ESAIM: Probability and Statistics May 2003, Vol. 7, 279–312
DOI: 10.1051/ps:2003013

ON ASYMPTOTIC MINIMAXITY OF KERNEL-BASED TESTS [*]

Michael Ermakov [1]

Abstract. In the problem of signal detection in Gaussian white noise we show asymptotic minimaxity of kernel-based tests. The test statistics equal L_2-norms of kernel estimates. The sets of alternatives are essentially nonparametric and are defined as the sets of all signals such that the L_2-norms of signal smoothed by the kernels exceed some constants $\rho_\epsilon > 0$. The constant ρ_ϵ depends on the power ϵ of noise and $\rho_\epsilon \to 0$ as $\epsilon \to 0$. Similar statements are proved also if an additional information on a signal smoothness is given. By theorems on asymptotic equivalence of statistical experiments these results are extended to the problems of testing nonparametric hypotheses on density and regression. The exact asymptotically minimax lower bounds of type II error probabilities are pointed out for all these settings. Similar results are also obtained for the problems of testing parametric hypotheses *versus* nonparametric sets of alternatives.

Mathematics Subject Classification. 62G10, 62G20.

Received March 21, 2002. Revised February 15, 2003.

1. Introduction

Suppose we observe a random process $Y_\epsilon(t), t \in [0,1]$, defined by a stochastic differential equation

$$\mathrm{d}Y_\epsilon(t) = S(t)\mathrm{d}t + \epsilon q(t)\mathrm{d}w(t), \quad \epsilon > 0 \tag{1.1}$$

where $\mathrm{d}w(t)$ is the standard Gaussian white noise and $q(t), t \in [0,1]$ is a weight function. The function S, called a signal, is unknown. The problem is to test a hypothesis that the signal $S(t)$ is absent, that is, $S(t) = 0$ for all $t \in [0,1]$.

We could not test this hypothesis without *a priori* information of parametric or nonparametric type (see Burnashev [5], Ermakov [12]). For nonparametric sets of alternatives *a priori* information is often given in terms of assumptions on a signal smoothness (see Ingster and Suslina [22], Ermakov [6], Spokoiny [31]). Such a setting can be considered as an analog of standard setting nonparametric estimation and obtained practically an adequate development. The optimal rates of distinguishability of hypotheses were pointed out for nonparametric sets of alternatives that can belong to a wide range of functional spaces (see Ingster and Suslina [22], Lepski and Spokoiny [25], Guerre and Lavergne [14]). The asymptotically minimax tests have been found for the nonparametric sets of alternatives in L_2 (see Ermakov [6]) and l_p (see Ingster [21]) spaces.

Keywords and phrases. Nonparametric hypothesis testing, kernel-based tests, goodness-of-fit tests, efficiency, asymptotic minimaxity, kernel estimator.

[*] *The support by RFFI-NNIO Grant 02-01-04001 and RFFI Grand 02-01-00262 is acknowledged.*

[1] Russian Academy of Sciences, Mechanical Engineering Problem Institute, Bolshoy Pr. V.O. 61, 199178 St. Petersburg, Russia; e-mail: `ermakov@random.ipme.ru`.

In nonparametric hypothesis testing the test statistics are often defined as the distances between the hypotheses and estimator of nonparametric parameter. We have no usually any information on a signal smoothness, such an information is not necessary in the problem of distinguishability of hypothesis and nonparametric sets of alternatives (see Ermakov [12]) and it seems desirable to represent the sets of alternatives in a more evident form depending also on distances between the hypotheses and alternatives, covering all possible alternatives. Thus it seems natural to consider the testing nonparametric hypotheses from the distance positions and to develop rigorous justification of this approach. From viewpoint of asymptotic minimaxity such an argumentation has been developed in Ermakov [10,11] in the case of standard goodness-of-fit tests. These results are based on the interpretation of test statistics of Kolmogorov, omega-square and chi-squared tests as the corresponding norms or seminorms (in the case of chi-squared tests) $N_n(\hat{F}_n - F_0)$ depending on a difference of empirical distribution function $\hat{F}_n$ of independent sample $X_1, \ldots, X_n$ and the distribution function F_0 of hypothesis. The corresponding norms or seminorms N_n are defined in the linear space generated by the differences of distribution functions. The sets of alternatives are the sets of all distribution functions F such that $N_n(F - F_0) > \rho_n > 0$ with $\rho_n \to 0$ as $n \to \infty$. In this setting asymptotic minimaxity of tests statistics $N_n(\hat{F}_n - F_0)$ has been proved and asymptotic behaviour of type II error probabilities has been studied. In the case of chi-squared tests we supposed that the number of cells grows with increasing sample size. Note that this approach can be naturally considered as a part of asymptotic theory of hypothesis testing on a value of functional (see Stein [32], Ermakov [8], Bickel *et al.* [2]).

In paper similar statements will be obtained for the test statistics based on the kernel estimator (see Bickel and Rosenblatt [1], Fan [13], Hart [19], Rayner and Best [29], Stute [33], Horowitz and Spokoiny [20] and references therein)

$$T(h, Y_\epsilon) = \int_0^1 \hat{S}_h^2(t) r(t)\mathrm{d}t = \int_0^1 \left(\frac{1}{h} \int_0^1 K\left(\frac{t-s}{h}\right) \mathrm{d}Y_\epsilon(s) \right)^2 r(t)\mathrm{d}t \tag{1.2}$$

where

$$\hat{S}_h(t) = \frac{1}{h} \int_0^1 K\left(\frac{t-s}{h}\right) \mathrm{d}Y_\epsilon(s)$$

is a kernel estimator of signal with a kernel K and $r(t), t \in [0,1]$ is a weight function. We suppose that the support of K is contained in $[-1,1]$, $K(t) = K(-t)$ for all $t \in (0,1)$, $\int_{-1}^1 K(t)\mathrm{d}t = 1$ and the function K is bounded. The functions $r(t), q(t)$ are supposed positive and continuous in $[0,1]$, $0 < c < r(t) < C < \infty, 0 < c < q(t) < C < \infty$ for all $t \in [0,1]$.

The sets $\Im_{\epsilon,h}$ of alternatives are as follows

$$\begin{aligned} \Im_{\epsilon,h} &= \Im_{\epsilon,h}(\rho_\epsilon) = \{S : T(h,S) > \rho_\epsilon(h) > 0, S \in L_2(0,1)\} \\ &= \left\{ S : \int_0^1 \left(\frac{1}{h} \int_0^1 K\left(\frac{t-s}{h}\right) S(s)\mathrm{d}s \right)^2 r(t)\mathrm{d}t > \rho_\epsilon(h) > 0,\ S \in L_2(0,1) \right\} \cdot \end{aligned} \tag{1.3}$$

The rates of convergence $\rho_\epsilon = \rho_\epsilon(h_\epsilon) \to 0$ and $h = h_\epsilon \to 0$ as $\epsilon \to 0$ will be defined later.

We also consider the sets of alternatives $\Im_{\epsilon,h_\epsilon}$ defined as the intersections of sets $\Im_{\epsilon,h_\epsilon}(\rho_\epsilon)$ with the balls in Sobolev space.

It is easy to see that, in the case of alternative S,

$$E_S(T(h, Y_\epsilon)) - \frac{\epsilon^2}{h^2} \int_0^1 r(t)\mathrm{d}t \int_0^1 K^2\left(\frac{t-s}{h}\right) q^2(s)\mathrm{d}s = \int_0^1 r(t)\mathrm{d}t \left(\frac{1}{h} \int_0^1 K\left(\frac{t-s}{h}\right) S(s)\mathrm{d}s \right)^2.$$

Thus the sets of alternatives are defined by the components of biases of test statistics $T(h, Y_\epsilon)$ caused by the presence of signal.

For any test L denote $\alpha(L) = E_0(L)$ its type I error probability and $\beta(L,S) = E_S(1-L)$ its type II error probability for the alternative $S \in \Im_{\epsilon,h}$. For any set of alternatives $\Im_\epsilon$ we put

$$\beta_\epsilon(L) = \beta_\epsilon(L,\Im_\epsilon) = \sup\{\beta(L,S) : S \in \Im_\epsilon\}\cdot \tag{1.4}$$

We say a family of tests U_ϵ with $\alpha(U_\epsilon) \le \alpha$, $0 < \alpha < 1$, $\epsilon > 0$ is asymptotically minimax for the sets of alternatives $\Im_\epsilon$, if for any family of tests $W_\epsilon, \alpha(W_\epsilon) \le \alpha$, it holds

$$\liminf_{\epsilon \to 0}(\beta_\epsilon(W_\epsilon,\Im_\epsilon) - \beta_\epsilon(U_\epsilon,\Im_\epsilon)) \ge 0.$$

The test statistics $R_\epsilon, \epsilon > 0$ generating the asymptotically minimax families of tests will be called asymptotically minimax as well.

In paper we prove asymptotic minimaxity of kernel-based test statistics $T_\epsilon(Y_\epsilon) = T(h_\epsilon, Y_\epsilon)$ for the sets of alternatives $\Im_{\epsilon,h_\epsilon}(\rho_\epsilon)$ and the intersections of sets $\Im_{\epsilon,h_\epsilon}(\rho_\epsilon)$ with the balls in Sobolev spaces. After that, using the theory of asymptotic equivalence of nonparametric statistical experiments (see Brown and Low [4], Nussbaum [27]) this statement is extended on the problems of hypothesis testing on regression and density. We show that similar results can be also obtained if the hypotheses are parametric. Such statements are proved for the problems of signal detection and hypothesis testing about density. We do not consider the same setting for parametric regression in order to do not increase extremely the scope of paper. The sets alternatives $\Im_{\epsilon,h_\epsilon}(\rho_\epsilon)$ are the largest sets $\Im_\epsilon$ such that the hypothesis and alternatives $S_\epsilon \in \Im_\epsilon$ are distinguishable if we apply the test statistics $T(h_\epsilon, Y_\epsilon)$. Thus we prove asymptotic minimaxity of test statistics $T(h_\epsilon, Y_\epsilon)$ for the largest among possible sets of alternatives. Moreover, it turns out, the lower bound of type II error probabilities is attained for all families of alternatives $S_\epsilon, \epsilon > 0$ such that $T(h_\epsilon, S_\epsilon) = \rho_\epsilon(1 + o(1))$.

The asymptotic behaviour of kernel-based test statistics has been intensively studied in many papers (see Konakov [24], Hall [16,17], Rayner and Best [29], Ghosh and Wei-Min Huang [15], Fan [13], Hart [19], Stute [33], Horowitz and Spokoiny [20] and references therein). Thus the results on asymptotic minimaxity represent the essential complement to the existing theory. We find the distance of hypothesis from the signal given in the Gaussian noise and can analyse the type II error probabilities for all possible sets of alternatives defined in terms of the same distance. The more detailed discussion of the role of asymptotic minimaxity in the distance approach for testing nonparametric hypotheses one can find in Ermakov [11].

The reasonings in the paper are based on the same approach as in Ermakov [7]. In Ermakov [7] the sets of alternatives were defined as the intersections of exteriors of balls and ellipsoid in L_2. The requirement that the signal belongs to ellipsoid was caused the smoothness assumptions. The problem was reduced to minimization of variance of test statistics. The minimum of variance was attained on the intersection of boundaries of balls and ellipsoid. Thus we got the minimization problem with two restrictions of quadratic type. In the present paper the sets of alternatives $\Im_{\epsilon,h_\epsilon}(\rho_\epsilon)$ are interpreted as the exteriors of ellipsoids in $L_2(0,1)$. As a consequence one needs to solve the problem of variance minimization with only one restriction of quadratic type. At the same time in the present paper the operator that set the restriction is not diagonal. This cause the main differences in paper reasoning.

Note that, in the problems of signal detection with a given signal smoothness, asymptotically minimax test statistics or statistics having optimal rates of distinguishability (see Ermakov [7], Ingster and Suslina [22]) are often defined as seminorms $N_\epsilon(Y_\epsilon)$ of quadratic type. A simple analysis of the proofs in Ermakov [7] and Ingster and Suslina [22] shows that these test statistics $N_\epsilon(Y_\epsilon)$ are asymptotically minimax or have optimal rates of distinguishability for the more wider sets of alternatives $\{S : N_\epsilon(S) > \rho_\epsilon > 0, S \in L_2(0,1)\}$ then in the setting with a signal smoothness. Thus the results on signal detection with a given signal smoothness can be also interpreted in terms of distance approach.

The asymptotic minimaxity of tests statistics $T(h_\epsilon, Y_\epsilon)$ is proved for the sets of alternatives $\Im_{\epsilon,h_\epsilon}(\rho_\epsilon)$ having two variable parameters h_ϵ and ρ_ϵ. If $\rho_\epsilon = \rho_\epsilon(h_\epsilon)$ satisfies more strong restrictions as a function of h_ϵ we prove asymptotic minimaxity of $T(h_\epsilon, Y_\epsilon)$ for the more narrow sets. These sets of alternatives are defined as intersections of $\Im_{\epsilon,h_\epsilon}(\rho_\epsilon)$ with the balls in Sobolev space.

Remark 1.1. If we test the hypothesis $S = S_0$, the test statistic $T(h, Y_\epsilon - S_0)$ has the following modified form

$$T(h, Y_\epsilon - S_0) = \int_0^1 \left(\int_0^1 K_h(t-s)\mathrm{d}Y_\epsilon(s) - \int_0^1 K_h(t-s)S_0(s)\mathrm{d}s \right)^2 r(t)\mathrm{d}t \tag{1.5}$$

where $K_h(t-s) = \frac{1}{h}K\left(\frac{t-s}{h}\right)$.

The sets of alternatives are as follows

$$\Im_{\epsilon,h}(\rho_\epsilon, S_0) = \left\{ S : \int_0^1 \left(\int_0^1 K_h(t-s)S(s)\mathrm{d}s - \int_0^1 K_h(t-s)S_0(s)\mathrm{d}s \right)^2 r(t)\mathrm{d}t > \rho_\epsilon(h) > 0, \quad S \in L_2(0,1) \right\}. \tag{1.6}$$

In this setting the kernel-based tests have often another form

$$\bar{T}(h, Y_\epsilon, S_0) = \int_0^1 \hat{S}_h^2(t) r(t)\mathrm{d}t = \int_0^1 \left(\int_0^1 K_h(t-s)\mathrm{d}Y_\epsilon(s) - S_0(t) \right)^2 r(t)\mathrm{d}t \tag{1.7}$$

and it seems natural, for such tests, to define the sets of alternatives in another form

$$\bar{\Im}_{\epsilon,h}(\rho_\epsilon, S_0) = \left\{ S : \int_0^1 \left(\int_0^1 K_h(t-s)S(s)\mathrm{d}s - S_0(t) \right)^2 r(t)\mathrm{d}t > \rho_\epsilon(h) > 0, \ S \in L_2(0,1) \right\} \tag{1.8}$$

as well.

The test statistic $\bar{T}(h, Y_\epsilon, S_0)$ contains additional bias term

$$E(\bar{T}(h, Y_\epsilon, S_0) - T(h, Y_\epsilon - S_0)) = \int_0^1 \left(\int_0^1 K_h(t-s)S_0(s)\mathrm{d}s - S_0(t) \right)^2 r(t)\mathrm{d}t. \tag{1.9}$$

Note that similar bias term

$$\int_0^1 \left(\int_0^1 K_h(t-s)S(s)\mathrm{d}s - S(t) \right)^2 r(t)\mathrm{d}t$$

caused the alternative is absent in test statistics $\bar{T}(h, Y_\epsilon, S_0)$ and $T(h, Y_\epsilon - S_0)$. Thus, using test statistics $T(h, Y_\epsilon - S_0)$, we simply delete the fast oscilating component both in hypothesis and alternatives. This is a standard procedure. If we test the hypothesis *versus* sets of alternatives defined in terms of series of ortogonal functions (see Ingster and Suslina [22], Lepskii and Spokoiny [25], Ermakov [7]), the tests statistics are also based on the first Fourier coefficients and estimates of these coefficients. The Fourier coefficients of higher orders are ignored both for the hypothesis and alternatives. Thus, using the test statistics $T(h, Y_\epsilon - S_0)$ instead of $\bar{T}(h, Y_\epsilon, S_0)$, we follow the same reasons. If we could not make any serious conclusions about very fast oscilating part of signal, we simply do not include this part in test statistics. The definition of sets of alternatives $\Im_{\epsilon,h}(\rho_\epsilon, S_0)$ follows the same reasons as well. Note that the bias term (1.9) have often the order $o(\rho_\epsilon(h_\epsilon))$ (see Rem. 2.2.3) and is unessential in the problems of hypothesis testing. In this case both test statistics $T(h, Y_\epsilon - S_0)$ and $\bar{T}(h, Y_\epsilon, S_0)$ are asymptotically minimax for both sets of alternatives $\Im_{\epsilon,h}(\rho_\epsilon, S_0)$ and $\bar{\Im}_{\epsilon,h}(\rho_\epsilon, S_0)$.

Remark 1.2. The asymptotic minimaxity is proved for a wide classes of sets of alternatives defined by the structure of kernel-based tests. All these sets of alternatives have the same optimal rates of distinguishability if *a priori* information is given, that signal belongs to a ball $W^{(\beta)}(P_0)$ in Sobolev space and $h_\epsilon \asymp \epsilon^{\frac{4}{4\beta+1}}$. Moreover we show that $T(h_\epsilon, S - S_0) \asymp ||S - S_0||$ if $S - S_0 \in W^{(\beta)}(P_0)$ and $h_\epsilon \asymp \epsilon^{\frac{4}{4\beta+1}}$ (see (2.8, 2.9)). Thus such a wide class of sets of alternatives arises as the consequence of requirement: for given procedure to enclose all distinguishable alternatives. Note that seminorm $T(h_\epsilon, S - S_0)$ has a rather evident interpretation. We compare the L_2-norms for differences of smoothed signals of hypothesis and alternatives obliterating the oscillations greater then h_ϵ.

