Journal of Multivariate Analysis 168 (2018) 304–322

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Generalizing distance covariance to measure and test multivariate mutual dependence via complete and incomplete V-statistics

Ze Jin *, David S. Matteson

Department of Statistical Science, Cornell University, Ithaca, NY 14853, USA

ARTICLE INFO

Article history:
Received 29 December 2017
Available online 20 August 2018

AMS 2000 subject classifications:
62G10
62H20
62G20
60E10

Keywords:
Characteristic functions
Distance covariance
Multivariate analysis
Mutual independence
V-statistics

ABSTRACT

We propose three new measures of mutual dependence between multiple random vectors. Each measure is zero if and only if the random vectors are mutually independent. The first generalizes distance covariance from pairwise dependence to mutual dependence, while the other two measures are sums of squared distance covariances. The proposed measures share similar properties and asymptotic distributions with distance covariance, and capture non-linear and non-monotone mutual dependence between the random vectors. Inspired by complete and incomplete V-statistics, we define empirical and simplified empirical measures as a trade-off between the complexity and statistical power when testing mutual independence. The implementation of corresponding tests is demonstrated by both simulation results and real data examples.

1. Introduction

Let $X = (X_1, \ldots, X_d)$ be a set of variables where each component is a random vector, and let $\mathbf{X} = \{X^1, \ldots, X^n\}$ be a random sample from F_X, the joint distribution of X. We are interested in testing the hypotheses

$$\mathcal{H}_0 : X_1, \ldots, X_d \text{ are mutually independent} \quad \text{vs.} \quad \mathcal{H}_A : X_1, \ldots, X_d \text{ are dependent.}$$

This problem has many applications, including independent component analysis [16,26], graphical models [8,10,22,23], naive Bayes classifiers [38,40], causal inference [5,25], etc. It has been studied under different settings and assumptions, including pairwise ($d = 2$) and mutual ($d \geq 2$) independence, univariate ($X_1, \ldots, X_d \in \mathbb{R}$) and multivariate ($X_1 \in \mathbb{R}^{p_1}, \ldots, X_d \in \mathbb{R}^{p_d}$) components, and more. Here we consider the general case where $X_1, \ldots, X_d$ are not assumed jointly normal.

The most extensively studied case is pairwise independence with univariate components ($X_1, X_2 \in \mathbb{R}$). Rank correlation is considered as a nonparametric counterpart to Pearson's product–moment correlation [28], including Kendall's tau [19], Spearman's rho [32], etc. Bergsma and Dassios [2] proposed a test based on an extension of Kendall's tau, testing an equivalent condition to $\mathcal{H}_0$. Additionally, Hoeffding [15] proposed a nonparametric test based on marginal and joint distribution functions, testing a necessary condition to investigate $\mathcal{H}_0$.

* Corresponding author.
E-mail addresses: zj58@cornell.edu (Z. Jin), matteson@cornell.edu (D.S. Matteson).

https://doi.org/10.1016/j.jmva.2018.08.006
0047-259X/

For pairwise independence with multivariate components ($X_1 \in \mathbb{R}^{p_1}, X_2 \in \mathbb{R}^{p_2}$), Székely et al., Székely and Rizzo [37,34] proposed a test based on distance covariance with fixed p_1, p_2 and $n \to \infty$ testing an equivalent condition to $\mathcal{H}_0$; this has been extended to martingale difference divergence in [31] with [17] testing conditional mean independence. Under the same setting, Gretton et al. [13] proposed a test based on Hilbert−Schmidt independence criterion (HSIC), which is 0 if and only if pairwise independence holds. Further, Székely and Rizzo [35] proposed a t-test based on a modified distance covariance for the setting in which n is finite and $p_1, p_2 \to \infty$, testing an equivalent condition to $\mathcal{H}_0$ as well.

For mutual independence with univariate components ($X_1, \ldots, X_d \in \mathbb{R}$), one natural way to extend the pairwise rank correlation to multiple components is to collect the rank correlations between all pairs of components, and examine the norm ($\mathcal{L}_2, \mathcal{L}_\infty$) of this collection. Leung and Drton [21] proposed a test based on the $\mathcal{L}_2$ norm with $n, d \to \infty$, and $d/n \to \gamma \in (0, \infty)$, and Han et al. [14] proposed a test based on the $\mathcal{L}_\infty$ norm with $n, d \to \infty$, and $d/n \to \gamma \in [0, \infty]$. Each is testing a necessary condition to $\mathcal{H}_0$, in general.

The challenging scenario of mutual independence with multivariate components ($X_1 \in \mathbb{R}^{p_1}, \ldots, X_d \in \mathbb{R}^{p_d}$) has not been well studied. Using a combinatorial formula of Möbius, Genest and Rémillard [12], Genest et al. [11] and Kojadinovic and Holmes [20] proposed tests based on ranks and Cramér−von Mises statistics, testing a necessary and sufficient condition to $\mathcal{H}_0$. Bilodeau and Lafaye de Micheaux [3] proposed a test based on characteristic functions under the assumption of normal margins and made a connection to V-statistics, Beran et al. [1] proposed a test based on half-space probabilities, Bilodeau and Nangue [4] and Fan et al. [9] proposed tests based on characteristic functions, testing an equivalent condition to $\mathcal{H}_0$, all with fixed $d, p_1, \ldots, p_d$ and $n \to \infty$. Under the same setting, Pfister et al. [30] proposed a test based on d-variable Hilbert−Schmidt independence criterion (dHSIC), which originates from HSIC and is 0 if and only if mutual independence holds. Yao et al. [39] proposed a test based on distance covariance between all pairs of components with $n, d \to \infty$, testing a necessary condition to $\mathcal{H}_0$. Inspired by distance covariance in Székely et al. [37], we propose new tests based on three measures of mutual dependence, i.e., complete measure, asymmetric measure and symmetric measure, with fixed $d, p_1, \ldots, p_d$ and $n \to \infty$ in this paper, testing an equivalent condition to $\mathcal{H}_0$. All computational complexities in this paper make no reference to the dimensions $d, p_1, \ldots, p_d$, as they are treated as constants.

Our measures of mutual dependence involve V-statistics, and are 0 if and only if mutual independence holds. They belong to energy statistics [36], and share many statistical properties with distance covariance. Our complete measure and dHSIC [30] both contain V-statistics with a similar structure. The main difference is that Pfister et al. [30] pursue kernel methods and overcome the computation bottleneck by resampling and Gamma approximation, while we take advantage of characteristic functions and resort to incomplete V-statistics. Our asymmetric and symmetric measures, and measures in Bilodeau and Nangue [4] and Fan et al. [9] all use characteristic functions. The main difference is that Bilodeau and Nangue [4] and Fan et al. [9] include all pairwise dependences from the Möbius decomposition, while we only consider a subset of pairwise dependences from it.

The weakness of testing mutual independence by a necessary condition, all pairwise independences, motivates our work on measures of mutual dependence, which is demonstrated by examples in Section 6: If we directly test mutual independence based on the measures of mutual dependence proposed in this paper, we successfully detect mutual dependence. Alternatively, if we check all pairwise independences based on distance covariance, we fail to detect any pairwise dependence, and mistakenly conclude that mutual independence holds probably because the mutual effect averages out when we narrow down to a pair.

The rest of this paper is organized as follows. In Section 2, we give a brief overview of distance covariance. In Section 3, we generalize distance covariance to complete measure of mutual dependence, with its properties and asymptotic distributions derived. In Section 4, we propose asymmetric and symmetric measures of mutual dependence, defined as sums of squared distance covariances. We present simulation results in Section 5, followed by synthetic and real data analysis in Section 6; an accompanying R package `EDMeasure` [18] is available on CRAN. Finally, Section 7 is the summary of our work. All proofs have been moved to Appendix.

The following notations will be used throughout this paper. Let $(\cdot, \ldots, \cdot)$ denote a concatenation of (vector) components into a vector. Let $t = (t_1, \ldots, t_d)$, $t^0 = (t_1^0, \ldots, t_d^0)$, $X = (X_1, \ldots, X_d) \in \mathbb{R}^p$, where for each $j \in \{1, \ldots, d\}$, $t_j, t_j^0, X_j \in \mathbb{R}^{p_j}$ and $p = p_1 + \cdots + p_d$ is the total dimension.

The assumed "X" under $\mathcal{H}_0$ is denoted by $\widetilde{X} = (\widetilde{X}_1, \ldots, \widetilde{X}_d)$, where for each $j \in \{1, \ldots, d\}$, $\widetilde{X}_j \overset{d}{=} X_j$, $\widetilde{X}_1, \ldots, \widetilde{X}_d$ are mutually independent, and $X, \widetilde{X}$ are independent. Let X', X'' be independent copies of X, i.e., $X, X', X'' \overset{\text{iid}}{\sim} F_X$, and $\widetilde{X}', \widetilde{X}''$ be independent copies of $\widetilde{X}$, i.e., $\widetilde{X}, \widetilde{X}', \widetilde{X}'' \overset{\text{iid}}{\sim} F_{\widetilde{X}}$.

The Euclidean norm of vector $X \in \mathbb{R}^p$ is denoted by $|X|_p$. Let the weighted $\mathcal{L}_2$ norm $\|\cdot\|_w$ of complex-valued function $\eta(t)$ be defined by $\|\eta(t)\|_w^2 = \int_{\mathbb{R}^p} |\eta(t)|^2 w(t)\, dt$ where $|\eta(t)|^2 = \eta(t)\overline{\eta(t)}$, $\overline{\eta(t)}$ is the complex conjugate of $\eta(t)$, and $w(t)$ is any positive weight function for which the integral exists.

Given the iid sample $\mathbf{X}$ from F_X, let $\mathbf{X}_j = \{X_j^k : k \in \{1, \ldots, n\}\}$ denote the corresponding iid sample from F_{X_j}, with $j \in \{1, \ldots, d\}$, such that $\mathbf{X} = \{\mathbf{X}_1, \ldots, \mathbf{X}_d\}$. Denote the joint characteristic functions of X and $\widetilde{X}$ as

$$\phi_X(t) = \mathrm{E}(e^{i\langle t, X\rangle}) \quad \text{and} \quad \phi_{\widetilde{X}}(t) = \prod_{j=1}^d \mathrm{E}(e^{i\langle t_j, X_j\rangle}),$$

and denote the empirical versions of $\phi_X(t)$ and $\phi_{\widetilde{X}}(t)$ as

$$\phi_X^n(t) = \frac{1}{n}\sum_{k=1}^n e^{i\langle t, X^k\rangle} \quad \text{and} \quad \phi_{\widetilde{X}}^n(t) = \prod_{j=1}^d \left(\frac{1}{n}\sum_{k=1}^n e^{i\langle t_j, X_j^k\rangle}\right).$$

For illustration purpose, we make a toy example with two components ($d = 2$), two dimensions each ($p = 4$), and two samples ($n = 2$), to exemplify the definitions of empirical measures proposed in this paper.

2. Distance covariance

Székely et al. [37] proposed distance covariance to capture non-linear and non-monotone pairwise dependence between two random vectors ($X_1 \in \mathbb{R}^{p_1}, X_2 \in \mathbb{R}^{p_2}$). The vectors X_1, X_2 are pairwise independent if and only if $\phi_X(t) = \phi_{X_1}(t_1)\phi_{X_2}(t_2)$, for all t, which is equivalent to $\int_{\mathbb{R}^p} |\phi_X(t) - \phi_{\widetilde{X}}(t)|^2 w(t)\,dt = 0$ for all $w(t) > 0$ if the integral exists. The weight functions of the form

$$w_0(t,m) = \left\{K(p_1; m)K(p_2; m)|t_1|_{p_1}^{p_1+m}|t_2|_{p_2}^{p_2+m}\right\}^{-1}$$

make the integral a finite and meaningful quantity composed of mth moments according to Lemma 1 in [33], where $K(q,m) = 2\pi^{q/2}\Gamma(1-m/2)/[m2^m\Gamma\{(q+m)/2\}]$, and Γ is the gamma function.

The non-negative distance covariance $\mathcal{V}(X)$ is defined by

$$\mathcal{V}^2(X) = \|\phi_X(t) - \phi_{\widetilde{X}}(t)\|_{w_0}^2 = \int_{\mathbb{R}^p} |\phi_X(t) - \phi_{\widetilde{X}}(t)|^2 w_0(t)\,dt,$$

where

$$w_0(t) = \left(K_{p_1}K_{p_2}|t_1|_{p_1}^{p_1+1}|t_2|_{p_2}^{p_2+1}\right)^{-1},$$

with $m = 1$ and $K_q = K(q, 1)$, while any following result can be generalized to $m \in (0, 2)$. If $\mathrm{E}|X|_p < \infty$, then $\mathcal{V}(X) \in [0,\infty)$, and $\mathcal{V}(X) = 0$ if and only if X_1, X_2 are pairwise independent.

The non-negative empirical distance covariance $\mathcal{V}_n(\mathbf{X})$ is defined by

$$\mathcal{V}_n^2(\mathbf{X}) = \|\phi_X^n(t) - \phi_{\widetilde{X}}^n(t)\|_{w_0}^2 = \int_{\mathbb{R}^p} |\phi_X^n(t) - \phi_{\widetilde{X}}^n(t)|^2 w_0(t)\,dt.$$

Calculating $\mathcal{V}_n^2(\mathbf{X})$ via the symmetry of Euclidean distances has the time complexity $O(n^2)$. Some asymptotic properties of $\mathcal{V}_n(\mathbf{X})$ are as follows when $\mathrm{E}|X|_p < \infty$.

(i) $\mathcal{V}_n(\mathbf{X}) \xrightarrow{\text{a.s.}} \mathcal{V}(X)$ as $n \to \infty$.

(ii) Under $\mathcal{H}_0$, $n\mathcal{V}_n^2(\mathbf{X}) \rightsquigarrow \|\zeta(t)\|_{w_0}^2$ as $n \to \infty$, where $\zeta(t)$ is a complex-valued Gaussian process with mean zero and covariance function

$$R(t,t^0) = \left\{\phi_{X_1}(t_1 - t_1^0) - \phi_{X_1}(t_1)\overline{\phi_{X_1}(t_1^0)}\right\}\left\{\phi_{X_2}(t_2 - t_2^0) - \phi_{X_2}(t_2)\overline{\phi_{X_2}(t_2^0)}\right\}.$$

(iii) Under $\mathcal{H}_A$, $n\mathcal{V}_n^2(\mathbf{X}) \xrightarrow{\text{a.s.}} \infty$ as $n \to \infty$.

3. Complete measure of mutual dependence

Generalizing the idea of distance covariance, we propose complete measure of mutual dependence to capture non-linear and non-monotone mutual dependence between multiple random vectors ($X_1 \in \mathbb{R}^{p_1}, \ldots, X_d \in \mathbb{R}^{p_d}$). Such vectors $X_1, \ldots, X_d$ are mutually independent if and only if $\phi_X(t) = \phi_{X_1}(t_1)\cdots\phi_{X_d}(t_d) = \phi_{\widetilde{X}}(t)$, for all t, which is equivalent to $\int_{\mathbb{R}^p} |\phi_X(t) - \phi_{\widetilde{X}}(t)|^2 w(t)\,dt = 0$ for all $w(t) > 0$ if the integral exists. In the following, we will present two weights $w_1(t)$, $w_2(t)$, and elaborate on the reason why we disregard $w_2(t)$ for computational efficiency later.

