Research Article

Björn Böttcher*

Dependence and dependence structures: estimation and visualization using the unifying concept of distance multivariance

https://doi.org/10.1515/stat-2020-0001
Received Aug 26, 2019; accepted Oct 28, 2019

Abstract: Distance multivariance is a multivariate dependence measure, which can detect dependencies between an arbitrary number of random vectors each of which can have a distinct dimension. Here we discuss several new aspects, present a concise overview and use it as the basis for several new results and concepts: in particular, we show that distance multivariance unifies (and extends) distance covariance and the Hilbert-Schmidt independence criterion HSIC, moreover also the classical linear dependence measures: covariance, Pearson's correlation and the RV coefficient appear as limiting cases. Based on distance multivariance several new measures are defined: a multicorrelation which satisfies a natural set of multivariate dependence measure axioms and m-multivariance which is a dependence measure yielding tests for pairwise independence and independence of higher order. These tests are computationally feasible and under very mild moment conditions they are consistent against all alternatives. Moreover, a general visualization scheme for higher order dependencies is proposed, including consistent estimators (based on distance multivariance) for the dependence structure.
Many illustrative examples are provided. All functions for the use of distance multivariance in applications are published in the R-package `multivariance`.

Keywords: multivariate dependence, testing independence, higher order dependence, multivariate dependence measure axioms, distance covariance, Hilbert Schmidt independence criterion

MSC: Primary: 62H15, Secondary: 62H20.

Contents

***Corresponding Author: Björn Böttcher:** TU Dresden, Fakultät Mathematik, Institut für Mathematische Stochastik, 01062 Dresden, Germany; Email: bjoern.boettcher@tu-dresden.de

1 Introduction

The detection of dependence is a common statistical task, which is crucial in many applications. There have been many methods employed and proposed, (see e.g. Josse and Holmes 2016; Tjøstheim *et al.* 2018; Liu *et al.* 2018) for recent surveys. Usually these focus on the (functional) dependence of pairs of variables. Thus when the dependence of many variables is studied the resulting networks (correlation networks, graphical models) only show the pairwise dependence. As long as pairwise dependence is present, this might be sufficient (and also for the detection of such dependencies total multivariance and m-multivariance provide efficient tests). But recall that pairwise independence does not imply the independence of all variables if more than two variables are considered. Thus, in particular if all variables are pairwise independent many methods of classical statistical inference would have discarded the data. Although there might be some higher order dependence present. This can only be detected directly with a multivariate dependence measure. The classical examples featuring 3-dependence are a dice in the shape of a tetrahedron with specially coloured sides (see Example 10.1) and certain events in multiple coin throws (Examples 10.2). In Example 10.3 a generalization to higher orders is presented.

To avoid misconceptions when talking about independence one should note that the term "mutual independence" is ambiguous, some authors use it as a synonym for pairwise independence, others for independence. For the latter also the terms "total independence" or "joint independence" are used. We use the terms: pairwise independence, its extension m-independence (see Section 2) and independence. Another misconception might be triggered by the term "dependence measure". Formally, such a measure assigns a value to "dependence". In our setting and in most of the cited papers these values are values of scaled test statistics used in independence tests and their only meaningful comparison is based on the corresponding p-values! Therefore the whole theory is based on independence tests. (Section 3.6 might be viewed as a starting point for a direct comparison of the values of these measures, but it does not provide a meaningful interpretation in general.)

In Böttcher *et al.* (2018, 2019) the basics of distance multivariance and total distance multivariance were developed, which can be used to detect multivariate dependence. Incidentally, a variant of total distance multivariance based on the Euclidean distance was simultaneously developed in Chakraborty and Zhang (2019). Moreover, distance multivariance names and extends a concept introduced in Bilodeau and Guetsop Nangue (2017). Here we recall and extend the main definitions and properties (Sections 2 and 4). In particular, the moment conditions required in Böttcher *et al.* (2019) for the independence tests are considerably relaxed (Theorem 2.5, Tests 4.1 and 4.3), invariance properties are explicitly discussed (Propositions 2.3 and 2.4) and resampling tests are introduced (Section 4). Moreover, on this basis the following new results and concepts are developed:

- A general **scheme for the visualization of higher order dependence** which can be used with any multivariate dependence measure (Section 6). For the setting of multivariance we provide explicit **consistent estimators for the (higher order) dependence structure**. In particular the method for the clustered dependence structure is based on the fact that multivariance is a truly multivariate dependence measure: On the one hand it can detect the dependence of multiple (more than 2) random variables. On the other hand each random variable can be multivariate and each can have a different dimension.

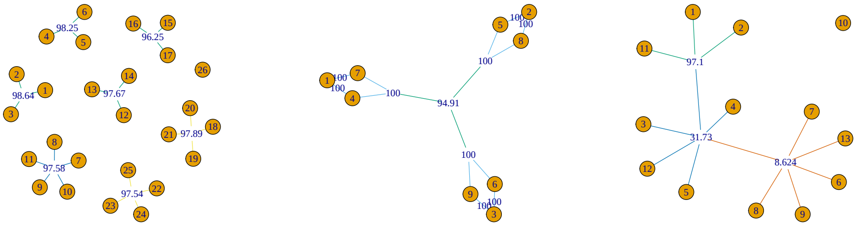


Figure 1: Visualized dependence structures of Examples 10.5, 10.6 and 10.7.

- **Global tests for pairwise (and higher order) independence**: Pairwise independence is a fundamental requisite for many standard tools, e.g. for the law of large numbers in its basic form (a result which is used in almost every statistical estimation). Recently in Yao *et al.* (2017) a test for pairwise independence of identically distributed random variables was proposed. In contrast, we derive in Section 5 a test for pairwise (and higher order) independence which is applicable to any mixture of marginal distributions and dimensions.
- **The framework of multivariance provides a unified theory**. Multivariance unifies several dependence measures and links these to classical theory, in particular:
  - We show in Section 3.1 that the RV-coefficient and, in particular, covariance (the most classical dependence measure of all) are covered as limiting cases by the framework of multivariance.
  - In Sejdinovic *et al.* (2013b) it was shown that independence testing methods based on reproducing kernels and methods based on distances are equivalent in a bivariate setting. We show in Section 3.2 that both are covered by the framework of multivariance. In particular a new, explicit and very elementary relation between these methods is presented. Moreover, this transfers also to the setting of multiple random vectors.
  - In independence testing Hilbert space methods require characteristic kernels. Multivariance requires the full support of the underlying measures. We show that these assumptions are equivalent (Proposition 2.2).
- **Multivariate correlations**: A formalization of desired properties of dependence measures goes back at least to Renyi's axioms (Rényi 1959). We discuss in Section 3.6 a multivariate extension of the axioms and provide several multicorrelations, e.g. (14), (15) and (56).

Recently also several other dependence measures for the detection of dependence of multiple random vectors were proposed, (e.g. Yao *et al.* 2017) proposed tests for pairwise independence, banded independence and independence based on distance covariance or based on the approach of Pfister *et al.* (2017). The latter presented tests for independence of multiple random vectors using kernels. In Jin and Matteson (2018) also distance covariance (a measure for the dependence of two random vectors; introduced in Székely *et al.* (2007)) was generalized to tests of independence of multiple random vectors. All these approaches are related to distance multivariance, see Section 3 for a detailed discussion. Empirical comparisons can be found in Examples 7.2 and 7.3.

It is remarkable that, although the above measures are able to detect higher order dependencies, all real data examples which were provided so far feature only pairwise dependence. Certainly the predominant statistical methods cause a publication bias for such datasets. Nevertheless, we want to point out that many available datasets feature higher order dependence. Based on a data driven study we collected over 350 datasets featuring statistically significant higher order dependencies[1]. All of these datasets are distributed as part of various R-packages without the context of higher order dependence. This indicates that higher order dependence can be detected frequently, but what remains open are intrinsic explanations of higher order dependence within each field of research of the underlying data. For illustration we discuss one of these datasets in further details in Section 7.2.

Besides the real data examples the presentation of this paper is complemented by a comprehensive collection of further examples (in Sections 7 and 10): illustrating higher order dependencies (Section 10.1), discussing various properties of distance multivariance (Section 10.2), comparisons to other dependence measures (Section 7.1). Technical details and further results are collected in Section 9.

The R code for the evaluation of distance multivariance and the corresponding tests is provided in the R-package `multivariance` (Böttcher 2019). Finally, based on (some of) the results of this paper we have the following recommendations for questions common in independence testing:

1. *Are at least some variables dependent? Detection of any kind of dependence:*

[1] A collection of datasets featuring higher order dependence, http://www.math.tu-dresden.de/~boettch/research/hod/

(a) The global independence test based on total multivariance can be used to detect any kind of dependence, alternatively 2-multivariance can be used to test for pairwise (in)dependence. The latter and m-multivariance can also be used (via a multiple testing approach) to reduce the statistical curse of dimension which total multivariance might suffer. For all settings fast distribution free (conservative) tests exist and these are applicable for large samples and a large number of random vectors. The computation of the test statistic takes in its current implementation for 100 variables with 1000 samples each (or for 1000 variables with 300 samples each) less than 2 seconds on a dated i7-6500U CPU Laptop. Slower, but approximately sharp, are the corresponding resampling tests. Faster approximately sharp tests are discussed in Berschneider and Böttcher (2019).

(b) As a complementary approach to the global tests one could perform multiple tests as suggested in Bilodeau and Guetsop Nangue (2017). This requires $2^n - n - 1$ individual tests, where n denotes the number of random vectors. Hence it is only applicable for small n. Bilodeau and Guetsop Nangue (2017) also provides a multiple testing approach to m-dependence.

2. *Which variables depend on each other? Dependence structure:* Especially if some dependence was detected the algorithm of Section 6 can be used to analyse which variables depend on each other, yielding either a full or clustered dependence structure. The method is based on multiple tests, but variables are clustered (or related tuples are excluded from further tests) as soon as a positive detection occurred. This can considerably reduce the computation time in comparison to 1.(b).

2 Distance multivariance

In the following distance multivariance is introduced. Some parts are essential for the (less technical) comparison to other dependence measures in Section 3, other parts are required for the introduction of m-multivariance (Section 5). Furthermore, several new results are included which make distance multivariance more accessible and applicable. Tests using distance multivariance will be discussed in Section 4.

Let X_i be $\mathbb{R}^{d_i}$ valued random variables with characteristic functions $f_{X_i}(t_i) = \mathbb{E}(\mathrm{e}^{\mathrm{i}t_i\cdot X_i})$ for $i = 1, \dots, n$, where $t_i \cdot X_i$ denotes the standard inner product $t_i^T X_i$. Then **distance multivariance** is defined by

$$M_\rho(X_1, \dots, X_n) := M_\rho(X_i, i \in \{1, \dots, n\}) := \sqrt{\int \left| \mathbb{E}\left(\prod_{i=1}^n (\mathrm{e}^{\mathrm{i}X_i\cdot t_i} - f_{X_i}(t_i)) \right) \right|^2 \rho(dt)} \tag{1}$$

and **total distance multivariance** is

$$\overline{M_\rho}(X_1, \dots, X_n) := \sqrt{\sum_{\substack{1\leq i_1<\ldots<i_m\leq n \\ 2\leq m\leq n}} M^2_{\otimes_{k=1}^m \rho_{i_k}}(X_{i_1}, \dots, X_{i_m})}, \tag{2}$$

where $\rho = \otimes_{i=1}^n \rho_i$ and each ρ_i is a symmetric measure with full topological support[2] on $\mathbb{R}^{d_i}$ such that $\int 1 \wedge |t_i|^2 \rho_i(dt_i) < \infty$ and $t = (t_1, \dots, t_n)$ with $t_i \in \mathbb{R}^{d_i}$. To simplify notation and definitions we will just use the term 'multivariance' instead of 'distance multivariance', and we will drop the subscript ρ if the measure is the full measure ρ.

Random variables $X_1, \dots, X_n$ are called ***m*-independent**, if $X_{i_1}, \dots, X_{i_m}$ are independent for any distinct $i_j \in \{1, \dots, n\}$ for $j = 1, \dots, m$. This concept is essential for the statement (and proof) of the following theorem. It is also the basis for the estimators for m-independence which will be developed in Section 5.

Theorem 2.1 (Characterization of independence, (Böttcher *et al.* 2019, Theorem 3.4.)). *For random variables $X_1, \dots, X_n$ the following are equivalent:*

[2] A measure ρ has **full topological support** on $\mathbb{R}^d$ if and only if $\rho(O) > 0$ for all open sets $O \subset \mathbb{R}^d$, $O \neq \emptyset$. See also Prop. 2.2.

1. *$X_1, \dots, X_n$ are independent,*
2. *$M(X_1, \dots, X_n) = 0$ and $X_1, \dots, X_n$ are $(n-1)$-independent,*
3. *$\overline{M}(X_1, \dots, X_n) = 0$.*

For statistical applications the following representations, which require the moment condition (5), are very useful. Let $(X_1', \dots, X_n')$ be an independent copy of $(X_1, \dots, X_n)$ and $\psi_i(y_i) := \int_{\mathbb{R}^{d_i}\setminus\{0\}} 1 - \cos(y_i \cdot t_i)\,\rho_i(dt_i)$ then

$$M_\rho^2(X_1, \dots, X_n) = \mathbb{E}\left(\prod_{i=1}^n \Psi_i(X_i, X_i')\right) \quad \text{and} \quad \overline{M_\rho}^2(X_1, \dots, X_n) = \mathbb{E}\left(\prod_{i=1}^n (1 + \Psi_i(X_i, X_i'))\right) - 1 \tag{3}$$

where

$$\Psi_i(X_i, X_i') := -\psi_i(X_i - X_i') + \mathbb{E}(\psi_i(X_i - X_i') \mid X_i) + \mathbb{E}(\psi_i(X_i - X_i') \mid X_i') - \mathbb{E}(\psi_i(X_i - X_i')). \tag{4}$$

Note that in Böttcher *et al.* (2019) a technical looking moment condition was required for the above representations, we show in Section 9.2 that the following more comprehensible condition is equivalent to it (for non constant random variables)

$$\textbf{finite joint } \boldsymbol{\psi}\textbf{-moments}\text{: for all } S \subset \{1, \dots, n\} : \mathbb{E}\left(\prod_{i \in S} \psi_i(X_i)\right) < \infty. \tag{5}$$

A direct consequence of (3) is the factorization of M and $\overline{M}$ for independent subsets, i.e., if $(X_i)_{i\in I}$ and $(X_i)_{i\in I^c}$ are independent for some $I \subset \{1, \dots, n\}$ then

$$M(X_i, i \in I \cup I^c) = M_{\otimes_{i\in I}\rho_i}(X_i, i \in I) \cdot M_{\otimes_{i\in I^c}\rho_i}(X_i, i \in I^c), \tag{6}$$

$$\overline{M}^2(X_i, i \in I \cup I^c) + 1 = (\overline{M}^2_{\otimes_{i\in I}\rho_i}(X_i, i \in I) + 1) \cdot (\overline{M}^2_{\otimes_{i\in I^c}\rho_i}(X_i, i \in I^c) + 1). \tag{7}$$

Furthermore, the expectations in (3) yield strongly consistent (see (Böttcher *et al.* 2019, Theorem 4.3) and Corollary 2.7) and numerically feasible estimators. Hereto denote samples of $(X_1, \dots, X_n)$ by $\boldsymbol{x}^{(k)} = (x_1^{(k)}, \dots, x_n^{(k)}) \in \mathbb{R}^{d_1} \times \dots \times \mathbb{R}^{d_n}$ for $k = 1, \dots, N$. Then **sample multivariance** ${}^N\!M$ is defined by

$${}^N\!M^2(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)}) := \frac{1}{N^2} \sum_{j,k=1}^N (A_1)_{jk} \cdot \ldots \cdot (A_n)_{jk} \tag{8}$$

and **sample total multivariance** ${}^N\overline{M}$ is defined by

$${}^N\overline{M}^2(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)}) := \left[\frac{1}{N^2} \sum_{j,k=1}^N (1 + (A_1)_{jk}) \cdot \ldots \cdot (1 + (A_n)_{jk})\right] - 1, \tag{9}$$

where $(A_i)_{jk}$ denotes the element in the j-th row and k-th column of the doubly centred distance matrix A_i defined by

$$A_i := -CB_iC \text{ with } B_i := \left(\psi_i(x_i^{(j)} - x_i^{(k)})\right)_{j,k=1,\dots,N} \text{ and } C := \left(\delta_{jk} - \frac{1}{N}\right)_{j,k=1,\dots,N}. \tag{10}$$

The matrices A_i are positive definite (Böttcher *et al.* 2019, Remark 4.2.b), since the considered distances $\psi_i(.-.)$ are given by

$$\psi_i(y_i) := \int_{\mathbb{R}^{d_i}\setminus\{0\}} 1 - \cos(y_i \cdot t_i)\,\rho_i(dt_i) \text{ for } y_i \in \mathbb{R}^{d_i}. \tag{11}$$

Functions defined via (11) appear in various areas: e.g. they are called variogram (e.g. Matheron 1963), continuous negative definite function (e.g. Berg and Forst 1975) or characteristic exponent of a Lévy process with Lévy measure ρ_i (e.g. Sato 1999), and they are closely related to the symbol of generators of Markov processes (Jacob 2001; Böttcher *et al.* 2013). The choice of ρ_i and ψ_i is discussed in more detail in Remark 2.8, the standard choices are the Euclidean distance $\psi_i(t_i) = |t_i|$ and for $\alpha_i \in (0, 2)$ stable distances $\psi_i(t_i) = |t_i|^{\alpha_i}$ and bounded functions of the form $\psi_i(t_i) = 1 - \exp(-\delta_i |t_i|^{\alpha_i})$ with $\delta_i > 0$. But also other functions like Minkowski distances are possible, various examples are given in (Böttcher *et al.* 2018, Table 1).

We call a function ψ **characterizing** if for any random vector X the function $z \mapsto \mathbb{E}(\psi(X-z))$ characterizes the distribution of X uniquely, or equivalently, if for finite measures μ the function $\mu \mapsto \int \psi(x - .)\mu(dx)$ is injective. The following Proposition provides a characterization of the required support property of ρ in terms of ψ, it actually solves an open problem of Böttcher *et al.* (2018). Most notably, it also links the setting of multivariance to other dependence measures, see Section 3.

Proposition 2.2. *Let ψ_i be given by* (11) *for a symmetric measure ρ_i such that $\int 1 \wedge |t_i|^2 \rho_i(t_i) < \infty$. Then ψ_i is characterizing if and only if ρ_i has full topological support.*

Proof. The statement is a consequence of Theorem 9.1 (see Section 9). Hereto note that the distributions of two random vectors coincide if and only if their characteristic functions coincide on a dense subset, i.e., μ almost surely for a measure μ with full topological support. □

There are important scaled versions of the estimators in (8) and (9):

- **normalized sample (total) multivariance**: We write $\mathcal{M}$ instead of M, if each A_i in (8) and (9) is replaced by

$$\mathcal{A}_i := \begin{cases} \frac{1}{{}^N a_i} A_i & \text{if } {}^N a_i > 0 \\ 0 & \text{if } {}^N a_i = 0 \end{cases}, \text{ where } {}^N a_i := \frac{1}{N^2} \sum_{j,k=1}^{N} \psi_i(x_i^{(j)} - x_i^{(k)}) \text{ which estimates } \mathbb{E}(\psi_i(X_i - X_i')). \tag{12}$$

  In the case of normalized sample total multivariance the sum in (9) is additionally scaled by the number of summands in the definition of total multivariance (2), i.e.,

$${}^N\overline{\mathcal{M}}^2(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)}) := \frac{1}{2^n - n - 1} \left\{ \left[\frac{1}{N^2} \sum_{j,k=1}^{N} \prod_{i=1}^{n} (1 + (\mathcal{A}_i)_{jk}) \right] - 1 \right\}. \tag{13}$$

  By this scaling the test statistics for multivariance and total multivariance have expectation 1 (in the case of independent variables).

- **sample multicorrelation**: We write $\mathcal{R}$ instead of M, if each A_i in (8) is replaced by

$$\mathcal{B}_i := \begin{cases} \frac{1}{{}^N b_i} A_i & \text{if } {}^N b_i > 0 \\ 0 & \text{if } {}^N b_i = 0 \end{cases}, \quad \text{where } {}^N b_i := \left(\frac{1}{N^2} \sum_{j,k=1}^{N} |(A_i)_{jk}|^n \right)^{1/n} \tag{14}$$
$$\text{which estimates } \left(\mathbb{E}(|\Psi_i(X_i, X_i')|^n)\right)^{1/n}.$$

- **unnormalized sample multicorrelation**: We write Mcor instead of M, if each A_i in (8) is replaced by

$$\mathcal{C}_i := \begin{cases} \frac{1}{{}^N c_i} A_i & \text{if } {}^N c_i > 0 \\ 0 & \text{if } {}^N c_i = 0 \end{cases}, \quad \text{where } {}^N c_i := \left(\frac{1}{N^2} \sum_{j,k=1}^{N} (A_i)_{jk}^n \right)^{1/n} \tag{15}$$
$$\text{which estimates } \left(M^2_{\otimes_{k=1}^n \rho_i}(\underbrace{X_i, \dots, X_i}_{n\text{-times}}) \right)^{1/n}.$$

Note that Mcor is newly introduced in this paper, see in particular Table 1 for a comparison. For even n multicorrelation $\mathcal{R}$ and Mcor coincide, but for odd n they differ. Only $\mathcal{R}$ is always bounded by 1, hence Mcor is called *unnormalized*. But only for Mcor the value 1 has an explicit interpretation. The population versions of the above sample measures are given by scaling Ψ_i in (3) with the final terms of (12), (14) and (15), e.g. normalized multivariance and normalized total multivariance are given by

$$\mathcal{M}_\rho^2(X_1, \dots, X_n) = \frac{M^2(X_1, \dots, X_n)}{\prod_{i=1}^n \mathbb{E}(\psi_i(X_i - X_i'))} \quad \text{and} \quad \overline{\mathcal{M}}_\rho^2(X_1, \dots, X_n) = \frac{\mathbb{E}\left(\prod_{i=1}^n \left(1 + \frac{\Psi_i(X_i, X_i')}{\mathbb{E}(\psi_i(X_i - X_i'))}\right)\right) - 1}{2^n - n - 1}, \tag{16}$$

where implicitly the finiteness of the corresponding moments is assumed, i.e.,

$$\textbf{finite first } \boldsymbol{\psi}\textbf{-moments}\text{: for all } i = 1, \dots, n : \ \mathbb{E}(\psi_i(X_i)) < \infty. \tag{17}$$

For the scaling of the multicorrelations one has to assume

$$\textbf{finite } \boldsymbol{\psi}\textbf{-moments of order } \boldsymbol{n}\text{: for all } i = 1, \dots, n : \ \mathbb{E}(\psi_i^n(X_i)) < \infty. \tag{18}$$

Note that the scaling factors given in (14) and (15) depend on n, thus the corresponding total multicorrelations do not have such a simple representation as $\overline{\mathcal{M}}$ (or its sample version (13)) in fact the following holds (analogously also for Mcor):

$$\overline{\mathcal{R}}_\rho^2(X_1, \dots, X_n) := \frac{1}{2^n - n - 1} \sum_{\substack{1 \le i_1 < \ldots < i_m \le n \\ 2 \le m \le n}} \frac{M^2_{\otimes_{k=1}^m \rho_{i_k}}(X_{i_1}, \dots, X_{i_m})}{\prod_{k=1}^m \left(\mathbb{E}(|\Psi_{i_k}(X_{i_k}, X'_{i_k})|^m)\right)^{1/m}} \tag{19}$$

$$\ge \frac{\mathbb{E}\left(\prod_{i=1}^n \left(1 + \frac{\Psi_i(X_i, X_i')}{(\mathbb{E}(|\Psi_i(X_i, X_i')|^n))^{1/n}}\right)\right) - 1}{2^n - n - 1}. \tag{20}$$

Therefore the total multicorrelations seem more of a theoretic interest, but the corresponding m-multicorrelations (which will be defined in Remark 5.5.3) have efficient estimators. Moreover, also the lower bound in (20) has an efficient sample version analogously to (13). For a comparison of these multicorrelations see Section 3.6.

The introduced dependence measures and their sample versions feature the following properties.

