

Conditional Distance Variance and Correlation

Barnabás Póczos BAPOCZOS@CS.CMU.EDU and **Jeff Schneider** SCHNEIDE@CS.CMU.EDU
Carnegie Mellon University

Abstract

Recently a new dependence measure, the distance correlation, has been proposed to measure the dependence between continuous random variables. A nice property of this measure is that it can be consistently estimated with the empirical average of the products of certain distances between the sample points. Here we generalize this quantity to measure the conditional dependence between random variables, and show that this can also be estimated with a statistic using a weighted empirical average of the products of distances between the sample points. We demonstrate the applicability of the estimators with numerical experiments on real and simulated data sets.

1. Introduction

Measuring conditional dependencies is crucial in many applications of machine learning and statistics. There are many problems where we want to know how the dependence of two random variables changes if we observe other random variables. Correlated random variables might become independent when we observe a third random variable, and the opposite situation is also possible when independent variables become dependent after observing other random variables.

The estimation of certain dependence and conditional dependence measures are easy in a few cases: For example, (i) when the random variables have discrete distributions with finite possible values, (ii) when there is a known simple relationship between them (e.g. a linear model describes their behavior), or (iii) if they have joint distributions that belong to a parametric family that is easy to estimate (e.g. Gauss distributions). In this paper we consider the more challenging nonparametric estimation problem when the random variables have continuous distributions, and we do not know have any other information about them.

A simple and elegant method, *the distance variance and correlation*, has been introduced recently to measure dependence between continuous vector-valued random variables (Székely et al., 2007). These quantities can be efficiently estimated using only certain Euclidean distances between the sample points. The *main contribution* of this paper is to generalize these quantities for the *conditional* case and provide consistent estimators for them. We will define the conditional covariance $\mathcal{V}(X, Y|Z)$ and conditional correlation $\mathcal{R}(X, Y|Z)$ quantities between X , Y , and Z random variables. These quantities are nonnegative and achieve zero if and only if X and Y are conditionally independent given Z . We will also show that $\mathcal{R}(X, Y|Z) \leq 1$, and it achieves this upper bound when there is a conditional linear relationship between X and Y given Z . Our goal is to consistently estimate these quantities. We will see in the subsequent sections that the problem is quite challenging. In the ‘unconditional’ case, it is enough to plug the empirical characteristic functions into the definition of distance covariance $V(X, Y)$. In the conditional case, however, it is not obvious if there are simple estimations for the conditional characteristic functions. Moreover, even if we could replace the

$\mathcal{F}_{X,Y|Z=z}(t, s)$ conditional characteristic functions with an appropriate estimate $\hat{\mathcal{F}}_{X,Y|Z=z}(t, s)$, we will still need to calculate its multidimensional integral with respect to $w(t, s)dt ds$, where $w(t, s)$ is a weight function. Even if the joint, conditional, and marginal densities of the random vectors were known, the problem would be still challenging because we need to calculate this integral. We also want to develop an estimator that is simple and uses only Euclidean distances between the sample points. In this paper we will show that all of these requirements are possible.

The derived estimators have several potential applications. In many scientific areas (e.g., epidemiology, psychology, pharmacoinformatics, econometrics) it is crucial to detect confounding variables and not to infer causation from apparent correlations (Pearl, 1998; Montgomery, 2005; Baumgarten and Olsen, 2004). Conditional dependencies play a central role in Bayesian network research as well. The structure learning algorithms of Bayesian nets need to fit a graph to a set of random variables in such a way that this graph satisfies the local Markov property, that is, each variable is conditionally independent of its non-descendants given its parent variables. To assess how well the local Markov property is satisfied in the fitted graph, we need to estimate the conditional dependencies of the nodes.

In our daily life we can also easily encounter examples when people infer causation from observed correlations. Many times, however, there is a hidden factor that is responsible for this correlation. A famous example from introductory statistical books is that there is nonzero correlation between the reading skills of children and their shoe size. Here the underlying common factor is obviously the age. We can find several similar examples in our daily life and in many ancient legends too. According to a northern European legend, the stork is responsible for delivering babies to parents. Indeed, one can show that highly statistically significant correlation exists between stork populations and human birth rates across Europe (Matthews, 2000). Conditional dependence estimators can help us to reveal the underlying hidden factors.

The paper is organized as follows. In the next section we summarize some related work. In Section 3 we review the definitions and properties of distance based variance, covariance, and correlation. We generalize them for the conditional case and formally introduce our estimation problem in Section 4. The proposed estimator is derived in Section 5; here we also discuss some of its theoretical properties and prove the consistency of the estimator. We demonstrate the consistency of the estimator by numerical experiments in Section 6 and also show the applicability of the estimator in real data sets. We finish the paper with a short discussion and draw conclusions. The proofs of the lemmas and theorems can be found in the Appendix, and there we also provide a few more numerical experiments.

Notation: Let $\mathcal{B}(x, R)$ denote a closed ball around $x \in \mathbb{R}^d$ with radius R , and let $\text{Vol}(\mathcal{B}(x, R)) = \lambda R^d$ be its volume, where λ stands for the volume of a d -dimensional unit ball. For brevity, $\mathcal{B}$ will denote the unit ball centered at $0 \in \mathbb{R}^d$, and $\mathcal{B}^c = \mathbb{R}^d \setminus \mathcal{B}$ stands for the complement of $\mathcal{B}$. For $x, t \in \mathbb{R}^d$, $\langle t, x \rangle$ denotes their inner products, and $|x|$ stands for the Euclidean norm of x . We use $X_n \rightarrow_p X$ to denote convergence of random variables in probability. If $y \in \mathbb{R}^{d_y}$, $z \in \mathbb{R}^{d_z}$ are column vectors, then $x = [y; z] \in \mathbb{R}^{d_y+d_z}$ is a column vector with components y and z .

2. Related work

Although the estimation of conditional dependence is a fundamental problem in statistics and machine learning, we know very little about how to estimate it efficiently. Recently, Fukumizu et al. (2008) proposed a new method for estimating conditional dependence based on reproducing kernel Hilbert

spaces (RKHS). There also exist methods for conditional independence tests (see e.g., [Bouezmarni et al. \(2009\)](#); [Su and White \(2008\)](#)); however, the goal of these methods is simply to reject or accept the hypothesis that the random variables are conditionally independent, and their primary goal is not to measure the strength of this dependence.

For the estimation of conditional dependence, we will use distance based statistics between the sample points and generalize the estimator of [Székely et al. \(2007\)](#). There are other estimators of information theoretic quantities that use similar statistics. [Hero and Michel \(1999\)](#) derived a strongly consistent estimator for the Rényi entropy using Euclidean functionals ([Steele, 1997](#); [Yukich, 1998](#)). [Póczos et al. \(2010\)](#) and [Pál et al. \(2010\)](#) combined these ideas with copula methods and proposed tools for Rényi mutual information estimation. [Leonenko et al. \(2008\)](#) and [Goria et al. \(2005\)](#) applied k -nearest neighbor based statistics for Shannon and Rényi- α entropy estimation. Inspired by these results, [Wang et al. \(2009\)](#) and [Pérez-Cruz \(2008\)](#) provided an estimator for the KL-divergence. These estimators use *power weighted* Euclidean distances. In contrast, our method applies the standard Euclidean distances. [Hero et al. \(2002a,b\)](#) also investigated the Rényi divergence estimation problem but assumed that one of the two density functions is known. Recently, [Sricharan et al. \(2010\)](#) proposed k nearest neighbor based methods for estimating non-linear functionals of density. Other interesting nonparametric dependence measures include the kernel mutual information ([Gretton et al., 2003](#)) and the Schweizer-Wolf measure ([Schweizer and Wolff, 1981](#)). Nonetheless, none of these above mentioned papers consider the *conditional dependence* estimation problem.

3. Distance Covariance and Correlation

To be able to define the distance variance and correlation, we will need the following lemma ([Székely and Rizzo, 2005](#)).

Lemma 1 *If $x \in \mathbb{R}^d$ and $0 < \alpha < 2$, then $\int_{\mathbb{R}^d} \frac{1 - \cos\langle t, x \rangle}{|t|^{d+\alpha}} dt = C(d, \alpha)|x|^\alpha$, where $C(d, \alpha) \doteq \frac{2\pi^{d/2}\Gamma(1-\alpha/2)}{\alpha 2^\alpha \Gamma((d+\alpha)/2)}$. The integral is defined in the principal value sense, i.e., $\lim_{\epsilon \rightarrow 0} \int_{\mathbb{R}^d \setminus \{\epsilon B \cup (\mathbb{R}^d \setminus \epsilon^{-1} B)\}} f(t) dt$. When $\alpha = 1$, then $c_d \doteq C(d, 1) = \frac{\pi^{(1+d)/2}}{\Gamma((1+d)/2)}$.*

We will also need the following related lemma.

Lemma 2 *If $x \in \mathbb{R}^d$, then $\int_{|t| < y} \frac{1 - \cos\langle t, x \rangle}{|t|^{d+1}} dt = |x|G(|x|y)$, where $G \leq c_d$ is a bounded continuous function, $\lim_{y \rightarrow \infty} G(y) = c_d$, and $\lim_{y \rightarrow 0} G(y) = 0$.*

The following lemma states that although $|t|^{-d-\alpha}$ is not integrable on $\mathbb{R}^d$, it is integrable on the $\{t : |t| > y\}$ domain for all $y > 0$.

Lemma 3 *Let $0 \leq \alpha$ and $0 < y$. Then $\int_{|t| > y} \frac{1}{|t|^{d+\alpha}} dt < \infty$.*

Now we are ready to define the distance covariance quantity. Let $(X, Y) \sim p_{X,Y}$ be random variables, where $X \in \mathbb{R}^{d_x}$, $Y \in \mathbb{R}^{d_y}$. Suppose that we have N i.i.d. samples drawn from the $p_{X,Y}$ distribution; they are denoted by $\{(X_n; Y_n)\}_{n=1}^N$, where $(X_n; Y_n) \in \mathbb{R}^{d_{xy}}$, $d_{xy} \doteq d_x + d_y$. Using this $\{(X_n; Y_n)\}_{n=1}^N$ sample, our goal is to estimate the distance covariance $\mathcal{V}(X, Y)$, which is defined as follows.

Definition 4 (Distance covariance) *Let $\mathcal{V}(X, Y) \doteq \iint w(t, s) |\mathcal{F}_{X,Y}(t, s) - \mathcal{F}_X(t)\mathcal{F}_Y(s)|^2 dt ds$. Here $w(t, s) \doteq (c_{d_x} c_{d_y} |t|^{1+d_x} |s|^{1+d_y})^{-1}$, and $\mathcal{F}_{X,Y}(t, s) \doteq \mathbb{E}[e^{i\langle t, X \rangle + i\langle s, Y \rangle}]$, $\mathcal{F}_X(t) \doteq \mathbb{E}[e^{i\langle t, X \rangle}]$, $\mathcal{F}_Y(s) \doteq \mathbb{E}[e^{i\langle s, Y \rangle}]$ denote the joint and marginal characteristic functions, respectively.*

$\mathcal{V}(X, Y)$ is a measure of dependence: $\mathcal{V}(X, Y) \geq 0$, and $\mathcal{V}(X, Y) = 0$ if and only if X and Y are independent. Here $w(t, s)$ is a non-integrable, nonnegative weight function. $\mathcal{V}(X, Y)$ measures how far $\mathcal{F}_{X,Y}$ is from $\mathcal{F}_X \mathcal{F}_Y$ with respect to this weight function.

Recently, Székely et al. (2007) proposed an elegant estimator for $\mathcal{V}(X, Y)$. We review this estimator briefly. Let $\mathcal{F}_{X,Y}^N(t, s) \doteq \frac{1}{N} \sum_{n=1}^N e^{i\langle t, X_n \rangle + i\langle s, Y_n \rangle}$, $\mathcal{F}_X^N(t) \doteq \frac{1}{N} \sum_{n=1}^N e^{i\langle t, X_n \rangle}$, $\mathcal{F}_Y^N(s) \doteq \frac{1}{N} \sum_{n=1}^N e^{i\langle s, Y_n \rangle}$ denote the empirical characteristic functions of $p_{X,Y}$, p_X , and p_Y . By plugging them into the definition of $\mathcal{V}(X, Y)$, we arrive at the following estimator:

$$\mathcal{V}_N(X, Y) \doteq \iint w(t, s) |\mathcal{F}_{X,Y}^N(t, s) - \mathcal{F}_X^N(t) \mathcal{F}_Y^N(s)|^2 dt ds. \quad (1)$$

Thanks to the cleverly chosen weights $w(t, s)$ and Lemma 1, Székely et al. (2007) proved that the integral in (1) has a closed form solution, and $\mathcal{V}_N(X, Y)$ can be calculated using the distances between the sample points: $\mathcal{V}_N(X, Y) = \frac{1}{N^2} \sum_{k,l=1}^N A_{k,l} B_{k,l}$, where $A_{k,l} \doteq a_{kl} - \bar{a}_{k\cdot} - \bar{a}_{\cdot l} + \bar{a}_{\cdot\cdot}$, $B_{k,l} \doteq b_{kl} - \bar{b}_{k\cdot} - \bar{b}_{\cdot l} + \bar{b}_{\cdot\cdot}$, $a_{kl} = |X_k - X_l|$, $\bar{a}_{k\cdot} = \frac{1}{N} \sum_{l=1}^N a_{kl}$, $\bar{a}_{\cdot l} = \frac{1}{N} \sum_{k=1}^N a_{kl}$, $\bar{a}_{\cdot\cdot} = \frac{1}{N^2} \sum_{k,l=1}^N a_{kl}$, $b_{kl} = |Y_k - Y_l|$, and similarly define the $\bar{b}_{k\cdot}$, $\bar{b}_{\cdot l}$, $\bar{b}_{\cdot\cdot}$ quantities too.

For brevity, introduce the $\mathcal{V}(X) \doteq \mathcal{V}(X, X)$, $\mathcal{V}_N(X) \doteq \mathcal{V}_N(X, X)$ shorthands. $\mathcal{V}(X)$ is called distance variance, and $\mathcal{V}_N(X)$ is its estimation:

Definition 5 (Distance variance)

$$\mathcal{V}(X) \doteq \iint w(t, s) |\mathcal{F}_{X,X}(t, s) - \mathcal{F}_X(t) \mathcal{F}_X(s)|^2 dt ds.$$

The distance correlation is defined by the distance covariance and distance variance as follows:

Definition 6 (Distance correlation) Let $\mathcal{R}(X, Y) \doteq \frac{\mathcal{V}(X,Y)}{\sqrt{\mathcal{V}(X)\mathcal{V}(Y)}}$ if $\mathcal{V}(X)\mathcal{V}(Y) > 0$, and let $\mathcal{R}(X, Y) \doteq 0$ otherwise. Similarly, $\mathcal{R}_N(X, Y)$, the estimator of $\mathcal{R}(X, Y)$, is defined by $\mathcal{R}_N(X, Y) \doteq \frac{\mathcal{V}_N(X,Y)}{\sqrt{\mathcal{V}_N(X)\mathcal{V}_N(Y)}}$, if $\mathcal{V}_N(X)\mathcal{V}_N(Y) > 0$, and $\mathcal{R}_N(X, Y) \doteq 0$ when $\mathcal{V}_N(X)\mathcal{V}_N(Y) = 0$.

One can prove that if $\mathbb{E}[|X| + |Y|] < \infty$, then $0 \leq \mathcal{R}(X, Y) \leq 1$, and $\mathcal{R}(X, Y) = 0$ if and only if X and Y are independent. If $\mathcal{R}_N(X, Y) = 1$, then there exist a real vector a , a real number b , and an orthogonal matrix C such that $Y = a + bXC$ (Székely et al., 2007).

4. Formal Problem Setup

Now we are ready to formally define the goal of this paper. Our goal is to generalize Definition 5 and Definition 6 for measuring *conditional* dependencies. Let $(X, Y, Z) \sim p_{X,Y,Z}$ be random variables, $X \in \mathbb{R}^{d_x}$, $Y \in \mathbb{R}^{d_y}$, $Z \in \mathbb{R}^{d_z}$. Suppose we have N i.i.d. samples from the distribution of $p_{X,Y,Z}$. They are denoted by $\{(X_n; Y_n; Z_n)\}_{n=1}^N$, where $(X_n; Y_n; Z_n) \in \mathbb{R}^{d_{xyz}}$, $d_{xyz} \doteq d_x + d_y + d_z$. To avoid more cumbersome notations and to simplify the problem somewhat, we will also assume that we have an additional $\tilde{Z}_t$ ($t = 1, 2, \dots, T$) i.i.d. sample from $p_Z(z)$.¹ Using these $\{(X_n; Y_n; Z_n)\}_{n=1}^N$, $\{\tilde{Z}_t\}_{t=1}^T$ samples, our goal is to estimate the conditional distance covariance $\mathcal{V}(X, Y|Z)$, variance $\mathcal{V}(X|Z)$, and correlation $\mathcal{R}(X, Y|Z)$. These quantities are defined as follows.

