The Annals of Statistics
2002, Vol. 30, No. 2, 498–527

ON NONPARAMETRIC TESTS OF POSITIVITY/MONOTONICITY/CONVEXITY

BY ANATOLI JUDITSKY AND ARKADI NEMIROVSKI

INRIA Rhone-Alpes and Technion—Israel Institute of Technology

We consider the problem of estimating the distance from an unknown signal, observed in a white-noise model, to convex cones of positive/monotone/convex functions. We show that, when the unknown function belongs to a Hölder class, the risk of estimating the L_r-distance, $1 \le r < \infty$, from the signal to a cone is essentially the same (up to a logarithmic factor) as that of estimating the signal itself. The same risk bounds hold for the test of positivity, monotonicity and convexity of the unknown signal.

We also provide an estimate for the distance to the cone of positive functions for which risk is, by a logarithmic factor, smaller than that of the "plug-in" estimate.

1. Introduction. Let a nonparametric signal $f\colon\ [0,1] \to \mathbb{R}$ be observed according to the standard "signal + white noise" model, so that the observation is a realization of the Gaussian random process $X_n^f(\cdot)$ on [0, 1]:

$$dX_n^f = f(t)\,dt + n^{-1/2}\,dW(t), \qquad 0 \le t \le 1,\ X_n^f(0) = 0; \tag{1}$$

here $W(\cdot)$ is the standard Wiener process and the parameter n plays the role of "volume of observations." We know that f is regular (belongs to a given Hölder ball Σ, see below) and are interested in doing either of the following:

(I) Estimate the distance

$$\Phi_r[f] = \inf\{\|f - g\|_r \mid g \in \mathcal{M}\}$$

from f to a given cone $\mathcal{M}$ in the space of functions on [0, 1]; here r, $1 \le r < \infty$, is fixed, and $\|\cdot\|_r$ is the standard L_r-norm on the unit segment. To be more specific, we consider the particular cases when $\mathcal{M}$ is the cone of (i) nonpositive, (ii) monotone or (iii) concave functions.

(II) Decide whether or not the signal belongs to the cone $\mathcal{M}$; more precisely, distinguish between the hypotheses

$$H_0\colon\ f \in \mathcal{M}$$

and

$$H_\varepsilon\colon\ \Phi_r[f] \ge \varepsilon.$$

Received March 2000; revised September 2001.

AMS 2000 subject classifications. 62G10, 62G08, 90C25.

Key words and phrases. Tests of convexity, nonparametric test, estimation of nonsmooth functionals, minimax risk.

We quantify our abilities to handle (I) and (II) by the minimax risks defined as follows:

1. For (I), the minimax risk is

$$\mathcal{R}^*_{\mathrm{est}}(n) = \inf_{\widehat{F}} \sup_{f \in \Sigma} E\{|\widehat{F}[X_n^f] - \Phi_r[f]|\},$$

   where the infimum is taken over all estimates [Borel, w.r.t. the uniform metric, functionals of observation (1)] and E stands for expectation over the noise affecting the observation.
2. For (II), the minimax risk ("resolution") is

$$\mathcal{R}^*(n) = \inf\{\varepsilon \mid H_0 \text{ and } H_\varepsilon \text{ are } p\text{-testable via observation (1)}\},$$

   where p-testability of the pair H_0, H_ε via observation (1) means that there exists a test—a Borel functional $T[\cdot]$ taking values in $\{0, 1\}$—such that

$$\sup_{f \in \Sigma} P_{f \text{ meets } H_0}\{T[X_n^f] = 1\} + \sup_{f \in \Sigma} P_{f \text{ meets } H_\varepsilon}\{T[X_n^f] = 0\} \le p, \tag{2}$$

   for p small (say $p \le 1/8$). Here the probability is taken w.r.t. the distribution of noise affecting the observation.

As usual, we are interested in the asymptotic behavior of the minimax risks as $n \to \infty$.

We focus on the case when the set Σ of signals in question is the Hölder ball $\Sigma_\rho(\beta, L)$ specified by the triple of parameters $\beta > 0$, $L > 0$ and $\rho \ge 2L$. It is defined as the set of all bounded continuous functions $f\colon [0, 1] \to [-\rho, \rho]$ which are $m = \lfloor \beta - 0 \rfloor = \max\{m \in \mathbb{Z} : m < \beta\}$ times continuously differentiable and such that the mth derivative is Hölder continuous, with exponent $\gamma = \beta - m$ and constant $2L$:

$$|f^{(m)}(t) - f^{(m)}(t')| \le 2L|t - t'|^{\beta - m} \qquad \forall t, t' \in [0, 1].$$

Note that the minimax risk of recovering $\Phi_r[f]$, $f \in \Sigma$, cannot be worse than the minimax risk

$$\mathcal{R}(n) = \inf_{\widehat{f}} \sup_{f \in \Sigma} E\{\|\widehat{f}(X_n^f) - f\|_r\} = O\big(L^{1/(2\beta+1)} n^{-\beta/(2\beta+1)}\big) \tag{3}$$

of recovering signals $f \in \Sigma$, the estimation error being measured in the $\|\cdot\|_r$-norm. This observation is an immediate consequence of the fact that $\Phi_r[\cdot]$ is Lipschitz continuous with Lipschitz constant equal to 1 in the $\|\cdot\|_r$-norm. So the worst-case risk over $f \in \Sigma$ of estimating $\Phi_r[f]$ by the "plug-in" estimate

$$\widehat{F}(Y) = \Phi_r[\widehat{f}(Y)],$$

associated with an estimate $\widehat{f}(\cdot)$ of f, is not worse than the worst-case $\|\cdot\|_r$-risk of $\widehat{f}$ itself:

$$\sup_{f \in \Sigma} E\{|\Phi_r[\widehat{f}(X_n^f)] - \Phi[f]|\} \le \sup_{f \in \Sigma} E\{\|\widehat{f}(X_n^f) - f\|_r\}.$$

Taking the infimum over $\widehat{f}$ we come to

$$\mathcal{R}^*_{\mathrm{est}}(n) \le \mathcal{R}(n),$$

as claimed.

Now, the minimax risk $\mathcal{R}^*(n)$ in the hypothesis testing problem (II) cannot be "essentially worse" than $\mathcal{R}^*_{\mathrm{est}}(n)$ and thus it cannot be essentially larger than $\mathcal{R}(n)$:

$$\mathcal{R}^*(n) \le 32\mathcal{R}^*_{\mathrm{est}}(n) \qquad [\le 32\mathcal{R}(n)]. \tag{4}$$

Indeed, given an estimate $\widehat{F}$ of $\Phi_r[f]$ with certain worst-case (w.r.t. $f \in \Sigma$) risk R, we can convert it into the following test of H_0 against H_{32R}: given X_n^f, set $T = \{\widehat{F}(X_n^f) \le 16R\}$ and claim that the true hypothesis is H_0 if $T = 1$ or H_{32R} if $T = 0$. It follows immediately from the Chebyshev inequality that the resulting test satisfies (2) with $p = 1/8$, $\varepsilon = 32R$; thus $\mathcal{R}^*(n) \le 32R$. Since R can be chosen arbitrarily close to $\mathcal{R}^*_{\mathrm{est}}(n)$ it implies the inequality (4).

Our results can be summarized as follows: suppose that the "degree of regularity" β of our signals "is coherent" with the hypothesis H_0 (namely, $\beta \ge 1$ when $\mathcal{M}$ is the cone of nonincreasing functions and $\beta \ge 2$ when $\mathcal{M}$ is the cone of concave functions). The latter assumption, roughly speaking, says that the degree of regularity of functions from $\mathcal{M}$ is not better than the a priori known degree of regularity of our signals. Then the main result of the present paper (Theorem 1 below) states that the "plug-in" estimate of $\Phi_r[\cdot]$ and the associated test for distinguishing between H_0 and H_ε are "nearly optimal." That is, their performance cannot be improved by more than logarithmic factors. Specifically, we prove that for any sufficiently large observation sample length n one has

$$\mathcal{R}^*(n) \ge C(\beta, r)\mathcal{R}(n)\big(\ln \mathcal{R}(n)\big)^{-\theta(\beta)}, \qquad C(\beta, r) > 0. \tag{5}$$

Note that (5) combines with (4) to imply that all three quantities $\mathcal{R}^*(n)$, $\mathcal{R}^*_{\mathrm{est}}(n)$, $\mathcal{R}(n)$ are "nearly" (up to factors logarithmic in these quantities) the same:

$$\begin{aligned} O\left(\frac{L^{1/(2\beta+1)} n^{-\beta/(2\beta+1)}}{\ln^{\theta}(n)}\right) &\le \mathcal{R}^*(n) \le O(1)\mathcal{R}^*_{\mathrm{est}}(n) \\ &\le O\big(L^{1/(2\beta+1)} n^{-\beta/(2\beta+1)}\big). \end{aligned} \tag{6}$$

The outlined result deserves some comments. From the sizeable literature on estimating functionals of nonparametric signals and nonparametric hypothesis testing (see, e.g., [1–27, 30, 37] and references therein) the following are known:

1. Typically, the minimax risk associated with a nonparametric hypothesis testing problem used to "decide whether the signal underlying observations belongs to a given set X or is at a 'large' $\|\cdot\|_r$-distance from the set" is essentially less than the risk at which one can estimate the $\|\cdot\|_r$-distance from the signal to the set X. For example, when $X = \{0\}$ ("decide whether the signal f

underlying the observations is nontrivial"), the minimax risk, on $\Sigma_\rho(\beta, L)$, in the testing problem is $O(n^{-2\beta/(4\beta+1)})$ for $1 \le r \le 2$ and $O(n^{-\beta/(2\beta+1-1/r)})$ for $r > 2$ [20], [37]. However, the minimax risk, on the same set $\Sigma_\rho(\beta, L)$, of recovering $\|f\|_r$ is, up to factors logarithmic in n, the plug-in risk $O(n^{-\beta/(2\beta+1)})$, except for the case when r is an even integer [29]. To the best of our knowledge, the phenomenon observed in the current paper, which we refer to as the "near optimality" of the simplest plug-in schemes, both in hypothesis testing and estimating the distance from f to X is new.

2. The minimax risk, w.r.t. $\Sigma_\rho(\beta, L)$, of recovering a "good" functional $F[f]$ is essentially better than the risk $\mathcal{R}(n) = O(n^{-\beta/(2\beta+1)})$ of recovering the signal f itself. For example, when F is smooth (Frechet differentiable on L_2 with Hölder continuous, with exponent γ, gradient), F can be recovered (cf. [11], [12], [16]) on $\Sigma_\rho(\beta, L)$ with "parametric" worst-case risk $O(n^{-1/2})$. Even a nonsmooth (although otherwise "simple") functional $F[f] = \|f\|_2$ can be recovered on $\Sigma_\rho(\beta, L)$ with risk $O(n^{-4\beta/(4\beta+1)})$ [17], which still is much better than the risk $\mathcal{R}(n)$ of a nonparametric estimate $\widehat{f}$ of $f \in \Sigma_\rho(\beta, L)$ in the $\|\cdot\|_2$-norm. Relation (6) says that this nice phenomenon disappears when instead of estimating the $\|\cdot\|_2$-distance from f to 0 (or from f to a linear subspace in L_2—the latter problem turns out to be essentially as good as the one of estimating $\|f\|_2$) we want to estimate the $\|\cdot\|_2$-distance from f to a less trivial (although quite natural) convex cone, say, the one of nonpositive functions.

