The Annals of Statistics
2013, Vol. 41, No. 3, 1431–1461
DOI: 10.1214/13-AOS1114

THE TWO-SAMPLE PROBLEM FOR POISSON PROCESSES: ADAPTIVE TESTS WITH A NONASYMPTOTIC WILD BOOTSTRAP APPROACH[1]

BY MAGALIE FROMONT, BÉATRICE LAURENT AND PATRICIA REYNAUD-BOURET

IRMAR (Université Européenne de Bretagne) - CREST (Ensai), IMT, INSA Toulouse, Université de Toulouse and Université de Nice Sophia-Antipolis, CNRS

Considering two independent Poisson processes, we address the question of testing equality of their respective intensities. We first propose testing procedures whose test statistics are U-statistics based on single kernel functions. The corresponding critical values are constructed from a nonasymptotic wild bootstrap approach, leading to level α tests. Various choices for the kernel functions are possible, including projection, approximation or reproducing kernels. In this last case, we obtain a parametric rate of testing for a weak metric defined in the RKHS associated with the considered reproducing kernel. Then we introduce, in the other cases, aggregated or multiple kernel testing procedures, which allow us to import ideas coming from model selection, thresholding and/or approximation kernels adaptive estimation. These multiple kernel tests are proved to be of level α, and to satisfy nonasymptotic oracle-type conditions for the classical $\mathbb{L}_2$-norm. From these conditions, we deduce that they 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.

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 nonatomic positive 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) "$f = g$" versus 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 examples, [12, 22, 48] or [55], whose applications were mainly turned to biology and medicine, and less frequently to

Received March 2012; revised February 2013.

[1]Supported by 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).

MSC2010 subject classifications. Primary 62G09, 62G10, 62G55; secondary 62G20.

Key words and phrases. Two-sample problem, Poisson process, bootstrap, adaptive tests, minimax separation rates, kernel methods, aggregation methods, multiple kernel.

reliability. More recent papers like [11, 39, 45] and [10] give interesting numerical comparisons of various testing procedures. As for nonhomogeneous Poisson processes, though a lot of references on the problem of testing proportionality of the hazard rates of the processes exist (see [17], e.g., and the references therein), very few papers are devoted to a comparison of the intensities themselves. Bovett and Saw [6] and Deshpande et al. [16], respectively, proposed conditional and unconditional procedures to test the null hypothesis "f/g is constant" versus "it is increasing." Deshpande et al. [16] 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) versus (H_1), whose test statistics are U-statistics based on a single kernel function, chosen either as a projection kernel, or as an approximation kernel, or as a reproducing kernel. Without any parametric or monotony assumption on f or g, these single kernel tests satisfy specific nonasymptotic performance properties.

In particular, for every α in [0, 1], they are of level α; that is, they have a probability of first kind error at most equal to α. For special values of α, they are even of size α; that is, their probability of first kind error is exactly equal to α, since they involve very sharp critical values obtained via a nonasymptotic wild bootstrap approach. In the classical two-sample problem for i.i.d. samples, the choice of the critical values in testing procedures is a crucial question. Indeed, the asymptotic distributions of many test statistics are not free from the common unknown density under the null hypothesis. In such cases, some bootstrap methods are often used to build data-driven critical values. By bootstrap methods, we mean the original ones introduced by Efron [18] of course, but also more general weighted bootstrap approaches such as the precursor Fisher's [19] permutation, the m out of n bootstrap introduced by Bretagnolle [7], the general exchangeably weighted bootstrap studied in [46] and including the Bayesian bootstrap of Rubin [53], for instance, as well as the wild bootstrap detailed in [43]. Except in the cases where the permutation approach is used, authors generally prove that the obtained tests are (only) asymptotically of level α; see, among many other papers, [47, 50, 51] and more recently [34] for a complete and very interesting discussion. In this work, we adopt one of these general weighted bootstrap approaches, but from a nonasymptotic point of view. The critical values of our tests are constructed from wild bootstrapped U-statistics, which are based on Rademacher variables. The use of Rademacher variables is well known in the bootstrap community since the work of Mammen [43], but also particularly in the statistical learning community since the works of Koltchinskii [36] and Bartlett et al. [5], followed by [37]. It was notably proposed for the construction of general confidence bands in a recent paper by Lounici and Nickl [42]. The main particularity of our study, as compared with previous ones, is that we prove here that, under (H_0), given the data, the considered wild bootstrapped U-statistics exactly have the same distributions as our test statistics. The corresponding tests are consequently of level α for every α in [0, 1], and even of size α for particular values of α. Note that as in [52] or in [27], it is also possible to randomize these tests in order to turn them into size α tests for every α. In this sense, our bootstrap method can be viewed as an adapted version of the permutation bootstrap method in a Poisson framework. As usual, even when permutation methods are considered, the wild bootstrapped critical values of our tests are not computed exactly in practice, but just approximated through a Monte Carlo method. We also address this question from a nonasymptotic point of view, since we also focus on controlling the loss due to the Monte Carlo approximation.

A nonasymptotic study of the second kind error of our single kernel tests is also performed. Given any β in $[0, 1]$, depending on the chosen kernel, we obtain nonasymptotic conditions which guarantee that the probability of second kind error is at most equal to β. This can be done via a sharp control of the wild bootstrapped critical values under the alternative, which results from concentration inequalities for Rademacher chaoses [14, 41].

In order to deduce from these conditions recognizable asymptotic rates of testing, we assume that the measure μ on $\mathbb{X}$ satisfies $d\mu = n\,d\nu$, where n can be seen as a growing number, whereas the measure ν is held fixed. Typically, n may be 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 intensities f and g w.r.t. ν. The reader may also assume for sake of simplicity that $\mathbb{X}$ is a measurable subset of $\mathbb{R}^d$ and that ν is the Lebesgue measure, but it is not required: ν may be any nonatomic positive σ-finite measure on any measurable set $\mathbb{X}$. With this normalization, when a reproducing kernel is considered, we obtain a parametric rate of testing for a weak metric defined in the associated RKHS, in the spirit of [61] or [23] for more classical weak metrics in i.i.d. samples frameworks. Our results complete those of Gretton et al. [25], who introduced reproducing kernels in the two-sample problem for i.i.d. samples. When a projection or an approximation kernel is considered, we obtain the following condition: the probability of second kind error of the test is at most equal to β as soon as the $\mathbb{L}_2$-distance w.r.t. ν between f and g is larger than a bound, which reproduces a bias–variance decomposition. This bound can be proved to be optimal with an appropriate choice of the kernel and/or the parameters of the kernel such as the vectorial space for a projection kernel or the bandwidth for an approximation kernel, choice which highly depends on the alternative.

In order to avoid choosing such a particular kernel and/or its parameters, we propose to aggregate several of the previous single kernel tests, making sure that the resulting multiple kernel test is still of level α. We also establish oracle-type conditions, which guarantee that the probability of second kind error is at most equal to β. The idea of considering multiple kernel methods instead of single kernel ones has led to recent developments in learning theory, under the name of "multiple kernel learning" problems; see [2, 9, 40, 44] or [38], for instance. Multiple kernel testing procedures for the classical two-sample problem for i.i.d. samples, which is closely related to the present problem, have even been proposed in [58] and [26].

However, the aggregation approaches developed in these papers differ from the one we consider here, which was inspired by adaptive estimation methods such as model selection, thresholding or approximation kernels methods, and was used in many papers devoted to adaptive testing in various classical one-sample frameworks; see [56] or [57] for adaptive tests related to thresholding methods, [32] for adaptive tests related to model selection methods, [28] for adaptive tests related to approximation kernels methods or [4] for adaptive tests related to both model selection and thresholding methods, for instance. In a Poisson process framework, we proposed in [20] an adaptive test of homogeneity also based on both model selection and thresholding approaches. In the two-sample problem for i.i.d. samples, Butucea and Tribouley [8] proposed an adaptive test based on a thresholding approach.

We complete the study by proving that some of our multiple kernel tests are also adaptive in a nonasymptotic minimax sense over various classes $\mathcal{S}_\delta$ of alternatives (f, g) for which $(f - g)$ is smooth with parameter δ. For clarity's sake, let us here recall a few definitions. For any level α test Φ_α, with values in $\{0, 1\}$ [rejecting (H_0) when $\Phi_\alpha = 1$], one defines its uniform separation rate $\rho(\Phi_\alpha, \mathcal{S}_\delta, \beta)$ over $\mathcal{S}_\delta$ as

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

where $\|f - g\|^2 = \int (f - g)^2 \, d\nu$ and $\mathbb{P}_{f,g}$ denotes the joint distribution of (N^1, N^{-1}). A level α test Φ_α is said to be minimax over a particular class $\mathcal{S}_\delta$ if its uniform separation rate achieves its best possible value over $\mathcal{S}_\delta$, which is called the minimax separation rate over $\mathcal{S}_\delta$ (see [3]) up to a multiplicative factor. It is said to be minimax adaptive if its uniform separation rates achieve (up to a possible unavoidable small loss) the minimax separation rates over several classes $\mathcal{S}_\delta$ simultaneously. A great number of papers deal with the computation of the minimax separation rates over various classes of alternatives, or more precisely with the computation of their asymptotic equivalents, that are the minimax rates of testing defined in the key series of papers due to Ingster [31]. The question of the minimax adaptivity has also been widely studied since the work of Spokoiny [56], who first brought out a context where minimax adaptive testing without a small loss of efficiency is impossible. For the problem of testing the goodness-of-fit of a Poisson process, Ingster and Kutoyants [33] derived the minimax rate of testing over a Sobolev or a Besov ball. For the problem of testing the homogeneity of a Poisson process, we derived in [20] similar minimax results considering classical Besov bodies, and we moreover obtained new minimax adaptivity results considering weak Besov bodies.

