June 25, 2018

DISTANCE COVARIANCE FOR DISCRETIZED STOCHASTIC PROCESSES

HEROLD DEHLING, MUNEYA MATSUI, THOMAS MIKOSCH, GENNADY SAMORODNITSKY,
AND LALEH TAFAKORI

ABSTRACT. Given an iid 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.

1. INTRODUCTION

1.1. **Distance covariance and distance correlation for vectors.** In a series of papers, Székely et al. [14, 15, 16, 17] introduced *distance covariance* and *distance correlation.* They are measures of the dependence between two vectors $\mathbf{X}$ and $\mathbf{Y}$, possibly with different dimensions. These measures have the desirable property that they are zero if and only if $\mathbf{X}$ and $\mathbf{Y}$ are independent. This is in contrast to many other dependence measures where one can only make statements about certain aspects of the dependence between $\mathbf{X}$ and $\mathbf{Y}$. For example, the correlation and covariance between two real-valued random variables X and Y allow one to make statements about their linear dependence; only in the case of joint Gaussianity of (X, Y) their correlation determines the full dependence structure between X and Y.

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

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

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 [6] in the univariate case and, later, Székely et al. [14, 15, 16, 17] 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* $\mathbf{X}$ *and* $\mathbf{Y}$ 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

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

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 iid copies $(\mathbf{X}_i, \mathbf{Y}_i)$, $i = 1, 2, \ldots$, of $(\mathbf{X}, \mathbf{Y})$ we have

$$
\begin{aligned}
T_\beta(\mathbf{X}, \mathbf{Y}) &= \mathbb{E}[|\mathbf{X}_1 - \mathbf{X}_2|^\beta |\mathbf{Y}_1 - \mathbf{Y}_2|^\beta] + \mathbb{E}[|\mathbf{X}_1 - \mathbf{X}_2|^\beta]\mathbb{E}[|\mathbf{Y}_1 - \mathbf{Y}_2|^\beta] \\
&\quad - 2\,\mathbb{E}[|\mathbf{X}_1 - \mathbf{X}_2|^\beta |\mathbf{Y}_1 - \mathbf{Y}_3|^\beta] \,.
\end{aligned}
$$
(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, i.e., 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 $R_\beta(\mathbf{X}, \mathbf{Y})$:

$$
\begin{aligned}
T_{n,\beta}(\mathbf{X}, \mathbf{Y}) &= \frac{1}{n^2} \sum_{k,l=1}^n |\mathbf{X}_k - \mathbf{X}_l|^\beta |\mathbf{Y}_k - \mathbf{Y}_l|^\beta + \frac{1}{n^2} \sum_{k,l=1}^n |\mathbf{X}_k - \mathbf{X}_l|^\beta \frac{1}{n^2} \sum_{k,l=1}^n |\mathbf{Y}_k - \mathbf{Y}_l|^\beta \\
&\quad - 2 \frac{1}{n^3} \sum_{k,l,m=1}^n |\mathbf{X}_k - \mathbf{X}_l|^\beta |\mathbf{Y}_k - \mathbf{Y}_m|^\beta \,, \\
R_{n,\beta}(\mathbf{X}, \mathbf{Y}) &= \frac{T_{n,\beta}(\mathbf{X}, \mathbf{Y})}{\sqrt{T_{n,\beta}(\mathbf{X}, \mathbf{X})}\sqrt{T_{n,\beta}(\mathbf{Y}, \mathbf{Y})}} \,.
\end{aligned}
$$

The quantity $T_{n,\beta}(\mathbf{X}, \mathbf{Y})$ is a V-statistic; cf. Székely et al. [14], Lyons [9]. Therefore standard theory yields a.s. consistency,

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

under suitable moment conditions; see Hoffmann–Jørgensen [7], 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 $n\,T_{n,\beta}(\mathbf{X}, \mathbf{Y})$ to a weighted sum of iid χ^2-variables; see Serfling [18], Lyons [9], Arcones and Giné [1]. Moreover, V-statistics theory also ensures that $T_{n,\beta}(\mathbf{X}, \mathbf{X}) \overset{\text{a.s.}}{\to} T_\beta(\mathbf{X}, \mathbf{X})$ and $T_{n,\beta}(\mathbf{Y}, \mathbf{Y}) \overset{\text{a.s.}}{\to} T_\beta(\mathbf{Y}, \mathbf{Y})$. 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 [16] considered the situation when $\mathbf{X}$ and $\mathbf{Y}$ are independent and have iid 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. [8] 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 $[0, 1]$. Via simulations the resulting estimator was compared with the distance correlation $R_{n,\beta}(\mathbf{X}, \mathbf{Y})$ where the components of the iid 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 [8] and [16] worked under quite distinct conditions. Székely and Rizzo [16] considered vectors $\mathbf{X}$ and $\mathbf{Y}$ with iid components whose dimensions increase to infinity for a fixed sample size n. In [8], $\mathbf{X}$ 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 [16] and [8]. We consider two processes X and Y on $[0, 1]$, which we assume to be stochastically continuous, measurable and bounded. In contrast to [8],

- 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

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

- 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 $[0, 1]$,

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

We consider a vector of weighted discretizations

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

and define

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

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

$$
\begin{aligned}
|\mathbf{Z}_p - \mathbf{Z}'_p|^2 &= \sum_{i=1}^{p} (Z(t_i) - Z'(t_i))^2 |\Delta_i| = \|Z^{(p)} - (Z')^{(p)}\|_2^2 \\
&\to \int_0^1 (Z(t) - Z'(t))^2 \, dt = \|Z - Z'\|_2^2, \quad p \to \infty,
\end{aligned}
$$

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 iid 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

$$
\begin{aligned}
T_\beta(X, Y) &= \mathbb{E}\left[\|X_1 - X_2\|_2^\beta \|Y_1 - Y_2\|_2^\beta \right] + \mathbb{E}\left[\|X_1 - X_2\|_2^\beta \right] \mathbb{E}\left[\|Y_1 - Y_2\|_2^\beta \right] \\
(1.4) &\qquad - 2\,\mathbb{E}\left[\|X_1 - X_2\|_2^\beta \|Y_1 - Y_3\|_2^\beta \right],
\end{aligned}
$$

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

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

$$-2\frac{1}{n^3}\sum_{k,l,m=1}^{n}\|X_k-X_l\|_2^\beta\|Y_k-Y_m\|_2^\beta$$

$$(1.5)\qquad\qquad =:\ \mathrm{I}_1+\mathrm{I}_3-2\mathrm{I}_2\,.$$

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

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

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.** 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 $\big((X_i^{(p)},Y_i^{(p)})\big)_{i=1,\dots,n}$ consisting of discretizations taken from an iid sequence $((X_i,Y_i))_{i=1,2,\dots}$ on the same partition $(t_i)_{i=0,\dots,p}$ 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 $\mathbf{X}=(X(j/p))_{j=1,\dots,p}$ and $\mathbf{Y}=(Y(j/p))_{j=1,\dots,p}$.

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

$$(1.6)\qquad\qquad n\left(T_{n,\beta}(X^{(p)},Y^{(p)})-T_{n,\beta}(X,Y)\right)\overset{\mathbb{P}}{\to}0\,,\qquad n\to\infty\,,$$

provided $\delta_n\to0$ 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 $n\,T_{n,\beta}(X,Y)$ to derive the distributional limit of $n\,T_{n,\beta}(X^{(p)},Y^{(p)})$. This limit has a weighted χ^2-distribution which is not easily evaluated. We will show that bootstrap versions of the degenerate V-statistics $n\,T_{n,\beta}(X,Y)$ and $n\,T_{n,\beta}(X^{(p)},Y^{(p)})$ are close in the sense of Mallows metrics and have the same distributional limit as $n\,T_{n,\beta}(X,Y)$.

