Nonparametric homogeneity tests

Cristina Butucea
Modal'X, Université Paris X, Bât. G, F92001 Nanterre Cedex
LPMA, Université Paris VI, 175, rue du Chevaleret, F75013 Paris
E-mail: cristina.butucea@u-paris10.fr
Karine Tribouley
Modal'X, Université Paris X, Bât. G, F92001 Nanterre Cedex
URA CNRS 743, Université Paris XI, F91405 Orsay Cedex
E-mail: karine.tribouley@u-paris10.fr

December 11, 2003

Abstract

We test whether two independent samples of i.i.d. random variables $X_1, \dots, X_n$ and $Y_1, \dots, Y_m$ having common probability density f and, respectively, g are issued from the same population. The null hypothesis $f = g$ is opposed to a large nonparametric class of smooth alternatives f and g. We consider several problems, according to the distance between the populations' densities: pointwise, interval-wise, L_2 and L_∞ norms. We propose test procedures that attain parametric rates in some cases. In other problems, the procedures adapt automatically to the smoothnesses of the underlying densities. After a numerical study of these tests, we prove their theoretical properties in the classical minimax approach.

Keywords Nonparametric test, Homogeneity Test, Wavelet estimator, Minimax rates, Adaptivity.

1 Introduction

Let $X_1, \dots X_n$, n i.i.d. variables with density f, and $Y_1, \dots Y_m$, m i.i.d. variables with density g, be two independent samples. We study in this paper nonparametric tests for deciding whether the samples are issued from the same probability law. Thus, the null hypothesis is

$$H_0 : f = g. \tag{1}$$

The test problem is well posed when the alternative is given. In this paper, we consider a large nonparametric class $\Lambda_{n,m}$ consisting of couples of density functions f and g of some given regularity which are far enough from each other in terms of some distance (or semi-distance). The alternative writes

$$H_1 : (f, g) \in \Lambda_{n,m}(C) = \mathcal{R} \cap \mathcal{S}_{n,m}(C). \tag{2}$$

The space $\mathcal{R}$ is a class of regularity allowing to derive optimality properties. The space $\mathcal{S}_{n,m}(C)$ gives the geometry of the problem and is defined with some loss function for the difference of the two densities $l(f - g)$:

$$\mathcal{S}_{n,m}(C) = \{f, g \text{ densities}, \ |l(f - g)| \geq C r_{n,m}\},$$

for a constant $C > 0$ and a sequence $r_{n,m} > 0$ tending to 0 when $n \wedge m \to \infty$. This sequence measures the rate of separation between the test hypotheses H_0 and H_1. This space is essential for the test procedure because the test statistic is built with estimators $T(X_1, \ldots, X_n, Y_1, \ldots, Y_m)$ of $l(f-g)$ using the two samples. For some particular distances, we may use nonparametric estimators of both densities and plug them into the distance, that is use $T_j(d) = l(\hat{f}_{j,n} - \hat{g}_{j,m})$ where j is the smoothing parameter of the estimation method varying in J. Nevertheless, this is not always an optimal estimator since it can be highly biased, in the case of L_p distances, $1 \leq p \leq \infty$ for example. We describe here a family $\{T_j(d)\}_{j \in J}$ of wavelet estimators of the quantity $l(f-g)$ for particular losses l. For each smoothing parameter $j \in J$, we define the **test statistic** comparing the estimator $T_j(d)$ to a **critical value**, $t_{j,n,m} > 0$

$$D_j(d) = \begin{cases} 0 & \text{if} \quad |T_j(d)| \leq t_{j,n,m} \\ 1 & \text{if} \quad |T_j(d)| > t_{j,n,m}. \end{cases}$$

which means that we decide H_0 if $|T_j(d)| \leq t_{j,n,m}$ and H_1 otherwise. The smoothing parameter j and the critical value $t_{j,n,m}$ have to be chosen such that the test statistic $D_j(d)$ is the best among the family $\{D_j(d)\}_{j \in J}$. Roughly speaking, the best test statistic minimizes the separating rate $r_{n,m}$, at fixed probability errors of first type and second type. As usual in non parametric setting, the optimal parameter j and thus the optimal critical value depends on the space $\mathcal{R}$ via the indices of regularity of the densities f, g. In view of practical applications, we need data driven procedures. When these indices are unknown, we build an **adaptive** test statistic

$$D(d) = \max_{j \in J} D_j(d)$$

where J is a set of indices only depending on n and m. This consists to accept H_0 if, for every level $j \in J$, the decision given by $D_j(d)$ is to accept H_0.

Such tests are known in the minimax and adaptive literature for the one sample problem of goodness-of-fit tests. In goodness-of-fit tests we compare the given sample to a entirely known distribution. This problem was solved for different regularity classes (Hölder or Sobolev or Besov) associated with various geometries that we shall also consider for our problem: pointwise, quadratic and supremum norm. For fixed smoothness of the unknown density, i.e. minimax testing, there is a rich literature summed-up in Ingster [5] and in Ingster and Suslina [7]. Optimal test procedures include orthogonal projection, kernel estimates or χ^2 procedure. Goodness-of-fit tests with alternatives of variable smoothness, into some given interval, were introduced by Spokoiny [15], for L_2 distance, in the Gaussian white noise model and generalized by Spokoiny [16] to L_p distances. Ingster [6] proved that a collection of χ^2 tests attains the adaptive rates of goodness-of-fit tests in L_2 distance as well for a density model.

To our knowledge, nonparametric comparison of different samples is known only for the regression model. We refer to Munk and Dette [12] and references there in for testing that two or more regression functions are equal, in L_2 distance. Their method is based on an asymptotically normal estimator of the squared L_2 norm. The power of this test is approximated theoretically and evaluated empirically. Nonparametric methods for comparing two or more regression functions were introduced in Dette and Neumeyer [3]. Three test statistics were proposed and their asymptotic normality allowed evaluation of the test errors. No optimality properties were studied in the minimax approach. In the case of the density framework, the most famous comparison tests are based on the distribution functions: see by instance the Kolmogorov-Smirnov test and the Cramer-Von Mises test. Special tables are computed allowing to choose the critical values in view to keep the first type error bounded. Generally, the alternative is not explicitly given and results on the power of the tests can be given if the alternative is restricted to a small family of densities like the Gaussian family.

We present in this paper several results. Four loss functions $l(f-g)$ are considered to quantify the distance between the densities:

$$(f-g)(x_0), \quad \int_A (f-g), \quad \int (f-g)^2 \quad \text{and} \quad \sup_x |f-g|(x),$$

where x_0 is a given point and A is a given interval. In the sequel, we refer to these loss functions as point wise, interval wise, quadratic or supremum problems. Our test procedures are based on wavelet decompositions. We first give data-driven procedures which means that no a priori knowledge of unknown quantities is required. The critical values are chosen by bootstrap methods. We study the empirical qualities of our procedures concerning the errors of first and second type. These qualities are varying with the procedures. To sum up, if the choices of x_0 or A are lucky, the point wise and the interval wise procedures are excellent: the prescribed levels α are generally respected and the empirical powers are high. These tests are able to detect differences between densities which are not detectable by the classical tests build on the repartition functions (by instance, for oscillating densities). The procedure based on the L_∞ norm gives good results but they have to be improved using better estimation methods of the quantile of the test statistic. The study of a bootstrap procedure will be the subject of a further work. Our opinion on the quadratic procedure is not really positive. First, from a practical point of view, the usual wavelet algorithm can not be employed because the test statistic is a $U-$ statistic and therefore the procedure is quite untractable for large n, m. Next, the test is very conservative. In view to give explanations on these differences between procedures, we study each procedure from a theoretical point of view. We exhibit the optimal (for the rates) procedures in minimax approach in each setup. We give full proofs of how the procedures attain the testing rates and of the optimality of these rates. We stress that there is an important difference with goodness-of-fit tests where we can simply transform the sample via the distribution function under the null hypothesis and then fit a uniform density. Indeed, no density is available under H_0 and such a transform is unknown. Moreover, this implies that the test statistics are not free under H_0. Next, we study in a theoretical way the data-driven procedures. We prove that our procedures achieve the optimal rates up to an extra $\log\log$ term. Generally, adaptivity to smoothness implies a small loss in the minimax rates. We believe that these losses are also the least possible as the adaptive lower bounds seem to confirm. This will be the subject of further scientific communication.

The paper is organized as follows. In Section 2, we present our test procedures. In Section 3, we give empirical results based on experiments. We focus in Section 4 on theoretical results. Theorem 1 and Theorem 2 provide optimality results for the rates in minimax approach and Theorem 3 gives rates for the adaptive procedures. These results about optimality are proven in Sections 6 for the upper bounds and Section 7 for the lower bounds. Section 5 is devoted to a discussion on particular points: we compare the empirical results with the theoretical results, we comment the difference between the rates of estimation and the rates of test. In the proofs of our theorems, we use three results concerning the control of the bias terms, the asymptotic distributions of the test statistics and exponential inequalities for these statistics, see Section 6. The proofs of these results are postponed to Appendix A.

2 Test procedures

We propose different test procedures, each of them being associated with different separation spaces $\mathcal{S}_{n,m}$. We focus on four distances measuring how far apart the density functions are. We restrict ourselves to the usual distances in the non parametric setting: pointwise, interval-wise, L_2 and L_∞ distances. In each context, we suggest different estimators T_j of $l(f-g)$ based on wavelet expansions,

j being the tuning parameter of the method. Next, we derive families of tests statistics $\{D_j(d)\}_j$ as explained in the introduction.

Let ϕ and ψ be a scaling function and an associated wavelet function compactly supported. For any function h, we denote by $h_{j,k}(x)$ the function $2^{j/2}h(2^j x - k)$. For any j, k, the scaling and wavelet coefficients of the functions f and g are defined respectively by

$$\alpha_{j,k} = \int \phi_{j,k} f, \qquad \beta_{j,k} = \int \psi_{j,k} f, \qquad a_{j,k} = \int \phi_{j,k} g, \qquad b_{j,k} = \int \psi_{j,k} g$$

and the scaling coefficients are estimated by their empirical counterparts

$$\hat{\alpha}_{j,k} = \frac{1}{n}\sum_{i=1}^{n} \phi_{j,k}(X_i), \quad \hat{a}_{j,k} = \frac{1}{m}\sum_{i=1}^{m} \phi_{j,k}(Y_i).$$

We fix $j^* \geq 0$. Let x_0 be a given point. Motivated by the wavelet expansion

$$f(x_0) - g(x_0) \quad = \quad \sum_k (\alpha_{j^*k} - a_{j^*k})\, \phi_{j^*k}(x_0) + \sum_{j=j^*}^{\infty} \sum_k (\beta_{jk} - b_{jk})\, \psi_{jk}(x_0).$$

the low frequencies part is estimated by

$$T_{j^*}(x_0) \quad = \quad \sum_k (\hat{\alpha}_{j^*k} - \hat{a}_{j^*k}) \phi_{j^*k}(x_0). \tag{3}$$

The test statistic $D_{j^*}(x_0)$ based on $T_{j^*}(x_0)$ leads to consider **fixed point alternative** for which the separation space is

$$\mathcal{S}_{n,m}(C) = \{f, g \text{ densities, } |f(x_0) - g(x_0)| \geq C r_{n,m}\}.$$

Let A be a given interval. We get

$$\int_A f - \int_A g \quad = \quad \sum_k (\alpha_{j^*k} - a_{j^*k}) \int_A \phi_{j^*k} + \sum_{j=j^*}^{\infty} \sum_k (\beta_{jk} - b_{jk}) \int_A \psi_{jk},$$

and we propose the test statistic $D_{j^*}(A)$ based on the estimator

$$T_{j^*}(A) \quad = \quad \sum_k (\hat{\alpha}_{j^*k} - \hat{a}_{j^*k}) \int_A \phi_{j^*k} \tag{4}$$

which leads to consider **interval-wise alternative** for which the separation space is

$$\mathcal{S}_{n,m}(C) = \{f, g \text{ densities, } |\int_A f - \int_A g| \geq C r_{n,m}\}.$$

Since the wavelet basis is orthonormal, the expansion holds

$$\|f - g\|_2^2 \quad = \quad \sum_k (\alpha_{j^*k} - a_{j^*k})^2 + \sum_{j=j^*}^{\infty} \sum_k (\beta_{jk} - b_{jk})^2\,,$$

leading to the following estimator

$$T_{j^*}(L_2) \quad = \quad \frac{1}{(n \wedge m)((n \wedge m) - 1)} \sum_{i_1=1}^{n\wedge m} \sum_{i_2=1}^{n \wedge m} U_{i_1 i_2} \tag{5}$$

where

$$U_{i_1 i_2} = \sum_k (\phi_{j^*k}(X_{i_1}) - \phi_{j^*k}(Y_{i_1}))\, (\phi_{j^*k}(X_{i_2}) - \phi_{j^*k}(Y_{i_2}))\, 1_{\{i_1 \neq i_2\}}.$$

We consider the test statistic $D_{j^*}(L_2)$ based on $T_{j^*}(L_2)$; the **quadratic alternative** is associated with the separation space

$$\mathcal{S}_{n,m}(C) = \{f, g \text{ densities}, \ \|f-g\|_2^2 \geq C r_{n,m}\}.$$

Motivated by the above expansion on the wavelet basis

$$\begin{aligned}
\|f-g\|_\infty &= \sup_x \left| \sum_k (\alpha_{j^*k} - a_{j^*k}) \, \phi_{j^*k}(x) + \sum_{j=j^*}^{\infty} \sum_k (\beta_{jk} - b_{jk}) \, \psi_{jk}(x) \right| \\
&\leq 2^{j^*/2} \sup_k |\alpha_{j^*k} - a_{j^*k}| \, \|\phi\|_\infty + \sup_x \left| \sum_{j=j^*}^{\infty} \sum_k (\beta_{jk} - b_{jk}) \, \psi_{jk}(x) \right|,
\end{aligned}$$

we consider the over estimator defined by

$$T_{j^*}(L_\infty) = \sup_k |T_{j^*k}| \tag{6}$$

for

$$T_{j^*k} = \hat{\alpha}_{j^*k} - \hat{a}_{j^*k}. \tag{7}$$

The test statistic $D_{j^*}(L_\infty)$ is based on $T_{j^*}(L_\infty)$; the **supremum alternative** is associated with the separation space

$$\mathcal{S}_{n,m}(C) = \{f, g \text{ densities}, \ \|f-g\|_\infty \geq C r_{n,m}\}.$$

Note that we can easily generalize the results to an alternative

$$\mathcal{S}_{n,m}(C) = \{f, g \text{ densities}, \ \left| \int \Phi(f-g) \right| \geq C r_{n,m}\},$$

based on a smooth functional Φ (at least 4 times continuously differentiable). Indeed, it is enough to develop this functional around fixed $\hat{f} - \hat{g}$ at order 3 ($\hat{f}$ and $\hat{g}$ are preliminary estimators) and then estimate successively the norms up to the order 3 (see Kerkyacharian and Picard [9] and Tribouley [18]).

