A statistical test of isomorphism between metric-measure spaces using the distance-to-a-measure signature

Claire Brécheteau

Electronic Journal of Statistics
Vol. 0 (0000)
ISSN: 1935-7524
DOI: 10.1214/154957804100000000

A statistical test of isomorphism between metric-measure spaces using the distance-to-a-measure signature*

Claire Brécheteau

Laboratoire de Mathématiques
Université Paris-Sud
91405 Orsay Cedex France
e-mail: claire.brecheteau@inria.fr
url: https://www.math.sciences.univ-nantes.fr/ brecheteau/

Abstract: We introduce the notion of DTM-signature, a measure on $\mathbb{R}$ that can be associated to any metric-measure space. This signature is based on the function distance to a measure (DTM) introduced in 2009 by Chazal, Cohen-Steiner and Mérigot. It leads to a pseudo-metric between metric-measure spaces, that is bounded above by the Gromov-Wasserstein distance. This pseudo-metric is used to build a statistical test of isomorphism between two metric-measure spaces, from the observation of two N-samples.

The test is based on subsampling methods and comes with theoretical guarantees. It is proven to be of the correct level asymptotically. Also, when the measures are supported on compact subsets of $\mathbb{R}^d$, rates of convergence are derived for the L_1-Wasserstein distance between the distribution of the test statistic and its subsampling approximation. These rates depend on some parameter $\rho > 1$. In addition, we prove that the power is bounded above by $\exp(-CN^{1/\rho})$, with C proportional to the square of the aforementioned pseudo-metric between the metric-measure spaces. Under some geometrical assumptions, we also derive lower bounds for this pseudo-metric.

An algorithm is proposed for the implementation of this statistical test, and its performance is compared to the performance of other methods through numerical experiments.

MSC 2010 subject classifications: Primary 62G10; secondary 62G09.
Keywords and phrases: statistical test, subsampling, metric-measure spaces, distance to a measure, (Gromov)-Wasserstein distances.

1. Introduction

Very often data comes in the form of a set of points from a metric space. A natural question, given two such sets of data, is to decide whether they are similar. For example, do they come from the same distribution? Are their shapes similar? From the seminal two-samples tests of Kolmogorov-Smirnov specific to measures on $\mathbb{R}$ or even on $\mathbb{R}^d$, to the more recent kernel two-sample tests by Gretton et al. [29], where the data are sent into a reproducing kernel Hilbert space and then compared through the maximum mean discrepancy, the literature is abundant and proficient on the subject of two-sample testing. Note that an overview of Wasserstein-distance-based two-sample tests appears in [40].

Unfortunately, testing equality of two measures from samples may be compromised when the data are not embedded into the same space, or if the two systems of coordinates in which

*This work was partially supported by the ANR project TopData and GUDHI; the Inria teams Datashape and Select; and the LMO at Université Paris-Sud, Université Paris-Saclay.

the data are represented are different. To overcome this issue, an idea is to forget about the embedding and only consider the set of points together with the distances between pairs. A natural framework to compare data is then to assume that they come from a measure on a metric space and to consider two such metric-measure spaces as being the same when they are equal up to some isomorphism, as defined below.

Definition 1.1 (mm-space)**.** *A* ***metric-measure space*** *(**mm-space**) is a triple* $(\mathcal{X}, \delta, \mu)$*, with* $\mathcal{X}$ *a set,* δ *a metric on* $\mathcal{X}$ *and* μ *a probability measure on* $\mathcal{X}$ *equipped with its Borel* σ*-algebra.*

Definition 1.2 (Isomorphism between mm-spaces)**.** *Two mm-spaces* $(\mathcal{X}, \delta, \mu)$ *and* $(\mathcal{Y}, \gamma, \nu)$ *are said to be* ***isomorphic*** *if there exist a one-to-one and onto isometry* $\phi\colon \mathcal{X} \to \mathcal{Y}$ *preserving measures, that is, such that* $\nu(\phi(A)) = \mu(A)$ *for any Borel subset* A *of* $\mathcal{X}$*.*

Such a map ϕ *is called an* ***isomorphism*** *between the mm-spaces* $(\mathcal{X}, \delta, \mu)$ *and* $(\mathcal{Y}, \gamma, \nu)$*.*

In this paper, we address the question of the comparison of general mm-spaces, up to an isomorphism. In other terms, we aim to design a metric or at least a pseudo-metric on the quotient space of mm-spaces by the relation of isomorphism. A suitable pseudo-metric should be stable under some perturbations and under sampling, discriminative, and easy to implement when dealing with discrete spaces. The purpose is then to use such a pseudo-metric to build a statistical test of isomorphism between two mm-spaces, from the observation of two samples. In a sense, it generalises the scale-location tests of Fromont et al. [27], which aim at testing equality to a fixed distribution on $\mathbb{R}$, up to translation and dilatation.

In his work on metric-measure spaces [30], Gromov proposes a first characterisation of such spaces. Indeed, he proves in Theorem $3\frac{1}{2}.5$ of [30] that any mm-space can be recovered, up to an isomorphism, from the knowledge, for all sizes N, of the distribution of the $N \times N$-matrix of distances associated to a N-sample. More recently, in [35], Mémoli proposes metrics on the quotient space of mm-spaces by the relation of isomorphism: the Gromov–Wasserstein distances; see Section 3.1 for a definition.

Unfortunately, even when dealing with discrete mm-spaces, the computation of these Gromov–Wasserstein distances is extremely costly. An alternative is to build a **signature** from each mm-space, that is to say, a mathematical object satisfying the property that we assign the same signature to two isomorphic mm-spaces. The mm-spaces are then compared through their signatures. In [35], Mémoli gives an overview of signatures, such as the shape distribution, the eccentricity or what he calls local distribution of distances.

Shape signatures are widely used for classification or pre-classification tasks; see for instance [38]. With a more topological point of view, persistence diagrams have been used for this purpose in [12, 19]. In [16] and [24], the authors derive bootstrapped confidence intervals for landscapes and persistence diagrams. Both are topological objects which are invariant under isomorphism. Thus, such confidence intervals could be used to build a statistical test of isomorphism. However, these topological signatures are hard to compute, especially in high dimension, and the intervals are too conservative. Moreover, such signatures are almost blind to the measures. This is why it is of interest to propose another method based on a more tractable signature.

As far as we know, the construction of well-founded isomorphism statistical tests from signatures to compare mm-spaces has not been considered in the literature.

The signature we introduce in this paper is based on the distance to a measure, which is defined in [13] as a generalisation of the function distance to a compact set and can be defined as follows.

Let $(\mathcal{X}, \delta)$ be a metric space, equipped with a Borel probability measure μ. Given m in $[0,1]$, the **pseudo-distance function** is defined at any point x of $\mathcal{X}$, by

$$\delta_{\mu,m}(x) = \inf\{r > 0 \mid \mu(\overline{\mathrm{B}}(x,r)) > m\},$$

with $\overline{\mathrm{B}}(x,r) = \{y \in \mathcal{X} \mid \delta(x,y) \leq r\}$. The function **distance to the measure** (DTM) μ with mass parameter m and denoted $\mathrm{d}_{\mu,m}$ is then defined for all x in $\mathcal{X}$ by

$$\mathrm{d}_{\mu,m}(x) = \frac{1}{m}\int_{l=0}^{m} \delta_{\mu,l}(x)\,\mathrm{d}l. \tag{1.1}$$

This function can be easily computed when the measure of interest is uniform on a finite set of N points: $\hat{\mu}_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_i}$ with the X_is in a metric space $(\mathcal{X}, \delta)$. Indeed, in this case, $\delta_{\hat{\mu}_N,l}(x)$ is the distance between x and its $\lceil lN \rceil$ nearest neighbour, denoted by $X^{(\lceil lN \rceil)}$. As a consequence, the distance to the measure $\hat{\mu}_N$ with mass parameter $m = \frac{k}{N}$ for some k in $\{1, \ldots, N\}$ at a point x of $\mathcal{X}$ satisfies:

$$\mathrm{d}_{\hat{\mu}_N,m}(x) = \frac{1}{k}\sum_{i=1}^{k} \delta(X^{(i)}, x).$$

The distance to the measure $\hat{\mu}_N$ at $x \in \mathcal{X}$ is thus equal to the mean of the distances to its k-nearest neighbours in $\{X_1, X_2 \ldots, X_N\}$.

The DTM-signature is then defined as follows.

Definition 1.3 (DTM-signature)**.** *The **DTM-signature** associated to some mm-space $(\mathcal{X}, \delta, \mu)$, denoted $\mathrm{d}_{\mu,m}(\mu)$, is the distribution of the real-valued random variable $\mathrm{d}_{\mu,m}(X)$ where X is some random variable from the distribution μ.*

More generally, it will be of interest in this paper to consider the push-forward of a measure with the distance-to-a-measure function associated to another measure.

Notation 1.1. *Let μ and μ' be two Borel probability measures on the same metric space $(\mathcal{X}, \delta)$. We denote by $\mathrm{d}_{\mu,m}(\mu')$ the distribution of the random variable $\mathrm{d}_{\mu,m}(X')$, where X' is some random variable from the distribution μ'.*

An important example consists of approximating the DTM-signature associated to some mm-space $(\mathcal{X}, \delta, \mu)$ with an estimator built from a N-sample from μ. In principle, we would like to approximate $\mathrm{d}_{\mu,m}(\mu)$ with the measure $\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_N)$. But unfortunately, the distribution of the test statistic (see Section 2) based on the measure $\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_N)$ is difficult to approximate by bootstrapping or subsampling methods. This is why the methods proposed in this paper rely on a test statistic based on the measure $\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_n)$ for some n smaller than N. Indeed, choosing a n smaller than N helps introduce more randomness in the subsampling distribution, allowing a better approximation of the distribution of the test statistic under the hypothesis of isomorphism.

Definition 1.4 (empirical DTM-signature)**.** *Given a N-sample X_1, X_2,$\ldots X_N$ from a Borel probability measure μ, with the notation $\hat{\mu}_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_i}$ and $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$ for some $n \leq N$, we define the **empirical DTM-signature** as the discrete Borel probability measure on $\mathbb{R}$, $\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_n)$.*

Note that it corresponds to the discrete distribution $\frac{1}{n}\sum_{i=1}^{n} \delta_{\mathrm{d}_{\hat{\mu}_N,m}(X_i)}$.

We propose to compare two DTM-signatures or empirical DTM-signatures by means of the L_1-Wasserstein distance W_1, which is defined in the following way; see [42].

Definition 1.5 (Wasserstein distance)**.** *The L_1-**Wasserstein distance** between two Borel probability measures μ and ν over the same metric space $(\mathcal{X}, \delta)$ is defined as:*

$$W_1(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{X}} \delta(x,y)\mathrm{d}\pi(x,y),$$

*where $\Pi(\mu,\nu)$ stands for the set of **transport plans** between μ and ν, that is the set of Borel probability measures π on $\mathcal{X}\times\mathcal{Y}$ satisfying $\pi(A\times\mathcal{Y}) = \mu(A)$ and $\pi(\mathcal{X}\times B) = \nu(B)$ for all Borel sets A in $\mathcal{X}$ and B in $\mathcal{Y}$.*

For two probability measures μ and ν over $\mathbb{R}$, the L_1-Wasserstein distance can be rewritten as the L_1-norm between the cumulative distribution functions of the measures, $F_\mu : t \mapsto \mu((-\infty, t])$ and F_ν, or equivalently, as the L_1-norm between the quantile functions, $F_\mu^{-1} : s \mapsto \inf\{x \in \mathbb{R} \,|\, F(x) \geq s\}$ and F_ν^{-1}; see for instance [8, Theorem 2.9 and Theorem 2.10] and the references therein. Thus, for empirical measures on $\mathbb{R}$, its computation is easy. Its complexity is the same as the complexity of a sort. Then, the computation of the statistic and the subsampling distribution, see Section 2, will also be easy.

As mentioned above, the strategy we use to build the test is subsampling, which is close to bootstrap. Such methods were first introduced by Efron [22] in 1979, mainly to derive confidence intervals, but were often used since then even in the domain of topological data analysis for the function distance to a measure [14] ; see [41] for a main reference on asymptotic bootstrap and [3] for non-asymptotic bootstrap. But as aforementioned, the choice of $n = N$ is unsuccessful and bootstrap fails experimentally and theoretically at least for our choice of statistic. Just as Politis and Romano in [39], we decide alternatively to use only a small part of the sample to approximate the distribution of the statistic, although we also use only a small part of the points to build the statistic. Thus, this method relates to subsampling.

In order to prove that the distribution of the statistic and the subsampling distribution are close, we derive an upper-bound for the Wasserstein distance between the two. Such a method was already used in the paper [5]. It is then enough to prove the convergence of the distribution of the statistic to some continuous distribution to establish that our test is asymptotically of the proper level, meaning that it is valid.

The paper is organized as follows. In Section 2, we construct the statistical test and provide the main results of the paper. In there, we state assumptions under which the test is proven to be asymptotically of the correct level. As well, we derive some non-asymptotic bounds for the expectation of the L_1-Wasserstein distance between the distribution of the statistic and the subsampling distribution. We also provide a lower-bound for the power of the test. This lower bound depends on some discriminative quantity, a pseudo-distance between mm-spaces, which is studied in Section 3 in different contexts. In this section, the pseudo-distance is proven to be bounded above by the Gromov-Wasserstein and by L_1-Wasserstein distances. Thus, the statistical test is stable under Wasserstein noise. In Section 4, we propose an algorithm to implement the test. Moreover, some numerical experiments illustrate the fact that our method works. We give an example for which our method even performs better than some other method. Finally, in Section 5 we expose three ideas of

isomorphism testing methods linked to our test, and show that one does not work at all whereas the other two could lead to major improvements.

2. Main results

Let $(\mathcal{X}, \delta, \mu)$ and $(\mathcal{Y}, \gamma, \nu)$ be two mm-spaces.

In this section, we build statistical tests of the null hypothesis

$$H_0 \text{ “The mm-spaces } (\mathcal{X}, \delta, \mu) \text{ and } (\mathcal{Y}, \gamma, \nu) \text{ are isomorphic”},$$

against its alternative:

$$H_1 \text{ “The mm-spaces } (\mathcal{X}, \delta, \mu) \text{ and } (\mathcal{Y}, \gamma, \nu) \text{ are not isomorphic”}.$$

The statistical tests we propose are based on the observation of two samples, an N-sample from μ and an N'-sample from ν. To simplify notation, we assume that $N' = N$, but the methods proposed also work when N' is different from N. More importantly, we have to keep the same n in both cases, as defined below.

Given an N-sample $X_1, X_2, \dots X_N$ from the measure μ, we denote $\hat{\mu}_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_i}$ and $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$ for some $n \leq N$. As well, we define $\hat{\nu}_N$ and $\hat{\nu}_n$ from an N-sample from ν.
We recall that $\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_n)$ is the discrete distribution $\frac{1}{n}\sum_{i=1}^{n} \delta_{\mathrm{d}_{\hat{\mu}_N,m}(X_i)}$, and that we compare signatures with W_1, the L_1-Wasserstein distance.

The **test statistic** is then defined as

$$T_{N,n,m}(\mu,\nu) = \sqrt{n} W_1\left(\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_n), \mathrm{d}_{\hat{\nu}_N,m}(\hat{\nu}_n)\right),$$

and its distribution is denoted by $\mathcal{L}_{N,n,m}(\mu,\nu)$.

Note that for two isomorphic mm-spaces $(\mathcal{X}, \delta, \mu)$ and $(\mathcal{Y}, \gamma, \nu)$, the following distributions $\mathcal{L}_{N,n,m}(\mu,\mu)$, $\mathcal{L}_{N,n,m}(\nu,\nu)$, and $\frac{1}{2}\mathcal{L}_{N,n,m}(\mu,\mu) + \frac{1}{2}\mathcal{L}_{N,n,m}(\nu,\nu)$ are equal. The notation $\frac{1}{2}\mathcal{L}_1 + \frac{1}{2}\mathcal{L}_2$ stands for the distribution of a random variable that is generated according to $\mathcal{L}_1$ with probability $\frac{1}{2}$ and according to $\mathcal{L}_2$ with probability $\frac{1}{2}$. The three aforementioned distributions correspond to the distribution of the test statistic $T_{N,n,m}(\mu,\nu)$; see Lemma C.1 in the Appendix.

For some $\alpha \in (0,1)$, we denote by $\mathrm{q}_{1-\alpha} = \inf\{x \in \mathbb{R} \,|\, F(x) \geq 1-\alpha\}$, the $1-\alpha$**-quantile** of a distribution with cumulative distribution function F.
The $1-\alpha$-quantile $\mathrm{q}_{1-\alpha,N,n,m}$ of $\frac{1}{2}\mathcal{L}_{N,n,m}(\mu,\mu) + \frac{1}{2}\mathcal{L}_{N,n,m}(\nu,\nu)$ will be approximated by the $1-\alpha$-quantile $\hat{\mathrm{q}}_{1-\alpha,N,n,m}$ of $\frac{1}{2}\mathcal{L}^*_{N,n,m}(\hat{\mu}_N,\hat{\mu}_N) + \frac{1}{2}\mathcal{L}^*_{N,n,m}(\hat{\nu}_N,\hat{\nu}_N)$. Here $\mathcal{L}^*_{N,n,m}(\hat{\mu}_N,\hat{\mu}_N)$ stands for the distribution of $\sqrt{n} W_1\left(\mathrm{d}_{\hat{\mu}_N,m}(\mu_n^*), \mathrm{d}_{\hat{\mu}_N,m}(\mu_n'^*)\right)$ conditionally to $\hat{\mu}_N$, where μ_n^* and $\mu_n'^*$ are two empirical measures from independent n-samples from $\hat{\mu}_N$.

The **test** we deal with in this paper is then

$$\boxed{\phi_{N,n,m} = \mathbb{1}_{T_{N,n,m}(\mu,\nu) \geq \hat{\mathrm{q}}_{1-\alpha,N,n,m}}}.$$

The null hypothesis H_0 is rejected if $\phi_{N,n,m} = 1$, that is if the L_1-Wasserstein distance between the two empirical signatures $\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_n)$ and $\mathrm{d}_{\hat{\nu}_N,m}(\hat{\nu}_n)$ is too high.

Note that it is equivalent to compute a **p-value** $\hat{p}_{N,n,m}$ from the subsampling distribution and the test statistic:

$$\hat{p}_{N,n,m} = 1 - F^*(T_{N,n,m}(\mu,\nu)),$$

with $F^*(t) = \mathbb{P}(T \leq t)$ for T a random variable from the distribution $\frac{1}{2}\mathcal{L}^*_{N,n,m}(\hat{\mu}_N,\hat{\mu}_N) + \frac{1}{2}\mathcal{L}^*_{N,n,m}(\hat{\nu}_N,\hat{\nu}_N)$.

The statistical test consists in rejecting the hypothesis H_0 if the p-value is not larger than α, that is

$$\boxed{\phi_{N,n,m} = \mathbb{1}_{\hat{p}_{N,n,m} \leq \alpha}}.$$

2.1. A test of asymptotic level α

In this section, we consider two isomorphic mm-spaces $(\mathcal{X},\delta,\mu)$ and $(\mathcal{Y},\gamma,\nu)$ (we may write $(\mu,\nu) \sim H_0$). In order to assert the validity of the test $\phi_{N,n,m}$, the probability $\mathbb{P}_{(\mu,\nu)\sim H_0}(\phi_{N,n,m} = 1)$ of rejecting H_0 must be bounded above by α. In this section, under mild assumptions, we prove that the test $\phi_{N,n,m}$ is of asymptotic level α, that is such that

$$\limsup_{N\to\infty} \mathbb{P}_{(\mu,\nu)\sim H_0}(\phi_{N,n,m} = 1) \leq \alpha.$$

We will prove (Lemma 2.1 and Lemma 2.2) that the test is of asymptotic level α when the distribution $\mathcal{L}_{N,n,m}(\mu,\mu)$ converges weakly to some atomless distribution $\tilde{\mathcal{L}}$, and when its approximation $\mathcal{L}^*_{N,n,m}(\hat{\mu}_N,\hat{\mu}_N)$ from the sample satisfies that $W_1(\mathcal{L}^*_{N,n,m}(\hat{\mu}_N,\hat{\mu}_N),\tilde{\mathcal{L}})$ converges in probability to 0.

Let $\mathbb{G}_{\mu,m}$ and $\mathbb{G}'_{\mu,m}$ be two independent Gaussian processes with covariance kernel $\kappa(s,t) = F_{\mathrm{d}_{\mu,m}(\mu)}(s)\left(1 - F_{\mathrm{d}_{\mu,m}(\mu)}(t)\right)$ for $s \leq t$, with $F_{\mathrm{d}_{\mu,m}(\mu)}$ the cumulative distribution function of $\mathrm{d}_{\mu,m}(\mu)$. The limit distribution $\tilde{\mathcal{L}}$ is actually given by the distribution of $\|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1$, the integral on $\mathbb{R}$ of the absolute value of the difference between the two Gaussian processes. When the distributions μ and ν are compactly supported, these convergences occur under assumptions outlined in the following theorem.

Theorem 2.1. *Let $(\mathcal{X},\delta,\mu)$ and $(\mathcal{Y},\gamma,\nu)$ be two mm-spaces, with μ and ν compactly supported. Let n be such that $\frac{n}{N} = o(1)$ and assume that when N goes to infinity, $\sqrt{n}\mathbb{E}[\|\mathrm{d}_{\mu,m} - \mathrm{d}_{\hat{\mu}_N,m}\|_{\infty,\mathcal{X}}]$ goes to 0.*

Then, if $\mathcal{L}(\|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1)$ is atomless, the statistical test

$$\phi_{N,n,m} = \mathbb{1}_{\sqrt{n}W_1(\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_n),\mathrm{d}_{\hat{\nu}_N,m}(\hat{\nu}_n)) \geq \hat{\mathrm{q}}_{1-\alpha,N,n,m}}$$

is of asymptotic level α.

The first assumption of Theorem 2.1 can be readily checked in some specific case. Among the distributions supported on compact subsets of $\mathbb{R}^d$, Theorem 2.2 deals with distributions that are regular in the following sense. We say that a measure μ is (a,b)-**standard** with positive parameters a and b, if for any positive radius r and any point x of the support of μ, we have that $\mu(\mathrm{B}(x,r)) \geq \min\{1, ar^b\}$, with $\mathrm{B}(x,r) = \{y \in \mathcal{X} \mid \delta(x,y) < r\}$. The assumption of (a,b)-standardness has been widely used in the context of set estimation and

Topological Data Analysis [18, 20, 21, 17, 15, 24]. Uniform measures on open subsets of $\mathbb{R}^d$ are particular cases of such regular distributions:

Example 2.1. *Let O be a non-empty bounded open subset of $\mathbb{R}^d$. Then, the uniform measure on O, μ_O is (a,d)-standard with*

$$a = \frac{\omega_d}{\mathrm{Leb}_d(O)} \left(\frac{\mathrm{Reach}(O)}{\mathcal{D}(O)} \right)^d .$$

Here, $\mathcal{D}(O)$ stands for the diameter of O, ω_d for $\mathrm{Leb}_d(\mathrm{B}(0,1))$, the Lebesgue volume of the unit d-dimensional ball, and $\mathrm{Reach}(O)$ is the reach of the open set O as defined in Section 3.2.3.

Proof. Proof in the Appendix, in Section A.1. □

Uniform measures on regular compact submanifolds are also standard. In [37] (Lemma 5.3), the authors give a bound for a depending on the reach of the submanifold. More generally, distributions supported on regular open subsets of $\mathbb{R}^d$ or manifolds, with density bounded from below by some positive constant, are (a,b)-standard distributions.

For compactly-supported distributions on $\mathbb{R}^d$ and among them, for (a,b)-standard distributions, assumptions of Theorem 2.1 are satisfied as soon as n remains small enough with respect to N, as follows.

Theorem 2.2. *Let μ and ν be two Borel probability measures supported on compact subsets of $\mathbb{R}^d$. We set $N = cn^\rho$ with some $c > 0$ and $\rho > 1$.*

The statistical test

$$\phi_{N,n,m} = \mathbb{1}_{\sqrt{n} W_1(\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_n), \mathrm{d}_{\hat{\nu}_N,m}(\hat{\nu}_n)) \geq \hat{\mathrm{q}}_{1-\alpha,N,n,m}}$$

is of asymptotic level α

- *in the general case, if $\rho > \frac{\max\{d,2\}}{2}$,*
- *in the (a,b)-standard case, if $\rho > 1$,*

with the additional assumption that $\mathcal{L}(\|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1)$ is atomless.

