



Adaptive test of independence based on HSIC measures

Mélanie Albert, Béatrice Laurent, Amandine Marrel, Anouar Meynaoui

► To cite this version:

Mélanie Albert, Béatrice Laurent, Amandine Marrel, Anouar Meynaoui. Adaptive test of independence based on HSIC measures. 2021. hal-02020084v4

HAL Id: hal-02020084

<https://hal.science/hal-02020084v4>

Preprint submitted on 8 Jan 2021

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.

ADAPTIVE TEST OF INDEPENDENCE BASED ON HSIC MEASURES.

Mélanie Albert^{*,1}, Béatrice Laurent^{†,1}, Amandine Marrel^{‡,2}, and Anouar Meynaoui^{§,1,2}

¹Institut de Mathématiques de Toulouse ; UMR5219, Université de Toulouse ; CNRS, INSA, F-31077 Toulouse, France.

²CEA, DEN, DER, F-13108 Saint-Paul-lez-Durance, France.

Abstract: The Hilbert-Schmidt Independence Criterion (HSIC) is a dependence measure based on reproducing kernel Hilbert spaces that is widely used to test independence between two random vectors. Remains the delicate choice of the kernel. In this work, we develop a new HSIC-based aggregated procedure which avoids such a kernel choice, and provide theoretical guarantees for this procedure. To achieve this, on the one hand, we introduce non-asymptotic single tests based on Gaussian kernels with a given bandwidth, which are of prescribed level. Then, we aggregate several single tests with different bandwidths, and prove sharp upper bounds for the uniform separation rate of the aggregated procedure over Sobolev balls. On the other hand, we provide a lower bound for the non-asymptotic minimax separation rate of testing over Sobolev balls, and deduce that the aggregated procedure is adaptive in the minimax sense over such regularity spaces. Finally, from a practical point of view, we perform numerical studies in order to assess the efficiency of our aggregated procedure and compare it to existing tests in the literature.

Mathematics Subject Classification: Primary: 62G10; secondary: 62G09

Keywords: nonparametric test of independence, Hilbert-Schmidt Independence Criterion, permutation methods, uniform separation rates, aggregated tests, non-asymptotic minimax and adaptive tests

1 Introduction to independence testing

Many nonparametric approaches to test independence between two continuous random vectors have been explored in the last few decades. Among them, [Hoeffding, 1948] introduces a test based on the difference between the joint distribution function and the product of the marginal distribution functions. This test has good properties in the asymptotic framework since it is consistent. Yet, it only applies to univariate random variables. Lately, [Weihs et al., 2018] extend Hoeffding's test to the case of multivariate random variables, but still in an asymptotic framework. Another classical method for testing independence is based on comparing the joint density and the product of the marginal densities [Rosenblatt, 1975, Ahmad and Li, 1997]. For this, an intermediate step is to estimate these densities using, e.g., the kernel-based method of Parzen-Rosenblatt [Parzen, 1962]. More recently, many approaches based on reproducing kernel Hilbert spaces (RKHS) have been developed (see [Aronszajn, 1950] for more details). One of the first RKHS measures is the kernel canonical correlation (KCC) [Bach and Jordan, 2002]. Yet, the estimation of the KCC is not practical since it requires an extra regularization which has to be

*Electronic address: melisande.albert@insa-toulouse.fr

†Electronic address: beatrice.laurent@insa-toulouse.fr

‡Electronic address: amandine.marrel@cea.fr

§Electronic address: anouar.meynaoui@gmail.com

adjusted. Other dependence measures, easier to estimate have been studied later. For instance, the kernel mutual information (KMI) [Gretton et al., 2003, Gretton et al., 2005b] and the constrained covariance (COCO) [Gretton et al., 2005c, Gretton et al., 2005b] are widely used, since they are relatively easy to interpret and to implement. Finally, one of the most interesting kernel dependence measure is the Hilbert-Schmidt independence criterion (HSIC) [Gretton et al., 2005a]. The HSIC has a very low computational cost and seems to numerically outperform all previous RKHS measures [Gretton et al., 2005a]. A first independence test based on the HSIC is developed using large deviation inequalities [Gretton et al., 2005a]. Then, other tests are constructed in [Gretton et al., 2008, Li and Yuan, 2019] using an approximation of the null distribution of the HSIC estimator either by an asymptotic Gamma distribution or by a permutation approach. A generalization to joint and mutual independence testing is presented in [Pfister et al., 2018]. We also mention the RKHS-based test [Póczos et al., 2012], using the copula-based kernel dependency measure. Yet, this test is more conservative than the test of [Gretton et al., 2008], since it is based on large deviation inequalities. Lately, based on characteristic functions, [Székely et al., 2007] introduce the distance covariance which has good properties and can be used in high dimensional frameworks [Székely and Rizzo, 2013, Yao et al., 2018]. Furthermore, it has been shown that the distance covariance coincides with the HSIC for a specific choice of kernels. Other tests have emerged based for instance on a sample space partitioning [Heller et al., 2016] or based on binary expansion [Zhang, 2019] and very recently extended to any arbitrary dimension [Lee et al., 2019]. Finally, the authors of [Berrett and Samworth, 2019] introduce a new test based on nearest neighbour methods and kernel mutual information which seems to achieve comparable results with the classical tests based on HSIC. In this paper, we focus on HSIC measures to test independence.

1.1 Adaptive independence tests

To study the non-asymptotic performances of testing, we consider the *uniform separation rate* as defined in [Baraud, 2002]. For any α -level test Δ_α with values in $\{0, 1\}$, which rejects independence when $\Delta_\alpha = 1$, the uniform separation rate $\rho(\Delta_\alpha, \mathcal{C}_\delta, \beta)$ of Δ_α , over a class \mathcal{C}_δ of regular alternatives f (such that the difference between the density f and the product of its marginals $f_1 \otimes f_2$ satisfies smoothness assumptions), with respect to (w.r.t.) the \mathbb{L}_2 -norm, is defined for all β in $(0, 1)$ by

$$\rho(\Delta_\alpha, \mathcal{C}_\delta, \beta) = \inf \left\{ \rho > 0; \sup_{f \in \mathcal{F}_\rho(\mathcal{C}_\delta)} P_f(\Delta_\alpha = 0) \leq \beta \right\}, \quad (1.1)$$

where $\mathcal{F}_\rho(\mathcal{C}_\delta) = \{f; f - f_1 \otimes f_2 \in \mathcal{C}_\delta, \|f - f_1 \otimes f_2\|_2 > \rho\}$ and $\|\cdot\|_2$ designates the usual \mathbb{L}_2 -norm. The uniform separation rate is then the smallest value in the sense of the \mathbb{L}_2 -norm of $f - f_1 \otimes f_2$ allowing to control the second kind error of the test by β . A test of level α having the optimal performances, should then have a uniform separation rate as small as possible over \mathcal{C}_δ . To quantify this, let us define, as in [Baraud, 2002], the *non-asymptotic minimax rate of testing* by

$$\rho(\mathcal{C}_\delta, \alpha, \beta) = \inf_{\Delta_\alpha} \rho(\Delta_\alpha, \mathcal{C}_\delta, \beta), \quad (1.2)$$

where the infimum is taken over all α -level tests. If the uniform separation rate of a test is upper bounded up to a constant by the non-asymptotic minimax rate of testing, then this test is said to be *optimal in the minimax sense*. The problem of non-asymptotic minimax rate of testing was raised in many papers over the past years. Among them, we

mention for example [Ingster and Suslina, 1998, Laurent et al., 2012] for minimax signal detection testing. Concerning independence testing, optimality in the minimax sense defined as above is closely related to *the asymptotic minimax rate* as introduced in the notable works of Ingster [Ingster, 1989, Ingster, 1993b], of Yodé [Yodé, 2004, Yodé, 2011] or very recently of [Li and Yuan, 2019]. Lately, [Berrett and Samworth, 2019] study upper bounds w.r.t. mutual information, and [Ramdas et al., 2016] obtain minimax lower bounds for linear independence testing. In the non-asymptotic framework considered in this paper, [Albert, 2015] obtains upper bounds w.r.t. the \mathbb{L}_2 distance over weak Besov spaces. Concurrent with our work, and independently, [Berrett et al., 2020] and [Kim et al., 2020] also obtain minimax separation rates for independence tests based on permuted U -statistics.

Furthermore, beyond the problem of minimax optimality, the straightforward practical construction of a minimax test usually depends on the unknown smoothness parameter δ of the regularity space \mathcal{C}_δ . The objective is then to construct a minimax test which does not need any smoothness assumption to be implemented. These tests are called *minimax adaptive*. The problem of adaptivity has received a great attention in the literature. We mention for instance the works of [Baraud et al., 2003] for linear regression model testing with Gaussian noise and of [Ingster, 2000, Fromont et al., 2006, Balakrishnan et al., 2019] for goodness-of-fit testing. The authors of [Fromont et al., 2013] consider an interesting approach for testing the equality of two Poisson processes intensities, which consists in aggregating several single kernel-based tests, and prove that it is adaptive over several regularity spaces. This paper lies in the lineage of these works.

1.2 Mathematical framework and notation

In this work, we study the problem of testing the independence between two continuous real random vectors $X = (X^{(1)}, \dots, X^{(p)})$ and $Y = (Y^{(1)}, \dots, Y^{(q)})$. The couple (X, Y) is assumed to have a joint density f w.r.t. Lebesgue measure on $\mathbb{R}^p \times \mathbb{R}^q$, with marginal density functions f_1 and f_2 . To avoid any misunderstanding, let us highlight that f_1 and f_2 are assumed to be unknown and are not fixed *a priori*. We denote by $f_1 \otimes f_2 : (x, y) \in \mathbb{R}^p \times \mathbb{R}^q \mapsto f_1(x)f_2(y)$ the product of the marginal densities. We also assume that we observe a n -sample $(X_1, Y_1), \dots, (X_n, Y_n)$ of independent and identically distributed (i.i.d.) random variables with common density f . The probability measure associated to this n -sample is denoted P_f . By analogy, $P_{f_1 \otimes f_2}$ designates the probability measure associated to a n -sample with common density $f_1 \otimes f_2$.

We address here the question of testing the null hypothesis (\mathcal{H}_0) : “ X and Y are independent” against the alternative (\mathcal{H}_1) : “ X and Y are dependent”. That is equivalent to test

$$(\mathcal{H}_0): “f = f_1 \otimes f_2” \quad \text{against} \quad (\mathcal{H}_1): “f \neq f_1 \otimes f_2”.$$

Throughout this document, we consider the following assumption:

$$\begin{aligned} \mathcal{A}_1 : & \text{ the density } f, \text{ and its marginal densities } f_1 \text{ and } f_2 \text{ are bounded,} \\ & \text{ and denote } M_f = \max \{ \|f\|_\infty, \|f_1\|_\infty, \|f_2\|_\infty \}. \end{aligned}$$

Moreover, the generic notation $C(a, b, \dots)$ denotes a positive constant depending only on its arguments (a, b, \dots) and that may vary from line to line. Finally, the dimensions p and q are assumed to be fixed, and do not depend on the sample size.

1.3 Review on HSIC measures

The definition of the HSIC is derived from the notion of cross-covariance operator [Baker, 1973, Fukumizu et al., 2004], which can be seen as a generalization of the classical covariance, measuring many forms of dependence between X and Y (not only linear ones). For this, [Gretton et al., 2005a] associate to X an RKHS \mathcal{F} composed of functions mapping from \mathbb{R}^p to \mathbb{R} (\mathcal{F} is a set of transformations for X), and characterized by a scalar product $\langle \cdot, \cdot \rangle_{\mathcal{F}}$. The same operation is carried out for Y , considering an RKHS denoted \mathcal{G} and a scalar product $\langle \cdot, \cdot \rangle_{\mathcal{G}}$. The cross-covariance operator $C_{X,Y}$ associated to \mathcal{F} and \mathcal{G} is the operator mapping from \mathcal{G} to \mathcal{F} and verifying for all $(F, G) \in \mathcal{F} \times \mathcal{G}$,

$$\langle F, C_{X,Y}(G) \rangle_{\mathcal{F}} = \text{Cov}(F(X), G(Y)).$$

Designating by $(u_i)_i$ and $(v_j)_j$ respectively orthonormal bases of \mathcal{F} and \mathcal{G} , the HSIC between X and Y is the square Hilbert-Schmidt norm of the operator $C_{X,Y}$ defined as in [Gretton et al., 2005a] by

$$\text{HSIC}(X, Y) = \|C_{X,Y}\|_{\text{HS}}^2 = \sum_{i,j} \langle u_i, C_{X,Y}(v_j) \rangle_{\mathcal{F}}^2 = \sum_{i,j} \text{Cov}(u_i(X), v_j(Y))^2.$$

The fundamental idea behind this definition is that $\text{HSIC}(X, Y)$ equals zero if and only if $\text{Cov}(F(X), G(Y)) = 0$ for all (F, G) in $\mathcal{F} \times \mathcal{G}$. Furthermore, X and Y are independent if and only if $\text{Cov}(F(X), G(Y)) = 0$ for all bounded and continuous functions F and G (see e.g. [Jacod and Protter, 2012]). It follows that, for well chosen RKHS, the nullity of the HSIC characterizes independence. Authors of [Gretton et al., 2005c] show that a sufficient condition so that the nullity of the associated HSIC characterizes independence is that the RKHS \mathcal{F} (resp. \mathcal{G}) induced by a kernel k (resp. l) is dense in the space of bounded and continuous functions mapping from \mathbb{R}^p (resp. \mathbb{R}^q) to \mathbb{R} . Such kernels are called *universal* [Micchelli et al., 2006]. However, the universality is a very limiting condition and only adapted to compact domains. Recently, a wider class of kernels called *characteristic kernels* has been introduced in [Fukumizu et al., 2008, Sriperumbudur et al., 2010]. These kernels characterize independence on compact as well as non-compact sets. Among them, one of the most commonly used is the Gaussian kernel [Steinwart, 2001], which we consider in this paper. It is defined as follows. Let g_d be the density of the standard Gaussian distribution on \mathbb{R}^d defined for all $x = (x^{(1)}, \dots, x^{(d)})$ in \mathbb{R}^d by

$$g_d(x) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2} \sum_{i=1}^d [x^{(i)}]^2\right). \quad (1.3)$$

For any bandwidths $\lambda = (\lambda_1, \dots, \lambda_p)$ in $(0, +\infty)^p$ and $\mu = (\mu_1, \dots, \mu_q)$ in $(0, +\infty)^q$, we denote for any x in \mathbb{R}^p and y in \mathbb{R}^q ,

$$\varphi_{\lambda}(x) = \frac{1}{\lambda_1 \dots \lambda_p} g_p\left(\frac{x^{(1)}}{\lambda_1}, \dots, \frac{x^{(p)}}{\lambda_p}\right), \quad \phi_{\mu}(y) = \frac{1}{\mu_1 \dots \mu_q} g_q\left(\frac{y^{(1)}}{\mu_1}, \dots, \frac{y^{(q)}}{\mu_q}\right). \quad (1.4)$$

Finally, the Gaussian kernels are defined for x, x' in \mathbb{R}^p and y, y' in \mathbb{R}^q by

$$k_{\lambda}(x, x') = \varphi_{\lambda}(x - x'), \quad \text{and} \quad l_{\mu}(y, y') = \phi_{\mu}(y - y').$$

A very convenient form of $\text{HSIC}(X, Y)$ is expressed in [Gretton et al., 2005a] using kernels k and l respectively associated to \mathcal{F} and \mathcal{G} ,

$$\begin{aligned} \text{HSIC}(X, Y) &= \mathbb{E}[k(X, X')l(Y, Y')] + \mathbb{E}[k(X, X')] \mathbb{E}[l(Y, Y')] \\ &\quad - 2\mathbb{E}[\mathbb{E}[k(X, X') | X] \mathbb{E}[l(Y, Y') | Y]], \end{aligned} \quad (1.5)$$

where (X', Y') is an i.i.d. copy of (X, Y) . Note that $\text{HSIC}(X, Y)$ only depends on the density f of (X, Y) . Hence, in the following, we denote by $\text{HSIC}_{\lambda, \mu}(f)$ the HSIC measure defined in (1.5), where the kernels k and l are respectively the Gaussian kernels k_λ and l_μ .

Given an i.i.d. n -sample $(X_i, Y_i)_{1 \leq i \leq n}$ with common density f , an estimator of $\text{HSIC}_{\lambda, \mu}(f)$ can be obtained by estimating each expectation of Equation (1.5). For this, we introduce the following U -statistics, respectively of order 2, 3 and 4,

$$\begin{aligned}\widehat{\text{HSIC}}_{\lambda, \mu}^{(2)} &= \frac{1}{n(n-1)} \sum_{(i, j) \in \mathbf{i}_2^n} k_\lambda(X_i, X_j) l_\mu(Y_i, Y_j), \\ \widehat{\text{HSIC}}_{\lambda, \mu}^{(3)} &= \frac{1}{n(n-1)(n-2)} \sum_{(i, j, r) \in \mathbf{i}_3^n} k_\lambda(X_i, X_j) l_\mu(Y_j, Y_r), \\ \widehat{\text{HSIC}}_{\lambda, \mu}^{(4)} &= \frac{1}{n(n-1)(n-2)(n-3)} \sum_{(i, j, r, s) \in \mathbf{i}_4^n} k_\lambda(X_i, X_j) l_\mu(Y_r, Y_s),\end{aligned}$$

where \mathbf{i}_r^n is the set of all r -tuples drawn without replacement from $\{1, \dots, n\}$. We estimate $\text{HSIC}_{\lambda, \mu}(f)$ by the U -statistic

$$\widehat{\text{HSIC}}_{\lambda, \mu} = \widehat{\text{HSIC}}_{\lambda, \mu}^{(2)} + \widehat{\text{HSIC}}_{\lambda, \mu}^{(4)} - 2\widehat{\text{HSIC}}_{\lambda, \mu}^{(3)}. \quad (1.6)$$

Similar estimators have been used to construct independence tests (see e.g. [Gretton et al., 2008]). Yet, only a heuristic choice of the bandwidths λ and μ is considered with no theoretical guarantees. To avoid this choice, following the work of [Fromont et al., 2013], we introduce in this paper an aggregated procedure based on Gaussian kernel HSIC measures and prove that it is minimax adaptive over Sobolev balls. Note that in the continuity of our work, [Kim et al., 2020] obtain minimax adaptive results over Hölder spaces for two-sample and independence tests based on permutations.

The structure of this paper is as follows. In Section 2, we first present a theoretical non-asymptotic HSIC-based test of prescribed level α as well as a permutation-based HSIC-test that is implemented in practice. We then provide theoretical conditions based on concentration inequalities for U -statistics, allowing to control the second kind error of the theoretical test by a given β . This last step leads us to sharp upper bounds of the uniform separation rate over Sobolev balls, and an optimal bandwidth choice (depending on the regularity parameter) in order to obtain a minimax optimal test. In Section 3, we introduce an aggregated procedure avoiding the bandwidth choice. We prove both an oracle-type inequality and sharp upper bounds for the uniform separation rate over Sobolev balls. Lower bounds over Sobolev spaces are obtained in Section 4. Finally, a comparison of the permutation-based test with the theoretical test first, and then with other existing tests, is presented in a simulation study in Section 5.

2 Single HSIC-based tests of independence

The aim of this section is to sharply upper bound the (non-asymptotic) uniform separation rate of HSIC-based tests over Sobolev balls which are well adapted to kernel-based tests. For this, theoretical conditions allowing to control the second kind error are first given in terms of $\text{HSIC}_{\lambda, \mu}(f)$ and then in terms of the \mathbb{L}_2 -norm of $f - f_1 \otimes f_2$. In this section, we consider fixed bandwidths (λ, μ) .

2.1 The testing procedures

Consider the notations introduced in Sections 1.2 and 1.3.

A theoretical test of independence Since Gaussian kernels are characteristic, testing the independence between X and Y is equivalent to testing

$$(\mathcal{H}_0) : \text{HSIC}_{\lambda,\mu}(f) = 0 \quad \text{against} \quad (\mathcal{H}_1) : \text{HSIC}_{\lambda,\mu}(f) > 0.$$

The statistic $\widehat{\text{HSIC}}_{\lambda,\mu}$ defined in Equation (1.6) is then a natural test statistic since it is an unbiased estimator of $\text{HSIC}_{\lambda,\mu}(f)$. For a prescribed level α in $(0, 1)$, we consider the theoretical statistical test $\Delta_\alpha^{\lambda,\mu}$ defined by

$$\Delta_\alpha^{\lambda,\mu} = \mathbb{1}_{\widehat{\text{HSIC}}_{\lambda,\mu} > q_{1-\alpha}^{\lambda,\mu}}, \quad (2.1)$$

where $q_{1-\alpha}^{\lambda,\mu}$ denotes the $(1 - \alpha)$ -quantile of $\widehat{\text{HSIC}}_{\lambda,\mu}$ under $P_{f_1 \otimes f_2}$. We reject (\mathcal{H}_0) if $\Delta_\alpha^{\lambda,\mu} = 1$. By definition of the quantile, this theoretical test is of non-asymptotic level α , that is for all densities f_1 and f_2 , $P_{f_1 \otimes f_2}(\Delta_\alpha^{\lambda,\mu} = 1) \leq \alpha$. Note that the analytical computation of the quantile $q_{1-\alpha}^{\lambda,\mu}$ is not possible since its value depends on the unknown marginal densities f_1 and f_2 . In practice, this quantile is approached by permutation with a Monte Carlo approximation as described in the following paragraph.

A permutation-based test of independence Let $\mathbb{Z}_n = (X_i, Y_i)_{1 \leq i \leq n}$ denote the original sample and compute the test statistic $\widehat{\text{HSIC}}_{\lambda,\mu}(\mathbb{Z}_n)$ defined by Equation (1.6). Then, let τ_1, \dots, τ_B be B i.i.d. random permutations of $\{1, \dots, n\}$, independent of \mathbb{Z}_n . We define for each permutation τ_b the corresponding permuted sample $\mathbb{Z}_n^{\tau_b} = (X_i, Y_{\tau_b(i)})_{1 \leq i \leq n}$ and compute the permuted test statistic on this new sample

$$\widehat{H}_{\lambda,\mu}^{\star b} = \widehat{\text{HSIC}}_{\lambda,\mu}(\mathbb{Z}_n^{\tau_b}).$$

Under $P_{f_1 \otimes f_2}$, each permuted sample $\mathbb{Z}_n^{\tau_b}$ has the same distribution as the original sample \mathbb{Z}_n . Hence, the random variables $\{\widehat{H}_{\lambda,\mu}^{\star b}\}_{1 \leq b \leq B}$, have the same distribution as $\widehat{\text{HSIC}}_{\lambda,\mu}$. We apply a trick, based on [Romano and Wolf, 2005, Lemma 1], which consists in adding the original sample to the Monte Carlo sample in order to obtain a test of non-asymptotic level α . To do so, denote

$$\widehat{H}_{\lambda,\mu}^{\star B+1} = \widehat{\text{HSIC}}_{\lambda,\mu} \quad \text{and} \quad \widehat{H}_{\lambda,\mu}^{\star(1)} \leq \widehat{H}_{\lambda,\mu}^{\star(2)} \leq \dots \leq \widehat{H}_{\lambda,\mu}^{\star(B+1)}$$

the order statistic. Then, the permuted quantile with Monte Carlo approximation $\hat{q}_{1-\alpha}^{\lambda,\mu}$ is thus defined as

$$\hat{q}_{1-\alpha}^{\lambda,\mu} = \widehat{H}_{\lambda,\mu}^{\star(\lceil (B+1)(1-\alpha) \rceil)}. \quad (2.2)$$

where $\lceil \cdot \rceil$ denotes the ceiling function. The permuted test with Monte Carlo approximation $\widehat{\Delta}_\alpha^{\lambda,\mu}$ performed in practice is then defined as

$$\widehat{\Delta}_\alpha^{\lambda,\mu} = \mathbb{1}_{\widehat{\text{HSIC}}_{\lambda,\mu} > \hat{q}_{1-\alpha}^{\lambda,\mu}}. \quad (2.3)$$

Proposition 1. *Let α in $(0, 1)$ and consider the permuted test with Monte Carlo approximation $\widehat{\Delta}_\alpha^{\lambda,\mu}$ defined by Equation (2.3). Then, for all B , $P_{f_1 \otimes f_2}(\widehat{\Delta}_\alpha^{\lambda,\mu} = 1) \leq \alpha$.*

Hence, both the theoretical test $\Delta_\alpha^{\lambda,\mu}$ and the permuted test $\widehat{\Delta}_\alpha^{\lambda,\mu}$ are of prescribed non-asymptotic level α . A comparison in terms of power is done on simulated data in Section B.1 justifying the restriction of the following theoretical study to the theoretical test.

2.2 Control of the second kind error in terms of HSIC

For an arbitrarily small β given in $(0, 1)$, Lemma 1 provides a first non-asymptotic condition on the alternative f ensuring that the probability under P_f of second kind error of the theoretical test $\Delta_\alpha^{\lambda, \mu}$ defined in Equation (2.1) is at most equal to β . This condition is given for the value of $\widehat{\text{HSIC}}_{\lambda, \mu}(f)$. It involves the variance of the estimator $\widehat{\text{HSIC}}_{\lambda, \mu}$ which is finite since this estimator is bounded.

Lemma 1. *Let α, β in $(0, 1)$ and $(X_i, Y_i)_{1 \leq i \leq n}$ be an i.i.d. sample with distribution P_f . Consider the test statistic $\widehat{\text{HSIC}}_{\lambda, \mu}$ defined by (1.6) and denote $q_{1-\alpha}^{\lambda, \mu}$ its $(1 - \alpha)$ -quantile under $P_{f_1 \otimes f_2}$. Then $P_f(\widehat{\text{HSIC}}_{\lambda, \mu} \leq q_{1-\alpha}^{\lambda, \mu}) \leq \beta$ as soon as*

$$\widehat{\text{HSIC}}_{\lambda, \mu}(f) \geq \sqrt{\frac{\text{Var}_f(\widehat{\text{HSIC}}_{\lambda, \mu})}{\beta}} + q_{1-\alpha}^{\lambda, \mu}.$$

Lemma 1 gives a threshold for $\widehat{\text{HSIC}}_{\lambda, \mu}(f)$ from which the dependence between X and Y is detectable with probability greater than $1 - \beta$. In order to express the order of magnitude of this threshold w.r.t. n, λ and μ , we establish sharp upper bounds for both the variance $\text{Var}_f(\widehat{\text{HSIC}}_{\lambda, \mu})$ and the quantile $q_{1-\alpha}^{\lambda, \mu}$. Proposition 2 gives an upper bound for the variance.

Proposition 2. *Let f be a density satisfying Assumption \mathcal{A}_1 , and $(X_i, Y_i)_{1 \leq i \leq n}$ be an i.i.d. sample with distribution P_f . Consider the test statistic $\widehat{\text{HSIC}}_{\lambda, \mu}$ defined by (1.6). Then,*

$$\text{Var}_f(\widehat{\text{HSIC}}_{\lambda, \mu}) \leq C(M_f, p, q) \left\{ \frac{1}{n} + \frac{1}{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q n^2} \right\}.$$

Proposition 3 provides an upper bound for the quantiles. It requires the following assumptions on the bandwidths (λ, μ) :

$$\mathcal{A}_2(\alpha) : \max \left\{ \prod_{i=1}^p \lambda_i, \prod_{j=1}^q \mu_j \right\} < 1 \quad \text{and} \quad n \sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q} > \log \left(\frac{1}{\alpha} \right) > 1.$$

Note that larger sample sizes allow for smaller bandwidths.

Proposition 3. *Let α in $(0, 1)$. Let f be a density satisfying \mathcal{A}_1 and $(X_i, Y_i)_{1 \leq i \leq n}$ be an i.i.d. sample with distribution P_f . Consider bandwidths (λ, μ) satisfy Assumptions $\mathcal{A}_2(\alpha)$. Denote $\widehat{\text{HSIC}}_{\lambda, \mu}$ the test statistic defined by (1.6) and $q_{1-\alpha}^{\lambda, \mu}$ its $(1 - \alpha)$ -quantile under $P_{f_1 \otimes f_2}$. Then,*

$$q_{1-\alpha}^{\lambda, \mu} \leq \frac{C(M_f, p, q)}{n \sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}} \log \left(\frac{1}{\alpha} \right).$$

Combining Lemma 1 with Propositions 2 and 3, Corollary 1 provides a sufficient condition on $\widehat{\text{HSIC}}_{\lambda, \mu}(f)$ depending on the bandwidths λ, μ and the sample size n in order to control the second kind error rate by β .

Corollary 1. *Under the assumptions of Lemma 1, Propositions 2 and 3, one has $P_f(\Delta_\alpha^{\lambda, \mu} = 0) \leq \beta$ as soon as*

$$\widehat{\text{HSIC}}_{\lambda, \mu}(f) > C(M_f, p, q, \beta) \left\{ \frac{1}{\sqrt{n}} + \frac{1}{n \sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}} \log \left(\frac{1}{\alpha} \right) \right\}.$$

Note that the right-hand side term given in Corollary 1 depends on the unknown density f . However, this dependence is weak since it only involves the infinite norm of f and its marginals.

2.3 Control of the second kind error in terms of \mathbb{L}_2 -norm

For the sake of interpretation, and in order to upper bound the non-asymptotic uniform separation rates w.r.t. the \mathbb{L}_2 -norm, we now want to express the condition on $\text{HSIC}_{\lambda,\mu}(f)$ obtained in Corollary 1 in terms of the \mathbb{L}_2 norm of the difference $f - f_1 \otimes f_2$. To do so, we first give in Lemma 2 a link between $\text{HSIC}_{\lambda,\mu}(f)$ and $\|f - f_1 \otimes f_2\|_2^2$.

Lemma 2. *Let $\psi = f - f_1 \otimes f_2$. The HSIC measure of f associated to kernels k_λ and l_μ and defined in Equation (1.5) can be written as*

$$\text{HSIC}_{\lambda,\mu}(f) = \langle \psi, \psi * (\varphi_\lambda \otimes \phi_\mu) \rangle_2,$$

where φ_λ and ϕ_μ are the functions defined in Equation (1.4), and $\langle \cdot, \cdot \rangle_2$ denotes the usual scalar product in the \mathbb{L}_2 space. One can easily deduce that

$$\text{HSIC}_{\lambda,\mu}(f) = \frac{1}{2} \left(\|\psi\|_2^2 + \|\psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2 - \|\psi - \psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2 \right). \quad (2.4)$$

Theorem 1 gives a sufficient condition on $\|f - f_1 \otimes f_2\|_2^2$, for the second kind error of the test $\Delta_\alpha^{\lambda,\mu}$ to be upper bounded by β .

Theorem 1. *Let α, β in $(0, 1)$ and consider the test $\Delta_\alpha^{\lambda,\mu}$ defined by (2.1). Assume that the density f satisfies \mathcal{A}_1 and that the bandwidths (λ, μ) satisfy $\mathcal{A}_2(\alpha)$. Then, $P_f(\Delta_\alpha^{\lambda,\mu} = 0) \leq \beta$ as soon as*

$$\|\psi\|_2^2 > \|\psi - \psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2 + \frac{C(M_f, p, q, \beta)}{n\sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}} \log\left(\frac{1}{\alpha}\right). \quad (2.5)$$

where $\psi = f - f_1 \otimes f_2$, and $C(\cdot)$ denotes a positive constant depending only on its arguments.

In Condition (2.5) appears a compromise between a bias term, namely $\|\psi - \psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2$, and a term induced by the square root of the variance of the estimator $\widehat{\text{HSIC}}_{\lambda,\mu}$. Note that, due to Proposition 3, this variance term also controls the quantile term. Comparing the conditions on the HSIC given in Corollary 1 and on the \mathbb{L}_2 -norm $\|f - f_1 \otimes f_2\|_2^2$ given in Theorem 1, the meticulous reader may notice that the term in $1/\sqrt{n}$ has been removed. This suppression seems to be necessary to obtain optimal separation rates according to the literature in other testing frameworks. This derives from quite tricky computations that we point out here and that directly prove Theorem 1. By combining Lemmas 1 and 2, direct computations lead to the condition

$$\|\psi\|_2^2 > \|\psi - \psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2 - \|\psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2 + 2\sqrt{\frac{\text{Var}_f(\widehat{\text{HSIC}}_{\lambda,\mu})}{\beta}} + 2q_{1-\alpha}^{\lambda,\mu}. \quad (2.6)$$

If one directly considers the upper bound of the variance $\text{Var}_f(\widehat{\text{HSIC}}_{\lambda,\mu})$ given in Proposition 2, one would get the unwanted $1/\sqrt{n}$ term. The idea is to take advantage of the negative term $-\|\psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2$ to compensate such term. To do so, we need a more refined control of the variance given in the technical Proposition 4.

Proposition 4. *Let f be a density satisfying Assumption \mathcal{A}_1 , and $(X_i, Y_i)_{1 \leq i \leq n}$ be an i.i.d. sample with distribution P_f . Consider the test statistic $\widehat{\text{HSIC}}_{\lambda,\mu}$ defined by (1.6). Then,*

$$\text{Var}_f(\widehat{\text{HSIC}}_{\lambda,\mu}) \leq C(M_f) \frac{\|\psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2}{n} + \frac{C(M_f, p, q)}{n^2 \lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}.$$

Finally, using standard inequalities such as $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ or such as $2\sqrt{ab} \leq ca + b/c$ for all positive a , b and c , one can prove that

$$2\sqrt{\frac{\text{Var}_f(\widehat{\text{HSIC}}_{\lambda,\mu})}{\beta}} \leq \|\psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2 + \frac{C(M_f, \beta)}{n} + \frac{C(M_f, p, q, \beta)}{n\sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}},$$

which leads to Theorem 1 when combined with Equation (2.6) and Proposition 3. Notice that such trick is already present in [Fromont et al., 2013].

2.4 Uniform separation rate over Sobolev balls

The bias term in Theorem 1 comes from the fact that we do not estimate $\|f - f_1 \otimes f_2\|_2^2$ but $\text{HSIC}_{\lambda,\mu}(f)$. In order to have a control of the bias term w.r.t λ and μ , we assume that $f - f_1 \otimes f_2$ belongs to some class of regular functions.

The Sobolev ball $\mathcal{S}_d^\delta(R)$ in dimension d in \mathbb{N}^* , with regularity parameter $\delta > 0$ and radius $R > 0$, is defined by

$$\mathcal{S}_d^\delta(R) = \left\{ s : \mathbb{R}^d \rightarrow \mathbb{R} ; s \in \mathcal{L}_1(\mathbb{R}^d) \cap \mathcal{L}_2(\mathbb{R}^d), \int_{\mathbb{R}^d} \|u\|^{2\delta} |\hat{s}(u)|^2 du \leq (2\pi)^d R^2 \right\}, \quad (2.7)$$

where $\|\cdot\|$ denotes the Euclidean norm associated to the usual scalar product $\langle \cdot, \cdot \rangle$ in \mathbb{R}^d , and \hat{s} denotes the Fourier transform of s , defined on \mathbb{R}^d by $\hat{s}(u) = \int_{\mathbb{R}^d} s(x) e^{i\langle x, u \rangle} dx$. Lemma 3 gives an upper bound for the bias term in the case where $f - f_1 \otimes f_2$ belongs to particular Sobolev balls.

Lemma 3. *Assume that $\psi = f - f_1 \otimes f_2$ belongs to the Sobolev ball $\mathcal{S}_{p+q}^\delta(R)$ with positive parameters δ and R , defined in (2.7). Let φ_λ and ϕ_μ be the functions defined in (1.4). Then, there exists T_δ in $[0, 1]$ such that*

$$\|\psi - \psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2 \leq (1 - e^{-T_\delta^2/2}) \|\psi\|_2^2 + C(p, q, \delta, R) \left[\sum_{i=1}^p \lambda_i^{2\delta} + \sum_{j=1}^q \mu_j^{2\delta} \right].$$

Moreover, if δ belongs to $(0, 2]$, then $T_\delta = 0$ and the term with $\|\psi\|_2$ vanishes.

In the following, we study optimality over $\mathcal{S}_{p+q}^\delta(R, R')$ defined by

$$\mathcal{S}_{p+q}^\delta(R, R') = \mathcal{S}_{p+q}^\delta(R) \cap \{f; \max\{\|f\|_\infty, \|f_1\|_\infty, \|f_2\|_\infty\} \leq R'\}. \quad (2.8)$$

One can deduce from Theorem 1 upper bounds for the uniform separation rates, defined in (1.1), of the test $\Delta_\alpha^{\lambda,\mu}$ over Sobolev balls.

