STATISTICS IN TRANSITION new series, September 2020
Vol. 21, No. 3, pp. 21–37, DOI 10.21307/stattrans-2020-042
Received – 17.12.2019; accepted – 11.05.2020

Measuring and Testing Mutual Dependence of Multivariate Functional Data

Mirosław Krzyśko[1], **Łukasz Smaga**[2]

ABSTRACT

This paper considers new measures of mutual dependence between multiple multivariate random processes representing multidimensional functional data. In the case of two processes, the extension of functional distance correlation is used by selecting appropriate weight function in the weighted distance between characteristic functions of joint and marginal distributions. For multiple random processes, two measures are sums of squared measures for pairwise dependence. The dependence measures are zero if and only if the random processes are mutually independent. This property is used to construct permutation tests for mutual independence of random processes. The finite sample properties of these tests are investigated in simulation studies. The use of the tests and the results of simulation studies are illustrated with an example based on real data.

Key words: characteristic function, dependence measure, distance covariance, multivariate functional data, permutation method, test of independence..

1. Introduction

In recent years, statistical methods for analysing data expressed as functions or curves have received much attention. Such data are called functional data, which can be univariate and multivariate, and appear in many application domains as, for instance, chemometrics, economics, medicine, meteorology. For analysis of such data (i.e. the so called functional data analysis), there is currently a wide spectrum of models and methods as, for example, clustering and classification, functional principal component analysis, hypothesis testing, regression models. For an overview, we refer to the following monographs and recent review papers: Ramsay and Silverman (2005), Ferrary and Vieu (2006), Horváth and Kokoszka (2012), Zhang (2013), Kokoszka and Reimherr (2017) and Cuevas (2014), Wang et al. (2016) respectively.

This paper addresses the correlation analysis and testing independence for functional data in both univariate and multivariate cases. For functional time series, independence testing was considered by Horváth and Rice (2015). We would like to explore the association between two or more sets of functional variables. For two multivariate variables, the canonical correlation in the framework of canonical correlation analysis (CCA) was first proposed for this problem by Hotelling (1936). For functional data, this method was

[1]Interfaculty Institute of Mathematics and Statistics, The President Stanisław Wojciechowski State University of Applied Sciences in Kalisz, Poland. E-mail: mkrzysko@amu.edu.pl. ORCID: https://orcid.org/0000-0001-8075-4432
[2]Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Poznań, Poland.
E-mail: ls@amu.edu.pl. ORCID: https://orcid.org/0000-0002-2442-8816.

extended by Leurgans et al. (1993), He et al. (2004), Krzyśko and Waszak (2013) and Krzyśko and Smaga (2019). Unfortunately, the association viewed by canonical correlation is not a global measurement, since the intensity of the relationship is expressed component by component (see, Górecki et al., 2017, for more details). This was one of the reasons for constructing other association measures. The two very popular of them are the ρV coefficient by Escoufier (1970, 1973) and the distance correlation $dCor$ coefficient proposed in Székely et al. (2007). Their functional extensions were investigated in Górecki et al. (2016, 2017). Moreover, Górecki et al. (2019) used the functional distance correlation coefficient, among others, to construct variable selection procedures for classification of functional data. Unfortunately, the ρV coefficient may not detect non-linear dependence between two sets of variables, and it is difficult to evaluate the magnitude of the relationship just by considering its value. In these directions, the distance correlation coefficient seems to perform better and, moreover, (under mild conditions) it is equal to zero if and only if the random vectors are independent, which is not true for the ρV coefficient in general. Recently, Chen et al. (2019) proposed other distance-based coefficients with similar properties to distance correlation coefficient, which can even result in more powerful test for independence of two random vectors. In this paper, we adapt their results to a functional data framework by defining the functional version of their coefficient using a basis function representation of functional observations. In contrast to Górecki et al. (2016), we allow non-orthogonal basis making our results more general. In particular, we redefine the functional distance correlation coefficient in more generality.

The above considerations concern the case of two sets of variables only. Sometimes, there is a need of measure association or test independence of more than two sets of features. In this direction, very good results were obtained by Jin and Matteson (2018) in the case of multivariate data. They proposed a few methods, but the best of them are two procedures based on sums of squared distance covariance coefficients. Thus, in this paper, we extend these methods for functional data using also the functional versions of coefficients by Chen et al. (2019).

The remainder of this paper is organized as follows. In Section 2, we propose permutation tests of independence and dependence measures of multiple random processes. The finite sample properties of the testing procedures are investigated in simulation studies in Section 3. In Section 4, the real data example is presented. Finally, Section 5 is the summary of our work.

2. Methodology

In this section, we first present the basis representation of functional data, which is a kind of dimension reduction method. Then, using this representation and characteristic function apparatus, we propose tests of independence and dependence measures of two random processes. Finally, we extend these results for more than two processes.

2.1. Basis representation of functional data

Let $\mathbf{X} = (X_1, \ldots, X_p)^\top$ be a random process belonging to the Hilbert space $L_2^p(I)$ of p-dimensional vectors of square integrable functions defined on the interval $I = [a,b]$, $a, b \in \mathbb{R}$. This space is endowed with the following inner product:

$$\langle \mathbf{f}, \mathbf{g} \rangle_p = \int_I \mathbf{f}^\top(t)\mathbf{g}(t)dt$$

for $\mathbf{f}, \mathbf{g} \in L_2^p(I)$. For $i = 1, \ldots, p$, let $\{\phi_{ij}\}_{j=1}^\infty$ be basis in $L_2^1(I)$. Then each element of $L_2^1(I)$ can be represented as an infinite linear combination of basis functions. Such representation is difficult to apply in practice. Moreover, only a number of the first coefficients in this representation is usually the largest and the most important (Ramsey and Silverman, 2005). Therefore, we assume that each component of the process $\mathbf{X}$ can be represented by a finite number of basis functions, i.e.

$$X_i(t) = \sum_{j=1}^{B_i} \alpha_{ij}\phi_{ij}(t), \tag{1}$$

for $t \in I$ and $i = 1, \ldots, p$. The linear combination of basis functions in the right hand side of equality (1) will be called the basis representation of the process X_i.

The choice of the basis is usually not very crucial. However, some suggestions for this subject can be found in the literature (see, for example, Horváth and Kokoszka, 2012). The value of B_i determine the degree of smoothness of the basis representation, i.e. small value cause more smoothness. This value can be chosen deterministically or taking into account the problem at hand or using the Bayesian Information Criterion (BIC). The coefficients α_{ij} are usually estimated by the least squares method. For details about the practical construction of the basis representation, see, for example, Krzyśko and Waszak (2013).

Finally, let us introduce the following matrix form of the basis representation of a random process $\mathbf{X}$. Let

$$\alpha = (\alpha_{11}, \ldots, \alpha_{1B_1}, \ldots, \alpha_{p1}, \ldots, \alpha_{pB_p})^\top$$

and

$$\Phi(t) = \operatorname{diag}\left(\phi_1^\top(t), \ldots, \phi_p^\top(t)\right)$$

is the block diagonal matrix of

$$\phi_i^\top(t) = (\phi_{i1}(t), \ldots, \phi_{iB_i}(t)),$$

for $i = 1, \ldots, p$. Then the representation (1) can be expressed as follows:

$$\mathbf{X}(t) = \Phi(t)\alpha$$

which can be seen as the basis representation of the process $\mathbf{X}$. This means that the process $\mathbf{X}$ belongs to the finite dimensional subspace, say $\mathscr{L}_2^p(I)$, of the space $L_2^p(I)$.