In the present two-sample problem for Poisson processes, no previous minimax result is available to our knowledge. As in [20], we here prove that the aggregation of single projection kernel tests leads to minimax adaptive tests over some classes of alternatives for which $(f - g)$ belongs to a Besov or a weak Besov body. Such a result can be linked to the minimax results obtained by Butucea and Tribouley [8], noting however that the classes of alternatives they consider impose both f and g to belong to a Besov space, which is more restrictive than only imposing some regularity assumptions on $(f-g)$. Then, when considering the aggregation of single approximation kernel tests, we obtain upper bounds for the uniform separation rates over some classes of alternatives based on multivariate Sobolev or anisotropic Nikol'skii–Besov balls. These upper bounds, which are conjectured to be optimal from results of Horowitz and Spokoiny [28] or Ingster and Stepanova [30] in other frameworks, are completely new in our Poisson setting, and even in a general setting for anisotropic Nikol'skii–Besov balls.

The paper is organized as follows. In Section 2, we introduce single kernel testing procedures. As explained above, the corresponding critical values are constructed from a wild bootstrap approach, leading to level α tests. We then give conditions ensuring that these single kernel tests also have a probability of second kind error at most equal to β, and we study the cost due to the Monte Carlo approximation of the wild bootstrapped critical values. In Section 3, we construct multiple kernel testing procedures of level α, by aggregating several of the single kernel tests introduced in Section 2. Oracle-type conditions are obtained, ensuring that these multiple kernel tests have a probability of second kind error at most equal to β. From these conditions, some of our tests are also proved to be minimax adaptive over various classes of alternatives based on classical and weak Besov bodies in the univariate case, or Sobolev and anistropic Nikol'skii–Besov balls in the multivariate case. The major proofs are given in Section 4, whereas a simulation study and the other proofs may be found in supplementary materials [21].

Let us now introduce some notation that will be used throughout the paper. For any measurable function h, let, when they exist $\|h\|_\infty = \sup_{x\in\mathbb{X}} |h(x)|$, and $\|h\|_1 = \int_{\mathbb{X}} |h(x)|\,d\nu_x$. Recalling that $\|h\| = (\int_{\mathbb{X}} h(x)^2\,d\nu_x)^{1/2}$, we introduce the scalar product $\langle\cdot,\cdot\rangle$ 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}), $\mathbb{E}_{f,g}$ stands for 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 will introduce some constants that we do not intend to evaluate here and that are denoted by $C(\alpha, \beta, \ldots)$, meaning that they may depend on $\alpha, \beta, \ldots$. Though they are denoted in the same way, they may vary from one line to another.

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

ASSUMPTION 1. $\|f\|_1 < +\infty$ and $\|g\|_1 < +\infty$.

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

2. Single kernel testing procedures with nonasymptotic wild bootstrapped critical values.

2.1. *Single kernel test statistics.* Since f and g are assumed to satisfy Assumptions 1 and 2, they are also assumed to belong to $\mathbb{L}^2(\mathbb{X}, d\nu)$. Hence, testing (H_0) "$f = g$" versus (H_1) "$f \neq g$" here amounts to testing that "$\|f - g\| = 0$" versus "$\|f - g\| > 0$." Considering a well-chosen finite dimensional subspace S of $\mathbb{L}^2(\mathbb{X}, d\nu)$, if Π_S denotes the orthogonal projection onto S for $\langle\cdot,\cdot\rangle$, any estimator of an increasing function of $\|\Pi_S(f - g)\|^2$ may thus be a relevant candidate to be a test statistic. Let $\{\varphi_\lambda, \lambda \in \Lambda\}$ be an orthonormal basis of S for $\langle\cdot,\cdot\rangle$, and let

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

where N is the pooled Poisson process whose point measure is given by $dN = dN^1 + dN^{-1}$. Since $\mathbb{E}[(\int \varphi_\lambda \, dN^1)^2] = (\int \varphi_\lambda(x) f(x) \, d\mu_x)^2 + \int \varphi_\lambda^2(x) f(x) \, d\mu_x$, and similarly for $\mathbb{E}[(\int \varphi_\lambda \, dN^{-1})^2]$, and recalling that $d\mu = n \, d\nu$, it is easy to see that $\hat{T}$ is an unbiased estimator of $n^2\|\Pi_S(f - g)\|^2$, and thus also a possible test statistic, whose large values lead to reject (H_0).

Let $(\varepsilon_x^0)_{x\in N}$ be the marks of the points from the pooled process N, defined by $\varepsilon_x^0 = 1$ if the point x of N belongs to N^1 and $\varepsilon_x^0 = -1$ if the point x of N belongs to N^{-1}. Then $\hat{T}$ can also be expressed as

$$\hat{T} = \sum_{\lambda\in\Lambda} \sum_{x\neq x'\in N} \varphi_\lambda(x)\varphi_\lambda(x')\varepsilon_x^0\varepsilon_{x'}^0 = \sum_{x\neq x'\in N} \left(\sum_{\lambda\in\Lambda} \varphi_\lambda(x)\varphi_\lambda(x')\right)\varepsilon_x^0\varepsilon_{x'}^0.$$

Starting from this remark, we can thus generalize the test statistic $\hat{T}$ by replacing in its expression the function: $(x, x') \in \mathbb{X}^2 \mapsto \sum_{\lambda\in\Lambda} \varphi_\lambda(x)\varphi_\lambda(x') \in \mathbb{R}$ by a general symmetric kernel function. So, let K be any symmetric kernel function $\mathbb{X} \times \mathbb{X} \to \mathbb{R}$ satisfying the following:

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

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

$$\hat{T}_K = \sum_{x\neq x'\in N} K(x, x')\varepsilon_x^0\varepsilon_{x'}^0 = \int_{\mathbb{X}^{[2]}} K(x, x')\varepsilon_x^0\varepsilon_{x'}^0 \, dN_x \, dN_{x'}. \tag{2.1}$$

Since for every x in N, $\mathbb{E}[\varepsilon_x^0|N] = (f(x) - g(x))/(f(x) + g(x))$ (see Proposition 1 below, e.g.),

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

$$= \int_{\mathbb{X}^2} K(x,x')(f-g)(x)(f-g)(x')\,d\mu_x d\mu_{x'}$$
$$= n^2 \int_{\mathbb{X}^2} K(x,x')(f-g)(x)(f-g)(x')\,d\nu_x d\nu_{x'}.$$

In the following, we use the notation

$$K[p](x') = \int_{\mathbb{X}} K(x,x')p(x)\,d\nu_x. \tag{2.2}$$

With this notation, $\hat{T}_K$ is then an unbiased estimator of

$$\mathcal{E}_K = n^2 \langle K[f-g], f-g\rangle, \tag{2.3}$$

whose existence is ensured thanks to Assumptions 1 and 3.

We have chosen to consider and study in this paper three possible examples of kernel functions. For each example, we give a simpler expression of $\mathcal{E}_K$, which allows us to justify the choice of $\hat{T}_K$ as test statistic.

Projection kernel case. Our first choice for K is a symmetric kernel function based on an orthonormal family $\{\varphi_\lambda, \lambda \in \Lambda\}$ for $\langle\cdot,\cdot\rangle$,

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

When the cardinality of Λ is finite, $\hat{T}_K$ corresponds to the above natural test statistic $\hat{T}$. When the cardinality of Λ is infinite, we assume that $\sup_{x,x'\in\mathbb{X}} \sum_{\lambda\in\Lambda} |\varphi_\lambda(x)\varphi_\lambda(x')| < +\infty$, 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 $\{\varphi_\lambda, \lambda \in \Lambda\}$ correspond to indicator functions with disjoint supports, this condition will be satisfied.

We check in these cases that for every s in $\mathbb{L}^2(\mathbb{X}, d\nu)$, $K[s] = \Pi_S(s)$, where S is the subspace of $\mathbb{L}^2(\mathbb{X}, d\nu)$ generated by $\{\varphi_\lambda, \lambda \in \Lambda\}$, and Π_S denotes as above the orthogonal projection onto S for $\langle\cdot,\cdot\rangle$. This justifies that such a kernel function K is called a projection kernel and that

$$\mathcal{E}_K = n^2 \|\Pi_S(f-g)\|^2.$$

Approximation kernel case. 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. Note 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,

$$\mathcal{E}_K = n^2 \langle k_h * (f-g), f-g\rangle,$$

where $k_h(u_1,\dots,u_d)=\frac{1}{\prod_{i=1}^d h_i}k(\frac{u_1}{h_1},\dots,\frac{u_d}{h_d})$ and $*$ is the usual convolution operator with respect to the measure ν.

Reproducing kernel case. Our third choice for K is a general reproducing kernel (see [54], e.g.) 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 test statistic close to the one of Gretton et al. [25] for the classical two-sample problem for i.i.d. samples of equal sizes. We will, however, see that the corresponding critical value is not constructed here in the same way as in [25]. While Gretton et al. derive their critical value from either concentration inequalities, or asymptotic arguments, or an asymptotic Efron's bootstrap approach, we construct our critical value from a nonasymptotic wild bootstrap approach.

In this case, it is easy to see that

$$\mathcal{E}_K=n^2\|m_f-m_g\|^2_{\mathcal{H}_K},$$

where $m_f=\int_{\mathbb{X}}K(\cdot,x)f(x)\,d\nu_x$ and $m_g=\int_{\mathbb{X}}K(\cdot,x)g(x)\,d\nu_x$. Note that in a "density" context where $\int_{\mathbb{X}}f(x)\,d\nu_x=\int_{\mathbb{X}}g(x)\,d\nu_x=1$, $\mathcal{E}_K$ is n^2 times the so-called squared maximum mean discrepancy on the unit ball in the RKHS $\mathcal{H}_K$ (see [25]) between the distributions $f\,d\nu$ and $g\,d\nu$, and that the functions m_f and m_g are known (see [59] e.g.) as the mean embeddings in $\mathcal{H}_K$ of the distributions $f\,d\nu$ and $g\,d\nu$, respectively. Moreover, in this context, assuming that the kernel K is characteristic (see also [59]), the map which assigns its mean embedding in $\mathcal{H}_K$ to any probability distribution is injective by definition, so $\mathcal{E}_K=0$ if and only if $f=g$.

We want to mention here that the introduction of reproducing kernels is particularly pertinent 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 not easy to deal with. In such situations, the use of reproducing kernels may be the only possible way to compute a meaningful test; see [25] where such kernels are used for microarrays data and graphs.

