

# Using space filling curves to compare two multivariate distributions with distribution-free tests

Cédric Heuchenne <sup>a,b</sup> and Gilles Mordant <sup>b,c</sup>

<sup>a</sup>HEC Liège, Management School of the University of Liège, ULiège, Liège, Belgium; <sup>b</sup>ISBA, Institute of Statistics, Biostatistics and Actuarial Sciences, Catholic University of Louvain, Louvain-la-Neuve, Belgium; <sup>c</sup>Institut für Mathematische Stochastik, Georg-August-Universität Göttingen, Göttingen

## ARTICLE HISTORY

Compiled June 7, 2022

## ABSTRACT

This paper proposes new tests to compare two multivariate probability distributions. Since basic ranks do not canonically exist in  $\mathbb{R}^d$ , it is impossible to have a natural multivariate generalization of rank-based tests such as the two-sample Kolmogorov–Smirnov test. We thus rely on recent measure transportation theory to transform this  $d$ –dimensional problem into one-dimensional classical tests by using space filling curves. To foster lower computation time, we develop distribution-free tests so as to avoid computing critical values for any particular problem. We demonstrate their theoretical validity and compare them to each other and to the existing distribution-free techniques via extensive simulations. We show that they are computationally efficient and most of the time outperform existing techniques when measuring the corresponding power functions. Finally, we apply the proposed tests to a data set about stars’ luminosity and temperature.

## AMS CLASSIFICATION

62G10, 62H15

## KEYWORDS

Multivariate two-sample test, Space filling curves, Distribution-free tests, Kolmogorov–Smirnov test.

## 1. Introduction

Comparing two samples drawn from multivariate distributions is a natural problem arising in statistics and machine learning. The motivation may come from various fields. For example, in physics, one may want to compare real-world observations with those from a simulated model or assess the difference between two sets of multivariate measurements on different instruments. In medicine, one may wish to perform a test to detect any difference between a control and a test group. In machine learning applications, comparing probability distributions is at the heart of generative adversarial networks. It is also required to assess fairness of machine learning procedure, e.g., to ensure that the distribution of outcomes is agnostic to the sex of the individuals.

Among the theoretical properties one may require from such a test, distribution-freeness and consistency are certainly the most important when the multivariate distributions are unknown and assumed continuous. With a particular focus on practice, one could also add fast computability requirements. Given these goals statements, we start by shortly reviewing the literature about multidimensional testing. For a more comprehensive review, we refer to [1].

A strand of the literature managed to extend the Kolmogorov–Smirnov test to higher dimensions. [2] gave the impulse by proposing to “compare the cumulative distribution functions of the two samples with all possible orderings”. Because of the computational burden, a lighter version of the latter was developed by [3], reducing the numbers of orderings to consider. More recently, [4] proposed a multivariate equivalent of the Kolmogorov–Smirnov test based on the Rosenblatt coupling ([5]). As the corresponding statistic is not-relabelling invariant, i.e., depends on the order in which variables appear in the coupling, they propose to test all possible ways of constructing the Rosenblatt coupling. They then use the Bonferroni inequality to obtain an upper bound on the global significance level. It is worth noting that one requires simulations to obtain critical values for the test and the latter still depend on the dependency structures of the data. None of these tests is distribution-free.

Graph-based statistical tests were also developed, such as [6], where a minimum spanning tree is built and edges connecting two vertices belonging to different groups are deleted. The remaining trees are called runs and rejection occurs when there are too few of them. Other tests of the same vein are [7], [8] (they both rely on  $k$ -nearest neighbours) or more recently [9]. None of them is distribution free. This implies that  $p$ -values are estimated by Fisher permutations, i.e., the procedure by which the distribution under the null is estimated from a random reassignment of labels to the data points. [10] proposed a distribution-free test based on non-bipartite matching graph, i.e., a graph linking observations two by two, based on shortest distances. The statistics relies on the number of pairs formed by one element of each sample. More recently, [11] proposed a distribution-free test based on a minimum Hamiltonian path. A Hamiltonian path is a path connecting points where each vertex has at most two edges. The minimization is performed over the sum of lengths of edges. Their statistic is based on the number of runs and is somewhat similar to [6], the only difference being the considered type of graph.

Another important part of the literature is linked to data depth. The latter concept aims at proposing a generalization of ranks for dimensions larger than one. [12] use it to test the equality of distributions and their methodology works for all depth measures, such as Mahalanobis’ depth, simplicial depth, Tukey’s depth, to name but a few. They propose various techniques but only their Mann-Whitney test based on depths instead of ranks is distribution-free. It is also known (see [13]) that this type of test has no power against location alternatives. Finally, computing these depths turns out to be computationally costly.

Finally, tests based on energy statistics, i.e., functions of interpoint distances were also developed by Székely in a series of papers (see [14] for an explanation of the ideas). Those tests have thorough theoretical foundations but the results are only asymptotic.

The aim of this paper is twofold. First, we fill a gap in the literature regarding the use of filling curves in statistics. To the best of our knowledge, space filling curves have never been used in statistics, except to compute a Wasserstein distance for approximate bayesian computation in a very recent work of [15]. They also noticed — and used — the very convenient computational aspects of the Hilbert space filling curve. However, the fact that space-filling curves are measure preserving transformations is not fully

exploited in their work. Second, we propose a methodology for bringing back the two-sample problem of comparing multivariate distributions to a one-dimensional one in a computationally effective way.

The idea of using space filling curves in statistics arises from their definition and properties that imply the possibility to transport any measure living on a finite dimensional space onto one living on the real line. Further, they turn out to be very useful in many other disciplines such as, e.g., parallel computing, sorting of databases or quasi-Monte Carlo methods. As the list is long and not relevant to the present paper, we refer the reader to the introduction of [16]. In the sequel, we set the notations used throughout the paper. Section 2 formalises the idea of transporting density mass relying on space filling curves. Section 3 establishes that the ranks are maximal ancillary, guaranteeing that tests based on the ranks will be efficient in a sense made precise later on. Section 4 proposes an analysis of what it implies to perform a 1-d test on the transformed dataset. Section 5 proposes some simulations while Section 6 analyses a real data set on stars' luminosity and temperature.

### 1.1. Notations

Let us first define  $\mathcal{S}_n$ , the set of permutations of  $\{1, \dots, n\}$ . In the sequel, we will consider two i.i.d. multivariate samples<sup>1</sup>. We will denote them by  $\{X_i^d\}_{i=1}^{N_X}$  and  $\{Y_i^d\}_{i=1}^{N_Y}$ , respectively, with  $X_i^d \in \mathbb{R}^d$ , for each  $i \in \{1, \dots, N_X\}$ , and  $Y_i^d \in \mathbb{R}^d$ , for each  $i \in \{1, \dots, N_Y\}$ . We define  $N = N_X + N_Y$ . We will denote the probability measure, distribution function and density associated with each random variable  $X^d$  (resp.  $Y^d$ ) by  $P_{X^d}$  (resp.  $P_{Y^d}$ ),  $F_d$  (resp.  $G_d$ ) and  $f_d$  (resp.  $g_d$ ). For a given cumulative distribution function  $F(\cdot)$ ,  $\hat{F}(\cdot)$  will denote the empirical cumulative distribution function. We use  $\lambda_i$  to denote the  $i$ -dimensional Lebesgue measure.

## 2. Space filling curve and general idea

In a  $d$ -dimensional space, with  $d \geq 2$ , there exists no canonical ordering. Bringing back problems to the one-dimensional case could thus benefit from the nice properties of rank-based statistics. Space filling curves, defined below, seem nice candidates for performing such a task. Also, following this line of thought leverages the isomorphism between any Borel-measure space  $(\mathbb{R}^d, \text{for instance})$  and the  $[0, 1]$  interval endowed with the Lebesgue measure: this explains why the notion of dimension in this context does not matter in probability theory; our approach becomes naturally similar to using ranks in one dimension.

First, we recall the definition of space filling-curves and a few properties on which the theoretical developments of this paper are built.

**Definition 2.1.** [17] A space filling curve is a surjective mapping  $\varphi : [0, 1] \mapsto [0, 1]^d$  such that for all  $x \in [0, 1]^d$ , one can quickly compute a  $t \in [0, 1]$  such that  $\varphi(t) = x$ .

In what follows, we will restrict our attention to Hilbert space filling curves, see [18,

---

<sup>1</sup>Even though we retain the assumption of i.i.d. data in the sequel, it should be noted that all of the following results hold for exchangeable data, i.e., data for which the indexing is immaterial. More precisely, the joint distribution of  $\{X_{\sigma i}^d\}_{i=1}^{N_X}$  should be invariant with respect to the permutation  $\sigma \in \mathcal{S}_n$ . This property ensures that the ranks obtained via our procedure are uniformly distributed on the set of permutations, the main requirement for the validity of rank-based statistics.

