



HAL
open science

The two-sample problem for Poisson processes: adaptive tests with a non-asymptotic wild bootstrap approach

Patricia Reynaud-Bouret, Magalie Fromont, Béatrice Laurent

► To cite this version:

Patricia Reynaud-Bouret, Magalie Fromont, Béatrice Laurent. The two-sample problem for Poisson processes: adaptive tests with a non-asymptotic wild bootstrap approach. 2012. hal-00679102v1

HAL Id: hal-00679102

<https://hal.science/hal-00679102v1>

Preprint submitted on 14 Mar 2012 (v1), last revised 13 Nov 2012 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THE TWO-SAMPLE PROBLEM FOR POISSON PROCESSES: ADAPTIVE TESTS WITH A NON-ASYMPTOTIC WILD BOOTSTRAP APPROACH

BY MAGALIE FROMONT^{*}, BÉATRICE LAURENT[†] AND PATRICIA REYNAUD-BOURET[‡]

^{*}*CREST Ensai - IRMAR*, [†]*IMT, INSA Toulouse* and [‡]*Laboratoire Jean-Alexandre Dieudonné, CNRS, Université de Nice Sophia-Antipolis*

Abstract Considering two independent Poisson processes, we address the question of testing equality of their respective intensities. We construct multiple testing procedures from the aggregation of single tests whose testing statistics come from model selection, thresholding and/or kernel estimation methods. The corresponding critical values are computed through a non-asymptotic wild bootstrap approach. The obtained tests are proved to be exactly of level α , and to satisfy non-asymptotic oracle type inequalities. From these oracle type inequalities, we deduce that our tests are adaptive in the minimax sense over a large variety of classes of alternatives based on classical and weak Besov bodies in the univariate case, but also Sobolev and anisotropic Nikol'skii-Besov balls in the multivariate case. A simulation study furthermore shows that they strongly perform in practice.

1. Introduction. We consider the two-sample problem for general Poisson processes. Let N^1 and N^{-1} be two independent Poisson processes observed on a measurable space \mathbb{X} , whose intensities with respect to some non-atomic σ -finite measure μ on \mathbb{X} are denoted by f and g . Given the observation of N^1 and N^{-1} , we address the question of testing the null hypothesis (H_0) " $f = g$ " against the alternative (H_1) " $f \neq g$ ".

Many papers deal with the two-sample problem for homogeneous Poisson processes such as, among others, the historical ones of [47], [13], [23], or [54], whose applications were mainly turned to biology and medicine, and less frequently to reliability. More recent papers like [39], [44], [12], and [11] give interesting numerical comparisons of various testing procedures. As for non-homogeneous Poisson processes, though we can find a lot of references on the problem of testing proportionality of the hazard rates of the processes (see [17] for instance for references), we found very few papers devoted to a

AMS 2000 subject classifications: Primary 62G09, 62G10, 62G55; secondary 62G20

Keywords and phrases: two-sample problem, Poisson process, bootstrap, adaptive tests, minimax separation rates, kernel methods, aggregation methods

comparison of the intensities themselves. Bovett and Saw [8] and Deshpande et al. [18] respectively proposed conditional and unconditional procedures to test the null hypothesis " f/g is constant" against "it is increasing". Deshpande et al. [18] considered their test from a usual asymptotic point of view, proving that it is consistent against several large classes of alternatives.

We propose in this paper to construct testing procedures of (H_0) against (H_1) which satisfy specific non-asymptotic performance properties.

More precisely, given some fixed α and β in $(0, 1)$, we prove that our tests are exactly of level α and that they have a probability of second kind error at most equal to β under "optimal" conditions for various classes of alternatives. In order to specify what is the meaning of "optimal" conditions for a class \mathcal{S}_δ of alternatives (f, g) for which $(f - g)$ is smooth with parameter δ , we introduce for any level α test Φ_α with values in $\{0, 1\}$ (rejecting (H_0) when $\Phi_\alpha = 1$) its uniform separation rate $\rho(\Phi_\alpha, \mathcal{S}_\delta, \beta)$ over \mathcal{S}_δ defined by

$$(1.1) \quad \rho(\Phi_\alpha, \mathcal{S}_\delta, \beta) = \inf \left\{ \rho > 0, \sup_{(f,g) \in \mathcal{S}_\delta, \|f-g\| > \rho} \mathbb{P}_{f,g}(\Phi_\alpha = 0) \leq \beta \right\}.$$

$\mathbb{P}_{f,g}$ here denotes the joint distribution of (N^1, N^{-1}) , and $\|\cdot\|$ is a \mathbb{L}^2 -norm which will be defined more precisely later in the paper. Considering some classes of alternatives based on classical Besov bodies, weak Besov bodies as in [21] in the univariate case, but also Sobolev and anisotropic Nikol'skii-Besov balls in the multivariate case, we prove that the uniform separation rates of our testing procedures achieve their best possible values over many of such classes simultaneously. In other words, we prove that our level α tests Φ_α have a uniform separation rate over \mathcal{S}_δ of the same order (up to a possible unavoidable small loss) as

$$(1.2) \quad \underline{\rho}(\mathcal{S}_\delta, \alpha, \beta) = \inf_{\Phi_\alpha} \rho(\Phi_\alpha, \mathcal{S}_\delta, \beta),$$

where the infimum is taken over all level α tests Φ_α and where \mathcal{S}_δ is a class of alternatives (f, g) for which $(f - g)$ belongs to a classical or weak Besov body, a Sobolev or an anisotropic Nikol'skii-Besov ball with smoothness parameter δ , for many δ simultaneously. Such tests are said to be adaptive in a minimax sense.

Note that the quantity $\underline{\rho}(\mathcal{S}_\delta, \alpha, \beta)$ introduced by Baraud [4] as the (α, β) -minimax rate of testing over \mathcal{S}_δ or the minimax separation rate over \mathcal{S}_δ is a stronger version of the - asymptotic - minimax rate of testing usually considered. In order to deduce from our results recognizable asymptotic rates of testing, we assume that the measure μ on \mathbb{X} satisfies $d\mu = nd\nu$, where n can be seen as a growing number whereas the measure ν is held fixed.

Typically, n is an integer and the above assumption amounts to considering the Poisson processes N^1 and N^{-1} as n pooled i.i.d. Poisson processes with respective intensity f and g w.r.t. ν . In the most general case, one can always assume that n is just a positive real number and not necessarily an integer. The reader may also assume for sake of simplicity that ν is the Lebesgue measure when \mathbb{X} is a measurable subset of \mathbb{R}^d , but it is not required at this stage: ν may be any positive measure on \mathbb{X} .

Many papers are devoted to the computation of asymptotic minimax rates of testing in various classical statistical models, hence we obviously do not give an exhaustive bibliography on this topic. We can cite for instance the key references that are the series of papers due to Ingster [31], and the paper by Spokoiny [56] who first proved that adaptive testing without a small loss of efficiency is impossible when considering classical Besov balls as classes of alternatives in the Gaussian white noise model.

In the one-sample Poisson process model, Ingster and Kutoyants [33] proved that for the goodness-of-fit testing problem, the minimax rate of testing over a Sobolev or a Besov ball $\mathcal{B}_{2,q}^{(\delta)}(R)$ (with $1 \leq q < +\infty$ and smoothness parameter $\delta > 0$) for the Sobolev or Besov norm or semi-norm is of order $n^{-2\delta/(4\delta+1)}$. For the problem of testing that the intensity is constant, we obtained in [21] new lower bounds for the (α, β) -minimax rates of testing over classical and weak Besov bodies. These lower bounds are similar to Ingster and Kutoyants' ones for classical Besov bodies, but for weak Besov bodies, the bounds are somewhat surprising: they are so large that they coincide with the minimax estimation rates and that there is no additional loss of efficiency due to adaptivity.

In the present two-sample problem for Poisson processes, no previous minimax results are available to our knowledge. However, we prove in Section 4.3.1 that some lower bounds can be deduced from the ones in the one-sample Poisson process model. Hence, our lower bounds in [21] also hold in the present problem. We can also refer to some known results in the classical i.i.d. two-sample problem, which is closely related to the present problem. We thus cite for instance the results obtained in [30], and [10]. Ingster [30] propose adaptive procedures whose testing statistics are related to a model selection estimation approach, while Butucea and Tribouley [10] propose an adaptive procedure whose testing statistic is based on a thresholding approach.

We consider in this paper a rather large variety of classes of alternatives, on univariate but also on multivariate spaces. Hence, in order to construct some general testing procedures, which are adaptive in the minimax sense over several of these classes simultaneously, we call for various approaches,

including model selection, thresholding, and kernel ones. Our testing procedures actually consist in multiple testing procedures obtained from the aggregation of model selection, thresholding, and/or kernel based single tests. The aggregation of model selection and thresholding based tests was already proposed in different statistical models in [56], [57], [5] and [21] where the considered classes of alternatives were univariate Besov bodies or balls. Our choice of introducing additional kernel based tests allows us to obtain minimax results for classes of alternatives based on multivariate Sobolev or anisotropic Nikol'skii-Besov balls. Notice that by kernel, we mean either a classical approximation kernel or a Mercer kernel (also called learning kernel), more usual in statistical learning theory. Approximation kernels were used for instance in [28] for the Gaussian white noise testing problem, while Mercer kernels were used for instance in [27] for the multivariate i.i.d. two-sample testing problem.

As far as here, we only mentioned the fact that our tests are proved to be exactly of level α , but we did not explain how we ensure that. In the classical i.i.d. two-sample problem, the question of the choice for the critical values in testing procedures is a well-known crucial question. Indeed, the asymptotic distributions of many testing statistics are not free from the common unknown density under the null hypothesis. In such cases, general bootstrap methods are often used to build data driven critical values. Except in the cases where the permutation bootstrap method is used, authors generally prove that the obtained tests are (only) asymptotically of level α .

In this work, we adopt one of these general bootstrap approach, but from the non-asymptotic point of view: the critical values of our tests are constructed from a wild bootstrap method, that can be viewed as an adapted version of the permutation bootstrap method in a Poisson framework. We prove that the corresponding tests are actually exactly of level α , and then we also address the question of the loss due to the Monte Carlo approximation of the wild bootstrapped critical values.

Section 2 is entirely devoted to bootstrap approaches, with general considerations first, and with a detailed description of our own wild bootstrap approach in the present two-sample problem for Poisson processes then. The rest of the paper is organized as follows. In Section 3, we introduce single tests, whose statistics are each based on one so-called kernel function defined from either an orthonormal basis of a subspace of $\mathbb{L}^2(\mathbb{X})$, or an approximation kernel, or a Mercer kernel. The corresponding critical values are constructed from the wild bootstrap approach described in Section 2, leading to exact level α single tests. We then give conditions such that these single tests also have a probability of second kind error at most equal to β ,

and finally we study the cost due to the Monte Carlo approximation of the wild bootstrapped critical values. In Section 4, we construct exact level α multiple tests by aggregating several of the single tests introduced in Section 3. The advantage of such multiple tests is that they satisfy non-asymptotic oracle type inequalities, hence we can prove that they are also adaptive in the minimax sense over various classes of alternatives based on classical and weak Besov bodies in the univariate case, or Sobolev and anisotropic Nikol'skii-Besov balls in the multivariate case. Then, a simulation study is proposed in Section 5. The proofs of our theoretical results are finally postponed to Section 6.

Let us now introduce some notations that will be used all along the paper. For any measurable function h , we denote by $\|h\|_\infty = \sup_{x \in \mathbb{X}} |h(x)|$, $\|h\|_\mu = (\int_{\mathbb{X}} h(x)^2 d\mu_x)^{1/2}$, $\|h\|_{1,\mu} = \int_{\mathbb{X}} |h(x)| d\mu_x$, and $\|h\| = (\int_{\mathbb{X}} h(x)^2 d\nu_x)^{1/2}$, $\|h\|_1 = \int_{\mathbb{X}} |h(x)| d\nu_x$ when they exist. We denote by $\langle \cdot, \cdot \rangle_\mu$ the scalar product associated with $\|\cdot\|_\mu$ and $\langle \cdot, \cdot \rangle$ the scalar product associated with $\|\cdot\|$. We denote by dN^1 and dN^{-1} the point measures associated with N^1 and N^{-1} respectively, and to suit for the notation $\mathbb{P}_{f,g}$ of the joint distribution of (N^1, N^{-1}) , we denote by $\mathbb{E}_{f,g}$ the corresponding expectation. We set for any event \mathcal{A} based on (N^1, N^{-1}) , $\mathbb{P}_{(H_0)}(\mathcal{A}) = \sup_{\{f,g\}, f=g} \mathbb{P}_{f,g}(\mathcal{A})$.

Furthermore, we introduce all along the paper some constants, that we do not intend to evaluate here, and that are denoted by $C(\alpha, \beta, \dots)$ meaning that they may depend on α, β, \dots . Though they are denoted in the same way, they may vary from one line to another line.

Finally, let us make the two following assumptions, which together imply that f and g belong to $\mathbb{L}^2(\mathbb{X}, d\mu)$, and which will be satisfied all along the paper, except when specified.

ASSUMPTION 1. $\|f\|_{1,\mu} < +\infty$ and $\|g\|_{1,\mu} < +\infty$.

ASSUMPTION 2. $\|f\|_\infty < +\infty$ and $\|g\|_\infty < +\infty$.

2. A wild bootstrap approach for marked Poisson processes.

2.1. *General considerations on bootstrap methods.* Bootstrap methods were introduced first by Efron [19] whose aim was to generalize and improve the ideas of the jackknife from Quenouille [48] and Tukey [59]. These methods were originally developed for an i.i.d. sample $X = (X_1, \dots, X_n)$ from a distribution P , and a root $R_n = R_n(X; P)$, defined as a functional of the sample X and the common distribution P , whose probabilistic characteristics having a particular interest from a statistical point of view (distribution,

or expectation, variance, quantiles, etc.) are unknown, and have to be estimated. Denoting by \mathbb{P}_n the empirical measure associated with X defined by $\mathbb{P}_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$, Efron's original idea was to replace in the expression of $R_n(X; P)$, P by \mathbb{P}_n , and X by an i.i.d. sample from \mathbb{P}_n denoted by $X_n^* = (X_{n,1}^*, \dots, X_{n,n}^*)$ and called a bootstrap sample from X . The conditional distribution of the resulting "bootstrapped" root $R_n^* = R_n(X_n^*; \mathbb{P}_n)$ given X is then proposed as an estimator of the distribution of R_n .

This intuitive estimation method was justified theoretically by asymptotic arguments that were first specific to the considered root and its probabilistic characteristics of interest (see [19] and for instance many papers on the bootstrap of the mean, and linear or related statistics like [55], [7], [3], [24] among others). These arguments generally rely on a result of consistency such as:

$$\mathcal{L}(R_n^*|X) \simeq \mathcal{L}(R_n),$$

meaning that the conditional distribution of R_n^* given X converges in probability or almost surely to the asymptotic distribution of R_n . Then, considering the empirical process $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P)$, some general results on the consistency of the bootstrapped empirical process $\mathbb{G}_n^* = \sqrt{n}(\mathbb{P}_n^* - \mathbb{P}_n)$ where $\mathbb{P}_n^* = \frac{1}{n} \sum_{i=1}^n \delta_{X_{n,i}^*}$, were obtained in [7], [25], or [38].

From a practical point of view, taking advantage from the fact that a realization of the bootstrap sample X_n^* given $(X_1, \dots, X_n) = (x_1, \dots, x_n)$ can be simulated by simply taking n values with replacement in the set $\{x_1, \dots, x_n\}$, statisticians often do not strive to compute exactly the probabilistic characteristics of the bootstrapped root R_n^* , but rather approximate them by Monte Carlo procedures. This explains the frequent confusion between the term of bootstrap and the one of resampling, which is more related to the mechanism at stake in the Monte Carlo procedures following the bootstrap estimation. If we introduce for every $i = 1 \dots n$, the random variable $M_{n,i}$ defined as the number of times that X_i appears in the bootstrap sample X_n^* , it is easy to see that the bootstrapped empirical process satisfies:

$$\mathbb{G}_n^* = \sqrt{n}(\mathbb{P}_n^* - \mathbb{P}_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (M_{n,i} - 1) \delta_{X_i},$$

and that any linear root R_n^* can be expressed as a function of X and the $M_{n,i}$'s only. The random vector $(M_{n,1}, \dots, M_{n,n})$ which has a multinomial distribution with parameters $(n, n^{-1}, \dots, n^{-1})$ is viewed as a resampling plan, and the $M_{n,i}$'s as the resampling weights of the bootstrap method.

Starting from this observation, many authors proposed to study other types of resampling weights, and to replace $(M_{n,1}, \dots, M_{n,n})$ by any exchangeable random (or not) vector $(W_{n,1}, \dots, W_{n,n})$. This allowed to see

some well-known methods such as Fisher's permutation ones or cross validation ones as some bootstrap methods (see [51], [45], and [2] for more details). This also led to various new types of bootstrap methods such as the m out of n bootstrap introduced by Bretagnolle [9], the Bayesian bootstrap of Rubin [52] and Lo [41], whose resampling weights have a Dirichlet distribution, Weng's [61] bootstrap, and the wild bootstrap whose weights are i.i.d. variables with expectation and variance equal to 1, and which is detailed in [42]. Præstgaard and Wellner [46] proved an analogue of Giné and Zinn's theorem from [25] for the general exchangeable weighted bootstrapped empirical process under appropriate conditions on the weights $(W_{n,1}, \dots, W_{n,n})$.

It is now admitted that these new types of bootstrap methods are very useful in many situations, and particularly when the root R_n is a degenerate U -statistic, and/or a testing statistic in the i.i.d. two-sample problem.

2.2. Bootstrap methods turned towards degenerate U -statistics. Let us recall that a U -statistic of order p ($1 \leq p \leq n$) can be defined by

$$U_{p,n}(h) = \binom{n}{2}^{-1} \sum_{1 \leq i_1 < \dots < i_p \leq n} h(X_{i_1}, \dots, X_{i_p})$$

for some symmetric measurable function h . $U_{p,n}(h)$ is said to be degenerate when $h \in \mathbb{L}^1(dP)$ and $\int h(x_1, \dots, x_p) dP(x_i) = 0$ for every $i \in \{1, \dots, p\}$. Bootstrap methods were widely investigated for U -statistics since they often appear in general statistical questions. In particular, for non-degenerate U -statistics of order 2, the consistency of Efron's bootstrap is proved in [7]. This becomes more complicated for degenerated U -statistics. Bretagnolle [9] actually proved that a naive use of Efron's bootstrap for $R_n(X; P) = n(U_{p,n}(h) - \mathbb{E}[h(X_{i_1}, \dots, X_{i_p})])$ fails in this case. This can be briefly explained by the fact that the distribution of $R_n(X; P)$ is not a uniformly continuous function of P , so that replacing naively P by \mathbb{P}_n in the expression of $R_n(X; P)$ results in losing the degeneracy property. Bretagnolle solved the problem by reducing the size of the bootstrap sample and introducing the so-called m out of n bootstrap. Arcones and Giné [1] gave another solution which consists in forcing in some sense the bootstrapped root to keep the degeneracy property. For instance, when $p = 2$, it consists in taking \bar{R}_n^*

instead of R_n^* where

$$\bar{R}_n^* = n \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \left[h(X_{n,i}^*, X_{n,j}^*) - \frac{1}{n} \sum_{k=1}^n h(X_k, X_{n,j}^*) - \frac{1}{n} \sum_{l=1}^n h(X_{n,i}^*, X_l) + \frac{1}{n^2} \sum_{k,l=1}^n h(X_k, X_l) \right].$$

Noticing that

$$\bar{R}_n^* = n \binom{n}{2}^{-1} \sum_{k,l=1}^n (M_{n,k} - 1)(M_{n,l} - 1) h(X_k, X_l),$$

this naturally led to generalizations of Arcones and Giné's results to other kinds of bootstrapped U -statistics, in particular Bayesian and wild bootstrapped U -statistics (see [29], [16], and [35]).