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. 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

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

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

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

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

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

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.*

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

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

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

(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

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

2.3. **Discussion of the conditions and examples.**

Remark 2.1. In the proofs we will need the conditions

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

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

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

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

Remark 2.2. In the case of an equidistant partition we have $\delta_n = 1/p$. Then the growth condition (A2) reads as

(2.2) $$\frac{p}{n^{\frac{2}{(\gamma_X \wedge \gamma_Y)(\beta \wedge 1)}}} \to \infty\,, \qquad n \to \infty\,,$$

while (B4) takes on the form

(2.3) $$\frac{p}{n^{\frac{\beta}{(\beta/2+\gamma_X \wedge \gamma_Y-1)(\beta \wedge 1)}}} \to \infty\,, \qquad n \to \infty\,,$$

provided one can ensure that $\beta/2 + \gamma_X \wedge \gamma_Y > 1$. The message from (2.2) is that we need to choose p the larger the smaller $\gamma_X \wedge \gamma_Y$ is, i.e., the rougher the sample paths. Similarly, for $\beta < 1$, p needs to be chosen the larger the smaller β is. Similar comments apply to (2.3).

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 $c_X > 0$,

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

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 $\mathbb{E}\big[|X(1)|^{2(2\beta-1)}\big] < \infty$, 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*

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

where L is a two-sided Lévy process on $\mathbb{R}$ with mean zero and finite variance, introduced in Marquardt [11]. 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, i.e., 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

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

Hence, if $c_X = \sup_{x \in [0,1]} \mathbb{E}[Z_X^2(x)] < \infty$, then

$$\mathrm{var}\big(\Delta X(s,t]\big) \le c_X\,(t-s)\,,$$

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 we conclude from (2.2) that (A2) holds if

$$(2.4) \qquad\qquad\qquad \frac{p}{n^{\frac{2}{\beta \wedge 1}}} \to \infty\,, \qquad n \to \infty\,.$$

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 [13]. 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 [12], 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 $t \in [0,1]$,

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

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

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

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

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

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

and, hence,

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

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, as the proof of Lemma 7.3 shows, 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 $\big[$(B1),(B2) and $p\,\delta_n^{\beta/2+\gamma_X \wedge \gamma_Y} \to 0\big]$ 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 iid sequence of standard normal random variables (N_i), a constant c, and a square summable sequence (λ_i).

(3) *If either* (A1),(A3) *or* $\left[\beta \in (0,1) \text{ and } (B1)\text{-}(B3) \text{ and } p\,\delta_n^{\beta+\gamma_X'\wedge\gamma_Y'} \to 0\right]$ *hold then*

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

(4) *If either* (A1)-(A3) *or* $\left[\beta \in (0,1) \text{ and } (B1)\text{-}(B4) \text{ and } p\,\delta_n^{\beta+\gamma_X'\wedge\gamma_Y'} \to 0\right]$ *hold then*

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

for an iid 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.

Remark 3.2. 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 [9], 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 $T_{n,\beta}(X^{(p)}, Y^{(p)}) \xrightarrow{\mathbb{P}} T_\beta(X,Y) = 0$ 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 [9], 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

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

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* $\mathbb{E}[\|X\|_2^\beta] + \mathbb{E}[\|Y\|_2^\beta] < \infty$. *Let* Y' *be a copy of* Y *independent of everything else. Then*

$$c_0^2\,T_\beta(X,Y) =$$
$$\int_{\mathbb{R}^2} |st|^{-(1+\beta/2)}\mathbb{E}\,\mathbb{E}\left[e^{is\int X(u)\,dB_1(u)}e^{it\int Y(u)\,dB_2(u)} - e^{is\int X(u)\,dB_1(u)}e^{it\int Y'(u)\,dB_2(u)}\,\Big|\,\mathcal{F}_B\right]^2 ds\,dt$$
(4.1)

where

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

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

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

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

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

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

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

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

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

By change of variables,

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

Thus we obtain

$$\mathbb{E}\big[\|X - X'\|_2^{\beta}\|Y - Y'\|_2^{\beta} + \|X - X'\|_2^{\beta}\|Y'' - Y'''\|_2^{\beta} - 2\|X - X'\|_2^{\beta}\|Y - Y''\|_2^{\beta}\big] = T_{\beta}(X, Y)\,.$$

$\square$

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

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

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 $\mathbb{E}[\|X\|_2^{\beta}] + \mathbb{E}[\|Y\|_2^{\beta}] < \infty$, 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 $\big(a(t),\, 0 \le t \le 1\big)$ and $\big(b(t),\, 0 \le t \le 1\big)$ be functions in $L^2[0, 1]$, and

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

Since the law of the bivariate process

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

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

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

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

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

are independent.

For every fixed realization of the processes X and Y,

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

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 $\big(a(t),\, 0 \leq t \leq 1\big)$ and $\big(b(t),\, 0 \leq t \leq 1\big)$, we conclude that the sums

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

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

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

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 3.2 that the limit distribution of $n\, T_{n,\beta}(X, Y)$ is not available. Theorem 3.1 states that the discretization $n\, T_{n,\beta}(X^{(p)}, Y^{(p)})$ has the same asymptotic properties as $n\, T_{n,\beta}(X, Y)$ 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 $n\, T_{n,\beta}(X^{(p)}, Y^{(p)})$. 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 [5], 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 iid sequence with marginal laws $\mathcal{L}(X^{(1)}) = F$ and $\mathcal{L}(X^{(2)}) = G$, respectively. On the subset of probability measures on S,

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

we define the semi-metric

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

where the infimum is taken over all random elements $\left(X_1^{(1)}, \ldots, X_m^{(1)}, X_1^{(2)}, \ldots, X_m^{(2)} \right)$ in S^{2m} such that $(X_i^{(1)}, X_i^{(2)})$, $i = 1, \ldots, m$, are iid 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

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

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

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

where (Z_i) are iid with common law H. The proof of the following result is completely analogous to that of Lemma 2.2 in Dehling and Mikosch [5].

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

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

For an S-valued iid 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 iid 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 [5].

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 $1 \leq i_1 \leq \ldots \leq i_m \leq m$, 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 iid copies of the pair $(Z_n^*, \tilde{Z})$, which we denote by $(Z_{n1}^*, \tilde{Z}_1), \ldots, (Z_{nm}^*, \tilde{Z}_m)$. Then

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

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

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

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

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

By Lebesgue's dominated convergence theorem, we obtain

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

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

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

This finishes the proof. $\qquad\square$

In what follows, (Z_i) will stand for the iid 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 Appendix A, 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 $n T_n(X, Y)$ 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 iid 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, i.e., we will show that suitable bootstrap versions of $n T_{n,\beta}(X, Y)$ and $n T_{n,\beta}(X^{(p)}, Y^{(p)})$ 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

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

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

$$
\begin{aligned}
d_2\big(\mathcal{L}(U_n(Z)); \mathcal{L}(U_n(Z^{(p)}))\big) &\le c\,\big\{\mathbb{E}\big[\big(h(Z_1,\ldots,Z_4) - h(Z_1^{(p)},\ldots,Z_4^{(p)})\big)^2\big]\big\}^{1/2}\\
&= c\,\big\{\mathbb{E}\big[\big(f(Z_1,\ldots,Z_4) - f(Z_1^{(p)},\ldots,Z_4^{(p)})\big)^2\big]\big\}^{1/2}\\
&\le c\,\big(\mathbb{E}I_1^2 + \mathbb{E}I_2^2 + \mathbb{E}I_3^2\big)^{1/2}\,,
\end{aligned}
$$

where

$$
\begin{aligned}
I_1 &= \|X_1 - X_2\|_2\|Y_1 - Y_2\|_2 - \|X_1^{(p)} - X_2^{(p)}\|_2\|Y_1^{(p)} - Y_2^{(p)}\|_2\,,\\
I_2 &= \|X_1 - X_2\|_2\|Y_3 - Y_4\|_2 - \|X_1^{(p)} - X_2^{(p)}\|_2\|Y_3^{(p)} - Y_4^{(p)}\|_2\,,\\
I_3 &= \|X_1 - X_2\|_2\|Y_1 - Y_3\|_2 - \|X_1^{(p)} - X_2^{(p)}\|_2\|Y_1^{(p)} - Y_3^{(p)}\|_2\,.
\end{aligned}
$$

