Journal of Multivariate Analysis 173 (2019) 125–144

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

On some characterizations and multidimensional criteria for testing homogeneity, symmetry and independence

Feifei Chen [a], Simos G. Meintanis [b,c,1], Lixing Zhu [c,d,*]

[a] School of Statistics, Renmin University of China, Beijing, PR China
[b] Department of Economics, National and Kapodistrian University of Athens, Athens, Greece
[c] Unit for Business Mathematics and Informatics, North–West University, Potchefstroom, South Africa
[d] Department of Mathematics, Hong Kong Baptist University, Hong Kong

ARTICLE INFO

Article history:
Received 26 April 2018
Received in revised form 13 February 2019
Accepted 13 February 2019
Available online 22 February 2019

AMS 2010 subject classifications:
primary 62G10
secondary 62H15

Keywords:
Characteristic function
Distance correlation
Independence testing
Symmetry testing
Two-sample problem

ABSTRACT

We propose three new characterizations and corresponding distance-based weighted test criteria for the two-sample problem, and for testing symmetry and independence with multivariate data. All quantities have the common feature of involving characteristic functions, and it is seen that these quantities are intimately related to some earlier methods, thereby generalizing them. The connection rests on a special choice of the weight function involved. Equivalent expressions of the distances in terms of densities are given as well as a Bayesian interpretation of the weight function is involved. The asymptotic behavior of the tests is investigated both under the null hypothesis and under alternatives, and affine invariant versions of the test criteria are suggested. Numerical studies are conducted to examine the performances of the criteria. It is shown that the normal weight function, which is the hitherto most often used, is seriously suboptimal. The procedures are biased in the sense that the corresponding test criteria degenerate in high dimension and hence a bias correction is required as the dimension tends to infinity.

1. Introduction

In this paper we are concerned with three major problems of statistical inference, namely those of testing homogeneity, testing symmetry and testing for independence. Specifically, and in the context of each problem, we will formulate a population measure which characterizes the underlying stochastic property of homogeneity, symmetry or independence in the sense of taking value zero under the corresponding null hypothesis while being strictly positive otherwise. We also propose empirical versions of the three population measures and study several aspects of the resulting test criteria including asymptotic as well as finite-sample behavior. The common theme of the aforementioned population measures is that they are expressed as weighted L_2-type distances involving the characteristic function (CF) of the underlying law. Particular instances of these population measures have appeared as characterizations earlier in the literature and corresponding test statistics have gained considerable popularity. We will show that most of these special cases result

* Corresponding author at: Department of Mathematics, Hong Kong Baptist University, Hong Kong.
E-mail address: lzhu@hkbu.edu.hk (L. Zhu).

1 On sabbatical leave from the University of Athens.

https://doi.org/10.1016/j.jmva.2019.02.006

from choosing a particular member of the spherical stable family of distributions in the aforementioned weighting scheme. In this connection we will make repeated use of the fact that the CF of a spherical stable law is given by

$$\varphi_Z(t)=\int_{\mathbb{R}^p}\cos(t^\top z)f_{\alpha,p}(z)dz=\exp(-\|t\|^\alpha), \tag{1}$$

where $f_{\alpha,p}$ stands for the density of a spherical stable law in $\mathbb{R}^p$ with characteristic exponent $\alpha\in(0,2]$, and $\varphi_Z(t)$ denotes the CF of a given random variable Z, $\|\cdot\|$ denotes the Euclidean norm. When clear from the context we will suppress the dimension and simply write f_α. The spherical stable family includes the multivariate standard Gaussian and Cauchy distributions as special cases, for $\alpha=2$ and $\alpha=1$, respectively. Further information regarding these distributions, including expressions for corresponding distribution functions and densities, may be found in [32,48].

The remainder of the paper is as follows. In Section 2 we introduce the three null hypotheses separately and in each case we deduce the appropriate distance measure and the corresponding characterization; we also provide connections between these quantities and the existing literature. Section 3 provides interpretations of these distance measures in terms of density deviations. In Section 4, three test criteria are computed, each for the respective null hypothesis considered and it is seen that earlier test criteria are particular instances of our quantities. A Bayesian type interpretation for the choice of the weight function and affine invariant versions of the tests statistics are also discussed. Asymptotic results can be found in Section 5, while the results of a series of Monte Carlo trials are presented and discussed in Section 6. The paper concludes in Section 7 with a summary of findings. All proofs of the theoretical results are presented in the Appendix.

2. Population distances

2.1. Homogeneity distance

Let X and Y be two independent random vectors of dimension p and suppose that we wish to test the null hypothesis

$$\mathcal{H}_0: X=_d Y, \tag{2}$$

where $=_d$ stands for equality in law.

The approach followed here is based on the uniqueness property between a given distribution and its CF, stating that the null hypothesis $\mathcal{H}_0$ in (2) holds if and only if

$$\forall_{t\in\mathbb{R}^p}\quad \varphi_X(t)=\varphi_Y(t). \tag{3}$$

The natural L_2-type distance corresponding to (3) is then given by

$$\mathcal{D}_w=\int_{\mathbb{R}^p}|\varphi_X(t)-\varphi_Y(t)|^2\, w(t)dt, \tag{4}$$

where $w>0$, a.e., is a weight function the role of which we will emphasize in this work. To begin with we note that in (4) as well as in the other distance measures that follow, a weight function is necessary when the CFs involved are non-integrable. In any case w is an indispensable part of the sample versions considered in Section 3 even with integrable population CFs.

On the basis of w, two types of homogeneity criteria have been considered in the literature. The first is based on a certain non-integrable weight function initially suggested by Székely and Rizzo [40]. Subsequently this approach gained considerable popularity, and the test criterion was later followed up by several authors, and for different problems; see, e.g., [26,36]. The second type of CF-based two-sample test statistics involves an integrable weight function and was suggested in the univariate case in [27], and later extended to vectorial observations in [22].

Here we will show that the two aforementioned approaches are intimately related, with the connecting element being a very special type of weight function. To see this, first recall from [40] that (3) is equivalent to

$$\mathrm{E}\{\cos t^\top(X-X_1)+\cos t^\top(Y-Y_1)-2\cos t^\top(X-Y)\}=0, \tag{5}$$

where $X_1=_d X$ (resp. $Y_1=_d Y$) with X_1 (resp. Y_1) independent of X (resp. Y).

The approach based on integrable weight functions admits an interesting specification whereby we replace the weight function w by the density, say f_T, of a random variable T with corresponding CF φ_T. We assume that f_T is positive with probability 1. Based on this specification we have the following novel characterization.

Proposition 1. *Let X and Y be two arbitrary independent vectors in $\mathbb{R}^p$. Then for any density f_T, which is positive with probability 1, with CF φ_T, the quantity $\mathcal{D}_f$ defined by*

$$\mathcal{D}_f=\mathrm{E}[\mathrm{Re}\{\varphi_T(X-X_1)+\varphi_T(Y-Y_1)-2\varphi_T(X-Y)\}], \tag{6}$$

where $\mathrm{Re}(z)$ *denotes the real part of a complex number* z*, is equal to zero if and only if* $X =_d Y$*, and it is otherwise strictly positive.*

Note that the distance $\mathcal{D}_f$ is a (squared) maximum mean discrepancy distance in the sense of [17], with associated kernel $k(x,y) = \mathrm{Re}\{\varphi_T(x-y)\}$.

Now if we replace f_T by the density of the spherical stable distribution f_α, then in view of (1) we can write the distance resulting from (6), say $\mathcal{D}_\alpha$, as

$$\mathcal{D}_\alpha = \mathrm{E}(e^{-\|X-X_1\|^\alpha} + e^{-\|Y-Y_1\|^\alpha} - 2e^{-\|X-Y\|^\alpha}), \tag{7}$$

which leads to the following result.

Corollary 1. *Let* X *and* Y *be two arbitrary independent vectors in* $\mathbb{R}^p$*. Then for any fixed* $\alpha \in (0,2]$*, the quantity* $\mathcal{D}_\alpha$ *defined in (7) is equal to zero if and only if* $X =_d Y$*, and it is otherwise strictly positive.*

An application of the approximation $e^x \approx 1 + x$, to (7) leads to

$$\widetilde{\mathcal{D}}_\alpha = \mathrm{E}\left\{2\|X-Y\|^\alpha - \|X-X_1\|^\alpha - \|Y-Y_1\|^\alpha\right\}, \tag{8}$$

with $\widetilde{\mathcal{D}}_\alpha$ being precisely the generalized two-sample energy statistic of Székely and Rizzo [42]. Note that this energy statistic makes use of the integral

$$\int_{\mathbb{R}^p} \frac{1-\cos(t^\top z)}{C\|z\|^{\alpha+p}}\, dz = \|t\|^\alpha, \tag{9}$$

where $\alpha \in (0,2)$ and C denotes a constant depending only on α and p.

There exist two more directions leading from (7) to (8) which provide further insight. The first comes from the fact that the Székely and Rizzo distance $\widetilde{\mathcal{D}}_\alpha$ results by means of the integral in (9) if in (A.1) in the Appendix we replace $\cos(\cdot)$ by $1-\cos(\cdot)$ and the weight function w by $\|t\|^{-(\alpha+p)}$. In this connection, recall that the density of a spherical stable distribution satisfies $f_\alpha(t) \approx \|t\|^{-(\alpha+p)}$ as $t \to \infty$; see [32]. Hence $\widetilde{\mathcal{D}}_\alpha$ may be viewed as resulting by way of approximation from $\mathcal{D}_\alpha$ when the density f_α which is used as weight function is replaced by its approximation for large arguments t.

The other direction leading from $\mathcal{D}_\alpha$ to $\widetilde{\mathcal{D}}_\alpha$ is as follows. Choose, instead of f_α, the density of the random variable $T/\gamma^{1/\alpha}$ as weight function in (6), for some $\gamma > 0$, and note that the CF corresponding to this density is given by $e^{-\|u\|^\alpha/\gamma}$. This yields a corresponding distance, say $\Delta_{\alpha,\gamma}$, analogous to that in (7) but with $\|\cdot\|^\alpha$ being replaced by $\|\cdot\|^\alpha/\gamma$ throughout Eq. (7). Now if we take a two-term expansion $e^{-\|x\|^\alpha/\gamma} = 1 - \|x\|^\alpha/\gamma + o(1/\gamma)$, as $\gamma \to \infty$, in the new distance $\Delta_{\alpha,\gamma}$, this will lead to

$$\lim_{\gamma\to\infty} \gamma\Delta_{\alpha,\gamma} = \widetilde{\mathcal{D}}_\alpha. \tag{10}$$

Eq. (10) shows that the energy population distance in (8) is a limiting instance of the distance in (4), in the case when the density of a properly scaled spherical stable density is used as weight function. Specifically this special case is recovered in the limit as the aforementioned scale parameter involved grows large.

The connections given above bring together two seemingly different approaches for two-sample testing: The approach utilizing (1) which employs an integrable weight function and the approach of employing the non-integrable weight function figuring in (9). However we must stress that $\mathcal{D}_\alpha$ is defined for any pair of random vectors without the moment condition $\mathrm{E}\|X\|^\alpha, \mathrm{E}\|Y\|^\alpha < \infty$, which is implicit in the definition of $\widetilde{\mathcal{D}}_\alpha$. Also while in $\widetilde{\mathcal{D}}_\alpha$, the value $\alpha = 2$ is excluded from the characterization, it is clear that this restriction no longer applies to $\mathcal{D}_\alpha$.

2.2. Symmetry distance

Suppose that X is an arbitrary random vector of dimension p, and that we wish to test the null hypothesis

$$\mathcal{H}_0 : X =_d -X, \tag{11}$$

of symmetry around the origin. In the Fourier domain, the null hypothesis (11) is restated as

$$\forall_{t\in\mathbb{R}^p} \quad \mathrm{Im}\{\varphi_X(t)\} = 0, \tag{12}$$

with a corresponding distance

$$\int_{\mathbb{R}^p} [\mathrm{Im}\{\varphi_X(t)\}]^2 w(t) dt,$$

where $\mathrm{Im}(z)$ denotes the imaginary part of a complex number z. Now observe that (12) holds if and only if

$$\mathrm{E}\{\cos t^\top(X-X_1) - \cos t^\top(X+X_1)\} = 0, \tag{13}$$

where X_1 is defined below (5), while the same quantity is positive otherwise. This observation leads us to the following characterization.

Proposition 2. *Let $X \in \mathbb{R}^p$ be an arbitrary random vector. Then for any density f_T, which is positive with probability* 1*, with CF φ_T, the quantity D_f defined by*

$$D_f = \mathrm{E}[\mathrm{Re}\{\varphi_T(X - X_1) - \varphi_T(X + X_1)\}], \tag{14}$$

is equal to zero if and only if X is symmetric around zero, and it is otherwise strictly positive.