Proposition 2.3 (Invariance properties of multivariance)**.** *The following properties hold for* $M, \overline{M}, \mathcal{M}, \overline{\mathcal{M}}, \mathcal{R}, \overline{\mathcal{R}}$, Mcor, $\overline{M\text{cor}}$ *and the corresponding sample versions, to avoid redundancy we only explicitly state them for* M*:*

(a) ***trivial for single variables****, i.e.,* $M_{\rho_i}(X_i) = 0$ *for all* $i \in \{1, \dots, n\}$*.*

(b) ***permutation invariant****, i.e.,* $M(X_1, \dots, X_n) = M_{\otimes_{i=1}^n \rho_{k_i}}(X_{k_1}, \dots, X_{k_n})$ *for all permutations* $k_1, \dots, k_n$ *of* $1, \dots, n$*. Moreover, the sample versions are in addition invariant with respect to permutations of the samples, i.e., the equality* ${}^N M(x^{(1)}, \dots, x^{(N)}) = {}^N M(x^{(l_1)}, \dots, x^{(l_N)})$ *holds for all permutations* $l_1, \dots, l_N$ *of* $1, \dots, N$*. (This should not be confused with the permutations for a resampling test, where components of the samples are permuted separately, see* (42)*.)*

(c) ***symmetric in each variable****, i.e.,* $M(X_1, \dots, X_n) = M(c_1 X_1, \dots, c_n X_n)$ *for all* $c_i \in \{-1, 1\}$*.*

(d) ***translation invariant****, i.e.,* $M(X_1 - r_1, \dots, X_n - r_n) = M(X_1, \dots, X_n)$ *for all* $r_i \in \mathbb{R}^{d_i}$*.*
Note that the latter and (c) imply that for dichotomous 0-1 coded data a swap of the coding does not change the value of the multivariance.

(e) ***rotation invariant for isotropic*** $\boldsymbol{\psi_i}$*, i.e., if* $\psi_i(x_i) = g_i(|x_i|)$ *for some* g_i *and all* $i = 1, \dots, n$*, then* $M(X_1, \dots, X_n) = M(R_1 X_1, \dots, R_n X_n)$ *for all rotations* R_i *on* $\mathbb{R}^{d_i}$*.*
Note that in this case, since also (c) and (d) hold, M *is invariant with respect to Euclidean isometries.*

Proof. For multivariance M the property (a) follows by direct calculation using (3) and (4), (b) is obvious by (3), (c) holds since ψ_i is symmetric and for (d) note that the translations cancel in (4). Moreover, since a rotation preserves Euclidean distances also (e) holds. Total multivariance $\overline{M}$ is just a sum of multivariances, hence it inherits these properties.

For sample multivariance ${}^N M$ the same arguments apply using (8) and (10). For the sample permutation invariance in (b) note that permutations of samples correspond to permutations of rows and columns of the centred distance matrices. Analogously also the scaling factors given in (12), (14) and (15) have these properties. Therefore the properties hold also for all scaled and sample versions of (total) multivariance. □

Moreover, for special functions ψ_i the scaled dependence measures feature also scale invariance.

Proposition 2.4 (Scale invariance of scaled multivariance for $\psi_i(x_i) = |x_i|^{\alpha_i}$)**.** *Let $\psi_i(x_i) = |x_i|^{\alpha_i}$ with $\alpha_i \in (0, 2)$. Then the scaled measures $\mathcal{M}, \overline{\mathcal{M}}, \mathcal{R}, \overline{\mathcal{R}}$, Mcor, $\overline{M\text{cor}}$ and the corresponding sample versions are scale invariant, that is,*

$$\mathcal{M}(r_1X_1, \ldots, r_nX_n) = \mathcal{M}(X_1, \ldots, X_n) \text{ for all } r_i \in \mathbb{R}\backslash\{0\}. \tag{21}$$

Proof. For $\psi_i(x_i) = |x_i|^{\alpha_i}$ note that Ψ_i given in (4) satisfies $\Psi_i(r_iX_i, r_iX_i') = |r_i|^{\alpha_i}\Psi_i(X_i, X_i')$. Thus multivariance is α-homogeneous, i.e., $M(r_1X_1, \ldots, r_nX_n) = M(X_1, \ldots, X_n)\prod_{i=1}^n |r_i|^{\alpha_i}$. The same holds (using (10)) for ${}^NM$ and also for the scaling factors given in (12), (14) and (15). Thus the factors $|r_i|^{\alpha_i}$ cancel by the scaling. □

The key for statistical tests based on multivariance is the following convergence result. The presented result relaxes the required moments considerably in comparison to (Böttcher *et al.* 2019, Thm. 4.5, 4.10, Cor. 4.16, 4.18), moreover also a new parameter β is introduced which will be useful in Section 6.

Theorem 2.5 (Asymptotics of sample multivariance)**.** *Let X_i, $i = 1, \ldots, n$ be non-constant random variables and let $\boldsymbol{X}^{(k)}$, $k = 1, \ldots, N$ be independent copies of $\boldsymbol{X} = (X_1, \ldots, X_n)$. Let either of the following conditions hold*

$$\text{all } \psi_i \text{ are bounded} \tag{22}$$

or

$$\text{for all } i = 1, \ldots, n: \ \mathbb{E}(\psi_i(X_i)) < \infty \text{ and } \mathbb{E}\left[(\log(1 + |X_i|^2))^{1+\varepsilon}\right] < \infty \text{ for some } \varepsilon > 0. \tag{23}$$

Then for any $\beta > 0$

$$N^\beta \cdot {}^N\mathcal{M}^2(\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(N)}) \xrightarrow[N\to\infty]{a.e.} \infty \qquad \text{if } X_1, \ldots, X_n \text{ are dependent but } (n-1)\text{-independent}, \tag{24}$$

$$N \cdot {}^N\mathcal{M}^2(\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(N)}) \xrightarrow[N\to\infty]{d} Q \qquad \text{if } X_1, \ldots, X_n \text{ are independent}, \tag{25}$$

$$N^\beta \cdot {}^N\overline{\mathcal{M}}^2(\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(N)}) \xrightarrow[N\to\infty]{a.e.} \infty \qquad \text{if } X_1, \ldots, X_n \text{ are dependent}, \tag{26}$$

$$N \cdot {}^N\overline{\mathcal{M}}^2(\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(N)}) \xrightarrow[N\to\infty]{d} \overline{Q} \qquad \text{if } X_1, \ldots, X_n \text{ are independent}, \tag{27}$$

where Q and $\overline{Q}$ are Gaussian quadratic forms with $\mathbb{E}Q = 1 = \mathbb{E}\overline{Q}$.

Proof. Here we explain the main new ideas, the details are provided in Section 9.3.

For the convergence in (25) and (27) exist at least two methods of proof: As in Böttcher *et al.* (2019) the convergence of empirical characteristic functions can be used. For this step a slightly relaxed (but technical) version of the log moment condition (see Remark 2.6) is necessary and sufficient, cf. (Böttcher *et al.* 2019, Remark 4.6.b). An alternative approach (Theorem 9.3 in Section 9) uses the theory of degenerate V-statistics, this requires moments of second order with respect to ψ_i, but no further condition. Thus, in particular, for bounded ψ_i the latter removes the log moment condition.

For $\beta = 1$ the divergence in (24) and (26) was proved in Böttcher *et al.* (2019) under the condition (5), which ensures that the representations (3) of the limits of sample (total) multivariance are well defined and finite. Using the characteristic function representation (1) (which is always well defined, but possibly infinite) the divergence can be proved without (5), see Section 9 Lemma 9.5 ff.. Moreover, the arguments used therein work for any $\beta > 0$. □

Remark 2.6. *The log moment condition $\mathbb{E}\left[(\log(1 + |X_i|^2))^{1+\varepsilon}\right] < \infty$ in* (23) *can be slightly relaxed to (Csörgő 1985, Condition* (*)*). But the latter is practically infeasible, thus we opted for a comprehensible condition. Moreover the log moment condition is trivially satisfied, if ψ_i satisfies a minimal growth, i.e., $\psi_i(x_i) \geq c\log(1+|x_i|^2)^{1+\varepsilon}$ for some $c, \varepsilon > 0$. Also the condition* (22) *is stated here for clarity, in fact* (A10) *is sufficient.*

Note that in Theorem 2.5 the parameter β was only considered in the dependent cases. In the case of independent random variables one obtains the following result.

Corollary 2.7 (Strong consistency of N^β-scaled multivariance for independent random variables). *Let X_i, $i = 1, \dots, n$ be independent random variables and let $\boldsymbol{X}^{(k)}$, $k = 1, \dots, N$ be independent copies of $\boldsymbol{X} = (X_1, \dots, X_n)$. If either* (22) *or* (23) *holds, then for any $\beta < 1$*

$$N^\beta \cdot {}^{N}\mathcal{M}^2(\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)}) \xrightarrow[N\to\infty]{a.e.} 0, \tag{28}$$

$$N^\beta \cdot {}^{N}\overline{\mathcal{M}}^2(\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)}) \xrightarrow[N\to\infty]{a.e.} 0. \tag{29}$$

Proof. The statements (28) and (29) are a direct consequence of (25) and (27), if one considers convergence in probability instead of 'a.e.', see e.g. (Böttcher *et al.* 2019, proof of Corollary 4.7) for the case $\beta = 0$.

For almost sure convergence one has to look at the proof(s) of (25) and (27). Therein a key step is an application of the central limit theorem, which requires (in the given setting) exactly the factor N for convergence (in distribution) to a standard normally distributed random variable. Using therein, for N replaced by N^β with $\beta < 1$, Marcinkiewicz's law of large numbers, e.g. (Kallenberg 1997, Theorem 3.23), (or the law of the iterated logarithm) yields the limit 0 almost surely. □

The choice of ψ_i is intertwined with the invariance properties (Propositions 2.3 and 2.4) and the moment conditions (22) and (23). For the population measures also condition (5) and for the scaled measures also (17) and (18) have to be considered. In particular, it is possible to choose ψ_i (or to transform the random variables) such that these conditions are satisfied regardless of the underlying distributions.

Remark 2.8.

1. *(Comments on choosing ψ) Based on Propositions 2.3 and 2.4 the canonical choice for ψ_i is $\psi_i(x_i) = |x_i|^{\alpha_i}$ with $\alpha_i \in (0, 2)$, classically with $\alpha_i = 1$ (other choices might provide higher power in tests; a general α_i selection procedure is to our knowledge not yet available).*

   Nevertheless there are many other options for ψ_i, see (Böttcher et al. 2018, Table 1) for various examples, and there are at least a few reasons why one might choose a ψ_i which is not (a power of) the Euclidean distance:

   (a) *For unbounded ψ_i condition* (23) *is required in Theorem 2.5. If the existence of these moments is unknown for the underlying distribution the convergence results might not hold. Here the use of a slower growing or bounded ψ_i is a safer approach, see Example 10.13.*

   (b) *The empirical size/power of the tests (details of theses are given in Section 4) can depend on the functions ψ_i used, see Example 10.11. Especially if some information on the dependence scale is known the parameter $\delta_i > 0$ in $\psi_i(x_i) = 1 - e^{-\delta_i |x_i|^{\alpha_i}}$ can be adapted accordingly, see (Böttcher et al. 2019, Example 5.2) for an example using multivariance. Adaptive procedures for δ_i can be found in Guetsop Nangue (2017).*

   (c) *A non-linear dependence of multivariance on sample distances might be desired, e.g. there might be application based reasons to use the Minkowski distance (Han et al. 2011, Section 2.4.4).*

   An alternative approach to ensure the moment conditions is the following.

2. *(Transformation to bounded random variables) Recall a basic result on independence: For i = 1, .., n let $X_i : \Omega \to \mathbb{R}^{d_i}$ be random variables and $f_i : \mathbb{R}^{d_i} \to D_i \subset \mathbb{R}^{s_i}$ be measurable functions, then:*

   $$X_i,\, i = 1, \dots, n \text{ are independent} \quad \Rightarrow \quad f_i(X_i),\, i = 1, \dots, n \text{ are independent.}$$

   Moreover, if $d_i = s_i$ and f_i are bijective then also the converse implication holds. Thus one way to ensure all moment conditions in Theorem 2.5 – and preserve the (in)dependence – is to transform the random variables by bounded (bounded D_i) bijective functions f_i. But beware that with this approach the multivariance is neither translation invariant nor homogeneous, cf. Example 10.13.

3 Comparison of multivariance to other dependence measures

In this section we compare multivariance to other dependence measures for random vectors $X_i \in \mathbb{R}^{d_i}$. We only consider dependence measures which are closely related, in the sense that they are also based on characteristic functions or appear as special cases. In the papers introducing and discussing these measures comparisons with further dependence measures can be found.

Recall that multivariance (squared), $M^2(X_1, \dots, X_n)$, is structurally of the form

$$\int \left| \mathbb{E}\left(\prod_{i=1}^{n} (\mathrm{e}^{\mathrm{i}X_i \cdot t_i} - f_{X_i}(t_i)) \right) \right|^2 \rho(dt) = \mathbb{E}\left[\prod_{i=1}^{n} \Psi_i(X_i, X_i') \right] \tag{30}$$

with Ψ_i given in (4) and $(X_1', \dots, X_n')$ being an independent copy of $(X_1, \dots, X_n)$.

3.1 Classical covariance, Pearson's correlation and the RV coefficient are limiting cases of multivariance

Let $n = 2$ and $\psi_i(x_i) = |x_i|^2$. Note that $|.|^2$ is not characterizing in the sense of Proposition 2.2. It actually does not correspond to a Lévy measure, cf. (Böttcher *et al.* 2018, Table 1). Thus the characteristic function representation (left hand side of (30)) does not hold and a value 0 of the right hand side does not characterize independence. Nevertheless, $|.|^2$ is a continuous negative definite function and it is the limit for $\alpha_i \uparrow 2$ of $|.|^{\alpha_i}$ which are valid functions for multivariance. Moreover, the right hand side of (30) is also for $|.|^2$ well defined, and it corresponds to classical linear dependence measures: Hereto denote by $X_{i,k}$ the components of the vectors X_i, i.e., $X_i = (X_{i,1}, \dots, X_{i,d_i})$ where $X_{i,k} \in \mathbb{R}$. By direct (but extensive calculation) the expectation representation in (30) of $M^2(X_1, X_2)$ with $\psi_i(x_i) = |x_i|^2 = \sum_{k=1}^{d_i} x_{i,k}^2$ simplifies to

$$\sum_{k=1}^{d_1} \sum_{l=1}^{d_2} (2\,\mathrm{Cov}(X_{1,k}, X_{2,l}))^2. \tag{31}$$

Especially for $d_1 = d_2 = 1$ the (absolute value of) classical covariance is recovered. Note that for $n = 2$ and $\psi_i(.) = |.|^2$ the scaling constants in (12), (14) and (15) become $2\,\mathrm{Var}(X_i)$, thus normalized multivariance coincides in this setting with both multicorrelations and with the absolute value of classical correlation. For arbitrary d_1 and d_2 the multicorrelations (squared) also coincide with the extension of correlation to random vectors developed in Escoufier (1973). The corresponding sample versions ${}^{N}\mathcal{M}$, ${}^{N}\mathcal{R}$ and ${}^{N}M\mathrm{cor}$ coincide for $d_1 = d_2 = 1$ with (the absolute value of) Pearson's correlation coefficient and ${}^{N}\mathcal{R}^2$ and ${}^{N}M\mathrm{cor}^2$ coincide for arbitrary d_1 and d_2 with the RV coefficient of Robert and Escoufier (1976) (see also Josse and Holmes 2016).

Note that also for $n > 2$ the right hand side of (30) with $\psi_i(x_i) = |x_i|^2$ is a well defined expression, which can be understood as an extension of covariance, Pearson's correlation and the RV coefficient to more than two random vectors.

3.2 Multivariance unifies distance covariance and HSIC

In the case of two random variables (that is, $n = 2$) multivariance coincides with generalized distance covariance (Böttcher *et al.* 2018) and the following (simplified) representations hold (using (Böttcher *et al.* 2018, Eq. (30)), direct calculations, (Josse and Holmes 2016, Eq. (3.2)) and the notation $\overline{\psi} = 1 - \psi$)

$$M^2(X_1, X_2) = \iint |f_{(X_1,X_2)}(t_1, t_2) - f_{X_1}(t_1) f_{X_2}(t_2)|^2 \rho_1(dt_1)\rho_2(dt_2) = \mathbb{E}\left[\prod_{i=1}^{2} \Psi_i(X_i, X_i') \right] \tag{32}$$

$$= \mathbb{E}\left[\prod_{i=1}^{2} (\overline{\psi_i}(X_i - X_i')) \right] - 2\mathbb{E}\left[\prod_{i=1}^{2} \mathbb{E}[\overline{\psi_i}(X_i - X_i') \mid X_i] \right] + \prod_{i=1}^{2} \mathbb{E}\left[\overline{\psi_i}(X_i - X_i') \right] \tag{33}$$

$$= \mathbb{E}\left[\prod_{i=1}^{2}\psi_i(X_i - X_i')\right] - 2\mathbb{E}\left[\prod_{i=1}^{2}\mathbb{E}[\psi_i(X_i - X_i') \mid X_i]\right] + \prod_{i=1}^{2}\mathbb{E}\left[\psi_i(X_i - X_i')\right] \tag{34}$$

$$= \operatorname{Cov}\left(\psi_1(X_1 - X_1'), \psi_2(X_2 - X_2')\right) - 2\operatorname{Cov}\left(\psi_1(X_1 - X_1'), \psi_2(X_2 - X_2'')\right). \tag{35}$$

The last line is included to emphasize that further interesting representations exist – this one actually provides a characterization of independence using (the classical linear dependence measure) covariance. Other equivalent representations are Brownian distance covariance (Székely and Rizzo 2009) (for $\psi(.) = |.|$) and its generalization Gaussian distance covariance (Böttcher *et al.* 2018, Section 7).

Note that (34) is for $\psi_i(x_i) = |x_i|^{\alpha_i}$ distance covariance (Székely *et al.* 2007) and (33) is for $\psi_i(x_i) = 1 - e^{-\delta_i\tilde{\psi}_i(x)}$ (where $\tilde{\psi}_i$ can be any real-valued continuous negative definite function, e.g. $|.|^{\alpha_i}$, and $\delta_i > 0$) the Hilbert Schmidt Independence Criterion (HSIC, (Gretton *et al.* 2008)) with kernel $k_i(x,y) = e^{-\delta_i\tilde{\psi}_i(x-y)}$.[3] For the latter just note that for any continuous positive definite function ϕ the function $\phi(0)-\phi$ is continuous negative definite (cf. (Jacob 2001, Corollary 3.6.10)), i.e., it fits into the framework of multivariance. The equivalence of kernel based approaches and distance based approaches was noted in Sejdinovic *et al.* (2013b), see also Shen and Vogelstein (2018) for a recent discussion. But note that the approach in Sejdinovic *et al.* (2013b) to the correspondence of kernels and distance functions only works for the case $n = 2$, whereas the above correspondence also extends to the multivariate setting.

In other words, in the case $n = 2$ multivariance with bounded measures ρ_i coincides with HSIC and special cases of unbounded ρ_i yield distance covariance. Therefore, in general, multivariance is an extension of these measures to more than two variables. But note that there is also at least one alternative extension as we will discuss in the next section.

As discussed in Remark 2.8, the cases with bounded measures have the advantage that most moment conditions are trivially satisfied and that in the case of HSIC the parameters δ_i provide a somehow natural bandwidth selection parameter. In contrast, using unbounded measures ρ_i corresponding to $|.|^{\alpha_i}$ provide (scaled) measures with superior invariance properties (Propositions 2.3 and 2.4). Note, that also in this case the parameters α_i offer some variability.

As a side remark, note that by the above it is straight forward that multivariance with $\tilde{\psi}_i$ is the derivative (in the bandwidth parameter at $\delta_i = 0$) of multivariance corresponding to $1 - e^{-\delta_i\tilde{\psi}_i(x)}$, this relation of distance covariance and HSIC was noted in Bilodeau and Guetsop Nangue (2017). Incidentally, it is also the key for relating Lévy processes to their generators, e.g. see the introduction of Böttcher *et al.* (2013).

Finally note that also the other measures discussed in the next section reduce for the case $n = 2$ to the above setting, thus they are included (or closely related as Jin and Matteson (2018), which considers a joint measure ρ without product structure).

3.3 Independence of more than two random vectors

As a consequence of Theorem 2.1 the multivariances of all subfamilies of the variables $X_1, \ldots, X_n$ characterize jointly their independence. In fact, this was suggested in Bilodeau and Guetsop Nangue (2017) as an approach to independence via multiple testing, i.e., via computing the p-value for each of these $2^n - n - 1$ multivariances separately. The approach is complementary to the global test using total multivariance.

In Bilodeau and Guetsop Nangue (2017) multivariance is considered in disguise: expanding the integrand of (30) and using the linearity of the expectation yields $\mathbb{E}\left(\prod_{i=1}^{n}(\mathrm{e}^{\mathrm{i}X_i\cdot t_i} - f_{X_i}(t_i))\right) = \sum_{S\subset\{1,\ldots n\}}\mathbb{E}(\prod_{i\in S}(\mathrm{e}^{\mathrm{i}X_i\cdot t_i})$ $\prod_{i\in S^c}(-f_{X_i}(t_i))$. This representation of the product is also called Möbius transformation of the characteristic

[3] HSIC (and dHSIC in Pfister *et al.* (2017)) require bounded, continuous, symmetric, positive definite kernels k_i. If k_i is additionally translation invariant, then $k_i(x_i, x_i') = k_i(x_i - x_i', 0) =: \phi(x_i - x_i')$ and ϕ is a continuous positive definite function. For the non translation invariant case see Section 3.5. Moreover, note that we assume here $\phi(0) = 1$ to avoid distracting constants in the presentation. HSIC and dHSIC additionally require that the kernel is characterizing, which is by Proposition 2.2 equivalent to the full support property of ρ.

functions. Without the characteristic function representation (with ψ_i based on kernels k_i) the multivariance of 3 random variables appeared before under the name "(complete) Lancaster interaction" in Sejdinovic *et al.* (2013a).

Other popular multivariate dependence measures based on characteristic functions are of the following form, which is here stated using our setting (with the notation $\overline{\psi} = 1 - \psi$ and $\rho_i(\mathbb{R}^{d_i}) = 1$; to reformulate it for positive definite kernels use the correspondence provided in Section 3.2):

$$\int \left| \mathbb{E}\left[\prod_{i=1}^{n} \mathrm{e}^{\mathrm{i}X_i\cdot t_i}\right] - \prod_{i=1}^{n} f_{X_i}(t_i)\right|^2 \rho(dt) \tag{36a}$$

$$= \mathbb{E}\left[\prod_{i=1}^{n}(\overline{\psi_i}(X_i - X_i'))\right] - 2\mathbb{E}\left[\prod_{i=1}^{n} \mathbb{E}[\overline{\psi_i}(X_i - X_i') \mid X_i]\right] + \prod_{i=1}^{n} \mathbb{E}\left[\overline{\psi_i}(X_i - X_i')\right]. \tag{36b}$$

Such dependence measures go back at least to (Kankainen 1995, (1.3)). It is important to note that the equality in (36) does not hold in general for unbounded measures ρ_i, e.g. for $n = 3$, X_1, X_2 dependent (satisfying (5)) and X_3 constant the term (36a) is infinite but (36b) is finite. Nevertheless, dependence measures of type (36) for $\rho = \otimes_{i=1}^{n}\rho_i$ with bounded and unbounded ρ_i were recently discussed in Fan *et al.* (2017) (in the unbounded case (Fan *et al.* 2017, Lemma 1a) provides only a rather complicated sample version, which actually corresponds to (36b), a proof can be found in Section 9.4), for finite ρ_i representation (36) corresponds to the also recently introduced measure dHSIC of Pfister *et al.* (2017) and for an unbounded (joint measure) ρ associated to $\psi(.) = |.|$ it was considered in Jin and Matteson (2018) (in this case (36b) has a slightly different form).

The above illustrates that also for measures derived via (36) various approaches can be unified using the framework of continuous negative definite functions and Lévy measures.

To compare (36) with multivariance, note that in (Böttcher *et al.* 2019, Section 3.5) it was shown that for any given multivariance there exist special kernels (beyond the restrictions of the above papers) which turn (36b) into multivariance. With the usual kernels the following holds: $\mathbb{E}\left(\prod_{i=1}^{n}(\mathrm{e}^{\mathrm{i}X_i\cdot t_i} - f_{X_i}(t_i))\right) = \mathbb{E}\left[\prod_{i=1}^{n} \mathrm{e}^{\mathrm{i}X_i\cdot t_i}\right] - \prod_{i=1}^{n} f_{X_i}(t_i)$ if the given random variables are $(n-1)$-independent (Böttcher *et al.* 2018, Corollary 3.3). Thus the left hand sides of (30) and (36) coincide in the case of $(n-1)$-independence. Without $(n-1)$-independence multivariance does not characterize independence, but total multivariance $\overline{M}(X_1, \ldots, X_n)$, given by

$$\sum_{\substack{S\subset\{1,\ldots,n\}\\|S|>1}} \int \left| \mathbb{E}\left(\prod_{i\in S}(\mathrm{e}^{\mathrm{i}X_i\cdot t_i} - f_{X_i}(t_i))\right)\right|^2 \underset{i\in S}{\otimes} \rho_i(dt_i) = \mathbb{E}\left(\prod_{i=1}^{n}(\Psi_i(X_i, X_i') + 1)\right) - 1, \tag{37}$$

does characterize independence.