We use letter C as a generic notation for positive constants. We put $K_h(t) = \frac{1}{h}K\left(\frac{t}{h}\right)$,

$$K_{2,h}(t) = \frac{1}{h}K_2\left(\frac{t}{h}\right) = \frac{1}{h^2}\int_{-1}^{1} K\left(\frac{t-s}{h}\right)K\left(\frac{s}{h}\right)\mathrm{d}s,$$

$$K_{i,h}(t) = \frac{1}{h}K_i\left(\frac{t}{h}\right) = \frac{1}{h^2}\int_{-1}^{1} K_{i-1}\left(\frac{t-s}{h}\right)K\left(\frac{s}{h}\right)\mathrm{d}s$$

for $i = 3, 4$. If $h = 1$, the index h will be omitted, that is, $K_{2,1} = K_2, K_{4,1} = K_4$ and so on. Denote $\chi(A)$ the indicator of an event A, $[x]$ the whole part of $x \in R^1$ and $||\cdot|| - L_2$-norm in [0,1].

In paper the three settings are considered: the signal detection in Gaussian white noise, the hypothesis testing on regression and density. It will be convenient to make use of similar or the same notation in the statements and in the proofs of the related results.

2. Main results

The results on signal detection, testing hypotheses on nonparametric regression and density will be given in three subsections.

2.1. Nonparametric signal detection

Define $x_\alpha, 0 < \alpha < 1$, by the equation

$$\alpha = 1 - \Phi(x_\alpha) = \frac{1}{\sqrt{2\pi}}\int_{x_\alpha}^{\infty} \exp\left\{-\frac{x^2}{2}\right\}\mathrm{d}x.$$

Denote

$$d_\epsilon(h) = \frac{\epsilon^2}{h}\int_0^1 r(t)\mathrm{d}t\int_{-1}^{1} K^2(u)q^2(t-uh)\mathrm{d}u,$$
$$\sigma^2 = 2\int_{-2}^{2} K_2^2(v)\mathrm{d}v\int_0^1 q^4(t)r^2(t)\mathrm{d}t.$$

Hereafter we suppose that $q(t) = 0$ if $t \notin [0,1]$.

Note that, if $q(t)$ satisfies Hoelder condition: $|q(t) - q(s)| < C|t-s|^\kappa, \kappa > 1/2$ for all $t, s \in [0,1]$, one can make use of the more simple formula

$$d_\epsilon(h) = \frac{\epsilon^2}{h}\int_0^1 q^2(t)r(t)\mathrm{d}t\int_{-1}^{1} K^2(u)\mathrm{d}u\left(1 + o\left(h^{1/2}\right)\right).$$

Theorem 2.1.1. *Let $\epsilon^2 h_\epsilon^{-1/2} \to 0$, $h_\epsilon \to 0$ as $\epsilon \to 0$ and*

$$0 < \liminf_{\epsilon\to 0} \epsilon^{-2}\rho_\epsilon(h_\epsilon)h_\epsilon^{1/2} \le \limsup_{\epsilon\to 0} \epsilon^{-2}\rho_\epsilon(h_\epsilon)h_\epsilon^{1/2} < \infty. \tag{2.1}$$

Then the family of kernel-based tests

$$L_\epsilon = \chi\left\{\epsilon^{-2}h_\epsilon^{1/2}\sigma^{-1}(T_\epsilon(Y_\epsilon) - d_\epsilon(h_\epsilon)) > x_\alpha\right\}$$

is asymptotically minimax for the sets of alternatives $\Im_{\epsilon,h_\epsilon}(\rho_\epsilon)$.

It holds

$$\beta_{\epsilon,h_\epsilon}(L_\epsilon) = \Phi\left(x_\alpha - \frac{h_\epsilon^{1/2}\rho_\epsilon(h_\epsilon)}{\epsilon^2\sigma}\right)(1+o(1)) \tag{2.2}$$

as $\epsilon \to 0$.

Moreover for each $S_\epsilon \in L_2(0,1), \epsilon > 0$ *such that* $T(h_\epsilon, S_\epsilon) = \rho_\epsilon(h_\epsilon)(1+o(1))$ *it holds*

$$\beta_{\epsilon,h_\epsilon}(L_\epsilon, S_\epsilon) = \Phi\left(x_\alpha - \frac{h_\epsilon^{1/2}\rho_\epsilon(h_\epsilon)}{\epsilon^2\sigma}\right)(1+o(1)) \tag{2.3}$$

as $\epsilon \to 0$.

Remark 2.1.1. In the kernel estimation, to preserve the optimal rate of convergence (see Hardle [18]), a modification of kernel estimator is often introduced near the boundary of interval [0,1]. The same problem can arise in testing nonparametric hypotheses if *a priori* information on a signal smoothness is given. If we are not interesting very seriously the signal behaviour near the boundary, one can use the test statistics

$$\tilde{T}(h, Y_\epsilon) = \int_h^{1-h}\left(\int_0^1 K_h(t-s)\mathrm{d}Y_\epsilon\right)^2 r(t)\mathrm{d}t$$

with the sets of alternatives

$$\tilde{\Im}_{\epsilon,h} = \{S : \tilde{T}(h,S) > \rho_\epsilon(h) > 0\}\cdot$$

For the test statistics $\tilde{T}(h, Y_\epsilon)$ the similar statements of Theorem 2.1 holds. One needs only to replace the sets of alternatives $\Im_{\epsilon,h_\epsilon}(\rho_\epsilon)$ by the sets $\tilde{\Im}_{\epsilon,h_\epsilon}(\rho_\epsilon) = \{S : \tilde{T}(h,S) > \rho_\epsilon(h) > 0\}$ and $d_\epsilon(h_\epsilon)$ by

$$\tilde{d}_\epsilon(h_\epsilon) = \frac{\epsilon^2}{h_\epsilon}\int_{h_\epsilon}^{1-h_\epsilon} r(t)\mathrm{d}t \int_{-1}^1 K^2(u)q^2(t-uh_\epsilon)\mathrm{d}u.$$

Similar modification of statements holds for the settings Theorems 2.2 and 2.3 as well.

Remark 2.1.2. As follows from (2.2) and (2.3) the lower bounds of type II error probabilities are attained for all families of alternatives $S_\epsilon, \epsilon > 0$ such that

$$0 < \liminf_{\epsilon\to 0} \epsilon^{-2}h_\epsilon^{1/2}T(h_\epsilon, S_\epsilon) \le \limsup_{\epsilon\to 0}\epsilon^{-2}h_\epsilon^{1/2}T(h_\epsilon,S_\epsilon) < \infty.$$

Thus the test statistics give "optimal distinguishability for all alternatives having a given distance from the hypothesis in the sense of $T^{1/2}(h_\epsilon, S)$-seminorm". Note that the same situation takes place in the case of chi-squared tests as well (see Ermakov [11]).

A similar statement is valid if *a priori* information on a signal smoothness is given that the signal S belongs to a ball in Sobolev space

$$S \in W_2^{(\beta)}(P_0) = \left\{S : \int_0^1 (S^2(s) + (S^{(\beta)}(s))^2)\mathrm{d}s < P_0\right\}\cdot$$

Hereafter $S^{(\beta)}$ denotes β-derivative of S.

The sets of alternatives equal $\Im_{\epsilon,h_\epsilon}(\rho_\epsilon, \beta, P_0) = \Im_{\epsilon,h_\epsilon}(\rho_\epsilon) \cap W_2^{(\beta)}(P_0)$.

Make the following additional assumption:

A. There exists the bounded β-derivative $K^{(\beta)}$ of kernel K, that is, $\sup_{s\in(-1,1)}|K^{(\beta)}(s)| < C < \infty$ and $K^{(\beta)}(-1) = K^{(\beta)}(1) = 0$, $K^{(i)}(-1) = K^{(i)}(1) = 0$ for all $i, 0 \le i \le \beta$. The function $r(t)$ has bounded β-derivatives on $(0,1)$.

Denote

$$C_\beta(K) = \int_{-1}^{1} (K^{(\beta)}(s))^2 \mathrm{d}s \int_0^1 q^2(t) r(t) \mathrm{d}t.$$

Theorem 2.1.2. *Let the assumptions of Theorem 2.1 be satisfied, let A hold and let*

$$\limsup_{\epsilon\to 0} \rho_\epsilon h_\epsilon^{-2\beta} C_\beta(K) < \frac{1}{2} P_0. \tag{2.4}$$

Then the family of kernel-based tests $L_\epsilon, \epsilon > 0$ is asymptotically minimax for the sets of alternatives $\Im_{\epsilon,h_\epsilon}(\rho_\epsilon, \beta, P_0)$ and

$$\beta_{\epsilon,h_\epsilon}(L_\epsilon) = \beta_{\epsilon,h_\epsilon}(L_\epsilon, \Im_{\epsilon,h_\epsilon}(\rho_\epsilon,\beta,P_0)) = \Phi\left(x_\alpha - \frac{h_\epsilon^{1/2}\rho_\epsilon(h_\epsilon)}{\epsilon^2\sigma}\right)(1+o(1)). \tag{2.5}$$

As follows from (2.3) the lower bound in (2.5) is attained for each family of signals $S_\epsilon \in W_2^{(\beta)}(P_0)$ such that $T(h_\epsilon, S_\epsilon) = \rho_\epsilon(h_\epsilon)(1+o(1))$.

By (2.1, 2.4) we get the following bounds for the rate of convergence h_ϵ and ρ_ϵ to zero

$$\limsup_{\epsilon\to 0} \epsilon^2 h_\epsilon^{-2\beta-1/2} < \infty \tag{2.6}$$

and

$$\limsup_{\epsilon\to 0} \epsilon^{-\frac{8\beta}{4\beta+1}} \rho_\epsilon > 0. \tag{2.7}$$

The proof of Theorem 2.1.2 is similar to that of Theorem 2.1.1. It suffices to test only that the realizations of random process generated by the Bayes *a priori* measures belongs to the ball $W_2^{(\beta)}(P_0)$ in Sobolev space. A similar statements can be obtained also for the balls in other functional spaces, using the same arguments and the fact that, by (2.3), the corresponding lower bound is attained.

We say that the sets of alternatives $\Im_\epsilon$ are distinguishable if, for each $0 < \alpha < 1$, there exists a family of tests $U_\epsilon, \alpha(U_\epsilon) = \alpha$ such that

$$\liminf_{\epsilon\to 0} \beta_{\epsilon,h_\epsilon}(U_\epsilon, \Im_\epsilon) < 1 - \alpha.$$

It follows from (2.1, 2.6, 2.7) that the optimal rate of distinguishability for the sets of alternatives $\Im_{\epsilon,h_\epsilon}(\rho_\epsilon,\beta,P_0)$ equals $\epsilon^{\frac{8\beta}{4\beta+1}}$. This rate is attained if $h_\epsilon \asymp \epsilon^{\frac{4}{4\beta+1}}$.

Define the sets

$$Q_1(\rho_\epsilon, \beta, P_0) = \left\{ S : S \in W_2^{(\beta)}(P_0), ||S||^2 > \rho_\epsilon \right\}.$$

Denote $Q(\rho_\epsilon,\beta,P_0)$ the set of all $S \in Q_1(\rho_\epsilon,\beta,P_0)$ such that there exist β-derivatives $S^{(\beta)}(0) = 0$ and $S^{(\beta)}(1) = 0$ of S and $S(0) = S(1) = S^{(i)}(0) = S^{(i)}(1) = 0$ for all $0 < i \le [\beta]$.

Denote

$$\hat{K}(\omega) = \int_{-\infty}^{\infty} K(t)\exp\{i\omega t\}\mathrm{d}t, \quad \hat{S}(\omega) = \int_{-\infty}^{\infty} S(t)\exp\{i\omega t\}\mathrm{d}t$$

the Fourier transforms of $K(t)$ and $S(t)$.

Suppose $\hat{K}(\omega)|\omega|^{-\beta} \to 0$ as $\omega \to \infty$. Suppose also that $r(t) = 1$ for all $t \in (0,1)$.

Denote $\omega_{0\epsilon}$ such that $P_0\omega_{0\epsilon}^{-2\beta} = \rho_\epsilon h_\epsilon^{-2\beta}$ and denote $\omega_0 = \inf\{\omega : \hat{K}(\omega) = 0, \omega > 0\}$.

We show that,

$$\Im_{\epsilon,h_\epsilon}(\rho_\epsilon,\beta,P_0) \subset Q_1(\rho_\epsilon,\beta,P_0) \tag{2.8}$$

and, if $\omega_{0\epsilon} < \omega_0 - \delta$ with $\delta > 0$,

$$Q(\rho_\epsilon,\beta,P_0) \subset \Im_{\epsilon,h_\epsilon}(C_\epsilon\rho_\epsilon,\beta,P_0) \tag{2.9}$$

with $C_\epsilon = |\hat{K}(\omega_{1\epsilon})|^{-2}(1+o(1))$ where $\omega_{1\epsilon} = \arg\inf_\omega\{\hat{K}(\omega) = \inf\{\hat{K}(u) : |u| < \omega_{0\epsilon}\}\}$.

The optimal order of distinguishability for the sets of alternatives $Q(\rho_\epsilon, \beta, P_0)$ equals $\epsilon^{\frac{8\beta}{4\beta+1}}$ (see Ingster and Suslina [22], Ermakov [7]). Thus if $h_\epsilon \asymp \epsilon^{\frac{4}{4\beta+1}}$ and $\omega_{0\epsilon} < \omega_0 - \delta$ with $\delta > 0$, then the orders of distinguishability coinside for the sets of alternatives $\Im_{\epsilon,h_\epsilon}(\rho_{1\epsilon}, \beta, P_0)$ and $Q(\rho_\epsilon, \beta, P_0)$.

In hypothesis testing with *a priori* information on a signal smoothness the optimal rates of distinguishability is often proved for the test statistics admitting the interpretation as seminorms in functional spaces (see Ermakov [7], Ingster and Suslina [22]). Theorems 2.1.1, 2.1.2 and (2.8) show that, in this case, one can expect asymptotic minimaxity of these tests statistics for essentially more wider sets of alternatives $\Im_\epsilon(\rho_\epsilon)$ generated by these seminorms. For such sets of alternatives we do not need to make any assumptions of smoothness type. Moreover the statements of type (2.9) hold.

By Young inequality, we get $T_\epsilon(S, h_\epsilon) < ||S||^2$. This implies (2.8).

By Parseval identity, we get

$$T_\epsilon(S, h_\epsilon) = \int_{-\infty}^{\infty} |\hat{K}(h_\epsilon\omega)\hat{S}(\omega)|^2 \mathrm{d}\omega,$$

$$\int_{-\infty}^{\infty} (S^{(\beta)}(t))^2 \mathrm{d}t = \int_{-\infty}^{\infty} |\omega|^{2\beta} |\hat{S}(\omega)|^2 \mathrm{d}\omega$$

and $||S|| = ||\hat{S}||$.

Hence, we get

$$\inf\{T_\epsilon(S, h_\epsilon) : ||S||^2 > \rho_\epsilon, S \in W_2^{(\beta)}(P_0), \operatorname{supp} S \subset (0,1)\} \ge \inf\left\{\int_{-\infty}^{\infty} |\hat{K}\,(h_\epsilon\omega)\,\hat{S}(\omega)|^2 \mathrm{d}\omega :\, ||\hat{S}||^2 > \rho_\epsilon,\right.$$

$$\left.\int_{-\infty}^{\infty} |\omega|^{2\beta} |\hat{S}(\omega)|^2 \mathrm{d}\omega < P_0\right\} + o\left(h_\epsilon^{2\beta}\right) = |\hat{K}(\omega_{1\epsilon})|^2 \rho_\epsilon^2 + o\left(h_\epsilon^{2\beta}\right). \tag{2.10}$$

This implies (2.9).

Theorems 2.1.1 and 2.1.2 admit the interpretation from the confidence estimation viewpoint.

We say that the family of confidence sets $U_\epsilon(Y_\epsilon)$ with confidence coefficient $1 - \alpha$ is $\Im_\epsilon(\rho_\epsilon)$-asymptotically minimax if for any other confidence sets $U_{1\epsilon}(Y_\epsilon)$ with the same confidence coefficient $1 - \alpha$

$$\liminf_{\epsilon \to 0} \sup_{S \in \Im_\epsilon(\rho_\epsilon)} (P_S(S \in U_{1\epsilon}(Y_\epsilon)) - P_S(S \in U_\epsilon(Y_\epsilon))) \ge 0$$

for each family $\rho_\epsilon \to 0$ as $\epsilon \to 0$.

Define the confidence sets

$$U_\epsilon(h_\epsilon, Y_\epsilon, x_\alpha) = \{S : T_{h_\epsilon}(Y_\epsilon - S) \le x_\alpha, S \in L_2(0,1)\}$$

with x_α defined by the equation $1 - \Phi(x_\alpha) = \alpha$.

Theorem 2.1.3. *Let the assumptions of Theorem 2.1.1 be satisfied. Then* $U_\epsilon(h_\epsilon, Y_\epsilon, x_\alpha)$ *are* $\Im_{\epsilon,h_\epsilon}(\rho_\epsilon)$*-asymptotically minimax confidence sets and* $U_\epsilon(h_\epsilon, Y_\epsilon, x_\alpha) \cap W_2^{(\beta)}(P_0)$ *are* $\Im_{\epsilon,h_\epsilon}(\rho_\epsilon) \cap W_2^{(\beta)}(P_0)$*-asymptotically minimax confidence sets.*

The proof is omitted. The reasoning are akin to the proof of similar statement on the relation of uniformly most powerful tests and uniformly most accurate confidence intervals.

2.2. **Testing hypotheses on nonparametric regression**

We shall follow to the setting in Brown and Low [4].

Let $H(\cdot)$ be an increasing c.d.f. in $[0,1]$. Let $S(\cdot) : [0,1] \to R^1$ and $\lambda^2(\cdot) : [0,1] \to (0,1)$ be measurable functions.

