Asymptotic Behaviour of the Empirical Distance Covariance for Dependent Data

Marius Kroll

Department of Mathematics, Ruhr-Universität Bochum, 44780 Bochum, Germany
E-Mail: marius.kroll@rub.de

July 20, 2020

Abstract

We give two asymptotic results for the distance covariance on separable metric spaces without any iid assumptions. In particular, we show the almost sure convergence of the empirical distance covariance under ergodicity for any measure with finite first moments. We further give a result concerning the asymptotic distribution of the empirical distance covariance under the assumption of absolute regularity and extend these results to certain types of pseudometric spaces. In the process, we derive a general theorem concerning the asymptotic distribution of degenerate V-statistics of order 2 under a strong mixing condition.

Keywords and phrases  Distance covariance, distance correlation, negative type, test of independence, mixing conditions.

MSC2020 subject classifications  62H20, 62G20, 60F05, 30L05

1 Introduction

In [10], Lyons introduced the concept of distance covariance for separable metric spaces, generalising the work done by Székely, Rizzo and Bakirov in [15]. In this very general case, the distance covariance of a measure \( \theta \) (on

*Supported by the German Research Council (DFG) via Research Training Group RTG 2131 (High dimensional phenomena in probability – fluctuations and discontinuity)
the product space $\mathcal{X} \times \mathcal{Y}$ of separable metric spaces $\mathcal{X}$ and $\mathcal{Y}$) with marginal distributions $\mu$ on $\mathcal{X}$ and $\nu$ on $\mathcal{Y}$ is defined as

$$\text{dcov}(\theta) := \int \delta_{\theta}(z, z') \, d\theta^2(z, z')$$

for $z = (x, y), z' = (x', y')$, where

$$\delta_{\theta}(z, z') := d_{\mu}(x, x')d_{\nu}(y, y')$$
$$d_{\mu}(x, x') := d(x, x') - a_{\mu}(x) - a_{\mu}(x') + D(\mu)$$
$$a_{\mu}(x) := \int d(x, x') \, d\mu(x')$$
$$D(\mu) := \int d(x, x') \, d\mu^2(x, x')$$

To examine the properties of this object, Lyons made use of the concept of (strong) negative type. A metric space $\mathcal{X}$ is said to be of negative type, if there exists a mapping $\phi : \mathcal{X} \to H$ to a Hilbert space $H$, such that $d(x, x') = \|\phi(x) - \phi(x')\|^2$ for all $x, x' \in \mathcal{X}$. It is of strong negative type if it is of negative type and $D(\mu_1 - \mu_2) = 0$ if and only if $\mu_1 = \mu_2$ for all probability measures $\mu_1, \mu_2$ with finite first moments. Lyons showed that the distance covariance is non-negative if $\mathcal{X}$ and $\mathcal{Y}$ are of negative type, and that the property $\text{dcov}(\theta) = 0 \iff \theta = \mu \otimes \nu$ holds if $\mathcal{X}$ and $\mathcal{Y}$ are of strong negative type.

This means that the distance covariance completely characterises independence of random variables in metric spaces of strong negative type. Estimators for the distance covariance and their asymptotic behaviour are therefore of great interest for tests of independence.

A special case for real-valued random variables follows from choosing the embedding

$$\phi : \mathbb{R}^d \to L^2(w_d) := \left\{ f : \mathbb{R}^d \to \mathbb{C} \mid \int |f|^2 w_d \, d\lambda^d < \infty \right\}$$
$$x \mapsto \frac{1}{\sqrt{2}}(1 - \exp(i\langle \cdot, x \rangle))$$

with $w_d(s) = \Gamma((d+1)/2)\pi^{-(d+1)/2}\|s\|^{-d+1}/2$, which Lyons in [10] refers to as the Fourier embedding. This results in the square of the distance covariance as introduced in [15], i.e.

$$\text{dcov}(\theta) = \int |\varphi_{\mathcal{X},\mathcal{Y}}(s, t) - \varphi_{\mathcal{X}}(s)\varphi_{\mathcal{Y}}(t)|^2 w_p(s)w_q(t) \, d(s, t)$$
where $\varphi_Z$ denotes the characteristic function of a random variable $Z$, and the vector $(X, Y) \in \mathbb{R}^{p+q}$ has distribution $\theta$.

Two of the main results of [10] are Proposition 2.6 and Theorem 2.7, which describe the asymptotic behaviour of $\text{dcov}(\theta_n)$, where $\theta_n$ is the empirical measure from $n$ iid-samples of $\theta$. Theorem 2.7, under sufficient moment assumptions, describes the asymptotic distribution of the sequence $n\text{dcov}(\theta_n)$, if $\theta = \mu \otimes \nu$. Proposition 2.6 gives the almost sure convergence $\text{dcov}(\theta_n) \xrightarrow{a.s.} \text{dcov}(\theta)$ for any measure $\theta$ with finite first moments. However, as noted by Jakobsen in [8], Lyons’ proof of Proposition 2.6 was incorrect and actually required $\theta$ to have finite $5/3$-moments. Lyons later acknowledged this in [11] (iii), showing that Proposition 2.6 as written in [10] is still correct in the case of spaces of negative type, but leaving the question of whether finite first moments are sufficient in the general case of separable metric spaces unanswered.

In Section 2, we show that one can indeed obtain the almost sure convergence of the estimator $\text{dcov}(\theta_n)$ under finite first moment assumption. We further generalise the theory presented in [10] in that we drop the iid assumption regarding the samples which constitute the empirical measure $\theta_n$. In Theorem 1, we show the almost sure convergence of $\text{dcov}(\theta_n)$ under assumption of ergodicity and finite first moments. In Theorem 3, we give an asymptotic result similar to Theorem 2.7 in [10], assuming absolute regularity. For this we make use of Theorem 2, which is a general result concerning the asymptotic distribution of degenerate V-statistics under the assumption of $\alpha$-mixing data. For definitions of $\alpha$-mixing and absolute regularity, as well as basic properties of these mixing conditions, see [4].

A further generalisation can be achieved by raising the metrics of the underlying metric spaces to the $\beta$-th power. We will denote this with $\text{dcov}_\beta$. Typically, $\beta$ is chosen between 0 and 2, where the choice $\beta = 1$ results in the regular distance covariance. An equivalent way of describing this is to use the regular definitions of distance covariance, but to consider pseudometric spaces of a particular kind instead of metric spaces, namely those which result from raising some metric to the $\beta$-th power (here, by a pseudometric we refer to a metric for which the triangle inequality need not hold). In Section 3, we generalise the results for metric spaces deduced in Section 2 to pseudometric spaces of this kind.

We now summarise some of the notation used in [10], as well as some basic properties of the distance covariance that will prove useful for our purposes.

Let $X$ and $Y$ be random variables with values in separable metric spaces $\mathcal{X}$ and $\mathcal{Y}$, respectively. Let $\theta = \mathcal{L}(X,Y)$ denote the distribution of $Z :=$. 
Distance Covariance for Dependent Data

\((X,Y)\), and \(\theta_n\) the empirical distribution of \(Z_1,...,Z_n \sim \theta\), where \((Z_k)_{k \in \mathbb{N}}\) is a strictly stationary and ergodic sequence with \(\mathcal{L}(Z_1) = \theta\). The variables \(Z_k\) have a representation \(Z_k = \Psi \circ T^k\) for a random variable \(\Psi\) with values in \(\mathcal{X} \times \mathcal{Y}\) and an ergodic, measure-preserving function \(T\) on the underlying probability space, the probability measure of which we will refer to as \(\rho\).