The approach (36) and total multivariance (37) require similar moment conditions[4] (the variant in Jin and Matteson (2018) requires a joint first moment) and the computational complexity of the sample versions is similar (the variant in Jin and Matteson (2018) has a higher complexity, but they also provide an approximate estimator with the same complexity). Total multivariance needs one product of doubly centred distance matrices whereas (36b) needs three products of different distance matrices (which actually coincide with those used for the double centring). Nevertheless, both approaches differ: In the Section 9.5 we calculate explicitly the difference of the population measures for the case $n = 3$, indicating that it is by no means theoretically obvious which approach might be more advantageous. Here certainly further investigations are required. A practical difference is the fact that the current implementation of dHSIC (Pfister and Peters 2019) requires $N > 2n$, for multivariance there is no such explicit restriction.

Generally, papers on dependence measures differ not only in their measures, but also in their methods of testing. For the approach (36) various methods have been proposed, of which the resampling method seems

[4] Based on the method of proof and based on the focus of the papers (sample or population versions; bounded or unbounded ψ_i) the stated conditions differ. But it seems a reasonable guess that these can be unified to those of multivariance, cf. the discussion in the proof of Theorem 2.5.

most popular. For multivariance we introduce the resampling method in Section 4. Similarly to the other measures there are also further (and faster) methods available for multivariance: Distribution free tests are used in Böttcher *et al.* (2019) (see Theorem 4.4) and in Berschneider and Böttcher (2019) tests based on moments of the finite sample or limit distribution and/or using eigenvalues of the associated Gaussian process are developed, see also Guetsop Nangue (2017).

3.4 Pairwise independence

In Section 5 we introduce m-multivariance. In particular, 2-multivariance provides a global test for pairwise independence without any condition (when using bounded ψ_i) or under the mild moment condition (23), see Test 5.4. A related approach to pairwise independence using distance covariance was developed in Yao *et al.* (2017), but in contrast it required assumptions which are necessary for applications of a (generalized) central limit theorem. The methods are compared in Example 7.3.

3.5 Generalizations

The setting of HSIC and also extensions of distance covariance are applicable to more general spaces than $\mathbb{R}^d$. In this settings the representation via characteristic functions and the characterization of independence (might) fail. Nevertheless, the representations given in (3) can canonically be extend to negative definite kernels $n(x_i, x_i')$ replacing $\psi(x_i - x_i')$. Thus it seems a natural guess that the key properties required for testing can be recovered in this setting, but to our knowledge this has not been studied yet. In the bivariate setting (Lyons 2013) considers the case where $n(x_i, x_i')$ is a metric, hereto note that in general $\psi(x_i - x_i')$ does not yield a metric, but $\sqrt{\psi(x_i - x_i')}$ does, cf. (Böttcher *et al.* 2018, Remark 3.8.b)).

For distance covariance exist also further modifications, like the affinely invariant distance correlation in Edelmann (2015). Also this extension seems possible for multivariance. It is only defined for random vectors with non singular covariance matrices and in this setting it would be a candidate to satisfy the set of axioms given in the next section (Móri and Székely 2018, Example 3).

3.6 Axiomatic classification of dependence measures

Rényi (1959) proposed a set of axioms which a dependence measure should satisfy. These have been challenged over the years, most recently e.g. in Móri and Székely (2018). They propose "four simple axioms" which a dependence measure d should satisfy, and distance correlation is called the "simplest and most appealing" measure which satisfies these axioms. All axioms were proposed for pairwise comparisons of random variables or vectors. We present here a multivariate extension to n non-constant random vectors (constants are removed to avoid technical difficulties, cf. Móri and Székely (2018)):

(A1) *characterization of independence*: $d(X_1, \dots, X_n) = 0$ if and only if X_i are independent.

(A2) *invariance*: $d(X_1, \dots, X_n) = d(S_1(X_1), \dots, S_n(X_n))$ for all similarity transforms[5] S_i.

(A3) *reference value*: $d(X_1, \dots, X_n) = 1$ if $X_1, \dots, X_n$ are related by similarity transforms (see (38) for details).

(A4) *continuity*: $d(X_1^{(k)}, \dots, X_n^{(k)}) \xrightarrow{k\to\infty} d(X_1, \dots, X_n)$, if $(X_1^{(k)}, \dots, X_n^{(k)}) \xrightarrow{k\to\infty} (X_1, \dots, X_n)$ in distribution (under a uniform moment condition, which ensures the finiteness of the measures).

[5] A **similarity transform** is any combination of translations, rotations, and reflections and non zero scalings (using the same scaling factor for all components of a vector), cf. Móri and Székely (2018).

Note that Móri and Székely (2018) uses a further common axiom — *normalization*: $d(\dots) \in [0, 1]$ — which was only indirectly assumed and (A3) was stronger: it contained “if and only if” with a seemingly more restrictive relation which actually forced explicitly the dimensions of the random vectors to be identical. Note that in the related (original) axiom (Rényi 1959, Axiom E) also only the “if” part was required and a footnote explicitly advised against strengthening it.

In the setting of multivariance we say that random variables X_i and X_k are related by similarity transforms S_i and S_k if

$$\psi_i(S_i(X_i) - S_i(X_i')) = \psi_k(S_k(X_k) - S_k(X_k')). \tag{38}$$

A prerequisite for the continuity (A4) is the finiteness of the measure d, cf. Móri and Székely (2018). Thus all considerations for (normalized) multivariance are under the moment condition (5) and for the multicorrelation we have to assume (18). Based on Propositions 2.3 and 2.4 the invariance with respect to similarity transforms holds for $\psi(x) = |x|^\alpha$, and it seems (cf. (Böttcher *et al.* 2018, Section 5)) that for other unbounded and all bounded ψ the invariance fails. Therefore we only consider $\psi(x) = |x|^\alpha$. Table 1 indicates which axioms are satisfied by the measures, all properties follow by direct calculations (the continuity uses the dominated convergence theorem; for the normalization a generalized Hölder inequality is used, see also (Böttcher *et al.* 2019, Proposition 4.13)). For multicorrelation the properties vary as the number of variables is even or odd, and $\overline{\mathcal{R}}$ yields always a measure with values in $[0, 1]$ whereas Mcor yields always the reference value 1 for variables related by similarity transforms. Note that for a multivariate normal distribution the value of total distance multivariance is (for the special case $\psi(x) = |x|$) linked to its correlation by (Chakraborty and Zhang 2019, Proposition 2).

By Table 1 the four axioms are simultaneously satisfied by $\overline{M\text{cor}}$. But recall that $\overline{\mathcal{R}}$ and $\overline{M\text{cor}}$ lack efficient sample versions. In the sample setting also $N \cdot {}^N\mathcal{M}^2$ and $N \cdot {}^N\overline{\mathcal{M}}^2$ provide statistically interpretable values (indirectly: via the corresponding p-value; yielding also a rough direct interpretation: they are positive and their expectation is one for independent random variables. Thus values much larger than one hint at dependence). Moreover, normalized multivariance requires only the moment condition (5) whereas multicor-

Table 1: Dependence measure axioms which are satisfied by (variants) of (total) multivariance for $\psi_i(x_i) = |x_i|^{\alpha_i}$ with $\alpha_i \in (0, 2)$.

axioms	(A1) characterization of independence	(A2) invariance	(A3) reference value	(A4) continuity	normalization
multivariance					
M	$n = 2$	–	–	✓	–
$\overline{M}$	✓	–	–	✓	–
normalized multivariance					
$\mathcal{M}$	$n = 2$	✓	–	✓	–
$\overline{\mathcal{M}}$	✓	✓	–	✓	–
multicorrelation					
$\mathcal{R}$	$n = 2$	✓	n even	✓	✓
$\overline{\mathcal{R}}$	✓	✓	$n = 2$	✓	✓
Mcor	$n = 2$	✓	✓	✓	n even
$\overline{M\text{cor}}$	✓	✓	✓	✓	$n = 2$
2-multivariance	char. of pairwise independence		(A3) and iff		
(Section 5)					
M_2	✓	–	–	✓	–
$\mathcal{M}_2$	✓	✓	–	✓	–
$M\text{cor}_2$ (= $\mathcal{R}_2$)	✓	✓	✓	✓	✓

relation requires the more restrictive condition (18). Finally, note that in the case $n = 2$ the multicorrelations coincide. Thus, in particular, $M\mathrm{cor}_2$ (defined in Section 5) provides a measure with efficient sample estimator. For this measure a value of 0 only characterizes pairwise independence, but the value 1 occurs if and only if the random variables are related by similarity transforms.

A first discussion of the behaviour of (total) multivariance when one enlarges the family of random variables can be found in (Böttcher *et al.* 2019, Proposition 3.7, Remark 3.8), which translates directly to multicorrelation.

4 Testing independence using multivariance

In this section we extend the discussion of (Böttcher *et al.* 2019, Section 4.5). We use the notation of Section 2, in particular $\boldsymbol{x}^{(k)} = (x_1^{(k)}, \dots, x_n^{(k)})$ are samples of independent copies of $(X_1, \dots, X_n)$. Based on Theorem 2.5, and recalling the fact that constant random variables are always independent, the following structure of a test for independence is obvious.

Test 4.1 (Test for n-independence, given $(n-1)$-independence)**.** *Let ψ_i be bounded or* (23) *be satisfied. Then a test for independence is given by: Reject n-independence if $X_1, \dots, X_n$ are $(n-1)$-independent and*

$$N \cdot {}^N\mathcal{M}^2(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)}) > R. \tag{39}$$

The value R will be discussed below.

Remark 4.2. *Note that also without the assumption of* $(n-1)$*-independence* (39) *provides a test for independence for which the type I error can be controlled by the choice of R, since the distribution of the test statistic under the hypothesis of independence is known, see* (25)*. But in this case it is unknown if the test statistic diverges if the hypothesis does not hold. Thus one can not control the Type II error and it will not be consistent against all alternatives (regardless of the satisfied moment conditions). A trivial example hereto would be the case where one random variable is constant, and thus the test statistic is always 0. But note that with the assumption of* $(n-1)$*-independence this problem does not appear, since the* $(n-1)$*-independence implies (given that at least one random variable is constant) that the random variables are independent.*

Analogous to Test 4.1, using total multivariance instead of multivariance, one gets the test for independence.

Test 4.3 (Test for (n-)independence)**.** *Let ψ_i be bounded or* (23) *be satisfied. Then a test for independence is given by: Reject independence if*

$$N \cdot {}^N\overline{\mathcal{M}}^2(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)}) > R. \tag{40}$$

To get a test with significance level $\alpha \in (0,1)$ the natural choice for R in (39) and (40) is the $(1-\alpha)$-quantile of the (limiting) distributions of the test statistics under H_0, i.e., assuming that the X_i are independent. To find this distribution explicitly or at least to have good estimates is non trivial, see Berschneider and Böttcher (2019) for an extensive discussion. As a starting point, one can follow (Székely *et al.* 2007, Theorem 6) where a general estimate for quadratic forms of Gaussian random variables given in Székely and Bakirov (2003) is used to construct a test for independence based on distance covariance. In our setting this directly yields the following result.

Theorem 4.4 (Rejection level for the distribution-free tests)**.** *Let $\alpha \in (0, 0.215]$. Then Test 4.1 and 4.3 with*

$$R := F_{\chi_1^2}^{-1}(1-\alpha) \tag{41}$$

are (conservative) tests with significance level α. Here $F_{\chi_1^2}$ is the distribution function of the Chi-squared distribution with one degree of freedom.

In the case of univariate Bernoulli random variables the significance level α is achieved (in the limit) by Test 4.1 with R given in (41), see (Berschneider and Böttcher 2019, Remark 4.27). But for other cases it might be very conservative, e.g. Example 10.11 (Figure A13). Recall that total multivariance is the sum of $2^n - n - 1$ distance multivariances (this is the number of summands in (2)). Thus one distance multivariance with a large value might be averaged out by many small summands, see Example 10.14. Hereto m-multivariance (which will be introduced in the next section) provides an intermediate remedy. It is also the sum of multivariances, but it has less summands. Thus the 'averaging out' (also known as 'statistical curse of dimension') will be still present but less dramatic.

Note that R in Theorem 4.4 is provided by a general estimate for quadratic forms. It yields in general conservative tests, since it does not consider the specific underlying (marginal) distributions. Less conservative tests can be constructed if the distributions are known or by estimating these distributions. The latter can be done by a resampling approach or by a spectral approach, similarly to the case of distance covariance (see (Sejdinovic *et al.* 2013b, Section 7.3.)). Methods related to the spectral approach are developed in Berschneider and Böttcher (2019).

In the following the resampling approach for $\mathcal{M}$ is introduced. The procedure is certainly standard to experts, never the less it seems important to recall it (to avoid ambiguity): Suppose we are given i.i.d. samples[6] $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}$ with unknown dependence, i.e., for each i the dependence of the components $x_1^{(i)}, \ldots, x_n^{(i)}$ is unknown. Now, resampling each component separately yields (almost) independent components. Thus Test 4.1 (respectively Test 4.3 with $\overline{\mathcal{M}}$) becomes a **resampling test** (resampling without replacement / permutation test) with $L \in \mathbb{N}$ replications using the rejection level R given by

$$R_{\text{rs}} := Q_{1-\alpha}\left(\left\{N \cdot {}^{N}\mathcal{M}^2\left(x_1^{(p_1^{(l)}(i))}, \ldots, x_n^{(p_n^{(l)}(i))},\ i = 1, \ldots, N\right),\quad l = 1, \ldots, L\right\}\right) \tag{42}$$

where each $p_k^{(l)}(1), \ldots, p_k^{(l)}(N)$ is a random permutation of $1, \ldots, N$ (and these are i.i.d. for $k = 1, \ldots, n$ and $l = 1, \ldots, L$) and $\boldsymbol{x}^{(i)} = (x_1^{(i)}, \ldots, x_n^{(i)})$ are the samples given for the test. Here $Q_{1-\alpha}(S)$ denotes the empirical $(1-\alpha)$-quantile of the samples in the set S. Instead of random permutations one could allow $p_k^{(l)}(1), \ldots, p_k^{(l)}(N)$ to be any sample of $1, \ldots, N$, this would also be a resampling test (resampling with replacement / bootstrap test), but note that the permutation test can be implemented more efficiently. Similarly, Test 4.1 (respectively Test 4.3 with $\overline{\mathcal{M}}$) becomes a **Monte Carlo test** with $L \in N$ replications using

$$R_{\text{MC}} := Q_{1-\alpha}\left(\left\{N \cdot {}^{N}\mathcal{M}^2(x_1^{(i,l)}, \ldots, x_n^{(i,l)},\ i = 1, \ldots, N),\quad l = 1, \ldots, L\right\}\right) \tag{43}$$

where $x_k^{(i,l)}$, $k = 1, \ldots, n$, $i = 1, \ldots, N$, $l = 1, \ldots, L$ are independent samples and for each fixed k the $x_k^{(i,l)}$, $i = 1, \ldots, N$, $l = 1, \ldots, L$ are i.i.d. samples of X_k.

Remark 4.5. *In (Pfister et al. 2017, Section 3.2) two related resampling tests are introduced for dHSIC. But note that they use slightly different terminology, i.e., therein the 'permutation test' considers samples as in* (42) *but instead of random permutations all permutations are considered. For the 'bootstrap test' they use all resamplings of the sample distribution of each variable. This yields $L = (N!)^n$ and $L = N^{Nd}$, respectively. Which is infeasible even for relatively small N, thus in (Pfister et al. 2017, Section 4.2) they also use randomly selected samples instead of all samples, and they call the resulting estimators 'Monte-Carlo approximations' of the estimators.*

5 *m*-multivariance

Pairwise independence is the prime requirement for various fundamental tools in stochastics, e.g. the classical law of large numbers. Especially when working with many variables (n large) a multiple testing approach

[6] Here we use a common abuse of terminology: An *independent sample* is a sample based on independent random variables. Analogously, an i.i.d. (independent and identically distributed) sample, is a sample of i.i.d. random variables. Moreover, note that here the random variables are in general random vectors with possibly dependent components.

might not be feasible. Thus a global test for pairwise independence has many applications, see also the motivation in Yao *et al.* (2017). Here we construct such a test, together with further generalizations which allow the successive testing of 2-independence, 3-independence, etc.

Define for $m \in \{2, \dots, n\}$ the ***m*-multivariance** M_m by

$$M^2_{m,\rho}(X_1, \dots, X_n) := \sum_{1 \le i_1 < \ldots < i_m \le n} M^2_{\otimes_{k=1}^m \rho_{i_k}}(X_{i_1}, \dots, X_{i_m}). \tag{44}$$

Instantly Theorem 2.1 yields the following characterization.

Proposition 5.1 (Characterization of m-independence)**.** *For random variables $X_1, \dots, X_n$ taking values in $\mathbb{R}^{d_1}, \dots, \mathbb{R}^{d_n}$ the following are equivalent:*

1. *$X_1, \dots, X_n$ are m-independent,*
2. *$M_{m,\rho}(X_1, \dots, X_n) = 0$ and $X_1, \dots, X_n$ are $(m-1)$-independent.*

In particular, $M_2(X_1, \dots, X_n) = 0$ characterizes pairwise independence.

Similar to multivariance, using (8), a strongly consistent estimator for M_m is the **sample m-multivariance**

$$^N\!M_m(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)}) = \sqrt{\sum_{1 \le i_1 < \ldots < i_m \le n} \frac{1}{N^2} \sum_{j,k=1}^{N} (A_{i_1})_{jk} \cdot \ldots \cdot (A_{i_m})_{jk}}. \tag{45}$$

Analogous to the case of normalized total multivariance the **normalized sample m-multivariance** $^N\mathcal{M}_m$ is given by

$$^N\mathcal{M}^2_m(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)}) = \binom{n}{m}^{-1} \sum_{1 \le i_1 < \ldots < i_m \le n} \frac{1}{N^2} \sum_{j,k=1}^{N} (\mathcal{A}_{i_1})_{jk} \cdot \ldots \cdot (\mathcal{A}_{i_m})_{jk}, \tag{46}$$

where $\mathcal{A}_i$ are the normalized matrices defined in (12). For (sample) m-multivariance the invariance properties (Propositions 2.3 and 2.4) hold analogously. To ensure that the expectation representation of m-multivariance (analogous to (3)) is finite the following condition (which is weaker than (5)) is required:

finite joint ψ-moments for families of size m:

$$\text{for all } S \subset \{1, \dots, n\} \text{ with } |S| \le m\text{: } \mathbb{E}\left(\prod_{i \in S} \psi_i(X_i)\right) < \infty. \tag{47}$$

Note that the sum $\sum_{1 \le i_1 < \ldots < i_m \le n}$ has $\binom{n}{m}$ summands, which might be a lot to compute. These sums can be simplified using the multinomial theorem, $(A_1 + \ldots + A_n)^m = \sum_{k_1 + \ldots + k_n = m} \frac{m!}{k_1! \cdot \ldots \cdot k_n!} A_1^{k_1} \cdot \ldots \cdot A_n^{k_n}$. In particular, for $m = 2, 3$ the following expressions of sample m-multivariance are easier to evaluate (analogous representations hold for the normalized sample m-multivariance):

$$^N\!M_2(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)}) = \sqrt{\frac{1}{2}\frac{1}{N^2} \sum_{k,l=1}^{N} \left(\left((A_1 + \ldots + A_n)_{kl}\right)^2 - \sum_{i=1}^{n} \left((A_i)_{kl}\right)^2 \right)}, \tag{48}$$

$$^N\!M_3(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N)}) = \sqrt{\frac{1}{3}\frac{1}{N^2} \sum_{k,l=1}^{N} \left(\left(\left(\sum_{i=1}^{n} A_i\right)_{kl} \right)^3 - 3 \left(\sum_{i=1}^{n} A_i\right)_{kl} \sum_{i=1}^{n} \left((A_i)_{kl}\right)^2 + 2 \sum_{i=1}^{n} \left((A_i)_{kl}\right)^3 \right)}. \tag{49}$$

Thus at least for small m these estimators are easy to compute and – analogous to the case of (total) multivariance – these can be used to test m-independence by the next results.

Theorem 5.2. *(Asymptotics of sample m-multivariance) Let X_i, $i = 1, \dots, n$ be non-constant random variables and let $\boldsymbol{X}^{(k)}$, $k = 1, \dots, N$ be independent copies of $(X_1, \dots, X_n)$. If either the ψ_i are bounded or* (23) *holds, then for $m \le n$*

$$N \cdot {}^N\mathcal{M}^2_m(\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)}) \xrightarrow[N \to \infty]{d} Q \qquad \text{if } X_1, \dots, X_n \text{ are } m\text{-independent}, \tag{50}$$

$$N \cdot {}^{N}\mathcal{M}_m^2(\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(N)}) \xrightarrow[N\to\infty]{a.e.} \infty \qquad \text{if } X_1, \ldots, X_n \text{ are } m\text{-dependent but } (m-1)\text{-independent.} \tag{51}$$

where Q is a Gaussian quadratic form with $\mathbb{E}Q = 1$.

Proof. Let the assumptions of Theorem 5.2 be satisfied. Then (50) holds, since in this case (25) implies the convergence of each of the $\binom{n}{m}$ summands of (46) to a Gaussian quadratic form with expectation 1. Thus, due to the normalizing factor in (46), the limiting distribution has expectation 1. Further note that given (23) all these quadratic forms can be expressed as a stochastic integral with respect to the same process, cf. (Böttcher *et al.* 2019, Supplement, Eq. (S.15)). This yields (by the same arguments as in the case of total multivariance (Böttcher *et al.* 2019, Section 4.3)) that the limiting distribution is in fact the distribution of a Gaussian quadratic form. If all ψ_i are bounded, a proof analogous to the one for the convergence of total multivariance in Theorem 9.3 shows the result.

The divergence (51) follows by (24), since the latter implies under the given assumptions that at least one summand of (46) diverges. □

Analogous to the cases of multivariance and total multivariance the above theorem immediately yields a test for m-independence which is (under the given moment conditions) consistent against all alternatives.

Test 5.3 (Test for m-independence, given $(m-1)$-independence)**.** *If either the ψ_i are bounded or* (23) *holds, then a test for m-independence is given by: Reject m-independence if $X_1, \ldots, X_n$ are $(m-1)$-independent and*

$$N \cdot {}^{N}\mathcal{M}_m^2(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}) > R, \tag{52}$$

with R as discussed in Section 4. (Note that one has to replace $\mathcal{M}$ by $\mathcal{M}_m$ in (42) *and* (43) *to get R for the resampling test and the Monte Carlo test, respectively.)*

For a test of m-independence (without controllable type II error) one can drop in Test 5.3 the assumption of $(m-1)$-independence, cf. Remark 4.2.

As a special case, for $m = 2$, the Test 5.3 becomes a test for pairwise independence.

Test 5.4 (Test for pairwise independence)**.** *If either the ψ_i are bounded or* (23) *holds, then a test for pairwise independence is given by: Reject pairwise independence if*

$$N \cdot {}^{N}\mathcal{M}_2^2(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}) > R, \tag{53}$$

with R as discussed in Section 4. (Note that one has to replace $\mathcal{M}$ by $\mathcal{M}_2$ in (42) *and* (43) *to get R for the resampling test and the Monte Carlo test, respectively.)*

Examples of the use of m-multivariance are given in the Sections 7 and 10, e.g. Example 7.3. To roundup this section we discuss some related estimators.

Remark 5.5.

1. *Analogous to the case of total multivariance one can define **total m-multivariance** for $\boldsymbol{X} = (X_1, \ldots, X_n)$ by*

$$\overline{M}_{m,\rho}^2(\boldsymbol{X}) := \sum_{\substack{1\le i_1<\ldots<i_r\le n\\ 2\le l\le m}} M^2_{\bigotimes_{k=1}^{l}\rho_{i_k}}(X_{i_1}, \ldots, X_{i_l}) = \sum_{l=2}^{m} M_{l,\rho}^2(\boldsymbol{X}) \tag{54}$$

   and calculate its sample version. There might be computationally simpler representations using formulas for $(A_1 + \ldots + A_n + 1)^m$. Moreover, also the complements of these measures, e.g. $\overline{M} - M_3 - M_2 = \overline{M} - \overline{M}_3$, might be of interest for multiple testing of higher order dependencies with disjoint hypotheses.

