HHS Public Access
Author manuscript
J Am Stat Assoc. Author manuscript; available in PMC 2016 February 10.

Published in final edited form as:
J Am Stat Assoc. 2015 ; 110(512): 1726–1734. doi:10.1080/01621459.2014.993081.

CONDITIONAL DISTANCE CORRELATION

Xueqin Wang, **Wenliang Pan**, **Wenhao Hu**, **Yuan Tian**, and **Heping Zhang**
Sun Yat-Sen University, North Carolina State University, Yale University

Abstract

Statistical inference on conditional dependence is essential in many fields including genetic association studies and graphical models. The classic measures focus on linear conditional correlations, and are incapable of characterizing non-linear conditional relationship including non-monotonic relationship. To overcome this limitation, we introduces a nonparametric measure of conditional dependence for multivariate random variables with arbitrary dimensions. Our measure possesses the necessary and intuitive properties as a correlation index. Briefly, it is zero almost surely if and only if two multivariate random variables are conditionally independent given a third random variable. More importantly, the sample version of this measure can be expressed elegantly as the root of a V or U-process with random kernels and has desirable theoretical properties. Based on the sample version, we propose a test for conditional independence, which is proven to be more powerful than some recently developed tests through our numerical simulations. The advantage of our test is even greater when the relationship between the multivariate random variables given the third random variable cannot be expressed in a linear or monotonic function of one random variable versus the other. We also show that the sample measure is consistent and weakly convergent, and the test statistic is asymptotically normal. By applying our test in a real data analysis, we are able to identify two conditionally associated gene expressions, which otherwise cannot be revealed. Thus, our measure of conditional dependence is not only an ideal concept, but also has important practical utility.

Keywords

Conditional distance correlation; Conditional independence test; $U(V)$ process with random kernel; Local bootstrap

1. INTRODUCTION

The concept of conditional dependence or independence is fundamental to statistical inference such as casual inference, graphical models, dimension reduction and Bayesian

Xueqin Wang is Professor, Department of Statistical Science, School of Mathematics and Computational Science, Southern China Research Center of Statistical Science, Sun Yat-Sen University, Guangzhou, 510275, China; Zhongshan School of Medicine, Sun Yat-Sen University, Guangzhou, 510080, China; and Xinhua College, Sun Yat-Sen University, Guangzhou, 510520, China (wangxq88@mail.sysu.edu.cn). Wenliang Pan is a Ph.D. student and Yuan Tian is a Master student, Department of Statistical Science, School of Mathematics and Computational Science, Sun Yat-Sen University, Guangzhou, 510275, China (sysu.wenliang@gmail.com and erin.tien@gmail.com). Wenhao Hu is a doctoral student, Department of Statistics, North Carolina State University, Raleigh, NC 27695, U.S.A. (whu6@ncsu.edu). Heping Zhang is Susan Dwight Bliss Professor, Department of Biostatistics Yale University School of Public Health, New Haven, CT 06520-8034, USA (heping.zhang@yale.edu).

network analysis (Ackley et al., 1985; Zhang et al., 2012). Many scientific questions must be answered through statistical inference on conditional dependence.

The most commonly used method to measure conditional dependence is the classical partial correlation (Lawrance, 1976) between two multivariate random variables conditioning on a third multivariate random variable. A zero correlation is equivalent to the conditional independence, under the condition that the underlying joint distribution of the three multivariate random variables is normal. Without this condition, the statement may not hold, even though the converse is always true. More importantly, this measure suffers from some serious drawbacks; for example, the joint normality condition implies the linearity between random variables, and as a result, the partial correlation does not depend on the value of the conditioning random variable. This may not be realistic in nonlinear world (Speed, 2011). Due to these limitations, the partial correlation cannot serve as a general measure of conditional dependence, and an appropriately defined and broadly applicable measure is warranted. Various measures have been proposed in literature, but none of them have the ideal properties. Consequently, testing for conditional independence among multivariate random variables with arbitrary dimensions is usually chanllenging in practice.

Linton and Gozalo (1996) proposed a Kolmogorov-Smirnov type of test statistic and a Cramér von-Mises type of test statistic based on a generalization of the empirical distribution function. Gretton et al. (2005) provided a conditional independence test derived from the normalized conditional cross-covariance operator under the framework of Reproducing Kernel Hilbert Space (RKHS). A series of test statistics have been developed using various difference measures between conditional densities including the use of smoothing empirical likelihood (Su and White, 2003), conditional characteristic function (Su and White, 2007), and weighted Hellinger distance (Su and White, 2008). Huang (2010) suggested another test of conditional association based on the maximal nonlinear conditional correlation.

We develop a proper measure of conditional dependence for multivariate random variables with arbitrary dimensions. Specifically, our proposed measure is zero almost surely if and only if the two multivariate random variables are conditionally independent given a third multivariate random variable, and varies with the value of conditional random variable. Importantly, although our measure is based on conditional characteristic function, it can be simplified as a concise expression of conditional moments, and its sample version is convenient to use in practice. Some additional features will be outlined below.

In Section 2 we propose a nonparametric measure of conditional correlation (covariance) for multivariate random variables by replacing characteristic function used in the definition of distance correlation (covariance) (Székely et al., 2007) with conditional characteristic functions. This definition is conceptually straightforward but computationally difficult. Hence, we explore two alternative types of the sample conditional distance covariance (SCDCov) in Section 3: the roots of V and U-processes with random kernel. It is important to observe that our two sample measures have the same asymptotic limit when defined differently. This property enables us to introduce a conditional distance independence test (CDIT) for random variables with arbitrary dimensions in Section 4. We show that the CDIT is asymptotically normal. Monte Carlo simulation studies in Section 5 suggest that the CDIT is more powerful than some tests cited above based on nonlinearity and nonmonotonicity, and has a comparable power to the partial correlation test and kernel based conditional independence test in the multivariate normal case. We offer our conclusions in Section 6 and defer all technical details to the Appendix.

2. CONDITIONAL DISTANCE COVARIANCE AND CORRELATION

2.1 Definition

Let X, Y, and Z be p, q, and r dimensional random vectors in Euclidean spaces $\mathbb{R}^p$, $\mathbb{R}^q$, and $\mathbb{R}^r$, respectively. For vectors $t \in \mathbb{R}^p$ and $s \in \mathbb{R}^q$, the conditional joint characteristic function of X, Y given Z is defined as,

$$\phi_{X,Y|Z}(t,s)=E[\exp(i\langle t,X\rangle+i\langle s,Y\rangle)|Z],$$

where E denotes the expectation, i is the complex number, and $\langle\cdot,\cdot\rangle$ is the inner product of the two cooresponding vectors. In addition, the conditional marginal characteristic functions of X, Y given Z are, respectively,

$$\phi_{X|Z}(t)=\phi_{X,Y|Z}(t,0), \text{ and } \phi_{Y|Z}(s)=\phi_{Y|Z}(0,s).$$

Note that if X is independent of Y given Z, denoted by $X \perp Y|Z$, then $\varphi_{X,Y|Z} = \varphi_{X|Z}\varphi_{Y|Z}$.

The key idea is that we measure the conditional dependence through "distance" between $\varphi_{X,Y|Z}$ and $\varphi_{X|Z}\varphi_{Y|Z}$, which can be defined by replacing the characteristic functions with the conditional characteristic functions in the correlation definition of Székely et al. (2007). In what follow, we define conditional distance covariance (CDCov), conditional distance variance (CDVar), and conditional distance correlation (CDCor).

Definition 1 (CDCov)—The conditional distance covariance $\mathcal{D}(X, Y|Z)$ between random vectors X and Y with finite moments given Z is defined as the square root of

$$\begin{aligned}&\mathscr{D}^2(X,Y|Z)\\&=\|\phi_{X,Y|Z}(t,s)-\phi_{X|Z}(t)\phi_{Y|Z}(s)\|^2\\&=\frac{1}{c_pc_q}\int_{\mathbb{R}^{p+q}}\frac{|\phi_{X,Y|Z}(t,s)-\phi_{X|Z}(t)\phi_{Y|Z}(s)|^2}{|t|_p^{p+1}|s|_q^{q+1}}dtds,\end{aligned} \quad (1)$$

where $c_p=\frac{\pi^{(p+1)/2}}{\Gamma((p+1)/2)}$ and $c_q=\frac{\pi^{(q+1)/2}}{\Gamma((q+1)/2)}$.

The CDVar is the square root of

$$\mathscr{D}^2(X|Z)=\mathscr{D}^2(X,X|Z). \quad (2)$$

Definition 2 (CDCor)—The conditional distance correlation between random vectors X and Y with finite moments given Z is defined as the square root of

$$\rho^2(X,Y|Z)=\frac{\mathscr{D}^2(X,Y|Z)}{\sqrt{\mathscr{D}^2(X|Z)\mathscr{D}^2(Y|Z)}} \quad (3)$$

if $\mathcal{D}^2(X|Z)\ \mathcal{D}^2(Y|Z) > 0$, or 0 otherwise.

A Working Example: To help the understanding of these definitions, we go through the calculation through the following working example. Suppose that (X, Y, Z) is a random vector generated from a 3-dimensional multinormal distribution with the mean vector $\mu = (0, 0, 0)$ and covariance matrix $\sum=\begin{pmatrix} 1 & 0.36 & 0.6 \\ 0.36 & 1 & 0.6 \\ 0.6 & 0.6 & 1 \end{pmatrix}$. We can see that $X \perp Y|Z$ since $(X, Y)|Z$ follows a 2-dimensional multinormal distribution with the covariance matrix $\begin{pmatrix} 0.74 & 0 \\ 0 & 0.74 \end{pmatrix}$. The mean vector of $(X, Y)|Z$ is (Z, Z). Hence, the conditional characteristic function is

$$\begin{aligned}\phi_{X,Y|Z}(t,s)&=E[\exp(i\langle t,X\rangle+i\langle s,Y\rangle)|Z]\\ &=\int\int\exp(itx+isy)\cdot f_{X,Y|Z}\,\mathrm{d}x\mathrm{d}y\\ &=\int\int\exp(itx+isy)\cdot\frac{1}{2\pi\cdot0.74}\exp\{-\frac{1}{2.0.74}((x-Z)^2+(y-Z)^2)\}\,\mathrm{d}x\mathrm{d}y\\ &=\int\exp(itx)\frac{1}{\sqrt{2\pi\cdot0.74}}\exp\{-\frac{(x-Z)^2}{2\cdot0.74}\}\,\mathrm{d}x\cdot\int\exp(isy)\frac{1}{\sqrt{2\pi\cdot0.74}}\exp\{-\frac{(y-Z)^2}{2\cdot0.74}\}\,\mathrm{d}y\\ &=\phi_{X|Z}(t)\phi_{Y|Z}(s).\end{aligned}$$

Thus, $\mathcal{D}^2(X, Y|Z) = 0$ and CDCov=0.

2.2 Properties of CDCov and CDCor

In this section, we present the analogous properties of CDCov (CDCor) to those of unconditional distance covariance (DCov) or distance correlation (DCor) in the sense of "almost surely" as stated in Theorems 1 and 2. These properties ensure that CDCov (CDCor) is a proper measurement of conditional dependence.

Let $E(|X|_p|Z)$ denote the p-th conditional moment of X given Z.

Theorem 1 (Properties of CDCov)—For random vector $(X, Y, Z) \in \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}^r$ such that $E(|X|_p + |Y|_q|Z) < \infty$ and $\sigma(Z)$ is complete, the following properties hold:

- **i.** $\mathcal{D}(X, Y|Z) \geq 0$, and $\mathcal{D}(X, Y|Z) = 0$ if and only if X and Y are conditionally independent given Z.
- **ii.** $\mathcal{D}(X|Z) = 0$ implies that $X = E(X|Z)$.

iii. $\mathscr{D}(a_1+b_1C_1X, a_2+b_2C_2Y|Z)=\sqrt{|b_1b_2|}\mathscr{D}(X, Y|Z)$ for any constant vectors $a_1 \in \mathbb{R}^p$, $a_2 \in \mathbb{R}^q$, scalars b_1, b_2, $p \times p$ orthonormal matrix C_1, and $q \times q$ orthonormal matrix C_2.

iv. $\mathcal{D}(a + bCX|Z) = |b|\ \mathcal{D}(X|Z)$ for any constant vectors a in $\mathbb{R}^p$, scalar b and $p \times p$ orthonormal matrices C.

v. If random vectors (X_1, Y_1) and (X_2, Y_2) are conditionally independent given Z, then

$$\mathscr{D}(X_1+X_2, Y_1+Y_2|Z) \leq \mathscr{D}(X_1, Y_1|Z)+\mathscr{D}(X_2, Y_2|Z).$$

Equality holds if and only if X_1 and Y_1 are both the functions of Z, or X_2 and Y_2 are both the functions of Z, or X_1, Y_1 are conditionally independent given Z and also X_2, Y_2 are conditionally independent given Z.

vi. If X and Y are conditionally independent given Z, then $\mathcal{D}(X + Y|Z) \leq \mathcal{D}(X|Z) + \mathcal{D}(Y|Z)$. Equality holds if and only if either X or Y is a function of Z.

Part (i) in Theorem 1 indicates that CDCov is nonnegative and that $\mathcal{D}(X, Y|Z) = 0$ is the necessary and sufficient condition of conditional independence. Part (ii) assures that $\mathcal{D}(X|Z) = 0$ iff random variable X is $\sigma(Z)$-measurable. Parts (iii) and (iv) present the properties of the CDCov and the CDVar under a linear transformation. Parts (v) and (vi) provide Minkowski-type inequalities for the CDCov and the CDVar.