Section 2.3, page 18], even though other curves could be considered. The reason for this choice is twofold.

First, the Hilbert curve is often said to preserve the best locality, i.e., two points close to each other on the curve will be close to each other in the space. Indeed, the Hilbert space filling curve is Hölder continuous with exponent  $1/d$  (see [18]).

Second, the possibility to compute the preimage of a point in a fixed number of operations makes it an appealing candidate.

**Definition 2.2.** A pseudo inverse of the Hilbert curve is a bijective function  $\bar{\varphi}^{-1}$  such that

$$\lambda_1 \{x : \bar{\varphi}^{-1}(\varphi(x)) = x\} = 1.$$

The bar in the above notation is used to show that  $\bar{\varphi}^{-1}$  is not an inverse, as the filling curve is not bijective.

**Property 2.3** (Bi-measure preserving property). *The filling space Hilbert curve maps an interval to a region with an area equal to the length of the interval. In general we have for  $d$ -dimensional Hilbert curves that for any Borel set  $A \subset [0, 1]$ :*

$$\lambda_1(A) = \lambda_d(\varphi(A)).$$

Because of the inherent measurability of both  $\varphi$  and  $\bar{\varphi}^{-1}$ , we have that the application  $X^1 := \bar{\varphi}^{-1}(X^d)$  is a random variable. Indeed, a composition of two measurable applications is indeed measurable and therefore  $X^1$  is a measurable function from a probability space to  $[0, 1]$  (or even  $\mathbb{R}$  theoretically).

With this result at hand, we propose the following procedure to compare two distributions in practice.

- Compute the minimum and the range of each component of the observations independently of their group and for each component, subtract the corresponding minimum and divide by the corresponding range so as to include all the data in the  $[0, 1]^d$  hypercube. Alternatively, one can apply any function  $f : \mathbb{R} \rightarrow (0, 1)$  to each component.
- Compute the preimage of each point using  $\bar{\varphi}^{-1}$ . An algorithm is provided in Appendix A.
- Perform a distributions equality test.

The rescaling phase is theoretically not required, given that  $\mathbb{R}^d$  and  $\mathbb{R}$  are equinumerous and there exists an isomorphism between them. This step is only performed for practical, computational reasons. From now on, we will consider that all observations have been rescaled, which explains the use of the support  $[0, 1]^d$  whenever needed. The possibility to rescale will always be guaranteed in practice.

Finally, note that the refinement of the space filling curve we propose to use is always the finest possible, so that one always tries to be as close as possible to the limiting, yet theoretical, curve. Therefore, the power of the test should not depend on any arbitrary choice and be comparable with other tests performed on the same machine.

As stated above, the bi-measurability is crucial for our testing purposes. Together with the following Theorem, It guarantees the consistency of the proposed test.

**Theorem 2.4.** (*Consistency of the proposed procedure*) Consider two i.i.d. samples  $\{Y_i^d\}_{i=1}^n, \{X_j^d\}_{j=1}^m$  from random variables  $Y^d$  and  $X^d$ , respectively. Further consider a test  $\phi$  for two  $\mathbb{R}$ -valued samples  $\phi : (\{Y_i\}_{i=1}^n, \{X_j\}_{j=1}^m) \rightarrow \{0, 1\}$ , that is measurable with respect to the Lebesgue measure and consistent i.e., if  $\exists B \in \mathcal{B}(\mathbb{R}) : P_{X^1}(B) \neq P_{Y^1}(B) \Rightarrow \lim_{n,m \rightarrow \infty} \phi(\{Y_i\}, \{X_j\}) = 1$  a.s. Then, if  $\exists B' \in \mathcal{B}(\mathbb{R}^d) : P_{X^d}(B') \neq P_{Y^d}(B')$ , we have

$$\lim_{n,m \rightarrow \infty} \phi\left(\{\bar{\varphi}^{-1}(Y_i^d)\}, \{\bar{\varphi}^{-1}(X_j^d)\}\right) = 1 \text{ a.s.}$$

**Proof.** Exploiting bi-measurability of the space filling curve with respect to Lebesgue measure, the claim follows. Indeed, we have found a measurable mapping to our decision space that will asymptotically map to “rejection” with probability one, whenever the two measures differ on a Borel set in  $\mathbb{R}^d$ .  $\square$

**Remark 1. Computational complexity.** Among the various techniques described in Section 1, our procedure is computationally very efficient. Indeed, the ‘transport step’ only requires  $\mathcal{O}(N)$  operations where the constant depends on the refinement level of the curve. Further, the number of refinement steps required remains very limited as can be seen from the algorithm presented in Appendix A. In this recursive algorithm, the number of steps required to attain a fixed precision is a linear function of the logarithm of the precision.

The second step of the procedure depends on the rank-based test chosen and usually requires a sorting step that is  $\mathcal{O}(N \log(N))$ . In the case of the Kolmogorov–Smirnov test, there even exists an algorithm that requires  $\mathcal{O}(N)$  operations (see [19]).

Overall, the complexity of the proposed procedure is  $\mathcal{O}(N \log(N))$ .

### 3. Maximal ancillarity result

Implicitly, the procedure gives rise to an ordering in  $\mathbb{R}^d$  and studying the properties of the latter is necessary to understand whether the tests based on such ranks are efficient. In this regard, establishing that the ranks are maximal ancillary is an important step. The structure of this section owes a lot to the presentation in [20] to which we refer for an introduction to the field of multivariate, maximal ancillary ranks. Before stating the main result, we need to introduce some elements. Let us use the notation  $X_{(\cdot)}$  for the vector of order statistics and  $F_{SFC}^{(n)}$  for the function mapping each sample point to its “space-filling” rank, i.e., the rank on  $[0, 1]$  of  $\bar{\varphi}^{-1}(X_i)$  for a given sample  $\{X_i\}_{i=1}^n =: X^{(n)}$ . We can thus define

$$F_{SFC}^{(n)}(X^{(n)}) := (F_{SFC}^{(n)}(X_1^{(n)}), \dots, F_{SFC}^{(n)}(X_n^{(n)})).$$

Finally, let us use the notation  $\mathcal{P}^d$  for the set of atomeless measures over  $[0, 1]^d$ . Recall that the choice of the latter set is purely induced by practice as, in theory, the Hilbert space-filling curve could be constructed as a surjective mapping from  $\mathbb{R}$  to  $\mathbb{R}^d$ .

**Theorem 3.1.** Let  $X^{(n)}$  be i.i.d. with distribution  $P \in \mathcal{P}^d$ . We have that

- (1)  $X_{(\cdot)}$  is minimum and complete and hence sufficient for  $\mathcal{P}^d$ .
- (2)  $F_{SFC}^{(n)}(X^{(n)})$  is uniformly distributed over the  $n!$  permutations of the set  $\{1, \dots, n\}$ .

- (3) For all  $P \in \mathcal{P}^d$ ,  $F_{SFC}^{(n)}(X^{(n)})$  and  $X_{(\cdot)}$  are mutually independent.  
(4)  $F_{SFC}^{(n)}(X^{(n)})$  is strongly  $\mathcal{P}^{d(n)}$ -essential maximal ancillary.

**Proof.** For part *i*), notice that the ordering only induces a permutation of the indices similarly to what holds for univariate ranks. The likelihoods of  $X_{(\cdot)}$  and  $\{X_i\}_{i=1}^n$  are thus trivially the same.

Regarding part *ii*), the fact that the ranks are uniformly distributed over the  $n!$  permutations of the set  $\{1, \dots, n\}$  is inherited from the very same property holding for the univariate ranks of the transported sample,  $\{\bar{\varphi}^{-1}(X_i)\}_{i=1}^n$ .

Part *iii*): Owing to *i*) and *ii*), this is an immediate consequence of Basu's Second Theorem, see [20, Appendix E.].

