arXiv:submit/2078314 [math.PR] 21 Nov 2017

# DETECTING INDEPENDENCE OF RANDOM VECTORS II. DISTANCE MULTIVARIANCE AND GAUSSIAN MULTIVARIANCE

BJÖRN BÖTTCHER, MARTIN KELLER-RESSEL, AND RENÉ L. SCHILLING

Abstract. 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. They extend distance covariance (introduced by Szekely, Rizzo and Bakirov) and generalized distance covariance (introduced in part I) 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. Based on our theoretical results, we present a test for independence of multiple random vectors which is consistent against all alternatives.

## Contents



## 1. Introduction

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

2010 *Mathematics Subject Classification.* 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.

We are grateful to our colleague, Ulrich Brehm, for insightful discussions on (elementary) symmetric polynomials. MKR acknowledges funding from the German Research Foundation (DFG) under grant ZUK 64 and KE 1736/1-1.

Székely, Rizzo and Bakirov [SRB07, SR09a] and its generalizations presented in the companion paper [BKRS17] to the present paper. 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 $X_1,\dots,X_n$ taking 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. To our knowledge, the problem of generalizing distance covariance to more than two random variables has so far been addressed only in a short paragraph '*How to (not) extend [distance covariance]* $\mathcal{V}(X,Y)$ *to more than two random variables*' by Bakirov and Székely [BS11]. Our approach is different from the approach suggested in [BS11]; it is, in fact, closer to the two approaches that were advised against in [BS11]. We will discuss and compare these approaches in greater detail at the end of Section 2, once the necessary concepts have been introduced.

Similar to distance covariance and its generalizations in [BKRS17], 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 and Section 2.2 of [BKRS17] for a general discussion. There are, however, further – up to moment conditions equivalent – definitions of distance multivariance. In particular, it 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 [SR09a, Def. 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

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

where the $A_i$ are doubly centred distance matrices of the sample points of $X_i$, i.e. $A_i := CB_iC$ 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

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

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. Section 2.1 of [BKRS17]. 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 [SR09a, Def. 3] is recovered.

Finally, we show in Theorems 4.6 and 4.11 asymptotic properties of sample distance multivariance as $N$ tends to infinity; these results are multivariate analogues of those in [SR09a, Thm. 5]. Based on these results, we formulate two new distribution-free tests for the joint independence of $n$ random variables in Section 4.5. 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 accurately between pairwise independence and higher-order dependence of random variables. The example also illustrates the practical validity of the two tests that are proposed.

For the immediate use of distance multivariance in applications all necessary functions are provided in the R package `multivariance` [Böt17].

## 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 $d = d_1 + \dots + d_n$. The characteristic function of $X_i$ is denoted by

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

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* $\rho_i$, i.e. Borel measures $\rho_i$ defined on $\mathbb{R}^{d_i} \setminus \{0\}$ such that

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

Note that the measures $\rho_i$ need not be finite. Such measures appear in the Lévy–Khintchine representation of infinitely divisible distributions, see the discussion in the companion paper [BKRS17, Sec. 2.1]. Throughout this paper we assume that $\rho_i$, $i = 1, \dots, n$ are symmetric Lévy measures with full topological support (cf. [BKRS17, Def. 2.3]), and we set $\rho := \rho_{d_1} \otimes \dots \otimes \rho_{d_n}$. To keep notation simple, we write $\int \cdots \rho_i(dt_i)$ and $\int_{\mathbb{R}^{d_i}} \cdots \rho_i(dt_i)$ instead of the formally correct $\int_{\mathbb{R}^{d_i}\setminus\{0\}} \cdots \rho_i(dt_i)$.

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

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

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 [SRB07, Def. 1], see also the discussion in [BKRS17, Sec. 2.2].

b) Both $M_\rho$ 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 $\rho$-integral is positive, and so the integral is always well-defined in $[0, +\infty]$. Just as in the bivariate case, see [BKRS17, Thm. 3.7, Rem. 3.8], we need moment conditions on the random variables $X_i$ to guarantee finiteness of $M_\rho$ and $\overline{M}_\rho$, see Proposition 3.7 below.

c) At first sight, total distance multivariance seems to suffer from the 'curse of dimension', since the sum (2.3) extends over all subfamilies (comprising at least two members) of $(X_1, \dots, X_n)$, i.e. $2^n - 1 - n$ 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 $M_\rho$ and its computation requires only $\mathcal{O}(n)$ operations.

Each Lévy measure $\rho_i$ is associated with a unique real-valued *continuous negative definite function*

$$\psi_i(y_i) := \int_{\mathbb{R}^{d_i}} (1 - \cos(y t_i))\ \rho_i(\mathrm{d}t_i) \quad \text{for} \quad y_i \in \mathbb{R}^{d_i}, \tag{2.5}$$

see [BKRS17, Sec. 2.1] for details. The functions $\psi_i$ 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. [SR09a]), since $\psi_i$ yields well-known distance functions (or rather 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

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

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 $|x|_{d_i, p} := \left(\sum_{j=1}^{d_i} |x_j|^p\right)^{1/p}$, for $p \in (1, 2]$ can be represented in the form (2.5); see [BKRS17, 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 random 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

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

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

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

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

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

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

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

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

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

These moment conditions are ordered from weak to strong, cf. Lemma 3.6 below. Also note that b) and a) trivially hold (for any choice of $p_i$) if the functions $\psi_i$ are bounded.

Let us briefly compare our approach with the methods of Bakirov and Székely.

*Comparison with* [BS11]. As mentioned in the introduction, 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 [BS11]. In the notation of our paper they discuss for three random variables $X, Y, Z$ the following objects:

a) Gaussian Covariance $\mathcal{G}(X,Y,Z) = \mathbb{E}\left(X^G X'^G Y^G Y'^G Z^G Z'^G\right)$ (cf. Section 3.3) where $G$ is a Brownian motion. This approach is dismissed in [BS11] since it does not characterize the independence of $X, Y, Z$.

b) The quantity

$$\int_{\mathbb{R}^d} \left| \mathbb{E}\left[e^{i(X\cdot t_1 + Y\cdot t_2 + Z\cdot t_3)}\right] - f_X(t_1) f_Y(t_2) f_Z(t_3) \right|^2 \rho(dt_1, dt_2, dt_3); \tag{2.7}$$

– 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 3.13. Consistent with the remarks of [BS11], 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, i.e. it does not suffer from the particular integrability problems as (2.7). However, under certain conditions, it coincides with (2.7), 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 4.6, 4.11 and Section 4 of the companion paper [BKRS17].

## 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.** The random variables $X_1, \dots, X_n$ are called *$m$-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:

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

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

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

$$\prod_{i=1}^{n}(a_i - b_i) = \sum_{S \subset \{1,\dots,n\}} \left(\prod_{i \in S} a_i\right)\left(\prod_{i \in S^c} b_i\right)(-1)^{|S^c|}, \tag{3.2}$$

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

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

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

If we use the random variables $Z_i := \mathrm{e}^{\mathrm{i} X_i \cdot t_i}$, Lemma 3.2 yields the following result.

