The Pennsylvania State University
The Graduate School
Eberly College of Science

COPULA VERSIONS OF RKHS-BASED AND DISTANCE-BASED CRITERIA

A Dissertation in
Statistics
by
Junli Lin

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2017

The dissertation of Junli Lin was reviewed and approved* by the following:

Michael Akritas
Professor of Statistics
Dissertation Advisor
Chair of Committee

Jesse Barlow
Professor of Computer Science & Engineering and Statistics

Matthew Reimherr
Professor of Statistics

Bharath Sriperumbudur
Professor of Statistics

David Hunter
Professor of Statistics
Department Head

*Signatures are on file in the Graduate School.

Abstract

Four general classes of statistics in hypothesis testing and corresponding measures are those based on reproducing kernels or distances. Among the most popular criteria for independence between two random vectors X and Y are the distance covariance (dCov) and the Hilbert-Schmidt independence criterion (HSIC). Among the most popular criteria for equal distributions of two random vectors X and Y are the Maximum Mean Discrepancy (MMD) and an energy distance (eD) criterion. Copula versions of these criteria are introduced. The estimators of the proposed criteria belong in the class of rank transform statistics and share the important property of being invariant under monotone transformations of each variable. The asymptotic theory is established under alternative hypothesis for the first two proposed statistics, and under null hypothesis for all the four proposed statistics, in which general distributions are allowed by employing mid-ranks. Dealing with the non-differentiability of the Euclidean norm, in combination with mid-ranks, presents methodological and notation challenges which are dealt with by novel arguments. Conservative tests, as well as linear time statistics for the first two proposed methods are also developed. Simulation studies suggest superior performance of the proposed statistics for certain classes of distributions.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my sincere gratitude to my advisor, Dr. Michael Akritas, for his continuous and valuable support of my Ph.D. study and research, and for his immense knowledge and patience. I could not have imagined accomplishing such achievement and completing this dissertation without his guidance.

Besides my advisor, I am also sincerely grateful to the rest of my thesis committee: Dr. Barlow, Dr. Reimherr and Dr. Sriperumbudur for their tremendous help in my Ph.D. study and in improving my dissertation.

Furthermore, I would like to thank my mother for her love and dedication throughout my life, who gives me strength and courage to pursue my dream.

Finally, I would also like to say thank you to all in the Department of Statistics, including those who helped and supported me during my Ph.D. study and in my life.

Dedication

This work is dedicated to my mother, Hua Huang.

Chapter 1 Introduction and Literature Review

1.1 Measures of Association

The measures of associations have been a major topic in the fields of statistics. In this section, several typical measures of associations are introduced.

1.1.1 Pearson Correlation Coefficient

Pearson correlation coefficient is the simplest measure of association, which measures the linear association between two random variables. For two random variables X and Y, the Pearson correlation coefficient of X and Y is defined as

$$\rho_{\text{Pearson}}(X,Y) = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}. \tag{1.1}$$

It is well-defined if both $\text{var}(X)$ and $\text{var}(Y)$ are finite. The value of a Pearson correlation coefficient $\rho_{\text{Pearson}}(X,Y)$ is always between -1 and 1, inclusively. $\rho_{\text{Pearson}}(X,Y) = -1$ or 1 implies that X and Y are totally positively or negatively linearly correlated, in which case $aX + bY = 0$ with probability 1 for some $a, b \in \mathbb{R}$. When $\rho_{\text{Pearson}}(X,Y) = 0$, X and Y are called uncorrelated, which means that X and Y do not have linear association. A Pearson correlation coefficient is easy to calculated, and it is invariant under non-degenerate linear transformations on X and on Y. In general, Pearson correlation coefficient cannot be used as a criterion of independence between two random variables, because two uncorrelated random variables are not necessarily independent. For example, X and Y are uncorrelated but dependent in each of the following four cases:

1. $X \sim \text{Uniform}(-1, 1)$ and $Y = X^2$.

2. $X \sim \text{Uniform}(0, 4\pi)$ and $Y = \cos(X)$.

3. (X, Y) is uniformly distributed on the edges of the unit square in a two-dimensional space. To be specific, let U_1, U_2 be two independent Uniform(0, 1) random variables, $X = U_1 I(U_2 < 0.5) + I(U_2 > 0.75)$ and $Y = U_1 I(U_2 > 0.5) + I(U_2 < 0.25)$.

4. (X, Y) is uniformly distributed on the unit circle in a two-dimensional space. To be specific, let U be a Uniform(0, 1) random variable, $X = \cos(2\pi U)$ and $Y = \sin(2\pi U)$.

The shape of the distributions in the above four cases are shown in Figure 1.1.

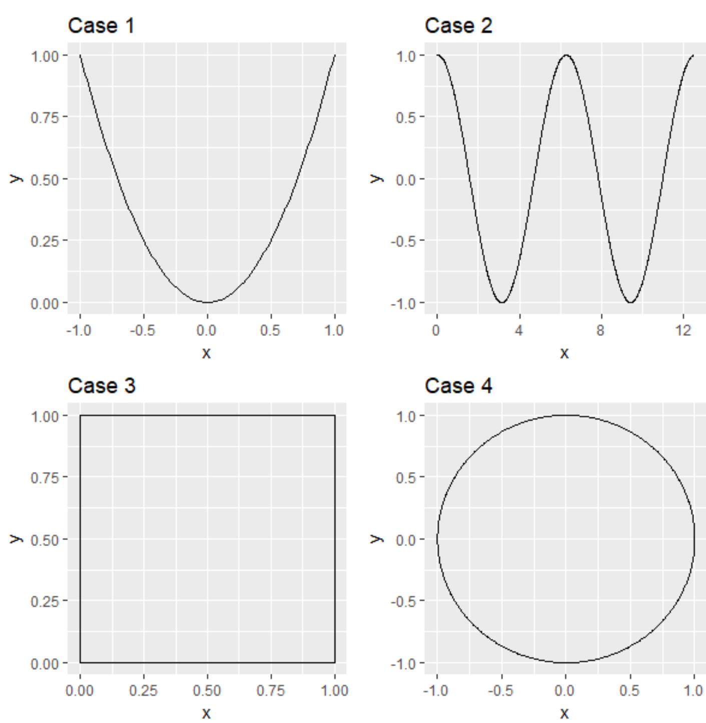

Figure 1.1. The illustration of the distributions in four different cases. In each case, the Pearson correlation coefficient of X and Y are 0, but X and Y are not independent.

Despite the fact that independent random variables may be uncorrelated, if the joint distribution of the two random variables is a multivariate normal distribution, Pearson correlation coefficient can be used as an independence criterion, because in this case the two random variable are independent if and only if they are uncorrelated. In this sense Pearson correlation coefficient is considered a parametric approach for testing independence of two random variables.

Given i.i.d. observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ from (X, Y), the Pearson correlation coefficient can be estimated by the sample Pearson correlation coefficient, which is defined as

$$\hat{\rho}_{\text{Pearson}}(X, Y) = \frac{\hat{\sigma}_{X,Y}}{\hat{\sigma}_X \hat{\sigma}_Y},$$

where

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i,$$

$$\hat{\sigma}_X^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2, \quad \hat{\sigma}_Y^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2,$$

$$\hat{\sigma}_{X,Y} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) \ .$$

It can be shown that $\hat{\rho}_{\text{Pearson}}(X, Y)$ is a consistent estimator for $\rho_{\text{Pearson}}(X, Y))$ and it is asymptotically normal.

1.1.2 Spearman's Rank Correlation Coefficient

Spearman's rank correlation coefficient (proposed in [24] and [25]) is one of the most commonly used measure of association between two random variables. It measures the proportion of the association between two random variables that can be described by monotone functions. Given i.i.d. observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ from (X, Y), denote the rank statistics of $X_1, \ldots, X_n$ by $R_{X_1}, \ldots, R_{X_n}$, and the rank statistics of $Y_1, \ldots, Y_n$ by $R_{Y_1}, \ldots, R_{Y_n}$. Then the sample Spearman's rank correlation coefficient is defined as

$$\hat{\rho}_{\text{Spearman}}(X, Y) = \frac{\hat{\sigma}_{R_X, R_Y}}{\hat{\sigma}_{R_X} \hat{\sigma}_{R_Y}},$$

where

$$\bar{R}_X = \frac{1}{n}\sum_{i=1}^{n} R_{X_i}, \quad \bar{R}_Y = \frac{1}{n}\sum_{i=1}^{n} R_{Y_i},$$

$$\hat{\sigma}^2_{R_X} = \frac{1}{n}\sum_{i=1}^{n}(R_{X_i} - \bar{R}_X)^2, \quad \hat{\sigma}^2_{R_Y} = \frac{1}{n}\sum_{i=1}^{n}(R_{Y_i} - \bar{R}_Y)^2,$$

$$\hat{\sigma}_{R_X,R_Y} = \frac{1}{n}\sum_{i=1}^{n}(R_{X_i} - \bar{R}_X)(R_{Y_i} - \bar{R}_Y).$$

The sample Spearman's rank correlation coefficient takes its extreme possible values -1 or 1 if and only if the observations of one random variable is a perfect monotone function of the observations of the other random variable with probability 1. In [16] it was shown that under mild conditions, $\sqrt{n}[\hat{\rho}_{\text{Spearman}}(X,Y) - E(\hat{\rho}_{\text{Spearman}}(X,Y))]$ is asymptotically normal.

Nevertheless, Spearman's rank correlation coefficient cannot be used as a criteria of independence in general. For each case of the distribution of (X,Y) in Figure 1.1, it can be verified that $E(\hat{\rho}_{\text{Spearman}}(X,Y)) = 0$.

1.1.3 RV coefficient

RV coefficient proposed in [6] is a measure of linear association between two random vectors. For two random vectors $\mathbf{X}$ and $\mathbf{Y}$, the RV coefficient of $\mathbf{X}$ and $\mathbf{Y}$ is defined as

$$\rho_{RV}(\mathbf{X},\mathbf{Y}) = \frac{\text{tr}(\Sigma_{\mathbf{XY}}\Sigma_{\mathbf{YX}})}{\sqrt{\text{tr}(\Sigma_{\mathbf{XX}})\text{tr}(\Sigma_{\mathbf{YY}})}}, \tag{1.2}$$

where $\text{tr}(\cdot)$ is the trace function, and

$$\begin{aligned}
\Sigma_{\mathbf{XX}} &= E[(\mathbf{X} - E(\mathbf{X}))(\mathbf{X} - E(\mathbf{X}))^T],\\
\Sigma_{\mathbf{YY}} &= E[(\mathbf{Y} - E(\mathbf{Y}))(\mathbf{Y} - E(\mathbf{Y}))^T],\\
\Sigma_{\mathbf{XY}} &= E[(\mathbf{X} - E(\mathbf{X}))(\mathbf{Y} - E(\mathbf{Y}))^T].
\end{aligned}$$

The RV coefficient of $\mathbf{X}$ and $\mathbf{Y}$ is well-defined if all components of $\Sigma_{\mathbf{XX}}$ and $\Sigma_{\mathbf{YY}}$ are finite. RV coefficient can be derived in the following sense. The linear association between two random vectors $\mathbf{X} \in \mathbb{R}^p$ and $\mathbf{Y} \in \mathbb{R}^q$ can be expressed mathematically by the following bivariate mapping:

$$C_{\text{linear}} : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}, \quad (\mathbf{a},\mathbf{b}) \mapsto \text{cov}(\mathbf{a}^T\mathbf{X}, \mathbf{b}^T\mathbf{Y}) = \mathbf{a}^T\Sigma_{\mathbf{XY}}\mathbf{b}.$$

Thus under the natural basis of $\mathbb{R}^p$ and $\mathbb{R}^p$, the bivariate mapping C_{linear} can be represented by the matrix $\Sigma_{\mathbf{XY}}$. The Hilbert-Schmidt norm of the matrix $\Sigma_{\mathbf{XY}}$,

$\|\Sigma_{\mathbf{XY}}\|_{HS}$, is zero if and only if the operator C_{linear} is a zero operator, if and only if $\mathbf{X}$ and $\mathbf{Y}$ do not have linear association. Therefore, $\|\Sigma_{\mathbf{XY}}\|_{HS}^2$ can be used as a criterion for the linear association between two random vectors $\mathbf{X}$ and $\mathbf{Y}$. Note that

$$\|\Sigma_{\mathbf{XY}}\|_{HS}^2 = \text{tr}(\Sigma_{\mathbf{XY}}\Sigma_{\mathbf{YX}}).$$

Thus the definition of RV coefficient in (1.2) is the "standardized" version of $\|C_{\text{linear}}\|_{HS}^2$. If $\mathbf{X}$ and $\mathbf{Y}$ are "standardized" such that $\text{tr}(\Sigma_{\mathbf{XX}}) = \text{tr}(\Sigma_{\mathbf{YY}}) = 1$, the RV coefficient of $\mathbf{X}$ and $\mathbf{Y}$ is essentially the sum of the eigenvalues of the matrix $\Sigma_{\mathbf{XY}}\Sigma_{\mathbf{YX}}$. In this sense, RV coefficient can be considered a multivariate generalization of Pearson correlation coefficient because the latter is the "standardized" version of the covariance. In fact, when $\mathbf{X}$ and $\mathbf{Y}$ are both random variables, the RV coefficient defined in (1.2) is reduced to the square of the Pearson correlation coefficient defined in (1.1).

Similar to Pearson correlation coefficient, RV coefficient cannot be used as a criterion of independence because $\Sigma_{\mathbf{XY}} = 0$ does not imply the independence of $\mathbf{X}$ and $\mathbf{Y}$, unless $(\mathbf{X}, \mathbf{Y})$ is normally distributed.

1.1.4 Canonical Correlation Coefficients

Canonical correlation analysis was first proposed by [17]. For any two random vectors $\mathbf{X}$ and $\mathbf{Y}$, canonical correlation analysis studies the Pearson correlation $\rho_{\text{Pearson}}(\mathbf{a}^T\mathbf{X}, \mathbf{b}^T\mathbf{Y})$ of any linear function of $\mathbf{X}$ and any linear function of $\mathbf{Y}$. Specifically, the first pair of canonical correlation vectors, $(\mathbf{a}_1, \mathbf{b}_1)$, is obtained by solving

$$\max \rho_{\text{Pearson}}(\mathbf{a}^T\mathbf{X}, \mathbf{b}^T\mathbf{Y}) = \frac{\mathbf{a}^T\Sigma_{\mathbf{XY}}\mathbf{b}}{\sqrt{\mathbf{a}^T\Sigma_{\mathbf{XX}}\mathbf{a} \cdot \mathbf{b}^T\Sigma_{\mathbf{YY}}\mathbf{b}}}. \tag{1.3}$$

The ith ($i = 2, \ldots, \min\{p, q\}$) pair of canonical correlation vectors $(\mathbf{a}_i, \mathbf{b}_i)$ is obtained by solving (1.3) under the constraints that $\mathbf{a}_i^T\mathbf{X}$ is uncorrelated with $\mathbf{a}_j^T\mathbf{X}$ for $j = 1, \ldots, i-1$, and that $\mathbf{b}_i^T\mathbf{Y}$ is uncorrelated with $\mathbf{b}_j^T\mathbf{Y}$ for $j = 1, \ldots, i-1$. The ith canonical correlation coefficients ($i = 1, \ldots, \min\{p, q\}$) is defined as the maximum value obtained in solving the ith optimization problem above. To be specific,

$$\rho_{\text{canonical},i} = \rho_{\text{Pearson}}(\mathbf{a}_i^T\mathbf{X}, \mathbf{b}_i^T\mathbf{Y}) = \frac{\mathbf{a}_i^T\Sigma_{\mathbf{XY}}\mathbf{b}_i}{\sqrt{\mathbf{a}_i^T\Sigma_{\mathbf{XX}}\mathbf{a}_i \cdot \mathbf{b}_i^T\Sigma_{\mathbf{YY}}\mathbf{b}_i}}.$$

It can be shown that the canonical correlation coefficients are the square root of the first $\min\{p, q\}$ eigenvalues of the matrix $\Sigma_{\mathbf{XX}}^{-1/2}\Sigma_{\mathbf{XY}}\Sigma_{\mathbf{YY}}^{-1}\Sigma_{\mathbf{YX}}\Sigma_{\mathbf{XX}}^{-1/2}$. This matrix is reduced to $\Sigma_{\mathbf{XY}}\Sigma_{\mathbf{YX}}$ when the covariance matrices of $\mathbf{X}$ and of $\mathbf{Y}$ are both identity matrices. It is the same matrix mentioned in Section 1.1.3. Unlike RV coefficient that combines all eigenvalues into one scalar, canonical correlation analysis studies the individual eigenvalues.

Similar to RV coefficient, canonical correlation coefficients cannot be used as a criteria of independence because $\Sigma_{\mathbf{XY}} = 0$ does not imply the independence of $\mathbf{X}$ and $\mathbf{Y}$, unless $(\mathbf{X}, \mathbf{Y})$ is normally distributed. In that case, $\rho_{\text{canonical},1}$ is used as the criterion of independence of $\mathbf{X}$ and $\mathbf{Y}$.

1.2 Criteria of Independence

The question of independence between two variables, or two groups of variables, arises frequently in such diverse areas as psychology, marketing, environmental sciences, astronomy, "-omics" studies; concrete applications can be found in several of the cited papers. All the coefficients introduced in Section 1.1 are widely used for answering univariate and multidimensional association questions.

It is well known, however, that none of the aforementioned association measures implies independence. [8] introduced the maximal correlation coefficient between two random variables $X,\ Y$ as

$$\rho_M(X, Y) = \sup_{f,g} \operatorname{Corr}(f(X), g(Y)),$$

where $f,\ g$ range over all Borel functions with $f(X)$ and $g(Y)$ square integrable. Clearly, $\rho_M(X, Y) = 0$ if and only if X and Y are independent; moreover, [21] showed that ρ_M satisfies all of his axioms for nonparametric measures of dependence. Except in some rare cases, however, the maximal correlation is difficult to calculate; see [20] and references therein. [3] introduced the *$\mathcal{F}$-correlation* as the maximal correlation when f and g range over a vector space $\mathcal{F}$ of functions (f and g can also range over different function spaces),

$$\rho_{\mathcal{F}} = \sup_{f,g\in\mathcal{F}} \operatorname{Corr}(f(X), g(Y))$$

and showed that if $\mathcal{F}$ is the reproducing kernel Hilbert space (RKHS) corresponding to a Gaussian kernel on $\mathbb{R}$, $\rho_{\mathcal{F}} = 0$ if and only if X and Y are independent; extension of this result to higher dimensions is not discussed. An estimator of $\rho_{\mathcal{F}}$ based on RKHS techniques and regularization is presented and applied for implementing kernel independent component analysis. [13] introduced the similar concept of *constrained covariance* as

$$\mathrm{COCO}(X, Y) = \sup_{f \in F,\ g \in G} \mathrm{Cov}(f(X), g(Y)),$$

where F and G are subspaces of function spaces $\mathcal{F}$ and $\mathcal{G}$, and showed that if $\mathcal{F}$ and $\mathcal{G}$ are RKHSs corresponding to universal kernels on compact domains, and F and G are the unit balls in the corresponding RKHSs, then $\mathrm{COCO}(X, Y) = 0$ if and only if the random vectors X and Y are independent. As [3] demonstrated, COCO can be recast as a generalized eigenvalue problem which can be easily estimated using RKHS methods; see also Lemma 1 in [13]. In particular, COCO can be estimated without the use of regularization, and thus it is a more convenient statistic for testing for independence.

[11] introduced the Hilbert-Schmidt independence criterion (HSIC) which is an empirical estimate of the Hilbert-Schmidt norm of the cross-covariance operator on RKHS with universal kernels. HSIC can be thought of as using the entire spectrum of the cross-covariance operator whereas COCO uses only the largest eigenvalue. Thus, HSIC characterizes the independence of two random vectors under the same conditions that COCO does. The assumption of bounded support can be eliminated with the use of characteristic kernels; see [26], [9] and references therein. The bias and rates of convergence of the HSIC statistic are studied in [11], and its asymptotic normality was derived in [12] both under the null and under the alternative hypothesis.

Using a different approach to the problem of independence testing, [30] introduced the distance covariance (dCov) as a measure of discrepancy between the joint characteristic function of $(\mathbf{X}, \mathbf{Y})$ and the product of their marginal characteristic functions for a specific weight function. The general (population version) form of this statistic is

$$\mathcal{V}^2(\mathbf{X}, \mathbf{Y}; w) = \int_{\mathbb{R}^{p+q}} |\phi_{\mathbf{X},\mathbf{Y}}(s,t) - \phi_{\mathbf{X}}(s)\phi_{\mathbf{Y}}(t)|^2 w(s,t)\,ds\,dt, \tag{1.4}$$

where $\phi_{\mathbf{X}}$ is the characteristic function of the p-dimensional vector $\mathbf{X}$, $\phi_{\mathbf{Y}}$ is the characteristic function of the q dimensional vector $\mathbf{Y}$, $\phi_{\mathbf{X},\mathbf{Y}}$ is their joint characteristic function, and w is a weight function. Not every choice of the weight function leads to a measure of dependence. Moreover, empirical characteristic functions may have large noise in the higher frequencies which could have a negative effect on the estimate of $\mathcal{V}^2(\mathbf{X}, \mathbf{Y}; w)$. [30] proposed the weight function

$$w(s,t) = [C_p C_q \|t\|_p^{1+p} \|s\|_q^{1+q}]^{-1}, \tag{1.5}$$

where $\|\cdot\|_d$ denotes the Euclidean norm for $\mathbb{R}^d$, and $C_d = \pi^{(1+d)/2}/\Gamma((1+d)/2)$. The discrepancy measure $\mathcal{V}^2(\mathbf{X}, \mathbf{Y}; w)$ with the weight function in (1.5), which will be denoted simply as $\mathcal{V}^2(\mathbf{X}, \mathbf{Y})$, is the definition of dCov. In their Theorem 3, [30] show that under the assumption of finite first moments dCov characterizes the independence of $\mathbf{X}$ and $\mathbf{Y}$. The asymptotic null distribution of the empirical version of dCov is also obtained there. [29] established the interesting and surprising result that dCov equals the Brownian covariance of $\mathbf{X}$ and $\mathbf{Y}$. A connection between HSIC and dCov was established in [22] where it was shown that a generalized version of dCov is a special case of HSIC for a particular choice of kernel.

HSIC and dCov are both criteria of dependence, while Pearson correlation coefficient and Spearman's rank correlation coefficient cannot only measure certain types of dependence. To illustrate the differences between these four methods, data are simulated from Case 4 in Figure 1.1, with different sample sizes. The p-values from the four methods are listed in Table 1.1. It shows that neither Pearson correlation coefficient nor Spearman's rank correlation fails to measure the dependence between the two random variables, while HSIC and dCov indeed measure such dependence.

Table 1.1. P-values of testing for association in four different cases of distributions with four different methods

Sample Size	Pearson	Spearman	HSIC	dCov
25	0.8419	0.9136	0.2727	0.5994
50	0.9535	0.8498	0.0076	0.1278
75	0.9496	0.9420	0.0033	0.1178
100	0.9938	0.9494	0.0001	0.0814

1.3 Some Backgrounds of Kernel Methods

Kernel methods has become popular to use linear algorithm to analyze non-linear problems. In this section, some backgrounds of kernel methods will be introduced.

1.3.1 Reproducing Kernel Hilbert Space (RKHS)

In this section, the reproducing kernel Hilbert space (RKHS) will be defined through a positive definite kernel.

Definition 1.3.1. *Let $\mathcal{X}$ be a non-empty set in $\mathbb{R}^d$. A symmetric function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called a positive definite kernel if*

$$\sum_{i,j=1}^{n} a_i a_j k(\mathbf{x}_i, \mathbf{x}_j) \geq 0$$

for any $n \in \mathbb{N}$ and $a_i \in \mathbb{R}$, $\mathbf{x}_i \in \mathcal{X}$ for $i = 1 \ldots, n$.

The examples of positive definite kernel include

1. Linear kernel: $k(\mathbf{x}_1, \mathbf{x}_2) = \langle \mathbf{x}_1, \mathbf{x}_2 \rangle$ for $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^d$, where $d \in \mathbb{N}$.
2. Gaussian kernel: $k(\mathbf{x}_1, \mathbf{x}_2) = e^{-\frac{\|\mathbf{x}_1 - \mathbf{x}_2\|^2}{2\sigma^2}}$ for $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^d$, where $\sigma > 0$, $d \in \mathbb{N}$.
3. $k(\mathbf{x}_1, \mathbf{x}_2) = e^{\langle \mathbf{x}_1, \mathbf{x}_2 \rangle}$ for $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^d$, where $d \in \mathbb{N}$.
4. V. Vovk's infinite polynomial kernel: $k(\mathbf{x}_1, \mathbf{x}_2) = (1 - \langle \mathbf{x}_1, \mathbf{x}_2 \rangle)^{-\alpha}$ for $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X} = \{\mathbf{x} \in \mathbb{R}^d : |\alpha \langle \mathbf{x}, \mathbf{x} \rangle| < 1\}$, where $\alpha > 0$ and $d \in \mathbb{N}$.

Given a positive definite kernel k, an RKHS $\mathcal{H}$ can be constructed as follows. First define a class of function on $\mathcal{X}$ as

$$\mathcal{H}_1 = \left\{ \sum_{i=1}^{n} a_i k(\cdot, \mathbf{x}_i) : n \in \mathbb{N}, a_i \in \mathbb{R}, \mathbf{x}_i \in \mathcal{X} \text{ for } i = 1, \ldots, n \right\}.$$

Define a dot product $\langle \cdot, \cdot \rangle_{\mathcal{H}_1}$ on $\mathcal{H}_1$ as

$$\left\langle \sum_{i=1}^{n} a_i k(\cdot, \mathbf{x}_i), \sum_{j=1}^{n'} a'_j k(\cdot, \mathbf{x}'_j) \right\rangle_{\mathcal{H}_1} = \sum_{i=1}^{n} \sum_{j=1}^{n'} a_i a'_j k(\mathbf{x}_i, \mathbf{x}'_j).$$

for any $n, n' \in \mathbb{N}$, $a_i \in \mathbb{R}$, $\mathbf{x}_i \in \mathcal{X}$, $i = 1, \ldots, n$, $a'_j \in \mathbb{R}$, $\mathbf{x}'_j \in \mathcal{X}$, $j = 1, \ldots, n$. It can be verify that $\langle \cdot, \cdot \rangle_{\mathcal{H}_1}$ is well-defined, and is indeed a dot product. Thus $\mathcal{H}_1$ is a inner product space. Complete the space $\mathcal{H}_1$ to obtain a Hilbert space $\mathcal{H}$ equipped with an inner product $\langle \cdot, \cdot \rangle$. Then $\mathcal{H}$ is called an RKHS induced by the kernel k. In this case, k is called a reproducing kernel of $\mathcal{H}$ because of the following reproducing property

$$\forall \mathbf{x} \in \mathcal{X}, k(\mathbf{x}, \cdot) \in \mathcal{F}, \quad \text{and } \forall f \in \mathcal{F}, \langle f(\cdot), k(\mathbf{x}, \cdot) \rangle_{\mathcal{F}} = f(\mathbf{x})$$

In [2] it is shown that a Hilbert space $\mathcal{H}$ is a RKHS if and only if there exists a unique positive definite kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ such that $\forall \mathbf{x} \in \mathcal{X}$, $k(\mathbf{x}, \cdot) \in \mathcal{F}$, and $\forall f \in \mathcal{F}$, $\langle f(\cdot), k(\mathbf{x}, \cdot) \rangle_{\mathcal{F}} = f(\mathbf{x})$. This enables us to define a RKHS through defining the reproducing kernel k. In practice, a positive definite kernel is easier to construct than an RKHS in general.

As a special case of the reproducing property, if $\phi_{\mathbf{x}} \in \mathcal{F}$ is defined as $\phi_{\mathbf{x}} = k(\cdot, \mathbf{x})$ for any $\mathbf{x} \in \mathcal{X}$,

$$\langle \phi_{\mathbf{x}_1}, \phi_{\mathbf{x}_2} \rangle = k(\mathbf{x}_1, \mathbf{x}_2).$$

$\phi_{\mathbf{x}} : \mathcal{X} \to \mathcal{F}$ is called feature map because of this property.

1.3.2 Universal Kernels

In some applications of kernel methods, an RKHS is constructed so that it is dense in $C(\mathcal{X})$, the space of all real-value continuous random variables. The reproducing kernel of such RKHS is called universal kernels, as defined in [27].

Definition 1.3.2. *A positive definite kernel k defined on a non-empty compact set $\mathcal{X} \subset \mathbb{R}^d$ is called a universal kernel if the induced RKHS $\mathcal{H}$ satisfies the following property: for any $f \in C(\mathcal{X})$ and $\epsilon > 0$, there exists $g \in \mathcal{H}$ such that $\|f - g\|_\infty < \epsilon$.*

A RKHS induced by a universal kernel enables us to approximate any continuous function on $\mathcal{X}$ by using an element in the RKHS. The Gaussian kernel is an example of universal kernel on any compact subset of its domain; see [27] for additional examples.

1.3.3 Mean Element and Cross-Covariance Operators

A continuous kernel induces a separable RKHS; see [18] or Theorem 7 in [14]. We will only consider continuous kernels because separability implies the existence of a countable orthonormal basis. Let $\mathbf{u}_i, i \geq 1$ and $\mathbf{v}_j, j \geq 1$, be orthonormal bases for $\mathcal{F}$ and $\mathcal{G}$, respectively. A linear operator $C : \mathcal{G} \to \mathcal{F}$ for which

$$\|C\|_{HS} = \sqrt{\sum_{i,j} \langle C(\mathbf{v}_j), \mathbf{u}_i \rangle_{\mathcal{F}}^2}, \tag{1.6}$$

is finite is called a Hilbert-Schmidt operator. When finite, the expression in (1.6) does not depend on the choice of orthonormal bases and is called the HS-norm. The set of Hilbert-Schmidt operators is a separable Hilbert space with the HS-norm being the corresponding norm.

Let $\mathbf{X} \in \mathcal{X} \subset \mathbb{R}^p$ be a random vector, and $\mathcal{F}$ be a class of functions on $\mathcal{X}$ which is an RKHS induced by a reproducing kernel k. Consider $E[f(\mathbf{X})]$ as a functional of $f \in \mathcal{F}$. It is a linear functional, so it may be expressed as the inner-product with an element in $\mathcal{F}$. The mean element of $\mathbf{X}$ on $\mathcal{F}$ is defined to be an element $\mu[\mathbf{X}]$ in $\mathcal{F}$ such that

$$\langle \mu[\mathbf{X}], f \rangle_{\mathcal{F}} = E[f(\mathbf{X})], \quad \forall\ f \in \mathcal{F}.$$

Under the assumption that $E[\sqrt{k(\mathbf{X}, \mathbf{X})}] < \infty$, $\mu[\mathbf{X}]$ uniquely exists, with the explicit expression (See Theorem 1 in [26])

$$\mu[\mathbf{X}] = E[k(\cdot, \mathbf{X})]$$

Now also let $\mathbf{Y} \in \mathcal{Y} \subset \mathbb{R}^q$ be a random vector, and $\mathcal{G}$ be a class of functions on $\mathcal{Y}$ which is an RKHS induced by a reproducing kernel l. The mean element of $\mathbf{Y}$ on $\mathcal{G}$ is $\mu[\mathbf{Y}] = E[k(\cdot, \mathbf{Y})]$. Consider $\mathrm{cov}[f(\mathbf{X}), g(\mathbf{Y})]$ as a bivariate function of $(f, g) \in \mathcal{F} \times \mathcal{G}$. It is a bilinear function, so it may be expressed in a quadratic form. The cross-covariance operator from $\mathcal{G}$ to $\mathcal{F}$, $C_{\mathbf{XY}} : \mathcal{G} \to \mathcal{F}$ is an linear operator such that

$$\langle f, C_{\mathbf{XY}} g \rangle_{\mathcal{F}} = \mathrm{cov}[f(\mathbf{X}), g(\mathbf{Y})], \quad \forall\ f \in \mathcal{F},\ g \in \mathcal{G}. \tag{1.7}$$

Under the assumption that $E[k(\mathbf{X}, \mathbf{X})] < \infty$ and $E[l(\mathbf{Y}, \mathbf{Y})] < \infty$, such an operator exists and is unique (see Theorem 1 in [7]); under some additional moment requirements, see (3.1), $C_{\mathbf{XY}}$ is also a HS operator. If $\mathbf{X} = \mathbf{Y}$ and $\mathcal{F} = \mathcal{G}$, $C_{\mathbf{XX}}$ is called *covariance operator*. To give an explicit form of $C_{\mathbf{XY}}$, define the tensor operator $f \otimes g : \mathcal{G} \to \mathcal{F}$ for any $f \in \mathcal{F}$ and $g \in \mathcal{G}$ as

$$(f \otimes g)(h) = f\langle g, h\rangle_{\mathcal{G}}, \quad \forall h \in \mathcal{G}.$$

Define the feature maps $\phi_{\mathbf{x}} = k(\cdot, \mathbf{x})$ for any $\mathbf{x} \in \mathcal{X}$ and $\psi_{\mathbf{x}} = l(\cdot, \mathbf{y})$ for any $\mathbf{y} \in \mathcal{Y}$. Then we have

$$\begin{aligned}
\mathrm{cov}[f(\mathbf{X}), g(\mathbf{Y})] &= E[f(\mathbf{X})g(\mathbf{Y})] - E[f(\mathbf{X})]E[g(\mathbf{Y})] \\
&= E[\langle f, \phi_{\mathbf{X}}\rangle_{\mathcal{F}}\langle g, \psi_{\mathbf{Y}}\rangle_{\mathcal{G}}] - \langle f, \mu_{\mathbf{X}}\rangle_{\mathcal{F}}\langle g, \mu_{\mathbf{Y}}\rangle_{\mathcal{G}} \\
&= E[\langle f, \phi_{\mathbf{X}}\langle \psi_{\mathbf{Y}}, g\rangle_{\mathcal{G}}\rangle_{\mathcal{F}}] - \langle f, \mu_{\mathbf{X}}\langle \mu_{\mathbf{Y}}, g\rangle_{\mathcal{G}}\rangle_{\mathcal{F}} \\
&= E[\langle f, (\phi_{\mathbf{X}} \otimes \psi_{\mathbf{Y}})g\rangle_{\mathcal{F}}] - \langle f, (\mu_{\mathbf{X}} \otimes \mu_{\mathbf{Y}})g\rangle_{\mathcal{F}} \\
&= \langle f, \{E[\phi_{\mathbf{X}} \otimes \psi_{\mathbf{Y}}] - (\mu_{\mathbf{X}} \otimes \mu_{\mathbf{Y}})\}\, g\rangle_{\mathcal{F}}.
\end{aligned}$$

Thus the explicit form of $C_{\mathbf{XY}}$ is

$$C_{\mathbf{XY}} = E[\phi_{\mathbf{X}} \otimes \psi_{\mathbf{Y}}] - \mu_{\mathbf{X}} \otimes \mu_{\mathbf{Y}}. \tag{1.8}$$

1.4 V-statistics

All the criteria of association and independence can be estimated using samples, and the corresponding statistics belong to a class of statistics called V-statistics. This section will describe basic V-statistics results that will be used later.

1.4.1 Definition

Let $\mathbf{z} \in \mathcal{Z} \subset \mathbb{R}^d$ be a random vector, and $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ be independent replicates of $\mathbf{Z}$. A statistic V_n is called a V-statistic of degree c if it has the following representation

$$V_n = \frac{1}{n^c} \sum_{i_1, \ldots, i_c = 1}^{n} h(\mathbf{Z}_{i_1}, \ldots, \mathbf{Z}_{i_c}),$$

where $c \in \mathbb{N}$ and $h : \mathcal{X}^c \to \mathbb{R}$. h is called the kernel of the V-statistics V_n. Any V-statistic V_n can also be written as a V-statistic with a symmetric kernel. To be specific,

$$V_n = \frac{1}{n^c} \sum_{i_1,\ldots,i_c=1}^{n} \tilde{h}(\mathbf{Z}_{i_1}, \ldots, \mathbf{Z}_{i_c}),$$

where

$$\tilde{h}(\mathbf{z}_1, \ldots, \mathbf{z}_c) = \frac{1}{(n)_c} \sum_{(i_1,\ldots,i_c) \in I_n^c} h(\mathbf{z}_{i_1}, \ldots, \mathbf{z}_{i_c}),$$

for any $\mathbf{z}_1, \ldots, \mathbf{z}_c \in \mathcal{Z}$, I_n^c denotes the set of all permutations of $\{1, \ldots, n\}$ and $(n)_c$ denotes the number of elements in I_n^c. $\tilde{h}$ is called symmetric because it is invariant under any permutation of the c arguments. By defining

$$T(G) = \int \cdots \int \tilde{h}(\mathbf{z}_1, \ldots, \mathbf{z}_c) \prod_{j=1}^{c} dG(\mathbf{z}_j),$$

where $G : \mathcal{Z}^d \to \mathbb{R}$, the expression in (1.4.1) can also be expressed as

$$V_n = T(F_n) = \int \cdots \int \tilde{h}(\mathbf{z}_1, \ldots, \mathbf{z}_c) \prod_{j=1}^{c} dF_n(\mathbf{z}_j) = \sum_{i_1,\ldots,i_c=1}^{n} \tilde{h}(\mathbf{Z}_{i_1}, \ldots, \mathbf{Z}_{i_c}),$$

where F_n is the empirical distribution function of $\mathbf{Z}$.

In many cases, the expression

$$T(F) = \int \cdots \int \tilde{h}(\mathbf{z}_1, \ldots, \mathbf{z}_c) \prod_{j=1}^{c} dF(\mathbf{z}_j) = E[\tilde{h}(\mathbf{Z}_{i_1}, \ldots, \mathbf{Z}_{i_c})]$$

is the parameter of interest when it is finite, where F is the distribution function of F, and the V-statistic $T(F_n)$ is an estimator for $T(F)$.

1.4.2 Asymptotic Distribution of $T(F_n)$

Using the notations in Section 1.4.1, suppose that $T(F)$ is finite. Then

$$\begin{aligned}
&T(F_n) - T(F) \\
&= \int \cdots \int \tilde{h}(\mathbf{z}_1, \ldots, \mathbf{z}_c) \left[\prod_{j=1}^{c} dF_n(\mathbf{z}_j) - \prod_{j=1}^{c} dF(\mathbf{z}_j) \right] \\
&= \int \cdots \int \tilde{h}(\mathbf{z}_1, \ldots, \mathbf{z}_c) \left\{ \prod_{j=1}^{c} \{[dF_n(\mathbf{z}_j) - dF(\mathbf{z}_j)] + dF(\mathbf{z}_j)\} - \prod_{j=1}^{c} dF(\mathbf{z}_j) \right\}
\end{aligned}$$

$$= \sum_{k=1}^{c} \binom{c}{k} \int \cdots \int \tilde{h}(\mathbf{z}_1, \ldots, \mathbf{z}_c) \left\{ \prod_{j=1}^{k} [dF_n(\mathbf{z}_j) - dF(\mathbf{z}_j)] \prod_{i=k+1}^{c} dF(\mathbf{z}_i) \right\}$$

$$= \sum_{k=1}^{c} \binom{c}{k} \int \cdots \int \tilde{h}^{(k)}(\mathbf{z}_1, \ldots, \mathbf{z}_k) \prod_{j=1}^{k} [dF_n(\mathbf{z}_j) - dF(\mathbf{z}_j)], \quad (1.9)$$

where

$$\tilde{h}^{(k)}(\mathbf{z}_1, \ldots, \mathbf{z}_k) = E[\tilde{h}(\mathbf{Z}_1, \ldots, \mathbf{Z}_c) | \mathbf{Z}_1 = \mathbf{z}_1, \ldots, \mathbf{Z}_k = \mathbf{z}_k]. \quad (1.10)$$

is the kth projection of the symmetric kernel $\tilde{h}$. Its existence is guaranteed by the fact that $T(F)$ is finite. By Lemma B in Section 6.3.2 in [23], if $E\{[\tilde{h}(\mathbf{Z}_{i_1}, \ldots, \mathbf{Z}_{i_c})]^2\} < \infty$ for all $1 \leq i_1, \ldots, i_c \leq m$,

$$E\left\{\left[\int \cdots \int \tilde{h}^{(k)}(\mathbf{z}_1, \ldots, \mathbf{z}_k) \prod_{j=1}^{k} [dF_n(\mathbf{z}_j) - dF(\mathbf{z}_j)]\right]^2\right\} = O(n^{-k}) \quad (1.11)$$

for $k = 1, \ldots, c$. Therefore, all terms in (1.9) for which $k > 1$ are $O_P(n^{-1})$. Thus

$$T(F_n) - T(F) = c \int \tilde{h}^{(1)}(\mathbf{z}_1)[dF_n(\mathbf{z}_1) - dF(\mathbf{z}_1)] + O_P(n^{-1})$$
$$= \frac{c}{n} \sum_{i=1}^{n} \left\{ \tilde{h}^{(1)}(\mathbf{Z}) - E[\tilde{h}^{(1)}(\mathbf{Z})] \right\} + O_P(n^{-1}).$$

The asymptotic normality of $T(F_n)$ follows from Central Limit Theorem. This result is summarized in the following lemma.