Theorem 2. *Let α, β in $(0, 1)$, and positive parameters δ, R and R' . Consider bandwidths (λ, μ) satisfying Assumptions $\mathcal{A}_2(\alpha)$ and denote $\Delta_\alpha^{\lambda,\mu}$ the test defined by (2.1). Then, the uniform separation rate defined in (1.1) of the test $\Delta_\alpha^{\lambda,\mu}$ over the Sobolev ball $\mathcal{S}_{p+q}^\delta(R, R')$ defined in Equation (2.8) can be upper bounded as follows*

$$\begin{aligned} \left[\rho \left(\Delta_\alpha^{\lambda,\mu}, \mathcal{S}_{p+q}^\delta(R, R'), \beta \right) \right]^2 &\leq C(p, q, \delta, R) \left[\sum_{i=1}^p \lambda_i^{2\delta} + \sum_{j=1}^q \mu_j^{2\delta} \right] \\ &\quad + \frac{C(R', p, q, \beta)}{n\sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}} \log \left(\frac{1}{\alpha} \right), \end{aligned} \quad (2.9)$$

where $C(\cdot)$ denote positive constants depending only on their arguments.

One can now determine optimal bandwidths (λ^*, μ^*) which minimize the right-hand side of Equation (2.9). To do so, the idea is to find for which (λ, μ) both terms in the right-hand side of (2.9) are of the same order w.r.t. n . We also provide an upper bound for the uniform separation rate of the optimized test $\Delta_\alpha^{\lambda^*, \mu^*}$ over Sobolev balls.

Corollary 2. *Let α in $(0, 1/e)$, β in $(0, 1)$, and $\delta, R, R' > 0$. Define for all i in $\{1, \dots, p\}$ and for all j in $\{1, \dots, q\}$,*

$$\lambda_i^* = \mu_j^* = n^{-2/(4\delta+p+q)}.$$

If $n > (\log(1/\alpha))^{1+(p+q)/(4\delta)}$, then, (λ^, μ^*) satisfy $\mathcal{A}_2(\alpha)$ and the uniform separation rate of the optimized test $\Delta_\alpha^{\lambda^*, \mu^*}$ over the Sobolev ball $\mathcal{S}_{p+q}^\delta(R, R')$ is controlled as follows*

$$\rho(\Delta_\alpha^{\lambda^*, \mu^*}, \mathcal{S}_{p+q}^\delta(R, R'), \beta) \leq C(p, q, \alpha, \beta, \delta, R, R') n^{-2\delta/(4\delta+p+q)}.$$

Note that, in the definition of the Sobolev ball $\mathcal{S}_{p+q}^\delta(R, R')$, we have the same regularity parameter $\delta > 0$ for all directions in \mathbb{R}^{p+q} . This corresponds to an isotropic regularity condition. Similar results over anisotropic Nikol'skii regularities are obtained in Appendix A in the supplementary material.

Moreover, the test $\Delta_\alpha^{\lambda^*, \mu^*}$ with the optimized bandwidths depends on the regularity parameter δ and cannot be computed in practice. In the next section, for the purpose of adaptivity, we build an aggregated testing procedure taking into account a collection of bandwidths. In particular, this avoids the delicate choice of arbitrary bandwidths. We then prove an oracle-type inequality and show that the uniform separation rate of this aggregated procedure is of the same order as the smallest uniform separation rate of the tests in the chosen collection, up to a logarithmic term.

Finally, note that subsequently, [Kim et al., 2020] generalize Theorem 2 to the permuted tests. However, they obtain a polynomial dependence in α , that is $1/\sqrt{\alpha}$ instead of $\log(1/\alpha)$, which leads to minimax optimal tests for an optimized bandwidth choice, as in Corollary 2. Yet, the dependence in α is not sharp enough to provide adaptive tests by aggregating.

3 Aggregated HSIC-based test of independence

In Section 2, we consider single tests based on Gaussian kernels associated to given bandwidths (λ, μ) . However, there is as yet no justified method to choose λ and μ with theoretical guarantees. In many cases, authors choose these parameters w.r.t the available data $(X_i, Y_i)_{1 \leq i \leq n}$ by taking for example λ (resp. μ) as the empirical median (see [Zhang et al., 2011]) or the empirical mean (see, e.g., [De Lozzo and Marrel, 2017, Marrel et al., 2020]) of $(\|X_i - X_j\|)_{1 \leq i < j \leq n}$ (resp. $(\|Y_i - Y_j\|)_{1 \leq i < j \leq n}$). To avoid this arbitrary choice, we consider in this section an aggregated testing procedure combining a collection of single tests based on different bandwidths.

3.1 The aggregated testing procedure

Consider now a finite or countable collection $\mathcal{W} \subset (0, +\infty)^p \times (0, +\infty)^q$ of bandwidths (λ, μ) and a collection of positive weights $\{\omega_{\lambda, \mu}\}_{(\lambda, \mu) \in \mathcal{W}}$ such that $\sum_{(\lambda, \mu) \in \mathcal{W}} e^{-\omega_{\lambda, \mu}} \leq 1$.

For a given α in $(0, 1)$, the aggregated test rejects (\mathcal{H}_0) if there is at least one (λ, μ) in \mathcal{W} such that the corresponding single test with corrected level $u_\alpha \exp(-\omega_{\lambda, \mu})$ rejects (\mathcal{H}_0) , that is

$$\exists (\lambda, \mu) \in \mathcal{W} ; \widehat{\text{HSIC}}_{\lambda, \mu} > q_{1-u_\alpha e^{-\omega_{\lambda, \mu}}}^{\lambda, \mu},$$

where u_α is the less conservative value such that the aggregated test is of level α . More precisely, this level correction is defined by

$$u_\alpha = \sup \left\{ u > 0 ; P_{f_1 \otimes f_2} \left(\sup_{(\lambda, \mu) \in \mathcal{W}} \left\{ \widehat{\text{HSIC}}_{\lambda, \mu} - q_{1-ue^{-\omega_{\lambda, \mu}}}^{\lambda, \mu} \right\} > 0 \right) \leq \alpha \right\}. \quad (3.1)$$

We should mention here that the supremum in Equation (3.1) exists since the function

$$u \mapsto P_{f_1 \otimes f_2} \left(\sup_{(\lambda, \mu) \in \mathcal{W}} \left\{ \widehat{\text{HSIC}}_{\lambda, \mu} - q_{1-ue^{-\omega_{\lambda, \mu}}}^{\lambda, \mu} \right\} > 0 \right)$$

is well defined for u in the interval $(0, \inf_{(\lambda, \mu) \in \mathcal{W}} \{\exp(\omega_{\lambda, \mu})\})$, non-decreasing, and converges to 0 and 1 respectively at the boundaries of this interval. Moreover, we can choose weights such that $\sum_{(\lambda, \mu) \in \mathcal{W}} e^{-\omega_{\lambda, \mu}} = 1$. Yet, in practice, it just changes the value of u_α and leads to the same test.

The (theoretical) aggregated test Δ_α is then defined by

$$\Delta_\alpha = \mathbb{1}_{\sup_{(\lambda, \mu) \in \mathcal{W}} \left\{ \widehat{\text{HSIC}}_{\lambda, \mu} - q_{1-u_\alpha e^{-\omega_{\lambda, \mu}}}^{\lambda, \mu} \right\} > 0}, \quad (3.2)$$

and rejects (\mathcal{H}_0) if $\Delta_\alpha = 1$. By definition of u_α , the test Δ_α is of level α .

For computational limitations, the collection \mathcal{W} is finite in practice. Moreover, as for the quantile, the correction u_α of the level is not analytically computable since it depends on the unknown marginal densities f_1 and f_2 . In practice, it can also be approached by a permutation method with Monte Carlo approximation, as done in [Albert, 2015]. More precisely, consider the notations of Section 2.1. First, generate B_1 independent and uniformly distributed random permutations of $\{1, \dots, n\}$, denoted $\tau_1, \dots, \tau_{B_1}$, independent of \mathbb{Z}_n and compute for each (λ, μ) in \mathcal{W} and each $u > 0$ the permuted quantile with Monte Carlo approximation $\hat{q}_{1-ue^{-\omega_{\lambda, \mu}}}^{\lambda, \mu}$ as defined in (2.2). Second, in order to estimate the probabilities under $P_{f_1 \otimes f_2}$ in (3.1), generate B_2 independent and uniformly distributed random permutations of $\{1, \dots, n\}$, denoted $\kappa_1, \dots, \kappa_{B_2}$, independent of \mathbb{Z}_n and of $\tau_1, \dots, \tau_{B_1}$. Denote for all permutation κ_b , the corresponding permuted statistic

$$\widehat{H}_{\lambda, \mu}^{\kappa_b} = \widehat{\text{HSIC}}_{\lambda, \mu}(\mathbb{Z}_n^{\kappa_b})$$

Then, the correction u_α is approached by Monte Carlo as follows:

$$\hat{u}_\alpha = \sup \left\{ u > 0 ; \frac{1}{B_2} \sum_{b=1}^{B_2} \mathbb{1}_{\max_{(\lambda, \mu) \in \mathcal{W}} \left\{ \widehat{H}_{\lambda, \mu}^{\kappa_b} - \hat{q}_{1-ue^{-\omega_{\lambda, \mu}}}^{\lambda, \mu} \right\} > 0} \leq \alpha \right\}. \quad (3.3)$$

In the end, the permuted aggregated test $\hat{\Delta}_\alpha$ with Monte Carlo approximation is defined by

$$\hat{\Delta}_\alpha = \mathbb{1}_{\max_{(\lambda, \mu) \in \mathcal{W}} \left\{ \widehat{\text{HSIC}}_{\lambda, \mu} - \hat{q}_{1-\hat{u}_\alpha e^{-\omega_{\lambda, \mu}}}^{\lambda, \mu} \right\} > 0}. \quad (3.4)$$

As for the single tests, a comparison in terms of power is done on simulated data in Section 5.1 justifying the restriction of the following theoretical study to the theoretical aggregated test.

3.2 Oracle-type conditions for the uniform separation rate over Sobolev balls

As a reminder, our goal is to construct a testing procedure with a uniform separation rate as small as possible and whose implementation does not require any information about the regularity of the difference $f - f_1 \otimes f_2$.

The main advantage of the aggregated procedure is that its second kind error is upper bounded by the smallest error of the single tests (with corrected levels) in the collection. The main argument is highlighted in Lemma 4.

Lemma 4. *Let α, β in $(0, 1)$, and consider the aggregated test Δ_α defined in Equation (3.2). Then, $u_\alpha \geq \alpha$ and*

$$P_f(\Delta_\alpha = 0) \leq \inf_{(\lambda, \mu) \in \mathcal{W}} \left\{ P_f \left(\Delta_{\alpha e^{-\omega_{\lambda, \mu}}}^{\lambda, \mu} = 0 \right) \right\}.$$

According to Lemma 4, if there exists at least one single test $\Delta_{\alpha e^{-\omega_{\lambda, \mu}}}^{\lambda, \mu}$ with a probability of second kind error at most equal to β , then the same control holds for the aggregated test Δ_α . Theorem 3 gives an oracle-type inequality for the uniform separation rate of the aggregated testing procedure Δ_α , showing the interest of this procedure.

Theorem 3. *Let α, β in $(0, 1)$. Consider a finite or countable collection $\mathcal{W} \subset (0, +\infty)^p \times (0, +\infty)^q$ of bandwidths (λ, μ) and a collection of positive weights $\{\omega_{\lambda, \mu}\}_{(\lambda, \mu) \in \mathcal{W}}$ such that $\sum_{(\lambda, \mu) \in \mathcal{W}} e^{-\omega_{\lambda, \mu}} \leq 1$ and such that all (λ, μ) in \mathcal{W} verifies Assumption $\mathcal{A}_2(\alpha e^{-\omega_{\lambda, \mu}})$. Then, the uniform separation rate over Sobolev balls $\mathcal{S}_{p+q}^\delta(R, R')$ with positive parameters δ, R and R' of the aggregated test Δ_α defined in Equation (3.2) can be upper bounded as follows*

$$\left[\rho \left(\Delta_\alpha, \mathcal{S}_{p+q}^\delta(R, R'), \beta \right) \right]^2 \leq C(p, q, \beta, \delta, R, R') \times \inf_{(\lambda, \mu) \in \mathcal{W}} \left\{ \left[\sum_{i=1}^p \lambda_i^{2\delta} + \sum_{j=1}^q \mu_j^{2\delta} \right] + \frac{1}{n \sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}} \left(\log \left(\frac{1}{\alpha} \right) + \omega_{\lambda, \mu} \right) \right\},$$

where $C(\cdot)$ is a positive constant depending only on its arguments.

Theorem 3 can be interpreted as an oracle-type condition for the uniform separation rate of the test Δ_α . Indeed, without knowing the regularity of $f - f_1 \otimes f_2$, we prove that the uniform separation rate of Δ_α is of the same order as the smallest uniform separation rate of the single tests corresponding to bandwidths (λ, μ) in \mathcal{W} , up to an additional term $\omega_{\lambda, \mu}$ due to the correction of the individual levels.

3.3 Uniform separation rate over Sobolev balls

In this section, we consider the aggregated test for a particular choice of bandwidths collection \mathcal{W} defined by

$$\mathcal{W} = \left\{ (2^{-m} \mathbf{1}_{p+q}, m \in \{1, \dots, M_n^{p,q}\}) \right\}, \quad (3.5)$$

where $\mathbf{1}_{p+q} = (1, 1, \dots, 1) \in \mathbb{R}^{p+q}$ and, denoting $\lfloor \cdot \rfloor$ the floor function,

$$M_n^{p,q} = \left\lceil \log_2 \left(\left[\frac{n}{\log(n)} \right]^{\frac{2}{p+q}} \right) \right\rceil.$$

In addition, we associate to every $(\lambda, \mu) = 2^{-m} \mathbf{1}_{p+q}$ in \mathcal{W} the positive weight

$$\omega_{\lambda, \mu} = 2 \log \left(m \times \frac{\pi}{\sqrt{6}} \right), \quad (3.6)$$

so that $\sum_{(\lambda, \mu) \in \mathcal{W}} e^{-\omega_{\lambda, \mu}} \leq 1$. Corollary 3 justifies that this particular choice of bandwidths collection and associated weights is well adapted to Sobolev regularities.

Corollary 3. *Let α, β in $(0, 1)$. Consider the aggregated test Δ_α defined in (3.2), with the particular choice of the collection \mathcal{W} and the weights $(\omega_{\lambda, \mu})_{(\lambda, \mu) \in \mathcal{W}}$ defined in (3.5) and (3.6). Assume that $\log \log(n) > 1$. Under the assumptions of Theorem 3, for any $\delta, R, R' > 0$, there exists a positive constant $C(p, q, \alpha, \delta)$ such that for all $n \geq C(p, q, \alpha, \delta)$, the uniform separation rate over the Sobolev ball $\mathcal{S}_{p+q}^\delta(R, R')$ of Δ_α can be upper bounded as follows:*

$$\rho \left(\Delta_\alpha, \mathcal{S}_{p+q}^\delta(R, R'), \beta \right) \leq C(p, q, \alpha, \beta, \delta, R, R') \left(\frac{\log \log(n)}{n} \right)^{2\delta/(4\delta+p+q)}. \quad (3.7)$$

According to Corollary 3, the uniform separation rate of the aggregated procedure over Sobolev balls is of the same order as the one of the optimized test $\Delta_\alpha^{\lambda^*, \mu^*}$ (given in Corollary 2), up to a $\log \log(n)$ factor. Note that this logarithmic loss is usually the price to pay for aggregated tests (see, e.g., [Spokoiny, 1996, Ingster, 2000]). Similar results over Nikol'skii-Besov spaces are also obtained in the supplementary material.

4 Lower bound for the non-asymptotic minimax rate over Sobolev balls

In this section, we present a general method based on a Bayesian approach to lower bound the non-asymptotic minimax rate of testing as defined in (1.2). The general idea of this method is due to [Ingster, 1993a] and relies on Lemma 5.

Lemma 5. *Let α, β, η in $(0, 1)$ such that $\alpha + \beta + \eta < 1$. Let \mathcal{C}_δ denote some regularity space, and recall that for all positive ρ , the set $\mathcal{F}_\rho(\mathcal{C}_\delta)$ is defined by*

$$\mathcal{F}_\rho(\mathcal{C}_\delta) = \{f ; f - f_1 \otimes f_2 \in \mathcal{C}_\delta, \|f - f_1 \otimes f_2\|_2 \geq \rho\}.$$

Let us denote

$$\beta[\mathcal{F}_\rho(\mathcal{C}_\delta)] = \inf_{\Delta_\alpha} \sup_{f \in \mathcal{F}_\rho(\mathcal{C}_\delta)} P_f(\Delta_\alpha = 0),$$

where the infimum is taken over all α -level tests of (\mathcal{H}_0) against (\mathcal{H}_1) .

Let $\rho_* > 0$ and consider a probability measure ν_{ρ_*} defined on the set of densities in $\mathbb{L}_2(\mathbb{R}^p \times \mathbb{R}^q)$ such that $\nu_{\rho_*}(\mathcal{F}_{\rho_*}(\mathcal{C}_\delta)) \geq 1 - \eta$. Define the associated probability measure $P_{\nu_{\rho_*}}$ for all measurable set A in $\mathbb{R}^{n(p+q)}$ by

$$P_{\nu_{\rho_*}}(A) = \int_{\mathbb{L}_2(\mathbb{R}^p \times \mathbb{R}^q)} P_f(A) d\nu_{\rho_*}(f).$$

Assume there exists a density f_0 that satisfies (\mathcal{H}_0) such that the probability measure $P_{\nu_{\rho_*}}$ is absolutely continuous w.r.t. P_{f_0} and verifies

$$\mathbb{E}_{P_{f_0}}[L_{\nu_{\rho_*}}^2(\mathbb{Z}_n)] < 1 + 4(1 - \alpha - \beta - \eta)^2, \quad (4.1)$$

where the likelihood ratio $L_{\nu_{\rho_*}}$ is defined by $L_{\nu_{\rho_*}} = dP_{\nu_{\rho_*}}/dP_{f_0}$. Then, for all $\rho \leq \rho_*$ we have that $\beta[\mathcal{F}_\rho(\mathcal{C}_\delta)] > \beta$. It follows that

$$\rho(\mathcal{C}_\delta, \alpha, \beta) = \inf_{\Delta_\alpha} \rho(\Delta_\alpha, \mathcal{C}_\delta, \beta) \geq \rho_*.$$

We aim at proving that

$$\rho_n^* = Cn^{-2\delta/(4\delta+p+q)}$$

is a lower bound for the non-asymptotic minimax rate of testing, defined in (1.2), over Sobolev balls $\mathcal{S}_{p+q}^\delta(R, R')$, for some positive constant C , that is, $\rho(\mathcal{S}_{p+q}^\delta(R, R'), \alpha, \beta) \geq \rho_n^*$. According to Lemma 5, it is sufficient to find a probability distribution $\nu_{\rho_n^*}$ such that $\nu_{\rho_n^*}(\mathcal{F}_{\rho_n^*}(\mathcal{S}_{p+q}^\delta(R, R')))) \geq 1 - \eta$ and such that Condition (4.1) holds.

To do so, we generalize the construction of [Butucea, 2007] to our multidimensional framework. The idea is to construct a finite set of alternatives $(f_\theta)_\theta$ by perturbing the uniform density on $[0, 1]^p \times [0, 1]^q$, and define $\nu_{\rho_n^*}$ as a uniform mixture of these alternatives. For this, consider the function G defined for all t in \mathbb{R} by

$$G(t) = e^{-1/[1-(4t+3)^2]} \mathbb{1}_{(-1, -1/2)}(t) - e^{-1/[1-(4t+1)^2]} \mathbb{1}_{(-1/2, 0)}(t). \quad (4.2)$$

One may notice that G is continuous, with support in $[-1, 0]$ and that $\int_{\mathbb{R}} G(t)dt = 0$. The function G together with its Fourier transform has valuable properties for our study.

Let h_n be in $(0, 1]$ to be specified later such that $M_n := 1/h_n$ is an integer. Denote $I_{n,p,q} = \{1, \dots, M_n\}^p \times \{1, \dots, M_n\}^q$. For all $\theta = (\theta_{(j,l)})_{(j,l) \in I_{n,p,q}}$ in $\{-1, 1\}^{M_n^{p+q}}$, define for all (x, y) in $\mathbb{R}^p \times \mathbb{R}^q$,

$$f_\theta(x, y) = \mathbb{1}_{[0,1]^{p+q}}(x, y) + C_0 h_n^{\delta+(p+q)} \sum_{(j,l) \in I_{n,p,q}} \theta_{(j,l)} \prod_{r=1}^p G_{h_n}(x_r - j_r h_n) \prod_{s=1}^q G_{h_n}(y_s - l_s h_n), \quad (4.3)$$

where for all $h > 0$, $G_h(\cdot) = (1/h)G(\cdot/h)$ and C_0 is a constant depending on $(p, q, \delta, R, R', \eta)$ that will be specified later. One may notice that for all θ , the alternative f_θ is supported in $[0, 1]^{p+q}$. Moreover, since the integral of G over \mathbb{R} equals 0, the marginals $f_{\theta,1}$ and $f_{\theta,2}$ of f_θ are respectively the uniform densities on $[0, 1]^p$ and $[0, 1]^q$. Lemmas 6 and 7 justify the choice of these alternatives.

Lemma 6. *Let $\delta > 0$, $R > 0$ and $R' \geq 1$. Consider h_n in $(0, 1]$ such that $M_n := 1/h_n$ is an integer. Then, for all $\theta = (\theta_{(j,l)})_{(j,l) \in I_{n,p,q}}$ in $\{-1, 1\}^{M_n^{p+q}}$, the function f_θ defined in Equation (4.3) satisfies the following properties.*

1. If $C_0 \leq \min\{1, R' - 1\}e^{p+q}$, then the function f_θ is a density function and

$$\max\{\|f_\theta\|_\infty, \|f_{\theta,1}\|_\infty, \|f_{\theta,2}\|_\infty\} \leq R'.$$

2. The function f_θ is such that $\|f_\theta - f_{\theta,1} \otimes f_{\theta,2}\|_2 = C_0 \|G\|_2^{p+q} h_n^\delta$.

Let us now consider a uniform mixture $\nu_{\rho_n^*}$ of the alternatives (f_θ) , for θ in $\{-1, 1\}^{M_n^{p+q}}$. Note that this is equivalent to considering a random alternative f_Θ where $\Theta = (\Theta_{(j,l)})_{(j,l) \in I_{n,p,q}}$ with i.i.d. Rademacher components $\Theta_{(j,l)}$. The aim of Lemma 7 is to prove that, for a well chosen constant C_0 , the random function $f_\Theta - f_{\Theta,1} \otimes f_{\Theta,2}$ belongs to the Sobolev ball $\mathcal{S}_{p+q}^\delta(R, R')$ with high probability.

Lemma 7. Let $\delta > 0$, $R > 0$ and $R' \geq 1$. Let Θ be the random vector $\Theta = (\Theta_{(j,l)})_{(j,l) \in I_{n,p,q}}$ with i.i.d. Rademacher components $\Theta_{(j,l)}$. Consider f_Θ defined by (4.3), where the vector θ is replaced by the random vector Θ . Then, there exists a positive constant $C(p, q, \delta, \eta)$ such that, if $C_0^2 \leq (2\pi)^{p+q} R^2 / [2C(p, q, \delta, \eta)]$, we have that

$$\mathbb{P}(f_\Theta - f_{\Theta,1} \otimes f_{\Theta,2} \in \mathcal{S}_{p+q}^\delta(R)) \geq 1 - \eta.$$

Following Lemma 5, let $P_{\nu_{\rho_n}^*}$ be the probability measure defined for all measurable set A in $\mathbb{R}^{n(p+q)}$ by

$$P_{\nu_{\rho_n}^*}(A) = \int_{\{-1,1\}^{M_n^{p+q}}} P_{f_\theta}(A) \pi(d\theta) = \frac{1}{2^{M_n^{p+q}}} \sum_{\theta \in \{-1,1\}^{M_n^{p+q}}} P_{f_\theta}(A), \quad (4.4)$$

where π is the distribution of a (M_n^{p+q}) -sample of i.i.d. Rademacher random variables. Proposition 5 justifies the use of these alternatives and this probability measure to prove the lower bound.

Proposition 5. Let α, β, η in $(0, 1)$ such that $\alpha + \beta + \eta < 1$, and let $\delta, R > 0$ and $R' \geq 1$. Denote f_0 the uniform density on $[0, 1]^{p+q}$. Assume that $C_0 = C_0(p, q, \delta, R, R', \eta)$ satisfies the assumptions of Lemmas 6 and 7. There exists some positive constant $C(p, q, \alpha, \beta, \delta, R, R', \eta)$ such that, if we set

$$M_n = \left\lceil \frac{1}{C(p, q, \alpha, \beta, \delta, R, R', \eta) n^{-2/(4\delta+p+q)}} \right\rceil \quad \text{and} \quad h_n = \frac{1}{M_n}, \quad (4.5)$$

we have $\nu_{\rho_n}^*(\mathcal{F}_{\rho_n}^*(\mathcal{S}_{p+q}^\delta(R, R')))) \geq 1 - \eta$. Furthermore, if we define $P_{\nu_{\rho_n}^*}$ by Equations (4.3) and (4.4), then we have, for n large enough,

$$\mathbb{E}_{P_{f_0}} \left[\left(\frac{dP_{\nu_{\rho_n}^*}}{dP_{f_0}}(\mathbb{Z}_n) \right)^2 \right] < 1 + 4(1 - \alpha - \eta - \beta)^2.$$

Finally, combining Lemmas 5, 6 and 7 with Proposition 5 leads to a lower bound for the non-asymptotic minimax rate of testing in Theorem 4.

Theorem 4. Consider α, β, η in $(0, 1)$ such that $\alpha + \beta + \eta < 1$. Let $\delta > 0$, $R > 0$ and $R' \geq 1$. Then, there exists a positive constant $C(p, q, \alpha, \beta, \delta, R, R', \eta)$ such that, for n large enough,

$$\rho(\mathcal{S}_{p+q}^\delta(R, R'), \alpha, \beta) \geq C(p, q, \alpha, \beta, \delta, R, R', \eta) n^{-2\delta/(4\delta+p+q)}.$$

Theorem 4 proves that the optimized test $\Delta_\alpha^{\lambda^*, \mu^*}$ introduced in Corollary 2 is optimal in the minimax sense over Sobolev balls since the upper and lower bounds coincide up to constants. Moreover, the aggregated testing procedure defined in Corollary 3 is optimal up to a $\log \log(n)$ term over Sobolev balls. Note that this logarithmic term obtained in the upper bound (3.7) is sometimes unavoidable for adaptivity (c.f. [Ingster, 2000] for the test of uniformity on $[0, 1]$). It seems reasonable to conjecture that it is also the case for independence testing. Hence, since the aggregated testing procedure does not depend on the prior knowledge of the regularity parameter δ , we may conclude that it is adaptive.

5 Numerical simulations

In this section, numerical simulations are performed in order to study the practical validity of our testing procedures. More precisely, we first compare the *theoretical* aggregated test Δ_α defined in (3.2) (studied in theory) and the *permuted* aggregated test $\hat{\Delta}_\alpha$ defined in (3.4) (applied in practice) in terms of power. A similar verification for the single tests, together with a comparison of the power for different bandwidth collections and weights choices, are also carried out in Appendix B.1 in the supplementary material. Then, we compare the *permuted* aggregated test with existing nonparametric independence tests on simulated data.

5.1 Comparison between the theoretical and the permuted aggregated tests

In this section, we numerically illustrate that the power of the permuted aggregated HSIC test approximates very well the power of the theoretical aggregated test, as soon as enough permutations are used to estimate the quantile under the null hypothesis.

All along this section, we rely on the following data generating mechanism inspired from the Ishigami function [Ishigami and Homma, 1990]. Let

$$X = U_1 \quad \text{and} \quad Y = \sin(U_1) + 4 \sin^2(U_2) + 0.5 U_3^4 \sin(U_1). \quad (5.1)$$

where U_1 , U_2 and U_3 are independent uniform random variables on $[0, 1]$.

The practical implementation of the theoretical and permuted aggregated testing procedures are described in Algorithms 1 and 2. They both require the estimation of the value of u_α defined in Equation (3.1). A very straightforward approach to do so is to proceed by dichotomy on the interval $[\alpha, M]$, where $M = \inf_{(\lambda, \mu) \in \mathcal{W}} \{e^{\omega_{\lambda, \mu}}\}$ (u_α belonging to this interval as mentioned in Section 3.1). More precisely, we need to estimate for different values of u , the probability

$$P(u) = P_{f_1 \otimes f_2} \left(\sup_{(\lambda, \mu) \in \mathcal{W}} \left\{ \widehat{\text{HSIC}}_{\lambda, \mu} - q_{1-ue^{-\omega_{\lambda, \mu}}}^{\lambda, \mu} \right\} > 0 \right). \quad (5.2)$$

In the theoretical case, this probability is approached by Monte Carlo independently on the observation (provided that we can simulate under the null hypothesis) whereas in the permuted case, it is based on samples obtained by permuting the observation. The algorithmic complexity of Algorithm 2 is $O((B_1 + B_2) |\mathcal{W}| n^2)$, corresponding to the estimation of the HSIC for all the permutations in Step 1, and all the windows in the collection \mathcal{W} .

Theoretical power For a given sample size n and a given collection of bandwidths \mathcal{W} with associated weights, we estimate the power of the theoretical aggregated test as follows. Since the approximation of the value of u_α and of the quantiles can be done independently of the observation, we run Steps 1 to 4 of Algorithm 1 only once. Then, we generate 1000 i.i.d. samples (observations) and for each one, we apply Step 5 of Algorithm 1. Finally, we estimate the theoretical power by $\hat{\pi}_{th}(n, \alpha)$ which is the proportion of times the aggregated procedure rejects the null hypothesis.

Algorithm 1 *Theoretical aggregated procedure*

Input: The observed n -sample, a prescribed level α , a collection of bandwidths \mathcal{W} and a family of weights $(\omega_{\lambda,\mu})_{(\lambda,\mu) \in \mathcal{W}}$.

1. Simulate a first set, denoted set (A), of 500.000 n -samples under the null hypothesis (to estimate the quantiles) and a second set, denoted (B) of 1000 n -samples also under the null hypothesis (to estimate the probabilities $P(u)$ defined in Equation (5.2) for different values of u).
2. Set $u_{min} = \alpha$ and $u_{max} = M$, where $M = \inf_{(\lambda,\mu) \in \mathcal{W}} \{e^{\omega_{\lambda,\mu}}\}$.
3. While $(u_{max} - u_{min}) > 10^{-3}u_{min}$, repeat the following steps.
 - (a) Set $u = (u_{min} + u_{max})/2$.
 - (b) For all (λ, μ) in \mathcal{W} , compute the Monte Carlo estimator $\tilde{q}_{1-ue^{-\omega_{\lambda,\mu}}}^{\lambda,\mu}$ of the quantile $q_{1-ue^{-\omega_{\lambda,\mu}}}^{\lambda,\mu}$ using the 500.000 samples of set (A).
 - (c) Estimate the probability $P(u)$ by Monte Carlo using the 1000 samples of set (B). More precisely, consider \hat{P}_u as the ratio of times at least one $\widehat{\text{HSIC}}_{\lambda,\mu}$ is greater than $\tilde{q}_{1-ue^{-\omega_{\lambda,\mu}}}^{\lambda,\mu}$.
 - (d) If $\hat{P}_u \leq \alpha$, then set $u_{min} = u$. Else set $u_{max} = u$ and repeat Step 3.
4. Set $\tilde{u}_\alpha = u$ and the quantiles with corrected levels $\left(\tilde{q}_{1-\tilde{u}_\alpha e^{-\omega_{\lambda,\mu}}}^{\lambda,\mu}\right)_{(\lambda,\mu) \in \mathcal{W}}$.
5. Finally, compute the observed statistics $(\widehat{\text{HSIC}}_{\lambda,\mu})_{(\lambda,\mu) \in \mathcal{W}}$ (on the given observation) and reject the null hypothesis if there is at least one (λ, μ) such that

$$\widehat{\text{HSIC}}_{\lambda,\mu} > \tilde{q}_{1-\tilde{u}_\alpha e^{-\omega_{\lambda,\mu}}}^{\lambda,\mu}.$$

Permuted power Unlike the theoretical case, we do not assume we are able to simulate under the null hypothesis to estimate the quantiles and to compute the correction u_α . Note that for the permuted test, Step 3 of Algorithm 2 depends on the observation and needs to be done for each new observation. Hence, for a given sample size n , a given collection of bandwidths \mathcal{W} and associated weights, we generate 1000 i.i.d. samples and for each one, we apply Steps 1 to 5 of Algorithm 2. Finally, we estimate the power of the permuted aggregated test by $\hat{\pi}(n, \alpha, B_1, B_2)$ which is the ratio of times the null hypothesis is rejected.

Numerical results In all the following, the prescribed level of the tests is set to $\alpha = 0.05$ and we consider sample sizes n in $\{50, 100, 200\}$. We consider six different collections of bandwidths $(\mathcal{W}_r)_{2 \leq r \leq 7}$, defined for all r by

$$\mathcal{W}_r = \{1, 1/2, \dots, 1/2^{r-1}\}^2. \quad (5.3)$$

Algorithm 2 *Permuted aggregated procedure*

Input: The observed n -sample \mathbb{Z}_n , a prescribed level α , a collection of bandwidths \mathcal{W} and a family of weights $(\omega_{\lambda,\mu})_{(\lambda,\mu) \in \mathcal{W}}$.

1. Generate a first set, say (A') , of B_1 i.i.d. random permutations of $\{1, \dots, n\}$ (to estimate the quantiles), and independently generate a second set, denoted (B') , of B_2 i.i.d. random permutations of $\{1, \dots, n\}$ (to estimate the probabilities $P(u)$ defined in Equation (5.2)), all independent of \mathbb{Z}_n .
2. Set $u_{min} = \alpha$ and $u_{max} = M$, where $M = \inf_{(\lambda,\mu) \in \mathcal{W}} \{e^{\omega_{\lambda,\mu}}\}$.
3. While $(u_{max} - u_{min}) > 10^{-3}u_{min}$, repeat the following steps.
 - (a) Set $u = (u_{min} + u_{max})/2$.
 - (b) For all (λ, μ) in \mathcal{W} , compute the permuted quantile with Monte Carlo approximation $\hat{q}_{1-ue^{-\omega_{\lambda,\mu}}}^{\lambda,\mu}$ as defined in (2.2) using the set (A') .
 - (c) Estimate $P(u)$ by permutation with Monte Carlo approximation using the set (B') . More precisely, consider

$$\hat{P}_u^*(\mathbb{Z}_n) = \frac{1}{B_2} \sum_{b=1}^{B_2} \mathbb{1}_{\max_{(\lambda,\mu) \in \mathcal{W}} \{\widehat{H}_{\lambda,\mu}^{\kappa_b} - \hat{q}_{1-ue^{-\omega_{\lambda,\mu}}}^{\lambda,\mu}\} > 0},$$

where $(\kappa_b)_{1 \leq b \leq B_2}$ denote the permutations of set (B') and $\widehat{H}_{\lambda,\mu}^{\kappa_b}$ is the statistic computed on the b th permuted sample $\mathbb{Z}_n^{\kappa_b}$, namely $\widehat{\text{HSIC}}_{\lambda,\mu}(\mathbb{Z}_n^{\kappa_b})$.

- (d) If $\hat{P}_u^*(\mathbb{Z}_n) \leq \alpha$, then set $u_{min} = u$. Else set $u_{max} = u$ and repeat Step 3.
4. Set $\hat{u}_\alpha = u$ and the quantiles with corrected levels $(\hat{q}_{1-\hat{u}_\alpha e^{-\omega_{\lambda,\mu}}}^{\lambda,\mu})_{(\lambda,\mu) \in \mathcal{W}}$.
5. Finally, compute the observed statistics $(\widehat{\text{HSIC}}_{\lambda,\mu})_{(\lambda,\mu) \in \mathcal{W}}$ (on the given observation) and reject the null hypothesis if there is at least one (λ, μ) such that

$$\widehat{\text{HSIC}}_{\lambda,\mu} > \hat{q}_{1-\hat{u}_\alpha e^{-\omega_{\lambda,\mu}}}^{\lambda,\mu}.$$

Note that, the case $r = 1$ would correspond to the single test with $\lambda = \mu = 1$. Moreover, for each r , we consider uniform weights defined for all (λ, μ) in the collection \mathcal{W}_r by

$$\omega_{\lambda,\mu} = \log(r^2). \quad (5.4)$$

For the permuted aggregated procedure, the number B_1 of permutations used to estimate the quantiles varies in $\{100, 200, 500, 1000, \dots, 5000\}$ and the number of permutations used to estimate the probabilities $P(u)$ is set to $B_2 = 500$.