A simple wild bootstrap method consists for instance in replacing the resampling weights $(M_{n,1}, \dots, M_{n,n})$ by $(W_{n,1}, \dots, W_{n,n}) = 2(B_1, \dots, B_n)$, where the B_k 's are i.i.d. random variables from the Bernoulli distribution with parameter $1/2$. Then the wild bootstrapped root can be written as

$$(2.1) \quad \bar{R}_n^\varepsilon = n \binom{n}{2}^{-1} \sum_{k,l=1}^n \varepsilon_k \varepsilon_l h(X_k, X_l),$$

where $(\varepsilon_1, \dots, \varepsilon_n) = (2B_1 - 1, \dots, 2B_n - 1)$ is an i.i.d. sample from the Rademacher distribution.

2.3. Bootstrap methods turned towards the i.i.d. two-sample problem. Recently, Janssen and Pauls [36] gave a complete study on the use of various bootstrap methods in statistical testing contexts.

Given a prescribed level $\alpha \in (0, 1)$, following the Neyman and Pearson principle, a test has at least to be exactly or asymptotically of level α . For instance, when the test rejects the null hypothesis if $R_n > c_{\alpha,n}$, R_n has to be a relevant statistic allowing to distinguish between the null and alternative hypotheses, and $c_{\alpha,n}$ is a critical value which has to guarantee that the test is actually exactly or asymptotically of level α . In the easiest cases, the distribution of R_n under the null hypothesis is exactly or asymptotically known, so that its quantiles or its asymptotic quantiles are also known or at least easily approximated by classical Monte Carlo methods. The $(1 - \alpha)$ quantile is then chosen as the critical value $c_{\alpha,n}$. When the distribution of R_n under the null hypothesis depends on the unknown underlying distribution,

one has however to turn to alternative approaches, such as bootstrap ones. A usual practice is to estimate the quantiles or asymptotic quantiles of R_n under the null hypothesis by the conditional quantiles of R_n^* given the observed variables and to approximate them by a Monte Carlo method. In the i.i.d. two-sample problem, one considers $X = (X_1^1, \dots, X_{n_1}^1, X_1^{-1}, \dots, X_{n_1}^{-1})$ where $X^1 = (X_1^1, \dots, X_{n_1}^1)$ and $X^{-1} = (X_1^{-1}, \dots, X_{n_1}^{-1})$ are two independent i.i.d. samples from respective unknown distributions P^1 and P^{-1} , and one aims at testing (H_0) " $P^1 = P^{-1}$ " against (H_1) " $P^1 \neq P^{-1}$ ". In this case, it is frequent that the distribution or the asymptotic distribution of the testing statistic under (H_0) is not free from the common distribution $P^1 = P^{-1}$. For instance, Kolmogorov-Smirnov statistics whose distributions or asymptotic distributions under (H_0) are free from $P^1 = P^{-1}$ under continuity hypotheses when the data are univariate, however fail to be distribution free under (H_0) when the data are multivariate. Bootstrap methods are now commonly used to overcome this difficulty (see [6] for the precursor permutation approach, [50], [51], [20] chapters 15 and 16, [45], and more recently [36] for interesting and helpful discussions). For instance, let us consider as in [45] the general two-sample Kolmogorov-Smirnov statistic indexed by a class $\mathcal{F}_{n_1+n_1}$ contained in a Donsker class under P^1 and P^{-1} , and denoted by R_{n_1, n_1} . Let R_{n_1, n_1}^* be the corresponding permutation or Efron's bootstrapped statistic and $c_{n_1, n_1}^*(1 - \alpha)$ be the $(1 - \alpha)$ quantile of the conditional distribution of R_{n_1, n_1}^* given X . Præstgaard proved that under (H_0) and appropriate conditions on $\mathcal{F}_{n_1+n_1}$, when $n_1, n_1 \rightarrow \infty$ with $n_1/(n_1 + n_1) \rightarrow \lambda \in (0, 1)$,

$$(2.2) \quad \mathcal{L}(R_{n_1, n_1}^* | X) \simeq \mathcal{L}(R_{n_1, n_1})$$

in probability and that $c_{n_1, n_1}^*(1 - \alpha)$ converges to the $(1 - \alpha)$ quantile of the asymptotic distribution in probability. It follows that the test rejecting (H_0) when $R_{n_1, n_1} > c_{n_1, n_1}^*(1 - \alpha)$ is asymptotically of level α . With similar arguments, Præstgaard also proved that the test is consistent against any alternative.

2.4. Bootstrap methods turned towards our two-sample problem for Poisson processes. Let us come back to the present two-sample problem for Poisson processes, and recall that we aim at testing (H_0) " $f = g$ " against (H_1) " $f \neq g$ ", from the observation of two independent Poisson processes N^1 and N^{-1} with respective intensities f and g . The single testing statistics that we consider in this paper are degenerate U -statistics of order 2. In order to construct adequate critical values, we have chosen to use the simple wild bootstrap approach described in (2.1). Of course, we are not in an i.i.d.

one-sample framework here, but we can easily imagine to generalize (2.1) to the case where X is replaced by (N^1, N^{-1}) , and P by $\mathbb{P}_{f,g}$.

More precisely, one of the single testing statistics considered in this paper is defined by

$$(2.3) \quad \hat{T} = \sum_{\lambda \in \Lambda} \sum_{x \neq x' \in N} \psi_\lambda(x) \psi_\lambda(x') \varepsilon_x^0 \varepsilon_{x'}^0,$$

where N is the pooled Poisson process such that $dN = dN^1 + dN^{-1}$, $\varepsilon_x^0 = 1$ if x belongs to N^1 and $\varepsilon_x^0 = -1$ if x belongs to N^{-1} , and $\{\psi_\lambda, \lambda \in \Lambda\}$ is an orthonormal basis of a finite dimensional subspace S of $\mathbb{L}^2(\mathbb{X}, d\mu)$. We introduce

$$(2.4) \quad \hat{T}^\varepsilon = \sum_{\lambda \in \Lambda} \sum_{i \neq i' \in \{1, \dots, N_n\}} \psi_\lambda(X_i) \psi_\lambda(X_{i'}) \varepsilon_i \varepsilon_{i'},$$

where N_n denotes the size of the pooled process N , the X_i 's denote the points of N , and $(\varepsilon_i)_{i \in \mathbb{N}}$ is a sequence of i.i.d. Rademacher variables independent of N . Under (H_0) , \hat{T}^ε has the same distribution as the true wild bootstrapped version $\sum_{\lambda \in \Lambda} \sum_{i \neq i' \in \{1, \dots, N_n\}} \psi_\lambda(X_i) \psi_\lambda(X_{i'}) \varepsilon_{X_i}^0 \varepsilon_{X_{i'}}^0 \varepsilon_i \varepsilon_{i'}$ of \hat{T} .

We then introduce the $(1 - \alpha)$ quantile of \hat{T}^ε conditionally to N that we denote by $q_{1-\alpha}^{(N)}$. Finally, we decide to reject (H_0) when $\hat{T} > q_{1-\alpha}^{(N)}$, where the conditional quantile $q_{1-\alpha}^{(N)}$ is in fact approximated by a classical Monte Carlo method.

Notice that this test is very close to Wellner's [60] permutation test in an i.i.d. two-sample framework, and we can see it as an adapted version of Wellner's test to our Poisson process framework. However, one of the particularities of our work, as compared to Wellner's and most of previous authors' ones, is that we control the probabilities of first and second kind errors of our tests from a non-asymptotic point of view. Whereas, for instance, Præstgaard [45] proves a strong result like (2.2) and the convergence of $c_{n_1, n-1}^* (1 - \alpha)$ to the $(1 - \alpha)$ asymptotic quantile in probability, we prove that under (H_0) , conditionally to N , \hat{T} and \hat{T}^ε exactly have the same distribution. Hence, under (H_0) , $\mathbb{P} \left(\hat{T} > q_{1-\alpha}^{(N)} | N \right) \leq \alpha$, as a consequence

$$\mathbb{P}_{(H_0)} \left(\hat{T} > q_{1-\alpha}^{(N)} \right) \leq \alpha,$$

which means that our test is exactly of level α . In the same way, whereas many authors prove that their bootstrapped tests are consistent against some or all alternatives, we give precise conditions on the alternatives which guarantee that our test has a prescribed probability of second kind error.

Such non-asymptotic results derive from the following proposition, which can be deduced from a general result of [14], but whose quite easy and complete proof is given in Section 6 for sake of understanding.

PROPOSITION 1.

(i) Let N^1 and N^{-1} be two independent Poisson processes on a metric space \mathbb{X} with intensities f and g with respect to some measure μ on \mathbb{X} and such that Assumption 1 is satisfied. Then the pooled process N whose point measure is given by $dN = dN^1 + dN^{-1}$ is a Poisson process on \mathbb{X} with intensity $f + g$ with respect to μ . Let $(\varepsilon_x^0)_{x \in N}$ be defined by $\varepsilon_x^0 = 1$ if x belongs to N^1 and $\varepsilon_x^0 = -1$ if x belongs to N^{-1} . Then conditionally to N , the variables $(\varepsilon_x^0)_{x \in N}$ are i.i.d. and

$$(2.5) \quad \forall x \in N, \mathbb{P}(\varepsilon_x^0 = 1|N) = \frac{f(x)}{f(x) + g(x)}, \quad \mathbb{P}(\varepsilon_x^0 = -1|N) = \frac{g(x)}{f(x) + g(x)},$$

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

(ii) Respectively, let N be a Poisson process on \mathbb{X} with intensity $f + g$ with respect to some measure μ . Let $(\varepsilon_x)_{x \in N}$ be a family of random variables with values in $\{-1, 1\}$ such that, conditionally to N , the variables $(\varepsilon_x)_{x \in N}$ are i.i.d. and

$$\forall x \in N, \mathbb{P}(\varepsilon_x = 1|N) = \frac{f(x)}{f(x) + g(x)}, \quad \mathbb{P}(\varepsilon_x = -1|N) = \frac{g(x)}{f(x) + g(x)},$$

with the convention that $0/0 = 1/2$. Then the point processes N^1 and N^{-1} , respectively defined by the point measures $dN_x^1 = \mathbf{1}_{\varepsilon_x=1}dN_x$ and $dN_x^{-1} = \mathbf{1}_{\varepsilon_x=-1}dN_x$ are two independent Poisson processes with respective intensities f and g with respect to μ on \mathbb{X} .

Proposition 1 allows us to exactly validate the wild bootstrap approach in our context.

Of course, looking at the performance of our test and the validity of the wild bootstrap approach from a non-asymptotic point of view poses the additional question of the exact loss due to the Monte Carlo approximation of $q_{1-\alpha}^{(N)}$. We address this question in Section 3.3.

3. Single testing procedures based on a general kernel function.

3.1. *Description of the single testing procedures.* Assuming that f and g belong to $\mathbb{L}^2(\mathbb{X}, d\mu)$ (which in particular occurs when Assumptions 1 and 2 are satisfied), a natural idea to test (H_0) " $f = g$ " against (H_1) " $f \neq g$ "

is to use an estimator of $\|f - g\|_\mu^2$ as testing statistic. We consider a finite dimensional subspace S of $L^2(\mathbb{X}, d\mu)$, an orthonormal basis $\{\psi_\lambda, \lambda \in \Lambda\}$ of S for $\langle \cdot, \cdot \rangle_\mu$, and the random variable \hat{T} defined by (2.3). \hat{T} can be rewritten as

$$\hat{T} = \sum_{\lambda \in \Lambda} \left(\left(\int_{\mathbb{X}} \psi_\lambda dN^1 - \int_{\mathbb{X}} \psi_\lambda dN^{-1} \right)^2 - \int_{\mathbb{X}} \psi_\lambda^2 dN \right).$$

Since $\mathbb{E} \left[\left(\int_{\mathbb{X}} \psi_\lambda dN^1 \right)^2 \right] = \left(\int \psi_\lambda(x) f(x) d\mu_x \right)^2 + \int \psi_\lambda^2(x) f(x) d\mu_x$, and similarly for $\mathbb{E} \left[\left(\int_{\mathbb{X}} \psi_\lambda dN^{-1} \right)^2 \right]$, it is easy to see that

$$\mathbb{E}_{f,g}[\hat{T}] = \|\Pi_S(f - g)\|_\mu^2,$$

where Π_S is the orthogonal projection onto S for $\langle \cdot, \cdot \rangle_\mu$. Hence \hat{T} is an unbiased estimator of $\|\Pi_S(f - g)\|_\mu^2$, and the choice of \hat{T} as testing statistic may be relevant.

From Proposition 1, we deduce that under (H_0) and conditionally to N , \hat{T} and \hat{T}_ε defined by (2.4) have the same distribution. Hence, as we have seen in Section 2.4, given a prescribed level α in $(0, 1)$, if $q_{1-\alpha}^{(N)}$ is the $(1 - \alpha)$ quantile of \hat{T}_ε conditionally to N , the test which rejects (H_0) when $\hat{T} > q_{1-\alpha}^{(N)}$ is exactly of level α .

We want to generalize this test, starting from the remark that \hat{T} can also be expressed as

$$\hat{T} = \sum_{x \neq x' \in N} K(x, x') \varepsilon_x^0 \varepsilon_{x'}^0,$$

where $K(x, x') = \sum_{\lambda \in \Lambda} \psi_\lambda(x) \psi_\lambda(x')$.

Let now K be any symmetric kernel function: $\mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}$ satisfying the following assumption.

ASSUMPTION 3. $\int_{\mathbb{X}^2} K^2(x, x') (f + g)(x) (f + g)(x') d\mu_x d\mu_{x'} < +\infty$.

Denoting by $\mathbb{X}^{[2]}$ the set $\{(x, x') \in \mathbb{X}^2, x \neq x'\}$, we introduce the statistic

$$(3.1) \quad \hat{T}_K = \sum_{x \neq x' \in N} K(x, x') \varepsilon_x^0 \varepsilon_{x'}^0 = \int_{\mathbb{X}^{[2]}} K(x, x') \varepsilon_x^0 \varepsilon_{x'}^0 dN_x dN_{x'}.$$

In the following, we use the notation \diamond_μ for the operator defined by:

$$(3.2) \quad K \diamond_\mu p(x') = \int_{\mathbb{X}} K(x, x') p(x) d\mu_x,$$

and by analogy

$$(3.3) \quad K \diamond p(x') = \int_{\mathbb{X}} K(x, x') p(x) d\nu_x.$$

Let us now introduce the quantity

$$\mathcal{E}_K = \langle K \diamond_{\mu} (f - g), f - g \rangle_{\mu},$$

which is well-defined thanks to Assumptions 1 and 3.

Since for every x in N , $\mathbb{E}[\varepsilon_x^0 | N] = (f(x) - g(x)) / (f(x) + g(x))$,

$$\begin{aligned} \mathbb{E}_{f,g}[\hat{T}_K] &= \mathbb{E}_{f,g} \left[\mathbb{E} \left[\int_{\mathbb{X}^{[2]}} K(x, x') \varepsilon_x^0 \varepsilon_{x'}^0 dN_x dN_{x'} \mid N \right] \right], \\ &= \mathbb{E}_{f,g} \left[\int_{\mathbb{X}^{[2]}} K(x, x') \frac{f(x) - g(x)}{f(x) + g(x)} \frac{f(x') - g(x')}{f(x') + g(x')} dN_x dN_{x'} \right] \\ &= \int_{\mathbb{X}^2} K(x, x') (f - g)(x) (f - g)(x') d\mu_x d\mu_{x'} \\ &= \mathcal{E}_K. \end{aligned}$$

Hence, the statistic \hat{T}_K is an unbiased estimator of \mathcal{E}_K .

We have chosen to consider and study in this paper three possible examples of symmetric kernel function K . For each example, by giving a simple expression of \mathcal{E}_K , we explain here why \hat{T}_K may be a relevant statistic for the problem of testing (H_0) " $f = g$ " against (H_1) " $f \neq g$ ".

1. Our first choice for K is a symmetric kernel function based on an orthonormal family $\{\psi_{\lambda}, \lambda \in \Lambda\}$ for $\langle \cdot, \cdot \rangle_{\mu}$:

$$K(x, x') = \sum_{\lambda \in \Lambda} \psi_{\lambda}(x) \psi_{\lambda}(x').$$

When the cardinality of Λ is finite, \hat{T}_K corresponds to our first natural testing statistic \hat{T} . When the cardinality of Λ is infinite, we assume that

$$\sup_{x, x' \in \mathbb{X}} \sum_{\lambda \in \Lambda} |\psi_{\lambda}(x) \psi_{\lambda}(x')| < +\infty,$$

which ensures that $K(x, x')$ is defined for all x, x' in \mathbb{X} and that Assumption 3 holds. Typically, if $\mathbb{X} = \mathbb{R}^d$ and if the functions $\{\psi_{\lambda}, \lambda \in \Lambda\}$ correspond to indicator functions with disjoint supports, this condition will be satisfied.

We check in these cases that

$$\mathcal{E}_K = \|\Pi_S(f - g)\|_{\mu}^2,$$

where Π_S denotes the orthogonal projection onto the subspace S generated by $\{\psi_\lambda, \lambda \in \Lambda\}$, that is \hat{T}_K is an unbiased estimator of $\|\Pi_S(f - g)\|_\mu^2$. Hence, when $\{\psi_\lambda, \lambda \in \Lambda\}$ is well-chosen, \hat{T}_K can be viewed as a relevant estimator of $\|f - g\|_\mu^2$.

It may also be interesting to consider an orthonormal family $\{\varphi_\lambda, \lambda \in \Lambda\}$ for $\langle \cdot, \cdot \rangle$ and

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

Recalling that $\mu = n\nu$, it is easy to pass from this case to the previous one by setting $\psi_\lambda = \varphi_\lambda/\sqrt{n}$: the new \hat{T}_K computed with the φ_λ 's is just n times the previous \hat{T}_K computed with the ψ_λ 's, as well as \mathcal{E}_K .

2. When $\mathbb{X} = \mathbb{R}^d$ 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 $k(-x) = k(x)$: for $x = (x_1, \dots, x_d)$, $x' = (x'_1, \dots, x'_d)$ in \mathbb{X} ,

$$K(x, x') = \frac{1}{\prod_{i=1}^d h_i} k\left(\frac{x_1 - x'_1}{h_1}, \dots, \frac{x_d - x'_d}{h_d}\right),$$

where $h = (h_1, \dots, h_d)$ is a vector of d positive bandwidths. Notice that the assumption that $k \in \mathbb{L}^2(\mathbb{R}^d)$ together with Assumption 2 ensure that Assumption 3 holds. Then, in this case,

$$\mathcal{E}_K = \langle k_h *_\mu (f - g), f - g \rangle_\mu,$$

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 $*_\mu$ is the usual convolution operator with respect to the measure μ .

3. Our third choice for K is a general Mercer or learning kernel (see [53]) such that

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

where θ and \mathcal{H}_K are a representation function and a RKHS associated with K . Here, $\langle \cdot, \cdot \rangle_{\mathcal{H}_K}$ denotes the scalar product of \mathcal{H}_K . We also choose K such that it satisfies Assumption 3. This choice leads to a testing statistic close to the one in [27] for the classical i.i.d. two-sample problem when the sizes of the i.i.d. samples are equal. We will however see that the statistic is not used here exactly in the same way as in [27], since the critical value of our test is not chosen in the same way. While Gretton et al. [27] choose their critical values from either concentration inequalities, or asymptotic arguments or an asymptotic Efron's bootstrap approach, we will choose our

critical value from the exact wild bootstrap approach described above, and which is precisely justified by Proposition 1.

In this case, it is easy to see that

$$\mathcal{E}_K = \left\| \int_{\mathbb{X}} \theta(x)(f - g)(x)d\mu_x \right\|_{\mathcal{H}_K}^2 .$$

From Lemma 4 in [27], we know that this quantity corresponds to the squared Maximum Mean Discrepancy on the unit ball in the RKHS \mathcal{H}_K :

$$\mathcal{E}_K = \left(\sup_{\|r\|_{\mathcal{H}_K} \leq 1} \int r(x)(f - g)(x)d\mu_x \right)^2 .$$

Moreover, when K is a universal kernel, such as the Gaussian and the Laplacian kernels, $\mathcal{E}_K = 0$ if and only if $f = g$. Note that in the density case (i.e. when $\int_{\mathbb{X}} f(x)d\mu_x = \int_{\mathbb{X}} g(x)d\mu_x = 1$) this is sufficient to say that the kernel is characteristic in the sense of [22].