2.2. Two sets of functional data

Assume that $\mathbf{X}$ and $\mathbf{Y}$ are two random processes belonging to the Hilbert spaces $L_2^p(I_1)$ and $L_2^q(I_2)$ respectively, where $I_1 = [a_1, b_1]$ and $I_2 = [a_2, b_2]$, $a_1, b_1, a_2, b_2 \in \mathbb{R}$. We would like to test the following hypotheses:

$$H_0 : \mathbf{X}, \mathbf{Y} \text{ are independent vs. } H_1 : \mathbf{X}, \mathbf{Y} \text{ are dependent}$$

and in the case of rejecting the null hypothesis, to measure the correlation between the processes $\mathbf{X}$ and $\mathbf{Y}$. For this purpose, we use the concept of characteristic function. Namely, (roughly speaking) we want to use the fact that the null hypothesis H_0 is equivalent to the equality of the characteristic function of the joint distribution of $\mathbf{X}$ and $\mathbf{Y}$ with the product of the characteristic functions of the distributions of $\mathbf{X}$ and $\mathbf{Y}$.

Let us first recall the definition of the characteristic function of a random process (Bosq, 2000, p. 37) in our framework. The characteristic functions of the processes $\mathbf{X}$ and $\mathbf{Y}$ are as follows:

$$\varphi_{\mathbf{X}}(\mathbf{u}) = E(\exp(i\langle \mathbf{u}, \mathbf{X}\rangle_p)), \quad \varphi_{\mathbf{Y}}(\mathbf{v}) = E(\exp(i\langle \mathbf{v}, \mathbf{Y}\rangle_q))$$

for $\mathbf{u} \in L_2^p(I_1)$ and $\mathbf{v} \in L_2^q(I_2)$, where $i^2 = -1$. (Of course, we assume that for all $\mathbf{u} \in L_2^p(I_1)$ the integral $\langle \mathbf{u}, \mathbf{X}\rangle_p$ converges for almost all realizations of $\mathbf{X}$, and the same applies to $\mathbf{v} \in L_2^q(I_2)$ and $\mathbf{Y}$.) Then the joint characteristic function of the pair of processes $\mathbf{X}$ and $\mathbf{Y}$ is of the form

$$\varphi_{\mathbf{X},\mathbf{Y}}(\mathbf{u}, \mathbf{v}) = E(\exp(i\langle \mathbf{u}, \mathbf{X}\rangle_p + i\langle \mathbf{v}, \mathbf{Y}\rangle_q)).$$

The next step is to combine these definitions with the basis representation of the processes $\mathbf{X}$ and $\mathbf{Y}$ (see Section 2.1). Suppose that $\mathbf{X} \in \mathscr{L}_2^p(I_1)$ and $\mathbf{Y} \in \mathscr{L}_2^q(I_2)$ and

$$\mathbf{X}(t) = \Phi_1(t)\alpha, \quad \mathbf{Y}(s) = \Phi_2(s)\beta,$$

where $\Phi_1(t) = \mathrm{diag}(\phi_{11}^\top(t), \ldots, \phi_{1p}^\top(t))$, $\Phi_2(s) = \mathrm{diag}(\phi_{21}^\top(s), \ldots, \phi_{2q}^\top(s))$, $\alpha \in \mathbb{R}^{K_x}$ and $\beta \in \mathbb{R}^{K_y}$ are random vectors, $K_x = B_1^x + \cdots + B_p^x$ and $K_y = B_1^y + \cdots + B_q^y$. Moreover, we assume that the functions $\mathbf{u} \in \mathscr{L}_2^p(I_1)$ and $\mathbf{v} \in \mathscr{L}_2^q(I_2)$, and they are represented as follows:

$$\mathbf{u}(t) = \Phi_1(t)\gamma, \quad \mathbf{v}(s) = \Phi_2(s)\delta,$$

where $\gamma \in \mathbb{R}^{K_x}$ and $\delta \in \mathbb{R}^{K_y}$ are constant vectors. Then we have

$$\langle \mathbf{u}, \mathbf{X}\rangle_p = \int_{I_1} \mathbf{u}^\top(t)\mathbf{X}(t)dt = \gamma^\top \int_{I_1} \Phi_1^\top(t)\Phi_1(t)dt\alpha = \gamma^\top \mathbf{J}_{\Phi_1}\alpha,$$

where $\mathbf{J}_{\Phi_1} = \mathrm{diag}(\mathbf{J}_{\phi_{11}}, \ldots, \mathbf{J}_{\phi_{1p}})$ and $\mathbf{J}_{\phi_{1i}} = \int_{I_i} \phi_{1i}(t)\phi_{1i}^\top(t)dt$ is the $B_i^x \times B_i^x$ cross product matrix, $i = 1, \ldots, p$. Analogously, we obtain $\langle \mathbf{v}, \mathbf{Y}\rangle_q = \delta^\top \mathbf{J}_{\Phi_2}\beta$. Therefore, the characteristic functions of the random processes $\mathbf{X}$ and $\mathbf{Y}$ are the characteristic functions of the random vectors $\mathbf{J}_{\Phi_1}\alpha$ and $\mathbf{J}_{\Phi_2}\beta$, i.e.

$$\varphi_{\mathbf{X}}(\mathbf{u}) = E(\exp(i\gamma^\top \mathbf{J}_{\Phi_1}\alpha)) = \varphi_{\mathbf{J}_{\Phi_1}\alpha}(\gamma), \quad \varphi_{\mathbf{Y}}(\mathbf{v}) = E(\exp(i\delta^\top \mathbf{J}_{\Phi_2}\beta)) = \varphi_{\mathbf{J}_{\Phi_2}\beta}(\delta).$$

Furthermore, the joint characteristic function of random processes $\mathbf{X}$ and $\mathbf{Y}$ is the joint characteristic function of random vectors $\mathbf{J}_{\Phi_1}\alpha$ and $\mathbf{J}_{\Phi_2}\beta$, i.e.

$$\varphi_{\mathbf{X},\mathbf{Y}}(\mathbf{u},\mathbf{v}) = E(\exp(i\gamma^\top \mathbf{J}_{\Phi_1}\alpha + i\delta^\top \mathbf{J}_{\Phi_2}\beta)) = \varphi_{\mathbf{J}_{\Phi_1}\alpha,\mathbf{J}_{\Phi_2}\beta}(\gamma,\delta).$$

These relations imply that for our purpose, we can use the distance methods for random vectors, which are based on

$$D_w = \int_{\mathbb{R}^{K_x+K_y}} |\varphi_{\mathbf{J}_{\Phi_1}\alpha,\mathbf{J}_{\Phi_2}\beta}(\gamma,\delta) - \varphi_{\mathbf{J}_{\Phi_1}\alpha}(\gamma)\varphi_{\mathbf{J}_{\Phi_2}\beta}(\delta)|^2 w(\gamma,\delta)d\gamma d\delta,$$

where $|z|$ is the modulus of $z \in \mathbb{C}$, and w is a weight function, which is positive almost everywhere. Different choices of the function w may result in plenty different methods. In the following, we consider two of them, which seem to be meaningful.