If we consider \(X\) to be of negative type via an embedding \(\phi\), we denote the Bochner integral \(\int \phi \, d\mu\) with \(\beta_\phi(\mu)\), and we write \(\hat{\phi}\) for the centered embedding \(\phi - \beta_\phi(\mu)\). If both \(X\) and \(Y\) are of negative type via embeddings \(\phi : \mathcal{X} \to H_1\) and \(\psi : \mathcal{Y} \to H_2\), we can consider the embedding

\[
\phi \otimes \psi : \mathcal{X} \times \mathcal{Y} \to H_1 \otimes H_2 \\
(x, y) \mapsto \phi(x) \otimes \psi(y)
\]

where \(H_1 \otimes H_2\) is the tensor product of the Hilbert spaces \(H_1\) and \(H_2\), equipped with the inner product \(\langle u_1 \otimes v_1, u_2 \otimes v_2 \rangle = \langle u_1, u_2 \rangle \langle v_1, v_2 \rangle\). We do not adopt different notations for the metrics of \(X\) and \(Y\) or the inner products on \(H_1, H_2\) or \(H_1 \otimes H_2\), as it is clear from their arguments which metric or inner product we consider.

By Proposition 3.5 in [10], we have that

\[
\delta_\theta(z, z') = 4((\hat{\phi} \otimes \hat{\psi})(z), (\hat{\phi} \otimes \hat{\psi})(z))
\]

for all \(z, z' \in \mathcal{X} \times \mathcal{Y}\), whenever \(X\) and \(Y\) are of negative type via embeddings \(\phi\) and \(\psi\), respectively.

2 Results for metric spaces

We now present our results in the case of separable metric spaces. It should be kept in mind that while we consider the usual distance correlation, Theorems 1 and 3 also hold for \(\text{dcov}_\beta\) (under appropriate moment conditions). However, we postpone discussion of this until Section 3 so as to avoid confusion by abstraction.

**Lemma 1.** Let \(X\) be a metrizable topological space, \((\mu_n)_{n \in \mathbb{N}}\) a sequence of measures on \(X\) with weak limit \(\mu\) and \(h : \mathcal{X} \to \mathbb{R}\) a \(\mu\)-a.s. continuous function which fulfills the following uniform integrability condition:

\[
\lim_{M \to \infty} \limsup_{n \to \infty} \int_{\{|h| > M\}} |h| \, d\mu_n = 0
\]

Furthermore, we require \(h\) to be dominated by some \(\mu\)-integrable function \(g\), i.e. \(|h| \leq g\) \(\mu\)-a.s. Then \(\int h \, d\mu_n \to \int h \, d\mu\).
Proof. Without loss of generality, suppose that $\mathcal{X}$ is a metric space. We can decompose the integral with respect to $\mu_n$ into a truncated part and a tail part:

$$\int h \, d\mu_n = \int_{\{|h| \leq M\}} h \, d\mu_n + \int_{\{|h| > M\}} h \, d\mu_n$$

The truncated integral converges, because it is the integral of an almost surely continuous and bounded function and $\mu_n \Rightarrow \mu$, while the uniform integrability condition (2) implies that the tail integral vanishes in the limit $M, n \to \infty$. More precisely, we have the inequality

$$\limsup_{n \to \infty} \int h \, d\mu_n \leq \limsup_{M \to \infty} \lim_{n \to \infty} \int_{\{|h| \leq M\}} h \, d\mu_n$$

$$+ \lim_{M \to \infty} \limsup_{n \to \infty} \int_{\{|h| > M\}} h \, d\mu_n$$

(3)

The second summand vanishes by assumption due to (2). For the first summand, note that for any fixed $M$, the limes superior in $n$ of the integral converges to $\int_{\{|h| \leq M\}} h \, d\mu$, since $h1_{\{|h| \leq M\}}$ is bounded and almost surely continuous. Furthermore, since $|h1_{\{|h| \leq M\}}| \leq |h| \leq g$, we can employ the dominated convergence theorem to obtain

$$\lim_{M \to \infty} \limsup_{n \to \infty} \int_{\{|h| \leq M\}} h \, d\mu_n = \int h \, d\mu$$

Therefore, the summands in (3) are indeed well-defined. This gives us

$$\limsup_{n \to \infty} \int h \, d\mu_n \leq \limsup_{M \to \infty} \lim_{n \to \infty} \int_{\{|h| \leq M\}} h \, d\mu_n + 0 = \int h \, d\mu$$

Since $0 \leq \liminf_{n \to \infty} \int_{\{|h| > M\}} h \, d\mu_n \leq \limsup_{n \to \infty} \int_{\{|h| > M\}} h \, d\mu_n$ for any $M$, we can use an almost identical argument to obtain

$$\liminf_{n \to \infty} \int h \, d\mu_n \geq \limsup_{M \to \infty} \liminf_{n \to \infty} \int_{\{|h| \leq M\}} h \, d\mu_n + 0 = \int h \, d\mu$$

and thus $\lim_{n \to \infty} \int h \, d\mu_n$ exists and is equal to $\int h \, d\mu$. \qed

In proving Theorem 1, we will make use of the following general result, which is a generalisation of Theorem U (ii) from [1].

Lemma 2. Let $(X_k)_{k \in \mathbb{N}}$ be a strictly stationary and ergodic process with values in a separable metrizable topological space $\mathcal{X}$ and marginal distribution...
Distance Covariance for Dependent Data

Let $h : X^d \to \mathbb{R}$ be a measurable function, and let $f : X \to \mathbb{R}$ be integrable with respect to $\theta$, so that $|h| \leq f \otimes \ldots \otimes f$, where the product denoted by $\otimes$ is taken $d$ times and $(f \otimes \ldots \otimes f)(x_1, \ldots, x_d) := \prod_{k=1}^{d} f(x_k)$. If $h$ is $\theta^d$-a.e. continuous, then $V_h(Z_1, \ldots, Z_n) \to \int h \, d\theta^d$ a.s., where $V_h(Z_1, \ldots, Z_n)$ denotes the $V$-statistics with kernel $h$, and the convergence holds with respect to $\rho$.

Proof. Without loss of generality, suppose that $X$ is a metric space. Let $\theta_n := n^{-1} \sum_{k=1}^{n} \delta_{X_k}$ denote the empirical measure of $X_1, \ldots, X_n$. We have $V_h(Z_1, \ldots, Z_n) = \int h \, d\theta_n$. Furthermore, $\theta_n \Rightarrow \theta$ a.s. (with respect to $\rho$), since $X$ is separable, and therefore $\theta_n \Rightarrow \theta^d$ a.s. by Theorem 2.8 (ii) in [3].

We now wish to employ Lemma 1. Hence, we need to show that the sequence of integrals fulfills the following uniform integrability condition:

$$\lim_{M \to \infty} \limsup_{n \to \infty} \int_{\{|h| > M\}} |h| \, d\theta_n^d = 0$$

We have

$$\int_{\{|h| > M\}} |h| \, d\theta_n^d \leq \int_{\{f \otimes \ldots \otimes f > M\}} f \otimes \ldots \otimes f \, d\theta_n^d$$

and since $\{f \otimes \ldots \otimes f > M\} \subseteq \bigcup_{i=1}^{d} M_i$ with $M_i := \{x \in X^d \mid f(x_i) > M_i^{1/d}\}$, the right-hand side is dominated by

$$\sum_{i=1}^{d} \int_{M_i} f \otimes \ldots \otimes f \, d\theta_n^d = d \left( \int f \, d\theta_n \right)^{d-1} \int_{\{f > M^{1/d}\}} f \, d\theta_n$$

which, due to Birkhoff's pointwise ergodic theorem, almost surely converges to $d \left( \mathbb{E}_{\theta} f \right)^{d-1} \mathbb{E}_{\theta} [1_{\{f > M^{1/d}\}}, f]$, where $1_A$ denotes the indicator function of a set $A$. Thus, almost surely,

$$\lim_{M \to \infty} \limsup_{n \to \infty} \int_{\{|h| > M\}} |h| \, d\theta_n^d \leq \lim_{M \to \infty} d \left( \mathbb{E}_{\theta} f \right)^{d-1} \mathbb{E}_{\theta} [1_{\{f > M^{1/d}\}}, f] = 0$$

since $f$ is assumed to be integrable.