For each triplet (r, n, B_1) , the empirical power of both the theoretical and the permuted aggregated testing procedures, respectively denoted $\hat{\pi}_{th}(n, \alpha)$ and $\hat{\pi}(n, \alpha, B_1, B_2)$, are obtained from 1000 different samples as described above. To compare them, we consider the relative absolute

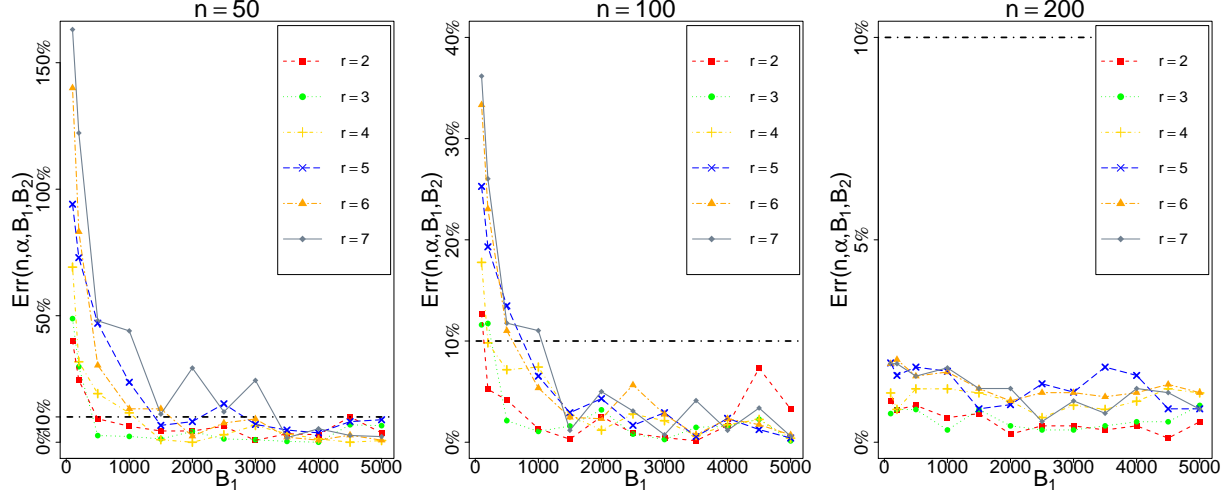


Figure 1: Absolute relative error between the empirical power of the theoretical and the permuted aggregated HSIC procedures w.r.t. the number B_1 of permutations, estimated from 1000 samples of sizes $n = 50, 100$ and 200 , with $B_2 = 500$, the bandwidth collections \mathcal{W}_r and associated weights resp. defined in (5.3) and (5.4) and prescribed level $\alpha = 0.05$.

error defined by

$$Err(n, \alpha, B_1, B_2) = \frac{|\hat{\pi}(n, \alpha, B_1, B_2) - \hat{\pi}_{th}(n, \alpha)|}{\hat{\pi}_{th}(n, \alpha)}.$$

Results are given in Figure 1. Notice that, regardless of the sample size n , the required number B_1 of permutations to well approximate the theoretical power increases with r . In fact, the supremum in Equation (3.3) becomes more difficult to estimate as the number r^2 of aggregated tests increases. Unsurprisingly, for a given B_1 , the accuracy of the power estimation increases with n as in the case of single tests. In particular, we observe that for $n = 50$, the largest error becomes less than 10% from $B_1 = 3500$, while this threshold seems to be achieved from $B_1 = 3000$ for $r = 4, 5, 6$ and from $B_1 = 500$ for $r = 2, 3$. For larger sample sizes $n = 100$ and 200 , a good approximation of the theoretical test seems to be achieved from small values of B_1 , even for a relatively large number of aggregated tests. In particular, for $n = 200$, an error smaller than 10% is reached for all values of B_1 .

All these results show that both theoretical and permuted tests have comparable powers provided that the sample size and the number of permutations are large enough. In the following, we numerically study the power of the permuted tests, which are used in practice.

5.2 Comparison with existing tests

To complete this simulation study, we compare our aggregated procedure with some existing reference tests of independence. For this, we simulate accordingly to the data generating mechanisms of [Berrett and Samworth, 2019], and a basic Gaussian model.

- (i) For l in $\{1, \dots, 10\}$, define the joint density $f_{[l]}$ of (X, Y) for all (x, y) in $[-\pi, \pi]$ by

$$f_{[l]}(x, y) = [1 + \sin(lx) \sin(ly)] / (4\pi^2).$$

- (ii) For l in $\{1, \dots, 10\}$, let $X = L \cos \Theta + \varepsilon_1/4$ and $Y = L \sin \Theta + \varepsilon_2/4$, where L , Θ , ε_1 and ε_2 are independent, with L is uniformly distributed on $\{1, \dots, l\}$, Θ is uniformly distributed on $[0, 2\pi]$ and $\varepsilon_1, \varepsilon_2$ are standard normal random variables.
- (iii) For ρ in $\{0.1, 0.2, \dots, 1\}$, let X be a uniform random variable on $[-1, 1]$ and define $Y = |X|^\rho \varepsilon$, where ε is a standard normal random variable independent with X .
- (iv) For ρ in $\{0, 0.1, \dots, 0.9\}$, let (X, Y) be a centered Gaussian vector such that $\text{Var}(X) = \text{Var}(Y) = 1$ and $\text{Cov}(X, Y) = \rho$.

We also consider the bivariate case $X = (X^{(1)}, X^{(2)})$ and $Y = (Y^{(1)}, Y^{(2)})$ where $(X^{(1)}, Y^{(1)})$ is generated according to mechanisms (i), (ii) or (iii), while $X^{(2)}, Y^{(2)}$ are independent uniform random variables on $[0, 1]$ and independent from $(X^{(1)}, Y^{(1)})$.

The numerical study of the impact of the bandwidth collection and the associated weights on the power of the aggregated procedure done in Appendix B.2 of the supplementary material suggests the following methodological choices. Inspired by usual heuristic bandwidths (see, e.g., [De Lozzo and Marrel, 2017]), define

$$\tilde{\lambda}^2 = \frac{1}{2n(n-1)} \sum_{1 \leq i \neq j \leq n} \|X_i - X_j\|^2 \text{ and } \tilde{\mu}^2 = \frac{1}{2n(n-1)} \sum_{1 \leq i \neq j \leq n} \|Y_i - Y_j\|^2, \quad (5.5)$$

where $\|\cdot\|$ denotes the Euclidean norm. Note that, in the univariate case ($p = q = 1$), $\tilde{\lambda}$ and $\tilde{\mu}$ are the empirical standard deviation of X and Y respectively. In the univariate case, we consider the collections defined by

$$\widetilde{\mathcal{W}} = \{2^{-m}(\tilde{\lambda}, \tilde{\mu}); 0 \leq m \leq 6\}. \quad (5.6)$$

Similarly, in the bivariate case, the bandwidth collections are defined by

$$\widetilde{\mathcal{W}} = \{2^{-m}(\tilde{\lambda}, \tilde{\lambda}, \tilde{\mu}, \tilde{\mu}); 0 \leq m \leq 6\}. \quad (5.7)$$

We also consider exponential weights, that are defined, by analogy with Equation (3.6), for all bandwidths $2^{-m}(\tilde{\lambda}, \tilde{\mu})$ or $2^{-m}(\tilde{\lambda}, \tilde{\lambda}, \tilde{\mu}, \tilde{\mu})$ as

$$\omega_{[m]} = 2 \log(m+1) + \log \left(\sum_{m'=0}^6 \frac{1}{(m'+1)^2} \right). \quad (5.8)$$

Note that the last term in (5.8) ensures that $\sum_{(\lambda, \mu) \in \widetilde{\mathcal{W}}} e^{-\omega_{\lambda, \mu}} = 1$.

In Figure 2, we compare our permutation-based aggregated HSIC test with the mutual information test (MINT) of [Berrett and Samworth, 2019] implemented in the R package `IndepTest`, the permutation-based HSIC single test (HSIC) implemented in the R package `dHSIC` [Pfister et al., 2018] with $B = 1000$ permutations, the distance covariance of [Székely et al., 2007] implemented in the R package `energy`, the test of [Heller et al., 2016] (HHG), the D -test of [Hoeffding, 1948] implemented in the R package `Hmisc` and the binary expansion test (BET) of [Zhang, 2019].

For each example, we simulate samples with size $n = 200$. In line with the results obtained in Section 5.1, Algorithm 2 is applied with $B_1 = 3000$ and $B_2 = 500$. The power of the different tests

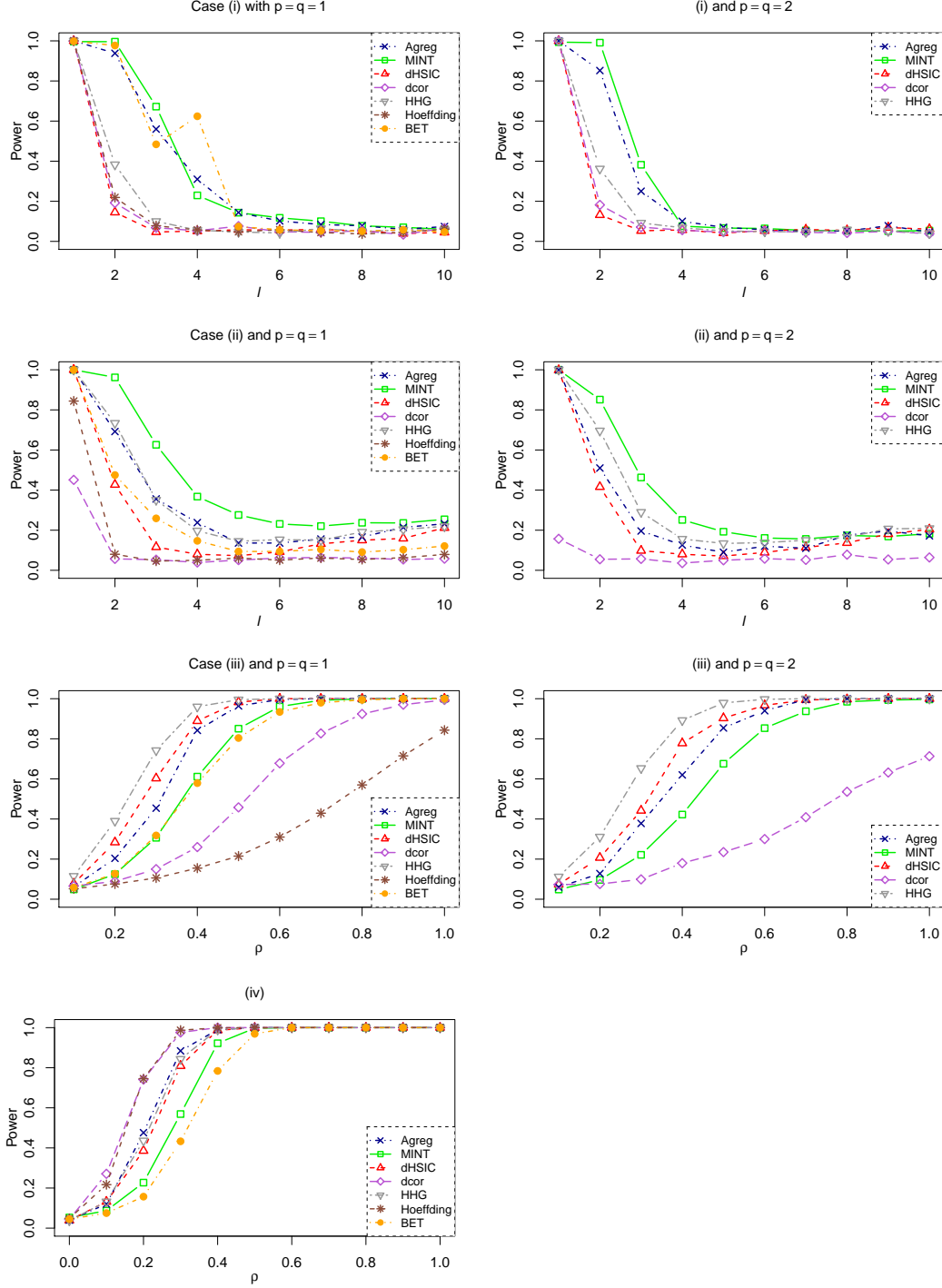


Figure 2: Power curves of the permuted aggregated HSIC test with $B_1 = 3000$ and $B_2 = 500$, bandwidth collection $\bar{\mathcal{W}}$ defined in (5.6) or (5.7) and exponential weights (5.8). It is compared to the MINT, the single HSIC test, the distance covariance, Heller, Heller and Gorfine's test (HHG), Hoeffding's D -test and the BET. The empirical power is estimated from 1000 samples of size $n = 200$. The prescribed level is $\alpha = 0.05$.

is estimated using 1000 different samples of (X, Y) and is represented w.r.t l for simulated data from (i) and (ii) and w.r.t ρ for (iii) and (iv).

As expected, no procedure of testing constantly yields the best performances in all cases. Indeed, it is well known that no uniformly most powerful test exists. However, as for the MINT procedure, the HSIC aggregated procedure seems to yield competitive results on all examples, contrarily to other procedures such as the distance covariance or Hoeffding’s D -test which perform very well in the Gaussian case, but are not very powerful in the other cases. Moreover, in most cases, the HSIC aggregated procedure performs better than the single HISC test, which illustrates the benefits of aggregation.

SUPPLEMENTARY MATERIAL

Supplement to “Adaptive test of independence based on HSIC measures”

This Supplement contains sharp upper bounds for the uniform separation rates over Nikol’skii-Besov balls, a further numerical study and all the proofs.

References

- [Ahmad and Li, 1997] Ahmad, I. A. and Li, Q. (1997). Testing independence by nonparametric kernel method. *Statistics & probability letters*, 34(2):201–210.
- [Albert, 2015] Albert, M. (2015). *Tests of independence by bootstrap and permutation: an asymptotic and non-asymptotic study. Application to neurosciences*. PhD thesis, Université Nice Sophia Antipolis.
- [Aronszajn, 1950] Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68(3):337–404.
- [Bach and Jordan, 2002] Bach, F. R. and Jordan, M. I. (2002). Kernel independent component analysis. *Journal of Machine Learning Research*, 3(Jul):1–48.
- [Baker, 1973] Baker, C. R. (1973). Joint measures and cross-covariance operators. *Transactions of the American Mathematical Society*, 186:273–289.
- [Balakrishnan et al., 2019] Balakrishnan, S., Wasserman, L., et al. (2019). Hypothesis testing for densities and high-dimensional multinomials: Sharp local minimax rates. *The Annals of Statistics*, 47(4):1893–1927.
- [Baraud, 2002] Baraud, Y. (2002). Non-asymptotic minimax rates of testing in signal detection. *Bernoulli*, 8(5):577–606.
- [Baraud et al., 2003] Baraud, Y., Huet, S., and Laurent, B. (2003). Adaptive tests of linear hypotheses by model selection. *The Annals of Statistics*, 31(1):225–251.
- [Berrett et al., 2020] Berrett, T. B., Kontoyiannis, I., and Samworth, R. J. (2020). Optimal rates for independence testing via U -statistic permutation tests. *arXiv preprint arXiv:2001.05513*.
- [Berrett and Samworth, 2019] Berrett, T. B. and Samworth, R. J. (2019). Nonparametric independence testing via mutual information. *Biometrika*, 106(3):547–566.

- [Butucea, 2007] Butucea, C. (2007). Goodness-of-fit testing and quadratic functional estimation from indirect observations. *The Annals of Statistics*, 35(5):1907–1930.
- [De Lozzo and Marrel, 2017] De Lozzo, M. and Marrel, A. (2017). Sensitivity analysis with dependence and variance-based measures for spatio-temporal numerical simulators. *Stochastic environmental research and risk assessment*, 31(6):1437–1453.
- [Fromont et al., 2006] Fromont, M., Laurent, B., et al. (2006). Adaptive goodness-of-fit tests in a density model. *The Annals of Statistics*, 34(2):680–720.
- [Fromont et al., 2013] Fromont, M., Laurent, B., and Reynaud-Bouret, P. (2013). The two-sample problem for Poisson processes: Adaptive tests with a nonasymptotic wild bootstrap approach. *The Annals of Statistics*, 41(3):1431–1461.
- [Fukumizu et al., 2004] Fukumizu, K., Bach, F. R., and Jordan, M. I. (2004). Dimensionality reduction for supervised learning with reproducing kernel Hilbert spaces. *Journal of Machine Learning Research*, 5(Jan):73–99.
- [Fukumizu et al., 2008] Fukumizu, K., Gretton, A., Sun, X., and Schölkopf, B. (2008). Kernel measures of conditional dependence. In *Advances in Neural Information Processing Systems*, pages 489–496.
- [Gretton et al., 2005a] Gretton, A., Bousquet, O., Smola, A., and Schölkopf, B. (2005a). Measuring statistical dependence with Hilbert-Schmidt norms. In *International conference on algorithmic learning theory*, pages 63–77. Springer.
- [Gretton et al., 2008] Gretton, A., Fukumizu, K., Teo, C. H., Song, L., Schölkopf, B., and Smola, A. J. (2008). A kernel statistical test of independence. In *Advances in Neural Information Processing Systems*, pages 585–592.
- [Gretton et al., 2005b] Gretton, A., Herbrich, R., Smola, A., Bousquet, O., and Schölkopf, B. (2005b). Kernel methods for measuring independence. *Journal of Machine Learning Research*, 6(Dec):2075–2129.
- [Gretton et al., 2003] Gretton, A., Herbrich, R., and Smola, A. J. (2003). The kernel mutual information. In *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP’03). 2003 IEEE International Conference on*, volume 4, pages IV–880. IEEE.
- [Gretton et al., 2005c] Gretton, A., Smola, A. J., Bousquet, O., Herbrich, R., Belitski, A., Augath, M., Murayama, Y., Pauls, J., Schölkopf, B., and Logothetis, N. K. (2005c). Kernel constrained covariance for dependence measurement. In *AISTATS*, volume 10, pages 112–119.
- [Heller et al., 2016] Heller, R., Heller, Y., Kaufman, S., Brill, B., and Gorfine, M. (2016). Consistent distribution-free K -sample and independence tests for univariate random variables. *The Journal of Machine Learning Research*, 17(1):978–1031.
- [Hoeffding, 1948] Hoeffding, W. (1948). A non-parametric test of independence. *The Annals of Mathematical Statistics*, pages 546–557.
- [Ingster, 1989] Ingster, Y. I. (1989). An asymptotically minimax test of the hypothesis of independence. *J. Soviet Math*, 44:466–476.

- [Ingster, 1993a] Ingster, Y. I. (1993a). Asymptotically minimax hypothesis testing for nonparametric alternatives. i, ii, iii. *Math. Methods Statist*, 2(2):85–114.
- [Ingster, 1993b] Ingster, Y. I. (1993b). Minimax testing of the hypothesis of independence for ellipsoids in l_p . *Zapiski Nauchnykh Seminarov POMI*, 207:77–97.
- [Ingster, 2000] Ingster, Y. I. (2000). Adaptive chi-square tests. *Journal of Mathematical Sciences*, 99(2):1110–1119.
- [Ingster and Suslina, 1998] Ingster, Y. I. and Suslina, I. A. (1998). Minimax detection of a signal for Besov bodies and balls. *Problemy Peredachi Informatsii*, 34(1):56–68.
- [Ishigami and Homma, 1990] Ishigami, T. and Homma, T. (1990). An importance quantification technique in uncertainty analysis for computer models. In *[1990] Proceedings. First International Symposium on Uncertainty Modeling and Analysis*, pages 398–403. IEEE.
- [Jacod and Protter, 2012] Jacod, J. and Protter, P. (2012). *Probability essentials*. Springer Science & Business Media.
- [Kim et al., 2020] Kim, I., Balakrishnan, S., and Wasserman, L. (2020). Minimax optimality of permutation tests. *arXiv preprint arXiv:2003.13208*.
- [Laurent et al., 2012] Laurent, B., Loubes, J.-M., and Marteau, C. (2012). Non asymptotic minimax rates of testing in signal detection with heterogeneous variances. *Electronic Journal of Statistics*, 6:91–122.
- [Lee et al., 2019] Lee, D., Zhang, K., and Kosorok, M. R. (2019). Testing independence with the binary expansion randomized ensemble test. *arXiv preprint arXiv:1912.03662*.
- [Li and Yuan, 2019] Li, T. and Yuan, M. (2019). On the optimality of gaussian kernel based nonparametric tests against smooth alternatives. *arXiv preprint arXiv:1909.03302*.
- [Marrel et al., 2020] Marrel, A., Raguet, H., and Chabridon, V. (2020). Statistical developments for target and conditional sensitivity analysis: application on safety studies for nuclear reactor. *HAL preprint hal-02541142v2*.
- [Micchelli et al., 2006] Micchelli, C. A., Xu, Y., and Zhang, H. (2006). Universal kernels. *Journal of Machine Learning Research*, 7(Dec):2651–2667.
- [Parzen, 1962] Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076.
- [Pfister et al., 2018] Pfister, N., Bühlmann, P., Schölkopf, B., and Peters, J. (2018). Kernel-based tests for joint independence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(1):5–31.
- [Póczos et al., 2012] Póczos, B., Ghahramani, Z., and Schneider, J. (2012). Copula-based kernel dependency measures. *arXiv preprint arXiv:1206.4682*.
- [Ramdas et al., 2016] Ramdas, A., Isenberg, D., Singh, A., and Wasserman, L. (2016). Minimax lower bounds for linear independence testing. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 965–969. IEEE.

- [Romano and Wolf, 2005] Romano, J. P. and Wolf, M. (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association*, 100(469):94–108.
- [Rosenblatt, 1975] Rosenblatt, M. (1975). A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *The Annals of Statistics*, pages 1–14.
- [Spokoiny, 1996] Spokoiny, V. G. (1996). Adaptive hypothesis testing using wavelets. *The Annals of Statistics*, 24(6):2477–2498.
- [Sriperumbudur et al., 2010] Sriperumbudur, B. K., Gretton, A., Fukumizu, K., Schölkopf, B., and Lanckriet, G. R. (2010). Hilbert space embeddings and metrics on probability measures. *Journal of Machine Learning Research*, 11(Apr):1517–1561.
- [Steinwart, 2001] Steinwart, I. (2001). On the influence of the kernel on the consistency of support vector machines. *Journal of Machine Learning Research*, 2(Nov):67–93.
- [Székely and Rizzo, 2013] Székely, G. J. and Rizzo, M. L. (2013). The distance correlation t -test of independence in high dimension. *Journal of Multivariate Analysis*, 117:193–213.
- [Székely et al., 2007] Székely, G. J., Rizzo, M. L., and Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, 35(6):2769–2794.
- [Weihs et al., 2018] Weihs, L., Drton, M., and Meinshausen, N. (2018). Symmetric rank covariances: a generalized framework for nonparametric measures of dependence. *Biometrika*, 105(3):547–562.
- [Yao et al., 2018] Yao, S., Zhang, X., and Shao, X. (2018). Testing mutual independence in high dimension via distance covariance. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(3):455–480.
- [Yodé, 2004] Yodé, A. (2004). Asymptotically minimax test of independence. *Mathematical Methods of Statistics*, 13(2):201–234.
- [Yodé, 2011] Yodé, A. (2011). Adaptive minimax test of independence. *Mathematical Methods of Statistics*, 20(3):246.
- [Zhang, 2019] Zhang, K. (2019). Bet on independence. *Journal of the American Statistical Association*, 114(528):1620–1637.
- [Zhang et al., 2011] Zhang, K., Peters, J., Janzing, D., and Schölkopf, B. (2011). Kernel-based conditional independence test and application in causal discovery. In *Proceedings of the 27th Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 804–813. AUAI Press.

SUPPLEMENT TO “ADAPTIVE TEST OF INDEPENDENCE BASED ON HSIC MEASURES”

Appendix [A](#) contains sharp upper bounds for the uniform separation rates over anisotropic Nikol'skii-Besov balls of the single and the aggregated HSIC tests developed in the main article. In Appendix [B](#), further simulations illustrate a comparison in terms of power between the theoretical and permuted single tests on the one hand, and the impact of the bandwidth collection and the weights choices on the power on the other hand. Finally, Appendix [C](#) is devoted to all the proofs.

The references of Equations, Theorems, Propositions, etc, that use only numbers such as (3.1) for instance, refer to the main article *Adaptive test of independence based on HSIC measures*.

A Control of the uniform separation rate over anisotropic Nikol'skii-Besov balls

In this section, we consider anisotropic Nikol'skii-Besov balls which allow to take into account different regularity parameters in each direction in \mathbb{R}^{p+q} . The anisotropic Nikol'skii-Besov ball $\mathcal{N}_{2,d}^\delta(R)$ in dimension d in \mathbb{N}^* , with regularity parameter $\delta = (\delta_1, \dots, \delta_d)$ in $(0, +\infty)^d$ and positive radius R , is defined by

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

where $\lfloor \delta_i \rfloor$ denotes the floor function of δ_i if δ_i is not integer and $\lfloor \delta_i \rfloor = \delta_i - 1$ if δ_i is an integer.

As in the Sobolev case, we study optimality over $\mathcal{N}_{2,d}^\delta(R, R')$ defined by

$$\mathcal{N}_{2,d}^\delta(R, R') = \mathcal{N}_{2,d}^\delta(R) \cap \{ f ; \max \{ \|f\|_\infty, \|f_1\|_\infty, \|f_2\|_\infty \} \leq R' \}. \quad (\text{A.1})$$

As in the Sobolev case, we prove upper bounds for the uniform separation rate of the tests defined in the main article over these new regularity spaces. Section [A.1](#) is devoted to the single test $\Delta_\alpha^{\lambda, \mu}$ with fixed bandwidths defined in Equation (2.1), and the study of the aggregated test Δ_α defined in Equation (3.2) is done in Sections [A.2](#) and [A.3](#).

A.1 Uniform separation rate of the single tests over Nikol'skii-Besov balls

In this section, we consider a fixed bandwidth (λ, μ) . Lemma [8](#) provides an upper bound of the bias term, similar to that of Lemma [3](#), in the case when $f - f_1 \otimes f_2$ belongs to an anisotropic Nikol'skii-Besov ball.

Lemma 8. Let $\psi = f - f_1 \otimes f_2$ and assume that ψ belongs to $\mathcal{N}_{2,p+q}^\delta(R)$, where the regularity parameter $\delta = (\nu_1, \dots, \nu_p, \gamma_1, \dots, \gamma_q)$ belongs to $(0, 2]^{p+q}$. Let φ_λ and ϕ_μ be the functions defined in (1.4). Then, the bias term can be controlled as follows

$$\|\psi - \psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2 \leq C(\delta, R) \left[\sum_{i=1}^p \lambda_i^{2\nu_i} + \sum_{j=1}^q \mu_j^{2\gamma_j} \right].$$

In the Nikol'skii-Besov case, the control of the bias term requires a restriction on the regularity parameter to $(0, 2]^{p+q}$, which comes from the fact that the Gaussian kernel is of order 2. In order to extend the range of the upper bound, kernels of higher order should be considered. This generalization lies beyond the scope of this article and requires further developments. As in Section 2.4, one can deduce from Theorem 1 upper bounds for the uniform separation rates of the single test $\Delta_\alpha^{\lambda, \mu}$ over Nikol'skii-Besov balls.

Theorem 5. Let α, β in $(0, 1)$, $\delta = (\nu_1, \dots, \nu_p, \gamma_1, \dots, \gamma_q)$ in $(0, 2]^{p+q}$ and $R, R' > 0$. Consider bandwidths (λ, μ) satisfying Assumptions $\mathcal{A}_2(\alpha)$ and denote $\Delta_\alpha^{\lambda, \mu}$ the test defined by (2.1). Then, the uniform separation rate defined in (1.1) of the test $\Delta_\alpha^{\lambda, \mu}$ over the Nikol'skii-Besov ball $\mathcal{N}_{2,p+q}^\delta(R, R')$ defined in Equation (A.1) can be upper bounded as follows

$$\begin{aligned} \left[\rho \left(\Delta_\alpha^{\lambda, \mu}, \mathcal{N}_{2,p+q}^\delta(R, R'), \beta \right) \right]^2 &\leq C(\delta, R) \left[\sum_{i=1}^p \lambda_i^{2\nu_i} + \sum_{j=1}^q \mu_j^{2\gamma_j} \right] \\ &\quad + \frac{C(R', p, q, \beta)}{n \sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}} \log \left(\frac{1}{\alpha} \right). \end{aligned} \quad (\text{A.2})$$

where $C(\cdot)$ are positive constants depending only on their arguments.

As in Section 2.4, we can deduce optimal bandwidths (λ^*, μ^*) which minimize the right-hand side of Equation (A.2) and compute an upper bound for the uniform separation rate of the optimized test $\Delta_\alpha^{\lambda^*, \mu^*}$ over Nikol'skii-Besov balls.

Corollary 4. Let α in $(0, 1/e)$, β in $(0, 1)$, $\delta = (\nu_1, \dots, \nu_p, \gamma_1, \dots, \gamma_q)$ in $(0, 2]^{p+q}$ and $R, R' > 0$. Define for all i in $\{1, \dots, p\}$ and for all j in $\{1, \dots, q\}$,

$$\lambda_i^* = n^{-2\eta/[\nu_i(1+4\eta)]} \quad \text{and} \quad \mu_j^* = n^{-2\eta/[\gamma_j(1+4\eta)]} \quad \text{where} \quad \frac{1}{\eta} = \sum_{i=1}^p \frac{1}{\nu_i} + \sum_{j=1}^q \frac{1}{\gamma_j}.$$

If $n \geq (\log(1/\alpha))^{1+1/(4\eta)}$, then, (λ^*, μ^*) satisfy $\mathcal{A}_2(\alpha)$ and the uniform separation rate of the optimized test $\Delta_\alpha^{\lambda^*, \mu^*}$ over the Nikol'skii-Besov ball $\mathcal{N}_{2,p+q}^\delta(R, R')$ is controlled as follows

$$\rho \left(\Delta_\alpha^{\lambda^*, \mu^*}, \mathcal{N}_{2,p+q}^\delta(R, R'), \beta \right) \leq C(p, q, \alpha, \beta, \delta, R, R') n^{-2\eta/(1+4\eta)}. \quad (\text{A.3})$$

Notice that the upper bound obtained for Nikol'skii-Besov balls in Corollary 4 is analogue to that obtained for Sobolev balls in Corollary 2. Indeed, if we consider the same regularities in all directions in the case of Nikol'skii-Besov balls: $\nu_1 = \dots = \nu_p = \gamma_1 = \dots = \gamma_q$, we obtain a similar upper bound. These upper bounds obtained in Corollaries 2 and 4 coincide with the asymptotic minimax

separation rate of testing mutual independence w.r.t. the \mathbb{L}_2 -norm over isotropic Nikol'skii-Besov spaces [Ingster, 1989]. This suggests that the test $\Delta_\alpha^{\lambda^*, \mu^*}$ with optimal bandwidths is optimal in the minimax sense over Nikol'skii-Besov balls with regularity parameter δ in $(0, 2]^{p+q}$. Yet, as in the Sobolev case, it cannot be adaptive since the optimal bandwidths (λ^*, μ^*) depend on the regularity δ .

Finally, note that subsequently, [Kim et al., 2020] also generalized Theorem 5 to the permuted tests, which suggests that the permuted test with bandwidths (λ^*, μ^*) defined in Corollary 4 is optimal in the minimax sense. However, as in the Sobolev case, they only obtain a polynomial dependence in α which is not sharp enough to provide adaptive tests by aggregating as done in Section A.3.

A.2 Oracle-type conditions for the uniform separation rate over Nikol'skii-Besov balls

Theorem 6 is equivalent to Theorem 3 over Nikol'skii-Besov balls and provides an oracle-type inequality for the uniform separation rate of the aggregated testing procedure Δ_α .

Theorem 6. *Let α, β in $(0, 1)$. Consider a finite or countable collection $\mathcal{W} \subset (0, +\infty)^p \times (0, +\infty)^q$ of bandwidths (λ, μ) and a collection of positive weights $\{\omega_{\lambda, \mu}\}_{(\lambda, \mu) \in \mathcal{W}}$ such that $\sum_{(\lambda, \mu) \in \mathcal{W}} e^{-\omega_{\lambda, \mu}} \leq 1$ and such that all (λ, μ) in \mathcal{W} verifies Assumption $\mathcal{A}_2(\alpha e^{-\omega_{\lambda, \mu}})$. Then, the uniform separation rate over Nikol'skii-Besov balls $\mathcal{N}_{2, p+q}^\delta(R, R')$ with $\delta = (\nu_1, \dots, \nu_p, \gamma_1, \dots, \gamma_q)$ in $(0, 2]^{p+q}$ and $R, R' > 0$ of the aggregated test Δ_α defined in Equation (3.2) can be upper bounded as follows*

$$\begin{aligned} \left[\rho \left(\Delta_\alpha, \mathcal{N}_{2, p+q}^\delta(R, R'), \beta \right) \right]^2 \leq C(p, q, \beta, \delta, R, R') \inf_{(\lambda, \mu) \in \mathcal{W}} \left\{ \left[\sum_{i=1}^p \lambda_i^{2\nu_i} + \sum_{j=1}^q \mu_j^{2\gamma_j} \right] \right. \\ \left. + \frac{1}{n \sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}} \left(\log \left(\frac{1}{\alpha} \right) + \omega_{\lambda, \mu} \right) \right\}, \end{aligned}$$

where $C(\cdot)$ is a positive constant depending only on its arguments.

As in the Sobolev case, Theorem 6 can be interpreted as an oracle-type condition for the uniform separation rate of the aggregated test Δ_α over Nikol'skii-Besov balls. Indeed, without knowing the regularity δ of $f - f_1 \otimes f_2$, the uniform separation rate of Δ_α is of the same order as the smallest uniform separation rate of the single tests corresponding to bandwidths (λ, μ) in \mathcal{W} , up to an additional term $\omega_{\lambda, \mu}$ due to the level corrections.

A.3 Control of the uniform separation rate of the aggregated procedure

In this section, we provide an upper bound for the uniform separation rate of the aggregated testing procedure Δ_α over Nikol'skii-Besov balls for the following specific choice of bandwidth collection

and weights. Let

$$\mathcal{W} = \left\{ \left(2^{-m_{1,1}}, \dots, 2^{-m_{1,p}}, 2^{-m_{2,1}}, \dots, 2^{-m_{2,q}} \right), \right. \\ \left. (m_{1,1}, \dots, m_{1,p}, m_{2,1}, \dots, m_{2,q}) \in (\mathbb{N}^*)^{p+q} ; \sum_{i=1}^p m_{1,i} + \sum_{j=1}^q m_{2,j} \leq 2 \log_2 \left(\frac{n}{\log(n)} \right) \right\}, \quad (\text{A.4})$$

In addition, we associate to every bandwidths $(\lambda, \mu) = (2^{-m_{1,1}}, \dots, 2^{-m_{1,p}}, 2^{-m_{2,1}}, \dots, 2^{-m_{2,q}})$ in \mathcal{W} the positive weight

$$\omega_{\lambda, \mu} = 2 \sum_{i=1}^p \log \left(m_{1,i} \times \frac{\pi}{\sqrt{6}} \right) + 2 \sum_{j=1}^q \log \left(m_{2,j} \times \frac{\pi}{\sqrt{6}} \right), \quad (\text{A.5})$$

so that $\sum_{(\lambda, \mu) \in \mathcal{W}} e^{-\omega_{\lambda, \mu}} \leq 1$.

Corollary 5. *Let α, β in $(0, 1)$. Consider the aggregated test Δ_α defined in (3.2), with the particular choice of the collection \mathcal{W} and the weights $(\omega_{\lambda, \mu})_{(\lambda, \mu) \in \mathcal{W}}$ defined in (A.4) and (A.5). Assume that $\log \log(n) > 1$. Then, under the assumptions of Theorem 6, for any $\delta = (\nu_1, \dots, \nu_p, \gamma_1, \dots, \gamma_q)$ in $(0, 2]^{p+q}$ and positive radii R, R' , there exists a positive constant $C(p, q, \alpha, \delta)$ such that for all $n \geq C(p, q, \alpha, \delta)$, the uniform separation rate over the Nikol'skii-Besov ball $\mathcal{N}_{2, p+q}^\delta(R, R')$ of Δ_α can be upper bounded as follows:*

$$\rho \left(\Delta_\alpha, \mathcal{N}_{2, p+q}^\delta(R, R'), \beta \right) \leq C(p, q, \alpha, \beta, \delta, R, R') \left(\frac{\log \log(n)}{n} \right)^{2\eta/(1+4\eta)},$$

where $\frac{1}{\eta} = \sum_{i=1}^p \frac{1}{\nu_i} + \sum_{j=1}^q \frac{1}{\gamma_j}$.