Morally, checking the assumption "$\mathcal{L}(\|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1)$ is atomless" boils down to verify the following assumption:

$$A_m : \text{“}\mathrm{d}_{\mu,m}(\mu) \text{ is not a Dirac mass”}.$$

The reason of this assertion is the following. The process $(\mathbb{G}_{\mu,m}(t) - \mathbb{G}'_{\mu,m}(t))_{t\in\mathbb{R}}$ is a Gaussian process with covariance kernel given by $2F_{\mathrm{d}_{\mu,m}(\mu)}(s)\left(1 - F_{\mathrm{d}_{\mu,m}(\mu)}(t)\right)$ for $s \leq t$. Set I, the smallest interval containing the support of the measure $\mathrm{d}_{\mu,m}(\mu)$. Note that for every $t \notin I$, $\mathbb{G}_{\mu,m}(t) - \mathbb{G}'_{\mu,m}(t) = 0$ since then its variance $2F_{\mathrm{d}_{\mu,m}(\mu)}(t)\left(1 - F_{\mathrm{d}_{\mu,m}(\mu)}(t)\right)$ is equal to 0. As a consequence, $\|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1$ is the integral of $|\mathbb{G}_{\mu,m}(t) - \mathbb{G}'_{\mu,m}(t)|$ over the interval I. This interval is reduced to a single point when $\mathrm{d}_{\mu,m}(\mu)$ is a Dirac mass. In this case, $\|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1$ is constant equal to 0. This interval is non trivial as soon as $\mathrm{d}_{\mu,m}(\mu)$ is not a Dirac mass. In this case, $\mathcal{L}(\|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1)$ is atomless, as the distribution of the integral of continuous random variables. A rigorous proof of this intuitive result is out of the scope of this paper. Even so, it should be noted that $(\mathbb{G}_{\mu,m}(t) - \mathbb{G}'_{\mu,m}(t))_{t\in\mathbb{R}}$ has

the same distribution as $\left(\sqrt{2}B(F_{\mathrm{d}_{\mu,m}(\mu)}(t))\right)_{t\in\mathbb{R}}$, where $(B(s))_{s\in[0,1]}$ is the Brownian bridge. Continuity of $\|B\|_1$ follows from former work in the literature. Indeed, Johnson and Killeen [31] derived an expression for the cumulative distribution function of $\mathcal{L}(\|B\|_1)$ that depends on the Airy function and is continuous, and Rice [32] derived an expression for its density.

The measures μ for which assumption A_m is satisfied for no $m \in [0,1]$ are such that $\mathrm{d}_{\mu,m}(x) = \mathrm{d}_{\mu,m}(y)$ for every $x, y \in \mathrm{Supp}(\mu)$ and $m \in [0,1]$, that is:

$$\forall x, y \in \mathrm{Supp}(\mu),\ \forall r \geq 0,\ \mu(\overline{\mathrm{B}}(x,r)) = \mu(\overline{\mathrm{B}}(y,r)). \tag{2.1}$$

For instance, discrete distributions for which A_m is never satisfied are exactly the distributions that are uniform on a finite set of points (Equation (2.1) with $r = 0$) and that satisfy that: "For every $x, y \in \mathrm{Supp}(\mu)$, for every $d > 0$, the number of points in $\mathrm{Supp}(\mu)$ at distance d to y is the same as the number of points in $\mathrm{Supp}(\mu)$ at distance d to x". Another example of measure μ such that A_m is never satisfied is given by any uniform distribution on a sphere in $\mathbb{R}^d$. In this case, (2.1) is trivially satisfied.

Characterizing all distributions for which A_m is not satisfied for some fixed $m \in [0,1]$ is not simple. The examples are more abundant than the above-mentioned ones. For instance, $\mu = 0.2\delta_0 + 0.2\delta_1 + 0.3\delta_5 + 0.3\delta_7$ does not satisfy $A_{0.4}$. For such examples, the isomorphism test on two samples from μ with parameter m will not work. Nonetheless, it is possible to circumvent this issue by modifying the datasets, as follows.

Set $\mathcal{D}$, a continuous and compactly-supported isotropic distribution on $\mathbb{R}^d$; for instance, the restriction of the normal distribution $\mathcal{N}(0,1)$ to a ball centered at 0. Then, the convolution of two isomorphic distributions μ and ν (in $\mathbb{R}^d$) with $\mathcal{D}$ will be isomorphic and will satisfy assumption A_m for every $m \in (0,1)$. For practice, this strategy consists in replacing the sample $(X_i)_{i\in[\![1,N]\!]}$ from μ with the sample $(X_i + Z_i)_{i\in[\![1,N]\!]}$ for $Z_i \sim \mathcal{D}$ independent random variables, independent from $(X_i)_{i\in[\![1,N]\!]}$. For the testing purpose, the same procedure should be applied to the sample from ν.

As aforementioned, the proof of Theorem 2.1 consists in showing that the distribution of the test statistic

$$\frac{1}{2}\mathcal{L}_{N,n,m}(\mu,\mu) + \frac{1}{2}\mathcal{L}_{N,n,m}(\nu,\nu)$$

under the hypothesis H_0 converges weakly to the fixed distribution

$$\mathcal{L} = \frac{1}{2}\mathcal{L}(\|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1) + \frac{1}{2}\mathcal{L}(\|\mathbb{G}_{\nu,m} - \mathbb{G}'_{\nu,m}\|_1)$$

when n and N go to ∞. But also, that the L_1-Wasserstein metric between the subsampling distribution

$$\frac{1}{2}\mathcal{L}^*_{N,n,m}(\hat{\mu}_N,\hat{\mu}_N) + \frac{1}{2}\mathcal{L}^*_{N,n,m}(\hat{\nu}_N,\hat{\nu}_N)$$

and $\mathcal{L}$ converges to 0 in probability.

Note that it is sufficient to prove these convergences for $\mathcal{L}_{N,n,m}(\mu,\mu)$ and $\mathcal{L}^*_{N,n,m}(\hat{\mu}_N,\hat{\mu}_N)$. Indeed, the L_1-Wasserstein distance W_1 is a metric for weak convergence and satisfies that for any distributions $\mathcal{L}_1$, $\mathcal{L}'_1$, $\mathcal{L}_2$ and $\mathcal{L}'_2$, $W_1(\frac{1}{2}\mathcal{L}_1 + \frac{1}{2}\mathcal{L}'_1, \frac{1}{2}\mathcal{L}_2 + \frac{1}{2}\mathcal{L}'_2) \leq \frac{1}{2}W_1(\mathcal{L}_1,\mathcal{L}'_1) + \frac{1}{2}W_1(\mathcal{L}_2,\mathcal{L}'_2)$. This is a straightforward consequence of the definition of the L_1-Wasserstein distance with transport plans.

Then, Theorem 2.1 follows from the following two lemmas.

Lemma 2.1. *For μ a measure supported on a compact set, we choose n as a function of N such that: $\frac{n}{N} = o(1)$ and $\sqrt{n}\mathbb{E}[\|\mathrm{d}_{\mu,m} - \mathrm{d}_{\hat{\mu}_N,m}\|_{\infty,\mathcal{X}}]$ goes to zero or more specifically $\frac{\sqrt{n}}{m}\mathbb{E}[W_1(\mu,\hat{\mu}_N)]$ goes to zero when N goes to ∞. Then we have that*

$$\mathcal{L}_{N,n,m}(\mu,\mu) \rightsquigarrow \mathcal{L}\left(\|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1\right), \tag{2.2}$$

and

$$W_1\left(\mathcal{L}^*_{N,n,m}(\hat{\mu}_N,\hat{\mu}_N), \mathcal{L}\left(\|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1\right)\right) \to 0 \text{ in probability,} \tag{2.3}$$

when N goes to infinity.

Proof. Proof in the Appendix, in Section C.3. □

Convergence of the distribution of the statistic and convergence of its approximation via subsampling to a fixed distribution is not sufficient to prove that the test has the correct level. Continuity of the limit distribution $\mathcal{L}$ is also required.

Lemma 2.2. *If the two convergences in Lemma 2.1 occur, and if the $1-\alpha$-quantile $\mathrm{q}_{1-\alpha}$ of the distribution $\frac{1}{2}\mathcal{L}(\|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1) + \frac{1}{2}\mathcal{L}(\|\mathbb{G}_{\nu,m} - \mathbb{G}'_{\nu,m}\|_1)$ is a point of continuity of its cumulative distribution function, then the asymptotic level of the test at (μ,ν) is α.*

Proof. Proof in the Appendix, in Section C.3. □

Note that under hypothesis H_0, the distributions $\frac{1}{2}\mathcal{L}(\|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1) + \frac{1}{2}\mathcal{L}(\|\mathbb{G}_{\nu,m} - \mathbb{G}'_{\nu,m}\|_1)$ and $\mathcal{L}(\|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1)$ are equal. Thus, the second assumption of Lemma 2.2 may be replaced by "the $1-\alpha$-quantile of the distribution $\mathcal{L}(\|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1)$ is a point of continuity of its cumulative distribution function".

Theorem 2.2 stems from the following Proposition 2.1 and Proposition 2.2. Moreover, in these propositions, upper bounds for the expectation of $W_1\left(\mathcal{L}_{N,n,m}(\mu,\mu), \mathcal{L}^*_{N,n,m}(\hat{\mu}_N,\hat{\mu}_N)\right)$ are provided for compactly-supported distributions on $\mathbb{R}^d$, and among them, for the (a,b)-standard ones. Somehow, these bounds aim at quantifying the performance of the test when the sample size N goes to ∞.

2.1.1. The case of measures supported on a compact subset of $\mathbb{R}^d$

Set $N = cn^\rho$ for some positive constants ρ and c. Then the test is asymptotically valid for two measures supported on a compact subset of the Euclidean space $\mathbb{R}^d$ if we assume that $\rho > \frac{\max\{d,2\}}{2}$.

Proposition 2.1. *Let μ be some Borel probability measure supported on some compact subset of $\mathbb{R}^d$. Under the assumption*

$$\rho > \frac{\max\{d,2\}}{2},$$

the two convergences of Lemma 2.1 occur.

Moreover, a bound for the expectation of $W_1\left(\mathcal{L}_{N,n,m}(\mu,\mu), \mathcal{L}^*_{N,n,m}(\hat{\mu}_N,\hat{\mu}_N)\right)$ is of order:

$$N^{\frac{1}{2\rho} - \frac{1}{\max\{d,2\}}} \left(\log(1+N)\right)^{\mathbb{1}_{d=2}}.$$

Furthermore, $W_1\left(\mathcal{L}_{N,n,m}(\mu,\mu), \mathcal{L}^*_{N,n,m}(\hat{\mu}_N,\hat{\mu}_N)\right) \to 0$ a.e. when n goes to ∞.

Proof. This proposition is based on rates of convergence for the Wasserstein distance between a measure μ on $\mathbb{R}^d$ and its empirical version $\hat{\mu}_N$; see [26] for general dimensions and [8] for $d = 1$. Proof in the Appendix, in Section C.4. □

2.1.2. The case of (a, b)-standard measures supported on a compact subset of $\mathbb{R}^d$

The test is asymptotically valid for two (a, b)-standard measures supported on compact connected subsets of $\mathbb{R}^d$ provided that $\rho > 1$:

Proposition 2.2. *Let μ be an (a, b)-standard measure supported on a connected compact subset of $\mathbb{R}^d$. The two convergences of Lemma 2.1 occur if the assumption $\rho > 1$ is satisfied. Moreover, a bound for the expectation of $W_1\left(\mathcal{L}_{N,n,m}(\mu, \mu), \mathcal{L}^*_{N,n,m}(\hat{\mu}_N, \hat{\mu}_N)\right)$ is of order $N^{\frac{1}{2\rho}-\frac{1}{2}}$ up to a logarithmic term.*

Proof. This proposition is based on rates of convergence for the infinity norm between the distance to a measure and its empirical version; see [18]. Proof in the Appendix, in Section C.5. □

Note that we can achieve a rate close to the parametric rate for standard measures, whereas for general measures, the rate gets worse when the dimension increases. In any case, we need ρ to be as large as possible for the subsampling distribution to be a good enough approximation of the distribution of the statistic, that is, to have a type I error close enough to α, keeping in mind that n should go to ∞ with N.

2.2. The power of the test

The **power** of the test $\phi_{N,n,m} = \mathbb{1}_{\sqrt{n}W_1\left(\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_n), \mathrm{d}_{\hat{\nu}_N,m}(\hat{\nu}_n)\right) \geq \hat{\mathrm{q}}_{1-\alpha,N,n,m}}$ is defined for two mm-spaces $(\mathcal{X}, \delta, \mu)$ and $(\mathcal{Y}, \gamma, \nu)$ by

$$1 - \mathbb{P}_{(\mu,\nu)}\left(\phi_{N,n,m} = 0\right),$$

where $\mathbb{P}_{(\mu,\nu)}\left(\phi_{N,n,m} = 0\right)$ stands for the probability that $\phi_{N,n,m} = 0$ when the test is built from samples from two general mm-spaces $(\mathcal{X}, \delta, \mu)$ and $(\mathcal{Y}, \gamma, \nu)$.

If the spaces are not isomorphic, we want the test to reject H_0 with high probability. It means that we want the power to be as large as possible. Here, we give a lower bound for the power, or more precisely an upper bound for $\mathbb{P}_{(\mu,\nu)}\left(\phi_{N,n,m} = 0\right)$, the **type II error**.

Theorem 2.3. *Let μ and ν be two Borel measures supported on $\mathcal{X}$ and $\mathcal{Y}$, two compact subsets of $\mathbb{R}^d$. We assume that the mm-spaces $(\mathcal{X}, \delta, \mu)$ and $(\mathcal{Y}, \gamma, \nu)$ are non-isomorphic and that the DTM-signature is discriminative for some m in $(0, 1]$, meaning that the pseudo-metric $W_1\left(\mathrm{d}_{\mu,m}(\mu), \mathrm{d}_{\nu,m}(\nu)\right)$ is positive. We choose $N = n^\rho$ with $\rho > 1$. Then for all positive ϵ, there exists N_0 depending on μ and ν such that for all $N \geq N_0$, the type II error*

$$\mathbb{P}_{(\mu,\nu)}\left(\sqrt{n}W_1\left(\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_n), \mathrm{d}_{\hat{\nu}_N,m}(\hat{\nu}_n)\right) < \hat{\mathrm{q}}_{1-\alpha,N,n,m}\right)$$

is bounded above by

$$4\exp\left(-\frac{W_1^2\left(\mathrm{d}_{\mu,m}(\mu), \mathrm{d}_{\nu,m}(\nu)\right)}{(2+\epsilon)\max\left\{\mathcal{D}^2_{\mu,m}, \mathcal{D}^2_{\nu,m}\right\}} N^{\frac{1}{\rho}}\right),$$

with $\mathcal{D}_{\mu,m}$ the diameter of the support of the measure $\mathrm{d}_{\mu,m}(\mu)$.

Proof. Proof in the Appendix, in Section C.6. □

In order to have a high power, that is to reject H_0 more often when the mm-spaces are not isomorphic, we need n to be big enough, that is ρ small enough. Recall that n has to be small enough for the law of the statistic and its subsampling version to be close. This means that some compromise must be made. Moreover, the choice of m for the test should depend on the geometry of the mm-spaces. The tuning of these parameters from the data is still an open question.

Moreover, note that the power of the test is strongly related to $\frac{W_1(\mathrm{d}_{\mu,m}(\mu),\mathrm{d}_{\nu,m}(\nu))}{\max\left\{\mathcal{D}^2_{\mu,m},\mathcal{D}^2_{\nu,m}\right\}}$. The test is powerful when the pseudo-metric is high with respect to the diameters of the signatures supports and does not discriminate between measures when it is low. In the following section, we derive some upper-bounds and lower-bounds for the pseudo-metric $W_1\left(\mathrm{d}_{\mu,m}(\mu),\mathrm{d}_{\nu,m}(\nu)\right)$, under some geometric assumptions.

3. Stability and discriminative properties of the DTM-signatures

3.1. Stability of the DTM-signatures

In this section, we prove stability results for the DTM-signature. These results all rely on the stability of the distance-to-a-measure function itself.

Proposition 3.1 (Stability, in [13] for $\mathbb{R}^d$, in [9] for metric spaces)**.** *For two mm-spaces $(\mathcal{X},\delta,\mu)$ and $(\mathcal{Y},\delta,\nu)$ embedded into the same metric space, we have that*

$$\|\mathrm{d}_{\mu,m}-\mathrm{d}_{\nu,m}\|_{\infty,\mathcal{X}\cup\mathcal{Y}}\leq\frac{1}{m}W_1(\mu,\nu).$$

In [35], Mémoli proposes a metric on the quotient space of mm-spaces by the relation of isomorphism, the Gromov–Wasserstein distance.

Definition 3.1 (Gromov–Wasserstein distance)**.** *The **Gromov–Wasserstein distance** between two mm-spaces $(\mathcal{X},\delta,\mu)$ and $(\mathcal{Y},\gamma,\nu)$ denoted $GW(\mathcal{X},\mathcal{Y})$ is defined by the expression*

$$\inf_{\pi\in\Pi(\mu,\nu)}\frac{1}{2}\int_{\mathcal{X}\times\mathcal{Y}}\int_{\mathcal{X}\times\mathcal{Y}}\Gamma_{\mathcal{X},\mathcal{Y}}\left(x,y,x',y'\right)\pi(\mathrm{d}x\times\mathrm{d}y)\pi(\mathrm{d}x'\times\mathrm{d}y'),$$

*with $\Gamma_{\mathcal{X},\mathcal{Y}}(x,y,x',y')=|\delta(x,x')-\gamma(y,y')|$. Here $\Pi(\mu,\nu)$ stands for the set of **transport plans** between μ and ν, that is the set of Borel probability measures π on $\mathcal{X}\times\mathcal{Y}$ satisfying $\pi(A\times\mathcal{Y})=\mu(A)$ and $\pi(\mathcal{X}\times B)=\nu(B)$ for all Borel sets A in $\mathcal{X}$ and B in $\mathcal{Y}$.*

The DTM-signature turns out to be stable with respect to this Gromov-Wasserstein distance.

Proposition 3.2. *We have that:*

$$W_1(\mathrm{d}_{\mu,m}(\mu),\mathrm{d}_{\nu,m}(\nu))\leq\frac{1}{m}GW(\mathcal{X},\mathcal{Y}).$$

Proof. Proof in the Appendix, in Section B. The proof is relatively similar to the ones given by Mémoli in [35] for other signatures. □

It follows directly that two isomorphic mm-spaces have the same DTM-signature. Whenever the two mm-spaces are embedded into the same metric space, we also get stability with respect to the L_1-Wasserstein distance.

Proposition 3.3. *If $(\mathcal{X}, \delta, \mu)$ and $(\mathcal{Y}, \delta, \nu)$ are two metric measure spaces embedded into some metric space $(\mathcal{Z}, \delta)$, then we can bound $W_1(\mathrm{d}_{\mu,m}(\mu), \mathrm{d}_{\nu,m}(\nu))$ above by*

$$W_1(\mu,\nu) + \min\left\{\|\mathrm{d}_{\mu,m} - \mathrm{d}_{\nu,m}\|_{\infty,\mathrm{Supp}(\mu)}, \|\mathrm{d}_{\mu,m} - \mathrm{d}_{\nu,m}\|_{\infty,\mathrm{Supp}(\nu)}\right\},$$

and more generally by

$$\left(1 + \frac{1}{m}\right) W_1(\mu,\nu).$$

Proof. First notice that:

$$\begin{aligned} W_1(\mathrm{d}_{\mu,m}(\mu), \mathrm{d}_{\nu,m}(\mu)) &\leq \int_{\mathcal{X}} |\mathrm{d}_{\mu,m}(x) - \mathrm{d}_{\nu,m}(x)| \, \mathrm{d}\mu(x) \\ &\leq \|\mathrm{d}_{\mu,m} - \mathrm{d}_{\nu,m}\|_{\infty,\mathrm{Supp}(\mu)}. \end{aligned}$$

Then, for all π in $\Pi(\mu,\nu)$:

$$W_1(\mathrm{d}_{\nu,m}(\mu), \mathrm{d}_{\nu,m}(\nu)) \leq \int_{\mathcal{X}\times\mathcal{Y}} |\mathrm{d}_{\nu,m}(x) - \mathrm{d}_{\nu,m}(y)| \, \mathrm{d}\pi(x,y).$$

Thus, since $\mathrm{d}_{\nu,m}$ is 1-Lipschitz:

$$W_1(\mathrm{d}_{\nu,m}(\mu), \mathrm{d}_{\nu,m}(\nu)) \leq W_1(\mu,\nu).$$

Then, the result follows from Proposition 3.1. □

According to [8], a measure μ and its empirical version $\hat{\mu}_N$ are close in terms of the Wasserstein metric W_1. As a consequence, Proposition 3.3 entails that the signature $\mathrm{d}_{\mu,m}(\mu)$ and the empirical signature $\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_N)$ are close when N gets large. This property is essential for the purpose of testing. The signature $\delta_{\mu,m}(\mu)$, built from the pseudo-metric $\delta_{\mu,m}$, does not satisfy this stability property. Indeed, for $m < \frac{1}{2} - \epsilon$, the measures $\mu = (m+\epsilon)\delta_0 + (1-m-\epsilon)\delta_1$ and $\nu = (m-\epsilon)\delta_0 + (1-m+\epsilon)\delta_1$ are very close in terms of W_1, nonetheless, their signatures $\delta_{\mu,m}(\mu) = \delta_0$ and $\delta_{\nu,m}(\nu) = (1-m+\epsilon)\delta_0 + (m-\epsilon)\delta_1$ are very different. This explains why, in this paper, we consider the DTM $\mathrm{d}_{\mu,m}$ and not the pseudo-metric $\delta_{\mu,m}$.

3.2. Discriminative properties of the DTM-signatures

The DTM-signature is stable but unfortunately does not always discriminate between mm-spaces. Indeed, in the following counter-example from [35] (example 5.6), there are two non-isomorphic mm-spaces sharing the same signatures for all values of m.

Example 3.1. *We consider two graphs made of 9 vertices each, clustered in three groups of 3 vertices, such that each vertex is at distance 1 exactly from each vertex of its group and at distance 2 from any other vertex. We assign a mass to each vertex; the distribution is the following, for the first graph (Figure 1):*

$$\mu = \left\{\left(\frac{23}{140}, \frac{1}{105}, \frac{67}{420}\right), \left(\frac{3}{28}, \frac{1}{28}, \frac{4}{21}\right), \left(\frac{2}{15}, \frac{1}{15}, \frac{2}{15}\right)\right\},$$

and for the second graph (Figure 2):

$$\nu = \left\{\left(\frac{3}{28}, \frac{1}{15}, \frac{67}{420}\right), \left(\frac{2}{15}, \frac{4}{21}, \frac{1}{105}\right), \left(\frac{23}{140}, \frac{2}{15}, \frac{1}{28}\right)\right\}.$$

The mm-spaces ensuing are not isomorphic since any one-to-one and onto measure-preserving map would send at least one pair of vertices at distance 1 from each other to a pair of vertices at distance 2 from each other, and thus it would not be an isometry.

Moreover, note that the DTM-signatures associated to the graphs are equal since the total mass of each cluster is exactly equal to $\frac{1}{3}$.

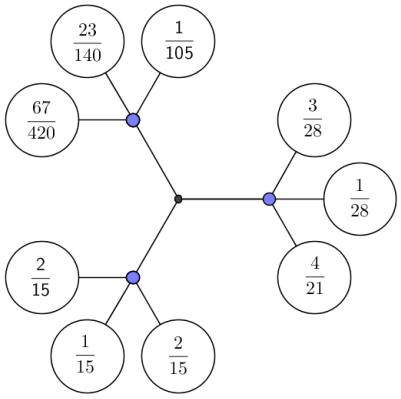


FIGURE 1. μ

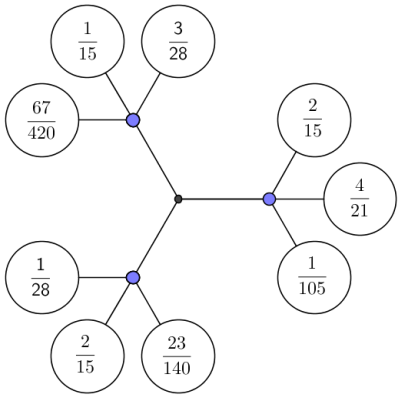


FIGURE 2. ν

Nevertheless, the signature can be discriminative in some cases. In the following, we give lower bounds for the L_1-Wasserstein distance between two signatures under different alternatives.

We will prove in Proposition 3.4 that this pseudo-distance is proportional to $|1-\lambda|$ when we consider a metric space $(\mathcal{X}, \delta, \mu)$ and its dilatation $(\mathcal{X}, \lambda\delta, \mu)$ with a factor $\lambda > 0$. More generally, it is possible to discriminate between two uniform distributions that are supported on compact subsets of $\mathbb{R}^d$, with different Lebesgue volumes $Leb_d(O)$ and $Leb_d(O')$. In Proposition 3.5 we provide a lower bound for the distance between such signatures, that is proportional to $|Leb_d(O)^{\frac{1}{d}} - Leb_d(O')^{\frac{1}{d}}|$. Note that this bound is kind of optimal, since we recover the factor $|1-\lambda|$ of Proposition 3.4 when O' is the image of O with a dilatation of parameter λ.