**Corollary 3.3.** *If the random variables $X_1, \dots, X_m$ are $(m-1)$-independent, then*

$$\begin{aligned}
&\mathbb{E}\left[\prod_{k=1}^{m}\left(\mathrm{e}^{\mathrm{i} X_{i_k} \cdot t_{i_k}} - f_{X_{i_k}}(t_{i_k})\right)\right] \\
&\qquad = f_{(X_{i_1},\dots,X_{i_m})}(t_{i_1}, \dots, t_{i_m}) - f_{X_{i_1}}(t_{i_1}) \cdot \ldots \cdot f_{X_{i_m}}(t_{i_m}).
\end{aligned} \tag{3.3}$$

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, i.e.*

$$X_1, \dots, X_n \quad \text{are independent} \implies M_\rho(X_1, \dots, X_n) = 0. \tag{3.4}$$

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

b) *Total distance multivariance characterizes independence, i.e.*

$$X_1, \dots, X_n \quad \text{are independent} \iff \overline{M}_\rho(X_1, \dots, X_n) = 0. \tag{3.5}$$

*Proof.* Suppose that $X_1, \dots, X_n$ are independent. For all indices $\{i_1, \dots, i_m\} \subset \{1, \dots, n\}$ we have

$$\mathbb{E}\left[\prod_{k=1}^m \left(\mathrm{e}^{\mathrm{i}X_{i_k}\cdot t_{i_k}} - f_{X_{i_k}}(t_{i_k})\right)\right] = \prod_{k=1}^m \mathbb{E}\left(\mathrm{e}^{\mathrm{i}X_{i_k}\cdot t_{i_k}} - f_{X_{i_k}}(t_{i_k})\right) = 0, \tag{3.6}$$

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

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

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

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

i.e. 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

$$M_{\bigotimes_{k=1}^m \rho_{i_k}}(X_{i_1}, \dots, X_{i_m}) = 0 \quad \text{for any} \quad \{i_1, \dots, i_m\} \subset \{1, \dots, n\}. \tag{3.7}$$

Starting with subsets of size 2, we note that

$$\overline{M}_{\rho_{i_1}\otimes\rho_{i_2}}(X_{i_1}, X_{i_2}) = M_{\rho_{i_1}\otimes\rho_{i_2}}(X_{i_1}, X_{i_2}) = \|f_{(X_{i_1},X_{i_2})} - f_{X_{i_1}} f_{X_{i_2}}\|_{L^2(\rho_{i_1}\otimes\rho_{i_2})} = 0 \tag{3.8}$$

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, (3.7) 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 $M_\rho$, total multivariance $\overline{M}_\rho$ and generalized distance covariance $V$ as defined in the companion paper [BKRS17, Def. 3.1] coincide, i.e.

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

The following properties are straightforward.

**Proposition 3.5.** *Distance multivariance enjoys the following properties.*

$$M_{\rho_i}(X_i) = 0 \quad \text{for all} \quad i = 1, \dots, n, \tag{3.9}$$

$$M_\rho(X_1, \dots, X_n) = M_\rho(c_1X_1, \dots, c_nX_n) \quad \text{for} \quad c_i \in \{-1, +1\}. \tag{3.10}$$

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

$$M_\rho(X_1, \dots, X_n) = M_{\bigotimes_{i\in S}\rho_i}(X_i, i \in S) \cdot M_{\bigotimes_{i\in S^c}\rho_i}(X_i, i \in S^c). \tag{3.11}$$

*Proof.* If $n = 1$, the expectation in (2.2) becomes $\mathbb{E}\left(e^{\mathrm{i}X_i t_i} - \mathbb{E}e^{\mathrm{i}X_i t_i}\right) = 0$ and (3.9) follows. Property (3.10) follows from the symmetry of the measures $\rho_i$. For the last property, note that the assumption of independence allows us to factorize the following expression

$$\mathbb{E}\left[\bigotimes_{i=1}^n \left(\mathrm{e}^{\mathrm{i}X_i\cdot\bullet} - f_{X_i}(\bullet)\right)\right] = \mathbb{E}\left[\bigotimes_{i\in S} \left(\mathrm{e}^{\mathrm{i}X_i\cdot\bullet} - f_{X_i}(\bullet)\right)\right] \cdot \mathbb{E}\left[\bigotimes_{i\in S^c} \left(\mathrm{e}^{\mathrm{i}X_i\cdot\bullet} - f_{X_i}(\bullet)\right)\right].$$

Since also $\rho$ can be factorized into $\bigotimes_{i\in S}\rho_i$ and $\bigotimes_{i\in S^c}\rho_i$, (3.11) follows. □

Next we show the relation between the moment conditions in Definition 2.3 and a related upper bound for the multivariance $M_\rho$.

**Lemma 3.6.** *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*

$$\mathbb{E}\left(\prod_{i=1}^n \psi_i(X_{k_i,i} - X'_{l_i,i})\right) \le 4^n \prod_{i=1}^n \left(\mathbb{E}\psi_i^{p_i}(X_i)\right)^{1/p_i} \tag{3.12}$$

*holds for all* $k_i, l_i \in \{0,1\}, i = 1,\dots,n$ *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 [BKRS17, Eq. (2.6)].

The other implication follows directly from (3.12). To show (3.12), note that the generalized Hölder inequality for $n$-fold products (cf. [Sch16, 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. [BKRS17, Eq. (2.5)]) and the Minkowski inequality for the $L^{p_i}$-norm yields the bound

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

see also Proposition 2.4 in the companion paper [BKRS17]. □

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 $\psi_i$ given in

**Proposition 3.7.** *Multivariance* $M_\rho = M_\rho^2(X_1,\dots,X_n)$ *can be written as*

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

*or*

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

*where*

$$\mathrm{sgn}(l,k) := (-1)^{\sum_{j=1}^n (k_j+l_j)} = \begin{cases} +1, & \textit{if } (k,l) \textit{ contains an even number of ‘1’s,} \\ -1, & \textit{if } (k,l) \textit{ contains an odd number of ‘1’s.} \end{cases}$$

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

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

**Remark 3.8.** a) The representations (3.13) and (3.14) have an interesting structural resemblance to the Leibniz' formula for determinants. The representaion (3.15) is the analogue of [BKRS17, Cor. 3.5] for the bivariate case.

b) In the bivariate case $n = 2$, (3.15) also holds under the weaker moment condition $\mathbb{E}\psi_1(X_1) + \mathbb{E}\psi_2(X_2) < \infty$, cf. [BKRS17, Thm. 3.7].

*Proof of Proposition 3.7.* Using (2.6), we can rewrite $M_\rho$ in the following way:

$$\begin{aligned} M_\rho^2 &= \int \left| \mathbb{E}\left[ \prod_{i=1}^n \left( \mathrm{e}^{\mathrm{i}X_i \cdot t_i} - f_{X_i}(t_i) \right) \right] \right|^2 \rho(\mathrm{d}t) \\ &= \int \left| \mathbb{E}\left[ \prod_{i=1}^n \left( \mathrm{e}^{\mathrm{i}X_{1,i} \cdot t_i} - \mathrm{e}^{\mathrm{i}X_{0,i} \cdot t_i} \right) \right] \right|^2 \rho(\mathrm{d}t) \\ &= \int \mathbb{E}\left[ \prod_{i=1}^n \left( \mathrm{e}^{\mathrm{i}X_{1,i} \cdot t_i} - \mathrm{e}^{\mathrm{i}X_{0,i} \cdot t_i} \right) \left( \mathrm{e}^{-\mathrm{i}X'_{1,i} \cdot t_i} - \mathrm{e}^{-\mathrm{i}X'_{0,i} \cdot t_i} \right) \right] \rho(\mathrm{d}t) \\ &= \int \mathbb{E}\left[ \sum_{k,l \in \{0,1\}^n} (-1)^{\sum_{j=1}^n (k_j + l_j)} \prod_{i=1}^n \mathrm{e}^{\mathrm{i}(X_{k_i,i} - X'_{l_i,i}) \cdot t_i} \right] \rho(\mathrm{d}t) \end{aligned} \tag{3.16}$$

and the penultimate line already gives (3.13). By (3.10),

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

Applying this to (3.16) shows that the imaginary part of the complex exponential cancels for $i = 1$. Repeated applications to $i = 2, \dots, n$ removes the other imaginary terms, and we obtain

$$M_\rho^2 = \int \mathbb{E}\left[ \sum_{k,l \in \{1,2\}^n} \operatorname{sgn}(k,l) \prod_{i=1}^n \cos((X_{k_i,i} - X'_{l_i,i}) \cdot t_i) \right] \rho(\mathrm{d}t). \tag{3.17}$$

It remains to show that (3.17) 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 $m \in \{1, \dots, n\}$, we get with $k' = (k_1, \dots, k_{m-1}, k_{m+1}, \dots, k_n)$, $l' = (l_1, \dots, l_{m-1}, l_{m+1}, \dots, l_n)$,

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

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 (3.17).