Finally, part *iv*) of the above claim follows from the injectivity of

$$\mathbb{R}^{nd} \setminus N \ni z \mapsto \left( z_{(\cdot)}, F_{SFC}^{(n)}(z^{(n)}) \right)$$

using the same reasoning as in [20, Proposition 2.5], where  $N$  is a null set. This null set in  $[0, 1]^{nd}$  corresponds to the case where there would exist  $(i, j) : z_i = z_j$ , which is Lebesgue negligible.  $\square$

## 4. Analysis of the procedure for various tests

### 4.1. Mann–Whitney–Wilcoxon test

The Mann–Whitney–Wilcoxon test relies on the possibility to order observations and give them a rank. Using  $X_i^1 = \bar{\varphi}^{-1}(X_i^d)$ ,  $i \in \{1, \dots, N_X\}$ ,  $Y_i^1 = \bar{\varphi}^{-1}(Y_i^d)$ ,  $i \in \{1, \dots, N_Y\}$ , we propose a test relying on the Mann–Whitney statistic:

$$U_{N_X, N_Y} = \sum_{i=1}^{N_X} \sum_{j=1}^{N_Y} \mathbb{1}(X_i^1 \leq Y_j^1)$$

The critical values are tabulated for small sample sizes and one can recourse to the central limit theorem whenever the sample sizes are sufficiently large. As mentioned above, when  $P_{X^d} \equiv P_{Y^d}$ , we obviously have  $P_{X^1} \equiv P_{Y^1}$ , which justifies the use of this statistic and its usual distribution.

### 4.2. Multivariate Kolmogorov–Smirnov test

Using the same argument, we can define a Kolmogorov–Smirnov type statistic

$$\sup_{z \in [0, 1]} |\hat{G}(z) - \hat{F}(z)| \sim D_{N_X, N_Y},$$

where

$$\hat{F}(x) := \frac{1}{N_X} \sum_{i=1}^{N_X} \mathbb{1}[X_i^1 \leq x], \quad \hat{G}(x) := \frac{1}{N_Y} \sum_{i=1}^{N_Y} \mathbb{1}[Y_i^1 \leq x]$$

and  $D_{N_X, N_Y}$  is the usual distribution of the statistic of the Kolmogorov–Smirnov test. The null hypothesis is rejected as soon as

$$D_{N_X, N_Y} > c(\alpha) \sqrt{\frac{N}{N_X N_Y}}$$

where  $c(\alpha)$  is a real number depending on the level of the test  $\alpha$ .

#### 4.2.1. Further considerations

We now consider what it implies for the proposed test to reject. Naturally, it means that  $\|\hat{F}(x) - \hat{G}(x)\|_\infty$  is larger than a certain threshold depending the chosen type I error. Considering fixed alternatives, the power of the test will depend on  $\|F(x) - G(x)\|_\infty$ . The latter can be rewritten for  $A^* = [0, y]$  and  $\sup_y$ ,

$$\begin{aligned} & \sup_y \left| \int_{A^*} d(\bar{\varphi}^{-1} \# P_{X^d}) - \int_{A^*} d(\bar{\varphi}^{-1} \# P_{Y^d}) \right| \\ &= \sup_y \left| \int_{\varphi(A^*)} f_d(\mathbf{x}) d\mathbf{x} - \int_{\varphi(A^*)} g_d(\mathbf{x}) d\mathbf{x} \right| \quad (\varphi \text{ is a measurable function}) \\ &= \sup_y \left| \int_{\varphi(A^*)} (f_d(\mathbf{x}) - g_d(\mathbf{x})) d\mathbf{x} \right| \end{aligned}$$

Consequently, the power of the test directly depends on the distributions difference in the  $d$ -dimensional space.

**Remark 3.1** As an extension of the Kolmogorov–Smirnov test, we will also apply the procedure of Section 2 to Kuiper’s test:

$$V_n = \sup_x [\hat{F}(x) - \hat{G}(x)] - \inf_x [\hat{F}(x) - \hat{G}(x)]$$

Even though it does not always achieve a better power than the Kolmogorov–Smirnov test, it does when  $\hat{F}(x)$  crosses  $\hat{G}(x)$ . For this reason, it should, in most cases, show a good performance in our procedure since in some respect, the space filling curve picks data in a way that does not suggest the superiority of a distribution with respect to the other on the whole segment  $[0, 1]$ .

#### 4.3. Wald–Wolfowitz test

The Wald–Wolfowitz ([21]) test counts the number of runs when the samples are jointly ordered. A run is defined as a block of elements from the same sample. Under the null hypothesis, this number of runs should be large in such a way that for sample sizes  $m$  and  $n$ :

$$\mathbb{P}(R = 2r) = 2 \frac{\binom{m-1}{r-1} \binom{n-1}{r-1}}{\binom{m+n}{m}}$$

and

$$\mathbb{P}(R = 2r + 1) = \frac{\binom{m-1}{r} \binom{n-1}{r-1} + \binom{m-1}{r-1} \binom{n-1}{r}}{\binom{m+n}{m}}.$$

Applying a Wald and Wolfowitz [21] test to the preimage of points obtained by the space filling curve yields another test of the Friedman and Rafsky ([6]) type. In fact, instead of computing exactly a minimum spanning tree as in the latter test, it is approximated by a graph built by means of a space filling curve. Figure 1 displays an example of runs under the null and alternative hypotheses of the resulting Wald and Wolfowitz test performed on preimages.

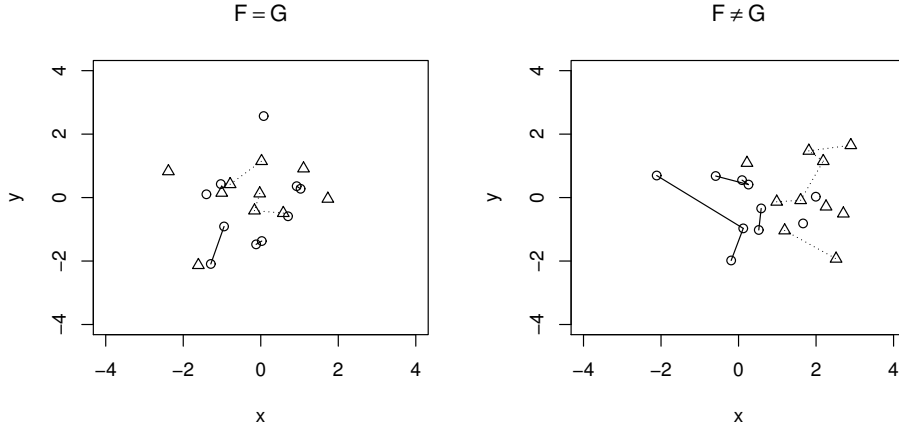


**Figure 1.** Runs for the Wald and Wolfowitz [21] test using space filling curves for two bivariate Gaussian distributions. Both distributions are the same on the left hand side and only differ through their means with a difference of means  $\Delta\mu = (1.5, 0)'$  on the right hand side. One type of point is used per sample.

Our procedure has a great advantage compared to the Friedman and Rafsky [6] test: it is distribution-free. Indeed, the test proposed by Friedman and Rafsky is only distribution-free conditionally on the sample. This implies that p-values have to be approximated for each statistical test one aims at performing.

In some respect, the combination of space filling curves and the Wald and Wolfowitz test can be seen as procedure analogous to [11]. Indeed, their test relying on the shortest Hamiltonian path requires solving the travelling salesman problem for which they naturally propose heuristics instead of finding the optimum path. Previously, [17] had used space filling curves as a computationally efficient heuristic to solve the travelling salesman problem.

| ID | Test                                        |
|----|---------------------------------------------|
| A  | Kolmogorov–Smirnov over space filling curve |
| B  | Kuiper’s test over SFC                      |
| C  | Wald–Wolfowitz over SFC                     |
| D  | Wilk’s empty cell test over SFC             |
| E  | Rosenbaum (2005)                            |
| F  | Biswas et al. (2014)                        |
| G  | Mann-Whitney based on Euclidean depth       |
| H  | Mann-Whitney based on Mahalanobis’ depth    |
| I  | Mann-Whitney based on Tukey’s depth         |

**Table 1.** List of tests used for simulation

#### 4.4. Wilk’s two-sample cells test

This test is also based on ranks and can be found in Wilk’s (1962). Given two samples  $\{X_i\}_{i=1}^m$  and  $\{Y_i\}_{i=1}^n$ , define intervals or cells  $I_i = (X_{(i-1)}, X_{(i)}]$ , for  $1 \leq i \leq m+1$  and where  $X_{(0)} = -\infty$  and  $X_{(m+1)} = +\infty$ . Let us denote by  $r_i$ ,  $i = 1, \dots, m+1$ , the number of elements of  $\{Y_i\}_{i=1}^n$  falling in  $I_i$  and  $S := \#\{i : r_i = 0\}$ . If the distributions of the observations in both samples are the same, we have for  $s_0 = 0, \dots, m+1$ ,

$$\mathbb{P}(S = s_0) = \frac{\binom{m+1}{s_0} \binom{n-1}{m-s_0}}{\binom{m+n}{m}}.$$

Applying the general idea of Section 2 here thus amounts to counting the number of sets of the form  $\bar{\varphi} \left( (X_{(i-1)}^1, X_{(i)}^1] \right)$  in which no element of  $\{Y_i^d\}_{i=i}^{N_Y}$  falls. The complexity of this test is also  $\mathcal{O}(N \log(N))$ .

