JMLR: Workshop and Conference Proceedings vol 23 (2012) 23.1–23.22 25th Annual Conference on Learning Theory

Kernels Based Tests with Non-asymptotic Bootstrap Approaches for Two-sample Problems

Magalie Fromont MFROMONT@ENSAI.FR
CREST Ensai / IRMAR - Campus de Ker-Lann - BP 37203 - 35172 Bruz Cedex (FRANCE)

Béatrice Laurent BEATRICE.LAURENT@INSA-TOULOUSE.FR
IMT, INSA Toulouse - 135 av. de Rangueil - 31077 Toulouse Cedex 4 (FRANCE)

Matthieu Lerasle MLERASLE@UNICE.FR
CNRS Université de Nice Sophia-Antipolis - 06108 Nice Cedex 2 (FRANCE)

Patricia Reynaud-Bouret PATRICIA.REYNAUD-BOURET@UNICE.FR
CNRS Université de Nice Sophia-Antipolis - 06108 Nice Cedex 2 (FRANCE)

Editor: Shie Mannor, Nathan Srebro, Robert C. Williamson

Abstract

Considering either two independent i.i.d. samples, or two independent samples generated from a heteroscedastic regression model, or two independent Poisson processes, we address the question of testing equality of their respective distributions. We first propose single testing procedures based on a general symmetric kernel. The corresponding critical values are chosen from a wild or permutation bootstrap approach, and the obtained tests are exactly (and not just asymptotically) of level α. We then introduce an aggregation method, which enables to overcome the difficulty of choosing a kernel and/or the parameters of the kernel. We derive non-asymptotic properties for the aggregated tests, proving that they may be optimal in a classical statistical sense.

Keywords: Two-sample problem, kernel methods, density model, regression model, Poisson process, wild bootstrap, permutation test, adaptive tests, aggregation methods.

1. Introduction

We study in this paper some classical problems of testing the null hypothesis that two independent sets of random variables are equally distributed, problems which are usually referred to as two-sample problems. Three different frameworks are considered: either the sets of random variables are i.i.d. samples from a density model, or samples from a heteroscedastic regression model, or non-homogeneous Poisson processes. Many papers deal with the i.i.d. two-sample problem, from the historical tests of Kolmogorov-Smirnov and Cramer von Mises and their extensions, to the more recent tests of Li (1999), Gretton et al. (2008) and Gretton et al. (2010), which are the closest ones to the present study. As for the two-sample problem of testing the equality of signals in non-parametric regression, we can cite among many others the papers by Hall and Hart (1990), King et al. (1991), or the more recent one by Franke and Halim (2007). When non-homogeneous Poisson processes are considered, Bovett and Saw (1980) and Deshpande et al. (1999) respectively propose conditional and unconditional tests for the two-sample problem, but for a restrictive alternative hypothesis.

In these frameworks, non-parametric tests usually consist in rejecting the null hypothesis when an estimator of a distance between the distributions, chosen as testing statistic, is larger than a certain critical value. The question of the choice of a critical value ensuring that the test is, exactly or at least asymptotically, of the desired level α is a frequent major question. Indeed, the exact or asymptotic distributions of many testing statistics are not free from the common unknown distribution under the null hypothesis. In such cases, general bootstrap methods are often used to build data driven critical values. Except when the permutation bootstrap method is used, authors generally prove that the obtained tests are (only) asymptotically of level α. We here construct testing procedures which satisfy specific non-asymptotic statistical performances properties, thus justifying their use for moderate or even small sample sizes.

The testing statistics that we first introduce can be viewed as extensions of the ones proposed by Li (1999), or Gretton et al. (2008) and Gretton et al. (2010) in the density model, when the sample sizes are equal. Li (1999)'s statistics are based on approximation kernels from the classical non-parametric kernel estimation approach, whereas Gretton et al. (2008) and Gretton et al. (2010)'s ones consist in unbiased estimators of the *Maximum Mean Discrepancy*[1], based on so-called *characteristic* kernels (see Fukumizu et al. (2009)). Here, we consider very general kernels, including projection kernels that are related to model selection and thresholding estimation approaches, approximation kernels and Mercer kernels. This will enable us to recover general well-known statistical properties. Li (1999) and Gretton et al. (2008), among other rough bounds, choose the corresponding critical values from an asymptotic Efron's bootstrap method, that may have a high computational cost. On the contrary, Gretton et al. (2010) choose them from asymptotic estimates for the null distribution of the testing statistic, which have a smaller computational cost, and good performances in practice when the sample sizes are large. In all cases, the resulting tests are proved to be asymptotically of level α. Though these tests are investigated via a Monte Carlo simulation study for moderate or small sample sizes by Li (1999), they are not theoretically justified for finite sample sizes, neither by Li (1999), nor by Gretton et al. (2008) and Gretton et al. (2010).

The first main contribution of our work at this stage consists in proposing a new choice for the critical values, based on wild or permutation bootstrap approaches, thus following the original ideas of Mammen (1992) and Bickel (1968), but applied to U-statistics as in Arcones and Giné (1992). This choice may lead to high computational costs, but it ensures that our tests are exactly (and not only asymptotically) of level α.

The second main contribution consists in deriving, for Poisson processes and in the particular cases of projection and approximation kernels, optimal non-asymptotic conditions on the alternatives, which guarantee that the probability of second kind error is at most equal to a prescribed level β.

The testing procedures that we introduce hereafter are also intended to overcome the question of calibrating the choice of the kernel and/or the parameters of the kernel, which is crucial when using such kernel methods, and which remains unresolved by Gretton et al. (2008). They are based on an aggregation approach, that is now well-known in adaptive testing (see Baraud et al. (2003) for instance), but not used in statistical learning theory yet to our knowledge. Instead of considering a particular single kernel, we consider a whole collection of kernels, and the corresponding collection of tests, each with an adapted level of significance. We then reject the null hypothesis when it is rejected by one of the tests in the collection. The aggregated tests are constructed to be of level α, and in the particular Poisson framework, the loss in second kind error due to the aggregation, when unavoidable, is as small as possible. The last results are expressed in the form of non-asymptotic

1. The MMD is defined as the distance between the two embedded probability distributions in a RKHS

oracle type inequalities, that may also lead to optimal results from the adaptive minimax point of view. This is the third, and probably most important, main contribution of this work.

The paper is organized as follows. We describe in Section 2 the three considered two-sample problems, ending with the Poisson process framework, which will be the most convenient to obtain interesting and strong results. We then introduce in Section 3 our testing procedures based on single kernel functions. Section 4 is devoted to the study of the probabilities of first kind error for the three frameworks, and the study of the probability of second kind error for Poisson processes. We present in Section 5 our aggregation approach and the oracle type inequalities satisfied by the aggregated tests in the Poisson process model.

2. Two-sample problems

Let $Z^{(1)}$ and $Z^{(2)}$ be two independent sets of random variables observed in a measurable space $\mathcal{Z}$, whose - possibly random - cardinalities are respectively denoted by N_1 and N_2, and whose distributions respectively depend on unknown functions s_1 and s_2. Let $Z^{(1)} = \{Z_1^{(1)}, \dots, Z_{N_1}^{(1)}\}$ and $Z^{(2)} = \{Z_1^{(2)}, \dots, Z_{N_2}^{(2)}\}$. From the observation of $\{Z_1^{(1)}, \dots, Z_{N_1}^{(1)}\}$ and $\{Z_1^{(2)}, \dots, Z_{N_2}^{(2)}\}$, we consider the general two-sample problem which consists in testing (H_0) "$s_1 = s_2$" against (H_1) "$s_1 \neq s_2$".

Let us introduce a few notations. As usual, $\mathbb{P}_{s_1,s_2}$ denotes the joint distribution of $(Z^{(1)}, Z^{(2)})$, and $\mathbb{E}_{s_1,s_2}$ the corresponding expectation. We set for any event $\mathcal{A}$ based on $(Z^{(1)}, Z^{(2)})$, $\mathbb{P}_{(H_0)}(\mathcal{A}) = \sup_{(s_1,s_2), s_1=s_2} \mathbb{P}_{s_1,s_2}(\mathcal{A})$. Furthermore, we will need to consider the pooled set $Z = Z^{(1)} \cup Z^{(2)}$, with cardinality $N = N_1 + N_2$, whose elements are denoted by $\{Z_1, \dots, Z_N\}$.

A density model. Here $N_1 = n_1$ and $N_2 = n_2$ are fixed integers. We assume that $(Z_1^{(1)}, \dots, Z_{n_1}^{(1)})$ and $(Z_1^{(2)}, \dots, Z_{n_2}^{(2)})$ are two independent samples of i.i.d. random variables, observed in a measurable space $\mathcal{Z}$, with respective densities s_1 and s_2 with respect to some non-atomic σ-finite measure ν on $\mathcal{Z}$. Let us also assume that $s_1, s_2 \in \mathbb{L}^2(\mathcal{Z}, d\nu)$. The two-sample problem corresponding to this density model is the most classical one in statistics, but it is also well-known by the learning community. Indeed, many papers of learning theory now deal with it, such as Gretton et al. (2008) or Gretton et al. (2010) for instance.

A heteroscedastic regression model. Here $N_1 = n_1$ and $N_2 = n_2$ are also fixed integers. We assume that $(Z_1^{(1)}, \dots, Z_{n_1}^{(1)})$ and $(Z_1^{(2)}, \dots, Z_{n_2}^{(2)})$ are two independent samples of i.i.d. random variables such that for every $i \in \{1, \dots, n_1\}$, $Z_i^{(1)} = (X_i^{(1)}, Y_i^{(1)})$, with $Y_i^{(1)} = s_1(X_i^{(1)}) + \sigma(X_i^{(1)})\xi_i^{(1)}$, and for every $i \in \{1, \dots, n_2\}$, $Z_i^{(2)} = (X_i^{(2)}, Y_i^{(2)})$, with $Y_i^{(2)} = s_2(X_i^{(2)}) + \sigma(X_i^{(2)})\xi_i^{(2)}$. Here $X_i^{(1)}$ and $X_i^{(2)}$ are observed in a measurable space $\mathcal{X}$, and $Y_i^{(1)}$ and $Y_i^{(2)}$ take their values in a measurable subset $\mathcal{Y}$ of $\mathbb{R}$. We set $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$. The couples $(X_i^{(1)}, \xi_i^{(1)})$ and $(X_i^{(2)}, \xi_i^{(2)})$ are assumed to be identically distributed, and $E[\xi_i^{(1)} | X_i^{(1)}] = 0$, $E[(\xi_i^{(1)})^2 | X_i^{(1)}] = 1$. Here, s_1, s_2, and σ are assumed to be in $\mathbb{L}^2(\mathcal{X}, P_X)$ where P_X denotes the known common distribution of the $X_i^{(1)}$'s and $X_i^{(2)}$'s. Note that since the variance function σ^2 is the same for both signals, the corresponding two-sample problem amounts to the problem of testing the equality of densities for the samples $(Z_1^{(1)}, \dots, Z_{n_1}^{(1)})$ and $(Z_1^{(2)}, \dots, Z_{n_2}^{(2)})$. This problem is also quite classical in statistics, at least in signal detection issues.