Moreover, two uniform distributions on compact sets with the same Lebesgue volume might also have different signatures, as enhanced by Example 3.3. Indeed, according to Proposition 3.6, whenever the inner offsets $O_\epsilon = \{x \in O \mid \inf_{y\in\partial O} \|x-y\|_2 \geq \epsilon\}$ and O'_ϵ have different Lebesgue volume for some $\epsilon > 0$, the signatures associated with some parameter m, depending on ϵ, will be different. Consequently, it is possible to discriminate between a "thin" subset of $\mathbb{R}^d$ or a set with not regular boundary and a "fat" set or a set with a regular boundary (for instance, a set with a large reach, as defined in Section 3.2.3) such as a ball.

The signatures are also sensitive to the density of distributions. A distribution with density bounded above by some constant C (a uniform distribution for instance) can be discriminated from distributions which density is larger than C on some sets that are large enough. Lower bounds for the distance between such signatures are derived in Proposition 3.7, Proposition 3.8 and Proposition 3.9. The proofs are based on the fact that two signatures $\mathrm{d}_{\mu,m}(\mu)$ and $\mathrm{d}_{\nu,m}(\nu)$ are different whenever the set $\{x \mid \mathrm{d}_{\nu,m}(x) > \inf_{x\in\mathrm{Supp}(\mu)} \mathrm{d}_{\mu,m}(x)\}$ has

positive ν-measure. One might use such a strategy to prove that two signatures are different in many other situations than the following situations handled in this paper.

3.2.1. When the distances are multiplied by some positive real number λ

Let λ be some positive real number. The DTM-signature discriminates between two mm-spaces isomorphic up to a dilatation of parameter λ, for $\lambda \neq 1$.

Proposition 3.4. *Let $(\mathcal{X}, \delta, \mu)$ and $(\mathcal{Y}, \gamma, \nu) = (\mathcal{X}, \lambda\delta, \mu)$ be two mm-spaces. We have*

$$W_1(\mathrm{d}_{\mu,m}(\mu), \mathrm{d}_{\nu,m}(\nu)) = |1-\lambda|\, \mathbb{E}_\mu[\mathrm{d}_{\mu,m}(X)],$$

for X a random variable of law μ.

Proof. First notice that $F^{-1}_{\mathrm{d}_{\nu,m}(\nu)} = \lambda F^{-1}_{\mathrm{d}_{\mu,m}(\mu)}$. Then,

$$\begin{aligned} W_1(\mathrm{d}_{\mu,m}(\mu), \mathrm{d}_{\nu,m}(\nu)) &= \int_0^1 \left| F^{-1}_{\mathrm{d}_{\mu,m}(\mu)}(s) - F^{-1}_{\mathrm{d}_{\nu,m}(\nu)}(s) \right| \mathrm{d}s \\ &= |1-\lambda| \int_0^1 \left| F^{-1}_{\mathrm{d}_{\mu,m}(\mu)}(s) \right| \mathrm{d}s \\ &= |1-\lambda|\, \mathbb{E}_\mu\left[\mathrm{d}_{\mu,m}(X)\right]. \end{aligned}$$

□

3.2.2. The case of uniform measures on non-empty bounded open subsets of $\mathbb{R}^d$

The DTM-signature discriminates between two uniform measures over two non-empty bounded open subsets of $\mathbb{R}^d$ with different Lebesgue volume, provided that m is small enough.

Proposition 3.5. *Let $(O, \|\cdot\|_2, \mu_O)$ and $(O', \|\cdot\|_2, \mu_{O'})$ be two mm-spaces, for O and O' two non-empty bounded open subsets of $\mathbb{R}^d$ satisfying $O = \left(\overline{O}\right)^\circ$ and $O' = \left(\overline{O'}\right)^\circ$. Here, $\overline{O}$ stands for the closure of O, C° for the interior of a set C, and $\|.\|_2$ for the Euclidean norm. The measure μ_O refers to the uniform measure on the open set O with respect to the Lebesgue measure Leb_d on $\mathbb{R}^d$, that is, $\mu_O(B) = \frac{\mathrm{Leb}_d(B\cap O)}{\mathrm{Leb}_d(O)}$ for all Borel set B.*

If $\mathrm{Leb}_d(O) \leq \mathrm{Leb}_d(O')$, then, a lower bound for $W_1(\mathrm{d}_{\mu_O,m}(\mu_O), \mathrm{d}_{\mu_{O'},m}(\mu_{O'}))$ is given by

$$\mu_O\left(O_{\epsilon(m,O)}\right) \frac{d}{d+1} \left(\frac{m}{\omega_d}\right)^{\frac{1}{d}} \left(\mathrm{Leb}_d(O')^{\frac{1}{d}} - \mathrm{Leb}_d(O)^{\frac{1}{d}}\right).$$

Here, $O_\epsilon = \{x \in O \,|\, \inf_{y\in\partial O} \|x-y\|_2 \geq \epsilon\}$ with ∂O the boundary of O, $\epsilon(m,O) = \left(\frac{m\mathrm{Leb}_d(O)}{\omega_d}\right)^{\frac{1}{d}}$ is the radius of any ball $B \subset O$ such that $\mu_O(B) = m$, and ω_d equals $\mathrm{Leb}_d(\mathrm{B}(0,1))$, the Lebesgue volume of the unit d-dimensional ball.

Proof. Proof in the Appendix, in Section A.2. □

This proposition can be applied to a simple case.

Example 3.2. *Let O be the open unit ball in $\mathbb{R}^d$, and O' the hypercube of radius 1 in $\mathbb{R}^d$. Then, the DTM-signature discriminates between μ_O and $\mu_{O'}$ whenever*

$$m < \frac{\omega_d}{2^d}.$$

Proof. Proof in the Appendix, in Section A.2. □

Proposition 3.5 states that two uniform distributions on compact sets with different Lebesgue volume have different signatures. Nonetheless, when these Lebesgue volumes are equal, it might also be possible to detect difference between the measures by considering the volumes of the inner offsets.

Proposition 3.6. *Set O, a subset of $\mathbb{R}^d$. Then, the set of points x in $\mathbb{R}^d$ for which $\mathrm{d}_{\mu_O,m}(x)$ is minimal is given by the inner offset $O_{\epsilon(m)}$, with $\epsilon(m) = \left(\frac{m\mathrm{Leb}_d(O)}{\omega_d}\right)^{\frac{1}{d}}$.*

As a consequence, if O and O' are two subsets of $\mathbb{R}^d$ such that $Leb_d(O) = Leb_d(O')$ but $Leb_d(O_{\epsilon(m)}) \neq Leb_d(O'_{\epsilon(m)})$, then the signatures $\mathrm{d}_{\mu_O,m}(\mu_O)$ and $\mathrm{d}_{\mu_{O'},m}(\mu_{O'})$ are different.

Proof. This proposition is a direct consequence of Proposition A.1 in the Appendix. □

This proposition can be applied to the following simple cases.

Example 3.3. *The uniform distributions on the rectangle $O = [0,1] \times [0,\pi]$ and the ball $O' = B(0,1)$ in $\mathbb{R}^2$ have a different signature for every $m \in (0,1)$.*

Proof. The volume of O_ϵ is given by $\mathrm{Leb}_2(O_\epsilon) = \pi(1-\epsilon)^2$ for $\epsilon \leq 1$. The volume of ${O'}_\epsilon$ is given by $\mathrm{Leb}_2({O'}_\epsilon) = (1-2\epsilon)(\pi - 2\epsilon)$ for $\epsilon \leq \frac{1}{2}$. For $1 > \epsilon > \frac{1}{2}$, $\mathrm{Leb}_2({O'}_\epsilon) = 0 < \mathrm{Leb}_2(O_\epsilon)$. Moreover, for $\epsilon \in [0, \frac{1}{2})$, $\mathrm{Leb}_2(O_\epsilon) > \mathrm{Leb}_2({O'}_\epsilon)$. In this example, $m = \epsilon^2$. □

More generally, "thin" sets and "fat" sets can be discriminated:

Example 3.4. *Set $\epsilon > 1$. The uniform distributions on the rectangle $O = [0,\epsilon] \times \left[0, \frac{\pi}{\epsilon}\right]$ and the ball $O' = B(0,1)$ in $\mathbb{R}^2$ have a different signature for $m = \epsilon^2$.*

Proof. The set O_ϵ is empty but ${O'}_\epsilon$ is not empty. □

Example 3.5. *Set $1 > \epsilon' > \epsilon > 0$. The uniform distributions on the rectangle $O = [0,\epsilon] \times \left[0, \frac{1}{\epsilon}\right]$ and the rectangle $O' = [0,\epsilon'] \times \left[0, \frac{1}{\epsilon'}\right]$ in $\mathbb{R}^2$ have a different signature for m such that $\sqrt{\frac{m}{\pi}} \in [\epsilon, \epsilon']$.*

Proof. For $m \in \left(\pi\left(\frac{\epsilon}{2}\right)^2, \pi\left(\frac{\epsilon'}{2}\right)^2\right)$, we get that $\epsilon(m) = \sqrt{\frac{m}{\pi}} \in \left(\frac{\epsilon}{2}, \frac{\epsilon'}{2}\right)$. For such values of m, $O_{\epsilon(m)}$ is empty but ${O'}_{\epsilon(m)}$ is not empty. □

The signatures of two sets that have the same volume and are very close (in terms of the Hausdorff metric for instance) might be different provided that the boundary of the first set is regular whereas the boundary of the second set is not. For instance, a rectangle and a biscuit-nantais-shaped rectangle have different signatures since the inner-offsets of the biscuit-nantais-shaped rectangle have smaller Lebesgue volume than the inner-offsets of the rectangle.

The aforementioned examples deal with measures with a support of dimension d in $\mathbb{R}^d$. Nonetheless, sometimes, it is also possible to prove that uniform distributions on submanifolds of dimension $d' < d$ of $\mathbb{R}^d$ have different signatures.

Example 3.6. *The uniform distributions on the circle* $O = \{x \in \mathbb{R}^2 \mid \|x\| = 1\}$ *and on the segment* $O' = [0, 2\pi] \times \{0\}$ *have a different signature for every* $m \in (0, 1]$.

Proof. We can compute $\delta_{\mu_O,l}(x) = 2\sin(\frac{\pi l}{2})$ for $x \in \mathrm{Supp}(\mu_O)$ and $\delta_{\mu_{O'},l}(x) \geq \pi l$ for $x \in \mathrm{Supp}(\mu_{O'})$. Then, $\sup_{x\in\mathrm{Supp}(\mu_O)} \mathrm{d}_{\mu_O,m}(x) < \inf_{x\in\mathrm{Supp}(\mu_{O'})} \mathrm{d}_{\mu_{O'},m}(x)$ for every $m \in (0, 1]$. □

For the same reason, a segment and a spiral with the same length have different signatures, whatever the value of m. This kind of example is investigated in the simulations part. In the following, we highlight the fact that signatures catch the density variation of distributions.

3.2.3. The case of two measures on the same open subset of $\mathbb{R}^d$ *where one of them is uniform*

Let $(O, \|\cdot\|_2, \mu_O)$ and $(O, \|\cdot\|_2, \nu)$ be two mm-spaces with O a non-empty bounded open subset of $\mathbb{R}^d$, $\mu_O = \frac{\mathrm{Leb}_d(.\cap O)}{\mathrm{Leb}_d(O)}$ and ν a measure absolutely continuous with respect to μ_O. Thanks to the Radon-Nikodym theorem, there is some μ_O-measurable function f on O such that, for all Borel sets A in O,

$$\nu(A) = \int_A f(\omega)\mathrm{d}\mu_O(\omega).$$

We can consider the λ-super-level sets of the function f denoted by $\{f \geq \lambda\}$. Again, we will denote by $\{f \geq \lambda\}_\epsilon$ the set of points belonging to $\{f \geq \lambda\}$ whose distance to $\partial\{f \geq \lambda\}$ is at least ϵ.

Then we get the following lower bound for the L_1-Wasserstein distance between the two signatures:

Proposition 3.7. *Under these hypotheses, a lower bound for* $W_1(\mathrm{d}_{\mu_O,m}(\mu_O), \mathrm{d}_{\nu,m}(\nu))$ *is given by*

$$\frac{d_{\min}}{d\,\mathrm{Leb}_d(O)} \int_{\lambda=1}^{\infty} \frac{1}{\lambda^{\frac{d+1}{d}}} \max_{\lambda'\geq\lambda} \lambda' \mathrm{Leb}_d\left(\{f \geq \lambda'\}_{\left(\frac{m}{\omega_d}\frac{\mathrm{Leb}_d(O)}{\lambda'}\right)^{\frac{1}{d}}}\right) \mathrm{d}\lambda,$$

with $d_{\min} = \left(\frac{m\mathrm{Leb}_d(O)}{\omega_d}\right)^{\frac{1}{d}} \frac{d}{d+1}$ *and* ω_d *stands for* $\mathrm{Leb}_d(\mathrm{B}(0, 1))$*, the Lebesgue volume of the unit d-dimensional ball.*

Proof. Proof in the Appendix, in Section A.3. □

It should be noted that when $\nu = \mu_O$, f is constant, equal to 1. Then, for $\lambda' > 1$, the sets $\{f \geq \lambda'\}$ are empty. As a consequence, the lower bound obtained in Proposition 3.7 is zero. Another simple example is the following.

Example 3.7. *The distribution* μ *with density* $\frac{1}{\pi}$ *on the ball* $O = \mathrm{B}(0, 1) = \{x \in \mathbb{R}^2 \mid \|x\| < 1\}$*, and the measure* ν *with density* $\frac{1}{2\pi}$ *on* $\mathrm{B}(0, 1)\backslash\mathrm{B}(0, \frac{1}{2})$ *and* $\frac{5}{2\pi}$ *on* $\mathrm{B}(0, \frac{1}{2})$ *have different signatures if* $m \leq \frac{5}{8}$*. Moreover, a lower bound for* $W_1(\mathrm{d}_{\mu,m}(\mu), \mathrm{d}_{\nu,m}(\nu))$ *is given by* $\frac{2\sqrt{m}}{3}\left(1 - \sqrt{\frac{2}{5}}\right)\left(\frac{\sqrt{5}}{2\sqrt{2}} - \sqrt{m}\right)^2$.

Proof. We have that $\mathrm{Leb}_2(O) = \pi$, $d_{\min} = \frac{2}{3}\sqrt{m}$. The density of ν with respect to μ is given by f which is equal to $\frac{1}{2}$ on $\mathrm{B}(0, 1)\backslash\mathrm{B}(0, \frac{1}{2})$ and $\frac{5}{2}$ on $\mathrm{B}(0, \frac{1}{2})$. As a consequence, $\{f \geq \lambda\}$ is empty as soon as $\lambda \geq \frac{5}{2}$. For $\lambda' \in \left(1, \frac{5}{2}\right)$, $\{f \geq \lambda'\}_{\left(\frac{m}{\omega_d}\frac{\mathrm{Leb}_d(O)}{\lambda'}\right)^{\frac{1}{d}}} =$

$\mathrm{B}\left(0, \frac{1}{2} - \left(\frac{m}{\omega_d}\frac{\mathrm{Leb}_d(O)}{\lambda'}\right)^{\frac{1}{d}}\right) = \mathrm{B}\left(0, \frac{1}{2} - \sqrt{\frac{m}{\lambda'}}\right)$. This set is non-empty if and only if $\lambda' \geq 4m$. Then, $\lambda' \mathrm{Leb}_2(\mathrm{B}\left(0, \frac{1}{2} - \sqrt{\frac{m}{\lambda'}}\right)) = \pi\left(\frac{\sqrt{\lambda'}}{2} - \sqrt{m}\right)^2$. It is maximal at $\lambda' = \frac{5}{2}$. Note that we must assume that $\frac{5}{2} \geq 4m$, that is, $m \leq \frac{5}{8}$.

Then, $W_1(\mathrm{d}_{\mu,m}(\mu), \mathrm{d}_{\nu,m}(\nu)) \geq \frac{2\sqrt{m}}{3}\left(1 - \sqrt{\frac{2}{5}}\right)\left(\frac{\sqrt{5}}{2\sqrt{2}} - \sqrt{m}\right)^2$. □

The proof of Proposition 3.7 extends to non-uniform distributions in $\mathbb{R}^d$, that are not necessarily supported on the same set.

Proposition 3.8. *Let μ and ν be two distributions on $\mathbb{R}^d$ with respective densities g and f with respect to the Lebesgue measure Leb_d. Assume that $g(x) \leq g_{\max}$ for every $x \in \mathbb{R}^d$. Then, a lower bound for $W_1(\mathrm{d}_{\mu,m}(\mu), \mathrm{d}_{\nu,m}(\nu))$ is given by*

$$\frac{d_{\min}\, g_{\max}^{\frac{1}{d}}}{d} \int_{\lambda=g_{\max}}^{+\infty} \frac{1}{\lambda^{\frac{d+1}{d}}} \max_{\lambda' \geq \lambda} \lambda' Leb_d\left(\{f \geq \lambda'\}_{\lambda'^{-\frac{1}{d}}\left(\frac{m}{\omega_d}\right)^{\frac{1}{d}}}\right) \mathrm{d}\lambda,$$

with $d_{\min} = \left(\frac{m}{g_{\max}\omega_d}\right)^{\frac{1}{d}} \frac{d}{d+1}$, a lower bound for the function $\mathrm{d}_{\mu,m}$.

Proof. Proof in the Appendix, in Section A.3. □

Actually, Proposition 3.7 is a consequence of Proposition 3.8. It suffices to replace $g_{\max}$ with $\frac{1}{\mathrm{Leb}_d(O)}$, the value of the density of μ_O with respect to the Lebesgue measure, on O. In the following, we work with the framework of Proposition 3.7.

When the density f is Hölder

In order to get additional results about discrimination, we need to define a quantity characterising the complexity of the set O. This is the notion of reach of an open set, defined from its medial axis.

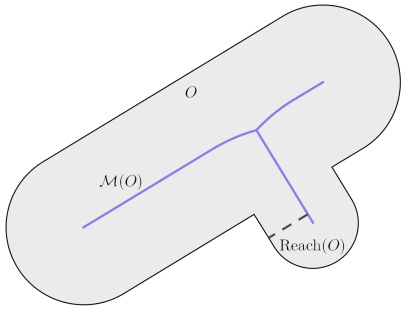


FIGURE 3. *Medial Axis and Reach of an open set*

The **medial axis** $\mathcal{M}(O)$ of O as the set of points in O having at least two projections onto ∂O. That is,

$$\mathcal{M}(O) = \{y \in O \mid \exists x', x'' \in \partial O,\, x' \neq x'',\, \|y - x'\|_2 = \|y - x''\|_2 = \mathrm{d}(y, \partial O)\},$$

with $\mathrm{d}(y, \partial O) = \inf\{\|x - y\|_2 \mid x \in \partial O\}$.

Its **reach**, $\mathrm{Reach}(O)$, is then defined as the distance between its boundary ∂O and its medial axis $\mathcal{M}(O)$. That is,

$$\mathrm{Reach}(O) = \inf\{\|x - y\|_2 \mid x \in \partial O,\ y \in \mathcal{M}(O)\}.$$

In the following, we assume that $\mathrm{Reach}(O) > 0$ and that f is Hölder on O, with positive parameters $\chi \in (0, 1]$ and $L > 0$, that is:

$$\forall x, y \in O,\ |f(x) - f(y)| \leq L\|x - y\|_2^{\chi}.$$

Then for m small enough, the DTM-signature is discriminative.

Proposition 3.9. *Under the assumptions of Proposition 3.7, if one of the following conditions is satisfied, then the quantity $W_1(\mathrm{d}_{\mu_O,m}(\mu_O), \mathrm{d}_{\nu,m}(\nu))$ is positive:*

$$m < \frac{\omega_d}{\mathrm{Leb}_d(O)} \min\left\{\mathrm{Reach}(O)^d, \left(\frac{\|f\|_{\infty,O} - 1}{2L}\right)^{\frac{d}{\chi}}\right\};$$

$$m \in \left[\frac{\omega_d}{\mathrm{Leb}_d(O)}(\mathrm{Reach}(O))^d, (\|f\|_{\infty,O} - 2L(\mathrm{Reach}(O))^{\chi})\,(\mathrm{Reach}(O))^d \frac{\omega_d}{\mathrm{Leb}_d(O)}\right);$$

$$m \in \left[\frac{\omega_d}{\mathrm{Leb}_d(O)}\left(\frac{d}{\chi}\right)^{\frac{d}{\chi}}(2L)^{-\frac{d}{\chi}}, \min\left\{m_0, \frac{\omega_d}{\mathrm{Leb}_d(O)}(\mathrm{Reach}(O))^{d+\chi}\frac{\chi}{d}2L\right\}\right),$$

with $m_0 = \|f\|_{\infty,O}^{\frac{d}{\chi}+1} \frac{\omega_d}{\mathrm{Leb}_d(O)}\left(\frac{d}{\chi}\right)^{\frac{d}{\chi}}(2L)^{-\frac{d}{\chi}}\left(\frac{\chi}{d+\chi}\right)^{\frac{\chi}{d+\chi}}$.

Moreover, under any of these conditions, we get the following lower bound on the pseudo-distance $W_1(\mathrm{d}_{\mu_O,m}(\mu_O), \mathrm{d}_{\nu,m}(\nu))$:

$$\frac{1}{1+d}\left(\frac{m\mathrm{Leb}_d(O)}{\omega_d}\right)^{\frac{1}{d}} \int_{\lambda=1}^{\infty} \frac{1}{\lambda^{1+\frac{1}{d}}} \sup_{\lambda' \geq \lambda} \nu\left(\{f \geq \lambda' + L\epsilon(\lambda')^{\chi}\} \cap O_{\epsilon(\lambda')}\right) \mathrm{d}\lambda,$$

with $\epsilon(\lambda') = \lambda'^{-\frac{1}{d}}\left(\frac{m\mathrm{Leb}_d(O)}{\omega_d}\right)^{\frac{1}{d}}$.

Here, ω_d stands for $\mathrm{Leb}_d(\mathrm{B}(0, 1))$, the Lebesgue volume of the unit d-dimensional ball.

Proof. Proof in the Appendix, in Section A.3. □

This proposition displays different intervals of values for m for which the DTM-signatures are discriminative. These intervals depend on the reach of O. Indeed, if m is small enough with respect to $\mathrm{Reach}(O)$, then the distance to the measure μ_O is easier to approximate on the whole set O, and is even known on most of the set (see Proposition A.1), and thus easier to compare with the distance to the measure ν.

The Hölder hypothesis provides some continuity of the density. Then, the function distance to the measure μ_O will take close values at any two points close enough. The main idea is to use the fact that the density of ν is not constant. Then, we can, for particular values of m, reveal a set of points of positive μ_O-measure for which the distance to the measure ν is smaller than the minimum of the distance to the measure μ_O.

Morally, this proposition consists in proving that $W_1(\mathrm{d}_{\mu_O,m}(\mu_O), \mathrm{d}_{\nu,m}(\nu))$ is positive whenever:

$$\|f\|_{\infty,O} > \inf\left\{\lambda + 2L\epsilon(\lambda)^\chi \,|\, \lambda \geq 1, \epsilon(\lambda) \leq \mathrm{Reach}(O)\right\}.$$

Note that $\epsilon(\lambda) \leq \mathrm{Reach}(O)$ is equivalent to $\lambda \geq \frac{m\mathrm{Leb}_d(O)}{\omega_d \mathrm{Reach}(O)^d}$. For each of the three intervals (a,b), the value of a or b is computed from the inequality $\|f\|_{\infty,O} > \lambda + 2L\epsilon(\lambda)^\chi$ for some λ well chosen, as follow. For the first interval, we take $\lambda = 1$; for the second one, we take λ such that $\epsilon(\lambda) = \mathrm{Reach}(O)$; for the last one, we take the λ that minimizes the function $\lambda \mapsto \lambda + 2L\epsilon(\lambda)^\chi$ on $\mathbb{R}_+$. The second value b or a is computed according to the additional constraint for the λs: $\lambda \geq 1$ and $\epsilon(\lambda) \leq \mathrm{Reach}(O)$.

In particular, the first interval is empty when $\frac{m\mathrm{Leb}_d(O)}{\omega_d \mathrm{Reach}(O)^d} > 1$ and the second one is empty when $\frac{m\mathrm{Leb}_d(O)}{\omega_d \mathrm{Reach}(O)^d} < 1$. The last interval is empty when the function $\lambda + 2L\epsilon(\lambda)^\chi$ attains its minimum at a point λ that is smaller than 1 or smaller than $\frac{m\mathrm{Leb}_d(O)}{\omega_d \mathrm{Reach}(O)^d}$. Since $\lambda \mapsto \lambda + 2L\epsilon(\lambda)^\chi$ is decreasing and then increasing, there is at least one non-empty interval if and only if $\|f\|_{\infty,O} > \inf\left\{\lambda + 2L\epsilon(\lambda)^\chi \,|\, \lambda \geq 1, \epsilon(\lambda) \leq \mathrm{Reach}(O)\right\}$.