The outlined "near optimality" result states that the plug-in estimate or test related to $\|\cdot\|_r$-distances from the cones of nonpositive, nonincreasing or concave functions is optimal up to logarithmic-in-n factors. A natural question is whether one can replace "logarithmic-in-n" by "constant" here, that is, whether there is a possibility of measuring the distances from nonparametric signals to the cones in question asymptotically better than we can recover the signals themselves. We demonstrate that such a possibility does exist in the case of (i), that is, when the associated cone is the cone of nonpositive functions. However, we do not know whether such a possibility exists in cases (ii) and (iii). Furthermore, the answer to the following question would be of major interest:

(?) Let $F[f]$ be a convex functional on the space $C[0,1]$ of real-valued continuous functions on $[0,1]$ such that F is Lipschitz continuous, with constant 1 w.r.t. the norm $\|\cdot\|_r$ (variant: let $F[f]$ be the $\|\cdot\|_r$-distance from f to a given convex closed, in the uniform norm, subset X of $C[0,1]$). Can $F[f]$ be estimated via the observation (1) with the worst-case risk over $\Sigma_\rho(\beta, L)$ which is "better in order than the plug-in risk." In other words, is $o(\mathcal{R}(n))$ as $n \to \infty$?

In the case when F is the $\|\cdot\|_r$-distance from f to a closed convex set X, we may pose a similar question about the risk of testing the hypothesis "$f \in X$" versus the alternative "the $\|\cdot\|_r$-distance from f to X is greater than or equal to ε."

Recall that in all particular cases when the answer to (?) is known, it is affirmative (e.g., the $\|\cdot\|_r$-norm of $f \in \Sigma_\rho(\beta, L)$ indeed can be recovered asymptotically better than f; see [8], [17], [18], [36]). Is it a "law of nature" or merely a consequence of "simplicity" of the functionals F (sets X) which have been studied?

There is another important point to be stressed about the result in (5). Up to now we have only considered the case when the distance to the cone is measured with the L_r-norm with $r < \infty$. On the other hand, the L_∞-distance is a more sensitive measure of discrepancy of functions on $[0, 1]$. Note that the minimax rates of convergence for the problems (i) and (ii) for the risks associated with the L_∞-distance are well known (cf., for instance, [3], and the references therein). They are of order $O((n/\log n)^{-\beta/(2\beta+1)})$. Furthermore, in the problem of the estimation of the distance to the cone of nonpositive functions (testing the hypotheses about the nonpositivity of f) even the sharp constant in the minimax risk has been obtained in [30]. Quite recently, those results were extended to the case of monotone functions in [6]. As we have already seen, in the problem studied in the literature, the risks $\mathcal{R}^*_{\text{est}}$ and $\mathcal{R}^*$ for $r < \infty$ are essentially smaller than the risks associated with the L_∞-distance, which motivated the choice of L_r-norm with relatively small r in the test problems. An important consequence of the lower bound in (5) is that in the problems considered in this paper there is no serious motivation for using a distance other than L_∞.

The rest of the paper is organized as follows. Relation (5) is carefully stated in Section 2. Then a "better-than-plug-in" estimate of the distance from a signal to the cone of nonpositive functions is built in Section 3. The proofs of the results are collected in Sections 4 and 5.

2. Lower bound on $\mathcal{R}^*(n)$. Let us fix a class $\Sigma_\rho(\beta, L)$ with $\rho \geq 2L$, an $r \in [1, \infty)$ and a nonnegative integer $q \leq \beta$, and let $\mathcal{M}_q$ be the closure, in the uniform norm on $[0, 1]$, of the set of all C^∞ functions with qth derivative nonpositive everywhere. Let, further,

$$\Phi_r[f] = \inf\{\|f - g\|_r \mid g \in \mathcal{M}_q\}$$

be the $\|\cdot\|_r$-distance from a continuous function $f\colon [0, 1] \to \mathbb{R}$ to $\mathcal{M}_q$. Note that, for $q = 0,\ 1$ and 2, $\Phi_r[f]$ is the $\|\cdot\|_r$-distance from f to the cones of nonpositive, nonincreasing and concave functions, respectively—the cones in which we indeed are interested.

Given observation $X_n^f(t)$ in (1) of a signal $f \in \Sigma_\rho(\beta, L)$ we are interested in distinguishing between the following two hypotheses:

$$H_0\colon\ f \in \mathcal{M}_q;$$
$$H_\varepsilon\colon\ \Phi_r[f] \geq \varepsilon.$$

Recall that we say the hypotheses H_0 and H_ε are *p-testable*, n being the size of the observation sample, if there exists a test $T[\cdot]$ [a functional of observation (1) taking values 0 or 1] satisfying (2), and we denote by $\mathcal{R}^*(n)$ the infimum of those $\varepsilon > 0$ for which the hypotheses H_0 and H_ε are $1/8$-testable.

Our goal is to establish the following.

THEOREM 1. *Let $\beta \geq q$. There exist $C = C(\beta, r) > 0$, $\vartheta = \vartheta(\beta, r)$ and $M = M(\beta, \rho, L, r, q)$ such that for all $n \geq M$ it holds that*

$$L^{-1/(2\beta+1)} n^{\beta/(2\beta+1)} \mathcal{R}^*(n) \geq C\left(\frac{1}{\ln n}\right)^{\vartheta}. \tag{7}$$

REMARK. When speaking about the cones of nonincreasing ($q = 1$) and ($q = 2$) concave functions, we have assumed that the regularity of f is compatible with the regularity of functions from the cone in question (i.e., respectively, $\beta \geq 1$ and $\beta \geq 2$). What happens when this assumption is violated? It can be easily shown that here the plug-in estimates or tests become "essentially nonoptimal." Say, when $\beta < 1$, the minimax [on $\Sigma_\rho(\beta, L)$] risk of recovering the $\|\cdot\|_2$-distance from f to the cone of nonincreasing functions is $O(n^{-((1/3)\wedge 2\beta/(4\beta+1))})$. This phenomenon is quite natural. Indeed, the degree of regularity of a monotone function is "nearly 1." (To be more precise, the L_1-norm of the derivative f' is bounded.) So the projection $\widehat{f_n}$ of f on the cone of nonincreasing functions can be recovered with the rate $n^{-1/3}$ (see also [34]). Now, to measure the distance from a "poorly regular" signal f to the cone of monotone functions is essentially the same as to measure the distance from f to "the point" $\widehat{f_n}$ (cf. [28]). Thus the minimax risk in this latter problem, as we have just mentioned, is significantly less than $\mathcal{R}(n)$.

3. Distance to the cone of positive functions. Let us denote $f_+(t) = f(t)\mathbb{1}_{f(t)>0}$. We consider the problem of estimating from the observation (1) the functional

$$\Phi_r[f] = \|f_+\|_r = \left(\int_0^1 f_+^r(t)\, dt\right)^{1/r}.$$

We consider the maximal risk of the estimator $\widehat{\Phi}_n$:

$$\mathcal{R}^*_{\text{est}}(\widehat{\Phi}_n) = \sup_{f \in \Sigma} E_f \left|\widehat{\Phi}_n - \Phi_r[f]\right|,$$

where $\Sigma = \Sigma_1(\beta, L)$ is the Hölder ball of functions such that $\|f\|_\infty < 1$.

THEOREM 2. *For any $1 \leq r < \infty$ there exist an estimator $\widehat{\Phi}_n$ and a constant C such that, for $n \geq n_0(L, \beta, r)$,*

$$\mathcal{R}^*_{\text{est}}(\widehat{\Phi}_n) \leq C L^{1/(2\beta+1)} (n \ln n)^{-2\beta/(2\beta+1)}.$$

The proof of the theorem is presented in Section 5. The idea of the estimation algorithm described below is as follows:

1. We start with the classical kernel smoothing $\widehat{f}_h(t)$ of the observation $X(t)$ with some "bandwidth" h. The parameter h is chosen in such a way that the smoothing $f_h(t)$ of the function f is at the desired distance from f, so that to estimate the distance $\Phi_r[f]$ it suffices to estimate the functional $\Phi_r[f_h]$.
2. To this end we use the technique suggested in [29]: we approximate the function $t^r \mathbb{1}_{t>0}$ with a trigonometric polynomial T_N of high order N and construct the unbiased estimate $T_N^*(\widehat{f}_h(t))$ of $T_N(f_h(t))$ [here $T_N^*(\cdot)$ is a trigonometric polynomial of degree N with specially corrected coefficients]. Finally, we set $\widehat{\Phi}_n = (\int_0^1 T_N^*(\widehat{f}_h(t))\,dt)^{1/r}$.

Let m be the largest integer smaller than β and let $K(\cdot)$ be a kernel of order m with compact support. That is, K satisfies the conditions

$$\begin{aligned} K(t) &= 0 \qquad \text{for } |t| > 1,\\ \int K(t)\,dt &= 1,\\ \int t^i K(t)\,dt &= 0 \qquad \text{for } i = 1, \dots, m. \end{aligned}$$

Along with the kernel K we consider the corrected *left* kernel $K^+(t)$, supported on $[0, 2]$ and the *right* kernel $K^-(t)$, with its support in $[-2, 0]$, which satisfy the moment relations above and have the same L_2-norm.

Let $[r]$ stand for the integer part of r. We define a partition function $r(t)$: $\mathbb{R}^+ \to \mathbb{R}^+$, $r(t) \geq 0$ such that its $([r]+1)$st derivative $r^{([r]+1)}(t)$ is bounded and

$$\begin{aligned} r(t) &= 1 \qquad \text{if } t \leq 1,\\ 0 < r(t) &< 1 \qquad \text{if } 1 < t < 2,\\ r(t) &= 0 \qquad \text{if } t \geq 2. \end{aligned} \tag{8}$$

ALGORITHM 1.
Step 1. Set

$$h = (L^2 n \log n)^{-1/(2\beta+1)},$$

and compute the smoothing

$$\widehat{f}(t) = \frac{1}{h}\int_0^1 K\left(\frac{t-u}{h}\right) dX(u).$$

This is a standard kernel estimator of $f(t)$. In fact, to correct the boundary effects we have to use, for $t \in [-1, -1+h]$, the corrected kernel K^+ and, for $t \in [1-h, 1]$, the kernel K^-. To simplify the notation in what follows we use the same notation K for all three kernels.

