Distance covariance for discretized stochastic processes

HEROLD DEHLING[1], MUNEYA MATSUI[2], THOMAS MIKOSCH[3],
GENNADY SAMORODNITSKY[4] and LALEH TAFAKORI[5]

[1]*Department of Mathematics, Ruhr-Universität Bochum, 44780 Bochum, Germany.*
E-mail: herold.dehling@rub.de
[2]*Department of Business Administration, Nanzan University, 18 Yamazato-cho, Showa-ku, Nagoya 466-8673, Japan. E-mail: mmuneya@gmail.com*
[3]*Department of Mathematics, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark. E-mail: mikosch@math.ku.dk*
[4]*School of Operations Research and Information Engineering, Cornell University, 220 Rhodes Hall, Ithaca, NY 14853, USA. E-mail: gs18@cornell.edu*
[5]*Department of Mathematical Sciences, School of Science, RMIT University, Melbourne, VIC, Australia.*
E-mail: laleh.tafakori@rmit.edu.au

Given an i.i.d. sequence of pairs of stochastic processes on the unit interval we construct a measure of independence for the components of the pairs. We define distance covariance and distance correlation based on approximations of the component processes at finitely many discretization points. Assuming that the mesh of the discretization converges to zero as a suitable function of the sample size, we show that the sample distance covariance and correlation converge to limits which are zero if and only if the component processes are independent. To construct a test for independence of the discretized component processes, we show consistency of the bootstrap for the corresponding sample distance covariance/correlation.

Keywords: distance covariance; empirical characteristic function; stochastic process; test of independence

1. Introduction

1.1. Distance covariance and distance correlation for vectors

In a series of papers, Székely et al. [22] and Székely and Rizzo [19–21] introduced *distance covariance* and *distance correlation*. They are measures of the dependence between two vectors **X** and **Y**, possibly with different dimensions. These measures have the desirable property that they are zero if and only if **X** and **Y** are independent. This is in contrast to many other dependence measures where one can only make statements about certain aspects of the dependence between **X** and **Y**. For example, the correlation and covariance between two real-valued random variables X and Y allow one to make statements about their linear dependence.

The distance covariance between a p-dimensional vector **X** and a q-dimensional vector **Y** is a weighted version of the squared distance between the joint characteristic function $\varphi_{\mathbf{X},\mathbf{Y}}$ of **X**, **Y**

and the product of the marginal characteristic functions $\varphi_{\mathbf{X}}$, $\varphi_{\mathbf{Y}}$ of these vectors. We know that $\mathbf{X}$ and $\mathbf{Y}$ are independent if and only if

$$\varphi_{\mathbf{X},\mathbf{Y}}(\mathbf{s},\mathbf{t}) = \varphi_{\mathbf{X}}(\mathbf{s})\varphi_{\mathbf{Y}}(\mathbf{t}), \quad \mathbf{s} \in \mathbb{R}^p, \mathbf{t} \in \mathbb{R}^q. \tag{1.1}$$

However, this identity is difficult to check if one has data at the disposal; a replacement of the corresponding characteristic functions by empirical versions does not lead to powerful statistical tools for detecting independence between $\mathbf{X}$ and $\mathbf{Y}$. First, Feuerverger [9] in the univariate case and, later, Székely et al. [22] and Székely and Rizzo [19–21] in the general multivariate case recommended to use a weighted L^2-distance between $\varphi_{\mathbf{X},\mathbf{Y}}$ and $\varphi_{\mathbf{X}}\varphi_{\mathbf{Y}}$: for $\beta \in (0,2)$, the *distance covariance between X and Y* is given by

$$T_\beta(\mathbf{X},\mathbf{Y}) = c_p c_q \int_{\mathbb{R}^{p+q}} \left|\varphi_{\mathbf{X},\mathbf{Y}}(\mathbf{s},\mathbf{t}) - \varphi_{\mathbf{X}}(\mathbf{s})\varphi_{\mathbf{Y}}(\mathbf{t})\right|^2 |\mathbf{s}|^{-(p+\beta)}|\mathbf{t}|^{-(q+\beta)}\, d\mathbf{s}\, d\mathbf{t},$$

where the constants c_d for $d \geq 1$ are chosen such that

$$c_d \int_{\mathbb{R}^d} \left(1 - \cos(\mathbf{s}'\mathbf{x})\right)|\mathbf{x}|^{-(d+\beta)}\, d\mathbf{x} = |\mathbf{s}|^\beta.$$

Here and in what follows we suppress the dependence of the Euclidean norm $|\cdot|$ on the dimension; it will always be clear from the context what the dimension is. The quantity $T_\beta(\mathbf{X},\mathbf{Y})$ is finite under suitable moment conditions on $\mathbf{X},\mathbf{Y}$. The corresponding *distance correlation* is given by

$$R_\beta(\mathbf{X},\mathbf{Y}) = \frac{T_\beta(\mathbf{X},\mathbf{Y})}{\sqrt{T_\beta(\mathbf{X},\mathbf{X})}\sqrt{T_\beta(\mathbf{Y},\mathbf{Y})}}.$$

An advantage of choosing the particular weight function $|\mathbf{s}|^{-(p+\beta)}|\mathbf{t}|^{-(q+\beta)}$ is that the distance covariance has an explicit form: for i.i.d. copies $(\mathbf{X}_i,\mathbf{Y}_i)$, $i = 1,2,\ldots,$ of $(\mathbf{X},\mathbf{Y})$ we have

$$T_\beta(\mathbf{X},\mathbf{Y}) = \mathbb{E}\left[|\mathbf{X}_1 - \mathbf{X}_2|^\beta |\mathbf{Y}_1 - \mathbf{Y}_2|^\beta\right] + \mathbb{E}\left[|\mathbf{X}_1 - \mathbf{X}_2|^\beta\right]\mathbb{E}\left[|\mathbf{Y}_1 - \mathbf{Y}_2|^\beta\right]$$
$$- 2\mathbb{E}\left[|\mathbf{X}_1 - \mathbf{X}_2|^\beta |\mathbf{Y}_1 - \mathbf{Y}_3|^\beta\right]. \tag{1.2}$$

The weight function ensures that $T_\beta(c\mathbf{X}, c\mathbf{Y}) = c^{2\beta} T_\beta(\mathbf{X},\mathbf{Y})$ for any constant c, hence $R_\beta(c\mathbf{X}, c\mathbf{Y})$ does not depend on c, that is, the distance correlation is scale invariant. A corresponding theory can be built on non-homogeneous kernels as well; see the discussion and references in Davis et al. [4] who consider auto- and cross-distance correlation functions for time series.

It is clear from the construction that $T_\beta(\mathbf{X},\mathbf{Y}) = R_\beta(\mathbf{X},\mathbf{Y}) = 0$ if and only if (1.1) holds. This observation motivates the construction of sample versions of $T_\beta(\mathbf{X},\mathbf{Y})$ and $R_\beta(\mathbf{X},\mathbf{Y})$ and one hopes that these have properties similar to their deterministic counterparts. In particular, one would like to test independence between $\mathbf{X}$ and $\mathbf{Y}$.

Replacing the characteristic functions in $T_\beta(\mathbf{X}, \mathbf{Y})$ and $R_\beta(\mathbf{X}, \mathbf{Y})$ by their sample analogs and taking into account (1.2), we obtain the sample versions of $T_\beta(\mathbf{X}, \mathbf{Y})$ and $R_\beta(\mathbf{X}, \mathbf{Y})$:

$$T_{n,\beta}(\mathbf{X}, \mathbf{Y}) = \frac{1}{n^2} \sum_{k,l=1}^{n} |\mathbf{X}_k - \mathbf{X}_l|^\beta |\mathbf{Y}_k - \mathbf{Y}_l|^\beta$$

$$+ \frac{1}{n^2} \sum_{k,l=1}^{n} |\mathbf{X}_k - \mathbf{X}_l|^\beta \frac{1}{n^2} \sum_{k,l=1}^{n} |\mathbf{Y}_k - \mathbf{Y}_l|^\beta$$

$$- 2\frac{1}{n^3} \sum_{k,l,m=1}^{n} |\mathbf{X}_k - \mathbf{X}_l|^\beta |\mathbf{Y}_k - \mathbf{Y}_m|^\beta,$$

$$R_{n,\beta}(\mathbf{X}, \mathbf{Y}) = \frac{T_{n,\beta}(\mathbf{X}, \mathbf{Y})}{\sqrt{T_{n,\beta}(\mathbf{X}, \mathbf{X})}\sqrt{T_{n,\beta}(\mathbf{Y}, \mathbf{Y})}}.$$

The quantity $T_{n,\beta}(\mathbf{X}, \mathbf{Y})$ is a V-statistic; cf. Székely et al. [22] and Lyons [12]. Therefore standard theory yields a.s. consistency,

$$T_{n,\beta}(\mathbf{X}, \mathbf{Y}) \stackrel{\text{a.s.}}{\to} T_\beta(\mathbf{X}, \mathbf{Y}), \quad n \to \infty,$$

under suitable moment conditions; see Hoffmann-Jørgensen [10] and Serfling [18]. If $\mathbf{X}$ and $\mathbf{Y}$ are independent the V-statistic $T_{n,\beta}(\mathbf{X}, \mathbf{Y})$ is degenerate of order 1. Under suitable moment conditions, one also has the weak convergence of $nT_{n,\beta}(\mathbf{X}, \mathbf{Y})$ to a weighted sum of i.i.d. χ^2-variables; see Serfling [18], Lyons [12] and Arcones and Giné [1]. Moreover, V-statistics theory also ensures that $T_{n,\beta}(\mathbf{X}, \mathbf{X}) \stackrel{\text{a.s.}}{\to} T_\beta(\mathbf{X}, \mathbf{X})$ and $T_{n,\beta}(\mathbf{Y}, \mathbf{Y}) \stackrel{\text{a.s.}}{\to} T_\beta(\mathbf{Y}, \mathbf{Y})$. Hence, $R_{n,\beta}(\mathbf{X}, \mathbf{Y})$ is an a.s. consistent estimator of $R_\beta(\mathbf{X}, \mathbf{Y})$ and, modulo a change of scale, $nR_{n,\beta}(\mathbf{X}, \mathbf{Y})$ has the same weak limit as $T_{n,\beta}(\mathbf{X}, \mathbf{Y})$.

1.2. Distance covariance and distance correlation for stochastic processes

Székely and Rizzo [20] considered the situation when $\mathbf{X}$ and $\mathbf{Y}$ are independent and have i.i.d. components, n is fixed, $p = q \to \infty$. Under these conditions, $R_{n,\beta}(\mathbf{X}, \mathbf{Y})$ converges to 1. In this way, they justified the empirical observation that $R_{n,\beta}(\mathbf{X}, \mathbf{Y})$ is close to 1 if p, q are large relative to n.

Matsui et al. [15] considered a version of the distance covariance for stochastic processes X, Y on $[0, 1]$, where it was assumed that the two processes are observed at a Poisson number of points in $[0, 1]$ and the Poisson intensity is fixed. Via simulations the resulting estimator was compared with the distance correlation $R_{n,\beta}(\mathbf{X}, \mathbf{Y})$ where the components of the i.i.d. vectors $(\mathbf{X}_i, \mathbf{Y}_i)$ consist of a Poisson number of the discretizations of (X_i, Y_i), respectively. Both types of estimators exhibited a similar behavior for independent X and Y, approaching zero for moderate sizes n, p, q. A possible explanation for this phenomenon is that Matsui et al. [15] and Székely and Rizzo [20] worked under quite distinct conditions. Székely and Rizzo [20] considered vectors $\mathbf{X}$ and $\mathbf{Y}$ with i.i.d. components whose dimensions increase to infinity for a fixed sample size n. In

Matsui et al. [15], $\mathbf{X}$ and $\mathbf{Y}$ can be understood as vectors of discretizations of genuine stochastic processes X, Y on $[0, 1]$, such as Brownian motion, fractional Brownian motion, Lévy processes, etc. In these cases, the components of $\mathbf{X}_i$ and $\mathbf{Y}_i$ are dependent.

In this paper, we again take up the theme of Székely and Rizzo [20] and Matsui et al. [15]. We consider two processes X and Y on $[0, 1]$, which we assume to be stochastically continuous, measurable and bounded. In contrast to Matsui et al. [15],

- we consider discretizations of these processes at a partition $0 = t_0 < t_1 < \cdots < t_p = 1$ of $[0, 1]$, assuming that $p = p_n \to \infty$ as $n \to \infty$ and the mesh satisfies

$$\delta_n = \max_{i=1,\ldots,p} (t_i - t_{i-1}) \to 0, \quad n \to \infty,$$

- we normalize the points $X(t_i)$ and $Y(t_i)$ by $\sqrt{t_i - t_{i-1}}$.

In the sequel, we suppress the dependence of p on n. It will be convenient to write for any partition (t_i) and a process Z on $[0, 1]$,

$$\Delta_i = (t_{i-1}, t_i], \qquad |\Delta_i| = t_i - t_{i-1}, \quad i = 1, \ldots, p, \qquad \Delta Z(s, t] = Z(t) - Z(s), \quad s < t.$$

We consider a vector of weighted discretizations

$$\mathbf{Z}_p = \left(|\Delta_1|^{1/2} Z(t_1), \ldots, |\Delta_p|^{1/2} Z(t_p) \right), \tag{1.3}$$

and define

$$Z^{(p)}(t) = \sum_{i=1}^{p} Z(t_i) \mathbf{1}(t \in \Delta_i), \quad t \in [0, 1].$$

For stochastically continuous, measurable and bounded processes Z and Z' we have

$$\left| \mathbf{Z}_p - \mathbf{Z}'_p \right|^2 = \sum_{i=1}^{p} \left(Z(t_i) - Z'(t_i) \right)^2 |\Delta_i| = \left\| Z^{(p)} - \left(Z' \right)^{(p)} \right\|_2^2$$

$$\to \int_0^1 \left(Z(t) - Z'(t) \right)^2 dt = \left\| Z - Z' \right\|_2^2, \quad p \to \infty,$$

in probability, where $\|\xi\|_2$ denotes the L^2-norm of a process ξ on $[0, 1]$.

