Detecting independence of random vectors: generalized distance covariance and Gaussian covariance

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

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

bjoern.boettcher@tu-dresden.de (B. Böttcher), martin.keller-ressel@tu-dresden.de (M. Keller-Ressel), rene.schilling@tu-dresden.de (R.L. Schilling)

Abstract   Distance covariance is a quantity to measure the dependence of two random vectors. We show that the original concept introduced and developed by Székely, Rizzo and Bakirov can be embedded into a more general framework based on symmetric Lévy measures and the corresponding real-valued continuous negative definite functions. The Lévy measures replace the weight functions used in the original definition of distance covariance. All essential properties of distance covariance are preserved in this new framework.

From a practical point of view this allows less restrictive moment conditions on the underlying random variables and one can use other distance functions than Euclidean distance, e.g. Minkowski distance. Most importantly, it serves as the basic building block for distance multivariance, a quantity to measure and estimate dependence of multiple random vectors, which is introduced in a follow-up paper [Distance Multivariance: New dependence measures for random vectors (submitted). Revised version of arXiv: 1711.07775v1] to the present article.

---

*Corresponding author.

VTeX

1 Introduction

The concept of distance covariance was introduced by Székely, Rizzo and Bakirov [37] as a measure of dependence between two random vectors of arbitrary dimensions. Their starting point is to consider a weighted L^2-integral of the difference of the (joint) characteristic functions f_X, f_Y and $f_{(X,Y)}$ of the ($\mathbb{R}^m$- and $\mathbb{R}^n$-valued) random variables X, Y and (X, Y),

$$\mathcal{V}^2(X, Y; w) = \iint_{\mathbb{R}^{m+n}} |f_{(X,Y)}(s, t) - f_X(s) f_Y(t)|^2 \, w(s, t) \, \mathrm{d}s \, \mathrm{d}t. \quad (1)$$

The weight w is given by $w(s, t) := c_{\alpha,m} |s|^{-m-\alpha} c_{\alpha,n} |t|^{-n-\alpha}$ for $\alpha \in (0, 2)$.

We are going to embed this into a more general framework. In order to illustrate the new features of our approach we need to recall some results on distance covariance. Among several other interesting properties, [37] shows that distance covariance characterizes independence, in the sense that $V^2(X, Y; w) = 0$ if, and only if, X and Y are independent. Moreover, they show that in the case $\alpha = 1$ the distance covariance $^N V^2(X, Y; w)$ of the empirical distributions of two samples $(x_1, x_2, \ldots, x_N)$ and $(y_1, y_2, \ldots, y_N)$ takes a surprisingly simple form. It can be represented as

$$^N V^2(X, Y; w) = \frac{1}{N^2} \sum_{k,l=1}^{N} A_{kl} B_{kl}, \quad (2)$$

where A and B are double centrings (cf. Lemma 4.2) of the Euclidean distance matrices of the samples, i.e. of $(|x_k - x_l|)_{k,l=1,\ldots,N}$ and $(|y_k - y_l|)_{k,l=1,\ldots,N}$. If $\alpha \neq 1$ then Euclidean distance has to be replaced by its power with exponent α. The connection between the weight function w in (1) and the (centred) Euclidean distance matrices in (2) is given by the Lévy–Khintchine representation of negative definite functions, i.e.

$$|x|^\alpha = c_p \int_{\mathbb{R}^m \setminus \{0\}} (1 - \cos s \cdot x) \frac{\mathrm{d}s}{|s|^{m+\alpha}}, \quad x \in \mathbb{R}^m,$$

where c_p is a suitable constant, cf. Section 2.1, Table 1. Finally, the representation (2) of $^N V^2(X, Y; w)$ and its asymptotic properties as $N \to \infty$ are used by Székely, Rizzo and Bakirov to develop a statistical test for independence in [37].

Yet another interesting representation of distance covariance is given in the follow-up paper [34]: Let $(X_{\text{cop}}, Y_{\text{cop}})$ be an independent copy of (X, Y) and let W and W' be Brownian random fields on $\mathbb{R}^m$ and $\mathbb{R}^n$, independent from each other and from $X, Y, X_{\text{cop}}, Y_{\text{cop}}$. The paper [34] defines the *Brownian covariance*

$$\mathcal{W}^2(X, Y) = \mathbb{E}\big[X^W X_{\text{cop}}^W Y^{W'} Y_{\text{cop}}^{W'} \big], \quad (3)$$

where $X^W := W(X) - \mathbb{E}[W(X) \mid W]$ for any random variable X and random field W with matching dimensions. Surprisingly, as shown in [34], Brownian covariance coincides with distance covariance, i.e. $\mathcal{W}^2(X, Y) = \mathcal{V}^2(X, Y; w)$ when $\alpha = 1$ is chosen for the kernel w.

The paper [34] was accompanied by a series of discussion papers [25, 6, 22, 11, 15, 18, 26, 17, 35] where various extensions, applications and open questions were suggested. Let us highlight the three problems which we are going to address:

a) Can the weight function w in (1) be replaced by other weight functions? (Cf. [15, 18])

b) Can the Euclidean distance (or its α-power) in (2) be replaced by other distances? (Cf. [22, 18, 23])

c) Can the Brownian random fields W, W' in (3) be replaced by other random fields? (Cf. [22, 26])

While insights and partial results on these questions can be found in all of the mentioned discussion papers, a definitive and unifying answer was missing for a long time. In the present paper we propose a generalization of distance covariance which resolves these closely related questions. In a follow-up paper [9] we extend our results to the detection of independence of d random variables $(X^1, X^2, \ldots, X^d)$, answering a question of [15, 1].

More precisely, we introduce in Definition 3.1 the *generalized* distance covariance

$$V^2(X, Y) = \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} |f_{(X,Y)}(s, t) - f_X(s) f_Y(t)|^2 \, \mu(\mathrm{d}s) \, \nu(\mathrm{d}t),$$

where μ and ν are symmetric Lévy measures, as a natural extension of distance covariance of Székely et al. [37]. The Lévy measures μ and ν are linked to negative definite functions Φ and Ψ by the well-known Lévy–Khintchine representation, cf. Section 2 where examples and important properties of negative definite functions are discussed. In Section 3 we show that several different representations (related to [23]) of $V^2(X, Y)$ in terms of the functions Φ and Ψ can be given. In Section 4 we turn to the finite-sample properties of generalized distance covariance and show that the representation (2) of ${}^N V^2(X, Y)$ remains valid, with the Euclidean distance matrices replaced by the matrices

$$\big(\Phi(x_k - x_l)\big)_{k,l=1,\ldots,N} \quad \text{and} \quad \big(\Psi(y_k - y_l)\big)_{k,l=1,\ldots,N}.$$

We also show asymptotic properties of ${}^N V^2(X, Y)$ as N tends to infinity, paralleling those of [34, 36] for Euclidean distance covariance. After some remarks on uniqueness and normalization, we show in Section 7 that the representation (3) remains also valid, when the Brownian random fields W and W' are replaced by centered Gaussian random fields G_Φ and G_Ψ with covariance kernel

$$\mathbb{E}\big[G_\Phi(x) G_\Phi(x')\big] = \Phi(x) + \Phi(x') - \Phi(x - x')$$

and analogously for G_Ψ.

To use generalized distance covariance (and distance multivariance) in applications all necessary functions and tests are provided in the R package `multivariance` [8]. Extensive examples and simulations can be found in [7], therefore we concentrate in the current paper on the theoretical foundations.

Notation. Most of our notation is standard or self-explanatory. Throughout we use positive (and negative) in the non-strict sense, i.e. $x \geq 0$ (resp. $x \leq 0$) and we write $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$ for the maximum and minimum. For a vector $x \in \mathbb{R}^d$ the Euclidean norm is denoted by $|x|$.

2 Fundamental results

In this section we collect some tools and concepts which will be needed in the sequel.

2.1 Negative definite functions

A function $\Theta : \mathbb{R}^d \to \mathbb{C}$ is called *negative definite* (in the sense of Schoenberg) if the matrix $(\Theta(x_i) + \overline{\Theta(x_j)} - \Theta(x_i - x_j))_{i,j} \in \mathbb{C}^{m \times m}$ is positive semidefinite hermitian for every $m \in \mathbb{N}$ and $x_1, \dots, x_m \in \mathbb{R}^d$. It is not hard to see, cf. Berg & Forst [3] or Jacob [19], that this is equivalent to saying that $\Theta(0) \geq 0$, $\Theta(-x) = \overline{\Theta(x)}$ and the matrix $(-\Theta(x_i - x_j))_{i,j} \in \mathbb{C}^{m \times m}$ is conditionally positive definite, i.e.

$$\sum_{i,j=1}^{m} \left[-\Theta(x_i - x_j) \right] \lambda_i \bar{\lambda}_j \geq 0 \quad \forall \lambda_1, \dots, \lambda_m \in \mathbb{C} \text{ such that } \sum_{k=1}^{m} \lambda_k = 0.$$

Because of this equivalence, the function $-\Theta$ is also called *conditionally positive definite* (and some authors call Θ *conditionally negative definite*).

Negative definite functions appear naturally in several contexts, for instance in probability theory as characteristic exponents (i.e. logarithms of characteristic functions) of infinitely divisible laws or Lévy processes, cf. Sato [28] or [10], in harmonic analysis in connection with non-local operators, cf. Berg & Forst [3] or Jacob [19] and in geometry when it comes to characterize certain metrics in Euclidean spaces, cf. Benyamini & Lindenstrauss [2].

The following theorem, compiled from Berg & Forst [3, Sec. 7, pp. 39–48, Thm. 10.8, p. 75] and Jacob [19, Sec. 3.6–7, pp. 120–155], summarizes some basic equivalences and connections.

Theorem 2.1. *For a function $\Theta : \mathbb{R}^d \to \mathbb{C}$ with $\Theta(0) = 0$ the following assertions are equivalent*

    a) *Θ is negative definite.*

    b) *$-\Theta$ is conditionally positive definite.*

    c) *$e^{-t\Theta}$ is positive definite for every $t > 0$.*

    d) *$t^{-1}(1 - e^{-t\Theta})$ is negative definite for every $t > 0$.*

If Θ is continuous, the assertions a)–d) are also equivalent to

    e) *Θ has the following integral representation*

$$\Theta(x) = \mathrm{i}l \cdot x + \frac{1}{2} x \cdot Qx$$
$$+ \int_{\mathbb{R}^d \setminus \{0\}} \left(1 - e^{\mathrm{i}x \cdot r} + \mathrm{i}x \cdot r \mathbb{1}_{(0,1)}(|r|) \right) \rho(\mathrm{d}r), \tag{4}$$

    where $l \in \mathbb{R}^d$, $Q \in \mathbb{R}^{d \times d}$ is symmetric and positive semidefinite and ρ is a measure on $\mathbb{R}^d \setminus \{0\}$ such that $\int_{\mathbb{R}^d \setminus \{0\}} (1 \wedge |s|^2) \rho(\mathrm{d}s) < \infty$.

Table 1. Some real-valued continuous negative definite functions (cndfs) on $\mathbb{R}^d$ and the corresponding Lévy measures and infinitely divisible distributions (IDD)

cndf $\Theta(x)$	Lévy measure $\rho(\mathrm{d}r)$	IDD
$\int (1 - \cos x \cdot r)\, \rho(\mathrm{d}r)$	ρ finite measure	compound Poisson (CP)
$\frac{|x|^2}{\lambda^2 + |x|^2},\ \lambda > 0,\ x \in \mathbb{R}$	$2\lambda^{-1} e^{-\lambda |r|}\, \mathrm{d}r$	CP with exponential (1-d)
$\frac{|x|^2}{\lambda^2 + |x|^2},\ \lambda > 0,\ x \in \mathbb{R}^d$	$\int_0^\infty e^{\lambda^2 u - \frac{|r|^2}{4u}} \frac{\lambda^2\, \mathrm{d}u}{(4\pi u)^{d/2}}\, \mathrm{d}r$	CP, cf. [29, Lem. 6.1]
$1 - e^{-\frac{1}{2}|x|^2}$	$(2\pi)^{-d/2} e^{-\frac{1}{2}|r|^2}\, \mathrm{d}r$	CP with normal
$|x|$	$\dfrac{\Gamma\left(\frac{1+d}{2}\right)}{\pi^{\frac{1+d}{2}}} \dfrac{\mathrm{d}r}{|r|^{d+1}}$	Cauchy
$\frac{1}{2}|x|^2$	no Lévy measure	normal
$|x|^\alpha,\ \alpha \in (0,2)$	$\dfrac{\alpha 2^{\alpha-1}\Gamma\left(\frac{\alpha+d}{2}\right)}{\pi^{d/2}\Gamma\left(1-\frac{\alpha}{2}\right)} \dfrac{\mathrm{d}r}{|r|^{d+\alpha}}$	α-stable
$\sqrt[p]{\sum_{k=1}^d |x_k|^p},\ p \in [1,2]$	see Lemma 2.2	see Lemma 2.2
$\ln(1 + \frac{x^2}{2}),\ x \in \mathbb{R}$	$\dfrac{1}{|r|} e^{-\sqrt{\frac{1}{2}}|r|}\, \mathrm{d}r$	variance Gamma (1-d)
$\ln \cosh(x),\ x \in \mathbb{R}$	$\dfrac{\mathrm{d}r}{2r \sinh(\pi r/2)}$	Meixner (1-d)
$|x|^\alpha + |x|^\beta,\ \alpha, \beta \in (0,2)$		mixture of stable
$(1 + |x|^\alpha)^{\frac{1}{\beta}} - 1,$ $\alpha \in (0,2), \beta \geq \frac{\alpha}{2}$		relativistic stable

We will frequently use the abbreviation *cndf* instead of *continuous negative definite function*. The representation (4) is the *Lévy–Khintchine formula* and any measure ρ satisfying

$$\rho \text{ is a measure on } \mathbb{R}^d \setminus \{0\} \text{ such that } \int_{\mathbb{R}^d \setminus \{0\}} \left(1 \wedge |r|^2\right) \rho(\mathrm{d}r) < \infty \qquad (5)$$

is commonly called *Lévy measure*. To keep notation simple, we will write $\int \cdots \rho(\mathrm{d}r)$ or $\int_{\mathbb{R}^d} \cdots \rho(\mathrm{d}r)$ instead of the more precise $\int_{\mathbb{R}^d \setminus \{0\}} \cdots \rho(\mathrm{d}r)$.