Note that the estimate $\widehat{f}(t)$ possesses deterministic and stochastic components:

$$\widehat{f}(t) = f_h(t) + \lambda_h \xi_h(t),$$

where

$$f_h(t) = \frac{1}{h}\int_0^1 K\left(\frac{t-u}{h}\right) f(u)\,du \tag{9}$$

and

$$\lambda_h = \left[E\left(\int_0^1 K\left(\frac{t-u}{h}\right) n^{-1/2}\,dW(u)\right)^2\right]^{1/2} = \frac{\|K\|_2}{\sqrt{nh}},$$

$$\xi_h(t) = \frac{1}{h\lambda_h}\int_0^1 K\left(\frac{t-u}{h}\right) n^{-1/2}\,dW(u) = \frac{1}{\|K\|_2\sqrt{h}}\int_0^1 K\left(\frac{t-u}{h}\right) dW(u).$$

Here $\|\cdot\|_2$ stands for the L_2-norm and $\xi_h(t)$ is clearly $N(0,1)$. Hence

$$E\,\widehat{f_h}(t) = f_h(t),$$

$$\operatorname{Var}\widehat{f_h}(t) = E\big(\widehat{f_h}(t) - f_h(t)\big)^2 = \lambda_h^2.$$

Step 2. Define the function $q(t)$ as

$$q(t) = t^r \mathbb{1}_{t>0} r(t).$$

Let α_k, $k = 0, 1, \dots$, be Fourier coefficients of $q(\cdot)$: $\alpha_k = \int_0^2 q(t)\psi_k(t)\,dt$, where

$$\psi_{2k}(t) = 2^{-1/2}\cos(\pi k t/2), \qquad \psi_{2k+1}(t) = 2^{-1/2}\sin(\pi k t/2).$$

Set

$$N = \big[\theta L^{-1/(2\beta+1)}(n\log n)^{\beta/(2\beta+1)}\big],$$

where $0 < \theta < \kappa$, with

$$\kappa = \begin{cases} \dfrac{4}{\pi\|K\|_2\sqrt{r(2\beta+1)}}, & \text{for } r > 2, \\[2ex] \dfrac{4}{\pi\|K\|_2\sqrt{2(2\beta+1)}}, & \text{for } 1 \le r \le 2. \end{cases} \tag{10}$$

For $k = 0, \dots, N$ we define the functions

$$\phi_{2k,\lambda}(t) = \exp(\pi^2 k^2 \lambda^2/8)\psi_{2k}(t), \qquad \phi_{2k+1,\lambda}(t) = \exp(\pi^2 k^2\lambda^2/8)\psi_{2k+1}(t)$$

and the polynomial

$$T^*_{N,\lambda}(t) = \sum_{k=0}^N \alpha_k \phi_{k,\lambda}(t). \tag{11}$$

Finally, we define the estimator $\widehat{\Phi}_n$ of Φ_n as

$$\widehat{\Phi}_n = \left(\int_0^1 T^*_{N,\lambda_h}\big(\widehat{f_h}(t)\big)\,dt\right)^{1/r}.$$

4. Proof of Theorem 1. The idea of the proof realized below can be summarized as follows. We construct a parametric family $\{G_u(t)\}$ of functions from $\mathcal{F}$ and two discrete (and asymmetric) prior distributions μ_0 and μ_1 on the parameter set which depend on n. We ensure that when drawn from the measure μ_0 random functions $G_u(t)$ are in the cone $\mathcal{M}_q$, while those generated from μ_1 "typically" have a "significant non-$\mathcal{M}_q$" part: $\Phi_r[G_u] > \delta_*$ with certain $\delta_* > 0$. Then assuming that the hypotheses H_0 and H_{δ_*} are testable, it should be possible to distinguish well between the two hypotheses on observations (1) saying that the observations are associated with random G distributed according to priors μ_0 and μ_1, respectively. On the other hand, we will show that our priors result in the distributions of observations (1) too close to each other to allow for reliable identification. Thus, the assumption that it is possible to distinguish well between the hypotheses H_0^N and $H_{\delta_*}^N$ leads to a contradiction, whence $\mathcal{R}^*(n) \geq \delta_*$. Note that similar ideas are used in the study [19].

The technique used here is close to that in [29] and is rather nonstandard. For instance, the prior measures μ_0 and μ_1 possess a number of moments, increasing in n, that coincide. Their construction is based on the theory of Chebyshev polynomials.

4.1. *Setup.*

4.1.1. *Function $G_u(\cdot)$.* Let us partition the segment $[0, 1]$ into $q + 1$ equal segments Δ_ℓ, $\ell = 0, 1, \ldots, q$, and let t_ℓ be the left endpoints of these segments. Given a C^∞-function g which vanishes outside Δ_0, consider the $(q+1) \times (q+1)$ matrix $A[g]$ with entries

$$a_{i\ell}[g] = \int_{\Delta_\ell} t^i g(t - t_\ell)\,dt, \qquad \ell, i = 0, 1, \ldots, q.$$

Of course, one can choose g to be positive on the interior of Δ_0 and such that the matrix $A[g]$ is nonsingular (look what happens if g is a probability density close, in the weak sense, to the unit mass placed at the midpoint of Δ_0). Let us once and for all fix a nonnegative C^∞-function g which vanishes outside of Δ_0, is positive on the interior of the segment and is such that the matrix $A[g]$ is nonsingular. Since $A[g]$ is nonsingular, there exist reals ω_ℓ, $\ell = 0, 1, \ldots, q$, such that

$$\sum_{\ell=0}^{q} \omega_\ell a_{i\ell}[g] = \begin{cases} 0, & i < q, \\ 1, & i = q. \end{cases} \tag{12}$$

It is clear that among the reals ω_ℓ, $\ell = 0, \ldots, q$, some are positive (otherwise $\sum_\ell \omega_\ell a_{q\ell}[g]$ would be nonpositive). Let $\ell_* \in \{0, 1, \ldots, q\}$ be such that ω_{ℓ_*} is positive. Also let $\omega = \max_\ell |\omega_\ell|$. We define the function $G_u(t)$ (t is a real argument, u is a real parameter) as follows:

1. When $q = 0$, we set

$$G_u(t) = ug(t).$$

2. When $q > 0$, we set $G_u(t) = 0$ for $t \le 0$ and

$$G_u(t) = \int_0^t \frac{(t-\tau)^{q-1}}{(q-1)!} \left(\sum_{\ell=0}^{q} [\omega_\ell u - \widehat{\omega}_\ell] g(\tau - t_\ell) \right) d\tau,$$

$$\widehat{\omega}_\ell = \begin{cases} 0, & \ell = \ell_*, \\ \omega, & \text{otherwise}, \end{cases}$$

for $t \ge 0$. In other words, $G_u(\cdot)$ vanishes on the nonpositive ray; when $t \ge 0$, $G_u(\cdot)$ satisfies the relation

$$\begin{aligned} &\frac{d^q}{dt^q} G_u(t) = \sum_{\ell=0}^{q} [\omega_\ell u - \widehat{\omega}_\ell] g(t - t_\ell); \\ &\frac{d^i}{dt^i}\bigg|_{t=0} G_u(t) = 0, \qquad i = 0, 1, \dots, q-1. \end{aligned} \tag{13}$$

Note that by construction

$$G_u(t) = G(t) + uD(t),$$

where both G and D are C^∞-functions vanishing on the nonpositive ray.

We make the following claim.

G.1. $\frac{d^q}{dt^q} G_u(t)$ and $D(\cdot)$ vanish outside of [0, 1].

Claim G.1 is evident when $q = 0$; let us verify this claim for $q \ge 1$. The fact that $\frac{d^q}{dt^q} G_u(t)$ vanishes when $t \ge 1$ is readily given by (13), so that all we need is to verify that $D(t)$ vanishes outside of [0, 1]. We already know that $D(t) = 0$ when $t \le 0$. For $t \ge 1$ we have

$$\begin{aligned} D(t) &= \int_0^t \frac{(t-\tau)^{q-1}}{(q-1)!} \Bigg[\underbrace{\sum_{\ell=0}^{q} \omega_\ell g(\tau - t_\ell)}_{h(\tau)} \Bigg] d\tau \\ &= \int_0^1 \frac{(t-\tau)^{q-1}}{(q-1)!} h(\tau)\, d\tau \quad [\text{since } h(\cdot) \text{ vanishes outside of } [0,1]] \\ &= \sum_{i=0}^{q-1} c_i(t) \int_0^1 \tau^i h(\tau)\, d\tau \\ &= \sum_{i=0}^{q-1} c_i(t) \sum_{\ell=0}^{q} \omega_\ell a_{i\ell}[g] \\ &= 0 \quad [\text{see } (12)]. \end{aligned}$$

G.2. When $-1 \le u \le 0$, $G_u(t) \in \mathcal{M}_q$, and when $0 \le u \le 1$, we have

$$\Phi_r[G_u(\cdot)] \ge \kappa_1 u \tag{14}$$

(from now on, $\kappa_i > 0$ depend on β, r only).

Indeed, when $-1 \le u \le 0$, the right-hand side in (13) is nonpositive (since g is nonnegative, $\omega_{\ell_*} > 0$ and $\widehat{\omega}_\ell \ge |\omega_\ell|$ when $\ell \ne \ell_*$), that is, $g_u(\cdot) \in \mathcal{M}_q$ for the indicated u. Now let $0 \le u \le 1$. By (13), the qth derivative of $G_u(\cdot)$ is nonnegative on Δ_{ℓ_*} and is greater than or equal to $\kappa_{1,1}u$ on a permanently fixed segment $\Delta \subset \Delta_{\ell_*}$. Hereafter let ψ be a fixed nontrivial nonnegative C^∞-function which vanishes outside of Δ, and let $\theta(t) = (-1)^q \frac{d^q}{dt^q}\psi(t)$. Whenever $h(\cdot) \in \mathcal{M}_q$, we have $\frac{d^q}{dt^q}[G_u(t) - h(t)] \ge \kappa_{1,1}u$ on Δ, whence

$$\int [G_u(t) - h(t)]\theta(t)\,dt = \int_\Delta \left(\frac{d^q}{dt^q}[G_u(t) - h(t)]\right)\psi(t)\,dt \ge \kappa_{1,2}u,$$

so that $\|G_u(\cdot) - h(\cdot)\|_r \ge \frac{\kappa_{1,2}u}{\|\theta\|_{r/(r-1)}}$, and (14) follows.

We set

$$\begin{aligned} &\text{(a)} \quad \kappa_0^{-1} = 1 + \max\nolimits_{0\le k\le\beta+1} \max_{\substack{|u|\le 1\\ 0\le t\le 1}} \left|\frac{d^k}{dt^k} G_u(t)\right|, \\ &\text{(b)} \quad \kappa_* = \kappa_0 \|D(\cdot)\|_2. \end{aligned} \tag{15}$$

4.1.2. *Parameters N, h and $\alpha(N)$.* Given n, let us set

$$\begin{aligned} &\text{(a)} \quad N = \lfloor (200L\kappa_*)^{2/(2\beta+1)} (n\ln n)^{1/(2\beta+1)} \rfloor, \\ &\qquad\quad h = N^{-1}; \\ &\text{(b)} \quad \alpha(N) = L\kappa_* n^{1/2} N^{-\beta-1/2}. \end{aligned} \tag{16}$$

Note that for all sufficiently large values of n the following holds:

$$\frac{0.001}{\sqrt{\ln n}} \le \alpha(N) \le \frac{0.01}{\sqrt{\ln N}}. \tag{17}$$

4.2. *Translation to the space of sequences.* We are about to translate our problem of hypothesis testing to the space of sequences. Let $I_1, \ldots, I_N$ be the partition of the segment $[0, 1]$ into N segments of length h each, and let t_i be the left endpoint of I_i. We associate with a point $\theta = (\theta_1, \ldots, \theta_N)$ from the unit cube $B_N = [-1, 1]^N$ the function

$$\begin{aligned} f_\theta(t) &= L\kappa_0 h^\beta \sum_{i=1}^N G_{\theta_i}\big(h^{-1}(t - t_i)\big) \\ &= L\kappa_0 h^\beta \Bigg[\underbrace{\sum_{i=1}^N G\big(h^{-1}(t - t_i)\big)}_{F(t)} + \sum_{i=1}^N \theta_i D\big(h^{-1}(t - t_i)\big)\Bigg] \end{aligned} \tag{18}$$

and claim first that $f_\theta \in \Sigma_\rho(\beta, L)$.

Indeed, by construction f_θ is a C^∞-function on the axis which vanishes on the nonpositive ray. For $\beta + 1 \geq p \geq q$ we have

$$f_\theta^{(p)}(t) = L\kappa_0 h^{\beta - p} \sum_{i=1}^{N} H_i^p\big(h^{-1}(t - t_i)\big), \qquad H_i^p(t) = \frac{d^p}{dt^p} G_{\theta_i}(t).$$

Since $p \geq q$, the functions H_i^p vanish outside the respective segments I_i (by G.1). Taking into account the origin of κ_0, we conclude that whenever $\beta + 1 \geq p \geq q$ the following hold:

$$\text{(19)} \qquad f_\theta^{(p)}(t_i) = 0, \qquad i = 1, \ldots, N; \qquad |f_\theta^{(p)}(t)| \leq L h^{\beta - p}, \qquad 0 \leq t \leq 1.$$

Now consider two cases: (a) β is not an integer; (b) β is an integer.