Finally, by Lemma 3.6, 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

$$\begin{aligned} M_\rho^2 &= \mathbb{E}\left(\sum_{k,l\in\{0,1\}^n} (-1)^{\sum_{j=1}^n (k_j+l_j)} \prod_{i=1}^n (-\psi_i(X_{k_i,i} - X'_{l_i,i}))\right) \\ &= \mathbb{E}\left(\prod_{i=1}^n \left[-\psi_i(X_{1,i} - X'_{1,i}) + \psi_i(X_{1,i} - X'_{0,i}) + \psi_i(X_{0,i} - X'_{1,i}) - \psi_i(X_{0,i} - X'_{0,i})\right]\right) \qquad (3.18) \\ &= \mathbb{E}\left(\mathbb{E}\left(\prod_{i=1}^n \left[-\psi_i(X_{1,i} - X'_{1,i}) + \psi_i(X_{1,i} - X'_{0,i}) + \psi_i(X_{0,i} - X'_{1,i}) - \psi_i(X_{0,i} - X'_{0,i})\right] \,\middle|\, \boldsymbol{X}_1, \boldsymbol{X}'_1\right)\right) \\ &= \mathbb{E}\left(\prod_{i=1}^n \left[-\psi_i(X_i - X'_i) + \mathbb{E}(\psi_i(X_i - X'_i) \mid X_i) + \mathbb{E}(\psi_i(X_i - X'_i) \mid X'_i) - \mathbb{E}\psi_i(X_i - X'_i)\right]\right). \qquad \square \end{aligned}$$

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

$$\mathrm{C}_\mathcal{F} X := X - \mathbb{E}(X \mid \mathcal{F}). \qquad (3.19)$$

*Then $\mathrm{C}$ is a linear operator and*

$$\mathrm{C}_{\{\emptyset,\Omega\}} X = X - \mathbb{E}X, \qquad (3.20)$$

$$\mathrm{C}_\mathcal{F} \mathrm{C}_{\mathcal{F}'} X = X - \mathbb{E}(X \mid \mathcal{F}') - \mathbb{E}(X \mid \mathcal{F}) + \mathbb{E}(\mathbb{E}(X \mid \mathcal{F}') \mid \mathcal{F}) = \mathrm{C}_{\mathcal{F}'} \mathrm{C}_\mathcal{F} X, \qquad (3.21)$$

$$\mathrm{C}_\mathcal{F} \mathrm{C}_{\mathcal{F}'} X = 0 \quad \text{if } X \text{ is } \mathcal{F}\text{- or } \mathcal{F}'\text{-measurable.} \qquad (3.22)$$

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

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

$$M_\rho^2(X_1, \dots, X_n) = \mathbb{E}\left(\prod_{i=1}^n -\mathrm{C}_{X_i} \mathrm{C}_{X'_i} \psi_i(X_i - X'_i)\right). \qquad (3.23)$$

*Proof.* The representation (3.23) follows from (3.15) and

$$\begin{aligned} &\mathrm{C}_{X_i} \mathrm{C}_{X'_i} \psi_i(X_i - X'_i) \\ &\quad = \mathrm{C}_{X_i} \left(\psi_i(X_i - X'_i) - \mathbb{E}[\psi_i(X_i - X'_i) \mid X'_i]\right) \\ &\quad = \psi_i(X_i - X'_i) - \mathbb{E}[\psi_i(X_i - X'_i) \mid X'_i] - \mathbb{E}[\psi_i(X_i - X'_i) \mid X_i] + \mathbb{E}\psi_i(X_i - X'_i). \ \square \end{aligned}$$

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 [BKRS17, Sec. 2.1]. Therefore, we can associate with any cndf $\psi$ some Gaussian random field indexed by $\mathbb{R}^d$.

**Definition 3.11.** Assume that the random variables $X_1, \dots, X_n$ satisfy one of the moment conditions in Definition 2.3 and let $G_1, \dots, G_n$ be independent, stationary Gaussian random fields with mean and covariance structure

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

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

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

**Remark 3.12.** a) Using the centering operator C from Proposition 3.9, 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 [BKRS17, Sec. 7].

c) If $\psi_i$ 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 $\psi_1$ and $\psi_2$ are given by the Euclidean norm, then $\mathcal{G}(X_1, X_2)$ coincides with the *Brownian covariance* of Székely and Rizzo [SR09a].

d) If $\psi_i(x) = |x|^\alpha$, then $G_i$ is a fractional Brownian field with Hurst exponent $H = \frac{\alpha}{2}$, cf. [SR09a, Sec. 4].

e) 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, i.e., $\mathbb{E}(|G_i(t)|^n) = 2^{\frac{n}{2}}\Gamma(\frac{n+1}{2})\pi^{-\frac{1}{2}}[\mathbb{E}G_i(t)^2]^{\frac{n}{2}}$, and $\mathbb{E}G_i(t)^2 = 2\psi_i(t)$ imply

$$\mathbb{E}|X_i^{G_i}|^n \le 2^n \mathbb{E}|G_i(X_i)|^n = 2^{2n}\Gamma(\tfrac{n+1}{2})\pi^{-1}\mathbb{E}(\psi_i(X_i)^{\frac{n}{2}}). \tag{3.27}$$

**Theorem 3.13.** *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, i.e.*

$$M_\rho(X_1, \dots, X_n) = \mathcal{G}(X_1, \dots, X_n). \tag{3.28}$$

*Proof.* By Corollary 3.10 we can represent squared multivariance in the product form (3.23). Each of the factors can be rewritten as

$$\begin{aligned}
&- C_X C_{X'} \psi(X - X') = \\
&\quad = C_X C_{X'} \left(\psi(X) + \psi(X') - \psi(X - X')\right) \\
&\quad = C_X C_{X'} \mathbb{E}(G(X)G(X') \mid X, X') \\
&\quad = \mathbb{E}(G(X)G(X') \mid X, X') - \mathbb{E}(G(X)G(X') \mid X) \\
&\qquad\qquad - \mathbb{E}(G(X)G(X') \mid X') + \mathbb{E}(G(X)G(X')) \\
&\quad = \mathbb{E}\left[\left(G(X) - \mathbb{E}(G(X) \mid G)\right)\left(G(X') - \mathbb{E}(G(X') \mid G)\right) \mid X, X'\right] \\
&\quad = \mathbb{E}(X^G X'^G \mid X, X'),
\end{aligned} \tag{3.29}$$

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

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

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

Note that for the penultimate equality the integrability, $\mathbb{E}\left(\prod_{i=1}^n |X_i^{G_i} X_i'^{G_i}|\right) < \infty$, is required. But this is implied by (3.27) and

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

where we used successively the independence of the $G_i$, the conditional Hölder inequality [CT97, 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 [Sch16, p. 133, Pr. 13.5] and the conditional Jensen inequality [CT97, 7.1.4]. □

## 4. Statistical properties of distance multivariance

**4.1. Sample distance multivariance.** We now consider a sample $(\boldsymbol{x}^{(1)},\dots,\boldsymbol{x}^{(N)})$ of $N$ observations of the random vector $\boldsymbol{X} = (X_1,\dots,X_n)$. Every observation $\boldsymbol{x}^{(j)}$ is a vector in $\mathbb{R}^d$, $d = d_1 + \dots + d_n$, of the form $\boldsymbol{x}^{(j)} = \left(x_1^{(j)},\dots,x_n^{(j)}\right)$, 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*

$$
\begin{aligned}
{}^{N}\!M_\rho^2(\boldsymbol{x}^{(1)},\dots,\boldsymbol{x}^{(N)}) &= \frac{1}{N^2}\sum_{j,k=1}^N (A_1\circ\dots\circ A_n)_{jk}\\
&= \frac{1}{N^2}\sum_{j,k=1}^N (A_1)_{jk}\cdot\ldots\cdot(A_n)_{jk};
\end{aligned}
\tag{4.1}
$$