Thus, for each of the three above choices for K , we have seen that it is actually pertinent to take \hat{T}_K as testing statistic and to reject (H_0) when \hat{T}_K is larger than a critical value to be defined.

The critical value that we define here is constructed from the same wild bootstrap approach as above. Hence, we introduce

$$(3.4) \quad \hat{T}_K^\varepsilon = \sum_{i \neq i' \in \{1, \dots, N_n\}} K(X_i, X_{i'}) \varepsilon_i \varepsilon_{i'} ,$$

where $(\varepsilon_i)_{i \in \mathbb{N}}$ is a sequence of i.i.d. Rademacher variables independent of N , and we denote by $q_{K, 1-\alpha}^{(N)}$ the $(1 - \alpha)$ quantile of \hat{T}_K^ε conditionally to N . We finally consider the test that rejects (H_0) when $\hat{T}_K > q_{K, 1-\alpha}^{(N)}$, whose test function is defined by

$$(3.5) \quad \Phi_{K, \alpha} = \mathbf{1}_{\hat{T}_K > q_{K, 1-\alpha}^{(N)}} .$$

3.2. Probabilities of first and second kind errors. In this section we study the properties of the above single test $\Phi_{K, \alpha}$ defined by (3.5), through its probabilities of first and second kind errors.

Firstly, from Proposition 1, we deduce that under (H_0) , \hat{T}_K^ε and \hat{T}_K have exactly the same distribution conditionally to N . As a result, given α in

$(0, 1)$, under (H_0) , $q_{K,1-\alpha}^{(N)}$ is also the $(1 - \alpha)$ quantile of \hat{T}_K conditionally to N and under (H_0) ,

$$\mathbb{P}\left(\hat{T}_K > q_{K,1-\alpha}^{(N)} \mid N\right) \leq \alpha.$$

By taking the expectation over N , we obtain that

$$\mathbb{P}_{(H_0)}(\Phi_{K,\alpha} = 1) \leq \alpha.$$

Secondly, given β in $(0, 1)$, we here want to bring out an exact condition on the alternative (f, g) which will guarantee that

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

Let us introduce the $1 - \beta/2$ quantile of the conditional quantile $q_{K,1-\alpha}^{(N)}$ that we denote by $q_{1-\beta/2}^\alpha$. Then for any (f, g) ,

$$\mathbb{P}_{f,g}(\Phi_{K,\alpha} = 0) \leq \mathbb{P}_{f,g}(\hat{T}_K \leq q_{1-\beta/2}^\alpha) + \beta/2,$$

and a condition which guarantees $\mathbb{P}_{f,g}(\hat{T}_K \leq q_{1-\beta/2}^\alpha) \leq \beta/2$ will be enough to ensure that $\mathbb{P}_{f,g}(\Phi_{K,\alpha} = 0) \leq \beta$. The following proposition gives such a condition. It is crucial in the proofs and understanding of the results of the paper.

PROPOSITION 2. *Let α, β be fixed levels in $(0, 1)$. For any symmetric kernel function K satisfying Assumption 3, we recall that*

$$\mathbb{E}_{f,g}[\hat{T}_K] = \mathcal{E}_K = \langle K \diamond_\mu (f - g), f - g \rangle_\mu,$$

with \diamond_μ defined by (3.2), and we introduce

$$\begin{aligned} A_K &= \int_{\mathbb{X}} (K \diamond_\mu (f - g)(x))^2 (f + g)(x) d\mu_x, \\ B_K &= \int_{\mathbb{X}^2} K^2(x, x') (f + g)(x) (f + g)(x') d\mu_x d\mu_{x'}. \end{aligned}$$

If

$$(3.6) \quad \mathcal{E}_K > 2\sqrt{\frac{2A_K + B_K}{\beta}} + q_{1-\beta/2}^\alpha,$$

then $\mathbb{P}_{f,g}(\hat{T}_K \leq q_{1-\beta/2}^\alpha) \leq \beta/2$, so that

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

Moreover, there exists some positive constant κ such that, for every K ,

$$(3.7) \quad q_{1-\beta/2}^\alpha \leq \kappa \ln(2/\alpha) \sqrt{\frac{2B_K}{\beta}}.$$

To prove the first part of this result, we simply use classical Markov inequalities since obtaining precise constants and dependency in β is not crucial here (see Section 6). The control of $q_{1-\beta/2}^\alpha$ is obtained from a property of Rademacher chaos combined with an exponential inequality (see [15] and [40]).

Note that in the Mercer kernel case, we are taking here the work of [27] further by giving, in the Poissonian context, a non-asymptotic condition to guarantee a certain power and this condition non surprisingly is about the squared Maximum Mean Discrepancy \mathcal{E}_K being large enough.

Furthermore, considering the asymptotic setting, i.e. where $d\mu = nd\nu$, allows to better understand Proposition 2 and to deduce from it recognizable properties in terms of uniform separation rates. Considering each of our three choices for the kernel K , and evaluating A_K and B_K in these cases, we actually obtain the following theorem.

THEOREM 3.1. *Let α, β be fixed levels in $(0, 1)$. Let $\kappa > 0$ be the constant of Proposition 2. Let $\Phi_{K,\alpha}$ be the test function defined by (3.5), where K is a symmetric kernel function chosen as in one of the three following different cases.*

1. *Let $\{\varphi_\lambda, \lambda \in \Lambda\}$ be an orthonormal basis of a linear subspace S of $\mathbb{L}^2(\mathbb{X}, d\nu)$. We assume either that the dimension of S is finite, equal to D , or that the two following conditions hold*

$$(3.8) \quad \sup_{x, x' \in \mathbb{X}} \sum_{\lambda \in \Lambda} |\varphi_\lambda(x)\varphi_\lambda(x')| = D < +\infty$$

$$(3.9) \quad \int_{\mathbb{X}^2} \left(\sum_{\lambda \in \Lambda} |\varphi_\lambda(x)\varphi_\lambda(x')| \right)^2 (f+g)(x') d\nu_x d\nu_{x'} < +\infty.$$

In both cases, we define $K(x, x') = \sum_{\lambda \in \Lambda} \varphi_\lambda(x)\varphi_\lambda(x')$, and we assume that

$$(3.10) \quad \|f-g\|^2 \geq \|(f-g) - \Pi_S(f-g)\|^2 + \frac{(4 + 2\sqrt{2}\kappa \ln(2/\alpha))}{n\sqrt{\beta}} M(f, g) \sqrt{D} + \frac{8\|f+g\|_\infty}{\beta n},$$

with $M(f, g) = \|f+g\|_\infty$ if the dimension of S is finite and $M(f, g) = \sqrt{\|f+g\|_\infty \|f+g\|_1}$ if the dimension of S is infinite.

2. Let $\mathbb{X} = \mathbb{R}^d$ and ν be the Lebesgue measure on \mathbb{X} . Let k be an approximation kernel in $\mathbb{L}^2(\mathbb{R}^d)$, such that $k(x) = k(-x)$, $h = (h_1, \dots, h_d)$ with $h_i > 0$ for every $i = 1 \dots d$, and $K(x, x') = k_h(x - x')$, with $k_h(u_1, \dots, u_d) = \frac{1}{\prod_{i=1}^d h_i} k\left(\frac{u_1}{h_1}, \dots, \frac{u_d}{h_d}\right)$. We assume that

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

3. Let K be a Mercer kernel associated with a representation function θ and a RKHS \mathcal{H}_K such that Assumption 3 holds.

We define $C_K = \int_{\mathbb{X}^2} K^2(x, x')(f + g)(x)(f + g)(x') d\nu_x d\nu_{x'}$, and we assume that

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

Then in all these three cases, one can guarantee that

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

Comments.

1. In the first case, we can see that the right hand side of (3.10) 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 for this kind of test (see [4] for instance), and we know that this leads to sharp upper bounds for the uniform separation rates over particular classes of alternatives.

2. The second case also reproduces a bias-variance decomposition when $k \in \mathbb{L}^1(\mathbb{R}^d)$ and $\int_{\mathbb{R}^d} k(x) d\nu_x = 1$: the bias is here $\|(f - g) - k_h * (f - g)\|$. When $h_1 = \dots = h_d$, the variance term is of order $h_1^{-d/2}/n$. As usual in the approximation kernel estimation theory, this coincides with what is found in the first case through the equivalence $h_1^{-d} \sim D$ (see [58] for instance for more details).

3. The third case is unusual, since the term $\|(f - g) - r^{-1}K \diamond (f - g)\|$ can not always be viewed as a bias term. Indeed, Mercer kernels are frequently used in statistical learning, but their approximation capacity is not necessarily considered. In particular when we consider the first case where the family

$\{\varphi_\lambda, \lambda \in \Lambda\}$ is orthonormal w.r.t. $\langle \cdot, \cdot \rangle$ or the second case with $k \in \mathbb{L}^1(\mathbb{R}^d)$ and $\int_{\mathbb{R}^d} k(x) d\nu_x = 1$, these conditions give a link between the measure ν and the statistics leading to potentially fruitful approximation properties. When Mercer kernels are involved, the only usual normalization consists in ensuring that $K(x, x) = 1$. This normalization corresponds to a normalization of the vector $\theta(x)$ in the RKHS. However since the RKHS has in general no link with the underlying measure ν , there is in general no classical normalization for our present problem. That is the reason why we forced such a normalization via the factor r . Note that performing the procedure with K or $r^{-1}K$ exactly leads to the same test. Thus, there are special cases for which one can go a bit further, and recover the results given in the first and second cases. Let us consider $K(x, x') = c\langle \varphi(x), \varphi(x') \rangle_{\mathbb{R}^\Lambda}$ with $\varphi(x) = (\varphi_\lambda(x))_{\lambda \in \Lambda}$ and $c > 0$. K is a Mercer kernel usually known as a projection kernel with $\mathcal{H}_K = \mathbb{R}^\Lambda$ and $\theta(x) = \sqrt{c}\varphi(x)$. It is not mandatory in statistical learning for instance that the family $(\varphi_\lambda(x))_{\lambda \in \Lambda}$ is orthonormal for $\langle \cdot, \cdot \rangle$. However, when we add this assumption, taking $r = c$ exactly leads to the result given in the first case. Let us now consider $\mathbb{X} = \mathbb{R}$ and ν the Lebesgue measure on \mathbb{R} . Let $K(x, x') = \exp(-(x - x')^2 / (2h^2))$ be the classical one-dimensional Gaussian Mercer kernel with variance h^2 . Let us denote by k the density of the standard Gaussian distribution and let $k_h = (1/h)k(\cdot/h)$. Then we have $K(x, x') = h\sqrt{2\pi}k_h(x - x')$. Applying the result stated in the third case with $r = h\sqrt{2\pi}$ also leads to the result stated in the second case.

When we are not in these special cases, though we can not really give a statistical interpretation of our result, we however think that the introduction of Mercer kernels may help if the space \mathbb{X} is unusual or pretty large with respect to the (mean) number of observations and/or if 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 [27] where such kernels are used for microarrays data and graphs).

3.3. Performances of the Monte Carlo procedure.

3.3.1. *Probability of first kind error.* In practice, a Monte Carlo procedure will be used to estimate the conditional quantiles $q_{K,1-\alpha}^{(N)}$. It is therefore natural to address the following question: what can we say about the probabilities of first and second kind errors of the test built with these Monte Carlo approximations? Recall that we consider the test $\Phi_{K,\alpha}$ rejecting (H_0) when $\hat{T}_K > q_{K,1-\alpha}^{(N)}$, where \hat{T}_K is defined by (3.1), and $q_{K,1-\alpha}^{(N)}$ is the $(1-\alpha)$ quantile of \hat{T}_K^ε defined by (3.4) conditionally to N . $q_{K,1-\alpha}^{(N)}$ is estimated by $\hat{q}_{K,1-\alpha}^{(N)}$ via the Monte Carlo method as follows. Conditionally to N , we consider a set of

B independent sequences $\{\varepsilon^b, 1 \leq b \leq B\}$, where $\varepsilon^b = (\varepsilon_x^b)_{x \in N}$ is a sequence of i.i.d. Rademacher random variables. We define for $1 \leq b \leq B$:

$$\hat{T}_K^{\varepsilon^b} = \sum_{x \neq x' \in N} K(x, x') \varepsilon_x^b \varepsilon_{x'}^b.$$

Under (H_0) , conditionally to N , the variables $\hat{T}_K^{\varepsilon^b}$ have the same distribution function as \hat{T}_K , which is denoted by F_K . We denote by $F_{K,B}$ the empirical distribution function (conditionally to N) of the sample $(\hat{T}_K^{\varepsilon^b}, 1 \leq b \leq B)$:

$$\forall x \in \mathbb{R}, F_{K,B}(x) = \frac{1}{B} \sum_{b=1}^B \mathbf{1}_{\hat{T}_K^{\varepsilon^b} \leq x}.$$

Then, $\hat{q}_{K,1-\alpha}^{(N)}$ is defined by

$$\hat{q}_{K,1-\alpha}^{(N)} = F_{K,B}^{-1}(1 - \alpha) = \inf \{t \in \mathbb{R}, F_{K,B}(t) \geq 1 - \alpha\}.$$

We finally consider the test given by

$$(3.11) \quad \hat{\Phi}_{K,\alpha} = \mathbf{1}_{\hat{T}_K > \hat{q}_{K,1-\alpha}^{(N)}}.$$

The following proposition gives an upper bound for the level of $\hat{\Phi}_{K,\alpha}$.

PROPOSITION 3. *Let $\hat{\Phi}_{K,\alpha}$ be the test defined by (3.11). Under the null hypothesis (H_0) "f = g",*

$$\mathbb{P}_{f,g} \left(\hat{\Phi}_{K,\alpha} = 1 \mid N \right) \leq \alpha + \inf_{\lambda > 0} \left(\frac{\lambda}{\sqrt{B}} + 2e^{-2\lambda^2} \right).$$

For example, if we take $\alpha = 0.05$ and $B = 200000$, the level of the test $\hat{\Phi}_{K,\alpha}$ is controlled by 5.5%.

3.3.2. Probability of second kind error.

PROPOSITION 4. *Let α and β be some fixed levels in $(0, 1)$ such that $\alpha > \sqrt{\ln B / 4B}$ and $\beta > 4/\sqrt{B}$. Let $\hat{\Phi}_{K,\alpha}$ be the test defined by (3.11). Let \mathcal{E}_K , A_K and B_K be defined as in Proposition 2. There exists some constant $\kappa > 0$ such that if*

$$(3.12) \quad \mathcal{E}_K > 2\sqrt{\frac{2A_K + B_K}{\beta}} + \kappa \ln \left(\frac{2}{\alpha - \frac{\sqrt{\ln B}}{2\sqrt{B}}} \right) \sqrt{\frac{B_K}{\frac{\beta}{2} - \frac{2}{\sqrt{B}}}},$$

then

$$\mathbb{P}_{f,g}(\hat{\Phi}_{K,\alpha} = 0) \leq \beta.$$

When comparing the condition (3.12) with (3.6) and (3.7) in Proposition 2, we can notice that they asymptotically coincide when $B \rightarrow +\infty$. Moreover, when $\alpha = \beta = 0.05$ and $B \geq 6400$, the multiplicative factor of $\kappa\sqrt{B_K}$ is multiplied by a factor of order 2 in (3.12) compared with (3.6). If even $B = 200000$, this factor passes from 23.4 in (3.6) to 26.4 in (3.12).

4. Multiple testing procedures. 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 when the kernel K is based on such an orthonormal family, the bandwidth h when K is based on an approximation kernel, the parameters of K when it is a Mercer kernel. 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, in [27], the parameter of the kernel is chosen from a heuristic method.

In order to avoid choosing particular kernels or parameters, we propose in this section to consider some collections of kernel functions instead of a single one, and to define multiple testing procedures by aggregating the corresponding single tests. We propose an adapted choice for the critical value. Then, we prove that these multiple procedures have strong statistical performances since they satisfy oracle type inequalities and are adaptive on several classes of alternatives simultaneously.

4.1. *Description of the multiple testing procedures.* Let us introduce a finite collection $\{K_m, m \in \mathcal{M}\}$ of symmetric kernel functions: $\mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}$ satisfying Assumption 3. For every m in \mathcal{M} , let \hat{T}_{K_m} and $\hat{T}_{K_m}^\varepsilon$ be respectively defined by (3.1) and (3.4) with $K = K_m$, and 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 $u \in (0, 1)$, we denote by $q_{m,1-u}^{(N)}$ the $(1-u)$ quantile of $\hat{T}_{K_m}^\varepsilon$ conditionally to the pooled process N . Given $\alpha \in (0, 1)$, we consider the test which rejects (H_0) when there exists at least one m in \mathcal{M} such that

$$\hat{T}_{K_m} > q_{m,1-u_\alpha^{(N)}}^{(N)} e^{-w_m},$$

where $u_\alpha^{(N)}$ is defined by

$$(4.1) \quad u_\alpha^{(N)} = \sup \left\{ u > 0, \mathbb{P} \left(\sup_{m \in \mathcal{M}} \left(\hat{T}_{K_m}^\varepsilon - q_{m,1-ue^{-w_m}}^{(N)} \right) > 0 \mid N \right) \leq \alpha \right\}.$$

Let Φ_α be the corresponding test function defined by

$$(4.2) \quad \Phi_\alpha = \mathbf{1}_{\sup_{m \in \mathcal{M}} \left(\hat{T}_{K_m} - q_{m, 1-u_\alpha^{(N)}}^{(N)} e^{-w_m} \right) > 0}.$$

Note that given the observation of the pooled process N , $u_\alpha^{(N)}$ and the quantile $q_{m, 1-u_\alpha^{(N)}}^{(N)} e^{-w_m}$ can be estimated by a Monte Carlo procedure (see Section 5).

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 $\hat{T}_{K_m} > q_{m, 1-u_\alpha^{(N)}}^{(N)} e^{-w_m}$. We can thus combine the results of Theorem 3.1.

Even if the multiple testing procedure may be performed for collections of kernel functions mixing the three possible choices of kernel functions in Theorem 3.1, we have decided to focus on the first two choices, where we know that the performance of the corresponding tests is already optimal if the space S or the approximation kernel k is well-chosen. This enables us to go a little bit further in the interpretation of the results as oracle inequalities.

4.2. Oracle type inequalities for the probability of second kind error.

4.2.1. Multiple testing procedures based on orthonormal families.

THEOREM 4.1. *Let α, β be fixed levels in $(0, 1)$. Let $\{S_m, m \in \mathcal{M}\}$ be a finite collection of linear subspaces of $\mathbb{L}^2(\mathbb{X}, d\nu)$ and for all m in \mathcal{M} , let $\{\varphi_\lambda, \lambda \in \Lambda_m\}$ be an orthonormal basis of S_m for $\langle \cdot, \cdot \rangle$. We assume either that S_m has finite dimension D_m or that conditions (3.8) and (3.9) hold for $\Lambda = \Lambda_m$ and $D = D_m$. We set, for all $m \in \mathcal{M}$, $K_m(x, x') = \sum_{\lambda \in \Lambda_m} \varphi_\lambda(x) \varphi_\lambda(x')$. Let Φ_α be the test function defined by (4.2) with the collection of kernels $\{K_m, m \in \mathcal{M}\}$ and a collection $\{w_m, m \in \mathcal{M}\}$ of positive numbers such that $\sum_{m \in \mathcal{M}} e^{-w_m} \leq 1$.*

Then the test Φ_α is a level α test. Moreover, there exists some positive constant κ such that if

$$(4.3) \quad \|f - g\|^2 \geq \inf_{m \in \mathcal{M}} \left\{ \|(f - g) - \Pi_{S_m}(f - g)\|^2 + \frac{4 + 2\sqrt{2}\kappa(\ln(2/\alpha) + w_m)}{n\sqrt{\beta}} M'(f, g) \sqrt{D_m} \right\} + \frac{8\|f + g\|_\infty}{\beta n},$$

with $M'(f, g) = \max\left(\|f + g\|_\infty, \sqrt{\|f + g\|_\infty \|f + g\|_1}\right)$, then

$$\mathbb{P}_{f,g}(\Phi_\alpha = 0) \leq \beta.$$

Comparing this result with the one obtained in Theorem 3.1 for the single test based on an orthonormal family, one can see that considering the multiple testing procedure allows to obtain the infimum over all m in \mathcal{M} in the right hand side of (4.3) at the price of the additional term w_m . This result can be viewed as an oracle type inequality: indeed, without knowing $(f - g)$, 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 . It will be used to prove that our multiple testing procedures are adaptive over various classes of alternatives.