In the case of (a), the largest integer, m, which is less than or equal to β, is greater than or equal to q, so that (19) is valid for $p = m$ and $p = m + 1$. Let $t \leq t'$ be two points of $[0, 1]$. If t, t' are from the same segment of the partition $[0, 1] = I_1 \cup \cdots \cup I_N$, then from (19) as applied to $p = m + 1$ it follows that

$$\begin{aligned} |f_\theta^{(m)}(t) - f_\theta^{(m)}(t')| &\leq L h^{\beta - m - 1} |t' - t| \\ &\leq L h^{\beta - m - 1} |t' - t|^{\beta - m} h^{1 - \beta + m} = L |t - t'|^{\beta - m}. \end{aligned}$$

If the points $t < t'$ belong to two distinct segments of the partition, then let $t_+ \leq t_-$ be, respectively, the right endpoint of the segment containing t and the left endpoint of the segment containing t'. Taking into account that $f_\theta^{(m)}$ vanishes at $t_\pm$ by (19) and applying the above computation, we get

$$\begin{aligned} |f_\theta^{(m)}(t)| &= |f_\theta^{(m)}(t) - f_\theta^{(m)}(t_+)| \leq L|t - t_+|^{\beta - m} \leq L|t - t'|^{\beta - m}, \\ |f_\theta^{(m)}(t')| &= |f_\theta^{(m)}(t') - f_\theta^{(m)}(t_-)| \leq L|t' - t_-|^{\beta - m} \leq L|t - t'|^{\beta - m}, \end{aligned}$$

whence

$$|f_\theta^{(m)}(t) - f_\theta^{(m)}(t')| \leq 2L|t - t'|^{\beta - m} \qquad \forall t, t' \in [0, 1],$$

so that $f_\theta \in \mathcal{H}(\beta, L)$.

In the case of (b), (19) is valid for $p = \beta$ (recall that $q \leq \beta$). It follows that $|f_\theta^{(\beta)}(t)| \leq L$; that is, $f_\theta^{(m)} = f_\theta^{(\beta - 1)}$ is Lipschitz continuous with constant L, whence $f_\theta \in \mathcal{H}(\beta, L)$.

We have seen that $f_\theta \in \mathcal{H}(\beta, L)$. This inclusion combines with the fact that f_θ is a C^∞-function vanishing on the nonpositive ray to yield that $|f_\theta(t)| \leq 2L \leq \rho$ on $[0, 1]$. Thus, $f_\theta \in \Sigma_\rho(\beta, L)$, as claimed.

We further claim that

$$\text{(20)} \qquad \begin{aligned} &\text{(a)} \quad \theta \leq 0 \quad \Rightarrow \quad f_\theta \in \mathcal{M}_q; \\ &\text{(b)} \quad \forall \theta \in B_N, \qquad \Phi_r[f_\theta] \geq \kappa_2 L h^\beta \Psi_r(\theta), \end{aligned}$$

where

$$\Psi_r(\theta) = \left(\frac{1}{N}\sum_{i=1}^{N}([\theta_i]_+)^r\right)^{1/r} \qquad (a_+ = \max[0, a]). \tag{21}$$

This is an immediate consequence of G.1 and G.2.

For $i = 1, \dots, N$, let

$$Y_i = Y_i^\theta = \frac{\sqrt{n}}{\|D(\cdot)\|_2\sqrt{h}} \int_{(i-1)h}^{ih} D\big(h^{-1}(t - t_i)\big)[dX_n^{f_\theta}(t) - L\kappa_0 h^\beta F(t)\,dt]. \tag{22}$$

From (18) and (15) it immediately follows that

$$Y_i = \alpha(N)\theta_i + \xi_i, \qquad i = 1, \dots, N, \tag{23}$$

where $\alpha(N)$ is given by (16b) and

$$\xi_i = \frac{1}{\|D(\cdot)\|_2\sqrt{h}} \int_{(i-1)h}^{ih} D\big(h^{-1}(t - t_i)\big)\,dW(t),$$

so that ξ_i are independent $\mathcal{N}(0, 1)$ random variables. It is straightforward to see that the set of statistics Y_i, $i = 1, \dots, N$, is sufficient for the parametric submodel (with $f \in \Sigma^N = \{f_\theta\}_{\theta \in B_N}$). Therefore, when restricting f to belong to Σ^N and setting $s_i = \alpha(N)\theta_i$, $i = 1, \dots, N$, the original "signal + white noise" model (1) becomes the "sequence space" model

$$Y_i = s_i + \xi_i, \qquad i = 1, \dots, N, \tag{24}$$

with $s = (s_1, \dots, s_N)$ from the cube $S_N = [-\alpha(N), \alpha(N)]^N$. With this transformation, the original testing problem (reduced to Σ^N) becomes the problem of testing, via observations (24) of an $s \in S_N$, the hypothesis

$$H_0\colon\ f^s \equiv f_{\alpha^{-1}(N)s} \in \mathcal{M}_q$$

versus the alternative

$$H_\varepsilon\colon\ \Phi_r[f^s] \ge \varepsilon.$$

Now consider the problem of testing, via observations (24) of an $s \in S_N$, the hypothesis

$$H_0^N\colon\ s \le 0$$

versus the alternative

$$H_\delta^N\colon\ \Psi_r(s) \ge \delta$$

with

$$\delta = \delta(\varepsilon) = \frac{\alpha(N)}{L\kappa_2 h^\beta}\varepsilon \quad \text{[see (20)]}.$$

We claim that if the pair of hypotheses H_0, H_ε is $\frac{1}{8}$-testable, then so is the pair H_0^N, $H_{\delta(\varepsilon)}^N$. Indeed, whenever $s \in S_n$ meets H_0^N, f^s meets H_0 [see (20a)], and whenever $s \in S_N$ meets $H_{\delta(\varepsilon)}^N$, f^s meets H_ε [see (20b)]. Thus, denoting by $\mathcal{R}_s(N)$ the infimum of those $\delta > 0$ for which the hypotheses H_0^N and H_δ^N are $\frac{1}{8}$-testable, we get the relation

$$\mathcal{R}^*(n) \geq L\kappa_2 h^\beta \alpha^{-1}(N)\mathcal{R}_s(N) = \kappa_3\sqrt{\frac{N}{n}}\mathcal{R}_s(N). \tag{25}$$

4.3. *In the sequence space.* We intend to establish the following.

PROPOSITION 4.1. *For all sufficiently large values of N one has*

$$\mathcal{R}_s(N) \geq \kappa_4(\ln N)^{-2}\alpha(N). \tag{26}$$

4.3.1. *From Proposition* 4.1 *to Theorem* 1. Postponing for the moment the proof of Proposition 4.1, let us derive from it the statement of Theorem 1. We have

$$\begin{aligned}\mathcal{R}^*(n) &\geq \kappa_3\sqrt{\frac{N}{n}}\mathcal{R}_s(N) \quad \text{[see (25)]}\\ &\geq \kappa_{4,1}\sqrt{\frac{N}{n}}\alpha(N)(\ln N)^{-2} \quad \text{[see (26)]}\\ &= \kappa_{4,2}L^{1/(2\beta+1)}(n\ln n)^{-\beta/(2\beta+1)}(\ln N)^{-2} \quad \text{[by (16)]}.\end{aligned}$$

Taking into account (16), we conclude that for all sufficiently large values of n one has

$$\mathcal{R}^*(n) \geq \kappa_{4,3}L^{1/(2\beta+1)}n^{-\beta/(2\beta+1)}(\ln n)^{-(5\beta+2)/(2\beta+1)},$$

as required in (7).

4.4. *Proof of Proposition* 4.1.

4.4.1. *Preliminaries.* Let $\mu_{N,0}$ and $\mu_{N,1}$ be two probability measures on the parameter set S_N, and let $P_{N,0}$ $P_{N,1}$ be the corresponding distributions of observations (24):

$$P_{N,j} = \mu_{N,j} * \mathcal{L}, \qquad j = 0, 1,$$

where $\mathcal{L}$ is the distribution of observation noises ξ in (24) (i.e., $\mathcal{L}$ is the N-dimensional Gaussian distribution with zero mean and unit covariance matrix). Also let

$$\mathcal{K}(P_{N,0}, P_{N,1}) = \int \log\left(\frac{dP_{N,1}}{dP_{N,0}}\right) dP_{N,1} \tag{27}$$

be the Kullback distance between $P_{N,0}$ and $P_{N,1}$. We need the following statement (which can be obtained from the Fano inequality; we, however, prefer to present a direct proof).

LEMMA 4.1. *One has*

$$R(N) \equiv \inf_T \big(P_{N,0}\{T=1\} + P_{N,1}\{T=0\}\big) \geq \exp\{-e^{-1} - \mathcal{K}(P_{N,0}, P_{N,1})\}, \tag{28}$$

the infimum being taken over all tests [*functions of observations* (24) *taking values* 0, 1].

PROOF. Consider a test $T(\cdot)$ for distinguishing between two hypotheses, H_0 and H_1, on the distribution of observations (24), saying, respectively, that the distribution is $P_{N,0}$ and $P_{N,1}$. Let $A = \{Y \in \mathbb{R}^N \colon T(Y) = 1\}$, $B = \mathbb{R}^N \backslash A$, $\phi = P_{N,0}(A)$, $\psi = P_{N,1}(B)$, $p = \phi + \psi$. We have

$$\begin{aligned}
\mathcal{K}(P_{N,0}, P_{N,1}) &= -\int_A \ln\left(\frac{dP_{N,0}}{dP_{N,1}}\right) dP_{N,1} - \int_B \ln\left(\frac{dP_{N,0}}{dP_{N,1}}\right) dP_{N,1} \\
&\geq -P_{N,1}(A) \ln\left(\frac{P_{N,0}(A)}{P_{N,1}(A)}\right) - P_{N,1}(B) \ln\left(\frac{P_{N,0}(B)}{P_{N,1}(B)}\right) \\
&\qquad\qquad \text{(Jensen's inequality)} \\
&= (1-\psi) \ln\left(\frac{1-\psi}{\phi}\right) + \psi \ln\left(\frac{\psi}{1-\phi}\right) \\
&= \underbrace{(1-\psi) \ln\left(\frac{1-\psi}{p-\psi}\right)}_{g_1(\psi)} + \underbrace{\psi \ln\left(\frac{\psi}{1-p+\psi}\right)}_{g_2(\psi)}.
\end{aligned}$$

We claim that

$$g_1(\psi) + g_2(\psi) \geq (p+1) \ln \frac{1}{p}, \qquad 0 \leq \psi \leq p, \tag{29}$$

whence, by the preceding computation,

$$\mathcal{K}(P_{N,0}, P_{N,1}) \geq (1+p) \ln \frac{1}{p}. \tag{30}$$

To justify (29), observe that for $0 \leq \psi \leq p$ it holds that

$$\begin{aligned}
g_1'(\psi) &= \ln\Bigg(\underbrace{\frac{p-\psi}{1-\psi}}_{h}\Bigg) + \underbrace{\frac{1-p}{p-\psi}}_{(1-h)/h} \\
&\geq \ln(1) = 0 \quad [\text{concavity of } \ln(\cdot)], \\
&\Downarrow \\
g_1(\psi) &\geq g(0) = \ln \frac{1}{p}, \qquad 0 \leq \psi \leq p,
\end{aligned}$$

$$
\begin{aligned}
g_2'(\psi) &= \ln\Bigl(\underbrace{\frac{\psi}{1-p+\psi}}_{1-h}\Bigr) + \underbrace{\frac{1-p}{1-p+\psi}}_{h} \\
&\le \ln(1) = 0 \quad [\text{concavity of } \ln(\cdot)], \\
&\Downarrow \\
g_2(\psi) &\ge g_2(p) = -p\ln\frac{1}{p}, \qquad 0 \le \psi \le p,
\end{aligned}
$$

and (29) follows.