We study in the next section the experimental qualities of the four families of tests $\{D_{j^*}, j^* \geq 0\}$ built with the estimators $T_{j^*}(x_0)$, $T_{j^*}(A)$, $T_{j^*}(L_2)$ and $T_{j^*}(L_\infty)$.

3 Numerical study

3.1 Presentation

This part is a joint work with Y. Misiti [1]. We test the four adaptive procedures described in Section 2 with $n = m = 500$ and using $DB3$ for the pointwise alternatives, quadratic alternatives and supremum alternatives. For the interval-wise alternative, the Haar basis is used. We fix the probability error of first type: $\alpha = [10\%, 5\%, 1\%]$ and we compute the empirical probability to choose the alternative. The empirical mean is computed with 100 repetitions of the algorithm.

Let us present the procedure. We denote $N = \left(n^{-1} + m^{-1}\right)^{-1}$ and we choose the set $J = \{j_0, \ldots, j_\infty\}$ of tuning parameters as follows

$$2^{j_0} = 1, \quad 2^{j_\infty} = N \text{ or } 2^{j_\infty} = N^2 \text{ in the quadratic problem.}$$

[1] URA CNRS 743, Université Paris XI

For each tuning parameter j varying in J, we compute the estimator $T_j(d)$ and compare with a critical value $t_{\alpha,j}$. If $|T_j(d)| > t_{\alpha,j}$ for at least one level j, we accept H_1. Otherwise, we accept H_0. The difficulty is to choose the critical value. We decide to use the normal approximation in the case of the point-wise, interval-wise, quadratic alternatives and we estimate the variance σ_j^2 of the statistic $|T_j(d)|$ using bootstrap methods. We take then $t_{\alpha,j} = q_j \sqrt{\hat{\sigma}_j^2}$ where q_j is the $(\alpha/\#J)$-quantile of the standard normal distribution and $\hat{\sigma}_j^2$ is the bootstrap estimator of σ_j^2. In the case of the supremum alternative, the normal approximation fails and we directly estimate by bootstrap method the $(\alpha/\#J)$-quantile. The usual method fails and we use over resampling with $b_q = q\log(q)^{-1}, q = n, m$. For the estimation of the variance, the size of the resampling is $B = 100$ and for the estimation of the quantile, $B = 400$.

We compare our test procedures with the Kolmogorov-Smirnov test and with the Cramer von Mises test with the same n, m. See the lines KS and CM in the tables.

3.2 Empirical results

First, we deal with standard unimodal densities: G denotes the Normal density function, C is the Cauchy density, t_5 is the Student with 5 freedom degrees and E is the Laplace density (symetrized exponential density). These densities are translated by 0.5 to avoid problems with the wavelet function in 0 (0 is the extremity of the support of the mother of the wavelet function). For the procedure associated with a pointwise alternative, we consider three specific points x_0: $x_0 = 0.5$ is the mode of the densities; $x_0 = 2$ is an extremal quantile and $x_0 = 1.5$ is a generic point. For the interval wise problem we consider a 'modal' interval $A = [0, 1]$ and a 'tailed' interval $A = [-4, -2]$. The empirical powers are given in Table 5 and the empirical first type errors are given in Table 6. In the problem of testing f against g, $\hat{\eta}$ is the empirical power of the test, $\hat{\alpha}_1$ (respectively $\hat{\alpha}_2$) is the empirical level of the test f against f (respectively g against g).

The results concerning the pointwise procedure are relatively good: the empirical power is generally high while the empirical level is a good estimator of the prescribed level α. Obviously, the results are depending on the point x_0 at which we evaluate the test statistic. The graphs of the densities above those tables point out that the studied densities have significant differences which are localized at the mode and at the tails and no difference at the generic point $x_0 = 1.5$. Hence, it is expected that the test built with the generic point is less powerful than the test build with the modal point. Any way, remark that the power is large. By instance, in the problem of testing the Cauchy against the Laplace, we obtain, for a prescribed level $\alpha = 0.10$

$$\hat{\alpha}_1 = 0.13, \hat{\alpha}_2 = 0.14 \quad \text{and} \quad \hat{\eta} = 0.91.$$

Observe that, for the same test problem, the results obtained with procedures based on the repartition function are poor.

$$\hat{\alpha}_1 = 0.01, \hat{\alpha}_2 = 0.14 \quad \text{and} \quad \hat{\eta} = 0.49$$

for the KS test. The procedure gives excellent powers when x_0 is a mode. But, we observe that, in this case, the error of the first type is over estimated. For instance, when testing the Laplace against the Gaussian, we get

$$\hat{\alpha}_1 = 0.16, \hat{\alpha}_2 = 0.13 \quad \text{and} \quad \hat{\eta} = 1.00$$

The pointwise test is disappointing when x_0 is a tail point. For the comparison between the Gaussian and the Laplace, we have

$$\hat{\alpha}_1 = 0.06, \hat{\alpha}_2 = 0.06 \quad \text{and} \quad \hat{\eta} = 0.42.$$

It is probably due to the fact that the number of data is too small (because the estimator is localized on the tails) and the test statistic itself, and even more significantly, the variance estimator, are not accurate enough estimators in this case.

The results concerning the interval wise procedure are excellent. This procedure allows to use more data than in the previous procedure: this explains the results on the test built with the tail interval. In the previous example, we improved a lot the empirical power with no degradation of the empirical error of first type

$$\hat{\alpha}_1 = 0.05, \hat{\alpha}_2 = 0.13 \quad \text{and} \quad \hat{\eta} = 1.00.$$

The power obtained when we compare the Gaussian density with the Student t_5 density is good: in fact, this is the only procedure which detects a difference between these two densities, and the empirical error of first type is almost equal to the prescribed theoretical error as well

$$\hat{\alpha}_1 = 0.09, \hat{\alpha}_2 = 0.13 \quad \text{and} \quad \hat{\eta} = 0.64.$$

Observe that even the CM test does not succeed to separate these densities

$$\hat{\alpha}_1 = 0.09, \hat{\alpha}_2 = 0.14 \quad \text{and} \quad \hat{\eta} = 0.22.$$

The results obtained with the modal interval are also very good. In the same way than for the pointwise procedure, the estimated power depends on the interval A (chosen by the user). The advantage of this procedure is that it is easier to choose a tail interval or a modal interval than a mode point (with a rough estimation of the densities).

Observing the graphs of the densities, it is expected that the test Laplace/Cauchy and the test Gaussian/Cauchy are the best suited for the procedure using the supremum distance. The results in both cases are excellent. We give the results when comparing the Laplace and the Cauchy densities:

$$\hat{\alpha}_1 = 0.04, \hat{\alpha}_2 = 0.09 \quad \text{and} \quad \hat{\eta} = 0.94.$$

We think that this procedure gives in some cases less satisfying results because of the difficulty to estimate correctly the quantiles of the test statistic $T_j = \sup_k |T_{jk}|$.

At last, the procedure associated with the quadratic alternative gives bad results. From an algorithmic point of view, this method has to be rejected for two reasons. First, since the test statistic is based on $U-$statistics, we can not apply Mallat's algorithm to compute the statistics and the practical interest of the wavelets methods is precisely the use of this fast algorithm. Secondly, the number of levels we have to take into account is considerable. For instance, for our simulations, all the procedures (except the quadratic one) use $\#J = 7$ levels while the quadratic procedure needs twice as many levels. We do not use all the levels: we stop at $j = 10$. To consider so many levels implies that the test is very conservative: the decision is always H_0. For instance, when testing the Laplace against the Cauchy, the results are similar to those concerning the KS test or the CM test (except for $\hat{\alpha}_2$)

$$\hat{\alpha}_1 = 0.00, \hat{\alpha}_2 = 0.00 \quad \text{and} \quad \hat{\eta} = 0.41.$$

The test concerning the Gaussian and the Cauchy gives the worse results of our study

$$\hat{\alpha}_1 = 0.09, \hat{\alpha}_2 = 0.00 \quad \text{and} \quad \hat{\eta} = 0.45.$$

Since the empirical level is always zero although it is not required, we decide to modify the procedure. We stop at $j = 6$ (hence $\#J = 7$) and we compute the critical value using the α quantile of the standard normal distribution instead of the $\alpha/7$ quantile. The results are given in the last lines of the tables (refered as L_2(bis)). We remark that the empirical powers are larger while the estimations of the prescribed levels are better. In the case of the test between the Gaussian and the Cauchy, we obtain

$$\hat{\alpha}_1 = 0.11, \hat{\alpha}_2 = 0.10 \quad \text{and} \quad \hat{\eta} = 0.99$$

which is excellent.

Now, we consider mixtures of densities: G_3 is a mixture of Gaussian variables with small variance and E_3 is a mixture of Laplace variables. We choose these particular hypotheses so that the KS test and the CM test fail, see Table 7. The pointwise and the interval wise procedures give excellent results. The empirical levels (except for $x_0 = 1$) are generally good estimators of the prescribed level. The empirical powers can be high. By instance, for $x_0 = 2.5$ and $\alpha = 0.1$, we get

$$\hat{\alpha}_1 = 0.07, \hat{\alpha}_2 = 0.08 \quad \text{and} \quad \hat{\eta} = 0.52.$$

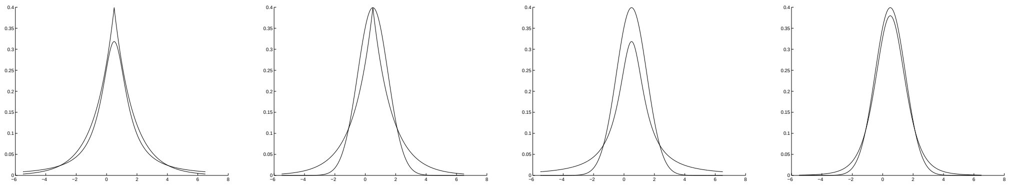
The tests associated with the supremum distance and to the quadratic distance are very conservative and therefore they are not more powerful than the KS test or the CM test. The second method for the quadratic problem improves a lot the results

$$\hat{\alpha}_1 = 0.01, \hat{\alpha}_2 = 0.02 \quad \text{and} \quad \hat{\eta} = 0.23.$$

Note that the prescribed level is under estimated.

We decide to explore the differences between the quadratic procedure (with no modifications) and the supremum problem in very simple situations. Since the considered densities are regular, we decide to restrict ourselves to a smaller number of indices j: we take $\#J = 6$. First, we test the Gaussian $\mathcal{N}(0,1)$ against the Gaussian $\mathcal{N}(m,1)$ for m varying between 0 and 1. Next, we test the $\mathcal{N}(0,1)$ against the Gaussian $\mathcal{N}(0,\sigma^2)$ for σ^2 varying between 1 and 1.4. Obviously, the usual tests (Kolmogorov-Smirnov and Cramer von Mises) give excellent results because the first test (respectively the second) consists in testing a location parameter (respectively a scaling parameter). We do not give the results for these tests because our aim is to compare the behaviour of both procedures (the quadratic procedure and the supremum procedure) The empirical powers are given in Fig.1 and Fig.2. We observe again that the quadratic procedure is very conservative: the prescribed level is $\alpha = 0.1$ estimated by $\hat{\alpha} = 0$. In the translation problem, the quadratic procedure is the best while the supremum procedure gives better results in the scaling problem.

3.3 Simulation results

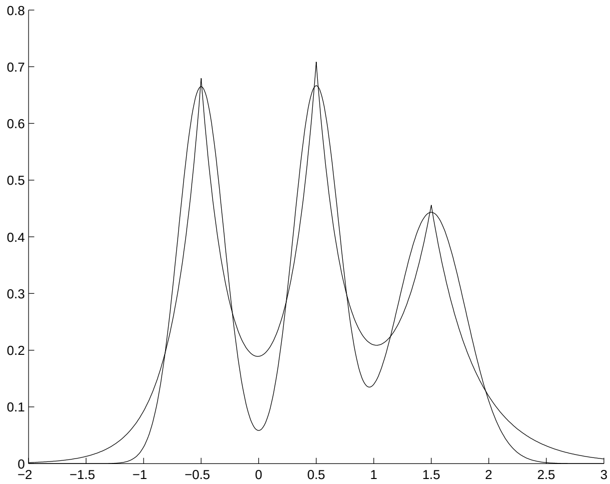


	E	vs	C	E	vs	G	G	vs	C	G	vs	t_5
α	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
KS	49	16	03	99	89	52	100	100	100	17	08	01
CM	54	24	02	98	89	41	100	100	100	22	09	03
$x = 0.5$	98	97	94	100	100	100	100	100	100	35	27	10
$x = 1.5$	91	81	63	86	67	51	100	98	86	23	11	01
$x = -2$	11	05	02	69	33	17	42	33	17	15	08	01
$A = [0, 1]$	100	100	99	100	100	100	84	75	55	10	08	02
$A = [-4, -2]$	13	09	03	100	100	99	100	100	100	64	54	27
L_∞	94	89	75	73	55	29	100	100	100	11	04	03
L_2	41	28	06	57	35	03	45	27	03	01	01	00
L_2(bis)	99	94	67	97	88	37	98	89	42	11	02	00

Table 5: Empirical power (in %) associated with tests of prescribed levels $\alpha = 10\%, 5\%, 1\%$ for testing $f = g$ when actually f is the Laplace against g the Cauchy, the Laplace against the Gaussian, the Cauchy against the Gaussian and the Student t_5 against the Gaussian, respectively.

	E	vs	E	C	vs	C	G	vs	G	t_5	vs	t_5
α	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
KS	01	00	00	14	09	02	14	06	02	05	02	00
CM	00	00	00	11	04	00	09	06	00	14	07	03
$x = 0.5$	12	03	01	16	11	01	13	09	01	21	13	01
$x = 1.5$	13	04	01	14	06	01	15	08	03	13	07	01
$x = -2$	09	06	01	06	01	00	06	04	01	06	01	00
$A = [0, 1]$	06	02	00	08	04	01	06	05	03	07	06	02
$A = [-4, -2]$	04	02	00	05	03	01	13	02	01	09	04	01
L_∞	04	03	00	09	05	02	14	12	05	03	03	02
L_2	00	00	00	00	00	00	01	00	00	00	00	00
L_2(bis)	05	02	00	11	03	01	10	01	00	05	02	00

Table 6: Empirical error of first type associated with tests of prescribed levels $\alpha = 10\%, 5\%, 1\%$ for testing $f = g$ when indeed $f = g$ is the Laplace, the Cauchy, the Gaussian and the Student t_5, respectively.