Thus, for each of the three above choices for K, considering a test which rejects (H_0) when $\hat{T}_K$ is "large enough" seems to be reasonable. It remains to explain what we mean by "large enough," that is, to define the critical values used in our tests.

2.2. *Critical values based on a nonasymptotic wild bootstrap approach*. The critical values we use here are based on a nonasymptotic wild bootstrap approach that we present and justify in this section. To do this, we start from the remark that under (H_0), the test statistic $\hat{T}_K$ is a degenerate U-statistic of order 2, for which adequate bootstrap methods were developed in particular in [7] and [1]. Bretagnolle [7] first noticed that a naive application of Efron's original bootstrap fails for degenerate U-statistics, since it leads the bootstrapped statistic to lose the degeneracy property. He therefore introduced the more appropriate m of n bootstrap, while Arcones and Giné [1] preferred to keep on using Efron's original bootstrap, but by forcing the bootstrapped statistic to satisfy the degeneracy property through a centering trick. The results of Arcones and Giné were then generalized to other kinds of bootstrap methods, and in particular Bayesian and wild bootstrapped U-statistics were introduced in [29, 35] and [15].

Following [15], we introduce a sequence $(\varepsilon_i)_{i\in\mathbb{N}}$ of i.i.d. Rademacher variables independent of N. Denoting by N_n the size of the pooled process N, and by $\{X_1,\dots,X_{N_n}\}$ the points of N, a wild bootstrapped version of $\hat{T}_K$ may be expressed as $\sum_{i\neq i'\in\{1,\dots,N_n\}} K(X_i,X_{i'})\varepsilon^0_{X_i}\varepsilon^0_{X_{i'}}\varepsilon_i\varepsilon_{i'}$. We consider in fact the simpler version

$$\hat{T}_K^{\varepsilon}=\sum_{i\neq i'\in\{1,\dots,N_n\}} K(X_i,X_{i'})\varepsilon_i\varepsilon_{i'}, \tag{2.4}$$

that can be proved to have, under (H_0), conditionally on N, the same distribution as the above wild bootstrapped version of $\hat{T}_K$. We now choose the quantile of the conditional distribution of $\hat{T}_K^{\varepsilon}$ given N as critical value for our test.

More precisely, for α in $(0,1)$, if $q^{(N)}_{K,1-\alpha}$ denotes the $(1-\alpha)$ quantile of the distribution of $\hat{T}_K^{\varepsilon}$ conditionally on N, we consider the test that rejects (H_0) when $\hat{T}_K>q^{(N)}_{K,1-\alpha}$. The corresponding test function is defined by

$$\Phi_{K,\alpha}=\mathbf{1}_{\hat{T}_K>q^{(N)}_{K,1-\alpha}}. \tag{2.5}$$

Note that in practice, the true conditional quantile $q^{(N)}_{K,1-\alpha}$ is not exactly computed, but in fact just approximated by a classical Monte Carlo method.

Of course, such bootstrap tests are not completely new in the statistical scene. However, the main particularities of our work is that we justify our test from a nonasymptotic point of view. We actually prove that under (H_0), conditionally on N, $\hat{T}_K$ and $\hat{T}_K^{\varepsilon}$ exactly have the same distribution. As a consequence the test defined by $\Phi_{K,\alpha}$ is of level α, that is, it has a probability of first kind error at most equal to α. We will briefly see in the next section that it may even be randomized to be of size α, that is, to have a probability of first kind error exactly equal to α.

In the same way, instead of focusing as many previous authors on the consistence against some alternatives, we give precise conditions on the alternatives which guarantee that $\Phi_{K,\alpha}$ has a probability of second kind error controlled by a prescribed value β in $(0,1)$. These results are detailed in the next section.

Furthermore, we do not forget that studying our tests from a nonasymptotic point of view poses the additional question of the exact loss in probabilities of first and second kind errors due to the Monte Carlo approximation of $q^{(N)}_{K,1-\alpha}$. We also address this question in Section 2.4.

Such a nonasymptotic approach is actually conceivable thanks to the following proposition, which is well known in the point processes literature and which can be deduced from a general result of [13]. A complete proof is given in the supplementary materials [21] for sake of understanding.

PROPOSITION 1. *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. Moreover, let $(\varepsilon^0_x)_{x\in N}$ be defined by $\varepsilon^0_x = 1$ if x belongs to N^1 and $\varepsilon^0_x = -1$ if x belongs to N^{-1}. Then, conditionally on N, the variables $(\varepsilon^0_x)_{x\in N}$ are independent and for every x in N,*

$$\mathbb{P}(\varepsilon^0_x = 1|N) = \frac{f(x)}{f(x)+g(x)}, \qquad \mathbb{P}(\varepsilon^0_x = -1|N) = \frac{g(x)}{f(x)+g(x)}. \tag{2.6}$$

2.3. *Probabilities of first and second kind errors.* We here study the probabilities of first and second kind errors of the test $\Phi_{K,\alpha}$ defined by (2.5).

From Proposition 1, we deduce that under (H_0), $\hat{T}_K$ and $\hat{T}^{\varepsilon}_K$ exactly have the same distribution conditionally on N. As a result, given α in (0, 1), under (H_0),

$$\mathbb{P}(\hat{T}_K > q^{(N)}_{K,1-\alpha}|N) \le \alpha. \tag{2.7}$$

By taking the expectation over N, we obtain that

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

In fact, inequality (2.7) can be turned in an equality only for some particular values of α, due to the discreteness of the conditional distribution of $\hat{T}_K$ given N. To go a little further, from Proposition 1, we deduce that the randomization hypothesis as defined by Romano and Wolf [52] and introduced by Hoeffding [27] is satisfied. From the construction of Hoeffding [27], one can therefore randomize $\Phi_{K,\alpha}$ to obtain a test $\Psi_{K,\alpha}$ such that $\Psi_{K,\alpha} \ge \Phi_{K,\alpha}$ a.s. and such that under (H_0), $\mathbb{E}(\Psi_{K,\alpha}|N) = \alpha$ for every α. Thus, by using the classical tool of randomization, one can circumvent the trouble due to the atoms of the discrete conditional distribution of $\hat{T}_K$ given N, and obtain a test with a probability of first kind error exactly equal to α for every α. Note that the randomized test $\Psi_{K,\alpha}$ necessarily has a probability of second kind error smaller than $\Phi_{K,\alpha}$'s one, since $\Psi_{K,\alpha} \ge \Phi_{K,\alpha}$ a.s.

However, in practice, since the conditional quantile $q^{(N)}_{K,1-\alpha}$ is approximated by a Monte Carlo method as we have explained above, we do not have access to the true randomized version of $\Phi_{K,\alpha}$. This explains why we have decided to focus in the following on the nonrandomized test $\Phi_{K,\alpha}$.

Given β in $(0,1)$, we now aim at bringing out a nonasymptotic condition on the alternative (f,g) which will guarantee that $\mathbb{P}_{f,g}(\Phi_{K,\alpha}=0)\leq\beta$. Denoting by $q^{\alpha}_{K,1-\beta/2}$ the $(1-\beta/2)$ quantile of the conditional quantile $q^{(N)}_{K,1-\alpha}$,

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

Thus, a condition which guarantees that $\mathbb{P}_{f,g}(\hat{T}_K\leq q^{\alpha}_{K,1-\beta/2})\leq\beta/2$ will be enough to ensure that $\mathbb{P}_{f,g}(\Phi_{K,\alpha}=0)\leq\beta$. The following proposition gives such a condition.

PROPOSITION 2. *Let α,β be fixed levels in $(0,1)$, and let us recall that for any symmetric kernel function K satisfying Assumption 3, $\mathbb{E}_{f,g}[\hat{T}_K]=\mathcal{E}_K$, with $\mathcal{E}_K$ given in (2.3). Markov's inequality implies that if*

$$\mathcal{E}_K>\sqrt{\frac{2\operatorname{Var}(\hat{T}_K)}{\beta}}+q^{\alpha}_{K,1-\beta/2}, \tag{2.8}$$

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

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

Setting $A_K=\int_{\mathbb{X}}(K[f-g](x))^2(f+g)(x)\,d\nu_x$ and $B_K=\int_{\mathbb{X}^2}K^2(x,x')(f+g)(x)(f+g)(x')\,d\nu_x d\nu_{x'}$, we have

$$\operatorname{Var}(\hat{T}_K)=4n^3A_K+2n^2B_K.$$

Moreover, there exists some constant $\kappa>0$ such that, for every K,

$$q^{\alpha}_{K,1-\beta/2}\leq\kappa\ln(2/\alpha)n\sqrt{\frac{2B_K}{\beta}}. \tag{2.9}$$

Comments. In this proposition, we simply use Markov's inequality since obtaining precise constants and dependency in β is not crucial here; see Section 4. The computation of $\operatorname{Var}(\hat{T}_K)$ is obtained from factorial moment measures, while the control of $q^{\alpha}_{K,1-\beta/2}$ derives from a property of Rademacher chaoses combined with an exponential inequality; see [14] and [41].

The following theorem allows us to better understand Proposition 2, and to deduce from it more recognizable properties in terms of uniform separation rates.

THEOREM 1. *Let α,β be fixed levels in $(0,1)$. Let K be a symmetric kernel function satisfying Assumption 3, and $\Phi_{K,\alpha}$ be the test defined by (2.5). Let C_K be an upper bound for $\int_{\mathbb{X}^2}K^2(x,x')(f+g)(x)(f+g)(x')\,d\nu_x\,d\nu_{x'}$. Then, we have*

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

$$
\begin{aligned}
&\|f-g\|^2 \\
&\geq \inf_{r>0}\left[\|(f-g) - r^{-1}K[f-g]\|^2 + \frac{4+2\sqrt{2}\kappa\ln(2/\alpha)}{nr\sqrt{\beta}}\sqrt{C_K}\right] \\
&\quad + \frac{8\|f+g\|_\infty}{\beta n}.
\end{aligned}
\tag{2.10}
$$

For instance, C_K *can be taken as follows*:

- $C_K = \|f+g\|_\infty^2 D$ *when* K *is chosen as in the* projection kernel case, *considering an orthonormal basis* $\{\varphi_\lambda, \lambda \in \Lambda\}$ *of a* D*-dimensional subspace* S *of* $\mathbb{L}^2(\mathbb{X}, d\nu)$;
- $C_K = \|f+g\|_\infty \|f+g\|_1 D$ *when* K *is chosen as in the* projection kernel case, *considering an orthonormal basis* $\{\varphi_\lambda, \lambda \in \Lambda\}$ *of a possibly infinite dimensional subspace* S *of* $\mathbb{L}^2(\mathbb{X}, d\nu)$, *which satisfies*

$$
\sup_{x,x' \in \mathbb{X}} \sum_{\lambda \in \Lambda} |\varphi_\lambda(x)\varphi_\lambda(x')| = D < +\infty, \tag{2.11}
$$

$$
\int_{\mathbb{X}^2} \left(\sum_{\lambda \in \Lambda} |\varphi_\lambda(x)\varphi_\lambda(x')|\right)^2 (f+g)(x')\, d\nu_x\, d\nu_{x'} < +\infty; \tag{2.12}
$$

- $C_K = \|f+g\|_\infty \|f+g\|_1 \|k\|^2 / \prod_{i=1}^d h_i$ *when* K *is chosen as in the approximation kernel case*.