As in the case of Sobolev regularity, according to Corollary 5, the uniform separation rate of the aggregated procedure over Nikol'skii-Besov balls is of the same order as the one of the optimized test $\Delta_\alpha^{\lambda^*, \mu^*}$ (given in Corollary 4), up to a $\log \log(n)$ factor which is, once again a usual price to pay for aggregated tests (see, e.g., [Spokoiny, 1996, Ingster, 2000].)

B Further numerical simulations

B.1 Single tests comparison

Similarly to Section 5.1, the objective here is to check that the permutation approach does not impact the power of the single HSIC test. To do so, we numerically illustrate that the power of the permuted single HSIC tests approximates very well the power of the theoretical tests, as soon as enough permutations are used for the estimation of the quantile under the null hypothesis.

In order to evaluate the accuracy of permuted single HSIC tests, we choose the kernel bandwidth associated to X (resp. Y) to be the empirical standard deviation s (resp. s') of X (resp. Y), which is a usual choice in the literature on single HSIC-test (see, e.g. [De Lozzo and Marrel, 2017]).

As in Section 5.1, we rely on the data generating mechanism inspired from the Ishigami function [Ishigami and Homma, 1990] defined in (5.1). In the following, we illustrate the power for the three sample sizes n in $\{50, 100, 200\}$ and the two levels α in $\{0.05, 0.001\}$.

For each sample size n and level α , we first estimate the power of the *theoretical* test. To achieve this, we simulate 500.000 n -samples under the null hypothesis¹ and compute the Monte Carlo estimator, denoted $\hat{q}_{1-\alpha}^{MC}$, of the theoretical $(1 - \alpha)$ -quantile of $\widehat{\text{HSIC}}_{s,s'}$ under the null hypothesis. Then, we generate 1000 different n -samples of (X, Y) under the alternative according to (5.1) and we estimate the power of the *theoretical* test by $\hat{\pi}_{th}(n, \alpha)$ which is the ratio of times that the observed test statistic $\widehat{\text{HSIC}}_{s,s'}$ exceeds the quantile $\hat{q}_{1-\alpha}^{MC}$.

The second step consists in estimating the power of the *permuted* tests for several values of the number of permutations B . The chosen values of B are $\{10, 20, \dots, 100, 200, \dots, 2500\}$. For each value of n , α and B , we generate 1000 n -sample of (X, Y) according to (5.1). For each n -sample, we compute the permuted quantile $\hat{q}_{1-\alpha}$ defined in Equation (2.2) using B random permutations of this sample. Thereafter, we estimate the power of the *permuted* test, by $\hat{\pi}(n, \alpha, B)$ which is the ratio of times the value of $\widehat{\text{HSIC}}_{s,s'}$ exceeds the permuted quantile $\hat{q}_{1-\alpha}^{s,s'}$ (computed on the corresponding sample).

As in Section 5.1, to compare the empirical powers of *theoretical* and *permuted* tests (resp. $\hat{\pi}_{th}(n, \alpha)$ and $\hat{\pi}(n, \alpha, B)$), we consider the relative absolute error $Err(n, \alpha, B)$ defined as

$$Err(n, \alpha, B) = \frac{|\hat{\pi}(n, \alpha, B) - \hat{\pi}_{th}(n, \alpha)|}{\hat{\pi}_{th}(n, \alpha)}.$$

The results obtained for $\alpha = 0.05$ and different n values are given by Figure 3. We can see that the accuracy of the permuted approach tends to increase as n increases. This is probably due to the fact that the power of the theoretical test increases as the sample size increases. Another explanation may be that, on the one hand, the power of the theoretical test is more difficult to estimate for small sample sizes, which explains the fluctuations observed for $n = 50$. On the other hand, as n increases, the approximation of the distribution of $\widehat{\text{HSIC}}_{s,s'}$ under the null hypothesis based on B permutations becomes more accurate, and this for any value of B larger than 500. Hence, the approximation of the quantile by permutation becomes more accurate, and thus, there are less fluctuations for larger sample sizes.

Generally, the permutation approach allows to obtain the power of the theoretical test with an acceptable precision, even for small values of B . In particular, we observe for $n = 50$ that aside from very small values of B and two outliers, the absolute relative error is always less than 10%. Moreover, from $n = 100$ this error is mostly less than 10% and no observed error is greater than 5% for $n = 200$.

Since the aggregated procedure requires an individual level correction, we also study the impact of the level on the accuracy of the permutation approximation. We show in Figure 4 the relative absolute error of the power w.r.t. n and B for the extreme level value $\alpha = 0.001$. Contrary to the case $\alpha = 0.05$, we observe here much less precision of the power approximation. In particular, for $n = 50$, $B = 2000$ permutations are required to obtain satisfactory accuracy (against $B = 30$ for $\alpha = 0.05$). Similar observations are done for $n = 100$ and 200 with respectively $B = 1200$ and $B = 500$ permutations required (against $B = 30$ and $B = 10$ for $\alpha = 0.05$). This slow convergence

¹To generate an independent n -sample of (X, Y) under the null hypothesis, we first generate an independent $2n$ -sample of (X, Y) according to (5.1). Only the first n elements are used to compute the marginal sample of Y and the remaining n elements are considered to be the marginal sample of X .

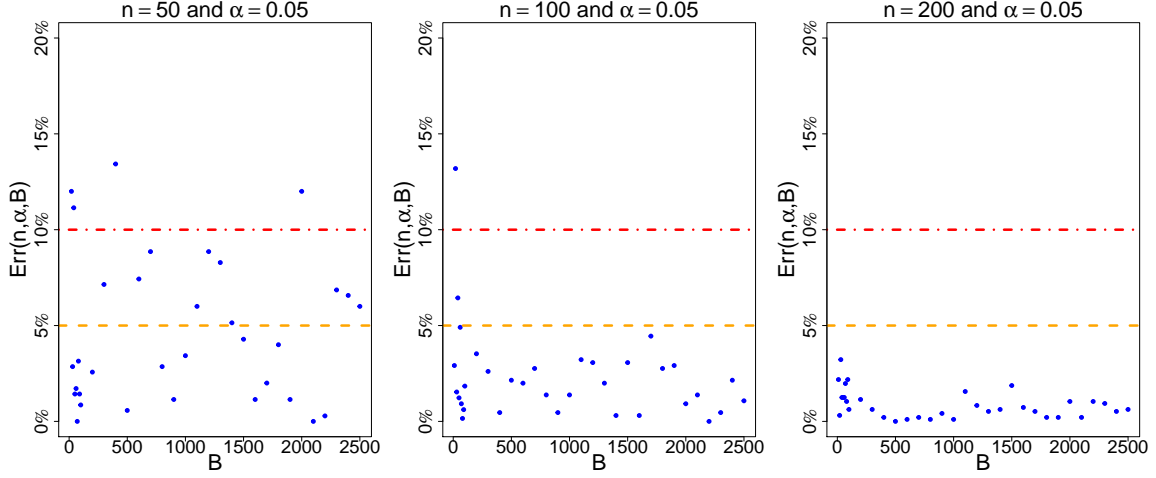


Figure 3: Absolute relative error between the empirical powers of the theoretical and permuted HSIC-tests, w.r.t the number B of permutations, for samples generated according to Ishigami's data generating mechanism defined in Equation (5.1) with sizes $n = 50, 100$ and 200 . The presumed level is $\alpha = 0.05$. The red (resp. orange) dashed line represents the error threshold of 10% (resp. 5%).

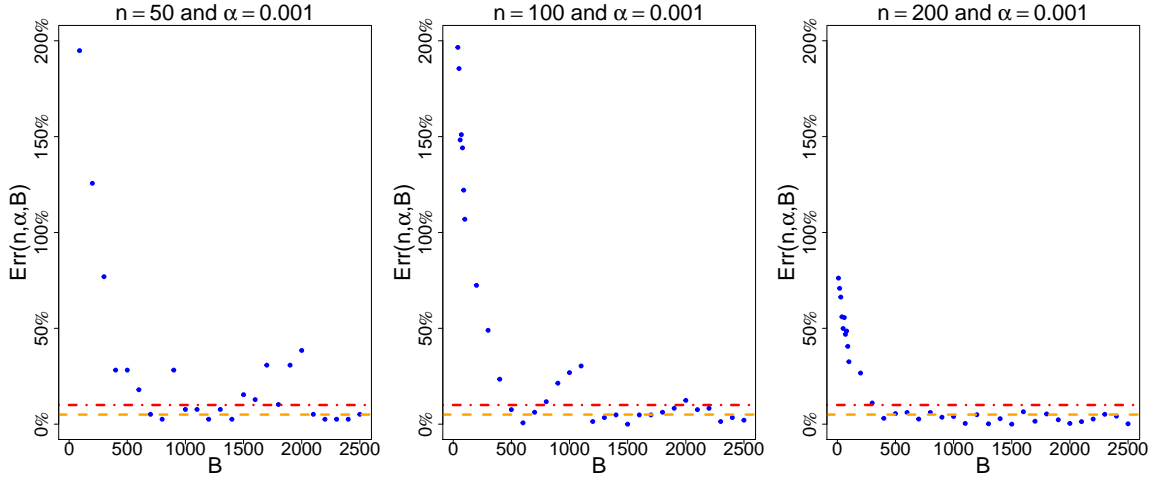


Figure 4: Absolute relative error between the empirical powers of the theoretical and the permuted HSIC-tests, w.r.t the number B of permutations, for samples generated according to Ishigami's data generating mechanism defined in Equation 5.1 of Section 5.1 with sizes $n = 50, 100$ and 200 . The presumed level of tests is $\alpha = 0.001$. The red (resp. orange) dashed line represents the error threshold of 10% (resp. 5%).

results from the difficulty of estimating extreme quantiles. Moreover, this phenomenon seems more significant for small sample sizes. Indeed, as in the previous case, the lowest the power of the test, the biggest its sensitivity to the quantile estimation error.

Similar results for the aggregated procedures are illustrated in Section 5.1 of the main article.

B.2 Choice of the collections of bandwidths and the weights

In our aggregated procedure, the collection of bandwidths \mathcal{W} , together with the weights have to be chosen. There is no universal best collection that would ensure optimal test power. To determine the collection, we first study the impact of the bandwidth choice on single HSIC-based tests. This leads us to particular forms of collections. Then, we investigate different choices of the collections \mathcal{W} and together with different weights (including the single test case).

B.2.1 Impact of the bandwidths choice on the power of the single tests

The optimal bandwidth depends on the intrinsic characteristics of X and Y and their dependence structure. Consequently, it seems relevant to consider the possible bandwidths relatively to the standard deviations of X and Y . Moreover, as already mentioned, the standard deviation is a usual choice for the bandwidth in the literature on single HSIC-test. We assume here that the exact values of standard deviations of X and Y , respectively denoted s and s' , are known. In such a way, we are able to construct collections which do not depend on the observation. In practice, when only a n -sample of (X, Y) is available, we estimate these standard deviations by the usual empirical estimators. Practice shows that the effect of this estimation does not significantly impact the single tests performance. Indeed, standard deviation estimators converge in most cases rapidly w.r.t. n . More particularly, this estimation error is small compared to the estimation error of the quantiles.

For this, we consider the univariate mechanism of dependence (ii) with $l = 2$ defined in Section 5.2. Moreover, we consider, as possible bandwidths λ and μ , multiple or dyadic fractions of s and s' respectively. For each couple (λ, μ) , the power of the permuted single HSIC tests (with $B = 1000$) is estimated as explained above. Figure 5 shows the obtained power maps w.r.t. (λ, μ) , for different sample sizes. First, we can observe that the bandwidths significantly impact the power: in this case, there is an optimal area around $(\lambda, \mu) = (s/4, s'/4)$ with a power close to one for $n = 200$. The power decreases progressively as we move away from this area, until being null for very high and very low values of bandwidths. We can also see that the regularity of the maps increases with the sample size (just like the power for each point). Similar conclusions can be observed for other values of l and the other data generating mechanisms (i) and (iii) with one or several areas with higher power, but are not presented here. These results illustrate that an arbitrary choice of bandwidths is not relevant and justify the interest of considering several bandwidths through an aggregation strategy. Note that, according to our experience, it might be appropriate to consider bandwidths higher than standard deviations. However, in Section B.2.2, we consider aggregating procedures based on collections $\mathcal{W}_r^{s, s'}$ of types

$$\mathcal{W}_r^{s, s'} = \{s, s/2, \dots, s/2^{r-1}\} \times \{s', s'/2, \dots, s'/2^{r-1}\}, \quad (\text{B.1})$$

where r belongs to \mathbb{N}^* . Note that in this univariate case, these collections generalize to other sizes r , in an anisotropic way, the ones considered in Section 5.2 introduced in Equation (5.6).

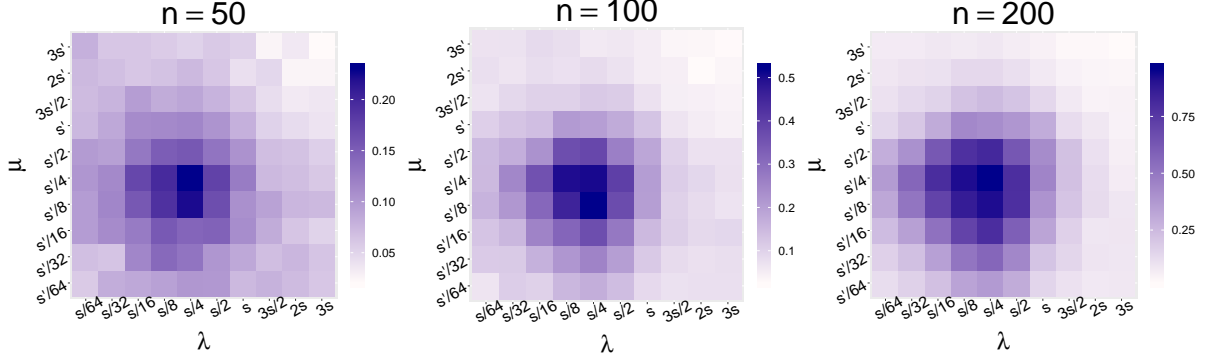


Figure 5: Power map of the permuted single HSIC test w.r.t. to kernel bandwidths λ and μ respectively associated to X and Y , for sample generated according to the univariate mechanism of dependence (ii) with $l = 2$ defined in Section 5.2 with sizes $n = 50, 100$ and 200 , $B = 1000$ and $\alpha = 0.05$.

B.2.2 Impact of the weights choice on the power of the aggregated procedure

Following the results of Section B.2.1, we consider bandwidth collections $\mathcal{W}_r^{s,s'}$ as defined in Equation (B.1), where s and s' are respectively the empirical standard deviations of the X_i 's and the Y_i 's. By now, let us compare two possible choices of weights: uniform and exponential weights. On the one hand, we recall that uniform weights depend only on the cardinality of the collection, and are defined in Equation (5.4) for all (λ, μ) in $\mathcal{W}_r^{s,s'}$ by

$$\omega_{\lambda,\mu} = \log(r^2).$$

On the other hand, in analogy with Equation (A.5), we consider the exponential weights defined for all bandwidths $(s/2^{m_1}, s'/2^{m_2})$ in $\mathcal{W}_r^{s,s'}$ by

$$\omega_{s/2^{m_1}, s'/2^{m_2}} = 2 \log(m_1 + 1) + 2 \log(m_2 + 1) + \log \left(\sum_{1 \leq u, v \leq r} \frac{1}{u^2 v^2} \right).$$

The results obtained with the two types of weights are given in Figure 6, for different values of r and sample sizes n . In this case, the uniform weights seem to give a better power than the exponential ones. However, we can observe a different behavior w.r.t. r . For the uniform weights, the power increases until a specific r ($r = 3$ or 4 w.r.t n), before decreasing with r , to being lower than the power with exponential weights. On the contrary, the power with exponential weights has a more robust behavior, since it increases with r until it stabilizes. This is a crucial advantage in favor of exponential weights, as the optimal r is unknown in practice. It prevents deterioration of the quality of the test, when too large collection sizes have been chosen. We can also observe that the two aggregated strategies yield a greater power than the single test (which corresponds to the case $r = 1$), as soon as the collection \mathcal{W} is large enough.

Similar conclusions have been drawn from the other analytical examples, which are not presented here for the sake of brevity. Thus, from our experience, we recommend in practice the use of the aggregated procedure with exponential weights with $r = 5$ or 6 .

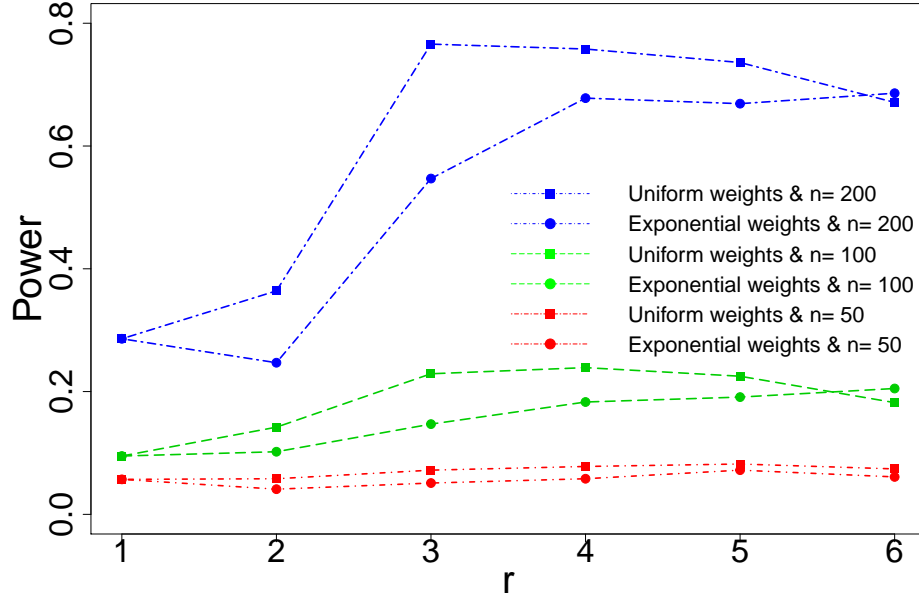


Figure 6: Empirical power of the permuted aggregated procedures with uniform and exponential weights, w.r.t. the number r of aggregated bandwidths in each direction, for samples generated according to the univariate mechanism of dependence (ii) with $l = 2$ defined in Section 5.2 of sizes $n = 50, 100$ and 200 , $B_1 = 3000$, $B_2 = 500$ and $\alpha = 0.05$.

C Proofs

All along the proofs, we set $Z = (X, Y)$ and $Z_i = (X_i, Y_i)$ for all i in $\{1, \dots, n\}$. We also denote by A, B and C positive universal constants whose values may change from line to line. Moreover, the generic notation $C(a, b, \dots)$ denotes a positive constant depending only on its arguments (a, b, \dots) and that may vary from line to line.

C.1 Proof of Proposition 1

Let α be in $(0, 1)$. In order to prove that the permuted test with Monte Carlo approximation $\widehat{\Delta}_\alpha^{\lambda, \mu}$ defined in Equation (2.3) is of prescribed level α , we use Lemma 1 of [Romano and Wolf, 2005] recalled here.

Lemma 9 ([Romano and Wolf, 2005, Lemma 1]). *Let R_1, \dots, R_{B+1} be $(B + 1)$ exchangeable random variables. Then, for all u in $(0, 1)$*

$$\mathbb{P}\left(\frac{1}{B+1} \left[1 + \sum_{b=1}^B \mathbb{1}_{R_b \geq R_{B+1}}\right] \leq u\right) \leq u.$$

Recall that for all $1 \leq b \leq B$,

$$\widehat{H}_{\lambda, \mu}^{\star b} = \widehat{\text{HSIC}}_{\lambda, \mu}(\mathbb{Z}_n^{\tau_b}) \quad \text{and} \quad \widehat{H}_{\lambda, \mu}^{\star B+1} = \widehat{\text{HSIC}}_{\lambda, \mu}(\mathbb{Z}_n) = \widehat{\text{HSIC}}_{\lambda, \mu}(\mathbb{Z}_n^{\tau_{B+1}}),$$

where $\tau_{B+1} = \text{id}$ is the identity permutation of $\{1, \dots, n\}$ (deterministic).

Assume that $f = f_1 \otimes f_2$. Then the random variables $\widehat{H}_{\lambda,\mu}^{\star 1}, \dots, \widehat{H}_{\lambda,\mu}^{\star B}$ and $\widehat{H}_{\lambda,\mu}^{\star B+1}$ are exchangeable. Indeed, let π be a (deterministic) permutation of $\{1, \dots, B+1\}$ and let us prove that

$$\left(\widehat{H}_{\lambda,\mu}^{\star 1}, \dots, \widehat{H}_{\lambda,\mu}^{\star B}, \widehat{H}_{\lambda,\mu}^{\star B+1}\right) \quad \text{and} \quad \left(\widehat{H}_{\lambda,\mu}^{\star \pi(1)}, \dots, \widehat{H}_{\lambda,\mu}^{\star \pi(B+1)}\right) \quad \text{have the same distribution.} \quad (\text{C.1})$$

Case 1. If $\pi(B+1) = B+1$, then, since the permutations $(\tau_b)_{1 \leq b \leq B}$ are i.i.d., they are exchangeable. Hence, $(\tau_{\pi(1)}, \dots, \tau_{\pi(B)})$ is an i.i.d. sample of uniform permutations of $\{1, \dots, n\}$, independent of \mathbb{Z}_n and (C.1) holds by construction.

Case 2. If $\pi(B+1) \neq B+1$, then

$$\widehat{H}_{\lambda,\mu}^{\star \pi(B+1)} = \widehat{\text{HSIC}}_{\lambda,\mu} \left(\mathbb{Z}_n^{\tau_{\pi(B+1)}} \right) = \widehat{\text{HSIC}}_{\lambda,\mu} \left(\tilde{\mathbb{Z}}_n \right), \quad \text{where} \quad \tilde{\mathbb{Z}}_n = \mathbb{Z}_n^{\tau_{\pi(B+1)}}.$$

In particular, for all b in $\{1, \dots, B\}$,

$$\begin{cases} \widehat{H}_{\lambda,\mu}^{\star \pi(b)} = \widehat{\text{HSIC}}_{\lambda,\mu} \left(\mathbb{Z}_n^{\tau_{\pi(b)}} \right) = \widehat{\text{HSIC}}_{\lambda,\mu} \left(\tilde{\mathbb{Z}}_n^{\tau_{\pi(b)} \circ \tau_{\pi(B+1)}^{-1}} \right) & \text{if } \pi(b) \neq B+1, \\ \widehat{H}_{\lambda,\mu}^{\star \pi(b)} = \widehat{\text{HSIC}}_{\lambda,\mu} \left(\mathbb{Z}_n \right) = \widehat{\text{HSIC}}_{\lambda,\mu} \left(\tilde{\mathbb{Z}}_n^{\text{id} \circ \tau_{\pi(B+1)}^{-1}} \right) & \text{if } \pi(b) = B+1. \end{cases}$$

Therefore, in order to prove (C.1), it is sufficient to prove that $\{\tau_{\pi(1)} \circ \tau_{\pi(B+1)}^{-1}, \dots, \tau_{\pi(B)} \circ \tau_{\pi(B+1)}^{-1}\}$ is an i.i.d. sample of uniform permutations of $\{1, \dots, n\}$ independent of $\tilde{\mathbb{Z}}_n$. Let A be a mesurable set, and $\sigma_1, \dots, \sigma_B$ be (fixed) permutations of $\{1, \dots, n\}$. Then

$$\begin{aligned} & \mathbb{P} \left(\tilde{\mathbb{Z}}_n \in A, \tau_{\pi(1)} \circ \tau_{\pi(B+1)}^{-1} = \sigma_1, \dots, \tau_{\pi(B)} \circ \tau_{\pi(B+1)}^{-1} = \sigma_B \right) \\ &= \mathbb{P} \left(\mathbb{Z}_n^{\tau_{\pi(B+1)}} \in A, \tau_{\pi(1)} = \sigma_1 \circ \tau_{\pi(B+1)}, \dots, \tau_{\pi(B)} = \sigma_B \circ \tau_{\pi(B+1)} \right) \\ &= \mathbb{E} \left[\mathbb{P} \left(\mathbb{Z}_n^{\tau_{\pi(B+1)}} \in A, \tau_{\pi(1)} = \sigma_1 \circ \tau_{\pi(B+1)}, \dots, \tau_{\pi(B)} = \sigma_B \circ \tau_{\pi(B+1)} \middle| \tau_{\pi(B+1)} \right) \right]. \end{aligned}$$

This leads to

$$\begin{aligned} & \mathbb{P} \left(\tilde{\mathbb{Z}}_n \in A, \tau_{\pi(1)} \circ \tau_{\pi(B+1)}^{-1} = \sigma_1, \dots, \tau_{\pi(B)} \circ \tau_{\pi(B+1)}^{-1} = \sigma_B \right) \\ &= \mathbb{E} \left[\mathbb{P}(\mathbb{Z}_n \in A) \times \left(\prod_{\substack{b=1 \\ b \neq \pi^{-1}(B+1)}}^B \mathbb{P}(\tau_{\pi(b)} = \sigma_b \circ \tau_{\pi(B+1)} \middle| \tau_{\pi(B+1)}) \right) \times \right. \\ & \quad \left. \mathbb{P}(\text{id} = \sigma_{\pi^{-1}(B+1)} \circ \tau_{\pi(B+1)} \middle| \tau_{\pi(B+1)}) \right], \quad (\text{C.2}) \end{aligned}$$

where (C.2) holds by independence of all permutations τ_b and of \mathbb{Z}_n and since, if $f = f_1 \otimes f_2$,

$\mathbb{Z}_n^{\tau_{\pi(B+1)}}$ and \mathbb{Z}_n have the same distribution. Hence,

$$\begin{aligned}
& \mathbb{P}(\tilde{\mathbb{Z}}_n \in A, \tau_{\pi(1)} \circ \tau_{\pi(B+1)}^{-1} = \sigma_1, \dots, \tau_{\pi(B)} \circ \tau_{\pi(B+1)}^{-1} = \sigma_B) \\
&= \mathbb{E} \left[\mathbb{P}(\mathbb{Z}_n \in A) \left(\frac{1}{n!} \right)^{B-1} \mathbb{P}(\text{id} = \sigma_{\pi^{-1}(B+1)} \circ \tau_{\pi(B+1)} \mid \tau_{\pi(B+1)}) \right], \\
&= \mathbb{P}(\mathbb{Z}_n \in A) \left(\frac{1}{n!} \right)^{B-1} \mathbb{P}(\tau_{\pi(B+1)} = \sigma_{\pi^{-1}(B+1)}^{-1}), \\
&= \mathbb{P}(\mathbb{Z}_n \in A) \left(\frac{1}{n!} \right)^B,
\end{aligned}$$

This ends the proof of the exchangeability of the $(\widehat{H}_{\lambda,\mu}^{\star b})_{1 \leq b \leq B+1}$.

Then, by applying Lemma 9 to the $(\widehat{H}_{\lambda,\mu}^{\star b})_{1 \leq b \leq B+1}$, we obtain

$$\begin{aligned}
P_{f_1 \otimes f_2}(\widehat{\Delta}_\alpha^{\lambda,\mu} = 1) &= P_{f_1 \otimes f_2}(\widehat{\text{HSIC}}_{\lambda,\mu} > \widehat{q}_{1-\alpha}^{\lambda,\mu}) \\
&= P_{f_1 \otimes f_2}(\widehat{H}_{\lambda,\mu}^{\star B+1} > \widehat{H}_{\lambda,\mu}^{\star \lceil (B+1)(1-\alpha) \rceil}) \\
&= P_{f_1 \otimes f_2} \left(\sum_{b=1}^{B+1} \mathbb{1}_{\widehat{H}_{\lambda,\mu}^{\star b} < \widehat{H}_{\lambda,\mu}^{\star B+1}} \geq \lceil (B+1)(1-\alpha) \rceil \right) \\
&= P_{f_1 \otimes f_2} \left(\sum_{b=1}^{B+1} \mathbb{1}_{\widehat{H}_{\lambda,\mu}^{\star b} \geq \widehat{H}_{\lambda,\mu}^{\star B+1}} \leq \lfloor \alpha(B+1) \rfloor \right), \tag{C.3}
\end{aligned}$$

where (C.3) comes from the fact that $B+1 - \lceil (B+1)(1-\alpha) \rceil = \lfloor \alpha(B+1) \rfloor$. Then,

$$\begin{aligned}
P_{f_1 \otimes f_2}(\widehat{\Delta}_\alpha^{\lambda,\mu} = 1) &= P_{f_1 \otimes f_2} \left(\sum_{b=1}^{B+1} \mathbb{1}_{\widehat{H}_{\lambda,\mu}^{\star b} \geq \widehat{H}_{\lambda,\mu}^{\star B+1}} \leq \alpha(B+1) \right) \\
&= P_{f_1 \otimes f_2} \left(\frac{1}{B+1} \left(1 + \sum_{b=1}^B \mathbb{1}_{\widehat{H}_{\lambda,\mu}^{\star b} \geq \widehat{H}_{\lambda,\mu}^{\star B+1}} \right) \leq \alpha \right) \\
&\leq \alpha, \tag{C.4}
\end{aligned}$$

where (C.4) is obtained from Lemma 9.

C.2 Proof of Lemma 1

Let α and β be in $(0, 1)$. We aim here to give a condition on $\text{HSIC}_{\lambda,\mu}(f)$ w.r.t. the variance $\text{Var}_f(\widehat{\text{HSIC}}_{\lambda,\mu})$ and the quantile $q_{1-\alpha}^{\lambda,\mu}$, so that the statistical test $\Delta_\alpha^{\lambda,\mu}$ defined in Equation (2.1) has a second kind error controlled by β . For this, we use Chebyshev's inequality. Since $\widehat{\text{HSIC}}_{\lambda,\mu}$ is an unbiased estimator of $\text{HSIC}_{\lambda,\mu}(f)$,

$$P_f \left(\left| \widehat{\text{HSIC}}_{\lambda,\mu} - \text{HSIC}_{\lambda,\mu}(f) \right| \geq \sqrt{\frac{\text{Var}_f(\widehat{\text{HSIC}}_{\lambda,\mu})}{\beta}} \right) \leq \beta.$$

We then have the following inequality:

$$P_f \left(\widehat{\text{HSIC}}_{\lambda, \mu} \leq \text{HSIC}_{\lambda, \mu}(f) - \sqrt{\frac{\text{Var}_f(\widehat{\text{HSIC}}_{\lambda, \mu})}{\beta}} \right) \leq \beta.$$

Consequently, one has $P_f \left(\widehat{\text{HSIC}}_{\lambda, \mu} \leq q_{1-\alpha}^{\lambda, \mu} \right) \leq \beta$, as soon as

$$\text{HSIC}_{\lambda, \mu}(f) \geq \sqrt{\frac{\text{Var}_f(\widehat{\text{HSIC}}_{\lambda, \mu})}{\beta}} + q_{1-\alpha}^{\lambda, \mu}.$$

C.3 Proof of Proposition 2

In order to control the variance $\text{Var}_f(\widehat{\text{HSIC}}_{\lambda, \mu})$ w.r.t. the bandwidths λ, μ and the sample size n , let us first give the following lemma for a general U -statistic of any order r in $\{1, \dots, n\}$.

Lemma 10. *Let h be a symmetric function with $r \leq n$ inputs, V_1, \dots, V_n be independent and identically distributed random variables and U_n be the U -statistic defined by*

$$U_n = \frac{(n-r)!}{n!} \sum_{(i_1, \dots, i_r) \in \mathbf{i}_r^n} h(V_{i_1}, \dots, V_{i_r}),$$

where \mathbf{i}_r^n is the set of all r -tuples drawn without replacement from $\{1, \dots, n\}$. The following inequality gives an upper bound of the variance of U_n ,

$$\text{Var}(U_n) \leq C(r) \left(\frac{\sigma^2}{n} + \frac{s^2}{n^2} \right), \quad (\text{C.5})$$

where $\sigma^2 = \text{Var}(\mathbb{E}[h(V_1, \dots, V_r) \mid V_1])$ and $s^2 = \text{Var}(h(V_1, \dots, V_r))$.

Proof of Lemma 10. First, using Hoeffding's decomposition (see e.g. [Serfling, 2009, Lemma A, p. 183]), the variance of U_n can be decomposed as

$$\text{Var}(U_n) = \binom{n}{r}^{-1} \sum_{c=1}^r \binom{r}{c} \binom{n-r}{r-c} \zeta_c,$$

where $\zeta_c = \text{Var}(\mathbb{E}[h(V_1, \dots, V_r) \mid V_1, \dots, V_c])$.

Let us now prove that, for all $n \in \mathbb{N}^*$, $r \in \{1, \dots, n\}$ and $c \in \{1, \dots, r\}$,

$$\binom{n}{r}^{-1} \binom{r}{c} \binom{n-r}{r-c} \leq \frac{C(r, c)}{n^c}. \quad (\text{C.6})$$

We first write

$$\binom{n}{r}^{-1} \binom{r}{c} \binom{n-r}{r-c} = \binom{r}{c} \times \frac{r!}{(r-c)!} \times \frac{(n-r)!}{(n+c-2r)!} \times \frac{(n-r)!}{n!}. \quad (\text{C.7})$$

Moreover,

$$\begin{aligned} n! &= (n-r)! \times (n-r+1) \times \dots \times (n-r+r) \\ &\geq (n-r)! \times (n-r+1)^r, \end{aligned}$$

and

$$\begin{aligned} (n-r)! &= (n-2r+c)! \times (n-2r+c+1) \times \dots \times (n-2r+c+r-c) \\ &\leq (n-2r+c)! \times (n-r+1)^{r-c}. \end{aligned}$$

Then, we have

$$\frac{(n-r)!}{(n+c-2r)!} \times \frac{(n-r)!}{n!} \leq \frac{1}{(n-r+1)^c}.$$

Furthermore, using that $n \geq r$, one can write

$$\begin{aligned} \frac{n-r+1}{n} &= 1 - \frac{r-1}{n} \\ &\geq 1 - \frac{r-1}{r} = \frac{1}{r}. \end{aligned}$$

This leads to, $\frac{1}{n-r+1} \leq \frac{r}{n}$. Finally, Equation (C.7) leads to Equation (C.6).

By upper bounding each term in Hoeffding's decomposition of the variance of U_n according to Equation (C.6), we obtain

$$\text{Var}(U_n) \leq C(r) \sum_{c=1}^r \frac{\zeta_c}{n^c}. \quad (\text{C.8})$$

On the one hand, $\zeta_1 = \sigma^2$. On the other hand, using the law of total variance (see e.g. [Weiss, 2006]), $\zeta_c \leq s^2$ for all c in $\{2, \dots, r\}$. By injecting this last inequality in Equation (C.8), we obtain for all n in \mathbb{N}^* ,

$$\text{Var}(U_n) \leq C(r) \left(\frac{\sigma^2}{n} + \frac{s^2}{n^2} \right),$$

which achieves the proof of Lemma 10. \square

Let us now apply Lemma 10 in order to control the variance of $\widehat{\text{HSIC}}_{\lambda, \mu}$ w.r.t λ , μ and n . For this, we first recall that $\widehat{\text{HSIC}}_{\lambda, \mu}$ can be written as a single U -statistic of order 4 as

$$\widehat{\text{HSIC}}_{\lambda, \mu} = \frac{1}{n(n-1)(n-2)(n-3)} \sum_{(i,j,q,r) \in \mathbf{i}_4^n} h_{i,j,q,r},$$

where the general term $h_{i,j,q,r}$ of $\widehat{\text{HSIC}}_{\lambda, \mu}$ is defined as in [Gretton et al., 2008] by

$$h_{i,j,q,r} = \frac{1}{4!} \sum_{(t,u,v,w)}^{(i,j,q,r)} \left[k_\lambda(X_t, X_u) l_\mu(Y_t, Y_u) + k_\lambda(X_t, X_u) l_\mu(Y_v, Y_w) - 2k_\lambda(X_t, X_u) l_\mu(Y_t, Y_v) \right].$$

where the sum represents all ordered quadruples (t, u, v, w) drawn without replacement from (i, j, q, r) .

