

A FAMILY OF CORRELATION COEFFICIENTS AND THEIR PROPERTIES

Muhong Gao and Qizhai Li*

*University of International Business and Economics,
and Academy of Mathematics and System Science, Chinese Academy of Sciences*

Supplementary Material

This supplementary material provides detailed proofs of the theorems. Specifically, the proofs of the theorems presented in Sections 2, 3, and 4 are collected in Sections [S1](#), [S2](#), and [S3](#), respectively.

Notations in proof For a Borel set $A \subseteq \mathbb{R}$ and $x \in \mathbb{R}$, let $I_A(x) := I(x \in A)$ be the indicator function of x belonging to A . Let $\pi(i)$ be the rank of X_i . If there are ties among X_i 's, then break ties at random so that π is a permutation of $\{1, \dots, n\}$. Let

$$N(i) := \begin{cases} \pi^{-1}(\pi(i) + 1), & \text{if } \pi(i) < n, \\ i, & \text{if } \pi(i) = n. \end{cases}$$

Then $X_{N(i)}$ is the right nearest neighbor of X_i . For convenience, we also write $X_{n,i} := X_{N(i)}$ and $Y_{n,i} := Y_{N(i)}$, for $i = 1, \dots, n$. Using these notations, the variation statistic in

*Corresponding author

Definition 1 can be equivalently written as $\zeta_n^{(h,F)}(X, Y) = \frac{1}{n} \sum_{i=1}^n h(F(Y_i), F(Y_{N(i)})) = \frac{1}{n} \sum_{i=1}^n h(F(Y_i), F(Y_{n,i}))$. Let $\tilde{\mu}_x := \mu_{Y|X=x}$ (or $\mu_{Y|x}$) denote the conditional law of Y given $X = x$. Refer to Lemma 1 below for the existence of $\tilde{\mu}_x$.

Lemma 1 (Theorem 2.1.22 and Exercise 4.1.18 in [Durrett \(2019\)](#)). *For each Borel set $A \subseteq \mathbb{R}$, there is a measurable map*

$$x \longmapsto \tilde{\mu}_x(A), \quad \text{from } \mathbb{R} \text{ into } [0, 1],$$

such that

1. *for each A , $\tilde{\mu}_X(A)$ is a version of $P(Y \in A | X)$;*
2. *with probability one, $\tilde{\mu}_X$ is a probability measure on $\mathbb{R}$.*

S1 Proofs of Theorems in Section 2

S1.1 Proofs of Theorem 1 and Corollary 1

Before proving Theorem 1, we first present Lemmas [A.1](#) – [A.5](#).

Lemma A.1. Let $F(\cdot)$ be a c.d.f. function. Let $h(\cdot, \cdot) : \mathbb{R}^2 \rightarrow [0, \infty)$ be a non-negative continuous function. For any $\epsilon > 0$, there exists a simple function

$$q(y, z) = \sum_{j=1}^m c_j \mathbf{I}_{B_j}(y) \mathbf{I}_{C_j}(z),$$

with $c_j > 0$ and B_j, C_j being Borel subsets of $\mathbb{R}$, such that

$$\sup_{(y,z) \in \mathbb{R}^2} |q(y, z) - h(F(y), F(z))| < \epsilon. \quad (\text{A.1})$$

Proof. Since h is continuous and $A := [0, 1] \times [0, 1]$ is a compact subset of $\mathbb{R}^2$, we have that h is uniformly continuous in A . Then there exists $\delta > 0$, such that

$$|h(y_1, z_1) - h(y_2, z_2)| < \epsilon \quad (\text{A.2})$$

holds for any $(y_1, z_1), (y_2, z_2) \in A$ with $|y_1 - y_2| \leq \delta$ and $|z_1 - z_2| \leq \delta$.

Let $\ell = \lceil 1/\delta \rceil - 1$ (i.e., the maximum integer not exceeding $1/\delta - 1$). Let

$$w_i = \inf\{y \in \mathbb{R} : F(y) \geq i \cdot \delta\}, \quad i = 1, \dots, \ell.$$

Since the c.d.f. function $F(\cdot)$ is non-decreasing, $w_1 \leq w_2 \leq \dots \leq w_\ell$ is also a non-decreasing sequence in $\mathbb{R}$. Let $w_1^* < \dots < w_{\ell^*}^*$ be the strictly-increasing sequence obtained from removing the ties in $\{w_i\}_{i=1}^\ell$. Partition $\mathbb{R}$ into non-overlapping intervals $\{W_i\}_{i=1}^{\ell^*+1}$, where $W_1 = (-\infty, w_1^*)$, $W_{\ell^*+1} = [w_{\ell^*}^*, \infty)$, and $W_i = [w_{i-1}^*, w_i^*)$, for $i = 2, \dots, \ell^*$. Then for any $i = 1, \dots, \ell^* + 1$, we have

$$\sup_{y_1, y_2 \in W_i} |F(y_1) - F(y_2)| \leq \delta. \quad (\text{A.3})$$

Let

$$q(y, z) = \sum_{i=1}^{\ell^*+1} \sum_{j=1}^{\ell^*+1} c_{i,j} \mathbf{I}_{W_i}(y) \mathbf{I}_{W_j}(z), \quad (\text{A.4})$$

where $c_{i,j} = h(F(u_i), F(v_j))$, with (u_i, v_j) being an arbitrary interior point in $W_i \times W_j$.

Combining (A.2) and (A.3), we have that for any $(y, z) \in W_i \times W_j$,

$$|q(y, z) - h(F(y), F(z))| = |h(F(u_i), F(v_j)) - h(F(y), F(z))| < \epsilon.$$

This combined with the fact that $\cup_{i=1}^{\ell^*+1} \cup_{j=1}^{\ell^*+1} W_i \times W_j = \mathbb{R}^2$ yields that $\sup_{(y,z) \in \mathbb{R}^2} |q(y, z) - h(F(y), F(z))| < \epsilon$. We prove that the simple function $q(y, z)$ in (A.4) satisfies (A.1).

□

Lemma A.2. Let $F(\cdot)$ be a c.d.f. function. Let $h(\cdot, \cdot) : \mathbb{R}^2 \rightarrow [0, \infty)$ be a non-negative continuous function. Define the bivariate function $g : \mathbb{R}^2 \mapsto [0, \infty)$,

$$g(x_1, x_2) := \int \int h(F(y), F(z)) \, d\tilde{\mu}_{x_1}(y) \, d\tilde{\mu}_{x_2}(z), \quad (x_1, x_2) \in \mathbb{R}^2. \quad (\text{A.5})$$

Then $g(X_1, X_{n,1}) - g(X_1, X_1)$ tends to 0 in probability as $n \rightarrow \infty$.

Proof. Fix some $\epsilon > 0$ and $\delta > 0$. From Lemma A.1, there exists a simple function $q(y, z) = \sum_{j=1}^m c_j \mathbf{I}_{B_j}(y) \mathbf{I}_{C_j}(z)$ such that $\sup_{(y,z) \in \mathbb{R}^2} |q(y, z) - h(F(y), F(z))| < \epsilon$. Define

the bivariate function

$$\begin{aligned}
r(x_1, x_2) : &= \int \int q(y, z) \, d\tilde{\mu}_{x_1}(y) \, d\tilde{\mu}_{x_2}(z) \\
&= \sum_{j=1}^m c_j \int \int \mathbf{I}_{B_j}(y) \mathbf{I}_{C_j}(z) \, d\tilde{\mu}_{x_1}(y) \, d\tilde{\mu}_{x_2}(z) \\
&= \sum_{j=1}^m c_j \cdot \tilde{\mu}_{x_1}(B_j) \tilde{\mu}_{x_2}(C_j).
\end{aligned} \tag{A.6}$$

Then we have

$$\begin{aligned}
&\sup_{(x_1, x_2) \in \mathbb{R}^2} |g(x_1, x_2) - r(x_1, x_2)| \\
&\leq \sup_{(x_1, x_2) \in \mathbb{R}^2} \int \int |h(F(y), F(z)) - q(y, z)| \, d\tilde{\mu}_{x_1}(y) \, d\tilde{\mu}_{x_2}(z) \\
&< \epsilon.
\end{aligned} \tag{A.7}$$

Note that

$$r(X_1, X_{n,1}) - r(X_1, X_1) = \sum_{j=1}^m c_j \cdot \tilde{\mu}_{X_1}(B_j) \cdot \{\tilde{\mu}_{X_{n,1}}(C_j) - \tilde{\mu}_{X_1}(C_j)\}. \tag{A.8}$$

By Corollary A.9 in [Chatterjee \(2021\)](#) and the fact that $x \mapsto \tilde{\mu}_x(A)$ is a measurable map for any Borel set A , we have that $\tilde{\mu}_{X_{n,1}}(C_j) - \tilde{\mu}_{X_1}(C_j) \xrightarrow{P} 0$ as $n \rightarrow \infty$. This combined with (A.8) gives that $r(X_1, X_{n,1}) - r(X_1, X_1) \xrightarrow{P} 0$. Therefore, there exists some integer $N > 0$ such that for any integer $n > N$,

$$\mathbb{P}(|r(X_1, X_{n,1}) - r(X_1, X_1)| > \epsilon) < \delta.$$

Finally, note that

$$\begin{aligned}
& |g(X_1, X_{n,1}) - g(X_1, X_1)| \\
& \leq |r(X_1, X_{n,1}) - r(X_1, X_1)| + |g(X_1, X_{n,1}) - r(X_1, X_{n,1})| + |g(X_1, X_1) - r(X_1, X_1)| \\
& \leq |r(X_1, X_{n,1}) - r(X_1, X_1)| + 2\epsilon.
\end{aligned}$$

Combining the facts above, for any $n > N$,

$$\mathbb{P}(|g(X_1, X_{n,1}) - g(X_1, X_1)| > 3\epsilon) < \delta.$$

The proof is completed given that ϵ and δ are arbitrary. $\square$

Lemma A.3. The function $g(x_1, x_2)$ defined in (A.5) is measurable in $\mathbb{R}^2$.