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

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

and

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

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

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. $\mathbb{E}[|X(t) - X(s)|^4] \le c\,|t - s|^{\widetilde{\gamma}_X}$ *and* $\mathbb{E}[|Y(t) - Y(s)|^4] \le c\,|t - s|^{\widetilde{\gamma}_Y}$ *hold.*
3a. $\sum_{n=1}^{\infty} \delta_n^{\gamma_X \wedge \gamma_Y} < \infty$.
3b. $\sum_{n=1}^{\infty} \big(\delta_n^{2(\gamma_X \wedge \gamma_Y)} + n^{-1}\delta_n^{\widetilde{\gamma}_X \wedge \widetilde{\gamma}_Y}\big) < \infty$.

If either 1, 2a, 3a *or* 1, 2a, 2b, 3b *hold then*

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

for a.e. realization of (Z_i).

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

$$
\begin{aligned}
d_{2,h}(n) &:= d_{2,h}\big(\mathcal{L}(Z^*), \mathcal{L}(Z^{(p)*})\big)\\
&\le \Big(\mathbb{E}_{F_n}\big[\big(h(Z_1^*, Z_2^*, Z_3^*, Z_4^*) - h(Z_1^{(p)*}, Z_2^{(p)*}, Z_3^{(p)*}, Z_4^{(p)*})\big)^2\big]\Big)^{1/2}\\
&= \frac{1}{n^2}\Big(\sum_{1 \le i_1, i_2, i_3, i_4 \le n} \big(h(Z_{i_1}, Z_{i_2}, Z_{i_3}, Z_{i_4}) - h(Z_{i_1}^{(p)}, Z_{i_2}^{(p)}, Z_{i_3}^{(p)}, Z_{i_4}^{(p)})\big)^2\Big)^{1/2}
\end{aligned}
$$

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

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

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

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

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

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

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,

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

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

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*
$$d_2\big(\mathcal{L}(U_n(Z)),\mathcal{L}(U_n(Z^{(p)*}))\big)\to 0$$
for a.e. realization of (Z_i).

For an application of the bootstrapped sample distance correlation $nR_n(X^{(p)},Y^{(p)})$ we still miss one step in the derivation of the bootstrap consistency: we also need to prove that the denominator quantities converge a.s.
$$T_n(X^{(p)},X^{(p)})\stackrel{\text{a.s.}}{\to}T(X,X)\quad\text{and}\quad T_n(Y^{(p)},Y^{(p)})\stackrel{\text{a.s.}}{\to}T(Y,Y)\,,\qquad n\to\infty\,.$$

In Lemma 7.4 we provide sufficient conditions for this to hold.

6. Simulations

In this section we illustrate the theoretical results in a small simulation study.

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

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.
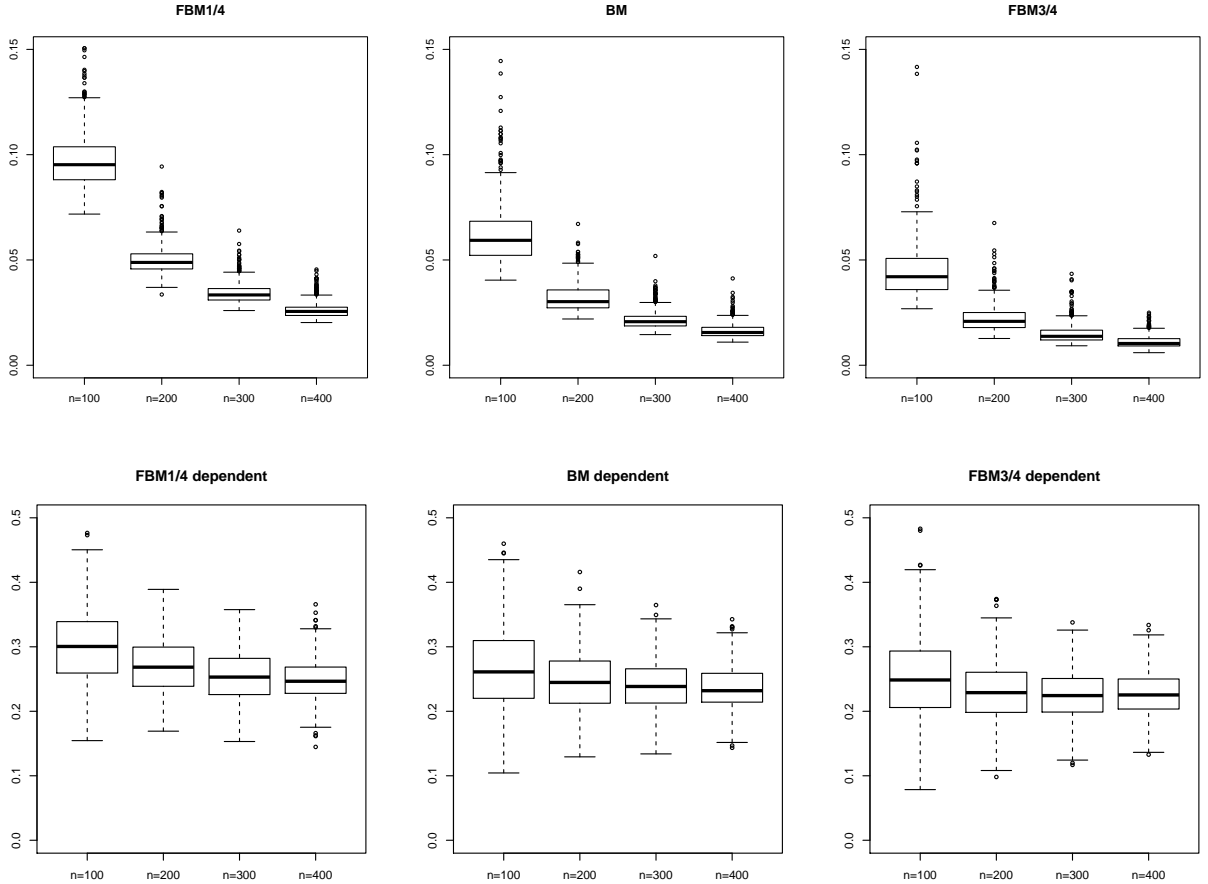

FIGURE 1. Boxplots for $R_n(X^{(p)}, Y^{(p)})$ simulated fBMs X, Y with $H = 1/4, 1/2, 3/4$ (from left to right), $p = 100$ and increasing sample sizes n. Top: iid fBMs X, Y. Each boxplot is based on 500 replications. Bottom: identically distributed fBMs X, Y with correlation $\rho = 0.5$. Each boxplot is based on 300 replications.