1. For brevity, we will omit the subscript Z and simply write $p(z)$

Definition 7 (Conditional distance covariance) *Let*

$$\mathcal{V}(X, Y|Z = z) \doteq \iint w(t, s) |\mathcal{F}_{X,Y|Z=z}(t, s) - \mathcal{F}_{X|Z=z}(t)\mathcal{F}_{Y|Z=z}(s)|^2 dt ds,$$

and let $\mathcal{V}(X, Y|Z) \doteq \int p(z)\mathcal{V}(X, Y|Z = z)dz$. Here $\mathcal{F}_{X,Y|Z=z}(t, s) \doteq \mathbb{E}[e^{i\langle t, X \rangle + i\langle s, Y \rangle} | Z = z]$, $\mathcal{F}_{X|Z=z}(t) \doteq \mathbb{E}[e^{i\langle t, X \rangle} | Z = z]$, $\mathcal{F}_{Y|Z=z}(s) \doteq \mathbb{E}[e^{i\langle s, Y \rangle} | Z = z]$ denote the joint and marginal conditional characteristic functions.

The next lemma states that under certain conditions $\mathcal{V}(X, Y|Z = z)$ is well-defined and finite. This is not immediately obvious, since $\iint w(t, s) dt ds = \infty$.

Lemma 8 ($\mathcal{V}(X, Y|Z = z)$ is well-defined) *If $\mathbb{E}[X|Z = z]$, and $\mathbb{E}[Y|Z = z]$ exist and are finite, then $\mathcal{V}(X, Y|Z = z)$ also exists and is finite.*

Definition 9 (Conditional distance variance) *Let*

$$\mathcal{V}(X|Z = z) \doteq \iint w(t, s) \left| \int e^{i\langle t+s, x \rangle} p(x|z) dx - \left[\int e^{i\langle t, x \rangle} p(x|z) dx \right] \left[\int e^{i\langle s, x \rangle} p(x|z) dx \right] \right|^2 dt ds.$$

The conditional distance variance is defined as $\mathcal{V}(X|Z) \doteq \int p(z)\mathcal{V}(X|Z = z)dz$.

Definition 10 (Conditional distance correlation) *If $\mathcal{V}(X|Z = z)\mathcal{V}(Y|Z = z) > 0$, then $\mathcal{R}(X, Y|Z = z) \doteq \frac{\mathcal{V}(X, Y|Z=z)}{\sqrt{\mathcal{V}(X|Z=z)\mathcal{V}(Y|Z=z)}}$. Otherwise, $\mathcal{R}(X, Y|Z = z) \doteq 0$. Let $\mathcal{R}(X, Y|Z) \doteq \int p(z)\mathcal{R}(X, Y|Z = z)dz$.*

Since $0 \leq \mathcal{R}(X, Y) \leq 1$, thus it is easy to see that $0 \leq \mathcal{R}(X, Y|Z = z) \leq 1$ under some slight conditions, and therefore $0 \leq \mathcal{R}(X, Y|Z) \leq 1$. Suppose that $p(z) > 0$ and $\mathcal{V}(X|Z = z)\mathcal{V}(Y|Z = z) > 0$ for all z . In this case, $0 = \mathcal{R}(X, Y|Z) = \mathcal{V}(X, Y|Z)$ if and only if X and Y are conditionally independent given Z .

Our goal is to consistently estimate $\mathcal{V}(X, Y|Z)$, $\mathcal{V}(X|Z)$, and $\mathcal{R}(X, Y|Z)$. We are faced with a couple of problems. In the “unconditional” case, it was enough to plug the empirical characteristic functions into Definition 4. In the conditional case, however, it is not obvious if there are simple estimations for $\mathcal{F}_{X,Y|Z=z}(t, s)$, $\mathcal{F}_{X|Z=z}(t)$, and $\mathcal{F}_{Y|Z=z}(s)$. Even if we could replace the $\mathcal{F}_{X,Y|Z=z}(t, s)$ characteristic function with an estimate $\mathcal{F}_{X,Y|Z=z}^N(t, s)$, it is still not obvious how its multidimensional integral can be calculated with respect to $w(t, s)p(z)dz dt ds$. Similarly, the problem would still be quite challenging if we knew the underlying densities, because we need to efficiently calculate the multidimensional integral with respect to $w(t, s)dt ds$. We also want to develop a simple estimator that uses only distances between the sample points. In the following sections we will show that these requirements are all possible.

5. Estimation

In this section we derive estimators for $\mathcal{V}(X, Y|Z)$, $\mathcal{V}(X, Y|Z = z)$, $\mathcal{V}(X|Z)$, $\mathcal{V}(X|Z = z)$, $\mathcal{R}(X, Y|Z)$, and $\mathcal{R}(X, Y|Z = z)$. By definition, $\mathcal{V}(X, Y|Z) \doteq \int p(z) \iint w(t, s) \Lambda(t, s) dt ds dz$, where

$$\Lambda(t, s) \doteq \left| \iint e^{i\langle t, x \rangle + i\langle s, y \rangle} p(x, y|z) dx dy - \left(\int e^{i\langle t, x \rangle} p(x|z) dx \right) \left(\int e^{i\langle s, y \rangle} p(y|z) dy \right) \right|^2.$$

Assume that we already have a $\hat{p} = \hat{p}_N$ density estimator for density p . For each given $Z = z_0$, thanks to the law of large numbers $\Lambda(t, s)$ can be estimated with the following quantity (see Appendix C for details). The quality of this estimation will be discussed in Lemma 11.

$\Lambda_N(t, s) \doteq$

$$\begin{aligned} & \left\{ \frac{1}{N} \sum_{n=1}^N \cos(\langle t, X_n \rangle + \langle s, Y_n \rangle) \frac{\hat{p}(z_0|X_n, Y_n)}{\hat{p}(z_0)} - \frac{1}{N^2} \left[\sum_{n=1}^N \cos \langle t, X_n \rangle \frac{\hat{p}(z_0|X_n)}{\hat{p}(z_0)} \right] \left[\sum_{n=1}^N \cos \langle s, Y_n \rangle \frac{\hat{p}(z_0|Y_n)}{\hat{p}(z_0)} \right] \right. \\ & + \frac{1}{N^2} \left[\sum_{n=1}^N \sin \langle t, X_n \rangle \frac{\hat{p}(z_0|X_n)}{\hat{p}(z_0)} \right] \left[\sum_{n=1}^N \sin \langle s, Y_n \rangle \frac{\hat{p}(z_0|Y_n)}{\hat{p}(z_0)} \right] \Big\}^2 \\ & + \left\{ \frac{1}{N} \sum_{n=1}^N \sin(\langle t, X_n \rangle + \langle s, Y_n \rangle) \frac{\hat{p}(z_0|X_n, Y_n)}{\hat{p}(z_0)} - \frac{1}{N^2} \left[\sum_{n=1}^N \sin \langle t, X_n \rangle \frac{\hat{p}(z_0|X_n)}{\hat{p}(z_0)} \right] \left[\sum_{n=1}^N \cos \langle s, Y_n \rangle \frac{\hat{p}(z_0|Y_n)}{\hat{p}(z_0)} \right] \right. \\ & \left. - \frac{1}{N^2} \left[\sum_{n=1}^N \cos \langle t, X_n \rangle \frac{\hat{p}(z_0|X_n)}{\hat{p}(z_0)} \right] \left[\sum_{n=1}^N \sin \langle s, Y_n \rangle \frac{\hat{p}(z_0|Y_n)}{\hat{p}(z_0)} \right] \right\}^2. \end{aligned}$$

Introduce the following shorthand notations.

$$\begin{aligned} \tilde{a}_{1,n} &\doteq \cos(\langle t, X_n \rangle + \langle s, Y_n \rangle) \frac{\hat{p}(z_0|X_n, Y_n)}{\hat{p}(z_0)}, & a_1 &\doteq \iint \cos(\langle t, x \rangle + \langle s, y \rangle) \frac{p(z_0, x, y)}{p(z_0)} dx dy, \\ \tilde{b}_{1,n} &\doteq \sin(\langle t, X_n \rangle + \langle s, Y_n \rangle) \frac{\hat{p}(z_0|X_n, Y_n)}{\hat{p}(z_0)}, & b_1 &\doteq \iint \sin(\langle t, x \rangle + \langle s, y \rangle) \frac{p(z_0, x, y)}{p(z_0)} dx dy, \\ \tilde{a}_{2,n} &\doteq \cos \langle t, X_n \rangle \frac{\hat{p}(z_0|X_n)}{\hat{p}(z_0)}, & a_2 &\doteq \int \cos \langle t, x \rangle \frac{p(z_0, x)}{p(z_0)} dx, & \tilde{a}_{3,n} &\doteq \cos \langle s, Y_n \rangle \frac{\hat{p}(z_0|Y_n)}{\hat{p}(z_0)}, \\ a_3 &\doteq \int \cos \langle s, y \rangle \frac{p(z_0, y)}{p(z_0)} dy, & \tilde{a}_{4,n} &\doteq \sin \langle t, X_n \rangle \frac{\hat{p}(z_0|X_n)}{\hat{p}(z_0)}, & a_4 &\doteq \int \sin \langle t, x \rangle \frac{p(z_0, x)}{p(z_0)} dx, \\ \tilde{a}_{5,n} &\doteq \sin \langle s, Y_n \rangle \frac{\hat{p}(z_0|Y_n)}{\hat{p}(z_0)}, & a_5 &\doteq \int \sin \langle s, y \rangle \frac{p(z_0, y)}{p(z_0)} dy \end{aligned}$$

Similarly, let

$$\begin{aligned} \tilde{b}_{2,n} &\doteq \sin \langle t, X_n \rangle \frac{\hat{p}(z_0|X_n)}{\hat{p}(z_0)}, & b_2 &\doteq \int \sin \langle t, x \rangle \frac{p(z_0, x)}{p(z_0)} dx, & \tilde{b}_{3,n} &\doteq \cos \langle s, Y_n \rangle \frac{\hat{p}(z_0|Y_n)}{\hat{p}(z_0)}, \\ b_3 &\doteq \int \cos \langle s, y \rangle \frac{p(z_0, y)}{p(z_0)} dy, & \tilde{b}_{4,n} &\doteq \cos \langle t, X_n \rangle \frac{\hat{p}(z_0|X_n)}{\hat{p}(z_0)}, & b_4 &\doteq \int \cos \langle t, x \rangle \frac{p(z_0, x)}{p(z_0)} dx, \\ \tilde{b}_{5,n} &\doteq \sin \langle s, Y_n \rangle \frac{\hat{p}(z_0|Y_n)}{\hat{p}(z_0)}, & b_5 &\doteq \int \sin \langle s, y \rangle \frac{p(z_0, y)}{p(z_0)} dy, \end{aligned}$$

and let $a_{i,N} \doteq \sum_{i=1}^N \tilde{a}_{i,n}/N$ and $b_{i,N} \doteq \sum_{i=1}^N \tilde{b}_{i,n}/N$. The following lemma states that $\Lambda_N(t, s)$ is a good approximation of $\Lambda(t, s)$ when N is large enough.

Lemma 11 (Bounding $|\Lambda_N(t, s) - \Lambda(t, s)|$) *Assume that there exist $\sigma^2 < \infty$ and $K < \infty$ such that for all t, s, i it holds that $\max(\text{Var}[\tilde{a}_{i,n}], \text{Var}[\tilde{b}_{i,n}]) \leq \sigma^2 < \infty$, and $\max(a_i, b_i) \leq K < \infty$. Let $\hat{p}_N$ be a density estimator such that almost surely $\lim_{N \rightarrow \infty} \sup_x |\hat{p}_N(x) - p(x)| = 0$. Then for all $\epsilon, \delta > 0$, there exists $N_0 = N_0(\epsilon, \delta)$ such that $\Pr(\max_{t,s} |\Lambda(t, s) - \Lambda_N(t, s)| < \epsilon) \geq 1 - \delta$ if $N > N_0$.*

Proof The proof can be found in Appendix D. ■

For brevity, let $\alpha_1 \doteq [\cos \langle t, X_k - X_l \rangle \cos \langle s, Y_k - Y_l \rangle - \sin \langle t, X_k - X_l \rangle \sin \langle s, Y_k - Y_l \rangle]$, $\alpha_2 \doteq [\cos \langle t, X_n - X_k \rangle \cos \langle s, Y_n - Y_l \rangle - \sin \langle t, X_n - X_k \rangle \sin \langle s, Y_n - Y_l \rangle]$. After some algebraic manipulation (see the the Appendix E for details), we have that $\Lambda_N = S_1 + S_2 - 2S_3$, where

$$S_1 \doteq \frac{1}{N^2} \sum_{k,l} \frac{\hat{p}(z_0|X_k, Y_k) \hat{p}(z_0|X_l, Y_l)}{\hat{p}^2(z_0)} \alpha_1,$$

$$S_2 \doteq \frac{1}{N^4} \left[\sum_{k,l}^N \frac{\hat{p}(z_0|X_k)\hat{p}(z_0|X_l)}{\hat{p}^2(z_0)} \cos\langle t, X_k - X_l \rangle \right] \left[\sum_{m,n}^N \frac{\hat{p}(z_0|Y_n)\hat{p}(z_0|Y_m)}{\hat{p}^2(z_0)} \cos\langle s, Y_m - Y_n \rangle \right],$$

$$S_3 \doteq \frac{1}{N^3} \sum_{n,k,l}^N \frac{\hat{p}(z_0|X_n, Y_n)\hat{p}(z_0|X_k)\hat{p}(z_0|Y_l)}{\hat{p}^3(z_0)} \alpha_2.$$

The next step is to calculate $\Omega_N(z_0) \doteq \iint w(t, s) \Lambda_N dt ds$, which is an approximation of $\mathcal{V}(X, Y|Z = z_0)$. The problem is that $w(t, s)$ is not integrable on the whole $\mathbb{R}^{d_x} \times \mathbb{R}^{d_y}$ domain, thus it is not even obvious if this quantity exists. To deal with this issue, we will integrate it on the smaller $D(\psi) \doteq \{(t, s) : |t| > \psi, |s| > \psi\}$ domain and then study its behavior as $\psi \rightarrow 0$. Therefore, we have that $\Omega_N(z_0) = \lim_{\psi \rightarrow 0} \Omega_{\psi, N}(z_0)$, where

$$\Omega_{\psi, N}(z_0) \doteq \iint_{D(\psi)} w(t, s) \Lambda_N dt ds = \iint_{D(\psi)} w(t, s) (S_1 + S_2 - 2S_3) dt ds. \quad (2)$$

To calculate this quantity, we exploit the facts that $\iint_{D(\psi)} w(t, s) \sin(\langle t, u \rangle) \sin(\langle s, v \rangle) dt ds = 0$, $\cos u \cos v = 1 + (1 - \cos u)(1 - \cos v) - (1 - \cos u) - (1 - \cos v)$, and apply Lemma 2. One can see that the following statement holds.

Lemma 12

$$\iint_{D(\psi)} w(t, s) (1 - \cos\langle t, x \rangle) (1 - \cos\langle s, y \rangle) = |x| (1 - G(|x|\psi)/c_d) |y| (1 - G(|y|\psi)/c_d),$$

where c_d and G were defined in Lemma 1, and Lemma 2.

For brevity, introduce the following shorthands: $\bar{p}(k, l) \doteq \hat{p}(z_0|X_k, Y_k)\hat{p}(z_0|X_l, Y_l)/\hat{p}^2(z_0)$. Now, we have that

Lemma 13

$$\begin{aligned} & \iint_{D(\psi)} w(t, s) S_1 dt ds \\ &= V_{1, N} + \frac{1}{N^2} \sum_{k,l}^N \bar{p}(k, l) |X_k - X_l| |Y_k - Y_l| (1 - G(|X_k - X_l|\psi)/c_d) (1 - G(|Y_k - Y_l|\psi)/c_d), \\ & \iint_{D(\psi)} w(t, s) S_2 dt ds = V_{2, N} + \frac{1}{N^4} \sum_{k,l,m,n}^N \frac{\hat{p}(z_0|X_k)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|X_l)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|Y_n)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|Y_m)}{\hat{p}(z_0)} \\ & \quad \times |X_k - X_l| |Y_m - Y_n| (1 - G(|X_k - X_l|\psi)/c_d) (1 - G(|Y_m - Y_n|\psi)/c_d), \\ & \iint_{D(\psi)} w(t, s) S_3 dt ds = V_{3, N} + \frac{1}{N^3} \sum_{n,k,l}^N \frac{\hat{p}(z_0|X_n, Y_n)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|X_k)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|Y_l)}{\hat{p}(z_0)} \\ & \quad \times |X_n - X_k| |Y_n - Y_l| (1 - G(|X_n - X_k|\psi)/c_d) (1 - G(|Y_n - Y_l|\psi)/c_d). \end{aligned}$$

For each ψ , $\lim_{N \rightarrow \infty} V_{1, N} + V_{2, N} - 2V_{3, N} = 0$ almost surely, and thus asymptotically the contribution of the sum of $V_{1, N}$, $V_{2, N}$ and $V_{3, N}$ terms is negligible.

Proof The proof can be found in Appendix F. ■

Let $\mathcal{V}_\psi(X, Y|Z = z) \doteq \iint_{D(\psi)} w(t, s) |\mathcal{F}_{X,Y|Z=z}(t, s) - \mathcal{F}_{X|Z=z}(t) \mathcal{F}_{Y|Z=z}(s)|^2 dt ds$. Note that by definition $\lim_{\psi \rightarrow 0} \mathcal{V}_\psi(X, Y|Z = z) = \mathcal{V}(X, Y|Z = z)$. Below we show that $\mathcal{V}_\psi(X, Y|Z = z)$ can be consistently estimated for each ψ .

