DISTANCE COVARIANCE FOR STOCHASTIC PROCESSES*

BY

MUNEYA MATSUI (NAGOYA), THOMAS MIKOSCH (COPENHAGEN),
AND GENNADY SAMORODNITSKY (ITHACA)

The authors of this paper would like to congratulate Tomasz Rolski on his 70th birthday. We would like to express our gratitude for his longstanding contributions to applied probability theory as an author, editor, and organizer. Tomasz kept applied probability going in Poland and beyond, even in difficult historical times. The applied probability community, including ourselves, has benefitted a lot from his enthusiastic, energetic and reliable work.

Sto lat niech żyje nam! Zdrowia, szczęścia, pomyślności!

Abstract. The distance covariance of two random vectors is a measure of their dependence. The empirical distance covariance and correlation can be used as statistical tools for testing whether two random vectors are independent. We propose an analog of the distance covariance for two stochastic processes defined on some interval. Their empirical analogs can be used to test the independence of two processes.

2010 AMS Mathematics Subject Classification: Primary: 62E20; Secondary: 62G20, 62M99, 60F05, 60F25.

Key words and phrases: Empirical characteristic function, distance covariance, stochastic process, test of independence.

1. DISTANCE COVARIANCE FOR PROCESSES ON $[0, 1]$

We consider a real-valued stochastic process $X = \big(X(t)\big)_{t \in [0,1]}$ with sample paths in a measurable space S such that X is measurable as a map from its probability space into S. We assume that the probability measure P_X generated by X

on S is uniquely determined by its finite-dimensional distributions. Examples include processes with continuous or càdlàg sample paths on $[0, 1]$. The probability measure P_X is then determined by the totality of the characteristic functions

$$\varphi_X(\mathbf{x}_k; \mathbf{s}_k) = \varphi_X^{(k)}(\mathbf{x}_k; \mathbf{s}_k) = \int_S e^{i\,(s_1\,f(x_1)+\ldots+s_k\,f(x_k))}\,P_X(df), \quad k \geqslant 1,$$

where $\mathbf{x}_k = (x_1, \ldots, x_k)' \in [0, 1]^k, \mathbf{s}_k = (s_1, \ldots, s_k)' \in \mathbb{R}^k$. In particular, for two such processes, X and Y, the measures P_X and P_Y coincide if and only if

$$\varphi_X(\mathbf{x}_k; \mathbf{s}_k) = \varphi_Y(\mathbf{x}_k; \mathbf{s}_k) \quad \text{for all } \mathbf{x}_k \in [0, 1]^k, \mathbf{s}_k \in \mathbb{R}^k, k \geqslant 1.$$

We now turn from the general question of identifying the distributions of X and Y to a more specific but related one: given two processes X, Y on $[0, 1]$ with values in S as above and defined on the same probability space, we intend to find some means to verify whether X and Y are independent. Motivated by the discussion above, we need to show that the joint law of (X, Y) on $S \times S$, denoted by $P_{X,Y}$, coincides with the product measure $P_X \otimes P_Y$. Assuming, once again, that a probability measure on $S \times S$ is determined by the finite-dimensional distributions (as is the case with the aforementioned examples), we need to show that the joint characteristic functions of (X, Y) factorize, i.e.,

$$(1.1) \quad \varphi_{X,Y}(\mathbf{x}_k; \mathbf{s}_k, \mathbf{t}_k) = \int_{S^2} \exp\Big(i \sum_{j=1}^{k} \big(s_j f(x_j) + t_j h(x_j)\big)\Big) P_{X,Y}(df, dh)$$

$$= \varphi_X(\mathbf{x}_k; \mathbf{s}_k)\, \varphi_Y(\mathbf{x}_k; \mathbf{t}_k) \quad \mathbf{x}_k \in [0, 1]^k, \ \mathbf{s}_k, \mathbf{t}_k \in \mathbb{R}^k, \ k \geqslant 1.$$

Clearly, this condition is hard to check, and therefore we try to get a more compact equivalent condition which can also be used for some statistical test of independence between X and Y.

For this reason, we consider a unit rate Poisson process $N = \big(N(t)\big)_{t \in [0,1]}$ with arrivals $0 < T_1 < T_2 < \ldots < T_{N(1)} \leqslant 1$, write $\mathbf{T}_N = (T_1, \ldots, T_{N(1)})'$, and correspondingly $\mathbf{s}_N, \mathbf{t}_N$ for any vectors in $\mathbb{R}^{N(1)}$. Then, for any positive probability density function g on $\mathbb{R}$, we define

$$(1.2) \quad d(P_{X,Y}, P_X \otimes P_Y)$$

$$= \mathbb{E}_N\Big[\int_{\mathbb{R}^{2\,N(1)}} |\varphi_{X,Y}(\mathbf{T}_N; \mathbf{s}_N, \mathbf{t}_N) - \varphi_X(\mathbf{T}_N; \mathbf{s}_N)\,\varphi_Y(\mathbf{T}_N; \mathbf{t}_N)|^2$$

$$\times \prod_{j=1}^{N(1)} g(s_j) g(t_j)\, d\mathbf{s}_N\, d\mathbf{t}_N\Big]$$

$$= \sum_{k=1}^{\infty} \mathbb{P}\big(N(1) = k\big) \int_{[0,1]^k} \Big[\int_{\mathbb{R}^{2k}} |\varphi_{X,Y}(\mathbf{x}_k; \mathbf{s}_k, \mathbf{t}_k) - \varphi_X(\mathbf{x}_k; \mathbf{s}_k)\varphi_Y(\mathbf{x}_k; \mathbf{t}_k)|^2$$

$$\times \prod_{j=1}^{k} g(s_j) g(t_j)\, d\mathbf{s}_k d\mathbf{t}_k \Big] d\mathbf{x}_k,$$

where in the last step we used the order statistics property of the homogeneous Poisson process. Here we interpret the summand corresponding to $k = 0$ as zero, and we also suppress the dependence on g in the notation. Now, the right-hand integrals vanish if and only if (1.1) is satisfied for Lebesgue a.e. $\mathbf{x}_k, \mathbf{s}_k, \mathbf{t}_k$, hence if and only if (1.1) holds for any $\mathbf{x}_k, \mathbf{s}_k, \mathbf{t}_k$. We summarize:

LEMMA 1.1. *Let us assume that g is a positive probability density on $\mathbb{R}$. Then $d(P_{X,Y}, P_X \otimes P_Y) = 0$ if and only if $P_{X,Y} = P_X \otimes P_Y$.*

REMARK 1.1. Lemma 1.1 can easily be extended in several directions.

1. The statement remains valid if the Poisson probabilities $\big(\mathbb{P}(N(1) = k)\big)_{k \geqslant 1}$ are replaced by any summable sequence of nonnegative numbers with infinitely many positive terms.

2. Obvious modifications of Lemma 1.1 are valid, e.g., for random fields X, Y on $[0, 1]^d$ (in this case we can sample the values of the random fields at the points of a Poisson random measure on $[0, 1]^d$ whose mean measure is the d-dimensional Lebesgue measure). Moreover, the values of X, Y may be multivariate.

3. The positive probability density $\prod_{j=1}^k g(s_j)g(t_j)$ on $\mathbb{R}^{2k}$ can be replaced by any positive measurable function provided the infinite series in (1.2) is finite. This idea will be exploited in Section 3 below.

4. Returning to our original problem about identifying the laws of X and Y, similar calculations show that the quantity

$$d(P_X, P_Y) = \sum_{k=1}^{\infty} \mathbb{P}\big(N(1) = k\big) \int\limits_{[0,1]^k} \big[\int\limits_{\mathbb{R}^k} |\varphi_X(\mathbf{x}_k; \mathbf{s}_k) - \varphi_Y(\mathbf{x}_k; \mathbf{s}_k)|^2$$
$$\times \prod_{j=1}^k g(s_j) \, d\mathbf{s}_k \big] d\mathbf{x}_k$$

vanishes if and only if $X \stackrel{d}{=} Y$, where $\stackrel{d}{=}$ means that all finite-dimensional distributions of X and Y coincide. The quantity $d(P_X, P_Y)$ can be taken as the basis for a goodness-of-fit test for the distributions of X and Y.