Lemma 1 therefore gives us

$$V_h(Z_1, \ldots, Z_n) = \int h \, d\theta_n^d \xrightarrow{n \to \infty} \int h \, d\theta^d$$

where the almost sure convergence holds with respect to the measure $\rho$ on the underlying probability space. 

\[\square\]
Note that the following result does not require any assumptions beyond the separability of the metric spaces $\mathcal{X}$ and $\mathcal{Y}$ and the ergodicity of the samples generating the empirical measure $\theta_n$. Thus, Proposition 2.6 in [10], requiring iid samples, is a special case of our result. It follows that the original assumption of first moments is indeed correct.

**Theorem 1.** If $X$ and $Y$ have finite first moments, i.e. $E d(X, x_0), E d(Y, y_0) < \infty$ for some fixed (but arbitrary) $z_0 = (x_0, y_0) \in \mathcal{X} \times \mathcal{Y}$, then

$$\text{dcov}(\theta_n) \xrightarrow{a.s.} \text{dcov}(\theta)$$

where the almost sure convergence holds with respect to $\rho$.

**Proof.** We follow the idea of the proof of Proposition 2.6 in [10]. Consider the symmetric kernel $\bar{h}$, defined as the symmetrisation of $h$, where

$$h(z_1, ..., z_6) := f(x_1, ..., x_4)f(y_1, y_2, y_5, y_6)$$

$$f(x_1, ..., x_4) := d(x_1, x_2) - d(x_1, x_3) - d(x_2, x_4) + d(x_3, x_4)$$

As shown in the proof of Proposition 2.6 in [10], we have

$$|h(z_1, ..., z_6)| \leq 4d(x_2, x_3)d(y_1, y_6)$$

(4)

Let $z_0 = (x_0, y_0)$ be an arbitrary but fixed point in $\mathcal{X} \times \mathcal{Y}$. Since $a + b \leq ab$ for all real $a, b \geq 2$, we have

$$d(x, x') \leq d(x, x_0) + d(x', x_0) \leq (2 \vee d(x, x_0))(2 \vee d(x', x_0))$$

for all $x, x' \in \mathcal{X}$. Now, for $z = (x, y) \in \mathcal{X} \times \mathcal{Y}$, let $\varphi_i(z)$ be defined as $2 \vee d(x, x_0)$ if $i = 2, 3$ and as $2 \vee d(y, y_0)$ if $i = 1, 6$, and write $\varphi$ for the maximum over all these $\varphi_i$. Using (4), this gives us

$$|h(z_1, ..., z_6)| \leq 4\varphi(z_1)\varphi(z_2)\varphi(z_3)\varphi(z_6)$$

The functions $\varphi_i$ are continuous and measurable, since the underlying metric spaces are separable. They are also integrable because $X$ and $Y$ are assumed to have finite first moments. Using Lemma 2 therefore gives us $V_{\bar{h}}(Z_1, ..., Z_n) \rightarrow \int \bar{h} d\theta^0$ almost surely with respect to $\rho$, where $V_{\bar{h}}(Z_1, ..., Z_n)$ denotes the $V$-statistics with kernel $\bar{h}$. Since the $V$-statistics with kernel $h$ are equal to $\text{dcov}(\theta_n)$, and $\int \bar{h} d\theta^0 = \text{dcov}(\theta)$ (cf. [10]), this is what we wanted to show. \qed
Theorem 2. Let \( Z \) be a \( \sigma \)-compact metrizable topological space, \( (Z_k)_{k \in \mathbb{N}} \) a sequence of \( Z \)-valued random variables with marginal distribution \( L(Z_1) = \theta \). Consider a continuous, symmetric, degenerate and positive semidefinite kernel \( h : Z^2 \rightarrow \mathbb{R} \) with finite \((2 + \varepsilon)\)-moments with respect to \( \theta \) and finite \((1 + \frac{\varepsilon}{2})\)-moments on the diagonal, i.e. \( \mathbb{E}|h(Z_1, Z_1)|^{1+\varepsilon/2} < \infty \). Furthermore, let the sequence \( (Z_k)_{k \in \mathbb{N}} \) satisfy an \( \alpha \)-mixing condition such that \( \alpha(n) = O(n^{-r}) \) for some \( r > 1 + 2\varepsilon^{-1} \). Then, with \( V = V_h(Z_1, \ldots, Z_n) \) denoting the \( V \)-statistics with kernel \( h \),

\[
nV \xrightarrow{D} \sum_{k=1}^{\infty} \lambda_k \zeta_k^2
\]

where \((\lambda_k, \varphi_k)\) are pairs of the non-negative eigenvalues and matching eigenfunctions of the integral operator

\[
f \mapsto \int h(\cdot, z)f(z) \, d\theta(z)
\]

and \((\zeta_k)_{k \in \mathbb{N}}\) is a sequence of centered Gaussian random variables whose covariance structure is given by

\[
\text{Cov}(\zeta_i, \zeta_j) = \lim_{n \to \infty} \frac{1}{n} \sum_{t,u=1}^{n} \text{Cov}(\varphi_i(Z_t), \varphi_j(Z_u)) \tag{5}
\]

Proof. We note that the conditions of Theorem 2 in [14] are satisfied by Propositions 1-3 and Assumption 1 ibid., the latter of which is a consequence of \( \mathbb{E}|h(Z_1, Z_1)|^{1+\varepsilon/2} < \infty \). Hence we get

\[
h(z, z') = \sum_{k=1}^{\infty} \lambda_k \varphi_k(z) \varphi_k(z')
\]

for all \( z, z' \in \text{supp}(\theta) \). The \( \varphi_k \) are centered and form an orthonormal basis of \( L^2(\theta) \). Writing \( V(K) \) for the \( V \)-statistics for the truncated kernel \( \sum_{k=1}^{K} \lambda_k \varphi_k(z) \varphi_k(z') \), we note that \( nV(K) = \sum_{k=1}^{K} \lambda_k \zeta_{n,k}^2 \), where \( \zeta_{n,k} := n^{-1/2} \sum_{t=1}^{n} \varphi_k(Z_t) \). Using the Cramér-Wold theorem, we will now show that, for any \( K \in \mathbb{N} \), \((\zeta_{n,k})_{1 \leq k \leq K}\) weakly converges to \((\zeta_k)_{1 \leq k \leq K}\), where the \( \zeta_k \) are centered Gaussian variables with their covariances given in (5).

Let \( c_1, \ldots, c_K \) be real constants and set \( \xi_t := \sum_{k=1}^{K} c_k \varphi_k(Z_t) \). Then the \( \xi_t \) are centered random variables with \( \mathbb{E}\xi_t^2 = \sum_{k=1}^{K} c_k^2 \).