Now if we replace f_T by the density of the spherical stable distribution f_α, then in view of (1) we can write the distance resulting from (14), say D_α, as

$$D_\alpha = \mathrm{E}(e^{-\|X-X_1\|^\alpha} - e^{-\|X+X_1\|^\alpha}), \tag{15}$$

which leads to the following result.

Corollary 2. *Let $X \in \mathbb{R}^p$ be an arbitrary random vector. Then for any fixed $\alpha \in (0, 2]$, the quantity D_α defined in (15) is equal to zero if and only if X is symmetric around zero, and it is otherwise strictly positive.*

By following the approximation arguments of Section 2.1, we can see that D_α in (15) may be approximated by the distance $\widetilde{D}_\alpha = \mathrm{E}(\|X + X_1\|^\alpha - \|X - X_1\|^\alpha)$, analogously as in (10). In turn the last distance results (apart from sign) by means of (9), if we write $1 - \cos(\cdot)$ instead of $\cos(\cdot)$ and replace $w(t)$ by $(C\|t\|)^{-(\alpha+p)}$ in (A.3) in the Appendix.

Hence here we have two levels of generalization. At the first level, the distance $\widetilde{D}_\alpha$ generalizes the symmetry distance and the corresponding characterization in [7,39,44], from $\widetilde{D}_1 = \mathrm{E}(\|X + X_1\| - \|X - X_1\|)$ to $\widetilde{D}_\alpha$ with arbitrary $\alpha \in (0, 2]$, while at the second level $\widetilde{D}_\alpha$ is obtained from D_α by means of a simple limiting argument.

2.3. Independence distance

Suppose now that X and Y are two arbitrary random vectors, of dimensions p and q, respectively, and that we wish to test the null hypothesis

$$\mathcal{H}_0 : X \perp Y, \tag{16}$$

where $\perp$ stands for independence in law. In the Fourier domain (16) may be stated as

$$\forall_{(t,s)\in\mathbb{R}^{p+q}} \quad \varphi_{X,Y}(t, s) = \varphi_X(t)\varphi_Y(s),$$

with corresponding distance given by

$$\mathbb{D}_w = \int_{\mathbb{R}^{p+q}} |\varphi_{X,Y}(t, s) - \varphi_X(t)\varphi_Y(s)|^2 w(t, s)dtds. \tag{17}$$

Unlike in Sections 2.1–2.2, we will elaborate here more since the derivations are slightly more involved. Specifically, compute

$$\begin{aligned} |\varphi_{X,Y}(t, s) - \varphi_X(t)\varphi_Y(s)|^2 &= \{\varphi_{X,Y}(t, s) - \varphi_X(t)\varphi_Y(s)\}\{\overline{\varphi_{X,Y}(t, s) - \varphi_X(t)\varphi_Y(s)}\} \\ &= |\varphi_{X,Y}(t, s)|^2 + |\varphi_X(t)|^2|\varphi_Y(s)|^2 - 2\mathrm{Re}\{\varphi_{X,Y}(t, s)\overline{\varphi_X(t)\varphi_Y(s)}\}, \end{aligned} \tag{18}$$

where $\bar{z}$ denotes the conjugate of a complex number z, and further notice that

$$|\varphi_{X,Y}(t, s)|^2 = \mathrm{E}[\cos\{t^\top(X - X_1) + s^\top(Y - Y_1)\}], \tag{19}$$

$$|\varphi_X(t)|^2|\varphi_Y(s)|^2 = \mathrm{E}\{\cos t^\top(X - X_1)\}\mathrm{E}\{\cos s^\top(Y - Y_1)\}, \tag{20}$$

$$\mathrm{Re}\{\varphi_{X,Y}(t, s)\overline{\varphi_X(t)\varphi_Y(s)}\} = \mathrm{E}[\cos\{t^\top(X - X_1) + s^\top(Y - Y_2)\}], \tag{21}$$

with (X_1, Y_1), (X_2, Y_2) being independent copies of (X, Y).

We will also assume that the weight function may be decomposed as $w(t, s) = w_p(t)w_q(s)$, where w_p and w_q are functions defined in the corresponding dimensions. We note in passing that while the product decomposition is not formally essential, it considerably simplifies expressions without giving up much generality, and therefore has been typically followed by researchers dating back to [37], and was later adopted in [12,43], among others.

With these observations, we are led to the following novel characterization.

Proposition 3. *Let $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ be two arbitrary random vectors. Then for any pair of densities f_{T_p} and f_{T_q}, which are positive with probability 1, with corresponding CFs φ_{T_p} and φ_{T_q}, the quantity $\mathbb{D}_f$ defined by*

$$\begin{aligned}\mathbb{D}_f =& \mathrm{E}\left[\mathrm{Re}\{\varphi_{T_p}(X-X_1)\}\mathrm{Re}\{\varphi_{T_q}(Y-Y_1)\}\right] + \mathrm{E}\left[\mathrm{Re}\{\varphi_{T_p}(X-X_1)\}\right]\mathrm{E}\left[\mathrm{Re}\{\varphi_{T_q}(Y-Y_1)\}\right] \\ &- 2\mathrm{E}\left[\mathrm{Re}\{\varphi_{T_p}(X-X_1)\}\mathrm{Re}\{\varphi_{T_q}(Y-Y_2)\}\right],\end{aligned} \tag{22}$$

is equal to zero if and only if X and Y are mutually independent, and it is otherwise strictly positive.

Now if replace f_{T_p} by $f_{\alpha,p}$ and f_{T_q} by $f_{\alpha,q}$, i.e., the densities of spherical stable distributions (in the corresponding dimensions), we can write the distance resulting from (22), say $\mathbb{D}_\alpha$, as

$$\mathbb{D}_\alpha = \mathrm{E}(e^{-(\|X-X_1\|^\alpha+\|Y-Y_1\|^\alpha)}) + \mathrm{E}(e^{-\|X-X_1\|^\alpha})\mathrm{E}(e^{-\|Y-Y_1\|^\alpha}) - 2\mathrm{E}(e^{-(\|X-X_1\|^\alpha+\|Y-Y_2\|^\alpha)}), \tag{23}$$

which leads to the following result.

Corollary 3. *Let $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ be two arbitrary random vectors. Then for any fixed $\alpha \in (0, 2]$, the quantity $\mathbb{D}_\alpha$ defined in (23) is equal to zero if and only if X and Y are mutually independent, and it is otherwise strictly positive.*

We note that for $\alpha = 1$, the distance measure $\mathbb{D}_\alpha$ may be heuristically described as an "exponentiated" version of the distance covariance of [43] given by

$$\mathrm{E}(\|X-X_1\|\|Y-Y_1\|) + \mathrm{E}(\|X-X_1\|)\mathrm{E}(\|Y-Y_1\|) - 2\mathrm{E}(\|X-X_1\|\|Y-Y_2\|).$$

At the same time our approximation techniques of Section 2.1 do not lead in this case to any sort of formal relation between $\mathbb{D}_\alpha$ and the distance covariance of [43] or its generalized version in [42] being defined for arbitrary $\alpha \in (0, 2]$.

3. Interpretations

While L_2 distances based on CFs are by all means formally acceptable, it would be interesting to related these distances with distances involving more traditional statistical quantities. To this end we will make use of the Parseval–Plancherel identity which states that

$$\|\varphi_X - \varphi_Y\|^2 = (2\pi)^p\|f_X - f_Y\|^2 \tag{24}$$

where

$$\|\varphi_X - \varphi_Y\|^2 = \int_{\mathbb{R}^p} |\varphi_X(t) - \varphi_Y(t)|^2 dt, \ \ \|f_X - f_Y\|^2 = \int_{\mathbb{R}^p} \{f_X(t) - f_Y(t)\}^2 dt,$$

for any given pair of CFs and corresponding densities, provided that both sides of (24) exist.

Now for the distance $\mathcal{D}_w$ in (4), consider as a tentative weight function the density f_α and assume that this density is proportional to φ_Z^2, where φ_Z is the CF of some symmetric around zero distribution. Prominent examples of this formulation will be given in the last paragraph of this section. Then replace the weight function in (4) by φ_Z^2, and apply (24) to the resulting quantity to get

$$\mathcal{D}_w = \mathcal{D}_{\varphi_Z^2} = \|\varphi_{X\star Z} - \varphi_{Y\star Z}\|^2 = (2\pi)^p\|f_{X\star Z} - f_{Y\star Z}\|^2,$$

where $\star$ denotes convolution. Hence the weighted L_2 distance between the CFs of a pair of variables X and Y may be equivalently interpreted as an L_2 distance between two densities: The density of the convolution $X\star Z$ and the density of the convolution $Y \star Z$, where Z is the random variable, the CF of which we have used as weight function in formulating the distance $\mathcal{D}_w$ as $\mathcal{D}_{\varphi_Z^2}$.

Along the same lines, expressions involving distances between densities may be obtained in the case of symmetry and independence but these are not so straightforward and illuminating to interpret. For example the interpretation for the symmetry distance D_w in (A.3) is recovered by noting that

$$\{\mathrm{Im}(\varphi_X(t))\}^2 = [|\varphi_X(t)|^2 - \{\varphi_X^2(t) + \varphi_X^2(-t)\}/2]/2.$$

In turn if we observe that $|\varphi_X(t)|^2$ is the CF of $X \star (-X_1)$ and $\{\varphi_X^2(t) + \varphi_X^2(-t)\}/2$ is the CF of a random variable which is equal to $X \star X_1$ and $(-X) \star (-X_1)$, each with probability 1/2, this interpretation ultimately rests on the fact that a given random variable X is symmetrically distributed around zero if and only if it satisfies $X =_d Y$, where Y is such that $\Pr(Y = X) = 1 - \Pr(Y = -X_1) = 1/2$.

We now go back to the original formulation and consider the question of which members of the spherical stable family qualify for the interpretation given here. The most prominent example is the multivariate standard normal density which, recall, is a special case of the spherical stable family for $\alpha = 2$, and could serve both as a density as well as a CF within the multivariate normality class. The other, less well-known member, is the multivariate standard Cauchy distribution, resulting from the same family for $\alpha = 1$, which after scaling has a density that coincides with the CF of a symmetric multivariate generalized Laplace distribution; see [24]. Note finally that both the normal and the generalized Laplace are infinitely divisible distributions so that their CFs have squares that are also CFs and belong to the same family.

4. Sample distances: Computations and affine invariance

4.1. Computations and related issues

In this section we will provide sample analogs of the distance measures introduced in Sections 2.1–2.3, in the form of V-statistics. To begin with the symmetry statistic, suppose that $X_1, \ldots, X_n$ are independent copies of the random variable X. Then the test statistic corresponding to the symmetry distance measure D_α in (15) is given by

$$D^s_{n,\alpha} = \frac{1}{n^2} \sum_{1\le j,k\le n} (e^{-\|X_j - X_k\|^\alpha} - e^{-\|X_j + X_k\|^\alpha}). \tag{25}$$

Likewise suppose that $Y_1, \ldots, Y_n$ are independent copies of the random variable Y. Then the test statistic corresponding to the homogeneity distance measure $\mathcal{D}_\alpha$ in (7) is given by

$$\mathcal{D}^h_{n,\alpha} = \frac{1}{n^2} \sum_{1\le j,k\le n} (e^{-\|X_j - X_k\|^\alpha} + e^{-\|Y_j - Y_k\|^\alpha} - 2e^{-\|X_j - Y_k\|^\alpha}). \tag{26}$$

In turn the test criterion corresponding to the independence distance measure $\mathbb{D}_\alpha$ in (23) is given by

$$\mathbb{D}^I_{n,\alpha} = \frac{1}{n^2} \sum_{1\le j,k\le n} \left\{e^{-(\|X_j - X_k\|^\alpha + \|Y_j - Y_k\|^\alpha)}\right\} + \left(\frac{1}{n^2} \sum_{1\le j,k\le n} e^{-\|X_j - X_k\|^\alpha}\right)\left(\frac{1}{n^2} \sum_{1\le j,k\le n} e^{-\|Y_j - Y_k\|^\alpha}\right) - \frac{2}{n^3} \sum_{1\le j,k,\ell\le n} (e^{-(\|X_j - X_k\|^\alpha + \|Y_j - Y_\ell\|^\alpha)}). \tag{27}$$

The symmetry statistic $D^s_{n,\alpha}$ dates back to the criterion of [13] for univariate symmetry, and was later generalized to arbitrary dimension in [15,20]; see also [30]. In fact $D^s_{n,\alpha}$ is a extended version of the statistic of [20] from $\alpha = 2$ to any $\alpha \in (0, 2]$, and, with proper standardization, continues to satisfy several nice properties of that statistic such as affine invariance. Particular cases of the homogeneity statistic in (26) have appeared in [1,2,4,22,34]. Most of these statistics are special cases of $\mathcal{D}^h_{n,\alpha}$, most often for $\alpha = 2$ but also for $\alpha = 1$. Finally, and beyond the already mentioned heuristic connection with the distance covariance, various authors, e.g., [6,10,21,23,28,38,42], have used statistics analogous to $\mathbb{D}^I_{n,\alpha}$ for testing independence of the components of a given random vector. Their statistics are often special cases of $\mathbb{D}^I_{n,\alpha}$, for $\alpha = 2$. By contrast, Bilodeau and Guetso Nangue [5] make use of the full spectrum of values $\alpha \in (0, 2]$, while Fan et al. [11] propose several weight functions, including the Gaussian as well as the Székely–Rizzo weight function.