2. *The simple form of the sample* 2*-multivariance in* (48) *might suggest other generalizations. For example one could also consider*

$${}^{N}\widetilde{M}_3(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}) := \sqrt{\frac{1}{2}\frac{1}{N^2}\sum_{k,l=1}^{N}\left(\left((A_1 + \ldots + A_n)_{kl}\right)^3 - \sum_{i=1}^{n}\left((A_i)_{kl}\right)^3\right)} \tag{55}$$

as an estimator for 3-independence. In fact in the case of 2-independence this provides (assuming (47) *and using (Böttcher et al. 2019, Corollary 4.7)) a weakly consistent estimator for $\mathcal{M}_3$. Hereto just note that the sums of all mixed terms of the form $((A_i)_{kl})^2(A_j)_{kl}$ with $i \neq j$ are estimators for multivariances like $\mathcal{M}(X_i, X_i, X_j)$, and the factorization for independent subsets* (6) *yields $\mathcal{M}(X_i, X_i, X_j) = \mathcal{M}(X_i, X_i)\mathcal{M}(X_j) = 0$. But note that the estimators for these terms squared and scaled by N do usually not vanish for $N \to \infty$. Thus a result like Theorem 5.2 fails to hold.*

3. *A further natural extension is to introduce the corresponding global scaled measures of m-dependence, i.e. m-multicorrelations. These require finite ψ-moments of order m (cf.* (18)*). E.g.* ***2-multicorrelation*** *is given by*

$$\mathcal{M}\mathrm{cor}_{2,\rho}(X_1, \dots, X_n) := \sqrt{\binom{n}{2}^{-1} \sum_{1\le i<j\le n} \frac{M^2_{\rho_i\otimes\rho_j}(X_i, X_j)}{\sqrt{M^2_{\rho_i\otimes\rho_i}(X_i, X_i)M^2_{\rho_j\otimes\rho_j}(X_j, X_j)}}} \tag{56}$$

and it coincides with the (analogously defined) $\mathcal{R}_2$ since for $n = 2$ the scaling factors in (14) *and* (15) *coincide. Moreover these factors have for each summand in* (56) *the same exponent, thus (in contrast to $\overline{\mathcal{R}}$ and $\overline{\mathcal{M}\mathrm{cor}}$) one gets efficient sample representations by replacing the A_i in* (48) *by those in* (14) *(or equivalently* (15)*). This correlation satisfies all the dependence measure axioms of Section 3.6 when one replaces (A1) by the characterization of pairwise independence, see Table 1.*

6 Dependence structure visualization and detection

In this section two visualizations of higher order dependencies of random variables $X_1, \dots, X_n$ using undirected graphs are introduced: the full and the clustered dependence structure. For each the population version and estimation procedures are discussed. The latter can be based on independence tests with a fixed significance level or on a consistent estimator. In Sections 7 and 10 various examples are presented. The implementation of the visualizations in R relies in particular on the package `igraph` (Csardi and Nepusz 2006).

A dependence structure graph consists of three elements (cf. Figure 2):

- **Circled nodes** denote random variables.
- **Edges** denote dependencies.
- **Non-circled nodes** ('dependency nodes') are primarily used to denote the dependence of the connected nodes and a label might denote the value of a dependence measure or a related quantity of the connected nodes (in our sample setting it is the value of the test statistic $N \cdot {}^N\mathcal{M}^2$ or the order of dependence). Secondarily these might be used to represent the 'random variable' which consists of all components of a connected cluster, e.g. in Figure 2 the node with label '97.1' represents the cluster of X_1, X_2, X_{11}.

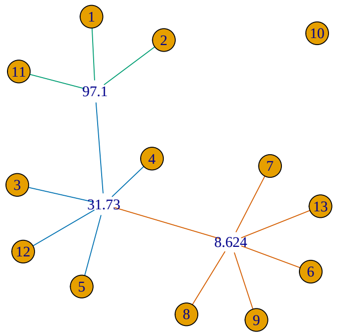


Figure 2: Visualized clustered dependence structure based on samples of Example 10.5.

A visualization of the **full dependence structure** is constructed by adding the corresponding 'dependency nodes' and edges for any m-tuple of $X_i, \dots, X_n$ which is m-dependent but $(m-1)$-independent. In general this graph can be very overloaded, see Example 10.9.

The direct approach to the full dependence structure based on samples is to test successively all $(m-1)$-independent m-tuples for m-independence for $m = 2, \dots, n$, adjust the p-values appropriately for multiple testing and add the significant dependency nodes and edges. For such a test procedure a direct visualization of the tests p-values was introduced in Genest and Rémillard (2004): the dependogram. Note that the full dependence structure visualizes the (lowest order) significant findings in a dependogram, see Example 7.1 for more details. In practice a visualization of the full dependence structure is only feasible for small n, since for n random variables there are $2^n - n - 1 = \sum_{k=2}^{n} \binom{n}{k}$ tuples to consider.

To overcome (or at least to reduce) the drawbacks of the full dependence structure one can alternatively use a clustered dependence structure. Hereto each set of connected vertices in an undirected graph will be called cluster. Then the **clustered dependence structure** graph is constructed by the following algorithm:

0. Include the circled nodes for $X_1, \dots, X_n$.
1. Interpret clusters as random vectors: Let k be the number of clusters currently in the graph and let $Y_i,\ i = 1, \dots, k$ be random variables which have as components the connected X_j of cluster i, e.g. $Y_1 = (X_1, X_2, X_{11})$ if X_1, X_2, X_{11} are connected via some edges. (In the very first run each random variable is its own cluster, i.e. $k := n$ and $Y_i := X_i$.) Moreover, set $m = 2$ (order of tuples to be tested next).
2. Add successively edges (and dependency nodes) for dependent clusters: If $m > k$ the graph construction is finished, otherwise: For all m-dependent subsets of $Y_1, \dots, Y_k$ add the corresponding dependency nodes and edges (connected to some non-circled node representing the cluster, if the cluster consists of more than one random variable) to the graph. If new nodes were introduced, go to step 2 otherwise repeat this step with m increased by one.

Since dependence and independence are not transitive, some information might be lost in the clustered dependence structure. Nevertheless, note that clustering preserves dependence, e.g. if at least one of the random variables $X_i, i \in I$ is dependent with one of $X_k, k \in K$ then also $(X_i, i \in I \cup J)$ is dependent with $(X_k, k \in K \cup L)$.

The visualization algorithm for a clustered dependence structure based on samples is analogous to the above, just in step 2 the m-independence has to be tested. Here one can (we do so) choose to skip sets of variables which have been tested before, i.e., sets which remained unchanged after the last cluster detection. But in any case the p-values have to be adjusted appropriately for multiple testing.

The ***appropriate*** **adjustment of p-values** due to multiple testing is the basis for many debates. For the full dependence structure and for the clustered dependence structure the situation is complicated by the fact that the total number of tests is unknown at the beginning, and the result of the tests in one step influence (by indicating that some tuples are lower order dependent or by clustering) the data for the tests thereafter. Thus adjusting p-values after clustering would usually require new tests. An approach which avoids this uses Holm's method separately for each set of multiple tests in step 2 of the algorithm, but one has to keep in mind that by this the global type I error bound increases with each set of tests.

In general one might also distinguish between visualizations of the results of **tests using a given significance level** (in this case there is a bound for a type I error based on the significance level and it depends also on the correction for multiple testing used) or visualizations using **consistent estimators** (if these exist they might also be based on tests, but then the significance level or rejection level is adapted based on the sample size, which might make it harder to get explicit error estimates). In the case of tests with a fixed significance level, the range of significance levels which yield the same results might give an additional indication of the *reliability* of the detection.

Remark 6.1 (Comment on detection errors). *The probability of a type I error can be estimated and/or bounded by the choice of the rejection level or significance level. But a type II error bound or estimate might not be available. In this case one has to keep in mind that, due to the successive estimation/testing procedure a type II error (i.e., a not detected dependence) for some tuple can yield a detected higher order dependence for a superset of the tuple. Thus in this case the higher order dependence still indicates that the components of the tuple are not independent (but they might not be lower order independent).*

All of the above applies to the use of any multivariate dependence measure or test in the dependence structure detection algorithms. Now we turn explicitly to the case of multivariance.

6.1 Dependence structure detection using distance multivariance

For consistent estimates using multivariance the following observation is essential, it is based on Theorem 2.5 and Corollary 2.7.

Corollary 6.2 (Consistent dependence estimation). *Let X_i, $i = 1, \ldots, n$ be $(n-1)$-independent non-constant random variables and let $\boldsymbol{X}^{(k)}$, $k = 1, \ldots, N$ be independent copies of $(X_1, \ldots, X_n)$. If either the ψ_i are bounded or (23) holds, then for any $\beta \in (0,1)$*

$$N^\beta \cdot \mathcal{M}^2(\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(N)}) \xrightarrow[N\to\infty]{a.e.} \begin{cases} \infty & \text{if } X_i, \ldots, X_n \text{ are dependent}, \\ 0 & \text{if } X_i, \ldots, X_n \text{ are independent}. \end{cases} \tag{57}$$

Hence using $R = R(N) := N^{1-\beta} \cdot C$ for any fixed constants $\beta \in (0,1)$, $C > 0$ in the independence Tests 4.1, 4.3, 5.3, 5.4 provides strongly consistent tests, in the sense that (under the assumptions of these tests) the test result converges almost surely to the correct statement as $N \to \infty$. Clearly the convergence speed of the estimator depends on the choice of β and C (see below for a rough error estimate).

Therefore there are several options for the dependence structure detection using Test 4.1:

(a) **conservative / distribution free**: the very fast but conservative rejection level given in Theorem 4.4,
(b) **resampling**: the slow but approximately sharp rejection level provided by the resampling approach (42) (or by the Monte Carlo approach (43)),
(c) **consistent**: the value $R := N^{1-\beta} \cdot C$ for tests corresponding to the consistent estimator.

Of the above options (a) and (b) provide (almost) directly also the corresponding p-values; but only (a) and (c) are feasible, since for the resampling approach (b) the sample size would have to be adapted (increased!) if the p-values are adjusted – yielding in general an extremely slow algorithm. For the consistent estimator (c) the corresponding p-value can only be estimated (using one of the other methods) and the convergence rates have not been analysed in detail yet, thus the actual type I error for a given finite sample is not directly available. Nevertheless note that fast and approximately sharp methods to estimate the p-values of multivariance are developed in the preprint (Berschneider and Böttcher 2019), see also Guetsop Nangue (2017). Moreover, an approximation of an upper bound for the type I error of the consistent estimator is given by the following elementary calculation: If one performs k independent sharp tests with significance levels γ_i then the probability of a type I error is $1 - \prod_{i=1}^k (1-\gamma_i)$. In the setting of multivariance the tests are in the limit (under H_0) independent and $\gamma_i \leq F_{\chi_1^2}(N^{1-\beta} \cdot C)$, thus posterior to testing the number of tests performed is known, say k, and the bound becomes $1 - (1 - F_{\chi_1^2}(N^{1-\beta} \cdot C))^k$ for the consistent estimator. Concerning β and C note that for the estimator discussed in Corollary 6.2 with β close to 1 the convergence to 0 (in the case of independence) becomes slower, for β close to 0 the divergence to ∞ (in the case of dependence) becomes slower, here $\beta = 1/2$ seems a balanced choice. For the value of C an optimal recommendation is still open – in our studies $C = 2$ seems a reasonable choice. Naturally, the constant C could also be based on a rejection level for a fixed sample size, e.g. choose C such that $5 \cdot C$ is the rejection level for a sample of size 25 for a significance level of 0.05. Then at least for this sample size the probability of a type I error is known, but this would still require some p-value estimation.

Finally, note that the suggested procedures are basic algorithms. There are certainly several variants and extensions possible, e.g. a further speed-up might be obtained by using total multivariance and m-multivariance for initial tests of independence (but beware of the problem of multiple vs. single tests). Furthermore, if pairwise dependence is detected (and clustered) this can be further analysed in the framework of graphical models. Hereto also note that (Pfister *et al.* 2017, Section 5.2) (see also (Chakraborty and Zhang 2019, Section 6)) provides a method for the detection of causal relations of variables using multivariate dependence measures, this can be used to refine an undirected graph (visualizing the dependence structure) into a directed graph. Moreover, clearly the visual layout allows many variants, e.g. one might also use different line types, thicknesses or colours to indicate the value of the dependence measure or the order of dependence, also the denoted values could for example be replaced by p-values or by some other quantity (or symbol) describing the dependence.

7 Empirical studies

In Section 10 a comprehensive collection of illustrating examples is provided. These discuss in detail several (toy-)examples of higher order dependence and their visualization, including the full and clustered dependence structure detection. Moreover also the detection power, empirical size and various other properties of multivariance are studied.

Here in the current section we will only discuss two types of examples: Comparisons with other multivariate dependence measures (showing that the new tests are competitive) and two basic real data examples (indicating the possibilities and some limitations of the methods).

For each example the dependence structure is visualized and the presented tables compare the power of the independence tests introduced in Sections 4 and 5 with those of the cited papers. Additionally the tables include a test called 'Comb' which combines the tests of m- and total multivariance by Holm's method. This provides a reference for readers with an interest in a joint test procedure, rather than comparing individual tests in their realm. For a full explanation of the setting, terms and parameter values of the studies we refer to the introduction of Section 10.

7.1 Empirical comparison of multivariance with other dependence measures

As discussed in Section 3.3 there are several dependence measures which are closely related to distance multivariance and its variants. For these empirical power comparisons are provided in Examples 7.2 and 7.3. But we begin with an example of a different visualization of higher order dependence which was proposed alongside a copula based dependence measure.

Example 7.1 (Dependogram vs. visualization). *In Genest and Rémillard (2004) copula based higher order dependence tests were proposed together with a dependogram, which provides a graphical representation of the test results of multiple testing. Our proposed full dependence structure visualization is closely related, it provides a visualization of all significant dependencies.*

In Figure 3 the dependogram and the corresponding dependence structure (which is here actually detected using the same samples and distance multivariance) are depicted for the example provided in (Genest and Rémillard 2004, Section 4.2): Let $Z_i, i = 1, \dots, 5$ be independent standard normal variables and let $X_1 := |Z_1| \operatorname{sign}(Z_2 Z_3)$, $X_i := Z_i$ for $i = 2, 3, 4$ and $X_5 := Z_4/2 + \sqrt{3} Z_5/2$. Now consider $N = 50$ samples of $(X_1, \dots, X_5)$.

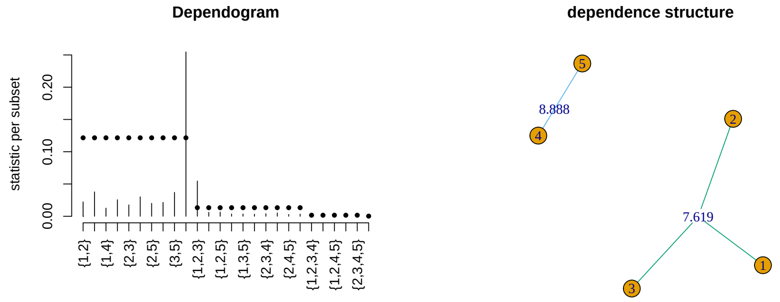
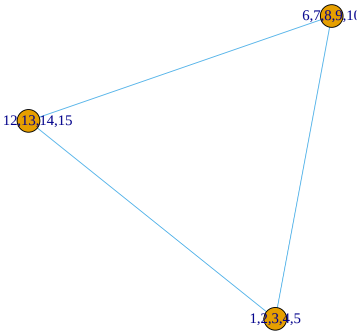


Figure 3: Ex. 7.1: dependogram (see Genest and Rémillard (2004); implemented in Hofert *et al.* (2018)) and the corresponding dependence structure.

Example 7.2 (Comparison with the methods of Jin and Matteson (2018)). *Here we compare tests based on multivariance with those presented in Jin and Matteson (2018). In general, one should note that the computation of sample multivariance has complexity* $O(N^2)$ *whereas the exact sample versions of Jin and Matteson (2018) (e.g.* $\mathcal{Q}_N, \mathcal{S}_N$*) have higher complexity. To reduce the complexity they introduce approximate estimators (e.g.* $\mathcal{Q}^\star_N, \mathcal{J}^\star_N$*) which have the same complexity as ours. Note that these approximate estimators are not permutation invariant with respect to the order of the samples. In fact their positive finding (significant p-values) in the real data example (Jin and Matteson 2018, 6.2 Financial data) is an artefact due to this shortcoming. Their test yields for the same data with permuted samples p-values about 0.3 and above. Therefore we strongly advise against the use of their approximate estimators in the given form. This problem can be reduced by permuting the samples prior to the use of their estimators.*

Nevertheless, we decided to use their measures for a comparison, since these are the most recent dependence measures related to the approach discussed in Section 3.3 corresponding to (36)*. Moreover Jin and Matteson (2018) also provides several variants and comparative tables including independence tests based on other measures. The following tables are computed with their parameter settings, e.g.* $\alpha = 0.1$*. We only include tests based on their best exact and approximate estimators (for each particular example), for further comparisons see the full tables in Jin and Matteson (2018).*

The example (Jin and Matteson 2018, Example 3) considers random variables X_i *with values in* $\mathbb{R}^5$ *such that* $(X_1, X_2, X_3) \sim N_{15}(0, \Sigma)$ *with* $\Sigma_{ij} = 1$ *for* $i = j$ *and* 0.1 *otherwise. For this example tests based on total multivariance and* 2*-multivariance match the power of tests based on their the exact estimator and outperform the approximate estimator (Figure 4).*

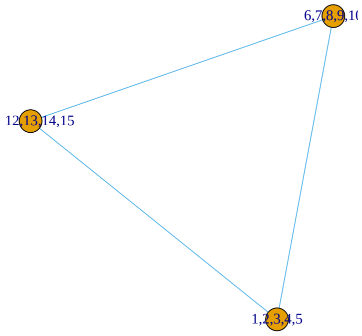


	resampling			*		
N	${}^N\overline{\mathcal{M}}$	${}^N\mathcal{M}_2$	Comb	$\mathcal{Q}^\star_N$	$\mathcal{S}_N$	$dHSIC$
25	0.408	0.417	0.359	0.220	0.418	0.982
50	0.712	0.722	0.631	0.378	0.719	1.000
100	0.960	0.970	0.941	0.707	0.961	1.000
150	0.995	0.995	0.993	0.873	0.996	1.000
200	1.000	1.000	1.000	0.946	1.000	1.000
300	1.000	1.000	1.000	0.997	1.000	1.000
500	1.000	1.000	1.000	1.000	1.000	1.000

* values from (Jin and Matteson 2018, Table 6)

Figure 4: Dependence structure sketch and empirical power of the (in)dependence tests for (Jin and Matteson 2018, Example 3) (Ex. 7.2).

As second example (Jin and Matteson 2018, Example 4) we consider $(Y_1, \ldots, Y_{15}) \sim N_{15}(0, \Sigma)$ *with* $\Sigma_{ij} = 1$ *for* $i = j$ *and* 0.4 *otherwise and set* $X_i := (\ln(Y_{5i}^2), \ldots, \ln(Y_{5i+4}^2))$ *for* $i = 1, 2, 3$*. Again, tests based on total multivariance and* 2*-multivariance are close to the power of the tests based on the exact estimator and they outperform the approximate estimator (Figure 5).*



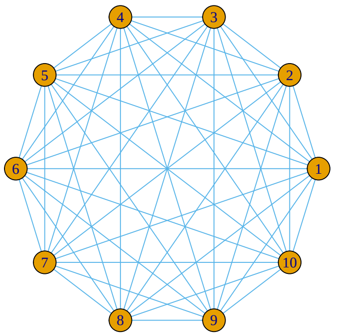
	resampling			*		
N	${}^N\overline{\mathcal{M}}$	${}^N\mathcal{M}_2$	Comb	$\mathcal{R}_N$	$\mathcal{J}^\star_N$	$dHSIC$
25	0.256	0.290	0.221	0.294	0.169	0.267
50	0.452	0.495	0.413	0.504	0.320	0.441
100	0.780	0.817	0.732	0.824	0.579	0.745
150	0.930	0.941	0.902	0.942	0.770	0.906
200	0.990	0.993	0.983	0.987	0.905	0.963
300	0.999	0.999	0.999	0.999	0.982	0.997
500	1.000	1.000	1.000	1.000	1.000	1.000

* values from (Jin and Matteson 2018, Table 8)

Figure 5: Dependence structure sketch and empirical power of the (in)dependence tests for (Jin and Matteson 2018, Example 4) (Ex. 7.2).

Finally, we discuss (Jin and Matteson 2018, Example 5): for dimensions $n \in \{5, 10, 15, 20, 25, 30, 50\}$ and sample size $N = 100$ we consider $(X_1, \dots, X_n) \sim N_n(0, \Sigma)$ with $\Sigma_{ij} = 1$ for $i = j$ and 0.1 otherwise. Here the test based on 2-multivariance is close to the power of the test based on the exact estimator and it outperforms the approximate estimator (Figure 6).

One might argue that the comparison with 2-multivariance is unjust, since it provides only a test for pairwise independence, whereas the other measures yield tests for independence. Hereto note that also the combination of the tests in 'Comb' has a higher detection rate than the tests based on the approximate estimators.



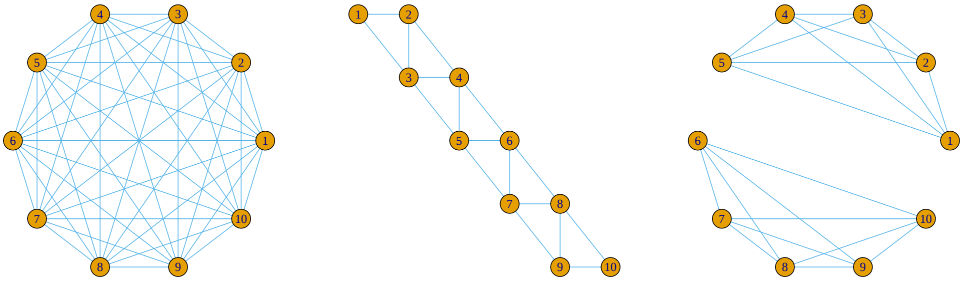
	resampling			distribution-free		*	
n	${}^{100}\overline{\mathcal{M}}$	${}^{100}\mathcal{M}_2$	Comb	${}^{100}\overline{\mathcal{M}}$	${}^{100}\mathcal{M}_2$	$\mathcal{Q}^\star_{100}$	$\mathcal{S}_{100}$
5	0.423	0.515	0.409	0.000	0.000	0.298	0.557
10	0.252	0.873	0.780	0.003	0.000	0.557	0.915
15	0.374	0.972	0.946	0.012	0.000	0.822	0.982
20	0.443	0.995	0.988	0.054	0.000	0.924	0.999
25	0.532	1.000	0.999	0.164	0.000	0.977	0.999
30	0.588	1.000	1.000	0.234	0.000	0.980	1.000
50	0.821	1.000	1.000	0.657	0.000	0.998	1.000

* values from (Jin and Matteson 2018, Table 10)

Figure 6: Dependence structure sketch ($n = 10$) and empirical power of the (in)dependence tests for (Jin and Matteson 2018, Example 5) (Ex. 7.2).