Note that, by definition, \( \varphi_k(z) = \lambda_k^{-1} \mathbb{E}[h(z, Z_1)\varphi_k(Z_1)] \) and thus

\[
|\varphi_k(z)| \leq |\lambda_k|^{-1} \|h(z, \cdot)\|_2
\]

(6)
Here, we have used the Cauchy-Schwarz inequality and the fact that the eigenfunctions $\varphi_k$ form an orthonormal basis of $L^2(\theta)$. This gives us

$$
\int |\varphi_k(z)|^{2+\varepsilon} \, d\theta(z) \leq \lambda_k^{- (2+\varepsilon)} \int \|h(z,.)\|^{2+\varepsilon} \, d\theta(z)
$$

$$
= \lambda_k^{- (2+\varepsilon)} \int \left( \int |h(z, z')|^2 \, d\theta(z') \right)^{\frac{2+\varepsilon}{2}} \, d\theta(z)
$$

$$
\leq \lambda_k^{- (2+\varepsilon)} \int |h(z, z')|^{2+\varepsilon} \, d\theta^2(z, z')
$$

by Jensen’s inequality, which implies $\|\varphi_k\|_{2+\varepsilon} \leq \lambda_k^{-1} \|h\|_{2+\varepsilon}$. Since our kernel $h$ has finite $(2 + \varepsilon)$-moments by assumption, this property translates to the eigenfunctions $\varphi_k$. Using Theorem 3.7 and Remark 1.8 in [4] therefore gives us

$$
|\text{Cov}(\varphi_k(Z_t), \varphi_l(Z_u))| \leq C \alpha(\sigma(Z_t), \sigma(Z_u))^{\varepsilon/(2+\varepsilon)} \leq C \alpha(|t - u|)^{\varepsilon/(2+\varepsilon)}
$$

for all $1 \leq k, l \leq K$, where $C$ is a positive constant depending on the corresponding eigenfunctions and $\varepsilon$-values. From this and the fact that $\alpha(n) = O(n^{-r})$ with $r > 1 + 2\varepsilon^{-1}$ it follows that, for any $k, l$, the series

$$
\sum_{d=1}^\infty \text{Cov}(\varphi_k(Z_1), \varphi_l(Z_{1+d}))
$$

and the series

$$
\lim_n n^{-1} \sum_{d=1}^{n-1} \text{Cov}(\varphi_k(Z_1), \varphi_l(Z_{1+d}))
$$

converge, since $d/n < 1$ for all $1 \leq d < n$. Thus, with $S_n$ denoting the sum over $\xi_1, ..., \xi_n$, we have that

$$
n^{-1} \sigma_n^2 := n^{-1} \mathbb{E} S_n^2 = n^{-1} \sum_{t,u=1}^{n} \sum_{k,l=1}^{K} c_k c_l \text{Cov}(\varphi_k(Z_t), \varphi_l(Z_u))
$$

$$
= \sum_{k=1}^{K} c_k^2 + n^{-1} \sum_{t \neq u} \sum_{k,l=1}^{K} c_k c_l \text{Cov}(\varphi_k(Z_t), \varphi_l(Z_u))
$$

$$
= \sum_{k=1}^{K} c_k^2 + n^{-1} 2(n-1) \sum_{d=1}^{n-1} \sum_{k,l=1}^{K} c_k c_l \text{Cov}(\varphi_k(Z_1), \varphi_l(Z_{1+d}))
$$

$$
\xrightarrow{n \to \infty} \sigma^2 < \infty
$$

where we have made use of the stationarity of the process $(Z_k)_{k \in \mathbb{N}}$ and the fact that the eigenfunctions $\varphi_k$ form an orthonormal basis of $L^2$. If $\xi_1, ..., \xi_K$ are Gaussian random variables with their covariance function given by (5), the limit $\sigma^2$ is the variance of the linear combination $\sum_{k=1}^{K} c_k \xi_k$.

We now show the uniform integrability of the sequence $(S_n^2 \sigma_n^{-2})_{n \in \mathbb{N}}$. It suffices to show that $\mathbb{E} |S_n \sigma_n^{-1}|^{2+\delta}$ is uniformly bounded in $n$ for some $\delta > 0$. 


Since $h$ has finite $(2 + \varepsilon)$-moments, we get

$$\sup_{n \in \mathbb{N}} \mathbb{E} \left| \sum_{k=1}^{K} c_k \varphi_k(Z_n) \right|^{2+\varepsilon} \leq \sup_{n \in \mathbb{N}} \left\{ K^{1+\varepsilon} \sum_{k=1}^{K} \mathbb{E} \left[ |c_k \varphi(Z_n)|^{2+\varepsilon} \right] \right\} < M(\varepsilon) < \infty$$

Here, we have made use of (6) and the stationarity of the sequence $(Z_n)$, which ensures that the upper bound $M(\varepsilon)$ is indeed uniform in $n$. Since $h(n) = O(n^{-r})$ with $r > 1 + 2\varepsilon^{-1}$ and $\sigma_n$ has rate of growth $\theta(\sqrt{n})$, Theorem 2.1 in [13] gives us $\mathbb{E}|S_n \sigma_n^{-1}|^{2+\delta} = O(1)$ for some $\delta > 0$. This implies uniform integrability of $(S_n^2 \sigma_n^{-2})_{n \in \mathbb{N}}$.

Using Theorem 10.2 from [4] therefore gives us

$$\sum_{k=1}^{K} c_k \zeta_{n,k} = \frac{S_n}{\sqrt{n}} \cdot \frac{\sigma_n}{\sqrt{n}} \xrightarrow{n \to \infty} \mathcal{N}(0, \sigma^2) = \mathcal{L} \left( \sum_{k=1}^{K} c_k \zeta_k \right)$$

and so, by the Cramér-Wold theorem, the vectors $(\zeta_{n,k})_{1 \leq k \leq K}$ converge to Gaussian vectors $(\zeta_k)_{1 \leq k \leq K}$ with the covariance structure described in (5) for any $K \in \mathbb{N}$.

Now, applying the continuous mapping theorem gives us

$$nV^{(K)} = \sum_{k=1}^{K} \lambda_k \zeta_{n,k}^2 \xrightarrow{n \to \infty} \sum_{k=1}^{K} \lambda_k \zeta_k^2 =: \zeta^{(K)}$$

and the summability of the eigenvalues $\lambda_k$, which is due to the identity $\sum_{k=1}^{\infty} \lambda_k = \mathbb{E} h(Z_1, Z_1) < \infty$, implies that

$$\mathbb{E} \left| \zeta - \zeta^{(K)} \right| = \sum_{k=1}^{\infty} \lambda_k \xrightarrow{K \to \infty} 0$$

We will now show that

$$\lim_{K \to \infty} \limsup_{n \to \infty} \mathbb{E}|nV - nV^{(K)}| = 0$$

We consider the Hilbert space $H$ of all real-valued sequences $(a_k)_{k \in \mathbb{N}}$ for which the series $\sum_{k=1}^{\infty} \lambda_k a_k^2$ converges, equipped with the inner product given by $\langle (a_k), (b_k) \rangle_H := \sum_{k=1}^{\infty} \lambda_k a_k b_k$. Then, writing $T_K(Z_t)$ for the $H$-valued random variable $(0^K, (\varphi_k(Z_t))_{k > K})$, where $0^K$ denotes the $K$-dimensional
zero vector, we get

\[\mathbb{E}|nV - nV(K)| = \mathbb{E}\left[\sum_{k > K} \lambda_k \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varphi_k(Z_t)\right)^2\right] = \mathbb{E}\left[\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} T_K(Z_t)\right)^2\right] = \text{Var} \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} T_K(Z_t)\right) = \frac{1}{n} \sum_{s,t=1}^{n} \text{Cov}(T_K(Z_s), T_K(Z_t))\]

Here, we define the covariance of two \(H\)-valued random variables \(X\) and \(Y\) as the real number \(\text{Cov}(X,Y) := \mathbb{E} \langle X,Y \rangle_H - \langle \mathbb{E} X, \mathbb{E} Y \rangle_H\). We aim to employ a covariance inequality for Hilbert-space valued random variables.

