Supplement to

"Model-Free Conditional Feature Screening with FDR Control"

Zhaoxue Tong[1], Zhanrui Cai[2], Songshan Yang[3] and Runze Li[1]

[1]Pennsylvania State University

[2]Carnegie Mellon University

[3]Renmin University of China

## S.1 Technical Lemmas

Lemma 1 can help us understand the model-free property of $\widehat{\rho}(X_k, Y \mid \mathbf{z})$.

**Lemma 1.** *Under Condition (C1) - (C4),*

1. (Consistency) $\widehat{\rho}(X_k, Y \mid \mathbf{z}) \xrightarrow{p} \rho(X_k, Y \mid \mathbf{z})$ *for* $k \in \mathcal{M}$*;*
2. (Distribution Free) $n\widehat{\rho}(X_k, Y \mid \mathbf{z}) \xrightarrow{d} \mathcal{N}_k$ *for* $k \in \mathcal{M}^c$*, where* $\mathcal{N}_k$ *is some non-degenerate random variable that does not depend on the distribution of* $X_k, Y$ *and* $\mathbf{z}$*.*

This is a direct result of Theorem 3 in (Cai et al., 2022). $\mathcal{N}_k$ is actually an infinite sum of weighted $\chi^2$, and the weights are real numbers associated with the distribution of $U, V$ and $\mathbf{w}$. Since, when $k \in \mathcal{M}^c$, $U, V$ and $\mathbf{w}$ are mutually independent and all follow Uniform$(0, 1)$ distribution, we know that $\widehat{\rho}(X_k, Y \mid \mathbf{z})$ is model-free in the sense that its distribution does not depend on the distribution of $X_k, Y$ and $\mathbf{z}$.

## S.2 Proofs

We begin by introducing some notations. Let $P$ be a probability measure. The $L_r(P)$-norm of a function $f$ is denoted as $\|f\|_{L_r(P)} = \left(\int |f|^r \, dP\right)^{\frac{1}{r}}$. For simplicity, we denote $c_i$, $c$, $C_i$, and $C$ as some positive constants that may take different values (independent of $n$ and $p$) in each appearance throughout this section.

**Definition 1** (Covering Number). *Let $\mathcal{T}$ be some subset of a metric space $(T, D)$, where $T$ is a set and $D$ is a metric on $T$. For $\epsilon > 0$, the $\epsilon$-covering number $N(\epsilon, \mathcal{T}, D)$ of $\mathcal{T}$ is the minimum number of balls with radius $\epsilon$, needed to cover $\mathcal{T}$. Specifically, it is the smallest value of $N$, such that there exist $t_1, ..., t_N$ in $T$, and for all $s \in \mathcal{T}$.*

$$\min_{j=1,\dots,N} D(s, t_j) \leq \epsilon$$

**Definition 2** (Envelope Function). *An envelope $F$ of a collection of functions $\mathcal{F}$ is any function $x \mapsto F(x)$ such that $|f(x)| \leq F(x)$, for any $x$ and $f \in \mathcal{F}$.*

**Definition 3** (Graph). *The graph of a real-valued function $f$ on a set $S$ is defined as the subset $\{(s,t) : 0 \leq t \leq f(s) \text{ or } f(s) \leq t \leq 0\}$ of $S \otimes \mathbb{R}$, where $\otimes$ denotes product $\sigma$-field. The subgraph of $f$ is defined as the subset $\{(s,u) : f(s) \geq u\}$ of $S \otimes \mathbb{R}$.*

**Definition 4** (VC Class). *Let $\mathcal{F}$ be a uniformly bounded collection of measurable functions on a measurable space $(S, \mathcal{S})$ and let $F$ be an envelope of $\mathcal{F}$. $\mathcal{F}$ is called a bounded measurable VC class of functions if*

1. (Bounded Measurable) *the class $\mathcal{F}$ is separable or is image admissible Suslin (Dudley, 2014),*
2. (VC Class) *and there exist some positive numbers $A$ and $v$ such that*

$$N(\epsilon \|F\|_{L_2(P)}, \mathcal{F}, L_2(P)) \leq \left(\frac{A}{\epsilon}\right)^v \tag{S1}$$

*for any probability measure $P$ on $(S, \mathcal{S})$ and any $\epsilon \in (0, 1)$. We refer to $A$ and $v$ as the VC characteristics of the class $\mathcal{F}$.*

By the definition of $\epsilon$-covering number, we assume in what follows that $A \geq 3\sqrt{e}$ and $v \geq 1$ for the convenience of proof. Interested readers are referred to Giné and Guillou (2001); Giné and Guillou (2002) for details.

### S.2.1 Technical Lemmas

**Lemma 2.** (Hoeffding's Lemma) *If $pr(a \leq Y \leq b) = 1$, then*

$$\mathbb{E}\left[\exp\left\{s(Y - \mathbb{E}Y)\right\}\right] \leq \exp\left\{s^2(b-a)^2/8\right\}, \text{ for all } s > 0.$$

**Lemma 3.** *Under (C1.1), the class of functions $\mathcal{F} = \left\{\mathbf{x} \mapsto K(\mathbf{t} - \mathbf{x}) : \mathbf{t} \in \mathbb{R}^d\right\}$ is a bounded measurable VC class of functions.*

***Proof.*** The following proof is adapted from Giné and Guillou (2002) Page 911.

Let $\rho$ denote a polynomial on $\mathbb{R}^d \times \mathbb{R}$ and $\varphi$ denote a real measurable function. Then the family of sets $\{\{(\mathbf{s}, u) : \rho((\mathbf{t} - \mathbf{s})/h, u) \geq \varphi(u)\} : \mathbf{t} \in \mathbb{R}^d, h > 0\}$ is contained in the family of positivity sets (see definition in Dudley (2014) Section 4.2 Page 179) of a finite dimensional space of functions. By Theorem 4.6 and 4.8 in Dudley (2014), $\mathcal{F}$ is a bounded VC class of measurable functions.

And since the map $(\mathbf{t}, \mathbf{x}) \mapsto K(\mathbf{t} - \mathbf{x})$ is jointly measurable, the class $\mathcal{F}$ is image admissible Suslin, hence measurable. Thus $\mathcal{F}$ is a bounded measurable VC class of (measurable) functions.

□

**Lemma 4.** *Under (C1.1), (C1.3) and (C3), consider $\widehat{f}_{\widetilde{\mathbf{z}}}(\mathbf{z})$, $\widehat{F}_{\widetilde{z}}(z)$ and $\widehat{g}(\mathbf{z}, x; \widetilde{X})$ defined in equation groups (7) and (8) with $\mathbf{z} \in \mathbb{R}^d$, $z \in \mathbb{R}$ and $x \in \mathbb{R}$. For any fixed $\epsilon > 0$, large enough $n \in \mathbb{N}$, $d\theta < 1/2$, and some positive constants $C_1$ and $C_2$, we have*

$$pr\left\{\sup_{\mathbf{z}}\left|\widehat{f}_{\widetilde{\mathbf{z}}}(\mathbf{z}) - \mathbb{E}_{\widetilde{\mathbf{z}}}\left\{\widehat{f}_{\widetilde{\mathbf{z}}}(\mathbf{z})\right\}\right| \geq \epsilon\right\} \leq C_1\left(n^{\frac{1}{2}-\theta d}\epsilon\right)^{C_2}\exp\left(-2M_K^{-2}n^{1-2\theta d}\epsilon^2\right), \quad \text{(S2)}$$

$$pr\left\{\sup_{z}\left|\widehat{F}_{\widetilde{z}}(z)-\mathbb{E}_{\widetilde{z}}\left\{\widehat{F}_{\widetilde{z}}(z)\right\}\right|\geq\epsilon\right\}\leq C_1\left(n^{\frac{1}{2}-\theta}\epsilon\right)^{C_2}\exp\left(-2M_K^{-2}n^{1-2\theta}\epsilon^2\right), \tag{S3}$$

$$pr\left\{\sup_{\mathbf{z},x}\left|\widehat{g}(\mathbf{z},x;\widetilde{X})-\mathbb{E}_{\widetilde{\mathbf{z}},\widetilde{X}}\left\{\widehat{g}(\mathbf{z},x;\widetilde{X})\right\}\right|\geq\epsilon\right\}\leq C_1\left(n^{\frac{1}{2}-\theta d}\epsilon\right)^{C_2}\exp\left(-2M_K^{-2}n^{1-2\theta d}\epsilon^2\right). \tag{S4}$$

***Proof.*** We first prove (S2). By Lemma 3 and Condition (C1.1), the class of functions $\mathcal{F}_1 := \{\mathbf{z}\mapsto K_h(\check{\mathbf{z}}-\mathbf{z}) : \check{\mathbf{z}}\in\mathbb{R}^d\}$ is a uniformly bounded measurable VC class of functions with VC characteristics $A$ and $v$. Thus, the covering number of $\mathcal{F}_1$ satisfies condition (2.14.6) in Theorem 2.14.9 in van der Vaart and Wellner (1996). For all $m_1\in\mathcal{F}_1$, we have $\|m_1\|_\infty\leq B_U := h^{-d}M_K = n^{\theta d}M_K$ by Condition (C1.3). Denote $\mathbb{P}_n m_1 = n^{-1}\sum_{i=1}^n m_1(\widetilde{\mathbf{z}}_i)$ and $Pm_1 = \mathbb{E}_{\widetilde{\mathbf{z}}}\{m_1(\widetilde{\mathbf{z}})\}$. Without loss of generality, assume $K(\cdot)\geq 0$. We have

$$\begin{aligned}
\mathrm{pr}\left\{\sup_{\mathbf{z}}\left|\widehat{f}_{\widetilde{\mathbf{z}}}(\mathbf{z})-\mathbb{E}_{\widetilde{\mathbf{z}}}\left\{\widehat{f}_{\widetilde{\mathbf{z}}}(\mathbf{z})\right\}\right|\geq\epsilon\right\} &= \mathrm{pr}\left\{\sup_{m_1\in\mathcal{F}_1}|(\mathbb{P}_n-P)m_1|\geq\epsilon\right\}\\
&= \mathrm{pr}\left\{\sqrt{n}\sup_{m_1\in\mathcal{F}_1}\left|(\mathbb{P}_n-P)\frac{m_1}{B_U}\right|\geq\sqrt{n}\frac{\epsilon}{B_U}\right\}\\
&\leq\left(\frac{C\sqrt{n}\epsilon}{\sqrt{v}B_U}\right)^v\exp\left(-2\frac{n\epsilon^2}{B_U^2}\right)\\
&=\left(\frac{C}{\sqrt{v}M_K}\right)^v\left(n^{\frac{1}{2}-\theta d}\epsilon\right)^v\exp\left(-\frac{2}{M_K^2}n^{1-2\theta d}\epsilon^2\right),
\end{aligned}$$

where $C$ depends on $A$ and $v$ only. The inequality holds due to Theorem 2.14.9 in van der Vaart and Wellner (1996). Hence (S2) is proved.

Next, we prove (S3). By Lemma 2.6.16 in van der Vaart and Wellner (1996), the class of functions $\mathcal{F}_2 := \{z\mapsto \mathbb{1}(z\leq\check{z}) : \check{z}\in\mathbb{R}\}$ is a uniformly bounded measurable VC class of functions. Repeating the proof for (S2) will show that (S3) holds.