Lemma 1.4.1. *Let $\{\mathbf{Z}_i\}$ be a sequence of random vectors in $\mathbb{R}^d$ on the distribution $F_{\mathbf{Z}}$. Consider the V-statistic of order c,*

$$V_n = \frac{1}{n^c} \sum_{i_1, \ldots, i_c = 1}^{n} \tilde{h}(\mathbf{Z}_{i_1}, \ldots, \mathbf{Z}_{i_c})$$

where $\tilde{h}$ is symmetric. Let $\tilde{h}^{(1)}(\mathbf{z}_1) = E(\tilde{h}(\mathbf{Z}_{i_1}, \ldots, \mathbf{Z}_{i_c}) | \mathbf{Z}_1 = \mathbf{z}_1)$. Assume that $E\{[\tilde{h}(\mathbf{Z}_{i_1}, \ldots, \mathbf{Z}_{i_c})]^2\} < \infty$ for all $1 \leq i_1, \ldots, i_c \leq m$. Then

$$V_n - E(\tilde{h}(\mathbf{Z}_1, \ldots, \mathbf{Z}_c)) = \frac{c}{n} \sum_{i=1}^{n} \left\{ \tilde{h}^{(1)}(\mathbf{Z}) - E[\tilde{h}^{(1)}(\mathbf{Z})] \right\} + O_P(n^{-1}) \ ,$$

and thus

$$\sqrt{n}[V_n - E(\tilde{h}(\mathbf{Z}_1, \ldots, \mathbf{Z}_c))] \xrightarrow{D} N(0, c^2 \text{var}(\tilde{h}^{(1)}(\mathbf{Z}))) \ .$$

1.4.3 Asymptotic Distribution of $T(F_n)$ When $\tilde{h}^{(1)} = 0$

Suppose that $\tilde{h}^{(1)} = 0$, which is usually called the first order degenerate case. From (1.9) and (1.11),

$$\begin{aligned}
& T(F_n) - T(F) \\
&= \binom{c}{2} \iint \tilde{h}^{(2)}(\mathbf{z}_1, \mathbf{z}_2) \prod_{j=1}^{2} [dF_n(\mathbf{z}_j) - dF(\mathbf{z}_j)] + O_P(n^{-3/2}) \\
&= \binom{c}{2} \iint [\tilde{h}^{(2)}(\mathbf{z}_1, \mathbf{z}_2) - \tilde{h}^{(1)}(\mathbf{z}_1) - \tilde{h}^{(1)}(\mathbf{z}_2) + E(\tilde{h})] \prod_{j=1}^{2} dF_n(\mathbf{z}_j) + O_P(n^{-3/2}) \\
&= \frac{c(c-1)}{2} \frac{1}{n^2} \sum_{i_1, i_2 = 1}^{n} \left[\tilde{h}^{(2)}(\mathbf{Z}_{i_1}, \mathbf{Z}_{i_2}) - \tilde{h}^{(1)}(\mathbf{Z}_{i_1}) - \tilde{h}^{(1)}(\mathbf{Z}_{i_2}) + E(\tilde{h})\right] + O_P(n^{-3/2}) \\
&= \frac{c(c-1)}{2} \frac{1}{n^2} \sum_{i_1, i_2 = 1}^{n} \tilde{h}^{(2)}(\mathbf{Z}_{i_1}, \mathbf{Z}_{i_2}) + O_P(n^{-3/2}).
\end{aligned}$$

Therefore, it suffices to analyze the asymptotic distribution of $\frac{1}{n^2} \sum_{i_1, i_2=1}^{n} h_2(\mathbf{Z}_{i_1}, \mathbf{Z}_{i_2})$, which can be obtained by Lemma B in Section 6.4.1 in [23]. This leads us to the follow lemma.

Lemma 1.4.2. *Let $\{\mathbf{Z}_i\}$ be a sequence of random vectors in $\mathbb{R}^d$ on the distribution $F_{\mathbf{Z}}$. Consider the V-statistic of order c,*

$$V_n = \frac{1}{n^c} \sum_{i_1, \ldots, i_c = 1}^{n} \tilde{h}(\mathbf{Z}_{i_1}, \ldots, \mathbf{Z}_{i_c})$$

where $\tilde{h}$ is symmetric. Let $h^{(k)}(\mathbf{z}_1, \cdots, \mathbf{z}_k) = E(\mathbf{Z}_{i_1}, \ldots, \mathbf{Z}_{i_c} | \mathbf{Z}_1 = \mathbf{z}_1, \ldots, \mathbf{Z}_k = \mathbf{z}_k)$ for $k = 1, 2$. Assume that $h^{(1)} = 0$, $E\{[\tilde{h}(\mathbf{Z}_{i_1}, \ldots, \mathbf{Z}_{i_c})]^2\} < \infty$ for all $1 \leq i_1, \ldots, i_c \leq m$. Denote by $\{\mu_i\}$ the eigenvalues of the operator A defined on $L_2(\mathbb{R}^d, F_{\mathbf{Z}})$ by

$$(Ag)(\mathbf{z}) = \int_{-\infty}^{\infty} h_2(F; \mathbf{z}, \mathbf{z}') g(\mathbf{z}') dF_{\mathbf{Z}}(\mathbf{z}'), \quad \mathbf{z} \in \mathbb{R}^d, \ g \in L_2(\mathbb{R}^d, F_{\mathbf{Z}})$$

Then

$$V_n - E(\tilde{h}(\mathbf{Z}_1, \ldots, \mathbf{Z}_c)) = \frac{c(c-1)}{2} \frac{1}{n^2} \sum_{i_1, i_2 = 1}^{n} \tilde{h}^{(2)}(\mathbf{Z}_{i_1}, \mathbf{Z}_{i_2}) + O_P(n^{-3/2}) \ ,$$

and thus

$$n[V_n - E(\tilde{h}(\mathbf{Z}_1, \ldots, \mathbf{Z}_c))] \xrightarrow{D} \sum_{i=1}^{\infty} \mu_k \chi^2_{1k} \ ,$$

where $\{\chi^2_{1k}\}$ *are independent* χ^2_1 *random variables.*

Chapter 2
Copula Version of dCov (CdCov)

2.1 Notations

The notations introduced in this section will be used for the rest of this dissertation.

For any random variable X, and independent replicates $X_1, \dots, X_n$, let its mid-cdf F_X and empirical mid-cdf $\hat{F}_X$ be defined as

$$F_X(x) = \frac{1}{2}\left[F_X^+(x) + F_X^-(x)\right], \quad \hat{F}_X(x) = \frac{1}{2n}\sum_{i=1}^{n}\left[I(X_i \le x) + I(X_i < x)\right] \quad (2.1)$$

for any $x \in \mathbb{R}$, where $F_X^+(x) = P(X \le x)$ and $F_X^-(x) = P(X < x)$.

For any random vector $\mathbf{X} = (X_1, \dots, X_d)$, and independent replicates $\mathbf{X}_1, \dots, \mathbf{X}_n$, let $\mathbf{F_X}$ and $\hat{\mathbf{F}}_\mathbf{X}$ be defined as

$$\mathbf{F_X}(\mathbf{x}) = (F_{X_1}(x_1), \dots, F_{X_d}(x_d)), \quad \hat{\mathbf{F}}_\mathbf{X}(\mathbf{x}) = (\hat{F}_{X_1}(x_1), \dots, \hat{F}_{X_d}(x_d)) \quad (2.2)$$

for any $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$. Moreover, if we define

$$\vec{I}_{\mathbf{w}_1}(\mathbf{w}_2) = \left(\frac{I(w_{11} \le w_{21}) + I(w_{11} < w_{21})}{2}, \dots, \frac{I(w_{1d} \le w_{2d}) + I(w_{1d} < w_{2d})}{2}\right). \quad (2.3)$$

for any two vectors $\mathbf{w}_1 = (w_{11}, \dots, w_{1d}) \in \mathbb{R}^d$ and $\mathbf{w}_2 = (w_{21}, \dots, w_{2d}) \in \mathbb{R}^d$, then we have the representation $\hat{F}_\mathbf{X}(\mathbf{x}) = n^{-1}\sum_i \vec{I}_{\mathbf{x}_i}(\mathbf{x})$.

For vectors $\mathbf{w}_i = (w_{i,1}, \dots, w_{i,d})$, $i \ge 1$, and any m-dimensional multi-index

$\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m) \in \mathbb{N}^m$, define $\mathbf{w}_{\boldsymbol{\beta}}$ to be the md-dimensional vector

$$\mathbf{w}_{\boldsymbol{\beta}} = (\mathbf{w}_{\beta_1}, \ldots, \mathbf{w}_{\beta_m}). \tag{2.4}$$

For example if $m = 4$, $\mathbf{w}_{(1,2,3,4)}$ and $\mathbf{w}_{(1,1,3,4)}$ denote the $4d$-dimensional vectors $(\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \mathbf{w}_4)$ and $(\mathbf{w}_1, \mathbf{w}_1, \mathbf{w}_3, \mathbf{w}_4)$, respectively. For any two functions f_1, f_2 such that $f_1 : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$, $f_2 : \mathbb{R}^q \times \mathbb{R}^q \to \mathbb{R}$, define the function $h_{f_1,f_2} : \mathbb{R}^{4(p+q)} \to \mathbb{R}$ by

$$h_{f_1,f_2}(\mathbf{w}_{(1,2,3,4)}) = f_1(\mathbf{u}_1, \mathbf{u}_2)[f_2(\mathbf{v}_3, \mathbf{v}_4) - 2f_2(\mathbf{v}_1, \mathbf{v}_3) + f_2(\mathbf{v}_1, \mathbf{v}_2)], \tag{2.5}$$

where $\mathbf{w}_i = (\mathbf{u}_i, \mathbf{v}_i),\ i \in \{1, \ldots, 4\}$, with $\mathbf{u}_i \in \mathbb{R}^p,\ \mathbf{v}_i \in \mathbb{R}^q$. Next, for any function $\mathbf{G}_\mathbf{Z} : \mathbb{R}^{p+q} \to \mathbb{R}^{p+q}$, let $\mathbf{G}_\mathbf{X}$ and $\mathbf{G}_\mathbf{Y}$ be the first p components and last q components of $\mathbf{G}_\mathbf{Z}$, respectively, and define

$$\begin{aligned}
f_1(\mathbf{x}_1, \mathbf{x}_2; \mathbf{G}_\mathbf{X}) &= f_1(\mathbf{G}_\mathbf{X}(\mathbf{x}_1), \mathbf{G}_\mathbf{X}(\mathbf{x}_2)), \\
f_2(\mathbf{y}_1, \mathbf{y}_2; \mathbf{G}_\mathbf{Y}) &= f_2(\mathbf{G}_\mathbf{Y}(\mathbf{y}_1), \mathbf{G}_\mathbf{Y}(\mathbf{y}_2)), \\
h_{f_1,f_2}(\mathbf{z}_{(1,2,3,4)}; \mathbf{G}_\mathbf{Z}) &= h_{f_1,f_2}(\mathbf{G}_\mathbf{Z}(\mathbf{z}_1), \mathbf{G}_\mathbf{Z}(\mathbf{z}_2), \mathbf{G}_\mathbf{Z}(\mathbf{z}_3), \mathbf{G}_\mathbf{Z}(\mathbf{z}_4)),
\end{aligned} \tag{2.6}$$

for any $\mathbf{z}_i = (\mathbf{x}_i, \mathbf{y}_i), \mathbf{x}_i \in \mathbb{R}^p, \mathbf{y}_i \in \mathbb{R}^q, i \in \mathbb{N}_4$.

Consider the $(p+q)$-dimensional random vector $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$, where $\mathbf{X} = (X_1, \ldots, X_p) \in \mathcal{X}$ and $\mathbf{Y} = (Y_1, \ldots, Y_q) \in \mathcal{Y}$. Using the notation in (2.2), the copula transformation of $\mathbf{Z}$ is

$$\mathbf{F}_\mathbf{Z}(\mathbf{Z}) = (F_{X_1}(X_1), \ldots, F_{X_p}(X_p), F_{Y_1}(Y_1), \ldots, F_{Y_q}(Y_q)) = (\mathbf{F}_\mathbf{X}(\mathbf{X}), \mathbf{F}_\mathbf{Y}(\mathbf{Y})).$$

Define

$$\gamma_{f_1,f_2}(\mathbf{X}, \mathbf{Y}; \mathbf{G}_\mathbf{Z}) = E\left[h_{f_1,f_2}(\mathbf{Z}_{(1,2,3,4)}; \mathbf{G}_\mathbf{Z})\right]. \tag{2.7}$$

From now on, let $\mathbf{Z}_j = (\mathbf{X}_j, \mathbf{Y}_j),\ j = 1, \ldots, n$, be n independent copies of $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$. For simplicity, we will use the following notations in all that follows:

$$\mathbf{U}_i = \mathbf{F}_\mathbf{X}(\mathbf{X}_i), \quad \mathbf{V}_i = \mathbf{F}_\mathbf{Y}(\mathbf{Y}_i), \quad \mathbf{W}_i = (\mathbf{U}_i, \mathbf{V}_i) \tag{2.8}$$

$$\hat{\mathbf{U}}_i = \hat{\mathbf{F}}_\mathbf{X}(\mathbf{X}_i), \quad \hat{\mathbf{V}}_i = \hat{\mathbf{F}}_\mathbf{Y}(\mathbf{Y}_i), \quad \hat{\mathbf{W}}_i = (\hat{\mathbf{U}}_i, \hat{\mathbf{V}}_i) \tag{2.9}$$

$$\mathbf{U}_{ij} = \mathbf{F}_{\mathbf{X}}(\mathbf{X}_i) - \mathbf{F}_{\mathbf{X}}(\mathbf{X}_j), \quad \mathbf{V}_{ij} = \mathbf{F}_{\mathbf{Y}}(\mathbf{Y}_i) - \mathbf{F}_{\mathbf{Y}}(\mathbf{Y}_j) \tag{2.10}$$

$$\hat{\mathbf{U}}_{ij} = \hat{\mathbf{F}}_{\mathbf{X}}(\mathbf{X}_i) - \hat{\mathbf{F}}_{\mathbf{X}}(\mathbf{X}_j), \quad \hat{\mathbf{V}}_{ij} = \hat{\mathbf{F}}_{\mathbf{Y}}(\mathbf{Y}_i) - \hat{\mathbf{F}}_{\mathbf{Y}}(\mathbf{Y}_j)\ . \tag{2.11}$$

The V-statistic corresponding to the function defined in (2.7) is defined as

$$\hat{\gamma}_{f_1,f_2}(\mathbf{X},\mathbf{Y};\mathbf{G}_{\mathbf{Z}}) = \frac{1}{n^4}\sum_{i,j,q,r=1}^{n} h_{f_1,f_2}(\mathbf{Z}_{(i,j,q,r)};\mathbf{G}_{\mathbf{Z}}) \tag{2.12}$$

We will now introduce some compact notation for writing the expression of the multivariate Taylor expansion with Lagrange remainder given in [1]. Let $f:\mathbb{R}^d \to \mathbb{R}$ have continuous Mth-order derivatives. For any $\mathbf{w} = (w_1,\dots,w_d) \in \mathbb{R}^d$, any $m \le M$, and any m-dimensional multi-index $\boldsymbol{\alpha} = (\alpha_1,\dots,\alpha_m) \in \mathbb{N}_d^m$, where for any $d \in \mathbb{N}$ we denote $\mathbb{N}_d = \{1,\dots,d\}$, define

$$|\boldsymbol{\alpha}| = m, \quad \mathbf{w}^{(\boldsymbol{\alpha})} = \prod_{i=1}^{|\boldsymbol{\alpha}|} w_{\alpha_i}, \quad D^{(\boldsymbol{\alpha})} f(\mathbf{w}) = \frac{\partial^{|\boldsymbol{\alpha}|} f(\mathbf{w})}{\partial w_{\alpha_1}\cdots\partial w_{\alpha_m}}.$$

Note that in the above notation, $\mathbf{w}^{(\alpha)} = w_\alpha$, for any $\alpha \in \mathbb{N}_d$. Then,

$$f(\mathbf{w}_2)-f(\mathbf{w}_1) = \sum_{m=1}^{M-1}\frac{1}{m!}\sum_{\boldsymbol{\alpha}\in\mathbb{N}_d^m} D^{(\boldsymbol{\alpha})} f(\mathbf{w})(\mathbf{w}_2-\mathbf{w}_1)^{(\boldsymbol{\alpha})} + \frac{1}{M!}\sum_{\boldsymbol{\alpha}\in\mathbb{N}_d^M} D^{(\boldsymbol{\alpha})} f(\tilde{\mathbf{w}})(\mathbf{w}_2-\mathbf{w}_1)^{(\boldsymbol{\alpha})} \tag{2.13}$$

for some $\tilde{\mathbf{w}}$ in the line segment

$$[\mathbf{w}_1,\mathbf{w}_2] = \{(1-\lambda)\mathbf{w}_1 + \lambda\mathbf{w}_2 : \lambda \in [0,1]\}$$

determined by $\mathbf{w}_1,\ \mathbf{w}_2$. The usefulness of (2.13) for our purposes can be seen by applying it to expand $h_{f_1,f_2}(\mathbf{Z}_{(i,j,r,s)};\hat{\mathbf{F}}_{\mathbf{Z}}) - h_{f_1,f_2}(\mathbf{Z}_{(i,j,r,s)};\mathbf{F}_{\mathbf{Z}})$; see the notation in (2.6). Assuming that h_{f_1,f_2} is Mth-order continuously differentiable on $[0,1]^{4(p+q)}$, then for any $i,j,r,s \in \mathbb{N}_n$,

$$\begin{aligned} &h_{f_1,f_2}(\hat{\mathbf{W}}_{(i,j,r,s)}) - h_{f_1,f_2}(\mathbf{W}_{(i,j,r,s)}) \\ &= n^{-(M-1)} \sum_{t_1,\dots,t_{M-1}=1}^{n}\ \sum_{m=1}^{M-1}\frac{1}{m!}\cdot\sum_{\boldsymbol{\alpha}\in\mathbb{N}_{4(p+q)}^m} D^{(\boldsymbol{\alpha})} h_{f_1,f_2}(\mathbf{W}_{(i,j,r,s)}) \\ &\qquad\qquad \cdot\prod_{c=1}^{m}\left[\vec{I}_{\mathbf{Z}_{(t_c,t_c,t_c,t_c)}}(\mathbf{Z}_{(i,j,r,s)}) - \mathbf{W}_{(i,j,r,s)}\right]^{(\alpha_c)} \end{aligned}$$

$$+\frac{1}{M!}\sum_{\boldsymbol{\alpha}\in\mathbb{N}^M_{4(p+q)}} D^{(\boldsymbol{\alpha})}h_{f_1,f_2}(\tilde{\mathbf{W}}_{(i,j,r,s)})\left[\hat{\mathbf{W}}_{(i,j,r,s)}-\mathbf{W}_{(i,j,r,s)}\right]^{(\boldsymbol{\alpha})} . \tag{2.14}$$

See Lemma A.1.1 for a detailed derivation of this equality, but for now note that the first term on the right hand side of (2.14) is a V-statistics of order $M-1$, and that the summation over the observations results from the empirical distribution functions involved in the definition (2.9) of $\hat{\mathbf{W}}_i$. Using the form of the Taylor expansion in (2.13) is critical for bringing out the summation over the observations and the formation of the V-statistic.

Finally, for a vector $\mathbf{w}=(w_1,\ldots,w_d)\in\mathbb{R}^d$, $|\mathbf{w}|_\infty=\max\{|w_1|,\ldots,|w_d|\}$, for a vector valued function $\mathbf{G}$, $\|\mathbf{G}\|_\infty=\sup_{\mathbf{w}}|\mathbf{G}_{\mathbf{Z}}(\mathbf{w})|_\infty$, I_n^m will denote the set of all m-permutations $(i_1,\ldots,i_m)$ of the numbers $1,\ldots,n$, and $(n)_m$ will denote the number of such permutations..

2.2 Copula dCov Criterion and Test Statistics

Let d_p and d_q be the Euclidean distance in $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively. To be specific,

$$d_p(\mathbf{u}_1,\mathbf{u}_2)=|\mathbf{u}_1-\mathbf{u}_2|_p\,,\quad d_q(\mathbf{v}_1,\mathbf{v}_2)=|\mathbf{v}_1-\mathbf{v}_2|_q\ , \tag{2.15}$$

where $|\cdot|_d$ is the Euclidean norm in $\mathbb{R}^d$.

If f_1, f_2, and $G_{\mathbf{Z}}$ in (2.7) are replaced by d_p and d_q, and $\mathbf{F}_{\mathbf{Z}}$, respectively, $\gamma_{d_p,d_q}(\mathbf{X},\mathbf{Y};\mathbf{F}_{\mathbf{Z}})$ is the dCov discrepancy measure $\mathcal{V}^2(\mathbf{F}_{\mathbf{X}}(\mathbf{X}),\mathbf{F}_{\mathbf{Y}}(\mathbf{Y}))$ defined in (1.4) with the weight function in (1.5).

The *copula dCov independence criterion* (CdCov) states that

$$\gamma_{d_p,d_q}(\mathbf{X},\mathbf{Y};\mathbf{F}_{\mathbf{Z}})=0 \iff \mathbf{X}\perp\!\!\!\perp\mathbf{Y}. \tag{2.16}$$

The equivalence in (2.16) follows the following lemma.

Lemma 2.2.1. *For any random variable X with $F=F_X$ as defined in (2.1), the equality $F^{-1}(F(X))=X$ holds with probability one, where $F^{-1}(s)=\inf\{x:F(x)\geq s\}$, for any $s\in[0,1]$.*

Proof. Since $F^{-1}(F(X))\leq X$ always holds, we will show that $P(\{F^{-1}(F(X))<$

$X\}) = 0$. Let $A = \{x : \exists \tilde{x} < x \text{ s.t. } F(\tilde{x}) = F(x)\}$, and notice that

$$\{X \in A\} = \{\omega : \exists \tilde{x} < X(\omega) \text{ s.t. } F(\tilde{x}) = F(X(\omega))\} = \{\omega : F^{-1}(F(X(\omega)) < X(\omega)\}$$

Thus it suffices to show that $P(X \in A) = 0$. This will follow by showing first that the image of A under F, $\Lambda = F(A)$, is countable, expressing the set A as $A = \bigcup_{\lambda \in \Lambda} E_\lambda$, where $E_\lambda = \{x \in A : F(x) = \lambda\}$, and showing that $P(X \in E_\lambda) = 0$ for all $\lambda \in \Lambda$. For any $\lambda \in \Lambda$, there exists $x \in A$ such that $F(x) = \lambda$. Because $x \in A$, there exists $\tilde{x} < x$ such that $F(\tilde{x}) = \lambda$. Thus the open interval $(\tilde{x}, x) \subset E_\lambda$ because F is non-decreasing, which implies that E_λ consists of at least one rational number. Since E_λ's are disjoint, Λ is countable. Next, to show that $P(X \in E_\lambda) = 0$, for any $\lambda \in \Lambda$, note that E_λ is an interval because F is non-decreasing. Since the value of F is constant on the interval E_λ, it can be shown that $P(X \in E_\lambda) = 0$ regardless of whether E_λ is open, closed or half-open. $\square$

Replacing f_1, f_2, and $G_{\mathbf{Z}}$ in (2.12) by d_p, d_q, and $\hat{\mathbf{F}}_{\mathbf{Z}}$, respectively, leads to the proposed *CdCov test statistic*

$$\hat{\gamma}_{d_p,d_q}(\mathbf{X}, \mathbf{Y}; \hat{\mathbf{F}}_{\mathbf{Z}}) = \frac{1}{n^4} \sum_{i,j,r,s=1}^{n} h_{d_p,d_q}(\mathbf{Z}_{(i,j,r,s)}; \hat{\mathbf{F}}_{\mathbf{Z}}). \tag{2.17}$$

In order to compute the observed value of the test statistic, the mid-rank of all the observations for each component of $\mathbf{X}$ and $\mathbf{Y}$ can be used. Suppose that $W_1, \ldots, W_n$ are n observations from a random variable W, and $\hat{F}_W$ be defined as in (2.1). Let $W_1^R, \ldots, W_n^R$ be the mid-ranks of $W_1, \ldots, W_n$ by assigning each set of tied observations the average of the ranks. Then $\hat{F}_W(W_i)$ can be expressed by a linear function of W_i^R,

$$\hat{F}_W(W_i) = \frac{W_i^R}{n} - \frac{1}{2n}. \tag{2.18}$$

Denote $\mathbf{X}_i = (X_{i,1}, \ldots, X_{i,p})$ and $\mathbf{Y}_i = (Y_{i,1}, \ldots, Y_{i,q})$ for $i = 1, \ldots, n$. The steps of the algorithm to compute the CdCov test statistic is as follows.

1. For each $j = 1, \ldots, p$, let $X_{1,j}^R, \ldots, X_{n,j}^R$ be the mid-ranks of $X_{1,j}, \ldots, X_{n,j}$. Let $X_{i,j}^F = \frac{X_{i,j}^R}{n} - \frac{1}{2n}$ for $i = 1, \ldots, n$.

2. For each $j = 1, \ldots, q$, let $Y_{1,j}^R, \ldots, Y_{n,j}^R$ be the mid-ranks of $Y_{1,j}, \ldots, Y_{n,j}$. Let $Y_{i,j}^F = \frac{Y_{i,j}^R}{n} - \frac{1}{2n}$ for $i = 1, \ldots, n$.

3. For each $i = 1, \dots, n$, write $\mathbf{X}_i^F = (X_{i,1}^F, \dots, X_{i,p}^F)$.

4. For each $i = 1, \dots, n$, write $\mathbf{Y}_i^F = (Y_{i,1}^F, \dots, Y_{i,q}^F)$.

5. Construct the distance matrix $\mathbf{A} = (d_p(\mathbf{X}_i^F, \mathbf{X}_j^F))_{i,j=1,\dots,n}$ for $\mathbf{X}_1^F, \dots, \mathbf{X}_n^F$.

6. Construct the distance matrix $\mathbf{B} = (d_q(\mathbf{Y}_i^F, \mathbf{Y}_j^F))_{i,j=1,\dots,n}$ for $\mathbf{Y}_1^F, \dots, \mathbf{Y}_n^F$.

7. Define $\mathbf{H} = \mathbf{I}_n - \mathbf{1}_n\mathbf{1}_n^T$, where $\mathbf{I}_n$ is the n-dimensional identity matrix, and $\mathbf{1}_n$ is the n-dimensional vector with all components being 1. Then

$$\hat{\gamma}_{d_p,d_q}(\mathbf{X}, \mathbf{Y}; \hat{\mathbf{F}}_{\mathbf{Z}}) = \frac{1}{n^2}\text{tr}(\mathbf{HAHB}) \tag{2.19}$$

This algorithm costs $O(n^2)$.

A permutation approach can be used to construct a test for independence between $\mathbf{X}$ and $\mathbf{Y}$. To be specific, a few more steps will be added after the above algorithm as follows.

8. Set the number of permutations N (for example, $N = 1000$)

9. For $r = 1, \dots, N$,

   (a) Generate a random permutation $(s_1, \dots, s_n)$ of $1, \dots, n$.

   (b) Apply the above permutation to both columns and rows of $\mathbf{B}$ to obtain $\mathbf{B}_r$. To be specific, denote $\mathbf{B} = (b_{ij})_{i,j=1,\dots n}$. Then $\mathbf{B}_r = (b_{s_i s_j})_{i,j=1,\dots n}$.

   (c) Compute
$$\hat{\gamma}_r = \frac{1}{n^2}\text{tr}(\mathbf{HAHB}_r) \tag{2.20}$$

10. The estimated p-value is the proportion of $\hat{\gamma}_1, \dots, \hat{\gamma}_N$ that is less than $\hat{\gamma}_{d_p,d_q}(\mathbf{X}, \mathbf{Y}; \hat{\mathbf{F}}_{\mathbf{Z}})$.

$$p\text{-value} = \frac{1}{N}\sum_{r=1}^{N} I(\hat{\gamma}_r < \hat{\gamma}_{d_p,d_q}(\mathbf{X}, \mathbf{Y}; \hat{\mathbf{F}}_{\mathbf{Z}}))$$

These three steps cost $O(Nn^2)$.

The copula version of a linear-time version of the above test statistic is introduced in Section 2.5, which reduces the cost of the algorithm.

2.3 Asymptotic Theory for the CdCov Test Statistic

The asymptotic theory of $\hat{\gamma}_{d_p,d_q}(\mathbf{X},\mathbf{Y};\hat{\mathbf{F}}_{\mathbf{Z}})$, defined in (2.17), depends on whether $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$ or not. The result for the dependent case are given in Theorem 2.3.1, while the independent case is treated in Theorem 2.3.2.

With $h_{f_1,f_2}(\mathbf{z}_{(1,2,3,4)};\mathbf{G}_{\mathbf{Z}})$ as defined in (2.6), (2.5), set

$$\tilde{h}_{f_1,f_2}(\mathbf{z}_{(1,2,3,4)};\mathbf{G}_{\mathbf{Z}}) = \frac{1}{4!}\sum_{(i_1,\ldots,i_4)\in I_4^4} h_{f_1,f_2}(\mathbf{z}_{(i_1,i_2,i_3,i_4)};\mathbf{G}_{\mathbf{Z}}),$$
$$\tilde{h}^{(1)}_{f_1,f_2}(\mathbf{z};\mathbf{G}_{\mathbf{Z}}) = E\left[\tilde{h}_{f_1,f_2}(\mathbf{Z}_{(1,2,3,4)};\mathbf{G}_{\mathbf{Z}})|\mathbf{Z}_1=\mathbf{z}\right]. \tag{2.21}$$

Next, let

$$B_{ijrs} = \left\{\min_{\substack{a_1,a_2\in\{i,j,r,s\}\\ a_1<a_2}} \min\left\{|\mathbf{U}_{a_1a_2}|_p, |\mathbf{V}_{a_1a_2}|_q\right\} > 0\right\}, \tag{2.22}$$

where $\mathbf{U}_{a_1a_2},\ \mathbf{V}_{a_1a_2}$ are defined in (2.10), and define

$$\begin{aligned}
\eta_{d_p,d_q}(\mathbf{Z}_{(i,j,r,s,t)}) &= \sum_{\alpha\in\mathbb{N}_{4(p+q)}} D^{(\alpha)} h_{d_p,d_q}(\mathbf{W}_{(i,j,r,s)})\left[\vec{I}_{\mathbf{Z}_{(t,t,t,t)}}(\mathbf{Z}_{(i,j,r,s)}) - \mathbf{W}_{(i,j,r,s)}\right]^{(\alpha)} I_{B_{ijrs}},\\
\tilde{\eta}_{d_p,d_q}(\mathbf{Z}_{(1,2,3,4,5)}) &= \frac{1}{5!}\sum_{(i_1,\ldots,i_5)\in I_5^5} \tilde{\eta}_{d_p,d_q}(\mathbf{Z}_{(i_1,i_2,i_3,i_4,i_5)}),\\
\tilde{\eta}^{(1)}_{d_p,d_q}(\mathbf{z}) &= E\left[\tilde{\eta}_{d_p,d_q}(\mathbf{Z}_{(1,2,3,4,5)})|\mathbf{Z}_1=\mathbf{z}\right].
\end{aligned} \tag{2.23}$$

Note that η_{d_p,d_q} is the kernel of the V-statistics obtained in the Taylor expansion in (2.14) for $M=2$, with an indicator random variable applied to avoid non-differential points.

Theorem 2.3.1. *Assume that $\gamma_{d_p,d_q}(\mathbf{X},\mathbf{Y};\mathbf{F}_{\mathbf{Z}}) \neq 0$. Then, under no further assumptions,*

$$\hat{\gamma}_{d_p,d_q}(\mathbf{X},\mathbf{Y};\hat{\mathbf{F}}_{\mathbf{Z}}) - \hat{\gamma}_{d_p,d_q}(\mathbf{X},\mathbf{Y};\mathbf{F}_{\mathbf{Z}}) = \frac{5}{n}\sum_{i=1}^{n}\tilde{\eta}^{(1)}_{d_p,d_q}(\mathbf{Z}_i) + o_P(n^{-1/2}), \tag{2.24}$$

where $\hat{\gamma}_{d_p,d_q}(\mathbf{X},\mathbf{Y};\mathbf{F}_{\mathbf{Z}})$ is defined by replacing $f_1,\ f_2$, and $G_{\mathbf{Z}}$ in (2.12) with $d_p,\ d_q$, *and* $\mathbf{F}_{\mathbf{Z}}$*, respectively, and* $\tilde{\eta}^{(1)}_{d_p,d_q}$ *defined in (2.23). Moreover,*

$$\sqrt{n}\left\{\hat{\gamma}_{d_p,d_q}(\mathbf{X},\mathbf{Y};\hat{\mathbf{F}}_{\mathbf{Z}}) - \gamma_{d_p,d_q}(\mathbf{X},\mathbf{Y};\mathbf{F}_{\mathbf{Z}})\right\} \xrightarrow{D} N(0,\sigma^2_{d_p,d_q}) \ , \tag{2.25}$$

where $\sigma^2_{d_p,d_q} = Var\left[4\tilde{h}^{(1)}_{d_p,d_q}(\mathbf{Z};\mathbf{F}_{\mathbf{Z}}) + 5\tilde{\eta}^{(1)}_{d_p,d_q}(\mathbf{Z})\right]$*, with* $\tilde{h}^{(1)}_{d_p,d_q}$ *defined by replacing* f_1, f_2 *with* d_p, d_q *in (2.21).*

We remark that Theorem 2.3.1 continues to hold also under $\gamma_{d_p,d_q}(\mathbf{X},\mathbf{Y};\mathbf{F}_{\mathbf{Z}}) = 0$, in which case $\tilde{h}^{(1)}_{d_p,d_q}(\mathbf{Z};\mathbf{F}_{\mathbf{Z}})$ and $\tilde{\eta}^{(1)}_{d_p,d_q}(\mathbf{Z})$ are both zero; see Lemma A.1.4.

Consider now the independent case, so $\gamma_{d_p,d_q}(\mathbf{X},\mathbf{Y};\mathbf{F}_{\mathbf{Z}}) = 0$, and set

$$\begin{aligned}&\zeta_{d_p,d_q}(\mathbf{Z}_{(i,j,r,s,t_1,t_2)})\\ &= \frac{1}{2!}\sum_{\boldsymbol{\alpha}\in\mathbb{N}^2_{4(p+q)}} D^{(\boldsymbol{\alpha})}h_{d_p,d_q}(\mathbf{W}_{(i,j,r,s)})\prod_{c=1}^{2}\left[\vec{I}_{\mathbf{Z}_{(t_c,t_c,t_c,t_c)}}(\mathbf{Z}_{(i,j,r,s)}) - \mathbf{W}_{(i,j,r,s)}\right]^{(\alpha_c)} I_{B_{ijrs}},\end{aligned} \tag{2.26}$$

where B_{ijrs} is defined in (2.22). Note that $\eta_{d_p,d_q}(\mathbf{Z}_{(i,j,r,s,t_1)}) + \zeta_{d_p,d_q}(\mathbf{Z}_{(i,j,r,s,t_1,t_2)})$, where η_{d_p,d_q} is defined in (2.23), is the kernel of the V-statistics obtained in the Taylor expansion in (2.14) for $M = 3$, with an indicator random variable applied to avoid non-differential points. With $h_{f_1,f_2}(\mathbf{z}_{(1,2,3,4)};\mathbf{G}_{\mathbf{Z}})$ as defined in (2.6), (2.5), define

$$\zeta_{d_p,d_q,Total}(\mathbf{Z}_{(i,j,r,s,t_1,t_2)}) = h_{d_p,d_q}(\mathbf{Z}_{(i,j,r,s)};\mathbf{F}_{\mathbf{Z}}) + \eta_{d_p,d_q}(\mathbf{Z}_{(i,j,r,s,t_1)}) + \zeta_{d_p,d_q}(\mathbf{Z}_{(i,j,r,s,t_1,t_2)}), \tag{2.27}$$

$$\tilde{\zeta}_{d_p,d_q,Total}(\mathbf{Z}_{(1,\ldots,6)}) = \frac{1}{6!}\sum_{(i_1,\ldots,i_6)\in I^6_6} \zeta_{d_p,d_q,Total}(\mathbf{Z}_{(i_1,\ldots,i_6)}), \tag{2.28}$$

$$\tilde{\zeta}^{(1)}_{d_p,d_q,Total}(\mathbf{z}_1) = E(\tilde{\zeta}_{d_p,d_q,Total}(\mathbf{Z}_{(1,\ldots,6)})|\mathbf{Z}_1 = \mathbf{z}_1), \tag{2.29}$$

$$\tilde{\zeta}^{(2)}_{d_p,d_q,Total}(\mathbf{z}_1,\mathbf{z}_2) = E(\tilde{\zeta}_{d_p,d_q,Total}(\mathbf{Z}_{(1,\ldots,6)})|\mathbf{Z}_1 = \mathbf{z}_1,\mathbf{Z}_2 = \mathbf{z}_2). \tag{2.30}$$

Theorem 2.3.2. *Assume that* $\gamma_{d_p,d_q}(\mathbf{X},\mathbf{Y};\mathbf{F}_{\mathbf{Z}}) = 0$*. Then, under no further assumptions,*

$$\hat{\gamma}_{d_p,d_q}(\mathbf{X},\mathbf{Y};\hat{\mathbf{F}}_{\mathbf{Z}}) = \frac{15}{n^2}\sum_{i,j=1}^{n}\tilde{\zeta}^{(2)}_{d_p,d_q,Total}(\mathbf{Z}_i,\mathbf{Z}_j) + o_P(n^{-1}). \tag{2.31}$$

Moreover, if $\{\mu_i\}$ *denote the eigenvalues of the operator* A *defined on* $L_2(\mathbb{R}^{p+q}, P_{\mathbf{Z}})$ *by*

$$(Ag)(\mathbf{z}) = \int_{-\infty}^{\infty} 15\tilde{\zeta}^{(2)}_{d_p,d_q,Total}(\mathbf{z},\mathbf{z}')g(\mathbf{z}')dP_{\mathbf{Z}}(\mathbf{z}'), \quad \mathbf{z} \in \mathbb{R}^{p+q},\ g \in L_2(\mathbb{R}^{p+q}, P_{\mathbf{Z}}), \tag{2.32}$$

then

$$n\hat{\gamma}_{d_p,d_q}(\mathbf{X},\mathbf{Y};\hat{\mathbf{F}}_{\mathbf{Z}}) \xrightarrow{D} \sum_{i=1}^{\infty} \mu_i \chi^2_{1i} \tag{2.33}$$

where $\{\chi^2_{1i}\}$ *are independent* χ^2_1 *random variables.*

2.4 Conservative Test with the CdCov Test Statistic

This section presents a result in the spirit of Theorem 6 of [30]. Set $\mu_{sum} = \sum_{i=1}^{\infty} \mu_i$ and

$$\hat{\mu}_{sum} = \frac{1}{n^4} \sum_{i,j=1}^{n} d_p(\hat{\mathbf{F}}_{\mathbf{X}}(\mathbf{X}_i), \hat{\mathbf{F}}_{\mathbf{X}}(\mathbf{X}_j)) \sum_{i,j=1}^{n} d_q(\hat{\mathbf{F}}_{\mathbf{Y}}(\mathbf{Y}_i), \hat{\mathbf{F}}_{\mathbf{Y}}(\mathbf{Y}_j)) \tag{2.34}$$

Theorem 2.4.1. *If* $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$*, with no additional assumptions,*

$$\frac{n\hat{\gamma}_{d_p,d_q}(\mathbf{X},\mathbf{Y};\hat{\mathbf{F}}_{\mathbf{Z}})}{\hat{\mu}_{sum}} \xrightarrow{D} T \tag{2.35}$$

where $T = \sum_{i=1}^{\infty} \frac{\mu_i}{\mu_{sum}} \chi^2_{1i}$ *and has mean* $E(T) = 1$.

By similar arguments as in Theorem 6 of [30], a conservative test with significance level $\alpha \in (0, 0.215]$ can be constructed in which the independence of $\mathbf{X}$ and $\mathbf{Y}$ is rejected when

$$\frac{n\hat{\gamma}_{d_p,d_q}(\mathbf{X},\mathbf{Y};\hat{\mathbf{F}}_{\mathbf{Z}})}{\hat{\mu}_{sum}} \geq \left[\Phi^{-1}(1-\alpha/2)\right]^2 s$$

where Φ is the cdf of the standard normal distribution. This conservative test has asymptotic significant level at most α, and it costs much less time than the permutation approach. However, it can be quite conservative for some distributions.

2.5 Copula Version of Linear-Time Statistics

This section presents another estimator for $\gamma_{d_p,d_q}(\mathbf{X},\mathbf{Y};\mathbf{F_Z})$ for which the computation time will be much shorter.

