The Annals of Statistics
2019, Vol. 47, No. 5, 2757–2789
https://doi.org/10.1214/18-AOS1764

DISTANCE MULTIVARIANCE: NEW DEPENDENCE MEASURES FOR RANDOM VECTORS

BY BJÖRN BÖTTCHER, MARTIN KELLER-RESSEL[1] AND RENÉ L. SCHILLING

TU Dresden

We introduce two new measures for the dependence of $n \geq 2$ random variables: *distance multivariance* and *total distance multivariance*. Both measures are based on the weighted L^2-distance of quantities related to the characteristic functions of the underlying random variables. These extend distance covariance (introduced by Székely, Rizzo and Bakirov) from pairs of random variables to n-tuplets of random variables. We show that total distance multivariance can be used to detect the independence of n random variables and has a simple finite-sample representation in terms of distance matrices of the sample points, where distance is measured by a continuous negative definite function. Under some mild moment conditions, this leads to a test for independence of multiple random vectors which is consistent against all alternatives.

1. Introduction and related work. Distance multivariance $M_\rho(X_1, X_2, \dots, X_n)$ and total distance multivariance $\overline{M}_\rho(X_1, X_2, \dots, X_n)$ are new measures for the dependence of random variables $X_1, \dots, X_n$. They are closely related to distance covariance, as introduced by Székely, Rizzo and Bakirov [23, 25] and its generalizations presented in [8]. Distance multivariance inherits many of the features of distance covariance; in particular, see Theorem 3.4 below:

- $M_\rho(X_1, \dots, X_n)$ and $\overline{M}_\rho(X_1, \dots, X_n)$ are defined for random variables $X_1, \dots, X_n$ with values in spaces of arbitrary dimensions $\mathbb{R}^{d_1}, \dots, \mathbb{R}^{d_n}$;
- if each subfamily of $X_1, \dots, X_n$ with $n-1$ elements is independent, then $M_\rho(X_1, \dots, X_n) = 0$ characterizes the independence of $X_1, \dots, X_n$;
- $\overline{M}_\rho(X_1, \dots, X_n) = 0$ characterizes the independence of $X_1, \dots, X_n$.

We emphasize that measuring the dependence of n random variables is different from measuring their pairwise dependence, and for this reason bivariate dependence measures, such as distance covariance, cannot be used directly to detect overall independence. A classical example, Bernstein's coins, is discussed in Section 5. The extension of distance covariance to more than two random variables was addressed in a short paragraph in Bakirov and Székely [1]. Our approach is

Received December 2017; revised September 2018.
[1] Supported by German Research Foundation (DFG) Grant ZUK 64 and KE 1736/1-1.
MSC2010 subject classifications. Primary 62H20; secondary 60E10, 62G10, 62G15, 62G20.
Key words and phrases. Dependence measure, stochastic independence, negative definite function, characteristic function, Gaussian random field, statistical test of independence.

different from the approach suggested in [1]; it is, in fact, closer to the two approaches that were advised against in [1]. We will discuss and compare these approaches in greater detail in Section 3.4, once the necessary concepts have been introduced. Recently, Yao *et al.* [26] introduced measures for pairwise dependence based on distance covariance. In contrast, distance multivariance does not only detect pairwise dependence, but any type of multivariate dependence. Jin and Matteson [17] present measures for multivariate independence which also use distance covariance. The resulting exact estimators are computationally more complex than those of distance multivariance; [6] shows that the approximate estimators of [17] have less empirical power but are computationally of the same order as distance multivariance.

Another line of research considers dependence measures based on reproducing kernel Hilbert spaces, notably the Hilbert–Schmidt independence criterion (HSIC) of [15], which has been shown to be equivalent to distance covariance in [21]. Subsequently, HSIC has been extended from a bivariate dependence measure to a multivariate dependence measure, dHSIC, in [18]. We compare dHSIC to distance multivariance in Section 3.5.

Similar to distance covariance in [25] and its generalizations given in [8], distance multivariance can be defined as a weighted L^2-norm of quantities related to the characteristic functions of $X_1, \dots, X_n$; cf. Definition 2.2 below. There are, however, further definitions of distance multivariance which are equivalent up to moment conditions. In particular, multivariance can be equivalently defined as *Gaussian multivariance* by evaluating a Gaussian random field at the instances $(X_1, \dots, X_n)$ and taking certain expectations; see Section 3.3. This generalizes Székely-and-Rizzo's [23], Definition 4, Brownian covariance which is recovered using $n = 2$ and multiparameter Brownian motion as random field.

The sample versions of both distance multivariance and total distance multivariance have simple expressions in terms of the distance matrices of the sample points; this means that we can compute these statistics efficiently even for large samples and in high dimensions. In concrete terms, as we show in Theorem 4.1, the square of the distance multivariance computed from samples $\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)}$ of the random vector $\boldsymbol{X} = (X_1, \dots, X_n)$ can be written as

$$^{N}M_\rho^2\big(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)}\big) = \frac{1}{N^2} \sum_{j,k=1}^{N} (A_1)_{jk} \cdot \dots \cdot (A_n)_{jk},$$

where the A_i are doubly centred distance matrices of the sample points of X_i, that is, $A_i := -C B_i C$ where C is the centering matrix $C = I - \frac{1}{N}\mathbb{1}$, $\mathbb{1} = (1)_{j,k=1,\dots,N}$, $I = (\delta_{jk})_{j,k=1,\dots,N}$, and B_i are the distance matrices of the sample points. The square of the sample *total* distance multivariance has a similar form

$$^{N}\overline{M}_\rho^2\big(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)}\big) = \frac{1}{N^2} \sum_{j,k=1}^{N} \big(1 + (A_1)_{jk}\big) \cdot \dots \cdot \big(1 + (A_n)_{jk}\big) - 1.$$

The (quasi-)distance that is used to compute B_i can be chosen, under mild restrictions, from the class of real-valued continuous negative definite functions; cf. [3], Chapter II, [16], Section 3.2. In particular, we may use Euclidean and p-Minkowski distances with exponent $p \in (1,2]$. In the bivariate case, and using Euclidean distance, the sample distance covariance of Székely and Rizzo [23], Definition 3, is recovered.

Finally, we show in Theorems 4.5 and 4.10 asymptotic properties of sample distance multivariance as N tends to infinity; these results are multivariate analogues of those in [23], Theorem 5. Based on these results, we formulate two new distribution-free tests for the joint independence of n random variables in Section 4.5. These tests are conservative, and a resampling approach can be used to construct tests achieving the nominal size; further results in this direction can be found in [6]. The paper concludes in Section 5 with an extended example based on *Bernstein's coins*, which demonstrates numerically that (total) distance multivariance is able to distinguish between pairwise independence and higher-order dependence of random variables. The example also illustrates the practical validity of the two tests that are proposed. A further example with sinusoidal dependence is discussed, illustrating the influence of the underlying distance on the dependence measure.

For the immediate use of distance multivariance in applications all necessary functions are provided in the R package `multivariance`, [7].

2. Preliminaries.

We consider a d-dimensional random vector $\boldsymbol{X} = (X_1, \dots, X_n)$, whose components X_i are random variables taking values in $\mathbb{R}^{d_i}$, $i = 1, \dots, n$, and where $d = d_1 + \cdots + d_n$. The characteristic function of X_i is denoted by

$$f_{X_i}(t_i) := \mathbb{E} e^{\mathrm{i} X_i \cdot t_i}, \qquad t_i \in \mathbb{R}^{d_i},$$

and we write $\boldsymbol{t} = (t_1, \dots, t_n)$. In order to define the distance multivariance of $(X_1, \dots, X_n)$, we use *Lévy measures* ρ_i, that is, Borel measures ρ_i defined on $\mathbb{R}^{d_i} \setminus \{0\}$ such that

$$\int_{\mathbb{R}^{d_i} \setminus \{0\}} \min\{|t_i|^2, 1\} \rho_i(\mathrm{d}t_i) < \infty. \tag{2.1}$$

Note that the measures ρ_i need not be finite. Such measures appear in the Lévy–Khintchine representation of infinitely divisible distributions; see [19]. Throughout this paper, we assume that ρ_i, $i = 1, \dots, n$ are symmetric Lévy measures with full topological support (cf. [8], Definition 2.3), and we set $\rho := \rho_{d_1} \otimes \cdots \otimes \rho_{d_n}$. To keep notation simple, we write $\int \dots \rho_i(\mathrm{d}t_i)$ and $\int_{\mathbb{R}^{d_i}} \dots \rho_i(\mathrm{d}t_i)$ instead of the formally correct $\int_{\mathbb{R}^{d_i} \setminus \{0\}} \dots \rho_i(\mathrm{d}t_i)$.

DEFINITION 2.1. Let $(X_i)_{i=1,\dots,n}$ be random variables with values in $\mathbb{R}^{d_i}$ and let the measures ρ_i be given as above. With $\rho := \rho_1 \otimes \cdots \otimes \rho_n$, we define

(a) *Distance multivariance* $M_\rho \in [0,\infty]$ by

$$M_\rho^2(X_1,\dots,X_n) := \int_{\mathbb{R}^d} \left| \mathbb{E}\left(\prod_{i=1}^n \left(e^{iX_i\cdot t_i} - f_{X_i}(t_i)\right)\right)\right|^2 \rho(dt_1,\dots,dt_n), \tag{2.2}$$

(b) *Total distance multivariance* $\overline{M}_\rho \in [0,\infty]$ by

$$\overline{M}_\rho^2(X_1,\dots,X_n) := \sum_{\substack{1\le i_1<\dots<i_m\le n\\ 2\le m\le n}} M^2_{\bigotimes_{j=1}^m \rho_{i_j}}(X_{i_1},\dots,X_{i_m}). \tag{2.3}$$

REMARK 2.2. (a) Using the tensor product for functions

$$(g_1\otimes\cdots\otimes g_n)(x_1,\dots,x_n) = g_1(x_1)\cdot\ldots\cdot g_n(x_n),$$

distance multivariance can be written in a compact way as

$$M_\rho(X_1,\dots,X_n) = \left\| \mathbb{E}\left[\bigotimes_{i=1}^n \left(e^{iX_i\cdot\bullet} - f_{X_i}(\bullet)\right)\right]\right\|_{L^2(\rho)}. \tag{2.4}$$

Thus, distance multivariance is the weighted L^2-norm of a quantity related to the characteristic functions of the X_i, analogous to the definition of distance covariance in Székely, Rizzo and Bakirov [25], Definition 1.

(b) Both expressions M_ρ and $\overline{M}_\rho$ are always well defined in $[0,+\infty]$: For each $\boldsymbol{t} = (t_1,\dots,t_n)$, the product appearing in the integrand of (2.2) can be bounded in absolute value by 2^n; therefore, the expectation exists. The integrand of the ρ-integral is positive, and so the integral is always well defined in $[0,+\infty]$. Just as in the bivariate case (see [8], Theorem 3.7, Remark 3.8), we need moment conditions on the random variables X_i to guarantee finiteness of M_ρ and $\overline{M}_\rho$; see Proposition 3.9 below.

(c) At first sight, total distance multivariance seems to suffer from a computational curse of dimension, since the sum (2.3) extends over all subfamilies (comprising at least two members) of $(X_1,\dots,X_n)$, that is, 2^n-1-n terms are summed. We will, however, show in Theorem 4.1, that the finite sample version of $\overline{M}_\rho$ has the same computational complexity as M_ρ and its computation requires only $\mathcal{O}(nN^2)$ operations given a sample of size N.

Each Lévy measure ρ_i uniquely defines a real-valued *continuous negative definite function*

$$\psi_i(y_i) := \int_{\mathbb{R}^{d_i}} \left(1-\cos(yt_i)\right)\rho_i(dt_i) \qquad \text{for } y_i\in\mathbb{R}^{d_i}; \tag{2.5}$$

see, for example, [16], Corollary 3.7.9. The functions ψ_i will play a key role in the finite-sample representation of distance multivariance and also appear in moment conditions. They are also the reason for the terms *distance* multivariance (and *distance* covariance, cf. [23]), since ψ_i yields well-known distance functions (and in many cases norms) in several important special cases. In particular, $x \mapsto |x|^\alpha$ where $|\cdot|$ is the standard d_i-dimensional Euclidean norm and $\alpha \in (0,2)$, can be represented using

$$\rho_i(\mathrm{d}t_i) = c_{\alpha,d_i}|t_i|^{-d_i-\alpha}\mathrm{d}t_i, \qquad \alpha \in (0,2), c_{\alpha,d_i} = \frac{\alpha 2^{\alpha-1}\Gamma(\frac{\alpha+d_i}{2})}{\pi^{d_i/2}\Gamma(1-\frac{\alpha}{2})},$$

since

$$|y_i|^\alpha = c_{d_i,\alpha}\int_{\mathbb{R}^{d_i}}(1-\cos y\cdot t_i)\frac{\mathrm{d}t_i}{|t_i|^{d_i+\alpha}}.$$

Also other Minkowski distances $|x|_{d_i,p} := (\sum_{j=1}^{d_i}|x_j|^p)^{1/p}$, for $p \in (1,2]$ can be written in the form (2.5); see [8], Lemma 2.2 and Table 1, for this and further examples.

For the following results and proofs, it will be useful to introduce some notation for various distributional copies of the vector $\boldsymbol{X} = (X_1,\dots,X_n)$. Recall that $\mathcal{L}(X_i)$ denotes the law of X_i and define the random vectors

$$\begin{aligned} \boldsymbol{X}_0 &= (X_{0,1},\dots,X_{0,n}) &&\sim \mathcal{L}(X_1)\otimes\cdots\otimes\mathcal{L}(X_n),\\ \boldsymbol{X}'_0 &= (X'_{0,1},\dots,X'_{0,n}) &&\sim \mathcal{L}(X_1)\otimes\cdots\otimes\mathcal{L}(X_n),\\ \boldsymbol{X}_1 &= (X_{1,1},\dots,X_{1,n}) &&\sim \mathcal{L}(X_1,\dots,X_n),\\ \boldsymbol{X}'_1 &= (X'_{1,1},\dots,X'_{1,n}) &&\sim \mathcal{L}(X_1,\dots,X_n), \end{aligned} \tag{2.6}$$

such that the random vectors $\boldsymbol{X}_0$, $\boldsymbol{X}'_0$, $\boldsymbol{X}_1$, $\boldsymbol{X}'_1$ are independent. Note that the subscript “1”—as in $\boldsymbol{X}_1$ and $\boldsymbol{X}'_1$—indicates that these vectors have the *same distribution* as $\boldsymbol{X}$, while the subscript “0”—as in $\boldsymbol{X}_0$ and $\boldsymbol{X}'_0$—means that these random vectors have the *same marginal distributions* as $\boldsymbol{X}$, but their coordinates are independent.

DEFINITION 2.3. We introduce the following moment conditions:
(a) The *mixed moment condition* holds if

$$\mathbb{E}\left(\prod_{i=1}^n \psi_i(X_{k_i,i}-X'_{l_i,i})\right) < \infty \qquad \text{for all } k_i, l_i \in \{0,1\}, i = 1,\dots,n.$$

(b) The *psi-moment condition* holds if there exist $p_i \in [1,\infty)$ satisfying $\sum_{i=1}^n p_i^{-1} = 1$ such that

$$\mathbb{E}\psi_i^{p_i}(X_i) < \infty \qquad \text{for all } i = 1,\dots,n.$$

In particular, one may choose $p_1 = \cdots = p_n = n$. (The case $p_i = \infty$ is also admissible, but this means that ψ_i must be bounded or X_i must have compact support.)

(c) The $2p$-*moment condition* holds if there exist $p_i \in [1, \infty)$ satisfying $\sum_{i=1}^n p_i^{-1} = 1$ such that

$$\mathbb{E}\big[|X_i|^{2p_i}\big] < \infty \qquad \text{for all } i = 1, \dots, n$$

(the case $p_i = \infty$ is also admissible, but this means that X_i is a.s. bounded).

As shown in Lemma S.1 in the Supplementary Material [9], these moment conditions are ordered from weak to strong, that is, (c) implies (b) and (b) implies (a). Also note that (b) and (a) trivially hold (for any choice of p_i) if the functions ψ_i are bounded.

3. Distance multivariance and total distance multivariance.

3.1. *Total distance multivariance characterizes independence.* We need the concept of m-independence of $n \geq m$ random variables.

DEFINITION 3.1. Random variables $X_1, \dots, X_n$ are *m-independent* (for some $m \leq n$) if for any sub-family $\{i_1, \dots, i_m\} \subset \{1, \dots, n\}$ the random variables $X_{i_1}, \dots, X_{i_m}$ are independent.

The condition of $(n-1)$-independence allows certain factorizations of expectations of products; the proof of the following lemma is given in the Supplementary Material [9].

