Supplement to "Sliced Independence Test"

Yilin Zhang, Canyi Chen and Liping Zhu

Renmin University of China

Supplementary Material

In this Supplement Material, we provide proofs for Theorem 1-3 and zero-independence equivalence of equation (4.1) in the main context.

S.1. Proof of Theorem 1

For notation clarity, we re-define $s_i(X_{(h,j)}) \overset{\text{def}}{=} s(T_i; X_{(h,j)})$, $\varepsilon_i(X_{(h,j)}) \overset{\text{def}}{=} \varepsilon(T_i; X_{(h,j)})$ and $\nu_i(Y_j) \overset{\text{def}}{=} 1(Y_j \geq T_i) - E\{1(Y_j \geq T_i) \mid T_i\}$. Let E_T be the expectation taking over the subscript T given all other random variables.

We define $\Lambda \overset{\text{def}}{=} E\Big[\text{var}\{1(Y \geq T) \mid X, T\}\Big]$, $W \overset{\text{def}}{=} E\Big[\text{var}\{1(Y \geq T) \mid T\}\Big]$,

$$\widehat{\Lambda} \overset{\text{def}}{=} \{n^2(c-1)\}^{-1} \sum_{i=1}^{n} \sum_{h=1}^{H} \sum_{j<l}^{c} \{1(Y_{(h,j)} \geq T_i) - 1(Y_{(h,l)} \geq T_i)\}^2 \text{ and}$$

$$\widehat{W} \overset{\text{def}}{=} \{n^2(n-1)\}^{-1} \sum_{i=1}^{n} \sum_{j<l}^{n} \{1(Y_j \geq T_i) - 1(Y_l \geq T_i)\}^2, \text{ where } \sum_{j<l}^{c} \overset{\text{def}}{=} \sum_{1 \leq j<l \leq c}.$$

It follows that $\mathcal{S}(X,Y) = (W - \Lambda)/W$ and $\widehat{\mathcal{S}}(X,Y) = (\widehat{W} - \widehat{\Lambda})/\widehat{W}$. Thus,

$$n^{1/2}\{\widehat{\mathcal{S}}(X,Y) - \mathcal{S}(X,Y)\} = n^{1/2}\{(\widehat{W} - W)\Lambda - W(\widehat{\Lambda} - \Lambda)\}/(\widehat{W}W).$$

Here, $\widehat{W}$ is a two-sample $U-$statistic. By Theorem 12.6 of van der Vaart (1998), $\widehat{W} = W + O_p(n^{-1/2})$. Thus, $\widehat{W}$ converges to W in probability. By Slutsky's Lemma, it suffices to prove the asymptotic normality for $n^{1/2}\{(\widehat{W} - W)\Lambda - W(\widehat{\Lambda} - \Lambda)\}$.

Following Theorem 1 of Zhu and Ng (1995), we divide $\widehat{\Lambda}$ into three parts. Define

$$\begin{aligned}
I_1 &\stackrel{\text{def}}{=} \{n^2(c-1)\}^{-1}\sum_{i=1}^{n}\sum_{h=1}^{H}\sum_{j<l}^{c}\{s_i(X_{(h,j)}) - s_i(X_{(h,l)})\}^2,\\
I_2 &\stackrel{\text{def}}{=} \{n^2(c-1)\}^{-1}\sum_{i=1}^{n}\sum_{h=1}^{H}\sum_{j<l}^{c}\{s_i(X_{(h,j)}) - s_i(X_{(h,l)})\}\{\varepsilon_i(X_{(h,j)}) - \varepsilon_i(X_{(h,l)})\} \text{ and}\\
I_3 &\stackrel{\text{def}}{=} \{n^2(c-1)\}^{-1}\sum_{i=1}^{n}\sum_{h=1}^{H}\sum_{j<l}^{c}\{\varepsilon_i(X_{(h,j)}) - \varepsilon_i(X_{(h,l)})\}^2.
\end{aligned}$$

We have $\widehat{\Lambda}$ equals to

$$\begin{aligned}
&\{n^2(c-1)\}^{-1}\sum_{i=1}^{n}\sum_{h=1}^{H}\sum_{j<l}^{c}\left[\left\{s_i(X_{(h,j)}) + \varepsilon_i(X_{(h,j)})\right\} - \left\{s_i(X_{(h,l)}) + \varepsilon_i(X_{(h,l)})\right\}\right]^2\\
= &\{n^2(c-1)\}^{-1}\sum_{i=1}^{n}\sum_{h=1}^{H}\sum_{j<l}^{c}\Bigg[\left\{s_i(X_{(h,j)}) - s_i(X_{(h,l)})\right\}^2\\
&+ 2\left\{s_i(X_{(h,j)}) - s_i(X_{(h,l)})\right\}\left\{\varepsilon_i(X_{(h,j)}) - \varepsilon_i(X_{(h,l)})\right\} + \left\{\varepsilon_i(X_{(h,j)}) - \varepsilon_i(X_{(h,l)})\right\}^2\Bigg]\\
= &I_1 + 2I_2 + I_3,
\end{aligned}$$

It follows immediately that $n^{1/2}\{\Lambda(\widehat{W}-W)-W(\widehat{\Lambda}-\Lambda)\} = n^{1/2}\{\Lambda(\widehat{W}-W)-W(I_3-\Lambda)\} - n^{1/2}WI_1 - 2n^{1/2}WI_2$. We shall prove the first term is asymptotically normal and the other two quantities converge in probability to 0.

In what follows, we will prove the asymptotic normality for $n^{1/2}\{\Lambda(\widehat{W}-W)-W(I_3-\Lambda)\}$ in three steps. We first decompose it as

$$\begin{aligned} & n^{1/2}\{\Lambda(\widehat{W}-W)-W(I_3-\Lambda)\} \\ = \; & n^{1/2}\{\Lambda(\widehat{W}_1-\widehat{W}_2-W)-W(I_4-I_5-\Lambda)\}. \end{aligned} \tag{S.1.1}$$

In the above display,

$$\begin{aligned} \widehat{W}_1 &\stackrel{\text{def}}{=} \{n^2(n-1)\}^{-1}\sum_{i=1}^{n}\sum_{j<l}^{n}\{\nu_i^2(Y_j)+\nu_i^2(Y_l)\}, \\ \widehat{W}_2 &\stackrel{\text{def}}{=} \{n^2(n-1)\}^{-1}\sum_{i=1}^{n}\sum_{j\neq l}^{n}\nu_i(Y_j)\nu_i(Y_l), \\ I_4 &\stackrel{\text{def}}{=} \{n^2(c-1)\}^{-1}\sum_{i=1}^{n}\sum_{h=1}^{H}\sum_{j<l}^{c}\{\varepsilon_i^2(X_{(h,j)})+\varepsilon_i^2(X_{(h,l)})\} \text{ and} \\ I_5 &\stackrel{\text{def}}{=} \{n^2(c-1)\}^{-1}\sum_{i=1}^{n}\sum_{h=1}^{H}\sum_{j\neq l}^{c}\varepsilon_i(X_{(h,j)})\varepsilon_i(X_{(h,l)}). \end{aligned}$$