In what follows, we refer to the quantities $d(P_{X,Y}, P_X \otimes P_Y)$ as *distance covariance* between the stochastic processes X and Y. This name is motivated by work on *distance covariance* for random vectors $\mathbf{X} \in \mathbb{R}^p, \mathbf{Y} \in \mathbb{R}^q$ (possibly of different dimensions) defined by

$$T(\mathbf{X}, \mathbf{Y}) = \int\limits_{\mathbb{R}^{p+q}} |\varphi_{\mathbf{X}, \mathbf{Y}}(\mathbf{s}, \mathbf{t}) - \varphi_{\mathbf{X}}(\mathbf{s}) \, \varphi_{\mathbf{Y}}(\mathbf{t})|^2 \, \mu(d\mathbf{s}, d\mathbf{t}),$$

where μ is a (possibly infinite) measure on $\mathbb{R}^{p+q}$; see, e.g., [1], [2], [6], [8], [9]. The authors of the quoted papers coined the names *distance covariance* and *distance correlation* for the standardized version

$$R(\mathbf{X}, \mathbf{Y}) = T(\mathbf{X}, \mathbf{Y}) / \sqrt{T(\mathbf{X}, \mathbf{X})T(\mathbf{Y}, \mathbf{Y})};$$

they chose some special infinite measures μ which lead to an elegant form of $T(\mathbf{X}, \mathbf{Y})$ and $R(\mathbf{X}, \mathbf{Y})$; see Section 3 for more information on this approach. The goal in the above-cited literature was to find a statistical tool for testing independence between the vectors $\mathbf{X}$ and $\mathbf{Y}$ using the fact that $R(\mathbf{X}, \mathbf{Y}) = 0$ if and only if $\mathbf{X}, \mathbf{Y}$ are independent provided μ has a positive Lebesgue density on $\mathbb{R}^{p+q}$. The sample versions $T_n(\mathbf{X}, \mathbf{Y})$ and $R_n(\mathbf{X}, \mathbf{Y}) = T_n(\mathbf{X}, \mathbf{Y})/\sqrt{T_n(\mathbf{X}, \mathbf{X})T_n(\mathbf{Y}, \mathbf{Y})}$, constructed from an i.i.d. sample $(\mathbf{X}_i, \mathbf{Y}_i)$, $i = 1, \ldots, n$, of copies of $(\mathbf{X}, \mathbf{Y})$, are then used as test statistics for checking independence of $\mathbf{X}$ and $\mathbf{Y}$.

For stochastic processes X, Y on $[0, 1]$ one might be tempted to test their independence based on independent observations $\mathbf{X}_i = \big(X_i(x_1), \ldots, X_i(x_k)\big)'$, $\mathbf{Y}_i = \big(Y_i(x_1), \ldots, Y_i(x_k)\big)'$, $i = 1, \ldots, n$, of the processes X, Y at the locations $\mathbf{x}_k$ in $[0, 1]^k$. However, it is observed in [7] that the empirical distance correlation $R_n(\mathbf{X}, \mathbf{Y})$ has the tendency to be very close to one even for relatively small values k. Our approach avoids the high dimensionality of the vectors $\mathbf{X}_i$ and $\mathbf{Y}_i$ by randomizing their dimension k.

Our paper is organized as follows. In Section 2 we study some of the theoretical properties of the distance covariance between two stochastic processes X, Y on $[0, 1]$ where we assume that g is a positive probability density. We find a tractable representation of this distance covariance from which we derive the corresponding sample version. In Section 3 we choose the non-integrable weight function g from the paper [9]. Again, we find a suitable representation of this distance covariance, derive the corresponding sample version and show that it is a consistent estimator of its deterministic counterpart. In Section 4 we conduct a small simulation study based on the sample distance correlation introduced in Section 2. We compare the small sample behavior of the sample distance correlation with the corresponding sample distance correlation of [9] for independent and dependent Brownian and fractional Brownian sample paths.

2. PROPERTIES OF DISTANCE COVARIANCE

2.1. Distance correlation. In the context of stochastic processes X, Y one may be interested in standardizing the distance covariance

$$T(X, Y) = d(P_{X,Y}, P_X \otimes P_Y),$$

i.e., in the *distance correlation*

$$R(X, Y) = \frac{T(X, Y)}{\sqrt{T(X, X)\, T(Y, Y)}}.$$

However, it is not obvious that $R(X, Y)$ assumes only values between zero and one. This property is guaranteed by a Cauchy–Schwarz argument.

LEMMA 2.1. *Assume that $g(s) = g(-s)$. Then $0 \leqslant R(X, Y) \leqslant 1$.*

We have $R(X, X) = 1$, In general, the relation $R(X, Y) = 1$ does not imply $X = Y$ a.s. For example, if X is symmetric, then $R(X, -X) = 1$ as well.

P r o o f. Assume that (X', Y') is an independent copy of (X, Y). Applying the Cauchy–Schwarz inequality first to the k-dimensional integral with respect to the product of k copies of g, then to the expectation with respect to the law of (X, Y), next with respect to the Lebesgue measure on $[0, 1]^k$ and, finally, with respect to the law of N, and using the symmetry of the density g, we obtain

$$
T(X, Y)
$$

$$
= \sum_{k=1}^{\infty} \mathbb{P}\big(N(1) = k\big) \int_{[0,1]^k} d\mathbf{x}_k
$$

$$
\times \mathbb{E}\Big[\int_{\mathbb{R}^{2k}} \Big[\Big(\exp\big(i \sum_{j=1}^{k} s_j X_j\big) - \varphi_X(\mathbf{x}_k; \mathbf{s}_k) \Big) \Big(\exp\big(i \sum_{j=1}^{k} t_j Y_j\big) - \varphi_Y(\mathbf{x}_k; \mathbf{t}_k) \Big)
$$

$$
\times \Big(\exp\big(-i \sum_{j=1}^{k} s_j X_j'\big) - \varphi_X(\mathbf{x}_k; -\mathbf{s}_k) \Big) \Big(\exp\big(-i \sum_{j=1}^{k} t_j Y_j'\big) - \varphi_Y(\mathbf{x}_k; -\mathbf{t}_k) \Big) \Big]
$$

$$
\times \prod_{j=1}^{k} g(s_j) g(t_j) \, d\mathbf{s}_k d\mathbf{t}_k \Big]
$$

$$
\leqslant \sum_{k=1}^{\infty} \mathbb{P}\big(N(1) = k\big) \int_{[0,1]^k} d\mathbf{x}_k
$$

$$
\times \Big(\mathbb{E}\Big[\Big| \int_{\mathbb{R}^{k}} \Big(\exp\big(i \sum_{j=1}^{k} s_j X_j\big) - \varphi_X(\mathbf{x}_k; \mathbf{s}_k) \Big)
$$

$$
\times \Big(\exp\big(-i \sum_{j=1}^{k} s_j X_j'\big) - \varphi_X(\mathbf{x}_k; -\mathbf{s}_k) \Big) \prod_{j=1}^{k} g(s_j) \, d\mathbf{s}_k \Big|^2 \Big] \Big)^{1/2}
$$

$$
\times \Big(\mathbb{E}\Big[\Big| \int_{\mathbb{R}^{k}} \Big(\exp\big(i \sum_{j=1}^{k} t_j Y_j\big) - \varphi_Y(\mathbf{x}_k; \mathbf{t}_k) \Big)
$$

$$
\times \Big(\exp\big(-i \sum_{j=1}^{k} t_j Y_j'\big) - \varphi_Y(\mathbf{x}_k; -\mathbf{t}_k) \Big) \prod_{j=1}^{k} g(t_j) \, d\mathbf{t}_k \Big|^2 \Big] \Big)^{1/2}
$$

$$
= \sum_{k=1}^{\infty} \mathbb{P}\big(N(1) = k\big) \int_{[0,1]^k} d\mathbf{x}_k
$$

$$
\times \Big[\int_{\mathbb{R}^{2k}} |\varphi_{X,X}(\mathbf{x}_k; \mathbf{s}_k, \mathbf{t}_k) - \varphi_X(\mathbf{x}_k; \mathbf{s}_k) \varphi_X(\mathbf{x}_k; \mathbf{t}_k)|^2 \prod_{j=1}^{k} g(s_j) g(t_j) \, d\mathbf{s}_k d\mathbf{t}_k \Big]^{1/2}
$$