LEMMA 3.2. *Let $Z_1, \dots, Z_n$ be $\mathbb{C}$-valued random variables which are $(n-1)$-independent. Then*

$$\mathbb{E}\left(\prod_{i=1}^n (Z_i - \mathbb{E}Z_i)\right) = \mathbb{E}\left(\prod_{i=1}^n Z_i - \prod_{i=1}^n \mathbb{E}Z_i\right). \tag{3.1}$$

If we use the random variables $Z_i := e^{iX_i \cdot t_i}$, Lemma 3.2 yields the following result for characteristic functions.

COROLLARY 3.3. *Let $X_1, \dots, X_m$ be $(m-1)$-independent random variables, then*

$$\begin{aligned} &\mathbb{E}\left[\prod_{k=1}^m \left(e^{iX_{i_k} \cdot t_{i_k}} - f_{X_{i_k}}(t_{i_k})\right)\right] \\ &\quad = f_{(X_{i_1}, \dots, X_{i_m})}(t_{i_1}, \dots, t_{i_m}) - f_{X_{i_1}}(t_{i_1}) \cdot \ldots \cdot f_{X_{i_m}}(t_{i_m}). \end{aligned} \tag{3.2}$$

This enables us to show that independence is indeed characterized by total distance multivariance.

THEOREM 3.4. (a) *Distance multivariance vanishes for independent random variables, that is,*

$$X_1, \dots, X_n \text{ are independent} \quad \Longrightarrow \quad M_\rho(X_1, \dots, X_n) = 0. \tag{3.3}$$

If $X_1, \dots, X_n$ are $(n-1)$-independent, then also the converse holds.

(b) *Total distance multivariance characterizes independence, that is,*

$$X_1, \dots, X_n \text{ are independent} \quad \Longleftrightarrow \quad \overline{M}_\rho(X_1, \dots, X_n) = 0. \tag{3.4}$$

REMARK 3.5. Note that multivariance is not just a building block of total multivariance, but has applications in its own right. The characterization of n-independence by $(n-1)$-independence and $M_\rho(X_1, \dots, X_n) = 0$ can be used to detect (higher order) dependence structures; this is used in [6]. Other applications can be found in the setting of independent component analysis (ICA). The algorithm of [11]) aims to transform the input signal into pairwise independent random variables which, if all assumptions of ICA are satisfied, are also mutually independent. Thus, distance multivariance can be used to test the validity of assumptions by testing for higher order dependence, given pairwise independence [2].

PROOF OF THEOREM 3.4. Suppose that $X_1, \dots, X_n$ are independent. We have for all indices $\{i_1, \dots, i_m\} \subset \{1, \dots, n\}$

$$\mathbb{E}\left[\prod_{k=1}^m \left(e^{iX_{i_k} \cdot t_{i_k}} - f_{X_{i_k}}(t_{i_k})\right)\right] = \prod_{k=1}^m \mathbb{E}\left(e^{iX_{i_k} \cdot t_{i_k}} - f_{X_{i_k}}(t_{i_k})\right) = 0, \tag{3.5}$$

and, so, $M_{\bigotimes_{k=1}^m \rho_{i_k}}(X_{i_1}, \dots, X_{i_m}) = 0$; this implies $\overline{M}_\rho(X_1, \dots, X_n) = 0$.

For the converse statements, suppose first that $X_1, \dots, X_n$ are $(n-1)$-independent and consider

$$\kappa(t_1, \dots, t_n) := \mathbb{E}\left[\prod_{i=1}^n \left(e^{iX_i \cdot t_i} - f_{X_i}(t_i)\right)\right].$$

By definition, $M_\rho(X_1, \dots, X_n)$ is the $L^2(\rho)$-norm of κ. Since ρ has full topological support and κ is continuous, $M_\rho = 0$ implies that $\kappa \equiv 0$ everywhere on $\mathbb{R}^d$. By Corollary 3.3, it follows that

$$f_{(X_1, \dots, X_n)}(t_1, \dots, t_n) = f_{X_1}(t_1) \cdot \dots \cdot f_{X_n}(t_n) \qquad \text{for all } t_1, \dots, t_n,$$

that is, the joint characteristic function of $X_1, \dots, X_n$ factorizes, and we conclude that $X_1, \dots, X_n$ are independent.

Finally, suppose that $\overline{M}_\rho(X_1, \dots, X_n) = 0$, and thus that

$$M_{\bigotimes_{k=1}^m \rho_{i_k}}(X_{i_1}, \dots, X_{i_m}) = 0 \qquad \text{for any } \{i_1, \dots, i_m\} \subset \{1, \dots, n\}. \tag{3.6}$$

Starting with subsets of size 2, we note that

$$\begin{aligned}\overline{M}_{\rho_{i_1}\otimes\rho_{i_2}}(X_{i_1}, X_{i_2}) &= M_{\rho_{i_1}\otimes\rho_{i_2}}(X_{i_1}, X_{i_2}) \\ &= \| f_{(X_{i_1}, X_{i_2})} - f_{X_{i_1}} f_{X_{i_2}} \|_{L^2(\rho_{i_1}\otimes\rho_{i_2})} = 0 \end{aligned} \tag{3.7}$$

for all $\{i_1, i_2\} \subset \{1, \dots, n\}$; this means that the random variables $X_1, \dots, X_n$ are pairwise independent, hence $X_1, \dots, X_n$ are 2-independent. Continuing with subsets of size 3, (3.6) together with the first part of the proof implies 3-independence of $X_1, \dots, X_n$. Repeating this argument finally yields the independence of $X_1, \dots, X_n$. □

3.2. *Further properties and representations of multivariance.* Directly from Definition 2.2, we see that for two random variables $X = X_1$ and $Y = X_2$ and Lévy measures $\rho = \rho_1 \otimes \rho_2$ the notions of multivariance M_ρ, total multivariance $\overline{M}_\rho$ and generalized distance covariance V as defined in [8], Definition 3.1, coincide, that is,

$$M_\rho(X, Y) = \overline{M}_\rho(X, Y) = V(X, Y).$$

The following properties are straightforward.

PROPOSITION 3.6. *Distance multivariance enjoys the following properties*:

$$M_{\rho_i}(X_i) = 0 \qquad \textit{for all } i = 1, \dots, n, \tag{3.8}$$

$$M_\rho(X_1, \dots, X_n) = M_\rho(c_1 X_1, \dots, c_n X_n) \qquad \textit{for } c_i \in \{-1, +1\}. \tag{3.9}$$

Let $S \subset \{1, \dots, n\}$. *If* $(X_i, i \in S)$ *is independent of* $(X_i, i \in S^c)$, *then*

$$M_\rho(X_1, \dots, X_n) = M_{\bigotimes_{i\in S}\rho_i}(X_i, i \in S) \cdot M_{\bigotimes_{i\in S^c}\rho_i}(X_i, i \in S^c). \tag{3.10}$$

PROOF. If $n = 1$, the expectation in (2.2) becomes $\mathbb{E}(e^{\mathrm{i}X_i t_i} - \mathbb{E}\mathrm{e}^{\mathrm{i}X_i t_i}) = 0$ and (3.8) follows. Property (3.9) follows from the symmetry of the measures ρ_i. For the last property, note that the assumption of independence allows us to factorize the following expression:

$$\begin{aligned}&\mathbb{E}\Bigg[\bigotimes_{i=1}^n (\mathrm{e}^{\mathrm{i}X_i\cdot\bullet} - f_{X_i}(\bullet))\Bigg] \\ &\quad = \mathbb{E}\Bigg[\bigotimes_{i\in S} (\mathrm{e}^{\mathrm{i}X_i\cdot\bullet} - f_{X_i}(\bullet))\Bigg] \cdot \mathbb{E}\Bigg[\bigotimes_{i\in S^c} (\mathrm{e}^{\mathrm{i}X_i\cdot\bullet} - f_{X_i}(\bullet))\Bigg].\end{aligned}$$

Since also ρ can be factorized into $\bigotimes_{i\in S}\rho_i$ and $\bigotimes_{i\in S^c}\rho_i$, (3.10) follows. □

Another relevant aspect is the behavior of (total) distance multivariance, when an independent component is added to a given random vector.

PROPOSITION 3.7. *Let X_{n+1} be independent from $(X_1, \dots, X_n)$. Then*

$$M_\rho(X_1, \dots, X_{n+1}) = 0, \tag{3.11}$$

$$\overline{M}_\rho(X_1, \dots, X_{n+1}) = \overline{M}_\rho(X_1, \dots, X_n). \tag{3.12}$$

PROOF. The first equation follows from (3.10) by taking $S = \{1, \dots, n\}$. If we insert this into (2.3), we see that all summands containing the index $i = n + 1$ do not contribute to total distance multivariance. Hence, (3.12) follows. □

REMARK 3.8. In this context, it is interesting to anticipate *normalized* total distance multivariance $\overline{\mathcal{M}}_\rho$ which will be defined in (4.28). If X_{n+1} is independent from $(X_1, \dots, X_n)$, it is easy to check that

$$\overline{\mathcal{M}}_\rho(X_1, \dots, X_{n+1}) = r(n) \cdot \overline{\mathcal{M}}_\rho(X_1, \dots, X_n),$$

where $r(n) = \sqrt{(2^n - n - 1)}/\sqrt{(2^{n+1} - n - 2)}$. Note that $r(n)$ is strictly increasing from $r(2) = 1/2$ to $\lim_{n\to\infty} r(n) = 1/\sqrt{2}$. Thus, the addition of an independent component affects $\overline{\mathcal{M}}_\rho$ by a factor from $[1/2, 1/\sqrt{2})$.

We now turn to different representations of multivariance. The representation as $L^2(\rho)$-norm in (2.2) is always well defined, but may have infinite value. Under suitable moment conditions, multivariance is finite and can be represented in terms of the continuous negative definite functions ψ_i given in (2.5). The proof of the following proposition can be found in the Supplementary Material [9].

PROPOSITION 3.9. *Multivariance $M_\rho = M_\rho^2(X_1, \dots, X_n)$ can be written as*

$$M_\rho^2 = \int \mathbb{E}\left(\sum_{k,l \in \{0,1\}^n} \operatorname{sgn}(k, l) \prod_{i=1}^n \mathrm{e}^{\mathrm{i}(X_{k_i,i} - X'_{l_i,i}) \cdot t_i} \right) \rho(\mathrm{d}t), \tag{3.13}$$

or

$$M_\rho^2 = \int \mathbb{E}\left(\sum_{k,l \in \{0,1\}^n} \operatorname{sgn}(k, l) \prod_{i=1}^n \left[\cos\left((X_{k_i,i} - X'_{l_i,i}) \cdot t_i\right) - 1\right] \right) \rho(\mathrm{d}t), \tag{3.14}$$

where

$$\operatorname{sgn}(k, l) := (-1)^{\sum_{j=1}^n (k_j + l_j)} = \begin{cases} +1 & \text{if } (k, l) \text{ contains an even no. of "1"s}, \\ -1 & \text{if } (k, l) \text{ contains an odd no. of "1"s}. \end{cases}$$

If one of the moment conditions in Definition 2.3 holds, then the distance multivariance $M_\rho(X_1, \dots, X_n)$ is finite, and the following representation holds:

$$\begin{aligned} M_\rho^2 = \mathbb{E}\Bigg(\prod_{i=1}^n \big[& -\psi_i(X_i - X_i') + \mathbb{E}(\psi_i(X_i - X_i') \mid X_i) \\ & + \mathbb{E}(\psi_i(X_i - X_i') \mid X_i') - \mathbb{E}\psi_i(X_i - X_i')\big] \Bigg). \end{aligned} \tag{3.15}$$

REMARK 3.10. (a) The representations (3.13) and (3.14) have an interesting structural resemblance to the Leibniz' formula for determinants; (3.15) is the analogue of [8], Corollary 3.5, for the bivariate case.

(b) In the bivariate case $n = 2$, distance multivariance is also finite under the weaker moment condition $\mathbb{E}\psi_1(X_1) + \mathbb{E}\psi_2(X_2) < \infty$; cf. [8], Theorem 3.7.

We introduce yet another representation of distance multivariance, which helps to clarify the relation to the finite-sample form and the representation as *Gaussian multivariance*, given in Section 3.3 below. For this, we need the centering operator $\mathrm{C}_{\mathcal{F}}$.

PROPOSITION 3.11. *Let X be an integrable random variable on $(\Omega, \mathcal{A}, \mathbb{P})$ and $\mathcal{F}, \mathcal{F}'$ be sub-σ-algebras of $\mathcal{A}$. Set*

$$\mathrm{C}_{\mathcal{F}} X := X - \mathbb{E}(X \mid \mathcal{F}). \tag{3.16}$$

Then C *is a linear operator and*

$$\mathrm{C}_{\{\varnothing,\Omega\}} X = X - \mathbb{E}X, \tag{3.17}$$

$$\mathrm{C}_{\mathcal{F}} \mathrm{C}_{\mathcal{F}'} X = X - \mathbb{E}(X \mid \mathcal{F}') - \mathbb{E}(X \mid \mathcal{F}) + \mathbb{E}(\mathbb{E}(X \mid \mathcal{F}') \mid \mathcal{F}), \tag{3.18}$$

$$\mathrm{C}_{\mathcal{F}} \mathrm{C}_{\mathcal{F}'} X = 0 \qquad \text{if } X \text{ is } \mathcal{F}'\text{-measurable}. \tag{3.19}$$

If $\mathcal{F}'$ and $\mathcal{F}$ are independent, then $\mathbb{E}(\mathrm{C}_{\mathcal{F}'} X \mid \mathcal{F}) = \mathrm{C}_{\{\varnothing,\Omega\}} \mathbb{E}(X \mid \mathcal{F})$.

All assertions of the proposition follow directly from the properties of conditional expectations, and we omit the proof. Geometrically, $\mathrm{C}_{\mathcal{F}} X$ can be interpreted as the residual from the orthogonal projection of X onto the set of $\mathcal{F}$-measurable functions. We will use the shorthand $\mathrm{C}_X := \mathrm{C}_{\sigma(X)}$.

COROLLARY 3.12. *If one of the moment conditions in Definition 2.3 holds, then*

$$M_\rho^2(X_1, \dots, X_n) = \mathbb{E}\left(\prod_{i=1}^n -\mathrm{C}_{X_i} \mathrm{C}_{X_i'} \psi_i(X_i - X_i')\right) \tag{3.20}$$

and

$$\overline{M}_\rho^2(X_1, \dots, X_n) = \mathbb{E}\left(\prod_{i=1}^n (1 - \mathrm{C}_{X_i} \mathrm{C}_{X_i'} \psi_i(X_i - X_i'))\right) - 1. \tag{3.21}$$

The factors can be written explicitly as

$$\begin{aligned} \mathrm{C}_{X_i} \mathrm{C}_{X_i'} \psi_i(X_i - X_i') = {} & \psi_i(X_i - X_i') - \mathbb{E}[\psi_i(X_i - X_i') \mid X_i'] \\ & - \mathbb{E}[\psi_i(X_i - X_i') \mid X_i] + \mathbb{E}\psi_i(X_i - X_i'). \end{aligned} \tag{3.22}$$

PROOF. The identity (3.22) follows directly from the definition of the double centering operator in Proposition 3.11. The representation (3.20) is an immediate consequence of (3.15) in Proposition 3.9. For representation (3.21) of the total multivariance, write $a_i := -\mathrm{C}_{X_i}\mathrm{C}_{X_i'}\psi_i(X_i - X_i')$. We can expand the product

$$\prod_{i=1}^{n}(1+a_i) = \sum_{m=0}^{n} e_m(a_1,\dots,a_n),$$

where the function $e_m(a_1,\dots,a_n)$ is the mth elementary symmetric polynomial in $(a_1,\dots,a_n)$, that is,

$$e_m(a_1,\dots,a_n) = \sum_{1\le i_1<\dots<i_m\le n} a_{i_1}\cdot\dots\cdot a_{i_m}.$$

In particular, $e_0(a_1,\dots,a_n)=1$ and $e_1(a_1,\dots,a_n)=a_1+\dots+a_n$. Taking expectations yields

$$\begin{aligned}\mathbb{E}\left[\prod_{i=1}^{n}(1+a_i)\right]-1 &= \sum_{m=1}^{n}\mathbb{E}\big[e_m(a_1,\dots,a_n)\big]-1\\ &= \sum_{\substack{1\le i_1<\dots<i_m\le n\\ 2\le m\le n}}\mathbb{E}[a_{i_1}\cdot\dots\cdot a_{i_m}]\\ &= \sum_{\substack{1\le i_1<\dots<i_m\le n\\ 2\le m\le n}} M_\rho^2(X_{i_1},\dots,X_{i_m})\\ &= \overline{M}_\rho^2(X_1,\dots,X_n), \end{aligned} \tag{3.23}$$

as claimed. Note that the first elementary symmetric polynomial e_1 does not contribute since $\mathbb{E}[a_i]=0$ for all $i\in\{1,\dots,n\}$. □

3.3. *Gaussian multivariance.* Recall that for a real-valued negative definite function $\psi:\mathbb{R}^d\to\mathbb{R}$ the matrix $(\psi(\xi_j)+\psi(\xi_k)-\psi(\xi_j-\xi_k))_{j,k=1,\dots,n}$, $\xi_1,\dots,\xi_n\in\mathbb{R}^d$, $n\in\mathbb{N}$, is positive semidefinite; see [16], Definition 3.6.6. Therefore, we can associate with any cndf ψ some Gaussian random field indexed by $\mathbb{R}^d$.