Theorem 2 (Properties of CDCor)

i. If $E(|X|_p + |Y|_q|Z) < \infty$, then $0 \leq \rho(X, Y|Z) \leq 1$, and $\rho(X, Y|Z) = 0$ if and only if X and Y are conditionally independent given Z.

ii. Orthogonal invariance: $\rho(a_1 + b_1C_1X, a_2 + b_2C_2Y|Z) = \rho(X, Y|Z)$ for any constant vectors $a_1 \in \mathbb{R}^p$, $a_2 \in \mathbb{R}^q$, scalars b_1, b_2, $p \times p$ orthonormal matrix C_1, and $q \times q$ orthonormal matrix C_2.

3. SAMPLE CONDITIONAL DISTANCE CORRELATION MEASURES

We have defined the conditional distance measures that have desirable theoretical properties. However, to explore their practical use, we need to introduce the sample forms to avoid complicated integrations involving conditional characteristic functions. Specifically, we consider two approaches. The first approach is to make use of V-process with random kernels, and verify that this V-process is identical to the form defined by plugging in the empirical characteristic functions in (1). The second approach is to estimate the CDCov by a U-process with random kernels. These two sample forms are constructed to have similar asymptotic properties and converge to the same theoretical conditional distance measures. The use of V-process or U-process allows us to employ the established theory of V-process and U-process.

3.1 Plug-in sample conditional distance covariance

Consider a kernel function K such as the Gaussian kernel (Li and Racine, 2007). Let $\omega_k(Z) = K_H(Z - Z_k)$ and $\omega(Z)/n = \sum_{k=1}^{n} \omega_k(Z)/n$. The theory below assumes that $\omega(Z)/n$ is a consistent estimator for the density of Z, which is known for the Gaussian kernel under certain regularity conditions.

Next, we define the joint empirical conditional characteristic function of $(X, Y)|Z$:

$$\phi^n_{X,Y|Z}(t,s) = \frac{1}{\omega(Z)} \sum_{k=1}^{n} \exp(i\langle t, X_k\rangle + i\langle s, Y_k\rangle)\omega_k(Z),$$

and the marginal empirical conditional characteristic functions of $X|Z$ and $Y|Z$:

$$\phi^n_{X|Z}(t) = \phi^n_{X,Y|Z}(t,0)$$

and

$$\phi^n_{Y|Z}(s) = \phi^n_{X,Y|Z}(0,s),$$

respectively. Then the sample conditional dependence is defined as follows.

Definition 3 (Plug-in sample conditional distance covariance, pSCDCov)— A sample conditional distance covariance $\mathscr{D}_n(\mathbf{W}_n|Z)$ between X_i's and Y_i's is defined as

$$\begin{aligned}\mathscr{D}^2_n(\mathbf{W}_n|Z) &= \mathscr{D}^2_n(\mathbf{X}_n, \mathbf{Y}_n, \mathbf{Z}_n|Z)\\ &= \|\phi^n_{X,Y|Z}(t,s) - \phi^n_{X|Z}(t)\phi^n_{Y|Z}(s)\|^2,\end{aligned} \quad (4)$$

where the norm $\|\cdot\|$ is defined in (1).

3.2 V-process based sample conditional distance covariance

Suppose that $W_i = (X_i, Y_i, Z_i)$, $i = 1, \cdots, n$ are sampled *iid.* from a random vector $W = (X, Y, Z)$ in $\mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}^r$. Let $\mathbf{X}_n = \{X_1, \cdots, X_n\}$, $\mathbf{Y}_n = \{Y_1, \cdots, Y_n\}$, $\mathbf{Z}_n = \{Z_1, \cdots, Z_n\}$, and $\mathbf{W}_n = (\mathbf{X}_n, \mathbf{Y}_n, \mathbf{Z}_n)$. Denote the Euclidean distance of X_i and X_j in $\mathbb{R}^p$ as $d^X_{ij} = d(X_i, X_j)$, and similarly, d^Y_{ij} for Y. Let

$$d_{ijkl} = (d^X_{ij} + d^X_{kl} - d^X_{ik} - d^X_{jl})(d^Y_{ij} + d^Y_{kl} - d^Y_{ik} - d^Y_{jl}).$$

Note that d_{ijkl} is not symmetric with respect to $\{i, j, k, l\}$, and we introduce a symmetric form as follows:

$$d^s_{ijkl}=d_{ijkl}+d_{ijlk}+d_{ilkj}.$$

Moreover, for the specific value of $\mathcal{D}^2(X, Y|Z = z)$ when given $Z = z$, the pairwise distances can be related to the CDCov by the following lemma.

Lemma 1— $\mathcal{D}^2(X, Y|Z = z)$ can be rewritten as:

$$\mathscr{D}^2(X,Y|Z=z)=\frac{1}{12}E[d^s_{1234}|Z_1=z, Z_2=z, Z_3=z, Z_4=z].$$

This expectation representation in Lemma 1 reveals that we can estimate our conditional dependence measures by using V process or U process. We begin with the definition of a V-type sample conditional distance covariance (vSCDCov) as the square root of

$$\mathscr{V}_n(\mathbf{W}_n|Z)=\frac{1}{n^4}\sum_{i,j,k,l}\psi_n(W_i,W_j,W_k,W_l;Z), \quad (5)$$

where $\psi_n()$ is the symmetric random kernel of degree 4 defined in Schick (1997):

$$\psi_n(W_i,W_j,W_k,W_l;Z)=\frac{n^4\omega_i(Z)\omega_j(Z)\omega_k(Z)\omega_l(Z)}{12\omega^4(Z)}d^s_{ijkl}. \quad (6)$$

Let $\mathbf{W}_{\mathbf{X}_n} = (\mathbf{X}_n, \mathbf{X}_n, \mathbf{Z}_n)$ and $\mathbf{W}_{\mathbf{Y}_n} = (\mathbf{Y}_n, \mathbf{Y}_n, \mathbf{Z}_n)$. Then, the V-type sample conditional distance correlation (vSCDCor), $\rho_n^{(v)}(\mathbf{W}_n|Z)$, between X_i's and Y_i's can be defined as the square root of

$$\frac{\mathscr{V}_n(\mathbf{W}_n|Z)}{\sqrt{\mathscr{V}_n(\mathbf{W}_{\mathbf{X}_n}|Z)\mathscr{V}_n(\mathbf{W}_{\mathbf{Y}_n}|Z)}}, \quad (7)$$

if $\mathcal{V}_n(\mathbf{W}_{\mathbf{X}_n}|Z)\ \mathcal{V}_n(\mathbf{W}_{\mathbf{Y}_n}|Z) > 0$ or 0 otherwise.

Theorem 3 below states that pSCDCov and vSCDCov are actually the same, giving rise to the same definition of sample conditional distance covariance, which we simply refer to as SCDCov from now on.

Theorem 3—If $\mathbf{W}_n = \{W_1, \cdots, W_n\}$ is a sample from the joint distribution of (X,Y,Z), then

$$\mathscr{V}_n(\mathbf{W}_n|Z)=\mathscr{D}^2_n(\mathbf{W}_n|Z).$$

3.3 U-process based sample conditional distance covariance

Here, we use U-process to further define an analogous unbiased sample conditional distance covariance (uSCDCov) as the square root of

$$\mathscr{U}_n(\mathbf{W}_n|Z)=\frac{1}{C_n^4}\sum_{i<j<k<l}\psi_n(W_i,W_j,W_k,W_l;Z). \quad (8)$$

Similarly, we can define the U-type sample conditional distance correlation (uSCDCor), $\rho_n^{(u)}(\mathbf{W}_n|Z)$.

3.4 Large sample theory of $\mathscr{V}_n$ and $\mathscr{U}_n$

We now present the asymptotic theory for $\mathscr{V}_n$ and $\mathscr{U}_n$ and assure that vSCDCov and uSCDCov converge to CDCov, and their companion correlations vSCDCor and uSCDCor converge to CDCor. In addition, the sample measures have desirable properties under linear transformations and the correlation measures are between 0 and 1.

Theorem 4—If $E(|X|_p + |Y|_q|Z) < \infty$, and if $\omega(Z)/n$ is a consistent density function estimator of Z, then $\mathscr{V}_n(\mathbf{W}_n|Z)$($\mathscr{U}_n(\mathbf{W}_n|Z)$) converges to $\mathscr{D}^2(X, Y|Z)$ in probability at each point of Z, that is,

$$\mathscr{V}_n(\mathbf{W}_n|Z)\xrightarrow[n\to\infty]{P}\mathscr{D}^2(X,Y|Z),$$
$$\mathscr{U}_n(\mathbf{W}_n|Z)\xrightarrow[n\to\infty]{P}\mathscr{D}^2(X,Y|Z).$$

The following corollary states the consistency of the sample conditional distance correlations.

Corollary 1—Under the assumptions of Theorem 4, it is in probability at each point of Z that

$$\rho_n^{(v)}(\mathbf{W}_n|Z)\xrightarrow[n\to\infty]{P}\rho(X,Y|Z),$$
$$\rho_n^{(u)}(\mathbf{W}_n|Z)\xrightarrow[n\to\infty]{P}\rho(X,Y|Z).$$

Theorem 5—For all constant vectors $a_1 \in \mathbb{R}^p$, $a_2 \in \mathbb{R}^q$, scalars b_1, b_2 and $p \times p$ orthonormal matrix C_1, $q \times q$ orthonormal matrix C_2, the following equations hold at each point of Z,

i. $\mathscr{V}_n(a_1 + b_1C_1\mathbf{X}_n, a_2 + b_2C_2\mathbf{Y}_n, \mathbf{Z}_n|Z) = |b_1b_2|\, \mathscr{V}_n(\mathbf{X}_n, \mathbf{Y}_n, \mathbf{Z}_n|Z)$.

ii. $\mathscr{U}_n(a_1 + b_1C_1\mathbf{X}_n, a_2 + b_2C_2\mathbf{Y}_n, \mathbf{Z}_n|Z) = |b_1b_2|\, \mathscr{U}_n(\mathbf{X}_n, \mathbf{Y}_n, \mathbf{Z}_n|Z)$.

iii. $\rho_n^{(v)}(a_1+b_1C_1\mathbf{X}_n, a_2+b_2C_2\boldsymbol{Y}_n, \boldsymbol{Z}_n|Z)=\rho_n(\mathbf{X}_n, \boldsymbol{Y}_n, \boldsymbol{Z}_n|Z)$.

iv. $\rho_n^{(u)}(a_1+b_1C_1\mathbf{X}_n, a_2+b_2C_2\boldsymbol{Y}_n, \boldsymbol{Z}_n|Z)=\rho_n^{'}(\mathbf{X}_n, \boldsymbol{Y}_n, \boldsymbol{Z}_n|Z)$.

v. $0 \leq \rho_n, \rho_n^{'} \leq 1$.

The above theorem confirms that the sample measures retain the linearity property and Cauchy inequality.

4. CONDITIONAL DISTANCE INDEPENDENCE TEST

4.1 The hypothesis

In this section, we discuss how to test conditional independence which can be stated as

$$X \perp Y|Z, \text{ or equivalently, } \mathscr{D}^2(X,Y|Z)=0 \text{ almost surely}.$$

To this end, we introduce quantities that can represent the conditional independence. Specifically, let

$$\mathscr{S}_a=E[\mathscr{D}^2(X,Y|Z)a(Z)], \quad (9)$$

where $a(\cdot)$ is a certain nonnegative function with the same support as the probability density function of Z. For computational considerations, we choose $a(Z)$ to be $12f^4(Z)$. Consequently, $X \perp Y|Z$ if and only if $\mathscr{S}_a = 0$.

4.2 The kernel function and bandwidth selection

In the previous section, we introduced a general kernel function K. Here, we choose the Gaussian kernel

$$K_H(\mathbf{u})=|H|^{-1}K(H^{-1}\cdot\mathbf{u})=(2\pi)^{-\frac{r}{2}}|H|^{-1}\exp(-\frac{1}{2}\mathbf{u}'H^{-2}\mathbf{u})$$

in $\mathbb{R}^r$, where H is a diagnoal matrix diag$\{h, \cdots, h\}$ determined by bandwidth h. There are other kernels, but the Gaussian kerne is commonly used in statistical learning, and its theoretical and computation properties are well established.

It is well known that the use of kernel function usually involves a challenging choice of bandwidth selection. This issue is not entirely solved, although there are many competing methods as presented in Li and Racine (2007), including so-called Rule-of-Thumb, Plugin, Least Square CV and Likelihood CV. For convenience of computation, we consider the Rule-of-Thumb method as introduced by Wand and Jones (1994, 1995). In fact, we found this bandwidth selection method works well in our simulation studies. In other words, there could be a more optimal method, but the gain in performance may not be outweigh the increased complexity by exploring a potentially better method.

With the Gaussian kernel, $\omega(Z)/n$ is known to be consistent under the following regularity conditions:

(C1) $\int_{\mathbb{R}^r} \mathbf{u}K(\mathbf{u})d\mathbf{u} = 0$, $\int_{\mathbb{R}^r} K(\mathbf{u})d\mathbf{u} = 1$, $\int_{\mathbb{R}^r} |K(\mathbf{u})|d\mathbf{u} < \infty$, $\int_{\mathbb{R}^r} K^2(\mathbf{u})d\mathbf{u} > 0$, $\int_{\mathbb{R}^r} \mathbf{u}^2K(\mathbf{u})d\mathbf{u} < \infty$.

(C2) $h^r \to 0$ and $nh^r \to \infty$, as $n \to \infty$. This requires h to be choosen appropriately according to n.