$$
\times \Big[\int_{\mathbb{R}^{2k}} |\varphi_{Y,Y}(\mathbf{x}_k; \mathbf{s}_k, \mathbf{t}_k) - \varphi_Y(\mathbf{x}_k; \mathbf{s}_k) \varphi_Y(\mathbf{x}_k; \mathbf{t}_k)|^2 \prod_{j=1}^{k} g(s_j) g(t_j) \, d\mathbf{s}_k d\mathbf{t}_k \Big]^{1/2}
$$

$$\leqslant \sum_{k=1}^{\infty} \mathbb{P}\big(N(1) = k\big)$$

$$\times \Big[\int\limits_{[0,1]^k} d\mathbf{x}_k \int\limits_{\mathbb{R}^{2k}} |\varphi_{X,X}(\mathbf{x}_k; \mathbf{s}_k, \mathbf{t}_k) - \varphi_X(\mathbf{x}_k; \mathbf{s}_k)\varphi_X(\mathbf{x}_k; \mathbf{t}_k)|^2$$

$$\times \prod_{j=1}^{k} g(s_j)g(t_j)\, d\mathbf{s}_k d\mathbf{t}_k \Big]^{1/2}$$

$$\times \Big[\int\limits_{[0,1]^k} d\mathbf{x}_k \int\limits_{\mathbb{R}^{2k}} |\varphi_{Y,Y}(\mathbf{x}_k; \mathbf{s}_k, \mathbf{t}_k) - \varphi_Y(\mathbf{x}_k; \mathbf{s}_k)\varphi_Y(\mathbf{x}_k; \mathbf{t}_k)|^2$$

$$\times \prod_{j=1}^{k} g(s_j)g(t_j)\, d\mathbf{s}_k d\mathbf{t}_k \Big]^{1/2}$$

$$\leqslant \sqrt{T(X,X)}\,\sqrt{T(Y,Y)}.$$

This proves that $0 \leqslant R(X, Y) \leqslant 1$. ∎

2.2. Representations. Our next goal is to find explicit expressions for the quantities $d(P_{X,Y}, P_X \otimes P_Y)$. We observe that

$$|\varphi_{X,Y}(\mathbf{x}_k; \mathbf{s}_k, \mathbf{t}_k) - \varphi_X(\mathbf{x}_k; \mathbf{s}_k)\,\varphi_Y(\mathbf{x}_k; \mathbf{t}_k)|^2$$
$$= |\varphi_{X,Y}(\mathbf{x}_k; \mathbf{s}_k, \mathbf{t}_k)|^2 + |\varphi_X(\mathbf{x}_k; \mathbf{s}_k)|^2 |\varphi_Y(\mathbf{x}_k; \mathbf{t}_k)|^2$$
$$- 2\operatorname{Re}\{\varphi_{X,Y}(\mathbf{x}_k; \mathbf{s}_k, \mathbf{t}_k)\varphi_X(\mathbf{x}_k; -\mathbf{s}_k)\varphi_Y(\mathbf{x}_k; -\mathbf{t}_k)\}.$$

This expression suggests to decompose (1.2) into three distinct parts, the first one being

$$\sum_{k=1}^{\infty} \frac{e^{-1}}{k!} \int\limits_{[0,1]^k} \Big[\int\limits_{\mathbb{R}^{2k}} |\varphi_{X,Y}(\mathbf{x}_k; \mathbf{s}_k, \mathbf{t}_k)|^2 \prod_{j=1}^{k} g(s_j)g(t_j)\, d\mathbf{s}_k\, d\mathbf{t}_k \Big] d\mathbf{x}_k$$

$$= \int\limits_{S^2} \sum_{k=1}^{\infty} \frac{e^{-1}}{k!} \Big\{ \int\limits_{[0,1]^k} \Big[\int\limits_{\mathbb{R}^{2k}} \exp\Big(i \sum_{r=1}^{k} \big(s_r\big(f(x_r)-f'(x_r)\big)+t_r\big(h(x_r)-h'(x_r)\big)\big)\Big)$$

$$\times \prod_{j=1}^{k} g(s_j)g(t_j)\, d\mathbf{s}_k d\mathbf{t}_k \Big] d\mathbf{x}_k \Big\} P_{X,Y}\big(d(f,h)\big) P_{X,Y}\big(d(f',h')\big)$$

$$= \int\limits_{S^2} \sum_{k=1}^{\infty} \frac{e^{-1}}{k!} \Big(\int\limits_{[0,1]} \Big[\int\limits_{\mathbb{R}} e^{is\,(f(x)-f'(x))}\, g(s)ds \int\limits_{\mathbb{R}} e^{it\,(h(x)-h'(x))}\, g(t)dt \Big] dx \Big)^k$$

$$\times P_{X,Y}\big(d(f,h)\big) P_{X,Y}\big(d(f',h')\big)$$

$$= e^{-1} \int\limits_{S^2} \Big[\exp\Big(\int\limits_{[0,1]} \Big[\int\limits_{\mathbb{R}^2} e^{is\,(f(x)-f'(x))+it\,(h(x)-h'(x))} g(s)\, g(t)\, ds\, dt \Big] dx \Big) - 1 \Big]$$

$$\times P_{X,Y}\big(d(f,h)\big) P_{X,Y}\big(d(f',h')\big).$$

Similar calculations yield

$$
\begin{aligned}
&d(P_{X,Y}, P_X \otimes P_Y) \\
&= e^{-1} \int_{S^2} \Big[\exp\Big(\int_{[0,1]} \int_{\mathbb{R}} e^{is\,(f(x)-f'(x))} g(s)\,ds \int_{\mathbb{R}} e^{is\,(h(x)-h'(x))} g(s)\,ds dx \Big) \Big] \\
&\quad \times \big[P_{X,Y}\big(d(f,h)\big)\, P_{X,Y}\big(d(f',h')\big) + P_X \otimes P_Y\big(d(f,h)\big) P_X \otimes P_Y\big(d(f',h')\big) \\
&\quad - P_{X,Y}\big(d(f,h)\big)\, P_X \otimes P_Y\big(d(f',h')\big) - P_{X,Y}\big(d(f',h')\big)\, P_X \otimes P_Y\big(d(f,h)\big) \big].
\end{aligned}
$$

We summarize our results:

LEMMA 2.2. *The distance covariance between the processes X and Y on $[0,1]$ with values in S can be written in the form*

$$
\begin{aligned}
&e^1 T(X,Y) \\
&= \mathbb{E}\Big[\exp\Big(\int_{[0,1]} \int_{\mathbb{R}} e^{is\,(X(x)-X'(x))} g(s)\,ds \int_{\mathbb{R}} e^{is(Y(x)-Y'(x))} g(ds) ds\, dx \Big) \Big] \\
&\quad + \mathbb{E}\Big[\exp\Big(\int_{[0,1]} \int_{\mathbb{R}} e^{is\,(X(x)-X'(x))} g(s) ds \int_{\mathbb{R}} e^{is(Y''(x)-Y'''(x))} g(s)\,ds dx \Big) \Big] \\
&\quad - 2\mathrm{Re}\,\mathbb{E}\Big[\exp\Big(\int_{[0,1]} \int_{\mathbb{R}} e^{is\,(X(x)-X'(x))} g(s) ds \int_{\mathbb{R}} e^{is(Y(x)-Y''(x))} g(s)\,ds dx \Big) \Big],
\end{aligned}
$$

where (X', Y') is an independent copy of (X, Y), and Y'', Y''' are independent copies of Y which are also independent of X, X', Y, Y'.

EXAMPLE 2.1. Let g be the density of a suitably scaled symmetric α-stable law on $\mathbb{R}$, $\alpha \in (0, 2]$. Then

$$
\int_{\mathbb{R}} e^{is\,(f(x)-f'(x))} g(s)\,ds = e^{-|f(x)-f'(x)|^\alpha},
$$

and so for a uniform random variable U on $(0, 1)$ which is independent of X, Y, X', Y', Y'', Y''',

$$
\begin{aligned}
(2.1) \quad d(P_{X,Y}, P_X \otimes P_Y) = e^{-1}\mathbb{E}\big[&\exp\big(\mathbb{E}_U e^{-|X(U)-X'(U)|^\alpha - |Y(U)-Y'(U)|^\alpha} \big) \\
&+ \exp\big(\mathbb{E}_U e^{-|X(U)-X'(U)|^\alpha - |Y''(U)-Y'''(U)|^\alpha} \big) \\
&- 2\exp\big(\mathbb{E}_U e^{-|X(U)-X'(U)|^\alpha - |Y(U)-Y''(U)|^\alpha} \big) \big],
\end{aligned}
$$

where $\mathbb{E}_U$ denotes expectation with respect to U.

