DETECT COMPLETE DEPENDENCE VIA TRACE CORRELATION IN THE PRESENCE OF MATRIX-VALUED RANDOM OBJECTS

Delin Zhao and Liping Zhu

Renmin University of China

Supplementary Material

This supplementary material includes the proofs for the propositions and theorems of the main paper. Section S1 presents the proofs of Propositions 1 and 2. Section S2 provides the proofs of Theorems 1 and 2. Additionally, Section S3 contains some technical lemmas. All notations used in this supplementary material are consistent with those used in the main text.

S1 Proofs of Propositions 1 and 2

Proof of Proposition 1: Denote the "metric" induced by $\omega_1(x)$ and $\omega_2(\mathbf{B})$ as

$$d_{\text{weight}}(\mathbf{X}_1, \mathbf{X}_2) = \int_x \int_{\mathbf{B}} \Big\{ I(\langle \mathbf{B}, \mathbf{X}_1 \rangle \le x) - I(\langle \mathbf{B}, \mathbf{X}_2 \rangle \le x) \Big\}^2 \omega_1(x)\omega_2(\mathbf{B})(d\mathbf{B})(dx).$$

Denote $\mathrm{supp}(\omega)$ as the support of ω, then $\mathrm{supp}(\omega_2) = \mathbb{R}^{p\times q}$. From the form of $d_{\mathrm{weight}}(\mathbf{X}_1, \mathbf{X}_2)$, symmetry, non-negativity and triangle inequality hold trivially. In addition, if $\mathbf{X}_1 = \mathbf{X}_2$, $d_{\mathrm{weight}}(\mathbf{X}_1, \mathbf{X}_2) = 0$ is also obvious. For the converse, if $d_{\mathrm{weight}}(\mathbf{X}_1, \mathbf{X}_2) = 0$, there exists some set $\mathcal{A} \subseteq \mathbb{R}^{p\times q}$ with $\omega_2(\mathcal{A}^c) = 0$, s.t. for any $\mathbf{B} \in \mathcal{A}$, $\int_x \{I(\langle \mathbf{B}, \mathbf{X}_1\rangle \leq x) - I(\langle \mathbf{B}, \mathbf{X}_2\rangle \leq x)\}^2\ \omega_1(x)(dx) = 0$. Assume $\mathbf{X}_1 \neq \mathbf{X}_2$, the set $\mathcal{B} = \{\mathbf{B} \in \mathbb{R}^{p\times q} : \langle \mathbf{B}, \mathbf{X}_1\rangle = \langle \mathbf{B}, \mathbf{X}_2\rangle\}$ can only have measure of 0. Then for any $\mathbf{B} \in \mathcal{A} \setminus \mathcal{B}$, where $\omega_2\{(\mathcal{A} \setminus \mathcal{B})^c\} = 0$, $\int_x \{I(\langle \mathbf{B}, \mathbf{X}_1\rangle \leq x) - I(\langle \mathbf{B}, \mathbf{X}_2\rangle \leq x)\}^2\ \omega_1(x)(dx) = 0$. However, for arbitrary $\mathbf{B} \in \mathcal{A}\setminus\mathcal{B}$, we can always find a set of x with positive measure falling between $\langle \mathbf{B}, \mathbf{X}_1\rangle$ and $\langle \mathbf{B}, \mathbf{X}_2\rangle$, thus $\int_x \{I(\langle \mathbf{B}, \mathbf{X}_1\rangle \leq x) - I(\langle \mathbf{B}, \mathbf{X}_2\rangle \leq x)\}^2\ \omega_1(x)(dx) > 0$, which implies a contradiction. Therefore, $(\mathbb{R}^{p\times q}, d_{\mathrm{weight}})$ is a metric space.

Observe that

$$\int_x \int_{\mathbf{B}} \mathrm{var}\left\{I(\langle \mathbf{B}, \mathbf{X}\rangle \leq x)\right\}\ \omega_1(x)\omega_2(\mathbf{B})(d\mathbf{B})(dx) = E\{d_{\mathrm{weight}}(\mathbf{X}_1, \mathbf{X}_2)\},$$

and

$$\int_x \int_{\mathbf{B}} E\left[\mathrm{var}\left\{I(\langle \mathbf{B}, \mathbf{X}\rangle \leq x) \mid Y\right\}\right]\ \omega_1(x)\omega_2(\mathbf{B})(d\mathbf{B})(dx) = E\{\widetilde{d}_{\mathrm{weight}}(Y)\},$$

provided $E\left\{d_{\mathrm{weight}}(\mathbf{X}_1, \mathbf{X}_2) \mid \mathbf{X}_1\right\} < \infty$ for some $\mathbf{X}_1 \in \mathbb{R}^{p\times q}$, from triangle inequality of $d_{\mathrm{weight}}(\mathbf{X}_1, \mathbf{X}_2)$ and Fubini's lemma. Therefore, (2.1) can be represented as

$$\mathrm{T}_{\text{weight}}(\mathbf{X} \mid Y) = 1 - E\{\widetilde{d}_{\text{weight}}(Y)\}/E\{d_{\text{weight}}(\mathbf{X}_1, \mathbf{X}_2)\}.$$

Next, we show its properties: $\mathrm{T}_{\text{weight}}(\mathbf{X} \mid Y) \in [0, 1]$ is obvious by noting that

$$\begin{aligned}\operatorname{var}\{I(\langle \mathbf{B}, \mathbf{X}\rangle \leq x)\} &= \operatorname{var}[E\{I(\langle \mathbf{B}, \mathbf{X}\rangle \leq x) \mid Y\}] \\ &\quad + E[\operatorname{var}\{I(\langle \mathbf{B}, \mathbf{X}\rangle \leq x) \mid Y\}] \\ &\geq \operatorname{var}[E\{I(\langle \mathbf{B}, \mathbf{X}\rangle \leq x) \mid Y\}] \geq 0.\end{aligned}$$

Independence $\Rightarrow \mathrm{T}_{\text{weight}}(\mathbf{X} \mid Y) = 0$ and complete dependence $\Rightarrow \mathrm{T}_{\text{weight}}(\mathbf{X} \mid Y) = 1$ follow directly from the form of $\mathrm{T}_{\text{weight}}(\mathbf{X} \mid Y)$. The converse of the latter can be derived from non-negativity and identity of indiscernibles of $d_{\text{weight}}(\mathbf{X}_1, \mathbf{X}_2)$. For the former, we know that

$$\begin{aligned}\mathrm{T}_{\text{weight}}(\mathbf{X} \mid Y) = 0 \quad \Leftrightarrow \quad & \int_x \int_{\mathbf{B}} \operatorname{var}[E\{I(\langle \mathbf{B}, \mathbf{X}\rangle \leq x) \mid Y\}] \\ & \qquad \omega_1(x)\omega_2(\mathbf{B})(d\mathbf{B})(dx) = 0.\end{aligned}$$

Denote

$$\begin{aligned}Q_1(\mathbf{B}) &= \int_x \operatorname{var}[E\{I(\langle \mathbf{B}, \mathbf{X}\rangle \leq x) \mid Y\}]\, \omega_1(x)(dx), \\ Q_2(x, \mathbf{B}) &= \operatorname{var}[E\{I(\langle \mathbf{B}, \mathbf{X}\rangle \leq x) \mid Y\}].\end{aligned}$$