(C3) The density function of Z and the conditional density function $f(\cdot|z)$ are twice differentiable and all of the derivatives are bounded.

4.3 The test and its properties

Let

$$\mathscr{S}_n = \frac{1}{C_n^5 |H|^4} \sum_{i<j<k<l<u} d_{ijkl}^s K_{iu} K_{ju} K_{ku} K_{lu}, \quad (10)$$

where $K_{iu} = K(H^{-1} \cdot (Z_i - Z_u))$.

The theorem below states that $\mathscr{S}_n$ is a consistent estimate of $\mathscr{S}_a$, and we propose it as a conditional distance independence test (CDIT). We should also note that as proven in the appendix, this theorem holds for any kernel function K satisfying the same regularity conditions as above.

Theorem 6 (Consistency)—Assume that conditions (C1)–(C3) hold and the second moments of X and Y exist, then as $n \to \infty$, we have

$$\mathscr{S}_n \xrightarrow[n\to\infty]{P} \mathscr{S}_a.$$

Moreover, using the theorey of U statistic discussed in Fan and Li (1996) and Lee (1990), we have the following asymptotic normality.

Theorem 7 (Weak convergence)—Assume that conditions (C1)–(C3) hold and the second moments of X and Y exist. If X and Y are conditionally independent given Z and if $n^2h^{(r+4)^2/4} \to 0$ as $n \to \infty$, we have

$$nh^{r/2}\mathscr{S}_n \xrightarrow[n\to\infty]{d} N(0, \sigma^2),$$

where σ^2 will be given in (A.9).

Theorem 7 validates that $nh^{r/2}\, \mathscr{S}_n$ is asymptotically normally distributed. Although we can use this asymptotic normality to compute p-values, we need to be cautious of its serious limitations in practice. Firstly, as it becomes clear in (A.9), the variance parameter σ^2 can be difficult to compute, and it is even more complicated for unknown distribution of random variables. Secondly, the sample size in practice may be too small (Su and White, 2007).

In practice, we consider the local bootstrap procedure proposed by Paparoditis and Politis (2000) as an alernative. For a given sample $\mathbf{W}_n = \{(X_i, Y_i, Z_i) : i = 1, \cdots, n\}$, we draw a local bootstrap sample $\mathbf{W}_\mathbf{n}^* = \{(X_i^*, Y_i, Z_i): i=1, \cdots, n\}$, and then compute the bootstrap statistic. The specific steps are as follows.

i. For $i = 1, \cdots, n$, draw X_i^* from

$$\hat{F}_{X|Z=Z_i} = \frac{\sum_{j=1}^n K_{ij} I_{(-\infty, X_j)}(x)}{\sum_{j=1}^n K_{ij}}.$$

ii. Compute $\mathscr{S}_n^*$ by using the local bootstrap sample

$$\mathbf{W}_\mathbf{n}^* = \{(X_i^*, Y_i, Z_i): i=1, \cdots, n\}.$$

iii. Repeat (a) and (b) for B times, and obtain $\mathscr{S}_{nk}^*$, $k = 1, \cdots, B$. And then the p-value of the test is given by

$$p = \frac{\sum_{k=1}^B [1 + I(|nh^{r/2}\mathscr{S}_n| > |nh^{r/2}\mathscr{S}_{nk}^*|)]}{B+1}.$$

5. EMPIRICAL RESULTS

In this section, we conduct several simulation experiments to demonstrate the performance of CDIT in comparison to CI.test (Scutari, 2009) and KCI.test (Fukumizu et al., 2008). The computation is based on 1000 random samples, and we use the significance level of 0.05.

Examples 1–4 consider the case that X and Y are conditionally independent given Z. Specifically, X, Y and Z are all univariate in Examples 1–3, X and Y remain univariate but Z is multivariate in Example 4. Simulation results are summarized in Table 1.

We consider various conditional dependence cases in Examples 5–12 and report the results in Table 2. As for Examples 1–4, we construct corresponding conditional dependence cases in Examples 5–8. Examples 9–12 evaluate the power of our proposed test with X, Y being multivariate for which CI.test and KCI.test are not applicable.

Ex1 (X, Y, Z) follows the multivariate normal distribution with zero mean vector μ and covariance matrix $\sum = \begin{pmatrix} 1 & 0.36 & 0.6 \\ 0.36 & 1 & 0.6 \\ 0.6 & 0.6 & 1 \end{pmatrix}$.

Ex2 X_1, Y_1, Z are *i.i.d.* random variables from the binomial distribution $B(10, 0.5)$. Define $X = X_1 + Z$ and $Y = Y_1 + Z$.

Ex3 $X_1, Y_1, Z \overset{i.i.d.}{\sim} N(0, 1)$, and define

$$Z_1=0.5(Z^3/7+Z/2),\ Z_2=(Z^3/2+Z)/3,$$
$$X_2=Z_1+\tanh X_1,\ X=X_2+X_2^3/3,$$
$$Y_2=Z_2+Y_1,\ Y=Y_2+\tanh(Y_2/3).$$

Therefore X and Y are conditionally independent given Z.

Ex4 $X_1, Y_1, Z_1, Z_2 \overset{i.i.d.}{\sim} B(10, 0.5)$, and define $X = X_1 + Z_1 + Z_2$ and $Y = Y_1 + Z_1 + Z_2$, $Z = (Z_1, Z_2)$.

Table 1 presents the type I error rates for Examples 1–4. From this table, we can see that the empirical type I error rates of the CDIT and KCI.test are under reasonable control. However, the CI.test is out of control in Example 3, which in fact contributes to an inflated power in Example 7.

Ex5 (X, Y, Z) has the multivariate normal distribution with zero mean vector μ and covariance matrix $\sum=\begin{pmatrix} 1 & 0.7 & 0.6 \\ 0.7 & 1 & 0.6 \\ 0.6 & 0.6 & 1 \end{pmatrix}$. Therefore, the conditional covariance matrix of X and Y given Z is

$$\sum(X,Y|Z)=\sum_{11}-\sum_{12}\sum_{22}^{-1}\sum_{21}=\begin{pmatrix} 0.64 & 0.34 \\ 0.34 & 0.64 \end{pmatrix}.$$

Ex6 $X_1, Z \overset{i.i.d.}{\sim} B(10, 0.5)$, and define $X = X_1 + Z$, $Y = (X_1 - 5)^4 + Z$.

Ex7 $X_1, Y_1, Z, \varepsilon \overset{i.i.d.}{\sim} N(0, 1)$, and define

$$Z_1=0.5(Z^3/7+Z/2),\ Z_2=(Z^3/2+Z)/3,$$
$$X_2=Z_1+\tanh X_1,\ X_3=X_2+X_2^3/3,$$
$$Y_2=Z_2+Y_1,\ Y_3=Y_2+\tanh(Y_2/3).$$

We standardize X_3, Y_3 and define

$$X=X_3+\cosh\varepsilon,\ Y=Y_3+\cosh\varepsilon^2.$$

Therefore X and Y are not conditionally independent given Z.

Ex8 $X_1, Z_1, Z_2 \overset{i.i.d.}{\sim} B(10, 0.5)$, and define $X = X_1 + Z_1 + Z_2$ and $Y = (X_1 - 5)^4 + Z_1 + Z_2$, $Z = (Z_1, Z_2)$.

Ex9 Supposed $Z_1, \cdots, Z_6 \overset{i.i.d.}{\sim} t(1)$, the t-distribution with degree freedom 1, and define

$$X_i=Z_i, i=1,2,3, \quad X_4=Z_4+Z_5,$$
$$Y=Z_1+Z_2+Z_3+Z_4+Z_5+Z_6.$$

Therefore $X = (X_1, X_2, X_3, X_4)$ and Y are not conditionally independent given $Z = Z_5$.

Ex10 Supposed $Z_1, \cdots, Z_{13} \overset{i.i.d.}{\sim} t(1)$, and define

$$X_i = Z_i, i = 1, 2, \cdots, 9, \quad X_{10} = Z_{10} + Z_{11};$$
$$Y_1 = Z_1 Z_2 + Z_3 Z_4 + Z_5 Z_{11} + Z_{12},$$
$$Y_2 = Z_6 Z_7 + Z_8 Z_9 + Z_{10} Z_{11} + Z_{13}.$$

Therefore $X = (X_1, \cdots X_{10})$ and $Y = (Y_1, Y_2)$ are not conditionally independent given $Z = Z_{11}$.

Ex11 Supposed $Z_1, \cdots, Z_4 \overset{i.i.d.}{\sim} t(2)$, and define

$$X_i = Z_i, i = 1, 2, 3, 4,$$
$$Y_1 = \sin Z_1 + \cos Z_2 + Z_3^2 + Z_4^2;$$
$$Y_2 = Z_1^2 + Z_2^2 + Z_3 + Z_4.$$

Therefore $X = (X_1, X_2, X_3, X_4)$ and $Y = (Y_1, Y_2)$ are not conditionally independent given $Z = (Z_1, Z_2)$.

Ex12 Supposed $Z_1, \cdots, Z_4 \overset{i.i.d.}{\sim} t(2)$, and define

$$X_i = Z_i, i = 1, 2, 3, 4,$$
$$Y_1 = Z_1 Z_2 + Z_3^3 + Z_4^2,$$
$$Y_2 = Z_1^3 + Z_2^2 + Z_3 Z_4.$$

Therefore $X = (X_1, X_2, X_3, X_4)$ and $Y = (Y_1, Y_2)$ are not conditionally independent given $Z = (Z_1, Z_2)$.

We can see from Table 2 that CI.test is the most powerful test for Example 5, due to the fact that the data are the generated from a multivariate normal distribution. Unlike KCI, the CDIT performs closely to CI.test and becomes closer as the sample size increases. In the nonlinear cases, the CDIT performs the best. Importantly, it has competitive or better performance than KCI when X, Y and Z are all multivariate, whereas CI.test are not applicable.

6. APPLICATION ON GENE REGULATION

In this section, we use our CDIT to re-analyze the data reported in Scheetz et al. (2006). The data were collected to study gene regulations in the mammalian eye, including 18976 probes of sufficient signal among 120 rats. Previous research documented that the mutations in gene *IMPG1* can cause the macular dystrophies (Manes et al., 2013) and gene *CFH* is important for age-related macular dystrophies (Hageman et al., 2005; Haines et al., 2005). Moreover, a significant association between *IMPG1* and *CFH* was reported (Dcor(*IMPG1*, *CFH*)=0.3105, p-value=10^{-6}), but it has not be investigated which gene remains associated with *IMPG1* beyond the known important role of *CFH*.

To this end, we treat the signal from *IMPG1* as the response variable and the signals from the other probes as covariates. *CFH* serves as the conditional variable. Bonferroni correction is used and 2.6×10^{-6} significant level is set ($0.05/18976 \approx 2.6 \times 10^{-6}$). Under this criterion, the p-value corresponding to gene *LOC*361100 is smaller than the threshold.

Table 3 indicates that *IMPG1* and *LOC361100* are significantly associated given *CFH* after the Bonferroni correction (the p-values is less than 0.05 after the correction), even though their marginal association is not.

To verify this finding, we fit the model

$$Y_{IMPG1} - g(X_{CFH}) = \beta_0 + \beta_1 (X_{LOC361100} - h(X_{CFH})) + \varepsilon,$$

where $g(\cdot)$ and $h(\cdot)$ are the fitted generalized additive models. Thus the test for the conditional dependence becomes the test for $\beta_1 = 0$. Denote

$$e(IMPG1|CFH) \triangleq Y_{IMPG1} - g(X_{CFH}),$$
$$e(LOC361100|CFH) \triangleq X_{LOC361100} - h(X_{CFH}).$$

$\hat{\beta_1} = 0.3966$, the p-value of H_0: $\beta_1 = 0$ is 2.786×10^{-6}. This result suggests a significant dependence between *LOC361100* and *IMPG1*, indicating that gene *LOC361100* may have latent relationship with macular dystrophies. Fig 1 displays the dependence between $e(IMPG1|CFH)$ and $e(LOC361100|CFH)$, as indicated by the increasing regression line.

It has been reported that the *LOC361100* gene expresses in tissues including kidney, lung, brain and eyes. According to Tsutsui et al. (2001, 1993), it participates in the process of DNA topological change and brain development. However, we did not find any reported role of *LOC361100* in the macular dystrophies. For this reason, the significant conditional association detected here by CDIT warrants further investigation and replication.

7. DISCUSSION

As discussed above, inference on conditional association is an important topic in scientific research. In addition to our earlier introduction, Li and his colleagues (Li, 2002), for example, introduced liquid association to evaluate the association of two gene expressions X and Y, given gene Z expression. Because liquid association assumes the normality of *Z* and is only sensitive to linear dependence, its application can be limited. In contrast, our CDIT is distribution free, sensitive to linear and non-linear dependence, and applicable to multivariate random variables. Therefore, CDIT is potentially useful for detecting the functional association between groups of genes.

Our conditional distance correlation can not only measure the nonlinear conditional correlation between two random variables given a third random variable, but also allow the arbitrary dimensions of random variables. Moreover, it captures the adaptive change of the conditional correlation dependent on the conditional random variable.

We presented theoretically desirable properties for our conditional distance correlation, and introduced two sample versions of our conditional distance correlation to simplify the computation. Importantly, these sample forms have the same asymptotic limit. Built on these desirable results, we proposed and demonstrated CDIT as a powerful test for conditional dependence. In the present work, we incorporated the concept of weights into the conditional dependence to construct the sample version, taking into account the conditional variable. Furthermore, our theoretical results make use of mainly V-statistic and large sample theory.