The triplet (l, Q, ρ) uniquely determines Θ; moreover Θ is real (hence, positive) if, and only if, $l = 0$ and ρ is symmetric, i.e. $\rho(B) = \rho(-B)$ for any Borel set $B \subset \mathbb{R}^d \setminus \{0\}$. In this case (4) becomes

$$\Theta(x) = \frac{1}{2} x \cdot Qx + \int_{\mathbb{R}^d} (1 - \cos x \cdot r)\, \rho(\mathrm{d}r). \qquad (6)$$

Using the representation (4) it is straightforward to see that we have $\sup_x |\Theta(x)| < \infty$ if ρ is a finite measure, i.e., $\rho(\mathbb{R}^d \setminus \{0\}) < \infty$, and $Q = 0$. The converse is also true, see [29, pp. 1390–1391, Lem. 6.2].

Table 1 contains some examples of continuous negative definite functions along with the corresponding Lévy measures and infinitely divisible laws.

A measure ρ on a topological space X is said to have *full* (*topological*) *support*, if $\mu(G) > 0$ for any open set $G \subset X$; for Lévy measures we have $X = \mathbb{R}^d \setminus \{0\}$.

Lemma 2.2. *Let* $p \in [1, 2]$*. The Minkowski distance function*

$$\ell_p(x) := \left(|x_1|^p + \cdots + |x_d|^p\right)^{1/p}, \quad x = (x_1, \ldots, x_d) \in \mathbb{R}^d$$

is a continuous negative definite function on $\mathbb{R}^d$*. If* $p \in (1, 2]$*, the Lévy measure has full support.*

It is interesting to note that the Minkowski distances for $p > 2$ and $d \geq 2$ are never negative definite functions. This is the consequence of Schoenberg's problem, cf. Zastavnyi [40, p. 56, Eq. (3)].

Proof of Lemma 2.2. Since each $x_i \mapsto |x_i|^p$, with $p \in [1, 2]$, is a one-dimensional continuous negative definite function, we can use the formula (6) to see that

$$\ell_p^p(x) = \begin{cases} \int_{\mathbb{R}^d} (1 - \cos x \cdot r) \sum_{i=1}^d \dfrac{c_p \, dx_i}{|x_i|^{1+p}} \otimes \delta_0(dx_{(i)}), & \text{if } p \in [1, 2), \\ x \cdot x, & \text{if } p = 2, \end{cases}$$

where $x_{(i)} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots x_d) \in \mathbb{R}^{d-1}$ and $c_p = \dfrac{p2^{p-1}\Gamma(\frac{p+1}{2})}{\pi^{1/2}\Gamma(1-\frac{p}{2})}$ is the constant of the one-dimensional p-stable Lévy measure, cf. Table 1.

This means that ℓ_p^p is itself a continuous negative definite function, but its Lévy measure is concentrated on the coordinate axes. Writing $\ell_p(x) = f_p(\ell_p^p(x))$ with

$$f_p(\tau) = \tau^{1/p} = \gamma_{1/p} \int_0^\infty \left(1 - e^{-\tau t}\right) \frac{dt}{t^{1+1/p}}, \quad \frac{1}{p} \in \left[\frac{1}{2}, 1\right], \ \gamma_{1/p} = \frac{1}{p\Gamma(1 - \frac{1}{p})},$$

shows that ℓ_p can be represented as a combination of the Bernstein function f_p and the negative definite function ℓ_p^p. In other words, ℓ_p is subordinate to ℓ_p^p in the sense of Bochner (cf. Sato [28, Chap. 30] or [30, Chap. 5, Chap. 13.1]) and it is possible to find the corresponding Lévy–Khintchine representation, cf. [28, Thm. 30.1]. We have

$$\ell_p(x) = \begin{cases} \int_{\mathbb{R}^d} (1 - \cos x \cdot r) \sum_{i=1}^d \dfrac{c_p \, dx_i}{|x_i|^2} \otimes \delta_0(dx_{(i)}), & \text{if } p = 1, \\ \int_{\mathbb{R}^d} (1 - \cos x \cdot r) \int_0^\infty \prod_{i=1}^d g_t(x_i) \dfrac{dt}{t^{1+1/p}}, & \text{if } p \in (1, 2), \\ \sqrt{x \cdot x}, & \text{if } p = 2, \end{cases}$$

where $x_i \mapsto g_t(x_i)$ is the probability density of the random variable $t^p X$ where X is a one-dimensional, symmetric $1/p$-stable random variable.

Although the $1/p$-stable density is known explicitly only for $1/p \in \{1, 2\}$, one can show – this follows, e.g. from [28, Thm. 15.10] – that it is strictly positive, i.e. the Lévy measure of ℓ_p, $p \in (1, 2)$ has full support. For $p = 1$ the measure does not have full support, since it is concentrated on the coordinate axes. For $p = 2$, note that $\ell_2(x) = |x|$ corresponds to the Cauchy distribution with Lévy measure given in Table 1, which has full support. $\square$

Using the Lévy–Khintchine representation (6) it is not hard to see, cf. [19, Lem. 3.6.21], that square roots of real-valued cndfs are subadditive, i.e.

$$\sqrt{\Theta(x+y)} \le \sqrt{\Theta(x)} + \sqrt{\Theta(y)}, \quad x, y \in \mathbb{R}^d \tag{7}$$

and, consequently,

$$\Theta(x+y) \le 2\big(\Theta(x) + \Theta(y)\big), \quad x, y \in \mathbb{R}^d. \tag{8}$$

Using a standard argument, e.g. [10, p. 44], we can derive from (7), (8) that cndfs grow at most quadratically as $x \to \infty$,

$$\Theta(x) \le 2 \sup_{|y| \le 1} \Theta(y)\big(1 + |x|^2\big). \tag{9}$$

We will assume that $\Theta(0) = 0$ is the only zero of the function Θ – incidentally, this means that $x \mapsto e^{-\Theta(x)}$ is the characteristic function of a(n infinitely divisible) random variable the distribution of which is non-lattice. This and (7) show that $(x, y) \mapsto \sqrt{\Theta(x-y)}$ is a metric on $\mathbb{R}^d$ and $(x, y) \mapsto \Theta(x-y)$ is a quasi-metric, i.e. a function which enjoys all properties of a metric, but the triangle inequality holds with a multiplicative constant $c > 1$. Metric measure spaces of this type have been investigated by Jacob et al. [20]. Historically, the notion of negative definiteness has been introduced by I.J. Schoenberg [31] in a geometric context: he observed that for a real-valued cndf Θ the function $d_\Theta(x, y) := \sqrt{\Theta(x-y)}$ is a metric on $\mathbb{R}^d$ and that these are the only metrics such that $(\mathbb{R}^d, d_\Theta)$ can be isometrically embedded into a Hilbert space. In other words: d_Θ behaves like a standard Euclidean metric in a possibly infinite-dimensional space.

2.2 Measuring independence of random variables with metrics

Let X, Y be random variables with values in $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively, and write $\mathcal{L}(X)$ and $\mathcal{L}(Y)$ for the corresponding probability laws. For any metric $d(\cdot, \cdot)$ defined on the family of $(m+n)$-dimensional probability distributions we have

$$d\big(\mathcal{L}(X, Y), \mathcal{L}(X) \otimes \mathcal{L}(Y)\big) = 0 \quad \text{if, and only if, } X, Y \text{ are independent.} \tag{10}$$

This equivalence can obviously be extended to finitely many random variables X_i, $i = 1, \ldots, n$, taking values in $\mathbb{R}^{d_i}$, respectively: Set $d := d_1 + \cdots + d_n$, take any metric $d(\cdot, \cdot)$ on the d-dimensional probability distributions and consider $d(\mathcal{L}(X_1, \ldots, X_n), \bigotimes_{i=1}^n \mathcal{L}(X_i))$. Moreover, the random variables $X_i, i = 1, \ldots, n$, are independent if, and only if, $(X_1, \ldots, X_{k-1})$ and X_k are independent for all $2 \le k \le n$.[2] In other words: $X_1, \ldots, X_n$ are independent if, and only if, for metrics on the $d_1 + \cdots + d_k$-dimensional probability distributions the distance of $\mathcal{L}(X_1, \ldots, X_k)$ and $\mathcal{L}(X_1, \ldots, X_{k-1}) \otimes \mathcal{L}(X_k)$ is zero for $k = 2, \ldots, n$. Thus, as in (10), only the concept of independence of pairs of random variables is needed. In [9, Sec. 3.1] we use a variant of this idea to characterize multivariate independence.

---

[2]This is an immediate consequence of the characterization of independence using characteristic functions: X, Y are independent if, and only if, $\mathbb{E}e^{i\xi \cdot X + i\eta \cdot Y} = \mathbb{E}e^{i\xi \cdot X}\mathbb{E}e^{i\eta \cdot Y}$ for all ξ, η.

Thus (10) is a good starting point for the construction of (new) estimators for independence. For this it is crucial that the (empirical) distance be computationally feasible. For discrete distributions with finitely many values this yields the classical chi-squared test of independence (using the χ^2-distance). For more general distributions other commonly used distances (e.g. relative entropy, Hellinger distance, total variation, Prokhorov distance, Wasserstein distance) might be employed (e.g. [4]), provided that they are computationally feasible. It turns out that the latter is, in particular, satisfied by the following distance.

Definition 2.3. Let U, V be d-dimensional random variables and denote by f_U, f_V their characteristic functions. For any symmetric measure ρ on $\mathbb{R}^d \setminus \{0\}$ with full support we define the distance

$$d_\rho\big(\mathcal{L}(U), \mathcal{L}(V)\big) := \| f_U - f_V \|_{L^2(\rho)} = \left(\int |f_U(r) - f_V(r)|^2 \, \rho(\mathrm{d}r) \right)^{1/2}. \quad (11)$$

The assumption that ρ has full support, i.e. $\rho(G) > 0$ for every nonempty open set $G \subset \mathbb{R}^d \setminus \{0\}$, ensures that $d_\rho(\mathcal{L}(U), \mathcal{L}(V)) = 0$ if, and only if, $\mathcal{L}(U) = \mathcal{L}(V)$, hence $d_\rho(\mathcal{L}(U), \mathcal{L}(V))$ is a metric. The symmetry assumption on ρ is not essential since the integrand appearing in (11) is even; therefore, we can always replace $\rho(\mathrm{d}r)$ by its symmetrization $\frac{1}{2}(\rho(\mathrm{d}r) + \rho(-\mathrm{d}r))$.

Currently it is unknown how the fact that the Lévy measure ρ has full support can be expressed in terms of the cndf $\Theta(u) = \int (1 - \cos u \cdot r)\, \rho(\mathrm{d}r)$ given by ρ, see (6).

Note that $d_\rho(\mathcal{L}(U), \mathcal{L}(V))$ is always well-defined in $[0, \infty]$. Any of the following conditions ensure that $d_\rho(\mathcal{L}(U), \mathcal{L}(V))$ is finite:

a) ρ is a finite measure;

b) ρ is a symmetric Lévy measure (cf. (5)) and $\mathbb{E}|U| + \mathbb{E}|V| < \infty$.

   Indeed, $d_\rho(\mathcal{L}(U), \mathcal{L}(V)) < \infty$ follows from the integrability properties (5) of the Lévy measure ρ and the elementary estimates

$$|f_U(r) - f_V(r)| \leq \mathbb{E}|e^{\mathrm{i}r \cdot U} - e^{\mathrm{i}r \cdot V}| \leq \mathbb{E}|r \cdot (U - V)| \leq |r| \cdot \mathbb{E}|U - V|$$
$$\text{and} \quad |f_U(r) - f_V(r)| \leq 2.$$

We obtain further sufficient conditions for $V(X, Y) < \infty$ in terms of moments of the real-valued cndf Θ, see (6), whose Lévy measure is ρ and with $Q = 0$.

Proposition 2.4. Let ρ be a symmetric Lévy measure on $\mathbb{R}^d \setminus \{0\}$ with full support and denote by $\Theta(u) = \int (1 - \cos u \cdot r)\, \rho(\mathrm{d}r)$, $u \in \mathbb{R}^d$, the real-valued cndf with Lévy triplet $(l = 0, Q = 0, \rho)$. For all d-dimensional random variables U, V the following assertions hold:

a) Assume that (U', V') is an i.i.d. copy of (U, V). Then

$$\mathbb{E}\Theta\big(U - U'\big) + \mathbb{E}\Theta\big(V - V'\big) \leq 2\mathbb{E}\Theta\big(U - V'\big). \quad (12)$$

b) Let U' be an i.i.d. copy of U. Then

$$\mathbb{E}\Theta(U - V) \leq 2\big(\mathbb{E}\Theta(U) + \mathbb{E}\Theta(V)\big) \quad (13)$$

and for $V = U'$ one has $\mathbb{E}\Theta(U - U') \leq 4\mathbb{E}\Theta(U)$.

c) *In* $[0, \infty]$ *we always have*

$$d_\rho^2\big(\mathcal{L}(U), \mathcal{L}(V)\big) \le 4\big(\mathbb{E}\Theta(U) + \mathbb{E}\Theta(V)\big). \tag{14}$$

d) *Let* (U', V') *be an i.i.d. copy of* (U, V) *and assume* $\mathbb{E}\Theta(U) + \mathbb{E}\Theta(V) < \infty$.
Then the following equality holds and all terms are finite

$$d_\rho^2\big(\mathcal{L}(U), \mathcal{L}(V)\big) = 2\mathbb{E}\Theta\big(U - V'\big) - \mathbb{E}\Theta\big(U - U'\big) - \mathbb{E}\Theta\big(V - V'\big). \tag{15}$$

Proof. Let us assume first that ρ is a finite Lévy measure, thus, Θ a bounded cndf. We denote by U' an i.i.d. copy of U. Since $\mathcal{L}(U - U')$ is symmetric, we can use Tonelli's theorem to get

$$\int \big(1 - f_U(r)\overline{f_U(r)}\big)\rho(dr) = \int \big(1 - \mathbb{E}\big(e^{iU \cdot r}e^{-iU' \cdot r}\big)\big)\rho(dr)$$

$$= \int \mathbb{E}\big(1 - \cos\big[(U - U') \cdot r\big]\big)\rho(dr)$$

$$= \mathbb{E}\Theta\big(U - U'\big)$$

$$= \int \Theta\big(u - u'\big)\mathbb{P}_U \otimes \mathbb{P}_U\big(du, du'\big). \tag{16}$$

Now we consider an i.i.d. copy (U', V') of (U, V) and use the above equality in (11). This yields

$$d_\rho^2\big(\mathcal{L}(U), \mathcal{L}(V)\big)$$

$$= \int \big(f_U(r)\overline{f_U(r)} - f_U(r)\overline{f_V(r)} - f_V(r)\overline{f_U(r)} + f_V(r)\overline{f_V(r)}\big)\rho(dr)$$

$$= \int \big(f_U(r)\overline{f_U(r)} - 1 + 2 - 2\operatorname{Re}\big(f_U(r)\overline{f_V(r)}\big) + f_V(r)\overline{f_V(r)} - 1\big)\rho(dr)$$

$$= 2\mathbb{E}\Theta\big(U - V'\big) - \mathbb{E}\Theta\big(U - U'\big) - \mathbb{E}\Theta\big(V - V'\big). \tag{17}$$

This proves (15) and, since $d_\rho(\mathcal{L}(U), \mathcal{L}(V)) \ge 0$, also (12). Combining Part b) and (15) yields (14), while (13) immediately follows from the subadditivity of a cndf (8).