Proof. By Lemma A.1, there exists a sequence of simple functions

$$q_i(y, z) := \sum_{j=1}^{m_i} c_{i,j} \mathbf{I}_{B_{i,j}}(y) \mathbf{I}_{C_{i,j}}(z), \quad i = 1, 2, \dots,$$

such that $q_i(y, z)$ converges uniformly to $h(F(y), F(z))$. Define the bivariate functions

$$r_i(x_1, x_2) := \int \int q_i(y, z) \, d\tilde{\mu}_{x_1}(y) \, d\tilde{\mu}_{x_2}(z), \quad i = 1, 2, \dots$$

From (A.7), we know that

$$\sup_{(x_1, x_2) \in \mathbb{R}^2} |g(x_1, x_2) - r_i(x_1, x_2)| \leq \sup_{(y, z) \in \mathbb{R}^2} |h(F(y), F(z)) - q_i(y, z)| \rightarrow 0,$$

as $i \rightarrow \infty$. This proves that $r_i(x_1, x_2)$ converges uniformly to $g(x_1, x_2)$. By (A.6), we know $\{r_i\}_{i=1}^\infty$ are measurable functions in $\mathbb{R}^2$. Combining the facts above, we have that $g(x_1, x_2)$ is measurable in $\mathbb{R}^2$. $\square$

Lemma A.4. Let $F(\cdot)$ be a c.d.f. function. Let $h(\cdot, \cdot) : \mathbb{R}^2 \rightarrow [0, \infty)$ be a non-negative continuous function. Then we have $\lim_{n \rightarrow \infty} \mathbb{E}(\zeta_n^{(h, F)}(X, Y)) = \zeta^{(h, F)}$.

Proof. Let $\mathcal{F}$ be the σ -algebra generated by $\{X_i\}_{i=1}^n$ and the randomness used for breaking ties in the selection of π . Then for each $i = 1, \dots, n$, we have

$$\begin{aligned} \mathbb{E}\left(h(F(Y_i), F(Y_{n,i})) \mid \mathcal{F}\right) &= \mathbb{E}\left(h(F(Y_i), F(Y_{n,i})) \mid X_i, X_{n,i}\right) \\ &= \int \int h(F(y), F(z)) \, d\tilde{\mu}_{X_i}(y) \, d\tilde{\mu}_{X_{n,i}}(z) = g(X_i, X_{n,i}), \end{aligned}$$

where the function $g(\cdot, \cdot)$ is defined in (A.5). Since $h(\cdot, \cdot)$ is continuous in $[0, 1] \times [0, 1]$, we have that $g(\cdot, \cdot)$ is bounded in $\mathbb{R}^2$. This combined with Lemma A.2 gives that

$$\lim_{n \rightarrow \infty} \mathbb{E}\left(h(F(Y_1), F(Y_{n,1}))\right) = \lim_{n \rightarrow \infty} \mathbb{E}(g(X_1, X_{n,1})) = \mathbb{E}(g(X_1, X_1)) = \zeta^{(h, F)}.$$

Note that $h(F(Y_i), F(Y_{n,i}))$ are identically distributed for $i = 1, 2, \dots, n$. It follows that

$E(\zeta_n^{(h,F)}(X, Y)) = 1/n \cdot \sum_{i=1}^n E(h(F(Y_i), F(Y_{n,i}))) = E(h(F(Y_1), F(Y_{n,1}))) \rightarrow \zeta^{(h,F)}$, as $n \rightarrow \infty$. This completes the proof. $\square$

Lemma A.5. Let $h(\cdot, \cdot) : \mathbb{R}^2 \rightarrow [0, \infty)$ be a non-negative continuous function. There exists a constant $C \in (0, \infty)$, such that for any integer $n \geq 1$ and any real number $t \geq 0$,

$$P(|\zeta_n^{(h,F)}(X, Y) - E(\zeta_n^{(h,F)}(X, Y))| \geq t) \leq 2e^{-C \cdot n \cdot t^2}.$$

Proof. Assume continuity of X , and suppose that for some $i_0 \leq n$, (X_{i_0}, Y_{i_0}) is replaced by a different value (X'_{i_0}, Y'_{i_0}) . There can be at most one index i_1 such that $N(i_1) = i_0$ (i.e., X_{i_0} is the right-nearest neighbor of X_{i_1} before the replacement). Also, there can be at most one index i_2 such that X'_{i_0} is the right-nearest neighbor of X_{i_2} after the replacement. Combining these facts, there are at most 3 pairs of $(Y_i, Y_{N(i)})$ that changes after the replacement; they are $(Y_{i_0}, Y_{N(i_0)})$, $(Y_{i_1}, Y_{N(i_1)})$, and $(Y_{i_2}, Y_{N(i_2)})$. Since h is continuous in $[0, 1] \times [0, 1]$, there exists a constant $c \in (0, \infty)$ such that $\sup_{(y,z) \in \mathbb{R}^2} h(F(y), F(z)) < c$. Therefore, after the replacement, $\zeta_n^{(h,F)}(X, Y)$ changes by at most $\Delta_n := 3c/n$. By the bounded difference concentration inequality ([McDiarmid et al., 1989](#)), we have

$$P(|\zeta_n^{(h,F)}(X, Y) - E(\zeta_n^{(h,F)}(X, Y))| \geq t) \leq 2e^{-2 \cdot t^2 / (n \Delta_n^2)} = 2e^{-C \cdot n \cdot t^2},$$

where $C = 2 \cdot (3c)^{-2}$ is a positive constant. For the general case of X , the result follows from the similar proof as that of Lemma A.11 in [Chatterjee \(2021\)](#). $\square$

Proof of Theorem 1. Note that for any $t > 0$, we have $\sum_{n=1}^{\infty} 2e^{-C \cdot n \cdot t^2} < \infty$. By Lemma [A.5](#) and the Borel-Cantelli lemma, $|\zeta_n^{(h,F)}(X, Y) - E(\zeta_n^{(h,F)}(X, Y))| \rightarrow 0$ a.s., as $n \rightarrow \infty$. This combined with Lemma [A.4](#) yields that $\zeta_n^{(h,F)}(X, Y) \rightarrow \zeta^{(h,F)}$ a.s., which proves result (i). The proof of result (ii) directly follows from the strong law of large numbers for U-statistics ([Hoeffding, 1961](#)). It remains to prove result (iii).

If $\chi^{(h,F)} > 0$, then result (iii) directly follows from combining results (i) and (ii). If $\chi^{(h,F)} = 0$, then we have $P(h(F(Y), F(Z)) = 0) = 1$ for random variables $Y, Z \sim_{\text{i.i.d.}} \mu_Y$. Thus, $P(\chi_n^{(h,F)}(Y) = 0) = 1$, which gives that $\lim_{n \rightarrow \infty} \xi_n^{(h,F)}(X, Y) = 0 = \xi^{(h,F)}$ a.s.. This completes the proof. $\square$

Proof of Corollary 1. Since h is Lipschitz continuous, there exists a constant $K_h \in (0, \infty)$, such that $|h(y_1, z_1) - h(y_2, z_2)| \leq K_h (|y_1 - y_2| + |z_1 - z_2|)$, for any $(y_1, z_1), (y_2, z_2) \in [0, 1]^2$. Then

$$|\zeta_n^{(h, \widehat{F}_n)}(X, Y) - \zeta_n^{(h,F)}(X, Y)| \leq 2K_h \cdot \sup_{x \in \mathbb{R}} |\widehat{F}_n(x) - F(x)| \rightarrow 0, \text{ a.s..}$$

Similarly, we also have $|\chi_n^{(h, \widehat{F}_n)}(Y) - \chi_n^{(h,F)}(Y)| \rightarrow 0$ a.s.. Combining these two results

gives that

$$|\xi_n^{(h, \widehat{F}_n)}(X, Y) - \xi_n^{(h, F)}(X, Y)| \rightarrow 0, \text{ a.s..}$$

The remaining of proof follows from the convergence result in Theorem 1. $\square$

Lemma A.6. If h is not continuous, then Theorem 1 fails to hold.

Proof. We provide an example where a discontinuous h leads to the violation of Theorem 1. Let $h(y, z) = \mathbf{I}(y \neq z)$. Obviously, this h function is not continuous in $[0, 1]^2$. Let $F = \Phi$ be the c.d.f. of standard normal distribution. Let X and Y be continuous random variables, satisfying $Y = f(X)$ for some measurable function f . In such case, we have $\zeta_n^{(h, F)}(X, Y) = \chi_n^{(h, F)}(X, Y) \equiv 1$, implying that $\xi_n^{(h, F)}(X, Y) \equiv 1 - \zeta_n^{(h, F)}(X, Y) / \chi_n^{(h, F)}(X, Y) = 0$. On the other hand, from the definitions of $\zeta^{(h, F)}$ and $\chi^{(h, F)}$, we have $\zeta^{(h, F)} = 0$ and $\chi^{(h, F)} = 1$, which gives that $\xi^{(h, F)} = 1$. Therefore, the convergence of $\xi_n^{(h, F)}(X, Y) \rightarrow \xi^{(h, F)}$ fails to hold, and Theorem 1 is violated. $\square$

Lemma A.7. For $\xi^{(h, F)}$ in (4.4), let $h(y, z) = |y - z|$, and let $F = F_Y$ be the c.d.f. of Y . Then we have

$$\xi^{(h, F_Y)} = \frac{\int \text{var}(\mathbf{E}(\mathbf{I}(Y \geq t) \mid X)) \, d\mu_Y(t)}{\int \text{var}(\mathbf{I}(Y \geq t)) \, d\mu_Y(t)}.$$

Proof. Recall the function $g(\cdot, \cdot)$ defined in (A.5). When taking $h(y, z) = |y - z|$ and

$F = F_Y$, we have

$$\begin{aligned}
 g(X, X) &= \int \int |F_Y(y) - F_Y(z)| \, d\tilde{\mu}_X(y) \, d\tilde{\mu}_X(z) \\
 &= \int \int \int \{I(t \leq y) - I(t \leq z)\}^2 \, d\mu_Y(t) \, d\tilde{\mu}_X(y) \, d\tilde{\mu}_X(z) \\
 &= \int \int \int \{I(t \leq y) + I(t \leq z) - 2I(t \leq y) \cdot I(t \leq z)\} \, d\tilde{\mu}_X(y) \, d\tilde{\mu}_X(z) \, d\mu_Y(t) \\
 &= 2 \int \left\{ \int I(t \leq y)^2 \, d\tilde{\mu}_X(y) - \left(\int I(t \leq y) \, d\tilde{\mu}_X(y) \right)^2 \right\} \, d\mu_Y(t) \\
 &= 2 \int \text{var}(I(Y \geq t) \mid X) \, d\mu_Y(t).
 \end{aligned}$$

