
DOI: https://doi.org/10.4213/tvp3003

Use of the all-Russian mathematical portal Math-Net.Ru implies that you have read and agreed to these terms of use
http://www.mathnet.ru/eng/agreement

Download details:
IP: 54.70.40.11
December 1, 2018, 15:24:17
LEBLANC F.*, LEPSKI O. V.**

TEST OF SYMMETRY IN NONPARAMETRIC REGRESSION

В модели непараметрической регрессии со случайным планом изучается задача проверки гипотезы о симметрии функции регрессии. В качестве альтернатив рассматриваются гладкие функции, принадлежащие некоторому шару в пространстве Гёльдера $H(\beta, L)$ и удаленные в $L_2$-норме от множества симметричных функций на величину $\sqrt{r_n}/2$. Показано, что гипотеза симметрии может быть состоятельно протестирована, если $r_n = O(n^{-4\beta/(4\beta+1)})$.

Ключевые слова и фразы: минимаксное различение гипотез, минимаксное решающее правило, гёльдеровский класс.

1. Introduction. As far as we know, the problem of testing symmetry of a curve in the nonparametric estimation has not been treated in the literature, except the case of density. A test for symmetry of the density in i.i.d. sampling was proposed by Y. Ingster [6]. Later another test for symmetry was proposed by I. A. Ahmad and Q. Li [1]. Let us also mention the papers [5], [2], [4], where the hypothesis of symmetry has appeared in order to treat some «econometric data».

The mathematical methods used in the paper to expand our test is a contribution to the minimax hypothesis testing approach introduced by Y. Ingster in the eighties (see for instance [6], [7]) and next developed by M. S. Ermakov [3] and more recently by O. Lepski and V. Spokoiny [9], [11], in the problem of detection of a signal in the Gaussian white noise model.

To provide the minimax theory we have to assume that the unknown function, under both hypotheses, belongs to a certain class of function defined by a ball in a set of smooth functions. Herein we consider a ball in a general Hölder space of the regularity $\beta > 1$.

Further on, the hypothesis of symmetry is referred to as null hypothesis. In the case under consideration, this hypothesis is composite:

$$H_0: f \in \Omega_0(L, \beta, M) = \Sigma(L, \beta) \cap \{f \mid \forall x, f(x) = f(-x)\},$$

*LMC-IMAG, Université Joseph Fourier, BP 53, 38041 Grenoble Cedex 09, France.
**LATP, Université de Provence, CMI, 39, rue F. Joliot Curie, 13543 Marseille Cedex 13, France; e-mail: Oleg.Lepski@cmi.univ-mrs.fr

1) The research was carried out within SFB 373 at Humboldt University, Berlin.
where \( \Sigma(L, \beta) \) is the ball of radius \( L \) in the Hölder space of regularity \( \beta > 1 \).

To study the power of the tests we introduce an alternative by

\[
H_1: f \in \Omega_1(L, \beta, M) = \Sigma(L, \beta) \cap \left\{ f \mid \|f(\cdot) - f(-\cdot)\|_2 \geq r_n \right\}.
\]

In particular, it implies that when \( f \) belongs to \( \Omega_1(L, \beta, M) \), it is separated from the set of symmetric functions by the \( L_2 \)-distance \( \sqrt{r_n/2} \).

We want to find the smallest value of the sensitivity parameter \( r_n \) that still allows one to test \( H_0 \) versus \( H_1 \) with the prescribed errors probabilities.

The quality of test is measured by the errors probabilities of the two types. The first one defines the critical area of the test. This probability is to reject the null hypothesis when it is true. We denote it by \( P_{H_0}\{T_n > T\} \), where \( T_n \) is a statistics based on the observations \( (X_i, Y_i), i = 1, \ldots, n \), and \( \{T_n > T\} \) defines the reject area of the null hypothesis.

The error probability of the second type (which permit to measure the power of the test) is the probability to accept the null hypothesis when the alternative holds. This error probability is denoted by \( P_{H_1}\{T_n \leq T\} \) and the complementary probability measures the power of the test.

The constant \( T = T(\alpha_1) \) defines the critical area of the test and it should be chosen such that the inequality

\[
\lim_{n \to \infty} \sup_{f \in \Omega_0(L, \beta, M)} P_{H_0}\{T_n > T\} \leq \alpha_1
\]

holds, where \( \alpha_1 \) is a given positive constant. In the other words, the error probability of the first type does not exceed, this is understood in the asymptotic sence, and uniformly over the set determining the null hypothesis the prescribed level \( \alpha_1 \). Moreover, we are also seeking a statistics \( T_n \) satisfying the following inequality with given level \( \alpha_2 > 0 \)

\[
\lim_{n \to \infty} \sup_{f \in \Omega_1(L, \beta, M)} P_{H_1}\{T_n \leq T\} \leq \alpha_2
\]

while the radius \( r_n \) should be as small as possible. Thus, we are interested in the least possible rate of \( r_n \) that guarantees the errors probabilities at the levels \( \alpha_1 \) and \( \alpha_2 \), respectively. This rate is called minimax. Its meaning is that under the other rate \( r'_n \) such that \( r'_n/r_n \) tends to zero the sum of the errors probabilities of the two types tends to one. Thus, under \( r'_n \), no consistent testing is possible. The main result of this paper claims that for \( r_n = Fn^{-4\beta/(4\beta+1)} \) with \( F \) large enough there exists a test based on \( T_n \) such that

\[
\lim_{n \to \infty} \sup_{f \in \Omega_0(L, \beta, M)} P_{H_0}\{T_n > T\} + \lim_{n \to \infty} \sup_{f \in \Omega_1(L, \beta, M)} P_{H_1}\{T_n \leq T\} \leq \alpha,
\]

where \( \alpha > 0 \) is fixed.
A first test satisfying the previous condition was developed by O. Lepski and F. Leblanc (see [8]). In this procedure the constant $T$ which defines the critical area of the test is not random but does depend on an unknown parameter involved in the definition of the considered Hölder class. Although this test achieves the minimax rate of testing it is hardly implementable for real data. Herein another minimax procedure of testing is developed, where the value $T$ which defines the critical area of the test does depend on the design. Hence in practical case, the test procedure requires to compute the statistics $T_n$ and the random value $T = T(X_1, \ldots, X_n)$.

The result that the rate $r_n = F n^{-\beta/((4\beta + 1))}$ cannot be improved in minimax sense might be obtained using the same arguments as in [6].

The paper is organized as follows. In Section 2 we present the problem setup and the test procedure. Then, in Section 3 we state Theorem 1, being the main result, which is proved in Section 4. The Appendix contains some auxiliary results.

2. Problem statement and test procedure. We consider the classical regression model:

$$Y_i = f(X_i) + \varepsilon_i, \quad i = 1, \ldots, n,$$

where the $X_i$'s are i.i.d. real random variables defined in a probability space $(\Omega, \mathcal{F}, P)$ and taking values in $[-1, 1]$, the probability density $p$ of $X_i$ being unknown. The variables $\varepsilon_i$'s are also i.i.d. centered random variables independent of $X_i$'s and such that $E \varepsilon_i^2 = \sigma^2$ and $E \varepsilon_i^4 = \mu_4 < \infty$. We suppose that $\sigma^2$ and $\mu_4$ are known.

We denote by $\Sigma(L, \beta, M), \beta = m + \alpha$ with $m > 0, 0 < \alpha \leq 1$, the set of functions $f$, with the $m$-th derivative satisfying for any $(x, y) \in [-1, 1]^2$:

$$|f^{(m)}(x) - f^{(m)}(y)| \leq L |x - y|^\alpha \quad \text{and such that } \|f\|_\infty \leq M.$$

The null hypothesis is presented by

$$H_0: f \in \Omega_0(L, \beta, M) = \Sigma(L, \beta, M) \cap \{f \mid \forall x, f(x) = f(-x)\}$$

and the alternative hypothesis is defined as:

$$H_1: f \in \Omega_1(L, \beta, M) = \Sigma(L, \beta, M) \cap \{f \mid \|f(\cdot) - f(-\cdot)\|_2^2 \geq F n^{-\beta/((4\beta + 1))}\}.$$

The statistic which underlies the test is based on a kernel estimator. Let $K$ be a convolution kernel. We denote $K_h(\cdot) = h^{-1} K(\cdot/h)$.

In what follows we denote $E_r$ the conditional expectation with respect to the noise given the design values.

In order to avoid technical problems due to the randomness of the design we base our statistic on a reconstructed design. These new observations will
be obtained as an average of initial observations over small and symmetric subintervals of size \( O(\ln n/n) \).