DEFINITION 3.13. Assume that $X_1,\dots,X_n$ satisfy one of the moment conditions in Definition 2.3 and let $G_1,\dots,G_n$ be independent (also independent of $X_1,\dots,X_n$), stationary Gaussian random fields with

$$\mathbb{E}G_i(\xi)=0 \quad\text{and}\quad \mathbb{E}\big(G_i(\xi)G_i(\eta)\big)=\psi_i(\xi)+\psi_i(\eta)-\psi_i(\xi-\eta) \tag{3.24}$$

for $\xi,\eta\in\mathbb{R}^{d_i}$. The *Gaussian multivariance* of $(X_1,\dots,X_n)$ is defined by

$$\mathcal{G}^2(X_1,\dots,X_n)=\mathbb{E}\left(\prod_{i=1}^{n}X_i^{G_i}X_i'^{G_i}\right), \tag{3.25}$$

where $(X'_1, \dots, X'_n)$ is an independent copy of $(X_1, \dots, X_n)$ and

$$X_i^{G_i} := G_i(X_i) - \mathbb{E}\big(G_i(X_i) \mid G_i\big). \tag{3.26}$$

REMARK 3.14. (a) Using the centering operator C from Proposition 3.11, we can write (3.26) as $X_i^{G_i} = \mathrm{C}_{G_i}\, G_i(X_i)$.

(b) In the bivariate case $n = 2$, Gaussian multivariance coincides with the Gaussian covariance defined in [8], Section 7.

(c) If ψ_i is given by the Euclidean norm, then G_i is a Brownian field indexed by $\mathbb{R}^{d_i}$. In particular, if $n = 2$ and both ψ_1 and ψ_2 are given by the Euclidean norm, then $\mathcal{G}(X_1, X_2)$ coincides with the *Brownian covariance* of Székely and Rizzo [23].

(d) If $\psi_i(x) = |x|^\alpha$, then G_i is a fractional Brownian field with Hurst exponent $H = \frac{\alpha}{2}$; cf. [23], Section 4.

THEOREM 3.15. *Suppose that one of the moment conditions of Definition 2.3 holds and $\mathbb{E}(\psi_i(X_i)^{\frac{n}{2}}) < \infty$ for $i = 1, \dots, n$. Then distance multivariance and Gaussian multivariance coincide*, *that is*,

$$M_\rho(X_1, \dots, X_n) = \mathcal{G}(X_1, \dots, X_n). \tag{3.27}$$

PROOF. By Corollary 3.12, we can represent squared multivariance in the product form (3.20). Each of the factors can be rewritten as

$$\begin{aligned}
&-C_X C_{X'} \psi(X - X') \\
&\quad = C_X C_{X'}\big(\psi(X) + \psi(X') - \psi(X - X')\big) \\
&\quad = C_X C_{X'} \mathbb{E}\big(G(X) G(X') \mid X, X'\big) \\
&\quad = \mathbb{E}\big(G(X)G(X') \mid X, X'\big) - \mathbb{E}\big(G(X)G(X') \mid X\big) \\
&\qquad - \mathbb{E}\big(G(X)G(X') \mid X'\big) + \mathbb{E}\big(G(X)G(X')\big) \\
&\quad = \mathbb{E}\big[\big(G(X) - \mathbb{E}(G(X) \mid G)\big)\big(G(X') - \mathbb{E}(G(X') \mid G)\big) \mid X, X'\big] \\
&\quad = \mathbb{E}\big(X^G X'^G \mid X, X'\big), \qquad (3.28)
\end{aligned}$$

where we have used the covariance structure (3.24) of the Gaussian process G in the third line. Putting everything together, we have

$$\begin{aligned}
M_\rho^2(X_1, \dots, X_n) &= \mathbb{E}\left(\prod_{i=1}^n -C_{X_i} C_{X'_i} \psi_i(X_i - X'_i)\right) \\
&= \mathbb{E}\left(\prod_{i=1}^n \mathbb{E}\big(X_i^{G_i} X_i'^{G_i} \mid X_i, X'_i\big)\right) = \mathbb{E}\left(\prod_{i=1}^n X_i^{G_i} X_i'^{G_i}\right) \\
&= \mathcal{G}^2(X_1, \dots, X_n).
\end{aligned}$$

Note that for the penultimate equality the absolute integrability of the integrand, that is, $\mathbb{E}(\prod_{i=1}^n |X_i^{G_i} X_i'^{G_i}|) < \infty$, is required.

Writing $\mathcal{F} := \sigma(X_i, i = 1, \dots, n)$ and $\mathcal{F}' := \sigma(X_i', i = 1, \dots, n)$, we obtain

$$\begin{aligned}
\mathbb{E}\left(\prod_{i=1}^n |X_i^{G_i} X_i'^{G_i}|\right) &= \mathbb{E}\left(\prod_{i=1}^n \mathbb{E}(|X_i^{G_i} X_i'^{G_i}| \mid \mathcal{F}, \mathcal{F}')\right) \\
&\le \mathbb{E}\left(\prod_{i=1}^n \sqrt{\mathbb{E}(|X_i^{G_i}|^2 \mid \mathcal{F}, \mathcal{F}')\mathbb{E}(|X_i'^{G_i}|^2 \mid \mathcal{F}, \mathcal{F}')}\right) \\
&= \mathbb{E}\left(\sqrt{\prod_{i=1}^n \mathbb{E}(|X_i^{G_i}|^2 \mid \mathcal{F})}\right) \cdot \mathbb{E}\left(\sqrt{\prod_{i=1}^n \mathbb{E}(|X_i'^{G_i}|^2 \mid \mathcal{F}')}\right) \\
&= \mathbb{E}\left(\sqrt{\prod_{i=1}^n \mathbb{E}(|X_i^{G_i}|^2 \mid \mathcal{F})}\right)^2 \le \left(\prod_{i=1}^n \mathbb{E}[(\mathbb{E}(|X_i^{G_i}|^2 \mid \mathcal{F}))^{\frac{n}{2}}]\right)^{\frac{2}{n}} \\
&\le \left(\prod_{i=1}^n \mathbb{E}[\mathbb{E}(|X_i^{G_i}|^n \mid \mathcal{F})]\right)^{\frac{2}{n}} = \left(\prod_{i=1}^n \mathbb{E}(|X_i^{G_i}|^n)\right)^{\frac{2}{n}},
\end{aligned}$$

where we used successively the independence of the G_i, the conditional Hölder inequality [10], 7.2.4, the independence and identical distribution of $(X_i, i = 1, \dots, n)$ and $(X_i', i = 1, \dots, n)$, the generalized Hölder inequality [20], page 133, Proposition 13.5, and the conditional Jensen inequality [10], 7.1.4.

Finally, note that for $n \in \mathbb{N}$ the elementary inequality $|a + b|^n \le 2^{n-1}(|a|^n + |b|^n)$ and the formula for absolute moments of Gaussian random variables, that is, $\mathbb{E}(|G_i(t)|^n) = 2^{\frac{n}{2}}\Gamma(\frac{n+1}{2})\pi^{-\frac{1}{2}}[\mathbb{E}G_i(t)^2]^{\frac{n}{2}}$, and $\mathbb{E}[G_i(t)^2] = 2\psi_i(t)$ imply

$$\mathbb{E}|X_i^{G_i}|^n \le 2^n \mathbb{E}|G_i(X_i)|^n = 2^{2n}\Gamma\left(\frac{n+1}{2}\right)\pi^{-\frac{1}{2}}\mathbb{E}(\psi_i(X_i)^{\frac{n}{2}}), \tag{3.29}$$

which proves the desired integrability. □

We conclude this section by comparing (total) distance multivariance to related approaches in [1] and to the multivariate Hilbert–Schmidt independence criterion (dHSIC) of [18].

3.4. *Comparison with* [1]. The problem of generalizing distance covariance of two random variables X, Y to multiple variables has been discussed in a short paragraph "*How to (not) extend [distance covariance] $\mathcal{V}(X, Y)$ to more than two random variables*" in [1]. In the notation of our paper, they discuss for three random variables X, Y, Z the following objects:

(a) Gaussian Covariance $\mathcal{G}(X, Y, Z) = \mathbb{E}(X^G X'^G Y^G Y'^G Z^G Z'^G)$ (cf. Section 3.3) where G is a Brownian motion. This approach is dismissed in [1] since it does not characterize the independence of X, Y, Z.

(b) The quantity

$$\int_{\mathbb{R}^d} \left|\mathbb{E}\left[e^{i(X\cdot t_1+Y\cdot t_2+Z\cdot t_3)}\right] - f_X(t_1) f_Y(t_2) f_Z(t_3)\right|^2 \rho(dt_1, dt_2, dt_3); \tag{3.30}$$

this should be compared with the similar, yet different expression (2.4). Bakirov and Székely dismiss this approach, since the integral can become infinite if $Z \equiv 0$, even if X and Y are bounded and independent; note that in this case the three random variables X, Y, Z are actually independent.

(c) The (bivariate) distance covariance of $U \sim \mathcal{L}(X, Y, Z)$ and $V \sim \mathcal{L}(X) \otimes \mathcal{L}(Y) \otimes \mathcal{L}(Z)$. Bakirov and Székely recommend to use this approach, since it is able to detect independence of X, Y, Z, but they do not follow up this approach with a deeper discussion.

Comparing with our results, let us add a few comments. The approach (a) is equivalent to the calculation of distance multivariance $M_\rho(X, Y, Z)$ (based on Euclidean distance), by Theorem 3.15. Consistent with the remarks of [1], distance multivariance cannot characterize independence; cf. Theorem 3.4. It serves, however, as a building block of *total distance multivariance*, which *does* characterize independence.

If $Z \equiv 0$, the expression (2.2) is zero, that is, it does not suffer from the particular integrability problems as (3.30). However, under certain conditions, it coincides with (3.30); see Corollary 3.3.

Compared with (c), our approach has the advantage that both distance multivariance and total distance multivariance have a very simple and efficient finite-sample representation, which retains all the benefits of the bivariate distance covariance; cf. Theorem 4.1. Also the asymptotic properties of the estimators are similar to the bivariate case; cf. Theorems 4.5, 4.10 and Section 4 in [8].

3.5. *Comparison with* dHSIC. The multivariate Hilbert–Schmidt independence criterion (dHSIC) was recently introduced in [18]. Using our notation, dHSIC is given by

$$\begin{aligned} \mathrm{dHSIC}(X_1, \ldots, X_n) :=& \mathbb{E}\left[\prod_{i=1}^n k_i(X_i, X_i')\right] + \prod_{i=1}^n \mathbb{E}[k_i(X_i, X_i')] \\ &- 2\mathbb{E}\left[\prod_{i=1}^n \mathbb{E}[k_i(X_i, X_i') \mid X_i]\right], \end{aligned} \tag{3.31}$$

where the k_i are continuous, bounded, characteristic, positive semidefinite kernels on $\mathbb{R}^{d_i}$. Here, a kernel $k(x, y)$ is said to be characteristic, if

$$\mu \mapsto \Pi(\mu) = \int k(x, \cdot)\mu(dx)$$

from the finite Borel measures to a suitable Hilbert space is an injective map; see [18], Section 2.1) for details.

Note that any continuous negative definite function ψ_i gives rise to a continuous positive semidefinite kernel under the correspondence

$$k_i(x, y) = \psi_i(x) + \psi_i(y) - \psi_i(y - x); \tag{3.32}$$

see [21]. In the bivariate case ($n = 2$), it is shown in [21] that dHSIC is equivalent to distance covariance with (quasi-)distance ψ_i. This raises the question whether equivalence of dHSIC and (total) distance multivariance still holds in the case $n > 2$. It can be easily shown by numerical experiments that they are not identical, at least not under the correspondence (3.32). Nevertheless, the experiments show a strong positive association between dHSIC and total multivariance. Clarifying the exact nature of this association remains an open question, but we present the following related result: Given the marginal distributions $\mathcal{L}(X_1), \dots, \mathcal{L}(X_n)$, we can find kernels k_i, *depending on these distributions*, such that dHSIC coincides formally with (total) distance multivariance on the random vector $(X_1, \dots, X_n)$. Note that, in general, these kernels are unbounded and its sample versions depend on all samples, thus they are beyond the restrictions imposed in [18].

PROPOSITION 3.16. *Let $X_1, \dots, X_n$ satisfy one of the moment conditions of Definition 2.3 and define the kernels*

$$\begin{aligned} k_i^\lambda(x_i, x_i') := &- \psi_i(x_i - x_i') + \mathbb{E}(\psi_i(x_i - X_i')) \\ &+ \mathbb{E}(\psi_i(X_i - x_i')) - \mathbb{E}(\psi_i(X_i - X_i')) + \lambda, \end{aligned} \tag{3.33}$$

where $\lambda \geq 0$ and write dHSIC$^\lambda$ *for the corresponding quantity defined in* (3.31). *Then*

$$\text{dHSIC}^0(X_1, \dots, X_n) = M_\rho^2(X_1, \dots, X_n), \tag{3.34}$$

$$\text{dHSIC}^1(X_1, \dots, X_n) = \overline{M}_\rho^2(X_1, \dots, X_n). \tag{3.35}$$

The kernel k_i^0 is not characteristic in the sense of [18], *Section* 2.1.

PROOF. Observe that $\mathbb{E}[k_i^\lambda(X_i, X_i')] = \mathbb{E}[k_i^\lambda(X_i, X_i') \mid X_i] = \lambda$, such that (3.31) simplifies to

$$\text{dHSIC}^\lambda(X_1, \dots, X_n) := \mathbb{E}\left[\prod_{i=1}^n (\lambda + k_i^0(X_i, X_i'))\right] - \lambda^n.$$

This is equal to (3.20) for $\lambda = 0$ and to (3.21) for $\lambda = 1$. It remains to show that k_i^0 is not characteristic. To this end, denote by μ_i the distribution of X_i. Then

$$\Pi(\mu_i)(y) = \int k^\lambda(x, y)\mu_i(x) = \mathbb{E}[k_i^\lambda(X_i, y)] = \lambda.$$

If $\lambda = 0$, then $\Pi(\mu_i) = 0 = \Pi(\mathbf{0})$, where $\mathbf{0}$ is the measure of mass zero. This shows that Π is not injective and, therefore, that k_i^0 is not characteristic. □

4. Statistical properties of distance multivariance.

4.1. *Sample distance multivariance.* We now consider a sample of N observations $(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)})$ of the random vector $\boldsymbol{X} = (X_1, \dots, X_n)$. Every observation $\boldsymbol{x}^{(j)}$ is a vector in $\mathbb{R}^d$, $d = d_1 + \cdots + d_n$, of the form $\boldsymbol{x}^{(j)} = (x_1^{(j)}, \dots, x_n^{(j)})$, with each $x_i^{(j)}$ in $\mathbb{R}^{d_i}$. Given such a sample, we denote by $(\hat{X}_1, \dots, \hat{X}_n)$ the random vector with the corresponding empirical distribution. Evaluating distance multivariance at this vector, we obtain the sample distance multivariance

$$^{N}M_\rho^2(x^{(1)}, \dots, x^{(N)}) := M_\rho^2(\hat{X}_1, \dots, \hat{X}_n),$$

which turns out to have a surprisingly simple representation.

Recall that the Hadamard (or Schur) product of two matrices $A, B \in \mathbb{R}^{N \times N}$ is the $N \times N$-matrix $A \circ B$ with entries $(A \circ B)_{jk} = A_{jk} B_{jk}$.

THEOREM 4.1. *Let $(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)})$ be a sample of size N.*

(a) *The sample distance multivariance can be written as*

$$^{N}M_\rho^2(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)}) = \frac{1}{N^2} \sum_{j,k=1}^{N} (A_1 \circ \cdots \circ A_n)_{jk}$$

$$= \frac{1}{N^2} \sum_{j,k=1}^{N} (A_1)_{jk} \cdot \ldots \cdot (A_n)_{jk}; \tag{4.1}$$

here, $A_i := -C B_i C$ where $B_i = (\psi_i(x_i^{(j)} - x_i^{(k)}))_{j,k=1,\dots,N}$ is the distance matrix and $C = I - \frac{1}{N}\mathbb{1}$ the centering matrix.