Then there exists $\mathcal{D} \subseteq \mathbb{R}^{p\times q}$ with $\omega_2(\mathcal{D}^c) = 0$, $Q_1(\mathbf{B}) = 0$ for any $\mathbf{B} \in \mathcal{D}$. Given $\mathbf{B} \in \mathcal{D}$, there exists $\mathcal{T} \subseteq \mathbb{R}$ with $\omega_1(\mathcal{T}^c) = 0$, $Q_2(x, \mathbf{B}) = 0$

for any $x \in \mathcal{T}$. Since $\omega_1(\mathcal{T}^c) = 0$, $\mathcal{T}$ is a dense subset of $\mathbb{R}$ and has itself a countable dense subset, denoted as $\mathcal{Q}$. Thus the countability of $\mathcal{Q}$ implies there exists a common set $\mathcal{Y} \subseteq \text{supp}(F_Y)$ with $F_Y(\mathcal{Y}^c) = 0$ s.t. $F_{\langle \mathbf{B},\mathbf{X}\rangle | Y=y}(x) = F_{\langle \mathbf{B},\mathbf{X}\rangle}(x)$ for any $x \in \mathcal{Q}$ and any $y \in \mathcal{Y}$. According to Resnick (2019, Lemma 8.1.1) that a probability is determined on a dense set (since $\mathcal{Q}$ is dense in $\mathbb{R}$), we conclude that $\langle \mathbf{B}, \mathbf{X}\rangle$ and Y are independent for any $\mathbf{B} \in \mathcal{D}$. Using the continuity of charateristic function, we can deduce that $\langle \mathbf{B}, \mathbf{X}\rangle$ and Y are independent for any $\mathbf{B} \in \mathbb{R}^{p\times q}$, thus $\mathbf{X}$ and Y are independent.

Noting that when $\omega_1(x)$ and $\omega_2(\mathbf{B})$ are standard normal densities,

$$\begin{aligned} d_{\text{weight}}(\mathbf{X}_1, \mathbf{X}_2) &= \text{pr}\left(x - \langle \mathbf{B}, \mathbf{X}_1\rangle \geq 0\right) + \text{pr}\left(x - \langle \mathbf{B}, \mathbf{X}_2\rangle \geq 0\right) \\ &\quad -2\text{pr}\left(x - \langle \mathbf{B}, \mathbf{X}_1\rangle \geq 0, x - \langle \mathbf{B}, \mathbf{X}_2\rangle \geq 0\right) \\ &= \pi^{-1} d_{\text{normal}}(\mathbf{X}_1, \mathbf{X}_2), \end{aligned}$$

where the last equality follows from Lemma 1 because $(x - \langle \mathbf{B}, \mathbf{X}_1\rangle)$ and $(x - \langle \mathbf{B}, \mathbf{X}_2\rangle)$ are bivariate normal with mean zero and correlation

$$\rho = (1 + \langle \mathbf{X}_1, \mathbf{X}_2\rangle)\left(1 + \|\mathbf{X}_1\|^2\right)^{-1/2}\left(1 + \|\mathbf{X}_2\|^2\right)^{-1/2}.$$

□

Proof of Proposition 2: When $K(\mathbf{X}_1, \mathbf{X}_2) = \langle \mathbf{\Psi}(\mathbf{X}_1), \mathbf{\Psi}(\mathbf{X}_2)\rangle$ but $\mathbf{\Psi}$ is infinite-dimensional, to formulate $\text{T}\{\mathbf{\Psi}(\mathbf{X}) \mid Y\}$ more rigorously, we follow van Zanten and van der Vaart (2008, Section 2.2) to introduce the notions of Gaussian random element W on the Banach space $(\mathcal{F}, \|\cdot\|)$ and the reproducing kernel Hilbert space $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ attached to W. In our case, $\mathcal{F} = C(\mathcal{X})$, which is the space of all continuous functions from a compact $\mathcal{X} \subset \mathbb{R}^{p\times q}$ to $\mathbb{R}$, equipped with the uniform norm $\|f\|_\infty = \sup_{\mathbf{X}\in\mathcal{X}} |f(\mathbf{X})|$. Since in this case for every kernel there exists a Gaussian process whose covariance function equals the kernel, we can equivalently define the trace correlation in the reproducing kernel Hilbert space as

$$\begin{aligned}\mathrm{T}\{\mathbf{\Psi}(\mathbf{X}) \mid Y\} \;&=\; E_{W,x}\left(\mathrm{var}\left[E\left\{I(\langle W, \Psi(\mathbf{X})\rangle_{\mathcal{H}} \leq x) \mid Y\right\}\right]\right) \Big/ \\ &\qquad E_{W,x}\left[\mathrm{var}\left\{I(\langle W, \Psi(\mathbf{X})\rangle_{\mathcal{H}} \leq x)\right\}\right],\end{aligned}$$

where W and x have respectively Gaussian distribution on $C(\mathcal{X})$ and standard Gaussian distribution on $\mathbb{R}$, and $K(\mathbf{X}_1, \mathbf{X}_2)$ and $\Psi(\mathbf{X}) = K(\cdot, \mathbf{X})$ are respectively the reproducing kernel and canonical feature map of $\mathcal{H}$ attached to W.

Therefore, the only nontrivial part of Proposition 2 is $\mathrm{T}\{\mathbf{\Psi}(\mathbf{X}) \mid Y\} = 0$ implies independence: define

$$Q(f) = \int_x \mathrm{var}\left(E\left[I\{f(\mathbf{X}) \leq x\} \mid Y\right]\right)\ \omega_1(x)(dx).$$

$\mathrm{T}\{\mathbf{\Psi}(\mathbf{X}) \mid Y\} = 0$ implies $Q(g) = 0$ for any $g \in \mathcal{E}$, where $P\{C(\mathcal{X}) \setminus \mathcal{E}\} = 0$. For any continuous function $f \in C(\mathcal{X})$, by the universality of K, $\overline{\mathcal{H}} = C(\mathcal{X})$, where $\overline{\mathcal{H}}$ denotes the closure of $\mathcal{H}$ in $C(\mathcal{X})$ w.r.t. $\|\cdot\|_\infty$. Therefore, either $f \in \mathcal{E}$, then $Q(f) = 0$, or $f \in C(\mathcal{X}) \setminus \mathcal{E}$. For the latter case, we claim that there exists a sequence of functions $\{f_n\} \in \mathcal{E}$ converging in $\|\cdot\|_\infty$ to f. Otherwise, there exists some $\delta > 0$, such that $\{h \in C(\mathcal{X}) : \|h - f\|_\infty < \delta\} \subset C(\mathcal{X}) \setminus \mathcal{E}$. However, the former set has positive probability according to van Zanten and van der Vaart (2008, Lemma 5.1), then $P\{C(\mathcal{X}) \setminus \mathcal{E}\} = 0$ will be contradicted. More to the point, $f_n \to f$ pointwisely, then $Q(f) = 0$ for $f \in C(\mathcal{X}) \setminus \mathcal{E}$ due to Fubini's lemma and dominated convergence theorem. Then $\mathbf{X}$ and Y are independent according to Lemma 2. The equivalent form of $\mathrm{T}\{\boldsymbol{\Psi}(\mathbf{X}) \mid Y\}$ can be derived similar to $\mathrm{T}(\mathbf{X} \mid Y)$, by applying Lemma 1 and the reproducing kernel formula (Da Prato and Zabczyk, 2014, Page 41):

$$\int_{\mathcal{F}} \langle h, x\rangle_{\mathcal{H}} \langle g, x\rangle_{\mathcal{H}} \, P(\mathrm{d}x) = \langle h, g\rangle_{\mathcal{H}}, \text{ for } h, g \in \mathcal{H}.$$

□

S2 Proofs of Theorems 1 and 2

Proof of Theorem 1:

(1) $\widehat{\mathrm{T}}(\mathbf{X} \mid Y) \xrightarrow{p} \mathrm{T}(\mathbf{X} \mid Y)$**.**