The equation (S.1.1) can be further divided into three parts as follows,

$$n^{1/2}\{\Lambda(\widehat{W}_1-W)-W(I_4-\Lambda)\} + n^{1/2}WI_5 - n^{1/2}\Lambda\widehat{W}_2.$$

Step 1. We derive the asymptotic normality for $n^{1/2}\{\Lambda(\widehat{W}_1-W)-W(I_4-$

$\Lambda)\}$.

$$\Lambda\widehat{W}_1 - WI_4 = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\{\Lambda\nu_i^2(Y_j) - W\varepsilon_i^2(X_j)\}.$$

which is a two-sample $U-$statistic. Denote

$$\begin{aligned}\zeta_1 \stackrel{\text{def}}{=} & \operatorname{var}\Big[E\{\Lambda W^{-1}\nu_i^2(Y_j) - \varepsilon_i^2(X_j) \mid (X_j, Y_j)\}\Big] \\ & +\operatorname{var}\Big[E\{\Lambda W^{-1}\nu_i^2(Y_j) - \varepsilon_i^2(X_j) \mid T_i\}\Big]. \quad \text{(S.1.2)}\end{aligned}$$

Apparently, $E\{\Lambda W^{-1}\nu_i^2(Y_j) - \varepsilon_i^2(X_j)\}^2 < \infty$. By Theorem 12.6 of van der Vaart (1998), $n^{1/2}\{\Lambda(\widehat{W}_1 - W) - W(I_4 - \Lambda)\} \xrightarrow{d} \mathcal{N}(0, W^2\zeta_1)$, as $n \to \infty$. If X and Y are independent, $\Lambda\nu_i^2(Y_j) - W\varepsilon_i^2(X_j) = 0$ and hence $n^{1/2}\{\Lambda(\widehat{W}_1 - W) - W(I_4 - \Lambda)\} = 0$.

Step 2. We prove the asymptotic normality for $n^{1/2}WI_5$. I_5 equals to

$$\begin{aligned} & \{n^2(c-1)\}^{-1}\sum_{h=1}^{H}\sum_{j\neq l}^{c}\sum_{i=1}^{n}[\varepsilon_i(X_{(h,j)})\varepsilon_i(X_{(h,l)}) - E_{T_i}\{\varepsilon_i(X_{(h,j)})\varepsilon_i(X_{(h,l)})\}] \\ & +\{n(c-1)\}^{-1}\sum_{h=1}^{H}\sum_{j\neq l}^{c} E_{T_i}\{\varepsilon_i(X_{(h,j)})\varepsilon_i(X_{(h,l)})\} \\ = & \{n(c-1)\}^{-1}\sum_{h=1}^{H}\sum_{j\neq l}^{c} E_{T_i}\{\varepsilon_i(X_{(h,j)})\varepsilon_i(X_{(h,l)})\} + O_p(n^{-1}), \quad \text{(S.1.3)}\end{aligned}$$

where the last equality is established by Chebyshev's inequality and the fact that

$$
\begin{aligned}
& E\left[\{n^2(c-1)\}^{-1}\sum_{h=1}^{H}\sum_{j\neq l}^{c}\sum_{i=1}^{n}\varepsilon_i(X_{(h,j)})\varepsilon_i(X_{(h,l)}) - E_{T_i}\{\varepsilon_i(X_{(h,j)})\varepsilon_i(X_{(h,l)})\}\big)\right]^2 \\
= & \{n^4(c-1)^2\}^{-1}\sum_{i=1}^{n}\sum_{h=1}^{H}\sum_{j\neq l}^{c} E\left(2[\varepsilon_i(X_{(h,j)})\varepsilon_i(X_{(h,l)}) - E_{T_i}\{\varepsilon_i(X_{(h,j)})\varepsilon_i(X_{(h,l)})\}]^2\right) \\
= & O(n^{-2}).
\end{aligned}
$$

The first equation of above display is because of the conditional un-correlation, and the second is by the boundedness of $\varepsilon_i(X_{(d,p)})$.

Parallel to the arguments used in Hsing and Carroll (1992, Theorem 2.3) and Zhu and Ng (1995, Theorem 1), we can prove that, using Central Limit Theorem, $n^{1/2}I_5$ converges to normal distribution, given the fact that $\sum_{j\neq l}^{c} E_{T_i}\{\varepsilon_i(X_{(h,j)})\varepsilon_i(X_{(h,l)})\}$ for $h = 1, \ldots, H$ are conditional independent crossing slice. The mean of the normal distribution is 0 and the asymptotic variance can be calculated as follows.

$$
\begin{aligned}
& nE\left[\{n(c-1)\}^{-1}\sum_{h=1}^{H}\sum_{j\neq l}^{c}E_{T_i}\{\varepsilon_i(X_{(h,j)})\varepsilon_i(X_{(h,l)})\}\right]^2 \\
= \;& \{n(c-1)^2\}^{-1}\sum_{h=1}^{H}\sum_{j\neq l}^{c}\sum_{d=1}^{H}\sum_{p\neq q}^{c}E\left[E_{T_i}\{\varepsilon_i(X_{(h,j)})\varepsilon_i(X_{(h,l)})\}E_{T_i}\{\varepsilon_i(X_{(d,p)})\varepsilon_i(X_{(d,q)})\}\right] \\
= \;& \{n(c-1)^2\}^{-1}\sum_{h=1}^{H}\sum_{j\neq l}^{c}E\left[2E_{T_i}\{\varepsilon_i(X_{(h,j)})\varepsilon_i(X_{(h,l)})\}^2\right] \\
= \;& 2\{n(c-1)^2\}^{-1}\sum_{h=1}^{H}\sum_{j\neq l}^{c}E\left[E\{\varepsilon_i(X_{(h,j)})\varepsilon_i(X_{(h,l)})\varepsilon_k(X_{(h,j)})\varepsilon_k(X_{(h,l)})\mid T_i,T_k\}\right] \\
= \;& 2\{n(c-1)^2\}^{-1}\sum_{h=1}^{H}\sum_{j\neq l}^{c}E\{V(T_i,T_k;X_{(h,j)})V(T_i,T_k;X_{(h,l)})\}, \qquad \text{(S.1.4)}
\end{aligned}
$$

where the second equation is because the expectation does not equal to zero only when $(h,j)=(d,p),(h,l)=(d,q)$ or $(h,j)=(d,q),(h,l)=(d,p)$.