If ρ is an arbitrary Lévy measure, its truncation $\rho_\epsilon(dr) := \mathbb{1}_{(\epsilon, \infty)}(|r|)\, \rho(dr)$ is a finite Lévy measure and the corresponding cndf Θ_ϵ is bounded. In particular, we have a)–d) for ρ_ϵ and Θ_ϵ. Using monotone convergence we get

$$\Theta(u) = \sup_{\epsilon > 0} \Theta_\epsilon(u) = \sup_{\epsilon > 0} \int_{|r| > \epsilon} (1 - \cos u \cdot r)\, \rho(dr).$$

Again by monotone convergence we see that the assertions a)–c) remain valid for general Lévy measures – if we allow the expressions to attain values in $[0, \infty]$. Because of (13), the moment condition assumed in Part d) ensures that the limits

$$\lim_{\epsilon \to 0} \mathbb{E}\Theta\big(U - V'\big) = \sup_{\epsilon > 0} \mathbb{E}\Theta\big(U - V'\big) \quad \text{etc.}$$

are finite, and (15) carries over to the general situation. $\qquad\square$

Remark 2.5. a) Since U and V play symmetric roles in (17) it is clear that

$$d_\rho^2\big(\mathcal{L}(U), \mathcal{L}(V)\big)$$
$$= \mathbb{E}\Theta\big(U - V'\big) + \mathbb{E}\Theta\big(V - U'\big) - \mathbb{E}\Theta\big(U - U'\big) - \mathbb{E}\Theta\big(V - V'\big)$$
$$= \int \Theta\big(u - u'\big)\,(2\mathbb{P}_U \otimes \mathbb{P}_V - \mathbb{P}_U \otimes \mathbb{P}_U - \mathbb{P}_V \otimes \mathbb{P}_V)\big(\mathrm{d}u, \mathrm{d}u'\big). \qquad (18)$$

b) While $d_\rho(\mathcal{L}(U), \mathcal{L}(V)) \in [0, \infty]$ is always defined, the right-hand side of (15) needs attention. If we do not assume the moment condition $\mathbb{E}\Theta(U) + \mathbb{E}\Theta(V) < \infty$, we still have

$$d_\rho^2\big(\mathcal{L}(U), \mathcal{L}(V)\big) = \lim_{\epsilon \to 0}\big(2\mathbb{E}\Theta_\epsilon\big(U - V'\big) - \Theta_\epsilon\big(U - U'\big) - \Theta_\epsilon\big(V - V'\big)\big), \quad (19)$$

but it is not clear whether the limits exist for each term.

The moment condition $\mathbb{E}\Theta(U) + \mathbb{E}\Theta(V) < \infty$ is sharp in the sense that it follows from $\mathbb{E}\Theta(U - V') < \infty$: Since U and V' are independent, Tonelli's theorem entails that $\mathbb{E}\Theta(u - V') < \infty$ for some $u \in \mathbb{R}^d$. Using the symmetry and sub-additivity of $\sqrt{\Theta}$, see (8), we get $\Theta(V') \leq 2(\Theta(u - V') + \Theta(u))$, i.e. $\mathbb{E}\Theta(V') < \infty$; $\mathbb{E}\Theta(U) < \infty$ follows in a similar fashion.

c) Since a cndf Θ grows at most quadratically at infinity, see (9), it is clear that

$$\mathbb{E}\big(|U|^2\big) + \mathbb{E}\big(|V|^2\big) < \infty \quad \text{implies} \quad \mathbb{E}\Theta(U) + \mathbb{E}\Theta(V) < \infty. \qquad (20)$$

One should compare this to the condition $\mathbb{E}|U| + \mathbb{E}|V| < \infty$ which ensures the finiteness of $d_\rho^2(\mathcal{L}(U), \mathcal{L}(V))$, but not necessarily the finiteness of the terms appearing in the representation (15).

d) As described at the beginning of this section, a measure of independence of $X_1, \ldots, X_n$ is given by $d_\rho(\mathcal{L}(X_1, \ldots, X_n), \bigotimes_{i=1}^n \mathcal{L}(X_i))$. This can be estimated by empirical estimators for (15). For the 1-stable (i.e. Cauchy) cndf, see Table 1, this direct approach to (multivariate) independence has recently been proposed by [21] – but the *exact* estimators become computationally challenging even for small samples. A further approximation recovers a computationally feasible estimation, resulting in a loss of power compared with our approach, cf. [7].

It is worth mentioning that the metric d_ρ can be used to describe convergence in distribution.

Lemma 2.6. *Let ρ be a finite symmetric measure with full support, then d_ρ given in* (11) *is a metric which characterizes convergence in distribution, i.e. for random variables X_n, $n \in \mathbb{N}$, and X one has*

$$X_n \xrightarrow[n \to \infty]{d} X \iff d_\rho\big(\mathcal{L}(X_n), \mathcal{L}(X)\big) \xrightarrow[n \to \infty]{} 0. \qquad (21)$$

The proof below shows that the implication "$\Longleftarrow$" does not need the finiteness of the Lévy measure ρ.

Proof. Convergence in distribution implies pointwise convergence of the characteristic functions. Therefore, we see by dominated convergence and because of the obvious estimates $|f_{X_n}| \leq 1$ and $|f_X| \leq 1$ that

$$\lim_{n \to \infty} \int_{\mathbb{R}^d} |f_{X_n}(r) - f_X(r)|^2\, \rho(\mathrm{d}r) = \int_{\mathbb{R}^d} \lim_{n \to \infty} |f_{X_n}(r) - f_X(r)|^2\, \rho(\mathrm{d}r) = 0.$$

Conversely, assume that $\lim_{n\to\infty} d_\rho(\mathcal{L}(X_n), \mathcal{L}(X)) = 0$. If we interpret this as convergence in $L^2(\rho)$, we see that there is a Lebesgue a.e. convergent subsequence $f_{X_{n(k)}} \to f_X$; since $f_{X_{n(k)}}$ and f_X are characteristic functions, this convergence is already pointwise, hence locally uniform, see Sasvári [27, Thm. 1.5.2]. By Lévy's continuity theorem, this entails the convergence in distribution of the corresponding random variables. Since the limit does not depend on the subsequence, the whole sequence must converge in distribution. $\square$

2.3 An elementary estimate for log-moments

Later on we need certain log-moments of the norm of a random vector. The following lemma allows us to formulate these moment conditions in terms of the coordinate processes.

Lemma 2.7. *Let X, Y be one-dimensional random variables and $\epsilon > 0$. Then $\mathbb{E} \log^{1+\epsilon}(1 \vee \sqrt{X^2 + Y^2})$ is finite if, and only if, the moments $\mathbb{E} \log^{1+\epsilon}(1 + X^2)$ and $\mathbb{E} \log^{1+\epsilon}(1 + Y^2)$ are finite.*

Proof. Assume that $\mathbb{E} \log^{1+\epsilon}(1 + X^2) + \mathbb{E} \log^{1+\epsilon}(1 + Y^2) < \infty$. Since

$$1 \vee \sqrt{X^2 + Y^2} = \sqrt{(X^2 + Y^2) \vee 1} \le \sqrt{(1 + X^2)(1 + Y^2)},$$

we can use the elementary estimate $(a + b)^{1+\epsilon} \le 2^\epsilon (a^{1+\epsilon} + b^{1+\epsilon})$, $a, b \ge 0$, to get

$$\mathbb{E} \log^{1+\epsilon}\left(1 \vee \sqrt{X^2 + Y^2}\right) \le \mathbb{E}\left[\left(\frac{1}{2} \log(1 + X^2) + \frac{1}{2} \log(1 + Y^2)\right)^{1+\epsilon}\right]$$

$$\le \frac{1}{2} \mathbb{E} \log^{1+\epsilon}(1 + X^2) + \frac{1}{2} \mathbb{E} \log^{1+\epsilon}(1 + Y^2).$$

Conversely, assume that $\mathbb{E} \log^{1+\epsilon}(1 \vee \sqrt{X^2 + Y^2}) < \infty$. Then we have

$$\begin{aligned}
&\mathbb{E} \log^{1+\epsilon}(1 + X^2) \\
&= \mathbb{E}\left[\mathbb{1}_{\{|X|<1\}} \log^{1+\epsilon}(1 + X^2)\right] + \mathbb{E}\left[\mathbb{1}_{\{|X|\ge 1\}} \log^{1+\epsilon}(1 + X^2)\right] \\
&\le \log^{1+\epsilon} 2 + \mathbb{E} \log^{1+\epsilon}\left[(2X^2) \vee 2\right] \\
&\le \log^{1+\epsilon} 2 + \mathbb{E}\left[\left(\log 2 + \log(1 \vee X^2)\right)^{1+\epsilon}\right] \\
&\le (1 + 2^\epsilon) \log^{1+\epsilon} 2 + \mathbb{E} \log^{1+\epsilon}\left(1 \vee (X^2 + Y^2)\right) \\
&\le (1 + 2^\epsilon) \log^{1+\epsilon} 2 + 2^{1+\epsilon} \mathbb{E} \log^{1+\epsilon}\left(1 \vee \sqrt{X^2 + Y^2}\right),
\end{aligned}$$

and $\mathbb{E} \log^{1+\epsilon}(1 + Y^2) < \infty$ follows similarly. $\square$

3 Generalized distance covariance

Székely et al. [37, 34] introduced distance covariance for two random variables X and Y with values in $\mathbb{R}^m$ and $\mathbb{R}^n$ as

$$\mathcal{V}^2(X, Y; w) := \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} |f_{(X,Y)}(x, y) - f_X(x) f_Y(y)|^2 \, w(x, y) \, dx \, dy,$$

with the weight $w(x, y) = w_{\alpha,m}(x) w_{\alpha,n}(y)$ where $w_{\alpha,m}(x) = c(p, \alpha)|x|^{-m-\alpha}$, $m, n \in \mathbb{N}$, $\alpha \in (0, 2)$. It is well known from the study of infinitely divisible distributions (see also Székely & Rizzo [33]) that $w_{\alpha,m}(x)$ is the density of an m-dimensional α-stable Lévy measure, and the corresponding cndf is just $|x|^{\alpha}$.

We are going to extend distance covariance to products of Lévy measures.

Definition 3.1. Let X and Y be random variables with values in $\mathbb{R}^m$ and $\mathbb{R}^n$ and $\rho := \mu \otimes \nu$ where μ and ν are symmetric Lévy measures on $\mathbb{R}^m \setminus \{0\}$ and $\mathbb{R}^n \setminus \{0\}$, both having full support. The *generalized distance covariance* $V(X, Y)$ is defined as

$$V^2(X, Y) = \iint_{\mathbb{R}^{m+n}} |f_{(X,Y)}(s, t) - f_X(s) f_Y(t)|^2 \, \mu(ds) \, \nu(dt). \qquad (22)$$

By definition, $V(X, Y) = \| f_{(X,Y)} - f_X \otimes f_Y \|_{L^2(\rho)}$ and, in view of the discussion in Section 2.2, we have

$$V(X, Y) = d_\rho \big(\mathcal{L}(U), \mathcal{L}(V) \big), \qquad (23)$$

where $U := (X_1, Y_1)$, $V := (X_2, Y_3)$ and $(X_1, Y_1), (X_2, Y_2), (X_3, Y_3)$ are i.i.d. copies of (X, Y).[3] It is clear that the product measure ρ inherits the properties "symmetry" and "full support" from its marginals μ and ν.

From the discussion following Definition 2.3 we immediately get the next lemma.

Lemma 3.2. *Let $V^2(X, Y)$ be generalized distance covariance of the m- resp. n-dimensional random variables X and Y, cf. (22). The random variables X and Y are independent if, and only if, $V^2(X, Y) = 0$.*

3.1 Generalized distance covariance with finite Lévy measures

Fix the dimensions $m, n \in \mathbb{N}$, set $d := m + n$, and assume that the measure ρ is of the form $\rho = \mu \otimes \nu$ where μ and ν are finite symmetric Lévy measures on $\mathbb{R}^m \setminus \{0\}$ and $\mathbb{R}^n \setminus \{0\}$, respectively. If we integrate the elementary estimates

$$1 \wedge |s|^2 \leq 1 \wedge \big(|s|^2 + |t|^2 \big) \leq \big(1 \wedge |s|^2 \big) + \big(1 \wedge |t|^2 \big),$$
$$1 \wedge |t|^2 \leq 1 \wedge \big(|s|^2 + |t|^2 \big) \leq \big(1 \wedge |s|^2 \big) + \big(1 \wedge |t|^2 \big),$$

with respect to $\rho(ds, dt) = \mu(ds) \, \nu(dt)$, it follows that ρ is a Lévy measure if μ and ν are finite Lévy measures.[4] We also assume that μ and ν, hence ρ, have full support.

Since $V(X, Y)$ is a metric in the sense of Section 2.2 we can use all results from the previous section to derive various representations of generalized distance covariance.

We write Φ, Ψ and Θ for the bounded cndfs induced by $\mu(ds)$, $\nu(dt)$, and $\rho(dr) = \mu(ds) \, \nu(dt)$,

$$\Phi(x) = \int_{\mathbb{R}^m} (1 - \cos x \cdot s) \, \mu(ds), \quad \Psi(y) = \int_{\mathbb{R}^n} (1 - \cos y \cdot t) \, \nu(dt)$$

$$\text{and} \quad \Theta(x, y) = \Theta(u) = \int_{\mathbb{R}^{m+n}} (1 - \cos u \cdot r) \, \rho(dr), \qquad (24)$$

---

[3] This is a convenient way to say that $\mathcal{L}(U) = \mathcal{L}((X, Y))$ while $\mathcal{L}(V) = \mathcal{L}(X) \otimes \mathcal{L}(Y)$.
[4] This argument also shows that the product measure ρ can only be a Lévy measure, if the marginals are finite measures. In this case, ρ is itself a finite measure.

with $u = (x, y) \in \mathbb{R}^{m+n}$ and $r = (s, t) \in \mathbb{R}^{m+n}$. The symmetry in each variable and the elementary identity

$$\frac{1}{2}\big[1 - \cos(x \cdot s - y \cdot t) + 1 - \cos(x \cdot s + y \cdot t)\big]$$
$$= (1 - \cos y \cdot t) + (1 - \cos x \cdot s) - (1 - \cos x \cdot s)(1 - \cos y \cdot t)$$

yield the representation

$$\Theta(x, y) = \Psi(y)\mu\big(\mathbb{R}^m\big) + \Phi(x)\nu\big(\mathbb{R}^n\big) - \Phi(x)\Psi(y). \tag{25}$$

We can now easily apply the results from Section 2.2. In order to do so, we consider six i.i.d. copies (X_i, Y_i), $i = 1, \dots, 6$, of the random vector (X, Y), and set $U := (X_1, Y_1)$, $V := (X_2, Y_3)$, $U' := (X_4, Y_4)$, $V' := (X_5, Y_6)$. This is a convenient way to say that

$$\mathcal{L}(U) = \mathcal{L}(U') = \mathcal{L}((X, Y)) \quad \text{and} \quad \mathcal{L}(V) = \mathcal{L}(V') = \mathcal{L}(X) \otimes \mathcal{L}(Y)$$

and U, U', V and V' are independent.[5]

The following formulae follow directly from Proposition 2.4.d) and Remark 2.5.a).