*here,* $A_i := CB_iC$ *where* $B_i = \left(-\psi_i\big(x_i^{(j)} - x_i^{(k)}\big)\right)_{j,k=1,\dots,N}$ *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*

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

**Remark 4.2.** a) If $n$ is even, then $B_i$ can be replaced by $-B_i$. This explains the different sign used in the case $n = 2$, cf. [SR09a, Def. 3] and [BKRS17, Lem. 4.2, Rem. 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 [BKRS17, Sec. 4] is recovered. If in addition $\psi_i(x) = |x|$, i.e. the Euclidean distance, then we get the sample distance covariance of Székely *et al.* [SRB07, SR09a].

b) Since the $\psi_i$ are continuous negative definite functions, the matrices $B_i$ are conditionally positive definite matrices, i.e. $\lambda^\top B_i\lambda \geq 0$ for all non-zero $\lambda$ in $\mathbb{R}^N$ with $\lambda_1 + \cdots + \lambda_N = 0$, cf. [BKRS17, Thm. 2.1]. As double centerings of conditionally negative 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 [BF75, Lem. 3.2]. This gives a simple explanation as to why ${}^N M_\rho^2$ is always a non-negative number.

c) Important special cases are when the $\psi_i$ 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 $\psi_i$ itself, defines a distance. Still, $\psi_i$ always defines a 'quasi-metric', i.e. a metric with a relaxed triangle inequality, cf. [BKRS17, Sec. 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 $\mathcal{O}(n)$.

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 (3.23), when the centering operator is replaced by the centering matrix.

*Proof of Theorem 4.1.* Since the empirical distribution has finite support, the moment conditions of Definition 2.3 are trivially satisfied. Therefore, we can use the representation (3.18) to get

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

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

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

$$\mathbb{E}\left(\psi_i\left(\hat{X}_i - \hat{X}_i'\right) \;\middle|\; \hat{X}_i = x_i^{(j)}\right) = \frac{1}{N}\sum_{l=1}^{N}\psi_i\left(x_i^{(j)} - x_i^{(l)}\right) = -\frac{1}{N}\left(\mathbb{1}_N^\top B\right)_j \tag{4.4b}$$

$$\mathbb{E}\left(\psi_i\left(\hat{X}_i - \hat{X}_i'\right) \;\middle|\; \hat{X}_i' = x_i^{(k)}\right) = \frac{1}{N}\sum_{m=1}^{N}\psi_i\left(x_i^{(m)} - x_i^{(k)}\right) = -\frac{1}{N}\left(B\mathbb{1}_N\right)_k \tag{4.4c}$$

$$-\mathbb{E}\psi_i\left(\hat{X}_i - \hat{X}_i'\right) = \frac{1}{N^2}\sum_{l,m=1}^{N}\psi_i\left(x_i^{(m)} - x_i^{(l)}\right) = \frac{1}{N^2}\mathbb{1}_N^\top B\mathbb{1}_N. \tag{4.4d}$$

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.

For the total multivariance, we fix $j, k$ and write $a_i = (A_i)_{jk}$ for the $(j,k)$th entry of the matrix $A_i$. We can expand the product

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

where $e_m(a_1, \dots, a_n)$ is the $m$th elementary symmetric polynomial in $(a_1, \dots, a_n)$, i.e.

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

In particular, $e_0(a_1, \dots, a_n) = 1$ and $e_1(a_1, \dots, a_n) = a_1 + \dots + a_n$. Summing over all matrix elements, and dividing by $N^2$, yields

$$\begin{aligned}
&\frac{1}{N^2}\sum_{j,k=1}^{N}(1 + (A_1)_{jk}) \cdot \ldots \cdot (1 + (A_n)_{jk}) - 1 \\
&\quad = \frac{1}{N^2}\sum_{j,k=1}^{N}\sum_{m=1}^{n} e_m\left((A_1)_{jk}, \dots, (A_n)_{jk}\right) \\
&\quad = \sum_{\substack{1 \le i_1 < \dots < i_m \le n \\ 2 \le m \le n}} \frac{1}{N^2}\sum_{j,k=1}^{N}(A_{i_1})_{jk} \cdot \ldots \cdot (A_{i_m})_{jk} \\
&\quad = \sum_{\substack{1 \le i_1 < \dots < i_m \le n \\ 2 \le m \le n}} M_\rho^2(\hat{X}_{i_1}, \dots, \hat{X}_{i_m}) = \overline{M}_\rho^2(\hat{X}_1, \dots, \hat{X}_n),
\end{aligned} \tag{4.5}$$

as claimed. Note that the first elementary symmetric polynomial $e_1$ does not contribute since the sum of the entries of each matrix $A_i$ is zero, see Remark 4.2.e). □

4.2. **Estimating distance multivariance.** In this section we study the properties of the sample distance multivariance ${}^N\!M_\rho$ as an estimator of $M_\rho$. 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 $M_\rho$)**.** *Let one of the moment conditions of Definition 2.3 be satisfied. Then ${}^N\!M_\rho$ is a strongly consistent estimator of $M_\rho$, i.e.*

$${}^N\!M_\rho(\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)}) \xrightarrow[N\to\infty]{} M_\rho(X_1, \dots, X_n) \quad a.s. \tag{4.6}$$

*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 [BKRS17, Thm. 4.4]. For $n \geq 3$ the arguments used in [BKRS17] break down. However, we show a weak consistency result under independence and relaxed moment conditions in Corollary 4.8 below.

Our next goal is to derive the distributional asymptotics for the estimator ${}^N\!M_\rho$. We start with a crucial technical lemma.

**Lemma 4.5.** *Let $\boldsymbol{X}^{(l)} := (X_1^{(l)}, \dots, X_n^{(l)})$ be iid copies of $\boldsymbol{X} = (X_1, \dots, X_n)$ and set*

$$Z_N(t) := \frac{1}{N} \sum_{l=1}^{N} \prod_{i=1}^{n} \left( \mathrm{e}^{\mathrm{i} X_i^{(l)} \cdot t_i} - \frac{1}{N} \sum_{k=1}^{N} \mathrm{e}^{\mathrm{i} X_i^{(k)} \cdot t_i} \right). \tag{4.7}$$

*Then*

$${}^N\!M_\rho(\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)}) = \| Z_N(\bullet) \|_{L^2(\rho)}. \tag{4.8}$$

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

$$\mathbb{E} Z_N(t) = 0, \tag{4.9}$$

$$\mathbb{E}\big(Z_N(t)\overline{Z_N(t')}\big) = \frac{(N-1)^n + (-1)^n (N-1)}{N^{n+1}} \prod_{i=1}^{n} \left[ f_{X_i}(t_i - t_i') - f_{X_i}(t_i)\overline{f_{X_i}(t_i')} \right], \tag{4.10}$$

$$\mathbb{E}\left( \left| \sqrt{N} Z_N(t) \right|^2 \right) = \frac{(N-1)^n + (-1)^n (N-1)}{N^n} \prod_{i=1}^{n} \left( 1 - |f_{X_i}(t_i)|^2 \right). \tag{4.11}$$

*Proof.* The equality (4.8) 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

$$\begin{aligned} \mathbb{E}(Z_N(t)\overline{Z_N(t')}) = \frac{1}{N^2} \sum_{l,l'=1}^{N} \mathbb{E} \Bigg[ \prod_{i=1}^{n} \left( \mathrm{e}^{\mathrm{i} X_i^{(l)} \cdot t_i} - \frac{1}{N} \sum_{k=1}^{N} \mathrm{e}^{\mathrm{i} X_i^{(k)} \cdot t_i} \right) \times \\ \times \prod_{i'=1}^{n} \left( \mathrm{e}^{-\mathrm{i} X_{i'}^{(l')} \cdot t_{i'}'} - \frac{1}{N} \sum_{k'=1}^{N} \mathrm{e}^{-\mathrm{i} X_{i'}^{(k')} \cdot t_{i'}'} \right) \Bigg]. \end{aligned} \tag{4.12}$$

