n° 2003-44

**Adaptive Minimax Testing in
the Discrete Regression Scheme**

G. GAYRAUD¹
C. POUET²

Les documents de travail ne reflètent pas la position de l'INSEE et n'engagent que leurs auteurs.

Working papers do not reflect the position of INSEE but only the views of the authors.

¹ CREST, Laboratoire de Statistique et Université de Rouen, Laboratoire de Mathématiques Raphael Salem.

² Centre de Mathématiques et Informatique, Université de Provence, 39 rue Joliot Curie, 13453 Marseille Cedex 13.

Adaptive minimax testing in the discrete regression scheme

Ghislaine Gayraud ^{*}
Christophe Pouet [†]

October 29, 2003

Adaptive minimax testing in the discrete regression scheme

BY

GHISLAINE GAYRAUD AND CHRISTOPHE POUET

Résumé

A partir de données observées sur un découpage régulier de $[0, 1]$, nous nous intéressons à un problème de test non-paramétrique concernant la fonction de régression f . Sous l'hypothèse nulle, la fonction f appartient à un espace fonctionnel donné (paramétrique ou bien non-paramétrique) dont on contrôle l'entropie. Sous l'hypothèse alternative, les fonctions sont séparées en norme L_2 de n'importe quelle fonction de l'hypothèse nulle. Sous l'alternative, les fonctions appartiennent à une grande variété de classes de Hölder et comme nous supposons la régularité des fonctions de l'alternative inconnue, notre approche est adaptative. Nous montrons que la version adaptative entraîne une perte de performance dans la vitesse minimax de test par rapport à la version non-adaptative; cette perte est optimale et elle est de l'ordre d'un terme en $\sqrt{\log \log}$. On propose une statistique de test ne dépendant que des observations qui atteint la vitesse de test optimale et on prouve qu'aucun test n'est consistant si la distance qui sépare l'hypothèse nulle de l'alternative est d'un ordre plus petit que la perte optimale.

Mots clés : Adaptation, hypothèse nulle composée, test d'hypothèses non-paramétrique, vitesse minimax de test, modèle de régression.

Summary

We consider the problem of testing hypotheses on the regression function from n observations on the regular grid on $[0, 1]$. We wish to test the null hypothesis that the regression function belongs to a given functional class (parametric or even nonparametric) against a composite nonparametric alternative. The functions under the alternative are separated in the L_2 -norm from any functions in the null hypothesis. We assume that the regression function belongs to a wide range of Hölder classes but as the smoothness parameter of the regression function is unknown, an adaptive approach is considered. It leads to an optimal but unavoidable loss of order $\sqrt{\log(\log n)}$ in the minimax rate of testing compared with the non-adaptive setting. We propose a smoothness-free test that achieves the optimal rate, and finally we prove the lower bound showing that no test can be consistent if in the distance between the functions under the null hypothesis and those in the alternative, the loss is of order smaller than the optimal loss.

Key words : Adaptive testing, composite null hypothesis, nonparametric hypotheses testing, minimax rate of testing, regression model.

1 Introduction

Discrete regression models are frequently used in statistics and econometrics. If the model is misspecified, it can lead to very inaccurate results. Therefore, the problem of hypothesis testing about regression function f is essential since it allows us to check the specification of the model. Nonparametric methods seem to be natural if you do not have in mind specific departures from the model. There have been a number of papers, initiated by Ingster (1982), concerning nonparametric hypothesis testing on the signal function in the regression model or in the Gaussian white noise model, via an asymptotical minimax approach : Ermakov (1990), Ingster (1990,1993), Härdle and Mammen (1993) , Lepskii (1993), Suslina (1993), Hart (1997), Härdle and Kneip (1999), Lepskii and Spokoiny (1999), Gayraud and Pouet (2001). In all of these papers, the authors specify a functional class to which f belongs; the considered functional classes are characterized by a smoothness parameter. Then, the goal of these papers is to determine the minimal (optimal) distance between the null hypothesis and the set of alternatives for which testing with prescribed error probabilities is still possible. However, the test procedures depend heavily on the smoothness assumption which is typically unknown as well as the regression function f : this makes them unnatural and unattractive from a practical point of view. It is our first motivation to extend the non-adaptive case of our previous paper (Gayraud and Pouet 2001) to the adaptive case i.e. the case where the smoothness parameter is also supposed unknown. Four other papers must be mentioned here in the context of adaptive resolution of hypothesis testing about the regression function : in Baraud, Huet and Laurent (2003), no assumption on f is required since the distance they used to separate the null hypothesis and the alternative is a discrete one; it avoids to quantify the approximation of the L_2 -norm, when it is used as the distance as in our study, by a discrete sum of squared terms. The same occurs in Horowitz and Spokoiny (2001) who give results for a composite null hypothesis and for Hölder spaces. Here also, we consider a functional class of Hölderian functions with a smoothness parameter ranging over a wide scale. But we are more general than Horowitz and Spokoiny (2001) in the choice of the class under the null hypothesis since they take a parametric family of given functions whereas in our study only a control of the entropy of the class which could even be nonparametric is required. The closest paper to our approach is Spokoiny (1996) and also its extension Spokoiny (1998) (the nontrivial extension of this second paper is to consider an arbitrary norm L_r instead of the L_2 -norm), although its model is the Gaussian white noise and its null hypothesis is defined by the unique null function. In our study, we do not specify the law of the errors except for the resolution of the lower bound and we get the upper bound under an assumption on the control of the tail distribution of the errors. Moreover Spokoiny (1996, 1998) consider a collection of Besov balls; it allows to decompose the regression function f on an orthonormal basis of wavelets which gives interesting properties of independence for the resolution of the lower bound.

To summarize, this paper is devoted to obtain the optimal minimax adaptive rates in testing $\mathcal{F}_0$, a given functional class of regression functions, against a functional class composed of Hölder balls. Our study gives the possibility to consider a rich class $\mathcal{F}_0$ since only a control of its entropy is required : parametric but also nonparametric functional classes are allowed. On the contrary of other papers (Horowitz and Spokoiny (2001), Guerre and Lavergne (2002)) in which a parametric family under the null hypothesis is considered, no existence of an estimate under the null hypothesis is needed. Concerning the upper bound, we do not specify the law of the errors; just a control of the tail distribution is required. It is worth mentioning an interesting link between the distribution tail of the errors and the entropy of the class $\mathcal{F}_0$. Besides, we prove that the adaptive case leads to a loss of efficiency of order $\sqrt{\log(\log n)}$.

The paper is organized as follows. In Section 2, we state the testing problem and the minimax approach to solve it. In Section 3, we explain the proposed adaptive test procedure. Section 4 is devoted to the statement of the results. In Section 5, some comments are given, in addition with some figures which illustrate the link between the distribution tail of the error and the control of the entropy of $\mathcal{F}_0$. The proofs are postponed in Section 6 and those of the lemmas are given in Appendix.

2 Model and adaptive minimax framework

We consider the usual regression model

$$Y_i = f(x_i) + \xi_i, \quad i = 1, \dots, n,$$

where f defined on $[0, 1]$ belongs to the class $\Sigma(\beta, C, M)$, $\beta > 1/4$, $C > 0$, $M > 0$, which is defined as follows :

$$\Sigma(\beta, C, M) = \begin{cases} \{f \in H(\beta, C) : \|f\|_\infty \leq M\} & \text{if } \frac{1}{4} < \beta \leq 1, \\ \{f \in H(\beta, C) : \|f\|_\infty \leq M, \|f'\|_\infty \leq M\} & \text{if } \beta > 1, \end{cases}$$

where $H(\beta, C)$ denotes the Hölderian class of functions with β as the smoothness parameter and C as the Hölder constant and $\|\cdot\|_\infty$ denotes the supremum norm. Besides, the real random variables ξ_i , $i \in \{1, \dots, n\}$ are i.i.d. with zero mean and unknown variance $\sigma^2 > 0$, and the points x_i are deterministic equispaced on $[0, 1]$ and numbered in such a way that $|x_i - x_{i-1}| = \frac{1}{n}$, $\forall i = 2, \dots, n$.

Given the sample $Y_1, \dots, Y_n$ and if we suppose β known, one could consider the following test problem

$$H_0 : f \in \mathcal{F}_0 \subset \Sigma(\beta, C, M),$$

against

$$H_1 : f \in \Lambda_{n,\beta}(v(n, \beta)) = \left\{ f \in \Sigma(\beta, C, M) : \inf_{f_0 \in \mathcal{F}_0} \|f - f_0\|_2 \geq v(n, \beta) \right\},$$

where $\|\cdot\|_2$ denotes the L_2 -norm and $v(n, \beta)$ is a sequence of positive numbers decreasing to zero as n goes to infinity.

In the problem of testing H_0 against H_1 assuming β known, the minimax rate of testing is $v(n, \beta) = n^{-2\beta/(4\beta+1)}$ and a test function which achieves the minimax rate of testing is a statistic which depends obviously on β (Gayraud and Pouet, 2001).

In this paper, we are interested in adaptive testing i.e. we suppose that β is unknown. Following the terminology of Spokoiny (1996), we assume that β belongs to a set $\mathcal{T} = \{\beta : 1/4 < \beta_\star \leq \beta \leq \beta^\star\}$, where $\beta_\star$ and $\beta^\star$ are fixed. We suppose that the given functional class $\mathcal{F}_0$ under H_0 is included in the greatest regular class $\Sigma(\beta^\star, C, M)$. Consequently, we search for an universal test Δ_n such that for n large enough and for any given α_0 and α_1 in $(0, 1)$

$$\sup_{f \in \mathcal{F}_0} \mathbb{P}_f(\Delta_n = 1) \leq \alpha_0 + o_n(1), \quad (1)$$

and

$$\exists A > 0, \text{ such that } \sup_{\beta \in \mathcal{T}} \sup_{f \in \Lambda_n(Av(nt_n^{-1}, \beta))} \mathbb{P}_f(\Delta_n = 0) \leq \alpha_1 + o_n(1), \quad (2)$$

where t_n is a sequence of positive numbers increasing to infinity with n .

