Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2009, Vol. 45, No. 2, 352–372
DOI: 10.1214/08-AIHP166

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Adaptive goodness-of-fit testing from indirect observations

Cristina Butucea[a], Catherine Matias[b] and Christophe Pouet[c]

[a]*Laboratoire Paul Painlevé (UMR CNRS 8524), Université des Sciences et Technologies de Lille 1, 59655 Villeneuve d'Ascq Cedex, France. E-mail: cristina.butucea@math.univ-lille1.fr*
[b]*Laboratoire Statistique et Génome (UMR CNRS 8071), Tour Evry 2, 523 pl. des Terrasses de l'Agora, 91 000 Evry, France. E-mail: cmatias@genopole.cnrs.fr*
[c]*Laboratoire d'Analyse, Topologie, Probabilités (UMR CNRS 6632), Centre de Mathématiques et Informatique, Université de Provence, 39 rue F. Joliot-Curie, 13453 Marseille cedex 13, France. E-mail: pouet@cmi.univ-mrs.fr*

Received 22 March 2007; revised 29 November 2007; accepted 30 January 2008

Abstract. In a convolution model, we observe random variables whose distribution is the convolution of some unknown density f and some known noise density g. We assume that g is polynomially smooth. We provide goodness-of-fit testing procedures for the test H_0: $f = f_0$, where the alternative H_1 is expressed with respect to $\mathbb{L}_2$-norm (i.e. has the form $\psi_n^{-2}\|f - f_0\|_2^2 \geq \mathcal{C}$). Our procedure is adaptive with respect to the unknown smoothness parameter τ of f. Different testing rates (ψ_n) are obtained according to whether f_0 is polynomially or exponentially smooth. A price for adaptation is noted and for computing this, we provide a non-uniform Berry–Esseen type theorem for degenerate U-statistics. In the case of polynomially smooth f_0, we prove that the price for adaptation is optimal. We emphasise the fact that the alternative may contain functions smoother than the null density to be tested, which is new in the context of goodness-of-fit tests.

Résumé. Dans un modèle de convolution, les observations sont des variables aléatoires réelles dont la distribution est la convoluée entre une densité inconnue f et une densité de bruit g supposée entièrement connue. Nous supposons que g est de régularité polynomiale. Nous proposons un test d'adéquation de l'hypothèse $H_0 : f = f_0$ lorsque l'alternative H_1 est exprimée à partir de la norme $\mathbb{L}_2$ (i.e. de la forme $\psi_n^{-2}\|f - f_0\|_2^2 \geq \mathcal{C}$). Cette procédure est adaptative par rapport au paramètre inconnu τ qui décrit la régularité de f. Nous obtenons différentes vitesses de test (ψ_n) en fonction du type de régularité de f_0 (polynomiale ou bien exponentielle). L'adaptativité induit une perte sur la vitesse de test, perte qui est calculée grâce à un théorème de type Berry–Esseen non-uniforme pour des U-statistiques dégénérées. Dans le cas d'une régularité polynomiale pour f_0, nous prouvons que cette perte est optimale. Soulignons que l'alternative peut éventuellement inclure des densités qui sont plus régulières que la densité à tester sous l'hypothèse nulle, ce qui est un point de vue nouveau pour les tests d'adaptation.

MSC: 62F12; 62G05; 62G10; 62G20

Keywords: Adaptive nonparametric tests; Convolution model; Goodness-of-fit tests; Infinitely differentiable functions; Partially known noise; Quadratic functional estimation; Sobolev classes; Stable laws

1. Introduction

Convolution model

Consider the convolution model where the observed sample $\{Y_j\}_{1\leq j\leq n}$ comes from the independent sum of independent and identically distributed (i.i.d.) random variables X_j and i.i.d. noise variables ε_j. Variables X_j have unknown density f and Fourier transform Φ (where $\Phi(u) = \int \exp(\mathrm{i}xu) f(x)\,\mathrm{d}x$) and the noise variables ε_j have known density g and Fourier transform Φ^g

$$Y_j = X_j + \varepsilon_j, \quad 1 \leq j \leq n. \tag{1}$$

The density of the observations is denoted by p and its Fourier transform Φ^p. Note that we have $p = f * g$, where $*$ denotes the convolution product and $\Phi^p = \Phi \cdot \Phi^g$.

The underlying unknown density f is always supposed to belong to $\mathbb{L}_1 \cap \mathbb{L}_2$. We shall consider probability density functions belonging to the class

$$\mathcal{F}(\alpha, r, \beta, L) = \left\{ f : \mathbb{R} \to \mathbb{R}_+, \int f = 1, \frac{1}{2\pi} \int \left|\Phi(u)\right|^2 |u|^{2\beta} \exp\bigl(2\alpha |u|^r\bigr) \,\mathrm{d}u \le L \right\}, \tag{2}$$

for L a positive constant, $\alpha > 0$, $0 \le r \le 2$, $\beta \ge 0$ and either $r > 0$ or $r = 0$ and then $\beta > 0$. Note that the case $r = 0$ corresponds to Sobolev densities whereas $r > 0$ corresponds to infinitely differentiable (or supersmooth) densities.

We consider noise distributions whose Fourier transform does not vanish on $\mathbb{R}$: $\Phi^g(u) \neq 0$, $\forall u \in \mathbb{R}$. Typically, in nonparametric estimation in convolution models the distinction of two different behaviours for the noise distribution occurs: for some constant $c_g > 0$, polynomially smooth noise

$$\left|\Phi^g(u)\right| \sim c_g |u|^{-\sigma}, \quad |u| \to \infty, \sigma > 1; \tag{3}$$

exponentially smooth noise $|\Phi^g(u)| \sim c_g \exp(-\gamma |u|^s)$, $|u| \to \infty$, $\gamma, s > 0$.

The exponentially smooth noise case is studied in a separate article and in a more general semiparametric framework [3].

There is a huge literature on convolution models published during the past two decades and focusing mainly on estimation problems. Our purpose here is to provide goodness-of-fit testing procedures on f, for the test of the hypothesis H_0: $f = f_0$, with alternatives expressed with respect to $\mathbb{L}_2$-norm, and being adaptive with respect to the unknown smoothness parameter of f.

Nonparametric goodness-of-fit testing has been studied extensively in the context of direct observations (namely a sample distributed from the density f to be tested), but also for regression or in the Gaussian white noise model. We refer to [11] for an overview on the subject. Analytic densities (namely densities in $\mathcal{F}(\alpha, r, \beta, L)$ with $r = 1$ and $\beta = 0$) first appeared in [13] who gives goodness-of-fit testing procedures with respect to pointwise and $\mathbb{L}_\infty$-risks in the Gaussian white noise model. Procedures with respect to $\mathbb{L}_2$-risk are given in [11] for Sobolev and analytic densities in the same model.

However, in the case of direct observations, there are few adaptive procedures. The pioneering work of [14] introduced adaptive testing procedures over scales of Besov classes and with respect to $\mathbb{L}_2$-risk. Let us also mention [4] and [5] for adaptive goodness-of-fit tests with a composite null hypothesis. Up to our knowledge, adaptive procedures do not exist in the case of indirect observations. The convolution model provides an interesting setup where observations may come from a signal observed through some noise.

There are two natural but very different approaches for the goodness-of-fit testing problem in the noisy setup. One can think of testing either the resulting density p or the initial density f. As density g is known, the null hypothesis H_0 may be expressed equivalently in the form $f = f_0$ or $p = p_0$. Moreover, testing p would result in better rates of testing than those obtained for f (as the convoluted density p is smoother than f). However, the alternative hypotheses in those two setups are not in a one-to-one correspondence. Here, we would like to emphasise that we only consider the latter problem of goodness-of-fit testing on f. Indeed, we think it more appropriate to express the alternatives by using the $\mathbb{L}_2$-distance between f and the null density f_0, which is always larger than the $\mathbb{L}_2$-distance between p and p_0. Moreover, there are cases where aspects of the underlying density f (apart from its smoothness) may be relevant to the statistician, like its modality, symmetry, monotonicity on some interval and these features may be strongly perturbed after convolution with some noise.

These two different points of view arise from a more general issue: how is the direct observations case related to the noisy one? As we want to focus on alternatives of the form $\psi_n^{-2} \| f - f_0 \|_2^2 \ge \mathcal{C}$ (rather than $\psi_n^{-2} \| p - p_0 \|_2^2 \ge \mathcal{C}$), results from the direct observations case cannot be used directly in our setting. Moreover, adaptivity of our procedure relies on the construction of a grid over the set of densities f. Then, using the corresponding grid on the set of densities p would not necessarily lead to an adaptive procedure.

However, one can compare the rates obtained in the two settings. Indeed, one can note that the rate we obtain for polynomially smooth densities in the alternative, namely $(n/\sqrt{\log\log n})^{-2\beta/(4\beta+4\sigma+1)}$, corresponds to the rate obtained by [14] in the Gaussian white noise setting, namely $(n/\sqrt{\log\log(n)})^{-2\beta/(4\beta+1)}$. Moreover, the rate we get for supersmooth densities in the alternative, namely $n^{-1/2}(\log n)^{(4\sigma+1)/(4r)}(\log\log\log n)^{1/4}$, shows an extra $\log\log\log n$

factor with respect to the non-adaptive result in [11], in the particular case of supersmooth densities with $r = 1$, namely $n^{-1/2}(\log n)^{1/4}$. Thus, we conjecture that the loss for adaptation on the direct observations case should be at most $(\log\log\log n)^{1/4}$. We deduce these rates when f_0 is smoother than the functions belonging to the alternative hypothesis, which is the usual setup for goodness-of-fit testing. Moreover, we also derive rates of testing when f_0 is less smooth than the functions belonging to the alternative hypothesis, which is a new setup. In the latter case, we observe that the testing rate is the minimax testing rate associated to the smoothness of f_0.

Nonparametric goodness-of-fit tests in convolution models were studied in [2] and in [9]. The approach used in [2] is based on a minimax point of view combined with estimation of the quadratic functional $\int f^2$. Assuming the smoothness parameter of f to be known, the authors of [9] define a version of the Bickel–Rosenblatt test statistic and study its asymptotic distribution under the null hypothesis and under fixed and local alternatives, while [2] provides a different goodness-of-fit testing procedure attaining the minimax rate of testing in each of the three following setups: Sobolev densities and polynomial noise, supersmooth densities and polynomial noise and Sobolev densities and exponential noise. The case of supersmooth densities and exponential noise is also studied but the optimality of the procedure is not established in the case $r > s$.

Our goal here is to provide adaptive versions of these last procedures with respect to the parameter $\tau = (\alpha, r, \beta)$. As we restrict our attention to testing problems where alternatives are expressed with respect to $\mathbb{L}_2$-norm (namely, the alternative has the form H_1: $\psi_n^{-2}\|f - f_0\|_2^2 \geq \mathcal{C}$), the problem is strongly related to asymptotically minimax estimation of $\int f^2$ and $\int (f - f_0)^2$. Our test statistic is based on a collection of kernel estimators of $\int (f - f_0)^2$ for convolution models, with a given set of regularity parameters τ. Then, adaptation to a scale of classes is obtained by rejecting the null hypothesis whenever at least one of the tests in the collection does, see for example [5].

Notation, definitions, assumptions

In the sequel, $\|\cdot\|_2$ denotes the $\mathbb{L}_2$-norm, $\bar{M}$ is the complex conjugate of M and $\langle M, N\rangle = \int M(x)\bar{N}(x)\,\mathrm{d}x$ is the scalar product of complex-valued functions in $\mathbb{L}_2(\mathbb{R})$. Moreover, probability and expectation with respect to the distribution of $Y_1, \ldots, Y_n$ induced by the unknown density f will be denoted by $\mathbb{P}_f$ and $\mathbb{E}_f$.