The independence of $X_i$, $X_{i'}$ for $i \neq i'$ implies

$$\begin{aligned} &\mathbb{E} \left[ \prod_{i=1}^{n} \left( \mathrm{e}^{\mathrm{i} X_i^{(l)} \cdot t_i} - \frac{1}{N} \sum_{k=1}^{N} \mathrm{e}^{\mathrm{i} X_i^{(k)} \cdot t_i} \right) \cdot \prod_{i'=1}^{n} \left( \mathrm{e}^{-\mathrm{i} X_{i'}^{(l')} \cdot t_{i'}'} - \frac{1}{N} \sum_{k'=1}^{N} \mathrm{e}^{-\mathrm{i} X_{i'}^{(k')} \cdot t_{i'}'} \right) \right] \\ &\quad = \prod_{i=1}^{n} \mathbb{E} \left[ \left( \mathrm{e}^{\mathrm{i} X_i^{(l)} \cdot t_i} - \frac{1}{N} \sum_{k=1}^{N} \mathrm{e}^{\mathrm{i} X_i^{(k)} \cdot t_i} \right) \cdot \left( \mathrm{e}^{-\mathrm{i} X_i^{(l')} \cdot t_i'} - \frac{1}{N} \sum_{k'=1}^{N} \mathrm{e}^{-\mathrm{i} X_i^{(k')} \cdot t_i'} \right) \right] \end{aligned}$$

and each factor simplifies to

$$\begin{aligned}
&\mathbb{E}\left[\left(\mathrm{e}^{\mathrm{i}X_i^{(l)}\cdot t_i} - \frac{1}{N}\sum_{k=1}^N \mathrm{e}^{\mathrm{i}X_i^{(k)}\cdot t_i}\right)\cdot\left(\mathrm{e}^{-\mathrm{i}X_i^{(l')}\cdot t_i'} - \frac{1}{N}\sum_{k'=1}^N \mathrm{e}^{-\mathrm{i}X_i^{(k')}\cdot t_i'}\right)\right]\\
&= \mathbb{E}\mathrm{e}^{\mathrm{i}X_i^{(l)}\cdot t_i - \mathrm{i}X_i^{(l')}\cdot t_i'} - 2\frac{N-1}{N} f_{X_i}(t_i)\overline{f_{X_i}(t_i')} - \frac{2}{N} f_{X_i}(t_i - t_i')\\
&\quad + \frac{N^2-N}{N^2} f_{X_i}(t_i)\overline{f_{X_i}(t_i')} + \frac{N}{N^2} f_{X_i}(t_i - t_i')\\
&= \mathbb{E}\mathrm{e}^{\mathrm{i}X_i^{(l)}\cdot t_i - \mathrm{i}X_i^{(l')}\cdot t_i'} - \frac{N-1}{N} f_{X_i}(t_i)\overline{f_{X_i}(t_i')} - \frac{1}{N} f_{X_i}(t_i - t_i').
\end{aligned}$$

Thus, splitting the sum in (4.12) into $l = l'$ and $l \neq l'$ yields

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

For $t' = t$ this reduces to

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

**Theorem 4.6** (Asymptotic distribution of ${}^{N}\!M_\rho$)**.**

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

$$N\cdot {}^{N}\!M_\rho^2(\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)}) \xrightarrow[N\to\infty]{d} \|\mathbb{G}\|^2_{L^2(\rho)} \tag{4.13}$$

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

$$\operatorname{Cov}(\mathbb{G}(t), \mathbb{G}(t')) = \mathbb{E}\big[\mathbb{G}(t)\overline{\mathbb{G}(t')}\big] = \prod_{i=1}^n \Big(f_{X_i}(t_i - t_i') - f_{X_i}(t_i)\overline{f_{X_i}(t_i')}\Big). \tag{4.14}$$

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*

$$N\cdot {}^{N}\!M_\rho^2(\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)}) \xrightarrow[N\to\infty]{} \infty \quad a.s. \tag{4.15}$$

**Remark 4.7.** a) The complex-valued Gaussian process $\mathbb{G}$ has to be distinguished from the Gaussian processes $G_i$ that appear in Definition 3.11 of the Gaussian multivariance.

b) Using the results of [Csö85], the log-moment condition in a) can be relaxed by a weaker (but more involved) integral test cf. [Csö85, Condition ($\star$)].

From [BKRS17, Lem. 2.7] it is readily seen that the log-moment condition in Theorem 4.6.a) is equivalent to $\mathbb{E}\left[\log^{1+\epsilon}\left(1 \vee \sqrt{|X_1|^2+\dots+|X_n|^2}\right)\right] < \infty$.

c) The expectation of the limit in (4.13) can be calculated as

$$\mathbb{E}(\|\mathbb{G}\|^2_{L^2(\rho)}) = \prod_{i=1}^n \int_{\mathbb{R}^{d_i}} \left(1 - |f_{X_i}(t_i)|^2\right) \rho_i(dt_i) = \prod_{i=1}^n \mathbb{E}\psi_i(X_i - X_i'). \tag{4.16}$$

d) From Lemma 4.5 it can be seen that ${}^N\!M_\rho$ is a biased estimator of $M_\rho$, since in the case of independence

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

while $M_\rho^2(X_1, \dots, X_n) = 0$. For bivariate distance covariance, this bias has already been discussed by Cope [Cop09] and Székely and Rizzo [SR09b].

*Proof of Theorem 4.6.* 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 (4.15) follows.

For part a), let $Z_N(t)$ be defined as in (4.7). Then ${}^N\!M_\rho(\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)}) = \|Z_N(.)\|_{L^2(\rho)}$ by Lemma 4.5.

If $\sqrt{N} Z_N$ converges in distribution to a Gaussian process then, by Lemma 4.5, this process is centred and has the covariance structure (4.14), 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

$$\sqrt{N}\left({}^N\!f(t) - f(t)\right) = \int \mathrm{e}^{\mathrm{i}x\cdot t} \, \mathrm{d}\Big(\sqrt{N}({}^N\!F_{\boldsymbol{X}}(x) - F_{\boldsymbol{X}}(x))\Big) \xrightarrow[N\to\infty]{d} \int \mathrm{e}^{\mathrm{i}x\cdot t} \, \mathrm{d}B(x), \tag{4.17}$$

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 (4.7) as

$$Z_N(t) = \sum_{S \subset \{1,\dots,n\}} (-1)^{n-|S|} \left({}^N\!f_S(t_S) \cdot \prod_{j \in S^c} {}^N\!f_j(t_j)\right). \tag{4.18}$$

In addition, we have the simple identity, cf. (3.2),

$$\prod_{j=1}^N \left({}^N\!f_i(t_i) - f_i(t_i)\right) = \sum_{S \subset \{1,\dots,n\}} (-1)^{n-|S|} \left(\prod_{j \in S} f_j(t_j) \cdot \prod_{j \in S^c} {}^N\!f_j(t_j)\right). \tag{4.19}$$

Subtracting (4.19) from (4.18) and rearranging the resulting equation yields

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

By (4.17), we have that

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

Moreover, 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

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

Together with (4.18) this yields the convergence

$$\sqrt{N} Z_N(t) \xrightarrow[N\to\infty]{d} \sum_{S\subset\{1,\dots,n\}} (-1)^{n-|S|} \int e^{\mathrm{i} x_S\cdot t_S}\, dB(x) \cdot \prod_{j\in S^c} f_j(t_j), \tag{4.20}$$

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 $T > 0$,

$$\sqrt{N} Z_N \xrightarrow[N\to\infty]{d} \mathbb{G} \quad \text{on} \quad \mathcal{C}_T := (C(B_T^d), \|\bullet\|_{B_T^d}), \tag{4.21}$$

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 (4.13), it remains to show that also the $L^2(\rho)$-norms of both sides of (4.21) converge, and that $T$ can be sent to infinity. To this end, we apply the truncation argument that is used in [BKRS17, Thm. 4.5 *et seq.*].