The most famous method of this kind was proposed by Székely et al. (2007). Górecki et al. (2016) used their methodology and considered the following functional distance covariance of random processes $\mathbf{X}$ and $\mathbf{Y}$:

$$FdCov(\mathbf{X},\mathbf{Y}) = dCov(\mathbf{J}_{\Phi_1}\alpha,\mathbf{J}_{\Phi_2}\beta) = \mathcal{V}_{\mathbf{J}_{\Phi_1}\alpha,\mathbf{J}_{\Phi_2}\beta} = \sqrt{D_{w_0}},$$

where

$$w_0(\gamma,\delta) = \frac{1}{C_{K_x}C_{K_y}\|\gamma\|_{K_x}^{K_x+1}\|\delta\|_{K_y}^{K_y+1}},$$

and

$$C_l = \frac{\pi^{(l+1)/2}}{\Gamma((l+1)/2)}$$

and $\|\cdot\|_l$ is the standard Euclidean norm in $\mathbb{R}^l$. The functional distance correlation between random processes $\mathbf{X}$ and $\mathbf{Y}$ is defined as follows:

$$FdCor(\mathbf{X},\mathbf{Y}) = \frac{FdCov(\mathbf{X},\mathbf{Y})}{\sqrt{FdCov(\mathbf{X},\mathbf{X})FdCov(\mathbf{Y},\mathbf{Y})}},$$

when $FdCov(\mathbf{X},\mathbf{X})$ and $FdCov(\mathbf{Y},\mathbf{Y})$ are positive, otherwise $FdCor(\mathbf{X},\mathbf{Y}) = 0$. Note that Górecki et al. (2016) used orthonormal basis, which implies the matrices $\mathbf{J}_{\Phi_1}$ and $\mathbf{J}_{\Phi_2}$ are identity matrices. Thus the above definition is a bit more general. For distributions with finite first moments, $FdCor(\mathbf{X},\mathbf{Y}) \in [0,1]$ and $FdCor(\mathbf{X},\mathbf{Y}) = 0$ if and only if $\mathbf{X}$ and $\mathbf{Y}$ are independent. The distance covariance by Székely et al. (2007) is implemented in the R package energy (R Core Team, 2019; Rizzo and Székely, 2019), which can be also used to calculate the functional distance covariance.

Recently, Chen et al. (2019) proposed other choice of weight function, which resulted in a kind of generalization of distance covariance. Namely, their weight functions are products of density functions. Let us now describe the details. Similarly as Székely et al. (2007), assume that the weight function $w(\gamma,\delta) = w_{K_x}(\gamma)w_{K_y}(\delta)$, where w_{K_x} and w_{K_y} are functions defined in the corresponding dimensions. This considerably simplifies expressions without giving up much generality. Let f_{K_x} and f_{K_y} be densities and let φ_{K_x} and φ_{K_y} be characteristic functions of $K_x \times 1$ and $K_y \times 1$ random vectors respectively. Chen et al. (2019) proved that when the densities f_{K_x} and f_{K_y} are positive with probability 1, then taking $w_{K_x} = f_{K_x}$ and $w_{K_y} = f_{K_y}$, D_w is as follows:

$$
\begin{aligned}
D_f =& E\left(Re\{\varphi_{K_x}(\mathbf{J}_{\Phi_1}(\alpha-\alpha_1))\}Re\{\varphi_{K_y}(\mathbf{J}_{\Phi_2}(\beta-\beta_1))\}\right)\\
&+E\left(Re\{\varphi_{K_x}(\mathbf{J}_{\Phi_1}(\alpha-\alpha_1))\}\right)E\left(Re\{\varphi_{K_y}(\mathbf{J}_{\Phi_2}(\beta-\beta_1))\}\right)\\
&-2E\left(Re\{\varphi_{K_x}(\mathbf{J}_{\Phi_1}(\alpha-\alpha_1))\}Re\{\varphi_{K_y}(\mathbf{J}_{\Phi_2}(\beta-\beta_2))\}\right),
\end{aligned}
$$

where $Re(z)$ denotes the real part of $z \in \mathbb{C}$, $\alpha_1 \stackrel{d}{=} \alpha$, $\beta_m \stackrel{d}{=} \beta$ for $m = 1,2$ and $\stackrel{d}{=}$ stands for equality in distribution. Moreover, D_f is equal to zero if and only if $\mathbf{J}_{\Phi_1}\alpha$ and $\mathbf{J}_{\Phi_2}\beta$ are mutually independent, and it is strictly positive otherwise.

There are many possible choices of the densities f_{K_x} and f_{K_y}. To greatly simplify D_f, the densities of spherical stable distributions can be used. The characteristic function of a spherical stable distribution with exponent $\alpha \in (0,2]$ is $\varphi_\alpha(\mathbf{t}) = \exp(-\|\mathbf{t}\|^\alpha)$. For $\alpha = 1$ and $\alpha = 2$, we have the multivariate standard Cauchy and normal distributions respectively. Further details about spherical stable distributions can be found in Zolotarev (1981) and Nolan (2013). A recent application of spherical stable distributions in the change-point methods to multivariate time-series can be found in Hlávka et al. (2020). When f_{K_x} and f_{K_y} are the densities of spherical stable distributions with the same exponent α, D_f can be written as

$$
\begin{aligned}
D_\alpha =& E\left(\exp(-(\|\mathbf{J}_{\Phi_1}(\alpha-\alpha_1)\|^\alpha+\|\mathbf{J}_{\Phi_2}(\beta-\beta_1)\|^\alpha))\right)\\
&+E\left(\exp(-\|\mathbf{J}_{\Phi_1}(\alpha-\alpha_1)\|^\alpha)\right)E\left(\exp(-\|\mathbf{J}_{\Phi_2}(\beta-\beta_1)\|^\alpha)\right)\\
&-2E\left(\exp(-(\|\mathbf{J}_{\Phi_1}(\alpha-\alpha_1)\|^\alpha+\|\mathbf{J}_{\Phi_2}(\beta-\beta_2)\|^\alpha))\right).
\end{aligned}
$$

Thus, we can define the functional distance covariance and correlation with exponent α of random processes $\mathbf{X}$ and $\mathbf{Y}$ as

$$
FdCov_\alpha(\mathbf{X},\mathbf{Y}) = \sqrt{D_\alpha}, \quad FdCor_\alpha(\mathbf{X},\mathbf{Y}) = \frac{FdCov_\alpha(\mathbf{X},\mathbf{Y})}{\sqrt{FdCov_\alpha(\mathbf{X},\mathbf{X})FdCov_\alpha(\mathbf{Y},\mathbf{Y})}}
$$

respectively. Similarly to $FdCor(\mathbf{X},\mathbf{Y})$, $FdCor_\alpha(\mathbf{X},\mathbf{Y}) \in [0,1]$ and $FdCor_\alpha(\mathbf{X},\mathbf{Y}) = 0$ if and only if $\mathbf{X}$ and $\mathbf{Y}$ are independent.