We focus here on three particular examples. The first example involves a nested collection of linear subspaces of $\mathbb{L}^2([0, 1])$, as in model selection estimation approaches. In the second example, we consider a collection of one dimensional linear subspaces of $\mathbb{L}^2([0, 1])$, and our testing procedure is hence related to a thresholding estimation approach. In the third example, we introduce a collection of infinite dimensional linear subspaces of $\mathbb{L}^2(\mathbb{R})$.

Example 1. Let $\mathbb{X} = [0, 1]$ and ν be the Lebesgue measure on $[0, 1]$. Let $\{\varphi_0, \varphi_{(j,k)}, j \in \mathbb{N}, k \in \{0, \dots, 2^j - 1\}\}$ be the Haar basis of $\mathbb{L}^2(\mathbb{X}, d\nu)$ with

$$(4.4) \quad \varphi_0(x) = \mathbf{1}_{[0,1]}(x) \quad \text{and} \quad \varphi_{(j,k)}(x) = 2^{j/2} \psi(2^j x - k),$$

where $\psi(x) = \mathbf{1}_{[0,1/2)}(x) - \mathbf{1}_{[1/2,1)}(x)$. The collection of linear subspaces $\{S_m, m \in \mathcal{M}\}$ is chosen as a collection of nested subspaces generated by subsets of the Haar basis. More precisely, we denote by S_0 the subspace of $\mathbb{L}^2([0, 1])$ generated by φ_0 , and we define $K_0(x, x') = \varphi_0(x)\varphi_0(x')$. We also consider for $J \geq 1$ the subspaces S_J generated by $\{\varphi_\lambda, \lambda \in \{0\} \cup \Lambda_J\}$ with $\Lambda_J = \{(j, k), j \in \{0, \dots, J - 1\}, k \in \{0, \dots, 2^j - 1\}\}$, and $K_J(x, x') = \sum_{\lambda \in \{0\} \cup \Lambda_J} \varphi_\lambda(x)\varphi_\lambda(x')$. Let for some $\bar{J} \geq 1$,

$$\mathcal{M}_{\bar{J}} = \{J, 0 \leq J \leq \bar{J}\},$$

and for every J in $\mathcal{M}_{\bar{J}}$, $w_J = 2(\ln(J + 1) + \ln(\pi/\sqrt{6}))$. Let $\Phi_\alpha^{(1)}$ be the test function defined by (4.2) with the collection of kernel functions $\{K_J, J \in \mathcal{M}_{\bar{J}}\}$ and with $\{w_J, J \in \mathcal{M}_{\bar{J}}\}$. We obtain from Theorem 4.1 that there

exists some constant $C(\alpha, \beta, \|f\|_\infty, \|g\|_\infty) > 0$ such that if

$$(4.5) \quad \|f - g\|^2 \geq C(\alpha, \beta, \|f\|_\infty, \|g\|_\infty) \inf_{J \in \mathcal{M}_{\tilde{J}}} \left\{ \|(f - g) - \Pi_{S_J}(f - g)\|^2 + (\ln(J + 2)) \frac{2^{J/2}}{n} \right\},$$

then $\mathbb{P}_{f,g}(\Phi_\alpha^{(1)} = 0) \leq \beta$.

Example 2. Let $\mathbb{X} = [0, 1]$ and ν be the Lebesgue measure on $[0, 1]$. Let $\{\varphi_0, \varphi_{(j,k)}, j \in \mathbb{N}, k \in \{0, \dots, 2^j - 1\}\}$ still be the Haar basis of $\mathbb{L}^2(\mathbb{X}, d\nu)$ defined by (4.4). Let for some $\tilde{J} \geq 1$,

$$\Lambda_{\tilde{J}} = \{(j, k), j \in \{0, \dots, \tilde{J} - 1\}, k \in \{0, \dots, 2^j - 1\}\}.$$

For any λ in $\{0\} \cup \Lambda_{\tilde{J}}$, we consider the subspace \tilde{S}_λ of $\mathbb{L}^2([0, 1])$ generated by φ_λ , and $K_\lambda(x, x') = \varphi_\lambda(x)\varphi_\lambda(x')$. Let $\Phi_\alpha^{(2)}$ be the test defined by (4.2) with the collection of kernel functions $\{K_\lambda, \lambda \in \{0\} \cup \Lambda_{\tilde{J}}\}$, with $w_0 = \ln(2)$, and $w_{(j,k)} = \ln(2^j) + 2(\ln(j + 1) + \ln(\pi/\sqrt{3}))$ for $j \in \mathbb{N}, k \in \{0, \dots, 2^j - 1\}$. We obtain from Theorem 4.1 and Pythagore's theorem that there is some constant $C(\alpha, \beta, \|f\|_\infty, \|g\|_\infty) > 0$ such that if there exists λ in $\{0\} \cup \Lambda_{\tilde{J}}$ for which

$$\|\Pi_{\tilde{S}_\lambda}(f - g)\|^2 \geq C(\alpha, \beta, \|f\|_\infty, \|g\|_\infty) \frac{w_\lambda}{n},$$

then $\mathbb{P}_{f,g}(\Phi_\alpha^{(2)} = 0) \leq \beta$. Notice that if $\mathcal{M}_{\tilde{J}} = \{m, m \subset \{0\} \cup \Lambda_{\tilde{J}}\}$, the above condition is equivalent to saying that there exists m in $\mathcal{M}_{\tilde{J}}$ such that

$$\|\Pi_{S_m}(f - g)\|^2 \geq C(\alpha, \beta, \|f\|_\infty, \|g\|_\infty) \frac{\sum_{\lambda \in m} w_\lambda}{n},$$

where S_m is the linear subspace generated by $\{\varphi_\lambda, \lambda \in m\}$. Hence, we obtain that there exists some constant $C(\alpha, \beta, \|f\|_\infty, \|g\|_\infty) > 0$ such that if

$$(4.6) \quad \|f - g\|^2 \geq C(\alpha, \beta, \|f\|_\infty, \|g\|_\infty) \inf_{m \in \mathcal{M}} \left\{ \|(f - g) - \Pi_{S_m}(f - g)\|^2 + \frac{\sum_{\lambda \in m} w_\lambda}{n} \right\},$$

then $\mathbb{P}_{f,g}(\Phi_\alpha^{(2)} = 0) \leq \beta$.

Example 3. Let $\mathbb{X} = \mathbb{R}$ and ν be the Lebesgue measure on \mathbb{R} . We define, for all j in \mathbb{N} and k in \mathbb{Z} , $\varphi_{j,k} = 2^{j/2} \mathbf{1}_{\left[\frac{k}{2^j}, \frac{k+1}{2^j}\right)}$. For every j in \mathbb{N} , we denote by S_j the linear subspace of $\mathbb{L}^2(\mathbb{R})$ generated by $\{\varphi_{j,k}, k \in \mathbb{Z}\}$ and $K_j(x, x') = \sum_{k \in \mathbb{Z}} \varphi_{j,k}(x) \varphi_{j,k}(x')$. Then (3.8) holds and

$$\sum_{k \in \mathbb{Z}} |\varphi_{j,k}(x) \varphi_{j,k}(x')| = 2^j \mathbf{1}_{x \in \left[\frac{k(x')}{2^j}, \frac{k(x')+1}{2^j}\right)}$$

where $k(x')$ is the integer defined by $x' \in \left[\frac{k(x')}{2^j}, \frac{k(x')+1}{2^j}\right)$. This ensures that (3.9) holds:

$$\begin{aligned} & \int_{\mathbb{R}^2} \left(\sum_{k \in \mathbb{Z}} |\varphi_{j,k}(x) \varphi_{j,k}(x')| \right)^2 (f+g)(x') d\nu_x d\nu_{x'} \\ &= 2^{2j} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbf{1}_{x \in \left[\frac{k(x')}{2^j}, \frac{k(x')+1}{2^j}\right)} d\nu_x \right) (f+g)(x') d\nu_{x'} \\ &= 2^j \int_{\mathbb{R}} (f+g)(x') d\nu_{x'} < +\infty. \end{aligned}$$

Let for some $\bar{j} \geq 1$, $\mathcal{M}_{\bar{j}} = \{j, 0 \leq j \leq \bar{j}\}$, and for every j in $\mathcal{M}_{\bar{j}}$, $w_j = 2(\ln(j+1) + \ln(\pi/\sqrt{6}))$. Let $\Phi_{\alpha}^{(3)}$ be defined by (4.2) with the collections $\{K_j, j \in \mathcal{M}_{\bar{j}}\}$ and $\{w_j, j \in \mathcal{M}_{\bar{j}}\}$. We obtain from Theorem 4.1 that there exists some positive constant $C(\alpha, \beta, \|f+g\|_{\infty}, \|f+g\|_1)$ such that if

$$\|f-g\|^2 \geq C(\alpha, \beta, \|f+g\|_{\infty}, \|f+g\|_1) \inf_{j \in \mathcal{M}_{\bar{j}}} \left\{ \|(f-g) - \Pi_{S_j}(f-g)\|^2 + \ln(j+2) \frac{2^{j/2}}{n} \right\},$$

then $\mathbb{P}_{f,g}(\Phi_{\alpha}^{(3)} = 0) \leq \beta$. Note that this example allows us to consider two independent Poisson processes that are defined on the whole real line, as well as the two examples presented in the next section.

4.2.2. Multiple testing procedures based on approximation kernels.

THEOREM 4.2. *Let α, β be fixed levels in $(0, 1)$, $\mathbb{X} = \mathbb{R}^d$ and let ν be 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 $\int_{\mathbb{X}} k_{m_1}^2(x) d\nu_x < +\infty$, $k_{m_1}(x) = k_{m_1}(-x)$, and a collection $\{h_{m_2}, m_2 \in \mathcal{M}_2\}$, where each h_{m_2} is a vector of d positive numbers*

$(h_{m_2,1}, \dots, h_{m_2,d})$. We set $\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2$, and for all $m = (m_1, m_2)$ in \mathcal{M} , $x = (x_1, \dots, x_d)$, $x' = (x'_1, \dots, x'_d)$ in \mathbb{R}^d ,

$$K_m(x, x') = k_{m_1, h_{m_2}}(x - x') = \frac{1}{\prod_{i=1}^d h_{m_2, i}} k_{m_1} \left(\frac{x_1 - x'_1}{h_{m_2, 1}}, \dots, \frac{x_d - x'_d}{h_{m_2, d}} \right).$$

Let Φ_α be the test defined by (4.2) with $\{K_m, m \in \mathcal{M}\}$ and a collection $\{w_m, m \in \mathcal{M}\}$ of positive numbers such that $\sum_{m \in \mathcal{M}} e^{-w_m} \leq 1$. Then Φ_α is a level α test. Moreover, there exists $\kappa > 0$ such that if

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

then

$$\mathbb{P}_{f, g}(\Phi_\alpha = 0) \leq \beta.$$

We focus here on two particular examples. The first example involves a collection of non necessarily integrable approximation kernels with a collection of bandwidths vectors whose components are the same in every direction. The second example involves a single integrable approximation kernel, but with a collection of bandwidths vectors whose components may differ according to every direction.

Example 4. Let $\mathbb{X} = \mathbb{R}^d$ and ν be the Lebesgue measure on \mathbb{R}^d . We set $\mathcal{M}_1 = \mathbb{N} \setminus \{0\}$ and $\mathcal{M}_2 = \mathbb{N}$. For m_1 in \mathcal{M}_1 , let k_{m_1} be a kernel such that $\int k_{m_1}^2(x) d\nu_x < +\infty$ and $k_{m_1}(x) = k_{m_1}(-x)$, non necessarily integrable, whose Fourier transform is defined when $k_{m_1} \in \mathbb{L}^1(\mathbb{R}^d) \cap \mathbb{L}^2(\mathbb{R}^d)$ by $\widehat{k_{m_1}}(u) = \int_{\mathbb{R}^d} k_{m_1}(x) e^{i\langle x, u \rangle} d\nu_x$ and is extended to $k_{m_1} \in \mathbb{L}^2(\mathbb{R}^d)$ in the Plancherel sense. We assume that for every m_1 in \mathcal{M}_1 , $\|\widehat{k_{m_1}}\|_\infty < +\infty$, and

$$(4.7) \quad \text{Ess sup}_{u \in \mathbb{R}^d \setminus \{0\}} \frac{|1 - \widehat{k_{m_1}}(u)|}{\|u\|_d^{m_1}} \leq C,$$

for some $C > 0$, where $\|u\|_d$ denotes the euclidean norm of u . Note that the sinc kernel, the spline type kernel and Pinsker's kernel given in [58] for instance satisfy this condition which can be viewed as an extension of the integrability condition (see [58] p. 26-27 for more details). For m_2 in \mathcal{M}_2 , let $h_{m_2} = (2^{-m_2}, \dots, 2^{-m_2})$ and for $m = (m_1, m_2)$ in $\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2$, let

$$K_m(x, x') = k_{m_1, h_{m_2}}(x - x') = \frac{1}{2^{-dm_2}} k_{m_1} \left(\frac{x_1 - x'_1}{2^{-m_2}}, \dots, \frac{x_d - x'_d}{2^{-m_2}} \right).$$

We take $w_{(m_1, m_2)} = 2(\ln(m_1(m_2 + 1)) + \ln(\pi^2/6))$, so $\sum_{m \in \mathcal{M}} e^{-w_m} \leq 1$. Let $\Phi_\alpha^{(4)}$ be the test function defined by (4.2) with the collection of kernel functions $\{K_m, m \in \mathcal{M}\}$ and $\{w_m, m \in \mathcal{M}\}$. We obtain from Theorem 4.2 that there exists some positive constant $C(\alpha, \beta)$ such that if

$$\|f - g\|^2 \geq C(\alpha, \beta) \left(\inf_{(m_1, m_2) \in \mathcal{M}} \left\{ \|(f - g) - k_{m_1, h_{m_2}} * (f - g)\|^2 + \frac{w_{(m_1, m_2)}}{n} \sqrt{\frac{\|f + g\|_\infty \|f + g\|_1 \|k_{m_1}\|^2}{2^{-dm_2}}} \right\} + \frac{\|f + g\|_\infty}{n} \right),$$

then $\mathbb{P}_{f, g}(\Phi_\alpha^{(4)} = 0) \leq \beta$.

Example 5. Let $\mathbb{X} = \mathbb{R}^d$ and ν be the Lebesgue measure on \mathbb{R}^d . Let $\mathcal{M}_1 = \{1\}$ and $\mathcal{M}_2 = \mathbb{N}^d$. For $x = (x_1, \dots, x_d)$ in \mathbb{R}^d , let $k_1(x) = \prod_{i=1}^d k_{1,i}(x_i)$ where the $k_{1,i}$'s are real valued kernels such that $k_{1,i} \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$, $k_{1,i}(x_i) = k_{1,i}(-x_i)$, and $\int_{\mathbb{R}} k_{1,i}(x_i) dx_i = 1$. For $m_2 = (m_{2,1}, \dots, m_{2,d})$ in \mathcal{M}_2 , $h_{m_2, i} = 2^{-m_{2,i}}$ and for $m = (m_1, m_2)$ in $\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2$,

$$K_m(x, x') = k_{m_1, h_{m_2}}(x - x') = \prod_{i=1}^d \frac{1}{h_{m_2, i}} k_{1, i} \left(\frac{x_i - x'_i}{h_{m_2, i}} \right).$$

We also set $w_{(1, m_2)} = 2 \sum_{i=1}^d (\ln(m_{2,i} + 1) + \ln(\pi/\sqrt{6}))$, so that $\sum_{m \in \mathcal{M}_1 \times \mathcal{M}_2} e^{-w_m} = 1$. Let $\Phi_\alpha^{(5)}$ be the test function defined by (4.2) with the collections $\{K_m, m \in \mathcal{M}\}$ and $\{w_m, m \in \mathcal{M}\}$. We deduce from Theorem 4.2 that there exists some positive constant $C(\alpha, \beta)$ such that if

$$\|f - g\|^2 \geq C(\alpha, \beta) \left(\inf_{m_2 \in \mathcal{M}_2} \left\{ \|(f - g) - k_{1, h_{m_2}} * (f - g)\|^2 + \frac{w_{(1, m_2)}}{n} \sqrt{\frac{\|f + g\|_\infty \|f + g\|_1 \|k_1\|^2}{\prod_{i=1}^d h_{m_2, i}}} \right\} + \frac{\|f + g\|_\infty}{n} \right),$$

then $\mathbb{P}_{f, g}(\Phi_\alpha^{(5)} = 0) \leq \beta$.

4.3. *Uniform separation rates over various classes of alternatives.* We here evaluate the uniform separation rates, defined by (1.1) with the norm $\|\cdot\|$ of $\mathbb{L}^2(\mathbb{X}, d\nu)$, of the multiple testing procedures introduced above over several classes of alternatives based on Besov and weak Besov bodies when $\mathbb{X} \subset \mathbb{R}$, or Sobolev and anisotropic Besov-Nikol'skii balls when $\mathbb{X} = \mathbb{R}^d$.

4.3.1. *Uniform separation rates for Besov and weak Besov bodies.* Let $\Phi_\alpha^{(1)}$ and $\Phi_\alpha^{(2)}$ be the tests defined in *Example 1* and *Example 2* in Section 4.2.1.

Recall that these tests are constructed from the Haar basis $\{\varphi_0, \varphi_{(j,k)}, j \in \mathbb{N}, k \in \{0, \dots, 2^j - 1\}\}$ of $\mathbb{L}^2(\mathbb{X}, d\nu)$ defined by (4.4). Let

$$\Psi_\alpha = \max(\Phi_{\alpha/2}^{(1)}, \Phi_{\alpha/2}^{(2)}).$$

Like in [21], we define for $\delta > 0$, $R > 0$ the Besov body $\mathcal{B}_{2,\infty}^\delta(R)$ as follows :

$$(4.8) \quad \mathcal{B}_{2,\infty}^\delta(R) = \left\{ s \in \mathbb{L}^2([0, 1]), s = \alpha_0 \varphi_0 + \sum_{j \in \mathbb{N}} \sum_{k=0}^{2^j-1} \alpha_{(j,k)} \varphi_{(j,k)}, \right. \\ \left. \alpha_0^2 \leq R^2 \text{ and } \forall j \in \mathbb{N}, \sum_{k=0}^{2^j-1} \alpha_{(j,k)}^2 \leq R^2 2^{-2j\delta} \right\},$$

We also consider weak Besov bodies given for $\gamma > 0$, $R' > 0$ by

$$(4.9) \quad \mathcal{W}_\gamma(R') = \left\{ s \in \mathbb{L}^2([0, 1]), s = \alpha_0 \varphi_0 + \sum_{j \in \mathbb{N}} \sum_{k=0}^{2^j-1} \alpha_{(j,k)} \varphi_{(j,k)}, \right. \\ \left. \forall t > 0, \alpha_0^2 \mathbf{1}_{\alpha_0^2 \leq t} + \sum_{j \in \mathbb{N}} \sum_{k=0}^{2^j-1} \alpha_{(j,k)}^2 \mathbf{1}_{\alpha_{(j,k)}^2 \leq t} \leq R'^2 t^{\frac{2\gamma}{1+2\gamma}} \right\}.$$

Let us introduce

$$\mathcal{B}_{\delta,\gamma,\infty}(R, R', R'') = \{(f, g), (f - g) \in \mathcal{B}_{2,\infty}^\delta(R) \cap \mathcal{W}_\gamma(R'), \\ \max(\|f\|_\infty, \|g\|_\infty) \leq R''\}.$$

COROLLARY 1. *Assume that $\ln \ln n \geq 1$, $2^{\tilde{J}} \geq n^2$, and $\tilde{J} = +\infty$. Then,*

(i) *if $\delta \geq \gamma/2$*

$$\rho(\Psi_\alpha, \mathcal{B}_{\delta,\gamma,\infty}(R, R', R''), \beta) \leq C(\delta, \gamma, R, R', R'', \alpha, \beta) \left(\frac{\ln \ln n}{n} \right)^{\frac{2\delta}{4\delta+1}},$$

(ii) *if $\delta < \gamma/2$*

$$\rho(\Psi_\alpha, \mathcal{B}_{\delta,\gamma,\infty}(R, R', R''), \beta) \leq C(\delta, \gamma, R, R', R'', \alpha, \beta) \left(\frac{\ln n}{n} \right)^{\frac{\gamma}{2\gamma+1}}.$$

Comments. Notice that our classes of alternatives are not defined in the same way as in [10] in the classical i.i.d. two-sample problem, since the classes of alternatives (f, g) of [10] are such that f and g both belong to a Besov ball. Here the smoothness condition is only required on the difference $(f - g)$. Our classes are consequently larger than those of [10]. In particular, the functions f and g might be very irregular but as long as their difference is smooth, the test will be powerful.