Remark 1. Here we define the test criteria in terms of V-statistics as it is common with most researchers working with CF-based statistics; see, e.g., [1,4,6,22,26,34,40]. Clearly though, there is an analogous formulation using a corresponding U-statistic. We illustrate this U/V dichotomy only for the homogeneity statistic, but similar arguments apply to the other two test criteria. To this end, assume that the null hypothesis $\mathcal{H}_0 : X =_d Y$ is true, and observe that

$$\mathcal{D}^h_{n,\alpha} = \mathcal{U}^h_{n,\alpha} + \{2 - 2\mathrm{E}(e^{-\|X-Y\|^\alpha})\}/n + o_p(1/n),$$

where $\mathcal{U}^h_{n,\alpha}$ is the corresponding U-statistic. Hence when multiplied by the normalizing constant n, the V-statistic differs from the U-statistic asymptotically by the constant $2 - 2\mathrm{E}(e^{-\|X-Y\|^\alpha})$. In this connection, and although the U-statistic is an unbiased estimate of its target, we note that the corresponding V-statistic is asymptotically equivalent and is expected to achieve the same power, at least for large sample size n.

We now provide an interesting Bayesian-type interpretation of the weight function. Specifically, replacing in Section 2 the weight function $w(t)$ by a density is like treating the argument t as a random quantity with the density $f_T(t)$ acting as a Bayesian prior on it. In this context the hitherto popular choice $\alpha = 2$ currently dominating the literature may be interpreted as choosing the multivariate standard normal density, which is a medium-tailed prior. This is equivalent to emphasizing values of the underlying CF for t close to the origin. In contrast $\alpha < 2$ yields a heavy-tailed density (and progressively more so as α decreases towards zero), which amounts to putting considerable emphasis also on values of the argument t away from zero. Recall, however, that the behavior of a given CF around the origin reflects the tail-properties of the underlying law, while the same behavior for large t is related to the smoothness of the corresponding density. Hence taking a smaller α is like shifting the emphasis from the tail properties of the underlying distribution to the smoothness properties of its density.

This observation invites a possible data-dependent choice of the weight function, whereby α is first estimated from the data and then from this estimate we choose the weight function in (1) with α replaced by $\hat{\alpha}$. Such an estimator could be either parametric (i.e., imposing a stable distribution on the data) or entirely nonparametric as in our context. Parametric/nonparametric tail-index estimation is a widely studied problem in the univariate setting (see for instance [8,29,33], but definitely less so in the vectorial context; nevertheless see [9,32]. In either case though the asymptotics of such data-driven test criteria will be different from those presented in Section 5, and need to be studied separately, but we anticipate that this problem will be far from trivial.

We close this section by pointing out that while, due to its popularity, the case in point here is the spherical stable density as weight function, our approach is nevertheless quite general. Specifically, Propositions 1–3 of Section 2 apply and yield corresponding characterizations and test statistics for any appropriately chosen CF. In this regard (6), (14), and (22) are the key equations here that are characterization-producing, in the sense that by replacing the corresponding CFs involved therein (φ_T, in (6) and (14), and φ_{T_p} and φ_{T_q} in (22)) by specific instances leads to corresponding characterizations which in turn then, and by proper estimation, render new test criteria. In particular mixtures of spherical stable distributions, mixtures of normal distributions, as well as multivariate generalized Laplace distributions would yield qualitatively much the same results if used as weight functions, only with aggravated computational formulae.

4.2. Affine invariance

When proposing measures of goodness-of-fit, affine invariance is one of the important features that should be taken into account. In this connection recall that conditionally on a set of available data, say $X_1, \ldots, X_n \in \mathbb{R}^p$, an arbitrary test statistic $T_n = T_n(X_1, \ldots, X_n)$ is affine invariant if

$$T_n(AX_1 + a, \ldots, AX_n + a) = T_n(X_1, \ldots, X_n), \tag{28}$$

for each nonsingular $p \times p$ matrix A and each vector $a \in \mathbb{R}^p$. Property (28) with $a = 0$ is termed scale invariance while the same property with A replaced by the identity matrix is termed location invariance. Clearly a test statistic should be affine invariant only if the corresponding null hypothesis is also affine invariant, and in this sense, as it will be seen below, the property is specific to the null hypothesis being tested. For instance, Henze [19] shows that any affine invariant test for multivariate normality should depend on the original observations $X_1, \ldots, X_n$ only via the corresponding Mahalanobis squared radii $D_X^{(jk)} = (X_j - \bar{X})^\top S_X^{-1}(X_k - \bar{X})$, where $\bar{X}$ and S_X stand for the sample mean and the sample covariance matrix, respectively, of $X_1, \ldots, X_n$. In what follows we discuss this aspect of the new statistics.

It should be stated right from the outset that our test statistics as formulated in the previous subsection are not affine invariant, although the homogeneity statistic in (26) as well as the independence statistic in (27) are by construction location invariant. In this connection the aim of the following discussion is to investigate whether specific data transformations, such as the usual location/scale standardization, may result to affine invariant versions of the test criteria figuring in Eqs. (25)–(27).

We begin with the symmetry statistic by first noting that the null hypothesis in (11) of symmetry around zero is scale equivariant but not location invariant. So it only makes sense to talk about a scale invariant test statistic. To this end, the test statistic in (25) will be scale invariant if instead of the original observations we compute the test criterion based on the standardized data defined, for all $j \in \{1, \ldots, n\}$, by $Z_j = S_X^{-1/2} X_j$. Analogous considerations apply in the case of testing the null hypothesis of symmetry around an unknown center.

Turning to the case of homogeneity testing, we note first that the corresponding null hypothesis in (2) is affine invariant and therefore we require that our test criterion satisfies

$$T_n(AX_1 + a, \ldots, AX_n + a, AY_1 + a, \ldots, AY_n + a) = T_n(X_1, \ldots, X_n, Y_1, \ldots, Y_n),$$

for each vector a and matrix A. However, the homogeneity statistic in (26) depends solely on distances between observations and hence it is clearly location invariant. Hence without loss of generality we may take $a = 0$ in the above equation, so that no data centering is necessary.

There exist two approaches to affine invariance of the homogeneity statistic. First we could use the standardized data $Z_{X,j} = S_X^{-1/2} X_j$ and $Z_{Y,j} = S_Y^{-1/2} Y_j$ with $j \in \{1, \ldots, n\}$, by estimating the covariance matrix separately for each of the two samples. This approach however leaves the terms $\|X_j - Y_k\|$ in (26) non-invariant, with the degree of non-invariance depending on the size of discrepancy between the two sample covariance matrices S_X and S_Y. Then again if we use a pooled covariance matrix estimator, say $S_{X,Y} = S(X_1, \ldots, X_n, Y_1, \ldots, Y_n)$, affine invariance requires that this estimator satisfies $S(AX_1 + a, \ldots, AX_n + a, AY_1 + a, \ldots, AY_n + a) = AS(X_1, \ldots, X_n, Y_1, \ldots, Y_n)A^\top$.

Clearly both approaches leave something to be desired, with the first leading to a test criterion that is really not affine invariant while the second approach may be viewed as imposing a common covariance structure on the two samples which would not make sense if the vectors are incommensurable. Nevertheless if the user is interested in shape differences between the two populations, i.e., in differences beyond location/scale, the latter standardization may seem a reasonable approach.

We finally consider the independence criterion in (27), by first noting that in the context of the corresponding null hypothesis in (16), affine invariance generalizes to

$$T_n(AX_1 + a, \ldots, AX_n + a, BY_1 + b, \ldots, BY_n + b) = T_n(X_1, \ldots, X_n, Y_1, \ldots, Y_n),$$

for each pair of constants a and b and each pair of nonsingular matrices A and B, in the corresponding dimensions. Clearly though, the test statistic in (27) is invariant under arbitrary location shifts of the vectors $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$, so for affine invariance it suffices to consider $a = b = 0$, in the above equation. In this connection if instead of the original observations we employ the standardized data $Z_{X,j} = S_X^{-1/2} X_j$ and $Z_{Y,j} = S_Y^{-1/2} Y_j$ with $j \in \{1, \dots, n\}$, then it may readily be shown that the test criterion resulting from (27) satisfies the above affine invariance equation, i.e., this type of scale standardization renders affine invariance.

We close this section by noting that the asymptotic properties in the case of the affine invariant versions of our test criteria will be different from those presented in the next section. Specifically although these limit distributions are obtained without any moment conditions on the underlying random vectors, this is not the case for their affine invariant versions. In particular, and as already discussed, such versions require proper standardization, and this standardization proves to be asymptotically influential. In this connection, the limit null distribution of the resulting affine invariant test statistic is obtained only under suitable moment conditions which relate to the specific standardization involved; see for instance [20,21,35].

By way of example consider the affine invariant version of the symmetry statistic. If in (30) instead of the original observations we employ the standardized data $Z_{X,j} = S_X^{-1/2} X_j$ with $j \in \{1, \dots, n\}$, the resulting sample distance measure now equals

$$\int_{\mathbb{R}^p} \left\{ \frac{1}{n} \sum_{j=1}^n \sin(t^\top Z_{X,j}) \right\}^2 w(t)dt = \int_{\mathbb{R}^p} \left\{ \frac{1}{n} \sum_{j=1}^n Z_{n,j}^s(t) \right\}^2 w(t)dt, \tag{29}$$

where $Z_{n,j}^s(t) = \sin(t^\top X_j) + t^\top \Delta_j \cos(t^\top X_j) + \epsilon_{n,j}(t)$ with $\Delta_j = (S_X^{-1/2} - I_p) X_j$. Note that due to affine invariance, and without loss of generality, we have assumed that the covariance matrix of X_j is equal to the identity matrix I_p. We then need to impose the condition that $\mathrm{E}\|X\|^4 < \infty$ in order to proceed as in Theorem 3.1 in [20] and show that the part of the test criterion in the right-hand side of (29) involving the errors $\epsilon_{n,j}(t)$ is asymptotically negligible. However by the same token, the part of the same quantity involving S_X has an effect on the null distribution, which differs from the one figuring in (31) in the covariance matrix of the limit process. In this connection, if an affine invariant version of the test statistic is desired then it is probably advisable to use instead of permutation tests, bootstrap tests which are more appropriate when parameters are being estimated; see for instance [14].

5. Asymptotic behavior

In this section, we will study the asymptotic behavior of our test statistics under both the null hypotheses as well as under general alternatives. To this end let

$$\begin{aligned}
\mathcal{D}_{n,w} &= \int_{\mathbb{R}^p} |\varphi_{n,X}(t) - \varphi_{n,Y}(t)|^2 w(t)dt, \\
D_{n,w} &= \int_{\mathbb{R}^p} [\mathrm{Im}\{\varphi_{n,X}(t)\}]^2 w(t)dt, \\
\mathbb{D}_{n,w} &= \int_{\mathbb{R}^{p+q}} |\varphi_{n,X,Y}(t,s) - \varphi_{n,X}(t)\varphi_{n,Y}(s)|^2 w(t,s)dtds,
\end{aligned} \tag{30}$$

denote the sample analogs of the distance measures introduced in Section 2, where $\varphi_{n,Z}(t) = \sum_{j=1}^n e^{it^\top Z_j}/n$, denotes the empirical CF computed on the basis of independent copies $Z_1, \dots, Z_n$ of a given random variable Z. A convenient setting for asymptotic theory is the separable Hilbert space of measurable real-valued functions defined on $\mathbb{R}^d$; we use the generic notation d for dimension, with d varying depending on context. In this context we impose the following assumptions.

Assumption (C). Let $w(t)$ be a measurable non-negative function on $\mathbb{R}^d$ such that, for all $t \in \mathbb{R}^d$, $w(t) = w(-t)$ and

$$0 < \int_{\mathbb{R}^d} w(t)dt < \infty.$$

Now denote by $\mathfrak{L}^2$ the separable Hilbert space of measurable real-valued functions on $\mathbb{R}^d$ that are square integrable with respect to $w(t)dt$, with the inner product and norm in $\mathfrak{L}^2$ respectively defined by

$$\langle f, g \rangle_w = \int_{\mathbb{R}^d} f(t)g(t)w(t)dt \quad \text{and} \quad \|f\|_w = \left\{ \int_{\mathbb{R}^d} f^2(t)w(t)dt \right\}^{1/2}.$$

We now formulate the asymptotic distribution of the homogeneity test statistic under the null hypothesis in the following theorem.