2.3. Sample distance covariance. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be an i.i.d. sample with distribution $P_{X,Y}$ and let $P_{n,X,Y}$ be the corresponding empirical distribution with marginals $P_{n,X}$ and $P_{n,Y}$. Then we can define the *sample distance*

covariance $T_n(X, Y)$ given by

$$
\begin{aligned}
e^1 T_n(X, Y) &= e^1 d(P_{n,X,Y}, P_{n,X} \otimes P_{n,Y}) \\
&= \frac{1}{n^2} \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} \exp\Big(\int_{[0,1]} \int_{\mathbb{R}} e^{is\,(X_{j_1}(x) - X_{j_2}(x))} g(s)\, ds \\
&\qquad\qquad\qquad \times \int_{\mathbb{R}} e^{is\,(Y_{j_1}(x) - Y_{j_2}(x))} g(s)\, ds dx \Big) \\
&\quad + \frac{1}{n^4} \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} \sum_{j_3=1}^{n} \sum_{j_4=1}^{n} \exp\Big(\int_{[0,1]} \int_{\mathbb{R}} e^{is\,(X_{j_1}(x) - X_{j_2}(x))} g(s)\, ds \\
&\qquad\qquad\qquad \times \int_{\mathbb{R}} e^{is\,(Y_{j_3}(x) - Y_{j_4}(x))} g(s) ds dx \Big) \\
&\quad - 2\,\mathrm{Re}\, \frac{1}{n^3} \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} \sum_{j_3=1}^{n} \exp\Big(\int_{[0,1]} \int_{\mathbb{R}} e^{is\,(X_{j_1}(x) - X_{j_2}(x))} g(s)\, ds \\
&\qquad\qquad\qquad \times \int_{\mathbb{R}} e^{is\,(Y_{j_1}(x) - Y_{j_3}(x))} g(s)\, ds dx \Big).
\end{aligned}
$$

REMARK 2.1. This estimator is the exact sample analog of the distance covariance. However, this estimator is of V-statistics-type and leads to an additional bias. For practical purposes, one should avoid summation over diagonal and subdiagonal terms, making the estimator of U-statistics-type. Then, for example, the first expression would turn into

$$
\begin{aligned}
\frac{1}{n(n-1)} \sum_{j_1=1}^{n} \sum_{j_2=1, j_2 \neq j_1}^{n} &\exp\Big(\int_{[0,1]} \int_{\mathbb{R}} e^{is\,(X_{j_1}(x) - X_{j_2}(x))} g(s)\, ds \\
&\times \int_{\mathbb{R}} e^{is\,(Y_{j_1}(x) - Y_{j_2}(x))} g(s)\, ds dx \Big).
\end{aligned}
$$

Since the bias is asymptotically negligible and we are interested only in asymptotic results, we stick to the original version of the sample distance covariance. In Section 3 we consider a distinct version of distance covariance; see (3.3). By virtue of its construction, diagonal and subdiagonal terms vanish in its sample version, i.e., a bias problem does not appear.

EXAMPLE 2.2. Assume that g is the density of a suitably scaled symmetric α-stable random variable. Then

$$
\begin{aligned}
&e^1 d(P_{n,X,Y}, P_{n,X} \otimes P_{n,Y}) \\
&= \frac{1}{n^2} \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} \exp\Big(\int_{[0,1]} e^{-|X_{j_1}(x) - X_{j_2}(x)|^\alpha - |Y_{j_1}(x) - Y_{j_2}(x)|^\alpha}\, dx \Big) \\
&\quad + \frac{1}{n^4} \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} \sum_{j_3=1}^{n} \sum_{j_4=1}^{n} \exp\Big(\int_{[0,1]} e^{-|X_{j_1}(x) - X_{j_2}(x)|^\alpha - |Y_{j_3}(x) - Y_{j_4}(x)|^\alpha}\, dx \Big) \\
&\quad - \frac{2}{n^3} \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} \sum_{j_3=1}^{n} \exp\Big(\int_{[0,1]} e^{-|X_{j_1}(x) - X_{j_2}(x)|^\alpha - |Y_{j_1}(x) - Y_{j_3}(x)|^\alpha}\, dx \Big).
\end{aligned}
$$

REMARK 2.2. The form of the sample distance covariance indicates that one needs to involve numerical methods for its calculation. In addition, in general we cannot assume that the sample paths of (X_i, Y_i) are completely observed. Then we need to approximate the path-dependent integrals appearing in the exponents of the expressions above by appropriate sums on a grid. These problems are not studied further in this paper.

The following result is an immediate consequence of the strong law of large numbers for U-statistics (see [3]) and the observation that $d(P_{n,X,Y}, P_{n,X} \otimes P_{n,Y})$ is a linear combination of U-statistics.

PROPOSITION 2.1. *Assume that* $\big((X_i, Y_i)\big)_{i=1,\ldots,n}$ *is an i.i.d. sequence of* S^2-*valued random elements. Then*

$$d(P_{n,X,Y}, P_{n,X} \otimes P_{n,Y}) \overset{a.s.}{\to} d(P_{X,Y}, P_X \otimes P_Y), \quad n \to \infty.$$

3. DISTANCE COVARIANCE WITH INFINITE WEIGHT MEASURES

So far we assumed that g is a positive integrable density. In the aforementioned literature (see, e.g., [9]) positive weight functions g were used which are not integrable over $\mathbb{R}$. In what follows, we consider an approach with suitable positive non-integrable weight functions which lead to a distance covariance for stochastic processes. Due to positivity of this weight function, Lemma 1.1 remains valid.

To begin, note that if the function g is not necessarily integrable but is symmetric, then appealing to (1.2) and using the symmetry of both the cosine function and the function g we have

(3.1) $d(P_{X,Y}, P_X \otimes P_Y)$

$$= \sum_{k=1}^{\infty} \mathbb{P}\big(N(1) = k\big) \int_{[0,1]^k} \mathbb{E}\bigg[\int_{\mathbb{R}^{2k}} \Big(\cos\big(\mathbf{s}_k'(\mathbf{X}_k - \mathbf{X}_k')\big) \cos\big(\mathbf{t}_k'(\mathbf{Y}_k - \mathbf{Y}_k')\big)$$

$$+ \cos\big(\mathbf{s}_k'(\mathbf{X}_k - \mathbf{X}_k')\big) \cos\big(\mathbf{t}_k'(\mathbf{Y}_k'' - \mathbf{Y}_k''')\big)$$

$$- 2\cos\big(\mathbf{s}_k'(\mathbf{X}_k - \mathbf{X}_k')\big) \cos\big(\mathbf{t}_k'(\mathbf{Y}_k - \mathbf{Y}_k'')\big)\Big) \prod_{j=1}^{k} g(s_j)g(t_j) \, d\mathbf{s}_k d\mathbf{t}_k\bigg] d\mathbf{x}_k,$$

where

$$\mathbf{X}_k = \big(X(x_1), \ldots, X(x_k)\big)', \quad \mathbf{Y}_k = \big(Y(x_1), \ldots, Y(x_k)\big)',$$

and $(\mathbf{X}_k', \mathbf{Y}_k')$ is an independent copy of $(\mathbf{X}_k, \mathbf{Y}_k)$, while $\mathbf{Y}_k'', \mathbf{Y}_k'''$ are i.i.d. copies of $\mathbf{Y}_k$ independent of everything else. Since