One could ask if there is a possibility to get (2) with alternatives defined in a closer neighborhood of H_0 for example with an order of $o(v(nt_n^{-1}, \beta))$ i.e. if one could distinguish the null hypothesis and the alternative when the alternative is much closer to the null hypothesis. The answer is negative and it is done by proving the relation of the lower bound i.e. for n large enough and for any $\alpha_2 \in (0, 1)$, there exists a positive constant a such that

$$\inf_{\Delta_n} \{ \sup_{f \in \mathcal{F}_0} \mathbb{P}_f(\Delta_n = 1) + \sup_{\beta \in \mathcal{T}} \sup_{f \in \Lambda_n(av(nt_n^{-1}, \beta))} \mathbb{P}_f(\Delta_n = 0) \} \geq \alpha_2 + o_n(1), \quad (3)$$

where $\inf_{\Delta_n}$ denotes the infimum under all possible test statistics.

Actually, relations (1), (2) and (3) mean that the non-adaptive rate $v(n, \beta)$ is contaminated by the term t_n in the adaptive setting and it could not be avoided.

3 Test Procedure.

In the case of unknown β in $\mathcal{T}$, and for some f_0 in $\mathcal{F}_0$, we consider

$$T_{n, \beta, f_0} = \frac{1}{\sqrt{m_\beta}} \sum_{k=1}^{m_\beta} \frac{m_\beta}{n \hat{\sigma}_n^2} \sum_{\substack{i \in I_k, j \in I_k \\ i \neq j}} (Y_i - f_0(x_i))(Y_j - f_0(x_j)), \quad (4)$$

where $m_\beta = (t_n^{-1}n)^{2/(4\beta+1)}$ with $t_n = \sqrt{\log(\log n)}$, $\hat{\sigma}_n^2 = \frac{1}{2(n-1)} \sum_{i=2}^n (Y_i - Y_{i-1})^2$ is an estimate of σ^2 and $I_k = \{i : x_i \in A_k\}$, with $A_k = [\frac{k-1}{m_\beta}, \frac{k}{m_\beta}]$, $\forall k \in \{1, \dots, m_\beta - 1\}$ and $A_{m_\beta} = [\frac{m_\beta-1}{m_\beta}, 1]$. The quantity m_β is supposed to be integer, if it is not the case, m_β will be chosen equal to the integer part of $(t_n^{-1}n)^{2/(4\beta+1)}$; therefore one could prove that if any β_1 and β_2 in $\mathcal{T}$ are much closer in the Euclidean distance than the order $\frac{1}{\log n}$, both of them provide the same m_{β_1} . Consequently, the range of adaptation $\mathcal{T}$ can be translated into a range $\mathcal{M}$ of the form $\mathcal{M} = \{m_\beta : m_{\beta_\star} \leq m_\beta \leq m_{\beta^\star}\}$ in which two consecutively m_{β_j} and $m_{\beta_{j+1}}$ correspond to a subset $\mathcal{T}^\star \subset \mathcal{T}$ of $\beta \in \mathcal{T}$ which are at least distant in Euclidean distance from $c \frac{1}{\log n}$ where c is an adapted positive constant which makes $m_{\beta_{j+1}}$ an integer and it depends on $\beta_\star, \beta^\star$ and n . Since $\beta_\star$ and $\beta^\star$ are fixed, the cardinality of the set $\mathcal{M}$ is of order $\log n$ as n large enough. Now we are able to define the adaptive test :

$$\Delta_n = \mathbb{I}_{\sup_{m_\beta \in \mathcal{M}} \inf_{f_0 \in \mathcal{F}_0} T_{n, \beta, f_0} \geq \rho_n}, \quad (5)$$

where $\rho_n = 4\sqrt{\log(\log n)}$ and T_{n, β, f_0} is given by relation (4).

4 Assumptions and Results

Assume the following :

(A1.Sup) The random variables $\xi_1, \dots, \xi_n$ are i.i.d., with zero mean and unknown finite variance $\sigma^2 > 0$.

(A1.Inf) The random variables $\xi_1, \dots, \xi_n$ are i.i.d. centered Gaussian with variance equal to $\sigma^2 > 0$.

(A2.Sup) $\lim_{y \rightarrow +\infty} \mathbb{P}(|\xi_i| \geq y) y^{\left(\frac{1}{\kappa}(2 + \frac{1}{2\beta_\star}) \vee 3\right)} = 0$, $\forall i \in \{1, \dots, n\}$, where $\kappa < 1$ is a positive constant.

(A3) The δ -entropy of the class $\mathcal{F}_0$ calculated in the supremum norm is bounded by δ^{-r} , where r is less than $\frac{2\beta_*(1-\kappa)(1+4\beta^*)}{(16+8\kappa)\beta_*\beta^*+2(2\beta^*+\kappa\beta_*)}$.

Remark 1 *Assumption (A2.Sup) deals with the distribution tail of the ξ_i 's. One must note that there exists a relation between the weight of the distribution tail (A2.Sup) and the entropy of the class $\mathcal{F}_0$. This point will be developed in Section 5. Moreover Assumption (A2.Sup) guarantees the existence of the third moment of the ξ_i 's which is needed to apply Berry-Esseen's Theorem (see the proof of Theorem 1).*

Theorem 1 deals with the upper bound. It states that our test can achieve asymptotically any given first-type and second-type errors under some conditions.

Theorem 1 (Behavior of Δ_n under the null hypothesis and under the alternative)

Under H_0 and according to (A1.Sup)-(A2.Sup), the first type error of Δ_n defined by relation (5) is of order $o_n(1)$ as n goes to infinity. This means that relation (1) holds with $\alpha_0 = 0$.

Suppose that the assumptions (A1.Sup), (A2.Sup) and (A3) hold. Then, there exists a constant $A > 0$ large enough such that relation (2) holds with $\alpha_1 = 0$.

Theorem 2 deals with the lower bound which is stated in the Gaussian case. It gives the optimal rate of testing H_0 against the range of alternatives when $\beta \in \mathcal{T}$, which could not be improved without losing the distinguishability between the hypotheses.

Theorem 2 (Lower bound)

Suppose that (A1.Inf)-(A3) hold and that there exists $\tilde{f}_0 \in \mathcal{F}_0$ such that $\tilde{f}_0$ belongs to all $\Sigma(\beta, C', L')$, when $\beta \in \mathcal{T}$, $C' < C$ and $M' < M$. Then, for any given $\alpha_2 \in (0, 1)$, there exists a positive constant a such that as n large enough,

$$\inf_{\Delta_n} \{ \sup_{f \in \mathcal{F}_0} \mathbb{P}_f(\Delta_n = 1) + \sup_{\beta \in \mathcal{T}} \sup_{f \in \Lambda_n(av(nt_n^{-1}), \beta))} \mathbb{P}_f(\Delta_n = 0) \} \geq \alpha_2 + o_n(1), \quad (6)$$

where $\inf_{\Delta_n}$ means the infimum under all possible test functions.

Proposition 1 *Under Assumption (A1.Sup), the estimate $\hat{\sigma}_n^2$ satisfies*

$$\forall \gamma > 0 \sup_{\beta \in \mathcal{T}} \sup_{f \in \Sigma(\beta, C, M)} \mathbb{P}_f(|\frac{\hat{\sigma}_n^2}{\sigma^2} - 1| > \gamma) \xrightarrow{n \rightarrow \infty} 0.$$

5 Comments

Lower bound

The construction of the lower bound is rather classical. Yet, there is a significative difference with Spokoiny (1996). In his paper, the parametric subset in the alternatives is built using wavelets. It provides good property of independence through the orthogonality between the different levels of resolution. Here we have no natural basis for constructing the parametric family in the alternative, which makes the proof rather technical. To cope with the dependence we met, we split the main sum over the range of β into several sums where the summands are almost independent.

Link between the smoothness parameter, the entropy and the distribution tail of the errors

In this part, the link between the smoothness parameter, the entropy and the distribution tail of the errors is highlighted by several points of view. It is illustrated via the three figures below.

Figure 1. When the behavior of the tail is known (i.e. κ is given), there is a relationship between the minimum smoothness, also the maximum one and the entropy of $\mathcal{F}_0$. In the case of $\kappa = 0.95$ (heavy tail), r must be much smaller than in the case of $\kappa = 0.05$ (light tail). This phenomena indicates that the behavior of the tail plays an important role in the problem. If a class is rich, light tails are needed in order to provide good localization. The role of $\beta_\star$ and $\beta^\star$ are important in the small values domain. When $\beta_\star$ is large, it seems to be almost no difference for r . We have to keep in mind another condition : $\mathcal{F}_0$ is included in $H(\beta^\star, C)$. It leads to a link between r and $\beta^\star$ defined by $r < \frac{1}{\beta^\star}$ (see Chapter 15 in Lorentz, Golitschek and Makovoz (1996) for the calculation of the entropy of a Hölder ball).

Figure 2. We suppose that the true smoothness is $\beta_0 = 1.2$ and we illustrate the effect of the smoothness range on r and κ . First, in all cases the heavier the tail is, the poorer the class $\mathcal{F}_0$ has to be. A small range for β allows large values for r . Also, we can see that the choice of $\beta_\star$ is much more important than the choice of $\beta^\star$. Thus the statistician has to be more careful in the choice of $\beta_\star$ than in the choice of $\beta^\star$.

Figure 3. We consider different situations that can be encountered. We show the effect of the assumptions (A2.Sup) and (A3) on the choice of $\beta_\star$ and $\beta^\star$. For a given pair of (κ, r) , the allowed area for $(\beta_\star, \beta^\star)$ is the intersection between the part of the plane under the horizontal line $(\beta^\star < \frac{1}{r})$, the one over the line $\beta_\star = \beta^\star$ since $\beta_\star \leq \beta^\star$ and the one under the third curve (Assumption (A3)). The range of adaptation is restricted by those assumptions. For example it is impossible to consider a rich class ($r = 0.3$) and a heavy tail ($\kappa = 0.912$) since Assumption (A3) is never satisfied (this is the reason why it does not appear on the graph). On the contrary for a poor class ($r = 0.03$) and a light tail ($\kappa = 0.15$), Assumption (A3) is always satisfied (this is the reason why it does not appear in the graph), we have a wide choice of $(\beta_\star, \beta^\star)$. For the two other graphs (a rich class and a light tail or a poor class and a heavy tail), Assumption (A3) plays a role for small values of $\beta_\star$.