Theorem 1 (*Limit Null Distribution of the Homogeneity Statistic*)**.** *Let $X_1, \dots, X_n$ and $Y_1, \dots, Y_n$ be independent copies of X and Y. Under $\mathcal{H}_0$, suppose $\varphi_X(t) = \varphi_Y(t) = \varphi(t)$, then under Assumption (C), as $n \to \infty$,*

$$n\mathcal{D}_{n,w} \rightsquigarrow \|Z^h(t)\|_w^2,$$

where $\rightsquigarrow$ means the convergence in distribution, and $\{Z^h(t) : t \in \mathbb{R}^p\}$ is a zero-mean Gaussian process with covariance kernel defined, for all $s, t \in \mathbb{R}^p$, by

$$\mathrm{E}\{Z^h(s)Z^h(t)\} = 2[\mathrm{Re}\{\varphi(s-t)\} + \mathrm{Im}\{\varphi(s+t)\} - \mathrm{Re}\{\varphi(t)\}\mathrm{Re}\{\varphi(s)\} - \mathrm{Im}\{\varphi(s)\}\mathrm{Re}\{\varphi(t)\} - \mathrm{Im}\{\varphi(t)\}\mathrm{Im}\{\varphi(s)\}].$$

The next results show that the test based on $\mathcal{D}_{n,w}$ is consistent against general alternatives.

Theorem 2 (*Stochastic Limit of the Homogeneity Statistic*)**.** *Let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ be independent copies of X and Y. Then under Assumption (C), as $n \to \infty$,*

$$\mathcal{D}_{n,w} \xrightarrow{wp1} \|D^h(t)\|_w^2,$$

where $\xrightarrow{wp1}$ means convergence with probability 1, and $D^h(t) = |\varphi_X(t) - \varphi_Y(t)|$.

If the weight function is positive with probability 1, then in view of Proposition 1, the quantity $\|D^h(t)\|_w^2$ is equal to zero if and only if the null hypothesis of homogeneity figuring in (2) holds true, which implies the consistency of the test that rejects this hypothesis for large values of $\mathcal{D}_{n,w}$. As a particular case, Corollary 1 implies the consistency of the test based on $\mathcal{D}_{n,\alpha}^h$ which results by replacing $w(t)$ by the spherical stable density.

Now we give the analogous asymptotic results for the symmetry test and the independence test.

Theorem 3 (*Limit Null Distribution of the Symmetry Statistic*)**.** *Let $X_1, \ldots, X_n$ be independent copies of X. Under $\mathcal{H}_0$ and Assumption (C), as $n \to \infty$,*

$$nD_{n,w} \rightsquigarrow \|Z^s(t)\|_w^2, \tag{31}$$

where $\{Z^s(t) : t \in \mathbb{R}^p\}$ is a zero-mean Gaussian process with covariance function given, for all $s, t \in \mathbb{R}^p$, by $\mathrm{E}\{Z^s(s)Z^s(t)\} = \mathrm{E}\{\sin(s^\top X)\sin(t^\top X)\}$.

Theorem 4 (*Stochastic Limit of the Symmetry Statistic*)**.** *Let $X_1, \ldots, X_n$ be independent copies of X. Then under Assumption (C), as $n \to \infty$,*

$$D_{n,w} \xrightarrow{wp1} \|D^s(t)\|_w^2,$$

where $D^s(t) = \mathrm{Im}\{\varphi_X(t)\}$.

Theorem 5 (*Limit Null Distribution of the Independence Statistic*)**.** *Let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ be independent copies of X and Y. Under $\mathbb{H}_0$ and Assumption (C), as $n \to \infty$,*

$$n\mathbb{D}_{n,w} \rightsquigarrow \|Z^I(t,s)\|_w^2,$$

where $\{Z^I(t,s) : (t,s) \in \mathbb{R}^{p+q}\}$ is a zero-mean Gaussian process with the same covariance kernel as the process $\{Z_0^I(t,s) : (t,s) \in \mathbb{R}^{p+q}\}$ defined as

$$\begin{aligned} Z_0^I(t,s) = {} & \left[\cos(s^\top Y) - \mathrm{Re}\{\varphi_Y(s)\}\right]\left[\cos(t^\top X) + \sin(t^\top X) - \mathrm{Re}\{\varphi_X(t)\} - \mathrm{Im}\{\varphi_X(t)\}\right] \\ & + \left[\sin(s^\top Y) - \mathrm{Im}\{\varphi_Y(s)\}\right]\left[\cos(t^\top X) - \sin(t^\top X) - \mathrm{Re}\{\varphi_X(t)\} + \mathrm{Im}\{\varphi_X(t)\}\right]. \end{aligned} \tag{32}$$

Theorem 6 (*Stochastic Limit of the Independence Statistic*)**.** *Let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ be independent copies of X and Y. Then under Assumption (C), as $n \to \infty$,*

$$\mathbb{D}_{n,w} \xrightarrow{wp1} \|D^I(t,s)\|_w^2,$$

where $D^I(t,s) = |\varphi_{X,Y}(t,s) - \varphi_X(t)\varphi_Y(s)|$.

The asymptotic null distribution of each test criterion is obtained by means of the Hilbert space Central Limit Theorem coupled with the continuous mapping theorem, while the corresponding stochastic limit makes use of the point-wise consistency of the empirical CF and dominated convergence. Detailed proofs are given in the Appendix.

The asymptotic null distributions obtained in this section are highly non-standard. Typically the test criteria have the same limit distribution as an infinite linear combination of independent chi-squared distributions. The coefficients though of this linear combination are the eigenvalues of a complicated integral operator, and thus they are extremely hard to solve for analytically. The only fully analytic representation of the limit null distribution is by Baringhaus [3]; it corresponds to the special case of the fixed-location symmetry statistic in (25) with $\alpha = 2$. Otherwise one possible way for directly approximating this distribution is to numerically approximate a finite number of coefficients, replace the infinite sum by a finite one involving these numerically produced coefficients, and finally simulate by Monte Carlo the resulting (finite-sum) distribution of independent chi-squared variates; see, e.g., [11,30]. However, such a numerical approximation is a formidable task, and therefore most authors have resorted to properly chosen resampling techniques, tailored to each testing situation, in order to compute critical points and actually carry out the tests. These resampling techniques will be discussed in the next section.

6. Simulations

One of the most generally accepted strategies of comparing different tests is by studying their relative efficiency via comparison of Bahadur slopes; see [31]. This approach allows comparison of any given test against the likelihood ratio test, which is the best possible test in the direction of any specific alternative. However, the related theory is highly technical since it requires the study of large deviations which is a formidable task in its own right. This line of research was carried out by Tenreiro [45], thereby producing some technical guidance for the efficiency of a CF-based statistic for multivariate normality against certain types of skewed alternatives. It should be emphasized though that Tenreiro's analytical results apply in the very specific parametric context of normality testing and were obtained only in the univariate case, and that the corresponding multivariate case was investigated by means of a Monte Carlo study.

In view of the above discussion, and while a theoretical investigation of test efficiency is certainly an interesting subject for further research, here we resort to a Monte Carlo study in order to investigate the power properties of our tests. In the simulation studies, all tests are performed at the 5% significance level using 2000 trials. Since, as was shown in Section 5, the asymptotic behavior under the null hypothesis is not distribution-free in any of the three cases, we perform the tests via permutation resampling schemes. In this connection we point out that permutation tests are applicable and can be exact if the randomization hypothesis holds. For the hypotheses of homogeneity and independence in (2) and (16), this hypothesis reads as

$$\Upsilon^{(2n)} =_d \Upsilon^{(\pi_{2n})} \quad \text{and} \quad \Upsilon^{(2n)} =_d \Upsilon^{(\pi_n)}, \tag{33}$$

respectively, where $\Upsilon^{(2n)} = (X_1, \ldots, X_n, Y_1, \ldots, Y_n)$, $\Upsilon^{(\pi_{2n})}$ denotes the vector produced by an arbitrary permutation of all elements of $\Upsilon^{(2n)}$, while $\Upsilon^{(\pi_n)}$ denotes the vector produced by an arbitrary permutation of only the last n elements of $\Upsilon^{(2n)}$. Clearly under (2) (resp. (16)), the left-hand (resp. right-hand) side of (33) holds true. As for the symmetry hypothesis, the randomization hypothesis holds true under a different group-transformation. Specifically we have

$$X^{(n)} =_d X^{(\pi_n)},$$

where $X^{(n)} = (X_1, \ldots, X_n)$ and $X^{(\pi_n)} = (X_{\pi(1)}, \ldots, X_{\pi(n)})$, where $X_{\pi(j)} = U_j X_j$ with U_j being iid and independent of $(X_1, \ldots, X_n)$, and taking values ± 1 with equal probability 0.5. This type of resampling is often labeled as wild bootstrap. See [25] for more information on the randomization hypothesis, and [16] for a book-length treatment of permutation tests.

Our Monte Carlo approximation uses $N = 1000$ permutations, which is a good surrogate for the exact permutation distribution. The characteristic exponent values for the spherical stable density weight function are $\alpha \in \{0.5, 1.0, 1.5, 2.0\}$.

(i) Permutation-based homogeneity testing: We follow the permutation resampling scheme described above which has also been used in [22].
(ii) Permutation-based symmetry testing: We follow the wild bootstrap resampling scheme which has also been used in [20,47].
(iii) Permutation-based independence testing: We follow the permutation resampling scheme described above which has also been adopted in [43].

The Monte Carlo results are produced using the R software on the Windows platform. The average CPU time for a single Monte Carlo trial using 1000 permutations was 0.37 s, 0.94 s, and 5.12 s, for sample size $n = 20$, 40, and 100, respectively, using an Intel(R) Core(TM) i5-2500 CPU, at 3.30 GHz.

6.1. Comparison with energy statistics

For comparison purposes, we present three examples that compare the empirical size and power of our proposed tests with those of the energy statistics of [42]. The latter statistics are based on (6.1), (6.2) and (7.11) in [42] for testing homogeneity, symmetry and independence, respectively. All tests are implemented as permutation tests, with the software for the application of the energy statistics being available in the `energy` package for R.

The distributions considered are: The uniform distribution $\mathcal{U}(a, b)$ in the interval (a, b); the Fisher–Snedecor distribution $\mathcal{F}(n_1, n_2,\ \delta)$ with degrees of freedom (n_1, n_2) and non-centrality parameter δ; the Pareto distribution $\mathcal{P}(a, s)$ with shape parameter a and scale parameter s; the Student t-distribution $t(m)$ with degrees of freedom m; the Cauchy distribution $\mathcal{C}(\ell, s)$ with location parameter ℓ and scale parameter s. In this notation, $\mathcal{C}_p(\ell, s)$ denotes a p-dimensional Cauchy distribution, and analogously for other distributions.