Despite the useful progress reported above, there remain important open issues. For example, we adopt Euclidean distance for simplicity, but the approach can be extended for α-distance: $(d_{ij}^X)^\alpha = \|X_i - X_j\|_p^\alpha$ and $(d_{ij}^Y)^\alpha = \|Y_i - Y_j\|_q^\alpha$, $0 < \alpha < 2$. A proper α may potentially improve the power with the type-I error under control. Morever, the power of our test diminishes dramatically as the dimensions of the random variables increase. While the dimensions of the random variables are expected to reduce the power, it would be useful to limit the power loss.

One limitation of our test is that it depends on a kernel function, which requires the selection of a bandwidth. The choice of the bandwidth is a challenging problem, and has been a difficult and active research topic on its own. It is beyond the scope of this article to address the problem of bandwidth selection. Instead, we took a practical approach after considering several existing choices. We used the Rule-of-Thumb method (Wand and Jones, 1995) because it was convienient for our numerical studies, and performed well. However, it could be a worthy effort to consider different methods. We also add that we presented asymptotic theory for various statistics based on the conditional distance. While it offers insight into the behavior of the important statistics, the theoretical result may not necessarily practical due to computational complexity and sample size limitation. Sample re-use methods such as bootstrapping may provide a useful alternative.

Acknowledgments

Wang's research is partially supported by National Science Foundation of China (NSFC) for Excellent Young Scholar (11322108), NCET(12-0559), NSFC(11001280), and RFDP(20110171110037). Zhang's research is partially supported by U.S. National Institute on Drug Abuse (R01 DA016750), a 1000-plan scholarship from Chinese Government, and International Collaborative Research Fund from NSFC(11328103).

References

Ackley H, Hinton E, Sejnowski J. A learning algorithm for Boltzmann machines. Cognitive Science. 1985:147–169.

Fan Y, Li Q. Consistent model specification tests: omitted variables and semiparametric functional forms. Econometrica: Journal of the Econometric Society. 1996:865–890.

Fukumizu K, Gretton A, Sun X, Schölkopf B. Kernel measures of conditional dependence. Conference on Neural Information Processing Systems. 2008

Gretton A, Bousquet O, Smola A, Schölkopf B. Measuring statistical dependence with Hilbert-Schmidt norms. Proceedings Algorithmic Learning Theory. 2005:63–77.

Hageman GS, Anderson DH, Johnson LV, Hancox LS, Taiber AJ, Hardisty LI, Hageman JL, Stockman HA, Borchardt JD, Gehrs KM, et al. A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102(20):7227–7232. [PubMed: 15870199]

Haines JL, Hauser MA, Schmidt S, Scott WK, Olson LM, Gallins P, Spencer KL, Kwan SY, Noureddine M, Gilbert JR, et al. Complement factor H 41 variant increases the risk of age-related macular degeneration. Science. 2005; 308(5720):419–421. [PubMed: 15761120]

Hall P. Central limit theorem for integrated square error of multivariate nonparametric density estimators. Journal of multivariate analysis. 1984; 14(1):1–16.

Huang T. Testing conditional independence using maximal nonlinear conditional correlation. The Annals of Statistics. 2010; 38(4):2047–2091.

Lawrance A. On conditional and partial correlation. The American Statistician. 1976; 30(3):146–149.

Lee, A. Statistics: Textbooks and Monographs M. Dekker. 1990. U-Statistics: Theory and Practice.

Li K-C. Genome-wide coexpression dynamics: Theory and application. Proceedings of The National Academy of Sciences. 2002; 991:16875–16880.

Li, Q.; Racine, JS. Nonparametric econometrics: Theory and practice. Princeton University Press; 2007.

Linton O, Gozalo P. Conditional Independence Restrictions: Testing and Estimation. Cowles Foundation Discussion Papers. 1996:1140–1186.

Manes G, Meunier I, Avila-Fernández A, Banfi S, Le Meur G, Zanlonghi X, Corton M, Simonelli F, Brabet P, Labesse G, et al. Mutations in IMPG1 Cause Vitelliform Macular Dystrophies. The American Journal of Human Genetics. 2013; 93(3):571–578. [PubMed: 23993198]

Paparoditis E, Politis D. The local bootstrap for kernel estimators under general dependence conditions. Annals of the Institute of Statistical Mathematics. 2000; 52(1):139–159.

Scheetz TE, Kim K-YA, Swiderski RE, Philp AR, Braun TA, Knudtson KL, Dorrance AM, DiBona GF, Huang J, Casavant TL, et al. Regulation of gene expression in the mammalian eye and its relevance to eye disease. Proceedings of the National Academy of Sciences. 2006; 103(39):14429–14434.

Schick A. On U-statistics with random kernels. Statistics & probability letters. 1997; 34(3):275–283.

Scutari M. Learning Bayesian networks with the bnlearn R package. 2009 arXiv preprint arXiv: 0908.3817.

Speed T. A Correlation for the 21st Century. science. 2011; 334(6062):1502–1503. [PubMed: 22174235]

Su L, White H. Testing Conditional Independence Via Empirical Likelihood. University of California at San Diego, Economics Working Paper Series. 2003

Su L, White H. A consistent characteristic function-based test for conditional independence. Journal of Econometrics. 2007; 141(2):807–834.

Su L, White H. A nonparametric Hellinger metric test for conditional independence. Econometric Theory. 2008; 24(4):829.

Székely G, Rizzo M. Hierarchical clustering via joint between-within distances: Extending Ward's minimum variance method. Journal of classification. 2005; 22(2):151–183.

Székely G, Rizzo M, Bakirov N. Measuring and testing dependence by correlation of distances. The Annals of Statistics. 2007; 35(6):2769–2794.

Tsutsui K, Okada S, Watanabe M, Shohmori T, Seki S, Inoue Y. Molecular cloning of partial cDNAs for rat DNA topoisomerase II isoforms and their differential expression in brain development. Journal of Biological Chemistry. 1993; 268(25):19076–19083. [PubMed: 8395528]

Tsutsui K, Sano K, Kikuchi A, Tokunaga A. Involvement of DNA topoisomerase IIbeta in neuronal differentiation. The Journal of Biological Chemistry. 2001; 276(8):5769–78. [PubMed: 11106659]

Wand M, Jones M. Multivariate plug-in bandwidth selection. Computational Statistics. 1994; 9(2):97–116.

Wand, M.; Jones, M. Kernel Smoothing. Chapman & Hall; 1995.

Zhang X, Zhao X, He K, Lu L, Cao Y, Liu J, Hao J, Liu Z, Chen L. Inferring gene regulatory networks from gene expression data by path consistency algorithm based on conditional mutual information. Bioinformatics. 2012; 28(1):98–104. [PubMed: 22088843]

APPENDIX A. TECHNICAL DETAILS

We begin with Lemma 2 from Székely and Rizzo (2005) for the convenience of presenting our proofs.

Lemma 2

If $0 < \alpha < 2$, then for all x in $\mathbb{R}^{\beta}$

$$\int_{\mathbb{R}^{\beta}} \frac{1-\cos\langle t,x\rangle}{|t|_{\beta}^{\beta+\alpha}} dt = C(\beta,\alpha)|x|_{\beta}^{\alpha},$$

where

$$C(\beta,\alpha) = \frac{2\pi^{\beta/2}\Gamma(1-\alpha/2)}{\alpha 2^{\alpha}\Gamma((\alpha+\beta)/2)}.$$

and $\Gamma(\cdot)$ is the complete gamma function. The integrals at 0 and ∞ are meant in the principal value sense: $lim_{\varepsilon\to 0}\int_{\mathbb{R}^{\beta}/\{\varepsilon B+\varepsilon^{-1}B^{c}\}}$, where B is the unit ball (centered at 0) in $\mathbb{R}^{\beta}$ and B^c is the complement of B.

In the following proofs, we take $\alpha = 1$ for clarity.

A.1 Proof of Theorem 1

- **i.** Statement (i) is obvious.
- **ii.** If $\mathcal{D}(X|Z) = 0$, then $\varphi_{X,X|Z}(t, s) = \varphi_{X|Z}(t)\varphi_{X|Z}(s)$ for all t, s; that is, X is conditionally independent of X given Z. This suggests that $\forall A \in \sigma(X)$, $P(A \cap A|Z) = P(A|Z)P(A|Z)$, then $P(A|Z) = 1$, or $P(A|Z) = 0$.
  1. If $\forall B \in \sigma(Z)$, $P(A \cap B) = P(B)$ and $P(A \cap B^c) = P(B^c)$, then

     $$P(A\cap B)+P(A\cap B^c)=P(B)+P(B^c).$$

     Thus $P(A) = 1$, and $A \in \sigma(Z)$ follows from the completeness.
  2. If $\forall B \in \sigma(Z)$, $P(A \cap B) = 0$ and $P(A \cap B^c) = 0$, then

     $$P(A\cap B)+P(A\cap B^c)=0.$$

     Thus $P(A) = 0$, and $A \in \sigma(Z)$ follows from the completeness.
  3. If $\exists B \in \sigma(Z)$, $P(A \cap B) = P(B)$ and $P(A \cap B^c) = 0$, then

     $$P(B\backslash A)=P(B)-P(A\cap B)=0.$$

and

$$P(A\backslash B)=P(A\cap B^c)=0.$$

Thus $P(A\triangle B) \le P(A\backslash B)+P(B\backslash A) = 0$, and $A \in \sigma(Z)$ follows from the completeness.

Therefore, $\sigma(X) \subset \sigma(Z)$, and there exists a Borel function f, such that $X = f(Z)$ and $E(X|Z) = E(f(Z)|Z) = f(Z) = X$.

iii.

$$\begin{aligned}&\mathscr{D}^2(a_1+b_1C_1X, a_2+b_2C_2Y|Z)\\ &=\int_{\mathbb{R}^{p+q}}\frac{|\phi_{a_1+b_1C_1X,a_2+b_2C_2Y|Z}(t,s)-\phi_{a_1+b_1C_1X|Z}(t)\phi_{a_2+b_2C_2Y|Z}(s)|^2}{c_pc_q|t|_p^{p+1}|s|_q^{q+1}}\mathrm{d}t\mathrm{d}s\\ &=\int_{\mathbb{R}^{p+q}}\frac{|e^{i\langle t,a_1\rangle+i\langle s,a_2\rangle}(\phi_{b_1C_1X,b_2C_2Y|Z}(t,s)-\phi_{b_1C_1X|Z}(t)\phi_{b_2C_2Y|Z}(s))|^2}{c_pc_q|t|_p^{p+1}|s|_q^{q+1}}\mathrm{d}t\mathrm{d}s.\end{aligned}$$

Denote $t' = b_1C_1t$ and $s' = b_2C_2s$. Then,

$$\begin{aligned}&\mathscr{D}^2(a_1+b_1C_1X, a_2+b_2C_2Y|Z)\\ &=|b_1||b_2|\int_{\mathbb{R}^p}\int_{\mathbb{R}^q}\frac{\phi_{X,Y|Z}(t',s')-\phi_{X|Z}(t',s')\phi_{Y|Z}(t',s')}{c_pc_q|t'|_p^{p+1}|s'|_q^{q+1}}\mathrm{d}t'\mathrm{d}s'\\ &=|b_1||b_2|\mathscr{D}^2(X,Y|Z).\end{aligned}$$

iv. This follows from (iii) for $a_1 = a_2 = a$, $b_1 = b_2 = b$ and $Y = X$.

v. Since random vectors (X_1, Y_1) and (X_2, Y_2) are conditionally independent given Z and the norm of a characteristic function cannot be greater than 1, we have

$$\begin{aligned}&\mathscr{D}(X_1+X_2, Y_1+Y_2|Z)\\ &=\|\phi_{X_1+X_2,Y_1+Y_2|Z}(t,s)-\phi_{X_1+X_2|Z}(t)\phi_{Y_1+Y_2|Z}(s)\|\\ &=\|\phi_{X_1,Y_1|Z}(t,s)\phi_{X_2,Y_2|Z}(t,s)-\phi_{X_1|Z}(t)\phi_{X_2|Z}(t)\phi_{Y_1|Z}(s)\phi_{Y_2|Z}(s)\|\\ &\le \|\phi_{X_1,Y_1|Z}(t,s)(\phi_{X_2,Y_2|Z}(t,s)-\phi_{X_2|Z}(t)\phi_{Y_2|Z}(s))\|\\ &+\|\phi_{X_2|Z}(t)\phi_{Y_2|Z}(s)(\phi_{X_1,Y_1|Z}(t,s)-\phi_{X_1|Z}(t)\phi_{Y_1|Z}(s))\|\\ &\le \|\phi_{X_2,Y_2|Z}(t,s)-\phi_{X_2|Z}(t)\phi_{Y_2|Z}(s)\|+\|\phi_{X_1,Y_1|Z}(t,s)-\phi_{X_1|Z}(t)\phi_{Y_1|Z}(s)\|\\ &=\mathscr{D}(X_1,Y_1|Z)+\mathscr{D}(X_2,Y_2|Z).\end{aligned}$$

Note that $\mathcal{D}(X_1+X_2, Y_1+Y_2|Z) = \mathcal{D}(X_1, Y_1|Z)+ \mathcal{D}(X_2, Y_2|Z)$ holds if and only if all equations above hold; or equivalently to (1) X_1 and Y_1 are both the function of Z, (2) X_2 and Y_2 are both the function of Z, or (3) X_1, X_2, Y_1, Y_2 are mutually conditionally independent given Z.

vi. This is a special case of (v) for $X_1 = Y_1 = X$ and $X_2 = Y_2 = Y$.