## 5. Power of the procedure and simulation results

In this section, we assess the power of the tests listed in Table 1 through a simulation study. Each test considered in the comparison is distribution-free. All the estimates of the power are obtained using 1000 independent samples for each setting and the test level is set to 5%. Even though our procedure theoretically works for any number of dimensions, we consider here bivariate distributions, the quality of the test depending on the number of bits on which one can encode (see Section 7). Finally, observe that we also present the capability of the test to meet the prescribed type I error as our simulations help understand how the power varies around the null hypothesis in the various proposed settings.

In all the settings considered, we used a precision of  $10^{-25}$  in the computation of the preimage of each point.

### 5.1. Weibull-based bivariate distributions

The power of the test is first evaluated for Weibull distribution based models. Formally, the family of distributions we use is  $\{\text{Weib}(2, \eta * X_i), X_i\}$ , where  $X_i \sim U[0, 1]$  and

the parameters are respectively shape and scale. The tests are performed for two distributions with  $\eta = 3$  and  $\eta$  ranging from 2.5 to 3.5 by step of 0.1, respectively. Results are reported in Table 2.

In most cases, the tests based on space filling curves exhibit better power behavior than the other tests considered. Kuiper’s test over space filling curves is the one with the most desirable properties. Indeed, even though it slightly underperforms in small sample cases, it outperforms the other tests with  $n$  growing, in particular for alternatives close to the null hypothesis.

### 5.2. Bivariate normals

The first setting involves bivariate normal distributions with varying location and scale parameters and a constant covariance matrix,  $I_2$ , i.e., the identity matrix with two rows. Results for a sequence of increasing differences in location along the x-axis are presented in Table 3. Note that results for depth-based tests are not reported in the cases where their power is not larger than the level of the test.

We next consider a difference of locations with constant norm equal to one, considering various possible rotations. The first distribution considered is a bivariate Gaussian with mean  $\mathbf{0}$  and the identity as covariance matrix. The alternatives will have the same scale and their location will be given by the couple  $(\cos(\alpha), \sin(\alpha))$ , where  $\alpha$  is an angle. This setting is obtained by considering 11 angles  $\alpha$ , defining 10 equal parts of the interval  $[0, \pi]$ . The location couples are written  $[\Delta x \ \Delta y]'$  in Table 4.

From these two tables, the main conclusion is that the proposed testing procedures exhibit a good power with respect to other tests; Kuiper’s test over space filling curves is clearly outperforming tests from the literature and does not, unlike the Kolmogorov–Smirnov test, lose power when the alternative is rotated.

### 5.3. Gaussian Mixtures

The last simulation setting we consider relies on Gaussian mixtures. The mixture is based on three bivariate Gaussian distributions with the following means  $\mu = \{(-2, 2), (0, 0), (2, -2)\}$  and all have the same covariance matrix  $\Sigma$  :

$$\Sigma = \begin{bmatrix} 0.8 & -0.72 \\ -0.72 & 0.8 \end{bmatrix}.$$

The vector of respective proportions  $\Pi$  depends on a constant  $\lambda$ , i.e.,  $\Pi = \{\frac{1-\lambda}{2}, \lambda, \frac{1-\lambda}{2}\}$ . The tests are performed on two distributions, the first one with parameter  $\lambda = 0.3$  and the second one for which its  $\lambda$  ranges from 0 to 1 by step of 0.1. The results are presented in Table 5.

In this last simulation setting, under the alternatives, one of the samples is concentrated in the middle of the other. Due to this, a center-outward ranking of points summarises well the information and the tests based on data-depth are more powerful than the other tests. Kuiper’s test over space filling curves still shows good performance when compared to the other tests not based on data depth. In conclusion, even though Kuiper’s test over space filling curves is not more powerful than all existing tests, it gives reasonable results in all considered settings. Its power when the alternative is close to the null in medium size samples is an improvement with respect to Biswas et al. and, unlike depth based tests, it does not exhibit tremendous lack of power against

|    |             | Varying $\eta$ |     |     |     |     |   |     |     |     |     |     |
|----|-------------|----------------|-----|-----|-----|-----|---|-----|-----|-----|-----|-----|
| ID | Sample size | 2.5            | 2.6 | 2.7 | 2.8 | 2.9 | 3 | 3.1 | 3.2 | 3.3 | 3.4 | 3.5 |
| A  | 8           | 9              | 5   | 4   | 3   | 2   | 2 | 2   | 2   | 2   | 2   | 3   |
| B  | 8           | 14             | 9   | 5   | 3   | 2   | 2 | 1   | 2   | 3   | 4   | 4   |
| C  | 8           | 25             | 14  | 8   | 6   | 3   | 3 | 4   | 3   | 5   | 6   | 7   |
| D  | 8           | 25             | 14  | 7   | 5   | 2   | 2 | 2   | 1   | 2   | 2   | 3   |
| E  | 8           | 8              | 3   | 2   | 1   | 1   | 1 | 1   | 1   | 1   | 1   | 1   |
| F  | 8           | 31             | 15  | 8   | 5   | 3   | 3 | 3   | 3   | 5   | 5   | 6   |
| G  | 8           | 5              | 6   | 5   | 6   | 6   | 4 | 4   | 5   | 5   | 6   | 6   |
| H  | 8           | 4              | 3   | 6   | 5   | 5   | 5 | 5   | 5   | 5   | 5   | 4   |
| I  | 8           | 2              | 2   | 3   | 4   | 4   | 3 | 4   | 4   | 3   | 4   | 3   |
| A  | 16          | 31             | 15  | 10  | 7   | 4   | 4 | 4   | 6   | 5   | 9   | 8   |
| B  | 16          | 49             | 26  | 13  | 6   | 3   | 3 | 3   | 5   | 5   | 8   | 11  |
| C  | 16          | 53             | 27  | 11  | 6   | 2   | 3 | 2   | 4   | 7   | 8   | 9   |
| D  | 16          | 53             | 27  | 11  | 6   | 2   | 2 | 1   | 2   | 3   | 4   | 5   |
| E  | 16          | 60             | 31  | 16  | 9   | 5   | 4 | 5   | 7   | 8   | 12  | 13  |
| F  | 16          | 60             | 26  | 13  | 5   | 3   | 2 | 2   | 3   | 5   | 7   | 9   |
| G  | 16          | 10             | 9   | 9   | 8   | 6   | 4 | 4   | 7   | 6   | 7   | 8   |
| H  | 16          | 7              | 7   | 6   | 6   | 6   | 5 | 5   | 5   | 5   | 8   | 6   |
| I  | 16          | 2              | 4   | 5   | 6   | 6   | 6 | 5   | 6   | 5   | 7   | 6   |
| A  | 32          | 71             | 43  | 20  | 9   | 5   | 5 | 5   | 6   | 11  | 16  | 24  |
| B  | 32          | 89             | 65  | 29  | 12  | 5   | 3 | 5   | 9   | 14  | 21  | 35  |
| C  | 32          | 95             | 68  | 32  | 12  | 5   | 5 | 7   | 9   | 16  | 22  | 32  |
| D  | 32          | 97             | 77  | 39  | 18  | 9   | 6 | 8   | 9   | 17  | 23  | 32  |
| E  | 32          | 85             | 48  | 16  | 6   | 4   | 4 | 4   | 4   | 10  | 13  | 15  |
| F  | 32          | 97             | 68  | 30  | 13  | 7   | 6 | 7   | 8   | 15  | 21  | 30  |
| G  | 32          | 15             | 17  | 13  | 8   | 6   | 4 | 6   | 6   | 8   | 10  | 13  |
| H  | 32          | 10             | 11  | 10  | 7   | 6   | 4 | 6   | 7   | 8   | 11  | 10  |
| I  | 32          | 2              | 5   | 7   | 7   | 6   | 5 | 6   | 7   | 6   | 7   | 6   |
| A  | 64          | 98             | 76  | 40  | 14  | 5   | 4 | 6   | 10  | 16  | 28  | 38  |
| B  | 64          | 100            | 94  | 65  | 27  | 7   | 3 | 6   | 16  | 30  | 49  | 64  |
| C  | 64          | 100            | 92  | 51  | 19  | 8   | 4 | 7   | 11  | 22  | 35  | 52  |
| D  | 64          | 100            | 96  | 65  | 29  | 14  | 9 | 10  | 16  | 28  | 42  | 60  |
| E  | 64          | 100            | 78  | 33  | 11  | 4   | 3 | 4   | 6   | 12  | 23  | 33  |
| F  | 64          | 100            | 93  | 51  | 17  | 6   | 5 | 7   | 10  | 19  | 33  | 50  |
| G  | 64          | 35             | 33  | 23  | 14  | 8   | 6 | 6   | 10  | 13  | 17  | 23  |
| H  | 64          | 18             | 20  | 14  | 12  | 9   | 5 | 6   | 10  | 12  | 15  | 19  |
| I  | 64          | 3              | 8   | 11  | 10  | 7   | 6 | 5   | 10  | 11  | 12  | 13  |
| A  | 128         | 100            | 100 | 81  | 33  | 8   | 4 | 8   | 15  | 40  | 64  | 83  |
| B  | 128         | 100            | 100 | 96  | 57  | 12  | 3 | 12  | 32  | 62  | 85  | 95  |
| C  | 128         | 100            | 100 | 81  | 33  | 9   | 4 | 7   | 15  | 37  | 62  | 83  |
| D  | 128         | 100            | 100 | 87  | 41  | 12  | 7 | 9   | 18  | 42  | 66  | 86  |
| E  | 128         | 100            | 97  | 52  | 13  | 3   | 2 | 3   | 6   | 15  | 33  | 49  |
| F  | 128         | 100            | 100 | 81  | 30  | 7   | 5 | 8   | 16  | 35  | 58  | 78  |
| G  | 128         | 62             | 60  | 42  | 22  | 9   | 5 | 8   | 13  | 22  | 33  | 41  |
| H  | 128         | 36             | 38  | 33  | 18  | 7   | 5 | 7   | 13  | 21  | 24  | 32  |
| I  | 128         | 5              | 18  | 22  | 16  | 8   | 5 | 7   | 12  | 18  | 19  | 23  |