We denote more generally by $\tau = (\alpha, r, \beta)$ the smoothness parameter of the unknown density f and by $\mathcal{F}(\tau, L)$ the corresponding class. As the density f is unknown, the a priori knowledge of its smoothness parameter τ could appear unrealistic. Thus, we assume that τ belongs to a closed subset $\mathcal{T}$, included in $(0, +\infty) \times (0, 2] \times (0, +\infty)$. For a given density f_0 in the class $\mathcal{F}(\tau_0)$, we want to test the hypothesis

$$H_0\colon \quad f = f_0$$

from observations $Y_1, \ldots, Y_n$ given by (1). We extend the results of [2] by giving the family of sequences $\Psi_n = \{\psi_{n,\tau}\}_{\tau\in\mathcal{T}}$ which separates (with respect to $\mathbb{L}_2$-norm) the null hypothesis from a larger alternative

$$H_1(\mathcal{C}, \Psi_n)\colon \quad f \in \bigcup_{\tau\in\mathcal{T}} \bigl\{f \in \mathcal{F}(\tau, L) \text{ and } \psi_{n,\tau}^{-2}\|f - f_0\|_2^2 \geq \mathcal{C}\bigr\}.$$

We recall that the usual procedure is to construct, for any $0 < \epsilon < 1$, a test statistic $\Delta_n^\star$ (an arbitrary function, with values in $\{0, 1\}$, which is measurable with respect to $Y_1, \ldots, Y_n$ and such that we accept H_0 if $\Delta_n^\star = 0$ and reject it otherwise). We prove then that there exists some $\mathcal{C}^0 > 0$ such that

$$\limsup_{n\to\infty}\Bigl\{\mathbb{P}_0\bigl[\Delta_n^\star = 1\bigr] + \sup_{f\in H_1(\mathcal{C},\Psi_n)} \mathbb{P}_f\bigl[\Delta_n^\star = 0\bigr]\Bigr\} \leq \epsilon, \tag{4}$$

holds for all $\mathcal{C} > \mathcal{C}^0$. This part is called the upper bound of the testing rate. Then, we prove the minimax optimality of this procedure, i.e. the lower bound

$$\liminf_{n\to\infty}\inf_{\Delta_n}\Bigl\{\mathbb{P}_0[\Delta_n = 1] + \sup_{f\in H_1(\mathcal{C},\Psi_n)} \mathbb{P}_f[\Delta_n = 0]\Bigr\} \geq \epsilon, \tag{5}$$

for some $\mathcal{C}_0 > 0$ and for all $0 < \mathcal{C} < \mathcal{C}_0$, where the infimum is taken over all test statistics Δ_n.

Let us first remark that as we use noisy observations (and unlike what happens with direct observations), this test cannot be reduced to testing uniformity of the distribution density of the observed sample (i.e. $f_0 = 1$ with support on the finite interval [0; 1]). As a consequence, additional assumptions used in [2] on the tail behaviour of f_0 (ensuring it does not vanish arbitrarily fast) are needed to obtain the optimality result of the testing procedure in the case of Sobolev density ($r = 0$) observed with polynomial noise ((T) and (P)). We recall these assumptions here for reader's convenience.

Assumption (T). $\exists c_0 > 0, \forall x \in \mathbb{R}, f_0(x) \geq c_0(1 + |x|^2)^{-1}$.

Moreover, we also need to control the derivatives of known Fourier transform Φ^g when establishing optimality results.

Assumption (P) (Polynomial noise). *If the noise satisfies* (3), *then assume that* Φ^g *is three times continuously differentiable and there exist* A_1, A_2 *such that*

$$\left|\left(\Phi^g\right)'(u)\right| \leq \frac{A_1}{|u|^{\sigma+1}} \quad \text{and} \quad \left|\left(\Phi^g\right)''(u)\right| \leq \frac{A_2}{|u|^{\sigma+2}}, \quad |u| \to \infty.$$

Remark 1. *We can generalise Assumption* (T) *and assume the existence of some* $p \geq 1$ *such that* $f_0(x)$ *is bounded from below by* $c_0(1 + |x|^p)^{-2}$ *for large enough* x. *In such a case, we obtain the same results if the Fourier transform* Φ^g *of the noise density is assumed to be* p *times continuously differentiable, with derivatives up to order* p *satisfying the same kind of bounds as in Assumption* (P).

Let us give some comments on the proofs. In the case of Sobolev null density f_0, the fact that our testing procedure attains the minimax rate of testing (upper bound of the testing rate (4)), relies on a very sharp control on the approximation of the distribution of some U-statistic by the Gaussian distribution. Indeed, in our context, the classical approach using a central limit theorem is not sufficient, nor are the classical exponential inequalities on U-statistics (see for instance [6] or [10]). Thus, we had to establish a new Berry–Esseen inequality for degenerate U-statistics of order 2. We took into account the fact that in our case, as in most statistical problems, the function defining the U-statistic is depending on the number n of observations. This approach appeared to be powerful and is very promising to tackle other similar problems.

Concerning the minimax optimality of our procedure (lower bound of the testing rate (5), established for Sobolev null densities f_0), we used an approach proposed by [5] but had to combine it with the use of some specific kernel, previously introduced in [2].

Roadmap

In Section 2, we provide a goodness-of-fit testing procedure for the test H_0: $f = f_0$, in two different cases: the density f_0 to be tested is either ordinary smooth ($r_0 = 0$) or supersmooth ($r_0 > 0$). The procedures are adaptive with respect to the smoothness parameter (α, r, β) of f. The auxiliary result on a Berry–Esseen inequality for degenerate U-statistics of order 2 is described in Section 3. In some cases, a loss for adaptation is noted with respect to known testing rates for fixed known parameters. When the loss is of order $\log\log n$ to some power, we prove that this price is unavoidable. Proofs are postponed to Section 4.

2. Test procedures and main results

The unknown density f belongs to the class $\mathcal{F}(\alpha, r, \beta, L)$. We are interested in adaptive, with respect to the parameter $\tau = (\alpha, r, \beta)$, goodness-of-fit testing procedures. We assume that this unknown parameter belongs to the following set

$$\mathcal{T} = \left\{\tau = (\alpha, r, \beta); \tau \in [\underline{\alpha}; +\infty) \times [\underline{r}; \bar{r}] \times [\underline{\beta}; \bar{\beta}]\right\},$$

where $\underline{\alpha} > 0$, $0 \leq \underline{r} \leq \bar{r} \leq 2$, $0 \leq \underline{\beta} \leq \bar{\beta}$ and either $\underline{r} > 0$ and $\alpha \in [\underline{\alpha}, \bar{\alpha}]$ or both $\underline{r} = \bar{r} = 0$ and $\underline{\beta} > 0$.

Let us introduce some notation. We consider a preliminary kernel J, with Fourier transform Φ^J, defined by

$$\forall x \in \mathbb{R}, \quad J(x) = \frac{\sin(x)}{\pi x}, \qquad \forall u \in \mathbb{R}, \quad \Phi^J(u) = 1_{|u|\le 1},$$

where 1_A is the indicator function of the set A. For any bandwidth $h = h_n \to 0$ as n tends to infinity, we define the rescaled kernel J_h by

$$\forall x \in \mathbb{R}, \quad J_h(x) = h^{-1} J\left(\frac{x}{h}\right) \quad \text{and} \quad \forall u \in \mathbb{R}, \quad \Phi^{J_h}(u) = \Phi^J(hu) = 1_{|u|\le 1/h}.$$

Now, the deconvolution kernel K_h with bandwidth h is defined via its Fourier transform Φ^{K_h} as

$$\Phi^{K_h}(u) = \left(\Phi^g(u)\right)^{-1} \Phi^J(uh) = \left(\Phi^g(u)\right)^{-1} \Phi^{J_h}(u), \quad \forall u \in \mathbb{R}. \tag{6}$$

Next, the quadratic functional $\int (f - f_0)^2$ is estimated by the statistic $T_{n,h}$:

$$T_{n,h} = \frac{2}{n(n-1)} \sum\sum_{1\le k<j\le n} \left\langle K_h(\cdot - Y_k) - f_0, K_h(\cdot - Y_j) - f_0 \right\rangle. \tag{7}$$

Note that $T_{n,h}$ may take negative values, but its expected value is positive.

In order to construct a testing procedure which is adaptive with respect to the parameter τ we introduce a sequence of finite regular grids over the set $\mathcal{T}$ of unknown parameters: $\mathcal{T}_N = \{\tau_i; 1 \le i \le N\}$. For each grid point τ_i we choose a testing threshold $t_{n,i}^2$ and a bandwidth h_n^i giving a test statistic T_{n,h_n^i}.

The test rejects the null hypothesis as soon as at least one of the single tests based on the parameter τ_i is rejected.

$$\Delta_n^\star = \begin{cases} 1 & \text{if } \sup_{1\le i\le N} |T_{n,h_n^i}| t_{n,i}^{-2} > \mathcal{C}^\star, \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

for some constant $\mathcal{C}^\star > 0$ and finite sequences of bandwidths $\{h_n^i\}_{1\le i\le N}$ and thresholds $\{t_{n,i}^2\}_{1\le i\le N}$.

We note that our asymptotic results work for large enough constant $\mathcal{C}^\star$. In practice we may choose it by Monte Carlo simulation under the null hypothesis, for known f_0, such that we control the first-type error of the test and bound it from above, e.g. by $\epsilon/2$.

Typically, the structure of the grid accounts for two different phenomena. A first part of the points is dedicated to the adaptation with respect to β in case $\bar{r} = \underline{r} = 0$, whereas the rest of the points are used to adapt the procedure with respect to r, in case $\underline{r} > 0$ (whatever the value of β).

In the two next theorems, the parameter σ is fixed and defined in (3). We note that the testing procedures and the associated convergence rates are different according to whether the tested density f_0 (which is known) is polynomially or exponentially smooth. Therefore, we separate the two different cases where f_0 belongs to a Sobolev class ($r_0 = 0$, $\alpha_0 \ge \underline{\alpha}$ and we assume $\beta_0 = \bar{\beta}$) and where f_0 is a supersmooth function ($\alpha_0 \in [\underline{\alpha}, \bar{\alpha}]$, $r_0 > 0$ and $\beta_0 \in [\underline{\beta}, \bar{\beta}]$ and then we focus on $r_0 = \bar{r}$ and $\alpha_0 = \bar{\alpha}$). Note that in the first case, the alternative contains functions f which are smoother ($r > 0$) than the null hypothesis f_0. To our knowledge, this kind of result is new in goodness-of-fit testing.

When f_0 belongs to Sobolev class $\mathcal{F}(\alpha_0, 0, \bar{\beta}, L)$, the grid is defined as follows. Let N and choose $\mathcal{T}_N = \{\tau_i; 1 \le i \le N+1\}$ such that

$$\begin{cases} \forall 1 \le i \le N, \quad \tau_i = (0; 0; \beta_i) \quad \text{and} \quad \beta_1 = \underline{\beta} < \beta_2 < \cdots < \beta_N = \bar{\beta}, \\ \forall 1 \le i \le N-1, \quad \beta_{i+1} - \beta_i = (\bar{\beta} - \underline{\beta})/(N-1) \\ \text{and} \quad \tau_{N+1} = (\underline{\alpha}; \bar{r}; 0). \end{cases}$$

In this case, the first N points are dedicated to the adaptation with respect to β when $\bar{r} = \underline{r} = 0$, whereas the last point τ_{N+1} is used to adapt the procedure with respect to r (whatever the value of β).

Theorem 1. *Assume $f_0 \in \mathcal{F}(\alpha_0, 0, \bar{\beta}, L)$. The test statistic $\Delta_n^\star$ given by (8) with parameters*

$$N = \lceil \log n \rceil, \qquad \forall 1 \le i \le N: \quad \begin{cases} h_n^i = \big(\frac{n}{\sqrt{\log\log n}}\big)^{-2/(4\beta_i+4\sigma+1)}, \\ t_{n,i}^2 = \big(\frac{n}{\sqrt{\log\log n}}\big)^{-4\beta_i/(4\beta_i+4\sigma+1)}, \end{cases}$$

$$h_n^{N+1} = n^{-2/(4\bar{\beta}+4\sigma+1)}, \qquad t_{n,N+1}^2 = n^{-4\bar{\beta}/(4\bar{\beta}+4\sigma+1)},$$

and any large enough positive constant $\mathcal{C}^\star$, satisfies (4) for any $\epsilon \in (0,1)$, with testing rate $\Psi_n = \{\psi_{n,\tau}\}_{\tau \in \mathcal{T}}$ given by

$$\psi_{n,\tau} = \left(\frac{n}{\sqrt{\log\log n}}\right)^{-2\beta/(4\beta+4\sigma+1)} 1_{r=0} + n^{-2\bar{\beta}/(4\bar{\beta}+4\sigma+1)} 1_{r>0}, \quad \forall \tau = (\alpha, r, \beta) \in \mathcal{T}.$$

Moreover, if $f_0 \in \mathcal{F}(\alpha_0, 0, \bar{\beta}, cL)$ for some $0 < c < 1$ and if Assumptions (T) and (P) hold, then this testing rate is adaptive minimax over the family of classes $\{\mathcal{F}(\tau, L), \tau \in [\underline{\alpha}, \infty) \times \{0\} \times [\underline{\beta}, \bar{\beta}]\}$ (i.e. (5) holds).