For this, let us first consider the \((2 + \varepsilon)\)-moments of \(T_K(Z_1)\). For any \(p > 0\), we get

\[\|T_K(Z_1)\|_p^p = \int \|T_K(z)\|_H^p \, d\theta(z) = \int \left(\sum_{k > K} \lambda_k \varphi_k(z)^2\right)^{p/2} \, d\theta(z) \leq \int \left(\sum_{k=1}^{\infty} \lambda_k \varphi_k(z)^2\right)^{p/2} \, d\theta(z) = \int h(z,z)^{p/2} \, d\theta(z) = \|h(Z_1,Z_1)\|_{p/2}^{p/2}\]

Since \(h\) has finite \((1 + \frac{\varepsilon}{2})\)-moments on the diagonal by assumption, this implies the \((2 + \varepsilon)\)-integrability of \(T_K(Z_1)\).

Lemma 2.2 in \cite{7} and the stationarity of the process \((Z_t)_{t \in \mathbb{N}}\) therefore gives us

\[|\text{Cov}(T_K(Z_s), T_K(Z_t))| \leq 15 \|T_K(Z_1)\|_{2+\varepsilon}^2 \alpha(|s - t|)^{\varepsilon/(2+\varepsilon)}\]

and we have shown before that \(n^{-1} \sum_{s,t=1}^{n} \alpha(|s - t|)^{\varepsilon/(2+\varepsilon)}\) converges to a finite limit \(c\). Furthermore, from \(\|T_K(Z_1)\|_{2+\varepsilon}^2 = \sum_{k > K} \lambda_k \frac{K \rightarrow \infty}{K \rightarrow \infty} 0\) and \(\|T_K(Z_1)\|_{2+\varepsilon} \leq \|T_1(Z_1)\|_{2+\varepsilon}\) (i.e. the sequence \((T_K(Z_1))_{K \in \mathbb{N}}\) is uniformly \((2 + \varepsilon)\)-integrable) it follows by Vitali’s Theorem that \(T_K(Z_1)\) \(\overset{(2+\varepsilon)}{\underset{K \rightarrow \infty}{\rightarrow}} 0\).

Putting all of the above together, we get

\[\lim_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{E}|nV - nV(K)| \leq 15c \lim_{K \rightarrow \infty} \|T_K(Z_1)\|_{2+\varepsilon}^2 = 0\]

By Theorem 3.2 in \cite{3}, \cite{7}, \cite{8} and \cite{9}, the latter of which we have just shown, imply \(nV \overset{D}{\underset{n \rightarrow \infty}{\rightarrow}} \zeta\). \(\square\)
Lemma 3. If $\mu$ has finite $q$-moments, then there exists an upper bound $M \in \mathbb{R}$ such that, for any collection of indices $i_1, \ldots, i_4$,
\[ \mathbb{E} \left[ f(X_{i_1}, \ldots, X_{i_4})^{2p} \right] \leq M(p) < \infty \]
for any $p < q$, where $f$ is the function from the proof of Theorem 4.

Proof. First, consider any two indices $i_1, i_2$. Then, due to (17), we have
\[ \mathbb{E}[d(X_{i_1}, X_{i_2})^q] \leq 2^{q-1}\mathbb{E}[d(X_{i_1}, x_0)^q + d(x_0, X_{i_2})^q] = 2^q \int d(x, x_0)^q \, d\mu(x) =: M_0 < \infty \]
where $x_0$ is some arbitrary point in $\mathcal{X}$.

Now, let $i_1, \ldots, i_4$ be fixed but arbitrary indices. Then, with a similar bound to the one used in Lemma 5,
\[ \mathbb{E}[f(X_{i_1}, \ldots, X_{i_4})^{2p}] \leq 4^p \mathbb{E}[d(X_{i_2}, X_{i_3})^p d(X_{i_1}, X_{i_4})^p] \]
\[ \leq 4^p \left\{ \mathbb{E}[d(X_{i_2}, X_{i_3})^p d(X_{i_1}, X_{i_4})^p] - \mathbb{E}[d(X_{i_2}, X_{i_3})^p] \mathbb{E}[d(X_{i_1}, X_{i_4})^p] \right\} \]
\[ + 4^p \mathbb{E}[d(X_{i_2}, X_{i_3})^p] \mathbb{E}[d(X_{i_1}, X_{i_4})^p] \]
We use Lemma 1 from [16] for the function $h(x_1, \ldots, x_4) := d(x_1, x_2)^p d(x_3, x_4)^p$ and the reordered collection $(i_2, i_3, i_1, i_4)$. Their assumptions are satisfied with $\delta := \frac{q}{p} - 1$, because
\[ \int h^{1+\delta} \, d(\mathcal{L}(X_{i_2}, X_{i_3}) \otimes \mathcal{L}(X_{i_1}, X_{i_4})) = \mathbb{E}[d(X_{i_2}, X_{i_3})^p] \mathbb{E}[d(X_{i_1}, X_{i_4})^p] \leq M_0^2 \]
due to (10). Thus, Lemma 1 in [16] gives us
\[ |\mathbb{E}[d(X_{i_2}, X_{i_3})^p d(X_{i_1}, X_{i_4})^p] - \mathbb{E}[d(X_{i_2}, X_{i_3})^p] \mathbb{E}[d(X_{i_1}, X_{i_4})^p]| \]
\[ \leq 4M_0^{\frac{q}{p+1}} \beta(|i_1 - i_3|)^{\delta} \]
where $\beta(n)$ is the $\beta$-mixing coefficient of the sequence $(Z_k)_{k \in \mathbb{N}}$. Because $\beta(n) \leq 1$ for all $n \in \mathbb{N}$, (10), (11) and (12) give us
\[ \mathbb{E}[f(X_{i_1}, \ldots, X_{i_4})^{2p}] \leq 4^{p+1} M_0^{\frac{q}{p+1}} + 4p M_0^2 =: M(p) < \infty \]

The following lemma is an adaptation of Lemma 2 in [16] in the sense that our result is implicitly contained in their proof. Another variant of this lemma (for U-statistics) can be found in [2]. Since both of these lemmas are slightly different from our version, we include a proof for the sake of completeness. However, it should be noted that all three proofs apply the same technique.
Lemma 4. Let $h$ be a symmetric and degenerate kernel of order $c \geq 2$. Here, we understand degeneracy as $E\beta = 0$. The sum converges due to our assumptions on $V$-statistic with kernel $r > c/p/(p-2)$, then $E[V^2] = O(n^{-c})$, where $V = V_h(Z_1, ..., Z_n)$ is the $V$-statistic with kernel $h$.

Proof. We will follow the basic idea of the proof of Lemma 2 in [16]. First, consider the special case of $c = 2$. We have

$$E \left[ \left( \sum_{1 \leq i, j \leq n} h(Z_{i1}, Z_{i2}) \right)^2 \right] = \sum_{1 \leq i, j, k, l \leq n} E[h(Z_{i1}, Z_{i2})h(Z_{k3}, Z_{l4})]$$

Now due to the degeneracy of our kernel $h$, we can employ Lemma 1 in [16] to obtain

$$E[h(Z_{i1}, Z_{i2})h(Z_{i3}, Z_{i4})] \leq M \cdot \beta (\max \{|i_2 - i_1|, |i_4 - i_3|\})^{(p-2)/p}$$

whenever $(i_1, i_2) \neq (i_3, i_4)$. Here, $M$ is some constant uniform in $i_1, ..., i_4$ and $n$.