| ID | Sample size | Varying $\eta$ |     |     |     |     |   |     |     |     |     |     |
|----|-------------|----------------|-----|-----|-----|-----|---|-----|-----|-----|-----|-----|
|    |             | 2.5            | 2.6 | 2.7 | 2.8 | 2.9 | 3 | 3.1 | 3.2 | 3.3 | 3.4 | 3.5 |
| A  | 256         | 100            | 100 | 99  | 69  | 14  | 5 | 12  | 39  | 77  | 96  | 100 |
| B  | 256         | 100            | 100 | 100 | 87  | 24  | 4 | 19  | 60  | 93  | 100 | 100 |
| C  | 256         | 100            | 100 | 99  | 54  | 11  | 5 | 10  | 28  | 63  | 87  | 98  |
| D  | 256         | 100            | 100 | 99  | 58  | 12  | 6 | 10  | 28  | 63  | 87  | 98  |
| E  | 256         | 100            | 100 | 90  | 34  | 9   | 5 | 7   | 18  | 41  | 68  | 91  |
| F  | 256         | 100            | 100 | 98  | 49  | 12  | 5 | 10  | 26  | 60  | 87  | 98  |
| G  | 256         | 93             | 90  | 72  | 36  | 13  | 5 | 10  | 25  | 41  | 56  | 70  |
| H  | 256         | 66             | 70  | 57  | 30  | 13  | 6 | 11  | 24  | 34  | 45  | 57  |
| I  | 256         | 8              | 37  | 43  | 30  | 12  | 5 | 10  | 20  | 30  | 37  | 41  |
| A  | 512         | 100            | 100 | 100 | 97  | 26  | 5 | 22  | 75  | 98  | 100 | 100 |
| B  | 512         | 100            | 100 | 100 | 100 | 47  | 4 | 37  | 94  | 100 | 100 | 100 |
| C  | 512         | 100            | 100 | 100 | 81  | 17  | 6 | 12  | 43  | 87  | 99  | 100 |
| D  | 512         | 100            | 100 | 100 | 82  | 19  | 6 | 12  | 43  | 87  | 99  | 100 |
| E  | 512         | 100            | 100 | 99  | 55  | 9   | 4 | 8   | 26  | 63  | 93  | 99  |
| F  | 512         | 100            | 100 | 100 | 78  | 14  | 6 | 12  | 40  | 88  | 99  | 100 |
| G  | 512         | 100            | 100 | 95  | 66  | 19  | 5 | 16  | 45  | 70  | 86  | 94  |
| H  | 512         | 92             | 95  | 89  | 58  | 17  | 5 | 15  | 39  | 61  | 76  | 86  |
| I  | 512         | 21             | 69  | 79  | 55  | 17  | 5 | 13  | 38  | 56  | 68  | 73  |

**Table 2.** Power comparison (in %). Weibull based bivariate distributions(see above).

location alternatives. For this reason, we suggest to use this proposed test. Naturally, if one has an *a priori* on what type of alternative may occur, one should choose the test accordingly. We now show one toy example where the test could be applied.

## 6. Application to stellar evolution

When characterising stars, two factors are deemed important: luminosity and effective temperature. A graph representation of these two components is called Hertzsprung–Russel diagram. From this diagram, clusters of stars were built and characterisation of the lifecycle of stars can be studied by looking at which stars move from one cluster to the other. White dwarfs, supergiants or giants are examples of these clusters’ names that are often represented on the diagram.

Assume that we would like to test if two groups of stars can be said to be in the same state. To generate this situation, we use the COMBO-17 dataset [22, pp. 242] and apply a clustering algorithm<sup>2</sup>. The output of the clustering algorithm is reproduced in Figure 2. On the graph, the points’ shape shows the group a point was assigned to, the  $y$ -axis measures the absolute magnitude, which is a measure of luminosity of a star with respect to a reference and the  $x$ -axis is a measure of surface temperature of the star. The two groups are then centred: indeed, beyond the locations difference between those groups, we look for other types of alternatives in this toy example.

Applying the test proposed in this paper, that is applying Kuiper’s test over SFC, we obtain that the p-value returned is 0. Based on our simulation study, we will compare the test only with those who exhibited a good power. We obtain that the test by Biswas et al. had a p-value of order  $10^{-6}$ , while the best test relying on data depth

---

<sup>2</sup>To be precise, we use the `hclust` function from R. 2 clusters are built.