(3.2) $\cos u \cos v = 1 - (1 - \cos u) - (1 - \cos v) + (1 - \cos u)(1 - \cos v),$

we have

$$d(P_{X,Y}, P_X \otimes P_Y) = \sum_{k=1}^{\infty} \mathbb{P}\big(N(1) = k\big) \int_{[0,1]^k} \mathbb{E}\Big\{ \int_{\mathbb{R}^{2k}} \Big[\Big(1 - \cos\big(\mathbf{s}_k'(\mathbf{X}_k - \mathbf{X}_k')\big)\Big)$$

$$\times \Big(1 - \cos\big(\mathbf{t}_k'(\mathbf{Y}_k - \mathbf{Y}_k')\big)\Big)$$

$$+ \Big(1 - \cos\big(\mathbf{s}_k'(\mathbf{X}_k - \mathbf{X}_k')\big)\Big)\Big(1 - \cos\big(\mathbf{t}_k'(\mathbf{Y}_k'' - \mathbf{Y}_k''')\big)\Big)$$

$$- 2\Big(1 - \cos\big(\mathbf{s}_k'(\mathbf{X}_k - \mathbf{X}_k')\big)\Big)\Big(1 - \cos\big(\mathbf{t}_k'(\mathbf{Y}_k - \mathbf{Y}_k'')\big)\Big)\Big]$$

$$\times \prod_{j=1}^{k} g(s_j)g(t_j)\, d\mathbf{s}_k d\mathbf{t}_k \Big\} d\mathbf{x}_k.$$

Next we replace the product kernels $\prod_{j=1}^{k} g(s_j)$ above by other positive measurable functions on $\mathbb{R}^k$. Inspired by [9] we choose the functions

$$g_k(\mathbf{s}) = c_k\, |\mathbf{s}|^{-k-\alpha}, \quad \mathbf{s} \in \mathbb{R}^k,\ \alpha \in (0, 2),$$

where the constant $c_k = c_k(\alpha) > 0$ is such that

$$\int_{\mathbb{R}^k} \big(1 - \cos(\mathbf{s}'\mathbf{x})\big)\, g_k(\mathbf{s})\, d\mathbf{s} = |\mathbf{x}|^{\alpha}, \quad x \in \mathbb{R}^k.$$

The corresponding distance covariance between X and Y becomes

$$d(P_{X,Y}, P_X \otimes P_Y) = \sum_{k=1}^{\infty} \mathbb{P}\big(N(1) = k\big) \int_{[0,1]^k} \mathbb{E}\Big\{ \int_{\mathbb{R}^{2k}} \Big[\Big(1 - \cos\big(\mathbf{s}_k'(\mathbf{X}_k - \mathbf{X}_k')\big)\Big)$$

$$\times \Big(1 - \cos\big(\mathbf{t}_k'(\mathbf{Y}_k - \mathbf{Y}_k')\big)\Big)$$

$$+ \Big(1 - \cos\big(\mathbf{s}_k'(\mathbf{X}_k - \mathbf{X}_k')\big)\Big)\Big(1 - \cos\big(\mathbf{t}_k'(\mathbf{Y}_k'' - \mathbf{Y}_k''')\big)\Big)$$

$$- 2\Big(1 - \cos\big(\mathbf{s}_k'(\mathbf{X}_k - \mathbf{X}_k')\big)\Big)\Big(1 - \cos\big(\mathbf{t}_k'(\mathbf{Y}_k - \mathbf{Y}_k'')\big)\Big)\Big]$$

$$\times g_k(\mathbf{s}_k)g_k(\mathbf{t}_k)\, d\mathbf{s}_k d\mathbf{t}_k \Big\} d\mathbf{x}_k.$$

By Fubini's theorem and the order statistics property of the Poisson process,

$$(3.3)\quad d(P_{X,Y}, P_X \otimes P_Y)$$

$$= \sum_{k=1}^{\infty} \mathbb{P}\big(N(1) = k\big) \int_{[0,1]^k} \big(\mathbb{E}[|\mathbf{X}_k - \mathbf{X}_k'|^{\alpha}|\mathbf{Y}_k - \mathbf{Y}_k'|^{\alpha}]$$

$$+ \mathbb{E}[|\mathbf{X}_k - \mathbf{X}_k'|^{\alpha}]\,\mathbb{E}[|\mathbf{Y}_k'' - \mathbf{Y}_k'''|^{\alpha}] - 2\mathbb{E}[|\mathbf{X}_k - \mathbf{X}_k'|^{\alpha}|\mathbf{Y}_k - \mathbf{Y}_k''|^{\alpha}]\big) d\mathbf{x}_k$$

$$= \mathbb{E}[|\mathbf{X}_N - \mathbf{X}_N'|^{\alpha}|\mathbf{Y}_N - \mathbf{Y}_N'|^{\alpha}] + \mathbb{E}[|\mathbf{X}_N - \mathbf{X}_N'|^{\alpha}\,|\mathbf{Y}_N'' - \mathbf{Y}_N'''|^{\alpha}]$$

$$- 2\,\mathbb{E}[|\mathbf{X}_N - \mathbf{X}_N'|^{\alpha}|\mathbf{Y}_N - \mathbf{Y}_N''|^{\alpha}]$$

$$:= I_1 + I_2 - 2\,I_3,$$

where $\mathbf{X}_N = \big(X(T_1), \ldots, X(T_{N(1)})\big)'$, $\mathbf{Y}_N = \big(Y(T_1), \ldots, Y(T_{N(1)})\big)'$, etc. In

particular, all the expectations are finite if

$$(3.4) \qquad \sup_{0 \leqslant x \leqslant 1} \mathbb{E}[|X(x)|^\alpha + |Y(x)|^\alpha + |X(x)Y(x)|^\alpha] < \infty.$$

An empirical version of I_1 is then given by

$$\hat{I}_1 = \frac{1}{l_n} \frac{1}{n^2} \sum_{1 \leqslant i,j \leqslant n} \sum_{k=1}^{l_n} |\mathbf{X}_{i,N_k} - \mathbf{X}_{j,N_k}|^\alpha |\mathbf{Y}_{i,N_k} - \mathbf{Y}_{j,N_k}|^\alpha,$$

where $((X_k, Y_k))$ are i.i.d. copies of (X, Y) independent of the i.i.d. copies (N_i) of the homogeneous Poisson process N. The empirical versions $\hat{I}_2, \hat{I}_3$ of I_2, I_3 are defined in an analogous way. The integer sequence (l_n) is such that $l_n \to \infty$.

By the strong law of large numbers for U-statistics, for fixed l, as $n \to \infty$,

$$\frac{1}{l} \sum_{k=1}^{l} A_{nk} = \frac{1}{l} \frac{1}{n^2} \sum_{1 \leqslant i,j \leqslant n} \sum_{k=1}^{l} |\mathbf{X}_{i,N_k} - \mathbf{X}_{j,N_k}|^\alpha |\mathbf{Y}_{i,N_k} - \mathbf{Y}_{j,N_k}|^\alpha$$

$$\overset{\text{a.s.}}{\to} \frac{1}{l} \sum_{k=1}^{l} \mathbb{E}[|\mathbf{X}_{N_k} - \mathbf{X}'_{N_k}|^\alpha |\mathbf{Y}_{N_k} - \mathbf{Y}'_{N_k}|^\alpha \mid N_k] := \frac{1}{l} \sum_{k=1}^{l} A_k.$$

Therefore, we can choose a sequence $\epsilon_n \downarrow 0$ such that

$$\mathbb{P}\left(\frac{1}{l} \Big| \sum_{k=1}^{l} (A_{nk} - A_k) \Big| > \epsilon_n \right) \to 0,$$

and then also choose an integer sequence (r_n) such that $r_n \to \infty$ and

$$r_n \, \mathbb{P}\left(\frac{1}{l} \Big| \sum_{k=1}^{l} (A_{nk} - A_k) \Big| > \epsilon_n \right) \to 0.$$

Let us note that the sequence (r_n) can be chosen to be monotone and such that $r_n - r_{n-1} \in \{0, 1\}$ for each n. Then

$$\mathbb{P}\left(\frac{1}{r_n l} \Big| \sum_{s=1}^{r_n} \sum_{k=(s-1)l+1}^{sl} (A_{nk} - A_k) \Big| > \epsilon_n \right)$$

$$\leqslant \mathbb{P}\left(\frac{1}{l} \sup_{s=1,\dots,r_n} \Big| \sum_{k=(s-1)l+1}^{sl} (A_{nk} - A_k) \Big| > \epsilon_n \right) \to 0.$$