Proposition 3.3. *Let X and Y be random variables with values in $\mathbb{R}^m$ and $\mathbb{R}^n$ and assume that μ, ν are finite symmetric Lévy measures on $\mathbb{R}^m \setminus \{0\}$ and $\mathbb{R}^n \setminus \{0\}$ with full support. Generalized distance covariance has the following representations*

$$V^2(X, Y) = d_\rho^2\big(\mathcal{L}(U), \mathcal{L}(V)\big) \tag{26}$$
$$= 2\mathbb{E}\Theta\big(U - V'\big) - \mathbb{E}\Theta\big(U - U'\big) - \mathbb{E}\Theta\big(V - V'\big) \tag{27}$$
$$= 2\mathbb{E}\Theta(X_1 - X_5, Y_1 - Y_6) - \mathbb{E}\Theta(X_1 - X_4, Y_1 - Y_4)$$
$$\qquad - \mathbb{E}\Theta(X_2 - X_5, Y_3 - Y_6) \tag{28}$$
$$= \mathbb{E}\Phi(X_1 - X_4)\Psi(Y_1 - Y_4) + \mathbb{E}\Phi(X_2 - X_5)\mathbb{E}\Psi(Y_3 - Y_6)$$
$$\qquad - 2\mathbb{E}\Phi(X_1 - X_5)\Psi(Y_1 - Y_6). \tag{29}$$

The latter equality follows from (25) since the terms depending only on one of the variables cancel as the random variables (X_i, Y_i) are i.i.d. This gives rise to various further representations of $V(X, Y)$.

Corollary 3.4. *Let (X, Y), (X_i, Y_i), Φ, Ψ and μ, ν be as in Proposition 3.3. Generalized distance covariance has the following representations*

$$V^2(X, Y)$$
$$= \mathbb{E}\Phi(X_1 - X_4)\Psi(Y_1 - Y_4) - 2\mathbb{E}\Phi(X_1 - X_2)\Psi(Y_1 - Y_3)$$
$$\qquad + \mathbb{E}\Phi(X_1 - X_2)\mathbb{E}\Psi(Y_3 - Y_4) \tag{30}$$
$$= \mathbb{E}\big[\Phi(X_1 - X_4) \cdot \big\{\Psi(Y_1 - Y_4) - 2\Psi(Y_1 - Y_3) + \Psi(Y_2 - Y_3)\big\}\big] \tag{31}$$

---

[5] In other words: in an expression of the form $f(X_i, X_j, Y_k, Y_l)$ all random variables are independent if, and only if, all indices are different. As soon as two indices coincide, we have (some kind of) dependence.

$$= \mathbb{E}\big[\{\Phi(X_1 - X_4) - \Phi(X_4 - X_2)\} \cdot \{\Psi(Y_4 - Y_1) - \Psi(Y_1 - Y_3)\}\big] \tag{32}$$

$$= \mathbb{E}\big[\{\Phi(X_1 - X_4) - \mathbb{E}\big(\Phi(X_4 - X_1) \mid X_4\big)\}$$
$$\cdot \{\Psi(Y_4 - Y_1) - \mathbb{E}\big(\Psi(Y_1 - Y_4) \mid Y_1\big)\}\big]. \tag{33}$$

Corollary 3.4 shows, in particular, that $V(X, Y)$ can be written as a function of (X_i, Y_i), $i = 1, \ldots, 4$,

$$V^2(X, Y) = \mathbb{E}\big[g\big((X_1, Y_1), \ldots, (X_4, Y_4)\big)\big] \tag{34}$$

for an appropriate function g; for instance, the formula (30) follows for

$$g\big((x_1, y_1), \ldots, (x_4, y_4)\big)$$
$$= \Phi(x_1 - x_4)\Psi(y_1 - y_4) - 2\Phi(x_1 - x_2)\Psi(y_1 - y_3) + \Phi(x_1 - x_2)\Psi(y_3 - y_4).$$

Corollary 3.5. *Let (X, Y), (X_i, Y_i), Φ, Ψ and μ, ν be as in Proposition 3.3 and write*

$$\overline{\overline{\Phi}} = \Phi(X_1 - X_4) - \mathbb{E}\big(\Phi(X_4 - X_1) \mid X_4\big)$$
$$- \mathbb{E}\big(\Phi(X_4 - X_1) \mid X_1\big) + \mathbb{E}\Phi(X_1 - X_4),$$

$$\overline{\overline{\Psi}} = \Psi(Y_1 - Y_4) - \mathbb{E}\big(\Psi(Y_4 - Y_1) \mid Y_4\big)$$
$$- \mathbb{E}\big(\Psi(Y_4 - Y_1) \mid Y_1\big) + \mathbb{E}\Psi(Y_1 - Y_4),$$

for the "doubly centered" versions of $\Phi(X_1 - X_4)$ and $\Psi(Y_1 - Y_4)$. Generalized distance covariance has the following representation

$$V^2(X, Y) = \mathbb{E}[\,\overline{\overline{\Phi}} \cdot \overline{\overline{\Psi}}\,]. \tag{35}$$

Proof. Denote by $\overline{\Phi} = \Phi(X_1 - X_4) - \mathbb{E}(\Phi(X_4 - X_1) \mid X_4)$ and $\overline{\Psi} = \Psi(Y_1 - Y_4) - \mathbb{E}(\Psi(Y_4 - Y_1) \mid Y_1)$ the centered random variables appearing in (33). Clearly,

$$\overline{\overline{\Phi}} = \overline{\Phi} - \mathbb{E}(\overline{\Phi} \mid X_1) \quad \text{and} \quad \overline{\overline{\Psi}} = \overline{\Psi} - \mathbb{E}(\overline{\Psi} \mid Y_4).$$

Thus,

$$\mathbb{E}[\overline{\overline{\Phi}} \cdot \overline{\overline{\Psi}}] = \mathbb{E}\big[\big(\overline{\Phi} - \mathbb{E}(\overline{\Phi} \mid X_1)\big) \cdot \big(\overline{\Psi} - \mathbb{E}(\overline{\Psi} \mid Y_4)\big)\big]$$
$$= \mathbb{E}[\overline{\Phi} \cdot \overline{\Psi}] + \mathbb{E}\big[\mathbb{E}(\overline{\Phi} \mid X_1) \cdot \mathbb{E}(\overline{\Psi} \mid Y_4)\big]$$
$$- \mathbb{E}\big[\mathbb{E}(\overline{\Phi} \mid X_1) \cdot \overline{\Psi}\big] - \mathbb{E}\big[\overline{\Phi} \cdot \mathbb{E}(\overline{\Psi} \mid Y_4)\big].$$

Since X_1 and Y_4 are independent, we have

$$\mathbb{E}\big[\mathbb{E}(\overline{\Phi} \mid X_1) \cdot \mathbb{E}(\overline{\Psi} \mid Y_4)\big] = \mathbb{E}\big[\mathbb{E}(\overline{\Phi} \mid X_1)\big] \cdot \mathbb{E}\big[\mathbb{E}(\overline{\Psi} \mid Y_4)\big] = \mathbb{E}[\overline{\Phi}] \cdot \mathbb{E}[\overline{\Psi}] = 0.$$

Using the tower property and the independence of (X_1, Y_1) and Y_4 we get

$$\mathbb{E}\big[\mathbb{E}(\overline{\Phi} \mid X_1) \cdot \overline{\Psi}\big] = \mathbb{E}\big[\mathbb{E}\big[\mathbb{E}(\overline{\Phi} \mid X_1) \mid Y_1, Y_4\big] \cdot \overline{\Psi}\big]$$
$$= \mathbb{E}\big[\mathbb{E}\big[\mathbb{E}(\overline{\Phi} \mid X_1) \mid Y_1\big] \cdot \overline{\Psi}\big] = 0,$$

where we use, for the last equality, that $\overline{\Psi}$ is orthogonal to the L^2-space of Y_1-measurable functions. In a similar fashion we see $\mathbb{E}[\overline{\Phi} \cdot \mathbb{E}(\overline{\Psi} \mid Y_4)] = 0$, and the assertion follows because of (33). $\qquad\square$

In Section 4 we will encounter further representations of the generalized distance covariance if X and Y have discrete distributions with finitely many values, as it is the case for empirical distributions.

3.2 Generalized distance covariance with arbitrary Lévy measures

So far, we have been considering finite Lévy measures μ, ν and bounded cndfs Φ and Ψ (24). We will now extend our considerations to products of unbounded Lévy measures. The measure $\rho := \mu \otimes \nu$ satisfies the integrability condition

$$\iint_{\mathbb{R}^{m+n}} (1 \wedge |x|^2)(1 \wedge |y|^2) \rho(dx, dy) = \int_{\mathbb{R}^m} (1 \wedge |x|^2) \mu(dx) \int_{\mathbb{R}^n} (1 \wedge |y|^2) \nu(dy) < \infty. \quad (36)$$

Other than in the case of finite marginals, ρ is no longer a Lévy measure, see the footnote on page 364. Thus, the function Θ defined in (24) need not be a cndf and we cannot directly apply Proposition 2.4; instead we need the following result ensuring the finiteness of $V(X, Y)$.

Lemma 3.6. *Let X, X' be i.i.d. random variables on $\mathbb{R}^m$ and Y, Y' be i.i.d. random variables on $\mathbb{R}^n$; μ and ν are symmetric Lévy measures on $\mathbb{R}^m$ and $\mathbb{R}^n$ with full support and with corresponding cndfs Φ and Ψ as in (24). Then*

$$V^2(X, Y) \leq \mathbb{E}\Phi(X - X') \cdot \mathbb{E}\Psi(Y - Y') \leq 16 \, \mathbb{E}\Phi(X) \cdot \mathbb{E}\Psi(Y). \quad (37)$$

Proof. Following Székely et al. [37, p. 2772] we get

$$
\begin{aligned}
|f_{X,Y}(s, t) - f_X(s) f_Y(t)|^2 &= \left| \mathbb{E}\left((e^{iX \cdot s} - f_X(s))(e^{iY \cdot t} - f_Y(t)) \right) \right|^2 \\
&\leq \left[\mathbb{E}\left(|e^{iX \cdot s} - f_X(s)||e^{iY \cdot t} - f_Y(t)| \right) \right]^2 \\
&\leq \mathbb{E}\left[|e^{iX \cdot s} - f_X(s)|^2 \right] \cdot \mathbb{E}\left[|e^{iY \cdot t} - f_Y(t)|^2 \right] \quad (38)
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{E}\left[|e^{iX \cdot s} - f_X(s)|^2 \right] &= \mathbb{E}\left[(e^{isX} - f_X(s))(e^{-isX} - \overline{f_X(s)}) \right] \\
&= 1 - |f_X(s)|^2 = 1 - f_X(s)\overline{f_X(s)}. \quad (39)
\end{aligned}
$$

Using (16) for $\rho = \mu, \Theta = \Phi, U = X$ and (13) for $\Theta = \Phi, U = X, V = X'$ shows

$$\int_{\mathbb{R}^m} \mathbb{E}\left[|e^{iX \cdot s} - f_X(s)|^2 \right] \mu(ds) = \mathbb{E}\Phi(X - X') \leq 4\mathbb{E}\Phi(X),$$

and an analogous argument for ν and Y yields the bound (37). $\qquad \square$

Looking at the various representations (29)–(33) of $V(X, Y)$ it is clear that these make sense as soon as all expectations in these expressions are finite, i.e. some moment condition in terms of Φ and Ψ should be enough to ensure the finiteness of $V(X, Y)$ and all terms appearing in the respective representations.

In order to use the results of the previous section we fix $\epsilon > 0$ and consider the *finite* symmetric Lévy measures

$$\mu_\epsilon(ds) := \frac{|s|^2}{\epsilon^2 + |s|^2} \mu(ds) \quad \text{and} \quad \nu_\epsilon(dt) := \frac{|t|^2}{\epsilon^2 + |t|^2} \nu(dt), \quad (40)$$

and the corresponding cndfs Φ_ϵ and Ψ_ϵ given by (24); the product measure $\rho_\epsilon :=$ $\mu_\epsilon \otimes \nu_\epsilon$ is a finite Lévy measure and the corresponding cndf Θ_ϵ is also bounded (it can be expressed by Φ_ϵ and Ψ_ϵ through the formula (25)).

This allows us to derive the representations (29)–(33) for each $\epsilon > 0$ and with Φ_ϵ and Ψ_ϵ. Since we have $\Phi = \sup_\epsilon \Phi_\epsilon$ and $\Psi = \sup_\epsilon \Psi_\epsilon$, we can use monotone convergence to get the representations for the cndfs Φ and Ψ with Lévy measures μ and ν, respectively. Of course, this requires the existence of certain (mixed) Φ-Ψ moments of the random variables (X, Y).

Theorem 3.7. *Let μ and ν be symmetric Lévy measures on $\mathbb{R}^m \setminus \{0\}$ and $\mathbb{R}^n \setminus \{0\}$ with full support and corresponding cndfs Φ and Ψ given by (24). For any random vector (X, Y) with values in $\mathbb{R}^{m+n}$ satisfying the following moment condition*

$$\mathbb{E}\Phi(X) + \mathbb{E}\Psi(Y) < \infty, \tag{41}$$

the generalized distance correlation $V(X, Y)$ is finite. If additionally

$$\mathbb{E}\big(\Phi(X)\Psi(Y)\big) < \infty, \tag{42}$$

then also the representations (29)–(33) *hold with all terms finite.*[6]

Proof. We only have to check the finiteness. Using Lemma 3.6, we see that (41) guarantees that $V(X, Y) < \infty$. The finiteness of (all the terms appearing in) the representations (29)–(33) follows from the monotone convergence argument, since the moment condition (42) ensures the finiteness of the limiting expectations. $\square$

Remark 3.8. a) Using the Hölder inequality the following condition implies (41) and (42):

$$\mathbb{E}\Phi^p(X) + \mathbb{E}\Psi^q(Y) < \infty \quad \text{for some } p, q > 1 \text{ with } \frac{1}{p} + \frac{1}{q} = 1. \tag{43}$$

If one of Ψ or Φ is bounded then (41) implies (42), and if both are bounded then the expectations are trivially finite.