Finally, we prove (S4). By the proof of Lemma 3 and Section 5 in Nolan and Pollard (1987), the subgraph of any $m_1\in\mathcal{F}_1$, $\{(\mathbf{z},u) : m_1(\mathbf{z})\geq u, u\in\mathbb{R}\}$, is a polynomial class (Pollard, 2012, Definition II.13). Similarly, the subgraph of any $m_2\in\mathcal{F}_2$, $\{(x,u) : m_2(x)\geq u, u\in\mathbb{R}\}$ is also a polynomial class. Consider the class of functions $\mathcal{F}_3 := \{(\mathbf{z},x)\mapsto K_h(\check{\mathbf{z}}-\mathbf{z})\mathbb{1}(x\leq\check{x}) : \check{\mathbf{z}}\in\mathbb{R}^d, \check{x}\in\mathbb{R}\} = \mathcal{F}_1\cdot\mathcal{F}_2 = \{m_1\cdot m_2 : m_1\in\mathcal{F}_1, m_2\in\mathcal{F}_2\}$. The subgraph of any $m_3\in\mathcal{F}_3$ can be represented as

$$
\begin{aligned}
& \{(\mathbf{z}, x, t) : m_1(\mathbf{z}) m_2(x) \geq t, t \in \mathbb{R}\} \\
= \quad & (\{(\mathbf{z}, x, t) : m_1(\mathbf{z}) \geq t, x \in \mathbb{R}, t > 0\} \cap \{(\mathbf{z}, x, t) : m_2(x) = 1, \mathbf{z} \in \mathbb{R}, t > 0\}) \\
& \cup (\{(\mathbf{z}, x, t) : m_1(\mathbf{z}) \geq t, x \in \mathbb{R}, t \leq 0\} \cap \{(\mathbf{z}, x, t) : m_2(x) = 1, \mathbf{z} \in \mathbb{R}, t \leq 0\}) \\
& \cup \{(\mathbf{z}, x, t) : m_2(x) = 0, \mathbf{z} \in \mathbb{R}, t \leq 0\},
\end{aligned}
$$

which is a finite number of Boolean operations among sets of polynomial class. By Lemma 18 in Nolan and Pollard (1987), $\{(\mathbf{z}, x, t) : m_1(\mathbf{z}) m_2(x) \geq t, t \in \mathbb{R}\}$ is also a polynomial class, so $\mathcal{F}_3$ is a uniformly bounded measurable VC class of functions. Repeating the proof for (S2) will show that (S4) holds. $\square$

**Lemma 5.** *Suppose that Condition (C1) to (C4) are fulfilled. Consider* $\widehat{f}_{\widetilde{\mathbf{z}}}(\mathbf{z})$, $\widehat{F}_{\widetilde{z}}(z)$ *and* $\widehat{g}(\mathbf{z}, x; \widetilde{X})$ *defined in Lemma 4. Then, for any* $0 < \gamma + d\theta \leq 1/2$ *and* $0 < \gamma \leq 2\theta$*, we have*

$$pr\left\{\sup_{\mathbf{z},x} \left|\widehat{g}(\mathbf{z}, x; \widetilde{X}) - g(\mathbf{z}, x; \widetilde{X})\right| \geq \tau n^{-\gamma}\right\} \leq C_3 n^{C_4(1-2\gamma-2\theta d)} \exp\left(-C n^{1-2\gamma-2\theta d}\right), \quad \text{(S5)}$$

$$pr\left\{\sup_{\mathbf{z}} \left|\widehat{f}_{\widetilde{\mathbf{z}}}(\mathbf{z}) - f_{\widetilde{\mathbf{z}}}(\mathbf{z})\right| \geq \tau n^{-\gamma}\right\} \leq C_3 n^{C_4(1-2\gamma-2\theta d)} \exp\left(-C n^{1-2\gamma-2\theta d}\right), \quad \text{(S6)}$$

$$pr\left\{\sup_{z} \left|\widehat{F}_{\widetilde{z}}(z) - F_{\widetilde{z}}(z)\right| \geq \tau n^{-\gamma}\right\} \leq C_3 n^{C_4(1-2\gamma-2\theta)} \exp\left(-C n^{1-2\gamma-2\theta}\right), \quad \text{(S7)}$$

*for some positive constants* $\tau, C_3, C_4$ *and* $C$*.*

***Proof.*** It suffices to prove (S5) since (S6) and (S7) can be proved similarly. The proof consists of two steps:

**Step 1.** We prove that, for $0 < \gamma \leq 2\theta$, there exists some $\tau > 0$ such that

$$\sup_{\mathbf{z},x} \left|\mathbb{E}_{\widetilde{\mathbf{z}},\widetilde{X}}\left\{\widehat{g}(\mathbf{z}, x; \widetilde{X})\right\} - g(\mathbf{z}, x; \widetilde{X})\right| \leq \tau n^{-\gamma}/2 \quad \text{(S8)}$$

Note that

$$
\begin{aligned}
\mathbb{E}_{\widetilde{\mathbf{z}},\widetilde{X}}\left\{\widehat{g}(\mathbf{z}, x; \widetilde{X})\right\} &= \mathbb{E}_{\widetilde{\mathbf{z}},\widetilde{X}}\left\{K_h(\mathbf{z} - \widetilde{\mathbf{z}}) \mathbb{1}(\widetilde{X} < x)\right\} \\
&= \mathbb{E}_{\widetilde{\mathbf{z}}}\left[\mathbb{E}_{\widetilde{\mathbf{z}},\widetilde{X}}\left\{K_h(\mathbf{z} - \widetilde{\mathbf{z}}) \mathbb{1}(\widetilde{X} < x) \mid \widetilde{\mathbf{z}}\right\}\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \mathbb{E}_{\widetilde{\mathbf{z}}}\left[K_h(\mathbf{z}-\widetilde{\mathbf{z}})\mathbb{E}_{\widetilde{\mathbf{z}},\widetilde{X}}\left\{\mathbb{1}(\widetilde{X}<x)\mid\widetilde{\mathbf{z}}\right\}\right] \\
&= \int_{\mathbb{R}^d} K_h(\mathbf{z}-\mathbf{u})\mathbb{E}_{\widetilde{X}|\widetilde{\mathbf{z}}}\left\{\mathbb{1}(\widetilde{X}<x)\mid\widetilde{\mathbf{z}}=\mathbf{u}\right\} f_{\widetilde{\mathbf{z}}}(\mathbf{u})\, d\mathbf{u}.
\end{aligned}
$$

Expanding $g(\mathbf{z},x;\widetilde{X})$ with respect to $\mathbf{z}$ in a Taylor series (Chacón and Duong, 2018, Equation (2.5)) using Condition (C1.2) and (C4) gives that

$$
\begin{aligned}
&\sup_{\mathbf{z},x}\left|\mathbb{E}_{\widetilde{\mathbf{z}},\widetilde{X}}\left\{\widehat{g}(\mathbf{z},x;\widetilde{X})\right\}-g(\mathbf{z},x;\widetilde{X})\right| \\
=\;& \sup_{\mathbf{z},x}\left|\int_{\mathbb{R}^d} K_h(\mathbf{z}-\mathbf{u})\left[\mathbb{E}_{\widetilde{X}|\widetilde{\mathbf{z}}}\left\{\mathbb{1}(\widetilde{X}<x)\mid\widetilde{\mathbf{z}}=\mathbf{u}\right\} f_{\widetilde{\mathbf{z}}}(\mathbf{u})-\mathbb{E}_{\widetilde{X}|\widetilde{\mathbf{z}}}\left\{\mathbb{1}(\widetilde{X}<x)\mid\widetilde{\mathbf{z}}=\mathbf{z}\right\} f_{\widetilde{\mathbf{z}}}(\mathbf{z})\right] d\mathbf{u}\right| \\
=\;& \sup_{\mathbf{z},x}\left|\int_{\mathbb{R}^d} K(\mathbf{t})\left[\mathbb{E}_{\widetilde{X}|\widetilde{\mathbf{z}}}\left\{\mathbb{1}(\widetilde{X}<x)\mid\widetilde{\mathbf{z}}=\mathbf{z}+h\mathbf{t}\right\} f_{\widetilde{\mathbf{z}}}(\mathbf{z}+h\mathbf{t})-\mathbb{E}_{\widetilde{X}|\widetilde{\mathbf{z}}}\left\{\mathbb{1}(\widetilde{X}<x)\mid\widetilde{\mathbf{z}}=\mathbf{z}\right\} f_{\widetilde{\mathbf{z}}}(\mathbf{z})\right] d\mathbf{t}\right| \\
=\;& \sup_{\mathbf{z},x}\left|\int_{\mathbb{R}^d} K(\mathbf{t})\left\{F_{\widetilde{X}|\widetilde{\mathbf{z}}}(x\mid\mathbf{z}+h\mathbf{t})f_{\widetilde{\mathbf{z}}}(\mathbf{z}+h\mathbf{t})-F_{\widetilde{X}|\widetilde{\mathbf{z}}}(x\mid\mathbf{z})f_{\widetilde{\mathbf{z}}}(\mathbf{z})\right\} d\mathbf{t}\right| \\
=\;& \sup_{\mathbf{z},x}\left|\int_{\mathbb{R}^d} K(\mathbf{t})\left\{g(\mathbf{z}+h\mathbf{t},x;\widetilde{X})-g(\mathbf{z},x;\widetilde{X})\right\} d\mathbf{t}\right| \\
\le\;& \alpha\kappa h^2=\alpha\kappa n^{-2\theta}
\end{aligned}
$$

holds for some constant $\alpha>0$, where $\mathbf{t}=(\mathbf{u}-\mathbf{z})/h$. Then for $0<\gamma\le 2\theta$, there exists some $\tau>0$ such that

$$
\sup_{\mathbf{z},x}\left|\mathbb{E}_{\widetilde{\mathbf{z}},\widetilde{X}}\left\{\widehat{g}(\mathbf{z},x;\widetilde{X})\right\}-g(\mathbf{z},x;\widetilde{X})\right|\le\tau n^{-\gamma}/2.
$$

**Step 2.** By (S8) and Lemma 4, we have

$$
\begin{aligned}
&\operatorname{pr}\left\{\sup_{\mathbf{z},x}\left|\widehat{g}(\mathbf{z},x;\widetilde{X})-g(\mathbf{z},x;\widetilde{X})\right|\ge\tau n^{-\gamma}\right\} \\
\le\;& \operatorname{pr}\left\{\sup_{\mathbf{z},x}\left|\widehat{g}(\mathbf{z},x;\widetilde{X})-\mathbb{E}_{\widetilde{\mathbf{z}},\widetilde{X}}\left\{\widehat{g}(\mathbf{z},x;\widetilde{X})\right\}\right|\ge\tau n^{-\gamma}/2\right\} \\
\le\;& C_3 n^{C_4(1-2\gamma-2\theta d)}\exp\left(-Cn^{1-2\gamma-2\theta d}\right)
\end{aligned}
$$

for some positive constants $\tau, C_3, C_4$ and $C$. This is because we can set the $\epsilon$ in the LHS of (S4) as $\tau n^{-\gamma}/2$ and plug $\epsilon$ into the RHS of (S4). Hence (S5) is proved.