Figure 1: κ fixed and r as the function of β_* and $\beta^* - \beta_*$

Figure 2: $\beta_0 = 1.2$ fixed and r as the function of κ

Figure 3: κ and r fixed, the range for β^* as function of β_*

6 Proofs

6.1 Proof of Theorem 1

The first type error : for any function f_0 in $\mathcal{F}_0$, we get

$$\begin{aligned}
 \mathbb{P}_{f_0}(\Delta_n = 1) &= \mathbb{P}_{f_0} \left(\sup_{m_\beta \in \mathcal{M}} \inf_{f \in \mathcal{F}_0} T_{n,\beta,f} \geq \rho_n \right) \\
 &\leq \mathbb{P}_{f_0} \left(\sup_{m_\beta \in \mathcal{M}} T_{n,\beta,f_0} \geq \rho_n \right) \\
 &\leq \mathbb{P}_{f_0} \left(\left\{ \sup_{m_\beta \in \mathcal{M}} \frac{\hat{\sigma}_n^2}{\sigma^2} T_{n,\beta,f_0} \geq \frac{\hat{\sigma}_n^2}{\sigma^2} \rho_n \right\} \cap \left\{ \left| \frac{\hat{\sigma}_n^2}{\sigma^2} - 1 \right| \leq \gamma \right\} \right) + \mathbb{P}_{f_0} \left(\left| \frac{\hat{\sigma}_n^2}{\sigma^2} - 1 \right| > \gamma \right) \\
 &\leq \sum_{m_\beta \in \mathcal{M}} \mathbb{P}_{f_0} \left(\frac{\hat{\sigma}_n^2}{\sigma^2} T_{n,\beta,f_0} \geq \rho_n (1 - \gamma) \right) + \mathbb{P}_{f_0} \left(\left| \frac{\hat{\sigma}_n^2}{\sigma^2} - 1 \right| > \gamma \right). \tag{7}
 \end{aligned}$$

Due to Proposition 1, the second term in right hand side of (7) is of order $o_n(1)$ as n large enough, and since (7) holds for any $\gamma > 0$, we take it equal to $\frac{1}{2}$. Next, we focus only on the first term in the right hand side of (7).

Note that under the distribution $\mathbb{P}_0$, the test statistic (4) can be rewritten as the following

$$\begin{aligned}\frac{\hat{\sigma}_n^2}{\sigma^2} T_{n,\beta,f_0} &= \frac{1}{\sqrt{m_\beta}} \sum_{k=1}^{m_\beta} \frac{m_\beta}{n\sigma^2} \sum_{\substack{i \in I_k, j \in I_k \\ i \neq j}} \xi_i \xi_j \\ &= \frac{\sqrt{2(1 - \frac{m_\beta}{n})}}{\sqrt{m_\beta}} \sum_{k=1}^{m_\beta} \eta_k,\end{aligned}$$

where $\eta_k = \frac{m_\beta}{\sigma^2 n \sqrt{2(1 - \frac{m_\beta}{n})}} \left(\sum_{\substack{i \in I_k, j \in I_k \\ i \neq j}} \xi_i \xi_j \right)$. Note that the η_k 's are i.i.d. centered variables

with an unit variance. Next we apply two results :

-the first one concerns the control of the tail distribution of an empirical mean of $X_1, \dots, X_N$, N i.i.d. observations with zero mean and a finite variance σ^2 . Then, with an extra condition on the existence of the third order moment which is given by Assumption (A2.Sup), the relations

$$\frac{1 - \mathbb{P}(\sigma^{-1} N^{-1/2} \sum_{i=1}^N X_i \leq x)}{1 - \Phi(x)} \xrightarrow{N \rightarrow \infty} 1, \quad (8)$$

$$\frac{\mathbb{P}(\sigma^{-1} N^{-1/2} \sum_{i=1}^N X_i \leq -x)}{\Phi(-x)} \xrightarrow{N \rightarrow \infty} 1, \quad (9)$$

hold in the interval $0 \leq x \leq (1 - \varepsilon)(\log(N))^{1/2}$, where $0 < \varepsilon < 1$.

-the second one is the well-known inequality

$$1 - \Phi(x) \leq \exp(-\frac{x^2}{2}) \frac{1}{\sqrt{2\pi}x}, \quad \forall x > 0. \quad (10)$$

Due to the order of ρ_n and applying results (8) and (10) to the second term in the right hand side of (7), we get

$$\begin{aligned}\sum_{m_\beta \in \mathcal{M}} \mathbb{P} \left(\frac{\sqrt{2(1 - \frac{m_\beta}{n})}}{\sqrt{m_\beta}} \sum_{k=1}^{m_\beta} \eta_k > \frac{\rho_n}{2\sqrt{2(1 - \frac{m_\beta}{n})}} \right) &\leq \sum_{m_\beta \in \mathcal{M}} \exp(-\log(\log n)) \frac{1}{2\sqrt{\pi} \sqrt{\log(\log n)}} \\ &\leq \sum_{m_\beta \in \mathcal{M}} \frac{1}{(\log n)} \frac{1}{2\sqrt{\pi} \sqrt{\log(\log n)}}. \quad (11)\end{aligned}$$

Finally, the right hand of (11) goes to zero as n goes to infinity like $\frac{1}{\log(\log n)}$ since $\text{card}(\mathcal{M}) = O(\log n)$.

The second type error : For all $\beta \in \mathcal{T}$, we have

$$\begin{aligned}\sup_{f \in \Lambda_n(Av(nt_n^{-1}, \beta))} \mathbb{P}_f \left(\sup_{m_\beta \in \mathcal{M}} \inf_{f_0 \in \mathcal{F}_0} T_{n,\beta,f_0} \leq \rho_n \right) &\leq \sup_{f \in \Lambda_n(Av(nt_n^{-1}, \beta))} \mathbb{P}_f \left(\left| \frac{\hat{\sigma}_n^2}{\sigma^2} - 1 \right| > \gamma \right) + \\ &\quad \sup_{f \in \Lambda_n(Av(nt_n^{-1}, \beta))} \mathbb{P}_f \left(\sup_{m_\beta \in \mathcal{M}} \inf_{f_0 \in \mathcal{F}_0} \frac{\hat{\sigma}_n^2}{\sigma^2} T_{n,\beta,f_0} \leq \rho_n(1 + \gamma) \right). \quad (12)\end{aligned}$$

Due to Proposition 1, the first term in the right hand side of (12) is $o_n(1)$ as n large enough and we could take γ equal to 1; therefore we focus only on the second term in the right hand side of (12) :

$$\begin{aligned} \sup_{f \in \Lambda_n(Av(nt_n^{-1}, \beta))} \mathbb{P}_f \left(\sup_{m_\beta \in \mathcal{M}} \inf_{f_0 \in \mathcal{F}_0} \frac{\hat{\sigma}_n^2}{\sigma^2} T_{n, \beta, f_0} \leq 2\rho_n \right) &\leq \sup_{f \in \Lambda_n(Av(nt_n^{-1}, \beta))} \mathbb{P}_f \left(\inf_{f_0 \in \mathcal{F}_0} \frac{\hat{\sigma}_n^2}{\sigma^2} T_{n, \beta_L, f_0} \leq 2\rho_n \right) \\ &= \sup_{f \in \Lambda_n(Av(nt_n^{-1}, \beta))} \mathbb{P}_f \left(\inf_{f_0 \in \mathcal{F}_0} \frac{\sqrt{m_L}}{n} \frac{\hat{\sigma}_n^2}{\sigma^2} T_{n, \beta_L, f_0} \leq \frac{2\rho_n \sqrt{m_L}}{n} \right), \quad (13) \end{aligned}$$

where β_L in T_{n, β_L, f_0} is $\beta_j \in \mathcal{T}^*$ since the true regularity β of the function f , belongs to $[\beta_j, \beta_{j+1}]$, m_L is associated with β_L . Under $f \in \Lambda_n(Av(nt_n^{-1}, \beta))$, rewrite $\frac{\sqrt{m_L}}{n} \frac{\hat{\sigma}_n^2}{\sigma^2} T_{n, \beta_L, f_0}$ in the following way

$$\begin{aligned} \frac{\sqrt{m_L}}{n} \frac{\hat{\sigma}_n^2}{\sigma^2} T_{n, \beta_L, f_0} &= \frac{1}{m_L} \sum_{k=1}^{m_L} \frac{m_L^2}{\sigma^2 n^2} \left(\sum_{\substack{i \in I_k, j \in I_k \\ i \neq j}} (\xi_i + f(x_i) - f_0(x_i)) (\xi_j + f(x_j) - f_0(x_j)) \right) \\ &= \frac{1}{m_L} \sum_{k=1}^{m_L} \frac{m_L^2}{n^2 \sigma^2} \left(\sum_{\substack{i \in I_k, j \in I_k \\ i \neq j}} \xi_i \xi_j \right) + 2 \frac{1}{m_L} \sum_{k=1}^{m_L} \frac{m_L^2}{n^2 \sigma^2} \left(\sum_{\substack{i \in I_k, j \in I_k \\ i \neq j}} (f(x_i) - f_0(x_i)) \xi_j \right) \\ &\quad + \frac{1}{m_L} \sum_{k=1}^{m_L} \frac{m_L^2}{\sigma^2 n^2} \left(\sum_{\substack{i \in I_k, j \in I_k \\ i \neq j}} (f(x_i) - f_0(x_i)) (f(x_j) - f_0(x_j)) \right). \end{aligned}$$

Set

$$\begin{aligned} T_{n, 1, \beta_L} &= \frac{1}{m_L} \sum_{k=1}^{m_L} \frac{m_L^2}{n^2 \sigma^2} \left(\sum_{\substack{i \in I_k, j \in I_k \\ i \neq j}} \xi_i \xi_j \right), \\ T_{n, 2, \beta_L, f_0} &= 2 \frac{1}{m_L} \sum_{k=1}^{m_L} \frac{m_L^2}{\sigma^2 n^2} \left(\sum_{\substack{i \in I_k, j \in I_k \\ i \neq j}} (f(x_i) - f_0(x_i)) \xi_j \right), \\ T_{n, 3, \beta_L, f_0} &= \frac{1}{m_L} \sum_{k=1}^{m_L} \frac{m_L^2}{n^2 \sigma^2} \left(\sum_{\substack{i \in I_k, j \in I_k \\ i \neq j}} (f(x_i) - f_0(x_i)) (f(x_j) - f_0(x_j)) \right). \end{aligned}$$