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 iid geometric BMs (left), X, Y are iid 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

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

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. [13, 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.
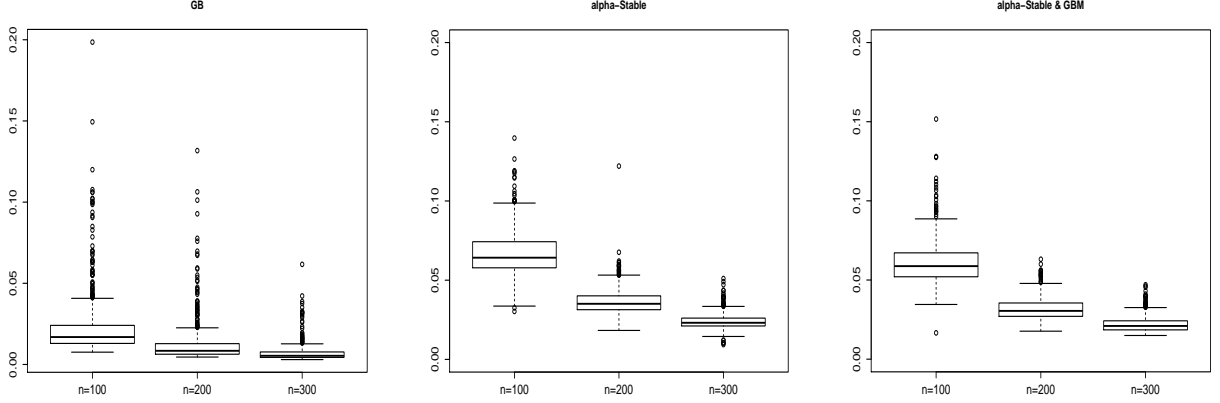

FIGURE 2. Boxplots for $R_n(X^{(p)}, Y^{(p)})$ for simulated independent non-Gaussian processes X, Y, $p = 100$ and increasing sample size n. Each boxplot is based on 500 replications. Left: iid geometric BMs X, Y. Middle: iid α-stable Lévy motions X, Y. Right: independent geometric BM X and α-stable Lévy motion Y.

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 $p = 300$ (middle) while X, Y are independent BMs: there is hardly any difference between the left and middle graphs for a given n. In the right graph we choose iid α-stable Lévy motions X, Y with the same parameters as before. 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 – in view of the definition of the distance correlation and the independence of $X^{(p)}$ and $Y^{(p)}$ for any p. However, it is perhaps unexpected that n and p may have similar size and still provide good approximations to zero.

In Figure 4 we visualize how the bootstrap works for the normalized sample distance correlations $nR_n(X^{(p)}, Y^{(p)})$ for iid fBMs X, Y. We show histograms based on 500 replications of $nR_n(X^{(p)}, Y^{(p)})$ and compare with the histograms based on 200 replications of the bootstrap version generated from a single sample. We see that the distributions of $nR_n(X^{(p)}, Y^{(p)})$ and its bootstrap version are close to each other and get more concentrated.
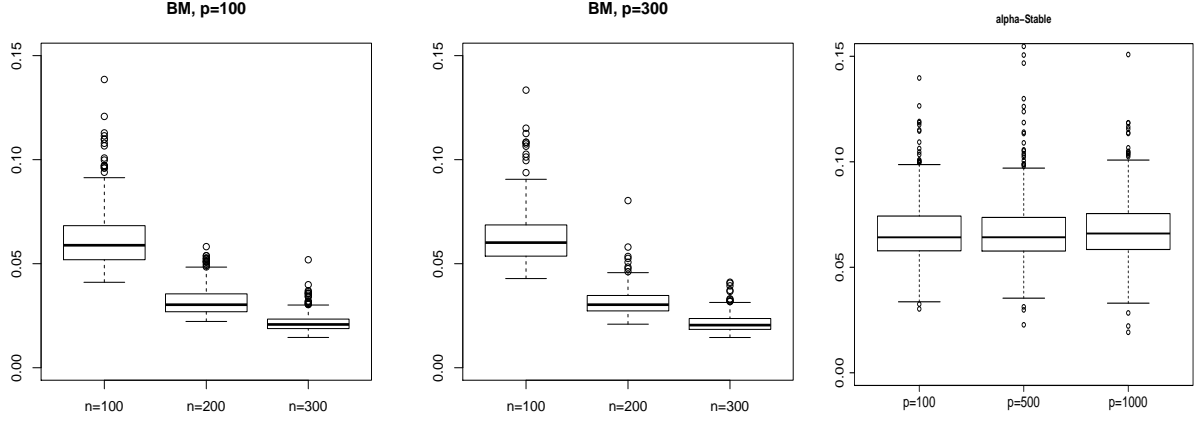
FIGURE 3. Boxplots for $R_n(X^{(p)}, Y^{(p)})$ for different p. Left and middle: X, Y are iid BMs. For each $p = 100$ (left) and $p = 300$ (middle) we take three distinct sample sizes $n = 100, 200, 300$. The boxplots are based on 300 replications. Right: X, Y are iid α-stable Lévy motions, $n = 100$ is fixed while $p = 100, 500, 1000$. The boxplots are based on 500 replications.

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 $n \geq 1$,*

$$\mathbb{E}\big[|T_{n,\beta}(X^{(p)}, Y^{(p)}) - T_{n,\beta}(X, Y)|\big] \leq c \, \delta_n^{(\gamma_X \wedge \gamma_Y)\,(\beta \wedge 1)/2} \,.$$

2. *If also* (B1),(B2) *hold then there is c such that*

$$\mathbb{E}\big[|T_{n,\beta}(X^{(p)}, Y^{(p)}) - T_{n,\beta}(X, Y)|\big] \leq c \, \big(p \, \delta_n^{(\beta/2 + \gamma_X \wedge \gamma_Y)}\big)^{(\beta \wedge 1)/\beta} \,.$$

Proof. We start with the decomposition

$$(7.1) \qquad T_{n,\beta}(X^{(p)}, Y^{(p)}) - T_{n,\beta}(X, Y) = I_1 + I_2 - 2I_3 \,,$$

where

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

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

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

(7.2)

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.
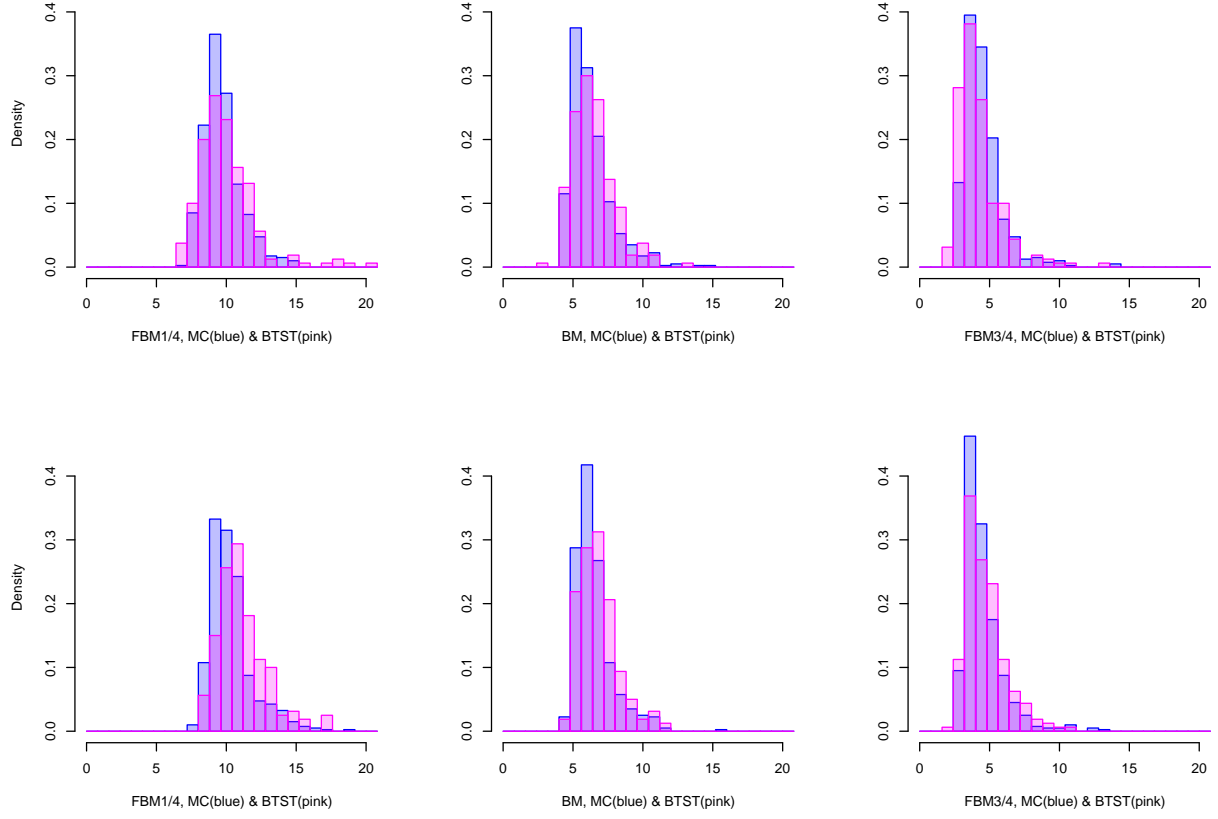
FIGURE 4. Comparison of histograms for $nR_n(X^{(p)}, Y^{(p)})$ based on Monte Carlo simulation (blue) and bootstrap (pink) for iid fBMs X, Y with $H = 1/4$, $H = 1/2$, $H = 3/4$ (from left to right). The sample size is $n = 100$ (top) and $n = 300$ (bottom) and $p = 100$. The histograms of $nR_n(X^{(p)}, Y^{(p)})$ and the bootstrap version are based on 500 and 200 replications, respectively.