In practice, $FdCor(\mathbf{X},\mathbf{Y})$ and $FdCor_\alpha(\mathbf{X},\mathbf{Y})$ have to be estimated. Assume that $\mathbf{X}_1,\dots,\mathbf{X}_n$ and $\mathbf{Y}_1,\dots,\mathbf{Y}_n$ are independent realizations of random processes $\mathbf{X}$ and $\mathbf{Y}$ respectively. Let $\mathbf{X}_i(t) = \Phi_1(t)\alpha_i$ and $\mathbf{Y}_i(s) = \Phi_2(s)\beta_i$, $i = 1,\dots,n$ be the basis representations of the observations. The estimator of $FdCor(\mathbf{X},\mathbf{Y})$, say $\widehat{FdCor}(\mathbf{X},\mathbf{Y})$, was derived in Górecki et al.

(2016), so we omit it to save space. The estimator of $FdCov^2_\alpha(\mathbf{X},\mathbf{Y})$ is as follows:

$$\begin{aligned}\widehat{FdCov}^2_\alpha(\mathbf{X},\mathbf{Y}) =& \frac{1}{n^2}\sum_{1\le j,k\le n}\exp\left(-(\|\mathbf{J}_{\Phi_1}(\alpha_j-\alpha_k)\|^\alpha+\|\mathbf{J}_{\Phi_2}(\beta_j-\beta_k)\|^\alpha)\right)\\ &+\frac{1}{n^4}\sum_{1\le j,k\le n}\exp\left(-\|\mathbf{J}_{\Phi_1}(\alpha_j-\alpha_k)\|^\alpha\right)\sum_{1\le j,k\le n}\exp\left(-\|\mathbf{J}_{\Phi_2}(\beta_j-\beta_k)\|^\alpha\right)\\ &-\frac{2}{n^3}\sum_{1\le j,k,l\le n}\exp\left(-(\|\mathbf{J}_{\Phi_1}(\alpha_j-\alpha_k)\|^\alpha+\|\mathbf{J}_{\Phi_2}(\beta_j-\beta_l)\|^\alpha)\right).\end{aligned}$$

To sum up, both $FdCor(\mathbf{X},\mathbf{Y})$ and $FdCor_\alpha(\mathbf{X},\mathbf{Y})$ can be used as measures of dependence of random processes $\mathbf{X}$ and $\mathbf{Y}$. Moreover, since they both are equal to zero if and only if the processes $\mathbf{X}$ and $\mathbf{Y}$ are independent, testing the null hypothesis H_0 is equivalent to testing $H_0^{dCor} : FdCor(\mathbf{X},\mathbf{Y}) = 0$ or $H_0^\alpha : FdCor_\alpha(\mathbf{X},\mathbf{Y}) = 0$. For testing these hypotheses, we propose permutation tests based on test statistics $\widehat{FdCor}(\mathbf{X},\mathbf{Y})$ and $\widehat{FdCor}_\alpha(\mathbf{X},\mathbf{Y})$, because the asymptotic null distributions of $n\widehat{FdCor}(\mathbf{X},\mathbf{Y})$ and $n\widehat{FdCor}_\alpha(\mathbf{X},\mathbf{Y})$ are complicated and not distribution free and the convergence rate may be slow (see Székely et al., 2007; Chen et al., 2019). In the permutation method, the test statistics are recalculated many times with the permutation samples $\mathbf{X}_1,\ldots,\mathbf{X}_n,\mathbf{Y}_{\pi(1)},\ldots,\mathbf{Y}_{\pi(n)}$, where a permutation π is uniformly chosen from the symmetric group $\mathscr{S}_n$, the set of all $n!$ permutations of $(1,\ldots,n)$.

In the next section, we show how the above results can be extended for measuring and testing mutual dependence of more than two random processes.

2.3. Multiple sets of functional data

Let $\mathbf{X}_1,\ldots,\mathbf{X}_d$ be d random processes belonging to $L_2^{p_1}(I_1),\ldots,L_2^{p_d}(I_d)$ respectively, where $I_l = [a_l,b_l]$, $a_l,b_l\in\mathbb{R}$, $l = 1,\ldots,d$. Of interest is to test the following hypotheses

$$H_0 : \mathbf{X}_1,\ldots,\mathbf{X}_d \text{ are independent vs. } H_1 : \mathbf{X}_1,\ldots,\mathbf{X}_d \text{ are dependent}$$

and in the case of rejecting the null hypothesis, to measure the correlation between the processes $\mathbf{X}_1,\ldots,\mathbf{X}_d$.

The methods based on characteristic functions of Section 2.2 can be extended for case $d > 2$. Namely, for random vectors, this was recently done by Jin and Matteson (2018), whose results could be directly applied to functional data in much the same way as presented in Section 2.2. However, such tests may not perform well as was already shown in Jin and Matteson (2018) for random vectors. Fortunately, they also proposed some alternatives to these methods, which have better finite sample properties. Therefore, we are limited only to these alternative methods, which are asymmetric and symmetric measures of mutual dependence to capture mutual dependence via aggregating pairwise dependence.

Assume that $\mathbf{X}_l \in \mathscr{L}_2^{p_l}(I_l)$ and we have the following basis representation of the processes $\mathbf{X}_l$:

$$\mathbf{X}_l(t_l) = \Phi_l(t_l)\alpha_l,$$

where $t_l \in I_l$ and $\alpha_l \in \mathbb{R}^{K_l}$ are random vectors, $l = 1, \ldots, d$. Let

$$\alpha_{c^+} = \left(\alpha_{c+1}^\top, \ldots, \alpha_d^\top\right)^\top, \quad c = 1, \ldots, d-1,$$
$$\alpha_{-c} = \left(\alpha_1^\top, \ldots, \alpha_{c-1}^\top, \alpha_{c+1}^\top, \ldots, \alpha_d^\top\right)^\top, \quad c = 1, \ldots, d.$$

Thus, α_{c^+} denotes the subset of processes on the right of α_c, while α_{-c} denotes the subset of processes except α_c. Let Cor be a dependence measure for two random vectors such that it is equal to zero if and only if the random vectors are independent. Then the asymmetric and symmetric measures of mutual dependence of random vectors $\alpha_1, \ldots, \alpha_d$ are defined by

$$R(\alpha_1, \ldots, \alpha_d) = \frac{1}{d-1} \sum_{c=1}^{d-1} Cor^2(\alpha_c, \alpha_{c^+}), \quad S(\alpha_1, \ldots, \alpha_d) = \frac{1}{d} \sum_{c=1}^{d} Cor^2(\alpha_c, \alpha_{-c}).$$

Under mild condition, Jin and Matteson (2018) showed that

$$R(\alpha_1, \ldots, \alpha_d) \in [0, \infty), \quad S(\alpha_1, \ldots, \alpha_d) \in [0, \infty)$$

and

$$R(\alpha_1, \ldots, \alpha_d) = 0, \quad S(\alpha_1, \ldots, \alpha_d) = 0$$

if and only if $\alpha_1, \ldots, \alpha_d$ are mutually independent.