	G_3	vs	E_3	G_3	vs	G_3	E_3	vs	E_3
α	10%	5%	1%	10%	5%	1%	10%	5%	1%
KS	00	00	00	02	00	00	00	00	00
CM	00	00	00	02	00	00	00	00	00
$x = 1.0$	87	74	48	11	02	00	21	08	03
$x = -0.5$	35	27	08	06	03	01	08	04	00
$x = 0.0$	26	17	08	10	05	01	12	09	01
$x = 2.5$	52	38	14	07	02	01	08	05	01
$A = [-2, -1]$	97	91	69	10	05	01	00	00	00
$A = [-0.2, 0.2]$	84	75	47	06	01	00	07	03	01
$A = [0.8, 1.2]$	19	12	05	04	02	01	04	02	00
L_∞	08	07	05	00	00	00	00	00	00
L_2	04	00	00	00	00	00	00	00	00
L_2(bis)	23	14	05	01	00	00	02	00	00

Table 7: The first column gives the empirical power (in %) associated with tests of prescribed levels $\alpha = 10\%, 5\%, 1\%$ for testing the Gaussian mixture against the Laplace mixture; the other columns give the empirical error of the first type (in %) associated with tests of prescribed levels $\alpha = 10\%, 5\%, 1\%$ for testing the Gaussian mixture against the Gaussian mixture and the Laplace mixture against the Laplace mixture.

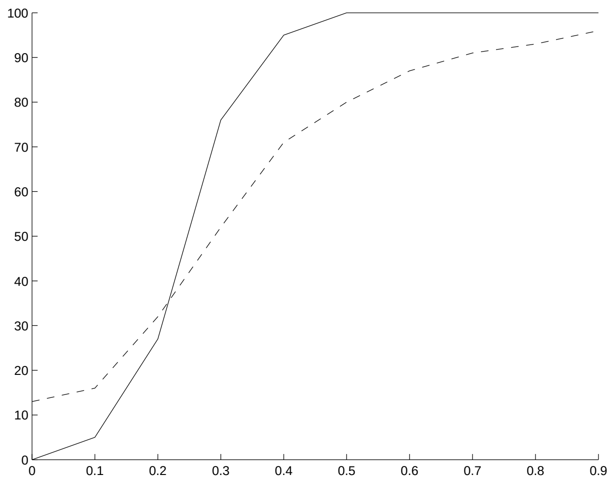


Fig 3: Comparison of the $\mathcal{N}(0,1)$ with the $\mathcal{N}(m,1), m = 0, 0.1, \ldots 1$. The powers (in %) of the tests are given: the quadratic alternative with the solid line, the supremum alternative with the dashed line. The prescribed level is $\alpha = 10\%$.

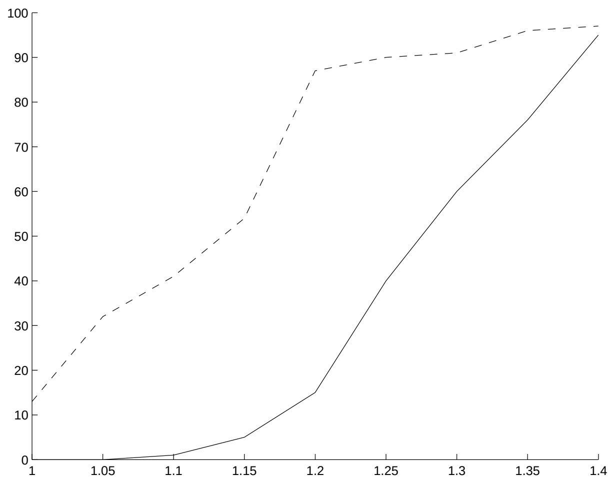


Fig 4: Comparison of the $\mathcal{N}(0,1)$ with the $\mathcal{N}(0,\sigma^2), \sigma^2 = 1, 1.05, \ldots 1.4$. The powers (in %) of the tests are given: the quadratic alternative with the solid line, the supremum alternative with the dashed line. The prescribed level is $\alpha = 10\%$.

3.4 Conclusion

The procedures associated with the pointwise distance and the interval wise distance are excellent and give good results even when the usual KS test fails. For the pointwise procedure the point x_0 has to be chosen carefully. The interval wise procedure answers partially to this constraint. Moreover, it improves the performances of the pointwise procedure on the tails of the densities. However, the interval A has to be short to generalize the point wise procedure. The procedure associated with the supremum alternative is interesting and could be improved by using a more accurate estimation of the quantiles. It can detect differences which are not seen by the L_2 type procedures (our quadratic procedure and the KS test). The quadratic procedure gives reasonably good results, if we decide to modify the procedure (i.e. to use only the first levels j and to take the α quantiles instead of the

$\alpha/\#J$ quantiles). But we we are still sceptical about the practical interest of this procedure: it does not use the advantages of the wavelet methods and the computation times are very long.

To conclude, our prescriptions are the following: if one has an a priori idea on the points where the densities are different, one chooses the point wise procedure; if one thinks that the tails are not similar, one takes the interval wise procedure; if one knows that on a small interval the densities are very far apart (like in the example of the scaling parameter for Gaussian densities), one uses the supremum procedure. If the quantities $f(x)$ and $g(x)$ are almost equally distant from each other when x walks along the support (like in the example of the shifted parameter for Gaussian densities or in the case of the mixtures), one takes the quadratic procedure.

In the next section, we study the theoretical properties of the procedures in order to explain these contrasting results.

4 Theoretical results

In the previous section, we focused on the probability errors of the test procedures. In this section, the point of view is quite different. Given the sum of probability errors, say γ, we study the separation rate $r_{n,m}$ between the null hypothesis and the alternative. Roughly speaking, we want to answer the question: " How far must be f from g to be able to detect a difference between them both?". First, we define the optimality criterion for the separation rate. Next, we give the regularity assumptions on the densities f et g. This allows us to define entirely the alternative of the test problem giving the space $\mathcal{R}$. For each test problem

$$H_0 : f = g \quad \text{against} \quad H_1 : (f, g) \in \Lambda_{n,m}(C) = \mathcal{R} \cap \mathcal{S}_{n,m}(C) \tag{8}$$

we give the best rate $r_{n,m}$ separating H_0 and H_1. The optimal choice of the tuning parameter j^* of our method and the critical value $t_{j^*,n,m}$ are given and this allows us to construct explicitly the optimal test procedure D_{j^*}. Finally, we consider the adaptive procedures $D = \max_{j \in J} D_j$ and we study their rates.

4.1 Definition of the optimality criterion

Let $0 < \gamma < 1$. A sequence $r_{n,m}$ is a **minimax rate of testing** for the problem (8), at level γ, if both statements are satisfied:

1. there exists a constant $C^* > 0$ and a test statistic $D^* = D_{j^*}$, called **rate optimal**, such that

$$\limsup_{n \wedge m \to \infty} \left(P_0 \left[D = 1 \right] + \sup_{f,g \in \Lambda_{n,m}(C)} P_{f,g} \left[D = 0 \right] \right) \leq \gamma \tag{9}$$

for all $C > C^*$;

2. there exists a constant $C_* > 0$ such that

$$\liminf_{n \wedge m \to \infty} \inf_{V} \left(P_0 \left[V = 1 \right] + \sup_{f,g \in \Lambda_{n,m}(C)} P_{f,g} \left[V = 0 \right] \right) \geq \gamma \tag{10}$$

for all $C < C_*$ where the infimum is taken over all test statistics V.

4.2 Assumptions

We complete the definition of each considered alternative giving the smoothness classes where the density functions belong to. Remark that under the alternative, densities may have different smoothnesses. Obviously, the spaces $\mathcal{R}$ and $\mathcal{S}_{n,m}(C)$ are related because the distance d measuring how far apart the density functions are has an effect on the required regularity.

Roughly speaking, we assume that the densities belong to a ball of either a Sobolev space or a Hölder space. In order to unify the notation and because we deal with wavelet methods, we express the regularity assumptions in terms of Besov spaces. For more details about the Besov spaces, see e.g. Triebel [18]; a characterization in terms of wavelet coefficients is given later on, in Relation (12). For the interval wise problem, we consider the Haar basis and put $DB = 1$. In all other problems, we assume that the scaling function and the wavelet function are compactly supported on $[0, 2DB - 1]$ for DB large enough and that the q-th moment of the wavelet ψ vanishes for $q = 0, \dots, DB$. We assume also that there exists a point x such that $\phi(x) \neq 0$ and $\psi(x) \neq 0$. By instance $x = 1$ for $DB3$, see the Daubechies's wavelets (Daubechies, [1]).

Let $0 < s_f, s_g \leq DB$ and $R > 0$. For $p \geq 1$, we define the collection of smoothness classes:

$$\mathcal{R}(p) \quad = \quad \left\{ f \in B^{s_f}_{p\infty}(R), g \in B^{s_g}_{p\infty}(R) \right\}$$

and we consider $p = 2$ for the quadratic alternative, respectively, $p = \infty$ for the pointwise, interval-wise and supremum alternatives. Moreover, for the L_∞ problem, we suppose that the densities have bounded (by $L > 0$) supports. In the L_2 problem, we add the assumption that the densities f, g are uniformly bounded.

4.3 Main results

Theorem 1 *Let $0 < \gamma < 1$ be the prescribed risk of the test. Denote by*

$$s = s_f \wedge s_g, \quad N = \left(\frac{1}{n} + \frac{1}{m} \right)^{-1}.$$

Then, for j^*, the corresponding estimator T_{j^*} defined in (3), (4), (5) or (6), and the critical value $t_{j^*,n,m}$ provide a test statistic

$$D_{j^*}(d) = I_{\{|T_{j^*}| > c\, t_{j^*,n,m}\}}$$

which achieves the rate $r^*_{n,m}$ of testing in each setup, respectively.

The quantities $j^*, t_{j^*,n,m}, r^*_{n,m}$ are described in Table 1, the constant c is depending on $\|\phi\|_\infty$, $\|f\|_\infty$ or $\|f\|_2$.

l	2^{j^*}	$t_{j^*,n,m}$	$r^*_{n,m}$	Restrictions
x_0	$N^{\frac{1}{2s+1}}$	$\left(\frac{2^{j^*}}{N}\right)^{\frac{1}{2}} \gamma^{-\frac{1}{2}}$	$N^{-\frac{s}{2s+1}}$	f, g bounded away from $[x_0 - \frac{2DB-1}{2^{j^*}}, x_0 + \frac{2DB-1}{2^{j^*}}]$
A	$N^{\frac{1}{s \wedge 1}}$	$N^{-\frac{1}{2}} \gamma^{-\frac{1}{2}}$	$N^{-\frac{1}{2}}$	$\lvert A\rvert > 2^{-j^*}$ $A \subset (\mathrm{supp}(f) \cap \mathrm{supp}(g))$
L_2	$N^{\frac{2}{4s+1}}$	$\left(\frac{2^{j^*}}{N^2}\right)^{\frac{1}{2}} \gamma^{-\frac{1}{2}}$	$N^{-\frac{4s}{4s+1}}$	
L_∞	$\left(\frac{N}{\log N}\right)^{\frac{1}{2s+1}}$	$\left(\frac{j^* + \log_2(L/\gamma)}{N}\right)^{\frac{1}{2}}$	$\left(\frac{N}{\log N}\right)^{-\frac{s}{2s+1}}$	

Table 1: Optimal parameters, critical values, minimax rates and restrictions

In other words, Theorem 1 ensures that (9) is satisfied for $D_{j^*}(d)$ and $r^*_{n,m}$. We have now to prove that (10) is satisfied by the rates $r^*_{n,m}$ in each setup, respectively. This is done in the following Theorem.

Theorem 2 *Let γ be the prescribed risk of test.*

In the pointwise problem, we consider a point x_0 such that f and g are bounded from below on $I_0 = [x_0 - \frac{2DB-1}{\log(N)}, x_0 + \frac{2DB-1}{\log(N)}]$. In the interval wise problem, we consider an interval A nested in supp$(f)\cap$supp(g) and such that $|A| > N^{-1/2}$.

Then, the rates $r^*_{n,m}$ given in Table 1 are minimax rates of testing.

This means that no other test can achieve faster separating rates than our testing procedures and therefore we have the corollary

Corollary 1 *The procedures described in Theorem 1 are optimal.*

Unfortunately, the procedures $D_{j^*}(d)$ depend on the extra parameter $s = s_f \wedge s_g$ which is generally unknown. Remark that the critical value depends also on unknown quantities. We propose data driven procedures in the sense that the test statistic is data driven. We have seen in the simulation part that the critical value has to be estimated by other methods (bootstrap, ...) Anyway, it does not affect the rates of testing. The rates of these procedures are stated in the following theorem.

Theorem 3 *Let $0 < \gamma < 1$ be the prescribed risk of the test. We consider the set of indices $J = \{j_0, \dots j_\infty\}$ where*

$$2^{j_0} = \log N, 2^{j_\infty} = \begin{cases} N/\log N, & \\ N^2/\log^3 N, & \text{in the quadratic problem.} \end{cases} \tag{11}$$

For $\tilde{t}_{j,n,m}$ given in Table 2 and $c > 0$ constant depending on $\|\phi\|_\infty, \|f\|_\infty$ or $\|f\|_2$, the test statistic

$$D(d) = I_{\{\max_{j\in J}\left(|T_j| - c\,\tilde{t}_{j,n,m}\right) > 0\}}$$

achieves the rate $\tilde{r}_{n,m}$.

l	$\tilde{t}_{j,n,m}$	$\tilde{r}_{n,m}$	restrictions
x_0	$\left(\frac{2^j \log\log N}{N}\right)^{\frac{1}{2}}$	$\left(\frac{N}{\sqrt{\log\log N}}\right)^{-\frac{s}{2s+1}}$	f, g bounded away from 0 on $[x_0 - \frac{2DB-1}{\log(N)}, x_0 + \frac{2DB-1}{\log(N)}]$
A	$\left(\frac{N}{\log\log N}\right)^{-\frac{1}{2}}$	$\left(\frac{N}{\sqrt{\log\log N}}\right)^{-\frac{1}{2}}$	$A \subset \text{supp}(f) \cap \text{supp}(g)$ $\lvert A\rvert > N^{-1/2}$
L_2	$\left(\frac{2^j\sqrt{\log\log N}}{N^2}\right)^{\frac{1}{2}}$	$\left(\frac{N}{\sqrt{\log\log N}}\right)^{-\frac{4s}{4s+1}}$	
L_∞	$\left(\frac{j+\log_2(L\log N)}{N}\right)^{\frac{1}{2}}$	$\left(\frac{N}{\log N}\right)^{-\frac{s}{2s+1}}$	

Table 2: Critical values, adaptive rates and restrictions

Such test procedures are commonly known in the minimax literature as adaptive to the smoothness: the test statistic $D(d)$ does not depend on the indices of regularity unlike the optimal test $D_{j^*}(d)$. There is usually a small loss in the rate due to generality of the class where the unknown functions belong.