We decompose $E\{V(T_i,T_k;X_{(h,j)})V(T_i,T_k;X_{(h,l)})\}=E\{V(T_i,T_k;X_{(h,j)})^2\}-E\Big[V(T_i,T_k;X_{(h,j)})\{V(T_i,T_k;X_{(h,l)})-V(T_i,T_k;X_{(h,j)})\}\Big]$, Thus, by Lemma 4, the above equality equals to

$$
\begin{aligned}
& 2\{n(c-1)\}^{-1}\sum_{h=1}^{H}\sum_{j=1}^{c}E\{V(T_i,T_k;X_{(h,j)})^2\} \\
& -2\{n(c-1)^2\}^{-1}\sum_{h=1}^{H}\sum_{j\neq l}^{c}E[V(T_i,T_k;X_{(h,j)})\{V(T_i,T_k;X_{(h,l)})-V(T_i,T_k;X_{(h,j)})\}] \\
= \;& 2(c-1)^{-1}E\{V(T_1,T_2;X)^2\}+o(1).
\end{aligned}
$$

Therefore, $n^{1/2}WI_5\xrightarrow{d}\mathcal{N}(0,2(c-1)^{-1}W^2E\{V(T_1,T_2;X)^2\})$.

If X and Y are independent, the asympototic variance can be further simplified.

$$
\begin{aligned}
& nE\left[\{n(c-1)\}^{-1}\sum_{h=1}^{H}\sum_{j\neq l}^{c} E_{T_i}\{\varepsilon_i(X_{(h,j)})\varepsilon_i(X_{(h,l)})\}\right]^2 \\
= \; & 2\{n(c-1)^2\}^{-1}\sum_{h=1}^{H}\sum_{j\neq l}^{c} E\Big(\big[E_{T_i}\{\nu_i(Y_{(h,j)})\nu_i(Y_{(h,l)})\}\big]^2\Big) \\
= \; & 2(c-1)^{-1}E\Big(\big[E_{T_i}\{\nu_i(Y_1)\nu_i(Y_2)\}\big]^2\Big). \qquad \text{(S.1.5)}
\end{aligned}
$$

In particular, if Y is continuous, the above equation can be further simplified to $2(c-1)^{-1}E\Big[F\{\min(Y_1,Y_2)\} - F(Y_1) - F(Y_2) + F(Y_2)^2/2 + F(Y_2)^2/2 + 1/3\Big]^2 = \{45(c-1)\}^{-1}$, where $F(\cdot)$ is the distribution function of Y. Thus,

$$
n^{1/2}WI_5 \xrightarrow{d} \mathcal{N}\left(0, 2(c-1)^{-1}W^2E\Big[E_{T_i}\{\nu_i(Y_1)\nu_i(Y_2)\}\Big]^2\right).
$$

When Y is continuous, the asymptotic variance reduces to $W^2/\{45(c-1)\}$.

Step 3. We combine $n^{1/2}\{\Lambda(\widehat{W}_1 - W) - W(I_4 - \Lambda)\}$, $n^{1/2}WI_5$ and $n^{1/2}\Lambda\widehat{W}_2$ to derive the asymptotic normality for $n^{1/2}\{\Lambda(\widehat{W} - W) - W(I_3 - \Lambda)\}$. In the previous steps, we show that $n^{1/2}\{\Lambda(\widehat{W}_1 - W) - W(I_4 - \Lambda)\}$ and $n^{1/2}WI_5$ are asymptotically normal. Thus, their covariance can be decomposed that $n\cdot\mathrm{cov}(\Lambda\widehat{W}_1 - WI_4, I_5) = n\Lambda\cdot\mathrm{cov}(\widehat{W}_1 - W, I_5) - nW\cdot\mathrm{cov}(I_4 - \Lambda, I_5)$. We have

$$
\begin{aligned}
& nW\cdot\mathrm{cov}(I_4 - \Lambda, I_5) \\
= \; & W\{n^2(c-1)\}^{-1}\sum_{p=1}^{n}\sum_{q=1}^{n}\sum_{h=1}^{H}\sum_{j\neq l}^{c} E[\varepsilon_i(X_{(h,j)})\varepsilon_i(X_{(h,l)})\{\varepsilon_p(X_q)^2 - \Lambda\}] = 0,
\end{aligned}
$$

where the last equation is due to the conditional uncorrelation. We also have

$$
\begin{aligned}
& n\Lambda \cdot \operatorname{cov}(\widehat{W}_1 - W, I_5) \\
= \quad & \Lambda\{n^2(c-1)\}^{-1} \sum_{p=1}^{n} \sum_{q=1}^{n} \sum_{h=1}^{H} \sum_{j \neq l}^{c} E\left[\{\nu_p^2(Y_q) - W\}\varepsilon_i(X_{(h,j)})\varepsilon_i(X_{(h,l)})\right] = 0.
\end{aligned}
$$

Thus, $n \cdot \operatorname{cov}(\Lambda\widehat{W}_1 - WI_4, I_5) = 0$. Note that

$$
\begin{aligned}
& E\left[\{n^2(n-1)\}^{-1} \sum_{i=1}^{n} \sum_{j \neq l}^{n} \nu_i(Y_j)\nu_i(Y_l)\right]^2 \\
& = \{n^4(n-1)^2\}^{-1} \sum_{i=1}^{n} \sum_{r=1}^{n} \sum_{j \neq l}^{n} 2E\{\nu_i(Y_j)\nu_i(Y_l)\nu_r(Y_j)\nu_r(Y_l)\} = O_p(n^{-2}).
\end{aligned}
$$

This, together with Chebyshev's inequality, gives that $\widehat{W}_2 = O_p(n^{-1})$.

Therefore, using Delta's method, we have

$$
n^{1/2}\{\Lambda(\widehat{W} - W) - W(I_3 - \Lambda)\} \xrightarrow{d} \mathcal{N}\left(0, W^2\left[\zeta_1 + 2(c-1)^{-1}E\{V(X, T_1, T_2)^2\}\right]\right).
$$

If X and Y are independent and Y is continuous, the asymptotic variance reduces to $\{45(c-1)\}^{-1}W^2$.

Noting that $s_i(X_{(h,j)}) - s_i(X_{(h,l)}) = 0$ for any h, j and l under the case that X and Y are independent, we have $n^{1/2}\widehat{\mathcal{S}}(X, Y) = n^{1/2}\{\Lambda(\widehat{W} - W) - W(I_3 - \Lambda)\}/(\widehat{W}W)$. Given the fact that $W = 1/6$ under independence, the proof for Theorem 1 (i) is accomplished directly using Slutsky's Lemma.