Example 1. Homogeneity test: We consider three scenarios in this example. In Example 1(a), $X \sim \mathcal{U}_p(-1, 1)$ and $Y \sim \mathcal{U}_q(-1, 1+C)$, the notation $\mathcal{U}_p(-1, 1)$ stands for the coordinates of $X \in \mathbb{R}^p$ are independent and identically distributed as uniform distribution $\mathcal{U}(-1, 1)$. In the following, we just suppose that the coordinates of $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ are iid. We choose $C = 0 : 0.25 : 1$, here $0 : 0.25 : 1$ means a vector of evenly spaced points in the interval $[0, 1]$ with spacing equal

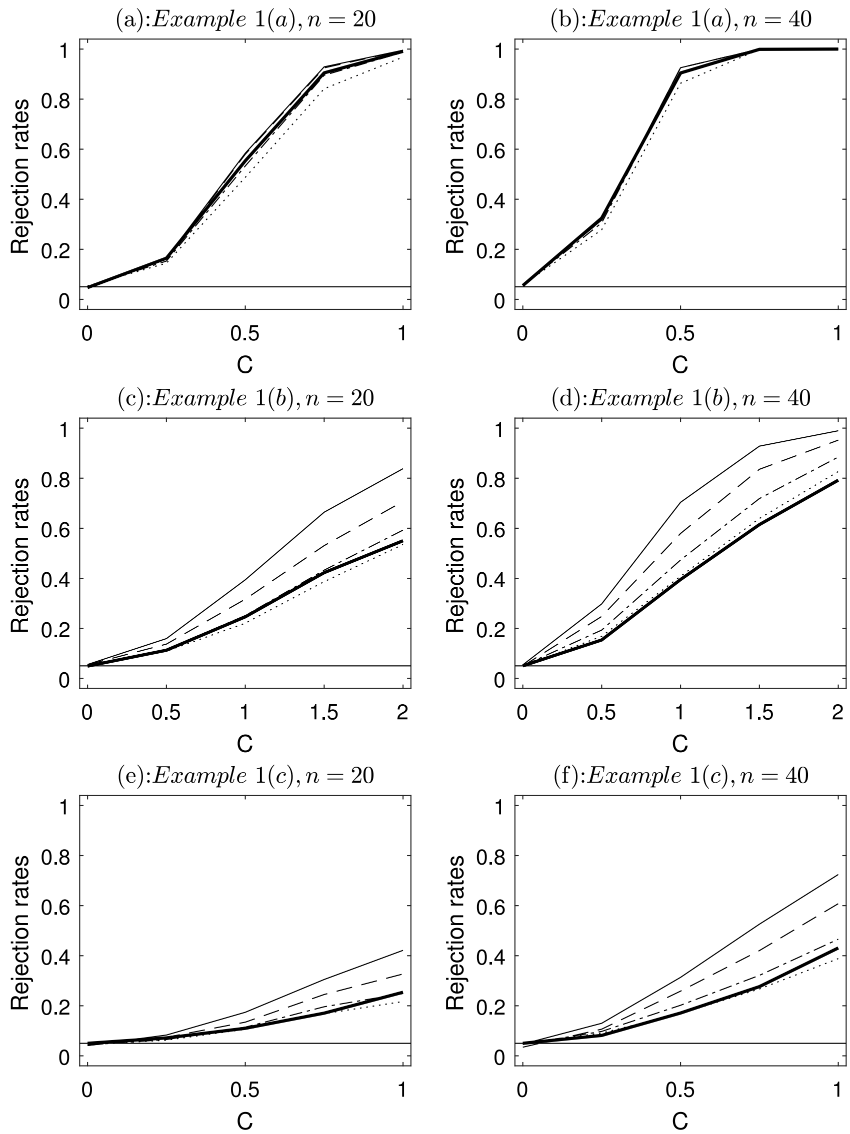

Fig. 1. Simulation results for Example 1. Rejection rates of the homogeneity test against the value of C at the 5% significance with dimensions $p = q = 5$ and sample size $n \in \{20, 40\}$. Characteristic exponent $\alpha = 0.5$ (solid line), $\alpha = 1.0$ (dashed line), $\alpha = 1.5$ (dash dot line) and $\alpha = 2.0$ (dotted line), energy statistics (thick line).

to 0.25, so that the choice of $C = 0$ corresponds to the null hypothesis, while with $C > 0$ we are in the alternative. The dimensions of X and Y considered are $p = q = 5$, with sample size $n \in \{20, 40\}$. In Example 1(b) we repeat Example 1(a) under identical conditions except that the coordinates of the random variables X and Y are independently generated from $\mathcal{F}(1, 2)$ and $\mathcal{F}(1, 2, C)$, and $C = 0 : 0.5 : 2$. In Example 1(c), $X \sim \mathcal{P}_p(1, 1)$, $Y \sim \mathcal{P}_q(1, 1+C)$ and $C = 0 : 0.25 : 1$. Note that the expectations of $\mathcal{F}(n_1, n_2, \delta)$ and $\mathcal{P}(a, s)$ exist only when $n_2 > 2$ and $a > 1$ respectively, hence X and Y do not have finite expectation in the latter two scenarios.

Example 2. Symmetry test. We consider two scenarios in this example, with finite (resp. infinite) expectation in the first (resp. second) scenario. That is, X comes either from $\mathcal{U}_p(-1, 1)$ or from $\mathcal{U}_p(-1, 1 - C)$ with equal probability 0.5 in Example 2(a), while X comes either from $\mathcal{C}_p(0, 1)$ or from $\mathcal{C}_p(C, 1)$ with equal probability 0.5 in Example 2(b). The dimension is $p = 5$ with sample size $n \in \{20, 40\}$, $C = 0 : 0.25 : 1$.

Example 3. Independence test: We generate two independent column vectors $X_0 \sim \mathcal{U}_p(0, 1), Y_0 \sim \mathcal{U}_q(0, 1)$ in Example 3(a) and $X_0 \sim t_p(1), Y_0 \sim t_q(1)$ in Example 3(b), with X_0 and Y_0 not having finite expectation in the latter

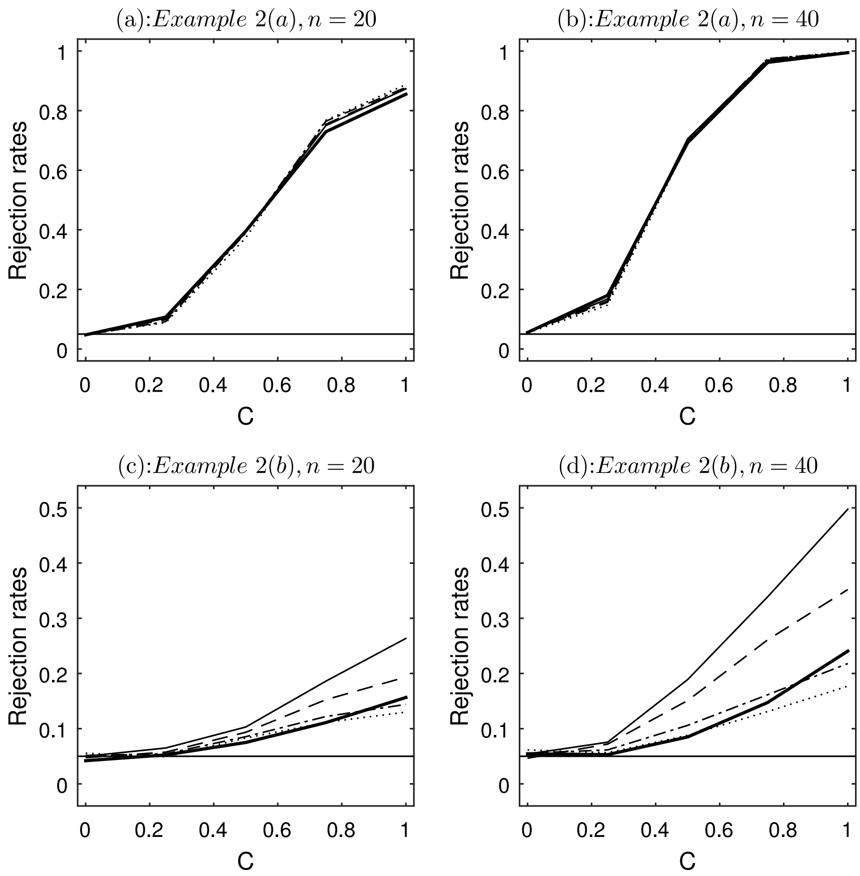

Fig. 2. Simulation results for Example 2. Rejection rates of the symmetry test against the value of C at the 5% significance with dimension $p = 5$ and sample size $n \in \{20, 40\}$. Characteristic exponent $\alpha = 0.5$ (solid line), $\alpha = 1.0$ (dashed line), $\alpha = 1.5$ (dash dot line) and $\alpha = 2.0$ (dotted line), energy statistics (thick line).

case. We set $X = X_0 + \Sigma Y_0$ and $Y = \Sigma^\top X_0 + Y_0$, where Σ is a $p \times q$ matrix with all elements equal to ρ. Thus ρ quantifies the dependence between X and Y. We take $\rho = 0 : 0.05 : 0.2$, hence $\rho = 0$ means X and Y are independent and $\rho > 0$ means the opposite. The dimensions are $p = q = 5$ with sample size $n \in \{20, 40\}$.

Figs. 1–3 display the comparative results of the tests for homogeneity, symmetry and independence, respectively. The following conclusions can be drawn. Both groups of tests have better performance with increasing sample size and distance from the null hypothesis. This shows the consistency of the tests. As expected, all proposed tests in this paper as well as the energy tests in [42] control the size well in lower dimension. It is concluded that the three tests are robust with respect to the choice of the characteristic exponent α in keeping the nominal size in lower dimensions. At the same time, however, the power varies considerably with the value of the characteristic exponent α, with higher power achieved for smaller values of α in almost all cases. This is a most interesting finding given the fact that, as already mentioned, most test criteria in the literature are confined to the value $\alpha = 2$. In addition, our proposed test criteria of homogeneity and symmetry have better empirical power when the random variables do not have finite expectation, such as in Example 1(b), (c) and in Example 2(b). Figs. 1 and 2 suggest this clearly.

6.2. Simulation experiments in high dimension

In this subsection, we concentrate on the behavior of our proposed tests in high dimension. Although the three distance-based weighted test criteria can be applied to test multivariate homogeneity, symmetry and independence in arbitrary dimension, an interesting phenomenon in high dimension appears in the numerical study. With higher dimension the power gets lower, and even degenerates to zero, especially with characteristic exponent $\alpha = 2$. To illuminate this phenomenon, we compute the empirical size and power in the following three examples, and present the plots of power curves with varying dimensions. The dimensions considered are $p = q = 1, 10 : 10 : 100$, with sample size $n = 40$.

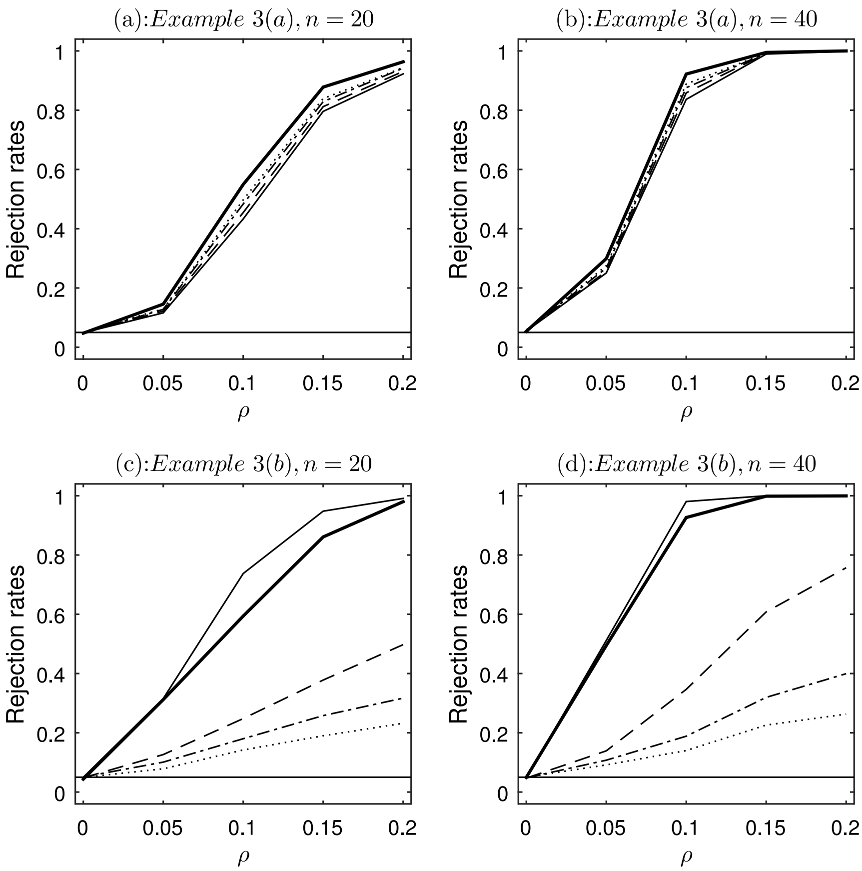

Fig. 3. Simulation results for Example 3. Rejection rates of the independence test against the value of ρ at the 5% significance with dimensions $p = q = 5$ and sample size $n \in \{20, 40\}$. Characteristic exponent $\alpha = 0.5$ (solid line), $\alpha = 1.0$ (dashed line), $\alpha = 1.5$ (dash dot line) and $\alpha = 2.0$ (dotted line), energy statistics (thick line).

Example 4. Homogeneity test: X and Y have iid components respectively following the multivariate normal distributions $\mathcal{N}_p(0, 1)$ and $\mathcal{N}_q(\mu, \sigma^2)$. The notation $\mathcal{N}_p(\mu, \sigma^2)$ stands for a p-dimensional normal distribution with all the components of the mean vector equal to μ, and a diagonal covariance matrix with all diagonal elements equal to σ^2.

Example 5. Symmetry test: The distribution of X is the mixture of $\mathcal{N}_p(0, 1)$ and $\mathcal{N}_p(\mu, 1)$, denoted by $\mathcal{MN}_p(\mu, 1)$. We consider a balanced mixture whereby an observation comes either from $\mathcal{N}_p(0, 1)$ or from $\mathcal{N}_p(\mu, 1)$ with equal probability 0.5.