Let us now see why the above upper bounds are nearly optimal in the minimax sense. Recall that the optimal value for the uniform separation rate of a level α test over a class of alternatives \mathcal{S}_δ is given by the minimax separation rate $\underline{\rho}(\mathcal{S}_\delta, \alpha, \beta)$ over \mathcal{S}_δ defined by (1.2).

We deduce from the results given in [21] the following lower bounds.

PROPOSITION 5. *Assume that $R > 0$, $R' > 0$, and $R'' \geq 2$, and fix some levels α and β in $(0, 1)$ such that $\alpha + \beta \leq 0.59$.*

(i) *When $\delta \geq \max(\gamma/2, \gamma/(1 + 2\gamma))$, then*

$$\liminf_{n \rightarrow +\infty} n^{\frac{2\delta}{1+4\delta}} \underline{\rho}(\mathcal{B}_{\delta, \gamma, \infty}(R, R', R''), \alpha, \beta) > 0.$$

(ii) *When $\delta < \gamma/2$ and $\gamma > 1/2$, then*

$$\liminf_{n \rightarrow +\infty} \left(\frac{n}{\ln n} \right)^{\frac{\gamma}{1+2\gamma}} \underline{\rho}(\mathcal{B}_{\delta, \gamma, \infty}(R, R', R''), \alpha, \beta) > 0.$$

(iii) *When $\delta < \gamma/(1 + 2\gamma)$ and $\gamma \leq 1/2$, then*

$$\liminf_{n \rightarrow +\infty} \left(n^{\frac{1}{4}} \wedge n^{\frac{2\gamma}{(1+4\delta)(1+2\gamma)}} \right) \underline{\rho}(\mathcal{B}_{\delta, \gamma, \infty}(R, R', R''), \alpha, \beta) > 0.$$

Hence the test Ψ_α is adaptive in the minimax sense over $\mathcal{B}_{\delta, \gamma, \infty}(R, R', R'')$, up to a $\ln \ln n$ factor in the case (i) and exactly in the case (ii). The exact rate is however unknown in the case (iii).

Comment. For sake of simplicity we only considered the Haar basis, but any wavelet basis may be considered. The proofs of the results could easily be extended to compactly supported wavelets.

4.3.2. *Uniform separation rates for Sobolev and anisotropic Nikol'skii-Besov balls.*

Sobolev balls. Let us introduce for $\delta > 0$ the Sobolev ball $\mathcal{S}_d^\delta(R)$ defined by

$$\mathcal{S}_d^\delta(R) = \left\{ s : \mathbb{R}^d \rightarrow \mathbb{R}, s \in \mathbb{L}^1(\mathbb{R}^d) \cap \mathbb{L}^2(\mathbb{R}^d), \int_{\mathbb{R}^d} \|u\|_d^{2\delta} |\hat{s}(u)|^2 du \leq (2\pi)^d R^2 \right\},$$

where $\|u\|_d$ denotes the euclidean norm of u and \hat{s} denotes the Fourier transform of s : $\hat{s}(u) = \int_{\mathbb{R}^d} s(x) e^{i\langle x, u \rangle} dx$.

We here evaluate the uniform separation rate of $\Phi_\alpha^{(4)}$ defined in *Example 4* of Section 4.2.2 over the classes

$$\begin{aligned} \mathcal{S}_d^\delta(R, R', R'') &= \{(f, g), (f - g) \in \mathcal{S}_d^\delta(R), \\ &\quad \max(\|f\|_1, \|g\|_1) \leq R', \max(\|f\|_\infty, \|g\|_\infty) \leq R''\}, \end{aligned}$$

for every $\delta > 0$, and $R, R', R'' > 0$.

COROLLARY 2. *Assume that $\ln \ln n \geq 1$. For any $\delta > 0$, $R, R', R'' > 0$,*

$$\rho(\Phi_\alpha^{(4)}, \mathcal{S}_d^\delta(R, R', R''), \beta) \leq C(\delta, \alpha, \beta, R, R', R'', d) \left(\frac{\ln \ln n}{n} \right)^{\frac{2\delta}{d+4\delta}}.$$

Comments. From [49], we know that, in the density model, the minimax adaptive estimation rate over $\mathcal{S}_d^\delta(R)$ is of order $n^{-\frac{\delta}{d+2\delta}}$ when $\delta > d/2$. Rigollet and Tsybakov construct some aggregated density estimators, based on Pinsker's kernel, that achieve this rate with exact constants. In the same way, the test $\Phi_\alpha^{(4)}$ consists in an aggregation of some tests based on a collection of kernels, that may be for instance a collection of Pinsker's kernels. It achieves over $\mathcal{S}_d^\delta(R, R', R'')$ a uniform separation rate of order $n^{-\frac{2\delta}{d+4\delta}}$ up to a $\ln \ln n$ factor. This rate is now known to be the optimal adaptive minimax rate of testing when $d = 1$ in several models (see [56] in a Gaussian model or [32] in the density model for instance). From the results of [28], we can conjecture that our rates are also optimal when $d > 1$.

Nikol'skii-Besov anisotropic balls. Let $\delta = (\delta_1, \dots, \delta_d)$ with $\delta_i > 0$ for every $i = 1 \dots d$ and $R > 0$. For all t in \mathbb{R} , we denote by $\lfloor t \rfloor$ the integer such that $t - \lfloor t \rfloor \in (0, 1]$. We consider the anisotropic Nikol'skii-Besov ball $\mathcal{N}_{2,d}^\delta(R)$ defined by:

$$(4.10) \quad \mathcal{N}_{2,d}^\delta(R) = \left\{ s : \mathbb{R}^d \rightarrow \mathbb{R}, s \text{ has continuous partial derivatives } D_i^{\lfloor \delta_i \rfloor} \right. \\ \left. \text{of order } \lfloor \delta_i \rfloor \text{ w.r.t } u_i, \text{ and } \|D_i^{\lfloor \delta_i \rfloor} s(u_1, \dots, u_i + v, \dots, u_d) - D_i^{\lfloor \delta_i \rfloor} s(u_1, \dots, u_d)\|_2 \right. \\ \left. \leq R |v|^{\delta_i - \lfloor \delta_i \rfloor}, \forall i = 1 \dots d, u_1, \dots, u_d, v \in \mathbb{R} \right\}.$$

Let $\Phi_\alpha^{(5)}$ be the test defined in *Example 5* in Section 4.2.2. Assume furthermore that $\int_{\mathbb{R}} |k_{1,i}(x_i)| |x_i|^{\delta_i} dx_i < +\infty$, and $\int_{\mathbb{R}} k_{1,i}(x_i) x_i^j dx_i = 0$ for every $i = 1 \dots d$ and $j = 1 \dots \lfloor \delta_i \rfloor$. We here evaluate the uniform separation rate of $\Phi_\alpha^{(5)}$ over the classes

$$\mathcal{N}_{2,d}^\delta(R, R', R'') = \{(f, g), (f - g) \in \mathcal{N}_{2,d}^\delta(R), \\ \max(\|f\|_1, \|g\|_1) \leq R', \max(\|f\|_\infty, \|g\|_\infty) \leq R''\},$$

for every $\delta \in]0, +\infty[^d$ and $R, R', R'' > 0$.

COROLLARY 3. *Assume that $\ln \ln n \geq 1$. For any $\delta_i > 0$, $i = 1 \dots d$, and $R, R', R'' > 0$, setting $1/\bar{\delta} = \sum_{i=1}^d 1/\delta_i$,*

$$\rho(\Phi_\alpha^{(5)}, \mathcal{N}_{2,d}^\delta(R, R', R''), \beta) \leq C(\delta, \alpha, \beta, R, R', R'', d) \left(\frac{\ln \ln n}{n} \right)^{\frac{2\bar{\delta}}{1+4\bar{\delta}}}.$$

Comments. When $d = 1$, from [32], we know that in the density model, the adaptive minimax rate of testing over a Nikol'skii class with smoothness parameter δ is of order $(\ln \ln n/n)^{2\delta/(1+4\delta)}$. We find here an upper bound similar to this univariate rate, but where δ is replaced by $\bar{\delta}$. Such results were obtained in a multivariate density estimation context in [26] where the adaptive minimax estimation rates over the anisotropic Nikol'skii classes are proved to be of order $n^{-\bar{\delta}/(1+2\bar{\delta})}$, and where adaptive kernel density estimators are proposed. Moreover, the minimax rates of testing obtained recently in [34] over anisotropic periodic Sobolev balls, but in the Gaussian white noise model, are of the same order as the upper bounds obtained here.

5. Simulation study.

5.1. *Presentation of the simulation study.* In this section, we study our testing procedures from a practical point of view. We consider $\mathbb{X} = [0, 1]$ or $\mathbb{X} = \mathbb{R}$, $n = 100$ and ν the Lebesgue measure on \mathbb{X} . N^1 and N^{-1} denote two independent Poisson processes with intensities f and g on \mathbb{X} with respect to μ with $d\mu = 100d\nu$. We focus on several couples of intensities (f, g) defined on \mathbb{X} and such that $\int_{\mathbb{X}} f(x) d\nu_x = \int_{\mathbb{X}} g(x) d\nu_x = 1$. We choose $\alpha = 0.05$. Conditionally to the number of points of both processes N^1 and N^{-1} , the points of N^1 and N^{-1} form two independent samples of i.i.d. variables with densities f and g with respect to ν . Hence, conditionally to the number of points of N^1 and N^{-1} , any test for the classical i.i.d. two-sample problem can be used here. We compare our tests to the conditional Kolmogorov-Smirnov test. Thus we consider five testing procedures, that we respectively denote by KS, Ne, Th, G, E.

KS corresponds to the conditional Kolmogorov-Smirnov test.

The tests Ne and Th respectively correspond to the test functions $\Phi_\alpha^{(1)}$ and $\Phi_\alpha^{(2)}$ defined in *Example 1* and *Example 2* in Section 4.2.1 with $\bar{J} = 7$ and $\tilde{J} = 6$.

The tests G and E are similar to the test defined by $\Phi_\alpha^{(5)}$ in *Example 5* in Section 4.2.2. For the test G, we consider the standard Gaussian approximation kernel defined by $k(x) = (2\pi)^{-1/2} \exp(-x^2/2)$ for all $x \in \mathbb{R}$ and for the test E, we consider the Epanechnikov approximation kernel defined by $k(x) = (3/4)(1 - x^2)\mathbf{1}_{|x| \leq 1}$. For both tests, we consider $\{h_m, m \in \mathcal{M}\} = \{1/24, 1/16, 1/12, 1/8, 1/4, 1/2\}$ and the associated collection of kernel functions $\{K_m, m \in \mathcal{M}\}$ given for all m in \mathcal{M} by $K_m(x, x') = \frac{1}{h_m} k\left(\frac{x-x'}{h_m}\right)$. We take for both tests $w_m = 1/|\mathcal{M}| = 1/6$.

Let us recall that our tests reject (H_0) when there exists m in \mathcal{M} such that

$$\hat{T}_{K_m} > q_{m, 1-u_\alpha^{(N)}}^{(N)} e^{-w_m}$$

where N is the pooled process obtained from N^1 and N^{-1} , and $u_\alpha^{(N)}$ is defined by (4.1). Hence, for each observation of the process N whose number of points is denoted by N_n , we have to estimate $u_\alpha^{(N)}$ and the quantiles $q_{m, 1-u_\alpha^{(N)}}^{(N)} e^{-w_m}$. These estimations are done by classical Monte Carlo methods based on the simulation of 400000 independent samples of size N_n of i.i.d. Rademacher variables (see Section 3.3 for the theoretical study of these Monte Carlo methods when only one test is considered). Half of the samples is used to estimate the distribution of each $\hat{T}_{K_m}^\varepsilon$. The other half is used to approximate the conditional probabilities occurring in (4.1). The point $u_\alpha^{(N)}$ is obtained by dichotomy, such that the estimated conditional probability occurring in (4.1) is less than α , but as close as possible to α . By monotony arguments, this is equivalent to make u varying on a regular grid of $[0, 1]$ with bandwidth 2^{-16} , and to choose $u_\alpha^{(N)}$ as the largest value of the u 's on the grid such that the estimated conditional probabilities in (4.1) are less than α .

5.2. Simulation results. We use 5000 simulations to estimate the level of significance, and 1000 to estimate the powers of each tests.

For the estimation of the levels, we consider three common intensities. The first one is the uniform density on $[0, 1]$, the second one is the Beta density with parameters $(2, 5)$, and the third one is a Laplace density with parameter 7.

Let

$$\begin{aligned} f_1(x) &= \mathbf{1}_{[0,1]}(x), \\ f_{2,2,5}(x) &= \frac{x(1-x)^4}{\int_0^1 x(1-x)^4 dx} \mathbf{1}_{[0,1]}(x) \\ f_{3,7}(x) &= \frac{7}{2} e^{-7|x-1/2|}. \end{aligned}$$

Taking f as one of these three functions, we realize 5000 simulations of two independent Poisson processes N^1 and N^{-1} both with intensity f w.r.t. to μ . For each simulation, we determine the conclusions of the tests Ks, Ne, Th, G and E, where the quantiles of our four last tests are estimated by the Monte Carlo methods described above. The levels of the tests are estimated by the number of rejections for these tests divided by 5000. The results are given in the following table:

f	KS	Ne	Th	G	E
f_1	0.053	0.049	0.045	0.053	0.053
$f_{2,2,5}$	0.053	0.047	0.043	0.051	0.050
$f_{3,7}$	0.0422	0.0492	0.0438	0.054	0.055

In order to evaluate the power of the tests, we first consider couples of intensities (f, g) such that $f = f_1$ and g is successively equal to intensities that are classical examples in wavelet settings, and are defined by

$$\begin{aligned} g_{1,a,\varepsilon}(x) &= (1 + \varepsilon) \mathbf{1}_{[0,a]}(x) + (1 - \varepsilon) \mathbf{1}_{[a,2a]}(x) + \mathbf{1}_{[2a,1]}(x) \\ g_{2,\eta}(x) &= \left(1 + \eta \sum_j \frac{h_j}{2} (1 + \operatorname{sgn}(x - p_j)) \right) \frac{\mathbf{1}_{[0,1]}(x)}{C_2(\eta)}, \\ g_{3,\varepsilon}(x) &= (1 - \varepsilon) \mathbf{1}_{[0,1]}(x) + \varepsilon \left(\sum_j g_j \left(1 + \frac{|x - p_j|}{w_j} \right)^{-4} \right) \frac{\mathbf{1}_{[0,1]}(x)}{0.284}, \end{aligned}$$

where p, h, g, w, ε are defined as in [21]¹, $0 < \varepsilon \leq 1$, $0 < a < 1/2$, $\eta > 0$ and $C_2(\eta)$ is such that $\int_0^1 g_{2,\eta}(x) dx = 1$.

1

$$\begin{aligned} p = & \left(\begin{array}{cccccccccccc} 0.1 & 0.13 & 0.15 & 0.23 & 0.25 & 0.4 & 0.44 & 0.65 & 0.76 & 0.78 & 0.81 \end{array} \right) \\ h = & \left(\begin{array}{cccccccccccc} 4 & -4 & 3 & -3 & 5 & -5 & 2 & 4 & -4 & 2 & -3 \end{array} \right) \\ g = & \left(\begin{array}{cccccccccccc} 4 & 5 & 3 & 4 & 5 & 4.2 & 2.1 & 4.3 & 3.1 & 5.1 & 4.2 \end{array} \right) \\ w = & \left(\begin{array}{cccccccccccc} 0.005 & 0.005 & 0.006 & 0.01 & 0.01 & 0.03 & 0.01 & 0.01 & 0.005 & 0.008 & 0.005 \end{array} \right) \end{aligned}$$

We then consider couples of intensities (f, g) such that f is equal to the above Laplace density $f_{3,7}$ with parameter 7, or to the Laplace density $f_{3,10}$ with parameter 10:

$$f_{3,10}(x) = 5e^{-10|x-1/2|},$$

and $g = g_{4,1/2,1/4}$ is the density of a Gaussian variable with expectation $1/2$ and standard deviation $1/4$.

For each couple (f, g) , we realize 1000 simulations of two independent Poisson processes N^1 and N^{-1} with respective intensities f and g w.r.t. μ . For each simulation, we determine the conclusions of the tests Ks, Ne, Th, G and E, where the quantiles of our four last tests are estimated by the Monte Carlo methods described above. The powers of the tests are estimated by the number of rejections for these tests divided by 1000. The results are summarized in Figures 1 and 2 where in each column, the estimated power is represented as a dot for every test. The triangles represent the upper and lower bounds of an asymptotic confidence interval with confidence level 99%, with variance estimation.

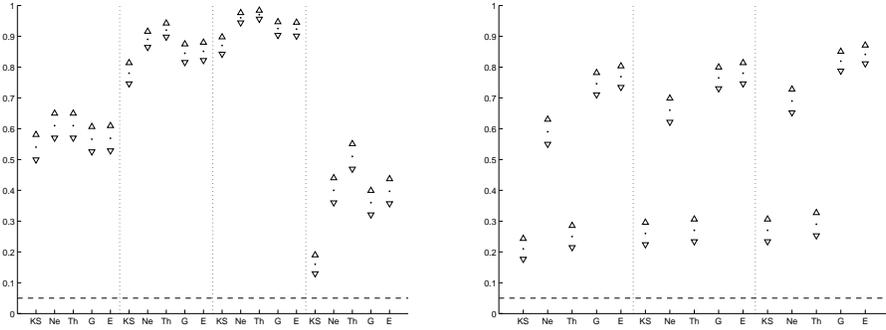


FIGURE 1. Left: $(f, g) = (f_1, g_{1,a,\varepsilon})$. Each column corresponds respectively to $(a, \varepsilon) = (1/4, 0.7)$, $(1/4, 0.9)$, $(1/4, 1)$ and $(1/8, 1)$. Right: $(f, g) = (f_1, g_{2,\eta})$. Each column corresponds respectively to $\eta = 4, 8$ and 15 .

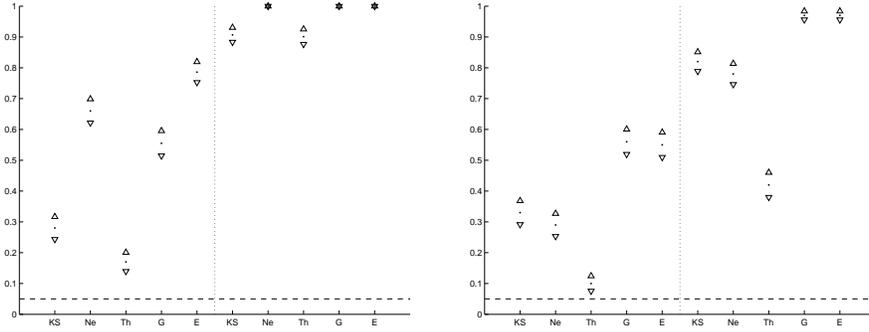


FIGURE 2. *Left:* $(f, g) = (f_1, g_{3,\varepsilon})$. The two columns correspond respectively to $\varepsilon = 0.5$ and 1. *Right:* $(f, g) = (f_{3,\lambda}, g_{4,1/2,1/4})$. The two columns correspond respectively to $\lambda = 7$ and $\lambda = 10$.