Recall that $T_2 = \sum_{h=1}^{H} \widetilde{d}(h) p_h$, and $\widehat{T}_2 = \sum_{h=1}^{H} \widehat{d}(h) \widehat{p}_h$ with $\widehat{p}_h = n_h/n$, and the quantity of interest can be written as

$$\mathrm{T}(\mathbf{X} \mid Y) - \widehat{\mathrm{T}}(\mathbf{X} \mid Y) = \left\{ \left(\widehat{T}_2 - T_2 \right) T_1 - \left(\widehat{T}_1 - T_1 \right) T_2 \right\} \Big/ \left(\widehat{T}_1 \times T_1 \right).$$

We have $\widehat{T}_1 - T_1 = O_p(n^{-1/2})$ from standard U-statistic theory (Serfling, 2009, Theorem 5.5.1A). As for

$$\widehat{T}_2 - T_2 = \sum_{h=1}^{H} \left\{ \widehat{d}(h) - \widetilde{d}(h) \right\} \widehat{p}_h + \sum_{h=1}^{H} \widetilde{d}(h) \left(\widehat{p}_h - p_h \right) = H_1 + H_2,$$

we have

$$E(H_2) = 0, \quad \mathrm{var}(H_2) = n^{-1} \mathrm{var}\left\{ \widetilde{d}(Y) \right\} = O(n^{-1}),$$

and for some constant $C > 0$,

$$\begin{aligned} & E(H_1) = E\{E(H_1 \mid \mathcal{F}_n)\} = 0, \quad \mathrm{var}(H_1) = E\{\mathrm{var}(H_1 \mid \mathcal{F}_n)\} \\ & \leq C \cdot E\Big[\sum_{h=1}^{H} \frac{n_h}{n^2} \mathrm{var}\{d(\mathbf{X}_i, \mathbf{X}_j) I(Y_i = h) I(Y_j = h) \mid \mathcal{F}_n\} \Big] = O(n^{-1}), \end{aligned}$$

where $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$. Therefore, $\widehat{T}_2 - T_2 = O_p(n^{-1/2})$ by Chebyshev's inequality. We thus conclude that $\widehat{\mathrm{T}}(\mathbf{X} \mid Y) \xrightarrow{p} \mathrm{T}(\mathbf{X} \mid Y)$.

(2) The Asymptotic Distributions.

Case (i) Assume X is independent of Y.

When H is fixed. Let $d_U(\mathbf{X}_i, \mathbf{X}_j) = d(\mathbf{X}_i, \mathbf{X}_j) - d_1(\mathbf{X}_i) - d_1(\mathbf{X}_j) + T_1$, then

$$\begin{aligned}\widehat{T}_1-\widehat{T}_2 &= \{n(n-1)\}^{-1}\sum_{i\neq j}^{n} d(\mathbf{X}_i,\mathbf{X}_j) \\ &\quad -\sum_{h=1}^{H}\frac{n-1}{n_h-1}\{n(n-1)\}^{-1}\sum_{i\neq j}^{n} \\ &\qquad d(\mathbf{X}_i,\mathbf{X}_j)I(Y_i=h)I(Y_j=h) \\ &= \{n(n-1)\}^{-1}\sum_{i\neq j}^{n} d_U(\mathbf{X}_i,\mathbf{X}_j) \\ &\quad -\sum_{h=1}^{H}\frac{n-1}{n_h-1}\{n(n-1)\}^{-1}\sum_{i\neq j}^{n} \\ &\qquad d_U(\mathbf{X}_i,\mathbf{X}_j)I(Y_i=h)I(Y_j=h),\end{aligned}$$

which we denote as $U_n^{(0)}-\sum_{h=1}^{H}\frac{n-1}{n_h-1}U_n^{(h)}$. Thus it is sufficient to show that

$$n\left(U_n^{(0)}-\sum_{h=1}^{H}\frac{n-1}{n_h-1}U_n^{(h)}\right)\xrightarrow{d}(H-1)T_1(Q-1),$$

where $Q=\sum_{i=1}^{\infty}\lambda_i Z_i^2$, $Z_i\overset{i.i.d}{\sim}\mathcal{N}(0,1)$ and λ_i are positive constants with $\sum_{i=1}^{\infty}\lambda_i=1$. The proof should be similar to that of Ke and Yin (2020, Theorem 7). Therefore, we skip these details.

When H is divergent. Define the projection of $\widehat{T}_1$ as $\widetilde{T}_1$, such that $\widetilde{T}_1-T_1=(n/2)^{-1}\sum_{i=1}^{n}\widetilde{d}_1(\mathbf{X}_i)$, where $\widetilde{d}_1(\mathbf{X}_i)=d_1(\mathbf{X}_i)-T_1$. From Serfling

(2009, Theorem 5.3.2), $\widehat{T}_1 - \widetilde{T}_1 = O_p(n^{-1})$, we have that

$$\begin{aligned}\widehat{T}_1 - T_1 &= \frac{2}{n}\sum_{i=1}^{n} \widetilde{d}_1(\mathbf{X}_i) + O_p(n^{-1}) \\ &= \sum_{h=1}^{H} \widehat{p}_h \{n_h(n_h-1)\}^{-1} \sum_{1\le i\ne j\le n_h} \\ &\qquad \Big\{\widetilde{d}_1(\mathbf{X}_{(h,i)}) + \widetilde{d}_1(\mathbf{X}_{(h,j)})\Big\} + O_p(n^{-1}).\end{aligned}$$

To apply Lemma 3, we denote $(\widehat{T}_2 - T_2)T_1 - (\widetilde{T}_1 - T_1)T_2 = \sum_{h=1}^{H} G_h$, and let $s_n = n^{-1/2}$ with $c_n^{-1} = \sum_{h=1}^{H} n_h/\{n(n_h-1)\}$, $m_n = H$, $\mathcal{F}_n = \sigma(Y_1,\ldots,Y_n)$, and $X_{n,h} = c_n^{1/2} G_h$. Assumption (C1) holds by definition. Assumption (C4) holds trivially since $E\big(X_{n,h} \mid \mathcal{F}_n\big) = (\widehat{p}_h - p_h)\widetilde{d}(h)T_1$, and we have $s_n^{-1}\sum_{h=1}^{m_n} E\left(X_{n,h} \mid \mathcal{F}_n\right) \equiv 0$ by independence. Next, denote $\widetilde{X}_{n,h} = X_{n,h} - E\left(X_{n,h} \mid \mathcal{F}_n\right)$, then

$$\begin{aligned}\widetilde{X}_{n,h} &= c_n^{1/2}\{n_h(n_h-1)\}^{-1} T_1 \sum_{1\le i\ne j\le n_h} \widehat{p}_h \\ &\qquad \big\{d(\mathbf{X}_{(h,i)}, \mathbf{X}_{(h,j)}) - d_1(\mathbf{X}_{(h,i)}) - d_1(\mathbf{X}_{(h,j)}) + T_1\big\}.\end{aligned}$$

Consider Assumption (C2), denote $\sigma^2 = \mathrm{var}\{d(\mathbf{X}_i,\mathbf{X}_j) - d_1(\mathbf{X}_i) - d_1(\mathbf{X}_j)\}$. We find that

$$E\big(\widetilde{X}_{n,h}^2 \mid \mathcal{F}_n\big) = 2c_n\big[n_h/\{n^2(n_h-1)\}\big]\sigma^2 T_1^2,$$

then the left hand side of Assumption (C2) converges in probability to $2\sigma^2 T_1^2$. As for Assumption (C3), we note that by Serfling (2009, Lemma 5.2.2.B), there exits some constant $C > 0$, such that $E\big(\widetilde{X}_{n,h}^4 \mid \mathcal{F}_n\big) \le$

$Cn^{-4}c_n^2$. Therefore, by Cauchy-Schwarz inequality and Chebyshev's inequality,

$$
\begin{aligned}
& s_n^{-2}\sum_{h=1}^{H} E\{\widetilde{X}_{n,h}^2 I(|\widetilde{X}_{n,h}| > \epsilon s_n) \mid \mathcal{F}_n\} \\
\leq\ & s_n^{-2}\sum_{h=1}^{H}\sqrt{E(\widetilde{X}_{n,h}^4 \mid \mathcal{F}_n)\mathrm{pr}(|\widetilde{X}_{n,h}| > \epsilon s_n \mid \mathcal{F}_n)} \\
\leq\ & \sqrt{2C\sigma^2 T_1^2}(nc_n)^{3/2}\sum_{h=1}^{H}\sqrt{n_h/\{n(n_h-1)\}}n^{-5/2}/\epsilon = O(H^{-1/2}) \to 0
\end{aligned}
$$

since H is divergent. By Lemma 3 and Slutsky's lemma, $(nc_n)^{1/2}\widehat{\mathrm{T}}(\mathbf{X} \mid Y) \xrightarrow{d} \mathcal{N}(0, 2\sigma^2/T_1^2)$.