Example 6. Independence test: The distribution of X is $\mathcal{N}_p(0, 1)$, while Y is taken to have the same dimension as X, with three scenarios. (a) $Y \sim \mathcal{N}_p(0, 1)$, $\text{cov}(X_j, Y_k) = \rho \times \delta_{jk}$ for all $j, k \in \{1, \ldots, p\}$. (b) $Y_{kj} = X_{kj}\varepsilon_{kj}$ for all $k \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, p\}$, where ε_{kj} are independent $\mathcal{N}(0, 1)$ variables and independent of X (the subscript kj denotes the jth component of the kth observation). (c) $Y_{kj} = \ln(X_{kj}^2)$ for all $k \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, p\}$.

Figs. 4–6 display the results of the tests for homogeneity, symmetry and independence, respectively. It can be seen clearly from the figures that the empirical size and power decrease, and often degenerate to zero. This phenomenon is more pronounced with characteristic exponent $\alpha = 2$, but this tendency definitely persists even if $\alpha < 2$, but for larger dimension p. Note that Székely and Rizzo [41] have already noticed this phenomenon when their distance correlation was used to deal with independence test. That is, even under independence, the empirical distance correlation coefficient approaches unity as the dimension tends to infinity when the sample size is fixed. It seems our version of distance-based statistics also have such a problem; i.e., the suggested criteria are biased in high dimension.

In order to further investigate this phenomenon, we repeated the experiment with Example 6, but with sample size $n = 100$. The results displayed in Fig. 7 obviously show that the empirical size and power still degenerates to zero in most cases, despite the increase of the sample size n.

As a final point we focus on the behavior of the sample distances as the dimension tends to infinity but the sample size is held fixed. Consider for instance the homogeneity test and write the sample homogeneity distance in the following

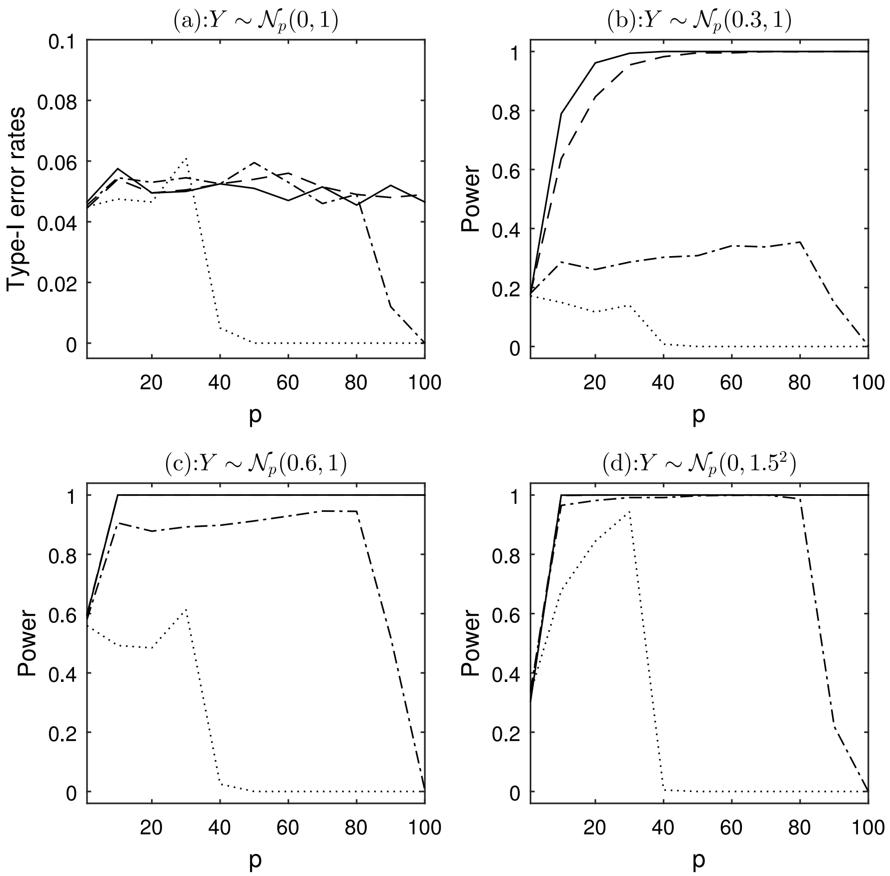

Fig. 4. Simulation results for Example 4. Rejection rates of the homogeneity test against dimension p at the 5% significance with sample size $n = 40$. Characteristic exponent $\alpha = 0.5$ (solid line), $\alpha = 1.0$ (dashed line), $\alpha = 1.5$ (dash dot line) and $\alpha = 2.0$ (dotted line).

form:

$$\begin{aligned}\mathcal{D}^h_{n,\alpha} &= \frac{1}{n^2}\sum_{1\le j,k\le n}\left(e^{-\|X_j-X_k\|^\alpha}+e^{-\|Y_j-Y_k\|^\alpha}-2e^{-\|X_j-Y_k\|^\alpha}\right)\\ &= \frac{1}{n^2}\sum_{j\ne k}\left(e^{-\|X_j-X_k\|^\alpha}+e^{-\|Y_j-Y_k\|^\alpha}-2e^{-\|X_j-Y_k\|^\alpha}\right)+\frac{1}{n^2}\sum_{j=1}^{n}\left(2-2e^{-\|X_j-Y_j\|^\alpha}\right).\end{aligned}$$

Now as an Associate Editor kindly pointed out to us, the quantity $\|X\|^\alpha$ behaves like $\{p+O_p(\sqrt{p})\}^\alpha = p^\alpha\{1+O_p(1/\sqrt{p})\}$, and hence our homogeneity statistic contains sums of terms each of which is of the order $\exp(-p^\alpha)$ or equal to 2. When the sample size n is fixed, we have, as $p \to \infty$.

$$n\mathcal{D}^h_{n,\alpha} \longrightarrow 2.$$

Likewise we find, as $p \to \infty$,

$$nD^s_{n,\alpha} \longrightarrow 1, \quad n\mathbb{D}^I_{n,\alpha} \longrightarrow 1/n.$$

This is clearly the reason why our tests degenerate in high dimension, especially when the characteristic exponent α is equal to two, in which case $\exp(-p^\alpha)$ approaches zero at the fastest possible rate. This observation calls for proper high-dimensional modifications of the test criteria, which is definitely a worthwhile subject for future research.

7. Conclusions

In this paper, we introduce three L_2-type distance measures for distance between distributions, distribution asymmetry, and variable dependence. All three distances are in the Fourier domain, i.e., they are expressed in terms of population characteristic functions, and require no moment condition, but we also provide interpretation in the domain of densities.

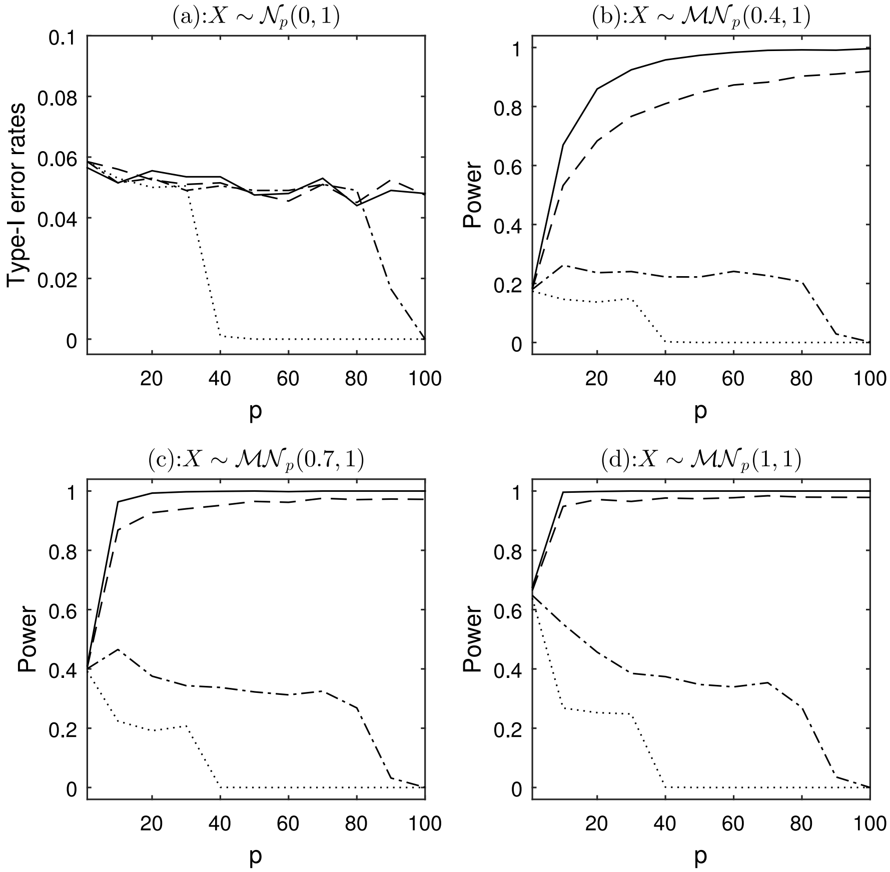

Fig. 5. Simulation results for Example 5. Rejection rates of the symmetry test against dimension p at the 5% significance with sample size $n = 40$. Characteristic exponent $\alpha = 0.5$ (solid line), $\alpha = 1.0$ (dashed line), $\alpha = 1.5$ (dash dot line) and $\alpha = 2.0$ (dotted line).

These measures lead to novel characterizations which generalize corresponding statements encountered earlier in the literature. Empirical counterparts of the characterizations are given yielding generalized versions of already existing test criteria, for which we study the asymptotic null distribution and consistency.

One interesting finding of our Monte Carlo study is that the hitherto emphasis on such test statistics with weight function the density of the normal distribution cannot be supported on the basis of our Monte Carlo results. In fact with this choice of weight function, the CF-test criteria exhibit size distortions as well as low power. Another interesting finding that requires further research concerns testing for homogeneity, symmetry and independence with increasing dimension. It seems the dimensionality is still a very serious issue even with our method. Thus, some more theoretical and numerical investigation on our test criteria should be conducted in the future.

Acknowledgments

The authors wish to sincerely thank the Editor-in-Chief, an Associate Editor and anonymous referees for their constructive criticism that led to an improved version of this work. This research was supported by a grant from The University Grants Council of Hong Kong. Chen's research was supported by the Outstanding Innovative Talents Cultivation Funded Programs 2017 of Renmin University of China. The work of Simos Meintanis was supported by grant number KA: 70/4/7658 of the program 'Kapodistrias' of the special account for research grants (EΛKE) of the National and Kapodistrian University of Athens.

Appendix

Proof of Proposition 1. Clearly one has, for all $t \in \mathbb{R}^p$, $d^h(t) = |\varphi_X(t) - \varphi_Y(t)|^2 \geq 0$, and also notice that $d^h(t)$ may equivalently be written as the left-hand side of (5). Hence for any weight function w which is positive with probability 1,

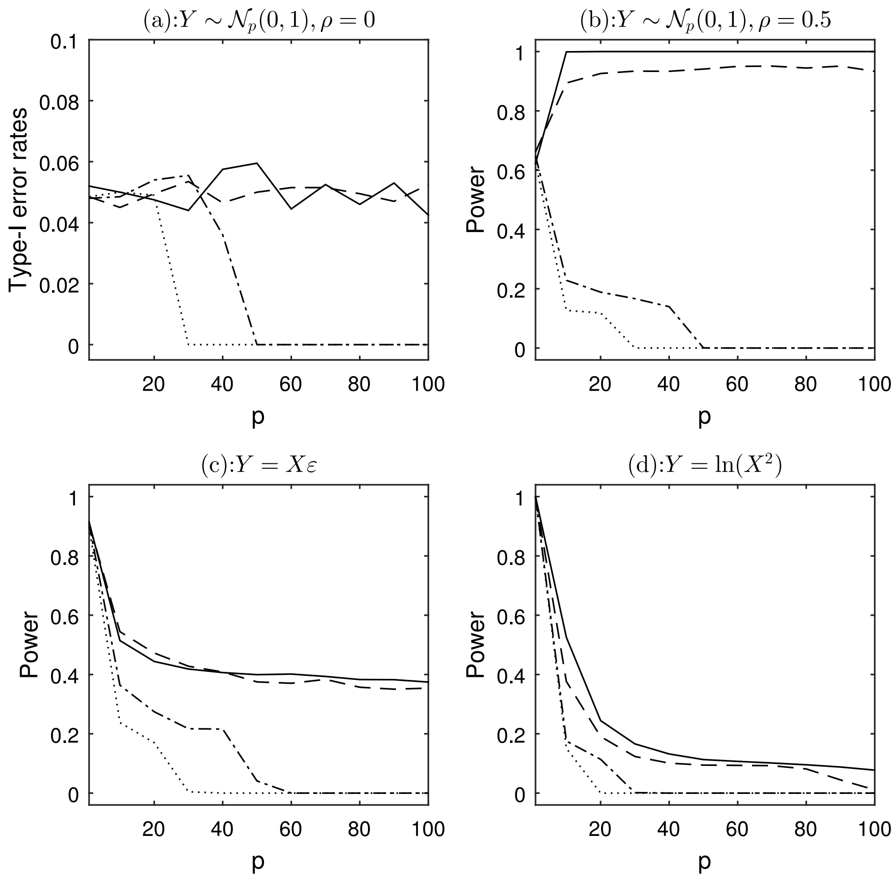

Fig. 6. Simulation results for Example 6. Rejection rates of the independence test against dimension p at the 5% significance with sample size $n = 40$. Characteristic exponent $\alpha = 0.5$ (solid line), $\alpha = 1.0$ (dashed line), $\alpha = 1.5$ (dash dot line) and $\alpha = 2.0$ (dotted line).