Lemma 14 *Assume that conditions of Lemma 11 hold. Then for all $\epsilon, \delta, \psi > 0$, there exists $N_0 = N_0(\epsilon, \delta, \psi)$ such that if $N > N_0$, then*

$$\Pr \left(\left| \mathcal{V}_\psi(X, Y|Z = z_0) - \iint_{D(\psi)} \Lambda_N(t, s) w(t, s) dt ds \right| < \epsilon \right) \geq 1 - \delta.$$

Let $H(x) \doteq x(1 - G(x\psi)/c_d)$ and introduce the following estimator for $\mathcal{V}_\psi(X, Y|Z = z_0)$:

$$\begin{aligned} \hat{\Omega}_{\psi, N}(z_0) &\doteq \frac{1}{N^2} \sum_{k, l} \frac{\hat{p}(z_0, X_k, Y_k) \hat{p}(z_0, X_l, Y_l)}{\hat{p}(X_k, Y_k) \hat{p}(X_l, Y_l) \hat{p}^2(z_0)} H(|X_k - X_l|) H(|Y_k - Y_l|) \\ &+ \frac{1}{N^4} \left[\sum_{k, l} \frac{\hat{p}(z_0, X_k) \hat{p}(z_0, X_l)}{\hat{p}(X_k) \hat{p}(X_l) \hat{p}^2(z_0)} H(|X_k - X_l|) \right] \left[\sum_{m, n} \frac{\hat{p}(z_0, Y_n) \hat{p}(z_0, Y_m)}{\hat{p}(Y_n) \hat{p}(Y_m) \hat{p}^2(z_0)} H(|Y_m - Y_n|) \right] \\ &- \frac{2}{N^3} \sum_{n, k, l} \frac{\hat{p}(z_0, X_n, Y_n) \hat{p}(z_0, X_k) \hat{p}(z_0, Y_l)}{\hat{p}(X_n, Y_n) \hat{p}(X_k) \hat{p}(Y_l) \hat{p}^3(z_0)} H(|X_n - X_k|) H(|Y_n - Y_l|). \end{aligned}$$

We have the following weak consistency theorem for this estimator.

Theorem 15 (Weak-consistency) *Under the conditions of Lemma 11, for all $\epsilon, \delta, \psi > 0$, there exists $N_0 = N_0(\epsilon, \delta, \psi)$ such that if $N > N_0$, then $\Pr(|\hat{\Omega}_{\psi, N}(z_0) - \mathcal{V}_\psi(X, Y|Z = z_0)| < \epsilon) \geq 1 - \delta$.*

Proof $\lim_{N \rightarrow \infty} \hat{\Omega}_{\psi, N}(z_0) = \lim_{N \rightarrow \infty} \iint_{D(\psi)} \Lambda_N(t, s) w(t, s) dt ds$ almost surely (since in the limit $V_{1, N} + V_{2, N} - 2V_{3, N}$ diminishes), and we use Lemma 14 above. $\blacksquare$

In practice, we are mostly interested in the $\lim_{\psi \rightarrow 0} \mathcal{V}_\psi(X, Y|Z = z_0) = \mathcal{V}(X, Y|Z = z_0)$ limit case. In this case, the estimator has a simple form:

$$\begin{aligned} \hat{\mathcal{V}}_N(X, Y|Z = z_0) &\doteq \frac{1}{N^2} \sum_{k, l} \frac{\hat{p}(z_0, X_k, Y_k) \hat{p}(z_0, X_l, Y_l)}{\hat{p}(X_k, Y_k) \hat{p}(X_l, Y_l) \hat{p}^2(z_0)} |X_k - X_l| |Y_k - Y_l| \\ &+ \frac{1}{N^4} \left[\sum_{k, l} \frac{\hat{p}(z_0, X_k) \hat{p}(z_0, X_l)}{\hat{p}(X_k) \hat{p}(X_l) \hat{p}^2(z_0)} |X_k - X_l| \right] \left[\sum_{m, n} \frac{\hat{p}(z_0, Y_n) \hat{p}(z_0, Y_m)}{\hat{p}(Y_n) \hat{p}(Y_m) \hat{p}^2(z_0)} |Y_m - Y_n| \right] \\ &- \frac{2}{N^3} \sum_{n, k, l} \frac{\hat{p}(z_0, X_n, Y_n) \hat{p}(z_0, X_k) \hat{p}(z_0, Y_l)}{\hat{p}(X_n, Y_n) \hat{p}(X_k) \hat{p}(Y_l) \hat{p}^3(z_0)} |X_n - X_k| |Y_n - Y_l|. \end{aligned} \tag{3}$$

Note that $\hat{\mathcal{V}}_N(X, Y|Z = z_0)$ can be calculated in $\mathcal{O}(N^2)$ time if the densities (i.e. the $\hat{p}$ terms in (3)) are already estimated. To see this, introduce the following notations:

$$\bar{a}_n = \frac{1}{N} \sum_{k=1}^N \sqrt{\frac{\hat{p}(z_0|X_n, Y_n)}{\hat{p}(z_0)}} \frac{\hat{p}(z_0|X_k)}{\hat{p}(z_0)} |X_n - X_k|, \quad \bar{b}_n = \frac{1}{N} \sum_{l=1}^N \sqrt{\frac{\hat{p}(z_0|X_n, Y_n)}{\hat{p}(z_0)}} \frac{\hat{p}(z_0|Y_l)}{\hat{p}(z_0)} |Y_n - Y_l|.$$

Now, the last term in $\hat{\mathcal{V}}_N(X, Y|Z = z_0)$ is simply $\frac{1}{N} \sum_{n=1}^N \bar{a}_n \bar{b}_n$. $\hat{\mathcal{V}}_N(X, Y|Z = z_0)$ has a simple form; it is a weighted average of the product of sample distances. In the next section we show that these weights can be calculated using sample distances as well.

5.1. k -NN Based Density Estimators

Let $X_{1:N} \doteq (X_1, \dots, X_N)$ be an i.i.d. sample from a distribution with density p , and let $\rho_k(x)$ denote the Euclidean distance of the k th nearest neighbor of x in the sample $X_{1:N} \setminus x$. k -NN density estimators operate using only distances between the observations in a given sample ($X_{1:N}$) and their k th nearest neighbors (breaking ties arbitrarily). Loftsgaarden and Quesenberry (1965) define the k -NN based density estimators of density p as follows.

Definition 16 (k -NN based density estimators)

$$\hat{p}_k(x) = \frac{k/N}{\text{Vol}(\mathcal{B}(x, \rho_k(x)))} = \frac{k}{N \lambda \rho_k^d(x)} \quad (4)$$

The following theorems show the consistency of these density estimators.

Theorem 17 (k -NN density estimators, convergence in probability) *If $k(N)$ denotes the number of neighbors applied at sample size N , $\lim_{N \rightarrow \infty} k(N) = \infty$, and $\lim_{N \rightarrow \infty} N/k(N) = \infty$, then $\hat{p}_{k(N)}(x) \rightarrow_p p(x)$ for almost all x .*

Theorem 18 (k -NN density estimators, a. s. convergence in sup norm) *If $\lim_{N \rightarrow \infty} \frac{k(N)}{\log(N)} = \infty$ and $\lim_{N \rightarrow \infty} \frac{N}{k(N)} = \infty$, then $\lim_{N \rightarrow \infty} \sup_x |\hat{p}_{k(N)}(x) - p(x)| = 0$ almost surely.*

If we apply k -NN density estimators for $p(Z)$, $p(X)$, $p(Y)$, $p(X, Y)$, $p(Z, X)$, $p(Z, Y)$, and $p(Z, X, Y)$, then we arrive at an estimator that uses only sample distances for the estimation. Note, however, that we need to consider different concatenations of the original $X_{1:N}$, $Y_{1:N}$, $Z_{1:N}$, z_0 samples. For example, the estimation of $p(Z, X)$ is a $d_z + d_x$ dimensional problem that uses the $\{[Z_n; X_n]\}_{n=1}^N$ sample set, where $[Z_n; X_n] \in \mathbb{R}^{d_z + d_x}$. In Equation (3), $\hat{\mathcal{V}}_N(X, Y|Z = z_0)$ is an estimator for $\mathcal{V}(X, Y|Z = z_0)$. $\mathcal{V}(X, Y|Z)$ can be estimated as follows:

$$\hat{\mathcal{V}}_{N,T}(X, Y|Z) \doteq \frac{1}{T} \sum_{t=1}^T \hat{\mathcal{V}}_N(X, Y|Z = \tilde{Z}_t) \quad (5)$$

This estimator can be calculated in $\mathcal{O}(TN^2)$ time. Similarly, we can derive estimators for $\mathcal{V}(X|Z = z_0)$ too.

$$\begin{aligned} \hat{\mathcal{V}}_N(X|Z = z_0) &\doteq \frac{1}{N^2} \sum_{k,l} \frac{\hat{p}(z_0, X_k) \hat{p}(z_0, X_l)}{\hat{p}(X_k) \hat{p}(X_l) \hat{p}^2(z_0)} |X_k - X_l|^2 + \frac{1}{N^4} \left(\sum_{k,l} \frac{\hat{p}(z_0, X_k) \hat{p}(z_0, X_l)}{\hat{p}(X_k) \hat{p}(X_l) \hat{p}^2(z_0)} |X_k - X_l| \right)^2 \\ &\quad - \frac{2}{N^3} \sum_{n,k,l} \frac{\hat{p}(z_0, X_n) \hat{p}(z_0, X_k) \hat{p}(z_0, X_l)}{\hat{p}(X_n) \hat{p}(X_k) \hat{p}(X_l) \hat{p}^3(z_0)} |X_n - X_k| |X_n - X_l|. \end{aligned} \quad (6)$$

The estimator of $\mathcal{V}(X|Z)$ is given by $\hat{\mathcal{V}}_{N,T}(X|Z) \doteq \frac{1}{T} \sum_{t=1}^T \hat{\mathcal{V}}_N(X|Z = \tilde{Z}_t)$. Similarly, the estimation of $\mathcal{R}(|X, Y|Z = z_0)$, $\hat{\mathcal{R}}_N(X, Y|Z = z_0)$, can be calculated by plugging the $\hat{\mathcal{V}}_N(X, Y|Z = z_0)$, $\hat{\mathcal{V}}_N(X|Z = z_0)$, and $\hat{\mathcal{V}}_N(Y|Z = z_0)$ estimates into Definition 10. Finally, $\hat{\mathcal{R}}_{N,T}(X, Y|Z) \doteq \frac{1}{T} \sum_{t=1}^T \hat{\mathcal{R}}_N(X, Y|Z = \tilde{Z}_t)$. We note that these estimators are not robust. If the densities (and hence the weights of the sample distances) are estimated poorly, then this might result in bad $\hat{\mathcal{V}}_N(X|Z = \tilde{Z}_t)$, $\hat{\mathcal{R}}_N(X, Y|Z = \tilde{Z}_t)$ estimates. To have a more robust estimator, we might use a robust mean by removing outliers, or even using the median instead of the simple empirical average ($\frac{1}{T} \sum_{t=1}^T \hat{\mathcal{R}}(X, Y|Z = \tilde{Z}_t)$) can result in a more robust estimation.

5.2. Semiparametric Estimation

If we knew that the underlying X, Y, Z variables belong to a parametric distribution family, then we can achieve a faster estimation compared to the totally nonparametric case. In this case, we can estimate the parameters of these densities first and then use them to calculate the weights of the sample distances in (3). We call this problem “semiparametric estimation”, because a part of the estimation problem is parametric, but we still use nonparametric methods to calculate the integral in Definition 7 with respect to $w(t, s)dt ds$.

6. Numerical Experiments to Demonstrate Consistency

In this section we demonstrate the consistency of the estimators with numerical experiments. Even for known density functions, generally it is difficult to calculate the conditional distance covariance and correlation in closed forms. However, when X and Y are 1-dimensional Gaussian variables, then this is possible.

Let $p(x, y, z)$ denote the joint density of $\mathcal{N}(\mu, \Sigma)$, Gaussian distribution with expected value μ and covariance matrix Σ . Its characteristic function is given by $\mathcal{F}(a) = e^{ia^T\mu + a^T\Sigma a}$. It is known that in this case the $p(X, Y|Z = z)$ distribution is also Gaussian. Its mean and covariance matrix is denoted by $\mu_{XY|z}$ and $\Sigma_{XY|z}$, respectively. Using these notations, $\mathcal{V}(X, Y|Z)$ can be calculated as follows.

Lemma 19

$$\begin{aligned} \mathcal{V}(X, Y|Z) &= \\ &= \iiint p(z) \left(c_{d_x} c_{d_y} |t|_{d_x}^{1+d_x} |s|_{d_y}^{1+d_y} \right)^{-1} e^{-t^T \Sigma_{X|z} t - s^T \Sigma_{Y|z} s} \left(1 + e^{-2t^T \rho_{XY|z} s} - 2e^{-t^T \rho_{XY|z} s} \right) dt ds dz, \end{aligned}$$

where $\Sigma_{XY|Z} = \begin{pmatrix} \Sigma_{X|z} & \rho_{XY|z} \\ \rho_{XY|z}^T & \Sigma_{Y|z} \end{pmatrix}$. Interestingly, for Gaussian variables the $\Sigma_{XY|Z}$ covariance matrix of $p(X, Y|Z = z)$ distribution does not depend on the actual value of z .

Lemma 20 *In the $d_x = d_y = 1$ special case, $\mathcal{V}(X, Y|Z)$ has a simple form:*

$$\begin{aligned} \mathcal{V}(X, Y|Z) &= \frac{4}{\pi} \left[\sqrt{\Sigma_{X|z} \Sigma_{Y|z} - \rho_{XY|z}^2} - \sqrt{\Sigma_{X|z} \Sigma_{Y|z}} + \rho_{XY|z} \arcsin \left(\rho_{XY|z} / \sqrt{\Sigma_{X|z} \Sigma_{Y|z}} \right) \right] \\ &- \frac{4}{\pi} \left[\sqrt{4\Sigma_{X|z} \Sigma_{Y|z} - \rho_{XY|z}^2} - \sqrt{4\Sigma_{X|z} \Sigma_{Y|z}} + \rho_{XY|z} \arcsin \left(\rho_{XY|z} / \sqrt{4\Sigma_{X|z} \Sigma_{Y|z}} \right) \right]. \end{aligned}$$

Similarly, we can calculate the $\mathcal{V}(X|Z)$, $\mathcal{R}(X, Y|Z)$ quantities in closed form too. For the details, see Appendix G.

6.1. Gaussian Distributions

The following experiment demonstrates that the proposed estimator (5) can consistently estimate $\mathcal{V}(X, Y|Z)$. In Figure 1(a) we display the performances of the proposed $\hat{\mathcal{V}}_N(X, Y|Z)$ estimator when the joint density of (X, Y, Z) is zero-mean Gaussian with a randomly chosen 3-dimensional nonsingular covariance matrix. Our results show that when we increase the sample size N , then $\hat{\mathcal{V}}_N(X, Y|Z)$ converges to $\mathcal{V}(X, Y|Z)$. The number of instances were varied between 50 and 2 500.

The means and standard deviations of the estimations are shown using error bars calculated from 10 independent runs. The number of nearest neighbors in then k -NN density estimator was set to $k = \lfloor \sqrt{N} \rfloor$. We also show results for the case when the densities—and thus the weights of the sample distances in (3)—are known. We can observe that in this case, when we do not need to estimate the densities, the convergence is faster.

We were also interested in studying the case when $\mathcal{V}(X, Y) > 0$, but $\mathcal{V}(X, Y|Z) = 0$. In other words, there is dependence between X and Y , but X and Y are independent given Z . To study this problem, we repeated the previous experiment with the exception that in this case we sampled A, B, Z random variables from the 1-dimensional standard normal distributions, and then set $X = A + Z$, $Y = B + Z$. It is easy to verify that in this case $\mathcal{V}(X, Y) > 0$, but $\mathcal{V}(X, Y|Z) = 0$. In Figure 1(b) we show five independent experiments. The results demonstrate that when we increase the sample sizes N , then the estimator converges to the true $\mathcal{V}(X, Y|Z) = 0$ value.

6.2. Nongaussian case

In this section we show that the proposed estimators can be used in the non-Gaussian case too. Generally, it is difficult to find a closed form expression for $\mathcal{V}(X, Y|Z)$. However when X and Y are conditionally independent given Z , then we know that $\mathcal{V}(X, Y|Z) = 0$. We set A, B, C to be independent $Beta(1, 3)$ variables and then set $Z = 5C$, $X = 5A + 5Z$, and $Y = 5B + 5Z$. In this case $\mathcal{V}(X, Y) > 0$, but $\mathcal{V}(X, Y|Z) = 0$. Figure 1(c) shows that as we increase the sample size, the estimator converges to the right quantity.

6.3. Semiparametric estimation

In Figure 1(d) we demonstrate the semiparametric approach. In this experiment we knew that the samples had joint Gaussian distribution, and we estimated their covariance matrices using maximum likelihood estimations. We plugged these parameters into their densities and then used these estimated densities to calculate the weights of the sample distances in (5).

7. Experiments on Medical Data

The next experiment demonstrates that the proposed estimator might be useful to detect confounder variables in medical data too.² We used the medical data published in Edwards (2000) (Section 3.1.4.). The data were taken from 35 patients and consist of three variables: digoxin clearance (X), urine flow (Y), and creatinine clearance (Z) (Fig. 2(a)). From medical knowledge we know that X should be independent of Y given Z . It was presented in Fukumizu et al. (2008) that there is a strong linear correlation between X and Y (Fig. 2(b)), and a partial correlation based test was not able to show the conditional independence of X and Y given Z . Fig. 2(c) shows that the distance covariance estimator was able to detect the large dependence between variables X and Y , but the *conditional* distance covariance estimator also shows that this dependence vanishes when we observe variable Z .