We put all components together instead of separating them, and choose the weight function

$$w_1(t) = (K_p|t|_p^{p+1})^{-1}.$$

Definition 1. The complete measure of mutual dependence $\mathcal{Q}(X)$ is defined by

$$\mathcal{Q}(X) = \|\phi_X(t) - \phi_{\widetilde{X}}(t)\|_{w_1}^2 = \int_{\mathbb{R}^p} |\phi_X(t) - \phi_{\widetilde{X}}(t)|^2 w_1(t)\,dt.$$

We can show an equivalence to mutual independence based on $\mathcal{Q}(X)$ according to Lemma 1 in [33].

Theorem 1. *If* $\mathrm{E}|X|_p < \infty$, *then* $\mathcal{Q}(X) \in [0, \infty)$, *and* $\mathcal{Q}(X) = 0$ *if and only if* $X_1, \ldots, X_d$ *are mutually independent. In addition,* $\mathcal{Q}(X)$ *has an interpretation as expectations*

$$\mathcal{Q}(X) = \mathrm{E}|X - \widetilde{X}'|_p + \mathrm{E}|X' - \widetilde{X}|_p - \mathrm{E}|X - X'|_p - \mathrm{E}|\widetilde{X} - \widetilde{X}'|_p.$$

It is straightforward to estimate $\mathcal{Q}(X)$ by replacing the characteristic functions with the empirical characteristic functions from the sample.

Definition 2. The empirical complete measure of mutual dependence $\mathcal{Q}_n(\mathbf{X})$ is defined by

$$\mathcal{Q}_n(\mathbf{X}) = \|\phi_X^n(t) - \phi_{\widetilde{X}}^n(t)\|_{w_1}^2 = \int_{\mathbb{R}^p} |\phi_X^n(t) - \phi_{\widetilde{X}}^n(t)|^2 w_1(t)\, dt.$$

Lemma 1. $\mathcal{Q}_n(\mathbf{X})$ *has an interpretation as a complete V-statistic, viz.*

$$\mathcal{Q}_n(\mathbf{X}) = \frac{2}{n^{d+1}} \sum_{k,\ell_1,\ldots,\ell_d=1}^{n} |X^k - (X_1^{\ell_1}, \ldots, X_d^{\ell_d})|_p + \frac{1}{n^2} \sum_{k,\ell=1}^{n} |X^k - X^\ell|_p$$
$$- \frac{1}{n^{2d}} \sum_{k_1,\ldots,k_d,\ell_1,\ldots,\ell_d=1}^{n} |(X_1^{k_1}, \ldots, X_d^{k_d}) - (X_1^{\ell_1}, \ldots, X_d^{\ell_d})|_p,$$

whose naive implementation has the time complexity $O(n^{2d})$.

With respect to the toy example, the first summation term in $\mathcal{Q}_n(\mathbf{X})$ contains 8 summands $|(X_1^k - X_1^{\ell_1}, X_2^k - X_2^{\ell_2})|_4$, for all $k, \ell_1, \ell_2 \in \{1, 2\}$, including $|(X_1^1 - X_1^1, X_2^1 - X_2^2)|_4$ and $|(X_1^1 - X_1^2, X_2^1 - X_2^2)|_4$.

In view of the definition of distance covariance, it seems natural to define the measure using the weight function

$$w_2(t) = (K_{p_1} \cdots K_{p_d} |t_1|_{p_1}^{p_1+1} \cdots |t_d|_{p_d}^{p_d+1})^{-1},$$

which equals $w_0(t)$ when $d = 2$. Given the weight function $w_2(t)$, we can define the squared distance covariance of mutual dependence $\mathcal{U}(X) = \|\phi_X(t) - \phi_{\widetilde{X}}(t)\|_{w_2}^2$ and its empirical counterpart $\mathcal{U}_n(\mathbf{X}) = \|\phi_X^n(t) - \phi_{\widetilde{X}}^n(t)\|_{w_2}^2$, which equal $\mathcal{V}^2(X)$ and $\mathcal{V}_n^2(\mathbf{X})$ when $d = 2$. The naive implementation of $\mathcal{U}_n(\mathbf{X})$ has the time complexity $O(n^{d+1})$.

The reason to favor $w_1(t)$ instead of $w_2(t)$ is a trade-off between the moment condition and time complexity. We often cannot afford the time complexity of $\mathcal{Q}_n(\mathbf{X})$ or $\mathcal{U}_n(\mathbf{X})$, and have to simplify them through incomplete V-statistics. An incomplete V-statistic is obtained by sampling the terms of a complete V-statistic, where the summation extends over only a subset of the tuple of indices. To simplify by replacing complete V-statistics with incomplete V-statistics, $\mathcal{U}_n(\mathbf{X})$ requires the additional dth moment condition $\mathrm{E}(|X_1|_{p_1} \cdots |X_d|_{p_d}) < \infty$, while $\mathcal{Q}_n(\mathbf{X})$ does not require any other condition in addition to the first moment condition $\mathrm{E}|X|_p < \infty$. Thus, we can reduce the complexity of $\mathcal{Q}_n(\mathbf{X})$ to $O(n^2)$ with a weaker condition, which makes $\mathcal{Q}(X)$ and $\mathcal{Q}_n(\mathbf{X})$ from $w_1(t)$ a more general solution. As an example, suppose $X_1 = \cdots = X_d \in \mathbb{R}^1$, then $\mathrm{E}|X|_p < \infty$ only requires finite first moment as $\mathrm{E}|X_1| < \infty$, while $\mathrm{E}(|X_1|_{p_1} \cdots |X_d|_{p_d}) < \infty$ requires finite dth moment as $\mathrm{E}|X_1|^d < \infty$.

Moreover, we define the simplified empirical version of $\phi_{\widetilde{X}}(t)$ as

$$\phi_{\widetilde{X}}^{n\star}(t) = \frac{1}{n} \sum_{k=1}^{n} e^{i \sum_{j=1}^{d} \langle t_j, X_j^{k+j-1} \rangle} = \frac{1}{n} \sum_{k=1}^{n} e^{i \langle t, (X_1^k, \ldots, X_d^{k+d-1}) \rangle},$$

in order to substitute $\phi_{\widetilde{X}}^n(t)$ for simplification, where X_j^{n+k} is interpreted as X_j^k for $k > 0$.

Definition 3. The simplified empirical complete measure of mutual dependence $\mathcal{Q}_n^\star(\mathbf{X})$ is defined by

$$\mathcal{Q}_n^\star(\mathbf{X}) = \|\phi_X^n(t) - \phi_{\widetilde{X}}^{n\star}(t)\|_{w_1}^2 = \int_{\mathbb{R}^p} |\phi_X^n(t) - \phi_{\widetilde{X}}^{n\star}(t)|^2 w_1(t)\, dt.$$

Lemma 2. $\mathcal{Q}_n^\star(\mathbf{X})$ *has an interpretation as an incomplete V-statistic, viz.*

$$\mathcal{Q}_n^\star(\mathbf{X}) = \frac{2}{n^2} \sum_{k,\ell=1}^{n} |X^k - (X_1^\ell, \ldots, X_d^{\ell+d-1})|_p + \frac{1}{n^2} \sum_{k,\ell=1}^{n} |X^k - X^\ell|_p$$
$$- \frac{1}{n^2} \sum_{k,\ell=1}^{n} |(X_1^k, \ldots, X_d^{k+d-1}) - (X_1^\ell, \ldots, X_d^{\ell+d-1})|_p,$$

whose naive implementation has the time complexity $O(n^2)$.

With respect to the toy example, the first summation term in $\mathcal{Q}_n^\star(\mathbf{X})$ contains 4 summands $|(X_1^k - X_1^\ell, X_2^k - X_2^{\ell+1})|_4$, for all $k, \ell \in \{1, 2\}$, including $|(X_1^1 - X_1^1, X_2^1 - X_2^2)|_4$ but not $|(X_1^1 - X_1^2, X_2^1 - X_2^2)|_4$.

Using a similar derivation to Theorems 2 and 5 in [37], some asymptotic distributions of $\mathcal{Q}_n(\mathbf{X})$, $\mathcal{Q}_n^\star(\mathbf{X})$ are obtained as follows.

Theorem 2. *If* $\mathrm{E}|X|_p < \infty$, *then, as* $n \to \infty$, $\mathcal{Q}_n(\mathbf{X}) \xrightarrow{\text{a.s.}} \mathcal{Q}(X)$ *and* $\mathcal{Q}_n^\star(\mathbf{X}) \xrightarrow{\text{a.s.}} \mathcal{Q}(X)$.

Theorem 3. *If* $\mathrm{E}|X|_p < \infty$, *then under* $\mathcal{H}_0$, *we have, as* $n \to \infty$, $n\mathcal{Q}_n(\mathbf{X}) \rightsquigarrow \|\zeta(t)\|_{w_1}^2$ *and* $n\mathcal{Q}_n^\star(\mathbf{X}) \rightsquigarrow \|\zeta^\star(t)\|_{w_1}^2$, *where* $\zeta(t)$, $\zeta^\star(t)$ *are complex-valued Gaussian processes with mean zero and covariance functions*

$$R(t, t^0) = \prod_{j=1}^d \phi_{X_j}(t_j - t_j^0) + (d-1)\prod_{j=1}^d \phi_{X_j}(t_j)\overline{\phi_{X_j}(t_j^0)} - \sum_{j=1}^d \phi_{X_j}(t_j - t_j^0)\prod_{\ell \neq j}\phi_{X_\ell}(t_\ell)\overline{\phi_{X_\ell}(t_\ell^0)},$$
$$R^\star(t, t^0) = 2R(t, t^0).$$

Under $\mathcal{H}_A$, *we have* $n\mathcal{Q}_n(\mathbf{X}) \xrightarrow{\text{a.s.}} \infty$ *and* $n\mathcal{Q}_n^\star(\mathbf{X}) \xrightarrow{\text{a.s.}} \infty$ *as* $n \to \infty$.

Theorems 2 and 3 are closely connected in the sense that $n\mathcal{Q}_n(\mathbf{X})$, $n\mathcal{Q}_n^\star(\mathbf{X})$ diverges to infinity under $\mathcal{H}_A$ as $\mathcal{Q}_n(\mathbf{X})$, $\mathcal{Q}_n^\star(\mathbf{X})$ converges to $\mathcal{Q}(\mathbf{X})$, $\mathcal{Q}^\star(\mathbf{X})$. Furthermore, $n\mathcal{Q}_n(\mathbf{X})$, $n\mathcal{Q}_n^\star(\mathbf{X})$ converges to a proper random variable under $\mathcal{H}_0$, which implies $\mathcal{Q}_n(\mathbf{X})$, $\mathcal{Q}_n^\star(\mathbf{X})$ converges to 0 under $\mathcal{H}_0$.

Therefore, a mutual independence test can be proposed based on the weak convergence of $n\mathcal{Q}_n(\mathbf{X})$, $n\mathcal{Q}_n^\star(\mathbf{X})$ in Theorem 3. Since the asymptotic null distributions of $n\mathcal{Q}_n(\mathbf{X})$, $n\mathcal{Q}_n^\star(\mathbf{X})$ depend on F_X, they will not be used in practice, and a permutation procedure will be used to approximate them instead.

4. Asymmetric and symmetric measures of mutual dependence

As an alternative, we now propose the asymmetric and symmetric measures of mutual dependence to capture mutual dependence via aggregating pairwise dependences. The subset of components on the right of X_c is denoted by $X_{c^+} = (X_{c+1}, \ldots, X_d)$, with $t_{c^+} = (t_{c+1}, \ldots, t_d)$ for $c \in \{0, \ldots, d-1\}$. The subset of components except X_c is denoted by $X_{-c} = (X_1, \ldots, X_{c-1}, X_{c^+})$, with $t_{-c} = (t_1, \ldots, t_{c-1}, t_{c^+})$ for $c \in \{1, \ldots, d-1\}$.

We denote pairwise independence by $\perp\!\!\!\perp$. The collection of pairwise independences implied by mutual independence includes "one versus others on the right"

$$\{X_1 \perp\!\!\!\perp X_{1^+}, X_2 \perp\!\!\!\perp X_{2^+}, \ldots, X_{d-1} \perp\!\!\!\perp X_d\}, \tag{1}$$

"one versus all the others"

$$\{X_1 \perp\!\!\!\perp X_{-1}, X_2 \perp\!\!\!\perp X_{-2}, \ldots, X_d \perp\!\!\!\perp X_{-d}\}, \tag{2}$$

and many others, e.g., $(X_1, X_2) \perp\!\!\!\perp X_{2^+}$. In fact, the number of pairwise independences resulting from mutual independence is at least $2^{d-1} - 1$, which grows exponentially with the number of components d. Therefore, we cannot test mutual independence simply by checking all pairwise independences even with moderate d.

Fortunately, we have two options to test only a small subset of all pairwise independences to fulfill the task. The first one is that $\mathcal{H}_0$ holds if and only if (1) holds, which can be verified via the sequential decomposition of distribution functions. This option is asymmetric and not unique, having $d!$ feasible subsets with respect to different orders of $X_1, \ldots, X_d$. The second one is that $\mathcal{H}_0$ holds if and only if (2) holds, which can be verified via the stepwise decomposition of distribution functions and the fact that $X_j \perp\!\!\!\perp X_{-j}$ implies $X_j \perp\!\!\!\perp X_{j^+}$. This option is symmetric and unique, having only one feasible subset.

To shed light on why these two options are necessary and sufficient conditions to mutual independence, we present the following inequality that the mutual dependence can be bounded by a sum of several pairwise dependences as

$$\left|\phi_X(t) - \prod_{j=1}^d \phi_{X_j}(t_j)\right| \leq \sum_{c=1}^{d-1} |\phi_{(X_c, X_{c^+})}\{(t_c, t_{c^+})\} - \phi_{X_c}(t_c)\phi_{X_{c^+}}(t_{c^+})|^2.$$

In consideration of these two options, we test a set of pairwise independences in place of mutual independence, where we use $\mathcal{V}^2(X)$ to test pairwise independence.

Definition 4. The asymmetric and symmetric measures of mutual dependence $\mathcal{R}(X)$, $\mathcal{S}(X)$ are defined by

$$\mathcal{R}(X) = \sum_{c=1}^{d-1} \mathcal{V}^2\{(X_c, X_{c^+})\} \quad \text{and} \quad \mathcal{S}(X) = \sum_{c=1}^{d} \mathcal{V}^2\{(X_c, X_{-c})\}.$$

We can show an equivalence to mutual independence based on $\mathcal{R}(X)$, $\mathcal{S}(X)$ according to Theorem 3 in [37].