We note that our testing procedure attains the polynomial rate $n^{-2\bar{\beta}/(4\bar{\beta}+4\sigma+1)}$ over the union of all classes containing functions smoother than f_0. Up to our knowledge, this phenomenon has never been identified in the literature. Note moreover that this rate is known to be a minimax testing rate over the class $\mathcal{F}(0, 0, \bar{\beta}, L)$ by results in [2]. Therefore we prove that the loss of some power of $\log\log n$ with respect to the minimax rate is unavoidable. A loss appears when the alternative contains classes of functions less smooth than f_0.

The proof that our adaptive procedure attains the above rate relies on the Berry–Esseen inequality presented in Section 3.

When f_0 belongs to class $\mathcal{F}(\bar{\alpha}, \bar{r}, \beta_0, L)$ of infinitely differentiable functions, the grid is defined as follows. Let N_1, N_2 and choose $\mathcal{T}_N = \{\tau_i; 1 \le i \le N = N_1 + N_2\}$ such that

$$\begin{cases} \forall 1 \le i \le N_1, \quad \tau_i = (0; 0; \beta_i) \quad \text{and} \quad \beta_1 = \underline{\beta} < \beta_2 < \cdots < \beta_{N_1} = \bar{\beta}, \\ \forall 1 \le i \le N_1 - 1, \quad \beta_{i+1} - \beta_i = (\bar{\beta} - \underline{\beta})/(N_1 - 1) \\ \text{and} \quad \forall 1 \le i \le N_2, \quad \tau_{N_1+i} = (\bar{\alpha}; r_i; \bar{\beta_0}) \quad \text{and} \quad r_1 = \underline{r} < r_2 < \cdots < r_{N_2} = \bar{r}, \\ \forall 1 \le i \le N_2 - 1, \quad r_{i+1} - r_i = (\bar{r} - \underline{r})/(N_2 - 1). \end{cases}$$

In this case, the first N_1 points are used for adaptation with respect to β in case $\bar{r} = \underline{r} = 0$, whereas the last N_2 points are used to adapt the procedure with respect to r (whatever the value of β).

Theorem 2. *Assume $f_0 \in \mathcal{F}(\bar{\alpha}, \bar{r}, \beta_0, L)$ for some $\beta_0 \in [\underline{\beta}, \bar{\beta}]$. The test statistic $\Delta_n^\star$ given by (8) with $\mathcal{C}^\star$ large enough and*

$$N_1 = \lceil \log n \rceil, \qquad \forall 1 \le i \le N_1: \quad \begin{cases} h_n^i = \big(\frac{n}{\sqrt{\log\log n}}\big)^{-2/(4\beta_i+4\sigma+1)}, \\ t_{n,i}^2 = \big(\frac{n}{\sqrt{\log\log n}}\big)^{-4\beta_i/(4\beta_i+4\sigma+1)}, \end{cases}$$

$$N_2 = \big\lceil \log\log n/(\bar{r} - \underline{r}) \big\rceil, \qquad \forall 1 \le i \le N_2: \quad \begin{cases} h_n^{N_1+i} = \big(\frac{\log n}{2c}\big)^{-1/r_i}, \quad c < \underline{\alpha}\exp\big(-\frac{1}{\underline{r}}\big), \\ t_{n,N_1+i}^2 = \frac{(\log n)^{(4\sigma+1)/(2r_i)}}{n}\sqrt{\log\log\log n}, \end{cases}$$

satisfies (4), with testing rate $\Psi_n = \{\psi_{n,\tau}\}_{\tau \in \mathcal{T}}$ given by

$$\psi_{n,\tau} = \left(\frac{n}{\sqrt{\log\log n}}\right)^{-2\beta/(4\beta+4\sigma+1)} 1_{r=0} + \frac{(\log n)^{(4\sigma+1)/(4r)}}{\sqrt{n}}(\log\log\log n)^{1/4} 1_{r \in [\underline{r}, \bar{r}]}.$$

We note that if Assumptions (T) and (P) hold for f_0 in $\mathcal{F}(\bar{\alpha}, \bar{r}, \beta_0, L)$, the same optimality proof as in Theorem 1 gives us that the loss of the $\log\log n$ to some power factor is optimal over alternatives in $\bigcup_{\alpha \in [\underline{\alpha}, \bar{\alpha}], \beta \in [\underline{\beta}, \bar{\beta}]} \mathcal{F}(\alpha, 0, \beta, L)$. A loss of a $(\log\log\log n)^{1/4}$ factor appears over alternatives of supersmooth densities (less smooth than f_0) with respect to the minimax rate in [2]. We do not prove that this loss is optimal.

3. Auxiliary result: Berry–Esseen inequality for degenerate U-statistics of order 2

This section is dedicated to the statement of a non-uniform Berry–Esseen type theorem for degenerate U-statistics. It draws its inspiration from [7] which provides a central limit theorem for degenerate U-statistics. Given a sample $Y_1, \dots, Y_n$ of i.i.d. random variables, we shall consider U-statistics of the form

$$U_n = \sum\sum_{1 \le i < j \le n} H(Y_i, Y_j),$$

where H is a symmetric function. We recall that degenerate U-statistic means

$$\mathbb{E}\{H(Y_1, Y_2)|Y_1\} = 0, \quad \text{almost surely.}$$

Thus, the statistic U_n is centered.

Limit theorems for degenerate U-statistics when H is fixed (independent of the sample size n) are well known and can be found in any monograph on the subject (see for instance [12]). In that case, the limit distribution is a linear combination of independent and centered $\chi^2(1)$ (chi-square with one degree of freedom) distributions. However, as noticed in [7], a normal distribution may result in some cases where H depends on n. In such a context, [7] provides a central limit theorem. But this result is not enough for our purpose (namely, optimality in Theorem 1). Indeed, we need to control the convergence to zero of the difference between the cumulative distribution function (cdf) of our U-statistic, and the cdf of the Gaussian distribution. Such a result may be derived using classical martingale methods.

In the rest of this section, n is fixed. Denote by $\mathcal{F}_i$ the σ-field generated by the random variables $\{Y_1, \dots, Y_i\}$. Define

$$v_n^2 = \mathbb{E}(U_n^2), \qquad Z_i = \frac{1}{v_n} \sum_{j=1}^{i-1} H(Y_i, Y_j), \quad 2 \le i \le n,$$

and note that as the U-statistic is degenerate, we have $\mathbb{E}(Z_i|Y_1, \dots, Y_{i-1}) = 0$. Thus,

$$S_k = \sum_{i=2}^{k} Z_i, \quad 2 \le k \le n,$$

is a centered martingale (with respect to the filtration $\{\mathcal{F}_k\}_{k\ge 2}$) and $S_n = v_n^{-1} U_n$. We use a non-uniform Berry–Esseen type theorem for martingales provided by [8], Theorem 3.9. Denote by ϕ the cdf of the standard normal distribution and introduce the conditional variance of the increments Z_j's,

$$V_n^2 = \sum_{i=2}^{n} \mathbb{E}(Z_i^2|\mathcal{F}_{i-1}) = \frac{1}{v_n^2} \sum_{i=2}^{n} \mathbb{E}\left\{\left(\sum_{j=1}^{i-1} H(Y_i, Y_j)\right)^2 \middle| \mathcal{F}_{i-1}\right\}.$$

Theorem 3. *Fix $0 < \delta \le 1$ and define*

$$L_n = \sum_{i=2}^{n} \mathbb{E}|Z_i|^{2+2\delta} + \mathbb{E}\left|V_n^2 - 1\right|^{1+\delta}.$$

There exists a positive constant C (depending only on δ) such that for any $0 < \epsilon < 1/2$ and any real x

$$\left|\mathbb{P}(U_n \le x) - \phi\left(\frac{x}{v_n}\right)\right| \le 16\epsilon^{1/2} \exp\left(-\frac{x^2}{4v_n^2}\right) + \frac{C}{\epsilon^{1+\delta}} L_n.$$

4. Proofs

We use C to denote an absolute constant, the values of which may change along the lines.

Proof of Theorem 1 (Upper bound). Let us give the sketch of proof concerning the upper-bound of the test. The statistic T_{n,h^i} will be abbreviated by $T_{n,i}$. We first need to control the first-type error of the test.

$$\begin{aligned}\mathbb{P}_0\big(\Delta_n^\star = 1\big) &= \mathbb{P}_0\big(\exists i \in \{1,\dots,N+1\} \text{ such that } |T_{n,i}| > \mathcal{C}^\star t_{n,i}^2\big)\\ &\leq \sum_{i=1}^{N+1} \mathbb{P}_0\big(\big|T_{n,i} - \mathbb{E}_0(T_{n,i})\big| > \mathcal{C}^\star t_{n,i}^2 - \mathbb{E}_0(T_{n,i})\big).\end{aligned}$$

The proof relies on the two following lemmas.

Lemma 1. *For any large enough $\mathcal{C}^\star > 0$, we have*

$$\sum_{i=1}^{N} \mathbb{P}_0\big(\big|T_{n,i} - \mathbb{E}_0(T_{n,i})\big| > \mathcal{C}^\star t_{n,i}^2 - \mathbb{E}_0(T_{n,i})\big) = \mathrm{o}(1).$$

Lemma 2. *For large enough $\mathcal{C}^\star$, there is some $\epsilon \in (0,1)$, such that*

$$\mathbb{P}_0\big(\big|T_{n,N+1} - \mathbb{E}_0(T_{n,N+1})\big| > \mathcal{C}^\star t_{n,N+1}^2 - \mathbb{E}_0(T_{n,N+1})\big) \leq \epsilon.$$

Lemma 1 relies on the Berry–Esseen type theorem (Theorem 3) presented in Section 3. Its proof is postponed to the end of the present section as the proof of Lemma 2.

Thus, the first type error term is as small as we need, as soon as we choose a large enough constant $\mathcal{C}^\star > 0$ in (8). We now focus on the second-type error of the test. We write

$$\begin{aligned}\sup_{\tau\in\mathcal{T}} \sup_{f\in\mathcal{F}(\tau,L)} \mathbb{P}_f\big(\Delta_n^\star = 0\big) &\leq 1_{\underline{r}>0} \sup_{r\in[\underline{r};\bar{r}],\alpha\geq\underline{\alpha},\beta\in[\underline{\beta},\bar{\beta}]} \sup_{\substack{f\in\mathcal{F}(\tau,L)\\ \|f-f_0\|_2^2\geq\mathcal{C}\psi_{n,\tau}^2}} \mathbb{P}_f\big(|T_{n,N+1}| \leq \mathcal{C}^\star t_{n,N+1}^2\big)\\ &\quad + 1_{\underline{r}=\bar{r}=0} \sup_{\alpha\geq\underline{\alpha},\beta\in[\underline{\beta};\bar{\beta}]} \sup_{\substack{f\in\mathcal{F}(\alpha,0,\beta,L)\\ \|f-f_0\|_2^2\geq\mathcal{C}\psi_{n,(\alpha,0,\beta)}^2}} \mathbb{P}_f\big(\forall 1\leq i\leq N, |T_{n,i}| \leq \mathcal{C}^\star t_{n,i}^2\big).\end{aligned}$$

Note that when the function f in the alternative is supersmooth ($\underline{r} > 0$), we only need the last test (with index $N+1$), whereas when it is ordinary smooth ($\underline{r} = \bar{r} = 0$), we use the family of tests with indexes $i \leq N$. In this second case, we use in fact only the test based on parameter β_f defined as the smallest point on the grid larger than β (see the proof of Lemma 4).

Lemma 3. *Fix $\underline{r} > 0$, for any $\alpha \geq \underline{\alpha}$, $r \in [\underline{r};\bar{r}]$, $\beta \in [\underline{\beta};\bar{\beta}]$. For any $\epsilon \in (0;1)$, there exists some large enough $\mathcal{C}^0$ such that for any $\mathcal{C} > \mathcal{C}^0$ and any $f \in \mathcal{F}(\alpha,r,\beta,L)$ such that $\|f-f_0\|_2^2 \geq \mathcal{C}\psi_{n,(\alpha,r,\beta)}$, we have*

$$\mathbb{P}_f\big(|T_{n,N+1}| \leq \mathcal{C}^\star t_{n,N+1}^2\big) \leq \epsilon.$$

Lemma 4. *We have*

$$\sup_{\alpha\geq\underline{\alpha}} \sup_{\beta\in[\underline{\beta};\bar{\beta}]} \sup_{\substack{f\in\mathcal{F}(\alpha,0,\beta,L)\\ \|f-f_0\|_2^2\geq\mathcal{C}\psi_{n,(\alpha,0,\beta)}^2}} \mathbb{P}_f\big(\forall 1\leq i\leq N, |T_{n,i}| \leq \mathcal{C}^\star t_{n,i}^2\big) = \mathrm{o}(1).$$

The proofs of Lemma 3 and Lemma 4 are postponed to the end of the present section. Thus, the second type error of the test converges to zero. This ends the proof of (4). □

Proof of Theorem 1 (Lower bound). As we already noted after the theorem statement, our test procedure attains the minimax rate associated to the class $\mathcal{F}(\alpha_0, 0, \bar{\beta}, L)$ where f_0 belongs, whenever the alternative f belongs to classes of functions smoother than f_0. Therefore, the lower bound we need to prove concerns the optimality of the loss of order $(\log\log n)^{1/2}$ due to alternatives less smooth than f_0.