For $\beta \in (0, 2]$, we introduce a stochastic process analog $T_\beta(X, Y)$ of $T_\beta(\mathbf{X}, \mathbf{Y})$ from (1.2). Consider an i.i.d. sequence (X_i, Y_i), $i = 1, 2, \ldots$, of processes X_i, Y_i on $[0, 1]$ with generic element (X, Y) which is also stochastically continuous, measurable and bounded. Define

$$T_\beta(X, Y) = \mathbb{E}\left[\|X_1 - X_2\|_2^\beta \|Y_1 - Y_2\|_2^\beta \right] + \mathbb{E}\left[\|X_1 - X_2\|_2^\beta \right] \mathbb{E}\left[\|Y_1 - Y_2\|_2^\beta \right]$$

$$- 2\mathbb{E}\left[\|X_1 - X_2\|_2^\beta \|Y_1 - Y_3\|_2^\beta \right], \tag{1.4}$$

where we assume that all moments involved are finite. Of course, $T_\beta(X, Y) = 0$ for independent X, Y. The converse is not obvious; we prove it in Section 4.

The sample analog of $T_\beta(X, Y)$ is given by

$$
\begin{aligned}
T_{n,\beta}(X, Y) &= \frac{1}{n^2} \sum_{k,l=1}^{n} \|X_k - X_l\|_2^\beta \|Y_k - Y_l\|_2^\beta \\
&\quad + \frac{1}{n^2} \sum_{k,l=1}^{n} \|X_k - X_l\|_2^\beta \frac{1}{n^2} \sum_{k,l=1}^{n} \|Y_k - Y_l\|_2^\beta \\
&\quad - 2\frac{1}{n^3} \sum_{k,l,m=1}^{n} \|X_k - X_l\|_2^\beta \|Y_k - Y_m\|_2^\beta \\
&=: I_1 + I_3 - 2I_2.
\end{aligned}
\tag{1.5}
$$

Assuming that the moments in $T_\beta(X, Y)$ are finite, the strong law of large numbers for V-statistics yields

$$
T_{n,\beta}(X, Y) \xrightarrow{\text{a.s.}} T_\beta(X, Y), \quad n \to \infty.
$$

This fact and the observation that $T_\beta(X, Y)$ vanishes for independent X, Y encourage one to call $T_\beta(X, Y)$ the *distance covariance between* X, Y, and $T_{n,\beta}(X, Y)$ its sample version. The corresponding distance and sample distance correlations $R_\beta(X, Y)$ and $R_{n,\beta}(X, Y)$ are defined in the natural way.

1.3. Objectives

We assume that for any $i \geq 1$, we are given a pair of stochastic processes $X_i = (X_i(t))_{0 \leq t \leq 1}$ and $Y_i = (Y_i(t))_{0 \leq t \leq 1}$, and that $(X_i, Y_i)_{i \geq 1}$ are mutually independent pairs. We want to test the hypothesis that the marginals X and Y are independent against the alternative that the two processes are dependent. As an example consider daily measurement curves of the concentration of nitrogen oxide (NO_x) pollutants taken at two different stations in neighboring cities. NO_x data have been widely studied in the literature on functional data analysis, see Horváth and Kokoszka [11] for detailed references. Rescale time in such a way that $[0, 1]$ corresponds to a full day. There is day-to-day variation in the NO_x curves. Dependence of the processes would indicate that there is some common factor behind the nitrogen oxide concentration at the two sites, e.g. air pressure and temperature. As a second example, imagine that the coastline of a country (like the Netherlands) can be mapped to the interval $[0, 1]$ and, at each location $s \in [0, 1]$ and on each day i, we have an observation of the height of sea waves, X_i, and the corresponding wind-speed, Y_i. An interesting question is whether the processes X_i and Y_i are independent. The coastline example was the starting point of functional extreme value analysis initiated by de Haan and co-workers; see de Haan and Lin [6] and Part III in the monograph by de Haan and Ferreira [5]. The latter example also indicates that the i.i.d. assumption on $(X_i, Y_i)_{i \geq 1}$ may not be satisfied in a perfect way, as for many other examples of functional time series data. This may be the starting point for further investigations on distance correlation for serially dependent data.

Typically, we will not have complete sample paths of (X_i, Y_i) at our disposal. In this paper, we assume that we observe a sample $((X_i^{(p)}, Y_i^{(p)}))_{i=1,\ldots,n}$ consisting of discretizations taken from an i.i.d. sequence $((X_i, Y_i))_{i=1,2,\ldots}$ on the same partition $(t_i)_{i=0,\ldots,p}$ of $[0, 1]$. We can define the corresponding sample distance covariance $T_{n,\beta}(X^{(p)}, Y^{(p)})$ and sample distance correlation $R_{n,\beta}(X^{(p)}, Y^{(p)})$. In view of the discussion above, we see that the latter quantities coincide with the corresponding quantities $T_{n,\beta}(\mathbf{X}_p, \mathbf{Y}_p)$ and $R_{n,\beta}(\mathbf{X}_p, \mathbf{Y}_p)$ where $\mathbf{X}_p$, and $\mathbf{Y}_p$ are defined through (1.3). In the case of an equidistant partition with mesh $\delta_n = 1/p$ we also observe that $R_{n,\beta}(\mathbf{X}_p, \mathbf{Y}_p)$ is exactly the classical sample distance correlation $R_{n,\beta}(\mathbf{X}, \mathbf{Y})$ of the vectors $\mathbf{X} = (X(j/p))_{j=1,\ldots,p}$ and $\mathbf{Y} = (Y(j/p))_{j=1,\ldots,p}$.

The main goal of this paper is to show that for independent X, Y,

$$n\big(T_{n,\beta}\big(X^{(p)}, Y^{(p)}\big) - T_{n,\beta}(X, Y)\big) \xrightarrow{\mathbb{P}} 0, \quad n \to \infty, \tag{1.6}$$

provided $\delta_n \to 0$ and $p = p_n \to \infty$ sufficiently fast. In turn, we will be able to exploit the existing limit theory for the normalized degenerate V-statistic $nT_{n,\beta}(X, Y)$ to derive the distributional limit of $nT_{n,\beta}(X^{(p)}, Y^{(p)})$. This limit has a weighted χ^2-distribution which is not easily evaluated. We will show that bootstrap versions of the degenerate V-statistics $nT_{n,\beta}(X, Y)$ and $nT_{n,\beta}(X^{(p)}, Y^{(p)})$ are close in the sense of Mallows metrics and have the same distributional limit as $nT_{n,\beta}(X, Y)$.

The paper is organized as follows. In Section 2, we introduce various technical conditions and discuss their applicability to some classes of stochastic processes. The main results of Theorem 3.1 yield sufficient conditions for (1.6) and the corresponding versions for the distance correlations, assuming independence between X, Y. The proof is given in Section 7 while some technical details are omitted and given in the Supplementary Material [7]. The bootstrap for $T_{n,\beta}(X^{(p)}, Y^{(p)})$ is discussed in Section 5. There we show that a suitable bootstrap version of $T_{n,\beta}(X^{(p)}, Y^{(p)})$ is consistent. The results of Section 4 may be of independent interest. There we show that $T_\beta(X, Y) = 0$ implies independence of the integrals $\int X \, dB_1$ and $\int Y \, dB_2$ conditional on $B = (B_1, B_2)$ which has independent Brownian motion components on $[0, 1]$ and is independent of (X, Y). In turn, the conditional independence of these integrals implies independence of X, Y. We give a small simulation study in Section 6 which shows that the theoretical results work for small and moderate values of n and p.

2. Technical conditions

To derive the results in Section 3, we assume various conditions on the smoothness and moments of the processes X, Y and their relation with the parameters of the partition, in particular p and δ_n. Throughout $\beta \in (0, 2)$ is fixed. If any of the processes X, Y have finite expectation we assume that they are centered.

We will work under two distinct settings: (1) finite variance of X, Y and (2) X, Y have finite βth moment.

2.1. The finite variance case

If X, Y have finite second moments we will work under the set of conditions (A):

(A1) *Smoothness of increments.* There exist $\gamma_X, \gamma_Y > 0$ and $c > 0$ such that

$$\text{var}\big(\Delta X(s,t]\big) \le c|t-s|^{\gamma_X} \quad \text{and} \quad \text{var}\big(\Delta Y(s,t]\big) \le c|t-s|^{\gamma_Y}, \quad s < t.$$

(A2) *Growth condition on $p = p_n \to \infty$.* We have

$$\delta_n = o\big(n^{-2/((\gamma_X \wedge \gamma_Y)(\beta \wedge 1))}\big), \quad n \to \infty.$$

(A3) *Additional moment conditions.* If $\beta \in (1,2)$, we have

$$\max_{0 \le t \le 1} \mathbb{E}\big[|X(t)|^{2(2\beta-1)}\big] + \max_{0 \le t \le 1} \mathbb{E}\big[|Y(t)|^{2(2\beta-1)}\big] < \infty.$$

2.2. The finite βth moment case

If X, Y possibly have infinite second moments we will work under the set of conditions (B):

(B1) *Finite βth moment.*

$$\mathbb{E}\Big[\max_{t \in (0,1]} |X(t)|^\beta\Big] < \infty \quad \text{and} \quad \mathbb{E}\Big[\max_{t \in (0,1]} |Y(t)|^\beta\Big] < \infty.$$

(B2) *Smoothness of increments.* There exist $\gamma_X, \gamma_Y > 0$ and $c > 0$ such that

$$\max_{i=1,\dots,p} \mathbb{E}\Big[\max_{t \in \Delta_i} |\Delta X(t,t_i]|^\beta\Big] \le c\delta_n^{\gamma_X} \quad \text{and} \quad \max_{i=1,\dots,p} \mathbb{E}\Big[\max_{t \in \Delta_i} |\Delta Y(t,t_i]|^\beta\Big] \le c\delta_n^{\gamma_Y}.$$

(B3) *Additional moment and smoothness conditions.* If $\beta \in (0,1)$, we also have

$$\mathbb{E}\Big[\max_{0 \le t \le 1} |X(t)|^{2\beta}\Big] < \infty \quad \text{and} \quad \mathbb{E}\Big[\max_{0 \le t \le 1} |Y(t)|^{2\beta}\Big] < \infty,$$

and there exist $\gamma_X', \gamma_Y' > 0$ and $c > 0$ such that

$$\max_{i=1,\dots,p} \mathbb{E}\Big[\max_{t \in \Delta_i} |\Delta X(t,t_i]|^{2\beta}\Big] \le c\delta_n^{\gamma_X'} \quad \text{and} \quad \max_{i=1,\dots,p} \mathbb{E}\Big[\max_{t \in \Delta_i} |\Delta Y(t,t_i]|^{2\beta}\Big] \le c\delta_n^{\gamma_Y'}.$$

(B4) *Growth condition on $p = p_n \to \infty$.* We have

$$\delta_n = o\big((pn^{\beta/(\beta \wedge 1)})^{-\frac{1}{\beta/2+\gamma_X \wedge \gamma_Y}}\big).$$

2.3. Discussion of the conditions and examples

Remark 2.1. In the proofs, we will need the conditions

$$\mathbb{E}\big[\|X\|_2^\beta\big] < \infty \quad \text{and} \quad \mathbb{E}\big[\|Y\|_2^\beta\big] < \infty \quad \text{for some } \beta \in (0,2). \tag{2.1}$$

If (A1) holds (in particular, $\sup_{t \in [0,1]}[\text{var}(X(t)) + \text{var}(Y(t))] < \infty$) (2.1) is automatic because by Jensen's inequality

$$\mathbb{E}\big[\|X\|_2^\beta\big] = \mathbb{E}\Bigg[\bigg(\int_0^1 (X(t))^2 \, dt\bigg)^{\beta/2}\Bigg] \le \bigg(\int_0^1 \text{var}(X(t)) \, dt\bigg)^{\beta/2} < \infty.$$

The same argument also shows that $\mathbb{E}[\|X\|_2^2] < \infty$ under (A1). If (B1) holds then (2.1) follows.

Remark 2.2. The procedures developed in this paper depend on the mesh δ_n, respectively the dimension $p = 1/\delta_n$ in the equidistant case, and the exponent β. We will now discuss briefly some aspects of the choice of these parameters. In practice, the mesh will most often be given by the data, which are observations of a stochastic process at given discrete times, and thus there is no choice for the statistician. In those cases when we could, at least in principle, observe the processes at all times $t \in [0, 1]$, the choice of the mesh becomes a task for the statistician. Several aspects have to be taken into account. A fine mesh, and thus a large value of p, will increase the numerical burden involved in calculating the test statistic and performing the bootstrap. On the other hand, one could expect the power to increase as the mesh becomes finer, though we do not have a rigorous proof. Our simulations indicate that a moderate size of p, say $p = 100$, is a reasonable choice, and that, e.g. $p = 10,000$ does not yield a more powerful test. The choice of β is related to the moment assumptions. Our theoretical results require the existence of the (2β)th moments, and this requirement will put a restriction on the choice of β in the case of very heavy tailed processes. Beyond such restrictions, it is an open question which choice of β yields the most powerful test. We suspect that the answer will depend on the specific choice of the alternative.

Example 2.3. Assume that X, Y are sample continuous self-similar processes with stationary increments and a finite variance. If the corresponding Hurst exponents are H_X, $H_Y \in (0, 1)$, then for some $c_X > 0$,

$$\text{var}\big(\Delta X(s, t]\big) = \text{var}\big(X(0, t - s]\big) = c_X(t - s)^{2H_X}, \quad s < t,$$

and similarly for Y. That is, we can choose $\gamma_X = 2H_X$ and $\gamma_Y = 2H_Y$ in (A1). Furthermore, (A3) holds for X if $\beta \in (1, 2)$ and $\mathbb{E}[|X(1)|^{2(2\beta-1)}] < \infty$, and similarly for Y. A special case is that of Gaussian X and Y which then are *fractional Brownian motions*, and (A3) trivially holds. A process with the same covariance structure is the *fractional Lévy process*

$$X(t) = \int_{\mathbb{R}} \big((t - s)_+^{H_X - 0.5} - (-s)_+^{H_X - 0.5}\big) \, dL(s), \quad t \in \mathbb{R}, H_X \in (0.5, 1),$$

where L is a two-sided Lévy process on $\mathbb{R}$ with mean zero and finite variance, introduced in Marquardt [14]. This process is not self-similar (unless L is a Brownian motion) but has stationary increments. Here (A1) holds with $\gamma_X = 2H_X$ and $\gamma_Y = 2H_Y$. Furthermore, (A3) holds if $\mathbb{E}[|L(1)|^{2(2\beta-1)}] < \infty$.

Notice also that *any centered Gaussian processes X and Y* satisfying (A1) have automatically continuous sample paths and (A3) is satisfied.