Comments.

(1) When K is chosen as in the *projection kernel case*, then $K[f-g] = \Pi_S(f-g)$. Hence by taking $r=1$ in (2.10), the right-hand side of the inequality 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 [3], e.g.), and we know that this leads to sharp upper bounds for the uniform separation rates over particular classes of alternatives.

(2) When K is chosen as in the *approximation kernel case* with k in $\mathbb{L}^1(\mathbb{R}^d)$, $\int_{\mathbb{R}^d} k(x)\, d\nu_x = 1$, and $h_1 = \cdots = h_d$, then $K[f-g] = k_h * (f-g)$, and $\|(f-g) - K[f-g]\|$ is a bias term. Hence by taking $r=1$ in the inequality (2.10), we still reproduce a bias–variance decomposition, but with a variance term of order $h_1^{-d/2}/n$, which coincides with the above variance term in the *projection kernel case* through the equivalence $h_1^{-d} \sim D$. This equivalence is usual in the approximation estimation theory; see [60], for instance, for more details.

(3) When K is chosen as in the *reproducing kernel case*, if K is proportional to a kernel from the two above cases, then one can appropriately choose the constant

r such that $\|(f-g)-r^{-1}K[f-g]\|$ is still a bias term. We thus recover for such kernel functions, such as the Gaussian and Laplacian kernels, which are commonly used in statistical learning theory, the same bias–variance decomposition as above. However, in some cases, one cannot find any normalization constant r for which $\|(f-g)-r^{-1}K[f-g]\|$ can be viewed as a bias term, and the result cannot be interpreted from a statistical point of view. In these cases in particular, the $\mathbb{L}^2$-norm which is considered in Theorem 1 is not the appropriate one to obtain relevant uniform separation rates, since it does not necessarily have any link with the norm of the RKHS $\mathcal{H}_K$. We give in the following theorem a more adequate result for the specific *reproducing kernel case*.

THEOREM 2. *Let α, β be fixed levels in $(0,1)$, and $\kappa > 0$ be the constant of Proposition 2. Let $\mathbb{X} = \mathbb{R}^d$ and K be a kernel function on $\mathbb{X} \times \mathbb{X}$ chosen as in the reproducing kernel case. Let $\Phi_{K,\alpha}$ be the test function defined by (2.5). We assume furthermore that $\int_{\mathbb{X}} f(x)\,d\nu_x = \int_{\mathbb{X}} g(x)\,d\nu_x = 1$, that K is a bounded measurable characteristic kernel and that $K(x,x)$ is constant equal to κ_0. Let m_f and m_g be the mean embeddings of the distributions $f\,d\nu$ and $g\,d\nu$, respectively, in $\mathcal{H}_K$. We have $\mathbb{P}_{f,g}(\Phi_{K,\alpha}=0) \leq \beta$ if*

$$\|m_f - m_g\|^2_{\mathcal{H}_K} \geq \frac{4\kappa_0}{n}\left(\frac{4}{\beta} + \frac{2+\kappa\sqrt{2}\ln(2/\alpha)}{\sqrt{\beta}}\right).$$

Comments.

(1) The assumption that $K(x,x)$ is constant is usual, since it is satisfied by any normalized or translation-invariant kernel; see [54] pages 46–47, 57, or [59], for instance. Moreover, as specified in [59], for instance, bounded continuous characteristic and translation-invariant reproducing kernels exist, at least in $\mathbb{R}^d$, where Bochner's theorem enables one to characterize them.

(2) The result that we have here is in fact comparable to the one obtained by Wellner [61] for two-sample tests in an i.i.d. samples framework. While Wellner's test is based on the estimation of a weak distance between $f\,d\nu$ and $g\,d\nu$, associated with the Sobolev norm with negative index, our test statistic is an unbiased estimator of $\mathcal{E}_K = n^2\|m_f - m_g\|^2_{\mathcal{H}_K}$, where $\|m_f - m_g\|_{\mathcal{H}_K} = \sup_{\|r\|_{\mathcal{H}_K}\leq 1}\int_{\mathbb{X}}(f-g)(x)r(x)\,d\nu_x$ defines a weak distance between the distributions $f\,d\nu$ and $g\,d\nu$. As in [61] (or [23] beforehand for the problem of testing uniformity), we obtain a uniform separation rate for this weak distance of the same order as the usual parametric separation rate, that is, of order $n^{-1/2}$.

2.4. *Performance of the Monte Carlo approximation.*

2.4.1. *Probability of first kind error.* In practice, a Monte Carlo method is used to approximate the conditional quantiles $q^{(N)}_{K,1-\alpha}$. It is therefore necessary 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 (2.1), and $q_{K,1-\alpha}^{(N)}$ is the $(1-\alpha)$ quantile of $\hat{T}_K^{\varepsilon}$ defined by (2.4) conditionally on N. The conditional quantile $q_{K,1-\alpha}^{(N)}$ is estimated by $\hat{q}_{K,1-\alpha}^{(N)}$ via the Monte Carlo method as follows. Conditionally on N, we consider a set of B independent sequences $\{\varepsilon^b, 1 \le b \le B\}$, where $\varepsilon^b = (\varepsilon_x^b)_{x\in N}$ is a sequence of i.i.d. Rademacher random variables. We define, for $1 \le b \le B$, $\hat{T}_K^{\varepsilon^b} = \sum_{x\neq x' \in N} K(x, x')\varepsilon_x^b \varepsilon_{x'}^b$. Under (H_0), conditionally on 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 on N) of the sample $(\hat{T}_K^{\varepsilon^b}, 1 \le b \le B)$,

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

Then, $\hat{q}_{K,1-\alpha}^{(N)}$ is defined by $\hat{q}_{K,1-\alpha}^{(N)} = \inf\{t \in \mathbb{R}, F_{K,B}(t) \ge 1-\alpha\}$. We finally consider the test given by

$$\hat{\Phi}_{K,\alpha} = \mathbf{1}_{\hat{T}_K > \hat{q}_{K,1-\alpha}^{(N)}}. \tag{2.13}$$

PROPOSITION 3. *Let α be some fixed level in $(0,1)$ and $\hat{\Phi}_{K,\alpha}$ be the test defined by* (2.13). *Under* (H_0),

$$\mathbb{P}(\hat{\Phi}_{K,\alpha} = 1|N) \le \frac{\lfloor B\alpha \rfloor + 1}{B+1}.$$

Comment. For example, if $B = 200$ and $\alpha = 0.05$, $\hat{\Phi}_{K,\alpha}$ is of level 5.5%.

2.4.2. *Probability of second kind error.*

PROPOSITION 4. *Let α and β be fixed levels in $(0,1)$ such that $\alpha_B = \alpha - \sqrt{\ln B/(2B)} > 0$ and $\beta_B = \beta - 2/B > 0$. Let $\hat{\Phi}_{K,\alpha}$ be the test given in* (2.13). *Let $\mathcal{E}_K$, A_K, B_K and κ as in Proposition* 2, *and let $q_{K,1-\beta_B/2}^{\alpha_B}$ be the $(1-\beta_B/2)$ quantile of $q_{K,1-\alpha_B}^{(N)}$. If*

$$\mathcal{E}_K > \sqrt{\frac{2\operatorname{Var}(\hat{T}_K)}{\beta}} + q_{K,1-\beta_B/2}^{\alpha_B}, \tag{2.14}$$

with $\operatorname{Var}(\hat{T}_K) = 4n^3 A_K + 2n^2 B_K$, *then* $\mathbb{P}_{f,g}(\hat{\Phi}_{K,\alpha} = 0) \le \beta$. *Moreover,*

$$q_{K,1-\beta_B/2}^{\alpha_B} \le \kappa \ln(2/\alpha_B) n \sqrt{\frac{2B_K}{\beta_B}}. \tag{2.15}$$

Comments. When comparing (2.14) and (2.15) with (2.8) and (2.9) in Proposition 2, we notice that they asymptotically coincide when $B \to +\infty$. Moreover, if $\alpha = \beta = 0.05$ and $B \geq 6000$, the multiplicative factor of $\kappa n\sqrt{B_K}$ is multiplied by a factor of order 1.2 in (2.15) compared with (2.9). If even $B = 200{,}000$, this factor passes from 23.4 in (2.9) to 24.1 in (2.15).

3. Multiple kernel testing procedures. In the above section, we consider testing procedures based on a single kernel function K. Using such single kernel tests, however, leads to the natural question of the choice of the kernel, and/or its parameters: the orthonormal family when K is a projection kernel, the vector of bandwidths h when K is based on an approximation kernel, the parameters of K when it is a reproducing 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 [25], 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 kernel testing procedures by aggregating the corresponding single kernel tests. We propose an adapted choice for the critical value. Then, we prove that these multiple kernel tests satisfy strong statistical properties, such as oracle-type properties and minimax adaptivity properties over many classes of alternatives.