Case (ii) Assume X is dependent but not completely dependent upon Y.

When H is fixed. We have

$$
\begin{aligned}
& (\widehat{T}_1 - T_1)T_2 - (\widehat{T}_2 - T_2)T_1 \\
=\ & T_2(n/2)^{-1}\sum_{i=1}^{n}\widetilde{d}_1(\mathbf{X}_i) - T_1\sum_{h=1}^{H}\Big[p_h\frac{n-1}{n_h-1}\frac{1}{n(n-1)} \\
& \qquad \sum_{i\neq j}^{n} d(\mathbf{X}_i,\mathbf{X}_j)I(Y_i=h)I(Y_j=h)/p_h - \widetilde{d}(h)p_h\Big] \\
=\ & T_2(n/2)^{-1}\sum_{i=1}^{n}\widetilde{d}_1(\mathbf{X}_i) - T_1\sum_{h=1}^{H}\big[p_h\frac{n-1}{n_h-1}\frac{2}{n}\sum_{i=1}^{n} \\
& \{d_2(\mathbf{X}_i,h)I(Y_i=h) - \widetilde{d}(h)p_h\} + (p_h\frac{n-1}{n_h-1} - 1)\widetilde{d}(h)p_h\big] + O_p(n^{-1}) \\
=\ & n^{-1}\sum_{i=1}^{n}\big[2T_2\widetilde{d}_1(\mathbf{X}_i) - T_1\{2d_2(\mathbf{X}_i,Y_i) - \widetilde{d}(Y_i) - T_2\}\big] + o_p(n^{-1/2}),
\end{aligned}
$$

where $\widetilde{d}_1(\mathbf{X}_i) = d_1(\mathbf{X}_i) - T_1$, $d_2(\mathbf{X}_i, h) = E\{d(\mathbf{X}_i,\mathbf{X}_j) \mid \mathbf{X}_i, Y_i = Y_j =$

$h\}$, the second equality follows from Serfling (2009, Theorem 5.3.2), and the third equality follows from the delta method. Therefore, by Slutsky's lemma, $n^{1/2}\{\widehat{\mathrm{T}}(\mathbf{X} \mid Y) - \mathrm{T}(\mathbf{X} \mid Y)\} \xrightarrow{d} \mathcal{N}(0, \tau_*^2/T_1^2)$, with $\tau_*^2 = \mathrm{var}[2\{1 - \mathrm{T}(\mathbf{X} \mid Y)\}d_1(\mathbf{X}) - 2d_2(\mathbf{X}, Y) + \widetilde{d}(Y)]$.

When H is divergent. Recall that $\tau_1 = \mathrm{var}\Big[\widetilde{d}(Y) - 2\{1 - \mathrm{T}(\mathbf{X} \mid Y)\}d_1(\mathbf{X})\Big]$, $\tau_2 = E\{V_2(Y)\} - 2\{1 - \mathrm{T}(\mathbf{X} \mid Y)\}E\{V_1(Y)\}$ and $\tau^2 = \tau_1 + 4\tau_2$. Following the notations of Case (i) when H is divergent, we now let $X_{n,h} = G_h/\tau$ instead, then:

$$
\begin{aligned}
\sum_{h=1}^{m_n} E\left(X_{n,h} \mid \mathcal{F}_n\right) &= T_1 \sum_{h=1}^{H} (\widehat{p}_h - p_h)\widetilde{d}(h)\Big/\tau \\
&\qquad - 2T_2 n^{-1} \sum_{i=1}^{n} E\left\{\widetilde{d}_1(\mathbf{X}_i) \mid Y_i\right\}\Big/\tau \\
&= n^{-1} \sum_{i=1}^{n} \Big[T_1\{\widetilde{d}(Y_i) - T_2\} - 2T_2 E\left\{\widetilde{d}_1(\mathbf{X}_i) \mid Y_i\right\}\Big]\Big/\tau.
\end{aligned}
$$

Define $X_{n,i} = \big[T_1\{\widetilde{d}(Y_i) - T_2\} - 2T_2 E\{\widetilde{d}_1(\mathbf{X}_i) \mid Y_i\}\big]\big/\big(\tau n^{1/2}\big)$. According to Lindeberg-Feller CLT for triangular arrays (Resnick, 2019, Exercise 9.9.1), since $E(X_{n,i}) = 0$,

$$
\begin{aligned}
\sum_{i=1}^{n} E\big(X_{n,i}^2\big) &= \mathrm{var}\big[T_1\widetilde{d}(Y) - 2T_2 E\{d_1(\mathbf{X}) \mid Y\}\big]/\tau^2 \\
&\to \lim_{H\to\infty} \mathrm{var}\big[T_1\widetilde{d}(Y) - 2T_2 E\{d_1(\mathbf{X}) \mid Y\}\big]/\tau^2,
\end{aligned}
$$

and by Cauchy-Schwarz inequality and Chebyshev's inequality,

$$\begin{aligned}\sum_{i=1}^{n} E\big\{X_{n,i}^2 I(|X_{n,i}|>\epsilon)\big\} &\le \sum_{i=1}^{n} E^{1/2}\big(X_{n,i}^4\big)\mathrm{pr}^{1/2}\big(|X_{n,i}|>\epsilon\big)\\ &\le Cn/\big\{\big(\tau n^{1/2}\big)^3\epsilon\big\}\to 0,\end{aligned}$$

as $n, H\to\infty$, the left hand side of Assumption (C4) converges in distribution to a non-degenerate normal distribution:

$$s_n^{-1}\sum_{h=1}^{m_n} E\left(X_{n,h}\mid \mathcal{F}_n\right)\xrightarrow{d}\mathcal{N}\Big(0,\lim_{H\to\infty}\mathrm{var}\big[T_1\widetilde{d}(Y)-2T_2E\{d_1(\mathbf{X})\mid Y\}\big]\big/\tau^2\Big).$$