Suppose that $n = 4m$, $m \in \mathbb{N}$. The copula versions of the linear-time statistic for estimating $\gamma_{d_p,d_q}(\mathbf{X},\mathbf{Y};\mathbf{F_Z})$ is defined as

$$\hat{\gamma}_{d_p,d_q,linear}(\mathbf{X},\mathbf{Y};\hat{\mathbf{F}}_\mathbf{Z}) = \frac{1}{m}\sum_{i=1}^{m} h_{d_p,d_q}(\mathbf{Z}_{(4i-3,4i-2,4i-1,4i)};\hat{\mathbf{F}}_\mathbf{Z}). \tag{2.36}$$

As shown in the proof of Theorem 3.2.1,

$$\begin{aligned}
&\hat{\gamma}_{d_p,d_q,linear}(\mathbf{X},\mathbf{Y};\hat{\mathbf{F}}_\mathbf{Z}) - \hat{\gamma}_{d_p,d_q,linear}(\mathbf{X},\mathbf{Y};\mathbf{F_Z})\\
&= \frac{1}{m}\sum_{i=1}^{m}\Big[h_{d_p,d_q}(\mathbf{Z}_{(4i-3,4i-2,4i-1,4i)};\hat{\mathbf{F}}_\mathbf{Z}) - h_{d_p,d_q}(\mathbf{Z}_{(4i-3,4i-2,4i-1,4i)};\mathbf{F_Z})\Big]\\
&= \frac{1}{m}\sum_{i=1}^{m}\frac{1}{n}\sum_{s=1}^{n}\eta_{d_p,d_q}(\mathbf{Z}_{(4i-3,4i-2,4i-1,4i,s)}) + O_P(n^{-1})\\
&= \frac{1}{m}\sum_{i=1}^{m}\frac{1}{4m}\sum_{j=1}^{m}\sum_{r=1}^{4}\eta_{d_p,d_q}(\mathbf{Z}_{(4i-3,4i-2,4i-1,4i,4(j-1)+r)}) + O_P(n^{-1})\\
&= \frac{1}{m^2}\sum_{i,j=1}^{m}\eta_{d_p,d_q,linear}(\mathbf{Z}_{(4i-3,4i-2,4i-1,4i,4j-3,4j-2,4j-1,4j)}) + O_P(n^{-1}),
\end{aligned}$$

where

$$\eta_{d_p,d_q,linear}(\mathbf{Z}_{(4i-3,4i-2,4i-1,4i,4j-3,4j-2,4j-1,4j)}) = \frac{1}{4}\sum_{r=1}^{4}\eta_{d_p,d_q}(\mathbf{Z}_{(4i-3,4i-2,4i-1,4i,4(j-1)+r)}).$$

Therefore,

$$\hat{\gamma}_{d_p,d_q,linear}(\mathbf{X},\mathbf{Y};\hat{\mathbf{F}}_\mathbf{Z}) = \frac{1}{m^2}\sum_{i,j=1}^{m}\zeta_{d_p,d_q,linear}(\mathbf{Z}_{(4i-3,4i-2,4i-1,4i,4j-3,4j-2,4j-1,4j)}) + O_P(n^{-1}),$$

where

$$\begin{aligned}
&\zeta_{d_p,d_q,linear}(\mathbf{Z}_{(4i-3,4i-2,4i-1,4i,4j-3,4j-2,4j-1,4j)})\\
&= h_{d_p,d_q}(\mathbf{Z}_{(4i-3,4i-2,4i-1,4i)};\mathbf{F_Z}) + \eta_{d_p,d_q,linear}(\mathbf{Z}_{(4i-3,4i-2,4i-1,4i,4j-3,4j-2,4j-1,4j)}) + O_P(n^{-1}).
\end{aligned}$$

Thus if we define

$$\begin{aligned}&\tilde{\zeta}_{d_p,d_q,linear}(\mathbf{Z}_{(4i-3,4i-2,4i-1,4i,4j-3,4j-2,4j-1,4j)})\\ =&\frac{1}{2}\Big[\zeta_{d_p,d_q,linear}(\mathbf{Z}_{(4i-3,4i-2,4i-1,4i,4j-3,4j-2,4j-1,4j)})\\ &\quad+\zeta_{d_p,d_q,linear}(\mathbf{Z}_{(4j-3,4j-2,4j-1,4j,4i-3,4i-2,4i-1,4i)})\Big],\end{aligned}$$

and

$$\tilde{\zeta}^{(1)}_{d_p,d_q,linear}(\mathbf{Z}_{(1,\ldots,4)}) = E\left[\tilde{\zeta}_{d_p,d_q,linear}(\mathbf{Z}_{(1,\ldots,8)})|\mathbf{Z}_{(1,\ldots,4)}\right],$$

then we obtain the asymptotic distribution of $\hat{\gamma}_{d_p,d_q,linear}(\mathbf{X},\mathbf{Y};\hat{\mathbf{F}}_{\mathbf{Z}})$,

$$\sqrt{n}\left[\hat{\gamma}_{d_p,d_q,linear}(\mathbf{X},\mathbf{Y};\hat{\mathbf{F}}_{\mathbf{Z}}) - \gamma_{d_p,d_q}(\mathbf{X},\mathbf{Y};\mathbf{F}_{\mathbf{Z}})\right] \xrightarrow{D} N(0,\sigma^2_{d_p,d_q,linear}) \ ,$$

where

$$\sigma^2_{d_p,d_q,linear} = \mathrm{var}\left(\tilde{\zeta}^{(1)}_{d_p,d_q,linear}(\mathbf{Z}_{(1,\ldots,4)})\right) \ .$$

This version of estimator $\hat{\gamma}_{d_p,d_q,linear}(\mathbf{X},\mathbf{Y};\hat{\mathbf{F}}_{\mathbf{Z}})$ is “linear" in the sense that the computation cost is $O(n)$ given the ranks of the observations. With a ranking method that costs $O(n\log n)$, a permutation test using $\hat{\gamma}_{d_p,d_q,linear}(\mathbf{X},\mathbf{Y};\hat{\mathbf{F}}_{\mathbf{Z}})$ costs $O(Nn + n\log n)$ where N is the number of permutations.

2.A Appendix: Proofs of Theorems

2.A.1 Proof of Theorem 2.3.1

Let $\delta_n = n^{-1/4}$, define

$$A_{ijrs} = \left\{\min_{\substack{a_1,a_2\in\{i,j,r,s\}\\ a_1<a_2}} \min\left\{|\mathbf{U}_{a_1a_2}|_p, \left|\hat{\mathbf{U}}_{a_1a_2}\right|_p, |\mathbf{V}_{a_1a_2}|_q, \left|\hat{\mathbf{V}}_{a_1a_2}\right|_q\right\} \geq \delta_n\right\} \ , \quad (2.37)$$

and write

$$\Delta\gamma_1 := \hat{\gamma}_{d_p,d_q}(\mathbf{X},\mathbf{Y};\hat{\mathbf{F}}_{\mathbf{Z}}) - \hat{\gamma}_{d_p,d_q}(\mathbf{X},\mathbf{Y};\mathbf{F}_{\mathbf{Z}}) = \Delta\gamma_2 + R_2, \quad (2.38)$$

where

$$\Delta\gamma_2 = \frac{1}{n^4}\sum_{i,j,r,s=1}^{n}\left[h_{d_p,d_q}(\mathbf{Z}_{(i,j,r,s)};\hat{\mathbf{F}}_{\mathbf{Z}}) - h_{d_p,d_q}(\mathbf{Z}_{(i,j,r,s)};\mathbf{F}_{\mathbf{Z}})\right] I_{A_{ijrs}},$$

$$R_2 = \frac{1}{n^4}\sum_{i,j,r,s=1}^{n}\left[h_{d_p,d_q}(\mathbf{Z}_{(i,j,r,s)};\hat{\mathbf{F}}_{\mathbf{Z}}) - h_{d_p,d_q}(\mathbf{Z}_{(i,j,r,s)};\mathbf{F}_{\mathbf{Z}})\right] I_{A^c_{ijrs}}. \quad (2.39)$$

Using Lemma A.1.3 it can be shown that $R_2 = o_p(n^{-1/2})$. See Section A.2.1 for details. Thus,

$$\Delta\gamma_1 = \Delta\gamma_2 + o_P(n^{-1/2}). \quad (2.40)$$

Next, using the Taylor expansion in (2.14) for $M = 2$, we write

$$\Delta\gamma_2 = \Delta\gamma_3 + R_3,$$

where

$$\Delta\gamma_3 = \frac{1}{n^5}\sum_{i,j,r,s,t=1}^{n}\sum_{\alpha\in\mathbb{N}_{4(p+q)}} D^{(\alpha)}h_{d_p,d_q}(\mathbf{W}_{(i,j,r,s)})\left[\vec{I}_{\mathbf{Z}_{(t,t,t,t)}}(\mathbf{Z}_{(i,j,r,s)}) - \mathbf{W}_{(i,j,r,s)}\right]^{(\alpha)} I_{A_{ijrs}}$$

$$R_3 = \frac{1}{n^4}\sum_{i,j,r,s=1}^{n}\frac{1}{2}\sum_{\boldsymbol{\alpha}\in\mathbb{N}^2_{4(p+q)}} D^{(\boldsymbol{\alpha})}h_{d_p,d_q}(\tilde{\mathbf{W}}_{(i,j,r,s)})\left[\hat{\mathbf{W}}_{(i,j,r,s)} - \mathbf{W}_{(i,j,r,s)}\right]^{(\boldsymbol{\alpha})} I_{A_{ijrs}} \quad (2.41)$$

for some $\tilde{\mathbf{W}}_{(i,j,r,s)} \in [\mathbf{W}_{(i,j,r,s)}, \hat{\mathbf{W}}_{(i,j,r,s)}]$. Using Lemma A.1.2 it can be shown that $R_3 = o_P(n^{-1/2})$. See Section A.2.2 for details. Thus,

$$\Delta\gamma_2 = \Delta\gamma_3 + o_P(n^{-1/2}). \quad (2.42)$$

Finally, write

$$\Delta\gamma_3 = \frac{1}{n^5}\sum_{i,j,r,s,t=1}^{n}\eta_{d_p,d_q}(\mathbf{Z}_{(i,j,r,s,t)}) - R_4, \quad (2.43)$$

where, η_{d_p,d_q} is defined in (2.23) and

$$R_4 = \frac{1}{n^5}\sum_{i,j,r,s,t=1}^{n}\sum_{\alpha\in\mathbb{N}_{4(p+q)}} D^{(\alpha)}h_{d_p,d_q}(\mathbf{W}_{(i,j,r,s)})\left[\vec{I}_{\mathbf{Z}_{(t,t,t,t)}}(\mathbf{Z}_{(i,j,r,s)}) - \mathbf{W}_{(i,j,r,s)}\right]^{(\alpha)} (I_{B_{ijrs}} - I_{A_{ijrs}}), \quad (2.44)$$

where B_{ijrs} is defined in (2.22). It can be shown that $R_4 = o_P(n^{-1/2})$; see Section

A.2.2 for details. Thus,

$$\Delta\gamma_3 = \frac{1}{n^5}\sum_{i,j,r,s,t=1}^{n}\eta_{d_p,d_q}(\mathbf{Z}_{(i,j,r,s,t)}) + o_P(n^{-1/2}). \tag{2.45}$$

Combining (2.45), (2.42), (2.40) and (2.38) we have

$$\hat{\gamma}_{d_p,d_q}(\mathbf{X},\mathbf{Y};\hat{\mathbf{F}}_{\mathbf{Z}}) - \hat{\gamma}_{d_p,d_q}(\mathbf{X},\mathbf{Y};\mathbf{F}_{\mathbf{Z}}) = \frac{1}{n^5}\sum_{i,j,r,s,t=1}^{n}\eta_{d_p,d_q}(\mathbf{Z}_{(i,j,r,s,t)}) + o_P(n^{-1/2}) \ . \tag{2.46}$$

By Lemma A.1.5,

$$\hat{\gamma}_{d_p,d_q}(\mathbf{X},\mathbf{Y};\mathbf{F}_{\mathbf{Z}}) - \gamma_{d_p,d_q}(\mathbf{X},\mathbf{Y};\mathbf{F}_{\mathbf{Z}}) = \frac{4}{n}\sum_{i=1}^{n}\tilde{h}^{(1)}_{d_p,d_q}(\mathbf{Z}_i;\mathbf{F}_{\mathbf{Z}}) + O_P(n^{-1}), \tag{2.47}$$

while by a similar V-statistics result we have

$$\frac{1}{n^5}\sum_{i,j,r,s,t=1}^{n}\eta_{d_p,d_q}(\mathbf{Z}_{(i,j,r,s,t)}) = \frac{5}{n}\sum_{i=1}^{n}\tilde{\eta}^{(1)}_{d_p,d_q}(\mathbf{Z}_i) + O_P(n^{-1}). \tag{2.48}$$

Relations (2.46) and (2.48) imply that (2.24) holds. Relations (2.46), (2.47) and (2.48) yield

$$\hat{\gamma}_{d_p,d_q}(\mathbf{X},\mathbf{Y};\hat{\mathbf{F}}_{\mathbf{Z}}) - \gamma_{d_p,d_q}(\mathbf{X},\mathbf{Y};\mathbf{F}_{\mathbf{Z}}) = \frac{1}{n}\sum_{i=1}^{n}\left[4\tilde{h}^{(1)}_{d_p,d_q}(\mathbf{Z}_i;\mathbf{F}_{\mathbf{Z}}) + 5\tilde{\eta}^{(1)}_{d_p,d_q}(\mathbf{Z}_i)\right] + o_P(n^{-1/2})$$

and thus (2.25) follows by the CLT.

2.A.2 Proof of Theorem 2.3.2

Let A_{ijrs} be as defined in (2.37) , but now set $\delta_n = n^{-1/8}$. Moreover, let $\Delta\gamma_1$ be as defined in (2.38), and write it again as $\Delta\gamma_1 = \Delta\gamma_2 + R_2$, where $\Delta\gamma_2$ and R_2 are as given in (2.39). Using the independence assumption and Lemma A.1.3 it can be shown that $R_2 = o_P(n^{-1})$. See Section A.3.1 for a detailed derivation. Thus,

$$\Delta\gamma_1 = \Delta\gamma_2 + o_P(n^{-1}). \tag{2.49}$$

Using the Taylor expansion in (2.14) for $M = 3$, we write

$$\Delta\gamma_2 = \Delta\gamma_3 + R_3,$$

where now $\Delta\gamma_3$ and R_3 are defined by

$$\begin{aligned}
\Delta\gamma_3 &= \frac{1}{n^6} \sum_{i,j,r,s,t_1,t_2=1}^{n} \sum_{m=1}^{2} \frac{1}{m!} \cdot \sum_{\boldsymbol{\alpha} \in \mathbb{N}^m_{4(p+q)}} D^{(\boldsymbol{\alpha})} h_{d_p,d_q}(\mathbf{W}_{(i,j,r,s)}) \cdot \qquad (2.50)\\
&\quad \prod_{c=1}^{m} \left[\vec{I}_{\mathbf{Z}_{(t_c,t_c,t_c,t_c)}}(\mathbf{Z}_{(i,j,r,s)}) - \mathbf{W}_{(i,j,r,s)}\right]^{(\alpha_c)} I_{A_{ijrs}}\\
R_3 &= \frac{1}{n^4} \sum_{i,j,r,s=1}^{n} \frac{1}{6} \sum_{\boldsymbol{\alpha} \in \mathbb{N}^3_{4(p+q)}} D^{(\boldsymbol{\alpha})} h_{d_p,d_q}(\tilde{\mathbf{W}}_{(i,j,r,s)}) \qquad (2.51)\\
&\quad \left[\hat{\mathbf{W}}_{(i,j,r,s)} - \mathbf{W}_{(i,j,r,s)}\right]^{(\boldsymbol{\alpha})} I_{A_{ijrs}} \qquad (2.52)
\end{aligned}$$

for some $\tilde{\mathbf{W}}_{(i,j,r,s)} \in [\mathbf{W}_{(i,j,r,s)}, \hat{\mathbf{W}}_{(i,j,r,s)}]$. It can be shown that $R_3 = o_P(n^{-1})$. See Section A.3.2 for details. Thus,

$$\Delta\gamma_2 = \Delta\gamma_3 + o_P(n^{-1}). \qquad (2.53)$$

Finally, write

$$\Delta\gamma_3 = \Delta\gamma_4 - R_4,$$

where

$$\begin{aligned}
\Delta\gamma_4 &= \frac{1}{n^6} \sum_{i,j,r,s,t_1,t_2=1}^{n} \sum_{m=1}^{2} \frac{1}{m!} \cdot \sum_{\boldsymbol{\alpha} \in \mathbb{N}^m_{4(p+q)}} D^{(\boldsymbol{\alpha})} h_{d_p,d_q}(\mathbf{W}_{(i,j,r,s)}) \cdot \prod_{c=1}^{m} \qquad (2.54)\\
&\quad \left[\vec{I}_{\mathbf{Z}_{(t_c,t_c,t_c,t_c)}}(\mathbf{Z}_{(i,j,r,s)}) - \mathbf{W}_{(i,j,r,s)}\right]^{(\alpha_c)} I_{B_{ijrs}}\\
R_4 &= \Delta\gamma_4 - \Delta\gamma_3\ , \qquad (2.55)
\end{aligned}$$

where B_{ijrs} is defined in (2.22). In Section A.3.3 it is shown that

$$R_4 = o_P(n^{-1})\ . \qquad (2.56)$$

Combining the above results, we have

$$\Delta\gamma_1 = \Delta\gamma_2 + o_P(n^{-1/2}) = \Delta\gamma_3 + o_P(n^{-1/2}) = \Delta\gamma_4 + o_P(n^{-1/2})$$

which means that

$$\hat{\gamma}_{d_p,d_q}(\mathbf{X}, \mathbf{Y}; \hat{\mathbf{F}}_\mathbf{Z}) - \hat{\gamma}_{d_p,d_q}(\mathbf{X}, \mathbf{Y}; \mathbf{F}_\mathbf{Z}) = \Delta\gamma_4 + o_P(n^{-1/2}). \tag{2.57}$$

Thus by the definition of $\Delta\gamma_4$ in (2.54), and recalling the definition of $\hat{\gamma}_{d_p,d_q}(\mathbf{X}, \mathbf{Y}; \mathbf{F}_\mathbf{Z})$ given in Theorem 2.3.1, relation (2.57) yields

$$\begin{aligned} \hat{\gamma}_{d_p,d_q}(\mathbf{X}, \mathbf{Y}; \hat{\mathbf{F}}_\mathbf{Z}) &= \frac{1}{n^6} \sum_{i,j,r,s,t_1,t_2=1}^{n} \zeta_{d_p,d_q,Total}(\mathbf{Z}_{(i,j,r,s,t_1,t_2)}) + o_P(n^{-1}) \\ &= \frac{1}{n^6} \sum_{i,j,r,s,t_1,t_2=1}^{n} \tilde{\zeta}_{d_p,d_q,Total}(\mathbf{Z}_{(i,j,r,s,t_1,t_2)}) + o_P(n^{-1}) \ , \end{aligned} \tag{2.58}$$

where $\zeta_{d_p,d_q,Total}$ is defined in (2.27), and $\tilde{\zeta}_{d_p,d_q,Total}$ is defined in (2.28). Let $\tilde{\zeta}^{(1)}_{d_p,d_q,Total}$ and $\tilde{\zeta}^{(2)}_{d_p,d_q,Total}$ be defined in (2.29) and (2.30), respectively. In Lemma A.1.4 it is shown that $\tilde{\zeta}^{(1)}_{d_p,d_q,Total} = 0$ so that, by Lemma 1.4.2,

$$\frac{1}{n^6} \sum_{i,j,r,s,t_1,t_2=1}^{n} \tilde{\zeta}_{d_p,d_q,Total}(\mathbf{Z}_{(i,j,r,s,t_1,t_2)}) = \frac{15}{n^2} \sum_{i_1,i_2=1}^{n} \tilde{\zeta}^{(2)}_{d_p,d_q,Total}(\mathbf{Z}_{i_1}, \mathbf{Z}_{i_2}) + O_P(n^{-3/2}) \ . \tag{2.59}$$

Relations (2.58) and (2.59) yield (2.31). By a generalization of Theorem B in Section 6.4.1 in [23] to vector valued observations, we have

$$n \cdot \frac{15}{n^2} \sum_{i_1,i_2=1}^{n} \tilde{\zeta}^{(2)}_{d_p,d_q,Total}(\mathbf{Z}_{i_1}, \mathbf{Z}_{i_2}) \xrightarrow{D} \sum_{i=1}^{\infty} \mu_i \chi^2_{1i} \ . \tag{2.60}$$

Then (2.33) follows from (2.31) and (2.60).

2.A.3 Proof of Theorems 2.4.1

Proof of Theorem 2.4.1. By Theorem 2.3.2 and Slutsky's Theorem, it suffices to show that, under $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$, $\hat{\mu}_{sum} \xrightarrow{P} \mu_{sum}$. From Page 1087 in [5], it follows that the eigenvalues $\{\mu_i\}$ obtained from (2.32) satisfy

$$\sum_{i=1}^{\infty} \mu_i = E\left[15\tilde{\zeta}^{(2)}_{d_p,d_q,Total}(\mathbf{Z},\mathbf{Z})\right] , \tag{2.61}$$

where $\tilde{\zeta}^{(2)}_{d_p,d_q,Total}$ is defined in (2.30). Thus, by the consistency of V-statistics and the definition of $\hat{\mu}_{sum}$ in (2.34), $\hat{\mu}_{sum} \xrightarrow{P} \mu_{sum}$ will follow by showing that, under $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$,

$$E\left[15\tilde{\zeta}^{(2)}_{d_p,d_q,Total}(\mathbf{Z},\mathbf{Z})\right] = E[d_p(\mathbf{U}_1,\mathbf{U}_2)]E[d_q(\mathbf{V}_1,\mathbf{V}_2)]. \tag{2.62}$$

To show (2.62), start by re-expressing $\tilde{\zeta}^{(2)}_{d_p,d_q,Total}(\mathbf{z},\mathbf{z})$ as

$$\begin{aligned}\tilde{\zeta}^{(2)}_{d_p,d_q,Total}(\mathbf{z},\mathbf{z}) &= E\left[\frac{1}{6!}\sum_{(i_1,\ldots,i_6)\in I_6^6} \zeta_{d_p,d_q,Total}(\mathbf{Z}_{(i_1,\ldots,i_6)})\middle|\mathbf{Z}_1=\mathbf{z},\mathbf{Z}_2=\mathbf{z}\right] \\ &= \frac{1}{(6)_2}\sum_{(i_1,i_2)\in I_6^2} E\left[\zeta_{d_p,d_q,Total}(\mathbf{Z}_{(1,\ldots,6)})\middle|\mathbf{Z}_{i_1}=\mathbf{z},\mathbf{Z}_{i_2}=\mathbf{z}\right] \end{aligned} \tag{2.63}$$

Next, from the definition of $\zeta_{d_p,d_q,Total}$ in (2.27), and considering the definition of ζ_{d_p,d_q} in (2.26), we can write

$$\zeta_{d_p,d_q,Total} = \sum_{j=1}^{6} \zeta_{d_p,d_q,j} , \tag{2.64}$$

where $\zeta_{d_p,d_q,1}$ and $\zeta_{d_p,d_q,2}$ correspond to the first order partial derivatives in (2.26), $\zeta_{d_p,d_q,3}$, $\zeta_{d_p,d_q,4}$ and $\zeta_{d_p,d_q,5}$ correspond to second order partial derivatives in (2.26), and $\zeta_{d_p,d_q,6}$ corresponds to h_{d_p,d_q}; see Section A.4.1 for the exact expressions of $\zeta_{d_p,d_q,j}$, $j = 1,\ldots,5$.

From (2.63) and (2.64) we have

$$\tilde{\zeta}^{(2)}_{d_p,d_q,Total}(\mathbf{z},\mathbf{z}) = \sum_{j=1}^{6}\frac{1}{(6)_2}\sum_{(i_1,i_2)\in I_6^2} E\left[\zeta_{d_p,d_q,j}(\mathbf{Z}_{(1,\ldots,6)})\middle|\mathbf{Z}_{i_1}=\mathbf{z},\mathbf{Z}_{i_2}=\mathbf{z}\right] . \tag{2.65}$$

In Section A.4.1 it is shown that the first five terms in (2.65) are zero, i.e., that

$$\sum_{(i_1,i_2)\in I_6^2} E\left[\zeta_{d_p,d_q,j}(\mathbf{Z}_{(1,\ldots,6)})\middle|\mathbf{Z}_{i_1}=\mathbf{z},\mathbf{Z}_{i_2}=\mathbf{z}\right]=0, \quad j=1,\ldots,5. \tag{2.66}$$

Moreover, in Section A.4.2 it is shown that the last term in (2.65) is

$$\begin{aligned} & \frac{1}{(6)_2}\sum_{(i_1,i_2)\in I_6^2} E\left[\zeta_{d_p,d_q,6}(\mathbf{Z}_1,\ldots,\mathbf{Z}_6)\middle|\mathbf{Z}_{i_1}=\mathbf{z},\mathbf{Z}_{i_2}=\mathbf{z}\right] \\ = & \frac{1}{15}\{d_p(\mathbf{u},\mathbf{u})-2E[d_p(\mathbf{u},\mathbf{U}_1)]+E[d_p(\mathbf{U}_1,\mathbf{U}_2)]\} \\ & \cdot\{d_q(\mathbf{v},\mathbf{v})-2E[d_q(\mathbf{v},\mathbf{V}_1)]+E[d_q(\mathbf{V}_1,\mathbf{V}_2)]\} \end{aligned} \tag{2.67}$$

Hence from (2.65) and (2.66), we have

$$\begin{aligned} 15\tilde{\zeta}^{(2)}_{d_p,d_q,Total}(\mathbf{z},\mathbf{z}) = & \{d_p(\mathbf{u},\mathbf{u})-2E[d_p(\mathbf{u},\mathbf{U}_1)]+E[d_p(\mathbf{U}_1,\mathbf{U}_2)]\} \\ & \cdot\{d_q(\mathbf{v},\mathbf{v})-2E[d_q(\mathbf{v},\mathbf{V}_1)]+E[d_q(\mathbf{V}_1,\mathbf{V}_2)]\} \end{aligned}$$

from which (2.62) is easily seen to follow under $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$. Finally $E(T)=1$ follows trivially. □

Chapter 3
Copula Version of HSIC (CHSIC)

3.1 Copula Version of HSIC (CHSIC)

Let $C_{\mathbf{XY}}$ be the cross-covariance operator defined in (1.7). Clearly, $C_{\mathbf{XY}}$ depends on the spaces $\mathcal{F}$ and $\mathcal{G}$, and it contains information of covariances between $f(\mathbf{X})$ and $g(\mathbf{Y})$ for any $f \in \mathcal{F}$ and $g \in \mathcal{G}$. In particular, if $C_{\mathbf{XY}} = 0$, all these covariances are zero. Equivalently, $\|C_{\mathbf{XY}}\|_{HS} = 0$ implies that these covariances are zero. However, $\|C_{\mathbf{XY}}\|_{HS} = 0$ does not necessarily imply independence. Assume now that $\mathcal{X}$ and $\mathcal{Y}$ are compact. In [11] it is shown that if $\mathcal{F}$ is dense in the space $\mathcal{C}_b(\mathcal{X})$ of continuous functions on $\mathcal{X}$ in the sup-norm metric, and similarly $\mathcal{G}$ is dense in $\mathcal{C}_b(\mathcal{Y})$, then $\|C_{\mathbf{XY}}\| = 0$ characterizes the independence of $\mathbf{X}$ and $\mathbf{Y}$. This is followed by the assumption that the kernel k is universal. From now on we will assume that the kernels k and l are universal kernels. [11] define the Hilbert-Schmidt Independence Criterion (HSIC) as the squared HS-norm of the associated cross-covariance operator $C_{\mathbf{XY}}$:

$$\text{HSIC} = \|C_{\mathbf{XY}}\|_{HS}^2,$$

and, using the explicit form of $C_{\mathbf{XY}}$ given in (1.8), they show that

$$\begin{aligned} \|C_{\mathbf{XY}}\|_{HS}^2 &= E_{\mathbf{X},\mathbf{X}',\mathbf{Y},\mathbf{Y}'}[k(\mathbf{X},\mathbf{X}')l(\mathbf{Y},\mathbf{Y}')] + E_{\mathbf{X},\mathbf{X}'}[k(\mathbf{X},\mathbf{X}')]E_{\mathbf{Y},\mathbf{Y}'}[l(\mathbf{Y},\mathbf{Y}')] \\ &\quad -2E_{\mathbf{X},\mathbf{Y}}[E_{\mathbf{X}'}[k(\mathbf{X},\mathbf{X}')]E_{\mathbf{Y}'}[l(\mathbf{Y},\mathbf{Y}')]], \end{aligned} \tag{3.1}$$

where $(\mathbf{X}', \mathbf{Y}')$ is an independent copy of $(\mathbf{X}, \mathbf{Y})$. In particular, the finiteness of the above integrals implies that the HS-norm of $C_{\mathbf{XY}}$ exists.

If $f_1,\ f_2$, and $G_{\mathbf{Z}}$ in (2.7) are replaced by k and l, and $\mathbf{F_Z}$, respectively, $\gamma_{k,l}(\mathbf{X},\mathbf{Y};\mathbf{F_Z})$ is the HSIC (3.1) for the vectors $\mathbf{F_X}(\mathbf{X}),\ \mathbf{F_Y}(\mathbf{Y})$.

The *copula HS independence criterion* (CHSIC) states

$$\gamma_{k,l}(\mathbf{X},\mathbf{Y};\mathbf{F_Z}) = 0 \iff \mathbf{X} \perp\!\!\!\perp \mathbf{Y}. \tag{3.2}$$

The equivalence in (3.2) follows Lemma 2.2.1.

Replacing $f_1,\ f_2$, and $G_{\mathbf{Z}}$ in (2.12) by universal kernels $k,\ l$, and $\hat{\mathbf{F}}_{\mathbf{Z}}$, respectively, leads to the proposed *CHSIC test statistic*

$$\hat{\gamma}_{k,l}(\mathbf{X},\mathbf{Y};\hat{\mathbf{F}}_{\mathbf{Z}}) = \frac{1}{n^4}\sum_{i,j,r,s=1}^{n} h_{k,j}(\mathbf{Z}_{(i,j,r,s)};\hat{\mathbf{F}}_{\mathbf{Z}}). \tag{3.3}$$

In order to compute the observed value of the test statistic, the mid-rank of all the observations for each component of $\mathbf{X}$ and $\mathbf{Y}$ can be used as in 2.18. Denote $\mathbf{X}_i = (X_{i,1},\ldots,X_{i,p})$ and $\mathbf{Y}_i = (Y_{i,1},\ldots,Y_{i,q})$ for $i = 1,\ldots,n$. The steps of the algorithm to compute the CHSIC test statistic is as follows.

1. For each $j = 1,\ldots,p$, let $X^R_{1,j},\ldots,X^R_{n,j}$ be the mid-ranks of $X_{1,j},\ldots,X_{n,j}$. Let $X^F_{i,j} = \frac{X^R_{i,j}}{n} - \frac{1}{2n}$ for $i = 1,\ldots,n$.

2. For each $j = 1,\ldots,q$, let $Y^R_{1,j},\ldots,Y^R_{n,j}$ be the mid-ranks of $Y_{1,j},\ldots,Y_{n,j}$. Let $Y^F_{i,j} = \frac{Y^R_{i,j}}{n} - \frac{1}{2n}$ for $i = 1,\ldots,n$.

3. For each $i = 1,\ldots,n$, write $\mathbf{X}^F_i = (X^F_{i,1},\ldots,X^F_{i,p})$.

4. For each $i = 1,\ldots,n$, write $\mathbf{Y}^F_i = (Y^F_{i,1},\ldots,Y^F_{i,q})$.

5. Construct the Gram matrix $\mathbf{K} = (k(\mathbf{X}^F_i,\mathbf{X}^F_j))_{i,j=1,\ldots,n}$ for $\mathbf{X}^F_1,\ldots,\mathbf{X}^F_n$.

6. Construct the Gram matrix $\mathbf{L} = (l(\mathbf{Y}^F_i,\mathbf{Y}^F_j))_{i,j=1,\ldots,n}$ for $\mathbf{Y}^F_1,\ldots,\mathbf{Y}^F_n$.

7. Define $\mathbf{H} = \mathbf{I}_n - \mathbf{1}_n\mathbf{1}_n^T$, where $\mathbf{I}_n$ is the n-dimensional identity matrix, and $\mathbf{1}_n$ is the n-dimensional vector with all components being 1. Then

$$\hat{\gamma}_{d_p,d_q}(\mathbf{X},\mathbf{Y};\hat{\mathbf{F}}_{\mathbf{Z}}) = \frac{1}{n^2}\mathrm{tr}(\mathbf{HKHL}) \tag{3.4}$$

This algorithm is very similar to the algorithm to compute the CdCov test statistic. It costs $O(n^2)$.

A permutation approach can be used to construct a test for independence between $\mathbf{X}$ and $\mathbf{Y}$. To be specific, a few more steps will be added after the above algorithm as follows.

8. Set the number of permutations N (for example, $N = 1000$)

9. For $r = 1, \ldots, N$,

   (a) Generate a random permutation $(s_1, \ldots, s_n)$ of $1, \ldots, n$.

   (b) Apply the above permutation to both columns and rows of $\mathbf{L}$ to obtain $\mathbf{L}_r$. To be specific, denote $\mathbf{L} = (l_{ij})_{i,j=1,\ldots n}$. Then $\mathbf{L}_r = (l_{s_i s_j})_{i,j=1,\ldots n}$.

   (c) Compute

$$\hat{\gamma}_r = \frac{1}{n^2}\mathrm{tr}(\mathbf{HKHL}_r) \tag{3.5}$$

10. The estimated p-value is the proportion of $\hat{\gamma}_1, \ldots, \hat{\gamma}_N$ that is less than $\hat{\gamma}_{k,l}(\mathbf{X}, \mathbf{Y}; \hat{\mathbf{F}}_{\mathbf{Z}})$.

$$p\text{-value} = \frac{1}{n}\sum_{r=1}^{n} I(\hat{\gamma}_{k,l}(\mathbf{X}, \mathbf{Y}; \hat{\mathbf{F}}_{\mathbf{Z}}) < \hat{\gamma}_r)$$

These three steps cost $O(Nn^2)$.

The copula version of a linear-time version of the above test statistic iss introduced in Section 2.5, which reduces the cost of the algorithm.

8. Set the number of permutations N (for example, $N = 1000$)

9. For $r = 1, \ldots, N$,

   (a) Generate a random permutation $(s_1, \ldots, s_n)$ of $1, \ldots, n$.

   (b) Apply the above permutation to both columns and rows of $\mathbf{B}$ to obtain $\mathbf{B}_r$. To be specific, denote $\mathbf{B} = (b_{ij})_{i,j=1,\ldots n}$. Then $\mathbf{B}_r = (b_{s_i s_j})_{i,j=1,\ldots n}$.

   (c) Compute

$$\hat{\gamma}_r = \frac{1}{n^2}\mathrm{tr}(\mathbf{HAHB}_r) \tag{3.6}$$

   (d) The estimated p-value is the proportion of $\hat{\gamma}_1, \ldots, \hat{\gamma}_N$ that is less than $\hat{\gamma}_{d_p,d_q}(\mathbf{X}, \mathbf{Y}; \hat{\mathbf{F}}_{\mathbf{Z}})$.

$$p\text{-value} = \frac{1}{N}\sum_{r=1}^{N} I(\hat{\gamma}_r < \hat{\gamma}_{d_p,d_q}(\mathbf{X}, \mathbf{Y}; \hat{\mathbf{F}}_{\mathbf{Z}}))$$

These two steps cost $O(Nn^2)$.

The copula version of a linear-time version of the above test statistics are introduced in Section 2.5, which reduce the cost of the algorithm.

3.2 Asymptotic Theory for the CHSIC Test Statistic

The asymptotic theory of $\hat{\gamma}_{k,l}(\mathbf{X},\mathbf{Y};\hat{\mathbf{F}}_{\mathbf{Z}})$, defined in (3.3), is given in Theorem 3.2.1 for the dependent case, and in Theorem 3.2.2 for the independent case.

Consider first the dependent case and define

$$
\begin{aligned}
\eta_{k,l}(\mathbf{Z}_{(i,j,r,s,t)}) &= \sum_{\alpha\in\mathbb{N}_{4(p+q)}} D^{(\alpha)}h_{k,l}(\mathbf{W}_{(i,j,r,s)})\left[\vec{I}_{\mathbf{Z}_{(t,t,t,t)}}(\mathbf{Z}_{(i,j,r,s)})-\mathbf{W}_{(i,j,r,s)}\right]^{(\alpha)},\\
\tilde{\eta}_{k,l}(\mathbf{Z}_{(1,2,3,4,5)}) &= \frac{1}{5!}\sum_{(i_1,\ldots,i_5)\in I_5^5}\tilde{\eta}_{k,l}(\mathbf{Z}_{(i_1,i_2,i_3,i_4,i_5)}),\\
\tilde{\eta}^{(1)}_{k,l}(\mathbf{z}) &= E\left[\tilde{\eta}_{k,l}(\mathbf{Z}_{(1,2,3,4,5)})|\mathbf{Z}_1=\mathbf{z}\right]. \qquad (3.7)
\end{aligned}
$$

Thus, $\eta_{k,l}$ is the kernel of the V-statistic obtained in (2.14) for $M=2$.

Theorem 3.2.1. *Suppose that all second-order partial derivatives of k and l exist on $(0,1)^{2p}$ and $(0,1)^{2q}$, respectively, and they are all bounded by M_2. Then*

$$
\hat{\gamma}_{k,l}(\mathbf{X},\mathbf{Y};\hat{\mathbf{F}}_{\mathbf{Z}})-\hat{\gamma}_{k,l}(\mathbf{X},\mathbf{Y};\mathbf{F}_{\mathbf{Z}})=\frac{5}{n}\sum_{i=1}^{n}\tilde{\eta}^{(1)}_{k,l}(\mathbf{Z}_i)+O_P(n^{-1}), \qquad (3.8)
$$

where $\hat{\gamma}_{k,l}(\mathbf{X},\mathbf{Y};\mathbf{F}_{\mathbf{Z}})$ is defined by replacing f_1, f_2, and $G_{\mathbf{Z}}$ in (2.12) by k,l, and $\mathbf{F}_{\mathbf{Z}}$, respectively, and $\tilde{\eta}^{(1)}_{k,l}$ is defined in (3.7). Moreover,

$$
\sqrt{n}\left\{\hat{\gamma}_{k,l}(\mathbf{X},\mathbf{Y};\hat{\mathbf{F}}_{\mathbf{Z}})-\gamma_{k,l}(\mathbf{X},\mathbf{Y};\mathbf{F}_{\mathbf{Z}})\right\}\xrightarrow{D}N(0,\sigma^2_{k,l}), \qquad (3.9)
$$

where $\sigma^2_{k,l}=var\left[4\tilde{h}^{(1)}_{k,l}(\mathbf{Z};\mathbf{F}_{\mathbf{Z}})+5\tilde{\eta}^{(1)}_{k,l}(\mathbf{Z})\right]$, with $\tilde{h}^{(1)}_{k,l}$ defined by replacing f_1,f_2 with k,l in (2.21).

Consider now the independence case, so $\gamma_{k,l}(\mathbf{X},\mathbf{Y};\mathbf{F}_{\mathbf{Z}})=0$, and set

$$
\zeta_{k,l}(\mathbf{Z}_{(i,j,r,s,t_1,t_2)})=\frac{1}{2!}\sum_{\boldsymbol{\alpha}\in\mathbb{N}^2_{4(p+q)}} D^{(\alpha)}h_{k,l}(\mathbf{W}_{(i,j,r,s)})\prod_{c=1}^{2}\left[\vec{I}_{\mathbf{Z}_{(t_c,t_c,t_c,t_c)}}(\mathbf{Z}_{(i,j,r,s)})-\mathbf{W}_{(i,j,r,s)}\right]^{(\alpha_c)}.
$$

Thus, $\eta_{k,l}(\mathbf{Z}_{(i,j,r,s,t_1)}) + \zeta_{k,l}(\mathbf{Z}_{(i,j,r,s,t_1,t_2)})$, where $\eta_{k,l}$ is defined in (3.7), is the kernel of the V-statistic obtained in the Taylor expansion in (2.14) for $M = 3$. With $h_{f_1,f_2}(\mathbf{z}_{(1,2,3,4)}; \mathbf{G_Z})$ as defined in (2.6), (2.5), define

$$\zeta_{k,l,Total}(\mathbf{Z}_{(i,j,r,s,t_1,t_2)}) = h_{k,l}(\mathbf{Z}_{(i,j,r,s)}; \mathbf{F_Z}) + \eta_{k,l}(\mathbf{Z}_{(i,j,r,s,t_1)}) + \zeta_{k,l}(\mathbf{Z}_{(i,j,r,s,t_1,t_2)}), \quad (3.10)$$

$$\tilde{\zeta}_{k,l,Total}(\mathbf{Z}_{(1,\dots,6)}) = \frac{1}{6!} \sum_{(i_1,\dots,i_6)\in I_6^6} \zeta_{k,l,Total}(\mathbf{Z}_{(i_1,\dots,i_6)}), \quad (3.11)$$

$$\tilde{\zeta}^{(1)}_{k,l,Total}(\mathbf{z}_1) = E(\tilde{\zeta}_{k,l,Total}(\mathbf{Z}_{(1,\dots,6)})|\mathbf{Z}_1 = \mathbf{z}_1) \quad (3.12)$$

$$\tilde{\zeta}^{(2)}_{k,l,Total}(\mathbf{z}_1, \mathbf{z}_2) = E(\tilde{\zeta}_{k,l,Total}(\mathbf{Z}_{(1,\dots,6)})|\mathbf{Z}_1 = \mathbf{z}_1, \mathbf{Z}_2 = \mathbf{z}_2). \quad (3.13)$$

Theorem 3.2.2. *Suppose that all third-order partial derivatives of k and l exist on $(0,1)^{2p}$ and $(0,1)^{2q}$, respectively, and they are all bounded by M_3. Then, if $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$,*

$$\hat{\gamma}_{k,l}(\mathbf{X}, \mathbf{Y}; \hat{\mathbf{F}}_\mathbf{Z}) = \frac{1}{n^2} \sum_{i,j=1}^{n} \tilde{\zeta}^{(2)}_{k,l,Total}(\mathbf{Z}_i, \mathbf{Z}_j) + O_P(n^{-3/2}) \ . \quad (3.14)$$

Moreover, if $\{\lambda_i\}$ are the eigenvalues of the operator A defined on $L_2(\mathbb{R}^{p+q}, P_\mathbf{Z})$ by

$$(Ag)(\mathbf{z}) = \int_{-\infty}^{\infty} 15\tilde{\zeta}^{(2)}_{k,l,Total}(\mathbf{z}, \mathbf{z}')g(\mathbf{z}')dP_\mathbf{Z}(\mathbf{z}'), \quad \mathbf{z} \in \mathbb{R}^{p+q}, \ g \in L_2(\mathbb{R}^{p+q}, P_\mathbf{Z}) \quad (3.15)$$

we have

$$n\hat{\gamma}_{k,l}(\mathbf{X}, \mathbf{Y}; \hat{\mathbf{F}}_\mathbf{Z}) \xrightarrow{D} \sum_{i=1}^{\infty} \lambda_i \chi^2_{1i} \quad (3.16)$$

where $\{\chi^2_{1i}\}$ are independent χ^2_1 random variables.