3.1. *Description of the multiple kernel testing procedures*. Let us introduce a finite collection $\{K_m, m \in \mathcal{M}\}$ of symmetric kernel functions: $\mathbb{X} \times \mathbb{X} \to \mathbb{R}$ satisfying Assumption 3. For every m in $\mathcal{M}$, let $\hat{T}_{K_m}$ and $\hat{T}^{\varepsilon}_{K_m}$ be defined by (2.1) and (2.4), respectively, 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 $q^{(N)}_{m,1-u}$ the $(1-u)$ quantile of $\hat{T}^{\varepsilon}_{K_m}$ conditionally on the pooled process N. Given α in (0, 1), we consider the test which rejects (H_0) when there exists at least one m in $\mathcal{M}$ such that

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

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

$$u^{(N)}_\alpha = \sup\Big\{u > 0, \mathbb{P}\Big(\sup_{m\in\mathcal{M}} \big(\hat{T}^{\varepsilon}_{K_m} - q^{(N)}_{m,1-ue^{-w_m}}\big) > 0 \big| N\Big) \leq \alpha\Big\}. \tag{3.1}$$

Let Φ_α be the corresponding test function defined by

$$\Phi_\alpha = \mathbf{1}_{\sup_{m\in\mathcal{M}}(\hat{T}_{K_m} - q^{(N)}_{m,1-u^{(N)}_\alpha e^{-w_m}})>0}. \tag{3.2}$$

Note that given the pooled process N, $u^{(N)}_\alpha$ and the quantile $q^{(N)}_{m,1-u^{(N)}_\alpha e^{-w_m}}$ can be estimated by a Monte Carlo method.

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 β in $(0, 1)$ if one can guarantee it for one of the single kernel tests rejecting (H_0) when $\hat{T}_{K_m} > q^{(N)}_{m,1-u^{(N)}_\alpha e^{-w_m}}$. We can thus combine the results of Theorem 1.

Recall that the idea of considering multiple kernel testing procedures, based on a collection of kernels instead of a single one, has already been developed in [58] (and also very recently in [26]) for the two-sample problem with i.i.d. samples. Adapting these procedures to our Poisson framework would, however, not lead to optimal nonasymptotic results. For instance, adapting the test of [58] to our Poisson framework would lead to use $\sup_{m\in\mathcal{M}} \hat{T}_{K_m}$ as test statistic, and compute its critical value by a classical bootstrap method. Such a test would not achieve the nonasymptotic properties, expressed as oracle type conditions or minimax adaptivity results, that we obtain in the following, focusing on collections of kernels chosen as in either the *projection kernel case*, or the *approximation kernel case*.

3.2. *Oracle-type conditions for the probability of second kind error.*

3.2.1. *Multiple kernel testing procedures based on projection kernels.*

THEOREM 3. *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 the conditions (2.11) and (2.12) hold with $\Lambda = \Lambda_m$ 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 defined by (3.2) 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 $\sum_{m\in\mathcal{M}} e^{-w_m} \le 1$.*

Then Φ_α is a level α test. Moreover, $\mathbb{P}_{f,g}(\Phi_\alpha = 0) \le \beta$ if

$$\begin{aligned}
&\|f-g\|^2\\
&\quad\ge \inf_{m\in\mathcal{M}}\biggl\{\|(f-g)-\Pi_{S_m}(f-g)\|^2\\
&\qquad\qquad + \frac{4+2\sqrt{2}\kappa(\ln(2/\alpha)+w_m)}{n\sqrt{\beta}} M(f,g)\sqrt{D_m}\biggr\}\\
&\qquad + \frac{8\|f+g\|_\infty}{\beta n},
\end{aligned}\tag{3.3}$$

where $\kappa > 0$ and $M(f, g) = \max(\|f + g\|_\infty, \sqrt{\|f + g\|_\infty\|f + g\|_1})$.

Comments. Comparing this result with the one obtained in Theorem 1 for the single kernel test based on a projection kernel, one can see that considering the multiple kernel testing procedure allows one to obtain the infimum over all m in

$\mathcal{M}$ on the right-hand side of (3.3) at the price of the additional term w_m. This result can be viewed as an oracle-type property: 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 kernel tests, up to the factor w_m. It will be used to prove that our multiple kernel testing procedures are adaptive over various classes of alternatives.

We focus here on two 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.

Multiple kernel case—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 $\mathbb{L}^2([0,1])$ with

$$\varphi_0(x)=\mathbf{1}_{[0,1]}(x) \quad \text{and} \quad \varphi_{(j,k)}(x)=2^{j/2}\psi\big(2^jx-k\big), \tag{3.4}$$

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 defined by (3.2) with the collection of kernels $\{K_J, J\in\mathcal{M}_{\bar{J}}\}$ and with $\{w_J, J\in\mathcal{M}_{\bar{J}}\}$. We obtain from Theorem 3 that there exists $C(\alpha,\beta,\|f\|_\infty,\|g\|_\infty)>0$ such that $\mathbb{P}_{f,g}(\Phi_\alpha^{(1)}=0)\leq\beta$ if

$$\begin{aligned}\|f-g\|^2 &\\ &\geq C\big(\alpha,\beta,\|f\|_\infty,\|g\|_\infty\big)\\ &\quad\times\inf_{J\in\mathcal{M}_{\bar{J}}}\bigg\{\big\|(f-g)-\Pi_{S_J}(f-g)\big\|^2+\big(\ln(J+2)\big)\frac{2^{J/2}}{n}\bigg\}.\end{aligned} \tag{3.5}$$

Multiple kernel case—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 $\mathbb{L}^2([0,1])$ defined by (3.4). Let for some $\tilde{J}\geq 1$,

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

For any λ in $\{0\}\cup\Lambda_{\tilde{J}}$, we consider the subspace $\tilde{S}_\lambda$ 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 (3.2) with the collection of kernels $\{K_\lambda, \lambda\in\{0\}\cup\Lambda_{\tilde{J}}\}$, 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 3 and Pythagoras' 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

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

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

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

where S_m is generated by $\{\varphi_\lambda,\lambda\in m\}$. Hence, there exists some constant $C(\alpha,\beta,\|f\|_\infty,\|g\|_\infty)>0$ such that $\mathbb{P}_{f,g}(\Phi_\alpha^{(2)}=0)\le\beta$ if

$$\begin{aligned}\|f-g\|^2&\\ &\ge C\big(\alpha,\beta,\|f\|_\infty,\|g\|_\infty\big)\inf_{m\in\mathcal{M}_{\tilde{J}}}\bigg\{\|(f-g)-\Pi_{S_m}(f-g)\|^2\\ &\qquad+\frac{\sum_{\lambda\in m}w_\lambda}{n}\bigg\}.\end{aligned}\tag{3.6}$$

3.2.2. *Multiple kernel testing procedures based on approximation kernels.*

THEOREM 4. *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 bandwidths $(h_{m_2,1},\dots,h_{m_2,d})$. 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 $\mathbb{R}^d$,*

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

Let Φ_α be the test defined by (3.2) with $\{K_m,m\in\mathcal{M}\}$ and a collection $\{w_m,m\in\mathcal{M}\}$ of positive numbers such that $\sum_{m\in\mathcal{M}}e^{-w_m}\le1$.

Then Φ_α is a level α test. Moreover, there exists $\kappa>0$ such that if

$$\begin{aligned}\|f-g\|^2&\\ &\ge\inf_{(m_1,m_2)\in\mathcal{M}}\bigg\{\|(f-g)-k_{m_1,h_{m_2}}*(f-g)\|^2\\ &\qquad+\frac{4+2\sqrt{2}\kappa(\ln(2/\alpha)+w_m)}{n\sqrt{\beta}}\sqrt{\frac{\|f+g\|_\infty\|f+g\|_1\|k_{m_1}\|^2}{\prod_{i=1}^d h_{m_2,i}}}\bigg\}\\ &\quad+\frac{8\|f+g\|_\infty}{\beta n},\end{aligned}$$

then

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

We focus here on two particular examples. The first example involves a collection of nonnecessarily 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.

Multiple kernel case—Example 3. 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 $k_{m_1}(x) = k_{m_1}(-x)$, nonnecessarily 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

$$\operatorname*{Ess\,sup}_{u \in \mathbb{R}^d \setminus \{0\}} \frac{|1 - \widehat{k}_{m_1}(u)|}{\|u\|_d^{m_1}} \leq C \tag{3.7}$$

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 [60], for instance, satisfy this condition which can be viewed as an extension of the integrability condition (see [60] pages 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 $\Phi_\alpha^{(3)}$ be the test defined by (3.2) with the collection of kernels $\{K_m, m \in \mathcal{M}\}$ and $\{w_m, m \in \mathcal{M}\}$. We obtain from Theorem 4 that there exists $C(\alpha, \beta) > 0$ such that $\mathbb{P}_{f,g}(\Phi_\alpha^{(3)} = 0) \leq \beta$ if

$$\begin{aligned} &\|f - g\|^2 \\ &\quad \geq C(\alpha,\beta)\Bigg(\inf_{(m_1,m_2)\in\mathcal{M}}\Bigg\{\big\|(f-g) - k_{m_1,h_{m_2}} * (f-g)\big\|^2 \\ &\qquad\qquad + \frac{w_{(m_1,m_2)}}{n}\sqrt{\frac{\|f+g\|_\infty \|f+g\|_1 \|k_{m_1}\|^2}{2^{-dm_2}}}\Bigg\} \\ &\qquad\qquad\qquad + \frac{\|f+g\|_\infty}{n}\Bigg). \end{aligned} \tag{3.8}$$

Multiple kernel case—Example 4. 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 $\Phi_\alpha^{(4)}$ be the test defined by (3.2) with the collections $\{K_m, m \in \mathcal{M}\}$ and $\{w_m, m \in \mathcal{M}\}$. We deduce from Theorem 4 that there exists $C(\alpha, \beta) > 0$ such that $\mathbb{P}_{f,g}(\Phi_\alpha^{(4)} = 0) \le \beta$ if

$$\begin{aligned} \|f - g\|^2 & \\ \ge C(\alpha, \beta)\Bigg(\inf_{m_2\in\mathcal{M}_2}\Bigg\{ & \|(f - g) - k_{1,h_{m_2}} * (f - g)\|^2 \\ & + \frac{w_{(1,m_2)}}{n}\sqrt{\frac{\|f + g\|_\infty \|f + g\|_1 \|k_1\|^2}{\prod_{i=1}^d h_{m_2,i}}}\Bigg\} \\ & + \frac{\|f + g\|_\infty}{n}\Bigg). \end{aligned} \tag{3.9}$$

3.3. *Uniform separation rates over various classes of alternatives.* We here evaluate the uniform separation rates, defined by (1.1), of the multiple kernel testing procedures introduced above over several classes of alternatives based on Besov and weak Besov bodies when $\mathbb{X} = [0, 1]$, or Sobolev and anisotropic Besov–Nikol'skii balls when $\mathbb{X} = \mathbb{R}^d$.