This proposition can be applied to concrete cases, proving the existence of some mass parameters m for which the DTM-signature is discriminative.

Example 3.8. *Let O be the open unit ball $\mathrm{B}(0,1)$ in $\mathbb{R}^d$, μ_O the uniform measure on O and ν the multivariate normal distribution $\mathcal{N}(0,\sigma^2 I)$ restricted to $\mathrm{B}(0,1)$.*

The signatures are discriminative, provided that σ is small enough, for all m smaller than 0.23 if $d = 1$; 0.30 if $d = 2$; 0.68 if $d = 3$; and for all values of m if $d \geq 4$.

Proof. The density f of the multivariate normal distribution $\mathcal{N}(0,\sigma^2 I)$ restricted to $\mathrm{B}(0,1)$ is Lipschitz, as a consequence, it is possible to apply Proposition 3.9 with the parameter $\chi = 1$. The proof is deferred to the Appendix, in Section A.3. □

The previous examples provide several relevant cases where the DTM-signature turns out to be discriminative. Thus, the test of isomorphism will be powerful for some distributions.

4. Numerical experiments

In this section, we first describe the procedure to implement the statistical test of isomorphism. Then, we illustrate the validity of the method by providing some numerical approximations of the type-I error and the power of the test for various examples. We also compare our test to a more basic statistical test of isomorphism.

4.1. The algorithm

The procedure for the statistical test is as follows.
In the code, if $\mathbb{Z} = \{Z_1, Z_2, \ldots, Z_n\}$, then we use the notation $\mathbb{1}_{\mathbb{Z}}$ for the measure $\frac{1}{n}\sum_{i=1}^n \delta_{Z_i}$.

Algorithm 1. *Test Procedure*

Input :
$\mathbb{X} = \{X_1, X_2, \ldots, X_N\}$ N-sample from μ;
$\mathbb{Y} = \{Y_1, Y_2, \ldots, Y_{N'}\}$ N'-sample from ν;
parameter n, mass parameter m, level α, number of subsampling repetitions N_{sub} ;

`# Compute` T `the test statistic`
Let σ be a random permutation of $\{1, 2, \dots, N\}$;
Let σ' be a random permutation of $\{1, 2, \dots, N'\}$ independent of σ;
Define $\mathbb{X}_n = \{X_{\sigma(1)}, X_{\sigma(2)}, \dots, X_{\sigma(n)}\}$;
Define $\mathbb{Y}_n = \{Y_{\sigma'(1)}, Y_{\sigma'(2)}, \dots, Y_{\sigma'(n)}\}$;

Define the test statistic $T = \sqrt{n} W_1(\mathrm{d}_{\mathbb{1}_{\mathbb{X}},m}(\mathbb{1}_{\mathbb{X}_n}), \mathrm{d}_{\mathbb{1}_{\mathbb{Y}},m}(\mathbb{1}_{\mathbb{Y}_n}))$;

`# Compute` W_{sub}`, a` N_{sub}`-sample from the subsampling law`
Compute $dtm\mathbb{X} = \{\mathrm{d}_{\mathbb{1}_{\mathbb{X}},m}(X_1), \mathrm{d}_{\mathbb{1}_{\mathbb{X}},m}(X_2), \dots, \mathrm{d}_{\mathbb{1}_{\mathbb{X}},m}(X_N)\}$;
Compute $dtm\mathbb{Y} = \{\mathrm{d}_{\mathbb{1}_{\mathbb{Y}},m}(Y_1), \mathrm{d}_{\mathbb{1}_{\mathbb{Y}},m}(Y_2), \dots, \mathrm{d}_{\mathbb{1}_{\mathbb{Y}},m}(Y_{N'})\}$;
Let W_{sub} be empty;
for j **in** $1 \,..\, \lfloor N_{sub}/2 \rfloor$:
  Let $dtm\mathbb{X}_1$ and $dtm\mathbb{X}_2$ be two independent n-samples from $dtm\mathbb{X}$ with replacement;
  Let $dtm\mathbb{Y}_1$ and $dtm\mathbb{Y}_2$ be two independent n-samples from $dtm\mathbb{Y}$ with replacement;
  Add $\sqrt{n} W_1(\mathbb{1}_{dtm\mathbb{X}_1}, \mathbb{1}_{dtm\mathbb{X}_2})$ and $\sqrt{n} W_1(\mathbb{1}_{dtm\mathbb{Y}_1}, \mathbb{1}_{dtm\mathbb{Y}_2})$ to W_{sub};

`# Compute` $pval$`, the p-value of the statistical test`
Let $pval$ be equal to the mean number of elements in W_{sub} bigger than T;

Output:
The hypothesis retained is H_1 if $pval \leq \alpha$, H_0 if not.

Recall that the L_1-Wasserstein distance W_1 is simply equal to the L_1-norm of the difference between the cumulative distribution functions, which is easy to implement in the discrete case. As explained in the Introduction, in order to compute the distance to an empirical measure on a N-sample at a point x, it is sufficient to search for its $k = \lceil mN \rceil$-nearest neighbours in the sample, where $m \in [0, 1]$ is the mass parameter. The distance to the empirical measure can also be implemented by the `R` function `dtm` with tuning parameter $r = 1$, from the package `TDA` [23].

4.2. An example in $\mathbb{R}^2$

In this subsection, we will compare the statistical test of this paper (**DTM**) with the statistical test (**KS**) which consists in applying a Kolmogorov-Smirnov two-sample test to the $\frac{N}{2}$-sample

$$\{\delta(X_1, X_2), \delta(X_3, X_4), \dots, \delta(X_{N-1}, X_N)\}$$

and the $\frac{N'}{2}$-sample

$$\{\gamma(Y_1, Y_2), \gamma(Y_3, Y_4), \dots, \gamma(Y_{N'-1}, Y_{N'})\}$$

given an N-sample $\mathbb{X} = \{X_1, X_2, \dots, X_N\}$ from an mm-space $(\mathcal{X}, \delta, \mu)$ and an N'-sample $\mathbb{Y} = \{Y_1, Y_2, \dots, Y_{N'}\}$ from an mm-space $(\mathcal{Y}, \gamma, \nu)$.

We apply our isomorphism test to measures supported on spirals in $\mathbb{R}^2$. For some shape parameter $v \in \mathbb{R}_+$, the measure μ_v is the distribution of the random vector $(R\sin(vR) + 0.03S, R\cos(vR) + 0.03S')$, with R, S and S' independent random variables, S and S' from

the standard normal distribution $\mathcal{N}(0,1)$ and R uniform on $(0,1)$. In the following experiments, we choose $\mu = \mu_{10}$ and $\nu = \mu_v$ for $v \in \{15, 20, 30, 40, 100\}$.

First, from the measure $\mu = \mu_{10}$ we get an $N = 2000$-sample $\mathbb{X} = \{X_1, X_2, \ldots, X_N\}$. As well, we get an $N = 2000$-sample $\mathbb{Y} = \{Y_1, Y_2, \ldots, Y_N\}$ from the measure $\nu = \mu_{20}$. This leads to the empirical measures $\hat{\mu}_N$ and $\hat{\nu}_N$. In Figure 4, we plot the cumulative distribution function of the measure $\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_N)$, that is, the function F defined for all t in $\mathbb{R}$ by the proportion of the X_i in $\mathcal{X}$ satisfying $\mathrm{d}_{\hat{\mu}_N,m}(X_i) \leq t$. It approximates the true cumulative distribution function associated to the DTM-signature $\mathrm{d}_{\mu,m}(\mu)$. As well, we plot the cumulative distribution function of the measure $\mathrm{d}_{\hat{\nu}_N,m}(\hat{\nu}_N)$.

The signatures are different. Thus, for the choice of parameter $m = 0.05$, the DTM-signature discriminates well between the measures $\mu = \mu_{10}$ and $\nu = \mu_{20}$. The signature with parameter $m = 0.05$ provides a local information about the measure μ_v. The spiral with $v = 10$ is less coiled that the spiral with $v = 20$. It means that at a point of the spiral, catching 5 percent of points requires a larger radius for $v = 20$ than for $v = 10$. As a consequence, $\mathrm{d}_{\mu_{10},0.05}(\mu_{10})$ takes smaller values than $\mathrm{d}_{\mu_{20},0.05}(\mu_{20})$, as illustrated by Figure 4. Note that a m close to 1 would not be relevant in such an example. Indeed, $\mathrm{d}_{\mu,1}(x)$ roughly corresponds to the distance of point x to the expectation of the measure μ. Since the spirals have the same diameter, the signatures would be very close for m close to 1.

Note that a small m would not be appropriated to discriminate between a spiral and its uncoiled version (a noisy sample generated around a segment). Indeed, the local behavior of the measure would be the same. In this case, a large choice of m would be more appropriate.

In Figure 5, for $m = 0.05$ and $n = 20$, we first generate 1000 independent realisations of the random variable $\sqrt{n}W_1(\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_n), \mathrm{d}_{\hat{\mu}'_N,m}(\hat{\mu}'_n))$, where $\hat{\mu}_N$ and $\hat{\mu}'_N$ are independent empirical measures from μ_{10}, $N = 2000$, and $\hat{\mu}_n$ and $\hat{\mu}'_n$ are the empirical measures associated to the n first points of the samples. We plot the empirical cumulative distribution function associated to this 1000-sample.

As well, from two fixed N-samples from μ_{10}, leading to two empirical distributions $\hat{\mu}_N$ and $\hat{\mu}'_N$, we generate a set of $N_{sub} = 1000$ random variables $\sqrt{n}W_1(\mathrm{d}_{\hat{\mu}_N,m}(\mu^*_n), \mathrm{d}_{\hat{\mu}'_N,m}({\mu'}^*_n))$, as explained in the Algorithm in Section 4.1, and we plot its cumulative distribution function. Note that the two cumulative distribution functions are close. This means that the $1-\alpha$-quantile of the distribution of the test statistic is well approximated by the $1-\alpha$-quantile of the subsampling distribution.

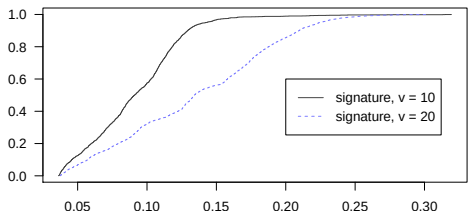


FIGURE 4. *DTM-signature estimates,* $m = 0.05$

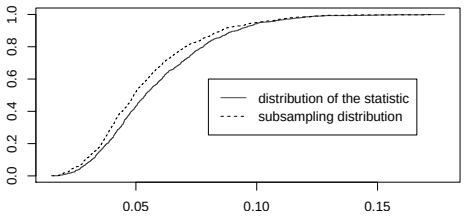


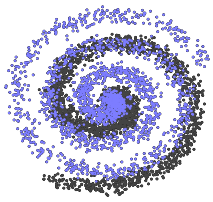
FIGURE 5. *Subsampling validity,* $v = 10$, $m = 0.05$

In order to approximate the type-I error and the power, we repeat the procedure of test **DTM** described in Section 4.1 1000 times independently. At each step, we sample $N = 2000$ points from the measures $\mu = \mu_{10}$ and $\nu = \mu_v$ to approximate the power, or twice μ_v to

approximate for the type-I error. We select the parameters $\alpha = 0.05$, $m = 0.05$, $n = 20$, and repeat subsampling $N_{sub} = 1000$ times. Then, we retain either H_0 or H_1. The type-I error or power approximation is simply equal to the mean number of times the hypothesis H_0 was rejected among the 1000 independent experiments. We also approximate the power for the method **KS** after repeating this test procedure 1000 times independently. Note that by construction, the test **KS** is truly of level $\alpha = 0.05$. Figure 8 contains the numerical values we obtained using the R software.

v	15	20	30	40	100
type-I error **DTM**	0.043	0.049	0.050	0.051	0.050
power **DTM**	0.525	0.884	0.987	0.977	0.985
power **KS**	0.768	0.402	0.465	0.414	0.422

FIGURE 6. *Type-I error and power approximations*



It turns out that our isomorphism test **DTM** is of level close to $\alpha = 0.05$ and is powerful. For parameters $v \geq 20$, our test is even more discriminative than the test **KS**.

4.3. An example in $\mathbb{R}^{28\times 28}$

In this subsection, we use our statistical test of isomorphism to compare the distribution of the digits "2" and the distribution of the digits "5" from the MNIST handwritten digits database. Each digit is represented by a picture of 28×28 pixels with grey levels, meaning that each digit $X = (x_1, x_2, \ldots, x_{28\times 28})$ can be seen as an element of $\mathbb{R}^{28\times 28}$, where x_i is equal to the grey level of the i-th pixel. We equip $\mathbb{R}^{28\times 28}$ with the Euclidean metric.

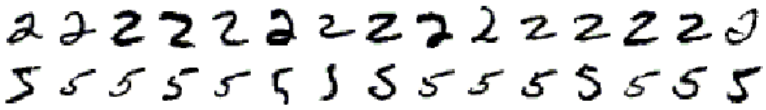


FIGURE 7. *MNIST database of handwritten digits*

The interest of the statistical test here is not to test whether a "2" and a "5" are isometric, but whether the distribution of the "2"s and the distribution of the "5"s (which are distributions on compact subsets of $\mathbb{R}^{28\times 28}$) are equal up to an isomorphism. These distributions are clearly not equal, since their supports are different, but it is not clear that there is no rigid transformation between the set of "2" and the set of "5" preserving the measures, as for instance a simple permutation of the pixels.

The statistical test is based on the observation of $N = 5958$ "2" and $N' = 5421$ "5". In order to prove the validity of the test, we repeat 1000 times the experiment consisting in randomly splitting the set of "2" in two parts and applying the statistical test to these two samples. The type-I error approximation is equal to the mean number of times the hypothesis H_0 was rejected. We do the same with the set of "5". And we repeat these experiments for different values of $n \in \{10, 20, 30, 50, 75, 100, 200\}$ and for a fixed $m = 0.1$, we repeat subsampling $N_{sub} = 1000$ times.

These results are encouraging since they prove that the test does not discriminate between two samples of "2" (respectively, between two samples of "5") with probability 0.95.

n	10	20	30	50	75	100	200
"2"	0.052	0.052	0.051	0.052	0.058	0.048	0.069
"5"	0.064	0.07	0.043	0.044	0.047	0.045	0.054

FIGURE 8. *Type-I error approximations*

Thus, the type-I error is of order 0.05. We choose the parameter $n = 100$ to make the test between the sample of "2" and the sample of "5". We get a p-value equal to 0. It means that we reject H_0 at any level α. So we can conclude that the distribution of the "2" and the distribution of the "5" in the MNIST database are not isomorphic.

Unlike the spirals, there is a priori no intuition about how to choose a parameter m that discriminates between the distribution of the "2" and the distribution of the "5". In this case, one may choose some intermediate parameter m (for instance $m = 0.1$), which does not contains too local or global information about μ. A better idea is to refer to Theorem 2.3 and the remarks below. One may plot $\frac{W_1\left(\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_N),\mathrm{d}_{\hat{\nu}_N,m}(\hat{\nu}_N)\right)}{\max\left\{\mathcal{D}^2_{\hat{\mu}_N,m},\mathcal{D}^2_{\hat{\nu}_N,m}\right\}}$ as a function of m. A suitable choice of m to discriminate between the distributions would be any maximum m of this function.

5. A discussion of other methods

5.1. The Kolmogorov-Smirnov test applied to empirical DTM-signatures

In order to test isomorphism between two mm-spaces $(\mathcal{X}, \delta, \mu)$ and $(\mathcal{Y}, \gamma, \nu)$ from two N-samples $\{X_1, X_2, \ldots, X_N\}$ and $\{Y_1, Y_2, \ldots, Y_N\}$, the idea of applying a Kolmogorov-Smirnov test to the set of points $\mathcal{D}_P = \{\mathrm{d}_{\hat{\mu}_N,m}(X_1), \mathrm{d}_{\hat{\mu}_N,m}(X_2), \ldots, \mathrm{d}_{\hat{\mu}_N,m}(X_N)\}$ on one hand, and the set of points $\mathcal{D}_Q = \{\mathrm{d}_{\hat{\nu}_N,m}(Y_1), \mathrm{d}_{\hat{\nu}_N,m}(Y_2), \ldots, \mathrm{d}_{\hat{\nu}_N,m}(Y_N)\}$ on the other fails drastically. Indeed, using a Kolmogorov-Smirnov test requires independence in the data, which the data $\mathcal{D}_P$ and $\mathcal{D}_Q$ do not satisfy.

In the following figure, we have plotted the cumulative distribution functions associated to the uniform measures on $\mathcal{D}_P$ and $\mathcal{D}_Q$, where μ and ν are equal and defined as in Section 4.2, with $v = 10$. The p-value was equal to $4.10^{-6} \leq 0.05$, thus the hypothesis H_0 was rejected.

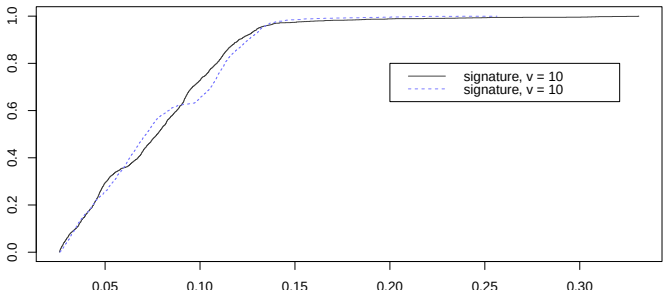


FIGURE 9. *DTM-signature estimates,* $m = 0.05$

We repeated the experiment 1000 times independently, and the proportion of rejected hypotheses H_0 was equal to 0.926, instead of 0.05 for a statistical test of level 0.05.

Thus, applying a Kolmogorov-Smirnov test to empirical DTM-signatures is to be avoided.

5.2. A different value of n for the test statistic and for the subsampling distribution

In [39], Politis and Romano propose subsampling methods consisting in approximating the distribution of a statistic with values of the statistic built from smaller subsets of the data.

For our statistical test, since the distribution of the statistic and the subsampling distribution converge weakly to the same distribution under some assumptions, see Lemma 2.1, we can imagine fixing some parameters n and l smaller than N, choosing as a test statistic

$$T = \sqrt{l} W_1\left(\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_l), \mathrm{d}_{\hat{\nu}_N,m}(\hat{\nu}_l)\right)$$

and approximating its distribution with the subsampling distribution

$$\frac{1}{2}\mathcal{L}^*_{N,n,m}(\hat{\mu}_N, \hat{\mu}_N) + \frac{1}{2}\mathcal{L}^*_{N,n,m}(\hat{\nu}_N, \hat{\nu}_N).$$

If we do so, consider (a,b)-standard measures supported on compact subsets of $\mathbb{R}^d$ and choose $N = n^\rho$ and $l = n^\beta$ with $1 < \beta < \rho$, then the test is asymptotically of level α, provided that the cumulative distribution function of $\|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1$ is continuous. Indeed, the hypothesis of Lemma 2.1, $\sqrt{l}\,\mathbb{E}[\|\mathrm{d}_{\mu,m} - \mathrm{d}_{\hat{\mu}_N,m}\|_{\infty,\mathcal{X}}] \to 0$, will be satisfied then. Moreover, the proof of Theorem 2.3, which provides an upper-bound for the type-II error, can be generalised to this case, leading to the upper-bound

$$4\exp\left(-\frac{W_1^2\left(\mathrm{d}_{\mu,m}(\mu), \mathrm{d}_{\nu,m}(\nu)\right)}{(2+\epsilon)\max\left\{\mathcal{D}^2_{\mu,m}, \mathcal{D}^2_{\nu,m}\right\}} n^\beta\right)$$

for the type-II error of the test $\phi_{N,n,m} = \mathbb{1}_{\sqrt{l}W_1\left(\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_l), \mathrm{d}_{\hat{\nu}_N,m}(\hat{\nu}_l)\right) \geq \hat{\mathrm{q}}_{1-\alpha,N,n,m}}$ if n is big enough. Note that it is a real improvement for the power.

Nonetheless, the upper-bounds for the L_1-Wasserstein distance between the test statistic distribution and the subsampling distribution in Proposition 2.1 and Proposition 2.2 cannot be generalised easily.

The following experiments emphasize the fact that the subsampling distribution does not necessarily well approximate the distribution of the test statistic if n and l are different. For the parameters $n = 20$ and $l \in \{20, 50, 200\}$, we have repeated 1000 times the experiment consisting of computing the p-value of our statistical test from two 2000-samples on a spiral with shape parameter $v = 10$, subsampling $N_{sub} = 1000$ times and with mass parameter $m = 0.05$. We sorted these p-values and plotted the associated cumulative distribution function. In this experiment, the hypothesis H_0 is satisfied, so the p-values should be uniformly distributed. Moreover they are independent. Thus, the curve we obtained should lie close to the diagonal. This is not the case when l is too far from n.

However, we get a power equal to 1 when choosing $l = 200$ or $l = 50$ instead of $l = 20$, which is much better than 0.884 which was obtained in Section 4.2 from the same experiment but with $l = n = 20$.

Such a procedure should not be used, despite the improvement of power. Indeed, we have not proved the existence of a non-asymptotic control of a distance between the distribution of the test statistic and the subsampling distribution. Moreover, the experiments emphasize that these distributions are too different to get a test of type-I error not greater than α.

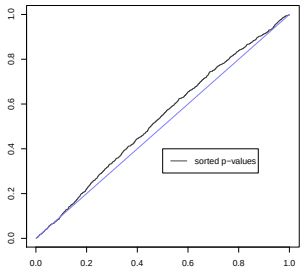


FIGURE 10. $l = 20$

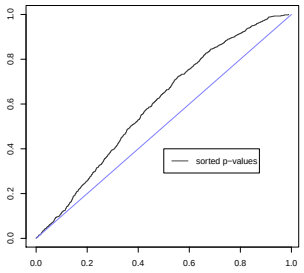


FIGURE 11. $l = 50$

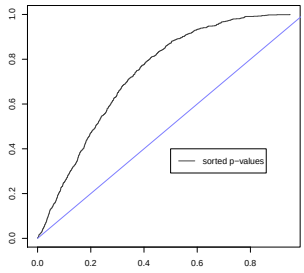


FIGURE 12. $l = 200$

5.3. *The one-sample Kolmogorov-Smirnov test of uniformity applied to p-values*

A major problem of the statistical test proposed in this paper is that the hypothesis retained truly depends on the arbitrary selection of the two n-samples to build the test statistic among the $\binom{N}{n}^2$ possible pairs of n-samples. Indeed, the p-value defined in Section 2 is random in the sense that different p-values can be associated to the same two N-samples. Moreover, the power is not that high because of n, which can be very small in comparison to N.

As an example, in Figure 13, we split an $N = 2000$-sample $\mathbb{X} = \{X_1, X_2, \ldots, X_N\}$ from the distribution μ_{10} on the spiral with shape parameter $v = 10$, into $\frac{N}{n} = 100$ disjointed subsets

$$\begin{aligned}
&\mathbb{X}_1 = \{X_1, X_2, \ldots, X_n\},\\
&\mathbb{X}_2 = \{X_{n+1}, X_{n+2}, \ldots, X_{2n}\},\\
&\ldots\\
&\mathbb{X}_{\frac{N}{n}} = \{X_{N-n+1}, X_{N-n+2}, \ldots, X_N\},
\end{aligned}$$

with $n = 20$.

As well, we split $\mathbb{Y}$, an $N = 2000$-sample from μ_{10} into $\frac{N}{n} = 100$ disjointed subsets $\mathbb{Y}_1$, $\mathbb{Y}_2$,... $\mathbb{Y}_{\frac{N}{n}}$.

Then, with the notation in the algorithm in Section 4.1, we consider the p-values $\hat{p}_1$, $\hat{p}_2$,... $\hat{p}_{\frac{N}{n}}$ with $\hat{p}_i$ associated to $T_i = \sqrt{n} W_1(\mathrm{d}_{\mathbb{1}_{\mathbb{X}},m}(\mathbb{1}_{\mathbb{X}_i}), \mathrm{d}_{\mathbb{1}_{\mathbb{Y}},m}(\mathbb{1}_{\mathbb{Y}_i}))$.

Note that the $\frac{N}{n}$ p-values would be independent if we replaced $\mathrm{d}_{\mathbb{1}_{\mathbb{X}},m}$ by $\mathrm{d}_{\mu_{10},m}$ in the computation of the test statistic, and if the subsampling distribution was replaced with the true distribution of the statistic. In practice, when N is big enough, we are close to these assumptions. Then the p-values $\hat{p}_1$, $\hat{p}_2$,... $\hat{p}_{\frac{N}{n}}$ should behave like independent random variables uniformly distributed on $[0, 1]$.

In figure 13, we have sorted these $\frac{N}{n} = 100$ p-values, which were built after repeating subsampling $N_{sub} = 1000$ times and with mass parameter $m = 0.05$. They seem to be uniform on $[0, 1]$; indeed their associated cumulative distribution function lies close to the diagonal.