Example 2.4. Assume that X and Y are Itô integrals, that is, there are two Brownian motions B_X, B_Y and predictable processes Z_X, Z_Y with respect to the corresponding Brownian filtrations such that

$$X(t) = \int_0^t Z_X(s) \, dB_X(s), \qquad Y(t) = \int_0^t Z_Y(s) \, dB_Y(s), \quad 0 \le t \le 1.$$

Then we have

$$\mathrm{var}\big(\Delta X(s,t)\big) = \int_s^t \mathbb{E}\big[Z_X^2(x)\big] \, dx, \quad s < t.$$

Hence, if $c_X = \sup_{x \in [0,1]} \mathbb{E}[Z_X^2(x)] < \infty$, then $\mathrm{var}(\Delta X(s,t)) \le c_X(t-s)$, and one can choose $\gamma_X = 1$ in (A1). Moreover, (A3) holds for X if $\beta \in (1,2)$ and $\mathbb{E}[|X(1)|^{2(2\beta-1)}] < \infty$. This follows from an application of Doob's maximal inequality for martingales. Similar arguments apply to the process Y. A special case is that of zero drift geometric Brownian motions; a simple computation shows that nothing changes even when the drift is not zero.

In the equidistant case, and when $\gamma_X = \gamma_Y = 1$, (A2) holds if

$$\frac{p}{n^{\frac{2}{\beta \wedge 1}}} \to \infty, \quad n \to \infty. \tag{2.2}$$

Example 2.5. For $\alpha \in (0,2)$ sample continuous self-similar SαS processes with stationary increments provide a family of examples with an infinite second moment. For such processes (B1) is satisfied for $\beta < \alpha$ and (B2) is satisfied with $\gamma_X = \gamma_Y = \beta H$, where H is the Hurst exponent. This follows from continuity, self-similarity and stationarity of the increments. Similarly, (B3) holds if $\beta < \alpha/2$ and $\gamma_X' = \gamma_Y' = 2\beta H$. Such processes include the *fractional harmonizable α-stable motions* and, if $1 < \alpha < 2$ and $1/\alpha < H < 1$, also the *linear fractional stable motions*; see Chapter 7 in Samorodnitsky and Taqqu [17]. Another example is that of the γ-*Mittag Leffler fractional SαS motion*, which is an integral of a γ-Mittag Leffler process with respect to a suitable SαS random measure; see Samorodnitsky [16], Section 8.4. Here $H = \gamma + (1-\gamma)/\alpha$.

Example 2.6. Lévy processes are stochastically continuous and bounded by definition. If X is a Lévy process with finite second moment (A1) holds because $\mathrm{var}(\Delta X(s,t)) = c(t-s)$, for $s < t$ and a constant c. Moreover, (A3) holds for X if $\mathbb{E}[|X(1)|^{2(2\beta-1)}] < \infty$. Indeed, an application of Lévy's maximal inequality yields for $t \in [0,1]$,

$$\mathbb{E}\big[|X(t)|^{2(2\beta-1)}\big] \le \mathbb{E}\Big[\max_{0 \le t \le 1} |X(t)|^{2(2\beta-1)}\Big] \le c\,\mathbb{E}\big[|X(1)|^{2(2\beta-1)}\big].$$

Similarly, for X, (B1) holds if $\mathbb{E}[|X(1)|^\beta] < \infty$, (B2) is satisfied if $\mathbb{E}[|\Delta X(s,t]|^\beta] \le c(t-s)^{\gamma_X}$, and (B3) holds if $\mathbb{E}[|\Delta X(s,t]|^{2\beta}] \le c(t-s)^{\gamma_X'}$.

3. Main results

We would like to use the distance covariance to test for independence of two stochastically continuous bounded stochastic processes X, Y on $[0,1]$. By the strong law of large numbers for

V-statistics we have

$$T_{n,\beta}(X, Y) \overset{\text{a.s.}}{\to} T_\beta(X, Y), \tag{3.1}$$

where the limit is defined in (1.4). If X, Y are independent then $T_\beta(X, Y) = 0$, and in Section 4 we prove that, conversely, $T_\beta(X, Y) = 0$ implies independence of X, Y. The following theorem establishes, in particular, that under appropriate conditions, if X, Y are independent, then also

$$T_{n,\beta}(X^{(p)}, Y^{(p)}) - T_{n,\beta}(X, Y) \overset{\mathbb{P}}{\to} 0 \tag{3.2}$$

and, hence,

$$T_{n,\beta}(X^{(p)}, Y^{(p)}) \overset{\mathbb{P}}{\to} 0. \tag{3.3}$$

This relation can be used in testing for independence of X, Y. Note that, if X, Y are dependent the results of Section 4 will imply that $T_\beta(X, Y) > 0$ and so, by (3.1) and (3.2), we see that $n T_{n,\beta}(X^{(p)}, Y^{(p)}) \overset{\mathbb{P}}{\to} \infty$.

In fact, the limiting equivalence (3.2) holds for dependent X, Y as well, see the proof of Lemma S.1 in the Supplementary Material [7], as long as one imposes more restrictive moment conditions (due to the use of Hölder-type inequalities for products of dependent random variables).

In the theorem below we assume, without loss of generality, that $\mathbb{E}[X(t)] = \mathbb{E}[Y(t)] = 0$ for any $t \in [0, 1]$, provided the expectations are finite. Indeed, $T_{n,\beta}$ contains expressions of the type $X_k - X_l, Y_k - Y_l$ or their discrete approximations. Therefore, we can always mean-correct X_k and Y_k, without changing the value of $T_{n,\beta}$.

Theorem 3.1. *Assume the following conditions*:

1. *X, Y are independent stochastically continuous bounded processes on $[0, 1]$ defined on the same probability space.*
2. *If X, Y have finite expectations, then these are assumed to be equal to 0.*
3. *$\delta_n \to 0$ as $n \to \infty$.*
4. *$\beta \in (0, 2)$.*

Then the following statements hold.

(1) *If either (A1) or [(B1), (B2) and $p\delta_n^{\beta/2 + \gamma_X \wedge \gamma_Y} \to 0$] are satisfied then (3.2) (and, hence, (3.3)) hold.*

(2) *If either (A1), (A2) or (B1), (B2), (B4) hold, then*

$$n T_{n,\beta}(X^{(p)}, Y^{(p)}) \overset{d}{\to} \sum_{i=1}^{\infty} \lambda_i (N_i^2 - 1) + c$$

for an i.i.d. sequence of standard normal random variables (N_i), a constant c, and a square summable sequence (λ_i).

(3) *If either* (A1), (A3) *or* [$\beta \in (0, 1)$ *and* (B1)–(B3) *and* $p\delta_n^{\beta+\gamma'_X \wedge \gamma'_Y} \to 0$] *hold then*

$$R_{n,\beta}\big(X^{(p)}, Y^{(p)}\big) \overset{\mathbb{P}}{\to} 0.$$

(4) *If either* (A1)–(A3) *or* [$\beta \in (0, 1)$ *and* (B1)–(B4) *and* $p\delta_n^{\beta+\gamma'_X \wedge \gamma'_Y} \to 0$] *hold then*

$$n R_{n,\beta}\big(X^{(p)}, Y^{(p)}\big) \overset{d}{\to} \sum_{i=1}^{\infty} \lambda_i \big(N_i^2 - 1\big) + c$$

for an i.i.d. sequence of standard normal random variables (N_i), *a constant* c, *and a square summable sequence* (λ_i).

The proof is given in Section 7 while some technical details are omitted and given in the Supplementary Material [7].

Remark 3.2. In Section S.3 of the Supplementary Material [7], we discuss the asymptotic behavior of $T_{n,\beta}(X^{(p)}, Y^{(p)})$ and $R_{n,\beta}(X^{(p)}, Y^{(p)})$ for dependent processes X, Y. In this case $T_\beta(X, Y)$ is positive. We prove central limit theory with Gaussian limits for

$$\sqrt{n}\big(T_{n,\beta}\big(X^{(p)}, Y^{(p)}\big) - T_\beta(X, Y), R_{n,\beta}\big(X^{(p)}, Y^{(p)}\big) - R_\beta(X, Y)\big).$$

In particular, if one used the normalization n for the independent case, one would get $n T_{n,\beta}(X^{(p)}, Y^{(p)}) \overset{\mathbb{P}}{\to} \infty$ and $n R_{n,\beta}(X^{(p)}, Y^{(p)}) \overset{\mathbb{P}}{\to} \infty$. This observation allows one to clearly distinguish between the independent case and the alternative of dependent X, Y.

The distinct asymptotic behavior of $T_{n,\beta}(X^{(p)}, Y^{(p)})$ and $R_{n,\beta}(X^{(p)}, Y^{(p)})$ in the independent and dependent cases is explained by the V-statistic structure underlying the sample distance covariance $T_{n,\beta}(X^{(p)}, Y^{(p)})$. Indeed, this quantity is approximated by the non-degenerate V-statistic $T_{n,\beta}(X, Y)$. In view of classical limit theory (see Arcones and Giné [1]) non-degenerate V-statistics satisfy the central limit theorem with normalization $\sqrt{n}$.

Remark 3.3. The numbers λ_i in parts (2) and (4) of the theorem are the eigenvalues of certain integral operators. This follows from limit theory for degenerate V-statistics; see Serfling [18], Lyons [12] and Arcones and Giné [1]. Unfortunately, neither the λ_i nor the distribution of the limit are available. Arcones and Giné [1] proved the consistency of a bootstrap version of degenerate U- and V-statistics. These latter results apply to $T_{n,\beta}(X, Y)$ but not to $T_{n,\beta}(X^{(p)}, Y^{(p)})$. In Section 5, we argue that the bootstrap also works for a modification of the latter quantity.

4. The condition $T_\beta(X, Y) = 0$ and independence of X and Y

The results in the previous section tell us that $T_{n,\beta}(X^{(p)}, Y^{(p)}) \overset{\mathbb{P}}{\to} T_\beta(X, Y) = 0$ for independent X, Y under various conditions on X, Y and the size of the mesh δ_n of the partition (t_i). An important question is whether, conversely, $T_\beta(X, Y) = 0$ also implies independence of X, Y. In

the case $\beta \in (0, 1]$ an affirmative answer to this question follows from Lyons [12], based on the fact that the metric obtained by raising the separable Hilbert space distance to the power $\beta \in (0, 1]$ is of the strong negative type. In the sequel, we extend the converse statement to all $\beta \in (0, 2)$. Our approach is based on studying the conditional independence of certain stochastic integrals.

Let B_1 and B_2 be independent Brownian motions on $[0, 1]$, independent of a pair (X, Y) of stochastically continuous bounded stochastic processes $[0, 1]$. The stochastic integrals

$$Z_1 = \int_0^1 X \, dB_1 \quad \text{and} \quad Z_2 = \int_0^1 Y \, dB_2$$

are well defined (and are, given (X, Y), independent normal random variables).

The next lemma demonstrates a connection between such stochastic integrals and distance covariances. Let $\mathcal{F}_B$ denote the σ-field generated by $B = (B_1, B_2)$.

Lemma 4.1. *Let $\beta \in (0, 2)$ and assume that $\mathbb{E}[\|X\|_2^\beta] + \mathbb{E}[\|Y\|_2^\beta] + \mathbb{E}[\|X\|_2^\beta \|Y\|_2^\beta] < \infty$. Let Y' be a copy of Y independent of everything else. Then*

$$c_0^2 T_\beta(X, Y) = \int_{\mathbb{R}^2} |st|^{-(1+\beta)} \mathbb{E}\big|\mathbb{E}\big[e^{is \int X(u) \, dB_1(u)} e^{it \int Y(u) \, dB_2(u)}$$

$$- e^{is \int X(u) \, dB_1(u)} e^{it \int Y'(u) \, dB_2(u)} \mid \mathcal{F}_B\big]\big|^2 \, ds \, dt, \tag{4.1}$$

where

$$c_0 = \int_{\mathbb{R}} \frac{1 - e^{-\frac{s^2}{2}}}{|s|^{1+\beta}} \, ds.$$

Proof. Consider an independent copy (X', Y') of (X, Y) and let Y'', Y''' be independent copies of Y which are independent of everything else. The expectation on the right-hand side in (4.1) can be written as

$$\mathbb{E}\big[e^{is \int (X-X') \, dB_1 + it \int (Y-Y') \, dB_2} + e^{is \int (X-X') \, dB_1 + it \int (Y''-Y''') \, dB_2}$$

$$- e^{is \int (X-X') \, dB_1 - it \int (Y-Y'') \, dB_2} - e^{-is \int (X-X') \, dB_1 + it \int (Y-Y'') \, dB_2}\big]$$

$$= \mathbb{E}\big[e^{-\frac{s^2}{2} \int (X(u)-X'(u))^2 \, du - \frac{t^2}{2} \int (Y(u)-Y'(u))^2 \, du}$$

$$+ e^{-\frac{s^2}{2} \int (X(u)-X'(u))^2 \, du - \frac{t^2}{2} \int (Y''(u)-Y'''(u))^2 \, du}$$

$$- 2e^{-\frac{s^2}{2} \int (X(u)-X'(u))^2 \, du - \frac{t^2}{2} \int (Y(u)-Y''(u))^2 \, du}\big]$$

$$= \mathbb{E}\big[\big(1 - e^{-\frac{s^2}{2} \int (X(u)-X'(u))^2 \, du}\big)\big(1 - e^{-\frac{t^2}{2} \int (Y(u)-Y'(u))^2 \, du}\big)$$

$$+ \big(1 - e^{-\frac{s^2}{2} \int (X(u)-X'(u))^2 \, du}\big)\big(1 - e^{-\frac{t^2}{2} \int (Y''(u)-Y'''(u))^2 \, du}\big)$$

$$- 2\big(1 - e^{-\frac{s^2}{2} \int (X(u)-X'(u))^2 \, du}\big)\big(1 - e^{-\frac{t^2}{2} \int (Y(u)-Y''(u))^2 \, du}\big)\big].$$

By change of variables,

$$\int_{\mathbb{R}} \frac{1 - e^{-\frac{s^2}{2} \int (X(u) - X'(u))^2 \, du}}{|s|^{1+\beta}} \, ds = c_0 \|X - X'\|_2^{\beta}.$$

Thus $T_\beta(X, Y)$ coincides with

$$\mathbb{E}\big[\|X - X'\|_2^{\beta} \|Y - Y'\|_2^{\beta} + \|X - X'\|_2^{\beta} \|Y'' - Y'''\|_2^{\beta} - 2\|X - X'\|_2^{\beta} \|Y - Y''\|_2^{\beta} \big]. \qquad \Box$$

An immediate corollary of Lemma 4.1 is that $T_\beta(X, Y) = 0$ implies that, for a.e. s, t,

$$\mathbb{E}\big[e^{is \int X(u) \, dB_1(u)} e^{it \int Y(u) \, dB_2(u)} - e^{is \int X(u) \, dB_1(u)} e^{it \int Y'(u) \, dB_2(u)} \mid \mathcal{F}_B \big] = 0$$

with probability 1. By Fubini's theorem, on an event of probability 1, this equality holds for all rational s, t, hence for all real s, t. We conclude that the stochastic integrals Z_1, Z_2 are conditionally independent given $\mathcal{F}_B$.