Now, define the event

$$D_{1, \beta} = \left\{ \inf_{f_0 \in \mathcal{F}_0} (T_{n, 2, \beta_L, f_0} + T_{n, 3, \beta_L, f_0}) < \frac{5}{2} \rho_n \frac{\sqrt{m_L}}{n} \right\},$$

and consider again relation (13), $\forall f \in \Lambda_n(Av(nt_n^{-1}, \beta))$,

$$\begin{aligned}
\mathbb{P}_f \left(\inf_{f_0 \in \mathcal{F}_0} \frac{\sqrt{m_L}}{n} \frac{\hat{\sigma}_n^2}{\sigma^2} T_{n, \beta_L, f_0} \leq 2\rho_n \frac{\sqrt{m_L}}{n} \right) &= \mathbb{P}_f \left(T_{n, 1, \beta_L} + \inf_{f_0 \in \mathcal{F}_0} (T_{n, 2, \beta_L, f_0} + T_{n, 3, \beta_L, f_0}) \leq 2\rho_n \frac{\sqrt{m_L}}{n} \right) \\
&\leq \mathbb{P}_f \left(\left\{ T_{n, 1, \beta_L} + \inf_{f_0 \in \mathcal{F}_0} (T_{n, 2, \beta_L, f_0} + T_{n, 3, \beta_L, f_0}) \leq 2\rho_n \frac{\sqrt{m_L}}{n} \right\} \cap D_{1, \beta}^c \right) \\
&\quad + \mathbb{P}_f(D_{1, \beta}) \\
&\leq \mathbb{P} \left(\frac{\sqrt{2(1 - \frac{m_\beta}{n})}}{\sqrt{m_\beta}} \sum_{k=1}^{m_L} \eta_k \leq -\frac{\rho_n}{2} \right) + \mathbb{P}_f(D_{1, \beta}), \quad (14)
\end{aligned}$$

where the η_k 's are defined in the proof of the first type error. Note that the first term in the right-hand side (RHS) of (14) is independent of f and is similar to the one considered in the proof of the first type error; therefore due to relation (9) the limit as n goes to infinity of

$$\mathbb{P} \left(\frac{\sqrt{2(1 - \frac{m_\beta}{n})}}{\sqrt{m_\beta}} \sum_{k=1}^{m_L} \eta_k \leq -\frac{\rho_n}{2} \right) \text{ is } o_n(1). \text{ The proof follows provided that for all } \beta \in \mathcal{T},$$

$$\sup_{f \in \Lambda_n(Av(nt_n^{-1}, \beta))} \mathbb{P}_f(D_{1, \beta}) = o_n(1), \quad \text{as } n \rightarrow +\infty.$$

Lemma 1 *For a sufficiently large positive constant A and under (A1.Sup)-(A2.Sup)-(A3), we get*

$$\sup_{\beta \in \mathcal{T}} \sup_{f \in \Lambda_n(Av(nt_n^{-1}, \beta))} \mathbb{P}_f(D_{1, \beta}) = o_n(1), \quad \text{as } n \rightarrow +\infty.$$

The proof of Lemma 1 is given in Appendix. From (12), (13), (14) and applying Lemma 1, there exists a positive constant A large enough such that relation (2) holds with $\alpha_1 = 0$.

6.2 Proof of Theorem 2

First, let ϕ be a function belonging to all $H(\beta, 1)$, for $\beta \in \mathcal{T}$, with support in $\mathbb{R}$ which satisfies $\phi(x) = 0$ for $x \notin]0, 1[$, $\int_0^1 \phi(x) dx = 0$ and $\int_0^1 \phi(x)^2 dx = \lambda^2$.

Let $\kappa_{n, \beta}^2 = \frac{[1 + (v(nt_n^{-1}, \beta))^{1/(4\beta)}]}{\lambda^2} (av(nt_n^{-1}, \beta))^2$. Denote $\Xi_\beta = \{-1; +1\}^{m_\beta}$, where $m_\beta = [h_\beta^{-1}]$ ($[\cdot]$ denotes the integer part), h_β will be chosen later and is a sequence of positive numbers decreasing to zero as n goes to infinity. Define $\eta_\beta = (\eta_{1, \beta}, \dots, \eta_{m_\beta, \beta}) \in \Xi_\beta$, $z_{k, \beta} = (k-1)h_\beta$ and $A_{k, \beta} = [z_{k, \beta}, z_{k+1, \beta}[$ for $k = 1, \dots, m_\beta - 1$ and $A_{m_\beta, \beta} = [z_{m_\beta-1, \beta}, 1]$. Let $I_{k, \beta} = \{i : x_i \in A_{k, \beta}\}$ for $k = \{1, \dots, m_\beta\}$. Set

$$f_{\eta_\beta}(x) = \tilde{f}_0(x) + \sum_{k=1}^{m_\beta} \eta_{k, \beta} \kappa_{n, \beta} \phi \left(\frac{x - z_{k, \beta}}{h_\beta} \right),$$

where $\tilde{f}_0$ is the function belonging to $\mathcal{F}_0$ whose properties are specified in Theorem 2. The considered parametric set of functions, $\mathcal{F}_\beta$, is $\{f_{\eta_\beta}, \eta_\beta \in \Xi_\beta\}$. Clearly, there is a one-to-one application between Ξ_β and $\mathcal{F}_\beta$.

Second, let us define the set of β 's say $\mathcal{T}'$ on which we will consider the whole parametric family $\mathcal{F}$. Recall that β^* is the greatest regularity and consider $h_0 := h_{\beta^*} = (nt_n^{-1})^{-2/(4\beta^*+1)}$ the length of each A_{k, β^*} , $k \in \{1, \dots, m_{\beta^*}\}$. Next, let us define the scale on h_β , $\beta \in \mathcal{T}'$ such as $h_0 = 2h_{\beta_1}$

and $h_{\beta_j} = 2h_{\beta_{j+1}}$, for all $j \in \{1, \dots, c \log n\}$, where c is a positive constant depending only on $\beta_\star$, $\beta^\star$ and n . Then the whole parametric family is defined as follows

$$\mathcal{F} = \bigcup_{j \in \{0, \dots, c \log n\}} \mathcal{F}_{\beta_j}.$$

We have to check two conditions on the set $\mathcal{F}$: for all $j \in \{0, \dots, c \log n\}$,

$$\sup_{0 \leq x \leq 1} |f_{\eta_{\beta_j}}(x)| \leq M, \text{ and also if } \beta_j > 1 \sup_{0 \leq x \leq 1} |f'_{\eta_{\beta_j}}(x)| \leq M, \quad (15)$$

$$f_{\eta_{\beta_j}} \in \Sigma(\beta_j, L, M). \quad (16)$$

Relations (15) and (16) are derived from hypotheses on ϕ and $\tilde{f}_0$ exactly as relations (16) and (17) in Gayraud and Pouet (2001).

Third, we define a probability measure π on $\mathcal{F}$ as

$$\pi = \frac{1}{1 + c \log n} \sum_{j \in \{0, \dots, c \log n\}} \pi_{\beta_j},$$

where π_{β_j} is the probability measure on $\mathcal{F}_{\beta_j}$. For all $j \in \{0, \dots, c \log n\}$, let $\eta_1, \dots, \eta_{m_{\beta_j}}$ be i.i.d. taking values 1 and -1 with probability $1/2$, and let π_{β_j} be the corresponding probability measure. Note that π is such that

$$\frac{d\mathbb{P}_\pi}{d\mathbb{P}_{\tilde{f}_0}} = \frac{1}{1 + c \log n} \sum_{j \in \{0, \dots, c \log n\}} \frac{d\mathbb{P}_{\pi_{\beta_j}}}{d\mathbb{P}_{\tilde{f}_0}}, \quad \text{where} \quad \frac{d\mathbb{P}_{\pi_{\beta_j}}}{d\mathbb{P}_{\tilde{f}_0}} = E_{\pi_{\beta_j}} \left[\frac{d\mathbb{P}_{f_{\eta_{\beta_j}}}}{d\mathbb{P}_{\tilde{f}_0}} \right].$$

Fourth, for any $j \in \{0, \dots, c \log n\}$, denote $\Lambda_{n, \beta_j} = \mathcal{F}_{\beta_j} \cap \Lambda_{n, \beta_j}(av(nt_n^{-1}, \beta_j))$. For any test function Δ_n , the term $\sup_{f_0 \in \mathcal{F}_0} \mathbb{P}_{f_0}(\Delta_n = 1) + \sup_{j \in \{0, \dots, c \log n\}} \sup_{f \in \Lambda_{n, \beta_j}} \mathbb{P}_f(\Delta_n = 0)$ is bounded from below

by

$$\begin{aligned} & \mathbb{P}_{\tilde{f}_0}(\Delta_n = 1) + \frac{1}{c \log n + 1} \sum_{j \in \{0, \dots, c \log n\}} \sup_{f \in \Lambda_{n, \beta_j}} \mathbb{P}_f(\Delta_n = 0) \geq \\ & \mathbb{P}_{\tilde{f}_0}(\Delta_n = 1) + \frac{1}{c \log n + 1} \sum_{j \in \{0, \dots, c \log n\}} \int_{\Lambda_{n, \beta_j}} \mathbb{P}_f(\Delta_n = 0) d\pi_{\beta_j}(f) = \\ & \mathbb{P}_{\tilde{f}_0}(\Delta_n = 1) + \frac{1}{c \log n + 1} \sum_{j \in \{0, \dots, c \log n\}} \left\{ \int \mathbb{P}_f(\Delta_n = 0) d\pi_{\beta_j}(f) - \int_{(\Lambda_{n, \beta_j})^c} \mathbb{P}_f(\Delta_n = 0) d\pi_{\beta_j}(f) \right\} \geq \\ & 1 - \frac{1}{2} \text{var}(\mathbb{P}_{\tilde{f}_0}, \mathbb{P}_\pi) - \frac{1}{c \log n + 1} \sum_{j \in \{0, \dots, c \log n\}} \int_{(\Lambda_{n, \beta_j})^c} d\pi_{\beta_j}(f) = \\ & (1 - \frac{1}{2} \text{var}(\mathbb{P}_{\tilde{f}_0}, \mathbb{P}_\pi))(1 + o_n(1)), \quad \text{as } n \text{ large enough,} \end{aligned} \quad (17)$$

where $\text{var}(P, Q)$ is the total variation distance between the distributions P and Q (see Le Cam (1986), Chapter 4 for definition and properties) and where relation (17) holds provided that

$$\lim_{n \rightarrow \infty} \inf_{j \in \{0, \dots, c \log n\}} \pi_{\beta_j}(f \in \Lambda_{n, \beta_j}) = 1. \quad (18)$$