Let us define the partition \((I_j \cup I_{-j})_{j=1,\ldots,N}\) of \([-1,1]\) as: for any \( j = 1, \ldots, N-1 \) with \( N = \lfloor An/\ln n \rfloor \), \( A \) being a positive constant,

\[
I_j = \left[ (j-1) \frac{\ln n}{An}, j \frac{\ln n}{An} \right], \quad I_{-j} = \left[ -(j-1) \frac{\ln n}{An}, (j-1) \frac{\ln n}{An} \right],
\]

\[
I_N = \left[ (N-1) \frac{\ln n}{An}, 1 \right] \quad \text{and} \quad I_{-N} = \left[ -1, -(N-1) \frac{\ln n}{An} \right],
\]

where \([ \cdot ]\) denotes the integer part. From now on without loss of generality we can suppose that \( N = An/\ln n \) so that

\[
\forall j = 1, \ldots, N \quad I_j = \left[ \frac{j-1}{N}, \frac{j}{N} \right] \quad \text{and} \quad I_{-j} = \left[ -\frac{j}{N}, -\frac{j-1}{N} \right].
\]

Note that when the unknown density \( p \) of the design is bounded from below the statistics \( n^{-1} \sum_{k=1}^{n} \mathbb{I}_{(X_k \in I_j)} \) is nonequal to zero with overwhelming probability. Let us put

\[
A_j = \left\{ \frac{1}{n} \sum_{k=1}^{n} \mathbb{I}_{(X_k \in I_j)} \neq 0, \quad \frac{1}{n} \sum_{k=1}^{n} \mathbb{I}_{(X_k \in I_{-j})} \neq 0 \right\} \quad \text{and} \quad A_n = \bigcap_{j=1}^{N} A_j.
\]

We define the reconstructed observations \( Z_j \) obtained when using the observations \((X_k, Y_k)\) such that \( X_k \) belongs to \( I_j \cup I_{-j} \), that is

\[
Z_j = \frac{1}{n} \sum_{k=1}^{n} Y_k \left\{ \frac{\mathbb{I}_{(X_k \in I_j)}}{n^{-1} \sum_{k=1}^{n} \mathbb{I}_{(X_k \in I_j)}} - \frac{\mathbb{I}_{(X_k \in I_{-j})}}{n^{-1} \sum_{k=1}^{n} \mathbb{I}_{(X_k \in I_{-j})}} \right\} \mathbb{I}_{A_j}.
\]

We will put \( Z_j = 0 \) when the subinterval \( I_j \cup I_{-j} \) does not contain any point of the design. To simplify notation we will write \( Z_j = n^{-1} \sum_{k=1}^{n} u_{j,k} Y_k \) with

\[
v_{j,k} = \frac{\mathbb{I}_{(X_k \in I_j)}}{n^{-1} \sum_{k=1}^{n} \mathbb{I}_{(X_k \in I_j)}} \quad u_{j,k} = v_{j,k} - v_{-j,k}.
\]

Next we estimate the difference \( f(i/N) - f(-i/N) \) by \( f_i \), where

\[
f_i = \frac{1}{N} \sum_{j=1}^{N} K_h \left( \frac{i-j}{N} \right) Z_j.
\]

The statistic \( \hat{T}_n \) (which is an estimator of the \( L_2 \)-norm of \( f(\cdot) - f(-\cdot) \)) which is used in the test is defined as follows:

\[
\hat{T}_n = \frac{1}{N} \sum_{i=1}^{N} \left\{ \left( \frac{1}{N} \sum_{j=1}^{N} K_h \left( \frac{i-j}{N} \right) \frac{1}{n} \sum_{k=1}^{n} u_{j,k} Y_k \right)^2 \right. 
\]

\[
\left. - \frac{1}{N^2} \sum_{j=1}^{N} K_h^2 \left( \frac{i-j}{N} \right) \frac{\sigma^2}{n^2} \sum_{k=1}^{n} u_{j,k}^2 \right\}. \quad (2)
\]
Since the negative term in expression 2 is equal to the empirical average of the terms \( \mathbb{E}_\varepsilon \eta_i^2 \) (see Proposition 3) if we set

\[
\eta_i = \frac{1}{N} \sum_{j=1}^{N} K_h \left( \frac{i-j}{N} \right) \zeta_j, \quad \text{with} \quad \zeta_j = \frac{1}{n} \sum_{k=1}^{n} u_{j,k} e_k,
\]

the statistic \( \widehat{T}_n \) can also be written as

\[
\widehat{T}_n = \frac{1}{N} \sum_{i=1}^{N} (f_i^2 - \mathbb{E}_\varepsilon \eta_i^2).
\] (3)

Finally, we define the following statistic which only depends on the design

\[
\Theta = \left\{ \frac{1}{N^2} \sum_{i=1}^{N} \sum_{q=1}^{N} 2\sigma^4 \left[ \sum_{j=1}^{N} K_h \left( \frac{i-j}{N} \right) K_h \left( \frac{q-j}{N} \right) \sum_{k=1}^{n} u_{j,k}^2 \right]^2 + \frac{1}{N^2} \sum_{i=1}^{N} \sum_{q=1}^{N} \frac{\mu_4 - 3\sigma^4}{n^4 N^4} \left[ \sum_{j=1}^{N} K_h^2 \left( \frac{i-j}{N} \right) K_h^2 \left( \frac{q-j}{N} \right) \sum_{k=1}^{n} u_{j,k}^4 \right] \right\}^{1/2}. \] (4)

3. Main result. We make the following assumptions on the model and the convolution kernel:

(D1): The support of the density \( p \) is \([-1,1]\), and for all \( x \), \( p_1 > p(x) > p_0 > 0 \) for some constants \( p_0 \) and \( p_1 \).

(K1): \( K(\cdot) \) is continuously differentiable symmetric and compactly supported over \([-1,1]\). We put \( K = \max(\|K\|_{\infty}, \|K'\|_{\infty}) \).

(K2): \( \int u^k K(u) \, du = 0 \) for any \( k = 1, \ldots, m \) and \( \int_1^1 K(u) \, du = 1 \).

(K3): The kernel \( K \) is corrected around the boundaries of the interval \([-1,1]\) so that: \( \forall y < h \) or \( \forall y > 1 - h \), \( \int_0^1 K_h(y-x) \, dx = 1 \) and \( \int_0^1 (y-x)^k K_h(y-x) \, dx = 0 \) for all \( k = 1, \ldots, m \).

We denote by \( \mathbf{P}_{H_0}(C) \) (respectively, \( \mathbf{P}_{H_1}(C) \)) the probability of the set \( C \) when \( f \in \Omega_0(L, \beta, M) \) (respectively, \( f \in \Omega_1(L, \beta, M) \)). Now we state the following result.

**Theorem 1.** Assume that (D1), (K1), (K2), and (K3) are satisfied and that \( 0 < A \leq p_0(1 - e^{-1}) \). Choose \( 0 < \alpha_1 < 1 \), \( 0 < \alpha_2 < 1 \), and \( 1/(4\beta + 1) > \delta > 0 \). Put \( T_\delta(\alpha_1) = \Theta/\sqrt{\alpha_1} + n^{-\delta}(n\sqrt{h})^{-1} \).

If \( \beta > 1 \) and \( h = n^{-2/(4\beta + 1)} \), then we have

\[
\lim_{n \to \infty} \sup_{f \in \Omega_0(L, \beta, M)} \mathbf{P}_{H_0}(\widehat{T}_n > T_\delta(\alpha_1)) \leq \alpha_1
\] (5)

and if \( F > 16L^2 + C \) (where \( C = C(\alpha_2, \alpha_1, \sigma, p_0, A) \) is an explicit positive constant) we also have

\[
\lim_{n \to \infty} \sup_{f \in \Omega_1(L, \beta, M)} \mathbf{P}_{H_1}(\widehat{T}_n \leq T_\delta(\alpha_1)) \leq \alpha_2.
\] (6)
4. Proof of Theorem 1. We give here the main points which allow us to prove the result. The structure of the proof is as follows: we first state several claims (Propositions 1–5) and show how the theorem follows from them. Next in the Appendix we prove the propositions.

We use the following notation:

\[ \zeta_j = \frac{1}{n} \sum_{k=1}^{n} u_{j,k} \xi_k, \quad \nu_j = \beta_j + \tilde{\beta}_j, \quad \beta_j = f\left(\frac{j}{N}\right) - f\left(-\frac{j}{N}\right), \]

\[ \tilde{\beta}_j = \frac{1}{n} \sum_{k=1}^{n} u_{j,k} \left( f(X_k) - f\left(\frac{j}{N}\right) \right) - \frac{1}{n} \sum_{k=1}^{n} v_{-j,k} \left( f(X_k) - f\left(-\frac{j}{N}\right) \right), \]

\[ b_i = \frac{1}{N} \sum_{j=1}^{N} K_h \left( \frac{i-j}{N} \right) \beta_j, \quad r_i = \frac{1}{N} \sum_{j=1}^{N} K_h \left( \frac{i-j}{N} \right) \tilde{\beta}_j, \]

\[ \eta_i = \frac{1}{N} \sum_{j=1}^{N} K_h \left( \frac{i-j}{N} \right) \zeta_j. \]

Let us denote by \( T_n \) and \( \overline{T}_n \) the following statistics:

\[ T_n = \frac{1}{N} \sum_{i=1}^{N} \left( (b_i + \eta_i + r_i)^2 - \mathbb{E}_\xi \eta_i^2 \right), \]

\[ \overline{T}_n = \frac{1}{N} \sum_{i=1}^{N} \left( (b_i + \eta_i)^2 - \mathbb{E}_\xi \eta_i^2 \right) = S_n + B_n + C_n, \]

with

\[ S_n = \frac{1}{N} \sum_{i=1}^{N} (\eta_i^2 - \mathbb{E}_\xi \eta_i^2), \quad B_n = \frac{1}{N} \sum_{i=1}^{N} b_i^2, \quad C_n = \frac{2}{N} \sum_{i=1}^{N} \eta_i b_i. \quad (7) \]

Note that, over the set \( A_n \), \( n^{-1} \sum_{k=1}^{n} v_{j,k} = 1 \) and \( n^{-1} \sum_{k=1}^{n} v_{-j,k} = 1 \). Hence \( Z_j = \eta_j + \beta_j + \tilde{\beta}_j \) and \( f_i = b_i + r_i + \eta_i \). So the statistic \( \overline{T}_n \) is equal to \( T_n \) over the set \( A_n \). Next since the statistics \( \overline{T}_n \) and \( T_n \) are rather close (as shown in Proposition 2) it will be enough to prove (5) and (6) when \( \overline{T}_n \) is substituted with \( T_n \). We have then

\[ T_n - \overline{T}_n = \frac{1}{N} \sum_{i=1}^{N} r_i^2 + \frac{2}{N} \sum_{i=1}^{N} b_i r_i + \frac{2}{N} \sum_{i=1}^{N} \eta_i r_i = R_n + U_n + V_n. \quad (8) \]

Note that, although \( \overline{T}_n \) is equal to \( T_n \) only over \( A_n \), the statistics \( S_n, C_n, R_n, U_n \) and \( V_n \) are well defined over \( A_n \cup A_n' \). To prove the result we use the following subresults.