The independent random variables $(x_{ni}, Y_{ni}), 1 \le i \le n$ are observed with

$$x_{ni} = H^{-1}\left(\frac{i}{n+1}\right)$$

and

$$Y_{ni} = S(x_{ni}) + \lambda(x_{ni})\xi_{ni}, \quad \xi_{ni} \sim N(0,1).$$

Suppose the functions $\lambda^2(\cdot)$ and $H(\cdot)$ are continuously differentiable and such that

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\log\lambda(t)\right| < C, \qquad 0 < c < p(t) \doteq \frac{\mathrm{d}H}{\mathrm{d}t}(t) < C, \qquad t \in [0,1]. \tag{2.11}$$

Denote $q(t) = \lambda(t)p^{-1/2}(t)$.

The problem is to test a hypothesis $S(t) = S_0(t), t \in [0,1]$ for a given function $S_0(t), t \in [0,1]$.

Let $h_n > 0, h_n \to 0$ as $n \to \infty$ be a given sequence. Define the kernel-based test statistics

$$T_n(Y_n) = \int_0^1 \left(\frac{1}{n}\sum_{i=1}^n K_{h_n}(t - x_{ni})Y_{ni} - \int_0^1 K_{h_n}(t-s)S_0(s)\mathrm{d}s\right)^2 r(t)\mathrm{d}t$$

with $Y_n = \{Y_{ni}\}_{i=1}^n$.

Define the functional $\bar{T}_n(S) = T_n(S_n)$ where $S_n = \{S(x_{ni})\}_{i=1}^n$.

We fix a sequence $c_n > 0, c_n \to 0$ as $n \to \infty$ and denote

$$\Im_n(h_n, c_n) = \left\{S : \int_0^1 \left(\frac{1}{n}\sum_{i=1}^n K_{h_n}(t - x_{ni})S(x_{ni}) - \int_0^1 K_{h_n}(t-s)S(s)\mathrm{d}s\right)^2 r(t)\mathrm{d}t \right.$$
$$\left. < c_n \int_0^1 \left(\int_0^1 K_{h_n}(t-s)(S(s) - S_0(s))\mathrm{d}s\right)^2 r(t)\mathrm{d}t, \ \ S \in L_2(0,1)\right\}. \tag{2.12}$$

The sets $\Im_{nh_n}$ of alternatives equal

$$\Im_{nh_n}(\rho_n) = \{S : \bar{T}_n(S) > \rho_n(h_n) > 0, S \in \Im_n(h_n, c_n)\}$$

or

$$\Im_{nh_n}(\rho_n, \beta, P_0) = \Im_{nh_n}(\rho_n) \cap W_2^{(\beta)}(P_0)$$

where

$$W_2^{(\beta)}(P_0) = \left\{S : \int_0^1 (S(s) - S_0(s))^2 + (S^{(\beta)}(s) - S_0^{(\beta)}(s))^2)\mathrm{d}s < P_0, \ \ S \in L_2(0,1)\right\}.$$

Remark 2.2.1. We test a hypothesis using the discrete observations. Thus it seems natural to make some assumptions on approximation properties of the following type

$$\int_0^1 \left(\frac{1}{n}\sum_{i=1}^n K_{h_n}(t - x_{ni})S(x_{ni}) - \int_0^1 K_{h_n}(t-s)S(s)\mathrm{d}s\right)^2 r(t)\mathrm{d}t = o(\rho_n(h_n))$$

if

$$\int_0^1 \left(\int_0^1 K_{h_n}(t-s)(S(s) - S_0(s))\mathrm{d}s\right)^2 r(t)\mathrm{d}t = O(\rho_n(h_n)).$$

The inequality in (2.12) can be interpreted as an extension of this assumption on the more distant alternatives.

Assume as follows:

A1. There exists $\gamma > 0$ such that

$$\int_{-\infty}^{\infty} (K(u_1 - s) - K(u_2 - s))^2 \mathrm{d}s < C|u_1 - u_2|^{1+\gamma} \tag{2.13}$$

for all $u_1, u_2 \in [0, 1]$.

A2. There exists $\kappa > 1/2$ such that

$$|S_0(u_1) - S_0(u_2)| < C|u_1 - u_2|^{\kappa}$$

for all $u_1, u_2 \in [0, 1]$.

Theorem 2.2. *Assume A1, A2 and (2.3). Let the assumptions of Theorem 2.1.1 be satisfied with $\epsilon = n^{-1/2}$. Let $n^{-1} h_n^{-3/2-\omega} \to 0$ as $n \to \infty$ with $\omega > 0$. Then the sequence of tests*

$$L_n = \chi \left\{ n h_n^{1/2} \sigma^{-1} (T_n(Y_n) - d_n(h_n)) > x_\alpha \right\} \tag{2.14}$$

is asymptotically minimax for the sets of alternatives $\Im_{nh_n}(\rho_n)$ and (2.2) holds.

The lower bound (2.2) is attained for any sequence $S_n \in \Im(h_n, c_n)$ such that $T_n(S_n) = \rho_n(h_n)$, that is, equation (2.3) holds.

Let A and (2.4) hold also. Then the sequence of tests L_n is asymptotically minimax for the sets of alternatives $\Im_{nh_n}(\rho_n, \beta, P_0)$ and (2.2) holds with

$$\beta_{nh_n}(L_n) = \beta_{nh_n}(L_n, \Im(\rho_n, \beta, P_0)).$$

Remark 2.2.2. The main goal of paper is to prove lower bounds of minimax type for the kernel-based tests and to show that these lower bounds are principally attained. In some settings the assumptions are rather strong. In theorems we pointed out that the asymptotic of type II error probabilities are the same for all sequences of alternatives S_n having a given distance $T_n(S_n) = \rho_n(h_n)$. One can suppose that the statements of such a type can be proved for essentially more wider assumptions and for essentially more wider classes of statistical models. The proof of lower bounds are more difficult and can be considered as serious additional argument for the analysis of kernel-based tests in distance terms.

Remark 2.2.3. The procedure of hypothesis testing is based on the comparison of kernel estimator with the smoothed signal $K_{h_n} * S_0$. The smoothing may cause the losses of information about the signal S_0. Such a losses will be absent if

$$||S_0||^2 - ||K_{h_n} * S_0|| = \int (1 - \hat{K}^2(h_n\omega)) \hat{S}_0^2(\omega) \mathrm{d}\omega = o(\rho_n) = o\left(h^{-1/2} n^{-1}\right). \tag{2.15}$$

Let $\hat{K}(\omega) = 1 - C|\omega|^\gamma (1 + o(1))$ in some vicinity of $\omega = 0$. Then

$$||S_0||^2 - ||K_{h_n} * S_0|| = C h_n^\gamma \int |\omega|^\gamma \hat{S}_0^2(\omega) \mathrm{d}\omega (1 + o(1)).$$

Thus it suffices to put $h_n = o(n^{-\frac{2}{1+2\gamma}})$ and (2.15) will be hold. If $\gamma = 1$, we get $h_n = o(n^{-2/3}), \rho_n = O(n^{-2/3}), \beta = 1/4$ and assumptions of Theorem 2.2 do not fulfilled. If $\gamma = 2$, we get $h_n = o(n^{-2/5}), \rho_n = O(n^{-4/5})$ and $\beta = 1$. Thus all the assumptions of Theorems 2.2 and Theorems 2.3, 3.2 given below are satisfied. Therefore, if we apply the hypothesis testing procedure with $h_n \asymp n^{-\lambda}, \lambda > 2/5$, we test the hypothesis *versus* alternatives having more serious fluctuation then the signal S_0.

Remark 2.2.4. The difference between the rates of consistent distinguishability $n^{-\frac{4\beta}{4\beta+1}}$ (or $n^{-1}h_n^{-1/2}$) in testing nonparametric hypotheses and $n^{-1/2}$ in testing parametric hypotheses is essentially smaller ($n^{-\frac{1}{4\beta+1}}$) then the corresponding difference ($n^{-\frac{1}{2\beta+1}}$) in estimation theory. If the sample size $n \le 2000$, the choice of bandwidth $O(n^{-\frac{2}{4\beta+1}})$ for the smoothness parameter $\beta \ge 2$ is approximately the same as in the testing with the kernel-based tests of parametric hypotheses. Thus, for sufficiently smooth signals, there exists small difference in interpretation of results of kernel-based procedure for parametric and nonparametric settings. The most essential difference is that we get uniform estimates of distinguishability in terms of the sets $\Im_{n,h_n}(\rho_n)$ for nonparametric setting. If we want to test the hypothesis *versus* fast oscilating nonparametric sets of signals, the definition of sets $\Im_{n,h_n}(\rho_n)$ shows clearly the types of oscillations that can be distinguished. This is the signals with oscillation width $\asymp 2h_n$ or $3h_n$ and the amplitude $\asymp \frac{\rho_n}{3l\sigma^2 h_n}$ where l is the number of oscillation peaks.

2.3. Nonparametric hypothesis testing on a density

Let $X_1,\dots,X_n$ be i.i.d.r.v.'s with c.d.f. $F(x)$, $x \in [0,1]$. The problem is to test a hypothesis $F(x) = F_0(x), x \in (0,1)$, where F_0 is a given c.d.f. We suppose $F_0(x)$ is absolutely continuous w.r.t. Lebesgue measure and has the density $f_0(x) = \frac{\mathrm{d}F_0}{\mathrm{d}x}(x)$, $x \in (0,1)$.

Denote $\hat{F}_n$ the empirical c.d.f. of $X_1,\dots,X_n$.

The kernel-based test statistics are defined as follows

$$T_n(\hat{F}_n) = \int_0^1 \left(\frac{1}{n}\sum_{i=1}^n K_{h_n}(t-X_i) - \int_0^1 K_{h_n}(t-s) f_0(s)\mathrm{d}s \right)^2 r(t)\mathrm{d}t$$
$$= \int_0^1 \left(\int_0^1 K_{h_n}(t-s)\mathrm{d}(\hat{F}_n(s) - F_0(s)) \right)^2 r(t)\mathrm{d}t.$$

The functionals T_n defining the sets of alternatives equal

$$T_n(F) \doteq T_n(F,F_0) \doteq \int_0^1 \left(\int_0^1 K_{h_n}(t-s)\mathrm{d}(F(s) - F_0(s)) \right)^2 r(t)\mathrm{d}t.$$

Make the following assumptions:

B. The density f_0 satisfies the Hoelder condition

$$|f_0(x) - f_0(y)| < C|x-y|^{\kappa}, \quad x,y \in [0,1] \tag{2.16}$$

with $\kappa > 1/2$ and $f_0(x) > c > 0$ for all $x \in [0,1]$.

C.

$$|r(x) - r(y)| < C|x-y|^{\kappa_1} \quad \text{for all } x,y \in [0,1] \quad \text{and } \kappa_1 > \frac{1}{2}. \tag{2.17}$$

We fix values $\zeta > \frac{1}{2}$ and $C > 0, c > 0$ and define the set $\Im = \Im(C,c,\zeta)$ of all distribution functions such that

$$F(h) + 1 - F(1-h) < Ch^{\zeta} \tag{2.18}$$

for all $0 < h < c$.

The sets of alternatives equal

$$\Im_{nh_n} = \Im_{nh_n}(\rho_n) = \{F : T_n(F) > \rho_n(h_n) > 0, F \in \Im\}$$

or

$$\Im_{nh_n} = \Im_{nh_n}(\rho_n,\beta,P_0) = \Im_{nh_n}(\rho_n) \cap W_2^{(\beta)}(P_0)$$

where

$$W_2^{(\beta)}(P_0) = \left\{ f : \int_0^1 (f(s) - f_0(s))^2 + (f^{(\beta)}(s) - f_0^{(\beta)}(s))^2 \mathrm{d}s < P_0, f(s) = \frac{\mathrm{d}F}{\mathrm{d}s}(s),\ F \in \Im \right\}.$$

In what follows, we shall make use of the same notation as in the problem of signal detection putting $\epsilon = n^{-1/2}$ and $q(t) = f_0^{1/2}(t), t \in [0,1]$. In particular

$$d_n(h_n) \doteq d_n(h_n, f_0) \doteq \frac{1}{nh_n} \int_{-1}^1 K^2(s)\mathrm{d}s \int_0^1 r(t) f_0(t)\mathrm{d}t,$$

$$\sigma^2 = \sigma^2(f_0) = 2\int_{-2}^2 \left(\int_{-1}^1 K(u+v)K(u)\mathrm{d}u \right)^2 \mathrm{d}v \int_0^1 f_0^2(t) r^2(t)\mathrm{d}t.$$

Theorem 2.3. *Assume A1, B, C and let the assumptions of Theorem 2.1 be satisfied with $\epsilon = n^{-1/2}$. Let $n^{-1}h_n^{-3/2-\omega} \to 0$ as $n \to \infty$ with $\omega > 0$. Then the sequence of tests*

$$L_n = \chi\left\{ nh_n^{1/2}\sigma^{-1}(T_n(\hat{F}_n) - d_n(h_n)) > x_\alpha \right\}$$

is asymptotically minimax and (2.2) holds.

Let A and (2.4) hold also. Then the sequence of tests L_n is asymptotically minimax for the sets of alternatives $\Im_{nh_n}(\rho_n, \beta, P_0)$ and (2.2) holds with

$$\beta_{nh_n}(L_n) = \beta_{nh_n}(L_n, \Im(\rho_n, \beta, P_0)).$$

Remark 2.3.1. The tests based on kernel estimators of density are usually treated as nonparametric tests for testing hypothesis on a density. In this setting we apply these tests for a more wide sets of alternatives defined on the sets of distribution functions.

Remark 2.3.2. The proofs of lower bounds in Theorems 2.2 and 2.3 are based on the statements about asymptotic equivalence of statistical experiments (see Brown and Low [4], Nussbaum [27]). The problem of hypothesis testing on a density is asymptotically equivalent to the problem of signal detection

$$\mathrm{d}Y(t) = f(t)\mathrm{d}t + \frac{1}{\sqrt{n}} f_0^{1/2}(t)\mathrm{d}w(t)$$

in the Gaussian white noise with the weight function $f_0^{1/2}(t)$ (see Nussbaum [27]). Since our model (1.1) of signal detection also contains the weight function $q(t)$ we can apply the theorem on asymptotic equivalence of statistical experiments putting $q(t) = f_0^{1/2}(t)$.

Remark 2.3.3. It is easy to see from the proof of Theorem 2.3 that the assumptions of theorem can be weaken. In the definition of sets $\Im_{nh_n}$ of alternatives the set $\Im = \Im(\zeta, C, c)$ can be replaced by the set of all distribution functions. In such a setting the statement of Theorem 2.3 holds for the sequence of test statistics

$$\hat{T}_n(\hat{F}_n) = T_n(\hat{F}_n) - \int_0^1 r(t)\mathrm{d}t \int_0^1 \left(K_{h_n}(t-x) - \int_0^1 K_{h_n}(t-s) f_0(s)\mathrm{d}s \right)^2 \mathrm{d}\hat{F}_n(x) \tag{2.19}$$

generating the sequence of tests

$$L_{1n} = \chi(nh_n^{1/2}\sigma^{-1}\hat{T}_n(\hat{F}_n) > x_\alpha).$$

The last addendum in the right-hand side of (2.19) deletes the component of bias $E_F(T_n(\hat{F}_n))$ having the order greater then $n^{-1}h_n^{-1/2} = O((\mathrm{Var}(T_n(\hat{F}_n)))^{1/2})$. Without deleting this term we need to estimate more accurately the boundary effects in asymptotic of $E_F(T_n(\hat{F}_n))$ and to assume (2.17, 2.18).

3. Main results. Parametric hypothesis

We begin with the study of problem of signal detection.

Suppose we observe a random process $Y_\epsilon(t)$ defined by a stochastic differential equation (1.1) with an unknown signal $S(t)$. The problem is to test a parametric hypothesis $S(t) = S(t,\theta), \theta \in \Theta \subset R^l$ *versus* nonparametric sets of alternatives

$$S \in \Im_{\epsilon,h_\epsilon}(\Theta) = \Im_{\epsilon,h_\epsilon}(\Theta,\rho_\epsilon) = \left\{S : \inf_{\theta\in\Theta} \int_0^1 \left(\int_0^1 K_{h_\epsilon}(t-s)(S(s)-S(s,\theta))\mathrm{d}s\right)^2 r(t)\mathrm{d}t > \rho_\epsilon(h_\epsilon) > 0,\ S \in L_2(0,1)\right\}$$

or

$$S \in \Im_{\epsilon,h_\epsilon}(\Theta) = \Im_{\epsilon,h_\epsilon}(\Theta,\rho_\epsilon,\beta,P_0) = \Im_{\epsilon,h_\epsilon}(\Theta,\rho_\epsilon) \cap W_2^{(\beta)}(P_0,\Theta)$$

where

$$W_\beta(P_0,\Theta) = \left\{S : \int_0^1 (S(s)-S(s,\theta))^2 + (S^{(\beta)}(s) - S^{(\beta)}(s,\theta))^2 \mathrm{d}s < P_0,\ \text{with}\, \theta = \tilde\theta(S) = \mathrm{argmin}_\theta\, T_\epsilon(S,\theta)\right\}\cdot$$

Thus, in the case of sets of alternatives $\Im_{\epsilon,h_\epsilon}(\Theta,\rho_\epsilon,\beta,P_0)$, we assume that there exists β-derivative $S^{(\beta)}(s,\theta)$ of a signal $S(s,\theta), \theta \in \Theta$ and $\int_0^1((S^{(\beta)}(s,\theta))^2\mathrm{d}s < \infty$.

Suppose the set Θ is a closure of bounded open set in R^l.