In all cases, the tests G and E based on the approximation kernels are more powerful (sometimes even about 4 times more powerful) than the KS test. This is also the case for the test Ne, except for the last example. The test Th is more powerful than the KS test for the couples $(f, g) = (f_1, g_{1,a,\varepsilon})$, but it fails to improve the KS test for the other alternatives. We conjecture that the test Th consists in the aggregation of too many single tests. We can finally notice that the test E strongly performs for every considered alternative, except in a sparse case, where the test E is less powerful than the test Th (see Figure 1). Our conclusion is that the test E is a good practical choice, except maybe when sparse processes are involved. Aggregating the tests E and Th in such cases would probably be a good compromise.

6. Proofs.

6.1. *Proof of Proposition 1.* All along the proof, \int denotes $\int_{\mathbb{X}}$. One of the key arguments of the proof is that the marked point processes are characterized by their Laplace functional (see for instance [14]).

To obtain the first point of the result, this key argument makes sufficient to compute $\mathbb{E} \left[\exp \left(\int h dN \right) \right]$ for a bounded measurable function h on \mathbb{X} . By independency,

$$\mathbb{E} \left[\exp \left(\int h dN \right) \right] = \mathbb{E} \left[\exp \left(\int h dN^1 \right) \right] \mathbb{E} \left[\exp \left(\int h dN^{-1} \right) \right].$$

Since the Laplace functional of N^1 is given by

$$\mathbb{E} \left[\exp \left(\int h dN^1 \right) \right] = \exp \left(\int (e^h - 1) f d\mu \right),$$

and the Laplace functional of N^{-1} has the same form, replacing f by g ,

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

which is the Laplace functional of Poisson process with intensity $(f + g)$ w.r.t. μ .

Let us now prove (2.5). The distribution of $(\varepsilon_x^0)_{x \in N}$ conditionally to N is characterized by the function

$$t = (t_x)_{x \in N} \mapsto \Phi(t, N) = \mathbb{E} \left[\exp \left(\sum_{x \in N} t_x \varepsilon_x^0 \right) \middle| N \right].$$

Let λ be a bounded measurable function defined on \mathbb{X} , and define

$$\mathbb{E}_\lambda = \mathbb{E} \left[\exp \left(\int \lambda dN \right) \exp \left(\sum_{x \in N} t_x \varepsilon_x^0 \right) \right].$$

Since N^1 and N^{-1} are independent,

$$\begin{aligned} \mathbb{E}_\lambda &= \mathbb{E} \left[\exp \left(\int (\lambda(x) + t_x) dN_x^1 \right) \exp \left(\int (\lambda(x) - t_x) dN_x^{-1} \right) \right] \\ &= \mathbb{E} \left[\exp \left(\int (\lambda(x) + t_x) dN_x^1 \right) \right] \mathbb{E} \left[\exp \left(\int (\lambda(x) - t_x) dN_x^{-1} \right) \right]. \end{aligned}$$

Then,

$$\begin{aligned} \mathbb{E}_\lambda &= \exp \left[\int (e^{(\lambda(x)+t_x)} - 1) f(x) + (e^{(\lambda(x)-t_x)} - 1) g(x) \right] d\mu_x \\ &= \exp \int (e^{h(x)} - 1) (f + g)(x) d\mu_x \\ &= \mathbb{E} \left[\exp \left(\int h dN \right) \right] \end{aligned}$$

where

$$h(x) = \lambda(x) + \ln \left(\frac{e^{t_x} f(x) + e^{-t_x} g(x)}{(f + g)(x)} \right).$$

Hence, for every bounded measurable function λ defined on \mathbb{X} ,

$$\begin{aligned} \mathbb{E} \left[\exp \left(\int \lambda dN \right) \exp \left(\sum_{x \in N} t_x \varepsilon_x^0 \right) \right] &= \\ &= \mathbb{E} \left[\exp \left(\int \lambda dN \right) \prod_{x \in N} \left(e^{t_x} \frac{f(x)}{(f + g)(x)} + e^{-t_x} \frac{g(x)}{(f + g)(x)} \right) \right]. \end{aligned}$$

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

$$\Phi(t, N) = \mathbb{E} \left[\exp \left(\sum_{x \in N} t_x \varepsilon_x^0 \right) \middle| N \right] = \prod_{x \in N} \left(e^{t_x} \frac{f(x)}{(f+g)(x)} + e^{-t_x} \frac{g(x)}{(f+g)(x)} \right),$$

which concludes the proof of (2.5).

To prove the second point of the result, let h_1 and h_{-1} be two bounded measurable functions on \mathbb{X} .

$$\begin{aligned} & \mathbb{E} \left[\exp \left(\int h_1 dN^1 + \int h_{-1} dN^{-1} \right) \right] \\ &= \mathbb{E} \left[\mathbb{E} \left[\exp \left(\int h_1 dN^1 + \int h_{-1} dN^{-1} \right) \middle| N \right] \right] \\ &= \mathbb{E} \left[\mathbb{E} \left[\prod_{x \in N} \exp (h_1(x) \mathbf{1}_{\varepsilon_x=1} + h_{-1}(x) \mathbf{1}_{\varepsilon_x=-1}) \middle| N \right] \right]. \end{aligned}$$

Remark that there is almost surely a finite number of points in N and that if x belongs to N , then $f(x) + g(x) > 0$. Moreover

$$\begin{aligned} \mathbb{E} [\exp (h_1(x) \mathbf{1}_{\varepsilon_x=1} + h_{-1}(x) \mathbf{1}_{\varepsilon_x=-1})] &= e^{h_1(x)} \frac{f(x)}{f(x) + g(x)} \\ &\quad + e^{h_{-1}(x)} \frac{g(x)}{f(x) + g(x)}. \end{aligned}$$

Then using the expression of the Laplace functional of N , with the function

$$h = \ln \left(e^{h_1(x)} \frac{f(x)}{f(x) + g(x)} + e^{h_{-1}(x)} \frac{g(x)}{f(x) + g(x)} \right),$$

leads to

$$\begin{aligned} & \mathbb{E} \left[\exp \left(\int h_1 dN^1 + \int h_{-1} dN^{-1} \right) \right] \\ &= \exp \left(\int \left(e^{h_1(x)} \frac{f(x)}{f(x) + g(x)} + e^{h_{-1}(x)} \frac{g(x)}{f(x) + g(x)} - 1 \right) (f+g)(x) d\mu_x \right). \end{aligned}$$

Finally we have that

$$\begin{aligned} \mathbb{E} \left[\exp \left(\int h_1 dN^1 + \int h_{-1} dN^{-1} \right) \right] &= \\ & \exp \left(\int (e^{h_1} - 1) f d\mu \right) \exp \left(\int (e^{h_{-1}} - 1) g d\mu \right). \end{aligned}$$

We here recognize the product of the Laplace functionals of two Poisson processes with respective intensities f and g . This gives the independence and concludes the proof.

6.2. *Proof of Proposition 2.* Let us prove the first part of Proposition 2. Recall that $q_{1-\beta/2}^\alpha$ denotes the $1-\beta/2$ quantile of $q_{K,1-\alpha}^{(N)}$, which is the $(1-\alpha)$ quantile of \hat{T}_K^ε conditionally to N .

We here want to find a condition on \hat{T}_K , or more precisely on $\mathcal{E}_K = E_{f,g}[\hat{T}_K]$, ensuring that $\mathbb{P}_{f,g}(\hat{T}_K \leq q_{1-\beta/2}^\alpha) \leq \beta/2$. From Markov's inequality, we have that for any $x > 0$,

$$\mathbb{P}_{f,g} \left(\left| -\hat{T}_K + \mathcal{E}_K \right| \geq x \right) \leq \frac{\text{Var}(\hat{T}_K)}{x^2}.$$

Let us compute $\text{Var}(\hat{T}_K) = \mathbb{E}_{f,g}[\hat{T}_K^2] - \mathcal{E}_K^2$. Let $\mathbb{X}^{[3]}$ and $\mathbb{X}^{[4]}$ be the sets $\{(x, y, u) \in \mathbb{X}^3, x, y, u \text{ all different}\}$ and $\{(x, y, u, v) \in \mathbb{X}^4, x, y, u, v \text{ all different}\}$ respectively. Since

$$\mathbb{E}_{f,g}[\hat{T}_K^2] = \mathbb{E}_{f,g} \left[\mathbb{E} \left[\left(\int_{\mathbb{X}^{[2]}} K(x, x') \varepsilon_x^0 \varepsilon_{x'}^0 dN_x dN_{x'} \right)^2 \middle| N \right] \right],$$

by using (2.5),

$$\begin{aligned} \mathbb{E}_{f,g}[\hat{T}_K^2] &= \mathbb{E}_{f,g} \left[\int_{\mathbb{X}^{[4]}} K(x, y) K(u, v) \frac{f-g}{f+g}(x) \frac{f-g}{f+g}(y) \right. \\ &\quad \left. \frac{f-g}{f+g}(u) \frac{f-g}{f+g}(v) dN_x dN_y dN_u dN_v \right] \\ &\quad + 4 \mathbb{E}_{f,g} \left[\int_{\mathbb{X}^{[3]}} K(x, y) K(x, u) \frac{f-g}{f+g}(y) \frac{f-g}{f+g}(u) dN_x dN_y dN_u \right] \\ &\quad + 2 \mathbb{E}_{f,g} \left[\int_{\mathbb{X}^{[2]}} K^2(x, y) dN_x dN_y \right]. \end{aligned}$$

Now, from Lemma 5.4 III in [14] on factorial moments measures applied to

Poisson processes, we deduce that

$$\begin{aligned} \mathbb{E}_{f,g}[\hat{T}_K^2] &= \int_{\mathbb{X}^4} \left(K(x,y)K(u,v)(f-g)(x)(f-g)(y) \right. \\ &\quad \left. (f-g)(u)(f-g)(v) \right) d\mu_x d\mu_y d\mu_u d\mu_v \\ &\quad + 4 \int_{\mathbb{X}^3} K(x,y)K(x,u)(f+g)(x)(f-g)(y)(f-g)(u) d\mu_x d\mu_y d\mu_u \\ &\quad + 2 \int_{\mathbb{X}^2} K^2(x,y)(f+g)(x)(f+g)(y) d\mu_x d\mu_y \end{aligned}$$

In order to apply this lemma, we have to verify that the three integrals appearing in the above expression are finite. This follows easily from Assumptions 1, 2 et 3. We finally obtain that

$$\mathbb{E}_{f,g}[\hat{T}_K^2] = \mathcal{E}_K^2 + 4A_K + 2B_K.$$

Hence, we obtain that for any $x > 0$,

$$\mathbb{P}_{f,g} \left(\left| -\hat{T}_K + \mathcal{E}_K \right| \geq x \right) \leq \frac{4A_K + 2B_K}{x^2}.$$

Taking $x = \sqrt{(8A_K + 4B_K)/\beta}$ in the above inequality leads to

$$(6.1) \quad \mathbb{P}_{f,g} \left(\left| -\hat{T}_K + \mathcal{E}_K \right| \geq \sqrt{\frac{8A_K + 4B_K}{\beta}} \right) \leq \frac{\beta}{2}.$$

Therefore, if $\mathcal{E}_K > \sqrt{\frac{8A_K + 4B_K}{\beta}} + q_{1-\beta/2}^\alpha$, then $\mathbb{P}_{f,g}(\hat{T}_K \leq q_{1-\beta/2}^\alpha) \leq \beta/2$, so

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

Let us now give a sharp upper bound for $q_{1-\beta/2}^\alpha$.

Reasoning conditionally to N , we recognize in \hat{T}_K^ε a homogeneous Rademacher chaos as defined by de la Peña and Giné [15]: it is of the form

$$X = \sum_{i \neq i'} x_{i,i'} \varepsilon_i \varepsilon_{i'},$$

where the $x_{i,i'}$'s are some real deterministic numbers and $(\varepsilon_i)_{i \in \mathbb{N}}$ is a sequence of i.i.d. Rademacher variables. Corollary 3.2.6 of [15] states that there exists some absolute constant $\kappa > 0$ such that

$$(6.2) \quad \mathbb{E} \left(\exp \left[\frac{|X|}{\kappa \sigma} \right] \right) \leq 2,$$

where $\sigma^2 = \mathbb{E}[X^2] = \sum_{i \neq i'} x_{i,i'}^2$.

Hence by Markov's inequality again, for all positive ξ ,

$$\mathbb{P}(|X| \geq \xi) \leq 2e^{-\frac{\xi}{\kappa\sigma}},$$

which implies that $\mathbb{P}(|X| \geq \kappa\sigma \ln(2/\alpha)) \leq \alpha$, (see also [40] for more precise constants). Applying this result to \hat{T}_K^ε with $\sigma^2 = \sum_{x \neq x' \in N} K^2(x, x')$ leads to

$$(6.3) \quad q_{K,1-\alpha}^{(N)} \leq \kappa \ln(2/\alpha) \sqrt{\int_{\mathbb{X}^{[2]}} K^2(x, y) dN_x dN_y}.$$

Hence $q_{1-\beta/2}^\alpha$ is upper bounded by the $(1 - \beta/2)$ quantile of $\kappa \ln(2/\alpha) \sqrt{\int_{\mathbb{X}^{[2]}} K^2(x, y) dN_x dN_y}$.

Using Markov's inequality again and Lemma 5.4 III in [14], we obtain that for any positive ξ ,

$$\mathbb{P}_{f,g} \left(\int_{\mathbb{X}^{[2]}} K^2(x, y) dN_x dN_y \geq \xi \right) \leq \frac{B_K}{\xi}.$$

By taking $\xi = 2B_K/\beta$, we can finally see that

$$(6.4) \quad q_{1-\beta/2}^\alpha \leq \kappa \ln \left(\frac{2}{\alpha} \right) \sqrt{\frac{2B_K}{\beta}}.$$

6.3. *Proof of Theorem 3.1.* We here want to apply the result of Proposition 2 to the three possible choices of the kernel function K .

First, notice that for all $r > 0$

$$\begin{aligned} \mathcal{E}_K &= n^2 \langle K \diamond (f - g), f - g \rangle = n^2 r \langle r^{-1} K \diamond (f - g), f - g \rangle, \\ &= \frac{n^2 r}{2} (\|f - g\|^2 + r^{-2} \|K \diamond (f - g)\|^2 - \|(f - g) - r^{-1} K \diamond (f - g)\|^2). \end{aligned}$$

Then,

$$\begin{aligned} A_K &= n^3 \int_{\mathbb{X}} (K \diamond (f - g)(x))^2 (f + g)(x) d\nu_x \\ &\leq n^3 \|K \diamond (f - g)\|^2 \|f + g\|_\infty, \end{aligned}$$

and $B_K = n^2 C_K$, where $C_K = \int_{\mathbb{X}^2} K^2(x, x') (f + g)(x) (f + g)(x') d\nu_x d\nu_{x'}$. From Proposition 2, we deduce that if

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

then

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

By using the elementary inequality $2ab \leq a^2 + b^2$ with $a = \|K \diamond (f - g)\|/r$ and $b = 2\sqrt{2}\sqrt{\|f + g\|_\infty/(n\beta)}$ in the right hand side of the previous inequality, we obtain that if

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

then

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

Since this holds for all $r > 0$, one can take the infimum. This exactly leads to the result in the third case, that is in the case of a kernel function based on a Mercer kernel. Let us now take $r = 1$ and control C_K in the two other cases.

The case of a kernel function based on an orthonormal family. We consider an orthonormal basis $\{\varphi_\lambda, \lambda \in \Lambda\}$ of a subspace S of $\mathbb{L}^2(\mathbb{X}, d\nu)$ and

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

In this case,

$$K \diamond (f - g) = \sum_{\lambda \in \Lambda} \left(\int_{\mathbb{X}} \varphi_\lambda(x)(f - g)(x) d\nu_x \right) \varphi_\lambda = \Pi_S(f - g).$$

Moreover, when the dimension of S is finite, equal to D ,

$$\begin{aligned} C_K &\leq \|f + g\|_\infty^2 \int_{\mathbb{X}} \left(\sum_{\lambda \in \Lambda} \varphi_\lambda(x)\varphi_\lambda(x') \right)^2 d\nu_x d\nu_{x'} \\ &\leq \|f + g\|_\infty^2 D. \end{aligned}$$

When the dimension of S is infinite,

$$\begin{aligned}
C_K &= \int_{\mathbb{X}^2} \left(\sum_{\lambda \in \Lambda} \varphi_\lambda(x) \varphi_\lambda(x') \right)^2 (f+g)(x)(f+g)(x') d\nu_x d\nu_{x'} \\
&\leq \|f+g\|_\infty \int_{\mathbb{X}^2} \left(\sum_{\lambda \in \Lambda} \varphi_\lambda(x) \varphi_\lambda(x') \right)^2 (f+g)(x') d\nu_x d\nu_{x'} \\
&\leq \|f+g\|_\infty \int_{\mathbb{X}^2} \left(\sum_{\lambda, \lambda' \in \Lambda} \varphi_\lambda(x) \varphi_\lambda(x') \varphi_{\lambda'}(x) \varphi_{\lambda'}(x') \right) (f+g)(x') d\nu_x d\nu_{x'} \\
&\leq \|f+g\|_\infty \sum_{\lambda, \lambda' \in \Lambda} \int_{\mathbb{X}} (\varphi_\lambda(x) \varphi_{\lambda'}(x) d\nu_x) \int_{\mathbb{X}} \varphi_\lambda(x') \varphi_{\lambda'}(x') (f+g)(x') d\nu_{x'},
\end{aligned}$$

where we have used the assumption

$$\int_{\mathbb{X}^2} \left(\sum_{\lambda \in \Lambda} |\varphi_\lambda(x) \varphi_\lambda(x')| \right)^2 (f+g)(x') d\nu_x d\nu_{x'} < +\infty$$

to invert the sum and the integral. Hence we have, by orthogonality, and since $\sum_{\lambda \in \Lambda} \varphi_\lambda^2(x) \leq D$,

$$\begin{aligned}
C_K &\leq \|f+g\|_\infty \sum_{\lambda \in \Lambda} \int_{\mathbb{X}} \varphi_\lambda^2(x') (f+g)(x') d\nu_{x'} \\
&\leq D \|f+g\|_\infty \|f+g\|_1.
\end{aligned}$$

Hence we have

$$C_K \leq M^2(f, g) D.$$

This concludes the proof in the first case.

The case of a kernel based on an approximation kernel. Assume now that $\mathbb{X} = \mathbb{R}^d$ and introduce an approximation kernel such that $\int k^2(x) d\nu_x < +\infty$ and $k(-x) = k(x)$, $h = (h_1, \dots, h_d)$, with $h_i > 0$ for every i , and $K(x, x') = k_h(x - x')$, with $k_h(x_1, \dots, x_d) = \frac{1}{\prod_{i=1}^d h_i} k\left(\frac{x_1}{h_1}, \dots, \frac{x_d}{h_d}\right)$. In this case,

$$K \diamond (f - g) = k_h * (f - g),$$

and

$$\begin{aligned}
C_K &= \int_{\mathbb{X}} k_h^2(x-x')(f+g)(x)(f+g)(x')d\nu_x d\nu_{x'} \\
&\leq \|f+g\|_\infty \int_{\mathbb{X}} k_h^2(x-x')(f+g)(x)d\nu_x d\nu_{x'}, \\
&\leq \frac{\|f+g\|_\infty \|f+g\|_1 \|k\|^2}{\prod_{i=1}^d h_i}.
\end{aligned}$$

This ends the proof in the second case.