We use this randomness to propose the following method (**DTM-KS**) to improve the power of our statistical test. We apply a one-sample Kolmogorov-Smirnov test of uniformity on $[0, 1]$ to the sample $\hat{p}_1$, $\hat{p}_2$,... $\hat{p}_{\frac{N}{n}}$. Then we get a p-value p_{KS}. Thanks to the previous

heuristic, this p-value should be close to uniform on $[0,1]$ if the hypothesis H_0 of the isomorphism test was satisfied, and should be small if most of the p-values $\hat{p}_1$, $\hat{p}_2$,... $\hat{p}_{\frac{N}{n}}$ were small. We can hope for a power improvement.

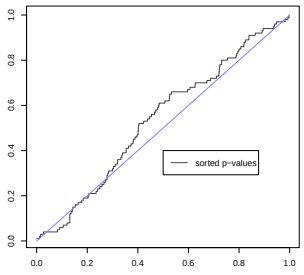


FIGURE 13. *Sorted p-values* $\hat{p}_1$, $\hat{p}_2$,... $\hat{p}_{\frac{N}{n}}$

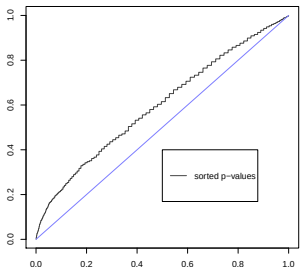


FIGURE 14. *Sorted p-values* p_{KS} ***DTM-KS***.

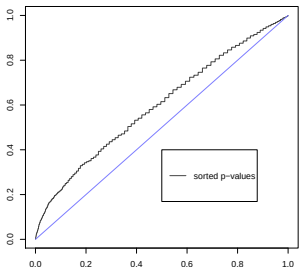


FIGURE 15. *Sorted p-values* p_{KS} ***DTM-KS2***.

In Figure 16, we evaluate the type-I error and the power for this new method, with the same procedure as in Section 4.2 and the same parameters.

In Figure 17, we evaluate the type-I error and the power for the testing method (**DTM-KS2**) consisting in applying a one-dimensional Kolmogorov-Smirnov test of uniformity on $[0,1]$ to the p-values $\hat{p}'_1$, $\hat{p}'_2$,... $\hat{p}'_{100}$, where $\hat{p}'_i$ is obtained from the test statistic

$$T'_i = \sqrt{n} W_1(\mathrm{d}_{\mathbb{1}_{\mathbb{X}},m}(\mathbb{1}_{\mathbb{X}'_i}), \mathrm{d}_{\mathbb{1}_{\mathbb{Y}},m}(\mathbb{1}_{\mathbb{Y}'_i}))$$

with $\mathbb{X}'_1$, $\mathbb{X}'_2$,... $\mathbb{X}'_{100}$ and $\mathbb{Y}'_1$, $\mathbb{Y}'_2$,... $\mathbb{Y}'_{100}$ independent n-samples without replacement from $\mathbb{X}$ and $\mathbb{Y}$ respectively. The procedure is the same as in Section 4.2 and the parameters are the same as well.

These procedures lead to major improvements for the power, but the type-I error degrades.

6. Concluding remarks and perspectives

This paper opens a new horizon of statistical tests based on shape signatures. It could be of interest to adapt these kind of methods to other signatures, if possible. In future it could even be interesting to build statistical tests based on many different signatures, leading to an even better discrimination. Regarding the test proposed in this paper itself, the geometric and statistical problem of the choice of the best parameters to use in practice is still an open, tough and engaging question.

Acknowledgements The author is extremely grateful to Frédéric Chazal, Pascal Massart and Bertrand Michel for introducing her to the distance to a measure, for their valuable comments and advise, and for proofreading. Many thanks also to Małgorzata Bogdan for her interest and pertinent comments. As well, Eddie Aamari, Marc Glisse, Clément Levrard and members of the Datashape and Select Inria teams should be thanked for their interest, and Cormac Walsh for the english. Last but not least, the author is extremely grateful to the anonymous referees for their valuable comments and suggestions.

v	15	20	30	40	100
type-I error **DTM-KS**	0.186	0.131	0.096	0.076	0.074
power **DTM-KS**	1	1	1	1	1

FIGURE 16. *Type-I error and power approximations* ***DTM-KS***

v	15	20	30	40	100
type-I error **DTM-KS2**	0.198	0.145	0.093	0.073	0.088
power **DTM-KS2**	1	1	1	1	1

FIGURE 17. *Type-I error and power approximations* ***DTM-KS2***

Appendix

Some notations and definitions: For O a non-empty bounded open subset of $\mathbb{R}^d$, we define the **uniform measure** μ_O for any Borel set A of $\mathbb{R}^d$, by

$$\mu_O(A) = \frac{\mathrm{Leb}_d(O \cap A)}{\mathrm{Leb}_d(O)},$$

with Leb_d the Lebesgue measure on $\mathbb{R}^d$.

We also define the **medial axis** of O, $\mathcal{M}(O)$ as the set of points in O having at least two projections onto ∂O. That is,

$$\mathcal{M}(O) = \{y \in O \,|\, \exists\, x', x'' \in \partial O,\, x' \neq x'',\, \|y - x'\|_2 = \|y - x''\|_2 = \mathrm{d}(y, \partial O)\},$$

with $\mathrm{d}(y, \partial O) = \inf\{\|x - y\|_2 \,|\, x \in \partial O\}$.

Its **reach**, $\mathrm{Reach}(O)$, is the distance between its boundary ∂O and its medial axis $\mathcal{M}(O)$. That is,

$$\mathrm{Reach}(O) = \inf\{\|x - y\|_2 \,|\, x \in \partial O,\, y \in \mathcal{M}(O)\}.$$

If K is a compact subset of $\mathbb{R}^d$, it is standard to define its reach as $\mathrm{Reach}\,(K^c)$, the reach of its complement in $\mathbb{R}^d$. See [25] to get more familiar with these notions.

In the following, ω_d stands for $\mathrm{Leb}_d(\mathrm{B}(0, 1))$, the Lebesgue volume of the unit d-dimensional ball.

Appendix A: Uniform measures on open subsets of $\mathbb{R}^d$

In this part, we focus on some mm-spaces $(O, \|\cdot\|_2, \mu_O)$ where O stands for a non-empty bounded open subset of $\mathbb{R}^d$ satisfying $\left(\overline{O}\right)^\circ = O$. Then, $\epsilon(m, O)$ is defined for some mass parameter m in $[0, 1]$ by

$$\epsilon(m, O) = \left(\frac{m\mathrm{Leb}_d(O)}{\omega_d}\right)^{\frac{1}{d}}.$$

This is the radius of a ball included in O, with μ_O measure equal to m. For some positive ϵ, O_ϵ stands for the set of points in O which distance to ∂O is not smaller than ϵ:

$$O_\epsilon = \left\{x \in O,\, \inf_{y \in \partial O} \|x - y\|_2 \geq \epsilon\right\}.$$

A.1. The distance to uniform measures

Here, we derive some properties of the spaces $(O, \|\cdot\|_2, \mu_O)$. We give a lower bound for the minimum of the distance to the measure μ_O and give a description of the points attaining this bound. First, we state some technical lemma proposed by Lieutier in [33].

Lemma A.1. *If we define the* ***skeleton*** $\mathrm{Sk}(O)$ *of the open set O as the set of centres of maximal balls (for the inclusion) included in O, then we get:*

$$\mathcal{M}(O) \subset \mathrm{Sk}(O) \subset \overline{\mathcal{M}(O)}.$$

Now we can formulate some technical lemma:

Lemma A.2. *For any x in O, there exist a maximal ball for the inclusion, included in O and containing x.*

Proof. Let us consider the class $\mathcal{S} = \{\mathrm{B}(y,r) \,|\, r > 0 \text{ and } x \in \mathrm{B}(y,r) \subset O\}$ of all non-empty open balls included in O and containing x. We are going to show that this class contains a maximal element by using the Zorn's lemma. For this, we need to show that the partially-ordered set $\mathcal{S}$ is inductive, which means that any non-empty totally-ordered subclass $\mathcal{T}$ of $\mathcal{S}$ is bounded above by some element of $\mathcal{S}$. Let $\mathcal{T}$ be a non-empty totally-ordered subclass of $\mathcal{S}$. Set $R = \sup\{r > 0 \,|\, \exists\, y \in O, \mathrm{B}(y,r) \in \mathcal{T}\}$ the supremum of the radii of all balls in $\mathcal{T}$. Since $\mathcal{T}$ is non-empty and O is bounded, R is positive and finite. Let $(y_k)_{k\in\mathbb{N}}$ be a sequence of centres of balls in $\mathcal{T}$ converging to a point y in $\mathbb{R}^d$ such that the sequence of associated radii $(r_k)_{k\in\mathbb{N}}$ is non decreasing with R as a limit. Since $\mathcal{T}$ is totally-ordered and the radii non decreasing, for every $K \in \mathbb{N}$, $\bigcup_{k\leq K} \mathrm{B}(y_k, r_k) = \mathrm{B}(y_K, r_K)$. Then, the union $\bigcup_{k\in\mathbb{N}} \mathrm{B}(y_k, r_k)$ is equal to $\mathrm{B}(y, R)$. Thus, $\mathrm{B}(y, R)$ belongs to $\mathcal{S}$ and upper bounds $\mathcal{T}$. So the class $\mathcal{S}$ is inductive and thanks to the Zorn's lemma, it contains a maximal element. □

Proof of Example 2.1: For any point x in O and $r > 0$, thanks to Lemma A.2 there exist a maximal ball $\mathrm{B}(x', r')$ included in $O \cap \mathrm{B}(x, r)$ which contains x. Assume for the sake of contradiction that $r' < \min\left\{\frac{r}{2}, \mathrm{Reach}(O)\right\}$.

Since $r' < \frac{r}{2}$, the ball $\overline{\mathrm{B}}(x', r')$ is included in $\mathrm{B}(x, r)$ thus $\mathrm{B}(x', r')$ is maximal in O. So x' belongs to $\mathrm{Sk}(O)$, and thanks to Lemma A.1, to $\overline{\mathcal{M}(O)}$. But $r' < \mathrm{Reach}(O)$; this is a contradiction.

It follows that:

$$\mu_O(\mathrm{B}(x,r)) \geq \mu_O\left(\mathrm{B}\left(x', \min\left\{\mathrm{Reach}(O), \frac{r}{2}\right\}\right)\right).$$

So, for $r \leq 2\mathrm{Reach}(O)$, since $2\mathrm{Reach}(O) \leq \mathcal{D}(O)$ by considering a point on $\mathrm{Sk}(O)$, we get:

$$\mu_O(\mathrm{B}(x,r)) \geq r^d \left(\frac{\mathrm{Reach}(O)}{\mathcal{D}(O)}\right)^d \frac{\omega_d}{\mathrm{Leb}_d(O)},$$

which is also true for r in $[2\mathrm{Reach}(O), \mathcal{D}(O)]$, whereas for $r \geq \mathcal{D}(O)$ we have $\mu_O(\mathrm{B}(x,r)) = 1$. The choice of a in the lemma is thus relevant. ■

We now focus on the set of points in $\mathbb{R}^d$ minimizing the distance to the measure μ_O. For this, we need some lemma.

Lemma A.3. *If x in $\mathbb{R}^d$ satisfies $\mu_O(\mathrm{B}(x,\epsilon)) = \frac{\omega_d \epsilon^d}{\mathrm{Leb}_d(O)}$, then $\mathrm{B}(x,\epsilon) \subset O$.*

Proof. If x in $\mathbb{R}^d$ satisfies $\mu_O(\mathrm{B}(x,\epsilon)) = \frac{\omega_d \epsilon^d}{\mathrm{Leb}_d(O)}$, then, $\mathrm{Leb}_d(O^c \cap \mathrm{B}(x,\epsilon)) = 0$. Assume for the sake of contradiction that the set $O^c \cap \mathrm{B}(x,\epsilon)$ is not empty. Since $\left(\overline{O}\right)^\circ = O$, then the open subset $(O^c)^\circ \cap \mathrm{B}(x,\epsilon)$ of $O^c \cap \mathrm{B}(x,\epsilon)$ is not empty, thus of positive Lebesgue measure, which is absurd. So $\mathrm{B}(x,\epsilon) \subset O$. □

Proposition A.1. *The constant* $\mathrm{d}_{\min} = \frac{d}{d+1}\left(\frac{m\mathrm{Leb}_d(O)}{\omega_d}\right)^{\frac{1}{d}}$ *is a lower bound for the distance to the measure* μ_O *over* $\mathbb{R}^d$*. Moreover, the set of points attaining this bound is exactly* $O_{\epsilon(m,O)}$*.*

Proof. Note that for all positive l smaller than m, we have:

$$\delta_{\mu,l}(x) \geq \left(\frac{l\mathrm{Leb}_d(O)}{\omega_d}\right)^{\frac{1}{d}}.$$

Moreover, these inequalities are equalities for all points x in $O_{\epsilon(m,O)}$. By integrating, we get the lower bound $\mathrm{d}_{\min}$ for $x \mapsto \mathrm{d}_{\mu,m}(x)$, and it is attained on $O_{\epsilon(m,O)}$.

Now take some point x in $\mathbb{R}^d$ satisfying $\mathrm{d}_{\mu,m}(x) = \mathrm{d}_{\min}$. For almost all l smaller than m, we have: $\delta_{\mu,l}(x) = \left(\frac{l\mathrm{Leb}_d(O)}{\omega_d}\right)^{\frac{1}{d}}$. In particular we get for these values of l that:

$$\mu\left(\overline{\mathrm{B}}\left(x, \left(\frac{m\mathrm{Leb}_d(O)}{\omega_d}\right)^{\frac{1}{d}}\right)\right) > l.$$

So, $\mu\left(\mathrm{B}\left(x, \left(\frac{m\mathrm{Leb}_d(O)}{\omega_d}\right)^{\frac{1}{d}}\right)\right) = m$, and thanks to Lemma A.3, we get that $x \in O_{\epsilon(m,O)}$. □

Proposition A.2. *If* $\mathrm{Reach}(O) \geq \epsilon(m,O)$*, then:*

$$\{x \in \mathbb{R}^d \,|\, \mathrm{d}_{\mu,m}(x) = \mathrm{d}_{\min}\}^{\epsilon(m,O)} = O,$$

where for any set A*, the notation* A^ϵ *stands for* $\bigcup_{x\in A} \overline{\mathrm{B}}(x,\epsilon)$*, the* ϵ***-offset*** *of* A*.*

Proof. Recall that, thanks to Proposition A.1 that $\{x \in \mathbb{R}^d \,|\, \mathrm{d}_{\mu,m}(x) = \mathrm{d}_{\min}\} = O_{\epsilon(m,O)}$. Moreover, $O^{\epsilon(m,O)}_{\epsilon(m,O)} \subset O$. Assume for the sake of contradiction that the set $O \backslash O^{\epsilon(m,O)}_{\epsilon(m,O)}$ is non-empty. Take a point x in this set and consider $\mathrm{B}(x',r')$ a maximal ball containing x and included in O given by Lemma A.2. Since $x \notin O^{\epsilon(m,O)}_{\epsilon(m,O)}$, we get that $r' < \epsilon(m,O)$. Moreover, x' belongs to $\mathrm{Sk}(O)$ and so, thanks to Lemma A.1, to $\overline{\mathcal{M}}(O)$. Then, by continuity of the function distance to the compact set ∂O, $r' = \mathrm{d}_{\partial O}(x') \geq \mathrm{Reach}(O) \geq \epsilon(m,O)$, which is a contradiction. So, $O^{\epsilon(m,O)}_{\epsilon(m,O)} = O$. □

A.2. The DTM-signature to discriminate between two uniform measures with different density.

Proof of Proposition 3.5: This proposition comes from the fact that for m small enough, for both mm-space, the distance-to-a-measure function will attain its minimum on a set of positive measure. Moreover, the two minima are different since the Lebesgue measure of the open sets are different.

More precisely, if the set $O_{\epsilon(m,O)}$ is non-empty, then the minimal value of the distance to a measure is given by

$$\min_{x\in\mathbb{R}^d}(\mathrm{d}_{\mu_O,m}(x)) = \mathrm{d}_{\min} := \frac{d}{d+1}\left(\frac{m\mathrm{Leb}_d(O)}{\omega_d}\right)^{\frac{1}{d}}.$$

Moreover, the points at minimal distance are exactly the points of $O_{\epsilon(m,O)}$. This is Proposition A.1. So, $F_{\mathrm{d}_{\mu_O,m}(\mu_O)}(\mathrm{d}_{\min}) = \mu_O\left(O_{\epsilon(m,O)}\right)$ and $F_{\mathrm{d}_{\mu_O,m}(\mu_O)}(x) = 0$ for every $x < \mathrm{d}_{\min}$. The same holds for O': $F_{\mathrm{d}_{\mu_{O'},m}(\mu_{O'})}(\mathrm{d}'_{\min}) = \mu_{O'}\left(O'_{\epsilon(m,O')}\right)$ and $F_{\mathrm{d}_{\mu_{O'},m}(\mu_{O'})}(x) = 0$ for every $x < \mathrm{d}'_{\min} := \frac{d}{d+1}\left(\frac{m\mathrm{Leb}_d(O')}{\omega_d}\right)^{\frac{1}{d}}$.

Then, definition of the L_1-Wasserstein metric as the L_1-norm between the cumulative distribution functions yields that:

$$W_1(\mathrm{d}_{\mu_O,m}(\mu_O), \mathrm{d}_{\mu_{O'},m}(\mu_{O'})) \geq \int_{\min(\mathrm{d}_{\min},\mathrm{d}'_{\min})}^{\max(\mathrm{d}_{\min},\mathrm{d}'_{\min})} \left|F_{\mathrm{d}_{\mu_O,m}(\mu_O)}(x) - F_{\mathrm{d}_{\mu_{O'},m}(\mu_{O'})}(x)\right| \mathrm{d}x.$$

Assume for instance that $\mathrm{d}_{\min} \leq \mathrm{d}'_{\min}$. Since cumulative distribution functions are non-decreasing, it comes that

$$\int_{\min(\mathrm{d}_{\min},\mathrm{d}'_{\min})}^{\max(\mathrm{d}_{\min},\mathrm{d}'_{\min})} \left|F_{\mathrm{d}_{\mu_O,m}(\mu_O)}(x) - F_{\mathrm{d}_{\mu_{O'},m}(\mu_{O'})}(x)\right| \mathrm{d}x \geq (\mathrm{d}'_{\min} - \mathrm{d}_{\min})\mu_O\left(O_{\epsilon(m,O)}\right).$$

■

Proof of Example 3.2: Since $\mathrm{Leb}_d(O) = \omega_d$, $\epsilon(m,O) = m^{\frac{1}{d}}$, and $O_{\epsilon(m,O)}$ is the ball of radius $1-\epsilon(m,O)$, thus, $\mu_O\left(O_{\epsilon(m,O)}\right) = \left(1-m^{\frac{1}{d}}\right)^d$.

Also, $\mathrm{Leb}_d(O') = 2^d$, $\epsilon(m,O') = 2\left(\frac{m}{\omega_d}\right)^{\frac{1}{d}}$, and $O'_{\epsilon(m,O')}$ is the hypercube of radius $1-\epsilon(m,O')$, thus, $\mu_{O'}\left(O'_{\epsilon(m,O')}\right) = \left(1-2\left(\frac{m}{\omega_d}\right)^{\frac{1}{d}}\right)^d$.

Thus, when m is smaller than $\frac{\omega_d}{2^d}$, the DTM-signature discriminates between μ_O and $\mu_{O'}$. Moreover, the L_1-Wasserstein distance between the signatures is bounded below by

$$\left(1-2\left(\frac{m}{\omega_d}\right)^{\frac{1}{d}}\right)^d \frac{d}{d+1}\left(\frac{m}{\omega_d}\right)^{\frac{1}{d}}\left(2-\omega_d^{\frac{1}{d}}\right).$$

■

A.3. The DTM-signature to discriminate between uniform and non uniform measures.

Proof of Proposition 3.7: As for Proposition A.1, we get that for any point x in O:

$$\mathrm{d}_{\mu_O,m}(x) \geq \mathrm{d}_{\min} := \left(\frac{m\mathrm{Leb}_d(O)}{\omega_d}\right)^{\frac{1}{d}} \frac{d}{1+d}.$$

We will lower bound the L_1-Wasserstein distance between $\mathrm{d}_{\mu_O,m}(\mu_O)$ and $\mathrm{d}_{\nu,m}(\nu)$ by the integral of $F_{\mathrm{d}_{\nu,m}(\nu)}$ over the interval $[0, \mathrm{d}_{\min}]$, since $F_{\mathrm{d}_{\mu_O,m}(\mu_O)}$ equals zero on this interval. We thus need to lower bound $F_{\mathrm{d}_{\nu,m}(\nu)}(t)$ for all $t \leq \mathrm{d}_{\min}$.

As for Proposition A.1, for $\lambda \geq 1$, any point x of $\{f \geq \lambda\}_{\lambda^{-\frac{1}{d}}\left(\frac{m\mathrm{Leb}_d(O)}{\omega_d}\right)^{\frac{1}{d}}}$ satisfies $\mathrm{d}_{\nu,m}(x) \leq \frac{\mathrm{d}_{\min}}{\lambda^{\frac{1}{d}}}$. Thus,

$$F_{\mathrm{d}_{\nu,m}(\nu)}\left(\frac{\mathrm{d}_{\min}}{\lambda^{\frac{1}{d}}}\right) \geq \nu\left(\{f \geq \lambda\}_{\lambda^{-\frac{1}{d}}\left(\frac{m\mathrm{Leb}_d(O)}{\omega_d}\right)^{\frac{1}{d}}}\right).$$

And we get by denoting $\lambda(t)$ the real number λ satisfying $t = \frac{\mathrm{d}_{\min}}{\lambda^{\frac{1}{d}}}$, that:

$$W_1(\mathrm{d}_{\mu_O,m}(\mu_O), \mathrm{d}_{\nu,m}(\nu)) \geq \int_{t=0}^{\mathrm{d}_{\min}} \nu\left(\{f \geq \lambda(t)\}_{\lambda(t)^{-\frac{1}{d}}\left(\frac{m\mathrm{Leb}_d(O)}{\omega_d}\right)^{\frac{1}{d}}}\right) \mathrm{d}t.$$

Since a cumulative distribution function is non decreasing, we get:

$$\begin{aligned}
&W_1(\mathrm{d}_{\mu_O,m}(\mu_O), \mathrm{d}_{\nu,m}(\nu)) \geq \\
&\int_{t=0}^{\mathrm{d}_{\min}} \sup_{t' \leq t} \nu\left(\{f \geq \lambda(t')\}_{\lambda(t')^{-\frac{1}{d}}\left(\frac{m\mathrm{Leb}_d(O)}{\omega_d}\right)^{\frac{1}{d}}}\right) \mathrm{d}t \\
&= \int_{\lambda=1}^{\infty} \mathrm{d}_{\min} \frac{1}{d} \frac{1}{\lambda^{\frac{1}{d}}} \frac{1}{\lambda} \sup_{\lambda' \geq \lambda} \nu\left(\{f \geq \lambda'\}_{\lambda'^{-\frac{1}{d}}\left(\frac{m\mathrm{Leb}_d(O)}{\omega_d}\right)^{\frac{1}{d}}}\right) \mathrm{d}\lambda \\
&\geq \frac{1}{d+1}\left(\frac{m\mathrm{Leb}_d(O)}{\omega_d}\right)^{\frac{1}{d}} \int_{\lambda=1}^{\infty} \frac{1}{\lambda^{1+\frac{1}{d}}} \sup_{\lambda' \geq \lambda} \lambda' \mu_O\left(\{f \geq \lambda'\}_{\left(\frac{m\mathrm{Leb}_d(O)}{\lambda'\omega_d}\right)^{\frac{1}{d}}}\right) \mathrm{d}\lambda.
\end{aligned}$$

■

Proof of Proposition 3.8: Since $\mu(B(x,r)) \leq \omega_d r^d g_{\max}$, for every $x \in \mathbb{R}^d$, $\mathrm{d}_{\mu,m}(x) \leq d_{\min}$. For every $x \in \{f \geq \lambda\}_{\lambda^{-\frac{1}{d}}\left(\frac{m}{\omega_d}\right)^{\frac{1}{d}}}$, $\nu\left(B\left(x, \lambda^{-\frac{1}{d}}\left(\frac{m}{\omega_d}\right)^{\frac{1}{d}}\right)\right) \geq m$, thus, $\mathrm{d}_{\nu,m}(x) \leq \left(\frac{m}{\lambda\omega_d}\right)^{\frac{1}{d}} \frac{d}{d+1} = \left(\frac{g_{\max}}{\lambda}\right)^{\frac{1}{d}} d_{\min}$.

$$\begin{aligned}
W_1(\mathrm{d}_{\mu,m}(\mu), \mathrm{d}_{\nu,m}(\nu)) &\leq \int_{t=0}^{d_{\min}} \max_{\lambda' \geq \lambda} \nu\left(\{f \geq \lambda'\}_{\lambda'^{-\frac{1}{d}}\left(\frac{m}{\omega_d}\right)^{\frac{1}{d}}}\right) \mathrm{d}t \\
&= \int_{\lambda=g_{\max}}^{+\infty} \frac{d_{\min} g_{\max}^{\frac{1}{d}}}{d\lambda^{\frac{d+1}{d}}} \max_{\lambda' \geq \lambda} \lambda' Leb_d\left(\{f \geq \lambda'\}_{\lambda'^{-\frac{1}{d}}\left(\frac{m}{\omega_d}\right)^{\frac{1}{d}}}\right) \mathrm{d}\lambda
\end{aligned}$$

with $\lambda(t) = g_{\max}\left(\frac{d_{\min}}{t}\right)^d$. ■

Now we assume that the density f is Hölder over O with parameters χ in $[0,1]$ and L positive.