|    |             | Difference in location |    |    |    |     |     |     |     |     |     |     |
|----|-------------|------------------------|----|----|----|-----|-----|-----|-----|-----|-----|-----|
| ID | Sample size | 0                      | .1 | .2 | .3 | .4  | .5  | .6  | .7  | .8  | .9  | 1   |
| A  | 8           | 2                      | 2  | 1  | 3  | 3   | 6   | 8   | 8   | 11  | 14  | 18  |
| B  | 8           | 2                      | 2  | 2  | 2  | 3   | 3   | 5   | 4   | 8   | 9   | 10  |
| C  | 8           | 3                      | 4  | 3  | 4  | 5   | 6   | 7   | 6   | 11  | 11  | 14  |
| D  | 8           | 2                      | 2  | 2  | 3  | 4   | 4   | 5   | 5   | 8   | 9   | 11  |
| E  | 8           | 1                      | 1  | 0  | 1  | 1   | 0   | 1   | 1   | 1   | 1   | 1   |
| F  | 8           | 3                      | 3  | 3  | 3  | 5   | 4   | 5   | 6   | 10  | 10  | 15  |
| A  | 16          | 3                      | 4  | 5  | 8  | 10  | 15  | 18  | 28  | 33  | 38  | 47  |
| B  | 16          | 3                      | 2  | 3  | 4  | 6   | 8   | 10  | 15  | 18  | 22  | 30  |
| C  | 16          | 3                      | 3  | 3  | 4  | 4   | 5   | 6   | 9   | 10  | 13  | 19  |
| D  | 16          | 2                      | 2  | 2  | 2  | 3   | 4   | 5   | 8   | 9   | 10  | 16  |
| E  | 16          | 4                      | 5  | 5  | 6  | 6   | 7   | 9   | 12  | 14  | 17  | 21  |
| F  | 16          | 3                      | 3  | 3  | 3  | 4   | 4   | 7   | 8   | 12  | 14  | 18  |
| A  | 32          | 5                      | 6  | 9  | 14 | 22  | 27  | 38  | 52  | 63  | 75  | 81  |
| B  | 32          | 4                      | 5  | 6  | 8  | 14  | 17  | 26  | 37  | 46  | 61  | 68  |
| C  | 32          | 6                      | 5  | 5  | 7  | 9   | 11  | 18  | 23  | 29  | 41  | 43  |
| D  | 32          | 9                      | 6  | 6  | 9  | 12  | 13  | 21  | 27  | 32  | 43  | 47  |
| E  | 32          | 3                      | 3  | 3  | 4  | 5   | 7   | 9   | 9   | 13  | 19  | 21  |
| F  | 32          | 6                      | 6  | 7  | 8  | 10  | 11  | 16  | 21  | 26  | 38  | 44  |
| A  | 64          | 3                      | 5  | 12 | 20 | 31  | 49  | 67  | 78  | 89  | 97  | 98  |
| B  | 64          | 3                      | 4  | 9  | 13 | 24  | 37  | 56  | 69  | 82  | 93  | 96  |
| C  | 64          | 5                      | 4  | 6  | 7  | 10  | 14  | 22  | 29  | 42  | 53  | 64  |
| D  | 64          | 8                      | 8  | 10 | 11 | 15  | 20  | 32  | 38  | 52  | 66  | 75  |
| E  | 64          | 4                      | 4  | 5  | 5  | 6   | 12  | 13  | 18  | 25  | 34  | 44  |
| F  | 64          | 4                      | 5  | 7  | 6  | 11  | 15  | 23  | 31  | 43  | 57  | 66  |
| A  | 128         | 5                      | 9  | 23 | 41 | 66  | 84  | 95  | 99  | 100 | 100 | 100 |
| B  | 128         | 3                      | 7  | 17 | 30 | 54  | 75  | 91  | 97  | 99  | 100 | 100 |
| C  | 128         | 4                      | 6  | 7  | 10 | 15  | 19  | 34  | 49  | 64  | 80  | 91  |
| D  | 128         | 7                      | 7  | 11 | 14 | 19  | 25  | 41  | 57  | 72  | 85  | 94  |
| E  | 128         | 2                      | 3  | 3  | 5  | 6   | 8   | 13  | 24  | 36  | 46  | 65  |
| F  | 128         | 5                      | 5  | 7  | 11 | 15  | 19  | 38  | 51  | 64  | 81  | 92  |
| A  | 256         | 5                      | 12 | 39 | 70 | 92  | 99  | 100 | 100 | 100 | 100 | 100 |
| B  | 256         | 4                      | 8  | 28 | 59 | 87  | 98  | 100 | 100 | 100 | 100 | 100 |
| C  | 256         | 4                      | 7  | 9  | 13 | 22  | 39  | 59  | 78  | 92  | 99  | 100 |
| D  | 256         | 5                      | 7  | 9  | 14 | 23  | 40  | 60  | 79  | 93  | 99  | 100 |
| E  | 256         | 5                      | 4  | 6  | 11 | 15  | 25  | 37  | 54  | 70  | 87  | 95  |
| F  | 256         | 4                      | 6  | 7  | 13 | 23  | 39  | 59  | 77  | 92  | 98  | 99  |
| A  | 512         | 3                      | 21 | 67 | 94 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| B  | 512         | 4                      | 17 | 56 | 90 | 99  | 100 | 100 | 100 | 100 | 100 | 100 |
| C  | 512         | 4                      | 5  | 8  | 17 | 33  | 60  | 81  | 95  | 99  | 100 | 100 |
| D  | 512         | 5                      | 5  | 9  | 18 | 34  | 61  | 82  | 95  | 100 | 100 | 100 |
| E  | 512         | 4                      | 4  | 7  | 12 | 21  | 36  | 56  | 78  | 92  | 98  | 100 |
| F  | 512         | 5                      | 5  | 8  | 19 | 36  | 59  | 80  | 96  | 99  | 100 | 100 |

**Table 3.** Power comparison (in %). Independent bivariate Gaussian distributions with location differences along the x-axis.

|    |             | Different settings: $\begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}$ |              |              |              |              |        |               |               |               |               |           |
|----|-------------|--------------------------------------------------------------------------|--------------|--------------|--------------|--------------|--------|---------------|---------------|---------------|---------------|-----------|
| ID | Sample size | 1<br>0                                                                   | 0.95<br>0.31 | 0.81<br>0.59 | 0.59<br>0.81 | 0.31<br>0.95 | 0<br>1 | -0.31<br>0.95 | -0.59<br>0.81 | -0.81<br>0.59 | -0.95<br>0.31 | -1.0<br>0 |
| A  | 8           | 18                                                                       | 16           | 13           | 9            | 6            | 5      | 6             | 9             | 10            | 15            | 18        |
| B  | 8           | 10                                                                       | 10           | 10           | 11           | 10           | 12     | 11            | 10            | 9             | 10            | 11        |
| C  | 8           | 13                                                                       | 13           | 13           | 14           | 13           | 15     | 14            | 13            | 11            | 12            | 13        |
| D  | 8           | 10                                                                       | 8            | 6            | 5            | 3            | 5      | 4             | 5             | 6             | 7             | 11        |
| E  | 8           | 2                                                                        | 2            | 3            | 3            | 2            | 2      | 2             | 2             | 2             | 1             | 2         |
| F  | 8           | 9                                                                        | 11           | 13           | 13           | 12           | 14     | 13            | 13            | 12            | 12            | 12        |
| A  | 16          | 47                                                                       | 46           | 39           | 26           | 18           | 15     | 17            | 27            | 36            | 47            | 46        |
| B  | 16          | 31                                                                       | 31           | 31           | 33           | 32           | 30     | 31            | 31            | 32            | 30            | 28        |
| C  | 16          | 18                                                                       | 18           | 20           | 19           | 21           | 19     | 19            | 20            | 18            | 16            | 15        |
| D  | 16          | 14                                                                       | 13           | 13           | 9            | 10           | 9      | 9             | 11            | 12            | 13            | 13        |
| E  | 16          | 18                                                                       | 18           | 19           | 18           | 19           | 18     | 17            | 17            | 17            | 21            | 21        |
| F  | 16          | 16                                                                       | 18           | 17           | 16           | 19           | 17     | 17            | 18            | 16            | 17            | 16        |
| A  | 32          | 84                                                                       | 79           | 71           | 56           | 44           | 39     | 40            | 54            | 71            | 79            | 83        |
| B  | 32          | 72                                                                       | 68           | 71           | 70           | 70           | 71     | 70            | 67            | 69            | 69            | 68        |
| C  | 32          | 46                                                                       | 45           | 48           | 42           | 38           | 43     | 44            | 44            | 45            | 46            | 46        |
| D  | 32          | 50                                                                       | 47           | 49           | 43           | 38           | 43     | 44            | 45            | 46            | 48            | 49        |
| E  | 32          | 23                                                                       | 21           | 22           | 22           | 23           | 21     | 23            | 21            | 24            | 24            | 22        |
| F  | 32          | 47                                                                       | 46           | 47           | 46           | 44           | 47     | 47            | 45            | 49            | 47            | 44        |
| A  | 64          | 98                                                                       | 98           | 94           | 86           | 71           | 67     | 72            | 83            | 94            | 97            | 98        |
| B  | 64          | 96                                                                       | 97           | 97           | 96           | 96           | 96     | 95            | 96            | 95            | 96            | 96        |
| C  | 64          | 69                                                                       | 62           | 67           | 71           | 70           | 70     | 69            | 69            | 68            | 68            | 65        |
| D  | 64          | 78                                                                       | 71           | 74           | 75           | 74           | 72     | 72            | 73            | 74            | 77            | 76        |
| E  | 64          | 43                                                                       | 39           | 41           | 43           | 42           | 42     | 43            | 42            | 40            | 43            | 43        |
| F  | 64          | 68                                                                       | 65           | 67           | 69           | 70           | 65     | 69            | 66            | 66            | 70            | 67        |
| A  | 128         | 100                                                                      | 100          | 100          | 100          | 99           | 99     | 99            | 100           | 100           | 100           | 100       |
| B  | 128         | 100                                                                      | 100          | 100          | 100          | 100          | 100    | 100           | 100           | 100           | 100           | 100       |
| C  | 128         | 90                                                                       | 91           | 91           | 92           | 92           | 90     | 93            | 92            | 92            | 92            | 90        |
| D  | 128         | 93                                                                       | 93           | 93           | 93           | 93           | 92     | 93            | 93            | 93            | 94            | 93        |
| E  | 128         | 63                                                                       | 64           | 61           | 63           | 63           | 63     | 63            | 64            | 61            | 65            | 60        |
| F  | 128         | 92                                                                       | 92           | 92           | 91           | 91           | 90     | 92            | 91            | 90            | 92            | 91        |