This means that

$$\frac{1}{r_n l} \sum_{k=1}^{r_n l} (A_{nk} - A_k) \overset{\mathbb{P}}{\to} 0, \qquad n \to \infty.$$

However, by the strong law of large numbers, as $n \to \infty$,

$$\frac{1}{r_n l} \sum_{k=1}^{r_n l} A_k \overset{\text{a.s.}}{\to} \mathbb{E}[A_1] = \mathbb{E}[|\mathbf{X}_N - \mathbf{X}'_N|^\alpha |\mathbf{Y}_N - \mathbf{Y}'_N|^\alpha].$$

Hence, for every l there is an (r_n) such that

$$\frac{1}{r_n\,l}\sum_{k=1}^{r_n l} A_{nk} \xrightarrow{\mathbb{P}} \mathbb{E}[A_1], \quad n \to \infty.$$

We conclude that

$$\sup_{l\,r_{n-1}\leqslant v\leqslant l\,r_n}\left|\frac{1}{v}\sum_{k=1}^{v} A_{nk} - \frac{1}{l\,r_n}\sum_{k=1}^{lr_n} A_{nk}\right|$$

$$\leqslant \frac{r_n - r_{n-1}}{l\,r_{n-1}r_n}\sum_{k=1}^{l\,r_n} A_{nk} + \frac{1}{lr_n}\sum_{k=lr_{n-1}+1}^{lr_n} A_{nk}.$$

The right-hand side converges in probability to zero, hence we have the law of large numbers for $\hat{I}_1$. Similar arguments apply to $\hat{I}_2$, $\hat{I}_3$. We summarize:

PROPOSITION 3.1. *Let* $\alpha \in (0, 2)$ *and assume that* (3.4) *holds. Then for any integer sequence* (l_n) *with* $l_n \to \infty$,

$$d(P_{n,X,Y}, P_{n,X} \otimes P_{n,Y}) = \frac{1}{l_n}\frac{1}{n^2}\sum_{1\leqslant i,j\leqslant n}\sum_{k=1}^{l_n}|\mathbf{X}_{i,N_k}-\mathbf{X}_{j,N_k}|^{\alpha}|\mathbf{Y}_{i,N_k}-\mathbf{Y}_{j,N_k}|^{\alpha}$$

$$+\frac{1}{l_n}\frac{1}{n^2}\sum_{1\leqslant i,j\leqslant n}\sum_{k=1}^{l_n}|\mathbf{X}_{i,N_k}-\mathbf{X}_{j,N_k}|^{\alpha}\frac{1}{l_n}\frac{1}{n^2}\sum_{1\leqslant i,j\leqslant n}\sum_{k=1}^{l_n}|\mathbf{Y}_{i,N_k}-\mathbf{Y}_{j,N_k}|^{\alpha}$$

$$-2\frac{1}{l_n}\frac{1}{n^3}\sum_{1\leqslant i,j,l\leqslant n}\sum_{k=1}^{l_n}|\mathbf{X}_{i,N_k}-\mathbf{X}_{j,N_k}|^{\alpha}|\mathbf{Y}_{i,N_k}-\mathbf{Y}_{l,N_k}|^{\alpha}$$

$$\xrightarrow{\mathbb{P}} d(P_{X,Y}, P_X \otimes P_Y).$$

4. A SIMULATION STUDY

In what follows, we conduct a small simulation study for the sample distance correlation $R_n(X, Y)$ from Section 2 for the standard normal density g. This choice implies that

$$e^1 T_n(X, Y) = e^1 d(P_{n,X,Y}, P_{n,X} \otimes P_{n,Y})$$

$$=\frac{1}{n^2}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n}\exp\Big(\int_{[0,1]} e^{-|X_{j_1}(x)-X_{j_2}(x)|^2/2-|Y_{j_1}(x)-Y_{j_2}(x)|^2/2}\,dx\Big)$$

$$+\frac{1}{n^4}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n}\sum_{j_3=1}^{n}\sum_{j_4=1}^{n}\exp\Big(\int_{[0,1]} e^{-|X_{j_1}(x)-X_{j_2}(x)|^2/2-|Y_{j_3}(x)-Y_{j_4}(x)|^2/2}\,dx\Big)$$

$$-\frac{2}{n^3}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n}\sum_{j_3=1}^{n}\exp\Big(\int_{[0,1]} e^{-|X_{j_1}(x)-X_{j_2}(x)|^2/2-|Y_{j_1}(x)-Y_{j_3}(x)|^2/2}\,dx\Big).$$

As a matter of fact, simulations of this quantity are highly complex. We choose a moderate sample size $n = 100$ and approximate the integrals on $[0, 1]$ by their Riemann sums at an equidistant grid with mesh $1/50$. For (X, Y), we take a bivariate Brownian motion (B_1, B_2) with correlation $\rho \in [0, 1]$, i.e.,

$$\operatorname{cov}\big(B_1(s), B_2(t)\big) = \rho \min(s, t), \quad s, t \in [0, 1],$$

and a bivariate fractional Brownian motion (W_1, W_2) with correlation $\rho \in [0, 1]$, i.e.,

$$\operatorname{cov}\big(W_1(s), W_2(t)\big) = \frac{\rho}{2}\{|s|^{2H} + |t|^{2H} - |t - s|^{2H}\}, \quad s, t \in [0, 1],$$

where we assume that the Hurst parameters of W_1 and W_2 are the same; see [4] for more general cross-correlation structures of vector-fractional Brownian motions.

We compare the behavior of the sample distance correlation

$$R_n(X, Y) = \frac{T_n(X, Y)}{\sqrt{T_n(X, X)}\sqrt{T_n(Y, Y)}}$$

of the aforementioned stochastic processes with the corresponding sample distance correlation from Székely and Rizzo [9],

$$R_n^{\mathrm{Sz}}(\mathbf{X}, \mathbf{Y}) = \frac{T_n^{\mathrm{Sz}}(\mathbf{X}, \mathbf{Y})}{\sqrt{T_n^{\mathrm{Sz}}(\mathbf{X}, \mathbf{X})}\sqrt{T_n^{\mathrm{Sz}}(\mathbf{Y}, \mathbf{Y})}},$$

where for a sample $(\mathbf{X}_i, \mathbf{Y}_i), i = 1, \ldots, n$, of independent copies of the vector $(\mathbf{X}, \mathbf{Y})$,

$$\begin{aligned}
T_n^{\mathrm{Sz}}(\mathbf{X}, \mathbf{Y}) = {} & \frac{1}{n^2} \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} |\mathbf{X}_{j_1} - \mathbf{X}_{j_2}||\mathbf{Y}_{j_1} - \mathbf{Y}_{j_2}| \\
& + \frac{1}{n^4} \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} \sum_{j_3=1}^{n} \sum_{j_4=1}^{n} |\mathbf{X}_{j_1} - \mathbf{X}_{j_2}||\mathbf{Y}_{j_3} - \mathbf{Y}_{j_4}| \\
& - 2\frac{1}{n^3} \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} \sum_{j_3=1}^{n} |\mathbf{X}_{j_1} - \mathbf{X}_{j_2}||\mathbf{Y}_{j_1} - \mathbf{Y}_{j_3}|.
\end{aligned}$$

We calculate the sample distance correlation $R_n^{\mathrm{Sz}}(\mathbf{X}, \mathbf{Y})$ based on $n = 100$ i.i.d. simulations of the vector $(\mathbf{X}, \mathbf{Y}) = \big(X(i/50), Y(i/50)\big)_{i=1,\ldots,50}$. The calculation of $R_n(X, Y)$ and $R_n^{\mathrm{Sz}}(\mathbf{X}, \mathbf{Y})$ is based on the same simulated sample paths $\big((X_i, Y_i)\big)_{i=1,\ldots,n}$.