□

**Lemma 6.** *Suppose the conditions in Lemma 5 hold, we have*

$$pr\left\{\sup_{\mathbf{z},x}\left|\frac{\widehat{g}(\mathbf{z},x;\widetilde{X})}{\widehat{f}_{\widetilde{\mathbf{z}}}(\mathbf{z})}-\frac{g(\mathbf{z},x;\widetilde{X})}{f_{\widetilde{\mathbf{z}}}(\mathbf{z})}\right|\geq\tau n^{-\gamma}\right\} \quad\leq\quad C_5 n^{C_6(1-2\gamma-2\theta d)}\exp\left(-Cn^{1-2\gamma-2d\theta}\right)$$

*for some positive constants* $\tau, C_5, C_6$ *and* $C$.

***Proof.*** Under Condition (C3), there exists some constant $\delta_0\in(0,1)$ such that $M_f:=M_L-\delta_0>0$. We have

$$\begin{aligned}
& \mathrm{pr}\left\{\sup_{\mathbf{z},x}\left|\frac{\widehat{g}(\mathbf{z},x;\widetilde{X})}{\widehat{f}_{\widetilde{\mathbf{z}}}(\mathbf{z})}-\frac{g(\mathbf{z},x;\widetilde{X})}{f_{\widetilde{\mathbf{z}}}(\mathbf{z})}\right|\geq\tau n^{-\gamma}\right\}\\
= & \mathrm{pr}\left\{\sup_{\mathbf{z},x}\left|\frac{\widehat{g}(\mathbf{z},x;\widetilde{X})-g(\mathbf{z},x;\widetilde{X})}{\widehat{f}_{\widetilde{\mathbf{z}}}(\mathbf{z})}+g(\mathbf{z},x;\widetilde{X})\left(\frac{1}{\widehat{f}_{\widetilde{\mathbf{z}}}(\mathbf{z})}-\frac{1}{f_{\widetilde{\mathbf{z}}}(\mathbf{z})}\right)\right|\geq\tau n^{-\gamma}\right\}\\
\leq & \mathrm{pr}\left\{\sup_{\mathbf{z},x}\left|\frac{\widehat{g}(\mathbf{z},x;\widetilde{X})-g(\mathbf{z},x;\widetilde{X})}{\widehat{f}_{\widetilde{\mathbf{z}}}(\mathbf{z})}\right|\geq\tau n^{-\gamma}/2,\left|\widehat{f}_{\widetilde{\mathbf{z}}}(\mathbf{z})\right|\geq M_f\right\}+\\
& \mathrm{pr}\left\{\sup_{\mathbf{z},x}\left|g(\mathbf{z},x;\widetilde{X})\right|\frac{\left|\widehat{f}_{\widetilde{\mathbf{z}}}(\mathbf{z})-f_{\widetilde{\mathbf{z}}}(\mathbf{z})\right|}{\left|\widehat{f}_{\widetilde{\mathbf{z}}}(\mathbf{z})\right|\left|f_{\widetilde{\mathbf{z}}}(\mathbf{z})\right|}\geq\tau n^{-\gamma}/2,\left|\widehat{f}_{\widetilde{\mathbf{z}}}(\mathbf{z})\right|\geq M_f\right\}+\\
& \mathrm{pr}\left\{\left|\widehat{f}_{\widetilde{\mathbf{z}}}(\mathbf{z})\right|<M_f\right\}.
\end{aligned}\tag{S9}$$

For the first term of the RHS of (S9), by Lemma 5, for some positive constants $C_5', C_6'$ and $C'$,

$$\begin{aligned}
& \mathrm{pr}\left\{\sup_{\mathbf{z},x}\left|\frac{\widehat{g}(\mathbf{z},x;\widetilde{X})-g(\mathbf{z},x;\widetilde{X})}{\widehat{f}_{\widetilde{\mathbf{z}}}(\mathbf{z})}\right|\geq\tau n^{-\gamma}/2,\left|\widehat{f}_{\widetilde{\mathbf{z}}}(\mathbf{z})\right|\geq M_f\right\}\\
\leq & \mathrm{pr}\left\{\sup_{\mathbf{z},x}\left|\widehat{g}(\mathbf{z},x;\widetilde{X})-g(\mathbf{z},x;\widetilde{X})\right|\geq\tau M_f n^{-\gamma}/2\right\}\\
\leq & C_5' n^{C_6'(1-2\gamma-2\theta d)}\exp\left(-C'n^{1-2\gamma-2\theta d}\right).
\end{aligned}$$

For the second term of the RHS of (S9), by Lemma 5, for some positive constants $C_5'', C_6''$ and $C''$,

$$\begin{aligned}
& \mathrm{pr}\left\{\sup_{\mathbf{z},x}\left|g(\mathbf{z},x;\widetilde{X})\right|\frac{\left|\widehat{f}_{\widetilde{\mathbf{z}}}(\mathbf{z})-f_{\widetilde{\mathbf{z}}}(\mathbf{z})\right|}{\left|\widehat{f}_{\widetilde{\mathbf{z}}}(\mathbf{z})\right|\left|f_{\widetilde{\mathbf{z}}}(\mathbf{z})\right|}\geq\tau n^{-\gamma}/2,\left|\widehat{f}_{\widetilde{\mathbf{z}}}(\mathbf{z})\right|\geq M_f\right\}\\
\leq & \mathrm{pr}\left\{\sup_{\mathbf{z},x}\left|\widehat{f}_{\widetilde{\mathbf{z}}}(\mathbf{z})-f_{\widetilde{\mathbf{z}}}(\mathbf{z})\right|\geq\tau M_f M_L M_U^{-1} n^{-\gamma}/2\right\}\\
\leq & C_5'' n^{C_6''(1-2\gamma-2\theta d)}\exp\left(-C''n^{1-2\gamma-2\theta d}\right).
\end{aligned}$$

For the third term of the RHS of (S9), we have

$$\begin{aligned}
\mathrm{pr}\left\{\left|\widehat{f}_{\widetilde{\mathbf{z}}}(\mathbf{z})\right| < M_f\right\} &= \mathrm{pr}\left\{\left|f_{\widetilde{\mathbf{z}}}(\mathbf{z}) + \widehat{f}_{\widetilde{\mathbf{z}}}(\mathbf{z}) - f_{\widetilde{\mathbf{z}}}(\mathbf{z})\right| < M_f\right\} \\
&\leq \mathrm{pr}\left\{\left|f_{\widetilde{\mathbf{z}}}(\mathbf{z})\right| - \left|\widehat{f}_{\widetilde{\mathbf{z}}}(\mathbf{z}) - f_{\widetilde{\mathbf{z}}}(\mathbf{z})\right| < M_L - \delta_0\right\} \\
&\leq \mathrm{pr}\left\{\left|\widehat{f}_{\widetilde{\mathbf{z}}}(\mathbf{z}) - f_{\widetilde{\mathbf{z}}}(\mathbf{z})\right| > \delta_0\right\},
\end{aligned}$$

then by Lemma 5, for some positive constants $C_5''', C_6'''$ and $C'''$, choose $\delta_0 = \tau n^{-\gamma}$ for some proper $\tau$ and $\gamma$,

$$\mathrm{pr}\left\{\left|\widehat{f}_{\widetilde{\mathbf{z}}}(\mathbf{z})\right| < M_f\right\} \leq \mathrm{pr}\left\{\left|\widehat{f}_{\widetilde{\mathbf{z}}}(\mathbf{z}) - f_{\widetilde{\mathbf{z}}}(\mathbf{z})\right| > \tau n^{-\gamma}\right\} \leq C_5''' n^{C_6'''(1-2\gamma-2\theta d)} \exp\left(-C''' n^{1-2\gamma-2\theta d}\right).$$

Hence we have

$$\mathrm{pr}\left\{\sup_{\mathbf{z},x}\left|\frac{\widehat{g}(\mathbf{z},x;\widetilde{X})}{\widehat{f}_{\widetilde{\mathbf{z}}}(\mathbf{z})} - \frac{g(\mathbf{z},x;\widetilde{X})}{f_{\widetilde{\mathbf{z}}}(\mathbf{z})}\right| \geq \tau n^{-\gamma}\right\} \leq C_5 n^{C_6(1-2\gamma-2\theta d)} \exp\left(-C n^{1-2\gamma-2\theta d}\right)$$

for some positive constants $C_5, C_6$ and $C$.

$\square$

**Lemma 7.** *Suppose that $A(u)$ and $B(u)$ are functions of vector $u \in \mathbb{U}$. They satisfy*

$$\sup_{u\in\mathbb{U}}|A(u)| \leq M_A < \infty, \quad \sup_{u\in\mathbb{U}}|B(u)| \leq M_B < \infty.$$
$$\sup_{u\in\mathbb{U}}\left|\widehat{A}(u)\right| \leq M_{\widehat{A}} < \infty, \quad \sup_{u\in\mathbb{U}}\left|\widehat{B}(u)\right| \leq M_{\widehat{B}} < \infty.$$

*Suppose we have*

$$\begin{aligned}
\sup_{u\in\mathbb{U}} pr\left\{\left|\widehat{A}(u) - A(u)\right| \geq \tau n^{-\gamma}\right\} &\leq C_7' n^{C_8'(1-\eta)} \exp\left(-C' n^{1-\eta}\right), \\
\sup_{u\in\mathbb{U}} pr\left\{\left|\widehat{B}(u) - B(u)\right| \geq \tau n^{-\gamma}\right\} &\leq C_7'' n^{C_8''(1-\eta)} \exp\left(-C'' n^{1-\eta}\right),
\end{aligned}$$

*where $\tau, \gamma, C_7', C_7'', C_8', C_8''$ and $\eta < 1$ are positive constants. Then we have*

$$\sup_{u\in\mathbb{U}} pr\left\{\left|\widehat{A}(u)\widehat{B}(u) - A(u)B(u)\right| \geq \tau n^{-\gamma}\right\} \leq C_7 n^{C_8(1-\eta)} \exp\left(-C n^{1-\eta}\right)$$

*for some positive constants $C_7, C_8$ and $C$.*

***Proof.*** For some proper constants $C_9, C_9', C_{10}, C_{10}', C_{11}$ and $C_{11}'$,

$$\begin{aligned}
& \sup_{u\in\mathbb{U}} \mathrm{pr}\left\{\left|\widehat{A}(u)\widehat{B}(u) - A(u)B(u)\right| \geq \tau n^{-\gamma}\right\} \\
= & \sup_{u\in\mathbb{U}} \mathrm{pr}\left\{\left|\widehat{A}(u)\widehat{B}(u) - \widehat{A}(u)B(u) + \widehat{A}(u)B(u) - A(u)B(u)\right| \geq \tau n^{-\gamma}\right\} \\
\leq & \sup_{u\in\mathbb{U}} \mathrm{pr}\left\{\left|\widehat{A}(u)\right|\left|\widehat{B}(u) - B(u)\right| \geq \tau n^{-\gamma}/2\right\} + \sup_{u\in\mathbb{U}} \mathrm{pr}\left\{|B(u)|\left|\widehat{A}(u) - A(u)\right| \geq \tau n^{-\gamma}/2\right\} \\
\leq & \sup_{u\in\mathbb{U}} \mathrm{pr}\left\{\left|\widehat{B}(u) - B(u)\right| \geq \tau M_{\widehat{A}}^{-1} n^{-\gamma}/2\right\} + \sup_{u\in\mathbb{U}} \mathrm{pr}\left\{\left|\widehat{A}(u) - A(u)\right| \geq \tau M_{B}^{-1} n^{-\gamma}/2\right\} \\
\leq & C_9 n^{C_{10}(1-\eta)} \exp\left(-C_{11} n^{1-\eta}\right) + C_9' n^{C_{10}'(1-\eta)} \exp\left(-C_{11}' n^{1-\eta}\right) \\
\leq & C_7 n^{C_8(1-\eta)} \exp\left(-C n^{1-\eta}\right),
\end{aligned}$$

where $C_7, C_8$ and $C$ are some proper positive constants.