Example 7.3 (Comparison with the methods of Yao *et al.* (2017)). *In Yao et al. (2017) several measures of dependence were introduced. The main contribution is a measure dCov for pairwise dependence, which is closely related to ${}^N\mathcal{M}_2$. The examples in Yao et al. (2017) use the parameters $N \in \{60, 100\}$ and $n \in \{50, 100, 200, 400, 800\}$ and $\alpha = 0.05$, which we also use here to provide values which can be compared to other dependence measures given in their tables. Let X_i, $i = 1, \dots, n$ be random variables with values in $\mathbb{R}$ such that $(X_1, \dots, X_n) \sim N(0, \Sigma)$. We consider (Yao et al. 2017, Example 2), hereto let $\Sigma \in \mathbb{R}^{n \times n}$, $\Sigma_{ij} = 1$ for $i = j$ and otherwise (for $i \neq j$) set:*

1. *auto-regressive structure: $\Sigma_{ij} = (0.25)^{|i-j|}$,*
2. *band structure: $\Sigma_{ij} = 0.25$ for $0 < |i - j| < 3$ and 0 otherwise,*
3. *block structure: $\Sigma = I_{\lfloor n/5 \rfloor} \otimes A$ where $I_k \in \mathbb{R}^{k \times k}$ is the identity matrix and $A \in \mathbb{R}^{k \times k}$ with $A_{ij} = 1$ for $i = j$ and 0.25 otherwise.*



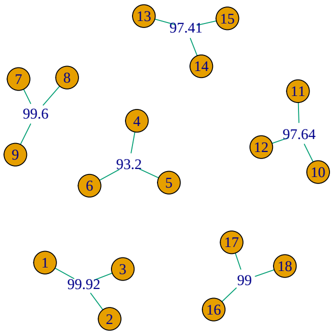
		auto-regressive				band structure				block structure			
		resampling			*	resampling			*	resampling			*
n	N	${}^N\overline{\mathcal{M}}$	${}^N\mathcal{M}_2$	Comb	$dCov$	${}^N\overline{\mathcal{M}}$	${}^N\mathcal{M}_2$	Comb	$dCov$	${}^N\overline{\mathcal{M}}$	${}^N\mathcal{M}_2$	Comb	$dCov$
50	60	0.052	0.898	0.768	0.886	0.159	0.999	0.999	1.000	0.232	0.998	0.994	0.999
100	60	0.108	0.873	0.807	0.906	0.192	1.000	1.000	0.999	0.234	1.000	0.998	1.000
200	60	0.104	0.896	0.765	0.909	0.167	1.000	0.999	1.000	0.139	0.999	0.998	1.000
400	60	0.111	0.924	0.812	0.909	0.174	0.998	0.998	1.000	0.177	1.000	0.999	1.000
800	60	0.101	0.937	0.843	0.908	0.105	1.000	1.000	1.000	0.128	1.000	1.000	1.000
50	100	0.115	0.999	0.996	0.998	0.137	1.000	1.000	1.000	0.195	1.000	1.000	1.000
100	100	0.071	0.999	0.986	0.999	0.153	1.000	1.000	1.000	0.170	1.000	1.000	1.000
200	100	0.128	1.000	0.999	1.000	0.142	1.000	1.000	1.000	0.222	1.000	1.000	1.000
400	100	0.073	1.000	1.000	0.999	0.168	1.000	1.000	1.000	0.169	1.000	1.000	1.000
800	100	0.084	1.000	0.998	0.999	0.139	1.000	1.000	1.000	0.191	1.000	1.000	1.000

* the $dCov$ values are from (Yao *et al.* 2017, Table 2)

Figure 7: Dependence structure sketches ($n = 10$) and empirical power of the (in)dependence tests for (Yao *et al.* 2017, Example 2.a) (Ex. 7.3).

In all cases the performance of tests based on 2-multivariance is very similar to their tests, see Figure 7. Note that due to computation time restrictions we used for the table in Figure 7 the resampling distribution of one sample to compute all resampling p-values (instead of resampling each sample separately).

In (Yao et al. 2017, Example 6) random variables $(X_1, \dots, X_n)$ are considered where the 3-tuples (X_1, X_2, X_3) and (X_4, X_5, X_6) and ... are independent and each 3-tuple consists of pairwise independent but 3-dependent Bernoulli random variables (as explicitly constructed in Example 10.2). Here only the sample sizes and dimensions $(N, n) \in \{(60, 18), (100, 36), (200, 72)\}$ are used. Figure 8 shows that the test based on 3-multivariance (and also the combined test 'Comb') clearly outperforms all tests included in their table (of which we only cite two in our table).



		resampling			(Yao *et al.* 2017, Table 4)	
n	N	${}^N\overline{\mathcal{M}}$	${}^N\mathcal{M}_3$	Comb	$dCov$	$dHSIC(3)$
18	60	0.112	1.000	1.000	0.051	0.708
36	100	0.044	1.000	1.000	0.048	0.314
72	200	0.047	1.000	1.000	0.057	0.073

Figure 8: Dependence structure of (Yao *et al.* 2017, Example 6) (with $n = 18$) and the empirical power comparison. Note that here dHSIC(3) denotes dHSIC with a special choice of the bandwidth parameter, see Yao *et al.* (2017) for details. (Ex. 7.3).

7.2 Real data examples

As stated in the introduction, looking at other papers on multivariate dependence measures (e.g. those discussed in Section 3.3) one notices that although these are capable of detecting dependencies of higher order the real data examples feature pairwise dependence. From our point of view this seems first of all to be due to the fact that the concept of higher order dependencies is not popular (or even unknown) in applied statistics. Therefore, on the one hand there is a very strong publication bias for datasets with pairwise dependencies, on the other hand even if datasets statistically feature higher order dependencies an explanation by field experts is yet missing. Nevertheless, we can refer to a collection of more than 350 datasets which feature higher order dependence[7].

In the following we present two examples for which 2-multivariance and total multivariance detect some dependence. In terms of dependence structure detection they are more delicate: The first example illustrates the difference between the clustered and full dependence structure and it indicates an application of higher order dependencies to model selection. The second example discusses detected higher order dependencies which are actually based on pairwise dependence. It illustrates results caused by a small sample size, a conservative detection method and by a relatively high (for multiple testing) significance level (see also Remark 6.1).

Example 7.4 (Quine's student survey data). *We consider a classical data set of a student survey (Aitkin 1978) (see also (Venables and Ripley 2002, R-package: MASS, dataset: quine)), which contains 146 samples of the variables: age (actually the class level), gender, cultural background, type of learner and the number of days school was missed. The dataset was extensively used in Aitkin (1978) to discuss model selection in a multi-factor analysis of variance to model the number missed school days.*

The conservative tests using 2-multivariance and total multivariance detect no dependencies (p-values: 0.0767, 0.1565), the corresponding resampling tests reject independence with actual p-values of 0.00.

The dependence structure detection yields the structures shown in Figure 9. Here the full structure provides a refinement of the clustered structure. For the detection we used resampling tests with 10000 resamples and significance level $\alpha = 0.01$. *Based on the actually performed multiple tests the approximate probability of a type I error is 0.0297 for the clustered structure and 0.0199 for the full structure. By the large number of resamples used this example might just seem to be an (impractical) proof of concept, but note that the same results can also be obtained with the faster methods developed in Berschneider and Böttcher (2019).*

For the variables: age, gender and missed-days 3-dependence (with lower order independence) was detected (Figure 9). To judge if this is really a sensible finding in terms of the field of study is beyond our expertise. Nevertheless the found dependencies naturally suggest candidates for a minimal model for the number of missed days: Based on the detected full dependence structure the missed days depend only on the cultural background and on the interaction term of age and gender.

[7] A collection of datasets featuring higher order dependence, http://www.math.tu-dresden.de/~boettch/research/hod/

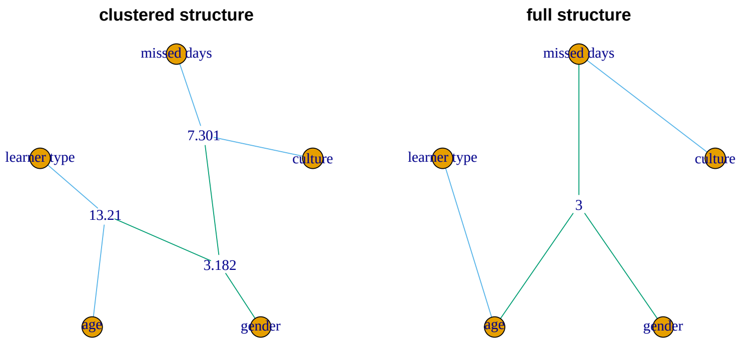


Figure 9: Student survey data (Ex. 7.4): detected dependence structures.

A preliminary attempt to use other multivariate dependence measures for the detection of the dependence structure in this example did not yield unanimous results. Hereto recall that any test based on resampling adds randomness to its results. Moreover, the implementations and presets of the dependence measures might not (or not yet) be appropriate for this particular task.

Example 7.5 (Decathlon)**.** *The results of decathlon athletes from 1985 to 2006 are provided by Unwin (2015). To consider these (and smaller subsets) as independent samples we only keep the personal best of each athlete, leaving 2709 samples, and order these by the achieved total points in increasing order (the field is denser for lower points, constituting more to the required i.i.d. setting for the sample subsets tested). It is well known that the 10 disciplines are dependent, e.g. Cox and Dunn (2002); Woolf et al. (2007). We are interested how many samples (using the real measurements of the results in each discipline) are required to detect a dependence using the presented methods: with tests based on 2-multivariance M_2 using the resampling method dependence is, for a significance level of $\alpha = 0.05$, first detected for $N = 5$ and finally for all $N > 11$ (the conservative method detects dependence first for $N = 154$). Using the test based on total-multivariance $\overline{M}$ with the resampling method dependence is detected also for all $N > 11$ (the conservative method detects dependence first for $N = 2603$ – thus the conservative test are very conservative in this setting!).*

Next we try to detect the clustered dependence structure based on conservative tests, thus it is interesting to see which structures are detected for various sample sizes, see Figure 10. The detection of the higher order dependence indicates early on that these variables are dependent, but due to the conservative tests, the actual lower order dependence is missed. With increasing sample size only the dominant pairwise dependencies are detected. Also note that due to the repeated testing and the given significance level the probability of a type I error is large.

Notably there are some natural variants: 1. Instead of the results one could consider the achieved points in each discipline, which are obtained by non linear transformations of the results. This yields almost the same inference. 2. Starting with the elite athletes instead of our order causes a change in the detection: In this case a resampling test based on 2-multivariance detects a dependence for all $N > 25$ but total-multivariance requires much more samples: $N > 177$. Thus here a curse of dimension is at work (compare with Example 10.14), which might indicate that for top athletes some disciplines are less dependent than for other athletes. We leave further analysis and interpretation to field experts. The setting also naturally yields to clustering methods (for dependent random variables) based on distance multivariance, a topic which is beyond the current paper.

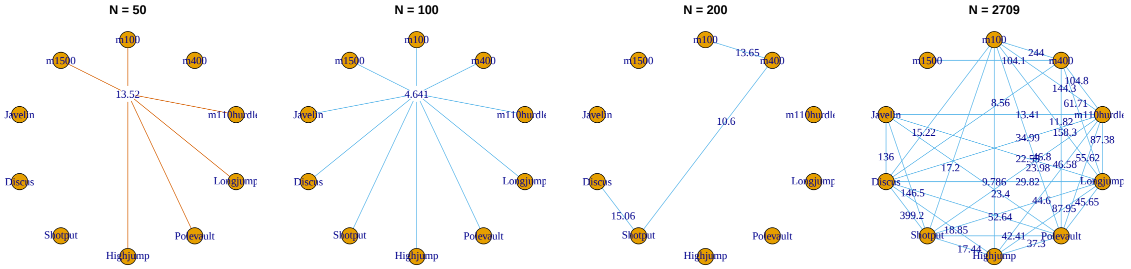


Figure 10: Decathlon (Ex. 7.5): detected dependence structures (based on conservative p-values) for 50, 100, 200 and all 2709 samples.

8 Conclusion

The framework of distance multivariance is a powerful tool for the detection of dependence. It provides a unified theory covering several known dependence measures. In particular, measures well established in applications appear as limiting cases, e.g. the RV coefficient.

Besides useful extensions of the theory (e.g. a relaxation of the required moment assumptions, introduction of approximately sharp resampling tests), also new tests and visualization procedures for higher order dependencies have been developed. As indicated in the introduction and in the example section: higher order dependencies seem not yet to be in the focus of applied statistics. We hope, that the presented results (and their ready-to-use implementations in the R package `multivariance`) help and inspire practitioners to study higher order dependencies.

The presented results yield also to several new theoretic research questions and topics, e.g.:

- The expositions in Section 3.3 provide a roadmap to unify further dependence measures using continuous negative definite functions.
- As hinted in Section 3.5 a natural follow-up question is the construction of an extension of multivariance to other spaces than $\mathbb{R}^d$.
- The flexibility of distance multivariance due to the continuous negative definite functions (see Equation (11) ff.) raises the general question of optimal or adaptive ρ_i or ψ_i selection procedures.
- Optimization of the dependence structure detection algorithm, e.g. parameter selection, adaptive procedures, improved error control.
- There might be multiple testing procedures for testing independence particularly tailored to the introduced methods – we just used the classical method of Holm.
- Clustering methods based on distance multivariance.
- Finally, as stated in Section 3 there are several related – but different – multivariate dependence measures. A general classification of situations where a particular measure provides the best performance seems still open. In particular, also their performance in the general dependence structure detection algorithms of Section 6 has to be investigated.

Acknowledgement: We are grateful for detailed proofreading comments by Georg Berschneider and some remarks by Réne Schilling and other colleagues. Moreover, we want to thank several referees for their detailed reports, which helped to bring the paper to its present form.

Financial Support: This research received no specific grant from any funding agency, commercial or nonprofit sectors.

Conflict of Interests: The authors have no conflicts of interest to disclose.

Ethics Statement: This research did not required ethical approval.

References

Aitkin, M. (1978), "The analysis of unbalanced cross-classifications", *Journal of the Royal Statistical Society: Series A (General)*, 141(2), 195–211.

Berg, C., and Forst, G. (1975), *Potential Theory on Locally Compact Abelian Groups*, Berlin: Springer.

Berschneider, G., and Böttcher, B. (2019), *On complex Gaussian random fields, Gaussian quadratic forms and sample distance multivariance*, arXiv:1808.07280v2.

Bilodeau, M., and Guetsop Nangue, A. (2017), "Tests of mutual or serial independence of random vectors with applications", *The Journal of Machine Learning Research*, 18(1), 2518–2557.

Böttcher, B. (2019), *multivariance: Measuring Multivariate Dependence Using Distance Multivariance*. R package version 2.2.0.

Böttcher, B., Keller-Ressel, M., and Schilling, R. L. (2018), "Detecting independence of random vectors: Generalized distance covariance and Gaussian covariance", *Modern Stochastics: Theory and Applications*, 5(3), 353–383.

Böttcher, B., Keller-Ressel, M., and Schilling, R. L. (2019), "Distance multivariance: New dependence measures for random vectors", *The Annals of Statistics*, 47(5). 2757–2789.

Böttcher, B., Schilling, R. L., and Wang J. (2013), *Lévy-Type Processes: Construction, Approximation and Sample Path Properties*, volume 2099 of *Lecture Notes in Mathematics, Lévy Matters*, Springer.

Chakraborty, S., and Zhang, X. (2019), "Distance metrics for measuring joint dependence with application to causal inference", *Journal of the American Statistical Association*, 114(528), 1638-1650.

Cox, T. F., and Dunn, R. T. (2002), "An analysis of decathlon data", *Journal of the Royal Statistical Society. Series D (The Statistician)*, 51(2), 179–187.

Csardi, G., and Nepusz, T. (2006), "The igraph software package for complex network research", *InterJournal*, Complex Systems, 1695.

Csörgő, S. (1985), "Testing for independence by the empirical characteristic function", *Journal of Multivariate Analysis*, 16(3), 290–299.

Edelmann D. (2015), *Structures of Multivariate Dependence*, PhD thesis, Universität Heidelberg.

Escoufier, Y. (1973), "Le traitement des variables vectorielles", *Biometrics*, 29(4), 751–760.

Fan, Y., de Micheaux, P. L., Penev, S., and Salopek, D. (2017), "Multivariate nonparametric test of independence", *Journal of Multivariate Analysis*, 153, 189–210.

Genest, C., and Rémillard, B. (2004), "Test of independence and randomness based on the empirical copula process", *Test*, 13(2), 335–369.

Gretton, A., Fukumizu, K., Teo, C. H., Song, L., Schölkopf, B., and Smola, A. J. (2008), "A kernel statistical test of independence", *Advances in Neural Information Processing Systems*, 20, 585–592.

Guetsop Nangue, A. (2017), *Tests de permutation d'ind´ependance en analyse multivariée*, PhD thesis, Université de Montréal.

Han, J., Pei, J., and Kamber, M. (2011), *Data mining: concepts and techniques,* Burlington: Morgan Kaufmann.

Hofert, M., Kojadinovic, I., Maechler, M., and Yan, J. (2018), *copula: Multivariate Dependence with Copulas*, R package version 0.999-19.

Jacob, N. (2001), *Pseudo-Differential Operators and Markov Processes I. Fourier Analysis and Semigroups*, London: Imperial College Press.

Jin, Z., and Matteson, D. S. (2018), "Generalizing distance covariance to measure and test multivariate mutual dependence via complete and incomplete V-statistics", *Journal of Multivariate Analysis*, 168, 304–322.

Josse, J., and Holmes, S. (2016), "Measuring multivariate association and beyond", *Statistics Surveys*, 10, 132-167.

Kallenberg, O. (1997), *Foundations of Modern Probability*, New York, Berlin, Heidelberg: Springer.

Kankainen, A. (1995), *Consistent testing of total independence based on the empirical characteristic function*, PhD thesis, University of Jyväskylä.

Korolyuk, V. S., and Borovskich, Y. V. (1994), *Theory of U-statistics*, volume 273, Dordrecht: Springer Science & Business Media.

Liu, Y., de la Pena, V., and Zheng, T. (2018), "Kernel-based measures of association", *Wiley Interdisciplinary Reviews: Computational Statistics*, 10(2), e1422.

Lyons, R. (2013), "Distance covariance in metric spaces", *The Annals of Probability*, 41(5), 3284-3305.

Matheron, G. (1963), "Principles of Geostatistics", *Economic geology*, 58(8), 1246–1266.

Móri, T. F., and Székely, G. J. (2018), "Four simple axioms of dependence measures", *Metrika*, 82, 1–16.

Pfister, N., Bühlmann, P., Schölkopf, B., and Peters, J. (2017), "Kernel-based tests for joint independence", *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80, 5–31.

Pfister, N., and Peters, J. (2019), *dHSIC: Independence Testing via Hilbert Schmidt Independence Criterion*, R package version 2.1.

Robert, P., and Escoufier, Y. (1976), "A Unifying Tool for Linear Multivariate Statistical Methods: The RV- Coefficient", *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 25(3), 257- 265.

Rényi, A. (1959), “On measures of dependence”, *Acta mathematica hungarica*, 10(3-4), 441–451.

Sato, K. (1999), *Lévy Processes and Infinitely Divisible Distributions*, Cambridge: Cambridge University Press.

Sejdinovic, D., Gretton, A., and Bergsma, W. (2013), “A Kernel Test for Three-Variable Interactions” in *Advances in Neural Information Processing Systems (NeurIPS)*, volume 26, pp. 1124–1132.

Sejdinovic, D., Sriperumbudur, B., Gretton, A., and Fukumizu, K. (2013), “Equivalence of distance-based and RKHS-based statistics in hypothesis testing”, *Annals of Statistics*, 41(5), 2263–2291.

Shen, C., and Vogelstein, J. T. (2018), “The exact equivalence of distance and kernel methods for hypothesis testing”, *CoRR*, abs/1806.05514.

Székely, G. J., and Bakirov, N. K. (2003), “Extremal probabilities for Gaussian quadratic forms”, *Probability Theory and Related Fields*, 126(2), 184–202.

Székely, G. J., and Rizzo, M. L. (2009), “Brownian distance covariance”, *Annals of Applied Statistics*, 3(4), 1236– 1265.

Székely, G. J., Rizzo, M. L., and Bakirov, N. K. (2007), “Measuring and testing dependence by correlation of distances”, *The Annals of Statistics*, 35(6), 2769–2794.

Tjøstheim, D., Otneim, H., and Støve, B. (2018), Statistical dependence: Beyond pearson’s ρ. arXiv:1809.10455v1.

Unwin, A. (2015), *GDAdata: Datasets for the Book Graphical Data Analysis with R,* R package version 0.93.

Venables, W. N., and Ripley, B. D. (2002), *Modern Applied Statistics with S*. New York: Springer, fourth edition.

Woolf, A., Ansley, L., and Bidgood, P. (2007), “Grouping of decathlon disciplines”, *Journal of Quantitative Analysis in Sports*, 3(4).

Yao, S., Zhang, X., and Shao, X. (2017), “Testing mutual independence in high dimension via distance covariance”, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80, 455–480.

Zinger, A., Kakosyan, A. V., and Klebanov, L. B. (1992), “A characterization of distributions by mean values of statistics and certain probabilistic metrics”, *Journal of Mathematical Sciences*, 59(4), 914–920.

9 Appendix - Further results and proofs

Here we collect several results which are essential for (parts of) the previous sections, but which were postponed to this section due to their technicality.

9.1 A theorem characterizing the support of Lévy measures

Note that in (Böttcher *et al.* 2018, after Definition 2.3) it was stated that it is unknown how to characterize the (full) support of Lévy measures in terms of the corresponding continuous negative definite function. The following result provides a characterization (via Proposition 2.2), it is related to (Zinger *et al.* 1992, Corollary 2).

Theorem 9.1. *Let* $\psi(x) := \int_{\mathbb{R}^d} 1 - \cos(x \cdot t)\, \rho(dt)$ *where* ρ *is a symmetric measure integrating* $1 \wedge |.|^2$, *and* X, Y *be* $\mathbb{R}^d$*-valued random vectors with characteristic functions* f_X, f_Y, *and assume* $\mathbb{E}(\psi(X)) < \infty$ *and* $\mathbb{E}(\psi(Y)) < \infty$. *Then*

$$f_X = f_Y \ \rho\text{-a.s.} \quad \Leftrightarrow \quad \text{for all } z \in \mathbb{R}^d : \ \mathbb{E}(\psi(X - z)) = \mathbb{E}(\psi(Y - z)). \tag{A1}$$

Proof. Additionally to the stated assumptions let Z, X', Y' be independent random variables which are also independent of X, Y and satisfy $\mathbb{E}(\psi(Z)) < \infty$ and $X' \overset{d}{=} X$, $Y' \overset{d}{=} Y$. Note that

$$\int 1 - Re(f_X(t) f_Z(-t))\, \rho(dt) = \iiint 1 - \cos((x - z) \cdot t)\, \rho(dt) \mathbb{P}(X \in dx) \mathbb{P}(Z \in dz) = \mathbb{E}(\psi(X - Z)) < \infty \tag{A2}$$

by Tonelli and using the (generalized) triangle inequality for continuous negative definite functions (A9). Thus the following implications hold:

$$\begin{aligned}
f_X = f_Y \ \rho\text{-a.s.} \ \Rightarrow \quad & \text{for all } z \in \mathbb{R}^d : \ \int Re((f_X(t) - f_Y(t)) \mathrm{e}^{-\mathrm{i}z \cdot t})\, \rho(dt) = 0 && \text{(A3)} \\
\Leftrightarrow \quad & \text{for all } z \in \mathbb{R}^d : \ \mathbb{E}(\psi(X - z)) = \mathbb{E}(\psi(Y - z)) && \text{(A4)} \\
\Rightarrow \quad & \mathbb{E}(\psi(X - X')) = \mathbb{E}(\psi(Y - X')) \text{ and } \mathbb{E}(\psi(X - Y')) = \mathbb{E}(\psi(Y - Y')) && \text{(A5)} \\
\Leftrightarrow \quad & \int |f_X(t)|^2 - Re(f_Y(t) f_{X'}(-t))\, \rho(dt) = 0 && \text{(A6)} \\
& \text{and} \quad \int |f_Y(t)|^2 - Re(f_X(t) f_{Y'}(-t))\, \rho(dt) = 0 && \\
\Rightarrow \quad & \int |f_X(t) - f_Y(t)|^2\, \rho(dt) = 0 && \text{(A7)}
\end{aligned}$$

and the last line is equivalent to the start. This completes the proof. □

9.2 Moment condition

In Böttcher *et al.* (2019) the following condition was used:

$$\textbf{mixed } \boldsymbol{\psi}\textbf{-moment condition:} \quad \mathbb{E}\left(\prod_{i=1}^{n} \psi_i(X_{k_i,i} - X'_{l_i,i}) \right) < \infty \text{ for all } k_i, l_i \in \{0, 1\}, i = 1, \ldots, n \tag{A8}$$

where $\boldsymbol{X}_0, \boldsymbol{X}'_0, \boldsymbol{X}_1, \boldsymbol{X}'_1$ are independent and have the same marginal distributions as $\boldsymbol{X}$ (for the dimensions d_i), $\boldsymbol{X}_1, \boldsymbol{X}'_1$ have also the same joint distribution as $\boldsymbol{X}$, but the marginal distributions of $\boldsymbol{X}_0, \boldsymbol{X}'_0$ are independent (for further details see (Böttcher *et al.* 2019, Def. 2.3.a)).

We show that for non constant random vectors X_i the joint ψ-moment condition (5) and (A8) are equivalent. If a random vector is constant condition (A8) becomes trivial since the corresponding factor therein is equal to 0.

Recall the (generalized) triangle inequality which holds for any real-valued negative definite function ψ (Böttcher *et al.* 2018, Equation (8)):

$$\psi(x+y) \leq 2\psi(x) + 2\psi(y). \tag{A9}$$

By this inequality (5) implies (A8). For the converse implication we begin with the following observation.