3.3 Conservative Test with the CHSIC Statistic

This section presents a result in the spirit of Theorem 6 of [30]. Set $\lambda_{sum} = \sum_{i=1}^{\infty} \lambda_i$ and

$$\hat{\lambda}_{sum} = \left\{ \frac{1}{n} \sum_{i=1}^{n} k(\hat{\mathbf{F}}_\mathbf{X}(\mathbf{X}_i), \hat{\mathbf{F}}_\mathbf{X}(\mathbf{X}_i)) - \frac{1}{n^2} \sum_{i,j=1}^{n} k(\hat{\mathbf{F}}_\mathbf{X}(\mathbf{X}_i), \hat{\mathbf{F}}_\mathbf{X}(\mathbf{X}_j)) \right\} \cdot \left\{ \frac{1}{n} \sum_{i=1}^{n} l(\hat{\mathbf{F}}_\mathbf{Y}(\mathbf{Y}_i), \hat{\mathbf{F}}_\mathbf{Y}(\mathbf{Y}_i)) - \frac{1}{n^2} \sum_{i,j=1}^{n} l(\hat{\mathbf{F}}_\mathbf{Y}(\mathbf{Y}_i), \hat{\mathbf{F}}_\mathbf{Y}(\mathbf{Y}_j)) \right\}. \quad (3.17)$$

Theorem 3.3.1. *Suppose that all assumptions in Theorem 3.2.2 are satisfied. If*

$\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$, *then*

$$\frac{n\hat{\gamma}_{k,l}(\mathbf{X},\mathbf{Y};\hat{\mathbf{F}}_{\mathbf{Z}})}{\hat{\lambda}_{sum}} \xrightarrow{D} T \tag{3.18}$$

where $T = \sum_{i=1}^{\infty} \frac{\lambda_i}{\lambda_{sum}} \chi^2_{1i}$ *and has mean* $E(T) = 1$.

By similar arguments as in Theorem 6 of [30], a conservative test with significance level $\alpha \in (0, 0.215]$ can be constructed in which the independence of $\mathbf{X}$ and $\mathbf{Y}$ is rejected when

$$\frac{n\hat{\gamma}_{k,l}(\mathbf{X},\mathbf{Y};\hat{\mathbf{F}}_{\mathbf{Z}})}{\hat{\lambda}_{sum}} \geq \left[\Phi^{-1}(1-\alpha/2)\right]^2$$

where Φ is the cdf of the standard normal distribution. This conservative test has asymptotic significant level at most α, and it costs much less time than the permutation approach. However, it can be quite conservative for some distributions.

3.4 Copula Versions of Linear-Time Statistic

This section presents another estimator for $\gamma_{k,l}(\mathbf{X},\mathbf{Y};\mathbf{F}_{\mathbf{Z}})$ for which the computation time will be much shorter.

Suppose that $n = 4m$, $m \in \mathbb{N}$. Similar to Section 2.5, the copula versions of the linear-time statistic for estimating $\gamma_{k,l}(\mathbf{X},\mathbf{Y};\mathbf{F}_{\mathbf{Z}})$ is defined as

$$\hat{\gamma}_{k,l,linear}(\mathbf{X},\mathbf{Y};\hat{\mathbf{F}}_{\mathbf{Z}}) = \frac{1}{m}\sum_{i=1}^{m} h_{k,l}(\mathbf{Z}_{(4i-3,4i-2,4i-1,4i)};\hat{\mathbf{F}}_{\mathbf{Z}}). \tag{3.19}$$

By the same arguments as in Section 2.5, it can be shown that

$$\sqrt{n}\left[\hat{\gamma}_{k,l,linear}(\mathbf{X},\mathbf{Y};\hat{\mathbf{F}}_{\mathbf{Z}}) - \gamma_{k,l}(\mathbf{X},\mathbf{Y};\mathbf{F}_{\mathbf{Z}})\right]$$

is asymptotic normal.

This version of estimator $\hat{\gamma}_{k,l,linear}(\mathbf{X},\mathbf{Y};\hat{\mathbf{F}}_{\mathbf{Z}})$ is “linear" in the sense that the computation cost is $O(n)$ given the ranks of the observations. With a ranking method that costs $O(n\log n)$, a permutation test using $\hat{\gamma}_{k,l,linear}(\mathbf{X},\mathbf{Y};\hat{\mathbf{F}}_{\mathbf{Z}})$ costs $O(Nn + n\log n)$ where N is the number of permutations.

3.5 Simulation Studies

In this section, the performance of the four methods CdCov, CHSIC, dCov and HSIC are evaluated and compared.

3.5.1 Simulation 1

In Simulation 1, the significance levels of the four methods are studied. The setting is as follows:

- $p = 6$, $q = 4$. All 10 components of $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$ are i.i.d. with one of the following 4 distributions:

  standard normal, standard exponential, standard lognormal and t_2.

- The sample size n equals one of the following:

  25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200.

- For each combination of the above parameters, the simulation is repeated 2000 times.

- The Gaussian kernels are used. The permutation approach is used to compute the p-values. The number of permutation is $N = 1000$.

- The null hypothesis is rejected when the computed p-value is less than 0.05.

The results are summarized in Figure 3.1. Given the distribution of $\mathbf{Z}$ and the sample size, the value of each dot in the figure is the proportion of simulations in which the null hypothesis that $\mathbf{X}$ and $\mathbf{Y}$ are independent is rejected, which estimate the significant level of the test. As expected, all these values are close to 0.05 .

3.5.2 Simulation 2

In Simulation 2, the powers of the four methods are studied. The setting is as follows:

- $p = 6$, $q = 4$. Let $\mathbf{Z} = (\mathbf{X}, \mathbf{Y}) = \mathbf{\Sigma}^{1/2}\mathbf{W}$. $\mathbf{W}$ is a 10-dimensional random vector whose components are i.i.d. with one of the following 4 distributions:

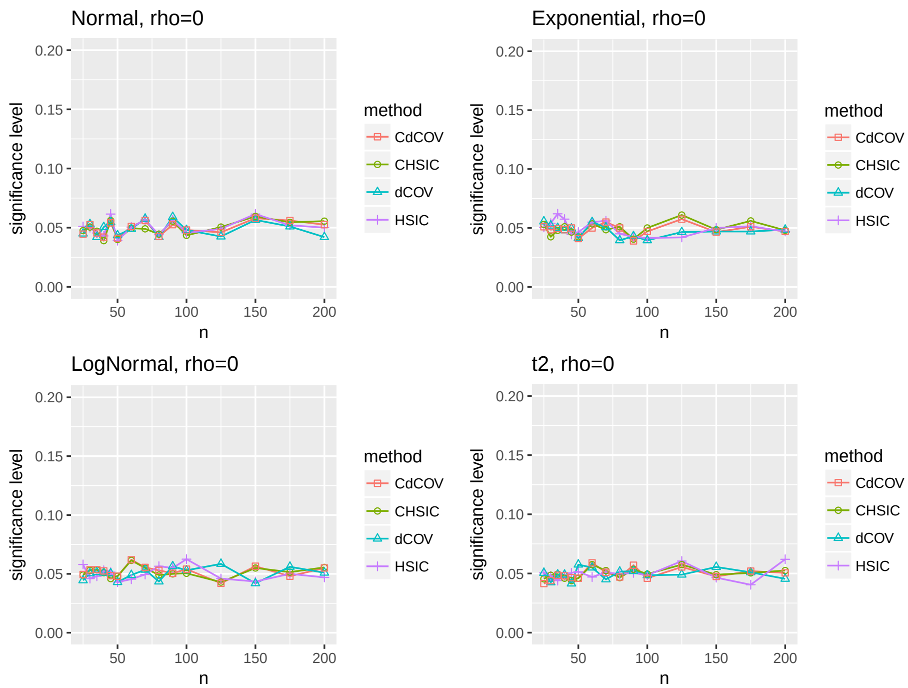

Figure 3.1. Summarized results in Simulation 1

standard normal, standard exponential, standard lognormal and t_2.

$\mathbf{\Sigma}$ is a 10×10 matrix with diagonal elements being 1 and off-diagonal elements being ρ, where ρ equals one of the following:

0.05, 0.1 .

- The sample size n equals one of the following:

  25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200.

- For each combination of the above parameters, the simulation is repeated 2000 times.

- The Gaussian kernels are used. The permutation approach is used to compute the p-values. The number of permutation is $N = 1000$.

- The null hypothesis is rejected when the computed p-value is less than 0.05.

The results are summarized in Figure 3.2 and Figure 3.3. Given the distribution of $\mathbf{Z}$ and the sample size, the value of each dot in the figure is the proportion of simulations in which the null hypothesis that $\mathbf{X}$ and $\mathbf{Y}$ are independent is rejected,

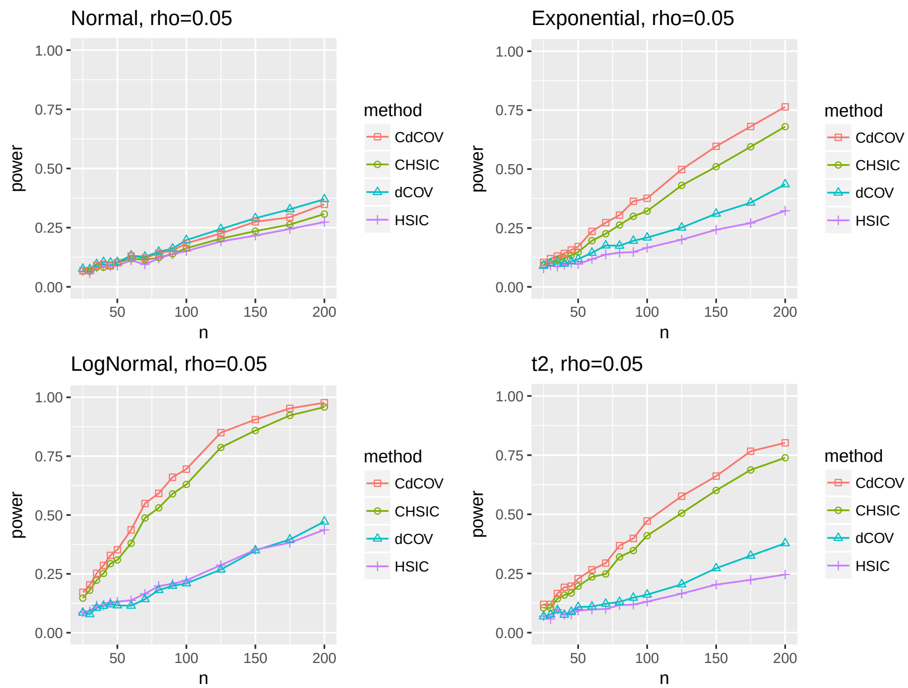

Figure 3.2. Summarized results in Simulation 2 (Part 1)

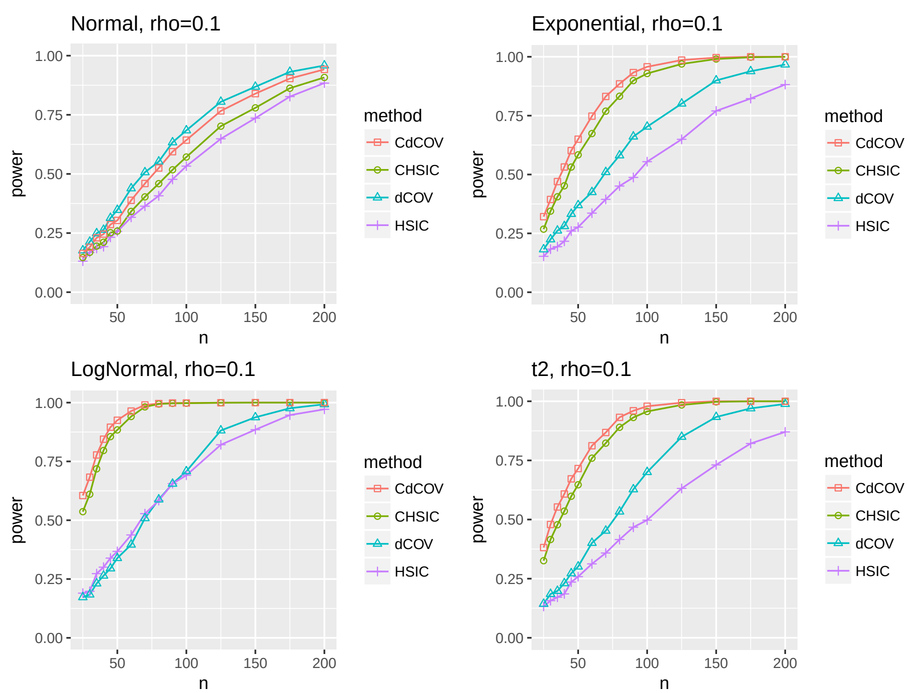

Figure 3.3. Summarized results in Simulation 2 (Part 2)

which estimate the power of the test. As expected, the powers increase as n increases. In the case of normal distribution, the performance of different methods are similar. In the other cases, the powers of CdCov and CHSIC are much higher than dCov and HSIC.

3.5.3 Simulation 3

In Simulation 3, the effect of the choice of the kernels on CHSIC is studied. The setting is as follows:

- $p = 6$, $q = 4$. Let $\mathbf{Z} = (\mathbf{X}, \mathbf{Y}) = \mathbf{\Sigma}^{1/2}\mathbf{W}$. $\mathbf{W}$ is a 10-dimensional vector whose components are i.i.d. standard exponential distribution.

  $\mathbf{\Sigma}$ is a 10×10 matrix with diagonal elements being 1 and off-diagonal elements being $\rho = 0.05$.

- The sample size n equals one of the following:

  10, 20, 40, 80, 160, 320.

- For each combination of the above parameters, the simulation is repeated 1000 times.

- One of the following kernels is used:

  kernels 2, 3 and 4 in Section 1.3.1

- The permutation approach is used to compute the p-values. The number of permutation is $N = 1000$.

- The null hypothesis is rejected when the computed p-value is less than 0.05.

The results are summarized in Figure 3.4. The value of each dot in the figure is the proportion of simulations in which the null hypothesis that $\mathbf{X}$ and $\mathbf{Y}$ are independent is rejected, which estimates the power of the test. In these results, the choice of kernels does not have significant effect on the power of the test.

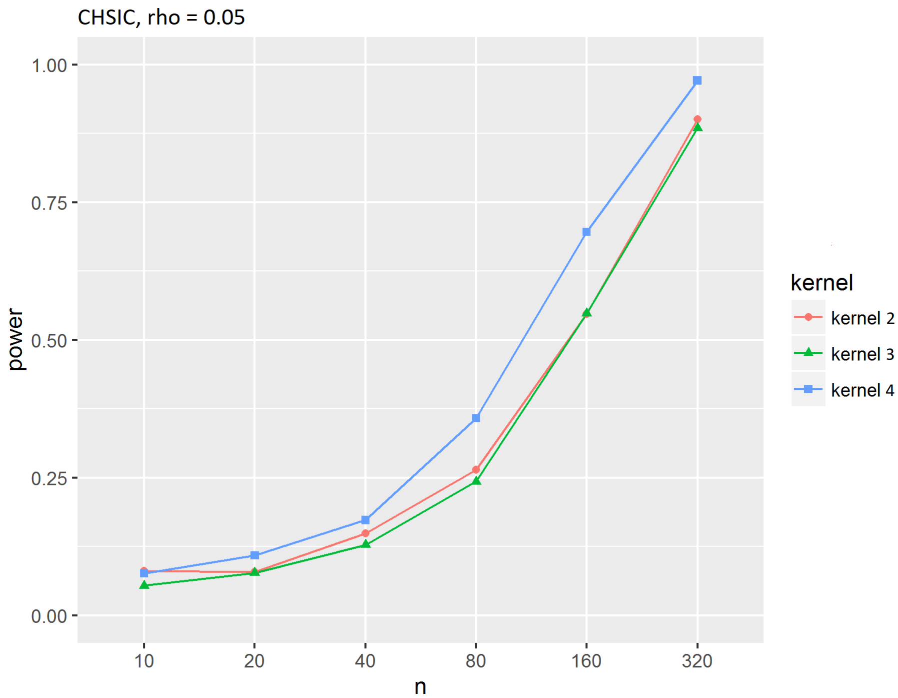

Figure 3.4. Summarized results in Simulation 3

3.A Appendix: Proof of Theorems

3.A.1 Proof of Theorem 3.2.1

Using the Taylor expansion in (2.14) for $M = 2$, write

$$\hat{\gamma}_{k,l}(\mathbf{X},\mathbf{Y};\hat{\mathbf{F}}_{\mathbf{Z}}) - \hat{\gamma}_{k,l}(\mathbf{X},\mathbf{Y};\mathbf{F}_{\mathbf{Z}}) = \Delta\gamma_2 + R_2, \tag{3.20}$$

where

$$\begin{aligned} \Delta\gamma_2 &= \frac{1}{n^5}\sum_{i,j,r,s,t_1=1}^{n}\sum_{\alpha\in\mathbb{N}_{4(p+q)}} D^{(\alpha)}h_{k,l}(\mathbf{W}_{(i,j,r,s)})\left[\vec{I}_{\mathbf{Z}_{(t_1,t_1,t_1,t_1)}}(\mathbf{Z}_{(i,j,r,s)}) - \mathbf{W}_{(i,j,r,s)}\right]^{(\alpha)}, \\ R_2 &= \frac{1}{n^4}\sum_{i,j,r,s=1}^{n}\frac{1}{2}\sum_{\boldsymbol{\alpha}\in\mathbb{N}^2_{4(p+q)}} D^{(\boldsymbol{\alpha})}h_{k,l}(\tilde{\mathbf{W}}_{(i,j,r,s)})\left[\hat{\mathbf{W}}_{(i,j,r,s)} - \mathbf{W}_{(i,j,r,s)}\right]^{(\boldsymbol{\alpha})} \end{aligned} \tag{3.21}$$

for some $\tilde{\mathbf{W}}_{(i,j,r,s)} \in [\mathbf{W}_{(i,j,r,s)}, \hat{\mathbf{W}}_{(i,j,r,s)}]$. Since for any $|\boldsymbol{\alpha}| = 2$, $\|D^{(\boldsymbol{\alpha})}h_{k,l}(\cdot)\|_\infty \le C$, for some finite constant C that depends on the bounds of k, l as well as the bounds of their first and second order derivatives,

$$|R_2| \le \frac{1}{n^4} \sum_{i,j,r,s=1}^{n} \frac{1}{2} \sum_{\boldsymbol{\alpha} \in \mathbb{N}^2_{4(p+q)}} |D^{(\boldsymbol{\alpha})} h_{k,l}(\tilde{\mathbf{W}}_{(i,j,r,s)})| \cdot \|\hat{\mathbf{F}}_{\mathbf{Z}} - \mathbf{F}_{\mathbf{Z}}\|^2_\infty \le \frac{1}{2}[4(p+q)]^2 C \|\hat{\mathbf{F}}_{\mathbf{Z}} - \mathbf{F}_{\mathbf{Z}}\|^2_\infty$$

which is $O_P(n^{-1})$. Thus

$$\hat{\gamma}_{k,l}(\mathbf{X}, \mathbf{Y}; \hat{\mathbf{F}}_{\mathbf{Z}}) - \hat{\gamma}_{k,l}(\mathbf{X}, \mathbf{Y}; \mathbf{F}_{\mathbf{Z}}) = \Delta\gamma_2 + O_P(n^{-1}). \tag{3.22}$$

By Lemma A.1.5,

$$\hat{\gamma}_{k,l}(\mathbf{X}, \mathbf{Y}; \mathbf{F}_{\mathbf{Z}}) - \gamma_{k,l}(\mathbf{X}, \mathbf{Y}; \mathbf{F}_{\mathbf{Z}}) = \frac{4}{n} \sum_{i=1}^{n} \tilde{h}^{(1)}_{k,l}(\mathbf{Z}_i; \mathbf{F}_{\mathbf{Z}}) + O_P(n^{-1}). \tag{3.23}$$

By the definition of $\Delta\gamma_2$ and a similar V-statistics result,

$$\Delta\gamma_2 = \frac{1}{n^5} \sum_{i,j,r,s,t=1}^{n} \eta_{k,l}(\mathbf{Z}_{(i,j,r,s,t)}) = \frac{5}{n} \sum_{i=1}^{n} \tilde{\eta}^{(1)}_{k,l}(\mathbf{Z}_i) + O_P(n^{-1}). \tag{3.24}$$

Relations (3.22) and (3.24) imply that (3.8) holds. Relations (3.22), (3.23) and (3.24) yield

$$\hat{\gamma}_{k,l}(\mathbf{X}, \mathbf{Y}; \hat{\mathbf{F}}_{\mathbf{Z}}) - \gamma_{k,l}(\mathbf{X}, \mathbf{Y}; \mathbf{F}_{\mathbf{Z}}) = \frac{1}{n} \sum_{i=1}^{n} \left[4\tilde{h}^{(1)}_{k,l}(\mathbf{Z}_i; \mathbf{F}_{\mathbf{Z}}) + 5\tilde{\eta}^{(1)}_{k,l}(\mathbf{Z}_i)\right] + O_P(n^{-1})$$

and thus (3.9) follows by the CLT.

3.A.2 Proof of Theorem 3.2.2

Using the Taylor expansion in (2.14) for $M = 3$, the difference $\Delta\gamma_1 = \hat{\gamma}_{k,l}(\mathbf{X}, \mathbf{Y}; \hat{\mathbf{F}}_{\mathbf{Z}}) - \hat{\gamma}_{k,l}(\mathbf{X}, \mathbf{Y}; \mathbf{F}_{\mathbf{Z}})$ can be written as

$$\Delta\gamma_1 = \Delta\gamma_2 + R_2, \tag{3.25}$$

where

$$\Delta\gamma_2 = \frac{1}{n^6} \sum_{i,j,r,s,t_1,t_2=1}^{n} \sum_{m=1}^{2} \frac{1}{m!} \cdot \sum_{\boldsymbol{\alpha} \in \mathbb{N}^m_{4(p+q)}} D^{(\boldsymbol{\alpha})} h_{k,l}(\mathbf{W}_{(i,j,r,s)})$$

$$\cdot \prod_{c=1}^{m} \left[\vec{I}_{\mathbf{Z}_{(t_c,t_c,t_c,t_c)}}(\mathbf{Z}_{(i,j,r,s)}) - \mathbf{W}_{(i,j,r,s)} \right]^{(\alpha_c)}$$

$$R_2 = \frac{1}{n^4} \sum_{i,j,r,s=1}^{n} \frac{1}{6} \sum_{\boldsymbol{\alpha} \in \mathbb{N}^3_{4(p+q)}} D^{(\boldsymbol{\alpha})} h_{k,l}(\tilde{\mathbf{W}}_{(i,j,r,s)}) \left[\hat{\mathbf{W}}_{(i,j,r,s)} - \mathbf{W}_{(i,j,r,s)} \right]^{(\boldsymbol{\alpha})} \quad (3.26)$$

for some $\tilde{\mathbf{W}}_{(i,j,r,s)} \in [\mathbf{W}_{(i,j,r,s)}, \hat{\mathbf{W}}_{(i,j,r,s)}]$. Since $|D^{(\boldsymbol{\alpha})} h_{k,l}(\tilde{\mathbf{W}}_{(i,j,r,s)})| \leq M_3$ for any $|\boldsymbol{\alpha}| = 3$,

$$|R_2| \leq \frac{1}{n^4} \sum_{i,j,r,s=1}^{n} \frac{1}{6} \sum_{\boldsymbol{\alpha} \in \mathbb{N}^3_{4(p+q)}} |D^{(\boldsymbol{\alpha})} h_{k,l}(\tilde{\mathbf{W}}_{(i,j,r,s)})| \cdot \|\hat{\mathbf{F}}_{\mathbf{Z}} - \mathbf{F}_{\mathbf{Z}}\|_\infty^3$$
$$\leq C_{p,q} M_3 \|\hat{\mathbf{F}}_{\mathbf{Z}} - \mathbf{F}_{\mathbf{Z}}\|_\infty^3 = O_P(n^{-3/2})$$

which means that

$$\hat{\gamma}_{k,l}(\mathbf{X}, \mathbf{Y}; \hat{\mathbf{F}}_{\mathbf{Z}}) - \hat{\gamma}_{k,l}(\mathbf{X}, \mathbf{Y}; \mathbf{F}_{\mathbf{Z}}) = \Delta\gamma_2 + O_P(n^{-3/2}) \quad (3.27)$$

Thus by the definition of $\Delta\gamma_2$ in (3.26), and recalling the definition of $\hat{\gamma}_{k,l}(\mathbf{X}, \mathbf{Y}; \mathbf{F}_{\mathbf{Z}})$ given in Theorem 3.2.1,

$$\begin{aligned} \hat{\gamma}_{k,l}(\mathbf{X}, \mathbf{Y}; \hat{\mathbf{F}}_{\mathbf{Z}}) &= \frac{1}{n^6} \sum_{i,j,r,s,t_1,t_2=1}^{n} \zeta_{k,l,Total}(\mathbf{Z}_{(i,j,r,s,t_1,t_2)}) + O_P(n^{-3/2}) \\ &= \frac{1}{n^6} \sum_{i,j,r,s,t_1,t_2=1}^{n} \tilde{\zeta}_{k,l,Total}(\mathbf{Z}_{(i,j,r,s,t_1,t_2)}) + O_P(n^{-3/2}) \quad (3.28) \end{aligned}$$

where $\zeta_{k,l,Total}$ is defined in (3.10), and $\tilde{\zeta}_{d_p,d_q,Total}$ is defined in (3.11). Let $\tilde{\zeta}^{(1)}_{k,l,Total}$ and $\tilde{\zeta}^{(2)}_{k,l,Total}$ be defined in (3.12) and (3.13), respectively. In Lemma A.1.4 it is shown that $\tilde{\zeta}^{(1)}_{k,l,Total} = 0$ so that, by Lemma 1.4.2,

$$\frac{1}{n^6} \sum_{i,j,r,s,t_1,t_2=1}^{n} \tilde{\zeta}_{k,l,Total}(\mathbf{Z}_{(i,j,r,s,t_1,t_2)}) = \frac{15}{n^6} \sum_{i_1,i_2=1}^{n} \tilde{\zeta}^{(2)}_{k,l,Total}(\mathbf{Z}_{i_1}, \mathbf{Z}_{i_2}) + O_P(n^{-3/2}). \quad (3.29)$$

Relations (3.28) and (3.29) yield (3.14). By a generalization of Theorem B in Section 6.4.1 in [23] to random vector observations, we have

$$n \cdot \frac{1}{n^2} \sum_{i_1,i_2=1}^{n} 15\tilde{\zeta}^{(2)}_{d_p,d_q,Total}(\mathbf{Z}_{i_1}, \mathbf{Z}_{i_2}) \xrightarrow{D} \sum_{i=1}^{\infty} \mu_i \chi^2_{1i} \quad (3.30)$$

Then (2.33) follows from (3.14) and (3.30).

3.A.3 Proof of Theorem 3.3.1

The proof of Theorem 3.3.1 proceeds by similar arguments as of 2.4.1 and is omitted.

Chapter 4
Copula Version of Criteria in Two-Sample Problem

4.1 Introduction

In this chapter, we suppose that $\mathcal{X} = \mathcal{Y}$, the two random vectors are defined on the same domain, and $\mathbf{X}$ and $\mathbf{Y}$ are independent. In the two-sample problem, our goal is to test whether $\mathbf{X}$ and $\mathbf{Y}$ have the same distribution. The e-distance criterion and maximum mean discrepancy are two criteria for two-sample problem, which will be introduced in this section after the introduction of some notations.

For any functions f such that $f : \mathbb{R}^{2p} \times \mathbb{R}^{2p} \to \mathbb{R}$, define the function $g_f : \mathbb{R}^{4p} \to \mathbb{R}$ by

$$g_f(\mathbf{u}_1, \mathbf{u}_2, \mathbf{v}_1, \mathbf{v}_2) = f(\mathbf{u}_1, \mathbf{u}_2) + f(\mathbf{v}_1, \mathbf{v}_2) - 2f(\mathbf{u}_1, \mathbf{v}_1) \tag{4.1}$$

where $\mathbf{u}_i, \mathbf{v}_i \in \mathbb{R}^p,\ i = 1, 2$. Next, for any function $\mathbf{G} : \mathbb{R}^p \to \mathbb{R}^p$, define

$$\begin{aligned} g_f(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}_1, \mathbf{y}_2; \mathbf{G}) &= g_f(\mathbf{G}(\mathbf{x}_1), \mathbf{G}(\mathbf{x}_2), \mathbf{G}(\mathbf{y}_1), \mathbf{G}(\mathbf{y}_2)) \\ &= f(\mathbf{G}(\mathbf{x}_1), \mathbf{G}(\mathbf{x}_2)) + f(\mathbf{G}(\mathbf{y}_1), \mathbf{G}(\mathbf{y}_1)) - 2f(\mathbf{G}(\mathbf{x}_1), \mathbf{G}(\mathbf{y}_1)) \end{aligned}$$

for any $\mathbf{z}_i = (\mathbf{x}_i, \mathbf{y}_i), \mathbf{x}_i, \mathbf{y}_i \in \mathbb{R}^p, i = 1, 2$. Define

$$\xi_f(\mathbf{X}, \mathbf{Y}; \mathbf{G}) = E\left[g_f(\mathbf{X}, \mathbf{X}', \mathbf{Y}, \mathbf{Y}'; \mathbf{G})\right]. \tag{4.2}$$

where $(\mathbf{X}', \mathbf{Y}')$ is an independent copy of $(\mathbf{X}, \mathbf{Y})$.

The maximum mean discrepancy (MMD) criterion was introduced by [10]. Suppose that k is a universal kernel on $\mathcal{X}$. Then the MMD of $\mathbf{X}$ and $\mathbf{Y}$ can bes defined by using the notation in (4.2) as

$$\xi_k(\mathbf{X}, \mathbf{Y}; \mathrm{id}) = E[k(\mathbf{X}, \mathbf{X}')] + E[k(\mathbf{Y}, \mathbf{Y}')] - 2E[k(\mathbf{X}, \mathbf{Y})] \tag{4.3}$$

assuming that all expectations are finite, where $(\mathbf{X}', \mathbf{Y}')$ is an independent copy of $(\mathbf{X}, \mathbf{Y})$. In [10] it is shown that $\xi_k(\mathbf{X}, \mathbf{Y}; \mathrm{id})$ is always non-negative, and

$$\xi_k(\mathbf{X}, \mathbf{Y}; \mathrm{id}) = 0 \iff \mathbf{X} \stackrel{D}{=} \mathbf{Y}. \tag{4.4}$$

The e-distance (energy distance) criterion was introduced by [28] and [4]. Suppose that $\mathbf{X}$ and $\mathbf{Y}$ have finite first moments. Then the e-distance of $\mathbf{X}$ and $\mathbf{Y}$ can be defined by using the notation in (4.2) as

$$-\xi_{d_p}(\mathbf{X}, \mathbf{Y}; \mathrm{id}) = 2E[d_p(\mathbf{X}, \mathbf{Y})] - E[d_p(\mathbf{X}, \mathbf{X}')] - E[d_q(\mathbf{Y}, \mathbf{Y}')] \tag{4.5}$$

where $(\mathbf{X}', \mathbf{Y}')$ is an independent copy of $(\mathbf{X}, \mathbf{Y})$. In [4] it is shown that $-\xi_{d_p}(\mathbf{X}, \mathbf{Y}; \mathrm{id})$ is always non-negative, and

$$-\xi_{d_p}(\mathbf{X}, \mathbf{Y}; \mathrm{id}) = 0 \iff \mathbf{X} \stackrel{D}{=} \mathbf{Y}. \tag{4.6}$$

4.2 Copula Copula MMD (CMMD) and E-Distance Criterion (CeD)

Lemma 2.2.1 can be generalized to the following lemma.

Lemma 4.2.1. *Let X be a random variable with cdf F_X, $H : \mathbb{R} \to \mathbb{R}$ be a bounded non-decreasing function, and $G : \mathbb{R} \to [0,1]$ be a function such that $G = (1-\lambda)F_X + \lambda H$ for some $\lambda \in (0,1)$. Then $G^{-1}(G(X)) = X$ with probability 1, where $G^{-1}(s) = \inf\{x : G(x) \geq s\}$.*

For any $\lambda \in (0,1)$, define $\mathbf{G}_\lambda = (1-\lambda)\mathbf{F}_\mathbf{X} + \lambda \mathbf{F}_\mathbf{Y}$. Then by Lemma 4.2.1, the following two statements are equivalent.

$$\mathbf{G}_\lambda(\mathbf{X}) \stackrel{D}{=} \mathbf{G}_\lambda(\mathbf{Y}) \iff \mathbf{X} \stackrel{D}{=} \mathbf{Y}. \tag{4.7}$$

The *copula MMD criterion* (CMMD) states that

$$\xi_k(\mathbf{X}, \mathbf{Y}; \mathbf{G}_\lambda) = 0 \iff \mathbf{X} \overset{D}{=} \mathbf{Y}. \tag{4.8}$$

The *copula e-distance criterion* (CeD) states that

$$-\xi_f(\mathbf{X}, \mathbf{Y}; \mathbf{G}_\lambda) = 0 \iff \mathbf{X} \overset{D}{=} \mathbf{Y}. \tag{4.9}$$

From now on, let $\mathbf{X}_j, j = 1, \ldots, m$, be m independent copies of $\mathbf{X}$, and $\mathbf{Y}_j, j = 1, \ldots, n$, be n independent copies of $\mathbf{Y}$. The generalized V-statistic corresponding to the function defined in (4.2) is defined as

$$\hat{\xi}_f(\mathbf{X}, \mathbf{Y}; \mathbf{G}) = \frac{1}{m^2 n^2} \sum_{i,j=1}^{m} \sum_{r,s=1}^{n} g_f(\mathbf{X}_i, \mathbf{X}_j, \mathbf{Y}_r, \mathbf{Y}_s; \mathbf{G}) \tag{4.10}$$

Under the null hypothesis, $\mathbf{G}_\lambda = \mathbf{F}_\mathbf{X} = \mathbf{F}_\mathbf{Y}$ for any $\lambda \in (0, 1)$. Denote this common function by $\mathbf{F}$. Typically, $\mathbf{F}$ is estimated by the pooled empirical function $\hat{\mathbf{F}} = \frac{m}{m+n}\hat{\mathbf{F}}_\mathbf{X} + \frac{n}{m+n}\hat{\mathbf{F}}_\mathbf{Y}$, which is a weighted average of the two empirical functions from $\mathbf{X}_i, i = 1, \ldots, m$ and $\mathbf{Y}_j, j = 1, \ldots, n$, respectively.

Replacing f and $\mathbf{G}$ in (4.10) by a universal kernel k and $\hat{\mathbf{F}}$, respectively, leads to the proposed *CMMD test statistic*

$$\hat{\xi}_k(\mathbf{X}, \mathbf{Y}; \hat{\mathbf{F}}) = \frac{1}{n^2} \sum_{i,j=1}^{n} g_k(\mathbf{X}_i, \mathbf{X}_j, \mathbf{Y}_r, \mathbf{Y}_s; \hat{\mathbf{F}}). \tag{4.11}$$

Replacing f and $\mathbf{G}$ in (4.10) by d_p and $\hat{\mathbf{F}}$, respectively, leads to the proposed *CeD test statistic*

$$-\hat{\xi}_{d_p}(\mathbf{X}, \mathbf{Y}; \hat{\mathbf{F}}) = -\frac{1}{n^2} \sum_{i,j=1}^{n} g_{d_p}(\mathbf{X}_i, \mathbf{X}_j, \mathbf{Y}_r, \mathbf{Y}_s; \hat{\mathbf{F}}), \tag{4.12}$$

In order to compute the observed value of the CMMD test statistic, the pooled mid-rank of all the observations for each component of $\mathbf{X}$ and $\mathbf{Y}$ can be used as in (2.18). Denote $\mathbf{X}_i = (X_{i,1}, \ldots, X_{i,p})$ for $i = 1, \ldots, m$ and $\mathbf{Y}_j = (Y_{j,1}, \ldots, Y_{j,p})$ for $j = 1, \ldots, n$. The steps of the algorithm to compute the CMMD test statistic is as follows.

1. For each $r = 1, \ldots, p$, let $X^R_{1,r}, \ldots, X^R_{m,r}, Y^R_{1,r}, \ldots, Y^R_{n,r}$ be the mid-ranks of

$X_{1,r}, \ldots, X_{m,r}, Y_{1,r}, \ldots, Y_{n,r}$. Let $X_{i,r}^F = \frac{X_{i,r}^R}{m+n} - \frac{1}{2(m+n)}$ for $i = 1, \ldots, m$ and $Y_{j,r}^F = \frac{Y_{j,r}^R}{m+n} - \frac{1}{2(m+n)}$ for $j = 1, \ldots, n$.

2. For each $i = 1, \ldots, m$, write $\mathbf{X}_i^F = (X_{i,1}^F, \ldots, X_{i,p}^F)$.

3. For each $j = 1, \ldots, n$, write $\mathbf{Y}_j^F = (Y_{j,1}^F, \ldots, Y_{j,p}^F)$.

4. Construct three kernel matrices:

$$\mathbf{K_{XX}} = (k(\mathbf{X}_i^F, \mathbf{X}_j^F))_{i,j=1,\ldots,m},$$
$$\mathbf{K_{XY}} = (k(\mathbf{X}_i^F, \mathbf{Y}_j^F))_{i=1,\ldots,m; j=1,\ldots,n},$$
$$\mathbf{K_{YY}} = (k(\mathbf{Y}_i^F, \mathbf{Y}_j^F))_{i,j=1,\ldots,n},$$

which are combined to a $(m+n) \times (m+n)$ pooled kernel matrices:

$$\mathbf{K} = \begin{pmatrix} \mathbf{K_{XX}} & \mathbf{K_{XY}} \\ \mathbf{K_{XY}^T} & \mathbf{K_{YY}} \end{pmatrix} \tag{4.13}$$

5. Let $\bar{k}_{\mathbf{XX}}$ be the average of all elements in $\mathbf{K_{XX}}$, $\bar{k}_{\mathbf{XY}}$ be the average of all elements in $\mathbf{K_{XY}}$, and $\bar{k}_{\mathbf{YY}}$ be the average of all elements in $\mathbf{K_{YY}}$. Then

$$\hat{\xi}_k(\mathbf{X}, \mathbf{Y}; \hat{\mathbf{F}}) = \bar{k}_{\mathbf{XX}} + \bar{k}_{\mathbf{YY}} - 2\bar{k}_{\mathbf{XY}}. \tag{4.14}$$

This algorithm costs $O((m+n)^2)$.

A permutation approach can be used to construct a test for independence between $\mathbf{X}$ and $\mathbf{Y}$. To be specific, a few more steps will be added after the above algorithm as follows.

5. Set the number of permutations N (for example, $N = 1000$)

6. For $r = 1, \ldots, N$,

   (a) Generate a random permutation $(s_1, \ldots, s_{m+n})$ of $1, \ldots, m+n$.

   (b) Apply the above permutation to both columns and rows of $\mathbf{K}$ to obtain $\mathbf{K}_r$. To be specific, denote $\mathbf{K} = (k_{ij})_{i,j=1,\ldots m+n}$. Then $\mathbf{K}_r =$

$(k_{s_i s_j})_{i,j=1,\ldots m+n}$. Decompose $\mathbf{K}_r$ into four matrices as

$$\mathbf{K}_r = \begin{pmatrix} \mathbf{K}_{r,\mathbf{XX}} & \mathbf{K}_{r,\mathbf{XY}} \\ \mathbf{K}_{r,\mathbf{YX}} & \mathbf{K}_{r,\mathbf{YY}} \end{pmatrix} \tag{4.15}$$

where $vK_{r,\mathbf{XX}}$, $\mathbf{K}_{r,\mathbf{XY}}$, $\mathbf{K}_{r,\mathbf{YX}}$ and $\mathbf{K}_{r,\mathbf{YY}}$ are $m \times m$, $m \times n$, $n \times m$ and $n \times n$ matrices, respectively.