A Poisson process model. Let $Z^{(1)} = \{Z_1^{(1)}, \ldots, Z_{N_1}^{(1)}\}$ and $Z^{(2)} = \{Z_1^{(2)}, \ldots, Z_{N_2}^{(2)}\}$ be the points of two independent Poisson processes observed in $\mathcal{Z}$, with respective intensities s_1 and s_2 with respect to a non-atomic σ-finite measure μ on $\mathcal{Z}$. Notice that here N_1 and N_2 are Poisson distributed random variables.

This framework may seem really unusual at least in learning theory. Nevertheless, it is particularly adapted to some specific applications, in reliability or traffic studies for instance. It can also be merely viewed as a density model, but with a Poissonization of the numbers N_1 and N_2, that is N_1 and N_2 are assumed to be Poisson distributed random variables. This Poissonization trick allows us, among others, to introduce a rather simple testing procedure satisfying sharp properties, that can not be exactly transposed in the other models. This is the reason why we start with this framework in the next sections.

In order to emphasize the link with the density model, we assume that the measure μ on $\mathcal{Z}$ satisfies $d\mu = nd\nu$, where ν is a fixed non-atomic σ-finite measure which may typically be the Lebesgue measure when $\mathcal{Z}$ is a measurable subset of $\mathbb{R}^d$. This amounts to considering the Poisson processes $Z^{(1)}$ and $Z^{(2)}$ as n pooled i.i.d. Poisson processes with respective intensities s_1 and s_2 w.r.t. ν. We assume furthermore that s_1 and s_2 belong to $\mathbb{L}^1(\mathcal{Z}, d\nu) \cap \mathbb{L}^\infty(\mathcal{Z}) \subset \mathbb{L}^2(\mathcal{Z}, d\nu)$.

3. Single tests based on single kernels

For any measurable function h w.r.t. the measure ν on $\mathcal{Z}$ in the density and Poisson process models, or the measure P_X on $\mathcal{X}$ in the regression model, we define for $k = 1, 2$, $\|h\|_k = (\int_{\mathcal{Z}} |h|^k d\nu)^{1/k}$, or $\|h\|_k = (\int_{\mathcal{X}} |h|^k dP_X)^{1/k}$ when they exist, and we denote by $\langle ., . \rangle$ the scalar product associated with $\|.\|_2$. Moreover, for any real valued function h, $\|h\|_\infty = \sup_{z \in \mathcal{Z}} |h(z)|$.

3.1. Single tests for Poisson processes

Let us take a symmetric kernel function $K : \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}$, which will be chosen as in one of the three following examples, and which satisfies

$$\int_{\mathcal{Z}^2} K^2(z, z')(s_1 + s_2)(z)(s_1 + s_2)(z') d\nu_z d\nu_{z'} < +\infty. \tag{1}$$

We introduce the testing statistic T_K which is defined by

$$T_K = \sum_{i,j \in \{1,\ldots,N\}, i \neq j} K(Z_i, Z_j) \varepsilon_i^0 \varepsilon_j^0,$$

where the ε_i^0's are some marks on Z. More precisely, for every i in $\{1, \ldots, N\}$, $\varepsilon_i^0 = 1$ if Z_i belongs to $Z^{(1)}$, $\varepsilon_i^0 = -1$ if Z_i belongs to $Z^{(2)}$.
We have chosen to study and discuss three particular examples of kernel functions. For each example, we explain why T_K may be a relevant statistic for the problem of testing (H_0) "$s_1 = s_2$" against (H_1) "$s_1 \neq s_2$".

Example 1. Our first choice for K is a kernel function based on a finite orthonormal family $\{\varphi_\lambda, \lambda \in \Lambda\}$ for $\langle ., . \rangle$:

$$K(z, z') = \sum_{\lambda \in \Lambda} \varphi_\lambda(z) \varphi_\lambda(z').$$

This kernel is known as a projection kernel. Indeed, for every function f in $\mathbb{L}^2(\mathcal{Z}, d\nu)$,

$$\int_{\mathcal{Z}} K(z, z')f(z')d\nu_{z'} = \Pi_S(f)(z),$$

where S is the subspace of $\mathbb{L}^2(\mathcal{Z}, d\nu)$ generated by $\{\varphi_\lambda, \lambda \in \Lambda\}$, and Π_S is the orthogonal projection onto S for $\langle ., . \rangle$. By straightforward calculations, one can see that T_K is actually an unbiased estimator of $n^2\|\Pi_S(s_1 - s_2)\|^2$. Hence, when $\{\varphi_\lambda, \lambda \in \Lambda\}$ is well-chosen, T_K can also be viewed as a relevant estimator of $n^2\|s_1 - s_2\|^2$.
Let us give some typical examples of such kernels obtained from the Fourier basis and from the Haar basis in the case where $\mathcal{Z} = [0, 1]$ and ν is the Lebesgue measure. We first introduce the Fourier basis $(\varphi_j)_{j\geq 0}$ of $\mathbb{L}^2([0, 1], d\nu)$ defined by

$$\varphi_0(x) = \mathbf{1}_{[0,1]}(x),\ \forall j \geq 1,\ \varphi_{2j}(x) = \sqrt{2}\cos(2\pi jx),\ \varphi_{2j-1}(x) = \sqrt{2}\sin(2\pi jx).$$

Setting $\Lambda = \{0, 1, \dots, 2D\}$, the corresponding kernel is the Dirichlet kernel defined by:

$$K(z, z') = 1 + 2\sum_{j=1}^{D} \cos(2\pi j(z - z')).$$

Let now $\{\varphi_0,\ \varphi_{(j,k)},\ j \in \mathbb{N},\ k \in \{0, \dots, 2^j - 1\}\}$ be the Haar basis of $\mathbb{L}^2([0, 1], d\nu)$ with

$$\varphi_0(x) = \mathbf{1}_{[0,1]}(x) \quad \text{and} \quad \varphi_{(j,k)}(x) = 2^{j/2}\psi(2^j x - k),$$

where $\psi(x) = \mathbf{1}_{[0,1/2[}(x) - \mathbf{1}_{[1/2,1[}(x)$. If $\Lambda = \{0\} \cup \big\{(j, k), 0 \leq j \leq J, k \in \{0, \dots, 2^j - 1\}\big\}$, the corresponding kernel K corresponds to a projection kernel onto the space generated by all the functions of the Haar basis up to the level J.

Example 2. When $\mathcal{Z} = \mathbb{R}^d$ and ν is the Lebesgue measure, our second choice for K is a kernel function based on an approximation kernel k in $\mathbb{L}^2(\mathbb{R}^d)$, and such that $k(-z) = k(z)$: for $z = (z_1, \dots, z_d)$, $z' = (z'_1, \dots, z'_d)$ in $\mathcal{Z}$,

$$K(z, z') = \frac{1}{\prod_{i=1}^d h_i} k\left(\frac{z_1 - z'_1}{h_1}, \dots, \frac{z_d - z'_d}{h_d}\right),$$

where $h = (h_1, \dots, h_d)$ is a vector of d positive bandwidths. In this case, $\mathbb{E}_{s_1,s_2}[T_K] = n^2\langle k_h * (s_1 - s_2), s_1 - s_2\rangle$, where $k_h(u_1, \dots, u_d) = \frac{1}{\prod_{i=1}^d h_i} k\left(\frac{u_1}{h_1}, \dots, \frac{u_d}{h_d}\right)$ and $*$ is the usual convolution operator w.r.t. ν.

Example 3. Our third choice corresponds to a general Mercer or learning kernel (see Schölkopf and Smola (2002)) such that

$$K(z, z') = \langle \theta(z), \theta(z')\rangle_{\mathcal{H}_K},$$

where θ and $\mathcal{H}_K$ are a representation function and a RKHS associated with K. Here, $\langle ., . \rangle_{\mathcal{H}_K}$ denotes the scalar product of $\mathcal{H}_K$. Recall that K also has to satisfy (1). This choice of general Mercer kernels leads to a testing statistic close to the one of Gretton et al. (2008) for the classical i.i.d. two-sample problem when the sizes of the i.i.d. samples are equal. In this case, $\mathbb{E}_{s_1,s_2}[T_K] = n^2 \left\| \int_{\mathcal{Z}} \theta(z)(s_1 - s_2)(z)d\nu_z \right\|^2_{\mathcal{H}_K}$, where $\|.\|_{\mathcal{H}_K}$ is the norm associated with $\langle ., . \rangle_{\mathcal{H}_K}$. From Lemma

4 in Gretton et al. (2008), we know that this quantity corresponds to n^2 times the squared Maximum Mean Discrepancy on the unit ball in the RKHS $\mathcal{H}_K$. Moreover, when K is a universal kernel, such as the Gaussian and the Laplacian kernels, $\mathbb{E}_{s_1,s_2}[T_K] = 0$ if and only if $s_1 = s_2$. If $\int_{\mathcal{Z}} s_1 d\nu = \int_{\mathcal{Z}} s_2 d\nu = 1$, this is sufficient to say that the kernel is characteristic in the sense of Fukumizu et al. (2009).

Notice that the projection kernels of *Example 1* and the kernels based on some particular approximation kernels such as in *Example 2* are also Mercer kernels. These kernels (the Dirichlet or Gaussian ones for instance) are thus often used in learning theory, though they are not always normalized in the same way. This difference in normalization may pose some problems in the statistical interpretation of some results (see Comment 4 of Theorem 4), and we will take care of it. However, even when the Mercer kernels lead to some results that remain difficult to interpret from a statistical point of view, their introduction is helpful when the space $\mathcal{Z}$ is unusual or pretty large with respect to the (mean) number of observations and/or when the measure ν is not well specified or easy to deal with. In such situations, the use of Mercer kernels may be the only possible way to compute a meaningful test (see Gretton et al. (2008) where such kernels are used for microarrays data and graphs).