Thus, using Lemma 10, the variance of $\widehat{\text{HSIC}}_{\lambda, \mu}$ can be upper bounded as follows:

$$\text{Var}_f(\widehat{\text{HSIC}}_{\lambda, \mu}) \leq C \left(\frac{\sigma^2(\lambda, \mu)}{n} + \frac{s^2(\lambda, \mu)}{n^2} \right), \quad (\text{C.9})$$

where, recalling that $Z_i = (X_i, Y_i)$ for all i in $\{1, \dots, n\}$, $\sigma^2(\lambda, \mu) = \text{Var}_f(\mathbb{E}[h_{1,2,3,4} | Z_1])$ and $s^2(\lambda, \mu) = \text{Var}_f(h_{1,2,3,4})$.

C.3.1 Upper bound of $\sigma^2(\lambda, \mu)$

By now, we upper bound $\sigma^2(\lambda, \mu) = \text{Var}_f(\mathbb{E}[h_{1,2,3,4} | Z_1])$ w.r.t. λ and μ . For this, we first notice that in the cases where $k_\lambda(X_a, X_b)l_\mu(Y_c, Y_d)$ is independent from Z_1 , the variance of its expectation conditionally on Z_1 equals 0. This corresponds to the cases where a, b, c and d are all different from 1. We then have the following inequality:

$$\sigma^2(\lambda, \mu) \leq C \sum_{i=1}^6 \sigma_i^2(\lambda, \mu),$$

where

$$\begin{aligned} \sigma_1^2(\lambda, \mu) &= \text{Var}_f(\mathbb{E}[k_\lambda(X_1, X_2)l_\mu(Y_1, Y_2) | Z_1]), & \sigma_2^2(\lambda, \mu) &= \text{Var}_f(\mathbb{E}[k_\lambda(X_1, X_2)l_\mu(Y_3, Y_4) | X_1]), \\ \sigma_3^2(\lambda, \mu) &= \text{Var}_f(\mathbb{E}[k_\lambda(X_3, X_4)l_\mu(Y_1, Y_2) | Y_1]), & \sigma_4^2(\lambda, \mu) &= \text{Var}_f(\mathbb{E}[k_\lambda(X_1, X_2)l_\mu(Y_1, Y_3) | Z_1]), \\ \sigma_5^2(\lambda, \mu) &= \text{Var}_f(\mathbb{E}[k_\lambda(X_2, X_1)l_\mu(Y_2, Y_3) | X_1]), & \sigma_6^2(\lambda, \mu) &= \text{Var}_f(\mathbb{E}[k_\lambda(X_2, X_3)l_\mu(Y_2, Y_1) | Y_1]). \end{aligned}$$

Case 1. Upper bound of $\sigma_1^2(\lambda, \mu)$:

$$\begin{aligned} \sigma_1^2(\lambda, \mu) &\leq \mathbb{E} \left[\left(\mathbb{E}[k_\lambda(X_1, X_2)l_\mu(Y_1, Y_2) | Z_1] \right)^2 \right] \\ &\leq \mathbb{E}[k_\lambda(X_1, X_2)l_\mu(Y_1, Y_2)k_\lambda(X_1, X_3)l_\mu(Y_1, Y_3)]. \end{aligned}$$

Moreover, we have

$$\begin{aligned} &\mathbb{E}[k_\lambda(X_1, X_2)k_\lambda(X_1, X_3)l_\mu(Y_1, Y_2)l_\mu(Y_1, Y_3)] \\ &= \int_{(\mathbb{R}^p \times \mathbb{R}^q)^3} k_\lambda(x_1, x_2)k_\lambda(x_1, x_3)l_\mu(y_1, y_2)l_\mu(y_1, y_3) \prod_{k=1}^3 f(x_k, y_k) dx_k dy_k. \end{aligned}$$

Since k_λ and l_μ are nonnegative, one can upper bound $f(x_2, y_2)$ and $f(x_3, y_3)$ by $\|f\|_\infty$, and obtain

$$\begin{aligned} \sigma_1^2(\lambda, \mu) &\leq \|f\|_\infty^2 \int_{(\mathbb{R}^p \times \mathbb{R}^q)^3} k_\lambda(x_1, x_2)k_\lambda(x_1, x_3)l_\mu(y_1, y_2)l_\mu(y_1, y_3) f(x_1, y_1) \prod_{k=1}^3 dx_k dy_k \\ &= \|f\|_\infty^2 \int_{\mathbb{R}^p \times \mathbb{R}^q} \left[\int_{\mathbb{R}^p} k_\lambda(x_1, x) dx \right]^2 \left[\int_{\mathbb{R}^q} l_\mu(y_1, y) dy \right]^2 f(x_1, y_1) dx_1 dy_1. \end{aligned}$$

Finally, using that $\int_{\mathbb{R}^p} k_\lambda(\cdot, x) dx = \int_{\mathbb{R}^q} l_\mu(\cdot, y) dy = 1$, we write

$$\sigma_1^2(\lambda, \mu) \leq \|f\|_\infty^2. \quad (\text{C.10})$$

Case 2. Upper bound of $\sigma_2^2(\lambda, \mu)$:

$$\begin{aligned} \sigma_2^2(\lambda, \mu) &\leq \mathbb{E} \left[\left(\mathbb{E} [k_\lambda(X_1, X_2) l_\mu(Y_3, Y_4) \mid X_1] \right)^2 \right] \\ &\leq \mathbb{E} \left[\left(\mathbb{E} [k_\lambda(X_1, X_2) \mid X_1] \right)^2 \right] \left(\mathbb{E} [l_\mu(Y_3, Y_4)] \right)^2 \\ &\leq \mathbb{E} [k_\lambda(X_1, X_2) k_\lambda(X_1, X_3)] \left(\mathbb{E} [l_\mu(Y_3, Y_4)] \right)^2. \end{aligned}$$

Moreover, it is easy to see that by upper bounding $f_1(x_2)$ and $f_1(x_3)$ by $\|f_1\|_\infty$, and recalling that $\int_{\mathbb{R}^p} k_\lambda(x_1, x) dx = 1$, we have,

$$\begin{aligned} \mathbb{E} [k_\lambda(X_1, X_2) k_\lambda(X_1, X_3)] &= \int_{\mathbb{R}^p} \left[\int_{\mathbb{R}^p} k_\lambda(x_1, x_2) f_1(x_2) dx_2 \right] \left[\int_{\mathbb{R}^p} k_\lambda(x_1, x_3) f_1(x_3) dx_3 \right] f_1(x_1) dx_1 \\ &\leq \|f_1\|_\infty^2. \end{aligned}$$

Besides, upper bounding $f_2(y_3)$ by $\|f_2\|_\infty$ in the integral form of $\mathbb{E} [l_\mu(Y_3, Y_4)]$ gives

$$\mathbb{E} [l_\mu(Y_3, Y_4)] \leq \|f_2\|_\infty.$$

By combining these inequalities, we obtain

$$\sigma_2^2(\lambda, \mu) \leq \|f_1\|_\infty^2 \|f_2\|_\infty^2. \quad (\text{C.11})$$

Case 3. Upper bound of $\sigma_3^2(\lambda, \mu)$: this case is similar to case 2 by exchanging X by Y and k_λ by l_μ . Thus, we have the inequality

$$\sigma_3^2(\lambda, \mu) \leq \|f_1\|_\infty^2 \|f_2\|_\infty^2. \quad (\text{C.12})$$

Case 4. Upper bound of $\sigma_4^2(\lambda, \mu)$:

$$\begin{aligned} \sigma_4^2(\lambda, \mu) &\leq \mathbb{E} \left[\left(\mathbb{E} [k_\lambda(X_1, X_2) l_\mu(Y_1, Y_3) \mid Z_1] \right)^2 \right] \\ &\leq \mathbb{E} [k_\lambda(X_1, X_2) k_\lambda(X_1, X_4) l_\mu(Y_1, Y_3) l_\mu(Y_1, Y_5)]. \end{aligned}$$

By upper bounding $f_1(x_2)$, $f_1(x_4)$ by $\|f_1\|_\infty$ and $f_2(y_3)$, $f_2(y_5)$ by $\|f_2\|_\infty$ in the integral form of $\mathbb{E} [k_\lambda(X_1, X_2) k_\lambda(X_1, X_4) l_\mu(Y_1, Y_3) l_\mu(Y_1, Y_5)]$, we obtain

$$\sigma_4^2(\lambda, \mu) \leq \|f_1\|_\infty^2 \|f_2\|_\infty^2. \quad (\text{C.13})$$

Case 5. Upper bound of $\sigma_5^2(\lambda, \mu)$:

$$\begin{aligned} \sigma_5^2(\lambda, \mu) &\leq \mathbb{E} \left[\left(\mathbb{E} [k_\lambda(X_2, X_1) l_\mu(Y_2, Y_3) \mid X_1] \right)^2 \right] \\ &\leq \mathbb{E} [k_\lambda(X_2, X_1) k_\lambda(X_4, X_1) l_\mu(Y_2, Y_3) l_\mu(Y_4, Y_5)]. \end{aligned}$$

By upper bounding $f(x_2, y_2)$ and $f(x_4, y_4)$ by $\|f\|_\infty$ in the integral form of the last expectation, we have

$$\sigma_5^2(\lambda, \mu) \leq \|f\|_\infty^2. \quad (\text{C.14})$$

Case 6. Upper bound of $\sigma_6^2(\lambda, \mu)$: this case is similar to case 5 by exchanging X by Y and k_λ by l_μ . We have then the inequality

$$\sigma_6^2(\lambda, \mu) \leq \|f\|_\infty^2. \quad (\text{C.15})$$

Finally, by combining inequalities (C.10), (C.11), (C.12), (C.13), (C.14) and (C.15), we have the following inequality

$$\sigma^2(\lambda, \mu) \leq C(M_f). \quad (\text{C.16})$$

C.3.2 Upper bound of $s^2(\lambda, \mu)$

Let us first recall that the general term of the U -statistic $\widehat{\text{HSIC}}_{\lambda, \mu}$ is written as

$$h_{1,2,3,4} = \frac{1}{4!} \sum_{(u,v,w,t)}^{(1,2,3,4)} k_\lambda(X_u, X_v) [l_\mu(Y_u, Y_v) + l_\mu(Y_w, Y_t) - 2l_\mu(Y_u, Y_w)].$$

Moreover, all the terms of the last sum have the same distribution. We then have

$$\begin{aligned} s^2(\lambda, \mu) &= \text{Var}_f(h_{1,2,3,4}) \\ &\leq C \text{Var}_f(k_\lambda(X_1, X_2) [l_\mu(Y_1, Y_2) + l_\mu(Y_3, Y_4) - 2l_\mu(Y_1, Y_3)]). \end{aligned}$$

It follows that,

$$\begin{aligned} \text{Var}_f(h_{1,2,3,4}) &\leq C [\text{Var}_f(k_\lambda(X_1, X_2)l_\mu(Y_1, Y_2)) + \text{Var}_f(k_\lambda(X_1, X_2)l_\mu(Y_3, Y_4)) + \text{Var}_f(k_\lambda(X_1, X_2)l_\mu(Y_1, Y_3))] \\ &\leq C \left(\mathbb{E}[k_\lambda^2(X_1, X_2)l_\mu^2(Y_1, Y_2)] + \mathbb{E}[k_\lambda^2(X_1, X_2)l_\mu^2(Y_3, Y_4)] + \mathbb{E}[k_\lambda^2(X_1, X_2)l_\mu^2(Y_1, Y_3)] \right). \end{aligned}$$

In order to bring back to multivariate normal densities, we express k_λ^2 and l_μ^2 as

$$k_\lambda^2 = \frac{k_{\lambda'}}{(4\pi)^{\frac{p}{2}} \lambda_1 \dots \lambda_p} \quad \text{and} \quad l_\mu^2 = \frac{l_{\mu'}}{(4\pi)^{\frac{q}{2}} \mu_1 \dots \mu_q},$$

where $\lambda' = \frac{\lambda}{\sqrt{2}}$ and $\mu' = \frac{\mu}{\sqrt{2}}$.

Consequently, the expectation $\mathbb{E}[k_\lambda^2(X_1, X_2)l_\mu^2(Y_1, Y_2)]$ can be expressed as

$$\begin{aligned} &\mathbb{E}[k_\lambda^2(X_1, X_2)l_\mu^2(Y_1, Y_2)] \\ &= \frac{1}{(4\pi)^{\frac{p+q}{2}} \lambda_1 \dots \lambda_p \mu_1 \dots \mu_q} \mathbb{E}[k_{\lambda'}(X_1, X_2)l_{\mu'}(Y_1, Y_2)] \\ &= \frac{1}{(4\pi)^{\frac{p+q}{2}} \lambda_1 \dots \lambda_p \mu_1 \dots \mu_q} \int_{(\mathbb{R}^p \times \mathbb{R}^q)^2} k_{\lambda'}(x_1, x_2) l_{\mu'}(y_1, y_2) f(x_1, y_1) f(x_2, y_2) dx_1 dx_2 dy_1 dy_2. \end{aligned}$$

By upper bounding $f(x_2, y_2)$ by $\|f\|_\infty$ in the last integral, we have

$$\begin{aligned} & \int_{(\mathbb{R}^p \times \mathbb{R}^q)^2} k_{\lambda'}(x_1, x_2) l_{\mu'}(y_1, y_2) f(x_1, y_1) f(x_2, y_2) dx_1 dx_2 dy_1 dy_2 \\ & \leq \|f\|_\infty \int_{\mathbb{R}^p \times \mathbb{R}^q} \left[\int_{\mathbb{R}^p} k_{\lambda'}(x_1, x_2) dx_2 \right] \left[\int_{\mathbb{R}^q} l_{\mu'}(y_1, y_2) dy_2 \right] f(x_1, y_1) dx_1 dy_1 \\ & = \|f\|_\infty. \end{aligned}$$

This leads to,

$$\mathbb{E} [k_\lambda^2(X_1, X_2) l_\mu^2(Y_1, Y_2)] \leq \frac{\|f\|_\infty}{(4\pi)^{\frac{p+q}{2}} \lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}. \quad (\text{C.17})$$

We can easily show by similar argument that

$$\mathbb{E} [k_\lambda^2(X_1, X_2) l_\mu^2(Y_3, Y_4)] \leq \frac{\|f_1\|_\infty \|f_2\|_\infty}{(4\pi)^{\frac{p+q}{2}} \lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}. \quad (\text{C.18})$$

and

$$\mathbb{E} [k_\lambda^2(X_1, X_2) l_\mu^2(Y_1, Y_3)] \leq \frac{\|f\|_\infty}{(4\pi)^{\frac{p+q}{2}} \lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}. \quad (\text{C.19})$$

From Equations (C.17), (C.18) and (C.19), we have

$$s^2(\lambda, \mu) \leq \frac{C(M_f)}{(4\pi)^{\frac{p+q}{2}} \lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}. \quad (\text{C.20})$$

From Equations (C.16) and (C.20), we deduce the following inequality for $\text{Var}_f(\widehat{\text{HSIC}}_{\lambda, \mu})$

$$\text{Var}_f(\widehat{\text{HSIC}}_{\lambda, \mu}) \leq C(M_f, p, q) \left\{ \frac{1}{n} + \frac{1}{n^2 \lambda_1 \dots \lambda_p \mu_1 \dots \mu_q} \right\}.$$

C.4 Proof of Proposition 3

To give an upper bound for the quantile $q_{1-\alpha}^{\lambda, \mu}$ w.r.t λ and μ , we use concentration inequalities for general U -statistics. Recall that $\widehat{\text{HSIC}}_{\lambda, \mu}$ can be written as a U -statistic of order 4,

$$\widehat{\text{HSIC}}_{\lambda, \mu} = \frac{1}{n(n-1)(n-2)(n-3)} \sum_{(i,j,q,r) \in \mathbf{i}_4^n} h_{i,j,q,r},$$

with general term $h_{i,j,q,r}$ defined by

$$h_{i,j,q,r} = \frac{1}{4!} \sum_{(t,u,v,w)}^{(i,j,q,r)} \left[k_\lambda(X_t, X_u) l_\mu(Y_t, Y_u) + k_\lambda(X_t, X_u) l_\mu(Y_v, Y_w) - 2k_\lambda(X_t, X_u) l_\mu(Y_t, Y_v) \right].$$

where the sum represents all ordered quadruples (t, u, v, w) drawn without replacement from (i, j, q, r) .

However, sharp upper bounds are obtained only for degenerate U -statistics (see e.g. [Houdré and Reynaud-Bour]

We recall that a U -statistic of order r , denoted $U_n = U_n(V_1, \dots, V_r)$, is degenerate if $\mathbb{E}[U_n \mid$

$V_1, \dots, V_{r-1}] = 0$. Note that this implies that $\mathbb{E}[U_n \mid V_1, \dots, V_i] = 0$ for all i in $\{1, \dots, r-1\}$. Hence, the first step to upper bound $q_{1-\alpha}^{\lambda, \mu}$ is to write $\widehat{\text{HSIC}}_{\lambda, \mu}$ as a sum of degenerate U -statistics. For this, we rely on the ANOVA-decomposition (ANOVA for ANalyse Of VAriance, see e.g. [Sobol, 2001]) of the symmetric function $h_{i,j,q,r}$. We then write

$$h_{i,j,q,r} = \frac{1}{2!} \sum_{(t,u)}^{(i,j,q,r)} h_{t,u} + \frac{1}{3!} \sum_{(t,u,v)}^{(i,j,q,r)} h_{t,u,v} + \tilde{h}_{i,j,q,r}, \quad (\text{C.21})$$

where the first (resp. the second) sum represents all ordered pairs (t, u) (resp. triplets (t, u, v)) drawn without replacement from (i, j, q, r) and the terms $h_{t,u}$, $h_{t,u,v}$ and $\tilde{h}_{i,j,q,r}$ are defined as

$$\begin{aligned} h_{t,u} &= \mathbb{E}[h_{i,j,q,r} \mid Z_t, Z_u], \\ h_{t,u,v} &= \mathbb{E}[h_{i,j,q,r} \mid Z_t, Z_u, Z_v] - \frac{1}{2!} \sum_{(t',u')}^{(t,u,v)} h_{t',u'}, \\ \tilde{h}_{i,j,q,r} &= h_{i,j,q,r} - \frac{1}{3!} \sum_{(t,u,v)}^{(i,j,q,r)} h_{t,u,v} - \frac{1}{2!} \sum_{(t,u)}^{(i,j,q,r)} h_{t,u}. \end{aligned}$$

Hence, by summing all terms $h_{i,j,q,r}$ for (i, j, q, r) in \mathbf{i}_4^n and then dividing by $n(n-1)(n-2)(n-3)$, we have

$$\widehat{\text{HSIC}}_{\lambda, \mu} = 6 \widehat{\text{HSIC}}_{\lambda, \mu}^{(2,D)} + 4 \widehat{\text{HSIC}}_{\lambda, \mu}^{(3,D)} + \widehat{\text{HSIC}}_{\lambda, \mu}^{(4,D)}, \quad (\text{C.22})$$

where

$$\begin{aligned} \widehat{\text{HSIC}}_{\lambda, \mu}^{(2,D)} &= \frac{1}{n(n-1)} \sum_{(i,j) \in \mathbf{i}_2^n} h_{i,j}, & \widehat{\text{HSIC}}_{\lambda, \mu}^{(3,D)} &= \frac{1}{n(n-1)(n-2)} \sum_{(i,j,q) \in \mathbf{i}_3^n} h_{i,j,q} \\ \widehat{\text{HSIC}}_{\lambda, \mu}^{(4,D)} &= \frac{1}{n(n-1)(n-2)(n-3)} \sum_{(i,j,q,r) \in \mathbf{i}_4^n} \tilde{h}_{i,j,q,r}. \end{aligned}$$

Lemma 11. *Let us assume that $f = f_1 \otimes f_2$. Then, the U -statistics $\widehat{\text{HSIC}}_{\lambda, \mu}^{(2,D)}$, $\widehat{\text{HSIC}}_{\lambda, \mu}^{(3,D)}$ and $\widehat{\text{HSIC}}_{\lambda, \mu}^{(4,D)}$ are degenerated.*

Proof of Lemma 11. According to Theorem 2 of [Gretton et al., 2008], if $f = f_1 \otimes f_2$, we have

$$\mathbb{E}[h_{i,j,q,r} \mid Z_i] = 0.$$

We then easily show that $\widehat{\text{HSIC}}_{\lambda, \mu}^{(2,D)}$ is degenerated by writing

$$\mathbb{E}[h_{i,j} \mid Z_i] = \mathbb{E}[h_{i,j,q,r} \mid Z_i] = 0. \quad (\text{C.23})$$

Moreover, to prove that $\widehat{\text{HSIC}}_{\lambda, \mu}^{(3,D)}$ is degenerated, we have

$$\begin{aligned} \mathbb{E}[h_{i,j,q} \mid Z_i, Z_j] &= \mathbb{E}[h_{i,j,q,r} \mid Z_i, Z_j] - \mathbb{E}[h_{i,j} \mid Z_i, Z_j] - \mathbb{E}[h_{i,q} \mid Z_i] - \mathbb{E}[h_{j,q} \mid Z_j] \\ &= h_{i,j} - h_{i,j} \\ &= 0, \end{aligned} \quad (\text{C.24})$$

where (C.24) holds by definition of $h_{i,j}$ and Equation (C.23). Finally, from previous cases, and by definition of $h_{i,j,q}$, we obtain

$$\begin{aligned}\mathbb{E}[\tilde{h}_{i,j,q,r} \mid Z_i, Z_j, Z_q] &= \mathbb{E}[h_{i,j,q,r} \mid Z_i, Z_j, Z_q] - h_{i,j,q} - h_{i,j} - h_{i,q} - h_{j,q} \\ &= 0,\end{aligned}$$

which proves that $\widehat{\text{HSIC}}_{\lambda,\mu}^{(4,D)}$ is degenerated. \square

Once we have upper bounds of the $(1 - \alpha)$ -quantiles of $\widehat{\text{HSIC}}_{\lambda,\mu}^{(r,D)}$ with r in $\{2, 3, 4\}$ under the assumption $P_{f_1 \otimes f_2}$, an upper bound of the quantile $q_{1-\alpha}^{\lambda,\mu}$ is naturally obtained. In fact, we can easily show that,

$$q_{1-\alpha}^{\lambda,\mu} \leq 6 q_{1-\alpha/3,2}^{\lambda,\mu} + 4 q_{1-\alpha/3,3}^{\lambda,\mu} + q_{1-\alpha/3,4}^{\lambda,\mu} \quad (\text{C.25})$$

where $q_{1-\alpha,r}^{\lambda,\mu}$ is the $(1 - \alpha)$ -quantiles of $\widehat{\text{HSIC}}_{\lambda,\mu}^{(r,D)}$ under $P_{f_1 \otimes f_2}$.

C.4.1 Upper bound of $q_{1-\alpha,2}^{\lambda,\mu}$

First, [Gretton et al., 2008] page 10 prove that, under the hypothesis $f = f_1 \otimes f_2$, $h_{i,j} = \mathbb{E}[h_{i,j,q,r} \mid Z_i, Z_j]$ can be written as follows

$$h_{i,j} = h_{(2)}(Z_i, Z_j),$$

where, for all $z_1 = (x, y)$ and $z' = (x', y')$ in $\mathbb{R}^p \times \mathbb{R}^q$,

$$\begin{aligned}h_{(2)}(z, z') &= \frac{1}{6} \left\{ k_\lambda(x, x') - \mathbb{E}[k_\lambda(x, X')] - \mathbb{E}[k_\lambda(X, x')] + \mathbb{E}[k_\lambda(X, X')] \right\} \\ &\quad \times \left\{ l_\mu(y, y') - \mathbb{E}[l_\mu(y, Y')] - \mathbb{E}[l_\mu(Y, y')] + \mathbb{E}[l_\mu(Y, Y')] \right\},\end{aligned}$$

for (X, Y) and (X', Y') independent random variables with common density $f_1 \otimes f_2$.

To upper bound the quantile $q_{1-\alpha,2}^{\lambda,\mu}$, we use the concentration inequality for degenerated U -statistics of order 2 given in [Giné et al., 2000, p.15, Equation (3.5)]. We write for all $t > 0$,

$$\mathbb{P} \left(\left| \sum_{(i,j) \in i_2^n} h_{(2)}(Z_i, Z_j) \right| > t \right) \leq A \exp \left(-\frac{1}{A} \min \left\{ \frac{t}{M}, \left(\frac{t}{L} \right)^{2/3}, \left(\frac{t}{K} \right)^{1/2} \right\} \right), \quad (\text{C.26})$$

where $A > 1$ is an absolute constant,

$$K = \|h_{(2)}\|_\infty, \quad \text{and} \quad M^2 = \sum_{(i,j) \in i_2^n} \mathbb{E}[h_{(2)}^2(Z_i, Z_j)] = n(n-1) \mathbb{E}[h_{(2)}^2(Z_1, Z_2)],$$

$$L^2 = \max \left\{ \left\| \sum_{i=1}^n \mathbb{E}[h_{(2)}^2(Z_i, \cdot)] \right\|_\infty, \left\| \sum_{j=1}^n \mathbb{E}[h_{(2)}^2(\cdot, Z_j)] \right\|_\infty \right\} = n \left\| \mathbb{E}[h_{(2)}^2(Z_1, \cdot)] \right\|_\infty.$$

By setting $\varepsilon = \frac{t}{n^2}$, and using Equation (C.26), we obtain

$$\mathbb{P} \left(\frac{1}{n^2} \left| \sum_{(i,j) \in i_2^n} h_{i,j} \right| > \varepsilon \right) \leq A \exp \left(-\frac{1}{A} \min \left\{ \frac{n^2 \varepsilon}{M}, \left(\frac{n^2 \varepsilon}{L} \right)^{2/3}, \left(\frac{n^2 \varepsilon}{K} \right)^{1/2} \right\} \right).$$

Therefore, we have for all $\varepsilon > 0$,

$$\mathbb{P} \left(\frac{1}{n^2} \left| \sum_{(i,j) \in i_2^n} h_{i,j} \right| > \varepsilon \right) \leq A \max \left\{ \exp \left(-\frac{n^2 \varepsilon}{AM} \right), \exp \left(-\frac{n^{4/3} \varepsilon^{2/3}}{AL^{2/3}} \right), \exp \left(-\frac{n \varepsilon^{1/2}}{AK^{1/2}} \right) \right\}.$$

By adjusting the constant A , we can replace in the last inequality $\frac{1}{n^2} \sum_{(i,j) \in i_2^n} h_{i,j}$ by $\widehat{\text{HSIC}}_{\lambda, \mu}^{(2,D)}$,

$$\mathbb{P} \left(\left| \widehat{\text{HSIC}}_{\lambda, \mu}^{(2,D)} \right| > \varepsilon \right) \leq A \max \left\{ \exp \left(-\frac{n^2 \varepsilon}{AM} \right), \exp \left(-\frac{n^{4/3} \varepsilon^{2/3}}{AL^{2/3}} \right), \exp \left(-\frac{n \varepsilon^{1/2}}{AK^{1/2}} \right) \right\}.$$

Hence, if ε_α is a positive number verifying

$$\alpha = A \max \left\{ \exp \left(-\frac{n^2 \varepsilon_\alpha}{AM} \right), \exp \left(-\frac{n^{4/3} \varepsilon_\alpha^{2/3}}{AL^{2/3}} \right), \exp \left(-\frac{n \varepsilon_\alpha^{1/2}}{AK^{1/2}} \right) \right\},$$

then, by definition of the quantile,

$$q_{1-\alpha, 2}^{\lambda, \mu} \leq \varepsilon_\alpha. \quad (\text{C.27})$$

By now, we upper bound ε_α (and consequently $q_{1-\alpha, 2}^{\lambda, \mu}$), in the 3 cases considered bellow.

Case 1. If $\alpha = A \exp(-n^2 \varepsilon_\alpha / [AM])$, then ε_α is expressed as

$$\varepsilon_\alpha = \frac{AM}{n^2} \left(\log \left(\frac{1}{\alpha} \right) + \log(A) \right).$$

Since in $\mathcal{A}_2(\alpha)$, we assume that $\log(1/\alpha) > 1$, and since $A > 1$, we can then bound ε_α as

$$0 < \varepsilon_\alpha \leq \frac{CM}{n^2} \log \left(\frac{1}{\alpha} \right). \quad (\text{C.28})$$

for some absolute positive constant C .

Let us upper bound M w.r.t λ , μ and n . First notice that

$$M^2 = n(n-1) \mathbb{E} \left[h_{(2)}^2(Z_1, Z_2) \right] \leq n^2 \mathbb{E} \left[h_{1,2}^2 \right].$$

Moreover, by the law of total variance,

$$\begin{aligned} \mathbb{E} \left[h_{1,2}^2 \right] &= \text{Var} \left(\mathbb{E} \left[h_{1,2,3,4} \mid Z_1, Z_2 \right] \right) \\ &\leq \text{Var} \left(h_{1,2,3,4} \right). \end{aligned}$$

Furthermore, we have shown in Equation (C.20) (see Section C.3.2) that,

$$\text{Var}(h_{1,2,3,4}) \leq \frac{C(M_{f_1 \otimes f_2}, p, q)}{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q} \leq \frac{C(M_f, p, q)}{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q},$$

since we work under $P_{f_1 \otimes f_2}$.

Hence, we can upper bound M as follows,

$$M \leq \frac{C(M_f, p, q)n}{\sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}}. \quad (\text{C.29})$$

Consequently, by combining Equations (C.28) and (C.29), we obtain

$$q_{1-\alpha,2}^{\lambda,\mu} \leq \frac{C(M_f, p, q)}{n\sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}} \log\left(\frac{1}{\alpha}\right). \quad (\text{C.30})$$

Case 2. If $\alpha = A \exp\left(-n^{4/3} \varepsilon_\alpha^{2/3} / [AL^{2/3}]\right)$, then ε_α verifies

$$\varepsilon_\alpha^{2/3} = \frac{AL^{2/3}}{n^{4/3}} \left(\log\left(\frac{1}{\alpha}\right) + \log(A) \right).$$

Thus, since $\log(1/\alpha) > 1$, ε_α can be upper bounded as

$$\varepsilon_\alpha \leq \frac{CL}{n^2} \log\left(\frac{1}{\alpha}\right)^{3/2}, \quad (\text{C.31})$$

Let us upper bound L w.r.t n , λ and μ , where

$$L^2 = n \sup_{z \in \mathbb{R}^p \times \mathbb{R}^q} \left\{ \mathbb{E}[h_{(2)}^2(Z_1, z)] \right\}.$$

Yet, for all $z = (x, y) \in \mathbb{R}^p \times \mathbb{R}^q$,

$$\begin{aligned} h_{(2)}^2(Z_1, z) &= \frac{1}{36} \left\{ k_\lambda(X_1, x) - \mathbb{E}[k_\lambda(X_1, X_2) \mid X_1] - \mathbb{E}[k_\lambda(X_3, x)] + \mathbb{E}[k_\lambda(X_3, X_2)] \right\}^2 \\ &\quad \times \left\{ l_\mu(Y_1, y) - \mathbb{E}[l_\mu(Y_1, Y_2) \mid Y_1] - \mathbb{E}[l_\mu(Y_3, y)] + \mathbb{E}[l_\mu(Y_3, Y_2)] \right\}^2. \end{aligned}$$

Therefore, we have the following inequality for $h_{(2)}^2(Z_1, z)$,

$$\begin{aligned} h_{(2)}^2(Z_1, z) &\leq C \left\{ k_\lambda^2(X_1, x) + \mathbb{E}[k_\lambda(X_1, X_2) \mid X_1]^2 + \mathbb{E}[k_\lambda(X_3, x)]^2 + \mathbb{E}[k_\lambda(X_3, X_2)]^2 \right\} \\ &\quad \times \left\{ l_\mu^2(Y_1, y) + \mathbb{E}[l_\mu(Y_1, Y_2) \mid Y_1]^2 + \mathbb{E}[l_\mu(Y_3, y)]^2 + \mathbb{E}[l_\mu(Y_3, Y_2)]^2 \right\}. \end{aligned}$$

Using that (X_1, \dots, X_n) and (Y_1, \dots, Y_n) are independent, and Jensen's inequality,

$$\mathbb{E}[h_{(2)}^2(Z_1, z)] \leq C \left\{ \mathbb{E}[k_\lambda^2(X_1, x)] + \mathbb{E}[k_\lambda^2(X_1, X_2)] \right\} \times \left\{ \mathbb{E}[l_\mu^2(Y_1, y)] + \mathbb{E}[l_\mu^2(Y_1, Y_2)] \right\}$$

Moreover, by similar arguments as in Section C.3.2, one can prove that for all x in \mathbb{R}^p ,

$$\mathbb{E}[k_\lambda^2(X_1, x)] \leq \frac{C(\|f_1\|_\infty, p)}{\lambda_1 \dots \lambda_p}, \quad \text{and} \quad \mathbb{E}[k_\lambda^2(X_1, X_2)] \leq \frac{C(\|f_1\|_\infty, p)}{\lambda_1 \dots \lambda_p},$$

and for all y in \mathbb{R}^q ,

$$\mathbb{E}[l_\mu^2(Y_1, y)] \leq \frac{C(\|f_2\|_\infty, q)}{\mu_1 \dots \mu_q} \quad \text{and} \quad \mathbb{E}[l_\mu^2(Y_1, Y_2)] \leq \frac{C(\|f_2\|_\infty, q)}{\mu_1 \dots \mu_q}.$$

Hence, by taking the supremum over $z = (x, y)$ in $\mathbb{R}^p \times \mathbb{R}^q$, we obtain

$$L^2 \leq C(M_f, p, q) \frac{n}{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}. \quad (\text{C.32})$$

By combining Equations (C.31) and (C.32), we have

$$\varepsilon_\alpha \leq \frac{C(M_f, p, q)}{n^{3/2} \sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}} \left[\log \left(\frac{1}{\alpha} \right) \right]^{3/2}.$$

Moreover, since from $\mathcal{A}_2(\alpha)$ we have $\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q < 1$, we obtain

$$\varepsilon_\alpha \leq \frac{C(M_f)}{(n \sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q})^{3/2}} \left[\log \left(\frac{1}{\alpha} \right) \right]^{3/2}. \quad (\text{C.33})$$

Case 3. If $\alpha = A \exp \left(-n \varepsilon_\alpha^{1/2} / \left[AK^{1/2} \right] \right)$, then ε_α is expressed as

$$\varepsilon_\alpha^{1/2} = \frac{AK^{1/2}}{n} \left(\log \left(\frac{1}{\alpha} \right) + \log(A) \right).$$

Using that, from $\mathcal{A}_2(\alpha)$, $\log(1/\alpha) > 1$, we upper bound ε_α as

$$\varepsilon_\alpha \leq \frac{CK}{n^2} \left[\log \left(\frac{1}{\alpha} \right) \right]^2. \quad (\text{C.34})$$

Moreover, we can easily show that

$$K = \|h_{(2)}\|_\infty \leq C \left(\sup_{x, x' \in \mathbb{R}^p} k_\lambda(x, x') \right) \left(\sup_{y, y' \in \mathbb{R}^q} l_\mu(y, y') \right) = \frac{C(p, q)}{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}. \quad (\text{C.35})$$

By combining Equations (C.34) and (C.35), we obtain:

$$\varepsilon_\alpha \leq \frac{C}{n^2 \lambda_1 \dots \lambda_p \mu_1 \dots \mu_q} \left[\log \left(\frac{1}{\alpha} \right) \right]^2. \quad (\text{C.36})$$

Finally, using Equations (C.27), (C.30), (C.33) and (C.36) and the fact that, from Assumption $\mathcal{A}_2(\alpha)$,

$$\frac{1}{n \sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}} \log \left(\frac{1}{\alpha} \right) < 1,$$

we have the following inequality

$$q_{1-\alpha, 2}^{\lambda, \mu} \leq \frac{C(M_f, p, q)}{n \sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}} \log \left(\frac{1}{\alpha} \right). \quad (\text{C.37})$$

C.4.2 Upper bound of $q_{1-\alpha,3}^{\lambda,\mu}$

In this part, we give an upper bound for the $(1 - \alpha)$ -quantile of

$$\widehat{\text{HSIC}}_{\lambda,\mu}^{(3,D)} = \frac{1}{n(n-1)(n-2)} \sum_{(i,j,q) \in \mathbf{i}_3^n} h_{(3)}(Z_i, Z_j, Z_q),$$

where $h_{(3)}$ is define by

$$h_{(3)}(Z_i, Z_j, Z_q) = h_{i,j,q} = \mathbb{E}[h_{i,j,q,r} \mid Z_i, Z_j, Z_q] - \frac{1}{2!} \sum_{(t,u)}^{(i,j,q)} h_{t,u}.$$

For this, we use the concentration inequality (c), page 1501 of [Arcones and Gine, 1993]. We write for all $t > 0$,

$$\mathbb{P} \left(n^{-3/2} \left| \sum_{(i,j,q) \in \mathbf{i}_3^n} h_{(3)}(Z_i, Z_j, Z_q) \right| > t \right) \leq A \exp \left(- \frac{Bt^{2/3}}{M^{2/3} + K^{1/2}t^{1/6}n^{-1/4}} \right), \quad (\text{C.38})$$

where $K = \|h_{(3)}\|_\infty$, $M^2 = \mathbb{E}[h_{1,2,3}^2]$ and $A > 1$, $B > 0$ are absolute constant.