Then

$$\zeta^{(h, F_Y)} = E(g(X, X)) = 2 \int E(\text{var}(I(Y \geq t) \mid X)) \, d\mu_Y(t).$$

Similarly, we have

$$\begin{aligned}
 \chi^{(h, F_Y)} &= \int \int |F_Y(y) - F_Y(z)| \, d\mu_Y(y) \, d\mu_Y(z) \\
 &= 2 \int \left\{ \int I(t \leq y)^2 \, d\mu_Y(y) - \left(\int I(t \leq y) \, d\mu_Y(y) \right)^2 \right\} \, d\mu_Y(t) \\
 &= 2 \int \text{var}(I(Y \geq t)) \, d\mu_Y(t).
 \end{aligned}$$

Putting these together, we have

$$\begin{aligned}
 \xi^{(h, F_Y)} &= 1 - \frac{\int E(\text{var}(I(Y \geq t) \mid X)) \, d\mu_Y(t)}{\int \text{var}(I(Y \geq t)) \, d\mu_Y(t)} \\
 &= \frac{\int \text{var}(E(I(Y \geq t) \mid X)) \, d\mu_Y(t)}{\int \text{var}(I(Y \geq t)) \, d\mu_Y(t)}.
 \end{aligned}$$

This completes the proof. $\square$

S1.2 Proof of Theorem 2

The property (P3)' and the sufficiency part of property (P2) are proved in Remark 2.10 and Remark 2.12 respectively. It remains to prove the necessity part of property (P2), that is, $\xi^{(h,F)}$ reaches its maximum value of 1 only if $Y = f(X)$ a.s. for some measurable function f .

If $Y = f(X)$ does not hold a.s. for any measurable f , then there exists a Borel set $A \subseteq \mathbb{R}$ with $\mu_X(A) > 0$, such that for all $x \in A$, the probability measure $\tilde{\mu}_x$ is not a point mass. By conditions (C1) and (C2), $h(F(y), F(z)) > 0$ for all $y \neq z$. Then for all $x \in A$, we have

$$g(x, x) = \int \int h(F(y), F(z)) \, d\tilde{\mu}_x(y) \, d\tilde{\mu}_x(z) > 0. \quad (\text{A.9})$$

This implies that

$$\mathbb{P}(g(X, X) > 0) \geq \mathbb{P}(X \in A) = \mu_X(A) > 0,$$

and thus $\zeta^{(h,F)} = \mathbb{E}(g(X, X)) > 0$. It follows that $\xi^{(h,F)} = 1 - \zeta^{(h,F)} / \chi^{(h,F)} < 1$. This completes the proof. $\square$

S1.3 Proof of Theorem 3

The proof of Theorem 3 follows a similar approach to that of Theorem 2.2 in [Chatterjee \(2021\)](#), although with many modifications included. Before the proof, we first present the following notations, definitions, and Lemmas [A.8–A.12](#).

Let π be a rank vector for the X_i 's, chosen uniformly at random from all available choices in case of ties. Let $\tau = \pi^{-1}$, and define

$$D_n := \frac{1}{n} \sum_{i=1}^{n-1} a_i,$$

where $a_i := h(F(Y_{\tau(i)}), F(Y_{\tau(i+1)}))$. Also, for ease of presentation, we write

$$b_{i,j} := h(F(Y_i), F(Y_j)), \quad \text{for } i, j = 1, \dots, n.$$

Let E' and var' denote conditional expectation and conditional variance given $Y_1, \dots, Y_n$.

Lemma A.8. We have

$$E'(D_n) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n h(F(Y_i), F(Y_j)) = \chi_n^{(h,F)}(Y).$$

Proof. Note that X_i 's and Y_i 's are independent, and thus τ is a uniform random permutation on $\{1, 2, \dots, n\}$ which is independent of Y_i 's. Take any $i = 1, \dots, n-1$. Since $(\tau(i), \tau(i+1))$ is uniformly distributed over all pairs of (j, k) where j and k are distinct,

we have

$$E'(a_i) = \frac{1}{n \cdot (n-1)} \sum_{1 \leq j \neq k \leq n} h(F(Y_j), F(Y_k)).$$

Then

$$E'(D_n) = \frac{1}{n} \sum_{i=1}^{n-1} E'(a_i) = \frac{1}{n^2} \sum_{1 \leq j \neq k \leq n} h(F(Y_j), F(Y_k)) = \chi_n^{(h,F)}(Y),$$

where the last equation is due to $h(F(Y_j), F(Y_k)) = 0$ for all $j = k$. $\square$

Lemma A.9. We have

$$\text{var}'(D_n) = V_n + O(1/n^2), \tag{A.10}$$

where

$$V_n := \frac{1}{n^3} \sum_{p,q=1}^n b_{p,q}^2 - \frac{2}{n^4} \sum_{p,q,r=1}^n b_{p,q} b_{p,r} + \frac{1}{n^5} \sum_{p,q,r,s=1}^n b_{p,q} b_{r,s}. \tag{A.11}$$

Proof. This follows from the similar proof as that of Lemma B.2 in [Chatterjee \(2021\)](#).

(Note that the term $O(1/n^2)$ in (A.10) and the coefficients $1/n^3, 2/n^4, 1/n^5$ in V_n in (A.11) are of different orders in comparison to those outlined in [Chatterjee \(2021\)](#).

This is because that our a_i and $b_{i,j}$ are of order $O(1)$, while those in [Chatterjee \(2021\)](#) are of order $O(n)$.) $\square$

Lemma A.10. As $n \rightarrow \infty$, $n \cdot \text{var}'(D_n)$ converges almost surely to the deterministic limit

$$\sigma^2 := \mathbb{E}h(F(Y_1), F(Y_2))^2 - 2 \cdot \mathbb{E}(h(F(Y_1), F(Y_2)) \cdot h(F(Y_1), F(Y_3))) + (\mathbb{E}h(F(Y_1), F(Y_2)))^2,$$

where Y_1, Y_2, Y_3 are i.i.d. copies of Y .

Proof. By condition (C1), $h(F(y), F(z))$ is bounded from above by some constant $C_1 \in (0, \infty)$. Suppose that one Y_i is replaced by some other value Y'_i . Then $b_{p,q}$ does not change if $p \neq i$ and $q \neq i$; and changes by at most C_1 if one or both of p, q are equal to i . Thus, from the formulation of V_n in (A.11), replacing a single Y_i by Y'_i results in a change of V_n by no more than C_2/n^2 , where $C_2 \in (0, \infty)$ is some positive constant. By the bounded difference concentration inequality (McDiarmid et al., 1989), we have

$$\mathbb{P}(|V_n - \mathbb{E}(V_n)| \geq t) \leq 2e^{-C_2 t^2 \cdot n^3}, \quad \text{for } t \geq 0.$$

Therefore, by applying the Borel-Cantelli lemma, we have $n \cdot (V_n - \mathbb{E}(V_n)) \rightarrow 0$ a.s. as $n \rightarrow \infty$. On the other hand, by the strong law of large numbers for U-statistics (Hoeffding, 1961), we have that $n \cdot \mathbb{E}(V_n)$ converges a.s. to the limit σ^2 as $n \rightarrow \infty$. Putting the pieces together, we have

$$\lim_{n \rightarrow \infty} n \cdot \text{var}'(D_n) = \lim_{n \rightarrow \infty} n \cdot V_n = \lim_{n \rightarrow \infty} n \cdot \mathbb{E}(V_n) = \sigma^2, \quad \text{a.s..}$$

This completes the proof. $\square$

Lemma A.11. Assume that h and F satisfy conditions (C1) and (C2). Write $h^*(y, z) := h(F(y), F(z))$. For $n \geq 2$, let $g_n(y_1, y_2, \dots, y_n) := \min\{h^*(y_1, y_2), \dots, h^*(y_1, y_n)\}$. As $n \rightarrow \infty$, we have

$$g_n(Y_1, \dots, Y_n) \rightarrow 0, \quad \text{a.s..}$$

Proof. Note that $Y_1, Y_2, \dots, Y_n$ are i.i.d. copies of Y and h^* is bounded from above. For any fixed y in the support of Y , $h^*(y, Y_2), \dots, h^*(y, Y_n)$ is a sequence of bounded i.i.d. variables with minimum value of 0. We have $\min\{h^*(y, Y_2), \dots, h^*(y, Y_n)\} \rightarrow 0$ a.s. as $n \rightarrow \infty$. Therefore, $E(g_n(Y_1, \dots, Y_n) \mid Y_1) \rightarrow 0$ a.s.. Using the fact that g_n is non-negative and non-increasing with n , we have $g_n(Y_1, \dots, Y_n) \rightarrow 0$ almost surely as $n \rightarrow \infty$. $\square$

Lemma A.12. Suppose that Y is not a constant. Then the limit σ^2 in Lemma A.10 is strictly positive.

Proof. Let $h^*(y, z) = h(F(y), F(z))$ and $\phi(y) = E(h^*(y, Y'))$ where Y' represents an i.i.d. copy of Y . Following the proof of Lemma B.4 of Chatterjee (2021), we have

$$\sigma^2 = E(\{h^*(Y, Y') - \phi(Y) - \phi(Y') + E(h^*(Y, Y'))\}^2).$$

Thus, $\sigma^2 = 0$ if and only $h^*(Y, Y') - \phi(Y) - \phi(Y') + E(h^*(Y, Y')) = 0$ a.s.. Suppose

that this is true. Then for each $i \geq 2$, we have

$$h^*(Y_1, Y_i) = \phi(Y_1) + \phi(Y_i) - \mathbb{E}(h^*(Y_1, Y_i)), \quad \text{a.s..} \quad (\text{A.12})$$

Taking the minimum over $i = 1, \dots, n$ on both sides yields that

$$\min_{i=2, \dots, n} \{h^*(Y_1, Y_i)\} = \phi(Y_1) + \min_{i=2, \dots, n} \{\phi(Y_i)\} - \mathbb{E}(h^*(Y_1, Y_i)).$$

By Lemma A.11, we have that $\min_{i=2, \dots, n} \{h^*(Y_1, Y_i)\} \rightarrow 0$ a.s.. Since Y_i 's are i.i.d. and $\phi(Y_i)$ is bounded, $\min_{i=2, \dots, n} \{\phi(Y_i)\}$ converges to a deterministic limit (i.e., the minimum of the support of $\phi(Y)$) almost surely. This implies that $\phi(Y_1)$ equals a constant a.s.. By (A.12), $h^*(Y, Y')$ also equals a constant a.s..