Lemma 9.2. *For random variables $(X_1, \dots, X_n)$ the following are equivalent:*

(a) for all $S \subset \{1,\dots,n\} : \mathbb{E}\left(\prod_{i\in S} \psi_i(X_i)\right) < \infty$,
(b) for all $S \subset \{1,\dots,n\} : \mathbb{E}\left(\prod_{i\in S} \psi_i(X_i - x_i)\right) < \infty$ for some $(x_1,\dots,x_n)$,
(c) for all $S \subset \{1,\dots,n\} : \mathbb{E}\left(\prod_{i\in S} \psi_i(X_i - \tilde{x}_i)\right) < \infty$ for all $(\tilde{x}_1,\dots,\tilde{x}_n)$.

Proof. Obviously (c) with $\tilde{x}_i = 0, i = 1,\dots,n$ is (a) which implies (b). Finally, (c) follows from (b) by $\psi_i(X_i - \tilde{x}_i) \leq 2\psi_i(X_i - x_i) + 2\psi(x_i - \tilde{x}_i)$ applied to each component. Note that hereto it is essential that the expectations are finite for all subsets $S \subset \{1,\dots,n\}$. □

Now note that $\mathbb{E}(\psi_i(X_i - X_i')) > 0$ for non-constant random variables. Thus the expectations of independent components (i.e., for $k_i = l_i = 0$ in (A8)) which factor out in (A8) yield strictly positive factors. Therefore, due to the independence of $(X_1,\dots,X_n)$ and $(X_1',\dots,X_n')$, the condition (A8) implies for all $S \subset \{1,\dots,n\}$: $\mathbb{E}\left(\prod_{i\in S}\psi_i(X_i - x_i)\right) < \infty$ for $\mathbb{P}_{(X_1,\dots,X_n)}$-almost all $(x_1,\dots,x_n)$. Hence the joint ψ-moment condition (5) holds by Lemma 9.2.

Note that further moment conditions for the case $\psi_i(.) = |.|$ can be found in Chakraborty and Zhang (2019).

9.3 Proof of the asymptotics of sample distance multivariance (Theorem 2.5)

Here we are in the setting of Section 2. The asymptotics (25) and (27) of the test statistic were proved in (Böttcher *et al.* 2019, Thm. 4.5, 4.10, Cor. 4.16, 4.18) and (Böttcher *et al.* 2018, Cor. 4.8) under the condition (23). The following theorem provides a proof using an alternative condition. Combining the results yields the convergence statements (25) and (27) of Theorem 2.5.

Theorem 9.3. *Let X_i, $i = 1,\dots,n$ be non-constant random variables such that*

$$\mathbb{E}(\psi_i^2(X_i)) < \infty \text{ for all } i = 1,\dots,n \tag{A10}$$

and let $\boldsymbol{X}^{(k)}$, $k = 1,\dots,N$ be independent copies of $\boldsymbol{X} = (X_1,\dots,X_n)$. Then

$$N \cdot {}^N\mathcal{M}^2(\boldsymbol{X}^{(1)},\dots,\boldsymbol{X}^{(N)}) \xrightarrow[N\to\infty]{d} Q \qquad \text{if } X_1,\dots,X_n \text{ are independent,} \tag{A11}$$

$$N \cdot {}^N\overline{\mathcal{M}}^2(\boldsymbol{X}^{(1)},\dots,\boldsymbol{X}^{(N)}) \xrightarrow[N\to\infty]{d} \overline{Q} \qquad \text{if } X_1,\dots,X_n \text{ are independent,} \tag{A12}$$

where Q and $\overline{Q}$ are Gaussian quadratic forms with $\mathbb{E}Q = 1 = \mathbb{E}\overline{Q}$.

Proof. Let $\boldsymbol{X}', \boldsymbol{X}^{(k)}$, $k = 1,\dots,N$ be independent copies of $\boldsymbol{X} = (X_1,\dots,X_n)$ with independent components. Note, ${}^N M^2(\boldsymbol{X}^{(1)},\dots,\boldsymbol{X}^{(N)}) = N^{-2}\sum_{j,k=1}^N {}^N\Phi(j,k)$ with ${}^N\Phi(j,k) := {}^N\Phi_{\{1,\dots,n\}}(j,k)$ where

$$\begin{aligned} &{}^N\Phi_S(j,k) := {}^N\Phi_S(j,k;\boldsymbol{X}^{(1)},\dots,\boldsymbol{X}^{(N)}) := \\ &\prod_{i\in S}\left(-\psi_i(X_i^{(j)} - X_i^{(k)}) + N^{-1}\sum_{m=1}^N \psi_i(X_i^{(j)} - X_i^{(m)}) + N^{-1}\sum_{l=1}^N \psi_i(X_i^{(l)} - X_i^{(k)}) - N^{-2}\sum_{l,m=1}^N \psi_i(X_i^{(l)} - X_i^{(m)})\right). \end{aligned} \tag{A13}$$

Similarly, define $\Phi(\boldsymbol{x}^{(j)}, \boldsymbol{x}^{(k)}) := \Phi_{\{1,\dots,n\}}(\boldsymbol{x}^{(j)}, \boldsymbol{x}^{(k)})$ with

$$\Phi_S(\boldsymbol{x}^{(j)}, \boldsymbol{x}^{(k)}) := \prod_{i\in S}\left(-\psi_i(x_i^{(j)} - x_i^{(k)}) + \mathbb{E}(\psi_i(x_i^{(j)} - X_i)) + \mathbb{E}(\psi_i(X_i - x_i^{(k)})) - \mathbb{E}(\psi_i(X_i - X_i'))\right). \tag{A14}$$

Then $\mathbb{E}(\Phi(\boldsymbol{X},\boldsymbol{X})) = \prod_{i=1}^N \mathbb{E}(\psi_i(X_i - X_i'))$ and

$$\mathbb{E}(\Phi(\boldsymbol{x},\boldsymbol{X})) = 0,\ \mathbb{E}(\Phi(\boldsymbol{X},\boldsymbol{X}')) = 0,\ \mathbb{E}(|\Phi(\boldsymbol{X},\boldsymbol{X})|) < \infty \quad \text{and} \quad \mathbb{E}(\Phi(\boldsymbol{X},\boldsymbol{X}')^2) < \infty. \tag{A15}$$

where (A10) was used for the bounds. Therefore $N\cdot N^{-2}\sum_{j,k=1}^N \Phi(\boldsymbol{X}^{(j)},\boldsymbol{X}^{(k)})$ converges in distribution to a Gaussian quadratic form by (Korolyuk and Borovskich 1994, Thm. 4.3.2, p. 141). Note that in the limit in Korolyuk and Borovskich (1994) appear $\mathbb{E}(\Phi(\boldsymbol{X},\boldsymbol{X}))$ and a sum $\sum_{i=1}^\infty \lambda_i$, which cancel in our setting – this equality also implies that the limit for normalized multivariance has expectation 1, cf. (Berschneider and Böttcher 2019, Lemma 2.3 and Remark 4.9.1). Finally, (A11) follows by Slutsky's theorem since

$$N\cdot N^{-2}\sum_{j,k=1}^N \left({}^N\!\Phi(j,k) - \Phi(\boldsymbol{X}^{(j)},\boldsymbol{X}^{(k)})\right) \xrightarrow[N\to\infty]{\mathbb{P}} 0. \tag{A16}$$

To avoid a false impression, note that (A16) seems natural since the strong law of large numbers implies that the expectations in (A14) are approximated by the corresponding sums in (A13). But the additional factor N in (A16) makes the proof technical, which we only sketch here: For (A16) it is, using the Markov inequality, sufficient to show that the second moment of the left hand side converges to 0. This moment and its limit can be calculated explicitly based on and similar to (Berschneider and Böttcher 2019, Theorem 4.15), where the second moment of ${}^N\!\Phi_S(j,k)$ is analysed in-depth.

Considering analogously ${}^N\overline{\Phi}(j,k) := \sum_{\substack{S\subset\{1,\dots,n\}\\ |S|>1}} {}^N\!\Phi_S(j,k)$ instead of ${}^N\!\Phi$ yields the result for total multivariance. □

Remark 9.4. *Based on the methods developed in the preprint (Berschneider and Böttcher 2019, e.g. Section 7.7) the second order moment in* (A15) *seems to be already bounded under the weaker assumption* $\mathbb{E}(\psi_i(X_i)) < \infty$ *for all* $i = 1,\dots,n$. *To make this rigorous one would have to rewrite (or at least discuss) the steps in Korolyuk and Borovskich (1994) in much more detail, which is beyond the bounds of this paper. Moreover, this clearly also requires a discussion if (and why) the counterexample, which shows that the log moment condition in* (23) *(see also Remark 2.6) is necessary for the convergence of the empirical characteristic functions, is somehow compensated by the* $L^2(\rho)$ *norm.*

To prove the divergence in (24) and (26) we require further notations: let $\varepsilon > 0$ and $\rho_\varepsilon := \otimes_{i=1}^n \rho_{i,\varepsilon}$ with $\rho_{i,\varepsilon}(.) := \rho_i(.\cap B_{i,\varepsilon}^c)$ where $B_{i,\varepsilon}^c := \{x \in \mathbb{R}^{d_i} : |x| > \varepsilon\}$. Note that the corresponding continuous negative definite functions $\psi_{i,\varepsilon}(x_i) := \int 1-\cos(x_i\cdot t_i)\,\rho_{i,\varepsilon}(dt_i)$ are bounded (with non-full support; alternatively one could also use the truncation of (Böttcher *et al.* 2018, Eq. (40)) which preserves the full support). Moreover recall that by (Böttcher *et al.* 2019, Supplement, (S.7))

$${}^N\!M_\rho(x^{(1)},\dots,x^{(N)}) = \sqrt{\int \left|\frac{1}{N}\sum_{j=1}^N\prod_{i=1}^n\left(\mathrm{e}^{\mathrm{i}x_i^{(j)}\cdot t_i} - \frac{1}{N}\sum_{k=1}^N \mathrm{e}^{\mathrm{i}x_i^{(k)}\cdot t_i}\right)\right|^2 \rho(dt).} \tag{A17}$$

This and (1) yield by the monotone convergence theorem: $\sup_{\varepsilon>0} {}^N\!M_{\rho_\varepsilon}(\boldsymbol{X}^{(1)},\dots,\boldsymbol{X}^{(N)}) = {}^N\!M_\rho(\boldsymbol{X}^{(1)},\dots,\boldsymbol{X}^{(N)})$ and $\sup_{\varepsilon>0} M_{\rho_\varepsilon}(X_1,\dots,X_n) = M_\rho(X_1,\dots,X_n)$. Which are the key ingredients for the proof of the following Lemma, which in turn is the key to prove (24) and (26) without any further moment restrictions.

Lemma 9.5. *Let* $\boldsymbol{X}^{(k)}, k = 1,\dots,N$ *be independent copies of* $\boldsymbol{X} = (X_1,\dots,X_n)$. *Then, without any moment assumptions, we have*

$$\liminf_{N\to\infty} {}^N\!M_\rho(\boldsymbol{X}^{(1)},\dots,\boldsymbol{X}^{(N)}) \geq M_\rho(X_1,\dots,X_n). \tag{A18}$$

Proof. In this proof we omit $(\boldsymbol{X}^{(1)},\dots,\boldsymbol{X}^{(N)})$ and $(X_1,\dots,X_n)$ in the notation. Note that for ρ_ε instead of ρ the joint ψ-moment condition (5) is always satisfied, and therefore $\lim_{N\to\infty} {}^N\!M_{\rho_\varepsilon} = M_{\rho_\varepsilon}$ by (Böttcher *et al.* 2019, Theorem 4.3). Thus

$$M_\rho = \sup_{\varepsilon>0} M_{\rho_\varepsilon} = \sup_{\varepsilon>0}\lim_{N\to\infty} {}^N\!M_{\rho_\varepsilon} = \sup_{\varepsilon>0}\liminf_{N\to\infty} {}^N\!M_{\rho_\varepsilon} \leq \liminf_{N\to\infty}\sup_{\varepsilon>0} {}^N\!M_{\rho_\varepsilon} = \liminf_{N\to\infty} {}^N\!M_\rho. \qquad \square$$

Now the proof of the divergence (24) is identical to (Böttcher *et al.* 2019, Thm. 4.5.b) just replacing (Böttcher *et al.* 2019, Theorem 4.3) by Lemma 9.5: If the random variables are $(n-1)$–independent but dependent Theorem 2.1 implies $M > 0$. Thus $N \cdot {}^{N}M$ diverges for $N \to \infty$ by Lemma 9.5. This also applies in the case of dependence to at least one summand of total multivariance (and the remaining terms are all non negative) therefore also the divergence (26) follows.

9.4 The population representation of (Fan *et al.* 2017, Lemma 1a)

Note that the γ, β terms appearing in (Fan *et al.* 2017, Lemma 1a) correspond in our notation to $\gamma_{j,l} = \psi_l(x_l^{(j)})$, $\gamma_{j,j',l} = \psi_l(x_l^{(j)} - x_l^{(j')})$ and

$$\beta_{j,j',l} = \psi_l(x_l^{(j)}) + \psi_l(x_l^{(j)}) - \psi_l(x_l^{(j)} - x_l^{(j')}) =: \phi_l(x_l^{(j)}, x_l^{(j')}). \tag{A19}$$

Turning their sample sums into expectations and observing the independence implied by the indices yields the population version of their T_n:

$$\sum_{\substack{S\subset\{1,\ldots,n\}\\ |S|>0}} \sum_{\substack{S'\subset\{1,\ldots,n\}\\ |S'|>0}} (-1)^{|S|+|S'|} \Bigg[\mathbb{E}\Bigg(\prod_{l\in S\cap S'} \phi_l(X_l, X_l') \prod_{l\in S\setminus S'} \psi_l(X_l) \prod_{l\in S'\setminus S} \psi_l(X_l') \Bigg) \tag{A20}$$

$$- 2\mathbb{E}\Bigg(\prod_{l\in S\cap S'} \mathbb{E}(\phi_l(X_l, X_l') \mid X_l) \prod_{l\in S\setminus S'} \psi_l(X_l) \prod_{l\in S'\setminus S} \mathbb{E}\left(\psi_l(X_l')\right) \Bigg) \tag{A21}$$

$$+ \prod_{l\in S\cap S'} \mathbb{E}(\phi_l(X_l, X_l')) \prod_{l\in S\setminus S'} E(\psi_l(X_l)) \prod_{l\in S'\setminus S} \mathbb{E}(\psi_l(X_l')) \Bigg] \tag{A22}$$

now note that $(-1)^{|S|+|S'|} = (-1)^{|S\setminus S'|+|S'\setminus S|}$ can be distributed as factor -1 to each factor in the products corresponding to $S\setminus S'$ and $S'\setminus S$. Then the formula

$$\sum_{\substack{S\subset\{1,\ldots,n\}\\ |S|>0}} \sum_{\substack{S'\subset\{1,\ldots,n\}\\ |S'|>0}} \prod_{l\in S\cap S'} a_l \prod_{l\in S\setminus S'} (-b_l) \prod_{l\in S'\setminus S} (-c_l) = \prod_{i=1}^{n}(1 + a_i - b_i - c_i) - \prod_{i=1}^{n}(1 - b_i) - \prod_{i=1}^{n}(1 - c_i) + 1 \tag{A23}$$

yields

$$\mathbb{E}\left(\prod_{i=1}^{n} \left(1 + \phi(X_i, X_i') - \psi_i(X_i) - \psi_i(X_i')\right) - \prod_{i=1}^{n}(1 - \psi_i(X_i)) - \prod_{i=1}^{n}(1 - \psi_i(X_i')) + 1 \right) \tag{A24}$$

$$- 2\mathbb{E}\left(\prod_{i=1}^{n} \left(1 + \mathbb{E}(\phi(X_i, X_i') \mid X_i) - \psi_i(X_i) - \mathbb{E}(\psi_i(X_i'))\right) - \prod_{i=1}^{n}(1 - \psi_i(X_i)) - \prod_{i=1}^{n}(1 - \mathbb{E}(\psi_i(X_i'))) + 1 \right) \tag{A25}$$

$$+ \prod_{i=1}^{n}(1 + \mathbb{E}(\phi(X_i, X_i')) - \mathbb{E}(\psi_i(X_i)) - \mathbb{E}(\psi_i(X_i'))) - \prod_{i=1}^{n}(1 - \mathbb{E}(\psi_i(X_i))) - \prod_{i=1}^{n}(1 - \mathbb{E}(\psi_i(X_i'))) + 1. \tag{A26}$$

Finally, after using the linearity of the expectation, the last two products in the first row cancel with the second product in (A25), and the third product in (A25) cancels with the last two products in (A26); also the trailing ”+1” cancel. For the remaining products the linearity of the expectation and the definition of ϕ in (A19) yield

$$\mathbb{E}\left(\prod_{i=1}^{n}(1 - \psi_i(X_i - X_i')) \right) - 2\mathbb{E}\left(\prod_{i=1}^{n} \mathbb{E}(1 - \psi_i(X_i - X_i') \mid X_i) \right) + \prod_{i=1}^{n} \mathbb{E}(1 - \psi_i(X_i - X_i')). \tag{A27}$$

9.5 The difference of dHSIC and total multivariance for $n = 3$

Expanding the product in (3) yields by careful accounting the representation:

$$\begin{aligned} M(X_1, X_2, X_3) = &- \mathbb{E}\left(\prod_{i=1}^{3} \psi_i(X_i - X_i')\right) - 4\mathbb{E}\left(\prod_{i=1}^{3} \mathbb{E}(\psi_i(X_i - X_i') \mid X_i)\right) - 4\prod_{i=1}^{3} \mathbb{E}(\psi_i(X_i - X_i')) && \text{(A28)} \\ &+ \sum_{(i,j,k)\in\pi(1,2,3)} \Big[\mathbb{E}\left(\psi_i(X_i - X_i' \mid X_i)\psi_j(X_j - X_j')\psi_k(X_k - X_k')\right) && \text{(A29)} \\ &\qquad - \frac{1}{2}\mathbb{E}(\psi_i(X_i - X_i'))\,\mathbb{E}\left(\psi_j(X_j - X_j')\psi_k(X_k - X_k')\right) && \text{(A30)} \\ &\qquad - \mathbb{E}\left(\mathbb{E}(\psi_i(X_i - X_i') \mid X_i)\psi_j(X_j - X_j')\mathbb{E}(\psi_i(X_k - X_k') \mid X_k')\right) && \text{(A31)} \\ &\qquad + 2\mathbb{E}(\psi_i(X_i - X_i'))\,\mathbb{E}\left(\mathbb{E}(\psi_j(X_j - X_j') \mid X_j)\mathbb{E}(\psi_k(X_k - X_k') \mid X_k)\right)\Big], && \text{(A32)} \end{aligned}$$

where $\pi(1, 2, 3)$ is the set of all permutations of the vector $(1, 2, 3)$. Define

$$\begin{aligned} &H_k(X_1, \ldots, X_n) := && \text{(A33)} \\ &\sum_{\substack{S\subset\{1,\ldots,n\} \\ |S|=k}} \left[\mathbb{E}\left(\prod_{i\in S}(-\psi_i(X_i - X_i'))\right) - 2\mathbb{E}\left(\prod_{i\in S}\mathbb{E}\left(-\psi_i(X_i - X_i') \mid X_i\right)\right) + \prod_{i\in S}\mathbb{E}\left(-\psi_i(X_i - X_i')\right)\right] \end{aligned}$$

and note $H_0 = H_1 = 0$ and $H_2(X_1, \ldots, X_n) = M_2(X_1, \ldots, X_n)$ where M_2 is 2-multivariance defined in (48). Using $\prod_{i=1}^{n}(1 - \alpha_i) = 1 + \sum_{k=1}^{n} \sum_{\substack{S\subset\{1,\ldots,n\} \\ |S|=k}} \prod_{i\in S}(-\alpha_i)$ one finds for arbitrary n that dHSIC is equal to $\sum_{k=2}^{n} H_k$. Thus, recalling that $\overline{M}(X_1, \ldots, X_n) = \sum_{k=2}^{n} M_k(X_1, \ldots, X_n)$ and $M_n(X_1, \ldots, X_n) = M(X_1, \ldots, X_n)$, we find for $n = 3$

$$\begin{aligned} dHSIC(X_1, X_2, X_3) - \overline{M}(X_1, X_2, X_3) &= H_2(\ldots) + H_3(\ldots) - M_2(\ldots) - M(\ldots) && \text{(A34)} \\ &= H_3(X_1, X_2, X_3) - M(X_1, X_2, X_3). \end{aligned}$$

Thus the difference in (A34) has almost the same representation as given in (A28)-(A32), only in (A28) the factors change. We did not succeed to find any simplified representation of the remaining terms which would allow a useful distinction. Obviously the values of the measures differ, but it remains an open problem if based on this difference one of the measures should be preferred.

10 Appendix - Collection of examples

The examples are arranged in several subsections: 10.1 discusses dependencies of higher order, 10.2 illustrates various properties of multivariance. A comparisons of multivariance with other dependence measures and real data examples can be found in the main body of the paper, Sections 7.1 and 7.2.

If not mentioned otherwise: We use the Euclidean distance $\psi_i(x_i) = |x_i|$, $L = 300$ repetitions for the resampling tests (Tests 4.1, 4.3, 5.3 with (42)), sample sizes $N \in \{10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}$, significance level $\alpha = 0.05$ and 1000 runs to compute the empirical power of the tests.

The clustered dependence structure is detected using the conservative tests with significance level α (with Holm's correction for multiple tests) or using the consistent estimators of Corollary 6.2 with $\beta = \frac{1}{2}$ and $C = 2$. The rate of the correct detections (empirical power) can be found in the tables in the columns 'distribution-free detection' and 'consistent detection', respectively.

Moreover, the tables contain the power of the independence tests based on the normalized multivariances $\mathcal{M}$, $\overline{\mathcal{M}}$, $\mathcal{M}_2$ and $\mathcal{M}_3$, which are studied using the (conservative) distribution-free method and the resampling method. To avoid overloading we included only a selection of the test results in the tables. Hereto note that the test based on $\mathcal{M}_3$, should (to provide a test which is consistent against all alternatives) be preceded by a test for pairwise independence, e.g. using $\mathcal{M}_2$. Here we performed these tests independently. But we also include in the tables a test 'Comb' which combines the tests based on $\mathcal{M}_2$, $\mathcal{M}_3$ and for $n > 3$ also $\overline{\mathcal{M}}$ to a global test for the same significance level (rejecting independence if at least one p-value, adjusted by Holm's method, is significant).

For most examples the clustered dependence structure is illustrated using the test based scheme (with conservative p-value) of Section 6. Explicit values in the graphs are the values of the test statistic for a successful detection with $N = 100$ samples, if not stated otherwise.

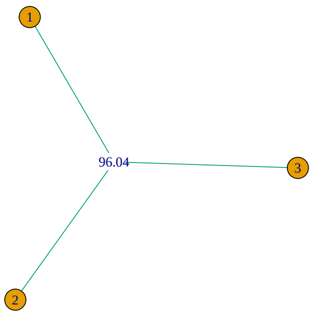
10.1 Detection and visualization of higher order dependencies

The generation of samples with higher order dependencies is explained by a detailed description of the two classical examples (Examples 10.1 and 10.2) and a basic example for dependence of arbitrary order (Example 10.3). These provide reference examples to detect and build more involved dependence structures, which also illustrate different aspects of higher order dependence: higher order dependencies with continuous marginal distributions (Example 10.4), disjoint clusters (Example 10.5), a mixture of pairwise and higher order dependence (Example 10.6), iterated dependencies (Example 10.7) and joint dependence of all variables (such that all are connected by dependencies of higher order) without any pairwise dependence (Example 10.8). The full dependence structures for the examples are collected in Example 10.9.