we have

$$\mathcal{D}_w = \int_{\mathbb{R}^p} \mathrm{E}\{\cos t^\top(X - X_1) + \cos t^\top(Y - Y_1) - 2\cos t^\top(X - Y)\} w(t)dt \geq 0. \tag{A.1}$$

Replace now the weight function w by the density f_T and invoke Fubini's Theorem to get

$$\begin{aligned} \mathcal{D}_f &= \mathrm{E}\int_{\mathbb{R}^p} \{\cos t^\top(X - X_1) + \cos t^\top(Y - Y_1) - 2\cos t^\top(X - Y)\} f_T(t)dt \\ &= \mathrm{E}[\mathrm{Re}\{\varphi_T(X - X_1) + \varphi_T(Y - Y_1) - 2\varphi_T(X - Y)\}] \geq 0. \end{aligned} \tag{A.2}$$

The uniqueness of the CF entails that $\mathcal{D}_f$ is positive unless (3) holds, and the proof is complete. □

Proof of Proposition 2. Observe that, for all $t \in \mathbb{R}^p$, $d^s(t) = |\varphi_X(t)|^2 - [\mathrm{Re}\{\varphi_X(t)\}]^2 \geq 0$, with equality holding if and only if $[\mathrm{Im}\{\varphi_X(t)\}]^2 = 0$, i.e., only under symmetry, and also notice that $d^s(t)$ may equivalently be written as the left-hand side of (13) up to a factor of 1/2. Hence for any weight function w which is positive with probability 1, we have

$$D_w = \int_{\mathbb{R}^p} \mathrm{E}\{\cos t^\top(X - X_1) - \cos t^\top(X + X_1)\} w(t)dt \geq 0. \tag{A.3}$$

As in (A.2), replace the weight function w by a density f_T and invoke Fubini's theorem to deduce

$$\begin{aligned} D_f &= \mathrm{E}\int_{\mathbb{R}^p} \{\cos t^\top(X - X_1) - \cos t^\top(X + X_1)\} f_T(t)dt \\ &= \mathrm{E}[\mathrm{Re}\{\varphi_T(X - X_1) - \varphi_T(X + X_1)\}] \geq 0. \end{aligned}$$

Clearly D_f is positive unless $d^s \equiv 0$, which only holds under symmetry. Thus the proof is complete. □

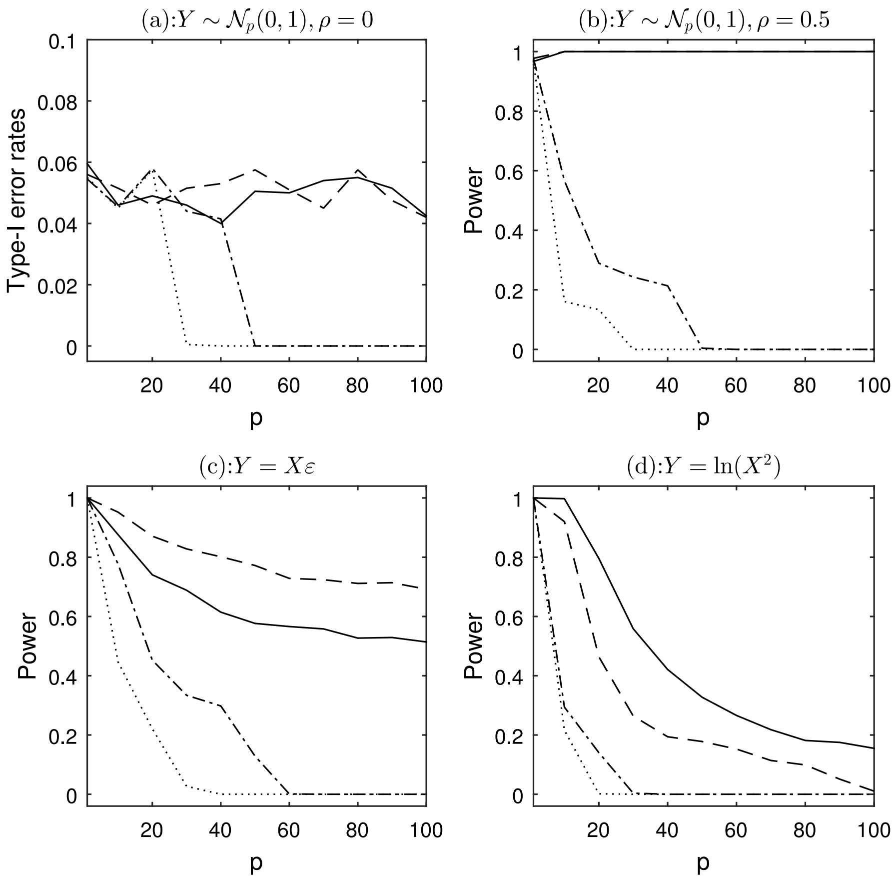

Fig. 7. Simulation results for Example 6. Rejection rates of the independence test against dimension p at the 5% significance with sample size $n = 100$. Characteristic exponent $\alpha = 0.5$ (solid line), $\alpha = 1.0$ (dashed line), $\alpha = 1.5$ (dash dot line) and $\alpha = 2.0$ (dotted line).

Proof of Proposition 3. Clearly, for all $t \in \mathbb{R}^p$,

$$d^I(t) = |\varphi_{X,Y}(t,s) - \varphi_X(t)\varphi_Y(s)|^2 \geq 0, \tag{A.4}$$

with equality holding if and only if X and Y are mutually independent, and also notice that from (18)–(21) it follows that $d^I(t)$ may equivalently be written as

$$\begin{aligned} d^I(t) = {} & \mathrm{E}[\cos\{t^\top(X - X_1) + s^\top(Y - Y_1)\}] + \mathrm{E}[\cos\{t^\top(X - X_1)\}]\mathrm{E}[\cos\{s^\top(Y - Y_1)\}] \\ & - 2\mathrm{E}[\cos\{t^\top(X - X_1) + s^\top(Y - Y_2)\}]. \end{aligned} \tag{A.5}$$

Now observe that $X - X_1$, $Y - Y_1$ and $Y - Y_2$ are symmetrically distributed around zero, and hence if we use the expansion $\cos(a + b) = \cos a \cos b - \sin a \sin b$, in (A.5) the terms containing $\mathrm{E}\{\sin(\cdot)\}$ vanish. Proceed further and assume the factorization $w(t,s) = w_p(t)w_q(s)$, with w_p, w_q satisfying

$$w_p : \mathbb{R}^p \mapsto [0, \infty), \quad w_q : \mathbb{R}^q \mapsto [0, \infty),$$

and in turn replace w_p and w_q by f_{T_p} and f_{T_q}, respectively, in (17). Then the proof is completed by using (A.4) and (A.5) and working analogously as in Propositions 1 and 2. $\square$

Proof of Theorem 1. The proof follows by adapting the arguments in [22,27]. Under $\mathcal{H}_0$, straightforward algebra yields

$$\begin{aligned} \mathcal{D}_{n,w} &= \int_{\mathbb{R}^p} |\varphi_{n,X}(t) - \varphi_{n,Y}(t)|^2 w(t)dt \\ &= \int_{\mathbb{R}^p} \left[\frac{1}{n}\sum_{j=1}^n \{\cos(t^\top X_j) - \cos(t^\top Y_j)\} + \frac{1}{n}\sum_{j=1}^n \{\sin(t^\top X_j) - \sin(t^\top Y_j)\}\right]^2 w(t)dt \end{aligned}$$

$$
\begin{aligned}
&= \int_{\mathbb{R}^p} \left[\frac{1}{n} \sum_{j=1}^{n} [\cos(t^\top X_j) + \sin(t^\top X_j) - \mathrm{Re}\{\varphi(t)\} - \mathrm{Im}\{\varphi(t)\}] \right. \\
&\qquad\qquad \left. - \frac{1}{n} \sum_{j=1}^{n} \left[\cos(t^\top Y_j) + \sin(t^\top Y_j) - \mathrm{Re}\{\varphi(t)\} - \mathrm{Im}\{\varphi(t)\} \right] \right]^2 w(t)dt \\
&= \int_{\mathbb{R}^p} \{Z^h_{n,X}(t) - Z^h_{n,Y}(t)\}^2 w(t)dt,
\end{aligned}
$$

where

$$
Z^h_{n,X}(t) = \frac{1}{n} \sum_{j=1}^{n} \left[\cos(t^\top X_j) + \sin(t^\top X_j) - \mathrm{Re}\{\varphi(t)\} - \mathrm{Im}\{\varphi(t)\} \right] = \frac{1}{n} \sum_{j=1}^{n} h(X_j, t),
$$

$$
Z^h_{n,Y}(t) = \frac{1}{n} \sum_{j=1}^{n} \left[\cos(t^\top Y_j) + \sin(t^\top Y_j) - \mathrm{Re}\{\varphi(t)\} - \mathrm{Im}\{\varphi(t)\} \right] = \frac{1}{n} \sum_{j=1}^{n} h(Y_j, t),
$$

where $h(X_1, t), \ldots, h(X_n, t)$ and $h(Y_1, t), \ldots, h(Y_n, t)$ are centered iid random elements of $\mathfrak{L}^2$, with $\mathrm{E}\|h(X_1, t)\|^2_w = \mathrm{E}\|h(Y_1, t)\|^2_w < \infty$. Note that this process involves the continuous sine and cosine functions and thus is continuous in the Hilbert space. Then, by the Hilbert space Central Limit Theorem (see, e.g., Section 1.8 in [46]), as $n \to \infty$,

$$
\sqrt{n}\, Z^h_{n,X}(t) \rightsquigarrow Z^h_1(t), \quad \sqrt{n}\, Z^h_{n,Y}(t) \rightsquigarrow Z^h_2(t),
$$

where $\{Z^h_1(t) : t \in \mathbb{R}^p\}$ and $\{Z^h_2(t) : t \in \mathbb{R}^p\}$ are independent Gaussian processes with zero mean and identical covariance function given, for all $s, t \in \mathbb{R}^p$, by

$$
\begin{aligned}
\mathrm{E}\{Z^h_1(s)Z^h_1(t)\} = \mathrm{E}\{Z^h_2(s)Z^h_2(t)\} &= \mathrm{Re}\{\varphi(s-t)\} + \mathrm{Im}\{\varphi(s+t)\} - \mathrm{Re}\{\varphi(s)\}\mathrm{Re}\{\varphi(t)\} - \mathrm{Re}\{\varphi(s)\}\mathrm{Im}\{\varphi(t)\} \\
&\quad - \mathrm{Im}\{\varphi(s)\}\mathrm{Re}\{\varphi(t)\} - \mathrm{Im}\{\varphi(s)\}\mathrm{Im}\{\varphi(t)\}
\end{aligned}
$$

Then, as $n \to \infty$, $\sqrt{n}\,\{Z^h_{n,X}(t) - Z^h_{n,Y}(t)\} \rightsquigarrow Z^h_1(t) - Z^h_2(t) = Z^h(t)$, where $\{Z^h(t) : t \in \mathbb{R}^p\}$ is a zero-mean Gaussian process with covariance function given, for all $s, t \in \mathbb{R}^p$, by

$$
\begin{aligned}
\mathrm{E}\{Z^h(s)Z^h(t)\} &= \mathrm{E}\{Z^h_1(s)Z^h_1(t)\} + \mathrm{E}\{Z^h_2(s)Z^h_2(t)\} \\
&= 2\Big[\mathrm{Re}\{\varphi(s-t)\} + \mathrm{Im}\{\varphi(s+t)\} - \mathrm{Re}\{\varphi(s)\}\mathrm{Re}\{\varphi(t)\} - \mathrm{Re}\{\varphi(s)\}\mathrm{Im}\{\varphi(t)\} \\
&\qquad - \mathrm{Im}\{\varphi(s)\}\mathrm{Re}\{\varphi(t)\} - \mathrm{Im}\{\varphi(s)\}\mathrm{Im}\{\varphi(t)\} \Big].
\end{aligned}
$$