Let $\hat\theta_\epsilon$ be an estimator of unknown parameter $\theta \in \Theta$. Define the test statistics

$$T_\epsilon(Y_\epsilon,\hat\theta_\epsilon) = \int_0^1 \left(\hat S_{h_\epsilon}(t) - \int_0^1 K_{h_\epsilon}(t-s)S(s,\hat\theta_\epsilon)\mathrm{d}s\right)^2 r(t)\mathrm{d}t.$$

For any test U denote $\alpha_\theta = E_\theta(U)$ its type I error probability for the hypothesis $\theta \in \Theta$. We put $\beta_{\epsilon,h_\epsilon}(U) = \beta_{\epsilon,h_\epsilon}(U,\Im_{\epsilon,h_\epsilon}(\Theta)) = \sup\{\beta(U,S) : S \in \Im_{\epsilon,h_\epsilon}(\Theta)\}$.

We say that a family of tests U_ϵ, $\epsilon > 0$, $\alpha_\theta(U_\epsilon) = E_\theta(U_\epsilon) \le \alpha > 0$, $\theta \in \Theta$ is uniformly asymptotically minimax on the sets of alternatives $\Im_{\epsilon,h_\epsilon}(\Theta)$ if the family of tests U_ϵ is asymptotically minimax for each fixed $\theta \in \Theta$ in the problems of testing the simple hypothesis $S(s) = S(s,\theta)$ *versus* $S \in \Im_{\epsilon,h_\epsilon}(\Theta)$.

For a wide class of estimators $\hat\theta_\epsilon$ we prove that the test statistics $T_\epsilon(Y_\epsilon,\hat\theta_\epsilon)$ generates uniformly asymptotically minimax families of tests.

Denote $u'v$ the inner product of $u, v \in R^l$.

Assume as follows:

D1. For all $\theta_1,\theta_2 \in \Theta, \theta_1 \ne \theta_2$

$$\int_0^1 (S(s,\theta_1) - S(s,\theta_2))^2 \mathrm{d}s \ne 0.$$

Suppose $S(s,\theta)$ is differentiable in $\theta \in \Theta$ and denote $S_{\theta_i}(s,\theta) = \frac{\partial S(s,\theta)}{\partial\theta_i}$ the partial derivatives of $S(s,\theta)$ for all $1 \le i \le l, s \in [0,1], \theta = (\theta_1,\dots,\theta_l) \in \Theta$. Denote $S_\theta(s,\theta) = \{S_{\theta_i}(s,\theta)\}_{i=1}^l$.

D2. There exists $\omega > 0$ such that for all $\theta_1,\theta_2 \in \Theta$

$$\int_0^1 (S(s,\theta_2) - S(s,\theta_1) - (\theta_2-\theta_1)'S_\theta(s,\theta_1))^2 \mathrm{d}s < C|\theta_2-\theta_1|^{2+\omega}.$$

D3. Uniformly in $\theta \in \Theta$ it holds $\int_0^1 S^2_{\theta_i}(s,\theta)\mathrm{d}s < C, \quad 1 \le i \le l.$

D4. There exists a functional $\bar{\theta} : L_2(0,1) \to \Theta$ such that, $\bar{\theta}(S(\cdot,\theta)) = \theta$ for all $\theta \in \Theta$ and for any $\delta > 0$

$$P_S(|\hat{\theta}_\epsilon - \bar{\theta}(S)| > \delta T_\epsilon^{1/2}(S, \bar{\theta}(S))) = o(1)$$

$$P_S\left(|\hat{\theta}_\epsilon - \bar{\theta}(S)|^{2+\omega} > \delta h_\epsilon^{1/2} T_\epsilon\left(S, \bar{\theta}(S)\right)\right) = o(1)$$

uniformly in $S \in L_2(0,1)$ as $\epsilon \to 0$.

D5. There exists $\lambda_1(\delta) \to 0$ as $\delta \to 0$ such that for all $\theta \in \Theta$

$$\sup\{|S(s,\theta) - S(t,\theta)| : |t-s| < \delta : \, t,s \in [0,1]\} < \lambda_1(\delta).$$

D6. There exists $\lambda_2(\delta) \to 0$ as $\delta \to 0$ such that

$$\sup\left\{\int_0^1 |S^{(\beta)}(s,\theta_1) - S^{(\beta)}(s,\theta_2)|^2 \mathrm{d}s : |\theta_1 - \theta_2| < \delta, \theta_1, \theta_2 \in \Theta\right\} < \lambda_2(\delta).$$

Theorem 3.1. *Assume D1–D5. Let* $\epsilon^2 h_\epsilon^{-1/2} \to 0, h_\epsilon \to 0$ *as* $\epsilon \to 0$ *and (2.1) holds. Then the family of tests*

$$L_\epsilon = \chi(\epsilon^{-2} h_\epsilon^{1/2} \sigma^{-1}(T_\epsilon(Y_\epsilon, \hat{\theta}_\epsilon) - d_\epsilon(h_\epsilon)) > x_\alpha)$$

is uniformly asymptotically minimax for the sets of alternatives $\Im_{\epsilon,h_\epsilon}(\Theta, \rho_\epsilon)$ *and (2.2, 2.3) hold.*

Let A, D6 and (2.4) hold also. Then the family of tests L_ϵ *is uniformly asymptotically minimax for the sets of alternatives* $\Im_{\epsilon,h_\epsilon}(\Theta, \rho_\epsilon, \beta, P_0)$ *and (2.2) holds with* $\beta_{\epsilon,h_\epsilon}(L_\epsilon) = \beta_{\epsilon,h_\epsilon}(L_\epsilon, \Im_{\epsilon,h_\epsilon}(\Theta, \rho_\epsilon, \beta, P_0))$.

The problem of testing parametric hypotheses on a density *versus* nonparametric sets of alternatives will be treated in the following setting. Let $X_1, \ldots, X_n$ be i.i.d.r.v.'s with c.d.f. $F(x), x \in [0,1]$. One needs to test a hypothesis $F = F_\theta$, $\theta \in \Theta$ *versus*

$$F \in \Im_{nh_n}(\Theta, \rho_n) = \{F : \inf\{T_n(F, F_\theta) : \theta \in \Theta\} > \rho_n(h_n), F \in \Im\}\cdot$$

Suppose that c.d.f.'s $F_\theta, \theta \in \Theta$ are absolutely continuous w.r.t. Lebesgue measure and have the densities $f(x,\theta) = \frac{\mathrm{d}F_\theta}{\mathrm{d}x}(x), x \in (0,1)$.

Let $\hat{\theta}_n$ be an estimator of θ. We shall test the hypothesis on the base of test statistics $\hat{T}_n = T_n(\hat{F}_n, F_{\hat{\theta}_n})$.

Make the following assumptions:

B1. There exists $\kappa > 1/2$ and $C > 0$ such that, for all $\theta \in \Theta$,

$$|f(x,\theta) - f(y,\theta)| < C|x-y|^\kappa, \quad x,y \in [0,1].$$

B2. There exist $C > c > 0$ such that $0 < c < f(x,\theta) < C < \infty$ for all $x \in [0,1]$ and $\theta \in \Theta$.

E1. For all $\theta \in \Theta$ it holds $F_\theta \in \Im$.

E2. The assumptions D1–D3, D5 hold with $S(s,\theta) = \sqrt{f(s,\theta)}, \theta \in \Theta$.

E3. For each c.d.f. $F(x) \in \Im$ there exists $\bar{\theta}(F) \in \Theta$ such that $\bar{\theta}(F_\theta) = \theta$ for all $\theta \in \Theta$ and for any $\delta > 0$

$$P_F(|\hat{\theta}_n - \bar{\theta}(F)|^2 > \delta T_n(F, \tilde{\theta}(F))) = o(1)$$

uniformly in $F \in \Im$.

Theorem 3.2. *Assume A1, B1, B2, C, E1–E3 and (2.1). Let $n^{-1}h_n^{-3/2-\omega} \to 0$ as $n \to \infty$ with $\omega > 0$. Then the sequence of tests*

$$L_n = \chi\{nh_n^{1/2}\sigma^{-1}(f_{\hat\theta_n})(T_n(\hat F_n, F_{\hat\theta_n}) - d_n(h_n, f_{\hat\theta_n})) > x_\alpha\}$$

is uniformly asymptotically minimax and

$$\beta_{nh_n}(L_n) = \sup_{\theta\in\Theta} \Phi\left(x_\alpha - nh_n^{1/2}\sigma(f_\theta)\rho_n(h_n)\right)(1+o(1)) \tag{3.1}$$

as $n \to \infty$.

We begin with the proof of Theorem 3.1. The proof of Theorem 2.1.1 is obtained by an easy modification of these arguments.

4. Proof of Theorem 3.1

To simplify notation we suppose that θ is one dimensional parameter, $\theta \in \Theta \subset R^1$.

First of all we study the asymptotic behaviour of test statistics $T(Y_\epsilon, \hat\theta_\epsilon)$ and prove the upper bound in (2.2).

Let $S(s) \in \Im_{\epsilon,h_\epsilon}(\Theta)$ be a true value of a signal. We have

$$\int_0^1 K_{h_\epsilon}(t-s)\left(\mathrm{d}Y_\epsilon(s) - S\left(s,\hat\theta_\epsilon\right)\mathrm{d}s\right) = g_{1h_\epsilon}(t) + g_{2h_\epsilon}(t) + \xi_\epsilon(t) \tag{4.1}$$

with

$$g_{1h_\epsilon}(t) = \int_0^1 K_{h_\epsilon}(t-s)(S(s) - S(s,\theta(S)))\mathrm{d}s,$$
$$g_{2h_\epsilon}(t) = \int_0^1 K_{h_\epsilon}(t-s)(S(s,\theta(S)) - S(s,\hat\theta_\epsilon))\mathrm{d}s,$$
$$\xi_\epsilon(t) = \epsilon\int_0^1 K_{h_\epsilon}(t-s)q(s)\mathrm{d}w(s).$$

Hence we get

$$T(Y_\epsilon,\hat\theta_\epsilon) = I_{1\epsilon} + I_{2\epsilon} + I_{3\epsilon} + I_{4\epsilon} + I_{5\epsilon} + I_{6\epsilon} \tag{4.2}$$

with

$$I_{1\epsilon} = \int_0^1 g_{1h_\epsilon}^2(t)r(t)\mathrm{d}t, \qquad I_{2\epsilon} = 2\int_0^1 g_{1h_\epsilon}(t)g_{2h_\epsilon}(t)r(t)\mathrm{d}t, \tag{4.3}$$

$$I_{3\epsilon} = \int_0^1 g_{2h_\epsilon}^2(t)r(t)\mathrm{d}t, \qquad I_{4\epsilon} = 2\int_0^1 g_{1h_\epsilon}(t)\xi_\epsilon(t)r(t)\mathrm{d}t, \tag{4.4}$$

$$I_{5\epsilon} = 2\int_0^1 g_{2h_\epsilon}(t)\xi_\epsilon(t)r(t)\mathrm{d}t, \quad I_{6\epsilon} = \int_0^1 \xi_\epsilon^2(t)r(t)\mathrm{d}t. \tag{4.5}$$

Since $S(s) \in \Im_{\epsilon,h_\epsilon}(\Theta,\rho_\epsilon(h_\epsilon))$ we have

$$I_{1\epsilon} > \rho_\epsilon(h_\epsilon). \tag{4.6}$$

Note that for any function $U \in L_2(0,1)$ it holds

$$\int_0^1 r(t)\left(\int_0^1 K_h(t-s)U(s)\mathrm{d}s\right)^2 \mathrm{d}t \le Ch^{-2}\int_0^1\left(\int_{t-h}^{t+h}|U(s)|\mathrm{d}s\right)^2\mathrm{d}t \tag{4.1}$$

$$\le Ch^{-1}\int_0^1\int_{t-h}^{t+h}U^2(s)\mathrm{d}s\mathrm{d}t < C\int_0^1 U^2(t)\mathrm{d}t. \tag{4.7}$$

Denote

$$W(s) = S\left(s, \hat{\theta}_\epsilon\right) - S\left(s, \bar{\theta}(S)\right) - \left(\hat{\theta}_\epsilon - \bar{\theta}(S)\right) S_\theta\left(s, \bar{\theta}(S)\right).$$

By (4.7), D2–D4, we get

$$I_{3\epsilon} < 2(\hat{\theta}_\epsilon - \bar{\theta}(S))^2 \int_0^1 r(t) \left(\int_0^1 K_{h_\epsilon}(t-s) S_\theta(s, \bar{\theta}(S)) \mathrm{d}s\right)^2 \mathrm{d}t + 2\int_0^1 r(t) \left(\int_0^1 K_{h_\epsilon}(t-s) W(s) \mathrm{d}s\right)^2 \mathrm{d}t \le C|\hat{\theta}_\epsilon - \bar{\theta}(S)|^2. \tag{4.8}$$

We have

$$E(I_{4\epsilon}) = 0. \tag{4.9}$$

Define the operators $\bar{K}_{2,qh}$ and $\bar{K}_{2,h}$ with the kernels $K_{2,qh}(t_1,t_2) = \int_0^1 K_h(t_1 - s) q^2(s) K_h(t_2 - s) \mathrm{d}s$ and $K_{2,h}(t_1,t_2)$ respectively. The operators $\bar{K}_{2,qh}$ and $\bar{K}_{2,h}$ are nonnegative. Since $\bar{K}_{2,qh} < C\bar{K}_{2,h}$ and the kernel K is bounded we get

$$\begin{aligned}
\mathrm{Var}(I_{4\epsilon}) &= 4\int_0^1 r(t_1)\mathrm{d}t_1 \int_0^1 r(t_2)\mathrm{d}t_2\, g_{1h_\epsilon}(t_1) g_{1h_\epsilon}(t_2) E(\xi_\epsilon(t_1)\xi_\epsilon(t_2)) \\
&= 4\epsilon^2 \int_0^1 r(t_1)\mathrm{d}t_1 \int_0^1 r(t_2)\mathrm{d}t_2 g_{1h_\epsilon}(t_1) g_{1h_\epsilon}(t_2) K_{2,qh_\epsilon}(t_1,t_2) \\
&\le C\epsilon^2 I_{1\epsilon}^{1/2} \left(\int_0^1 r(t_1)\mathrm{d}t_1 \left(\int_0^1 K_{2,h_\epsilon}(t_1,t_2) g_{1h_\epsilon}(t_2) r(t_2)\mathrm{d}t_2\right)^2\right)^{1/2} \\
&\le C\epsilon^2 I_{1\epsilon}^{1/2} \left(\int_0^1 r(t_1)\mathrm{d}t_1 \int_0^1 |K_{2,h_\epsilon}(t_1,t_2)| r(t_2)\mathrm{d}t_2 \int_0^1 |K_{2,h_\epsilon}(t_1,t_3)| g_{1h_\epsilon}^2(t_3) r(t_3)\mathrm{d}t_3\right)^{1/2} \\
&\le C\epsilon^2 I_{1\epsilon}^{1/2} \left(\int_0^1 r(t_1)\mathrm{d}t_1 \int_0^1 |K_{2,h_\epsilon}(t_1,t_3)| g_{1h_\epsilon}^2(t_3) r(t_3)\mathrm{d}t_3\right)^{1/2} \\
&\le C\epsilon^2 I_{1\epsilon}^{1/2} \left(h_\epsilon^{-1} \int_0^1 r(t_1)\mathrm{d}t_1 \int_{t_1 - 2h_\epsilon}^{t_1 + 2h_\epsilon} g_{1h_\epsilon}^2(t_3) r(t_3)\mathrm{d}t_3\right)^{1/2} \le C\epsilon^2 I_{1\epsilon}.
\end{aligned} \tag{4.10}$$

By Schwartz inequality, we get

$$I_{2\epsilon} \le 2 I_{1\epsilon}^{1/2} I_{3\epsilon}^{1/2}. \tag{4.11}$$

We have

$$I_{5\epsilon}^2 \le 2I_{51\epsilon}^2 + 2I_{52\epsilon}^2 \tag{4.12}$$

with

$$I_{51\epsilon} \doteq 2(\hat{\theta}_\epsilon - \bar{\theta}(S)) Q_\epsilon \doteq 2(\hat{\theta}_\epsilon - \bar{\theta}(S)) \int_0^1 \xi_\epsilon(t) r(t)\mathrm{d}t \int_0^1 K_{h_\epsilon}(t-s) S_\theta(s, \bar{\theta}(S))\mathrm{d}s, \tag{4.13}$$

$$I_{52\epsilon} = 2\int_0^1 r(t)\xi_\epsilon(t)\mathrm{d}t \int_0^1 K_{h_\epsilon}(t-s) W(s)\mathrm{d}s. \tag{4.14}$$

By (4.7), we get

$$
\begin{aligned}
EQ_\epsilon^2 &= \int_0^1 q^2(s)\left(\int_0^1 K_{h_\epsilon}(s-t)r(t)\int_0^1 K_{h_\epsilon}(t-s_1)S_\theta(s_1,\bar\theta(S))\mathrm{d}s_1\mathrm{d}t\right)^2\mathrm{d}s \\
&\le C\int_0^1 q^2(s)\left(\int_0^1 K_{h_\epsilon}(s-t)r(t)S_\theta(t,\bar\theta(S))\mathrm{d}t\right)^2\mathrm{d}s \\
&\le C\int_0^1 q^2(s)r^2(s)S_\theta^2(s,\bar\theta(S))\mathrm{d}s \le C<\infty. \qquad (4.15)
\end{aligned}
$$

By Schwartz inequality, we get

$$I_{52\epsilon} \le J_{51\epsilon}J_{52\epsilon} \qquad (4.16)$$

with

$$
\begin{aligned}
J_{51\epsilon}^2 &= \int_0^1 r(t)\mathrm{d}t\left(\int_0^1 K_{h_\epsilon}(t-s)W(s)\mathrm{d}s\right)^2, \\
J_{52\epsilon}^2 &= I_{6\epsilon} = \int_0^1 r(t)\xi_\epsilon^2(t)\mathrm{d}t.
\end{aligned}
$$