By setting $\varepsilon = \frac{t}{n^{3/2}}$ and using Equation (C.38), we have

$$\mathbb{P} \left(\frac{1}{n^3} \left| \sum_{(i,j,q) \in \mathbf{i}_3^n} h_{i,j,q} \right| > \varepsilon \right) \leq A \exp \left(- \frac{Bn\varepsilon^{2/3}}{M^{2/3} + K^{1/2}\varepsilon^{1/6}} \right).$$

Moreover, by adjusting the value of B , we can write

$$\mathbb{P} \left(\left| \widehat{\text{HSIC}}_{\lambda,\mu}^{(3,D)} \right| > \varepsilon \right) \leq A \exp \left(- \frac{Bn\varepsilon^{2/3}}{M^{2/3} + K^{1/2}\varepsilon^{1/6}} \right). \quad (\text{C.39})$$

Hence, if ε_α is a positive number verifying

$$A \exp \left(- \frac{Bn\varepsilon_\alpha^{2/3}}{M^{2/3} + K^{1/2}\varepsilon_\alpha^{1/6}} \right) = \alpha, \quad (\text{C.40})$$

then, we have the following inequality

$$q_{1-\alpha,3}^{\lambda,\mu} \leq \varepsilon_\alpha.$$

In order to upper bound ε_α in (C.40), we set $\gamma_\alpha = \varepsilon_\alpha^{1/6}$ and we obtain

$$Bn\gamma_\alpha^4 = K^{1/2} \log \left(\frac{A}{\alpha} \right) \gamma_\alpha + M^{2/3} \log \left(\frac{A}{\alpha} \right). \quad (\text{C.41})$$

The polynomial Equation (C.41) has no explicit solutions. However, it is possible to give an upper bound of its roots. Indeed,

$$Bn\gamma_\alpha^4 \leq 2 \max \{ K^{1/2}\gamma_\alpha, M^{2/3} \} \log \left(\frac{A}{\alpha} \right).$$

Case 1. If $K^{1/2}\gamma_\alpha \geq M^{2/3}$, then, γ_α verifies the following inequality,

$$\gamma_\alpha^3 \leq \frac{2K^{1/2}}{Bn} \log\left(\frac{A}{\alpha}\right) \leq \frac{CK^{1/2}}{n} \log\left(\frac{1}{\alpha}\right),$$

since $\log(1/\alpha) > 1$ in $\mathcal{A}_2(\alpha)$, one gets $\log(A/\alpha) \leq C \log(1/\alpha)$. Hence,

$$\varepsilon_\alpha \leq \frac{CK}{n^2} \left(\log\left(\frac{1}{\alpha}\right) \right)^2.$$

Moreover, once again, one can upper bound $K = \|h_{(3)}\|_\infty$ by

$$K \leq C \left[\sup_{x, x' \in \mathbb{R}^p} k_\lambda(x, x') \right] \left[\sup_{y, y' \in \mathbb{R}^q} l_\mu(y, y') \right] = \frac{C(p, q)}{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}.$$

Hence,

$$\varepsilon_\alpha \leq \frac{C(p, q)}{n^2 \lambda_1 \dots \lambda_p \mu_1 \dots \mu_q} \left(\log\left(\frac{1}{\alpha}\right) \right)^2,$$

and, since from Assumption $\mathcal{A}_2(\alpha)$,

$$\frac{1}{n \sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}} \log\left(\frac{1}{\alpha}\right) < 1,$$

we have the following inequality

$$\varepsilon_\alpha \leq \frac{C}{n \sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}} \log\left(\frac{1}{\alpha}\right). \quad (\text{C.42})$$

Case 2. If $K^{1/2}\gamma_\alpha \leq M^{2/3}$, then,

$$\gamma_\alpha^4 \leq \frac{2M^{2/3}}{Bn} \log\left(\frac{A}{\alpha}\right) \leq \frac{CM^{2/3}}{n} \log\left(\frac{1}{\alpha}\right),$$

since $\log(1/\alpha) > 1$ in $\mathcal{A}_2(\alpha)$. Therefore, ε_α can be upper bounded as

$$\varepsilon_\alpha \leq \frac{CM}{n^{3/2}} \left[\log\left(\frac{1}{\alpha}\right) \right]^{3/2}.$$

Moreover, using the law of total variance, one can upper bound $M^2 = \mathbb{E}[h_{1,2,3}^2]$ by

$$M^2 = \text{Var}(h_{1,2,3}) \leq C \text{Var}(h_{1,2,3,4}). \quad (\text{C.43})$$

Then, according Equation (C.20) (see Section C.3.2), under $P_{f_1 \otimes f_2}$, M can be upper bounded as

$$M \leq \frac{C(M_{f_1 \otimes f_2}, p, q)}{\sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}} \leq \frac{C(M_f, p, q)}{\sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}}.$$

Hence,

$$\varepsilon_\alpha \leq \frac{C(M_f, p, q)}{n^{3/2} \sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}} \left[\log\left(\frac{1}{\alpha}\right) \right]^{3/2},$$

Moreover, since both assumptions in $\mathcal{A}_2(\alpha)$ imply that $n^{-1} \log(1/\alpha) < 1$, we obtain

$$\varepsilon_\alpha \leq \frac{C(M_f, p, q)}{n\sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}} \log\left(\frac{1}{\alpha}\right). \quad (\text{C.44})$$

Finally, both (C.42) and (C.44) lead to

$$q_{1-\alpha,3}^{\lambda,\mu} \leq \frac{C(M_f, p, q)}{n\sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}} \log\left(\frac{1}{\alpha}\right). \quad (\text{C.45})$$

C.4.3 Upper bound of $q_{1-\alpha,4}^{\lambda,\mu}$

In this part, we give an upper bound for the $(1-\alpha)$ -quantile of

$$\widehat{\text{HSIC}}_{\lambda,\mu}^{(4,D)} = \frac{1}{n(n-1)(n-2)(n-3)} \sum_{(i,j,q,r) \in \mathbf{i}_4^n} h_{(4)}(Z_i, Z_j, Z_q, Z_r),$$

under $P_{f_1 \otimes f_2}$ where $h_{(4)}$ is define by

$$h_{(4)}(Z_i, Z_j, Z_q, Z_r) = \tilde{h}_{i,j,q,r} = h_{i,j,q,r} - \frac{1}{3!} \sum_{(t,u,v)}^{(i,j,q,r)} h_{t,u,v} - \frac{1}{2!} \sum_{(t,u)}^{(i,j,q,r)} h_{t,u}.$$

For this, we use the concentration inequality (d), page 1501 of [Arcones and Gine, 1993]. We have for all $t > 0$,

$$\mathbb{P}\left(\frac{1}{n^2} \left| \sum_{(i,j,q,r) \in \mathbf{i}_4^n} \tilde{h}_{i,j,q,r} \right| > t\right) \leq A \exp\left(-B\sqrt{\frac{t}{K}}\right),$$

where $A > 1$ and $B > 0$ are absolute constants and $K = \|h_{(4)}\|_\infty$.

By setting $\varepsilon = \frac{t}{n^2}$, we have

$$\mathbb{P}\left(\frac{1}{n^4} \left| \sum_{(i,j,q,r) \in \mathbf{i}_4^n} \tilde{h}_{i,j,q,r} \right| > \varepsilon\right) \leq A \exp\left(-Bn\sqrt{\frac{\varepsilon}{K}}\right).$$

Furthermore, by adjusting the constant B , we can replace $\frac{1}{n^4} \sum_{(i,j,q,r) \in \mathbf{i}_4^n} \tilde{h}_{i,j,q,r}$ by $\widehat{\text{HSIC}}_{\lambda,\mu}^{(4,D)}$ and obtain

$$\mathbb{P}\left(\left|\widehat{\text{HSIC}}_{\lambda,\mu}^{(4,D)}\right| > \varepsilon\right) \leq A \exp\left(-Bn\sqrt{\frac{\varepsilon}{K}}\right). \quad (\text{C.46})$$

Hence, if ε_α is a positive number verifying

$$A \exp\left(-Bn\sqrt{\frac{\varepsilon_\alpha}{K}}\right) = \alpha, \quad (\text{C.47})$$

then

$$q_{1-\alpha,4}^{\lambda,\mu} \leq \varepsilon_\alpha.$$

By resolving Equation (C.47), we obtain

$$\varepsilon_\alpha = \frac{BK}{n^2} \left[\log \left(\frac{A}{\alpha} \right) \right]^2.$$

Therefore, since $\log(1/\alpha) > 1$ in $\mathcal{A}_2(\alpha)$, we can easily show that

$$\varepsilon_\alpha \leq \frac{CK}{n^2} \left[\log \left(\frac{1}{\alpha} \right) \right]^2.$$

Moreover, as above, one can upper bound $K = \|h_{(4)}\|_\infty$ by

$$K \leq C \left[\sup_{x, x' \in \mathbb{R}^p} k_\lambda(x, x') \right] \left[\sup_{y, y' \in \mathbb{R}^q} l_\mu(y, y') \right] = \frac{C(p, q)}{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}.$$

Hence,

$$q_{1-\alpha, 4}^{\lambda, \mu} \leq \frac{C(M_f, p, q)}{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q n^2} \left(\log \left(\frac{1}{\alpha} \right) \right)^2.$$

Consequently, since from Assumption $\mathcal{A}_2(\alpha)$,

$$\frac{1}{n \sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}} \log \left(\frac{1}{\alpha} \right) < 1,$$

one finally obtains

$$q_{1-\alpha, 4}^{\lambda, \mu} \leq \frac{C(M_f, p, q)}{n \sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}} \log \left(\frac{1}{\alpha} \right). \quad (\text{C.48})$$

Finally, combining (C.25), (C.37), (C.45) and (C.48) ends the proof of Proposition 3.

C.5 Proof of Corollary 1

The proof of this corollary is immediately obtained from Lemma 1, Proposition 2 and Proposition 3.

C.6 Proof of Lemma 2

Recalling the formulation of $\text{HSIC}_{\lambda, \mu}(f)$ given in Equation (1.5) with $k = k_\lambda$ and $l = l_\mu$, we obtain

$$\begin{aligned} \text{HSIC}_{\lambda, \mu}(f) &= \int_{(\mathbb{R}^p \times \mathbb{R}^q)^2} k_\lambda(x, x') l_\mu(y, y') f(x, y) f(x', y') dx dy dx' dy' \\ &\quad - 2 \int_{(\mathbb{R}^p \times \mathbb{R}^q)^2} k_\lambda(x, x') l_\mu(y, y') f(x, y) f_1(x') f_2(y') dx dy dx' dy' \\ &\quad + \int_{(\mathbb{R}^p \times \mathbb{R}^q)^2} k_\lambda(x, x') l_\mu(y, y') f_1(x) f_2(y) f_1(x') f_2(y') dx dy dx' dy'. \end{aligned}$$

This expression can be factorized using the symmetry of the kernels k_λ and l_μ as

$$\begin{aligned} \text{HSIC}_{\lambda, \mu}(f) &= \int_{(\mathbb{R}^p \times \mathbb{R}^q)^2} k_\lambda(x, x') l_\mu(y, y') \left[f(x, y) - f_1(x) f_2(y) \right] \left[f(x', y') - f_1(x') f_2(y') \right] dx dy dx' dy' \\ &= \int_{(\mathbb{R}^p \times \mathbb{R}^q)^2} k_\lambda(x, x') l_\mu(y, y') \psi(x, y) \psi(x', y') dx dy dx' dy', \end{aligned}$$

where $\psi(x, y) = f(x, y) - f_1(x)f_2(y)$.

Thereafter, by replacing $k_\lambda(x, x')$ with $\varphi_\lambda(x - x')$ and replacing $l_\mu(y, y')$ with $\phi_\mu(y - y')$, where φ_λ and ϕ_μ are respectively the functions defined in Equation (1.4), one obtains

$$\begin{aligned} \text{HSIC}_{\lambda, \mu}(f) &= \int_{\mathbb{R}^p \times \mathbb{R}^q} \psi(x, y) \left[\int_{\mathbb{R}^p \times \mathbb{R}^q} \psi(x', y') \varphi_\lambda(x - x') \phi_\mu(y - y') dx' dy' \right] dx dy \\ &= \int_{\mathbb{R}^p \times \mathbb{R}^q} \psi(x, y) [\psi * (\varphi_\lambda \otimes \phi_\mu)](x, y) dx dy \\ &= \langle \psi, \psi * (\varphi_\lambda \otimes \phi_\mu) \rangle_2. \end{aligned}$$

C.7 Proof of Proposition 4

First recall that $\widehat{\text{HSIC}}_{\lambda, \mu}$ can be written as a U -statistic of order 4, that is

$$\widehat{\text{HSIC}}_{\lambda, \mu} = \frac{1}{n(n-1)(n-2)(n-3)} \sum_{(i, j, q, r) \in \mathbf{i}_4^n} h_{i, j, q, r},$$

where the general term $h_{i, j, q, r}$ of $\widehat{\text{HSIC}}_{\lambda, \mu}$ is defined by

$$h_{i, j, q, r} = \frac{1}{4!} \sum_{(t, u, v, w)}^{(i, j, q, r)} (k_{t, u} l_{t, u} + k_{t, u} l_{v, w} - 2k_{t, u} l_{t, v}). \quad (\text{C.49})$$

where the sum represents all ordered quadruples (t, u, v, w) drawn without replacement from (i, j, q, r) , and for all t, u in $\{1, \dots, n\}$,

$$k_{t, u} = k_\lambda(X_t, X_u) \quad \text{and} \quad l_{t, u} = l_\mu(Y_t, Y_u).$$

According to Equations (C.9) and (C.20), we already proved that

$$\text{Var}_f(\widehat{\text{HSIC}}_{\lambda, \mu}) \leq \frac{C}{n} \text{Var}_f(\mathbb{E}[h_{1, 2, 3, 4} \mid Z_1]) + \frac{C(M_f, p, q)}{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q n^2}. \quad (\text{C.50})$$

To prove the intended result, we need a sharper control of $\text{Var}_f(\mathbb{E}[h_{1, 2, 3, 4} \mid Z_1])$ in terms of $\|\psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2$, which is provided in Lemma 12.

Lemma 12. *For all λ in $(0, +\infty)^p$ and μ in $(0, +\infty)^q$, we have*

$$\text{Var}_f(\mathbb{E}[h_{1, 2, 3, 4} \mid Z_1]) \leq C(M_f) \|\psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2.$$

Finally, both Equation (C.50) and Lemma 12 end the proof of Proposition 4.

Proof of Lemma 12. The first step to upper bound $\text{Var}_f(\mathbb{E}[h_{1, 2, 3, 4} \mid Z_1])$ is to rewrite $h_{1, 2, 3, 4}$ by isolating all the terms depending on Z_1 .

$$\begin{aligned} h_{1, 2, 3, 4} &= \frac{1}{4!} \sum_{(t, u, v, w)}^{(1, 2, 3, 4)} [k_{t, u} l_{t, u} + k_{t, u} l_{v, w} - 2k_{t, u} l_{t, v}] \\ &= \frac{2}{4!} \sum_{(u, v, w)}^{(2, 3, 4)} [k_{1, u} l_{1, u} + k_{1, u} l_{v, w} + k_{u, v} l_{1, w} - k_{w, v} l_{w, 1} - k_{u, 1} l_{u, v} - k_{1, u} l_{1, v}] + R(Z_2, Z_3, Z_4), \end{aligned}$$

where the last sum represents all triplets (u, v, w) drawn without replacement from $(2, 3, 4)$ and $R(Z_2, Z_3, Z_4)$ is a random variable depending only on Z_2, Z_3 and Z_4 .

Then,

$$h_{1,2,3,4} = \frac{1}{12} \sum_{(u,v,w)}^{(2,3,4)} [k_{1,u}(l_{1,u} - l_{1,v}) - k_{u,1}(l_{u,v} - l_{v,w}) - (k_{w,v} - k_{u,v})l_{1,w}] + R(Z_2, Z_3, Z_4).$$

The random variable $R(Z_2, Z_3, Z_4)$ being independent from Z_1 , the variance of its expectation conditionally to Z_1 is equal to 0. It is then easy to see that $\text{Var}_f(\mathbb{E}[h_{1,2,3,4} | Z_1])$ can be upper bounded as follows:

$$\begin{aligned} \text{Var}_f(\mathbb{E}[h_{1,2,3,4} | Z_1]) &\leq C [\text{Var}_f(\mathbb{E}[k_{1,2}(l_{1,2} - l_{1,3}) | Z_1]) + \text{Var}_f(\mathbb{E}[k_{2,1}(l_{2,3} - l_{3,4}) | X_1]) \\ &\quad + \text{Var}_f(\mathbb{E}[(k_{2,3} - k_{4,3})l_{1,2} | Y_1])]. \end{aligned} \quad (\text{C.51})$$

By now, we reformulate the function $\psi * (\varphi_\lambda \otimes \phi_\mu)$ in a simpler form in order to link its \mathbb{L}_2 -norm with the upper bound given in Equation (C.51). For notational convenience, we denote $G_{\lambda,\mu} = \psi * (\varphi_\lambda \otimes \phi_\mu)$. Then

$$\begin{aligned} G_{\lambda,\mu}(x, y) &= \int_{\mathbb{R}^p \times \mathbb{R}^q} \psi(x', y') k_\lambda(x, x') l_\mu(y, y') \, dx' dy' \\ &= \int_{\mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}^q} k_\lambda(x, x') \left(l_\mu(y, y') - l_\mu(y, y'') \right) f(x', y') f_2(y'') \, dx' dy' dy'' \\ &= \mathbb{E} \left[k_\lambda(x, X') \left(l_\mu(y, Y') - l_\mu(y, Y'') \right) \right], \end{aligned}$$

where (X', Y') and Y'' are independent random variables with respective densities f and f_2 .

Thereafter, the conditional expectations in Equation (C.51) can all be expressed as follows:

$$\begin{aligned} \mathbb{E}[k_{1,2}(l_{1,2} - l_{1,3}) | Z_1] &= G_{\lambda,\mu}(X_1, Y_1), \\ \mathbb{E}[k_{2,1}(l_{2,3} - l_{3,4}) | X_1] &= \mathbb{E}[G_{\lambda,\mu}(X_1, Y_3) | X_1], \\ \mathbb{E}[(k_{2,3} - k_{4,3})l_{1,2} | Y_1] &= \mathbb{E}[G_{\lambda,\mu}(X_3, Y_1) | Y_1]. \end{aligned}$$

Thus, using the law of total variance [Weiss, 2006], we have the following upper bound:

$$\text{Var}_f(\mathbb{E}[h_{1,2,3,4} | Z_1]) \leq C \left[\text{Var}_f(G_{\lambda,\mu}(X_1, Y_1)) + \text{Var}_f(G_{\lambda,\mu}(X_1, Y_3)) + \text{Var}_f(G_{\lambda,\mu}(X_3, Y_1)) \right].$$

On the other hand, it is straightforward to upper bound the three variances in the last equation as

$$\begin{aligned} \text{Var}_f(G_{\lambda,\mu}(X_1, Y_1)) &\leq \|f\|_\infty \|G_{\lambda,\mu}\|_2^2, \\ \text{Var}_f(G_{\lambda,\mu}(X_1, Y_3)) &\leq \|f_1 \otimes f_2\|_\infty \|G_{\lambda,\mu}\|_2^2, \\ \text{Var}_f(G_{\lambda,\mu}(X_3, Y_1)) &\leq \|f_1 \otimes f_2\|_\infty \|G_{\lambda,\mu}\|_2^2. \end{aligned}$$

Finally, combining these inequalities with Equation (C.51) allows to upper bound $\text{Var}_f(\mathbb{E}[h_{1,2,3,4} | Z_1])$ as

$$\text{Var}_f(\mathbb{E}[h_{1,2,3,4} | Z_1]) \leq C(M_f) \|\psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2,$$

which ends the proof of Lemma 12. \square

C.8 Proof of Lemma 3

Recall that for any bandwidths $\lambda = (\lambda_1, \dots, \lambda_p)$ in $(0, +\infty)^p$ and $\mu = (\mu_1, \dots, \mu_q)$ in $(0, +\infty)^q$, φ_λ and ϕ_μ are defined in Equation (1.4) for any x in \mathbb{R}^p and y in \mathbb{R}^q ,

$$\varphi_\lambda(x) = \frac{1}{\lambda_1 \dots \lambda_p} g_p \left(\frac{x^{(1)}}{\lambda_1}, \dots, \frac{x^{(p)}}{\lambda_p} \right), \quad \phi_\mu(y) = \frac{1}{\mu_1 \dots \mu_q} g_q \left(\frac{y^{(1)}}{\mu_1}, \dots, \frac{y^{(q)}}{\mu_q} \right),$$

where g_p and g_q are the standard Gaussian density defined in Equation (1.3). The objective here is to provide an upper bound of the bias term $\|\psi - \psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2$ w.r.t λ and μ .

First of all, since $\psi - \psi * (\varphi_\lambda \otimes \phi_\mu)$ belongs to \mathbb{L}_2 , by Plancherel's theorem we obtain that

$$\begin{aligned} (2\pi)^{p+q} \|\psi - \psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2 &= \left\| \widehat{\psi - \psi * (\varphi_\lambda \otimes \phi_\mu)} \right\|_2^2 \\ &= \left\| \widehat{\psi} \left(1 - \widehat{\varphi_\lambda \otimes \phi_\mu} \right) \right\|_2^2 \\ &= \int_{\mathbb{R}^p \times \mathbb{R}^q} \left| 1 - \widehat{\varphi_\lambda \otimes \phi_\mu}(\xi, \zeta) \right|^2 \left| \widehat{\psi}(\xi, \zeta) \right|^2 d\xi d\zeta. \end{aligned}$$

Moreover, by definition of φ_λ and ϕ_μ (see Equation (1.4)),

$$\widehat{\varphi_\lambda \otimes \phi_\mu}(\xi, \zeta) = \widehat{g_p \otimes g_q}(\lambda\xi, \mu\zeta),$$

where $\lambda\xi = (\lambda_1\xi^{(1)}, \dots, \lambda_p\xi^{(p)})$ and $\mu\zeta = (\mu_1\zeta^{(1)}, \dots, \mu_q\zeta^{(q)})$. Besides, the Gaussian density satisfies for all (u, v) in $\mathbb{R}^p \times \mathbb{R}^q$,

$$\widehat{g_p \otimes g_q}(u, v) = (2\pi)^{(p+q)/2} g_p \otimes g_q(u, v) = \exp \left(-\frac{\|(u, v)\|^2}{2} \right).$$

Hence, the bias term satisfies

$$(2\pi)^{p+q} \|\psi - \psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2 = \int_{\mathbb{R}^p \times \mathbb{R}^q} \left[1 - \exp \left(-\frac{\|(\lambda\xi, \mu\zeta)\|^2}{2} \right) \right]^2 \left| \widehat{\psi}(\xi, \zeta) \right|^2 d\xi d\zeta.$$

In addition, for all $\delta > 0$, there exists T_δ in $[0, 1]$ such that

$$\forall x \geq T_\delta, \quad 1 - \exp(-x^2/2) \leq x^\delta. \quad (\text{C.52})$$

Indeed, the function $g_\delta : x \mapsto 1 - \exp(-x^2/2) - x^\delta$ is continuous on \mathbb{R}_+ , satisfies $g_\delta(0) = 0$, and for all $x \geq 1$, $g_\delta(x) < 0$ since

$$1 - \exp(-x^2/2) < 1 \leq x^\delta.$$

Note that if $\delta \leq 2$, then $T_\delta = 0$ since, in addition, for all x in $[0, 1]$,

$$1 - \exp(-x^2/2) \leq \frac{x^2}{2} \leq \frac{x^\delta}{2} \leq x^\delta.$$

Therefore, one can split the integral as

$$(2\pi)^{p+q} \|\psi - \psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2 = I_1 + I_2, \quad (\text{C.53})$$

where

$$\begin{aligned}
I_1 &= \int_{\|(\lambda\xi, \mu\zeta)\| < T_\delta} \left[1 - \exp\left(\frac{-\|(\lambda\xi, \mu\zeta)\|^2}{2}\right) \right]^2 |\widehat{\psi}(\xi, \zeta)|^2 d\xi d\zeta \\
&\leq \left(1 - e^{-T_\delta^2/2}\right)^2 \|\widehat{\psi}\|_2^2 \\
&\leq \left(1 - e^{-T_\delta^2/2}\right) (2\pi)^{p+q} \|\psi\|_2^2, \quad \text{since } \left(1 - e^{-T_\delta^2/2}\right) < 1,
\end{aligned} \tag{C.54}$$

and

$$\begin{aligned}
I_2 &= \int_{\|(\lambda\xi, \mu\zeta)\| \geq T_\delta} \left[1 - \exp\left(\frac{-\|(\lambda\xi, \mu\zeta)\|^2}{2}\right) \right]^2 |\widehat{\psi}(\xi, \zeta)|^2 d\xi d\zeta \\
&\leq \int_{\mathbb{R}^p \times \mathbb{R}^q} \|(\lambda\xi, \mu\zeta)\|^{2\delta} |\widehat{\psi}(\xi, \zeta)|^2 d\xi d\zeta,
\end{aligned}$$

by Equation (C.52). In addition, since for all $1 \leq i \leq p$, $\lambda_i^2 \leq \|(\lambda, \mu)\|^2$ and for all $1 \leq j \leq q$, $\mu_j^2 \leq \|(\lambda, \mu)\|^2$,

$$\|(\lambda\xi, \mu\zeta)\|^{2\delta} = \left(\sum_{i=1}^p \lambda_i^2 [\xi^{(i)}]^2 + \sum_{j=1}^q \mu_j^2 [\zeta^{(j)}]^2 \right)^\delta \leq \|(\lambda, \mu)\|^{2\delta} \|(\xi, \zeta)\|^{2\delta}.$$

Thus, since ψ belongs to $\mathcal{S}_{p+q}^\delta(R)$,

$$I_2 \leq (2\pi)^{p+q} R^{2\delta} \|(\lambda, \mu)\|^{2\delta}. \tag{C.55}$$

Thereafter, using Hölder's inequality if $\delta \geq 1$ and fact that $\|\cdot\|_{1/\delta} \leq \|\cdot\|_1$ if $\delta < 1$, it is straightforward to see that

$$\|(\lambda, \mu)\|^{2\delta} \leq C(p, q, \delta) \left[\sum_{i=1}^p \lambda_i^{2\delta} + \sum_{j=1}^q \mu_j^{2\delta} \right], \tag{C.56}$$

Finally, combining (C.54), (C.55) and (C.56) in (C.53) leads to

$$\|\psi - \psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2 \leq \left(1 - e^{-T_\delta^2/2}\right) \|\psi\|_2^2 + C(p, q, \delta, R) \left[\sum_{i=1}^p \lambda_i^{2\delta} + \sum_{j=1}^q \mu_j^{2\delta} \right].$$

Note once again that if $\delta \leq 2$, then $T_\delta = 0$, and one directly obtains that

$$\|\psi - \psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2 \leq C(p, q, \delta, R) \left[\sum_{i=1}^p \lambda_i^{2\delta} + \sum_{j=1}^q \mu_j^{2\delta} \right].$$

C.9 Proof of Theorem 2

Assume that ψ belongs to the Sobolev balls $\mathcal{S}_{p+q}^\delta(R, R')$ with $\delta, R, R' > 0$. One may notice that, by Lemma 3, since, $T_\delta \leq 1$, then

$$\|\psi - \psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2 \leq \left(1 - e^{-1/2}\right) \|\psi\|_2^2 + C(p, q, \delta, R) \left[\sum_{i=1}^p \lambda_i^{2\delta} + \sum_{j=1}^q \mu_j^{2\delta} \right].$$

Thus, since $M_f \leq R'$, one may easily deduce from Theorem 1 that $P_f(\Delta_\alpha^{\lambda, \mu} = 0) \leq \beta$ as soon as

$$e^{-1/2} \|\psi\|_2^2 > C(p, q, \delta, R) \left[\sum_{i=1}^p \lambda_i^{2\delta} + \sum_{j=1}^q \mu_j^{2\delta} \right] + \frac{C(R', p, q, \beta)}{n \sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}} \log \left(\frac{1}{\alpha} \right),$$

that is, since constants may vary from line to line,

$$\|\psi\|_2^2 > C(p, q, \delta, R) \left[\sum_{i=1}^p \lambda_i^{2\delta} + \sum_{j=1}^q \mu_j^{2\delta} \right] + \frac{C(R', p, q, \beta)}{n \sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}} \log \left(\frac{1}{\alpha} \right).$$

It now follows from the definition (1.1) of the uniform separation rate that

$$\left[\rho \left(\Delta_\alpha^{\lambda, \mu}, \mathcal{S}_{p+q}^\delta(R, R'), \beta \right) \right]^2 \leq C(p, q, \delta, R) \left[\sum_{i=1}^p \lambda_i^{2\delta} + \sum_{j=1}^q \mu_j^{2\delta} \right] + \frac{C(R', p, q, \beta)}{n \sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}} \log \left(\frac{1}{\alpha} \right).$$

C.10 Proof of Corollary 2

The objective here is to give the uniform separation rate having the smallest upper bound w.r.t. the sample size n , when ψ belongs to a Sobolev ball $\mathcal{S}_{p+q}^\delta(R, R')$. For this, we recall that according to Theorem 2, we have

$$\left[\rho \left(\Delta_\alpha^{\lambda, \mu}, \mathcal{S}_{p+q}^\delta(R, R'), \beta \right) \right]^2 \leq C(p, q, \delta, R) \left[\sum_{i=1}^p \lambda_i^{2\delta} + \sum_{j=1}^q \mu_j^{2\delta} \right] + \frac{C(R', p, q, \beta)}{n \sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}} \log \left(\frac{1}{\alpha} \right).$$

In order to have the smallest behavior of the right side of the last inequality w.r.t. n , one has then to choose bandwidths $\lambda^* = (\lambda_1^*, \dots, \lambda_p^*)$ and $\mu^* = (\mu_1^*, \dots, \mu_q^*)$ w.r.t. n in such a way that

$$\left[\sum_{i=1}^p \lambda_i^{*2\delta} + \sum_{j=1}^q \mu_j^{*2\delta} \right] \quad \text{and} \quad \frac{1}{n \sqrt{\lambda_1^* \dots \lambda_p^* \mu_1^* \dots \mu_q^*}}$$

have the same order. Thereafter, it is clear that all λ_i^* 's and μ_j^* 's have the same behavior w.r.t. n . It follows that for all i in $\{1, \dots, p\}$ and all j in $\{1, \dots, q\}$, we have

$$\lambda_i^* = \mu_j^* = n^{-2/(4\delta+p+q)}.$$

Consequently, the separation rate over $\mathcal{S}_{p+q}^\delta(R, R')$ can be upper bounded as

$$\rho \left(\Delta_\alpha^{\lambda^*, \mu^*}, \mathcal{S}_{p+q}^\delta(R, R'), \beta \right) \leq C(p, q, \alpha, \beta, \delta, R, R') n^{-2\delta/(4\delta+p+q)}.$$

C.11 Proof of Lemma 8

The objective here is to give an upper bound of the bias term $\|\psi - \psi * (\varphi_\lambda \otimes \phi_\mu)\|_2^2$ w.r.t. λ and μ , when ψ belongs to a Nikol'skii-Besov ball $\mathcal{N}_{2,p+q}^\delta(R)$, with $\delta = (\nu_1, \dots, \nu_p, \gamma_1, \dots, \gamma_q)$ in $(0, 2]^{p+q}$. We first set $b = \psi * (\varphi_\lambda \otimes \phi_\mu) - \psi$ and we write

$$\begin{aligned} b(x, y) &= \psi * (\varphi_\lambda \otimes \phi_\mu)(x, y) - \psi(x, y) \\ &= \int \psi(x', y') \varphi_\lambda(x - x') \phi_\mu(y - y') dx' dy' - \psi(x, y). \end{aligned}$$

Moreover, using Equation (1.4), the function b can be written in terms of the functions g_p and g_q defined in Equation (1.3) as

$$\begin{aligned} b(x, y) &= \frac{1}{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q} \int \psi(x', y') g_p \left(\frac{x_1 - x'_1}{\lambda_1}, \dots, \frac{x_p - x'_p}{\lambda_p} \right) g_q \left(\frac{y_1 - y'_1}{\mu_1}, \dots, \frac{y_q - y'_q}{\mu_q} \right) dx' dy' \\ &\quad - \psi(x, y) \\ &= \int \psi(x_1 + \lambda_1 u_1, \dots, x_p + \lambda_p u_p, y_1 + \mu_1 v_1, \dots, y_q + \mu_q v_q) g_p(u_1, \dots, u_p) g_q(v_1, \dots, v_q) du dv \\ &\quad - \psi(x, y). \end{aligned}$$

Thereafter, using that $\int_{\mathbb{R}^p} g_p = \int_{\mathbb{R}^q} g_q = 1$, the function b can be expressed as

$$b(x, y) = \int g_p(u_1, \dots, u_p) g_q(v_1, \dots, v_q) \left[\psi(x_1 + \lambda_1 u_1, \dots, y_q + \mu_q v_q) - \psi(x, y) \right] du dv.$$

Let us from now define for all i in $\{1, \dots, p\}$ and j in $\{1, \dots, q\}$, the functions $b_{1,i}$ and $b_{2,j}$ by

$$\begin{aligned} b_{1,i}(x, y) &= \int g_p(u_1, \dots, u_p) g_q(v_1, \dots, v_q) \omega_{1,i}(x, y, u_1, \dots, u_i) du dv, \\ b_{2,j}(x, y) &= \int g_p(u_1, \dots, u_p) g_q(v_1, \dots, v_q) \omega_{2,j}(x, y, u_1, \dots, u_p, v_1, \dots, v_j) du dv, \end{aligned}$$

where the function $\omega_{1,i}$ is defined as

$$\begin{aligned} \omega_{1,i}(x, y, u_1, \dots, u_i) &= \psi(x_1 + \lambda_1 u_1, \dots, x_i + \lambda_i u_i, x_{i+1}, \dots, x_p, y) \\ &\quad - \psi(x_1 + \lambda_1 u_1, \dots, x_{i-1} + \lambda_{i-1} u_{i-1}, x_i, \dots, x_p, y), \end{aligned}$$

while the function $\omega_{2,j}$ is defined as

$$\begin{aligned} \omega_{2,j}(x, y, u_1, \dots, u_p, v_1, \dots, v_j) &= \psi(x_1 + \lambda_1 u_1, \dots, x_p + \lambda_p u_p, y_1 + \mu_1 v_1, \dots, y_j + \mu_j v_j, y_{j+1}, \dots, y_q) \\ &\quad - \psi(x_1 + \lambda_1 u_1, \dots, x_p + \lambda_p u_p, y_1 + \mu_1 v_1, \dots, y_{j-1} + \mu_{j-1} v_{j-1}, y_j, \dots, y_q). \end{aligned}$$

It is then easy to see that the function b is the sum of all the functions $b_{1,i}$ and $b_{2,j}$

$$b(x, y) = \sum_{i=1}^p b_{1,i}(x, y) + \sum_{j=1}^q b_{2,j}(x, y).$$

One can then deduce that it would be sufficient for the control of the \mathbb{L}_2 -norm of b , to control the \mathbb{L}_2 -norms of all the functions $b_{1,i}$ and $b_{2,j}$. Using the triangular inequality, we have

$$\|b\|_2 \leq \sum_{i=1}^p \|b_{1,i}\|_2 + \sum_{j=1}^q \|b_{2,j}\|_2. \quad (\text{C.57})$$

By now, let us upper bound $\|b_{1,i}\|_2^2$ and $\|b_{2,j}\|_2^2$ for all i in $\{1, \dots, p\}$ and j in $\{1, \dots, q\}$. We distinguish two cases.

Case 1. Assume that $0 < \nu_i \leq 1$. We first recall that $\|b_{1,i}\|_2^2$ can be written as

$$\|b_{1,i}\|_2^2 = \int \left[\int g_p(u_1, \dots, u_p) g_q(v_1, \dots, v_q) \omega_{1,i}(x, y, u_1, \dots, u_i) du dv \right]^2 dx dy.$$

We use the following lemma from page 13 of [Tsybakov, 2009].