Example 10.1 (Coloured tetrahedron). *Consider a dice shaped as a tetrahedron with sides coloured red, green, blue and stripes of all three colours on the fourth side. The events that a particular colour is on the bottom side – when throwing this dice – are pairwise independent events. But they are not independent. Both properties follow by direct calculation:*

$$\mathbb{P}(red) = \mathbb{P}(green) = \mathbb{P}(blue) = \frac{2}{4}$$
$$\mathbb{P}(red\ and\ green) = \mathbb{P}(red\ and\ blue) = \mathbb{P}(green\ and\ blue) = \frac{1}{4}$$
$$\mathbb{P}(red)\mathbb{P}(green) = \mathbb{P}(red)\mathbb{P}(blue) = \mathbb{P}(green)\mathbb{P}(blue) = \frac{1}{4}$$
$$\mathbb{P}(red\ and\ green\ and\ blue) = \frac{1}{4} \neq \frac{1}{8} = \mathbb{P}(red)\mathbb{P}(green)\mathbb{P}(blue).$$

Thus this provides an example of three variables which are 2-independent, but dependent. In Figure A1 the empirical powers of the tests are denoted. Maybe it seems surprising that the empirical power of the test based dependence structure detection is not 1 albeit the others have power 1. Hereto recall that the distribution-free test *is sharp for Bernoulli random variables, thus (due to the correction for multiple tests) it is expected that in* 5% *of the cases already a (false) detection of pairwise dependence occurs. Furthermore, note that the distribution-free test for the normalized total multivariance has for* $N = 10$ *an empirical power of* 0 *due to averaging (see also Example 10.14).*



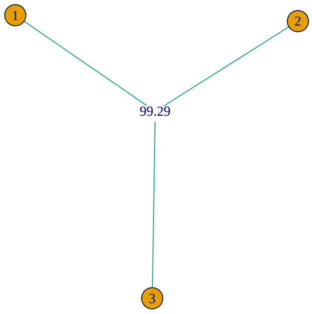
	resampling			distribution-free			consistent
N	$^{N}\mathcal{M}$	$^{N}\overline{\mathcal{M}}$	Comb	$^{N}\mathcal{M}$	$^{N}\overline{\mathcal{M}}$	detection	detection
10	0.891	0.854	0.900	0.926	0.000	0.928	0.726
20	1.000	1.000	1.000	1.000	1.000	0.955	0.991
30	1.000	1.000	1.000	1.000	1.000	0.958	0.999
40	1.000	1.000	1.000	1.000	1.000	0.943	0.999
50	1.000	1.000	1.000	1.000	1.000	0.960	0.999
60	1.000	1.000	1.000	1.000	1.000	0.958	1.000
70	1.000	1.000	1.000	1.000	1.000	0.958	1.000
80	1.000	1.000	1.000	1.000	1.000	0.943	1.000
90	1.000	1.000	1.000	1.000	1.000	0.952	1.000
100	1.000	1.000	1.000	1.000	1.000	0.949	1.000

Figure A1: Colored tetrahedron (Ex. 10.1): dependence structure, empirical power and detection rate.

Example 10.2 (Two coins — three events)**.** *Throw two fair coins and consider the three events: the first shows head, the second shows tail, both show the same. Then again a direct calculation shows pairwise independence, but dependence. The probability that all three events occur simultaneously is* 0*.*

Alternatively the same (but with a joint probability of 1/4 *as in Example 10.1) holds for the events: the first shows head, the second shows head, both show the same.*

Figure 10.2 shows the dependence structure, empirical power and detection rate for the case with joint probability 0*. The results are indistinguishable from Example 10.1.*



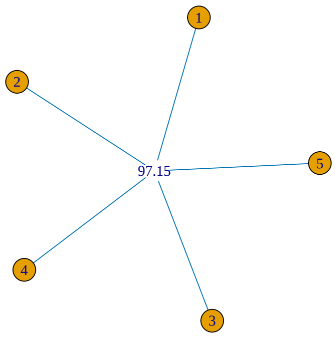
	resampling			distribution-free			consistent
N	$^{N}\mathcal{M}$	$^{N}\overline{\mathcal{M}}$	Comb	$^{N}\mathcal{M}$	$^{N}\overline{\mathcal{M}}$	detection	detection
10	0.913	0.862	0.918	0.929	0.000	0.933	0.714
20	1.000	1.000	1.000	1.000	1.000	0.957	0.988
30	1.000	1.000	1.000	1.000	1.000	0.960	0.997
40	1.000	1.000	1.000	1.000	1.000	0.961	0.999
50	1.000	1.000	1.000	1.000	1.000	0.948	1.000
60	1.000	1.000	1.000	1.000	1.000	0.961	1.000
70	1.000	1.000	1.000	1.000	1.000	0.951	1.000
80	1.000	1.000	1.000	1.000	1.000	0.948	1.000
90	1.000	1.000	1.000	1.000	1.000	0.954	1.000
100	1.000	1.000	1.000	1.000	1.000	0.962	0.999

Figure A2: Three events of two coins (Ex. 10.2): dependence structure, empirical power and detection rate.

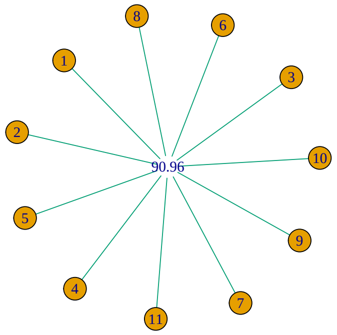
A simple generalization yields the next important example, featuring higher order dependence in its 'purest' form.

Example 10.3 (n coins — $(n+1)$ events)**.** *Throw* n *fair coins and consider the* $n+1$ *events: The first shows head, the second shows head, ..., the* n*-th shows head, there is an odd number of heads. Then by direct calculation* *these are n-independent, but dependent (the joint probability of the n+1 events is* 0 *for even n and it is* $(1/2)^n$ *for odd n). To get an intuition, note that given n of these events one can directly calculate the* $(n+1)$*th event. But given less, provides not enough information to determine any further event - any option is equally likely.*

Figure A3 shows the dependence structure and the empirical power of the tests. For total multivariance there is a loss of power compared to the previous examples due to the averaging (only one of the $2^n - n - 1$ *summands diverges, see also Example 10.14). Moreover one starts to see that the distribution-free method is conservative for total multivariance (recall that also with univariate Bernoulli marginals it is only sharp for multivariance, not for total multivariance). The low rate of successful detections of the test based dependence structure detection is again due to the sharp rejection level for Bernoulli random variables and the p-value adjustment due to multiple testing of all k-tuples for each* $k \in \{2, \dots, n+1\}$.



	resampling			distribution-free			consistent
N	${}^N\mathcal{M}$	${}^N\overline{\mathcal{M}}$	Comb	${}^N\mathcal{M}$	${}^N\overline{\mathcal{M}}$	detection	detection
10	0.814	0.165	0.076	0.792	0.001	0.755	0.348
20	0.999	0.610	0.359	1.000	0.000	0.915	0.938
30	1.000	0.961	0.812	1.000	0.000	0.902	0.990
40	1.000	1.000	0.997	1.000	0.001	0.905	0.993
50	1.000	1.000	1.000	1.000	0.007	0.896	0.999
60	1.000	1.000	1.000	1.000	0.040	0.905	1.000
70	1.000	1.000	1.000	1.000	0.137	0.906	1.000
80	1.000	1.000	1.000	1.000	0.453	0.900	0.999
90	1.000	1.000	1.000	1.000	0.901	0.886	0.999
100	1.000	1.000	1.000	1.000	1.000	0.909	1.000

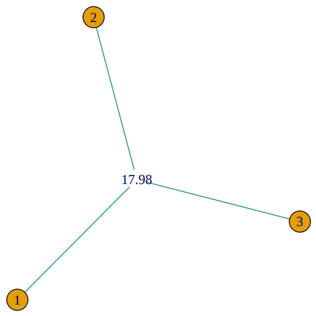


	resampling			distribution-free			consistent
N	${}^N\mathcal{M}$	${}^N\overline{\mathcal{M}}$	Comb	${}^N\mathcal{M}$	${}^N\overline{\mathcal{M}}$	detection	detection
10	0.353	0.045	0.046	0.340	0.002	0.294	0.001
20	0.987	0.056	0.044	0.991	0.000	0.895	0.065
30	1.000	0.075	0.059	1.000	0.000	0.805	0.385
40	1.000	0.114	0.077	1.000	0.000	0.805	0.671
50	1.000	0.155	0.091	1.000	0.000	0.749	0.834
60	1.000	0.179	0.102	1.000	0.000	0.751	0.905
70	1.000	0.220	0.122	1.000	0.000	0.731	0.949
80	1.000	0.261	0.151	1.000	0.000	0.760	0.975
90	1.000	0.291	0.160	1.000	0.000	0.763	0.983
100	1.000	0.337	0.226	1.000	0.000	0.738	0.990

Figure A3: Events of 4 and 10 coins (Ex. 10.3): dependence structure, empirical power and detection rate.

The previous examples only used dichotomous data. Obviously the same dependence structures can also appear (and be detected) for other marginal distributions. A basic example is the following.

Example 10.4 (Perturbed coins). *Let* (Y_1, Y_2, Y_3) *be the random variables corresponding to the events of* $n = 2$ *coins in Example 10.3 and* Z_1, Z_2, Z_3 *be i.i.d. standard normal random variables. Now set* $X_i := Y_i + rZ_i$ *for* $i = 1, 2, 3$ *and some fixed* $r \in \mathbb{R}$*. For these the same dependence structure as in Example 10.2 (Figure A2) is detected. Figure A4 shows the dependence structure and the empirical power for* $r \in \{0.25, 0.5, 0.75, 1\}$*. Note that the rate of successful detections of the test based dependence structure algorithm improves in comparison to the previous examples (for N large) whereas the consistent estimator requires larger samples. The former is due to the fact that only in the case of univariate Bernoulli distributed random variables the distribution-free method* *is sharp for multivariance. In all other cases it becomes conservative and therefore the rate of falsely detected pairwise dependencies is reduced. Increasing the value of r reduces the empirical power. This is expected, since the dependence structure becomes blurred by the variability of the Z_i's.*



	resampling			distribution-free			consistent
N	$^N\mathcal{M}$	$^N\overline{\mathcal{M}}$	Comb	$^N\mathcal{M}$	$^N\overline{\mathcal{M}}$	detection	detection
10	0.629	0.136	0.541	0.006	0.000	0.011	0.000
20	0.987	0.788	0.991	0.481	0.000	0.490	0.000
30	1.000	0.985	1.000	0.929	0.000	0.930	0.000
40	1.000	1.000	1.000	0.998	0.001	0.999	0.000
50	1.000	1.000	1.000	1.000	0.014	1.000	0.000
60	1.000	1.000	1.000	1.000	0.165	0.999	0.001
70	1.000	1.000	1.000	1.000	0.557	0.999	0.008
80	1.000	1.000	1.000	1.000	0.873	0.999	0.016
90	1.000	1.000	1.000	1.000	0.985	0.997	0.059
100	1.000	1.000	1.000	1.000	0.998	0.999	0.131

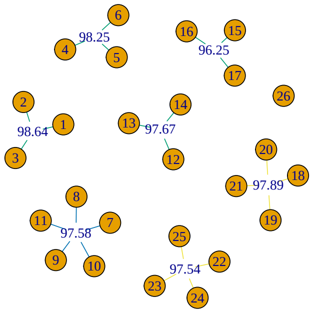
	resampling							
	$r = 0.25$		$r = 0.5$		$r = 0.75$		$r = 1$	
N	$^N\mathcal{M}$	$^N\overline{\mathcal{M}}$	$^N\mathcal{M}$	$^N\overline{\mathcal{M}}$	$^N\mathcal{M}$	$^N\overline{\mathcal{M}}$	$^N\mathcal{M}$	$^N\overline{\mathcal{M}}$
10	0.629	0.136	0.114	0.050	0.060	0.045	0.050	0.047
20	0.987	0.788	0.315	0.095	0.102	0.065	0.068	0.060
30	1.000	0.985	0.500	0.159	0.122	0.078	0.072	0.054
40	1.000	1.000	0.646	0.213	0.168	0.060	0.076	0.055
50	1.000	1.000	0.778	0.320	0.185	0.087	0.085	0.060
60	1.000	1.000	0.853	0.437	0.260	0.099	0.092	0.066
70	1.000	1.000	0.908	0.506	0.284	0.086	0.087	0.048
80	1.000	1.000	0.951	0.616	0.292	0.095	0.110	0.070
90	1.000	1.000	0.969	0.670	0.361	0.114	0.127	0.069
100	1.000	1.000	0.984	0.749	0.401	0.141	0.119	0.060

Figure A4: Normal perturbed events of 2 coins (Ex. 10.4): dependence structure, empirical power and detection rate.

Now the above examples will be used as building blocks to illustrate the dependence structure detection algorithm. For the following examples the visualized dependence structure is (at least to us) much more comprehensible than the literal description.

Example 10.5 (Several disjoint dependence clusters). *We look at samples of* $(X_1, \dots, X_{26})$ *where* (X_1, X_2, X_3) *are as in Example 10.3 with 2 coins,* $(X_7, \dots, X_{11})$ *are as in Example 10.3 with 4 coins,* (X_4, X_5, X_6) *and* (X_{12}, X_{13}, X_{14}) *and* (X_{15}, X_{16}, X_{17}) *are as in Example 10.1,* $(X_{18}, \dots, X_{21})$ *and* $(X_{22}, \dots, X_{25})$ *are as in Example 10.3 with 3 coins and* $X_{26} \sim N(0, 1)$*. Furthermore, each of these tuples is independent of the others. Note that we added* X_{26} *to make the detection much harder, since now the factorization for independent subsets* (6) *implies* $\mathcal{M}(X_1, \dots, X_{26}) = 0$.

Figure A5 shows that the detection algorithm and tests based on 3-multivariance (with resampling) perform well, whereas tests using total multivariance suffer from averaging (see also Example 10.14) and the distribution-free dependence tests are too conservative.

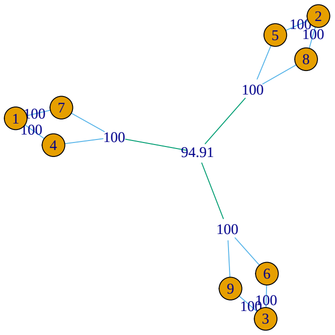


N	resampling			distribution-free			consistent
	${}^{N}\overline{\mathcal{M}}$	${}^{N}\mathcal{M}_3$	Comb	${}^{N}\overline{\mathcal{M}}$	${}^{N}\mathcal{M}_3$	detection	detection
10	0.044	0.073	0.049	0.015	0.000	0.000	0.000
20	0.047	0.214	0.116	0.000	0.000	0.000	0.007
30	0.041	0.424	0.242	0.000	0.000	0.908	0.177
40	0.045	0.654	0.465	0.000	0.000	0.945	0.508
50	0.039	0.831	0.667	0.000	0.000	0.950	0.716
60	0.053	0.947	0.855	0.000	0.000	0.920	0.835
70	0.053	0.986	0.955	0.000	0.000	0.920	0.911
80	0.047	0.998	0.989	0.000	0.000	0.916	0.966
90	0.034	1.000	0.999	0.000	0.000	0.915	0.973
100	0.051	1.000	1.000	0.000	0.000	0.923	0.981

Figure A5: The dependence structure with several clusters (Ex. 10.5).

Example 10.6 (Star dependence structure). *Consider samples of* $(X_1, X_2, X_3,\ X_1, X_2, X_3,\ X_1, X_2, X_3)$ *where* X_1, X_2, X_3 *are as in Example 10.3 with 2 coins. Then the structure in Figure A6 is detected. Here the graph was slightly cleaned up: vertices representing only pairwise dependence were reduced to edges with labels.*

The variables are Bernoulli distributed and thus (as e.g. in Example 10.3) the detection rate of 95% *reflects the* 5% *falsely detected pairwise dependencies.*

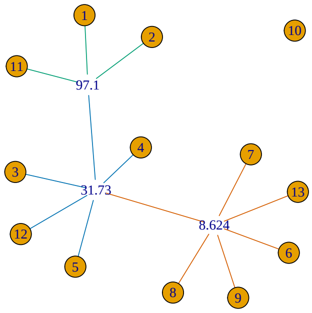


N	resampling				distribution-free				consistent
	${}^{N}\overline{\mathcal{M}}$	${}^{N}\mathcal{M}_2$	${}^{N}\mathcal{M}_3$	Comb	${}^{N}\overline{\mathcal{M}}$	${}^{N}\mathcal{M}_2$	${}^{N}\mathcal{M}_3$	detection	detection
10	1.000	1.000	0.999	1.000	0.269	0.237	0.049	0.000	0.679
20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.973	0.988
30	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.950	0.998
40	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.959	0.999
50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.960	1.000
60	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.963	1.000
70	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.954	1.000
80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.955	1.000
90	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.946	1.000
100	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.958	1.000

Figure A6: The star dependence structure of Ex. 10.6.

Example 10.7 (Iterated dependence structure). *Consider samples of random variables* $(X_1, \dots, X_{13})$ *where* $X_1, \dots, X_{10}$ *are independent but* X_1, X_2, X_{11} *are dependent (but all subtuples are independent), the same holds* *for $X_1, \ldots, X_5, X_{12}$ and $X_1, \ldots, X_9, X_{13}$. Such examples can be constructed by letting $X_{11} = f(X_1, X_2)$ for some (special) f, and analogously for the others. If such a structure is detected the graph looks like Figure A7.*

For the dependence we used $f(x_1, \ldots, x_k) = \sum_{i=1}^{k} x_i \mod 2$, and $X_i, i = 1, \ldots, 10$ were i.i.d. Bernoulli random variables. The dependence structure is reasonably detected given 100 samples by the test based algorithm, the consistent estimator requires a much large sample size. Tests based on total multivariance and 3-multivariance also detect the dependence, see Figure A7.

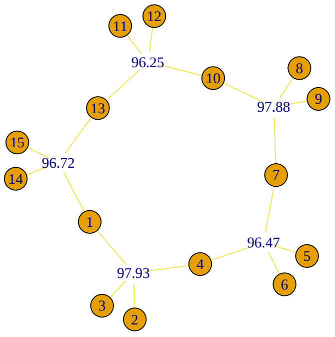


	resampling			distribution-free			consistent
N	${}^N\overline{\mathcal{M}}$	${}^N\mathcal{M}_3$	Comb	${}^N\overline{\mathcal{M}}$	${}^N\mathcal{M}_3$	detection	detection
10	0.039	0.070	0.063	0.004	0.000	0.000	0.000
20	0.061	0.161	0.107	0.000	0.000	0.000	0.000
30	0.093	0.266	0.156	0.000	0.000	0.000	0.000
40	0.152	0.395	0.258	0.000	0.000	0.000	0.000
50	0.277	0.543	0.412	0.000	0.000	0.000	0.000
60	0.416	0.742	0.590	0.000	0.000	0.001	0.000
70	0.541	0.844	0.742	0.000	0.000	0.042	0.000
80	0.667	0.918	0.853	0.000	0.000	0.556	0.000
90	0.733	0.970	0.929	0.000	0.000	0.858	0.000
100	0.842	0.984	0.972	0.000	0.000	0.889	0.000

Figure A7: The iterated dependence structure of Ex. 10.7.

Example 10.8 (Ring dependence structure). *The random variables $(X_1, \ldots, X_{15})$ are defined as follows. X_i are i.i.d. Bernoulli random variables for $i \in \{1, 2, 3, 5, 6, 8, 9, 11, 12, 14\}$, $X_k := (\sum_{i=k-3}^{k-1} X_i) \mod 2$ for $k \in \{4, 7, 10, 13\}$ and $X_{15} := (X_{13} + X_{14} + X_1) \mod 2$.*

Since here only quadruple dependence is present, only total multivariance is expected to detect it. The dependence structure detection works surprisingly well, also with small sample sizes, see Figure A8.



	resampling		distribution-free		consistent
N	${}^N\overline{\mathcal{M}}$	Comb	${}^N\overline{\mathcal{M}}$	detection	detection
10	0.040	0.063	0.002	0.000	0.010
20	0.050	0.082	0.000	0.054	0.404
30	0.089	0.120	0.000	0.939	0.750
40	0.180	0.157	0.000	0.938	0.903
50	0.305	0.256	0.000	0.948	0.961
60	0.481	0.383	0.000	0.928	0.971
70	0.620	0.510	0.000	0.925	0.983
80	0.760	0.672	0.000	0.933	0.994
90	0.851	0.769	0.000	0.931	0.998
100	0.924	0.887	0.000	0.915	0.997

Figure A8: The ring dependence structure of Ex. 10.8.

Example 10.9 (The full dependence structures). *For Examples 10.1 to 10.5 the clustered dependence structure and the full dependence structure coincide. For Examples 10.6, 10.7 and 10.8 the full dependence structures are given in Figure A9, A10 and A11, respectively. The full graph is not as easy to comprehend as the clustered graphs, to improve it we used the order of dependence as labels of the dependency nodes. Moreover, besides the* *full graph also individual graphs depicting only the dependence of a certain order (for tuples for which no lower order dependence was detected) are presented.*

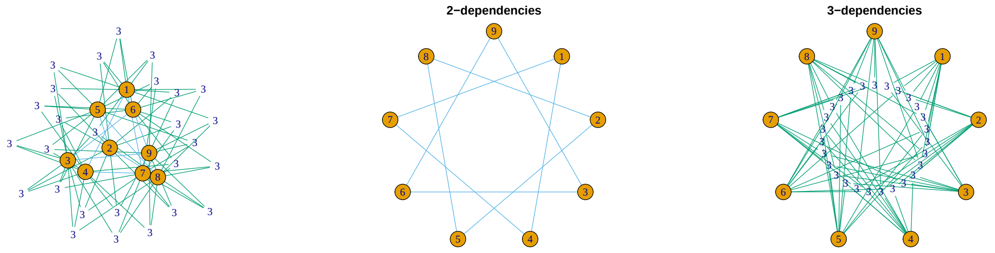


Figure A9: The full dependence structure of Ex. 10.6 (star dependence structure).

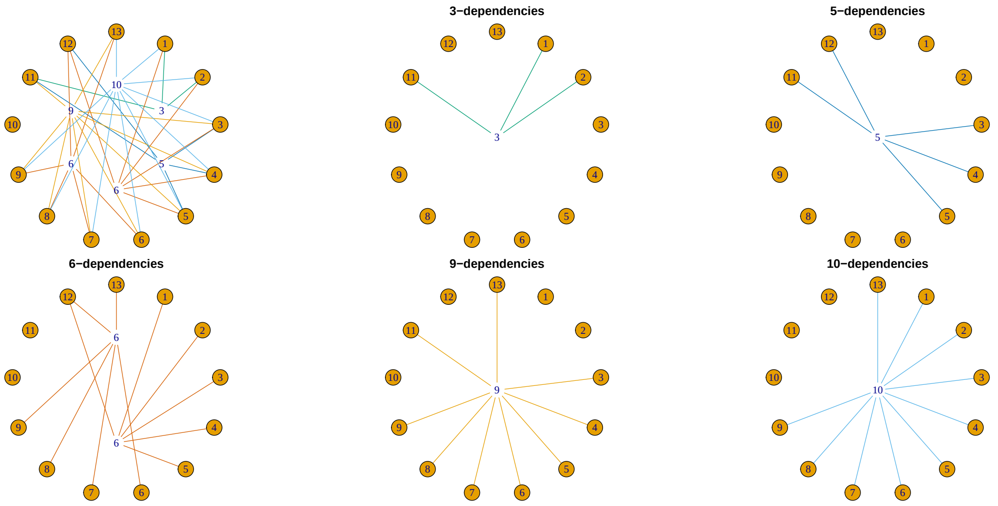


Figure A10: The full dependence structure of Ex. 10.7 (iterated dependence structure).

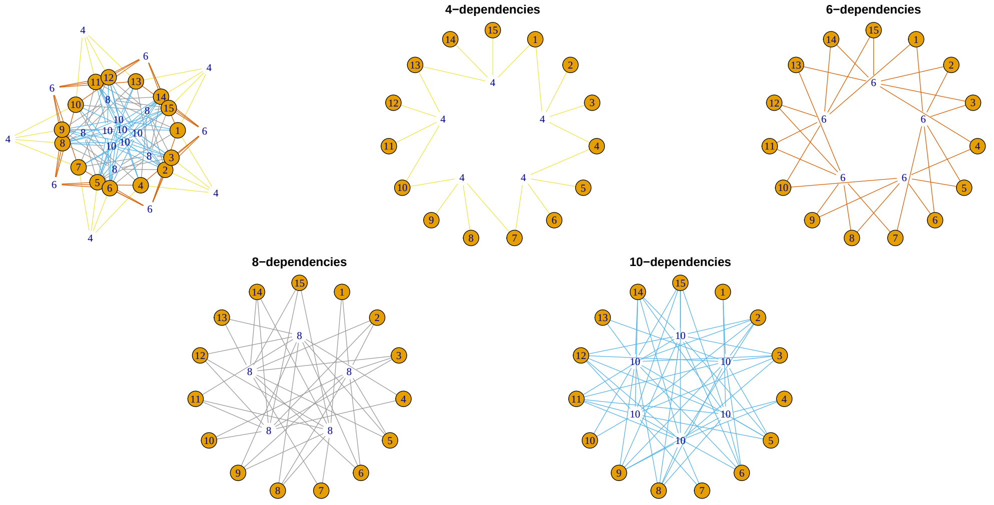


Figure A11: The full dependence structure of Ex. 10.8 (ring dependence structure).