Theorem 4. *If* $\mathrm{E}|X|_p < \infty$, *then* $\mathcal{R}(X), \mathcal{S}(X) \in [0, \infty)$, *and* $\mathcal{R}(X), \mathcal{S}(X) = 0$ *if and only if* $X_1, \ldots, X_d$ *are mutually independent.*

It is straightforward to estimate $\mathcal{R}(X)$, $\mathcal{S}(X)$ by replacing the characteristic functions with the empirical characteristic functions from the sample.

Definition 5. The empirical asymmetric and symmetric measures of mutual dependence $\mathcal{R}_n(\mathbf{X})$, $\mathcal{S}_n(\mathbf{X})$ are defined by

$$\mathcal{R}_n(\mathbf{X}) = \sum_{c=1}^{d-1} \mathcal{V}_n^2\{(\mathbf{X}_c, \mathbf{X}_{c+})\} \quad \text{and} \quad \mathcal{S}_n(\mathbf{X}) = \sum_{c=1}^{d} \mathcal{V}_n^2\{(\mathbf{X}_c, \mathbf{X}_{-c})\}.$$

The implementations of $\mathcal{R}_n(\mathbf{X})$, $\mathcal{S}_n(\mathbf{X})$ have the time complexity $O(n^2)$. Using a similar derivation to Theorems 2 and 5 in [37], some asymptotic properties of $\mathcal{R}_n(\mathbf{X})$, $\mathcal{S}_n(\mathbf{X})$ are obtained as follows.

Theorem 5. *If* $\mathrm{E}|X|_p < \infty$, *then* $\mathcal{R}_n(\mathbf{X}) \xrightarrow{\text{a.s.}} \mathcal{R}(X)$ *and* $\mathcal{S}_n(\mathbf{X}) \xrightarrow{\text{a.s.}} \mathcal{S}(X)$ *as* $n \to \infty$.

Theorem 6. *If* $\mathrm{E}|X|_p < \infty$, *then under* $\mathcal{H}_0$, *we have, as* $n \to \infty$,

$$n\mathcal{R}_n(\mathbf{X}) \rightsquigarrow \sum_{j=1}^{d-1} \|\zeta_j^R\{(t_j, t_{j+})\}\|_{w_0}^2 \quad \text{and} \quad n\mathcal{S}_n(\mathbf{X}) \rightsquigarrow \sum_{j=1}^{d} \|\zeta_j^S\{(t_j, t_{-j})\}\|_{w_0}^2,$$

where $\zeta_j^R\{(t_j, t_{j+})\}$, $\zeta_j^S\{(t_j, t_{-j})\}$ *are complex-valued Gaussian processes corresponding to the limiting distributions of* $n\mathcal{V}_n^2\{(\mathbf{X}_j, \mathbf{X}_{j+})\}$, $n\mathcal{V}_n^2\{(\mathbf{X}_j, \mathbf{X}_{-j})\}$. *Under* $\mathcal{H}_A$, *we have* $n\mathcal{R}_n(\mathbf{X}) \xrightarrow{\text{a.s.}} \infty$ *and* $n\mathcal{S}_n(\mathbf{X}) \xrightarrow{\text{a.s.}} \infty$ *as* $n \to \infty$.

It is surprising to find that the variables $\mathcal{V}_n^2\{(\mathbf{X}_c, \mathbf{X}_{c+})\}$ defined for all $c \in \{1, \ldots, d-1\}$ are mutually independent asymptotically, and that the variables $\mathcal{V}_n^2\{(\mathbf{X}_c, \mathbf{X}_{-c})\}$ defined for $c \in \{1, \ldots, d\}$ are mutually independent asymptotically as well, which is a crucial discovery behind Theorem 6. Theorems 5 and 6 are also closely connected in a similar way to Theorems 2 and 3. Similar to Theorem 3, the asymptotic results in Theorem 6 will not be used, but will be approximated by a permutation procedure in the tests.

$\mathcal{R}_n(\mathbf{X})$, $\mathcal{S}_n(\mathbf{X})$ only contain a subset of pairwise dependences from the Möbius decomposition used in [4,9,12], but we still obtain an equivalent condition to mutual independence. On the one hand, $\mathcal{R}_n(\mathbf{X})$, $\mathcal{S}_n(\mathbf{X})$ have much lower complexity when d gets large. On the other hand, we probably cannot narrow down to the smallest pair with significant dependence, while we can still find clues about the dependence structure. For example, the dependence between X_1 and X_2 is not directly included in $\mathcal{R}_n(\mathbf{X})$, $\mathcal{S}_n(\mathbf{X})$, but it is expected to be captured by the dependence between X_1 and X_{1+} included in $\mathcal{R}_n(\mathbf{X})$, $\mathcal{S}_n(\mathbf{X})$. Thus, we can observe the dependence between X_1 and X_{1+}, but not between X_1 and X_2 without further investigation.

Alternatively, we can plug in $\mathcal{Q}(X)$ instead of $\mathcal{V}^2(X)$ in Definition 4 and $\mathcal{Q}_n(\mathbf{X})$ instead of $\mathcal{V}_n^2(\mathbf{X})$ in Definition 5, and define the asymmetric and symmetric measures $\mathcal{J}(X)$, $\mathcal{I}(X)$ accordingly, which equal $\mathcal{Q}(X)$, $\mathcal{Q}_n(\mathbf{X})$ when $d = 2$. The naive implementations of $\mathcal{J}_n(\mathbf{X})$, $\mathcal{I}_n(\mathbf{X})$ have the time complexity $O(n^4)$. Similarly, we can replace $\mathcal{Q}_n(\mathbf{X})$ with $\mathcal{Q}_n^\star(\mathbf{X})$ to simplify them, and define the simplified empirical asymmetric and symmetric measures $\mathcal{J}_n^\star(\mathbf{X})$, $\mathcal{I}_n^\star(\mathbf{X})$, reducing their complexities to $O(n^2)$ without any other condition except the first moment condition $\mathrm{E}|X|_p < \infty$. Through the same derivations, we can show that $\mathcal{J}_n(\mathbf{X})$, $\mathcal{J}_n^\star(\mathbf{X})$, $\mathcal{I}_n(\mathbf{X})$, $\mathcal{I}_n^\star(\mathbf{X})$ have similar convergences as $\mathcal{R}_n(\mathbf{X})$, $\mathcal{S}_n(\mathbf{X})$ in Theorems 5 and 6.

5. Simulation studies

In this section, we evaluate the finite sample performance of proposed measures $\mathcal{Q}_n$, $\mathcal{R}_n$, $\mathcal{S}_n$, $\mathcal{J}_n$, $\mathcal{I}_n$, $\mathcal{Q}_n^\star$, $\mathcal{J}_n^\star$, $\mathcal{I}_n^\star$ by performing simulations similar to Székely et al. [37], and compare them with benchmark measures $\mathcal{V}_n^2$ [37], BN^h, BN^d [4], dHSIC [30], and HL^τ, HL^ρ [14] respectively in various scenarios. Note that BN^h is based on HSIC, BN^d is based on distance covariance, HL^τ is based on Kendall's τ, and HL^ρ is based on Spearman's ρ. We also include permutation tests based on finite-sample extensions of HL^τ, HL^ρ, denoted by HL_n^τ, HL_n^ρ. Moreover, dHSIC is implemented in the `R` package `dHSIC` [29] using the Gaussian kernel with a median heuristic to choose the bandwidth.

We test the null hypothesis $\mathcal{H}_0$ with significance level $\alpha = 0.1$ and examine the empirical size and power of each measure. In each scenario, we run 1000 repetitions with the adaptive permutation size $B = \lfloor 200 + 5000/n \rfloor$ where n is the sample size, for all empirical measures that require a permutation procedure to approximate their asymptotic distributions, i.e., $\mathcal{Q}_n$, $\mathcal{R}_n$, $\mathcal{S}_n$, $\mathcal{J}_n$, $\mathcal{I}_n$, $\mathcal{Q}_n^\star$, $\mathcal{J}_n^\star$, $\mathcal{I}_n^\star$, $\mathcal{V}_n^2$, BN^h, BN^d, dHSIC, HL_n^τ, HL_n^ρ.

In the following two examples, we fix $d = 2$ and change n from 25 to 500, and compare $\mathcal{Q}_n$, $\mathcal{R}_n$, $\mathcal{S}_n$, $\mathcal{J}_n$, $\mathcal{I}_n$, $\mathcal{Q}_n^\star$, $\mathcal{J}_n^\star$, $\mathcal{I}_n^\star$ to $\mathcal{V}_n^2$.

Example 1 (*Pairwise Multivariate Normal*). $X_1, X_2 \in \mathbb{R}^5$, $(X_1, X_2)^\top \sim \mathcal{N}_{10}(0, \Sigma)$, where $\Sigma_{ii} = 1$. Under $\mathcal{H}_0$, $\Sigma_{ij} = 0$, $i \neq j$. Under $\mathcal{H}_A$, $\Sigma_{ij} = 0.1$, $i \neq j$. The results are in Tables 1 and 2.

Example 2 (*Pairwise Multivariate Non-normal*). $X_1, X_2 \in \mathbb{R}^5$, $(Y_1, Y_2)^\top \sim \mathcal{N}_{10}(0, \Sigma)$, where $\Sigma_{ii} = 1$. $X_1 = \ln(Y_1^2)$, $X_2 = \ln(Y_2^2)$. Under $\mathcal{H}_0$, $\Sigma_{ij} = 0$, $i \neq j$. Under $\mathcal{H}_A$, $\Sigma_{ij} = 0.4$, $i \neq j$. The results are in Tables 3 and 4.

Table 1
Empirical size ($\alpha = 0.1$) in Example 1 with 1000 repetitions and $d = 2$.

n	$\mathcal{V}_n^2, \mathcal{R}_n, \mathcal{S}_n$	$\mathcal{Q}_n, \mathcal{J}_n, \mathcal{I}_n$	$\mathcal{Q}_n^\star, \mathcal{J}_n^\star$	$\mathcal{I}_n^\star$
25	0.106	0.102	0.108	0.111
30	0.098	0.115	0.086	0.114
35	0.095	0.101	0.084	0.101
50	0.101	0.101	0.111	0.106
70	0.114	0.109	0.090	0.102
100	0.104	0.105	0.118	0.117

Table 2
Empirical power ($\alpha = 0.1$) in Example 1 with 1000 repetitions and $d = 2$.

n	$\mathcal{V}_n^2, \mathcal{R}_n, \mathcal{S}_n$	$\mathcal{Q}_n, \mathcal{J}_n, \mathcal{I}_n$	$\mathcal{Q}_n^\star, \mathcal{J}_n^\star$	$\mathcal{I}_n^\star$
25	0.273	0.246	0.160	0.182
50	0.496	0.448	0.259	0.300
100	0.807	0.751	0.442	0.514
150	0.943	0.922	0.604	0.720
200	0.979	–	0.749	0.836
300	1.000	–	0.889	0.954
500	1.000	–	0.978	0.995

Table 3
Empirical size ($\alpha = 0.1$) in Example 2 with 1000 repetitions and $d = 2$.

n	$\mathcal{V}_n^2, \mathcal{R}_n, \mathcal{S}_n$	$\mathcal{Q}_n, \mathcal{J}_n, \mathcal{I}_n$	$\mathcal{Q}_n^\star, \mathcal{J}_n^\star$	$\mathcal{I}_n^\star$
25	0.088	0.093	0.091	0.092
30	0.098	0.104	0.108	0.110
35	0.104	0.102	0.104	0.099
50	0.097	0.098	0.093	0.097
70	0.094	0.097	0.089	0.097
100	0.092	0.092	0.114	0.099

Table 4
Empirical power ($\alpha = 0.1$) in Example 2 with 1000 repetitions and $d = 2$.

n	$\mathcal{V}_n^2, \mathcal{R}_n, \mathcal{S}_n$	$\mathcal{Q}_n, \mathcal{J}_n, \mathcal{I}_n$	$\mathcal{Q}_n^\star, \mathcal{J}_n^\star$	$\mathcal{I}_n^\star$
25	0.181	0.185	0.141	0.152
50	0.352	0.339	0.200	0.239
100	0.610	0.607	0.372	0.413
150	0.793	0.792	0.474	0.588
200	0.885	–	0.604	0.711
300	0.989	–	0.803	0.892
500	0.999	–	0.953	0.988

Table 5
Empirical size ($\alpha = 0.1$) in Example 3 with 1000 repetitions and $d = 3$.

n	BN^h	BN^d	dHSIC	$\mathcal{Q}_n$	$\mathcal{Q}_n^\star$	$\mathcal{R}_n$	$\mathcal{S}_n$	$\mathcal{J}_n$	$\mathcal{J}_n^\star$	$\mathcal{I}_n$	$\mathcal{I}_n^\star$
25	0.103	0.106	0.097	0.095	0.103	0.093	0.096	0.101	0.100	0.091	0.101
30	0.114	0.106	0.101	–	0.110	0.110	0.114	0.108	0.118	0.111	0.125
35	0.101	0.095	0.090	–	0.108	0.106	0.102	0.109	0.106	0.104	0.092
50	0.098	0.100	0.106	–	0.083	0.113	0.108	0.110	0.090	0.105	0.085
70	0.114	0.102	0.107	–	0.107	0.104	0.104	0.098	0.101	0.108	0.109
100	0.127	0.125	0.099	–	0.085	0.106	0.108	0.104	0.103	0.109	0.096

For both Examples 1 and 2, the empirical size of all measures is close to $\alpha = 0.1$. The empirical power of $\mathcal{Q}_n, \mathcal{R}_n, \mathcal{S}_n, \mathcal{J}_n, \mathcal{I}_n$ is almost the same as that of $\mathcal{V}_n^2$, while the empirical power of $\mathcal{Q}_n^\star, \mathcal{J}_n^\star, \mathcal{I}_n^\star$ is lower than that of $\mathcal{V}_n^2$, which makes sense because we trade-off testing power and time complexity for simplified measures.

In the following two examples, we fix $d = 3$ and change n from 25 to 500, and compare $\mathcal{Q}_n, \mathcal{R}_n, \mathcal{S}_n, \mathcal{J}_n, \mathcal{I}_n, \mathcal{Q}_n^\star, \mathcal{J}_n^\star, \mathcal{I}_n^\star$ to BN^h, BN^d, dHSIC.

Example 3 (*Mutual Multivariate Normal*). $X_1, X_2, X_3 \in \mathbb{R}^5$, $(X_1, X_2, X_3)^\top \sim \mathcal{N}_{15}(0, \Sigma)$ where $\Sigma_{ii} = 1$. Under $\mathcal{H}_0$, $\Sigma_{ij} = 0$, $i \neq j$. Under $\mathcal{H}_A$, $\Sigma_{ij} = 0.1$, $i \neq j$. The results are in Tables 5 and 6.