(b) *The sample total distance multivariance can be written as*

$$^{N}\overline{M}_\rho^2(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)}) = \left[\frac{1}{N^2} \sum_{j,k=1}^{N} \big(1 + (A_1)_{jk}\big) \cdot \ldots \cdot \big(1 + (A_n)_{jk}\big)\right] - 1. \tag{4.2}$$

REMARK 4.2. (a) If n is even, then A_i can be replaced by $-A_i$. This explains the different sign used in the case $n = 2$; cf. [23], Definition 3, and [8], Lemma 4.2, Remark 4.3.

If $n = 2$, then $\sum_{j,k=1}^{N} (A_1 \circ A_2)_{jk} = \operatorname{trace}(A_2^\top A_1)$ and the generalized sample distance covariance from [8], Section 4, is recovered. If in addition $\psi_i(x) = |x|$, that is, the Euclidean distance, then we get the sample distance covariance of Székely *et al.* [23, 25].

(b) Since the ψ_i are continuous negative definite functions, the matrices $-B_i$ are conditionally positive definite matrices, that is, $-\lambda^\top B_i \lambda \geq 0$ for all nonzero λ in $\mathbb{R}^N$ with $\lambda_1 + \cdots + \lambda_N = 0$. As the double centerings of conditionally positive definite matrices, the matrices A_i are positive definite. By Schur's theorem, the

N-fold Hadamard product of positive definite matrices is again positive definite; see Berg and Forst [3], Lemma 3.2. This gives a simple explanation as to why ${}^N M_\rho^2$ is always a nonnegative number.

(c) Important special cases are when the ψ_i are chosen as Euclidean distance, or as Minkowski distances. In these cases, each B_i is a distance matrix. In general, B_i need not be a distance matrix, since only $\sqrt{\psi_i}$, but not necessarily ψ_i itself, defines a distance. Still, ψ_i always defines a quasi-metric, that is, a metric with a relaxed triangle inequality; cf. [8], Section 2.

(d) Even though total distance multivariance is defined as the sum of the multivariances of all $2^n - 1 - n$ subfamilies of $\{X_1, \dots, X_n\}$ with at least two members (cf. (2.3)), its empirical version (4.2) has a computational complexity of only $\mathcal{O}(nN^2)$.

(e) The row and column sums of each A_i are zero. This is a consequence of the double centering $A_i = -CB_iC$.

(f) Equation (4.1) is a direct analogue of the representation (3.20), when the centering operator is replaced by the centering matrix. The same is true for (4.2) in relation to (3.21).

PROOF OF THEOREM 4.1. Since the support of the empirical distribution is finite, the moment conditions of Definition 2.3 are trivially satisfied. Therefore, we can use the representation (3.20) to get

$$\begin{aligned}
&{}^N M_\rho^2(x^{(1)}, \dots, x^{(N)}) \\
&= M_\rho^2(\hat{X}_1, \dots, \hat{X}_n) \\
&= \mathbb{E}\Bigg(\prod_{i=1}^n [-\psi_i(\hat{X}_i - \hat{X}_i') + \mathbb{E}(\psi_i(\hat{X}_i - \hat{X}_i') \mid \hat{X}_i) \\
&\qquad + \mathbb{E}(\psi_i(\hat{X}_i - \hat{X}_i') \mid \hat{X}_i') - \mathbb{E}\psi_i(\hat{X}_i - \hat{X}_i')]\Bigg) \\
&= \frac{1}{N^2}\sum_{j,k=1}^N \Bigg(\prod_{i=1}^n [-\psi_i(x_i^{(j)} - x_i^{(k)}) + \mathbb{E}(\psi_i(\hat{X}_i - \hat{X}_i') \mid \hat{X}_i = x_i^{(j)}) \\
&\qquad + \mathbb{E}(\psi_i(\hat{X}_i - \hat{X}_i') \mid \hat{X}_i' = x_i^{(k)}) - \mathbb{E}\psi_i(\hat{X}_i - \hat{X}_i')]\Bigg).
\end{aligned} \tag{4.3}$$

Denoting by $\mathbb{1}_N$ the column vector consisting of N ones, we can rewrite the individual terms in (4.3) as

$$\psi_i(x_i^{(j)} - x_i^{(k)}) = (B_i)_{jk}, \tag{4.4a}$$

$$\mathbb{E}(\psi_i(\hat{X}_i - \hat{X}_i') \mid \hat{X}_i = x_i^{(j)}) = \frac{1}{N}\sum_{l=1}^N (B_i)_{jl} = \frac{1}{N}(\mathbb{1}_N^\top B_i)_j, \tag{4.4b}$$

$$\mathbb{E}\big(\psi_i\big(\hat{X}_i - \hat{X}_i'\big) \mid \hat{X}_i' = x_i^{(k)}\big) = \frac{1}{N}\sum_{m=1}^{N}(B_i)_{mk} = \frac{1}{N}(B_i\mathbb{1}_N)_k, \tag{4.4c}$$

$$\mathbb{E}\psi_i\big(\hat{X}_i - \hat{X}_i'\big) = \frac{1}{N^2}\sum_{l,m=1}^{N}(B_i)_{ml} = \frac{1}{N^2}\mathbb{1}_N^\top B_i\mathbb{1}_N. \tag{4.4d}$$

This shows that each factor on the right-hand side of (4.3) is the (j,k)th entry of the matrix $A_i = -CB_iC$, and (4.1) follows. The representation (4.2) can be derived in complete analogy from (3.21). □

4.2. *Estimating distance multivariance*. In this section, we examine the properties of the sample distance multivariance ${}^N M_\rho$ as an estimator of M_ρ. The corresponding results for the sample total distance multivariance will be presented in the next section.

THEOREM 4.3 (${}^N M_\rho$ is a strongly consistent estimator for M_ρ). *Let one of the moment conditions of Definition 2.3 be satisfied. Then ${}^N M_\rho$ is a strongly consistent estimator of M_ρ, that is,*

$${}^N M_\rho\big(\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)}\big) \xrightarrow[N\to\infty]{} M_\rho(X_1, \dots, X_n) \qquad a.s. \tag{4.5}$$

PROOF. Inserting the representation (4.4) into (4.3), we see that ${}^N M_\rho$ is a V-statistic. Thus the convergence of the estimator ${}^N M_\rho$ is just the strong law of large numbers for V-statistics. □

REMARK 4.4. In the case of $n = 2$ strong consistency can be obtained under the weaker moment condition $\mathbb{E}\psi_i(X_i) < \infty$ for $i = 1, 2$; see [8], Theorem 4.4. For $n \geq 3$, the arguments used in [8] break down. However, we show a weak consistency result under independence and relaxed moment conditions in Corollary 4.7 below.

The next result is our main result on the asymptotics properties of the estimator ${}^N M_\rho$. The proof is technical and relegated to the Supplementary Material [9].

THEOREM 4.5 (Asymptotic distribution of ${}^N M_\rho$). (a) *Let $X_1, \dots, X_n$ be independent random variables such that the moments $\mathbb{E}\psi_i(X_i) < \infty$ and $\mathbb{E}[\log^{1+\epsilon}(1 + |X_i|^2)] < \infty$ exist for some $\epsilon > 0$ and all $i = 1, \dots, n$. Then*

$$N \cdot {}^N M_\rho^2\big(\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)}\big) \xrightarrow[N\to\infty]{d} \|\mathbb{G}\|^2_{L^2(\rho)}, \tag{4.6}$$

where $\mathbb{G}$ is a centred, that is, $\mathbb{E}\mathbb{G}(t) = 0$, $\mathbb{C}$-valued Gaussian process indexed by $\mathbb{R}^d$ with covariance function

$$\operatorname{Cov}\big(\mathbb{G}(t), \mathbb{G}(t')\big) = \mathbb{E}\big[\mathbb{G}(t)\overline{\mathbb{G}(t')}\big] = \prod_{i=1}^{n}\big(f_{X_i}\big(t_i - t_i'\big) - f_{X_i}(t_i)\overline{f_{X_i}\big(t_i'\big)}\big). \tag{4.7}$$

(b) *Suppose that the random variables* $X_1, \dots, X_n$ *are* $(n-1)$-*independent, but not* n-*independent and that one of the moment conditions of Definition* 2.3 *holds. Then*

$$N \cdot {}^N M_\rho^2(\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)}) \xrightarrow[N\to\infty]{} \infty \qquad a.s. \tag{4.8}$$

REMARK 4.6. (a) The complex-valued Gaussian process $\mathbb{G}$ has to be distinguished from the Gaussian processes G_i that appear in Definition 3.13 of the Gaussian multivariance.

(b) Using the results of [13], the log-moment condition in (a) can be relaxed by a weaker (but more involved) integral test; cf. [13], Condition ($\star$).

From [8], Lemma 2.7, it is readily seen that the log-moment condition in Theorem 4.5.a) is equivalent to $\mathbb{E}[\log^{1+\epsilon}(1 \vee \sqrt{|X_1|^2 + \cdots + |X_n|^2})] < \infty$.

(c) The expectation of the limit in (4.6) can be calculated as

$$\mathbb{E}(\|\mathbb{G}\|^2_{L^2(\rho)}) = \prod_{i=1}^n \int_{\mathbb{R}^{d_i}} (1 - |f_{X_i}(t_i)|^2)\rho_i(dt_i) = \prod_{i=1}^n \mathbb{E}\psi_i(X_i - X_i'). \tag{4.9}$$

(d) From Lemma S.2 in the Supplementary Material [9], it can be seen that ${}^N M_\rho$ is a biased estimator of M_ρ, since in the case of nondegenerate and independent random variables

$$\mathbb{E}[{}^N M_\rho^2(\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(n)})] = \frac{(N-1)^n + (-1)^n(N-1)}{N^{n+1}} \prod_{i=1}^n \mathbb{E}\psi_i(X_i - X_i') > 0,$$

while $M_\rho^2(X_1, \dots, X_n) = 0$. For bivariate distance covariance, this bias has already been discussed by Cope [12] and Székely and Rizzo [24].

Finally, we present a weak consistency result for ${}^N M_\rho$ under independence, which holds under milder moment conditions than the strong consistency result Theorem 4.3.

COROLLARY 4.7. *Suppose that* $X_1, \dots, X_n$ *are independent random variables with* $\mathbb{E}\psi_i(X_i) < \infty$ *and* $\mathbb{E}[\log^{1+\epsilon}(1 + |X_i|^2)] < \infty$ *for some* $\epsilon > 0$ *and all* $i = 1, \dots, n$. *Then*

$${}^N M_\rho(\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)}) \xrightarrow[N\to\infty]{} 0 \qquad \text{in probability.} \tag{4.10}$$

PROOF. The corollary is a direct consequence of Theorem 4.5 and the observation that

$$nZ_n \xrightarrow{d} Z \quad \Longrightarrow \quad Z_n \xrightarrow{d} 0 \quad \Longrightarrow \quad Z_n \xrightarrow{\mathbb{P}} 0;$$

the second implication follows since the d-limit is degenerated. □

4.3. *Estimating total distance multivariance.* To simplify notation, we write $\rho_S = \bigotimes_{i\in S}\rho_i$. Recall that

$$(4.11)\qquad {}^N\overline{M}^2_\rho(\boldsymbol{X}^{(1)},\dots,\boldsymbol{X}^{(N)}) = \sum_{\substack{S\subset\{1,\dots,n\}\\ |S|\ge 2}} {}^N M^2_{\rho_S}(\boldsymbol{X}^{(1)},\dots,\boldsymbol{X}^{(N)}).$$

Note that M_{ρ_S} depends only on the random variables $(X_i, i\in S)$, that is, $M_{\rho_S} = M_{\rho_S}(X_i, i\in S)$. This means that the sample version

$${}^N M_{\rho_S} = {}^N M_{\rho_S}(\boldsymbol{X}^{(1)},\dots,\boldsymbol{X}^{(N)})$$

is computed only from the S-coordinates of the samples $\boldsymbol{X}^{(1)},\dots,\boldsymbol{X}^{(N)}$. The results of this section are mostly direct consequences of the results of the previous section (replacing M_ρ by M_{ρ_S} and ${}^N M_\rho$ by ${}^N M_{\rho_S}$).

COROLLARY 4.8 (${}^N\overline{M}_\rho$ is a strongly consistent estimator of $\overline{M}_\rho$). *Assume that one of the moment conditions of Definition 2.3 is satisfied. Then*

$$(4.12)\qquad {}^N\overline{M}_\rho(\boldsymbol{X}^{(1)},\dots,\boldsymbol{X}^{(N)}) \xrightarrow[N\to\infty]{} \overline{M}_\rho(X_1,\dots,X_n)\qquad a.s.$$

PROOF. Apply Theorem 4.3 to each M_{ρ_S} in (4.11). □

COROLLARY 4.9. *Let $X_1,\dots,X_n$ be independent random variables with $\mathbb{E}\psi_i(X_i)<\infty$ and $\mathbb{E}[\log^{1+\epsilon}(1+|X_i|^2)]<\infty$ for some $\epsilon>0$ and all $i=1,\dots,n$. Then*

$$(4.13)\qquad {}^N\overline{M}_\rho(\boldsymbol{X}^{(1)},\dots,\boldsymbol{X}^{(N)}) \xrightarrow[N\to\infty]{} 0\qquad \textit{in probability}.$$

PROOF. Apply Corollary 4.7 to each M_{ρ_S} in (4.11). □

The next theorem is the analogue of the convergence result Theorem 4.5. For each $S\subset\{1,\dots,n\}$, we denote by $\mathbb{G}_S$ the centred Gaussian process

$$(4.14)\qquad \mathbb{G}_S(t_S) := \sum_{R\subset S}(-1)^{|S|-|R|}\int e^{\mathrm{i}x_R\cdot t_R}\,\mathrm{d}B(x)\cdot\prod_{j\in S\setminus R} f_j(t_j)$$

(cf. (S.15) in the Supplementary Material [9]), indexed by $t_S\in \times_{i\in S}\mathbb{R}^{d_i}$, and where B is the Brownian bridge from (S.12) in the Supplementary Material [9]. Applying Theorem 4.5 with $\{1,\dots,n\}$ replaced by S, we see that $\mathbb{G}_S$ has covariance structure

$$(4.15)\qquad \mathbb{E}(\mathbb{G}_S(t)\overline{\mathbb{G}_S(t')}) = \prod_{i\in S}\big(f_{X_i}(t_i-t_i') - f_{X_i}(t_i)\overline{f_{X_i}(t_i')}\big).$$

THEOREM 4.10 (Asymptotic distribution of ${}^N\overline{M}_\rho$). (a) *Suppose that* $X_1, \dots, X_n$ *are independent with* $\mathbb{E}\psi_i(X_i) < \infty$ *and* $\mathbb{E}[\log^{1+\epsilon}(1+|X_i|^2)] < \infty$ *for some* $\epsilon > 0$ *and all* $i = 1, \dots, n$. *Then*

$$N \cdot {}^N\overline{M}^2_\rho(\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)}) \xrightarrow[N\to\infty]{d} \sum_{\substack{S\subset\{1,\dots,n\}\\ |S|\geq 2}} \|\mathbb{G}_S\|^2_{L^2(\rho_S)}. \tag{4.16}$$

(b) *Suppose that the random variables* $X_1, \dots, X_n$ *are not independent and that one of the moment conditions of Definition* 2.3 *holds*. *Then*

$$N \cdot {}^N\overline{M}^2_\rho(\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)}) \xrightarrow[N\to\infty]{} \infty \qquad a.s. \tag{4.17}$$

REMARK 4.11. Note that the processes $(\mathbb{G}_S)$, $S \subset \{1, \dots, n\}$ on the right-hand side of (4.16) are *jointly Gaussian*. Therefore, the limit appearing in (4.16) is a quadratic form of centred Gaussian random variables. This fact will be used in Section 4.5 to construct a statistical test of (multivariate) independence. Further properties of the processes $\mathbb{G}_S$ are discussed in [5].

PROOF OF THEOREM 4.10. (a) For any $S \subset \{1, \dots, n\}$ with $|S| \geq 2$, we know from Theorem 4.5 that

$$N \cdot {}^N M^2_{\rho_S}(\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)}) \xrightarrow[N\to\infty]{} \|\mathbb{G}_S\|^2_{\rho_S},$$

and (4.16) follows.