Since continuous negative definite functions grow at most quadratically, we see that (41) and (42) also follow if

$$\mathbb{E}|X|^{2p} + \mathbb{E}|Y|^{2q} < \infty \quad \text{for some } p, q > 1 \text{ with } \frac{1}{p} + \frac{1}{q} = 1. \tag{44}$$

b) A slightly different set-up was employed by Lyons [23]: If the cndfs Φ and Ψ are subadditive, then the expectation in (32) is finite. This is a serious restriction on the class of cndfs since subadditivity means that Φ and Ψ can grow at most linearly at infinity, whereas general cndfs grow at most quadratically, cf. (9). Note, however, that square roots of real cndfs are always subadditive, cf. (7).

---

[6]As before, we denote in these formulae by (X_i, Y_i), $i = 1, \ldots, 6$, i.i.d. copies of (X, Y).

4 Estimating generalized distance covariance

Let $(x_i, y_i)_{i=1,\ldots,N}$ be a sample of (X, Y) and denote by $(\widehat{X}^{(N)}, \widehat{Y}^{(N)})$ the random variable which has the corresponding empirical distribution, i.e. the uniform distribution on $(x_i, y_i)_{i=1,\ldots,N}$; to be precise, repeated points will have the corresponding multiple weight. By definition, $(\widehat{X}^{(N)}, \widehat{Y}^{(N)})$ is a bounded random variable and for i.i.d. copies with $\mathcal{L}((\widehat{X}_i^{(N)}, \widehat{Y}_i^{(N)})) = \mathcal{L}((\widehat{X}^{(N)}, \widehat{Y}^{(N)}))$ for $i = 1, \ldots, 4$, we have using (34)

$$\mathbb{E}\big[g\big((\widehat{X}_1^{(N)}, \widehat{Y}_1^{(N)}), \ldots, (\widehat{X}_4^{(N)}, \widehat{Y}_4^{(N)})\big)\big]$$
$$= \int_{\mathbb{R}^n \setminus \{0\}} \int_{\mathbb{R}^m \setminus \{0\}} \big| f_{(\widehat{X}^{(N)}, \widehat{Y}^{(N)})}(s, t) - f_{\widehat{X}^{(N)}}(s) f_{\widehat{Y}^{(N)}}(t) \big|^2 \, \mu(\mathrm{d}s)\, \nu(\mathrm{d}t). \qquad (45)$$

The formulae (27)–(33) hold, in particular, for the empirical random variables $(\widehat{X}^{(N)}, \widehat{Y}^{(N)})$, and so we get

$$\mathbb{E}\big[g\big((\widehat{X}_1^{(N)}, \widehat{Y}_1^{(N)}), \ldots, (\widehat{X}_4^{(N)}, \widehat{Y}_4^{(N)})\big)\big]$$
$$= \frac{1}{N^4} \sum_{i,j,k,l=1}^{N} g\big((x_i, y_i), (x_j, y_j), (x_k, y_k), (x_l, y_l)\big) \qquad (46)$$
$$= \frac{1}{N^4} \sum_{i,j,k,l=1}^{N} \big[\Phi(x_i - x_k)\Psi(y_i - y_k) - 2\Phi(x_i - x_j)\Psi(y_i - y_l)$$
$$+ \Phi(x_i - x_j)\Psi(y_k - y_l)\big] \qquad (47)$$
$$= \frac{1}{N^2} \sum_{i,k=1}^{N} \Phi(x_i - x_k)\Psi(y_i - y_k) - \frac{2}{N^3} \sum_{i,j,l=1}^{N} \Phi(x_i - x_j)\Psi(y_i - y_l)$$
$$+ \frac{1}{N^2} \sum_{i,j=1}^{N} \Phi(x_i - x_j)\frac{1}{N^2} \sum_{k,l=1}^{N} \Psi(y_k - y_l) \qquad (48)$$
$$= \left(\frac{1}{N^2} - \frac{2}{N^3} + \frac{2}{N^4}\right) \sum_{\substack{i,k=1 \\ \text{distinct}}}^{N} \Phi(x_i - x_k)\Psi(y_i - y_k)$$
$$+ \left(\frac{4}{N^4} - \frac{2}{N^3}\right) \sum_{\substack{i,j,l=1 \\ \text{distinct}}}^{N} \Phi(x_i - x_j)\Psi(y_i - y_l)$$
$$+ \frac{1}{N^4} \sum_{\substack{i,j,k,l=1 \\ \text{distinct}}}^{N} \Phi(x_i - x_j)\Psi(y_k - y_l). \qquad (49)$$

The sum in (46) is – for functions g which are symmetric under permutations of their variables – a V-statistic.

Definition 4.1. The estimator $^NV^2 := {}^NV^2((x_1, y_1), \ldots, (x_N, y_N))$ of $V^2(X, Y)$ is defined by (47).

In abuse of notation, we also write ${}^N V^2((x_1, \ldots, x_N), (y_1, \ldots, y_N))$ and ${}^N V^2(\boldsymbol{x}, \boldsymbol{y})$ with $\boldsymbol{x} = (x_1, \ldots, x_N)$, $\boldsymbol{y} = (y_1, \ldots, y_N)$ instead of the precise ${}^N V^2((x_1, y_1), \ldots, (x_N, y_N))$.

Since all random variables are bounded, we could use any of the representations (46)–(49) to define an estimator. A computationally feasible representation is given in the following lemma.

Lemma 4.2. *The estimator ${}^N V^2$ of $V^2(X, Y)$ has the following representation using matrix notation:*

$$
{}^N V^2\big((x_1, y_1), \ldots, (x_N, y_N)\big) = \frac{1}{N^2} \operatorname{trace}\big(B^\top A\big) = \frac{1}{N^2} \sum_{k,l=1}^{N} A_{kl} B_{kl}, \tag{50}
$$

where

$$
\begin{aligned}
A &= C^\top a C, & a &= \big(\Phi(x_k - x_l)\big)_{k,l=1,\ldots,N}, \\
B &= C^\top b C, & b &= \big(\Psi(y_k - y_l)\big)_{k,l=1,\ldots,N},
\end{aligned} \tag{51}
$$

and $C = I - \frac{1}{N}\mathbb{1}$ with $\mathbb{1} = (1)_{k,l=1,\ldots,N}$ and $I = (\delta_{jk})_{j,k=1,\ldots,N}$.

Remark 4.3. If $\Phi(x) = |x|$ and $\Psi(y) = |y|$ the matrices a and b in (51) are Euclidean distance matrices. For general cndfs Φ and Ψ, the matrices $-a$ and $-b$ are conditionally positive definite (see Theorem 2.1), and A, B are positive definite. This gives a simple explanation as to why the right-hand side of (50) is positive.

Proof of Lemma 4.2. By definition,

$$
\operatorname{trace}\big(b^\top a\big) = \sum_{i,j=1}^{N} a_{ij} b_{ij} = \sum_{i,j=1}^{N} \Phi(x_i - x_j)\Psi(y_i - y_j), \tag{52}
$$

$$
\operatorname{trace}\big(b^\top \mathbb{1} a\big) = \sum_{i,j,k=1}^{N} a_{ij} b_{ik} = \sum_{i,j,k=1}^{N} \Phi(x_i - x_j)\Psi(y_i - y_k), \tag{53}
$$

$$
\operatorname{trace}\big(\mathbb{1} b^\top \mathbb{1} a\big) = \sum_{i,j,k,l=1}^{N} a_{ij} b_{kl} = \sum_{i,j=1}^{N} \Phi(x_i - x_j) \sum_{k,l=1}^{N} \Psi(y_k - y_l), \tag{54}
$$

and this allows us to rewrite (48) as

$$
{}^N V^2 = \frac{1}{N^2} \operatorname{trace}\big(b^\top a\big) - \frac{2}{N^3} \operatorname{trace}\big(b^\top \mathbb{1} a\big) + \frac{1}{N^4} \operatorname{trace}\big(\mathbb{1} b^\top \mathbb{1} a\big). \tag{55}
$$

Observe that $C = C^\top$ and $CC^\top = C$. Using this and the fact that the trace is invariant under cyclic permutations, we get

$$
\begin{aligned}
\operatorname{trace}\big(B^\top A\big) &= \operatorname{trace}\big(C^\top b^\top C C^\top a C\big) \\
&= \operatorname{trace}\big(C C^\top b^\top C C^\top a\big) = \operatorname{trace}\big(C b^\top C a\big).
\end{aligned}
$$

Plugging in the definition of C now gives

$$\text{trace}(B^\top A) = \text{trace}((b^\top - \tfrac{1}{N}\mathbb{1}b^\top)(a - \tfrac{1}{N}\mathbb{1}a))$$
$$= \text{trace}(b^\top a) - \tfrac{1}{N}\text{trace}(\mathbb{1}b^\top a) - \tfrac{1}{N}\text{trace}(b^\top \mathbb{1}a) + \tfrac{1}{N^2}\text{trace}(\mathbb{1}b^\top \mathbb{1}a)$$
$$= \text{trace}(b^\top a) - \tfrac{2}{N}\text{trace}(b^\top \mathbb{1}a) + \tfrac{1}{N^2}\text{trace}(\mathbb{1}b^\top \mathbb{1}a).$$

For the last equality we use that a and b are symmetric matrices. □

We will now show that $^N V^2$ is a consistent estimator for $V^2(X, Y)$.

Theorem 4.4 (Consistency). *Let $(X_i, Y_i)_{i=1,\dots,N}$ be i.i.d. copies of (X, Y) and write $X := (X_1, \dots, X_N)$, $Y := (Y_1, \dots, Y_N)$. If $\mathbb{E}[\Phi(X) + \Psi(Y)] < \infty$ holds, then*

$$^N V^2(X, Y) \xrightarrow[N\to\infty]{} V^2(X, Y) \quad a.s. \tag{56}$$

Proof. The moment condition $\mathbb{E}[\Phi(X) + \Psi(Y)] < \infty$ ensures that the generalized distance covariance $V^2(X, Y)$ is finite, cf. Lemma 3.6. Define μ_ϵ and ν_ϵ as in (40), and write $V_\epsilon^2(X, Y)$ for the corresponding generalized distance covariance and $^N V_\epsilon^2(X, Y)$ for its estimator. By the triangle inequality we obtain

$$|V^2(X, Y) - {}^N V^2(X, Y)|$$
$$\leq |V^2(X, Y) - V_\epsilon^2(X, Y)| + |V_\epsilon^2(X, Y) - {}^N V_\epsilon^2(X, Y)|$$
$$+ |^N V_\epsilon^2(X, Y) - {}^N V^2(X, Y)|.$$

We consider the three terms on the right-hand side separately. The first term vanishes as $\epsilon \to 0$, since

$$\lim_{\epsilon\to 0} V_\epsilon^2(X, Y) = V^2(X, Y)$$

by monotone convergence. For each $\epsilon > 0$, the second term converges to zero as $N \to \infty$, since

$$\lim_{N\to\infty} {}^N V_\epsilon^2(X, Y) = V_\epsilon^2(X, Y) \quad \text{a.s.}$$

by the strong law of large numbers (SLLN) for V-statistics; note that this is applicable since the functions Φ_ϵ and Ψ_ϵ are bounded (because of the finiteness of the Lévy measures μ_ϵ and ν_ϵ).

For the third term we set $\mu^\epsilon = \mu - \mu_\epsilon$, $\nu^\epsilon = \nu - \nu_\epsilon$ and write $\Phi^\epsilon, \Psi^\epsilon$ for the corresponding continuous negative definite functions. Lemma 3.6 yields the inequality

$$|^N V_\epsilon^2((x_1, y_1), \dots, (x_N, y_N)) - {}^N V^2((x_1, y_1), \dots, (x_N, y_N))|$$
$$= \iint \left| f_{\widehat{X}^{(N)}, \widehat{Y}^{(N)}}(s, t) - f_{\widehat{X}^{(N)}}(s) f_{\widehat{Y}^{(N)}}(t) \right|^2 \mu^\epsilon(ds)\, \nu^\epsilon(dt)$$
$$\leq 16\, \mathbb{E}\Phi^\epsilon(\widehat{X}^{(N)}) \cdot \mathbb{E}\Psi^\epsilon(\widehat{Y}^{(N)}) = 16 \sum_{i=1}^{N} \frac{1}{N}\Phi^\epsilon(x_i) \cdot \sum_{i=1}^{N} \frac{1}{N}\Psi^\epsilon(y_i).$$

From the representation (24) we know that $\Phi^\epsilon(x) \leq \Phi(x)$, hence also $\mathbb{E}\Phi^\epsilon(X) \leq \mathbb{E}\Phi(X)$ and this is finite by assumption. Therefore, we can use monotone convergence to conclude that $\lim_{\epsilon\to 0} \mathbb{E}\Phi^\epsilon(X) = 0$. Thus, the classical SLLN applies and proves

$$\lim_{\epsilon\to 0} \limsup_{N\to\infty} |^N V_\epsilon^2(X, Y) - {}^N V^2(X, Y)| \leq \lim_{\epsilon\to 0} \mathbb{E}\Phi^\epsilon(X) \cdot \mathbb{E}\Psi^\epsilon(Y) = 0 \quad \text{a.s.} \quad □$$

Next we study the behaviour of the estimator under the hypothesis of independence.