Example 4 (*Mutual Multivariate Non-normal*). $X_1, X_2, X_3 \in \mathbb{R}^5$. $(Y_1, Y_2, Y_3)^\top \sim \mathcal{N}_{15}(0, \Sigma)$ where $\Sigma_{ii} = 1$. $X_k = \ln(Y_k^2)$, $k = 1, 2, 3$. Under $\mathcal{H}_0$, $\Sigma_{ij} = 0$, $i \neq j$. Under $\mathcal{H}_A$, $\Sigma_{ij} = 0.4$, $i \neq j$. The results are in Tables 7 and 8.

Table 6
Empirical power ($\alpha = 0.1$) in Example 3 with 1000 repetitions and $d = 3$.

n	BN^h	BN^d	dHSIC	$\mathcal{Q}_n$	$\mathcal{Q}_n^\star$	$\mathcal{R}_n$	$\mathcal{S}_n$	$\mathcal{J}_n$	$\mathcal{J}_n^\star$	$\mathcal{I}_n$	$\mathcal{I}_n^\star$
25	0.992	0.998	0.982	0.383	0.220	0.402	0.418	0.360	0.199	0.384	0.228
50	1.000	1.000	1.000	–	0.378	0.707	0.719	0.651	0.338	0.671	0.389
100	1.000	1.000	1.000	–	0.707	0.956	0.961	0.940	0.643	0.946	0.767
150	1.000	1.000	1.000	–	0.873	0.996	0.996	0.993	0.830	0.994	0.921
200	1.000	1.000	1.000	–	0.946	1.000	1.000	–	0.930	–	0.972
300	1.000	1.000	1.000	–	0.997	1.000	1.000	–	0.996	–	0.999
500	1.000	1.000	1.000	–	1.000	1.000	1.000	–	1.000	–	1.000

Table 7
Empirical size ($\alpha = 0.1$) in Example 4 with 1000 repetitions and $d = 3$.

n	BN^h	BN^d	dHSIC	$\mathcal{Q}_n$	$\mathcal{Q}_n^\star$	$\mathcal{R}_n$	$\mathcal{S}_n$	$\mathcal{J}_n$	$\mathcal{J}_n^\star$	$\mathcal{I}_n$	$\mathcal{I}_n^\star$
25	0.099	0.105	0.102	0.089	0.098	0.096	0.097	0.096	0.099	0.092	0.108
30	0.092	0.089	0.087	–	0.098	0.102	0.100	0.094	0.099	0.095	0.108
35	0.105	0.104	0.087	–	0.116	0.116	0.122	0.123	0.117	0.123	0.113
50	0.098	0.096	0.107	–	0.091	0.112	0.109	0.102	0.097	0.113	0.088
70	0.127	0.127	0.101	–	0.084	0.103	0.105	0.096	0.112	0.102	0.116
100	0.105	0.103	0.110	–	0.112	0.105	0.105	0.109	0.099	0.104	0.107

Table 8
Empirical power ($\alpha = 0.1$) in Example 4 with 1000 repetitions and $d = 3$.

n	BN^h	BN^d	dHSIC	$\mathcal{Q}_n$	$\mathcal{Q}_n^\star$	$\mathcal{R}_n$	$\mathcal{S}_n$	$\mathcal{J}_n$	$\mathcal{J}_n^\star$	$\mathcal{I}_n$	$\mathcal{I}_n^\star$
25	0.285	0.268	0.267	0.289	0.164	0.294	0.287	0.291	0.154	0.287	0.169
50	0.479	0.479	0.441	–	0.280	0.504	0.510	0.490	0.278	0.501	0.320
100	0.768	0.760	0.745	–	0.521	0.824	0.826	0.807	0.498	0.816	0.579
150	0.919	0.929	0.906	–	0.689	0.942	0.942	0.937	0.679	0.941	0.770
200	0.982	0.987	0.963	–	0.838	0.987	0.986	–	0.826	–	0.905
300	0.999	0.999	0.997	–	0.957	0.999	0.999	–	0.956	–	0.982
500	1.000	1.000	1.000	–	1.000	1.000	1.000	—	1.000	–	1.000

Table 9
Empirical size ($\alpha = 0.1$) in Example 5 with 1000 repetitions and $n = 100$.

d	HL^τ	HL^ρ	HL_n^τ	HL_n^ρ	$\mathcal{Q}_n^\star$	$\mathcal{R}_n$	$\mathcal{S}_n$	$\mathcal{J}_n^\star$	$\mathcal{I}_n^\star$
5	0.076	0.066	0.113	0.105	0.097	0.091	0.091	0.094	0.104
10	0.077	0.070	0.104	0.097	0.107	0.092	0.094	0.119	0.107
15	0.094	0.087	0.116	0.113	0.109	0.093	0.093	0.108	0.100
20	0.077	0.066	0.089	0.089	0.096	0.099	0.118	0.115	0.101
25	0.074	0.058	0.086	0.091	0.097	0.090	0.082	0.095	0.097
30	0.091	0.082	0.110	0.114	0.109	0.092	0.104	0.105	0.109
50	0.080	0.061	0.088	0.087	0.087	0.091	0.088	0.095	0.087

For both Examples 3 and 4, the empirical size of all measures is close to $\alpha = 0.1$. The empirical power of $\mathcal{Q}_n$, $\mathcal{R}_n$, $\mathcal{S}_n$, $\mathcal{J}_n$, $\mathcal{I}_n$ is almost the same, the empirical power of $\mathcal{Q}_n^\star$, $\mathcal{J}_n^\star$, $\mathcal{I}_n^\star$ is almost the same, while the empirical power of $\mathcal{Q}_n^\star$, $\mathcal{J}_n^\star$, $\mathcal{I}_n^\star$ is lower than that of $\mathcal{Q}_n$, $\mathcal{R}_n$, $\mathcal{S}_n$, $\mathcal{J}_n$, $\mathcal{I}_n$, which makes sense since we trade-off testing power and time complexity for simplified measures. BN^h, BN^d, dHSIC outperform all other measures in normal Example 3, while $\mathcal{Q}_n$, $\mathcal{R}_n$, $\mathcal{S}_n$, $\mathcal{J}_n$, $\mathcal{I}_n$ achieve slightly better performance than BN^h, BN^d, dHSIC in non-normal Example 4.

To compare the computation time of these measures, we evaluate one case in Example 4 with $n = 25$ under $\mathcal{H}_0$. When running on Dell PowerEdge 2650 with 16GB RAM using a single core, $\mathcal{R}_n$, $\mathcal{S}_n$, $\mathcal{Q}_n^\star$, $\mathcal{J}_n^\star$, $\mathcal{I}_n^\star$ take 164.09, 117.57, 51.66, 71.39, 94.96 s respectively, while BN^h, BN^d, dHSIC take 207.16, 204.42, 70.40 s respectively.

In the last example, we change d from 5 to 50 and fix $n = 100$, and compare $\mathcal{R}_n$, $\mathcal{S}_n$, $\mathcal{Q}_n^\star$, $\mathcal{J}_n^\star$, $\mathcal{I}_n^\star$ to HL^τ, HL^ρ, HL_n^τ, HL_n^ρ.

Example 5 (*Mutual Univariate Normal High-Dimensional*)**.** $X_1, \ldots, X_d \in \mathbb{R}^1$. $(X_1, \ldots, X_d)^\top \sim \mathcal{N}_d(0, \Sigma)$, where $\Sigma_{ii} = 1$. Under $\mathcal{H}_0$, $\Sigma_{ij} = 0, i \neq j$. Under $\mathcal{H}_A$, $\Sigma_{ij} = 0.1, i \neq j$. The results are in Tables 9 and 10.

The empirical size of HL^τ, HL^ρ is much lower than $\alpha = 0.1$ and too conservative, while that of other measures is fairly close to $\alpha = 0.1$. The reason is probably that the convergence to asymptotic distributions of HL^τ, HL^ρ requires larger sample size n and number of components d. The measures $\mathcal{R}_n$, $\mathcal{S}_n$ have the highest empirical power, and outperform the simplified measures $\mathcal{Q}_n^\star$, $\mathcal{J}_n^\star$, $\mathcal{I}_n^\star$. The empirical power of simplified measures is similar to or even lower than that of benchmark measures when $d = 5$. However, the empirical power of simplified measures converges much faster than that of benchmark measures as d grows.

Table 10
Empirical power ($\alpha = 0.1$) in Example 5 with 1000 repetitions and $n = 100$.

d	HL^{τ}	HL^{ρ}	HL_n^{τ}	HL_n^{ρ}	$\mathcal{Q}_n^\star$	$\mathcal{R}_n$	$\mathcal{S}_n$	$\mathcal{J}_n^\star$	$\mathcal{I}_n^\star$
5	0.317	0.305	0.410	0.405	0.298	0.545	0.557	0.245	0.318
10	0.426	0.416	0.500	0.510	0.557	0.896	0.915	0.409	0.497
15	0.513	0.481	0.593	0.602	0.822	0.975	0.982	0.538	0.643
20	0.558	0.534	0.625	0.634	0.924	0.996	0.999	0.586	0.647
25	0.593	0.539	0.645	0.634	0.977	0.999	0.999	0.663	0.689
30	0.605	0.556	0.675	0.664	0.980	1.000	1.000	0.711	0.700
50	0.702	0.641	0.742	0.731	0.998	1.000	1.000	0.775	0.717

Moreover, $\mathcal{Q}_n^\star$ shows significant advantage over $\mathcal{J}_n^\star, \mathcal{I}_n^\star$. The reason is probably that $\mathcal{Q}_n^\star$ is based on truly mutual dependence while $\mathcal{J}_n^\star$, $\mathcal{I}_n^\star$ are based on pairwise dependences, and large d compared to n introduces much more noise to $\mathcal{J}_n^\star, \mathcal{I}_n^\star$ because of their summation structures, which makes them more difficult to detect mutual dependence.

The asymptotic analysis of our measures only allows small d compared to n, while our measures work well with large d compared to n in Example 5. However, this success relies on the underlying dependence structure, which is dense since each component is dependent on any other component. In contrast, if the dependence structure is sparse as each component is dependent on only a few of other components, then all measures are likely to fail.

6. Illustrative examples

We start with two examples comparing different methods to show the value of our mutual independence tests. In practice, people usually check all pairwise dependences to test mutual independence, due to the lack of reliable and universal mutual independence tests. It is very likely to miss the complicated mutual dependence structure, and make unsound decisions in corresponding applications assuming that mutual independence holds.

6.1. Synthetic data

We define a triplet of random vectors (X, Y, Z) on $\mathbb{R}^q \times \mathbb{R}^q \times \mathbb{R}^q$, where $X, Y \sim \mathcal{N}(0, I_q)$, $W \sim \mathcal{E}(1/\sqrt{2})$, the first element of Z is $Z_1 = \mathrm{sign}(X_1Y_1)W$ and the remaining $q-1$ elements are $Z_{2:q} \sim \mathcal{N}(0, I_{q-1})$, and $X, Y, W, Z_{2:q}$ are mutually independent. Clearly, (X, Y, Z) is a pairwise independent but mutually dependent triplet.

An iid sample of (X, Y, Z) is randomly generated with sample size $n = 500$ and dimension $q = 5$. On the one hand, we test the null hypothesis $\mathcal{H}_0 : X, Y, Z$ are mutually independent using proposed measures $\mathcal{R}_n$, $\mathcal{S}_n$, $\mathcal{Q}_n^\star$, $\mathcal{J}_n^\star$, $\mathcal{I}_n^\star$. On the other hand, we test the null hypotheses $\mathcal{H}_0^{(1)} : X \perp\!\!\!\perp Y$, $\mathcal{H}_0^{(2)} : Y \perp\!\!\!\perp Z$, and $\mathcal{H}_0^{(3)} : X \perp\!\!\!\perp Z$ using distance covariance $\mathcal{V}_n^2$. An adaptive permutation size $B = 210$ is used for all tests.

As expected, mutual dependence is successfully captured, as the p-values of mutual independence tests are 0.0143 ($\mathcal{Q}_n^\star$), 0.0286 ($\mathcal{J}_n^\star$), 0 ($\mathcal{I}_n^\star$), 0.0381 ($\mathcal{R}_n$) and 0 ($\mathcal{S}_n$). Meanwhile, the p-values of pairwise independence tests are 0.2905 (X, Y), 0.2619 (Y, Z), and 0.3048 (X, Z). According to the Bonferroni correction for multiple tests among all the pairs, the significance level should be adjusted as $\alpha/3$ for pairwise tests. As a result, no signal of pairwise dependence is detected, and we cannot reject mutual independence.

6.2. Financial data

Fama and French [6,7] proposed the Fama–French three-factor and five-factor models to explain the stock returns, and demonstrated that these factors comprising the stock returns are correlated according to long-term market research in finance. Thus, we apply our tests to a subset of these factors and confirm this argument as an application.

We collect the annual Fama–French five factors in the past 52 years between 1964 and 2015; the data are available at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. In particular, we are interested in whether mutual dependence among three factors, $X =$ Mkt-RF (excess return on the market), $Y =$ SMB (small minus big), and $Z =$ RF (risk-free return) exists, where annual returns are considered as nearly independent observations. Both histograms and pair plots of X, Y, Z are depicted in Fig. 1.

For one, we apply a single mutual independence test $\mathcal{H}_0 : X, Y, Z$ are mutually independent. For another, we apply three pairwise independence tests $\mathcal{H}_0^{(1)} : X \perp\!\!\!\perp Y$, $\mathcal{H}_0^{(2)} : Y \perp\!\!\!\perp Z$, and $\mathcal{H}_0^{(3)} : X \perp\!\!\!\perp Z$. An adaptive permutation size $B = 296$ is used for all tests.

The p-values of mutual independence tests are 0.0236 ($\mathcal{Q}_n^\star$), 0.0642 ($\mathcal{J}_n^\star$), 0.0541 ($\mathcal{I}_n^\star$), 0.1588 ($\mathcal{R}_n$) and 0.1486 ($\mathcal{S}_n$), indicating that mutual dependence is successfully captured. In the meanwhile, the p-values of pairwise independence tests using distance covariance $\mathcal{V}_n^2$ are 0.1419 (X, Y), 0.5743 (Y, Z) and 0.5405 (X, Z). Similarly, the significance level should be adjusted as $\alpha/3$ according to the Bonferroni correction, and thus we cannot reject mutual independence, since no signal of pairwise dependence is detected.