A.2 Proof of Theorem 2

i. By the Cauchy-Schwarz inequality,

$$\begin{aligned}0 &\le \int_{\mathbb{R}^{p+q}} \frac{|\phi_{X,Y|Z}(t,s)-\phi_{X|Z}(t)\phi_{Y|Z}(s)|^2}{c_pc_q|t|_p^{p+1}|s|_q^{q+1}}\mathrm{d}t\mathrm{d}s\\ &= \int_{\mathbb{R}^{p+q}} \frac{[E((\exp(i\langle t,X\rangle)-\phi_{X|Z}(t))(\exp(i\langle s,Y\rangle)-\phi_{Y|Z}(s))|Z)]^2}{c_pc_q|t|_p^{p+1}|s|_q^{q+1}}\mathrm{d}t\mathrm{d}s\\ &\le \int_{\mathbb{R}^{p+q}} \frac{E[(\exp(i\langle t,X\rangle)-\phi_{X|Z}(t))|Z]^2E[(\exp(i\langle s,Y\rangle)-\phi_{Y|Z}(s))|Z]^2}{c_pc_q|t|_p^{p+1}|s|_q^{q+1}}\mathrm{d}t\mathrm{d}s\\ &= \int_{\mathbb{R}^{p+q}} \frac{(1-|\phi_{X|Z}(t)|^2)(1-|\phi_{Y|Z}(s)|^2)}{c_pc_q|t|_p^{p+1}|s|_q^{q+1}}\mathrm{d}t\mathrm{d}s.\end{aligned}$$

This implies that $0 \le \rho(X, Y|Z) \le 1$.

ii. It directly follows from (iii) of Theorem 1.

A.3 Proof of Lemma 1

First, given the event $Z = z$, we consider

$$\begin{aligned}&|\phi_{X,Y|Z=z}(t,s)-\phi_{X|Z=z}(t)\phi_{Y|Z=z}(s)|^2\\ &=\phi_{X,Y|Z=z}(t,s)\overline{\phi_{X,Y|Z=z}(t,s)}+\phi_{X|Z=z}(t)\phi_{Y|Z=z}(s)\overline{\phi_{X|Z=z}(t)\phi_{Y|Z=z}(s)}\\ &-\phi_{X,Y|Z=z}(t,s)\overline{\phi_{X|Z=z}(t)\phi_{Y|Z=z}(s)}-\phi_{X|Z=z}(t)\phi_{Y|Z=z}(s)\overline{\phi_{X,Y|Z=z}(t,s)}\\ &=E[\exp(i\langle t,X_1-X_2\rangle+i\langle s,Y_1-Y_2\rangle)|Z_1=z,Z_2=z]\\ &+E[\exp(i\langle t,X_1-X_2\rangle)|Z_1=z,Z_2=z]E[\exp(i\langle s,Y_1-Y_2\rangle)|Z_1=z,Z_2=z]\\ &-2E[\exp(i\langle t,X_1-X_2\rangle+i\langle s,Y_1-Y_3\rangle)|Z_1=z,Z_2=z,Z_3=z].\end{aligned}$$

Using the following equation

$$\exp(i(u+v))=1-(1-\exp(iu))-(1-\exp(iv))+(1-\exp(iu))(1-\exp(iv)),$$

we can get that

$$\begin{aligned}&|\phi_{X,Y|Z=z}(t,s)-\phi_{X|Z=z}(t)\phi_{Y|Z=z}(s)|^2\\ &=E[(1-\exp(i\langle t,X_1-X_2\rangle))(1-\exp(i\langle s,Y_1-Y_2\rangle))|Z_1=z,Z_2=z]\\ &+E[(1-\exp(i\langle t,X_1-X_2\rangle))|Z_1=z,Z_2=z]E[(1-\exp(i\langle s,Y_1-Y_2\rangle))|Z_1=z,Z_2=z]\\ &-2E[(1-\exp(i\langle t,X_1-X_2\rangle))(1-\exp(i\langle s,Y_1-Y_3\rangle))|Z_1=z,Z_2=z,Z_3=z].\end{aligned}$$

According to Lemma 2, we have

$$\begin{aligned}&\mathscr{D}^2(X,Y|Z=z)\\ &=\|\phi_{X,Y|Z=z}(t,s)-\phi_{X|Z=z}(t)\phi_{Y|Z=z}(s)\|^2\\ &=\frac{1}{c_pc_q}\int_{\mathbb{R}^{p+q}}\frac{|\phi_{X,Y|Z=z}(t,s)-\phi_{X|Z=z}(t)\phi_{Y|Z=z}(s)|^2}{|t|_p^{p+1}|s|_q^{q+1}}\mathrm{d}t\mathrm{d}s\\ &=E[d_{12}^Xd_{12}^Y|Z_1=z,Z_2=z]+E[d_{12}^X|Z_1=z,Z_2=z]E[d_{12}^Y|Z_1=z,Z_2=z]\\ &-2E[d_{12}^Xd_{13}^Y|Z_1=z,Z_2=z,Z_3=z].\end{aligned}$$

By the definition of d_{ijkl}:

$$d_{ijkl}=(d^X_{ij}+d^X_{kl}-d^X_{ik}-d^X_{jl})(d^Y_{ij}+d^Y_{kl}-d^Y_{ik}-d^Y_{jl})$$

and d^s_{ijkl}:

$$d^s_{ijkl}=d_{ijkl}+d_{ijlk}+d_{ilkj}.$$

We can verify the following the identities after some algebra.

$$\begin{aligned}&\tfrac{1}{12}E[d^s_{1234}|Z_1=z,Z_2=z,Z_3=z,Z_4=z]\\&=\tfrac{1}{4}E[d_{1234}|Z_1=z,Z_2=z,Z_3=z,Z_4=z]\\&=E[d^X_{12}d^Y_{12}|Z_1=z,Z_2=z]+E[d^X_{12}|Z_1=z,Z_2=z]E[d^Y_{12}|Z_1=z,Z_2=z]\\&-2E[d^X_{12}d^Y_{13}|Z_1=z,Z_2=z,Z_3=z].\end{aligned}$$

Therefore,

$$\mathscr{D}^2(X,Y|Z=z)=\frac{1}{12}E[d^s_{1234}|Z_1=z,Z_2=z,Z_3=z,Z_4=z].$$

A.4 Proof of Theorem 3

Lemma 2 implies that there exist constants c_p and c_q such that for all X in $\mathbb{R}^p$, Y in $\mathbb{R}^q$,

$$\begin{aligned}&\int_{\mathbb{R}^p}\frac{1-\exp(i\langle t,X\rangle)}{|t|_p^{p+1}}\mathrm{d}t=c_p|X|_p,\\&\int_{\mathbb{R}^q}\frac{1-\exp(i\langle s,Y\rangle)}{|s|_q^{q+1}}\mathrm{d}s=c_q|Y|_q,\end{aligned}$$

where the integrals are considered in the sense of principal value.

We first simplify pSCDCov and prove the following algebraic identity:

$$\begin{aligned}\mathscr{D}^2_n(\mathbf{W}_n|Z)&=\|\phi^n_{X,Y|Z}(t,s)-\phi^n_{X|Z}(t)\phi^n_{Y|Z}(s)\|^2\\&=D_1+D_2-2D_3,\end{aligned}\quad\text{(A.1)}$$

where

$$D_1=\frac{1}{\omega^2(Z)}\sum_{k,l=1}^{n}d_{kl}^{X}d_{kl}^{Y}\omega_k(Z)\omega_l(Z),$$
$$D_2=\frac{1}{\omega^4(Z)}\sum_{k,l=1}^{n}d_{kl}^{X}\omega_k(Z)\omega_l(Z)\sum_{k,l=1}^{n}d_{kl}^{Y}\omega_k(Z)\omega_l(Z),$$
$$D_3=\frac{1}{\omega^3(Z)}\sum_{k,l,m=1}^{n}d_{kl}^{X}d_{km}^{Y}\omega_k(Z)\omega_l(Z)\omega_m(Z).$$

Note that $|\phi^n_{X,Y|Z}(t,s)|^2$, $|\phi^n_{X|Z}(t)\phi^n_{Y|Z}(s)|^2$ and $\phi^n_{X,Y|Z}(t,s)\overline{\phi^n_{X|Z}(t)\phi^n_{Y|Z}(s)}$ are involved in the derivation of the pSCDCov. Each of them can be expressed as the sum of a V-process with a random kernel and remainder as follows:

$$\begin{aligned}&\phi^n_{X,Y|Z}(t,s)\overline{\phi^n_{X,Y|Z}(t,s)}\\&=\frac{\sum_{k,l=1}^{n}\omega_k(Z)\omega_l(Z)\cos\langle X_k-X_l,t\rangle\cos\langle Y_k-Y_l,s\rangle}{(\omega(Z))^2}+V_1,\\&\phi^n_{X|Z}(t)\phi^n_{Y|Z}(s)\overline{\phi^n_{X|Z}(t)\phi^n_{Y|Z}(s)}\\&=\frac{\sum_{k,l=1}^{n}\omega_k(Z)\omega_l(Z)\cos\langle X_k-X_l,t\rangle}{(\omega(Z))^2}\frac{\sum_{k,l=1}^{n}\omega_k(Z)\omega_l(Z)\cos\langle Y_k-Y_l,s\rangle}{(\omega(Z))^2}+V_2,\end{aligned}$$

and

$$\begin{aligned}&\phi^n_{X,Y|Z}(t,s)\overline{\phi^n_{X|Z}(t)\phi^n_{Y|Z}(s)}\\&=\frac{\sum_{k,l,m=1}^{n}\omega_k(Z)\omega_l(Z)\omega_m(Z)\cos\langle X_k-X_l,t\rangle\cos\langle Y_k-Y_l,s\rangle}{(\omega(Z))^3}+V_3,\end{aligned}$$

where the integrals of V_1, V_2 and V_3 equal to zero.

Note that

$$\cos u\cos v=1-(1-\cos u)-(1-\cos v)+(1-\cos u)(1-\cos v).$$

We only need to calculate the following integral:

$$\begin{aligned}&\int_{\mathbb{R}^{p+q}}\frac{(1-\cos\langle X_k-X_l,t\rangle)(1-\cos\langle Y_k-Y_l,s\rangle)}{|t|_p^{p+1}|s|_q^{q+1}}\mathrm{d}t\mathrm{d}s\\&=\int_{\mathbb{R}^p}\frac{(1-\cos\langle X_k-X_l,t\rangle)}{|t|_p^{p+1}}\mathrm{d}t\int_{\mathbb{R}^q}\frac{(1-\cos\langle Y_k-Y_l,s\rangle)}{|s|_q^{q+1}}\mathrm{d}s\\&=c_pc_qd_{kl}^{X}d_{kl}^{Y}.\end{aligned}$$

The above notation is given in Section 3. Thus (A.1) holds.

Next, we verify the following algebraic identity:

$$\mathscr{V}_n(\mathbf{W}_n|Z)=D_1+D_2-2D_3. \quad \text{(A.2)}$$

By the definition of the pSCDCov in Definition 3 and denoting the numerator of $\mathscr{V}_n(\mathbf{W}_n|Z)$ by

$$\varphi_n(\mathbf{W}_i, \mathbf{W}_j, \mathbf{W}_k, \mathbf{W}_l; Z) = \omega_i(Z)\omega_j(Z)\omega_k(Z)\omega_l(Z) d_{ijkl}^s,$$

we have

$$\begin{aligned}
&\sum_{i,j,k,l} \varphi_n(\mathbf{W}_i, \mathbf{W}_j, \mathbf{W}_k, \mathbf{W}_l; Z) \\
&= 3 \sum_{i,j,k,l} \omega_i(Z)\omega_j(Z)\omega_k(Z)\omega_l(Z) d_{ijkl} \\
&= 3[\sum_{i,j,k,l} (d_{ij}^X d_{ij}^Y + d_{kl}^X d_{kl}^Y + d_{ik}^X d_{ik}^Y + d_{jl}^X d_{jl}^Y)\omega_i(Z)\omega_j(Z)\omega_k(Z)\omega_l(Z) \\
&+ \sum_{i,j,k,l} (d_{ij}^X d_{kl}^Y + d_{ij}^X d_{kl}^Y + d_{ik}^X d_{jl}^Y + d_{jl}^X d_{ik}^Y)\omega_i(Z)\omega_j(Z)\omega_k(Z)\omega_l(Z) \\
&- \sum_{i,j,k,l} (d_{ij}^X d_{ik}^Y + d_{ij}^X d_{jl}^Y + d_{kl}^X d_{ik}^Y + d_{kl}^X d_{jl}^Y \\
&+ d_{ik}^X d_{ij}^Y + d_{ik}^X d_{kl}^Y + d_{jl}^X d_{ij}^Y + d_{jl}^X d_{kl}^Y)\omega_i(Z)\omega_j(Z)\omega_k(Z)\omega_l(Z)] \\
&\triangleq T_1 + T_2 + T_3.
\end{aligned}$$

Using the symmetry of distance $d_{ij}^X = d_{ji}^X$ and $d_{ij}^Y = d_{ji}^Y$, the three terms in right hand side of the equation can be simplified as follows:

$$\begin{aligned}
T_1 &= 12\omega^2(Z)\sum_{i,j} d_{ij}^X d_{ij}^Y \omega_i(Z)\omega_j(Z), \\
T_2 &= 12\sum_{i,j} d_{ij}^X \omega_i(Z)\omega_j(Z)\sum_{k,l} d_{kl}^Y \omega_k(Z)\omega_l(Z), \\
T_3 &= 24\omega(Z)\sum_i \sum_{k,l} d_{ij}^X d_{ik}^Y \omega_i(Z)\omega_j(Z)\omega_k(Z).
\end{aligned}$$

Thus,

$$\begin{aligned}
&\sum_{i,j,k,l} \varphi_n(\mathbf{W}_i, \mathbf{W}_j, \mathbf{W}_k, \mathbf{W}_l; Z) \\
&= 12\omega^2(Z)\sum_{i,j} d_{ij}^X d_{ij}^Y \omega_i(Z)\omega_j(Z) + 12\sum_{i,j} d_{ij}^X \omega_i(Z)\omega_j(Z)\sum_{k,l} d_{kl}^Y \omega_k(Z)\omega_l(Z) \\
&- 24\omega(Z)\sum_i \sum_{k,l} d_{ij}^X d_{ik}^Y \omega_i(Z)\omega_j(Z)\omega_k(Z) \\
&= 12\omega^4(Z)(D_1 + D_2 - 2D_3).
\end{aligned}$$

We therefore obtain

$$\begin{aligned}
\mathscr{V}_n(\mathbf{W}_n|Z) &= \frac{1}{12\omega^4(Z)} \sum_{i,j,k,l} \varphi_n(\mathbf{W}_i, \mathbf{W}_j, \mathbf{W}_k, \mathbf{W}_l; Z) \\
&= D_1 + D_2 - 2D_3.
\end{aligned}$$

Thus, (A.1) and (A.2) imply that $\mathscr{D}_n^2(\mathbf{W}_n|Z) = \mathscr{V}_n(\mathbf{W}_n|Z)$.