By (4.7), D2, we get

$$J_{51\epsilon}^2 \le C\int_0^1 r(t)W^2(t)\mathrm{d}t \le C|\hat\theta_\epsilon-\bar\theta(S)|^{2+\omega}. \qquad (4.17)$$

Estimating similarly to (4.10), we get

$$E(J_{52\epsilon}^2) \le \epsilon^2\int_0^1 q^2(s)K_{2,rh_\epsilon}(s,s)\mathrm{d}s \le \epsilon^2 h_\epsilon^{-1} \qquad (4.18)$$

with

$$K_{2,rh_\epsilon}(y_1,y_2) = \int_0^1 K_{h_\epsilon}(y_1-t)r(t)K_{h_\epsilon}(y_2-t)\mathrm{d}t, \quad y_1,y_2\in[0,1].$$

By (4.16–4.18), we get

$$I_{52\epsilon} \le \epsilon h_\epsilon^{-1/2}|\hat\theta_\epsilon-\bar\theta(S)|^{1+\omega/2}. \qquad (4.19)$$

By D4, (4.13–4.15, 4.19), we get

$$I_{5\epsilon}^2 = O_P\left(\epsilon^2 h_\epsilon^{(\omega-1)/2}\right). \qquad (4.20)$$

By straightforward calculations, arguing similarly to Hall [16, 17], we get

$$E(I_{6\epsilon}) = d_\epsilon(h_\epsilon)(1+O(h_\epsilon)), \qquad (4.21)$$

$$\mathrm{Var}(I_{6\epsilon}) = 2\epsilon^4\int_0^1 r(t_1)\mathrm{d}t_1\int_0^1 r(t_2)\mathrm{d}t_2 K_{2,qh_\epsilon}^2(t_1,t_2). \qquad (4.22)$$

Putting $t_2=t_1+vh_\epsilon, s=t_1-uh_\epsilon$, we get

$$
\begin{aligned}
\mathrm{Var}(I_{6\epsilon}) = {}&2\epsilon^4\int_0^1 r(t_1)\mathrm{d}t_1\int_{-t_1/h_\epsilon}^{(1-t_1)/h_\epsilon} r(t_1+vh_\epsilon) \\
&\times\left(\int_{(t_1-1)/h_\epsilon}^{t_1/h_\epsilon} K(u)q^2(t_1-uh_\epsilon)K(u+v)\mathrm{d}u\right)^2\mathrm{d}v = \frac{\epsilon^4\sigma^2}{h_\epsilon}(1+o(1)). \qquad (4.23)
\end{aligned}
$$

By D4 (4.6, 4.8–4.11, 4.20–4.23) together, we get that $\epsilon^2 h_\epsilon^{-1/2} = O(I_{1\epsilon})$ implies

$$I_{2\epsilon} + I_{3\epsilon} + I_{4\epsilon} + I_{5\epsilon} = o_P(I_{1\epsilon} + I_{6\epsilon} - d_\epsilon(h_\epsilon)) \tag{4.24}$$

as $\epsilon \to 0$.

Lemma 4.1. *Let the assumptions of Theorem 3.1 be satisfied. Then the distributions of $h_\epsilon^{1/2}\epsilon^{-2}\sigma^{-1}(I_{6\epsilon}(h_\epsilon) - d_\epsilon(h_\epsilon))$ converge to the standard normal one.*

By (4.6, 4.24) and Lemma 4.1 we get (2.2) and (2.3). The proof of Lemma 4.1 will be given later.

It remains to prove the lower bounds for the type II error probabilities in the problems of testing a simple hypothesis $S = S(\theta_0), \theta_0 \in \Theta$ *versus* $S \in \Im_{\epsilon,h_\epsilon}(\Theta, \rho_\epsilon(h_\epsilon))$.

The proof of lower bounds is based on the wellknown fact that the Bayes risk does not exceed the minimax one. We fix $\delta > 0$ and introduce the family of Gaussian probability measures $\mu_{\epsilon\delta}$ which set by the random processes

$$\tilde S(t) = \tilde S_\epsilon(t) = S(t,\theta_0) + \tau r^{1/2}(t)\int_0^1 K_{h_\epsilon}(t-s)q(s)\mathrm{d}w_1(s)$$

where $\mathrm{d}w_1(s), s \in (0,1)$ is a Gaussian white noise and

$$\tau^2 = \tau^2_{\epsilon,\delta} = 2(1+\delta)\rho_\epsilon(h_\epsilon)h_\epsilon\sigma^{-2}.$$

The Bayes probability measure $\nu_{\epsilon\delta}$ is defined as the conditional probability measure of $\tilde S$ under the condition $\tilde S \in \Im_{\epsilon,h_\epsilon}(\Theta)$.

Lemma 4.2. *It holds*

$$((1+\delta)\rho_\epsilon(h_\epsilon))^{-1}\int_0^1\left(\int_0^1 K_{h_\epsilon}(t-s)(\tilde S(s) - S(s,\theta_0))\mathrm{d}s\right)^2 r(t)\mathrm{d}t \to 1 \tag{4.25}$$

and

$$((1+\delta)\rho_\epsilon(h_\epsilon))^{-1}\inf_{\theta\in\Theta}\int_0^1\left(\int_0^1 K_{h_\epsilon}(t-s)(\tilde S(s) - S(s,\theta))\mathrm{d}s\right)^2 r(t)\mathrm{d}t \to 1 \tag{4.26}$$

in probability as $\epsilon \to 0$.

This implies

$$P_{\mu_{\epsilon\delta}}(\tilde S \in \Im_{\epsilon h_\epsilon}(\Theta, \rho_\epsilon(h_\epsilon))) = 1 + o(1) \tag{4.27}$$

as $\epsilon \to 0$.

The proof of Lemma 4.2 will be given later.

Denote $\tilde U_\epsilon$ and U_ϵ a posteriory Bayes likelihood ratios generated by *a priori* Bayes probability measures $\mu_{\epsilon\delta}$ and $\nu_{\epsilon\delta}$ respectively. It is easy to see that (4.27) implies $\tilde U_\epsilon/U_\epsilon \to 1$ as $\epsilon \to 0$ in probability both in the case of hypothesis and Bayes alternatives $\nu_{\epsilon\delta}, \mu_{\epsilon\delta}$. This allows us to replace *a priori* Bayes probability measures $\nu_{\epsilon\delta}$ by *a priori* Bayes probability measure $\mu_{\epsilon\delta}$ in the further arguments. Therefore, for the proof of theorem, it suffices to find the representation of Bayes test statistic $D_{\epsilon\delta}(Y_\epsilon)$ corresponding to *a priori* probability measure $\mu_{\epsilon\delta}$ in a simple form and to show that, for the tests $U_{\epsilon\delta}$ generated by the test statistics $D_{\epsilon\delta}(Y_\epsilon)$, it holds

$$\lim_{\delta\to 0}\lim_{\epsilon\to 0}\left(\beta_{\epsilon,h_\epsilon}(L_\epsilon) - \int \beta(U_{\epsilon\delta}, S)\mathrm{d}\mu_{\epsilon\delta}\right) = 0. \tag{4.28}$$

Let us find Bayes *a posteriori* likelihood ratios in the case of *a priori* probability measures $\mu_{\epsilon\delta}$.

Let $\{\phi_j\}_1^\infty$ be an orthonormal system of functions in $L_2(0,1)$. Then (1.1) can be written as follows

$$y_j = s_j + \epsilon\xi_j, \quad 1 \le j < \infty$$

with $y_j = \int_0^1 \phi_j(t)\mathrm{d}Y_\epsilon(t)$, $s_j = \int_0^1 S(t)\phi_j(t)\mathrm{d}t$, $\xi_j = \int_0^1 \phi_j(t)q(t)\mathrm{d}w(t)$.

Define the operators Q, R such that $(Qu)(t) = q(t)u(t), (Ru)(t) = r(t)u(t)$ for any function $u \in L_2(0,1)$. Define also the operator K_h with the kernel $K_h(x-t)$ with $x,t \in [0,1]$ and the unit operator E. In (4.29–4.32) we shall make use of notation $Y_\epsilon = \{y_j\}_{j=1}^\infty, S = \{s_j\}_{j=1}^\infty, S_0 = \{s_{j0}\}_{j=1}^\infty$ with $s_{j0} = \int_0^1 S(t,\theta_0)\phi_j(t)\mathrm{d}t$.

The Bayes *a posteriori* likelihood ratio equals

$$\begin{aligned}
&\int \exp\left\{-\frac{1}{2\epsilon^2}(Y_\epsilon - S)'Q^{-2}(Y_\epsilon - S) - \frac{1}{2}\tau^{-2}(S-S_0)'Q^{-1}K_{h_\epsilon}^{-1}R^{-1}K_{h_\epsilon}^{-1}Q^{-1}(S-S_0)\right.\\
&\qquad\left. + \frac{1}{2\epsilon^2}(Y_\epsilon - S_0)'Q^{-2}(Y_\epsilon - S_0)\right\}\mathrm{d}\mu_{\epsilon\delta} = \int \exp\left\{\epsilon^{-2}(Y_\epsilon - S_0)'Q^{-2}(S-S_0)\right.\\
&\qquad\left. - \frac{1}{2}(S-S_0)'\left(\epsilon^{-2}Q^{-2} + \tau^{-2}Q^{-1}K_{h_\epsilon}^{-1}R^{-1}K_{h_\epsilon}^{-1}Q^{-1}\right)(S-S_0)\right\}\mathrm{d}\mu_{\epsilon\delta}\\
&\qquad = C\int \exp\left\{-\frac{1}{2}||(Y_\epsilon - S_0)'Q^{-1}\left(\epsilon^{-2}Q^{-2} + \tau^{-2}Q^{-1}K_{h_\epsilon}^{-1}R^{-1}K_{h_\epsilon}^{-1}Q^{-1}\right)^{-1/2}\right.\\
&\qquad\quad\left. - \left(\epsilon^{-2}Q^{-2} + \tau^{-2}Q^{-1}K_{h_\epsilon}^{-1}R^{-1}K_{h_\epsilon}^{-1}Q^{-1}\right)^{1/2}(S-S_0)||^2\right\}\mathrm{d}\mu_{\epsilon\delta}\\
&\qquad \times \exp\left\{-\frac{1}{2}(Y_\epsilon - S_0)'Q^{-1}\left(\epsilon^{-2}Q^{-2} + \tau^{-2}Q^{-1}K_{h_\epsilon}^{-1}R^{-1}K_{h_\epsilon}^{-1}Q^{-1}\right)^{-1}Q^{-1}(Y_\epsilon - S_0)\right\}\\
&\qquad = C\exp\left\{-\frac{1}{2}(Y_\epsilon - S_0)'Q^{-1}\left(\epsilon^{-2}Q^{-2} + \tau^{-2}Q^{-1}K_{h_\epsilon}^{-1}R^{-1}K_{h_\epsilon}^{-1}Q^{-1}\right)^{-1}Q^{-1}(Y_\epsilon - S_0)\right\}. \qquad (4.29)
\end{aligned}$$

Thus the Bayes test statistics can be defined as follows

$$\begin{aligned}
D_{\epsilon\delta} &= (Y_\epsilon - S_0)'Q^{-1}\left(\epsilon^{-2}Q^{-2} + \tau^{-2}Q^{-1}K_{h_\epsilon}^{-1}R^{-1}K_{h_\epsilon}^{-1}Q^{-1}\right)^{-1}Q^{-1}(Y_\epsilon - S_0)\\
&= (Y_\epsilon - S_0)'K_{h_\epsilon}R^{1/2}\left(\epsilon^{-2}\tau^2K_{h_\epsilon}RK_{h_\epsilon} + E\right)^{-1}R^{1/2}K_{h_\epsilon}(Y_\epsilon - S_0).
\end{aligned}$$

Denote

$$D_{1\epsilon\delta} = D_{1\epsilon\delta}(Y_\epsilon - S_0) = \epsilon^{-2}\tau^2(Y_\epsilon - S_0)'(K_{h_\epsilon}RK_{h_\epsilon})^2(Y_\epsilon - S_0).$$

We have

$$\begin{aligned}
T_\epsilon - D_{\epsilon\delta} - D_{1\epsilon\delta} &= \epsilon^{-4}\tau^4(Y_\epsilon - S_0)'(K_{h_\epsilon}R^{1/2})^3(\epsilon^{-2}\tau^2K_{h_\epsilon}RK_{h_\epsilon} + E)^{-1}(R^{1/2}K_{h_\epsilon})^3(Y_\epsilon - S_0))\\
&< \epsilon^{-4}\tau^4(Y_\epsilon - S_0)'(K_{h_\epsilon}RK_{h_\epsilon})^3(Y_\epsilon - S_0) \doteq D_{2\epsilon\delta}. \qquad (4.30)
\end{aligned}$$

We have

$$D_{1\epsilon\delta}(Y_\epsilon - S_0) \le 2D_{1\epsilon\delta}(Y_\epsilon - S) + 2D_{1\epsilon\delta}(S - S_0), \qquad (4.31)$$

$$D_{2\epsilon\delta}(Y_\epsilon - S_0) \le 2D_{2\epsilon\delta}(Y_\epsilon - S) + 2D_{2\epsilon\delta}(S - S_0). \qquad (4.32)$$

The unique difference of statistics $T_\epsilon = (Y_\epsilon - S)'K_{h_\epsilon}RK_{h_\epsilon}(Y_\epsilon - S)$ and $D_{1\epsilon\delta}(Y_\epsilon - S), D_{2\epsilon\delta}(Y_\epsilon - S)$ are the powers of the kernels. Hence, estimating similarly to (4.21–4.23), we get

$$d_{1\epsilon}(h_\epsilon) \doteq E_S[D_{1\epsilon\delta}(Y_\epsilon - S)] = \tau^2\int_0^1 q^2(t)\,\mathrm{d}t\int_0^1 \mathrm{d}t_1K_{2,rh_\epsilon}^2(t,t_1) < C\frac{\tau^2}{h_\epsilon} < C\epsilon^2h_\epsilon^{-1/2}, \qquad (4.33)$$

$$E_S[D_{2\epsilon\delta}(Y_\epsilon - S)] = \epsilon^{-2}\tau^4 \int_0^1 q^2(t)\,\mathrm{d}t \int_0^1 r(s)\mathrm{d}s \left(\int_0^1 \mathrm{d}t_1 K_{h_\epsilon}(t_1 - s) \times K_{2,rh_\epsilon}(t,t_1)\right)^2 < C\frac{\epsilon^{-2}\tau^4}{h_\epsilon} < C\epsilon^2, \quad (4.34)$$

$$\mathrm{Var}_S(D_{1\epsilon\delta}(Y_\epsilon - S)) < C\frac{\tau^4}{h_\epsilon} < C\epsilon^4, \quad \mathrm{Var}_S(D_{2\epsilon\delta}(Y_\epsilon - S)) < C\frac{\epsilon^{-4}\tau^8}{h_\epsilon} < C\epsilon^4 h_\epsilon^3. \quad (4.35)$$

By straightforward calculations, using (4.7), we get

$$D_{1\epsilon\delta}(S - S_0) < C\rho_\epsilon(h_\epsilon), \quad D_{2\epsilon\delta}(S - S_0) < C\rho_\epsilon(h_\epsilon) \quad (4.36)$$

if $\rho_\epsilon(h_\epsilon) < T_\epsilon(S,\theta_0) < C\rho_\epsilon(h_\epsilon)$.

By (4.30–4.36) we get

$$P_S\left(\epsilon^{-2}h_\epsilon^{1/2}\sigma^{-1}(T_\epsilon(Y_\epsilon,\theta_0) - d_\epsilon(h_\epsilon)) < x_\alpha\right) = P_S\left(\epsilon^{-2}h_\epsilon^{1/2}\sigma^{-1}(D_{\epsilon\delta}(Y_\epsilon,\theta_0) - d_\epsilon(h_\epsilon)) - d_{1\epsilon}(h_\epsilon)\right) < x_\alpha(1 + o(1))) \quad (4.37)$$

uniformly in $S:\ \rho_\epsilon(h_\epsilon) < T_\epsilon(S,\theta_0) < C\rho_\epsilon(h_\epsilon)$ as $\delta \to 0, \epsilon \to 0$.

By (4.25, 4.29, 4.37) we get (4.28). This completes the proof of Theorem 3.1 in the case of sets of alternatives $\Im_{\epsilon,h_\epsilon}(\Theta,\rho_\epsilon)$.

The Theorem 2.1.2 and Theorem 3.1 in the case of sets of alternatives $\Im_{\epsilon,h_\epsilon}(\Theta,\rho_\epsilon,\beta,P_0)$ follows from Lemma 4.3.

Lemma 4.3. *Let A and (2.4) hold additionally. Then*

$$(1+\delta)^{-1}\int_0^1 \left(\tilde{S}^{(\beta)}(t) - S^{(\beta)}(t,\theta_0)\right)^2 \mathrm{d}t \to P_0 \quad (4.38)$$

and

$$(1+\delta)^{-1}\int_0^1 \left(\tilde{S}^{(\beta)}(t) - S(s,\bar{\theta}_\epsilon)\right)^2 \mathrm{d}t \to P_0 \quad (4.39)$$

in probability as $\epsilon \to 0$.

Proof of Lemma 4.1. We have

$$\epsilon^{-2}h_\epsilon^{1/2}I_{6\epsilon} = 2J_{1\Delta\epsilon} + J_{2\Delta\epsilon} \quad (4.40)$$

where

$$J_{1\Delta\epsilon} = h_\epsilon^{1/2}\int_0^1 q(y_1)\mathrm{d}w(y_1)\int_{y_1-2h_\epsilon}^{y_1-\Delta} K_{2,rh_\epsilon}(y_1,y_2)q(y_2)\mathrm{d}w(y_2),$$

$$J_{2\Delta\epsilon} = h_\epsilon^{1/2}\int_0^1 q(y_1)\mathrm{d}w(y_1)\int_{y_1-\Delta}^{y_1+\Delta} K_{2,rh_\epsilon}(y_1,y_2)q(y_2)\mathrm{d}w(y_2)$$

and $\Delta = \Delta_\epsilon \to 0$, $\Delta_\epsilon/h_\epsilon \to 0$ as $\epsilon \to 0$.