□

### S.2.2 Proof of Theorem 1(i)

Denote $\rho(X_k, Y|Z)$ as $\rho_k$ and $\widehat{\rho}(X_k, Y|Z)$ as $\widehat{\rho}_k$. Define

$$\begin{aligned}
\widetilde{\rho}_k = \widetilde{\rho}(X_k, Y \mid \mathbf{z}) = n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \Big\{ & \left(e^{-|U_{ki}-U_{kj}|} + e^{-U_{ki}} + e^{U_{ki}-1} + e^{-U_{kj}} + e^{U_{kj}-1} + 2e^{-1} - 4\right) \\
& \left(e^{-|V_i - V_j|} + e^{-V_i} + e^{V_i - 1} + e^{-V_j} + e^{V_j-1} + 2e^{-1} - 4\right) e^{-\|\mathbf{w}_i - \mathbf{w}_j\|_1} \Big\}.
\end{aligned}$$

We have

$$\begin{aligned}
& \mathrm{pr}\left(\max_{k\in[p]} |\widehat{\rho}_k - \rho_k| \geq a n^{-\gamma}\right) \\
\leq & \mathrm{pr}\left(\max_{k\in[p]} |\widehat{\rho}_k - \widetilde{\rho}_k| > a n^{-\gamma}/2\right) + \mathrm{pr}\left(\max_{k\in[p]} |\widetilde{\rho}_k - \rho_k| \geq a n^{-\gamma}/2\right). \qquad \text{(S10)}
\end{aligned}$$

We decompose $\rho_k$ as

$$\rho_k = g_{k1} + g_{k2} + c_7 g_{k3} + g_{k4} + g_{k5} + c_7 g_{k6} + c_7 g_{k7} + c_7 g_{k8} + c_7^2 g_{k9},$$

where

$$g_{k1} = \mathbb{E}\left(e^{-|U_{k1}-U_{k2}|} e^{-|V_1 - V_2|} e^{-\|\mathbf{w}_1 - \mathbf{w}_2\|_1}\right),$$

$$
\begin{aligned}
g_{k2} &= \mathbb{E}\left[e^{-|U_{k1}-U_{k2}|}\left\{\sum_{i=1}^{2}\left(e^{-V_i}+e^{V_i-1}\right)\right\}e^{-\|\mathbf{w}_1-\mathbf{w}_2\|_1}\right],\\
g_{k3} &= \mathbb{E}\left(e^{-|U_{k1}-U_{k2}|}e^{-\|\mathbf{w}_1-\mathbf{w}_2\|_1}\right),\\
g_{k4} &= \mathbb{E}\left[\left\{\sum_{i=1}^{2}\left(e^{-U_{ki}}+e^{U_{ki}-1}\right)\right\}e^{-|V_1-V_2|}e^{-\|\mathbf{w}_1-\mathbf{w}_2\|_1}\right],\\
g_{k5} &= \mathbb{E}\left[\left\{\sum_{i=1}^{2}\left(e^{-U_{ki}}+e^{U_{ki}-1}\right)\right\}\left\{\sum_{i=1}^{2}\left(e^{-V_i}+e^{V_i-1}\right)\right\}e^{-\|\mathbf{w}_1-\mathbf{w}_2\|_1}\right],\\
g_{k6} &= \mathbb{E}\left[\left\{\sum_{i=1}^{2}\left(e^{-U_{ki}}+e^{U_{ki}-1}\right)\right\}e^{-\|\mathbf{w}_1-\mathbf{w}_2\|_1}\right],\\
g_{k7} &= \mathbb{E}\left(e^{-|V_1-V_2|}e^{-\|\mathbf{w}_1-\mathbf{w}_2\|_1}\right),\\
g_{k8} &= \mathbb{E}\left[\left\{\sum_{i=1}^{2}\left(e^{-V_i}+e^{V_i-1}\right)\right\}e^{-\|\mathbf{w}_1-\mathbf{w}_2\|_1}\right],\\
g_{k9} &= \mathbb{E}\left(e^{-\|\mathbf{w}_1-\mathbf{w}_2\|_1}\right)\\
c_7 &= 2e^{-1}-4.
\end{aligned}
$$

$\widetilde{\rho}_k$ can be correspondingly decomposed as

$$
\widetilde{\rho}_k = \widetilde{g}_{k1}+\widetilde{g}_{k2}+c_7\widetilde{g}_{k3}+\widetilde{g}_{k4}+\widetilde{g}_{k5}+c_7\widetilde{g}_{k6}+c_7\widetilde{g}_{k7}+c_7\widetilde{g}_{k8}+c_7^2\widetilde{g}_{k9},
$$

where

$$
\begin{aligned}
\widetilde{g}_{k1} &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(e^{-\left|U_{ki}-U_{kj}\right|}e^{-|V_i-V_j|}e^{-\|\mathbf{w}_i-\mathbf{w}_j\|_1}\right),\\
\widetilde{g}_{k2} &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{e^{-\left|U_{ki}-U_{kj}\right|}\left(e^{-V_i}+e^{V_i-1}+e^{-V_j}+e^{V_j-1}\right)e^{-\|\mathbf{w}_i-\mathbf{w}_j\|_1}\right\},\\
\widetilde{g}_{k3} &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(e^{-\left|U_{ki}-U_{kj}\right|}e^{-\|\mathbf{w}_i-\mathbf{w}_j\|_1}\right),\\
\widetilde{g}_{k4} &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{\left(e^{-U_{ki}}+e^{U_{ki}-1}+e^{-U_{kj}}+e^{U_{kj}-1}\right)e^{-|V_i-V_j|}e^{-\|\mathbf{w}_i-\mathbf{w}_j\|_1}\right\},\\
\widetilde{g}_{k5} &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{\left(e^{-U_{ki}}+e^{U_{ki}-1}+e^{-U_{kj}}+e^{U_{kj}-1}\right)\left(e^{-V_i}+e^{V_i-1}+e^{-V_j}+e^{V_j-1}\right)e^{-\|\mathbf{w}_i-\mathbf{w}_j\|_1}\right\},
\end{aligned}
$$

$$\widetilde{g}_{k6} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{\left(e^{-U_{ki}} + e^{U_{ki}-1} + e^{-U_{kj}} + e^{U_{kj}-1}\right)e^{-\|\mathbf{w}_i-\mathbf{w}_j\|_1}\right\},$$

$$\widetilde{g}_{k7} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(e^{-|V_i-V_j|}e^{-\|\mathbf{w}_i-\mathbf{w}_j\|_1}\right),$$

$$\widetilde{g}_{k8} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{\left(e^{-V_i} + e^{V_i-1} + e^{-V_j} + e^{V_j-1}\right)e^{-\|\mathbf{w}_i-\mathbf{w}_j\|_1}\right\},$$

$$\widetilde{g}_{k9} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(e^{-\|\mathbf{w}_i-\mathbf{w}_j\|_1}\right).$$

$\widehat{\rho}_k$ can be decomposed in the same way as

$$\widehat{\rho}_k = \widehat{g}_{k1} + \widehat{g}_{k2} + c_7\widehat{g}_{k3} + \widehat{g}_{k4} + \widehat{g}_{k5} + c_7\widehat{g}_{k6} + c_7\widehat{g}_{k7} + c_7\widehat{g}_{k8} + c_7^2\widehat{g}_{k9},$$

where

$$\widehat{g}_{k1} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(e^{-\left|\widehat{U}_{ki}-\widehat{U}_{kj}\right|}e^{-\left|\widehat{V}_i-\widehat{V}_j\right|}e^{-\|\widehat{\mathbf{w}}_i-\widehat{\mathbf{w}}_j\|_1}\right),$$

$$\widehat{g}_{k2} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{e^{-\left|\widehat{U}_{ki}-\widehat{U}_{kj}\right|}\left(e^{-\widehat{V}_i} + e^{\widehat{V}_i-1} + e^{-\widehat{V}_j} + e^{\widehat{V}_j-1}\right)e^{-\|\widehat{\mathbf{w}}_i-\widehat{\mathbf{w}}_j\|_1}\right\},$$

$$\widehat{g}_{k3} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(e^{-\left|\widehat{U}_{ki}-\widehat{U}_{kj}\right|}e^{-\|\widehat{\mathbf{w}}_i-\widehat{\mathbf{w}}_j\|_1}\right),$$

$$\widehat{g}_{k4} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{\left(e^{-\widehat{U}_{ki}} + e^{\widehat{U}_{ki}-1} + e^{-\widehat{U}_{kj}} + e^{\widehat{U}_{kj}-1}\right)e^{-\left|\widehat{V}_i-\widehat{V}_j\right|}e^{-\|\widehat{\mathbf{w}}_i-\widehat{\mathbf{w}}_j\|_1}\right\},$$

$$\widehat{g}_{k5} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{\left(e^{-\widehat{U}_{ki}} + e^{\widehat{U}_{ki}-1} + e^{-\widehat{U}_{kj}} + e^{\widehat{U}_{kj}-1}\right)\left(e^{-\widehat{V}_i} + e^{\widehat{V}_i-1} + e^{-\widehat{V}_j} + e^{\widehat{V}_j-1}\right)e^{-\|\widehat{\mathbf{w}}_i-\widehat{\mathbf{w}}_j\|_1}\right\},$$

$$\widehat{g}_{k6} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\left\{e^{-\widehat{U}_{ki}} + e^{\widehat{U}_{ki}-1} + e^{-\widehat{U}_{kj}} + e^{\widehat{U}_{kj}-1}\right\}e^{-\|\widehat{\mathbf{w}}_i-\widehat{\mathbf{w}}_j\|_1}\right],$$

$$\widehat{g}_{k7} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(e^{-\left|\widehat{V}_i-\widehat{V}_j\right|}e^{-\|\widehat{\mathbf{w}}_i-\widehat{\mathbf{w}}_j\|_1}\right),$$

$$\widehat{g}_{k8} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\left(e^{-\widehat{V}_i} + e^{\widehat{V}_i-1} + e^{-\widehat{V}_j} + e^{\widehat{V}_j-1}\right)e^{-\|\widehat{\mathbf{w}}_i-\widehat{\mathbf{w}}_j\|_1}\right],$$

$$\widehat{g}_{k9} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(e^{-\|\widehat{\mathbf{w}}_i-\widehat{\mathbf{w}}_j\|_1}\right).$$

The following proof consists of three steps:

**Step 1**. We deal with the first term of the RHS of (S10), i.e., we prove that

$$\mathrm{pr}\left(\max_{k\in[p]}|\widehat{\rho}_k-\widetilde{\rho}_k|\geq an^{-\gamma}\right)\leq cpn^{\widetilde{c}\{1-2\gamma-2\theta s\}}\exp\left\{-c'n^{1-2\gamma-2\theta s}\right\}$$

for some proper positive constants $c,\widetilde{c}$ and $c'$.