Corollary 2 *The adaptive procedures studied in Section 3 are nearly optimal. In particular, for the L_∞ setup the procedure is adaptive to the smoothness and rate optimal, i.e. without any loss in rate.*

5 Discussion

We want to stress again the fact that the applied approach in our simulation study and the theoretical approach giving minimax and adaptive rates of testing are substantially different. In the first approach, we fix densities we want to compare and estimate empirically the probability errors of our testing procedure. In the theoretical part, we study our estimation procedure at fixed sum of errors, γ, and over very large, non parametric, classes of possible densities to test on. The rate of testing is expressed in terms of increasing sample sizes n and m. These rates are interpreted as the minimal separation distances between the null hypothesis and the alternative so that the test can still be performed.

The testing rates are evaluated asymptotically. Therefore, the constants associated with these rates and/or procedures are rather large in the exponential inequalities given next, in Section 6. This is especially the case for the L_2 problem, possibly explaining the bad simulation results in this case. Nevertheless, it is impossible to use theoretical critical values as they are issued from our proofs.

The theoretical results show the empirical results in a different light. Simulation results are quite good for interval-wise problem, where almost parametric rates of convergence are expected in theory. Pointwise and supremum results are quite encouraging as well. We think that the L_2 procedure fails because the behaviour of the test statistic is different under H_0 and under H_1 (see Lemma 2) which is not the case for the other procedures. We show in the proofs that the test statistic $T_j(L_2)$ is degenerate under H_0 and non degenerate under H_1. It could explain the fact that the test is so conservative.

We conclude with a comment on the difference between the test problem and the estimation problem. We note that the test statistic is an estimator of the loss function $l(f-g)$. A priori, the testing rate seems related to the estimation rate of $f-g$ with loss function l. This is indeed the case for pointwise and L_∞ problems, where we test at the same rate as we estimate the function $(f-g)$. This is not the case for the L_2 problem, where testing is easier than estimating $(f-g)$. Indeed, the testing rate $N^{-4s/(4s+1)}$ is much faster that the estimating rate $N^{-2s/(2s+1)}$. But, we remark that the testing rate is the same as the rate for estimating $\|f-g\|_2^2$ (on the non parametric side). In adaptive estimation of such functionals a loss in the rate is unavoidable due to the generality of the class where the unknown functions belong. The loss is known to be of order $\log N$. In the goodness-of-fit problems, Spokoiny [15] showed first that the loss in adaptive testing is of order $\sqrt{\log\log N}$ and it is much smaller than for the adaptive estimation. The same phenomenon happens for pointwise and interval-wise loss functions. No additional loss appears for L_∞ problem.

6 Proof of the upper bounds

The sketch of the proof of the upper bound results of type (9) is the following. First we give an upper bound of the first-type error by using the Chebychev Inequality or an exponential inequality. We choose the critical value $t_{j,n,m}$ in such a way that the first-type error is upper bounded by $\gamma/2$. Next, we find an upper bound of the second-type error. In order to do this, we give the asymptotic law of our estimator T_j. As usual, the balance between bias and variance allows us to compute the optimal level j^* in the estimators expression. Finally, the minimal separation distance between hypotheses H_0 and H_1 is chosen such that the second-type error is bounded by $\gamma/2$ as well.

First, in Lemma 1, we evaluate the estimation bias. Then, Lemma 2 gives the asymptotic law of the estimator T_j. Under the null hypothesis H_0, the exponential inequalities are stated in Lemma 3. Finally, we study first and second-type errors. The proofs of Lemma 2 and Lemma 3 are postponed to Appendix A.

In the sequel, we put

$$s = s_f \wedge s_g, \quad N = \left(\frac{1}{n} + \frac{1}{m}\right)^{-1}.$$

6.1 Bias bounds

Let us recall the characterization of the Besov spaces thanks to the wavelet coefficients:

$$
\begin{aligned}
h \in B^s_{p\infty}(R) \quad &\Longrightarrow \quad \forall j \geq 0, \quad \sum_k |\beta_{jk}|^p \leq R\, 2^{-j(s+\frac{1}{2}-\frac{1}{p})p}, \\
h \in B^s_{\infty\infty}(R) \quad &\Longrightarrow \quad \forall j \geq 0, \quad \sup_k |\beta_{jk}| \leq R\, 2^{-j(s+\frac{1}{2})}.
\end{aligned} \tag{12}
$$

The following lemma is a direct application of (12).

Lemma 1 *Let $0 < s_f, s_g < DB$ and $R > 0$. Assume that $f \in B_{s_f p\infty}(R)$ and $g \in B_{s_g p\infty}(R)$ for p given in the table. The quantities B_{j_0}, b, b_{j_0} given in Table 1 satisfy*

$$\forall j_0 > 0, \quad |B_{j_0}| \quad \leq \quad b\, b_{j_0}.$$

l	p	B_{j_0}	b	b_{j_0}
x_0	∞	$\sum_{j=j_0}^{\infty} \sum_k (\beta_{jk} - b_{jk}) \psi_{jk}(x_0)$	$2(2DB-1)\|\psi\|_\infty R$	$2^{-j_0 s}$
A	∞	$\sum_{j=j_0}^{\infty} \sum_k (\beta_{jk} - b_{jk}) \int_A \psi_{jk}$	$2\|\psi\|_1 R$	$2^{-j_0 s}$
L_2	2	$\sum_{j=j_0}^{\infty} \sum_k (\beta_{jk} - b_{jk})^2$	$2R$	$2^{-2j_0 s}$
L_∞	∞	$\sup_x \lvert \sum_{j=j_0}^{\infty} \sum_k (\beta_{jk} - b_{jk}) \psi_{jk}(x) \rvert$	$2R(2DB-1)\|\psi\|_\infty$	$2^{-j_0 s}$

Table 3: Bias of test statistics

6.2 Asymptotic distribution

We establish the asymptotic normality of the statistics of interest. We remark the difference between the statistics associated either to the quadratic problem or the other problems. In the quadratic case, the order of the variance can be different under the null hypothesis and under the alternative. Note that, under H_1, on the parametric side (i.e. when the rate of convergence is $\sqrt{N}$), we give the exact constant u appearing in the variance. It is interesting to notice that u is depending on $l(f-g) = \|f-g\|_2^2$ and then can be bounded from below under H_1. This remark is fundamental for the study of the error of the second type.

Lemma 2 *In the quadratic problem, assume that the densities f and g are bounded. In supremum problem, assume that f and g are compactly supported. For j large enough and B_j given in Table 1, the statistics T_j defined in (3), (4), (5) and the statistic T_{jk} defined in (7), have the following properties*

$$ET_j = l(f-g) - B_j, \qquad\qquad ET_{jk} = l(f-g)$$

and, denoting either T_j or T_{jk} by S_{jk}

$$V(S_{jk}) = \begin{cases} v\, v_{j,n,m} + \frac{u}{N} & \textit{(in the quadratic problem)} \\ v\, v_{j,n,m} & \textit{(in the other problems)} \end{cases}$$

where

$$u = \int (f-g)^2(f+g) - 2\int (f-g)f \int (f-g)g - \left(\int (f-g)^2\right)^2$$

and the constant v is bounded by v_{max} (see Table 4). Assume in addition that

l	$l(h)$	$v\max$	$v_{j,n,m}$
x_0	$h(x_0)$	$(\|f\|_\infty \vee \|g\|_\infty)(2DB-1)^2\|\phi\|_\infty^2$	$2^j N^{-1}$
A	$\int_A h$	$[\int_A f(1-\int_A f)] \vee [\int_A g(1-\int_A g)]$	N^{-1}
L_2	$\int h^2$	$\|f\|_2^2 \|g\|_2^2$	$2^j N^{-2}$
L_∞	$\int h\phi_{jk}$	$(\|f\|_\infty \vee \|g\|_\infty)$	N^{-1}

Table 4: Asymptotic variance of test statistics

- *for the point wise problem: x_0 is such that f and g are bounded from below on $I_0 = [x_0 - 2^{-j}(2DB-1), x_0 + 2^{-j}(2DB-1)]$*
- *for the interval wise problem: A is such that $|A| > 2^{-j}$ and $\bar{A} \subset (supp\ (f) \cap supp\ (g))$*

then, under H_0 for the quadratic problem, under either H_0 or H_1 for the other problems, we have

$$(v_{j,n,m})^{-1/2}\ (S_{jk} - ES_{jk}) \xrightarrow{\mathcal{D}}_{n\wedge m \to +\infty} \mathcal{N}(0, v).$$

Moreover, under H_1 for the quadratic problem, we have

$$\begin{array}{lll} \textit{if } 2^j \le N & \textit{then} & \sqrt{N}\ (T_j - ET_j) \xrightarrow{\mathcal{D}}_{n\wedge m \to +\infty} \mathcal{N}(0, u) \\ \textit{if } 2^j > N & \textit{then} & (v_{j,n,m})^{-1/2}\,(T_j - ET_j) \xrightarrow{\mathcal{D}}_{n\wedge m \to +\infty} \mathcal{N}(0, v). \end{array}$$

6.3 Exponential inequalities

The following lemma is used for the particular case of the supremum problem to establish the optimality of the procedure $D_{j^*}(L_\infty)$ and for all problems $D(d)$ to study the rates of the adaptive procedures.

Lemma 3 is valid under H_0 and H_1 except in the case of the quadratic problem. In the L_2 framework, the $U-$statistic T_j has to be degenerate and this holds under H_0. Anyway, we only need the exponential inequality to bound the error of the first type.

Lemma 3 *The statistic S_{jk} denotes either T_j defined in (3), (4), (5) or T_{jk} defined in (7). Under the null hypothesis H_0, for j large enough, there exist constants $c, C, \tilde{C}$ depending on $\|f\|_\infty$ and $\|\phi\|_\infty$ such that*

$$\forall 0 < \lambda \le c\ c_{j,n,m}, \forall k, \quad P(|S_{jk}| \ge \lambda) \le C \exp\left(-\tilde{C} K(\lambda, j, n, m)\right)$$

where $v_{j,n,m}$ is given in Table 4 and $C = 2$ except in the quadratic problem.

l	$c_{j,n,m}$	$K(\lambda, j, n, m)$
x_0	$2^{j/2}$	$\frac{\lambda^2}{v_{j,n,m}}$
A	1	$\frac{\lambda^2}{v_{j,n,m}}$
L_2	$+\infty$	$\frac{\lambda^2}{v_{j,n,m}} \wedge \left(N\lambda \wedge N\frac{\lambda^{2/3}}{2^{j/3}} \wedge N\frac{\lambda^{1/2}}{2^{j/2}}\right)$
L_∞	$2^{-j/2}$	$\frac{\lambda^2}{v_{j,n,m}}$

6.4 Proof of Theorem 1 - Pointwise, interval wise and quadratic problems

Let j be varying between j_0 and j_∞ defined in (11). Let us remark that under H_0, the quantity ET_j is zero. Applying Lemma 2 and using Chebyshev Inequality, we obtain

$$P_0(D_j(d)=1) \quad = \quad P_0(|T_j| \geq t_{j,n,m}) \leq t_{n,m}^{-2}\, v\, v_{j,n,m}.$$

We choose

$$t_{j,n,m} = \sqrt{\frac{v\mathrm{max}}{(\gamma/2)}\, v_{j,n,m}} \tag{13}$$

in order to bound the risk of the first type by $\gamma/2$. Using Lemma 2, the estimator T_j satisfies the following equality

$$ET_j \quad = \quad l(f-g) \, - \, B_j.$$

Under the alternative H_1, we get

$$\begin{aligned}
& P_{f,g}(D_j(d)=0) = P_{f,g}(|T_j| \leq t_{j,n,m}) \\
\leq \;& P_{f,g}\left(-t_{j,n,m} - l(f-g) + B_j \leq T_j - ET_j \leq t_{j,n,m} - l(f-g) + B_j\right) \\
\leq \;& P_{f,g}\left(-t_{j,n,m} - l(f-g) + B_j \leq T_j - ET_j \text{ if } l(f-g) < 0\right) \\
& + P_{f,g}\left(T_j - ET_j \leq t_{j,n,m} - l(f-g) + B_j \text{ if } l(f-g) > 0\right) \\
\leq \;& 2P_{f,g}\left(\frac{l(f-g)}{\sqrt{V(T_j)}}\left(1 - \frac{|B_j|}{r_{n,m}} - \frac{t_{j,n,m}}{r_{n,m}}\right) \leq \chi\right).
\end{aligned} \tag{14}$$

Following Lemma 2, the random variable χ is asymptotically Gaussian. Let us denote q_γ the quantile verifying $2\,P(\mathcal{N}(0,1) > q_\gamma) = \gamma/2$. In the quadratic problem, we restrict the study for a level j such that $2^j > N$ which implies that $V(T_j) \leq v v_{j,n,m}$. The probability term is bounded as soon as

$$r_{n,m} \geq (4t_{j,n,m}) \vee (4|B_j|) \vee (2q_\gamma \sqrt{V(T_j)}).$$

Combining Lemma 1 and (13), we choose the smoothing index j^* such that

$$\sqrt{\frac{v\mathrm{max}}{(\gamma/2)}\, v_{j^*,n,m}} = b \quad b_{j^*}.$$

Remark that in the quadratic problem, the optimal index is given by

$$2^{j^*} = N^{\frac{2}{1+4s}}$$

and then is larger than N if $s < 1/4$. We explore the case where $s \geq 1/4$ at the end of this proof. We get

$$(4t_{j^*,n,m}) \vee (4|B_{j^*}|) \vee (2q_\gamma\sqrt{V(T_{j^*})}) \leq (4 \vee \sqrt{2\gamma} q_\gamma) t_{j^*,n,m}.$$

Using the critical value found in (13), the rates verify

$$r_{n,m} \quad \geq \quad (4 \vee \sqrt{2\gamma} q_\gamma)\sqrt{\frac{v\mathrm{max}}{(\gamma/2)}\, v_{j^*,n,m}}$$

Replacing $v_{j^*,n,m}$, we obtain the rate $r^*_{n,m}$ announced in Table 1. We finish the proof considering the quadratic problem when $s \geq 1/4$. Since $2^{j^*} < N$, we have $V(T_j) \leq \frac{u}{N}$. By Cauchy-Schwarz Inequality, we bound u

$$u \leq c_{fg}\; l(f-g),$$

for $c_{fg} = \|f-g\|_\infty + 2\|f\|_2\,\|g\|_2$ and (14) writes

$$\begin{aligned}
P_{f,g}(D_{j^*}=0) &\leq 2P_{f,g}\left(\frac{l(f-g)}{\sqrt{c_{fg}\,N^{-1}l(f-g))}}\left(1-\frac{|B_{j^*}|}{r^*_{n,m}}-\frac{t_{j^*,n,m}}{r^*_{n,m}}\right)\leq \chi\right)\\
&\leq 2P_{f,g}\left(\left(\frac{N\,r^*_{n,m}}{4c_{fg}}\right)^{1/2}\leq \chi\right),
\end{aligned}$$

Since $Nr^*_{n,m}$ tends to infinity, this probability is going to 0 and the proof is completed.