For the left hand side of Assumption (C2),

$$\begin{aligned}\sum_{h=1}^{H} E\left(\widetilde{X}_{n,h}^2\mid\mathcal{F}_n\right) &\cong n^{-1}\sum_{h=1}^{H}\Big(T_1^2\Big[\{n(n_h-1)/(2n_h)\}^{-1}V_0(h)\\ &\quad +4\frac{n_h(n_h-2)}{n(n_h-1)}V_2(h)\Big]-8T_1T_2\left\{(n_h/n)V_1(h)\right\}\\ &\quad +4T_2^2 n^{-1}\sum_{i=1}^{n_h}\mathrm{var}\{d_1(\mathbf{X}_{(h,i)})\mid Y_{(h,i)}\}\Big)\Big/\tau^2,\end{aligned}$$

where $V_0(h)=\mathrm{var}(\varepsilon_{i,j,h}\mid Y_i=Y_j=h)$. We remark that for $H=o(n)$,

$$\begin{aligned}\sum_{h=1}^{H}\frac{n_h}{n(n_h-1)}V_2(h) &\le 2\sum_{h=1}^{H} n^{-1}V_2(h)\\ &= O\big(H/n\big)=o(1),\ \text{and}\\ \sum_{h=1}^{H}\{(n_h/n)-p_h\}V_2(h) &= n^{-1}\sum_{i=1}^{n}[V_2(Y_i)-E\{V_2(Y)\}]\\ &= O_p\big(n^{-1/2}\big)=o_p(1),\end{aligned}$$

where the second argument follows from Chebyshev's inequality. Then

$$\sum_{h=1}^{H}\frac{n_h(n_h-2)}{n(n_h-1)}V_2(h)\overset{p}{\sim}\sum_{h=1}^{H}V_2(h)p_h=E\{V_2(Y)\}\to\lim_{H\to\infty}E\{V_2(Y)\}.$$

Following similar arguments, we can derive that

$$
\begin{aligned}
\sum_{h=1}^{H}(n_h/n)V_1(h) &\overset{p}{\sim} \sum_{h=1}^{H} V_1(h)p_h \\
&= E\{V_1(Y)\} \to \lim_{H\to\infty} E\{V_1(Y)\}, \\
\sum_{h=1}^{H} n^{-1}\sum_{i=1}^{n}\operatorname{var}\{d_1(\mathbf{X}_i) \mid Y_i\}I(Y_i = h) &\overset{p}{\sim} \sum_{h=1}^{H}\operatorname{var}\{d_1(\mathbf{X}) \mid Y = h\}p_h \\
&= E[\operatorname{var}\{d_1(\mathbf{X}) \mid Y\}] \\
&\to \lim_{H\to\infty} E[\operatorname{var}\{d_1(\mathbf{X}) \mid Y\}].
\end{aligned}
$$

Moreover, we have

$$
\sum_{h=1}^{H}\{n(n_h-1)/(2n_h)\}^{-1}V_0(h) \;\le\; 4\sum_{h=1}^{H} V_0(h)/n = O\big(H/n\big) \to 0,
$$

whenever $H = o(n)$. As for Assumption (C3), we note that by Serfling (2009, Lemma 5.2.2.A), given any $r \ge 2$, there exists some constant $C > 0$, such that $E\Big(\big|\widetilde{X}_{n,h}\big|^r \mid \mathcal{F}_n\Big) \le C n_h^{r/2}/\big(\tau^2 n\big)^r$. Therefore, by Markov's inequality,

$$
\begin{aligned}
& s_n^{-2}\sum_{h=1}^{H} E\big\{\widetilde{X}_{n,h}^2 I\big(\big|\widetilde{X}_{n,h}\big| > \epsilon s_n\big) \mid \mathcal{F}_n\big\} \\
\le\;& s_n^{-2}\sum_{h=1}^{H} E^{1/2}\big(\widetilde{X}_{n,h}^4 \mid \mathcal{F}_n\big)\operatorname{pr}^{1/2}\big(\big|\widetilde{X}_{n,h}\big| > \epsilon s_n \mid \mathcal{F}_n\big) \\
\le\;& C'\sum_{h=1}^{H}(n_h/n)^{\lfloor r/4+1\rfloor}/\big(\tau^{r+4}\epsilon^{r/2}\big).
\end{aligned}
$$

Note that there exists some sufficiently large r and constant $C > 0$, such that

$$
\begin{aligned}
E\Big\{\sum_{h=1}^{H}(n_h/n)^{\lfloor r/4+1\rfloor}\Big\} &\le C\Big(H/n+\sum_{h=1}^{H}p_h^{\lfloor r/4+1\rfloor}\Big)\\
&\le C\Big(H/n+\sum_{h=1}^{H}n^{-\lfloor r/4+1\rfloor\alpha}\Big)\to 0
\end{aligned}
$$

by assumption. This implies

$$
\sum_{h=1}^{H}(n_h/n)^{\lfloor r/4+1\rfloor}\Big/\Big(\tau^{r+4}\epsilon^{r/2}\Big)=o_p(1).
$$

Therefore, by Slutsky's lemma and Lemma 3, $n^{1/2}\{\widehat{\mathrm{T}}(\mathbf{X}\mid Y)-\mathrm{T}(\mathbf{X}\mid Y)\}(T_1/\tau)\xrightarrow{d}\mathcal{N}(0,1)$.

The relationship between τ^2 and τ_*^2.

$$
\begin{aligned}
\tau_*^2 &= \mathrm{var}[2\{1-\mathrm{T}(\mathbf{X}\mid Y)\}d_1(\mathbf{X})-2d_2(\mathbf{X},Y)+\widetilde{d}(Y)]\\
&= \mathrm{var}[\widetilde{d}(Y)-2\{1-\mathrm{T}(\mathbf{X}\mid Y)\}d_1(\mathbf{X})]\\
&+ 8\{1-\mathrm{T}(\mathbf{X}\mid Y)\}\mathrm{cov}\{d_1(\mathbf{X}),\widetilde{d}(Y)\}\\
&- 8\{1-\mathrm{T}(\mathbf{X}\mid Y)\}[E\{V_1(Y)\}+\mathrm{cov}\{d_1(\mathbf{X}),\widetilde{d}(Y)\}]\\
&+ 4[E\{V_2(Y)\}+\mathrm{var}\{\widetilde{d}(Y)\}]-4\mathrm{var}\{\widetilde{d}(Y)\}=\tau^2.
\end{aligned}
$$

Case (iii) Assume $\mathbf{X}$ is completely dependent upon Y.

When $\mathbf{X}$ is completely dependent upon Y, there exists a matrix of functions

$\mathbf{G} \in \mathbb{R}^{p\times q}$ such that $\mathrm{pr}\{\mathbf{X} = \mathbf{G}(Y)\} = 1$. Therefore, with probability 1,

$$\begin{aligned}\widehat{T}_2 &= \sum_{h=1}^{H} \{n(n_h-1)\}^{-1} \sum_{1\le i\ne j\le n_h} d(\mathbf{X}_{(h,i)}, \mathbf{X}_{(h,j)}) \\ &= \sum_{h=1}^{H} \{n(n_h-1)\}^{-1} \sum_{1\le i\ne j\le n_h} d\big(\mathbf{G}(h), \mathbf{G}(h)\big) = 0,\end{aligned}$$

which implies $\mathrm{pr}\{\widehat{\mathrm{T}}(\mathbf{X} \mid Y) = 1\} = 1$.

(3) The Asymptotic Null Variance of $(nc_n/2)^{1/2}\big(\widehat{T}_1 - \widehat{T}_2\big)/\sigma$.