When X and Y are dependent, under condition (C1), combining Lemmas 3 and 2, we have $I_1 = o_p(n^{-1/2})$ and $I_2 = o_p(n^{-1/2})$. Thus, $n^{1/2}\{\widehat{\mathcal{S}}(X, Y) -$

$\mathcal{S}(X,Y)\} = n^{1/2}\{\Lambda(\widehat{W}-W)-W(I_3-\Lambda)\}/(\widehat{W}W)+o_p(1)$, which converges to normal distribution with mean 0 and variance $[\zeta_1+2(c-1)^{-1}E\{V(X,T_1,T_2)^2\}]/W^2$.

□

S.2. Proof of Theorem 2

This proof follows almost the same path as that used to prove Theorem 1. We firstly prove Theorem 2 (i), the case that X and Y are independent. By the same arguments of Theorem 1, we have $I_1 = 0$, $I_2 = 0$, $\Lambda\widehat{W}_1 - WI_4 = 0$ and $\widehat{W}_2 = O_p(n^{-1})$. As for I_5, given $c = o(n)$, from (S.1.3), we can derive that

$$(nc)^{1/2}I_5 \quad = \quad H^{-1/2}\sum_{h=1}^{H}\left[(c-1)^{-1}\sum_{j\neq l}^{c}E_{T_i}\{\varepsilon_i(X_{(h,j)})\varepsilon_i(X_{(h,l)})\}\right] + o_p(1).$$

As H goes to infinity, applying the central limit theorem for triangular arrays derived in Hsing and Carroll (1992, Theorem A.4) to I_5, $(nc)^{1/2}I_5$ converges to normal distribution with mean 0 an variance derived from (S.1.5) as

$$\lim_{c\to\infty} E\{(nc)^{1/2}I_5\}^2 = \lim_{c\to\infty} c\{45(c-1)\}^{-1} = 45^{-1}.$$

The Lyapunov condition can be verified via direct calculation for the boundness of item. Using Slutsky's Lemma, we complete the proof for Theorem 2 (i).

When X and Y are dependent, following the proof of Lemma 3 and 2 , under condition (C1*), it can be similarly prove that $n^{1/2}I_1$ and $n^{1/2}I_2$ both equal to $o_p(cn^{-1/2+\max\{r,1/(2+b)\}})$. Under Condition (C2*), $n^{1/2}I_1$ and $n^{1/2}I_2$ are both $o_p(1)$.

We can also prove that $n^{1/2}I_5 = o_p(1)$, given the following fact. As c goes to infinity, from equation (S.1.4), we can derive that the asymptotic variance $nE(I_5^2)$ can be calculated as follows,

$$
\begin{aligned}
& 2\{n(c-1)^2\}^{-1}\sum_{h=1}^{H}\sum_{j\neq l}^{c} E\{\varepsilon_i(X_{(h,j)})\varepsilon_i(X_{(h,l)})\varepsilon_k(X_{(h,j)})\varepsilon_k(X_{(h,l)})\} + O(n^{-1/2}) \\
\leq\ & 2\{n(c-1)^2\}^{-1}\sum_{h=1}^{H}\sum_{j\neq l}^{c} 1 + O(n^{-1/2}) \\
\leq\ & 2/(c-1) + O(n^{-1/2}) \to 0.
\end{aligned}
$$

Other terms are the same as in Theorem 1, we have

$$
n^{1/2}\{\Lambda(\widehat{W}-W) - W(\widehat{\Lambda}-\Lambda)\} = n^{1/2}\{\Lambda(\widehat{W}_1 - W) - W(I_4-\Lambda)\} + o_p(1),
$$

which converges to normal distribution with mean 0 and variance $W^2\zeta_1$, which has been proved in Step 1 of Theorem 1. Using Slutsky's Lemma, we complete the proof of Theorem 2 (ii). □

S.3. Proof of zero-independence equivalence of (4.1)

The proof of zero-independence equivalence in the multivariate case is in spirit exactly the same as that in the univariate one expect for X replaced with $\mathbf{x}$. The proof in the univariate case was given in Lemma A.1, page 1 of the Supplement to Chatterjee (2020). We omitted the details here. $\square$

S.4. Proof of Theorem 3

Denote

$$\widetilde{\Lambda} \stackrel{\text{def}}{=} n^{-2}\sum_{i=1}^{n}\sum_{h=1}^{H}\sum_{j<l}^{n_h}\{1(Y_{(h,j)} \geq T_i) - 1(Y_{(h,l)} \geq T_i)\}^2/(n_h - 1).$$

With this notation, $\widehat{\mathcal{S}}(\mathbf{x}, Y) - \mathcal{S}(\mathbf{x}, Y) = \{(\widehat{W} - W)\Lambda - W(\widetilde{\Lambda} - \Lambda)\}/(\widehat{W}W)$. Following similar arguments in the proof of Theorem 1, we divide $\widetilde{\Lambda}$ into several parts that, in symbols, $\widetilde{\Lambda} = \widetilde{I}_1 + 2\widetilde{I}_2 + \widetilde{I}_3$, where

$$\begin{aligned}
\widetilde{I}_1 &\stackrel{\text{def}}{=} n^{-2}\sum_{i=1}^{n}\sum_{h=1}^{H}\sum_{j<l}^{n_h}(n_h - 1)^{-1}\{s_i(\mathbf{x}_{(h,j)}) - s_i(\mathbf{x}_{(h,l)})\}^2, \\
\widetilde{I}_2 &\stackrel{\text{def}}{=} n^{-2}\sum_{i=1}^{n}\sum_{h=1}^{H}\sum_{j<l}^{n_h}(n_h - 1)^{-1}\{s_i(\mathbf{x}_{(h,j)}) - s_i(\mathbf{x}_{(h,l)})\}\{\varepsilon_i(\mathbf{x}_{(h,j)}) - \varepsilon_i(\mathbf{x}_{(h,l)})\} \text{ and} \\
\widetilde{I}_3 &\stackrel{\text{def}}{=} n^{-2}\sum_{i=1}^{n}\sum_{h=1}^{H}\sum_{j<l}^{n_h}(n_h - 1)^{-1}\{\varepsilon_i(\mathbf{x}_{(h,j)}) - \varepsilon_i(\mathbf{x}_{(h,l)})\}^2.
\end{aligned}$$