(b) By Corollary 4.8, we have ${}^N\overline{M}_\rho \to \overline{M}_\rho$ almost surely. Moreover, $\overline{M}_\rho > 0$ by Theorem 3.4, since the random variables $(X_1, \dots, X_n)$ are not independent. Thus, $N \cdot {}^N\overline{M}^2_\rho \to \infty$ almost surely. □

4.4. *Normalizing and scaling distance multivariance*. With practical applications in mind, there are at least two reasons to consider rescaled versions of (total) distance multivariance:

- To obtain a *distance multicorrelation* whose value is bounded by 1—analogous to Székely-Rizzo-and-Bakirov's distance correlation [25], Definition 3;
- To normalize the asymptotic distribution of the sample (total) distance multivariance under independence; cf. Theorem 4.5 and Theorem 4.10.

We will use normalized multivariances as test statistics in two tests for independence in Section 4.5. For the scaling constants we use in the following the convention $0/0 := 0$. This ensures that we also cover the case of degenerated (i.e., constant) random variables.

Distance multicorrelation.

DEFINITION 4.12. Let $X_1, \dots, X_n$ be random variables with $\mathbb{E}\psi_i^n(X_i) < \infty$ for all $i = 1, \dots, n$. We set

$$a_i := \| \mathrm{C}_{X_i} \mathrm{C}_{X_i'} \psi_i(X_i - X_i') \|_{L^n(\mathbb{P})}$$

and define *distance multicorrelation* as

$$\mathcal{R}_\rho^2(X_1, \dots, X_n) := \frac{M_\rho^2(X_1, \dots, X_n)}{a_1 \cdot \dots \cdot a_n}. \tag{4.18}$$

For the sample version of distance multicorrelation, we define

$${}^N a_i := {}^N a_i(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)}) = \left(\frac{1}{N^2} \sum_{k,l=1}^N |(A_i)_{kl}|^n \right)^{1/n}, \tag{4.19}$$

where the A_i are the doubly centred matrices from Theorem 4.1, and set

$${}^N\mathcal{R}_\rho^2(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)}) := \frac{1}{N^2} \sum_{k,l=1}^N \frac{(A_1)_{kl}}{{}^N a_1} \cdot \dots \cdot \frac{(A_n)_{kl}}{{}^N a_n}. \tag{4.20}$$

Note that $a_i = 0$ if, and only if, X_i is degenerate, hence, $\mathcal{R}_\rho(X_1, \dots, X_n)$ is well defined as a finite nonnegative number.

PROPOSITION 4.13. (a) *Distance multicorrelation and its sample version satisfy*

$$0 \le \mathcal{R}_\rho(X_1, \dots, X_n) \le 1 \quad \textit{and} \quad 0 \le {}^N\mathcal{R}_\rho(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)}) \le 1. \tag{4.21}$$

(b) *For i.i.d. copies $\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)}$ of $\boldsymbol{X} = (X_1, \dots, X_n)$ it holds that*

$$\lim_{N \to \infty} {}^N\mathcal{R}_\rho(\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)}) = \mathcal{R}_\rho(X_1, \dots, X_n) \qquad \textit{a.s.}$$

(c) *For $n = 2$ and $\psi_1(x) = \psi_2(x) = |x|$, distance multicorrelation coincides with the distance correlation of* [25].

REMARK 4.14. Székely and Rizzo [23], Theorem 4(iv), show for the case $n = 2$ (i.e., for distance correlation) that ${}^N\mathcal{R}_\rho(\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)}) = 1$ implies that the sample points $(x_1^{(1)}, \dots, x_1^{(N)})$ and $(x_2^{(1)}, \dots, x_2^{(N)})$ can be transformed into each other by a Euclidean isometry composed with scaling by a nonnegative number. An analogous result seems not to hold for distance multicorrelation in the case $n > 2$.

PROOF OF PROPOSITION 4.13. By the generalized Hölder inequality for n-fold products (cf. [20], page 133, Problem 13.5), we have that

$$\begin{aligned} M_\rho^2(X_1, \dots, X_n) &= \mathbb{E}\left(\prod_{i=1}^n - \mathrm{C}_{X_i}\,\mathrm{C}_{X_i'}\,\psi_i(X_i - X_i')\right) \\ &\le \mathbb{E}\left(\prod_{i=1}^n \left|\mathrm{C}_{X_i}\,\mathrm{C}_{X_i'}\,\psi_i(X_i - X_i')\right|\right) \\ &\le \prod_{i=1}^n \left\|\mathrm{C}_{X_i}\,\mathrm{C}_{X_i'}\,\psi_i(X_i - X_i')\right\|_{L^n(\mathbb{P})} = a_1 \cdot \ldots \cdot a_n, \end{aligned}$$

and (4.21) follows. For the convergence result, note that

$$ {}^N a_i^n = \frac{1}{N^2} \sum_{k,l=1}^N \left|(A_i)_{kl}\right|^n \xrightarrow[N\to\infty]{} \mathbb{E}\left[\left|\mathrm{C}_{X_i}\,\mathrm{C}_{X_i'}\,\psi_i(X_i - X_i')\right|^n\right] = a_i^n \tag{4.22}$$

by the law of large numbers for V-statistics; cf. Theorem 4.3 and its proof. Part (c) follows from direct comparison with [23]. □

Normalized distance multivariance. Alternatively, we can normalize distance multivariance in such a way, that the limiting distribution under independence (cf. Theorems 4.5 and 4.10) has unit expectation.

DEFINITION 4.15. Let $X_1, \dots, X_n$ be random variables with $\mathbb{E}\psi_i(X_i) < \infty$ for all $i = 1, \dots, n$, set

$$b_i := \mathbb{E}\psi_i(X_i - X_i')$$

and define *normalized distance multivariance* as

$$\mathcal{M}_\rho^2(X_1, \dots, X_n) := \frac{M_\rho^2(X_1, \dots, X_n)}{b_1 \cdot \ldots \cdot b_n}. \tag{4.23}$$

For the sample version of normalized distance multicorrelation, we define

$${}^N b_i := {}^N b_i(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)}) := \frac{1}{N^2} \sum_{k,l=1}^N \psi_i(x_i^{(l)} - x_i^{(k)}) = \frac{1}{N^2} \sum_{k,l=1}^N (B_i)_{kl}, \tag{4.24}$$

and set

$${}^N\mathcal{M}_\rho^2(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)}) := \frac{1}{N^2} \sum_{k,l=1}^N \frac{(A_1)_{kl}}{{}^N b_1} \cdot \ldots \cdot \frac{(A_n)_{kl}}{{}^N b_n}. \tag{4.25}$$

COROLLARY 4.16. *Suppose that $X_1, \dots, X_n$ are nondegenerate independent random variables with $\mathbb{E}\psi_i(X_i) < \infty$ and $\mathbb{E}[\log^{1+\epsilon}(1 + |X_i|^2)] < \infty$ for some $\epsilon > 0$ and all $i = 1, \dots, n$. Then*

$$N \cdot {}^N\mathcal{M}^2_\rho(\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)}) \xrightarrow[N\to\infty]{d} Q, \tag{4.26}$$

where $Q = \|\mathbb{G}\|^2_\rho/(b_1 \cdot \dots \cdot b_n)$ and $\mathbb{E}Q = 1$.

PROOF. This follows from Theorem 4.5 in combination with

$$^N b_i = \frac{1}{N^2}\sum_{k,l=1}^{N} \psi_i(X_i^{(k)} - X_i^{(l)}) \xrightarrow[N\to\infty]{} \mathbb{E}\psi_i(X_i - X_i') = b_i, \tag{4.27}$$

under the assumption $\mathbb{E}\psi_i(X_i) < \infty$. □

It remains to find an analogous normalization for *total* distance multivariance. For a subset $S \subset \{1, \dots, n\}$, define $M_{\rho_S}(X_1, \dots, X_n)$ as in Section 4.3 and set $b_S = \prod_{i\in S} b_i$.

DEFINITION 4.17. For the random variables $X_1, \dots, X_n$, we define the *normalized total distance multivariance* as

$$\overline{\mathcal{M}}^2_\rho(X_1, \dots, X_n)^2 := \frac{1}{2^n - 1 - n}\sum_{\substack{S\subset\{1,\dots,n\}\\ |S|\ge 2}} \frac{M^2_{\rho_S}(X_i, i \in S)}{b_S}. \tag{4.28}$$

Its sample version becomes

$$\begin{aligned} &{}^N\overline{\mathcal{M}}^2_\rho(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)}) \\ &:= \frac{1}{2^n - 1 - n}\left\{\frac{1}{N^2}\sum_{k,l=1}^{N}\left(1 + \frac{(A_1)_{kl}}{^N b_1}\right)\cdot \dots \cdot \left(1 + \frac{(A_n)_{kl}}{^N b_1}\right) - 1\right\}. \end{aligned} \tag{4.29}$$

Similar to Corollary 4.16, we have the following result.

COROLLARY 4.18. *Suppose that $X_1, \dots, X_n$ are nondegenerate independent random variables with $\mathbb{E}\psi_i(X_i) < \infty$ and $\mathbb{E}[\log^{1+\epsilon}(1 + |X_i|^2)] < \infty$ for some $\epsilon > 0$ and all $i = 1, \dots, n$. Then*

$$N \cdot {}^N\overline{\mathcal{M}}^2_\rho(\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)}) \xrightarrow[N\to\infty]{d} \overline{Q}, \tag{4.30}$$

where

$$\overline{Q} = \frac{1}{2^n - n - 1}\sum_{\substack{S\subset\{1,\dots,n\}\\ |S|\ge 2}} \frac{\|\mathbb{G}_S\|^2_{L^2(\rho_S)}}{b_S} \quad \textit{and} \quad \mathbb{E}\overline{Q} = 1.$$

PROOF. Convergence follows from Theorem 4.10. Note that the sum runs over $2^n - n - 1$ subsets and $\mathbb{E}[\|\mathbb{G}_S\|^2_{L^2(\rho_S)}] = b_S$ by Corollary 4.16. □

4.5. *Two tests for independence.* Based on the normalized multivariance statistics $\mathcal{M}_\rho$ and $\overline{\mathcal{M}}_\rho$ and the convergence results of Corollaries 4.16 and 4.18, we can formulate two statistical tests for the independence of the random variables $X_1, \dots, X_n$. To assess a critical value for the test statistics, we use the same approach as Székely and Rizzo [23]: Both limiting random variables Q and $\overline{Q}$ are quadratic forms of centered Gaussian random variables, normalized to $\mathbb{E}Q = \mathbb{E}\overline{Q} = 1$. Hence, by [22], page 181,

$$\mathbb{P}\big(Q \geq \chi^2_{1-\alpha}(1)\big) \leq \alpha \quad \text{and} \quad \mathbb{P}\big(\overline{Q} \geq \chi^2_{1-\alpha}(1)\big) \leq \alpha, \tag{4.31}$$

for all $0 < \alpha \leq 0.215$, where $\chi^2_{1-\alpha}(1)$ denotes the $(1-\alpha)$-quantile of a chi-square distribution with one degree of freedom. Note that (4.31) is, in general, very rough, thus the following tests are, in general, quite conservative. The first test uses multivariance and, therefore, requires the a priori assumption of $(n-1)$-independence.

TEST A. *Let $\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)}$ be observations of the random vector $\boldsymbol{X} = (X_1, \dots, X_n)$, let $\alpha \in (0, 0.215)$, and suppose that the moment conditions of Corollary 4.16 and one of the moment conditions of Definition 2.3 hold. Under the assumption of $(n-1)$-independence, the null hypothesis*

$$\boldsymbol{H_0}: \quad (X_1, \dots, X_n) \text{ are independent}$$

is rejected against the alternative hypothesis

$$\boldsymbol{H_1}: \quad (X_1, \dots, X_n) \text{ are not independent}$$

at level α, if the normalized multivariance ${}^N\mathcal{M}(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)})$ satisfies

$$N \cdot {}^N\mathcal{M}^2_\rho\big(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)}\big) \geq \chi^2_{1-\alpha}(1).$$

The second test uses *total* multivariance, and hence does not require a-priori assumptions, except for the moment conditions. We emphasize that this test on mutual independence can be applied in very general settings: It is distribution-free and the random variables $X_1, \dots, X_n$ can take values in arbitrary dimensions.

TEST B. *Let $(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)})$ be observations of the random vector $\boldsymbol{X} = (X_1, \dots, X_n)$, let $\alpha \in (0, 0.215)$, and suppose that the moment conditions of Corollary 4.18 and one of the moment conditions of Definition 2.3 hold. The null hypothesis*

$$\boldsymbol{H_0}: \quad (X_1, \dots, X_n) \text{ are independent}$$

is rejected against the alternative hypothesis

$$\boldsymbol{H_1}: \quad (X_1, \dots, X_n) \text{ are not independent}$$

at level α, *if the normalized total multivariance* ${}^N\overline{\mathcal{M}}_\rho(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)})$ *satisfies*

$$N \cdot {}^N\overline{\mathcal{M}}^2_\rho(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)}) \geq \chi^2_{1-\alpha}(1).$$

Note that in Test A and Test B the moment conditions of Definition 2.3 ensure the divergence (for $N \to \infty$) of the test statistics in the case of dependence; cf. Theorem 4.5 and Theorem 4.10. Thus these tests are consistent against all alternatives.

In Section 5 below, we give a numerical example of both tests that also allows to assess their power for different sample sizes N.

REMARK 4.19. If the marginal distributions are known, it is possible to perform a Monte Carlo test, where the p-value is obtained from the empirical (Monte Carlo) distribution of the test statistic under H_0. Even without knowledge of the marginal distributions, resampling tests can be performed. These and further tests based on distance multivariance are discussed in [5, 6].

5. Examples. In this section, we present two basic examples which illustrate some key aspects of distance multivariance:

Bernstein's coins: This is a classical example of pairwise independence with higher order dependence. It shows that distance multivariance accurately detects multivariate dependence.

Sinusoidal dependence: This is a basic example which was considered in [21] to illustrate that distance covariance can perform poorly when used to detect small scale (local) dependencies. We show that the flexibility of generalized distance multivariance—due to the choice of the distance functions ψ_i—can be used to improve the power of the test considerably.

5.1. *Bernstein's coins.* The first example of pairwise independent, but not (totally) independent random variables is attributed to S. N. Bernstein; cf. [14], Section V.3. We illustrate this example by using two identical fair coins: coin I and coin II. Based on independent tosses of these two coins, define the following events:

$$A = \{\text{coin I shows heads}\}, \qquad B = \{\text{coin II shows tails}\},$$
$$C = \{\text{both coins show the same side}\}.$$

All events have probability $\frac{1}{2}$, and they are pairwise independent, since

$$\mathbb{P}(A \cap B) = \mathbb{P}(B \cap C) = \mathbb{P}(C \cap A) = \frac{1}{4}.$$

They are, however, not independent, since $A \cap B \cap C = \varnothing$; hence,

$$0 = \mathbb{P}(A \cap B \cap C) \neq \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C) = \frac{1}{8}.$$

Hence, the distance covariances[2] of the pairs (A, B), (B, C) and (C, A) should vanish, due to pairwise independence, while the distance multivariance and the total distance multivariance of the triplet (A, B, C) should detect their higher-order dependence. We discuss both the analytic approach and the numerical simulation of the relevant quantities.

Let ρ_A, ρ_B, ρ_C be one-dimensional symmetric Lévy measures with the corresponding continuous negative definite functions ψ_A, ψ_B and ψ_C. We write $\rho = \rho_A \otimes \rho_B \otimes \rho_C$ and $\rho_{AB} := \rho_A \otimes \rho_B$, etc.

Analytic approach. First, note that pairwise independence yields

$$\begin{aligned} M_{\rho_{AB}}(A, B)^2 &= \int_{\mathbb{R}^2} \big(f_{A,B}(r, s) - f_A(r) f_B(s)\big)^2 \rho_A \otimes \rho_B(\mathrm{d}r, \mathrm{d}s) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} 0 \rho_A(\mathrm{d}r) \rho_B(\mathrm{d}s) = 0, \end{aligned}$$

and similarly for $M_{\rho_{BC}}(B, C)$ and $M_{\rho_{AC}}(C, A)$. On the other hand, from the pairwise independence and Corollary 3.3 we obtain

$$\begin{aligned} M_\rho(A, B, C)^2 &= \int_{\mathbb{R}^3} \big(f_{A,B,C}(r, s, t) - f_A(r) f_B(s) f_C(t)\big)^2 \rho(\mathrm{d}r, \mathrm{d}s, \mathrm{d}t) \\ &= \frac{1}{64} \int_{\mathbb{R}} \int_{\mathbb{R}} \int_{\mathbb{R}} |1 - \mathrm{e}^{\mathrm{i}r}|^2 |1 - \mathrm{e}^{\mathrm{i}s}|^2 |1 - \mathrm{e}^{\mathrm{i}t}|^2 \rho_A(\mathrm{d}r) \rho_B(\mathrm{d}s) \rho_C(\mathrm{d}t) \\ &= \frac{1}{8} \psi_A(1) \psi_B(1) \psi_C(1). \end{aligned}$$

In particular, for $\psi(x) = |x|$ we obtain

$$M_\rho(A, B, C) = \overline{M}_\rho(A, B, C) = \frac{1}{2\sqrt{2}}.$$

We calculate the scaling factors from Section 4.4 as

$$a_A = a_B = a_C = b_A = b_B = b_C = \frac{1}{2},$$

which shows that multicorrelation and normalized multivariance coincide in this case, that is,

$$\mathcal{R}_\rho(A, B, C) = 1 = \mathcal{M}_\rho(A, B, C).$$

Finally, normalized total multivariance is given by

$$\overline{\mathcal{M}}_\rho(A, B, C) = \frac{1}{\sqrt{2^3 - 3 - 1}} \mathcal{M}_\rho(A, B, C) = \frac{1}{2}.$$

---

[2]In slight abuse of notation, we identify the events A, B, C with the random variables $\mathbb{1}_A(\omega), \mathbb{1}_B(\omega), \mathbb{1}_C(\omega)$.