Recall that $d_U(\mathbf{X}_i, \mathbf{X}_j) = d(\mathbf{X}_i, \mathbf{X}_j) - d_1(\mathbf{X}_i) - d_1(\mathbf{X}_j) + T_1$, $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$,

and

$$\begin{aligned}\widehat{T}_1 - \widehat{T}_2 &= \{n(n-1)\}^{-1} \sum_{i\ne j}^{n} d_U(\mathbf{X}_i, \mathbf{X}_j) \\ &\quad - \sum_{h=1}^{H} \frac{n-1}{n_h-1} \{n(n-1)\}^{-1} \sum_{i\ne j}^{n} \\ &\qquad d_U(\mathbf{X}_i, \mathbf{X}_j) I(Y_i = h) I(Y_j = h) \\ &= U_n^{(0)} - \sum_{h=1}^{H} \frac{n-1}{n_h-1} U_n^{(h)}.\end{aligned}$$

It's easy to check that

$$E\left(\widehat{T}_1 - \widehat{T}_2 \mid \mathcal{F}_n\right) = E\left(U_n^{(0)} \mid \mathcal{F}_n\right) - \sum_{h=1}^{H} \frac{n-1}{n_h-1} E\left(U_n^{(h)} \mid \mathcal{F}_n\right) = 0.$$

Moreover, $\mathrm{var}\big(U_n^{(0)} \mid \mathcal{F}_n\big) = 2\sigma^2/\{n(n-1)\}$, for $h_1 \ne h_2$, $\mathrm{cov}\big(U_n^{(h_1)}, U_n^{(h_2)} \mid \mathcal{F}_n\big) = 0$, and for $h = 1, \ldots, H$,

$$\mathrm{cov}\big(U_n^{(0)}, U_n^{(h)} \mid \mathcal{F}_n\big) = \frac{2n_h(n_h-1)\sigma^2}{\{n(n-1)\}^2},\ \mathrm{var}\big(U_n^{(h)} \mid \mathcal{F}_n\big) = \frac{2n_h(n_h-1)\sigma^2}{\{n(n-1)\}^2}.$$

Therefore,

$$\begin{aligned} \operatorname{var}\left(\widehat{T}_1 - \widehat{T}_2 \mid \mathcal{F}_n\right) &= \sum_{h=1}^{H} \frac{(n-1)^2}{(n_h-1)^2} \frac{2n_h(n_h-1)\sigma^2}{\{n(n-1)\}^2} - \frac{2\sigma^2}{n(n-1)} \\ &= 2n^{-1}\{c_n^{-1} - (n-1)^{-1}\}\sigma^2, \end{aligned}$$

that is,

$$\operatorname{var}\left\{(nc_n)^{1/2}\big(\widehat{T}_1 - \widehat{T}_2\big) \mid \mathcal{F}_n\right\} = 2\{1 - c_n/(n-1)\}\sigma^2.$$

By dominated convergence theorem, we have $\operatorname{var}\big\{(nc_n)^{1/2}\big(\widehat{T}_1 - \widehat{T}_2\big)\big\} \to 2\sigma^2$ if H is divergent, and $\operatorname{var}\big\{(nc_n)^{1/2}\big(\widehat{T}_1 - \widehat{T}_2\big)\big\} \to 2(1-H^{-1})\sigma^2$, $c_n \to n/H$ if H is fixed. □

Proof of Theorem 2: Following the same paradigm as the proof of Theorem 1, we now decompose $\widehat{T}_2 - T_2$ into three parts:

$$\begin{aligned} \widehat{T}_2 - T_2 &= \sum_{h=1}^{H}\{n(n_h-1)\}^{-1} \sum_{1\le i\neq j\le n_h} \big\{d(\mathbf{X}_{(h,i)}, \mathbf{X}_{(h,j)}) - m(Y_{(h,i)}, Y_{(h,j)})\big\} \\ &+ \sum_{h=1}^{H}\{n(n_h-1)\}^{-1} \sum_{1\le i\neq j\le n_h} \\ &\qquad \big[\big\{m(Y_{(h,i)}, Y_{(h,i)}) + m(Y_{(h,j)}, Y_{(h,j)})\big\}/2 - T_2\big] \\ &+ \sum_{h=1}^{H}\{n(n_h-1)\}^{-1} \sum_{1\le i\neq j\le n_h} \big[m(Y_{(h,i)}, Y_{(h,j)}) \\ &\qquad - \big\{m(Y_{(h,i)}, Y_{(h,i)}) + m(Y_{(h,j)}, Y_{(h,j)})\big\}/2\big] \\ &= D_1 + D_2 + D_3. \end{aligned}$$

We have $D_2 = n^{-1}\sum_{i=1}^{n}\big\{m(Y_i, Y_i) - T_2\big\} = O_p\big(n^{-1/2}\big)$ from classical CLT, and $D_1 = O_p\big(n^{-1/2}\big)$ from Chebyshev's inequality. If $\mathbf{X}$ is independent of

Y, $D_3 = 0$, and if $\mathbf{X}$ is dependent upon Y, $D_3 = o_p\big(n^{-1/2}\big)$ from Lemma 5 under Condition (A1) and (A3). In addition, $\widehat{T}_1 - T_1 = O_p\big(n^{-1/2}\big)$ from standard U-statistic theory (Serfling, 2009, Theorem 5.5.1A). Therefore, we conclude that $\widehat{\mathrm{T}}(\mathbf{X} \mid Y) - \mathrm{T}(\mathbf{X} \mid Y) = O_p\big(n^{-1/2}\big)$.

Next we show the asymptotic normality:

Case (i) Assume X is independent of Y. Similarly, we consider the projection $\widetilde{T}_1$ of U-statistic T_1, where $\widetilde{T}_1 - T_1 = (n/2)^{-1}\sum_{i=1}^{n}\widetilde{d}_1(\mathbf{X}_i)$, and we have

$$\begin{aligned}\widehat{T}_1 - T_1 &= (n/2)^{-1}\sum_{i=1}^{n}\widetilde{d}_1(\mathbf{X}_i) + O_p(n^{-1})\\ &= \sum_{h=1}^{H}\frac{n_h}{n}\frac{1}{n_h(n_h-1)}\sum_{1\le i\neq j\le n_h}\Big\{\widetilde{d}_1(\mathbf{X}_{(h,i)}) + \widetilde{d}_1(\mathbf{X}_{(h,j)})\Big\} + O_p(n^{-1}).\end{aligned}$$

Denote $T_1(D_1+D_2) - T_2(\widetilde{T}_1 - T_1) = \sum_{h=1}^{H} G_h$, we will again apply Lemma 3 with $s_n = (nc_n)^{-1/2}$, $m_n = H$, $\mathcal{F}_n = \sigma(Y_1,\ldots,Y_n)$, and $X_{n,h} = G_h$: Assumption (C1) holds by definition. Assumption (C4) can be checked because $\sum_{h=1}^{m_n} E\left(X_{n,h} \mid \mathcal{F}_n\right) \equiv 0$ by independence. Next, denote $\widetilde{X}_{n,h} = X_{n,h} - E\left(X_{n,h} \mid \mathcal{F}_n\right)$. Now consider Assumption (C2): we have

$$\begin{aligned}\widetilde{X}_{n,h} \;=\; & \{n(n_h-1)\}^{-1}\,T_1\sum_{1\le i\neq j\le n_h}\\ & \Big\{d(\mathbf{X}_{(h,i)},\mathbf{X}_{(h,j)}) - d_1(\mathbf{X}_{(h,i)}) - d_1(\mathbf{X}_{(h,j)}) + T_1\Big\}.\end{aligned}$$

Denote $\sigma^2 = \mathrm{var}\{d(\mathbf{X}_i,\mathbf{X}_j) - d_1(\mathbf{X}_i) - d_1(\mathbf{X}_j)\}$, then the left hand side of

Assumption (C2) is $2\sigma^2 T_1^2$. We note again that

$$
\begin{aligned}
& s_n^{-2} \sum_{h=1}^{H} E\big\{\widetilde{X}_{n,h}^2 I\big(\big|\widetilde{X}_{n,h}\big| > \epsilon s_n\big) \mid \mathcal{F}_n\big\} \\
\leq \;& s_n^{-2} \sum_{h=1}^{H} \sqrt{E\big(\widetilde{X}_{n,h}^4 \mid \mathcal{F}_n\big)\mathrm{pr}\big(\big|\widetilde{X}_{n,h}\big| > \epsilon s_n \mid \mathcal{F}_n\big)} \\
\leq \;& \sqrt{2C\sigma^2 T_1^2}(nc_n)^{3/2} \sum_{h=1}^{H} \sqrt{n_h/\{n(n_h-1)\}}\, n^{-5/2}/\epsilon = O\big(H^{-1/2}\big) \to 0,
\end{aligned}
$$

which implies Assumption (C3). Since $\mathrm{T}(\mathbf{X} \mid Y) = 0$, from Lemma 3 and Slutsky's lemma, $(nc_n)^{1/2}\widehat{\mathrm{T}}(\mathbf{X} \mid Y) \xrightarrow{d} \mathcal{N}(0, 2\sigma^2/T_1^2)$.