The proof of (18) follows from relation (19) in Gayraud and Pouet (2001) : we replace ψ_n by $v(nt_n^{-1}, \beta_j)$ and we consider the subset $\mathcal{F}_{0,j}^* \subset \mathcal{F}_0$ defined by $\mathcal{F}_{0,j}^* = \{f \in \mathcal{F}_0 : \|\tilde{f}_0 - f\|_2 \leq 2\kappa_{n,\beta_j}\lambda\}$ instead of Θ' . We construct an n^{-b} -net $\mathcal{N}_{0,\beta_j}^*$ in the supremum norm on the subset $\mathcal{F}_{0,j}^*$ with $b > 1/2$ and following the calculations until the inequality (22) in Gayraud and Pouet (2001), we get

$$\inf_{j \in \{0, \dots, c \log n\}} \pi_{\beta_j}(f \in \Lambda_{n,\beta_j}) \geq 1 - \exp(n^{br}) \exp(-n^{\frac{1}{4\beta^*+1}}), \quad (19)$$

where the last inequality is coarse since the diameter of the set $\mathcal{F}_{0,j}^*$ is not taken into account to calculate the entropy. The constraint $b < \frac{1}{r(4\beta^*+1)}$ is compatible with $b > \frac{1}{2}$ since a lower bound for the term $\frac{1}{r(4\beta^*+1)}$ is $\frac{1}{2r}$ which is strictly larger than $\frac{1}{2}$ (equivalent to $r < 1$). Therefore, the RHS of (19) tends to 1 as n goes to infinity : relation (18) is then satisfied. Fifth, we define a partition on $\mathcal{T}'$ into M_n sets $\mathcal{T}_l$, each containing N_n β 's with the following property :

$$\text{for all } \beta, \gamma \in \mathcal{T}_l : |\beta - \gamma| \geq c_1 \frac{\log(\log n)}{\log n},$$

where c_1 is a positive integer. Then, N_n is of order $\frac{\log n}{\log(\log n)}$ and M_n is of order $\log(\log n)$.

To conclude, it remains to prove in (17) that $\text{var}(\mathbb{P}_{\tilde{f}_0}, \mathbb{P}_\pi)$ goes to zero as n goes to infinity. We have,

$$\begin{aligned} \text{var}(\mathbb{P}_{\tilde{f}_0}, \mathbb{P}_\pi) &= \frac{1}{2} \int \left| \frac{d\mathbb{P}_\pi}{d\mathbb{P}_{\tilde{f}_0}} - 1 \right| d\mathbb{P}_{\tilde{f}_0} \\ \int \left| \frac{d\mathbb{P}_\pi}{d\mathbb{P}_{\tilde{f}_0}} - 1 \right| d\mathbb{P}_{\tilde{f}_0} &\leq \frac{1}{M_n} \sum_{l=1}^{M_n} \int \left| \frac{1}{N_n} \sum_{\beta \in \mathcal{T}_l} \frac{d\mathbb{P}_{\pi_\beta}}{d\mathbb{P}_{\tilde{f}_0}} - 1 \right| d\mathbb{P}_{\tilde{f}_0}, \\ &\leq \frac{1}{M_n} \sum_{l=1}^{M_n} \sqrt{\int \left(\frac{1}{N_n} \sum_{\beta \in \mathcal{T}_l} \frac{d\mathbb{P}_{\pi_\beta}}{d\mathbb{P}_{\tilde{f}_0}} - 1 \right)^2 d\mathbb{P}_{\tilde{f}_0}}. \end{aligned}$$

We focus only on the quantity

$$\begin{aligned} \int \left(\frac{1}{N_n} \sum_{\beta \in \mathcal{T}_l} \frac{d\mathbb{P}_{\pi_\beta}}{d\mathbb{P}_{\tilde{f}_0}} - 1 \right)^2 d\mathbb{P}_{\tilde{f}_0} &= T_{1,n} + T_{2,n}, \\ \text{where } T_{1,n} &= \frac{1}{N_n^2} \sum_{\beta \in \mathcal{T}_l} \int \left(\frac{d\mathbb{P}_{\pi_\beta}}{d\mathbb{P}_{\tilde{f}_0}} - 1 \right)^2 d\mathbb{P}_{\tilde{f}_0} \\ \text{and } T_{2,n} &= \frac{1}{N_n^2} \sum_{\substack{\beta, \gamma \in \mathcal{T}_l \\ \beta \neq \gamma}} \int \left(\frac{d\mathbb{P}_{\pi_\beta}}{d\mathbb{P}_{\tilde{f}_0}} - 1 \right) \left(\frac{d\mathbb{P}_{\pi_\gamma}}{d\mathbb{P}_{\tilde{f}_0}} - 1 \right) d\mathbb{P}_{\tilde{f}_0}. \end{aligned}$$

Let us study separately $T_{1,n}$ and $T_{2,n}$. One must first note that for any $l \in \{1, \dots, M_n\}$ and for any β in $\mathcal{T}_l$,

$$\begin{aligned} \frac{d\mathbb{P}_{\pi_\beta}}{d\mathbb{P}_{\tilde{f}_0}} &= E_{\pi_\beta} \left(\frac{d\mathbb{P}_{f_{\eta_\beta}}}{d\mathbb{P}_{\tilde{f}_0}} \right) \\ &= E_{\pi_\beta} \left(\prod_{i=1}^n \exp \left(\frac{1}{\sigma^2} \kappa_{n,\beta} (Y_i - \tilde{f}_0(x_i)) \sum_{s=1}^{m_\beta} \eta_{s,\beta} \phi \left(\frac{x_i - z_{s,\beta}}{h_\beta} \right) - \frac{1}{2\sigma^2} \kappa_{n,\beta}^2 \left(\sum_{s=1}^{m_\beta} \eta_{s,\beta} \phi \left(\frac{x_i - z_{s,\beta}}{h_\beta} \right) \right)^2 \right) \right) \\ &= \prod_{s=1}^{m_\beta} \exp \left(-\frac{1}{2\sigma^2} \kappa_{n,\beta}^2 \sum_{i \in I_{s,\beta}} \phi^2 \left(\frac{x_i - z_{s,\beta}}{h_\beta} \right) \right) E_{\pi_\beta} \left(\exp \left(\frac{1}{\sigma^2} \kappa_{n,\beta} \sum_{i \in I_{s,\beta}} (Y_i - \tilde{f}_0(x_i)) \eta_{s,\beta} \phi \left(\frac{x_i - z_{s,\beta}}{h_\beta} \right) \right) \right) \quad (20) \\ &= \prod_{s=1}^{m_\beta} \exp \left(-\frac{1}{2\sigma^2} \kappa_{n,\beta}^2 \sum_{i \in I_{s,\beta}} \phi^2 \left(\frac{x_i - z_{s,\beta}}{h_\beta} \right) \right) \cosh \left(\frac{1}{\sigma^2} \kappa_{n,\beta} \sum_{i \in I_{s,\beta}} (Y_i - \tilde{f}_0(x_i)) \phi \left(\frac{x_i - z_{s,\beta}}{h_\beta} \right) \right). \quad (21) \end{aligned}$$

Set $\frac{1}{\sigma^2} \kappa_{n,\beta} \sum_{i \in I_{s,\beta}} (Y_i - \tilde{f}_0(x_i)) \phi \left(\frac{x_i - z_{s,\beta}}{h_\beta} \right) = Z_s \kappa_{n,\beta} \frac{1}{\sigma} \sqrt{\sum_{i \in I_{s,\beta}} \phi^2 \left(\frac{x_i - z_{s,\beta}}{h_\beta} \right)}$, where Z_s is under $\mathbb{P}_{\tilde{f}_0}$

a standard normal real variable. Next applying the following equality,

$$\int \cosh^2(ux) \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{x^2}{2} \right) dx = \frac{1}{2} (1 + \exp(2u^2)),$$

we obtain

$$\int \left(\frac{d\mathbb{P}_{\pi_\beta}}{d\mathbb{P}_{\tilde{f}_0}} \right)^2 d\mathbb{P}_{\tilde{f}_0} = \prod_{s=1}^{m_\beta} \cosh \left(\kappa_{n,\beta}^2 \frac{1}{\sigma^2} \sum_{i \in I_{s,\beta}} \phi^2 \left(\frac{x_i - z_{s,\beta}}{h_\beta} \right) \right).$$

Then, from (21) and applying a classical majoration concerning cosh (see Ingster 1993), we get

$$\int \left(\frac{d\mathbb{P}_{\pi_\beta}}{d\mathbb{P}_{\tilde{f}_0}} \right)^2 d\mathbb{P}_{\tilde{f}_0} \leq \prod_{s=1}^{m_\beta} \exp \left(c_2 \kappa_{n,\beta}^4 \frac{1}{\sigma^4} \left(\sum_{i \in I_{s,\beta}} \phi^2 \left(\frac{x_i - z_{s,\beta}}{h_\beta} \right) \right)^2 \right). \quad (22)$$