By condition (C2) and the fact that Y is not a constant, there exists some $\delta_1 > 0$ such that $\mathbb{P}(|F(Y) - F(Y')| > \delta_1) > 0$. By condition (C1), $h(y, z)$ is uniformly continuous on $[0, 1] \times [0, 1]$, and satisfies $h(y, z) > 0$ for all (y, z) in the set $\mathcal{T}_1 := \{(y, z) \in [0, 1] \times [0, 1] : |y - z| \geq \delta_1\}$. Since $\mathcal{T}_1$ is a compact set, we have $\epsilon := \inf_{(y, z) \in \mathcal{T}_1} h(y, z) > 0$. Again, due to the continuity of h , there exists some $\delta_2 \in (0, \delta_1)$, such that $h(y, z) < \epsilon/2$ for any (y, z) in the set $\mathcal{T}_2 := \{(y, z) \in [0, 1] \times [0, 1] : |y - z| \leq \delta_2\}$. Also, due to the continuity of F , we have $\mathbb{P}(|F(Y) - F(Y')| \leq \delta_2) > 0$. Combining the results above yields that

$$\mathbb{P}(h^*(Y, Y') \geq \epsilon) \geq \mathbb{P}(|F(Y) - F(Y')| > \delta_1) > 0,$$

$$P(h^*(Y, Y') \leq \epsilon/2) \leq P(|F(Y) - F(Y')| \leq \delta_2) > 0.$$

This contradicts with the fact that $h^*(Y, Y')$ equals a constant almost surely. Thus, the assumption $\sigma^2 = 0$ is incorrect. The proof is completed. $\square$

Proof of Theorem 3. Notice that $Y_{[i]} = Y_{\tau(i)}$, which implies that $\zeta_n^{(h,F)}(X, Y) = D_n$. Also, Lemma A.8 provides that $\chi_n^{(h,F)}(Y) = E'(D_n)$. Therefore, we have

$$\xi_n^{(h,F)}(X, Y) = 1 - \frac{\zeta_n^{(h,F)}(X, Y)}{\chi_n^{(h,F)}(Y)} = -\frac{D_n - E'(D_n)}{E'(D_n)}.$$

Note that $E'(D_n) = \chi_n^{(h,F)}(Y) \rightarrow \chi^{(h,F)} = Eh(F(Y_1), F(Y_2))$ almost surely as $n \rightarrow \infty$.

Comparing the limits σ and σ in Theorem 3 and Lemma A.10 gives that $\sigma = \sigma/\chi^{(h,F)}$.

Putting these together, we get

$$\frac{\sqrt{n} \cdot \xi_n^{(h,F)}(X, Y)}{\sigma} = -\frac{\sqrt{n}\{D_n - E'(D_n)\}}{\sigma} \cdot \frac{\chi^{(h,F)}}{E'(D_n)} = -\frac{\sqrt{n}\{D_n - E'(D_n)\}}{\sigma} \cdot \{1 + o_P(1)\}.$$

It suffices to prove the central limit theorem:

$$\frac{\sqrt{n}\{D_n - E'(D_n)\}}{\sigma} \xrightarrow{D} N(0, 1), \quad \text{as } n \rightarrow \infty,$$

where $\sigma^2 = \lim_{n \rightarrow \infty} n \text{var}'(D_n) > 0$. The remaining proof follows from that of Theorem 2.2 in the supplementary material (on page 14) of Chatterjee (2021), by replacing the term $\min\{R(\tau(i)), R(\tau(i+1))\}$ in Chatterjee (2021) with our $h(F(Y_{\tau(i)}), F(Y_{\tau(i+1)}))$. $\square$

Lemma A.13. Let $T_n = n \cdot V_n / (\chi_n^{(h,F)}(Y))^2$, where V_n is defined in (A.11). Then as $n \rightarrow \infty$, T_n converges a.s to σ^2 . Thus, T_n is a consistent estimator of σ^2 .

Proof. Using the strong law of large numbers for U-statistics, we have $\chi_n^{(h,F)}(Y) \rightarrow Eh(F(Y_1), F(Y_2))$ a.s.. Also, Lemmas A.9 and A.10 verify that $n \cdot V_n \rightarrow \acute{\sigma}^2$ a.s., where $\acute{\sigma}^2$ defined in Lemma A.10 is the numerator of σ^2 in Theorem 3. Combining these two results completes the proof. $\square$

S2 Proof of theorems in Section 3

S2.1 Proof of Proposition 1

Before proving Proposition 1, we first present Lemmas B.1–B.3.

Lemma B.1. Let $\{\phi_i\}_{i=1,2,\dots}$ be a set of orthogonal basis functions in $L^2([0, 1])$. Then $\{\psi_{i,j}\}_{1 \leq i < j}$ with $\psi_{i,j}(x, y) := \phi_i(x)\phi_j(y)$ is a set of orthogonal basis functions in $L^2([0, 1]^2)$.

Proof. We first prove the statement: for any Borel sets $A, B \in \mathcal{B}([0, 1])$, the indicator function $f_{A,B}(x, y) := \mathbf{I}(x \in A, y \in B) = \mathbf{I}(x \in A)\mathbf{I}(y \in B)$ can be approximated by a finite linear combination of $\{\psi_{i,j}\}_{1 \leq i < j}$.

Write $f_A(x) := \mathbf{I}(x \in A)$ and $f_B(x) := \mathbf{I}(x \in B)$. Fix some $\epsilon > 0$. Since $\{\phi_i\}_{i=1,2,\dots}$ is a set of orthogonal bases in $L^2([0, 1])$, there exists some functions $\tilde{f}_A(x) = \sum_{i=1}^{N_1} d_i \phi_i(x)$ and $\tilde{f}_B(x) = \sum_{i=1}^{N_2} e_i \phi_i(x)$, such that $\|f_A - \tilde{f}_A\|_2 \leq \epsilon$ and $\|f_B - \tilde{f}_B\|_2 \leq \epsilon$. Let $\delta_A(x) =$

$|f_A(x) - \tilde{f}_A(x)|$ and $\delta_B(y) = |f_B(y) - \tilde{f}_B(y)|$. We have

$$\begin{aligned}
 \|f_{A,B} - \tilde{f}_A \tilde{f}_B\|_2^2 &= \int \int_{[0,1]^2} |f_A(x)f_B(y) - \tilde{f}_A(x)\tilde{f}_B(y)|^2 \, dx dy \\
 &\leq \int \int_{[0,1]^2} \{\delta_A(x) + \delta_B(y) + \delta_A(x)\delta_B(y)\}^2 \, dx dy \\
 &\leq 3 \int \int_{[0,1]^2} \delta_A^2(x) + \delta_B^2(y) + \delta_A^2(x)\delta_B^2(y) \, dx dy \\
 &= 3(\|\delta_A\|_2^2 + \|\delta_B\|_2^2 + \|\delta_A\|_2^2 \cdot \|\delta_B\|_2^2) \\
 &= 6\epsilon^2 + 3\epsilon^4.
 \end{aligned}$$

The previous statement is proved by noting that $\tilde{f}_A \tilde{f}_B$ is a finite linear combinations of $\{\psi_{i,j}\}_{1 \leq i < j}$ and ϵ is arbitrary.

By using Theorem 4.1.17 in Guo (2005), for any $f \in L^2([0,1]^2)$ and $\epsilon > 0$, there exists a simple function $\tilde{f}(x,y) = \sum_{i=1}^N c_i \mathbf{I}(x \in A_i, y \in B_i)$ such that $\|f - \tilde{f}\|_2 \leq \epsilon$. Combining this and the previous statement, the set of finite linear combinations of $\{\psi_{i,j}\}_{1 \leq i < j}$ can approximate any f , and thus is dense in $L^2([0,1]^2)$. By Theorem 4.2.14 of Guo (2005), $\{\psi_{i,j}\}_{1 \leq i < j}$ forms a set of orthogonal bases in $L^2([0,1]^2)$. $\square$

Lemma B.2. For a symmetric continuous function $h \in L^2([0,1]^2)$, the following two properties of h are equivalent:

1. For any bounded $f \in L^2([0,1])$, s.t. $\int_{[0,1]} f(x) dx = 0$, the integral

$$\int \int_{[0,1]^2} f(x)f(y)h(x,y) \, dx \, dy$$

is non-positive.

2. There exists a symmetric positive-definite continuous function $\varphi \in L^2([0, 1]^2)$, and a continuous function $g \in L^2([0, 1])$, such that

$$h(x, y) = -\varphi(x, y) + g(x) + g(y), \quad \text{for any } (x, y) \in [0, 1]^2.$$

Proof. We first prove property 2 leads to property 1. Assume that $f \in L^2([0, 1])$ satisfies $\int_{[0,1]} f(x)dx = 0$. Note that for any function $g \in L^2([0, 1])$, we have

$$\int \int_{[0,1]^2} f(x)f(y)g(x) \, dx \, dy = \int_{[0,1]} f(x)g(x) \, dx \cdot \int_{[0,1]} f(y)dy = 0.$$

It follows that

$$\begin{aligned} & \int \int_{[0,1]^2} f(x)f(y)h(x, y) \, dx \, dy \\ &= - \int \int_{[0,1]^2} f(x)f(y)\varphi(x, y) \, dx \, dy \leq 0, \end{aligned}$$

where the last inequality is due to the fact that φ is positive-definite. Thus property 1 is satisfied.