To deal with $\widetilde{I}_3$, we decompose it as two parts that $\widetilde{I}_3 = \widetilde{I}_4 - \widetilde{I}_5$, where

$$\begin{aligned}
\widetilde{I}_4 &\stackrel{\text{def}}{=} n^{-2}\sum_{i=1}^{n}\sum_{h=1}^{H}\sum_{j<l}^{n_h}\{\varepsilon_i^2(\mathbf{x}_{(h,j)}) + \varepsilon_i^2(\mathbf{x}_{(h,l)})\}/(n_h - 1) \text{ and} \\
\widetilde{I}_5 &\stackrel{\text{def}}{=} n^{-2}\sum_{i=1}^{n}\sum_{h=1}^{H}\sum_{j\neq l}^{n_h}\varepsilon_i(\mathbf{x}_{(h,j)})\varepsilon_i(\mathbf{x}_{(h,l)})/(n_h - 1).
\end{aligned}$$

Firstly, we consider the case when $\mathbf{x}$ and Y are independent. Under this case, we have $\widetilde{I}_1 = \widetilde{I}_2 = 0$,

$$\begin{aligned}
\widetilde{I}_4 &\stackrel{\text{def}}{=} n^{-2}\sum_{i=1}^{n}\sum_{h=1}^{H}\sum_{j=1}^{n_h}\nu_i^2(Y_{(h,j)}) \text{ and} \\
\widetilde{I}_5 &\stackrel{\text{def}}{=} 2n^{-2}\sum_{i=1}^{n}\sum_{h=1}^{H}\sum_{j<l}^{n_h}\nu_i(Y_{(h,j)})\nu_i(Y_{(h,l)})/(n_h-1).
\end{aligned}$$

Thus, we have $(nc_n)^{1/2}\{(\widehat{W}-W)\Lambda - W(\widetilde{\Lambda}-\Lambda)\} = (nc_n)^{1/2}\{(\widehat{W}_1+\widehat{W}_2)\Lambda - W(\widetilde{I}_4-\widetilde{I}_5)\}$. Given the fact that the sum of ordered observations is equal to that of the non-ordered, we have $\widetilde{I}_4 = \widehat{W}_1$. $\widehat{W}_2 = O_p(n^{-1})$ has been shown in the proof of Theorem 1. We also have $W = \Lambda$ under independence. Thus, $(nc_n)^{1/2}\{(\widehat{W}-W)\Lambda - W(\widetilde{\Lambda}-\Lambda)\} = (nc_n)^{1/2}W\widetilde{I}_5 + o_p(1)$. Define

$$\widetilde{I}_{5,1} \stackrel{\text{def}}{=} 2n^{-1}\sum_{h=1}^{H}\sum_{j<l}^{n_h} E_{T_i}\{\nu_i(Y_{(h,j)})\nu_i(Y_{(h,l)})\}/(n_h-1).$$

Following (S.1.3), we can further prove that

$$(nc_n)^{1/2}\{(\widehat{W}-W)\Lambda - W(\widetilde{\Lambda}-\Lambda)\} = (nc_n)^{1/2}W\widetilde{I}_{5,1} + o_p(1).$$

If $\mathbf{x}$ and Y are independent, $\widetilde{I}_{5,1}$ and

$$2n^{-1}\sum_{h=1}^{H}\sum_{j<l}^{n_h} E_{T_i}\{\nu_i(Y_{h,j})\nu_i(Y_{h,l})\}/(n_h-1)$$

follow exactly the same distribution. As H diverges to infinity, $(nc_n)^{1/2}W\widetilde{I}_{5,1}$ is asymptotically normal with mean 0 and variance of the form

$$
\begin{aligned}
& \lim_{n\to\infty} E(c_n n W^2 \widetilde{I}_{5,1}^2) \\
= & \lim_{n\to\infty} 4c_n n^{-1} W^2 \sum_{h=1}^{H} \sum_{j<l}^{n_h} E\Big[E_{T_i}^2\{\nu_i(Y_{h,j})\nu_i(Y_{h,l})\}\Big] \Big/ (n_h-1)^2 \\
= & \; 2W^2 E\Big[E_{T_i}^2\{\nu_i(Y_1)\nu_i(Y_2)\}\Big] = W^4\sigma^2.
\end{aligned}
$$

As a two-sample U-statistic, $\widehat{W}$ converges in probability to W as n diverges (van der Vaart, 1998, Theorem 12.6). The Slutsky's Theorem yields that $(nc_n)^{1/2}\widehat{\mathcal{S}}(\mathbf{x}, Y)$ is asymptotically normal with mean 0 and variance σ^2. In particular, if Y is continuous, $\sigma^2 = 4/5$. The proof is completed under independence.

Next we prove the consistency of $\widehat{\mathcal{S}}(\mathbf{x}, Y)$ when $\mathbf{x}$ and Y are not independent. As we have already shown that $\widehat{W}$ converges in probability to W, we only prove $\widetilde{\Lambda} - \Lambda$ converges in probability to 0 in the following context. Under Condition (C3)–(C4), we know $\widetilde{I}_1 = o_p(1)$ and $\widetilde{I}_2 = o_p(1)$. Thus, $\widetilde{\Lambda} - \Lambda = \widetilde{I}_4 - \widetilde{I}_5 - \Lambda + o_p(1)$. We have

$$
\widetilde{I}_4 = n^{-2} \sum_{i=1}^{n} \sum_{h=1}^{H} \sum_{j=1}^{n_h} \varepsilon_i^2(\mathbf{x}_{(h,j)}) = n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \varepsilon_i^2(\mathbf{x}_j).
$$

The above formula shows that $\widetilde{I}_4$ is a two sample U-statistics. From van der Vaart (1998, Theorem 12.6), $\widetilde{I}_4$ converges in probability to $E\{\varepsilon_i^2(X_j)\} = \Lambda$.