More precisely, we prove (5), where the alternative $H_1(\mathcal{C}, \Psi_n)$ is now restricted to $\bigcup_{\beta\in[\underline{\beta},\bar{\beta}]}\{f \in \mathcal{F}(0,0,\beta,L)$ and $\psi_{n,\beta}^{-2}\|f - f_0\|_2^2 \geq \mathcal{C}\}$ and $\psi_{n,\beta}$ denotes the rate $\psi_{n,\tau}$ when $\tau = (0,0,\beta,L)$.

The general approach for proving such a lower bound (5) is to exhibit a finite number of regularities $\{\beta_k\}_{1\leq k\leq K}$ and corresponding probability distributions $\{\pi_k\}_{1\leq k\leq K}$ on the alternatives $H_1(\mathcal{C}, \psi_{n,\beta_k})$ (more exactly, on parametric subsets of these alternatives) such that the distance between the distributions induced by f_0 (the density being tested) and the mean distribution of the alternatives is small.

We use a finite grid $\bar{\mathcal{B}} = \{\beta_1 < \beta_2 < \cdots < \beta_K\} \subset [\underline{\beta}, \bar{\beta}]$ such that

$$\forall\beta \in [\underline{\beta}, \bar{\beta}],\ \exists k:\quad |\beta_k - \beta| \leq \frac{1}{\log n}.$$

To each point β in this grid, we associate a bandwidth

$$h_\beta = (n\rho_n)^{-2/(4\beta+4\sigma+1)}, \quad \rho_n = (\log\log n)^{-1/2} \quad \text{and} \quad M_\beta = h_\beta^{-1}.$$

We use the same deconvolution kernel as in [2], constructed as follows. Let G be defined as in Lemma 2 in [2]. The function G is an infinitely differentiable function, compactly supported on $[-1, 0]$ and such that $\int G = 0$. Then, the deconvolution kernel H_β is defined via its Fourier transform Φ^{H_β} by

$$\Phi^{H_\beta}(u) = \Phi^G(h_\beta u)\big(\Phi^g(u)\big)^{-1}.$$

Note that the factor ρ_n in the bandwidth's expression corresponds to the loss for adaptation.

We also consider for each β, a probability distribution π_β (also denoted π_k when $\beta = \beta_k$) defined on $\{-1,+1\}^{M_\beta}$ which is in fact the product of Rademacher distributions on $\{-1,+1\}$ and a parametric subset of $H_1(\mathcal{C}, \psi_{n,\beta})$ containing the following functions

$$f_{\theta,\beta}(x) = f_0(x) + \sum_{j=1}^{M_\beta} \theta_j h_\beta^{\beta+\sigma+1} H_\beta(x - x_{j,\beta}), \qquad \begin{cases} \theta_j \text{ i.i.d. with } \mathbb{P}(\theta_j = \pm 1) = 1/2, \\ x_{j,\beta} = j h_\beta \in [0,1]. \end{cases}$$

Convolution of these functions with g induces another parametric set of functions

$$p_{\theta,\beta}(y) = p_0(y) + \sum_{j=1}^{M_\beta} \theta_j h_\beta^{\beta+\sigma+1} G_\beta(y - x_{j,\beta}),$$

where $G_\beta(y) = h_\beta^{-1} G(y/h_\beta) = H_\beta * g(y)$.

As established in [2] (Lemmas 2 and 4), for any β, any $\theta \in \{-1,+1\}^{M_\beta}$ and small enough h_β (i.e. large enough n) the function $f_{\theta,\beta}$ is a probability density and belongs to the Sobolev class $\mathcal{F}(0,0,\beta,L)$ and $p_{\theta,\beta}$ is also a probability density. Moreover we have

$$\frac{1}{K}\sum_{\beta\in\bar{\mathcal{B}}} \pi_\beta\big(\|f_{\theta,\beta} - f_0\|_2^2 \geq \mathcal{C}\psi_{n,\beta}^2\big) \underset{n\to+\infty}{\longrightarrow} 1,$$

which means that for each β, the random parametric family $\{f_{\theta,\beta}\}_\theta$ belongs almost surely (with respect to the measure π_β) to the alternative set $H_1(\mathcal{C}, \psi_{n,\beta})$. The subset of functions which are not in the alternative $H_1(\mathcal{C}, \Psi_n)$ is asymptotically negligible. We then have,

$$
\begin{aligned}
\gamma_n &\triangleq \inf_{\Delta_n}\Big\{\mathbb{P}_0(\Delta_n=1)+\sup_{f\in H_1(\mathcal{C},\Psi_n)}\mathbb{P}_f(\Delta_n=0)\Big\}\\
&\geq \inf_{\Delta_n}\left\{\mathbb{P}_0(\Delta_n=1)+\frac{1}{K}\sum_{k=1}^{K}\sup_{f\in H_1(\mathcal{C},\psi_{n,\beta_k})}\mathbb{P}_f(\Delta_n=0)\right\}\\
&\geq \inf_{\Delta_n}\left\{\mathbb{P}_0(\Delta_n=1)+\frac{1}{K}\sum_{k=1}^{K}\left(\int_\theta \mathbb{P}_{f_{\theta,\beta_k}}(\Delta_n=0)\pi_k(\mathrm{d}\theta)-\pi_k\big(\|f_{\theta,\beta_k}-f_0\|_2^2<\mathcal{C}\psi_{n,\beta_k}^2\big)\right)\right\}\\
&\geq \inf_{\Delta_n}\left\{\mathbb{P}_0(\Delta_n=1)+\frac{1}{K}\sum_{k=1}^{K}\left(\int_\theta \mathbb{P}_{f_{\theta,\beta_k}}(\Delta_n=0)\pi_k(\mathrm{d}\theta)\right)\right\}+\mathrm{o}(1).
\end{aligned}
$$

Let us denote by

$$
\pi=\frac{1}{K}\sum_{k=1}^{K}\pi_k\quad\text{and}\quad \mathbb{P}_\pi=\frac{1}{K}\sum_{k=1}^{K}\mathbb{P}_k=\frac{1}{K}\sum_{k=1}^{K}\int_\theta \mathbb{P}_{f_{\theta,\beta_k}}\pi_k(\mathrm{d}\theta).
$$

Those notations lead to

$$
\begin{aligned}
\gamma_n &\geq \inf_{\Delta_n}\big\{\mathbb{P}_0(\Delta_n=1)+\mathbb{P}_\pi(\Delta_n=0)\big\}\\
&\geq \inf_{\Delta_n}\left\{1-\int_{\Delta_n=0}\mathrm{d}\mathbb{P}_0+\int_{\Delta_n=0}\mathrm{d}\mathbb{P}_\pi\right\}\geq 1-\sup_A\int_A(\mathrm{d}\mathbb{P}_0-\mathrm{d}\mathbb{P}_\pi)\\
&\geq 1-\frac{1}{2}\|\mathbb{P}_\pi-\mathbb{P}_0\|_1, \qquad (9)
\end{aligned}
$$

where we used Scheffé's lemma.

The finite grid $\bar{\mathcal{B}}$ is split into subsets $\bar{\mathcal{B}}=\bigcup_l \bar{\mathcal{B}}_l$ with $\bar{\mathcal{B}}_l\cap\bar{\mathcal{B}}_k=\emptyset$ when $l\neq k$ and such that

$$
\forall l,\quad \forall\beta_1\neq\beta_2\in\bar{\mathcal{B}}_l,\quad \frac{c\log\log n}{\log n}\leq|\beta_1-\beta_2|.
$$

The number of subsets $\bar{\mathcal{B}}_l$ is denoted by $K_1=\mathrm{O}(\log\log n)$ and the cardinality $|\bar{\mathcal{B}}_l|$ of each subset $\bar{\mathcal{B}}_l$ is of the order $\mathrm{O}(\log n/\log\log n)$, uniformly with respect to l.

The lower bound (5) is then obtained from (9) in the following way

$$
\gamma_n\geq 1-\frac{1}{2K_1}\sum_{l=1}^{K_1}\left\|\frac{1}{|\bar{\mathcal{B}}_l|}\sum_{\beta\in\bar{\mathcal{B}}_l}\mathbb{P}_\beta-\mathbb{P}_0\right\|_1,
$$

where $\mathbb{P}_\beta=\int_\theta \mathbb{P}_{f_{\theta,\beta}}\pi_\beta(\mathrm{d}\theta)$.

Here we do not want to apply the triangular inequality to the whole set of indexes $\bar{\mathcal{B}}$. Indeed, this would lead to a lower bound equal to 0. Yet, if we do not apply some sort of triangular inequality, we cannot deal with the sum because of too much dependency. This is why we introduced the subsets $\bar{\mathcal{B}}_l$ with the property that two points in the same subset $\bar{\mathcal{B}}_l$ are far enough away from each other. This technique was already used in [5] for the discrete regression model.

Let us denote by ℓ_β the likelihood ratio

$$
\ell_\beta=\frac{\mathrm{d}\mathbb{P}_\beta}{\mathrm{d}\mathbb{P}_0}=\int\frac{\mathrm{d}\mathbb{P}_{f_{\theta,\beta}}}{\mathrm{d}\mathbb{P}_0}\pi_\beta(\mathrm{d}\theta).
$$

We thus have

$$\gamma_n \geq 1 - \frac{1}{2K_1}\sum_{l=1}^{K_1}\int\left(\frac{1}{|\bar{\mathcal{B}}_l|}\sum_{\beta\in\bar{\mathcal{B}}_l}\ell_\beta - 1\right)\mathrm{d}\mathbb{P}_0 = 1 - \frac{1}{2K_1}\sum_{l=1}^{K_1}\left\|\frac{1}{|\bar{\mathcal{B}}_l|}\sum_{\beta\in\bar{\mathcal{B}}_l}\ell_\beta - 1\right\|_{\mathbb{L}_1(\mathbb{P}_0)}.$$

Now we use the usual inequality between $\mathbb{L}_1$ and $\mathbb{L}_2$-distances to get that

$$\gamma_n \geq 1 - \frac{1}{2K_1}\sum_{l=1}^{K_1}\left\|\frac{1}{|\bar{\mathcal{B}}_l|}\sum_{\beta\in\bar{\mathcal{B}}_l}\ell_\beta - 1\right\|_{\mathbb{L}_2(\mathbb{P}_0)} = 1 - \frac{1}{2K_1}\sum_{l=1}^{K_1}\left\{\mathbb{E}_0\left(\frac{1}{|\bar{\mathcal{B}}_l|}\sum_{\beta\in\bar{\mathcal{B}}_l}\ell_\beta - 1\right)^2\right\}^{1/2}.$$

Let us focus on the expected value appearing in the lower bound. We have

$$\mathbb{E}_0\left(\frac{1}{|\bar{\mathcal{B}}_l|}\sum_{\beta\in\bar{\mathcal{B}}_l}\ell_\beta - 1\right)^2 = \frac{1}{|\bar{\mathcal{B}}_l|^2}\sum_{\beta\in\bar{\mathcal{B}}_l}Q_\beta + \frac{1}{|\bar{\mathcal{B}}_l|^2}\sum_{\substack{\beta,\nu\in\bar{\mathcal{B}}_l\\ \beta\neq\nu}}Q_{\beta,\nu},$$

where there are two quantities to evaluate

$$Q_\beta = \mathbb{E}_0\big((\ell_\beta - 1)^2\big) \quad\text{and}\quad Q_{\beta,\nu} = \mathbb{E}_0(\ell_\beta\ell_\nu - 1).$$

The first term Q_β is treated as in [2]. It corresponds to the computation of a χ^2-distance between the two models induced by $\mathbb{P}_\beta$ and $\mathbb{P}_0$ (see term Δ^2 in [2]). Indeed we have

$$Q_\beta \leq C M_\beta n^2 h_\beta^{4\beta+4\sigma+2} \leq C\frac{1}{\rho_n^2}.$$