By straightforward calculations, arguing similarly to (4.10, 4.62) we get

$$\mathrm{Var}(J_{2\Delta\epsilon}) = 2h_\epsilon\int_{-1}^1 q^2(y_1)\mathrm{d}y_1\int_{y_1-\Delta}^{y_1+\Delta} K^2_{2,rh_\epsilon}(y_1,y_2)q^2(y_2)\mathrm{d}y_2 < C\Delta. \quad (4.41)$$

Thus it suffices to study the limit behaviour of $J_{1\Delta\epsilon}$. One can write

$$J_{1\Delta\epsilon} = \sum_{j=1}^{C_\Delta} Z_{j\epsilon}$$

where $C_\Delta = [1/\Delta]$ and

$$Z_{j\epsilon} = h_\epsilon^{1/2} \int_{(j-1)\Delta}^{j\Delta} q(y_1)\mathrm{d}w(y_1) \int_{y_1-2h_\epsilon}^{y_1-\Delta} K_{2,rh_\epsilon}(y_1,y_2)q(y_2)\mathrm{d}w(y_2).$$

We can consider $J_{1\Delta\epsilon}$ as a sum of martingale differences $Z_{j\epsilon}$ and to apply corresponding Central Limit Theorem (see Brown [3]) to prove asymptotic normality. Thus it suffices to show that

$$\lim_{\epsilon\to 0} \sum_{j=2}^{C_\Delta} E\left\{Z_{j\epsilon}^2 \chi(|Z_{j\epsilon}| > C)\right\} = 0, \tag{4.42}$$

$$\lim_{\epsilon\to 0} \frac{2}{\sigma^2} \sum_{j=2}^{C_\Delta} E\left\{Z_{j\epsilon}^2 | F_{j-1,\epsilon}\right\} = 1 \tag{4.43}$$

where $F_{j-1,\epsilon}$ is the σ-field generated by the Wiener process $w(t)$, $0 \le t \le (j-1)\Delta_\epsilon$.

We have

$$\begin{aligned}\sum_{j=1}^{C_\Delta} E(Z_{j\epsilon}^4) &= 3h_\epsilon^2 \sum_{j=1}^{C_\Delta} \left(\int_{(j-1)\Delta}^{j\Delta} q^2(y_1)\mathrm{d}y_1 \int_{y_1-2h_\epsilon}^{y_1-\Delta} K_{2,rh_\epsilon}^2(y_1,y_2)q^2(y_2)\mathrm{d}y_2 \right)^2 \\ &= 3h_\epsilon^2 \Delta \int_0^1 q^4(y_1)\mathrm{d}y_1 \left(\int_{y_1-2h_\epsilon}^{y_1-\Delta} K_{2,rh_\epsilon}^2(y_1,y_2)q^2(y_2)\mathrm{d}y_2 \right)^2 (1+o(1)) \\ &= 3\Delta h_\epsilon^{-1} \int_0^1 q^8(y)r^4(y)\mathrm{d}y \left(\int_{-2}^0 K_2^2(u)\mathrm{d}u \right)^2 (1+o(1)). \end{aligned} \tag{4.44}$$

By Chebyshov inequality, equation (4.44) implies (4.42).

Denote

$$V_{j\epsilon} = E\left(Z_{j\epsilon}^2 | F_{j-1,\epsilon}\right) = h_\epsilon \int_{(j-1)\Delta}^{j\Delta} q^2(y_1)\mathrm{d}y_1 \left(\int_{y_1-2h_\epsilon}^{y_1-\Delta} K_{2,rh_\epsilon}(y_1,y_2)q(y_2)\mathrm{d}w(y_2) \right)^2.$$

Estimating similarly to (4.10), we get

$$\begin{aligned}\mathrm{Var}\left(\sum_{j=1}^{C_\Delta} V_{j\epsilon}\right) &< Ch_\epsilon^2 \int_0^1 q^2(y_1)\mathrm{d}y_1 \int_0^1 q^2(y_2)\mathrm{d}y_2 \left(\int_{z_1-2h_\epsilon}^{z_2-\Delta} K_{2,rh_\epsilon}(y_1,y_3)q^2(y_3)K_{2,rh_\epsilon}(y_2,y_3)\mathrm{d}y_3 \right)^2 \\ &< Ch_\epsilon^2 \int_{-4}^4 \mathrm{d}x_1 \int_{-4}^4 \mathrm{d}x_2 K_{4,h_\epsilon}^2(x_1,x_2) < Ch_\epsilon \end{aligned} \tag{4.45}$$

where $z_1 = z_1(y_1,y_2) = \max\{y_1,y_2\}, z_2 = z_2(y_1,y_2) = \min\{y_1,y_2\}$.

By Chebyshov inequality, equation (4.45) implies (4.43). This completes the proof of Lemma 4.1.

Proof of Lemma 4.2. Denote $\zeta(t) = \tilde{S}(t) - S(t,\theta_0)$ and $\bar{\theta}_\epsilon = \mathrm{argmin}_{\theta\in\Theta} M_\epsilon(\theta)$ where

$$M_\epsilon(\theta) = \int_0^1 r(t)\mathrm{d}t \left(\int_0^1 K_{h_\epsilon}(t-s)\left(\tilde{S}(s) - S(s,\theta)\right)\mathrm{d}s \right)^2.$$

We have

$$M_\epsilon(\bar{\theta}_\epsilon) = M_{1\epsilon} + 2M_{2\epsilon} + M_{3\epsilon} \tag{4.46}$$

with

$$M_{1\epsilon} = \int_0^1 r(t)\mathrm{d}t \left(\int_0^1 K_{h_\epsilon}(t-s)\zeta(s)\mathrm{d}s \right)^2,$$

$$M_{2\epsilon} = \int_0^1 r(t)\mathrm{d}t \int_0^1 K_{h_\epsilon}(t-s) \left(S(s,\theta_0) - S(s,\bar{\theta}_\epsilon) \right) \mathrm{d}s \int_0^1 K_{h_\epsilon}(t-s)\zeta(s)\mathrm{d}s,$$

$$M_{3\epsilon} = \int_0^1 r(t)\mathrm{d}t \left(\int_0^1 K_{h_\epsilon}(t-s) \left(S(s,\theta_0) - S(s,\bar{\theta}_\epsilon) \right) \mathrm{d}s \right)^2.$$

At first we shall prove (4.26), assuming that (4.25) holds, that is

$$((1+\delta)\rho_\epsilon(h_\epsilon))^{-1} M_{1\epsilon} \to 1 \tag{4.47}$$

in probability as $\epsilon \to 0$.

After that the proof of (4.25) will be given.

We have

$$M_{2\epsilon} \le M_{1\epsilon}^{1/2} M_{3\epsilon}^{1/2}. \tag{4.48}$$

Using the definition of $\bar{\theta}_\epsilon$ and (4.47, 4.48) together we get

$$M_{3\epsilon} \to 0 \tag{4.49}$$

in probability as $\epsilon \to 0$.

We have

$$M_{3\epsilon}^{1/2} \ge B_{1\epsilon}^{1/2} - B_{2\epsilon}^{1/2} - B_{3\epsilon}^{1/2} \tag{4.50}$$

with

$$B_{1\epsilon} = \int_0^1 (S(t,\theta_0) - S(t,\bar{\theta}_\epsilon))^2 r(t)\mathrm{d}t,$$

$$B_{2\epsilon} = \int_0^1 r(t)\mathrm{d}t \left(\int_0^1 K_{h_\epsilon}(t-s)(S(s,\theta_0) - S(t,\theta_0))\mathrm{d}s \right)^2,$$

$$B_{3\epsilon} = \int_0^1 r(t)\mathrm{d}t \left(\int_0^1 K_{h_\epsilon}(t-s)(S(s,\bar{\theta}_\epsilon) - S(t,\bar{\theta}_\epsilon))\mathrm{d}s \right)^2.$$

By D6,

$$B_{2\epsilon} < C\omega(h_\epsilon), \qquad B_{3\epsilon} < C\omega(h_\epsilon). \tag{4.51}$$

By (4.49–4.51), we get

$$B_{1\epsilon} \to 0 \tag{4.52}$$

in probability as $\epsilon \to 0$.

By D1–D3, equation (4.52) implies

$$\bar{\theta}_\epsilon \to \theta_0 \tag{4.53}$$

in probability as $\epsilon \to 0$.

Denote

$$V(s) = S(s,\theta_0) - S(s,\bar{\theta}_\epsilon) - (\theta_0 - \bar{\theta}_\epsilon) S_\theta(s,\theta_0).$$

We have

$$M_{3\epsilon} = M_{31\epsilon} + 2M_{32\epsilon} + M_{33\epsilon} \tag{4.54}$$

with

$$M_{31\epsilon} = (\bar{\theta}_\epsilon - \theta_0)^2 \int_0^1 r(t)\mathrm{d}t \left(\int_0^1 K_{h_\epsilon}(t-s) S_\theta(s,\theta_0)\mathrm{d}s \right)^2,$$
$$M_{32\epsilon} = (\bar{\theta}_\epsilon - \theta_0) \int_0^1 r(t)\mathrm{d}t \int_0^1 K_{h_\epsilon}(t-s) S_\theta(s,\theta_0)\mathrm{d}s \int_0^1 K_{h_\epsilon}(t-s) V(s)\mathrm{d}s,$$
$$M_{33\epsilon} = \int_0^1 r(t)\mathrm{d}t \left(\int_0^1 K_{h_\epsilon}(t-s) V(s)\mathrm{d}s \right)^2.$$

By (4.7), D2, we get

$$M_{33\epsilon} \le C \int_0^1 r(t)\mathrm{d}t \int_0^1 V^2(s)\mathrm{d}s \le C|\bar{\theta}_\epsilon - \theta_0|^{2+\omega}. \tag{4.55}$$

By Schwartz inequality, we get

$$M_{32\epsilon} < M_{31\epsilon}^{1/2} M_{33\epsilon}^{1/2}. \tag{4.56}$$

By (4.7), we get

$$M_{31\epsilon} \le C \left(\bar{\theta}_\epsilon - \theta_0\right)^2 \int_0^1 r(t) S_\theta^2(t,\theta_0)\mathrm{d}t \le C \left(\bar{\theta}_\epsilon - \theta_0\right)^2. \tag{4.57}$$

By (4.54–4.57), we get

$$M_{3\epsilon} \le C|\bar{\theta}_\epsilon - \theta_0|^2(1 + o_P(1)). \tag{4.58}$$

By (4.47, 4.48, 4.58) together, we get

$$M_{2\epsilon} = O_P\left(|\bar{\theta}_\epsilon - \theta_0|\rho_\epsilon^{1/2}(h_\epsilon)\right). \tag{4.59}$$

By (4.47), $M_{1\epsilon}$ can be represented as $M_{1\epsilon} = \rho_\epsilon(h_\epsilon)(1+\delta) + o_P(\rho_\epsilon(h_\epsilon))$ as $\epsilon \to 0$. Similarly, by (4.58, 4.59), we get $M_{2\epsilon} + M_{3\epsilon} = O_P(|\bar{\theta}_\epsilon - \theta_0|\rho_\epsilon^{1/2}(h_\epsilon) + |\bar{\theta}_\epsilon - \theta_0|^2)$. Hence, by definition of $M_\epsilon(\bar{\theta}_\epsilon) = \min\{M_\epsilon(\theta) : \theta \in \Theta\}$ and (4.46, 4.53) we get

$$|\bar{\theta}_\epsilon - \theta_0| = o_P\left(\rho_\epsilon^{1/2}(h_\epsilon)\right). \tag{4.60}$$

By (4.46, 4.47, 4.58–4.60) together, we get (4.26).

It remains to prove (4.25). Denote

$$\tilde{K}_{2,rh_\epsilon}(t_1,t_2) = \int_0^1 K_{h_\epsilon}(t_1 - s) r^{1/2}(s) K_{h_\epsilon}(t_2 - s)\mathrm{d}s.$$

By straightforward calculations, using the same technique as in (4.10), we get

$$E\left\{\int_0^1 \left(\int_0^1 K_{h_\epsilon}(t-s)(\tilde{S}(s)-S(s,\theta_0))\mathrm{d}s\right)^2 r(t)\mathrm{d}t\right\} =$$

$$\tau^2 E\left\{\int_0^1 \left(\int_0^1 K_{h_\epsilon}(t-s_1)r^{1/2}(s_1)\mathrm{d}s_1 \int_0^1 K_{h_\epsilon}(s_1-s_2)q(s_2)\mathrm{d}w_1(s_2)\right)^2 r(t)\mathrm{d}t\right\}$$

$$= \tau^2 \int_0^1 r(t)\mathrm{d}t \int_0^1 K_{h_\epsilon}(t-s_1)\mathrm{d}s_1 \int_0^1 K_{h_\epsilon}(t-s_3)\mathrm{d}s_3 r^{1/2}(s_1)r^{1/2}(s_3)$$

$$\times E\left(\int_0^1 K_{h_\epsilon}(s_1-s_2)q(s_2)\mathrm{d}w_1(s_2)\int_0^1 K_{h_\epsilon}(s_3-s_4)q(s_4)\mathrm{d}w_1(s_4)\right)$$

$$= \tau^2 \int_0^1 r(t)\mathrm{d}t \int_0^1 K_{h_\epsilon}(t-s_1)\mathrm{d}s_1 \int_0^1 K_{h_\epsilon}(t-s_3)\mathrm{d}s_3 r^{1/2}(s_1)r^{1/2}(s_3)$$

$$\times \int_0^1 K_{h_\epsilon}(s_1-s_2)q^2(s_2)K_{h_\epsilon}(s_2-s_3)\mathrm{d}s_2$$

$$= \tau^2 \int_0^1 r(t)\mathrm{d}t \int_0^1 q^2(s)\mathrm{d}s\tilde{K}^2_{2,rh_\epsilon}(t,s) = \tau^2\sigma^2/(2h_\epsilon)(1+o(1)). \quad (4.61)$$

Arguing similarly to (4.61), we get

$$\mathrm{Var}\left(\int_0^1 \left(\int_0^1 K_{h_\epsilon}(t-s)\left(\tilde{S}(s)-S(s,\theta_0)\right)\mathrm{d}s\right)^2 r(t)\mathrm{d}t\right)$$

$$= 2\tau^4 \int_0^1 r(t_1)\mathrm{d}t_1 \int_0^1 r(t_2)\mathrm{d}t_2 \left(\int_0^1 K_{h_\epsilon}(t_1-s_1)\mathrm{d}s_1 \int_0^1 K_{h_\epsilon}(t_2-s_3)\mathrm{d}s_3\right.$$

$$\left.\times\, r^{1/2}(s_1)r^{1/2}(s_3)\int_0^1 K_{h_\epsilon}(s_1-s_2)q^2(s_2)K_{h_\epsilon}(s_3-s_2)\mathrm{d}s_2\right)^2 (1+o(1))$$

$$= 2\tau^4 \int_0^1 r(t_1)\mathrm{d}t_1 \int_0^1 r(t_2)\mathrm{d}t_2 \left(\int \tilde{K}_{2,rh_\epsilon}(t_1,s_2)q^2(s_2)\tilde{K}_{2,rh_\epsilon}(t_2,s_2)\mathrm{d}s_2\right)^2 (1+o(1))$$

$$= 2\tau^4 h_\epsilon^{-1}\int_0^1 q^4(s)r^4(s)\mathrm{d}s \int_{-4}^4 K_4^2(v)\mathrm{d}v(1+o(1)). \quad (4.62)$$

By Chebyshov inequality, equations (4.61, 4.62) together imply (4.25).

Proof of Lemma 4.3. To simplify the reasoning we assume $r(t)=1$ for all $t\in[0,1]$. This does not cause any principal differences in the arguments.

We have

$$\tilde{\zeta}_\beta \doteq \tilde{S}^{(\beta)}(t) - S^{(\beta)}(t,\theta_0) = \tau h^{-\beta-1}\int_0^1 K^{(\beta)}\left(\frac{t-s}{h}\right) q(s)\mathrm{d}w_1(s). \quad (4.63)$$

Repeating similar estimates as in the proof of (4.25) we get (4.38).

We have

$$\int_0^1 \left(\tilde{S}^{(\beta)}(t) - S^{(\beta)}\left(t,\bar{\theta}_\epsilon\right)\right)^2 \mathrm{d}t \doteq D_{1\epsilon}+D_{2\epsilon}+D_{3\epsilon} \quad (4.64)$$

where

$$D_{1\epsilon} = \int_0^1 \left(\tilde{S}^{(\beta)}(t) - S^{(\beta)}(t,\theta_0)\right)^2 \mathrm{d}t,$$
$$D_{2\epsilon} = \int_0^1 \left(\tilde{S}^{(\beta)}(t) - S^{(\beta)}(t,\theta_0)\right)^2 \mathrm{d}t \left(S^{(\beta)}(t,\theta_0) - S^{(\beta)}\left(t,\bar{\theta}_\epsilon\right)\right)^2 \mathrm{d}t,$$
$$D_{3\epsilon} = \int_0^1 \left(S^{(\beta)}(t,\bar{\theta}_\epsilon) - S^{(\beta)}(t,\theta_0)\right)^2 \mathrm{d}t.$$

We have

$$|D_{2\epsilon}| < D_{1\epsilon}^{1/2} D_{3\epsilon}^{1/2}. \tag{4.65}$$

By B2, equation (4.60), we get

$$D_{3\epsilon} \to 0 \tag{4.66}$$

in probability as $\epsilon \to 0$.

By (4.38, 4.64–4.66) together, we get (4.39). This completes the proof of Lemma 4.3.

5. Proofs of Theorems 2.2, 2.3 and 3.2

The further arguments will be given in the notation of Theorems 2.2 and 2.3. In the case of Theorem 3.2 a modification of notation is unessential.