Next we prove that property 1 leads to property 2. Consider the Fourier orthogonal basis functions $\{\phi_k\}_{k=1}^\infty$ in $L^2([0, 1])$:

$$\phi_1(x) \equiv 1, \quad \phi_{2k}(x) = \sqrt{2} \cos((2x - 1)k\pi), \quad \phi_{2k+1}(x) = \sqrt{2} \sin((2x - 1)k\pi), \quad k = 1, 2, \dots$$

For $i, j = 1, 2, \dots$, define the function $\psi_{i,j}$ in $L^2([0, 1]^2)$:

$$\psi_{i,j}(x, y) := \phi_i(x)\phi_j(y), \quad (x, y) \in [0, 1]^2.$$

By Lemma B.1, $\{\psi_{i,j}\}_{1 \leq i < j}$ is a set of orthogonal basis functions in $L^2([0, 1]^2)$. Then h can be uniquely represented as the series $h = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} k_{i,j} \psi_{i,j}$, where $k_{i,i} = \langle h, \psi_{i,i} \rangle$ and $k_{i,j} = k_{j,i} = 1/2 \cdot \langle h, \psi_{i,j} \rangle$ for $j \neq i$. Let $h^\dagger := \sum_{i=2}^{\infty} \sum_{j=2}^{\infty} k_{i,j} \psi_{i,j}$, and $g^\dagger := \sum_{i=2}^{\infty} k_{1,i} \phi_i$. Given that $\|h\|_2 < \infty$ and $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} k_{i,j}^2 < \infty$, then $h^\dagger \in L^2([0, 1]^2)$ and $g^\dagger \in L^2([0, 1])$ are both well-defined. Moreover, we have the decomposition

$$h(x, y) = h^\dagger(x, y) + g^\dagger(x) + g^\dagger(y) + k_{1,1}, \quad (\text{B.1})$$

where $k_{1,1} = \int \int_{[0,1]^2} h(x, y) dx dy$ is a constant. Let $s(y) := 1/2 \cdot \int_{x \in [0,1]} h(x, y) dx$. We observe that $s(y)$ is continuous for $y \in [0, 1]$, since h is continuous. Note that for any integer $i \geq 2$, we have

$$\begin{aligned} k_{1,i} &= 1/2 \cdot \langle h, \psi_{1,i} \rangle = 1/2 \cdot \int \int_{[0,1]^2} h(x, y) \phi_1(x) \phi_i(y) dx dy \\ &= \int_{[0,1]} s(y) \phi_i(y) dy = \langle s, \phi_i \rangle. \end{aligned}$$

Therefore, $g^\dagger = \sum_{i=2}^{\infty} \langle s, \phi_i \rangle \cdot \phi_i = s - \langle s, \phi_1 \rangle \cdot \phi_1 = s - k_{1,1}$ is also a continuous function.

For any index set $I \subseteq \{1, 2, \dots\}$ with cardinality $|I| < \infty$, define the matrix $K_I := (k_{i,j})_{(i,j) \in I \times I}$. For integer $N \geq 3$, define the set $I_{2,N} := \{2, 3, \dots, N\}$. In what

follows, we prove the statement:

$$\text{for any } N \geq 3, K_{I_{2,N}} \text{ is a negative-definite matrix.} \quad (\text{B.2})$$

If the above statement is not true, then there exist some integer $N \geq 3$ and a vector $(c_2, c_3, \dots, c_N)^\top \in \mathbb{R}^{N-1}$, such that $\sum_{i=2}^N \sum_{j=2}^N c_i c_j k_{i,j} > 0$. Let $f := \sum_{i=2}^N c_i \phi_i$. It is clear that f is bounded and $\int_{[0,1]} f(x) dx = \langle f, \phi_1 \rangle = 0$. Using the decomposition (B.1), we have

$$\begin{aligned} & \int \int_{[0,1]^2} f(x) f(y) h(x, y) \, dx \, dy \\ &= \int \int_{[0,1]^2} f(x) f(y) \left\{ h^\dagger(x, y) + g^\dagger(x) + g^\dagger(y) + k_{1,1} \right\} \, dx \, dy \\ &= \int \int_{[0,1]^2} f(x) f(y) h^\dagger(x, y) \, dx \, dy \\ &= \sum_{i=2}^N \sum_{j=2}^N c_i c_j \int \int_{[0,1]^2} \phi_i(x) \phi_j(y) h^\dagger(x, y) \, dx \, dy \\ &= \sum_{i=2}^N \sum_{j=2}^N c_i c_j \cdot \langle h^\dagger, \psi_{i,j} \rangle \\ &= 2 \cdot \sum_{i=2}^N \sum_{j=2}^N c_i c_j k_{i,j} > 0, \end{aligned}$$

which is contradict with property 1 of h . Thus, the statement (B.2) is proved.

For integer $\ell \geq 3$, define $h_\ell := \sum_{i=2}^\ell \sum_{j=2}^\ell k_{i,j} \psi_{i,j}$. Using (B.2), $\{h_\ell\}_{\ell \geq 3}$ is a set of negative-definite functions. Note that $h_\ell \rightarrow_{L_2} h^\dagger$, as $\ell \rightarrow \infty$. Thus there exists a subsequence $\{\ell_1, \ell_2, \dots\}$ such that $h_{\ell_m} \rightarrow h^\dagger$ a.e. as $m \rightarrow \infty$. Then for any integer

$N \geq 1$, and real numbers $c_1, c_2, \dots, c_N \in \mathbb{R}$ and $x_1, x_2, \dots, x_N \in [0, 1]$ (with a possibly exception of null subset of $[0, 1]$), we have

$$\sum_{i=1}^N \sum_{j=1}^N c_i c_j h^\dagger(x_i, x_j) = \lim_{m \rightarrow \infty} \sum_{i=1}^N \sum_{j=1}^N c_i c_j h_{\ell_m}(x_i, x_j) \geq 0,$$

which proves that $h^\dagger$ is a.e. negative-definite. Combining this with (B.1) and the fact that both h and $g^\dagger$ are continuous functions, we have that $h^\dagger$ is a symmetric continuous negative-definite function in $L^2([0, 1]^2)$. Finally, property 2 of h is satisfied by using (B.1) and letting $\varphi = -h^\dagger$ and $g = g^\dagger + 1/2 \cdot k_{1,1}$. $\square$

Lemma B.3. Assume that h and F satisfy the regularity conditions (C1) and (C2). If $\xi^{(h,F)}$ satisfies property (P1), then h satisfies property 1 in Lemma B.2.

Proof. Suppose that h does not satisfy property 1 in Lemma B.2. Then there exists some bounded $f \in L^2([0, 1])$, s.t. $\int_{[0,1]} f(x) dx = 0$, and the integral $\int \int_{[0,1]^2} f(x) f(y) h(x, y) dx dy > 0$. Obviously, this f is not a constant. Let $f^+(x) = \max\{f(x), 0\}$ and $f^-(x) = \max\{-f(x), 0\}$ be the positive and negative parts of f . Since $f \in L^2([0, 1])$, we have $c := \int f^+(x) dx = \int f^-(x) dx \in (0, \infty)$. Consider the random variable X with probability distribution $P(X = 1) = 1/2$ and $P(X = 2) = 1/2$. Let Z be a continuous random

variable with the following conditional p.d.f.:

$$f_{Z|X=x}(z) = \begin{cases} f^+(z)/c, & \text{if } x = 1, \\ f^-(z)/c, & \text{if } x = 2. \end{cases}$$

Since F satisfies condition (C2), F^{-1} exists and we can define $Y = F^{-1}(Z)$. For this distribution of (X, Y) , the corresponding $\chi^{(h,F)}$ and $\zeta^{(h,F)}$ satisfy

$$\begin{aligned} & \chi^{(h,F)} - \zeta^{(h,F)} \\ &= \int \left[\int \int -h(F(y), F(z)) \cdot \{d\mu_{Y|x}(y) - d\mu_Y(y)\} \cdot \{d\mu_{Y|x}(z) - d\mu_Y(z)\} \right] d\mu_X(x) \\ &= \int \left[\int \int_{[0,1]^2} -h(y, z) \cdot \{d\mu_{Z|x}(y) - d\mu_Z(y)\} \cdot \{d\mu_{Z|x}(z) - d\mu_Z(z)\} \right] d\mu_X(x) \\ &= \frac{1}{2c} \int \int_{[0,1]^2} -h(y, z) \cdot \{f^+(y) - f^-(y)\} \cdot \{f^+(z) - f^-(z)\} dy dz \\ &\quad + \frac{1}{2c} \int \int_{[0,1]^2} -h(y, z) \cdot \{f^-(y) - f^+(y)\} \cdot \{f^-(z) - f^+(z)\} dy dz \\ &= \frac{1}{c} \int \int_{[0,1]^2} -h(y, z) \cdot f(y) \cdot f(z) dy dz < 0. \end{aligned}$$

It follows that $\xi^{(h,F)} = (\chi^{(h,F)} - \zeta^{(h,F)})/\chi^{(h,F)} < 0$, which violates property (P1). This completes the proof. $\square$

Proof of Proposition 1. We first prove the sufficiency part. Suppose that h satisfies condition (C3). Then there exists a positive-definite function φ , such that $h(y, z) =$

$1/2 \cdot \{\varphi(y, y) + \varphi(z, z)\} - \varphi(y, z)$. This combined with Theorem 1 gives that

$$\begin{aligned}
 & \chi^{(h,F)} - \zeta^{(h,F)} \\
 &= \int \left[\int \int -h(F(y), F(z)) \cdot \{\mathrm{d}\mu_{Y|x}(y) - \mathrm{d}\mu_Y(y)\} \cdot \{\mathrm{d}\mu_{Y|x}(z) - \mathrm{d}\mu_Y(z)\} \right] \mathrm{d}\mu_X(x) \\
 &= \int \left[\int \int \varphi(F(y), F(z)) \cdot \{\mathrm{d}\mu_{Y|x}(y) - \mathrm{d}\mu_Y(y)\} \cdot \{\mathrm{d}\mu_{Y|x}(z) - \mathrm{d}\mu_Y(z)\} \right] \mathrm{d}\mu_X(x) \\
 &\geq 0,
 \end{aligned}$$

where the last inequality is because φ is positive-definite. Thus $\xi^{(h,F)} = (\chi^{(h,F)} - \zeta^{(h,F)})/\chi^{(h,F)} \geq 0$. On the other hand, it is obvious that $\xi^{(h,F)} \leq 1$ since h is non-negative. Combining these proves that $\xi^{(h,F)}$ satisfies property (P1).