3.3.1. *Uniform separation rates for Besov and weak Besov bodies.* In this section, we adapt to the present setting the results that we obtained in [20].

Given α in $(0, 1)$, let $\Phi_{\alpha/2}^{(1)}$ and $\Phi_{\alpha/2}^{(2)}$ be the tests defined in the *multiple kernel case—Example 1* and the *multiple kernel case—Example 2* (with α replaced by $\alpha/2$), and let $\Psi_\alpha = \max(\Phi_{\alpha/2}^{(1)}, \Phi_{\alpha/2}^{(2)})$.

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 $\mathbb{L}^2([0, 1])$ defined by (3.4). We define for $\delta > 0$, $R > 0$ the Besov body $\mathcal{B}_{2,\infty}^\delta(R)$ as follows:

$$\mathcal{B}_{2,\infty}^\delta(R) = \Bigg\{ s = \alpha_0\varphi_0 + \sum_{j\in\mathbb{N}}\sum_{k=0}^{2^j-1} \alpha_{(j,k)}\varphi_{(j,k)}/\alpha_0^2 \le R^2, \forall j \in \mathbb{N}, \sum_{k=0}^{2^j-1} \alpha_{(j,k)}^2 \le R^2 2^{-2j\delta} \Bigg\}.$$

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

$$\mathcal{W}_\gamma(R') = \Bigg\{ s = \alpha_0\varphi_0 + \sum_{j\in\mathbb{N}} \sum_{k=0}^{2^j-1} \alpha_{(j,k)}\varphi_{(j,k)} \\ \Big/ \forall t > 0, \alpha_0^2 \mathbf{1}_{\alpha_0^2 \le t} + \sum_{j\in\mathbb{N}} \sum_{k=0}^{2^j-1} \alpha_{(j,k)}^2 \mathbf{1}_{\alpha_{(j,k)}^2 \le t} \le R'^2 t^{2\gamma/(1+2\gamma)} \Bigg\}.$$

COROLLARY 1. *Assume that* $\ln\ln n \ge 1$, $2^{\bar{J}} \ge n^2$, *and* $\tilde{J} = +\infty$. *Then, for any* $\delta > 0$, $\gamma > 0$, $R, R', R'' > 0$, *if*

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

$\rho(\Psi_\alpha, \mathcal{B}_{\delta,\gamma,\infty}(R, R', R''), \beta)$, *defined by* (1.1), *is upper bounded by*:

(i) $C(\delta, \gamma, R, R', R'', \alpha, \beta)(\frac{\ln\ln n}{n})^{2\delta/(4\delta+1)}$ *if* $\delta \ge \gamma/2$,
(ii) $C(\delta, \gamma, R, R', R'', \alpha, \beta)(\frac{\ln n}{n})^{\gamma/(2\gamma+1)}$ *if* $\delta < \gamma/2$.

Comments.

(1) Lower bounds for the minimax separation rates over $\mathcal{B}_{\delta,\gamma,\infty}(R, R', R'')$ are also available, proving that 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 if $\delta \ge \max(\gamma/2, \gamma/(1 + 2\gamma))$ and exactly if $\delta < \gamma/2$ and $\gamma > 1/2$. In the other cases, the exact rate is unknown.

(2) Let us mention here that our classes of alternatives are not defined in the same way as in [8] in the classical two-sample problem for i.i.d. samples, since the classes of alternatives (f, g) of [8] are such that f and g both belong to a Besov ball. Here the smoothness condition is only required on the difference $(f - g)$. In particular, the functions f and g might be very irregular, but as long as their difference is smooth, the probability of second kind error of the test will be controlled.

3.3.2. *Uniform separation rates for Sobolev and anisotropic Nikol'skii–Besov balls*. Let $\Phi_\alpha^{(3)}$ be defined as in the *multiple kernel case—Example 3*, and let us introduce for $\delta > 0$ the Sobolev ball $\mathcal{S}_d^\delta(R)$ defined by

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

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

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

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

then

$$\rho(\Phi_\alpha^{(3)}, \mathcal{S}_d^\delta(R, R', R''), \beta) \leq C(\delta, \alpha, \beta, R, R', R'', d)\left(\frac{\ln\ln n}{n}\right)^{2\delta/(d+4\delta)}.$$

Comments. From [49], we know that, in the density model, the minimax adaptive estimation rate over $\mathcal{S}_d^\delta(R)$ is of order $n^{-\delta/(d+2\delta)}$ 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 $\Phi_\alpha^{(3)}$ consists of 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 $n^{-2\delta/(d+4\delta)}$ 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 $d > 1$.

Let $\Phi_\alpha^{(4)}$ be the test defined in the *multiple kernel case—Example 4*. Let $\Delta = (\Delta_1, \dots, \Delta_d)$, where for every $i = 1, \dots, d$, Δ_i is a positive integer. 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 $j = 1, \dots, \Delta_i$.

For $\delta = (\delta_1, \dots, \delta_d) \in \prod_{i=1}^d (0, \Delta_i]$ and $R > 0$, we consider the anisotropic Nikol'skii–Besov ball $\mathcal{N}_{2,d}^\delta(R)$ defined by

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

COROLLARY 3. *Assume that* $\ln\ln n \geq 1$. *For any* $\delta = (\delta_1, \dots, \delta_d)$ *in* $\prod_{i=1}^d (0, \Delta_i]$ *and* $R, R', R'' > 0$, *if*

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

then, for $1/\bar{\delta} = \sum_{i=1}^d 1/\delta_i$,

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

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 [24] 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 [30] over anisotropic periodic Sobolev balls, but in the Gaussian white noise model, are of the same order as the upper bounds obtained here.

4. Proofs.

4.1. *Proof of Proposition* 2. Recall that $q^{\alpha}_{K,1-\beta/2}$ denotes the $1-\beta/2$ quantile of $q^{(N)}_{K,1-\alpha}$, which is the $(1-\alpha)$ quantile of $\hat{T}^{\varepsilon}_K$ conditionally on N. We here want to find a condition on $\mathcal{E}_K = \mathbb{E}_{f,g}[\hat{T}_K]$, ensuring that

$$\mathbb{P}_{f,g}(\hat{T}_K \le q^{\alpha}_{K,1-\beta/2}) \le \beta/2.$$

From Markov's inequality, we have that for any $x > 0$,

$$\mathbb{P}_{f,g}(|-\hat{T}_K + \mathcal{E}_K| \ge x) \le \frac{\operatorname{Var}(\hat{T}_K)}{x^2}.$$

This implies that

$$\mathbb{P}_{f,g}\left(|-\hat{T}_K + \mathcal{E}_K| \ge \sqrt{\frac{2\operatorname{Var}(\hat{T}_K)}{\beta}}\right) \le \frac{\beta}{2}. \tag{4.1}$$

Therefore, if $\mathcal{E}_K > \sqrt{\frac{2\operatorname{Var}(\hat{T}_K)}{\beta}} + q^{\alpha}_{K,1-\beta/2}$, then $\mathbb{P}_{f,g}(\hat{T}_K \le q^{\alpha}_{K,1-\beta/2}) \le \beta/2$, so $\mathbb{P}_{f,g}(\Phi_{K,\alpha} = 0) \le \beta$.

Let us compute $\operatorname{Var}(\hat{T}_K) = \mathbb{E}_{f,g}[\hat{T}_K^2] - \mathcal{E}_K^2$. 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

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

by using (2.6),

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

$$\times \frac{f-g}{f+g}(v)\,dN_x\,dN_y\,dN_u\,dN_v\Bigg]$$
$$+ 4\mathbb{E}_{f,g}\bigg[\int_{\mathbb{X}^{[3]}} K(x,y)K(x,u)\frac{f-g}{f+g}(y)\frac{f-g}{f+g}(u)\,dN_x\,dN_y\,dN_u\bigg]$$
$$+ 2\mathbb{E}_{f,g}\bigg[\int_{\mathbb{X}^{[2]}} K^2(x,y)\,dN_x\,dN_y\bigg].$$

Now, from Lemma 5.4 III in [13] on factorial moments measures applied to Poisson processes, we deduce that

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

Note that the three above integrals are finite, thanks to Assumptions 1, 2 and 3. We finally obtain that $\mathbb{E}_{f,g}[\hat{T}_K^2] = \mathcal{E}_K^2 + 4n^3 A_K + 2n^2 B_K$, hence

$$\operatorname{Var}(\hat{T}_K) = 4n^3 A_K + 2n^2 B_K.$$

Let us now give a sharp upper bound for $q^{\alpha}_{K,1-\beta/2}$. Reasoning conditionally on N, we recognize in $\hat{T}_K^{\varepsilon}$ a homogeneous Rademacher chaos, as defined by de la Peña and Giné [14], 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 [14] states that there exists some absolute constant $\kappa > 0$ such that if $\sigma^2 = \mathbb{E}[X^2] = \sum_{i\neq i'} x^2_{i,i'}$, then

$$\mathbb{E}\big[\exp\big(|X|/(\kappa\sigma)\big)\big] \le 2.$$

Hence by Markov's inequality,

$$\mathbb{P}\big(|X| \ge \kappa\sigma\ln(2/\alpha)\big) \le \alpha.$$

Note that one could find more precise constants with the results of [41].