Numerical simulation. To complement the analytical results by a numerical simulation, we have simulated 5000 replications of $N = 3, \dots, 30$ tosses of Bernstein's coins. We calculated the pairwise sample distance covariances ${}^N M_{\rho_{AB}}(A, B)$, ${}^N M_{\rho_{BC}}(B, C)$, ${}^N M_{\rho_{AC}}(C, A)$ as well as the sample distance multivariance ${}^N M_\rho(A, B, C)$ and the sample total distance multivariance ${}^N \overline{M}_\rho(A, B, C)$. We used Euclidean distance as underlying distance in all cases. Due to pairwise independence, the bivariate distance covariances should tend to zero for increasing N, while the multivariances should tend to the nonzero limits that we calculated analytically above.

Figure 1 shows the average values of the multivariance statistics over 5000 replications, along with their empirical 5% and 95% quantiles. Figure (a) uses no scaling, Figure (b) shows "normalized" quantities (cf. Section 4.4) and Figure (c) shows squared normalized quantities scaled by N, as they appear in Theorems 4.5 and 4.10. Also shown is the critical value $\chi^2_{0.95}(1)$ of the test proposed in Section 4.5. In summary, the numerical simulation shows that:

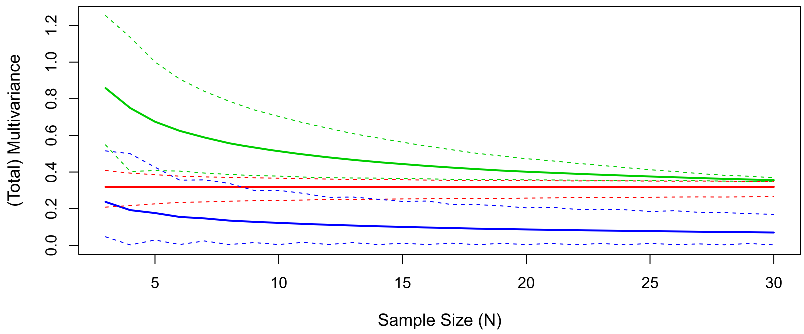
- (Total) distance multivariance is able to distinguish correctly pairwise independence of the events A, B, C from their higher-order dependence;
- The sample statistics converge quickly to their analytic limits and numerically confirm the asymptotic results from Theorems 4.5 and 4.10.
- The hypothesis of pairwise independence of A and B would be correctly accepted in about 95% of simulations, confirming the specificity of the proposed tests.
- Test A (with the a priori assumption of pairwise independence) has a power exceeding 95% for sample sizes $N > 5$. Test B (no a priori assumptions) has a power exceeding 95% for $N > 14$.

Note that all necessary functions and tests for such simulations and for the use of distance multivariance in applications are provided in the R package `multivariance` [7].

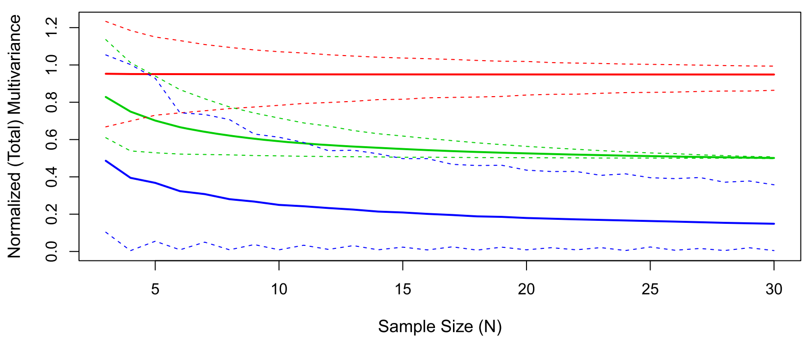
5.2. *Sinusoidal dependence.* In [21], page 2287, it was pointed out that for random variables X, Y with a common *sinusoidal density*

$$f_l(x, y) := \frac{1}{4\pi^2}\bigl(1 + \sin(lx)\sin(ly)\bigr) \qquad \text{on } [-\pi, \pi]^2 \text{ for some } l \in \mathbb{N} \tag{5.1}$$

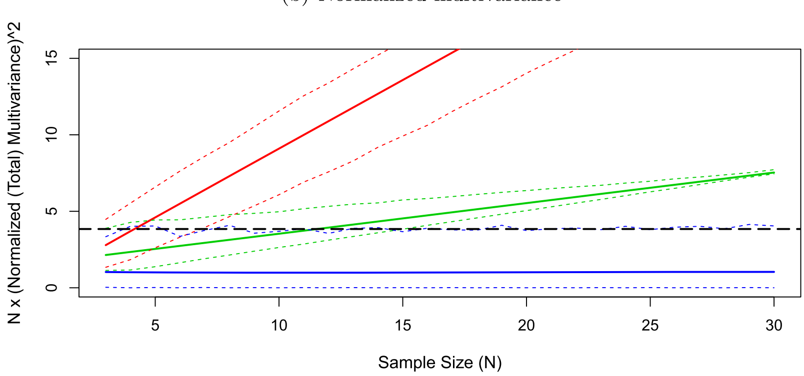
the detection of the dependence using distance covariance is poor for $l > 1$. It was also noted that choosing (in our notation) $\psi_i(x) = |x|^\alpha$ with some $\alpha \neq 1$ might improve the power; see Figure 2(a). Using the bounded continuous negative definite function $\psi_i(x) = \frac{1}{\gamma}(1 - \exp(-\gamma|x|))$ with $\gamma > 0$ can increase the power considerably for larger l; see Figure 2(b). Here, we used the same sample parameters as in [4] (5000 samples, $N = 200$, $\alpha = 0.05$). The p-values were calculated by Monte Carlo estimation with 10,000 replications.



(a) Multivariance without normalization

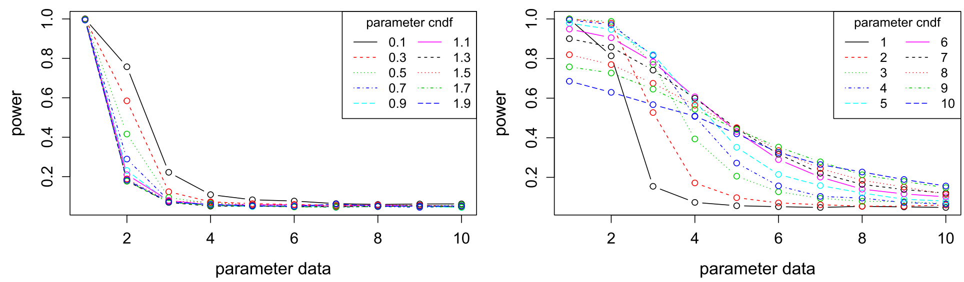


(b) Normalized multivariance



(c) Squared normalized multivariance scaled by sample size

FIG. 1. *These plots show sample distance covariance* ${}^N M_{\rho_{AB}}(A, B)$ *(blue), sample distance multivariance* ${}^N M_\rho(A, B, C)$ *(red) and sample total distance multivariance* ${}^N \overline{M}_\rho(A, B, C)$ *(green) for Bernstein's coin toss experiment (cf. Section* 5*), averaged over* 5000 *Monte-Carlo replications. Also shown are the empirical* 5% *and* 95% *quantiles (dashed). Different scalings are used in the plots* (a)–(c), *and plot* (c) *also shows the critical value (significance level* $\alpha = 5\%$*) of the independence tests from Section* 4.5 *(long dashes, black).*



(a) $\psi_i(x) = |x|^\alpha$ (b) $\psi_i(x) = \frac{1}{\gamma}(1 - \exp(-\gamma|x|))$

FIG. 2. *Power of tests based on distance multivariance for the sinusoidal example with density f_l given in* (5.1). *The parameter of the data is l and the parameter of the ψ_i is α and γ, respectively. Here*, (a) *is the alpha-stable case and* (b) *uses a bounded cndf.*

The following heuristic was used to choose the value of γ: Note that

$$\psi_i(x) := \frac{1}{\gamma}(1 - \exp(-\gamma|x|)) \tag{5.2}$$

is a bounded function which is strictly increasing for $x > 0$. Suppose we know that the local dependencies occur in a window of (Euclidean) distance δ. Thus, it seems reasonable to *neglect* all pairs which are further apart than δ by setting all their ψ_i-distances to (roughly) the same value, that is, we choose γ such that $\psi_i(\delta) \geq 0.99 \cdot \sup_x \psi_i(x)$. This is achieved by setting $\gamma := -\ln(0.01)/\delta$. For the sinusoidal example δ is the period of the sin functions, that is, $\delta = \pi/l$. Let us compare the resulting test with the methods `MINT` and `MINTav` which were proposed in [BS17] for a wide range of situations. Figure 3 shows in the setting of sinusoidal data that

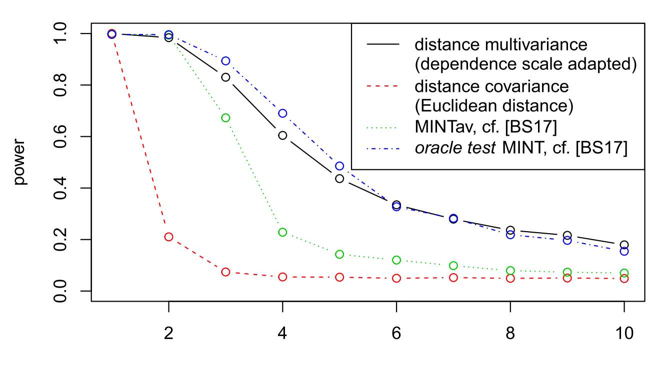


FIG. 3. *Comparison of the power of distance multivariance with distance adapted to the dependence scale, classical distance covariance with Euclidean distance*, `MINTav` *and* `MINT`. *The latter were recently introduced in* [4] *and it was shown that for this example they outperform many* (*all in their comparison*) *other dependence measures.*

our proposed test outperforms `MINTav` and has similar power as the *oracle test* `MINT`.

Note that `MINTav` uses no a priori information about the dependence scale, and that `MINT` computes the p-value using all possible parameters and selects a posteriori the parameter (for each setting) which yielded the highest power. In contrast, our test requires a heuristic parameter selection using certain a priori knowledge of the data generation mechanism.

Further extensions and details on resampling, Monte Carlo and other tests based on distance multivariance can be found in [5, 6].

Acknowledgments. We are grateful to Ulrich Brehm (TU Dresden), for insightful discussions on (elementary) symmetric polynomials and to Georg Berschneider (TU Dresden) who read and commented on the entire text. We would also like to thank Gabor J. Székely (NSF) for advice on the current literature. We thank the anonymous referees and the handling editor for their helpful comments.

NOTE ADDED IN PROOF. *After this paper had been accepted, the authors learned from Prof. Martin Bilodeau that his joint paper with Aurélien Guetsop Nangue, "Tests of mutual or serial independence of random vectors with applications," which has appeared in* The Journal of Machine Learning Research **18** (2017) *pp.* 1–40*—we will refer to this paper as* [*BGN*]*—contains a test of independence for several random variables which is also based on* (*empirical*) *characteristic functions. In this paper, generalizations of the Hilbert–Schmidt independence criterion* (*HSIC*) *and of distance covariance are investigated. The formula* [(7), *BGN*] *is formally equivalent to our definition* (2.2) *of multivariance*; *although* (2.2) *is not stated in* [*BGN*], *it can be derived with some calculations using the Möbius transform* [(1), *BGN*] *of characteristic functions. In order to extend HSIC from finite measure kernels to stable measure kernels needed for distance covariance*, [*Theorem* 4.*i*, *BGN*] *establishes the* α*-stable version of our formula* (3.15). *The test statistic* [(8), (9), *BGN*] *and its consistency and asymptotics* [*Theorems* 2 *and* 4.*ii*, *BGN*] *correspond to special cases of our Theorems* 4.1(*a*) *and* 4.3. *The approach to test independence of n random variables of* [*BGN*] *is complementary to ours*: [*BGN*] *propose to combine the* p*-values from* $2^n - n - 1$ *evaluations of their test statistics, while we propose a global test using* total *multivariance. We would like to point out that our results were obtained independently.*

SUPPLEMENTARY MATERIAL

Supplement to "Distance multivariance: New dependence measures for random vectors" (DOI: 10.1214/18-AOS1764SUPP; .pdf). It contains the proofs of some of the main results as well as a few auxiliary statements.

REFERENCES

[1] BAKIROV, N. K. and SZÉKELY, G. J. (2010). Brownian covariance central limit theorem for stationary sequences. *Teor. Veroyatn. Primen.* **55** 462–488. All references to this paper are based on the english translation (Nail K. Bakirov and Gábor J. Székely. Brownian covariance and central limit theorem for stationary sequences. *Theory of Probability & Its Applications*, 55(3):371–394, 2011). MR2768533

[2] BARTHOLOMÄUS, J. (2018). Private communication, TU Dresden.

[3] BERG, C. and FORST, G. (1975). *Potential Theory on Locally Compact Abelian Groups*. *Ergebnisse der Mathematik und ihrer Grenzgebiete* **87**. Springer, New York. MR0481057

[4] BERRETT, T. B. and SAMWORTH, R. J. (2017). Nonparametric independence testing via mutual information. Available at arXiv:1711.06642v1.

[5] BERSCHNEIDER, G. and BÖTTCHER, B. (2018). On complex Gaussian random fields, Gaussian quadratic forms and sample distance multivariance. Available at arXiv:1808.07280v1.

[6] BÖTTCHER, B. (2017). Dependence structures—estimation and visualization using distance multivariance. Available at arXiv:1712.06532v1.

[7] BÖTTCHER, B. (2017). multivariance: Measuring multivariate dependence using distance multivariance. R package version 1.0.5.

[8] BÖTTCHER, B., KELLER-RESSEL, M. and SCHILLING, R. L. (2018). Detecting independence of random vectors: Generalized distance covariance and Gaussian covariance. *Mod. Stoch. Theory Appl.* **5** 353–383.

[9] BÖTTCHER, B., KELLER-RESSEL, M. and SCHILLING, R. L. (2019). Supplement to "Distance multivariance: New dependence measures for random vectors." DOI:10.1214/18-AOS1764SUPP.

[10] CHOW, Y. S. and TEICHER, H. (1997). *Probability Theory*: *Independence*, *Interchangeability*, *Martingales*, 3rd ed. *Springer Texts in Statistics*. Springer, New York. MR1476912

[11] COMON, P. (1994). Independent component analysis, a new concept? *Signal Process.* **36** 287–314.