We first deal with $\widehat{g}_{k1}$ and $\widetilde{g}_{k1}$ with the goal to prove that

$$\mathrm{pr}\left(|\widehat{g}_{k1}-\widetilde{g}_{k1}|\geq an^{-\gamma}\right)\leq cpn^{\widetilde{c}\{1-2\gamma-2\theta s\}}\exp\left\{-c'n^{1-2\gamma-2\theta s}\right\}. \tag{S11}$$

By Lemma 6, for some positive constants $b_1', b_2'$ and $b_3'$, we have

$$\begin{aligned}\mathrm{pr}\left(\max_i\left|\widehat{U}_{ki}-U_{ki}\right|\geq an^{-\gamma}\right) &= \mathrm{pr}\left\{\max_i\left|\frac{\widehat{g}(\mathbf{z}_i,X_{ki};X_k)}{\widehat{f}_{\mathbf{z}}(\mathbf{z}_i)}-\frac{g(\mathbf{z}_i,X_{ki};X_k)}{f_{\mathbf{z}}(\mathbf{z}_i)}\right|\geq an^{-\gamma}\right\}\\ &\leq \mathrm{pr}\left\{\sup_{\mathbf{z},x}\left|\frac{\widehat{g}(\mathbf{z},x;X_k)}{\widehat{f}_{\mathbf{z}}(\mathbf{z})}-\frac{g(\mathbf{z},x;X_k)}{f_{\mathbf{z}}(\mathbf{z})}\right|\geq an^{-\gamma}\right\}\\ &\leq b_1'n^{b_2'(1-2\gamma-2\theta s)}\exp\left(-b_3'n^{1-2\gamma-2\theta s}\right).\end{aligned}$$

Then for some positive constants $b_1, b_2$ and $b_3$,

$$\begin{aligned}&\mathrm{pr}\left(\max_{i,j}\left|\left|\widehat{U}_{ki}-\widehat{U}_{kj}\right|-|U_{ki}-U_{kj}|\right|\geq an^{-\gamma}\right)\\ \leq\ &\mathrm{pr}\left(\max_i\left|\widehat{U}_{ki}-U_{ki}\right|\geq an^{-\gamma}/2\right)+\mathrm{pr}\left(\max_j\left|\widehat{U}_{kj}-U_{kj}\right|\geq an^{-\gamma}/2\right)\\ \leq\ &b_1n^{b_2(1-2\gamma-2\theta s)}\exp\left(-b_3n^{1-2\gamma-2\theta s}\right).\end{aligned}$$

Since $|e^{-x}-e^{-y}|\leq|x-y|$ for $x>0$ and $y>0$, we have

$$\begin{aligned}&\mathrm{pr}\left(\max_{i,j}\left|e^{\left|\widehat{U}_{ki}-\widehat{U}_{kj}\right|}-e^{\left|U_{ki}-U_{kj}\right|}\right|\geq an^{-\gamma}\right)\\ \leq\ &\mathrm{pr}\left(\max_{i,j}\left|\left|\widehat{U}_{ki}-\widehat{U}_{kj}\right|-|U_{ki}-U_{kj}|\right|\geq an^{-\gamma}\right)\\ \leq\ &b_1n^{b_2(1-2\gamma-2\theta s)}\exp\left(-b_3n^{1-2\gamma-2\theta s}\right).\end{aligned} \tag{S12}$$

Similarly, for some positive constants $b_4, b_5$ and $b_6$,

$$\begin{aligned}&\mathrm{pr}\left(\max_{i,j}\left|e^{-\left|\widehat{V}_i-\widehat{V}_j\right|}-e^{-|V_i-V_j|}\right|\geq an^{-\gamma}\right)\\ \leq\ &b_4n^{b_5\{1-2\gamma-2\theta s\}}\exp\left\{-b_6n^{1-2\gamma-2\theta s}\right\},\end{aligned} \tag{S13}$$

and for some positive constants $b_7, b_8$ and $b_9$,

$$\begin{aligned} & \mathrm{pr}\left(\max_{i,j}\left|e^{-\|\widehat{\mathbf{w}}_i-\widehat{\mathbf{w}}_j\|_1}-e^{-\|\mathbf{w}_i-\mathbf{w}_j\|_1}\right|\geq an^{-\gamma}\right) \\ \leq\ & b_7 n^{b_8\{1-2\gamma-2\theta(s-1)\}}\exp\left\{-b_9 n^{1-2\gamma-2\theta(s-1)}\right\}. \end{aligned} \tag{S14}$$

By Lemma 7, (S12), (S13) and (S14), for some positive constants $c_1, \widetilde{c}_1$ and $c_1'$, we have

$$\begin{aligned} & \mathrm{pr}\left(\max_{i,j}\left|e^{\left|\widehat{U}_{ki}-\widehat{U}_{kj}\right|}e^{-\left|\widehat{V}_i-\widehat{V}_j\right|}e^{-\|\widehat{\mathbf{w}}_i-\widehat{\mathbf{w}}_j\|_1}-e^{\left|U_{ki}-U_{kj}\right|}e^{-|V_i-V_j|}e^{-\|\mathbf{w}_i-\mathbf{w}_j\|_1}\right|\geq an^{-\gamma}\right) \\ \leq\ & c_1 n^{\widetilde{c}_1\{1-2\gamma-2\theta s\}}\exp\left\{-c_1' n^{1-2\gamma-2\theta s}\right\}. \end{aligned}$$

Now we have

$$\begin{aligned} & \mathrm{pr}(|\widehat{g}_{k1}-\widetilde{g}_{k1}|\geq an^{-\gamma}) \\ =\ & \mathrm{pr}\left\{\left|\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(e^{\left|\widehat{U}_{ki}-\widehat{U}_{kj}\right|}e^{-\left|\widehat{V}_i-\widehat{V}_j\right|}e^{-\|\widehat{\mathbf{w}}_i-\widehat{\mathbf{w}}_j\|_1}-e^{\left|U_{ki}-U_{kj}\right|}e^{-|V_i-V_j|}e^{-\|\mathbf{w}_i-\mathbf{w}_j\|_1}\right)\right|\geq an^{-\gamma}\right\} \\ \leq\ & \mathrm{pr}\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|e^{\left|\widehat{U}_{ki}-\widehat{U}_{kj}\right|}e^{-\left|\widehat{V}_i-\widehat{V}_j\right|}e^{-\|\widehat{\mathbf{w}}_i-\widehat{\mathbf{w}}_j\|_1}-e^{\left|U_{ki}-U_{kj}\right|}e^{-|V_i-V_j|}e^{-\|\mathbf{w}_i-\mathbf{w}_j\|_1}\right|\geq an^{-\gamma}\right) \\ \leq\ & \mathrm{pr}\left(\max_{i,j}\left|e^{\left|\widehat{U}_{ki}-\widehat{U}_{kj}\right|}e^{-\left|\widehat{V}_i-\widehat{V}_j\right|}e^{-\|\widehat{\mathbf{w}}_i-\widehat{\mathbf{w}}_j\|_1}-e^{\left|U_{ki}-U_{kj}\right|}e^{-|V_i-V_j|}e^{-\|\mathbf{w}_i-\mathbf{w}_j\|_1}\right|\geq an^{-\gamma}\right) \\ \leq\ & c_1 n^{\widetilde{c}_1\{1-2\gamma-2\theta s\}}\exp\left\{-c_1' n^{1-2\gamma-2\theta s}\right\}. \end{aligned}$$

Repeating the above scheme can give us results similar to (S11):

$$\begin{aligned} \mathrm{pr}(|\widehat{g}_{k2}-\widetilde{g}_{k2}|\geq an^{-\gamma}) &\leq c_2 n^{\widetilde{c}_2\{1-2\gamma-2\theta s\}}\exp\left\{-c_2' n^{1-2\gamma-2\theta s}\right\}, \\ \mathrm{pr}(|\widehat{g}_{k3}-\widetilde{g}_{k3}|\geq an^{-\gamma}) &\leq c_3 n^{\widetilde{c}_3\{1-2\gamma-2\theta s\}}\exp\left\{-c_3' n^{1-2\gamma-2\theta s}\right\}, \\ \mathrm{pr}(|\widehat{g}_{k4}-\widetilde{g}_{k4}|\geq an^{-\gamma}) &\leq c_4 n^{\widetilde{c}_4\{1-2\gamma-2\theta s\}}\exp\left\{-c_4' n^{1-2\gamma-2\theta s}\right\}, \\ \mathrm{pr}(|\widehat{g}_{k5}-\widetilde{g}_{k5}|\geq an^{-\gamma}) &\leq c_5 n^{\widetilde{c}_5\{1-2\gamma-2\theta s\}}\exp\left\{-c_5' n^{1-2\gamma-2\theta s}\right\}, \\ \mathrm{pr}(|\widehat{g}_{k6}-\widetilde{g}_{k6}|\geq an^{-\gamma}) &\leq c_6 n^{\widetilde{c}_6\{1-2\gamma-2\theta s\}}\exp\left\{-c_6' n^{1-2\gamma-2\theta s}\right\}, \\ \mathrm{pr}(|\widehat{g}_{k7}-\widetilde{g}_{k7}|\geq an^{-\gamma}) &\leq c_7 n^{\widetilde{c}_7\{1-2\gamma-2\theta s\}}\exp\left\{-c_7' n^{1-2\gamma-2\theta s}\right\}, \\ \mathrm{pr}(|\widehat{g}_{k8}-\widetilde{g}_{k8}|\geq an^{-\gamma}) &\leq c_8 n^{\widetilde{c}_8\{1-2\gamma-2\theta s\}}\exp\left\{-c_8' n^{1-2\gamma-2\theta s}\right\}, \\ \mathrm{pr}(|\widehat{g}_{k9}-\widetilde{g}_{k9}|\geq an^{-\gamma}) &\leq c_9 n^{\widetilde{c}_9\{1-2\gamma-2\theta(s-1)\}}\exp\left\{-c_9' n^{1-2\gamma-2\theta(s-1)}\right\}, \end{aligned}$$

where $\{c_i\}_{i=2}^9, \{\widetilde{c}_i\}_{i=2}^9$ and $\{c'_i\}_{i=2}^9$ are some positive constants. Then for some positive constants $c, \widetilde{c}$ and $c'$ we have