This upper bound goes to infinity very slowly. The number of β's in each $\bar{\mathcal{B}}_l$ compensates this behaviour

$$\frac{1}{|\bar{\mathcal{B}}_l|^2}\sum_{\beta\in\bar{\mathcal{B}}_l}Q_\beta \leq \frac{1}{|\bar{\mathcal{B}}_l|\rho_n^2} = \mathrm{O}\left(\frac{(\log\log n)^2}{\log n}\right) = \mathrm{o}(1).$$

The second term is a new one (with respect to non-adaptive case). As G is compactly supported and the points β and ν are far away from each other, we can prove that this term is asymptotically negligible. Recall the expression of the likelihood ratio for a fixed β

$$\ell_\beta = \int\frac{\mathrm{d}\mathbb{P}_{f_{\theta,\beta}}}{\mathrm{d}\mathbb{P}_0}\pi_\beta(\mathrm{d}\theta) = \int\prod_{r=1}^n\left(1 + \sum_{j=1}^{M_\beta}\theta_{j,\beta}h_\beta^{\beta+\sigma+1}\frac{G_\beta(Y_r - x_{j,\beta})}{p_0(Y_r)}\right)\pi_\beta(\mathrm{d}\theta).$$

Thus,

$$\begin{aligned}\ell_\beta\ell_\nu &= \int\frac{\mathrm{d}\mathbb{P}_{f_{\theta,\beta}}}{\mathrm{d}\mathbb{P}_0}\pi_\beta(\mathrm{d}\theta)\int\frac{\mathrm{d}\mathbb{P}_{f_{\theta,\nu}}}{\mathrm{d}\mathbb{P}_0}\pi_\nu(\mathrm{d}\theta)\\ &= \int\prod_{r=1}^n\left(1 + \sum_{j=1}^{M_\beta}\theta_{j,\beta}h_\beta^{\beta+\sigma+1}\frac{G_\beta(Y_r - x_{j,\beta})}{p_0(Y_r)}\right)\left(1 + \sum_{i=1}^{M_\nu}\theta_{i,\nu}h_\nu^{\nu+\sigma+1}\frac{G_\nu(Y_r - x_{i,\nu})}{p_0(Y_r)}\right)\pi_\beta(\mathrm{d}\theta_{\cdot,\beta})\pi_\nu(\mathrm{d}\theta_{\cdot,\nu}).\end{aligned}$$

The random variables Y_r are i.i.d. and $\mathbb{E}_0(\frac{G_\beta(Y_r - x_{j,\beta})}{p_0(Y_r)}) = 0$. Thus we have

$$\begin{aligned}\mathbb{E}_0(\ell_\beta\ell_\nu) = \int&\left[1 + \sum_{j=1}^{M_\beta}\sum_{i=1,i\subset j}^{M_\nu}\theta_{j,\beta}\theta_{i,\nu}h_\beta^{\beta+\sigma+1}h_\nu^{\nu+\sigma+1}\mathbb{E}_0\left(\frac{G_\beta(Y_1 - x_{j,\beta})G_\nu(Y_1 - x_{i,\nu})}{p_0^2(Y_1)}\right)\right]^n\\ &\times\pi_\beta(\mathrm{d}\theta_{\cdot,\beta})\pi_\nu(\mathrm{d}\theta_{\cdot,\nu}),\end{aligned}$$

where the second sum concerns only some indexes i, denoted by $i \subset j$. This notation stands for the set of indexes i such that $[(i-1)h_\beta; ih_\beta] \cap [(j-1)h_\nu; jh_\nu] \neq \emptyset$. From now on, we fix $\beta > \nu$. Denote by G' (resp. p_0') the first derivative of G (resp. p_0). (The density p_0 is continuously differentiable as it is the convolution product $f_0 * g$, where the noise density g is at least continuously differentiable.)

Lemma 5. *For any $\beta > \nu$ and any $(i, j) \in \{1, \dots, M_\nu\} \times \{1, \dots, M_\beta\}$, we have*

$$\mathbb{E}_0\left(\frac{G_\beta(Y_1 - x_{j,\beta})G_\nu(Y_1 - x_{i,\nu})}{p_0^2(Y_1)}\right) = \frac{h_\nu}{h_\beta^2} R_{i,j},$$

where R_{ij} satisfies

$$|R_{i,j}| \le \Bigl(\inf_{[0,1]} p_0\Bigr)^{-1} \|G\|_\infty \|G'\|_\infty \bigl(1 + \mathrm{o}(1)\bigr)$$

and $\mathrm{o}(1)$ *is uniform with respect to* (i, j).

The proof of this lemma is omitted. Applying Lemma 5, we get

$$Q_{\beta,\nu} + 1 = \int \left[1 + \sum_{j=1}^{M_\beta} \sum_{i=1, i\subset j}^{M_\nu} \theta_{j,\beta}\theta_{i,\nu} h_\beta^{\beta+\sigma+1} h_\nu^{\nu+\sigma+1} \frac{h_\nu}{(h_\beta)^2} R_{i,j}\right]^n \pi_\beta(\mathrm{d}\theta_{\cdot,\beta})\pi_\nu(\mathrm{d}\theta_{\cdot,\nu}).$$

Lemma 6. *Let U be a real valued random variable such that $\forall k \in \mathbb{N}, \mathbb{E}(U^{2k+1}) = 0$. We have, for any integer $n \ge 1$,*

$$\mathbb{E}(1 + U)^n \le 1 + \sum_{k=1}^{\lfloor n/2 \rfloor} \frac{n^{2k}}{(2k)!} \mathbb{E}\bigl(U^{2k}\bigr),$$

where $\lfloor x \rfloor$ is the largest integer which is smaller than x.

The proof of Lemma 6 is obvious and therefore omitted. Apply Lemma 6 to get the inequality

$$Q_{\beta,\nu} \le \sum_{k=1}^{\lfloor n/2 \rfloor} \frac{n^{2k}}{(2k)!} \bigl(h_\beta^{\beta+\sigma-1} h_\nu^{\nu+\sigma+2}\bigr)^{2k} \mathbb{E}_\pi \left(\sum_{j=1}^{M_\beta} \sum_{i=1, i\subset j}^{M_\nu} \theta_{j,\beta}\theta_{i,\nu} R_{i,j}\right)^{2k}.$$

But the θ's are i.i.d. Rademacher variables and the $R_{i,j}$'s are deterministic, thus

$$\mathbb{E}_\pi \left(\sum_{j=1}^{M_\beta} \sum_{i=1, i\subset j}^{M_\nu} \theta_{j,\beta}\theta_{i,\nu} R_{i,j}\right)^{2k} = \sum_{1 \le j_1, \dots, j_k \le M_\beta} \sum_{\substack{1 \le i_1, \dots, i_k \le M_\nu \\ \forall l, i_l \subset j_l}} \left(\prod_{l=1}^k R_{i_l, j_l}^2\right).$$

Using the bound on the $R_{i,j}$ given by Lemma 5,

$$\mathbb{E}_\pi \left(\sum_{j=1}^{M_\beta} \sum_{i=1, i\subset j}^{M_\nu} \theta_{j,\beta}\theta_{i,\nu} R_{i,j}\right)^{2k} \le \Bigl(\Bigl(\inf_{[0,1]} p_0\Bigr)^{-1} \|G\|_\infty \|G'\|_\infty \bigl(1 + \mathrm{o}(1)\bigr)\Bigr)^{2k} h_\nu^k.$$

Indeed, each index j_l may take at most $M_\beta = h_\beta^{-1}$ different values but the constraint $i_l \subset j_l$ implies that each index i_l

is limited to at most h_β/h_ν different values. Thus we get

$$
\begin{aligned}
Q_{\beta,\nu} &\le C \sum_{k=1}^{\lfloor n/2 \rfloor} \frac{n^{2k}}{(2k)!}\left(C h_\beta^{\beta+\sigma+1} h_\nu^{\nu+\sigma+1} \frac{h_\nu}{h_\beta^2}\right)^{2k} h_\nu^{-k} \\
&\le C \sum_{k=1}^{\lfloor n/2 \rfloor}\left(n^2 h_\beta^{2\beta+2\sigma+1/2} h_\nu^{2\nu+2\sigma+1/2} \frac{h_\nu^{5/2}}{h_\beta^{5/2}}\right)^k \le C \sum_{k=1}^{\lfloor n/2 \rfloor}\left(\frac{h_\nu^{5/2}}{\rho_n^2 h_\beta^{5/2}}\right)^k \le C \frac{1}{\rho_n^2}\frac{h_\nu^{5/2}}{h_\beta^{5/2}}.
\end{aligned}
$$

As $\beta > \nu$ both belong to some set $\bar{\mathcal{B}}_l$, we have $\beta - \nu \ge c(\log\log n)/(\log n)$ and according to the choice of the bandwidths,

$$
\begin{aligned}
\frac{h_\nu^{5/2}}{h_\beta^{5/2}} &= (n\rho_n)^{-20(\beta-\nu)/((4\beta+4\sigma+1)(4\nu+4\sigma+1))} \\
&\le \exp\left\{-\frac{20c\log\log n}{(4\bar{\beta}+4\sigma+1)^2}\bigl(1+\mathrm{o}(1)\bigr)\right\} \le (\log n)^{-w},
\end{aligned}
$$

where the constant w (depending on the constant c used in the construction of the sets $\bar{\mathcal{B}}_l$) can be tailored to our need. Therefore

$$
\frac{1}{|\bar{\mathcal{B}}_l|^2}\sum_{\substack{\beta,\nu\in|\bar{\mathcal{B}}_l| \\ \beta\ne\nu}} Q_{\beta,\nu} \le \frac{C}{\rho_n^2(\log n)^w}
$$

which goes to 0 as n goes to $+\infty$. We finally obtain the upper bound

$$
\mathbb{E}_0\left(\left(\frac{1}{|\bar{\mathcal{B}}_l|}\sum_{\beta\in|\bar{\mathcal{B}}_l|}\ell_\beta - 1\right)^2\right) \le \mathrm{O}\left(\frac{1}{|\bar{\mathcal{B}}_l|\rho_n^2}\right) + \mathrm{O}\left(\frac{1}{\rho_n^2(\log n)^w}\right) = \mathrm{o}(1),
$$

which leads to

$$
\gamma_n \ge 1 - \frac{1}{2}\frac{1}{K_1}\sum_{l=1}^{K_1}\left\{\mathrm{O}\left(\frac{1}{|\bar{\mathcal{B}}_l|\rho_n^2}\right) + \mathrm{O}\left(\frac{1}{\rho_n^2(\log n)^c}\right)\right\}^{1/2} = 1 + \mathrm{o}(1).
$$

□

Proof of Theorem 2. Assume now that $f_0 \in \mathcal{F}(\bar{\alpha}, \bar{r}, \beta_0, L)$, for some $\beta_0 \in [\underline{\beta}, \bar{\beta}]$. The proof follows the same lines as the proof of Theorem 1.

For the first-type error we write

$$
\begin{aligned}
\mathbb{P}_0\bigl(\Delta_n^* = 1\bigr) = &\sum_{i=1}^{N_1}\mathbb{P}_0\bigl(\bigl|T_{n,i} - \mathbb{E}_0(T_{n,i})\bigr| > \mathcal{C}^\star t_{n,i}^2 - \mathbb{E}_0(T_{n,i})\bigr) \\
&+ \sum_{i-N_1=1}^{N_2}\mathbb{P}_0\bigl(\bigl|T_{n,i} - \mathbb{E}_0(T_{n,i})\bigr| > \mathcal{C}^\star t_{n,i}^2 - \mathbb{E}_0(T_{n,i})\bigr).
\end{aligned}
$$

For the first N_1 terms we apply Lemma 1 with $\mathbb{E}_0(T_{n,i}) = \mathrm{o}(1)L(h_i)^{2\beta_0}\exp(-2\bar{\alpha}/h_i^{\bar{r}})$ which is smaller than $t_{n,i}^2$ for all $i = 1, \ldots, N_1$ and the same result follows. For the last N_2 terms we also use the Berry–Esseen inequality as in the proof of Lemma 1 for

$$
x = \mathcal{C}^\star t_{n,i}^2 - \mathbb{E}_0(T_{n,i}) \ge \mathcal{C}^\star t_{n,i}^2\bigl(1 - \mathrm{o}(1)\bigr)
$$

as $\mathbb{E}_0(T_{n,i}) = \mathrm{o}(1) h_i^{2\beta_0} \exp(-2\bar{\alpha}/h_i^{\bar{r}}) = \mathrm{o}(1/n)$. We get $x/v_n = \mathrm{O}(1)(\log\log\log n)^{1/2}$

$$\sum_{i-N_1=1}^{N_2} \mathbb{P}_0\bigl(\bigl|T_{n,i} - \mathbb{E}_0(T_{n,i})\bigr| > \mathcal{C}^\star t_{n,i}^2 - \mathbb{E}_0(T_{n,i})\bigr)$$

$$\leq N_2 \frac{v_n}{\mathcal{C}^\star t_{n,i}^2} \exp\biggl(-\frac{(\mathcal{C}^\star)^2 t_{n,i}^4}{4 v_n^2}\biggr) \leq C_1 \frac{(\log\log\log n)^{-1/2}}{(\log\log n)^{b-1}} = \mathrm{o}(1),$$

for some $b > 1$ for $\mathcal{C}^\star$ large enough. Indeed, all other calculations are similar as they are related mostly to the distribution of the noise which did not change.