In the framework of functional data, we can use $FdCor$ or $FdCor_\alpha$ as Cor above. Then for testing the null hypothesis H_0, we can verify

$$H_0^R : R(\alpha_1, \ldots, \alpha_d) = 0 \text{ or } H_0^S : S(\alpha_1, \ldots, \alpha_d) = 0.$$

For these purposes, we use permutation tests based on the following test statistics being estimators of R and S:

$$\hat{R} = \frac{1}{d-1} \sum_{c=1}^{d-1} \widehat{FdCor}^2(\mathbf{X}_c, \mathbf{X}_{c^+}), \quad \hat{S} = \frac{1}{d} \sum_{c=1}^{d} \widehat{FdCor}^2(\mathbf{X}_c, \mathbf{X}_{-c})$$

or

$$\hat{R}_\alpha = \frac{1}{d-1} \sum_{c=1}^{d-1} \widehat{FdCor}_\alpha^2(\mathbf{X}_c, \mathbf{X}_{c^+}), \quad \hat{S}_\alpha = \frac{1}{d} \sum_{c=1}^{d} \widehat{FdCor}_\alpha^2(\mathbf{X}_c, \mathbf{X}_{-c}).$$

The pooled permutation sample is constructed by separately permuting the samples corresponding to processes $\mathbf{X}_2, \ldots, \mathbf{X}_d$. More precisely, when

$$\mathbf{X}_{11}, \ldots, \mathbf{X}_{1n}, \mathbf{X}_{21}, \ldots, \mathbf{X}_{2n}, \ldots, \mathbf{X}_{d1}, \ldots, \mathbf{X}_{dn}$$

are the observations, the pooled permutation sample is as follows:

$$\mathbf{X}_{11}, \ldots, \mathbf{X}_{1n}, \mathbf{X}_{2\pi_1(1)}, \ldots, \mathbf{X}_{2\pi_1(n)}, \ldots, \mathbf{X}_{d\pi_{d-1}(1)}, \ldots, \mathbf{X}_{d\pi_{d-1}(n)},$$

where the permutations $\pi_1,\dots,\pi_{d-1}$ are uniformly chosen from the symmetric group $\mathscr{S}_n$.

Under appropriate conditions, we have that $\hat{R}$, $\hat{S}$, $\hat{R}_\alpha$ and $\hat{S}_\alpha$ belong to the interval $[0,1]$. Therefore, they can be used as measures of dependence of random processes $\mathbf{X}_1,\dots,\mathbf{X}_d$.

3. Simulation studies

In this section, the finite sample behaviour of the permutation tests $\hat{R}$, $\hat{S}$, $\hat{R}_\alpha$ and $\hat{S}_\alpha$ for $\alpha = 0.1, 0.5, 1, 1.5, 2$ is determined in simulation studies. We investigate both the control of the type I error and power of the tests.

3.1. Simulation experiments

We set the number of observations $n = 15$ and investigate $d = 3$ random processes $\mathbf{X}_1 = (X_{11}, X_{12})^\top$, $\mathbf{X}_2 = (X_{21}, X_{22})^\top$, $\mathbf{X}_3 = (X_{31}, X_{32})^\top$ with dimensions $p_1 = p_2 = p_3 = 2$. The functional observations corresponding to these processes are generated in the following three models:

Model 1. They are represented by their values in an equally spaced grid of 50 points $t_{1,1} = t_{2,1} = t_{3,1} = 0,\dots,t_{1,50} = t_{2,50} = t_{3,50} = 1$ in $I_1 = I_2 = I_3 = [0,1]$, which are generated in the following way:

$$\begin{bmatrix} \mathbf{X}_{1r}(t_{1,u}) \\ \mathbf{X}_{2r}(t_{2,u}) \\ \mathbf{X}_{3r}(t_{3,u}) \end{bmatrix} = \begin{bmatrix} \Phi_1(t_{1,u}) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Phi_2(t_{2,u}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Phi_3(t_{3,u}) \end{bmatrix} \begin{bmatrix} \alpha_{1,r} \\ \alpha_{2,r} \\ \alpha_{3,r} \end{bmatrix} + \varepsilon_{r,u},$$

where $r = 1,\dots,n$, $u = 1,\dots,50$, the matrices Φ_l are as in Section 2 and contain the Fourier basis functions only and $B_i^l = 5$, $i = 1,2$, $l = 1,2,3$, $(\alpha_{1,r}^\top, \alpha_{2,r}^\top, \alpha_{3,r}^\top)^\top$ are 30-dimensional random vectors, and $\varepsilon_{r,u}^\top = (\varepsilon_{r,u,1},\dots,\varepsilon_{r,u,6})$ are the measurement errors such that $\varepsilon_{r,u,v} \sim N(0, 0.025a_{r,v})$ and $a_{r,v}$ is the range of the v-th row of the following matrix:

$$\begin{bmatrix} \Phi_1(t_{1,1})\alpha_{1,r} & \dots & \Phi_1(t_{1,50})\alpha_{1,r} \\ \Phi_2(t_{2,1})\alpha_{2,r} & \dots & \Phi_2(t_{2,50})\alpha_{2,r} \\ \Phi_3(t_{3,1})\alpha_{3,r} & \dots & \Phi_3(t_{3,50})\alpha_{3,r} \end{bmatrix}.$$

The random vectors $(\alpha_{1,r}^\top, \alpha_{2,r}^\top, \alpha_{3,r}^\top)^\top$ are generated as $Z_r\Sigma_\rho^{1/2}$, where $\Sigma_\rho = (1-\rho)\mathbf{I}_{30} + \rho\mathbf{1}_{30}\mathbf{1}_{30}^\top$, $\rho = 0, 0.1$, $\mathbf{I}_a$ is the $a \times a$ identity matrix, $\mathbf{1}_a$ is the $a \times 1$ vector of ones, and Z_r are 30×1 random vectors with iid coordinates from the following distributions: the standard normal distribution N, the Student t-distribution t_3 with three degrees of freedom, the Fisher-Snedecor distribution $F_{1,5}$ with 1 and 5 degrees of freedom, the standard Cauchy distribution C, the log-normal distribution LN. When $\rho = 0$, the null hypothesis about independence is true and we study the type I error of tests, while for $\rho = 0.1$, the alternative holds and we investigate their power. Note that for Cauchy distribution C, the expected value does not exist, but this distribution was among others considered in similar simulations of Chen et al. (2019), so we also use it.

Model 2. First, for each $t \in \{0.04, 0.08, \dots, 1\}$, the observations for $X_{li}(t)$, $l = 1,2$, $i = 1,2$ are generated as independent random variables of normal distribution $N(0, 0.25)$ or other non-normal distributions considered in Model 1. Then, for the same t, $X_{3i}(t) = \rho X_{1i}(t) + \varepsilon_i(t)$ for $i = 1,2$, where $\varepsilon_i(t)$ are independent random variables of normal distribution $N(0, 0.1)$. We set $\rho = 0, 0.5$ and then the null, alternative hypothesis is true respectively.

Model 3. This model is similar to Model 2, but here we consider non-linear dependence instead of linear dependence. More precisely, we set $X_{3i}(t) = X_{1i}^{\rho}(t) + \varepsilon_i(t)$ and $\rho = 2,3$. For both values of ρ, the alternative hypothesis holds.