Proof of Proposition 3.9: First notice that for all positive λ, with $\epsilon(\lambda) = \lambda^{-\frac{1}{d}} \left(\frac{m\mathrm{Leb}_d(O)}{\omega_d}\right)^{\frac{1}{d}}$ we have:

$$\{f \geq \lambda + L\epsilon(\lambda)^\chi\} \cap O_{\epsilon(\lambda)} \subset \{f \geq \lambda\}_{\epsilon(\lambda)}.$$

According to Proposition 3.7, the aim is thus to show that for some λ bigger than 1, the set $\{f \geq \lambda + L\epsilon(\lambda)^\chi\} \cap O_{\epsilon(\lambda)}$ is non-empty. We thus focus on the supremum of f over $O_{\epsilon(\lambda)}$, which we denote by $\|f\|_{\infty,\epsilon(\lambda)}$ and try to prove that it is bigger than $\lambda + L\epsilon(\lambda)^\chi$.

Remind that if $\mathrm{Reach}(O) \geq \epsilon(\lambda)$, then thanks to Proposition A.2, the set $O^{\epsilon(\lambda)}_{\epsilon(\lambda)}$ equals O. Since f is Hölder, we can thus build some sequence $(y_n)_{n\in\mathbb{N}}$ in $O_{\epsilon(\lambda)}$, such that $f(y_n) \geq \|f\|_{\infty,O} - \frac{1}{n} - L\epsilon(\lambda)^\chi$. Finally we get:

$$\|f\|_{\infty,\epsilon(\lambda)} \geq \|f\|_{\infty,O} - L\epsilon(\lambda)^\chi.$$

So the quantity $W_1(\mathrm{d}_{\mu_O,m}(\mu_O), \mathrm{d}_{\nu,m}(\nu))$ is positive whenever:

$$\|f\|_{\infty,O} > \inf\left\{\lambda + 2L\epsilon(\lambda)^\chi \,|\, \lambda \geq 1, \epsilon(\lambda) \leq \mathrm{Reach}(O)\right\}.$$

With $\lambda_0 = 1$, we have $\lambda_0 + 2L\epsilon(\lambda_0)^\chi = 1 + 2L\left(\frac{m\mathrm{Leb}_d(O)}{\omega_d}\right)^{\frac{\chi}{d}}$.

With λ_1 satisfying $\epsilon(\lambda_1) = \mathrm{Reach}(O)$, we have:

$$\lambda_1 + 2L\epsilon(\lambda_1)^\chi = \frac{1}{(\mathrm{Reach}(O))^d}\frac{m\mathrm{Leb}_d(O)}{\omega_d} + 2L(\mathrm{Reach}(O)^\chi).$$

We also have that

$$\inf\left\{\lambda + 2L\epsilon(\lambda)^\chi \,|\, \lambda > 0\right\} = (2L)^{\frac{d}{d+\chi}} \left(\frac{\mathrm{Leb}_d(O)}{\omega_d}\right)^{\frac{\chi}{d+\chi}} m^{\frac{\chi}{d+\chi}} \left[\left(\frac{\chi}{d}\right)^{\frac{d}{\chi+d}} + \left(\frac{\chi}{d}\right)^{-\frac{\chi}{d+\chi}}\right].$$

The infimum is attained at $\lambda_2 = \left(\frac{\chi}{d}\right)^{\frac{d}{\chi+d}} (2L)^{\frac{d}{\chi+d}} \left(\frac{m\mathrm{Leb}_d(O)}{\omega_d}\right)^{\frac{\chi}{\chi+d}}$.

It proves the first part of the proposition.

The second part is a straightforward consequence of the proof of Proposition 3.7. ■

Proof of Example 3.8: The measure ν is absolutely continuous with respect to μ_O with density f defined by

$$f(x) = C_\sigma \frac{1}{(2\pi)^{\frac{d}{2}}\sigma^d} \exp -\frac{\|x\|_2^2}{2\sigma^2},$$

with

$$C_\sigma = \frac{1}{\int_{\mathrm{B}(0,1)} \frac{1}{(2\pi)^{\frac{d}{2}}\sigma^d} \exp -\frac{\|x\|_2^2}{2\sigma^2} \mathrm{d}x}.$$

Thanks to the previous proposition, we can prove that the signatures are discriminative provided that σ is small enough, for all m smaller than 0.23 if $d = 1$; 0.30 if $d = 2$; 0.68 if $d = 3$; and for all value of m is $d \geq 4$.

More precisely, the signatures are discriminative when

$$m \leq \sigma^d \frac{\exp\frac{d}{2}}{2^d} \left(1 - \frac{(2\pi)^{\frac{d}{2}}\sigma^d}{C_\sigma}\right)^d,$$

or when

$$m \in \left[\sigma^{d^2+d}\left(\frac{d}{2}\right)^d \frac{(2\pi)^{\frac{d^2}{2}}}{C_\sigma^d} \exp\frac{d}{2}, \min\left\{1, C_\sigma \exp\frac{d}{2}\left(\frac{d}{2}\right)^d \left(\frac{1}{d+1}\right)^{\frac{1}{d+1}}, \frac{2}{d} C_\sigma \frac{\exp -\frac{1}{2}}{(2\pi)^{\frac{d}{2}}\sigma^{d+1}}\right\}\right).$$

When σ is small enough, this last assumption can be rewritten as follows

$$m \in \left[\sigma^{d^2+d}\left(\frac{d}{2}\right)^d \frac{(2\pi)^{\frac{d^2}{2}}}{C_\sigma^d} \exp\frac{d}{2}, \min\left\{1, C_\sigma {C'}_d\right\}\right),$$

with ${C'}_d = \frac{2}{d}\frac{\exp -\frac{1}{2}}{(2\pi)^{\frac{d}{2}}\sigma^{d+1}}$. Note that ${C'}_d$ is much bigger than 1 when $d \geq 4$. Also, ${C'}_1 \simeq 0.23$, ${C'}_2 \simeq 0.30$, ${C'}_3 \simeq 0.68$.

In order to get these results, we use Proposition 3.9. Thanks to the mean value theorem on $\mathrm{B}(0,1)$, which is a convex subset of $\mathbb{R}^d$, the density f is Lipschitz with parameter

$$L = C_\sigma \frac{\exp -\frac{1}{2}}{(2\pi)^{\frac{d}{2}}\sigma^{d+1}}.$$

Thus we use $\chi = 1$ since f is Lipschitz. Moreover, the reach of a ball is equal to its radius, thus $\mathrm{Reach}(O) = 1$, and $\mathrm{Leb}_d(O) = \omega_d$ by definition.

Note that when σ goes to zero, the scaling parameter C_σ goes to 1. ■

Appendix B: Stability of the DTM-signature

Proof of Proposition 3.2: The proof is relatively similar to the ones given by Mémoli in [35] for other signatures.

For any map plan π between μ and ν Borel measures on $(\mathcal{X}, \delta)$ and $(\mathcal{Y}, \gamma)$, we get:

$$
\begin{aligned}
&W_1(\mathrm{d}_{\mu,m}(\mu), \mathrm{d}_{\nu,m}(\nu)) \leq \\
&\int_{\mathcal{X}\times\mathcal{Y}} |\mathrm{d}_{\mu,m}(x) - \mathrm{d}_{\nu,m}(y)| \,\mathrm{d}\pi(x,y) = \\
&\int_{\mathcal{X}\times\mathcal{Y}} \left| \frac{1}{m}\int_0^m \delta_{\mu,l}(x)\mathrm{d}l - \frac{1}{m}\int_0^m \delta_{\nu,l}(y)\mathrm{d}l \right| \mathrm{d}\pi(x,y) \leq \\
&\int_{\mathcal{X}\times\mathcal{Y}} \frac{1}{m}\int_0^m |\delta_{\mu,l}(x) - \delta_{\nu,l}(y)| \,\mathrm{d}l \,\mathrm{d}\pi(x,y) = \\
&\frac{1}{m}\int_{\mathcal{X}\times\mathcal{Y}}\int_0^m \left|\inf\{r>0 \,|\, \mu(\overline{\mathrm{B}}(x,r)) > l\} - \inf\{r>0 \,|\, \nu(\overline{\mathrm{B}}(y,r)) > l\}\right| \mathrm{d}l \,\mathrm{d}\pi(x,y) = \\
&\frac{1}{m}\int_{\mathcal{X}\times\mathcal{Y}}\int_0^m \left| \int_0^{+\infty} \left(\mathbb{1}_{\mu(\overline{\mathrm{B}}(x,r))\leq l} - \mathbb{1}_{\nu(\overline{\mathrm{B}}(y,r))\leq l} \right) \mathrm{d}r \right| \mathrm{d}l \,\mathrm{d}\pi(x,y) \leq \\
&\frac{1}{m}\int_{\mathcal{X}\times\mathcal{Y}}\int_0^{+\infty}\int_0^m \left| \mathbb{1}_{\mu(\overline{\mathrm{B}}(x,r))\leq l} - \mathbb{1}_{\nu(\overline{\mathrm{B}}(y,r))\leq l} \right| \mathrm{d}l \,\mathrm{d}r \,\mathrm{d}\pi(x,y) \leq \\
&\frac{1}{m}\int_{\mathcal{X}\times\mathcal{Y}}\int_0^{+\infty} \left|\mu(\overline{\mathrm{B}}(x,r)) \wedge m - \nu(\overline{\mathrm{B}}(y,r)) \wedge m\right| \mathrm{d}r \,\mathrm{d}\pi(x,y) \leq \\
&\frac{1}{m}\int_{\mathcal{X}\times\mathcal{Y}}\int_0^{+\infty} \left| \int_{\mathcal{X}\times\mathcal{Y}} \left(\mathbb{1}_{\delta(x,x')\leq r} - \mathbb{1}_{\gamma(y,y')\leq r} \right) \mathrm{d}\pi(x',y') \right| \wedge m \,\mathrm{d}r \,\mathrm{d}\pi(x,y) \leq \\
&\frac{1}{m}\int_{\mathcal{X}\times\mathcal{Y}}\int_{\mathcal{X}\times\mathcal{Y}}\int_0^{+\infty} \left| \mathbb{1}_{\delta(x,x')\leq r} - \mathbb{1}_{\gamma(y,y')\leq r} \right| \mathrm{d}r \,\mathrm{d}\pi(x',y') \,\mathrm{d}\pi(x,y) = \\
&\frac{1}{m}\int_{\mathcal{X}\times\mathcal{Y}}\int_{\mathcal{X}\times\mathcal{Y}} |\delta(x,x') - \gamma(y,y')| \,\mathrm{d}\pi(x',y') \,\mathrm{d}\pi(x,y),
\end{aligned}
$$

which concludes.

Appendix C: The test

C.1. A lemma

Lemma C.1 (Equality of empirical signatures under the isomorphic assumption)**.** *If $(\mathcal{X},\delta,\mu)$ and $(\mathcal{Y},\gamma,\nu)$ are two isomorphic mm-spaces, then the distributions of the random variables*

$$\sqrt{n} W_1(\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_n), \mathrm{d}_{\hat{\mu}'_N,m}(\hat{\mu}'_n))$$

and

$$\sqrt{n} W_1(\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_n), \mathrm{d}_{\hat{\nu}_N,m}(\hat{\nu}_n))$$

are equal. Here the empirical measures are all independent and the measures $\hat{\mu}'_N$ and $\hat{\mu}'_n$ are from samples from μ.

Proof. Note that for $(X'_1, X'_2, \ldots, X'_N)$ a N-sample of law μ and ϕ an isomorphism between $(\mathcal{X},\delta,\mu)$ and $(\mathcal{Y},\gamma,\nu)$, the tuple $(\phi(X'_1), \phi(X'_2), \ldots, \phi(X'_N))$ is a N-sample of law ν. Moreover, $\delta(X'_i, X'_j) = \gamma(\phi(X'_i), \phi(X'_j))$ for all i and j in $[\![1, N]\!]$. It follows that the distances and the nearest neighbours are preserved.

Thus, the distributions of $(\mathrm{d}_{\hat{\mu}_N,m}(X'_i))_{i\in[\![1,n]\!]}$ and $(\mathrm{d}_{\hat{\nu}_N,m}(Y_i))_{i\in[\![1,n]\!]}$ are equal.

The lemma follows from the equality:

$$W_1(\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_n), \mathrm{d}_{\hat{\nu}_N,m}(\hat{\nu}_n)) = \int_0^{+\infty} \frac{1}{n} \left| \sum_{i=1}^{n} \mathbb{1}_{\mathrm{d}_{\hat{\mu}_N,m}(X_i)\leq s} - \sum_{i=1}^{n} \mathbb{1}_{\mathrm{d}_{\hat{\nu}_N,m}(Y_i)\leq s} \right| \mathrm{d}s,$$

with $(X_1, X_2, \ldots, X_N)$ a N-sample from μ. □

C.2. L_1-Wasserstein distance between the laws of interest

Lemma C.2. *The quantity* $W_1\left(\mathcal{L}_{N,n,m}(\mu,\mu), \mathcal{L}^*_{N,n,m}(\hat{\mu}_N, \hat{\mu}_N)\right)$ *is bounded above by*

$$2\sqrt{n}\left(\mathbb{E}[\|\mathrm{d}_{\hat{\mu}_N,m} - \mathrm{d}_{\mu,m}\|_{\infty,\mathcal{X}}] + W_1(\mathrm{d}_{\mu,m}(\mu), \mathrm{d}_{\mu,m}(\hat{\mu}_N)) + \|\mathrm{d}_{\mu,m} - \mathrm{d}_{\hat{\mu}_N,m}\|_{\infty,\mathcal{X}}\right).$$

Proof. Let $(X_1, X_2, \ldots X_N)$ be a N-sample of law μ, and $\hat{\mu}_N$ the associated empirical measure. We can upper bound the L_1-Wasserstein distance between the subsampling distribution

$$\mathcal{L}^*(\sqrt{n}W_1(\mathrm{d}_{\hat{\mu}_N,m}(\mu_n^*), \mathrm{d}_{\hat{\mu}_N,m}(\mu_n'^*))|\hat{\mu}_N)$$

and the distribution of interest

$$\mathcal{L}(\sqrt{n}W_1(\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_n), \mathrm{d}_{\hat{\mu}'_N,m}(\hat{\mu}'_n))),$$

by

$$W_1\left(\mathcal{L}\left(\sqrt{n}W_1\left(\mathrm{d}_{\hat{\mu}_N,m}\left(\mu_n^*\right), \mathrm{d}_{\hat{\mu}_N,m}\left(\mu_n'^*\right)\right)|\hat{\mu}_N\right), \mathcal{L}\left(\sqrt{n}W_1\left(\mathrm{d}_{\mu,m}\left(\mu_n^*\right), \mathrm{d}_{\mu,m}\left(\mu_n'^*\right)\right)|\hat{\mu}_N\right)\right) \tag{C.1}$$

$$+ W_1\left(\mathcal{L}\left(\sqrt{n}W_1\left(\mathrm{d}_{\mu,m}\left(\mu_n^*\right), \mathrm{d}_{\mu,m}\left(\mu_n'^*\right)\right)|\hat{\mu}_N\right), \mathcal{L}\left(\sqrt{n}W_1\left(\mathrm{d}_{\mu,m}\left(\hat{\mu}_n\right), \mathrm{d}_{\mu,m}\left(\hat{\mu}'_n\right)\right)\right)\right) \tag{C.2}$$

$$+ W_1\left(\mathcal{L}\left(\sqrt{n}W_1\left(\mathrm{d}_{\mu,m}\left(\hat{\mu}_n\right), \mathrm{d}_{\mu,m}\left(\hat{\mu}'_n\right)\right)\right), \mathcal{L}\left(\sqrt{n}W_1\left(\mathrm{d}_{\hat{\mu}_N,m}\left(\hat{\mu}_n\right), \mathrm{d}_{\hat{\mu}'_N,m}\left(\hat{\mu}'_n\right)\right)\right)\right). \tag{C.3}$$

We bound the term C.1 by

$$2\sqrt{n}\|\mathrm{d}_{\mu,m} - \mathrm{d}_{\hat{\mu}_N,m}\|_{\infty,\mathcal{X}}.$$

the term C.2 by

$$2\sqrt{n}W_1\left(\mathrm{d}_{\mu,m}(\mu), \mathrm{d}_{\mu,m}\left(\hat{\mu}_N\right)\right)$$

and the term C.3 by

$$2\sqrt{n}\mathbb{E}[\|\mathrm{d}_{\mu,m} - \mathrm{d}_{\hat{\mu}_N,m}\|_{\infty,\mathcal{X}}].$$

This is proved in the three following lemmas. □

Lemma C.3 (Study of term C.3)**.** *We have*

$$W_1\left(\mathcal{L}(\sqrt{n}W_1(\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_n), \mathrm{d}_{\hat{\mu}'_N,m}(\hat{\mu}'_n))), \mathcal{L}(\sqrt{n}\,W_1(\mathrm{d}_{\mu,m}(\hat{\mu}_n), \mathrm{d}_{\mu,m}(\hat{\mu}'_n)))\right) \leq$$
$$2\sqrt{n}\mathbb{E}[\|\mathrm{d}_{\mu,m} - \mathrm{d}_{\hat{\mu}_N,m}\|_{\infty,\mathcal{X}}].$$

Proof. To bound this L_1-Wasserstein distance, we choose as a transport plan the law of the random vector

$$(\sqrt{n}W_1(\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_n),\mathrm{d}_{\hat{\mu}'_N,m}(\hat{\mu}'_n)),\sqrt{n}\,W_1(\mathrm{d}_{\mu,m}(\hat{\mu}_n),\mathrm{d}_{\mu,m}(\hat{\mu}'_n))),$$

with $\hat{\mu}_n$, $\hat{\mu}'_n$, $\hat{\mu}_{N-n}$ and $\hat{\mu}'_{N-n}$ independent empirical measures of law μ. Then the L_1-Wasserstein distance is bounded by

$$\mathbb{E}[|\sqrt{n}W_1(\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_n),\mathrm{d}_{\hat{\mu}'_N,m}(\hat{\mu}'_n))-\sqrt{n}\,W_1(\mathrm{d}_{\mu,m}(\hat{\mu}_n),\mathrm{d}_{\mu,m}(\hat{\mu}'_n))|],$$

which is not bigger than:

$$\sqrt{n}\mathbb{E}[W_1(\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_n),\mathrm{d}_{\mu,m}(\hat{\mu}_n))+\,W_1(\mathrm{d}_{\hat{\mu}'_N,m}(\hat{\mu}'_n),\mathrm{d}_{\mu,m}(\hat{\mu}'_n))].$$

We bound the term $\mathbb{E}[W_1(\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_n),\mathrm{d}_{\mu,m}(\hat{\mu}_n))]$ by $\mathbb{E}[\|\mathrm{d}_{\mu,m}-\mathrm{d}_{\hat{\mu}_N,m}\|_{\infty,\mathcal{X}}]$, thanks to Lemma C.6. □

Lemma C.4 (Study of term C.2)**.** *We have*

$$W_1\left(\mathcal{L}(\sqrt{n}W_1(\mathrm{d}_{\mu,m}(\hat{\mu}_n),\mathrm{d}_{\mu,m}(\hat{\mu}'_n)),\mathcal{L}(\sqrt{n}\,W_1(\mathrm{d}_{\mu,m}(\mu^*_n),\mathrm{d}_{\mu,m}(\mu'^*_n))|\hat{\mu}_N)\right)\leq$$
$$2\sqrt{n}W_1(\mathrm{d}_{\mu,m}(\mu),\mathrm{d}_{\mu,m}(\hat{\mu}_N)).$$

Proof. Let π be the optimal transport plan associated to $W_1\left(\mathrm{d}_{\mu,m}(\mu),\mathrm{d}_{\mu,m}(\hat{\mu}_N)\right)$; see the definition of the L_1-Wasserstein with transport plans.

From a n-sample of law π, we get two empirical distributions $\mathrm{d}_{\mu,m}(\hat{\mu}_n)$ and $\mathrm{d}_{\mu,m}(\mu^*_n)$. Independently, from another n-sample of law π, we get $\mathrm{d}_{\mu,m}(\hat{\mu}'_n)$ and $\mathrm{d}_{\mu,m}(\mu'^*_n)$.

The L_1-Wasserstein distance is then bounded by

$$\sqrt{n}\mathbb{E}_{\pi^{\otimes n}\otimes\pi^{\otimes n}}[W_1(\mathrm{d}_{\mu,m}(\hat{\mu}_n),\mathrm{d}_{\mu,m}(\mu^*_n))+\,W_1(\mathrm{d}_{\mu,m}(\hat{\mu}'_n),\mathrm{d}_{\mu,m}(\mu'^*_n))].$$

Now notice that, if we denote $\hat{\mu}_n=\sum_{i=1}^n\frac{1}{n}\delta_{Y_i}$ and $\mu^*_n=\sum_{i=1}^n\frac{1}{n}\delta_{Z_i}$, we have:

$$\begin{aligned}W_1(\mathrm{d}_{\mu,m}(\hat{\mu}_n),\mathrm{d}_{\mu,m}(\mu^*_n))&=\int_{t=0}^{+\infty}\left|\frac{1}{n}\sum_{i=1}^n\mathbb{1}_{\mathrm{d}_{\mu,m}(Y_i)\leq t}-\frac{1}{n}\sum_{i=1}^n\mathbb{1}_{\mathrm{d}_{\mu,m}(Z_i)\leq t}\right|\mathrm{d}t\\&\leq\frac{1}{n}\sum_{i=1}^n\int_{t=0}^{+\infty}\left|\mathbb{1}_{\mathrm{d}_{\mu,m}(Y_i)\leq t}-\mathbb{1}_{\mathrm{d}_{\mu,m}(Z_i)\leq t}\right|\mathrm{d}t\\&=\frac{1}{n}\sum_{i=1}^n\left|\mathrm{d}_{\mu,m}(Y_i)-\mathrm{d}_{\mu,m}(Z_i)\right|.\end{aligned}$$

So, the L_1-Wasserstein distance is not bigger than

$$2\sqrt{n}\mathbb{E}[|\mathrm{d}_{\mu,m}(Y)-\mathrm{d}_{\mu,m}(Z)|],$$

with $(\mathrm{d}_{\mu,m}(Y),\mathrm{d}_{\mu,m}(Z))$ of law π, so we get the upper bound:

$$2\sqrt{n}\left(W_1(\mathrm{d}_{\mu,m}(\mu),\mathrm{d}_{\mu,m}(\hat{\mu}_N)\right).$$

□

Lemma C.5 (Study of term C.1)**.** *We have*

$$W_1\left(\mathcal{L}(\sqrt{n}W_1(\mathrm{d}_{\mu,m}(\mu_n^*),\mathrm{d}_{\mu,m}(\mu_n'^*))|\hat{\mu}_N),\mathcal{L}(\sqrt{n}\,W_1(\mathrm{d}_{\hat{\mu}_N,m}(\mu_n^*),\mathrm{d}_{\hat{\mu}_N,m}(\mu_n'^*))|\hat{\mu}_N)\right)\leq 2\sqrt{n}\|\mathrm{d}_{\mu,m}-\mathrm{d}_{\hat{\mu}_N,m}\|_{\infty,\mathcal{X}}.$$

Proof. It is the same proof as for the first lemma, except that $\hat{\mu}_N$ is fixed. □

Lemma C.6. *Let ν, μ and μ' be some measures over some metric space $(\mathcal{X},\delta)$, we have:*

$$W_1(\mathrm{d}_{\mu,m}(\nu),\mathrm{d}_{\mu',m}(\nu))\leq\int_{\mathcal{X}}|\mathrm{d}_{\mu,m}(x)-\mathrm{d}_{\mu',m}(x)|\mathrm{d}\nu(x)\leq\|\mathrm{d}_{\mu,m}-\mathrm{d}_{\mu',m}\|_{\infty,\mathrm{Supp}(\nu)}.$$

Proof. We chose the transport plan $(\mathrm{d}_{\mu,m}(Y),\mathrm{d}_{\mu',m}(Y))$ for Y of law ν. □

Thanks to Proposition 3.1 and to the fact that the distance to a measure is 1-Lipschitz, we can derive another upper bound depending only on the L_1-Wasserstein distance between the measure μ and its empirical versions:

Corollary C.1. *The quantity $W_1\left(\mathcal{L}_{N,n,m}(\mu,\mu),\mathcal{L}^*_{N,n,m}(\hat{\mu}_N,\hat{\mu}_N)\right)$ is bounded above by*

$$2\frac{\sqrt{n}}{m}\mathbb{E}\left[W_1\left(\hat{\mu}_N,\mu\right)\right]+2\sqrt{n}\left(1+\frac{1}{m}\right)W_1(\hat{\mu}_N,\mu).$$

The rates of convergence of the L_1-Wasserstein distance between a Borel probability measure on the Euclidean space $\mathbb{R}^d$ and its empirical version are faster when the dimension d is low; see [26]. Thus, we prefer to use the first bound for regular measures. In this case, we use rates of convergence for the distance to a measure, derived in [18]. For regular measures, in some cases, the bound in Lemma C.2 is better than the bound in Corollary C.1.