The next theorem, which is the main result of this section, shows that this implies independence of X and Y.

Theorem 4.2. *If the stochastic integrals Z_1 and Z_2 are a.s. conditionally independent given $\mathcal{F}_B$ then X, Y are independent. In particular, if $\beta \in (0, 2)$ and $\mathbb{E}[\|X\|_2^{\beta} + \|Y\|_2^{\beta} + \|X\|_2^{\beta}\|Y\|_2^{\beta}] < \infty$, then $T_\beta(X, Y) = 0$ if and only if X, Y are independent.*

Proof. Only the fact that the conditional independence of the integrals implies independence of X and Y remains to be proved. Let $(a(t), 0 \le t \le 1)$ and $(b(t), 0 \le t \le 1)$ be functions in $L^2[0, 1]$, and

$$A_1(t) = \int_0^t a(s) \, ds \quad \text{and} \quad A_2(t) = \int_0^t b(s) \, ds, \quad 0 \le t \le 1.$$

Since the law of the bivariate process

$$\big(\tilde{B}_1(t), \tilde{B}_2(t), 0 \le t \le 1 \big) = \big(B_1(t) + A_1(t), B_2(t) + A_2(t), 0 \le t \le 1 \big),$$

is equivalent to the law of the standard bivariate Brownian motion, it follows that the integrals

$$\int_0^1 X(t) \, d\tilde{B}_1(t) = \int_0^1 X(t) \, dB_1(t) + \int_0^1 X(t) a(t) \, dt$$

and

$$\int_0^1 Y(t) \, d\tilde{B}_2(t) = \int_0^1 Y(t) \, dB_2(t) + \int_0^1 Y(t) b(t) \, dt$$

are a.s. conditionally independent given $\mathcal{F}_B$.

It is not difficult to construct a sequence (C_n) of events in $\mathcal{F}_B$, of positive probability, such that the conditional laws of the integrals

$$\int_0^1 X(t)\,dB_1(t) \quad \text{and} \quad \int_0^1 Y(t)\,dB_2(t)$$

given C_n converge to the degenerate law at zero as $n \to \infty$. One way for producing such a sequence of events is to let the two independent Brownian motions take values close to zero at the points $i/n, i = 0, 1, \ldots, n$. Letting $n \to \infty$, we conclude that the integrals

$$\int_0^1 X(t)a(t)\,dt \quad \text{and} \quad \int_0^1 Y(t)b(t)\,dt$$

are independent.

For every fixed realization of the processes X and Y,

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_t^{t+\varepsilon} X(s)\,ds = X(t) \quad \text{and} \quad \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_t^{t+\varepsilon} Y(t)\,ds = Y(s) \qquad (4.2)$$

for all t in a set of full Lebesgue measure. By Fubini's theorem there is a set M of full Lebesgue measure such that, for every $t \in M$, (4.2) holds a.s. By necessity, the set M is dense in $[0, 1]$.

To prove our claim it suffices to prove that for any points $0 = t_0 < t_1 < \cdots < t_k < t_{k+1} = 1$, $k \geq 1$, the random vectors $(X(t_1), \ldots, X(t_k))$ and $(Y(t_1), \ldots, Y(t_k))$ are independent. By stochastic continuity of the processes X and Y it is enough to restrict ourselves to the case when every $t_i \in M$. Let $0 < \varepsilon < \min_{i=1,\ldots,k}(t_{i+1} - t_i)$. Choosing piece-wise constant functions $(a(t), 0 \leq t \leq 1)$ and $(b(t), 0 \leq t \leq 1)$, we conclude that the sums

$$\sum_{i=1}^k \theta_i \int_{t_i}^{t_i+\varepsilon} X(t)\,dt \quad \text{and} \quad \sum_{i=1}^k \gamma_i \int_{t_i}^{t_i+\varepsilon} Y(t)\,dt$$

are independent for any choice of $\theta_1, \ldots, \theta_k$ and $\gamma_1, \ldots, \gamma_k$. Since all points (t_i) are in the set M, dividing by ε and letting $\varepsilon \to 0$ we conclude that

$$\sum_{i=1}^k \theta_i X(t_i) \quad \text{and} \quad \sum_{i=1}^k \gamma_i Y(t_i)$$

are independent for any choice of $\theta_1, \ldots, \theta_k$ and $\gamma_1, \ldots, \gamma_k$. By the Cramér–Wold device this implies that the vectors $(X(t_1), \ldots, X(t_k))$ and $(Y(t_1), \ldots, Y(t_k))$ are independent. $\square$

5. The bootstrap for the sample distance covariance

We mentioned in Remark 3.3 that the limit distribution of $nT_{n,\beta}(X, Y)$ is not available. Theorem 3.1 states that the discretization $nT_{n,\beta}(X^{(p)}, Y^{(p)})$ has the same asymptotic properties as

$nT_{n,\beta}(X, Y)$ under suitable conditions on the smoothness of the sample paths, moment conditions and the growth rate of $p = p_n \to \infty$.

In this section, we advocate the use of the bootstrap for approximating the distribution of $nT_{n,\beta}(X^{(p)}, Y^{(p)})$. The bootstrap can be made to work for the degenerate V-statistic $T_{n,\beta}(X, Y)$ as shown in Arcones and Giné [1]. In this case, the naive bootstrap does not work and one has to modify the degenerate kernel. Since the V-statistic $T_{n,\beta}(X^{(p)}, Y^{(p)})$ is degenerate for every fixed p we face the problem of approximating the distribution of the latter statistic by its bootstrap version. We will show that this approximation works.

We will make use of a modification of Lemma 2.2 in Dehling and Mikosch [8], which deals with U-statistics with a kernel defined on the Euclidean space. We work with a separable metric space S. For $m \geq 1$, let $h : S^m \mapsto \mathbb{R}$ be a symmetric function. Let $(X_i^{(1)}, X_i^{(2)})$, $i = 1, 2, \ldots$, be an $S \times S$-valued i.i.d. sequence with marginal laws $\mathcal{L}(X^{(1)}) = F$ and $\mathcal{L}(X^{(2)}) = G$, respectively. On the subset of probability measures on S,

$$\Gamma_{2,h} = \{H : \mathbb{E}[h^2(Z_1, \ldots, Z_m)] < \infty \text{ for i.i.d. } (Z_i) \text{ with common law } H\},$$

we define the semi-metric

$$d_{2,h}(F, G) = \inf\{(\mathbb{E}[(h(X_1^{(1)}, \ldots, X_m^{(1)}) - h(X_1^{(2)}, \ldots, X_m^{(2)}))^2])^{1/2}\},$$

where the infimum is taken over all random elements $(X_1^{(1)}, \ldots, X_m^{(1)}, X_1^{(2)}, \ldots, X_m^{(2)})$ in S^{2m} such that $(X_i^{(1)}, X_i^{(2)})$, $i = 1, \ldots, m$, are i.i.d. S^2-valued random elements, $X_i^{(1)}$ has law F and $X_i^{(2)}$ has law G. The fact that $d_{2,h}$ is a semi-metric can be shown using similar arguments as in the proof of Lemma 8.1 in Bickel and Freedman [2] that discusses the properties of the related Wasserstein metric d_2 on a subset of probability measures on $\mathbb{R}$, $\Gamma_2 = \{H : \mathbb{E}_H[Z^2] < \infty\}$, defined by

$$d_2(F, G) = \inf\{(\mathbb{E}[|A - B|^2])^{1/2} : \mathcal{L}(A) = F, \mathcal{L}(B) = G\}.$$

Let $m \geq 2$ and choose $H \in \Gamma_{2,h}$. Define a function on $S \times S$ by

$$h_2(x, y; H) = \mathbb{E}[h(x, y, Z_3, \ldots, Z_m)] - \mathbb{E}[h(x, Z_2, \ldots, Z_m)]$$
$$- \mathbb{E}[h(Z_1, y, Z_3, \ldots, Z_m)] + \mathbb{E}[h(Z_1, \ldots, Z_m)], \tag{5.1}$$

where (Z_i) are i.i.d. with common law H. The proof of the following result is completely analogous to that of Lemma 2.2 in Dehling and Mikosch [8].

Lemma 5.1. *Let F, G be in $\Gamma_{2,h}$, $(X_j^{(1)})$ i.i.d. with common law F, and $(X_j^{(2)})$ i.i.d. with common law G. Then for any $n \geq 1$,*

$$d_2\left(\mathcal{L}\left(\frac{1}{n} \sum_{1 \leq i \neq j \leq n} h_2(X_i^{(1)}, X_j^{(1)}; F)\right), \mathcal{L}\left(\frac{1}{n} \sum_{1 \leq i \neq j \leq n} h_2(X_i^{(2)}, X_j^{(2)}; G)\right)\right)$$
$$\leq 2^{5/2} d_{2,h}(F, G). \tag{5.2}$$

For an S-valued i.i.d. sequence (Z_i) with common law $F \in \Gamma_{2,h}$ and $n \geq 1$ we denote by F_n the empirical law of $Z_1, \ldots, Z_n$. Consider an i.i.d. sequence (Z_{ni}^*) with the law F_n, that is, given that law, independent of (Z_i). The following result is analogous to Theorem 2.1 in Dehling and Mikosch [8].

Corollary 5.2. *Under the aforementioned conditions, and if also $\mathbb{E}[|h(Z_{i_1}, \ldots, Z_{i_m})|^2] < \infty$ for all indices $1 \leq i_1 \leq \cdots \leq i_m \leq m$, we have*

$$d_2\left(\mathcal{L}\left(\frac{1}{n} \sum_{1 \leq i \neq j \leq n} h_2(Z_{ni}^*, Z_{nj}^*; F_n)\right), \mathcal{L}\left(\frac{1}{n} \sum_{1 \leq i \neq j \leq n} h_2(Z_i, Z_j; F)\right)\right) \to 0$$

for almost all realizations of (Z_i).

Proof. By (5.2), it suffices to show that $d_{2,h}(F_n, F) \to 0$, almost surely. By Varadarajan's theorem (see Billingsley [3], p. 29) the empirical distribution F_n converges weakly to the distribution F, for almost all realizations $(z_i)_{i \geq 1}$ of $(Z_i)_{i \geq 1}$. Thus, by Skorokhod's theorem, there exist a sequence of random variables $(Z_n^*)_{n \geq 1}$ such that Z_n^* has distribution F_n, and an F-distributed random variable $\tilde{Z}$ such that $Z_n^* \to \tilde{Z}$ almost surely. We now take m i.i.d. copies of the pair $(Z_n^*, \tilde{Z})$, which we denote by $(Z_{n1}^*, \tilde{Z}_1), \ldots, (Z_{nm}, \tilde{Z}_m)$. Then

$$\left(Z_{n1}^*, \ldots, Z_{nm}^*\right) \to (\tilde{Z}_1, \ldots, \tilde{Z}_m), \quad \text{almost surely.}$$

Moreover, by definition of $d_{2,h}$, we have

$$d_{2,h}(F_n, F) \leq \left(\mathbb{E}\left[\left(h\left(Z_{n1}^*, \ldots, Z_{nm}^*\right) - h(\tilde{Z}_1, \ldots, \tilde{Z}_m)\right)^2\right]\right)^{1/2}.$$

It suffices to show that the right-hand side converges to 0 as $n \to \infty$. For any $\epsilon > 0$, we can find a bounded continuous function $g : S^m \to \mathbb{R}$ such that

$$\mathbb{E}\left[\left(h(\tilde{Z}_1, \ldots, \tilde{Z}_m) - g(\tilde{Z}_1, \ldots, \tilde{Z}_m)\right)^2\right] \leq \epsilon.$$

By Lebesgue's dominated convergence theorem, we obtain

$$\mathbb{E}\left[\left(g\left(Z_{n,1}^*, \ldots, Z_{n,m}^*\right) - g(\tilde{Z}_1, \ldots, \tilde{Z}_m)\right)^2\right] \to 0.$$

The strong law of large numbers for U-statistics implies that

$$\mathbb{E}\left[\left(h\left(Z_{n,1}^*, \ldots, Z_{n,m}^*\right) - g\left(Z_{n,1}^*, \ldots, Z_{n,m}^*\right)\right)^2\right]$$
$$= \frac{1}{n^m} \sum_{1 \leq i_1, \ldots, i_m \leq n} \left(h(z_{i_1}, \ldots, z_{i_m}) - g(z_{i_1}, \ldots, z_{i_m})\right)^2$$
$$\to \mathbb{E}\left(h(Z_1, \ldots, Z_m) - g(Z_1, \ldots, Z_m)\right)^2 \leq \epsilon.$$

This finishes the proof. □

In what follows, (Z_i) will stand for the i.i.d. sequence of the pairs (X_i, Y_i), $i = 1, 2, \ldots$, used in the previous sections for defining the quantities $T_{n,\beta}(X, Y)$. Correspondingly, we write $(Z_i^{(p)})$ for the sequence of the discretizations $(X_i^{(p)}, Y_i^{(p)})$, $i = 1, 2, \ldots$, with generic element $Z^{(p)}$. For the ease of presentation, we focus on the case $\beta = 1$ and suppress β in the notation. We consider only the case when X, Y have finite second moments. A generic element $Z = (X, Y)$ has trajectory (x, y) assuming values in a function space S where x, y are defined on $[0, 1]$ and are Riemann square-integrable.