Theorem 4.5. *If X and Y are independent and satisfy the moment conditions*

$$\mathbb{E}\big[\Phi(X)+\Psi(Y)\big] < \infty \quad and \quad \mathbb{E}\big[\log^{1+\epsilon}\big(1+|X|^2\big)+\log^{1+\epsilon}\big(1+|Y|^2\big)\big] < \infty \tag{57}$$

for some $\epsilon > 0$, then

$$N \cdot {}^N V^2(X, Y) \xrightarrow[N\to\infty]{d} \iint |\mathbb{G}(s,t)|^2 \, \mu(ds)\, \nu(dt) = \|\mathbb{G}\|^2_{\mu\otimes\nu} \tag{58}$$

in distribution, where $(\mathbb{G}(s,t))_{(s,t)\in\mathbb{R}^{m+n}}$ is a complex-valued Gaussian random field with $\mathbb{E}(\mathbb{G}(s,t)) = 0$ and

$$\mathrm{Cov}\big(\mathbb{G}(s,t), \mathbb{G}(s',t')\big)$$
$$= \big(f_X(s-s') - f_X(s)\overline{f_X(s')}\big) \cdot \big(f_Y(t-t') - f_Y(t)\overline{f_Y(t')}\big). \tag{59}$$

Proof. Let X, Y and $\mathbb{G}$ be as above and let X', X_i and Y', Y_i ($i \in \mathbb{N}$) be independent random variables with laws $\mathcal{L}(X)$ and $\mathcal{L}(Y)$, respectively. Define for $N \in \mathbb{N}$, $s \in \mathbb{R}^m$, $t \in \mathbb{R}^n$

$$Z_N(s,t) := \frac{1}{N}\sum_{k=1}^{N} e^{is\cdot X_k + it\cdot Y_k} - \frac{1}{N^2}\sum_{k,l=1}^{N} e^{is\cdot X_k + it\cdot Y_l}. \tag{60}$$

Then

$$N \cdot {}^N V^2(X, Y) = \|\sqrt{N}Z_N\|^2_{\mu\otimes\nu}. \tag{61}$$

The essential idea is to show that (on an appropriate space) $\sqrt{N}Z_N \xrightarrow{d} \mathbb{G}$ and then to apply the continuous mapping theorem.

Since X and Y are independent, we have

$$\mathbb{E}\big(Z_N(s,t)\big) = 0, \tag{62}$$
$$\mathbb{E}\big(Z_N(s,t)\overline{Z_N(s',t')}\big)$$
$$= \tfrac{N-1}{N^2}\big(f_X(s-s') - f_X(s)\overline{f_X(s')}\big)\big(f_Y(t-t') - f_Y(t)\overline{f_Y(t')}\big), \tag{63}$$
$$\mathbb{E}|\sqrt{N}Z_N(s,t)|^2 = \tfrac{N-1}{N}\big(1 - |f_X(s)|^2\big)\big(1 - |f_Y(t)|^2\big). \tag{64}$$

The first identity (62) is obvious and (64) follows from (63) if we set $s = s'$ and $t = t'$. The proof of (63) is deferred to Lemma 4.6 following this proof.

The convergence $\sqrt{N}Z_N \xrightarrow{d} \mathbb{G}$ in $\mathcal{C}_T := (C(K_T), \|.\|_{K_T})$ with $K_T = \{x \in \mathbb{R}^{m+n} : |x| \leq T\}$, i.e. in the space of continuous functions on K_T equipped with the supremum norm, holds if $\mathbb{E}\log^{1+\epsilon}(\sqrt{|X|^2 + |Y|^2} \vee 1) < \infty$, see Csörgő [14, Thm. on p. 294] or Ushakov [38, Sec. 3.7]. This log-moment condition is equivalent to the log-moment condition (57), see Lemma 2.7.

In fact, the result in [14] is cast in a more general setting, proving the convergence for vectors $(X_1, \ldots, X_N)$, but only one-dimensional marginals are considered. The proof for multidimensional marginals is very similar, so we will only give an outline:

Let F_Z denote the distribution function of the random variable Z and $^N F_z$ the empirical distribution of a sample $z_1, \ldots, z_N$; if the sample is replaced by N independent copies of Z, we write $^N F_Z$. Using this notation and the independence of X and Y yields the representation

$$\sqrt{N} Z_N(s, t) = \sqrt{N} \left(\frac{1}{N} \sum_{k=1}^{N} e^{is \cdot X_k + it \cdot Y_k} - f_{(X,Y)}(s, t) \right)$$

$$- \sqrt{N} \left(\frac{1}{N} \sum_{k=1}^{N} e^{is \cdot X_k} - f_X(s) \right) \left(\frac{1}{N} \sum_{l=1}^{N} e^{it \cdot Y_l} \right)$$

$$- \sqrt{N} \left(\frac{1}{N} \sum_{l=1}^{N} e^{it \cdot Y_l} - f_Y(t) \right) f_X(s)$$

$$= \int e^{i(s \cdot x + t \cdot y)} \, d \left(\sqrt{N} \left({}^N F_{(X,Y)}(x, y) - F_{(X,Y)}(x, y) \right) \right)$$

$$- \int e^{is \cdot x} \, d \left(\sqrt{N} \left({}^N F_X(x) - F_X(x) \right) \right) \cdot \left(\frac{1}{N} \sum_{l=1}^{N} e^{it \cdot Y_l} \right)$$

$$- \int e^{it \cdot y} \, d \left(\sqrt{N} \left({}^N F_Y(y) - F_Y(y) \right) \right) \cdot f_X(s). \tag{65}$$

Note that for bivariate distributions the integrals with respect to a single variable reduce to integrals with respect to the corresponding marginal distribution, e.g. $\int_{\mathbb{R}^{m+n}} h(x) \, d F_{(X,Y)}(x, y) = \int_{\mathbb{R}^m} h(x) \, d F_X(x)$. Therefore, a straightforward calculation shows that $\sqrt{N} Z_N(s, t)$ equals

$$\int g(x, y) \, d \left(\sqrt{N} \left({}^N F_{(X,Y)}(x, y) - F_{(X,Y)}(x, y) \right) \right) \tag{66}$$

with the integrand

$$g(x, y) := e^{i(s \cdot x + t \cdot y)} - e^{is \cdot x} \left[\frac{1}{N} \sum_{l=1}^{N} e^{it \cdot Y_l} \right] - f_X(t) e^{it \cdot y}.$$

Following Csörgő [13] we obtain for $N \to \infty$ the limit

$$\int_{\mathbb{R}^{m+n}} \left(e^{is \cdot x + it \cdot y} - e^{is \cdot x} f_Y(t) - f_X(s) e^{it \cdot y} \right) dB(x, y), \tag{67}$$

where B is a Brownian bridge; as in [13, Eq. (3.2)] one can show that it is a Gaussian process indexed by $\mathbb{R}^{m+n}$ satisfying

$$\mathbb{E}(B(x, y)) = 0, \tag{68}$$

$$\mathbb{E}(B(x, y) B(x', y')) = \mathbb{P}(X \le x \wedge x', Y \le y \wedge y')$$
$$- \mathbb{P}(X \le x, Y \le y) \mathbb{P}(X \le x', Y \le y'), \tag{69}$$

$$\lim_{x \to -\infty} B(x, y) = \lim_{y \to -\infty} B(x, y) = \lim_{(x,y) \to (\infty, \infty)} B(x, y) = 0. \tag{70}$$

The limit (67) is continuous if, and only if, a rather complicated tail condition is satisfied [13, Thm. 3.1]; Csörgő [14, p. 294] shows that this condition is implied by the simpler moment condition (57), cf. Lemma 2.7. Thus, $\sqrt{N} Z_N \xrightarrow{d} \mathbb{G}$ in $\mathcal{C}_T :=$ $(C(K_T), \|.\|_{K_T})$.

Pick $\epsilon > 0$, set $\delta := 1/\epsilon$ and define

$$\mu_{\epsilon,\delta}(A) := \mu\big(A \cap \{\epsilon \le |s| < \delta\}\big) \quad \text{and} \quad \mu^{\epsilon,\delta} := \mu - \mu_{\epsilon,\delta}; \tag{71}$$

the measures $\nu_{\epsilon,\delta}$ and $\nu^{\epsilon,\delta}$ are defined analogously. Note that

$$\begin{aligned} \left| \|h\|_{\mu_{\epsilon,\delta} \otimes \nu_{\epsilon,\delta}} - \|h'\|_{\mu_{\epsilon,\delta} \otimes \nu_{\epsilon,\delta}} \right|^2 &\le \|h - h'\|_{\mu_{\epsilon,\delta} \otimes \nu_{\epsilon,\delta}}^2 \\ &= \int |h - h'|^2 \, d\mu_{\epsilon,\delta} \otimes \nu_{\epsilon,\delta} \\ &\le \|h - h'\|_{K_T}^2 \cdot \mu_{\epsilon,\delta}\big(\mathbb{R}^n\big) \cdot \nu_{\epsilon,\delta}\big(\mathbb{R}^m\big) \end{aligned} \tag{72}$$

shows that $h \mapsto \|h\|_{\mu_{\epsilon,\delta} \otimes \nu_{\epsilon,\delta}}^2$ is continuous on $\mathcal{C}_T$. Thus, the continuous mapping theorem implies

$$\|\sqrt{N} Z_N\|_{\mu_{\epsilon,\delta} \otimes \nu_{\epsilon,\delta}}^2 \xrightarrow[N \to \infty]{d} \|\mathbb{G}\|_{\mu_{\epsilon,\delta} \otimes \nu_{\epsilon,\delta}}^2. \tag{73}$$

By the triangle inequality we have

$$\begin{aligned} \left| N \cdot {}^N V^2(X, Y) - \|\mathbb{G}\|_{\mu \otimes \nu}^2 \right| &\le \left| N \cdot {}^N V^2(X, Y) - \|\sqrt{N} Z_N\|_{\mu_{\epsilon,\delta} \otimes \nu_{\epsilon,\delta}}^2 \right| \\ &\quad + \left| \|\sqrt{N} Z_N\|_{\mu_{\epsilon,\delta} \otimes \nu_{\epsilon,\delta}}^2 - \|\mathbb{G}\|_{\mu_{\epsilon,\delta} \otimes \nu_{\epsilon,\delta}}^2 \right| \\ &\quad + \left| \|\mathbb{G}\|_{\mu_{\epsilon,\delta} \otimes \nu_{\epsilon,\delta}}^2 - \|\mathbb{G}\|_{\mu \otimes \nu}^2 \right|. \end{aligned} \tag{74}$$

Thus, it remains to show that the first and last terms on the right-hand side vanish uniformly as $\epsilon \to 0$. Note that

$$\begin{aligned} &\|\mathbb{G}\|_{\mu \otimes \nu}^2 - \|\mathbb{G}\|_{\mu_{\epsilon,\delta} \otimes \nu_{\epsilon,\delta}}^2 \\ &= \|\mathbb{G}\|_{\mu^{\epsilon,\delta} \otimes \nu^{\epsilon,\delta}}^2 + \|\mathbb{G}\|_{\mu_{\epsilon,\delta} \otimes \nu^{\epsilon,\delta}}^2 + \|\mathbb{G}\|_{\mu^{\epsilon,\delta} \otimes \nu_{\epsilon,\delta}}^2 \xrightarrow[\epsilon \to 0]{} 0 \text{ a.s.} \end{aligned} \tag{75}$$

This follows from the dominated convergence theorem, since

$$\begin{aligned} \mathbb{E}\big(\|\mathbb{G}\|_{\mu \otimes \nu}^2\big) &= \big\|1 - |f_X|^2\big\|_\mu^2 \cdot \big\|1 - |f_Y|^2\big\|_\nu^2 \\ &= \mathbb{E}\Phi\big(X - X'\big) \cdot \mathbb{E}\Psi\big(Y - Y'\big) < \infty. \end{aligned} \tag{76}$$

Moreover,

$$\begin{aligned} &\mathbb{E}\big(\big| N \cdot {}^N V^2(X, Y) - \|\sqrt{N} Z_N\|_{\mu_{\epsilon,\delta} \otimes \nu_{\epsilon,\delta}}^2 \big|\big) \\ &= \mathbb{E}\|\sqrt{N} Z_N\|_{\mu^{\epsilon,\delta} \otimes \nu^{\epsilon,\delta}}^2 + \mathbb{E}\|\sqrt{N} Z_N\|_{\mu_{\epsilon,\delta} \otimes \nu^{\epsilon,\delta}}^2 + \mathbb{E}\|\sqrt{N} Z_N\|_{\mu^{\epsilon,\delta} \otimes \nu_{\epsilon,\delta}}^2 \end{aligned} \tag{77}$$

and for the first term we have

$$\big\|\mathbb{E}\big(|\sqrt{N} Z_N|^2\big)\big\|_{\mu^{\epsilon,\delta} \otimes \nu^{\epsilon,\delta}}^2 = \big(\tfrac{N-1}{N}\big)^2 \cdot \big\|1 - |f_X|^2\big\|_{\mu^{\epsilon,\delta}}^2 \cdot \big\|1 - |f_Y|^2\big\|_{\nu^{\epsilon,\delta}}^2 \xrightarrow[\epsilon \to 0]{} 0 \tag{78}$$

by dominated convergence, since we have $\|1 - |f_X|^2\|_{\mu^{\epsilon,\delta}}^2 \le \mathbb{E}\Phi(X - X') < \infty$ and $\|1 - |f_Y|^2\|_{\nu^{\epsilon,\delta}}^2 \le \mathbb{E}\Psi(Y - Y') < \infty$. The other summands are dealt with similarly.

The result follows since the convergence in (75) and (78) is uniform in N. $\qquad\square$

We still have to prove (63).

