Optimal Nonparametric Testing of Qualitative Hypotheses

Lutz Dümbgen and Vladimir G. Spokoiny

Medizinische Universität zu Lübeck
and
Weierstraß-Institut für Angewandte Analysis und Stochastik, Berlin

January 1998

Keywords and phrases: adaptivity, convexity, Lévy's modulus of continuity, monotonicity, positivity.

Abstract. Suppose one observes a process Y on the unit interval, where $dY = f + n^{-1/2}dW$ with an unknown function parameter f, given scale parameter $n \geq 1$ ("sample size") and standard Brownian motion W. We propose two classes of tests of qualitative nonparametric hypotheses about f such as monotonicity or convexity. These tests are asymptotically optimal and adaptive with respect to two different criteria. As a by-product we obtain an extension of Lévy's modulus of continuity of Brownian motion. It is of independent interest because of its potential applications to simultaneous confidence intervals in nonparametric curve estimation.

1 Introduction

Suppose that one observes a stochastic process Y on the unit interval $[0,1]$, where

$$Y(t) \;=\; \int_{[0,t]} f(x)\,dx + n^{-1/2}\,W(t).$$

Here f is an unknown function in $L^1[0,1]$, $n \geq 1$ is a given scale parameter, and W is standard Brownian motion. The present paper devises asymptotically optimal and adaptive tests of nonparametric hypotheses H_o about f as $n \to \infty$. Specifically we treat the following three hypotheses:

$$\begin{aligned} H_{\leq}: &\quad f \leq 0,\\ H_{\uparrow}: &\quad f \text{ is non-decreasing},\\ H_c: &\quad f \text{ is convex}. \end{aligned}$$

This is not a standard hypothesis testing problem because we do not assume any fixed or at least parametric structure of the model under the null hypothesis. In that sense we are speaking of testing a *qualitative* hypothesis. Such problems arise, for instance, in the statistical analysis of econometrical, medical or biological data.

Since the null is nonparametrically specified, it is natural to consider a nonparametric alternative. The theory of nonparametric testing of simple hypotheses is now well developed. We refer to the survey of Ingster (1993) where the reader can find more references. Under the nonparametric approach it is typically assumed that the alternative is smooth and deviates from the null hypothesis in some integral norm. In case of qualitative hypotheses such as monotonicity, the deviations of the alternative from the null are usually well localized. In this situation it is natural to measure the distance between the hypothesis and the alternative in supremum norm.

Below we consider two different approaches to define the alternative set. In the first case we consider

$$H_A(L, c) \ := \ \Big\{ g \in \mathcal{F}(L) : \Delta(g) \geq c \Big\}$$

with the Lipschitz function classes

$$\mathcal{F}(L) \ := \ \Big\{ g : |g(x) - g(y)| \leq L|x - y| \text{ for all } x, y \in [0,1] \Big\}$$

and the distance

$$\Delta(g) \ := \ \inf_{f_o \in H_o} \|g - f_o\|_\infty$$

from g to H_o in terms of supremum norm $\|\cdot\|_\infty$ on $[0,1]$.

The idea of the proposed testing procedure is to estimate nonparametrically the function f and then to use the distance $\Delta(\widehat{f})$ from the estimate $\widehat{f}$ to the set H_o as a test statistic. We apply kernel estimates $\widehat{f}_h$ where h is the bandwidth parameter.

It is well-known that the choice of the smoothing parameter h is crucial in nonparametric inference. Ingster (1993) showed that for testing simple hypotheses the proper choice of h is to be of order $L^{-2/3}(n^{-1}\log n)^{1/3}$ which allows one to detect an alternative f deviated from the null with the distance $\Delta(f)$ of order $(Ln^{-1}\log n)^{1/3}$. Lepski (1993) and Lepski and Tsybakov (1996) refined this result in showing that a proper choice of the kernel function and the bandwidth h leads to asymptotically optimal tests.