Let us first assume that $k := |i_2 - i_1| \geq |i_4 - i_3| =: l$. For any fixed value of $k$, we have at most $2(n-k)$ possible values for $i_1$. Furthermore, since $k \geq l \geq 0$, we have $k + 1$ possible values for $l$ and, for any fixed $l$, at most $2(n-l)$ possible values for $i_3$. Writing

$$I := \{(i_1, ..., i_4) \mid 1 \leq i_1, ..., i_4 \leq n, |i_2 - i_1| \geq |i_4 - i_3|, (i_1, i_2) \neq (i_3, i_4)\}$$

this gives us

$$\sum_{i_1, ..., i_4 \in I} E[h(Z_{i1}, Z_{i2})h(Z_{i3}, Z_{i4})] \leq \sum_{k=0}^{n-1} \sum_{l=0}^{n-k} \sum_{i_3=1}^{n-l} M \beta(k)^{(p-2)/p}$$

$$\leq 4Mn^2 \sum_{k=0}^{n-1} (k+1)^{(p-2)/p}$$

$$= O(n^2)$$

The sum converges due to our assumptions on $\beta(n)$. The same bound can be established for the cases where $|i_4 - i_3| \geq |i_2 - i_1|$. The only combinations missing are those where $(i_1, i_2) = (i_3, i_4)$, of which there are $n^2$. We can combine these results to get

$$\sum_{1 \leq i_1, ..., i_4 \leq n} E[h(Z_{i1}, Z_{i2})h(Z_{i3}, Z_{i4})] = O(n^2)$$
which proves the lemma in the case \( c = 2 \).

The proof for arbitrary \( c \) follows the same idea. We then obtain an upper bound of

\[
2^c M n^c \sum_{k=0}^{n-1} (k + 1)^{c-1} \beta(k)^{(p-2)/p} \leq 2^{2c-1} M n^c \sum_{k=0}^{n-1} (k^{c-1} + 1) \beta(k)^{(p-2)/p}
\]

which again is \( O(n^c) \) due to our bounds on \( \beta(n) \). \( \square \)

**Theorem 3.** Suppose that \( X \) and \( Y \) are of negative type via mappings \( \phi \) and \( \psi \), respectively, and that \( X \times Y \) is \( \sigma \)-compact. If \( X \) and \( Y \) are independent, have finite \( (1 + \varepsilon) \)-moments for some \( \varepsilon > 0 \), and the sequence \( (Z_k)_{k \in \mathbb{N}} \) is absolutely regular with mixing coefficients \( \beta(n) = O(n^{-r}) \) for some \( r > 6(1 + 2\varepsilon^{-1}) \), then

\[
n \cdot \text{dcov}(\theta_n) \xrightarrow{D} n \to \infty \zeta := \sum_{k=1}^{\infty} \lambda_k \zeta_k^2
\]

where the \( \zeta_k \) are centered Gaussian random variables whose covariance function given in (5) is determined by the dependence structure of the sequence \( (Z_k)_{k \in \mathbb{N}} \), and the parameters \( \lambda_k > 0 \) are determined by the underlying distribution \( \theta \).

**Proof.** Consider the identity \( \text{dcov}(\theta_n) = V_h(Z_1, ..., Z_n) =: V \) as given in Theorem 1. We will employ Hoeffding decomposition, i.e.

\[
V = \sum_{c=0}^{6} \binom{6}{c} V_{h_c}(Z_1, ..., Z_n)
\]

where

\[
\tilde{h}_c(z_1, ..., z_c) = \sum_{A \subseteq \{1, ..., 6\}} (-1)^{6-\#A} \int \tilde{h}(z_1, ..., z_6) \ d\theta^{6-c}(z_{c+1}, ..., z_6)
\]

for \( 0 \leq c \leq 6 \). It can be readily seen that under the assumption of independence of \( X \) and \( Y \), \( \tilde{h}_1 = 0 \) almost surely, and so the Hoeffding decomposition reduces to

\[
V = \sum_{c=2}^{6} \binom{6}{c} V_{h_c}(Z_1, ..., Z_n)
\]

We will show that the kernel \( h_2 \) satisfies the conditions of Theorem 2 and that, under our assumptions,

\[
nV - nV_{h_2}(Z_1, ..., Z_n) \xrightarrow{P} n \to \infty 0
\]
Application of some algebra shows that $\bar{h}_2 = \delta_\theta/15$, proceeding in the following way:

It can be easily checked that under independence of $X$ and $Y$, $\bar{h}$ is a degenerate kernel, since integrating over all but one argument of $f$ (with respect to either of the marginal distributions of $\theta$) yields a function which is 0 almost surely. Therefore,

$$\bar{h}_2(z_1, z_2) = \frac{1}{6!} \sum_{\sigma \in S_6} \int h(z_{\sigma(1)}, ..., z_{\sigma(6)}) \, d\theta^4(z_3, ..., z_6)$$

where $S_6$ is the symmetric group of all permutations operating on $\{1, ..., 6\}$. Notice that the summands are equal to $\delta_\theta(z_{\sigma(1)}, z_{\sigma(2)})$ if $\sigma(1), \sigma(2) \in \{1, 2\}$. This follows directly from the definitions of $d_\mu$ and $d_\nu$. Moreover, 1 and 2 are the only indices appearing in both $f(X_1, ..., X_4)$ and $f(Y_1, Y_2, Y_5, Y_6)$, so any permutation $\sigma$ with $\sigma(1), \sigma(2) \notin \{1, 2\}$ results in taking the integral of $f$ over all or all but one argument, either with respect to $\mu$ or with respect to $\nu$. But we have seen before that these integrals are 0 almost surely, and so, due to the independence of $X$ and $Y$, the same is true for the integral of $h$ with respect to $\theta$.

There are $2 \cdot 4!$ permutations of this kind, and so

$$\bar{h}_2(z_1, z_2) = \frac{2 \cdot 4!}{6!} \sum_{\sigma \in S_6} \delta_\theta(z_{\sigma(1)}, z_{\sigma(2)}) = \frac{1}{15} \delta_\theta(z_1, z_2)$$

We can therefore consider the object $\delta_\theta$ instead of $\bar{h}_2$.

By identity (1) we have, for any real constants $c_1, ..., c_m$ and $z_1, ..., z_m \in X \times Y$,

$$\sum_{i,j=1}^m c_i c_j \delta_\theta(z_i, z_j) = 4 \sum_{i,j=1}^m c_i c_j (\langle \hat{\phi} \otimes \hat{\psi} \rangle(z_i), \hat{\phi} \otimes \hat{\psi})(z_j))$$

$$= 4 \left( \sum_{i=1}^m c_i (\hat{\phi} \otimes \hat{\psi})(z_i), \sum_{i=1}^m c_i (\hat{\phi} \otimes \hat{\psi})(z_i) \right)$$

$$= \left\| 2 \sum_{i=1}^m c_i (\hat{\phi} \otimes \hat{\psi})(z_i) \right\|^2 \geq 0$$

so our kernel is positive semidefinite. It is furthermore continuous. By Lemma 5, $\delta_\theta$ has finite $(2 + \varepsilon)$-moments with respect to $\theta^2$ and finite $(1 + \frac{\varepsilon}{2})$-moments on the diagonal. Since $2\alpha(n) \leq \beta(n)$ (cf. 4), we have

$$nV_{\bar{h}_2}(Z_1, ..., Z_n) \xrightarrow{D}{\nrightarrow} \sum_{k=1}^{\infty} \lambda_k \xi_k^2$$

(15)
by Theorem 2. We will now prove (14). For this, we will first note that under our assumptions, the kernel $\hat{h}$ has finite $(2 + \varepsilon)$-moments with respect to $\theta^0$. This can be seen with a similar approach as in the proof of Lemma 5. Furthermore, Lemma 3 together with the independence of $X$ and $Y$ gives us the existence of an upper bound $M \in \mathbb{R}$ such that

$$E \left[ \bar{h}(Z_1, ..., Z_{i_6})^{2+\varepsilon} \right] \leq M < \infty$$

for any collection of indices $1 \leq i_1, ..., i_6 \leq n$.