Proposition 1. Assume (D1) is satisfied and \( 0 < A \leq p_0(1-1/e) \), then

\[ \lim_{n \to \infty} \mathbb{P}(A_n) = 1. \]
Proposition 2. Assume (D1) and (K1) are satisfied. We put $N = An/\ln n$ and $h$ as defined in Theorem 1. If $\delta < 1/(4\beta + 1)$ we have
\[
\sup_{f \in \Omega_0(L, \beta, M)} P \left\{ n \sqrt{h} |T_n - \bar{T}_n| > n^{-\delta} \right\} = o(1),
\]
\[
\sup_{f \in \Omega_1(L, \beta, M)} P \left\{ n \sqrt{h} |T_n - \bar{T}_n| > n^{-\delta} \right\} = o(1).
\]

Proposition 3.

i) $E \eta_i^2 = \frac{1}{N^2} \sum_{j=1}^{N} K_h \left( \frac{i-j}{N} \right) \frac{\sigma^2}{n^2} \sum_{k=1}^{n} u_{j,k}^2$

ii) $\Theta^2 = E \left( \frac{1}{N} \sum_{i=1}^{N} (\eta_i^2 - E \eta_i^2) \right) = E S_n^2$. (9)

Proposition 4. Suppose (D1), (K1), (K2), and (K3) are satisfied. Then
\[
n \sqrt{h} B_n \geq n \sqrt{h} \left( \frac{1}{4} \|f(\cdot) - f(\cdot)\| - 4L^2 h^{2\beta} + o(h^{2\beta}) \right)
\]
and hence if $h = n^{-2\beta/(4\beta + 1)}$, then under the alternative for any $f \in \Omega_1(L, \beta, M)$ we have $n \sqrt{h} B_n \geq F_0 + o(1)$, where $F_0 = F/4 - 4L^2$.

Proposition 5. Suppose (D1) and (K1) are satisfied. Set $\Lambda_n = \{|C_n| < \lambda B_n\}$ and $\Gamma_n = \{\Theta < B_n/\gamma\}$ for two positive constants $\gamma$ and $\lambda$. Moreover we put $c = 1 - 1/e - A/p_0$ and $\gamma_0 = \gamma_0^2 c^2 F_0^2 / (320 \sigma^4)$. Then under alternative we have for any $\lambda > 0$ and $0 < \gamma \leq \gamma_0$:
\[
\sup_{f \in \Omega_1(L, \beta, M)} P_{H_1}(\Gamma_n) = o(1) \quad \text{and} \quad \sup_{f \in \Omega_1(L, \beta, M)} P_{H_1}(\Lambda_n^c) = o(1). \quad (10)
\]

Now we show how the result of Theorem 1 follows from the previous propositions.

Proof of (5). Let us prove that the first type error is asymptotically smaller than a fixed value $\alpha_1$. Recall that on $A_n$, $\bar{T}_n = T_n$. Moreover, under the null hypothesis $\bar{T}_n = S_n + B_n + C_n = S_n$. We have
\[
P_{H_0}(\bar{T}_n > T_\delta(\alpha_1)) \leq P_{H_0}(\bar{T}_n > T_\delta(\alpha_1), A_n) + P_{H_0}(A_n^c)
\]
\[
\leq P_{H_0}\left\{ n \sqrt{h} T_n > n \sqrt{h} T_\delta(\alpha_1), A_n \right\} + P_{H_0}(A_n^c)
\]
\[
\leq P_{H_0}\left\{ n \sqrt{h} (T_n - \bar{T}_n) > n^{-\delta}, A_n \right\} + P_{H_0}\left\{ S_n > \frac{\Theta}{\sqrt{\alpha_1}}, A_n \right\} + P_{H_0}(A_n^c)
\]
\[
\leq P_{H_0}\left\{ n \sqrt{h} (T_n - \bar{T}_n) > n^{-\delta} \right\} + P_{H_0}\left\{ S_n > \frac{\Theta}{\sqrt{\alpha_1}} \right\} + P_{H_0}(A_n^c).
\]
Here we use the fact that $S_n$ is well defined over $A_n$ and by the Chebyshev inequality

$$P_{H_0} \{ \hat{T}_n > T_\delta(\alpha_1) \} \leq P_{H_0} \left\{ \left( T_n - \bar{T}_n \right) > \frac{n^{-\delta}}{n\sqrt{h}} \right\} + \frac{E_c S_n^2}{\Theta^2 \alpha_1^{-1}} + P_{H_0}(A_n^c). \quad (11)$$

Next, we substitute the bounds given by Propositions 1, 2, and 3 into inequality (11) and take the supremum over the set $\Omega_0(L, \beta, M)$. Finally we pass to the limit when $n$ tends to infinity to conclude the proof of (5).

Proof of (6).

$$P_{H_1} \{ \hat{T}_n \leq T_\delta(\alpha_1) \} \leq P_{H_1} \left\{ n\sqrt{h} \hat{T}_n \leq n\sqrt{h} T_\delta(\alpha_1), A_n \right\} + P_{H_1}(A_n^c)$$

$$\leq P_{H_1} \left\{ n\sqrt{h} T_n \leq n\sqrt{h} T_\delta(\alpha_1) + n^{-\delta} \right\}$$

$$+ P_{H_1} \left\{ n\sqrt{h} (T_n - \bar{T}_n) < -n^{-\delta} \right\} + P_{H_1}(A_n^c)$$

$$\leq P_{H_1} \left\{ S_n + B_n + C_n \leq T_\delta(\alpha_1) + (n\sqrt{h})^{-1} n^{-\delta} \right\}$$

$$+ P_{H_1} \left\{ n\sqrt{h}|T_n - \bar{T}_n| > n^{-\delta} \right\} + P_{H_1}(A_n^c)$$

$$\leq P_{H_1} \left\{ S_n \leq \frac{\Theta}{\sqrt{\alpha_1}} + 2(n\sqrt{h})^{-1} n^{-\delta} - B_n - C_n, \Lambda_n \right\}$$

$$+ P_{H_1} \left\{ n\sqrt{h}|T_n - \bar{T}_n| > n^{-\delta} \right\} + P_{H_1}(A_n^c)$$

$$\leq P_{H_1} \left\{ S_n \leq \frac{\Theta}{\sqrt{\alpha_1}} + 2(n\sqrt{h})^{-1} n^{-\delta} + (\lambda - 1) B_n \right\}$$

$$+ P_{H_1} \left\{ n\sqrt{h}|T_n - \bar{T}_n| > n^{-\delta} \right\} + P_{H_1}(A_n^c)$$

$$\leq P_{H_1} \left\{ S_n \leq \frac{\Theta}{\sqrt{\alpha_1}} + 2(n\sqrt{h})^{-1} n^{-\delta} + (\lambda - 1) B_n, \Gamma_n \right\}$$

$$+ P_{H_1} \left\{ n\sqrt{h}|T_n - \bar{T}_n| > n^{-\delta} \right\} + P_{H_1}(\Gamma_n^c) + P_{H_1}(A_n^c) + P_{H_1}(A_n^c). \quad (12)$$

Let us now choose $\gamma$ such that $1/\sqrt{\alpha_1} + \gamma(\lambda - 1) < 0$. Then for $n$ large enough

$$\Theta \left( \frac{1}{\sqrt{\alpha_1}} + \gamma(\lambda - 1) \right) + 2(n\sqrt{h})^{-1} n^{-\delta} \leq \Theta \left( \frac{1}{\sqrt{\alpha_1}} + \gamma(\lambda - 1) \right).$$
It follows from (12) that

\[
\mathbb{P}_{H_1}\{\hat{T}_n < T_\delta(\alpha_1)\} \leq \mathbb{P}_{H_1}\left\{ |S_n| \geq \frac{-\Theta}{2} \left( \frac{1}{\sqrt{\alpha_1}} + (\lambda - 1)\gamma \right) \right\} \\
+ \mathbb{P}_{H_1}\left\{ n\sqrt{h} |T_n - \bar{T}_n| > n^{-\delta} \right\} + \mathbb{P}_{H_1}(\Gamma_n^c) + \mathbb{P}_{H_1}(A_n^c) + \mathbb{P}_{H_1}(A_n^c).
\]