Since $n\mathcal{D}_{n,w} = \|\sqrt{n}\,\{Z^h_{n,X}(t) - Z^h_{n,Y}(t)\}\|^2_w$, and by invoking the Continuous Mapping Theorem analogously as in [18] or [27], we find that, as $n \to \infty$, $n\mathcal{D}_{n,w} \rightsquigarrow \|Z^h(t)\|^2_w$. □

Proof of Theorem 2. Observe that, for all t,

$$
|\varphi_{n,X}(t) - \varphi_{n,Y}(t)|^2 \le \left\{ |\varphi_{n,X}(t)| + |\varphi_{n,Y}(t)| \right\}^2 \le 4,
$$

and owing to the consistency of the empirical characteristic function, we find that, as $n \to \infty$ and for any fixed t,

$$
|\varphi_{n,X}(t) - \varphi_{n,Y}(t)|^2 \xrightarrow{wp1} |\varphi_X(t) - \varphi_Y(t)|^2.
$$

So by Lebesgue's Dominated Convergence Theorem, we have $\mathcal{D}_{n,w} \xrightarrow{wp1} \|D^h(t)\|^2_w$, as $n \to \infty$. □

Proof of Theorem 3. Straightforward algebra yields

$$
D_{n,w} = \int_{\mathbb{R}^p} \left\{ \frac{1}{n} \sum_{j=1}^{n} \sin(t^\top X_j) \right\}^2 w(t)dt.
$$

Now note that under $\mathcal{H}_0$, $\mathrm{E}\{\sin(t^\top X)\} = 0$, and hence $\sin(t^\top X_1), \ldots, \sin(t^\top X_n)$ are centered iid random elements of $\mathfrak{L}^2$, with $\mathrm{E}\|\sin(t^\top X_1)\|^2_w < \infty$. The proof can be completed similarly to the proof of Theorem 1 without additional difficulty. □

Proof of Theorem 4. Note that, for all t,

$$
\left\{ \frac{1}{n} \sum_{j=1}^{n} \sin(t^\top X_j) \right\}^2 \le 1,
$$

and hence by the Strong Law of Large Numbers we have, for fixed t, as $n \to \infty$,

$$\frac{1}{n}\sum_{j=1}^{n}\sin(t^\top X_j) \xrightarrow{wp1} \mathrm{E}\{\sin(t^\top X)\}.$$

The proof can be completed similarly to the proof of Theorem 2. □

Proof of Theorem 5. Straightforward algebra yields

$$\begin{aligned}
\mathbb{D}_{n,w} &= \int_{\mathbb{R}^{p+q}}\Bigg[\frac{1}{n}\sum_{j=1}^{n}\cos(s^\top Y_j)\left\{\cos(t^\top X_j)+\sin(t^\top X_j)\right\}+\frac{1}{n}\sum_{j=1}^{n}\sin(s^\top Y_j)\{\cos(t^\top X_j)-\sin(t^\top X_j)\}\\
&\quad -\frac{1}{n^2}\sum_{j=1}^{n}\cos(s^\top Y_j)\sum_{k=1}^{n}\{\cos(t^\top X_k)+\sin(t^\top X_k)\}-\frac{1}{n^2}\sum_{j=1}^{n}\sin(s^\top Y_j)\sum_{k=1}^{n}\{\cos(t^\top X_k)-\sin(t^\top X_k)\}\Bigg]^2 w(t,s)dtds\\
&= \int_{\mathbb{R}^{p+q}}\Bigg[\frac{1}{n}\sum_{j=1}^{n}[\cos(s^\top Y_j)-\mathrm{Re}\{\varphi_Y(s)\}]\times\{\cos(t^\top X_j)+\sin(t^\top X_j)\}\\
&\quad +\frac{1}{n}\sum_{j=1}^{n}[\sin(s^\top Y_j)-\mathrm{Im}\{\varphi_Y(s)\}]\times\{\cos(t^\top X_j)-\sin(t^\top X_j)\}\\
&\quad -\frac{1}{n^2}\sum_{j=1}^{n}[\cos(s^\top Y_j)-\mathrm{Re}\{\varphi_Y(s)\}]\sum_{k=1}^{n}\{\cos(t^\top X_k)+\sin(t^\top X_k)\}\\
&\quad -\frac{1}{n^2}\sum_{j=1}^{n}[\sin(s^\top Y_j)-\mathrm{Im}\{\varphi_Y(s)\}]\sum_{k=1}^{n}\{\cos(t^\top X_k)-\sin(t^\top X_k)\}\Bigg]^2 w(t,s)dtds\\
&= \int_{\mathbb{R}^{p+q}}\Bigg[\frac{1}{n}\sum_{j=1}^{n}[\cos(s^\top Y_j)-\mathrm{Re}\{\varphi_Y(s)\}][\cos(t^\top X_j)+\sin(t^\top X_j)-\mathrm{Re}\{\varphi_X(t)\}-\mathrm{Im}\{\varphi_X(t)\}]\\
&\quad +\frac{1}{n}\sum_{j=1}^{n}[\sin(s^\top Y_j)-\mathrm{Im}\{\varphi_Y(s)\}][\cos(t^\top X_j)-\sin(t^\top X_j)-\mathrm{Re}\{\varphi_X(t)\}+\mathrm{Im}\{\varphi_X(t)\}]\Bigg]^2 w(t,s)dtds+o_p(1/n).
\end{aligned}$$

The rest of the proof is similar to the proof of Theorem 1. □

Proof of Theorem 6. The proof can be carried out similarly to the proof of Theorem 2. □

References

[1] V. Alba Fernández, M.D. Jiméne Gamero, J. Muñoz García, A test for the two-sample problem based on empirical characteristic functions, Comput. Statist. Data Anal. 52 (2008) 3730–3748.
[2] B. Aslan, G. Zech, New test for the multivariate two-sample problem based on the concept of minimum energy, J. Stat. Comput. Simul. 75 (2005) 109–119.
[3] L. Baringhaus, Fibonacci numbers, Lucas numbers and integrals of certain Gaussian processes, Proc. Amer. Math. Soc. 124 (1996) 3875–3884.
[4] L. Baringhaus, C. Franz, On a new multivariate two-sample test, J. Multivariate Anal. 88 (2004) 190–206.
[5] M. Bilodeau, A. Guetso Nangue, Tests of mutual or serial independence of random vectors with applications, J. Mach. Learn. Res. 18 (2017) Paper No. 74, 40.
[6] M. Bilodeau, P. Lafaye de Micheaux, A multivariate empirical characteristic function test of independence with normal marginals, J. Multivariate Anal. 95 (2005) 345–369.
[7] A. Buja, B.F. Logan, J.A. Reeds, L.A. Shepp, Inequalities and positive-definite functions arising from a problem in multidimensional scaling, Ann. Statist. 22 (1994) 406–438.
[8] S. Csörgő, P. Deheuvels, D. Mason, Kernel estimates of the tail index of a distribution, Ann. Statist. 13 (1985) 1050–1077.
[9] A. Dematteo, S. Clémençon, On tail index estimation based on multivariate data, J. Nonparametr. Stat. 28 (2016) 152–176.
[10] Z. Du, Testing for serial independence of panel errors, Comput. Statist. Data Anal. 76 (2014) 248–261.
[11] Y. Fan, P. Lafaye de Micheaux, S. Penev, D. Salopek, Multivariate nonparametric test of independence, J. Multivariate Anal. 153 (2017) 189–210.
[12] A. Feuerverger, A consistent test for bivariate dependence, Int. Stat. Rev. 61 (1993) 419–433.
[13] A. Feuerverger, R.A. Mureika, The empirical characteristic function and its applications, Ann. Statist. 5 (1977) 88–97.
[14] S. Ghosh, J. Beran, Two-sample T_3 plot: A graphical comparison of two distributions, J. Comput. Graph. Statist. 9 (2000) 167–179.
[15] S. Ghosh, F.H. Ruymgaart, Applications of empirical characteristic functions in some multivariate problems, Canad. J. Statist. 20 (1992) 429–440.
[16] P. Good, Permutation, Parametric and Bootstrap Tests of Hypotheses, third ed., Springer, New York, 2005.
[17] A. Gretton, K.M. Borgwardt, M.J. Rasch, B. Schölkopf, A. Smola, A kernel two-sample test, J. Mach. Learn. Res. 13 (2012) 723–773.
[18] A.K. Gupta, N. Henze, B. Klar, Testing for affine equivalence of elliptically symmetric distributions, J. Multivariate Anal. 88 (2004) 222–242.
[19] N. Henze, Invariant tests for multivariate normality: A critical review, Statist. Pap. 43 (2002) 467–506.

[20] N. Henze, B. Klar, S.G. Meintanis, Invariant tests for symmetry about an unspecified point based on the empirical characteristic function, J. Multivariate Anal. 87 (2003) 275–297.

[21] Z. Hlávka, M. Hušková, S.G. Meintanis, Tests for independence in non-parametric heteroscedastic regression models, J. Multivariate Anal. 102 (2011) 816–827.

[22] M. Hušková, S.G. Meintanis, Tests for the multivariate k-sample problem based on the empirical characteristic function, J. Nonparametr. Stat. 20 (2008) 263–277.

[23] A. Kankainen, N.G. Ushakov, A consistent modification of a test for independence based on the empirical characteristic function, J. Math. Sci. (N. Y.) 89 (1996) 1486–1494, Stability Problems for Stochastic Models, Part 1 (Moscow, 1996).

[24] T.J. Kozubowski, K. Podgó, I. Rychlik, Multivariate generalized Laplace distribution and related random fields, J. Multivariate Anal. 113 (2013) 59–72.

[25] E.L. Lehmann, J.P. Romano, Testing Statistical Hypotheses, third ed., Springer, New York, 2005.

[26] D.S. Matteson, N.A. James, A nonparametric approach for multiple change point analysis of multivariate data, J. Amer. Statist. Assoc. 109 (2014) 334–345.

[27] S.G. Meintanis, Permutation tests for homogeneity based on the empirical characteristic function, J. Nonparametr. Stat. 17 (2005) 583–592.

[28] S.G. Meintanis, G. Iliopoulos, Fourier methods for testing multivariate independence, Comput. Statist. Data Anal. 52 (2008) 1884–1895.

[29] S. Mittnik, S.T. Rachev, Tail estimation of the stable index α, Appl. Math. Lett. 9 (1996) 53–56.

[30] J. Ngatchou-Wandji, Testing for symmetry in multivariate distributions, Stat. Methodol. 6 (2009) 230–250.

[31] Y. Nikitin, Asymptotic Efficiency of Nonparametric Tests, Cambridge University Press, Cambridge, 1995.

[32] J.P. Nolan, Multivariate elliptically contoured stable distributions: Theory and estimation, Comput. Statist. 28 (2013) 2067–2089.

[33] V. Paulauskas, A new estimator for a tail index, Acta Appl. Math. 79 (2003) 55–67.

[34] J.-F. Quessy, F. Éthier, Cramér–von Mises and characteristic function tests for the two and k-sample problems with dependent data, Comput. Statist. Data Anal. 56 (2012) 2097–2111.

[35] J.-F. Quessy, F. Éthier, Two bootstrap strategies for a k-problem up to location-scale with dependent samples, J. Probab. Stat. (2014) 523139, 12.

[36] M.L. Rizzo, G.J. Székely, DISCO analysis: A nonparametric extension of analysis of variance, Ann. Appl. Stat. 4 (2010) 1034–1055.

[37] M. Rosenblatt, A quadratic measure of deviation of two-dimensional density estimates and a test of independence, Ann. Statist. 3 (1975) 1–14.

[38] L. Su, H. White, A consistent characteristic function-based test for conditional independence, J. Econometrics 141 (2007) 807–834.

[39] G.J. Székely, T.F. Móri, A characteristic measure of asymmetry and its application for testing diagonal symmetry, Comm. Statist. Theory Methods 30 (2001) 1633–1639.

[40] G.J. Székely, M.L. Rizzo, Hierarchical clustering via joint between-within distances: Extending Ward's minimum variance method, J. Classification 22 (2005) 151–183.

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

[42] G.J. Székely, M.L. Rizzo, Energy statistics: A class of statistics based on distances, J. Statist. Plann. Inference 143 (2013) 1249–1272.

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

[44] G.J. Szekely, P.K. Sen, Characterization of diagonal symmetry: Location unknown, and a test based on allied U-processes, J. Statist. Plann. Inference 102 (2002) 349–358.

[45] C. Tenreiro, On the choice of the smoothing parameter for the BHEP goodness-of-fit test, Comput. Statist. Data Anal. 53 (2009) 1038–1053.

[46] A.W. van der Vaart, J.A. Wellner, Weak Convergence and Empirical Processes, With Applications to Statistics, Springer, New York, 1996.

[47] L.-X. Zhu, G. Neuhaus, Nonparametric Monte Carlo tests for multivariate distributions, Biometrika 87 (2000) 919–928.

[48] V.M. Zolotarev, 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, 1981, pp. 283–305.