Under the hypothesis that X, Y are independent, $T_n(X, Y)$ has representation as a V-statistic of order 4 with a 1-degenerate symmetric kernel $h_4 = h(x_1, x_2, x_3, x_4)$; see the Appendix, where we also show that, when scaled by n, the limits of $T_n(X, Y)$ and the corresponding normalized U-statistic (which is obtained by ignoring all summands $h(Z_{i_1}, Z_{i_2}, Z_{i_3}, Z_{i_4})$ with the property $i_j = i_k$ for $j \neq k$) differ by an additive constant. Applying the Hoeffding decomposition to this U-statistic, the limiting distribution of $nT_n(X, Y)$ coincides, up to a scale change, with the limiting distribution of the following normalized U-statistic:

$$U_n(Z) = \frac{1}{n} \sum_{1 \leq i \neq j \leq n} h_2(Z_i, Z_j; F_Z),$$

where $F_Z = F_X \times F_Y$ and h_2 is defined in (5.1). Arcones and Giné [1] proved that the correct bootstrap version of $nT_n(X, Y)$ is

$$U_n(Z^*) = \frac{1}{n} \sum_{1 \leq i \neq j \leq n} h_2(Z_{ni}^*, Z_{nj}^*; F_{n,Z}),$$

where $F_{n,Z}$ is the empirical distribution of the i.i.d. sample $Z_1, \ldots, Z_n$. The fact that the limiting distributions of $U_n(Z)$ and $U_n(Z^*)$ coincide follows from Corollary 5.2.

Our program for the remainder of this section is to show that we are allowed to replace $Z = (X, Y)$ by the corresponding discretizations $Z^{(p)} = (X^{(p)}, Y^{(p)})$ in the aforementioned U- and V-statistics, that is, we will show that suitable bootstrap versions of $nT_{n,\beta}(X, Y)$ and $nT_{n,\beta}(X^{(p)}, Y^{(p)})$ have the same limiting distribution. We start by showing that $U_n(Z)$ and $U_n(Z^{(p)})$ are close in the sense of the d_2-metric.

Lemma 5.3. *Assume the following conditions*:

1. *X, Y are independent and have finite second moments.*
2. *Condition* (A1) *holds.*
3. *$\delta_n \to 0$ as $n \to \infty$.*

Then $d_2(\mathcal{L}(U_n(Z)); \mathcal{L}(U_n(Z^{(p)}))) \leq c\delta_n^{(\gamma_X \wedge \gamma_Y)/2} \to 0$.

Proof. By (5.2), with h given by (A.1), we have

$$d_2\big(\mathcal{L}\big(U_n(Z)\big); \mathcal{L}\big(U_n\big(Z^{(p)}\big)\big)\big)$$
$$\leq c\big\{\mathbb{E}\big[\big(h(Z_1, \ldots, Z_4) - h\big(Z_1^{(p)}, \ldots, Z_4^{(p)}\big)\big)^2\big]\big\}^{1/2}$$

$$\leq c\big\{\mathbb{E}\big[\big(f(Z_1,\ldots,Z_4) - f\big(Z_1^{(p)},\ldots,Z_4^{(p)}\big)\big)^2\big]\big\}^{1/2}$$

$$\leq c\big(\mathbb{E}I_1^2 + \mathbb{E}I_2^2 + \mathbb{E}I_3^2\big)^{1/2},$$

where

$$I_1 = \big\|X_1 - X_2\big\|_2\big\|Y_1 - Y_2\big\|_2 - \big\|X_1^{(p)} - X_2^{(p)}\big\|_2\big\|Y_1^{(p)} - Y_2^{(p)}\big\|_2,$$

$$I_2 = \big\|X_1 - X_2\big\|_2\big\|Y_3 - Y_4\big\|_2 - \big\|X_1^{(p)} - X_2^{(p)}\big\|_2\big\|Y_3^{(p)} - Y_4^{(p)}\big\|_2,$$

$$I_3 = \big\|X_1 - X_2\big\|_2\big\|Y_1 - Y_3\big\|_2 - \big\|X_1^{(p)} - X_2^{(p)}\big\|_2\big\|Y_1^{(p)} - Y_3^{(p)}\big\|_2.$$

The second moments are estimated as in Proposition 7.1 below. We have by (7.4),

$$\mathbb{E}\big[\big(\big\|X_1 - X_2\big\|_2 - \big\|X_1^{(p)} - X_2^{(p)}\big\|_2\big)^2\big\|Y_1 - Y_2\big\|_2^2\big] \leq c\delta_n^{\gamma_X}$$

and

$$\mathbb{E}\big[\big\|X_1^{(p)} - X_2^{(p)}\big\|_2^2\big(\big\|Y_1 - Y_2\big\|_2 - \big\|Y_1^{(p)} - Y_2^{(p)}\big\|_2\big)^2\big] \leq c\delta_n^{\gamma_Y}.$$

That is, $\mathbb{E}[I_1^2] \leq c\delta_n^{\gamma_X \wedge \gamma_Y}$. The second moments of I_2, I_3 can be bounded by the same quantities. $\qquad\square$

Our next goal is to show that, under appropriate assumptions, the difference between the laws of $U_n(Z^*)$ and $U_n(Z^{(p)*})$ asymptotically vanishes.

Lemma 5.4. *Consider the following conditions*:

1. *X, Y are independent and have finite second moments.*
2a. *Condition* (A1) *holds.*
2b. *$\mathbb{E}[|X(t) - X(s)|^4] \leq c|t - s|^{\tilde{\gamma}_X}$ and $\mathbb{E}[|Y(t) - Y(s)|^4] \leq c|t - s|^{\tilde{\gamma}_Y}$ hold.*
3a. *$\sum_{n=1}^{\infty} \delta_n^{\gamma_X \wedge \gamma_Y} < \infty$.*
3b. *$\sum_{n=1}^{\infty} \big(\delta_n^{2(\gamma_X \wedge \gamma_Y)} + n^{-1}\delta_n^{\tilde{\gamma}_X \wedge \tilde{\gamma}_Y}\big) < \infty$.*

If either 1, 2a, 3a *or* 1, 2a, 2b, 3b *hold, then $d_2(\mathcal{L}(U_n(Z^*)), \mathcal{L}(U_n(Z^{(p)*}))) \to 0$, for a.e. realization of (Z_i).*

Proof. With h given by (A.1), by Lemma 5.1 it is enough to prove that $d_{2,h}(\mathcal{L}(Z^*), \mathcal{L}(Z^{(p)*})) \to 0$ for a.e. realization of (Z_i). We have

$$d_{2,h}(n) := d_{2,h}\big(\mathcal{L}(Z^*), \mathcal{L}(Z^{(p)*})\big)$$

$$\leq \big(\mathbb{E}_{F_n}\big[\big(h\big(Z_1^*, Z_2^*, Z_3^*, Z_4^*\big) - h\big(Z_1^{(p)*}, Z_2^{(p)*}, Z_3^{(p)*}, Z_4^{(p)*}\big)\big)^2\big]\big)^{1/2}$$

$$= \frac{1}{n^2}\bigg(\sum_{1 \leq i_1,i_2,i_3,i_4 \leq n} \big(h(Z_{i_1}, Z_{i_2}, Z_{i_3}, Z_{i_4}) - h\big(Z_{i_1}^{(p)}, Z_{i_2}^{(p)}, Z_{i_3}^{(p)}, Z_{i_4}^{(p)}\big)\big)^2\bigg)^{1/2}$$

$$\leq \frac{1}{n^2} \left(\sum_{1 \leq i_1, i_2, i_3, i_4 \leq n} \left(f(Z_{i_1}, Z_{i_2}, Z_{i_3}, Z_{i_4}) - f\left(Z_{i_1}^{(p)}, Z_{i_2}^{(p)}, Z_{i_3}^{(p)}, Z_{i_4}^{(p)}\right) \right)^2 \right)^{1/2}.$$

We first show that the right-hand side converges to zero under the assumption that 1, 2a, and 3a hold. Using (A1), we obtain

$$\mathbb{E}\left[d_{2,h}\left(\mathcal{L}(Z^*), \mathcal{L}(Z^{(p)*}) \right) \right]^2$$
$$\leq \sum_{1 \leq j_1, j_2, j_3, j_4 \leq 4} \mathbb{E}\left[\left(f(Z_{j_1}, \ldots, Z_{j_4}) - f\left(Z_{j_1}^{(p)}, \ldots, Z_{j_4}^{(p)}\right) \right)^2 \right] \leq c \delta_n^{\gamma_X \wedge \gamma_Y}.$$

Thus, if $\sum_n \delta_n^{\gamma_X \wedge \gamma_Y} < \infty$ applications of Markov's inequality and the Borel–Cantelli lemma yield that $d_{2,h}(\mathcal{L}(Z^*), \mathcal{L}(Z^{(p)*})) \to 0$ a.s. as $n \to \infty$.

Now assume that 1, 2a, 2b and 3b hold. Using standard calculations for U-statistics, we have

$$\mathrm{var}\left(d_{2,h}^2(n) \right) \leq c \sum_{1 \leq j_1, j_2, j_3, j_4 \leq 4} \left[n^{-1} \mathrm{var}\left(\left(h(Z_{j_1}, \ldots, Z_{j_4}) - h\left(Z_{j_1}^{(p)}, \ldots, Z_{j_4}^{(p)}\right) \right)^2 \right) \right.$$
$$\left. + \left(\mathbb{E}\left[\left(h(Z_{j_1}, \ldots, Z_{j_4}) - h\left(Z_{j_1}^{(p)}, \ldots, Z_{j_4}^{(p)}\right) \right)^2 \right] \right)^2 \right]$$
$$= J_1 + J_2.$$

We have $J_2 = O(\delta_n^{2(\gamma_X \wedge \gamma_Y)})$. We can handle J_1 similarly to the proof of Lemma 5.3. For example,

$$\mathbb{E}\left[\left\| X_1 - X_1^{(p)} \right\|_2^4 \right] = \mathbb{E}\left[\left(\int_0^1 \left(X(u) - X^{(p)}(u) \right)^2 du \right)^2 \right]$$
$$\leq c \int_0^1 \mathbb{E}\left[\left(X(u) - X^{(p)}(u) \right)^4 \right] du \leq c \delta_n^{\widetilde{\gamma}_X}.$$

Now $d_{2,h}(n) \overset{\text{a.s.}}{\to} 0$ as $n \to \infty$ follows by an application of Markov's inequality of order 2, the Borel–Cantelli lemma and since $\sum_n (n^{-1} \delta_n^{\widetilde{\gamma}_X \wedge \widetilde{\gamma}_Y} + \delta_n^{2(\gamma_X \wedge \gamma_Y)}) < \infty$. We omit further details. $\square$

Combining the previous arguments, a natural bootstrap version of the degenerate V-statistic $n T_n(X^{(p)}, Y^{(p)})$ is given by $U_n(Z^{(p)*})$.

Proposition 5.5. *Assume the conditions of Lemma* 5.4. *Then*

$$d_2\left(\mathcal{L}(U_n(Z)), \mathcal{L}(U_n(Z^{(p)*})) \right) \to 0$$

for a.e. realization of (Z_i).

For an application of the bootstrapped sample distance correlation $n R_n(X^{(p)}, Y^{(p)})$ we still miss one step in the derivation of the bootstrap consistency: we also need to prove that the

denominator quantities converge a.s.

$$T_n\big(X^{(p)}, X^{(p)}\big) \xrightarrow{\text{a.s.}} T(X, X) \quad \text{and} \quad T_n\big(Y^{(p)}, Y^{(p)}\big) \xrightarrow{\text{a.s.}} T(Y, Y), \quad n \to \infty.$$

In Lemma S.2 of the Supplementary Material [7], we provide sufficient conditions for this to hold.

6. Simulations

In this section, we illustrate the theoretical results in a small simulation study. Throughout we choose $\beta = 1$ and suppress the dependence on β in the notation.

We start with identically distributed fractional Brownian motions (fBM) X, Y on $[0, 1]$ with Hurst coefficient H and correlation ρ where the dependence between X and Y is given by the covariance function

$$\text{cov}\big(X(s), Y(t)\big) = \frac{\rho}{2}\big\{|s|^{2H} + |t|^{2H} - |t - s|^{2H}\big\}, \quad s, t \in [0, 1].$$

If $X = Y$, we also set $\rho = 1$. Note that, for $H = 1/2$, the right-hand side collapses into $\rho(s \wedge t)$, corresponding to Brownian motions X, Y. The top graph in Figure 1 nicely illustrates the consistency of the sample correlation $R_n(X^{(p)}, Y^{(p)})$ for independent X and Y ($\rho = 0$). In the top row we fix $p = 100$ and increase n from 100 to 400, and we choose $H = 1/4$, $H = 1/2$ (BM) and $H = 3/4$. Apparently, we can see the influence of the smoothness of the sample paths: the larger H the larger $\gamma_X = \gamma_Y = 2H$ (see Example 2.3), the smoother the sample paths and the closer $R_n(X^{(p)}, Y^{(p)})$ to zero; see also the upper bounds in Proposition 7.1. In the bottom row, we show $R_n(X^{(p)}, Y^{(p)})$ for dependent X and Y with $\rho = 0.5$. We again choose $H = 1/4$, $H = 1/2$ (BM) and $H = 3/4$, fix $p = 100$ and increase n from 100 to 300. In the bottom graphs the sample distance correlation converges to some positive constants; we see a clear difference between the independent and dependent cases.

In Figure 2, we illustrate the performance of the sample distance correlation $R_n(X^{(p)}, Y^{(p)})$ when X and Y are independent (possibly with distinct distributions) non-Gaussian processes. We treat three cases, including heavy-tailed processes: X, Y are i.i.d. geometric BMs (left), X, Y are i.i.d. α-stable Lévy motions (middle), X is a geometric BM and Y an α-stable Lévy motion (right). For geometric BM we choose the parametrization

$$X(t) = \exp\big((1 - 0.7^2/2)t + 0.7B(t)\big), \quad t \in [0, 1],$$

where $\mu = 1$ (drift), $\sigma = 0.7$ (volatility) and B is standard BM. The parameters of the α-stable Lévy motions are $(\alpha, \beta, \mu, \sigma) = (1.8, 0.3, 0, 1)$; cf. Samorodnitsky and Taqqu [17], Ex. 3.1.3. We fix $p = 100$ and increase n from 100 to 300. Also in these non-Gaussian settings the boxplots nicely illustrate consistency of $R_n(X^{(p)}, Y^{(p)})$ even in the heavy-tailed α-stable case.