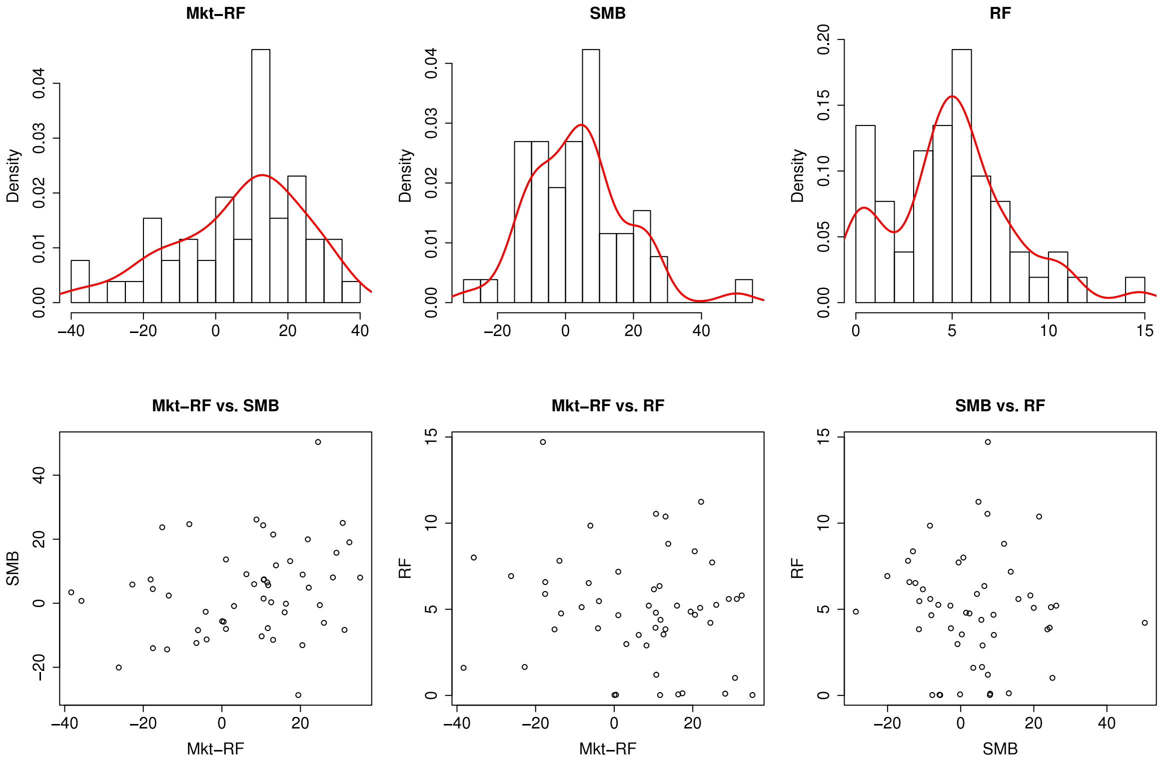

Fig. 1. Three annual Fama–French factors between 1964 and 2015: Mkt-RF (excess return on the market), SMB (small minus big) and RF (risk-free return). The correlations are corr(Mkt-RF, SMB) = 0.238, corr(Mkt-RF, RF) = −0.161, and corr(SMB, RF) = −0.0645. Red lines in the histograms are estimated kernel densities. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

7. Conclusion

We propose three measures of mutual dependence for random vectors based on the equivalence to mutual independence through characteristic functions, following the idea of distance covariance in Székely et al. [37].

When we select the weight function for the complete measure, we trade off between moment condition and time complexity. Then we simplify it by replacing complete V-statistics by incomplete V-statistics, as a trade-off between testing power and time complexity. These two trade-offs make the simplified complete measure both effective and efficient. The asymptotic distributions of our measures depend on the underlying distribution F_X. Thus, the corresponding tests are not distribution-free, and we use a permutation procedure to approximate the asymptotic distributions in practice.

We illustrate the value of our measures through both synthetic and financial data examples, where mutual independence tests based on our measures successfully capture the mutual dependence, while checking all pairwise independences as an alternative independences fails and mistakenly leads to the conclusion that mutual independence holds. Our measures achieve competitive or even better results than the benchmark measures in simulations with various examples. Although we do not allow large d compared to n in asymptotic analysis, our measures work well in a large d example since the dependence structure is dense. Lastly, it would be interesting to extend current results on continuous variables to categorical variables, as applied statisticians may rely on such measures to conduct sensitivity analyses [24] correspondingly.

Acknowledgments

We are grateful to Stanislav Volgushev for helpful comments on a preliminary draft of this paper. Useful comments and suggestions from the Editor-in-Chief are also gratefully acknowledged. This research was support from an NSF Award, USA (DMS-1455172), a Xerox PARC Faculty Research Award, USA, and Cornell University's Atkinson Center for a Sustainable Future, USA (AVF-2017).

Appendix

Proof of Theorem 1. By definition, $\mathcal{Q}(X) = \int_{\mathbb{R}^p} |\phi_X(t) - \phi_{\tilde{X}}(t)|^2 w_1(t)\, dt$. Since $w_1(t)$ is a positive weight function, it is clear that $\mathcal{Q}(X) \geq 0$ and that $X_1, \ldots, X_d$ are mutually independent if and only if $\mathcal{Q}(X) = 0$. By the boundedness property of characteristic functions and Fubini's theorem, we have

$$\begin{aligned}
|\phi_X(t)-\phi_{\widetilde{X}}(t)|^2 &= \phi_X(t)\overline{\phi_X(t)}+\phi_{\widetilde{X}}(t)\overline{\phi_{\widetilde{X}}(t)}-\phi_X(t)\overline{\phi_{\widetilde{X}}(t)}-\phi_{\widetilde{X}}(t)\overline{\phi_X(t)}\\
&= (\mathrm{E}^{i\langle t,X\rangle})\mathrm{E}(e^{-i\langle t,X\rangle})+\mathrm{E}(e^{i\langle t,\widetilde{X}\rangle})\mathrm{E}(e^{-i\langle t,\widetilde{X}\rangle})-\mathrm{E}(e^{i\langle t,X\rangle})\mathrm{E}(e^{-i\langle t,\widetilde{X}\rangle})-\mathrm{E}(e^{i\langle t,\widetilde{X}\rangle})\mathrm{E}(e^{-i\langle t,X\rangle})\\
&= \mathrm{E}(e^{i\langle t,X-X'\rangle})+\mathrm{E}(e^{i\langle t,\widetilde{X}-\widetilde{X}'\rangle})-\mathrm{E}(e^{i\langle t,X-\widetilde{X}'\rangle})-\mathrm{E}(e^{i\langle t,\widetilde{X}-X'\rangle})\\
&= \mathrm{E}(\cos\langle t,X-X'\rangle)+\mathrm{E}(\cos\langle t,\widetilde{X}-\widetilde{X}'\rangle)+\mathrm{E}(\cos\langle t,X-\widetilde{X}'\rangle)+\mathrm{E}(\cos\langle t,\widetilde{X}-X'\rangle)\\
&= \mathrm{E}(1-\cos\langle t,X-\widetilde{X}'\rangle)+\mathrm{E}(1-\cos\langle t,\widetilde{X}-X'\rangle)-\mathrm{E}(1-\cos\langle t,X-X'\rangle)-\mathrm{E}(1-\cos\langle t,\widetilde{X}-\widetilde{X}'\rangle).
\end{aligned}$$

Since $\mathrm{E}|X|_p<\infty$ implies $\mathrm{E}|\widetilde{X}|_p<\infty$, we have $\mathrm{E}(|X|_p+|\widetilde{X}|_p)<\infty$. Then the triangle inequality implies

$$\max(\mathrm{E}|X-X'|_p,\mathrm{E}|\widetilde{X}-\widetilde{X}'|_p,\mathrm{E}|X-\widetilde{X}'|_p,\mathrm{E}|\widetilde{X}-X'|_p)<\infty.$$

Therefore, by Fubini's theorem and Lemma 1, it follows that

$$\begin{aligned}
\mathcal{Q}(X)=\int|\phi_X(t)-\phi_{\widetilde{X}}(t)|^2\,w_1(t)\,dt &= \int\mathrm{E}(1-\cos\langle t,X-\widetilde{X}'\rangle)\,w_1(t)\,dt+\int\mathrm{E}(1-\cos\langle t,\widetilde{X}-X'\rangle)\,w_1(t)\,dt\\
&\quad-\int\mathrm{E}(1-\cos\langle t,X-X'\rangle)\,w_1(t)\,dt-\int\mathrm{E}(1-\cos\langle t,\widetilde{X}-\widetilde{X}'\rangle)\,w_1(t)\,dt\\
&= \mathrm{E}|X-\widetilde{X}'|_p+\mathrm{E}|\widetilde{X}-X'|_p-\mathrm{E}|X-X'|_p-\mathrm{E}|\widetilde{X}-\widetilde{X}'|_p<\infty.
\end{aligned}$$

This concludes the argument. □

Proof of Lemma 1. After a simple calculation, we have

$$\begin{aligned}
|\phi^n_X(t)-\phi^n_{\widetilde{X}}(t)|^2 &= \phi^n_X(t)\overline{\phi^n_X(t)}-\phi^n_X(t)\overline{\phi^n_{\widetilde{X}}(t)}-\phi^n_{\widetilde{X}}(t)\overline{\phi^n_X(t)}+\phi^n_{\widetilde{X}}(t)\overline{\phi^n_{\widetilde{X}}(t)}\\
&= \frac{1}{n^2}\sum_{k,\ell=1}^n\cos\langle t,X^k-X^\ell\rangle-\frac{2}{n^{d+1}}\sum_{k,\ell_1,\ldots,\ell_d=1}^n\cos\langle t,X^k-(X_1^{\ell_1},\ldots,X_d^{\ell_d})\rangle\\
&\quad+\frac{1}{n^{2d}}\sum_{k_1,\ldots,k_d,\ell_1,\ldots,\ell_d=1}^n\cos\langle t,(X_1^{k_1},\ldots,X_d^{k_d})-(X_1^{\ell_1},\ldots,X_d^{\ell_d})\rangle+V\\
&= -\frac{1}{n^2}\sum_{k,\ell=1}^n(1-\cos\langle t,X^k-X^\ell\rangle)+\frac{2}{n^{d+1}}\sum_{k,\ell_1,\ldots,\ell_d=1}^n\{1-\cos\langle t,X^k-(X_1^{\ell_1},\ldots,X_d^{\ell_d})\rangle\}\\
&\quad-\frac{1}{n^{2d}}\sum_{k_1,\ldots,k_d,\ell_1,\ldots,\ell_d=1}^n\{1-\cos\langle t,(X_1^{k_1},\ldots,X_d^{k_d})-(X_1^{\ell_1},\ldots,X_d^{\ell_d})\rangle\}+V,
\end{aligned}$$

where V is imaginary and thus 0 as $|\phi^n_X(t)-\phi^n_{\widetilde{X}}(t)|^2$ is real. By Lemma 1 in [33],

$$\begin{aligned}
\mathcal{Q}_n(\mathbf{X}) = \|\phi^n_X(t)-\phi^n_{\widetilde{X}}(t)\|^2_{w_1} &= -\frac{1}{n^2}\sum_{k,\ell=1}^n|X^k-X^\ell|_p+\frac{2}{n^{d+1}}\sum_{k,\ell_1,\ldots,\ell_d=1}^n|X^k-(X_1^{\ell_1},\ldots,X_d^{\ell_d})|_p\\
&\quad-\frac{1}{n^{2d}}\sum_{k_1,\ldots,k_d,\ell_1,\ldots,\ell_d=1}^n|(X_1^{k_1},\ldots,X_d^{k_d})-(X_1^{\ell_1},\ldots,X_d^{\ell_d})|_p.
\end{aligned}$$

This concludes the argument. □

Proof of Lemma 2. After a simple calculation, we have

$$\begin{aligned}
|\phi^n_X(t) - \phi^{n\star}_{\widetilde{X}}(t)|^2 &= \phi^n_X(t)\overline{\phi^n_X(t)} - \phi^n_X(t)\overline{\phi^{n\star}_{\widetilde{X}}(t)} - \phi^{n\star}_{\widetilde{X}}(t)\overline{\phi^n_X(t)} + \phi^{n\star}_{\widetilde{X}}(t)\overline{\phi^{n\star}_{\widetilde{X}}(t)} \\
&= \frac{1}{n^2}\sum_{k,\ell=1}^{n} \cos\langle t, X^k - X^\ell\rangle - \frac{2}{n^2}\sum_{k,\ell=1}^{n} \cos\langle t, X^k - (X^\ell_1, \dots, X^{\ell+d-1}_d)\rangle \\
&\quad + \frac{1}{n^2}\sum_{k,\ell=1}^{n} \cos\langle t, (X^k_1, \dots, X^{k+d-1}_d) - (X^\ell_1, \dots, X^{\ell+d-1}_d)\rangle + V^\star \\
&= -\frac{1}{n^2}\sum_{k,\ell=1}^{n} \{1 - \cos\langle t, X^k - X^\ell\rangle\} + \frac{2}{n^2}\sum_{k,\ell=1}^{n} \{1 - \cos\langle t, X^k - (X^\ell_1, \dots, X^{\ell+d-1}_d)\rangle\} \\
&\quad - \frac{1}{n^2}\sum_{k,\ell=1}^{n} \{1 - \cos\langle t, (X^k_1, \dots, X^{k+d-1}_d) - (X^\ell_1, \dots, X^{\ell+d-1}_d)\rangle\} + V^\star,
\end{aligned}$$

where $V^\star$ is imaginary and thus 0 as $|\phi^n_X(t) - \phi^{n\star}_{\widetilde{X}}(t)|^2$ is real. By Lemma 1 in [33],

$$\begin{aligned}
\mathcal{Q}^\star_n(\mathbf{X}) = \|\phi^n_X(t) - \phi^{n\star}_{\widetilde{X}}(t)\|^2_{w_1} &= -\frac{1}{n^2}\sum_{k,\ell=1}^{n} |X^k - X^\ell|_p + \frac{2}{n^2}\sum_{k,\ell=1}^{n} |X^k - (X^\ell_1, \dots, X^{\ell+d-1}_d)|_p \\
&\quad - \frac{1}{n^2}\sum_{k,\ell=1}^{n} |(X^k_1, \dots, X^{k+d-1}_d) - (X^\ell_1, \dots, X^{\ell+d-1}_d)|_p.
\end{aligned}$$

This concludes the argument. □

Proof of Theorem 2. Define

$$\mathcal{Q}_n = \|\phi^n_X(t) - \phi^n_{\widetilde{X}}(t)\|^2_{w_1} \equiv \|\xi_n(t)\|^2_{w_1} \quad \text{and} \quad \mathcal{Q}^\star_n = \|\phi^n_X(t) - \phi^{n\star}_{\widetilde{X}}(t)\|^2_{w_1} \equiv \|\xi^\star_n(t)\|^2_{w_1}.$$

For all $\delta \in (0, 1)$, define the region

$$D(\delta) = \{t = (t_1, \dots, t_d) : \delta \le |t|^2_p = |t_1|^2_{p_1} + \cdots + |t_d|^2_{p_d} \le 2/\delta\}, \tag{3}$$

and random variables

$$\mathcal{Q}_{n,\delta} = \int_{D(\delta)} |\xi_n(t)|^2 w_1(t)\, dt \quad \text{and} \quad \mathcal{Q}^\star_{n,\delta} = \int_{D(\delta)} |\xi^\star_n(t)|^2 w_1(t)\, dt.$$