The test statistics are calculated using the Fourier basis with $B_i^l = 5$, $i = 1,2$, $l = 1,2,3$. We use the least squares method to estimate the coefficients of the basis representation of generated functional data. The empirical sizes and powers (resp. p-values) of the permutation tests were estimated in 500 simulation runs (resp. $1{,}000$ permutation samples). For simplicity, the significance level is set to 5%. The simulation experiments as well as real data example of Section 4 were conducted in the R program (R Core Team, 2019).

3.2. Simulation results

The empirical sizes and powers of the permutation tests obtained in Models 1-3 are presented in Tables 1-3 respectively. Let us now discuss these simulation results.

The empirical sizes of all tests obtained in Models 1-2 (Tables 1-2 with $\rho = 0$) are usually very close to the level of significance of 5%. However, we can observe that the testing procedures $\widehat{FdCor}_\alpha$ with larger α (i.e. $\alpha = 1.5, 2$) tend to highly over-reject the null hypothesis in the case of Cauchy distribution C in Model 1. It seems that this can be explained by non-existence of the first moment of this distribution. Thus, the permutation tests seem to control the type I error level, except possibly tests based on $\widehat{FdCor}_{1.5}$ and $\widehat{FdCor}_2$.

In Model 1, all three processes $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$ are equally correlated, which is a similar scenario to that considered in Jin and Matteson (2018) for random vectors. Then both methods R and S perform very similarly in terms of size control and power. On the other hand, in Model 2, the processes $\mathbf{X}_1$ and $\mathbf{X}_3$ are correlated (when $\rho > 0$), and they are uncorrelated to process $\mathbf{X}_2$. Such setting was not considered by Jin and Matteson (2018). In this case, the testing procedures $\hat{S}$ and $\hat{S}_\alpha$ are much more powerful than the tests $\hat{R}$ and $\hat{R}_\alpha$ respectively. This perhaps can be explained by that the S method considers more comparisons between processes $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$ than the R method. In the case of Model 3, the processes $\mathbf{X}_1$ and $\mathbf{X}_3$ are non-linearly dependent (quadratically [$\rho = 2$] or cubically [$\rho = 3$]), and they are uncorrelated to process $\mathbf{X}_2$. Here, the comparison between methods R and S is more complicated and depends on the distribution of the data as well as the test statistic used. For tests $\widehat{FdCor}$ and $\widehat{FdCor}_{0.1}$, the methods R and S have similar empirical powers in most cases. For the other estimators (i.e. $\widehat{FdCor}_\alpha$, $\alpha = 0.5, 1, 1.5, 2$), the method R is usually more powerful than the method S. However, there are some exceptions, for example, under normal distribution N and cubic dependence or under Student distribution t_3 and quadratic dependence, the reverse is true.

Table 1: Empirical sizes ($\rho = 0$) and powers ($\rho = 0.1$) (as percentages) of all tests obtained in Model 1.

Distr.	ρ	Method	$\widehat{FdCor}$	$\widehat{FdCor}_{0.1}$	$\widehat{FdCor}_{0.5}$	$\widehat{FdCor}_{1}$	$\widehat{FdCor}_{1.5}$	$\widehat{FdCor}_{2}$
N	0	R	4.2	5.0	5.0	5.8	5.0	7.0
		S	4.4	5.4	5.4	6.4	6.4	4.6
	0.1	R	33.8	29.8	25.2	9.0	6.2	5.8
		S	35.2	30.8	25.2	9.8	5.2	4.0
t_3	0	R	5.6	5.2	5.2	4.6	4.4	5.2
		S	6.0	4.8	4.6	5.2	4.4	9.4
	0.1	R	26.6	24.0	19.2	5.2	4.8	4.8
		S	24.8	22.2	18.0	8.0	5.8	7.6
$F_{1,5}$	0	R	5.0	6.2	6.2	4.2	5.0	5.8
		S	5.0	4.8	5.0	3.4	4.6	7.0
	0.1	R	26.8	35.4	29.2	12.2	11.0	14.2
		S	24.4	30.6	29.8	10.8	8.8	14.8
C	0	R	4.4	4.6	5.6	3.0	19.4	80.4
		S	5.0	4.2	6.2	4.8	24.8	83.6
	0.1	R	43.0	65.8	36.4	11.8	35.0	88.8
		S	37.6	57.6	26.6	9.6	40.8	90.0
LN	0	R	4.2	5.0	4.8	4.6	4.8	4.8
		S	4.0	4.0	4.0	7.2	6.8	4.8
	0.1	R	24.8	29.6	26.4	10.6	8.2	7.4
		S	21.2	25.2	25.0	10.8	6.8	6.4

We can observe that the empirical powers of the tests $\widehat{FdCor}_{\alpha}$ usually decrease with the increasing α. There are only few exceptions (e.g. Model 3 and normal distribution N), but in these cases, the power loss between the most and the least powerful tests is not so large as in the remaining ones. Thus, among the tests $\widehat{FdCor}_{\alpha}$, the test $\widehat{FdCor}_{0.1}$ (i.e. with small α) is the most powerful in most scenarios.

In Models 1-2 and in Model 3 with normal distribution N, the tests $\widehat{FdCor}_{\alpha}$ with small α (e.g. $\alpha = 0.1$) are usually comparable with tests $\widehat{FdCor}$ in terms of power. Nevertheless, in some cases (e.g. under Fisher-Snedecor distribution $F_{1,5}$, Cauchy distribution C and the log-normal distribution LN), the tests $\widehat{FdCor}_{0.1}$ may have greater power than the tests $\widehat{FdCor}$. In Model 3 and non-normal distributions, the testing procedures $\widehat{FdCor}_{0.1}$ are much more powerful than the tests $\widehat{FdCor}$.

To sum up, the permutation test $\hat{S}_{0.1}$ seems to perform best. It maintains the type I error level very well and has power, which is greater than or comparable to power of the other tests considered. This test is followed by testing procedure $\hat{S}$. The test $\hat{S}_{0.1}$ overcomes the test $\hat{S}$ especially in the case of non-linear dependence.

Table 2: Empirical sizes ($\rho = 0$) and powers ($\rho = 0.5$) (as percentages) of all tests obtained in Model 2.

Distr.	ρ	Method	$\widehat{FdCor}$	$\widehat{FdCor}_{0.1}$	$\widehat{FdCor}_{0.5}$	$\widehat{FdCor}_{1}$	$\widehat{FdCor}_{1.5}$	$\widehat{FdCor}_{2}$
N	0	R	4.4	4.6	5.4	4.8	4.8	5.0
		S	4.0	4.8	4.6	4.6	3.8	4.2
	0.5	R	50.4	45.2	44.2	48.4	49.2	50.0
		S	77.6	69.2	70.6	76.0	80.6	82.2
t_3	0	R	5.6	5.4	5.8	5.8	6.4	5.8
		S	5.4	4.2	4.4	4.0	3.8	3.6
	0.5	R	48.2	47.2	41.8	36.8	33.0	29.0
		S	87.0	82.0	84.6	82.0	70.0	50.4
$F_{1,5}$	0	R	5.4	5.6	6.0	5.4	6.2	5.6
		S	4.6	5.8	6.0	6.4	6.2	4.4
	0.5	R	41.4	45.4	36.6	24.8	17.0	11.0
		S	67.4	67.8	72.6	56.8	23.0	9.6
C	0	R	5.0	5.2	6.2	3.8	5.2	4.8
		S	4.0	4.8	6.6	6.2	6.8	6.0
	0.5	R	34.4	44.0	31.8	8.2	6.2	5.2
		S	46.4	61.6	54.8	15.8	9.6	14.4
LN	0	R	4.6	4.0	4.0	3.8	4.2	4.2
		S	4.0	3.8	3.8	4.0	4.4	4.6
	0.5	R	48.8	46.4	41.0	35.8	30.0	25.6
		S	81.8	75.6	77.8	74.4	54.6	26.4

4. Real data example

In this section, we illustrate the use of the dependence measures and tests of independence for functional data proposed in Section 2 and the simulation results of Section 3. For this purpose, we consider the famous Canadian weather data, which are available in the R package fda (Ramsay et al., 2018).