Set

$$\rho_{i,\epsilon}(A) := \rho_i(A \cap (B_{1/\epsilon}^{d_i} \setminus B_\epsilon^{d_i})) \quad \text{and} \quad \rho_i^\epsilon := \rho_i - \rho_{i,\epsilon}, \tag{4.22}$$

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

$$\left|\|h\|_{\rho_\epsilon} - \|h'\|_{\rho_\epsilon}\right|^2 \le \|h - h'\|_{\rho_\epsilon} \le \sup_{|x|\le\sqrt{n}/\epsilon} |h(x) - h'(x)|^2 \cdot \prod_{i=1}^{n} \rho_{i,\epsilon}(\mathbb{R}^{d_i}), \tag{4.23}$$

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

$$\|\sqrt{N} Z_N\|_{\rho_\epsilon}^2 \xrightarrow[N\to\infty]{d} \|\mathbb{G}\|_{\rho_\epsilon}^2. \tag{4.24}$$

By the portmanteau theorem, the convergence (4.13) is equivalent to the statement

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

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

$$\begin{aligned}\left|\mathbb{E}h(N \cdot {}^{N}\!M^2) - \mathbb{E}h(\|\mathbb{G}\|_\rho^2)\right| \le L_h \mathbb{E}\left|N \cdot {}^{N}\!M^2 - \|\sqrt{N} Z_N\|_{\rho_\epsilon}^2\right| \\ + \left|\mathbb{E}h\left(\|\sqrt{N} Z_N\|_{\rho_\epsilon}^2\right) - \mathbb{E}h\left(\|\mathbb{G}\|_{\rho_\epsilon}^2\right)\right| \\ + L_h \mathbb{E}\left|\|\mathbb{G}\|_{\rho_\epsilon}^2 - \|\mathbb{G}\|_\rho^2\right|.\end{aligned} \tag{4.25}$$

The middle term tends to zero as $N \to \infty$, by (4.24). To estimate the other terms, set

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

For the first term on the right hand side of (4.25) we get the bound

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

Using (4.11) we see with the constant $C_N = [(N-1)^n + (-1)^n(N-1)]/N^n \le 1$

$$\left\|\mathbb{E}\left(|\sqrt{N} Z_N|^2\right)\right\|_{\mu^\epsilon}^2 = C_N \sum_{k=1}^n \left[\|1 - |f_{X_k}|^2\|_{\rho_k^\epsilon}^2 \prod_{\substack{i=1\\ i\ne k}}^n \|1 - |f_{X_i}|^2\|_{\rho_i}^2\right], \tag{4.26}$$

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

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

The last term in (4.25) can be estimated in a similar way. We have

$$\|\mathbb{G}\|_\rho^2 - \|\mathbb{G}\|_{\rho_\epsilon}^2 \le \|\mathbb{G}\|_{\mu^\epsilon}^2 \xrightarrow[\epsilon\to 0]{} 0 \quad \text{a.s.} \tag{4.27}$$

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

$$\mathbb{E}(\|\mathbb{G}\|_{\mu^\epsilon}^2) \le n\mathbb{E}(\|\mathbb{G}\|_\rho^2) = n \prod_{i=1}^n \|1 - |f_{X_i}|^2\|_{\rho_i}^2 = n \prod_{i=1}^n \mathbb{E}\psi_i(X_i - X_i') < \infty. \tag{4.28}$$

Together with (4.25) this shows the convergence result (4.13) and completes the proof. □

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

$${}^{N}\!M_\rho(\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(N)}) \xrightarrow[N\to\infty]{} 0 \quad \textit{in probability.} \tag{4.29}$$

*Proof.* The corollary is a direct consequence of Theorem 4.6 and the observation that

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

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

$$ {}^N\overline{M}^2_\rho(\boldsymbol{X}^{(1)},\dots,\boldsymbol{X}^{(N)}) = \sum_{\substack{S\subset\{1,\dots,n\}\\ |S|\geq 2}} {}^N M^2_{\rho_S}(\boldsymbol{X}^{(1)},\dots,\boldsymbol{X}^{(N)}). \tag{4.30}$$

Note that $M_{\rho_S}$ depends only on the random variables $(X_i, i\in S)$, i.e. $M_{\rho_S} = M_{\rho_S}(X_i, i\in S)$. 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 $M_\rho$ by $M_{\rho_S}$ and ${}^N M_\rho$ by ${}^N M_{\rho_S}$).

**Corollary 4.9** (${}^N\overline{M}_\rho$ is a strongly consistent estimator of $\overline{M}_\rho$)**.** *Assume that one of the moment conditions of Definition 2.3 is satisfied. Then*

$$ {}^N\overline{M}_\rho(\boldsymbol{X}^{(1)},\dots,\boldsymbol{X}^{(N)}) \xrightarrow[N\to\infty]{} \overline{M}_\rho(X_1,\dots,X_n) \text{ a.s.} \tag{4.31}$$

*Proof.* This follows by an application of Theorem 4.3 to each $M_{\rho_S}$ in (4.30). □

**Corollary 4.10.** *Suppose that $X_1,\dots,X_n$ are independent random variables with $\mathbb{E}\psi_i(X_i)<\infty$ and $\mathbb{E}\left[\log^{1+\epsilon}(1+|X_i|^2)\right]<\infty$ for some $\epsilon>0$ and all $i=1,\dots,n$. Then*

$$ {}^N\overline{M}_\rho(\boldsymbol{X}^{(1)},\dots,\boldsymbol{X}^{(N)}) \xrightarrow[N\to\infty]{} 0 \quad \text{in probability.} \tag{4.32}$$

*Proof.* This follows by an application of Corollary 4.8 to each $M_{\rho_S}$ in (4.30). □

The next theorem is the analogue of the convergence result Theorem 4.6. For each $S\subset\{1,\dots,n\}$, we denote by $\mathbb{G}_S$ the centred Gaussian process

$$ \mathbb{G}_S(t_S) := \sum_{R\subset S} (-1)^{n-|R|} \int e^{ix_R\cdot t_R}\,dB(x)\cdot \prod_{j\in R^c} f_j(t_j), \tag{4.33}$$

cf. (4.20), indexed by $t_S \in \times_{i\in S}\mathbb{R}^{d_i}$, and where $B$ is the Brownian bridge from (4.17). Applying Theorem 4.6 with $\{1,\dots,n\}$ replaced by $S$, we see that $\mathbb{G}_s$ has covariance structure

$$ \mathbb{E}(\mathbb{G}_S(t)\overline{\mathbb{G}_S(t')}) = \prod_{i\in S}\left(f_{X_i}(t_i-t_i') - f_{X_i}(t_i)\overline{f_{X_i}(t_i')}\right). \tag{4.34}$$

**Theorem 4.11** (Asymptotic distribution of ${}^N\overline{M}_\rho$)**.**

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

$$ N\cdot {}^N\overline{M}^2_\rho(\boldsymbol{X}^{(1)},\dots,\boldsymbol{X}^{(N)}) \xrightarrow[N\to\infty]{d} \sum_{\substack{S\subset\{1,\dots,n\}\\ |S|\geq 2}} \|\mathbb{G}_S\|^2_{L^2(\rho_S)}. \tag{4.35}$$

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*

$$ N\cdot {}^N\overline{M}^2_\rho(\boldsymbol{X}^{(1)},\dots,\boldsymbol{X}^{(N)}) \xrightarrow[N\to\infty]{} \infty \quad \text{a.s.} \tag{4.36}$$

**Remark 4.12.** Note that the processes $(\mathbb{G}_S), S\subset\{1,\dots,n\}$ on the right hand side of (4.35) are *jointly Gaussian.* Therefore, the limit appearing in (4.35) is a quadratic form of centred Gaussian random variables. This fact will be used in Subsection 4.5 to construct a statistical test of (multivariate) independence.