A.5 Proof of Theorem 4

We first show that $\mathscr{D}_n(\mathbf{W}_n|Z) \xrightarrow{a.s} \mathscr{D}(X, Y|Z)$. Denote $u_k = \exp(i\langle t, X_k\rangle) - \varphi_{X|Z}(t)$ and $v_k = \exp(i\langle s, Y_k\rangle) - \varphi_{Y|Z}(s)$, and define $\xi_n(t, s)$ as follows:

$$\xi_n(t,s)=\phi^n_{X,Y|Z}(t,s)-\phi^n_{X|Z}(t)\phi^n_{Y|Z}(s)$$
$$=\frac{\sum_{k=1}^n u_k v_k \omega_k(Z)}{\omega(Z)}-\frac{\sum_{k=1}^n u_k \omega_k(Z)}{\omega(Z)}\frac{\sum_{k=1}^n v_k \omega_k(Z)}{\omega(Z)}.$$

Also define a region with respect to a given $\delta > 0$:

$$O(\delta)=\{(t,s):\delta \leq |t|_p \leq 1/\delta, \delta \leq |s|_q \leq 1/\delta\}$$

and random variables

$$\mathscr{D}^2_{n,\delta}(\mathbf{W}_n|Z)-\int_{O(\delta)}\frac{|\xi_n(t,s)|^2}{c_p c_q |t|_p^{p+1}|s|_q^{q+1}}\mathrm{d}t\mathrm{d}s.$$

According to the Lebesgue dominated convergence theorem, for each $\delta > 0$, it follows that almost surely,

$$\lim_{n\to\infty}\mathscr{D}^2_{n,\delta}=\mathscr{D}^2_{\cdot,\delta}=\int_{O(\delta)}\frac{|\phi_{X,Y|Z}(t,s)-\phi_{X|Z}(t)\phi_{Y|Z}(s)|^2}{c_p c_q |t|_p^{p+1}|s|_q^{q+1}}\mathrm{d}t\mathrm{d}s.$$

And $\mathscr{D}^2_{\cdot,\delta}$ converges to $\mathcal{D}^2$ as δ tends to zero. Now it remains to prove that almost surely,

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}|\mathscr{D}^2_{n,\delta}-\mathscr{D}^2_n|=0. \quad \text{(A.3)}$$

Given $\delta > 0$,

$$|\mathscr{D}^2_{n,\delta}-\mathscr{D}^2_n| \leq \left(\int_{|t|_p<\delta}+\int_{|t|_p>1/\delta}+\int_{|s|_q<\delta}+\int_{|s|_q>1/\delta}\right)\frac{|\xi_n(t,s)|^2}{c_p c_q |t|_p^{p+1}|s|_q^{q+1}}\mathrm{d}s\mathrm{d}t. \quad \text{(A.4)}$$

We consider the first term in (A.4). For $r = (r_1, r_2, \cdots, r_p)$ in $\mathbb{R}^p$ define the function

$$G(y)=\int_{|r|<y}\frac{1-\cos r_1}{|r|_p^{p+1}}\mathrm{d}r.$$

According to Lemma 2, $G(y)$ is bounded by c_p and $\lim_{y\to\infty} G(y) = 0$. Applying the inequality $|x + y|^2 \le 2|x|^2 + 2|y|^2$ and the Cauchy-Schwarz inequality for sums, one can obtain that

$$
\begin{aligned}
&|\xi_n(t,s)|^2\\
&\le 2\left|\frac{\sum_{k=1}^n u_k v_k \omega_k(Z)}{\omega(Z)}\right|^2 + 2\left|\frac{\sum_{k=1}^n u_k \omega_k(Z)}{\omega(Z)}\right|^2 \left|\frac{\sum_{k=1}^n v_k \omega_k(Z)}{\omega(Z)}\right|^2\\
&= 2\frac{|\sum_{k=1}^n u_k\sqrt{\omega_k(Z)} v_k\sqrt{\omega_k(Z)}|^2}{(\omega(Z))^2}\\
&+ 2\frac{|\sum_{k=1}^n u_k\sqrt{\omega_k(Z)}\sqrt{\omega_k(Z)}|^2}{(\omega(Z))^2}\frac{|\sum_{k=1}^n v_k\sqrt{\omega_k(Z)}\sqrt{\omega_k(Z)}|^2}{(\omega(Z))^2}\\
&\le 4\frac{\sum_{k=1}^n |u_k|^2\omega_k(Z)}{\omega(Z)}\frac{\sum_{k=1}^n |v_k|^2\omega_k(Z)}{\omega(Z)}.
\end{aligned}
\tag{A.5}
$$

Note that

$$
\begin{aligned}
|v_k|^2 = &1 - \exp(i\langle s, Y_k\rangle)\overline{\phi_{Y|Z}(s)} + 1 - \exp(-i\langle s, Y_k\rangle)\phi_{Y|Z}(s)\\
&-(1 - |\phi_{Y|Z}(s)|^2),
\end{aligned}
$$

we have

$$
\begin{aligned}
\int_{\mathbb{R}^q}\frac{|v_k|^2}{c_q|s|_q^{q+1}}\mathrm{d}s &= (2E_Y(|Y_k - Y||Z) - E(|Y - Y'||Z))\\
&\le 2(|Y_k| + E(|Y||Z)),
\end{aligned}
$$

where $E_Y(\cdot|Z)$ is taken with respect to Y, and $Y' \overset{D}{=} Y$ is independent of Y_k. Moreover, after a suitable change of variables, we can obtain

$$
\begin{aligned}
&\int_{|t|_p<\delta}\frac{|u_k|^2}{c_p|t|_p^{p+1}}\mathrm{d}t\\
&\le 2E_X(|X_k - X|G(|X_k - X|\delta)|Z) - E(|X - X'|G(|X - X'|\delta)|Z)\\
&\le 2E_X(|X_k - X|G(|X_k - X|\delta)|Z),
\end{aligned}
$$

where $E_X(\cdot|Z)$ is taken with respect to X, and $X' \overset{D}{=} X$ is independent of X_k. Thus

$$
\begin{aligned}
&\int_{|t|_p<\delta}\frac{|\xi_n(t,s)|^2}{c_p c_q|t|_p^{p+1}|s|_q^{q+1}}\mathrm{d}t\mathrm{d}s\\
&\le 4\frac{\sum_{k=1}^n\int_{|t|_p<\delta}\frac{|u_k|^2}{c_p|t|_p^{p+1}}\mathrm{d}t\,\omega_k(Z)}{\omega(Z)}\frac{\sum_{k=1}^n\int_{\mathbb{R}^q}\frac{|v_k|^2}{c_q|s|_q^{q+1}}\mathrm{d}s\,\omega_k(Z)}{\omega(Z)}\\
&\le 16\frac{\sum_{k=1}^n(|Y_k| + E(|Y||Z))\omega_k(Z)}{\omega(Z)}\\
&\cdot\frac{\sum_{k=1}^n E_X(|X_k - X|G(|X_1 - X_2|\delta)|Z)\omega_k(Z)}{\omega(Z)}.
\end{aligned}
\tag{A.6}
$$

Since $\omega(Z)/n$ is a consistent density function estimator of Z,

$$\limsup_{n\to\infty}\int_{|t|_p<\delta}\frac{|\xi_n(t,s)|^2}{c_pc_q|t|_p^{p+1}|s|_q^{q+1}}\mathrm{d}t\mathrm{d}s \leq 32E(|Y||Z)E(|X_1-X_2|G(|X_1-X_2|\delta)|Z).$$

By Lebesgue dominated convergence theorem for conditional expectations,

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}\int_{|t|_p<\delta}\frac{|\xi_n(t,s)|^2}{c_pc_q|t|_p^{p+1}|s|_q^{q+1}}\mathrm{d}t\mathrm{d}s=0$$

almost surely.

Next, consider the second term in (A.4). Inequality (A.5) implies that $|u_k|^2 \leq 4$. Then,

$$\frac{1}{\omega(Z)}\sum_{k=1}^{n}\omega_k(Z)\int_{|t|_p>1/\delta}\frac{|u_k|^2}{c_p|t|_p^{p+1}}\mathrm{d}t \leq 4\int_{|t|_p>1/\delta}\frac{1}{c_p|t|_p^{1+p}}\mathrm{d}t.$$

Thus

$$\limsup_{n\to\infty}\int_{|t|_p>1/\delta}\frac{|\xi_n(t,s)|^2}{c_pc_q|t|_p^{p+1}|s|_q^{q+1}}\mathrm{d}t\mathrm{d}s \leq 16\delta\frac{\sum_{k=1}^{n}2(|Y_k|+E(|Y||Z))\omega_k(Z)}{\omega(Z)}.$$

We therefore have

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}\int_{|t|_p>1/\delta}\frac{|\xi_n(t,s)|^2}{c_pc_q|t|_p^{p+1}|s|_q^{q+1}}\mathrm{d}t\mathrm{d}s=0.$$

The remaining two terms in (A.4) can be dealt with similar to the first two terms, and likewise for (A.3).

According to Theorem 1 of Section 4.2 in Lee (1990), a V-type statistic can be simplified as follows:

$$\mathscr{V}_n(\mathbf{W}|Z)=\frac{(n-1)(n-2)(n-3)}{n^3}\mathscr{U}_n(\mathbf{W}|Z)+o_p(1).$$

Therefore

$$\mathscr{U}_n(\mathbf{W}_n|Z)\xrightarrow[n\to\infty]{P}\mathscr{D}^2(X,Y|Z).$$

A.6 Proof of Theorem 5

i. Consider the Euclidian distance in $\mathbb{R}^p$, with the orthogonal invariance of Euclidian distance, we have

$$\begin{aligned} d_{ij}^{a_1+b_1C_1X} &= d(a_1+b_1C_1X_i, a_1+b_1C_1X_j) \\ &= \|(a_1+b_1C_1X_i)-(a_1+b_1C_1X_j)\|_p \\ &= |b_1|\,\|X_i-X_j\|_p \\ &= |b_1| d_{ij}^X. \end{aligned}$$

By (A.2),

$$\begin{aligned} &\mathscr{D}_n^2(a_1+b_1C_1\mathbf{X}_n, a_2+b_2C_2\boldsymbol{Y}_n, \boldsymbol{Z}_n|Z) \\ &= \frac{1}{\omega^2(Z)}\sum_{i,j} d_{ij}^{a_1+b_1C_1X} d_{ij}^{a_2+b_2C_2Y}\omega_i(Z)\omega_j(Z) \\ &+ \frac{1}{\omega^4(Z)}\sum_{i,j} d_{ij}^{a_1+b_1C_1X}\omega_i(Z)\omega_j(Z)\sum_{k,l} d_{kl}^{a_2+b_2C_2Y}\omega_k(Z)\omega_l(Z) \\ &- \frac{2}{\omega^3(Z)}\sum_i\sum_{k,l} d_{ij}^{a_1+b_1C_1X} d_{ik}^{a_2+b_2C_2Y}\omega_i(Z)\omega_j(Z)\omega_k(Z) \\ &= |b_1b_2|\Big(\frac{1}{\omega^2(Z)}\sum_{i,j} d_{ij}^X d_{ij}^Y\omega_i(Z)\omega_j(Z) + \frac{1}{\omega^4(Z)}\sum_{i,j} d_{ij}^X\omega_i(Z)\omega_j(Z)\sum_{k,l} d_{kl}^Y\omega_k(Z)\omega_l(Z) \\ &- \frac{2}{\omega^3(Z)}\sum_i\sum_{k,l} d_{ij}^X d_{ik}^Y\omega_i(Z)\omega_j(Z)\omega_k(Z)\Big) \\ &= |b_1b_2|\mathscr{D}_n^2(\mathbf{X}_n, \boldsymbol{Y}_n, \boldsymbol{Z}_n|Z). \end{aligned}$$

Thus $\mathscr{D}_n(a_1+b_1C_1\mathbf{X}_n, a_2+b_2C_2\boldsymbol{Y}_n, \boldsymbol{Z}_n|Z) = \sqrt{|b_1b_2|}\mathscr{D}_n(\mathbf{X}_n, \boldsymbol{Y}_n, \boldsymbol{Z}_n|Z)$.

ii. It follows from a similar argument in Theorem 2.