$$\begin{aligned}
& \mathrm{pr}(|\widehat{\rho}_k - \widetilde{\rho}_k| \geq an^{-\gamma}) \\
= \ & \mathrm{pr}\Big\{|(\widehat{g}_{k1} - \widetilde{g}_{k1}) - (\widehat{g}_{k2} - \widetilde{g}_{k2}) + c_5(\widehat{g}_{k3} - \widetilde{g}_{k3}) + (\widehat{g}_{k4} - \widetilde{g}_{k4}) - (\widehat{g}_{k5} - \widetilde{g}_{k5}) \\
& \quad + c_5(\widehat{g}_{k6} - \widetilde{g}_{k6}) + c_6(\widehat{g}_{k7} - \widetilde{g}_{k7}) - c_6(\widehat{g}_{k8} - \widetilde{g}_{k8}) + c_5c_6(\widehat{g}_{k9} - \widetilde{g}_{k9})| \geq an^{-\gamma}\Big\} \\
\leq \ & \mathrm{pr}\left(|\widehat{g}_{k1} - \widetilde{g}_{k1}| \geq \frac{a}{9}n^{-\gamma}\right) + \mathrm{pr}\left(|\widehat{g}_{k2} - \widetilde{g}_{k2}| \geq \frac{a}{9}n^{-\gamma}\right) + \mathrm{pr}\left(|\widehat{g}_{k3} - \widetilde{g}_{k3}| \geq \frac{a}{9c_5}n^{-\gamma}\right) \\
& + \mathrm{pr}\left(|\widehat{g}_{k4} - \widetilde{g}_{k4}| \geq \frac{a}{9}n^{-\gamma}\right) + \mathrm{pr}\left(|\widehat{g}_{k5} - \widetilde{g}_{k5}| \geq \frac{a}{9}n^{-\gamma}\right) + \mathrm{pr}\left(|\widehat{g}_{k6} - \widetilde{g}_{k6}| \geq \frac{a}{9c_5}n^{-\gamma}\right) \\
& + \mathrm{pr}\left(|\widehat{g}_{k7} - \widetilde{g}_{k7}| \geq \frac{a}{9c_6}n^{-\gamma}\right) + \mathrm{pr}\left(|\widehat{g}_{k8} - \widetilde{g}_{k8}| \geq \frac{a}{9c_6}n^{-\gamma}\right) + \mathrm{pr}\left(|\widehat{g}_{k9} - \widetilde{g}_{k9}| \geq \frac{a}{9c_5c_6}n^{-\gamma}\right) \\
\leq \ & cn^{\widetilde{c}\{1-2\gamma-2\theta s\}}\exp\left\{-c'n^{1-2\gamma-2\theta s}\right\},
\end{aligned}$$

and hence

$$\mathrm{pr}\left(\max_{k\in[p]} |\widehat{\rho}_k - \widetilde{\rho}_k| \geq an^{-\gamma}\right) \leq cpn^{\widetilde{c}\{1-2\gamma-2\theta s\}}\exp\left\{-c'n^{1-2\gamma-2\theta s}\right\} \tag{S15}$$

for some proper positive constants $a, c, \widetilde{c}$ and $c'$.

**Step 2**. We deal with the second term of the RHS of (S10), i.e., we prove that

$$\mathrm{pr}\left(\max_{k\in[p]} |\widetilde{\rho}_k - \rho_k| \geq an^{-\gamma}\right) \leq cp\exp\left(-bn^{1-2\gamma}\right)$$

for some proper $a > 0$, $b > 0$ and $c > 0$.

By Condition (C2), $U_k = F_{X_k|\mathbf{Z}}(X_k \mid \mathbf{Z})$, $V = F_{Y|\mathbf{Z}}(Y \mid \mathbf{Z})$, $W_1 = F_{Z_1}(Z_1), ..., W_s = F_{Z_s|Z_1,...,Z_{s-1}}(\cdot \mid Z_1, ..., Z_{s-1})$ are all one to one transformation, so instead of seeing $\rho_k$ as a function of $X_k, Y$ and $\mathbf{z}$, we consider it as a function of $U_k, V$ and $\mathbf{w}$.

We first deal with $\widetilde{g}_{k1}$. Define $\widetilde{g}^*_{k1} = (n(n-1))^{-1}\sum_{i\neq j}\left(e^{-\left|U_{ki}-U_{kj}\right|}e^{-|V_i-V_j|}e^{-\|\mathbf{w}_i-\mathbf{w}_j\|_1}\right)$, which is a U-statistic. $\widetilde{g}_{k1}$ can be rewritten as

$$\widetilde{g}_{k1} = \frac{1}{n^2}\left\{n(n-1)\widetilde{g}^*_{k1} + \sum_{i=1}^{n}\sum_{j=i}\left(e^{-\left|U_{ki}-U_{kj}\right|}e^{-|V_i-V_j|}e^{-\|\mathbf{w}_i-\mathbf{w}_j\|_1}\right)\right\} = \frac{n-1}{n}\widetilde{g}^*_{k1} + \frac{1}{n}.$$

For any given $\epsilon > 0$, there is a large enough $n$ such that $\epsilon \geq 2(1-g_{k1})/(n+1)$. Then, we have

$$\begin{aligned}
\mathrm{pr}\left(|\widetilde{g}_{k1} - g_{k1}| \geq \epsilon\right) &= \mathrm{pr}\left\{\left|\frac{n-1}{n}(\widetilde{g}^{*}_{k1} - g_{k1}) + \frac{1}{n}(1-g_{k1})\right| \geq \epsilon\right\} \\
&\leq \mathrm{pr}\left(\frac{n-1}{n}\left|\widetilde{g}^{*}_{k1} - g_{k1}\right| + \frac{1}{n}\left|1-g_{k1}\right| \geq \epsilon\right) \\
&\leq \mathrm{pr}\left(\left|\widetilde{g}^{*}_{k1} - g_{k1}\right| \geq \frac{\epsilon}{2}\right). \qquad \text{(S16)}
\end{aligned}$$

To prove the uniform consistency of $\widetilde{g}_{k1}$, it suffices to show the uniform consistency of $\widetilde{g}^{*}_{k1}$.

By Markov's inequality, for any $t > 0$,

$$\mathrm{pr}\left(\widetilde{g}^{*}_{k1} - g_{k1} \geq \epsilon\right) = \mathrm{pr}\left[\exp\left\{t(\widetilde{g}^{*}_{k1} - g_{k1})\right\} \geq \exp(t\epsilon)\right] \leq \exp(-t\epsilon)\exp(-tg_{k1})\,\mathbb{E}\left\{\exp(t\widetilde{g}^{*}_{k1})\right\}. \qquad \text{(S17)}$$

Denote the kernel of $\widetilde{g}^{*}_{k1}$ as $h_1(U_{ki}, V_i, \mathbf{w}_i; U_{kj}, V_j, \mathbf{w}_j) = e^{-\left|U_{ki}-U_{kj}\right|}e^{-|V_i-V_j|}e^{-\|\mathbf{w}_i-\mathbf{w}_j\|_1}$. Since U-statistic can be represented as an average of averages of i.i.d. random variables, we can rewrite $\widetilde{g}^{*}_{k1}$ as $\widetilde{g}^{*}_{k1} = (n!)^{-1}\sum_{n!}\Omega_1(U_{ki_1}, V_{i_1}, \mathbf{w}_{i_1}; ...; U_{ki_n}, V_{i_n}, \mathbf{w}_{i_n})$, where $\sum_{n!}$ denotes the summation over all $n!$ permutations $(i_1, ..., i_n)$ of $(1, ..., n)$, and each $\Omega_1(U_{ki_1}, V_{i_1}, \mathbf{w}_{i_1}; ...; U_{ki_n}, V_{i_n}, \mathbf{w}_{i_n})$ is an average of $m = \lfloor n/2 \rfloor$, i.e., $m^{-1}\sum_{r=1}^{m} h_1(U_{k,i_{2r-1}}, V_{i_{2r-1}}, \mathbf{w}_{i_{2r-1}}; U_{k,i_{2r}}, V_{i_{2r}}, \mathbf{w}_{i_{2r}})$. Denote $h_1(U_{k,i_{2r-1}}, V_{i_{2r-1}}, \mathbf{w}_{i_{2r-1}}; U_{k,i_{2r}}, V_{i_{2r}}, \mathbf{w}_{i_{2r}})$ as $h_1^{(r)}$. By Jensen's inequality,

$$\begin{aligned}
\mathbb{E}\left\{\exp(t\widetilde{g}^{*}_{k1})\right\} &= \mathbb{E}\left\{\exp\left(t\frac{1}{n!}\sum_{n!}\frac{1}{m}\sum_{r=1}^{m}h_1^{(r)}\right)\right\} \\
&\leq \frac{1}{n!}\sum_{n!}\mathbb{E}\left\{\exp\left(\frac{t}{m}\sum_{r=1}^{m}h_1^{(r)}\right)\right\} \\
&= \mathbb{E}^{m}\left\{\exp\left(\frac{t}{m}h_1^{(r)}\right)\right\} \qquad \text{(S18)}
\end{aligned}$$

Combining (S17) and (S18), since $h_1^{(r)} \in [e^{-2(s+2)}, 1]$, then by Lemma 2,

$$\begin{aligned}
\mathrm{pr}\left(\widetilde{g}^{*}_{k1} - g_{k1} \geq \epsilon\right) &\leq \exp(-t\epsilon)\,\mathbb{E}^{m}\left[\exp\left\{\frac{t}{m}\left(h_1^{(r)} - g_1\right)\right\}\right] \\
&\leq \exp(-t\epsilon)\left[\exp\left\{\left(\frac{t}{m}\right)^2\left(1-e^{-2(s+2)}\right)^2/8\right\}\right]^{m} \\
&\leq \exp\left\{-t\epsilon + t^2\frac{\left(1-e^{-2(s+2)}\right)^2}{8m}\right\}.
\end{aligned}$$

By Chernoff's method, when $t = 4m\epsilon \left(1 - e^{-2(s+2)}\right)^{-2}$, we have

$$\mathrm{pr}\left(\widetilde{g}^*_{k1} - g_{k1} \geq \epsilon\right) \leq \exp\left\{\frac{-2m\epsilon^2}{\left(1 - e^{-2(s+2)}\right)^2}\right\}.$$

Then by the symmetry of U-statistic,

$$\mathrm{pr}\left(|\widetilde{g}^*_{k1} - g_{k1}| \geq \epsilon\right) \leq 2\exp\left\{\frac{-2m\epsilon^2}{\left(1 - e^{-2(s+2)}\right)^2}\right\}.$$

Choose $\epsilon = an^{-\gamma}$ and some proper $b_1 > 0$ and $c > 0$, by (S16) we have

$$\mathrm{pr}\left(|\widetilde{g}_{k1} - g_{k1}| \geq an^{-\gamma}\right) \leq c\,\exp\left(-b_1 n^{1-2\gamma}\right). \tag{S19}$$

Similarly, we have results like (S19) for $\widetilde{g}_{k2}, ..., \widetilde{g}_{k9}$ that hold for some proper $b_l > 0$ and $c > 0$:

$$\mathrm{pr}\left(|\widetilde{g}_{kl} - g_{kl}| \geq an^{-\gamma}\right) \leq c\,\exp\left(-b_l n^{1-2\gamma}\right), \quad l = 2, ..., 9. \tag{S20}$$