(c) Let $\bar{k}_{r,\mathbf{XX}}$ be the average of all elements in $\mathbf{K}_{r,\mathbf{XX}}$, $\bar{k}_{r,\mathbf{XY}}$ be the average of all elements in $\mathbf{K}_{r,\mathbf{XY}}$, and $\bar{k}_{r,\mathbf{YY}}$ be the average of all elements in $\mathbf{K}_{r,\mathbf{YY}}$. Compute

$$\hat{\gamma}_r = \bar{k}_{r,\mathbf{XX}} + \bar{k}_{r,\mathbf{YY}} - 2\bar{k}_{r,\mathbf{XY}} \ . \tag{4.16}$$

7. The estimated p-value is the proportion of $\hat{\gamma}_1, \ldots, \hat{\gamma}_N$ that is less than $\hat{\xi}_k(\mathbf{X}, \mathbf{Y}; \hat{\mathbf{F}})$.

$$p\text{-value} = \frac{1}{N} \sum_{r=1}^{N} I(\hat{\gamma}_r < \hat{\xi}_k(\mathbf{X}, \mathbf{Y}; \hat{\mathbf{F}}))$$

These three steps cost $O(N(m+n)^2)$.

The steps of the algorithm to compute the CeD test statistic and conduct the permutation test are almost the same as CMMD test statistic with k replaced by d_p.

4.3 Asymptotic Theory for the CMMD Test Statistic and CeD Test Statistic

In this section, the asymptotic results for the CMMD test statistic and CeD test statistic under the null hypothesis are given in Theorem 4.3.1 and 4.3.2.

For simplicity, let $\mathbf{U}_i = \mathbf{F}(\mathbf{X}_i)$ for $i = 1, \ldots, m$, and let $\mathbf{V}_j = \mathbf{F}(\mathbf{Y}_j)$ for $j = 1, \ldots, n$. The first theorem is about the CMMD test statistic, in which a universal kernel k is used. Define

$\psi_{k,1}(\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_{t_1}, \mathbf{Y}_r, \mathbf{Y}_s)$

$$= \sum_{\alpha \in \mathbb{N}_{4p}} D^{(\alpha)} g_k(\mathbf{U}_i, \mathbf{U}_j, \mathbf{V}_r, \mathbf{V}_s)(\vec{I}_{(\mathbf{x}_{t_1}, \mathbf{x}_{t_1}, \mathbf{x}_{t_1}, \mathbf{x}_{t_1})}(\mathbf{X}_i, \mathbf{X}_j, \mathbf{Y}_r, \mathbf{Y}_s) - (\mathbf{U}_i, \mathbf{U}_j, \mathbf{V}_r, \mathbf{V}_s))^{(\alpha)} , \tag{4.17}$$

$$\psi_{k,1,Total}(\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_{t_1}, \mathbf{Y}_r, \mathbf{Y}_s) = \psi_{k,1}(\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_{t_1}, \mathbf{Y}_r, \mathbf{Y}_s) + g_k(\mathbf{X}_i, \mathbf{X}_j, \mathbf{Y}_r, \mathbf{Y}_s; \mathbf{F}) , \tag{4.18}$$

$$\tilde{\psi}_{k,1,Total}(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{Y}_1, \mathbf{Y}_2) = \frac{1}{6} \sum_{(i_1,i_2,i_3) \in I_3^3} \sum_{(j_1,j_2) \in I_2^2} \psi_{k,1,Total}(\mathbf{X}_{i_1}, \mathbf{X}_{i_2}, \mathbf{X}_{i_3}, \mathbf{Y}_{j_1}, \mathbf{Y}_{j_2}) , \tag{4.19}$$

$$\tilde{\psi}^{(1,0)}_{k,1,Total}(\mathbf{x}) = E\left[\tilde{\psi}_{k,1,Total}(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{Y}_1, \mathbf{Y}_2) | \mathbf{X}_1 = \mathbf{x}\right] , \tag{4.20}$$

and

$$\tilde{\psi}^{(0,1)}_{k,1,Total}(\mathbf{y}) = E\left[\tilde{\psi}_{k,1,Total}(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{Y}_1, \mathbf{Y}_2) | \mathbf{Y}_1 = \mathbf{y}\right] . \tag{4.21}$$

Theorem 4.3.1. *Suppose that all second-order partial derivatives of k exist on $(0,1)^{2p}$, all bounded by M_2, and that $m, n \to \infty$ in such a way that $\frac{n}{m+n} \to \lambda$, $\lambda \in (0,1)$. Then, if $\mathbf{X} \overset{D}{=} \mathbf{Y}$,*

$$(m+n)^{-1/2} \left[\hat{\xi}_k(\mathbf{X}, \mathbf{Y}; \hat{\mathbf{F}}) - \xi_k(\mathbf{X}, \mathbf{Y}; \mathbf{F})\right] \xrightarrow{D} N(0, \sigma_k^2) \tag{4.22}$$

where

$$\sigma_k^2 = 9\text{var}\left[\tilde{\psi}^{(0,1)}_{k,1,Total}(\mathbf{X})\right] + 12\text{cov}\left[\tilde{\psi}^{(0,1)}_{k,1,Total}(\mathbf{X}), \tilde{\psi}^{(1,0)}_{k,1,Total}(\mathbf{X})\right] + \frac{\lambda^3 + (1-\lambda)^3}{\lambda(1-\lambda)} \text{var}\left[\tilde{\psi}^{(1,0)}_{k,1,Total}(\mathbf{X})\right]$$

The second theorem is about the CeD test statistic, in which the Euclidean distance d_p is used. Let C_{ijrs} be defined as

$$C_{ijrs} = \Big\{\min\Big\{|\mathbf{U}_i - \mathbf{U}_j|_p, |\mathbf{V}_r - \mathbf{V}_s|_p, |\mathbf{U}_i - \mathbf{V}_r|_p, \tag{4.23}$$

$$|\mathbf{U}_i - \mathbf{V}_s|_p, |\mathbf{U}_j - \mathbf{V}_r|_p, |\mathbf{U}_j - \mathbf{V}_s|_p\Big\} > 0\Big\}, \tag{4.24}$$

Define

$$\begin{aligned}
&\psi_{d_p,1}(\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_{t_1}, \mathbf{Y}_r, \mathbf{Y}_s) \\
&= \sum_{\alpha \in \mathbb{N}_{4p}} D^{(\alpha)} g_{d_p}(\mathbf{U}_i, \mathbf{U}_j, \mathbf{V}_r, \mathbf{V}_s) \qquad (4.25)\\
&\qquad \cdot (\vec{I}_{(\mathbf{X}_{t_1},\mathbf{X}_{t_1},\mathbf{X}_{t_1},\mathbf{X}_{t_1})}(\mathbf{X}_i, \mathbf{X}_j, \mathbf{Y}_r, \mathbf{Y}_s) - (\mathbf{U}_i, \mathbf{U}_j, \mathbf{V}_r, \mathbf{V}_s))^{(\alpha)} I_{C_{ijrs}} \ , \qquad (4.26)
\end{aligned}$$

$$\psi_{d_p,1,Total}(\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_{t_1}, \mathbf{Y}_r, \mathbf{Y}_s) = \psi_{d_p,1}(\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_{t_1}, \mathbf{Y}_r, \mathbf{Y}_s) + g_{d_p}(\mathbf{X}_i, \mathbf{X}_j, \mathbf{Y}_r, \mathbf{Y}_s; \mathbf{F}) \ , \qquad (4.27)$$

$$\tilde{\psi}_{d_p,1,Total}(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{Y}_1, \mathbf{Y}_2) = \frac{1}{6} \sum_{(i_1,i_2,i_3)\in I_3^3} \sum_{(j_1,j_2)\in I_2^2} \psi_{d_p,1,Total}(\mathbf{X}_{i_1}, \mathbf{X}_{i_2}, \mathbf{X}_{i_3}, \mathbf{Y}_{j_1}, \mathbf{Y}_{j_2}) \ , \qquad (4.28)$$

$$\tilde{\psi}^{(1,0)}_{d_p,1,Total}(\mathbf{x}) = E\left[\tilde{\psi}_{d_p,1,Total}(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{Y}_1, \mathbf{Y}_2) | \mathbf{X}_1 = \mathbf{x}\right] \ , \qquad (4.29)$$

and

$$\tilde{\psi}^{(0,1)}_{d_p,1,Total}(\mathbf{y}) = E\left[\tilde{\psi}_{d_p,1,Total}(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{Y}_1, \mathbf{Y}_2) | \mathbf{Y}_1 = \mathbf{y}\right] \ . \qquad (4.30)$$

Theorem 4.3.2. *Suppose that $m, n \to \infty$ in such a way that $\frac{n}{m+n} \to \lambda$, $\lambda \in (0,1)$. Then, if $\mathbf{X} \stackrel{D}{=} \mathbf{Y}$,*

$$(m+n)^{-1/2}\left[\hat{\xi}_{d_p}(\mathbf{X}, \mathbf{Y}; \hat{\mathbf{F}}) - \xi_{d_p}(\mathbf{X}, \mathbf{Y}; \mathbf{F})\right] \stackrel{D}{\to} N(0, \sigma^2_{d_p s})$$

where

$$\begin{aligned}
\sigma^2_{d_p} &= 9\text{var}\left[\tilde{\psi}^{(0,1)}_{d_p,1,Total}(\mathbf{X})\right] + 12\text{cov}\left[\tilde{\psi}^{(0,1)}_{d_p,1,Total}(\mathbf{X}), \tilde{\psi}^{(1,0)}_{d_p,1,Total}(\mathbf{X})\right] \\
&\quad + \frac{\lambda^3 + (1-\lambda)^3}{\lambda(1-\lambda)} \text{var}\left[\tilde{\psi}^{(1,0)}_{d_p,1,Total}(\mathbf{X})\right]
\end{aligned}$$

4.4 Simulation Studies

In this section, the performance of the four methods CMMD, Ced, MMD and eD are evaluated and compared.

4.4.1 Simulation 4

In Simulation 4, the significance levels of the four methods are studied. The setting is as follows:

- $p = 5$. Let $\mathbf{X} = \mathbf{\Sigma}^{1/2}\mathbf{W}_1$ and $\mathbf{Y} = \mathbf{\Sigma}^{1/2}\mathbf{W}_2$. $\mathbf{W}_1$ and $\mathbf{W}_2$ are 5-dimensional independent random vectors whose components are i.i.d. with one of the following 4 distributions:

  standard normal, standard exponential, standard lognormal and t_2.

  $\mathbf{\Sigma}$ is a 5×5 matrix with diagonal elements being 1 and off-diagonal elements being 0.1.

- The sample size m for $\mathbf{X}$ equals one of the following:

  25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100.

  The sample size for $\mathbf{Y}$ is $n = 2m$.

- For each combination of the above parameters, the simulation is repeated 1000 times.

- The Gaussian kernels are used. The permutation approach is used to compute the p-values. The number of permutation is $N = 1000$.

- The null hypothesis is rejected when the computed p-value is less than 0.05.

The results are summarized in Figure 4.1. Given the distributions of $\mathbf{X}$ and $\mathbf{Y}$, and the sample size, the value of each dot in the figure is the proportion of simulations in which the null hypothesis that $\mathbf{X}$ and $\mathbf{Y}$ have the same distribution is rejected, which estimate the significant level of the test. As expected, all these values are close to 0.05 .

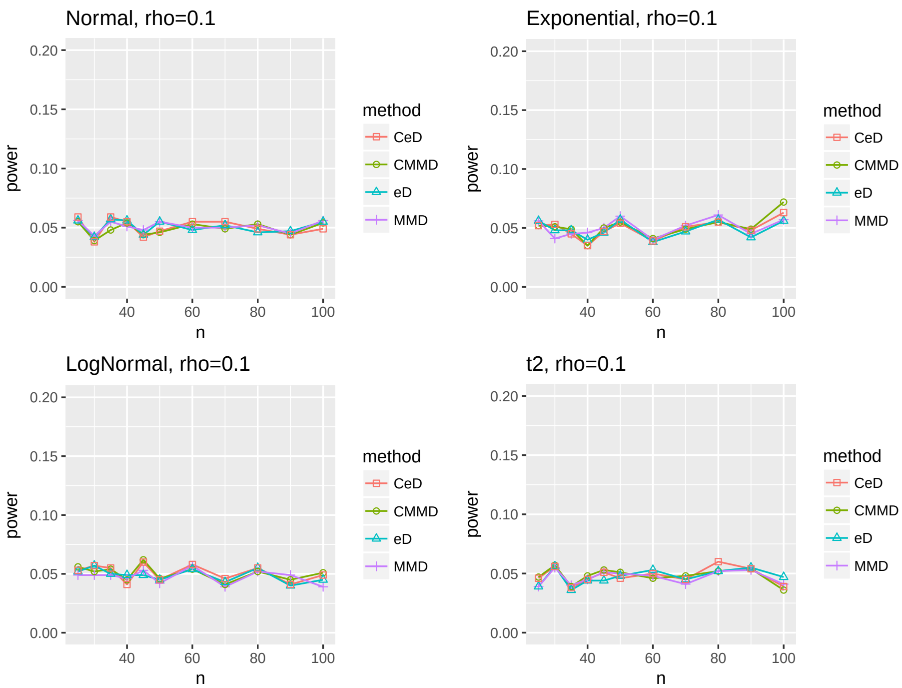

Figure 4.1. Summarized results in Simulation 4

4.4.2 Simulation 5

In Simulation 5, the powers of the four methods are studied. The setting is as follows:

- $p = 5$. Let $\mathbf{X} = \mathbf{\Sigma}_1^{1/2}\mathbf{W}_1$ and $\mathbf{Y} = \mathbf{\Sigma}_2^{1/2}\mathbf{W}_2$. $\mathbf{W}_1$ and $\mathbf{W}_2$ are independent 5-dimensional random vectors whose components are i.i.d. with one of the following 4 distributions:

  standard normal, standard exponential, standard lognormal and t_2.

  $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are both 5×5 matrices with diagonal elements being 1 and off-diagonal elements being 0.1 (for $\mathbf{\Sigma}_1$) and ρ (for $\mathbf{\Sigma}_2$), respectively, where ρ equals one of the following:

  0.5, 0.9 .

- The sample size m for $\mathbf{X}$ equals one of the following:

  25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100.

  The sample size for $\mathbf{Y}$ is $n = 2m$.

- For each combination of the above parameters, the simulation is repeated 1000 times.

- The Gaussian kernels are used. The permutation approach is used to compute the p-values. The number of permutation is $N = 1000$.

- The null hypothesis is rejected when the computed p-value is less than 0.05.

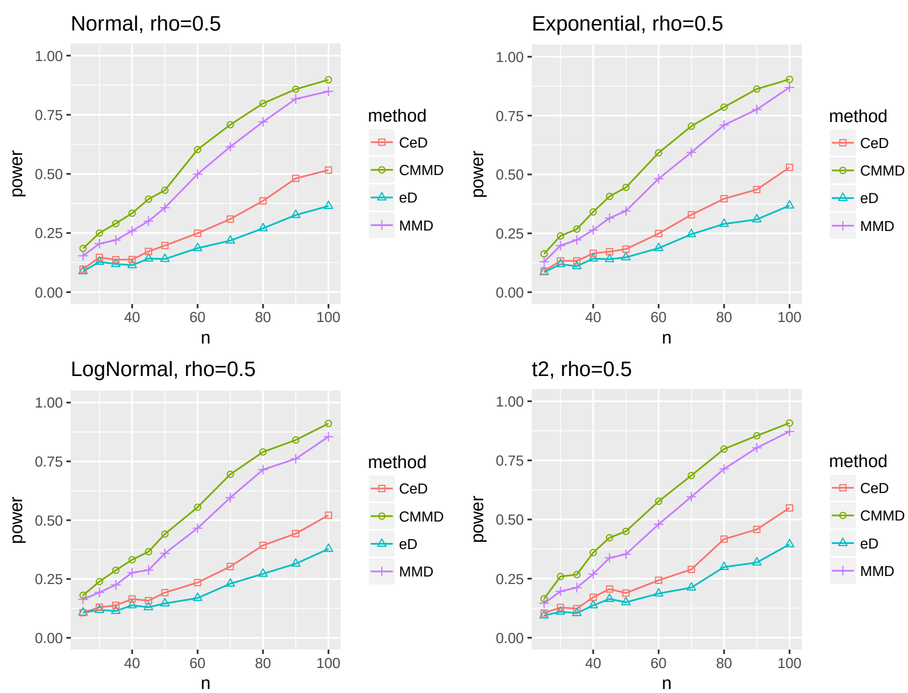

Figure 4.2. Summarized results in Simulation 5 (Part 1)

The results are summarized in Figure 4.2 and Figure 4.3. Given the distributions of $\mathbf{X}$ and $\mathbf{Y}$, and the sample size, the value of each dot in the figure is the proportion of simulations in which the null hypothesis that $\mathbf{X}$ and $\mathbf{Y}$ have the same distribution is rejected, which estimates the power of the test. As expected, the powers increase as n increases. In all the cases, the powers of CMMD and MMD are much higher than CeD and eD. Among CMMD and MMD, the power of the former is higher than the latter. Among CeD and eD, the power of the former is higher than the latter.

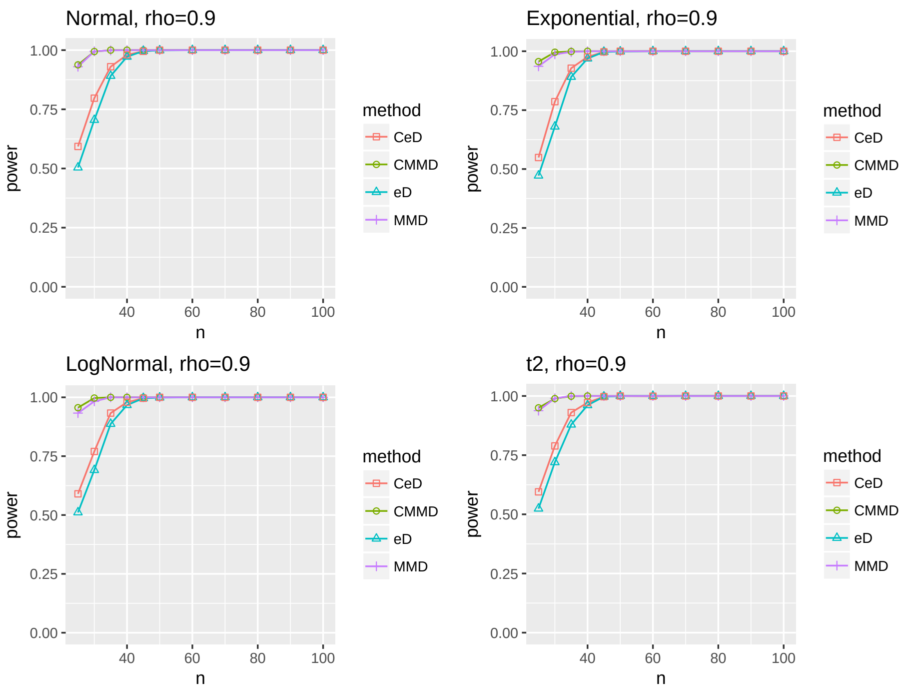

Figure 4.3. Summarized results in Simulation 5 (Part 2)

4.A Appendix: Proof of Theorems

4.A.1 Proof of Theorem 4.3.1

For simplicity, let $\hat{\mathbf{U}}_i = \hat{\mathbf{F}}(\mathbf{X}_i)$ for $i = 1, \dots, m$, and let $\hat{\mathbf{V}}_j = \hat{\mathbf{F}}(\mathbf{Y}_j)$ for $j = 1, \dots, n$. Notice that for any $\mathbf{x} \in \mathbb{R}^p$,

$$\hat{\mathbf{F}}(\mathbf{x}) = \frac{m}{m+n}\hat{\mathbf{F}}_{\mathbf{X}}(\mathbf{x}) + \frac{n}{m+n}\hat{\mathbf{F}}_{\mathbf{Y}}(\mathbf{x}) = \frac{m}{m+n}\frac{1}{m}\sum_{i=1}^{m}\vec{I}_{\mathbf{X}_i}(\mathbf{x}) + \frac{n}{m+n}\frac{1}{n}\sum_{i=1}^{n}\vec{I}_{\mathbf{Y}_i}(\mathbf{x}) \ .$$

Therefore, by similar derivation as in (2.14),

$$\begin{aligned} & g_k(\mathbf{X}_i, \mathbf{X}_j, \mathbf{Y}_r, \mathbf{Y}_s; \hat{\mathbf{F}}) - g_k(\mathbf{X}_i, \mathbf{X}_j, \mathbf{Y}_r, \mathbf{Y}_s; \mathbf{F}) \\ = & \sum_{\alpha \in \mathbb{N}_{4p}} D^{(\alpha)} g_k(\mathbf{U}_i, \mathbf{U}_j, \mathbf{V}_r, \mathbf{V}_s)((\hat{\mathbf{U}}_i, \hat{\mathbf{U}}_j, \hat{\mathbf{V}}_r, \hat{\mathbf{V}}_s) - (\mathbf{U}_i, \mathbf{U}_j, \mathbf{V}_r, \mathbf{V}_s))^{(\alpha)} \\ & + \frac{1}{2}\sum_{\boldsymbol{\alpha} \in \mathbb{N}_{4p}^2} D^{(\boldsymbol{\alpha})} g_k(\tilde{\mathbf{U}}_i, \tilde{\mathbf{U}}_j, \tilde{\mathbf{V}}_r, \tilde{\mathbf{V}}_s)((\hat{\mathbf{U}}_i, \hat{\mathbf{U}}_j, \hat{\mathbf{V}}_r, \hat{\mathbf{V}}_s) - (\mathbf{U}_i, \mathbf{U}_j, \mathbf{V}_r, \mathbf{V}_s))^{(\boldsymbol{\alpha})} \end{aligned}$$

$$
\begin{aligned}
&= \frac{m}{m+n}\frac{1}{m}\sum_{t_1=1}^{m}\sum_{\alpha\in\mathbb{N}_{4p}} D^{(\alpha)}g_k(\mathbf{U}_i,\mathbf{U}_j,\mathbf{V}_r,\mathbf{V}_s) \\
&\qquad\qquad \cdot (\vec{I}_{(\mathbf{X}_{t_1},\mathbf{X}_{t_1},\mathbf{X}_{t_1},\mathbf{X}_{t_1})}(\mathbf{X}_i,\mathbf{X}_j,\mathbf{Y}_r,\mathbf{Y}_s) - (\mathbf{U}_i,\mathbf{U}_j,\mathbf{V}_r,\mathbf{V}_s))^{(\alpha)} \\
&\quad + \frac{n}{m+n}\frac{1}{n}\sum_{t_2=1}^{m}\sum_{\alpha\in\mathbb{N}_{4p}} D^{(\alpha)}g_k(\mathbf{U}_i,\mathbf{U}_j,\mathbf{V}_r,\mathbf{V}_s) \\
&\qquad\qquad \cdot (\vec{I}_{(\mathbf{Y}_{t_2},\mathbf{Y}_{t_2},\mathbf{Y}_{t_2},\mathbf{Y}_{t_2})}(\mathbf{X}_i,\mathbf{X}_j,\mathbf{Y}_r,\mathbf{Y}_s) - (\mathbf{U}_i,\mathbf{U}_j,\mathbf{V}_r,\mathbf{V}_s))^{(\alpha)} \\
&\quad + \frac{1}{2}\sum_{\boldsymbol{\alpha}\in\mathbb{N}_{4p}^2} D^{(\boldsymbol{\alpha})}g_k(\tilde{\mathbf{U}}_i,\tilde{\mathbf{U}}_j,\tilde{\mathbf{V}}_r,\tilde{\mathbf{V}}_s)((\hat{\mathbf{U}}_i,\hat{\mathbf{U}}_j,\hat{\mathbf{V}}_r,\hat{\mathbf{V}}_s) - (\mathbf{U}_i,\mathbf{U}_j,\mathbf{V}_r,\mathbf{V}_s))^{(\boldsymbol{\alpha})} \\
&= \frac{m}{m+n}\frac{1}{m}\sum_{t_1=1}^{m}\psi_{k,1}(\mathbf{U}_i,\mathbf{U}_j,\mathbf{U}_{t_1},\mathbf{V}_r,\mathbf{V}_s) + \frac{n}{m+n}\frac{1}{n}\sum_{t_2=1}^{m}\psi_{k,2}(\mathbf{U}_i,\mathbf{U}_j,\mathbf{V}_r,\mathbf{V}_s,\mathbf{U}_{t_2}) \\
&\quad + r_1(\mathbf{U}_i,\mathbf{U}_j,\mathbf{V}_r,\mathbf{V}_s;\hat{\mathbf{F}}) \ ,
\end{aligned}
$$

where $\psi_{k,1}$ is defined in (4.17), and

$$
\begin{aligned}
&\psi_{k,2}(\mathbf{U}_i,\mathbf{U}_j,\mathbf{V}_r,\mathbf{V}_s,\mathbf{V}_{t_2}) \\
&= \sum_{\alpha\in\mathbb{N}_{4p}} D^{(\alpha)}g_k(\mathbf{U}_i,\mathbf{U}_j,\mathbf{V}_r,\mathbf{V}_s)(\vec{I}_{(\mathbf{Y}_{t_2},\mathbf{Y}_{t_2},\mathbf{Y}_{t_2},\mathbf{Y}_{t_2})}(\mathbf{X}_i,\mathbf{X}_j,\mathbf{Y}_r,\mathbf{Y}_s) - (\mathbf{U}_i,\mathbf{U}_j,\mathbf{V}_r,\mathbf{V}_s))^{(\alpha)} \ ,
\end{aligned}
\tag{4.31}
$$

$$
\begin{aligned}
&r_1(\mathbf{U}_i,\mathbf{U}_j,\mathbf{V}_r,\mathbf{V}_s;\hat{\mathbf{F}}) \\
&= \frac{1}{2}\sum_{\boldsymbol{\alpha}\in\mathbb{N}_{4p}^2} D^{(\boldsymbol{\alpha})}g_k(\tilde{\mathbf{U}}_i,\tilde{\mathbf{U}}_j,\tilde{\mathbf{V}}_r,\tilde{\mathbf{V}}_s)((\hat{\mathbf{U}}_i,\hat{\mathbf{U}}_j,\hat{\mathbf{V}}_r,\hat{\mathbf{V}}_s) - (\mathbf{U}_i,\mathbf{U}_j,\mathbf{V}_r,\mathbf{V}_s))^{(\boldsymbol{\alpha})} \ .
\end{aligned}
\tag{4.32}
$$

Because for any $|\boldsymbol{\alpha}| = 2$, $\|D^{(\boldsymbol{\alpha})}h_{k,l}(\cdot)\|_\infty \leq C$, for some finite constant C that depends on the bounds of k as well as the bounds of their first and second order derivatives, we have

$$
\begin{aligned}
&\left|\frac{1}{m^2n^2}\sum_{i,j=1}^{m}\sum_{r,s=1}^{n} r_1(\mathbf{U}_i,\mathbf{U}_j,\mathbf{V}_r,\mathbf{V}_s;\hat{\mathbf{F}})\right| \\
&\leq \frac{1}{m^2n^2}\sum_{i,j=1}^{m}\sum_{r,s=1}^{n}\frac{1}{2}\sum_{\boldsymbol{\alpha}\in\mathbb{N}_{4p}^2}|D^{(\boldsymbol{\alpha})}g_k(\tilde{\mathbf{U}}_i,\tilde{\mathbf{U}}_j,\tilde{\mathbf{V}}_r,\tilde{\mathbf{V}}_s)| \cdot \|\hat{\mathbf{F}} - \mathbf{F}\|_\infty^2 \\
&\leq \frac{1}{2}(4p)^2 C\|\hat{\mathbf{F}} - \mathbf{F}\|_\infty^2 \ .
\end{aligned}
$$

Hence

$$
\begin{aligned}
&\hat{\xi}_k(\mathbf{X},\mathbf{Y};\hat{\mathbf{F}}) - \hat{\xi}_k(\mathbf{X},\mathbf{Y};\mathbf{F}) \\
&= \frac{m}{m+n}\frac{1}{m^3n^2}\sum_{i,j,t_1=1}^{m}\sum_{r,s=1}^{n}\psi_{k,1}(\mathbf{X}_i,\mathbf{X}_j,\mathbf{X}_{t_1},\mathbf{Y}_r,\mathbf{Y}_s) \\
&\quad + \frac{n}{m+n}\frac{1}{m^2n^3}\sum_{i,j=1}^{m}\sum_{r,s,t_2=1}^{n}\psi_{k,2}(\mathbf{X}_i,\mathbf{X}_j,\mathbf{Y}_r,\mathbf{Y}_s,\mathbf{Y}_{t_2}) + O_P((m+n)^{-1}) \ .
\end{aligned}
$$

On the other hand, by analyzing the conditionals expectation of each term given $\mathbf{X}_i,\mathbf{X}_j,\mathbf{Y}_r,\mathbf{Y}_s$, it can be shown that

$$
E\left[\hat{\xi}_k(\mathbf{X},\mathbf{Y};\hat{\mathbf{F}})\right] = E\left[\hat{\xi}_k(\mathbf{X},\mathbf{Y};\mathbf{F})\right] + O_P((m+n)^{-1}) = \xi_k(X,Y;\mathbf{F}) + O_P((m+n)^{-1}) \tag{4.33}
$$

Hence

$$
\begin{aligned}
&\hat{\xi}_k(\mathbf{X},\mathbf{Y};\hat{\mathbf{F}}) \\
&= \frac{m}{m+n}\frac{1}{m^3n^2}\sum_{i,j,t_1=1}^{m}\sum_{r,s=1}^{n}[\psi_{k,1}(\mathbf{X}_i,\mathbf{X}_j,\mathbf{X}_{t_1},\mathbf{Y}_r,\mathbf{Y}_s) + g_k(\mathbf{X}_i,\mathbf{X}_j,\mathbf{Y}_r,\mathbf{Y}_s;\mathbf{F})] \\
&\quad + \frac{n}{m+n}\frac{1}{m^2n^3}\sum_{i,j=1}^{m}\sum_{r,s,t_2=1}^{n}[\psi_{k,2}(\mathbf{X}_i,\mathbf{X}_j,\mathbf{Y}_r,\mathbf{Y}_s,\mathbf{Y}_{t_2}) + g_k(\mathbf{X}_i,\mathbf{X}_j,\mathbf{Y}_r,\mathbf{Y}_s;ss\mathbf{F})] \\
&\quad + O_P((m+n)^{-1}) \\
&= \frac{m}{m+n}\frac{1}{m^3n^2}\sum_{i,j,t_1=1}^{m}\sum_{r,s=1}^{n}\psi_{k,1,total}(\mathbf{X}_i,\mathbf{X}_j,\mathbf{X}_{t_1},\mathbf{Y}_r,\mathbf{Y}_s) \\
&\quad + \frac{n}{m+n}\frac{1}{m^2n^3}\sum_{i,j=1}^{m}\sum_{r,s,t_2=1}^{n}\psi_{k,2,Total}(\mathbf{X}_i,\mathbf{X}_j,\mathbf{Y}_r,\mathbf{Y}_s,\mathbf{Y}_{t_2}) \\
&\quad + O_P((m+n)^{-1}) \\
&= \frac{m}{m+n}V_1 + \frac{n}{m+n}V_2 + O_P((m+n)^{-1}) \ ,
\end{aligned} \tag{4.34}
$$

where $\psi_{k,1,total}$ is defined in (4.19),

$$
\psi_{k,2,Total}(\mathbf{X}_i,\mathbf{X}_j,\mathbf{X}_{t_1},\mathbf{Y}_r,\mathbf{Y}_s) = \psi_{k,2}(\mathbf{X}_i,\mathbf{X}_j,\mathbf{X}_{t_1},\mathbf{Y}_r,\mathbf{Y}_s) + g_k(\mathbf{X}_i,\mathbf{X}_j,\mathbf{Y}_r,\mathbf{Y}_s;\mathbf{F}) \ ,
$$

and

$$V_1 = \frac{1}{m^3n^2} \sum_{i,j,t_1=1}^{m} \sum_{r,s=1}^{n} \psi_{k,1,total}(\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_{t_1}, \mathbf{Y}_r, \mathbf{Y}_s)$$

$$V_2 = \frac{1}{m^2n^3} \sum_{i,j=1}^{m} \sum_{r,s,t_2=1}^{n} \psi_{k,2,Total}(\mathbf{X}_i, \mathbf{X}_j, \mathbf{Y}_r, \mathbf{Y}_s, \mathbf{Y}_{t_2}) \ .$$

By Theorem 2.5.2 in [15], the projection of $(m+n)^{-1/2}(V_1 - E(V_1))$ is $(m+n)^{-1/2}(E(V_{1,p} - E(V_1))$, where $V_{1,p}$ can be written ass

$$V_{1,p} = \sum_{i=1}^{m} p_{11}(\mathbf{X}_i) + \sum_{j=1}^{n} p_{12}(\mathbf{Y}_j) \ . \tag{4.35}$$

The function p_{11} in (4.35) is defined by

$$\begin{aligned} p_{11}(\mathbf{x}) &= E[V_1|\mathbf{X}_1 = \mathbf{x}] \\ &= E\left[\frac{1}{m^3n^2} \sum_{i,j,t_1=1}^{m} \sum_{r,s=1}^{n} \psi_{k,1,total}(\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_{t_1}, \mathbf{Y}_r, \mathbf{Y}_s)\middle|\mathbf{X}_1 = \mathbf{x}\right] \\ &= \frac{1}{m^3} \sum_{i,j,t_1=1}^{m} E\left[\tilde{\psi}_{k,1,total}(\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_{t_1}, \mathbf{Y}_1, \mathbf{Y}_2)\middle|\mathbf{X}_1 = \mathbf{x}\right] \\ &= \frac{1}{m^3}\left[3(m-1)_2\tilde{\psi}^{(1,0)}_{k,1,Total}(\mathbf{x}) + 3(m-2)\tilde{\psi}^{(2,0)}_{k,1,Total}(\mathbf{x},\mathbf{x}) + \tilde{\psi}^{(3,0)}_{k,1,Total}(\mathbf{x},\mathbf{x},\mathbf{x})\right] \ , \end{aligned}$$

where $\psi^{(1,0)}_{k,1,Total}$ is defined in (4.20), and

$$\psi^{(2,0)}_{k,1,Total}(\mathbf{x}_1, \mathbf{x}_2) = E\left[\tilde{\psi}_{k,1,Total}(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{Y}_1, \mathbf{Y}_2)|\mathbf{X}_1 = \mathbf{x}_1, \mathbf{X}_2 = \mathbf{x}_2\right]$$

$$\psi^{(3,0)}_{k,1,Total}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = E\left[\tilde{\psi}_{k,1,Total}(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{Y}_1, \mathbf{Y}_2)|\mathbf{X}_1 = \mathbf{x}_1, \mathbf{X}_2 = \mathbf{x}_2, \mathbf{X}_3 = \mathbf{x}_3\right]$$

and similarly

$$p_{12}(\mathbf{y}) = E[V|\mathbf{Y}_1 = \mathbf{y}] = \frac{1}{n^2}\left[2(n-1)\tilde{\psi}^{(0,1)}_{k,1,Total}(\mathbf{y}) + \tilde{\psi}^{(0,2)}_{k,1,Total}(\mathbf{y},\mathbf{y})\right] \ .$$

where $\psi^{(0,1)}_{k,1,Total}$ is defined in (4.21), and

$$\psi^{(0,2)}_{k,1,Total}(\mathbf{y}_1, \mathbf{y}_2) = E\left[\tilde{\psi}_{k,1,Total}(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{Y}_1, \mathbf{Y}_2)|\mathbf{Y}_1 = \mathbf{y}_1, \mathbf{Y}_2 = \mathbf{y}_2\right] \ .$$

By a generalization of Theorem 3 on Page 40 in [19] to the case of V-statistics, it can be shown that

$$\mathrm{var}(V_1) - \mathrm{var}(V_{1,p}) = O((m+n)^{-2}) \ .$$

Again by Theorem 2.5.2 in [15],

$$\mathrm{var}(V_1 - V_{1,p}) = \mathrm{var}(V_1) - \mathrm{var}(V_{1,p}) = O((m+n)^{-2}) \ .$$

Therefore,

$$\begin{aligned}
V_1 &= V_{1,p} + O_P((m+n)^{-1}) \\
&= \sum_{i=1}^{m} p_{11}(\mathbf{X}_i) + \sum_{j=1}^{n} p_{12}(\mathbf{Y}_j) + O_P((m+n)^{-1}) \\
&= \frac{3(m-1)_2}{m^3} \sum_{i=1}^{m} \tilde{\psi}^{(1,0)}_{k,1,Total}(\mathbf{X}_i) + \frac{2(n-1)}{n^2} \sum_{j=1}^{n} \tilde{\psi}^{(0,1)}_{k,1,Total}(\mathbf{Y}_j) + O_P((m+n)^{-1}) \ .
\end{aligned}$$

By the same arguments, we have

$$V_2 = \frac{2(m-1)}{m^2} \sum_{i=1}^{m} \tilde{\psi}^{(1,0)}_{k,2,Total}(\mathbf{X}_i) + \frac{3(n-1)_2}{n^3} \sum_{j=1}^{n} \tilde{\psi}^{(0,1)}_{k,2,Total}(\mathbf{Y}_j) + O_P((m+n)^{-1}) \ ,$$

where $\psi^{(1,0)}_{k,2,Total}$ and $\psi^{(0,1)}_{k,2,Total}$ is defined in similar forms as in (4.20) and (4.21). Thus from (4.34), noticing that $\frac{n}{m+n} = \lambda + o(1)$, $V_1 = O((m+n)^{-1/2})$ and $V_2 = O((m+n)^{-1/2})$, we have

$$\begin{aligned}
&\hat{\xi}_k(\mathbf{X}, \mathbf{Y}; \hat{\mathbf{F}}) \\
= &\frac{m}{m+n} V_1 + \frac{n}{m+n} V_2 + O_P((m+n)^{-1}) \\
= &(1-\lambda) V_1 + \lambda V_2 + o_P((m+n)^{-1/2}) \\
= &(1-\lambda) \left[\frac{3}{m} \sum_{i=1}^{m} \tilde{\psi}^{(1,0)}_{k,1,Total}(\mathbf{X}_i) + \frac{2}{n} \sum_{j=1}^{n} \tilde{\psi}^{(0,1)}_{k,1,Total}(\mathbf{Y}_j) \right] \\
&\quad + \lambda \left[\frac{2}{m} \sum_{i=1}^{m} \tilde{\psi}^{(1,0)}_{k,2,Total}(\mathbf{X}_i) + \frac{3}{n} \sum_{j=1}^{n} \tilde{\psi}^{(0,1)}_{k,2,Total}(\mathbf{Y}_j) \right] + o_P((m+n)^{-1/2}) \\
= &\frac{1}{m} \sum_{i=1}^{m} \left[3(1-\lambda) \tilde{\psi}^{(1,0)}_{k,1,Total}(\mathbf{X}_i) + 2\lambda \tilde{\psi}^{(1,0)}_{k,2,Total}(\mathbf{X}_i) \right]
\end{aligned}$$

$$+\frac{1}{n}\sum_{j=1}^{n}\left[2(1-\lambda)\tilde{\psi}^{(0,1)}_{k,1,Total}(\mathbf{Y}_j)+3\lambda\tilde{\psi}^{(0,1)}_{k,2,Total}(\mathbf{Y}_j)\right]+o_P((m+n)^{-1/2}) \quad (4.36)$$

By central limit theorem,

$$\sqrt{m}\cdot\frac{1}{m}\sum_{i=1}^{m}\left[3(1-\lambda)\tilde{\psi}^{(1,0)}_{k,1,Total}(\mathbf{X}_i)+2\lambda\tilde{\psi}^{(1,0)}_{k,2,Total}(\mathbf{X}_i)\right]\xrightarrow{D}N(0,\sigma_1^2)$$

$$\sqrt{n}\cdot\frac{1}{n}\sum_{j=1}^{n}\left[2(1-\lambda)\tilde{\psi}^{(0,1)}_{k,1,Total}(\mathbf{Y}_j)+3\lambda\tilde{\psi}^{(0,1)}_{k,2,Total}(\mathbf{Y}_j)\right]\xrightarrow{D}N(0,\sigma_2^2)$$

where

$$\begin{aligned}\sigma_1^2&=\text{var}\left[3(1-\lambda)\tilde{\psi}^{(1,0)}_{k,1,Total}(\mathbf{X})+2\lambda\tilde{\psi}^{(1,0)}_{k,2,Total}(\mathbf{X})\right]\\&=\text{var}\left[3(1-\lambda)\tilde{\psi}^{(1,0)}_{k,1,Total}(\mathbf{X})+2\lambda\tilde{\psi}^{(0,1)}_{k,1,Total}(\mathbf{X})\right]\end{aligned}$$

$$\begin{aligned}\sigma_2^2&=\text{var}\left[2(1-\lambda)\tilde{\psi}^{(0,1)}_{k,1,Total}(\mathbf{Y})+3\lambda\tilde{\psi}^{(0,1)}_{k,2,Total}(\mathbf{Y})\right]\\&=\text{var}\left[2(1-\lambda)\tilde{\psi}^{(0,1)}_{k,1,Total}(\mathbf{X})+3\lambda\tilde{\psi}^{(1,0)}_{k,1,Total}(\mathbf{X})\right]\end{aligned}$$

The above two equalities follow the symmetry of g_k and the fact that $\mathbf{X}\overset{D}{=}\mathbf{Y}$. Thus by the independence of $\mathbf{X}_1,\ldots,\mathbf{X}_m,\mathbf{Y}_1,\ldots,\mathbf{Y}_n$, (4.33) and (4.36), (4.22) follows with

$$\begin{aligned}\sigma_k^2&=\frac{1}{1-\lambda}\sigma_1^2+\frac{1}{\lambda}\sigma_2^2\\&=9\text{var}\left[\tilde{\psi}^{(0,1)}_{k,1,Total}(\mathbf{X})\right]+12\text{cov}\left[\tilde{\psi}^{(0,1)}_{k,1,Total}(\mathbf{X}),\tilde{\psi}^{(1,0)}_{k,1,Total}(\mathbf{X})\right]\\&\quad+\frac{\lambda^3+(1-\lambda)^3}{\lambda(1-\lambda)}\text{var}\left[\tilde{\psi}^{(1,0)}_{k,1,Total}(\mathbf{X})\right]\end{aligned}$$

4.A.2 Proof of Theorem 4.3.2

The proof of Theorem 4.3.2 proceeds by similar arguments as of Theorem 4.3.1, with similar techniques used in the proof of Theorem 2.3.1 to avoid distances that are too close to 0, which is omitted.