6.4. *Proof of Proposition 3.* Assume that (H_0) " $f = g$ " is satisfied. Conditionally to N , \hat{T}_K and $\hat{q}_{K,1-\alpha}^{(N)}$ are independent. This leads lo

$$\begin{aligned}
\mathbb{P}\left(\hat{\Phi}_{K,\alpha} = 1 \mid N\right) &= \mathbb{E}\left[\mathbf{1}_{\hat{T}_K > \hat{q}_{K,1-\alpha}^{(N)}} \mid N\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}_{\hat{T}_K > \hat{q}_{K,1-\alpha}^{(N)}} \mid N, \varepsilon^1, \dots, \varepsilon^B\right] \mid N\right] \\
&= \mathbb{E}\left[1 - F_K(\hat{q}_{K,1-\alpha}^{(N)}) \mid N\right]
\end{aligned}$$

where F_K denotes the distribution function of \hat{T}_K under the null hypothesis, given N . We introduce the event

$$(6.5) \quad \Omega_\lambda = \left\{ \sup_{t \in \mathbb{R}} |F_{K,B}(t) - F_K(t)| \leq \frac{\lambda}{\sqrt{B}} \right\}.$$

We deduce from the Dvoretzky-Kiefer-Wolfowitz inequality (see [43]) that

$$\forall \lambda > 0, \mathbb{P}(\Omega_\lambda^C \mid N) \leq 2 \exp(-2\lambda^2).$$

Hence,

$$\begin{aligned}
\mathbb{P}\left(\hat{\Phi}_{K,\alpha} = 1 \mid N\right) &= \mathbb{E}\left[\left(1 - F_K(\hat{q}_{K,1-\alpha}^{(N)})\right) \mathbf{1}_{\Omega_\lambda} + \left(1 - F_K(\hat{q}_{K,1-\alpha}^{(N)})\right) \mathbf{1}_{\Omega_\lambda^C} \mid N\right] \\
&\leq \mathbb{E}\left[\left(1 - F_{K,B}(\hat{q}_{K,1-\alpha}^{(N)}) + \frac{\lambda}{\sqrt{B}}\right) \mathbf{1}_{\Omega_\lambda}\right] + \mathbb{P}(\Omega_\lambda^C \mid N).
\end{aligned}$$

By definition of $\hat{q}_{K,1-\alpha}^{(N)}$, we have $1 - F_{K,B}(\hat{q}_{K,1-\alpha}^{(N)}) \leq \alpha$, hence, we obtain that for all $\lambda > 0$,

$$\mathbb{P}\left(\hat{\Phi}_{K,\alpha} = 1 \mid N\right) \leq \alpha + \frac{\lambda}{\sqrt{B}} + 2e^{-2\lambda^2}.$$

This concludes the proof of Proposition 3.

6.5. *Proof of Proposition 4.* We set $\alpha_B = \alpha - \sqrt{\ln B/4B}$.

As in Section 3.2, we introduce the $1 - (\frac{\beta}{2} - \frac{2}{\sqrt{B}})$ quantile of the bootstrapped quantile $q_{K,1-\alpha_B}^{(N)}$ that we denote by $q_{1-(\frac{\beta}{2}-\frac{2}{\sqrt{B}})}^{\alpha_B}$.

We set for a moment $t = q_{1-(\frac{\beta}{2}-\frac{2}{\sqrt{B}})}^{\alpha_B}$. Recall that $\hat{q}_{K,1-\alpha}^{(N)} = F_{K,B}^{-1}(1-\alpha)$, and that $q_{K,1-u}^{(N)}$ denotes the $(1-u)$ quantile of \hat{T}_K^ε conditionnaly to N . We have

$$\mathbb{P}(\hat{q}_{K,1-\alpha}^{(N)} > t) = \mathbb{P}(F_{K,B}(t) < 1-\alpha).$$

Let us introduce the event Ω_λ defined by (6.5), with $\lambda = \sqrt{\ln B}/2$. We then have

$$\begin{aligned} \mathbb{P}(F_{K,B}(t) < 1-\alpha|N) &\leq \mathbb{P}(\{F_{K,B}(t) < 1-\alpha\} \cap \Omega_\lambda|N) + \mathbb{P}(\Omega_\lambda^C|N) \\ &\leq \mathbb{P}(F_K(t) < 1-\alpha + \lambda/\sqrt{B}|N) + 2e^{-2\lambda^2} \\ &\leq \mathbb{P}(q_{K,1-\alpha+\lambda/\sqrt{B}}^{(N)} > t|N) + 2e^{-2\lambda^2} \\ &\leq \mathbb{P}(q_{K,1-\alpha_B}^{(N)} > t|N) + \frac{2}{\sqrt{B}} \end{aligned}$$

by definition of λ . Now, the definition of t implies that

$$\mathbb{P}\left(q_{K,1-\alpha_B}^{(N)} > t\right) \leq \frac{\beta}{2} - \frac{2}{\sqrt{B}},$$

hence

$$\mathbb{P}(\hat{q}_{K,1-\alpha}^{(N)} > t) \leq \beta/2.$$

Let us now control the probability of second kind error of the test $\hat{\Phi}_{K,\alpha}$.

$$\begin{aligned} \mathbb{P}\left(\hat{T}_K \leq \hat{q}_{K,1-\alpha}^{(N)}\right) &\leq \mathbb{P}\left(\{\hat{T}_K \leq \hat{q}_{K,1-\alpha}^{(N)}\} \cap \{\hat{q}_{K,1-\alpha}^{(N)} \leq t\}\right) + \mathbb{P}\left(\hat{q}_{K,1-\alpha}^{(N)} > t\right) \\ &\leq \mathbb{P}\left(\hat{T}_K \leq t\right) + \beta/2. \end{aligned}$$

We finally deduce from (6.1) that if $\mathcal{E}_K > \sqrt{\frac{8A_K+4B_K}{\beta}} + q_{1-(\frac{\beta}{2}-\frac{2}{\sqrt{B}})}^{\alpha_B}$, then

$$\mathbb{P}_{f,g}(\hat{T}_K \leq q_{1-(\frac{\beta}{2}-\frac{2}{\sqrt{B}})}^{\alpha_B}) \leq \beta/2.$$

An upper bound for $q_{1-(\frac{\beta}{2}-\frac{2}{\sqrt{B}})}^{\alpha_B}$ is obtained in Section 6.2. This concludes the proof of Proposition 4.

6.6. *Proof of Theorem 4.1 and Theorem 4.2.* The proof is similar for both theorems.

It is clear from the definition of $u_\alpha^{(N)}$ that the test defined by Φ_α is of level α . Obviously, by Bonferonni's inequality, $u_\alpha^{(N)} \geq \alpha$, hence, setting $\alpha_m = \alpha e^{-w_m}$, we have

$$\begin{aligned} & \mathbb{P}_{f,g} \left(\exists m \in \mathcal{M}, \hat{T}_{K_m} > q_{K_m, 1-e^{-w_m} u_\alpha^{(N)}}^{(N)} \right) \\ & \geq \mathbb{P}_{f,g} \left(\exists m \in \mathcal{M}, \hat{T}_{K_m} > q_{K_m, 1-\alpha_m}^{(N)} \right) \\ & \geq 1 - \mathbb{P}_{f,g} \left(\forall m \in \mathcal{M}, \hat{T}_{K_m} \leq q_{K_m, 1-\alpha_m}^{(N)} \right) \\ & \geq 1 - \inf_{m \in \mathcal{M}} \mathbb{P}_{f,g} \left(\hat{T}_{K_m} \leq q_{K_m, 1-\alpha_m}^{(N)} \right) \\ & \geq 1 - \beta, \end{aligned}$$

as soon as there exists m in \mathcal{M} such that $\mathbb{P}_{f,g} \left(\hat{T}_{K_m} \leq q_{K_m, 1-\alpha_m}^{(N)} \right) \leq \beta$. We can now apply Theorem 3.1, replacing $\ln(2/\alpha)$ by $(\ln(2/\alpha) + w_m)$, to conclude the proof.

6.7. *Proof of Corollary 1.* It is well known (see [21] for instance) that with the notations of Example 1

$$\|(f - g) - \Pi_{S_J}(f - g)\|^2 \leq C(\delta) R^2 2^{-2J\delta},$$

when $(f - g)$ belongs to $\mathcal{B}_{2,\infty}^\delta(R)$. Hence the upper bound of (4.5) with

$$\bar{J} = \left\lceil \log_2 \left(\left(\frac{n}{\ln \ln n} \right)^{\frac{2}{4\delta+1}} \right) \right\rceil$$

leads to

$$\rho(\Phi_\alpha^{(1)}, \mathcal{B}_{\delta,\gamma,\infty}(R, R', R''), \beta) \leq C(\delta, \alpha, \beta, R, R'') \left(\frac{n}{\ln \ln n} \right)^{-\frac{2\delta}{4\delta+1}}.$$

Of course a similar upper bound applies to Ψ_α .

Let J be an integer that will be chosen later. As in [21], for any $m \subset \Lambda_J = \{(j, k), j \in \{0, \dots, J-1\}, k \in \{0, \dots, 2^j - 1\}\}$, one can write

$$\|(f - g) - \Pi_{S_m}(f - g)\|^2 = \|(f - g) - \Pi_{S_J}(f - g)\|^2 + \|\Pi_{S_m}(f - g) - \Pi_{S_J}(f - g)\|^2.$$

But for m such that $\{\alpha_\lambda, \lambda \in m\}$ is the set of the D largest coefficients among $\{\alpha_\lambda, \lambda \in \{0\} \cup \Lambda_J\}$,

$$\|\Pi_{S_m}(f - g) - \Pi_{S_J}(f - g)\|^2 \leq C(\gamma) R^{2+4\gamma} D^{-2\gamma},$$

see for instance [21] p.36. Taking

$$J = \lfloor \log_2 n^\epsilon \rfloor + 1$$

for some $\epsilon > 0$, the right hand side of (4.6) is upper bounded by

$$C(\delta, \gamma, \alpha, \beta, R, R', R'') \left[n^{-2\epsilon\delta} + D^{-2\gamma} + \frac{\epsilon D \ln n}{n} \right].$$

Taking $D = \lfloor (n/\ln n)^{1/(2\gamma+1)} \rfloor$, and $\epsilon > \gamma/(\delta(2\gamma+1))$, one obtains that when $\delta < \gamma/2$, $D \leq 2^J$, and

$$\rho(\Phi_\alpha^{(2)}, \mathcal{B}_{\delta, \gamma, \infty}(R, R', R''), \beta) \leq C(\delta, \gamma, \alpha, \beta, R, R', R'') \left(\frac{n}{\ln n} \right)^{-\frac{\gamma}{1+2\gamma}}.$$

Since this upper bound also applies to Ψ_α , one has

$$\begin{aligned} & \rho(\Psi_\alpha, \mathcal{B}_{\delta, \gamma, \infty}(R, R', R''), \beta) \\ & \leq C(\delta, \gamma, \alpha, \beta, R, R', R'') \inf \left\{ \left(\frac{n}{\ln \ln n} \right)^{-\frac{2\delta}{4\delta+1}}, \left(\frac{n}{\ln n} \right)^{-\frac{\gamma}{1+2\gamma}} \right\}. \end{aligned}$$

6.8. *Proof of Proposition 5.* As usual, we introduce a finite subset \mathcal{C} of $\mathcal{B}_{\delta, \gamma, \infty}(R, R', R'')$, composed of couples of intensities which are particularly difficult to distinguish. Here one can use the finite subset of possible intensities $\mathcal{S}_{M, D, r}$ that has been defined in [21] Equation (6.4), and define

$$\mathcal{C} = \{(f, g), f = \rho \mathbf{1}_{[0,1]} \text{ and } g \in \mathcal{S}_{M, D, r}\},$$

for some fixed positive ρ . Next the computations of the lower bounds of [21] can be completely reproduced once we remark that the likelihood ratio

$$\frac{d\mathbb{P}_{\rho \mathbf{1}_{[0,1]}, g}}{d\mathbb{P}_{\rho \mathbf{1}_{[0,1]}, \rho \mathbf{1}_{[0,1]}}}(N^1, N^{-1}) = \frac{d\mathbb{P}_{\rho \mathbf{1}_{[0,1]}}}{d\mathbb{P}_{\rho \mathbf{1}_{[0,1]}}}(N^1) \times \frac{d\mathbb{P}_g}{d\mathbb{P}_{\rho \mathbf{1}_{[0,1]}}}(N^{-1}),$$

where on the left hand side $\mathbb{P}_{f, g}$ represents the joint distribution of two independent Poisson processes N^1 and N^{-1} , with respective intensities f and g , and on the right hand side \mathbb{P}_f represents the distribution of one Poisson process N with intensity f . This means that the likelihood ratios that have been considered in [21] are exactly the ones we need here to compute the lower bound. The results are consequently identical.

6.9. *Proof of Corollary 2.* Considering Theorem 4.2 applied to *Example 4*, we mainly have to find a sharp upper bound for

$$\inf_{(m_1, m_2) \in \mathcal{M}} \left\{ \|(f - g) - k_{m_1, h_{m_2}} * (f - g)\|^2 + \frac{w_{(m_1, m_2)}}{n} \sqrt{\frac{\|f + g\|_\infty \|f + g\|_1 \|k_{m_1}\|^2}{2^{-dm_2}}} \right\},$$

when (f, g) belongs to $\mathcal{S}_d^\delta(R, R', R'')$.

Let us first control the bias term $\|(f - g) - k_{m_1, h_{m_2}} * (f - g)\|^2$. Plancherel's theorem gives that when $(f - g) \in \mathbb{L}^1(\mathbb{R}^d) \cap \mathbb{L}^2(\mathbb{R}^d)$,

$$\begin{aligned} (2\pi)^d \|(f - g) - k_{m_1, h_{m_2}} * (f - g)\|^2 &= \|\widehat{(f - g)} - \widehat{k_{m_1, h_{m_2}}} * \widehat{(f - g)}\|^2 \\ &= \|\widehat{(f - g)} - \widehat{k_{m_1, h_{m_2}}} \widehat{(f - g)}\|^2 \\ &= \|(1 - \widehat{k_{m_1, h_{m_2}}}) \widehat{(f - g)}\|^2 \\ &= \int_{\mathbb{R}^d} \left(1 - \widehat{k_{m_1, h_{m_2}}}\right)^2 (u) (\widehat{(f - g)})^2(u) d\nu_u \\ &= \int_{\mathbb{R}^d} \left|1 - \widehat{k_{m_1}}(2^{-m_2} u)\right|^2 (u) (\widehat{(f - g)})^2(u) d\nu_u. \end{aligned}$$

Assume now that $(f, g) \in \mathcal{S}_d^\delta(R, R', R'')$, and take $m_1^* = \min\{m_1 \in \mathcal{M}_1, m_1 \geq \delta\}$. Note that since $\|\widehat{k_{m_1^*}}\|_\infty < +\infty$ and $\widehat{k_{m_1^*}}$ satisfies the condition (4.7), there also exists some constant $C(\delta) > 0$ such that

$$\text{Ess sup}_{u \in \mathbb{R}^d \setminus \{0\}} \frac{|1 - \widehat{k_{m_1^*}}(u)|}{\|u\|_d^\delta} \leq C(\delta).$$

Then

$$\|(f - g) - k_{m_1^*, h_{m_2}} * (f - g)\|^2 \leq \frac{C(\delta)}{(2\pi)^d} \int_{\mathbb{R}^d} \|2^{-m_2} u\|_d^{2\delta} (\widehat{(f - g)})^2(u) d\nu_u,$$

and since $(f - g) \in \mathcal{S}_d^\delta(R)$,

$$\|(f - g) - k_{m_1^*, h_{m_2}} * (f - g)\|^2 \leq 2^{-2\delta m_2} C(\delta) R^2.$$

Furthermore, $\|k_{m_1^*}\|^2 \leq C(\delta)$, so

$$\begin{aligned} & \inf_{(m_1, m_2) \in \mathcal{M}} \left\{ \|(f - g) - k_{m_1, h_{m_2}} * (f - g)\|^2 \right. \\ & \quad \left. + \frac{w(m_1, m_2)}{n} \sqrt{\frac{\|f + g\|_\infty \|f + g\|_1 \|k_{m_1}\|^2}{2^{-dm_2}}} \right\} \\ & \leq C(\delta, \alpha, \beta, R) \inf_{m_2 \in \mathcal{M}_2} \left\{ 2^{-2\delta m_2} + \frac{w(m_1^*, m_2)}{n} \sqrt{\frac{\|f + g\|_\infty \|f + g\|_1}{2^{-dm_2}}} \right\}. \end{aligned}$$

Choosing

$$m_2^* = \left\lfloor \log_2 \left(\left(\frac{n}{\ln \ln n} \right)^{\frac{2}{d+4\delta}} \right) \right\rfloor$$

leads to

$$2^{-2\delta m_2^*} \leq 2^{2\delta} \left(\frac{\ln \ln n}{n} \right)^{\frac{4\delta}{d+4\delta}},$$

and since $w(m_1^*, m_2^*) \leq C(\delta, d) \ln \ln n$,

$$\frac{w(m_1^*, m_2^*)}{n} \sqrt{\frac{\|f + g\|_\infty \|f + g\|_1}{2^{-dm_2^*}}} \leq C(\delta, d) \sqrt{\|f + g\|_\infty \|f + g\|_1} \left(\frac{\ln \ln n}{n} \right)^{\frac{4\delta}{d+4\delta}}.$$

Noticing that

$$\frac{1}{n} \leq \left(\frac{\ln \ln n}{n} \right)^{\frac{4\delta}{d+4\delta}},$$

when $(f, g) \in \mathcal{S}_d^\delta(R, R', R'')$,

$$\begin{aligned} & C(\alpha, \beta) \left(\inf_{(m_1, m_2) \in \mathcal{M}} \left\{ \|(f - g) - k_{m_1, h_{m_2}} * (f - g)\|^2 \right. \right. \\ & \quad \left. \left. + \frac{w(m_1, m_2)}{n} \sqrt{\frac{\|f + g\|_\infty \|f + g\|_1 \|k_{m_1}\|^2}{2^{-dm_2}}} \right\} + \frac{\|f + g\|_\infty}{n} \right) \\ & \leq C(\delta, \alpha, \beta, R, R', R'', d) \left(\frac{\ln \ln n}{n} \right)^{\frac{4\delta}{d+4\delta}}. \end{aligned}$$

This concludes the proof of Corollary 2.

6.10. *Proof of Corollary 3.* As in the previous section, considering Theorem 4.2 applied to *Example 5*, we here have to find a sharp upper bound for

$$\inf_{m_2 \in \mathcal{M}_2} \left\{ \|(f - g) - k_{1, h_{m_2}} * (f - g)\|^2 + \frac{w(1, m_2)}{n} \sqrt{\frac{\|f + g\|_\infty \|f + g\|_1 \|k_1\|^2}{\prod_{i=1}^d h_{m_2, i}}} \right\}.$$

Let us first evaluate $\|(f - g) - k_{1, h} * (f - g)\|^2$ when $(f - g) \in \mathcal{N}_{2, d}^\delta(\mathbb{R})$, and $h = (h_1, \dots, h_d)$.