For any fixed δ, the weight function $w_1(t)$ is bounded on $D(\delta)$. Hence $\mathcal{Q}_{n,\delta}$ is a combination of V-statistics of bounded random variables. Similar to Theorem 2 in [37], it follows by the Strong Law of Large Numbers (SLLN) for V-statistics [27] that almost surely

$$\lim_{n\to\infty} \mathcal{Q}_{n,\delta} = \lim_{n\to\infty} \mathcal{Q}^\star_{n,\delta} = \mathcal{Q}_{\cdot,\delta} = \int_{D(\delta)} |\phi_X(t) - \phi_{\widetilde{X}}(t)|^2 w_1(t)\, dt.$$

Clearly $\mathcal{Q}_{\cdot,\delta} \to \mathcal{Q}$ as $\delta \to 0$. Hence, $\mathcal{Q}_{n,\delta} \to \mathcal{Q}$ a.s. and $\mathcal{Q}^\star_{n,\delta} \to \mathcal{Q}$ a.s. as $\delta \to 0$ and $n \to \infty$. In order to show $\mathcal{Q}_n \to \mathcal{Q}$ a.s. and $\mathcal{Q}^\star_n \to \mathcal{Q}$ a.s. as $n \to \infty$, it remains to prove that we have almost surely

$$\limsup_{\delta\to 0}\limsup_{n\to\infty} |\mathcal{Q}_{n,\delta} - \mathcal{Q}_n| = \limsup_{\delta\to 0}\limsup_{n\to\infty} |\mathcal{Q}^\star_{n,\delta} - \mathcal{Q}^\star_n| = 0.$$

To this end, define a mixture of $\widetilde{X}$ and X as $Y_{-c} = (\widetilde{X}_1, \ldots, \widetilde{X}_{c-1}, X_{c^+})$ for all $c \in \{1, \ldots, d-1\}$. By the Cauchy–Schwarz inequality,

$$
\begin{aligned}
|\xi_n(t)|^2 &= \left| \phi_X^n(t) - \prod_{j=1}^d \phi_{X_j}^n(t_j) \right|^2 \\
&= \left| \phi_X^n(t) - \prod_{j=1}^d \phi_{X_j}^n(t_j) - \sum_{c=1}^{d-2} \left\{ \prod_{j=1}^c \phi_{X_j}^n(t_j) \phi_{X_{c^+}}^n(t_{c^+}) \right\} + \sum_{c=1}^{d-2} \left\{ \prod_{j=1}^c \phi_{X_j}^n(t_j) \phi_{X_{c^+}}^n(t_{c^+}) \right\} \right|^2 \\
&\le \left[|\phi_X^n(t) - \phi_{X_1}^n(t_1)\phi_{X_{1^+}}^n(t_{1^+})| + \sum_{c=1}^{d-2} \left| \left\{ \prod_{j=1}^c \phi_{X_j}^n(t_j) \phi_{X_{c^+}}^n(t_{c^+}) \right\} - \left\{ \prod_{j=1}^c \phi_{X_j}^n(t_j) \phi_{X_{c+1}}^n(t_{c+1}) \phi_{X_{(c+1)^+}}^n(t_{(c+1)^+}) \right\} \right| \right]^2 \\
&= \left\{ \sum_{c=1}^{d-1} |\phi_{(X_c, Y_{-c})}^n(t_c, t_{-c}) - \phi_{X_c}^n(t_c)\phi_{Y_{-c}}^n(t_{-c})| \right\}^2 \\
&\le (d-1) \sum_{c=1}^{d-1} |\phi_{(X_c, Y_{-c})}^n(t_c, t_{-c}) - \phi_{X_c}^n(t_c)\phi_{Y_{-c}}^n(t_{-c})|^2,
\end{aligned}
$$

and

$$
\begin{aligned}
|\xi_n^\star(t)|^2 &= \left| \frac{1}{n} \sum_{k=1}^n e^{i\langle t, X^k \rangle} - \frac{1}{n} \sum_{k=1}^n e^{i \sum_{j=1}^d \langle t_j, X_j^{k+j-1} \rangle} \right|^2 \\
&= \left| \frac{1}{n} \sum_{k=1}^n \left(e^{i\langle t, X^k \rangle} - \sum_{c=2}^{d-1} e^{i\langle t, (X_1^k, \ldots, X_c^{k+c-1}, X_{c^+}^k) \rangle} + \sum_{c=2}^{d-1} e^{i\langle t, (X_1^k, \ldots, X_c^{k+c-1}, X_{c^+}^k) \rangle} - e^{i \sum_{j=1}^d \langle t_j, X_j^{k+j-1} \rangle} \right) \right|^2 \\
&= \left| \frac{1}{n} \sum_{k=1}^n \sum_{c=1}^{d-1} \left(e^{i\langle t, (X_1^k, \ldots, X_c^{k+c-1}, X_{c^+}^k) \rangle} - e^{i\langle t, (X_1^k, \ldots, X_{c+1}^{k+c}, X_{(c+1)^+}^k) \rangle} \right) \right|^2 \\
&\le (d-1) \sum_{c=1}^{d-1} \left| \frac{1}{n} \sum_{k=1}^n e^{i\langle t_{-(c+1)}, (X_1^k, \ldots, X_c^{k+c-1}, X_{(c+1)^+}^k) \rangle} (e^{i\langle t_{c+1}, X_{c+1}^k \rangle} - e^{i\langle t_{c+1}, X_{c+1}^{k+c} \rangle}) \right|^2 \\
&\le (d-1) \sum_{c=1}^{d-1} \left(\frac{1}{n} \sum_{k=1}^n |e^{i\langle t_{c+1}, X_{c+1}^k \rangle} - e^{i\langle t_{c+1}, X_{c+1}^{k+c} \rangle}|^2 \right) \\
&\le (d-1) \sum_{c=2}^{d} \frac{2}{n} \sum_{k=1}^n (|e^{i\langle t_c, X_c^k \rangle} - \phi_{X_c}(t_c)|^2 + |\phi_{X_c}(t_c) - e^{i\langle t_c, X_c^{k+c-1} \rangle}|^2) \\
&= 4(d-1) \sum_{c=2}^{d} \frac{1}{n} \sum_{k=1}^n |e^{i\langle t_c, X_c^k \rangle} - \phi_{X_c}(t_c)|^2.
\end{aligned}
$$

By the inequality $sa + (1-s)b \ge a^s b^{1-s}$, valid for all $s \in (0, 1)$ when $a, b > 0$, we have

$$
\begin{aligned}
|t|_p^{1+p} = (|t_c|_{p_c}^2 + |t_{-c}|_{p_{-c}}^2)^{(1+p)/2} &\ge \left(\frac{1+p_c}{2+p} |t_c|_{p_c}^2 + \frac{1+p_{-c}}{2+p} |t_{-c}|_{p_{-c}}^2 \right)^{(1+p)/2} \\
&\ge \left(|t_c|_{p_c}^{\frac{2(1+p_c)}{2+p}} |t_{-c}|_{p_{-c}}^{\frac{2(1+p_{-c})}{2+p}} \right)^{(1+p)/2} = |t_c|_{p_c}^{\frac{1+p_{-c}}{2+p}+p_c} |t_{-c}|_{p_{-c}}^{\frac{1+p_c}{2+p}+p_{-c}} \equiv |t_c|_{p_c}^{m_c+p_c} |t_{-c}|_{p_{-c}}^{m_{-c}+p_{-c}},
\end{aligned}
$$

where $p_{-c} = p - p_c$, $m_c \in (0, 1)$, $m_{-c} \in (0, 1)$ and consequently

$$
\begin{aligned}
w_1(t) = \frac{1}{K(p,1)|t|_p^{1+p}} &\le \frac{K(p_c, m_c)K(p_{-c}, m_{-c})}{K(p,1)} \frac{1}{K(p_c, m_c)|t_c|_{p_c}^{m_c+p_c}} \frac{1}{K(p_{-c}, m_{-c})|t_{-c}|_{p_{-c}}^{m_{-c}+p_{-c}}} \\
&\equiv C(p, p_c, p_{-c}) \frac{1}{K(p_c, m_c)|t_c|_{p_c}^{m_c+p_c}} \frac{1}{K(p_{-c}, m_{-c})|t_{-c}|_{p_{-c}}^{m_{-c}+p_{-c}}},
\end{aligned}
$$

where $C(p, p_c, p_{-c})$ is a constant depending only on p, p_c, p_{-c}.

By the fact that $\mathbb{R}^p \setminus D(\delta) \subset \{|t_c|^2_{p_c}, |t_{-c}|^2_{p_{-c}} < \delta\} \cup \{|t_c|^2_{p_c} > 1/\delta\} \cup \{|t_{-c}|^2_{p_{-c}} > 1/\delta\}$ and similar steps as in the proof of Theorem 2 in [37], we have almost surely

$$\begin{aligned}
\limsup_{\delta\to 0}\limsup_{n\to\infty}|\mathcal{Q}_{n,\delta}-\mathcal{Q}_n| &= \limsup_{\delta\to 0}\limsup_{n\to\infty}\int_{\mathbb{R}^p\setminus D(\delta)}|\xi_n(t)|^2 w_1(t)\,dt\\
&\le (d-1)\sum_{c=1}^{d-1}\limsup_{\delta\to 0}\limsup_{n\to\infty}\int_{\mathbb{R}^p\setminus D(\delta)}|\phi^n_{(X_c,Y_{-c})}(t_c,t_{-c})-\phi^n_{X_c}(t_c)\phi^n_{Y_{-c}}(t_{-c})|^2 w_1(t)\,dt\\
&\le C(p,p_c,p_{-c})(d-1)\sum_{c=1}^{d-1}\limsup_{\delta\to 0}\limsup_{n\to\infty}\int_{\mathbb{R}^p\setminus D(\delta)}|\phi^n_{(X_c,Y_{-c})}(t_c,t_{-c})\\
&\qquad -\phi^n_{X_c}(t_c)\phi^n_{Y_{-c}}(t_{-c})|^2\frac{1}{K(p_c,m_c)|t_c|^{m_c+p_c}_{p_c}}\frac{1}{K(p_{-c},m_{-c})|t_{-c}|^{m_{-c}+p_{-c}}_{p_{-c}}}\,dt_c\,dt_{-c}\\
&=0
\end{aligned}$$

and

$$\begin{aligned}
\limsup_{\delta\to 0}\limsup_{n\to\infty}|\mathcal{Q}^\star_{n,\delta}-\mathcal{Q}^\star_n| &= \limsup_{\delta\to 0}\limsup_{n\to\infty}\int_{\mathbb{R}^p\setminus D(\delta)}|\xi^\star_n(t)|^2 w_1(t)\,dt\\
&\le 4(d-1)\sum_{c=2}^{d}\limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\int_{\mathbb{R}^p\setminus D(\delta)}\left|e^{i\langle t_c,X_c^k\rangle}-\phi_{X_c}(t_c)\right|^2 w_1(t)\,dt\\
&\le C(p,p_c,p_{-c})4(d-1)\sum_{c=2}^{d}\limsup_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\int_{\mathbb{R}^p\setminus D(\delta)}\left|e^{i\langle t_c,X_c^k\rangle}-\phi_{X_c}(t_c)\right|^2\\
&\qquad\times\frac{1}{K(p_c,m_c)|t_c|^{m_c+p_c}_{p_c}}\frac{1}{K(p_{-c},m_{-c})|t_{-c}|^{m_{-c}+p_{-c}}_{p_{-c}}}\,dt_c\,dt_{-c}\\
&=0.
\end{aligned}$$

Therefore, we have almost surely

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}|\mathcal{Q}_{n,\delta}-\mathcal{Q}_n| = \limsup_{\delta\to 0}\limsup_{n\to\infty}|\mathcal{Q}^\star_{n,\delta}-\mathcal{Q}^\star_n| = 0,$$

and the proof is complete. □

Proof of Theorem 3. First assume that $\mathcal{H}_0$ holds. Let $\zeta(t)$ denote a complex-valued Gaussian process with mean zero and covariance function

$$R(t,t^0)=\prod_{j=1}^{d}\phi_{X_j}(t_j-t_j^0)+(d-1)\prod_{j=1}^{d}\phi_{X_j}(t_j)\overline{\phi_{X_j}(t_j^0)}\sum_{j=1}^{d}\phi_{X_j}(t_j-t_j^0)\prod_{j'\neq j}\phi_{X_{j'}}(t_{j'})\overline{\phi_{X_{j'}}(t_{j'}^0)}.$$

Define $n\mathcal{Q}_n = n\|\phi^n_X(t)-\phi^n_{\tilde{X}}(t)\|^2_{w_1}\equiv\|\zeta_n(t)\|^2_{w_1}$. After a simple calculation, we find $\mathrm{E}\{\zeta_n(t)\}=\mathrm{E}\{\zeta^\star_n(t)\}=0$ and

$$\begin{aligned}
\mathrm{E}\{\zeta_n(t)\overline{\zeta_n(t^0)}\} &= \left(1-\frac{1}{n^{d-1}}\right)\prod_{j=1}^{d}\phi_{X_j}(t_j-t_j^0)+\left\{n-1-\frac{(n-1)^d}{n^{d-1}}\right\}\prod_{j=1}^{d}\phi_{X_j}(t_j)\overline{\phi_{X_j}(t_j^0)}\\
&\quad-\frac{(n-1)^{d-1}}{n^{d-1}}\left\{\sum_{j=1}^{d}\phi_{X_j}(t_j-t_j^0)\prod_{j'\neq j}\phi_{X_j}(t_j)\overline{\phi_{X_j}(t_j^0)}\right\}+o_n(1),
\end{aligned}$$

which converges to $R(t,t^0)$ as $n\to\infty$. In particular, $\mathrm{E}|\zeta_n(t)|^2\to R(t,t)\le d$ as $n\to\infty$, and thus $\mathrm{E}|\zeta_n(t)|^2\le d+1$ for enough large n.