The Canadian weather data contain the daily temperature and precipitation records of 35 Canadian weather stations averaged over 1960 to 1994 for 365 days. The raw temperature and precipitation curves for 35 weather stations are presented in Figure 1. Thus, we have $n = 35$ observations of two random processes ($d = 2$) representing temperature and precipitation. These functional observations are discretized in 365 time points. For illustrative purposes, we would like to measure dependence and test independence of temperature and precipitation treated as functional data. From Figure 1, (rather positive) correlation between temperature and precipitation may be observed. More precisely, weather stations with large temperature are also characterized by higher precipitation (dashed lines). In contrast, weather stations with lower temperature record lower rainfall (solid lines). To theoretically confirm this relationship, we use the methods described in Section 2 in the following.

Table 3: Empirical powers (as percentages) of all tests obtained in Model 3.

Distr.	ρ	Method	$\widehat{FdCor}$	$\widehat{FdCor}_{0.1}$	$\widehat{FdCor}_{0.5}$	$\widehat{FdCor}_{1}$	$\widehat{FdCor}_{1.5}$	$\widehat{FdCor}_{2}$
N	2	R	6.8	7.0	7.6	6.4	6.2	6.4
		S	6.6	5.8	6.6	6.4	6.0	7.0
	3	R	11.4	9.6	10.2	10.2	11.8	13.2
		S	15.8	13.8	14.6	16.0	17.0	17.8
t_3	2	R	54.4	59.2	53.8	26.4	12.8	7.4
		S	52.0	56.0	53.6	30.2	15.8	8.8
	3	R	68.8	84.0	69.2	27.6	14.0	8.6
		S	69.0	82.8	64.0	6.8	3.6	3.4
$F_{1,5}$	2	R	91.0	98.8	98.4	46.6	16.8	12.6
		S	84.0	96.8	95.6	6.4	3.4	3.2
	3	R	75.6	97.2	68.6	19.0	11.2	15.0
		S	71.2	95.0	8.0	5.2	5.4	5.0
C	2	R	74.6	95.8	64.4	23.0	34.4	52.6
		S	58.8	89.6	11.6	5.8	10.0	24.2
	3	R	67.2	98.0	54.0	29.0	40.4	60.8
		S	61.4	94.0	5.6	4.8	8.8	27.2
LN	2	R	98.0	99.6	100.0	71.0	28.0	11.6
		S	96.6	99.4	99.6	24.6	6.6	5.4
	3	R	85.4	98.0	79.4	30.4	6.8	6.6
		S	82.6	95.6	23.6	6.8	5.6	6.4

We use the permutation tests $\widehat{FdCor}$ and $\widehat{FdCor}_\alpha$ with $\alpha = 0.1, 0.5, 1, 1.5, 2$ and $1,000$ permutation samples. For the basis representation of the weather data, we use the Fourier basis with different size (i.e. $B_1^l = 3, 5, \ldots, 15$ for $l = 1, 2$) and the least squares method to estimate coefficients. The Fourier basis is recommended for periodical data (see, for example, Horváth and Kokoszka, 2012), so it is sensible for temperature and precipitation data, since they have annual cycles.

The results of statistical analysis are depicted in Table 4. We observe quite big values of correlation coefficients, especially $\widehat{FdCor}_\alpha$'s. Moreover, these values seem to not depend on the basis size. The same is true for p-values of the tests $\widehat{FdCor}$ and $\widehat{FdCor}_\alpha$ with $\alpha = 0.1, 0.5$. However, this is not true for the remaining testing procedures. This follows from the fact that the tests $\widehat{FdCor}$ and $\widehat{FdCor}_\alpha$ with small α are more robust to increasing dimension than the tests $\widehat{FdCor}_\alpha$ with moderate and large α. This was observed for random vectors in simulation studies in Chen et al. (2019) and moves to the case of functional data. Moreover, the p-values of the tests $\widehat{FdCor}_\alpha$ usually increase with the increasing α. Finally, the tests $\widehat{FdCor}$ and $\widehat{FdCor}_\alpha$ with $\alpha = 0.1, 0.5$ reject the null hypothesis at level of significance of 5%, in contrast to the remaining tests. These confirm the simulation results of Section 3, since the tests $\widehat{FdCor}$ and $\widehat{FdCor}_\alpha$ with small α were observed there to be

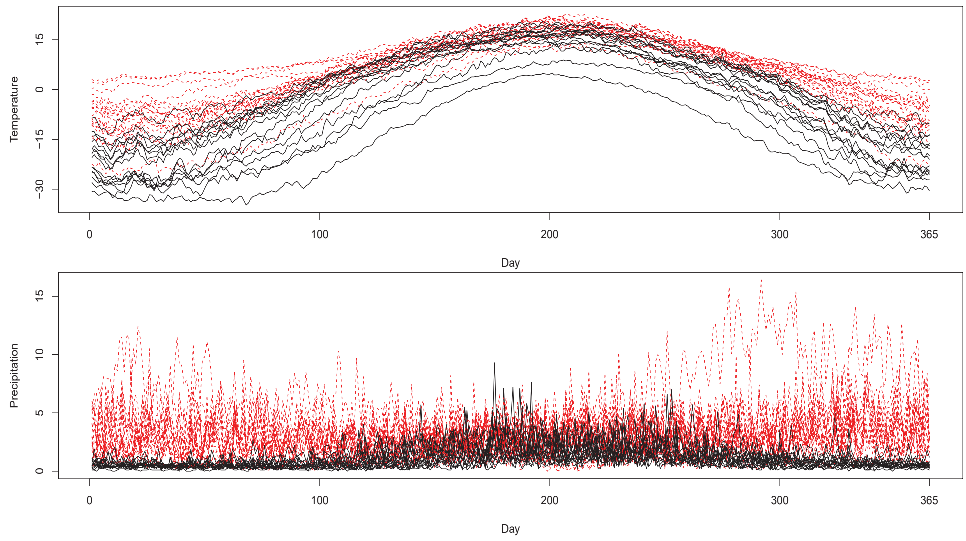

Figure 1: Temperature and precipitation for Canadian weather stations.

Table 4: Functional coefficients of correlation (FCor) and p-values of tests of independence for temperature and precipitation for Canadian weather stations.