Applying this result to $\hat{T}_K^{\varepsilon}$ with $\sigma^2 = \sum_{x\neq x'\in N} K^2(x,x')$ leads to

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

Hence $q^{\alpha}_{K,1-\beta/2}$ 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 [13], we obtain that

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

and

$$q^{\alpha}_{K,1-\beta/2} \leq \kappa \ln(2/\alpha) n \sqrt{\frac{2B_K}{\beta}}.$$

4.2. *Proof of Theorem* 1. First notice that for every $r > 0$, and every kernel function K satisfying Assumption 3,

$$\mathcal{E}_K = \frac{n^2 r}{2}\left(\|f-g\|^2 + r^{-2}\|K[f-g]\|^2 - \|(f-g) - r^{-1}K[f-g]\|^2\right).$$

With the notation of Proposition 2, let C_K be any upper bound for B_K. Since $A_K \leq \|K[f-g]\|^2\|f+g\|_\infty$, from Proposition 2, we deduce that $\mathbb{P}_{f,g}(\Phi_{K,\alpha} = 0) \leq \beta$ if

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

By using the elementary inequality $2ab \leq a^2 + b^2$ with $a = \|K[f-g]\|/r$ and $b = 2\sqrt{2}\sqrt{\|f+g\|_\infty/(n\beta)}$ on the right-hand side of the above condition, this condition can be replaced by

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

We can even add an infimum over r on the right-hand side of the condition, since r can be arbitrarily chosen. Let us now justify our choices for C_K.

Projection kernel case. We consider an orthonormal basis $\{\varphi_\lambda, \lambda \in \Lambda\}$ of a subspace S of $\mathbb{L}^2(\mathbb{X}, d\nu)$ and $K(x,x') = \sum_{\lambda\in\Lambda}\varphi_\lambda(x)\varphi_\lambda(x')$. When the dimension of S is finite, equal to D,

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

When the dimension of S is infinite,

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

where we have used assumption (2.12) to invert the sum and the integral. Hence we have, by orthogonality, and since by assumption (2.11), $\sum_{\lambda \in \Lambda} \varphi_\lambda^2(x) \le D$,

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

Approximation kernel case. 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 $k_h(x_1, \dots, x_d) = \frac{1}{\prod_{i=1}^d h_i} k(\frac{x_1}{h_1}, \dots, \frac{x_d}{h_d})$. In this case,

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

This ends the proof of Theorem 1.

4.3. *Proof of Theorem* 2. We first recall that when K is chosen as in the *reproducing kernel case*, under the assumptions of Theorem 2, $\mathcal{E}_K = n^2 \|m_f - m_g\|_{\mathcal{H}_K}^2$; see Section 2.1.

Since $A_K = \int_{\mathbb{X}} \langle \int_{\mathbb{X}} \theta(x)(f-g)(x)\, d\nu_x, \theta(y) \rangle_{\mathcal{H}_K}^2 (f+g)(y)\, d\nu_y$, by the Cauchy–Schwarz inequality for the norm $\|\cdot\|_{\mathcal{H}_K}$ in the RKHS, we obtain

$$A_K \le \int_{\mathbb{X}} \Big\| \int_{\mathbb{X}} \theta(x)(f-g)(x)\, d\nu_x \Big\|_{\mathcal{H}_K}^2 \|\theta(y)\|_{\mathcal{H}_K}^2 (f+g)(y)\, d\nu_y.$$

Now, since for every y in $\mathbb{X}$, $\|\theta(y)\|_{\mathcal{H}_K}^2 = K(y, y) = \kappa_0$,

$$\begin{aligned} A_K &\le \kappa_0 \left\| \int_{\mathbb{X}} \theta(x)(f-g)(x)\, d\nu_x \right\|_{\mathcal{H}_K}^2 \|f+g\|_1 \\ &\le \kappa_0 \|m_f - m_g\|_{\mathcal{H}_K}^2 \|f+g\|_1. \end{aligned}$$

This leads to

$$\begin{aligned} 2n\sqrt{\frac{2nA_K}{\beta}} &\le 2n\sqrt{\frac{2\kappa_0 n\|f+g\|_1}{\beta}} \|m_f - m_g\|_{\mathcal{H}_K} \\ &\le \frac{n^2}{2}\|m_f - m_g\|_{\mathcal{H}_K}^2 + 4\frac{\kappa_0 n\|f+g\|_1}{\beta}. \end{aligned}$$

Finally, noting that $B_K \le \kappa_0^2 \|f+g\|_1^2$ and that by the assumption $\|f+g\|_1 = 2$, we obtain the desired result from Proposition 2 and obvious calculations.

4.4. *Proof of Proposition* 3. First let us rewrite here a result due to Romano and Wolf [52].

LEMMA 1. *Let $Y_0, \dots, Y_B$ be $B+1$ exchangeable variables. Then for all $u \in [0, 1]$,*

$$\mathbb{P}\left(\frac{1}{B+1}\left(1 + \sum_{i=1}^{B} \mathbf{1}_{Y_i \ge Y_0}\right) \le u\right) \le u.$$

Assume that (H_0) is satisfied. Conditionally on N, the observed statistic $\hat{T}_K^{\varepsilon^0} := \hat{T}_K$ has the same distribution and is independent of the $\hat{T}_K^{\varepsilon^b}$'s for $b = 1, \dots, B$. Therefore the variables $\hat{T}_K^{\varepsilon^b}$'s for $b = 0, \dots, B$ are exchangeable variables given N. Hence applying Lemma 1, we obtain

$$\begin{aligned} \mathbb{P}(\hat{\Phi}_\alpha^K = 1|N) &= \mathbb{P}(\hat{T}_K > \hat{q}_{K,1-\alpha}^{(N)}|N) \\ &= \mathbb{P}\left(\sum_{b=1}^{B} \mathbf{1}_{\hat{T}_K^{\varepsilon^b} \ge \hat{T}_K^{\varepsilon^0}} \le \lfloor B\alpha \rfloor \Big| N\right) \\ &= \mathbb{P}\left(\frac{1}{B+1}\left(1 + \sum_{b=1}^{B} \mathbf{1}_{\hat{T}_K^{\varepsilon^b} \ge \hat{T}_K^{\varepsilon^0}}\right) \le \frac{\lfloor B\alpha \rfloor + 1}{B+1} \Big| N\right) \\ &\le \frac{\lfloor B\alpha \rfloor + 1}{B+1}. \end{aligned}$$

4.5. *Proof of Proposition 4.* Let $t = q^{\alpha_B}_{K,1-\beta_B/2}$. By definition of $\hat{q}^{(N)}_{K,1-\alpha}$,

$$\mathbb{P}_{f,g}\big(\hat{q}^{(N)}_{K,1-\alpha} > t\big) = \mathbb{P}_{f,g}\big(F_{K,B}(t) < 1-\alpha\big) = \mathbb{P}_{f,g}\left(\sum_{b=1}^{B} \mathbf{1}_{\hat{T}_K^{\varepsilon^b} \le t} < B(1-\alpha)\right).$$

We have

$$\begin{aligned}
&\mathbb{P}_{f,g}\left(\sum_{b=1}^{B} \mathbf{1}_{\hat{T}_K^{\varepsilon^b} \le t} < B(1-\alpha),\, F_K(t) \ge 1-\alpha_B\right)\\
&\quad \le \mathbb{P}_{f,g}\left(\sum_{b=1}^{B} \big(\mathbf{1}_{\hat{T}_K^{\varepsilon^b} \le t} - F_K(t)\big) < B(1-\alpha) - B(1-\alpha_B)\right).
\end{aligned}$$

So we can decompose as follows:

$$\begin{aligned}
\mathbb{P}_{f,g}\big(\hat{q}^{(N)}_{K,1-\alpha} > t\big) &\le \mathbb{P}_{f,g}\big(F_K(t) < 1-\alpha_B\big)\\
&\quad + \mathbb{P}_{f,g}\left(\sum_{b=1}^{B} \big(\mathbf{1}_{\hat{T}_K^{\varepsilon^b} \le t} - F_K(t)\big) < -B\sqrt{\frac{\ln B}{2B}}\right).
\end{aligned}$$

By Hoeffding's inequality applied to the second probability given N, we obtain

$$\mathbb{P}_{f,g}\big(\hat{q}^{(N)}_{K,1-\alpha} > t\big) \le \mathbb{P}_{f,g}\big(F_K(t) < 1-\alpha_B\big) + \frac{1}{B}.$$

But by definition of t, this becomes

$$\mathbb{P}_{f,g}\big(\hat{q}^{(N)}_{K,1-\alpha} > t\big) \le \frac{\beta}{2}.$$

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

$$\begin{aligned}
\mathbb{P}_{f,g}\big(\hat{T}_K \le \hat{q}^{(N)}_{K,1-\alpha}\big) &\le \mathbb{P}_{f,g}\big(\hat{T}_K \le \hat{q}^{(N)}_{K,1-\alpha}, \hat{q}^{(N)}_{K,1-\alpha} \le t\big) + \mathbb{P}_{f,g}\big(\hat{q}^{(N)}_{K,1-\alpha} > t\big)\\
&\le \mathbb{P}_{f,g}\big(\hat{T}_K \le t\big) + \beta/2.
\end{aligned}$$

We deduce from (4.1) that if

$$\mathcal{E}_K > \sqrt{\frac{2\operatorname{Var}(\hat{T}_K)}{\beta}} + t,$$

then $\mathbb{P}_{f,g}(\hat{T}_K \le t) \le \beta/2$, and $\mathbb{P}_{f,g}(\hat{\Phi}_{K,\alpha} = 0) \le \beta$. An upper bound for t is finally derived from (2.9), which concludes the proof.

Acknowledgments. We are grateful to the referees, whose discussion and comments allowed us to improve some of our results and to clarify some important points.

SUPPLEMENTARY MATERIAL

Simulation study and additional proofs (DOI: 10.1214/13-AOS1114SUPP; .pdf). A simulation study, the proofs of Proposition 1, Theorems 3 and 4, and of Corollaries 1, 2 and 3 are given in the supplementary material [21].

REFERENCES

[1] ARCONES, M. A. and GINÉ, E. (1992). On the bootstrap of U and V statistics. *Ann. Statist.* **20** 655–674. MR1165586

[2] BACH, F. R. (2008). Consistency of the group lasso and multiple kernel learning. *J. Mach. Learn. Res.* **9** 1179–1225. MR2417268

[3] BARAUD, Y. (2002). Non-asymptotic minimax rates of testing in signal detection. *Bernoulli* **8** 577–606. MR1935648

[4] BARAUD, Y., HUET, S. and LAURENT, B. (2003). Adaptive tests of linear hypotheses by model selection. *Ann. Statist.* **31** 225–251. MR1962505

[5] BARTLETT, P. L., BOUCHERON, S. and LUGOSI, G. (2002). Model selection and error estimation. *Machine Learning* **48** 85–113.