Lemma 4.6. *In the setting of (the proof of) Theorem 4.5 we have*

$$\mathbb{E}\big(Z_N(s,t)\overline{Z_N(s',t')}\big)$$
$$= \tfrac{N-1}{N^2}\big(f_X(s-s') - f_X(s)\overline{f_X(s')}\big)\big(f_Y(t-t') - f_Y(t)\overline{f_Y(t')}\big).$$

Proof. Observe that

$$Z_N(s,t) = \frac{1}{N}\sum_{k=1}^{N} e^{\mathrm{i}s\cdot X_k + \mathrm{i}t\cdot Y_k} - \frac{1}{N^2}\sum_{k,l=1}^{N} e^{\mathrm{i}s\cdot X_k + \mathrm{i}t\cdot Y_l}$$

$$= \frac{1}{N}\sum_{k=1}^{N}\left(e^{\mathrm{i}s\cdot X_k} - \tfrac{1}{N}\sum_{l=1}^{N} e^{\mathrm{i}s\cdot X_l}\right)\left(e^{\mathrm{i}t\cdot Y_k} - \tfrac{1}{N}\sum_{l=1}^{N} e^{\mathrm{i}t\cdot Y_l}\right).$$

Using this formula and the independence of the random variables $(X_1,\ldots,X_N)$ and $(Y_1,\ldots,Y_N)$ yields

$$\mathbb{E}\big(Z_N(s,t)\overline{Z_N(s',t')}\big)$$
$$= \frac{1}{N^2}\sum_{j,k=1}^{N}\mathbb{E}\Bigg[\left(e^{\mathrm{i}s\cdot X_k} - \tfrac{1}{N}\sum_{l=1}^{N} e^{\mathrm{i}s\cdot X_l}\right)\left(e^{\mathrm{i}t\cdot Y_k} - \tfrac{1}{N}\sum_{l=1}^{N} e^{\mathrm{i}t\cdot Y_l}\right)$$
$$\times\left(e^{-\mathrm{i}s'\cdot X_j} - \tfrac{1}{N}\sum_{l=1}^{N} e^{-\mathrm{i}s'\cdot X_l}\right)\left(e^{-\mathrm{i}t'\cdot Y_j} - \tfrac{1}{N}\sum_{l=1}^{N} e^{-\mathrm{i}t'\cdot Y_l}\right)\Bigg]$$
$$= \frac{1}{N^2}\sum_{j,k=1}^{N}\mathbb{E}\Bigg[\left(e^{\mathrm{i}s\cdot X_k} - \tfrac{1}{N}\sum_{l=1}^{N} e^{\mathrm{i}s\cdot X_l}\right)\left(e^{-\mathrm{i}s'\cdot X_j} - \tfrac{1}{N}\sum_{l=1}^{N} e^{-\mathrm{i}s'\cdot X_l}\right)\Bigg]$$
$$\times\mathbb{E}\Bigg[\left(e^{\mathrm{i}t\cdot Y_k} - \tfrac{1}{N}\sum_{l=1}^{N} e^{\mathrm{i}t\cdot Y_l}\right)\left(e^{-\mathrm{i}t'\cdot Y_j} - \tfrac{1}{N}\sum_{l=1}^{N} e^{-\mathrm{i}t'\cdot Y_l}\right)\Bigg].$$

A lengthy but otherwise straightforward calculation shows that

$$\mathbb{E}\Bigg[\left(e^{\mathrm{i}s\cdot X_k} - \tfrac{1}{N}\sum_{l=1}^{N} e^{\mathrm{i}s\cdot X_l}\right)\left(e^{-\mathrm{i}s'\cdot X_j} - \tfrac{1}{N}\sum_{l=1}^{N} e^{-\mathrm{i}s'\cdot X_l}\right)\Bigg]$$
$$= \mathbb{E}\big(e^{\mathrm{i}s\cdot X_k - \mathrm{i}s'\cdot X_j}\big) - \frac{N-1}{N} f_X(s)\overline{f_X(s')} - \frac{1}{N} f_X(s-s'),$$

and an analogous formula holds for the Y_i. Summing over $k, j = 1,\ldots,N$ and distinguishing between the cases $k = j$ and $k \neq j$ finally gives

$$\mathbb{E}\big(Z_N(s,t)\overline{Z_N(s',t')}\big)$$
$$= \left(\frac{(N-1)^2}{N^3} + \frac{N-1}{N^3}\right)\big(f_X(s-s') - f_X(s)\overline{f_X(s')}\big)\big(f_Y(t-t') - f_Y(t)\overline{f_Y(t')}\big),$$

and the lemma follows. $\qquad\square$

Remark 4.7. a) If we symmetrize the expression for $N \cdot {}^N V^2$ in a suitable way, we can transform it into a degenerate U-statistic. For random variables with bounded second moments we can then use classical results to show the convergence to $\sum_{k=1}^{\infty} \lambda_k X_k^2$, where λ_k are some coefficients and X_k are i.i.d. standard normal random variables, see e.g. Serfling [32, Sec. 5.5.2] or Witting & Müller-Funk [39, Satz 7.183]. In order to relax the bounds on the moments one would have to show convergence of the corresponding λ_k.

b) The log-moment condition can be slightly relaxed, but this leads to a much more involved expression, cf. Csörgő [14], and for the case $\epsilon = 0$ a counterexample is known, see Csörgő [12, p. 133]. Unfortunately, the convergence of the characteristic functions is stated without any moment condition in Murata [24, Thm. 4] which is based on Feuerverger & Mureika [16, Thm. 3.1].

Corollary 4.8. *Assume that X, Y are non-degenerate with $\mathbb{E}\Phi(X) + \mathbb{E}\Psi(Y) < \infty$ and set*

$$a_N := \frac{1}{N^2} \sum_{i,k=1}^{N} \Phi(X_i - X_k) \quad \text{and} \quad b_N := \frac{1}{N^2} \sum_{j,l=1}^{N} \Psi(Y_j - Y_l).$$

a) *If X, Y are independent random variables satisfying the log-moment conditions $\mathbb{E}\log^{1+\epsilon}(1 + |X|^2) < \infty$ and $\mathbb{E}\log^{1+\epsilon}(1 + |Y|^2) < \infty$ for some $\epsilon > 0$, then*

$$\frac{N \cdot {}^N V^2}{a_N b_N} \xrightarrow[N \to \infty]{d} \sum_{k=1}^{\infty} \alpha_k X_k^2, \tag{79}$$

where X_i are independent standard normal random variables and the coefficients α_k satisfy $\alpha_k \geq 0$ with $\sum_{k=1}^{\infty} \alpha_k = 1$.

b) *If the random variables X and Y are not independent, then*

$$\frac{N \cdot {}^N V^2}{a_N b_N} \xrightarrow[N \to \infty]{d} \infty. \tag{80}$$

Proof. We can almost literally follow the proof of Corollary 2 and Theorem 6 in Székely et al. [37]. Just note that $a_N b_N$ is an estimator for $\mathbb{E}\Phi(X - X') \cdot \mathbb{E}\Psi(Y - Y')$ and this is, by (64) and (76), the limit of the expectation of the numerator. $\qquad\square$

5 Remarks on the uniqueness of the Cauchy distance covariance

It is a natural question whether it is possible to extend distance covariance further by taking a measure ρ in the definition (22) which does not factorize. To ensure the finiteness of $V(X, Y)$ one still has to assume

$$\int \left(1 \wedge |s|^2\right)\left(1 \wedge |t|^2\right) \rho(\mathrm{d}s, \mathrm{d}t) < \infty, \tag{81}$$

see also Székely & Rizzo [36, Eq. (2.4)] and (22) at the beginning of Section 3.2. Furthermore, it is no restriction to assume that ρ is symmetric, since the integrand in

(22) is symmetric, hence we can always symmetrize any non-symmetric measure ρ without changing the value of the integral. Thus, the function

$$\Theta(x, y) := \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} \big(1 - \cos(x \cdot s)\big)\big(1 - \cos(x \cdot t)\big) \, \rho(\mathrm{d}s, \mathrm{d}t) \tag{82}$$

is well-defined and symmetric in each variable. The corresponding generalized distance covariance $V^2(X, Y)$ can be expressed by (28), if the expectations on the right-hand side are finite. Note, however, that the nice and computationally feasible representations of $V^2(X, Y)$ make essential use of the factorization of ρ which means that they are no longer available in this setting.

Let X and Y be random variables with values in $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively, such that for some $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$

$$\mathbb{P}(X = 0) = \mathbb{P}(X = x) = \mathbb{P}(Y = 0) = \mathbb{P}(Y = y) = \frac{1}{2}. \tag{83}$$

A direct calculation of (28), using $\Theta(0, y) = \Theta(x, 0) = 0$, gives

$$V^2(X, Y) = \gamma \cdot \Theta(x, y), \tag{84}$$

with

$$\begin{aligned} \gamma = {} & 2\mathbb{P}(X_1 \neq X_5, Y_1 \neq Y_6) - \mathbb{P}(X_1 \neq X_4, Y_1 \neq Y_4) \\ & - \mathbb{P}(X_2 \neq X_5)\mathbb{P}(Y_3 \neq Y_6), \end{aligned} \tag{85}$$

where (X_i, Y_i), $i = 1, \ldots, 6$, are i.i.d. copies of the random vector (X, Y).

Now suppose that $V^2(X, Y)$ is homogeneous and/or rotationally invariant, i.e. for some $\alpha, \beta \in (0, 2)$, all scalars $a, b > 0$ and orthogonal matrices $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{m \times m}$

$$V^2(aX, bY) = a^\alpha b^\beta V^2(X, Y), \tag{86}$$
$$V^2(AX, BY) = V^2(X, Y) \tag{87}$$

hold. The homogeneity, (86), yields

$$\Theta(x, y) = |x|^\alpha |y|^\beta \Theta\big(\tfrac{x}{|x|}, \tfrac{y}{|y|}\big), \tag{88}$$

and the rotational invariance, (87), shows that $\Theta(x/|x|, y/|y|)$ is a constant. In particular, homogeneity of degree $\alpha = \beta = 1$ and rotational invariance yield that $\Theta(x, y) = \mathrm{const} \cdot |x| \cdot |y|$. Since the Lévy–Khintchine formula furnishes a one-to-one correspondence between the cndf and its Lévy triplet, see (the comments following) Theorem 2.1, this determines ρ uniquely: it factorizes into two Cauchy Lévy measures. This means that – even in a larger class of weights – the assumptions (86) and (87) imply a unique (up to a constant) choice of weights, and we have recovered Székely-and-Rizzo's uniqueness result from [36].

Lemma 5.1. *Let $V^2(X, Y) := \|f_{X,Y} - f_X \otimes f_Y\|^2_{L^2(\rho)}$ be a generalized distance covariance as in Definition 2.3 and assume that the symmetric measure ρ satisfies*

the integrability condition (81). *If $V^2(X, Y)$ is homogeneous of order $\alpha \in (0, 2)$ and $\beta \in (0, 2)$ and rotationally invariant in each argument, then the measure ρ defining $V^2(X, Y)$ is – up to a multiplicative constant – of the form*

$$\rho(\mathrm{d}s, \mathrm{d}t) = c(\alpha, m)c(\beta, n)|s|^{-\alpha-m}|t|^{-\beta-n}\, \mathrm{d}s\, \mathrm{d}t.$$

Moreover, $V^2(X, Y)$ can be represented by (28) *with $\Theta(x, y) = C \cdot |x|^\alpha \cdot |y|^\beta$.*

For completeness, let us mention that the constants $c(\alpha, m)$ and $c(\beta, n)$ are of the form

$$c(\alpha, m) = \alpha 2^{\alpha-1}\pi^{-m/2}\Gamma\left(\tfrac{\alpha+m}{2}\right)/\Gamma\left(1 - \tfrac{\alpha}{2}\right),$$

see e.g. [10, p. 34, Example 2.4.d)] or [3, p. 184, Exercise 18.23].

6 Generalized distance correlation

We continue our discussion in the setting of Section 3.2. Let $\rho = \mu \otimes \nu$ and assume that μ and ν are symmetric Lévy measures on $\mathbb{R}^m \setminus \{0\}$ and $\mathbb{R}^n \setminus \{0\}$, each with full support. For m- and n-dimensional random variables X and Y the generalized distance covariance is, cf. Definition 3.1,

$$V(X, Y) = \sqrt{\iint |f_{(X,Y)}(s, t) - f_X(s)f_Y(t)|^2 \, \mu(\mathrm{d}s)\, \nu(\mathrm{d}t)}. \tag{89}$$

We set

$$V(X) := \sqrt{\int |f_{(X,X)}(s, t) - f_X(s)f_X(t)|^2 \, \mu(\mathrm{d}s)\, \mu(\mathrm{d}t)}$$

$$= \|f_{(X,X)} - f_X \otimes f_X\|_{L^2(\mu \otimes \mu)}, \tag{90}$$

$$V(Y) := \sqrt{\int |f_{(Y,Y)}(s, t) - f_Y(s)f_Y(t)|^2 \, \nu(\mathrm{d}s)\, \nu(\mathrm{d}t)}$$

$$= \|f_{(Y,Y)} - f_Y \otimes f_Y\|_{L^2(\nu \otimes \nu)}, \tag{91}$$

and define *generalized distance correlation* as

$$R(X, Y) := \begin{cases} \dfrac{V(X, Y)}{\sqrt{V(X)V(Y)}}, & \text{if } V(X) \cdot V(Y) > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{92}$$

Using the Cauchy–Schwarz inequality it follows from (38) that, whenever $R(X, Y)$ is well defined, one has

$$0 \le R(X, Y) \le 1 \quad \text{and} \quad R(X, Y) = 0 \text{ iff } X, Y \text{ are independent.} \tag{93}$$

The *sample distance correlation* is given by

$$^N R\big((x_1, y_1), \ldots, (x_N, y_N)\big) = \left(\frac{\frac{1}{N^2}\sum_{k,l=1}^N A_{kl}B_{kl}}{\sqrt{\frac{1}{N^2}\sum_{k,l=1}^N A_{kl}A_{kl}} \cdot \sqrt{\frac{1}{N^2}\sum_{k,l=1}^N B_{kl}B_{kl}}} \right)^{\frac{1}{2}}, \tag{94}$$

where we use the notation introduced in Lemma 4.2.

Example 6.1. For standard normal random variables X, Y with $\rho = \mathrm{Cor}(X, Y)$, $\Phi(x) = |x|$ and $\Psi(y) = |y|$ the distance correlation becomes

$$R(X, Y) = \left(\frac{\sqrt{1 - \rho^2} - \sqrt{4 - \rho^2} + \rho(\arcsin \rho - \arcsin \frac{\rho}{2}) + 1}{1 - \sqrt{3} + \pi/3} \right)^{1/2} \le |\rho|, \quad (95)$$

cf. Székely & Rizzo [34, Thm. 6].

7 Gaussian covariance

Let us finally show that the results on Brownian covariance of Székely & Rizzo [34, Sec. 3] do have an analogue for the generalized distance covariance.

For a symmetric Lévy measure with corresponding continuous negative definite function $\Phi : \mathbb{R}^m \to \mathbb{R}$ let G_Φ be the Gaussian field indexed by $\mathbb{R}^m$ with

$$\mathbb{E}G_\Phi(x) = 0 \quad \text{and} \quad \mathbb{E}\big(G_\Phi(x)G_\Phi(x')\big) = \Phi(x) + \Phi(x') - \Phi(x - x'). \quad (96)$$

Analogously we define the random field G_Ψ.

For a random variable Z with values in $\mathbb{R}^d$ and for a Gaussian random field G indexed by $\mathbb{R}^d$ we set

$$Z^G := G(Z) - \mathbb{E}\big(G(Z) \mid G\big). \quad (97)$$

Definition 7.1. Let G_Φ, G_Ψ be mean-zero Gaussian random fields indexed by $\mathbb{R}^m$ and $\mathbb{R}^n$ and with covariance structure given by the cndfs Φ and Ψ, respectively. For any two m- and n-dimensional random variables X and Y the *Gaussian covariance* is defined as

$$G^2(X, Y) := \mathrm{Cov}^2_{G_\Phi, G_\Psi}(X, Y) := \mathbb{E}\big(X_1^{G_\Phi} X_2^{G_\Phi} Y_1^{G_\Psi} Y_2^{G_\Psi}\big), \quad (98)$$

where (X_1, Y_1), (X_2, Y_2) are i.i.d. copies of (X, Y).

We can now identify Gaussian covariance and generalized distance covariance.