Then due to the Chebyshev inequality and upper bounds given in Propositions 1, 2, and 5, for any \(0 < \lambda < 1\) and any \((\sqrt{\alpha_1}(1 - \lambda))^{-1} \leq \gamma \leq \gamma_0 = p_0^2c^2F_0^2/(320\sigma^4)\) we get

\[
\mathbb{P}_{H_1}\{\hat{T}_n < T_\delta(\alpha_1)\} \leq 4 \left( \frac{1}{\sqrt{\alpha_1}} + (\lambda - 1)\gamma \right)^{-2} + o(1). \tag{13}
\]

When taking \(\lambda = \frac{1}{2}\) and \(\gamma = (1 - \lambda)^{-1}(1/\sqrt{\alpha_1} + 2/\sqrt{\alpha_2})\) the first term in the right-hand side of inequality (13) is bounded by \(\alpha_2\) (note that for this choice of the constant \(\gamma\) the condition \((\lambda - 1)\gamma + 1/\sqrt{\alpha_1}\) obviously holds). Moreover, if \(F > 16L^2 + C\) with \(C = 32\sqrt{10}(1/\sqrt{\alpha_1} + 2/\sqrt{\alpha_2})^{1/2}\sigma^2/(p_0(1 - 1/e) - A/p_0)\), then the condition \(\gamma \leq \gamma_0\) is satisfied. Hence, for any \(f \in \Omega_1(L, \beta, M)\) and for any \(n\) large enough \(\mathbb{P}_{H_1}\{\hat{T}_n < T_\delta(\alpha_1)\} \leq \alpha_2 + o(1)\).

Now we take the supremum over the set \(\Omega_1(L, \beta, M)\) and the limit when \(n\) tends to infinity in the previous inequality what implies (6).

5. Appendix.

5.1. Preliminary results. Let us state two lemmas which will be very useful in the proofs of our five propositions.

Lemma 1. Let \(\xi_k^{(n)}\), \(k = 1, \ldots, n\), be i.i.d. Bernoulli random variables such that \(\mathbb{P}\{\xi_k^{(n)} = 1\} = p_n\). Then \(\forall c \in \mathbb{R}\)

\[
\mathbb{P}\left\{ \sum_{k=1}^{n} \xi_k^{(n)} < cnp_n \right\} \leq \exp(-np_n(1-c-e^{-1})). \tag{14}
\]

As a corollary of Lemma 1 we have Lemma 2.

Lemma 2. Suppose that assumption (D1) is satisfied and \(N = An/\ln n\) with \(0 < A < (1-e^{-1})p_0\). Denote \(p_n^j = \mathbb{P}(I_j)\) for any \(j = -N, \ldots, N\),

\[
M_n = \bigcap_{j=-N}^{N} \left\{ \frac{1}{n} \sum_{k=1}^{n} \mathbb{I}(X_k \in I_j) \geq c'p_n^j \right\}.
\]

Then for any \(0 < c' \leq 1 - 1/e - A/p_0\), \(\lim_{n \to +\infty} \mathbb{P}(M_n) = 1\).

The proofs of Lemmas 1 and 2 can be omitted, because it is just the application of the exponential Markov inequality to the sum of independent bounded random variables.
To prove Propositions 2 and 5 we have to supply some upper bound for the terms \( u_{j,k} \) which contain random denominators. We state the following assertions which hold true at least on the set \( M_n \). Recall that

\[ v_{j,k} = \frac{\Pi_{\{X_k \in I_j\}}}{n^{-1} \sum_k \Pi_{\{X_k \in I_j\}}} \quad \text{and} \quad u_{j,k} = v_{j,k} - v_{-j,k}. \]

**Lemma 3.** Suppose that \((D1)\) is satisfied and set \( N = An/\ln n \) with \( A \leq p_0(1 - e^{-1}) \). Then
i) \( \forall k \in \{1, \ldots, n\} \) and \( \forall j \in \{1, \ldots, N\} \),

\[ \mathbb{E}(u_{j,k}^2 \Pi_{M_n}) \leq \frac{2A}{p_0 c^2 \ln n} = \frac{2}{p_0 c^2} N \]

and over the set \( M_n \):

ii) \[ \sum_k u_{j,k}^2 \leq \frac{2A}{p_0 c \ln n} = \frac{2}{p_0 c} nN, \]

iii) \[ \sum_k u_{j,k}^4 \leq \frac{4A^2}{p_0^3 c^3 (\ln n)^3} = \frac{4}{p_0^3 c^3} nN^3. \]

**Proof.** Since \( I_j \) and \( I_{-j} \) have an empty intersection we obviously have \( u_{j,k}^2 = v_{j,k}^2 + v_{-j,k}^2 \). Moreover, on the set \( M_n \) it holds: \( \forall j, \, c \Pi_{\{X_k \in I_j\}} \leq n^{-1} \sum_k \Pi_{\{X_k \in I_j\}} \), hence, \( v_{j,k}^2 \leq \Pi_{\{X_k \in I_j\}}/(c \Pi_{\{X_k \in I_j\}})^2 \).

Next, since \( p_0 \ln n/(An) \leq p_1 \ln n/(An) \), when summing up the expectation of \( v_{j,k}^2 \) and \( v_{-j,k}^2 \) we conclude i).

To prove ii) and iii) note that on the set \( M_n \)

\[ v_{j,k}^2 = \frac{\Pi_{\{X_k \in I_j\}}}{(n^{-1} \sum_k \Pi_{\{X_k \in I_j\}})^2} \quad \text{and} \quad v_{j,k}^4 = \frac{\Pi_{\{X_k \in I_j\}}}{(n^{-1} \sum_k \Pi_{\{X_k \in I_j\}})^4}. \]

When summing up over \( k \) we obtain

\[ \sum_k v_{j,k}^2 = \frac{1}{n} \sum_k \Pi_{\{X_k \in I_j\}} \leq \frac{A}{p_0 c \ln n} = \frac{1}{p_0 c} nN, \]

\[ \sum_k v_{j,k}^4 = \frac{1}{(n^{-1} \sum_k \Pi_{\{X_k \in I_j\}})^3} \leq \frac{A^3}{p_0^3 c^3 (\ln n)^3} = \frac{1}{p_0^3 c^3} nN^3. \]

**Lemma 4.** Let \( X = (X_1, \ldots, X_n) \) and \( \varepsilon = (\varepsilon_1, \ldots, \varepsilon_n) \) be two independent random vectors and \( M \) a measurable set which depends only on \( X \). Then for any statistics \( g(X, \varepsilon) \) and any \( c > 0 \) we have

\[ \mathbb{P}\{g(X, \varepsilon) > c\}, \ M \} \overset{\text{def}}{=} \mathbb{P}\{g(X, \varepsilon) > c\} \cap M \} \leq \mathbb{E}_X \left( \frac{\mathbb{E}_c g^2(X, \varepsilon)}{c^2} \Pi_M \right), \]

where \( \mathbb{E}_X \) (respectively, \( \mathbb{E}_c \)) denotes the expectation with respect to \( X \) (respectively, \( \varepsilon \)).
This lemma simply states the conditional Chebyshev inequality so we omit its proof.

The following lemma gives the moments of orders two and four of the reconstructed noise \((\zeta_j)_{j=1,\ldots,N}\) defined as \(\zeta_j = n^{-1} \sum_{k=1}^{n} u_{j,k} \varepsilon_k\).

**Lemma 5.** For any \((j, s, k, m) \in \{1, \ldots, N\}^4\) the following equalities hold:

i) \[E_e(\zeta_j \zeta_s) = 0 \text{ if } j \neq s,\]

ii) \[E_e(\zeta_j^2) = \sigma^2 n^{-2} \sum_{k=1}^{n} u_{j,k}^2,\]

iii) \[E_e(\zeta_j \zeta_s \zeta_k \zeta_m) = 0 \text{ if } j \neq s \neq p \neq m,\]

iv) \[E_e(\zeta_j^2 \zeta_s^2) = E_e(\zeta_j^2) E_e(\zeta_s^2) \text{ if } j \neq s,\]

v) \[E_e(\zeta_j \zeta_s^3) = 0 \text{ if } j \neq s,\]

vi) \[E_e(\zeta_j^4) = \frac{3\sigma^4}{n^4} \left( \sum_{k=1}^{n} u_{j,k}^2 \right)^2 + \frac{\mu_4 - 3\sigma^4}{n^4} \sum_{k=1}^{n} u_{j,k}^4 = 3 \left( E_e(\zeta_j^2) \right)^2 + \frac{\mu_4 - 3\sigma^4}{n^4} \sum_{k=1}^{n} u_{j,k}^4.\]

**Proof.** Recall that \((A_j)_{j=1,\ldots,N}\) is a partition of the interval \([-1,1]\). Hence the variables \((X_k, \varepsilon_k)\) which contribute to the construction of \(\zeta_j\) do not contribute to the construction of \(\zeta_s\) for \(s \neq j\). Furthermore, the variables \(X_k\) and \(\varepsilon_k\) are supposed to be mutually independent thus by construction \((\zeta_j)_{j}\) is a sequence of independent random variables.