Employing Lemma 4 therefore gives us

$$E \left[ V_{\bar{h}_c}(Z_1, ..., Z_6)^2 \right] = O(n^{-c})$$

for all $c \geq 2$. Now, together with (13), we have

$$E \left[ (nV - nV_{\bar{h}_c}(Z_1, ..., Z_6))^2 \right] = E \left[ \left( n \sum_{c=3}^{6} \binom{6}{c} V_{\bar{h}_c}(Z_1, ..., Z_6)^2 \right)^2 \right] \leq 4n^2 \sum_{c=3}^{6} E \left[ V_{\bar{h}_c}(Z_1, ..., Z_6)^2 \right]$$

with

$$= \sum_{c=3}^{6} O(n^{2-c}) = O(n^{-1})$$

This implies (14), which together with (15) proves the Theorem.

Using these two results, we can generalise Corollary 2.8 from [10].

**Corollary 1.** Under the assumptions of Theorem 3, we have

$$n \frac{d\text{cov}(\theta_n)}{D(\mu_n)D(\nu_n)} \overset{D}{\underset{n \to \infty}{\to}} \sum_{k=1}^{\infty} \lambda_k \xi_k^2 =: Q$$

with $EQ = 1$. If $d\text{cov}(\theta) > 0$, i.e. $\theta$ is not the product measure of its marginal distributions $\mu$ and $\nu$, the left hand side converges to $\infty$ almost surely.

**Proof.** We have the identity $D(\mu_n) = n^{-2} \sum_{k,l=1}^{n} d(X_k, X_l)$, and thus by Lemma 2 $D(\mu_n) \overset{a.s.}{\underset{n \to \infty}{\to}} D(\mu)$. The same holds for $D(\nu_n)$, and thus the convergence in distribution follows with the Slutsky theorem. Since $D(\mu)D(\nu) = E\delta_{\theta}(Z_1, Z_1) = \sum_{k=1}^{\infty} \lambda_k$, the expected value of the limiting distribution is equal to 1.

If $d\text{cov}(\theta) > 0$, the almost sure convergence follows by Theorem 2.
Remark. It would be desirable to achieve a result similar to Theorem 3 under the assumption of just $\alpha$-mixing. For example, Theorem 3.2 in [5] gives such a result under the supposition that $X$ and $Y$ are real-valued random vectors.

For our more general setting of (pseudo-)metric spaces, one only needs to show that [14] still holds in the case of $\alpha$-mixing, since Theorem 2 does not require absolute regularity. We consider it likely that this can indeed be derived from the amicable properties of the distance covariance.

3 Generalisation to pseudometric spaces

Let $(X', d)$ be a metric space and consider $d^\beta$ for $\beta \in (0, 2]$. Then $d^\beta$ is a pseudometric, i.e. the triangle inequality does not necessarily hold for $d^\beta$. We will develop parts of the theory of [10] for pseudometric spaces of this particular kind, which we will refer to as $\beta$-pseudometric spaces. This is of interest if one considers $\text{dcov}_\beta$, a generalisation of the usual distance covariance, which results from using the $\beta$-th power of the metrics on $X$ and $Y$ for the definition of $d_\mu$ and $d_\nu$. That is, $\text{dcov}_\beta$ with respect to $(X', d)$ and $(Y, d)$ is equivalent to the regular distance covariance with respect to the $\beta$-pseudometric spaces $(X', d^\beta)$ and $(Y, d^\beta)$. Obviously, for any constant $\beta > 0$, $d^\beta$ induces the same topology (and thus, the same Borel $\sigma$-algebra) as the original metric $d$. This means that any $\beta$-pseudometric space is a metrizable topological space.

This approach of viewing $\text{dcov}_\beta$ not as a different object on the same space, but as the same object on a different space might not be very intuitive at first. However, since the concept of (strong) negative type does not require a metric space, this characterisation allows us to still use the relation between (strong) negative type of the underlying space and the distance covariance. This leads to the question of whether $(X', d^\beta)$ is of (strong) negative type, given the original metric space $(X', d)$, for which some criteria are known – see for example Corollary 3 or, more generally, [9] and [12].

Note that if $\beta \in (0, 1]$, $d^\beta$ is indeed still a metric, and we can rely on the already developed theory for separable metric spaces. Thus, we get the following result.

**Corollary 2.** Let $\beta \in (0, 1]$. Theorems 1 and 3 still hold for $\text{dcov}_\beta$ if we replace the finite first moment condition of Theorem 1 and the finite $(1 + \varepsilon)$-moment condition of Theorem 3 by finite $\beta$- and $(1 + \varepsilon)\beta$-moment assumptions, respectively.

**Proof.** Theorem 1 follows immediately. For Theorem 2, we note that $d^\beta$
induces the same Borel σ-algebra as \( d \). Furthermore, by Remark 3.19 in [10], the resulting metric spaces are still of negative type.

For \( \beta \in (1, 2) \), while we cannot rely on the triangle inequality, the Jensen inequality gives us a result which we will call the weak triangle inequality. Specifically, for any \( \beta \in [1, 2] \):

\[
d^\beta(x, x') \leq 2^{\beta-1}\{d^\beta(x, x_0) + d^\beta(x_0, x')\}
\]

(17)

for all \( x, x', x_0 \in X \). This can be further bounded by replacing the factor \( 2^{\beta-1} \) by 2.

Like in the metric case, we say that a probability measure \( \mu \) has finite first moment if there exists an element \( x_0 \in X \) such that \( \int d(x, x_0) \, d\mu(x) < \infty \). Again, the choice of \( x_0 \) is arbitrary due to the weak triangle inequality. Thus, we can define the objects \( a_\mu \), \( D(\mu) \) and \( d_\mu \) as in the metric case.

**Lemma 5.** If \( \mu \) has finite \( \beta p \)-moment, then \( d_\mu^{(\beta)} \) has finite \( 2p \)-moment with respect to \( \mu^2 \) and finite \( p \)-moment on the diagonal for any \( p \geq 1 \).

**Proof.** We take inspiration from the proof of Proposition 2.6 in [10]. Define the functions

\[
f(x_1, ..., x_4) := d^\beta(x_1, x_2) - d^\beta(x_1, x_3) - d^\beta(x_2, x_4) + d^\beta(x_3, x_4)
\]

\[
h(x_1, ..., x_6) := f(x_1, ..., x_4) f(x_1, x_2, x_5, x_6)
\]

We have

\[
f(x_1, ..., x_4) \leq 2d^\beta(x_1, x_2) - d^\beta(x_1, x_3) - d^\beta(x_2, x_4) + 2d^\beta(x_3, x_4) =: f_+
\]

and, using the weak triangle inequality, \( |f_+| \leq 4d^\beta(x_2, x_3) \). Similarly, we have

\[
f(x_1, ..., x_4) \geq d^\beta(x_1, x_2) - 2d^\beta(x_1, x_3) - 2d^\beta(x_2, x_4) + d^\beta(x_3, x_4) =: f_-
\]

Again, \( |f_-| \leq 4d^\beta(x_2, x_3) \), and thus \( |f(x_1, ..., x_4)| \leq 4d^\beta(x_2, x_3) \). In the same way, one shows that the absolute value of \( f(x_1, ..., x_4) \) can also be bounded by \( 4d^\beta(x_1, x_4) \). Therefore \( |h(x_1, ..., x_6)| \leq 16d^\beta(x_2, x_3) d^\beta(x_1, x_4) \), and so

\[
\int |d_\mu^{(\beta)}(x_1, x_2)|^{2p} \, d\mu^2(x_1, x_2) = \int \left| \int h(x_1, ..., x_6) \, d\mu^4(x_3, ..., x_6) \right|^p \, d\mu^2(x_1, x_2)
\]

\[
\leq 16^p \int d^{2p}(x_2, x_3) d^{2p}(x_1, x_4) \, d\mu^4(x_1, ..., x_4)
\]

\[
= \left( 4^p/2 \int d^{2p}(x, x') \, d^2(x, x') \right)^2 < \infty
\]
Furthermore, we have

\[
\int |d_\mu^{(\beta)}(x, x)|^p \, d\mu(x) = \int \left| \int f(x, x, x_3, \ldots, x_6) \, d\mu^2(x_3, x_4) \right|^p \, d\mu(x) \\
\leq 4^p \int d_\beta^p(x, x_3) \, d\mu^2(x, x_3) < \infty
\]

i.e. \(d_\mu^{(\beta)}\) has finite \(p\)-moment on the diagonal.