If we denote by $\diamond$ the following operator:

$$K \diamond p(z) = \langle K(.,z), p \rangle, \tag{2}$$

by the Cauchy-Schwarz inequality, (1) ensures that $\langle K \diamond (s_1 - s_2), s_1 - s_2 \rangle$ is well-defined. Then, in the three above examples, T_K is an unbiased estimator of $n^2 \langle K \diamond (s_1 - s_2), s_1 - s_2 \rangle$ (see Appendix B for more details). It is therefore appropriate to take it as testing statistic, and reject (H_0) when T_K is larger than a critical value to be defined.

Since the distribution of T_K under (H_0) is unknown, we turn to a wild bootstrap approach, which comes from the original idea of Mammen (1992) applied to U-statistics as in Arcones and Giné (1992), but which is here proved to be exactly justified.

We introduce a sequence $(\varepsilon_i)_{i \in \mathbb{N}}$ of i.i.d. Rademacher variables independent of Z. Following Mammen (1992) and Arcones and Giné (1992), the wild bootstrapped version of T_K would be given by $\sum_{i,j \in \{1,\ldots,N\}, i \neq j} K(Z_i, Z_j) \varepsilon_i^0 \varepsilon_j^0 \varepsilon_i \varepsilon_j$. It is easy to see that under (H_0), this wild bootstrapped version of T_K has the same distribution as

$$T_K^\varepsilon = \sum_{i,j \in \{1,\ldots,N\}, i \neq j} K(Z_i, Z_j) \varepsilon_i \varepsilon_j.$$

Furthermore, from a corollary of a more general result of Daley and Vere-Jones (2008), whose statement and proof are given in Appendix C for sake of understanding, we prove that under (H_0), since s_1 and s_2 are assumed to belong to $\mathbb{L}^1(\mathcal{Z}, d\nu)$, T_K and T_K^ε exactly have the same distribution conditioned on Z (see also Proposition 2). Hence, we consider the $(1-\alpha)$ quantile of T_K^ε conditioned on Z denoted by $q_{K,1-\alpha}^{(Z)}$. We finally introduce the test that rejects (H_0) when $T_K > q_{K,1-\alpha}^{(Z)}$, whose test function is defined by

$$\Phi_{K,\alpha} = \mathbf{1}_{T_K > q_{K,1-\alpha}^{(Z)}}. \tag{3}$$

3.2. Single tests in the density and heteroscedastic regression models

To shorten the following mathematical expressions, let us define

$$\begin{aligned} a_{n_1,n_2} &= \left(\frac{1}{n_1(n_1-1)} - c_{n_1,n_2}\right)^{1/2} \\ b_{n_1,n_2} &= -a_{n_2,n_1} = -\left(\frac{1}{n_2(n_2-1)} - c_{n_1,n_2}\right)^{1/2}, \end{aligned}$$

where

$$c_{n_1,n_2} = \frac{1}{n_1 n_2 (n_1+n_2-2)}.$$

We consider a symmetric kernel function K chosen as in *Example 1*, *Example 2*, or *Example 3*, replacing $\mathcal{Z}$ by $\mathcal{X}$ and ν by P_X in the regression model, and we introduce the testing statistic defined by

$$\dot{T}_K = \sum_{i,j\in\{1,\ldots,N\}, i\neq j} K(Z_i, Z_j)\left(\varepsilon_i^0 \varepsilon_j^0 + c_{n_1,n_2}\right),$$

in the density model, or

$$\ddot{T}_K = \sum_{i,j\in\{1,\ldots,N\}, i\neq j} Y_i Y_j K(X_i, X_j)\left(\varepsilon_i^0 \varepsilon_j^0 + c_{n_1,n_2}\right),$$

in the regression model. The marks ε_i^0's are here defined by $\varepsilon_i^0 = a_{n_1,n_2}$ if $Z_i \in Z^{(1)}$ and $\varepsilon_i^0 = b_{n_1,n_2}$ if $Z_i \in Z^{(2)}$. We can prove by straightforward calculations (see Appendix B for a proof) the following result.

Proposition 1 *The statistics $\dot{T}_K$ and $\ddot{T}_K$ are unbiased estimators of $\langle K \diamond (s_1 - s_2), s_1 - s_2\rangle$.*

Let us now explain how we choose the corresponding critical values. Let $R = (R_1, \ldots, R_N)$ be a random vector uniformly distributed on the set of all permutations of $\{1, \ldots, N\}$ and independent of Z, and let $\varepsilon_i = a_{n_1,n_2}$ if $i \in \{R_1, \ldots, R_{n_1}\}$, and $\varepsilon_i = b_{n_1,n_2}$ if $i \in \{R_{n_1+1}, \ldots, R_N\}$.

We then define in the density model:

$$\dot{T}_K^{\varepsilon} = \sum_{i,j\in\{1,\ldots,N\}, i\neq j} K(Z_i, Z_j)\left(\varepsilon_i \varepsilon_j + c_{n_1,n_2}\right),$$

and denote by $\dot{q}_{K,1-\alpha}^{(Z)}$ the $(1-\alpha)$ quantile of $\dot{T}_K^{\varepsilon}$ conditioned on Z.

In the same way, we define in the regression model:

$$\ddot{T}_K^{\varepsilon} = \sum_{i,j\in\{1,\ldots,N\}, i\neq j} Y_i Y_j K(X_i, X_j)\left(\varepsilon_i \varepsilon_j + c_{n_1,n_2}\right),$$

and denote by $\ddot{q}_{K,1-\alpha}^{(Z)}$ the $(1-\alpha)$ quantile of $\ddot{T}_K^{\varepsilon}$ conditioned on Z.

We finally consider the tests that reject (H_0) when $\dot{T}_K > \dot{q}_{K,1-\alpha}^{(Z)}$ and $\ddot{T}_K > \ddot{q}_{K,1-\alpha}^{(Z)}$, whose test functions are respectively denoted by $\dot{\Phi}_{K,\alpha}$ and $\ddot{\Phi}_{K,\alpha}$.

Notice that rejecting (H_0) when $\dot{T}_K > \dot{q}_{K,1-\alpha}^{(Z)}$ is equivalent to rejecting (H_0) when $\sum_{i,j\in\{1,\ldots,N\}, i\neq j} K(Z_i, Z_j)\varepsilon_i^0\varepsilon_j^0$ is larger than the $(1-\alpha)$ quantile of the conditional distribution of $\sum_{i,j\in\{1,\ldots,N\}, i\neq j} K(Z_i, Z_j)\varepsilon_i\varepsilon_j$ given Z, which simplifies the implementation of $\dot{\Phi}_{K,\alpha}$. The implementation of $\ddot{\Phi}_{K,\alpha}$ can of course be simplified in the same way.

4. Probabilities of first and second kind errors of the single tests

Let $\mathbf{T}_K$, $\mathbf{T}_K^\varepsilon$, $\mathbf{q}_{K,1-\alpha}^{(Z)}$ and $\mathbf{\Phi}_{K,\alpha}$ be either T_K, T_K^ε, $q_{K,1-\alpha}^{(Z)}$ and $\Phi_{K,\alpha}$ respectively in the Poisson process model, or $\dot{T}_K$, $\dot{T}_K^\varepsilon$, $\dot{q}_{K,1-\alpha}^{(Z)}$ and $\dot{\Phi}_{K,\alpha}$ respectively in the density model, or $\ddot{T}_K$, $\ddot{T}_K^\varepsilon$, $\ddot{q}_{K,1-\alpha}^{(Z)}$ and $\ddot{\Phi}_{K,\alpha}$ respectively in the regression model.

4.1. Probabilities of first kind error

We here state a result which is at the origin of our choice of the various bootstrap methods involved in the construction of the critical values $\mathbf{q}_{K,1-\alpha}^{(Z)}$.

Proposition 2 *Under (H_0), $\mathbf{T}_K$ and $\mathbf{T}_K^\varepsilon$ have the same distribution conditioned on Z.*

As a consequence, given α in $(0,1)$, under (H_0),

$$\mathbb{P}_{s_1,s_2}\left(\mathbf{T}_K > \mathbf{q}_{K,1-\alpha}^{(Z)}\middle|Z\right) \leq \alpha. \tag{4}$$

By taking the expectation over Z, we thus obtain that

$$\mathbb{P}_{(H_0)}\left(\mathbf{\Phi}_{K,\alpha} = 1\right) \leq \alpha, \tag{5}$$

that is $\mathbf{\Phi}_{K,\alpha}$ is exactly of level α.

Notice that the property (4) is in fact stronger than the usual control of the probability of first kind error (without any conditioning) of the test, such as in (5).

4.2. Probability of second kind error for Poisson processes

In this section, we exclusively consider the Poisson process model.

Given β in $(0,1)$, we here bring out a sufficient condition on the alternative (s_1,s_2) which guarantees that

$$\mathbb{P}_{s_1,s_2}(\Phi_{K,\alpha} = 0) \leq \beta.$$

Proposition 3 *Let $\alpha,\beta \in (0,1)$. For any symmetric kernel K satisfying (1), let $A_K = n^2\int_{\mathcal{Z}}\left(K\diamond(s_1-s_2)\right)^2(s_1+s_2)d\nu$, and $B_K = n^2\int_{\mathcal{Z}^2}K^2(z,z')(s_1+s_2)(z)(s_1+s_2)(z')d\nu_z d\nu_{z'}$. There exists some absolute constant $\kappa > 0$ such that if*

$$\mathbb{E}_{s_1,s_2}[T_K] > 2\sqrt{\frac{2A_K+B_K}{\beta}} + \kappa\ln(2/\alpha)\sqrt{\frac{2B_K}{\beta}},$$

then $\mathbb{P}_{s_1,s_2}(\Phi_{K,\alpha} = 0) \leq \beta$.

We are now in a position to deduce from Proposition 3 recognizable properties in terms of uniform separation rates. Considering each of our three possible choices for the kernel K, and evaluating A_K and B_K in these cases, we actually obtain the following theorem.