Since \(E_e \varepsilon_k = 0\) we obviously have that \(E_e \zeta_j = 0\) for any \(j\). So the assertions i), iii)–v) are obvious.

We first prove ii):

\[E_e(\zeta_j^2) = \frac{1}{n^4} \sum_{k=1}^{n} \sum_{l=1}^{n} u_{j,k} u_{j,l} E_e(\varepsilon_k \varepsilon_l) = \frac{1}{n^2} \sum_{k=1}^{n} u_{j,k}^2 E_e(\varepsilon_k^2) = \frac{\sigma^2}{n^2} \sum_{k=1}^{n} u_{j,k}^2.\]

Let us now prove the last assertion vi). Let \((J_s)_{s=1,\ldots,5}\) be the partition of the set of indices \(J = \{(k, m, l, t) \in \{1, \ldots, n\}^4\}\) defined as

\[J_1 = \{(k, k, l, l) \mid k \neq l, (k, l) \in \{1, \ldots, n\}^2\},\]

\[J_2 = \{(k, l, k, l) \mid k \neq l, (k, l) \in \{1, \ldots, n\}^2\},\]

\[J_3 = \{(k, l, l, k) \mid k \neq l, (k, l) \in \{1, \ldots, n\}^2\},\]

\[J_4 = \{(k, k, k, k) \mid k \in \{1, \ldots, n\}\}.\]

Then we have

\[E_e(\zeta_j^4) = \frac{1}{n^4} \sum_{(k, m, l, t) \in J} u_{j,k} u_{j,m} u_{j,l} u_{j,t} E_e(\varepsilon_k \varepsilon_m \varepsilon_l \varepsilon_t) = \frac{1}{n^4} \sum_{(k, m, l, t) \in J_1 \cup J_2 \cup J_3} u_{j,k}^2 u_{j,l}^2 E_e(\varepsilon_k^2 \varepsilon_l^2) + \frac{1}{n^4} \sum_{(k, m, l, t) \in J_4} u_{j,k}^4 E_e(\varepsilon_k^4) = \Sigma_1 + \Sigma_2.\]
The summation over the set \( J_5 \) is equal to zero since the variables \( \varepsilon_k \) are independent and centered. The first summation \( \Sigma_1 \) can be rewritten as the following expression:

\[
\Sigma_1 = \frac{3}{n^2} \sum_{k=1}^{n} \sum_{i=1, j \neq k}^{n} u_{j,k}^2 u_{j,i}^2 \sigma^4 = \frac{3}{n^2} \sum_{k=1}^{n} \left( \sum_{i=1}^{n} u_{j,k}^2 u_{j,i}^2 - u_{j,k}^4 \right) 
\]

\[
= \frac{3\sigma^4}{n^4} \left( \sum_{k=1}^{n} u_{j,k}^2 \right)^2 - \frac{3\sigma^4}{n^4} \sum_{k=1}^{n} u_{j,k}^4. 
\] (15)

For the second term \( \Sigma_2 \) we have immediately

\[
\Sigma_2 = \frac{\mu_4}{n^4} \sum_{k=1}^{n} u_{j,k}^4. 
\] (16)

Finally, summing up the equalities (15) and (16) we conclude the proof of vi).

5.2. Proof of Proposition 1. Let us show that \( \lim P(A_n^c) = 0 \) when \( n \) tends to infinity. Note that

\[
P(A_n^c) = P\left\{ \bigcup_{j=1}^{N} A_j^c \right\} \leq \sum_{j=1}^{N} P(A_j^c)
\]

\[
\leq \sum_{j=1}^{N} P\left\{ \sum_{\{X_k \in I_j\}} = 0 \right\} + \sum_{j=1}^{N} P\left\{ \sum_{\{X_k \in I_{-j}\}} = 0 \right\}. 
\] (17)

We now estimate \( P\{\sum_{\{X_k \in I_j\}} = 0 \} \) using the Markov inequality, as in Lemma 1. For any \( \varepsilon < 1 \)

\[
P\left\{ \sum_{\{X_k \in I_j\}} = 0 \right\} \leq e^{\varepsilon} e^{-np_0(1-\varepsilon^{-1})}
\]

(recall that we have denoted \( p_0 = P(I_j) \)). The density \( p \) of the design being bounded from below by \( p_0 \), for any \( j = 1, \ldots, N \), we have \( p_n \geq p_0 \ln n/(An) \); therefore, for any \( j \)

\[
P\left\{ \sum_{\{X_k \in I_j\}} = 0 \right\} \leq e^{\varepsilon} n^{-p_0(1-1/\varepsilon)/A}.
\]

The second term in the right-hand side of (17) is estimated in the same way because \( P(I_{-j}) \) is also bounded from below by \( p_0 \ln n/(An) \). Finally, when substituting these bounds in (17) we obtain

\[
P(A_n^c) \leq 2e^\varepsilon N n^{-p_0(1-1/\varepsilon)/A} = 2e^\varepsilon A n^{1-p_0(1-1/\varepsilon)/A} \ln^{-1} n.
\]

If \( 0 < \varepsilon \leq p_0(1 - 1/\varepsilon) \) the last term of the previous inequality tends to zero as \( n \) goes to infinity.

5.3. Proof of Proposition 2. By equality (8) we have \( T_n - \hat{T}_n = R_n + U_n + V_n \) and under both hypotheses \( H_0 \) and \( H_1 \) we have

\[
P\{n\sqrt{h}|T_n - \hat{T}_n| > n^{-\delta}\} \leq P\{n\sqrt{h}R_n > n^{-\delta}/3\} + P\{n\sqrt{h}|U_n| > n^{-\delta}/3\}
\]

\[
+ P\{n\sqrt{h}|V_n| > n^{-\delta}/3\}. 
\] (18)
To complete the proof it suffices to show that each term in the right-hand side of (18) goes to zero as \(n\) tends to infinity.

We start with the following result on the elements \(b_i, r_i,\) and \(\eta_i\) which define \(R_n, U_n,\) and \(V_n.\)

**Lemma 6.** Let assumptions of Proposition 2 hold. Then
i) \(\forall f \in \Sigma(L, \beta, M)\) and \(i = 1, \ldots, N, |r_i| \leq 4KL/N;\)

ii) \(\forall f \in \Sigma(L, \beta, M)\) and \(i = 1, \ldots, N, |b_i| \leq 4K\|f\|_\infty \leq 4KM;\)

iii) \(\forall i = 1, \ldots, N, \mathbb{E}_x(\mathbb{E}_e(\eta_i^2) \mathbb{1}_{M_n}) \leq 4\sigma^2K^2(p_0c^2)^{-1}(nh)^{-1}.\)

**Proof.** We will use the notation introduced at the beginning of Section 4.

Let us first consider i). Since \(f\) belongs to the class \(\Sigma(\beta, L, M)\) with \(\beta > 1,\) it is bounded and satisfies the Lipschitz property \(|f(X_k) - f(j/N)| \leq L|X_k - j/N|.\) Moreover \(v_{j,k} = 0\) for any \(k\) such that \(|X_k - j/N| > 1/N.\) So that \(v_{j,k}|f(X_k) - f(j/N)|\) is no larger than \(L/N.\) The same bound holds true for the terms \(v_{j,k}|f(X_k) - f(-j/N)|.\) So \(|\tilde{\beta}_j|\) is bounded with \(2L/N\) for any \(j = 1, \ldots, N\) and

\[
|r_i| \leq \frac{2L}{N^2} \sum_{j=1}^{N} \left|K_h \left(\frac{i-j}{N}\right)\right|.
\]

Since \(K_h\) has a compact support, \(K_h((i-j)/N) = 0\) for any \(j\) such that \(|j-i| > Nh,\) and there are no more than \(2Nh\) nonvanishing terms in the right-hand side of the above inequality. Furthermore each of the terms of the summation is bounded by \(\|K_h\|_\infty \leq K/h\) and finally \(|r_i|\) admits the upper bound in i).

To prove ii) we use again the fact that \(K_h\) is compactly supported and that \(f\) is bounded:

\[
|b_i| \leq \frac{1}{N} \sum_{j=1}^{N} \left|K_h \left(\frac{i-j}{N}\right)\right| \left|f \left(\frac{j}{N}\right) - f \left(\frac{-j}{N}\right)\right| \leq 2\|f\|_\infty \frac{2Nh}{N} \frac{K}{h}.
\]

To show iii) we use the expression for \(\mathbb{E}_e\eta_i^2\) given in Proposition 3. When taking the expectation with respect to \(X\) over the set \(M_n\) we obtain

\[
\mathbb{E}_x(\mathbb{E}_e(\eta_i^2) \mathbb{1}_{M_n}) = \frac{\sigma^2}{n^2N^2} \sum_{j=1}^{N} \int_{1}^{N} K_h \left(\frac{i-j}{N}\right) \mathbb{E} \left(\sum_{k=1}^{n} u_{j,k}^2 \mathbb{1}_{M_n}\right),
\]

and by iii) of Lemma 3 and since the support of \(K\) is compact

\[
\mathbb{E}_x(\mathbb{E}_e(\eta_i^2) \mathbb{1}_{M_n}) \leq \frac{\sigma^2}{n^2N^2} \sum_{j=1}^{N} \int_{1}^{N} K_h \left(\frac{i-j}{N}\right) \frac{2N}{p_0c^2} \leq \frac{\sigma^2}{n^2N^2} \frac{2NhK^2}{h^2} \frac{2N}{p_0c^2}.
\]

Recall that \((Nh^2)^{-1} = o(1).\) We can now finish the proof of the proposition. By the Markov inequality and i) of Lemma 6 we obtain for the
positive term $R_n$:

$$
P\left\{ n\sqrt{h} R_n > \frac{n^{-\delta}}{3} \right\} \leq \frac{3n\sqrt{h}}{n^{-\delta}/3} \mathbf{E} R_n = \frac{3n\sqrt{h}}{N} \sum_{i=1}^{N} \mathbf{E} r_i^2 \\
\leq \frac{3n\sqrt{h}}{n^{-\delta}/3} \sup_i \mathbf{E} r_i^2 \leq \frac{3n\sqrt{h}}{n^{-\delta}/3} \frac{16K^2L^2}{N^2}.
$$

(19)

Due to the choice of $N = An/\ln n$ and of the bandwith $h = n^{-2/(4\beta+1)}$ as soon as $\delta < 1 + (4\beta + 1)^{-1}$ the last bound tends to zero, and for any $f \in \Sigma(L, \beta)$ $P\{n\sqrt{h} R_n > n^{-\delta}/3\} = o(1)$.