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

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

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

(7.3) $$=: I_{11} + I_{12}.$$

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

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

By Lyapunov's inequality,

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

$$\leq \; c \Big(\int_0^1 \mathrm{var}(Y^{(p)}(t)) \, dt \Big)^{\beta/2} < \infty \, .$$

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

$$(7.4) \qquad \mathbb{E}\big[\, \|X_1^{(p)} - X_2^{(p)}\|_2^\beta - \|X_1 - X_2\|_2^\beta \, \big]$$

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

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

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

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

$$\leq \; c \, \delta_n^{\gamma_X \beta/2}.$$

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

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

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

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

$$(7.5)$$

$$\leq \; c \, \Big(\mathbb{E}\big[\|X_1^{(p)} - X_2^{(p)}\|_2^2 \vee \|X_1 - X_2\|_2^2\big]\Big)^{(\beta-1)/2} \Big(\mathbb{E}\big[\|(X_1^{(p)} - X_1) - (X_2^{(p)} - X_2)\|_2^{2/(3-\beta)}\big]\Big)^{(3-\beta)/2}$$

$$= \; c \, P_1 \, P_2 \, .$$

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

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

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

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

Now we turn to I_2. Observe that

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

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

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

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

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),

$$\mathbb{E}\Big[\Big(\sum_{i=1}^{p} \int_{\Delta_i} \big(\Delta X_1(t, t_i) - \Delta X_2(t, t_i)\big)^2 \, dt\Big)^{\beta/2}\Big] \leq c \sum_{i=1}^{p} |\Delta_i|^{\beta/2} \mathbb{E}\big[\max_{t \in \Delta_i} \Delta X(t, t_i)^{\beta}\big]$$
$$\leq c \, p \, \delta_n^{\beta/2 + \gamma_X} \, .$$

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

$$\mathbb{E}\big[\, \|X_1^{(p)} - X_2^{(p)}\|_2^{\beta} - \|X_1 - X_2\|_2^{\beta}\,\big]$$
$$\leq c \, \mathbb{E}\big[\big(\|X_1^{(p)} - X_2^{(p)}\|_2^{\beta-1} \vee \|X_1 - X_2\|_2^{\beta-1}\big) \, \|(X_1^{(p)} - X_1) - (X_2^{(p)} - X_2)\|_2\big]$$
$$\leq c \, \Big(\mathbb{E}\big[\|X_1^{(p)} - X_2^{(p)}\|_2^{\beta} \vee \|X_1 - X_2\|_2^{\beta}\big]\Big)^{(\beta-1)/\beta} \Big(\mathbb{E}\big[\|(X_1^{(p)} - X_1) - (X_2^{(p)} - X_2)\|_2^{\beta}\big]\Big)^{1/\beta}$$
$$(7.6) \quad = \quad c \, \widetilde{P}_1 \widetilde{P}_2 \, .$$

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

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

We also have

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

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

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

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 Appendix A. (Lyons [9] 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

$$(7.7) \qquad\qquad\qquad T_{n,\beta}(X, Y) \xrightarrow{\text{a.s.}} T_{\beta}(X, Y) = 0 \, .$$

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 $n \, T_{n,\beta}(X, Y)$ converges in distribution to a series of independent weighted χ^2-distributed random variables, and $n \, T_{n,\beta}(X^{(p)}, Y^{(p)})$ has the same weak limit; we refer to Arcones and Giné [1], Serfling [18] for general limit theory on U- and V-statistics.

Remark 7.2. Following the aforementioned arguments, the strong law of large numbers (7.7) 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.

Our next goal is to prove the last two parts of Theorem 3.1. They will follow if we can show consistency of $T_{n,\beta}(X^{(p)}, X^{(p)})$ and $T_{n,\beta}(Y^{(p)}, Y^{(p)})$. This is the content of the following lemma.

Lemma 7.3. *Assume the following conditions:*

1. *X is defined on $[0,1]$ and has Riemann square-integrable sample paths.*
2. *If X has a finite first moment X is centered.*
3. *$\delta_n \to 0$ as $n \to \infty$.*
4. *$\beta \in (0,2)$.*

Moreover, consider the following conditions:

(1) *X has finite second moment and there exist $\gamma_X > 0$ and $c > 0$ such that*

$$\text{(7.8)} \qquad \text{var}\big(X(s,t]\big) \le c\,|t-s|^{\gamma_X}, \qquad s < t.$$

If $\beta \in (1,2)$ we also assume

$$\text{(7.9)} \qquad \max_{0 \le t \le 1} \mathbb{E}[|X(t)|^{2(2\beta-1)}] < \infty.$$

(2) *For some $\beta \in (0,1)$,*

$$\text{(7.10)} \qquad \mathbb{E}\big[\max_{0 \le t \le 1} |X(t)|^{2\beta}\big] < \infty,$$

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

$$\text{(7.11)} \qquad \max_{i=1,\ldots,p} \mathbb{E}\big[\max_{t \in \Delta_i} |\Delta X(t,t_i)|^{2\beta}\big] \le c\,\delta_n^{\gamma_X'}.$$

If either (1) or (2) hold then

$$T_{n,\beta}(X^{(p)}, X^{(p)}) - T_{n,\beta}(X,X) \xrightarrow{\mathbb{P}} 0.$$

Moreover, we also have

$$\text{(7.12)} \qquad T_{n,\beta}(X^{(p)}, X^{(p)}) \xrightarrow{\mathbb{P}} T_\beta(X,X),$$

where

$$T_\beta(X,X) = \mathbb{E}\big[\|X_1 - X_2\|_2^{2\beta}\big] + \big(\mathbb{E}\big[\|X_1 - X_2\|_2^\beta\big]\big)^2 - 2\,\mathbb{E}\big[\|X_1 - X_2\|_2^\beta \|X_1 - X_3\|_2^\beta\big].$$

Note that since (7.8) and (7.9) are respectively implied by conditions (A1) and (A2), while (7.10) and (7.11) are implied by the condition (B3), the conditions of Lemma 7.3 are included in (3), (4) of Theorem 3.1.