As for the second-type error,

$$\sup_{\tau\in\mathcal{T}} \sup_{f\in\mathcal{F}(\tau,L)} \mathbb{P}_f\bigl(\Delta_n^\star = 0\bigr)$$

$$\leq 1_{\underline{r}=\bar{r}=0} \sup_{\alpha\geq\underline{\alpha},\beta\in[\underline{\beta};\bar{\beta}]} \sup_{\substack{f\in\mathcal{F}(\alpha,0,\beta,L)\\ \|f-f_0\|_2^2\geq\mathcal{C}\psi_{n,(\alpha,0,\beta)}^2}} \mathbb{P}_f\bigl(\forall 1\leq i\leq N_1, |T_{n,i}|\leq \mathcal{C}^\star t_{n,i}^2\bigr)$$

$$+ 1_{\underline{r}>0} \sup_{r\in[\underline{r};\bar{r}],\alpha\in[\underline{\alpha},\bar{\alpha}],\beta\in[\underline{\beta},\bar{\beta}]} \sup_{\substack{f\in\mathcal{F}(\tau,L)\\ \|f-f_0\|_2^2\geq\mathcal{C}\psi_{n,\tau}^2}} \mathbb{P}_f\bigl(\forall N_1+1\leq i\leq N_1+N_2, |T_{n,i}|\leq \mathcal{C}^\star t_{n,i}^2\bigr).$$

For the first term in the previous sum we actually apply precisely Lemma 4. For the second term we mimic the proof of Lemma 4 and choose some f in $\mathcal{F}(\alpha, r, \beta, L)$ such that $\|f - f_0\|_2^2 \geq \mathcal{C}\psi_{n,r}^2$, where we denote $\psi_{n,r} = \psi_{n,\tau} 1_{r>0}$. We define r_f as the smallest point on the grid $\{r_1, \ldots, r_{N_2}\}$ such that $r \leq r_f$. We denote by h_f, $t_{n,f}^2$ and $T_{n,f}$ the bandwidth, the threshold and the test statistic associated to parameters $\bar{\alpha}$ and r_f (they do not depend on β). Then

$$\mathbb{P}_f\bigl(\forall N_1+1\leq i\leq N_1+N_2, |T_{n,i}|\leq \mathcal{C}^\star t_{n,i}^2\bigr)$$
$$\leq \mathbb{P}_f\bigl(\bigl|T_{n,f} - \mathbb{E}_f(T_{n,f})\bigr| \geq \|f - f_0\|_2^2 - \mathcal{C}^\star t_{n,f}^2 - B_f(T_{n,f})\bigr), \tag{10}$$

where, as in Theorem 1

$$\bigl|B_f(T_{n,f})\bigr| = \bigl|\|J_h * f - f\|_2^2 + 2\langle f - J_h * f, f_0\rangle\bigr|$$
$$\leq \biggl(L h_f^{2\beta} \exp\biggl(-\frac{2\alpha}{h_f^r}\biggr) + 2L h_f^{\beta+\beta_0} \exp\biggl(-\frac{\alpha}{h_f^r} - \frac{\bar{\alpha}}{h_f^{\bar{r}}}\biggr)\biggr)\bigl(1+\mathrm{o}(1)\bigr)$$
$$\leq L\bigl(h_f^{2\beta} + h_f^{\beta+\beta_0}\bigr) \exp\biggl(-\frac{2\alpha}{h_f^r}\biggr)\bigl(1+\mathrm{o}(1)\bigr)$$
$$\leq L\bigl(h_f^{\beta+\beta\wedge\beta_0}\bigr) \exp\biggl(-\frac{2\alpha}{h_f^r}\biggr)\bigl(1+\mathrm{o}(1)\bigr).$$

Using Markov's inequality, we get the following upper bound for (10)

$$\frac{\mathrm{Var}_f(T_{n,f})}{(\|f - f_0\|_2^2 - \mathcal{C}^\star t_{n,f}^2 - B_f(T_{n,f}))^2}. \tag{11}$$

The variance is bounded from above by

$$\mathbb{E}_f\bigl(T_{n,f} - \mathbb{E}_f(T_{n,f})\bigr)^2 \leq \frac{C}{n^2 h_f^{4\sigma+1}} + \frac{4\Omega_g^2(f - f_0)}{n}, \tag{12}$$

and similarly to [2] we show that $\Omega_g^2(f-f_0) \le \|f-f_0\|_2^2(\log\|f-f_0\|_2^{-2})^{2\sigma/r}$. We have

$$t_{n,f}^2\psi_{n,r}^{-2} = (\log n)^{(4\sigma+1)(1/r_f-1/r)/2} \le 1,$$

and thus $\|f-f_0\|_2^2 - \mathcal{C}^\star t_{n,f}^2 \ge (\mathcal{C}-\mathcal{C}^\star)\psi_{n,r}^2$. Moreover,

$$B_f(T_{n,f})\psi_{n,r}^{-2} \le C(\log\log\log n)^{-1/2}(\log n)^{-(\beta+\beta\wedge\beta_0)/r_f-(4\sigma+1)/(2r)}\exp\left\{-2\alpha\left(\frac{\log n}{2c}\right)^{r/r_f}+\log n\right\}.$$

The construction of the grid ensures that $-1/(\log\log n) \le r-r_f \le 0$ and thus

$$\begin{aligned}\exp\left\{-2\alpha\left(\frac{\log n}{2c}\right)^{r/r_f}+\log n\right\} &= \exp\left\{-\frac{\log n}{c}\left[\alpha\exp\left(\frac{r-r_f}{r_f}\log\log n\big(1+\mathrm{o}(1)\big)\right)-c\right]\right\}\\ &\le \exp\left\{-\frac{\log n}{c}\left[\underline{\alpha}\exp\left(\frac{-1}{\underline{r}}\big(1+\mathrm{o}(1)\big)\right)-c\right]\right\} = \mathrm{O}(1),\end{aligned}$$

as we chose the constant $c<\underline{\alpha}\exp(-1/\underline{r})$. Finally, we have $B_f(T_{n,f})\psi_{n,r}^{-2} = \mathrm{o}(1)$. Let us come back to (11). We distinguish two cases whether the first or the second term in (12) is dominant. If the first term in the variance dominates, we have the following bound for (11)

$$\frac{n^{-2}h_f^{-(4\sigma+1)}}{(\mathcal{C}-\mathcal{C}^\star)^2\psi_{n,\tau}^4} \le \frac{C}{\log\log\log n} \to 0.$$

On the other hand, if the second term in (12) is the larger one, the bound (11) writes

$$\begin{aligned}\frac{n^{-1}\|f-f_0\|_2^2(\log\|f-f_0\|_2^{-2})^{2\sigma/r}}{\|f-f_0\|_2^4(1-\mathcal{C}^\star/\mathcal{C}+\mathrm{o}(1))^2} &\le Cn^{-1}\psi_{n,r}^{-2}(\log\psi_{n,r}^{-2})^{2\sigma/r}\\ &= C(\log n)^{-1/(2r)}(\log\log\log n)^{-1/2} = \mathrm{o}(1).\end{aligned}$$

This finishes the proof. □

Proof of Theorem 3. This proof follows the lines of Theorem 3.9 in [8]. Combining the Skorokhod representation theorem and Lemma 3.3 in [8], there exists a nonnegative random variable T_n such that for any $0<\epsilon<1/2$ and any real x,

$$\big|\mathbb{P}(U_n\le x)-\phi(x)\big| = \left|\mathbb{P}\big(S_n\le v_n^{-1}x\big)-\phi\left(\frac{x}{v_n}\right)\right| \le 16\epsilon^{1/2}\exp\left\{-\frac{x^2}{4v_n^2}\right\}+\mathbb{P}\big(|T_n-1|>\epsilon\big).$$

Moreover, for any $\delta>0$,

$$\mathbb{P}\big(|T_n-1|>\epsilon\big) \le 4\epsilon^{-1-\delta}\mathbb{E}\big[\big|T_n-V_n^2\big|^{1+\delta}+\big|V_n^2-1\big|^{1+\delta}\big],$$

where $T_n-V_n^2$ is a sum of Martingale differences. In the same way as in [8], we obtain (as $\delta\le 1$)

$$\mathbb{P}\big(|T_n-1|>\epsilon\big) \le C\epsilon^{-1-\delta}\left[\sum_{i=1}^n\mathbb{E}|Z_i|^{2+2\delta}+\mathbb{E}\big|V_n^2-1\big|^{1+\delta}\right],$$

which concludes the proof. □

We now present the proofs of the lemmas.

Proof of Lemma 1. Let us set $\rho_n = (\log\log n)^{-1/2}$ and fix $1 \le i \le N$. We use the obvious notation $p_0 = f_0 * g$. As we have

$$\mathbb{E}_0(T_{n,i}) = \|K_{h^i} * p_0 - f_0\|_2^2 = \|J_{h^i} * f_0 - f_0\|_2^2$$

and

$$\langle K_h(\cdot - Y_1) - J_h * f_0, J_h * f_0 - f_0 \rangle = 0$$

we easily get

$$T_{n,i} - \mathbb{E}_0(T_{n,i}) = \frac{2}{n(n-1)} \sum\sum_{1 \le k < j \le n} \langle K_{h^i}(\cdot - Y_k) - J_{h^i} * f_0, K_{h^i}(\cdot - Y_j) - J_{h^i} * f_0 \rangle.$$

Let us set

$$H(Y_j, Y_k) = 2\{n(n-1)\}^{-1} \langle K_{h^i}(\cdot - Y_k) - J_{h^i} * f_0, K_{h^i}(\cdot - Y_j) - J_{h^i} * f_0 \rangle$$

and note that H is a symmetric function with $\mathbb{E}_0\{H(Y_1, Y_2)\} = 0$ and $\mathbb{E}_0\{H(Y_1, Y_2)|Y_1\} = 0$. As a consequence, $T_{n,i} - \mathbb{E}_0(T_{n,i})$ is a degenerate U-statistic. Using Theorem 3 (and the notation of Section 3) to control its cdf, we get that for any $0 < \delta \le 1$, for any $0 < \varepsilon < 1/2$ and any x

$$\begin{aligned}
&\left| \mathbb{P}_0\big(T_{n,i} - \mathbb{E}_0(T_{n,i}) > x\big) - \left(1 - \phi\left(\frac{x}{v_n}\right)\right) \right| \\
&\quad \le 16\varepsilon^{1/2} \exp\left(-\frac{x^2}{4v_n^2}\right) + \frac{C}{\varepsilon^{1+\delta}} \left\{ \sum_{i=2}^n \mathbb{E}_0|Z_i|^{2+2\delta} + \mathbb{E}_0\left|V_n^2 - 1\right|^{1+\delta} \right\},
\end{aligned}$$

where $v_n^2 = \mathrm{Var}_0(T_{n,i})$ and

$$Z_i = \frac{1}{v_n} \sum_{j=1}^{i-1} H(Y_i, Y_j) \quad \text{and} \quad V_n^2 = \sum_{i=2}^n \mathbb{E}_0\big(Z_i^2|\mathcal{F}_{i-1}\big)$$

as in Section 3. Choose $\delta = 1$ and consider ε as a constant (optimisation in ε is not necessary in our context), thus

$$\left| \mathbb{P}_0\big(T_{n,i} - \mathbb{E}_0(T_{n,i}) > x\big) - \left(1 - \phi\left(\frac{x}{v_n}\right)\right) \right| \le C \exp\left(-\frac{x^2}{4v_n^2}\right) + C \left\{ \sum_{i=2}^n \mathbb{E}_0|Z_i|^4 + \mathbb{E}_0\left|V_n^2 - 1\right|^2 \right\}. \tag{13}$$