2. Note, however, that we do not have hypothesis tests yet that could determine if the variables are conditionally independent with some significance level.

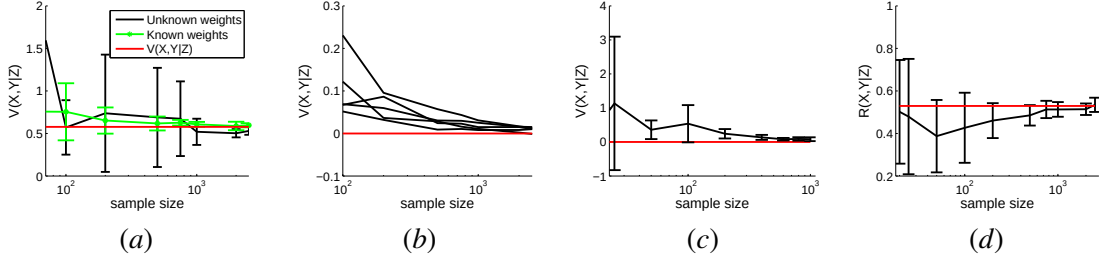


Figure 1: Estimated vs. true $\mathcal{V}(X, Y|Z)$ and $\mathcal{R}(X, Y|Z)$ as a function of the sample size. X, Y, Z were chosen to be 1d Gaussian variables. The red line indicates the true values. (a) We chose X, Y, Z such that $\mathcal{V}(X, Y|Z) > 0$. The means and standard deviations of the estimations are shown using error bars calculated from 10 independent runs. When the densities (and therefore the weights in (3)) are known, then the convergence is faster (green line). (b) X, Y, Z were chosen to be 1d Gaussian variables such that X and Y are conditionally independent given Z . We show five independent experiments. (c). A, B, C were chosen to be independent $Beta(1, 3)$ variables, and then we set $Z = 5C$, $X = 5A + Z$, $Y = 5B + Z$, so $\mathcal{V}(X, Y) > 0$, but $\mathcal{V}(X, Y|Z) = 0$. (d) Estimated vs. true $\mathcal{R}(X, Y|Z)$ as a function of the sample size.

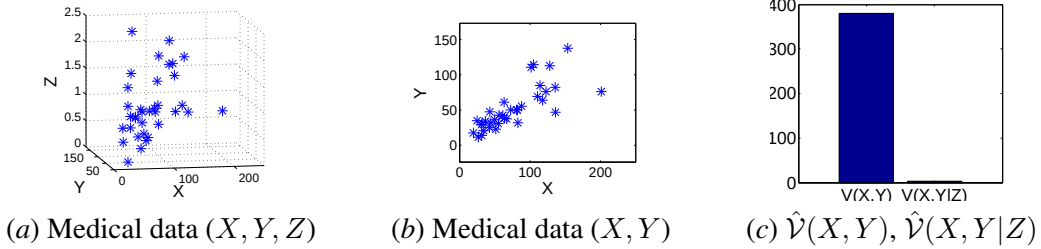


Figure 2: (a) The medical data set. (b) There is dependence between X and Y . (c) Estimated $\mathcal{V}(X, Y)$ and $\mathcal{V}(X, Y|Z)$ values.

8. Conclusion

We proposed new nonparametric estimators for the conditional distance covariance, variance and correlation. We proved the weak consistency of the estimators and demonstrated their consistency and applicability by numerical experiments on real and simulated datasets. There are several open questions left waiting for answers. Currently we do not know the convergence rates of the estimators, and how they depend on the parameters such as k , the dimension, and the densities. We do not know the asymptotic distributions of the estimators, and do not have conditional independence tests yet. [Gretton et al. \(2009\)](#) have shown recently that there is a connection between kernel methods and distance covariances. It is also known that certain conditional dependencies can be estimated with kernel methods ([Fukumizu et al., 2008](#)). However, it is still unknown if kernel methods and conditional distance covariance can be related to each other.

Acknowledgments

We thank John Lafferty for the helpful discussions.

References

- M. Baumgarten and C. Olsen. Confounding in epidemiology. http://www.collegeboard.com/prod_downloads/yes/4297_MODULE_10.pdf, 2004.
- T. Bouezmarni, J. Rombouts, and A. Taamouti. A nonparametric copula based test for conditional independence with applications to Granger causality. Technical report, Universidad Carlos III, Departamento de Economia, 2009.
- D. Edwards. *Introduction to graphical modelling*. Springer verlag, New York, 2000.
- K. Fukumizu, A. Gretton, X. Sun, and B. Schoelkopf. Kernel measures of conditional dependence. In *NIPS 20*, pages 489–496, Cambridge, MA, 2008. MIT Press.
- M. Goría, N. Leonenko, V. Mergel, and N. Inverardi. A new class of random vector entropy estimators and its applications in testing statistical hypotheses. *Journal of Nonparametric Statistics*, 17: 277–297, 2005.
- A. Gretton, R. Herbrich, and A. Smola. The kernel mutual information. In *Proc. ICASSP*, 2003.
- A. Gretton, K. Fukumizu, and B. Sriperumbudur. Discussion of: Brownian distance covariance. *The Annals of Applied Statistics*, 3(4):1285–1294, 2009.
- A. Hero and O. Michel. Asymptotic theory of greedy approximations to minimal k -point random graphs. *IEEE Trans. on Information Theory*, 45(6):1921–1938, 1999.
- A. Hero, B. Ma, O. Michel, and J. Gorman. Alpha-divergence for classification, indexing and retrieval, 2002a. Communications and Signal Processing Laboratory Technical Report CSPL-328.
- A. Hero, B. Ma, O. Michel, and J. Gorman. Applications of entropic spanning graphs. *IEEE Signal Processing Magazine*, 19(5):85–95, 2002b.
- N. Leonenko, L. Pronzato, and V. Savani. A class of Rényi information estimators for multidimensional densities. *Annals of Statistics*, 36(5):2153–2182, 2008.
- D. Loftsgaarden and C. Quesenberry. A nonparametric estimate of a multivariate density function. *Annals of Mathematical Statistics*, 36:1049–1051, 1965.
- R. Matthews. Storks deliver babies ($p=0.008$). *Teaching Statistics*, 22:36–38, 2000.
- D. Montgomery. *Design and Analysis of Experiments*. John Wiley and Sons, 2005.
- D. Pál, B. Póczos, and Cs. Szepesvári. Estimation of Rényi entropy and mutual information based on generalized nearest-neighbor graphs. In *NIPS 2010*, 2010.
- J. Pearl. Why there is no statistical test for confounding, why many think there is, and why they are almost right. UCLA Computer Science Department, Technical Report R-256, 1998.

- F. Pérez-Cruz. Estimation of information theoretic measures for continuous random variables. In *NIPS-2008*, volume 21, 2008.
- B. Póczos, S. Kirshner, and Cs. Szepesvári. REGO: Rank-based estimation of Rényi information using Euclidean graph optimization. In *AISTATS 2010*, 2010.
- B. Schweizer and E. Wolff. On nonparametric measures of dependence for random variables. *The Annals of Statistics*, 9, 1981.
- K. Sricharan, R. Raich, and A. Hero. Empirical estimation of entropy functionals with confidence. Technical Report, <http://arxiv.org/abs/1012.4188>, 2010.
- M. Steele. *Probability Theory and Combinatorial Optimization*. Society for Industrial and Applied Mathematics, 1997.
- L. Su and H. White. A nonparametric Hellinger metric test for conditional independence. *Econometric Theory*, 24:829–864, 2008.
- G. Székely and M. Rizzo. Hierarchical clustering via joint between-within distances: Extending Ward’s minimum variance method. *Journal of Classification*, 22(2):151–183, 2005.
- G. Székely, M. Rizzo, and N. Bakirov. Measuring and testing dependence by correlation of distances. *Annals of Statistics*, 35(6):2769–2794, 2007.
- Q. Wang, S. Kulkarni, and S. Verdú. Divergence estimation for multidimensional densities via k -nearest-neighbor distances. *IEEE Transactions on Information Theory*, 55(5), 2009.
- J. Yukich. *Probability Theory of Classical Euclidean Optimization Problems*. Springer, 1998.

Supplementary Material–Appendix

Appendix A. Other Numerical Experiments

In this section we study the behavior of the estimator and demonstrate its applicability on real data sets. Below we illustrate with image data that the conditional distance covariance and correlation can either be larger or smaller than the standard unconditional distance covariance and correlation. In other words, extra knowledge can either increase or decrease the dependence between random variables.

Increasing/decreasing conditional covariance

We chose a gray-scale image (Fig. 3(c)) of size 75×100 and considered its pixel values ($Z \in [0, 255]$) as if they were samples from a distribution. We also constructed two noisy versions of Z : $X = Z + A$ (Fig. 3(a)) and $Y = Z + B$ (Fig. 3(b)), where A and B were independent random noise variables with uniform $U[-5, 5]$ distributions. By construction, $\mathcal{V}(X, Y) > 0$, $\mathcal{R}(X, Y) > 0$, but $\mathcal{V}(X, Y|Z) = 0$, $\mathcal{R}(X, Y|Z) = 0$, that is, the observation of Z eliminates the dependence between X and Y . This is also confirmed by the estimated $\hat{\mathcal{V}}(X, Y|Z)$ value, which is much smaller than $\hat{\mathcal{V}}(X, Y)$ (Fig. 3(d)).

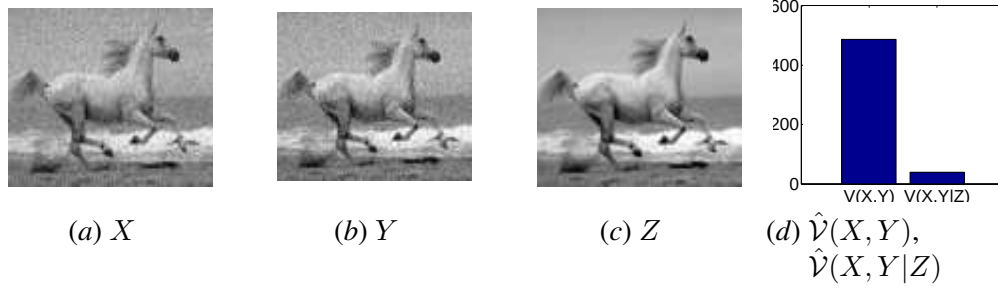


Figure 3: Demonstration that conditioning to a third variable (Z) can decrease the dependence between X and Y . (a), and (b): Noisy versions of the picture in (c). (d): Estimated $\mathcal{V}(X, Y)$ and $\mathcal{V}(X, Y|Z)$ values.

The following experiment demonstrates that the opposite situation can also occur. Similarly to the previous case, we chose two noisy images (Fig. 4(a) and Fig. 4(b)). We considered their pixel values as they were i.i.d. samples from two random variables X and Y , and then constructed their noisy sum: $Z = X + Y + A$, where A played the role of noise and it had uniform $U[-5, 5]$ distribution. Fig. 4(d) shows that $\hat{\mathcal{V}}(X, Y) \approx 0$ (i.e. the two original images were almost independent), but $\hat{\mathcal{V}}(X, Y|Z) > 0$ (i.e. having information about their sum increases the mutual information).

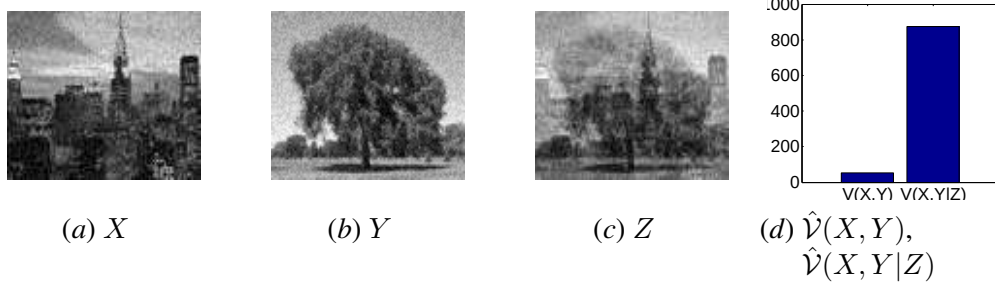


Figure 4: Demonstration that conditioning to a third variable (Z) can increase the dependence between X and Y . (c): Noisy sum of pictures in (a) and (b). (d): Estimated $\mathcal{V}(X, Y)$ and $\mathcal{V}(X, Y|Z)$.

Appendix B. Proofs of Lemmas

Proof of Lemma 2

Proof Let $x = R_x e_x$, where $R_x > 0$, $|x| = R_x$, and $e_x \in \mathbb{R}^d$ is a unit vector ($|e_x| = 1$). Introduce the following $G_1 : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}$ function.

$$G_1(y, x) \doteq \int_{|t| < y} \frac{1 - \cos \langle t, x \rangle}{|t|^{d+1} |x|} dt.$$

From Lemma 1, we can see that $G_1(y, x) \leq c_d$, and $\lim_{y \rightarrow \infty} G_1(y, x) = c_d$. Since $|x| = R_x$, after the $t' = R_x t$ integral transformation we have that

$$\begin{aligned} R_x G_1(y, x) &= \int_{|t'| < y R_x} \frac{1 - \cos \langle t' / R_x, R_x e_x \rangle}{|t'|^{d+1} / R_x^{d+1}} \frac{1}{R_x^d} dt' \\ &= \int_{|t'| < y R_x} \frac{1 - \cos \langle t', e_x \rangle}{|t'|^{d+1}} R_x dt'. \end{aligned} \quad (7)$$

Let Q_x be an orthonormal transformation that transforms e_x to a canonical unit vector e_0 (i.e., $e_0 = Q_x e_x$). After the $t'' = Q_x t'$ integral transformation and exploiting that $\langle t', e_x \rangle = \langle Q_x t', Q_x e_x \rangle$, we have that

$$G_1(y, x) = \int_{|t''| < y R_x} \frac{1 - \cos \langle t'', e_0 \rangle}{|t''|^{d+1}} dt''.$$

Since $G_1(y, x)$ does not depend on e_x (only on $y R_x$), thus $G_1(y, x) = G(y R_x) = G(y|x)$. ■

Proof of Lemma 3

Proof After d -dimensional spherical integral transformation, we have that

$$\int_{|t| > y} \frac{1}{|t|^{d+\alpha}} dt$$

$$\begin{aligned}
 &= \int_{r>y} \int_{\phi_1=0}^{\pi} \cdots \int_{\phi_{n-2}=0}^{\pi} \int_{\phi_{n-1}=0}^{2\pi} \frac{1}{r^{d+\alpha}} r^{d-1} \sin^{d-2}(\phi_1) \cdots \sin(\phi_{d-2}) \, dr \, d\phi_1 \cdots d\phi_{d-1} \\
 &\leq C \int_{r>y} \frac{1}{r^{1+\alpha}} dr < \infty, \text{ where } C < \infty.
 \end{aligned}$$

■

Proof of Lemma 8

Proof

$$\begin{aligned}
 &|\mathcal{F}_{X,Y|Z=z}(t,s) - \mathcal{F}_{X|Z=z}(t)\mathcal{F}_{Y|Z=z}(s)|^2 \\
 &= \left| \int \left[e^{i\langle t,x \rangle} - \int e^{i\langle t,\tilde{x} \rangle} p(\tilde{x}|z) d\tilde{x} \right] \left[e^{i\langle s,y \rangle} - \int e^{i\langle s,\tilde{y} \rangle} p(\tilde{y}|z) d\tilde{y} \right] p(x,y|z) dx dy \right|^2 \\
 &\leq \left[\int \left| e^{i\langle t,x \rangle} - \int e^{i\langle t,\tilde{x} \rangle} p(\tilde{x}|z) d\tilde{x} \right|^2 p(x|z) dx \right] \left[\int \left| e^{i\langle s,y \rangle} - \int e^{i\langle s,\tilde{y} \rangle} p(\tilde{y}|z) d\tilde{y} \right|^2 p(y|z) dy \right] \\
 &= (1 - |\mathcal{F}_{X|Z=z}(t)|^2) (1 - |\mathcal{F}_{Y|Z=z}(s)|^2).
 \end{aligned}$$

Here we used the Cauchy-Bunyakowsky inequality, and the following identity

$$\begin{aligned}
 &\int \left| e^{i\langle t,x \rangle} - \int e^{i\langle t,\tilde{x} \rangle} p(\tilde{x}|z) d\tilde{x} \right|^2 p(x|z) dx = \\
 &\mathbb{E} \left[(\cos\langle t, X \rangle - \mathbb{E}[\cos\langle t, \tilde{X} \rangle | Z = z])^2 + (\sin\langle t, X \rangle - \mathbb{E}[\sin\langle t, \tilde{X} \rangle | Z = z])^2 \middle| Z = z \right] \\
 &= 1 - (\mathbb{E}[\cos\langle t, X \rangle | Z = z])^2 - (\mathbb{E}[\sin\langle t, X \rangle | Z = z])^2 \\
 &= 1 - |\mathcal{F}_{X|Z=z}(t)|^2.
 \end{aligned}$$