The statements on asymptotic equivalence of statistical experiments (see Brown and Low [4] and Nussbaum [27]) can be applied to the proof of lower bounds if the realizations of random processes generated by the Bayes *a priori* measures belong to the Hoelder space

$$\Sigma(\beta, M) = \{S : |S(t) - S(s)| < M|t-s|^\beta, \quad t,s \in (0,1)\}$$

with $M > 0, \beta > \frac{1}{2}$.

In the problem of hypothesis testing on density we need also to suppose $f(t) > c > 0$ for all $t \in [0,1]$ (see Nussbaum [27]).

Denote

$$\bar{\Im}_{nh_n}(\beta, M) = \{S : S \in \Im_{nh_n}, S \in \Sigma(\beta, M), |S(t) - S_0(t)| < c_n, t \in [0,1]\}$$

where $c_n \to 0$ as $n \to \infty$.

The Bayes *a priori* probability measures $\nu_{ln\delta}$ of S and f in the problems of hypothesis testing on regression and density respectively are defined as the conditional probability measures of $\tilde{S} = \tilde{S}_n$ under the condition $\tilde{S}_n \in \bar{\Im}_{nh_n}(\beta, M_l)$ with $M_l \to \infty$ as $l \to \infty$.

For the proof of lower bounds it suffices to show that there exists $M_l \to \infty$ as $l \to \infty$ such that

$$P_{\mu_{n\delta}}(\tilde{S}_n \in \Im_{nh_n}(\beta, M_l)) = 1 + o(1) \tag{5.1}$$

as $l \to \infty, \delta \to 0, n \to \infty$.

Thus we need to prove that there exists $c_n \to 0$ as $n \to \infty$ such that

$$\lim_{n\to\infty} P(\sup\{|\tilde{S}_n(t) - S_0(t)| : t \in [0,1]\} < c_n) = 1 \tag{5.2}$$

and there exists $\omega_l \to 0, M_l \to \infty$ as $l \to \infty$ such that

$$\liminf_{n\to\infty} P\left(\tilde{S}_n \in \Sigma(\beta, M_l)\right) > 1 - \omega_l. \tag{5.3}$$

We begin with the proof of (5.2). By A, we get

$$E|\tilde{S}_n(t)-\tilde{S}_n(s)|^2 < C\tau^2\int_0^1 |K_{h_n}(t-u)-K_{h_n}(s-u)|^2 q^2(u)\mathrm{d}u$$
$$\le C\tau^2 h_n^{-2-\gamma}\min\left\{|t-s|^{1+\gamma}, h_n^{1+\gamma}\right\} < Cn^{-1}h_n^{-3/2-\gamma}\min\left(|t-s|^{1+\gamma}, h_n^{1+\gamma}\right). \quad (5.4)$$

By straightforward calculations, we get

$$E\left(\tilde{S}_n^2(t)\right) < C\tau^2 h_n^{-1} < Cn^{-1}h_n^{-1/2}. \quad (5.5)$$

By Theorem 7.1 in Piterbarg [28] and Slepian comparison principle (see Slepian [30]) we get (5.2) with $c_n = ch_n^{\gamma/2}$.

It follows from Theorem 1 (Sect. 15 in Lifshits [26]) that for any sequence $\phi_l > 0, \phi_l \to 0$ as $l \to \infty$ there exist sequences $M_{ln} \to \infty$ as $l \to \infty$ such that

$$P\left(\tilde{S}_n \in \Sigma(\beta, M_{ln})\right) > 1-\phi_l \quad (5.6)$$

with $\frac{1}{2} < \beta < \frac{1+\gamma}{2}$.

The proof of Theorem 1 Sect. 15 in Lifshits [26] is based on Borel–Cantelli lemma. In order to show that one can choose the values $M_{ln} = M_l$ which does not depend on n it suffices to make use of the following version of Borel–Cantelli lemma in Lifshits [26] arguments.

Lemma 5.1. *Let $A_{1n}, A_{2n}, \dots$ be sequences of events. Let there exist a sequence $\kappa_m \to 0$ as $m \to \infty$ such that for each n*

$$\sum_{i=m}^{\infty} P(A_{in}) < \kappa_m. \quad (5.7)$$

Denote $B_{mn} = \cup_{i=m}^{\infty} A_{in}$. Then $P(B_{mn}) \to 0$ as $m \to \infty$ uniformly in n.

Applying Lemma 5.1 in the reasoning the proof of Theorem 1 (Sect. 15 in Lifshits [26]) we get the version of this theorem with $M_l = M_{ln}$ which does not depend on n. Therefore (5.3) holds. By (5.2, 5.3), we can apply to the realizations of random processes generated by corresponding Bayes *a priori* measures the arguments of the proof of Theorem 2.1 and get the lower bounds in Theorems 2.2, 2.3 and 3.2 as corollaries of Theorem 4.1 in Brown and Low [4] and Theorems 2.1, 2.7 in Nussbaum [27] respectively. This completes the proof of lower bounds in Theorems 2.2, 2.3 and 3.2.

Proof of Theorem 2.2. Upper bound. The estimates are akin to (4.1–4.24).

Denote

$$g_h(t) = \frac{1}{n}\sum_{i=1}^{n} K_h(t-x_{ni})S(x_{ni})$$

and

$$g_{0h}(t) = \frac{1}{n}\sum_{i=1}^{n} K_h(t-x_{ni})S_0(x_{ni}).$$

We write

$$T_n(Y) = I_{1n} + I_{2n} + I_{3n} \quad (5.8)$$

with

$$I_{1n} = \int_0^1 (g_{h_n}(t) - g_{0h_n}(t))^2 r(t)\mathrm{d}t,$$

$$I_{2n} = \int_0^1 \left(\frac{1}{n}\sum_{i=1}^n K_{h_n}(t - x_{ni})(Y_{ni} - S(x_{ni}))\right)(g_{h_n}(t) - g_{0h_n}(t))r(t)\mathrm{d}t,$$

$$I_{3n} = \int_0^1 \left(\frac{1}{n}\sum_{i=1}^n K_{h_n}(t - x_{ni})(Y_{ni} - S(x_{ni}))\right)^2 r(t)\mathrm{d}t.$$

Observe that I_{3n} does not depend on S.

We write

$$I_{3n} = I_{31n} + I_{32n} \tag{5.9}$$

where

$$I_{31n} = \frac{1}{n^2}\sum_{i=1}^n \int_0^1 (K_{h_n}(t - x_{ni})(Y_{ni} - S(x_{ni})))^2 r(t)\mathrm{d}t,$$

$$I_{32n} = \frac{2}{n^2}\sum_{1\le i<j\le n} (Y_{ni} - S(x_{ni}))(Y_{nj} - S(x_{nj}))\int_0^1 K_{h_n}(t - x_{ni})K_{h_n}(t - x_{nj})r(t)\mathrm{d}t.$$

Denote $t_{ni} = \frac{2i-1}{2n+1}$ for all $1 \le i \le n$.

We have

$$E(I_{31n}) = \frac{1}{n^2}\sum_{i=1}^n \lambda^2(x_{ni})\int_0^1 K^2_{h_n}(t - x_{ni})r(t)\mathrm{d}t. \tag{5.10}$$

We have

$$|E(I_{31n}) - d_n(h_n)| < R_{n1} + R_{n2} \tag{5.11}$$

where, by (2.4),

$$\begin{aligned}
R_{n1} &= \frac{1}{n}\sum_{i=1}^n \lambda^2(x_{ni})\int_0^1 r(t)\mathrm{d}t \int_{t_{ni}}^{t_{n,i+1}} |K^2_{h_n}(t - x_{ni}) - K^2_{h_n}\left(t - H^{-1}(s)\right)|\mathrm{d}s \\
&\le \frac{C}{nh_n}\sum_{i=1}^n \lambda^2(x_{ni})\int_{t_{ni}}^{t_{n,i+1}} \mathrm{d}s \int_0^1 r(t)\mathrm{d}t|K_{h_n}(t - x_{ni}) - K_{h_n}\left(t - H^{-1}(s)\right)| \\
&\le \frac{C}{nh_n}\sum_{i=1}^n \int_{t_{ni}}^{t_{n,i+1}} \left(\int_0^1 \left(K_{h_n}(t - x_{ni}) - K_{h_n}\left(t - H^{-1}(s)\right)\right)^2 \mathrm{d}t\right)^{1/2} \mathrm{d}s \\
&\le \frac{C}{nh_n}\sum_{i=1}^n \int_{t_{ni}}^{t_{n,i+1}} h_n^{(1+\gamma)/2}\mathrm{d}s \le \frac{C}{nh_n^{1/2-\gamma/2}} = o\left(\frac{1}{nh_n^{1/2}}\right)
\end{aligned} \tag{5.12}$$

and

$$R_{2n} = \frac{1}{n}\sum_{i=1}^n \int_{t_{ni}}^{t_{n,i+1}} |\lambda^2(x_{ni}) - \lambda^2(H^{-1}(s))|\mathrm{d}s \int_0^1 K^2_{h_n}\left(t - H^{-1}(s)\right)r(t)\mathrm{d}t \le \frac{C}{n^2h_n} = o\left(n^{-1}h_n^{-1/2}\right). \tag{5.13}$$

By (5.11–5.13), we get

$$|E(I_{31n}) - d_n(h_n)| = o\left(n^{-1}h_n^{-1/2}\right). \tag{5.14}$$

Using a similar technique as in the estimation of addendum denoted I_{1n} in Hall [16], we get

$$\operatorname{Var}(I_{31n}) = O(n^{-1}). \tag{5.15}$$

Using the same reasoning and estimates as in analysis of I_{2n} in Hall [16], we get the following lemma:

Lemma 5.2. *The distributions of I_{32n} are asymptotically normal and*

$$E(I_{32n}) = 0, \quad \operatorname{Var}(I_{32n}) = n^{-2}h_n^{-1}\sigma^2(1+o(1)). \tag{5.16}$$

We have

$$E(I_{2n}) = 0. \tag{5.17}$$

Arguing similarly to (4.10), we get

$$\begin{aligned}\operatorname{Var}(I_{2n}) &= \frac{1}{n^2}\sum_{i=1}^n \lambda^2(x_{ni})\int_0^1 r(t)\mathrm{d}t\int_0^1 r(s)\mathrm{d}sK_{h_n}(t-x_{ni})K_{h_n}(s-x_{ni})(g_{h_n}(t)-g_{0h_n}(t))(g_{h_n}(s)-g_{0h_n}(s))\\ &\le \frac{1}{n^2}I_{1n}^{1/2}\left(\int_0^1 r(t)\mathrm{d}t\left(\sum_{i=1}^n \lambda^2(x_{ni})\int_0^1 r(s)K_{h_n}(s-x_{ni})\mathrm{d}sK_{h_n}(t-x_{ni})(g_{h_n}(t)-g_{0h_n}(t))\right)^2\right)^{1/2}.\end{aligned} \tag{5.18}$$

Since

$$D_n(t,s) \doteq \sum_{i=1}^n \lambda^2(x_{ni})r(t)K_{h_n}(t-x_{ni})r(s)K_{h_n}(s-x_{ni}) < \frac{Cn}{h_n}$$

if $|t-s| < 2h_n$ and $D_n(t,s) = 0$ if $|t-s| > 2h_n$, we get

$$\begin{aligned}\operatorname{Var}(I_{2n}) &< Cn^{-1}I_{1n}^{1/2}\left(\int_0^1 \mathrm{d}t h_n^{-2}\left(\int_{t-2h_n}^{t+2h_n}|g_{h_n}(s)-g_{0h_n}(s)|\mathrm{d}s\right)^2\right)^{1/2}\\ &\le Cn^{-1}I_{1n}^{1/2}\left(\int_0^1 \mathrm{d}t h_n^{-1}\int_{t-2h_n}^{t+2h_n}(g_{h_n}(s)-g_{0h_n}(s))^2\mathrm{d}s\right)^{1/2} \le \frac{C}{n}I_{1n}.\end{aligned} \tag{5.19}$$

Now the upper bound in Theorem 2.2 follows from Lemma 5.2 and (5.8, 5.9, 5.14–5.17, 5.19) together. This completes the proof of Theorem 2.2.

Proof of Theorem 3.2. Upper bound. We follow to the same arguments as in the proof of Theorem 3.1.

Let $F \in \Im_{nh_n}(\Theta)$ be a true c.d.f. Denote $\theta_0 = \theta(F)$.

We have

$$\int_0^1 K_{h_n}(t-s)\mathrm{d}\left(\hat{F}_n(s) - F\left(s,\hat{\theta}_n\right)\right) = g_{1h_n}(t) + g_{2h_n}(t) + \xi_n(t) \tag{5.20}$$

with

$$\begin{aligned} g_{1h_n}(t) &= \int_0^1 K_{h_n}(t-s)\mathrm{d}(F(s)-F(s,\theta_0)),\\ g_{2h_n}(t) &= \int_0^1 K_{h_n}(t-s)\mathrm{d}\left(F(s,\theta_0)-F\left(s,\hat{\theta}_n\right)\right),\\ \xi_n(t) &= \int_0^1 K_{h_n}(t-s)\mathrm{d}\left(\hat{F}_n(s)-F(s)\right).\end{aligned}$$

Hence we get

$$T(\hat{F}_n, F_{\hat{\theta}_n}) = I_{1n} + I_{2n} + I_{3n} + I_{4n} + I_{5n} + I_{6n} \tag{5.21}$$

with $I_{1n}, \dots, I_{6n}$ defined by (4.3–4.5) respectively with $\epsilon = n$.

Since $F \in \Im_{nh_n}(\Theta)$ then (4.6) holds.

Similarly to (4.8) we get

$$I_{3n} < C|\hat{\theta}_n - \theta_0|^2. \tag{5.22}$$

We have

$$E(I_{4n}) = 0. \tag{5.23}$$

By Schwartz inequality, we get

$$\begin{aligned} E(I_{4n}^2) = n^{-1} \int_0^1 \int_0^1 g_{1h_n}(t_1) g_{1h_n}(t_2) r(t_1) r(t_2) \Bigg(\int_0^1 K_{h_n}(t_1 - s) K_{h_n}(t_2 - s) \mathrm{d}F(s) \\ - \int_0^1 K_{h_n}(t_1 - s) \mathrm{d}F(s) \int_0^1 K_{h_n}(t_2 - s) \mathrm{d}F(s) \Bigg) \\ < n^{-1} \int_0^1 \mathrm{d}F(s) \left(\int_0^1 K_{h_n}(t - s) r(t) g_{1h_n}(t) \mathrm{d}t \right)^2 \\ < n^{-1} \int_0^1 \mathrm{d}F(s) \int_0^1 K_{h_n}^2(t - s) \mathrm{d}t \int_0^1 r^2(t) g_{1h_n}^2(t) \mathrm{d}t < Cn^{-1} h_n^{-1} T_n(F, F_{\theta_0}). \end{aligned} \tag{5.24}$$

We get

$$I_{2n} < I_{1n}^{1/2} I_{3n}^{1/2}. \tag{5.25}$$

Arguing similarly to (4.13–4.20), we get

$$I_{5n} = O_P\left(n^{-1} h_n^{(-1+\omega)/2} \right). \tag{5.26}$$

It remains to study the asymptotic behaviour of I_{6n}.