Next note that $U_n$ is equal to zero under $H_0$. However, under both hypotheses for $f \in \Sigma(L, \beta, M)$:

$$
P\left\{ n\sqrt{h} |U_n| > \frac{n^{-\delta}}{3} \right\} \leq \frac{3n\sqrt{h}}{n^{-\delta}/3} \sum_{i=1}^{N} |r_i||b_i| = O \left( \frac{n\sqrt{h}}{n^{-\delta}/N} \right)
$$

when applying i) and ii) of Lemma 6. Due to the choice of $\delta, h,$ and $N$ and since $\delta < (4\beta + 1)^{-1}$ we conclude that $P\{n\sqrt{h} |U_n| > n^{-\delta}/3\} = o(1)$.

To finish the proof of the proposition let us estimate the last term in (18):

$$
P\left\{ n\sqrt{h} |V_n| > \frac{n^{-\delta}}{3} \right\} \leq \left\{ \frac{2n\sqrt{h}}{N} \sum_{i=1}^{N} |\eta_i r_i| > \frac{n^{-\delta}}{3}, M_n \right\} + P(M_n^c)
$$

\[ \leq P \left\{ \bigcup_{i=1}^{N} \left\{ |\eta_i r_i| > \frac{n^{-\delta}}{6n\sqrt{h}}, M_n \right\} \right\} + P(M_n^c) \]

\[ \leq \sum_{i=1}^{N} P \left\{ |\eta_i r_i| > \frac{n^{-\delta}}{6n\sqrt{h}}, M_n \right\} + P(M_n^c). \]

Since $|r_i| \leq 4KL/N$ and $P(M_n^c) = o(1)$ we have

$$
P\left\{ n\sqrt{h} |V_n| > \frac{n^{-\delta}}{3} \right\} \leq \sum_{i=1}^{N} P \left\{ |\eta_i| > \frac{n^{-\delta}N}{24KLn\sqrt{h}}, M_n \right\} + o(1).
$$

Finally, applying Lemma 4 and assertion iii) of Lemma 6 we have

$$
P\left\{ n\sqrt{h} |V_n| > \frac{n^{-\delta}}{3} \right\} = O \left[ \frac{N}{nh} \frac{n^2h}{n^{-2\delta}N^2} \right].
$$

So for any $1/2 > (4\beta + 1)^{-1} > \delta > 0$, $P\{n\sqrt{h} |V_n| > n^{-\delta}/3\} = o(1)$.

**5.4. Proof of Proposition 3.** Let us first prove i). Using i) of Lemma 5 we obtain

$$
E_x \eta_i^2 = \frac{1}{N^2} \sum_{j=1}^{N} \sum_{s=1}^{N} K_h \left( \frac{i-j}{N} \right) K_h \left( \frac{i-s}{N} \right) E_x (\zeta_j \zeta_s)
$$

\[ = \frac{1}{N^2} \sum_{j=1}^{N} K_h^2 \left( \frac{i-j}{N} \right) E_x \zeta_j^2 = \frac{\sigma^2}{n^2N^2} \sum_{j=1}^{N} K_h^2 \left( \frac{i-j}{N} \right) \sum_{k=1}^{n} u_{j,k}^2. \]
Let us show that $E_n S_n^2 = \Theta^2$. Recall that $S_n$ is defined as $S_n = N^{-1} \sum_{i=1}^{N} (\eta_i^2 - E_n \eta_i^2)$. Thus

$$E_n S_n^2 = \frac{1}{N^2} \sum_{(i,q) \in \{1, \ldots, N\}^2} [E_n(\eta_i^2 \eta_q^2) - E_n \eta_i^2 E_n \eta_q^2].$$

We compute first the term $E_n(\eta_i^2 \eta_q^2)$: we take the expectation (with respect to noise) in the following expression

$$\eta_i^2 \eta_q^2 = \frac{1}{N^4} \sum_{j} K_h \left( \frac{i-j}{N} \right) K_h \left( \frac{i-s}{N} \right) K_h \left( \frac{q-p}{N} \right) K_h \left( \frac{q-m}{N} \right) \zeta_j \zeta_s \zeta_p \zeta_m,$$

where the sum is taken over $(j, s, p, m) \in J$. We will assume that the set of indices $J = \{1, \ldots, N\}^4$ is partitionned as $\bigcup_{r=1}^{5} J_r$ with

$$J_1 = \{(j, s, p, m) \in J \mid j = s, p = m, j \neq p\},$$
$$J_2 = \{(j, s, p, m) \in J \mid j = m, p = s, j \neq p\},$$
$$J_3 = \{(j, s, p, m) \in J \mid j = p, s = m, j \neq s\},$$
$$J_4 = \{(j, s, p, m) \in J \mid j = s = p = m\}.$$

We denote by $\Sigma_1$ the term which is obtained when taking the expectation of the summation over the set of indices $J_1$. The summations over $J_2 \cup J_3$ and $J_4$ are denoted by $\Sigma_{2,3}$ and $\Sigma_4$, respectively. Note that the summation over $J_5$ (the complementary set of $J_1 \cup J_2 \cup J_3 \cup J_4$ in $J$) is obviously vanishing due to the properties of the reconstructed noise $\zeta_k$ (i.e., $E_n \zeta_k = 0$). Simplest algebra gives

$$\Sigma_1 = E_n \eta_i^2 E_n \eta_q^2 - \frac{1}{N^4} \left\{ \sum_{j=1}^{N} K_h \left( \frac{i-j}{N} \right) K_h \left( \frac{q-j}{N} \right) (E_n \zeta_j^2)^2 \right\}. \quad (22)$$

Second $\Sigma_{2,3}$ satisfies

$$\Sigma_{2,3} = \frac{2}{N^4} \left\{ \sum_{j=1}^{N} K_h \left( \frac{i-j}{N} \right) K_h \left( \frac{q-j}{N} \right) (E_n \zeta_j^2)^2 \right\}$$

$$- \frac{2}{N^4} \left\{ \sum_{j=1}^{N} K_h^2 \left( \frac{i-j}{N} \right) K_h^2 \left( \frac{q-j}{N} \right) (E_n \zeta_j^2)^2 \right\}. \quad (23)$$

At last,

$$\Sigma_4 = \frac{1}{N^4} \left\{ \sum_{j=1}^{N} K_h^2 \left( \frac{i-j}{N} \right) K_h^2 \left( \frac{q-j}{N} \right) E_n \zeta_j^2 \right\}. \quad (24)$$

We have from vi) in Lemma 5

$$E_n \zeta_j^4 = 3(\zeta_j^2)^2 + (\mu_4 - 3\sigma^4) n^{-4} \sum_{k=1}^{n} u_{j,k}^4.$$
Then summing up the equalities (22)-(24) and substracting the term $E_{e} \eta^{2}_{i} E_{e} \eta^{2}_{q}$ we obtain

$$E_{e}(\eta^{2}_{i} \eta^{2}_{q}) - E_{e} \eta^{2}_{i} E_{e} \eta^{2}_{q} = \frac{2}{N^4} \left\{ \sum_{j=1}^{N} K_{h}\left( \frac{i-j}{N} \right) K_{h}\left( \frac{q-j}{N} \right) E_{e} \xi_{j}^{2} \right\}^{2} + \frac{\mu_{4} - 3\sigma^{4}}{N^4} \left\{ \sum_{j=1}^{N} K_{h}^{2}\left( \frac{i-j}{N} \right) K_{h}^{2}\left( \frac{q-j}{N} \right) \sum_{k=1}^{n} u_{j,k}^{4} \right\}.$$ 

Finally, we substitute the expression given in Lemma 5 for $E_{e} \xi_{j}^{2}$ and take the sum over the indices $i$ and $q$.