[6] BOVETT, J. M. and SAW, J. G. (1980). On comparing two Poisson intensity functions. *Comm. Statist. Theory Methods* **9** 943–948. MR0574314

[7] BRETAGNOLLE, J. (1983). Lois limites du bootstrap de certaines fonctionnelles. *Ann. Inst. H. Poincaré Sect. B* (*N.S.*) **19** 281–296. MR0725561

[8] BUTUCEA, C. and TRIBOULEY, K. (2006). Nonparametric homogeneity tests. *J. Statist. Plann. Inference* **136** 597–639. MR2181971

[9] CHAPELLE, O., VAPNIK, V., BOUSQUET, O. and MUKHERJEE, S. (2002). Choosing multiple parameters for support vector machines. *Machine Learing* **46** 131–159.

[10] CHIU, S. N. (2010). Parametric bootstrap and approximate tests for two Poisson variates. *J. Stat. Comput. Simul.* **80** 263–271. MR2752162

[11] CHIU, S. N. and WANG, L. (2009). Homogeneity tests for several Poisson populations. *Comput. Statist. Data Anal.* **53** 4266–4278. MR2744323

[12] COX, D. R. (1953). Some simple approximate tests for Poisson variates. *Biometrika* **40** 354–360. MR0058177

[13] DALEY, D. J. and VERE-JONES, D. (2008). *An Introduction to the Theory of Point Processes. Vol. II*: *General Theory and Structure*, 2nd ed. Springer, New York. MR2371524

[14] DE LA PEÑA, V. H. and GINÉ, E. (1999). *Decoupling*: *From Dependence to Independence*. Springer, New York. MR1666908

[15] DEHLING, H. and MIKOSCH, T. (1994). Random quadratic forms and the bootstrap for U-statistics. *J. Multivariate Anal.* **51** 392–413. MR1321305

[16] DESHPANDE, J. V., 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** 237–251. MR1721560

[17] DESHPANDE, J. V. and SENGUPTA, D. (1995). Testing the hypothesis of proportional hazards in two populations. *Biometrika* **82** 252–261.

[18] EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26. MR0515681

[19] FISHER, R. A. (1935). *The Design of Experiments*. Oliver & Boyd, Edinburgh.

[20] FROMONT, M., LAURENT, B. and REYNAUD-BOURET, P. (2011). Adaptive tests of homogeneity for a Poisson process. *Ann. Inst. Henri Poincaré Probab. Stat.* **47** 176–213. MR2779402

[21] FROMONT, M., LAURENT, B. and REYNAUD-BOURET, P. (2013). Supplement to "The two-sample problem for Poisson processes: Adaptive tests with a non-asymptotic wild bootstrap approach." DOI:10.1214/13-AOS1114SUPP.

[22] GAIL, M. (1974). Computations for designing comparative Poisson trials. *Biometrics* **30** 231–237.

[23] GINÉ M., E. (1975). Invariant tests for uniformity on compact Riemannian manifolds based on Sobolev norms. *Ann. Statist.* **3** 1243–1266. MR0388663

[24] GOLDENSHLUGER, A. and LEPSKI, O. (2011). Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39** 1608–1632. MR2850214

[25] GRETTON, A., BORGWARDT, K. M., RASCH, M. J., SCHÖLKOPF, B. and SMOLA, A. (2008). A kernel method for the two-sample problem. *J. Mach. Learn. Res.* **1** 1–10.

[26] GRETTON, A., SRIPERUMBUDUR, B. K., SEJDINOVIC, D., STRATHMANN, H., BALAKRISHNAN, S., PONTIL, M. and FUKUMIZU, K. (2012). Optimal kernel choice for large-scale two-sample tests. In *Advances in Neural Information Processing Systems (NIPS) 25* 1214–1222. Available at http://books.nips.cc/papers/files/nips25/NIPS2012_0592.pdf.

[27] HOEFFDING, W. (1952). The large-sample power of tests based on permutations of observations. *Ann. Math. Statistics* **23** 169–192. MR0057521

[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** 599–631. MR1828537

[29] HUŠKOVÁ, M. and JANSSEN, P. (1993). Consistency of the generalized bootstrap for degenerate U-statistics. *Ann. Statist.* **21** 1811–1823. MR1245770

[30] INGSTER, Y. and STEPANOVA, N. (2011). Estimation and detection of functions from anisotropic Sobolev classes. *Electron. J. Stat.* **5** 484–506. MR2813552

[31] INGSTER, Y. I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives. I, II, III. *Math. Methods Statist.* **2** 85–114, 171–189, 249–268.

[32] INGSTER, Y. I. (2000). Adaptive chi-square tests. *J.Math. Sci.* **99** 1110–1119.

[33] INGSTER, Y. I. and KUTOYANTS, Y. A. (2007). Nonparametric hypothesis testing for intensity of the Poisson process. *Math. Methods Statist.* **16** 217–245. MR2356819

[34] JANSSEN, A. and PAULS, T. (2003). How do bootstrap and permutation tests work? *Ann. Statist.* **31** 768–806. MR1994730

[35] JANSSEN, P. (1994). Weighted bootstrapping of U-statistics. *J. Statist. Plann. Inference* **38** 31–41. MR1256846

[36] KOLTCHINSKII, V. (2001). Rademacher penalties and structural risk minimization. *IEEE Trans. Inform. Theory* **47** 1902–1914. MR1842526

[37] KOLTCHINSKII, V. (2006). Local Rademacher complexities and oracle inequalities in risk minimization. *Ann. Statist.* **34** 2593–2656. MR2329442

[38] KOLTCHINSKII, V. and YUAN, M. (2010). Sparsity in multiple kernel learning. *Ann. Statist.* **38** 3660–3695. MR2766864

[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] LANCKRIET, G. R. G., CRISTIANINI, N., BARTLETT, P., EL GHAOUI, L. and JORDAN, M. I. (2003/04). Learning the kernel matrix with semidefinite programming. *J. Mach. Learn. Res.* **5** 27–72. MR2247973

[41] LATAŁA, R. (1999). Tail and moment estimates for some types of chaos. *Studia Math.* **135** 39–53. MR1686370

[42] LOUNICI, K. and NICKL, R. (2011). Global uniform risk bounds for wavelet deconvolution estimators. *Ann. Statist.* **39** 201–231. MR2797844

[43] MAMMEN, E. (1992). Bootstrap, wild bootstrap, and asymptotic normality. *Probab. Theory Related Fields* **93** 439–455. MR1183886

[44] MICCHELLI, C. A. and PONTIL, M. (2005). Learning the kernel function via regularization. *J. Mach. Learn. Res.* **6** 1099–1125. MR2249850

[45] NG, H. K. T., GU, K. and TANG, M. L. (2007). A comparative study of tests for the difference of two Poisson means. *Comput. Statist. Data Anal.* **51** 3085–3099. MR2345627
[46] PRÆSTGAARD, J. and WELLNER, J. A. (1993). Exchangeably weighted bootstraps of the general empirical process. *Ann. Probab.* **21** 2053–2086. MR1245301
[47] PRÆSTGAARD, J. T. (1995). Permutation and bootstrap Kolmogorov–Smirnov tests for the equality of two distributions. *Scand. J. Stat.* **22** 305–322. MR1363215
[48] 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
[49] RIGOLLET, P. and TSYBAKOV, A. B. (2007). Linear and convex aggregation of density estimators. *Math. Methods Statist.* **16** 260–280. MR2356821
[50] ROMANO, J. P. (1988). A bootstrap revival of some nonparametric distance tests. *J. Amer. Statist. Assoc.* **83** 698–708. MR0963797
[51] ROMANO, J. P. (1989). Bootstrap and randomization tests of some nonparametric hypotheses. *Ann. Statist.* **17** 141–159. MR0981441
[52] ROMANO, J. P. and WOLF, M. (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *J. Amer. Statist. Assoc.* **100** 94–108. MR2156821
[53] RUBIN, D. B. (1981). The Bayesian bootstrap. *Ann. Statist.* **9** 130–134. MR0600538
[54] SCHÖLKOPF, B. and SMOLA, A. J. (2002). *Learning with Kernels*. MIT Press, Cambridge, MA.
[55] SHIUE, W. and BAIN, L. J. (1982). Experiment size and power comparisons for two-sample Poisson tests. *J. R. Stat. Soc. Ser. C Appl. Stat.* **31** 130–134.
[56] SPOKOINY, V. G. (1996). Adaptive hypothesis testing using wavelets. *Ann. Statist.* **24** 2477–2498. MR1425962
[57] SPOKOINY, V. G. (1998). Adaptive and spatially adaptive testing of a nonparametric hypothesis. *Math. Methods Statist.* **7** 245–273. MR1651777
[58] SRIPERUMBUDUR, B. K., FUKUMIZU, K., GRETTON, A., LANCKRIET, G. R. G. and SCHÖLKOPF, B. (2009). Kernel choice and classifiability for RKHS embeddings of probability distributions. In *Advances in Neural Information Processing Systems (NIPS) 22* 1750–1758. Available at http://books.nips.cc/papers/files/nips22/NIPS2009_0893.pdf
[59] SRIPERUMBUDUR, B. K., FUKUMIZU, K. and LANCKRIET, G. R. G. (2011). Universality, characteristic kernels and RKHS embedding of measures. *J. Mach. Learn. Res.* **12** 2389–2410. MR2825431
[60] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer, New York. MR2724359
[61] WELLNER, J. A. (1979). Permutation tests for directional data. *Ann. Statist.* **7** 929–943. MR0536498

M. FROMONT
UNIVERSITÉ HAUTE BRETAGNE
PLACE DU RECTEUR H. LE MOAL
CS 24307
35043 RENNES CEDEX
FRANCE
E-MAIL: magalie.fromont@univ-rennes2.fr

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

P. 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