Then, since ϕ is bounded, the choice of m_β and $\kappa_{n,\beta}$ yields the order of the right-hand side term in (22) as n large enough

$$\exp \left(c_3 m_\beta \kappa_{n,\beta}^4 \frac{1}{\sigma^4} \frac{n^2}{m_\beta^2} \right) = O((\log n)^{c_4}),$$

where c_4 is a positive constant depending on σ , $\|\phi\|_\infty$, λ and a . Therefore if we choose a small enough so that c_4 is strictly less than one, we get

$$T_{1,n} \xrightarrow{n \rightarrow +\infty} 0. \quad (23)$$

Rewrite $T_{2,n}$ as follows

$$\begin{aligned} T_{2,n} &= \frac{1}{N_n^2} \sum_{\substack{\beta, \gamma \in \mathcal{T}_l \\ \beta \neq \gamma}} \int \left(\frac{d\mathbb{P}_{\pi_\beta}}{d\mathbb{P}_{\tilde{f}_0}} \frac{d\mathbb{P}_{\pi_\gamma}}{d\mathbb{P}_{\tilde{f}_0}} - \frac{d\mathbb{P}_{\pi_\beta}}{d\mathbb{P}_{\tilde{f}_0}} - \frac{d\mathbb{P}_{\pi_\gamma}}{d\mathbb{P}_{\tilde{f}_0}} + 1 \right) d\mathbb{P}_{\tilde{f}_0} \\ &= \frac{1}{N_n^2} \sum_{\substack{\beta, \gamma \in \mathcal{T}_l \\ \beta \neq \gamma}} \int \left(\frac{d\mathbb{P}_{\pi_\beta}}{d\mathbb{P}_{\tilde{f}_0}} \frac{d\mathbb{P}_{\pi_\gamma}}{d\mathbb{P}_{\tilde{f}_0}} - 1 \right) d\mathbb{P}_{\tilde{f}_0}. \end{aligned}$$

Since the number of pairs $(\beta, \gamma) \in \mathcal{T}_l \times \mathcal{T}_l$ is of order N_n^2 as n is large enough, we want to treat for any pair $\mathcal{T}_l \times \mathcal{T}_l$, the term $\int \left(\frac{d\mathbb{P}_{\pi_\beta}}{d\mathbb{P}_{\tilde{f}_0}} \frac{d\mathbb{P}_{\pi_\gamma}}{d\mathbb{P}_{\tilde{f}_0}} - 1 \right) d\mathbb{P}_{\tilde{f}_0}$ which will be denoted $T_{2,\gamma,\beta,n}$ and we want to show that for any pair $(\beta, \gamma) \in \mathcal{T}_l \times \mathcal{T}_l$, the term $T_{2,\gamma,\beta,n}$ goes to 1 as n is increasing which implies that

$$\begin{aligned} T_{2,n} &= \frac{1}{N_n^2} \sum_{\substack{\beta, \gamma \in \mathcal{T}_l \\ \beta \neq \gamma}} \int \left(\frac{d\mathbb{P}_{\pi_\beta}}{d\mathbb{P}_{\tilde{f}_0}} \frac{d\mathbb{P}_{\pi_\gamma}}{d\mathbb{P}_{\tilde{f}_0}} - 1 \right) d\mathbb{P}_{\tilde{f}_0} \leq \sup_{\substack{\beta, \gamma \in \mathcal{T}_l \\ \beta \neq \gamma}} T_{2,\gamma,\beta,n} - 1 \\ &= o_n(1), \text{ as } n \rightarrow +\infty. \end{aligned} \quad (24)$$

Without loss of generality, suppose that $\beta < \gamma$ so that by construction, there exists a positive integer $u_{\gamma,\beta}$ satisfying $h_\gamma = u_{\gamma,\beta} h_\beta$ and $m_\gamma < m_\beta$. For all $k \in \{1, \dots, m_\gamma\}$, the interval $A_{k,\gamma}$ contains $u_{\gamma,\beta}$ intervals $(A_{s,\beta})_s$, then denote $I_{s,\beta}^{k,\gamma} = \{i : x_i \in A_{s,\beta} \subset A_{k,\gamma}\}$ and obviously $\text{card}(I_{s,\beta}^{k,\gamma}) = \frac{n}{m_\beta}$. Denote $(A_{s,\beta,\gamma})_s$ the intervals included in $A_{k,\gamma}$ and denote also by $z_{s,\beta,\gamma}$ the endpoints of these $(A_{s,\beta,\gamma})_s$. From previous calculations (20) and applying Fubini, we get

$$\begin{aligned} T_{2,\gamma,\beta,n} &= \int \prod_{s=1}^{m_\beta} \exp \left(-\frac{1}{2\sigma^2} \kappa_{n,\beta}^2 \sum_{i \in I_{s,\beta}} \phi^2 \left(\frac{x_i - z_{s,\beta}}{h_\beta} \right) \right) E_{\pi_\beta} \left(\exp \left(\frac{1}{\sigma^2} \kappa_{n,\beta} \sum_{i \in I_{s,\beta}} (Y_i - \tilde{f}_0(x_i)) \eta_{s,\beta} \phi \left(\frac{x_i - z_{s,\beta}}{h_\beta} \right) \right) \right) \\ &\times \prod_{k=1}^{m_\gamma} \exp \left(-\frac{1}{2\sigma^2} \kappa_{n,\gamma}^2 \sum_{i \in I_{k,\gamma}} \phi^2 \left(\frac{x_i - z_{k,\gamma}}{h_\gamma} \right) \right) E_{\pi_\gamma} \left(\exp \left(\frac{1}{\sigma^2} \kappa_{n,\gamma} \sum_{i \in I_{k,\gamma}} (Y_i - \tilde{f}_0(x_i)) \eta_{k,\gamma} \phi \left(\frac{x_i - z_{k,\gamma}}{h_\gamma} \right) \right) \right) d\mathbb{P}_{\tilde{f}_0} \\ &= \prod_{k=1}^{m_\gamma} \prod_{s=1}^{u_{\gamma,\beta}} \exp \left(-\frac{1}{2\sigma^2} \sum_{i \in I_{s,\beta}^{k,\gamma}} \left\{ \kappa_{n,\beta}^2 \phi^2 \left(\frac{x_i - z_{s,\beta,\gamma}}{h_\beta} \right) + \kappa_{n,\gamma}^2 \phi^2 \left(\frac{x_i - z_{k,\gamma}}{h_\gamma} \right) \right\} \right) \\ &\times E_{\pi_\beta} E_{\pi_\gamma} \int \exp \left(\frac{1}{\sigma^2} \sum_{i \in I_{s,\beta}^{k,\gamma}} (Y_i - \tilde{f}_0(x_i)) \left\{ \kappa_{n,\beta} \eta_{s,\beta} \phi \left(\frac{x_i - z_{s,\beta,\gamma}}{h_\beta} \right) + \kappa_{n,\gamma} \eta_{k,\gamma} \phi \left(\frac{x_i - z_{k,\gamma}}{h_\gamma} \right) \right\} \right) d\mathbb{P}_{\tilde{f}_0} \\ &= \prod_{k=1}^{m_\gamma} \prod_{s=1}^{u_{\gamma,\beta}} E_{\pi_\beta} E_{\pi_\gamma} \exp \left(\frac{1}{\sigma^2} \sum_{i \in I_{s,\beta}^{k,\gamma}} \kappa_{n,\beta} \eta_{s,\beta} \phi \left(\frac{x_i - z_{s,\beta,\gamma}}{h_\beta} \right) \kappa_{n,\gamma} \eta_{k,\gamma} \phi \left(\frac{x_i - z_{k,\gamma}}{h_\gamma} \right) \right) \\ &= \prod_{k=1}^{m_\gamma} \prod_{s=1}^{u_{\gamma,\beta}} \cosh \left(\frac{1}{\sigma^2} \sum_{i \in I_{s,\beta}^{k,\gamma}} \kappa_{n,\beta} \phi \left(\frac{x_i - z_{s,\beta,\gamma}}{h_\beta} \right) \kappa_{n,\gamma} \phi \left(\frac{x_i - z_{k,\gamma}}{h_\gamma} \right) \right) \\ &\leq \exp \left(c_5 m_\gamma u_{\gamma,\beta} \kappa_{n,\beta}^2 \kappa_{n,\gamma}^2 \frac{n^2}{m_\beta^2} \right), \end{aligned} \quad (25)$$

where the last inequality holds due to (22), and the positive constant c_5 depends on $\|\phi\|_\infty$ and σ^2 . Since for any β we have $\kappa_{n,\gamma}^2 \frac{n}{\sqrt{m_\beta}} = c_6 t_n$ where c_6 is a positive constant depending on a

and λ , the right hand side in (25) becomes

$$\begin{aligned}
\exp \left(c_6 m_\gamma u_{\gamma,\beta} \kappa_{n,\beta}^2 \kappa_{n,\gamma}^2 \frac{n^2}{m_\beta^2} \right) &= \exp \left(c_6 m_\gamma \frac{m_\beta}{m_\gamma} \kappa_{n,\beta}^2 \kappa_{n,\gamma}^2 \frac{n}{\sqrt{m_\beta}} \frac{n}{\sqrt{m_\gamma}} \frac{1}{m_\beta} \sqrt{\frac{m_\gamma}{m_\beta}} \right) \\
&= \exp \left(c_7 t_n^2 \sqrt{\frac{m_\gamma}{m_\beta}} \right) \\
&= \exp \left(c_7 t_n^2 \sqrt{\frac{m_\gamma}{m_\beta}} \right) \\
&= o_n \left((\log(\log n)) (\log n)^{-\delta} \right) \text{ as } n \rightarrow \infty,
\end{aligned} \tag{26}$$

where δ is a positive constant depending on the positive integer appearing in the difference separating β and γ , and c_7 depends on c_5 and c_6 . Relation (26) implies (24), which together with (23) yield the result and if we take a sufficiently small (6) holds with $\alpha_2 = 1$.