Since $p\ln p \ge \exp\{-1\}$, (30) implies that $p \ge \exp\{-e^{-1} - \mathcal{K}(P_{N,0}, P_{N,1})\}$, as required in (28). $\square$

4.4.2. *Applying Lemma* 4.1. We are about to use Lemma 4.1 in the situation where the priors $\mu_{N,j}$ are of product structure

$$\mu_{N,j} = \mu_j^N, \qquad j = 0, 1;$$

here μ_0, μ_1 are probability measures on $[-\alpha(N), \alpha(N)]$. We assume that

$$
\text{(31)} \qquad
\begin{aligned}
&\text{(a)} \quad \operatorname{supp}\mu_0 \subset [-\alpha(N), 0], \\
&\text{(b)} \quad \int t_+ \mu_1(dt) = v_1 > 0.
\end{aligned}
$$

We claim that the following implication holds true:

$$
\text{(32)} \qquad
\left.
\begin{aligned}
&\text{(a)} \qquad \frac{4\alpha^{2r}(N)}{N v_1^{2r}} \le \frac{1}{4} \\
&\text{(b)} \quad \mathcal{K}(P_{N,0}, P_{N,1}) \le 0.3
\end{aligned}
\right\} \Rightarrow \mathcal{R}_s(N) \ge \tfrac{1}{2} v_1.
$$

Indeed, assume, contrary to what should be proved, that the premise in (32) is satisfied and that there exists a test T, based on observations (24), such that

$$
\text{(33)} \qquad \sup_{\substack{s\in S_N\\ s\le 0}} \operatorname*{Prob}_s\{T = 1\} + \sup_{\substack{s\in S_N\\ \Psi_r(s)>(1/2)v_1}} \operatorname*{Prob}_s\{T = 0\} \le \tfrac{1}{8},
$$

where Prob_s stands for the probability w.r.t. the distribution of observations (24) associated with s. Let us see how well the test T distinguishes between the distributions $P_{N,0}$ and $P_{N,1}$. In view of (31a), the prior $\mu_{N,0}$ "sits" on the set of nonpositive $s \in S_N$; applying (33), we therefore get

$$
\text{(34)} \qquad P_{N,0}\{T = 1\} = \int \operatorname*{Prob}_s\{T = 1\}\mu_{N,0}(ds) \le \tfrac{1}{8}.
$$

Now let $A = \{s \in S_N : \Psi_r(s) \le \frac{1}{2}v_1\}$. Applying (33), we get

$$
\text{(35)} \qquad P_{N,1}\{T = 0\} = \int \operatorname*{Prob}_s\{T = 0\}\mu_{N,1}(ds) \le \tfrac{1}{8} + \int_A \mu_{N,1}(ds) \equiv \tfrac{1}{8} + p.
$$

We are about to prove that $p \le \frac{1}{4}$. To this end, we first observe that

$$E_{\mu_{N,1}}\{\Psi_r^r(s)\} = \int (t_+)^r \mu_1(dt) \equiv v^r \ge v_1^r. \tag{36}$$

Setting $w = \frac{1}{2}v_1$, we have

$$\begin{aligned}
p &= \mu_{N,1}\left\{s : \left(\frac{1}{N}\sum_{i=1}^N([s_i]_+)^r\right)^{1/r} \le w\right\} \\
&= \mu_{N,1}\left\{s : \frac{1}{N}\sum_{i=1}^N([s_i]_+)^r \le w^r\right\} \\
&\le \mu_{N,1}\left\{s : \frac{1}{N}\sum_{i=1}^N([s_i]_+)^r \le \frac{1}{2^r}v^r\right\} \\
&\le \frac{N^{-1}\int[(t_+)^{2r} - v^{2r}]\mu_1(dt)}{(v^r - 2^{-r}v^r)^2} \quad \text{(the Chebyshev inequality)} \\
&\le \frac{4\alpha^{2r}(N)}{Nv^{2r}} \quad (\text{since supp } \mu_1 \subset [-\alpha(N), \alpha(N)]) \\
&\le \tfrac{1}{4} \quad [\text{by (32a) and in view of } v \ge v_1]
\end{aligned}$$

as claimed.

Since $p \le \frac{1}{4}$, (34) and (35) imply that

$$P_{N,0}\{T = 1\} + P_{N,1}\{T = 0\} \le \tfrac{1}{2},$$

whence, by (28),

$$\exp\{-e^{-1} - \mathcal{K}(P_{N,0}, P_{N,1})\} \le 0.5,$$

or, which is the same,

$$\mathcal{K}(P_{N,0}, P_{N,1}) \ge \ln 2 - e^{-1} = 0.325\ldots,$$

which contradicts (32b). Implication (32) is proved.

Note that the Kullback distance between the marginal measures $P_{N,0}$ and $P_{N,1}$, due to the product structure of model (24) and of the priors $\mu_{N,0}, \mu_{N,1}$, can be written as

$$\begin{aligned}
&\mathcal{K}(P_{N,0}, P_{N,1}) = N\mathcal{K}(p_{\mu_0}, p_{\mu_1}), \\
&\mathcal{K}(p_{\mu_0}, p_{\mu_1}) = \int \ln\left(\frac{p_{\mu_1}(y)}{p_{\mu_0}(y)}\right)p_{\mu_1}(y)\,dy,
\end{aligned} \tag{37}$$

where, for a finitely supported measure μ on the axis,

$$\begin{aligned}
p_\mu(y) &= \int \varphi(y - t)\mu(dt), \\
\varphi(y) &= \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{y^2}{2}\right\}
\end{aligned}$$

(φ is the standard Gaussian density on the axis). Thus, (32) can be rewritten as

$$\left.\begin{array}{ll}\text{(a)} & \dfrac{4\alpha^{2r}(N)}{N v_1^{2r}} \le \dfrac{1}{4} \\ \text{(b)} & N\mathcal{K}(p_{\mu_0}, p_{\mu_1}) \le 0.3\end{array}\right\} \Rightarrow \mathcal{R}_s(N) \ge \tfrac{1}{2}v_1. \tag{38}$$

4.4.3. *Specifying* μ_0, μ_1. It is time now to specify our choice of the measures μ_j, $j = 0, 1$. These measures will be "$\alpha(N)$-scalings" of probability measures ν_j on $[-1, 1]$:

$$\mu_j(A) = \nu_j(\alpha^{-1}(N)A) \qquad (\lambda A = \{t = \lambda a \mid a \in A\}). \tag{39}$$

The measures ν_j, $j = 0, 1$, are determined by two parameters: a positive integer m and a real $\sigma \in (0, 1]$; these parameters will be specified later.

Given m, σ, we define μ_j, $j = 0, 1$, as follows. Consider the Banach space $C[-1, 0]$ of continuous functions on $[-1, 0]$ equipped with the uniform norm, and let $\mathcal{P}_m$ be the subspace of this space comprising all polynomials of degree less than or equal to m. The value $p(\sigma)$ of a polynomial $p \in \mathcal{P}_m$ at the point σ is a linear functional on the finite-dimensional subspace $\mathcal{P}_m$ of $C[-1, 0]$; let ν_* be the norm of this linear functional on $\mathcal{P}_m$. By the Hahn–Banach theorem, we can extend the functional in question from $\mathcal{P}_m$ to the entire space $C[-1, 0]$, not increasing the norm of the functional. Taking into account that every continuous linear functional on $C[-1, 0]$ is an integral over a measure (not necessarily nonnegative) of bounded variation, we see that there exists a measure ν on $[-1, 0]$ such that

$$\begin{array}{ll}\text{(a)} & \displaystyle\int p(t)\nu(dt) = p(\sigma) \qquad \forall p \in \mathcal{P}_m, \\ \text{(b)} & \displaystyle\mathrm{Var}(\nu) \equiv \int |\nu(dt)| = \nu_*.\end{array} \tag{40}$$

The quantity ν_* can be computed explicitly. Indeed, by its origin, ν_* is the maximum of $p(\sigma)$, the maximum being taken over all polynomials $p(\cdot)$ with $\deg p \le m$ and $\max_{-1\le t\le 0} |p(t)| \le 1$. The corresponding extremal polynomial $\pi_m(t)$ is known (Markov's theorem); it is obtained from the Chebyshev polynomial $T_m(t) = \cos(m \arccos(t))$ by linear substitution of argument which maps the segment $[-1, 0]$ onto the segment $[-1, 1]$:

$$\pi_m(t) = T_m(2t + 1).$$

Thus, $\nu_* = \pi_m(\sigma) = T_m(1 + 2\sigma)$. Taking into account that $T_m(h) = \mathrm{ch}(m \operatorname{arccosh}(h))$ for $h > 1$, we get

$$\nu_* = \mathrm{ch}\big(m \operatorname{arccosh}(1 + 2\sigma)\big). \tag{41}$$

Now let $\nu = \nu_+ - \nu_-$ be the decomposition of ν into its positive and negative components, and let δ_σ be the unit mass placed at the point σ. We set

$$\nu_0 = \mathrm{Var}^{-1}(\nu_+)\nu_+, \qquad \nu_1 = [\mathrm{Var}(\nu_-) + 1]^{-1}[\nu_- + \delta_\sigma]. \tag{42}$$

By construction, $\nu_\pm$ are probability measures on $[-1, 1]$. We claim that the following hold:

$$
\begin{aligned}
&\text{(a)} \qquad \operatorname{supp}\nu_0 \subset [-1, 0];\\
&\text{(b)} \qquad \operatorname{supp}\nu_1 \subset [-1, 0] \cup \{\sigma\},\\
&\qquad\qquad \nu_1(\{\sigma\}) = \frac{2}{1 + \operatorname{ch}(m \operatorname{arccosh}(1 + 2\sigma))};\\
&\text{(c)} \quad \int t^i \nu_0(dt) = \int t^i \nu_1(dt), \qquad i = 0, 1, \dots, m.
\end{aligned}
\tag{43}
$$

Indeed, (a) is evident. To prove (b) and (c), observe first that (40a) applied with $p(t) \equiv 1$ implies that

$$
\operatorname{Var}(\nu_+) = \operatorname{Var}(\nu_-) + 1,
\tag{44}
$$

whence, in view of (40b),

$$
\operatorname{Var}(\nu_+) = \operatorname{Var}(\nu_-) + 1 = \frac{\nu_* + 1}{2} = \frac{\operatorname{ch}(m\ \operatorname{arccosh}(1 + 2\sigma)) + 1}{2};
$$

in particular, (43b) indeed takes place. Besides this, (44) ensures that

$$
\begin{aligned}
\int t^i [\nu_0(dt) - \nu_1(dt)] &= \frac{1}{\operatorname{Var}(\nu_+)} \int t^i [\nu_+(dt) - \nu_-(dt) - \delta_\sigma(dt)]\\
&= \frac{1}{\operatorname{Var}(\nu_+)} \left[\int t^i \nu(dt) - \sigma^i \right],
\end{aligned}
$$

and the latter quantity is 0 for $i \le m$ due to (40a). We have proved (43c).

Since the measures μ_j, $j = 0, 1$, are obtained from the measures ν_j by scaling (39), relation (43) implies that the following hold:

$$
\begin{aligned}
&\text{(a)} \qquad \operatorname{supp}\mu_0 \subset [-\alpha(N), 0];\\
&\text{(b)} \qquad \operatorname{supp}\mu_1 \subset [-\alpha(N), 0] \cup \{\sigma\alpha(N)\} \subset [-\alpha(N), \alpha(N)],\\
&\qquad \mu_1(\{\sigma\alpha(N)\}) = \frac{2}{1 + \operatorname{ch}(m \operatorname{arccosh}(1 + 2\sigma))};\\
&\text{(c)} \qquad \int t^i \mu_0(dt) = \int t^i \mu_1(dt), \qquad i = 0, 1, \dots, m.
\end{aligned}
\tag{45}
$$

4.4.4. *Bounding* $\mathcal{K}(p_{\mu_0}, p_{\mu_1})$. Now let us prove the following.