10.2 Empirical studies of properties of distance multivariance

Note that in the papers introducing distance multivariance (Böttcher *et al.* 2018, 2019) only two very elementary examples are contained. Thus simultaneously to illustrating the a measures and methods provided in the current paper we also provide the first detailed empirical study of distance multivariance. For further related examples see also Chakraborty and Zhang (2019); Bilodeau and Guetsop Nangue (2017). The following aspects of multivariance, total multivariance and m-multivariance are discussed: the empirical size of the tests (Example 10.10), the dependence of the distribution of the test statistic on marginal distributions, sample size, dimension and the choice of ψ (Example 10.11), the computational complexity (Example 10.12), the moment conditions (Example 10.13) and the statistical curse of dimensions (Example 10.14). The section closes with a generalization of total multivariance (Example 10.15).

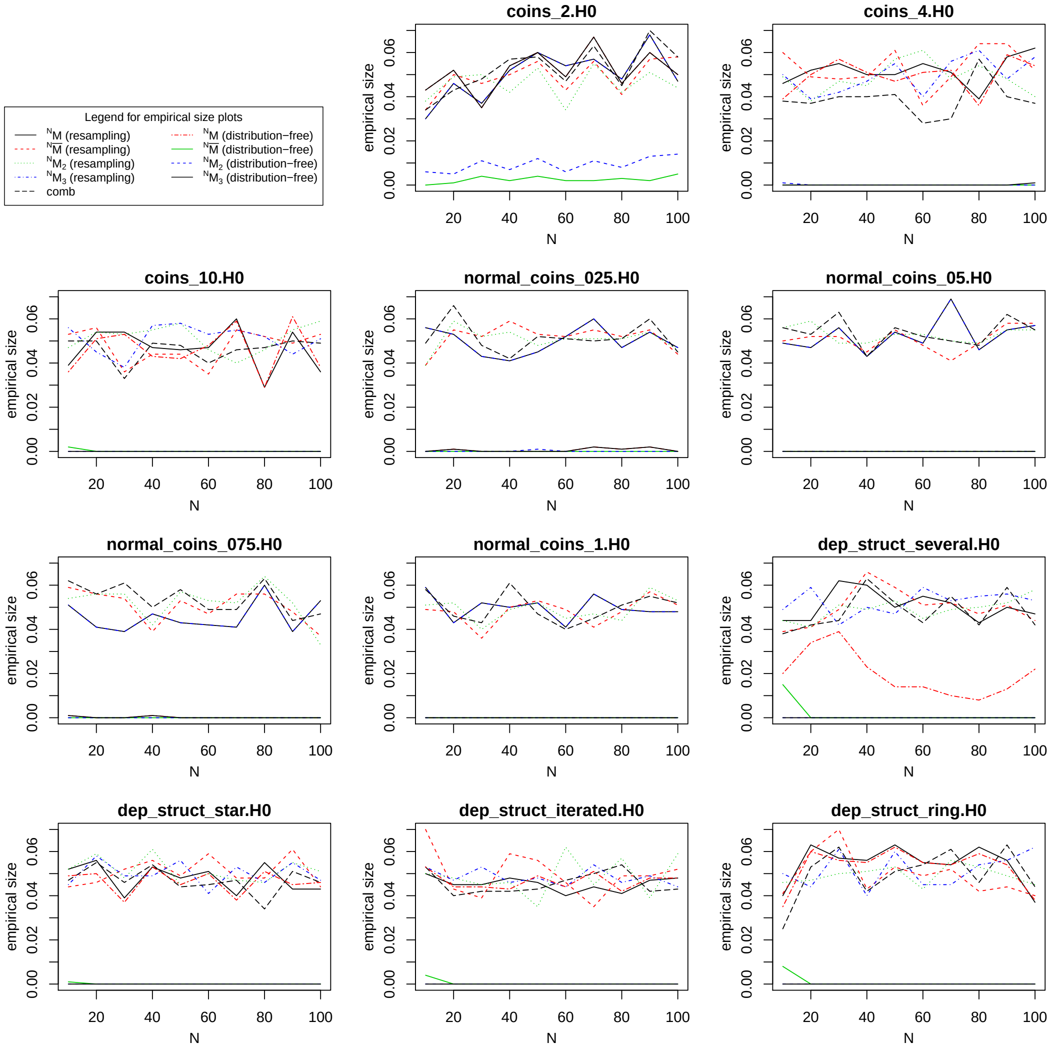


Figure A12: The empirical size of the tests for Examples 10.2 to 10.8 (Ex. 10.10).

Example 10.10 (Empirical size). *Here we consider the same settings as in the previous examples but with H_0 data, i.e., the marginal distributions remain as in the examples but the components are now independent. In Figure A12 the empirical sizes of the tests are depicted (no empirical size was above the depicted range). The resampling methods have (as expected for a sharp test) an empirical size close to 0.05. For Bernoulli marginals* *also the distribution-free method for multivariance is close to 0.05. In the other cases (and for m- and total multivariance) the tests are conservative.*

Next we analyze the effect of various parameters on the distribution of the test statistic.

Example 10.11 (Influence of sample size, marginal distributions and ψ_i). *The distribution of the test statistic $N \cdot {}^N\mathcal{M}^2$ under the hypothesis of independence depends on the marginal distributions of the random variables and also on the number of variables n as Figure A13 illustrates (see also Figure A15). The empirical distributions are based on 3000 samples each.*

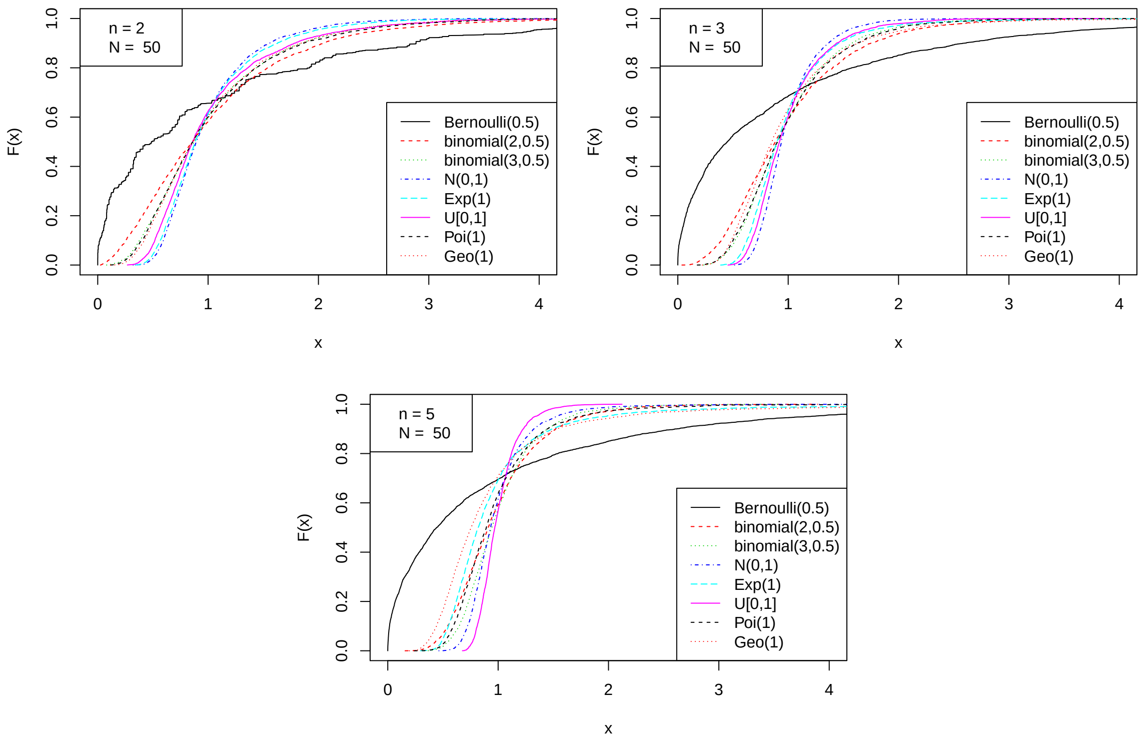


Figure A13: Empirical distribution of $N \cdot {}^N M_\rho^2(X_1, \dots, X_n)$ for i.i.d. X_i with various distributions (Ex. 10.11).

Moreover the distribution also clearly depends on the choice of the reference measure ρ or equivalently (see (11) *and Remark 2.8) on the distances determined by ψ_i. For Figure A14 we used $\psi_i(x_i) = |x_i|^\alpha$ with $\alpha \in \{0.5, 1, 1.5\}$, and the plots show that in general for $\alpha = 1.5$ the upper tail of the distribution of the test statistic comes closer to the distribution-free limit which is the χ_1^2-distribution. Note that the χ_1^2-distribution is matched in the case of Bernoulli distributed random variables, in this case the choice of ψ_i has no effect on the empirical distribution of the test statistic, since $\psi_i(0) = 0$ and $\psi_i(1) = 1$ for all α.*

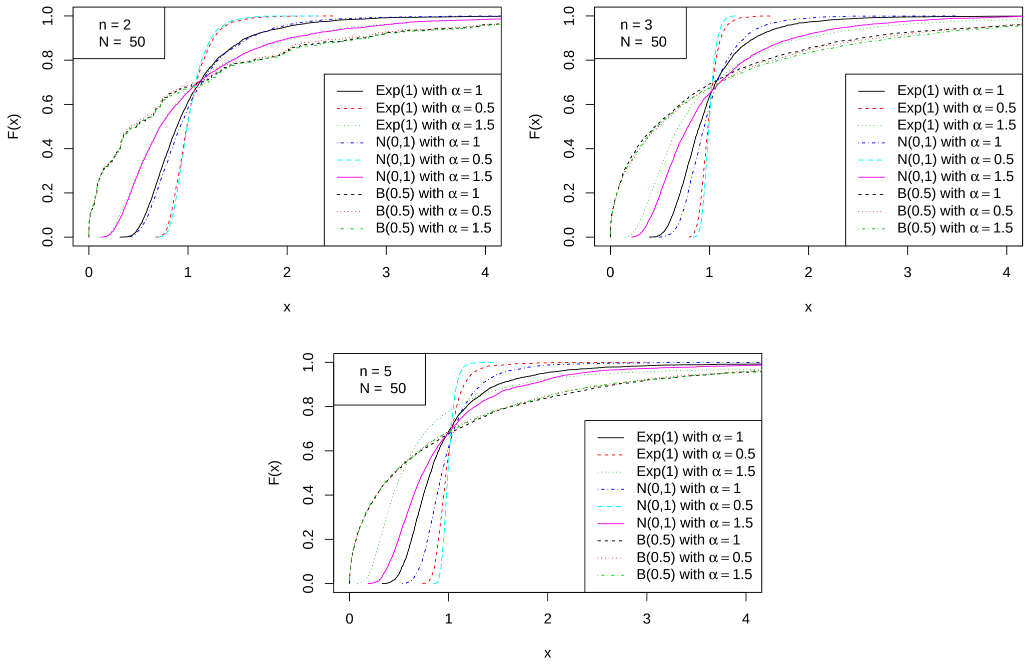


Figure A14: Empirical distribution of $N \cdot {}^{N}M_{\rho}^{2}(X_1, \ldots, X_n)$ for i.i.d. X_i with various distributions and for $\psi_i(x) = |x|^{\alpha}$ (Ex. 10.11).

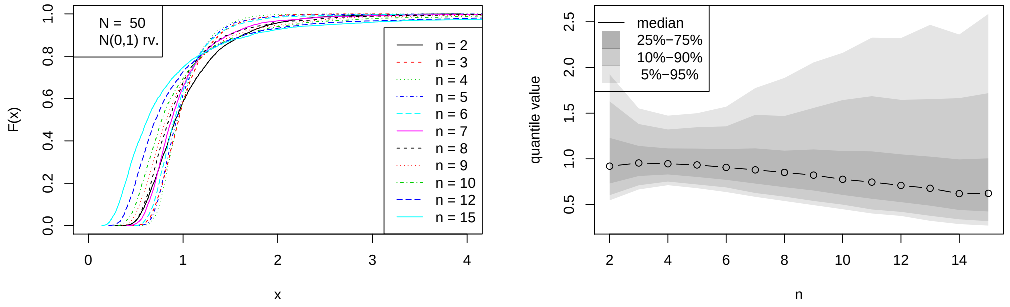


Figure A15: Dependence of the distribution of $N \cdot {}^{N}M_{\rho}^{2}(X_1, \ldots, X_n)$ for i.i.d. r.v. on the number of variables (Ex. 10.11).

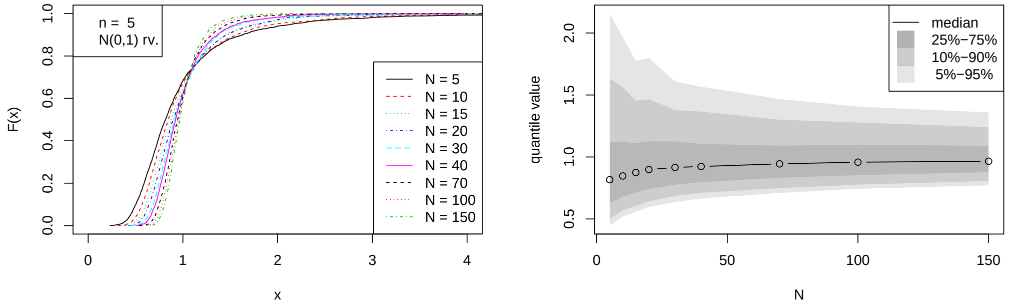


Figure A16: Dependence of the distribution of $N \cdot {}^{N}M_{\rho}^{2}(X_1, \ldots, X_n)$ for i.i.d. r.v. on the sample size (Ex. 10.11).

For independent normally distributed random variables the dependence of the test statistic on the number of variables n is depicted in Figure A15, and the dependence on the sample size N is illustrated in Figure A16. Roughly, the distribution spreads with the number of variables and shrinks to a limiting distribution (as stated in Theorem 2.5) with increasing sample size.

Example 10.12 (Computational complexity). *To illustrate that the theoretical complexity* $O(nN^2)$ *is met by the computations, we computed distance multivariance for various values of N and n (using i.i.d. normal samples). In Figure A17 the median of the computation time of 1000 repetitions for each combination of* $n \in \{2, 3, \dots, 10\}$ *and* $N \in \{10, 20, \dots, 100\}$ *is depicted. The linear growth in the dimension n and the non-linear (quadratic) growth in the number of variables N is clearly visible.*

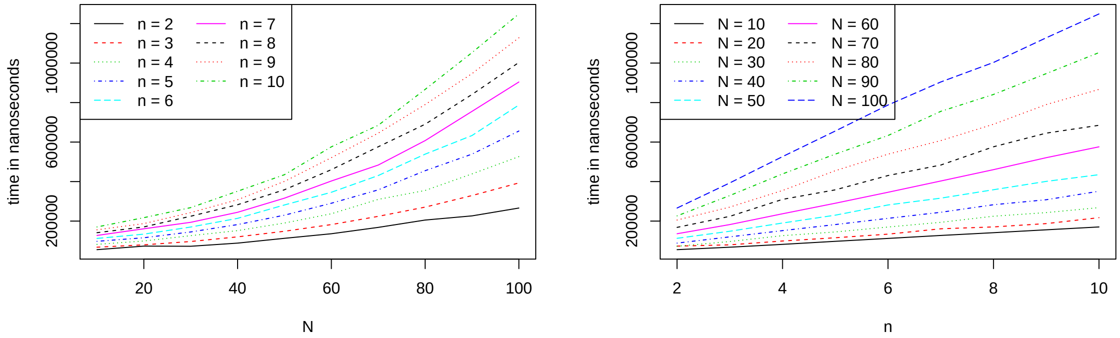
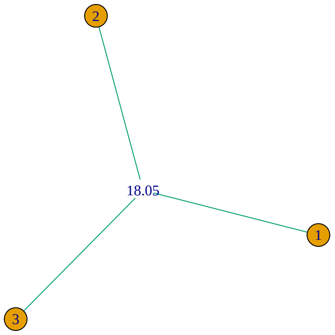


Figure A17: Dependence of the computation time of multivariance on sample size N and dimension n (Ex. 10.12).

Example 10.13 (Infinite moments – cf. Remark 2.8). *Similar to Example 10.4 let* (Y_1, Y_2, Y_3) *be the random variables corresponding to the events of* $n = 2$ *coins in Example 10.3 and* Z_1, Z_2, Z_3 *be independent Cauchy distributed random variables. Now set* $X_i := Y_i + rZ_i^3$ *for* $i = 1, 2, 3$ *and some fixed* $r \in \mathbb{R}$ *(here we only use* $r = 0.001$*). Note that* $\mathbb{E}(|X_i|^{\frac{1}{3}}) = \infty$*, thus clearly the moment condition* (23) *does not hold for the standard* $\psi_i(.) = |.|$*. Now we compare three methods: a) we don't care (thus we use the standard method); b) we use* $\psi_i(\cdot) = \ln(1 + \frac{|\cdot|^2}{2})$ *which increases slowly enough such that the moments exist; c) we consider the bounded random variables* $\arctan(X_i)$ *instead of* X_i *(cf. Remark 2.8.2). The results are shown in Figure A18. It turns out that method a) is not reliable, method b) works reasonably. In our setup method c) works best, but recall that this method destroys the translation and scale invariance of the test statistic, thus already if we shift our data it might not work anymore.*



	resampling					
	$\psi_i(\cdot) = \lvert\cdot\rvert$		$\psi_i(\cdot) = \ln(1 + \frac{\lvert\cdot\rvert^2}{2})$		$\arctan(X_i)$	
N	${}^N\mathcal{M}$	${}^N\overline{\mathcal{M}}$	${}^N\mathcal{M}$	${}^N\overline{\mathcal{M}}$	${}^N\mathcal{M}$	${}^N\overline{\mathcal{M}}$
10	0.254	0.148	0.260	0.178	0.509	0.229
20	0.244	0.169	0.266	0.158	0.919	0.693
30	0.169	0.106	0.262	0.137	0.990	0.913
40	0.139	0.102	0.237	0.148	1.000	0.987
50	0.106	0.075	0.200	0.115	1.000	0.996
60	0.080	0.060	0.199	0.085	1.000	0.998
70	0.077	0.060	0.208	0.099	1.000	1.000
80	0.087	0.064	0.225	0.102	1.000	1.000
90	0.071	0.064	0.207	0.087	1.000	1.000
100	0.067	0.057	0.241	0.093	1.000	1.000

Figure A18: Multivariance for samples of a distribution with infinite expectation (Ex. 10.13).

Example 10.14 ((total and m-)multivariance – statistical curse of dimensions). *Let $X_1, \dots, X_n$ be independent random variables and set $Y_1 := X_2$. Then (due to the independence of the X_i)*

$$\overline{M}(Y_1, X_2, \dots, X_n) - \overline{M}(X_1, \dots, X_n) = M(Y_1, X_2) - 0 = M(Y_1, X_2) > 0. \tag{A35}$$

But the corresponding difference of the estimators might be negative, as a direct calculation shows. Empirically we study this setting with X_i i.i.d. Bernoulli random variables. The empirical power of the independence test with resampling for ${}^N\mathcal{M}_2$ and ${}^N\overline{\mathcal{M}}$ is shown in Figure A19 for increasing n and various sample sizes. As expected the decrease of power is rapid for total multivariance and at least not as bad for 2-multivariance.

The resampling method was used, since the distribution-free test is not sharp in this setting. It is for univariate Bernoulli marginals only sharp for multivariance but not for total or m-multivariance ($m < n$).

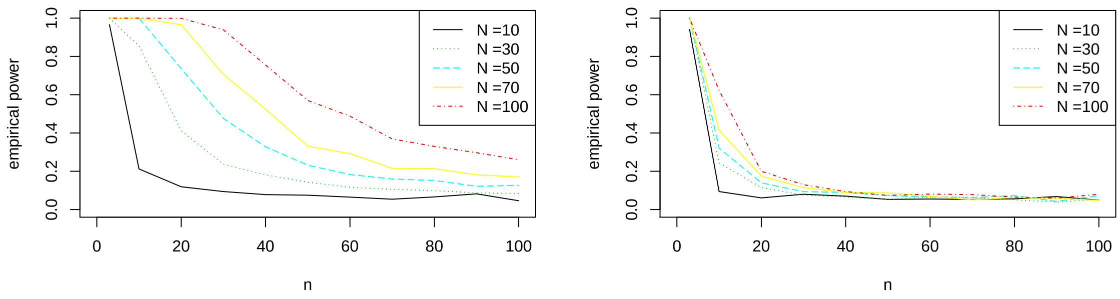


Figure A19: The curse of dimension for ${}^N\mathcal{M}_2$ (left) and ${}^N\overline{\mathcal{M}}$ (right) using the resampling rejection level (Ex. 10.14).

We close this section with an extension of total multivariance which introduces a further parameter to tune the power of the tests. Recall that we assumed, in order to avoid distracting constants, in Section 3 that the kernels of HSIC (and the related measures) satisfy $k_i(x_i, x_i) = 1$. Without this assumption additional constants appear naturally. As a special case one is led to the following dependence measure, which was incidentally suggested to us before (Martin Keller-Ressel, private communication, 2017) and it is for $\psi_i(x_i) = |x_i|$ a special case of the joint distance covariance developed in Chakraborty and Zhang (2019).

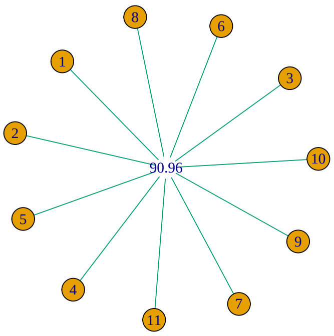
Example 10.15 (total distance multivariance with parameter λ). *Let $\lambda > 0$ and define **λ-total multivariance***

$$\overline{\mathcal{M}_\rho}^2(\lambda; X_1, \dots, X_n) := \sum_{\substack{1\le i_1<\dots<i_m\le n\\ 2\le m\le n}} M^2_{\otimes_{k=1}^m \rho_{i_k}}(X_{i_1}, \dots, X_{i_m})\lambda^{n-m} \tag{A36}$$

and its sample version

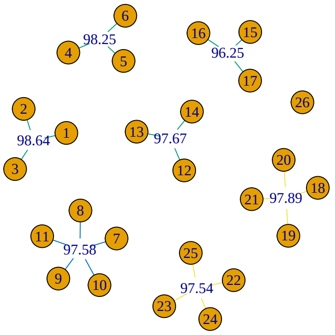
$${}^N\overline{M}^2(\lambda; \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}) := \left[\frac{1}{N^2}\sum_{j,k=1}^{N}(\lambda + (A_1)_{jk}) \cdot \ldots \cdot (\lambda + (A_n)_{jk})\right] - \lambda^n. \tag{A37}$$

Thus one puts the weight λ^{n-k} on the multivariance of each k-tuple for $k = 2, \dots, n$. Therefore with $\lambda < 1$ the n-tuple gets the biggest weight, with $\lambda > 1$ the 2-tuples (i.e., pairwise dependence) get the biggest weight. This might be used to improve the detection rate of total multivariance as Figure A20 and Figure A21 show. If the random variables are $(n-1)$-independent then clearly the detection improves when λ gets closer to 0, Figure A20. If some lower order dependence is present then some optimal λ seems to exist, Figure A21, but a priori its value seems unclear.



	resampling			
	$\lambda = 0.25$	$\lambda = 0.5$	$\lambda = 0.75$	$\lambda = 1$
N	${}^N\overline{\mathcal{M}}$	${}^N\overline{\mathcal{M}}$	${}^N\overline{\mathcal{M}}$	${}^N\overline{\mathcal{M}}$
10	0.080	0.050	0.062	0.050
20	0.763	0.160	0.059	0.057
30	0.994	0.666	0.127	0.069
40	1.000	0.962	0.304	0.110
50	1.000	1.000	0.528	0.141
60	1.000	1.000	0.772	0.201
70	1.000	1.000	0.927	0.232
80	1.000	1.000	0.985	0.280
90	1.000	1.000	1.000	0.280
100	1.000	1.000	1.000	0.311

Figure A20: Empirical power of tests based on λ-total multivariance for the events of 10 coins, compare to Figure A3 (Ex. 10.15).



	resampling			
	$\lambda = 1$	$\lambda = 2$	$\lambda = 4$	$\lambda = 8$
N	${}^N\overline{\mathcal{M}}$	${}^N\overline{\mathcal{M}}$	${}^N\overline{\mathcal{M}}$	${}^N\overline{\mathcal{M}}$
10	0.051	0.041	0.041	0.051
20	0.034	0.035	0.056	0.104
30	0.031	0.042	0.093	0.147
40	0.052	0.043	0.115	0.156
50	0.053	0.048	0.181	0.202
60	0.043	0.055	0.263	0.272
70	0.040	0.057	0.424	0.343
80	0.050	0.063	0.547	0.430
90	0.046	0.069	0.693	0.501
100	0.047	0.080	0.785	0.562

Figure A21: Empirical power of tests based on λ-total multivariance for the dependence in Example 10.5, compare to Figure A5 (Ex. 10.15).