	B_1^l	$\widehat{FdCor}$	$\widehat{FdCor}_{0.1}$	$\widehat{FdCor}_{0.5}$	$\widehat{FdCor}_1$	$\widehat{FdCor}_{1.5}$	$\widehat{FdCor}_2$
FCor	3	0.7379	0.9921	0.9825	0.9907	0.9914	0.9909
	5	0.7436	0.9935	0.9895	0.9975	0.9985	0.9988
	7	0.7449	0.9939	0.9913	0.9989	0.9998	0.9999
	9	0.7461	0.9942	0.9924	0.9993	0.9999	0.9999
	11	0.7464	0.9943	0.9932	0.9995	0.9999	0.9999
	13	0.7466	0.9944	0.9937	0.9996	0.9999	0.9999
	15	0.7468	0.9946	0.9942	0.9997	0.9999	0.9999
p-value	3	0.001	0.000	0.037	0.250	0.279	0.319
	5	0.001	0.000	0.015	0.339	0.352	0.306
	7	0.001	0.000	0.008	0.341	0.429	0.422
	9	0.001	0.000	0.002	0.228	0.407	0.411
	11	0.001	0.000	0.003	0.350	0.410	0.413
	13	0.001	0.000	0.006	0.360	0.353	0.325
	15	0.001	0.000	0.006	0.352	0.402	0.452

more powerful than the tests $\widehat{FdCor}_\alpha$ with moderate and large α. For these reasons, we should reject the null hypothesis about independence and conclude that there is a relationship between temperature and precipitation recorded in Canadian weather stations.

5. Conclusions

We have proposed new measures of mutual dependence for two or more sets of univariate and multivariate functional data. Our construction is based on the equivalence to mutual independence through characteristic functions and the basis function representation of the functional observations. Then, the problem is reduced to random vectors of basis expansion coefficients. We do not assume that the basis is orthogonal in contrast to the previous literature. For two sets of functional data, we follow the idea of functional distance correlation and construct functional versions of coefficients by Chen et al. (2019) indexed by hyperparameter $\alpha \in (0,2]$. In the case of more than two sets of functional data, we use the coefficients for pairs of sets and aggregate them by sums of their squares adapting the asymmetric and symmetric methods by Jin and Matteson (2018) to functional data framework. Simulation studies and real data example suggest that permutation tests based on new functional coefficients with small α and symmetric method perform best in terms of size control and power.

REFERENCES

BOSQ, D., (2000). *Linear Processes in Function Spaces. Theory and Applications*, Springer.

CHEN, F., MEINTANIS, S. G., ZHU, L., (2019). On some Characterizations and Multidimensional Criteria for Testing Homogeneity, Symmetry and Independence. *Journal of Multivariate Analysis*, 173, pp. 125–144.

CUEVAS, A., (2014). A Partial Overview of the Theory of Statistics with Functional Data. *Journal of Statistical Planning and Inference*, 147, pp. 1–23.

ESCOUFIER, Y., (1970). Echantillonnage dans une population de variables aléatoires réelles. Ph.D thesis, Université des Sciences et Techniques du Languedoc, Montpellier.

ESCOUFIER, Y., (1973). Le Traitement des Variables Vectorielles. *Biometrics*, 29, pp. 751–760.

FERRATY, F., VIEU, P., (2006). *Nonparametric Functional Data Analysis: Theory and Practice*, Springer: New York.

GÓRECKI, T., KRZYŚKO, M., RATAJCZAK, W., WOŁYŃSKI, W., (2016). An Extension of the Classical Distance Correlation Coefficient for Multivariate Functional Data with Applications. *Statistics in Transition new series*, 17, pp. 449–466.

GÓRECKI, T., KRZYŚKO, M., WOŁYŃSKI, W., (2017). Correlation Analysis for Multivariate Functional Data. *Studies in Classification, Data Analysis, and Knowledge Organization: Data Science*, pp. 243–258.

GÓRECKI, T., KRZYŚKO, M., WOŁYŃSKI, W., (2019). Variable Selection in Multivariate Functional Data Classification. *Statistics in Transition new series*, 20, pp. 123–138.

HE, G., MÜLLER, H. G., WANG, J. L., (2004). Methods of Canonical Analysis for Functional Data. *Journal of Statistical Planning and Inference*, 122, pp. 141–159.

HLÁVKA, Z., HUŠKOVÁ, M., MEINTANIS, S. G., (2020). Change-Point Methods for Multivariate Time-Series: Paired Vectorial Observations. *Statistical Papers*, DOI: https://doi.org/10.1007/s00362-020-01175-3.

HORVÁTH, L., KOKOSZKA, P., (2012). *Inference for Functional Data with Applications*, Springer.

HORVÁTH, L., RICE, G., (2015). Testing for Independence between Functional Time Series. *Journal of Econometrics*, 189, pp. 371–382.

HOTELLING, H., (1936). Relation Between Two Sets of Variables. *Biometrika*, 28, pp. 321–377.

JIN, Z., MATTESON, D. S., (2018). Generalizing Distance Covariance to Measure and Test Multivariate Mutual Dependence via Complete and Incomplete V-statistics. *Journal of Multivariate Analysis*, 168, pp. 304–322.

KOKOSZKA, P., REIMHERR, M., (2017). *Introduction to Functional Data Analysis*. Chapman and Hall/CRC.

KRZYŚKO, M., SMAGA, Ł., (2019). Robust Estimation in Canonical Correlation Analysis for Multivariate Functional Data. *Hacettepe Journal of Mathematics and Statistics*, 48, pp. 521–535.

KRZYŚKO, M., WASZAK, Ł., (2013). Canonical Correlation Analysis for Functional Data. *Biometrical Letters*, 50, pp. 95–105.

LEURGANS, S. E., MOYEED, R. A., SILVERMAN, B. W., (1993). Canonical Correlation Analysis when the Data are Curves. *Journal of the Royal Statistical Society. Series B (Methodological)*, 55, pp. 725–740.

NOLAN, J. P., (2013). Multivariate Elliptically Contoured Stable Distributions: Theory and Estimation. *Computational Statistics*, 28, pp. 2067–2089.

R CORE TEAM (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

RAMSAY, J. O., SILVERMAN, B. W. (2005). *Functional Data Analysis*, Second Edition, Springer.

RAMSAY, J. O., WICKHAM, H., GRAVES, S., HOOKER, G., (2018). fda: Functional Data Analysis. R package version 2.4.8. https://CRAN.R-project.org/ package=fda

RIZZO, M. L., SZÉKELY, G. J., (2019). energy: E-Statistics: Multivariate Inference via the Energy of Data. R package version 1.7-6.
https://CRAN.R-project.org/package=energy

SZÉKELY, G. J., RIZZO, M. L., BAKIROV, N. K., (2007). Measuring and Testing Dependence by Correlation of Distances. *The Annals of Statistics*, 35, pp. 2769–2794.

WANG, J. L., CHIOU, J. M., Müller, H. G., (2016). Functional Data Analysis. *Annual Review of Statistics and Its Application*, 3, pp. 257–295.

ZHANG, J. T., (2013). *Analysis of Variance for Functional Data*. Chapman & Hall: London.

ZOLOTAREV, V. M., (1981). Integral Transformations of Distributions and Estimates of Parameters of Multidimensional Spherically Symmetric Stable Laws. In *Contributions to Probability: A Collection of Papers Dedicated to Eugene Lukacs*, Academic Press, New York-London, pp. 283–305.