We want to apply this inequality at the point $x = \mathcal{C}^\star t_{n,i}^2 - \mathbb{E}_0(T_{n,i})$. First, note that

$$\mathbb{E}_0(T_{n,i}) = \|J_{h^i} * f_0 - f_0\|_2^2 = \frac{1}{2\pi} \int_{|u| > 1/(h^i)} \left|\Phi_0(u)\right|^2 \mathrm{d}u \le L\big(h^i\big)^{2\bar{\beta}} \le L t_{n,i}^2,$$

leading to

$$x \ge \big(\mathcal{C}^\star - L\big) t_{n,i}^2 = \big(\mathcal{C}^\star - L\big)(n\rho_n)^{-4\beta_i/(4\beta_i + 4\sigma + 1)}$$

and we choose $\mathcal{C}^\star > L$. Now, the variance term v_n^2 satisfies (see [2])

$$v_n^2 = \mathbb{E}_0\big(T_{n,i} - \mathbb{E}_0(T_{n,i})\big)^2 = \frac{C}{n^2 (h^i)^{4\sigma+1}} \big(1 + \mathrm{o}(1)\big).$$

Using the choice of the bandwidth h^i, we obtain a bound of the first term in (13)

$$C\exp\left(-\frac{x^2}{4v_n^2}\right) \le C\exp\left(-\frac{(\mathcal{C}^\star)^2}{C'}\rho_n^{-2}\right) = C(\log n)^{-b},$$

where $b = (\mathcal{C}^\star)^2/(C')$ can be chosen as large as we need. Let us deal with the other terms appearing in (13). For large enough n,

$$\left|\left\langle K_{h^i}(\cdot - Y_k) - J_{h^i} * f_0, K_{h^i}(\cdot - Y_j) - J_{h^i} * f_0\right\rangle\right| \le \frac{2}{\pi}\int_{|u|\le 1/h^i}\left|\Phi^g(u)\right|^{-2}\mathrm{d}u \le \frac{C}{(h^i)^{2\sigma+1}}$$

and thus, for any $p \ge 2$,

$$\mathbb{E}_0\left\{\left|H(Y_1, Y_2)\right|^{2p}\right\} \le Cn^{-4p}\left(h^i\right)^{-2p(2\sigma+1)}.$$

This leads to

$$\begin{aligned}
\sum_{i=2}^n \mathbb{E}_0|Z_i|^4 &\le \frac{1}{v_n^4}\sum_{i=2}^n\left(\sum_{j=1}^{i-1}\mathbb{E}_0\left(H(Y_i, Y_j)^4\right) + 3\sum\sum_{1\le j\ne k\le i-1}\mathbb{E}_0\left(H(Y_i, Y_j)^2 H(Y_i, Y_k)^2\right)\right)\\
&\le \frac{1}{v_n^4}\sum_{i=2}^n\left((i-1)\mathbb{E}_0\left(H(Y_1, Y_2)^4\right) + 3(i-1)(i-2)\mathbb{E}_0\left(H(Y_1, Y_2)^2 H(Y_1, Y_3)^2\right)\right)\\
&\le \frac{\mathrm{O}(1)}{v_n^4}n^2\mathbb{E}_0\left(H(Y_1, Y_2)^4\right) + \frac{\mathrm{O}(1)}{v_n^4}n^3\mathbb{E}_0\left(H(Y_1, Y_2)^2 H(Y_1, Y_3)^2\right)\\
&\le \mathrm{O}(1)\frac{n^3}{n^8(h^i)^{4(2\sigma+1)}}n^4\left(h^i\right)^{2(4\sigma+1)} = \frac{\mathrm{O}(1)}{n(h^i)^2}.
\end{aligned}$$

Moreover, following the lines of the proof of Theorem 1 in [7] we get

$$\mathbb{E}_0\left|V_n^2 - 1\right|^2 \le \frac{1}{v_n^4}\left(\mathbb{E}_0\left(G^2(Y_1, Y_2)\right) + \frac{1}{n}\mathbb{E}_0\left(H^4(Y_1, Y_2)\right)\right),$$

where $G(x, y) = \mathbb{E}_0(H(Y_1, x)H(Y_1, y))$. In [1] this last term was bounded from above for this model by Ch^i so

$$\mathbb{E}_0\left|V_n^2 - 1\right|^2 \le Ch^i.$$

Returning to (13) we finally get for $x = \mathcal{C}^\star t_{n,i}^2 - \mathbb{E}_0(T_{n,i})$,

$$\left|\mathbb{P}_0\left(T_{n,i} - \mathbb{E}_0(T_{n,i}) > x\right) - \left\{1 - \phi\left(\frac{x}{v_n}\right)\right\}\right| \le C\left((\log n)^{-b} + h^i\right) \le C(\log n)^{-b}.$$

Finally we obtain, for b large when $\mathcal{C}^\star$ is large

$$\begin{aligned}
&\sum_{i=1}^N \mathbb{P}_0\left(\left|T_{n,i} - \mathbb{E}_0(T_{n,i})\right| > \mathcal{C}^\star t_{n,i}^2 - \mathbb{E}_0(T_{n,i})\right)\\
&\quad\le N\left(1 - \phi\left(\frac{x}{v_n}\right) + C(\log n)^{-b}\right) \le CN\left(v_n x^{-1}\exp\left(-\frac{x^2}{2v_n^2}\right) + (\log n)^{-b}\right)\\
&\quad\le CN\rho_n(\log n)^{-b} \le C\frac{(\log\log n)^{-1/2}}{\log n^{b-1}}.
\end{aligned}$$

□

Proof of Lemma 2. Using a Markov inequality and the usual controls on bias and variance, we get

$$\mathbb{P}_0\big(\big|T_{n,N+1}-\mathbb{E}_0(T_{n,N+1})\big| > \mathcal{C}^\star t_{n,N+1}^2-\mathbb{E}_0(T_{n,N+1})\big) \le \frac{Cn^{-2}(h^{N+1})^{-(4\sigma+1)}}{(\mathcal{C}^\star t_{n,N+1}^2 - C(h^{N+1})^{2\bar\beta})^2}$$

which is $\mathrm{O}((\mathcal{C}^\star - C)^{-1})$ and by choosing $\mathcal{C}^\star$ large enough, this term is smaller than some $\epsilon > 0$. □

Proof of Lemma 3. Let us write

$$\mathbb{P}_f\big(|T_{n,N+1}| \le \mathcal{C}^\star t_{n,N+1}^2\big) \le \mathbb{P}_f\big(|T_{n,N+1}-\mathbb{E}_f T_{n,N+1}| \ge \|f-f_0\|_2^2 - \mathcal{C}^\star t_{n,N+1}^2 - B_f(T_{n,N+1})\big),$$

where

$$\begin{aligned}\big|B_f(T_{n,N+1})\big| &= \big|\mathbb{E}_f(T_{n,N+1}) - \|f-f_0\|_2^2\big|\\ &\le \int_{|u|\ge 1/h^{N+1}} \big|\Phi(u)\big|^2\,\mathrm{d}u + 2\bigg(\int_{|u|\ge 1/h^{N+1}}\big|\Phi(u)\big|^2\,\mathrm{d}u \int_{|u|\ge 1/h^{N+1}}\big|\Phi_0(u)\big|^2\,\mathrm{d}u\bigg)^{1/2}\\ &\le \bigg(L\big(h^{N+1}\big)^{2\beta}\exp\bigg\{-\frac{2\alpha}{(h^{N+1})^r}\bigg\} + 2L\big(h^{N+1}\big)^{\beta+\bar\beta}\exp\bigg\{-\frac{\alpha}{(h^{N+1})^r}\bigg\}\bigg)\\ &\le 2L\big(h^{N+1}\big)^{\beta+\bar\beta}\exp\bigg\{-\frac{\alpha}{(h^{N+1})^r}\bigg\}\big(1+\mathrm{o}(1)\big).\end{aligned}$$

In the same way as in the proof of Lemma 4, we have

$$\mathbb{E}_f\big(T_{n,N+1}-\mathbb{E}_f(T_{n,N+1})\big)^2 \le \frac{C}{n^2(h^{N+1})^{4\sigma+1}} + \frac{4\Omega_g^2(f-f_0)}{n}1_{\beta\ge\sigma} = w_{n,f}^2,$$

and $\Omega_g(f-f_0)$ is a constant depending on f and g (but not n) and satisfying $|\Omega_g^2(f-f_0)| \le C\|f-f_0\|_2^{2-2\sigma/\bar\beta}$. The rest of the proof follows the same lines as Lemma 4. Indeed, Markov's inequality leads the following bound on the second-type error term

$$\begin{aligned}&\frac{w_{n,f}^2}{(\|f-f_0\|_2^2 - \mathcal{C}^\star t_{n,N+1}^2 - 2L(h^{N+1})^{2\beta}\exp\{-\alpha/(h^{N+1})^r\}(1+\mathrm{o}(1)))^2}\\ &\quad\le \max\bigg(\frac{Cn^{-2}(h^{N+1})^{-4\sigma-1}}{(\mathcal{C}^0-\mathcal{C}^\star)^2\psi_{n,r}^4};\frac{C}{n\|f-f_0\|_2^{2+2\sigma/\bar\beta}(\mathcal{C}^0-\mathcal{C}^\star)^2}\bigg).\end{aligned}$$

The first term in the right-hand side is a constant which can be as small as we need, by choosing a large enough constant $\mathcal{C}^0$. The second term converges to zero. □

Proof of Lemma 4. When $\bar r = \underline{r} = 0$, let us fix some constant $\mathcal{C} > \mathcal{C}^0$ ($\mathcal{C}^0$ will be chosen later) and a density f belonging to $\mathcal{F}(\alpha,0,\beta,L)$ for some unknown $\alpha > \underline{\alpha}$ and $\beta \in [\underline{\beta};\bar\beta]$ which satisfies $\|f-f_0\|_2^2 \ge \mathcal{C}\psi_{n,(\alpha,0,\beta)}^2$ (choose β as the largest one). In this proof, we abbreviate $\psi_{n,(\alpha,0,\beta)}$ to $\psi_{n,\beta}$ since in this case, the rate only depends on β. We define β_f as the smallest point on the finite grid $\{\underline{\beta} = \beta_1 < \beta_2 < \cdots < \beta_N = \bar\beta\}$ such that $\beta \le \beta_f$

$$\begin{aligned}&\beta_f \in \{\underline{\beta} = \beta_0 < \beta_1 < \cdots < \beta_N = \bar\beta\},\quad f \in \mathcal{F}(\alpha,0,\beta,L), \|f-f_0\|_2^2 \ge \mathcal{C}\psi_{n,\beta}^2,\\ &\beta \le \beta_f \quad\text{and}\quad \forall\beta_i < \beta_f,\quad \text{we have } \beta > \beta_i.\end{aligned}\tag{14}$$

We shall abbreviate to h_f, $t_{n,f}^2$ and $T_{n,f}$ the bandwidth, the threshold (both defined in Theorem 1) and the statistic (7) corresponding to parameter β_f. We write

$$\begin{aligned}\mathbb{P}_f\big(\forall i \in \{1,\ldots,N\}, |T_{n,i}| \le \mathcal{C}^\star t_{n,i}^2\big) &\le \mathbb{P}_f\big(\big|T_{n,f}-\mathbb{E}_f(T_{n,f})\big| \ge -\mathcal{C}^\star t_{n,f}^2 + \mathbb{E}_f(T_{n,f})\big)\\ &\le \mathbb{P}_f\big(\big|T_{n,f}-\mathbb{E}_f(T_{n,f})\big| \ge \|f-f_0\|_2^2 - \mathcal{C}^\star t_{n,f}^2 + B_f(T_{n,f})\big),\end{aligned}\tag{15}$$

where

$$B_f(T_{n,f}) = \mathbb{E}_f(T_{n,f}) - \|f - f_0\|_2^2 = \|J_h * f\|_2^2 - \|f\|_2^2 + 2\langle f - J_h * f, f_0\rangle$$

is in fact a bias term. It satisfies

$$\begin{aligned}\left|B_f(T_{n,f})\right| &\le \int_{|u|\ge 1/h_f} \left|\Phi(u)\right|^2 \mathrm{d}u + 2\left(\int_{|u|\ge 1/h_f} \left|\Phi(u)\right|^2 \mathrm{d}u \int_{|u|\ge 1/h_f} \left|\Phi_0(u)\right|^2 \mathrm{d}u\right)^{1/2} \\ &\le L\mathrm{e}^{-2\underline{\alpha}}\left(h_f^{2\beta} + 2h_f^{\bar{\beta}+\beta}\right) \le 3\mathrm{e}^{-2\underline{\alpha}} L h_f^{2\beta},\end{aligned}$$

as f belongs to $\mathcal{F}(\alpha, 0, \beta, L) \subseteq \mathcal{F}(\underline{\alpha}, 0, \beta, L)$.