Theorem 4 *Let $\alpha, \beta \in (0,1)$, and $\kappa > 0$ be the constant of Proposition 3. Let $\Phi_{K,\alpha}$ be the test function defined by (3), where K may be chosen as in Section 3.1.*

1. When K is constructed as in Example 1 from an orthonormal basis $\{\varphi_\lambda, \lambda \in \Lambda\}$ of a D-dimensional linear subspace S of $\mathbb{L}^2(\mathcal{Z}, d\nu)$, we introduce the following condition:

$$\|s_1-s_2\|_2^2 \geq \|(s_1-s_2)-\Pi_S(s_1-s_2)\|_2^2+\frac{(4+2\sqrt{2}\kappa\ln(2/\alpha))\|s_1+s_2\|_\infty\sqrt{D}}{n\sqrt{\beta}}+\frac{8\|s_1+s_2\|_\infty}{\beta n}. \quad (6)$$

2. When $\mathcal{Z} = \mathbb{R}^d$, ν is the Lebesgue measure on $\mathcal{Z}$, and K is constructed as in Example 2 from an approximation kernel k in $\mathbb{L}^2(\mathbb{R}^d)$, such that $k(x) = k(-x)$, and $h = (h_1, \ldots, h_d)$ with $h_i > 0$, we introduce the following condition:

$$\|s_1 - s_2\|_2^2 \geq \|(s_1 - s_2) - k_h * (s_1 - s_2)\|_2^2 + \frac{4+2\sqrt{2}\kappa\ln(2/\alpha)}{n\sqrt{\beta}}\sqrt{\frac{\|s_1+s_2\|_\infty\|s_1+s_2\|_1\|k\|_2^2}{\prod_{i=1}^d h_i}} + \frac{8\|s_1+s_2\|_\infty}{\beta n}. \quad (7)$$

3. When K is a Mercer kernel associated with a representation function θ and a RKHS $\mathcal{H}_K$ such that (1) holds as in Example 3, we introduce the following condition:

$$\|s_1-s_2\|_2^2 \geq \inf_{r>0}\left[\left\|(s_1 - s_2) - r^{-1}K \diamond (s_1 - s_2)\right\|_2^2 + \frac{4+2\sqrt{2}\kappa\ln(2/\alpha)}{nr\sqrt{\beta}}\sqrt{C_K}\right] + \frac{8\|s_1+s_2\|_\infty}{\beta n}, \quad (8)$$

where $C_K = B_K/n^2 = \int_{\mathcal{Z}^2} K^2(z,z')(s_1+s_2)(z)(s_1+s_2)(z')d\nu_z d\nu_{z'}$.

If K is chosen as in one of these three cases, and if the condition (6), (7) or (8) on (s_1, s_2) is satisfied respectively, then $\mathbb{P}_{s_1,s_2}(\Phi_{K,\alpha} = 0) \leq \beta$.

Comments.

1. The proof of this result relies on two fundamental points. The first one is the fact that T_K is an unbiased estimator of $n^2\langle K \diamond (s_1 - s_2), s_1 - s_2\rangle$. The second one is the fact that the bootstrapped testing statistic T_K^ε, from which the critical value of the test is obtained, is a Rademacher chaos, which can be precisely controlled. When we consider the density and regression models, though the first point can be directly transposed (see Proposition 1), it is not the case for the second point, at this stage of our research.

2. In the first case, we see that the right hand side of (6) reproduces a bias-variance decomposition close to the bias-variance decomposition for projection estimators, with a variance term of order $\sqrt{D}/n$ instead of D/n. This is quite usual in statistical testing theory (see Baraud (2002) for instance), and we know that this leads to sharp upper bounds for the uniform separation rates of the test. Let us explain this point in the case where the kernel K is the projection kernel onto the space generated by the functions of the Haar basis up to the level J. We assume that $(s_1 - s_2)$ belongs to some Besov body of index $\delta > 0$ and radius $R > 0$. Let us recall that this implies that $\|(s_1 - s_2) - \Pi_S(s_1 - s_2)\|_2^2 \leq R^2 D^{-2\delta}$, which gives an upper bound for the right hand side of (6). If we choose the value of D (depending on R and δ) minimizing this upper bound, we obtain that $\mathbb{P}_{s_1,s_2}(\Phi_{K,\alpha} = 0) \leq \beta$ if $\|s_1 - s_2\|_2 \geq Cn^{-2\delta/(1+4\delta)}$, where C is a constant depending on α, β, R, δ and $\|s_1 + s_2\|_\infty$. It was proved in Fromont et al. (2011) that the uniform separation rate

over a Besov body of index δ and radius R for testing homogeneity of a Poisson process is precisely bounded from below by $n^{-2\delta/(1+4\delta)}$. By simple arguments, this lower bound can be extended to the present two-sample problem. In this sense, the results obtained in Theorem 4 are sharp.

3. The second case also reproduces a bias-variance decomposition when $k \in \mathbb{L}^1(\mathbb{R}^d)$ and $\int_{\mathbb{R}^d} k(z)d\nu_z = 1$: the bias is here $\|(s_1 - s_2) - k_h * (s_1 - s_2)\|_2$. When $h_1 = \ldots = h_d$, the variance term is of order $h_1^{-d/2}/n$. As usual in the approximation kernel estimation theory, this coincide with what is found in the first case through the equivalence $h_1^{-d} \sim D$ (see Tsybakov (2009) for instance for more details).

4. The third case is however unusual, since the term $\|(s_1 - s_2) - r^{-1}K \diamond (s_1 - s_2)\|_2$ can not always be viewed as a bias term. Indeed, when Mercer kernels K are used in learning theory, their approximation capacity is not always considered. Even when the Mercer kernel is based on a projection kernel or an approximation kernel (such as the classical Gaussian kernel for instance), its usual normalization does not necessarily lead to sharp results from the point of view of approximation or statistical theory. This is the reason why we wanted to keep the possibility of replacing the usual normalization by a possibly more adequate one, through the factor r^{-1}.

5. Multiple or aggregated tests based on collections of kernels

In the previous section, we have considered testing procedures based on a single kernel function K. Using such single tests however leads to the natural question of the choice of the kernel, and/or its parameters: the orthonormal family in *Example 1*, the approximation kernel and the bandwidth h in *Example 2*, the Mercer kernel and/or its parameters in *Example 3*. Authors often choose particular parameters regarding the performance properties that they target for their tests, or use a data driven method to choose these parameters which is not always justified. For instance, Gretton et al. (2008) and Gretton et al. (2010) choose the parameter of the kernel from a heuristic method.

In order to overcome these issues, we propose in this section to consider some collections of kernel functions instead of single ones, and to define multiple testing procedures by aggregating the corresponding single tests, with an adapted choice of the critical values.

5.1. The aggregation of single tests

Let us introduce a finite collection $\{K_m, m \in \mathcal{M}\}$ of symmetric kernels: $\mathcal{Z} \times \mathcal{Z} \to \mathbb{R}$ that satisfy (1) in the Poisson process and density models, or symmetric kernels: $\mathcal{X} \times \mathcal{X} \to \mathbb{R}$ such that $\int_{\mathcal{X}^2} K_m^2(x,x')(s_1+s_2)(x)(s_1+s_2)(x')dP_X(x)dP_X(x') < +\infty$ in the regression model. Let $\{w_m, m \in \mathcal{M}\}$ be a collection of positive numbers such that $\sum_{m\in\mathcal{M}} e^{-w_m} \leq 1$. For all m in $\mathcal{M}$, let $\mathbf{T}_{K_m}$ and $\mathbf{T}^{\varepsilon}_{K_m}$ be defined as in Section 4, just taking $K = K_m$. For $u \in (0,1)$, we denote by $\mathbf{q}^{(Z)}_{m,1-u}$ the $(1-u)$ quantile of $\mathbf{T}^{\varepsilon}_{K_m}$ conditioned on the pooled process Z, and we introduce for $\alpha \in (0,1)$:

$$\mathbf{u}^{(Z)}_{\alpha} = \sup\left\{ u > 0, \mathbb{P}\left(\sup_{m\in\mathcal{M}} \left(\mathbf{T}^{\varepsilon}_{K_m} - \mathbf{q}^{(Z)}_{m,1-ue^{-w_m}} \right) > 0 \,\middle|\, Z \right) \leq \alpha \right\}.$$

We now consider the test which rejects (H_0) when there exists at least one m in $\mathcal{M}$ such that $\mathbf{T}_{K_m} > \mathbf{q}^{(Z)}_{m,1-\mathbf{u}^{(Z)}_{\alpha}e^{-w_m}}$, whose test function is given by

$$\mathbf{\Phi}_{\alpha} = \mathbf{1}_{\sup_{m\in\mathcal{M}}\left(\mathbf{T}_{K_m} - \mathbf{q}^{(Z)}_{m,1-\mathbf{u}^{(Z)}_{\alpha}e^{-w_m}}\right)>0}. \tag{9}$$

Note that given the observation of the pooled process Z, $\mathbf{u}_\alpha^{(Z)}$ and $\mathbf{q}^{(Z)}_{m,1-u_\alpha^{(Z)}e^{-w_m}}$ can be estimated by a classical Monte Carlo procedure.

It is quite straightforward to see that this test is of level α and that one can guarantee a probability of second kind error at most equal to $\beta \in (0,1)$ if one can guarantee it for one of the single tests rejecting (H_0) when $\mathbf{T}_{K_m} > \mathbf{q}^{(Z)}_{m,1-\mathbf{u}_\alpha^{(Z)}e^{-w_m}}$.

We thus obtain from Theorem 4 an interesting result for Poisson processes.

5.2. Oracle type inequalities for Poisson processes

In this section, as in Section 4.2, we exclusively consider the Poisson process model.

Theorem 5 *Let $\alpha, \beta \in (0,1)$. Let $\{K_m, m \in \mathcal{M}\}$ be a collection of kernels, chosen as in one of the two following cases, and $\{w_m, m \in \mathcal{M}\}$ be a collection of positive numbers such that $\sum_{m\in\mathcal{M}} e^{-w_m} \le 1$.*

Case 1. Let $\{S_m, m \in \mathcal{M}\}$ be a finite collection of D_m-dimensional linear subspaces of $\mathbb{L}^2(\mathcal{Z}, d\nu)$, spanned by orthonormal bases denoted by $\{\varphi_\lambda, \lambda \in \Lambda_m\}$ respectively. We set, for all m in $\mathcal{M}$, $K_m(z,z') = \sum_{\lambda\in\Lambda_m} \varphi_\lambda(z)\varphi_\lambda(z')$, and we introduce the condition:

$$\|s_1-s_2\|_2^2 \ge \inf_{m\in\mathcal{M}}\left\{ \|(s_1-s_2)-\Pi_{S_m}(s_1-s_2)\|_2^2 + \frac{4+2\sqrt{2}\kappa\left(\ln(2/\alpha)+w_m\right)}{n\sqrt{\beta}}\|s_1+s_2\|_\infty\sqrt{D_m}\right\} + \frac{8\|s_1+s_2\|_\infty}{\beta n}. \quad (10)$$