**Table 4.** Power comparison (in %). Independent bivariate Gaussian distributions with fixed norm of the location difference and various relative positions.

|    |             | $\lambda$ |    |    |    |    |    |    |     |     |     |     |
|----|-------------|-----------|----|----|----|----|----|----|-----|-----|-----|-----|
| ID | Sample size | 0         | .1 | .2 | .3 | .4 | .5 | .6 | .7  | .8  | .9  | 1   |
| A  | 8           | 3         | 3  | 2  | 2  | 2  | 1  | 2  | 3   | 3   | 4   | 6   |
| B  | 8           | 3         | 2  | 2  | 1  | 2  | 1  | 2  | 5   | 5   | 6   | 11  |
| C  | 8           | 4         | 4  | 3  | 3  | 3  | 3  | 4  | 8   | 7   | 10  | 16  |
| D  | 8           | 2         | 2  | 2  | 2  | 3  | 2  | 3  | 7   | 7   | 10  | 16  |
| E  | 8           | 1         | 1  | 1  | 1  | 1  | 1  | 1  | 1   | 1   | 2   | 2   |
| F  | 8           | 7         | 5  | 4  | 3  | 4  | 4  | 5  | 9   | 16  | 21  | 32  |
| G  | 8           | 12        | 9  | 7  | 5  | 6  | 10 | 14 | 19  | 28  | 43  | 57  |
| H  | 8           | 12        | 8  | 7  | 4  | 5  | 9  | 14 | 20  | 31  | 42  | 59  |
| I  | 8           | 10        | 6  | 5  | 4  | 5  | 9  | 10 | 16  | 24  | 33  | 43  |
| A  | 16          | 5         | 4  | 4  | 3  | 4  | 4  | 4  | 8   | 7   | 11  | 13  |
| B  | 16          | 5         | 3  | 3  | 2  | 3  | 3  | 4  | 9   | 13  | 19  | 29  |
| C  | 16          | 5         | 3  | 3  | 2  | 3  | 3  | 4  | 6   | 11  | 18  | 28  |
| D  | 16          | 4         | 2  | 1  | 2  | 2  | 3  | 4  | 6   | 10  | 18  | 28  |
| E  | 16          | 8         | 4  | 5  | 5  | 5  | 6  | 9  | 10  | 10  | 20  | 27  |
| F  | 16          | 7         | 3  | 3  | 2  | 3  | 4  | 6  | 10  | 16  | 31  | 50  |
| G  | 16          | 22        | 11 | 8  | 5  | 6  | 12 | 24 | 38  | 55  | 75  | 88  |
| H  | 16          | 24        | 13 | 8  | 5  | 5  | 14 | 25 | 40  | 57  | 78  | 91  |
| I  | 16          | 22        | 12 | 8  | 4  | 5  | 14 | 24 | 36  | 53  | 73  | 85  |
| A  | 32          | 9         | 7  | 4  | 5  | 4  | 6  | 8  | 11  | 16  | 24  | 35  |
| B  | 32          | 11        | 7  | 4  | 3  | 4  | 6  | 11 | 20  | 34  | 50  | 67  |
| C  | 32          | 16        | 9  | 5  | 6  | 6  | 8  | 11 | 19  | 34  | 52  | 75  |
| D  | 32          | 19        | 11 | 7  | 7  | 7  | 10 | 16 | 24  | 40  | 60  | 83  |
| E  | 32          | 9         | 4  | 4  | 3  | 3  | 4  | 6  | 10  | 17  | 27  | 44  |
| F  | 32          | 20        | 9  | 7  | 6  | 5  | 8  | 12 | 22  | 37  | 58  | 83  |
| G  | 32          | 43        | 21 | 8  | 6  | 9  | 23 | 45 | 71  | 88  | 96  | 100 |
| H  | 32          | 44        | 24 | 10 | 5  | 9  | 26 | 48 | 74  | 90  | 97  | 100 |
| I  | 32          | 44        | 22 | 9  | 5  | 8  | 24 | 46 | 71  | 89  | 97  | 100 |
| A  | 64          | 13        | 8  | 4  | 3  | 5  | 6  | 12 | 19  | 31  | 47  | 68  |
| B  | 64          | 21        | 11 | 5  | 4  | 6  | 10 | 24 | 44  | 65  | 84  | 97  |
| C  | 64          | 24        | 9  | 7  | 4  | 5  | 10 | 18 | 31  | 50  | 80  | 98  |
| D  | 64          | 32        | 15 | 10 | 7  | 10 | 15 | 25 | 43  | 64  | 87  | 99  |
| E  | 64          | 14        | 7  | 4  | 3  | 4  | 5  | 10 | 18  | 31  | 57  | 83  |
| F  | 64          | 28        | 13 | 5  | 4  | 6  | 9  | 16 | 35  | 59  | 85  | 99  |
| G  | 64          | 73        | 39 | 14 | 5  | 14 | 40 | 74 | 96  | 100 | 100 | 100 |
| H  | 64          | 75        | 43 | 13 | 4  | 15 | 44 | 78 | 97  | 100 | 100 | 100 |
| I  | 64          | 75        | 42 | 13 | 4  | 15 | 43 | 76 | 96  | 100 | 100 | 100 |
| A  | 128         | 29        | 13 | 6  | 4  | 6  | 11 | 23 | 44  | 70  | 90  | 98  |
| B  | 128         | 52        | 22 | 8  | 4  | 7  | 22 | 51 | 81  | 95  | 99  | 100 |
| C  | 128         | 45        | 14 | 6  | 4  | 6  | 9  | 23 | 50  | 82  | 98  | 100 |
| D  | 128         | 50        | 19 | 8  | 6  | 8  | 14 | 29 | 57  | 87  | 99  | 100 |
| E  | 128         | 21        | 6  | 3  | 1  | 3  | 4  | 11 | 23  | 48  | 82  | 99  |
| F  | 128         | 47        | 13 | 6  | 3  | 6  | 10 | 24 | 54  | 82  | 98  | 100 |
| G  | 128         | 96        | 68 | 22 | 6  | 22 | 72 | 96 | 100 | 100 | 100 | 100 |
| H  | 128         | 98        | 72 | 27 | 4  | 25 | 74 | 98 | 100 | 100 | 100 | 100 |
| I  | 128         | 97        | 72 | 27 | 5  | 25 | 73 | 97 | 100 | 100 | 100 | 100 |

| ID | Sample size | $\lambda$ |     |    |    |    |     |     |     |     |     |     |
|----|-------------|-----------|-----|----|----|----|-----|-----|-----|-----|-----|-----|
|    |             | 0         | .1  | .2 | .3 | .4 | .5  | .6  | .7  | .8  | .9  | 1   |
| A  | 256         | 50        | 23  | 9  | 5  | 9  | 23  | 52  | 81  | 98  | 100 | 100 |
| B  | 256         | 85        | 44  | 13 | 4  | 11 | 45  | 85  | 98  | 100 | 100 | 100 |
| C  | 256         | 72        | 23  | 9  | 6  | 7  | 15  | 42  | 79  | 98  | 100 | 100 |
| D  | 256         | 73        | 24  | 9  | 7  | 8  | 16  | 46  | 81  | 98  | 100 | 100 |
| E  | 256         | 51        | 13  | 7  | 5  | 7  | 10  | 26  | 60  | 89  | 100 | 100 |
| F  | 256         | 77        | 22  | 9  | 6  | 8  | 16  | 43  | 79  | 98  | 100 | 100 |
| G  | 256         | 100       | 94  | 44 | 5  | 39 | 93  | 100 | 100 | 100 | 100 | 100 |
| H  | 256         | 100       | 96  | 46 | 5  | 43 | 95  | 100 | 100 | 100 | 100 | 100 |
| I  | 256         | 100       | 95  | 45 | 5  | 42 | 95  | 100 | 100 | 100 | 100 | 100 |
| A  | 512         | 89        | 46  | 13 | 5  | 12 | 42  | 87  | 100 | 100 | 100 | 100 |
| B  | 512         | 99        | 81  | 24 | 5  | 25 | 78  | 99  | 100 | 100 | 100 | 100 |
| C  | 512         | 93        | 34  | 9  | 4  | 8  | 24  | 61  | 95  | 100 | 100 | 100 |
| D  | 512         | 94        | 35  | 9  | 4  | 8  | 25  | 62  | 95  | 100 | 100 | 100 |
| E  | 512         | 77        | 21  | 6  | 4  | 7  | 13  | 40  | 80  | 99  | 100 | 100 |
| F  | 512         | 95        | 35  | 7  | 5  | 9  | 23  | 60  | 95  | 100 | 100 | 100 |
| G  | 512         | 100       | 100 | 72 | 6  | 71 | 100 | 100 | 100 | 100 | 100 | 100 |
| H  | 512         | 100       | 100 | 75 | 6  | 74 | 100 | 100 | 100 | 100 | 100 | 100 |
| I  | 512         | 100       | 100 | 74 | 6  | 72 | 100 | 100 | 100 | 100 | 100 | 100 |

**Table 5.** Power comparison (in %). Gaussian mixture with varying weights (see full description in the text).

was the one using Euclidean distance with a p-value of order  $10^{-7}$ .