*Proof of Theorem 4.11.* a) For any $S \subset \{1, \ldots, n\}$ with $|S| \geq 2$, we know from Theorem 4.6 that

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

and (4.35) follows.

b) By Corollary 4.9 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 $(X_1, \ldots, X_n)$ 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 absolute value is bounded by 1 – analogous to Székely-Rizzo-and-Bakirov's distance correlation [SRB07, Def. 3];
- To normalize the asymptotic distribution of the sample (total) distance multivariance under independence, cf. (4.13) in Theorem 4.6 and (4.35) in Theorem 4.11.

We will use normalized multivariances as test statistics in two tests for independence in Section 4.5.

*Distance multicorrelation.*

**Definition 4.13.** Let $X_1, \ldots, X_n$ be non-degenerate random variables and suppose that $\mathbb{E}\psi_i^n(X_i) < \infty$ for all $i = 1, \ldots, n$. We set

$$a_i := \left\| \mathrm{C}_{X_i} \mathrm{C}_{X_i'} \psi_i(X_i - X_i') \right\|_{L^n(\mathbb{P})}$$
$$= \mathbb{E}\Big[ \left|\psi_i(X_i - X_i') - \mathbb{E}\left(\psi_i(X_i - X_i') \mid X_i'\right) - \mathbb{E}\left(\psi_i(X_i - X_i') \mid X_i\right) + \mathbb{E}\psi_i(X_i - X_i')\right|^n \Big]^{\frac{1}{n}}$$

and define *distance multicorrelation* as

$$\mathcal{R}^2_\rho(X_1, \ldots, X_n) := \frac{M^2_\rho(X_1, \ldots, X_n)}{a_1 \cdot \ldots \cdot a_n}. \tag{4.37}$$

For the sample version of distance multicorrelation, we define

$${}^N a_i := {}^N a_i(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}) = \left( \frac{1}{N^2} \sum_{k,l=1}^N |(A_i)_{kl}|^n \right)^{1/n}, \tag{4.38}$$

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

$${}^N\mathcal{R}^2_\rho(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(1)}) := \frac{1}{N^2} \sum_{k,l=1}^N \frac{(A_1)_{kl}}{{}^N a_1} \cdot \ldots \cdot \frac{(A_n)_{kl}}{{}^N a_n}. \tag{4.39}$$

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

**Proposition 4.14.** a) *Distance multicorrelation and its sample version satisfy*

$$0 \leq \mathcal{R}_\rho(X_1, \ldots, X_n) \leq 1 \quad \textit{and} \quad 0 \leq {}^N\mathcal{R}_\rho(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}) \leq 1. \tag{4.40}$$

b) *For iid copies $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(N)}$ of $\boldsymbol{X} = (X_1, \ldots, X_n)$ it holds that*

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

c) *For $n = 2$ and $\psi_1(x) = \psi_2(x) = |x|$ distance multicorrelation coincides with the distance correlation of* [SRB07].

**Remark 4.15.** Székely and Rizzo [SR09a, Thm 4.(iv)] show for the case $n = 2$ (i.e. for distance correlation) that ${}^N\mathcal{R}_\rho(\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(N)}) = 1$ implies that the sample points $(x_1^{(1)}, \ldots, x_1^{(N)})$ and $(x_2^{(1)}, \ldots, x_2^{(N)})$ can be transformed into each other by a Euclidean isometry composed with scaling by a non-negative number. Unfortunately, we were not able to obtain an analogous result in the case $n > 2$.

*Proof of Proposition 4.14.* By the generalized Hölder inequality for $n$-fold products (cf. [Sch16, p. 133, Pr. 13.5]), we have that

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

and (4.40) follows. For the convergence result, note that

$${}^N a_i^n = \frac{1}{N^2} \sum_{k,l=1}^N |(A_i)_{kl}|^n \xrightarrow[N\to\infty]{} \mathbb{E}\left[\left|\mathrm{C}_{X_i}\, \mathrm{C}_{X_i'}\, \psi_i(X_i - X_i')\right|^n\right] = a_i^n \tag{4.41}$$

by the law of large numbers for V-statistics, cf. Theorem 4.3 and its proof. Part c) follows from direct comparison with [SR09a]. □

*Normalized distance multivariance.* Alternatively, we can normalize distance multivariance in such a way, that the limiting distribution under independence (cf. Theorems 4.6 and 4.11) has unit expectation.

**Definition 4.16.** Let $X_1, \ldots, X_n$ be random variables such that $\mathbb{E}\psi_i(X_i) < \infty$ for all $i = 1, \ldots, n$, set

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

and define *normalized distance multivariance* as

$$\mathcal{M}_\rho^2(X_1, \ldots, X_n) := \frac{M_\rho^2(X_1, \ldots, X_n)}{b_1 \cdot \ldots \cdot b_n}. \tag{4.42}$$

For the sample version of normalized distance multicorrelation, we define

$${}^N b_i := {}^N b_i(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}) := \frac{1}{N^2} \sum_{k,l=1}^N \psi_i\left(x_l^i - x_k^i\right) = -\frac{1}{N^2} \sum_{k,l=1}^N (B_i)_{kl}, \tag{4.43}$$

and set

$${}^N\mathcal{M}_\rho^2(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(1)}) := \frac{1}{N^2} \sum_{k,l=1}^N \frac{(A_1)_{kl}}{{}^N b_1} \cdot \ldots \cdot \frac{(A_n)_{kl}}{{}^N b_n}. \tag{4.44}$$

**Corollary 4.17.** *Suppose that $X_1, \ldots, X_n$ are independent random variables with $\mathbb{E}\psi_i(X_i) < \infty$ and $\mathbb{E}\left[\log^{1+\epsilon}(1 + |X_i|^2)\right] < \infty$ for some $\epsilon > 0$ and all $i = 1, \ldots, n$. Then*

$$N \cdot {}^N\mathcal{M}_\rho^2(\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(N)}) \xrightarrow[N\to\infty]{d} Q, \tag{4.45}$$

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

*Proof.* This follows from Theorem 4.6 in combination with

$${}^N b_i = \frac{1}{N^2} \sum_{k,l=1}^N \psi_i(X_i^{(k)} - X_i^{(l)}) \xrightarrow[N\to\infty]{} \mathbb{E}\psi_i(X_i - X_i') = b_i, \tag{4.46}$$

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

$$\overline{\mathcal{M}}_\rho^2(X_1,\dots,X_n)^2 := \frac{1}{2^n-1-n}\sum_{\substack{S\subset\{1,\dots,n\}\\|S|\geq 2}} \frac{M_{\rho_S}^2(X_i, i\in S)}{b_S}. \tag{4.47}$$

Its sample version becomes

$$^N\overline{\mathcal{M}}_\rho^2(\boldsymbol{x}^{(1)},\dots,\boldsymbol{x}^{(1)}) := \frac{1}{2^n-1-n}\left\{\frac{1}{N^2}\sum_{k,l=1}^{N}\left(1+\frac{(A_1)_{kl}}{^Nb_1}\right)\cdot\ldots\cdot\left(1+\frac{(A_1)_{kl}}{^Nb_1}\right)-1\right\}. \tag{4.48}$$

Similar to Corollary 4.17, we have the following result.