6.3 Proof of Proposition 1

For any $\beta \in \mathcal{T}$ and for any f in $\Sigma(\beta, C, M)$, we get

$$\begin{aligned}
\hat{\sigma}_n^2 &= \frac{1}{2(n-1)} \sum_{i=2}^n (Y_i - Y_{i-1})^2 \\
&= \frac{1}{2(n-1)} \sum_{i=2}^n (\xi_i - \xi_{i-1})^2 + \frac{1}{2(n-1)} \sum_{i=2}^n (f(x_i) - f(x_{i-1}))^2 \\
&\quad + \frac{1}{(n-1)} \sum_{i=2}^n (f(x_i) - f(x_{i-1})) (\xi_i - \xi_{i-1}).
\end{aligned} \tag{27}$$

It can be shown that under (A1.Sup), the first part in the RHS of (27) tends to σ^2 in probability as n goes to infinity. Indeed, if we split the first term in the RHS of (27) into two terms which depend on the quantities $(\xi_i - \xi_{i-1})^2$ with i even and with i odd respectively, then both terms are empirical mean of i.i.d. variables with expectation equal to $\sigma^2/2$; the weak large numbers law achieves the proof. For the third part, Chebyshev's inequality yields for any $\beta \in \mathcal{T}$ and for any f in $\Sigma(\beta, C, M)$ and any $\tau > 0$:

$$\mathbb{P}_f \left(\frac{1}{(n-1)} \sum_{i=2}^n (\xi_i - \xi_{i-1}) (f(x_i) - f(x_{i-1})) > \tau \right) = o_n(1) \text{ as } n \text{ goes to } +\infty.$$

Finally, for the second part,

$$|f(x_i) - f(x_{i-1})|^2 \leq \begin{cases} C^2 |x_i - x_{i-1}|^{2\beta} \leq C^2 n^{-2\beta_\star} & \text{if } \beta \leq 1, \\ M^2 |x_i - x_{i-1}|^2 \leq M^2 n^{-2} & \text{if } \beta > 1. \end{cases}$$

Therefore,

$$\frac{1}{2(n-1)} \sum_{i=2}^n (f(x_i) - f(x_{i-1}))^2 \leq \frac{n-1}{2(n-1)} (\sup(M, C))^2 n^{-2\beta_\star},$$

which goes to zero as $n \rightarrow +\infty$.

APPENDIX

Proof of Lemma 1 For all $\beta \in \mathcal{T}$ and for all $f \in \Lambda_n(Av(nt_n^{-1}, \beta))$, denote $\beta_L \in \mathcal{T}^*$ satisfying $\beta_L \leq \beta < \beta_{L+1}$, where β_{L+1} is also in $\mathcal{T}^*$; the proof of the lemma will be made in three steps :

-(a) First, we replace the infimum over $\mathcal{F}_0$ by an infimum over a set $\mathcal{N}_0$ whose cardinality is controlled by Assumption (A3). The set $\mathcal{N}_0$ will be defined latter.

-(b) Second, we obtain a lower bound for $T_{n,3,\beta_L,f_0^N}$, $f_0^N \in \mathcal{N}_0$.

-(c) Third, we use Bernstein's inequality.

(a) Due to Assumption (A3), and for all $f_0 \in \mathcal{F}_0$, there exists f_0^N such that $\sup_{x \in [0,1]} |f_0(x) - f_0^N(x)| \leq \delta$,

with $\delta = n^{-4\beta^*/(4\beta^*+1)-\kappa 2\beta_*/(4\beta^*+1)}$. We denote by $\mathcal{N}_0$ the smallest set including such functions f_0^N . Then,

$$\begin{aligned} |T_{n,2,\beta_L,f_0} - T_{n,2,\beta_L,f_0^N}| &= \left| 2 \frac{1}{m_L} \sum_{k=1}^{m_L} \frac{m_L^2}{\sigma^2 n^2} \left(\sum_{i \in I_k, j \in I_k, i \neq j} [f_0^N(x_i) - f_0(x_i)] \xi_j \right) \right| \\ &\leq \frac{2}{\sigma^2} \delta \sup_{i \in \{1, \dots, n\}} |\xi_i|. \end{aligned} \quad (28)$$

The same occurs for $T_{n,3,\beta_L,f_0}$ i.e.

$$\begin{aligned} |T_{n,3,\beta_L,f_0} - T_{n,3,\beta_L,f_0^N}| &= \left| \frac{1}{m_L} \sum_{k=1}^{m_L} \frac{m_L^2}{\sigma^2 n^2} \sum_{i \in I_k, j \in I_k, i \neq j} [f(x_i) - f_0^N(x_i)][f(x_j) - f_0^N(x_j)] - \right. \\ &\quad \left. [f(x_i) - f_0^N(x_i) + f_0^N(x_i) - f_0(x_i)][f(x_j) - f_0^N(x_j) + f_0^N(x_j) - f_0(x_j)] \right| \\ &\leq \delta^2 \frac{2}{\sigma^2} + \delta \frac{4}{\sigma^2} (2M + \delta) \\ &= \delta^2 \frac{6}{\sigma^2} + \delta \frac{8M}{\sigma^2}, \end{aligned} \quad (29)$$

where M is one of the constants which characterizes the functional class under the alternative.

Thus relations (28) and (29) lead to

$$\inf_{f_0 \in \mathcal{F}_0} (T_{n,2,\beta_L,f_0} + T_{n,3,\beta_L,f_0}) \geq \inf_{f_0^N \in \mathcal{N}_0} \left(T_{n,2,\beta_L,f_0^N} + T_{n,3,\beta_L,f_0^N} - \frac{2}{\sigma^2} \delta \sup_{i=1, \dots, n} |\xi_i| - \delta^2 \frac{6}{\sigma^2} - \delta \frac{8M}{\sigma^2} \right).$$

Denote $B(n) = n^{\kappa \frac{2\beta_*}{4\beta_*+1}}$, where $\kappa < 1$ is the positive constant defined in Assumption (A2.Sup). Next, consider $D_2 = \{ \sup_{i=1, \dots, n} |\xi_i| > B(n) \}$, then by Assumption (A2.Sup) and the choice of

$B(n)$, $\mathbb{P}(D_2) \rightarrow 0$ as $n \rightarrow +\infty$. Moreover the choice of $B(n)$ entails that $\frac{2}{\sigma^2} \delta B(n) + \delta^2 \frac{6}{\sigma^2} + \delta \frac{8M}{\sigma^2} = o_n(\rho_n \frac{\sqrt{m_L}}{n})$. To conclude step (a), just note that for all $\beta \in \mathcal{T}$ and uniformly over $\Lambda_n(Av(nt_n^{-1}, \beta))$, we have as n large enough

$$\mathbb{P}_f(D_{1,\beta}) \leq \mathbb{P}_f \left(\inf_{f_0^N \in \mathcal{N}_0} (T_{n,2,\beta_L,f_0^N} + T_{n,3,\beta_L,f_0^N}) < 3\rho_n \frac{\sqrt{m_L}}{n} \right). \quad (30)$$

(b) For all $\beta \in \mathcal{T}$ and for all $f \in \Lambda_n(Av(nt_n^{-1}, \beta))$, we get following from Gayraud and Pouet (2001), in both cases $\beta > 1$ and $\beta \leq 1$,

$$\begin{aligned} T_{n,3,\beta_L,f_0^N} &= T_{n,3,\beta_L,f_0^N} - T_{n,3,\beta_L,f_0} + T_{n,3,\beta_L,f_0} \\ &\geq \tilde{C}^2 A^2 m^{-2\beta} - \delta \\ &\geq \frac{\tilde{C}^2}{4} A^2 m_\beta^{-2\beta}, \end{aligned} \quad (31)$$

where the last inequality holds due to the value of δ and provided that $\beta_L \leq \beta$; the positive constant $\tilde{C}$ depends on β, β_L, C, M . For simplicity sake, denote $c^2 = \frac{\tilde{C}^2}{4}$. Then from (31),

$$\begin{aligned} \inf_{f_0^N \in \mathcal{N}_0} \left(T_{n,2,\beta_L,f_0^N} + T_{n,3,\beta_L,f_0^N} \right) &\geq cAm_\beta^{-\beta} \inf_{f_0^N \in \mathcal{N}_0} \left[\frac{T_{n,2,\beta_L,f_0^N}}{\sqrt{T_{n,3,\beta_L,f_0^N}}} + \sqrt{T_{n,3,\beta_L,f_0^N}} \right] \\ &\geq cAm_\beta^{-\beta} \left[\inf_{f_0^N \in \mathcal{N}_0} \frac{T_{n,2,\beta_L,f_0^N}}{\sqrt{T_{n,3,\beta_L,f_0^N}}} + cAm_\beta^{-\beta} \right]. \end{aligned}$$

Then, as $-\sup(-x) = \inf(x)$, we obtain

$$\begin{aligned} \mathbb{P}_f(D_{1,\beta}) &\leq \mathbb{P}_f \left(\frac{3\rho_n \frac{\sqrt{m_L}}{n}}{cAm_\beta^{-\beta}} \geq - \sup_{f_0^N \in \mathcal{N}_0} \left| \frac{T_{n,2,\beta_L,f_0^N}}{\sqrt{T_{n,3,\beta_L,f_0^N}}} \right| + cAm_\beta^{-\beta} \right) \\ &= \mathbb{P}_f \left(\sup_{f_0^N \in \mathcal{N}_0} \left| \frac{T_{n,2,\beta_L,f_0^N}}{\sqrt{T_{n,3,\beta_L,f_0^N}}} \right| \geq cAm_\beta^{-\beta} - \frac{12(\log(\log n))^{1/2} \sqrt{m_L}}{cAm_\beta^{-\beta} n} \right) \\ &\leq \sum_{f_0^N \in \mathcal{N}_0} \mathbb{P}_f \left(\left| \frac{T_{n,2,\beta_L,f_0^N}}{\sqrt{T_{n,3,\beta_L,f_0^N}}} \right| \geq cAm_\beta^{-\beta} - \frac{12(\log(\log n))^{1/2} \sqrt{m_L}}{ncAm_\beta^{-\beta}} \right). \end{aligned} \quad (32)$$

(c) Let us consider the event $D_{3,\beta} = \left\{ \left| \frac{T_{n,2,\beta_L,f_0^N}}{\sqrt{T_{n,3,\beta_L,f_0^N}}} \right| \geq cAm_\beta^{-\beta} - \frac{12(\log(\log n))^{1/2} \sqrt{m_L}}{ncAm_\beta^{-\beta}} \right\}$. We use the following lemma to prove that the event is negligible in probability.