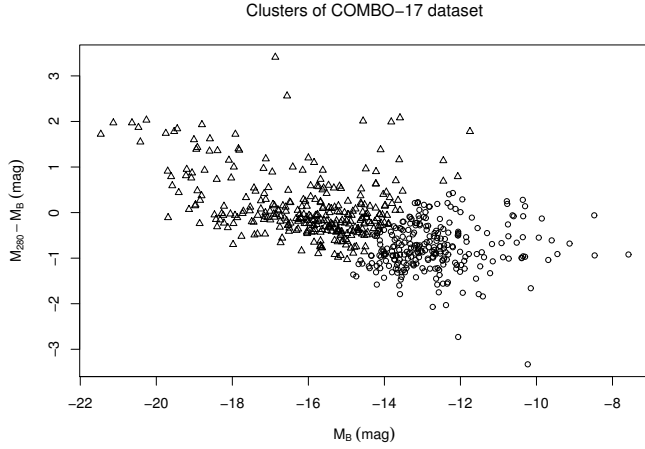


**Figure 2.** Output of the clustering algorithm for the COMBO-17 dataset.

## 7. Discussion and conclusion

The methodology, even though solely presented for the two-sample problem, can also be used for more than two groups; one could for instance use the test proposed by

[23], [24], or some  $k$ -sample generalisation of the Kolmogorov–Smirnov test, see [25] or [26], to name but a few.

As the computational complexity of our procedure is rather limited and given its distribution-freeness, it can be used —possibly in conjunction with other metric— for virtually any machine learning pipeline where two distributions need to be compared. Combining distances is rather appealing. Indeed, the fact that the distance relies on an (admittedly peculiar) ordering of the data might also make the procedures more robust to particular alternatives that would not have been accounted for. A thorough empirical analysis of the potential gains of the idea just mentioned certainly deserves to be investigated in an independent work.

In this paper, we proposed to use space filling curves for performing  $d$ -dimensional distributions equality tests. The proposed procedure allows for directly applying usual statistics, even though some may be more appropriate than others. For this reason, we propose an analysis of some classical tests performed on the preimages of the filling curve and interpret them in the original multidimensional space.

We then compare those proposed testing procedures with other distribution-free tests on various simulation settings. The results suggest that our methodology performs well compared to other techniques, especially when Kuiper’s test is applied on the space filling curve and little information is known about the alternative hypothesis.

Finally, the test can be used in higher dimensions; this is however left for further developments since the existing algorithm for computing the preimage of a point requires enough bits of encoding, which is not something one could test in R as it is mainly developed for 32 bits. These technical details are considered beyond the scope of the present paper.

## Acknowledgement(s)

The authors warmly thank Dr. Ghosh for sharing the code of his test procedure.

## References

- [1] Hallin M. On distribution and quantile functions, ranks and signs in  $\mathbb{R}^d$ . ECARES Working Papers. 2017;.
- [2] Peacock J. Two-dimensional goodness-of-fit testing in astronomy. Monthly Notices of the Royal Astronomical Society. 1983;202(3):615–627.
- [3] Fasano G, Franceschini A. A multidimensional version of the kolmogorov–smirnov test. Monthly Notices of the Royal Astronomical Society. 1987;225(1):155–170.
- [4] Justel A, Peña D, Zamar R. A multivariate kolmogorov-smirnov test of goodness of fit. Statistics & Probability Letters. 1997;35(3):251–259.
- [5] Rosenblatt M. Remarks on a multivariate transformation. The annals of mathematical statistics. 1952;23(3):470–472.
- [6] Friedman JH, Rafsky LC. Multivariate generalizations of the wald-wolfowitz and smirnov two-sample tests. The Annals of Statistics. 1979;:697–717.
- [7] Schilling MF. Multivariate two-sample tests based on nearest neighbors. Journal of the American Statistical Association. 1986;81(395):799–806.
- [8] Henze N. A multivariate two-sample test based on the number of nearest neighbor type coincidences. The Annals of Statistics. 1988;:772–783.

- [9] Liu Z, Modarres R. A triangle test for equality of distribution functions in high dimensions. *Journal of Nonparametric Statistics*. 2011;23(3):605–615.
- [10] Rosenbaum PR. An exact distribution-free test comparing two multivariate distributions based on adjacency. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2005;67(4):515–530.
- [11] Biswas M, Mukhopadhyay M, Ghosh AK. A distribution-free two-sample run test applicable to high-dimensional data. *Biometrika*. 2014;101(4):913–926.
- [12] Liu RY, Singh K. A quality index based on data depth and multivariate rank tests. *Journal of the American Statistical Association*. 1993;88(421):252–260.
- [13] Rousson V. On distribution-free tests for the multivariate two-sample location-scale model. *Journal of multivariate analysis*. 2002;80(1):43–57.
- [14] Székely GJ, Rizzo ML. The energy of data. *Annual Review of Statistics and Its Application*. 2017;4:447–479.
- [15] Bernton E, Jacob PE, Gerber M, et al. Approximate bayesian computation with the wasserstein distance. *Journal of the Royal Statistical Society: Series B*. 2019;81(2):235–269.
- [16] Haverkort HJ. Sixteen space-filling curves and traversals for d-dimensional cubes and simplices. CoRR. 2017;abs/1711.04473. Available from: <http://arxiv.org/abs/1711.04473>.
- [17] Gao J, Steele JM. General spacefilling curve heuristics and limit theory for the traveling salesman problem. *Journal of Complexity*. 1994;10(2):230–245.
- [18] Bader M. Space-filling curves: an introduction with applications in scientific computing. Vol. 9. Springer Science & Business Media; 2012.
- [19] Gonzalez T, Sahni S, Franta WR. An efficient algorithm for the kolmogorov-smirnov and lilliefors tests. *ACM Transactions on Mathematical Software (TOMS)*. 1977;3(1):60–64.
- [20] Hallin M, del Barrio E, Cuesta-Albertos J, et al. Distribution and quantile functions, ranks and signs in dimension  $d$ : A measure transportation approach. *The Annals of Statistics*. 2021;49(2):1139–1165.
- [21] Wald A, Wolfowitz J. Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics*. 1948;:326–339.
- [22] Feigelson ED, Babu GJ. Modern statistical methods for astronomy: with r applications. Cambridge University Press; 2012.
- [23] Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*. 1952;47(260):583–621.
- [24] van der Waerden B. Order tests for the two-sample problem and their power. In: *Indagationes Mathematicae (Proceedings)*; Vol. 55; Elsevier; 1952. p. 453–458.
- [25] Kiefer J. K-sample analogues of the kolmogorov-smirnov and cramér-v. mises tests. *The Annals of Mathematical Statistics*. 1959;:420–447.
- [26] Conover W. Several k-sample kolmogorov-smirnov tests. *The Annals of Mathematical Statistics*. 1965;36(3):1019–1026.

## Appendix A. An algorithm for computing preimage of a point in 2D

The algorithm presented in this section is from [18]. It is presented here to help to understand the sequential localisation of the points on the 1-dimensional space by refining the position in space. The fact that it is a pseudo inverse can be noticed from the fact that certain conditions include an equality sign, fixing a localisation to a point that could have had another a preimage than the one chosen.

```

double xy2d_Bader (double x, double y, double eps)
{
    if (eps>1){return 0;}
    if (x<0.5)
    {
        if(y<0.5)
        {
            return (0+xy2d_Bader(2*y,2*x,4*eps))/4;
        }
        else
        {
            return (1+xy2d_Bader(2*x,2*y-1,4*eps))/4;
        }
    }else
    {
        if(y>=0.5)
        {
            return (2+xy2d_Bader(2*x-1,2*y-1,4*eps))/4;
        }
        else
        {
            return (3+xy2d_Bader(1-2*y,2-2*x,4*eps))/4;
        }
    }
}

```

**Figure A1.** Hilbert index algorithm