LEMMA 4.2. *Let* $\alpha \in (0, 1]$, *let* $m > 0$ *be an integer and let* ϕ, ψ *be two probability measures on* $[-1, 1]$ *such that*

$$
\int t^i \phi(dt) = \int t^i \psi(dt), \qquad i = 0, 1, \dots, m.
\tag{46}
$$

For a probability measure γ *on* $[-1, 1]$, *let*

$$p_{\gamma,\alpha}(y) = \int \varphi(y - \alpha t)\gamma(dt) = \varphi(y)\int \exp\left\{\alpha t y - \frac{\alpha^2 t^2}{2}\right\}\gamma(dt),$$

$$\left[\varphi(y) = \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{y^2}{2}\right\}\right].$$

Then for every $T \geq 2$ *and all* $\alpha \leq \frac{1}{10T}$ *one has*

$$\begin{aligned}\mathcal{K}(p_{\phi,\alpha}, p_{\psi,\alpha}) &\equiv \int \ln\bigl(p_{\psi,\alpha}(y)/p_{\phi,\alpha}(y)\bigr)p_{\psi,\alpha}(y)\,dy \\ &\leq \frac{3}{2}(10\alpha T)^{m+1} + 12\alpha\exp\left\{-\frac{(T-1)^2}{2}\right\}.\end{aligned} \tag{47}$$

PROOF.

Step 1. We have

$$\begin{aligned}\mathcal{K}(\alpha) &\equiv \mathcal{K}(p_{\phi,\alpha}, p_{\psi,\alpha}) = \int \ln\left(\frac{g_\psi(\alpha, y)}{g_\phi(\alpha, y)}\right)g_\psi(\alpha, y)\varphi(y)\,dy, \\ g_\gamma(\alpha, y) &= \int \exp\left\{\alpha t y - \frac{\alpha^2 t^2}{2}\right\}\gamma(dy).\end{aligned} \tag{48}$$

For every $T \geq 2$, we have

$$\begin{aligned}\mathcal{K}(\alpha) &= \mathcal{K}_T(\alpha) + H_T(\alpha), \\ \mathcal{K}_T(\alpha) &= \int_{-T}^{T} \ln\left(\frac{g_\psi(\alpha, y)}{g_\phi(\alpha, y)}\right)g_\psi(\alpha, y)\varphi(y)\,dy, \\ H_T(\alpha) &= \int_{|y|\geq T} \ln\left(\frac{g_\psi(\alpha, y)}{g_\phi(\alpha, y)}\right)g_\psi(\alpha, y)\varphi(y)\,dy.\end{aligned} \tag{49}$$

Step 2. We claim that the following hold:

(i) One has

$$H_T(\alpha) \leq 12\alpha\exp\left\{-\frac{(T-1)^2}{2}\right\}. \tag{50}$$

Indeed, whatever a probabilistic measure γ on $[-1, 1]$ is, we clearly have

$$\exp\left\{-\frac{\alpha^2}{2} - \alpha|y|\right\} \leq g_\gamma(\alpha, y) \leq \exp\{\alpha|y|\},$$

whence

$$
\begin{aligned}
H_T(\alpha) &\le \int_{|y|\ge T} \ln\left(\frac{\exp\{\alpha|y|\}}{\exp\{-\alpha^2/2-\alpha|y|\}}\right)\exp\{\alpha|y|\}\varphi(y)\,dy\\
&= 2\int_{y\ge T}\left(2\alpha y+\frac{\alpha^2}{2}\right)\exp\left\{\alpha y-\frac{y^2}{2}\right\}\frac{dy}{\sqrt{2\pi}}\\
&= \sqrt{\tfrac{2}{\pi}}\int_{z\ge T-\alpha}\left(2\alpha z+\frac{5\alpha^2}{2}\right)\exp\left\{-\frac{z^2}{2}\right\}\exp\left\{\frac{\alpha^2}{2}\right\}dz\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\text{(substitution } z=y-\alpha)\\
&\le 6\exp\left\{\frac{\alpha^2}{2}\right\}\alpha\int_{z\ge T-1} z\exp\left\{-\frac{z^2}{2}\right\}dz\\
&\qquad\qquad\left(\text{since } T-\alpha\ge T-1\ge 1\ge\alpha,\text{ whence } 2\alpha z+\frac{5\alpha^2}{2}\le 6\alpha z\right)\\
&\le 12\alpha\exp\left\{-\frac{(T-1)^2}{2}\right\}.
\end{aligned}
$$

(ii) The function $\mathcal{K}_T(\cdot)$ can be extended, as an analytic function, to the circle

$$
D_T=\left\{z\in\mathbb{C}:|z|\le d_t\equiv\frac{1}{10T}\right\}
$$

and

$$
z\in D_T\Rightarrow|\mathcal{K}_T(z)|\le\tfrac{3}{2}. \tag{51}
$$

Indeed, if γ is a probability measure on $[-1,1]$ and $|y|\le T$, then the function of α

$$
h_y(\alpha)=g_\gamma(\alpha,y)=\int\exp\left\{\alpha t y-\frac{\alpha^2t^2}{2}\right\}\gamma(dt)
$$

is analytic, and the modulus of its derivative in the circle $|\alpha|\le a\le 1$ does not exceed $(T+1)\exp\{aT+a^2/2\}$, whence

$$
|\alpha|\le a\le 1\quad\Rightarrow\quad|h_y(\alpha)-h_y(0)|=|h_y(\alpha)-1|\le a(T+1)\exp\left\{aT+\frac{a^2}{2}\right\}.
$$

It follows that

$$
\begin{aligned}
&\forall(y\in\mathbb{R}:|y|\le T,\alpha\in\mathbb{C}:|\alpha|\le d_T),\\
&|g_\gamma(\alpha,y)-1|\le\frac{T+1}{10T}\exp\{0.105\}\le\frac{1}{5}\exp\{0.105\}\le\frac{1}{4}.
\end{aligned}
\tag{52}
$$

As a result,

$$
\forall(y\in\mathbb{R}:|y|\le T,\alpha\in\mathbb{C}:|\alpha|\le d_T),\qquad\left|\frac{g_\psi(\alpha,y)}{g_\phi(\alpha,y)}-1\right|\le\frac{1+1/4}{1-1/4}-1=\frac{2}{3}.
$$

We see that the function $\ln(g_\psi(\alpha, y)/g_\phi(\alpha, y))$, regarded as a function of α, can be extended analytically from the real axis to the circle D_T on the complex plane, and the modulus of the extension in this circle does not exceed the quantity $\sum_{j=1}^{\infty} \frac{1}{j}(\frac{2}{3})^j = -\ln(1-\frac{2}{3}) = \ln 3$:

$$(53) \qquad \forall(y \in \mathbb{R}\colon |y| \le T, \alpha \in \mathbb{C}\colon |\alpha| \le d_T), \qquad \left|\ln\left(\frac{g_\psi(\alpha, y)}{g_\phi(\alpha, y)}\right)\right| \le \ln 3.$$

Combining (52) as applied to $\gamma = \psi$ with (53), we see that the function $\mathcal{K}_T(\alpha)$ indeed can be extended, as an analytic function of α, to the circle D_T, and the modulus of the extension in this circle does not exceed the quantity

$$\int_{|y| \le T} (\ln 3)\left(1 + \frac{1}{4}\right)\varphi(y)\,dy \le \frac{5\ln 3}{4} \le \frac{3}{2},$$

as claimed.

(iii) The function $\mathcal{K}_T(\alpha)$ has a zero of order at least $m+1$ at the point $\alpha = 0$. Indeed, let

$$f(\alpha, y) = g_\psi(\alpha, y) - g_\phi(\alpha, y) = \int \exp\left\{\alpha t y - \frac{\alpha^2 t^2}{2}\right\}[\psi(dt) - \phi(dt)].$$

For $\ell = 0, 1, \dots, m$ we have

$$\begin{aligned}\frac{\partial^\ell}{\partial\alpha^\ell}\bigg|_{\alpha=0} f(\alpha, y) &= \int \frac{\partial^\ell}{\partial\alpha^\ell}\bigg|_{\alpha=0} \exp\{\alpha t y - \alpha^2 t^2\}[\psi(dt) - \phi(dt)] \\ &= \int t^\ell f_\ell(y)[\psi(dt) - \phi(dt)] \\ &= 0 \quad \text{[see (46)].}\end{aligned}$$

It follows that there exist $C < \infty$, $c > 0$ such that

$$\forall(\alpha, y \in \mathbb{R}: |y| \le T, |\alpha| \le c), \qquad |f(\alpha, y)| \le C|\alpha|^{m+1}.$$

Since

$$\ln\left(\frac{g_\psi(\alpha, y)}{g_\phi(\alpha, y)}\right) = \ln\left(1 + \frac{f(\alpha, y)}{g_\phi(\alpha, y)}\right),$$

we conclude that there exist $C' < \infty$, $c' > 0$ such that

$$\forall(\alpha, y \in \mathbb{R}: |y| \le T, |\alpha| \le c'), \qquad \left|\ln\left(\frac{g_\psi(\alpha, y)}{g_\phi(\alpha, y)}\right)\right| \le C'|\alpha|^{m+1},$$

whence

$$|\mathcal{K}_T(\alpha)| \le \int_{|y| \le T} \left|\ln\left(\frac{g_\psi(\alpha, y)}{g_\phi(\alpha, y)}\right)\right| g_\psi(\alpha, y)\varphi(y)\,dy \le o(|\alpha^m|), \qquad \alpha \to 0,$$

as claimed.

Step 3. By Step 2(ii) and (iii), the function $\mathcal{K}_T$ in the circle D_T satisfies the bound

$$|\mathcal{K}_T(\alpha)| \le \frac{3}{2}\left(\frac{|\alpha|}{d_T}\right)^{m+1}. \tag{54}$$

Combining (54) and Step 2(i), we come to (47). □

4.5. *Concluding the proof of Proposition* 4.1. Now let us specify the parameters m, σ underlying the construction of the measures $\nu_j, \mu_j, j = 0, 1$, as

$$m = \rfloor \ln N \lfloor, \qquad \sigma = \frac{1}{4m^2}. \tag{55}$$

Step 1. Observe that the data $[\alpha = \alpha(N), \ m, \psi = \nu_1, \ \phi = \nu_0]$ satisfy the premise of Lemma 4.2, and that the functions $p_{\phi,\alpha}(\cdot)$, $p_{\psi,\alpha}(\cdot)$ associated with this data are $p_{\mu_0}(y)$, $p_{\mu_1}(y)$, respectively. Setting

$$T = \sqrt{2\ln N} + 1$$

and taking into account (17), we see that $10\alpha(N)T \le 0.2$, provided that N is large enough. Thus, we may use (47) to get the estimate

$$\begin{aligned} N\mathcal{K}(p_{\mu_0}, p_{\mu_1}) &\le \frac{3}{2}N\big(10\alpha(N)T\big)^{m+1} + 12N\alpha(N)\exp\left\{-\frac{(T-1)^2}{2}\right\} \\ &\le \frac{3}{2}N(0.2)^{m+1} + 12N\alpha(N)\exp\{-\ln N\} \\ &= \frac{3}{2}N(0.2)^{m+1} + 12\alpha(N), \end{aligned} \tag{56}$$

and the concluding quantity, in view of (55) and (17), is less than or equal to 0.3 for all sufficiently large values of N. Thus, (38a) is valid for all sufficiently large values of N.