Let us study the variance term $\mathbb{E}_f(T_{n,f} - \mathbb{E}_f(T_{n,f}))^2$. According to [2], this term is upper-bounded by $w_{n,f}^2$ given by

$$\mathbb{E}_f\left(T_{n,f} - \mathbb{E}_f(T_{n,f})\right)^2 \le \frac{C}{n^2 h_f^{4\sigma+1}} + \frac{4\Omega_g^2(f - f_0)}{n} 1_{\beta\ge\sigma} = w_{n,f}^2,$$

and $\Omega_g(f - f_0)$ is a constant depending on f and g (but not n) and satisfying $|\Omega_g^2(f - f_0)| \le C\|f - f_0\|_2^{2-2\sigma/\beta}$ (see proof of Theorem 6 in [2]).

Using Markov's inequality, this leads to the following upper bound of (15)

$$\frac{w_{n,f}^2}{(\|f - f_0\|_2^2 - \mathcal{C}^\star t_{n,f}^2 - 3\mathrm{e}^{-2\underline{\alpha}} L h_f^{2\beta})^2}.$$

We will proceed differently when $\beta < \sigma$ and when $\beta \ge \sigma$. Let us first consider the term concerning $\beta < \sigma$. The point is to use that f satisfies $\|f - f_0\|_2^2 \ge \mathcal{C}\psi_{n,\beta}^2$. Note that we have $\beta_f \ge \beta$, constants $\mathcal{C} > \mathcal{C}^\star$ and

$$\psi_{n,\beta}^2 t_{n,f}^{-2} = (n\rho_n)^{4(\beta_f-\beta)(4\sigma+1)/\{(4\beta_f+4\sigma+1)(4\beta+4\sigma+1)\}},$$

ensuring that the term $\mathcal{C}\psi_{n,\beta}^2 - \mathcal{C}^\star t_{n,f}^2$ is always positive. Moreover, as $0 \ge \beta - \beta_f \ge -(\bar{\beta} - \underline{\beta})/\log n$, we have

$$\begin{aligned}\psi_{n,\beta}^2 h_f^{-2\beta} &= \exp\left\{\frac{16\beta(\beta - \beta_f)}{(4\beta_f + 4\sigma + 1)(4\beta + 4\sigma + 1)} \log(n\rho_n)\right\} \\ &\ge \exp\left\{-\frac{16\bar{\beta}(\bar{\beta} - \underline{\beta})}{(4\underline{\beta} + 4\sigma + 1)^2}\left(1 + \mathrm{o}(1)\right)\right\} =: \mathcal{C}_1\left(1 + \mathrm{o}(1)\right).\end{aligned}$$

Thus, we choose $\mathcal{C}^0 = \mathcal{C}^\star + 3\mathrm{e}^{-2\underline{\alpha}} L/\mathcal{C}_1$ such that for any $\mathcal{C} > \mathcal{C}^0$, we have

$$\|f - f_0\|_2^2 - \mathcal{C}^\star t_{n,f}^2 - 3\mathrm{e}^{-2\underline{\alpha}} L h_f^{2\beta} \ge \left(\mathcal{C} - \mathcal{C}^* - \frac{3\mathrm{e}^{-2\underline{\alpha}} L}{\mathcal{C}_1}\right)\psi_{n,\beta}^2 = a\psi_{n,\beta}^2,$$

with $a > 0$. Thus, we get

$$\begin{aligned}&\sup_{\alpha>\underline{\alpha}} \sup_{\beta\in[\underline{\beta};\bar{\beta}]} \sup_{\substack{f\in\mathcal{F}(\alpha,0,\beta,L)\\ \|f-f_0\|_2^2\ge\mathcal{C}\psi_{n,\beta}^2}} \mathbb{P}_f\left(\forall i \in \{1, \dots, N\}, |T_{n,i}| \le \mathcal{C}^\star t_{n,i}^2\right) \\ &\quad\le \max\left\{\sup_{\beta<\sigma} \sup_f \frac{C}{n^2 h_f^{4\sigma+1} \psi_{n,\beta}^4}; \sup_{\beta\ge\sigma} \sup_f \frac{C\|f - f_0\|_2^{2-2\sigma/\beta}}{n(\|f - f_0\|_2^2 - \mathcal{C}^\star t_{n,f}^2 - 3\mathrm{e}^{-2\underline{\alpha}} L h_f^{2\beta})^2}\right\}.\end{aligned}$$

Finally, this leads to the bound

$$\max\left\{\sup_{\beta<\sigma}\sup_f \frac{C}{n^2 h_f^{4\sigma+1}\psi_{n,\beta}^4}; \sup_{\beta\geq\sigma}\frac{C}{n\|f-f_0\|_2^{2+2\sigma/\beta}(a/\mathcal{C})^2}\right\}$$
$$\leq \max\left\{\sup_{\beta<\sigma}\sup_f \frac{C}{n^2 h_f^{4\sigma+1}\psi_{n,\beta}^4}; \sup_{\beta\geq\sigma}\frac{C}{n\psi_{n,\beta}^{2+2\sigma/\beta}}\right\}\leq \rho_n,$$

which converges to zero as n tends to infinity. □

Proof of Lemma 5. As $\beta>\nu$, the bandwidths satisfy $h_\nu h_\beta^{-1}=\mathrm{o}(1)$. Then, as G is compactly supported on $[-1,0]$, we have

$$\mathbb{E}_0\left(\frac{G_\beta(Y_1-x_{j,\beta})G_\nu(Y_1-x_{i,\nu})}{p_0^2(Y_1)}\right)=\int_{\mathbb{R}}\frac{G_\beta(y-x_{j,\beta})G_\nu(y-x_{i,\nu})}{p_0(y)}\,\mathrm{d}y$$
$$=\int_{[-1,0]}\frac{G_\beta(h_\nu u+x_{i,\nu}-x_{j,\beta})G(u)}{p_0(h_\nu u+x_{i,\nu})}\,\mathrm{d}u.$$

Apply the Taylor formula to get

$$G_\beta(h_\nu u+x_{i,\nu}-x_{j,\beta})=G_\beta(x_{i,\nu}-x_{j,\beta})+\frac{h_\nu}{h_\beta^2}uG'\left(\frac{h_\nu\tilde{u}_1+x_{i,\nu}-x_{j,\beta}}{h_\beta}\right)$$

and

$$\frac{1}{p_0(h_\nu u+x_{i,\nu})}=\frac{1}{p_0(x_{i,\nu})}-\frac{p_0'(h_\nu\tilde{u}_2+x_{i,\nu})}{p_0(h_\nu\tilde{u}_2+x_{i,\nu})^2}h_\nu u,$$

where $0\leq\tilde{u}_1\leq u$ and $0\leq\tilde{u}_2\leq u$. As $\int G=0$, we obtain

$$\int_{[-1,0]}\frac{G_\beta(h_\nu u+x_{i,\nu}-x_{j,\beta})G(u)}{p_0(h_\nu u+x_{i,\nu})}\,\mathrm{d}u$$
$$=\frac{1}{p_0(x_{i,\nu})}\frac{h_\nu}{h_\beta^2}\int_{[-1,0]}uG'\left(\frac{h_\nu\tilde{u}_1+x_{i,\nu}-x_{j,\beta}}{h_\beta}\right)G(u)\,\mathrm{d}u$$
$$-h_\nu G_\beta(x_{i,\nu}-x_{j,\beta})\int_{[-1,0]}\frac{p_0'(h_\nu\tilde{u}_2+x_{i,\nu})}{p_0(h_\nu\tilde{u}_2+x_{i,\nu})^2}uG(u)\,\mathrm{d}u$$
$$-\frac{h_\nu^2}{h_\beta^2}\int_{[-1,0]}\frac{p_0'(h_\nu\tilde{u}_2+x_{i,\nu})}{p_0(h_\nu\tilde{u}_2+x_{i,\nu})^2}u^2G'\left(\frac{h_\nu\tilde{u}_1+x_{i,\nu}-x_{j,\beta}}{h_\beta}\right)G(u)\,\mathrm{d}u.$$

This leads to

$$\mathbb{E}_0\left(\frac{G_\beta(Y_1-x_{j,\beta})G_\nu(Y_1-x_{i,\nu})}{p_0^2(Y_1)}\right)=\frac{h_\nu}{h_\beta^2}R_{i,j},$$

where

$$R_{i,j}=\frac{1}{p_0(x_{i,\nu})}\int_{[-1,0]}uG'\left(\frac{h_\nu\tilde{u}_1+x_{i,\nu}-x_{j,\beta}}{h_\beta}\right)G(u)\,\mathrm{d}u$$
$$-h_\beta G\left(\frac{x_{i,\nu}-x_{j,\beta}}{h_\beta}\right)\int_{[-1,0]}\frac{p_0'(h_\nu\tilde{u}_2+x_{i,\nu})}{p_0(h_\nu\tilde{u}_2+x_{i,\nu})^2}uG(u)\,\mathrm{d}u$$
$$-h_\nu\int_{[-1,0]}\frac{p_0'(h_\nu\tilde{u}_2+x_{i,\nu})}{p_0(h_\nu\tilde{u}_2+x_{i,\nu})^2}u^2G'\left(\frac{h_\nu\tilde{u}_1+x_{i,\nu}-x_{j,\beta}}{h_\beta}\right)G(u)\,\mathrm{d}u$$

satisfies

$$|R_{ij}| \le \Big(\inf_{[0,1]} p_0\Big)^{-1} \|G\|_\infty \big\|G'\big\|_\infty + \|G\|_\infty \big\|p_0'\big\|_\infty \Big(\inf_{[-1,1]} p_0\Big)^{-2} \big(h_\beta \|G\|_\infty + h_\nu \big\|G'\big\|_\infty\big),$$

which ends the proof of Lemma 5. □

References

[1] C. Butucea. Asymptotic normality of the integrated square error of a density estimator in the convolution model. *SORT* **28**(1) (2004) 9–26. MR2076033

[2] C. Butucea. Goodness-of-fit testing and quadratic functional estimation from indirect observations. *Ann. Statist.* **35**(5) (2007) 1907–1930. MR2363957

[3] C. Butucea, C. Matias and C. Pouet. Adapativity in convolution models with partially known noise distribution. Technical report, 2007. Available at http://arxiv.org/abs/0804.1056.

[4] M. Fromont and B. Laurent. Adaptive goodness-of-fit tests in a density model. *Ann. Statist.* **34**(2) (2006) 680–720. MR2281881

[5] G. Gayraud and C. Pouet. Adaptive minimax testing in the discrete regression scheme. *Probab. Theory Related Fields* **133**(4) (2005) 531–558. MR2197113

[6] E. Giné, R. Latała and J. Zinn. Exponential and moment inequalities for U-statistics. In *High Dimensional Probability, II (Seattle, WA, 1999)* 13–38. *Progr. Probab.* **47**. Birkhäuser Boston, Boston, MA, 2000. MR1857312

[7] P. Hall. Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivariate Anal.* **14**(3) (1984) 1–16. MR0734096

[8] P. Hall and C. Heyde. *Martingale Limit Theory and Its Application*. Academic Press, New York, 1980. MR0624435

[9] H. Holzmann, N. Bissantz and A. Munk. Density testing in a contaminated sample. *J. Multivariate Anal.* **98**(1) (2007) 57–75. MR2292917

[10] C. Houdré and P. Reynaud-Bouret. Exponential inequalities, with constants, for U-statistics of order two. In *Stochastic Inequalities and Applications* 55–69. *Progr. Probab.* **56**. Birkhäuser, Basel, 2003. MR2073426

[11] Y. Ingster and I. Suslina. *Nonparametric Goodness-Of-Fit Testing Under Gaussian Models*. Springer, New York, 2003. MR1991446

[12] V. Korolyuk and Y. Borovskikh. *Theory of U-statistics*. Kluwer Academic Publishers, Dordrecht, 1994. MR1472486

[13] C. Pouet. On testing non-parametric hypotheses for analytic regression functions in Gaussian noise. *Math. Methods Statist.* **8**(4) (1999) 536–549. MR1755899

[14] V. Spokoiny. Adaptive hypothesis testing using wavelets. *Ann. Statist.* **24**(6) (1996) 2477–2498. MR1425962