Chapter 5 Conclusions and Future Work

5.1 Conclusions

This dissertation focuses on copula version of RKHS-based and distance-based criteria for testing of independence or two-sample problem.

In Chapter 2, the copula version of distance covariance criterion (CdCov) is introduced with the corresponding test statistics proposed. CdCov is obtained by applying marginal mid-cdf on each variable to obtain a copula distribution, and the corresponding test statistics can be calculated using mid-ranks for each variable. It is shown that without any assumptions, the test statistic is asymptotically normal under the alternative hypothesis, and it is asymptotically distributed as mixed chi-squares distribution under the null hypothesis. To compute the p-value, the permutation test and a conservative test are introduced. The copula version of a linear-time statistic which costs less in time is also introduced.

In Chapter 3, the copula version of HSIC (CHSIC) is introduced with the corresponding test statistics proposed. CHSIC is obtained by applying marginal cdf on each variable to obtain a copula distribution, and the corresponding test statistics can be calculated using mid-ranks for each variable. It is shown that with proper kernels and no assumptions on the true distribution, the test statistic is asymptotically normal under the alternative hypothesis, and it is asymptotically distributed as mixed chi-squares distribution under the null hypothesis. To compute the p-value, the permutation test and a conservative test are introduced. The copula version of a linear-time statistic which costs less in time is also introduced.

In Chapter 4, the copula version of e-distance and MMD are introduced with the corresponding test statistics proposed. The criteria are obtained by applying a pooled marginal cdf on each variable to obtain a copula distribution, and the corresponding test statistics can be calculated using pooled mid-ranks for each variable. It is shown that with a proper kernel for CMMD and no assumptions for CeD, the two test statistics are asymptotically normal under the null hypothesis. To compute the p-value, the permutation tests are introduced.

5.2 Future Work

There are several possible future extensions of the work in this dissertation.

Since CdCov and CHSIC can serve as independence criteria of two random vectors, they can be applied to statistical procedures that utilize such criteria, including but not limit to variable selection, feature screening and discriminant analysis. The asymptotic properties of such statistical procedure with CdCov or CHSIC can be studied, and their performance may be superior to the original dCov and HSIC due to the fact that the original random vectors are all converted to copula distribution.

A universal kernel is used in CHSIC and CMMD, which in practice can be chosen from several candidate classes of functions. Notice that the distribution of the random vectors become copula distributions after the marginal mid-cdf transformations, and that when calculating the test statistic only the values of the kernel at finite many points are used given the sample size. Thus it may be possible that a criterion on the choice of the kernel can be formulated, and the best choice of the kernel can be chosen accordingly.

There is obvious similarity in the analysis of the asymptotic properties for the test statistics introduced in this dissertation. These test statistics are essentially obtained by a two-step process. The first step is replacing the expectation by empirical average, and the second step is replacing the marginal distribution by empirical marginal distributions. As a result, the test statistics are equivalent to a V-statistic asymptotically. It is possible that the copula transformation can be applied to a broader class of estimators, and the analysis can be generalized to a broader class of test statistics.

Appendix A
Detailed Proofs

A.1 Some Lemmas

Lemma A.1.1. *Suppose that h_{f_1,f_2} is Mth-order continuously differentiable on $[0,1]^{4(p+q)}$, $M \in \mathbb{N}^+$. Then for any $i, j, r, s \in \mathbb{N}_n$, (2.14) holds.*

Proof. Set Rem $= (M!)^{-1} \sum_{\boldsymbol{\alpha} \in \mathbb{N}^M_{4(p+q)}} D^{(\boldsymbol{\alpha})} h_{f_1,f_2}(\tilde{\mathbf{W}}_{(i,j,r,s)}) \left(\hat{\mathbf{W}}_{(i,j,r,s)} - \mathbf{W}_{(i,j,r,s)}\right)^{(\boldsymbol{\alpha})}$. By (2.13),

$$
\begin{aligned}
& h_{f_1,f_2}(\hat{\mathbf{W}}_{(i,j,r,s)}) - h_{f_1,f_2}(\mathbf{W}_{(i,j,r,s)}) \\
= & \sum_{m=1}^{M-1} \frac{1}{m!} \sum_{\boldsymbol{\alpha} \in \mathbb{N}^m_{4(p+q)}} D^{(\boldsymbol{\alpha})} h_{f_1,f_2}(\mathbf{W}_{(i,j,r,s)}) \left(\hat{\mathbf{W}}_{(i,j,r,s)} - \mathbf{W}_{(i,j,r,s)}\right)^{(\boldsymbol{\alpha})} + \text{Rem} \\
= & \sum_{m=1}^{M-1} \frac{1}{m!} \sum_{\boldsymbol{\alpha} \in \mathbb{N}^m_{4(p+q)}} D^{(\boldsymbol{\alpha})} h_{f_1,f_2}(\mathbf{W}_{(i,j,r,s)}) \left[n^{-1} \sum_{t=1}^{n} \vec{I}_{\mathbf{Z}_{(t,t,t,t)}}(\mathbf{Z}_{(i,j,r,s)}) - \mathbf{W}_{(i,j,r,s)} \right]^{(\boldsymbol{\alpha})} + \text{Rem} \\
= & \sum_{m=1}^{M-1} \frac{1}{m!} \sum_{\boldsymbol{\alpha} \in \mathbb{N}^m_{4(p+q)}} D^{(\boldsymbol{\alpha})} h_{f_1,f_2}(\mathbf{W}_{(i,j,r,s)}) \cdot n^{-m} \sum_{t_1,\ldots,t_m=1}^{n} \prod_{c=1}^{m} \left[\vec{I}_{\mathbf{Z}_{(t_c,t_c,t_c,t_c)}}(\mathbf{Z}_{(i,j,r,s)}) - \mathbf{W}_{(i,j,r,s)} \right]^{(\alpha_c)} \\
& + \text{Rem} \\
= & \, n^{-(M-1)} \sum_{t_1,\ldots,t_{M-1}=1}^{n} \sum_{m=1}^{M-1} \frac{1}{m!} \cdot \sum_{\boldsymbol{\alpha} \in \mathbb{N}^m_{4(p+q)}} D^{(\boldsymbol{\alpha})} h_{f_1,f_2}(\mathbf{W}_{(i,j,r,s)}) \\
& \cdot \prod_{c=1}^{m} \left[\vec{I}_{\mathbf{Z}_{(t_c,t_c,t_c,t_c)}}(\mathbf{Z}_{(i,j,r,s)}) - \mathbf{W}_{(i,j,r,s)} \right]^{(\alpha_c)} + \text{Rem} \ ,
\end{aligned}
\tag{A.1}
$$

where the last equality follows by the identity

$$n^{-m} \sum_{t_1,\ldots,t_m=1}^{n} g_m(t_1,\ldots,t_m) = n^{-(M-1)} \sum_{t_1,\ldots,t_{M-1}=1}^{n} g_m(t_1,\ldots,t_m)$$

which holds for any $m < M$, and any function g_m that depends on m and $t_1,\ldots,t_m$.

□

Lemma A.1.2. *If* $\mathbf{u}_1 = (u_{11},\ldots,u_{1r})$ *and* $\mathbf{u}_2 = (u_{21},\ldots,u_{2r})$ *are two vectors in* $\mathbb{R}^r$ *such that* $u_{1i}u_{2i} > 0$ *for any* $i = 1,\ldots,r$*, then for any* $\tilde{\mathbf{u}}$ *in the line segment* $[\mathbf{u}_1,\mathbf{u}_2]$,

$$|\tilde{\mathbf{u}}|_r \geq \frac{1}{\sqrt{2}} \min\{|\mathbf{u}_1|_r, |\mathbf{u}_2|_r\} \tag{A.2}$$

Proof. If the three points $\mathbf{u}_1$, $\mathbf{u}_2$ and $\mathbf{0}$ fall on a straight line in $\mathbb{R}^r$, because $u_{1i}u_{2i} > 0$ for any $i = 1,\ldots,p$, there exists $\lambda > 0$ such that $\mathbf{u}_2 = \lambda\mathbf{u}_1$. Assume without loss of generality that $|\mathbf{u}_1|_r \leq |\mathbf{u}_2|_r$, so $\lambda \geq 1$. Then for any $\tilde{\mathbf{u}} \in [\mathbf{u}_1,\mathbf{u}_2]$, there exists $\tilde{\lambda} \in [1,\lambda]$ such that $\tilde{\mathbf{u}} = \tilde{\lambda}\mathbf{u}_1$. Thus

$$|\tilde{\mathbf{u}}|_r = \tilde{\lambda}\,|\mathbf{u}_1|_r \geq |\mathbf{u}_1|_r \geq \frac{1}{\sqrt{2}} \min\{|\mathbf{u}_1|_r, |\mathbf{u}_2|_r\} \tag{A.3}$$

and (A.2) is proved in this case. For the rest of the proof, we assume that the three vertices $\mathbf{u}_1$, $\mathbf{u}_2$ and $\mathbf{0}$ do not fall on a straight line in $\mathbb{R}^r$. This also implies that $r \geq 2$. Consider the triangle with the three vertices $\mathbf{u}_1$, $\mathbf{u}_2$ and $\mathbf{0}$. Because $\mathbf{u}_1^T\mathbf{u}_2 = \sum_i u_{1i}u_{2i} > 0$, the angle of the triangle at $\mathbf{0}$ is acute. We consider two cases depending on whether the triangle is acute or not.

If the triangle is not acute, then either the angle at $\mathbf{u}_1$ or $\mathbf{u}_2$ is not acute. Because $\tilde{\mathbf{u}}$ is at the segment $[\mathbf{u}_1,\mathbf{u}_2]$, the length of $\tilde{\mathbf{u}}$ must be between the lengths of $\mathbf{u}_1$ and $\mathbf{u}_2$. Thus $|\tilde{\mathbf{u}}|_p \geq \min\{|\mathbf{u}_1|_p, |\mathbf{u}_2|_p\}$.

If the triangle is acute, the minimum value of $|\tilde{\mathbf{u}}|$ is h, the length of the altitude of the triangle towards the base $[\mathbf{u}_1,\mathbf{u}_2]$. Without loss of generality, suppose that $|\mathbf{u}_1|_p \leq |\mathbf{u}_2|_p$. Then $\theta_1 \geq \theta_2$, where θ_i is the angle of the triangle at the vertex $\mathbf{u}_i$, $i = 1,2$. Because the triangle is acute, $\theta_1 + \theta_2 \geq 90°$, and thus $\theta_1 \geq 45°$. By the law of sines, $|\tilde{\mathbf{u}}|_p \geq h = |\mathbf{u}_1|_p \sin\theta_1 \geq \frac{1}{\sqrt{2}}|\mathbf{u}_1|_p = \frac{1}{\sqrt{2}}\min\{|\mathbf{u}_1|_p, |\mathbf{u}_2|_p\}$.

In both cases, $|\tilde{\mathbf{u}}|_p \geq \frac{1}{\sqrt{2}}\min\{|\mathbf{u}_1|_p, |\mathbf{u}_2|_p\}$. Thus (A.2) is proved. □

Lemma A.1.3. *Suppose $X_1, \ldots, X_n$ are i.i.d. copies of a random variable X, and let the distribution function F_X of X, and the empirical distribution function $\hat{F}_X$, be as defined in (2.1). Then for $n \geq 2$ and $\delta > 0$,*

$$P(0 < |F_X(X_2) - F_X(X_1)| < \delta) \leq 4\delta \tag{A.4}$$

$$P(0 < |\hat{F}_X(X_2) - \hat{F}_X(X_1)| < \delta) \leq 4\delta \frac{n}{n-1} \tag{A.5}$$

Proof. Let $U_1, \ldots, U_n$ be i.i.d. Uniform(0,1) random variables. We will use the representation $X_i = F_X^{-1}(U_i)$, where F_X^{-1} is the quantile function of X. Let $p_X(x) = P(X = x)$, and write

$$\begin{aligned}
&P(0 < F_X(X_2) - F_X(X_1) < \delta) \\
&= P\left(X_1 < X_2,\ 0 < \frac{1}{2}p_X(X_1) + (F_X^-(X_2) - F_X^+(X_1)) + \frac{1}{2}p_X(X_2) < \delta\right) \\
&= P(X_1 < X_2, 0 < p_X(X_1) + 2(F_X^-(X_2) - F_X^+(X_1)) + p_X(X_2) < 2\delta) \\
&\leq P(X_1 < X_2, p_X(X_1) + (F_X^-(X_2) - F_X^+(X_1)) + p_X(X_2) < 2\delta) \\
&= P(X_1 < X_2, F_X^+(X_2) - F_X^-(X_1) < 2\delta) \\
&= P(F_X^{-1}(U_1) < F_X^{-1}(U_2), F_X^+(F_X^{-1}(U_2)) - F_X^-(F_X^{-1}(U_1)) < 2\delta)\ .
\end{aligned} \tag{A.6}$$

By the properties of the quantile function, $F_X^-(F_X^{-1}(u)) \leq u \leq F_X^+(F_X^{-1}(u))$, $\forall\ u \in [0, 1]$, so that

$$F_X^+(F_X^{-1}(U_2)) - F_X^-(F_X^{-1}(U_1)) \geq U_2 - U_1\ . \tag{A.7}$$

From (A.6), (A.7) and the fact that F_X^{-1} is non-decreasing, we have

$$\begin{aligned}
&P(0 < F_X(X_2) - F_X(X_1) < \delta) \\
&\leq P(0 < U_2 - U_1 < 2\delta) = E[P(0 < U_2 - U_1 < 2\delta | U_1)] \\
&= E[P(U_1 < U_2 < U_1 + 2\delta | U_1)] \leq E[2\delta] = 2\delta\ .
\end{aligned}$$

By symmetry, we have $P(0 < F_X(X_1) - F_X(X_2) < \delta) \leq 2\delta$, and thus (A.4) holds.

To show (A.5) note that when $x_1 < x_2$,

$$\hat{F}_X(x_2) - \hat{F}_X(x_1) \quad = \quad \frac{1}{2n}\sum_{i=1}^{n} I(X_i = x_2) + \frac{1}{n}\sum_{i=1}^{n} I(x_1 < X_i < x_2) + \frac{1}{2n}\sum_{i=1}^{n} I(X_i = x_1)$$

$$\geq \frac{1}{2n}\sum_{i=1}^{n} I(x_1 \leq X_i \leq x_2) \ .$$

Therefore,

$$\begin{aligned}
& P(0 < \hat{F}_X(X_2) - \hat{F}_X(X_1) < \delta) = P(X_1 < X_2,\ \hat{F}_X(X_2) - \hat{F}_X(X_1) < \delta) \\
\leq\ & P\left(X_1 < X_2,\ \frac{1}{2n}\sum_{i=1}^{n} I(X_1 \leq X_i \leq X_2) < \delta\right) \\
=\ & P\left(F_X^{-1}(U_1) < F_X^{-1}(U_2),\ \frac{1}{n}\sum_{i=1}^{n} I(F_X^{-1}(U_1) \leq F_X^{-1}(U_i) \leq F_X^{-1}(U_2)) < 2\delta\right) \\
\leq\ & P\left(U_1 < U_2,\ \frac{1}{n}\sum_{i=1}^{n} I(U_1 \leq U_i \leq U_2) < 2\delta\right) \\
=\ & P\left(U_1 < U_2,\ \frac{1}{n}\sum_{i=1}^{n} I(U_i \leq U_2) - \frac{1}{n}\sum_{i=1}^{n} I(U_i < U_1) < 2\delta\right) \\
\leq\ & P\left(U_1 < U_2,\ \frac{1}{n}(R_2 - R_1) < 2\delta\right) \\
=\ & P\left(0 < R_2 - R_1 < 2n\delta\right),
\end{aligned}$$

where $R_1, \ldots, R_n$ are the rank statistics of $U_1, \ldots, U_n$. By the same argument,

$$P(0 < \hat{F}_X(X_1) - \hat{F}_X(X_2) < \delta) \leq P\left(0 < R_1 - R_2 < 2n\delta\right). \tag{A.8}$$

Therefore,

$$\begin{aligned}
& P(0 < |\hat{F}_X(X_2) - \hat{F}_X(X_1)| < \delta) \\
=\ & P(0 < \hat{F}_X(X_2) - \hat{F}_X(X_1) < \delta) + P(0 < \hat{F}_X(X_1) - \hat{F}_X(X_2) < \delta) \\
\leq\ & P\left(0 < R_2 - R_1 < 2n\delta\right) + P\left(0 < R_1 - R_2 < 2n\delta\right) \\
=\ & P\left(|R_2 - R_1| < 2n\delta\right) \ = \ E[P(|R_2 - R_1| < 2n\delta | R_1)] \\
=\ & E[P(R_1 - 2n\delta < R_2 < R_1 + 2n\delta | R_1)] \\
\leq\ & E\left[4n\delta \cdot \frac{1}{n-1}\middle| R_1\right] \ = \ 4\delta\frac{n}{n-1} \ ,
\end{aligned}$$

where the last inequality follows from the fact that conditioning on R_1, the distribution of R_2 is uniform in the finite set consisting of the $n-1$ elements $\{1, 2, \ldots, n\} - \{R_1\}$, and the fact that the number of integers that are contained in the interval $(R_1 - 2n\delta, R_1 + 2n\delta)$ and unequal to R_1 is no more than $4n\delta$. Thus, (A.5) is proved. $\square$

Lemma A.1.4. *Let $\tilde{h}^{(1)}_{d_p,d_q}(\mathbf{z};\mathbf{F_Z})$, $\tilde{\eta}^{(1)}_{d_p,d_q}(\mathbf{z})$ and $\tilde{\zeta}^{(1)}_{d_p,d_q,Total}(\mathbf{z})$ be defined in (2.21), (2.23) and (2.30), respectively. If $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$ then*

a) $\tilde{h}^{(1)}_{d_p,d_q}(\mathbf{z};\mathbf{F_Z}) = 0$,

b) $\tilde{\eta}^{(1)}_{d_p,d_q}(\mathbf{z}) = 0$, and

c) $\tilde{\zeta}^{(1)}_{d_p,d_q,Total}(\mathbf{z}) = 0$.

Proof. a) Write $\tilde{h}^{(1)}_{d_p,d_q}(\mathbf{z};\mathbf{F_Z}) = J_1 - 2J_2 + J_3$, where

$$\begin{aligned}
J_1 &= \sum_{s=1}^{4} E\left[k(\mathbf{X}_1,\mathbf{X}_2;\mathbf{F_X})l(\mathbf{Y}_3,\mathbf{Y}_4;\mathbf{F_Y})|\mathbf{Z}_s=\mathbf{z}\right] \\
&= 2\left\{E[k(\mathbf{x},\mathbf{X}_2;\mathbf{F_X})l(\mathbf{Y}_3,\mathbf{Y}_4);\mathbf{F_Y})] + E[k(\mathbf{X}_1,\mathbf{X}_2;\mathbf{F_X})l(\mathbf{y},\mathbf{Y}_4);\mathbf{F_Y})]\right\} \\
&= 2E[k(\mathbf{x},\mathbf{X}_1;\mathbf{F_X})]E[l(\mathbf{Y}_1,\mathbf{Y}_2);\mathbf{F_Y})] + 2E[k(\mathbf{X}_1,\mathbf{X}_2;\mathbf{F_X})]E[l(\mathbf{y},\mathbf{Y}_1);\mathbf{F_Y})], \\
J_2 &= \sum_{s=1}^{4} E\left[k(\mathbf{X}_1,\mathbf{X}_2;\mathbf{F_X})l(\mathbf{Y}_1,\mathbf{Y}_3;\mathbf{F_Y})|\mathbf{Z}_s=\mathbf{z}\right] \\
&= E[k(\mathbf{x},\mathbf{X}_2;\mathbf{F_X})l(\mathbf{y},\mathbf{Y}_3;\mathbf{F_Y})] + E[k(\mathbf{X}_1,\mathbf{x};\mathbf{F_X})l(\mathbf{Y}_1,\mathbf{Y}_3;\mathbf{F_Y})] \\
&\quad + E[k(\mathbf{X}_1,\mathbf{X}_2)l(\mathbf{Y}_1,\mathbf{y};\mathbf{F_Y})] \\
&= E[k(\mathbf{x},\mathbf{X}_1;\mathbf{F_X})]E[l(\mathbf{y},\mathbf{Y}_1;\mathbf{F_Y})] + E[k(\mathbf{X}_1,\mathbf{x};\mathbf{F_X})]E[l(\mathbf{Y}_1,\mathbf{Y}_2;\mathbf{F_Y})] \\
&\quad + E[k(\mathbf{X}_1,\mathbf{X}_2;\mathbf{F_X})]E[l(\mathbf{Y}_1,\mathbf{y};\mathbf{F_Y})], \\
J_3 &= \sum_{s=1}^{4} E\left[k(\mathbf{X}_1,\mathbf{X}_2;\mathbf{F_X})l(\mathbf{Y}_1,\mathbf{Y}_2;\mathbf{F_Y})|\mathbf{Z}_s=\mathbf{z}\right] \\
&= 2E[k(\mathbf{x},\mathbf{X}_1;\mathbf{F_X})l(\mathbf{y},\mathbf{Y}_1;\mathbf{F_Y})] = 2E[k(\mathbf{x},\mathbf{X}_1;\mathbf{F_X})]E[l(\mathbf{y},\mathbf{Y}_1;\mathbf{F_Y})].
\end{aligned}$$

Since $J_1 - 2J_2 + J_3 = 0$, part a) is shown.

b) Let $\eta_{d_p,d_q}, \tilde{\eta}_{d_p,d_q}$ and $\tilde{\eta}^{(1)}_{d_p,d_q}$ be as defined in (2.23). Write

$$\eta_{d_p,d_q} = \sum_{j=1}^{2} \eta_{d_p,d_q,j} \ ,$$

where

$$\eta_{d_p,d_q,1}(\mathbf{Z}_{(1,\ldots,5)}) \ = \ \sum_{\alpha\in\mathbb{N}_{2p}} D^{(\alpha)}d_p(\mathbf{U}_1,\mathbf{U}_2)\left(d_q(\mathbf{V}_3,\mathbf{V}_4) - 2d_q(\mathbf{V}_1,\mathbf{V}_3) + d_q(\mathbf{V}_1,\mathbf{V}_2)\right)$$

$$
\begin{aligned}
& \cdot \left[\vec{I}_{\mathbf{X}_{(5,5)}}(\mathbf{X}_{(1,2)}) - \mathbf{U}_{(1,2)}\right]^{(\alpha)} I_{B_{1234}}, \\
\eta_{d_p,d_q,2}(\mathbf{Z}_{(1,\ldots,5)}) \quad = \quad & d_p(\mathbf{U}_1,\mathbf{U}_2) \sum_{\alpha\in\mathbb{N}_{4q}} D^{(\alpha)} \left(d_q(\mathbf{V}_3,\mathbf{V}_4) - 2d_q(\mathbf{V}_1,\mathbf{V}_3) + d_q(\mathbf{V}_1,\mathbf{V}_2)\right) \\
& \cdot \left[\vec{I}_{\mathbf{Y}_{(5,5,5,5)}}(\mathbf{Y}_{(1,2,3,4)}) - \mathbf{V}_{(1,2,3,4)}\right]^{(\alpha)} I_{B_{1234}}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
5\tilde{\eta}^{(1)}_{d_p,d_q}(\mathbf{z}) &= E\left[\frac{1}{5!}\sum_{(i_1,\ldots,i_5)\in I_5^5} \eta_{d_p,d_q}(\mathbf{Z}_{(i_1,\ldots,i_5)})\middle|\mathbf{Z}_1=\mathbf{z}\right] \qquad \text{(A.9)} \\
&= \frac{1}{5}\sum_{i\in I_5} E\left[\eta_{d_p,d_q}(\mathbf{Z}_{(1,\ldots,5)})\middle|\mathbf{Z}_i=\mathbf{z}\right] \\
&= \sum_{j=1}^{2}\frac{1}{5}\sum_{i\in I_5} E\left[\eta_{d_p,d_q,j}(\mathbf{Z}_{(1,\ldots,5)})\middle|\mathbf{Z}_i=\mathbf{z}\right] \qquad \text{(A.10)}
\end{aligned}
$$

By (A.9), part b) will follow by showing that

$$
E\left[\eta_{d_p,d_q,j}(\mathbf{Z}_{(1,\ldots,5)})\middle|\mathbf{Z}_i=\mathbf{z}\right] = 0,\ i=1,\ldots,5,\ j=1,2. \qquad \text{(A.11)}
$$

Relation (A.11) will be shown in detail for the case of $j=1$. The case of $j=2$ follows by similar arguments. Let

$$
C_{1234} = \left\{\min_{\substack{a_1,a_2\in\{1,2,3,4\}\\ a_1<a_2}} |\mathbf{U}_{a_1a_2}|_p > 0\right\}, \quad D_{1234} = \left\{\min_{\substack{a_1,a_2\in\{1,2,3,4\}\\ a_1<a_2}} |\mathbf{V}_{a_1a_2}|_q > 0\right\},
$$

so that $I_{B_{1234}} = I_{C_{1234}} I_{D_{1234}}$, and notice that under $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$, for any $i\in I_6$,

$$
\begin{aligned}
& E\left[\eta_{d_p,d_q,1}(\mathbf{Z}_{(1,\ldots,5)})\middle|\mathbf{Z}_i=\mathbf{z}\right] \\
&= \sum_{\alpha\in\mathbb{N}_{2p}} E\left\{D^{(\alpha)} d_p(\mathbf{U}_1,\mathbf{U}_2)\left[\vec{I}_{\mathbf{X}_{(5,5)}}(\mathbf{X}_{(1,2)}) - \mathbf{U}_{(1,2)}\right]^{(\alpha)} I_{C_{1234}}\middle|\mathbf{X}_i=\mathbf{x}\right\} \\
&\quad \cdot E\left\{\left(d_q(\mathbf{V}_3,\mathbf{V}_4) - 2d_q(\mathbf{V}_1,\mathbf{V}_3) + d_q(\mathbf{V}_1,\mathbf{V}_2)\right) I_{D_{1234}}\middle|\mathbf{Y}_i=\mathbf{y}\right\}, \qquad \text{(A.12)}
\end{aligned}
$$

If $i \neq 5$, each summand in (A.12) is 0 because

$$\sum_{\alpha \in \mathbb{N}_{2p}} E\left\{ D^{(\alpha)} d_p(\mathbf{U}_1, \mathbf{U}_2) \left[\vec{I}_{\mathbf{X}_{(5,5)}}(\mathbf{X}_{(1,2)}) - \mathbf{U}_{(1,2)} \right]^{(\alpha)} I_{C_{1234}} \middle| \mathbf{X}_i = \mathbf{x} \right\} = 0 \ ,$$

which follows by the fact that

$$E\left\{ \vec{I}_{\mathbf{X}_{(5,5)}}(\mathbf{X}_{(1,2)}) - \mathbf{U}_{(1,2)} \middle| \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4 \right\} = 0 \ .$$

If $i = 5$, each summand in (A.12) is 0 because

$$E\left\{ \left(d_q(\mathbf{V}_3, \mathbf{V}_4) - 2d_q(\mathbf{V}_1, \mathbf{V}_3) + d_q(\mathbf{V}_1, \mathbf{V}_2)\right) I_{D_{1234}} \middle| \mathbf{Y}_5 = \mathbf{y} \right\} = 0.$$

Thus part b) is shown.

c) By parts a) and b), and the definition of $\tilde{\zeta}^{(1)}_{d_p,d_q,Total}(\mathbf{z})$, it suffices to show that

$$\tilde{\zeta}^{(1)}_{d_p,d_q}(\mathbf{z}) := E(\tilde{\zeta}_{d_p,d_q}(\mathbf{Z}_{(1,\ldots,6)}|\mathbf{Z}_1 = \mathbf{z}) = 0,$$

where with $\zeta_{d_p,d_q}(\mathbf{Z}_{(1,\ldots,6)})$ as defined in (2.26),

$$\tilde{\zeta}_{d_p,d_q}(\mathbf{Z}_{(1,\ldots,6)}) = \frac{1}{6!} \sum_{(i_1,\ldots,i_6) \in I_6^6} \zeta_{d_p,d_q}(\mathbf{Z}_{(i_1,\ldots,i_6)}).$$

Write

$$\zeta_{d_p,d_q} = \sum_{j=1}^{3} \zeta_{d_p,d_q,j} \ , \tag{A.13}$$

where

$$\begin{aligned}
\zeta_{d_p,d_q,1}(\mathbf{Z}_{(1,\ldots,6)}) &= \frac{1}{2} \sum_{\boldsymbol{\alpha} \in \mathbb{N}_{2p}^2} D^{(\boldsymbol{\alpha})} d_p(\mathbf{U}_1, \mathbf{U}_2) \left(d_q(\mathbf{V}_3, \mathbf{V}_4) - 2d_q(\mathbf{V}_1, \mathbf{V}_3) + d_q(\mathbf{V}_1, \mathbf{V}_2)\right) \\
&\quad \cdot \left[\vec{I}_{\mathbf{X}_{(5,5)}}(\mathbf{X}_{(1,2)}) - \mathbf{U}_{(1,2)} \right]^{(\boldsymbol{\alpha})} I_{B_{1234}}, \\
\zeta_{d_p,d_q,2}(\mathbf{Z}_{(1,\ldots,6)}) &= \frac{1}{2} d_p(\mathbf{U}_1, \mathbf{U}_2) \sum_{\boldsymbol{\alpha} \in \mathbb{N}_{4p}^2} D^{(\boldsymbol{\alpha})} \left(d_q(\mathbf{V}_3, \mathbf{V}_4) - 2d_q(\mathbf{V}_1, \mathbf{V}_3) + d_q(\mathbf{V}_1, \mathbf{V}_2)\right) \\
&\quad \cdot \left[\vec{I}_{\mathbf{Y}_{(5,5,5,5)}}(\mathbf{Y}_{(1,2,3,4)}) - \mathbf{V}_{(1,2,3,4)} \right]^{(\boldsymbol{\alpha})} I_{B_{1234}},
\end{aligned}$$

$$\zeta_{d_p,d_q,3}(\mathbf{Z}_{(1,\ldots,6)}) = \sum_{\alpha\in\mathbb{N}_{2p}}\sum_{\beta\in\mathbb{N}_{4q}} D^{(\alpha)}d_p(\mathbf{U}_1,\mathbf{U}_2)D^{(\beta)}\left(d_q(\mathbf{V}_3,\mathbf{V}_4) - 2d_q(\mathbf{V}_1,\mathbf{V}_3) + d_q(\mathbf{V}_1,\mathbf{V}_2)\right)$$
$$\cdot\left[\vec{I}_{\mathbf{X}_{(5,5)}}(\mathbf{X}_{(1,2)}) - \mathbf{U}_{(1,2)}\right]^{(\alpha)}\left[\vec{I}_{\mathbf{Y}_{(6,6,6,6)}}(\mathbf{Y}_{(1,2,3,4)}) - \mathbf{V}_{(1,2,3,4)}\right]^{(\beta)} I_{B_{1234}}\ .$$

Note that

$$\tilde{\zeta}^{(1)}_{d_p,d_q}(\mathbf{z}) = \frac{1}{6}\sum_{i\in I_6} E\left[\zeta_{d_p,d_q}(\mathbf{Z}_{(1,\ldots,6)})\middle|\mathbf{Z}_i=\mathbf{z}\right] = \sum_{j=1}^{3}\frac{1}{6}\sum_{i\in I_6} E\left[\zeta_{d_p,d_q,j}(\mathbf{Z}_{(1,\ldots,6)})\middle|\mathbf{Z}_i=\mathbf{z}\right] \tag{A.14}$$

By (A.14), part c) will follow by showing that

$$E\left[\zeta_{d_p,d_q,j}(\mathbf{Z}_{(1,\ldots,6)})\middle|\mathbf{Z}_i=\mathbf{z}\right] = 0, \quad j=1,2,\ i=1,\ldots,6. \tag{A.15}$$

Each of the equations in (A.15) can be shown by an argument similar to that used for showing (A.11). □

Lemma A.1.5. *Under the notation and assumptions of Theorem 3.2.1*

$$\hat{\gamma}_{k,l}(\mathbf{X},\mathbf{Y};\mathbf{F}_\mathbf{Z}) - \gamma_{k,l}(\mathbf{X},\mathbf{Y};\mathbf{F}_\mathbf{Z}) = \frac{4}{n}\sum_{i=1}^{n}\tilde{h}^{(1)}_{k,l}(\mathbf{Z}_i;\mathbf{F}_\mathbf{Z}) + O_P(n^{-1}). \tag{A.16}$$

Moreover, if $\mathbf{X}\perp\!\!\!\perp\mathbf{Y}$,

$$\hat{\gamma}_{k,l}(\mathbf{X},\mathbf{Y};\mathbf{F}_\mathbf{Z}) = O_P(n^{-1}). \tag{A.17}$$

Proof. The U-statistic corresponding to $\hat{\gamma}_{k,l}$ is

$$\hat{\gamma}_{k,l,U}(\mathbf{X},\mathbf{Y};\mathbf{F}_\mathbf{Z}) = \frac{1}{(n)_4}\sum_{(i,j,q,r)\in I_n^4} h_{k,l}(\mathbf{Z}_i,\mathbf{Z}_j,\mathbf{Z}_q,\mathbf{Z}_r;\mathbf{F}_\mathbf{Z}),$$

where I_n^m denotes the set of all m-permutations $(i_1,\ldots,i_m)$ of the numbers $1,\ldots,n$, and $(n)_m$ denotes the number of such permutations. The number of terms in the difference of $n^4\hat{\gamma}_{k,l}(\mathbf{X},\mathbf{Y};\mathbf{F}_\mathbf{Z})$ and $\binom{n}{4}\hat{\gamma}_{k,l,U}(\mathbf{X},\mathbf{Y};\mathbf{F}_\mathbf{Z})$ is $n^4-(n)_4\leq 2n^3$. Each term is bounded by $4M_0^2$. Therefore,

$$|\hat{\gamma}_{k,l}(\mathbf{X},\mathbf{Y};\mathbf{F}_\mathbf{Z}) - \hat{\gamma}_{k,l,U}(\mathbf{X},\mathbf{Y};\mathbf{F}_\mathbf{Z})|$$

$$= \left| \left(\frac{1}{n^4} - \frac{1}{(n)_4}\right) \sum_{(i,j,q,r)\in I_n^4} h_{k,l}(\mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_q, \mathbf{Z}_r; \mathbf{F}_\mathbf{Z}) + \frac{1}{n^4} \sum_{\substack{i,j,q,r=1 \\ (i,j,q,r)\notin I_n^4}}^{n} h_{k,l}(\mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_q, \mathbf{Z}_r; \mathbf{F}_\mathbf{Z}) \right|$$
$$\leq (2n^{-1} + 2n^{-1})4M_0^2 \leq 8M_0^2 n^{-1}. \tag{A.18}$$

By generalizing the theorem in Section 5.3.2 in [23] to random vector observations,

$$\hat{\gamma}_{k,l,U}(\mathbf{X}, \mathbf{Y}; \mathbf{F}_\mathbf{Z}) - \gamma_{k,l}(\mathbf{X}, \mathbf{Y}; \mathbf{F}_\mathbf{Z}) = \frac{4}{n} \sum_{i=1}^{n} h_{k,l}^{(1)}(\mathbf{Z}_i; \mathbf{F}_\mathbf{Z}) + O_P(n^{-1}) \; . \tag{A.19}$$

Thus, (A.16) follows from (A.18) and (A.19). If $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$ then, by Lemma A.1.4, $\tilde{h}_{k,l}^{(1)} = 0$. Since also $\gamma_{k,l}(\mathbf{X}, \mathbf{Y}; \mathbf{F}_\mathbf{Z}) = 0$ under independence, (A.17) follows from (A.16). □

Lemma A.1.6. *Suppose that* $\mathbf{Z}$ *is a* $(p + q)$*-dimensional random vector, let* $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ *are i.i.d. copies of* $\mathbf{Z}$*, and let* $\mathbf{F}_\mathbf{Z}, \hat{\mathbf{F}}_\mathbf{Z}$ *be defined as in (2.1). Then for any* $\alpha > 0$*, there exists a finite positive constant* $C_{p,q,\alpha}$ *which does not depend on the distribution of* $\mathbf{Z}$*, such that*

$$E\left(\|\hat{\mathbf{F}}_\mathbf{Z} - \mathbf{F}_\mathbf{Z}\|_\infty^\alpha\right) \leq \frac{C_{p,q,\alpha}}{n^{\alpha/2}}, \quad n = 1, 2, \ldots \tag{A.20}$$

Proof. Write $\mathbf{Z} = (Z_1, \ldots, Z_{p+q})$, and $\mathbf{Z}_i = (Z_{i,1}, \ldots, Z_{i,p+q})$ for $i = 1, \ldots, n$. Define

$$D_{Z_i}^+ = \sup_z |\hat{F}_{Z_i}^+(z) - F_{Z_i}^+(z)|, \quad D_{Z_i}^- = \sup_z |\hat{F}_{Z_i}^-(z) - F_{Z_i}^-(z)|, \quad D_{Z_i} = \sup_z |\hat{F}_{Z_i}(z) - F_{Z_i}(z)|,$$

for any $i = 1, \ldots, n$, where

$$\hat{F}_{Z_i}^+(z) = \frac{1}{n} \sum_{i=1}^{n} I(Z_i \leq z), \quad \hat{F}_{Z_i}^-(z) = \frac{1}{n} \sum_{i=1}^{n} I(Z_i < z) \; .$$

Then $\hat{F}_{Z_i} = \frac{1}{2}(\hat{F}_{Z_i}^+ + \hat{F}_{Z_i}^-)$. By the Dvoretzky-Kiefer-Wolfowitz inequality, there exists a finite positive constant C such that

$$P(D_{Z_i}^+ > d) \leq Ce^{-2nd^2}, \quad P(D_{Z_i}^- > d) \leq Ce^{-2nd^2}, \quad d > 0, n \in \mathbb{N}.$$

The second inequality can be shown by considering the random variable $-Z_i$. Note that

$$
\begin{aligned}
D_{Z_i} &= \sup_z \left| \hat{F}_{Z_i}(z) - F_{Z_i}(z) \right| = \sup_x \left| \frac{1}{2}(\hat{F}^+_{Z_i}(z) - F^+_{Z_i}(z)) + \frac{1}{2}(\hat{F}^-_{Z_i}(z) - F^-_{Z_i}(z)) \right| \\
&\leq \frac{1}{2} \sup_x \left| \hat{F}^+_{Z_i}(z) - F^+_{Z_i}(z) \right| + \frac{1}{2} \sup_x \left| \hat{F}^-_{Z_i}(z) - F^-_{Z_i}(z) \right| = \frac{1}{2} D^+_n + \frac{1}{2} D^-_n \ .
\end{aligned}
$$

Thus

$$
P(D_n > d) \leq P(D^+_n > d) + P(D^-_n > d) \leq 2Ce^{-2nd^2}, \quad d > 0, n \in \mathbb{N} \ .
$$

Noting that

$$
\|\hat{\mathbf{F}}_{\mathbf{Z}} - \mathbf{F}_{\mathbf{Z}}\|_\infty = \max_{j=1,\ldots,p+q} D_{Z_j},
$$

we have

$$
P\left(\|\hat{\mathbf{F}}_{\mathbf{Z}} - \mathbf{F}_{\mathbf{Z}}\|_\infty > d \right) \leq \sum_{j=1}^{p+q} P(D_{Z_j} > d) \leq 2(p+q)Ce^{-2nd^2} \ .
$$

Therefore,

$$
\begin{aligned}
& E\left(\|\hat{\mathbf{F}}_{\mathbf{Z}} - \mathbf{F}_{\mathbf{Z}}\|^\alpha_\infty \right) \\
= & \int_0^\infty P\left(\|\hat{\mathbf{F}}_{\mathbf{Z}} - \mathbf{F}_{\mathbf{Z}}\|^\alpha_\infty > s \right) ds = \int_0^\infty P\left(\|\hat{\mathbf{F}}_{\mathbf{Z}} - \mathbf{F}_{\mathbf{Z}}\|_\infty > s^{1/\alpha} \right) ds \\
\leq & \int_0^\infty 2Ce^{-2ns^{2/\alpha}} ds = \int_0^\infty 2C(p+q)e^{-t} d\left((t/2n)^{\alpha/2} \right) \\
= & \ 2C(p+q)(\alpha/2)(2n)^{-\alpha/2} \int_0^\infty t^{\alpha/2-1} e^{-t} dt \\
= & \ \frac{C(p+q)\alpha\Gamma(\alpha/2)}{2^{\alpha/2}} \cdot \frac{1}{n^{\alpha/2}} \ .
\end{aligned}
$$

□

A.2 Derivations Needed for the Proof of Theorem 2.3.1

A.2.1 Proof of (2.40)

Let R_2 be given by (2.39). We will show that $R_2 = o_p(n^{-1/2})$. We have

$$\begin{aligned} R_2 &= \frac{1}{n^4} \sum_{i,j,r,s=1}^{n} \left[\left|\hat{\mathbf{U}}_{ij}\right|_p \left(\left|\hat{\mathbf{V}}_{rs}\right|_q - 2\left|\hat{\mathbf{V}}_{ir}\right|_q + \left|\hat{\mathbf{V}}_{ij}\right|_q \right) - |\mathbf{U}_{ij}|_p \left(|\mathbf{V}_{rs}|_q - 2|\mathbf{V}_{ir}|_q + |\mathbf{V}_{ij}|_q \right) \right] I_{A^c_{ijrs}} \\ &= \frac{1}{n^4} \sum_{i,j,r,s=1}^{n} \Bigg\{ \left(\left|\hat{\mathbf{U}}_{ij}\right|_p - |\mathbf{U}_{ij}|_p \right) \left(\left|\hat{\mathbf{V}}_{rs}\right|_q - 2\left|\hat{\mathbf{V}}_{ir}\right|_q + \left|\hat{\mathbf{V}}_{ij}\right|_q \right) \\ &\qquad + |\mathbf{U}_{ij}|_p \left[\left(\left|\hat{\mathbf{V}}_{rs}\right|_q - |\mathbf{V}_{rs}|_q \right) - 2\left(\left|\hat{\mathbf{V}}_{ir}\right|_q - |\mathbf{V}_{ir}|_q \right) + \left(\left|\hat{\mathbf{V}}_{ij}\right|_q - |\mathbf{V}_{ij}|_q \right) \right] \Bigg\} I_{A^c_{ijrs}} \\ &\leq 16\sqrt{pq} \|\hat{F}_{\mathbf{Z}} - F_{\mathbf{Z}}\|_\infty \cdot \frac{1}{n^4} \sum_{i,j,r,s=1}^{n} I_{A^c_{ijrs}}, \end{aligned} \tag{A.21}$$

where the inequality follows by the fact that for any $t_1, t_2 \in \mathbb{N}_n$,

$$\begin{aligned} \left| \left|\hat{\mathbf{U}}_{t_1 t_2}\right|_p - |\mathbf{U}_{t_1 t_2}|_p \right| &\leq \left|\hat{\mathbf{U}}_{t_1 t_2} - \mathbf{U}_{t_1 t_2}\right|_p = \left|\hat{\mathbf{F}}_{\mathbf{X}}(\mathbf{X}_{t_1}) - \hat{\mathbf{F}}_{\mathbf{X}}(\mathbf{X}_{t_2}) - \mathbf{F}_{\mathbf{X}}(\mathbf{X}_{t_1}) + \mathbf{F}_{\mathbf{X}}(\mathbf{X}_{t_2})\right|_p \\ &\leq 2\sqrt{p} \|\hat{\mathbf{F}}_{\mathbf{Z}} - \mathbf{F}_{\mathbf{Z}}\|_\infty, \end{aligned} \tag{A.22}$$

and similarly

$$\left| \left|\hat{\mathbf{V}}_{t_1 t_2}\right|_q - |\mathbf{V}_{t_1 t_2}|_q \right| \leq 2\sqrt{q} \|\hat{\mathbf{F}}_{\mathbf{Z}} - \mathbf{F}_{\mathbf{Z}}\|_\infty. \tag{A.23}$$

Next, write

$$\begin{aligned} E\left(\frac{1}{n^4} \sum_{i,j,r,s=1}^{n} I_{A^c_{ijrs}} \right) &= \frac{1}{n^4} \sum_{i,j,r,s=1}^{n} P\left(A^c_{ijrs}\right) \\ \leq \frac{1}{n^4} \sum_{i,j,r,s=1}^{n} \sum_{\substack{a,b \in \{i,j,r,s\} \\ a \neq b}} &\Big[P\left(|\mathbf{U}_{ab}|_p \leq \delta_n\right) + P\left(\left|\hat{\mathbf{U}}_{ab}\right|_p \leq \delta_n\right) \\ &+ P\left(|\mathbf{V}_{ab}|_p \leq \delta_n\right) + P\left(\left|\hat{\mathbf{V}}_{ab}\right|_p \leq \delta_n\right) \Big] \\ \leq 96\delta_n \frac{n}{n-1}, & \end{aligned} \tag{A.24}$$

where the last inequality follows by Lemma A.1.3, while the first inequality follows by noting that if the Euclidean norm of a vector is less than δ_n then the absolute value of each component is less than δ_n. Since also $\|\hat{F}_{\mathbf{Z}} - F_{\mathbf{Z}}\|_\infty = O_P(n^{-1/2})$, $R_2 = O_P(n^{-1/2}\delta_n) = o_P(n^{-1/2})$ follows from (A.21) and (A.24).