In practical applications the Lipschitz constant L is typically unknown. The problem of adaptive (data-driven) choice of a smoothing parameter for testing a simple hypothesis, where deviation from the null hypothesis is measured by some integral norm, was considered in Ledwina and Kallenberg (1995), Fan (1996), Spokoiny (1996), Ingster and Suslina (1997). The main message of Spokoiny (1996) is that the adaptive approach leads necessarily to suboptimal rates by a (log log)-factor. But here we consider supremum norm and reach a completely different conclusion: Adaptive testing is possible without any loss of efficiency. Combining kernel estimators with different bandwidths leads to a procedure which is asymptotically efficient in the minimax sense for arbitrary values of the unknown smoothness parameter L.

Our second type of alternative is described in terms of a test signal $\psi \in \mathcal{L}^\infty(\mathbf{R})$, where $\psi(x) = 0$ if $|x| > 1$. The idea is that f may be equal, or at least similar, to

$$g \ := \ f_o + \tilde{c}\psi\Big(\frac{\cdot - t}{h}\Big)$$

for some $f_o \in H_o$, $\tilde{c} > 0$, $h \in \,]0, 1/2]$ and $t \in [h, 1-h]$. The signal ψ is required to satisfy

$$(1.1) \qquad \int_{[0,1]} \psi\Big(\frac{x-t}{h}\Big) f_o(x)\, dx \ \le \ 0 \quad \text{for all } f_o \in H_o, h \in \,]0,1/2], t \in [h, 1-h].$$

The Cauchy-Schwarz inequality and (1.1) together imply that

$$\Psi_h(g) \ := \ \sup_{s \in [h, 1-h]} \frac{1}{h} \int_{[0,1]} \psi\Big(\frac{x-s}{h}\Big) g(x)\, dx \ \le \ \tilde{c} \int \psi(x)^2\, dx$$

with equality if $\int \psi((x-t)/h) f(x)\, dx = 0$. Thus we consider the alternative

$$\widetilde{H}_A(h, \tilde{c}) \ := \ \Big\{ g \in L^1[0,1] : \Psi_h(g) \ge \tilde{c} \int \psi(x)^2\, dx \Big\}.$$

This type of alternative assumes that a generally non-negative (or monotonous) function is contaminated at some place(s) by a signal with approximately known shape but location, scale and magnitude of this contamination are unknown. Again using kernel estimators with different bandwidths in a suitable way leads to asymptotically efficient tests in a minimax sense.

Section 2 contains the precise definition of the tests. Their asymptotic properties are presented in Section 3. These results depend on a new theorem about maxima of stochastic processes presented in Section 4. The latter can be viewed as an extension of Lévy's modulus of continuity of Brownian motion. It is of independent interest because of its potential applications to simultaneous confidence intervals in nonparametric curve estimation. Specific applications of such confidence intervals are introduced by Chaudhuri and Marron (1997) and Dümbgen (1998). All proofs are deferred to Section 6.

2 Definition of the tests

The kernel estimators use a fixed function k on $\mathbf{R}$ satisfying the following two requirements:

(2.1) $\quad k$ has bounded total variation $\mathrm{TV}(k)$;

(2.2) $\quad k(x) = 0$ if $|x| > 1$.

Then for $h \in\,]0, 1/2]$ and $t \in [h, 1-h]$ we define

$$k_{h,t}(x) \;:=\; \frac{1}{h} k\Big(\frac{x-t}{h}\Big)$$

and

$$\begin{aligned} \widehat{f}_h(t) &\;:=\; \int k_{h,t}(x)\, dY(x), \\ \bar{f}_h(t) &\;:=\; \int k_{h,t}(x) f(x)\, dx \;=\; \mathbb{E}\, \widehat{f}_h(t). \end{aligned}$$

Note that the distribution of the stochastic process

$$\Big(n^{1/2}(\widehat{f}_h - \bar{f}_h)(t)\Big)_{h\in\,]0,1/2],\ t\in[h,1-h]} \;=\; \Big(\int k_{h,t}(x)\, dW(x)\Big)_{h\in\,]0,1/2],\ t\in[h,1-h]}$$

does not depend on n or f_o. Here is a key result for this process.