A.7 proof of Theorem 6

Let $P_n(W_1, W_2, W_3, W_4, W_5) = d_{1234}^s K_{15}K_{25}K_{35}K_{45}$ and express $\mathscr{S}_n$ as a U-statistic with random kernel,

$$\mathscr{S}_n = \frac{1}{C_n^5 h^{4r}}\sum_{i<j<k<l<u}\mathscr{P}_n(W_i, W_j, W_k, W_l, W_u),$$

where

$$\begin{aligned} &\mathscr{P}_n(W_1, W_2, W_3, W_4, W_5) \\ &= \tfrac{1}{5}[P_n(W_1, W_2, W_3, W_4, W_5) + P_n(W_5, W_2, W_3, W_4, W_1) \\ &+ P_n(W_1, W_5, W_3, W_4, W_2) + P_n(W_1, W_2, W_5, W_4, W_3) \\ &+ P_n(W_1, W_2, W_3, W_5, W_4)]. \end{aligned}$$

Step 1

$\mathscr{S}_n = E\,\mathscr{P}_n(W_1, W_2, W_3, W_4, W_5)/h^{4r} + o_p(1)$.

Consider the H-decomposition in Lee (1990), we denote that

$$P_{nc}(W_1,\cdots,W_c)=E(P_n(W_1,\cdots,W_5)|W_1,\cdots,W_c),$$
$$\mathscr{P}_{nc}(W_1,\cdots,W_c)=E(\mathscr{P}_n(W_1,\cdots,W_5)|W_1,\cdots,W_c).$$

Further, let $h_n^{(1)}=\mathscr{P}_{n1}(W_1)/h^{4r}$ and

$$h_n^{(c)}(W_1,\cdots,W_c)=\mathscr{P}_{nc}(W_1,\cdots,W_c)/h^{4r}-\sum_{j=1}^{c-1}\sum_{(c,j)}h_n^{(j)}(W_{i_1},\cdots,W_{i_c}).$$

Then

$$\mathscr{S}_n=\sum_{c=1}^{5}\begin{pmatrix}5\\c\end{pmatrix}H_n^{(c)}, \quad \text{(A.7)}$$

where $H_n^{(c)}=\begin{pmatrix}n\\c\end{pmatrix}^{-1}\sum_{(n,c)}h_n^{(c)}(W_{i_1},\cdots,W_{i_c})$ satisfies the following properties:

1. $H_n^{(c)}(c=1,\cdots,5)$ are uncorrelated.
2. $Var(H_n^{(c)})=\begin{pmatrix}n\\c\end{pmatrix}Var(h_n^{(c)}(W_1,\cdots,W_c))$.
3. $Var(h_n^{(c)}(W_1,\cdots,W_c))=\sum_{j=1}^{c}(-1)^{c-j}\begin{pmatrix}c\\j\end{pmatrix}Var(\mathscr{P}_{nc}(W_1,\cdots,W_c)/h^{4r})$.

Note that $E\mathscr{P}_{n1}^2(W_1)$ can be expanded into several terms, and each of these terms can be shown to be of order h^{8r}. We only give the proof for the first term. Since

$$\begin{gathered}
E(P_{n1}^2(W_1))\\
=\int(\int d_{1234}^s f(x_2,y_2,z_2)f(x_3,y_3,z_3)f(x_4,y_4,z_4)f(x_5,y_5,z_5)\mathrm{d}x_2\mathrm{d}x_3\mathrm{d}x_4\mathrm{d}x_5\mathrm{d}y_2\mathrm{d}y_3\mathrm{d}y_4\mathrm{d}y_5\\
\cdot K_{15}K_{25}K_{35}K_{45}\mathrm{d}z_2\mathrm{d}z_3\mathrm{d}z_4\mathrm{d}z_5)^2 f(x_1,y_1,z_1)\mathrm{d}x_1\mathrm{d}y_1\mathrm{d}z_1\\
=h^{8r}\int(\int d_{1234}^s f(x_2,y_2,z_1+H(z_{51}-z_{52}))\\
\cdot f(x_3,y_3,z_1+H(z_{51}-z_{53}))f(x_4,y_4,z_1+H(z_{51}-z_{54}))\\
\cdot f(x_5,y_5,z_1+Hz_{51})\mathrm{d}x_2\mathrm{d}x_3\mathrm{d}x_4\mathrm{d}x_5\mathrm{d}y_2\mathrm{d}y_3\mathrm{d}y_4\mathrm{d}y_5\\
\cdot K(z_{51})K(z_{52})K(z_{53})K(z_{54})\mathrm{d}z_{51}\mathrm{d}z_{52}\mathrm{d}z_{53}\mathrm{d}z_{54})^2\\
\cdot f(x_1,y_1,z_1)\mathrm{d}x_1\mathrm{d}y_1\mathrm{d}z_1\\
=O_P(h^{8r}),
\end{gathered}$$

we have $E(\mathscr{P}_{n1}^2(W_1))=O_P(h^{8r})$. Therefore, $Var(H_n^{(1)})=O_P(\frac{1}{n})$.

Analogously to $Var(H_n^{(1)})$, we can obtain that $Var(H_n^{(c)})=O_P(\frac{1}{n^c h^{(c-1)r}})$.

According to (A.7), we have

$$\mathscr{S}_n = E\mathscr{P}_n(W_1, W_2, W_3, W_4, W_5)/h^{4r} + \sum_{c=1}^{5} O_P\left(\frac{1}{n^c h^{(c-1)r}}\right).$$

Thus, $\mathscr{S}_n = E\,\mathscr{P}_n(W_1, W_2, W_3, W_4, W_5)/h^{4r} + o_p(1)$ when $nh^r \to \infty$.

Step 2

$E\,\mathscr{P}_n(W_1, W_2, W_3, W_4, W_5)/h^{4r} = E[\,\mathscr{D}^2(X, Y|Z)12f^4(Z)] + O_P\,(h^2)$.

Due to the definition of $\mathscr{P}_n(W_1, W_2, W_3, W_4, W_5)$, it's easy to verify that

$$E\mathscr{P}_n(W_1, W_2, W_3, W_4, W_5) = E(d^s_{1234} K_{15} K_{25} K_{35} K_{45}),$$

thus we consider the expectation of $d^s_{1234} K_{15} K_{25} K_{35} K_{45}$ as follows,

$$\begin{gathered} E(d^s_{1234} K_{15} K_{25} K_{35} K_{45}) \\ = 12E[(\|X_1 - X_2\|_p \|Y_1 - Y_2\|_q + \|X_1 - X_2\|_p \|Y_3 - Y_4\|_q \\ -2\|X_1 - X_2\|_p \|Y_1 - Y_3\|_q) K_{15} K_{25} K_{35} K_{45}]. \end{gathered}$$

Consider the first term of the above formula,

$$\begin{gathered} E[(\|X_1 - X_2\|_p \|Y_1 - Y_2\|_q K_{15} K_{25} K_{35} K_{45}] \\ = \int \|x_1 - x_2\|_p \|y_1 - y_2\|_q K_{15} K_{25} K_{35} K_{45} \\ \cdot f(x_1, y_1, z_1) f(x_2, y_2, z_2) f(x_3, y_3, z_3) f(x_4, y_4, z_4) f(z_5) \\ \cdot \mathrm{d}x_1 \mathrm{d}x_2 \mathrm{d}x_3 \mathrm{d}x_4 \mathrm{d}y_1 \mathrm{d}y_2 \mathrm{d}y_3 \mathrm{d}y_4 \mathrm{d}z_1 \mathrm{d}z_2 \mathrm{d}z_3 \mathrm{d}z_4 \mathrm{d}z_5 \\ = \int \|x_1 - x_2\|_p \|y_1 - y_2\|_q f(x_1, y_1|z_1) f(x_2, y_2|z_2) f(z_1) f(z_2) \\ \cdot K_{15} K_{25} K_{35} K_{45} f(z_3) f(z_4) f(z_4) \mathrm{d}x_1 \mathrm{d}x_2 \mathrm{d}y_1 \mathrm{d}y_2 \mathrm{d}z_1 \mathrm{d}z_2 \mathrm{d}z_3 \mathrm{d}z_4 \mathrm{d}z_5. \end{gathered}$$

Denote $z_j = z_i + Hz_{ji}$, $1 \le i < j \le 5$, as the variable transformations between z_i's which we use in evaluating the integrals. With the Taylor expansion, we have

$$\begin{gathered} \int \|x_1 - x_2\|_p \|y_1 - y_2\|_q f(x_1, y_1|z_1) f(x_2, y_2|z_2) f(z_1) f(z_2) \\ \cdot K_{15} K_{25} K_{35} K_{45} f(z_3) f(z_4) f(z_5) \mathrm{d}x_1 \mathrm{d}x_2 \mathrm{d}y_1 \mathrm{d}y_2 \mathrm{d}z_1 \mathrm{d}z_2 \mathrm{d}z_3 \mathrm{d}z_4 \mathrm{d}z_5 \\ = \int \|x_1 - x_2\|_p \|y_1 - y_2\|_q f(x_1, y_1|z_5 + Hz_{51}) f(x_2, y_2|z_5 + Hz_{52}) \\ \cdot \mathrm{d}x_1 \mathrm{d}x_2 \mathrm{d}y_1 \mathrm{d}y_2 K(z_{51}) K(z_{52}) f(z_5 + Hz_{51}) f(z_5 + Hz_{52}) \mathrm{d}z_{51} \mathrm{d}z_{52} \\ \cdot K(z_{53}) K(z_{54}) f(z_5 + Hz_{53}) f(z_5 + Hz_{54}) f(z_5) \mathrm{d}z_{53} \mathrm{d}z_{54} \mathrm{d}z_5 \\ = h^{4r} \int \|x_1 - x_2\|_p \|y_1 - y_2\|_q f(x_1, y_1|z_5) f(x_2, y_2|z_5) \mathrm{d}x_1 \mathrm{d}x_2 \mathrm{d}y_1 \mathrm{d}y_2 \\ \cdot f(z_5) f(z_5) f(z_5) f(z_5) f(z_5) \mathrm{d}z_5 + O_P(h^{4r+2}) \\ = h^{4r} \int E(d^X_{12} d^Y_{12}|z_5) f^5(z_5) \mathrm{d}z_5 + O_P(h^{4r+2}). \end{gathered}$$

Author Manuscript

Similarly, we can verify that

$$
\begin{aligned}
& E[(\|X_1-X_2\|_p\|Y_1-Y_3\|_q K_{15}K_{25}K_{35}K_{45}] \\
=& \int\|x_1-x_2\|_p\|y_1-y_3\|_q K_{15}K_{25}K_{35}K_{45} f(x_1,y_1,z_1)f(x_2,y_2,z_2)f(x_3,y_3,z_3) \\
& \cdot f(x_4,y_4,z_4)f(z_5)\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3\mathrm{d}x_4\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}y_3\mathrm{d}y_4\mathrm{d}z_1\mathrm{d}z_2\mathrm{d}z_3\mathrm{d}z_4\mathrm{d}z_5 \\
=& h^{4r}\int\|x_1-x_2\|_p\|y_1-y_3\|_q f(x_1,y_1|z_5)f(x_2|z_5)f(y_3|z_5)\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}y_1\mathrm{d}y_3 \\
& \cdot f(z_5)f(z_5)f(z_5)f(z_5)f(z_5)\mathrm{d}z_5+O_P(h^{4r+2}) \\
=& h^{4r}\int E(d_{12}^X d_{13}^Y|z_5)f^5(z_5)\mathrm{d}z_5+O_P(h^{4r+2}),
\end{aligned}
$$

and

$$
\begin{aligned}
& E[(\|X_1-X_2\|_p\|Y_3-Y_4\|_q K_{15}K_{25}K_{35}K_{45}] \\
=& \int\|x_1-x_2\|_p\|y_3-y_4\|_q K_{15}K_{25}K_{35}K_{45} f(x_1,y_1,z_1)f(x_2,y_2,z_2)f(x_3,y_3,z_3) \\
& \cdot f(x_4,y_4,z_4)f(z_5)\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3\mathrm{d}x_4\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}y_3\mathrm{d}y_4\mathrm{d}z_1\mathrm{d}z_2\mathrm{d}z_3\mathrm{d}z_4\mathrm{d}z_5 \\
=& h^{4r}\int\|x_1-x_2\|_p\|y_3-y_4\|_q f(x_1|z_5)f(y_3|z_5)f(x_2|z_5)f(y_4|z_5)\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}y_3\mathrm{d}y_4 \\
& \cdot f(z_5)f(z_5)f(z_5)f(z_5)f(z_5)\mathrm{d}z_5+O_P(h^{4r+2}) \\
=& h^{4r}\int E(d_{12}^X d_{34}^Y|z_5)f^5(z_5)\mathrm{d}z_5+O_P(h^{4r+2}),
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
& E\mathscr{P}_n(W_1,W_2,W_3,W_4,W_5)/h^{4r} \\
=& E(d_{1234}^s K_{15}K_{25}K_{35}K_{45})/h^{4r} \\
=& 12\int[E(d_{12}^X d_{12}^Y|z_5)+E(d_{12}^X d_{34}^Y|z_5)-2E(d_{12}^X d_{13}^Y|z_5)]f^5(z_5)\mathrm{d}z_5+O_P(h^2) \\
=& 12E[\mathscr{D}^2(X,Y|Z)f^4(Z)]+O_P(h^2).
\end{aligned}
\tag{A.8}
$$

According to the results in step 1 and step 2, we can obtain that

$$\mathscr{S}_n \xrightarrow[n\to\infty]{P} 12E[\mathscr{D}^2(X,Y|Z)f^4(Z)].$$

A.8 Proof of Theorem 7

The notations in Theorem 7 are defined in Theorem 6. We use Lemma B.4 in Fan and Li (1996), which extends Theorem 1 in Hall (1984), to obtain the asymptotical distribution of $\mathscr{S}_n$ in the following three steps.