Next, we prove the necessity part. Suppose that $\xi^{(h,F)}$ satisfies property (P1). By Lemma B.3, h satisfies property 1 in Lemma B.2. By Lemma B.2, h also satisfies property 2, and thus can be decomposed as

$$h(y, z) = -\varphi(y, z) + g(y) + g(z), \quad \text{for any } (y, z) \in [0, 1]^2,$$

where φ is a symmetric continuous positive-definite function. By the regularity condition (C1), $h(y, y) = 0$ for all $y \in [0, 1]$. It follows that $0 = h(y, y) = -\varphi(y, y) + 2 \cdot g(y)$, which yields $g(y) = 1/2 \cdot \varphi(y, y)$. Therefore, $h(y, z) = -\varphi(y, z) + 1/2 \cdot \{\varphi(y, y) + \varphi(z, z)\}$ satisfies condition (C3). This completes the proof. $\square$

S2.2 Proof of Theorem 4

We first prove the necessity part. Suppose that $\xi^{(h,F)}$ satisfies properties (P1), (P2), and (P3). By Proposition 1, there exists a continuous positive-definite function φ such that $h(y, z) = 1/2 \cdot \{\varphi(y, y) + \varphi(z, z)\} - \varphi(y, z)$. It suffices to show that φ satisfies condition (C4). If this is not true, then there exist two distinct probability distributions μ_1 and μ_2 with $\mu_1((0, 1)) = \mu_2((0, 1)) = 1$, such that $E_{Z \sim \mu_1} \varphi(t, Z) = E_{Z \sim \mu_2} \varphi(t, Z)$ for all $t \in [0, 1]$. Let $Z_1 \sim \mu_1$ and $Z_2 \sim \mu_2$. By the fact that $\mu_1((0, 1)) = \mu_2((0, 1)) = 1$ and condition (C2) of F , both random variables $F^{-1}(Z_1)$ and $F^{-1}(Z_2)$ are well-defined. Denote by η_1 and η_2 the distribution laws of $F^{-1}(Z_1)$ and $F^{-1}(Z_2)$ respectively. Consider the random variable X with probability distribution $P(X = 1) = 1/2$ and $P(X = 2) = 1/2$. Let $Y|\{X = 1\} \sim \eta_1$ and $Y|\{X = 2\} \sim \eta_2$. Obviously, X and Y are not independent, and $\mu_Y = (\eta_1 + \eta_2)/2$. For this (X, Y) , when fixing $x = 1$, we have

$$\begin{aligned}
& \int \int \varphi(F(y), F(z)) \cdot \{d\mu_{Y|x}(y) - d\mu_Y(y)\} \cdot \{d\mu_{Y|x}(z) - d\mu_Y(z)\} \\
&= \frac{1}{4} \cdot \int \left[\int \varphi(F(y), F(z)) \cdot \{d\eta_1(z) - d\eta_2(z)\} \right] \{d\eta_1(y) - d\eta_2(y)\} \\
&= \frac{1}{4} \cdot \int_{[0,1]} \left[\int_{[0,1]} \varphi(y, z) \cdot \{d\mu_1(z) - d\mu_2(z)\} \right] \{d\mu_1(y) - d\mu_2(y)\} \\
&= \frac{1}{4} \cdot \int_{[0,1]} 0 \cdot \{d\mu_1(y) - d\mu_2(y)\} = 0.
\end{aligned}$$

Similarly, when fixing $x = 2$, the above equation also holds. Then

$$\chi^{(h,F)} - \zeta^{(h,F)}$$

$$\begin{aligned}
 &= \int \left[\int \int -h(F(y), F(z)) \cdot \{d\mu_{Y|x}(y) - d\mu_Y(y)\} \cdot \{d\mu_{Y|x}(z) - d\mu_Y(z)\} \right] d\mu_X(x) \\
 &= \int \left[\int \int \varphi(F(y), F(z)) \cdot \{d\mu_{Y|x}(y) - d\mu_Y(y)\} \cdot \{d\mu_{Y|x}(z) - d\mu_Y(z)\} \right] d\mu_X(x) \\
 &= 0.
 \end{aligned}$$

It follows that $\xi^{(h,F)} = (\chi^{(h,F)} - \zeta^{(h,F)})/\chi^{(h,F)} = 0$ holds for this pair of non-independent X and Y . This violates property (P3). Therefore, φ must satisfy condition (C4).

Next, we prove the sufficiency part. Suppose that h satisfies condition (C5). Using the results in Proposition 1 and Theorem 2, it suffices to prove that $\xi^{(h,F)} > 0$ for any non-independent X and Y . Note that φ is symmetric and positive-definite. By the Moore-Aronszajn Theorem ([Aronszajn, 1950](#)), there exists a unique reproducing kernel Hilbert space (RKHS) $\mathcal{H}_\varphi$, such that φ is its reproducing kernel. For $y \in [0, 1]$, write $\varphi_y := \varphi(\cdot, y) \in \mathcal{H}_\varphi$. Let $\langle \cdot, \cdot \rangle_{\mathcal{H}_\varphi}$ and $\|\cdot\|_{\mathcal{H}_\varphi}$ denote the inner product and norm respectively of this RKHS. Then by the Riesz representation theorem, we have $\varphi(y, z) = \langle \varphi_y, \varphi_z \rangle_{\mathcal{H}_\varphi}$ for any $y, z \in [0, 1]$. For any probability distribution μ , let $m_\varphi(\mu) := \int \varphi(\cdot, F(y)) d\mu(y) = E_{Y \sim \mu}(\varphi_{F(Y)})$ be the mean element with respect to μ . By Lemma 3.1 of [Muandet et al. \(2017\)](#) and the fact that φ is continuous, $m_\varphi(\mu)$ exists and belongs to $\mathcal{H}_\varphi$. For any x in the support of X , we have

$$\begin{aligned}
 &\int \int \varphi(F(y), F(z)) \cdot \{d\mu_{Y|x}(y) - d\mu_Y(y)\} \cdot \{d\mu_{Y|x}(z) - d\mu_Y(z)\} \\
 &= \int \left[\int \langle \varphi_{F(y)}, \varphi_{F(z)} \rangle_{\mathcal{H}_\varphi} \cdot \{d\mu_{Y|x}(y) - d\mu_Y(y)\} \right] \cdot \{d\mu_{Y|x}(z) - d\mu_Y(z)\}
 \end{aligned}$$

$$\begin{aligned}
 &= \int \langle m_\varphi(\mu_{Y|x}) - m_\varphi(\mu_Y), \varphi_{F(z)} \rangle_{\mathcal{H}_\varphi} \cdot \{d\mu_{Y|x}(z) - d\mu_Y(z)\} \\
 &= \|m_\varphi(\mu_{Y|x}) - m_\varphi(\mu_Y)\|_{\mathcal{H}_\varphi}.
 \end{aligned}$$

By condition (C2), F is a one-to-one map. Then for any $Y_1 \sim \mu_1$ and $Y_2 \sim \mu_2$ with distinct distributions μ_1 and μ_2 , the distributions of $F(Y_1)$ and $F(Y_2)$ are also distinct. Denote by η_1 and η_2 the distribution laws of $F(Y_1)$ and $F(Y_2)$ respectively. We have $\eta_1 \neq \eta_2$ and $\eta_1((0, 1)) = \eta_2((0, 1)) = 1$. Using this and condition (C4), $m_\varphi(\mu_1) = \int \varphi(\cdot, F(y))d\mu_1(y) = \int \varphi(\cdot, y)d\eta_1(y)$ is distinct from $m_\varphi(\mu_2) = \int \varphi(\cdot, F(y))d\mu_2(y) = \int \varphi(\cdot, y)d\eta_2(y)$. Therefore, if $\mu_{Y|x}$ and μ_Y are distinct, then $\|m_\varphi(\mu_{Y|x}) - m_\varphi(\mu_Y)\|_{\mathcal{H}_\varphi} > 0$. Since X and Y are not independent, there exists some Borel set $A \subseteq \mathbb{R}$ with $\mu_X(A) > 0$, such that $\mu_{Y|x} \neq \mu_Y$ for all $x \in A$. It follows that

$$\begin{aligned}
 &\chi^{(h,F)} - \zeta^{(h,F)} \\
 &= \int \left[\int \int \varphi(F(y), F(z)) \cdot \{d\mu_{Y|x}(y) - d\mu_Y(y)\} \cdot \{d\mu_{Y|x}(z) - d\mu_Y(z)\} \right] d\mu_X(x) \\
 &= \int \|m_\varphi(\mu_{Y|x}) - m_\varphi(\mu_Y)\|_{\mathcal{H}_\varphi} \cdot d\mu_X(x) \\
 &\geq \int_{x \in A} \|m_\varphi(\mu_{Y|x}) - m_\varphi(\mu_Y)\|_{\mathcal{H}_\varphi} \cdot d\mu_X(x) > 0.
 \end{aligned}$$

This completes the proof. □

S3 Proofs of theorems in Section 4

S3.1 Proof of Theorem 5

Before proving Theorem 5, we first present Lemmas C.1 and C.2.

Lemma C.1. Assume function h satisfies conditions (C1) and (C6). Then we have

$$\lim_{n \rightarrow \infty} \mathbb{E}(\tilde{\zeta}_n^{(h)}(X, Y)) = \tilde{\zeta}^{(h)}.$$

Proof. Let $\zeta_n^{*(h)}(X, Y) := \frac{1}{n} \sum_{i=1}^n h(F_Y(Y_i), F_Y(Y_{n,i}))$. Let $\Delta_n := \sup_{t \in \mathbb{R}} |F_{Y,n}(t) - F_Y(t)|$. By condition (C6), we have

$$\begin{aligned} |\tilde{\zeta}_n^{(h)}(X, Y) - \zeta_n^{*(h)}(X, Y)| &\leq \frac{1}{n} \sum_{i=1}^n \left| h(F_{Y,n}(Y_i), F_{Y,n}(Y_{n,i})) - h(F_Y(Y_i), F_Y(Y_{n,i})) \right| \\ &\leq \frac{1}{n} \sum_{i=1}^n K_h \cdot \{|F_{Y,n}(Y_i) - F_Y(Y_i)| + |F_{Y,n}(Y_{n,i}) - F_Y(Y_{n,i})|\} \\ &\leq 2 \cdot K_h \Delta_n. \end{aligned}$$

By the Glivenko-Cantelli theorem (Tucker, 1959), $\Delta_n \rightarrow 0$ almost surely as $n \rightarrow \infty$.