[12] COPE, L. (2009). Discussion of: Brownian distance covariance [MR2752127]. *Ann. Appl. Stat.* **3** 1279–1281. MR2752130

[13] CSÖRGŐ, S. (1985). Testing for independence by the empirical characteristic function. *J. Multivariate Anal.* **16** 290–299. MR0793494

[14] FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications*. *Vol. II*. 2nd ed. Wiley, New York. MR0270403

[15] GRETTON, A., BOUSQUET, O., SMOLA, A. and SCHÖLKOPF, B. (2005). Measuring statistical dependence with Hilbert–Schmidt norms. In *Algorithmic Learning Theory*. *Lecture Notes in Computer Science* **3734** 63–77. Springer, Berlin. MR2255909

[16] JACOB, N. (2001). *Pseudo Differential Operators and Markov Processes*. *Vol. I*: *Fourier Analysis and Semigroups*. Imperial College Press, London. MR1873235

[17] JIN, Z. and MATTESON, D. S. (2018). Generalizing distance covariance to measure and test multivariate mutual dependence via complete and incomplete V-statistics. *J. Multivariate Anal.* **168** 304–322. Available at arXiv:1709.02532. All references to this paper are based on the arxiv version 1709.02532v5 which differs from the published version. MR3858367

[18] PFISTER, N., BÜHLMANN, P., SCHÖLKOPF, B. and PETERS, J. (2018). Kernel-based tests for joint independence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 5–31. MR3744710

[19] SATO, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. *Cambridge Studies in Advanced Mathematics* **68**. Cambridge Univ. Press, Cambridge. Translated from the 1990 Japanese original, revised by the author. MR1739520

[20] SCHILLING, R. L. (2017). *Measures*, *Integrals and Martingales*, 2nd ed. Cambridge Univ. Press, Cambridge. MR3644418

[21] SEJDINOVIC, D., SRIPERUMBUDUR, B., GRETTON, A. and FUKUMIZU, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Statist.* **41** 2263–2291. MR3127866

[22] SZÉKELY, G. J. and BAKIROV, N. K. (2003). Extremal probabilities for Gaussian quadratic forms. *Probab. Theory Related Fields* **126** 184–202. MR1990053

[23] SZÉKELY, G. J. and RIZZO, M. L. (2009). Brownian distance covariance. *Ann. Appl. Stat.* **3** 1236–1265. MR2752127

[24] SZÉKELY, G. J. and RIZZO, M. L. (2009). Rejoinder: Brownian distance covariance. *Ann. Appl. Stat.* **3** 1303–1308. MR2752135

[25] SZÉKELY, G. J., RIZZO, M. L. and BAKIROV, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35** 2769–2794. MR2382665

[26] YAO, S., ZHANG, X. and SHAO, X. (2018). Testing mutual independence in high dimension via distance covariance. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 455–480. MR3798874

B. BÖTTCHER
M. KELLER-RESSEL
R. L. SCHILLING
FAKULTÄT MATHEMATIK
INSTITUT FÜR MATHEMATISCHE STOCHASTIK
TU DRESDEN
01062 DRESDEN
GERMANY
E-MAIL: bjoern.boettcher@tu-dresden.de
martin.keller-ressel@tu-dresden.de
rene.schilling@tu-dresden.de

SUPPLEMENT TO "DISTANCE MULTIVARIANCE: NEW DEPENDENCE MEASURES FOR RANDOM VECTORS"

By Björn Böttcher, Martin Keller-Ressel* and René L. Schilling

TU Dresden

Proofs and auxiliary results. Here we collect supplementary material to [BKRS19]. It contains the proofs of some of the main results as well as a few additional statements: Lemma S.1 discusses the moment conditions introduced in Definition 2.3 and Lemma S.2 analyses the estimator which is required for the proof of the main convergence result (Theorem 4.5).

Unless otherwise mentioned, all numbered references refer to [BKRS19].

S.1. *Proofs and auxiliary results for Section 2.*

Proof of Lemma 3.2. For arbitrary $a_i, b_i \in \mathbb{C}$, $i = 1, \dots, n$, we have

$$\prod_{i=1}^{n}(a_i - b_i) = \sum_{S\subset\{1,\dots,n\}} \left(\prod_{i\in S} a_i\right)\left(\prod_{i\in S^c} b_i\right)(-1)^{|S^c|}, \tag{S.1}$$

where $|S|$ denotes the cardinality of S and $S^c := \{1,\dots,n\}\setminus S$. Thus,

$$\begin{aligned}
\mathbb{E}\left(\prod_{i=1}^{n}(Z_i - \mathbb{E}Z_i)\right) &= \mathbb{E}\left[\sum_{S\subset\{1,\dots,n\}} \left(\prod_{i\in S} Z_i\right)\left(\prod_{i\in S^c} \mathbb{E}(Z_i)\right)(-1)^{|S^c|}\right] \\
&= \mathbb{E}\left(\prod_{i=1}^{n} Z_i\right) + \sum_{\substack{S\subset\{1,\dots,n\}\\ |S|\le n-1}} \mathbb{E}\left(\prod_{i\in S} Z_i\right)\left(\prod_{i\in S^c} \mathbb{E}\left(Z_i\right)\right)(-1)^{|S^c|} \\
&= \mathbb{E}\left(\prod_{i=1}^{n} Z_i\right) + \sum_{\substack{S\subset\{1,\dots,n\}\\ |S|\le n-1}} \left(\prod_{i=1}^{n} \mathbb{E}\left(Z_i\right)\right)(-1)^{|S^c|} \\
&= \mathbb{E}\left(\prod_{i=1}^{n} Z_i\right) - \prod_{i=1}^{n} \mathbb{E}(Z_i);
\end{aligned}$$

$(n-1)$-independence is used in the penultimate line. □

*Supported by the German Research Foundation (DFG) under grant ZUK 64 and KE 1736/1-1

LEMMA S.1. *The moment conditions in Definition 2.3 are ordered from weak to strong, i.e. c) implies b) and b) implies a). In particular, the estimate*

$$\mathbb{E}\left(\prod_{i=1}^n \psi_i(X_{k_i,i} - X'_{l_i,i})\right) \le 4^n \prod_{i=1}^n \left(\mathbb{E}\psi_i^{p_i}(X_i)\right)^{1/p_i} \tag{S.2}$$

holds for all $k_i, l_i \in \{0,1\}, i = 1, \ldots, n$ *and all* $p_i \in [1,\infty)$ *with* $\sum_{i=1}^n p_i^{-1} = 1$.

PROOF. The implication from c) to b) follows from the fact that every continuous negative definite function is quadratically bounded, i.e. $|\psi(x)| \le C(1+x^2)$ for some $C > 0$, see [Jac01, Lem. 3.6.22].

The other implication follows directly from (S.2). To show (S.2), note that the generalized Hölder inequality for n-fold products (cf. [Sch17, p. 133, Pr. 13.5]) gives

$$\mathbb{E}\left(\prod_{i=1}^n \psi_i(X_{k_i,i} - X'_{l_i,i})\right) \le \prod_{i=1}^n \left(\mathbb{E}\psi_i^{p_i}(X_{k_i,i} - X'_{l_i,i})\right)^{1/p_i}.$$

Using an inequality for continuous negative definite functions (cf. [BKRS18, Eq. (2.5)], see also [Jac01, Lem. 3.6.21]) and the Minkowski inequality for the L^{p_i}-norm yields the bound

$$\begin{aligned}\left(\mathbb{E}\psi_i^{p_i}(X_{k_i,i} - X'_{l_i,i})\right)^{1/p_i} &\le 2\left(\mathbb{E}\left[\psi_i(X_{k_i,i}) + \psi_i(X'_{l_i,i})\right]^{p_i}\right)^{1/p_i} \\ &\le 4\left(\mathbb{E}\psi_i^{p_i}(X_i)\right)^{1/p_i}.\end{aligned}$$

□

S.2. *Proofs and auxiliary results for Section 3.*

PROOF OF PROPOSITION 3.9. Using (2.6), we can rewrite M_ρ in the following way:

$$\begin{aligned}M_\rho^2 &= \int \left|\mathbb{E}\left[\prod_{i=1}^n \left(\mathrm{e}^{\mathrm{i}X_i\cdot t_i} - f_{X_i}(t_i)\right)\right]\right|^2 \rho(\mathrm{d}t) \\ &= \int \left|\mathbb{E}\left[\prod_{i=1}^n \left(\mathrm{e}^{\mathrm{i}X_{1,i}\cdot t_i} - \mathrm{e}^{\mathrm{i}X_{0,i}\cdot t_i}\right)\right]\right|^2 \rho(\mathrm{d}t) \\ &= \int \mathbb{E}\left[\prod_{i=1}^n \left(\mathrm{e}^{\mathrm{i}X_{1,i}\cdot t_i} - \mathrm{e}^{\mathrm{i}X_{0,i}\cdot t_i}\right)\left(\mathrm{e}^{-\mathrm{i}X'_{1,i}\cdot t_i} - \mathrm{e}^{-\mathrm{i}X'_{0,i}\cdot t_i}\right)\right] \rho(\mathrm{d}t) \\ &= \int \mathbb{E}\left[\sum_{k,l\in\{0,1\}^n} (-1)^{\sum_{j=1}^n (k_j+l_j)} \prod_{i=1}^n \mathrm{e}^{\mathrm{i}(X_{k_i,i} - X'_{l_i,i})\cdot t_i}\right] \rho(\mathrm{d}t)\end{aligned} \tag{S.3}$$

and the ultimate line already gives (3.13). By (3.9),

$$M_\rho^2(X_1, X_2, \ldots, X_n) = \frac{1}{2}\left(M_\rho^2(X_1, X_2, \ldots, X_n) + M_\rho^2(-X_1, X_2, \ldots, X_n)\right).$$

Applying this to (S.3) shows that the imaginary part of the complex exponential cancels for $i = 1$. Repeated applications to $i = 2, \ldots, n$ removes the other imaginary terms, and we obtain

$$M_\rho^2 = \int \mathbb{E}\left[\sum_{k,l\in\{0,1\}^n} \operatorname{sgn}(k,l) \prod_{i=1}^n \cos((X_{k_i,i} - X'_{l_i,i}) \cdot t_i)\right] \rho(\mathrm{d}t). \tag{S.4}$$

It remains to show that (S.4) is equal to (3.14). For this, we note that the product appearing in (3.14) is of the form

$$\prod_{i=1}^n \left[\cos((X_{k_i,i} - X'_{l_i,i}) \cdot t_i) - 1\right] = \prod_{i=1}^n \cos((X_{k_i,i} - X'_{l_i,i}) \cdot t_i) + \prod_{i=1}^n c(k_i, l_i)$$

where $c(k_i, l_i)$ is either $\cos((X_{k_i,i} - X'_{l_i,i}) \cdot t_i)$ or -1 and at least one factor in the second product is -1; if, say, $c(k_m, l_m) = -1$ for some $m \in \{1, \ldots, n\}$, we get with $k' = (k_1, \ldots, k_{m-1}, k_{m+1}, \ldots, k_n)$, $l' = (l_1, \ldots, l_{m-1}, l_{m+1}, \ldots, l_n)$,

$$\begin{aligned}&\sum_{k,l\in\{0,1\}^n} \operatorname{sgn}(k,l) \prod_{i=1}^n c(k_i, l_i)\\ &\quad = -\sum_{k_m,l_m\in\{0,1\}} (-1)^{k_m+l_m} \sum_{k',l'\in\{0,1\}^{n-1}} \operatorname{sgn}(k', l') \prod_{i\neq m} c(k_i, l_i).\end{aligned}$$

This expression is 0 since the inner sum does not depend on k_m, l_m and appears exactly four times, twice with positive and twice with negative sign. This shows that (3.14) is equal to (S.4).

Finally, by Lemma S.1, all moment conditions in Definition 2.3 imply the mixed moment condition 2.3.a), $\mathbb{E}\left(\prod_{i=1}^n \psi_i(X_{k_i,i} - X'_{l_i,i})\right) < \infty$ for all $k, l \in \{0,1\}^n$. Under this condition, Fubini's theorem together with the tower property for conditional expectations and the independence properties (2.6) of $\boldsymbol{X}_0, \boldsymbol{X}'_0$ yield

$$\begin{aligned}M_\rho^2 &= \mathbb{E}\left(\sum_{k,l\in\{0,1\}^n} (-1)^{\sum_{j=1}^n (k_j+l_j)} \prod_{i=1}^n (-\psi_i(X_{k_i,i} - X'_{l_i,i}))\right)\\ &= \mathbb{E}\left(\prod_{i=1}^n \Psi_{i,0,1}\right) = \mathbb{E}\left(\mathbb{E}\left(\prod_{i=1}^n \Psi_{i,0,1} \,\middle|\, \boldsymbol{X}_1, \boldsymbol{X}'_1\right)\right) = \mathbb{E}\left(\prod_{i=1}^n \overline{\overline{\Psi}}_i\right)\end{aligned} \tag{S.5}$$

where

$$\begin{aligned}\Psi_{i,0,1} &:= -\psi_i(X_{1,i} - X'_{1,i}) + \psi_i(X_{1,i} - X'_{0,i}) \\ &\quad + \psi_i(X_{0,i} - X'_{1,i}) - \psi_i(X_{0,i} - X'_{0,i}), \\ \overline{\overline{\Psi}}_i &:= -\psi_i(X_i - X'_i) + \mathbb{E}(\psi_i(X_i - X'_i) \mid X_i) \\ &\quad + \mathbb{E}(\psi_i(X_i - X'_i) \mid X'_i) - \mathbb{E}\psi_i(X_i - X'_i). \qquad \square\end{aligned}$$

S.3. *Proofs and auxiliary results for Section 4.*

LEMMA S.2. *Let* $\boldsymbol{X}^{(l)} := (X_1^{(l)}, \dots, X_n^{(l)})$ *be independent and identical distributed copies of* $\boldsymbol{X} = (X_1, \dots, X_n)$ *and set*

$$Z_N(t) := \frac{1}{N} \sum_{l=1}^{N} \prod_{i=1}^{n} \left(\mathrm{e}^{\mathrm{i} X_i^{(l)} \cdot t_i} - \frac{1}{N} \sum_{k=1}^{N} \mathrm{e}^{\mathrm{i} X_i^{(k)} \cdot t_i} \right). \tag{S.6}$$

Then

$$^{N}M_\rho(\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)}) = \|Z_N(\bullet)\|_{L^2(\rho)}. \tag{S.7}$$

If $X_1, \dots, X_n$ *are independent, then*

$$\mathbb{E} Z_N(t) = 0, \tag{S.8}$$

$$\mathbb{E}(Z_N(t)\overline{Z_N(t')}) = \frac{1}{N} \cdot C_N \cdot \prod_{i=1}^{n} \left[f_{X_i}(t_i - t'_i) - f_{X_i}(t_i)\overline{f_{X_i}(t'_i)} \right], \tag{S.9}$$

$$\mathbb{E}\left(\left| \sqrt{N} Z_N(t) \right|^2 \right) = C_N \cdot \prod_{i=1}^{n} \left(1 - |f_{X_i}(t_i)|^2 \right), \tag{S.10}$$

with constant $C_N := \frac{(N-1)^n + (-1)^n (N-1)}{N^n}$.

PROOF. The equality (S.7) follows by inserting the empirical characteristic function into the representation (2.4) of distance multivariance.

Assume that the random variables $X_1, \dots, X_n$ are independent. We obtain

$$\mathbb{E}\left(\mathrm{e}^{\mathrm{i} X_i^{(l)} \cdot t_i} - \frac{1}{N} \sum_{k=1}^{N} \mathrm{e}^{\mathrm{i} X_i^{(k)} \cdot t_i} \right) = 0, \quad i = 1, \dots, n,\ l = 1, \dots, N,$$

hence, $\mathbb{E} Z_N(t) = 0$. Next, consider

$$\begin{aligned}\mathbb{E}(Z_N(t)\overline{Z_N(t')}) = \frac{1}{N^2} \sum_{l,l'=1}^{N} \mathbb{E}\Bigg[&\prod_{i=1}^{n} \left(\mathrm{e}^{\mathrm{i} X_i^{(l)} \cdot t_i} - \frac{1}{N} \sum_{k=1}^{N} \mathrm{e}^{\mathrm{i} X_i^{(k)} \cdot t_i} \right) \times \\ &\times \prod_{i'=1}^{n} \left(\mathrm{e}^{-\mathrm{i} X_{i'}^{(l')} \cdot t'_{i'}} - \frac{1}{N} \sum_{k'=1}^{N} \mathrm{e}^{-\mathrm{i} X_{i'}^{(k')} \cdot t'_{i'}} \right) \Bigg].\end{aligned} \tag{S.11}$$