Now we have that

$$\begin{aligned}
 \mathcal{V}(X,Y|Z=z) &\doteq \iint w(t,s) |\mathcal{F}_{X,Y|Z=z}(t,s) - \mathcal{F}_{X|Z=z}(t)\mathcal{F}_{Y|Z=z}(s)|^2 dt ds \\
 &\leq \iint w(t,s) (1 - |\mathcal{F}_{X|Z=z}(t)|^2) (1 - |\mathcal{F}_{Y|Z=z}(s)|^2) dt ds \\
 &= \iint \frac{1 - |\mathcal{F}_{X|Z=z}(t)|^2}{c_{d_x}|t|^{1+d_x}} \frac{1 - |\mathcal{F}_{Y|Z=z}(s)|^2}{c_{d_y}|s|^{1+d_y}} dt ds \\
 &= \int \frac{1 - \mathbb{E}[\cos\langle t, X - \tilde{X} \rangle | Z = z]}{c_{d_x}|t|^{1+d_x}} dt \int \frac{1 - \mathbb{E}[\cos\langle t, Y - \tilde{Y} \rangle | Z = z]}{c_{d_y}|s|^{1+d_y}} ds \\
 &= \mathbb{E} \left[\int \frac{1 - \cos\langle t, X - \tilde{X} \rangle}{c_{d_x}|t|^{1+d_x}} dt \middle| Z = z \right] \mathbb{E} \left[\int \frac{1 - \cos\langle t, Y - \tilde{Y} \rangle}{c_{d_y}|s|^{1+d_y}} ds \middle| Z = z \right] \\
 &= \mathbb{E}[|X - \tilde{X}| | Z = z] \mathbb{E}[|Y - \tilde{Y}| | Z = z] < \infty.
 \end{aligned}$$

Here $\tilde{X}, \tilde{Y}$ denote random variables sampled from $p(X|Z = z)$ and $p(Y|Z = z)$, respectively, and independently from X and Y . In the derivation we used Lemma 1, and Lemma 21, Lemma 22 below. ■

Lemma 21 $\int \frac{1 - \mathbb{E}[\cos\langle t, X - \tilde{X} \rangle | Z = z]}{c_{d_x} |t|^{1+d_x}} dt = \mathbb{E} \left[\int \frac{1 - \cos\langle t, X - \tilde{X} \rangle}{c_{d_x} |t|^{1+d_x}} dt \middle| Z = z \right].$

Lemma 22 $|\mathcal{F}_{X|Z=z}(t)|^2 = \mathbb{E} [\cos\langle t, X - \tilde{X} \rangle | Z = z].$

Proof of Lemma 21

Proof

$$\begin{aligned} \int \frac{1 - \mathbb{E} [\cos\langle t, X - \tilde{X} \rangle | Z = z]}{|t|^{1+d_x}} dt &= \int \frac{1 - \iint \cos\langle t, x - \tilde{x} \rangle p(x, \tilde{x}|z) dx d\tilde{x}}{|t|^{1+d_x}} dt \\ &= \int \frac{\iint (1 - \cos\langle t, x - \tilde{x} \rangle) p(x, \tilde{x}|z) dx d\tilde{x}}{|t|^{1+d_x}} dt \\ &= \iint \frac{(1 - \cos\langle t, x - \tilde{x} \rangle) p(x, \tilde{x}|z)}{|t|^{1+d_x}} dx d\tilde{x} dt \\ &= \mathbb{E} \left[\int \frac{1 - \cos\langle t, X - \tilde{X} \rangle}{|t|^{1+d_x}} dt \middle| Z = z \right] \end{aligned}$$

■

Proof of Lemma 22

Proof

$$\begin{aligned} \mathbb{E} [\cos\langle t, X - \tilde{X} \rangle | Z = z] &= \mathbb{E} [\cos\langle t, X \rangle \cos\langle t, \tilde{X} \rangle + \sin\langle t, X \rangle \sin\langle t, \tilde{X} \rangle | Z = z] \\ &= \mathbb{E}^2 [\cos\langle t, X \rangle | Z = z] + \mathbb{E}^2 [\sin\langle t, X \rangle | Z = z] \\ &= |\mathcal{F}_{X|Z=z}(t)|^2. \end{aligned}$$

■

Proof of Lemma 12

Proof The statement follows from Lemma 1 and Lemma 2:

$$\int_{|t| > \psi} \frac{1 - \cos\langle t, x \rangle}{c_d |t|^{d+1}} dt = |x| \left(1 - \frac{G(|x|\psi)}{c_d} \right).$$

■

Proof of Lemma 14

Proof

$$\begin{aligned} \Pr(|\mathcal{V}_\psi(X, Y|Z = z_0) - \iint_{D(\psi)} \Lambda_N(t, s) w(t, s) dt ds| < \epsilon) \\ = \Pr(|\iint_{D(\psi)} |\Lambda(t, s) - \Lambda_N(t, s)| w(t, s) dt ds| < \epsilon) \end{aligned}$$

From Lemma 11 we have that for all $\kappa, \delta > 0$ there exists $N_0 = N_0(\kappa, \delta)$ such that $\Pr(|\Lambda(t, s) - \Lambda_N(t, s)| < \kappa) \geq 1 - \delta$ (for all t, s) if $N > N_0$. In turn, for each $\epsilon > 0$ with probability at least $1 - \delta$

$$\iint_{D(\psi)} |\Lambda(t, s) - \Lambda_N(t, s)| w(t, s) dt ds < \kappa \iint_{D(\psi)} w(t, s) dt ds < \epsilon,$$

if $N > N_0(\kappa, \delta)$, and $0 < \kappa < \epsilon / \iint_{D(\psi)} w(t, s) dt ds$. ■

Appendix C. Rewriting $\mathcal{V}(X, Y|Z)$

$\Lambda(t, s)$ can be rewritten as follows :

$$\begin{aligned} \Lambda(t, s) &\doteq \left| \iint e^{i\langle t, x \rangle + i\langle s, y \rangle} p(x, y|z_0) dx dy - \left(\int e^{i\langle t, x \rangle} p(x|z_0) dx \right) \left(\int e^{i\langle s, y \rangle} p(y|z_0) dy \right) \right|^2 \\ &= \left\{ \iint \cos(\langle t, x \rangle + \langle s, y \rangle) \frac{p(x, y, z_0)}{p(z_0)} dx dy - \left[\int \frac{\cos(\langle t, x \rangle) p(x, z_0)}{p(z_0)} dx \right] \left[\int \frac{\cos(\langle s, y \rangle) p(y, z_0)}{p(z_0)} dy \right] \right. \\ &\quad + \left[\int \frac{\sin(\langle t, x \rangle) p(x, z_0)}{p(z_0)} dx \right] \left[\int \frac{\sin(\langle s, y \rangle) p(y, z_0)}{p(z_0)} dy \right] \Big\}^2 \\ &\quad + \left\{ \iint \sin(\langle t, x \rangle + \langle s, y \rangle) \frac{p(x, y, z_0)}{p(z_0)} dx dy - \left[\int \frac{\sin(\langle t, x \rangle) p(x, z_0)}{p(z_0)} dx \right] \left[\int \frac{\cos(\langle s, y \rangle) p(y, z_0)}{p(z_0)} dy \right] \right. \\ &\quad \left. - \left[\int \frac{\cos(\langle t, x \rangle) p(x, z_0)}{p(z_0)} dx \right] \left[\int \frac{\sin(\langle s, y \rangle) p(y, z_0)}{p(z_0)} dy \right] \right\}^2. \end{aligned}$$

Assume that a $\hat{p} = \hat{p}_N$ estimator is available for density p . Using the law of large numbers, we can approximate the integrals in this equation with the following quantities:

$$\iint \cos(\langle t, x \rangle + \langle s, y \rangle) \frac{p(x, y, z_0)}{p(z_0)} dx dy \approx \frac{1}{N} \sum_{n=1}^N \cos(\langle t, X_n \rangle + \langle s, Y_n \rangle) \frac{\hat{p}(z_0|X_n, Y_n)}{\hat{p}(z_0)}, \quad (8)$$

$$\int \frac{p(x, z_0)}{p(z_0)} \cos \langle t, x \rangle dx \approx \frac{1}{N} \sum_{n=1}^N \frac{\hat{p}(z_0|X_n)}{\hat{p}(z_0)} \cos \langle t, X_n \rangle, \quad (9)$$

$$\int \frac{p(y, z_0)}{p(z_0)} \cos \langle s, y \rangle dy \approx \frac{1}{N} \sum_{n=1}^N \frac{\hat{p}(z_0|Y_n)}{\hat{p}(z_0)} \cos \langle s, Y_n \rangle. \quad (10)$$

We do similar approximations with the sin function too. By plugging these approximations into the formulas above, we arrive at the following equations:

$$\begin{aligned} \Lambda_N(t, s) \doteq & \left\{ \frac{1}{N} \sum_{n=1}^N \cos(\langle t, X_n \rangle + \langle s, Y_n \rangle) \frac{\hat{p}(z_0|X_n, Y_n)}{\hat{p}(z_0)} \right. \\ & - \frac{1}{N^2} \left[\sum_{n=1}^N \cos\langle t, X_n \rangle \frac{\hat{p}(z_0|X_n)}{\hat{p}(z_0)} \right] \left[\sum_{n=1}^N \cos\langle s, Y_n \rangle \frac{\hat{p}(z_0|Y_n)}{\hat{p}(z_0)} \right] \\ & + \frac{1}{N^2} \left[\sum_{n=1}^N \sin\langle t, X_n \rangle \frac{\hat{p}(z_0|X_n)}{\hat{p}(z_0)} \right] \left[\sum_{n=1}^N \sin\langle s, Y_n \rangle \frac{\hat{p}(z_0|Y_n)}{\hat{p}(z_0)} \right] \Big\}^2 \\ & + \left\{ \frac{1}{N} \sum_{n=1}^N \sin(\langle t, X_n \rangle + \langle s, Y_n \rangle) \frac{\hat{p}(z_0|X_n, Y_n)}{\hat{p}(z_0)} \right. \\ & - \frac{1}{N^2} \left[\sum_{n=1}^N \sin\langle t, X_n \rangle \frac{\hat{p}(z_0|X_n)}{\hat{p}(z_0)} \right] \left[\sum_{n=1}^N \cos\langle s, Y_n \rangle \frac{\hat{p}(z_0|Y_n)}{\hat{p}(z_0)} \right] \\ & \left. - \frac{1}{N^2} \left[\sum_{n=1}^N \cos\langle t, X_n \rangle \frac{\hat{p}(z_0|X_n)}{\hat{p}(z_0)} \right] \left[\sum_{n=1}^N \sin\langle s, Y_n \rangle \frac{\hat{p}(z_0|Y_n)}{\hat{p}(z_0)} \right] \right\}^2. \end{aligned}$$

$$\begin{aligned} \mathcal{V}(X, Y|Z) &= \iint w(t, s) \int p(z) \left| \iint e^{i\langle t, x \rangle + i\langle s, y \rangle} p(x, y|z) dx dy \right. \\ &\quad \left. - \left(\int e^{i\langle t, x \rangle} p(x|z) dx \right) \left(\int e^{i\langle s, y \rangle} p(y|z) dy \right) \right|^2 dz dt ds \\ &= \iint w(t, s) \int p(z) \left| \iint e^{i\langle t, x \rangle + i\langle s, y \rangle} \frac{p(x, y, z)}{p(z)} dx dy \right. \\ &\quad \left. - \left(\int e^{i\langle t, x \rangle} \frac{p(x, z)}{p(z)} dx \right) \left(\int e^{i\langle s, y \rangle} \frac{p(y, z)}{p(z)} dy \right) \right|^2 dz dt ds. \end{aligned}$$

The absolute value term inside the integral can be rewritten:

$$\begin{aligned} & \left| \iint e^{i\langle t, x \rangle + i\langle s, y \rangle} \frac{p(x, y, z)}{p(z)} dx dy - \left(\int e^{i\langle t, x \rangle} \frac{p(x, z)}{p(z)} dx \right) \left(\int e^{i\langle s, y \rangle} \frac{p(y, z)}{p(z)} dy \right) \right|^2 \\ &= \left| \iint \cos(\langle t, x \rangle + \langle s, y \rangle) + i \sin(\langle t, x \rangle + \langle s, y \rangle) \frac{p(x, y, z)}{p(z)} dx dy \right. \\ &\quad \left. - \left(\int \cos(\langle t, x \rangle) + i \sin(\langle t, x \rangle) \frac{p(x, z)}{p(z)} dx \right) \left(\int \cos(\langle s, y \rangle) + i \sin(\langle s, y \rangle) \frac{p(y, z)}{p(z)} dy \right) \right|^2 \\ &= \left| \iint \frac{\cos(\langle t, x \rangle + \langle s, y \rangle) p(x, y, z)}{p(z)} + i \frac{\sin(\langle t, x \rangle + \langle s, y \rangle) p(x, y, z)}{p(z)} dx dy \right. \\ &\quad \left. - \left(\int \frac{\cos(\langle t, x \rangle) p(x, z)}{p(z)} + i \frac{\sin(\langle t, x \rangle) p(x, z)}{p(z)} dx \right) \left(\int \frac{\cos(\langle s, y \rangle) p(y, z)}{p(z)} + i \frac{\sin(\langle s, y \rangle) p(y, z)}{p(z)} dy \right) \right|^2 \\ &= \left| \iint \frac{\cos(\langle t, x \rangle + \langle s, y \rangle) p(x, y, z)}{p(z)} + i \frac{\sin(\langle t, x \rangle + \langle s, y \rangle) p(x, y, z)}{p(z)} dx dy \right. \\ &\quad - \left[\int \frac{\cos(\langle t, x \rangle) p(x, z)}{p(z)} dx \right] \left[\int \frac{\cos(\langle s, y \rangle) p(y, z)}{p(z)} dy \right] \\ &\quad \left. + \left[\int \frac{\sin(\langle t, x \rangle) p(x, z)}{p(z)} dx \right] \left[\int \frac{\sin(\langle s, y \rangle) p(y, z)}{p(z)} dy \right] \right|^2 \end{aligned}$$

$$\begin{aligned}
 & -i \left[\int \frac{\sin(\langle t, x \rangle) p(x, z)}{p(z)} dy \right] \left[\int \frac{\cos(\langle s, y \rangle) p(y, z)}{p(z)} dy \right] \\
 & -i \left[\int \frac{\cos(\langle t, x \rangle) p(x, z)}{p(z)} dy \right] \left[\int \frac{\sin(\langle s, y \rangle) p(y, z)}{p(z)} dy \right] \Big|^2 \\
 & = \left\{ \iint \frac{\cos(\langle t, x \rangle + \langle s, y \rangle) p(x, y, z)}{p(z)} dx dy - \left[\int \frac{\cos(\langle t, x \rangle) p(x, z)}{p(z)} dx \right] \left[\int \frac{\cos(\langle s, y \rangle) p(y, z)}{p(z)} dy \right] \right. \\
 & \quad \left. + \left[\int \frac{\sin(\langle t, x \rangle) p(x, z)}{p(z)} dx \right] \left[\int \frac{\sin(\langle s, y \rangle) p(y, z)}{p(z)} dy \right] \right\}^2 \\
 & + \left\{ \frac{\sin(\langle t, x \rangle + \langle s, y \rangle) p(x, y, z)}{p(z)} dx dy - \left[\int \frac{\sin(\langle t, x \rangle) p(x, z)}{p(z)} dx \right] \left[\int \frac{\cos(\langle s, y \rangle) p(y, z)}{p(z)} dy \right] \right. \\
 & \quad \left. - \left[\int \frac{\cos(\langle t, x \rangle) p(x, z)}{p(z)} dx \right] \left[\int \frac{\sin(\langle s, y \rangle) p(y, z)}{p(z)} dy \right] \right\}^2.
 \end{aligned}$$

Appendix D. Proof of Lemma 11

The lemma follows from the Chebysev's inequality (weak law of large numbers) by bounding the errors of the approximations in (8)–(10). Assume that there exist $\sigma^2 < \infty$ and $K < \infty$ such that for all t, s it holds that $\max(\text{Var}[\tilde{a}_{i,1}], \text{Var}[\tilde{b}_{i,1}]) \leq \sigma^2 \leq \infty$, and $\max(a_i, b_i) < K < \infty$. First we prove that for all $\epsilon > 0, \delta > 0$ there exists $N_0 = N_0(\epsilon, \delta) < \infty$ such that if $N > N_0$, then for all t, s, i

$$\begin{aligned}
 \Pr(|a_{i,N} - a_i| \geq \epsilon) & \leq \delta \\
 \Pr(|b_{i,N} - b_i| \geq \epsilon) & \leq \delta
 \end{aligned}$$