Case (ii) Assume X is dependent but not completely dependent upon Y. We then apply Lemma 3 with $s_n = n^{-1/2}$, $m_n = H$, $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$, and $X_{n,h} = G_h$ instead: Assumption (C4) can be similarly checked with

$$
\begin{aligned}
s_n^{-1} \sum_{h=1}^{m_n} E\left(X_{n,h} \mid \mathcal{F}_n\right) \;\; \xrightarrow{d} \;\; & \mathcal{N}(0, T_1^2 \mathrm{var}[m(Y,Y) \\
& -2\{1 - \mathrm{T}(\mathbf{X} \mid Y)\}E\{d_1(\mathbf{X}) \mid Y\}]).
\end{aligned}
$$

The left hand side of Assumption (C2) under Condition (A2)-(A3) and

Lemma 5 is

$$
\begin{aligned}
& n\sum_{h=1}^{H}\{n(n_h-1)\}^{-2}\Big[T_1^2\{2\sum_{1\le i\neq j\le n_h}V_0(Y_{(h,i)},Y_{(h,j)}) \\
& +4\sum_{[i,j,k]}^{n_h}V_4(Y_{(h,i)},Y_{(h,j)},Y_{(h,k)})\}-8T_1T_2(n_h-1)\sum_{[i,j]}^{n_h}V_3(Y_{(h,i)},Y_{(h,j)}) \\
& +4T_2^2(n_h-1)^2\sum_{i=1}^{n_h}\mathrm{var}\{d_1(\mathbf{X}_{(h,i)})\mid Y_{(h,i)}\}\Big] \\
& \xrightarrow{p} 4T_1^2E\{V_4(Y,Y,Y)\}-8T_1T_2E\{V_3(Y,Y)\}+4T_2^2E[\mathrm{var}\{d_1(\mathbf{X})\mid Y\}],
\end{aligned}
$$

since $H=o(n)$ implies for some $C>0$,

$$
\begin{aligned}
& \sum_{h=1}^{H}n^{-1}(n_h-1)^{-1}\sum_{i=1}^{n_h}V_4(Y_{(h,i)},Y_{(h,i)},Y_{(h,i)}) \\
& \qquad\qquad \le C\sum_{h=1}^{H}\frac{n_h}{n(n_h-1)}=O\big(H/n\big)=o(1), \text{ and} \\
& \sum_{h=1}^{H}\{n(n_h-1)^2\}^{-1}\sum_{1\le i\neq j\le n_h}V_0(Y_{(h,l)},Y_{(h,j)})=O_p(H/n)=o_p(1),
\end{aligned}
$$

where $V_0(Y_i,Y_j)=\mathrm{var}(\varepsilon_{i,j}\mid Y_i,Y_j)$. Denote $\tau_3=\mathrm{var}\Big[m(Y,Y)-2\{1-\mathrm{T}(\mathbf{X}\mid Y)\}d_1(\mathbf{X})\Big]$, $\tau_4=E\{V_4(Y,Y,Y)\}-2\{1-\mathrm{T}(\mathbf{X}\mid Y)\}E\{V_3(Y,Y)\}$ and $\tau_s^2=\tau_3^2+4\tau_4^2$. Assumption (C3) holds since under Condition (A3),

$$
\begin{aligned}
& s_n^{-2}\sum_{h=1}^{H}E\big\{\widetilde{X}_{n,h}^2I\big(\big|\widetilde{X}_{n,h}\big|>\epsilon s_n\big)\mid\mathcal{F}_n\big\} \\
\le\ & s_n^{-2}\sum_{h=1}^{H}E^{1/2}\big(\widetilde{X}_{n,h}^4\mid\mathcal{F}_n\big)\mathrm{pr}^{1/2}\big(\big|\widetilde{X}_{n,h}\big|>\epsilon s_n\mid\mathcal{F}_n\big) \\
\le\ & C\sum_{h=1}^{H}(n_h/n)^{\lfloor r/4+1\rfloor}\big/\epsilon^{r/2}\le C'\sum_{h=1}^{H}n^{-\lfloor r/4+1\rfloor(1-\alpha)}\big/\epsilon^{r/2}\to 0,
\end{aligned}
$$

for some sufficiently large r. Therefore, we will eventually have by Slutsky's lemma and Lemma 3 that $n^{1/2}\{\widehat{\mathrm{T}}(\mathbf{X} \mid Y) - \mathrm{T}(\mathbf{X} \mid Y)\} \xrightarrow{d} \mathcal{N}(0, \tau_s^2/T_1^2)$.

Case (iii) Assume X is completely dependent upon Y. Now $T_2 = 0$, $D_1 = 0$, $D_2 = 0$ and $D_3 = o_p\big(\max_h n_h/n^{1-\gamma}\big)$ under Condition (A3). Therefore, $\widehat{T}_2 - T_2 = o_p\big(\max_h n_h/n^{1-\gamma}\big)$, and $\widehat{\mathrm{T}}(\mathbf{X} \mid Y) - 1 = o_p\big(\max_h n_h/n^{1-\gamma}\big)$.

□

S3 Technical Lemmas

Lemma 1. *(Gupta, 1963, Page 793) Let $(Z_1, Z_2)^\tau$ be bivariate normally distribution with mean zero, and correlation ρ, then*

$$pr(Z_1 \geq 0, Z_2 \geq 0) = 4^{-1} + (2\pi)^{-1} \arcsin(\rho).$$

Lemma 2. *Let $(\mathbf{X}, Y)$ be random variables on $\mathcal{X} \times \mathbb{R}$, then they are independent if and only if*

$$\int_x var\left(E\left[I\{f(\mathbf{X}) \leq x\} \mid Y\right]\right)\ \omega_1(x)(dx) = 0$$

for any bounded, continuous function $f(\cdot)$.

Proof of Lemma 2: The "only if" part is obvious. For the converse, we have $f(\mathbf{X})$ and Y are independent for any bounded, continuous function $f(\cdot)$ following similar arguments in the proof of Proposition 1. Thus

$E\{f(\mathbf{X})g(Y)\} = E\{f(\mathbf{X})\}E\{g(Y)\}$ for each pair (f, g) of bounded, continuous functions, to which we apply Jacod and Protter (2012, Theorem 10.1) to conclude $\mathbf{X}$ and Y are independent. □

Lemma 3. *(Hsing and Carroll, 1992, Theorem A.4) Let $\{s_n\}$ be a sequence of positive constants, $\{X_{n,k}\}$ a triangular array of random variables for $k = 1, \ldots, m_n$ and $n = 1, 2, 3, \ldots,$ and $\mathcal{F}_n$ a sequence of σ-fields. Define $\widetilde{X}_{n,k} = X_{n,k} - E\left(X_{n,k} \mid \mathcal{F}_n\right)$. Finally, assume that*

(C1) $X_{n,1}, \ldots, X_{n,m_n}$ are conditionally independent given $\mathcal{F}_n$.

(C2) $s_n^{-2} \sum_{k=1}^{m_n} E\left(\widetilde{X}_{n,k}^2 \mid \mathcal{F}_n\right) \xrightarrow{p} \sigma^2$.