In Figure 3, we study the influence of the size of p on the sample distance correlation for a given n. We choose $p = 100$ (left) and $p = 1000$ (middle) while X, Y are independent BMs (top) and dependent BMs with correlation $\rho = 0.5$ (bottom): there is hardly any difference between the left and middle graphs for a given n. In the right graphs, we choose i.i.d. (top) and dependent
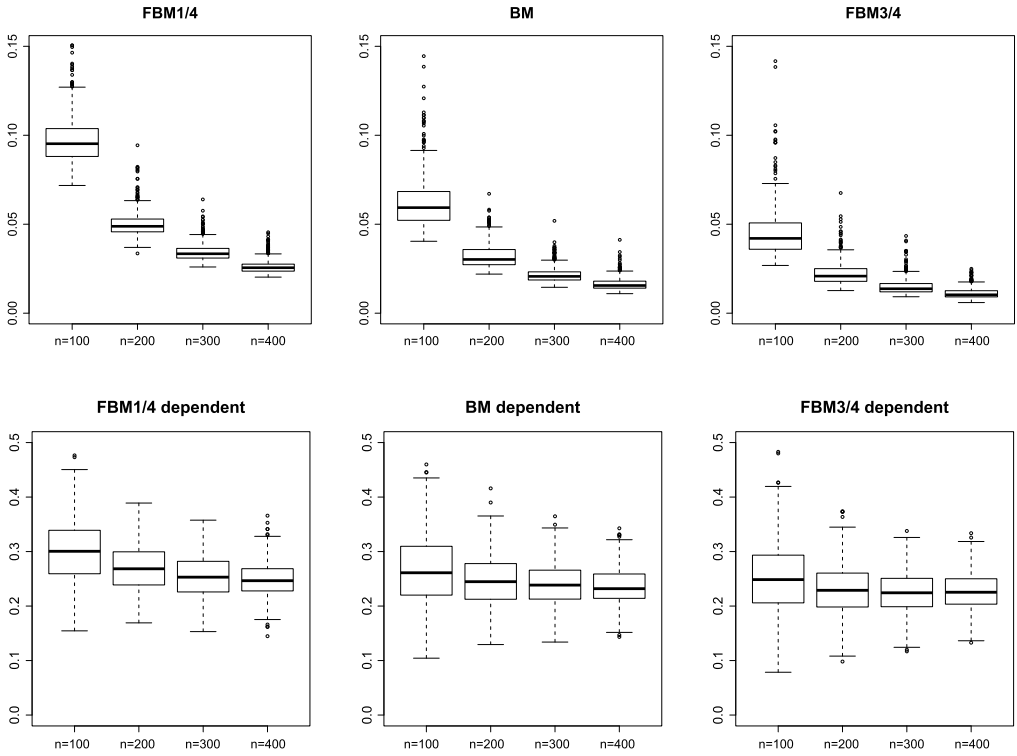
Figure 1. Boxplots for $R_n(X^{(p)}, Y^{(p)})$ simulated fBMs X, Y with $H = 1/4, 1/2, 3/4$ (from left to right), $p = 100$ and increasing sample sizes n. Top: i.i.d. fBMs X, Y. Each boxplot is based on 500 replications. Bottom: identically distributed fBMs X, Y with correlation $\rho = 0.5$. Each boxplot is based on 300 replications.
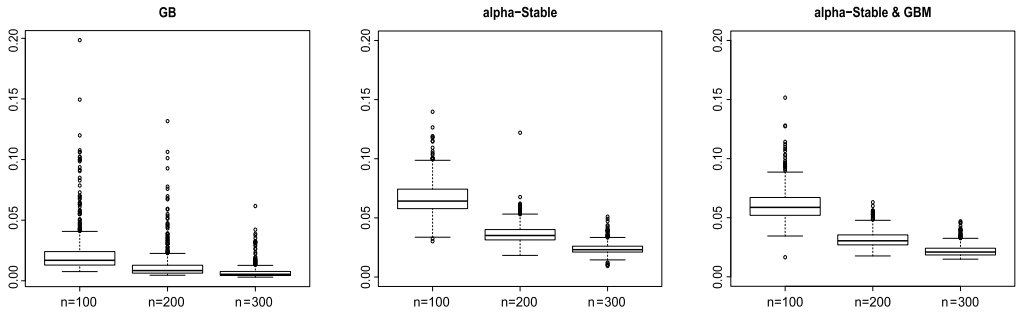

Figure 2. Boxplots for $R_n(X^{(p)}, Y^{(p)})$ for simulated independent non-Gaussian processes X, Y, $p = 100$ and increasing sample size n. Each boxplot is based on 500 replications. Left: i.i.d. geometric BMs X, Y. Middle: i.i.d. α-stable Lévy motions X, Y. Right: independent geometric BM X and α-stable Lévy motion Y.
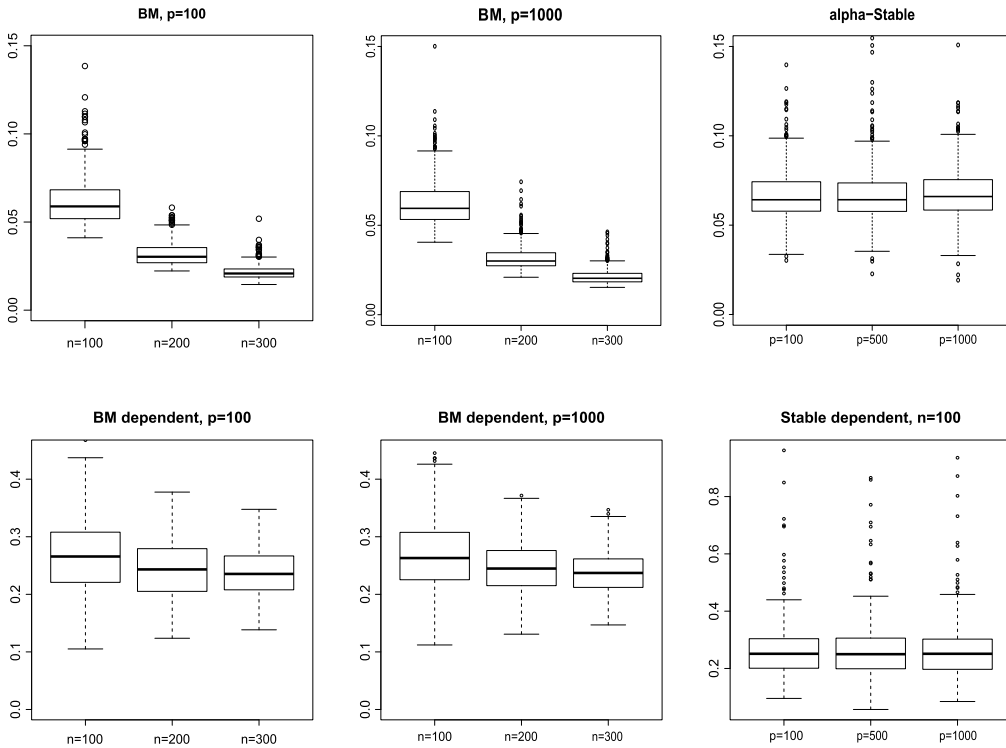
Figure 3. Boxplots for $R_n(X^{(p)}, Y^{(p)})$ for different p. Top left and middle: X, Y are i.i.d. BMs. For each $p = 100$ (left) and $p = 1000$ (middle) we take three distinct sample sizes $n = 100, 200, 300$. The boxplots are based on 300 replications. Top right: X, Y are i.i.d. α-stable Lévy motions, $n = 100$ is fixed while $p = 100, 500, 1000$. The boxplots are based on 500 replications. Bottom left and middle: X, Y are dependent BMs with correlation $\rho = 0.5$, where p and n are the same as those for i.i.d. BMs. Bottom right: X, Y are dependent α-stable Lévy motions given by (6.2) with $\rho = 0.5$. The other parameters are the same as at the top right.

(bottom) α-stable Lévy motions X, Y. For the i.i.d. case we take the same parameters as before, and for the dependent case we use (6.2) with $\rho = 0.5$. We increased p from 100 to 1000 and fix $n = 100$. Again, one can hardly see any difference between the boxplots. These observations are not surprising, and indicate that the discretized distance correlation with $p = 100$ is already very close to the limit obtained for $p \to \infty$.

We also examine some dependent heavy-tailed cases. We have chosen two simple stochastic process models for X, Y where we can control the tails and the dependence. First, we consider i.i.d. standard BMs B_1, B_2 which are subject to a joint heavy-tailed shock, $(X, Y) = A^{1/2}(B_1, B_2)$, where A is a Pareto(α) variable for some $\alpha > 0$ with density $f_\alpha(x) = \alpha(1 + x)^{-(\alpha+1)}, x > 0$. We also assume that A and (B_1, B_2) are independent. Notice that $A^{1/2}$ does not have a 2αth moment. Second, we consider $(X, Y) = (A_1^{1/2} B_1, A_2^{1/2} B_2)$ where A_1, A_2 are i.i.d. copies of A with density f_α, independent of (B_1, B_2) while B_1 and B_2 are
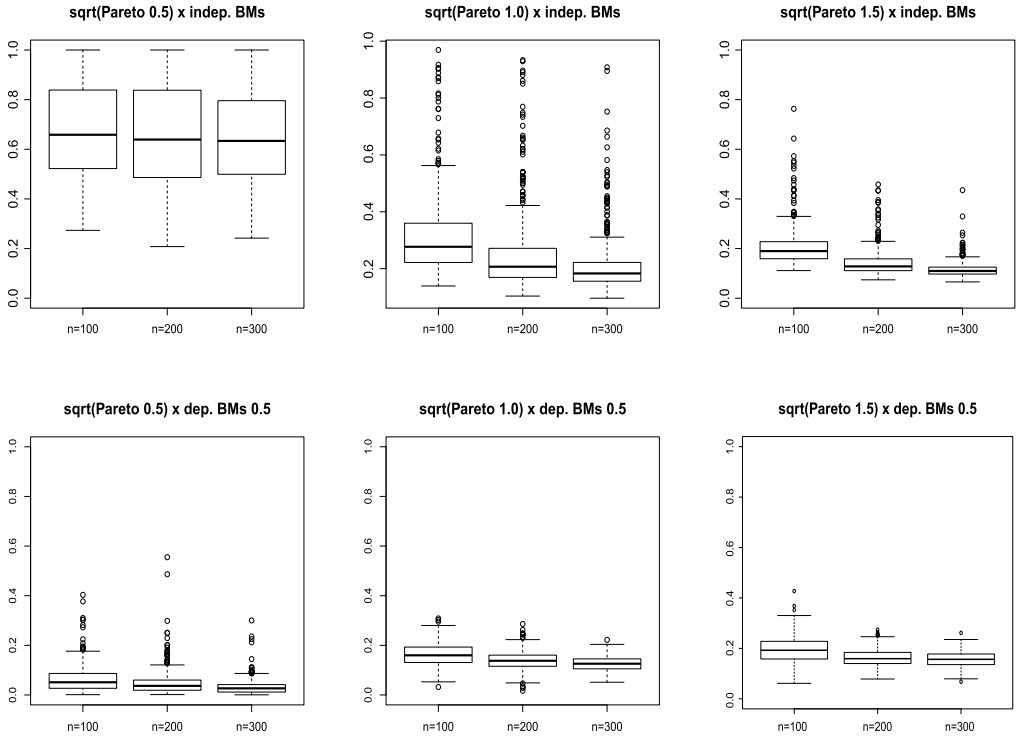
Figure 4. Boxplots of $R_n(X^{(p)}, Y^{(p)})$ for dependent heavy-tail cases. Top: $(X, Y) = A^{1/2}(B_1, B_2)$ for a Pareto(α) variable A independent of i.i.d. Brownian motions (B_1, B_2). Bottom: $(X, Y) = (A_1^{1/2} B_1, A_2^{1/2} B_2)$ for i.i.d. copies A_1, A_2 independent of the Brownian motions B_1, B_2 with correlation $\rho = 0.5$. From left to right: $\alpha = 0.5, 1.0, 1.5$. Sample sizes $n = 100, 200, 300$, $p = 100$, and each plot is based on 500 replications.

dependent BMs with correlation $\rho = 0.5$. We have chosen $2\alpha = 1, 2, 3$. In the case $\alpha = 0.5$ the theoretical results of this paper about consistency of $T_n(X^{(p)}, Y^{(p)})$ do not apply since $\mathbb{E}[\|X\|_2 + \|Y\|_2] = \infty$ while in the cases $\alpha = 1, 1.5$, $T_n(X^{(p)}, Y^{(p)}) \overset{\mathbb{P}}{\to} T(X, Y) > 0$.

The first/second model is examined in the top/bottom graphs of Figure 4, respectively. In the cases $\alpha = 1, 1.5$ the centers of the boxplots seem to stabilize with increasing sample size, pointing at the consistency of $R_n(X^{(p)}, Y^{(p)})$. In the top graphs (first model), we observe that the distributions of $R_n(X^{(p)}, Y^{(p)})$ have a rather wide range while the bottom boxplots (second model) are less spread and their center is much below those of the first model. Moreover, in the $\alpha = 0.5$ case the plot is close to zero. It could be taken as a false indication of independence between X and Y. We do not have a full explanation for the phenomena observed in Figure 4; in both heavy-tailed dependent models our assumptions for the existence of non-degenerate weak limits are not satisfied due to the lack of moments.

We have conducted a simulation study to illustrate the performance of the bootstrap procedure for the distance correlation based test for independence. Specifically, we have tested

Table 1. Bootstrap sizes of $nR_n(X^{(p)}, Y^{(p)})$ with independent BMs and α-stable LMs

	Brownian motions		1.8-stable Lévy motions	
n	$\xi = 0.025$	$\xi = 0.05$	$\xi = 0.025$	$\xi = 0.05$
100	2.9	5.7	1.8	5.8
200	2.9	5.0	3.2	6.8
300	3.1	5.6	4.4	7.0
400	3.6	6.8	3.4	5.4

Empirical rejection rates of bootstrap test based on $nR_n(X^{(p)}, Y^{(p)})$ for independent BMs ($M = 1000$ iterations) and independent α-stable LMs X, Y ($M = 500$).

for independence of two BMs and two α-stable Lévy motions X, Y. From a given sample $(X_1^{(p)}, Y_1^{(p)}), \ldots, (X_n^{(p)}, Y_n^{(p)})$ of $(X^{(p)}, Y^{(p)})$ we draw 500 bootstrap samples; we choose $p = 100$. From each sample we calculate the distance correlation, from these 500 distance correlations we determine the empirical $(1 - \xi)$-quantile $q_{1-\xi}^*$ and verify whether

$$R_n(X^{(p)}, Y^{(p)}) \geq q_{1-\xi}^*. \tag{6.1}$$

Then we repeat this procedure $M = 500$ and $M = 1000$ times and count the number of times when (6.1) is satisfied.