C.3. An asymptotic result with the convergence to the law of $\|\mathbb{G}_{\mu,m}-\mathbb{G}'_{\mu,m}\|_1$

Proof of Lemma 2.1: The random function $\sqrt{n}\left(F_{\mathrm{d}_{\mu,m}(\mu)}-F_{\mathrm{d}_{\mu,m}(\hat{\mu}_n)}\right)$ converges weakly in L_1 to some Gaussian prossess $\mathbb{G}_{\mu,m}$ with covariance kernel

$$\kappa(s,t)=F_{\mathrm{d}_{\mu,m}(\mu)}(s)\left(1-F_{\mathrm{d}_{\mu,m}(\mu)}(t)\right)$$

for $s\leq t$; see [4] or part 3.3 of [8]. Thanks to Theorem 2.8, p.23, in [7], since $L_1\times L_1$ is separable and $\hat{\mu}_n$ and $\hat{\mu}'_n$ are independent, the random vector

$$\left(\sqrt{n}\left(F_{\mathrm{d}_{\mu,m}(\mu)}-F_{\mathrm{d}_{\mu,m}(\hat{\mu}_n)}\right),\sqrt{n}\left(F_{\mathrm{d}_{\mu,m}(\mu)}-F_{\mathrm{d}_{\mu,m}(\hat{\mu}'_n)}\right)\right)$$

converges weakly to $(\mathbb{G}_{\mu,m},\mathbb{G}'_{\mu,m})$ with $\mathbb{G}_{\mu,m}$ and $\mathbb{G}'_{\mu,m}$ independent Gaussian processes. Since the map $(x,y)\mapsto x-y$ is continuous in L_1, the mapping theorem states that $\sqrt{n}\left(F_{\mathrm{d}_{\mu,m}(\hat{\mu}'_n)}-F_{\mathrm{d}_{\mu,m}(\hat{\mu}_n)}\right)$ converges weakly to the Gaussian process $\mathbb{G}_{\mu,m}-\mathbb{G}'_{\mu,m}$ in L_1. Once more we use the mapping theorem with the continuous map $x\mapsto\|x\|_1$ and the definition of the L_1-Wasserstein distance as the L_1-norm of the cumulative distribution functions to get that:

$$\sqrt{n}W_1(\mathrm{d}_{\mu,m}(\hat{\mu}_n),\mathrm{d}_{\mu,m}(\hat{\mu}'_n))\rightsquigarrow\|\mathbb{G}_{\mu,m}-\mathbb{G}'_{\mu,m}\|_1.$$

We then get the convergence of moments following the same method as for Theorem 2.4 in [4]. We have the bound $\mathbb{E}[\|t\mapsto\mathbb{1}_{\mathrm{d}_{\mu,m}(X_i)\leq t}-\mathbb{1}_{\mathrm{d}_{\mu,m}(Y_i)\leq t}\|_1]\leq\mathcal{D}_\mu<\infty$. Moreover,

the random function $\sqrt{n}\left(F_{\mathrm{d}_{\mu,m}(\hat{\mu}'_n)} - F_{\mathrm{d}_{\mu,m}(\hat{\mu}_n)}\right)$ converges weakly to the Gaussian process $\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}$ in L_1. So, thanks to Theorem 5.1 in [1] (cited in [2] p.136), we have:

$$\mathbb{E}[\sqrt{n}W_1(\mathrm{d}_{\mu,m}(\hat{\mu}_n), \mathrm{d}_{\mu,m}(\hat{\mu}'_n))] \to \mathbb{E}[\|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1].$$

We deduce that:

$$W_1\left(\mathcal{L}\left(\sqrt{n}W_1\left(\mathrm{d}_{\mu,m}\left(\hat{\mu}_n\right), \mathrm{d}_{\mu,m}\left(\hat{\mu}'_n\right)\right)\right), \mathcal{L}\left(\|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1\right)\right) \to 0.$$

Moreover, we have the bound:

$$W_1\left(\mathcal{L}\left(\sqrt{n}W_1\left(\mathrm{d}_{\mu,m}\left(\hat{\mu}_n\right), \mathrm{d}_{\mu,m}\left(\hat{\mu}'_n\right)\right)\right), \mathcal{L}_{N,n,m}(\mu,\mu)\right) \leq 2\sqrt{n}\mathbb{E}[\|\mathrm{d}_{\mu,m} - \mathrm{d}_{\hat{\mu}_N,m}\|_{\infty,\mathcal{X}}].$$

So, if $\sqrt{n}\mathbb{E}\left[\|\mathrm{d}_{\mu,m} - \mathrm{d}_{\hat{\mu}_N,m}\|_{\infty,\mathcal{X}}\right] \to 0$ when $N \to \infty$, we have that:

$$W_1\left(\mathcal{L}_{N,n,m}(\mu,\mu), \mathcal{L}\left(\|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1\right)\right) \to 0.$$

Finally, with the same arguments as for Lemma C.2, we get that:

$$\begin{aligned}
&W_1\left(\mathcal{L}^*_{N,n,m}(\hat{\mu}_N, \hat{\mu}_N), \mathcal{L}\left(\|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1\right)\right) \leq \\
&W_1\left(\mathcal{L}\left(\sqrt{n}W_1\left(\mathrm{d}_{\mu,m}\left(\hat{\mu}_n\right), \mathrm{d}_{\mu,m}\left(\hat{\mu}'_n\right)\right)\right), \mathcal{L}\left(\|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1\right)\right) \\
&+ 2\sqrt{n}W_1\left(\mathrm{d}_{\mu,m}(\mu), \mathrm{d}_{\mu,m}\left(\hat{\mu}_N\right)\right) + 2\sqrt{n}\|\mathrm{d}_{\mu,m} - \mathrm{d}_{\hat{\mu}_N,m}\|_{\infty,\mathcal{X}}.
\end{aligned}$$

Since μ is compactly-supported $\mathrm{d}_{\mu,m}(\mu)$ is also compactly supported. Moreover, Theorem 3.2 of Bobkov and Ledoux [8] states that for any probability P on $\mathbb{R}$ with cumulative distribution function given by F_P, $\sqrt{N}\mathbb{E}\left[W_1(P, P_N)\right] \leq \int_{-\infty}^{+\infty} \sqrt{F_P(s)(1-F_P(s))}\mathrm{d}s$. As a consequence, since $\frac{n}{N} = o(1)$, the expectation $\sqrt{n}\mathbb{E}\left[W_1(\mathrm{d}_{\mu,m}(\mu), \mathrm{d}_{\mu,m}(\hat{\mu}_N))\right]$ converges to zero.

Finally, if $\mathbb{E}\left[\sqrt{n}\|\mathrm{d}_{\mu,m} - \mathrm{d}_{\hat{\mu}_N,m}\|_{\infty,\mathcal{X}}\right]$ converge to 0, using the Markov inequality, we get that $W_1\left(\mathcal{L}^*_{N,n,m}(\hat{\mu}_N, \hat{\mu}_N), \mathcal{L}\left(\|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1\right)\right)$ converges to 0 in probability. ■

Proof of Lemma 2.2: Let $\epsilon < \alpha$ and η be two positive numbers.

The probability $\mathbb{P}_{(\mu,\nu)}\left(\phi_{N,n,m} = 1\right)$ is bounded above by

$$\mathbb{P}\left(\sqrt{n}W_1\left(\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_n), \mathrm{d}_{\hat{\nu}_N,m}(\hat{\nu}_n)\right) \geq \mathrm{q}_{1-(\alpha+\epsilon)} - \eta\right) + \mathbb{P}\left(\hat{\mathrm{q}}_{1-\alpha} < \mathrm{q}_{1-(\alpha+\epsilon)} - \eta\right).$$

For convenience, we use the notation $\mathcal{L} = \mathcal{L}\left(\frac{1}{2}\|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1 + \frac{1}{2}\|\mathbb{G}_{\nu,m} - \mathbb{G}'_{\nu,m}\|_1\right)$ and $\mathcal{L}^* = \frac{1}{2}\mathcal{L}^*_{N,n,m}(\hat{\mu}_N, \hat{\mu}_N) + \frac{1}{2}\mathcal{L}^*_{N,n,m}(\hat{\nu}_N, \hat{\nu}_N)$. Then, $\mathrm{q}_{1-(\alpha+\epsilon)}$ is the $1-(\alpha+\epsilon)$-quantile of $\mathcal{L}$ and $\hat{\mathrm{q}}_{1-\alpha}$, the $1-\alpha$-quantile of $\mathcal{L}^*$. If $\hat{\mathrm{q}}_{1-\alpha} < \mathrm{q}_{1-(\alpha+\epsilon)} - \eta$, then,

$$\begin{aligned}
W_1\left(\mathcal{L}, \mathcal{L}^*\right) &\geq \int_{\hat{\mathrm{q}}_{1-\alpha}}^{\mathrm{q}_{1-(\alpha+\epsilon)}} \left|F_{\mathcal{L}}(x) - F_{\mathcal{L}^*}(x)\right| \mathrm{d}x \\
&\geq \left(\mathrm{q}_{1-(\alpha+\epsilon)} - \hat{\mathrm{q}}_{1-\alpha}\right)\left((1-\alpha) - (1-\alpha-\epsilon)\right) \geq \epsilon\eta
\end{aligned}$$

since for $x \geq \hat{\mathrm{q}}_{1-\alpha}$, $F_{\mathcal{L}^*}(x) \geq 1-\alpha$, and for $x < \mathrm{q}_{1-(\alpha+\epsilon)}$, $F_{\mathcal{L}}(x) \leq 1-\alpha-\epsilon$.

Then, it comes that $\mathbb{P}\left(\hat{\mathrm{q}}_{1-\alpha} < \mathrm{q}_{1-\alpha+\epsilon} - \eta\right)$ is bounded above by $\mathbb{P}\left(W_1\left(\mathcal{L}, \mathcal{L}^*\right) \geq \epsilon\eta\right)$.

Since $W_1\left(\mathcal{L}, \mathcal{L}^*\right) \leq B_{N,n}$ with

$$B_{N,n} := \frac{1}{2}W_1(\mathcal{L}^*_{N,n,m}(\hat{\mu}_N, \hat{\mu}_N), \|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1) + \frac{1}{2}W_1(\mathcal{L}^*_{N,n,m}(\hat{\nu}_N, \hat{\nu}_N), \|\mathbb{G}_{\nu,m} - \mathbb{G}'_{\nu,m}\|_1),$$

it comes that $\mathbb{P}\left(\hat{\mathrm{q}}_{1-\alpha} < \mathrm{q}_{1-\alpha+\epsilon} - \eta\right) \leq P(B_{N,n} \geq \epsilon\eta)$, which goes to zero when N goes to ∞, since the second convergence in Lemma 2.1 is satisfied.

Finally, under the H_0 hypothesis, the distribution of $\sqrt{n}W_1\left(\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_n), \mathrm{d}_{\hat{\nu}_N,m}(\hat{\nu}_n)\right)$ is given by $\mathcal{L}_{N,n,m}(\mu,\mu)$, $\mathcal{L}_{N,n,m}(\nu,\nu)$ or equivalently by $\frac{1}{2}\mathcal{L}_{N,n,m}(\mu,\mu)+\frac{1}{2}\mathcal{L}_{N,n,m}(\nu,\nu)$. Then, the fact that the first convergence in Lemma 2.1 occurs and the Portmanteau lemma entail that $\limsup_{N\to\infty}\mathbb{P}_{(\mu,\nu)}\left(\phi_{N,n,m}=1\right)$ is bounded above by $\mathbb{P}\left(Z \geq \mathrm{q}_{1-(\alpha+\epsilon)} - \eta\right)$, with Z a random variable from the distribution $\frac{1}{2}\mathcal{L}(\|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1) + \frac{1}{2}\mathcal{L}(\|\mathbb{G}_{\nu,m} - \mathbb{G}'_{\nu,m}\|_1)$, which has the same distribution as $\mathcal{L}(\|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1)$ and $\mathcal{L}(\|\mathbb{G}_{\nu,m} - \mathbb{G}'_{\nu,m}\|_1)$ under hypothesis H_0.

We now make η and ϵ go to zero and under the continuity assumption,

$$\limsup_{N\to\infty}\mathbb{P}_{(\mu,\nu)}\left(\phi_{N,n,m}=1\right) \leq \alpha.$$

As well, we get that $\liminf_{N\to\infty}\mathbb{P}_{(\mu,\nu)}\left(\phi_{N,n,m}=1\right) \geq \alpha$.

C.4. The case of measures supported on a compact subset of $\mathbb{R}^d$

Proof of part 1 of Proposition 2.1: We need to show that under the assumption $\rho > \frac{\max\{d,2\}}{2}$,

$$\sqrt{n}\mathbb{E}[\|\mathrm{d}_{\mu,m} - \mathrm{d}_{\hat{\mu}_N,m}\|_{\infty,\mathcal{X}}] \to 0.$$

First consider $d > 2$.

Thanks to Theorem 1 of [26], there is some positive constant C depending on μ such that for N big enough:

$$\mathbb{E}[W_1(\hat{\mu}_N,\mu)] \leq CN^{-\frac{1}{d}}.$$

Then, according to Proposition 3.1, $\sqrt{n}\mathbb{E}\|\mathrm{d}_{\mu,m} - \mathrm{d}_{\hat{\mu}_N}\|_\infty \leq \frac{\sqrt{n}}{m}CN^{-\frac{1}{d}}$. It converges to 0 when $\rho > \frac{d}{2}$. For $d \leq 2$, we use the other bounds from Fournier and Guillin, the condition is then $\rho > 1$. ■

Proof of part 2 of Proposition 2.1: We may assume that the diameter $\mathcal{D}_\mu$ of the support of the measure μ equals 1. Indeed, if we apply a dilatation to the measure to make the diameter of its support be equal to 1, then the quantity $W_1\left(\mathcal{L}_{N,n,m}(\mu,\mu), \mathcal{L}^*_{N,n,m}(\hat{\mu}_N,\hat{\mu}_N)\right)$ is simply multiplied by the parameter of the dilatation. By using Corollary C.1 and Theorem 1 of [26], we have a bound for the expectation:

$$\mathbb{E}\left[W_1\left(\mathcal{L}_{N,n,m}(\mu,\mu), \mathcal{L}^*_{N,n,m}(\hat{\mu}_N,\hat{\mu}_N)\right)\right] \leq \begin{cases} C\frac{\sqrt{n}}{m}N^{-\frac{1}{d}} & \text{if d>2} \\ C\frac{\sqrt{n}}{m}N^{-\frac{1}{2}}\log(1+N) & \text{if d=2} \\ C\frac{\sqrt{n}}{m}N^{-\frac{1}{2}} & \text{if d<2} \end{cases}$$

for some positive constant C depending on μ. ■

Proof of part 3 of Proposition 2.1: First notice that for $\lambda > 1$,

$$\mathbb{P}\left(W_1\left(\mathcal{L}_{N,n,m}(\mu,\mu), \mathcal{L}^*_{N,n,m}(\hat{\mu}_N,\hat{\mu}_N)\right) \geq \lambda\right) = 0$$

under the assumption $\mathcal{D}_\mu = 1$. We thus focus on values of λ not bigger than 1. In this case, with the Theorem 2 of [26], we get easily that:

$$\mathbb{P}\left(W_1\left(\mathcal{L}_{N,n,m}(\mu,\mu),\mathcal{L}^*_{N,n,m}(\hat{\mu}_N,\hat{\mu}_N)\right)\geq\lambda\right)\leq$$
$$\begin{cases} C\exp\left(-C'\left(\lambda\frac{N^{\frac{1}{d}}m}{\sqrt{n}}-C''\right)^d\right) & \text{for d>2}\\ C\exp\left(-C'\left(\frac{\frac{\sqrt{N}m}{\sqrt{n}}\lambda-C''\sqrt{\frac{N}{N-n}}\log(1+N-n)}{\log\left(2+\frac{2\sqrt{N}}{\frac{\sqrt{N}m}{\sqrt{n}}\lambda-C''\sqrt{\frac{N}{N-n}}\log(1+N-n)}\right)}\right)^2\right) & \text{for d=2}\\ C\exp\left(-C'\left(\lambda\frac{\sqrt{N}m}{\sqrt{n}}-C''\right)^2\right) & \text{for d<2}\end{cases}$$

for some positive constants C, C' and C'' depending on μ.
We conclude the proof with the Borel–Cantelli Lemma. ■

C.5. The case of (a,b)-standard measures

Let μ be a Borel probability measure supported on a connected compact subset $\mathcal{X}$ of $\mathbb{R}^d$. We assume this measure to be (a,b)-standard for some positive numbers a and b. In this part, we derive rates of convergence in probability and in expectation for the quantity $\|\mathrm{d}_{\hat{\mu}_N,m}-\mathrm{d}_{\mu,m}\|_{\infty,\mathcal{X}}$. Thanks to these results, we can derive upper bounds and rates of convergence in expectation for $W_1\left(\mathcal{L}_{N,n,m}(\mu,\mu),\mathcal{L}^*_{N,n,m}(\hat{\mu}_N,\hat{\mu}_N)\right)$. We finally propose a choice for the parameter N depending on n for which the weak convergence $\mathcal{L}_{N,n,m}(\mu,\mu)\rightsquigarrow\|\mathbb{G}_{\mu,m}-\mathbb{G}'_{\mu,m}\|_1$ and the convergence in probability of $W_1\left(\mathcal{L}^*_{N,n,m}(\hat{\mu}_N,\hat{\mu}'_N),\|\mathbb{G}_{\mu,m}-\mathbb{G}'_{\mu,m}\|_1\right)$ to zero occur.

C.5.1. Upper bounds for $\mathbb{P}(\sqrt{n}\|\mathrm{d}_{\hat{\mu}_N,m}-\mathrm{d}_{\mu,m}\|_{\infty,\mathcal{X}}\geq\lambda)$

We use the bounds given in Theorem 1 of [18], with the bound for the modulus of continuity given by **Lemma 3** in [18]: $\omega(h)=\left(\frac{h}{a}\right)^{\frac{1}{b}}$. We directly get the following lemma:

Lemma C.7 (Upper bound for $|\mathrm{d}_{\hat{\mu}_N,m}(x)-\mathrm{d}_{\mu,m}(x)|$)**.** *Let x be a fixed point in $\mathcal{X}$ and λ a positive number. We have,*

$$\frac{1}{2}\mathbb{P}(|\mathrm{d}_{\hat{\mu}_N,m}(x)-\mathrm{d}_{\mu,m}(x)|\geq\lambda)\leq$$
$$\exp\left(-2a^{\frac{2}{b}}Nm^{\frac{2b-2}{b}}\lambda^2\right)+\exp\left(-\frac{a}{2^{b-1}}N^{\frac{b+1}{2}}m^b\lambda^b\right)+\exp\left(-a^{\frac{1}{b}}N^{\frac{b+1}{2b}}m\lambda\right).$$

In order to derive an upper bound for $\|\mathrm{d}_{\hat{\mu}_N,m}-\mathrm{d}_{\mu,m}\|_{\infty,\mathcal{X}}$, like in [18], we use the fact that the function distance to a measure is 1-Lipschitz and that $\mathcal{X}$ is compact, which means that we can compute a bound by upper-bounding the difference $|\mathrm{d}_{\hat{\mu}_N,m}(x)-\mathrm{d}_{\mu,m}(x)|$ over a finite number of points x of $\mathcal{X}$. Thanks to the following lemma, the minimal number of points needed for this purpose is not bigger than $\frac{(4\mathcal{D}_\mu\sqrt{d}+\lambda)^d}{\lambda^d}$:

Lemma C.8. *Let μ is a measure supported on $\mathcal{X}$ a compact subset of $\mathbb{R}^d$, and for $\lambda>0$ denote $N(\mu,\lambda)=\inf\{N\in\mathbb{N},\ \exists\ x_1,x_2\dots x_N\in\mathcal{X},\ \bigcup_{i\in[\![1,N]\!]}\mathrm{B}(x_i,\lambda)\supset\mathcal{X}\}$. Then, we have:*

$$N(\mu,\lambda)\leq\frac{\left(\mathcal{D}_\mu\sqrt{d}+\lambda\right)^d}{\lambda^d}.$$

Proof. The idea is to put a grid on the hypercube containing $\mathcal{X}$ with edges of length $\mathcal{D}_\mu$. The grid is a union of small hypercubes with edges of length equal to $\frac{\lambda}{\sqrt{d}}$, so that the number of such small hypercubes into which the big one is split is not superior to $\left(\frac{\mathcal{D}_\mu\sqrt{d}}{\lambda}+1\right)^d$.

Then, we decide that each time the intersection between $\mathcal{X}$ and some small hypercube is non-empty, we keep one of the elements of the intersection. We denote x_i the element associated to the i-th hypercube. Finally, each point x in $\mathcal{X}$ belongs to a small hypercube, and its distance to the corresponding x_i is smaller than $\sqrt{\sum_{k=1}^d \frac{\lambda^2}{d}} = \lambda$. □

We thus derive upper bounds for $\|\mathrm{d}_{\hat{\mu}_N,m} - \mathrm{d}_{\mu,m}\|_{\infty,\mathcal{X}}$:

Proposition C.1 (Upper bound for $\|\mathrm{d}_{\hat{\mu}_N,m} - \mathrm{d}_{\mu,m}\|_{\infty,\mathcal{X}}$)**.** *We have,*

$$\frac{\lambda^d}{2\left(4\mathcal{D}_\mu\sqrt{d}+\lambda\right)^d}\ \mathbb{P}(\|\mathrm{d}_{\hat{\mu}_N,m} - \mathrm{d}_{\mu,m}\|_{\infty,\mathcal{X}} \geq \lambda) \leq$$
$$\exp\left(-\frac{a^{\frac{2}{b}}}{2}Nm^{\frac{2b-2}{b}}\lambda^2\right) + \exp\left(-\frac{a}{2^{2b-1}}N^{\frac{b+1}{2}}m^b\lambda^b\right) + \exp\left(-\frac{a^{\frac{1}{b}}}{2}N^{\frac{b+1}{2b}}m\lambda\right).$$

Proof. Since the function distance to a measure is 1-Lipschitz, we get that:

$$\|\mathrm{d}_{\hat{\mu}_N,m} - \mathrm{d}_{\mu,m}\|_{\infty,\mathcal{X}} \leq \frac{\lambda}{2} + \sup_i\{|\mathrm{d}_{\hat{\mu}_N,m}(x_i) - \mathrm{d}_{\mu,m}(x_i)|\},$$

for the family $(x_i)_i$ associated to a grid which sides are of length equal to $\frac{\lambda}{4\sqrt{d}}$. We can thus bound the probability $\mathbb{P}(\|\mathrm{d}_{\hat{\mu}_N,m} - \mathrm{d}_{\mu,m}\|_{\infty,\mathcal{X}} \geq \lambda)$ by

$$\sum_{i=1}^{N\left(\mu,\frac{\lambda}{4}\right)} \mathbb{P}\left(|\mathrm{d}_{\hat{\mu}_N,m}(x_i) - \mathrm{d}_{\mu,m}(x_i)| \geq \frac{\lambda}{2}\right),$$

with $N\left(\mu,\frac{\lambda}{4}\right) \leq \frac{\left(4\mathcal{D}_\mu\sqrt{d}+\lambda\right)^d}{\lambda^d}$ thanks to Lemma C.8. □

C.5.2. Upper bounds for the expectation $\mathbb{E}[\|\mathrm{d}_{\hat{\mu}_N,m} - \mathrm{d}_{\mu,m}\|_{\infty,\mathcal{X}}]$

In order to get upper bounds for $\mathbb{E}[\|\mathrm{d}_{\hat{\mu}_N,m} - \mathrm{d}_{\mu,m}\|_{\infty,\mathcal{X}}]$, we use the same trick as used in [18], which is:

Lemma C.9. *Let X a random variable such that:*

$$\mathbb{P}(X \geq \lambda) \leq 1 \wedge D\lambda^{-q}\exp(-c\lambda^s)$$

for some integers q and s and some $D > 0$.