Lemma 13. Let $\rho : \mathbb{R}^d \times \mathbb{R}^{d'} \rightarrow \mathbb{R}$ be a Borel function, then we have the following inequality:

$$\int \left(\int \rho(\theta, z) d\theta \right)^2 dz \leq \left[\int \left(\int \rho^2(\theta, z) dz \right)^{1/2} d\theta \right]^2.$$

By applying Lemma 13 to the function

$$((u, v), (x, y)) \mapsto g_p(u_1, \dots, u_p) g_q(v_1, \dots, v_q) \omega_{1,i}(x, y, u_1, \dots, u_i),$$

we obtain

$$\begin{aligned} \|b_{1,i}\|_2^2 &\leq \left[\int \left(\int g_p^2(u_1, \dots, u_p) g_q^2(v_1, \dots, v_q) \omega_{1,i}^2(x, y, u_1, \dots, u_i) dx dy \right)^{1/2} dudv \right]^2 \\ &= \left[\int g_p(u_1, \dots, u_p) g_q(v_1, \dots, v_q) \left(\int \omega_{1,i}^2(x, y, u_1, \dots, u_i) dx dy \right)^{1/2} dudv \right]^2. \end{aligned} \quad (\text{C.58})$$

On the other hand, since ψ belongs to the Nikol'skii-Besov ball $\mathcal{N}_{2,p+q}^\delta(R)$, we have

$$\left(\int \omega_{1,i}^2(x, y, u_1, \dots, u_i) dx dy \right)^{1/2} \leq R \lambda_i^{\nu_i} |u_i|^{\nu_i}.$$

We then have by injecting this last inequation in Equation (C.58), that

$$\|b_{1,i}\|_2^2 \leq C(\nu_i, R) \lambda_i^{2\nu_i}.$$

Case 2. Now assume that $1 < \nu_i \leq 2$. In this case the function ψ has continuous first-order partial derivatives. Using Taylor expansion with integral form of the remainder w.r.t. the i^{th} variable of ψ , we have

$$\omega_{1,i}(x, y, u_1, \dots, u_i) = \lambda_i u_i \int_0^1 (1 - \tau) D_i^1 \psi(x_1 + \lambda_1 u_1, \dots, x_i + \tau \lambda_i u_i, x_{i+1}, \dots, y) d\tau,$$

where D_i^1 denotes the first-order partial derivative of ψ w.r.t. the i^{th} variable.

Thereafter, by injecting the last equation in the expression of $b_{1,i}$, we obtain

$$\begin{aligned} b_{1,i}(x, y) &= \int \lambda_i u_i g_p(u_1, \dots, u_p) g_q(v_1, \dots, v_q) \\ &\quad \left[\int_0^1 (1 - \tau) D_i^1 \psi(x_1 + \lambda_1 u_1, \dots, x_i + \tau \lambda_i u_i, x_{i+1}, \dots, y) d\tau \right] dudv. \end{aligned}$$

Furthermore, using the fact that g_p is of order 2, we have that $\int u_i g_p(u_1, \dots, u_p) du_i = 0$. The function $b_{1,i}$ can then be written as

$$b_{1,i}(x, y) = \int \lambda_i u_i g_p(u_1, \dots, u_p) g_q(v_1, \dots, v_q) \left[\int_0^1 (1 - \tau) D_i^1 \omega_{1,i}(x, y, u_1, \dots, \tau u_i) d\tau \right] dudv.$$

We have then the following equation for the \mathbb{L}_2 -norm of $b_{1,i}$:

$$\|b_{1,i}\|_2^2 = \int \left[\int \lambda_i u_i g_p(u_1, \dots, u_p) g_q(v_1, \dots, v_q) \left(\int_0^1 (1-\tau) D_i^1 \omega_{1,i}(x, y, u_1, \dots, \tau u_i) d\tau \right) dudv \right]^2 dx dy.$$

By now, we use as in Case 1 of Lemma 13 in order to upper bound $\|b_{1,i}\|_2^2$. We then obtain:

$$\|b_{1,i}\|_2^2 \leq \left(\int \left[\int \left(\lambda_i u_i g_p(u_1, \dots, u_p) g_q(v_1, \dots, v_q) \int_0^1 (1-\tau) D_i^1 \omega_{1,i}(x, y, u_1, \dots, \tau u_i) d\tau \right)^2 dx dy \right]^{1/2} dudv \right)^2$$

Then,

$$\|b_{1,i}\|_2^2 \leq \left(\int \lambda_i u_i g_p(u_1, \dots, u_p) g_q(v_1, \dots, v_q) \left[\int \left(\int_0^1 (1-\tau) D_i^1 \omega_{1,i}(x, y, u_1, \dots, \tau u_i) d\tau \right)^2 dx dy \right]^{1/2} dudv \right)^2$$

We apply a second time Lemma 13. For this, consider the function

$$\rho: ((x, y), \tau) \mapsto (1-\tau) D_i^1 \omega_{1,i}(x, y, u_1, \dots, \tau u_i),$$

we then have

$$\|b_{1,i}\|_2^2 \leq \left(\int \lambda_i u_i g_p(u_1, \dots, u_p) g_q(v_1, \dots, v_q) \left[\int_0^1 (1-\tau) \left(\int (D_i^1 \omega_{1,i}(x, y, u_1, \dots, \tau u_i))^2 dx dy \right)^{1/2} d\tau \right] dudv \right)^2. \quad (\text{C.59})$$

On the other hand, using that ψ belongs to the Nikol'skii-Besov ball $\mathcal{N}_{2,p+q}^\delta(R)$,

$$\left(\int (D_i^1 \omega_{1,i}(x, y, u_1, \dots, \tau u_i))^2 dx dy \right)^{1/2} \leq R \lambda_i^{\nu_i-1} |\tau u_i|^{\nu_i-1}.$$

We then obtain by injecting this last inequation in Equation (C.59), that

$$\|b_{1,i}\|_2^2 \leq C(\nu_i, R) \lambda_i^{2\nu_i}.$$

Besides, for all j in $\{1, \dots, q\}$, by similar arguments, one can prove that

$$\|b_{2,j}\|_2^2 \leq C(\gamma_j, R) \mu_j^{2\gamma_j}.$$

Consequently, according to Equation (C.57), we have the following upper bound of $\|b\|_2^2$

$$\|b\|_2^2 \leq C(\delta, R) \left[\sum_{i=1}^p \lambda_i^{2\nu_i} + \sum_{j=1}^q \mu_j^{2\gamma_j} \right].$$

C.12 Proof of Theorem 5

This proof is similar to the one of Theorem 2. Assume that ψ belongs to the Sobolev balls $\mathcal{N}_{2,p+q}^\delta(R, R')$ with $\delta = (\nu_1, \dots, \nu_p, \gamma_1, \dots, \gamma_q)$ in $(0, 2]^{p+q}$ and $R, R' > 0$. Then, since $M_f \leq R'$, one may easily deduce from Theorem 1 and Lemma 8 that $P_f(\Delta_\alpha^{\lambda, \mu} = 0) \leq \beta$ as soon as

$$\|\psi\|_2^2 > C(\delta, R) \left[\sum_{i=1}^p \lambda_i^{2\nu_i} + \sum_{j=1}^q \mu_j^{2\gamma_j} \right] + \frac{C(R', p, q, \beta)}{n \sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}} \log \left(\frac{1}{\alpha} \right).$$

One can then conclude from the definition (1.1) of the uniform separation rate that

$$\left[\rho \left(\Delta_\alpha^{\lambda, \mu}, \mathcal{N}_{2,p+q}^\delta(R, R'), \beta \right) \right]^2 \leq C(\delta, R) \left[\sum_{i=1}^p \lambda_i^{2\nu_i} + \sum_{j=1}^q \mu_j^{2\gamma_j} \right] + \frac{C(R', p, q, \beta)}{n \sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}} \log \left(\frac{1}{\alpha} \right).$$

C.13 Proof of Corollary 4

We aim here to give the uniform separation rate having the smallest upper bound w.r.t. the sample size n , when ψ belongs to a Nikol'skii-Besov ball $\mathcal{N}_{2,p+q}^\delta(R, R')$, with $\delta = (\nu_1, \dots, \nu_p, \gamma_1, \dots, \gamma_q)$ in $(0, 2]^{p+q}$. We first recall that Theorem 5 shows that

$$\left[\rho \left(\Delta_\alpha^{\lambda, \mu}, \mathcal{N}_{2,p+q}^\delta(R, R'), \beta \right) \right]^2 \leq C(\delta, R) \left[\sum_{i=1}^p \lambda_i^{2\nu_i} + \sum_{j=1}^q \mu_j^{2\gamma_j} \right] + \frac{C(R', p, q, \beta)}{n \sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}} \log \left(\frac{1}{\alpha} \right).$$

Hence, in order to minimize the right side of the last inequality w.r.t. n , we choose bandwidths $\lambda^* = (\lambda_1^*, \dots, \lambda_p^*)$ and $\mu^* = (\mu_1^*, \dots, \mu_q^*)$ w.r.t. n such that

$$\left[\sum_{i=1}^p \lambda_i^{*2\nu_i} + \sum_{j=1}^q \mu_j^{*2\gamma_j} \right] \quad \text{and} \quad \frac{1}{n \sqrt{\lambda_1^* \dots \lambda_p^* \mu_1^* \dots \mu_q^*}}$$

have the same order. Let us set for all i in $\{1, \dots, p\}$ and all j in $\{1, \dots, q\}$, $\lambda_i^* = n^{a_i}$ and $\mu_j^* = n^{b_j}$. Then, it is clear that for all i and all j

$$2a_i \nu_i = 2b_j \gamma_j = -\frac{1}{2} \left[\sum_{r=1}^p a_r + \sum_{s=1}^q b_s \right] - 1. \quad (\text{C.60})$$

One can first express all a_i 's and all b_j 's w.r.t a_1 as

$$a_i = a_1 \frac{\nu_1}{\nu_i} \quad \text{and} \quad b_j = a_1 \frac{\nu_1}{\gamma_j}.$$

Thereafter, using Equation (C.60) we have

$$2a_1 \nu_1 = \frac{-a_1 \nu_1}{2\eta} - 1.$$

Thus, we first write that $a_1 = \frac{-2\eta}{\nu_1(4\eta + 1)}$. We next obtain for all i and for all j that

$$a_i = \frac{-2\eta}{\nu_i(4\eta + 1)} \quad \text{and} \quad b_j = \frac{-2\eta}{\gamma_j(4\eta + 1)}.$$

Note that the condition $n \geq (\log(1/\alpha))^{1+1/(4\eta)}$ ensures that (λ^*, μ^*) satisfies Assumption $\mathcal{A}_2(\alpha)$. Consequently, the separation rate over $\mathcal{N}_{2,p+q}^\delta(R, R')$ can be upper bound as

$$\rho\left(\Delta_{\alpha}^{\lambda^*, \mu^*}, \mathcal{N}_{2,p+q}^\delta(R, R'), \beta\right) \leq C(p, q, \alpha, \beta, \delta, R, R') n^{-\frac{2\eta}{1+4\eta}}.$$

C.14 Proof of Lemma 4

Let α be in $(0, 1)$, we first prove that $u_\alpha \geq \alpha$. For this, we apply Bonferroni's inequality

$$\begin{aligned} P_{f_1 \otimes f_2} \left(\sup_{(\lambda, \mu) \in \mathcal{W}} \left\{ \widehat{\text{HSIC}}_{\lambda, \mu} - q_{1-\alpha e^{-\omega_{\lambda, \mu}}}^{\lambda, \mu} \right\} > 0 \right) &= P_{f_1 \otimes f_2} \left(\bigcup_{(\lambda, \mu) \in \mathcal{W}} \left\{ \widehat{\text{HSIC}}_{\lambda, \mu} > q_{1-\alpha e^{-\omega_{\lambda, \mu}}}^{\lambda, \mu} \right\} \right) \\ &\leq \sum_{(\lambda, \mu) \in \mathcal{W}} P_{f_1 \otimes f_2} \left(\widehat{\text{HSIC}}_{\lambda, \mu} > q_{1-\alpha e^{-\omega_{\lambda, \mu}}}^{\lambda, \mu} \right) \\ &\leq \sum_{(\lambda, \mu) \in \mathcal{W}} \alpha e^{-\omega_{\lambda, \mu}} \\ &\leq \alpha. \end{aligned}$$

Then, by definition of u_α we have $u_\alpha \geq \alpha$. Thereafter, we obtain

$$\begin{aligned} P_f(\Delta_\alpha = 0) &= P_f \left(\bigcap_{(\lambda, \mu) \in \mathcal{W}} \left\{ \widehat{\text{HSIC}}_{\lambda, \mu} \leq q_{1-u_\alpha e^{-\omega_{\lambda, \mu}}}^{\lambda, \mu} \right\} \right) \\ &\leq \inf_{(\lambda, \mu) \in \mathcal{W}} P_f \left(\widehat{\text{HSIC}}_{\lambda, \mu} \leq q_{1-u_\alpha e^{-\omega_{\lambda, \mu}}}^{\lambda, \mu} \right) \\ &\leq \inf_{(\lambda, \mu) \in \mathcal{W}} P_f \left(\widehat{\text{HSIC}}_{\lambda, \mu} \leq q_{1-\alpha e^{-\omega_{\lambda, \mu}}}^{\lambda, \mu} \right) \\ &= \inf_{(\lambda, \mu) \in \mathcal{W}} \left\{ P_f \left(\Delta_{\alpha e^{-\omega_{\lambda, \mu}}}^{\lambda, \mu} = 0 \right) \right\}, \end{aligned}$$

which concludes the proof.

C.15 Proof of Theorems 3 and 6

Let α and β be in $(0, 1)$. According to Lemma 4, $P_f(\Delta_\alpha = 0) \leq \beta$ as soon as there exists (λ, μ) in \mathcal{W} such that

$$P_f \left(\Delta_{\alpha e^{-\omega_{\lambda, \mu}}}^{\lambda, \mu} = 0 \right) \leq \beta.$$

Then, according to Theorem 2 if ψ belongs to $\mathcal{S}_{p+q}^\delta(R, R')$, or Theorem 5 if ψ belongs to $\mathcal{N}_{2,p+q}^\delta(R, R')$, one can take the infimum of the upper bounds for the uniform separation rates over $\mathcal{S}_{p+q}^\delta(R)$ (resp. over $\mathcal{N}_{2,p+q}^\delta(R, R')$) of the single tests over \mathcal{W} while replacing $\log(1/\alpha)$ by $\log(1/\alpha) + \omega_{\lambda, \mu}$.

C.16 Proof of Corollary 3

Assume that ψ belongs to $\mathcal{S}_{p+q}^\delta(R, R')$ with regularity parameter $\delta > 0$ and positive radiuses R, R' . Let us first verify that $\mathcal{A}_2(\alpha e^{-\omega_{\lambda, \mu}})$ holds for all (λ, μ) in \mathcal{W} . Let (λ, μ) in \mathcal{W} . Then, by definition of $M_n^{p,q}$,

$$n \sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q} \geq n 2^{-M_n^{p,q}(\frac{p+q}{2})} \geq \log(n).$$

Moreover,

$$\begin{aligned} \log\left(\frac{1}{\alpha}\right) + \omega_{\lambda,\mu} &\leq \log\left(\frac{1}{\alpha}\right) + 2\log\left(M_n^{p,q} \frac{\pi}{\sqrt{6}}\right) \\ &\leq \log\left(\frac{1}{\alpha}\right) + 2\log\left(\frac{\pi}{\sqrt{6}}\right) + \log\left(\frac{2}{p+q}\right) + \log(\log_2(n)). \end{aligned}$$

This implies that there exists $C(p, q, \alpha)$ such that for $n \geq C(p, q, \alpha)$, $\mathcal{A}_2(\alpha e^{-\omega_{\lambda,\mu}})$ holds, and this for all (λ, μ) in \mathcal{W} . Hence, using Theorem 3, we have the following inequality

$$\begin{aligned} \left[\rho\left(\Delta_\alpha, \mathcal{S}_{p+q}^\delta(R, R'), \beta\right)\right]^2 &\leq C(p, q, \beta, \delta, R, R') \inf_{(\lambda, \mu) \in \mathcal{W}} \left\{ \left[\sum_{i=1}^p \lambda_i^{2\delta} + \sum_{j=1}^q \mu_j^{2\delta} \right] + \right. \\ &\quad \left. + \frac{1}{n\sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q}} \left(\log\left(\frac{1}{\alpha}\right) + \omega_{\lambda,\mu} \right) \right\}. \end{aligned}$$

Let us take $(\lambda^*, \mu^*) = 2^{-m^*} \mathbf{1}_{p+q}$ with m^* satisfying the condition

$$\left(\frac{n}{\log \log(n)}\right)^{\frac{2}{p+q+4\delta}} < 2^{m^*} \leq 2 \left(\frac{n}{\log \log(n)}\right)^{\frac{2}{p+q+4\delta}}.$$

Note that there exists a positive constant $C(p, q, \delta)$ such that for $n \geq C(p, q, \delta)$, $m^* \in \{1, \dots, M_n^{p,q}\}$, which implies that $(\lambda^*, \mu^*) \in \mathcal{W}$. Noticing that

$$\begin{aligned} \left[\sum_{i=1}^p (\lambda_i^*)^{2\delta} + \sum_{j=1}^q (\mu_j^*)^{2\delta} \right] &+ \frac{1}{n\sqrt{\lambda_1^* \dots \lambda_p^* \mu_1^* \dots \mu_q^*}} \left(\log\left(\frac{1}{\alpha}\right) + \omega_{\lambda^*, \mu^*} \right) \\ &\leq C(p, q, \alpha, \delta) \left(\frac{\log \log(n)}{n} \right)^{4\delta/(4\delta+p+q)}, \end{aligned}$$

and applying Theorem 3, we obtain the desired result.

C.17 Proof of Corollary 5

Assume that ψ belongs to $\mathcal{N}_{2,p+q}^\delta(R, R')$ with regularity parameter $\delta = (\nu_1, \dots, \nu_p, \gamma_1, \dots, \gamma_q)$ in $(0, 2]^{p+q}$ and positive radiuses R, R' . Let us first verify that $\mathcal{A}_2(\alpha e^{-\omega_{\lambda,\mu}})$ holds for all (λ, μ) in \mathcal{W} . The condition $\sum_{i=1}^p m_{1,i} + \sum_{j=1}^q m_{2,j} \leq 2\log_2 \left[\left(\frac{n}{\log(n)} \right) \right]$ implies that for all (λ, μ) in \mathcal{W} ,

$$n\sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q} \geq \log(n).$$

Moreover, by definition of the weights $\omega_{\lambda,\mu}$, we have that for all (λ, μ) in \mathcal{W} ,

$$\begin{aligned} \log\left(\frac{1}{\alpha}\right) + \omega_{\lambda,\mu} &\leq \log\left(\frac{1}{\alpha}\right) + 2(p+q) \log\left(\frac{\pi}{\sqrt{6}}\right) + 2 \sum_{i=1}^p \log(m_{1,i}) + 2 \sum_{j=1}^q \log(m_{2,i}) \\ &\leq \log\left(\frac{1}{\alpha}\right) + 2(p+q) \log\left(\frac{2\pi}{\sqrt{6}}\right) + 2(p+q) \log(\log_2(n)). \end{aligned}$$

This implies that there exists some constant $C(p, q, \alpha)$ such that for $n \geq C(p, q, \alpha)$, $\mathcal{A}_2(\alpha e^{-\omega_{\lambda, \mu}})$ holds for all (λ, μ) in \mathcal{W} . Hence, using Theorem 6, we have the following inequality

$$\begin{aligned} \left[\rho \left(\Delta_\alpha, \mathcal{N}_{2,p+q}^\delta(R, R'), \beta \right) \right]^2 &\leq C(p, q, \beta, \delta, R, R') \inf_{(\lambda, \mu) \in \mathcal{W}} \left\{ \left[\sum_{i=1}^p \lambda_i^{2\nu_i} + \sum_{j=1}^q \mu_j^{2\gamma_j} \right] \right. \\ &\quad \left. + \frac{1}{\sqrt{\lambda_1 \dots \lambda_p \mu_1 \dots \mu_q n}} \left[\log \left(\frac{1}{\alpha} \right) + \omega_{\lambda, \mu} \right] \right\}. \end{aligned}$$

Let us take $\lambda^* = (2^{-m_{1,1}^*}, \dots, 2^{-m_{1,p}^*})$ and $\mu^* = (2^{-m_{2,1}^*}, \dots, 2^{-m_{2,q}^*})$, where the integers $m_{1,1}^*, \dots, m_{1,p}^*, m_{2,1}^*, \dots, m_{2,q}^*$ are defined by the inequalities

$$\left(\frac{n}{\log \log(n)} \right)^{\frac{2\eta}{\nu_i(1+4\eta)}} < 2^{m_{1,i}^*} \leq 2 \left(\frac{n}{\log \log(n)} \right)^{\frac{2\eta}{\nu_i(1+4\eta)}}$$

and

$$\left(\frac{n}{\log \log(n)} \right)^{\frac{2\eta}{\gamma_j(1+4\eta)}} < 2^{m_{2,j}^*} \leq 2 \left(\frac{n}{\log \log(n)} \right)^{\frac{2\eta}{\gamma_j(1+4\eta)}},$$

where $\eta^{-1} = \sum_{i=1}^p (\nu_i)^{-1} + \sum_{j=1}^q (\gamma_j)^{-1}$. Note that there exists a positive constant $C(\delta)$ such that for $n \geq C(\delta)$, (λ^*, μ^*) belongs to \mathcal{W} . Then, we obviously have

$$\begin{aligned} \left[\rho \left(\Delta_\alpha, \mathcal{N}_{2,p+q}^\delta(R, R'), \beta \right) \right]^2 &\leq C(p, q, \beta, \delta, R, R') \left\{ \left[\sum_{i=1}^p (\lambda_i^*)^{2\nu_i} + \sum_{j=1}^q (\mu_j^*)^{2\gamma_j} \right] \right. \\ &\quad \left. + \frac{1}{\sqrt{\lambda_1^* \dots \lambda_p^* \mu_1^* \dots \mu_q^* n}} \left(\log \left(\frac{1}{\alpha} \right) + \omega_{\lambda^*, \mu^*} \right) \right\}. \end{aligned}$$

By definition of the integers $m_{1,1}^*, \dots, m_{1,p}^*, m_{2,1}^*, \dots, m_{2,q}^*$, we have

$$(\lambda_i^*)^{-1/2} = 2^{m_{1,i}^*/2} \leq \sqrt{2} \left(\frac{n}{\log \log(n)} \right)^{\frac{\eta}{\nu_i(1+4\eta)}} \quad \text{and} \quad (\mu_j^*)^{-1/2} = 2^{m_{2,j}^*/2} \leq \sqrt{2} \left(\frac{n}{\log \log(n)} \right)^{\frac{\eta}{\gamma_j(1+4\eta)}}.$$

Therefore, we obtain

$$(\lambda_1^* \dots \lambda_p^* \mu_1^* \dots \mu_q^*)^{-1/2} \leq 2^{(p+q)/2} \left(\frac{n}{\log \log(n)} \right)^{\frac{1}{(1+4\eta)}}. \quad (\text{C.61})$$

Let us now upper bound $\omega_{\lambda^*, \mu^*}$. We first write

$$\begin{aligned} \omega_{\lambda^*, \mu^*} &= 2 \sum_{i=1}^p \log \left(m_{1,i}^* \times \frac{\pi}{\sqrt{6}} \right) + 2 \sum_{j=1}^q \log \left(m_{2,j}^* \times \frac{\pi}{\sqrt{6}} \right) \\ &= 2 \log (m_{1,1}^* \dots m_{1,p}^* m_{2,1}^* \dots m_{2,q}^*) + 2(p+q) \log \left(\frac{\pi}{\sqrt{6}} \right). \end{aligned}$$

Moreover, it is easy to see that for $n \geq C(\delta)$,

$$m_{1,i}^* \leq \frac{2\eta}{\nu_i(1+4\eta)} \log_2(n) \quad \text{and} \quad m_{2,j}^* \leq \frac{2\eta}{\gamma_j(1+4\eta)} \log_2(n).$$

Then, for $n \geq C(\delta)$,

$$\log(m_{1,1}^* \dots m_{1,p}^* m_{2,1}^* \dots m_{2,q}^*) \leq C(\delta) \log \log(n).$$

Thereafter, $\omega_{\lambda^*, \mu^*}$ can be upper bound as

$$\omega_{\lambda^*, \mu^*} \leq C(\delta, p, q) \log \log(n). \quad (\text{C.62})$$

From Equations (C.61) and (C.62), we have

$$\frac{1}{n \sqrt{\lambda_1^* \dots \lambda_p^* \mu_1^* \dots \mu_q^*}} \left(\log \left(\frac{1}{\alpha} \right) + \omega_{\lambda^*, \mu^*} \right) \leq C(\alpha, \delta, p, q) \left(\frac{\log \log(n)}{n} \right)^{\frac{4\eta}{1+4\eta}}. \quad (\text{C.63})$$

We aim now to upper bound $\sum_{i=1}^p (\lambda_i^*)^{2\nu_i} + \sum_{j=1}^q (\mu_j^*)^{2\gamma_j}$. By definition of the integers $m_{1,1}^*, \dots, m_{1,p}^*, m_{2,1}^*, \dots, m_{2,q}^*$,

$$(\lambda_i^*)^{2\nu_i} \leq \left(\frac{\log \log(n)}{n} \right)^{\frac{4\eta}{1+4\eta}} \quad \text{and} \quad (\mu_j^*)^{2\gamma_j} \leq \left(\frac{\log \log(n)}{n} \right)^{\frac{4\eta}{1+4\eta}}.$$

Therefore, we obtain

$$\sum_{i=1}^p (\lambda_i^*)^{2\nu_i} + \sum_{j=1}^q (\mu_j^*)^{2\gamma_j} \leq (p+q) \left(\frac{\log \log(n)}{n} \right)^{\frac{4\eta}{1+4\eta}}. \quad (\text{C.64})$$

Consequently, from Equations (C.63) and (C.64),

$$\rho \left(\Delta_\alpha, \mathcal{N}_{2,p+q}^\delta(R, R'), \beta \right) \leq C(p, q, \alpha, \beta, \delta, R, R') \left(\frac{\log \log(n)}{n} \right)^{\frac{2\eta}{1+4\eta}},$$

which ends the proof of Corollary 5

C.18 Proof of Lemma 5

Assume there exists a distribution f_0 that satisfies (\mathcal{H}_0) such that the probability measure $P_{\nu_{\rho_*}}$ is absolutely continuous w.r.t. P_{f_0} and verifies Equation (4.1).

Let us first lower bound $\beta[\mathcal{F}_{\rho_*}(\mathcal{C}_\delta)]$ w.r.t. the distributions $P_{\nu_{\rho_*}}$ and P_{f_0} . We recall that

$$\beta[\mathcal{F}_{\rho_*}(\mathcal{C}_\delta)] = \inf_{\Delta_\alpha} \sup_{f \in \mathcal{F}_{\rho_*}(\mathcal{C}_\delta)} P_f(\Delta_\alpha = 0).$$

Using the assumption $\nu_{\rho_*}(\mathcal{F}_{\rho_*}(\mathcal{C}_\delta)) \geq 1 - \eta$, we obtain the following inequalities

$$\begin{aligned} P_{\nu_{\rho_*}}(\Delta_\alpha = 0) &= \int_{\mathbb{L}_2(\mathbb{R}^p \times \mathbb{R}^q)} P_f(\Delta_\alpha = 0) d\nu_{\rho_*}(f) \\ &\leq \int_{\mathcal{F}_{\rho_*}(\mathcal{C}_\delta)} P_f(\Delta_\alpha = 0) d\nu_{\rho_*}(f) + \eta \\ &\leq \sup_{f \in \mathcal{F}_{\rho_*}(\mathcal{C}_\delta)} P_f(\Delta_\alpha = 0) + \eta. \end{aligned}$$

This leads to

$$\sup_{f \in \mathcal{F}_{\rho_*}(\mathcal{C}_\delta)} P_f(\Delta_\alpha = 0) \geq P_{\nu_{\rho_*}}(\Delta_\alpha = 0) - \eta.$$

Hence, we have

$$\begin{aligned}
\beta[\mathcal{F}_{\rho_*}(\mathcal{C}_\delta)] &\geq \inf_{\Delta_\alpha} P_{\nu_{\rho_*}}(\Delta_\alpha = 0) - \eta \\
&\geq 1 - \sup_{\Delta_\alpha} P_{\nu_{\rho_*}}(\Delta_\alpha = 1) - \eta \\
&\geq 1 - \alpha - \sup_{\Delta_\alpha} |P_{\nu_{\rho_*}}(\Delta_\alpha = 1) - P_{f_0}(\Delta_\alpha = 1)| - \eta.
\end{aligned}$$

We denote by $\|P_{\nu_{\rho_*}} - P_{f_0}\|_{TV}$ the total variation distance between the distributions $P_{\nu_{\rho_*}}$ and P_{f_0} . We recall that,

$$\|P_{\nu_{\rho_*}} - P_{f_0}\|_{TV} = \sup_{E \in \mathcal{E}} |P_{\nu_{\rho_*}}(E) - P_{f_0}(E)|,$$

where \mathcal{E} is the space of measurable sets in $\mathbb{R}^{n(p+q)}$. We then obtain

$$\beta[\mathcal{F}_{\rho_*}(\mathcal{C}_\delta)] \geq 1 - \alpha - \eta - \|P_{\nu_{\rho_*}} - P_{f_0}\|_{TV}.$$

Notice that,

$$\|P_{\nu_{\rho_*}} - P_{f_0}\|_{TV} = \sup_{E \in \mathcal{E}} [P_{\nu_{\rho_*}}(E) - P_{f_0}(E)] = \sup_{E \in \mathcal{E}} [P_{f_0}(E) - P_{\nu_{\rho_*}}(E)].$$

It is then straightforward to show that

$$\begin{aligned}
\|P_{\nu_{\rho_*}} - P_{f_0}\|_{TV} &= \frac{1}{2} \int_{\mathbb{R}^{n(p+q)}} |L_{\nu_{\rho_*}} - 1| \, dP_{f_0} \\
&= \frac{1}{2} \mathbb{E}_{P_{f_0}} [|L_{\nu_{\rho_*}}(\mathbb{Z}_n) - 1|] \\
&\leq \frac{1}{2} (\mathbb{E}_{P_{f_0}} [L_{\nu_{\rho_*}}^2(\mathbb{Z}_n)] - 1)^{1/2},
\end{aligned}$$

where the last inequality holds by applying Cauchy-Schwarz and the fact that $\mathbb{E}_{P_{f_0}} [L_{\nu_{\rho_*}}(\mathbb{Z}_n)] = 1$. Thus,

$$\beta[\mathcal{F}_{\rho_*}(\mathcal{C}_\delta)] \geq 1 - \alpha - \eta - \frac{1}{2} (\mathbb{E}_{P_{f_0}} [L_{\nu_{\rho_*}}^2(\mathbb{Z}_n)] - 1)^{1/2}.$$

If the condition (4.1) holds, we then obtain

$$\beta[\mathcal{F}_{\rho_*}(\mathcal{C}_\delta)] > \beta.$$

Furthermore, using that $\mathcal{F}_{\rho_*}(\mathcal{C}_\delta) \subset \mathcal{F}_\rho(\mathcal{C}_\delta)$ for all $\rho \leq \rho_*$, we have

$$\beta[\mathcal{F}_\rho(\mathcal{C}_\delta)] > \beta.$$

Let us now prove that this implies the lower bound

$$\rho(\mathcal{C}_\delta, \alpha, \beta) = \inf_{\Delta_\alpha} \rho(\Delta_\alpha, \mathcal{C}_\delta, \beta) \geq \rho_*. \quad (\text{C.65})$$

Assume $\beta[\mathcal{F}_{\rho_*}(\mathcal{C}_\delta)] > \beta$, then

$$\forall \Delta_\alpha, \quad \sup_{f \in \mathcal{F}_{\rho_*}(\mathcal{C}_\delta)} P_f(\Delta_\alpha = 0) > \beta.$$

In particular, since the family $\{\mathcal{F}_\rho(\mathcal{C}_\delta)\}_{\rho > 0}$ is non increasing for the inclusion,

$$\forall \Delta_\alpha, \quad \rho(\Delta_\alpha, \mathcal{C}_\delta, \beta) = \inf \left\{ \rho > 0 ; \quad \sup_{f \in \mathcal{F}_\rho(\mathcal{C}_\delta)} P_f(\Delta_\alpha = 0) \leq \beta \right\} > \rho_*,$$

which directly implies (C.65).

C.19 Proof of Lemma 6

Proof of 1 Assume that

$$C_0 \leq \min\{1, R' - 1\}e^{p+q}. \quad (\text{C.66})$$

Let us first prove that f_θ is a density function. First, it is obvious from Equation (4.3) that

$$\int_{\mathbb{R}^{p+q}} f_\theta(x, y) \, dx \, dy = 1,$$

since $\mathbb{1}_{[0,1]^{p+q}}$ is a probability density function and that $\int_{\mathbb{R}} G(x) \, dx = 0$. It remains to check that f_θ is a non-negative function under Assumption (C.66).

Let $j = (j_1, \dots, j_p)$ in $\{1, \dots, M_n\}^p$ and $l = (l_1, \dots, l_q)$ in $\{1, \dots, M_n\}^q$. Knowing that for all $1 \leq r \leq p$ and all $1 \leq s \leq q$, the supports of the functions $G_{h_n}(\cdot - j_r h_n)$ and $G_{h_n}(\cdot - l_s h_n)$ are respectively the intervals $[(j_r - 1)h_n, j_r h_n]$ and $[(l_s - 1)h_n, l_s h_n]$, we deduce that the support of the function

$$g_{n,j,l} : (x, y) \mapsto \prod_{r=1}^p G_{h_n}(x_r - j_r h_n) \prod_{s=1}^q G_{h_n}(y_s - l_s h_n) \quad (\text{C.67})$$

is the set

$$D_{(j,l)} = \prod_{r=1}^p [(j_r - 1)h_n, j_r h_n] \times \prod_{s=1}^q [(l_s - 1)h_n, l_s h_n]. \quad (\text{C.68})$$

These supports are then disjoint for different multi-indexes (j, l) in $I_{n,p,q}$ and have as union set $(0, 1]^{p+q}$ (since $M_n h_n = 1$). In particular, for all (x, y) in $(0, 1]^{p+q}$,

$$\begin{aligned} \left| \sum_{(j,l) \in I_{n,p,q}} \theta_{(j,l)} \prod_{r=1}^p G_{h_n}(x_r - j_r h_n) \prod_{s=1}^q G_{h_n}(y_s - l_s h_n) \right| &\leq \frac{1}{h_n^{p+q}} \left(\sup_{t \in [-1,0]} |G(t)| \right)^{p+q} \\ &= \frac{1}{(e h_n)^{p+q}}. \end{aligned} \quad (\text{C.69})$$

Hence, if (x, y) belongs to $[0, 1]^{p+q}$, then since $h_n \leq 1$,

$$f_\theta(x, y) \geq 1 - C_0 \frac{h_n^\delta}{e^{p+q}} \geq 1 - \frac{C_0}{e^{p+q}} \geq 0,$$

by equation (C.66). Otherwise, if $(x, y) \notin [0, 1]^{p+q}$, then $f_\theta(x, y) = 0$. In particular, for all (x, y) in \mathbb{R}^{p+q} , $f_\theta(x, y) \geq 0$.

Remains to prove that $\max\{\|f_\theta\|_\infty, \|f_{\theta,1}\|_\infty, \|f_{\theta,2}\|_\infty\} \leq R'$. On the one hand, since $f_{\theta,1} = \mathbb{1}_{[0,1]^p}$ and $f_{\theta,2} = \mathbb{1}_{[0,1]^q}$, we directly obtain

$$\|f_{\theta,1}\|_\infty = \|f_{\theta,2}\|_\infty = 1 \leq R'.$$

On the other hand, by (C.69), for all (x, y) in \mathbb{R}^{p+q} ,

$$|f_\theta(x, y)| \leq 1 + C_0 \frac{h_n^\delta}{e^{p+q}} \leq 1 + \frac{C_0}{e^{p+q}}.$$

Hence, assuming (C.66) directly leads to $\|f_\theta\|_\infty \leq R'$, which ends the proof of this point.