Then we have

$$\begin{aligned}
& \mathrm{pr}\left(|\widetilde{\rho}_k - \rho_k| \geq an^{-\gamma}\right) \\
= \; & \mathrm{pr}\Big\{|(\widetilde{g}_{k1} - g_{k1}) - (\widetilde{g}_{k2} - g_{k2}) + c_5(\widetilde{g}_{k3} - g_{k3}) + (\widetilde{g}_{k4} - g_{k4}) - (\widetilde{g}_{k5} - g_{k5}) \\
& \quad + c_5(\widetilde{g}_{k6} - g_{k6}) + c_6(\widetilde{g}_{k7} - g_{k7}) - c_6(\widetilde{g}_{k8} - g_{k8}) + c_5c_6(\widetilde{g}_{k9} - g_{k9})| \geq an^{-\gamma}\Big\} \\
\leq \; & \mathrm{pr}\left(|\widetilde{g}_{k1} - g_{k1}| \geq \frac{a}{9}n^{-\gamma}\right) + \mathrm{pr}\left(|\widetilde{g}_{k2} - g_{k2}| \geq \frac{a}{9}n^{-\gamma}\right) + \mathrm{pr}\left(|\widetilde{g}_{k3} - g_{k3}| \geq \frac{a}{9c_5}n^{-\gamma}\right) \\
& + \mathrm{pr}\left(|\widetilde{g}_{k4} - g_{k4}| \geq \frac{a}{9}n^{-\gamma}\right) + \mathrm{pr}\left(|\widetilde{g}_{k5} - g_{k5}| \geq \frac{a}{9}n^{-\gamma}\right) + \mathrm{pr}\left(|\widetilde{g}_{k6} - g_{k6}| \geq \frac{a}{9c_5}n^{-\gamma}\right) \\
& + \mathrm{pr}\left(|\widetilde{g}_{k7} - g_{k7}| \geq \frac{a}{9c_6}n^{-\gamma}\right) + \mathrm{pr}\left(|\widetilde{g}_{k8} - g_{k8}| \geq \frac{a}{9c_6}n^{-\gamma}\right) + \mathrm{pr}\left(|\widetilde{g}_{k9} - g_{k9}| \geq \frac{a}{9c_5c_6}n^{-\gamma}\right) \\
\leq \; & c\,\exp\left(-bn^{1-2\gamma}\right)
\end{aligned}$$

and hence

$$\mathrm{pr}\left(\max_{k\in[p]} |\widetilde{\rho}_k - \rho_k| \geq an^{-\gamma}\right) \leq cp\,\exp\left(-bn^{1-2\gamma}\right) \tag{S21}$$

for some proper $a > 0$, $b > 0$ and $c > 0$.

**Step 3**. By (S15) and (S21), we have

$$\mathrm{pr}\left(\max_{k\in[p]} |\widehat{\rho}_k - \rho_k| \geq an^{-\gamma}\right) \leq \mathrm{pr}\left(\max_{k\in[p]} |\widehat{\rho}_k - \widetilde{\rho}_k| \geq an^{-\gamma}/2\right) + \mathrm{pr}\left(\max_{k\in[p]} |\widetilde{\rho}_k - \rho_k| \geq an^{-\gamma}/2\right)$$

$$\leq \quad cpn^{\widetilde{c}\{1-2\gamma-2\theta s\}}\exp\left\{-c'n^{1-2\gamma-2\theta s}\right\}$$

for some proper positive constants $a, c, \widetilde{c}$ and $c'$.

### S.2.3 Proof of Theorem 1(ii)

By assumption (10), $\mathcal{M} \nsubseteq \widehat{\mathcal{M}}$ implies that there exists some $j \in \mathcal{M}$ such that $\widehat{\rho}_j < an^{-\gamma}$, which means $|\widehat{\rho}_j - \rho_j| \geq an^{-\gamma}$. So, we have

$$\begin{aligned}
\mathrm{pr}\left(\mathcal{M} \subseteq \widehat{\mathcal{M}}\right) &\geq 1 - \mathrm{pr}\left(|\widehat{\rho}_j - \rho_j| \geq an^{-\gamma} \text{ for some } j \in \mathcal{M}\right) \\
&\geq 1 - |\mathcal{M}| \max_{j\in\mathcal{M}} \mathrm{pr}\left(|\widehat{\rho}_j - \rho_j| \geq an^{-\gamma}\right) \\
&\geq 1 - c_1 \left|\mathcal{M}\right| n^{c_2\{1-2\gamma-2\theta s\}}\exp\left\{-c_3 n^{1-2\gamma-2\theta s}\right\}
\end{aligned}$$

for some proper positive constants $c_1, c_2$ and $c_3$.

### S.2.4 Proof of Theorem 1(iii)

Under Condition (C5), there exists some $\delta_1 = \min_{j\in\mathcal{M}} \rho_j - \max_{j\in\mathcal{M}^c} \rho_j > 0$. Thus

$$\begin{aligned}
\mathrm{pr}\left(\min_{j\in\mathcal{M}} \widehat{\rho}_j \leq \max_{j\in\mathcal{M}^c} \widehat{\rho}_j\right) &= \mathrm{pr}\left(\min_{j\in\mathcal{M}} \widehat{\rho}_j - \min_{j\in\mathcal{M}} \rho_j + \delta_1 \leq \max_{j\in\mathcal{M}^c} \widehat{\rho}_j - \max_{j\in\mathcal{M}^c} \rho_j\right) \\
&= \mathrm{pr}\left(\max_{j\in\mathcal{M}^c} \widehat{\rho}_j - \max_{j\in\mathcal{M}^c} \rho_j - (\min_{j\in\mathcal{M}} \widehat{\rho}_j - \min_{j\in\mathcal{M}} \rho_j) \geq \delta_1\right) \\
&\leq \mathrm{pr}\left(\left|\max_{j\in\mathcal{M}^c} \widehat{\rho}_j - \max_{j\in\mathcal{M}^c} \rho_j + (\min_{j\in\mathcal{M}} \rho_j - \min_{j\in\mathcal{M}} \widehat{\rho}_j)\right| \geq \delta_1\right) \\
&\leq \mathrm{pr}\left(\left|\max_{j\in\mathcal{M}^c} \widehat{\rho}_j - \max_{j\in\mathcal{M}^c} \rho_j\right| \geq \delta_1/2\right) + \mathrm{pr}\left(\left|\min_{j\in\mathcal{M}} \widehat{\rho}_j - \min_{j\in\mathcal{M}} \rho_j\right| \geq \delta_1/2\right) \\
&\leq 2\mathrm{pr}\left(\max_{j\in[p]} |\widehat{\rho}_j - \rho_j| \geq \delta_1/2\right).
\end{aligned}$$

For some $a, \gamma$ and $\theta$ as defined in Theorem 1(iii), choose $\delta_1/2 = an^{-\gamma}$, we have

$$\mathrm{pr}\left(\max_{j\in\mathcal{M}^c} \widehat{\rho}_j < \min_{j\in\mathcal{M}} \widehat{\rho}_j\right) \geq 1 - c_4 pn^{c_5\{1-2\gamma-2\theta s\}}\exp\left\{-c_6 n^{1-2\gamma-2\theta s}\right\}$$

for some proper positive constants $c_4, c_5$ and $c_6$.

### S.2.5 Lemmas for Proof of Theorem 2

**Lemma 8.** *Let* $\{X_i\}_{i=1}^n$ *be identically distributed as Bernoulli*$(p)$*, and denote* $S_n = X_1 + \cdots + X_n$*. Then for any* $j \in \{1, ..., n\}$*, we have pr*$\{X_j = 1 \mid S_n\} = S_n/n$*.*

***Proof.*** For any $j \in \{1, ..., n\}$, since $\{X_i\}_{i=1}^n$ are identically distributed, we have

$$\begin{aligned} S_n &= \mathbb{E}(S_n \mid S_n) = \mathbb{E}\left(\sum_{i=1}^n X_i \mid S_n\right) = \sum_{i=1}^n \mathbb{E}(X_i \mid S_n) = n\mathbb{E}(X_j \mid S_n) \\ \implies \quad & \mathrm{pr}\{X_j = 1 \mid S_n\} = \mathbb{E}(X_j \mid S_n) = \frac{S_n}{n} \end{aligned}$$

□

**Lemma 9.** *Consider two independent sequences of random variables* $\{A_{1i}\}_{i=1}^\infty$ *and* $\{A_{2i}\}_{i=1}^\infty$ *that converge in distribution to the same random variable* $A$*. Then we have that sgn*$(A_{1i} - A_{2i})$ *converges in distribution to Bernoulli*$(1/2)$*.*

***Proof.*** Let $\psi_{1n}, \psi_{2n}$ and $\psi$ be the characteristic functions of $A_{1n}, A_{2n}$ and $A$. Then we have, for every $t$, $\psi_{1n}(t) \to \psi(t)$ and $\psi_{2n}(t) \to \psi(t)$ as $n \to \infty$. Define $A_- = A_{1n} - A_{2n}$ and $A^- = A_{2n} - A_{1n}$, then the characteristic function of $A_-$ is $\psi_{n-}(t) = \psi_{1n}(t)\psi_{2n}(-t) \to \psi(t)\psi(-t)$, and the characteristic function of $A^-$ is $\psi_n^-(t) = \psi_{2n}(t)\psi_{1n}(-t) \to \psi(-t)\psi(t)$. Define the random variable that has characteristic function $\psi(-t)\psi(t)$ as $B$.

Consider $\mathrm{pr}(A_- \geq 0)$ and $\mathrm{pr}(A^- > 0)$. We have $\mathrm{pr}(A_- \geq 0) + \mathrm{pr}(A^- > 0) = 1$, $\mathrm{pr}(A_- \geq 0) \to \mathrm{pr}(B \geq 0)$ and $\mathrm{pr}(A^- > 0) \to \mathrm{pr}(B \geq 0)$, so $\mathrm{pr}(A_- \geq 0) \to 1/2$ and $\mathrm{pr}(A^- > 0) \to 1/2$, hence $\mathrm{pr}\{\mathrm{sgn}(A_{1i} - A_{2i}) = +1\} \to 1/2$ and $\mathrm{pr}\{\mathrm{sgn}(A_{1i} - A_{2i}) = -1\} \to 1/2$. □

**Lemma 10.** *For any* $B_k$ *defined in Theorem 2, we have that* $B_k$ *converges in distribution to Bernoulli*$(1/2)$*.*

***Proof.*** By definition (11), the distribution of $B_k$ is the same as the distribution of $\mathrm{sgn}(n_1\widehat{\rho}_{k1} - n_2\widehat{\rho}_{k2})$. By Lemma 1.2, $n_1\widehat{\rho}_{k1}$ and $n_2\widehat{\rho}_{k2}$ has the same limiting distribution, so by Lemma 9, $B_k$ converges in distribution to Bernoulli(1/2). □

**Lemma 11.** *Denote the index set of $\mathcal{M}^c$ as $\{1, ..., p_0\}$. For $\{S_k\}_{k=1}^{p_0}$ defined in Theorem 2, under Condition (C6), we have $S_{p_0}/p_0 \xrightarrow{p} 1/2$ as $(n, p_0) \to \infty$.*

***Proof.*** By Markov's inequality, given the constant $\lambda$ defined in Condition (C6), we have

$$\begin{aligned} & \mathrm{pr}\left\{\left|\frac{S_{p_0}}{p_0}-\frac{1}{2}\right| \geq c_n^{-\frac{1}{2}}\right\} \leq c_n \mathbb{E}\left(\frac{S_{p_0}}{p_0}-\frac{1}{2}\right)^2 \\ = \; & c_n\left\{\mathbb{E}\left(\frac{S_{p_0}}{p_0}-\mathbb{E}\frac{S_{p_0}}{p_0}\right)^2+\mathbb{E}\left(\mathbb{E}\frac{S_{p_0}}{p_0}-\frac{1}{2}\right)^2+2\mathbb{E}\left[\left(\frac{S_{p_0}}{p_0}-\mathbb{E}\frac{S_{p_0}}{p_0}\right)\left(\mathbb{E}\frac{S_{p_0}}{p_0}-\frac{1}{2}\right)\right]\right\} \\ = \; & c_n\left(A_1+A_2+A_3\right) \\ = \; & c_n\left(A_1+A_2\right). \end{aligned} \tag{S22}$$