6.5 Proof of Theorem 1 - Supremum problem

Let j be varying between j_0 and j_∞ defined in (11). Under H_0, the quantities $\alpha_{jk} - a_{jk}$ are zero for any k. Let us remark that, since f and g are compactly supported, the number of coefficients $T_{jk} = \hat{\alpha}_{jk} - \hat{a}_{jk}$ appearing in $\max_k |T_{jk}|$ is less than $2^j L + (2DB-1)$ where L (respectively $2DB-1$) is the length of the support of f and g (respectively of ϕ). Using Lemma 3, there exists a constant $\tilde{C}_0 > 0$ depending on $\|f\|_\infty$ and $\|\phi\|_\infty$ such that

$$\begin{aligned}
P_0(D_j(L_\infty)=1) &= P_0(\max_k |T_{jk}| \geq t_{j,n,m})\\
&\leq \sum_k P_0\left(|T_{jk}|\geq t_{j,n,m}\right)\\
&\leq 2(2^jL+(2DB-1))\exp\{-t^2_{j,n,m}\tilde{C}_0 v^{-1}_{j,n,m}\}
\end{aligned}$$

as soon as $t_{j,n,m} \leq c2^{-j/2}$. The choice

$$t_{j,n,m} = \sqrt{(\log(4L\gamma^{-1})+j)\,\tilde{C}_0^{-1}\,v_{j,n,m}}$$

is convenient because $2^j \leq n\log(n)^{-1}$ and allows to bound the risk of the first type for the test $D_j(L_\infty)$ by $\gamma/2$. Remark that the expansion (6) on the wavelet basis implies

$$\max_k |ET_{jk}| \geq 2^{j/2}\|\phi\|_\infty^{-1}(\|f-g\|_\infty - B_j)$$

and therefore, under H_1, there exists k^* such that

$$|ET_{jk^*}| \geq 2^{j/2}\|\phi\|_\infty^{-1}(r_{n,m} - |B_j|).$$

Using Lemma 2, there exists an asymptotically Gaussian variable χ such that

$$\begin{aligned}
&P_{f,g}(D_j(L_\infty)=0) = P_{f,g}(T_j(L_\infty)\leq t_{j,n,m})\\
\leq\ & P_{f,g}\left(\forall k, |T_{jk}|\leq t_{j,n,m}\right)\\
\leq\ & P_{f,g}\left(-t_{j,n,m}-ET_{jk^*}\leq T_{jk^*}-ET_{jk^*}\leq t_{j,n,m}-ET_{jk^*}\right)\\
\leq\ & P_{f,g}\left(-t_{j,n,m}-(\alpha_{jk^*}-a_{jk^*})\leq T_{jk^*}-ET_{jk^*}\quad \text{if } \alpha_{jk^*}<a_{jk^*}\right)\\
& +P_{f,g}\left(T_{jk^*}-ET_{jk^*}\leq t_{j,n,m}-(a_{jk^*}-a_{jk^*})\quad \text{if } \alpha_{jk^*}>a_{jk^*}\right)\\
\leq\ & 2P_{f,g}\left(\chi\geq \frac{2^{j/2}r_{n,m}}{\sqrt{V(T_{jk^*})}}\|\phi\|_\infty^{-1}\left(1-\frac{|B_j|\,\|\phi\|_\infty}{r_{n,m}}-\frac{2^{-j/2}t_{j,n,m}\|\phi\|_\infty}{r_{n,m}}\right)\right).
\end{aligned}$$

We finish the proof in the same way as previously.

6.6 Proof of Theorem 3

Put $\gamma_j = \gamma|J|^{-1}/2$ and note that there exists a constant c such that $\gamma_j = c(\log N)^{-1}$. We follow the lines of the proof in the non adaptive setting replacing γ with γ_j. In the supremum problem, taking

$$\tilde{t}_{j,n,m} \quad = \quad \left(\frac{j\log(2) + \log(4L/\gamma_j)}{N} \right)^{1/2},$$

we get

$$P_0(D(d)=1) \quad = \quad P_0(\exists j \in J, |T_j| > \tilde{t}_{j,n,m}) \le \sum_{j\in J} P_0(|T_j| > \tilde{t}_{j,n,m}) \le \sum_{j\in J}\gamma_j = \gamma/2.$$

In the other problems, we choose

$$\tilde{t}_{j,n,m} = \left(\tilde{C}_0^{-1} v_{j,n,m} \log(C/\gamma_j) \right)^{1/2}$$

where $C, \tilde{C}_0 > 0$ are the constants of Lemma 3. Applying Lemma 3, we have

$$P_0(D(d)=1) \quad \le \quad C\sum_{j\in J} \exp\left(-\tilde{C}_0 \frac{\tilde{t}^2_{j,n,m}}{v_{j,n,m}} \right),$$

as soon as $2^j \le \frac{N^2}{(\log\log N)^3}$ for the quadratic problem. We obtain then

$$P_0(D(d)=1) \le \gamma/2.$$

On the other hand,

$$P_{f,g}(D(d)=0) \quad = \quad P_{f,g}(\forall j \in J, |T_j| < \tilde{t}_{j,n,m}) \le P_{f,g}(|T_{j^{**}}| < \tilde{t}_{j^{**},n,m})$$

for $j^{**} \in J$ to determine. The choices

$$2^{j^{**}} = 2^{j^*}(\log|J|)^{-\epsilon}$$

for $\epsilon = (1+2s)^{-1}/2$ in the point wise problem, $\epsilon = (2s)^{-1}$ in the interval wise problem, $\epsilon = (1+4s)^{-1}$ in the quadratic problem, $\epsilon = 0$ in the supremum problem lead to the announced rates.

7 Proof of the lower bounds

The main idea in the proof of the lower bounds is to reduce substantially the large class of functions to a parametric subset. If this finite (but increasing with $n \wedge m$) set is well chosen, the distance between these functions is giving the optimal rate, while the distance between the resulting models decreases to 0 or it is upper bounded by some constant. The proofs for the pointwise and supremum problem are based on the following Lemma (which is proved at the end of this section)

Lemma 4 *Let $\{(f,g_1),\ldots,(f,g_M)\}$ be M-couples of density functions in the class $\Lambda_{n,m}(C) = \mathcal{R} \cup \mathcal{S}_{n,m}(C)$. Moreover, denote h the common density under the null hypothesis and assume that*

$$P_{h,h}\left[\frac{1}{M}\sum_{k=1}^{M} \frac{dP_{f,g_k}}{dP_{h,h}} \ge 1-\delta \right] \ge \frac{\gamma}{1-\delta}, \tag{15}$$

for some $0<\gamma<1$ and $0<\delta<1-\gamma$.

Then the lower bound in (10) *holds true.*

For the quadratic problem, this lemma based on M hypothesis is not enough but we consider a richer subfamily of experiments. We construct our hypotheses similarly to Ingster [5] or Pouet [14]. We write the proof based on Assouad's cube in a simpler manner, so that it is easy to see why this richer family has such a small Bayesian risk.

For the minimax pointwise and supremum setups, we follow the lines of proof in Lepski and Tsybakov [11]. In the following subsections, we need only to describe the choice of these particular functions in each nonparametric setup (i.e. except the interval-wise setup) and prove that they verify all needed conditions.

We mention that the cited proofs were given for the one-sample goodness-of-fit problem instead of the two-sample homogeneity test problem that we consider here. As we already mentioned these proofs need to be based on a general underlying density h under H_0 that cannot be reduced to a uniform density since h is unknown.

Assume for convenience, without loss of generality, that $n \geq m$ and $s_f \geq s_g$, note $s = s_f \wedge s_g = s_g$ and remark that N is of order $m \wedge n = m$, when $m \wedge n \to \infty$.

In all specific constructions further on, we consider

$$h = f,$$

i.e. we construct hypotheses in Lemma 4 based on the common density under the null hypothesis H_0. Then $g_1, \dots, g_M$ are basically this very density plus suitable perturbations.

We need to assume that this common unknown density under H_0 has a Besov norm $\|h\|_B < R$, according to the setup. This is not a restriction for our former results.

7.1 Proof of Theorem 2 - Pointwise problem

In this setup, it is enough to consider $M = 1$. Define g_1 by

$$g_1(x) = f(x) + 2^{-js}\psi(2^j(x - x_0)),$$

and choose

$$2^j = (m \wedge n)^{1/(2s+1)}.$$

Step 1. The functions g_1, f have the following properties:

- $g_1 \in B^{s_g}_{\infty,\infty}(R)$: we just need to take ψ in the class $B^{s_g}_{\infty,\infty}(R')$, with R' small enough to have
$$\|g_1\|_{B^s_{\infty\infty}} \leq \|f\|_{B^s_{\infty\infty}} + 2^{-js}\|\psi\|_{B^s_{\infty\infty}} \leq R.$$
- g_1 is a density: Let us denote $m_f > 0$ the bound from below of f on I_0. Note that, as $n \wedge m \to \infty$, the support of the perturbation is shrinking and the size of the perturbation $\|2^{-js}\psi(\cdot)\|_\infty$ decreases to 0. It means that for $n \wedge m$ large enough, g_1 is a positive function. Obviously, $\int g_1 = 1$ because f is a density and $\int \psi = 0$.
- The choice of j leads to $|g_1(x_0) - f(x_0)| = 2^{-js}|\psi(0)| \geq CN^{-s/(2s+1)}$ for some $C > 0$.

We conclude that $(f, g_1) \in \Lambda_{n,m}(C)$ for $r_{n,m} = N^{-s/(2s+1)}$.
Step 2. We check Relation (15). We write

$$\begin{aligned} P &= P_{h,h}\left(\frac{dP_{f,g_1}}{dP_{h,h}} \geq 1 - \delta\right) = P_{h,h}\left(\log\prod_{i=1}^m \frac{g_1(Y_i)}{f(Y_i)} \geq \log(1-\delta)\right) \\ &= P_{h,h}\left(\frac{\sum_{i=1}^m \log Z_i - \mu_m}{\sigma_m} \geq \frac{\log(1-\delta) - \mu_m}{\sigma_m}\right) \end{aligned}$$

where

$$Z_i = \frac{g_1}{f}(Y_i),\ \mu_m = \sum_{i=1}^{m} E_{h,h} \log Z_i,\ \sigma_m^2 = \sum_{i=1}^{m} V_{h,h}(\log Z_i)$$

satisfy the following lemma (proved in Appendix A)

Lemma 5 *The Z_i's are independent variables such that the $\log Z_i$'s have finite moments up to order 3 and*

$$\lim_{m\to\infty} \sigma_m^{-3} \sum_{i=1}^{m} E_{h,h}|Z_i - E_{h,h}Z_i|^3 = 0, \tag{16}$$

$$|\mu_m| \le m_f^{-1} \quad \textit{and} \quad \sigma_m^2 \le 2m_f^{-1} \tag{17}$$

Combining Lyapounov Theorem with this lemma, P is larger than $\gamma/(1-\delta)$ as soon as

$$\frac{\log(1-\delta) + m_f^{-1}}{2m_f^{-1}} \le q_{1-\gamma/(1-\delta)}.$$

Relation (15) is verified. We just have to apply Lemma 4 to obtain the lower bound.

7.2 Proof of Theorem 2 - Supremum problem

The construction is very similar, but we consider here an increasing number of perturbed functions: $M = 2^j$. Because f is at least a continuous density, we can find a compact set on which f is positive. Without loss of generality we take this compact set to be $[0,1]$ and let $m_f > 0$ be a lower bound of f on this interval. Choose

$$2^j = \left(\frac{m \wedge n}{\log(m \wedge n)}\right)^{1/(2s+1)}$$

and define

$$g_k(x) = f(x) + 2^{-js}\psi(2^j(x - x_k)), k = 1, \dots, M$$

where $x_k = 2^{-j-1}k$.

Step 1. As in the previous part, we can prove that, for any $k = 1, \dots M$, g_k are densities belonging to $B^{s_g}_{\infty\infty}(R)$. Moreover, since there exists a constant C such that

$$\|g_k - f\|_\infty = 2^{-js}\|\psi\|_\infty \ge C \left(\frac{N}{\log N}\right)^{-\frac{s}{2s+1}},$$

we deduce that $(f, g_k) \in \Lambda_{n,m}(C)$ for $r_{n,m} = \left(\frac{N}{\log N}\right)^{-\frac{s}{2s+1}}$.

Step 2. We have to verify Relation (15). For $k = 1, \dots M$, put $Z_i^{(k)} = (g_k/f)(Y_i)$ and $U_k = \prod_{i=1}^m Z_i^{(k)}$. Note that $E_{h,h}Z_i^{(k)} = 1$ and that U_k are independent variables since g_k have disjoint supports. We write

$$\begin{aligned}
& P_{h,h}\left(\frac{1}{M}\sum_{k=1}^{M} \frac{dP_{f,g_k}}{dP_{h,h}}(X_1,\dots,X_n,Y_1,\dots,Y_m) \ge 1-\delta\right) \\
= \quad & P_{h,h}\left(\frac{1}{M}\sum_{k=1}^{M}\prod_{i=1}^{m} Z_i^{(k)} - 1 \ge -\delta\right) \\
\ge \quad & 1 - P_{h,h}\left(\left|\sum_{k=1}^{M}(U_k - E_{h,h}U_k)\right| \ge M\delta\right) \\
\ge \quad & 1 - c\frac{M}{M^2\delta^2} Var_{h,h}(U_1),
\end{aligned}$$

We used successively Markov inequality and Rosenthal inequality for moments of the sum of independent variables. Finally,

$$\begin{aligned}
Var_{h,h}(U_1) &= E_{h,h}\left|\prod_{i=1}^{m}\left(1+\frac{2^{-js}\psi(2^j(Y_i-x_1))}{f(Y_1)}\right)-1\right|^2\\
&= E_{h,h}\prod_{i=1}^{m}\left(1+\frac{2^{-js}\psi(2^j(Y_i-x_1))}{f(Y_1)}\right)^2-1\\
&= \left(1+E_{h,h}\left[\frac{2^{-2js}\psi^2(2^j(Y_1-x_1))}{f^2(Y_1)}\right]\right)^m-1\\
&= \left(1+m_f^{-2}\,2^{-2js}\int\psi^2(2^j(y-x_1)h(y)dy\right)^m-1\\
&\leq c\,m2^{-j(2s+1)},
\end{aligned}$$

which gives

$$P_{h,h}\left(\frac{1}{M}\sum_{k=1}^{M}\frac{dP_{f,g_k}}{dP_{h,h}}\geq 1-\delta\right)\geq 1-c\frac{m2^{-j(2s+1)}}{M\delta^2}\geq 1-c\frac{\log m}{\delta^2}\left(\frac{\log m}{m}\right)^{\frac{1}{2s+1}}$$

and this is larger than $\gamma/(1-\delta)$ for m large enough. This ends the proof of Relation (15). We just have to apply Lemma 4 to obtain the lower bound.