This is easy to see. For example, for $a_{1,N}$ we have that

$$\begin{aligned}
 |a_{1,N} - a_1| & = \left| \frac{1}{N} \sum_{n=1}^N \cos(\langle t, X_n \rangle + \langle s, Y_n \rangle) \frac{\hat{p}(z_0 | X_n, Y_n)}{\hat{p}(z_0)} - \iint \cos(\langle t, x \rangle + \langle s, y \rangle) \frac{p(z_0, x, y)}{p(z_0)} dx dy \right| \\
 & \leq \left| \frac{1}{N} \sum_{n=1}^N \cos(\langle t, X_n \rangle + \langle s, Y_n \rangle) \frac{\hat{p}(z_0 | X_n, Y_n)}{\hat{p}(z_0)} - \frac{1}{N} \sum_{n=1}^N \cos(\langle t, X_n \rangle + \langle s, Y_n \rangle) \frac{p(z_0 | X_n, Y_n)}{p(z_0)} \right| \\
 & \quad + \left| \frac{1}{N} \sum_{n=1}^N \cos(\langle t, X_n \rangle + \langle s, Y_n \rangle) \frac{p(z_0 | X_n, Y_n)}{p(z_0)} - \iint \cos(\langle t, x \rangle + \langle s, y \rangle) \frac{p(z_0, x, y)}{p(z_0)} dx dy \right| \\
 & \leq \left| \frac{1}{N} \sum_{n=1}^N \cos(\langle t, X_n \rangle + \langle s, Y_n \rangle) \frac{p(z_0 | X_n, Y_n)}{p(z_0)} - \iint \cos(\langle t, x \rangle + \langle s, y \rangle) \frac{p(z_0, x, y)}{p(z_0)} dx dy \right| \\
 & \quad + \frac{1}{N} \sum_{n=1}^N \left| \frac{\hat{p}(z_0, X_n, Y_n)}{\hat{p}(z_0) \hat{p}(X_n, Y_n)} - \frac{p(z_0, X_n, Y_n)}{p(z_0) p(X_n, Y_n)} \right|. \tag{11}
 \end{aligned}$$

$$+ \frac{1}{N} \sum_{n=1}^N \left| \frac{\hat{p}(z_0, X_n, Y_n)}{\hat{p}(z_0) \hat{p}(X_n, Y_n)} - \frac{p(z_0, X_n, Y_n)}{p(z_0) p(X_n, Y_n)} \right|. \tag{12}$$

(11) can be upper bounded by the Markov inequality. For all t, s , we have that

$$\Pr \left(\left| \frac{1}{N} \sum_{n=1}^N \cos(\langle t, X_n \rangle + \langle s, Y_n \rangle) \frac{p(z_0 | X_n, Y_n)}{p(z_0)} - \iint \cos(\langle t, x \rangle + \langle s, y \rangle) \frac{p(z_0, x, y)}{p(z_0)} dx dy \right| \geq \epsilon \right) \leq \frac{\sigma^2}{N \epsilon^2}.$$

Since $\lim_{N \rightarrow \infty} \sup(x) |\hat{p}(x) - p(x)| = 0$ by the assumptions on the density estimator $\hat{p}$, thus with high probability (12) also becomes arbitrarily close to zero if N is large enough.

$$\Lambda_N(t, s) = (a_{1,N} - a_{2,N} a_{3,N} + a_{4,N} a_{5,N})^2 + (b_{1,N} - b_{2,N} b_{3,N} + b_{4,N} b_{5,N})^2,$$

$$\Lambda(t, s) = (a_1 - a_2 a_3 + a_4 a_5)^2 + (b_1 - b_2 b_3 + b_4 b_5)^2.$$

Therefore,

$$\begin{aligned} |\Lambda_N(t, s) - \Lambda(t, s)| &\leq |(a_{1,N} - a_{2,N} a_{3,N} + a_{4,N} a_{5,N})^2 - (a_1 - a_2 a_3 + a_4 a_5)^2| \\ &\quad + |(b_{1,N} - b_{2,N} b_{3,N} + b_{4,N} b_{5,N})^2 - (b_1 - b_2 b_3 + b_4 b_5)^2| \\ &\leq (|a_{1,N} - a_1| + |a_{2,N} a_{3,N} - a_2 a_3| + |a_{4,N} a_{5,N} - a_4 a_5|) \\ &\quad \times (|a_{1,N} + a_1| + |a_{2,N} a_{3,N} + a_2 a_3| + |a_{4,N} a_{5,N} + a_4 a_5|) \\ &\quad + (|b_{1,N} - b_1| + |b_{2,N} b_{3,N} - b_2 b_3| + |b_{4,N} b_{5,N} - b_4 b_5|) \\ &\quad \times (|b_{1,N} + b_1| + |b_{2,N} b_{3,N} + b_2 b_3| + |b_{4,N} b_{5,N} + b_4 b_5|) \end{aligned}$$

From the triangle inequality we have that

$$\begin{aligned} |a_{2,N} a_{3,N} - a_2 a_3| &\leq |(a_2 - a_{2,N}) a_3| + |(a_3 - a_{3,N})(a_{2,N} - a_2)| + |(a_3 - a_{3,N}) a_2| \\ |a_{4,N} a_{5,N} - a_4 a_5| &\leq |(a_4 - a_{4,N}) a_5| + |(a_5 - a_{5,N})(a_{4,N} - a_4)| + |(a_5 - a_{5,N}) a_4| \\ |b_{2,N} b_{3,N} - b_2 b_3| &\leq |(b_2 - b_{2,N}) b_3| + |(b_3 - b_{3,N})(b_{2,N} - b_2)| + |(b_3 - b_{3,N}) b_2| \\ |b_{4,N} b_{5,N} - b_4 b_5| &\leq |(b_4 - b_{4,N}) b_5| + |(b_5 - b_{5,N})(b_{4,N} - b_4)| + |(b_5 - b_{5,N}) b_4| \end{aligned}$$

Since $a_i, b_i < K$ for all t, s , one can see that for each $\epsilon, \delta > 0$ there exist $N_0(\epsilon, \delta)$, such that if $N > N_0$, then

$$\begin{aligned} \Pr(|a_{1,N} - a_1| + |a_{2,N} a_{3,N} - a_2 a_3| + |a_{4,N} a_{5,N} - a_4 a_5| > \epsilon) &\leq \delta, \\ \Pr(|a_{1,N} + a_1| + |a_{2,N} a_{3,N} + a_2 a_3| + |a_{4,N} a_{5,N} + a_4 a_5| > 10K + \epsilon) &\leq \delta. \end{aligned}$$

It immediately follows from this that with high probability $|\Lambda_N(t, s) - \Lambda(t, s)|$ can be arbitrarily small if N is large enough.

Appendix E. Rewriting Λ_N

We already know that

$$\begin{aligned} \Lambda_N &= \left\{ \frac{1}{N} \sum_{n=1}^N \cos(\langle t, X_n \rangle + \langle s, Y_n \rangle) \frac{\hat{p}(z_0 | X_n, Y_n)}{\hat{p}(z_0)} \right. \\ &\quad - \frac{1}{N^2} \left[\sum_{n=1}^N \cos \langle t, X_n \rangle \frac{\hat{p}(z_0 | X_n)}{\hat{p}(z_0)} \right] \left[\sum_{n=1}^N \cos \langle s, Y_n \rangle \frac{\hat{p}(z_0 | Y_n)}{\hat{p}(z_0)} \right] \\ &\quad + \frac{1}{N^2} \left[\sum_{n=1}^N \sin \langle t, X_n \rangle \frac{\hat{p}(z_0 | X_n)}{\hat{p}(z_0)} \right] \left[\sum_{n=1}^N \sin \langle s, Y_n \rangle \frac{\hat{p}(z_0 | Y_n)}{\hat{p}(z_0)} \right] \Big\}^2 \\ &\quad + \left\{ \frac{1}{N} \sum_{n=1}^N \sin(\langle t, X_n \rangle + \langle s, Y_n \rangle) \frac{\hat{p}(z_0 | X_n, Y_n)}{\hat{p}(z_0)} \right. \\ &\quad - \frac{1}{N^2} \left[\sum_{n=1}^N \sin \langle t, X_n \rangle \frac{\hat{p}(z_0 | X_n)}{\hat{p}(z_0)} \right] \left[\sum_{n=1}^N \cos \langle s, Y_n \rangle \frac{\hat{p}(z_0 | Y_n)}{\hat{p}(z_0)} \right] \\ &\quad - \frac{1}{N^2} \left[\sum_{n=1}^N \cos \langle t, X_n \rangle \frac{\hat{p}(z_0 | X_n)}{\hat{p}(z_0)} \right] \left[\sum_{n=1}^N \sin \langle s, Y_n \rangle \frac{\hat{p}(z_0 | Y_n)}{\hat{p}(z_0)} \right] \Big\}^2. \end{aligned}$$

It can be rewritten as follows.

$$\Lambda_N = \frac{1}{N^2} \sum_{k,l} \cos(\langle t, X_k \rangle + \langle s, Y_k \rangle) \cos(\langle t, X_l \rangle + \langle s, Y_l \rangle) \frac{\hat{p}(z_0|X_k, Y_k) \hat{p}(z_0|X_l, Y_l)}{\hat{p}(z_0) \hat{p}(z_0)} \quad (13)$$

$$+ \frac{1}{N^4} \left[\sum_{k,l} \cos(\langle t, X_k \rangle) \cos(\langle t, X_l \rangle) \frac{\hat{p}(z_0|X_k) \hat{p}(z_0|X_l)}{\hat{p}(z_0) \hat{p}(z_0)} \right] \left[\sum_{k,l} \cos(\langle s, Y_k \rangle) \cos(\langle s, Y_l \rangle) \frac{\hat{p}(z_0|Y_k) \hat{p}(z_0|Y_l)}{\hat{p}(z_0) \hat{p}(z_0)} \right] \quad (14)$$

$$+ \frac{1}{N^4} \left[\sum_{k,l} \sin(\langle t, X_k \rangle) \sin(\langle t, X_l \rangle) \frac{\hat{p}(z_0|X_k) \hat{p}(z_0|X_l)}{\hat{p}(z_0) \hat{p}(z_0)} \right] \left[\sum_{k,l} \sin(\langle s, Y_k \rangle) \sin(\langle s, Y_l \rangle) \frac{\hat{p}(z_0|Y_k) \hat{p}(z_0|Y_l)}{\hat{p}(z_0) \hat{p}(z_0)} \right] \quad (15)$$

$$- \frac{2}{N^3} \sum_{n,k,l} \cos(\langle t, X_n \rangle + \langle s, Y_n \rangle) \cos(\langle t, X_k \rangle) \cos(\langle s, Y_l \rangle) \frac{\hat{p}(z_0|X_n, Y_n)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|X_k)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|Y_l)}{\hat{p}(z_0)} \quad (16)$$

$$+ \frac{2}{N^3} \sum_{n,k,l} \cos(\langle t, X_n \rangle + \langle s, Y_n \rangle) \sin(\langle t, X_k \rangle) \sin(\langle s, Y_l \rangle) \frac{\hat{p}(z_0|X_n, Y_n)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|X_k)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|Y_l)}{\hat{p}(z_0)} \quad (17)$$

$$- \frac{2}{N^4} \sum_{m,n,k,l} \cos(\langle t, X_m \rangle) \cos(\langle s, Y_n \rangle) \sin(\langle t, X_k \rangle) \sin(\langle s, Y_l \rangle) \frac{\hat{p}(z_0|X_m)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|Y_n)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|X_k)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|Y_l)}{\hat{p}(z_0)} \quad (18)$$

$$+ \frac{1}{N^2} \sum_{k,l} \sin(\langle t, X_k \rangle + \langle s, Y_k \rangle) \sin(\langle t, X_l \rangle + \langle s, Y_l \rangle) \frac{\hat{p}(z_0|X_k, Y_k) \hat{p}(z_0|X_l, Y_l)}{\hat{p}(z_0) \hat{p}(z_0)} \quad (19)$$

$$+ \frac{1}{N^4} \left[\sum_{k,l} \sin(\langle t, X_k \rangle) \cos(\langle s, Y_l \rangle) \frac{\hat{p}(z_0|X_k) \hat{p}(z_0|Y_l)}{\hat{p}(z_0) \hat{p}(z_0)} \right] \left[\sum_{k,l} \sin(\langle t, X_k \rangle) \cos(\langle s, Y_l \rangle) \frac{\hat{p}(z_0|X_k) \hat{p}(z_0|Y_l)}{\hat{p}(z_0) \hat{p}(z_0)} \right] \quad (20)$$

$$+ \frac{1}{N^4} \left[\sum_{k,l} \cos(\langle t, X_k \rangle) \sin(\langle s, Y_l \rangle) \frac{\hat{p}(z_0|X_k) \hat{p}(z_0|Y_l)}{\hat{p}(z_0) \hat{p}(z_0)} \right] \left[\sum_{k,l} \cos(\langle t, X_k \rangle) \sin(\langle s, Y_l \rangle) \frac{\hat{p}(z_0|X_k) \hat{p}(z_0|Y_l)}{\hat{p}(z_0) \hat{p}(z_0)} \right] \quad (21)$$

$$- \frac{2}{N^3} \sum_{n,k,l} \sin(\langle t, X_n \rangle + \langle s, Y_n \rangle) \sin(\langle t, X_k \rangle) \cos(\langle s, Y_l \rangle) \frac{\hat{p}(z_0|X_n, Y_n)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|X_k)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|Y_l)}{\hat{p}(z_0)} \quad (22)$$

$$- \frac{2}{N^3} \sum_{n,k,l} \sin(\langle t, X_n \rangle + \langle s, Y_n \rangle) \cos(\langle t, X_k \rangle) \sin(\langle s, Y_l \rangle) \frac{\hat{p}(z_0|X_n, Y_n)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|X_k)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|Y_l)}{\hat{p}(z_0)} \quad (23)$$

$$+ \frac{2}{N^4} \sum_{m,n,k,l} \sin(\langle t, X_m \rangle) \cos(\langle s, Y_n \rangle) \cos(\langle t, X_k \rangle) \sin(\langle s, Y_l \rangle) \frac{\hat{p}(z_0|X_m)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|Y_n)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|X_k)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|Y_l)}{\hat{p}(z_0)}. \quad (24)$$

Observe the (18)+(24)=0. Now, let $S_1 \doteq (13) + (19)$.

$$S_1 = \frac{1}{N^2} \sum_{k,l} \frac{\hat{p}(z_0|X_k, Y_k) \hat{p}(z_0|X_l, Y_l)}{\hat{p}(z_0) \hat{p}(z_0)} \left[\cos(\langle t, X_k \rangle + \langle s, Y_k \rangle) \cos(\langle t, X_l \rangle + \langle s, Y_l \rangle) \right]$$

$$\begin{aligned}
 & + \sin(\langle t, X_k \rangle + \langle s, Y_k \rangle) \sin(\langle t, X_l \rangle + \langle s, Y_l \rangle) \Big] \\
 &= \frac{1}{N^2} \sum_{k,l}^N \frac{\hat{p}(z_0|X_k, Y_k) \hat{p}(z_0|X_l, Y_l)}{\hat{p}(z_0) \hat{p}(z_0)} [\cos(\langle t, X_k \rangle + \langle s, Y_k \rangle - \langle t, X_l \rangle - \langle s, Y_l \rangle)] \\
 &= \frac{1}{N^2} \sum_{k,l}^N \frac{\hat{p}(z_0|X_k, Y_k) \hat{p}(z_0|X_l, Y_l)}{\hat{p}(z_0) \hat{p}(z_0)} [\cos(\langle t, X_k - X_l \rangle + \langle s, Y_k - Y_l \rangle)] \\
 &= \frac{1}{N^2} \sum_{k,l}^N \frac{\hat{p}(z_0|X_k, Y_k) \hat{p}(z_0|X_l, Y_l)}{\hat{p}(z_0) \hat{p}(z_0)} [\cos(\langle t, X_k - X_l \rangle) \cos(\langle s, Y_k - Y_l \rangle) - \sin(\langle t, X_k - X_l \rangle) \sin(\langle s, Y_k - Y_l \rangle)].
 \end{aligned}$$

Similarly, let $S_2 = (14) + (15) + (20) + (21)$. We have that

$$\begin{aligned}
 S_2 &= \frac{1}{N^4} \sum_{k,l,m,n}^N \frac{\hat{p}(z_0|X_k)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|X_l)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|Y_m)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|Y_n)}{\hat{p}(z_0)} \\
 &\quad \times [\cos(\langle t, X_k \rangle) \cos(\langle t, X_l \rangle) \cos(\langle s, Y_m \rangle) \cos(\langle s, Y_n \rangle) + \sin(\langle t, X_k \rangle) \sin(\langle t, X_l \rangle) \sin(\langle s, Y_m \rangle) \sin(\langle s, Y_n \rangle) \\
 &\quad + \sin(\langle t, X_k \rangle) \sin(\langle t, X_l \rangle) \cos(\langle s, Y_m \rangle) \cos(\langle s, Y_n \rangle) + \cos(\langle t, X_k \rangle) \cos(\langle t, X_l \rangle) \sin(\langle s, Y_m \rangle) \sin(\langle s, Y_n \rangle)] \\
 &= \frac{1}{N^4} \sum_{k,l}^N \frac{\hat{p}(z_0|X_k)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|X_l)}{\hat{p}(z_0)} [\cos(\langle t, X_k \rangle) \cos(\langle t, X_l \rangle) + \sin(\langle t, X_k \rangle) \sin(\langle t, X_l \rangle)] \\
 &\quad \times \sum_{m,n}^N \frac{\hat{p}(z_0|Y_m)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|Y_n)}{\hat{p}(z_0)} [\cos(\langle s, Y_m \rangle) \cos(\langle s, Y_n \rangle) + \sin(\langle s, Y_m \rangle) \sin(\langle s, Y_n \rangle)] \\
 &= \frac{1}{N^4} \left[\sum_{k,l}^N \frac{\hat{p}(z_0|X_k)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|X_l)}{\hat{p}(z_0)} \cos(\langle t, X_k - X_l \rangle) \right] \times \left[\sum_{m,n}^N \frac{\hat{p}(z_0|Y_m)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|Y_n)}{\hat{p}(z_0)} \cos(\langle s, Y_m - Y_n \rangle) \right].
 \end{aligned}$$

Finally, let $-2S_3 = (16) + (17) + (23) + (22)$.