Lemma 2 For all $\beta \in \mathcal{T}$ and for any $f \in \Lambda_n(Av(nt_n^{-1}, \beta))$, we get as n large enough,

$$\sum_{f_0^N \in \mathcal{N}_0} \mathbb{P}_f(D_{3,\beta} \cap D_2^c) = o_n(1),$$

where D_2^c is the complement of the event $D_2 = \left\{ \sup_{i=1,\dots,n} |\xi_i| > B(n) \right\}$ and $B(n) = n^{\kappa \frac{2\beta_\star}{4\beta_\star+1}}$.

Then, from relation (32) in step (b), and applying Lemma 2, Lemma 1 is proved.

Proof of Lemma 2 The main device is Bernstein's inequality. Let

$$Z_j = \frac{m_L}{n^2 \sigma^2} \sum_{i \in I_k} \frac{f(x_i) - f_0^N(x_i)}{\sqrt{T_{n,3,\beta_L,f_0^N}}} \xi_j, \quad \text{if } j \in I_k.$$

The random variables $Z_1, \dots, Z_n$ are i.i.d. with zero mean since $\mathbb{E}(\xi_i) = 0, \forall i \in \{1, \dots, n\}$. Similarly to Gayraud and Pouet (2001), we obtain an upper bound for Z_j , for all $j = 1, \dots, n$, that is

$$|Z_j| \leq K_1 B(n) \frac{\sqrt{m_L}}{\sigma^2 n}, \quad (33)$$

where K_1 is a positive constant. Also, we get

$$\sum_{j=1}^n \mathbb{E}_f \left((Z_j - \mathbb{E}_f(Z_j))^2 \right) \leq \frac{K_2}{n}, \quad (34)$$

where K_2 is a positive constant.

Next from (33) and (34) and applying Bernstein's inequality, we get an exponential upper bound,

$$\begin{aligned} \mathbb{P}_f \left(\left| \frac{T_{n,2,\beta_L,f_0^N}}{\sqrt{T_{n,3,\beta_L,f_0^N}}} \right| \geq cAm_\beta^{-\beta} - \frac{12(\log(\log n))^{1/2} \sqrt{m_L}}{ncAm_\beta^{-\beta}} \right) \\ \leq \exp \left(- \frac{\left(cAm_\beta^{-\beta} - \frac{12(\log(\log n))^{1/2} \sqrt{m_L}}{ncAm_\beta^{-\beta}} \right)^2}{\frac{(cAm_\beta^{-\beta} - \frac{12(\log(\log n))^{1/2} \sqrt{m_L}}{ncAm_\beta^{-\beta}}) K_1 \frac{B(n) \sqrt{m_L}}{\sigma^2 n}}{\frac{K_2}{n} + \frac{1}{3}}} \right), \quad (35) \end{aligned}$$

where $B(n) = n^{\kappa \frac{2\beta_*}{4\beta_*+1}}$ and with κ satisfying Assumption (A2.Sup). Such a choice for $B(n)$ yields $\mathbb{P}(\sup_{i=1,\dots,n} |\xi_i| > B(n)) = o_n(1)$ as $n \rightarrow +\infty$.

Let us consider the squared term in the numerator of the exponential term in the previous inequality and let us denote it N^2 and show that N is of order $m_\beta^{-\beta}$:

$$\begin{aligned} N &= m_\beta^{-\beta} [cA - 12t_n^{(4\beta+1)/(4\beta+1)} (cA)^{-1} n^{1/(4\beta_L+1)} n^{-(4\beta+1)/(4\beta+1)} n^{4\beta/(4\beta+1)} t_n^{-1/(4\beta_L+1)} t_n^{-4\beta/(4\beta+1)}] \\ &= m_\beta^{-\beta} [cA - 12(cA)^{-1} n^{1/(4\beta_L+1)} n^{-1/(4\beta+1)} t_n^{-1/(4\beta_L+1)} t_n^{1/(4\beta+1)}] \\ &= m_\beta^{-\beta} [cA - 12(cA)^{-1} \exp\{\log(n)4(\beta - \beta_L)/\iota\} t_n^{-4(\beta - \beta_L)/\iota}] \\ &= m_\beta^{-\beta} [cA - 12(cA)^{-1} \exp\{4a_0\} \exp\{-2(\log(\log(\log n)))(\beta - \beta_L)/\iota\}] \\ &\geq m_\beta^{-\beta} c \frac{A}{2}, \end{aligned}$$

where $\iota = ((4\beta_L + 1)(4\beta + 1))$ and a_0 are positive constants and where the last relation holds since we can choose A sufficiently large enough and since the last exponential term in the above equation is a constant smaller than one as n is large enough. Actually choose A sufficiently large such that $cA - 12(cA)^{-1} \exp\{4a_0\} \exp\{-4/\iota\} \geq A/2$. Then we obtain

$$\begin{aligned} \frac{(cAm_\beta^{-\beta} - \frac{12\sqrt{\log(\log n)m_L}}{ncAm_\beta^{-\beta}})^2}{\frac{K_2}{n}} &\sim n^{1/(4\beta+1)} t_n^{4\beta/(4\beta+1)}, \\ \frac{(cAm_\beta^{-\beta} - \frac{12\sqrt{\log(\log n)m_L}}{ncAm_\beta^{-\beta}})^2}{\frac{(cAm_\beta^{-\beta} - \frac{12\sqrt{\log(\log n)m_L}}{ncAm_\beta^{-\beta}}) K_1 \frac{B(n) \sqrt{m_L}}{\sigma^2 n}}{\frac{K_2}{n} + \frac{1}{3}}}} &\sim n^{(2\beta)/(4\beta+1)} n^{4(\beta_L - \beta)/\iota} (B(n))^{-1} t_n^{2\beta/(4\beta+1)} t_n^{1/(4\beta_L+1)}. \end{aligned}$$

The choice of m , t_n and $B(n)$ entails that the right-hand side in (35), which is independent of the regression function f , is decreasing to 0 exponentially. Finally, due to the previous argument and due to Assumption (A3), Lemma 2 follows. More precisely, the RHS in (35) is at least of order

$$\exp\left(-n^{\frac{2\beta_\star}{4\beta_\star+1}-\kappa\frac{2\beta_\star}{4\beta_\star+1}}\right),$$

which would be compared to the cardinality of $\mathcal{N}_0$ which is $\exp(\delta^{-r})$, with $\delta = n^{-\frac{4\beta_\star}{4\beta_\star+1}-\kappa\frac{2\beta_\star}{4\beta_\star+1}}$. Denote $R(\beta_\star, \beta^\star) = \frac{4\beta^\star}{4\beta_\star+1} + \kappa\frac{2\beta_\star}{4\beta_\star+1} = \frac{(16+8\kappa)\beta_\star\beta^\star+2(2\beta^\star+\kappa\beta_\star)}{(4\beta_\star+1)(4\beta^\star+1)}$. Then the relation between r , and $\beta^\star$, $\beta_\star$ and κ is

$$\begin{aligned} R(\beta_\star, \beta^\star)r &< \frac{2\beta_\star(1-\kappa)}{4\beta_\star+1} \\ &\Leftrightarrow \\ r &< \frac{2\beta_\star(1-\kappa)(1+4\beta^\star)}{(16+8\kappa)\beta_\star\beta^\star+2(2\beta^\star+\kappa\beta_\star)}. \end{aligned} \quad (36)$$

Relation (36) is exactly the required condition in Assumption (A3).

Acknowledgements. We are grateful to Oleg Lepskii for helpful discussions.

References

- [1] Baraud, Y., Huet, S., Laurent, B.: *Adaptive tests of linear hypothesis by model selection*, Ann. Stat., **31**, 1, 225-251 (2003)
- [2] Ermakov, M.S.: Minimax detection of a signal in a white Gaussian noise, *Theory Probab. Appl.*, **35**, 667–679 (1990)
- [3] Gayraud, G., Pouet, Ch.: *Minimax Testing composite null hypotheses in the discrete regression scheme*, Math. Methods Stat., **4**, 375–394 (2001)
- [4] Guerre, E., Lavergne, P.: *Optimal minimax rates for nonparametric specification testing in regression models*, Econom. Theory, **18**, 1139–1171 (2002)
- [5] Härdle, W., Kneip, A.: *Testing a regression model when we have smooth alternatives in mind*, Scand. J. Stat., **26**, 2, 221-238 (1999)
- [6] Härdle, W., Mammen, E.: *Comparing nonparametric versus parametric fits*, Ann. Stat., **21**, 4, 1926-1947 (1993)
- [7] Hart, J.D.: *Nonparametric Smoothing and Lack-of-fit Tests*. Springer, N.Y. e.a. 1997
- [8] Horowitz, J.L., Spokoiny, V.G.: *An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative*, Econometrica, **69**, 599–631 (2001)
- [9] Ingster, Yu.I.: On minimax nonparametric detection of a signal in Gaussian white noise, *Probl. Inf. Trans.*, **18** (2), 61–73 (1982)
- [10] Ingster, Yu.I.: Minimax detection of signals in L_p -metrics, *Zap. Nauchn. Semin. LOMI*, **184**, 152–168 (1990)

- [11] Ingster, Yu.I.: *Asymptotically minimax hypothesis testing for nonparametric alternatives, I, II, III.*, Math. Methods Stat., **2**, 85-114; 171-189; 249-268 (1993)
- [12] Le Cam, L.M.: *Asymptotic methods in statistical decision theory*, Springer Series in Statistics, Springer, New-York 1986
- [13] Lepskii, O.V.: On asymptotically exact testing of nonparametric hypotheses, *Discussion paper, Université de Louvain* 1993
- [14] Lepskii, O.V., Spokoiny, V.G.: *Minimax nonparametric hypothesis testing: the case of an inhomogeneous alternative*, Bernoulli, **5**, 333-358 (1999)
- [15] Lorentz, G.G., Golitschek M.V., Makovoz Y.: *Constructive Approximation: advanced problems*, Springer Verlag, Berlin 1996
- [16] Spokoiny, V.G.: *Adaptive hypothesis testing using wavelets*, Ann. Stat., **24**, 6, 2477-2498 (1996)
- [17] Spokoiny, V.G.: *Adaptive and spatially adaptive testing of a nonparametric hypothesis.*, Math. Methods Stat., **7**, 3, 245-273 (1998)
- [18] Suslina, I.A.: Minimax detection of a signal for l_p -ball, *Zap. Nauchn. Semin. POMI*, **207**, 127-137 (1993)