For $x = (x_1, \dots, x_d) \in \mathbb{R}^d$, let $b(x) = k_{1, h} * (f - g)(x) - (f - g)(x)$. Then

$$b(x) = \int_{\mathbb{R}^d} k_1(u_1, \dots, u_d) (f - g)(x_1 + u_1 h_1, \dots, x_d + u_d h_d) du_1 \dots du_d - (f - g)(x),$$

and since $\int_{\mathbb{R}^d} k_1(u_1, \dots, u_d) du_1 \dots du_d = 1$,

$$\begin{aligned} b(x) &= \int_{\mathbb{R}^d} k_1(u_1, \dots, u_d) \left[(f - g)(x_1 + u_1 h_1, \dots, x_d + u_d h_d) \right. \\ &\quad \left. - (f - g)(x_1, \dots, x_d) \right] du_1 \dots du_d \\ &= \sum_{i=1}^d b_i(x), \end{aligned}$$

where for $i = 1 \dots d$,

$$\begin{aligned} b_i(x) &= \int_{\mathbb{R}^d} k_1(u_1, \dots, u_d) \left[(f - g)(x_1 + u_1 h_1, \dots, x_i + u_i h_i, x_{i+1}, \dots, x_d) \right. \\ &\quad \left. - (f - g)(x_1 + u_1 h_1, \dots, x_i, x_{i+1}, \dots, x_d) \right] du_1 \dots du_d, \end{aligned}$$

As in the proof of Proposition 1.5 p. 13 of [58], using the Taylor expansion of $(f - g)$ in the i th direction and the fact that $\int_{\mathbb{R}} k_{1, i}(u_i) u_i^j du_i = 0$ for $j = 1 \dots [\delta_i]$, we obtain that

$$\begin{aligned}
b_i(x) &= \int_{\mathbb{R}^d} k_1(u) \frac{(u_i h_i)^{[\delta_i]}}{([\delta_i]-1)!} \left[\int_0^1 (1-\tau)^{[\delta_i]-1} \right. \\
&\quad \left. D_i^{[\delta_i]}(f-g)(x_1 + u_1 h_1, \dots, x_i + \tau u_i h_i, x_{i+1}, \dots, x_d) d\tau \right] du \\
&= \int_{\mathbb{R}^d} k_1(u) \frac{(u_i h_i)^{[\delta_i]}}{([\delta_i]-1)!} \left[\int_0^1 (1-\tau)^{[\delta_i]-1} \right. \\
&\quad \left(D_i^{[\delta_i]}(f-g)(x_1 + u_1 h_1, \dots, x_i + \tau u_i h_i, x_{i+1}, \dots, x_d) \right. \\
&\quad \left. - D_i^{[\delta_i]}(f-g)(x_1 + u_1 h_1, \dots, x_i, \dots, x_d) \right) d\tau \Big] du.
\end{aligned}$$

Hence, by using twice Lemma 1.1 p. 13 of [58] extended to the spaces $\mathbb{R}^d \times \mathbb{R}$ and $\mathbb{R}^d \times \mathbb{R}^d$,

$$\begin{aligned}
\|b_i\|_2^2 &\leq \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} |k_1(u)| \frac{|u_i h_i|^{[\delta_i]}}{([\delta_i]-1)!} \left[\int_0^1 (1-\tau)^{[\delta_i]-1} \right. \right. \\
&\quad \left. \left| D_i^{[\delta_i]}(f-g)(x_1 + u_1 h_1, \dots, x_i + \tau u_i h_i, x_{i+1}, \dots, x_d) \right. \right. \\
&\quad \left. \left. - D_i^{[\delta_i]}(f-g)(x_1 + u_1 h_1, \dots, x_i, \dots, x_d) \right| d\tau \right] du \Big)^2 dx \\
&\leq \left[\int_{\mathbb{R}^d} |k_1(u)| \frac{|u_i h_i|^{[\delta_i]}}{([\delta_i]-1)!} \left(\int_{\mathbb{R}^d} \left[\int_0^1 (1-\tau)^{[\delta_i]-1} \right. \right. \right. \\
&\quad \left. \left. \left| D_i^{[\delta_i]}(f-g)(x_1 + u_1 h_1, \dots, x_i + \tau u_i h_i, x_{i+1}, \dots, x_d) \right. \right. \right. \\
&\quad \left. \left. \left. - D_i^{[\delta_i]}(f-g)(x_1 + u_1 h_1, \dots, x_i, \dots, x_d) \right| d\tau \right]^2 dx \right)^{1/2} du \Big]^2
\end{aligned}$$

and

$$\begin{aligned}
\|b_i\|_2^2 &\leq \left[\int_{\mathbb{R}^d} |k_1(u)| \frac{|u_i h_i|^{[\delta_i]}}{([\delta_i]-1)!} \left(\int_0^1 (1-\tau)^{[\delta_i]-1} \left(\int_{\mathbb{R}^d} \right. \right. \right. \\
&\quad \left. \left. \left| D_i^{[\delta_i]}(f-g)(x_1 + u_1 h_1, \dots, x_i + \tau u_i h_i, x_{i+1}, \dots, x_d) \right. \right. \right. \\
&\quad \left. \left. \left. - D_i^{[\delta_i]}(f-g)(x_1 + u_1 h_1, \dots, x_i, \dots, x_d) \right|^2 dx \right)^{1/2} d\tau \right) du \Big]^2.
\end{aligned}$$

Since $(f - g) \in \mathcal{N}_{2,d}^\delta(R)$,

$$\|b_i\|_2 \leq C(\delta_i)R \int_{\mathbb{R}^d} |k_1(u)| |u_i h_i|^{\delta_i} du \leq C(\delta_i) \left(\int_{\mathbb{R}^d} |k_1(u)| |u_i|^{\delta_i} du \right) R h_i^{\delta_i}.$$

So,

$$\|k_{1,h} * (f - g) - (f - g)\| \leq C(\delta)R \sum_{i=1}^d h_i^{\delta_i}.$$

Let us now find some m_2 in \mathcal{M}_2 giving a sharp upper bound for

$$\inf_{m_2 \in \mathcal{M}_2} \left\{ \|(f - g) - k_{1,h_{m_2}} * (f - g)\|^2 + \frac{w_{(1,m_2)}}{n} \sqrt{\frac{\|f + g\|_\infty \|f + g\|_1 \|k_1\|^2}{\prod_{i=1}^d h_{m_2,i}}} \right\}.$$

Let $1/\bar{\delta} = \sum_{i=1}^d 1/\delta_i$, and choose $m_2^* = (m_{2,1}^*, \dots, m_{2,d}^*)$ in \mathcal{M}_2 , with

$$m_{2,i}^* = \left\lfloor \log_2 \left(\left(\frac{n}{\ln \ln n} \right)^{\frac{2\bar{\delta}}{\delta_i(1+4\bar{\delta})}} \right) \right\rfloor,$$

for every $i = 1 \dots d$. Since $h_{m_2^*} = (2^{-m_{2,1}^*}, \dots, 2^{-m_{2,d}^*})$,

$$\|(f - g) - k_{1,h_{m_2^*}} * (f - g)\| \leq C(\delta, R) \sum_{i=1}^d 2^{-m_{2,i}^* \delta_i},$$

so

$$\|(f - g) - k_{1,h_{m_2^*}} * (f - g)\|^2 \leq C(\delta, R) d^2 \left(\frac{\ln \ln n}{n} \right)^{\frac{4\bar{\delta}}{1+4\bar{\delta}}}.$$

Moreover, it is easy to see that $w_{(1,m_2^*)} \leq C(\delta, d) \ln \ln n$, and hence

$$\begin{aligned} & \frac{w_{(1,m_2^*)}}{n} \sqrt{\frac{\|f + g\|_\infty \|f + g\|_1 \|k_1\|^2}{\prod_{i=1}^d 2^{-m_{2,i}^*}}} \\ & \leq C(\delta, \alpha, \beta, R', R'', d) \frac{\ln \ln n}{n} \left(\frac{n}{\ln \ln n} \right)^{\sum_{i=1}^d \frac{\bar{\delta}}{(1+4\bar{\delta})\delta_i}} \\ & \leq C(\delta, \alpha, \beta, R', R'', d) \left(\frac{\ln \ln n}{n} \right)^{\frac{4\bar{\delta}}{1+4\bar{\delta}}}. \end{aligned}$$

Noticing that when $\ln \ln n \geq 1$,

$$\frac{1}{n} \leq \left(\frac{\ln \ln n}{n} \right)^{\frac{4\bar{\delta}}{1+4\bar{\delta}}},$$

this ends the proof of Corollary 3.

Acknowledgment. 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.

- [1] ARCONES, M. A. AND GINÉ, E. (1992). *On the bootstrap of U and V statistics*, *Ann. Statist.* **20**, 2, 655–674. [MR1165586](#)
- [2] ARLOT, S. (2009). *Model selection by resampling penalization*, *Electron. J. Stat.* **3**, 557–624. [MR2519533](#)
- [3] ATHREYA, K. B. (1987). *Bootstrap of the mean in the infinite variance case*, *Ann. Statist.* **15**, 2, 724–731. [MR888436](#)
- [4] BARAUD, Y. (2002). *Non-asymptotic minimax rates of testing in signal detection*, *Bernoulli* **8**, 5, 577–606. [MR1935648](#)
- [5] BARAUD, Y. AND HUET, S. AND LAURENT, B. (2003). *Adaptive tests of linear hypotheses by model selection*, *Ann. Statist.* **31**, 1, 225–251. [MR1962505](#)
- [6] BICKEL, P. J. (1968). *A distribution free version of the Smirnov two sample test in the p -variate case*, *Ann. Math. Statist.* **40**, 1–23. [MR0256519](#)
- [7] BICKEL, P. J. AND FREEDMAN, D. A. (1981). *Some asymptotic theory for the bootstrap*, *Ann. Statist.* **9**, 6, 1196–1217. [MR630103](#)
- [8] BOVETT, J. M. AND SAW, J. G. (1980). *On comparing two Poisson intensity functions*, *Comm. Statist. A—Theory Methods* **9**, 9, 943–948. [MR574314](#)
- [9] BRETAGNOLLE, J. (1983). *Lois limites du bootstrap de certaines fonctionnelles*, *Ann. Inst. H. Poincaré Sect. B* **19**, 3, 281–296. [MR725561](#)
- [10] BUTUCEA, C. AND TRIBOULEY, K. (2006). *Nonparametric homogeneity tests*, *J. Statist. Plann. Inference* **136**, 3, 597–639. [MR2181971](#)
- [11] CHIU, S. N. (2010). *Parametric bootstrap and approximate tests for two Poisson variates*, *J. Stat. Comput. Simul.* **80**, 3–4, 263–271. [MR2752162](#)
- [12] CHIU, S. N. AND WANG, L. (2009). *Homogeneity tests for several Poisson populations*, *Comput. Statist. Data Anal.* **53**, 12, 4266–4278. [MR2744323](#)
- [13] COX, D. R. (1953). *Some simple approximate tests for Poisson variates*, *Biometrika* **40**, 354–360. [MR0058177](#)
- [14] DALEY, D. J. AND VERE-JONES, D. (2008). *An introduction to the theory of point processes. Vol. II*, Second Edition, Springer, New York. [MR2371524](#)
- [15] DE LA PEÑA, V. H. AND GINÉ, E. (1999). *From dependence to independence, Randomly stopped processes. U -statistics and processes. Martingales and beyond*, Springer, New York. [MR1666908](#)
- [16] DEHLING, H. AND MIKOSCH, T. (1994). *Random quadratic forms and the bootstrap for U -statistics*, *J. Multivariate Anal.* **51**, 2, 392–413. [MR1321305](#)
- [17] DESHPANDE, J. V. AND SENGUPTA, D. (1995). *Testing the hypothesis of proportional hazards in two populations*, *Biometrika* **82**, 252–261.
- [18] DESHPANDE, J. V. AND MUKHOPADHYAY, M. AND NAIK-NIMBALKAR, U. V. (1999). *Testing of two sample proportional intensity assumption for non-homogeneous Poisson processes*, *J. Statist. Plann. Inference* **81**, 2, 237–251. [MR62M07](#)
- [19] EFRON, B. (1979). *Bootstrap methods: another look at the jackknife*, *Ann. Statist.* **7**, 1, 1–26. [MR0515681](#)

- [20] EFRON, B. AND TIBSHIRANI, R. J. (1993). *An introduction to the bootstrap*, Monographs on Statistics and Applied Probability, Chapman and Hall, New York. [MR1270903](#)
- [21] FROMONT, M. AND LAURENT, B. AND REYNAUD-BOURET, P. (2011). *Adaptive tests of homogeneity for a Poisson process*, *Ann. Inst. Henri Poincaré Probab. Stat.* **47**, 1, 176–213. [MR2779402](#)
- [22] FUKUMIZU, K. AND SRIPERUMBUDUR, B. AND GRETTON, A. AND SCHÖLKOPF, B. (2009). *Characteristic kernels on groups and semigroups*, *Proceedings of the NIPS 2008 conference*, 473–480.
- [23] GAIL, M. (1974). *Computations for designing comparative Poisson trials*, *Biometrics* **30**, 2, 231–237.
- [24] GINÉ, E. AND ZINN, J. (1989). *Necessary conditions for the bootstrap of the mean*, *Ann. Statist.* **17**, 2, 684–691. [MR994259](#)
- [25] GINÉ, E. AND ZINN, J. (1990). *Bootstrapping general empirical measures*, *Ann. Probab.* **18**, 2, 851–869. [MR1055437](#)
- [26] GOLDENSHLUGER, A. AND LEPSKI, O. (2011). *Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality*, *Ann. Statist.* **29**, 3, 1608–1632. [MR2850214](#)
- [27] GRETTON, A. AND BORGWARDT, K. M. AND RASCH, M. J. AND SCHÖLKOPF, B. AND SMOLA, A. (2008). *A kernel method for the two-sample problem*, *J. Mach. Learn. Res.* **1**, 1–10.
- [28] HOROWITZ, J. L. AND SPOKOINY, V. G. (2001). *An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative*, *Econometrica* **69**, 3, 599–631. [MR1828537](#)
- [29] HUŠKOVÁ, M. AND JANSSEN, P. (1993). *Consistency of the generalized bootstrap for degenerate U -statistics*, *Ann. Statist.* **21**, 4, 1811–1823. [MR1245770](#)
- [30] INGSTER, Y. I. (1990). *Adaptive tests for minimax testing of nonparametric hypotheses*, *Probability theory and mathematical statistics, Vol. I (Vilnius, 1989)*, Mokslas, Vilnius, 539–549. [MR1153844](#)
- [31] INGSTER, Y. I. (1993). *Asymptotically minimax testing for nonparametric alternatives I-II-III*, *Math. Methods Statist.* **2**, 85–114, 171–189, 249–268. [MR1259685](#)
- [32] INGSTER, Y. I. (1997). *Adaptive chi-square tests*, *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* **244**, Veroyatn. i Stat. 2, 150–166, 333. [MR1700386](#)
- [33] INGSTER, Y. I. AND KUTOYANTS, Y. A. (2007). *Nonparametric hypothesis testing for intensity of the Poisson process*, *Math. Methods Statist.* **16**, 3, 217–245. [MR2356819](#)
- [34] INGSTER, Y. I. AND STEPANOVA, N. (2011). *Estimation and detection of functions from anisotropic Sobolev classes*, *Electron. J. Statist.* **5**, 484–506. [MR2813552](#)
- [35] JANSSEN, A. (1994). *Two-sample goodness-of-fit tests when ties are present*, *J. Statist. Plann. Inference* **39**, 3, 399–424. [MR1278591](#)
- [36] JANSSEN, A. AND PAULS, T. (2003). *How do bootstrap and permutation tests work?*, *Ann. Statist.* **31**, 3, 768–806. [MR1994730](#)
- [37] KINGMAN, J. F. C. (1993). *Poisson processes*, *Oxford Studies in Probability* **3**, The Clarendon Press Oxford University Press, New-York. [MR1207584](#)
- [38] KLAASSEN, C. A. J. AND WELLNER, J. A. (1992). *Kac empirical processes and the bootstrap*, *Probability in Banach spaces, 8 (Brunswick, ME, 1991)* **30**, 411–429, Birkhäuser Boston.
- [39] KRISHNAMOORTHY, K. AND THOMSON, J. (2004). *A more powerful test for comparing two Poisson means*, *J. Statist. Plann. Inference* **119**, 23–35. [MR2018448](#)
- [40] LATAŁA, R. (1999). *Tail and moment estimates for some types of chaos*, *Studia Math.* **135**, 1, 39–53. [MR1686370](#)

- [41] LO, A.Y. (1987). *A large sample study of the Bayesian bootstrap*, *Ann. Statist.* **15**, 1, 360–375. [MR885742](#)
- [42] MAMMEN, E. (1992). *Bootstrap, wild bootstrap, and asymptotic normality*, *Probab. Theory Related Fields* **93**, 4, 439–455. [MR1183886](#)
- [43] MASSART, P. (1990). *The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality*, *Ann. Probab.* **18**, 3, 1269–1283. [MR1062069](#)
- [44] NG, H. K. T. AND GU, K. AND TANG, M. L. (2007). *A comparative study of tests for the difference of two Poisson means*, *Comput. Statist. Data Anal.* **51**, 6, 3085–3099. [MR2345627](#)
- [45] PRAESTGAARD, J. T. (1995). *Permutation and bootstrap Kolmogorov-Smirnov tests for the equality of two distributions*, *Scand. J. Statist.* **22**, 3, 305–322. [MR1363215](#)
- [46] PRAESTGAARD, J. T. AND WELLNER, J. A. (1993). *Exchangeably weighted bootstraps of the general empirical process*, *Ann. Probab.* **21**, 4, 2053–2086. [MR1245301](#)
- [47] PRZYBOROWSKI, J. AND WILENSKI, H. (1940). *Homogeneity of results in testing samples from Poisson series with an application to testing clover seed for dodder*, *Biometrika* **31**, 313–323. [MR0002070](#)
- [48] QUENOUILLE, M. H. (1949). *Approximate tests of correlation in time-series*, *Proc. Cambridge Philos. Soc.* **45** .
- [49] RIGOLLET, P. AND TSYBAKOV, A. B. (2007). *Linear and convex aggregation of density estimators*, *Math. Methods Statist.* **16**, 3, 260–280. [MR2356821](#)
- [50] ROMANO, J. P. (1988). *A bootstrap revival of some nonparametric distance tests*, *J. Amer. Statist. Assoc.* **83**, 403, 698–708. [MR963797](#)
- [51] ROMANO, J. P. (1988). *Bootstrap and randomization tests of some nonparametric hypotheses*, *Ann. Statist.* **17**, 1, 141–159. [MR981441](#)
- [52] RUBIN, D. B. (1981). *The Bayesian bootstrap*, *Ann. Statist.* **9**, 1, 130–134. [MR600538](#)
- [53] SCHÖLKOPF, B., AND SMOLA, A. J. (2002). *Learning with kernels*, Cambridge (Mass.), MIT Press.
- [54] SHIUE, W. AND BAIN, L. J. (1982). *Experiment size and power comparisons for two-sample Poisson tests*, *J. Roy. Statist. Soc. Ser. C* **31**, 2, 130–134.
- [55] SINGH, K. (1981). *On the asymptotic accuracy of Efron’s bootstrap*, *Ann. Statist.* **9**, 6, 1187–1195. [MR630102](#)
- [56] SPOKOINY, V. G. (1996). *Adaptive hypothesis testing using wavelets*, *Ann. Statist.* **24**, 6, 2477–2498. [MR1425962](#)
- [57] SPOKOINY, V. G. (1998). *Adaptive and spatially adaptive testing of a nonparametric hypothesis*, *Math. Methods Statist.* **7**, 3, 245–273. [MR1651777](#)
- [58] TSYBAKOV, A. B. (2009). *Introduction to nonparametric estimation*, *Springer Series in Statistics* Springer, New York. [MR2724359](#)
- [59] TUKEY, J. W. (1958). *Bias and confidence in not quite large samples*, *Annals of Math. Statist.* **29**, 614.
- [60] WELLNER, J. A. (1979). *Permutation tests for directional data*, *Ann. Statist.* **7**, 5, 929–943. [MR536498](#)
- [61] WENG, C.-S. (1989). *On a second-order asymptotic property of the Bayesian bootstrap mean*, *Ann. Statist.* **17**, 2, 705–710. [MR994261](#)

MAGALIE FROMONT
 ENSAI - CAMPUS DE KER-LANN
 RUE BLAISE PASCAL - BP 37203
 35172 BRUZ CEDEX (FRANCE)
 E-MAIL: mfromont@ensai.fr

BÉATRICE LAURENT
 DÉPARTEMENT GMM -INSA
 135 AV. DE RANGUEIL
 31077 TOULOUSE CEDEX 4 (FRANCE)
 E-MAIL: Beatrice.Laurent@insa-toulouse.fr

PATRICIA REYNAUD-BOURET
LABORATOIRE JEAN-ALEXANDRE DIEUDONNÉ
CNRS - UNIVERSITÉ DE NICE SOPHIA-ANTIPOLIS
PARC VALROSE
06108 NICE CEDEX 2 (FRANCE)
E-MAIL: Patricia.Reynaud-Bouret@unice.fr