(C3) for every $c > 0$, $s_n^{-2} \sum_{k=1}^{m_n} E\left\{\widetilde{X}_{n,k}^2 I\left(\left|\widetilde{X}_{n,k}\right| > cs_n\right) \mid \mathcal{F}_n\right\} \xrightarrow{p} 0$.

(C4) $s_n^{-1} \sum_{k=1}^{m_n} E\left(X_{n,k} \mid \mathcal{F}_n\right)$ converges in distribution to some distribution G.

Then the limiting distribution of $s_n^{-1} \sum_{k=1}^{m_n} X_{n,k}$ is the convolution of G and $N\left(0, \sigma^2\right)$.

Lemma 4. *(Hsing and Carroll, 1992, Lemma A.1) Suppose that $Z_1, \ldots, Z_n$ are an i.i.d. sample and r is a positive constant. Let $Z_{(i)}$ be the ith order statistic. Then*

$$n^{-r}\left(\left|Z_{(n)}\right| + \left|Z_{(1)}\right|\right) = o_p(1),$$

if and only if $x^{1/r}P\{|Z|>x\}\to 0$ *as* $x\to\infty$.

Lemma 5. *Under Condition (A1)-(A3), let* $c_1=\max_h n_h$,

$$n^{-\gamma}\sum_{h=1}^{H}\frac{1}{n_h-1}\sum_{1\le i\neq j\le n_h}\big\{m(Y_{(h,i)},Y_{(h,j)})-m(Y_{(h,i)},Y_{(h,i)})\big\}=o_p(c_1),$$

$$n^{-\xi}\sum_{h=1}^{H}\frac{1}{n_h-1}\sum_{1\le i\neq j\le n_h}\big\{V_4(Y_{(h,i)},Y_{(h,j)})-V_4(Y_{(h,i)},Y_{(h,i)})\big\}=o_p(c_1),$$

$$n^{-\xi}\sum_{h=1}^{H}\frac{1}{(n_h-1)^2}\sum_{1\le i\neq j\neq k\le n_h}\big\{V_3(Y_{(h,i)},Y_{(h,j)},Y_{(h,k)})-V_3(Y_{(h,i)},Y_{(h,i)},Y_{(h,i)})\big\}=o_p(c_1).$$

This lemma is analogous to Hsing and Carroll (1992, LEMMA A.3).

Proof of Lemma 5: In what follows, we only prove the first argument because the others can be proved in a similar way. Let

$$D_h=\frac{1}{n_h-1}\sum_{1\le i\neq j\le n_h}|m(Y_{(h,i)},Y_{(h,j)})-m(Y_{(h,i)},Y_{(h,i)})|.$$

Step 1. If Y is boundedly supported, under the assumptions,

$$\begin{aligned}
n^{-\gamma}\sum_{h=1}^{H}D_h &\le n^{-\gamma}\sum_{h=1}^{H}\frac{1}{n_h-1}\sum_{1\le i\neq j\le n_h}\big|M(Y_{(h,i)})-M(Y_{(h,j)})\big|\\
&\le n^{-\gamma}\sum_{h=1}^{H}n_h\sum_{i=1}^{n_h-1}\big|M(Y_{(h,i+1)})-M(Y_{(h,i)})\big|\\
&\le c_1n^{-\gamma}\sum_{j=1}^{n-1}\big|M(Y_{(j+1)})-M(Y_{(j)})\big|=o(c_1),
\end{aligned}$$

where the last equality follows from the definition of total variation (c.f. Zhu and Ng, 1995, Page 729). If the support of Y is unbounded, it suffices to show that

$$c_1^{-1} n^{-\gamma} \sum_{h=[H\delta]}^{[H(1-\delta)]} D_h \xrightarrow{p} 0, \tag{S3.1}$$

and, for $t > 0$,

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \Big[\mathrm{pr}\Big\{ c_1^{-1} n^{-\gamma} \sum_{h=1}^{[H\delta]} D_h > t \Big\} + \mathrm{pr}\Big\{ c_1^{-1} n^{-\gamma} \sum_{h=[H(1-\delta)]}^{H} D_h > t \Big\} \Big] = 0. \tag{S3.2}$$

Step 2. We now show (S3.1). Fix $\delta \in (0, 1/2)$. Let F_Y denote the distribution function of Y and F_Y^{-1} the left-continuous inverse of F_Y. Define $A_n = I\left\{Y_{([n\delta])} > F_Y^{-1}(\beta)\right\}$ and $B_n = I\left\{Y_{([n(1-\delta)])} < F_Y^{-1}(1-\beta)\right\}$ for $0 < \beta < \delta$. Given any such β, we have $A_n \xrightarrow{p} 1$ and $B_n \xrightarrow{p} 1$. Thus (S3.1) follows from

$$c_1^{-1} n^{-\gamma} \sum_{[H\delta]}^{[H(1-\delta)]} D_h A_n B_n \longrightarrow 0,$$

which, in turn, follows from a similar procedure to **Step 1** by noting that under the event $\{A_n = 1, B_n = 1\}$, the Y's in the summation are boundedly supported.

Step 3. Then we show (S3.2). Choose $\delta > 0$ small enough so that $C_n \xrightarrow{p} 1$, where $C_n = I\left\{Y_{([n\delta])} < -B_0\right\}$. Under the non-expansive condition, we have that

$$\begin{aligned} c_1^{-1} n^{-\gamma} \sum_{h=1}^{[H\delta]} D_h C_n &\leq n^{-\gamma} \sum_{j=1}^{[n\delta]-1} \left| M(Y_{(j+1)}) - M(Y_{(j)}) \right| C_n \\ &\leq n^{-\gamma} \left| M(Y_{([n\delta])}) - M(Y_{(1)}) \right| \\ &\leq n^{-\gamma} \left\{ \left| M(Y)_{([n\delta])} \right| + \left| M(Y)_{(1)} \right| \right\} = o_p(1). \end{aligned}$$

where the two equalities follow from Lemma 4 and that, under the event $\{C_n = 1\}$, $M(Y)$ is non-decreasing, respectively. Together with $C_n \xrightarrow{p} 1$, we have

$$c_1^{-1} n^{-\gamma} \sum_{h=1}^{[H\delta]} D_h \xrightarrow{p} 0,$$

and the other tail can be handled similarly. Thus $n^{-\gamma} \sum_{h=1}^{H} D_h = o_p(c_1)$. □

References

Da Prato, G. and J. Zabczyk (2014). *Stochastic Equations in Infinite Dimensions.* Encyclopedia of Mathematics and its Applications. Cambridge University Press.

Gupta, S. S. (1963). Probability integrals of multivariate normal and multivariate t^1. *The Annals of Mathematical Statistics 34*(3), 792–828.

Hsing, T. and R. J. Carroll (1992). An asymptotic theory for sliced inverse regression. *The Annals of Statistics 20*(2), 1040–1061.

Jacod, J. and P. Protter (2012). *Probability Essentials.* Universitext. Springer Berlin Heidelberg.

REFERENCES

Ke, C. and X. Yin (2020). Expected conditional characteristic function-based measures for testing independence. *Journal of the American Statistical Association 115*(530), 985–996.

Resnick, S. (2019). *A Probability Path*. Birkhäuser Boston.

Serfling, R. (2009). *Approximation Theorems of Mathematical Statistics*. Wiley Series in Probability and Statistics. Wiley.

van Zanten, J. and A. van der Vaart (2008). Reproducing kernel hilbert spaces of gaussian priors. In *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*, Volume 3, pp. 200–222. Institute of Mathematical Statistics.

Zhu, L. and K. Ng (1995). Asymptotics of sliced inverse regression. *Statistica Sinica 5*(2), 727–736.