Using this and the fact that Δ_n is bounded by 1, it follows that

$$\lim_{n \rightarrow \infty} \mathbb{E}(|\tilde{\zeta}_n^{(h)}(X, Y) - \zeta_n^{*(h)}(X, Y)|) = 0.$$

Thus it suffices to show that $\mathbb{E}(\zeta_n^{*(h)}(X, Y)) \rightarrow \tilde{\zeta}^{(h)}$. This follows from Lemma A.4 (by letting $F(\cdot) = F_Y(\cdot)$). $\square$

Lemma C.2. Assume function h satisfies conditions (C1) and (C6). There exists a

constant $C \in (0, \infty)$, such that for any integer $n \geq 1$ and any real number $t \geq 0$,

$$P(|\tilde{\zeta}_n^{(h)}(X, Y) - E(\tilde{\zeta}_n^{(h)}(X, Y))| \geq t) \leq 2e^{-C \cdot n \cdot t^2}.$$

Proof. Assume continuity of X . Suppose that for some $i_0 \leq n$, (X_{i_0}, Y_{i_0}) is replaced by a different value (X'_{i_0}, Y'_{i_0}) . According to the proof of Lemma A.5, there are at most 3 pairs of $(Y_i, Y_{N(i)})$ that changes after the replacement. Since h is continuous in $[0, 1] \times [0, 1]$, there exists a constant $c \in (0, \infty)$ such that $\sup_{(y,z) \in \mathbb{R}^2} h(F(y), F(z)) < c$. Note that for each fixed $t \in \mathbb{R}$, $F_{Y,n}(t)$ changes by at most $1/n$ after the replacement. By condition (C6), $h(F_{Y,n}(Y_i), F_{Y,n}(Y_{N(i)}))$ changes by at most $2K_h/n$ for those unchanged pairs of $(Y_i, Y_{N(i)})$. In summary, after the replacement, $\tilde{\zeta}_n^{(h)}(X, Y)$ changes by at most

$$\Delta_n = \frac{1}{n} \left(3c + n \cdot \frac{2K_h}{n} \right) = \frac{2K_h + 3c}{n}.$$

By the bounded difference concentration inequality (McDiarmid et al., 1989), we have

$$P(|\tilde{\zeta}_n^{(h)}(X, Y) - E(\tilde{\zeta}_n^{(h)}(X, Y))| \geq t) \leq 2e^{-2 \cdot t^2 / (n \Delta_n^2)} = 2e^{-C \cdot n \cdot t^2},$$

where $C = 2 \cdot (2K_h + 3c)^{-2}$ is a positive constant. For the general case of X , the result follows from the similar proof as that of Lemma A.11 in Chatterjee (2021). $\square$

Proof of Theorem 5. Note that for any $t > 0$, we have $\sum_{n=1}^{\infty} 2e^{-C \cdot n \cdot t^2} < \infty$. By Lemma C.2 and the Borel-Cantelli lemma, $|\tilde{\zeta}_n^{(h)}(X, Y) - E(\tilde{\zeta}_n^{(h)}(X, Y))| \rightarrow 0$ a.s., as $n \rightarrow \infty$.

This combined with Lemma C.1 yields that $\tilde{\zeta}_n^{(h)}(X, Y) \rightarrow \tilde{\zeta}^{(h)}$ a.s. as $n \rightarrow \infty$, which proves result (i).

Let $\chi_n^{*(h)}(Y) := \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n h(F_Y(Y_i), F_Y(Y_{n,i}))$. Following a similar proof as that of Lemma C.1, we have

$$|\tilde{\chi}_n^{(h)}(Y) - \chi_n^{*(h)}(Y)| \rightarrow 0, \quad \text{a.s. as } n \rightarrow \infty.$$

Using the strong law of large numbers for U-statistics (Hoeffding, 1961), $\chi_n^{*(h)}(Y) \rightarrow \tilde{\chi}^{(h)}$ a.s. as $n \rightarrow \infty$. By condition (C1) of h and the fact that Y is not a constant, we have $\tilde{\chi}^{(h)} > 0$. Putting these together proves result (ii). Finally, result (iii) directly follows from combining results (i) and (ii). $\square$

S3.2 Proof of Theorem 6

The proof of Theorem 6 is similar to that of Theorem 3. Before the proof, we first present the following notations, definitions, and Lemmas C.3–C.5. Let π be a rank vector for the X_i 's, chosen uniformly at random from all available choices in case of ties. Let $\tau = \pi^{-1}$, and define

$$\tilde{D}_n := \frac{1}{n} \sum_{i=1}^{n-1} \tilde{a}_i,$$

where $\tilde{a}_i := h(F_{Y,n}(Y_{\tau(i)}), F_{Y,n}(Y_{\tau(i+1)}))$. Also, we write $\tilde{b}_{i,j} := h(F_{Y,n}(Y_i), F_{Y,n}(Y_j))$ for $i, j = 1, \dots, n$. Let E' and var' denote conditional expectation and conditional variance given $Y_1, \dots, Y_n$.

Lemma C.3. We have

$$\begin{aligned} E'(\tilde{D}_n) &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n h(F_{Y,n}(Y_i), F_{Y,n}(Y_j)) = \tilde{\chi}_n^{(h)}(Y), \\ \text{var}'(\tilde{D}_n) &= \tilde{V}_n + O(1/n^2), \end{aligned}$$

where

$$\tilde{V}_n := \frac{1}{n^3} \sum_{p,q=1}^n \tilde{b}_{p,q}^2 - \frac{2}{n^4} \sum_{p,q,r=1}^n \tilde{b}_{p,q} \tilde{b}_{p,r} + \frac{1}{n^5} \sum_{p,q,r,s=1}^n \tilde{b}_{p,q} \tilde{b}_{r,s}.$$

Proof. This directly follows from Lemmas A.8 and A.9, by substituting $F = F_{Y,n}$. $\square$

Lemma C.4. As $n \rightarrow \infty$, $n \cdot \text{var}'(\tilde{D}_n)$ converges almost surely to the deterministic limit

$$\begin{aligned} \check{\sigma}^2 &:= E h(F_Y(Y_1), F_Y(Y_2))^2 - 2 \cdot E(h(F_Y(Y_1), F_Y(Y_2)) \cdot h(F_Y(Y_1), F_Y(Y_3))) \\ &\quad + (E h(F_Y(Y_1), F_Y(Y_2)))^2, \end{aligned}$$

where Y_1, Y_2, Y_3 are i.i.d. copies of Y .

Proof. This follows the similar proof as that of Lemma A.10. Suppose that one Y_i is

replaced by some other value Y'_i . By condition (C6), for $p \neq i$ and $q \neq i$, $\tilde{b}_{p,q}$ changes by at most C_1/n , where $C_1 \in (0, \infty)$ is some constant; and if one or both of p, q are equal to i , then $\tilde{b}_{p,q}$ changes by at most C_2 , where $C_2 = \sup_{y,z \in \mathbb{R}} h(F_Y(y), F_Y(z)) \in (0, \infty)$ is a constant. Using these, replacing a single Y_i with Y'_i results in a modification of $\tilde{V}_n$ by no more than C_3/n^2 , where $C_3 \in (0, \infty)$ is some constant. By the bounded difference concentration inequality ([McDiarmid et al., 1989](#)), we have

$$P(|\tilde{V}_n - E(\tilde{V}_n)| \geq t) \leq 2e^{-C_3 t^2 \cdot n^3}, \quad \text{for } t \geq 0.$$

Therefore, $n \cdot (\tilde{V}_n - E(\tilde{V}_n)) \rightarrow 0$ almost surely as $n \rightarrow \infty$.

Let $\check{b}_{i,j} = h(F_Y(Y_i), F_Y(Y_j))$, for $i, j = 1, \dots, n$. Let

$$\check{V}_n = \frac{1}{n^3} \sum_{p,q=1}^n \check{b}_{p,q}^2 - \frac{2}{n^4} \sum_{p,q,r=1}^n \check{b}_{p,q} \check{b}_{p,r} + \frac{1}{n^5} \sum_{p,q,r,s=1}^n \check{b}_{p,q} \check{b}_{r,s}.$$

Let $\Delta_n = \sup_{t \in \mathbb{R}} |F_{Y,n}(t) - F_Y(t)|$. By condition (C6), we have $|\tilde{b}_{i,j} - \check{b}_{i,j}| \leq 2K_h \Delta_n$ for $i, j = 1, \dots, n$. Thus, it is easy to check that $|\tilde{V}_n - \check{V}_n| \leq C_4 \Delta_n / n$ for some constant $C_4 \in (0, \infty)$. By the Glivenko-Cantelli theorem ([Tucker, 1959](#)), we have $\Delta_n \leq 1$ and $\Delta_n \rightarrow 0$ a.s. as $n \rightarrow \infty$. This yields $\lim_{n \rightarrow \infty} n \cdot E(|\tilde{V}_n - \check{V}_n|) = 0$. On the other hand, by the strong law of large numbers for U-statistics ([Hoeffding, 1961](#)), we have that $n \cdot E(\check{V}_n)$

converges a.s. to the limit $\check{\sigma}^2$ as $n \rightarrow \infty$. Putting the pieces together, we have

$$\lim_{n \rightarrow \infty} n \cdot \text{var}'(\tilde{D}_n) = \lim_{n \rightarrow \infty} n \cdot \tilde{V}_n = \lim_{n \rightarrow \infty} n \cdot \mathbb{E}(\tilde{V}_n) = \lim_{n \rightarrow \infty} n \cdot \mathbb{E}(\check{V}_n) = \check{\sigma}^2, \quad \text{a.s..}$$

This completes the proof. □

Lemma C.5. Suppose that Y is not a constant. Then the limit $\check{\sigma}^2$ in Lemma C.4 is strictly positive.