**Corollary 4.19.** *Suppose that $X_1,\dots,X_n$ are independent random variables with $\mathbb{E}\psi_i(X_i)<\infty$ and $\mathbb{E}\left[\log^{1+\epsilon}(1+|X_i|^2)\right]<\infty$ for some $\epsilon>0$ and all $i=1,\dots,n$. Then*

$$N\cdot {^N\overline{\mathcal{M}}_\rho^2}(\boldsymbol{X}^{(1)},\dots,\boldsymbol{X}^{(N)}) \xrightarrow[N\to\infty]{d} \overline{Q}, \tag{4.49}$$

*where*

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

*Proof.* Convergence follows from Theorem 4.11. Note that $\mathbb{E}\left[\|\mathbb{G}_S\|^2_{L^2(\rho_S)}\right]=b_S$ by Corollary 4.17, with the sum running over $2^n-n-1$ subsets. □

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 4.17 and 4.19, 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 [SR09a]: Both limiting random variables $Q$ and $\overline{Q}$ are quadratic forms of centred Gaussian random variables, normalized to $\mathbb{E}Q=\mathbb{E}\overline{Q}=1$. Hence, by Székely and Bakirov [SB03, p. 181],

$$\mathbb{P}(Q\geq\chi^2_{1-\alpha}(1))\leq\alpha \quad\text{and}\quad \mathbb{P}(\overline{Q}\geq\chi^2_{1-\alpha}(1))\leq\alpha, \tag{4.50}$$

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 (4.50) is, in general, very rough, thus the following test is, 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 4.17 and one of the moment conditions of Definition 2.3 hold. Under the assumption of $(n-1)$-independence, the null hypothesis*

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

*is rejected against the alternative hypothesis*

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

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

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

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 4.19 and one of the moment conditions of Definition 2.3 hold. The null hypothesis*

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

*is rejected against the alternative hypothesis*

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

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

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

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 4.6 and Theorem 4.11. 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 $N$.

## 5. Example: Bernstein's coins

The first example of pairwise independent, but not (totally) independent random variables is commonly attributed to S.N. Bernstein, cf. [Fel71, Sec. 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 $A \cap B \cap C = \emptyset$, 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[1] 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 $\rho_A, \rho_B, \rho_C$ be one-dimensional symmetric Lévy measures with the corresponding continuous negative definite functions $\psi_A$, $\psi_B$ and $\psi_C$. We write $\rho = \rho_A \otimes \rho_B \otimes \rho_C$ and $\rho_{AB} := \rho_A \otimes \rho_B$ etc.

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

*Analytic Approach.* First, note that pairwise independence immediately yields

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

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

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

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, i.e.

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

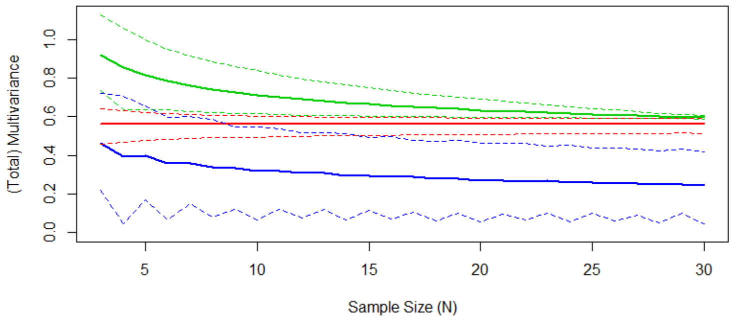
Finally, normalized total multivariance is given by

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

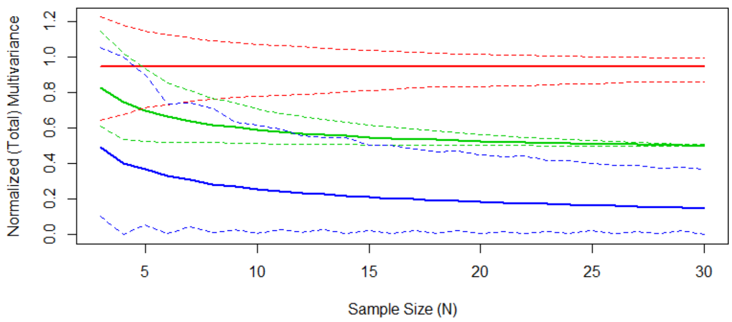
*Numerical Simulation.* To complement the analytical results by a numerical simulation, we have simulated 5000 replications of $N = 3,\dots,30$ tosses of Bernstein's coins. We calculated the pairwise sample distance covariances ${}^{N}M_{\rho_{AB}}(A,B)$, ${}^{N}M_{\rho_{BC}}(B,C)$, ${}^{N}M_{\rho_{AC}}(C,A)$ 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 non-zero 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 4.6 and 4.11. 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

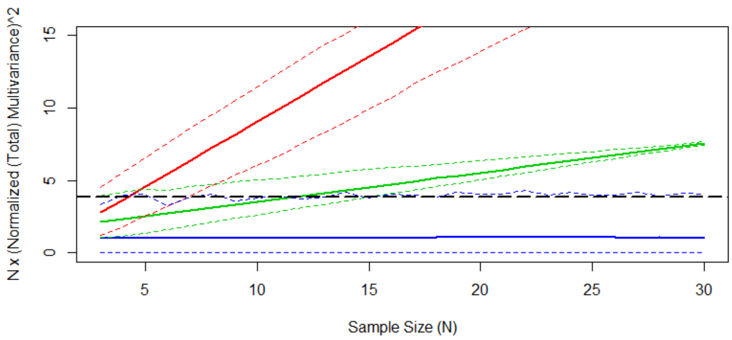
- (Total) distance multivariance is able to distinguish correctly pairwise independence of the events $A, B, C$ from their higher-order dependence;
- The sample statistics converge quickly to their analytic limits and give a numerical confirmation of the asymptotic results from Theorems 4.6 and 4.11.
- 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 $N > 14$.




(a) Multivariance without normalization




(b) Normalized multivariance




(c) Squared normalized multivariance scaled by sample size

FIGURE 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).

Note that all necessary functions for such simulations and for the use of distance multivariance in applications are provided in the R package `multivariance` [Böt17].

## References


[BF75] Christian Berg and Gunnar Forst. *Potential Theory on Locally Compact Abelian Groups*. Springer, Berlin, 1975.

[BKRS17] Björn Böttcher, Martin Keller-Ressel, and René L. Schilling. Detecting independence of random vectors I. Generalized distance covariance and Gaussian covariance. 2017.

[Böt17] Björn Böttcher. *multivariance: Measuring Multivariate Dependence Using Distance Multivariance*, 2017. R package version 1.0.5.

[BS11] 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.

[Cop09] Leslie Cope. Discussion of: Brownian distance covariance. *Annals of Applied Statistics*, 3(4):1279–1281, 12 2009.

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

[Csö85] Sándor Csörgő. Testing for independence by the empirical characteristic function. *Journal of Multivariate Analysis*, 16(3):290–299, 1985.

[CT97] Yuan Shih Chow and Henry Teicher. *Probability theory: independence, interchangeability, martingales*. Springer Science & Business Media, New York, 3rd edition, 1997.

[Fel71] William Feller. *An introduction to probability theory and its applications, volume I.* John Wiley & Sons, 3rd edition, 1971.

[SB03] Gábor J. Székely and Nail K. Bakirov. Extremal probabilities for Gaussian quadratic forms. *Probability theory and related fields*, 126(2):184–202, 2003.

[Sch16] René. L. Schilling. *Measures, Integrals and Martingales* (*second edition*). Cambridge University Press, 2016.

[SR09a] Gábor J Székely and Maria L Rizzo. Brownian distance covariance. *Annals of Applied Statistics*, 3(4):1236–1265, 2009.

[SR09b] Gábor J. Székely and Maria L. Rizzo. Rejoinder: Brownian distance covariance. *Annals of Applied Statistics*, 3(4):1303–1308, 12 2009.

[SRB07] Gábor J. Székely, Maria L. Rizzo, and Nail K. Bakirov. Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, 35(6):2769–2794, 2007.



TU Dresden, Fachrichtung Mathematik, Institut für Mathematische Stochastik, 01062 Dresden, Germany

*E-mail address*: `bjoern.boettcher@tu-dresden.de`
*URL*: `https://www.math.tu-dresden.de/~boettch/`

*E-mail address*: `martin.keller-ressel@tu-dresden.de`
*URL*: `https://www.math.tu-dresden.de/~mkeller/`

*E-mail address*: `rene.schilling@tu-dresden.de`
*URL*: `https://www.math.tu-dresden.de/sto/schilling/`