Case 2. If $\mathcal{Z} = \mathbb{R}^d$ and ν is the Lebesgue measure on $\mathbb{R}^d$, let $\{k_{m_1}, m_1 \in \mathcal{M}_1\}$ be a collection of approximation kernels such that $\int_{\mathcal{Z}} k_{m_1}^2(z)d\nu_z < \infty$, $k_{m_1}(z) = k_{m_1}(-z)$, and a collection $\{h_{m_2}, m_2 \in \mathcal{M}_2\}$ of vectors $h_{m_2} = (h_{m_2,1}, \ldots, h_{m_2,d})$ of d positive numbers. We set $\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2$, and for all $m = (m_1, m_2)$ in $\mathcal{M}$, $z = (z_1, \ldots, z_d)$, $z' = (z_1', \ldots, z_d')$ in $\mathbb{R}^d$,

$$K_m(z,z') = k_{m_1,h_{m_2}}(z-z') = \frac{1}{\prod_{i=1}^d h_{m_2,i}} k_{m_1}\left(\frac{z_1-z_1'}{h_{m_2,1}}, \ldots, \frac{z_d-z_d'}{h_{m_2,d}}\right).$$

We introduce the following condition:

$$\|s_1-s_2\|_2^2 \ge \inf_{(m_1,m_2)\in\mathcal{M}} \left\{ \|(s_1-s_2) - k_{m_1,h_{m_2}} * (s_1-s_2)\|_2^2 + \frac{4+2\sqrt{2}\kappa(\ln(2/\alpha)+w_m)}{n\sqrt{\beta}}\sqrt{\frac{\|s_1+s_2\|_\infty\|s_1+s_2\|_1\|k_{m_1}\|_2^2}{\prod_{i=1}^d h_{m_2,i}}}\right\} + \frac{8\|s_1+s_2\|_\infty}{\beta n}, \quad (11)$$

Let $\mathbf{\Phi}_\alpha$ be the test defined by (9). $\mathbf{\Phi}_\alpha$ is a level α test, and if either (10) in Case 1 or (11) in Case 2 is satisfied, then $\mathbb{P}_{s_1,s_2}(\mathbf{\Phi}_\alpha = 0) \le \beta$.

Comparing these results with the ones obtained in Theorem 4, one can see that considering the aggregated tests allows to obtain the infimum over all m in $\mathcal{M}$ in the right hand side of (10) and (11) at the price of the only additional term w_m. This result can be viewed as an oracle type inequality: indeed, without knowing $(s_1 - s_2)$, we know that the uniform separation rate of the aggregated test is of the same order as the smallest uniform separation rate in the collection of single tests, up to the factor w_m. By choosing a collection of kernels based on nested or more complicated linear subspaces generated by subsets of the Haar basis of $\mathbb{L}^2(\mathcal{Z}, d\nu)$, when $\mathcal{Z} = [0,1]$ and ν is the Lebesgue measure on $[0,1]$, as in the paper by Fromont et al. (2011), this can be used to prove that our test is adaptive in the minimax sense over classes of alternatives (s_1, s_2) such that $(s_1 - s_2)$ belongs to Besov or weak Besov bodies with various parameters (see Fromont et al. (2011) for more details).

Acknowledgments

The authors acknowledge the support of the French Agence Nationale de la Recherche (ANR), under grant ATLAS (JCJC06-137446) ”From Applications to Theory in Learning and Adaptive Statistics” and under grant Calibration (ANR-2011-BS01-010-02).

References

M. A. Arcones and E. Giné. On the bootstrap of U and V statistics. *Ann. Statist.*, 20(2):655–674, 1992.

Y. Baraud. Non-asymptotic minimax rates of testing in signal detection. *Bernoulli*, 8(5):577–606, 2002.

Y. Baraud, S. Huet, and B. Laurent. Adaptive tests of linear hypotheses by model selection. *Ann. Statist.*, 31(1):225–251, 2003.

P. J. Bickel. A distribution free version of the Smirnov two sample test in the p-variate case. *Ann. Math. Statist.*, 40:1–23, 1968.

J. M. Bovett and J. G. Saw. On comparing two Poisson intensity functions. *Comm. Statist. A—Theory Methods*, 9(9):943–948, 1980.

D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes. Vol. II*. Probability and its Applications (New York). Springer, New York, second edition, 2008. General theory and structure.

V. H. de la Peña and E. Giné. *Decoupling*. Probability and its Applications (New York). Springer-Verlag, New York, 1999. From dependence to independence, Randomly stopped processes. U-statistics and processes. Martingales and beyond.

J. V. Deshpande, M. Mukhopadhyay, and U. V. Naik-Nimbalkar. Testing of two sample proportional intensity assumption for non-homogeneous Poisson processes. *J. Statist. Plann. Inference*, 81(2): 237–251, 1999.

J. Franke and S. Halim. Wild bootstrap tests. *Signal proc. mag., IEEE*, 24(4):31–37, 2007.

M. Fromont, B. Laurent, and P. Reynaud-Bouret. Adaptive tests of homogeneity for a Poisson process. *Ann. Inst. Henri Poincaré Probab. Stat.*, 47(1):176–213, 2011.

K. Fukumizu, B. Sriperumbudur, A. Gretton, and B. Schölkopf. Characteristic kernels on groups and semigroups. *Advances in Neural Information Processing Systems 21, (NIPS 2008)*, pages 473–480, 2009.

A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola. A kernel method for the two-sample problem. *J. Mach. Learn. Res.*, 1:1–10, 2008.

A. Gretton, K. Fukumizu, Z. Harchaoui, and B. K. Sriperumbudur. A fast, consistent kernel two-sample test. In *Advances in Neural Information Processing Systems 22, (NIPS 2009)*, pages 673–681, 2010.

P. Hall and J. D. Hart. Bootstrap test for difference between means in nonparametric regression. *J. Amer. Statist. Assoc.*, 85(412):1039–1049, 1990.

E. King, J. D. Hart, and T. E. Wehrly. Testing the equality of two regression curves using linear smoothers. *Statist. Probab. Lett.*, 12(3):239–247, 1991.

Q. Li. Nonparametric testing the similarity of two unknown density functions: local power and bootstrap analysis. *J. Nonparametr. Statist.*, 11(1-3):189–213, 1999.

E. Mammen. Bootstrap, wild bootstrap, and asymptotic normality. *Probab. Theory Related Fields*, 93(4):439–455, 1992.

B. Schölkopf and A. J. Smola. *Learning with kernels*. Cambridge (Mass.) : MIT Press, 2002.

A. B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York, 2009. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.

Appendix A. Short simulation study

In this Appendix, we want to bring a part of an experimental study, which aims at illustrating the theoretical results of this paper. We consider the only Poisson process model here. Let $\mathcal{Z} = [0,1]$ or $\mathcal{Z} = \mathbb{R}$, $n = 100$ and ν be the Lebesgue measure on $\mathcal{Z}$. $Z^{(1)}$ and $Z^{(2)}$ denote two independent Poisson processes with intensities s_1 and s_2 on $\mathcal{Z}$ with respect to $\mu = 100\,\nu$. We focus on the multiple testing procedure $\Phi_{K,\alpha}$ defined by (9), with K chosen as: a projection kernel based on nested subsets of the Haar basis on $[0,1]$, the standard Gaussian kernel, or the Epanechnikov approximation kernel, and with $\alpha = 0.05$. The corresponding tests are denoted by Ne, G, and E respectively. To be more explicit, recall that we introduced the Haar basis $\{\varphi_0,\ \varphi_{(j,k)},\ j \in \mathbb{N},\ k \in \{0,\ldots,2^j-1\}\}$ in Section 3.1. Let $K_0(z,z') = \varphi_0(z)\varphi_0(z')$, and for $J \geq 1$, $K_J(z,z') = \sum_{\lambda \in \{0\}\cup\Lambda_J} \varphi_\lambda(z)\varphi_\lambda(z')$ with $\Lambda_J = \{(j,k),\ j \in \{0,\ldots,J-1\},\ k \in \{0,\ldots,2^j-1\}\}$. For $J \geq 1$, we take $w_J = 2\left(\ln(J+1) + \ln(\pi/\sqrt{6})\right)$. The test Ne then corresponds to the multiple testing procedure $\mathbf{\Phi}_\alpha$ defined by (9), with the collection of kernels $\{K_J, J = 0,\ldots,7\}$ and with the collection of weights $\{w_J, J = 0,\ldots,7\}$. Let us now describe precisely the tests G and E. Let k be defined by either $k(u) = (2\pi)^{-1/2}\exp(-u^2/2)$ for all $u \in \mathbb{R}$ in the Gaussian case, or $k(u) = (3/4)(1-u^2)\mathbf{1}_{|u|\leq 1}$ in the Epanechnikov case. Let $\{h_m, m \in \mathcal{M}\} = \{1/24, 1/16, 1/12, 1/8, 1/4, 1/2\}$ be a collection of bandwidths and $\{K_m, m \in \mathcal{M}\}$ be the corresponding collection of kernels given by

$K_m(z,z') = \frac{1}{h_m} k\left(\frac{z-z'}{h_m}\right)$ for all m in $\mathcal{M}$. The tests G and E then correspond to the multiple testing procedure $\mathbf{\Phi}_\alpha$ defined by (9), with the collection of kernels $\{K_m, m \in \mathcal{M}\}$ and $\{w_m, m \in \mathcal{M}\} = \{1/6, \ldots, 1/6\}$.

As we noticed in Section 2, the present Poisson process model merely corresponds to a Poissonization of the density model. Hence, conditionally on the number of points of $Z^{(1)}$ and $Z^{(2)}$, any test for the classical i.i.d. two-sample problem can be used here. We compare our tests with the Kolmogorov-Smirnov test and the test proposed by Gretton et al. (2008), based on a Gaussian kernel with a heuristic choice for the parameter of the kernel, and a critical value obtained from an Efron's bootstrap method. These two tests are respectively denoted by KS and M (for MMD abbreviation). The probabilities of first kind error of the five tests Ne, G, E, KS and M are estimated from 5000 simulations, and the probabilities of second kind error from 1000 simulations.

We focus on several intensities s_1 and s_2, taken among:

$$
\begin{aligned}
&f_1(x) = \mathbf{1}_{[0,1]}(x),\\
&f_{2,a,\varepsilon}(x) = (1+\varepsilon)\mathbf{1}_{[0,a)}(x) + (1-\varepsilon)\mathbf{1}_{[a,2a)}(x) + \mathbf{1}_{[2a,1)}(x),\\
&f_{3,\eta}(x) = \left(1 + \eta \sum_j \frac{h_j}{2}(1+\operatorname{sgn}(x-p_j))\right) \frac{\mathbf{1}_{[0,1]}(x)}{C_2(\eta)},\\
&f_{4,\varepsilon}(x) = (1-\varepsilon)\mathbf{1}_{[0,1]}(x) + \varepsilon\left(\sum_j g_j \left(1 + \frac{|x-p_j|}{w_j}\right)^{-4}\right) \frac{\mathbf{1}_{[0,1]}(x)}{0.284},\\
&f_{5,\lambda}(x) = \frac{\lambda}{2} e^{-\lambda|x-1/2|},\\
&f_{6,\mu,\sigma} = \frac{1}{\sqrt{2\pi}\sigma} e^{-|x-\mu|^2/\sigma^2},
\end{aligned}
$$

where p, h, g, w, ε are defined as in Fromont et al. (2011) [2], $0 < \varepsilon \leq 1$, $0 < a < 1/2$, $\eta > 0$ and $C_2(\eta)$ is such that $\int_0^1 g_{2,\eta}(x)dx = 1$.