Proof. This follows from the similar proof to that of Lemma A.12. The only difference is that we need to check the modified version of function $\tilde{h}^*(y, z) := h(F_Y(y), F_Y(z))$ is not a constant a.s. (instead of $h^*(y, z) = h(F(y), F(z))$ in Lemma A.12). If Y is continuous, then $U_1 := F_Y(Y)$ and $U_2 := F_Y(Y')$ are random variables that follow i.i.d. Unif[0,1] distribution. By condition (C1) of h , $\tilde{h}^*(Y, Y') = h(U_1, U_2)$ is not a constant almost surely. If Y is not continuous, then there exists some $y_0 \in \mathbb{R}$ s.t. $p_0 := \mathbb{P}(Y = y_0) \in (0, 1)$. Using condition (C1), we have

$$\begin{aligned} \mathbb{P}(\tilde{h}^*(Y, Y') = 0) &\geq \mathbb{P}(Y = Y' = y_0) = p_0^2 > 0, \\ \mathbb{P}(\tilde{h}^*(Y, Y') > 0) &\geq \mathbb{P}(Y = y_0, Y' \neq y_0) = p_0 \cdot (1 - p_0) > 0. \end{aligned}$$

This proves $\tilde{h}^*(Y, Y')$ is not a constant almost surely. □

Proof of Theorem 6. Similarly to the proof of Theorem 3, we have

$$\frac{\sqrt{n} \cdot \tilde{\xi}_n^{(h)}(X, Y)}{\tilde{\sigma}} = -\frac{\sqrt{n}\{\tilde{D}_n - E'(\tilde{D}_n)\}}{\check{\sigma}} \cdot \frac{\tilde{\chi}^{(h)}}{E'(\tilde{D}_n)} = -\frac{\sqrt{n}\{\tilde{D}_n - E'(\tilde{D}_n)\}}{\check{\sigma}} \cdot \{1 + o_P(1)\}.$$

It suffices to prove the central limit theorem:

$$\frac{\sqrt{n}\{\tilde{D}_n - E'(\tilde{D}_n)\}}{\check{\sigma}} \xrightarrow{\mathcal{D}} N(0, 1), \quad \text{as } n \rightarrow \infty,$$

where $\check{\sigma}^2 = \lim_{n \rightarrow \infty} n \text{var}'(\tilde{D}_n) > 0$. The remaining proof follows from that of Theorem 2.2 on page 14 of the supplementary material of [Chatterjee \(2021\)](#), by replacing the term $\min\{R(\tau(i)), R(\tau(i+1))\}$ in [Chatterjee \(2021\)](#) with our $h(F_{Y,n}(Y_{\tau(i)}), F_{Y,n}(Y_{\tau(i+1)}))$. $\square$

S3.3 Proof of Theorem 7

Proof of Theorem 7 result (i) The property (P3)' and the sufficiency part of property (P2) directly follows from the similar arguments as in Remark 2.12 and Remark 2.10 respectively (by substituting $F = F_Y$).

It remains to prove the necessity part of property (P2). This follows from the similar proof of Theorem 2. The only thing we need to verify in particular is the modified version of (A.9), that is, for all $x \in A$,

$$g(x, x) = \int \int h(F_Y(y), F_Y(z)) \, d\tilde{\mu}_x(y) \, d\tilde{\mu}_x(z) > 0.$$

By conditions (C1), we have $h(F_Y(y), F_Y(z)) > 0$ for all distinct y, z in the support of Y . This proves the above inequality. $\square$

Proof of Theorem 7 result (ii) Before the proof, we first show Lemma C.6.

Lemma C.6. Assume that h satisfies the regularity condition (C1). If $\tilde{\xi}^{(h)}$ satisfies property (P1), then h must satisfy property 1 in Lemma B.2.

Proof. Suppose that h does not satisfy property 1 in Lemma B.2. Then there exists some bounded $f \in L^2([0, 1])$, s.t. $\int_{[0,1]} f(x)dx = 0$, and $\int \int_{[0,1]^2} f(x)f(y)h(x, y) \, dx \, dy > 0$. Let $c := \sup_{x \in [0,1]} |f(x)| \in (0, \infty)$. Consider the random variable X with probability distribution $P(X = 1) = 1/2$ and $P(X = 2) = 1/2$. Let U be a continuous random variable with the following conditional p.d.f.:

$$f_{U|X=x}(u) = \begin{cases} 1 + f(u)/c, & \text{if } x = 1, \\ 1 - f(u)/c, & \text{if } x = 2. \end{cases}$$

Let $F_1(u) = P(U \leq u \mid X = 1)$ and $F_2(u) = P(U \leq u \mid X = 2)$ be the conditional c.d.f.'s of U . Clearly, $U \sim \text{Unif}[0,1]$, and $F_1(u) + F_2(u) = 2u$ for any $u \in [0, 1]$. Let Φ denote the c.d.f. of standard normal distribution, and let $Y = \Phi^{-1}(U)$. For any $y \in \mathbb{R}$, we have

$$P(Y \leq y) = 1/2 \cdot \{P(\Phi^{-1}(U) \leq y \mid X = 1) + P(\Phi^{-1}(U) \leq y \mid X = 2)\}$$

$$= 1/2 \cdot \{F_1(\Phi(y)) + F_2(\Phi(y))\} = \Phi(y).$$

Therefore, the c.d.f. of Y is $F_Y = \Phi$. For this distribution of (X, Y) , the corresponding $\tilde{\chi}^{(h)}$ and $\tilde{\zeta}^{(h)}$ satisfy

$$\begin{aligned} & \tilde{\chi}^{(h)} - \tilde{\zeta}^{(h)} \\ &= \int \left[\int \int -h(\Phi(y), \Phi(z)) \cdot \{d\mu_{Y|x}(y) - d\mu_Y(y)\} \cdot \{d\mu_{Y|x}(z) - d\mu_Y(z)\} \right] d\mu_X(x) \\ &= \int \left[\int \int_{[0,1]^2} -h(y, z) \cdot \{d\mu_{U|x}(y) - d\mu_U(y)\} \cdot \{d\mu_{U|x}(z) - d\mu_U(z)\} \right] d\mu_X(x) \\ &= \frac{1}{c} \int \int_{[0,1]^2} -h(y, z) \cdot f(y) \cdot f(z) \, dy \, dz < 0. \end{aligned}$$

It follows that $\tilde{\xi}^{(h)} = (\tilde{\chi}^{(h)} - \tilde{\zeta}^{(h)}) / \tilde{\chi}^{(h)} < 0$, which violates property (P1). This completes the proof. $\square$

Next, we prove result (ii) of Theorem 7. The sufficiency part directly follows from the proof of Proposition 1 (by letting $F = F_Y$). For the necessity part, Lemma C.6 proves that h satisfies property 1 in Lemma B.2. By Lemma B.2, h also satisfies property 2, and thus can be decomposed as $h(y, z) = -\varphi(y, z) + g(y) + g(z)$. The remaining proof is the same as that of Proposition 1. $\square$

Proof of Theorem 7 result (iii) This follows a similar proof to that of the sufficiency part of Theorem 4. Using results (i) and (ii) of Theorem 7, it suffices to prove that $\tilde{\xi}^{(h)} > 0$ for any non-independent X and Y . Let $\mathcal{H}_\varphi$ be the RKHS with φ being its

reproducing kernel. For probability distribution μ with $\mu((0, 1]) = 1$, let $m_\varphi(\mu) := \int \varphi(\cdot, u) d\mu(u) = E_{U \sim \mu}(\varphi_U)$ be the mean element with respect to μ . Let $U = F_Y(Y)$. For any x in the support of X , we have

$$\begin{aligned} & \int \int \varphi(F_Y(y), F_Y(z)) \cdot \{d\mu_{Y|x}(y) - d\mu_Y(y)\} \cdot \{d\mu_{Y|x}(z) - d\mu_Y(z)\} \\ &= \int \int_{[0,1]^2} \varphi(y, z) \cdot \{d\mu_{U|x}(y) - d\mu_U(y)\} \cdot \{d\mu_{U|x}(z) - d\mu_U(z)\} \\ &= \|m_\varphi(\mu_{U|x}) - m_\varphi(\mu_U)\|_{\mathcal{H}_\varphi}. \end{aligned}$$

By condition (C4)', for any distinct probability measures μ_1 and μ_2 with $\mu_1((0, 1]) = \mu_2((0, 1]) = 1$, we have $m_\varphi(\mu_1) \neq m_\varphi(\mu_2)$. Note that $\mu_{U|x}((0, 1]) = \mu_U((0, 1]) = 1$. Therefore, $\|m_\varphi(\mu_{U|x}) - m_\varphi(\mu_U)\|_{\mathcal{H}_\varphi} > 0$ if $\mu_{U|x} \neq \mu_U$. Since X and Y are not independent, there exists some Borel set $A \subseteq \mathbb{R}$ with $\mu_X(A) > 0$, such that $\mu_{U|x} \neq \mu_U$ for all $x \in A$. It follows that

$$\begin{aligned} & \tilde{\chi}^{(h)} - \tilde{\zeta}^{(h)} \\ &= \int \left[\int \int_{[0,1]^2} \varphi(y, z) \cdot \{d\mu_{U|x}(y) - d\mu_U(y)\} \cdot \{d\mu_{U|x}(z) - d\mu_U(z)\} \right] d\mu_X(x) \\ &= \int \|m_\varphi(\mu_{U|x}) - m_\varphi(\mu_U)\|_{\mathcal{H}_\varphi} \cdot d\mu_X(x) \\ &\geq \int_{x \in A} \|m_\varphi(\mu_{U|x}) - m_\varphi(\mu_U)\|_{\mathcal{H}_\varphi} \cdot d\mu_X(x) > 0. \end{aligned}$$

This completes the proof. □

Bibliography

- Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American mathematical society* 68(3), 337–404.
- Chatterjee, S. (2021). A new coefficient of correlation. *Journal of the American Statistical Association* 116(536), 2009–2022.
- Durrett, R. (2019). *Probability: theory and examples*, Volume 49. Cambridge university press.
- Guo, M. (2005). *Real and Functional Analysis*. Peking University Press.
- Hoeffding, W. (1961). The strong law of large numbers for u-statistics. Technical report, North Carolina State University. Dept. of Statistics.
- McDiarmid, C. et al. (1989). On the method of bounded differences. *Surveys in combinatorics* 141(1), 148–188.
- Muandet, K., K. Fukumizu, B. Sriperumbudur, and B. Schölkopf (2017, 01). Kernel mean embedding of distributions: A review and beyond. *Foundations and Trends in Machine Learning* 10, 1–141.
- Tucker, H. G. (1959). A generalization of the glivenko-cantelli theorem. *The Annals of Mathematical Statistics* 30(3), 828–830.