Proof of 2 Let us prove that, for all θ in $\{-1, 1\}^{M_n^{p+q}}$, f_θ satisfies

$$\|f_\theta - f_{\theta,1} \otimes f_{\theta,2}\|_2 = C(p, q, \delta, R, R', \eta) h_n^\delta.$$

Since, $\int_{\mathbb{R}} G(t) dt = 0$, we know that $f_{\theta,1} = \mathbb{1}_{[0,1]^p}$ and $f_{\theta,2} = \mathbb{1}_{[0,1]^q}$, thus $f_{\theta,1} \otimes f_{\theta,2} = \mathbb{1}_{[0,1]^{p+q}}$ and

$$f_\theta - f_{\theta,1} \otimes f_{\theta,2} = C_0 h_n^{\delta+(p+q)} \sum_{(j,l) \in I_{n,p,q}} \theta_{(j,l)} g_{n,j,l}(x, y),$$

where the functions $g_{n,j,l}$ are defined in (C.67), with disjoint supports.

In particular,

$$\|f_\theta - f_{\theta,1} \otimes f_{\theta,2}\|_2^2 = C_0^2 h_n^{2\delta+2(p+q)} \sum_{(j,l) \in I_{n,p,q}} \|g_{n,j,l}\|_2^2.$$

Moreover, for all $(j, l) \in I_{n,p,q}$,

$$\begin{aligned} \|g_{n,j,l}\|_2^2 &= \int_{\mathbb{R}^{p+q}} \left[\prod_{r=1}^p G_{h_n}^2(x_r - j_r h_n) \prod_{s=1}^q G_{h_n}^2(y_s - l_s h_n) \right] dx_1 \dots dx_p dy_1 \dots dy_q \\ &= \left[\prod_{r=1}^p \left(\int_{\mathbb{R}} G_{h_n}^2(x_r - j_r h_n) dx_r \right) \right] \times \left[\prod_{s=1}^q \left(\int_{\mathbb{R}} G_{h_n}^2(y_s - l_s h_n) dy_s \right) \right], \end{aligned}$$

and for all k in $\{1, \dots, M_n\}$, a simple change of variables implies that

$$\int_{\mathbb{R}} G_{h_n}^2(t - kh_n) dt = \frac{1}{h_n^2} \int_{\mathbb{R}} G^2\left(\frac{t - kh_n}{h_n}\right) dt = \frac{1}{h_n} \int_{\mathbb{R}} G^2(t) dt = \frac{\|G\|_2^2}{h_n},$$

since G belongs to $\mathbb{L}_2(\mathbb{R})$. We thus deduce that

$$\|g_{n,j,l}\|_2^2 = \frac{\|G\|_2^{2(p+q)}}{h_n^{p+q}}, \quad (\text{C.70})$$

and that, since the cardinality of $I_{n,p,q}$ equals M_n^{p+q} , recalling that $M_n h_n = 1$,

$$\|f_\theta - f_{\theta,1} \otimes f_{\theta,2}\|_2^2 = C_0^2 \|G\|_2^{2(p+q)} h_n^{2\delta+2(p+q)} \times \frac{M_n^{p+q}}{h_n^{p+q}} = C_0^2 \|G\|_2^{2(p+q)} h_n^{2\delta}.$$

C.20 Proof of Lemma 7

Let us prove that there exists a positive constant $C(p, q, \delta, \eta)$ such that, if $C_0^2 \leq (2\pi)^{p+q} R^2 / [2C(p, q, \delta, \eta)]$, then the random function $f_\Theta - f_{\Theta,1} \otimes f_{\Theta,2}$ belongs to the Sobolev ball $\mathcal{S}_{p+q}^\delta(R)$ with probability greater than $1 - \eta$. This point relies on Lemma [Butucea, 2007, Lemma 2] recalled below.

Lemma 14 ([Butucea, 2007]). *Let G be the function defined in Equation (4.2). Then G is an infinitely differentiable function such that $\int_{\mathbb{R}} G(x) dx = 0$. Its Fourier transform verifies*

$$|\widehat{G}(u)| \leq C \exp\left(-a\sqrt{|u|}\right) \quad \text{as } |u| \rightarrow \infty,$$

for some positive constants C and a . Moreover, \widehat{G} is an infinitely differentiable and bounded function.

According to the Fourier transform properties, we write, for all θ in $\{-1, 1\}^{M_n^{p+q}}$,

$$\widehat{f}_\theta(u, v) = \widehat{f}_{\theta,1} \otimes \widehat{f}_{\theta,2}(u, v) + C_0 h_n^{\delta+(p+q)} \sum_{(j,l) \in I_{n,p,q}} \theta_{j,l} \prod_{r=1}^p \exp(iu_r j_r h_n) \widehat{G}(h_n u_r) \prod_{s=1}^q \exp(iv_s l_s h_n) \widehat{G}(h_n v_s).$$

Then,

$$\left| \widehat{f}_\theta(u, v) - \widehat{f}_{\theta,1} \otimes \widehat{f}_{\theta,2}(u, v) \right|^2 = H_{1,n}(u, v) + H_{2,n}(u, v, \theta), \quad (\text{C.71})$$

where the functions $H_{1,n}$ and $H_{2,n}$ are respectively defined by

$$H_{1,n}(u, v) = C_0^2 M_n^{p+q} h_n^{2\delta+2(p+q)} \left(\prod_{r=1}^p \left| \widehat{G}(h_n u_r) \right|^2 \right) \left(\prod_{s=1}^q \left| \widehat{G}(h_n v_s) \right|^2 \right), \quad (\text{C.72})$$

$$H_{2,n}(u, v, \theta) = C_0^2 h_n^{2\delta+2(p+q)} \sum_{\substack{(j_1, l_1) \in I_{n,p,q} \\ (j_2, l_2) \in I_{n,p,q} \\ (j_1, l_1) \neq (j_2, l_2)}} \theta_{j_1, l_1} \theta_{j_2, l_2} \mathcal{G}_{j_1, l_1, j_2, l_2}(h_n u, h_n v), \quad (\text{C.73})$$

and, the function $\mathcal{G}_{j_1, l_1, j_2, l_2}$ is defined for all indexes $j_k = (j_{k,1}, \dots, j_{k,p})$ and $l_k = (l_{k,1}, \dots, l_{k,q})$, $k = 1, 2$, as

$$\mathcal{G}_{j_1, l_1, j_2, l_2} : (u, v) \mapsto \left(\prod_{r=1}^p \exp(iu_r (j_{1,r} - j_{2,r})) \left| \widehat{G}(u_r) \right|^2 \right) \left(\prod_{s=1}^q \exp(iv_s (l_{1,s} - l_{2,s})) \left| \widehat{G}(v_s) \right|^2 \right). \quad (\text{C.74})$$

By now, our aim is to prove that

$$\mathbb{P} \left(\int_{\mathbb{R}^{p+q}} \|(u, v)\|^{2\delta} \left| \widehat{f}_\theta(u, v) - \widehat{f}_{\theta,1} \otimes \widehat{f}_{\theta,2}(u, v) \right|^2 du dv \leq (2\pi)^{p+q} R^2 \right) \geq 1 - \eta.$$

First, by Equation (C.56), we have

$$\|(u, v)\|^{2\delta} \leq C(p, q, \delta) \left[\sum_{i=1}^p |u_i|^{2\delta} + \sum_{j=1}^q |v_j|^{2\delta} \right]. \quad (\text{C.75})$$

We then obtain from Equations (C.72) and (C.75) the following result,

$$\begin{aligned} & \int_{\mathbb{R}^{p+q}} \|(u, v)\|^{2\delta} H_{1,n}(u, v) du dv \\ & \leq C(p, q, \delta) C_0^2 M_n^{p+q} h_n^{2\delta+2(p+q)} \left(\int_{\mathbb{R}} |t|^{2\delta} \left| \widehat{G}(h_n t) \right|^2 dt \right) \left(\int_{\mathbb{R}} \left| \widehat{G}(h_n z) \right|^2 dz \right)^{p+q-1} \\ & = C(p, q, \delta) C_0^2 (M_n h_n)^{p+q} \left(\int_{\mathbb{R}} |t|^{2\delta} \left| \widehat{G}(t) \right|^2 dt \right) \left(\int_{\mathbb{R}} \left| \widehat{G}(z) \right|^2 dz \right)^{p+q-1}. \end{aligned}$$

The functions $t \mapsto |t|^{2\delta} \left| \widehat{G}(t) \right|^2$ and $z \mapsto \left| \widehat{G}(z) \right|^2$ being integrable according to Lemma 14, we have

$$\begin{aligned} \int_{\mathbb{R}^{p+q}} \|(u, v)\|^{2\delta} H_{1,n}(u, v) du dv & \leq C(p, q, \delta) C_0^2 (M_n h_n)^{p+q} \\ & \leq (2\pi)^{p+q} R^2 / 2 \end{aligned} \quad (\text{C.76})$$

provided that $C(p, q, \delta)C_0^2 \leq (2\pi)^{p+q}R^2/2$.

To complete the proof, let us now consider the random part. Starting from the expression of $H_{2,n}$ in (C.73), we write

$$\int_{\mathbb{R}^{p+q}} \|(u, v)\|^{2\delta} H_{2,n}(u, v, \Theta) du dv = C_0^2 h_n^{p+q} \sum_{\substack{(j_1, l_1) \in I_{n,p,q} \\ (j_2, l_2) \in I_{n,p,q} \\ (j_1, l_1) \neq (j_2, l_2)}} \Theta_{j_1, l_1} \Theta_{j_2, l_2} \int_{\mathbb{R}^{p+q}} \|(u, v)\|^{2\delta} \mathcal{G}_{j_1, l_1, j_2, l_2}(u, v) du dv.$$

Noting that $\mathcal{G}_{j_2, l_2, j_1, l_1} = \overline{\mathcal{G}_{j_1, l_1, j_2, l_2}}$, we get

$$\sum_{\substack{(j_1, l_1) \in I_{n,p,q} \\ (j_2, l_2) \in I_{n,p,q} \\ (j_1, l_1) \neq (j_2, l_2)}} \Theta_{j_1, l_1} \Theta_{j_2, l_2} \mathcal{G}_{j_1, l_1, j_2, l_2}(u, v) = \sum_{(j_1, l_1) \prec (j_2, l_2) \in I_{n,p,q}} \Theta_{j_1, l_1} \Theta_{j_2, l_2} (\mathcal{G}_{j_1, l_1, j_2, l_2}(u, v) + \overline{\mathcal{G}_{j_1, l_1, j_2, l_2}})$$

where $(j_1, l_1) \prec (j_2, l_2)$ means that (j_1, l_1) is strictly smaller than (j_2, l_2) in the lexicographic order. In the following, we prove that there exists some positive constant $C(p, q, \delta, \eta)$ such that, with probability greater than $1 - \eta$, we have

$$\sum_{(j_1, l_1) \prec (j_2, l_2) \in I_{n,p,q}} \Theta_{j_1, l_1} \Theta_{j_2, l_2} \int_{\mathbb{R}^{p+q}} \|(u, v)\|^{2\delta} (\mathcal{G}_{j_1, l_1, j_2, l_2}(u, v) + \overline{\mathcal{G}_{j_1, l_1, j_2, l_2}}) du dv \leq C(p, q, \delta, \eta) M_n^{p+q}. \quad (\text{C.77})$$

Since $M_n h_n = 1$, this implies that

$$\mathbb{P}\left(\int_{\mathbb{R}^{p+q}} \|(u, v)\|^{2\delta} H_{2,n}(u, v, \Theta) du dv \leq (2\pi)^{p+q} R^2/2\right) \geq 1 - \eta,$$

choosing C_0 such that $C(p, q, \delta, \eta)C_0^2 \leq (2\pi)^{p+q}R^2/2$, and concludes the proof.

In order to show that (C.77) holds with probability greater than $1 - \eta$, we use Hoeffding's inequality. For the sake of completeness, let us first recall this inequality.

Lemma 15. [*Hoeffding, 1963*] *Let Z_1, \dots, Z_n be independent real random variables such that for all i , $a \leq Z_i \leq b$. Then we have, for all $x > 0$,*

$$\mathbb{P}(|Z_1 + \dots + Z_n| \geq x) \leq 2 \exp\left(-\frac{2x^2}{n(b-a)^2}\right).$$

We apply Hoeffding's inequality to the variables $(Z_{j_1, l_1, j_2, l_2})_{(j_1, l_1) \in I_{n,p,q}, (j_2, l_2) \in I_{n,p,q}, (j_1, l_1) \prec (j_2, l_2)}$, where

$$Z_{j_1, l_1, j_2, l_2} = \Theta_{j_1, l_1} \Theta_{j_2, l_2} \int_{\mathbb{R}^{p+q}} \|(u, v)\|^{2\delta} (\mathcal{G}_{j_1, l_1, j_2, l_2}(u, v) + \overline{\mathcal{G}_{j_1, l_1, j_2, l_2}}(u, v)) du dv.$$

One easily verifies that the variables $(Z_{j_1, l_1, j_2, l_2})_{(j_1, l_1) \in I_{n,p,q}, (j_2, l_2) \in I_{n,p,q}, (j_1, l_1) \prec (j_2, l_2)}$ are independent. Furthermore,

$$|Z_{j_1, l_1, j_2, l_2}| \leq 2 \int_{\mathbb{R}^{p+q}} \|(u, v)\|^{2\delta} \prod_{r=1}^p |\widehat{G}(u_r)|^2 \prod_{s=1}^q |\widehat{G}(v_s)|^2 \leq C(p, q, \delta)$$

by Lemma 14. Hence, we obtain from Hoeffding's inequality that for all $x > 0$,

$$\mathbb{P} \left(\left| \sum_{\substack{(j_1, l_1) \in I_{n,p,q} \\ (j_2, l_2) \in I_{n,p,q} \\ (j_1, l_1) \neq (j_2, l_2)}} Z_{j_1, l_1, j_2, l_2} \right| \geq x \right) \leq 2 \exp \left(-\frac{x^2}{2C^2(p, q, \delta) M_n^{2p+2q}} \right).$$

We deduce from the above inequality that

$$\mathbb{P} \left(\left| \sum_{\substack{(j_1, l_1) \in I_{n,p,q} \\ (j_2, l_2) \in I_{n,p,q} \\ (j_1, l_1) \neq (j_2, l_2)}} Z_{j_1, l_1, j_2, l_2} \right| \geq C(p, q, \delta, \eta) M_n^{p+q} \right) \leq \eta,$$

which yields (C.77) and concludes the proof.

C.21 Proof of Proposition 5

Let $\mathbb{Z}_n = (X_i, Y_i)_{1 \leq i \leq n}$ be an i.i.d sample with common uniform distribution P_{f_0} on $[0, 1]^{p+q}$.

For simplicity, denote for all $1 \leq i \leq n$ and all (j, l) in $I_{n,p,q}$,

$$a_{i,j,l} = C_0 h_n^{\delta+(p+q)} g_{n,j,l}(X_i, Y_i) = C_0 h_n^{\delta+(p+q)} \prod_{r=1}^p G_{h_n}(X_i^{(r)} - j_r h_n) \prod_{s=1}^q G_{h_n}(Y_i^{(s)} - l_s h_n),$$

where $g_{n,j,l}$ is defined in Equation (C.67), such that $f_\theta(X_i, Y_i) = 1 + \sum_{(j,l) \in I_{n,p,q}} \theta_{(j,l)} a_{i,j,l}$. Note that $a_{i,j,l} \neq 0$ if and only if (X_i, Y_i) belongs to the set $D_{(j,l)}$ defined in Equation (C.68).

Then, since $f_0 = \mathbb{1}_{[0,1]^{p+q}}$, the likelihood ratio equals

$$\begin{aligned} L_{\nu_{\rho_n^*}}(\mathbb{Z}_n) &= \frac{dP_{\nu_{\rho_n^*}}}{dP_{f_0}}(\mathbb{Z}_n) = \int \prod_{i=1}^n \frac{f_\theta}{f_0}(X_i, Y_i) \pi(d\theta) \\ &= \mathbb{E}_\Theta \left[\prod_{i=1}^n \left(1 + \sum_{(j,l) \in I_{n,p,q}} \Theta_{(j,l)} a_{i,j,l} \right) \right], \end{aligned}$$

where $\Theta = (\Theta_{(j,l)})_{(j,l) \in I_{n,p,q}}$ has i.i.d. Rademacher components $\Theta_{(j,l)}$, and $\mathbb{E}_\Theta[\cdot]$ denotes the expectation w.r.t. Θ .

Noticing that for all $1 \leq i \leq n$, there exists a unique (j, l) in $I_{n,p,q}$ such that $a_{i,j,l} \neq 0$, we obtain

$$1 + \sum_{(j,l) \in I_{n,p,q}} \Theta_{(j,l)} a_{i,j,l} = \prod_{(j,l) \in I_{n,p,q}} (1 + \Theta_{(j,l)} a_{i,j,l}).$$

Thus,

$$\begin{aligned} L_{\nu_{\rho_n^*}}(\mathbb{Z}_n) &= \mathbb{E}_{\Theta} \left[\prod_{(j,l) \in I_{n,p,q}} \prod_{i=1}^n (1 + \Theta_{(j,l)} a_{i,j,l}) \right] \\ &= \prod_{(j,l) \in I_{n,p,q}} \left[\frac{1}{2} \prod_{i=1}^n (1 - a_{i,j,l}) + \frac{1}{2} \prod_{i=1}^n (1 + a_{i,j,l}) \right]. \end{aligned}$$

Moreover, for ε in $\{-1, 1\}$,

$$\prod_{i=1}^n (1 + \varepsilon a_{i,j,l}) = 1 + \sum_{k=1}^n \varepsilon^k \left(\sum_{1 \leq i_1 < \dots < i_k \leq n} a_{i_1,j,l} \dots a_{i_k,j,l} \right).$$

Hence, by cancelling the odd terms, we obtain

$$\begin{aligned} \frac{1}{2} \prod_{i=1}^n (1 - a_{i,j,l}) + \frac{1}{2} \prod_{i=1}^n (1 + a_{i,j,l}) &= 1 + \sum_{k=1}^{[n/2]} \sum_{1 \leq i_1 < \dots < i_{2k} \leq n} a_{i_1,j,l} \dots a_{i_{2k},j,l} \\ &= 1 + \sum_{k=1}^{[n/2]} A_{k,j,l}, \end{aligned}$$

where $[\cdot]$ denotes the integer part, and

$$A_{k,j,l} = \sum_{1 \leq i_1 < \dots < i_{2k} \leq n} a_{i_1,j,l} \dots a_{i_{2k},j,l}. \quad (\text{C.78})$$

Thus,

$$\left[L_{\nu_{\rho_n^*}}(\mathbb{Z}_n) \right]^2 = \prod_{(j,l) \in I_{n,p,q}} \left(1 + \sum_{k=1}^{[n/2]} A_{k,j,l} \right)^2 = \prod_{(j,l) \in I_{n,p,q}} (1 + B_{j,l}),$$

where

$$B_{j,l} = 2 \sum_{k=1}^{[n/2]} A_{k,j,l} + \sum_{k,k'=1}^{[n/2]} A_{k,j,l} A_{k',j,l}. \quad (\text{C.79})$$

Then,

$$\left[L_{\nu_{\rho_n^*}}(\mathbb{Z}_n) \right]^2 = 1 + \sum_{m=1}^{M_n^{p+q}} \frac{1}{m!} \sum_{(j_1,l_1), \dots, (j_m,l_m)}^{\neq} B_{j_1,l_1} \dots B_{j_m,l_m}, \quad (\text{C.80})$$

where \sum^{\neq} means that the indexes are all distinct.

After tedious computations, up to a possible permutation of the indexes $(j_1, l_1), \dots, (j_m, l_m)$, we can express the product $B_{j_1,l_1} \dots B_{j_m,l_m}$ as a sum of terms of the form

$$2^P \sum_{k_1=1}^{[n/2]} \sum_{k_2'', \dots, k_M''=1}^{[n/2]} 2A_{k_1,j_1,l_1} \times A_{k_2'',j_2',l_2'} \times \dots \times A_{k_M'',j_M',l_M'} \quad (\text{C.81})$$

or

$$2^Q \sum_{k_1, k'_1=1}^{[n/2]} \sum_{k''_2, \dots, k''_M=1}^{[n/2]} A_{k_1, j_1, l_1} A_{k'_1, j_1, l_1} \times A_{k''_2, j'_2, l'_2} \times \dots \times A_{k''_M, j'_M, l'_M} \quad (\text{C.82})$$

where P and Q are integers, $M \in \{m-1, \dots, 2m-2\}$ and $(j'_2, l'_2), \dots, (j'_M, l'_M)$ are drawn in $(j_2, l_2), \dots, (j_m, l_m)$ such that each (j_r, l_r) for $2 \leq r \leq m$ appears exactly once or twice. To be more precise, P and Q count the number of indexes $(j_r, l_r)_r$, appearing exactly once in the product.

First note that in Equation (C.81), the index (j_1, l_1) appears only once. Moreover

$$\begin{aligned} & \mathbb{E}_{f_0} [A_{k_1, j_1, l_1} \times A_{k''_2, j'_2, l'_2} \times \dots \times A_{k''_M, j'_M, l'_M}] \\ &= \sum_{i_{1,1} < \dots < i_{1,2k_1}} \sum_{i_{2,1} < \dots < i_{2,2k''_2}} \dots \sum_{i_{M,1} < \dots < i_{M,2k''_M}} \mathbb{E}_{f_0} [a_{i_{1,1}, j_1, l_1} \times \dots \times a_{i_{1,2k_1}, j_1, l_1} \times \\ & \quad \times a_{i_{2,1}, j'_2, l'_2} \times \dots \times a_{i_{2,2k''_2}, j'_2, l'_2} \times \dots \times a_{i_{M,1}, j'_M, l'_M} \times \dots \times a_{i_{M,2k''_M}, j'_M, l'_M}] \end{aligned}$$

If $i_{1,1}$ appears at least twice in the sums, that is there exists $2 \leq r \leq M$ and $1 \leq s \leq 2k''_r$ such that $i_{1,1} = i_{r,s}$, then, $a_{i_{1,1}, j_1, l_1} a_{i_{1,1}, j_r, l_r} = 0$ since $D_{j_1, l_1} \cap D_{j_r, l_r} = \emptyset$. Otherwise, if $i_{1,1}$ appears only once, by independence between the $(X_i, Y_i)_i$, we obtain that

$$\begin{aligned} & \mathbb{E}_{f_0} [a_{i_{1,1}, j_1, l_1} \times \dots \times a_{i_{1,2k_1}, j_1, l_1} \times a_{i_{2,1}, j'_2, l'_2} \times \dots \times a_{i_{M,2k''_M}, j'_M, l'_M}] \\ &= \mathbb{E}_{f_0} [a_{i_{1,1}, j_1, l_1}] \times \mathbb{E}_{f_0} [a_{i_{1,2}, j_1, l_1} \times \dots \times a_{i_{1,2k_1}, j_1, l_1} \times a_{i_{2,1}, j'_2, l'_2} \times \dots \times a_{i_{M,2k''_M}, j'_M, l'_M}] = 0 \end{aligned}$$

Hence,

$$\mathbb{E}_{f_0} [A_{k_1, j_1, l_1} \times A_{k''_2, j'_2, l'_2} \times \dots \times A_{k''_M, j'_M, l'_M}] = 0.$$

and thus, all the terms of the form (C.81) have a null expectation.

Let us now consider Equation (C.82) (where the index (j_1, l_1) appears twice).

$$\begin{aligned} & \mathbb{E}_{f_0} [A_{k_1, j_1, l_1} \times A_{k'_1, j_1, l_1} \times A_{k''_2, j'_2, l'_2} \times \dots \times A_{k''_M, j'_M, l'_M}] = \\ & \sum_{i_{1,1} < \dots < i_{1,2k_1}} \sum_{i'_{1,1} < \dots < i'_{1,2k'_1}} \sum_{i_{2,1} < \dots < i_{2,2k''_2}} \dots \sum_{i_{M,1} < \dots < i_{M,2k''_M}} \mathbb{E}_{f_0} [a_{i_{1,1}, j_1, l_1} \times \dots \times a_{i_{1,2k_1}, j_1, l_1} \times a_{i'_{1,1}, j_1, l_1} \times \\ & \quad \times \dots \times a_{i'_{1,2k'_1}, j_1, l_1} \times a_{i_{2,1}, j'_2, l'_2} \times \dots \times a_{i_{2,2k''_2}, j'_2, l'_2} \times \dots \times a_{i_{M,1}, j'_M, l'_M} \times \dots \times a_{i_{M,2k''_M}, j'_M, l'_M}] \end{aligned}$$

If there exists at least one index $i_{1, \cdot}$ or $i'_{1, \cdot}$ that can be isolated, then by independence,

$$\begin{aligned} & \mathbb{E}_{f_0} [a_{i_{1,1}, j_1, l_1} \times \dots \times a_{i_{1,2k_1}, j_1, l_1} \times a_{i'_{1,1}, j_1, l_1} \times \dots \times a_{i'_{1,2k'_1}, j_1, l_1} \times \\ & \quad \times a_{i_{2,1}, j'_2, l'_2} \times \dots \times a_{i_{2,2k''_2}, j'_2, l'_2} \times \dots \times a_{i_{M,1}, j'_M, l'_M} \times \dots \times a_{i_{M,2k''_M}, j'_M, l'_M}] = 0. \end{aligned}$$

Hence, the remaining terms are obtained for $k_1 = k'_1$ and $i_{1,s} = i'_{1,s}$ for all $1 \leq s \leq 2k_1$. These arguments are being valid for any index (j_r, l_r) , we obtain that $Q = 0$ and

$$\begin{aligned}
\mathbb{E}_{f_0} [B_{j_1, l_1} \dots B_{j_m, l_m}] &= \sum_{k_1, \dots, k_m=1}^{[n/2]} \sum_{i_{1,1} < \dots < i_{1,2k_1}} \dots \sum_{i_{m,1} < \dots < i_{m,2k_m}} \mathbb{E}_{f_0} \left[a_{i_{1,1}, j_1, l_1}^2 \times \dots \times a_{i_{1,2k_1}, j_1, l_1}^2 \times \dots \times \right. \\
&\quad \left. \times a_{i_{m,1}, j_m, l_m}^2 \times \dots \times a_{i_{m,2k_m}, j_m, l_m}^2 \right] \\
&= \sum_{k_1, \dots, k_m=1}^{[n/2]} \sum_{\substack{I_1, \dots, I_m \subset \{1, \dots, n\} \\ \text{Card}(I_r) = 2k_r}} \mathbb{E}_{f_0} \left[\prod_{i_1 \in I_1} a_{i_1, j_1, l_1}^2 \times \dots \times \prod_{i_m \in I_m} a_{i_m, j_m, l_m}^2 \right].
\end{aligned}$$

If the subsets I_r are not pairwise disjoint, the product $\prod_{i_1 \in I_1} a_{i_1, j_1, l_1}^2 \times \dots \times \prod_{i_m \in I_m} a_{i_m, j_m, l_m}^2 = 0$, since the supports $D_{(j_r, l_r)}$ are disjoint.

Thus,

$$\begin{aligned}
\mathbb{E}_{f_0} [B_{j_1, l_1} \dots B_{j_m, l_m}] &= \sum_{k_1, \dots, k_m=1}^{[n/2]} \sum_{\substack{I_1, \dots, I_m \subset \{1, \dots, n\} \\ I_r \cap I_s = \emptyset, \forall r \neq s \\ \text{Card}(I_r) = 2k_r}} \mathbb{E}_{f_0} \left[\prod_{i_1 \in I_1} a_{i_1, j_1, l_1}^2 \times \dots \times \prod_{i_m \in I_m} a_{i_m, j_m, l_m}^2 \right] \\
&= \sum_{k_1, \dots, k_m=1}^{[n/2]} \sum_{\substack{I_1, \dots, I_m \subset \{1, \dots, n\} \\ I_r \cap I_s = \emptyset, \forall r \neq s \\ \text{Card}(I_r) = 2k_r}} \prod_{i_1 \in I_1} \mathbb{E}_{f_0} [a_{i_1, j_1, l_1}^2] \times \dots \times \prod_{i_m \in I_m} \mathbb{E}_{f_0} [a_{i_m, j_m, l_m}^2],
\end{aligned}$$

by independence of the $(X_i, Y_i)_{1 \leq i \leq n}$.

Besides, for all $1 \leq i \leq n$,

$$\mathbb{E}_{f_0} [a_{i, j, l}^2] = C_0^2 h_n^{2\delta+2p+2q} \|g_{n, j, l}\|_2^2 = C(p, q, \delta, R, R', \eta) h_n^{2\delta+p+q},$$

since C_0 depends on p, q, δ, R, R' and η , and by Equation (C.70). Thus,

$$\begin{aligned}
\mathbb{E}_{f_0} [B_{j_1, l_1} \dots B_{j_m, l_m}] &= \sum_{k_1, \dots, k_m=1}^{[n/2]} \sum_{\substack{I_1, \dots, I_m \subset \{1, \dots, n\} \\ I_r \cap I_s = \emptyset, \forall r \neq s \\ \text{Card}(I_r) = 2k_r}} \left(C(p, q, \delta, R, R', \eta) h_n^{2\delta+p+q} \right)^{2k_1 + \dots + 2k_m} \\
&= \sum_{k_1, \dots, k_m=1}^{[n/2]} \binom{n}{2k_1, \dots, 2k_m, n - \sum_{r=1}^m 2k_r} \left(C(p, q, \delta, R, R', \eta) h_n^{2\delta+p+q} \right)^{2k_1 + \dots + 2k_m}.
\end{aligned}$$

Moreover, the multinomial coefficient can be upper bounded as follows

$$\binom{n}{2k_1, \dots, 2k_m, n - \sum_{r=1}^m 2k_r} \leq n^{2k_1 + \dots + 2k_m}.$$

Hence,

$$\begin{aligned}\mathbb{E}_{f_0}[B_{j_1, l_1} \dots B_{j_m, l_m}] &\leq \sum_{k_1=1}^{[n/2]} \dots \sum_{k_m=1}^{[n/2]} \left(C(p, q, \delta, R, R', \eta) n \times h_n^{2\delta+p+q} \right)^{2k_1 + \dots + 2k_m} \\ &= \left[\sum_{k=1}^{[n/2]} \left(C(p, q, \delta, R, R', \eta) n \times h_n^{2\delta+p+q} \right)^{2k} \right]^m.\end{aligned}$$

Furthermore, for h_n defined in (4.5) we have

$$h_n \leq C(p, q, \alpha, \beta, \delta, R, R', \eta) n^{-2/(4\delta+p+q)},$$

and thus, whatever the constant $C(p, q, \alpha, \beta, \delta, R, R', \eta)$ is,

$$C(p, q, \delta, R, R', \eta) n \times h_n^{2\delta+p+q} \leq C'(p, q, \alpha, \beta, \delta, R, R', \eta) n^{-(p+q)/(4\delta+p+q)} < 1/2$$

for n large enough. Thus, by property of geometric series, we get

$$\begin{aligned}\sum_{k=1}^{[n/2]} \left\{ \left[C(p, q, \delta, R, R', \eta) n \times h_n^{2\delta+p+q} \right]^2 \right\}^k &\leq \frac{\left[C(p, q, \delta, R, R', \eta) n \times h_n^{2\delta+p+q} \right]^2}{1 - \left[C(p, q, \delta, R, R', \eta) n \times h_n^{2\delta+p+q} \right]^2} \\ &\leq \frac{4}{3} \left[C(p, q, \delta, R, R', \eta) n \times h_n^{2\delta+p+q} \right]^2.\end{aligned}$$

We recall that the constants $C(\cdot)$ may vary from line to line. This being true for all (j, l) in $I_{n,p,q}$, from Equation (C.80), we deduce that

$$\begin{aligned}\mathbb{E}_{f_0} \left[\left\{ L_{\nu_{\rho_n^*}}(\mathbb{Z}_n) \right\}^2 \right] &\leq 1 + \sum_{m=1}^{M_n^{p+q}} \binom{M_n^{p+q}}{m} \left[C(p, q, \delta, R, R', \eta) n \times h_n^{2\delta+p+q} \right]^{2m} \\ &\leq 1 + \sum_{m=1}^{M_n^{p+q}} \left\{ M_n^{p+q} \left[C(p, q, \delta, R, R', \eta) n \times h_n^{2\delta+p+q} \right]^2 \right\}^m \\ &\leq 1 + \sum_{m=1}^{M_n^{p+q}} \left[C(p, q, \delta, R, R', \eta) n^2 \times h_n^{4\delta+p+q} \right]^m,\end{aligned}$$

since $\binom{M_n^{p+q}}{m} \leq [M_n^{p+q}]^m$ and $M_n h_n = 1$.

Finally, for h_n defined in (4.5), with

$$C(p, q, \alpha, \beta, \delta, R, R', \eta) = \left(\frac{1}{C(p, q, \delta, R, R', \eta)} \times \frac{4(1 - \alpha - \beta - \eta)^2}{1 + 4(1 - \alpha - \beta - \eta)^2} \right)^{1/(4\delta+p+q)},$$

we directly obtain that

$$C(p, q, \delta, R, R', \eta) n^2 \times h_n^{4\delta+p+q} \leq \frac{4(1 - \alpha - \beta - \eta)^2}{1 + 4(1 - \alpha - \beta - \eta)^2} < 1.$$

Hence, by property of the geometric series we obtain,

$$\begin{aligned}\mathbb{E}_{f_0} \left[\left(L_{\nu_{\rho_n^*}}(\mathbb{Z}_n) \right)^2 \right] &< 1 + \frac{[C(p, q, \delta, R, R', \eta) n^2 \times h_n^{4\delta+p+q}]}{1 - [C(p, q, \delta, R, R', \eta) n^2 \times h_n^{4\delta+p+q}]} \\ &< 1 + 4(1 - \alpha - \beta - \eta)^2,\end{aligned}$$

which ends the proof of Proposition 5.

References

- [Arcones and Gine, 1993] Arcones, M. A. and Gine, E. (1993). Limit theorems for U -processes. *The Annals of Probability*, pages 1494–1542.
- [Butucea, 2007] Butucea, C. (2007). Goodness-of-fit testing and quadratic functional estimation from indirect observations. *The Annals of Statistics*, 35(5):1907–1930.
- [De Lozzo and Marrel, 2017] De Lozzo, M. and Marrel, A. (2017). Sensitivity analysis with dependence and variance-based measures for spatio-temporal numerical simulators. *Stochastic environmental research and risk assessment*, 31(6):1437–1453.
- [Giné et al., 2000] Giné, E., Latała, R., and Zinn, J. (2000). Exponential and moment inequalities for U -statistics. In *High Dimensional Probability II*, pages 13–38. Springer.
- [Gretton et al., 2008] Gretton, A., Fukumizu, K., Teo, C. H., Song, L., Schölkopf, B., and Smola, A. J. (2008). A kernel statistical test of independence. In *Advances in Neural Information Processing Systems*, pages 585–592.
- [Hoeffding, 1963] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58:13–30.
- [Houdré and Reynaud-Bouret, 2003] Houdré, C. and Reynaud-Bouret, P. (2003). Exponential inequalities, with constants, for U -statistics of order two. In *Stochastic inequalities and applications*, pages 55–69. Springer.
- [Ingster, 1989] Ingster, Y. I. (1989). An asymptotically minimax test of the hypothesis of independence. *J. Soviet Math*, 44:466–476.
- [Ingster, 2000] Ingster, Y. I. (2000). Adaptive chi-square tests. *Journal of Mathematical Sciences*, 99(2):1110–1119.
- [Ishigami and Homma, 1990] Ishigami, T. and Homma, T. (1990). An importance quantification technique in uncertainty analysis for computer models. In *[1990] Proceedings. First International Symposium on Uncertainty Modeling and Analysis*, pages 398–403. IEEE.
- [Kim et al., 2020] Kim, I., Balakrishnan, S., and Wasserman, L. (2020). Minimax optimality of permutation tests. *arXiv preprint arXiv:2003.13208*.
- [Romano and Wolf, 2005] Romano, J. P. and Wolf, M. (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association*, 100(469):94–108.

- [Serfling, 2009] Serfling, R. J. (2009). *Approximation theorems of mathematical statistics*, volume 162. John Wiley & Sons.
- [Sobol, 2001] Sobol, I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and computers in simulation*, 55(1-3):271–280.
- [Spokoiny, 1996] Spokoiny, V. G. (1996). Adaptive hypothesis testing using wavelets. *The Annals of Statistics*, 24(6):2477–2498.
- [Tsybakov, 2009] Tsybakov, A. B. (2009). Introduction to nonparametric estimation. Revised and extended from the 2004 French original. Translated by Vladimir Zaiats.
- [Weiss, 2006] Weiss, N. A. (2006). *A course in probability*. Addison-Wesley.