Step 1

$\mathscr{S}_n$ is a degenerated U-statistic with random kernel.

According to (A.8), we have

$$
\begin{aligned}
& E\mathscr{P}_n(W_1, W_2, W_3, W_4, W_5) \\
=& 12h^{4r}\int[E(d_{12}^X d_{12}^Y|z_5)+E(d_{12}^X d_{34}^Y|z_5)-2E(d_{12}^X d_{13}^Y|z_5)]f^5(z_5)\mathrm{d}z_5+O_P(h^{4r+2}) \\
=& 12h^{4r}E[\mathscr{D}^2(X,Y|Z)f^4(Z)]+O_P(h^{4r+2}).
\end{aligned}
$$

Due to the conditional independence assumption, $\mathcal{D}^2(X, Y|Z) = 0$, thus

$$
E\mathscr{P}_n(W_1, W_2, W_3, W_4, W_5)=O_P(h^{4r+2}).
$$

Therefore, $\mathcal{S}_n$ is a degenerated U-statistic with random kernel.

Step 2

When $n \to \infty$,

$$
\frac{E\mathscr{G}_n^2(W_1, W_2)+n^{-1}E\mathscr{P}_{n2}^4(W_1, W_2)}{(E\mathscr{P}_{n2}^2(W_1, W_2))^2} \to 0,
$$

where

$$
\begin{aligned}
G_n(W_1, W_2)&=E(P_{n2}(W_1, W_3)P_{n2}(W_2, W_3)|W_1, W_2), \\
\mathscr{G}_n(W_1, W_2)&=E(\mathscr{P}_{n2}(W_1, W_3)\mathscr{P}_{n2}(W_2, W_3)|W_1, W_2).
\end{aligned}
$$

Note that $E(\mathscr{P}_{n2}^2(W_1, W_2))$ can be expanded into several terms, and each of these terms can be shown to be of order h^{14r}. We only give the proof for the first term. Since

$$
\begin{aligned}
& E(P_{n2}^2(W_1, W_2)) \\
=& \int(\int d_{1234}^s f(x_3, y_3, z_3)f(x_4, y_4, z_4)f(x_5, y_5, z_5)\mathrm{d}x_3\mathrm{d}x_4\mathrm{d}x_5\mathrm{d}y_3\mathrm{d}y_4\mathrm{d}y_5 \\
& \cdot K_{15}K_{25}K_{35}K_{45}\mathrm{d}z_3\mathrm{d}z_4\mathrm{d}z_5)^2 f(x_1, y_1, z_1)f(x_2, y_2, z_2)\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}z_1\mathrm{d}z_2 \\
=& h^{7r}\int(\int d_{1234}^s f(x_3, y_3, z_1+H(z_{21}+z_{52}-z_{53})) \\
& \cdot f(x_4, y_4, z_1+H(z_{21}+z_{52}-z_{54})) \\
& \cdot f(x_5, y_5, z_1+Hz_{21}+Hz_{52})\mathrm{d}x_3\mathrm{d}x_4\mathrm{d}x_5\mathrm{d}y_3\mathrm{d}y_4\mathrm{d}y_5 \\
& \cdot K(z_{52}+z_{21})K(z_{53})K(z_{54})K(z_{52})\mathrm{d}z_{53}\mathrm{d}z_{54}\mathrm{d}z_{52})^2 \\
& \cdot f(x_1, y_1, z_1)f(x_2, y_2, z_1+Hz_{21})\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}z_1\mathrm{d}z_{21} \\
=& O_p(h^{7r}),
\end{aligned}
$$

we have $(E\mathscr{P}_{n2}^2(W_1, W_2))^2=O_P(h^{14r})$.

Analogously to $E\mathscr{P}_{n2}^2(W_1, W_2)$, we can obtain that $E\mathscr{P}_{n2}^4(W_1, W_2)=O_P(h^{13r})$.

We can also expand $E\mathscr{G}_n^2(W_1, W_2)$ into several terms, and each of which is of order h^{15r}. We give the proof for the first term only here. Since

$$\begin{gathered} P_{n2}(W_1,W_3) \\ =h^{3r}[\int d^s_{1234} f(x_2,y_2,z_3+H(z_{53}-z_{52}))f(x_4,y_4,z_3+H(z_{53}-z_{54})) \\ \cdot f(x_5,y_5,z_3+Hz_{53})K(z_{53}+H^{-1}(z_3-z_1))K(z_{52})K(z_{54})K(z_{53}) \\ \cdot \mathrm{d}z_{52}\mathrm{d}z_{54}\mathrm{d}z_{53}\mathrm{d}x_2\mathrm{d}x_4\mathrm{d}x_5\mathrm{d}y_2\mathrm{d}y_4\mathrm{d}y_5] \end{gathered}$$

and

$$\begin{gathered} P_{n2}(W_2,W_3) \\ =h^{3r}[\int d^s_{1234} f(x_1,y_1,z_3+H(z_{53}-z_{51}))f(x_4,y_4,z_3+H(z_{53}-z_{54})) \\ \cdot f(x_5,y_5,z_3+Hz_{53})K(z_{53}+H^{-1}(z_3-z_2))K(z_{51})K(z_{54})K(z_{53}) \\ \cdot \mathrm{d}z_{51}\mathrm{d}z_{54}\mathrm{d}z_{53}\mathrm{d}x_1\mathrm{d}x_4\mathrm{d}x_5\mathrm{d}y_1\mathrm{d}y_4\mathrm{d}y_5], \end{gathered}$$

we can verify that $EG_n^2(W_1,W_2)=O_p(h^{15r})$, with one more transformation $z_3 = z_1 + Hz_{31}$ and $z_2 = z_1 + Hz_{21}$ in the integral. Furthermore, $E\mathscr{G}_n^2(W_1,W_2)=O_p(h^{15r})$.

Therefore, under the conditions $nh^r \to \infty$ and $h^r \to 0$, we have

$$\frac{E\mathscr{G}_n^2(W_1,W_2)+n^{-1}E\mathscr{P}_{n2}^4(W_1,W_2)}{(E\mathscr{P}_{n2}^2(W_1,W_2))^2} \to 0.$$

Step 3

The asymptotical distribution of $\mathcal{S}_n$ is $N(0, \sigma^2)$, where

$$\sigma^2=200\sigma'^2=72\sum\nolimits_1+32\sum\nolimits_2+96\sum\nolimits_{12}, \quad \text{(A.9)}$$

and

$$\begin{gathered} \sigma_1(x_1,y_1,x_2,y_2,z_1,z_2) \\ =\int d^s_{1234}\prod_{i=3}^{5} f(x_i,y_i,z_1)\mathrm{d}x_3\mathrm{d}x_4\mathrm{d}x_5\mathrm{d}y_3\mathrm{d}y_4\mathrm{d}y_5 \\ \cdot K(z_2)K(z_3)K(z_4)K(z_5)\mathrm{d}z_3\mathrm{d}z_4\mathrm{d}z_5, \\ \sigma_2(x_1,y_1,x_2,y_2,z_1,z_2) \\ =\int d^s_{2345}\prod_{i=1}^{2} f(x_i,y_i,z_1)\mathrm{d}x_3\mathrm{d}x_4\mathrm{d}x_5\mathrm{d}y_3\mathrm{d}y_4\mathrm{d}y_5 \\ \cdot K(z_2+z_5)K(z_3)K(z_4)K(z_5)\mathrm{d}z_3\mathrm{d}z_4\mathrm{d}z_5, \\ \sum\nolimits_1=\int \sigma_1^2\prod_{i=1}^{2} f(x_i,y_i,z_1)\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}z_1\mathrm{d}z_2, \\ \sum\nolimits_2=\int \sigma_2^2\prod_{i=1}^{2} f(x_i,y_i,z_1)\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}z_1\mathrm{d}z_2, \\ \sum\nolimits_{12}=\int \sigma_1\sigma_2\prod_{i=1}^{2} f(x_i,y_i,z_1)\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}z_1\mathrm{d}z_2. \end{gathered}$$

We need to verify the condition

$$E(\mathscr{P}_{nc}^2)/E(\mathscr{P}_{n2}^2)=O_P(n^{(c-2)}), c=3,4,5.$$

Similarly to the procedure of calculating $E(\mathscr{P}_{n2}^2)$, we have

$$E(\mathscr{P}_{n3}^2)=O_P(h^{6r}), E(\mathscr{P}_{n4}^2)=O_P(h^{5r}), E(\mathscr{P}_{n5}^2)=O_P(h^{4r}).$$

Thus the condition is satisfied when $nh^r \to \infty$.

According to Lemma B.4 in Fan and Li (1996), it follows that

$$\begin{aligned} nh^{r/2}\mathscr{S}_n=&\frac{nh^{r/2}}{C_n^5 h^{4r}}\sum_{i<j<k<l<u}\mathscr{P}_n(W_i,W_j,W_k,W_l,W_u)\\ \xrightarrow[n\to\infty]{d}&N(0,h^r\cdot\frac{1}{h^{8r}}\cdot h^{7r}2^{-1}5^2(5-1)^2{\sigma'}^2)+O_p(nh^{r/2}\cdot h^2)\\ \xrightarrow[n\to\infty]{d}&N(0,200{\sigma'}^2),\\ \xrightarrow[n\to\infty]{d}&N(0,\sigma^2).\end{aligned}$$

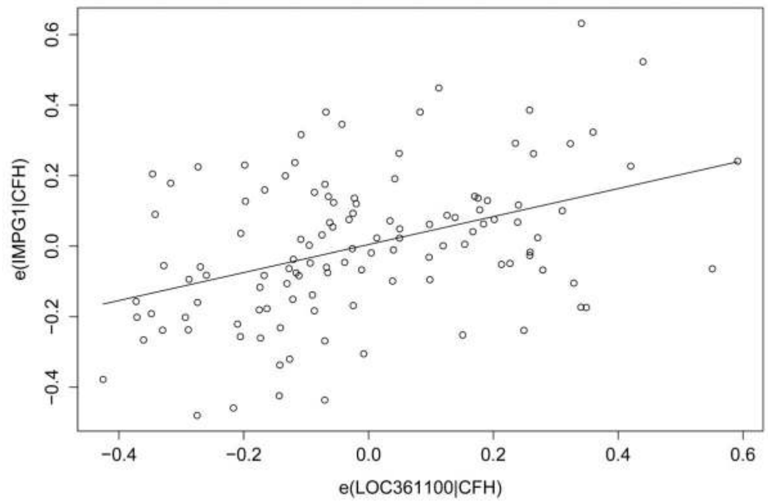

Figure 1.
Residual Plot of Regressing Y_{IMPG1} against $X_{LOC361100}$ Conditional on X_{CFH}

Table 1

Type-I Error Rates for CDIT, CI.test and KCI.test

	n					n				
	Example 1					Example 2				
Test	50	100	150	200	250	50	100	150	200	250
CDIT	0.035	0.034	0.050	0.057	0.048	0.046	0.053	0.055	0.048	0.058
CI.test	0.041	0.051	0.037	0.054	0.041	0.062	0.046	0.044	0.045	0.039
KCI	0.039	0.043	0.041	0.040	0.046	0.035	0.004	0.037	0.047	0.05
	Example 3					Example 4				
Test	50	100	150	200	250	50	100	150	200	250
CDIT	0.035	0.048	0.055	0.053	0.043	0.049	0.054	0.051	0.058	0.053
CI.test	0.222	0.363	0.482	0.603	0.677	0.043	0.064	0.066	0.050	0.053
KCI	0.058	0.047	0.057	0.061	0.054	0.037	0.035	0.058	0.039	0.049

Table 2

Empirical Power of CDIT, CI.test and KCI.test

	n					n				
	Example 5					Example 6				
Test	50	100	150	200	250	50	100	150	200	250
CDIT	0.898	0.993	1.000	1.000	1.000	0.752	0.995	1.000	1.000	1.000
CI.test	0.978	1.000	1.000	1.000	1.000	0.468	0.434	0.467	0.476	0.474
KCI	0.158	0.481	0.557	0.602	0.742	0.296	0.862	0.995	1.000	1.000
	Example 7					Example 8				
Test	50	100	150	200	250	50	100	150	200	250
CDIT	0.918	0.998	1.000	1.000	1.000	0.361	0.731	0.949	0.977	0.994
CI.test	0.953	0.984	0.983	0.995	0.987	0.456	0.476	0.464	0.461	0.485
KCI	0.574	0.947	0.998	1.000	1.000	0.089	0.401	0.685	1.000	1.000
	Example 9					Example 10				
Test	50	100	150	200	250	50	100	150	200	250
CDIT	0.802	0.955	0.975	0.983	0.994	0.355	0.789	0.912	0.935	0.965
KCI	0.189	0.419	0.517	0.579	0.777	0.023	0.350	0.557	0.718	0.846
	Example 11					Example 12				
Test	50	100	150	200	250	50	100	150	200	250
CDIT	0.768	0.973	0.994	0.995	1.000	0.812	0.956	0.976	0.995	0.995
KCI	0.300	0.590	0.744	0.786	0.936	0.115	0.227	0.334	0.396	0.476

Table 3

p-values of CDCov and DCov conditional on CFH

		p-value of DCov	p-value of CDCov
IMPG1	*LOC361100*	0.00835	1.0×10^{-6}