For all $\delta\in(0,1)$, define the region $D(\delta)$ as in (3). Given any $\epsilon>0$, we choose a partition $\{D^1(\delta),\ldots,D^N(\delta)\}$ of $D(\delta)$ into $N(\epsilon)$ measurable sets with diameter at most ϵ, and suppress the notation of $D(\delta)$, $D^\ell(\delta)$ as D, D^ℓ. Then for any fixed $t^\ell\in D^\ell$ with $\ell\in\{1,\ldots,N\}$, we define a sequence of random variables

$$Q_n(\delta)=\sum_{\ell=1}^{N}\int_{D^\ell}|\zeta_n(t^\ell)|^2 w_1(t)\,dt.$$

For any fixed $M > 0$, let $\beta(\epsilon) = \sup_{t,t^0} \mathrm{E}||\zeta_n(t)|^2 - |\zeta_n(t^0)|^2|$, where the supremum is taken over all $t = (t_1, \ldots, t_d)$ and $t^0 = (t_1^0, \ldots, t_d^0)$ such that

$$\max(|t|_p^2, |t^0|_p^2) \le M \quad \text{and} \quad |t - t^0|_p^2 = \sum_{j=1}^d |t_j - t_j^0|_p^2 \le \epsilon^2.$$

By the Continuous Mapping Theorem and $\zeta_n(t) \to \zeta_n(t^0)$ as $\epsilon \to 0$, we have $|\zeta_n(t)|^2 \to |\zeta_n(t^0)|^2$ as $\epsilon \to 0$. By Lebesgue's Dominated Convergence Theorem and the fact that $\mathrm{E}|\zeta_n(t)|^2 \le d + 1$ for large enough n, we have $\mathrm{E}||\zeta_n(t)|^2 - |\zeta_n(t^0)|^2| \to 0$ as $\epsilon \to 0$, which leads to $\beta(\epsilon) \to 0$ as $\epsilon \to 0$. Thus as $\epsilon \to 0$,

$$\begin{aligned} \mathrm{E}\left|\int_D |\zeta_n(t)|^2 w_1(t)\,dt - Q_n(\delta)\right| &= \mathrm{E}\left|\sum_{\ell=1}^N \int_{D^\ell} \{|\zeta_n(t)|^2 - |\zeta_n(t^\ell)|^2\} w_1(t)\,dt\right| \\ &\le \sum_{\ell=1}^N \int_{D^\ell} \mathrm{E}\left||\zeta_n(t)|^2 - |\zeta_n(t^\ell)|^2\right| w_1(t)\,dt \le \beta(\epsilon)\int_D 1 w_1(t)\,dt \to 0. \end{aligned}$$

By similar steps in Theorem 2, we have, as $\delta \to 0$,

$$\mathrm{E}\left|\int_D |\zeta_n(t)|^2 w_1(t)\,dt - \|\zeta_n\|_{w_1}^2\right| \to 0 \quad \text{and} \quad \mathrm{E}\left|\int_D |\zeta_n^\star(t)|^2 w_1(t)\,dt - \|\zeta_n^\star\|_{w_1}^2\right| \to 0.$$

Therefore, $\mathrm{E}|Q_n(\delta) - \|\zeta_n\|_{w_1}^2| \to 0$ as $\epsilon, \delta \to 0$ and $\mathrm{E}|Q_n^\star(\delta) - \|\zeta_n^\star\|_{w_1}^2| \to 0$ as $\epsilon, \delta \to 0$.

On the other hand, for any fixed $t^\ell \in D^\ell$ with $\ell \in \{1, \ldots, N\}$, define a random variable

$$Q(\delta) = \sum_{\ell=1}^N \int_{D^\ell} |\zeta(t^\ell)|^2 w_1(t)\,dt.$$

Similarly, we have $\mathrm{E}|Q(\delta) - \|\zeta\|_{w_1}^2| \to 0$ as $\epsilon, \delta \to 0$. By the multivariate Central Limit Theorem, the delta method and the Continuous Mapping Theorem, we have, as $n \to \infty$,

$$Q_n(\delta) = \sum_{\ell=1}^N \int_{D^\ell} |\zeta_n(t^\ell)|^2 w_1(t)\,dt \rightsquigarrow Q(\delta) = \sum_{\ell=1}^N \int_{D^\ell} |\zeta(t^\ell)|^2 w_1(t)\,dt.$$

Therefore, $\|\zeta_n\|_{w_1}^2 \to_{\mathcal{D}} \|\zeta\|_{w_1}^2$ as $\epsilon, \delta \to 0$, $n \to \infty$, since the $Q_n(\delta)$s have the following properties:

(a) $Q_n(\delta)$ converges in distribution to $Q(\delta)$ as $n \to \infty$;

(b) $\mathrm{E}|Q_n(\delta) - \|\zeta_n\|_{w_1}^2| \to 0$ as $\epsilon, \delta \to 0$;

(c) $\mathrm{E}|Q(\delta) - \|\zeta\|_{w_1}^2| \to 0$ as $\epsilon, \delta \to 0$.

Analogous to $\zeta(t)$, $\zeta_n(t)$, $\beta(\epsilon)$, $Q(\delta)$, $Q_n(\delta)$ for $\mathcal{Q}_n$, we can define $\zeta^\star(t)$, $\zeta_n^\star(t)$, $\beta^\star(\epsilon)$, $Q^\star(\delta)$, $Q_n^\star(\delta)$ for $\mathcal{Q}_n^\star$, and prove that $\|\zeta_n^\star\|_{w_1}^2 \to_{\mathcal{D}} \|\zeta^\star\|_{w_1}^2$ as $\epsilon, \delta \to 0$, $n \to \infty$ through the same derivations. The only differences are that

$$\mathrm{E}\{\zeta_n^\star(t)\overline{\zeta_n^\star(t^0)}\} = 2R(t, t^0) \quad \text{and} \quad \mathrm{E}|\zeta_n^\star(t)|^2 = 2R(t, t) \le 2d + 1$$

for large enough n.

Finally, assume that $\mathcal{H}_A$ holds. By Theorems 1 and 2, we have $\mathcal{Q}_n \to \mathcal{Q} > 0$ a.s. as $n \to \infty$. Therefore, $n\mathcal{Q}_n \to \infty$ a.s. as $n \to \infty$. Similarly, we can prove that $n\mathcal{Q}_n^\star \to \infty$ a.s. as $n \to \infty$ through the same derivations. □

Proof of Theorem 4. First observe that $\mathrm{E}|X|_p < \infty$, and hence $0 \le \mathcal{V}^2(X_c, X_{c^+}) < \infty$ for all $c \in \{1, \ldots, d-1\}$. Thus,

$$0 \le \mathcal{R}(X) = \sum_{c=1}^{d-1} \mathcal{V}^2(X_c, X_{c^+}) < \infty.$$

Similarly, we have

$$0 \le \mathcal{S}(X) = \sum_{c=1}^{d} \mathcal{V}^2(X_c, X_{-c}) < \infty.$$

We will now show that

$$\mathcal{R}(X) = \sum_{c=1}^{d-1} \mathcal{V}^2(X_c, X_{c^+}) = 0 \Leftrightarrow \mathcal{S}(X) = \sum_{c=1}^{d} \mathcal{V}^2(X_c, X_{-c}) = 0 \Leftrightarrow X_1, \ldots, X_d \text{ are mutually independent.}$$

First assume that $X_1, \ldots, X_d$ are mutually independent. Then X_c and X_{c^+} are independent for all $c \in \{1, \ldots, d-1\}$. By Theorem 3 in [37], $\mathcal{V}^2(X_c, X_{c^+}) = 0$ for all $c \in \{1, \ldots, d-1\}$. As a result, $\mathcal{R}(X) = 0$. Similarly, we can prove that $\mathcal{S}(X) = 0$, since X_c and X_{-c} are independent for all $c \in \{1, \ldots, d\}$.

To show the converse, first assume $\mathcal{R}(X) = 0$, then $\mathcal{V}^2(X_c, X_{c^+}) = 0$ for all $c \in \{1, \ldots, d-1\}$. By Theorem 3 in [37], X_c and X_{c^+} are independent, for all $c \in \{1, \ldots, d-1\}$. Thus, for all $t \in \mathbb{R}^p$, we have

$$\phi_{(X_j, X_{j^+})}(t_j, t_{j^+}) - \phi_{X_j}(t_j)\phi_{X_{j^+}}(t_{j^+}) = 0,$$

where for each $j \in \{1, \ldots, d\}$, ϕ_{X_j} and $\phi_{X_{j+}}$ denote the marginal and $\phi_{(X_j, X_{j^+})}$ denotes the joint characteristic function of X_j and X_{j^+} respectively. For all $t \in \mathbb{R}^p$, we have

$$\begin{aligned}
\left|\phi_X(t) - \prod_{j=1}^{d} \phi_{X_j}(t_j)\right| &= \left|\phi_X(t) - \prod_{j=1}^{d} \phi_{X_j}(t_j) - \sum_{c=1}^{d-2}\left\{\prod_{j=1}^{c} \phi_{X_j}(t_j)\phi_{X_{c^+}}(t_{c^+})\right\} + \sum_{c=1}^{d-2}\left\{\prod_{j=1}^{c} \phi_{X_j}(t_j)\phi_{X_{c^+}}(t_{c^+})\right\}\right| \\
&\le |\phi_X(t) - \phi_{X_1}(t_1)\phi_{X_{1^+}}(t_{1^+})| + \sum_{c=1}^{d-2}\left|\prod_{j=1}^{c} \phi_{X_j}(t_j)\phi_{X_{c^+}}(t_{c^+}) - \prod_{j=1}^{c} \phi_{X_j}(t_j)\phi_{X_{c+1}}(t_{c+1})\phi_{X_{(c+1)^+}}(t_{(c+1)^+})\right| \\
&\le |\phi_X(t) - \phi_{X_1}(t_1)\phi_{X_{1^+}}(t_{1^+})| + \sum_{c=1}^{d-2}\left|\prod_{j=1}^{c} \phi_{X_j}(t_j)\right| |\phi_{X_{c^+}}(t_{c^+}) - \phi_{X_{c+1}}(t_{c+1})\phi_{X_{(c+1)^+}}(t_{(c+1)^+})| \\
&\le |\phi_X(t) - \phi_{X_1}(t_1)\phi_{X_{1^+}}(t_{1^+})| + \sum_{c=1}^{d-2}|\phi_{X_{c^+}}(t_{c^+}) - \phi_{X_{c+1}}(t_{c+1})\phi_{X_{(c+1)^+}}(t_{(c+1)^+})| \\
&= \sum_{c=1}^{d-1}|\phi_{(X_c, X_{c^+})}(t_c, t_{c^+}) - \phi_{X_c}(t_c)\phi_{X_{c^+}}(t_{c^+})| = 0.
\end{aligned}$$

Therefore, for all $t \in \mathbb{R}^p$, we have $|\phi_X(t) - \prod_{j=1}^{d}\phi_{X_j}(t_j)| = 0$, which implies that $X_1, \ldots, X_d$ are mutually independent. Similarly, we can prove that $\mathcal{S}(X) = 0$ implies that $X_1, \ldots, X_d$ are mutually independent, since X_c and X_{-c} are independent implies that X_c and X_{c^+} are independent. □

Proof of Theorem 5. By Theorem 2 in [37], we have

$$\lim_{n\to\infty} \mathcal{V}_n^2(\mathbf{X_c}, \mathbf{X_{c^+}}) = \mathcal{V}^2(X_c, X_{c^+})$$

for all $c \in \{1, \ldots, d-1\}$, and

$$\lim_{n\to\infty} \mathcal{V}_n^2(\mathbf{X_c}, \mathbf{X_{-c}}) = \mathcal{V}^2(X_c, X_{-c})$$

for all $c \in \{1, \ldots, d\}$. Therefore, as $n \to \infty$, the limit of sum converges to the sum of limit as

$$\mathcal{R}_n(\mathbf{X}) \xrightarrow{\text{a.s.}} \mathcal{R}(X) \quad \text{and} \quad \mathcal{S}_n(\mathbf{X}) \xrightarrow{\text{a.s.}} \mathcal{S}(X).$$

This concludes the argument. □

Proof of Theorem 6. First assume that $\mathcal{H}_0$ holds. Define

$$n\mathcal{R}_n(\mathbf{X}) = n\sum_{c=1}^{d-1} \mathcal{V}_n^2(\mathbf{X}_c, \mathbf{X}_{c^+}) \equiv \sum_{c=1}^{d-1} \|\zeta_n^c(t_{(c-1)^+})\|_{w_0}^2,$$

which is the sum corresponding to the pairs $(X_{d-1}, X_d), (X_{d-2}, (X_{d-1}, X_d)), (X_{d-3}, (X_{d-2}, X_{d-1}, X_d)), \ldots, (X_1(X_2, \ldots, X_d))$. Any two of them can be reorganized as (X_1, X_2) and $(X_4, (X_1, X_2, X_3))$ where X_3 could be empty. Without loss of generality, we will show that $\phi^n_{(X_1,X_2)}(t_1, t_2) - \phi^n_{X_1}(t_1)\phi^n_{X_2}(t_2)$ and $\phi^n_{(X_1,X_2,X_3,X_4)}(s_1, s_2) - \phi^n_{(X_1,X_2,X_3)}(s_1) \times \phi^n_{X_4}(s_2)$ are uncorrelated. Then it follows that the variables $\zeta_n^c(t_{(c-1)^+})$ are uncorrelated for all $c \in \{1, \ldots, d-1\}$.

After a simple calculation, we find

$$\mathrm{E}\left\{\phi^n_{(X_1,X_2)}(t_1, t_2) - \phi^n_{X_1}(t_1)\phi^n_{X_2}(t_2)\right\} = \mathrm{E}\left\{\phi^n_{(X_1,X_2,X_3,X_4)}(s_1, s_2) - \phi^n_{(X_1,X_2,X_3)}(s_1)\phi^n_{X_4}(s_2)\right\} = 0$$

and

$$\mathrm{E}\left\{\phi^n_{(X_1,X_2)}(t_1, t_2) - \phi^n_{X_1}(t_1)\phi^n_{X_2}(t_2)\right\}\left\{\overline{\phi^n_{(X_1,X_2,X_3,X_4)}(s_1, s_2) - \phi^n_{(X_1,X_2,X_3)}(s_1)\phi^n_{X_4}(s_2)}\right\} = 0.$$

As a result

$$\mathrm{cov}\{\phi^n_{(X_1,X_2)}(t_1, t_2) - \phi^n_{X_1}(t_1)\phi^n_{X_2}(t_2), \overline{\phi^n_{(X_1,X_2,X_3,X_4)}(s_1, s_2) - \phi^n_{(X_1,X_2,X_3)}(s_1)\phi^n_{X_4}(s_2)}\} = 0.$$

Let $p_{(c-1)^+} = p_c + \cdots + p_d$. For all $\delta > 0$, define the region

$$D_c(\delta) = \left\{ t_{(c-1)^+} = (t_c, t_{c^+}) = (t_c, \ldots, t_d) : \delta \le |t_{(c-1)^+}|^2_{p_{(c-1)^+}} = \sum_{j=c}^{d} |t_j|^2_{p_j} \le 2/\delta \right\}.$$

Given arbitrary $\epsilon > 0$, we choose a partition $\{D_c^1, \ldots, D_c^{N_c}\}$ of $D_c(\delta)$ into $N_c(\epsilon)$ measurable sets with diameter at most ϵ, and define a sequence of random variables for any fixed $t^\ell_{(c-1)^+} \in D_c^\ell$ with $\ell \in \{1, \ldots, N_c\}$ as

$$Q_n^c(\delta) = \sum_{\ell=1}^{N_c} \int_{D_c^\ell} |\zeta_n^c(t^\ell_{(c-1)^+})|^2 w_0(t)\,dt.$$

Let $\zeta^c(t_{(c-1)^+}) = \zeta^c(t_c, t_{c^+})$ denote a complex-valued Gaussian process with mean zero and covariance function

$$R_c^\zeta(t_{(c-1)^+}, t^0_{(c-1)^+}) = \left\{ \phi_{X_c}(t_c - t_c^0) - \phi_{X_c}(t_c)\overline{\phi_{X_c}(t_c^0)} \right\} \left\{ \phi_{X_{c^+}}(t_{c^+} - t^0_{c^+}) - \phi_{X_{c^+}}(t_{c^+})\overline{\phi_{X_{c^+}}(t^0_{c^+})} \right\}.$$