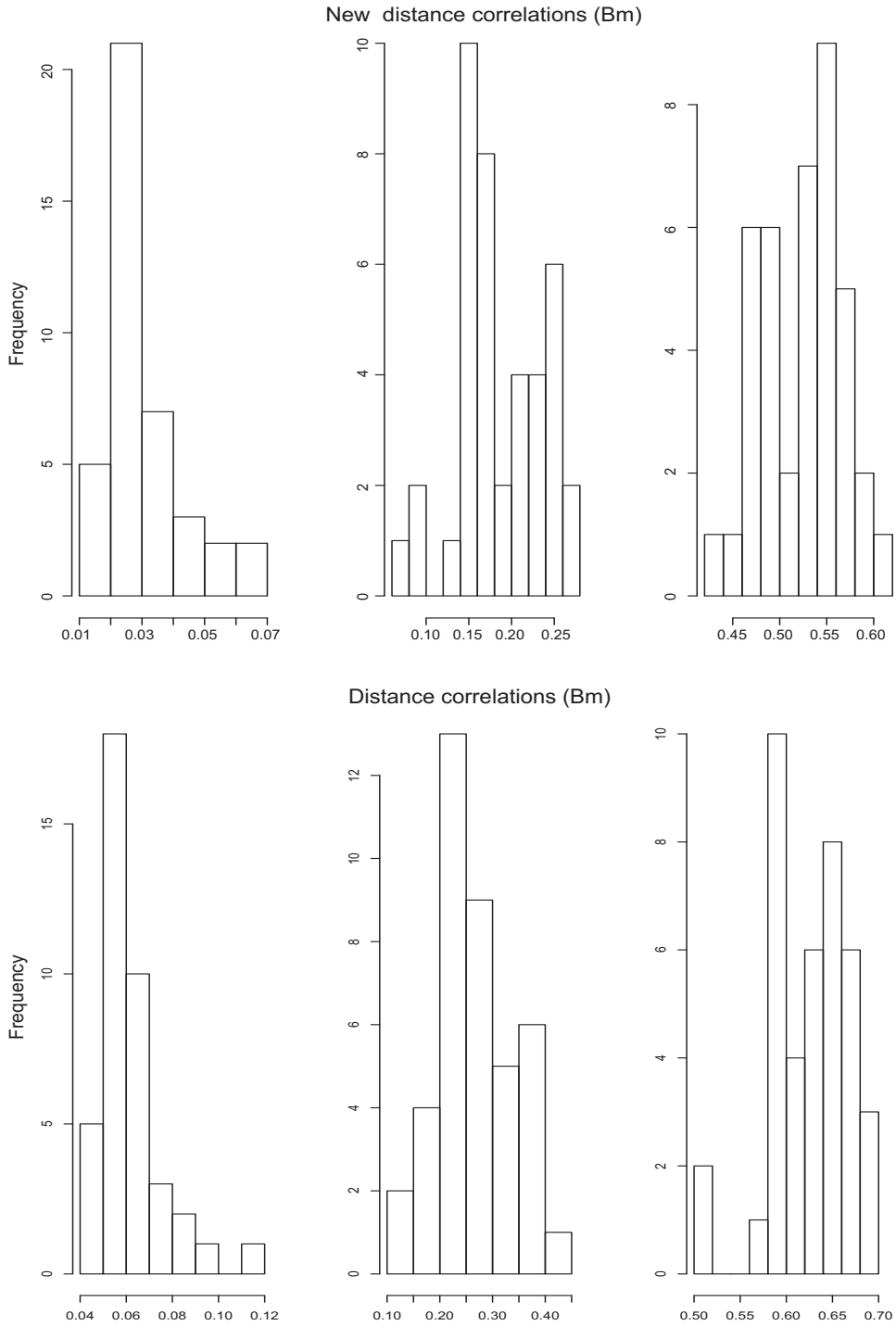New distance correlations (Bm)

Distance correlations (Bm)

FIGURE 1. Histograms of $R_n(B_1, B_2)$ (top) and $R_n^{\mathrm{Sz}}(B_1, B_2)$ (bottom) based on 40 samples. The correlations of B_1 and B_2 are respectively $\rho = 0$, 0.5, 0.8, from left to right.

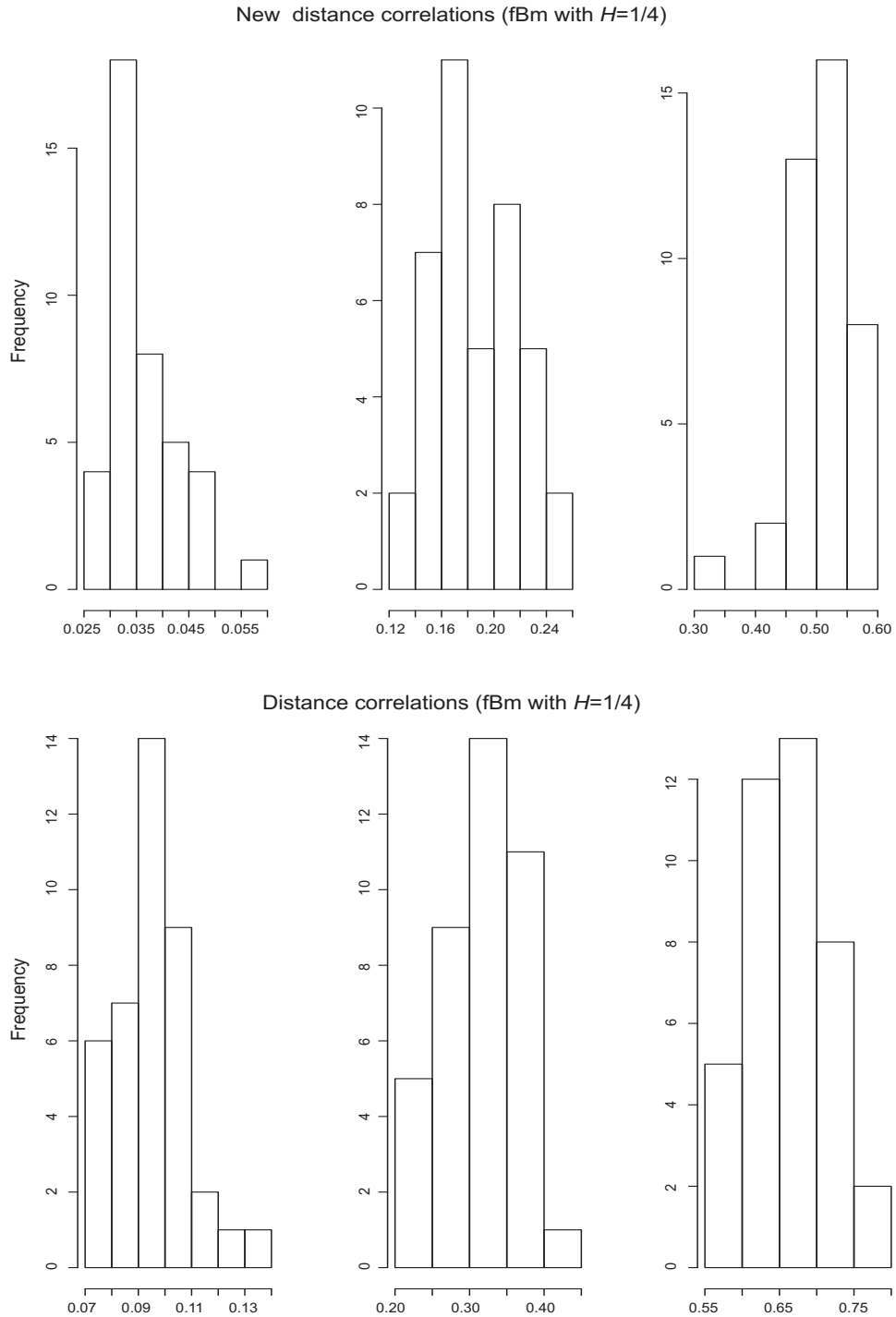New distance correlations (fBm with *H*=1/4)

Distance correlations (fBm with *H*=1/4)