We can now define \(\delta_\theta\) and \(dcov(\theta)\) analogously to the metric case. Since the relevant proofs do not make use of the triangle inequality, it follows from [10] that for pseudometric spaces of strong negative type \(\theta = \mu \otimes \nu\) if and only if \(dcov(\theta) = 0\). This, together with the next Lemma, gives a very easy proof of Theorem 4.2 in [6].

**Lemma 6.** If \((H, \|\cdot\|)\) is a separable Hilbert space, then \((H, \|\cdot\|^\beta)\) is of negative type for all \(\beta \in (0, 2]\), and of strong negative type for all \(\beta \in (0, 2)\).

**Proof.** Without loss of generality, assume \(H\) to be equal to \(L^2[0, 1]\). By Theorem 5 in [12], for any \(\beta \in (0, 2]\), there exists an embedding \(\Phi : H \to L^2[0, 1]\) with \(\|x - x'\|_2^{\beta/2} = \|\Phi(x) - \Phi(x')\|_2\) for all \(x, x' \in H\), which implies that \((H, \|\cdot\|^\beta)\) is of negative type. By Remark 3.19 in [10] (which, along with all its auxiliary results, also holds for pseudometric spaces), the space \((H, \|\cdot\|^\beta)\) therefore has strong negative type for all \(\beta \in (0, 2)\).

We can use this Lemma to adapt Corollary 5.9 from [9].

**Corollary 3.** Let \((X, d)\) be a metric space. If there exists an isometric embedding from \(X\) into a separable Hilbert space \(H\), then \((X, d^{\beta})\) is of negative type for all \(\beta \in (0, 2]\) and of strong negative type for all \(\beta \in (0, 2)\).

**Proof.** Fix \(\beta \in (0, 2]\), and let \(\varphi : X \to L^2[0, 1]\) be an isometric embedding. By Lemma 6 \((H, \|\cdot\|^\beta)\) is of negative type via some embedding \(\Phi\), which implies that \((X, d^{\beta})\) is of negative type via \((\Phi \circ \varphi)\). If \(\beta < 2\), then \((H, \|\cdot\|^\beta)\) is of strong negative type, and so, for any two probability measures \(\mu_1, \mu_2\) on \(X\), we have that

\[
D(\mu_1 - \mu_2) = \int \|\varphi(x) - \varphi(x')\|_H^\beta \, d(\mu_1^2 - \mu_2^2)(x, x') \\
= \int_{\varphi(X)} \|x - x'\|_H^\beta \, d\left((\mu_1^\varphi)^2 - (\mu_2^\varphi)^2\right)(x, x') = D(\mu_1^\varphi - \mu_2^\varphi)
\]

where \(\mu_i^\varphi\) denotes the pushforward of \(\mu_i\) via \(\varphi\). We can extend the last integral to the entire space \(H\), because the pushforward measures vanish on


\[ \varphi(\mathcal{X})^C. \]

Using the strong negative type of \( (H, \| \cdot \|_H^\beta) \), this gives us \( \mu_1^\varphi = \mu_2^\varphi \), which implies \( \mu_1 = \mu_2 \), since \( \varphi \) is injective.

\[ \square \]

**Corollary 4.** Let \( \beta \in (1, 2) \). Then, if we replace the finite first moment condition of Theorem 1 by a finite \( \beta \)-moment assumption, Theorem 1 still holds for \( \text{dCov}_\beta \). If we furthermore assume \( \mathcal{X} \) and \( \mathcal{Y} \) to be isometrically embeddable into separable Hilbert spaces, and replace the finite \((1 + \varepsilon)\)-condition with a finite \((1 + \varepsilon)\beta\)-moment assumption, then Theorem 1 still holds for \( \text{dCov}_\beta \).

**Proof.** We first consider Theorem 1. We can replace (4) by

\[ |h(z_1, \ldots, z_6)| \leq 16d^\beta(x_2, x_3)d^\beta(y_1, y_4) \]

as we have done in the proof of Lemma 5. This changes the original bound only by constant, which does not affect the remainder of the proof.

If \( \mathcal{X} \) and \( \mathcal{Y} \) are isometrically embeddable into separable Hilbert spaces, then by Corollary 3 the spaces resulting from raising their metrics to the power \( \beta \) are of negative type. By Lemma 5 the proof of Theorem 1 still holds for \( \beta \)-pseudometric spaces. We can therefore apply Theorem 1 to the spaces \((\mathcal{X}, d^\beta)\) and \((\mathcal{Y}, d^\beta)\).

\[ \square \]

**References**

[1] **Aaronson, J., Burton, R., Dehling, H., Gilat, D., Hill, T., and Weiss, B.** Strong laws for l- and u-statistics. *Transactions of the American Mathematical Society* 348, 7 (1996), 2845–2866.

[2] **Arcones, M. A.** The law of large numbers for u-statistics under absolute regularity. *Electronic Communications in Probability* 3 (1998), 13–19.

[3] **Billingsley, P.** *Convergence of Probability Measures*, 2 ed. John Wiley & Sons, New York – Chichester – Weinheim, 1999.

[4] **Bradley, R. C.** *Introduction to Strong Mixing Conditions*, vol. 1. Kendrick Press, Heber City, 2007.

[5] **Davis, R. A., Matsui, M., Mikosch, T., and Wan, P.** Applications of distance correlation to time series. *Bernoulli* 24, 4A (2018), 3087–3116.
[6] DEHLING, H., MATSUI, M., MIKOSCH, T., SAMORODNITSKY, G., AND TAFAKORI, L. Distance covariance for discretized stochastic processes. Available at https://arxiv.org/abs/1806.09369.

[7] DEHLING, H., AND PHILIPP, W. Almost sure invariance principles for weakly dependent vector-valued random variables. *The Annals of Probability* 10, 3 (1982), 689–701.

[8] JAKOBSEN, M. E. Distance covariance in metric spaces: Non-parametric independence testing in metric spaces. Master's thesis, University of Copenhagen, 2017. Available at https://arxiv.org/abs/1706.03490.

[9] LI, H., AND WESTON, A. Strict p-negative type of a metric space. *Positivity* 14, 3 (2010), 529–545.

[10] LYONS, R. Distance covariance in metric spaces. *The Annals of Probability* 41, 5 (2013), 3284–3305.

[11] LYONS, R. Errata to distance covariance in metric spaces. *The Annals of Probability* 46, 4 (2018), 2400–2405.

[12] SCHOENBERG, I. J. Metric spaces and positive definite functions. *Transactions of the American Mathematical Society* 44 (1938), 522–536.

[13] SOTRES, D. A., AND GHOSH, M. Strong convergence of linear rank statistics for mixing processes. *The Indian Journal of Statistics* 39 (1977), 1–11.

[14] SUN, H. Mercer theorem for rkhs on noncompact sets. *Journal of Complexity* 214 (2005), 337–349.

[15] SZEKELY, G. J., RIZZO, M. L., AND BAKIROV, N. K. Measuring and testing dependence by correlation of distances. *Annals of Statistics* 35, 6 (2007), 2769–2794.

[16] YOSHIHARA, K. Limiting behavior of u-statistics for stationary, absolutely regular processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 35 (1976), 237–252.