Denote

$$\sigma^2(h, F) = 2 \int_0^1 r^2(x) \mathrm{d}x \int_{-2}^2 \mathrm{d}u \left[\int_{-1}^1 K(z) K(z + u) \mathrm{d}F(x - zh) \right]^2.$$

For all $1 \le i \le n, 1 \le j \le n$ denote

$$H_n(X_i, X_j) = \int_0^1 \left(K_{h_n}(t - X_i) - \int_0^1 K_{h_n}(t - s) \mathrm{d}F(s) \right) \left(K_{h_n}(t - X_j) - \int_0^1 K_{h_n}(t - s) \mathrm{d}F(s) \right) r(t) \mathrm{d}t.$$

We have

$$I_{6n} = I_{61n} + I_{62n} \tag{5.27}$$

where

$$I_{61n} = 2n^{-2} \sum_{1 \le i < j \le n} H_n(X_i, X_j),$$

$$I_{62n} = n^{-2} \sum_{j=1}^n H_n(X_j, X_j).$$

It follows from (2.13) and (2.14) that

$$\begin{aligned} E(I_{62n}) &= n^{-1}\int_0^1\int_0^1 K^2_{h_n}(t-x)\mathrm{d}F(x)r(t)\mathrm{d}t\left(1+o\left(h_n^{1/2}\right)\right)\\ &= n^{-1}\int_0^1\int_0^1 K^2_{h_n}(t-x)r(x)\mathrm{d}F(x)\mathrm{d}t\left(1+o\left(h_n^{1/2}\right)\right)\\ &= n^{-1}h_n^{-1}\int_0^1 r(t)\mathrm{d}F(t)\int_{-1}^1 K^2(z)\mathrm{d}z\left(1+o\left(h_n^{1/2}\right)\right)\doteq e_n\left(1+o\left(h_n^{1/2}\right)\right). \end{aligned} \tag{5.28}$$

By straightforward calculations, we get

$$\mathrm{Var}(I_{62n}) = O\left(n^{-3}h_n^{-2}\right). \tag{5.29}$$

By (2.13), we get

$$\sup_{h_n<t<1-h_n}\left|\int_0^1 K_{h_n}(t-s)r(s)\mathrm{d}s - r(t)\right| < Ch_n^{\kappa_1}.$$

Hence, using (2.12) and Schwartz inequality, we get

$$\begin{aligned} \left|\int_0^1 r(t)(\mathrm{d}F(t)-\mathrm{d}F(t,\theta_0))\right| &< Ch_n^{\zeta}+Ch_n^{\kappa_1}+\left|\int_{h_n}^{1-h_n}\int_0^1 K_{h_n}(t-s)r(s)\mathrm{d}s(\mathrm{d}F(t)-\mathrm{d}F(t,\theta_0))\right|\\ &< Ch_n^{\zeta}+Ch_n^{\kappa_1}+\left|\int_0^1 r(s)\mathrm{d}s\int_{h_n}^{1-h_n} K_{h_n}(s-t)(\mathrm{d}F(t)-\mathrm{d}F(t,\theta_0))\right| < Ch_n^{\zeta}+Ch_n^{\kappa_1}+CI_{1n}^{1/2}. \end{aligned}$$

Therefore, using $n^{-1}h_n^{-3/2-\omega}\to 0$ as $n\to\infty$ and $I_{1n} > cn^{-1}h_n^{-1/2}$ we get

$$|d_n(h_n, f_{\theta_0}) - e_n| < Cn^{-1}h_n^{-1}I_{1n}^{1/2} + o\left(n^{-1}h_n^{-1/2}\right) = o(I_{1n}). \tag{5.30}$$

By the same arguments, using E2, E3, we get

$$|d_n(h_n, f_{\theta_0}) - d_n(h_n, f_{\hat\theta_n})| < Cn^{-1}h_n^{-1}T_n^{1/2}(\hat F_{\hat\theta_n}, F_{\theta_0}| \le Cn^{-1}h_n^{-1}|\hat\theta_n-\theta_0| = o_P(I_{1n}). \tag{5.31}$$

By Lemma 3 in Hall [17],

$$E(I_{61n}) = 0, \tag{5.32}$$

$$\begin{aligned} n^2h_n\mathrm{Var}(I_{61n}) &= \frac{1}{2}\int_0^1 r(t)\mathrm{d}t\int_{-t/h_n}^{h_n^{-1}-t/h_n} r(t+uh_n)\left[\int_{-1}^1 K(z)K(z+u)\mathrm{d}F(t-zh_n)\right.\\ &\quad\left. -\, h_n\left\{\int_{-1}^1 K(z)\mathrm{d}F(t-zh_n)\right\}\left\{\int_{-1}^1 K(z)\mathrm{d}F(t+uh_n-zh_n)\right\}\right]^2\mathrm{d}u(1+o(1))\\ &= \frac{1}{2}\int_0^1 r^2(t)\mathrm{d}t\int_{-t/h_n}^{h_n^{-1}-t/h_n}\left[\int_{-1}^1 K(z)K(z+u)\mathrm{d}F(t-zh_n)\right]^2\mathrm{d}u(1+o(1)) \doteq \frac{1}{4}\sigma^2(h_n,F)(1+o(1)). \end{aligned} \tag{5.33}$$

By B1, we get

$$\sigma^2(h,F_{\theta_0})-\sigma^2(f_{\theta_0}) = 2\int_0^1 r^2(t)\mathrm{d}t\int_{-2}^2\mathrm{d}u\left(\left(\int_{-1}^1 K(z)K(z+u)f_{\theta_0}(t-zh)\mathrm{d}z\right)^2 - \left(\int_{-1}^1 K(z)K(z+u)\mathrm{d}z\right)^2 f_{\theta_0}^2(t)\right)+O(h)$$
$$\le 2\int_0^1 r^2(t)\mathrm{d}t\int_{-2}^2\mathrm{d}u\int_{-1}^1 K(z)K(z+u)|f_{\theta_0}(t-zh)-f_{\theta_0}(t)|\mathrm{d}z \times\int_{-1}^1 K(z)K(z+u)(f_{\theta_0}(t-zh)+f_{\theta_0}(t))\mathrm{d}z+O(h)$$
$$\le Ch^\kappa\int_0^1 r^2(t)\mathrm{d}t\int_{-2}^2\mathrm{d}u\int_{-1}^1 K(z)K(z+u)|z|^\kappa\mathrm{d}z \times\int_{-1}^1 K(z)K(z+u)(f_{\theta_0}(t-zh)+f_{\theta_0}(t))\mathrm{d}z+O(h)\le Ch^\kappa. \quad (5.34)$$

By E1, we have

$$\frac{1}{2}\sigma^2(h_n,F)-\frac{1}{2}\sigma^2(h_n,F_{\theta_0}) = \int_0^1 r^2(t)\mathrm{d}t\int_{-2}^2\mathrm{d}u\int_{-1}^1 K(z_1)K(z_1+u)\mathrm{d}(F(t-z_1h_n)-F_{\theta_0}(t-z_1h_n)) \times\int_{-2}^2 K(z_2)K(z_2+u)\mathrm{d}(F(t-z_2h_n)+F_{\theta_0}(t-z_2h_n))+O(h_n)$$
$$= J_{1n}+J_{2n}+O(h_n). \quad (5.35)$$

with

$$J_{1n} = \int_0^1 r^2(x)\mathrm{d}x\int_{-1}^1\int_{-1}^1 K_2(z_1-z_2)K(z_1)\mathrm{d}(F(x-z_1h_n)-F_{\theta_0}(x-z_1h_n)) \times K(z_2)\mathrm{d}(F(x-z_2h_n)-F_{\theta_0}(x-z_2h_n))\le CI_{1n}=CT_n(F,F_{\theta_0}), \quad (5.36)$$

$$J_{2n} = 2\int_0^1 r^2(x)\mathrm{d}x\int_{-1}^1\int_{-2}^2 K_2(z_1-z_2)\times K(z_1)\mathrm{d}(F(x-z_1h_n)-F_{\theta_0}(x-z_1h_n))K(z_2)\mathrm{d}F_{\theta_0}(x-z_2h_n)+O(h_n). \quad (5.37)$$

Since the operator $\bar{K}_{2,h}$ is nonnegative, by Shwartz inequality, we get

$$J_{2n} < 2\int_0^1 r^2(x)\mathrm{d}x\left(\int_{-1}^1\int_{-2}^2 K_2(z_1-z_2)K(z_1)\mathrm{d}(F(x-z_1h_n)-F_{\theta_0}(x-z_1h_n)) \times K(z_2)\mathrm{d}(F(x-z_2h_n)-F_{\theta_0}(x-z_2h_n))\right)^{1/2} \times\left(\int_{-1}^1\int_{-2}^2 K_2(z_1-z_2)K(z_1)\mathrm{d}F_{\theta_0}(x-z_1h_n)K(z_2)\mathrm{d}F_{\theta_0}(x-z_2h_n)\right)^{1/2}+O(h_n)$$

$$\leq C\left(\int_0^1 r^2(x)\mathrm{d}x\int_{-1}^1\int_{-2}^2 K_2(z_1-z_2)K(z_1)\mathrm{d}(F(x-z_1h_n)-F_{\theta_0}(x-z_1h_n))\right.$$
$$\times\; K(z_2)\mathrm{d}(F(x-z_2h_n)-F_{\theta_0}(x-z_2h_n)))^{1/2}$$
$$\times\left(\int_0^1 r^2(x)\mathrm{d}x\int_{-1}^1\int_{-2}^2 K_2(z_1-z_2)K(z_1)\mathrm{d}F_{\theta_0}(x-z_1h_n)K(z_2)\mathrm{d}F_{\theta_0}(x-z_2h_n)\right)^{1/2}+O(h_n)$$
$$\leq C\left(\int_0^1 r^2(x)\mathrm{d}x\int_{-1}^1\int_{-2}^2 K_2(z_1-z_2)K(z_1)\mathrm{d}(F(x-z_1h_n)-F_{\theta_0}(x-z_1h_n))\right.$$
$$\times\; K(z_2)\mathrm{d}(F(x-z_2h_n)-F_{\theta_0}(x-z_2h_n)))^{1/2}\leq CT_n^{1/2}(F,F_{\theta_0})+O(h_n). \tag{5.38}$$

By (5.35–5.38), we get

$$|\sigma^2(h_n,F)-\sigma^2(h_n,F_{\theta_0})|<CI_{1n}+CI_{1n}^{1/2}+O(h_n). \tag{5.39}$$

Arguing similarly and using E2, E3, we get

$$|\sigma^2(h_n,F_{\hat\theta_n})-\sigma^2(h_n,F_{\theta_0})|=O_P\left(I_{1n}+I_{1n}^{1/2}\right)+O(h_n) \tag{5.40}$$

as $n\to\infty$.

We have

$$\beta(K_n,F)\leq\Lambda_{1n}+\Lambda_{2n}+\Lambda_{3n}+\Lambda_{4n} \tag{5.41}$$

with

$$\Lambda_{1n}=P_F\left(-T_n\left(\hat F_n,F_{\theta_0}\right)+d_n\left(h_n,f_{\theta_0}\right)<-x_\alpha n^{-1}h_n^{-1/2}\sigma\left(h_n,F_{\theta_0}\right)+O\left(c_nI_{1n}^{1/2}n^{-1/2}h_n^{-1/4}+c_nn^{-1}h_n^{-1/2}\right)\right),$$
$$\Lambda_{2n}=P_F\left(T_n\left(\hat F_n,F_{\theta_0}\right)-T_n\left(\hat F_n,F_{\hat\theta_n}\right)<c_nI_{1n}^{1/2}n^{-1/2}h_n^{-1/4}\right)=P_F\left(I_{2n}+I_{3n}+I_{5n}<c_nI_{1n}^{1/2}n^{-1/2}h_n^{-1/4}\right),$$
$$\Lambda_{3n}=P_F\left(|d_n\left(h_n,f_{\theta_0}\right)-d_n\left(h_n,f_{\hat\theta_n}\right)|>c_nn^{-1}h_n^{1/2}\right),$$
$$\Lambda_{4n}=P_F\left(|\sigma^2\left(h_n,F_{\hat\theta_n}\right)-\sigma^2\left(h_n,F_{\theta_0}\right)|>c_n\right).$$

By (5.22, 5.25, 5.26) and (5.31, 5.40) respectively, there exists $c_n\to 0$ as $n\to\infty$ such that

$$\Lambda_{2n}=o(1),\quad \Lambda_{3n}=o(1),\quad \Lambda_{4n}=o(1). \tag{5.42}$$

By Chebyshov inequality, using (5.41, 5.42) and (5.28, 5.30), we get

$$\begin{aligned}\beta(K_n,F)&=\Lambda_{1n}+o(1)=P_F(-T_n(\hat F_n,F_{\theta_0})+d_n(h_n,F_{\theta_0})+T_n(F,F_{\theta_0})(1+o(1))\\&<-x_\alpha n^{-1}h_n^{-1/2}\sigma(h_n,F_{\theta_0})+T_n(F,F_{\theta_0})(1+o(1)))+o(1)\\&\leq\frac{\mathrm{Var}_F(T_n(\hat F_n,F_{\theta_0}))}{(T_n(F,F_{\theta_0})(1+o(1))-x_\alpha n^{-1}h_n^{-1/2}\sigma(h_n,F_{\theta_0}))^2}.\end{aligned} \tag{5.43}$$

By (5.24, 5.27, 5.29, 5.33) together, we get

$$\mathrm{Var}_F\left(T_n\left(\hat F_n,F_{\theta_0}\right)\right)<2\mathrm{Var}_F(I_{4n})+2\mathrm{Var}(I_{6n})<Cn^{-1}h_n^{-1}I_{1n}+O\left(n^{-2}h_n^{-1}\right). \tag{5.44}$$

By (2.1, 5.43, 5.44) together, for any sequence of c.d.f. F_n,

$$\beta(K_n, F_n) \to 0 \tag{5.45}$$

if $nh_n^{1/2} T_n(F_n, F_{\theta_0}) \to \infty$ as $n \to \infty$.

Denote $\Gamma_C = \{F : nh_n^{1/2} T_n(F, F_{\theta_0}) < C\}$.

By (5.39, 5.40) together, we have

$$\sigma^2(h_n, F)(1 + o(1)) = \sigma^2 \left(f_{\hat{\theta}_n} \right) (1 + o(1)) \tag{5.46}$$

uniformly in $F \in \Gamma_C$.

The estimates in the proof of Lemma 3 in Hall [17] are uniform w.r.t. $F \in \Gamma_C$. Therefore the distributions of $2I_{61n}(\sigma(f_{\hat{\theta}_n})nh_n^{1/2})^{-1}$ converges to the standard normal one uniformly w.r.t. $F \in \Gamma_C$. Hence, using (5.21–5.31, 5.41, 5.42, 5.45, 5.46), we get (3.1). This completes the proof of Theorem 3.2.

Remark 5.1. The corresponding version of (5.20) for the test statistics $\hat{T}_n(\hat{F}_n)$ does not contain the addendum I_{32n}. Therefore, in the analysis of asymptotic behaviour of $\hat{T}_n(\hat{F}_n)$ we do not need to estimate $E(I_{32n})$. This allows to simplify the definition of sets of alternatives and to prove the statement of Remark 2.4.

References

[1] P.J. Bickel and M. Rosenblatt, On Some Global Measures of Deviation of Density Function Estimates. *Ann. Statist.* **1** (1973) 1071-1095.

[2] P. Bickel, C. Klaassen, Y. Ritov and J. Wellner, *Efficient and Adaptive Estimation for the Semiparametric Models*. John Hopkins University Press, Baltimore (1993).

[3] B.M. Brown, Martingale Central Limit Theorems. *Ann. Math. Statist.* **42** (1971) 59-66.

[4] L.D. Brown and M. Low, Asymptotic Equivalence of Nonparametric Regression and White Noise. *Ann. Statist.* **24** (1996) 2384-2398.

[5] M.V. Burnashev, On the Minimax Detection of an Inaccurately Known Signal in a White Gaussian Noise. *Theory Probab. Appl.* **24** (1979) 107-119.

[6] N.N. Chentsov, *Statistical Decision Rules and Optimal Inference*. Moskow, Nauka (1972).

[7] M.S. Ermakov, Minimax Detection of a Signal in Gaussian White Noise. *Theory Probab. Appl.* **35** (1990) 667-679.

[8] M.S. Ermakov, On Asymptotic Minimaxity of Rank Tests. *Statist. Probab. Lett.* **15** (1992) 191-196.

[9] M.S. Ermakov, Minimax Nonparametric Testing Hypotheses on a Density Function. *Theory Probab. Appl.* **39** (1994) 396-416.

[10] M.S. Ermakov, Asymptotic Minimaxity of Tests of Kolmogorov and Omega-squared Types. *Theory Probab. Appl.* **40** (1995) 54-67.

[11] M.S. Ermakov, Asymptotic Minimaxity of Chi-squared Tests. *Theory Probab. Appl.* **42** (1997) 668-695.

[12] M.S. Ermakov, On Distinquishability of Two Nonparametric Sets of Hypotheses. *Statist. Probab. Lett.* **48** (2000) 275-282.

[13] Y. Fan, Testing Goodness of Fit of a Parametric Density Function by Kernel Method. *Econometric Theory* **10** (1994) 316-356.

[14] E. Guerre and P. Lavergne, *Minimax Rates for Nonparametric Specification Testing in Regression Models*, Working Paper. Toulouse University of Social Sciences, Toulouse, France (1999).

[15] B.K. Ghosh and Wei-Mion Huang, The Power and Optimal Kernel of the Bickel–Rosenblatt Test for Goodness of Fit. *Ann. Statist.* **19** (1991) 999-1009.

[16] P. Hall, Integrated Square Error Properties of Kernel Estimators of Regression Function. *Ann. Statist.* **12** (1984) 241-260.

[17] P. Hall, Central Limit Theorem for Integrated Square Error of Multivariate Nonparametric Density Estimators. *J. Multivar. Anal.* **14** (1984) 1-16.

[18] W. Hardle, *Applied Nonparametric Regression*. Cambridge University Press, Cambridge (1989).

[19] J.D. Hart, *Nonparametric Smoothing and Lack-of-fit Tests*. Springer-Verlag, New York (1997).

[20] J.L. Horowitz and V.G. Spokoiny, *Adaptive, Rate-optimal Test of Parametric Model against a Nonparametric Alternative*, Vol. 542, Preprint. Weierstrass-Institute of Applied Analysis and Stochastic, Berlin (1999).

[21] Yu.I. Ingster, Minimax Detection of Signal in l_p-metrics. *Z. Nauchn. Sem. (POMI)* **184** (1990) 152-168.

[22] Yu.I. Ingster and I.A. Suslina, Minimax Detection of Signals for Besov Balls and Bodies. *Probl. Inform. Transm.* **34** (1998) 56-68.

[23] Yu.I. Ingster and I.A. Suslina, *Nonparametric Goodness-of-Fit Testing under Gaussian Model*. Springer-Verlag, New York, *Lecture Notes in Statist.* **169**.

[24] V.D. Konakov, On a Global Measure of Deviation for an Estimate of the Regression Line. *Theor. Probab. Appl.* **22** (1977) 858-868.

[25] O.V. Lepski and V.G. Spokoiny, Minimax Nonparametric Hypothesis Testing: The Case of an Inhomogeneous Alternative. *Bernoulli* **5** (1999) 333-358.

[26] M.A. Lifshits, *Gaussian Random Functions*. TViMS Kiev (1995).

[27] M. Nussbaum, Asymptotic Equivalence of Density Estimation and Gaussian White Noise. *Ann. Statist.* **24** (1996) 2399-2430.

[28] V.I. Piterbarg, *Asymptotic Methods in Theory of Gaussian Proceses and Fields*. Moskow University, Moskow (1988).

[29] J.C.W. Rayner and D.J. Best, *Smooth Tests of Goodness of Fit*. Oxford University Press, New York (1989).

[30] D. Slepian, The One-sided Barrier Problem for Gaussian Noise. *Bell System Tech. J.* **41** (1962) 463-501.

[31] V.G. Spokoiny, Adaptive Hypothesis Testing using Wavelets. *Ann. Statist.* **24** (1996) 2477-2498.

[32] Ch. Stein, Efficient Nonparametric Testing and Estimation, in *Third Berkeley Symp. Math. Statist. and Probab*, Vol. 1. Univ. California Press, Berkeley (1956) 187-195.

[33] W. Stute, Nonparametric Model Checks for Regression. *Ann. Statist.* **25** (1997) 613-641.