Step 2. The measure μ_0 clearly satisfies (31a). In view of (45), μ_2 satisfies (31b) with

$$v_1 = \frac{2\sigma\alpha(N)}{1 + \mathrm{ch}(m \operatorname{arccosh}(1 + 2\sigma))}.$$

We have $\mathrm{ch}(2\sqrt{\sigma}) \ge 1 + \frac{1}{2}(2\sqrt{\sigma})^2 = 1 + 2\sigma$, whence $\operatorname{arccosh}(1 + 2\sigma) \le 2\sqrt{\sigma} = m^{-1}$. It follows that for all sufficiently large values of N it holds that

$$v_1 \ge \frac{2\sigma\alpha(N)}{1 + \mathrm{ch}(1)} \ge 0.75\sigma\alpha(N) \ge \kappa_5(\ln N)^{-2}\alpha(N). \tag{57}$$

In view of (17), (57) and (16) the condition in (38a) is satisfied for all sufficiently large values of N.

Step 3. We see that when n (or, which is the same, N) is large enough, the measures μ_j, $j = 0, 1$, satisfy the premise in (38), and the corresponding v_1 satisfies (57). Applying (38), we get the bound

$$\mathcal{R}_s(N) \geq \kappa_6 (\ln N)^{-2} \alpha(N),$$

as required in Proposition 4.1. $\square$

5. Proof of Theorem 2. We start with a simple technical lemma.

LEMMA 1. *Let*

$$T_N(z) = \sum_{k=0}^{N} \alpha_k \psi_k(z) \tag{58}$$

be the Fourier polynomial of order $N > 1$ of the function q. Let j be an integer. Then there exists C_0 such that, for any $-2 \leq z \leq 2$,

$$\left|q^{(j)}(z) - T_N^{(j)}(z)\right| \leq \frac{C_0(\pi/4)^j}{N^{r-j}(r-j)} \qquad \textit{for } 0 \leq j < r. \tag{59}$$

Further, if r is an integer, $|T_N^{(r)}(z)| \leq C_0(\pi/4)^r(1 + \ln N)$, and

$$\left|T_N^{(j)}(z)\right| \leq C_0(\pi/4)^j N^{j-r} \qquad \textit{for } j > r. \tag{60}$$

PROOF. It can be easily verified that the Fourier coefficients α_k of q satisfy $|\alpha_k| \leq C_0 k^{-(r+1)}$ for some $C_0 < \infty$. Then, for any $z \in [-2, 2]$,

$$\begin{aligned} \left|q^{(j)}(z) - T_N^{(j)}(z)\right| &\leq \sum_{k=N+1}^{\infty} \left(\frac{\pi k}{4}\right)^j |\alpha_k| \leq C_0 \left(\frac{\pi}{4}\right)^j \sum_{k=N+1}^{\infty} k^{j-r-1} \\ &= C_0 \left(\frac{\pi}{4}\right)^j N^{j-r}. \end{aligned}$$

On the other hand, for $j \geq r$,

$$\left|T_N^{(j)}(z)\right| \leq \sum_{k=0}^{\infty} \left(\frac{\pi k}{4}\right)^j |\alpha_k| \leq C_0 \left(\frac{\pi k}{4}\right)^j \sum_{k=0}^{\infty} k^{j-r-1}.$$

The latter sum can be bounded by $(1 + \ln N)$ for $j = r$ and by N^{j-r} for $j > r$. $\square$

LEMMA 2. *Let the polynomial $T^*_{N,\lambda}(\cdot)$ be defined by* (11). *Suppose that ξ is a $N(0, 1)$ random variable. Then $T^*_{N,\lambda}(z + \lambda\xi)$ is an unbiased estimate of $T_N(z)$, that is*,

$$ET^*_{N,\lambda}(z + \lambda\xi) = T_N(z), \tag{61}$$

and there exists C_1, which depends only on r, such that for $z \in [-1, 1]$ the variance

$$E\big(T^*_{N,\lambda}(z+\lambda\xi) - T_N(z)\big)^2 \tag{62}$$
$$\leq C_1\Big(\lambda^2 z^{2r-2}\mathbb{1}_{z>0} + \lambda^{2r}\ln^2 N + N^{-2r}\exp\Big(\frac{\pi^2\lambda^2 N^2}{16}\Big)\Big).$$

PROOF. Note that

$$\begin{aligned}
&E\phi_{2k,\lambda}(z+\lambda\xi)\\
&\quad= 2^{-1/2}(2\pi)^{-1/2}\exp\left(\frac{\pi^2k^2\lambda^2}{8}\right)\int_{-\infty}^{\infty}\cos\Big(\frac{\pi k(z+\lambda x)}{2}\Big)\exp\Big(-\frac{x^2}{2}\Big)\,dx\\
&\quad= 2^{-1/2}(2\pi)^{-1/2}\mathrm{Re}\Big(\int_{-\infty}^{\infty}\exp\Big(\frac{pi^2k^2\lambda^2}{8}+\frac{i\pi k(z+\lambda x)}{2}-\frac{x^2}{2}\Big)\,dx\Big)\\
&\quad= \mathrm{Re}\Big(2^{-1/2}\exp\Big(\frac{i\pi kz}{2}\Big)(2\pi)^{-1/2}\int_{-\infty}^{\infty}\exp\Big(-\Big(x-\frac{i\pi k\lambda}{2}\Big)^2\Big/2\Big)\,dx\Big)\\
&\quad= 2^{-1/2}\cos\Big(\frac{\pi kz}{2}\Big) = \psi_{2k}(z).
\end{aligned}$$

We get the equality $E\phi_{2k+1,\lambda}(z+\lambda\xi) = \psi_{2k}(z)$ in the same way. This implies (61) by construction of the polynomial $T^*_{N,\lambda}$ in (11).

The result of Theorem 1.1 of [16] states that, for $z \in [-2, 2]$,

$$E\big(T^*_{N,\lambda}(z+\lambda\xi) - T_N(z)\big)^2 = \sum_{j=1}^{N}\frac{\lambda^{2j}|T_N^{(j)}(z)|^2}{j!} \equiv \sum_{j=1}^{N}\frac{I_j}{j!}. \tag{63}$$

Let us estimate the terms in the right-hand side of (63). Note that if $g(t) = t^r\mathbb{1}_{t>0}$,

$$g^{(j)}(t) = r(r-1)\cdots(r-j+1)t^{r-j}\mathbb{1}_{t>0} \qquad \text{for } j < r.$$

Since the function $q(t)$ and its derivatives coincide with those of $g(t)$ on $[-1, 1]$, we have, by (59),

$$\begin{aligned}
I_j &= \lambda^{2j}|T^{(j)}N(z)|^2\\
&\leq 2\lambda^{2j}\Big[z^{2r-2j}\mathbb{1}_{z>0}\big(r(r-1)\cdots(r-j+1)\big)^2 + C_0^2\Big(\frac{\pi}{4}\Big)^{2j}\frac{N^{2(j-r)}}{(r-j)^2}\Big] \qquad (64)\\
&\leq a_j\lambda^{2j}z^{2r-2j}\mathbb{1}_{z>0} + C_2N^{-2r}\Big(\frac{\pi^2\lambda^2N^2}{16}\Big)^j
\end{aligned}$$

for $1 \leq j < r$. Here the coefficients a_j depend only on r. In the case when r is integer,

$$I_r \leq C_0^2\lambda^{2r}\Big(\frac{\pi}{4}\Big)^{2r}(1+\ln N)^2.$$

On the other hand, for $j > r$ we get, from (60),

$$I_j \le C_0^2 \lambda^{2j} \left(\frac{\pi}{4}\right)^{2j} N^{2(j-2)} = C_0^2 N^{-2r} \left(\frac{\pi^2 \lambda^2 N^2}{16}\right)^j. \tag{65}$$

The bounds in (64) and (65) provide us with the estimate

$$\begin{aligned}
\sum_{j=1}^{N} \frac{I_j}{j!} &\le \sum_{j=1}^{r-1} a_j \lambda^{2j} z^{2r-2j} \mathbb{1}_{z>0} + C_2 N^{-2r} \sum_{j=1}^{N} \left(\frac{\pi^2 \lambda^2 N^2}{16}\right)^j \Big/ j! \\
&\quad + C_0^2 \lambda^{2r} \left(\frac{\pi}{4}\right)^{2r} \frac{(1 + \ln N)^2}{r!} \quad \text{(if } r \text{ is an integer)} \\
&\le C_1 \left(\lambda^2 z^{2r-2} \mathbb{1}_{z>0} + \lambda^{2r} \ln^2 N + N^{-2r} \exp\left(\frac{\pi^2 \lambda^2 N^2}{16}\right)\right). \qquad \square
\end{aligned}$$

Let us denote $\Psi_r[f_h] = \int_0^1 f_{h,+}^r(t)\, dt$, $\gamma_n(t) = T^*_{N,\lambda_h}(\widehat{f}_h(t))$ and $\widehat{\Psi}_n = \int_0^1 \gamma_n(t)\, dt$, so that $\Psi_r[f_h] = (\Psi_r[f_h])^{1/r}$ and $\widehat{\Phi}_n = (\widehat{\Psi}_n)^{1/r}$.

LEMMA 3. *We have the estimate*

$$E|\widehat{\Psi}_n - E\widehat{\Psi}_n| \le C_3 t (N^{-1} \Phi_t^{r-1}[f_h] + N^{-r}).$$

PROOF. Since the kernel K is compactly supported we conclude from the definition of $\widehat{f}_h(t)$ that $\widehat{f}_h(t)$ and $\widehat{f}_h(t')$ are independent random variables when $|t' - t| > 2h$. Consider the partition of the interval [0, 1] into bins of length $2h$. For the sake of simplicity we suppose that $h^{-1} = 4m$ for some integer m, so that there are exactly $2m$ bins. Let us denote

$$\xi_j = \int_{4jh}^{2(2j+1)h} \big(\gamma_n(t) - E\gamma_n(t)\big)\, dt, \qquad \eta_j = \int_{2(2j+1)h}^{4(j+1)h} \big(\gamma_n(t) - E\gamma_n(t)\big)\, dt.$$

Then we have, for any p, $1 \le p \le 2$,

$$\begin{aligned}
[E|\widehat{\Psi}_n - E\widehat{\Psi}_n|^p]^{1/p} &= \left[E\left|\int_0^1 \big(\gamma_n(t) - E\gamma_n(t)\big)\, dt\right|^p\right]^{1/p} \\
&\le \left[E\left|\sum_{j=0}^{m-1} \xi_j\right|^p\right]^{1/p} + \left[E\left|\sum_{j=0}^{m-1} \eta_j\right|^p\right]^{1/p} \qquad (66)\\
&\le \left[\sum_{j=0}^{m-1} E|\xi_j|^p\right]^{1/p} + \left[\sum_{j=0}^{m-1} E|\eta_j|^p\right]^{1/p},
\end{aligned}$$

where the last bound results from the Marcinkiewicz–Zygmund inequality. By the Jensen inequality we have, from (62),

$$\begin{aligned}
E|\xi_j|^p &\le (2h)^{p-1}\int_{4jh}^{2(2j+1)h} E|\gamma_n(t)-E\gamma_n(t)|^p\,dt \\
&\le (2h)^{p-1}\int_{4jh}^{2(2j+1)h} \bigl(E\bigl(\gamma_n(t)-E\gamma_n(t)\bigr)^2\bigr)^{p/2} \\
&\le C_4 h^{p-1}\biggl(\lambda_h^p\int_{4jh}^{2(2j+1)h} f_{h,+}^{p(r-1)}+\lambda_h^{rp}(\ln N)^p \\
&\qquad\qquad + N^{-rp}\exp\biggl(\frac{p\pi^2N^2\lambda_h^2}{32}\biggr)\biggr).
\end{aligned}$$