7.3 Proof of Theorem 2 - Quadratic problem

We start with the same construction as before. Nevertheless, a subfamily of $M = C^2\,2^j$ couples of densities is not rich enough in this setup. Remark that C is a fixed constant to be determined in the sequel. For $\theta_1,\ldots,\theta_M$ i.i.d. Bernoulli(1/2) random variables, let:

$$g_\theta(x)=f(x)+\sum_{k=1}^{M}\theta_k 2^{-js}\psi(2^j(x-x_k)),$$

and choose

$$2^j=(n\wedge m)^{\frac{2}{1+4s}}\,.$$

Step 1. The subfamily contains 2^M couples of densities belonging to $B^{s_g}_{2\infty}(R)$. Moreover,

$$\|g_\theta-f\|_{B^s_{2\infty}}=M^{1/2}2^{-j/2}=C.$$

The rate is given by

$$\|g_\theta-f\|_2=\left(\sum_{k=1}^{M}2^{-2js}\int\psi^2(2^j(x-x_k))dx\right)^{1/2}=M^{1/2}2^{-js}2^{-j/2}=Cm^{-2s/(4s+1)}.$$

We deduce that $(f,g_\theta)\in\Lambda_{n,m}(C)$ for $r_{n,m}=m^{-4s/(4s+1)}$.
Step 2. The proof of the theorem in this case is not based on Lemma 4, but is slightly more complicated. Indeed, when we give a lower bound of second-type error, we take the mean with respect to the measure $\pi(d\theta)$ due to the i.i.d. random variables θ. Keeping the same notations as previously, this gives:

$$\begin{aligned}
&\inf_V\left(P_{h,h}(V=1)+\max_\theta P_{f,g_\theta}(V=0)\right)\\
\geq\;&\inf_V\left(P_{h,h}(V=1)+\int P_{f,g_\theta}(V=0)\pi(d\theta)\right)\geq 1-\Delta/2,
\end{aligned}$$

where

$$\Delta^2 = E_{h,h}\left(\int\prod_{i=1}^{m}\frac{g_\theta}{f}(Y_i)\pi(d\theta)\right)^2 - 1.$$

For the last inequality, we refer to Ingster [5], Theorem 2.1. It is sufficient to upper bound:

$$\begin{aligned}
\Delta^2+1 &= E_{h,h}\left(\int\prod_{i=1}^{m}\frac{g_\theta}{f}(Y_i)\pi(d\theta)\right)^2\\
&= E_{h,h}\left(\int\prod_{i=1}^{m}\left(1+\sum_{k=1}^{M}\frac{\theta_k 2^{-js}\psi(2^j(Y_i-x_k))}{f(Y_i)}\right)\pi(d\theta)\right)^2\\
&= E_{h,h}\left(\int\prod_{i=1}^{m}\prod_{k=1}^{M}\left(1+\frac{\theta_k 2^{-js}\psi(2^j(Y_i-x_k))}{f(Y_i)}\right)\pi(d\theta)\right)^2.
\end{aligned}$$

Indeed, the previous equality is true since each Y_i can be in the support of a single perturbation $\psi(2^j(\cdot - x_k))$ at a time. Let us denote in the following $Y_{i,k}$ the random variable Y_i if it belongs to the support of the k-th perturbation and 0 otherwise. Thus, the variables under the sum are independent as k goes from 1 to M. We have, for Δ^2+1:

$$\begin{aligned}
& E_{h,h}\prod_{k=1}^{M}\left(\int\prod_{i=1}^{m}\left(1+\frac{\theta_k 2^{-js}\psi(2^j(Y_{i,k}-x_k))}{f(Y_{i,k})}\right)\pi(d\theta_k)\right)^2\\
= \; & E_{h,h}\prod_{k=1}^{M}\left[\frac{1}{2}\prod_{i=1}^{m}\left(1+\frac{2^{-js}\psi(2^j(Y_{i,k}-x_k))}{f(Y_{i,k})}\right)+\frac{1}{2}\prod_{i=1}^{m}\left(1-\frac{2^{-js}\psi(2^j(Y_{i,k}-x_k))}{f(Y_{i,k})}\right)\right]^2\\
\leq \; & \prod_{k=1}^{M}\Bigg\{\ \frac{1}{2}\left(1+E_{h,h}\left[\frac{2^{-2js}\psi^2(2^j(Y_{1,k}-x_k))}{f^2(Y_{1,k})}\right]\right)^m\\
& \qquad +\frac{1}{2}\left(1-E_{h,h}\left[\frac{2^{-2js}\psi^2(2^j(Y_{1,k}-x_k))}{f^2(Y_{1,k})}\right]\right)^m\ \Bigg\}\\
\leq \; & \left[1+m(m-1)\left(E_{h,h}\left[\frac{2^{-2js}\psi^2(2^j(Y_{1,k}-x_k))}{f^2(Y_{1,k})}\right]\right)^2\right]^M.
\end{aligned}$$

Here, we used successively the inequality $(a+b)^2\leq 2a^2+2b^2$ and the fact that

$$E_{h,h}\left[2^{-js}\psi(2^j(Y_{i,k}-x_k))f^{-1}(Y_{i,k})\right]=0.$$

We also use asymptotic approximations, since:

$$\begin{aligned}
E_{h,h}\left(\frac{2^{-2js}\psi^2(2^j(Y_{1,k}-x_k))}{f^2(Y_{1,k})}\right) &= \int\frac{2^{-2js}\psi^2(2^j(y-x_k))}{f(y)}1_{[x_k-\frac{2^{-j/2}}{2},x_k+\frac{2^{-j}}{2}]}(y)dy\\
&\leq c m_f^{-1}2^{-j(2s+1)}.
\end{aligned}$$

Remark that, since f is continuous and positive on $[0,1]$, there exists m_f such that $f(y)\geq m_f>0$ on the integration domain. On the whole, we get as a lower bound of the sum of the test errors:

$$\begin{aligned}
1-\Delta/2 &\geq 1-\frac{1}{2}\left((1+\frac{c}{m_f^2}m^2 2^{-2j(2s+1)})^M-1\right)^{1/2}\\
&\geq 1-c\,M^{1/2}m2^{-j(2s+1)}=1-c\,C.
\end{aligned}$$

We just have to choose C smaller than $(1-\gamma)c^{-1}$ to end the proof of Theorem 10.

7.4 Proof of Lemma 4

Let us first restrict the large class of functions $\{(f,g) \in \Lambda_{n,m}(C)\}$ to a finite subset $\{(f,g_1),\dots,(f,g_M)\}$ belonging to $\Lambda_{n,m}(C)$. For any statistic V, we have

$$\sup_{f,g\in\Lambda_{n,m}(C)} P_{f,g}(V=0) \quad \geq \quad \max_{(f,g_1),\dots,(f,g_M)} P_{f,g}(V=0) \geq \frac{1}{M}\sum_{k=1}^{M} P_{f,g_k}(V=0)$$

which implies

$$\begin{aligned} r &= \inf_V \left(P_{h,h}\left(V=1\right) + \sup_{f,g\in\Lambda_{n,m}(C)} P_{f,g}\left(V=0\right)\right) \\ &\geq \inf_V \left(E_{h,h}\left[I_{\{V=1\}}\right] + E_{h,h}\left[I_{\{V=0\}} \frac{1}{M}\sum_{k=1}^{M} \frac{dP_{f,g_k}}{dP_{h,h}}\right]\right). \end{aligned}$$

Use now the relation (15) and denote by $A = \{M^{-1}\sum_{k=1}^{M} dP_{f,g_k}/dP_{h,h} \geq 1-\delta\}$:

$$r \quad \geq \quad \inf_V \left[E_{h,h}\left(I_{[V=1]} + (1-\delta)I_{\{V=0\}}I_A\right)\right] \geq (1-\delta)P_{h,h}(A) \geq \gamma$$

which ends the proof.

References

[1] I. Daubechies, " Ten Lectures on Wavelets", SIAM: Philadelphia, 1992.

[2] W. Feller, "An Introduction to Probability Theory and Its Applications", Wiley, 1966.

[3] H. Dette and D. Neumeyer, Nonparametric analysis of covariance, *Ann. Statist.* **29** (2001) 1361-1400.

[4] E. Giné, R. Latala, J. Zinn, Exponential and moment inequaliries for $U-$statistics, *High dimensional probability, II*, 13–38, Progr. Probab., 47, *Birkhaüser Boston, Boston, MA,* 2000.

[5] Yu. I. Ingster, Asymptotically minimax hypothesis testing for nonparametric alternatives, I, II, III. *Math. Methods Statist.* **2** (1993) 85-114, 171-189, 248-268.

[6] Yu. I. Ingster, Adaptive chi-square tests, *J. Math. Sciences* **99** (2000) 1110-1120.

[7] Yu. I. Ingster and I. A. Suslina, "Nonparametric Goodness-of-Fit Testing Under Gaussian Models", Lecture Notes in Statistics, 169, Springer-Verlag New York, 2003.

[8] V. S. Koroljuk and Yu. V. Borovskich, "Theory of U-statistics", Kluwer Academic Publishers, 1994.

[9] G. Kerkyacharian and D. Picard, Estimating nonquadratic functionals of a density using Haar wavelets. *Ann. Statist.* **24** (1996) 485-507.

[10] O. V. Lepskii, Asymptotically minimax adaptive estimation I: Upper bounds. Optimally adaptive estimates, *Theory Probab. Appl.* **36** (1991) 682-697.

[11] O. V. Lepski and A. B. Tsybakov, Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probab. Theory Relat. Fields* **117** (2000) 17-48.

[12] A. Munk and H. Dette, Nonparametric comparison of several regression functions: exact and asymptotic theory, *Ann. Statist.* **26** (1998) 2339-2368.

[13] V. V. Petrov, “Limit Theorems of Probability Theory”, Clarendon Press, Oxford, 1995.

[14] C. Pouet, An asymptotically optimal test for a parametric set of regression functions against a non-parametric alternative, *J. Statist. Plann. Inf.* **98** (2001) 177-189.

[15] V. G. Spokoiny, Adaptive hypothesis testing using wavelets. *Ann. Statist.* **24** (1996) 2477-2498.

[16] V. G. Spokoiny, Adaptive and spatially adaptive testing of a nonparametric hypothesis. *Math. Methods Statist.* **7** (1998) 245-273.

[17] H, Triebel, *Theory of Function Spaces 2.* (1990) Basel: Birkaüser Verlag

[18] K. Tribouley, ”Adaptive Estimation of Integrated Functionals”. *Mathematical Methods of Statistics.* **9** (2000) 19–36.

8 Appendix. Upper bounds.

8.1 Proof of Lemma 2 in the supremum problem

For all k varying in

$$\mathcal{K} = \{2^j(\min_{i=1,\dots n} X_i \wedge \min_{i=1,\dots m} Y_i) - (2D-1), \dots, 2^j(\max_{i=1,\dots n} X_i \vee \max_{i=1,\dots m} Y_i)\},$$

we consider

$$T_{jk} = \frac{1}{n}\sum_{i=1}^{n} \phi_{jk}(X_i) - \frac{1}{m}\sum_{i=1}^{m} \phi_{jk}(Y_i).$$

The expectation and the variance are

$$ET_{jk} \quad = \quad \alpha_{jk} - a_{jk}, \quad V(T_k) = \frac{1}{n}v_1 + \frac{1}{m}v_2$$

with

$$v_l = \left(\int \phi_{jk}^2 h - \left(\int \phi_{jk} h\right)^2\right) \text{ and } h = f1_{\{l=1\}} + g1_{\{l=2\}}.$$

We have

$$v_l \quad \leq \quad \int \phi^2(t-k)h(2^{-j}t)dt \leq \|h\|_\infty.$$

Since $k \in \mathcal{K}$, $v_1 \wedge v_2 = V(\phi_{jk}(X)) \wedge V(\phi_{jk}(Y)) \neq 0$. It follows

$$V(T_k) \quad = \quad \left(\frac{1}{n} + \frac{1}{m}\right) v$$

where v is a constant such that $0 < v \leq \|f\|_\infty \vee \|g\|_\infty$. Since the X's (respectively the Y's) are independent and have the same distribution, the standard limit theorem holds. Since the X's and the Y's are also independent, we deduce the asymptotic normality of

$$\frac{T_k - ET_k}{\sqrt{V(T_k)}} = \frac{\overline{\phi_{jk}(X)} - E\phi_{jk}(X)}{\sqrt{n^{-1}v_1}} c_1 + \frac{\overline{\phi_{jk}(Y)} - E\phi_{jk}(Y)}{\sqrt{m^{-1}v_2}} c_2$$

where

$$c_1 = \left(\frac{n^{-1}v_1}{n^{-1}v_1 + m^{-1}v_2}\right)^{1/2} \quad \text{and} \quad c_2 = \left(\frac{m^{-1}v_2}{n^{-1}v_1 + m^{-1}v_2}\right)^{1/2}.$$

satisfy $c_1^2 + c_2^2 = 1$.

8.2 Proof of Lemma 2 in the point wise problem

Let us recall that

$$T_j(x_0) = \frac{1}{n}\sum_{i=1}^{n} Z_i^{(1)} - \frac{1}{m}\sum_{i=1}^{m} Z_i^{(2)}$$

where

$$Z_i^{(1)} = \sum_k \phi_{jk}(X_i)\phi_{jk}(x_0) \qquad Z_i^{(2)} = \sum_k \phi_{jk}(Y_i)\phi_{jk}(x_0).$$

The expectation is

$$ET_j(x_0) \;=\; EZ_1^{(1)} + EZ_1^{(2)} = \sum_k (\alpha_{jk} - a_{jk})\,\phi_{jk}(x_0).$$

We have

$$V\left(T_j(x_0)\right) = \frac{1}{n} V Z_1^{(1)} + \frac{1}{m} V Z_1^{(2)}.$$

with

$$V(Z_1^{(l)}) \;=\; \int \left(\sum_k \phi_{jk}(t)\phi_{jk}(x_0)\right)^2 h(t)dt \;-\; \left(\sum_k \phi_{jk}(x_0)\int \phi_{jk} h\right)^2$$

for $h = f1_{\{l=1\}} + g1_{\{l=2\}}$. Since the support of scaling function is $[0, 2D-1]$, the set of the indices k such that $\phi_{jk}(x_0) \neq 0$ is $\{k \in Z,\, 2^j - (2D-1) \leq k \leq 2^j\}$. We deduce that the number of terms in the sum is $2D-1$ and that the integral is on the interval $I_0 = [x_0 - 2^{-j}(2D-1), x_0 + 2^{-j}(2D-1)]$.