A.2.2 Proof of (2.42) and of (2.45)

Let R_3 be defined in (2.41). We will show that $R_3 = o_P(n^{-1/2})$. Note that the form of R_3 results the remainder of the Taylor expansion in (2.14) for $M = 2$. For the purposes of this proof it is more convenient to express this remainder term in terms of a reparametrization of the function h_{d_p,d_q}. For any $i, j, r, s \in \mathbb{N}_n$, define $\mathbf{W}_{ij} = \mathbf{W}_i - \mathbf{W}_j$, $\hat{\mathbf{W}}_{ij} = \hat{\mathbf{W}}_i - \hat{\mathbf{W}}_j$, and

$$\mathbf{W}_{ijrs} = (\mathbf{W}_{ij}, \mathbf{W}_{ir}, \mathbf{W}_{rs}),\ \hat{\mathbf{W}}_{ijrs} = (\hat{\mathbf{W}}_{ij}, \hat{\mathbf{W}}_{ir}, \hat{\mathbf{W}}_{rs})\ .$$

Because d_p and d_q are Euclidean norms, $h_{d_p,d_q}(\mathbf{W}_{(i,j,r,s)})$ is a function of $\mathbf{W}_{ijrs}$, i.e.,

$$h^*_{d_p,d_q}(\mathbf{W}_{ij}, \mathbf{W}_{ir}, \mathbf{W}_{rs}) = h_{d_p,d_q}(\mathbf{W}_{(i,j,r,s)})\ , \tag{A.25}$$

for a suitable function $h^*_{d_p,d_q}$. Thus, by a Taylor expansion with $M = 2$,

$$\begin{aligned} &h_{d_p,d_q}(\hat{\mathbf{W}}_{(i,j,r,s)}) - h_{d_p,d_q}(\mathbf{W}_{(i,j,r,s)}) = h^*_{d_p,d_q}(\hat{\mathbf{W}}_{ijrs}) - h^*_{d_p,d_q}(\mathbf{W}_{ijrs}) \\ &= \sum_{\alpha\in\mathbb{N}_{3(p+q)}} D^{(\alpha)} h^*_{d_p,d_q}(\mathbf{W}_{ijrs}) \left[\hat{\mathbf{W}}_{ijrs} - \mathbf{W}_{ijrs}\right]^{(\alpha)} \\ &\quad + \frac{1}{2} \sum_{\boldsymbol{\alpha}\in\mathbb{N}^2_{3(p+q)}} D^{(\boldsymbol{\alpha})} h^*_{d_p,d_q}(\tilde{\mathbf{W}}_{ijrs}) \left[\hat{\mathbf{W}}_{ijrs} - \mathbf{W}_{ijrs}\right]^{(\boldsymbol{\alpha})}, \end{aligned}$$

for some $\tilde{\mathbf{W}}_{ijrs} = (\tilde{\mathbf{W}}_{ij}, \tilde{\mathbf{W}}_{ir}, \tilde{\mathbf{W}}_{rs}) \in [\mathbf{W}_{ij}, \hat{\mathbf{W}}_{ij}] \times [\mathbf{W}_{ir}, \hat{\mathbf{W}}_{ir}] \times [\mathbf{W}_{rs}, \hat{\mathbf{W}}_{rs}]$. It can be shown by a change of variables that

$$\begin{aligned} &\sum_{\alpha\in\mathbb{N}_{3(p+q)}} D^{(\alpha)} h^*_{d_p,d_q}(\mathbf{W}_{ijrs}) \left[\hat{\mathbf{W}}_{ijrs} - \mathbf{W}_{ijrs}\right]^{(\alpha)} \\ &= \sum_{\alpha\in\mathbb{N}_{4(p+q)}} D^{(\alpha)} h_{d_p,d_q}(\mathbf{W}_{(i,j,r,s)}) \left[\hat{\mathbf{W}}_{(i,j,r,s)} - \mathbf{W}_{(i,j,r,s)}\right]^{(\alpha)}. \end{aligned}$$

Thus, an equivalent expression of the remainder term R_3 is

$$R_3 = \frac{1}{n^4}\sum_{i,j,r,s=1}^{n}\frac{1}{2}\sum_{\boldsymbol{\alpha}\in\mathbb{N}^2_{3(p+q)}} D^{(\boldsymbol{\alpha})}h^*_{d_p,d_q}(\tilde{\mathbf{W}}_{ijrs})\left[\hat{\mathbf{W}}_{ijrs}-\mathbf{W}_{ijrs}\right]^{(\boldsymbol{\alpha})} I_{A_{ijrs}}.$$

It can be shown by straightforward calculus that for any $|\boldsymbol{\alpha}|=2$

$$|D^{(\boldsymbol{\alpha})}h^*_{d_p,d_q}(\tilde{\mathbf{W}}_{ijrs})| \leq \frac{4(\sqrt{p}+\sqrt{q}+2)}{\min\left\{\left|\tilde{\mathbf{U}}_{ij}\right|_p,\left|\tilde{\mathbf{V}}_{ij}\right|_q,\left|\tilde{\mathbf{V}}_{ir}\right|_q,\left|\tilde{\mathbf{V}}_{rs}\right|_q\right\}},$$

where $(\tilde{\mathbf{U}}_{ab},\tilde{\mathbf{V}}_{ab})=\tilde{\mathbf{W}}_{ab}$ for any $a,b\in\{i,j,r,s\}$. By Lemma A.1.2,

$$\begin{aligned}\min\left\{\left|\tilde{\mathbf{U}}_{ij}\right|_p,\left|\tilde{\mathbf{V}}_{ij}\right|_q,\left|\tilde{\mathbf{V}}_{ir}\right|_q,\left|\tilde{\mathbf{V}}_{rs}\right|_q\right\} &\geq \min_{\substack{a,b\in\{i,j,r,s\}\\ a<b}}\min\left\{\left|\tilde{\mathbf{U}}_{ab}\right|_p,\left|\tilde{\mathbf{V}}_{ab}\right|_q\right\}\\ \geq\ 2^{-1/2}\min_{\substack{a,b\in\{i,j,r,s\}\\ a<b}}\min\left\{|\mathbf{U}_{ab}|_p,\left|\hat{\mathbf{U}}_{ab}\right|_p,|\mathbf{V}_{ab}|_q,\left|\hat{\mathbf{V}}_{ab}\right|_q\right\} &\geq 2^{-1/2}\delta_n, \qquad \text{(A.26)}\end{aligned}$$

where the last inequality holds when the event A_{ijrs} occurs. Therefore,

$$\begin{aligned}|R_3| &\leq \frac{1}{n^4}\sum_{i,j,r,s=1}^{n}\frac{1}{2}\sum_{\boldsymbol{\alpha}\in\mathbb{N}^2_{3(p+q)}}|D^{(\boldsymbol{\alpha})}h^*_{d_p,d_q}(\tilde{\mathbf{W}}_{(i,j,r,s)})|\cdot\|\hat{F}_{\mathbf{Z}}-F_{\mathbf{Z}}\|^2_\infty I_{A_{ijrs}}\\ &\leq 4[3(p+q)]^2\sqrt{2}(\sqrt{p}+\sqrt{q}+2)\delta_n^{-1}\|\hat{F}_{\mathbf{Z}}-F_{\mathbf{Z}}\|^2_\infty = o_P(n^{-1/2}), \qquad \text{(A.27)}\end{aligned}$$

since $\delta_n = n^{-1/4}$.

Next, let R_4 be defined in (2.44). We will show that $R_4 = o_P(n^{-1/2})$. By straightforward calculus it can be shown that for any $|\boldsymbol{\alpha}|=1$, $|D^{(\boldsymbol{\alpha})}h_{d_p,d_q}(\tilde{\mathbf{W}}_{(i,j,r,s)})| \leq 4(\sqrt{p}+\sqrt{q})$. Thus

$$\begin{aligned}|R_4| &\leq \frac{1}{n^4}\sum_{i,j,r,s=1}^{n}\sum_{\alpha\in\mathbb{N}_{4(p+q)}}|D^{(\alpha)}h_{d_p,d_q}(\mathbf{W}_{(i,j,r,s)})|\cdot\|\hat{F}_{\mathbf{Z}}-F_{\mathbf{Z}}\|_\infty I_{A^c_{ijrs}}\\ &\leq 16(\sqrt{p}+\sqrt{q})(p+q)\|\hat{F}_{\mathbf{Z}}-F_{\mathbf{Z}}\|_\infty\frac{1}{n^4}\sum_{i,j,r,s=1}^{n} I_{A^c_{ijrs}}.\end{aligned}$$

From (A.24), $\frac{1}{n^4}\sum_{i,j,r,s=1}^n I_{A^c_{ijrs}} = O_P(\delta_n)$. On the other hand, $\|\hat{F}_{\mathbf{Z}}-F_{\mathbf{Z}}\|_\infty = O_P(n^{-1/2})$. Therefore, $R_4 = O_P(n^{-1/2}\delta_n) = o_P(n^{-1/2})$ because $\delta_n = n^{-1/4}$.

A.3 Derivations Needed for the Proof of Theorem 3.2

A.3.1 Proof of (2.49)

Let R_2 be given by (2.39). We will show that, under the assumption of independence, $R_2 = o_p(n^{-1})$. Consider

$$E(R_2^2) = \frac{1}{n^8} \sum_{i,j,r,s=1}^{n} \sum_{i',j',r',s'=1}^{n} E\Big\{ \Big[h_{d_p,d_q}(\mathbf{Z}_{(i,j,r,s)}; \hat{F}_{\mathbf{Z}}) - h_{d_p,d_q}(\mathbf{Z}_{(i,j,r,s)}; F_{\mathbf{Z}})\Big] \cdot \Big[h_{d_p,d_q}(\mathbf{Z}_{i',j',r',s'}; \hat{F}_{\mathbf{Z}}) - h_{d_p,d_q}(\mathbf{Z}_{i',j',r',s'}; F_{\mathbf{Z}})\Big] I_{A^c_{ijrs}} I_{A^c_{i'j'r's'}} \Big\}. \tag{A.28}$$

First it will be shown that the terms in the summation in (A.28) for which $\{i, j, r, s\} \cap \{i', j', r', s'\} = \varnothing$ are all zero. To see this, write each term in the summation as

$$\begin{aligned}
& E\Big\{ \Big[\left|\hat{\mathbf{U}}_{ij}\right|_p \left(\left|\hat{\mathbf{V}}_{rs}\right|_q - 2\left|\hat{\mathbf{V}}_{ir}\right|_q + \left|\hat{\mathbf{V}}_{ij}\right|_q\right) - \left|\mathbf{U}_{ij}\right|_p \left(\left|\mathbf{V}_{rs}\right|_q - 2\left|\mathbf{V}_{ir}\right|_q + \left|\mathbf{V}_{ij}\right|_q\right)\Big] \\
& \quad \Big[\left|\hat{\mathbf{U}}_{i'j'}\right|_p \left(\left|\hat{\mathbf{V}}_{r's'}\right|_q - 2\left|\hat{\mathbf{V}}_{i'r'}\right|_q + \left|\hat{\mathbf{V}}_{i'j'}\right|_q\right) - \left|\mathbf{U}_{i'j'}\right|_p \left(\left|\mathbf{V}_{r's'}\right|_q - 2\left|\mathbf{V}_{i'r'}\right|_q + \left|\mathbf{V}_{i'j'}\right|_q\right)\Big] I_{A^c_{ijrs}} I_{A^c_{i'j'r's'}} \\
&= E\Big\{ \left|\hat{\mathbf{U}}_{ij}\right|_p \left|\hat{\mathbf{U}}_{i'j'}\right|_p \left(\left|\hat{\mathbf{V}}_{rs}\right|_q - 2\left|\hat{\mathbf{V}}_{ir}\right|_q + \left|\hat{\mathbf{V}}_{ij}\right|_q\right) \left(\left|\hat{\mathbf{V}}_{r's'}\right|_q - 2\left|\hat{\mathbf{V}}_{i'r'}\right|_q + \left|\hat{\mathbf{V}}_{i'j'}\right|_q\right) I_{A^c_{ijrs}} I_{A^c_{i'j'r's'}} \Big\} \\
& \quad - E\Big\{ \left|\hat{\mathbf{U}}_{ij}\right|_p \left|\mathbf{U}_{i'j'}\right|_p \left(\left|\hat{\mathbf{V}}_{rs}\right|_q - 2\left|\hat{\mathbf{V}}_{ir}\right|_q + \left|\hat{\mathbf{V}}_{ij}\right|_q\right) \left(\left|\mathbf{V}_{r's'}\right|_q - 2\left|\mathbf{V}_{i'r'}\right|_q + \left|\mathbf{V}_{i'j'}\right|_q\right) I_{A^c_{ijrs}} I_{A^c_{i'j'r's'}} \Big\} \\
& \quad - E\Big\{ \left|\mathbf{U}_{ij}\right|_p \left|\hat{\mathbf{U}}_{i'j'}\right|_p \left(\left|\mathbf{V}_{rs}\right|_q - 2\left|\mathbf{V}_{ir}\right|_q + \left|\mathbf{V}_{ij}\right|_q\right) \left(\left|\hat{\mathbf{V}}_{r's'}\right|_q - 2\left|\hat{\mathbf{V}}_{i'r'}\right|_q + \left|\hat{\mathbf{V}}_{i'j'}\right|_q\right) I_{A^c_{ijrs}} I_{A^c_{i'j'r's'}} \Big\} \\
& \quad + E\Big\{ \left|\mathbf{U}_{ij}\right|_p \left|\mathbf{U}_{i'j'}\right|_p \left(\left|\mathbf{V}_{rs}\right|_q - 2\left|\mathbf{V}_{ir}\right|_q + \left|\mathbf{V}_{ij}\right|_q\right) \left(\left|\mathbf{V}_{r's'}\right|_q - 2\left|\mathbf{V}_{i'r'}\right|_q + \left|\mathbf{V}_{i'j'}\right|_q\right) I_{A^c_{ijrs}} I_{A^c_{i'j'r's'}} \Big\}.
\end{aligned} \tag{A.29}$$

If $\{i, j, r, s\} \cap \{i', j', r', s'\} = \varnothing$, the first term on the right hand side of (A.29) is

$$E\Big\{ E\Big[\left|\hat{\mathbf{U}}_{ij}\right|_p \left|\hat{\mathbf{U}}_{i'j'}\right|_p \left(\left|\hat{\mathbf{V}}_{rs}\right|_q - 2\left|\hat{\mathbf{V}}_{ir}\right|_q + \left|\hat{\mathbf{V}}_{ij}\right|_q\right) \left(\left|\hat{\mathbf{V}}_{r's'}\right|_q - 2\left|\hat{\mathbf{V}}_{i'r'}\right|_q + \left|\hat{\mathbf{V}}_{i'j'}\right|_q\right) I_{A^c_{ijrs}} I_{A^c_{i'j'r's'}}$$

$$
\begin{aligned}
&\qquad \left|\hat{F}_{\mathbf{Z}}, \mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_r, \mathbf{X}_s, \mathbf{Z}_{i'}, \mathbf{Z}_{j'}, \mathbf{Z}_{r'}, \mathbf{Z}_{s'}\right]\Bigg\} \\
&= E\Bigg\{E\Bigg[\left(\left|\hat{\mathbf{V}}_{rs}\right|_q - 2\left|\hat{\mathbf{V}}_{ir}\right|_q + \left|\hat{\mathbf{V}}_{ij}\right|_q\right) I_{A^c_{ijrs}} \Bigg|\hat{F}_{\mathbf{Z}}, \mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_r, \mathbf{X}_s, \mathbf{Z}_{i'}, \mathbf{Z}_{j'}, \mathbf{Z}_{r'}, \mathbf{Z}_{s'}\Bigg] \\
&\qquad \left|\hat{\mathbf{U}}_{ij}\right|_p \left|\hat{\mathbf{U}}_{i'j'}\right|_p \left(\left|\hat{\mathbf{V}}_{r's'}\right|_q - 2\left|\hat{\mathbf{V}}_{i'r'}\right|_q + \left|\hat{\mathbf{V}}_{i'j'}\right|_q\right) I_{A^c_{i'j'r's'}}\Bigg\} \\
&= 0,
\end{aligned}
$$

since, by symmetry, the conditional expectations of $\left|\hat{\mathbf{V}}_{rs}\right|_q I_{A^c_{ijrs}}$, $\left|\hat{\mathbf{V}}_{ir}\right|_q I_{A^c_{ijrs}}$ and $\left|\hat{\mathbf{V}}_{ij}\right|_q I_{A^c_{ijrs}}$ are equal. That the other terms on the right hand side of (A.29) are zero if $\{i,j,r,s\} \cap \{i',j',r',s'\} = \varnothing$ can be shown by similar arguments. Thus the number of non-zero terms in (A.28) is at most $n^4 \cdot 4 \cdot 4n^3 = 16n^7$. Next use (A.22) and (A.23) to write

$$
\begin{aligned}
\left|\left|\hat{\mathbf{U}}_{ij}\right|_p \left|\hat{\mathbf{V}}_{rs}\right|_q - |\mathbf{U}_{ij}|_p |\mathbf{V}_{rs}|_q\right| &= \left|\hat{\mathbf{U}}_{ij}\right|_p \left|\left|\hat{\mathbf{V}}_{rs}\right|_q - |\mathbf{V}_{rs}|_q\right| + \left|\left|\hat{\mathbf{U}}_{ij}\right|_p - |\mathbf{U}_{ij}|_p\right| |\mathbf{V}_{rs}|_q \\
&\leq 2\sqrt{pq}\|\hat{F}_{\mathbf{Z}} - F_{\mathbf{Z}}\|_\infty
\end{aligned}
$$

for any $i,j,r,s = 1,\ldots,n$. Moreover,

$$
\begin{aligned}
& E\left[I_{A^c_{ijrs}}\right] \\
\leq & \sum_{\substack{a_1,a_2 \in \{i,j,r,s\} \\ a_1 < a_2}} \Bigg[P\left(|\mathbf{U}_{a_1a_2}|_p \leq \delta_n\right) + P\left(\left|\hat{\mathbf{U}}_{a_1a_2}\right|_p \leq \delta_n\right) \\
& \qquad + P\left(|\mathbf{V}_{a_1a_2}|_p \leq \delta_n\right) + P\left(\left|\hat{\mathbf{V}}_{a_1a_2}\right|_p \leq \delta_n\right)\Bigg] \\
\leq & \ 96\delta_n \frac{n}{n-1},
\end{aligned}
$$

where the last inequality follows by Lemma A.1.3. Therefore, by Cauchy-Schwartz inequality and the result in (A.20),

$$
\begin{aligned}
E(R_2^2) &\leq n^{-8} \cdot 16n^7 \cdot 4pqE\left[\|\hat{F}_{\mathbf{Z}} - F_{\mathbf{Z}}\|_\infty^2 I_{A^c_{ijrs}}\right] \\
&\leq 64pqn^{-1}\left[E\|\hat{F}_{\mathbf{Z}} - F_{\mathbf{Z}}\|_\infty^4\right]^{1/2}\left[E\left(I_{A^c_{ijrs}}\right)\right]^{1/2} \\
&\leq 64pqn^{-1} \cdot \frac{C_{p,q}}{n}\sqrt{96\delta_n \frac{n}{n-1}},
\end{aligned}
$$

which shows that $R_2 = o_P(n^{-1})$.

A.3.2 Proof of (2.53)

Let R_3 be defined in (2.50). To show that $R_3 = o_P(n^{-1})$ we will use the same reparametrization of the function h_{d_p,d_q} that was employed in Section A.2.2. Thus, we define $\mathbf{W}_{ij} = \mathbf{W}_i - \mathbf{W}_j$, $\hat{\mathbf{W}}_{ij} = \hat{\mathbf{W}}_i - \hat{\mathbf{W}}_j$, $\mathbf{W}_{ijrs} = (\mathbf{W}_{ij}, \mathbf{W}_{ir}, \mathbf{W}_{rs})$ and $\hat{\mathbf{W}}_{ijrs} = (\hat{\mathbf{W}}_{ij}, \hat{\mathbf{W}}_{ir}, \hat{\mathbf{W}}_{rs})$, and consider the function $h^*_{d_p,d_q}$ defined in (A.25).

Then by Taylor expansion with $M = 3$,

$$\begin{aligned}
&h^*_{d_p,d_q}(\hat{\mathbf{W}}_{ijrs}) - h^*_{d_p,d_q}(\mathbf{W}_{ijrs}) \\
&= \sum_{m=1}^{2} \frac{1}{m!} \sum_{\boldsymbol{\alpha} \in \mathbb{N}^m_{3(p+q)}} D^{(\boldsymbol{\alpha})} h^*_{d_p,d_q}(\mathbf{W}_{ijrs}) \left[\hat{\mathbf{W}}_{ijrs} - \mathbf{W}_{ijrs}\right]^{(\boldsymbol{\alpha})} \\
&\quad + \frac{1}{6} \sum_{\boldsymbol{\alpha} \in \mathbb{N}^3_{3(p+q)}} D^{(\boldsymbol{\alpha})} h^*_{d_p,d_q}(\tilde{\mathbf{W}}_{ijrs}) \left[\hat{\mathbf{W}}_{ijrs} - \mathbf{W}_{ijrs}\right]^{(\boldsymbol{\alpha})}
\end{aligned}$$

for some $\tilde{\mathbf{W}}_{ijrs} = (\tilde{\mathbf{W}}_{ij}, \tilde{\mathbf{W}}_{ir}, \tilde{\mathbf{W}}_{rs}) \in [\mathbf{W}_{ij}, \hat{\mathbf{W}}_{ij}] \times [\mathbf{W}_{ir}, \hat{\mathbf{W}}_{ir}] \times [\mathbf{W}_{rs}, \hat{\mathbf{W}}_{rs}]$. It can be shown by a change of variables that

$$\begin{aligned}
&\sum_{m=1}^{2} \frac{1}{m!} \sum_{\boldsymbol{\alpha} \in \mathbb{N}^m_{3(p+q)}} D^{(\boldsymbol{\alpha})} h^*_{d_p,d_q}(\mathbf{W}_{ijrs}) \left[\hat{\mathbf{W}}_{ijrs} - \mathbf{W}_{ijrs}\right]^{(\boldsymbol{\alpha})} \\
&= \sum_{m=1}^{2} \frac{1}{m!} \sum_{\boldsymbol{\alpha} \in \mathbb{N}^m_{4(p+q)}} D^{(\boldsymbol{\alpha})} h_{d_p,d_q}(\mathbf{W}_{(i,j,r,s)}) \left[\hat{\mathbf{W}}_{(i,j,r,s)} - \mathbf{W}_{(i,j,r,s)}\right]^{(\alpha)}.
\end{aligned}$$

Thus, an equivalent expression of the remainder term R_3 is

$$R_3 = \frac{1}{n^4} \sum_{i,j,r,s=1}^{n} \frac{1}{6} \sum_{\boldsymbol{\alpha} \in \mathbb{N}^3_{3(p+q)}} D^{(\boldsymbol{\alpha})} h^*_{d_p,d_q}(\tilde{\mathbf{W}}_{ijrs}) \left[\hat{\mathbf{W}}_{ijrs} - \mathbf{W}_{ijrs}\right]^{(\boldsymbol{\alpha})} I_{A_{ijrs}} \tag{A.30}$$

It can be shown by straightforward calculus that for any $|\boldsymbol{\alpha}| = 3$

$$|D^{(\boldsymbol{\alpha})} h^*_{d_p,d_q}(\tilde{\mathbf{W}}_{ijrs})| \leq \frac{12(\sqrt{p} + \sqrt{q} + 2)}{\left(\min\left\{\left|\tilde{\mathbf{U}}_{ij}\right|_p, \left|\tilde{\mathbf{V}}_{ij}\right|_q, \left|\tilde{\mathbf{V}}_{ir}\right|_q, \left|\tilde{\mathbf{V}}_{rs}\right|_q\right\}\right)^2} \tag{A.31}$$

where $(\tilde{\mathbf{U}}_{ab}, \tilde{\mathbf{V}}_{ab}) = (\tilde{\mathbf{W}}_{ab})$ for any $a_1, a_2 \in \{i, j, r, s\}$. Using the inequality in (A.26) which holds when the event A_{ijrs} occurs, we have

$$
\begin{aligned}
|R_3| &\le \frac{1}{n^4}\sum_{i,j,r,s=1}^{n}\frac{1}{6}\sum_{\boldsymbol{\alpha}\in\mathbb{N}^3_{3(p+q)}}|D^{(\boldsymbol{\alpha})}h_{d_p,d_q}(\tilde{\mathbf{W}}_{(i,j,r,s)})|\cdot\|\hat{F}_{\mathbf{Z}}-F_{\mathbf{Z}}\|^3_{\infty}I_{A_{ijrs}} \\
&\le 12[3(p+q)]^3\sqrt{2}(\sqrt{p}+\sqrt{q}+2)\delta_n^{-2}\|\hat{F}_{\mathbf{Z}}-F_{\mathbf{Z}}\|^3_{\infty}=o_P(n^{-1}),
\end{aligned}
$$

since $\delta_n = n^{-1/8}$.

A.3.3 Proof of (2.56)

Let R_4 be given by (2.54), and use the argument in the proof of Lemma A.1.1 to write

$$
R_4=\frac{1}{n^4}\sum_{i,j,r,s=1}^{n}\sum_{m=1}^{2}\frac{1}{m!}\sum_{\boldsymbol{\alpha}\in\mathbb{N}^m_{4(p+q)}}D^{(\boldsymbol{\alpha})}h_{d_p,d_q}(\mathbf{W}_{(i,j,r,s)})\left[\hat{\mathbf{W}}_{(i,j,r,s)}-\mathbf{W}_{(i,j,r,s)}\right]^{(\boldsymbol{\alpha})}(I_{B_{ijrs}}-I_{A_{ijrs}})\,. \tag{A.32}
$$

It will be convenient to partition the index sets $\mathbb{N}^m_{4(p+q)}$, $m=1,2$, as follows. Define

$$
\begin{aligned}
\mathcal{I}_{\mathbf{X}} &= \{(i_1-1)(p+q)+i_2 : i_1=1,\ldots,4,\ i_2=1,\ldots,p\}\,, \\
\mathcal{I}_{\mathbf{Y}} &= \{(i_1-1)(p+q)+i_2 : i_1=1,\ldots,4,\ i_2=p+1,\ldots,p+q\}\,,
\end{aligned}
$$

and note that

$$
\begin{aligned}
\mathbb{N}_{4(p+q)} &= \mathcal{I}_{\mathbf{X}}\cup\mathcal{I}_{\mathbf{Y}}, \\
\mathbb{N}^2_{4(p+q)} &= (\mathcal{I}_{\mathbf{X}}\times\mathcal{I}_{\mathbf{X}})\cup(\mathcal{I}_{\mathbf{X}}\times\mathcal{I}_{\mathbf{Y}})\cup(\mathcal{I}_{\mathbf{Y}}\times\mathcal{I}_{\mathbf{X}})\cup(\mathcal{I}_{\mathbf{Y}}\times\mathcal{I}_{\mathbf{Y}}).
\end{aligned}
$$

Recalling that $h_{d_p,d_q}(\mathbf{W}_{(i,j,r,s)})=d_p(\mathbf{U}_i,\mathbf{U}_j)\,(d_p(\mathbf{V}_r,\mathbf{V}_s)-2d_p(\mathbf{V}_i,\mathbf{V}_r)+d_p(\mathbf{V}_i,\mathbf{V}_j))$, the expression for R_4 in (A.32) can be written as

$$
R_4=R_{41}+R_{42}+R_{43}+R_{44}+R_{45}\ , \tag{A.33}
$$

where R_{41} corresponds to $\mathcal{I}_X$ and is given by

$$
\begin{aligned}
R_{41} \;=\; &\frac{1}{n^4}\sum_{i,j,r,s=1}^{n}\sum_{\alpha\in\mathbb{N}_{2p}}D^{(\alpha)}d_p(\mathbf{U}_i,\mathbf{U}_j)\,(d_q(\mathbf{V}_r,\mathbf{V}_s)-2d_q(\mathbf{V}_i,\mathbf{V}_r)+d_q(\mathbf{V}_i,\mathbf{V}_j)) \\
&\cdot\left[\hat{\mathbf{U}}_{(i,j)}-\mathbf{U}_{(i,j)}\right]^{(\alpha)}(I_{B_{ijrs}}-I_{A_{ijrs}}),
\end{aligned}
$$

R_{42} corresponds to $\mathcal{I}_Y$ and is given by

$$
\begin{aligned}
R_{42} \;=\; \frac{1}{n^4}\sum_{i,j,r,s=1}^{n} \Bigg\{ & \sum_{\alpha\in\mathbb{N}_{2q}} d_p(\mathbf{U}_i,\mathbf{U}_j)D^{(\alpha)}d_q(\mathbf{V}_r,\mathbf{V}_s)\left[\hat{\mathbf{V}}_{(r,s)}-\mathbf{V}_{(r,s)}\right]^{(\alpha)} \\
& -2\sum_{\alpha\in\mathbb{N}_{2q}} d_p(\mathbf{U}_i,\mathbf{U}_j)D^{(\boldsymbol{\alpha})}d_q(\mathbf{V}_i,\mathbf{V}_r)\left[\hat{\mathbf{V}}_{(i,r)}-\mathbf{V}_{(i,r)}\right]^{(\alpha)} \\
& +\sum_{\alpha\in\mathbb{N}_{2q}} d_p(\mathbf{U}_i,\mathbf{U}_j)D^{(\boldsymbol{\alpha})}d_q(\mathbf{V}_i,\mathbf{V}_j)\left[\hat{\mathbf{V}}_{(i,j)}-\mathbf{V}_{(i,j)}\right]^{(\alpha)} \Bigg\}(I_{B_{ijrs}}-I_{A_{ijrs}}),
\end{aligned}
$$

R_{43} corresponds to $\mathcal{I}_X \times \mathcal{I}_X$ and is given by

$$
\begin{aligned}
R_{43} \;=\; \frac{1}{n^4}\sum_{i,j,r,s=1}^{n} \frac{1}{2}\sum_{\boldsymbol{\alpha}\in\mathbb{N}_{2p}^2} & D^{(\boldsymbol{\alpha})}d_p(\mathbf{U}_i,\mathbf{U}_j)\left(d_q(\mathbf{V}_r,\mathbf{V}_s)-2d_q(\mathbf{V}_i,\mathbf{V}_r)+d_q(\mathbf{V}_i,\mathbf{V}_j)\right) \\
& \cdot\left[\hat{\mathbf{U}}_{(i,j)}-\mathbf{U}_{(i,j)}\right]^{(\boldsymbol{\alpha})}(I_{B_{ijrs}}-I_{A_{ijrs}}),
\end{aligned}
$$

R_{44} corresponds to $\mathcal{I}_Y \times \mathcal{I}_Y$ and is given by

$$
\begin{aligned}
R_{44} \;=\; \frac{1}{n^4}\sum_{i,j,r,s=1}^{n} \frac{1}{2}\Bigg\{ & \sum_{\boldsymbol{\alpha}\in\mathbb{N}_{2q}^2} d_p(\mathbf{U}_i,\mathbf{U}_j)D^{(\boldsymbol{\alpha})}d_q(\mathbf{V}_r,\mathbf{V}_s)\left[\hat{\mathbf{V}}_{(r,s)}-\mathbf{V}_{(r,s)}\right]^{(\boldsymbol{\alpha})} \\
& -2\sum_{\boldsymbol{\alpha}\in\mathbb{N}_{2q}^2} d_p(\mathbf{U}_i,\mathbf{U}_j)D^{(\boldsymbol{\alpha})}d_q(\mathbf{V}_i,\mathbf{V}_r)\left[\hat{\mathbf{V}}_{(i,r)}-\mathbf{V}_{(i,r)}\right]^{(\boldsymbol{\alpha})} \\
& +\sum_{\boldsymbol{\alpha}\in\mathbb{N}_{2q}^2} d_p(\mathbf{U}_i,\mathbf{U}_j)D^{(\boldsymbol{\alpha})}d_q(\mathbf{V}_i,\mathbf{V}_j)\left[\hat{\mathbf{V}}_{(i,j)}-\mathbf{V}_{(i,j)}\right]^{(\boldsymbol{\alpha})} \Bigg\}(I_{B_{ijrs}}-I_{A_{ijrs}}),
\end{aligned}
$$

and R_{45} corresponds to $\mathcal{I}_X \times \mathcal{I}_Y$ and $\mathcal{I}_Y \times \mathcal{I}_X$ and is given by

$$
\begin{aligned}
R_{45} =& \frac{1}{n^4}\sum_{i,j,r,s=1}^{n}\sum_{\alpha\in\mathbb{N}_{2p}}\sum_{\beta\in\mathbb{N}_{2q}} D^{(\alpha)}d_p(\mathbf{U}_i,\mathbf{U}_j)D^{(\beta)}d_q(\mathbf{V}_r,\mathbf{V}_s) \\
& \qquad \cdot\left[\hat{\mathbf{U}}_{(i,j)}-\mathbf{U}_{(i,j)}\right]^{(\alpha)}\left[\hat{\mathbf{V}}_{(r,s)}-\mathbf{V}_{(r,s)}\right]^{(\beta)}(I_{B_{ijrs}}-I_{A_{ijrs}}) \\
& -2\frac{1}{n^4}\sum_{i,j,r,s=1}^{n}\sum_{\alpha\in\mathbb{N}_{2p}}\sum_{\beta\in\mathbb{N}_{2q}} D^{(\alpha)}d_p(\mathbf{U}_i,\mathbf{U}_j)D^{(\beta)}d_q(\mathbf{V}_i,\mathbf{V}_r) \\
& \qquad \cdot\left[\hat{\mathbf{U}}_{(i,j)}-\mathbf{U}_{(i,j)}\right]^{(\alpha)}\left[\hat{\mathbf{V}}_{(i,r)}-\mathbf{V}_{(i,r)}\right]^{(\beta)}(I_{B_{ijrs}}-I_{A_{ijrs}}) \\
& +\frac{1}{n^4}\sum_{i,j,r,s=1}^{n}\sum_{\alpha\in\mathbb{N}_{2p}}\sum_{\beta\in\mathbb{N}_{2q}} D^{(\alpha)}d_p(\mathbf{U}_i,\mathbf{U}_j)D^{(\beta)}d_q(\mathbf{V}_i,\mathbf{V}_j)
\end{aligned}
$$

$$\cdot \left[\hat{\mathbf{U}}_{(i,j)} - \mathbf{U}_{(i,j)}\right]^{(\alpha)} \left[\hat{\mathbf{V}}_{(i,j)} - \mathbf{V}_{(i,j)}\right]^{(\beta)} (I_{B_{ijrs}} - I_{A_{ijrs}}).$$

We will show that R_{41}, R_{42}, R_{43}, R_{44} and R_{45} are all $o_P(n^{-1})$.