5.5. Proof of Proposition 4. Recall that when denoting $g(x) = f(x) - f(-x)$ the function $g$ belongs to $\Sigma(2L, \beta, 2M)$ and $B_{n}$ can be expressed using $g$ instead of $f$, that is

$$B_{n} = \frac{1}{N} \sum_{i=1}^{N} b_{i}^{2} \quad \text{with} \quad b_{i} = \frac{1}{N} \sum_{j=1}^{N} K_{h}\left( \frac{i-j}{N} \right) g\left( \frac{j}{N} \right).$$

**Lemma 7.** For any function $f \in \Sigma(L, \beta, M)$ the following hold true:

i) $$\frac{1}{N} \sum_{i=1}^{N} g^{2}\left( \frac{i}{N} \right) = \frac{1}{2} \| f(\cdot) - f(-\cdot) \|_{2}^{2} + O\left( \frac{LM}{N} \right),$$

ii) $$\frac{1}{N} \sum_{i=1}^{N} \left[ b_{i} - g\left( \frac{i}{N} \right) \right]^{2} \leq 4L^{2}h^{2\beta}(1 + o(1)).$$

**Proof.** The first assertion is almost obvious. Indeed

$$\int_{0}^{1} g^{2} = \frac{1}{2} \| f(\cdot) - f(-\cdot) \|_{2}^{2},$$

$$\frac{1}{N} \sum_{i=1}^{N} g^{2}\left( \frac{i}{N} \right) - \int_{0}^{1} g^{2}(x) \, dx = \sum_{i=1}^{N} \int_{(i-1)/n}^{i/n} \left[ g^{2}\left( \frac{i}{N} \right) - g^{2}(x) \right] \, dx.$$ 

Since $g \in \Sigma(2L, \beta, 2M)$ it is a Lipschitz function. Thus for any $x \in [0, 1]$ and any $i = 1, \ldots, N$,

$$\left| g^{2}\left( \frac{i}{N} \right) - g^{2}(x) \right| \leq 2L \left| \frac{i}{N} - x \right| 2 \| g \|_{\infty} = 4LM \left| \frac{i}{N} - x \right|.$$ 

The integral of this value over a small interval of size $1/N$ is of order $LM/N^{2}$, what gives i).

In order to prove the second assertion of the lemma we first estimate $|b_{i} - g(i/N)|$. We have for any $i = 1, \ldots, N$

$$\left| b_{i} - g\left( \frac{i}{N} \right) \right| \leq \left| \frac{1}{N} \sum_{j=1}^{N} g\left( \frac{j}{N} \right) K_{h}\left( \frac{i-j}{N} \right) \int_{0}^{1} g(x) K_{h}\left( \frac{x}{N} - x \right) \, dx \right| + \left| \int_{0}^{1} g(x) K_{h}\left( \frac{i}{N} - x \right) \, dx - g\left( \frac{i}{N} \right) \right| \overset{\text{def}}{=} |T_{1}(i)| + |T_{2}(i)|. \quad (25)$$
For any \( i = 1, \ldots, N \), \( T_1(i) \) can be expressed as follows:

\[
T_1(i) = \frac{1}{N} \sum_{j=1}^{N} g \left( \frac{j}{N} \right) K_h \left( \frac{i-j}{N} \right) - \int_{0}^{1} g(x) K_h \left( \frac{i}{N} - x \right) \, dx
\]

\[
= \sum_{j=1}^{N} \int_{(j-1)/N}^{j/N} \left[ h_i \left( \frac{j}{N} \right) - h_i(x) \right] \, dx,
\]  

(26)

where \( h_i(x) \) denotes \( g(x) K_h(i/N - x) \).

If \( j \in \{1, \ldots, N\} \) is such that \( |j-i| > 2Nh \), then \( h_i(j/N) = 0 \). Furthermore, for any \( x \in [(j-1)/N, j/N], |i/N - x| > h \), hence, \( g(x) K_h(i/N - x) = h_i(x) = 0 \) and

\[
\int_{(j-1)/N}^{j/N} \left[ h_i \left( \frac{j}{N} \right) - h_i(x) \right] \, dx = 0.
\]

Therefore, the nonvanishing terms in the last sum in (26) belong to the set of indices \( j \) such that \( |j-i| \leq 2Nh \), and for these indices we have

\[
\left| h_i \left( \frac{j}{N} \right) - h_i(x) \right| \leq \frac{2MK}{h^2} \left| \frac{j}{N} - x \right| (1 + o(1)),
\]

since \( g \) and \( K(\cdot) \) are both Lipschitz. Thus for any \( j \) such that \( |j-i| \leq 2Nh \) we have

\[
\int_{(j-1)/N}^{j/N} \left| h_i \left( \frac{j}{N} \right) - h_i(x) \right| \, dx \leq \frac{2MK}{h^2N^2} (1 + o(1)).
\]

It follows that, due to (26), for any \( i = 1, \ldots, N \),

\[
|T_1(i)| \leq \sum_{|j-i| \leq 2Nh} \frac{2MK}{h^2N^2} (1 + o(1)) \frac{8MK}{hN} (1 + o(1)) = O \left( \frac{M}{Nh} \right).
\]  

(27)

Let us consider \( T_2(i) \). Note that due to (K2) and (K3), for all \( i, \int_{0}^{1} K_h(i/N - x) \, dx = \int_{-1}^{1} K(u) \, du = 1 \). So we have

\[
T_2(i) = \int_{0}^{1} g(x) K_h \left( \frac{i}{N} - x \right) \, dx - \int_{0}^{1} g \left( \frac{i}{N} \right) K_h \left( \frac{i}{N} - x \right) \, dx
\]

\[
= \int_{0}^{1} \left[ g(x) - g \left( \frac{i}{N} \right) \right] K_h \left( \frac{i}{N} - x \right) \, dx.
\]  

(28)

Since the function \( g \) belongs to \( \Sigma(2L, \beta, 2M) \), using the standard arguments for the calculation of the bias term in nonparametric statistics, one obtains for all \( i = 1, \ldots, N \)

\[
|T_2(i)| \leq 2Lh^\beta.
\]  

(29)

Thus, substituting (27) and (29) in inequality (25) we get

\[
|b_i - g(i/N)| \leq 2Lh^\beta (1 + o(1)).
\]

Finally, taking the sum over \( i = 1, \ldots, N \) we arrive at

\[
\frac{1}{N} \sum_{i=1}^{N} \left| b_i - g \left( \frac{i}{N} \right) \right| \leq 4L^2h^{2\beta} (1 + o(1)).
\]
Now we finish the proof of Proposition 4. Note that

\[ B_n = \frac{1}{N} \sum_{i=1}^{N} \left[ b_i - g \left( \frac{i}{N} \right) \right] + \frac{1}{N} \sum_{i=1}^{N} g^2 \left( \frac{i}{N} \right) + \frac{2}{N} \sum_{i=1}^{N} \left[ b_i - g \left( \frac{i}{N} \right) \right] g \left( \frac{i}{N} \right). \]

Applying the inequality $2xy \geq -x^2/2 - 2y^2$ to the cross term of the previous equality we obtain

\[ B_n \geq \frac{1}{2N} \sum_{i=1}^{N} g^2 \left( \frac{i}{N} \right) - \frac{1}{N} \sum_{i=1}^{N} \left[ b_i - g \left( \frac{i}{N} \right) \right]^2. \]

By Lemma 7, for any $f \in \Sigma(\beta, L, M)$

\[ B_n \geq \frac{1}{4} \| f(\cdot) - f(-\cdot) \|_2^2 - 4L^2 h^{2\beta} + O \left( \frac{1}{N} \right) + o(h^{2\beta}). \]

To conclude the proof of the proposition we put $h^{-2\beta} = n^{\sqrt{h}}$. Then for any $f \in \Omega(L, \beta, M)$ we obtain $n^{\sqrt{h}} B_n \geq F/4 - 4L^2 + o(1)$.

**5.6. Proof of Proposition 5.** Let us prove first that for any $\lambda > 0$, $P(A_n^\lambda) = o(1)$. We have

\[ C_n = \frac{2}{N} \sum_{i=1}^{N} \eta_i b_i, \quad \text{hence,} \quad \mathbb{E}_e C_n^2 = \frac{4}{N^2} \sum_{j=1}^{N} \sum_{i=1}^{N} \mathbb{E}_e (\eta_i \eta_j) b_i b_j. \]

By Lemma 5, we have

\[ \mathbb{E}_e (\eta_i \eta_j) = \frac{1}{N^2} \sum_{k=1}^{N} \sum_{s=1}^{N} K_h \left( \frac{i-k}{N} \right) K_h \left( \frac{j-s}{N} \right) \mathbb{E}_e (\zeta_k \zeta_s) \]

\[ = \frac{\sigma^2}{n^2 N^2} \sum_{k=1}^{N} K_h \left( \frac{i-k}{N} \right) K_h \left( \frac{j-k}{N} \right) \sum_{s=1}^{n} u_{k,s}^2. \]

Then it follows from Lemma 3 that \( \mathbb{E}_e (C_n^2 \mathbb{I}_{M_n}) \leq \frac{8\sigma^2}{p_0c^2 n^2 N^2} \sum_{i,j} \sum_{k=1}^{N} \left| K_h \left( \frac{i-k}{N} \right) \right| \left| K_h \left( \frac{j-k}{N} \right) \right| \left| b_i | \left| b_j \right|. \]