We have:

$$\mathbb{E}[X] \leq \left(\frac{\ln c}{c}\right)^{\frac{1}{s}}\left(\frac{q}{s}\right)^{\frac{1}{s}}\left[1 + D\left(\frac{q}{s}\right)^{\frac{-q-s}{s}}\frac{(\ln c)^{\frac{-q-s}{s}}}{s}\right].$$

More particularly, if $c \geq \exp D^{\frac{s}{q+s}}\frac{s}{q}$, then:

$$\mathbb{E}[X] \leq 2\left(\frac{\ln c}{c}\right)^{\frac{1}{s}}\left(\frac{q}{s}\right)^{\frac{1}{s}}.$$

Proof. For any $\lambda_0 > 0$, that we can choose as $\lambda_0 = \frac{[\ln K]^{\frac{1}{s}}}{c^{\frac{1}{s}}}$, we get that:

$$\begin{aligned}
\mathbb{E}[X] &\leq \lambda_0 + \int_{\lambda_0}^{\infty} D\lambda^{-q} \exp(-c\lambda^s) \mathrm{d}\lambda \\
&\leq \lambda_0 + D\frac{\lambda_0^{-q-s+1}}{cs} \exp -c\lambda_0^s \\
&= \frac{[\ln K]^{\frac{1}{s}}}{c^{\frac{1}{s}}} + D\frac{[\ln K]^{\frac{-q-s+1}{s}}}{scc^{\frac{-q-s+1}{s}}} \frac{1}{K} \\
&= \frac{[\ln K]^{\frac{1}{s}}}{c^{\frac{1}{s}}} + \frac{[\ln K]^{\frac{1}{s}}}{c^{\frac{1}{s}}} D\frac{[\ln K]^{\frac{-q-s}{s}}}{sc^{\frac{-q}{s}}} \frac{1}{K} \\
&= \frac{[\ln K]^{\frac{1}{s}}}{c^{\frac{1}{s}}} \left[1 + D\frac{[\ln K]^{\frac{-q-s}{s}}}{sKc^{\frac{-q}{s}}}\right]
\end{aligned}$$

Finally, if we choose $K = c^{\frac{q}{s}}$, we get:

$$\mathbb{E}[X] \leq \left(\frac{q}{s}\right)^{\frac{1}{s}} \left[\frac{\ln c}{c}\right]^{\frac{1}{s}} \left[1 + D\left[\frac{q}{s}\right]^{\frac{-q-s}{s}} \frac{(\ln c)^{\frac{-q-s}{s}}}{s}\right].$$

□

From this lemma, we can derive the following lemma.

Lemma C.10. *We have,*

$$\begin{aligned}
&\mathbb{E}[\|\mathrm{d}_{\hat{\mu}_N,m} - \mathrm{d}_{\mu,m}\|_{\infty,\mathcal{X}}] \leq \\
&\square'_1 \frac{1}{N^{\frac{1}{2}} m^{\frac{b-1}{b}}} \left(\log\left(Nm^{\frac{2b-2}{b}}\right)\right)^{\frac{1}{2}} + \\
&\square'_2 \frac{1}{N^{\frac{b+1}{2b}} m} \left(\log\left(N^{\frac{b+1}{2}} m^b\right)\right)^{\frac{1}{b}} + \\
&\square'_3 \frac{1}{N^{\frac{b+1}{2b}} m} \log\left(N^{\frac{b+1}{2b}} m\right).
\end{aligned}$$

for some constants □ *depending on* a *and* b*. And hence,*

$$\begin{aligned}
&\sqrt{n}\mathbb{E}[\|\mathrm{d}_{\hat{\mu}_N,m} - \mathrm{d}_{\mu,m}\|_{\infty,\mathcal{X}}] \leq \\
&\square'_1 \frac{\sqrt{n}}{N^{\frac{1}{2}} m^{\frac{b-1}{b}}} \left(\log\left(Nm^{\frac{2b-2}{b}}\right)\right)^{\frac{1}{2}} + \\
&\square'_2 \frac{\sqrt{n}}{N^{\frac{b+1}{2b}} m} \left(\log\left(N^{\frac{b+1}{2}} m^b\right)\right)^{\frac{1}{b}} + \\
&\square'_3 \frac{\sqrt{n}}{N^{\frac{b+1}{2b}} m} \log\left(N^{\frac{b+1}{2b}} m\right).
\end{aligned}$$

C.5.3. Upper bounds for the expectation of $W_1\left(\mathcal{L}_{N,n,m}(\mu,\mu), \mathcal{L}^*_{N,n,m}(\hat{\mu}_N, \hat{\mu}_N)\right)$

Proof of part 2 of Proposition 2.2: For all $\lambda > 0$, for any (a,b)-standard measure μ supported on a connected compact subset of $\mathbb{R}^d$, we can use Lemma C.2 and Lemma C.10

together with the rates of convergence of the L_1-Wasserstein distance between empirical and true distribution in [8] to get the following result.

If $m \geq \frac{1}{2}$, then for n big enough we have, for some constants $\square$ depending on a and b:

$$\begin{aligned}
&\mathbb{E}\left[W_1\left(\mathcal{L}_{N,n,m}(\mu,\mu), \mathcal{L}^*_{N,n,m}(\hat{\mu}_N,\hat{\mu}_N)\right)\right] \leq \\
&\square'_1 \frac{n^{\frac{1}{2}}}{(N)^{\frac{1}{2}} m^{\frac{b-1}{b}}} \left(\log\left(N m^{\frac{2b-2}{b}}\right)\right)^{\frac{1}{2}} \\
&+ \square'_2 \frac{n^{\frac{1}{2}}}{(N)^{\frac{b+1}{2b}} m} \left(\log\left(N^{\frac{b+1}{2}} m^b\right)\right)^{\frac{1}{b}} \\
&+ \square'_3 \frac{n^{\frac{1}{2}}}{(N)^{\frac{b+1}{2b}} m} \log\left(N^{\frac{b+1}{2b}} m\right) \\
&+ \square'_4 \frac{n^{\frac{1}{2}}}{N^{\frac{1}{2}}}.
\end{aligned}$$

C.5.4. Convergence to the law of $\|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1$

Proof of part 1 of Proposition 2.2: In order to get these two results, we use Lemma 2.1. The convergence to zero of $\sqrt{n}\mathbb{E}[\|\mathrm{d}_{\mu,m} - \mathrm{d}_{\hat{\mu}_{N-n},m}\|_{\infty,\mathcal{X}}]$ is a direct consequence of Lemma C.10. ■

C.6. The power of the test

Proof of Theorem 2.3

Lemma C.11. *Let α, κ be two positive numbers and $\mathcal{L}$ and $\mathcal{L}^*$ two laws of real random variables. We denote $\mathrm{q}_{1-\alpha}$ (respectively $\mathrm{q}^*_{1-\alpha}$) the $1-\alpha$-quantile of the law $\mathcal{L}$ (respectively $\mathcal{L}^*$). If $W_1(\mathcal{L},\mathcal{L}^*) < \kappa$ then:*

$$\mathrm{q}^*_{1-\alpha} \leq 2\frac{\kappa}{\alpha} + \mathrm{q}_{1-\frac{\alpha}{2}}.$$

Proof. Suppose that $\mathrm{q}^*_{1-\alpha} > 2\frac{\kappa}{\alpha} + \mathrm{q}_{1-\frac{\alpha}{2}}$. Then,

$$\begin{aligned}
W_1(\mathcal{L},\mathcal{L}^*) &\geq \int_{\mathrm{q}_{1-\frac{\alpha}{2}}}^{\mathrm{q}^*_{1-\alpha}} |F_{\mathcal{L}}(x) - F_{\mathcal{L}^*}(x)|\,\mathrm{d}x \\
&\geq \int_{\mathrm{q}_{1-\frac{\alpha}{2}}}^{\mathrm{q}^*_{1-\alpha}} \left(1 - \frac{\alpha}{2} - (1-\alpha)\right)\mathrm{d}x \geq \kappa.
\end{aligned}$$

□

In this part we assume that m is fixed in $[0,1]$ and $N = cn^\rho$ for some $\rho > 1$ and $c > 0$. Recall that our aim is to bound above the type II error, that is:

$$\mathbb{P}_{(\mu,\nu)}\left(\sqrt{n} W_1\left(\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_n), \mathrm{d}_{\hat{\nu}_N,m}(\hat{\nu}_n)\right) < \hat{\mathrm{q}}_{1-\alpha}\right).$$

For some $\kappa = n^\gamma$ with γ in $\left[0, \frac{1}{2}\right)$ to be chosen later, we first bound above the quantile $\hat{\mathrm{q}}_{1-\alpha}$ with high probability.

As noticed in the proof of Lemma 2.1, the law of $\sqrt{n}W_1\left(\mathrm{d}_{\mu,m}(\hat{\mu}_n), \mathrm{d}_{\mu,m}(\hat{\mu}'_n)\right)$ converges to $\mathcal{L}(\|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1)$, there is also the convergence of the first moments. So, for n big enough, we have:

$$W_1(\mathcal{L}\left(\sqrt{n}W_1\left(\mathrm{d}_{\mu,m}(\hat{\mu}_n), \mathrm{d}_{\mu,m}(\hat{\mu}'_n)\right)\right), \mathcal{L}(\|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1)) \leq 1.$$

Then, under the assumption

$$W_1(\mathcal{L}\left(\sqrt{n}W_1\left(\mathrm{d}_{\mu,m}(\hat{\mu}_n), \mathrm{d}_{\mu,m}(\hat{\mu}'_n)\right)\right), \mathcal{L}^*_{N,n,m}(\hat{\mu}_N, \hat{\mu}_N)) \leq \kappa,$$

we have

$$W_1(\mathcal{L}(\|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1), \mathcal{L}^*_{N,n,m}(\hat{\mu}_N, \hat{\mu}_N)) \leq \kappa + 1.$$

We can do the same thing for ν. Thus we get that for n big enough and under the previous assumptions:

$$W_1\left(\frac{1}{2}\mathcal{L}(\|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1) + \frac{1}{2}\mathcal{L}(\|\mathbb{G}_{\nu,m} - \mathbb{G}'_{\nu,m}\|_1), \frac{1}{2}\mathcal{L}^*_{N,n,m}(\hat{\mu}_N, \hat{\mu}_N) + \frac{1}{2}\mathcal{L}^*_{N,n,m}(\hat{\nu}_N, \hat{\nu}_N)\right)$$

is bounded above with $\kappa + 1$. And thanks to Lemma C.11,

$$\hat{\mathrm{q}}_{1-\alpha} \leq \tilde{\mathrm{q}}_{1-\frac{\alpha}{2}} + 2\frac{\kappa+1}{\alpha},$$

with $\tilde{\mathrm{q}}_{1-\alpha}$ the $1-\alpha$-quantile of the law $\frac{1}{2}\mathcal{L}(\|\mathbb{G}_{\mu,m} - \mathbb{G}'_{\mu,m}\|_1) + \frac{1}{2}\mathcal{L}(\|\mathbb{G}_{\nu,m} - \mathbb{G}'_{\nu,m}\|_1)$.

We need to notice that with similar arguments as for Lemma C.2, we have:

$$W_1\left(\mathcal{L}\left(\sqrt{n}W_1\left(\mathrm{d}_{\mu,m}(\hat{\mu}_n), \mathrm{d}_{\mu,m}(\hat{\mu}'_n)\right)\right), \mathcal{L}^*_{N,n,m}(\hat{\mu}_N, \hat{\mu}_N)\right) \leq$$
$$2\sqrt{n}\mathcal{D}_\mu \|F_{\mathrm{d}_{\mu,m}(\mu)} - F_{\mathrm{d}_{\mu,m}(\hat{\mu}_N)}\|_{\infty,(0,\mathcal{D}_\mu)} + 2\frac{\sqrt{n}}{m}W_1(\mu, \hat{\mu}_N).$$

Now notice that

$$\sqrt{n}W_1\left(\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_n), \mathrm{d}_{\hat{\nu}_N,m}(\hat{\nu}_n)\right) \geq \sqrt{n}W_1\left(\mathrm{d}_{\mu,m}(\mu), \mathrm{d}_{\nu,m}(\nu)\right)$$
$$- \sqrt{n}W_1\left(\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_n), \mathrm{d}_{\mu,m}(\mu)\right) - \sqrt{n}W_1\left(\mathrm{d}_{\hat{\nu}_N,m}(\hat{\nu}_n), \mathrm{d}_{\nu,m}(\nu)\right),$$

but as well, thanks to Lemma C.6, the definition of the L_1-Wasserstein distance as the L_1-norm between the cumulative distribution functions and to Proposition 3.1:

$$\sqrt{n}W_1\left(\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_n), \mathrm{d}_{\mu,m}(\mu)\right) \leq$$
$$\frac{\sqrt{n}}{m}W_1(\mu, \hat{\mu}_N) + \sqrt{n}\mathcal{D}_{\mu,m}\|F_{\mathrm{d}_{\mu,m}(\hat{\mu}_n)} - F_{\mathrm{d}_{\mu,m}(\mu)}\|_{\infty,(0,\mathcal{D}_\mu)},$$

with $\mathcal{D}_{\mu,m}$ the diameter of the support of the measure $\mathrm{d}_{\mu,m}(\mu)$. So, we can finally upper

bound $\mathbb{P}_{(\mu,\nu)}\left(\sqrt{n}W_1\left(\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_n),\mathrm{d}_{\hat{\nu}_N,m}(\hat{\nu}_n)\right)<\hat{\mathrm{q}}_{1-\alpha}\right)$ by

$$\mathbb{P}\left(\sqrt{n}\mathcal{D}_\mu\|F_{\mathrm{d}_{\mu,m}(\mu)}-F_{\mathrm{d}_{\mu,m}(\hat{\mu}_N)}\|_{\infty,(0,\mathcal{D}_\mu)}\geq\frac{\kappa}{4}\right)+$$
$$\mathbb{P}\left(\sqrt{n}\mathcal{D}_\nu\|F_{\mathrm{d}_{\nu,m}(\nu)}-F_{\mathrm{d}_{\nu,m}(\hat{\nu}_N)}\|_{\infty,(0,\mathcal{D}_\nu)}\geq\frac{\kappa}{4}\right)+$$
$$2\mathbb{P}\left(\frac{\sqrt{n}}{m}W_1(\mu,\hat{\mu}_N)\geq\frac{\kappa}{4}\right)+2\mathbb{P}\left(\frac{\sqrt{n}}{m}W_1(\nu,\hat{\nu}_N)\geq\frac{\kappa}{4}\right)+$$
$$\mathbb{P}\left(\|F_{\mathrm{d}_{\mu,m}(\hat{\mu}_n)}-F_{\mathrm{d}_{\mu,m}(\mu)}\|_{\infty,(0,\mathcal{D}_\mu)}\geq\frac{W_1(\mathrm{d}_{\mu,m}(\mu),\mathrm{d}_{\nu,m}(\nu))}{2\mathcal{D}_{\mu,m}}-\frac{\tilde{\mathrm{q}}_{1-\frac{\alpha}{2}}}{2\mathcal{D}_{\mu,m}\sqrt{n}}-\frac{(4+\alpha)\kappa+4}{4\mathcal{D}_{\mu,m}\alpha\sqrt{n}}\right)+$$
$$\mathbb{P}\left(\|F_{\mathrm{d}_{\nu,m}(\hat{\nu}_n)}-F_{\mathrm{d}_{\nu,m}(\nu)}\|_{\infty,(0,\mathcal{D}_\nu)}\geq\frac{W_1(\mathrm{d}_{\mu,m}(\mu),\mathrm{d}_{\nu,m}(\nu))}{2\mathcal{D}_{\nu,m}}-\frac{\tilde{\mathrm{q}}_{1-\frac{\alpha}{2}}}{2\mathcal{D}_{\nu,m}\sqrt{n}}-\frac{(4+\alpha)\kappa+4}{4\mathcal{D}_{\nu,m}\alpha\sqrt{n}}\right).$$

For all positive ϵ, for n big enough, note that the sum of the last two terms can be bounded thanks to the DKW-Massart inequality [34], by

$$4\exp\left(-\frac{W_1^2\left(\mathrm{d}_{\mu,m}(\mu),\mathrm{d}_{\nu,m}(\nu)\right)}{(2+\epsilon)\max\left\{\mathcal{D}_\mu^2,\mathcal{D}_\nu^2\right\}}n\right).$$

Note also that thanks to the DKW-Massart inequality, the first term can be bounded above by

$$2\exp\left(-\frac{1}{8\mathcal{D}_\mu^2}cn^{\rho-1+2\gamma}\right).$$

The second term is similar. Thanks to Theorem 2 in [26], the third term is bounded above by

$$c_1\exp\left(-c_2m^dn^{\rho+d\gamma-\frac{d}{2}}\right),$$

for some fixed constants c_1 and c_2. The remaining terms are similar.

Since $\rho>1$, we can choose a positive γ satisfying: $\gamma<\frac{1}{2}$, $\rho+d\gamma-\frac{d}{2}>1$ and $\rho-1+2\gamma>1$. So the two last expressions are negligible in comparison to the first one.

So, for n big enough, $\mathbb{P}_{(\mu,\nu)}\left(\sqrt{n}W_1\left(\mathrm{d}_{\hat{\mu}_N,m}(\hat{\mu}_n),\mathrm{d}_{\hat{\nu}_N,m}(\hat{\nu}_n)\right)<\hat{\mathrm{q}}_{1-\alpha}\right)$ is bounded above by

$$4\exp\left(-\frac{W_1^2\left(\mathrm{d}_{\mu,m}(\mu),\mathrm{d}_{\nu,m}(\nu)\right)}{3\max\left\{\mathcal{D}_{\mu,m}^2,\mathcal{D}_{\nu,m}^2\right\}}n\right).$$

■

References

[1] de Acosta, A. and Giné, E. (1979). Convergence Of Moments And Related Functionals In The Central Limit Theorem In Banach Spaces. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **48** 213–231.

[2] Araujo, A. and Giné, E. (1980). *The Central Limit Theorem for Real and Banach Valued Random Variables* John Wiley & Sons Inc.

[3] Arlot, S. (2007). Rééchantillonnage et Sélection de Modèles PhD thesis. Université Paris-Sud – Paris XI.

[4] del Barrio, E., Giné, E. and Matrán, C. (1999). Central Limit Theorems For The Wasserstein Distance Between The Empirical And The True Distributions. *The Annals of Probability* **27** 1009–1071.
[5] del Barrio, E., Lescornel, H. and Loubes, J-M. (2015). A statistical analysis of a deformation model with Wasserstein barycenters : estimation procedure and goodness of fit test. *unpublished.*
[6] Bickel, P. and Doksum, K. (1977). *Mathematical statistics : basic ideas and selected topics*. Englewood Cliffs, N.J. Prentice Hall.
[7] Billingsley, P. (1999). *Convergence of Probability Measures* Wiley-Interscience.
[8] Bobkov, S. and Ledoux, M. (2014). One-Dimensional Empirical Measures, Order Statistics And Kantorovich Transport Distances. *the Memoirs of the AMS - American Mathematical Society. to be published*
[9] Buchet, M. (2014). Topological Inference From Measures. PhD thesis. Université Paris-Sud – Paris XI.
[10] Buet, B. and Leonardi, G. (2015). Recovering Measures From Approximate Values On Balls. *unpublished.*
[11] Cazals, F. and Lhéritier, A. (2015). Beyond Two-sample-tests: Localizing Data Discrepancies in High-dimensional Spaces. *IEEE/ACM DSAA*
[12] Chazal, F., Cohen-Steiner, D., Guibas L. J., Mémoli, F. and Oudot S. (2009). Gromov-Hausdorff Stable Signatures for Shapes using Persistence. *Computer Graphics Forum (proc. SGP 2009)* 1393–1403.
[13] Chazal, F., Cohen-Steiner, D. and Mérigot, Q. (2011). Geometric Inference for Probability Measures. *Foundations of Computational Mathematics* **11** 733–751.
[14] Chazal, F., Fasy, B., Lecci, F., Michel, B., Rinaldo, A. and Wasserman, L.(2018). Robust Topological Inference: Distance To a Measure and Kernel Distance. *Journal of Machine Learning Research* **18** 1–40.
[15] Chazal, F., Fasy, B. T., Lecci, F., Michel, B., Rinaldo, A. and Wasserman, L. (2015). Subsampling methods for persistent homology. *Proceedings of the 32nd International Conference on Machine Learning, PMLR* **37** 2143–2151.
[16] Chazal, F., Fasy, B., Lecci, F., Rinaldo, A., Singh, A. and Wasserman, L. (2013). On the Bootstrap for Persistence Diagrams and Landscapes *Modeling and Analysis of Information Systems* **20** 96–105.
[17] Chazal, F., Glisse, M., Labruère, C. and Michel, B. (2015). Convergence rates for persistence diagram estimation in topological data analysis *Journal of Machine Learning Research* **16** 3603—3635
[18] Chazal, F., Massart, P. and Michel, B. (2016). Rates Of Convergence For Robust Geometric Inference. *Electronic Journal of Statistics* **10** 2243–2286.
[19] Chazal, F., De Silva, V. and Oudot, S. (2014). Persistence stability for geometric complexes. *Geometriae Dedicata* **173** 193–214.
[20] Cuevas, A. (2009). Set estimation: another bridge between statistics and geometry. *Boletín de Estadística e Investigación Operativa* **25** 71–85
[21] Cuevas, A. and Rodríguez-Casal, A. (2004). On boundary estimation. *Advances in Applied Probability* 340—354
[22] Efron, B.(1979). Bootstrap Methods: Another Look at the Jackknife *Annals of Statistics* **7** 1–26.
[23] Fasy, B., Kim, J., Lecci, F. and Maria, C. (2014). Introduction to the R package TDA. *unpublished.*
[24] Fasy, B., Lecci, F., Rinaldo, A., Wasserman, L., Balakrishnan, S. and Singh, A. (2014).

Confidence Sets For Persistence Diagrams *The Annals of Statistics*. **42** 2301-–2339.
[25] Federer, H. (1959). Curvature Measures. *Transactions of the American Mathematical Society* **93** 418–491.
[26] Fournier, N. and Guillin, A. (2015). On The Rate Of Convergence In Wasserstein Distance Of The Empirical Measure. *Probability Theory & Related Fields* **162** 707–738.
[27] Fromont, M. and Laurent, B. (2006). Adaptive goodness-of-fit tests in a density model *Annals of Statistics* **34** 680–720.
[28] Fromont, M., Laurent, B., Lerasle, M. and Reynaud-Bouret, P.(2012). Kernels Based Tests with Non-asymptotic Bootstrap Approaches for Two-sample Problems *Journal of Machine Learning Research: Workshop and Conference proceedings* COLT 2012 **23** 23.1–23.22.
[29] Gretton, A., Borgwardt, K., Rasch, M., Schölkopf, B. and Smola, A. (2012). A Kernel Two-Sample Test. *Journal of Machine Learning Research* **13** 723–773.
[30] Gromov, M. (2003). *Metric Structures for* R*iemannian and Non-*R*iemannian Spaces*. Birkhäuser Basel.
[31] Johnson, B. McK. and Killeen, T. (1983). An Explicit Formula for the C.D.F. of the L_1 Norm of the Brownian Bridge *The Annals of Probability* **11** 807–808.
[32] Rice, S.O. (1982). The Integral of the Absolute Value of the Pinned Wiener Process–Calculation of Its Probability Density by Numerical Integration *The Annals of Probability* **10** 240–243.
[33] Lieutier, A. (2004). Any Open Bounded Subset of R^n Has the Same Homotopy Type Than Its Medial Axis. *Computer Aided Geometric Design* **36** 1029–1046.
[34] Massart, P. (1990). The Tight Constant in the Dvoretzky-Kiefer-Wolfowitz Inequality. *The Annals of Probability* **18** 1269–1283.
[35] Mémoli, F. (2011). Gromov–Wasserstein Distances and the Metric Approach to Object Matching. *Foundations of Computational Mathematics* **11** 417–487.
[36] von Luxburg, U. and Alamgir, M. (2013). Density estimation from unweighted k-nearest neighbor graphs: a roadmap. *Neural Information Processing Systems (NIPS)*
[37] Niyogi, P., Smale, S. and Weinberger, S. (2008). Finding the Homology of Submanifolds with High Confidence from Random Samples. *Discrete and Computational Geometry* **39** 419–441.
[38] Osada, R., Funkhouser, T., Chazelle, B. and Dobkin, D. (2002). Shape Distributions. *ACM Transactions on Graphics* **21** 807–832.
[39] Politis, D. and Romano, J. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics* **22** 2031–2050.
[40] Ramdas, A., Trillos, N. and Cuturi, M. (2015). On Wasserstein Two Sample Testing and Related Families of Nonparametric Tests. *unpublished.*
[41] van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes* Springer Series in Statistics
[42] Villani, C. (2003). *Topics in Optimal Transportation*. American Mathematical Society.