On one hand, we have

$$\begin{aligned} V(Z_1^{(l)}) &= 2^j \int \left(\sum_k \phi_{0k}(t)\phi(2^j x_0 - k)\right)^2 h(2^{-j}t)dt \\ &\leq 2^j \|h\|_\infty \sum_{k_1 k_2} \int |\phi_{0k_1}\phi_{0k_2}|\ \|\phi\|_\infty^2 \\ &\leq 2^j \|h\|_\infty (2D-1)^2 \|\phi\|_\infty^2. \end{aligned}$$

On the other hand,

$$\begin{aligned} \left(\sum_k \phi_{jk}(x_0)\int \phi_{jk} h\right)^2 &\leq \left(\sum_k \phi(2^j x_0 - k)\int \phi_{0k}(t) h(2^{-j}t)dt\right)^2 \\ &\leq \|h\|_\infty^2 (2D-1)^2 \|\phi\|_\infty^2\ \|\phi\|_1^2. \end{aligned}$$

It follows

$$\begin{aligned} V(Z_1^{(l)}) &\geq 2^j \left[\int \left(\sum_k \phi_{0k}(t)\phi(2^j x_0 - k)\right)^2 h(2^{-j}t)dt \;-\; 2^{-j}(2D-1)\,\|h\|_\infty\,\|\phi\|_1\,\|\phi\|_\infty\right] \\ &\geq 2^j \left[\inf_{x\in I_0} h(x)\sum_k \phi^2(2^j x_0 - k) \;-\; 2^{-j}(2D-1)\,\|h\|_\infty\,\|\phi\|_1\,\|\phi\|_\infty\right] \end{aligned}$$

Let k^* be such that $\phi^2(2^j x_0 - k) \neq 0$ (for instance $k^* = 2^j x_0 - 1$ if the scaling function satisfies $\phi(1) \neq 0$). We deduce, as soon as j is large enough,

$$V\left(T_j(x_0)\right) \;=\; 2^j \left(\frac{1}{n} + \frac{1}{m}\right) v(x_0),$$

where $0 < v(x_0) \leq \left(\|f\|_\infty \vee \|g\|_\infty\right)(2D-1)^2\,\|\phi\|_\infty^2$. The proof of the asymptotic normality is the same as in the previous subsection.

8.3 Proof of Lemma 2 in the interval wise problem

Let us recall that the Haar scaling function is $\phi_{jk}(x) = 2^{j/2} 1_{[0,1]}(2^j x - k)$. Let us put $A = [a, b]$ such that the diameter $|A| = (b - a)$ is more than 2^{-j} and let us denote $\mathcal{K} = \{2^j a, \ldots, 2^j b - 1\}$ the set of indices such that $\phi_{jk}(x) \neq 0$ for $x \in A$. We get

$$\int_A \phi_{jk} = 2^{-j/2} 1_{\{k \in \mathcal{K}\}}. \tag{18}$$

Let us recall that

$$T_j(A) = \frac{1}{n} \sum_{i=1}^{n} Z_i^{(1)} - \frac{1}{m} \sum_{i=1}^{m} Z_i^{(2)}$$

where

$$Z_i^{(1)} = \sum_k \phi_{jk}(X_i) \int_A \phi_{jk} \qquad Z_i^{(2)} = \sum_k \phi_{jk}(Y_i) \int_A \phi_{jk}.$$

The expectation is

$$ET_j(A) = EZ_1^{(1)} + EZ_1^{(2)} = \sum_k (\alpha_{j_0k} - a_{j_0k}) \int_A \phi_{j_0k}.$$

We have

$$V(T_j(A)) = \frac{1}{n} VZ_1^{(1)} + \frac{1}{m} VZ_1^{(2)}.$$

with

$$VZ_1^{(l)} = \sum_{k_1,k_2} \left(\int \phi_{j_0k_1}\phi_{j_0k_2} h - \int \phi_{j_0k_1} h \int \phi_{j_0k_2} h \right) \int_A \phi_{j_0k_1} \int_A \phi_{j_0k_2}$$

for $h = f 1_{\{l=1\}} + g 1_{\{l=2\}}$. Since the supports of $\phi_{j_0k_1}$and of $\phi_{j_0k_2}$ are disjoint when $k_1 \neq k_2$, applying Lemma 8 and using (18), we get

$$\begin{aligned}
VZ_1^{(l)} &= 2^{-j} \left(\sum_{k \in K} \int \phi^2(2^j x - k) h - \left(\int \sum_{k \in K} \phi_{j_0k} h \right)^2 \right) \\
&= \int_A h \int_0^1 \sum_k \phi_{0k}^2 - \left(\int_A h \right)^2 \left(\int_0^1 \sum_k \phi_{0k} \right)^2 \\
&= \left(\int_A h \right) \left(1 - \int_A h \right).
\end{aligned}$$

Since $\bar{A} \subset (\text{supp } (f) \cap \text{supp } (g))$, we obviously have $V(Z_1^l) > 0$ for $l = 1$ or $l = 2$. We deduce

$$V(T_j(A)) = \left(\frac{1}{n} + \frac{1}{m} \right) v(A),$$

for

$$0 < v(A) \le \left(\int_A f \right) \left(1 - \int_A f \right) \vee \left(\int_A g \right) \left(1 - \int_A g \right).$$

The proof of the asymptotic normality is the same as in the previous subsection.

8.4 Proof of Lemma 2 in the quadratic problem

Put $q = n \wedge m$. Let us recall that the estimator $T_j(L_2)$ is given by

$$T_j(L_2) = \frac{1}{q(q-1)} \sum_{i_1,i_2=1}^{q} U_{i_1,i_2}$$

where

$$U_{i_1,i_2} \quad = \quad \sum_k \left(\phi_{jk}(X_{i_1}) - \phi_{jk}(Y_{i_1})\right)\left(\phi_{jk}(X_{i_2}) - \phi_{jk}(Y_{i_2})\right) 1_{\{i_1 \neq i_2\}}.$$

Since

$$EU_{i_1,i_2} \quad = \quad \sum_k \left(\int \phi_{jk} f - \int \phi_{jk} g\right)^2,$$

we deduce the expectation of the estimator $T_j(L_2)$

$$ET_j(L_2) \quad = \quad \sum_k \left(\alpha_{jk} - a_{jk}\right)^2.$$

Let us compute the second moments of the U's.:

$$cov(EU_{i_1 i_2}, EU_{i_3 i_4}) \sum_{k_1 k_2} a_{k_1 k_2}^2 1_{\{i_1=i_3, i_2=i_4\}} + \left(\sum_k b_k^2\right)^2 1_{\{i_1 \neq i_2 \neq i_3 \neq i_4\}}$$
$$+ \sum_{k_1 k_2} a_{k_1 k_2} b_{k_1} b_{k_2} \left(1_{\{i_1=i_3, i_2 \neq i_4\}} + 1_{\{i_1 \neq i_3, i_2=i_4\}} + 1_{\{i_1=i_4, i_2 \neq i_3\}} + 1_{\{i_1 \neq i_4, i_2=i_3\}}\right)$$

where

$$a_{k_1 k_2} \quad = \quad \left(\int \phi_{jk_1}\phi_{jk_2} f - 2\int \phi_{jk_1} f \int \phi_{jk_2} g + \int \phi_{jk_1}\phi_{jk_2} g\right)$$
$$b_k \quad = \quad \left(\int \phi_{jk} f - \int \phi_{jk} g\right)$$

We deduce that there exists a constant $c > 0$ such that

$$V(T_j(L_2)) \quad = \quad \frac{1}{q(q-1)} \sum_{k_1 k_2} a_{k_1 k_2}^2 + 4\left(\frac{1}{q-1} + \frac{c}{q^2}\right) \sum_{k_1 k_2} \left(a_{k_1 k_2} b_{k_1} b_{k_2} - b_{k_1}^2 b_{k_2}^2\right).$$

The evaluation of each quantity is given in the following lemma which is proved at the end of this section.

Lemma 6 *Let $s_f, s_g > 0$. If f (respectively g) is a bounded function belonging to $\mathcal{B}^{s_f}_{2,\infty}(R)$ (respectively to $\mathcal{B}^{s_g}_{2,\infty}(R)$), there exists a constant $c > 0$ such that , for any level j,*

$$\sum_{k_1 k_2} a_{k_1 k_2}^2 \quad \leq \quad 2^j \, (2DB - 1)\|f\|_2^2 \|g\|_2^2$$
$$\sum_{k_1 k_2} \left(a_{k_1 k_2} b_{k_1} b_{k_2} - b_{k_1}^2 b_{k_2}^2\right) \quad = \quad \int (f-g)^2 (f+g) - 2\int (f-g) f \int (f-g) g$$
$$- \left(\int (f-g)^2\right)^2 + c 2^{-j(s_f \wedge s_g)}.$$

Remark that, under H_0, the b's are zero. We deduce that, under H_0

$$ET_j(L_2) = 0 \text{ and } V(T_j(L_2)) \leq \frac{2^j}{q^2} \|f\|_2 \, \|g\|_2.$$

Under H_1, applying Lemma 6, we obtain

$$V(T_j(L_2)) = \frac{2^j}{q^2} v + \frac{1}{q} u$$

where u, v are given in Lemma 2. Using Hoeffding Central Limit Theorem for second order non degenerate U-statistics (see for instance Koroljuk and Borovskich [8]), we prove the asymptotic normality of $T_j(L_2)$. The key point is that $T_j(L_2)$ is non degenerate under H_1 since:

$$EU_{1,2} - E(U_{1,2}/X_2, Y_2) \quad = \quad \sum_k b_k (b_k - (\phi_{jk}(X_1) - \phi_{jk}(Y_1)) \neq 0.$$

8.5 Proof of Lemma 3 in the point wise, interval wise and supremum problems

Recall the following result

Proposition 1 Bernstein's Inequality *(see [13], p. 57). Suppose that* $E\xi_i = 0$, $\sigma_i^2 = E\xi_i^2 < \infty\,(i = 1, \dots, m)$, $B = \sum_{i=1}^m \sigma_i^2$. *Suppose there exists positive constants* H *and* C_0 *such that*

$$|E\xi_i^p| \le C_0^p\, p^p \sigma_i^2 H^{p-2} \quad (i = 1, \dots, m)$$

for all integers $p \ge 2$. *Then, there exists a positive constant* $\tilde{C}_0$ *depending on* C_0 *such that*

$$\forall\, 0 \le x \le B/H, \quad P(\,|\sum_{i=1}^m \xi_i| \ge x) \le 2\exp -\tilde{C}_0\,\frac{x^2}{B}.$$

In fact, the above inequality has sightly different assumptions (about the moment condition of ξ) than the inequality in Petrov [13] p. 57. The proof is the same (using the concavity of the log function which leads to the inequality $q! \ge q^q \exp(-q+1), q > 2$).

Let us first consider the supremum problem. Suppose that $m \le n$, let m_f be a lower bound for f. Put

$$Z_i \;=\; \left(\frac{1}{n}\phi_{jk}(X_i) - \frac{1}{m}\phi_{jk}(Y_i)\right)\, 1_{\{1\le i\le m\}} + \frac{1}{n}\phi_{jk}(X_i)\, 1_{\{m+1\le i\le n\}}$$

and

$$\xi_i = Z_i - EZ_i.$$

Observe that, under H_0

$$\sum_{i=1}^n EZ_i = 0$$

and then, we have

$$T_{jk} = \sum_{i=1}^n (Z_i - EZ_i) = \sum_{i=1}^n \xi_i.$$

Let q be an integer larger than 1. Since

$$\begin{aligned}
EZ_i &= \left(\left(\frac{1}{n} + \frac{1}{m}\right) 1_{\{1\le i\le m\}} + \frac{1}{n}\, 1_{\{m+1\le i\le n\}}\right) \int \phi_{jk} f \\
EZ_i^2 &= \left(\left(\frac{1}{n^2} + \frac{1}{m^2}\right) 1_{\{1\le i\le m\}} + \frac{1}{n^2}\, 1_{\{m+1\le i\le n\}}\right) \int \phi_{jk}^2 f \\
|EZ_i^p| &\le 2^{j(\frac{p}{2}-1)} \left(\left(\frac{1}{n^p} + \frac{1}{m^p}\right) 1_{\{1\le i\le m\}} + \frac{1}{n^p}\, 1_{\{m+1\le i\le n\}}\right) 2^{p-1}\|f\|_\infty\, \|\phi\|_p^p
\end{aligned}$$

we deduce that there exists some constant $C > 0$ such that

$$\begin{aligned}
\sigma_i^2 &= \left(\left(\frac{1}{n^2} + \frac{1}{m^2}\right) 1_{\{1\le i\le m\}} + \frac{1}{n^2}\, 1_{\{m+1\le i\le n\}}\right) \left(\int \phi_{jk}^2 f - \left(\int \phi_{jk} f\right)^2\right) \\
&\ge \left(\left(\frac{1}{n} + \frac{1}{m}\right)^2 1_{\{1\le i\le m\}} + \frac{1}{n^2}\, 1_{\{m+1\le i\le n\}}\right) m_f \left(\frac{1}{2} - C2^{-j}\right) \\
E|\xi_i|^p &\le \left(\left(\frac{1}{n} + \frac{1}{m}\right)^p 1_{\{1\le i\le m\}} + \frac{1}{n^p}\, 1_{\{m+1\le i\le n\}}\right) 2^{p-1}\, 2^{j(\frac{p}{2}-1)}\, \|f\|_\infty\, \|\phi\|_p^p \frac{\sigma_i^2}{\sigma_i^2} \\
&\le 2^p \|f\|_\infty m_f^{-1} \|\phi\|_p^p \sigma_i^2 \left(2^{j/2}\left(\frac{1}{n} + \frac{1}{m}\right)^p 1_{\{1\le i\le m\}} + \frac{1}{n^p}\, 1_{\{m+1\le i\le n\}}\right)^{p-2}
\end{aligned}$$

We finish the proof applying Petrov Inequality with

$$B = \left(\frac{1}{n} + \frac{1}{m}\right) \|f\|_\infty \qquad \text{and} \qquad H = 2^{j/2}\left(\frac{1}{n} + \frac{1}{m}\right).$$

The proof is analogue in the point wise problem and in the interval wise problem, setting

$$\begin{aligned} Z_i &= \left(\frac{1}{n}\sum_k \phi_{jk}(X_i)\phi_{jk}(x_0) - \frac{1}{m}\sum_k \phi_{jk}(Y_i)\phi_{jk}(x_0)\right) 1_{\{1\le i\le m\}} \\ &+ \frac{1}{n}\sum_k \phi_{jk}(X_i)\phi_{jk}(x_0)\ 1_{\{m+1\le i\le n\}} \\ \tilde{Z}_i &= \left(\frac{1}{n}\sum_k \phi_{jk}(X_i)\int_A \phi_{jk} - \frac{1}{m}\sum_k \phi_{jk}(Y_i)\int_A \phi_{jk}\right) 1_{\{1\le i\le m\}} \\ &+ \frac{1}{n}\sum_k \phi_{jk}(X_i)\int_A \phi_{jk}\ 1_{\{m+1\le i\le n\}} \end{aligned}$$

with (B, H) being respectively

$$\left(2^j\left(\frac{1}{n} + \frac{1}{m}\right)\|f\|_\infty, 2^{j/2}\left(\frac{1}{n} + \frac{1}{m}\right)\right), \qquad \left(\left(\frac{1}{n} + \frac{1}{m}\right)\|f\|_\infty, \left(\frac{1}{n} + \frac{1}{m}\right)\right)$$

which leads to the announced result.