The same bound holds for $E|\eta_j|^p$. Let us take $p=r/(r-1)$ for $r>2$ and $p=2$ for $1\le r\le 2$. When substituting these bounds into (66) we obtain

$$\begin{aligned}
E|\widehat{\Psi}_n-E\widehat{\Psi}_n| &\le [E|\widehat{\Psi}_n-E\widehat{\Psi}_n|^p]^{1/p} \\
&\le C_5 h^{(1/r)\wedge(1/2)}\biggl(\lambda_h\|f_{h,+}\|_r^{r-1}+\lambda_h^r(\ln N) \\
&\qquad\qquad + N^{-r}\exp\biggl(\frac{\pi^2N^2\lambda_h^2}{32}\biggr)\biggr).
\end{aligned}$$

Now recall that λ_h and N are chosen so that $\lambda_h^2N^2=\theta^2\|K\|_2^2\ln n$. For the values of θ which satisfy (10) we have $\theta^2\pi^2\|K\|_2^2/32\le\frac{1}{2r(2\beta+1)}$ for $r>2$ and $\theta^2\pi^2\|K\|_2^2/32\le\frac{1}{4(2\beta+1)}$ for $1\le r\le 2$. Thus

$$h^{(1/r)\wedge(1/2)}\exp\biggl(\frac{\pi^2N^2\lambda_h^2}{32}\biggr)\le n^{-(1/(2r(2\beta+1))\wedge 1/(4(2\beta+1)))}$$

[here we use the notation $a\wedge b$ for $\min(a,b)$]. Furthermore, $h^{1/r\wedge 1/2}\lambda_h^r\ln N\le N^{-r}$ and $h^{1/r\wedge 1/2}\lambda_h\le N^{-1}$ for n large enough. When summing up we obtain the bound

$$E|\widehat{\Psi}_n-E\widehat{\Psi}_n|\le C_3(N^{-r}+N^{-1}\|f_{h,+}\|_r^{r-1}). \qquad \square$$

We can now complete the proof of the theorem. Indeed,

$$\begin{aligned}
E|\widehat{\Phi}_n-\Phi_r[f]| &= E|\widehat{\Psi}_n^{1/r}-\Psi_r^{1/r}[f]| \\
&\le E|(\widehat{\Psi}_n)^{1/r}-(E\widehat{\Psi}_n)^{1/r}|+|(E\widehat{\Psi}_n)^{1/r}-\Psi_r^{1/r}[f_h]| \\
&\quad + |\Phi_r[f_h]-\Phi_r[f]| \\
&= \delta_1+\delta_2+\delta_3.
\end{aligned} \tag{67}$$

Recall that $\Phi_r(f)$ is Lipschitz continuous in the norm $\|\cdot\|_r$ with unit Lipschitz constant. This implies that

$$\delta_3 = |\Phi_r[f_h] - \Phi_r[f]| \leq \|f_h(t) - f(t)\|_r \leq Lh^\beta = L^{1/(2\beta+1)}(n \ln n)^{-\beta/(2\beta+1)}.$$

Further, the bound (59) for $j = 0$ gives

$$\delta_2 \leq |E\widehat{\Psi}_n - \Psi_r[f_h]|^{1/r} \leq \max_{z\in[-1,1]} |z^r \mathbb{1}_{z>0} - T_N(z)|^{1/r} \leq C_6 N^{-1}. \tag{68}$$

On the other hand,

$$\begin{aligned}\delta_1 &= E|\widehat{\Psi}_n^{1/r} - (E\widehat{\Psi}_n)^{1/r}| \\ &\leq (E|\widehat{\Psi}_n - E\widehat{\Psi}_n|)^{1/r} \mathbb{1}_{\Phi_r[f_h]\leq 2C_6N^{-1}} + \frac{E|\widehat{\Psi}_n - E\widehat{\Psi}_n|}{(E\widehat{\Psi}_n)^{(r-1)/r}} \mathbb{1}_{\Phi_r[f_h]>2C_6N^{-1}}.\end{aligned}$$

However, due to (68) we conclude that, for $\Phi_r[f_h] > 2C_6N^{-1}$,

$$(E\widehat{\Psi}_n)^{1/r} \geq \Phi_r[f_h] - C_6N^{-1} \geq \Phi_r[f_h]/2.$$

Now, by Lemma 3, δ_1 can be bounded as follows:

$$\delta_1 \leq C_7\biggl(N^{-1} + \frac{N^{-1}(\Phi_r[f_h])^{r-1} + N^{-r}}{(\Phi_r[f_h])^{r-1}} \mathbb{1}_{\Phi_r[f_h]>2C_6N^{-1}}\biggr) \leq C_8 N^{-1}.$$

When substituting the obtained bounds into (67) we obtain the statement of the theorem. $\square$

REFERENCES

[1] BICKEL, P. J. and RITOV, Y. (1988). Estimating integrated squared density derivatives: sharp best order of convergence estimates. *Sankhyā Ser. A* **50** 381–393.

[2] BIRGÉ, L. and MASSART, P. (1995). Estimation of integral functionals of a density. *Ann. Statist.* **23** 11–29.

[3] DIACK, C. A. T. and THOMAS-AGNAN, C. (1998). A nonparametric test of the nonconvexity of regression. *J. Nonparametr. Statist.* **9** 335–362.

[4] DONOHO, D. L. and LIU, R. C. (1991). Geometrizing rates of convergence, III. *Ann. Statist.* **19** 668–701.

[5] DONOHO, D. L. and NUSSBAUM, M. (1990). Minimax quadratic estimation of a quadratic functional. *J. Complexity* **6** 290–323.

[6] DÜMBGEN, L. and SPOKOINY, V. (2001). Multiscale testing of qualitative hypotheses. *Ann. Statist.* **29** 124–152.

[7] EFROMOVICH, S. and LOW, M. (1996). On Bickel and Ritov's conjecture about adaptive estimation of the integral of the square of density derivative. *Ann. Statist.* **24** 682–686.

[8] ERMAKOV, M. (1990). Minimax detection of a signal in Gaussian white noise. *Theory Probab. Appl.* **35** 667–679.

[9] FAN, J. (1991). On the estimation of quadratic functionals. *Ann. Statist.* **19** 1273–1294.

[10] HALL, P. and MARRON, J. S. (1987). Estimation of integrated squared density derivatives. *Statist. Probab. Lett.* **6** 109–115.

[11] IBRAGIMOV, I. A. and KHASMINSKI, R. Z. (1981). *Statistical Estimation*: *Asymptotic Theory*. Springer, Berlin.

[12] IBRAGIMOV, I. A. and KHASMINSKI, R. Z. (1980). Some estimation problems for stochastic differential equations. *Stochastic Differential Systems. Lecture Notes Control and Inform. Sci.* **25** 1–12. Springer, New York.

[13] IBRAGIMOV, I. A. and KHASMINSKI, R. Z. (1980). On the estimation of distribution density. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov (LOMI)* **98** 61–85.

[14] IBRAGIMOV, I. A. and KHASMINSKI, R. Z. (1987). Estimation of linear functionals in Gaussian noise. *Theory Probab. Appl.* **32** 30–39.

[15] IBRAGIMOV, I. A. and KHASMINSKI, R. Z. (1991). Asymptotically normal families of distributions and efficient estimation. *Ann. Statist.* **19** 1681–1724.

[16] IBRAGIMOV, I. A., NEMIROVSKI, A. and KHASMINSKI, R. Z. (1986). Some problems of nonparametric estimation in Gaussian white noise. *Theory Probab. Appl.* **31** 391–406.

[17] INGSTER, YU. I. (1982). Minimax nonparametric detection of signals in Gaussian white noise. *Problems Inform. Transmission* **18** 130–140.

[18] INGSTER, YU. I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives, I–III. *Math. Methods Statist.* **2** 85–114, 171–189, 249–268.

[19] INGSTER, YU. I. (2000). On testing a hypothesis which is close to a simple hypothesis. *Theory Prob. Appl.* **45** 310–323.

[20] INGSTER, YU. I. and SUSLINA, I. A. (2000). Minimax nonparametric hypothesis testing for ellipsoids and Besov bodies. *ESAIM Probab. Statist.* **4** 53–135.

[21] KERKYACHARIAN and PICARD, D. (1996). Estimating nonquadratic functionals of a density using Haar wavelets. *Ann. Statist.* **24** 485–507.

[22] KHASMINSKI, R. Z. and IBRAGIMOV, I. A. (1979). On the nonparametric estimation of functionals. In *Proceedings of the Second Prague Symposium on Asymptotic Statistics* (P. Mandl and M. Huskova, eds.) 41–51. North-Holland, Amsterdam.

[23] KOROSTELEV, A. P. (1990). On the accuracy of estimation of non-smooth functionals of regression. *Theory Probab. Appl.* **35** 784–787.

[24] KOROSTELEV, A. P. and TSYBAKOV, A. B. (1993). *Minimax Theory of Image Reconstruction*. *Lecture Notes in Statist.* **82**. Springer, New York.

[25] KOSHEVNIK, YU. and LEVIT, B. YA. (1976). On a nonparametric analogue of the information matrix. *Theory Probab. Appl.* **21** 738–753.

[26] LAURENT, B. (1996). Efficient estimation of integral functionals of a density. *Ann. Statist.* **24** 659–681.

[27] LEHMANN, E. L. (1959). *Testing Statistical Hypothesis*. Wiley, New York.

[28] LEPSKI, O. (1999). How to improve the accuracy of estimation. *Math. Methods Statist.* **8** 441–486.

[29] LEPSKI, O., NEMIROVSKI, A. and SPOKOINY, V. (1999). On estimation of the L_r norm of a regression function. *Probab. Theory Related Fields* **113** 221–253.

[30] LEPSKIJ, O. (1993). Estimation of the maximum of a nonparametric signal to within a constant. *Theory Probab. Appl.* **38** 152–158.

[31] LEVIT, B. YA. (1974). On optimality of some statistical estimates. In *Proceedings of the Prague Symposium on Asymptotic Statistics* (J. Hajek, ed.) **2** 215–238. Univ. Karlova, Prague.

[32] LEVIT, B. YA. (1975). Efficiency of a class of nonparametric estimates. *Theory Probab. Appl.* **20** 738–754.

[33] LEVIT, B. YA. (1978). Asymptotically efficient estimation of nonlinear functionals. *Problems Inform. Transmission* **14** 204–209.

[34] NEMIROVSKI, A. S., POLYAK, B. T. and TSYBAKOV, A. B. (1984). Signal processing by the nonparametric maximum-likelihood method. *Problems Inform. Transmission* **20** 177–192.

[35] RUDIN, W. (1966) *Real and Complex Analysis*. McGraw-Hill, New York.

[36] SPOKOINY, V. (1996). Adaptive hypothesis testing using wavelets. *Ann. Statist.* **24** 2477–2498.

[37] SPOKOINY, V. (1998). Adaptive and spatially adaptive testing of a nonparametric hypothesis. *Math. Methods Statist.* **7** 245–273.

INRIA RHÔNE-ALPES
655 AVENUE DE L'EUROPE
T 38330 MONTBONNOT
ST. MARTIN
FRANCE
E-MAIL: anatoli.iouditski@inriaalpes.fr

FACULTY OF INDUSTRIAL ENGINEERING
AND MANAGEMENT
TECHNION—ISRAEL INSTITUTE OF TECHNOLOGY
TECHNION CITY, HAIFA 32000
ISRAEL
E-MAIL: nemirovs@ie.technion.ac.il