The independence of X_i, X_j for $i \neq j$ implies

$$\mathbb{E}\left[\prod_{i=1}^{n}\left(\mathrm{e}^{\mathrm{i}X_i^{(l)}\cdot t_i} - \frac{1}{N}\sum_{k=1}^{N}\mathrm{e}^{\mathrm{i}X_i^{(k)}\cdot t_i}\right)\cdot\prod_{i'=1}^{n}\left(\mathrm{e}^{-\mathrm{i}X_{i'}^{(l')}\cdot t'_{i'}} - \frac{1}{N}\sum_{k'=1}^{N}\mathrm{e}^{-\mathrm{i}X_{i'}^{(k')}\cdot t'_{i'}}\right)\right]$$
$$= \prod_{i=1}^{n}\mathbb{E}\left[\left(\mathrm{e}^{\mathrm{i}X_i^{(l)}\cdot t_i} - \frac{1}{N}\sum_{k=1}^{N}\mathrm{e}^{\mathrm{i}X_i^{(k)}\cdot t_i}\right)\cdot\left(\mathrm{e}^{-\mathrm{i}X_i^{(l')}\cdot t'_i} - \frac{1}{N}\sum_{k'=1}^{N}\mathrm{e}^{-\mathrm{i}X_i^{(k')}\cdot t'_i}\right)\right]$$

and each factor simplifies to

$$\begin{aligned}
&\mathbb{E}\left[\left(\mathrm{e}^{\mathrm{i}X_i^{(l)}\cdot t_i} - \frac{1}{N}\sum_{k=1}^{N}\mathrm{e}^{\mathrm{i}X_i^{(k)}\cdot t_i}\right)\cdot\left(\mathrm{e}^{-\mathrm{i}X_i^{(l')}\cdot t'_i} - \frac{1}{N}\sum_{k'=1}^{N}\mathrm{e}^{-\mathrm{i}X_i^{(k')}\cdot t'_i}\right)\right]\\
&= \mathbb{E}\mathrm{e}^{\mathrm{i}X_i^{(l)}\cdot t_i - \mathrm{i}X_i^{(l')}\cdot t'_i} - 2\frac{N-1}{N}f_{X_i}(t_i)\overline{f_{X_i}(t'_i)} - \frac{2}{N}f_{X_i}(t_i - t'_i)\\
&\qquad + \frac{N^2-N}{N^2}f_{X_i}(t_i)\overline{f_{X_i}(t'_i)} + \frac{N}{N^2}f_{X_i}(t_i - t'_i)\\
&= \mathbb{E}\mathrm{e}^{\mathrm{i}X_i^{(l)}\cdot t_i - \mathrm{i}X_i^{(l')}\cdot t'_i} - \frac{N-1}{N}f_{X_i}(t_i)\overline{f_{X_i}(t'_i)} - \frac{1}{N}f_{X_i}(t_i - t'_i).
\end{aligned}$$

Thus, splitting the sum in (S.11) into $l = l'$ and $l \neq l'$ yields

$$\begin{aligned}
&\mathbb{E}(Z_N(t)\overline{Z_N(t')})\\
&= \frac{1}{N^2}\sum_{l,l'=1}^{N}\prod_{i=1}^{n}\left[\mathbb{E}\mathrm{e}^{\mathrm{i}X_i^{(l)}\cdot t_i - \mathrm{i}X_i^{(l')}\cdot t'_i} - \frac{N-1}{N}f_{X_i}(t_i)\overline{f_{X_i}(t'_i)} - \frac{1}{N}f_{X_i}(t_i - t'_i)\right]\\
&= \frac{N}{N^2}\prod_{i=1}^{n}\left[-\frac{N-1}{N}f_{X_i}(t_i)\overline{f_{X_i}(t'_i)} - \left(\frac{1}{N} - 1\right)f_{X_i}(t_i - t'_i)\right]\\
&\qquad + \frac{N^2-N}{N^2}\prod_{i=1}^{n}\left[\left(-\frac{N-1}{N} + 1\right)f_{X_i}(t_i)\overline{f_{X_i}(t'_i)} - \frac{1}{N}f_{X_i}(t_i - t'_i)\right]\\
&= \left(\frac{1}{N}\left(\frac{N-1}{N}\right)^n + \frac{N-1}{N}\left(\frac{-1}{N}\right)^n\right)\prod_{i=1}^{n}\left[f_{X_i}(t_i - t'_i) - f_{X_i}(t_i)\overline{f_{X_i}(t'_i)}\right]\\
&= \frac{(N-1)^n + (-1)^n(N-1)}{N^{n+1}}\prod_{i=1}^{n}\left[f_{X_i}(t_i - t'_i) - f_{X_i}(t_i)\overline{f_{X_i}(t'_i)}\right].
\end{aligned}$$

For $t' = t$ this reduces to

$$\mathbb{E}\left(\left|\sqrt{N}Z_N(t)\right|^2\right) = N\cdot\frac{(N-1)^n + (-1)^n(N-1)}{N^{n+1}}\prod_{i=1}^{n}\left(1 - |f_{X_i}(t_i)|^2\right). \quad \square$$

Proof of Theorem 4.5. We start with part b), which is a simple consequence of the strong consistency of ${}^N\!M_\rho$. Indeed, by Theorem 4.3 we have

${}^N M_\rho \to M_\rho$ a.s., and from Theorem 3.4 we know that $M_\rho > 0$ under the conditions of b), such that (4.8) follows.

For part a), let $Z_N(t)$ be defined as in (S.6). Then ${}^N M_\rho(\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)}) = \|Z_N(.)\|_{L^2(\rho)}$ by Lemma S.2.

If $\sqrt{N} Z_N$ converges in distribution to a Gaussian process then, by Lemma S.2, this process is centred and has the covariance structure (4.7), i.e. it is distributed as $\mathbb{G}$. In order to show convergence, we introduce the following notation. Denote by $F_{\boldsymbol{X}}$ the distribution function of $\boldsymbol{X}$ and by ${}^N F_{\boldsymbol{X}}$ the empirical distribution function of the iid sequence $(\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)})$. For a subset $S \subset \{1, \dots, n\}$ we write $t_S := (t_i)_{i \in S}$ and denote the corresponding empirical characteristic function by

$$
{}^N f_S(t_S) := \frac{1}{N} \sum_{j=1}^{N} \exp\left(\mathrm{i} \sum_{i \in S} X_i^{(j)} \cdot t_i \right) = \int \mathrm{e}^{\mathrm{i} x_S \cdot t_S} \, \mathrm{d}({}^N F_{\boldsymbol{X}}(x)).
$$

If $S = \{i\}$ is a singleton, we write ${}^N f_i := {}^N f_{\{i\}}$. By [Csö81, Thm 3.1, p. 208] the log-moment condition is sufficient for the convergence

$$
\sqrt{N}\, ({}^N f(t) - f(t)) = \int \mathrm{e}^{\mathrm{i} x \cdot t} \, \mathrm{d}(\sqrt{N} ({}^N F_{\boldsymbol{X}}(x) - F_{\boldsymbol{X}}(x))) \xrightarrow[N \to \infty]{d} \int \mathrm{e}^{\mathrm{i} x \cdot t} \, \mathrm{d}B(x), \tag{S.12}
$$

where B is a Brownian bridge indexed by $\mathbb{R}^d$ (cf. [Csö81, Eq. (3.2)]) and the distributional convergence is uniform (in t) on compact subsets of $\mathbb{R}^d$. Next, we rewrite Z_N from (S.6) as

$$
Z_N(t) = \sum_{S \subset \{1, \dots, n\}} (-1)^{n - |S|} \left({}^N f_S(t_S) \cdot \prod_{j \in S^c} {}^N f_j(t_j) \right). \tag{S.13}
$$

In addition, we have the simple identity, cf. (S.1),

$$
\prod_{j=1}^{n} (f_j(t_j) - {}^N f_j(t_j)) = \sum_{S \subset \{1, \dots, n\}} (-1)^{n - |S|} \left(\prod_{j \in S} f_j(t_j) \cdot \prod_{j \in S^c} {}^N f_j(t_j) \right). \tag{S.14}
$$

Subtracting (S.14) from (S.13) and rearranging the resulting equation yields

$$
\begin{aligned}
\sqrt{N} Z_N(t) = \sum_{S \subset \{1, \dots, n\}} & (-1)^{n - |S|} \sqrt{N} ({}^N f_S(t_S) - f_S(t_S)) \cdot \prod_{j \in S^c} {}^N f_j(t_j) \\
& + \sqrt{N} \prod_{j=1}^{n} (f_j(t_j) - {}^N f_j(t_j)).
\end{aligned}
$$

By (S.12), we have that

$$\sqrt{N}\left({}^N f_S(t_S) - f_S(t_S)\right) \xrightarrow[N\to\infty]{d} \int \mathrm{e}^{\mathrm{i}x_S\cdot t_S}\,\mathrm{d}B(x).$$

By the Glivenko–Cantelli theorem the limit $\lim_{N\to\infty} {}^N f_j(t_j) = f_j(t_j)$ exists uniformly in t_j for all $j = 1,\dots,n$, and thus

$$\begin{aligned}&\sqrt{N}\prod_{j=1}^{n}\left({}^N f_j(t_j) - f_j(t_j)\right)\\ &= \sqrt{N}\left({}^N f_1(t_1) - f_1(t_1)\right)\cdot\prod_{j=2}^{n}\left({}^N f_j(t_j) - f_j(t_j)\right) \xrightarrow[N\to\infty]{} 0.\end{aligned}$$

Together with (S.13) this yields the convergence

$$\text{(S.15)}\qquad \sqrt{N}Z_N(t) \xrightarrow[N\to\infty]{d} \sum_{S\subset\{1,\dots,n\}} (-1)^{n-|S|}\int \mathrm{e}^{\mathrm{i}x_S\cdot t_S}\,\mathrm{d}B(x)\cdot\prod_{j\in S^c} f_j(t_j),$$

which takes place uniformly on compacts. The right hand side is a complex-valued Gaussian process indexed by $\mathbb{R}^d$; denoting this process by $\mathbb{G}$, we have thus shown that for each $T > 0$,

$$\text{(S.16)}\qquad \sqrt{N}Z_N \xrightarrow[N\to\infty]{d} \mathbb{G} \quad\text{on}\quad \mathcal{C}_T := (C(B_T^d), \|\bullet\|_{B_T^d}),$$

where $B_T^d := B_T^d(0) := \{x\in\mathbb{R}^d : |x| < T\}$ and $\|f\|_{B_T^d} := \sup_{x\in B_T^d}|f(x)|$. To obtain (4.6), it remains to show that also the $L^2(\rho)$-norms of both sides of (S.16) converge, and that T can be sent to infinity. To this end, we apply a truncation argument.

Set

$$\text{(S.17)}\qquad \rho_{i,\epsilon}(A) := \rho_i(A\cap(B_{1/\epsilon}^{d_i}\setminus B_\epsilon^{d_i})) \quad\text{and}\quad \rho_i^\epsilon := \rho_i - \rho_{i,\epsilon},$$

and note that the $\rho_{i,\epsilon}$ are finite measures for each $\epsilon > 0$, by (2.1). In addition, we define $\rho_\epsilon = \bigotimes_{i=1}^n \rho_{i,\epsilon}$ as well as $\rho^\epsilon = \bigotimes_{i=1}^n \rho_i^\epsilon$ and introduce, for this proof, the shorthand notation $\|\bullet\|_{\rho_\epsilon} = \|\bullet\|_{L^2(\rho_\epsilon)}$. Note that $|x_i| \le 1/\epsilon$, $x_i\in\mathbb{R}^{d_i}$, for all $i = 1,\dots,n$ implies $|x| \le \sqrt{n}/\epsilon$, $x = (x_1,\dots,x_n)\in\mathbb{R}^d$, and hence we have

(S.18)

$$\left|\|h\|_{\rho_\epsilon} - \|h'\|_{\rho_\epsilon}\right|^2 \le \|h - h'\|_{\rho_\epsilon} \le \sup_{|x|\le\sqrt{n}/\epsilon}\left|h(x) - h'(x)\right|^2\cdot\prod_{i=1}^{n}\rho_{i,\epsilon}(\mathbb{R}^{d_i}),$$

which shows that $\|\bullet\|^2_{\rho_\epsilon}$ is continuous on $\mathcal{C}_T$ for any $T \geq \sqrt{n}/\epsilon$. Thus, the continuous mapping theorem implies that for any $\epsilon > 0$

$$\|\sqrt{N} Z_N\|^2_{\rho_\epsilon} \xrightarrow[N\to\infty]{d} \|\mathbb{G}\|^2_{\rho_\epsilon}. \tag{S.19}$$

By the portmanteau theorem, the convergence (4.6) is equivalent to the statement

$$\lim_{N\to\infty} \mathbb{E}\left(h(N \cdot {}^N\!M^2) - h(\|\mathbb{G}\|^2_\rho)\right) = 0$$

for all bounded Lipschitz continuous functions $h : \mathbb{R} \to \mathbb{R}$. Denoting the Lipschitz constant of h by L_h, we see

$$\begin{aligned}\left|\mathbb{E}h(N \cdot {}^N\!M^2) - \mathbb{E}h(\|\mathbb{G}\|^2_\rho)\right| \leq L_h \mathbb{E}\left|N \cdot {}^N\!M^2 - \|\sqrt{N} Z_N\|^2_{\rho_\epsilon}\right| \\ + \left|\mathbb{E}h\left(\|\sqrt{N} Z_N\|^2_{\rho_\epsilon}\right) - \mathbb{E}h\left(\|\mathbb{G}\|^2_{\rho_\epsilon}\right)\right| \\ + L_h \mathbb{E}\left|\|\mathbb{G}\|^2_{\rho_\epsilon} - \|\mathbb{G}\|^2_\rho\right|.\end{aligned} \tag{S.20}$$

The middle term tends to zero as $N \to \infty$, by (S.19). To estimate the other terms, define μ^ϵ to be the measure given by

$$(\rho_1^\epsilon \otimes \rho_2 \otimes \ldots \otimes \rho_n) + (\rho_1 \otimes \rho_2^\epsilon \otimes \ldots \otimes \rho_n) + \cdots + (\rho_1 \otimes \ldots \otimes \rho_{n-1} \otimes \rho_n^\epsilon).$$

For the first term on the right hand side of (S.20) we get the bound

$$\mathbb{E}\left|N \cdot {}^N\!M^2 - \|\sqrt{N} Z_N\|^2_{\rho_\epsilon}\right| = \mathbb{E}\left|\|\sqrt{N} Z^N\|^2_\rho - \|\sqrt{N} Z_N\|^2_{\rho_\epsilon}\right| \leq \mathbb{E}\|\sqrt{N} Z_N\|^2_{\mu^\epsilon}.$$

Using (S.10) we see with $C_N := [(N-1)^n + (-1)^n(N-1)]/N^n \leq 1$

$$\left\|\mathbb{E}\left(|\sqrt{N} Z_N|^2\right)\right\|^2_{\mu^\epsilon} = C_N \sum_{k=1}^n \left[\|1 - |f_{X_k}|^2\|^2_{\rho_k^\epsilon} \prod_{\substack{i=1\\ i\neq k}}^n \|1 - |f_{X_i}|^2\|^2_{\rho_i}\right], \tag{S.21}$$

and this expression converges to 0 as $\epsilon \to 0$. This follows from dominated convergence, since

$$\sum_{k=1}^n \left[\|1 - |f_{X_k}|^2\|^2_{\rho_k^\epsilon} \prod_{\substack{i=1\\ i\neq k}}^n \|1 - |f_{X_i}|^2\|^2_{\rho_i}\right] \leq n \prod_{i=1}^n \mathbb{E}\psi_i(X_i - X_i') < \infty.$$

The last term in (S.20) can be estimated in a similar way. We have

$$\|\mathbb{G}\|^2_\rho - \|\mathbb{G}\|^2_{\rho_\epsilon} \leq \|\mathbb{G}\|^2_{\mu^\epsilon} \xrightarrow[\epsilon\to 0]{} 0 \quad \text{a.s.} \tag{S.22}$$

by dominated convergence, since $\lim_{\epsilon\to 0}\int g_i\,d\rho_i^\epsilon = 0$ for integrable g_i and

$$\begin{aligned}\mathbb{E}(\|\mathbb{G}\|^2_{\mu^\epsilon}) \le n\mathbb{E}(\|\mathbb{G}\|^2_{\rho}) &= n\prod_{i=1}^n \|1-|f_{X_i}|^2\|^2_{\rho_i} \\ &= n\prod_{i=1}^n \mathbb{E}\psi_i(X_i - X_i') < \infty.\end{aligned} \tag{S.23}$$

Together with (S.20) this shows the convergence result (4.6) and completes the proof. □

REFERENCES

[BKRS18] Björn Böttcher, Martin Keller-Ressel, and René L. Schilling. Detecting independence of random vectors: Generalized distance covariance and Gaussian covariance. *Modern Stochastics: Theory and Applications*, 5(3):353–383, 2018.

[BKRS19] Björn Böttcher, Martin Keller-Ressel, and René L. Schilling. Distance multivariance: New dependence measures for random vectors. *The Annals of Statistics*, 47(5), pp. 36, 2019.

[Csö81] Sándor Csörgő. Multivariate empirical characteristic functions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 55(2):203–229, 1981.

[Jac01] Niels Jacob. *Pseudo-Differential Operators and Markov Processes I. Fourier Analysis and Semigroups.* Imperial College Press, London, 2001.

[Sch17] René. L. Schilling. *Measures, Integrals and Martingales* (*second edition*). Cambridge University Press, 2017.

TU Dresden
Fakultät Mathematik
Institut für Mathematische Stochastik
01062 Dresden, Germany
E-mail: bjoern.boettcher@tu-dresden.de
E-mail: martin.keller-ressel@tu-dresden.de
E-mail: rene.schilling@tu-dresden.de