Theorem 7.2 (Gaussian covariance is generalized distance covariance). *Assume that X and Y satisfy $\mathbb{E}\Phi(X) + \mathbb{E}\Psi(Y) < \infty$. If G_Φ and G_Ψ are independent, then*

$$G^2(X, Y) = V^2(X, Y). \quad (99)$$

Proof. The proof is similar to Székely & Rizzo [34, Thm. 8]. By conditioning and the independence of G_Φ and G_Ψ we see

$$\begin{aligned}
&\mathbb{E}\big(X_1^{G_\Phi} X_2^{G_\Phi} Y_1^{G_\Psi} Y_2^{G_\Psi}\big) \\
&= \mathbb{E}\big(\mathbb{E}\big(X_1^{G_\Phi} X_2^{G_\Phi} \mid X_1, X_2, Y_1, Y_2\big) \cdot \mathbb{E}\big(Y_1^{G_\Psi} Y_2^{G_\Psi} \mid X_1, X_2, Y_1, Y_2\big)\big). \quad (100)
\end{aligned}$$

Using $\mathbb{E}(G_\Phi(x)G_\Phi(x')) = \Phi(x) + \Phi(x') - \Phi(x - x') =: \varphi(x, x')$ yields

$$\begin{aligned}
&\mathbb{E}\big(X_1^{G_\Phi} X_2^{G_\Phi} \mid X_1, X_2, Y_1, Y_2\big) \\
&= \varphi(X_1, X_2) - \mathbb{E}\big(\varphi(X_1, X_2) \mid X_1\big) - \mathbb{E}\big(\varphi(X_1, X_2) \mid X_2\big) + \mathbb{E}\varphi(X_1, X_2) \\
&= -\Phi(X_1 - X_2) + \mathbb{E}\big(\Phi(X_1 - X_2) \mid X_1\big) + \mathbb{E}\big(\Phi(X_1 - X_2) \mid X_2\big) \\
&\quad - \mathbb{E}\Phi(X_1 - X_2),
\end{aligned}$$

where the second equality is due to cancellations. An analogous calculation for $\mathbb{E}(Y_1^{G_\Psi} Y_2^{G_\Psi} \mid X_1, X_2, Y_1, Y_2)$ turns (100) into (35).

For (100) we have to make sure that $\mathbb{E}(|X_1^{G_\Phi} X_2^{G_\Phi} Y_1^{G_\Psi} Y_2^{G_\Psi}|) < \infty$. This follows from

$$
\begin{aligned}
&\mathbb{E}\left(|X_1^{G_\Phi} X_2^{G_\Phi} Y_1^{G_\Psi} Y_2^{G_\Psi}|\right) \\
&= \mathbb{E}\left(\mathbb{E}\left[|X_1^{G_\Phi} X_2^{G_\Phi}| \mid X_1, X_2, Y_1, Y_2\right] \mathbb{E}\left[|Y_1^{G_\Psi} Y_2^{G_\Psi}| \mid X_1, X_2, Y_1, Y_2\right]\right) \\
&\leq \mathbb{E}\Bigg(\sqrt{\mathbb{E}\left[|X_1^{G_\Phi}|^2 \mid X_1, X_2, Y_1, Y_2\right] \mathbb{E}\left[|X_2^{G_\Phi}|^2 \mid X_1, X_2, Y_1, Y_2\right]} \\
&\qquad\qquad \times \sqrt{\mathbb{E}\left[|Y_1^{G_\Psi}|^2 \mid X_1, X_2, Y_1, Y_2\right] \mathbb{E}\left[|Y_2^{G_\Psi}|^2 \mid X_1, X_2, Y_1, Y_2\right]}\Bigg) \\
&= \mathbb{E}\Bigg(\sqrt{\mathbb{E}\left[|X_1^{G_\Phi}|^2 \mid X_1, Y_1\right] \mathbb{E}\left[|X_2^{G_\Phi}|^2 \mid X_2, Y_2\right]} \\
&\qquad\qquad \times \sqrt{\mathbb{E}\left[|Y_1^{G_\Psi}|^2 \mid X_1, Y_1\right] \mathbb{E}\left[|Y_2^{G_\Psi}|^2 \mid X_2, Y_2\right]}\Bigg) \\
&= \mathbb{E}\left(\sqrt{\mathbb{E}\left[|X_1^{G_\Phi}|^2 \mid X_1, Y_1\right]}\right)^2 \mathbb{E}\left(\sqrt{\mathbb{E}\left[|Y_1^{G_\Psi}|^2 \mid X_1, Y_1\right]}\right)^2 \\
&\leq \mathbb{E}\left(|X_1^{G_\Phi}|^2\right) \mathbb{E}\left(|Y_1^{G_\Psi}|^2\right).
\end{aligned}
$$

In this calculation we use first the independence of G_Φ and G_Ψ, the conditional Cauchy–Schwarz inequality and the fact that the random variables (X_1, Y_1) and (X_2, Y_2) are i.i.d. In the final estimate we use again the Cauchy–Schwarz inequality. In order to see that the right-hand side is finite, we note that (96) and (97) yield

$$
\begin{aligned}
\mathbb{E}\left(|X^{G_\Phi}|^2\right) &= \mathbb{E}\left(|G_\Phi(X) - \mathbb{E}(G_\Phi(X) \mid G_\Phi)|^2\right) \\
&\leq 2\mathbb{E}\left(|G_\Phi(X)|^2\right) = 4\mathbb{E}\Phi(X) < \infty.
\end{aligned}
$$

A similar estimate for Y completes the proof. $\square$

8 Conclusion

We have shown that the concept of distance covariance introduced by Székely et al. [37] can be embedded into a more general framework based on Lévy measures, cf. Section 3. In this generalized setting the key results for statistical applications are: the convergence of the estimators and the fact that also the limit distribution of the (scaled) estimators is known, cf. Section 4. Moreover – for applications this is of major importance – the estimators have the numerically efficient representation (50).

The results allow the use of generalized distance covariance in the tests for independence developed for distance covariance, e.g. tests based on a general Gaussian quadratic form estimate or resampling tests. The test statistic is the function $T := \frac{N \cdot {}^N V^2}{a_N b_N}$ discussed in Corollary 4.8. Using the quadratic form estimate (see [37]

for details) its p-value can be estimated by $1 - F(T)$ where F is the distribution function of the Chi-squared distribution with 1 degree of freedom. This test and re-sampling tests are studied in detail in [9] and [7], respectively. In addition, these papers contain illustrating examples which show that the new flexibility provided by the choice of the Lévy measures (equivalently: by the choice of the continuous negative definite function) can be used to improve the power of these tests. Moreover, new test procedures using distance covariance and its generalizations are developed in [5].

Finally, the presented results are also the foundation for a new approach to testing and measuring multivariate dependence, i.e. the mutual (in)dependence of an arbitrary number of random vectors. This approach is developed in [9] accompanied by extensive examples and further applications in [7]. All functions required for the use of generalized distance covariance in applications are implemented in the R package `multivariance` [8].

References

[1] Bakirov, N.K., Székely, G.J.: Brownian covariance and central limit theorem for stationary sequences. Theory of Probability and its Applications **55**(3), 371–394 (2011) MR2768533. https://doi.org/10.1137/S0040585X97984954

[2] Banyamini, Y., Lindenstrauss, J.: Geometric Nonlinear Functional Analysis. American Mathematical Society, Providence (RI) (2000) MR1727673

[3] Berg, C., Forst, G.: Potential Theory on Locally Compact Abelian Groups. Springer, Berlin (1975) MR0481057

[4] Berrett, T.B., Samworth, R.J.: Nonparametric independence testing via mutual information. arXiv: 1711.06642v1 (2017)

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

[6] Bickel, P.J., Xu, Y.: Discussion of: Brownian distance covariance. The Annals of Applied Statistics **3**(4), 1266–1269 (2009) MR2752128. https://doi.org/10.1214/09-AOAS312A

[7] Böttcher, B.: Dependence structures - estimation and visualization using distance multivariance. arXiv: 1712.06532v1 (2017)

[8] Böttcher, B.: Multivariance: Measuring Multivariate Dependence Using Distance Multivariance. (2017). R package version 1.0.5

[9] Böttcher, B., Keller-Ressel, M., Schilling, R.L.: Distance Multivariance: New dependence measures for random vectors (submitted). Revised version of arXiv: 1711.07775v1 (2017)

[10] Böttcher, B., Schilling, R.L., Wang, J.: Lévy-Type Processes: Construction, Approximation and Sample Path Properties. Lecture Notes in Mathematics, Lévy Matters, vol. 2099. Springer (2013) MR3156646. https://doi.org/10.1007/978-3-319-02684-8

[11] Cope, L.: Discussion of: Brownian distance covariance. The Annals of Applied Statistics **3**(4), 1279–1281 (2009). doi:https://doi.org/10.1214/00-AOAS312C MR2752130. https://doi.org/10.1214/00-AOAS312C

[12] Csörgő, S.: Limit behaviour of the empirical characteristic function. The Annals of Probability, 130–144 (1981) MR0606802

[13] Csörgő, S.: Multivariate empirical characteristic functions. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **55**(2), 203–229 (1981). MR0608017. https://doi.org/10.1007/BF00535160

[14] Csörgő, S.: Testing for independence by the empirical characteristic function. Journal of Multivariate Analysis **16**(3), 290–299 (1985) MR0793494. https://doi.org/10.1016/0047-259X(85)90022-3

[15] Feuerverger, A.: Discussion of: Brownian distance covariance. The Annals of Applied Statistics **3**(4), 1282–1284 (2009) MR2752131. https://doi.org/10.1214/09-AOAS312D

[16] Feuerverger, A., Mureika, R.A.: The empirical characteristic function and its applications. The Annals of Statistics, 88–97 (1977) MR0428584

[17] Genovese, C.R.: Discussion of: Brownian distance covariance. The Annals of Applied Statistics **3**(4), 1299–1302 (2009). doi:https://doi.org/10.1214/09-AOAS312G MR2752134. https://doi.org/10.1214/09-AOAS312G

[18] Gretton, A., Fukumizu, K., Sriperumbudur, B.K.: Discussion of: Brownian distance covariance. The Annals of Applied Statistics **3**(4), 1285–1294 (2009). MR2752132. https://doi.org/10.1214/09-AOAS312E

[19] Jacob, N.: Pseudo-Differential Operators and Markov Processes I. Fourier Analysis and Semigroups. Imperial College Press, London (2001) MR1873235. https://doi.org/10.1142/9781860949746

[20] Jacob, N., Knopova, V., Landwehr, S., Schilling, R.L.: A geometric interpretation of the transition density of a symmetric Lévy process. Science China: Mathematics **55**, 1099–1126 (2012) MR2925579. https://doi.org/10.1007/s11425-012-4368-0

[21] Jin, Z., Matteson, D.S.: Generalizing Distance Covariance to Measure and Test Multivariate Mutual Dependence. arXiv: 1709.02532v1 (2017)

[22] Kosorok, M.R.: Discussion of: Brownian distance covariance. The Annals of Applied Statistics **3**(4), 1270–1278 (2009). doi:https://doi.org/10.1214/09-AOAS312B MR2752129. https://doi.org/10.1214/09-AOAS312B

[23] Lyons, R.: Distance covariance in metric spaces. The Annals of Probability **41**(5), 3284–3305 (2013) MR3127883. https://doi.org/10.1214/12-AOP803

[24] Murata, N.: Properties of the empirical characteristic function and its application to testing for independence. In: Third International Workshop on Independent Component Analysis and Signal Separation (ICA2001), pp. 295–300 (2001)

[25] Newton, M.A.: Introducing the discussion paper by Székely and Rizzo. The Annals of Applied Statistics **3**(4), 1233–1235 (2009). MR2752126. https://doi.org/10.1214/09-AOAS34INTRO

[26] Rémillard, B.: Discussion of: Brownian distance covariance. The Annals of Applied Statistics **3**(4), 1295–1298 (2009). doi:https://doi.org/10.1214/09-AOAS312F MR2752133. https://doi.org/10.1214/09-AOAS312F

[27] Sasvári, Z.: Positive Definite and Definitizable Functions. Akademie-Verlag, Berlin (1994) MR1270018

[28] Sato, K.: Lévy Processes and Infinitely Divisible Distributions. Cambridge University Press, Cambridge (1999) MR1739520

[29] Schilling, R., Schnurr, A.: The symbol associated with the solution of a stochastic differential equation. Electronic Journal of Probability **15**, 1369–1393 (2010) MR2721050. https://doi.org/10.1214/EJP.v15-807

[30] Schilling, R.L., Song, R., Vondraček, Z.: Bernstein Functions. Theory and Applications, 2nd edn. de Gruyter (2012) MR2978140. https://doi.org/10.1515/9783110269338

[31] Schoenberg, I.J.: Metric spaces and positive definite functions. Transactions of the American Mathematical Society **44**(3), 522–536 (1938) MR1501980. https://doi.org/10.2307/1989894

[32] Serfling, R.J.: Approximation Theorems of Mathematical Statistics. John Wiley & Sons (2009) MR0595165

[33] Székely, G.J., Rizzo, M.L.: Hierarchical clustering via joint between-within distances: Extending ward's minimum variance method. Journal of Classification **22**(2), 151–183 (2005) MR2231170. https://doi.org/10.1007/s00357-005-0012-9

[34] Székely, G.J., Rizzo, M.L.: Brownian distance covariance. The Annals of Applied Statistics **3**(4), 1236–1265 (2009) MR2752127. https://doi.org/10.1214/09-AOAS312

[35] Székely, G.J., Rizzo, M.L.: Rejoinder: Brownian distance covariance. The Annals of Applied Statistics **3**(4), 1303–1308 (2009). doi:https://doi.org/10.1214/09-AOAS312REJ MR2752135. https://doi.org/10.1214/09-AOAS312REJ

[36] Székely, G.J., Rizzo, M.L.: On the uniqueness of distance covariance. Statistics and Probability Letters **82**(12), 2278–2282 (2012) MR2979766. https://doi.org/10.1016/j.spl.2012.08.007

[37] Székely, G.J., Rizzo, M.L., Bakirov, N.K.: Measuring and testing dependence by correlation of distances. The Annals of Statistics **35**(6), 2769–2794 (2007) MR2382665. https://doi.org/10.1214/009053607000000505

[38] Ushakov, N.G.: Selected Topics in Characteristic Functions. VSP (1999) MR1745554. https://doi.org/10.1515/9783110935981

[39] Witting, H., Müller-Funk, U.: Mathematische Statistik II. Teubner, Stuttgart (1995) MR1363716. https://doi.org/10.1007/978-3-322-90152-1

[40] Zastavnyi, V.P.: On positive definiteness of some functions. Journal of Multivariate Analysis **73**(3), 55–81 (2000) MR1766121. https://doi.org/10.1006/jmva.1999.1864