Similarly as the deviation in (S.1.3), we can prove that

$$\widetilde{I}_5 = n^{-1}\sum_{h=1}^{H}\sum_{j\neq l}^{n_h} E_{T_i}\left\{\varepsilon_i(\mathbf{x}_{(h,j)})\varepsilon_i(\mathbf{x}_{(h,l)})\right\}/(n_h-1) + O_p(n^{-1}).$$

To prove $\widetilde{I}_5 = o_p(1)$, we derive that

$$\begin{aligned} & \mathrm{pr}\left[n^{-1}\sum_{h=1}^{H}\sum_{j\neq l}^{n_h}\left|E_{T_i}\left\{\varepsilon_i(\mathbf{x}_{(h,j)})\varepsilon_i(\mathbf{x}_{(h,l)})\right\}\right|/(n_h-1) > \eta\right] \\ \leq\ & E\left(\mathrm{pr}\left[n^{-1}\sum_{h=1}^{H}\sum_{j\neq l}^{n_h}\left|E_{T_i}\left\{\varepsilon_i(\mathbf{x}_{(h,j)})\varepsilon_i(\mathbf{x}_{(h,l)})\right\}\right|/(n_h-1) > \eta\,\middle|\,\{\mathbf{x}_i\}_{i=1}^n\right]\right). \end{aligned}$$

To control the probability inside the expectation, we use the concentration inequality for bounded difference function (Wainwright, 2019, Corollary 2.21). Given $\{\mathbf{x}_i\}_{i=1}^n$, we denote

$$g(Y_1,\ldots,Y_n) \stackrel{\mathrm{def}}{=} n^{-1}\sum_{h=1}^{H}\sum_{j\neq l}^{n_h} E_{T_i}\left\{\varepsilon_i(\mathbf{x}_{(h,j)})\varepsilon_i(\mathbf{x}_{(h,l)})\right\}/(n_h-1).$$

If Y_j is replaced by Y_j', the change of $g(Y_1,\ldots,Y_n)$ is bounded by n^{-1}. Therefore, we have

$$\mathrm{pr}\left[n^{-1}\sum_{h=1}^{H}\sum_{j\neq l}^{n_h}\left|E_{T_i}\left\{\varepsilon_i(\mathbf{x}_{(h,j)})\varepsilon_i(\mathbf{x}_{(h,l)})\right\}\right|/(n_h-1) > \eta\,\middle|\,\{\mathbf{x}_i\}_{i=1}^n\right] \leq 2\exp(-n\eta^2).$$

Thus, we have $\mathrm{pr}(|\widetilde{I}_5| > \eta) \leq 2\exp(-n\eta^2)$, which means that $\widetilde{I}_5 = o_p(1)$. This completes the proof. $\square$

S.5. Technical Lemmas

Lemma 1 (Lemma A.1 of Hsing and Carroll (1992))**.** Suppose $Z_1,\ldots,Z_n$ are an i.i.d. sample and r is a positive constant. Let $Z_{(i)}$ be the i-th order statistic. Then $n^{-1/r}(|Z_{(n)}|+|Z_{(1)}|) = o_p(1)$ if and only if $x^r pr(|Z| > x) \to 0$ as $x \to \infty$.

Lemma 2. Under Condition (C1), $I_2 = o_p(n^{-1/2})$.

Proof of Lemma 2: By definition,

$$I_2 = \{n^2(c-1)\}^{-1} \sum_{h=1}^{H} \sum_{j<l}^{c} \sum_{i=1}^{n} \{s_i(X_{(h,j)}) - s_i(X_{(h,l)})\}\{\varepsilon_i(X_{(h,j)}) - \varepsilon_i(X_{(h,l)})\}.$$

We have

$$\begin{aligned}
|I_2| &\leq \{n^2(c-1)\}^{-1} \sum_{i=1}^{n} \sum_{h=1}^{H} \sum_{j<l}^{c} |s_i(X_{(h,j)}) - s_i(X_{(h,l)})| \\
&\leq \{n^2(c-1)\}^{-1} \sum_{i=1}^{n} \sum_{m=1}^{c-1} \sum_{j=1}^{n-m} |s_i(X_{(j+m)}) - s_i(X_{(j)})| \\
&\leq \{n^2(c-1)\}^{-1} \sum_{i=1}^{n} \sum_{m=1}^{c-1} \sum_{k=1}^{m} \sum_{j=1}^{n-1} |s_i(X_{(j+1)}) - s_i(X_{(j)})| \\
&\leq 2c/n^2 \sum_{i=1}^{n} \sum_{j=1}^{n-1} |s_i(X_{(j+1)}) - s_i(X_{(j)})| \qquad \text{(S.5.1)}
\end{aligned}$$

By the boundedness of $\varepsilon_i(X_{(h,j)})$, the first inequality follows. The second is resulted from the additivity across different slices, the third follows from the triangle inequality, and the fourth uses the fact that the summations over m and k have $c(c-1)/2$ terms.

In what follows, we shall show that

$$S_n \stackrel{\text{def}}{=} \sum_{j=1}^{n-1} |s_i(X_{(j+1)}) - s_i(X_{(j)})| = o_p(n^{1/2}).$$

If X has a bounded support, by (C1),

$$\lim_{n\to\infty} n^{-1/2} \sup_{\Pi_n(B)} \sum_{j=1}^{n-1} |s_i(X_{(j+1)}) - s_i(X_{(j)})| = 0$$

almost surely, which implies $S_n = o_p(n^{1/2})$.

If the support of X is unbounded, it suffices to show that for $\delta \in (0, 1/2)$,

$$n^{-1/2} \sum_{j=[n\delta]}^{[n(1-\delta)]} |s_i(X_{(j+1)}) - s_i(X_{(j)})| \xrightarrow{pr} 0, \tag{S.5.2}$$

and, for $\eta > 0$,

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \left\{ \mathrm{pr}\left(n^{-1/2} \sum_{j=1}^{[n\delta]} |s_i(X_{(j+1)}) - s_i(X_{(j)})| > \eta \right) + \mathrm{pr}\left(n^{-1/2} \sum_{j=[n(1-\delta)]}^{n-1} |s_i(X_{(j+1)}) - s_i(X_{(j)})| > \eta \right) \right\} = 0. \tag{S.5.3}$$

We shall show (S.5.2) and (S.5.3) in the following two steps.

Step 1. We aim to show (S.5.2) holds. Let G be the distribution of X and $G^{\leftarrow}$ the left-continuous inverse of G. Define $A_n = 1\{X_{([n\delta])} > G^{\leftarrow}(\beta)\}$ and $B_n = 1\{X_{([n(1-\delta)])} < G^{\leftarrow}(1-\beta)\}$ for $0 < \beta < \delta$. For some $\beta > 0$, $\min\{E(A_n), E(B_n)\} \to 1$. Thus (S.5.2) follows from

$$n^{-1/2} \sum_{j=[n\delta]}^{[n(1-\delta)]} |s_i(X_{(j+1)}) - s_i(X_{(j)})| A_n B_n \xrightarrow{pr} 0,$$

which, in turns, follows from (C1) with $r = 1/2$.

Step 2. We aim to show (S.5.3). Set $\delta > 0$ small enough such that $E(C_n) \to 1$, where $C_n = 1(X_{([n\delta])} < -B_0)$. By the non-expansive condition in (C1), we have

$$\begin{aligned} n^{-1/2} \sum_{j=1}^{[n\delta]} |s_i(X_{(j+1)}) - s_i(X_{(j)})| C_n &\leq n^{-1/2} \sum_{j=1}^{[n\delta]} |M(X_{(j+1)}) - M(X_{(j)})| \\ &= n^{-1/2} |M(X_{(1)}) - M(X_{([n\delta])})|, \end{aligned}$$

which approaches zero by Lemma 1. The other tail can be dealt with similarly.

