P(|\hat{D}_{\mathbf{H}}(X_{1..n_1}, Y_{1..n_2}) - D_{\mathbf{H}}(\rho_X, \rho_Y)| > \varepsilon) \leq \sum_{k=1}^{-\log(\varepsilon/2)} (q_n(\rho_X, \mathcal{H}_k, \varepsilon/4) + q_n(\rho_Y, \mathcal{H}_k, \varepsilon/4)),$$

which, together with (6)implies the statement. $\square$