$$\begin{aligned}
 -2S_3 &= -\frac{2}{N^3} \sum_{n,k,l}^N \cos(\langle t, X_n \rangle + \langle s, Y_n \rangle) \cos(\langle t, X_k \rangle) \cos(\langle s, Y_l \rangle) \frac{\hat{p}(z_0|X_n, Y_n)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|X_k)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|Y_l)}{\hat{p}(z_0)} \\
 &\quad + \frac{2}{N^3} \sum_{n,k,l}^N \cos(\langle t, X_n \rangle + \langle s, Y_n \rangle) \sin(\langle t, X_k \rangle) \sin(\langle s, Y_l \rangle) \frac{\hat{p}(z_0|X_n, Y_n)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|X_k)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|Y_l)}{\hat{p}(z_0)} \\
 &\quad - \frac{2}{N^3} \sum_{n,k,l}^N \sin(\langle t, X_n \rangle + \langle s, Y_n \rangle) \sin(\langle t, X_k \rangle) \cos(\langle s, Y_l \rangle) \frac{\hat{p}(z_0|X_n, Y_n)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|X_k)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|Y_l)}{\hat{p}(z_0)} \\
 &\quad - \frac{2}{N^3} \sum_{n,k,l}^N \sin(\langle t, X_n \rangle + \langle s, Y_n \rangle) \cos(\langle t, X_k \rangle) \sin(\langle s, Y_l \rangle) \frac{\hat{p}(z_0|X_n, Y_n)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|X_k)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|Y_l)}{\hat{p}(z_0)} \\
 &= \frac{2}{N^3} \sum_{n,k,l}^N \frac{\hat{p}(z_0|X_n, Y_n)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|X_k)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|Y_l)}{\hat{p}(z_0)} \\
 &\quad \times \left\{ -[\cos(\langle t, X_n \rangle) \cos(\langle s, Y_n \rangle) - \sin(\langle t, X_n \rangle) \sin(\langle s, Y_n \rangle)] \cos(\langle t, X_k \rangle) \cos(\langle s, Y_l \rangle) \right. \\
 &\quad + [\cos(\langle t, X_n \rangle) \cos(\langle s, Y_n \rangle) - \sin(\langle t, X_n \rangle) \sin(\langle s, Y_n \rangle)] \sin(\langle t, X_k \rangle) \sin(\langle s, Y_l \rangle) \\
 &\quad - [\sin(\langle t, X_n \rangle) \cos(\langle s, Y_n \rangle) + \cos(\langle t, X_n \rangle) \sin(\langle s, Y_n \rangle)] \sin(\langle t, X_k \rangle) \cos(\langle s, Y_l \rangle) \\
 &\quad \left. - [\sin(\langle t, X_n \rangle) \cos(\langle s, Y_n \rangle) + \cos(\langle t, X_n \rangle) \sin(\langle s, Y_n \rangle)] \cos(\langle t, X_k \rangle) \sin(\langle s, Y_l \rangle) \right\}
 \end{aligned}$$

$$\begin{aligned}
 &= \frac{2}{N^3} \sum_{n,k,l}^N \frac{\hat{p}(z_0|X_n, Y_n)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|X_k)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|Y_l)}{\hat{p}(z_0)} \times \\
 &\quad \left\{ - [\cos(\langle t, X_n \rangle) \cos(\langle t, X_k \rangle) + \sin(\langle t, X_n \rangle) \sin(\langle t, X_k \rangle)] [\cos(\langle s, Y_n \rangle) \cos(\langle s, Y_l \rangle) + \sin(\langle s, Y_n \rangle) \sin(\langle s, Y_l \rangle)] \right. \\
 &\quad \left. + [\cos(\langle t, X_n \rangle) \sin(\langle t, X_k \rangle) - \sin(\langle t, X_n \rangle) \cos(\langle t, X_k \rangle)] [\cos(\langle s, Y_n \rangle) \sin(\langle s, Y_l \rangle) - \sin(\langle s, Y_n \rangle) \cos(\langle s, Y_l \rangle)] \right\} \\
 &= \frac{2}{N^3} \sum_{n,k,l}^N \frac{\hat{p}(z_0|X_n, Y_n)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|X_k)}{\hat{p}(z_0)} \frac{\hat{p}(z_0|Y_l)}{\hat{p}(z_0)} \\
 &\quad \times \left\{ - \cos(\langle t, X_n - X_k \rangle) \cos(\langle s, Y_n - Y_l \rangle) + \sin(\langle t, X_k - X_n \rangle) \sin(\langle s, Y_l - Y_n \rangle) \right\}.
 \end{aligned}$$

Appendix F. Integrals of S_1, S_2, S_3

Proof of Lemma 13

For brevity, let $C_t(X_k - X_l) \doteq \cos \langle t, X_k - X_l \rangle$, and $S_t(X_k - X_l) \doteq \sin \langle t, X_k - X_l \rangle$.

Proof

$$\begin{aligned}
 &\iint_{D(\psi)} w(t, s) S_1 \, dt \, ds = \\
 &= \iint_{D(\psi)} w(t, s) \frac{1}{N^2} \sum_{k,l}^N \bar{p}(k, l) [C_t(X_k - X_l) C_s(Y_k - Y_l) - S_t(X_k - X_l) S_s(Y_k - Y_l)] \, dt \, ds \\
 &= \frac{1}{N^2} \sum_{k,l}^N \bar{p}(k, l) \iint_{D(\psi)} w(t, s) C_t(X_k - X_l) C_s(Y_k - Y_l) \, dt \, ds \\
 &= \frac{1}{N^2} \sum_{k,l}^N \bar{p}(k, l) \iint_{D(\psi)} dt \, ds \, w(t, s) \{ -1 + C_t(X_k - X_l) \\
 &\quad + C_s(Y_k - Y_l) + (1 - C_t(X_k - X_l))(1 - C_s(Y_k - Y_l)) \} \\
 &= \frac{1}{N^2} \sum_{k,l}^N \bar{p}(k, l) \iint_{D(\psi)} w(t, s) \{ 1 - (1 - C_t(X_k - X_l)) - (1 - C_s(Y_k - Y_l)) \} \, dt \, ds \\
 &\quad + \frac{1}{N^2} \sum_{k,l}^N \bar{p}(k, l) \iint_{D(\psi)} w(t, s) \{ (1 - C_t(X_k - X_l))(1 - C_s(Y_k - Y_l)) \} \, dt \, ds \\
 &= V_{1,N} + \frac{1}{N^2} \sum_{k,l}^N \bar{p}(k, l) |X_k - X_l| |Y_k - Y_l| (1 - G(|X_k - X_l| \psi / c_d)) (1 - G(|Y_k - Y_l| \psi / c_d)).
 \end{aligned}$$

■

$$\iint_{D(\psi)} w(t, s) S_1 \, dt \, ds = \iint_{D(\psi)} w(t, s) \frac{1}{N^2} \sum_{k,l}^N \frac{\hat{p}(z_0|X_k, Y_k) \hat{p}(z_0|X_l, Y_l)}{\hat{p}^2(z_0)}$$

$$\begin{aligned}
 & \times [\cos\langle t, X_k - X_l \rangle \cos\langle s, Y_k - Y_l \rangle - \sin\langle t, X_k - X_l \rangle \sin\langle s, Y_k - Y_l \rangle] dt ds, \\
 &= \frac{1}{N^2} \sum_{k,l}^N \frac{\hat{p}(z_0|X_k, Y_k) \hat{p}(z_0|X_l, Y_l)}{\hat{p}^2(z_0)} \left[\iint_{D(\psi)} w(t, s) \cos\langle t, X_k - X_l \rangle \cos\langle s, Y_k - Y_l \rangle dt ds \right] \\
 &= \frac{1}{N^2} \sum_{k,l}^N \frac{\hat{p}(z_0|X_k, Y_k) \hat{p}(z_0|X_l, Y_l)}{\hat{p}^2(z_0)} \left[\iint_{D(\psi)} w(t, s) \right. \\
 & \quad \times \{1 - (1 - \cos\langle t, X_k - X_l \rangle) - (1 - \cos\langle s, Y_k - Y_l \rangle) + (1 - \cos\langle t, X_k - X_l \rangle)(1 - \cos\langle s, Y_k - Y_l \rangle)\} dt ds \left. \right] \\
 &= \frac{1}{N^2} \sum_{k,l}^N \frac{\hat{p}(z_0|X_k, Y_k) \hat{p}(z_0|X_l, Y_l)}{\hat{p}^2(z_0)} \left[\iint_{D(\psi)} w(t, s) \{1 - (1 - \cos\langle t, X_k - X_l \rangle) - (1 - \cos\langle s, Y_k - Y_l \rangle) dt ds \} \right] \\
 & \quad + \frac{1}{N^2} \sum_{k,l}^N \frac{\hat{p}(z_0|X_k, Y_k) \hat{p}(z_0|X_l, Y_l)}{\hat{p}^2(z_0)} \left[\iint_{D(\psi)} w(t, s) \{(1 - \cos\langle t, X_k - X_l \rangle)(1 - \cos\langle s, Y_k - Y_l \rangle)\} dt ds \right] \\
 &= V_{1,N} + \frac{1}{N^2} \sum_{k,l}^N \frac{\hat{p}(z_0|X_k, Y_k) \hat{p}(z_0|X_l, Y_l)}{\hat{p}^2(z_0)} |X_k - X_l| |Y_k - Y_l| (1 - G(|X_k - X_l|\psi)/c_d)(1 - G(|Y_k - Y_l|\psi)/c_d).
 \end{aligned}$$

$$\begin{aligned}
 & \iint_{D(\psi)} w(t, s) S_2 dt ds \\
 &= \iint_{D(\psi)} w(t, s) \frac{1}{N^4} \left[\sum_{k,l}^N \frac{\hat{p}(z_0|X_k) \hat{p}(z_0|X_l)}{\hat{p}^2(z_0)} \cos\langle t, X_k - X_l \rangle \right] \left[\sum_{m,n}^N \frac{\hat{p}(z_0|Y_n) \hat{p}(z_0|Y_m)}{\hat{p}^2(z_0)} \cos\langle s, Y_m - Y_n \rangle \right] dt ds \\
 &= \frac{1}{N^4} \sum_{k,l,m,n}^N \frac{\hat{p}(z_0|X_k) \hat{p}(z_0|X_l)}{\hat{p}^2(z_0)} \frac{\hat{p}(z_0|Y_n) \hat{p}(z_0|Y_m)}{\hat{p}^2(z_0)} \iint_{D(\psi)} w(t, s) \cos\langle t, X_k - X_l \rangle \cos\langle s, Y_m - Y_n \rangle dt ds \\
 &= \frac{1}{N^4} \sum_{k,l,m,n}^N \frac{\hat{p}(z_0|X_k) \hat{p}(z_0|X_l)}{\hat{p}^2(z_0)} \frac{\hat{p}(z_0|Y_n) \hat{p}(z_0|Y_m)}{\hat{p}^2(z_0)} \\
 & \quad \times \left[\iint_{D(\psi)} w(t, s) \{1 - (1 - \cos\langle t, X_k - X_l \rangle) - (1 - \cos\langle s, Y_m - Y_n \rangle) dt ds \} \right] \\
 &+ \frac{1}{N^4} \sum_{k,l,m,n}^N \frac{\hat{p}(z_0|X_k) \hat{p}(z_0|X_l)}{\hat{p}^2(z_0)} \frac{\hat{p}(z_0|Y_n) \hat{p}(z_0|Y_m)}{\hat{p}^2(z_0)} \\
 & \quad \times \left[\iint_{D(\psi)} w(t, s) \{(1 - \cos\langle t, X_k - X_l \rangle)(1 - \cos\langle s, Y_m - Y_n \rangle)\} dt ds \right] \\
 &= V_{2,N} + \frac{1}{N^4} \sum_{k,l,m,n}^N \frac{\hat{p}(z_0|X_k) \hat{p}(z_0|X_l)}{\hat{p}^2(z_0)} \frac{\hat{p}(z_0|Y_n) \hat{p}(z_0|Y_m)}{\hat{p}^2(z_0)} \\
 & \quad \times |X_k - X_l| |Y_m - Y_n| (1 - G(|X_k - X_l|\psi)/c_d)(1 - G(|Y_m - Y_n|\psi)/c_d).
 \end{aligned}$$

$$\iint_{D(\psi)} w(t, s) S_3 dt ds = \iint_{D(\psi)} w(t, s) \frac{1}{N^3} \sum_{n,k,l}^N \frac{\hat{p}(z_0|X_n, Y_n) \hat{p}(z_0|X_k) \hat{p}(z_0|Y_l)}{\hat{p}^3(z_0)}$$

$$\times [\cos\langle t, X_n - X_k \rangle \cos\langle s, Y_n - Y_l \rangle + \sin\langle t, X_n - X_k \rangle \sin\langle s, Y_n - Y_l \rangle] dt ds.$$

$$\begin{aligned} & \iint_{D(\psi)} w(t, s) S_3 dt ds \\ &= \frac{1}{N^3} \sum_{n,k,l}^N \frac{\hat{p}(z_0|X_n, Y_n) \hat{p}(z_0|X_k) \hat{p}(z_0|Y_l)}{\hat{p}^3(z_0)} \left[\iint_{D(\psi)} w(t, s) \cos\langle t, X_n - X_k \rangle \cos\langle s, Y_n - Y_l \rangle dt ds \right] \\ &= \frac{1}{N^3} \sum_{n,k,l}^N \frac{\hat{p}(z_0|X_n, Y_n) \hat{p}(z_0|X_k) \hat{p}(z_0|Y_l)}{\hat{p}^3(z_0)} \\ & \quad \times \left[\iint_{D(\psi)} w(t, s) \{1 - (1 - \cos\langle t, X_n - X_k \rangle) - (1 - \cos\langle s, Y_n - Y_l \rangle)\} dt ds \right] \\ & \quad + \frac{1}{N^3} \sum_{n,k,l}^N \frac{\hat{p}(z_0|X_n, Y_n) \hat{p}(z_0|X_k) \hat{p}(z_0|Y_l)}{\hat{p}^3(z_0)} \\ & \quad \times \left[\iint_{D(\psi)} w(t, s) \{(1 - \cos\langle t, X_n - X_k \rangle)(1 - \cos\langle s, Y_n - Y_l \rangle)\} dt ds \right] \\ &= V_{3,N} + \frac{1}{N^3} \sum_{n,k,l}^N \frac{\hat{p}(z_0|X_n, Y_n) \hat{p}(z_0|X_k) \hat{p}(z_0|Y_l)}{\hat{p}^3(z_0)} \\ & \quad \times |X_n - X_k| |Y_n - Y_l| (1 - G(|X_n - X_k|\psi)/c_d)(1 - G(|Y_n - Y_l|\psi)/c_d). \end{aligned}$$

We will need the following lemma.

Lemma 23 *If f is a nonnegative bounded function and $\lim_{N \rightarrow \infty} \sup_x |\hat{p}_N(x) - p(x)| = 0$, then*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{l=1}^N \hat{p}_N(z_0|X_l, Y_l) f(X_l, Y_l) = \iint p(z_0, x, y) f(x, y) dx dy.$$

Proof

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{l=1}^N \hat{p}(z_0|X_l, Y_l) f(X_l, Y_l) &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{l=1}^N [\hat{p}(z_0|X_l, Y_l) - p(z_0|X_l, Y_l)] f(X_l, Y_l) \\ & \quad + \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{l=1}^N p(z_0|X_l, Y_l) f(X_l, Y_l). \end{aligned}$$

The first term becomes negligible when N is large, since

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{l=1}^N |\hat{p}_N(z_0|X_l, Y_l) - p(z_0|X_l, Y_l)| f(X_l, Y_l) = 0 \quad (\text{almost surely}),$$

thanks to the $\lim_{N \rightarrow \infty} \sup_x |\hat{p}_N(x) - p(x)| = 0$ assumption and the boundedness of f . ■

It is easy to see that the following lemma holds.