The obtained estimated levels of the tests fluctuate between: 0.042 and 0.053 for KS, 0.048 and 0.052 for M, 0.047 and 0.049 for Ne, 0.051 and 0.054 for G, 0.05 and 0.55 for E.

The obtained estimated powers of the tests are represented in the two following figures. The dots represent the estimated powers, and the triangles represent the upper and lower bounds of asymptotic confidence intervals with confidence level 99%, with variance estimation.

The main point that we can notice here is that when the alternative intensities are very irregular, our tests perform better, even sometimes much better, than the two other ones, and that it is particularly true for the test E. The only case where the test M clearly outperforms ours involves more regular intensities. Of course, we do not always know whether the underlying intensities of

2.

p=	(	0.1	0.13	0.15	0.23	0.25	0.4	0.44	0.65	0.76	0.78	0.81	)
h=	(	4	-4	3	-3	5	-5	2	4	-4	2	-3	)
g=	(	4	5	3	4	5	4.2	2.1	4.3	3.1	5.1	4.2	)
w=	(	0.005	0.005	0.006	0.01	0.01	0.03	0.01	0.01	0.005	0.008	0.005	)

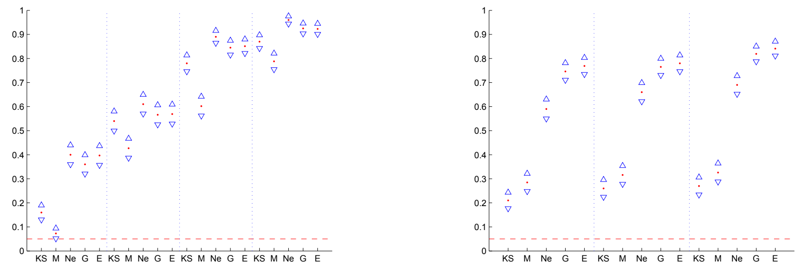

Figure 1: Left: $(s_1, s_2) = (f_1, f_{2,a,\varepsilon})$. Each column resp. corresponds to $(a, \varepsilon) = (1/8, 1)$, $(1/4, 0.7)$, $(1/4, 0.9)$, and $(1/4, 1)$. Right: $(s_1, s_2) = (f_1, f_{3,\eta})$. Each column resp. corresponds to $\eta = 4$, 8 and 15.

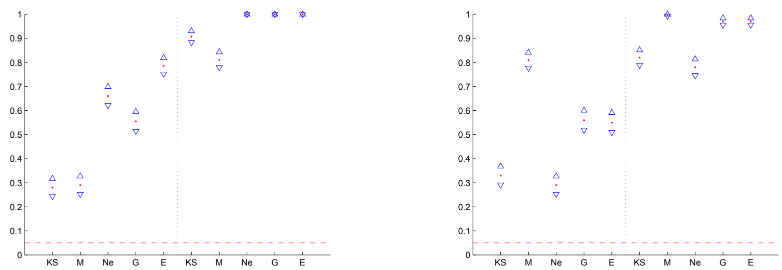

Figure 2: Left: $(s_1, s_2) = (f_1, f_{4,\varepsilon})$. The two columns resp. correspond to $\varepsilon = 0.5$ and 1. Right: $(s_1, s_2) = (f_{5,\lambda}, f_{6,1/2,1/4})$. The two columns resp. correspond to $\lambda = 7$ and $\lambda = 10$.

our problem are irregular or not. A good compromise would be in such cases to aggregate several of the studied tests, for instance Ne, M and E.

Appendix B. The testing statistics as unbiased estimators

The Poisson process model. Let $(\varepsilon_z^0)_{z \in Z}$ be defined by $\varepsilon_z^0 = 1$ if z belong to $Z^{(1)}$ and $\varepsilon_z^0 = -1$ if z belongs to $Z^{(2)}$, and let dZ be the point measure associated with the pooled process Z. Denoting by $\mathcal{Z}^{[2]}$ the set $\{(z, z') \in \mathcal{Z}^2, z \neq z'\}$, we have that

$$T_K = \sum_{z, z' \in Z, z \neq z'} K(z, z') \varepsilon_z^0 \varepsilon_{z'}^0 = \int_{\mathcal{Z}^{[2]}} K(z, z') \varepsilon_z^0 \varepsilon_{z'}^0 dZ_z dZ_{z'}.$$

Since, for every z in Z, $\mathbb{E}[\varepsilon_z^0|Z] = \frac{(s_1-s_2)(z)}{(s_1+s_2)(z)}$ (see (12) for a proof),

$$\begin{aligned}
\mathbb{E}_{s_1,s_2}[T_K] &= \mathbb{E}_{s_1,s_2}\left[\mathbb{E}\left[\int_{\mathcal{Z}^{[2]}} K(z,z')\varepsilon_z^0\varepsilon_{z'}^0 dZ_z dZ_{z'}\middle| Z\right]\right] \\
&= \mathbb{E}_{s_1,s_2}\left[\int_{\mathcal{Z}^{[2]}} K(z,z')\frac{(s_1-s_2)(z)}{(s_1+s_2)(z)}\frac{(s_1-s_2)(z')}{(s_1+s_2)(z')} dZ_z dZ_{z'}\right] \\
&= n^2\int_{\mathcal{Z}^2} K(z,z')(s_1-s_2)(z)(s_1-s_2)(z')d\nu_z d\nu_{z'}. \\
&= n^2\langle K\diamond(s_1-s_2), s_1-s_2\rangle.
\end{aligned}$$

The density model. Let us introduce

$$c_{n_1} = \frac{1}{n_1(n_1-1)} \text{ and } c_{n_2} = \frac{1}{n_2(n_2-1)}.$$

Now remark that $\dot{T}_K = \sum_{i,j\in\{1,\dots,N\}, i\neq j} K(Z_i,Z_j)\left(\varepsilon_i^0\varepsilon_j^0 + c_{n_1,n_2}\right)$ can be rewritten as

$$\begin{aligned}
\dot{T}_K &= \sum_{i,j\in\{1,\dots,n_1\}, i\neq j} K(Z_i^{(1)}, Z_j^{(1)})c_{n_1} + \sum_{i,j\in\{1,\dots,n_2\}, i\neq j} K(Z_i^{(2)}, Z_j^{(2)})c_{n_2} \\
&\quad -2\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} K(Z_i^{(1)}, Z_j^{(2)})\left(\sqrt{(c_{n_1}-c_{n_1,n_2})(c_{n_2}-c_{n_1,n_2})} - c_{n_1,n_2}\right).
\end{aligned}$$

Noticing that

$$(c_{n_1}-c_{n_1,n_2})(c_{n_2}-c_{n_1,n_2}) = \left((n_1n_2)^{-1} + c_{n_1,n_2}\right)^2,$$

we obtain by straightforward calculations that

$$\begin{aligned}
\mathbb{E}_{s_1,s_2}\left[\dot{T}_K\right] &= \int_{\mathcal{Z}^2} K(z,z')s_1(z)s_1(z')d\nu_z d\nu_{z'} + \int_{\mathcal{Z}^2} K(z,z')s_2(z)s_2(z')d\nu_z d\nu_{z'} \\
&\quad -2\int_{\mathcal{Z}^2} K(z,z')s_1(z)s_2(z')d\nu_z d\nu_{z'} \\
&= \int_{\mathcal{Z}^2} K(z,z')(s_1-s_2)(z)(s_1-s_2)(z')d\nu_z d\nu_{z'}. \\
&= \langle K\diamond(s_1-s_2), s_1-s_2\rangle.
\end{aligned}$$

$\dot{T}_K$ is thus an unbiased estimator of $\langle K\diamond(s_1-s_2), s_1-s_2\rangle$.

The regression model. We keep the same notations as above for c_{n_1} and c_{n_2}. Now remark that $\ddot{T}_K$ can be rewritten as

$$\begin{aligned}
\ddot{T}_K &= \sum_{i,j\in\{1,\dots,n_1\}, i\neq j} Y_i^{(1)}Y_j^{(1)}K(X_i^{(1)}, X_j^{(1)})c_{n_1} + \sum_{i,j\in\{1,\dots,n_2\}, i\neq j} Y_i^{(2)}Y_j^{(2)}K(X_i^{(2)}, X_j^{(2)})c_{n_2} \\
&\quad -2\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} Y_i^{(1)}Y_j^{(2)}K(X_i^{(1)}, X_j^{(2)})\left(\sqrt{(c_{n_1}-c_{n_1,n_2})(c_{n_2}-c_{n_1,n_2})} - c_{n_1,n_2}\right).
\end{aligned}$$

Hence, by similar arguments as above, we obtain that

$$\begin{aligned}
\mathbb{E}_{s_1,s_2}\left[\ddot{T}_K\right] &= \int_{\mathcal{X}^2} K(x,x')s_1(x)s_1(x')dP_X(x)dP_X(x') \\
&\quad + \int_{\mathcal{X}^2} K(x,x')s_2(x)s_2(x')dP_X(x)dP_X(x') \\
&\quad -2\int_{\mathcal{X}^2} K(x,x')s_1(x)s_2(x')dP_X(x)dP_X(x') \\
&= \int_{\mathcal{X}^2} K(x,x')(s_1-s_2)(x)(s_1-s_2)(x')dP_X(x)dP_X(x'). \\
&= \langle K(s_1-s_2), s_1-s_2\rangle.
\end{aligned}$$

$\ddot{T}_K$ is thus an unbiased estimator of $\langle K\diamond(s_1-s_2), s_1-s_2\rangle$.

Appendix C. Exact validity of the wild bootstrap approach in the Poisson process model

We here want to prove that under (H_0), the testing statistic T_K and its bootstrapped version T_K^ε exactly have the same distribution conditioned on Z. In this appendix, we prove a more precise result, which is known in the Poisson process field, but never explicitly stated as follows to our knowledge.