Combining steps 1-2, we have

$$I_2 = 2c/n^2 \sum_{i=1}^{n} o_p(n^{1/2}) = o_p(n^{-1/2}),$$

which completes the proof of Lemma 2. □

Lemma 3. Under Condition (C1), $I_1 = o_p(n^{-1/2})$.

Proof of Lemma 3: By definition,

$$\begin{aligned} I_1 &= \{n^2(c-1)\}^{-1} \sum_{h=1}^{H} \sum_{j<l}^{c} \sum_{i=1}^{n} \{s_i(X_{(h,j)}) - s_i(X_{(h,l)})\}^2 \\ &= \{n^2(c-1)\}^{-1} \sum_{i=1}^{n} \sum_{h=1}^{H} \sum_{j<l}^{c} \{s_i(X_{(h,j)}) - s_i(X_{(h,l)})\}^2. \end{aligned}$$

We apply a similar operation in (S.5.1) to obtain that

$$\begin{aligned} I_1 &= \{n^2(c-1)\}^{-1} \sum_{i=1}^{n} \sum_{h=1}^{H} \sum_{j<l}^{c} \{s_i(X_{(h,j)}) - s_i(X_{(h,l)})\}^2 \\ &\leq 2c/n^2 \sum_{i=1}^{n} \sum_{j=1}^{n-1} |s_i(X_{(j+1)}) - s_i(X_{(j)})|. \end{aligned}$$

The above term is the same as the bound term in (S.5.1). Following the same paradigm in proof of Lemma 2, we complete the proof of Lemma 3. $\square$

Lemma 4. Under Condition (C2), we have

$$\{n(c-1)^2\}^{-1}\sum_{h=1}^{H}\sum_{j\neq l}^{c} E[V(T_i,T_k;X_{(h,j)})\{V(T_i,T_k;X_{(h,l)})-V(T_i,T_k;X_{(h,j)})\}] = o(1).$$

Proof of Lemma 4: We mainly follow similar arguments for proving Lemma A.2 of Hsing and Carroll (1992). With straightforward algebraic calculations,

$$\begin{aligned}
&\left|\{n(c-1)^2\}^{-1}\sum_{h=1}^{H}\sum_{j\neq l}^{c} V(T_i,T_k;X_{(h,j)})\{V(T_i,T_k;X_{(h,l)})-V(T_i,T_k;X_{(h,j)})\}\right| \\
\leq\ & \{n(c-1)^2\}^{-1}\sum_{h=1}^{H}\sum_{j\neq l}^{c} |V(T_i,T_k;X_{(h,j)})||V(T_i,T_k;X_{(h,l)})-V(T_i,T_k;X_{(h,j)})| \\
\leq\ & \{n(c-1)^2\}^{-1}\sum_{h=1}^{H}\sum_{j\neq l}^{c} |\{V(T_i,T_k;X_{(h,l)})-V(T_i,T_k;X_{(h,j)})| \\
\leq\ & n^{-1}\sum_{h=1}^{H}\sum_{j=2}^{c} |\{V(T_i,T_k;X_{(h,j)})-V(T_i,T_k;X_{(h,j-1)})| \\
\leq\ & n^{-1}\sum_{j=2}^{n} |\{V(T_i,T_k;X_{(j)})-V(T_i,T_k;X_{(j-1)})|.
\end{aligned}$$

It suffices to prove the following two facts.

$$n^{-1}\sum_{j=[n\delta]}^{[n(1-\delta)]} |\{V(T_i,T_k;X_{(j)})-V(T_i,T_k;X_{(j-1)})\}| \xrightarrow{pr} 0 \qquad \text{(S.5.4)}$$

and for any $\eta > 0$,

$$\lim_{\delta\to 0}\limsup_{n\to\infty} \mathrm{pr}\left[n^{-1}\sum_{j=1}^{[n\delta]} |\{V(T_i,T_k;X_{(j)}) - V(T_i,T_k;X_{(j-1)})| > \eta\right] \qquad \text{(S.5.5)}$$

$$+\lim_{\delta\to 0}\limsup_{n\to\infty} \mathrm{pr}\left[n^{-1}\sum_{j=[n(1-\delta)]}^{n} |\{V(T_i,T_k;X_{(j)}) - V(T_i,T_k;X_{(j-1)})| > \eta\right] = 0$$

Under condition (C2), we can prove (S.5.4) following similar procedure to derive (S.5.2). In addition, (S.5.5) is true because $V(T_i, T_k; X_j)$ is bounded.

□

Lemma 5. Under Condition (C3)-(C4), $\widetilde{I}_2 = o_p(1)$ and $\widetilde{I}_1 = o_p(1)$.

Proof of Lemma 5: We only prove $\widetilde{I}_2 = o_p(1)$ and omit the proof for $\widetilde{I}_1 = o_p(1)$, which follows similar pattern. By definition,

$$\widetilde{I}_2 \;=\; n^{-2}\sum_{i=1}^{n}\sum_{h=1}^{H}\sum_{j<l}^{n_h}(n_h-1)^{-1}\{s_i(\mathbf{x}_{(h,j)}) - s_i(\mathbf{x}_{(h,l)})\}\{\varepsilon_i(\mathbf{x}_{(h,j)}) - \varepsilon_i(\mathbf{x}_{(h,l)})\}.$$

We have

$$\begin{aligned}
|\widetilde{I}_2| &\le n^{-2}\sum_{i=1}^{n}\sum_{h=1}^{H}\sum_{j<l}^{n_h}(n_h-1)^{-1}|s_i(\mathbf{x}_{(h,j)}) - s_i(\mathbf{x}_{(h,l)})| \\
&\le n^{-2}\sum_{i=1}^{n}\sum_{h=1}^{H}\sum_{j<l}^{n_h}(n_h-1)^{-1}\big\{|s_i(\mathbf{x}_{(h,j)}) - s_i(\mathbf{x}_{(h,0)})| + |s_i(\mathbf{x}_{(h,l)}) - s_i(\mathbf{x}_{(h,0)})|\big\} \\
&\le (C_3/n)\sum_{h=1}^{H}\sum_{j=1}^{n_h}\|\mathbf{x}_{(h,j)} - \mathbf{x}_{(h,0)}\| \le C_3\{n^{-1}\sum_{h=1}^{H}\sum_{j=1}^{n_h}\|\mathbf{x}_{(h,j)} - \mathbf{x}_{(h,0)}\|^2\}^{-1/2}.
\end{aligned}$$

Here, $\mathbf{x}_{(h,0)}$ is the sample mean of the data in h-th cluster. By the boundedness of $\varepsilon_i(X_{(h,j)})$, the first inequality follows. The second is resulted from the triangle inequality. And the third can be established from Condition (C4). The last uses Cauchy-Schwarz inequality. According to Lemma 6, we know $\widetilde{I}_2$ converges in probability to 0, as n diverges to infinity. Thus, we complete the proof of Lemma 5. $\square$

Lemma 6. Under Condition (C3), given $H = O(n^\delta)$, there exists H points for initial centers in K-means algorithm, that

$$n^{-1}\sum_{h=1}^{H}\sum_{j=1}^{n_h}\|\mathbf{x}_{(h,j)} - \mathbf{x}_{(h,0)}\|^2$$

converges in probability to 0, as n diverges to infinity.