By the multivariate Central Limit Theorem, the delta method and the Continuous Mapping Theorem, we have

$$\begin{pmatrix} Q_n^1(\delta) - \sum_{\ell=1}^{N_1} \int_{D_1^\ell} |\zeta^1(t^\ell)|^2 w_0(t)\,dt \\ \vdots \\ Q_n^{d-1}(\delta) - \sum_{\ell=1}^{N_{d-1}} \int_{D_{d-1}^\ell} |\zeta^{d-1}(t^\ell_{(d-2)^+})|^2 w_0(t)\,dt \end{pmatrix} \rightsquigarrow \begin{pmatrix} \sum_{\ell=1}^{N_1} \int_{D_1^\ell} |\zeta^1(t^\ell)|^2 w_0(t)\,dt \\ \vdots \\ \sum_{\ell=1}^{N_{d-1}} \int_{D_{d-1}^\ell} |\zeta^{d-1}(t^\ell_{(d-2)^+})|^2 w_0(t)\,dt \end{pmatrix}$$

as $n \to \infty$ with asymptotic mutual independence. Thus, the variables $Q_n^1(\delta), \ldots, Q_n^{d-1}(\delta)$ are asymptotically mutually independent. By similar steps in the proof of Theorem 5 in [37], we have $\mathrm{E}|Q_n^c(\delta) - \|\zeta_n^c(t_{(c-1)^+})\|^2_{w_0}| \to 0$ for all $c \in \{1, \ldots, d-1\}$ as $\epsilon, \delta \to 0$. Hence

$$\begin{pmatrix} \|\zeta_n^1(t)\|^2_{w_0} - Q_n^1(\delta) \\ \vdots \\ \|\zeta_n^{d-1}(t_{(d-2)^+})\|^2_{w_0} - Q_n^{d-1}(\delta) \end{pmatrix} \to_{\mathcal{P}} \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

as $\epsilon, \delta \to 0$. By the multivariate Slutsky's theorem, we also have

$$\begin{pmatrix} \|\zeta_n^1(t)\|^2_{w_0} \\ \vdots \\ \|\zeta_n^{d-1}(t_{(d-2)^+})\|^2_{w_0} \end{pmatrix} \rightsquigarrow \begin{pmatrix} \|\zeta^1(t)\|^2_{w_0} \\ \vdots \\ \|\zeta^{d-1}(t_{(d-2)^+})\|^2_{w_0} \end{pmatrix}$$

as $\epsilon, \delta \to 0$ and $n \to \infty$ with asymptotic mutual independence. Therefore, there is asymptotic mutual independence between the variables $\|\zeta_n^c(t_{(c-1)^+})\|^2_{w_0}$ defined for all $c \in \{1, \ldots, d-1\}$. Furthermore, as $n \to \infty$,

$$\sum_{c=1}^{d-1} \|\zeta_n^c(t_{(c-1)^+})\|^2_{w_0} \rightsquigarrow \sum_{c=1}^{d-1} \|\zeta^c(t_{(c-1)^+})\|^2_{w_0}.$$

Analogous to $\zeta_n^c(t_{(c-1)^+})$, $\zeta^c(t_{(c-1)^+})$, $R_c^\zeta(t_{(c-1)^+}, t^0_{(c-1)^+})$ for $\mathcal{R}_n(\mathbf{X})$, we can define $\eta_n^c(t)$, $\eta^c(t)$, $R_c^\eta(t, t^0)$ for $\mathcal{S}_n(\mathbf{X})$, and prove through similar derivations that the variables $\|\eta_n^c(t)\|^2_{w_0}$ defined for all $c \in \{1, \ldots, d\}$ are asymptotically mutually independent, and that as $n \to \infty$,

$$\sum_{c=1}^{d} \|\eta_n^c(t)\|^2_{w_0} \rightsquigarrow \sum_{c=1}^{d} \|\eta_c(t)\|^2_{w_0}.$$

The only difference is that one now needs to show that the variables $\phi^n_{(X_1,X_2,X_3)}(t_1, t_2, t_3) - \phi^n_{X_1}(t_1)\phi^n_{(X_2,X_3)}(t_2, t_3)$ and $\phi^n_{(X_1,X_2,X_3)}(s_1, s_2, s_3) - \phi^n_{X_2}(s_2)\phi^n_{(X_1,X_3)}(s_1, s_3)$ are asymptotically uncorrelated.

Finally, suppose that $\mathcal{H}_A$ holds. By Theorem 4, we then have $\mathcal{R}_n \to \mathcal{R} > 0$ a.s. as $n \to \infty$. Therefore, $n\mathcal{R}_n \to \infty$ a.s. as $n \to \infty$. Similarly, we can prove that $n\mathcal{S}_n \to \infty$ a.s. as $n \to \infty$ through the same derivations. Further note that under $\mathcal{H}_A$, the variables $\zeta_n^c(t_{(c-1)^+})$ defined for all $c \in \{1, \ldots, d-1\}$ are not asymptotically uncorrelated, and that the variables $\eta_n^c(t)$ defined for all $c \in \{1, \ldots, d\}$ are not asymptotically uncorrelated. □

Complete measure of mutual dependence using weight function w_2. Except that $\mathcal{U}_n(\mathbf{X})$ requires the additional dth moment condition $\mathrm{E}(|X_1|_{p_1}\cdots|X_d|_{p_d}) < \infty$ to be simplified, $\mathcal{U}(X)$ is in an extremely complicated form. Even when $d = 3$, $\mathcal{U}(X)$ already has 12 different terms as follows:

$$\begin{aligned}
\mathcal{U}(X) &= \|\phi_X(t) - \phi_{\widetilde{X}}(t)\|_{w_2}^2 \\
&= -\mathrm{E}|X_1 - X_1'|_{p_1}|X_2 - X_2'|_{p_2}|X_3 - X_3'|_{p_3} + 2\mathrm{E}|X_1 - X_1'|_{p_1}|X_2 - X_2''|_{p_2}|X_3 - X_3'''|_{p_3} \\
&- \ \mathrm{E}|X_1 - X_1'|_{p_1}\mathrm{E}|X_2 - X_2'|_{p_2}\mathrm{E}|X_3 - X_3'|_{p_3} \\
&\quad + \mathrm{E}|X_1 - X_1'|_{p_1}|X_2 - X_2'|_{p_2} + \mathrm{E}|X_1 - X_1'|_{p_1}|X_3 - X_3'|_{p_3} + \mathrm{E}|X_2 - X_2'|_{p_2}|X_3 - X_3'|_{p_3} \\
&\quad - 2\mathrm{E}|X_1 - X_1'|_{p_1}|X_2 - X_2''|_{p_2} - 2\mathrm{E}|X_1 - X_1'|_{p_1}|X_3 - X_3'''|_{p_3} - 2\mathrm{E}|X_2 - X_2''|_{p_2}|X_3 - X_3'''|_{p_3} \\
&\quad + \mathrm{E}|X_1 - X_1'|_{p_1}\mathrm{E}|X_2 - X_2'|_{p_2} + \mathrm{E}|X_1 - X_1'|_{p_1}\mathrm{E}|X_3 - X_3'|_{p_3} + \mathrm{E}|X_2 - X_2'|_{p_2}\mathrm{E}|X_3 - X_3'|_{p_3} \\
&= -\mathrm{E}|X_1 - X_1'|_{p_1}|X_2 - X_2'|_{p_2}|X_3 - X_3'|_{p_3} + 2\mathrm{E}|X_1 - X_1'|_{p_1}|X_2 - X_2''|_{p_2}|X_3 - X_3'''|_{p_3} \\
&\quad - \mathrm{E}|X_1 - X_1'|_{p_1}\mathrm{E}|X_2 - X_2'|_{p_2}\mathrm{E}|X_3 - X_3'|_{p_3} + \sum_{1\le i<j\le 3}\mathrm{E}|X_i - X_i'|_{p_i}|X_j - X_j'|_{p_j} \\
&\quad - 2\sum_{1\le i<j\le 3}\mathrm{E}|X_i - X_i'|_{p_i}|X_j - X_j''|_{p_j} + \sum_{1\le i<j\le 3}\mathrm{E}|X_i - X_i'|_{p_i}\mathrm{E}|X_j - X_j'|_{p_j}.
\end{aligned}$$

In general, the number of different terms in $\mathcal{U}(X)$ grows exponentially as d increases. Basically, we will see all combinations of all components in all moments as expectations.

References

[1] R. Beran, M. Bilodeau, P. Lafaye de Micheaux, Nonparametric tests of independence between random vectors, J. Multivariate Anal. 98 (2007) 1805–1824.
[2] W. Bergsma, A. Dassios, A consistent test of independence based on a sign covariance related to Kendall's tau, Bernoulli 20 (2014) 1006–1028.
[3] M. Bilodeau, P. Lafaye de Micheaux, A multivariate empirical characteristic function test of independence with normal marginals, J. Multivariate Anal. 95 (2005) 345–369.
[4] M. Bilodeau, A.G. Nangue, Tests of mutual or serial independence of random vectors with applications, J. Mach. Learn. Res. 18 (2017) 2518–2557.
[5] P. Ding, A paradox from randomization-based causal inference, Statist. Sci. 32 (2017) 331–345.
[6] E.F. Fama, K.R. French, Common risk factors in the returns on stocks and bonds, J. Financial Econom. 33 (1993) 3–56.
[7] E.F. Fama, K.R. French, A five-factor asset pricing model, J. Financial Econom. 16 (2015) 1–22.
[8] J. Fan, Y. Feng, L. Xia, A conditional dependence measure with applications to undirected graphical models, 2015. arXiv preprint arXiv:1501.01617.
[9] Y. Fan, P. Lafaye de Micheaux, S. Penev, D. Salopek, Multivariate nonparametric test of independence, J. Multivariate Anal. 153 (2017) 189–210.
[10] L. Gan, N.N. Narisetty, F. Liang, Bayesian regularization for graphical models with unequal shrinkage, J. Amer. Statist. Assoc. (2018) in press.
[11] C. Genest, J.-F. Quessy, B. Rémillard, Asymptotic local efficiency of Cramér–von Mises tests for multivariate independence, Ann. Statist. 35 (2007) 166–191.
[12] C. Genest, B. Rémillard, Test of independence and randomness based on the empirical copula process, Test 13 (2004) 335–369.
[13] A. Gretton, O. Bousquet, A. Smola, B. Scholkopf, Measuring statistical dependence with Hilbert–Schmidt norms, in: ALT, vol. 16, Springer, New York, 2005, pp. 63–78.
[14] F. Han, S. Chen, H. Liu, Distribution-free tests of independence in high dimensions, Biometrika 104 (2017) 813–828.
[15] W. Hoeffding, A non-parametric test of independence, Ann. Math. Stat. 19 (1948) 546–557.
[16] Z. Jin, D.S. Matteson, Independent component analysis via energy-based and kernel-based mutual dependence measures, 2018. arXiv preprint arXiv:1805.06639.
[17] Z. Jin, X. Yan, D.S. Matteson, Testing for conditional mean independence with covariates through martingale difference divergence, in: Proceedings of the 34th Conference on Uncertainty in Artificial Intelligence, AUAI, 2018.
[18] Z. Jin, S. Yao, D.S. Matteson, X. Shao, EDMeasure: Energy-Based Dependence Measures, 2018. R package version 1.2.
[19] M.G. Kendall, A new measure of rank correlation, Biometrika 30 (1938) 81–93.
[20] I. Kojadinovic, M. Holmes, Tests of independence among continuous random vectors based on Cramér–von Mises functionals of the empirical copula process, J. Multivariate Anal. 100 (2009) 1137–1154.
[21] D. Leung, M. Drton, Testing independence in high dimensions with sums of rank correlations, Ann. Statist. 46 (2018) 280–307.
[22] Z. Li, T.H. McCormick, An expectation conditional maximization approach for Gaussian graphical models, 2017. arXiv preprint arXiv:1709.06970.
[23] Z.R. Li, T.H. McCormick, S.J. Clark, Bayesian joint spike-and-slab graphical lasso, 2018. arXiv preprint arXiv:1805.07051.
[24] J. Lu, On finite-population Bayesian inferences for 2^K factorial designs with binary outcomes, 2018. arXiv preprint arXiv:1803.04499,.
[25] J. Lu, P. Ding, T. Dasgupta, Treatment effects on ordinal outcomes: Causal estimands and sharp bounds, 2015. arXiv preprint arXiv:1507.01542.
[26] D.S. Matteson, R.S. Tsay, Independent component analysis via distance covariance, J. Amer. Statist. Assoc. 112 (2017) 623–637.
[27] R. von Mises, On the asymptotic distribution of differentiable statistical functions, Ann. Math. Stat. 18 (1947) 309–348.
[28] K. Pearson, Note on regression and inheritance in the case of two parents, Proc. R. Soc. London 58 (1895) 240–242.
[29] N. Pfister, J. Peters, dHSIC: Independence Testing via Hilbert–Schmidt Independence Criterion, 2017, R package version 2.0.
[30] N. Pfister, P. Bühlmann, B. Schölkopf, J. Peters, Kernel-based tests for joint independence, J. R. Stat. Soc. Ser. B Stat. Methodol. 80 (2018) 5–31.
[31] X. Shao, J. Zhang, Martingale difference correlation and its use in high-dimensional variable screening, J. Amer. Statist. Assoc. 109 (2014) 1302–1318.
[32] C.E. Spearman, The proof and measurement of association between two things, Amer. J. Psych. 15 (1904) 72–101.
[33] G.J. Székely, M.L. Rizzo, A new test for multivariate normality, J. Multivariate Anal. 93 (2005) 58–80.
[34] G.J. Székely, M.L. Rizzo, Brownian distance covariance, Ann. Appl. Stat. 3 (2009) 1236–1265.
[35] G.J. Székely, M.L. Rizzo, The distance correlation t-test of independence in high dimension, J. Multivariate Anal. 117 (2013) 193–213.
[36] G.J. Székely, M.L. Rizzo, Energy statistics: A class of statistics based on distances, J. Statist. Plann. Inference 143 (2013) 1249–1272.
[37] G.J. Székely, M.L. Rizzo, N.K. Bakirov, Measuring and testing dependence by correlation of distances, Ann. Statist. 35 (2007) 2769–2794.

[38] R.J. Tibshirani, T. Hastie, B. Narasimhan, G. Chu, Diagnosis of multiple cancer types by shrunken centroids of gene expression, Proc. Nat. Acad. Sci. USA 99 (2002) 6567–6572.

[39] S. Yao, X. Zhang, X. Shao, Testing mutual independence in high dimension via distance covariance, J. R. Stat. Soc. Ser. B Stat. Methodol. 80 (2018) 455–480.

[40] B. Zhang, M. Dundar, M. Al Hasan, Bayesian non-exhaustive classification a case study: Online name disambiguation using temporal record streams, in: Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, ACM, 2016, pp. 1341–1350.