The empirical rejection rates in the case of independent BMs and independent 1.8-stable LMs X, Y are reported in Table 1: each cell corresponds to a given sample size n and a given test level ξ; $M = 1000$ iterations for BMs and $M = 500$ for stable LMs.

In Table 2, we conduct the corresponding simulations for dependent standard BMs with correlation $\rho \in (0, 1)$ and $M = 500$. We also consider dependent α-stable processes (X, Y): given

Table 2. Bootstrap power of $nR_n(X^{(p)}, Y^{(p)})$ with correlated BMs and α-stable LMs

	Dependent BMs						Dependent 1.8. stable Lévy motion					
n:	100		200		300		100		200		300	
$\rho \setminus \xi$	0.025	0.05	0.025	0.05	0.025	0.05	0.25	0.05	0.025	0.05	0.025	0.05
0.1	8.0	13.4	21.4	29.0	28.6	36.6	10.4	15.8	19.4	27.8	28.4	38.4
0.2	38.4	49.4	69.0	80.2	89.2	93.4	39.4	49.4	62.0	73.4	79.4	87.2
0.3	78.4	85.8	98.0	98.6	99.8	99.8	63.6	65.0	89.0	93.0	93.8	96.4
0.4	96.2	97.4	100	100	100	100	84.0	90.2	94.4	96.4	96.4	97.8
0.5	99.4	100	100	100	100	100	91.4	93.4	96.2	96.6	97.2	97.8

Empirical rejection rates of bootstrap test based on $nR_n(X^{(p)}, Y^{(p)})$ for correlated BMs and 1.8-stable LMs. In each cell, the rejection rate is given among $M = 500$ iterations.

Table 3. Comparison of powers with tests by Matsui et al. [15]

	Dependent BMs						Dependent 1.8. stable					
	Matsui et al.		DC $p = 100$		DC $p = 10^4$		Matsui et al.		DC $p = 100$		DC $p = 10^4$	
$\rho \setminus \xi$	0.025	0.05	0.025	0.05	0.025	0.05	0.025	0.05	0.025	0.05	0.025	0.05
0.1	7	9	6	13	4	12	6	14	15	21	20	25
0.2	38	41	32	51	36	48	42	53	45	50	56	63
0.3	81	84	71	84	73	81	82	88	72	83	82	86
0.4	99	100	98	100	97	98	99	99	94	98	96	98
0.5	100	100	100	100	100	100	100	100	99	99	99	99

Simulated powers of tests based on $n\overline{R}_n(X, Y)$ (Matsui et al. [15]) (left rows) and $nR_n(X^{(p)}, Y^{(p)})$ (middle and right rows) with two different $p = 100, 10^4$ for correlated BMs and 1.8-stable LMs; for $M = 100$ Monte-Carlo runs. The sample size is $n = 100$ in all cases.

two independent α-stable Lévy motions (X, Z), we take

$$Y = \rho X + Z \tag{6.2}$$

with $\rho = 0.1, \ldots, 0.5$, noticing the scale invariance of $R(X, Y)$. In agreement with the theory, the rejection rates increase as n and ρ increase.

Finally, we have compared the powers of our tests with the test proposed in Matsui et al. [15], taking the same BMs and 1.8-stable LMs alternatives as in Table 2. We have calculated the critical values for all tests by Monte-Carlo simulation for independent processes, taking sample size $n = 100$ and $M = 300$ runs. We have then simulated the powers for various dependent processes. For the Matsui et al. [15] test, we take standard Poisson intensity, resulting in the test statistic $n\overline{R}_n(X, Y)$; see Matsui et al. [15], p. 366, Section 4. We calculate the integrals in $n\overline{R}_n(X, Y)$ numerically by Riemann sums with 100 equidistant grid points. For the test statistics $nR_n(X^{(p)}, Y^{(p)})$, developed in the present paper, we take discretizations $p = 100, 10^4$. The rejection rates given in Table 3 are based on $M = 100$ Monte-Carlo runs. The small number of runs was chosen because the calculation of the test statistics $n\overline{R}_n(X, Y)$ is computationally very complex. In view of Table 3, we see that both statistics are comparable. It seems that for stable processes, our test is more powerful than the test of Matsui et al. [15]. For all alternatives, the choice of the discretization $p = 100$ or $p = 10^4$ seems to make little difference. This suggests that $p = 100$ is already a sufficiently fine discretization.

7. Proof of Theorem 3.1

We prove the theorem by a series of auxiliary results.

Proposition 7.1. *Assume the conditions* 1.–4. *of Theorem* 3.1.

1. *If also* (A1) *holds, then there is c such that for any* $n \geq 1$,
$$\mathbb{E}\big[\big|T_{n,\beta}\big(X^{(p)}, Y^{(p)}\big) - T_{n,\beta}(X, Y)\big|\big] \leq c\delta_n^{(\gamma_X \wedge \gamma_Y)(\beta \wedge 1)/2}.$$

2. *If also* (B1), (B2) *hold, then there is c such that*
$$\mathbb{E}\big[\big|T_{n,\beta}\big(X^{(p)}, Y^{(p)}\big) - T_{n,\beta}(X, Y)\big|\big] \leq c\big(p\delta_n^{(\beta/2+\gamma_X \wedge \gamma_Y)}\big)^{(\beta \wedge 1)/\beta}.$$

Proof. We start with the decomposition
$$T_{n,\beta}\big(X^{(p)}, Y^{(p)}\big) - T_{n,\beta}(X, Y) = I_1 + I_2 - 2I_3, \tag{7.1}$$

where

$$I_1 = \frac{1}{n^2} \sum_{k,l=1}^{n} \big(\big\|X_k^{(p)} - X_l^{(p)}\big\|_2^\beta \big\|Y_k^{(p)} - Y_l^{(p)}\big\|_2^\beta - \|X_k - X_l\|_2^\beta \|Y_l - Y_k\|_2^\beta\big),$$

$$I_2 = \frac{1}{n^4} \sum_{k,l=1}^{n} \big\|X_k^{(p)} - X_l^{(p)}\big\|_2^\beta \sum_{k,l=1}^{n} \big\|Y_k^{(p)} - Y_l^{(p)}\big\|_2^\beta$$
$$- \frac{1}{n^4} \sum_{k,l=1}^{n} \|X_k - X_l\|_2^\beta \sum_{k,l=1}^{n} \|Y_k - Y_l\|_2^\beta, \tag{7.2}$$

$$I_3 = \frac{1}{n^3} \sum_{k,l,m=1}^{n} \big\|X_k^{(p)} - X_l^{(p)}\big\|_2^\beta \big\|Y_k^{(p)} - Y_m^{(p)}\big\|_2^\beta$$
$$- \frac{1}{n^3} \sum_{k,l,m=1}^{n} \|X_k - X_l\|_2^\beta \|Y_k - Y_m\|_2^\beta.$$

We will find bounds for the absolute values of the expectations of these quantities. From now on, c denotes any positive constants whose values are not of interest.

First assume that (X, Y) *have finite second moment.* Observe that

$$|I_1| \leq \frac{1}{n^2} \sum_{k,l=1}^{n} \big|\big\|X_k^{(p)} - X_l^{(p)}\big\|_2^\beta - \|X_k - X_l\|_2^\beta\big|\big\|Y_k^{(p)} - Y_l^{(p)}\big\|_2^\beta$$
$$+ \frac{1}{n^2} \sum_{k,l=1}^{n} \big|\big\|Y_k^{(p)} - Y_l^{(p)}\big\|_2^\beta - \|Y_k - Y_l\|_2^\beta\big|\|X_k - X_l\|_2^\beta$$
$$=: I_{11} + I_{12}. \tag{7.3}$$

By a symmetry argument, interchanging the roles of X and Y, it suffices to consider I_{11}. Using the independence of X and Y, we have

$$\mathbb{E}[I_{11}] \leq \mathbb{E}\big[\big|\big\|X_1^{(p)} - X_2^{(p)}\big\|_2^\beta - \|X_1 - X_2\|_2^\beta\big|\big]\mathbb{E}\big[\big\|Y_1^{(p)} - Y_2^{(p)}\big\|_2^\beta\big].$$

By Lyapunov's inequality,

$$\mathbb{E}\big[\big\|Y_1^{(p)} - Y_2^{(p)}\big\|_2^\beta\big] \le \big(\mathbb{E}\big[\big\|Y_1^{(p)} - Y_2^{(p)}\big\|_2^2\big]\big)^{\beta/2}$$

$$\le c\left(\int_0^1 \mathrm{var}\big(Y^{(p)}(t)\big)\,dt\right)^{\beta/2} < \infty.$$

Assume $0 < \beta \le 1$. Then, by concavity and Jensen's inequality,

$$\mathbb{E}\big[\big|\big\|X_1^{(p)} - X_2^{(p)}\big\|_2^\beta - \big\|X_1 - X_2\big\|_2^\beta\big|\big]$$

$$\le \mathbb{E}\big[\big\|\big(X_1^{(p)} - X_2^{(p)}\big) - (X_1 - X_2)\big\|_2^\beta\big]$$

$$= \mathbb{E}\left[\left(\sum_{i=1}^p \int_{\Delta_i} \big(\Delta X_1(t, t_i] - \Delta X_2(t, t_i]\big)^2\,dt\right)^{\beta/2}\right]$$

$$\le \left(\sum_{i=1}^p \int_{\Delta_i} \mathrm{var}\big(\Delta X_1(t, t_i] - \Delta X_2(t, t_i]\big)\,dt\right)^{\beta/2}$$

$$= \left(\sum_{i=1}^p \int_{\Delta_i} \big(\mathrm{var}\big(\Delta X_1(t, t_i]\big) + \mathrm{var}\big(\Delta X_2(t, t_i]\big)\big)\,dt\right)^{\beta/2} \le c\delta_n^{\gamma_X \beta/2}. \tag{7.4}$$

The last step follows from (A1). If $1 < \beta < 2$, we use the inequality $|x^\beta - y^\beta| \le \beta(x \vee y)^{\beta-1} \times |y - x|$ for positive x, y and Hölder's inequality to obtain

$$\mathbb{E}\big[\big|\big\|X_1^{(p)} - X_2^{(p)}\big\|_2^\beta - \big\|X_1 - X_2\big\|_2^\beta\big|\big]$$

$$\le c\mathbb{E}\big[\big(\big\|X_1^{(p)} - X_2^{(p)}\big\|_2^{\beta-1} \vee \big\|X_1 - X_2\big\|_2^{\beta-1}\big)\big|\big\|X_1^{(p)} - X_2^{(p)}\big\|_2 - \big\|X_1 - X_2\big\|_2\big|\big]$$

$$\le c\mathbb{E}\big[\big(\big\|X_1^{(p)} - X_2^{(p)}\big\|_2^{\beta-1} \vee \big\|X_1 - X_2\big\|_2^{\beta-1}\big)\big\|\big(X_1^{(p)} - X_2^{(p)}\big) - (X_1 - X_2)\big\|_2\big]$$

$$\le c\big(\mathbb{E}\big[\big\|X_1^{(p)} - X_2^{(p)}\big\|_2^2 \vee \big\|X_1 - X_2\big\|_2^2\big]\big)^{(\beta-1)/2}$$

$$\quad \times \big(\mathbb{E}\big[\big\|\big(X_1^{(p)} - X_1\big) - \big(X_2^{(p)} - X_2\big)\big\|_2^{2/(3-\beta)}\big]\big)^{(3-\beta)/2}$$

$$= cP_1 P_2. \tag{7.5}$$

Since $(3 - \beta)^{-1} < 1$ the same arguments as in the case $0 < \beta < 1$ yield $P_2 \le c\delta_n^{\gamma_X/2}$. Moreover, we have

$$P_1^{2/(\beta-1)} \le \mathbb{E}\big[\big\|X_1^{(p)} - X_2^{(p)}\big\|_2^2\big] + \mathbb{E}\big[\big\|X_1 - X_2\big\|_2^2\big] = P_{11} + P_{12}.$$

It follows from Remark 2.1 that $P_{12} < \infty$ and a similar argument yields $P_{11} < \infty$.

Summarizing the previous bounds for $0 < \beta < 2$ under (A1), we have

$$\mathbb{E}[I_{11}] \le c\delta_n^{(\gamma_X \wedge \gamma_Y)(\beta \wedge 1)/2}.$$

Now we turn to I_2. Observe that

$$|I_2| \le \frac{1}{n^2} \sum_{k,l=1}^n \left| \|X_k^{(p)} - X_l^{(p)}\|_2^\beta - \|X_k - X_l\|_2^\beta \right| \frac{1}{n^2} \sum_{k,l=1}^n \|Y_k^{(p)} - Y_l^{(p)}\|_2^\beta$$

$$+ \frac{1}{n^2} \sum_{k,l=1}^n \|X_k - X_l\|_2^\beta \frac{1}{n^2} \sum_{k,l=1}^n \left| \|Y_k^{(p)} - Y_l^{(p)}\|_2^\beta - \|Y_k - Y_l\|_2^\beta \right|,$$

and a similar bound exists for $|I_3|$. The same arguments as above yield

$$\mathbb{E}\big[|I_2 + I_3|\big] \le c\delta_n^{(\gamma_X \wedge \gamma_Y)(\beta \wedge 1)/2}.$$

We omit further details.