Lemma 24 *For each ψ , $\lim_{N \rightarrow \infty} V_{1,N} + V_{2,N} - 2V_{3,N} = 0$ almost surely, and thus asymptotically the contributions of $V_{1,N}$, $V_{2,N}$ and $V_{3,N}$ terms are negligible.*

Proof

$$\begin{aligned} V_{1,N} &\doteq \frac{1}{N^2} \sum_{k,l}^N \frac{\hat{p}_N(z_0|X_k, Y_k) \hat{p}_N(z_0|X_l, Y_l)}{\hat{p}_N^2(z_0)} \\ &\quad \times \left[\iint_{D(\psi)} w(t, s) \{1 - (1 - \cos\langle t, X_k - X_l \rangle) - (1 - \cos\langle s, Y_k - Y_l \rangle)\} dt ds \right] \\ &= \frac{1}{N} \sum_k^N \frac{\hat{p}_N(z_0|X_k, Y_k)}{\hat{p}_N(z_0)} \frac{1}{N} \sum_l^N \frac{\hat{p}_N(z_0|X_l, Y_l)}{\hat{p}_N(z_0)} \\ &\quad \times \left[\iint_{D(\psi)} w(t, s) \{1 - (1 - \cos\langle t, X_k - X_l \rangle) - (1 - \cos\langle s, Y_k - Y_l \rangle)\} dt ds \right]. \end{aligned}$$

Since

$$\iint w(t, s) \{1 - (1 - \cos\langle t, X_k - X_l \rangle) - (1 - \cos\langle s, Y_k - Y_l \rangle)\} dt ds < \infty,$$

thus

$$\begin{aligned} \lim_{N \rightarrow \infty} \mathcal{V}_{1,N} &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_k^N \frac{\hat{p}_N(z_0|X_k, Y_k)}{\hat{p}_N(z_0)} \iint p(\tilde{x}, \tilde{y}) \frac{p(z_0|\tilde{x}, \tilde{y})}{p(z_0)} \\ &\quad \times \left[\iint_{D(\psi)} w(t, s) \{1 - (1 - \cos\langle t, X_k - \tilde{x} \rangle) - (1 - \cos\langle s, Y_k - \tilde{y} \rangle)\} dt ds \right] d\tilde{x} d\tilde{y} \\ &= \iint p(\hat{x}, \hat{y}) \frac{p(z_0|\hat{x}, \hat{y})}{p(z_0)} \iint p(\tilde{x}, \tilde{y}|z_0) \\ &\quad \times \left[\iint_{D(\psi)} w(t, s) \{1 - (1 - \cos\langle t, \hat{x} - \tilde{x} \rangle) - (1 - \cos\langle s, \hat{y} - \tilde{y} \rangle)\} dt ds \right] d\tilde{x} d\tilde{y} d\hat{x} d\hat{y} \\ &= \iint p(\hat{x}, \hat{y}|z_0) \iint p(\tilde{x}, \tilde{y}|z_0) \left[\iint_{D(\psi)} w(t, s) \{1 - (1 - \cos\langle t, \hat{x} - \tilde{x} \rangle) - (1 - \cos\langle s, \hat{y} - \tilde{y} \rangle)\} dt ds \right] d\tilde{x} d\tilde{y} d\hat{x} d\hat{y}. \end{aligned}$$

Here we used Lemma (23). Similarly,

$$\begin{aligned} V_{2,N} &\doteq \frac{1}{N^4} \sum_{k,l,m,n}^N \frac{\hat{p}(z_0|X_k) \hat{p}(z_0|X_l)}{\hat{p}^2(z_0)} \frac{\hat{p}(z_0|Y_n) \hat{p}(z_0|Y_m)}{\hat{p}^2(z_0)} \\ &\quad \times \left[\iint_{D(\psi)} w(t, s) \{1 - (1 - \cos\langle t, X_k - X_l \rangle) - (1 - \cos\langle s, Y_m - Y_n \rangle)\} dt ds \right]. \end{aligned}$$

$$\begin{aligned} \lim_{N \rightarrow \infty} V_{2,N} &= \iiint p(\hat{x}|z_0) p(\hat{y}|z_0) p(\tilde{x}|z_0) p(\tilde{y}|z_0) \\ &\quad \times \left[\iint_{D(\psi)} w(t, s) \{1 - (1 - \cos\langle t, \hat{x} - \tilde{x} \rangle) - (1 - \cos\langle s, \hat{y} - \tilde{y} \rangle)\} dt ds \right] d\tilde{x} d\tilde{y} d\hat{x} d\hat{y}. \end{aligned}$$

$$\begin{aligned}
 V_{3,N} &\doteq \frac{1}{N^3} \sum_{n,k,l}^N \frac{\hat{p}(z_0|X_n, Y_n) \hat{p}(z_0|X_k) \hat{p}(z_0|Y_l)}{\hat{p}^3(z_0)} \\
 &\quad \times \left[\iint_{D(\psi)} w(t, s) \{1 - (1 - \cos\langle t, X_n - X_k \rangle) - (1 - \cos\langle s, Y_n - Y_l \rangle) dt ds\} \right]. \\
 \lim_{N \rightarrow \infty} V_{3,N} &= \iiint p(\hat{x}, \hat{y}|z_0) p(\tilde{x}|z_0) p(\tilde{y}|z_0) \\
 &\quad \times \left[\iint_{D(\psi)} w(t, s) \{1 - (1 - \cos\langle t, \hat{x} - \tilde{x} \rangle) - (1 - \cos\langle s, \hat{y} - \tilde{y} \rangle) dt ds\} \right] d\tilde{x} d\tilde{y} d\hat{x} d\hat{y}.
 \end{aligned}$$

■

Therefore,

$$\lim_{N \rightarrow \infty} V_{1,N} + V_{2,N} - 2V_{3,N} = 0 \quad (\text{a.s.}).$$

Appendix G. Gaussian

Proof of Lemma 19

Proof

$$\begin{aligned}
 \mathcal{V}(X, Y|Z) &= \\
 &= \int p(z) \iint w(t, s) \left| e^{i[t;s]^T \mu_{XY|z} - \frac{1}{2}[t;s]^T \Sigma_{XY|z} [t;s]^T} \right. \\
 &\quad \left. - \left(e^{it^T \mu_{X|z} - \frac{1}{2}t^T \Sigma_{X|z} t} \right) \left(e^{is^T \mu_{Y|z} - \frac{1}{2}s^T \Sigma_{Y|z} s} \right) \right|^2 dt ds dz \\
 &= \int p(z) \iint \left(c_{d_x} c_{d_y} |t|_{d_x}^{1+d_x} |s|_{d_y}^{1+d_y} \right)^{-1} \left| e^{-\frac{1}{2}[t;s]^T \Sigma_{XY|z} [t;s]^T} - e^{-\frac{1}{2}t^T \Sigma_{X|z} t - \frac{1}{2}s^T \Sigma_{Y|z} s} \right|^2 dt ds dz \\
 &= \iiint p(z) \left(c_{d_x} c_{d_y} |t|_{d_x}^{1+d_x} |s|_{d_y}^{1+d_y} \right)^{-1} e^{-t^T \Sigma_{X|z} t - s^T \Sigma_{Y|z} s} \left(1 + e^{-2t^T \rho_{XY|z} s} - 2e^{-t^T \rho_{XY|z} s} \right) dt ds dz,
 \end{aligned}$$

■

1D Gaussian

Székely et al. (2007) calculated $\mathcal{V}(X, Y)$ for the 1-dimensional Gaussian case. We can use the similar tools to calculate $\mathcal{V}(X, Y|Z)$.

$$\mathcal{V}(X, Y|Z) = \iiint p(z) \frac{1}{\pi^2 t^2 s^2} e^{-t^2 \Sigma_{X|z} - s^2 \Sigma_{Y|z}} (1 + e^{-2ts \rho_{XY|z}} - 2e^{-ts \rho_{XY|z}}) dt ds dz,$$

Let

$$S(\rho_{XY|z}) \doteq \iint e^{-t^2 \Sigma_{X|z} - s^2 \Sigma_{Y|z} - 2ts \rho_{XY|z}} dt ds$$

$$\begin{aligned}
 &= \iint e^{-\frac{1}{2}[t;s]^T \begin{pmatrix} 2\Sigma_{X|z} & 2\rho_{XY|z} \\ 2\rho_{XY|z} & \Sigma_{Y|z} \end{pmatrix} [t;s]} dt ds \\
 &= (2\pi) \left| \begin{pmatrix} 2\Sigma_{X|z} & 2\rho_{XY|z} \\ 2\rho_{XY|z} & \Sigma_{Y|z} \end{pmatrix}^{-1} \right|^{1/2} \\
 &= \frac{\pi}{\sqrt{\Sigma_{X|z}\Sigma_{Y|z} - \rho_{XY|z}^2}},
 \end{aligned}$$

where we used the fact that $\int \exp(-\frac{1}{2}x^T \Sigma^{-1}x) = |2\pi \Sigma|^{1/2}$.

Let

$$\begin{aligned}
 G(\rho_{XY|z}) &\doteq \iint \frac{1}{t^2 s^2} e^{-t^2 \Sigma_{X|z} - s^2 \Sigma_{Y|z}} (1 + e^{-2ts\rho_{XY|z}} - 2e^{-ts\rho_{XY|z}}) dt ds dz, \\
 \mathcal{V}(X, Y|Z) &= \int p(z) \frac{1}{\pi^2} G(\rho_{XY|z}) dz.
 \end{aligned}$$

Now,

$$\begin{aligned}
 \frac{\partial^2}{\partial \rho_{XY|z}^2} G(\rho_{XY|z}) &= \iint \frac{1}{t^2 s^2} e^{-t^2 \Sigma_{X|z} - s^2 \Sigma_{Y|z}} \frac{\partial^2}{\partial \rho_{XY|z}^2} (1 + e^{-2ts\rho_{XY|z}} - 2e^{-ts\rho_{XY|z}}) dt ds dz, \\
 &= \iint \frac{1}{t^2 s^2} e^{-t^2 \Sigma_{X|z} - s^2 \Sigma_{Y|z}} (e^{-2ts\rho_{XY|z}} (4t^2 s^2) - 2e^{-ts\rho_{XY|z}} (t^2 s^2)) dt ds dz, \\
 &= \iint e^{-t^2 \Sigma_{X|z} - s^2 \Sigma_{Y|z}} (4e^{-2ts\rho_{XY|z}} - 2e^{-ts\rho_{XY|z}}) dt ds dz, \\
 &= 4S(\rho_{XY|z}) - 2S(\rho_{XY|z}/2) \\
 &= \frac{4\pi}{\sqrt{\Sigma_{X|z}\Sigma_{Y|z} - \rho_{XY|z}^2}} - \frac{2\pi}{\sqrt{\Sigma_{X|z}\Sigma_{Y|z} - \rho_{XY|z}^2/4}} \\
 &= \frac{4\pi}{\sqrt{\Sigma_{X|z}\Sigma_{Y|z} - \rho_{XY|z}^2}} - \frac{4\pi}{\sqrt{4\Sigma_{X|z}\Sigma_{Y|z} - \rho_{XY|z}^2}}.
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 G(\rho_{XY|z}) &= \int_0^{\rho_{XY|z}} \int_0^\lambda \frac{4\pi}{\sqrt{\Sigma_{X|z}\Sigma_{Y|z} - \kappa^2}} - \frac{4\pi}{\sqrt{4\Sigma_{X|z}\Sigma_{Y|z} - \kappa^2}} d\kappa d\lambda \\
 &= 4\pi \int_0^{\rho_{XY|z}} \arcsin\left(\frac{\lambda}{\sqrt{\Sigma_{X|z}\Sigma_{Y|z}}}\right) - \arcsin\left(\frac{\lambda}{\sqrt{4\Sigma_{X|z}\Sigma_{Y|z}}}\right) d\lambda \\
 &= 4\pi \left(\sqrt{\Sigma_{X|z}\Sigma_{Y|z} - \rho_{XY|z}^2} + \rho_{XY|z} \arcsin\left(\frac{\rho_{XY|z}}{\sqrt{\Sigma_{X|z}\Sigma_{Y|z}}}\right) - \sqrt{\Sigma_{X|z}\Sigma_{Y|z}} \right) \\
 &\quad - 4\pi \left(\sqrt{4\Sigma_{X|z}\Sigma_{Y|z} - \rho_{XY|z}^2} + \rho_{XY|z} \arcsin\left(\frac{\rho_{XY|z}}{\sqrt{4\Sigma_{X|z}\Sigma_{Y|z}}}\right) - \sqrt{4\Sigma_{X|z}\Sigma_{Y|z}} \right).
 \end{aligned}$$

Above we used the following facts:

$$\begin{aligned}
 \int \frac{1}{\sqrt{1 - x^2/b^2}} dx &= b \arcsin(x/b), \\
 \int \frac{1}{\sqrt{b^2 - x^2}} dx &= \arcsin(x/b),
 \end{aligned}$$

$$\int_0^\lambda \frac{1}{\sqrt{\Sigma_{X|z}\Sigma_{Y|z} - \kappa^2}} d\kappa = \arcsin\left(\frac{\lambda}{\sqrt{\Sigma_{X|z}\Sigma_{Y|z}}}\right),$$

$$\int_0^b \arcsin(x/a) dx = \sqrt{a^2 - b^2} + b \arcsin(b/a) - a.$$

Hence, we have that

$$\begin{aligned} \mathcal{V}(X, Y|Z) &= \int p(z) \frac{1}{\pi^2} G(\rho_{XY|z}) dz \\ &= \int p(z) \frac{4}{\pi} \left(\sqrt{\Sigma_{X|z}\Sigma_{Y|z} - \rho_{XY|z}^2} + \rho_{XY|z} \arcsin\left(\frac{\rho_{XY|z}}{\sqrt{\Sigma_{X|z}\Sigma_{Y|z}}}\right) - \sqrt{\Sigma_{X|z}\Sigma_{Y|z}} \right) dz \\ &\quad - \int p(z) \frac{4}{\pi} \left(\sqrt{4\Sigma_{X|z}\Sigma_{Y|z} - \rho_{XY|z}^2} + \rho_{XY|z} \arcsin\left(\frac{\rho_{XY|z}}{\sqrt{4\Sigma_{X|z}\Sigma_{Y|z}}}\right) - \sqrt{4\Sigma_{X|z}\Sigma_{Y|z}} \right) dz. \end{aligned}$$

Interestingly, when we deal with normal distribution, then these conditional covariances do not depend on the actual value of $Z = z$. In turn, we can treat them as constants in the integral above.

$$\begin{aligned} \mathcal{V}(X, Y|Z) &= \frac{4}{\pi} \left(\sqrt{\Sigma_{X|z}\Sigma_{Y|z} - \rho_{XY|z}^2} + \rho_{XY|z} \arcsin\left(\frac{\rho_{XY|z}}{\sqrt{\Sigma_{X|z}\Sigma_{Y|z}}}\right) - \sqrt{\Sigma_{X|z}\Sigma_{Y|z}} \right) \\ &\quad - \frac{4}{\pi} \left(\sqrt{4\Sigma_{X|z}\Sigma_{Y|z} - \rho_{XY|z}^2} + \rho_{XY|z} \arcsin\left(\frac{\rho_{XY|z}}{\sqrt{4\Sigma_{X|z}\Sigma_{Y|z}}}\right) - \sqrt{4\Sigma_{X|z}\Sigma_{Y|z}} \right). \end{aligned}$$

Introduce the following notation for the joint covariance matrix of X, Y, Z :

$$\Sigma_{[X;Y;Z]} = \begin{pmatrix} \Sigma_X & \Sigma_{XY} & \Sigma_{XZ} \\ \Sigma_{YX} & \Sigma_Y & \Sigma_{YZ} \\ \Sigma_{ZX} & \Sigma_{ZY} & \Sigma_Z \end{pmatrix}.$$

Let

$$\begin{aligned} \Sigma_{11} &\doteq \Sigma_{[X;Y]} = \begin{pmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{pmatrix}, \\ \Sigma_{12} &\doteq \begin{pmatrix} \Sigma_{XZ} \\ \Sigma_{YZ} \end{pmatrix}, \quad \Sigma_{21} \doteq \begin{pmatrix} \Sigma_{ZX} & \Sigma_{ZY} \end{pmatrix}, \\ \Sigma_{22} &= \Sigma_Z. \end{aligned}$$

Now, we have that

$$\Sigma_{[X;Y]|z} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \begin{pmatrix} \Sigma_{X|z} & \rho_{XY|z} \\ \rho_{YX|z} & \Sigma_{Y|z} \end{pmatrix}.$$

Using these formula we can easily calculate $\mathcal{V}(X|Z = z)$, too.

$$\begin{aligned} \mathcal{V}(X|Z = z) &= \iint w(t, s) \left| \iint e^{i\langle t, x \rangle + i\langle s, x \rangle} p(x|z) dx - \left(\int e^{i\langle t, x \rangle} p(x|z) dx \right) \left(\int e^{i\langle s, x \rangle} p(x|z) dx \right) \right|^2 dt ds \\ &= \iint w(t, s) \left| e^{i(t+s)^T \mu_{X|z} - \frac{1}{2}(t+s)^T \Sigma_{X|z} (t+s)} - \left(e^{it^T \mu_{X|z} - \frac{1}{2}t^T \Sigma_{X|z} t} \right) \left(e^{is^T \mu_{X|z} - \frac{1}{2}s^T \Sigma_{X|z} s} \right) \right|^2 dt ds \end{aligned}$$

$$\begin{aligned}
 &= \iint w(t, s) \left| e^{-\frac{1}{2}(t+s)^T \Sigma_{X|z}(t+s)} - e^{-\frac{1}{2}t^T \Sigma_{X|z}t - \frac{1}{2}s^T \Sigma_{X|z}s} \right|^2 dt ds \\
 &= \iint w(t, s) \left| e^{-\frac{1}{2}[t;s]^T \begin{pmatrix} \Sigma_{X|z} & \Sigma_{X|z} \\ \Sigma_{X|z} & \Sigma_{X|z} \end{pmatrix} [t;s]} - e^{-\frac{1}{2}t^T \Sigma_{X|z}t - \frac{1}{2}s^T \Sigma_{X|z}s} \right|^2 dt ds \\
 &= \frac{4}{\pi} \left(\sqrt{\Sigma_{X|z} \Sigma_{Y|z} - \rho_{XY|z}^2} + \rho_{XY|z} \arcsin \left(\frac{\rho_{XY|z}}{\sqrt{\Sigma_{X|z} \Sigma_{Y|z}}} \right) - \sqrt{\Sigma_{X|z} \Sigma_{Y|z}} \right) \\
 &\quad - \frac{4}{\pi} \left(\sqrt{4\Sigma_{X|z} \Sigma_{Y|z} - \rho_{XY|z}^2} + \rho_{XY|z} \arcsin \left(\frac{\rho_{XY|z}}{\sqrt{4\Sigma_{X|z} \Sigma_{Y|z}}} \right) - \sqrt{4\Sigma_{X|z} \Sigma_{Y|z}} \right),
 \end{aligned}$$

where $\rho_{XY|z} = \Sigma_{X|z} = \Sigma_{Y|z}$.