Proposition 6 *(i) Let $Z^{(1)}$ and $Z^{(2)}$ be two independent Poisson processes on a metric space $\mathcal{Z}$ with intensities s_1 and s_2 with respect to some measure μ on $\mathcal{Z}$ and such that $s_1, s_2 \in \mathbb{L}^1(\mathcal{Z}, d\mu)$. Denote by $dZ^{(1)}$ and $dZ^{(2)}$ the point measures respectively associated with $Z^{(1)}$ and $Z^{(2)}$. Then the pooled process Z whose point measure is given by $dZ = dZ^{(1)} + dZ^{(2)}$ is a Poisson process on $\mathcal{Z}$ with intensity s_1+s_2 with respect to μ. Let $\left(\varepsilon_z^0\right)_{z\in Z}$ be defined by $\varepsilon_z^0 = 1$ if z belongs to $Z^{(1)}$ and $\varepsilon_z^0 = -1$ if z belongs to $Z^{(2)}$. Then conditionally on Z, the variables $\left(\varepsilon_z^0\right)_{z\in Z}$ are i.i.d. and*

$$\forall\, z\in Z,\ \mathbb{P}\left(\varepsilon_z^0 = 1|Z\right) = \frac{s_1(z)}{(s_1+s_2)(z)},\ \mathbb{P}\left(\varepsilon_z^0 = -1|Z\right) = \frac{s_2(z)}{(s_1+s_2)(z)} \tag{12}$$

with the convention that $0/0 = 1/2$.

(ii) Respectively, let Z be a Poisson process on $\mathcal{Z}$ with intensity s_1+s_2 with respect to some measure μ. Let $(\varepsilon_z)_{z\in Z}$ be a family of random variables with values in $\{-1,1\}$ such that, conditionally on Z, the variables $(\varepsilon_z)_{z\in Z}$ are i.i.d. and

$$\forall\, z\in Z,\ \mathbb{P}\left(\varepsilon_z = 1|Z\right) = \frac{s_1(z)}{(s_1+s_2)(z)},\ \mathbb{P}\left(\varepsilon_z = -1|Z\right) = \frac{s_2(z)}{(s_1+s_2)(z)}$$

with the convention that $0/0 = 1/2$. Then the point processes $Z^{(1)}$ and $Z^{(2)}$, respectively defined by the point measures $dZ_z^{(1)} = \mathbf{1}_{\varepsilon_z=1}dZ_z$ and $dZ_z^{(2)} = \mathbf{1}_{\varepsilon_z=-1}dZ_z$ are two independent Poisson processes with respective intensities s_1 and s_2 with respect to μ.

All along the proof, $\int$ denotes $\int_{\mathcal{Z}}$. One of the key arguments of the proof is that the marked point processes are characterized by their Laplace functional (see for instance Daley and Vere-Jones (2008)).

To prove the first point of the result, this key argument makes sufficient to compute $\mathbb{E}\left[\exp\left(\int hdZ\right)\right]$ for a bounded measurable function h on $\mathcal{Z}$. Since $Z^{(1)}$ and $Z^{(2)}$ are independent,

$$\mathbb{E}\left[\exp\left(\int hdZ\right)\right] = \mathbb{E}\left[\exp\left(\int hdZ^{(1)}\right)\right]\mathbb{E}\left[\exp\left(\int hdZ^{(2)}\right)\right].$$

Since the Laplace functional of $Z^{(1)}$ is given by $\mathbb{E}\left[\exp\left(\int hdZ^{(1)}\right)\right] = \exp\left(\int\left(e^h-1\right)s_1 d\mu\right)$, and the Laplace functional of $Z^{(2)}$ has the same form, replacing s_1 by s_2,

$$\mathbb{E}\left[\exp\left(\int hdZ\right)\right] = \exp\left(\int\left(e^h-1\right)(s_1+s_2)\,d\mu\right),$$

which is the Laplace functional of a Poisson process with intensity (s_1+s_2) w.r.t. μ.
Let us now prove (12). The distribution of $(\varepsilon_z^0)_{z\in Z}$ conditioned on Z is characterized by the function:

$$t=(t_z)_{z\in Z}\mapsto \Phi(t,Z) = \mathbb{E}\left[\exp\left(\sum_{z\in Z} t_z\varepsilon_z^0\right)\middle| Z\right].$$

Let λ be a bounded measurable function defined on $\mathcal{Z}$, and define

$$\mathbb{E}_\lambda = \mathbb{E}\left[\exp\left(\int \lambda dZ\right)\exp\left(\sum_{z\in Z} t_z\varepsilon_z^0\right)\right].$$

By independency of $Z^{(1)}$ and $Z^{(2)}$ again,

$$\begin{aligned}\mathbb{E}_\lambda &= \mathbb{E}\left[\exp\left(\int(\lambda(z)+t_z)dZ_z^{(1)}\right)\exp\left(\int(\lambda(z)-t_z)dZ_z^{(2)}\right)\right]\\ &= \mathbb{E}\left[\exp\left(\int(\lambda(z)+t_z)dZ_z^{(1)}\right)\right]\mathbb{E}\left[\exp\left(\int(\lambda(z)-t_z)dZ_z^{(2)}\right)\right].\end{aligned}$$

Then

$$\begin{aligned}\mathbb{E}_\lambda &= \exp\left[\int(e^{(\lambda(z)+t_z)}-1)s_1(z)+(e^{(\lambda(z)-t_z)}-1)s_2(z)\right]d\mu_z\\ &= \exp\int(e^{h(z)}-1)(s_1+s_2)(z)d\mu_z\\ &= \mathbb{E}\left[\exp\left(\int hdZ\right)\right],\end{aligned}$$

where

$$h(z) = \lambda(z) + \ln\left(\frac{e^{t_z}s_1(z)+e^{-t_z}s_2(z)}{(s_1+s_2)(z)}\right).$$

Hence, for every bounded measurable function λ defined on $\mathcal{Z}$,

$$\mathbb{E}\left[\exp\left(\int\lambda dZ\right)\exp\left(\sum_{z\in Z}t_z\varepsilon_z^0\right)\right] = \mathbb{E}\left[\exp\left(\int\lambda dZ\right)\prod_{z\in Z}\left(e^{t_z}\frac{s_1(z)}{(s_1+s_2)(z)} + e^{-t_z}\frac{s_2(z)}{(s_1+s_2)(z)}\right)\right].$$

Recalling that the marked point processes are characterized by their Laplace functional, this implies that

$$\Phi(t,Z)=\mathbb{E}\left[\exp\left(\sum_{z\in Z}t_z\varepsilon_z^0\right)\middle|Z\right]=\prod_{z\in Z}\left(e^{t_z}\frac{s_1(z)}{(s_1+s_2)(z)}+e^{-t_z}\frac{s_2(z)}{(s_1+s_2)(z)}\right),$$

which concludes the proof of (12).

To prove the second point of the result, let h_1 and h_2 be two bounded measurable functions on $\mathcal{Z}$.

$$\begin{aligned}\mathbb{E}\left[\exp\left(\int h_1dZ^{(1)}+\int h_2dZ^{(2)}\right)\right] &= \mathbb{E}\left[\mathbb{E}\left[\exp\left(\int h_1dZ^{(1)}+\int h_2dZ^{(2)}\right)\middle|Z\right]\right]\\ &= \mathbb{E}\left[\mathbb{E}\left[\prod_{z\in Z}\exp\left(h_1\left(z\right)\mathbf{1}_{\varepsilon_z=1}+h_2\left(z\right)\mathbf{1}_{\varepsilon_z=-1}\right)\middle|Z\right]\right].\end{aligned}$$

Remark that there is almost surely a finite number of points in Z and that if z belongs to Z, then $s_1\left(z\right)+s_2\left(z\right)>0$. Moreover

$$\mathbb{E}\left[\exp\left(h_1\left(z\right)\mathbf{1}_{\varepsilon_z=1}+h_2\left(z\right)\mathbf{1}_{\varepsilon_z=-1}\right)\right] = e^{h_1(z)}\frac{s_1\left(z\right)}{s_1\left(z\right)+s_2\left(z\right)} + e^{h_2(z)}\frac{s_2\left(z\right)}{s_1\left(z\right)+s_2\left(z\right)}.$$

Then using the expression of the Laplace functional of Z with the function

$$h=\ln\left(e^{h_1(z)}\frac{s_1\left(z\right)}{s_1\left(z\right)+s_2\left(z\right)}+e^{h_2(z)}\frac{s_2\left(z\right)}{s_1\left(z\right)+s_2\left(z\right)}\right),$$

leads to

$$\begin{aligned}&\mathbb{E}\left[\exp\left(\int h_1dZ^{(1)}+\int h_2dZ^{(2)}\right)\right]\\ &\qquad=\exp\left(\int\left(e^{h_1(z)}\frac{s_1\left(z\right)}{s_1\left(z\right)+s_2\left(z\right)}+e^{h_2(z)}\frac{s_2\left(z\right)}{s_1\left(z\right)+s_2\left(z\right)}-1\right)(s_1+s_2)(z)d\mu_z\right).\end{aligned}$$

Finally we have that

$$\mathbb{E}\left[\exp\left(\int h_1dZ^{(1)}+\int h_2dZ^{(2)}\right)\right]=\exp\left(\int\left(e^{h_1}-1\right)s_1d\mu\right)\exp\left(\int\left(e^{h_2}-1\right)s_2d\mu\right).$$

We here recognize the product of the Laplace functionals of two Poisson processes with respective intensities s_1 and s_2 w.r.t. μ. This gives the independence and concludes the result.

Appendix D. Proof of Proposition 3

Given β in $(0,1)$, we here want to bring out an exact condition on the alternative (s_1,s_2) which ensures that

$$\mathbb{P}_{s_1,s_2}(\Phi_{K,\alpha}=0)\leq\beta. \tag{13}$$

Let us introduce the $(1-\beta/2)$ quantile of the conditional quantile $q^{(Z)}_{K,1-\alpha}$ that we denote by $q^{\alpha}_{1-\beta/2}$. Then for any (s_1, s_2),

$$\mathbb{P}_{s_1,s_2}(\Phi_{K,\alpha} = 0) \leq \mathbb{P}_{s_1,s_2}(T_K \leq q^{\alpha}_{1-\beta/2}) + \beta/2,$$

and a condition which guarantees $\mathbb{P}_{s_1,s_2}(T_K \leq q^{\alpha}_{1-\beta/2}) \leq \beta/2$ will be enough to ensure (13).

We denote by $\mathcal{Z}^{[3]}$ and $\mathcal{Z}^{[4]}$ the sets $\{(z_1, z_2, z_3) \in \mathcal{Z}^3,\ z_1, z_2, z_3 \text{ all different}\}$ and $\{(z_1, z_2, z_3, z_4) \in \mathcal{Z}^4,\ z_1, z_2, z_3, z_4 \text{ all different}\}$ respectively. Let us denote by dZ the point measure associated with the pooled process Z.