Proof of Lemma 6: For simplicity, we prove the case that there is no cluster deleted in each iteration. Under Condition (C3), we have

$$\mathrm{pr}\left(\sup_{i\in\{1,\ldots,n\}}\|\mathbf{x}_i\| > r\right) \le C_1 n\exp(-C_2 r^2).$$

Denote the ball $\mathcal{Z}_r \stackrel{\text{def}}{=} \{\mathbf{z}\in\mathbb{R}^p : \|\mathbf{z}\|\le r\}$ equipped with Euclidean distance. Let $\{\mathbf{z}_h\}_{h=1}^H$ be the ε-covering of $\mathcal{Z}_r$, that for any $\mathbf{z}\in\mathcal{Z}_r$, $\inf_{h\in\{1,\ldots,H\}}\|\mathbf{z} - \mathbf{z}_h\| \le \varepsilon$. According to the covering number (Wainwright, 2019, Example 5.8), such covering set exists, if we set $H \ge (1 + 2r/\varepsilon)^p$. And we denote these H points as the initial points for K-means algorithm.

For K-means clustering with standard algorithm, we know that the loss function decreases monotonically in each iteration. Therefore, the loss in the final iteration less or equal to the loss with initial points. That is,

$$n^{-1}\sum_{h=1}^{H}\sum_{j=1}^{n_h}\|\mathbf{x}_{(h,j)}-\mathbf{x}_{(h,0)}\|^2 \le n^{-1}\sum_{j=1}^{n}\inf_{h\in\{1,\ldots,H\}}\|\mathbf{x}_j-\mathbf{z}_h\|^2.$$

We consider the right side term and have

$$\begin{aligned} & \operatorname{pr}\left(n^{-1}\sum_{j=1}^{n}\inf_{h\in\{1,\ldots,H\}}\|\mathbf{x}_j-\mathbf{z}_h\|^2 > t\right) \\ \le\ & \operatorname{pr}\left\{n^{-1}\sum_{j=1}^{n}\inf_{h\in\{1,\ldots,H\}}\|\mathbf{x}_j-\mathbf{z}_h\|^2 1(\|\mathbf{x}_j\|\le r) > t\right\} + \operatorname{pr}\left(\sup_{i\in\{1,\ldots,n\}}\|\mathbf{x}_i\| > r\right). \end{aligned}$$

To deal with the first term, we use the concentration inequality for bounded difference function (Wainwright, 2019, Corollary 2.21). From the definition of covering set $\{\mathbf{z}_h\}_{h=1}^{H}$, if $\mathbf{x}_j$ is replaced by $\mathbf{x}_j'$, the change of the following function is bounded by ε^2/n.

$$g(\mathbf{x}_1,\ldots,\mathbf{x}_n) \stackrel{\text{def}}{=} n^{-1}\sum_{j=1}^{n}\min_h\|\mathbf{x}_j-\mathbf{z}_h\|^2 1(\|\mathbf{x}_j\|\le r).$$

Therefore,

$$\operatorname{pr}\Big\{\Big|g(\mathbf{x}_1,\ldots,\mathbf{x}_n)-E\left[g(\mathbf{x}_1,\ldots,\mathbf{x}_n)\right]\Big| > t/2\Big\} \le 2\exp\{-nt^2/(2\varepsilon^4)\}. \tag{S.5.6}$$

From the definition of covering set, we also have $g(\mathbf{x}_1,\ldots,\mathbf{x}_n)<\varepsilon^2$. Thus,

$$E\left[g(\mathbf{x}_1,\ldots,\mathbf{x}_n)\right] = E\left[\min_h\|\mathbf{x}-\mathbf{z}_h\|^2 1(\|\mathbf{x}\|\le r)\right] \le \epsilon^2. \tag{S.5.7}$$

Let $\varepsilon = n^{-\gamma_1}$, $t = n^{-\gamma_2}$ with $\delta^{1/p} > \gamma_1 \ge 2^{-1}\gamma_2 > 0$, $1+4\gamma_1-2\gamma_2 > 0$.

Combining the result of (S.5.6) and (S.5.7), we have

$$\begin{aligned}
& \operatorname{pr}\left\{n^{-1}\sum_{j=1}^{n}\inf_{h\in\{1,\ldots,H\}}\|\mathbf{x}_j-\mathbf{z}_h\|^2 1(\|\mathbf{x}_j\|\le r) > n^{-\gamma_2}\right\} \\
\le\ & n^{-2\gamma_1} + 2\exp(-n^{1+4\gamma_1-2\gamma_2}/2).
\end{aligned}$$

We choose $r = n^{\gamma_3}$ that $\gamma_3 > 0$ and $p(\gamma_1+\gamma_3)\le\delta$, we have

$$\begin{aligned}
& \operatorname{pr}\left(n^{-1}\sum_{j=1}^{n}\inf_{h\in\{1,\ldots,H\}}\|\mathbf{x}_j-\mathbf{z}_h\|^2 > n^{-\gamma_2}\right) \\
\le\ & n^{-2\gamma_1} + 2\exp(-n^{1+4\gamma_1-2\gamma_2}/2) + C_1 n\exp(-C_2 n^{2\gamma_3}).
\end{aligned}$$

As $n\to\infty$, we have

$$n^{-1}\sum_{j=1}^{n}\inf_{h\in\{1,\ldots,H\}}\|\mathbf{x}_j-\mathbf{z}_h\|^2 \xrightarrow{p} 0.$$

Here, $\xrightarrow{p}$ stands for "converge in probability". $\square$

References

Chatterjee, S. (2020). A new coefficient of correlation. *Journal of the American Statistical Association*, 1–21.

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

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge, UK: Cambridge University Press.

Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*, Volume 48. Cambridge University Press.

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