Consider $E(R_{41}^2)$. Under the null hypothesis that $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$,

$$E[R_{41}^2] = \frac{1}{n^8} \sum_{i,j,r,s=1}^{n} \sum_{i',j',r',s'=1}^{n} \sum_{\alpha \in \mathbb{N}_{2p}} \sum_{\alpha' \in \mathbb{N}_{2p}}$$
$$E\Bigg\{ D^{(\alpha)} d_p(\mathbf{U}_i, \mathbf{U}_j) D^{(\alpha')} d_p(\mathbf{U}_{i'}, \mathbf{U}_{j'}) \left[\hat{\mathbf{U}}_{(i,j)} - \mathbf{U}_{(i,j)}\right]^{(\alpha)} \left[\hat{\mathbf{U}}_{(i',j')} - \mathbf{U}_{(i',j')}\right]^{(\alpha)}$$
$$(d_q(\mathbf{V}_{r'}, \mathbf{V}_{s'}) - 2d_q(\mathbf{V}_{i'}, \mathbf{V}_{r'}) + d_q(\mathbf{V}_{i'}, \mathbf{V}_{j'}))\,(I_{B_{i'j'r's'}} - I_{A_{i'j'r's'}})$$
$$(d_q(\mathbf{V}_r, \mathbf{V}_s) - 2d_q(\mathbf{V}_i, \mathbf{V}_r) + d_q(\mathbf{V}_i, \mathbf{V}_j))\,(I_{B_{ijrs}} - I_{A_{ijrs}}) \Bigg\}. \qquad \text{(A.34)}$$

Notice that if $\{i, j, r, s\} \cap \{i', j', r', s'\} = \varnothing$, the expectation term in (A.34) is 0 because

$$E\Bigg\{ (d_q(\mathbf{V}_r, \mathbf{V}_s) - 2d_q(\mathbf{V}_i, \mathbf{V}_r) + d_q(\mathbf{V}_i, \mathbf{V}_j))\,(I_{B_{ijrs}} - I_{A_{ijrs}}) \,\Bigg|\, \hat{F}_{\mathbf{Z}}, \mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_r, \mathbf{X}_s, \mathbf{Z}_{i'}, \mathbf{Z}_{j'}, \mathbf{Z}_{r'}, \mathbf{Z}_{s'} \Bigg\} = 0.$$

Therefore, the number of non-zero terms is at most $n^4 \cdot 4n^3 \cdot 2p \cdot 2p = 16n^7p^2$. Next it can be shown by straightforward calculus that for any $|\boldsymbol{\alpha}| = 1$,

$$|D^{(\boldsymbol{\alpha})} d_p(\mathbf{U}_i, \mathbf{U}_j)| \leq 1.$$

It follows that each expectation in (A.34) is bounded by

$$(2\sqrt{q})^2 E\left[\|\hat{F}_{\mathbf{Z}} - F_{\mathbf{Z}}\|_\infty^2 I_{A_{ijrs}^c}\right]$$
$$\leq 4q \left[E\|\hat{F}_{\mathbf{Z}} - F_{\mathbf{Z}}\|_\infty^4\right]^{1/2} \left[E\left(I_{A_{ijrs}^c}\right)\right]^{1/2} \leq 4q \cdot \frac{C_{p,q}}{n} \sqrt{96\delta_n \frac{n}{n-1}}$$

and hence we obtain

$$E(R_{41}^2) \leq n^{-8} \cdot 16n^7p^2 \cdot 4q \cdot \frac{C_{p,q}}{n} \sqrt{96\delta_n \frac{n}{n-1}}.$$

This shows that $R_{41} = o_P(n^{-1})$. By the same argument, $R_{42} = o_P(n^{-1})$ because after permutation of the indexes (set (r, s, i, j) instead of (i, j, r, s) for the first summand, and (i, r, j, s) instead of (i, j, r, s) for the second summand), R_{42} can be written as

$$R_{42} = \frac{1}{n^4} \sum_{i,j,r,s=1}^{n} \sum_{\alpha \in \mathbb{N}_{2q}} D^{(\alpha)} d_q(\mathbf{V}_i, \mathbf{V}_j) \left(d_p(\mathbf{U}_r, \mathbf{U}_s) - 2d_p(\mathbf{U}_i, \mathbf{U}_r) + d_p(\mathbf{U}_i, \mathbf{U}_j) \right) \cdot \left[\hat{\mathbf{V}}_{(i,j)} - \mathbf{V}_{(i,j)} \right]^{(\alpha)} (I_{B_{ijrs}} - I_{A_{ijrs}}).$$

Next, consider $E(R_{43}^2)$. Under the null hypothesis that $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$,

$$\begin{aligned} E[R_{43}^2] = \frac{1}{n^8} \frac{1}{4} \sum_{i,j,r,s=1}^{n} \sum_{i',j',r',s'=1}^{n} \sum_{\boldsymbol{\alpha} \in \mathbb{N}_{2p}^2} \sum_{\boldsymbol{\alpha}' \in \mathbb{N}_{2p}^2} \\ E\Big\{ D^{(\boldsymbol{\alpha})} d_p(\mathbf{U}_i, \mathbf{U}_j) D^{(\boldsymbol{\alpha}')} d_p(\mathbf{U}_{i'}, \mathbf{U}_{j'}) \left[\hat{\mathbf{U}}_{(i,j)} - \mathbf{U}_{(i,j)} \right]^{(\boldsymbol{\alpha})} \left[\hat{\mathbf{U}}_{(i',j')} - \mathbf{U}_{(i',j')} \right]^{(\boldsymbol{\alpha})} \\ \left(d_q(\mathbf{V}_{r'}, \mathbf{V}_{s'}) - 2d_q(\mathbf{V}_{i'}, \mathbf{V}_{r'}) + d_q(\mathbf{V}_{i'}, \mathbf{V}_{j'}) \right) (I_{B_{i'j'r's'}} - I_{A_{i'j'r's'}}) \\ \left(d_q(\mathbf{V}_r, \mathbf{V}_s) - 2d_q(\mathbf{V}_i, \mathbf{V}_r) + d_q(\mathbf{V}_i, \mathbf{V}_j) \right) (I_{B_{ijrs}} - I_{A_{ijrs}}) \Big\}. \end{aligned} \tag{A.35}$$

Notice that if $\{i, j, r, s\} \cap \{i', j', r', s'\} = \varnothing$, the expectation term in (A.35) is 0 because

$$E\Big\{ \left(d_q(\mathbf{V}_r, \mathbf{V}_s) - 2d_q(\mathbf{V}_i, \mathbf{V}_r) + d_q(\mathbf{V}_i, \mathbf{V}_j) \right) (I_{B_{ijrs}} - I_{A_{ijrs}}) \Big| \hat{F}_{\mathbf{Z}}, \mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_r, \mathbf{X}_s, \mathbf{Z}_{i'}, \mathbf{Z}_{j'}, \mathbf{Z}_{r'}, \mathbf{Z}_{s'} \Big\} = 0.$$

Therefore, the number of non-zero terms is at most $n^4 \cdot 4n^3 \cdot 2p \cdot 2p = 16n^7p^2$. Next it can be shown by straightforward calculus that for any $|\boldsymbol{\alpha}| = 2$,

$$|D^{(\boldsymbol{\alpha})} d_p(\mathbf{U}_i, \mathbf{U}_j)| \leq \frac{1}{|\mathbf{U}_{ij}|},$$

Thus each expectation in (A.35) is bounded by

$$\begin{aligned} &(2\sqrt{q})^2 \delta_n^{-2} E\left[\|\hat{F}_{\mathbf{Z}} - F_{\mathbf{Z}}\|_\infty^4 I_{A_{ijrs}^c} \right] \\ \leq\ & 4q\delta_n^{-2} \left[E\|\hat{F}_{\mathbf{Z}} - F_{\mathbf{Z}}\|_\infty^8 \right]^{1/2} \left[E\left(I_{A_{ijrs}^c} \right) \right]^{1/2} \leq 4q\delta_n^{-2} \cdot \frac{C_{p,q}}{n^2} \sqrt{96\delta_n \frac{n}{n-1}} \end{aligned}$$

Therefore,

$$E(R_{43}^2) \le \frac{1}{n^8}\frac{1}{4} \cdot 16n^7p^2 \cdot 4q\delta_n^{-2} \cdot \frac{C_{p,q}}{n}\sqrt{96\delta_n \frac{n}{n-1}},$$

which shows that $R_{43} = o_P(n^{-1})$. By the same argument, $R_{44} = o_P(n^{-1})$ because after permutation of the indexes, R_{44} can be written as

$$R_{44} = \frac{1}{n^4}\sum_{i,j,r,s=1}^{n} \frac{1}{2}\sum_{\boldsymbol{\alpha}\in\mathbb{N}_{2q}^2} D^{(\boldsymbol{\alpha})} d_q(\mathbf{V}_i, \mathbf{V}_j)\left(d_p(\mathbf{U}_r, \mathbf{U}_s) - 2d_p(\mathbf{U}_i, \mathbf{U}_r) + d_p(\mathbf{U}_i, \mathbf{U}_j)\right) \cdot \left[\hat{\mathbf{V}}_{(i,j)} - \mathbf{V}_{(i,j)}\right]^{(\boldsymbol{\alpha})} (I_{B_{ijrs}} - I_{A_{ijrs}}),$$

Finally, consider R_{45}.

$$E|R_{45}| \le 3 \cdot \frac{1}{n^4} \cdot n^4 \cdot (2p)(2q) \cdot E\left[\|\hat{F}_{\mathbf{Z}} - F_{\mathbf{Z}}\|_\infty^2 I_{A_{ijrs}^c}\right] \le 12pq\frac{C_{p,q}}{n}\sqrt{96\delta_n\frac{n}{n-1}}$$

which shows that $R_{45} = o_P(n^{-1})$. This completes the proof.

A.4 Derivations Needed for the Proof of Theorem 2.4.1

A.4.1 Proof of (2.66)

The expressions for the $\zeta_{d_p,d_q,j}$, $j = 1, \ldots, 5$, functions that appear in (2.66) are

$$\begin{aligned}
\zeta_{d_p,d_q,1}(\mathbf{Z}_{(1,\ldots,6)}) &= \sum_{\alpha\in\mathbb{N}_{2p}} D^{(\alpha)} d_p(\mathbf{U}_1, \mathbf{U}_2)\left(d_q(\mathbf{V}_3, \mathbf{V}_4) - 2d_q(\mathbf{V}_1, \mathbf{V}_3) + d_q(\mathbf{V}_1, \mathbf{V}_2)\right) \\
&\quad \cdot \left[\vec{I}_{\mathbf{X}_{(5,5)}}(\mathbf{X}_{(1,2)}) - \mathbf{U}_{(1,2)}\right]^{(\alpha)} I_{B_{1234}}, \\
\zeta_{d_p,d_q,2}(\mathbf{Z}_{(1,\ldots,6)}) &= d_p(\mathbf{U}_1, \mathbf{U}_2)\sum_{\alpha\in\mathbb{N}_{4q}} D^{(\alpha)}\left(d_q(\mathbf{V}_3, \mathbf{V}_4) - 2d_q(\mathbf{V}_1, \mathbf{V}_3) + d_q(\mathbf{V}_1, \mathbf{V}_2)\right) \\
&\quad \cdot \left[\vec{I}_{\mathbf{Y}_{(5,5,5,5)}}(\mathbf{Y}_{(1,2,3,4)}) - \mathbf{V}_{(1,2,3,4)}\right]^{(\alpha)} I_{B_{1234}}, \\
\zeta_{d_p,d_q,3}(\mathbf{Z}_{(1,\ldots,6)}) &= \frac{1}{2}\sum_{\boldsymbol{\alpha}\in\mathbb{N}_{2p}^2} D^{(\boldsymbol{\alpha})} d_p(\mathbf{U}_1, \mathbf{U}_2)\left(d_q(\mathbf{V}_3, \mathbf{V}_4) - 2d_q(\mathbf{V}_1, \mathbf{V}_3) + d_q(\mathbf{V}_1, \mathbf{V}_2)\right) \\
&\quad \cdot \left[\vec{I}_{\mathbf{X}_{(5,5)}}(\mathbf{X}_{(1,2)}) - \mathbf{U}_{(1,2)}\right]^{(\boldsymbol{\alpha})} I_{B_{1234}},
\end{aligned}$$

$$\begin{aligned}\zeta_{d_p,d_q,4}(\mathbf{Z}_{(1,\ldots,6)}) &= \frac{1}{2}d_p(\mathbf{U}_1,\mathbf{U}_2)\sum_{\boldsymbol{\alpha}\in\mathbb{N}_{4p}^2} D^{(\boldsymbol{\alpha})}\left(d_q(\mathbf{V}_3,\mathbf{V}_4)-2d_q(\mathbf{V}_1,\mathbf{V}_3)+d_q(\mathbf{V}_1,\mathbf{V}_2)\right)\\ &\quad\cdot\left[\vec{I}_{\mathbf{Y}_{(5,5,5,5)}}(\mathbf{Y}_{(1,2,3,4)})-\mathbf{V}_{(1,2,3,4)}\right]^{(\boldsymbol{\alpha})} I_{B_{1234}},\\ \zeta_{d_p,d_q,5}(\mathbf{Z}_{(1,\ldots,6)}) &= \sum_{\alpha\in\mathbb{N}_{2p}}\sum_{\beta\in\mathbb{N}_{4q}} D^{(\alpha)}d_p(\mathbf{U}_1,\mathbf{U}_2)D^{(\beta)}\left(d_q(\mathbf{V}_3,\mathbf{V}_4)-2d_q(\mathbf{V}_1,\mathbf{V}_3)+d_q(\mathbf{V}_1,\mathbf{V}_2)\right)\\ &\quad\cdot\left[\vec{I}_{\mathbf{X}_{(5,5)}}(\mathbf{X}_{(1,2)})-\mathbf{U}_{(1,2)}\right]^{(\alpha)}\left[\vec{I}_{\mathbf{Y}_{(6,6,6,6)}}(\mathbf{Y}_{(1,2,3,4)})-\mathbf{V}_{(1,2,3,4)}\right]^{(\beta)} I_{B_{1234}},\end{aligned}$$

We will only show (2.66) for the case of $j=1$ and $j=5$; the cases for $j=2,\ 3$ and 4 follow by arguments similar to those for the case $j=1$.

Write $I_{B_{1234}} = I_{C_{1234}} I_{D_{1234}}$, where

$$C_{1234}=\left\{\min_{\substack{a_1,a_2\in\{1,2,3,4\}\\ a_1<a_2}}|\mathbf{U}_{a_1a_2}|_p>0\right\},\quad D_{1234}=\left\{\min_{\substack{a_1,a_2\in\{1,2,3,4\}\\ a_1<a_2}}|\mathbf{V}_{a_1a_2}|_q>0\right\},\tag{A.36}$$

For the case of $j=1$, notice that under $\mathbf{X}\perp\!\!\!\perp\mathbf{Y}$, for any $(i_1,i_2)\in I_6^2$,

$$\begin{aligned}&E\left[\zeta_{d_p,d_q,1}(\mathbf{Z}_{(1,\ldots,6)})\middle|\mathbf{Z}_{i_1}=\mathbf{z},\mathbf{Z}_{i_2}=\mathbf{z}\right]\\ &=\sum_{\alpha\in\mathbb{N}_{2p}}E\left\{D^{(\alpha)}d_p(\mathbf{U}_1,\mathbf{U}_2)\left[\vec{I}_{\mathbf{X}_{(5,5)}}(\mathbf{X}_{(1,2)})-\mathbf{U}_{(1,2)}\right]^{(\alpha)}I_{C_{1234}}\middle|\mathbf{X}_{i_1}=\mathbf{x},\mathbf{X}_{i_2}=\mathbf{x}\right\}\\ &\qquad\cdot E\left\{\left(d_q(\mathbf{V}_3,\mathbf{V}_4)-2d_q(\mathbf{V}_1,\mathbf{V}_3)+d_q(\mathbf{V}_1,\mathbf{V}_2)\right)I_{D_{1234}}\middle|\mathbf{Y}_{i_1}=\mathbf{y},\mathbf{Y}_{i_2}=\mathbf{y}\right\},\end{aligned}\tag{A.37}$$

If $5\notin\{i_1,i_2\}$, each summand in (A.37) is 0 because

$$\sum_{\alpha\in\mathbb{N}_{2p}}E\left\{D^{(\alpha)}d_p(\mathbf{U}_1,\mathbf{U}_2)\left[\vec{I}_{\mathbf{X}_{(5,5)}}(\mathbf{X}_{(1,2)})-\mathbf{U}_{(1,2)}\right]^{(\alpha)}I_{C_{1234}}\middle|\mathbf{X}_{i_1}=\mathbf{x},\mathbf{X}_{i_2}=\mathbf{x}\right\}=0\ ,\tag{A.38}$$

which follows by the fact that

$$E\left\{\vec{I}_{\mathbf{X}_{(5,5)}}(\mathbf{X}_{(1,2)})-\mathbf{U}_{(1,2)}\middle|\mathbf{X}_1,\mathbf{X}_2,\mathbf{X}_3,\mathbf{X}_4,\mathbf{X}_6\right\}=0\ .$$

If $\{i_1, i_2\} \cap \{1,2,3,4\} = \varnothing$, each summand in (A.37) is 0 because

$$E\left\{ \left(d_q(\mathbf{V}_3,\mathbf{V}_4) - 2d_q(\mathbf{V}_1,\mathbf{V}_3) + d_q(\mathbf{V}_1,\mathbf{V}_2)\right) I_{D_{1234}} \middle| \mathbf{Y}_{i_1} = \mathbf{y}, \mathbf{Y}_{i_2} = \mathbf{y} \right\} = 0. \tag{A.39}$$

Therefore, each summand in (A.37) is nonzero only if one of the two indexes i_1 and i_2 equals 5 and the other equals $1, 2, 3$ or 4. Thus

$$\begin{aligned}
&\sum_{(i_1,i_2)\in I_6^2} E\left[\zeta_{k,l,1}(\mathbf{Z}_{(1,\dots,6)}) \middle| \mathbf{Z}_{i_1} = \mathbf{z}, \mathbf{Z}_{i_2} = \mathbf{z}\right] = 2\sum_{i=1}^{4} E\left[\zeta_{k,l,1}(\mathbf{Z}_{(1,\dots,6)}) \middle| \mathbf{Z}_i = \mathbf{z}, \mathbf{Z}_5 = \mathbf{z}\right] \\
&= 2\sum_{i=1}^{4}\sum_{\alpha\in\mathbb{N}_{2p}} E\left\{ D^{(\alpha)} d_p(\mathbf{U}_1,\mathbf{U}_2) \left[\vec{I}_{\mathbf{X}_{(5,5)}}(\mathbf{X}_{(1,2)}) - \mathbf{U}_{(1,2)}\right]^{(\alpha)} I_{C_{1234}} \middle| \mathbf{X}_i = \mathbf{x}, \mathbf{X}_5 = \mathbf{x}\right\} \\
&\quad \cdot E\left\{ \left(d_q(\mathbf{V}_3,\mathbf{V}_4) - 2d_q(\mathbf{V}_1,\mathbf{V}_3) + d_q(\mathbf{V}_1,\mathbf{V}_2)\right) I_{D_{1234}} \middle| \mathbf{Y}_i = \mathbf{y}\right\}
\end{aligned}$$

For simplicity, define

$$A_i = \sum_{\alpha\in\mathbb{N}_{2p}} E\left\{ D^{(\alpha)} d_p(\mathbf{U}_1,\mathbf{U}_2) \left[\vec{I}_{\mathbf{X}_{(5,5)}}(\mathbf{X}_{(1,2)}) - \mathbf{U}_{(1,2)}\right]^{(\alpha)} I_{C_{1234}} \middle| \mathbf{X}_i = \mathbf{x}, \mathbf{X}_5 = \mathbf{x}\right\} \tag{A.40}$$

$$B_i = E\left\{ \left(d_q(\mathbf{V}_3,\mathbf{V}_4) - 2d_q(\mathbf{V}_1,\mathbf{V}_3) + d_q(\mathbf{V}_1,\mathbf{V}_2)\right) I_{D_{1234}} \middle| \mathbf{Y}_i = \mathbf{y}\right\} \tag{A.41}$$

Then by symmetry, it is easy to verify that $A_1 = A_2$, $A_3 = A_4$, $B_1 = -B_2$, $B_3 = -B_4$. Therefore,

$$2\sum_{(i_1,i_2)\in I_6^2} E\left[\zeta_{k,l,1}(\mathbf{Z}_{(1,\dots,6)}) \middle| \mathbf{Z}_{i_1} = \mathbf{z}, \mathbf{Z}_{i_2} = \mathbf{z}\right] = 2\sum_{i=1}^{4} A_i B_i = 0 \tag{A.42}$$

For the case of $j = 5$, notice that under $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$, for any $(i_1, i_2) \in I_6^2$,

$$\begin{aligned}
&E\left[\zeta_{d_p,d_q,5}(\mathbf{Z}_{(1,\dots,6)}) \middle| \mathbf{Z}_{i_1} = \mathbf{z}, \mathbf{Z}_{i_2} = \mathbf{z}\right] \\
&= \sum_{\alpha\in\mathbb{N}_{2p}} E\left\{ D^{(\alpha)} d_p(\mathbf{U}_1,\mathbf{U}_2) \left[\vec{I}_{\mathbf{X}_{(5,5)}}(\mathbf{X}_{(1,2)}) - \mathbf{U}_{(1,2)}\right]^{(\alpha)} I_{C_{1234}} \middle| \mathbf{X}_{i_1} = \mathbf{x}, \mathbf{X}_{i_2} = \mathbf{x}\right\} \\
&\quad \cdot \sum_{\beta\in\mathbb{N}_{4q}} E\Big\{ D^{(\beta)} \left(d_q(\mathbf{V}_3,\mathbf{V}_4) - 2d_q(\mathbf{V}_1,\mathbf{V}_3) + d_q(\mathbf{V}_1,\mathbf{V}_2)\right)
\end{aligned} \tag{A.43}$$

$$\cdot \left[\vec{I}_{\mathbf{Y}_{(6,6,6,6)}}(\mathbf{Y}_{(1,2,3,4)}) - \mathbf{V}_{(1,2,3,4)}\right]^{(\beta)} I_{D_{1234}} \Bigg| \mathbf{Y}_{i_1} = \mathbf{y}, \mathbf{Y}_{i_2} = \mathbf{y} \Bigg\}. \quad \text{(A.44)}$$

If $5 \notin \{i_1, i_2\}$ or $6 \notin \{i_1, i_2\}$, each summand in (A.37) is 0 by similar arguments as those to show (A.38). Therefore, each summand in (A.37) is nonzero only if one of $\{i_1, i_2\} = \{5, 6\}$. Thus

$$\sum_{(i_1,i_2)\in I_6^2} E\left[\zeta_{k,l,5}(\mathbf{Z}_{(1,\ldots,6)}) \middle| \mathbf{Z}_{i_1} = \mathbf{z}, \mathbf{Z}_{i_2} = \mathbf{z}\right] = 2E\left[\zeta_{k,l,5}(\mathbf{Z}_{(1,\ldots,6)}) \middle| \mathbf{Z}_5 = \mathbf{z}, \mathbf{Z}_6 = \mathbf{z}\right]$$

$$= 2\sum_{\alpha\in\mathbb{N}_{2p}} E\Bigg\{ D^{(\alpha)} d_p(\mathbf{U}_1, \mathbf{U}_2) \left[\vec{I}_{\mathbf{X}_{(5,5)}}(\mathbf{X}_{(1,2)}) - \mathbf{U}_{(1,2)}\right]^{(\alpha)} I_{C_{1234}} \Bigg| \mathbf{X}_5 = \mathbf{x}, \mathbf{X}_6 = \mathbf{x} \Bigg\}$$

$$\cdot \sum_{\beta\in\mathbb{N}_{4q}} E\Bigg\{ D^{(\beta)} \left(d_q(\mathbf{V}_3, \mathbf{V}_4) - 2d_q(\mathbf{V}_1, \mathbf{V}_3) + d_q(\mathbf{V}_1, \mathbf{V}_2)\right)$$

$$\cdot \left[\vec{I}_{\mathbf{Y}_{(6,6,6,6)}}(\mathbf{Y}_{(1,2,3,4)}) - \mathbf{V}_{(1,2,3,4)}\right]^{(\beta)} I_{D_{1234}} \Bigg| \mathbf{Y}_5 = \mathbf{y}, \mathbf{Y}_6 = \mathbf{y} \Bigg\}$$

$$= 0$$

because

$$\sum_{\beta\in\mathbb{N}_{4q}} E\Bigg\{ D^{(\beta)} d_q(\mathbf{V}_3, \mathbf{V}_4) \left[\vec{I}_{\mathbf{Y}_{(6,6,6,6)}}(\mathbf{Y}_{(1,2,3,4)}) - \mathbf{V}_{(1,2,3,4)}\right]^{(\beta)} I_{D_{1234}} \Bigg| \mathbf{Y}_5 = \mathbf{y}, \mathbf{Y}_6 = \mathbf{y} \Bigg\}$$

$$= \sum_{\beta\in\mathbb{N}_{4q}} E\Bigg\{ D^{(\beta)} d_q(\mathbf{V}_1, \mathbf{V}_3) \left[\vec{I}_{\mathbf{Y}_{(6,6,6,6)}}(\mathbf{Y}_{(1,2,3,4)}) - \mathbf{V}_{(1,2,3,4)}\right]^{(\beta)} I_{D_{1234}} \Bigg| \mathbf{Y}_5 = \mathbf{y}, \mathbf{Y}_6 = \mathbf{y} \Bigg\}$$

$$= \sum_{\beta\in\mathbb{N}_{4q}} E\Bigg\{ D^{(\beta)} d_q(\mathbf{V}_1, \mathbf{V}_2) \left[\vec{I}_{\mathbf{Y}_{(6,6,6,6)}}(\mathbf{Y}_{(1,2,3,4)}) - \mathbf{V}_{(1,2,3,4)}\right]^{(\beta)} I_{D_{1234}} \Bigg| \mathbf{Y}_5 = \mathbf{y}, \mathbf{Y}_6 = \mathbf{y} \Bigg\}.$$

On Wed, May 24, 2017 at 2:49 PM, roycelin3 <roycelin3@gmail.com> wrote:

$$\zeta_{d_p,d_q,1}(\mathbf{Z}_{(1,\ldots,6)})$$
$$= \sum_{\alpha\in\mathbb{N}_{2p}} D^{(\alpha)} d_p(\mathbf{U}_1, \mathbf{U}_2) \left(d_q(\mathbf{V}_3, \mathbf{V}_4) - 2d_q(\mathbf{V}_1, \mathbf{V}_3) + d_q(\mathbf{V}_1, \mathbf{V}_2)\right)$$
$$\cdot \left[\vec{I}_{\mathbf{X}_{(5,5)}}(\mathbf{X}_{(1,2)}) - \mathbf{U}_{(1,2)}\right]^{(\alpha)} I_{B_{1234}},$$

$$\zeta_{d_p,d_q,2}(\mathbf{Z}_{(1,\ldots,6)})$$

$$= d_p(\mathbf{U}_1, \mathbf{U}_2) \sum_{\alpha \in \mathbb{N}_{4q}} D^{(\alpha)} \left(d_q(\mathbf{V}_3, \mathbf{V}_4) - 2d_q(\mathbf{V}_1, \mathbf{V}_3) + d_q(\mathbf{V}_1, \mathbf{V}_2)\right)$$
$$\cdot \left[\vec{I}_{\mathbf{Y}_{(5,5,5,5)}}(\mathbf{Y}_{(1,2,3,4)}) - \mathbf{V}_{(1,2,3,4)}\right]^{(\alpha)} I_{B_{1234}},$$

$$\zeta_{d_p,d_q,3}(\mathbf{Z}_{(1,\ldots,6)})$$
$$= \frac{1}{2} \sum_{\boldsymbol{\alpha} \in \mathbb{N}_{2p}^2} D^{(\boldsymbol{\alpha})} d_p(\mathbf{U}_1, \mathbf{U}_2) \left(d_q(\mathbf{V}_3, \mathbf{V}_4) - 2d_q(\mathbf{V}_1, \mathbf{V}_3) + d_q(\mathbf{V}_1, \mathbf{V}_2)\right)$$
$$\cdot \left[\vec{I}_{\mathbf{X}_{(5,5)}}(\mathbf{X}_{(1,2)}) - \mathbf{U}_{(1,2)}\right]^{(\boldsymbol{\alpha})} I_{B_{1234}},$$

$$\zeta_{d_p,d_q,4}(\mathbf{Z}_{(1,\ldots,6)})$$
$$= \frac{1}{2} d_p(\mathbf{U}_1, \mathbf{U}_2) \sum_{\boldsymbol{\alpha} \in \mathbb{N}_{4p}^2} D^{(\boldsymbol{\alpha})} \left(d_q(\mathbf{V}_3, \mathbf{V}_4) - 2d_q(\mathbf{V}_1, \mathbf{V}_3) + d_q(\mathbf{V}_1, \mathbf{V}_2)\right)$$
$$\cdot \left[\vec{I}_{\mathbf{Y}_{(5,5,5,5)}}(\mathbf{Y}_{(1,2,3,4)}) - \mathbf{V}_{(1,2,3,4)}\right]^{(\boldsymbol{\alpha})} I_{B_{1234}},$$

$$\zeta_{d_p,d_q,5}(\mathbf{Z}_{(1,\ldots,6)})$$
$$= \sum_{\alpha \in \mathbb{N}_{2p}} \sum_{\beta \in \mathbb{N}_{4q}} D^{(\alpha)} d_p(\mathbf{U}_1, \mathbf{U}_2) D^{(\beta)} \left(d_q(\mathbf{V}_3, \mathbf{V}_4) - 2d_q(\mathbf{V}_1, \mathbf{V}_3) + d_q(\mathbf{V}_1, \mathbf{V}_2)\right)$$
$$\cdot \left[\vec{I}_{\mathbf{X}_{(5,5)}}(\mathbf{X}_{(1,2)}) - \mathbf{U}_{(1,2)}\right]^{(\alpha)} \left[\vec{I}_{\mathbf{Y}_{(5,5,5,5)}}(\mathbf{Y}_{(1,2,3,4)}) - \mathbf{V}_{(1,2,3,4)}\right]^{(\beta)} I_{B_{1234}},$$

A.4.2 Proof of (2.67)

For any $(i_1, i_2) \in I_6^2$ we have that, under $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$,

$$E[\zeta_{d_p,d_q,5}(\mathbf{Z}_{(1,\ldots,6)})|\mathbf{Z}_{i_1} = \mathbf{z}, \mathbf{Z}_{i_2} = \mathbf{z}]$$
$$= E[d_p(\mathbf{U}_1, \mathbf{U}_2)|\mathbf{X}_{i_1} = \mathbf{x}, \mathbf{X}_{i_2} = \mathbf{x}] E[d_q(\mathbf{V}_3, \mathbf{V}_4) - 2d_q(\mathbf{V}_1, \mathbf{V}_3) + d_q(\mathbf{V}_1, \mathbf{V}_2)|\mathbf{Y}_{i_1} = \mathbf{y}, \mathbf{Y}_{i_2} = \mathbf{y}]$$

which is easily seen to be zero if $\{i_1, i_2\} = \{5, 6\}$. Therefore

$$\frac{1}{(6)_2} \sum_{(i_1,i_2) \in I_6^2} E\left[\zeta_{d_p,d_q,5}(\mathbf{Z}_1, \ldots, \mathbf{Z}_6) \middle| \mathbf{Z}_{i_1} = \mathbf{z}, \mathbf{Z}_{i_2} = \mathbf{z}\right]$$
$$= \frac{1}{(6)_2} 2 \sum_{i_1=1}^{4} \sum_{i_2=5}^{6} \left\{ E[d_p(\mathbf{U}_1, \mathbf{U}_2)|\mathbf{X}_{i_1} = \mathbf{x}, \mathbf{X}_{i_2} = \mathbf{x}] \right.$$

$$
\begin{aligned}
&\cdot E[d_q(\mathbf{V}_3,\mathbf{V}_4)-2d_q(\mathbf{V}_1,\mathbf{V}_3)+d_q(\mathbf{V}_1,\mathbf{V}_2)|\mathbf{Y}_{i_1}=\mathbf{y},\mathbf{Y}_{i_2}=\mathbf{y}]\Big\}\\
&+\frac{1}{(6)_2}\sum_{(i_1,i_2)\in I_4^2}\Big\{E[d_p(\mathbf{U}_1,\mathbf{U}_2)|\mathbf{X}_{i_1}=\mathbf{x},\mathbf{X}_{i_2}=\mathbf{x}]\\
&\cdot E[d_q(\mathbf{V}_3,\mathbf{V}_4)-2d_q(\mathbf{V}_1,\mathbf{V}_3)+d_q(\mathbf{V}_1,\mathbf{V}_2)|\mathbf{Y}_{i_1}=\mathbf{y},\mathbf{Y}_{i_2}=\mathbf{y}]\Big\}\\
=&\frac{1}{(6)_2}4\sum_{i_1=1}^{4}\Big\{E[d_p(\mathbf{U}_1,\mathbf{U}_2)|\mathbf{X}_{i_1}=\mathbf{x}]\\
&\cdot E[d_q(\mathbf{V}_3,\mathbf{V}_4)-2d_q(\mathbf{V}_1,\mathbf{V}_3)+d_q(\mathbf{V}_1,\mathbf{V}_2)|\mathbf{Y}_{i_1}=\mathbf{y}]\Big\}\\
&+\frac{1}{(6)_2}\sum_{(i_1,i_2)\in I_4^2}\Big\{E[d_p(\mathbf{U}_1,\mathbf{U}_2)|\mathbf{X}_{i_1}=\mathbf{x},\mathbf{X}_{i_2}=\mathbf{x}]\\
&\cdot E[d_q(\mathbf{V}_3,\mathbf{V}_4)-2d_q(\mathbf{V}_1,\mathbf{V}_3)+d_q(\mathbf{V}_1,\mathbf{V}_2)|\mathbf{Y}_{i_1}=\mathbf{y},\mathbf{Y}_{i_2}=\mathbf{y}]\Big\}
\end{aligned}
$$

It can be easily seen that the first summation is 0 by writing

$$
\begin{aligned}
&\sum_{i_1=1}^{4}\Big\{E[d_p(\mathbf{U}_1,\mathbf{U}_2)|\mathbf{X}_{i_1}=\mathbf{x}]\\
&\cdot E[d_q(\mathbf{V}_3,\mathbf{V}_4)-2d_q(\mathbf{V}_1,\mathbf{V}_3)+d_q(\mathbf{V}_1,\mathbf{V}_2)|\mathbf{Y}_{i_1}=\mathbf{y}]\Big\}=\sum_{i_1=1}^{4}A_iB_i
\end{aligned}
$$

where for $i=1,\ldots,4$,

$$
\begin{aligned}
A_i&=E[d_p(\mathbf{U}_1,\mathbf{U}_2)|\mathbf{X}_{i_1}=\mathbf{x}],\\
B_i&=E[d_q(\mathbf{V}_3,\mathbf{V}_4)-2d_q(\mathbf{V}_1,\mathbf{V}_3)+d_q(\mathbf{V}_1,\mathbf{V}_2)|\mathbf{Y}_{i_1}=\mathbf{y}]
\end{aligned}
$$

and noting that $A_1=A_2$, $A_3=A_4=0$, $B_1=-B_2$.

For the second summation, each term equals to one of the following expression for different cases:

$$
\begin{aligned}
&(i_1,i_2)=(1,2)\text{ or }(2,1): && d_p(\mathbf{u},\mathbf{u})\left\{E[d_q(\mathbf{V}_1,\mathbf{V}_2)]-2E[d_q(\mathbf{v},\mathbf{V}_1)]+d_q(\mathbf{v},\mathbf{v})\right\}\\
&(i_1,i_2)=(1,3)\text{ or }(3,1): && E[d_p(\mathbf{u},\mathbf{U}_1)]\left\{2E[d_q(\mathbf{v},\mathbf{V}_1)]-2d_q(\mathbf{v},\mathbf{v})\right\}\\
&(i_1,i_2)=(1,4)\text{ or }(4,1): && 0\\
&(i_1,i_2)=(2,3)\text{ or }(3,2): && 0
\end{aligned}
$$

$$(i_1, i_2) = (2,4) \text{ or } (4,2): \quad E[d_p(\mathbf{u}, \mathbf{U}_1)]\,\{-2E[d_q(\mathbf{V}_1, \mathbf{V}_2)] + 2E[d_q(\mathbf{v}, \mathbf{V}_1)]\}$$

$$(i_1, i_2) = (3,4) \text{ or } (4,3): \quad E[d_p(\mathbf{U}_1, \mathbf{U}_2)]\,\{E[d_q(\mathbf{V}_1, \mathbf{V}_2)] - 2E[d_q(\mathbf{v}, \mathbf{V}_1)] + d_q(\mathbf{v}, \mathbf{v})\}$$

The proof of (2.67) follows by adding these expressions.

Bibliography

[1] T. M. Apostol. *Mathematical analysis*. Addison-Wesley Reading, MA, 2nds edition, 1974.

[2] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American mathematical society*, 68(3):337–404, 1950.

[3] F. Back and M. Jordan. Kernel independent component analysis. *Journal of Machine Learning Research*, 3:1–48, 2002.

[4] L. Baringhaus and C. Franz. On a new multivariate two-sample test. *Journal of multivariate analysis*, 88(1):190–206, 2004.

[5] N. Dunford and J. T. Schwartz. *Linear operators: Part II: Spectral Theory: Self Adjoint Operators in Hilbert Space*. Interscience Publishers, 1963.

[6] Y. Escoufier. Le traitement des variables vectorielles. *Biometrics*, pages 751–760, 1973.

[7] K. Fukumizu, F. R. Bach, and M. I. Jordan. Dimensionality reduction for supervised learning with reproducing kernel hilbert spaces. *Journal of Machine Learning Research*, 5(Jan):73–99, 2004.

[8] H. Gebelein. Das statistische problem der korrelation als variation und eigenwertproblem und sein zusammenhang mit der ausgleichrechnung. *Angew. Math. Mech.*, 21:364ÃŘ379, 1941.

[9] A. Gretton. A simpler condition for consistency of a kernel independence test. *arXiv preprint arXiv:1501.06103*, 2015.

[10] A. Gretton, K. M. Borgwardt, M. Rasch, B. Schölkopf, A. J. Smola, et al. A kernel method for the two-sample-problem. *Advances in neural information processing systems*, 19:513, 2007.

[11] A. Gretton, O. Bousquet, A. Smola, and B. Schölkopf. Measuring statistical dependence with hilbert-schmidt norms. In *International conference on algorithmic learning theory*, pages 63–77. Springer, 2005.

[12] A. Gretton, K. Fukumizu, C. H. Teo, L. Song, B. Schölkopf, and A. J. Smola. A kernel statistical test of independence. In *Advances in neural information processing systems*, pages 585–592, 2007.

[13] A. Gretton, A. J. Smola, O. Bousquet, R. Herbrich, A. Belitski, M. Augath, Y. Murayama, J. Pauls, B. Schölkopf, and N. K. Logothetis. Kernel constrained covariance for dependence measurement. In *AISTATS*, volume 10, pages 112–119, 2005.

[14] M. Hein and O. Bousquet. Kernels, associated structures and generalizations. *Max-Planck-Institut fuer biologische Kybernetik, Technical Report*, 2004.

[15] T. P. Hettmansperger and J. W. McKean. Statistical inference based on ranks. *Psychometrika*, 43(1):69–79, 1978.

[16] W. Hoeffding. A class of statistics with asymptotically normal distribution. *The annals of mathematical statistics*, pages 293–325, 1948.

[17] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28(3/4):321–377, 1936.

[18] M. G. Krein. Hermitian-positive kernels on homogeneous spaces. i. *Amer. Math. Soc. Transl*, 2:69–108, 1963.

[19] A. Lee. U-statistics. *Theory and Practice," Marcel Dekker, New York*, 1990.

[20] N. Papadatos and T. Xifara. A simple method for obtaining the maximal correlation coefficient and related characterizations. *Journal of Multivariate Analysis*, 118:102–114, 2013.

[21] A. Rényi. On measures of dependence. *Acta Mathematica Hungarica*, 10:441–451, 1959.

[22] D. Sejdinovic, B. K. Sriperumbudur, A. Gretton, and K. Fukumizu. Equivalence of distance-based and rkhs-based statistics in hypothesis testing. *The Annals of Statistics*, 41:2263 – 2291, 2013.

[23] R. J. Serfling. *Approximation theorems of mathematical statistics*, volume 162. John Wiley & Sons, 2009.

[24] C. Spearman. The proof and measurement of association between two things. *The American journal of psychology*, 15(1):72–101, 1904.

[25] C. Spearman. 'footrule' for measuring correlation. *British Journal of Psychology, 1904-1920*, 2(1):89–108, 1906.

[26] B. K. Sriperumbudur, A. Gretton, K. Fukumizu, B. Schölkopf, and G. R. Lanckriet. Hilbert space embeddings and metrics on probability measures. *Journal of Machine Learning Research*, 11(Apr):1517–1561, 2010.

[27] I. Steinwart. On the influence of the kernel on the consistency of support vector machines. *Journal of machine learning research*, 2(Nov):67–93, 2001.

[28] G. J. Székely and M. L. Rizzo. Testing for equal distributions in high dimension. *InterStat*, 5:1–6, 2004.

[29] G. J. Szekely and M. L. Rizzo. Brownian distance covariance. *The Annals of Applied Statistics*, 3:1236–1265, 2009.

[30] G. J. Székely, M. L. Rizzo, N. K. Bakirov, et al. Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, 35(6):2769–2794, 2007.

Vita

Junli Lin

Education:

- Ph.D. Statistics, Pennsylvania State University, 2017
  Dissertation Title: Copula Versions of RKHS-Based and Distance-Based Statistics
  Advisor: Dr. Michael Akritas

- B.S. Mathematics of Academic Talent Program, Tsinghua University, 2011

Teaching Experience:

Graduate Instructor in Pennsylvania State University:

- STAT 480: Introduction to SAS, Summer 2014
- STAT 401: Experimental Methods, Summer 2015
- STAT 301: Statistical Analysis I, Fall 2016
- STAT 301: Statistical Analysis I, Spring 2017

Awards and Fellowships:

- Distinction in the applied part of the Master qualifier exam, 2012
- Third Prize for Model Student of Academic Records, 2010
- Third Prize for Model Student of Academic Records, 2009
- National Encouragement Scholarship, 2008

Presentations:

- SMAC seminar, Pennsylvania State University, 11/2016
- 3rd Conference of the International Society for Non-Parametric Statistics (ISNPS), Avignon, France, 06/2016
- Joint Statistical Meetings, Seattle, U.S., 05/2015