The last equation holds since $A_3 = 0$. We prove (S22) $\to 0$ as $(n, p_0) \to \infty$ in two steps:

**Step 1.** We prove that $c_n A_1 \to 0$ as $(n, p_0) \to 0$.

Let $\{B'_k\}_{k=1}^{p_0}$ be a copy of $\{B_k\}_{k=1}^{p_0}$, such that $B'_k \perp\!\!\!\perp B'_j$ for all $k \neq j$ and $B'_k \overset{iid}{\sim} B_k$. Denote $S'_{p_0} = B'_1 + \cdots + B'_{p_0}$, $\mathbb{E}$ takes expectation with respect to $\{B_k\}_{k=1}^{p_0}$ and $\mathbb{E}'$ takes expectation with respect to $\{B'_k\}_{k=1}^{p_0}$. Let $\{\varepsilon_k\}_{k=1}^{p_0}$ be i.i.d. random variables with $\mathrm{pr}(\varepsilon_k = \pm 1) = 1/2$ for $k = 1, \ldots, p_0$, i.e., a sequence with Rademacher distribution, and $\mathbb{E}_\varepsilon$ takes expectation with respect to $\{\varepsilon_k\}_{k=1}^{p_0}$. Then we have

$$\begin{aligned} c_n A_1 &= \frac{c_n}{p_0^2}\mathbb{E}\left(S_{p_0}-\mathbb{E}S_{p_0}\right)^2 = \frac{c_n}{p_0^2}\mathbb{E}\left(S_{p_0}-\mathbb{E}'S'_{p_0}\right)^2 = \frac{c_n}{p_0^2}\mathbb{E}\left(\mathbb{E}'\left(S_{p_0}-S'_{p_0}\right)\right)^2 \\ &\leq \frac{c_n}{p_0^2}\mathbb{E}\mathbb{E}'\left(S_{p_0}-S'_{p_0}\right)^2 && \text{(S23)} \\ &= \frac{c_n}{p_0^2}\mathbb{E}\mathbb{E}'\left(\sum_{k=1}^{p_0}(B_k-B'_k)\right)^2 \\ &= \frac{c_n}{p_0^2}\mathbb{E}\mathbb{E}'\mathbb{E}_\varepsilon\left(\sum_{k=1}^{p_0}\varepsilon_k(B_k-B'_k)\right)^2 && \text{(S24)} \\ &= \frac{c_n}{p_0^2}\mathbb{E}\mathbb{E}'\left\{\mathbb{E}_\varepsilon\left(\sum_{k=1}^{p_0}\varepsilon_k B_k\right)^2+\mathbb{E}_\varepsilon\left(\sum_{k=1}^{p_0}\varepsilon_k B'_k\right)^2-2\mathbb{E}_\varepsilon\left(\sum_{k=1}^{p_0}\varepsilon_k B_k\right)\left(\sum_{k=1}^{p_0}\varepsilon_k B'_k\right)\right\} \\ &\leq \frac{c_n}{p_0^2}\mathbb{E}\mathbb{E}'\left\{\mathbb{E}_\varepsilon\left(\sum_{k=1}^{p_0}\varepsilon_k B_k\right)^2+\mathbb{E}_\varepsilon\left(\sum_{k=1}^{p_0}\varepsilon_k B'_k\right)^2\right\} && \text{(S25)} \end{aligned}$$

$$
\begin{aligned}
&= 2\frac{c_n}{p_0^2}\mathbb{E}\mathbb{E}_\varepsilon\left(\sum_{k=1}^{p_0}\varepsilon_k B_k\right)^2 \\
&\leq 2\frac{c_n}{p_0^2}\mathbb{E}\left(\sum_{k=1}^{p_0} B_k^2\right) && \text{(S26)} \\
&= 2\frac{c_n}{p_0}\mathbb{E}B_k^2 \\
&\leq 2\frac{c_n}{p_0} && \text{(S27)} \\
&\to 0 \text{ as } (n, p_0) \to \infty. && \text{(S28)}
\end{aligned}
$$

(S23) is by Jensen's inequality. (S24) is because $B_k - B_k'$ and $\varepsilon_k(B_k - B_k')$ have the same distribution for $k = 1, ..., p_0$. (S25) is because

$$
\mathbb{E}_\varepsilon\left(\sum_{k=1}^{p_0}\varepsilon_k B_k\right)\left(\sum_{k=1}^{p_0}\varepsilon_k B_k'\right) = \mathbb{E}_\varepsilon\left(\sum_{k=1}^{p_0}\varepsilon_k^2 B_k B_k'\right) = \sum_{k=1}^{p_0} B_k B_k' \geq 0.
$$

(S26) is by Khintchine inequality. (S27) is because $B_k \in \{0, 1\}$. (S28) holds by Condition (C6.2).

**Step 2.** We prove that $c_n A_2 \to 0$ as $(n, p_0) \to 0$.

$$
c_n A_2 = c_n\left(\mathbb{E}\frac{S_{p_0}}{p_0} - \frac{1}{2}\right)^2 = c_n\left(\mathbb{E}B_k - \frac{1}{2}\right)^2 = c_n\left\{o\left(c_n^{-1}\right)\right\}^2 = o\left\{c_n^{-1}\right\}.
$$

The third equation holds by Condition (C6.1), and $c_n \to \infty$ as $n \to \infty$ gives $c_n A_2 \to 0$ as $(n, p_0) \to 0$.

By Step 1 and 2, we have (S22) $\to 0$ as $(n, p_0) \to 0$, thus $S_{p_0}/p_0 \xrightarrow{p} 1/2$ as $(n, p_0) \to \infty$. □

## S.3 Extensions to Discrete Data

In this section, we discuss how to extend the proposed method to the discrete data setting. To begin with, we assume that $X$, $Y$ and $Z$ are discrete random variables. Define $F_{X|Z}(x \mid z) = P(X \leq x \mid Z = z)$, $F_{X|Z}(x- \mid z) = P(X < x \mid Z = z)$, $F_{Y|Z}(y \mid z) = P(Y \leq y \mid Z = z)$, and $F_{Y|Z}(y- \mid z) = P(Y < y \mid Z = z)$. We further let $U_X$ and $U_Y$ be two independent and identically distributed $U(0, 1)$ random variables, and apply the transformations

$$
U = (1 - U_X)F_{X|Z}(X- \mid Z) + U_X F_{X|Z}(X \mid Z), \tag{S29}
$$

$$V \;=\; (1-U_Y)F_{Y|Z}(Y- \mid Z) + U_Y F_{Y|Z}(Y \mid Z). \tag{S30}$$

According to Brockwell (2007), both $U$ and $V$ are uniformly distributed on $(0,1)$. In addition, $U \perp\!\!\!\perp Z$ and $V \perp\!\!\!\perp Z$. In the following proposition, we establish the equivalence between conditional independence and mutual independence. The following proposition is adapted from Theorem 8 of Cai et al. (2022).

**Proposition 1.** *For discrete random variables $X$, $Y$ and $Z$, $X \perp\!\!\!\perp Y \mid Z$ if and only if $U, V$ and $Z$ are mutually independent.*

We summarize all the cases of $X$, $Y$, and $Z$ in the following table. One can also follow the proof of Theorem 8 in Cai et al. (2022) and show that $X \perp\!\!\!\perp Y \mid Z$ if and only if $U, V$ and $W$ are mutually independent.

| | | |
|---|---|---|
| $U$ | $X$ is continuous, $U = F_{X\|Z}(X \mid Z)$. | $X$ is discrete, $U = (1-U_X)F_{X\|Z}(X- \mid Z) + U_X F_{X\|Z}(X \mid Z)$. |
| $V$ | $Y$ is continuous, $V = F_{Y\|Z}(X \mid Z)$. | $Y$ is discrete, $V = (1-V_X)F_{Y\|Z}(Y- \mid Z) + U_Y F_{Y\|Z}(Y \mid Z)$. |
| $W$ | $Z$ is continuous, $W = F_Z(Z)$. | $Z$ is discrete, $W = Z$. |

Note that for discrete data, the conditional cumulative distribution function is a step-wise function that does not require kernel estimations. After estimating $\widehat{U}$, $\widehat{V}$, and $\widehat{W}$, we can estimate the marginal screening utility by $V$-statistics similar to (6) in the main paper.

# S.4 FDR Control Performance of ISIS-SCAD

Table S1 reports the average number of selected predictors, the selection probability of true variables, the empirical FDR and the F1 score of ISIS-SCAD for Example 3.

As shown in Table 1, although the ISIS-SCAD has a higher probability of choosing the true variables, this comes at the expense of a high FDR rate and low F1 score. It shows that the proposed CIS-REDS is a better choice if researchers want to control the false discoveries.

Table S1: FDR control of ISIS-SCAD for Example 3. Seven true active predictors $\{X_k\}_{k=3}^{9}$ are to be identified. $\left|\widehat{\mathcal{M}}\right|$ is the average number of selected predictors, $P_k, k = 3, ..., 9$ report the probability that the active predictor $X_k$ is selected, $P_a$ stands for the probability that all active predictors are selected.

| Model | $\left\|\widehat{\mathcal{M}}\right\|$ | $P_2$ | $P_3$ | $P_4$ | $P_5$ | $P_6$ | $P_7$ | $P_8$ | $P_a$ | $\widehat{\mathrm{FDR}}$ | F1 Score |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 7 | 37.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.811 | 0.318 |
| 8 | 37.000 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 1.000 | 0.990 | 0.990 | 0.812 | 0.318 |
| 9 | 37.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.811 | 0.318 |
| 10 | 37.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.811 | 0.318 |

# References

Brockwell, A. (2007). "Universal residuals: A multivariate transformation." *Statistics & Probability Letters*, **77(14)**, 1473–1478.

Cai, Z., Li, R., and Zhang, Y. (2022). "A distribution free conditional independence test with applications to causal discovery." *Journal of Machine Learning Research*. Accepted.

Chacón, J.E. and Duong, T. (2018). *Multivariate kernel smoothing and its applications*. CRC Press.

Dudley, R.M. (2014). *Uniform central limit theorems*. Cambridge Studies in Advanced Mathematics. Cambridge University Press.

Giné, E. and Guillou, A. (2002). "Rates of strong uniform consistency for multivariate kernel density estimators." *Annales de l'I.H.P. Probabilités et statistiques*, **38(6)**, 907–921.

Giné, E. and Guillou, A. (2001). "On consistency of kernel density estimators for randomly censored data: rates holding uniformly over adaptive intervals." *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, **37(4)**, 503 – 522.

Nolan, D. and Pollard, D. (1987). "$U$ processes: rates of convergence." *The Annals of Statistics*, **15(2)**, 780 – 799.

Pollard, D. (2012). *Convergence of stochastic processes.* Springer Science & Business Media.

van der Vaart, A.W. and Wellner, J. (1996). *Weak convergence and empirical processes: with applications to statistics.* Springer-Verlag New York.