Proof. We assume condition (1). We use the decomposition (7.1) and follow the lines of the proof of Proposition 7.1. In this case,

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

$$\text{(7.13)} \qquad I_2 = \Big(\frac{1}{n^2} \sum_{k,l=1}^n \|X_k^{(p)} - X_l^{(p)}\|_2^\beta\Big)^2 - \Big(\frac{1}{n^2} \sum_{k,l=1}^n \|X_k - X_l\|_2^\beta\Big)^2,$$

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

We start by considering I_1. First assume that $\beta \le 1$. Observe that

$$\mathbb{E}[|I_1|] \le \mathbb{E}\big[\|(X_1^{(p)} - X_1) - (X_2^{(p)} - X_2)\|_2^\beta \big(\|X_1^{(p)} - X_2^{(p)}\|_2^\beta + \|X_1 - X_2\|_2^\beta\big)\big]$$

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

Similarly as in (7.4) the first expectation is bounded by $c \, \delta_n^{\beta \gamma_X}$, while the remaining two expectations are bounded, so that as in the proof of Proposition 7.1, we have that

$$\mathbb{E}[|I_1|] \leq c \, \delta_n^{\beta \gamma_X / 2} \, .$$

If $1 < \beta < 2$ we may proceed as for $\mathbb{E}[I_{11}]$ in the proof of Proposition 7.1 in the case $1 < \beta < 2$:

$$
\begin{aligned}
\mathbb{E}[|I_1|] & \leq & \mathbb{E}\big[\, \| X_1^{(p)} - X_2^{(p)} \|_2^{2\beta} - \| X_1 - X_2 \|_2^{2\beta} \, \big] \\
& \leq & c \, \mathbb{E}\big[\| X_1^{(p)} - X_2^{(p)} \|_2^{2\beta-1} \vee \| X_1 - X_2 \|_2^{2\beta-1} \, \big\| X_1^{(p)} - X_2^{(p)} \|_2 - \| X_1 - X_2 \|_2 \, \big] \\
& \leq & c \Big(\mathbb{E}\big[\| X_1^{(p)} - X_2^{(p)} \|_2^{2(2\beta-1)} \vee \| X_1 - X_2 \|_2^{2(2\beta-1)} \big] \Big)^{1/2} \\
& & \times \Big(\mathbb{E}\big[\| (X_1^{(p)} - X_1) - (X_2^{(p)} - X_2) \|_2^2 \big] \Big)^{1/2} \\
& = & c \, P_1 \, P_2 \, .
\end{aligned}
$$

We have $P_2 \leq c \, \delta_n^{\gamma_X / 2}$ and

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

We deal only with P_{12}; P_{11} can be bounded in a similar way. For $1 < 2\beta \leq 2$, the function $f(x) = |x|^{2\beta-1}$ is concave. Therefore

$$
\begin{aligned}
P_{12} & = & \mathbb{E}\Big[\Big(\int_0^1 (X_1(t) - X_2(t))^2 \, dt \Big)^{2\beta-1} \Big] \\
& \leq & \Big(\mathbb{E}\Big[\int_0^1 (X_1(t) - X_2(t))^2 \, dt \Big] \Big)^{2\beta-1} < \infty \, .
\end{aligned}
$$

In the last step we used (7.8).

If $2 < 2\beta < 4$ we have by Lyapunov's inequality and (7.9),

$$P_{12} = \mathbb{E}\Big[\Big(\int_0^1 (X_1(t) - X_2(t))^2 \, dt \Big)^{2\beta-1} \Big] \leq \mathbb{E}\Big[\int_0^1 |X_1(t) - X_2(t)|^{2(2\beta-1)} \, dt \Big] < \infty \, .$$

Thus we proved that

$$\mathbb{E}[|I_1|] \leq c \, \delta_n^{\gamma_X (\beta \wedge 1)/2} \, .$$

We can deal with I_2 in the same way by observing that

$$
\begin{aligned}
I_2 & = & \frac{1}{n^2} \sum_{k,l=1}^n \big(\| X_k^{(p)} - X_l^{(p)} \|_2^{\beta} - \| X_k - X_l \|_2^{\beta} \big) \frac{1}{n^2} \sum_{k,l=1}^n \big(\| X_k^{(p)} - X_l^{(p)} \|_2^{\beta} + \| X_k - X_l \|_2^{\beta} \big) \\
(7.14) \quad & = & \widetilde{P}_1 \widetilde{P}_2 \, .
\end{aligned}
$$

The expected value of $\widetilde{P}_2$ is bounded and hence $\widetilde{P}_2$ is stochastically bounded while similar calculations as for I_1 show that $\mathbb{E}[|\widetilde{P}_1|] \to 0$. Hence $I_2 \xrightarrow{\mathbb{P}} 0$. We have

$$
\begin{aligned}
I_3 & = & \frac{1}{n^3} \sum_{k,l,m=1}^n \big(\| X_k^{(p)} - X_l^{(p)} \|_2^{\beta} - \| X_k - X_l \|_2^{\beta} \big) \| X_k^{(p)} - X_m^{(p)} \|_2^{\beta} \\
& & + \frac{1}{n^3} \sum_{k,l,m=1}^n \| X_k - X_l \|_2^{\beta} \big(\| X_k^{(p)} - X_m^{(p)} \|_2^{\beta} - \| X_k - X_m \|_2^{\beta} \big) \\
(7.15) \quad & = & I_{31} + I_{32} \, .
\end{aligned}
$$

We will deal only with I_{32}; the other case is similar. Assume $0 < \beta \leq 1$. By the Cauchy-Schwarz inequality and using similar bounds as above,

$$
\begin{aligned}
\mathbb{E}[|I_{32}|] &\leq \left(\mathbb{E}\left[\|X_1 - X_2\|_2^{2\beta} \right] \right)^{1/2} \left(\mathbb{E}\left[\left| \|X_1^{(p)} - X_3^{(p)}\|_2^{\beta} - \|X_1 - X_3\|_2^{\beta} \right|^2 \right] \right)^{1/2} \\
&\leq \left(\mathbb{E}\left[\|X_1 - X_2\|_2^{2\beta} \right] \right)^{1/2} \left(\mathbb{E}\left[\|(X_1^{(p)} - X_1) - (X_3^{(p)} - X_3)\|_2^{2\beta} \right] \right)^{1/2} \to 0 \,.
\end{aligned}
$$
(7.16)

Now assume $1 < \beta < 2$. Then

$$
\begin{aligned}
\mathbb{E}[|I_{32}|] &\leq c\, \mathbb{E}\left[\|(X_1^{(p)} - X_1) - (X_3^{(p)} - X_3)\|_2 \left(\|X_1^{(p)} - X_3^{(p)}\|_2^{\beta-1} \vee \|X_1 - X_3\|_2^{\beta-1} \right) \|X_1 - X_2\|_2^{\beta} \right] \\
&\leq c\, \left(\mathbb{E}\left[\|(X_1^{(p)} - X_1) - (X_3^{(p)} - X_3)\|_2^2 \right] \right)^{1/2} \\
&\quad \times \left(\mathbb{E}\left[\left(\|X_1^{(p)} - X_3^{(p)}\|_2^{2(\beta-1)} \vee \|X_1 - X_3\|_2^{2(\beta-1)} \right) \|X_1 - X_2\|_2^{2\beta} \right] \right)^{1/2} \\
&\leq c\, \left(\mathbb{E}\left[\|(X_1^{(p)} - X_1) - (X_3^{(p)} - X_3)\|_2^2 \right] \right)^{1/2} \\
&\quad \times \left\{ \left(\mathbb{E}\left[\|X_1^{(p)} - X_3^{(p)}\|_2^{2(\beta-1)} \|X_1 - X_2\|_2^{2\beta} \right] \right)^{1/2} \right. \\
&\quad \left. + \left(\mathbb{E}\left[\|X_1 - X_3\|_2^{2(\beta-1)} \|X_1 - X_2\|_2^{2\beta} \right] \right)^{1/2} \right\} \,.
\end{aligned}
$$

The first factor is P_2 from above which is bounded by $c\delta_n^{\gamma_X/2}$. For the second term, we only consider $\mathbb{E}[\|X_1^{(p)} - X_3^{(p)}\|_2^{2(\beta-1)} \|X_1 - X_2\|_2^{2\beta}]$ by a symmetry argument. An application of Hölder's inequality to this quantity yields the bounds

$$
\left(\mathbb{E}\left[\|X_1^{(p)} - X_3^{(p)}\|_2^{2(2\beta-1)} \right] \right)^{\frac{\beta-1}{2\beta-1}} \left(\mathbb{E}\left[\|X_1 - X_2\|_2^{2(2\beta-1)} \right] \right)^{\frac{\beta}{2\beta-1}} = P_{11}^{\frac{\beta-1}{2\beta-1}} P_{12}^{\frac{\beta}{2\beta-1}} \,,
$$

where P_{11}, P_{12} are defined above and shown to be bounded. This concludes the proof under condition (1).