Next, assume that (X, Y) have finite βth moment for some $\beta \in (0, 2)$. We follow the patterns of the proof in the finite variance case. We start by bounding $\mathbb{E}[|I_1|]$. First, assume $\beta \in (0, 1]$. Following (7.4), we have by (B2),

$$\mathbb{E}\left[\left(\sum_{i=1}^p \int_{\Delta_i} \big(\Delta X_1(t, t_i) - \Delta X_2(t, t_i)\big)^2 \, dt \right)^{\beta/2} \right]$$

$$\le c \sum_{i=1}^p |\Delta_i|^{\beta/2} \mathbb{E}\left[\max_{t \in \Delta_i} |\Delta X(t, t_i)|^\beta \right] \le cp\delta_n^{\beta/2 + \gamma_X}.$$

Now assume $1 < \beta < 2$. Following (7.5), we have by Hölder's inequality,

$$\mathbb{E}\big[\big| \|X_1^{(p)} - X_2^{(p)}\|_2^\beta - \|X_1 - X_2\|_2^\beta \big| \big]$$

$$\le c\mathbb{E}\big[\big(\|X_1^{(p)} - X_2^{(p)}\|_2^{\beta-1} \vee \|X_1 - X_2\|_2^{\beta-1} \big) \|(X_1^{(p)} - X_1) - (X_2^{(p)} - X_2)\|_2 \big]$$

$$\le c\big(\mathbb{E}\big[\|X_1^{(p)} - X_2^{(p)}\|_2^\beta \vee \|X_1 - X_2\|_2^\beta \big] \big)^{(\beta-1)/\beta}$$

$$\times \big(\mathbb{E}\big[\|(X_1^{(p)} - X_1) - (X_2^{(p)} - X_2)\|_2^\beta \big] \big)^{1/\beta}$$

$$= c\widetilde{P}_1 \widetilde{P}_2.$$

Proceeding as for $0 < \beta < 1$, we have

$$\widetilde{P}_2 = \big(\mathbb{E}\big[\|(X_1^{(p)} - X_1) - (X_2^{(p)} - X_2)\|_2^\beta \big] \big)^{1/\beta} \le c\big(p\delta_n^{\beta/2 + \gamma_X} \big)^{1/\beta}.$$

We also have

$$\widetilde{P}_1^{\beta/(\beta-1)} \le \mathbb{E}\big[\|X_1^{(p)} - X_2^{(p)}\|_2^\beta \big] + \mathbb{E}\big[\|X_1 - X_2\|_2^\beta \big].$$

The right-hand side is finite by assumption (B1). Collecting bounds for $0 < \beta < 2$, we arrive at

$$\mathbb{E}\big[|I_1|\big] \le c\big(p\delta_n^{\beta/2 + \gamma_X \wedge \gamma_Y} \big)^{1 \wedge \beta^{-1}}.$$

The quantities $\mathbb{E}[|I_i|]$, $i = 2, 3$, can be bounded in a similar way. $\qquad\square$

Now we can finish the proof of the first two parts of Theorem 3.1. We assume that either (A1) or [(B1), (B2) and $p\delta_n^{\beta/2+\gamma_X\wedge\gamma_Y} \to 0$] are satisfied. Under these assumptions, it follows from Proposition 7.1 that $T_{n,\beta}(X, Y) - T_{n,\beta}(X^{(p)}, Y^{(p)}) \xrightarrow{\mathbb{P}} 0$. The quantity $T_{n,\beta}(X, Y)$ can be written as a V-statistic of order 4 of the sample $((X_i, Y_i))_{i=1,\dots,n}$; see the Appendix. (Lyons [12] used a V-statistics of order 6. The higher order leads to a higher numerical complexity for the calculation of the bootstrap quantities.) Since X, Y are assumed independent and $\mathbb{E}[\|X\|_2^\beta] + \mathbb{E}[\|Y\|_2^\beta] < \infty$ (see Remark 2.1) we may apply the strong law of large numbers to the V-statistic $T_{n,\beta}(X, Y)$ implying that

$$T_{n,\beta}(X, Y) \xrightarrow{\text{a.s.}} T_\beta(X, Y) = 0. \tag{7.6}$$

Hence the first parts of the theorem follow.

Under the corresponding growth conditions (A2) and (B4) on $\delta_n \to 0$, Proposition 7.1 also yields $n(T_{n,\beta}(X, Y) - T_{n,\beta}(X^{(p)}, Y^{(p)})) \xrightarrow{\mathbb{P}} 0$. Then we can use the fact that the V-statistic $T_{n,\beta}(X, Y)$ is degenerate of order 1 to conclude that $nT_{n,\beta}(X, Y)$ converges in distribution to a series of independent weighted χ^2-distributed random variables, and $nT_{n,\beta}(X^{(p)}, Y^{(p)})$ has the same weak limit; we refer to Arcones and Giné [1] and Serfling [18] for general limit theory on U- and V-statistics.

The proof of the following two parts of the theorem is given in Section S.1 of the Supplementary Material [7].

Remark 7.2. Following the aforementioned arguments, the strong law of large numbers (7.6) remains valid if X and Y are dependent and the corresponding moments in the definition of $T_\beta(X, Y)$ are finite. In this case $T_{n,\beta}(X, Y)$ is a non-degenerate V-statistic and it follows from the Hoeffding decomposition that $(\sqrt{n}(T_{n,\beta}(X, Y) - T_\beta(X, Y)))$ converges to a normal distribution provided sufficiently high moments of (X, Y) are satisfied.

Appendix: The sample distance covariance as a degenerate V-statistic

We assume that $Z_i = (X_i, Y_i)$, $i = 1, 2, \dots$, is an i.i.d. sequence with generic element (X, Y) whose components are Riemann square-integrable on $[0, 1]$, and $\mathbb{E}[\|X\|_2^\beta + \|Y\|_2^\beta + \|X\|_2^\beta\|Y\|_2^\beta] < \infty$ and for some $\beta \in (0, 2)$. Under the assumption of independence on X, Y Lyons [12,13] proved that $T_{n,\beta}(X, Y)$ has representation as a V-statistic of order 6 with degenerate kernel of order 1. In what follows, we will indicate that it can be written as a V-statistic of order 4 with symmetric degenerate kernel of order 1. This fact is useful for improving upon the complexity of the numerical approximation of the sample distance correlation and its bootstrap version.

We start with the kernel

$$f\big((x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)\big) \quad \big(=: f(z_1, z_2, z_3, z_4)\big)$$
$$= \|x_1 - x_2\|_2^\beta \|y_1 - y_2\|_2^\beta + \|x_1 - x_2\|_2^\beta \|y_3 - y_4\|_2^\beta - 2\|x_1 - x_2\|_2^\beta \|y_1 - y_3\|_2^\beta.$$

From this representation, it is obvious that

$$T_{n,\beta}(X, Y) = \frac{1}{n^4} \sum_{1 \leq i,j,k,l \leq n} f(Z_i, Z_j, Z_k, Z_l).$$

Then one can define the corresponding symmetric kernel via the usual symmetrization as

$$h(z_1, z_2, z_3, z_4) = \frac{1}{24} \sum_{(l_1, l_2, l_3, l_4) \text{ permutation of } (1,2,3,4)} f(z_{l_1}, z_{l_2}, z_{l_3}, z_{l_4}). \tag{A.1}$$

It is not difficult to see that the kernel h is at least 1-degenerate, by showing that, under the null hypothesis of independence of X and Y,

$$\mathbb{E}\big[f(z_1, Z_2, Z_3, Z_4)\big] + \mathbb{E}\big[f(Z_2, z_1, Z_3, Z_4)\big] + \mathbb{E}\big[f(Z_2, Z_3, z_1, Z_4)\big]$$
$$+ \mathbb{E}\big[f(Z_2, Z_3, Z_4, z_1)\big]$$
$$= 0.$$

Still under the null hypothesis of independence of X and Y,

$$\mathbb{E}\big[h\big(z_1, z_2, (X_3, Y_3), (X_4, Y_4)\big)\big]$$
$$= \frac{1}{6}\big(\|x_1 - x_2\|_2^\beta + \mathbb{E}\big[\|X_1 - X_2\|_2^\beta\big] - \mathbb{E}\big[\|x_1 - X\|_2^\beta\big] - \mathbb{E}\big[\|x_2 - X\|_2^\beta\big]\big)$$
$$\times \big(\|y_1 - y_2\|_2^\beta + \mathbb{E}\big[\|Y_1 - Y_2\|_2^\beta\big] - \mathbb{E}\big[\|y_1 - Y\|_2^\beta\big] - \mathbb{E}\big[\|y_2 - Y\|_2^\beta\big]\big),$$

and the right-hand side is not constant. Hence, the kernel h is precisely 1-degenerate. In summary:

Lemma A.1. *If X, Y are independent and $\mathbb{E}[\|X\|_2^\beta + \|Y\|_2^\beta] < \infty$ for some $\beta \in (0, 2)$ then $T_{n,\beta}(X, Y)$ has representation as a V-statistic with a symmetric kernel h of order 4 which is 1-degenerate. Moreover, the corresponding U-statistic $\widetilde{T}_{n,\beta}(X, Y)$, which is obtained from $T_{n,\beta}(X, Y)$ by restricting the summation to indices (i_1, i_2, i_3, i_4) with mutually distinct components, satisfies the relation that as $n \to \infty$*

$$n\big(T_{n,\beta}(X, Y) - \widetilde{T}_{n,\beta}(X, Y)\big) \xrightarrow{\mathbb{P}} \mathbb{E}\big[\|X_1 - X_2\|_2^\beta\big]\mathbb{E}\big[\|Y_1 - Y_2\|_2^\beta\big]. \tag{A.2}$$

Indeed, observe that $\Delta_n = T_{n,\beta} - \widetilde{T}_{n,\beta}$ is based on summation of the kernel h over indices (i_1, i_2, i_3, i_4) for which at least two components coincide. If more than 2 indices coincide the number of these summands in Δ_n is of the order $O(n^2)$. However, the normalization in $n\Delta_n$ is of the order n^3. Therefore, the sum of these terms is negligible as $n \to \infty$. Finally, the part of the sum corresponding to the case when exactly two indices coincide and the other indices are different, can be written as a U-statistic of order 3. By the law of large numbers, this U-statistic converges a.s. to $\mathbb{E}[\|X_1 - X_2\|_2^\beta]\mathbb{E}[\|Y_1 - Y_2\|_2^\beta]$.

Remark A.2. The additional moment assumption on $h(Z_{i_1}, Z_{i_2}, Z_{i_3}, Z_{i_4})$, $1 \leq i_1 \leq i_2 \leq i_3 \leq i_4 \leq 4$, required in Corollary 5.2 is satisfied for our kernel. Note that it suffices to consider the non-symmetric kernel f, and to show that $\mathbb{E}[(f(Z_{i_1}, Z_{i_2}, Z_{i_3}, Z_{i_4}))^2] < \infty$, for all indices $1 \leq i_1, \ldots, i_4 \leq 4$. For our specific kernel, this condition reads

$$\mathbb{E}\big[\big(\|X_{i_1} - X_{i_2}\|^\beta \big[\|Y_{i_1} - Y_{i_2}\|^\beta + \|Y_{i_3} - Y_{i_4}\|^\beta - 2\|Y_{i_1} - Y_{i_3}\|^\beta\big]\big)^2\big] < \infty,$$

and this holds under the moment conditions made in this paper.

Acknowledgments

Supplementary Material

Supplement to "Distance covariance for discretized stochastic processes" (DOI: 10.3150/20-BEJ1206SUPP; .pdf). The online supplement contains the proofs of parts (3) and (4) of Theorem 3.1, and further details of the proof of the bootstrap consistency. In addition, we study the asymptotic behaviour of the distance correlation under the alternative when X and Y are dependent.

References

[1] Arcones, M.A. and Giné, E. (1992). On the bootstrap of U and V statistics. *Ann. Statist.* **20** 655–674. MR1165586 https://doi.org/10.1214/aos/1176348650

[2] Bickel, P.J. and Freedman, D.A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217. MR0630103

[3] Billingsley, P. (1968). *Convergence of Probability Measures*. New York: Wiley. MR0233396

[4] Davis, R.A., Matsui, M., Mikosch, T. and Wan, P. (2018). Applications of distance correlation to time series. *Bernoulli* **24** 3087–3116. MR3779711 https://doi.org/10.3150/17-BEJ955

[5] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory*: *An Introduction*. *Springer Series in Operations Research and Financial Engineering*. New York: Springer. MR2234156 https://doi.org/10.1007/0-387-34471-3

[6] de Haan, L. and Lin, T. (2001). On convergence toward an extreme value distribution in *C*[0, 1]. *Ann. Probab.* **29** 467–483. MR1825160 https://doi.org/10.1214/aop/1008956340

[7] Dehling, H., Matsui, M., Mikosch, T., Samorodnitsky, G. and Tafakori, L. (2020). Supplement to "Distance covariance for discretized stochastic processes." https://doi.org/10.3150/20-BEJ1206SUPP

[8] Dehling, H. and Mikosch, T. (1994). Random quadratic forms and the bootstrap for *U*-statistics. *J. Multivariate Anal.* **51** 392–413. MR1321305 https://doi.org/10.1006/jmva.1994.1069

[9] Feuerverger, A. (1993). A consistent test for bivariate dependence. *Int. Stat. Rev.* **61** 419–433.

[10] Hoffmann-Jørgensen, J. (1994). *Probability with a View Toward Statistics*, *Vol. I*. *Chapman & Hall Probability Series*. New York: CRC Press. MR1278485

[11] Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. *Springer Series in Statistics*. New York: Springer. MR2920735 https://doi.org/10.1007/978-1-4614-3655-3

[12] Lyons, R. (2013). Distance covariance in metric spaces. *Ann. Probab.* **41** 3284–3305. MR3127883 https://doi.org/10.1214/12-AOP803

[13] Lyons, R. (2018). Errata to "Distance covariance in metric spaces" [MR3127883]. *Ann. Probab.* **46** 2400–2405. MR3813995 https://doi.org/10.1214/17-AOP1233

[14] Marquardt, T. (2006). Fractional Lévy processes with an application to long memory moving average processes. *Bernoulli* **12** 1099–1126. MR2274856 https://doi.org/10.3150/bj/1165269152

[15] Matsui, M., Mikosch, T. and Samorodnitsky, G. (2017). Distance covariance for stochastic processes. *Probab. Math. Statist.* **37** 355–372. MR3745391 https://doi.org/10.19195/0208-4147.37.2.9

[16] Samorodnitsky, G. (2016). *Stochastic Processes and Long Range Dependence*. *Springer Series in Operations Research and Financial Engineering*. Cham: Springer. MR3561100 https://doi.org/10.1007/978-3-319-45575-4

[17] Samorodnitsky, G. and Taqqu, M.S. (1994). *Stable Non-Gaussian Random Processes*: *Stochastic Models with Infinite Variance*. *Stochastic Modeling*. New York: CRC Press. MR1280932

[18] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. *Wiley Series in Probability and Mathematical Statistics*. New York: Wiley. MR0595165

[19] Székely, G.J. and Rizzo, M.L. (2009). Brownian distance covariance. *Ann. Appl. Stat.* **3** 1236–1265. MR2752127 https://doi.org/10.1214/09-AOAS312

[20] Székely, G.J. and Rizzo, M.L. (2013). The distance correlation *t*-test of independence in high dimension. *J. Multivariate Anal.* **117** 193–213. MR3053543 https://doi.org/10.1016/j.jmva.2013.02.012

[21] Székely, G.J. and Rizzo, M.L. (2014). Partial distance correlation with methods for dissimilarities. *Ann. Statist.* **42** 2382–2412. MR3269983 https://doi.org/10.1214/14-AOS1255

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