From Markov's inequality, we have that

$$\mathbb{P}_{s_1,s_2}\left(|-T_K + \mathbb{E}_{s_1,s_2}[T_K]| \geq t\right) \leq \frac{\mathrm{Var}(T_K)}{t^2}.$$

Since $\mathbb{E}_{s_1,s_2}[T_K^2] = \mathbb{E}_{s_1,s_2}\left[\mathbb{E}\left[\left(\int_{\mathcal{Z}^{[2]}} K(z,z')\varepsilon^0_z\varepsilon^0_{z'} dZ_z dZ_{z'}\right)^2 |Z\right]\right]$, by using (12), we obtain that

$$\begin{aligned}
\mathbb{E}_{s_1,s_2}[T_K^2] \quad = \quad & \mathbb{E}_{s_1,s_2}\Bigg[\int_{\mathcal{Z}^{[4]}} K(z_1,z_2)K(z_3,z_4)\frac{s_1-s_2}{s_1+s_2}(z_1)\frac{s_1-s_2}{s_1+s_2}(z_2) \\
& \frac{s_1-s_2}{s_1+s_2}(z_3)\frac{s_1-s_2}{s_1+s_2}(z_4) dZ_{z_1}dZ_{z_2}dZ_{z_3}dZ_{z_4}\Bigg] \\
& +4\mathbb{E}_{s_1,s_2}\left[\int_{\mathcal{Z}^{[3]}} K(z_1,z_2)K(z_1,z_3)\frac{s_1-s_2}{s_1+s_2}(z_2)\frac{s_1-s_2}{s_1+s_2}(z_3)dZ_{z_1}dZ_{z_2}dZ_{z_3}\right] \\
& +2\mathbb{E}_{s_1,s_2}\left[\int_{\mathcal{Z}^{[2]}} K^2(z_1,z_2)dZ_{z_1}dZ_{z_2}\right].
\end{aligned}$$

Now, from Lemma 5.4 III in Daley and Vere-Jones (2008) on factorial moments measures applied to Poisson processes, we deduce that

$$\begin{aligned}
\mathbb{E}_{s_1,s_2}[T_K^2] \quad = \quad & \int_{\mathcal{Z}^4}\Bigg(K(z_1,z_2)K(z_3,z_4)(s_1-s_2)(z_1)(s_1-s_2)(z_2) \\
& (s_1-s_2)(z_3)(s_1-s_2)(z_4)\Bigg)d\mu_{z_1}d\mu_{z_2}d\mu_{z_3}d\mu_{z_4} \\
& +4\int_{\mathcal{Z}^3} K(z_1,z_2)K(z_1,z_3)(s_1+s_2)(z_1)(s_1-s_2)(z_2)(s_1-s_2)(z_3)d\mu_{z_1}d\mu_{z_2}d\mu_{z_3} \\
& +2\int_{\mathcal{Z}^2} K^2(z_1,z_2)(s_1+s_2)(z_1)(s_1+s_2)(z_2)d\mu_{z_1}d\mu_{z_2},
\end{aligned}$$

where the integrals in the right hand side are finite from the assumption (1).
We finally obtain that

$$\mathbb{E}_{s_1,s_2}[T_K^2] = (\mathbb{E}_{s_1,s_2}[T_K])^2 + 4A_K + 2B_K,$$

so

$$\mathbb{P}_{s_1,s_2}\left(|-T_K + \mathbb{E}_{s_1,s_2}[T_K]| \geq \sqrt{\frac{8A_K + 4B_K}{\beta}}\right) \leq \frac{\beta}{2}.$$

Therefore if

$$\mathbb{E}_{s_1,s_2}[T_K] > \sqrt{\frac{8A_K + 4B_K}{\beta}} + q^{\alpha}_{1-\beta/2}, \tag{14}$$

then $\mathbb{P}_{s_1,s_2}(T_K \leq q^{\alpha}_{1-\beta/2}) \leq \beta/2$.

Let us now give a sharp upper bound for $q^{\alpha}_{1-\beta/2}$. Conditionally on Z, T_K^{ε} is a homogeneous Rademacher chaos of the form $\sum_{i\neq i'} z_{i,i'}\varepsilon_i\varepsilon_{i'}$, where the $z_{i,i'}$'s are some real deterministic numbers and $(\varepsilon_i)_{i\in\mathbb{N}}$ is a sequence of i.i.d. Rademacher variables. From Corollary 3.2.6 of de la Peña and Giné (1999), we deduce that there exists some absolute constant $\kappa > 0$ such that

$$\mathbb{E}\left(\exp\left[\frac{|\sum_{i\neq i'} z_{i,i'}\varepsilon_i\varepsilon_{i'}|}{\kappa\sum_{i\neq i'} z^2_{i,i'}}\right]\right) \leq 2,$$

hence by using Markov's inequality, we have that

$$\mathbb{P}\left(|\sum_{i\neq i'} z_{i,i'}\varepsilon_i\varepsilon_{i'}| \geq \kappa\ln(2/\alpha)\sum_{i\neq i'} z^2_{i,i'}\right) \leq \alpha.$$

Applying this result to T_K^{ε}, we obtain that $q^{\alpha}_{1-\beta/2}$ is upper bounded by the $(1-\beta/2)$ quantile of $\kappa\ln(2/\alpha)\sqrt{\int_{\mathcal{Z}^{[2]}} K^2(z,z')dZ_z dZ_{z'}}$. By using Markov's inequality again and Lemma 5.4 III in Daley and Vere-Jones (2008) again, we finally deduce that

$$\mathbb{P}_{s_1,s_2}\left(\int_{\mathcal{Z}^{[2]}} K^2(z,z')dZ_z dZ_{z'} \geq \frac{2B_K}{\beta}\right) \leq \frac{\beta}{2},$$

and that $q^{\alpha}_{1-\beta/2} \leq \kappa\ln(2/\alpha)\sqrt{\frac{2B_K}{\beta}}$, which allows to conclude with (14).

Appendix E. Proof of Theorem 4

First, notice that $A_K \leq n^3\|K\diamond(s_1-s_2)\|_2^2\|s_1+s_2\|_\infty$ and recall that $B_K = n^2C_K$. Then, for all $r>0$,

$$\mathbb{E}_{s_1,s_2}[T_K] = \frac{n^2r}{2}\left(\|s_1-s_2\|_2^2 + r^{-2}\|K\diamond(s_1-s_2)\|_2^2 - \|(s_1-s_2) - r^{-1}K\diamond(s_1-s_2)\|_2^2\right).$$

From Proposition 3, we deduce that $\mathbb{P}_{s_1,s_2}(\mathbf{\Phi}_{K,\alpha}=0) \leq \beta$ if

$$\begin{aligned}\|s_1-s_2\|_2^2 &+ r^{-2}\|K\diamond(s_1-s_2)\|_2^2 - \|(s_1-s_2) - r^{-1}K\diamond(s_1-s_2)\|_2^2 \\ &\geq 4\sqrt{\frac{2\|s_1+s_2\|_\infty}{n\beta}}\frac{\|K\diamond(s_1-s_2)\|_2}{r} + \frac{2}{nr\sqrt{\beta}}\left(2+\kappa\sqrt{2}\ln\left(\frac{2}{\alpha}\right)\right)\sqrt{C_K}.\end{aligned}$$

By using the elementary inequality $2ab \leq a^2+b^2$ with $a = \|K\diamond(s_1-s_2)\|_2/r$ and $b = 2\sqrt{2}\sqrt{\|s_1+s_2\|_\infty/(n\beta)}$ in the right hand side of the above condition, this condition can be replaced by:

$$\begin{aligned}\|s_1-s_2\|_2^2 \geq \|(s_1-s_2) - r^{-1}K\diamond(s_1-s_2)\|_2^2 &+ \frac{8\|s_1+s_2\|_\infty}{n\beta} \\ &+ \frac{2}{nr\sqrt{\beta}}\left(2+\kappa\sqrt{2}\ln\left(\frac{2}{\alpha}\right)\right)\sqrt{C_K}.\end{aligned}$$

We can even add an infimum over r in the right hand side of the condition, since r can be arbitrarily chosen. This exactly leads to the result in the case 3.

The results in the cases 1 and 2 are obtained by taking $r = 1$ and controlling C_K in these two cases.

Control of C_K in Case 1. We consider an orthonormal basis $\{\varphi_\lambda, \lambda \in \Lambda\}$ of a D-dimensional subspace S of $\mathbb{L}^2(\mathcal{Z}, d\nu)$ and $K(z,z') = \sum_{\lambda\in\Lambda} \varphi_\lambda(z)\varphi_\lambda(z')$. In this case,

$$K \diamond (s_1 - s_2) = \sum_{\lambda\in\Lambda} \left(\int_{\mathcal{Z}} \varphi_\lambda(z)(s_1 - s_2)(z) d\nu_z \right) \varphi_\lambda = \Pi_S(s_1 - s_2).$$

Moreover, since the dimension of S is assumed to be finite, equal to D,

$$\begin{aligned} C_K &\le \|s_1 + s_2\|_\infty^2 \int_{\mathcal{Z}} \left(\sum_{\lambda\in\Lambda} \varphi_\lambda(z)\varphi_\lambda(z') \right)^2 d\nu_z d\nu_{z'} \\ &\le \|s_1 + s_2\|_\infty^2 D. \end{aligned}$$

Control of C_K in Case 2. Assume now that $\mathcal{Z} = \mathbb{R}^d$ and introduce an approximation kernel such that $\int k^2(z) d\nu_z < +\infty$ and $k(-z) = k(z)$, $h = (h_1, \ldots, h_d)$, with $h_i > 0$ for every i, and $K(z,z') = k_h(z - z')$, with $k_h(z_1, \ldots, z_d) = \frac{1}{\prod_{i=1}^d h_i} k\left(\frac{z_1}{h_1}, \ldots, \frac{z_d}{h_d}\right)$. In this case, $K \diamond (s_1 - s_2) = k_h * (s_1 - s_2)$, and

$$\begin{aligned} C_K &= \int_{\mathcal{Z}} k_h^2(z - z')(s_1 + s_2)(z)(s_1 + s_2)(z') d\nu_z d\nu_{z'} \\ &\le \|s_1 + s_2\|_\infty \int_{\mathcal{Z}} k_h^2(z - z')(s_1 + s_2)(z) d\nu_z d\nu_{z'}, \\ &\le \frac{\|s_1 + s_2\|_\infty \|s_1 + s_2\|_1 \|k\|_2^2}{\prod_{i=1}^d h_i}. \end{aligned}$$