We assume condition (2). Now we prove the lemma under the condition that the moments of $X(t)$ of the order $2\beta \in (0, 2)$ are finite. We have for $2\beta \leq 1$ by concavity and in view of condition (7.11),

$$
\mathbb{E}[|I_1|] \leq \mathbb{E}\left[\|(X_1 - X_1^{(p)}) - (X_2 - X_2^{(p)})\|_2^{2\beta} \right] \leq c\,\delta_n^{\beta} \sum_{i=1}^p \mathbb{E}\left[\max_{t \in \Delta_i} |\Delta X(t, t_i)|^{2\beta} \right] \leq c\,p\,\delta_n^{\gamma_X'} \,.
$$
(7.17)

The right-hand side goes to zero by assumption. For $2\beta \in (1, 2)$ we have by Hölder's inequality,

$$
\begin{aligned}
\mathbb{E}[|I_1|] &\leq \mathbb{E}\left[\left| \|X_1^{(p)} - X_2^{(p)}\|_2^{2\beta} - \|X_1 - X_2\|_2^{2\beta} \right| \right] \\
&\leq c\, \mathbb{E}\left[\left(\|X_1^{(p)} - X_2^{(p)}\|_2^{2\beta-1} \vee \|X_1 - X_2\|_2^{2\beta-1} \right) \|(X_1^{(p)} - X_1) - (X_2^{(p)} - X_2)\|_2 \right] \\
&\leq c\, \left(\mathbb{E}\left[\|X_1^{(p)} - X_2^{(p)}\|_2^{2\beta} \vee \|X_1 - X_2\|_2^{2\beta} \right] \right)^{(2\beta-1)/(2\beta)} \\
&\quad \times \left(\mathbb{E}\left[\|(X_1^{(p)} - X_1) - (X_2^{(p)} - X_2)\|_2^{2\beta} \right] \right)^{1/(2\beta)} \\
&= c\, \widehat{P}_1 \widehat{P}_2 \,.
\end{aligned}
$$

The quantity $\widehat{P}_1$ is finite in view of (7.10) and $\widehat{P}_2 \to 0$ by the argument of (7.17).

For $I_2 = \widetilde{P}_1 \widetilde{P}_2$ we use (7.14). Since $\mathbb{E}[\|X_1 - X_2\|_2^{\beta}]$ and $\mathbb{E}[\|X_1^{(p)} - X_2^{(p)}\|_2^{\beta}]$ are finite the expectation of $\widetilde{P}_2$ is bounded while

$$
\mathbb{E}[|\widetilde{P}_1|] \leq 2\, \mathbb{E}[\|X - X^{(p)}\|_2^{\beta}] \leq 2 \left(\mathbb{E}[\|X - X^{(p)}\|_2^{2\beta}] \right)^{1/2} \,.
$$

The argument of (7.17) shows that the right-hand side converges to zero.

Finally, we use the decomposition $I_3 = I_{31} + I_{32}$. Inequality (7.16) and the bounds above show that $\mathbb{E}[|I_{32}|] \to 0$; the case $\mathbb{E}[|I_{31}|] \to 0$ follows in a similar way.

Collecting all bounds above, we proved $T_{n,\beta}(X^{(p)}, X^{(p)}) - T_{n,\beta}(X, X) \overset{\mathbb{P}}{\to} 0$ both under the conditions of (1) and (2). Then relation (7.12) is immediate. Indeed, under the assumption $\mathbb{E}[\|X\|_2^{2\beta}] < \infty$ the strong law of large numbers for U- and V-statistics yields $T_{n,\beta}(X, X) \overset{\text{a.s.}}{\to} T_\beta(X, X)$. $\qquad\square$

For the proof of the bootstrap consistency in Section 5 we need a.s. convergence of $T_{n,1}(X^{(p)}, Y^{(p)}) =: T_n(X^{(p)}, Y^{(p)})$. We give some sufficient conditions.

Lemma 7.4. *Assume the following conditions on the Riemann square-integrable process X on $[0, 1]$.*

1. $\mathbb{E}[\|X\|_2^2] < \infty$ *and* $\mathbb{E}[X(u)] = 0$ *for* $u \in [0, 1]$.
2. *(A.1) holds.*
3. $E[|X(t) - X(s)|^4] \leq c|t - s|^{\widetilde{\gamma}_X}$ *holds for some* $\widetilde{\gamma}_X > 0$.
4. $\sum_{n=1}^{\infty} n^{-1}\big(\delta_n^{\gamma_X} + \delta_n^{\widetilde{\gamma}_X}\big) < \infty$.

Then $T_n(X^{(p)}, X^{(p)}) \overset{\text{a.s.}}{\to} T(X, X)$ *holds as* $n \to \infty$.

Proof. From (7.1) recall the decomposition $T_n(X^{(p)}, X^{(p)}) - T_n(X, X) = I_1 + I_2 - 2I_3$; see also (7.13). Since $\mathbb{E}[\|X\|_2^2] < \infty$, by the strong law of large numbers for V-statistics, $T_n(X, X) \overset{\text{a.s.}}{\to} T(X, X)$. Therefore it suffices to show that

$$I_i \overset{\text{a.s.}}{\to} 0, \qquad i = 1, 3,$$

$$I_2' := \frac{1}{n^2}\sum_{k,l=1}^{n} \|X_k^{(p)} - X_l^{(p)}\|_2 - \frac{1}{n^2}\sum_{k,l=1}^{n} \|X_k - X_l\|_2 \overset{\text{a.s.}}{\to} 0.$$

We have

$$|I_2'| \leq \frac{1}{n}\sum_k^n \|X_k^{(p)} - X_k\|_2$$

$$= \frac{1}{n}\sum_k^n \big(\|X_k^{(p)} - X_k\|_2 - \mathbb{E}[\|X^{(p)} - X\|_2]\big) + \mathbb{E}[\|X^{(p)} - X\|_2].$$

By Jensen's inequality,

$$\mathbb{E}[\|X^{(p)} - X\|_2] \leq \Big(\int_0^1 \text{var}(X^{(p)}(u) - X(u))\,du\Big)^{1/2} \leq \delta_n^{\gamma_X/2} \to 0.$$

Moreover,

$$\text{var}\Big(\frac{1}{n}\sum_k^n \big(\|X_k^{(p)} - X_k\|_2\big)\Big) \leq n^{-1}\mathbb{E}[\|X^{(p)} - X\|_2^2] \leq n^{-1}\delta_n^{\gamma_X}.$$

Using Markov's inequality and the Borel-Cantelli lemma, we conclude that $I_2' \overset{\text{a.s.}}{\to} 0$ if $\sum_n n^{-1}\delta_n^{\gamma_X} < \infty$.