Since the right-hand side of the inequality above is deterministic it also gives an upper bound for $\mathbb{E}_X (\mathbb{E}_e C_n^2 \mathbb{I}_{M_n})$. Moreover, $\left| b_i \right| \left| b_j \right| \leq (b_i^2 + b_j^2)/2$, what implies

\[ \mathbb{E}_X (\mathbb{E}_e C_n^2 \mathbb{I}_{M_n}) \leq \frac{4\sigma^2}{p_0c^2 n N^2} \sum_{i,j} \sum_{k=1}^{N} \left| K_h \left( \frac{i-k}{N} \right) \right| \left| K_h \left( \frac{j-k}{N} \right) \right| \left( b_i^2 + b_j^2 \right) \]

\[ \leq \frac{8\sigma^2}{p_0c^2 n N^2} \sum_{i=1}^{N} \sum_{k=1}^{N} \left| K_h \left( \frac{i-k}{N} \right) \right| b_i^2 \sum_{j: |j-i| \leq Nh} \left| K_h \left( \frac{j-k}{N} \right) \right|.$$
\[ \leq \frac{8\sigma^2}{p_0c^2 nN^2} \sum_{i=1}^{N} \sum_{k=1}^{N} b_i^2 \sum_{k=1}^{N} K_h \left( \frac{i-k}{N} \right) \frac{2NhK}{h} \]

\[ \leq \frac{32\sigma^2K^2}{p_0c^2} \frac{1}{nN} \sum_{i=1}^{N} b_i^2 = \frac{32\sigma^2K^2 B_n}{p_0c^2 n} \]

(we use here that \( K(\cdot) \) is bounded and compactly supported). By Lemma 4 we have for any \( \lambda > 0 \)

\[ \mathbb{P}\{ |C_n| > \lambda B_n \} \leq \frac{\mathbb{E}_{X} \left( \mathbb{E}_C \mathbb{E}_{\Pi_{M_n}} \right)}{\lambda^2 B_n^2} \leq \frac{32\sigma^2K^2}{p_0c^2} \frac{1}{nB_n}. \]

Now we substitute \( (nB_n)^{-1} \) by its upper bound under the alternative \( \sqrt{h}/(F_0 + o(1)) \) (see Proposition 4), and since \( h \to 0 \) we conclude \( \sup_{f \in \Omega_1(L, \beta, M)} \mathbb{P}(\Lambda_n^c) = o(1) \).

Let us now prove that \( \mathbb{P}_{H_1}(\Gamma_n^c) = o(1) \) uniformly over the set of functions \( \Omega_1(L, \beta, M) \). The expression for \( \Theta \) is given by (4). When substituting \( \sum_k u_{j,k}^2 \) and \( \sum_k u_{j,k}^2 \) by the upper bounds given in ii) and iii) of Lemma 3 we have on \( M_n \): \( \Theta \to \theta_1^2 + \theta_2^2 \) with

\[ \theta_1^2 = \frac{8\sigma^4}{p_0^3c^3 n^2 N^4} \sum_{i=1}^{N} \sum_{q=1}^{N} \left( \sum_{j=1}^{N} K_h \left( \frac{i-j}{N} \right) K_h \left( \frac{q-j}{N} \right) \right)^2, \]

\[ \theta_2^2 = \frac{4|\mu_4 - 3\sigma^4|}{p_0^3c^3} \frac{1}{n^3 N^3} \sum_{i=1}^{N} \sum_{q=1}^{N} \sum_{j=1}^{N} K_h^2 \left( \frac{i-j}{N} \right) K_h \left( \frac{q-j}{N} \right). \]

Note that \( \theta_1^2 \) and \( \theta_2^2 \) are deterministic and when using the properties of the kernel \( K \) (bounded and compactly supported) we have the following upper bounds for \( \theta_1^2 \) and \( \theta_2^2 \):

\[ \theta_1^2 \leq \frac{8\sigma^4}{p_0^3c^3 n^2 N^4} \sum_{i=1}^{N} \sum_{q=1}^{N} \left( \sum_{j=1}^{N} K_h \left( \frac{i-j}{N} \right) K_h \left( \frac{q-j}{N} \right) \right)^2 \]

\[ \leq \frac{8\sigma^4}{p_0^3c^3 n^2 N^4} \sum_{i=1}^{N} \sum_{q=1}^{N} \left( 2Nh \frac{K^2}{h^2} \right)^2 \]

\[ \leq \frac{8\sigma^4}{p_0^3c^3 n^2 N^4} N \cdot 4Nh \cdot 4N^2 h^2 \frac{K^4}{h^4} = \frac{128\sigma^4 K^4}{p_0^3c^2 n^2 h} \]

and as \( nh \to \infty \),

\[ \theta_2^2 = \frac{4|\mu_4 - 3\sigma^4|}{p_0^3c^3} \frac{1}{n^3 N^3} \sum_{i=1}^{N} \sum_{q=1}^{N} \sum_{j=1}^{N} K_h^2 \left( \frac{i-j}{N} \right) K_h \left( \frac{q-j}{N} \right) \]

\[ \leq \frac{4|\mu_4 - 3\sigma^4|}{p_0^3c^3} \frac{1}{n^3 N^3} N \cdot 4Nh \cdot 2Nh \frac{K^4}{h^4} = \frac{4|\mu_4 - 3\sigma^4|}{p_0^3c^3} \frac{1}{n^3 h^2} \to o(1). \]

So on the set \( M_n \) for \( n \) large enough we have

\[ \Theta \leq \frac{2 \cdot 128\sigma^4}{p_0^3c^2} \frac{1}{n^2 h^2}, \]
and we conclude that
\[
P_{H_1}\left\{ \Theta \geq \frac{B_n}{\gamma} \right\} \leq P_{H_1}\left\{ \Theta^2 \geq \frac{B_n^2}{\gamma^2}, M_n \right\} + P_{H_1}(M_n^c),
\]
\[
1 \leq P_{H_1}\left\{ \frac{2 \cdot 128\sigma^4}{p_0^2c^2} \frac{1}{n^2h} \geq \frac{B_n^2}{\gamma^2} \geq \frac{F_0^2 + o(1)}{n^2\gamma^2h}, M_n \right\} + P_{H_1}(M_n^c).
\]
Now, the probability in the right-hand side of the above inequality is equal to zero as soon as
\[
\gamma \leq \gamma_0 = \frac{F_0^2 p_0^2 c^2}{(320\sigma^4)}, \sup_{f \in \Omega_1(L, s, M)} P_{H_1}(\Gamma_n^c) = o(1).
\]

Acknowledgments. Authors thank W. Härdle, C. Hafner, V. Spokoiny, and A. Juditsky for the suggestions and helpful remarks.

REFERENCES
3. Ермаков М. С. Минимаксное обнаружение сигнала в гауссовском белом шуме. — Теория вероятн. и ее примен., 1990, т. 35, в. 4, с. 704-715.

Поступила в редакцию 02. VII. 1999
LIMIT THEOREMS FOR CERTAIN FUNCTIONALS OF UNIONS OF RANDOM CLOSED SETS

Пусть $X_1, X_2, \ldots$ — последовательность случайных независимых, одинаково распределенных, замкнутых подмножеств некоторого локально компактного хаусдорфова сепарабельного пространства $E$. Для каждого случайного замкнутого множества $Y$ мы рассматриваем функционал $Q_Y(F)$, равный вероятности того, что $Y$ не пересекается с замкнутым подмножеством $F \subseteq E$. Цель статьи — установить предельные теоремы для случайных величин $Q_Y(X_1 \cup \cdots \cup X_n)$. Полученные результаты применяются для асимптотического анализа средней ширины выпуклых оболочек, порожденных равномерными выборками на многомерном шаре.

Ключевые слова и фразы: случайные множества, объединения случайных множеств, функционалы достижения, экстремальные значения, выпуклые оболочки, средняя ширина, периметр.

1. Introduction and main results. Consider a locally compact, Hausdorff and separable topological space $E$ and denote by $\mathcal{F} := \mathcal{F}(E)$ the family of all its closed subsets. We endow $\mathcal{F}$ with the vague (Fell) topology $\mathcal{T}_f$ generated by subfamilies of the form

$$\mathcal{F}^K_{G_1, \ldots, G_n} := \{F \in \mathcal{F} \mid F \cap K = \emptyset, F \cap G_1 \neq \emptyset, \ldots, F \cap G_n \neq \emptyset\}, \quad n \geq 0,$$

where $K$ runs through the class of compacts and $G_1, \ldots, G_n$ are open (see Chapter 1 in [9] or Section 1.1 in [10]). It is known that $(\mathcal{F}, \mathcal{T}_f)$ is Hausdorff, compact and separable (see ibidem).

The random elements taking values in $(\mathcal{F}, \mathcal{T}_f)$, where $\mathcal{T}_f$ is the Borel $\sigma$-field corresponding to $\mathcal{T}_f$, will be referred to as random closed sets (for extensive references see [6], [9] or [10]). Without a further mention we will assume all the considered random elements to be defined on a common probability space $(\Omega, \mathcal{F}, P)$.

For a random closed set $Y$ its avoidance functional $Q_Y: \mathcal{F} \to [0, 1]$ is defined by $Q_Y(F) := P\{F \cap Y = \emptyset\}$. A number of important functionals on $\mathcal{F}$ can be expressed in terms of $Q_Y$ for properly chosen $Y$. For example taking an arbitrary $E$-valued random element $V$ and setting $Y$ to be the

*Faculty of Mathematics and Computer Science, Nicholas Copernicus University, Toruń, Poland; e-mail: tomeks@mat.uni.torun.pl