8.6 Proof of Lemma 3 in the quadratic problem

We recall the following result from Gine et al. [4]

Proposition 2 *There exist universal constants $C', \tilde{C} > 0$ such that, if u is a bounded canonical kernel, completely degenerate, of the i.i.d. variables $Z_1, \dots Z_q$, then for all $x > 0$,*

$$P\left(\left|\sum_{1\le i_1\ne i_2\le q} u(Z_{i_1}, Z_{i_2})\right| \ge x\right) \le C' \exp\left(-\tilde{C}^{-1}\left(\frac{x^2}{D^2} \wedge \frac{x}{C} \wedge \frac{x^{2/3}}{B^{2/3}} \wedge \frac{x^{1/2}}{A^{1/2}}\right)\right)$$

where A, B, C, D are defined by

$$A = \|u\|_\infty, \quad B^2 = q\|Eu^2(Z,.)\|_\infty, \quad C^2 = q^2 Eu^2,$$

and

$$D^2 = q \sup\left\{Eu(Z_1, Z_2)u_1(Z_1)u_2(Z_2), Eu_1^2(Z) \le 1, Eu_2^2(Z) \le 1\right\}.$$

We apply this proposition with $Z = (X, Y)$ and

$$u(z_1, z_2) = \sum_k (\phi_{jk}(x_1) - \phi_{jk}(y_1))(\phi_{jk}(x_2) - \phi_{jk}(y_2))$$

which is is degenerate under H_0. The following Lemma, proven at the end of Section 8, gives evaluations for $A, B, C, D.$ evaluations

Lemma 7 *There exists some positive constant c such that*

$$A \le c2^j, \quad B^2 \le c2^j q, \quad C^2 \le cq^2 2^j, \quad D \le cq.$$

For all $x > 0$ and all level j, we obtain

$$P_0(|T_j(L_2)| \ge x) \le C' \exp\left[-\tilde{C}\left(q^2x^2 \wedge \frac{qx}{2^{j/2}} \wedge \frac{qx^{2/3}}{2^{j/3}} \wedge \frac{qx^{1/2}}{2^{j/2}}\right)\right] \tag{19}$$

We end the proof remarking that $q = n \wedge m$ has same order than N.

8.7 Proof of Lemma 6 and Lemma 7

8.7.1 Preliminary results

First, recall the following results. Since the wavelet basis is orthonormal, the Parseval Equality obviously holds:

$$\forall j \geq 0, \quad \int h^2 \quad = \quad \sum_k \left(\int \phi_{jk} h \right)^2 + \sum_{j=j}^{\infty} \sum_k \left(\int \psi_{jk} h \right)^2. \tag{20}$$

Lemma 8 (Meyer) *If θ is a bounded 1-periodic function, and $h \in L^1(R)$, then:*

$$\int_R h(t)\theta(2^j t)dt \rightarrow_{j \longrightarrow \infty} \int_{[0,1]} \theta(t)dt \int_R h(t)dt$$

8.7.2 Evaluation of $E = \sum_{k_1 k_2} a_{k_1 k_2}^2$

Let h and h' be either f or g. Using Cauchy-Schwarz Inequality, Parseval Inequality and Meyer Lemma, we get

$$\begin{aligned}
\sum_k \left(\int \phi_{jk} h \right)^2 &\leq \|h\|_2^2, \\
\left| \sum_{k_1 k_2} \int \phi_{jk_1}\phi_{jk_2} h \int \phi_{jk_1}\phi_{jk_2} h' \right| &\leq \left(\sum_{k_1 k_2} \left(\int \phi_{jk_1}\phi_{jk_2} h \right)^2 \sum_{k_1 k_2} \left(\int \phi_{jk_1}\phi_{jk_2} h' \right)^2 \right)^{1/2} \\
&\leq \left(\sum_k \int \phi_{jk}^2 h^2 \sum_k \int \phi_{jk}^2 h'^2 \right)^{1/2} \leq 2^j \|h\|_2 \|h'\|_2 \\
\left| \sum_{k_1 k_2} \int \phi_{jk_1}\phi_{jk_2} h \int \phi_{jk_1} h \int \phi_{jk_2} h' \right| &\leq 2^{-j} \|h\|_\infty \|h'\|_\infty \left| \sum_{k_1 k_2} \int \phi_{jk_1}\phi_{jk_2} h \right| \leq \|h\|_\infty \|h'\|_\infty.
\end{aligned}$$

We conclude

$$E \leq 2^j \|h\|_2 \|h'\|_2.$$

8.7.3 Evaluation of $F = \sum_{k_1 k_2} (a_{k_1 k_2} b_{k_1} b_{k_2} - b_{k_1}^2 b_{k_2}^2)$

For any function $p(.)$, let us denote $B_p(x) = p(x) - \sum_k (\int \phi_{jk} p)\phi_{jk}(x)$ and $B_2(p)^2 = \int p^2 - \sum_k \left(\int \phi_{jk} p \right)^2$. Remark that $\int B_p^2 = B_2(p)^2$ and applying Lemma 1, $|B_2(p)| \leq M 2^{-js_p}$ as soon as $p(.) \in \mathcal{B}_{2,\infty}^{s_p}(M)$.

Let us denote $h = f - g$, $s_h = s_f \wedge s_g$, h' being either f or g. Remark that

$$F \quad = \quad F_f + F_g - 2F_{fg} - F_h$$

for

$$\begin{aligned}
F_{h'} &= \sum_{k_1 k_2} \int \phi_{jk_1}\phi_{jk_2} h' \int \phi_{jk_1} h \; int \phi_{jk_2} h \\
F_{f,g} &= \sum_{k_1 k_2} \int \phi_{jk_1} f \int \phi_{jk_2} g \int \phi_{jk_1} h \int \phi_{jk_2} h \\
F_h &= \left(\sum_k \left(\int \phi_{jk} h \right)^2 \right)^2,
\end{aligned}$$

h' being either f or g. We get

$$\begin{aligned}
F_{h'} &= \int \left(\sum_{k_1} \left(\int \phi_{jk_1} h \right) \phi_{jk_1} \right) \left(\sum_{k_2} \left(\int \phi_{jk_2} h \right) \phi_{jk_2} \right) h' \\
&= \int (h - B_h)^2 h' = \int h^2 h' + r_{h'}
\end{aligned} \tag{21}$$

where

$$
\begin{aligned}
|r_h| &= \left|-2\int B_h hh' + \int B_h^2 hh'\right| \leq B_2(h)\left(2\|hh'\|_2 + B_2(h)\|hh'\|_\infty\right) \\
&\leq c2^{-js_h}.
\end{aligned}
$$

Moreover

$$
\begin{aligned}
F_{fg} &= \int \left(\sum_{k_1}\left(\int \phi_{jk_1} f\right)\phi_{jk_1}\right) h \int \left(\sum_{k_2}\left(\int \phi_{jk_2} g\right)\phi_{jk_2}\right) h \\
&= \int (f - B_f)\, h \int (g - B_g)\, h = \int fh \int gh + r_{f,g} \qquad (22)
\end{aligned}
$$

where

$$
\begin{aligned}
|r_{f,g}| &= \left|-\int B_g fh - \int B_f^2 gh + \int B_f B_g h^2\right| \\
&\leq B_2(g)\|fh\|_2 + B_2(f)\|gh\|_2 + B_2(f)B_2(g)\|h\|_\infty^2 \leq c2^{-js_h}.
\end{aligned}
$$

Combining (21), (22) with

$$
F_h = \|h\|_2^4 + B_2(h),
$$

we deduce

$$
F = \int h^2 f + \int h^2 g - 2\int hf \int hg - \left(\int h^2\right)^2 + c2^{-js_h}.
$$

8.7.4 Evaluation of $B^2 = q\|Eu^2(Z,.)\|_\infty$

Under H_0, the densities f and g are equal; we denote h the common density. With the same argument as previously, we get

$$
\begin{aligned}
B^2 &= q\|\sum_{k_1k_2} c_{k_1k_2}\left(\phi_{jk_1}(x) - \phi_{jk_1}(y)\right)\left(\phi_{jk_2}(x) - \phi_{jk_2}(y)\right)\|_\infty \\
&\leq q\max_{k_1k_2}|c_{k_1k_2}|(2DB-1)2^j\, 4\|\phi\|_\infty^2
\end{aligned}
$$

for

$$
c_{12} = 2\left(\int \phi_{jk_1}\phi_{jk_2} h - 2\left(\int \phi_{jk_1} h\right)^2\right)
$$

Since there exists a constant c depending on $\|h\|_2, \|\phi\|_2$ such that $|c_{k_1k_2}| \leq c$, we get $B^2 \leq cq\,2^j$.

8.7.5 Evaluation of $C^2 = q^2Eu^2$

By Lemma 6, we have $C^2 = q^2 2^j$.

8.7.6 Evaluation of $D = q\sup_{u_1u_2} Eu(Z_1,Z_2)u_1(Z_1)u_2(Z_2)$

Denote h the joint density of Z: $h(x,y) = f(x)g(y)$.

$$
\begin{aligned}
C_{u_1u_2} &= Eu(Z_1,Z_2)u_1(Z_1)u_2(Z_2) \\
&\leq \sum_{l=1,2}\sum_{l'=1,2}\left|\sum_k\left(\int \phi_{jk}(x_l)u_{i_1}(x,y)h(x,y)dxdy\right)\left(\int \phi_{jk}(x_{l'})u_{i_1}(x,y)h(x,y)dxdy\right)\right|
\end{aligned}
$$

where $x_1 = x$ and $x_2 = y$. Moreover, for u being either u_1 or u_2, put

$$c_f(k) = \left| \int \phi_{jk}(x) u(x,y) h(x,y) dxdy \right|$$

and remark that

$$\begin{aligned} c_f(k) &\leq \left(\int \phi_{jk}^2 f \int u^2(x,y) 1_{[-\frac{k}{2^j}, \frac{2DB-k}{2^j}]}(x) h(x,y) dxdy \right)^{1/2} \\ &\leq \|f\|_\infty \left(\int u^2(x,y) 1_{[-\frac{k}{2^j}, \frac{2DB-k}{2^j}]}(x) h(x,y) dxdy \right)^{1/2} \end{aligned}$$

and then

$$\sum_k c_f(k)^2 \quad \leq \quad \|f\|_\infty \int u^2(x,y) h(x,y) dxdy. \tag{23}$$

In the same way

$$\sum_k c_g(k)^2 \quad \leq \quad \|g\|_\infty \int u^2(x,y) h(x,y) dxdy, \tag{24}$$

where

$$c_f(k) = \left| \int \phi_{jk}(y) u(x,y) h(x,y) dxdy \right|.$$

Using Cauchy-Schwarz Inequalities and (23), (24), we deduce that

$$C_{u_1 u_2} \leq (\|f\|_\infty + \|g\|_\infty)^2 \sum_{l=1,2} E u_l^2,$$

and we obtain the announced result.

8.8 Proof of Lemma 5

We note first that $Z_i = 1 + t(Y_i)$ where $t(x) = 2^{-js}\psi(2^j(x - x_0))/f(x)$. We get

$$\|t\|_\infty \quad \leq \quad 2^{-js}\|\psi\|_\infty / m_f, \quad E_f t(Y_1) = 0, \quad E_f |t(Y_1)|^k \leq m_f^{-k+1} 2^{-j(ks+1)} \|\psi\|_k^2,$$

for $k = 1, 2 \ldots$. Then, for j large enough, $\|t\|_\infty \leq 1$ and since

$$\forall |x| < 1, \, x - x^2 \leq \log(1+x) \leq x + x^2,$$

we obtain

$$\begin{aligned} \mu_m &= m E_f \log(1 + t(Y_1)) \geq m \left(E_f t(Y_1) - E_f t^2(Y_1) \right) \\ &\geq -m 2^{-j(2s+1)} m_f^{-1} \geq -m_f^{-1}. \end{aligned}$$

Similarly $\mu_m \leq m_f^{-1}$. For the variance

$$\begin{aligned} \sigma_m^2 &\leq m E_f \left(\log(1 + t(Y_1))^2 \right) \\ &\leq m E_f \left(t(Y_1)^2 (1 + t(Y_1))^2 \right) \\ &\leq m m_f^{-1} 2^{-j(2s+1)} \left(1 + 2^{-js} m_f^{-1} \|\psi\|_\infty \right)^2 \leq 2 m_f^{-1}. \end{aligned}$$

Similarly, $\sigma_m^2 \geq m_f^{-1}$. At last, the 3rd order moment is finite and since

$$\forall |x| < 1, |x| \leq |\log(1+x)| \leq |x| + x^2,$$

we have

$$
\begin{aligned}
& \sigma_m^{-3} \sum_{i=1}^{m} E_{h,h} |\log Z_i - E_{h,h} \log Z_i|^3 \\
\leq \; & 4m\sigma_m^{-3} \left[\quad E_{h,h} |\, \log(1+t(Y_1))\,|^3 \; + \; |\, E_{h,h} \log(1+t(Y_1))\,|^3 \quad \right] \\
\leq \; & 4m\sigma_m^{-3} \left[\quad E_{h,h} |t(Y_1)|^3 (1+|t(Y_1)|)^3 \; + \; (\, E_{h,h} |t(Y_1)|\,(1+|t(Y_1)|)\,)^3 \quad \right] \\
\leq \; & 4mm_f^3 \left(1 + 2^{-js} m_f^{-1} \|\psi\|_\infty\right)^3 \left(m_f^{-2} 2^{-j(3s+1)} \|\psi\|_3^2 + 2^{-3j(s+1)}\right) \\
\leq \; & 8mm_f \|\psi\|_3^2 \, 2^{-3j(s+1)}
\end{aligned}
$$

which converges to 0 and completes the proof of Lemma (5).