The proof of $I_1 \overset{\text{a.s.}}{\to} 0$ is similar. We have by the Cauchy-Schwarz inequality,

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

$$\times \Big(\frac{1}{n^2}\sum_{k,l=1}^{n} \big(\|X_k^{(p)} - X_l^{(p)}\|_2 + \|X_k - X_l\|_2\big)^2\Big)^{1/2}$$

$$\leq \quad c\frac{1}{n^2}\sum_{k,l=1}^{n}\big(\|X_k^{(p)}-X_k\|_2+\|X_l-X_l^{(p)}\|_2\big)^2$$

$$+c\Big(\frac{1}{n^2}\sum_{k,l=1}^{n}\big(\|X_k^{(p)}-X_k\|_2+\|X_l-X_l^{(p)}\|_2\big)^2\Big)^{1/2}\Big(\frac{1}{n^2}\sum_{k,l=1}^{n}\|X_k-X_l\|_2^2\Big)^{1/2}.$$

Therefore it remains to show that

$$\frac{1}{n}\sum_{k=1}^{n}\big(\|X_k^{(p)}-X_k\|_2^2-\mathbb{E}[\|X^{(p)}-X\|_2^2]\big)+\mathbb{E}[\|X^{(p)}-X\|_2^2]\overset{\text{a.s.}}{\to}0.$$

But we have $\mathbb{E}[\|X^{(p)}-X\|_2^2]=O(\delta_n^{\gamma_X})$ and

$$\operatorname{var}\Big(\frac{1}{n}\sum_{k=1}^{n}\big(\|X_k^{(p)}-X_k\|_2^2\big)\Big) \quad\leq\quad n^{-1}\mathbb{E}[\|X_k^{(p)}-X_k\|_2^4]$$

$$\leq\quad n^{-1}\int_0^1\mathbb{E}[(X^{(p)}(u)-X(u))^4]\,du\leq n^{-1}\delta_n^{\widetilde{\gamma}_X}.$$

Since we assume $\sum_n n^{-1}\delta_n^{\widetilde{\gamma}_X}<\infty$ applications of Markov's inequality and the Borel-Cantelli lemma show that $I_1\overset{\text{a.s.}}{\to}0$.

Finally, we show $I_3\overset{\text{a.s.}}{\to}0$. We have

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

$$+\frac{2}{n^3}\sum_{k,l,m=1}^{n}\big(\|X_k^{(p)}-X_l^{(p)}\|_2-\|X_k-X_l\|_2\big)\|X_k-X_m\|_2$$

$$=\quad I_{31}+I_{32}.$$

The Cauchy-Schwarz inequality yields

$$|I_{31}| \quad\leq\quad \frac{1}{n^2}\sum_{k,l=1}^{n}\big(\|X_k^{(p)}-X_l^{(p)}\|_2-\|X_k-X_l\|_2\big)^2$$

$$\leq\quad c\frac{1}{n}\sum_{k=1}^{n}\|X_k^{(p)}-X_k\|_2^2\overset{\text{a.s.}}{\to}0,$$

$$|I_{32}| \quad\leq\quad c\Big(\frac{1}{n^2}\sum_{k,l=1}^{n}\big(\|X_k^{(p)}-X_l^{(p)}\|_2-\|X_k-X_l\|_2\big)^2\Big)^{1/2}\Big(\frac{1}{n^2}\sum_{k,l=1}^{n}\|X_k-X_l\|_2^2\Big)^{1/2}$$

$$\leq\quad c\Big(\frac{1}{n}\sum_{k=1}^{n}\big(\|X_k^{(p)}-X_k\|_2^2\big)^{1/2}\Big(\frac{1}{n^2}\sum_{k,l=1}^{n}\|X_k-X_l\|_2^2\Big)^{1/2}\overset{\text{a.s.}}{\to}0.$$

This proves the lemma. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

APPENDIX A. THE SAMPLE DISTANCE COVARIANCE AS A DEGENERATE V-STATISTIC

We assume that $Z_i=(X_i,Y_i)$, $i=1,2,\ldots$, is an iid sequence with generic element (X,Y) whose components are Riemann square-integrable on $[0,1]$, and $\mathbb{E}[\|X\|_2^\beta+\|Y\|_2^\beta]<\infty$ for some $\beta\in(0,2)$. Lyons [9, 10] 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

$$
\begin{aligned}
f((x_1,y_1),(x_2,y_2),(x_3,y_3),(x_4,y_4)) &= f(z_1,z_2,z_3.z_4) \\
&= \|x_1-x_2\|_2^\beta\|y_1-y_2\|_2^\beta + \|x_1-x_2\|_2^\beta\|y_3-y_4\|_2^\beta \\
&\quad -2\|x_1-x_2\|_2^\beta\|y_1-y_3\|_2^\beta.
\end{aligned}
$$

From this representation, it is obvious that

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

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

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

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,

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

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

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

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*

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

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 $h(Z_{i_1},Z_{i_2},Z_{i_3},Z_{i_4})$, $1\le i_1\le i_2\le i_3\le i_4\le 4$, 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 $1\le i_1,\ldots,i_4\le 4$. For our specific kernel, this condition reads

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

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

References

[1] Arcones, M.A. and Giné, E. (1992) On the bootstrap of U and V statistics. *Ann. Statist.* **20**, 655–674.

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

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

[4] Davis, R.A., Matsui, M., Mikosch, T. and Wan, P. (2018) Applications of distance correlation to time series. *Bernoulli* **24**, 3087–3116.

[5] Dehling, H. and Mikosch, T. (1994) Random quadratic forms and the bootstrap for U-statistics. *J. Multivar. Anal.* **51**, 392–413.

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

[7] Hoffmann-Jørgensen, J. (1994) *Probability with a View Towards Statistics.* Chapman & Hall, New York.

[8] Matsui, M., Mikosch, T. and Samorodnitsky G. (2017) Distance covariance for stochastic processes. *Probab. Math. Statist.* **37**, 355–372.

[9] Lyons, R. (2013) Distance covariance in metric spaces. *Ann. Probab.* **41**, 3284–3305.

[10] Lyons, R. (2018) Errata to "Distance covariance in metric spaces". *Ann. Probab.* **46**, 2400–2405.

[11] Marquardt, T. (2006). Fractional Lévy processes with an application to long memory moving average processes. *Bernoulli* **12**, 1099–1126.

[12] Samorodnitsky, G. (2016) *Stochastic Processes and Long Range Dependence.* Springer, Berlin.

[13] Samorodnitsky, G. and Taqqu, M.S. (1994) *Stable Non-Gaussian Random Processes. Stochastic Models with Infinite Variance.* Chapman & Hall, London.

[14] Székely, G.J., Rizzo, M.L. and Bakirov, N.K. (2007) Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35**, 2769–2794.

[15] Székely, G.J. and Rizzo, M.L. (2009) Brownian distance covariance. *Ann. Appl. Stat.* **3**, 1236–1265.

[16] Székely, G.J. and Rizzo, M.L. (2013) The distance correlation t-test of independence in high dimension. *J. Multivar. Anal.* **117**, 193–213.

[17] Székely, G.J. and Rizzo, M.L. (2014) Partial distance correlation with methods for dissimilarities. *Ann. Statist.* **42**, 2382–2412.

[18] Serfling, R.J. (1980) *Approximation Theorems of Mathematical Statistics.* John Wiley & Sons, New York.

Department of Mathematics, Ruhr-Universität Bochum, 44780 Bochum, Germany
E-mail address: herold.dehling@rub.de

Department of Business Administration, Nanzan University, 18 Yamazato-cho, Showa-ku, Nagoya 466-8673, Japan
E-mail address: mmuneya@gmail.com

Department of Mathematics, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark
E-mail address: mikosch@math.ku.dk

School of Operations Research and Information Engineering, Cornell University, 220 Rhodes Hall, Ithaca, NY 14853, U.S.A.
E-mail address: gs18@cornell.edu

University of Melbourne, School of Mathematics and Statistics, Richard Berry Building, Parkville, 3010, Melbourne, Australia
E-mail address: laleh.tafakori@unimelb.edu.au