FIGURE 2. Histograms of $R_n(W_1, W_2)$ (top) and $R_n^{\mathrm{Sz}}(W_1, W_2)$ (bottom) for $H = 0.25$ based on 40 samples. The correlations of W_1 and W_2 are respectively $\rho = 0$, 0.5, 0.8, from left to right.
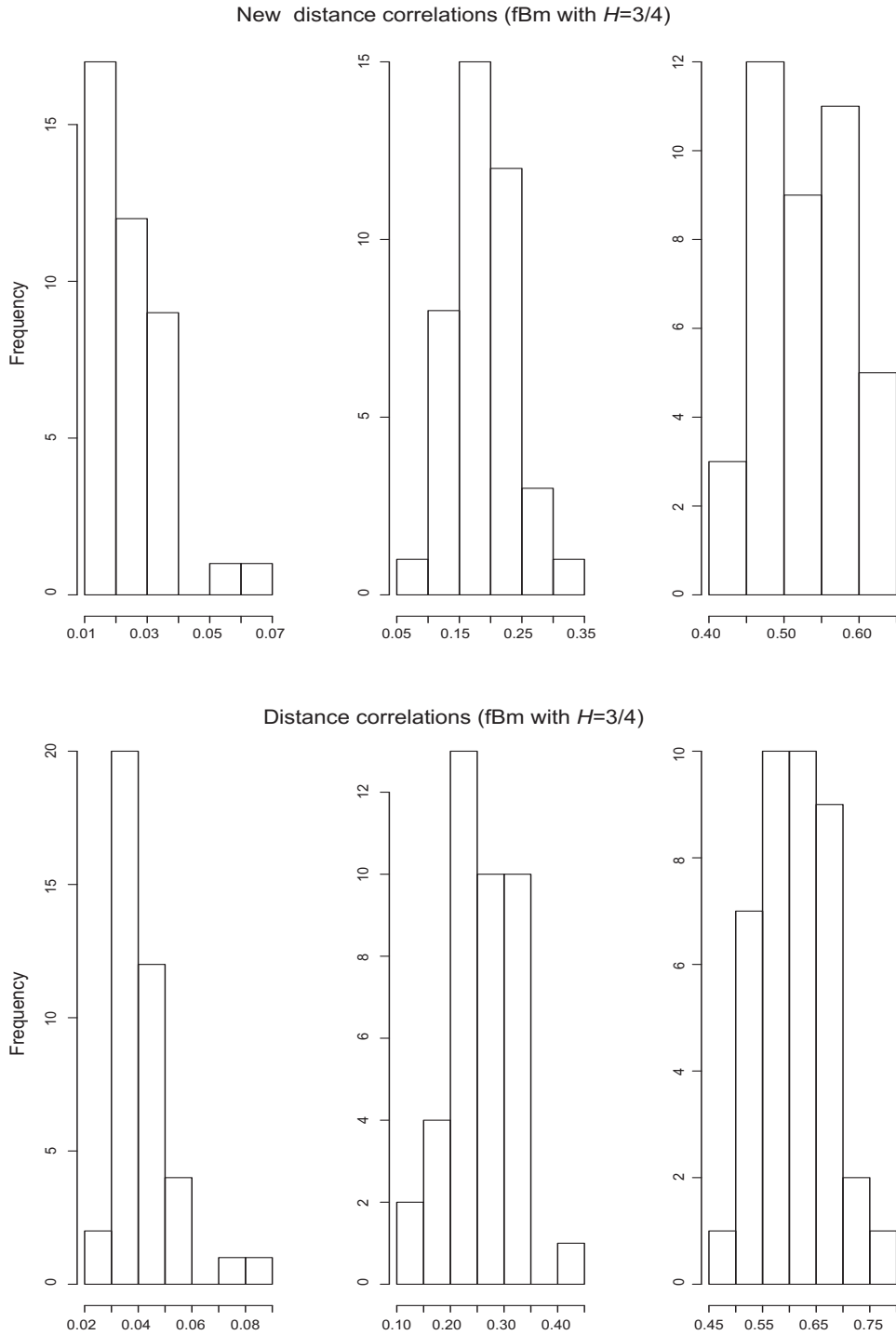
FIGURE 3. Histograms of $R_n(W_1, W_2)$ (top) and $R_n^{\mathrm{Sz}}(W_1, W_2)$ (bottom) for $H = 0.75$ based on 40 samples. The correlations of W_1 and W_2 are respectively $\rho = 0$, 0.5, 0.8, from left to right.
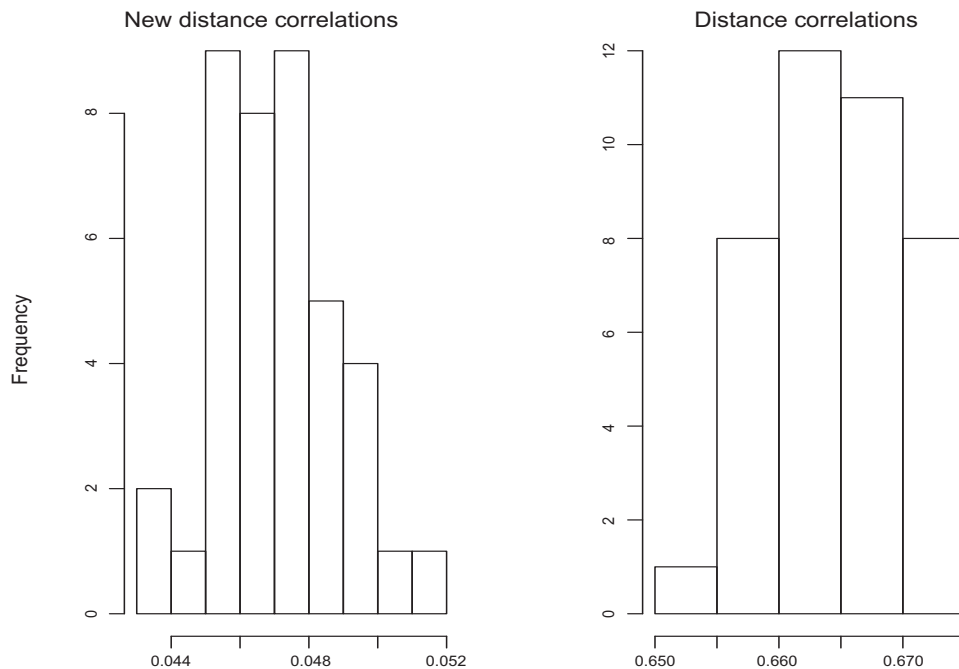
New distance correlations

Distance correlations

FIGURE 4. Histograms of $R_n(X, Y)$ (left) and $R_n^{\mathrm{Sz}}(X, Y)$ (right) based on 40 samples, where X and Y are independent piecewise constant processes based on i.i.d. normal random variables.

Figures 1–3 are based on forty independent simulations of $R_n(X, Y)$ and $R_n^{\mathrm{Sz}}(\mathbf{X}, \mathbf{Y})$. The three left (right) histograms show $R_n(X, Y)$ ($R_n^{\mathrm{Sz}}(\mathbf{X}, \mathbf{Y})$) for three different choices of processes (X, Y). Although it is difficult to judge from such a small simulation study with rather special stochastic processes, these graphs give one the impression that both sample distance correlations capture the independence or dependence of the processes X and Y quite well. The quantities $R_n^{\mathrm{Sz}}(\mathbf{X}, \mathbf{Y})$ have the tendency to be larger than $R_n(X, Y)$.

Finally, we consider two independent piecewise constant processes X and Y on $[0, 1]$, assuming i.i.d. standard normal values on the intervals $\big((i-1)/50, i/50\big]$, $i = 1, 2, \ldots, 50$. This is essentially the setting of Székely and Rizzo [7] who chose independent vectors of i.i.d. normal random variables for the construction of $R_n^{\mathrm{Sz}}(\mathbf{X}, \mathbf{Y})$. In the right histogram of Figure 4 one can see that $R_n^{\mathrm{Sz}}(\mathbf{X}, \mathbf{Y})$ is typically far from zero. This was observed in [7] where the case when the dimension of the vectors is large compared to the sample size was studied. On the other hand, our measure $R_n(X, Y)$ is quite in agreement with the independence hypothesis.

Of course, more investigations are needed to find out about the strengths and weaknesses of the distance covariances and correlation for processes introduced in this paper. One of the main problems will be to find reliable confidence bands for the estimator $R_n(X, Y)$. This is work in progress.

REFERENCES

[1] R. A. Davis, M. Matsui, T. Mikosch, and P. Wan, *Applications of distance correlation to time series*, Bernoulli (to appear).

[2] A. Feuerverger, *A consistent test for bivariate dependence*, Internat. Statist. Rev. 61 (1993), pp. 419–433.

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

[4] F. Lavancier, A. Philippe, and D. Surgailis, *Covariance function of vector self-similar processes*, Statist. Probab. Lett. 79 (2009), pp. 2415–2421.

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

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

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

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

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

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

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

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