Theorem 2.1 *Suppose that Conditions (2.1, 2.2) hold, and let $\tau^2 := \int k(x)^2\, dx$. Then*

$$\sup_{h\in\,]0,1/2],\ t\in[h,1-h]} D(h)^{-1}\Big((nh)^{1/2}|\widehat{f}_h(t) - \bar{f}_h(t)| - (2\tau^2 \log(1/h))^{1/2}\Big) \;<\; \infty$$

almost surely, where $D(h) := (\log(e/h))^{-1/2} \log\log(e^e/h)$.

2.1 Tests for $H_A(L, c)$

In connection with $H_A(L, c)$ we consider only the hypotheses $H_{\le}, H_{\uparrow}$ of non-positivity or monotonicity and utilize the triangular kernel

$$k(x) \;:=\; (1 - |x|)^+.$$

Here a^+ means $\max\{a, 0\}$ and we get $\int k(x)\, dx = 1$, $\tau^2 = \int k(x)^2 = 2/3$. The value $\widehat{f}_h$ is viewed as an estimator for f. The definition of $\widehat{f}_h(t)$ and $\bar{f}_h(t)$ is extended to $t \in [0, 1]$ via

$$k_{h,t}(x) \;:=\; \frac{1\{0 \le x \le 1\}}{c_{h,t}} k\Big(\frac{x-t}{h}\Big),$$

where

$$c_{h,t} \;:=\; \int 1\{0 \le x \le 1\} k\Big(\frac{x-t}{h}\Big)\, dx \;=\; h \int 1\Big\{y \le \frac{t \wedge (1-t)}{h}\Big\} k(y)\, dy \;\ge\; h/2,$$

so that $\int k_{h,t}(x)\, dx = 1$. Note that the following condition holds with $M = 4$.

(2.3) For arbitrary $h \in\,]0, 1/2]$ and $t \in [0, 1]$,

$$\mathrm{TV}(hk_{h,t}) \le M/h \quad \text{and} \quad \{x : k_{h,t}(x) \ne 0\} \subset [t-h, t+h] \cap [0, 1].$$

This condition implies a useful result for the boundary kernel estimators.

Theorem 2.2 *Suppose that Condition (2.3) holds. Then*

$$\sup_{h\in\,]0,1/2],\ t\in[0,h[\cup]1-h,1]} \frac{(nh)^{1/2}|\widehat{f}_h(t)-\bar{f}_h(t)|}{(\log\log(e^e/h))^{1/2}} \;<\; \infty$$

almost surely.

For fixed h a natural test statistic would be $\Delta(\widehat{f}_h)$. In order to combine $\Delta(\widehat{f}_h)$, $0 < h \le 1/2$, note that there are simple formulae for the distance $\Delta(g)$ from a function g to H_o, namely

$$\Delta(g) \;=\; \begin{cases} \sup\limits_{0\le t\le 1} g^+(t) & \text{if } H_o = H_{\le}, \\ \sup\limits_{0\le s<t\le 1} (g(s)-g(t))^+/2 & \text{if } H_o = H_{\uparrow}. \end{cases} \tag{2.4}$$

Thus for testing non-positivity we define the test statistic $T(n^{1/2}Y)$ by

$$T_{\le}(n^{1/2}Y) \;:=\; \sup_{h\in\,]0,1/2]}\ \sup_{0\le t\le 1} D(h)^{-1}\Big((nh)^{1/2}\widehat{f}_h(t) - (2\tau^2\log(1/h))^{1/2}\Big)$$

and for testing monotonicity by

$$T_{\uparrow}(n^{1/2}Y) \;:=\; \sup_{h\in\,]0,1/2]}\ \sup_{0\le s<t\le 1} D(h)^{-1}\Big((nh)^{1/2}(\widehat{f}_h(s)-\widehat{f}_h(t))/2 - (2\tau^2\log(1/h))^{1/2}\Big).$$

Note that in case of $f = 0$, the test statistic $T(n^{1/2}Y)$ equals $T(W)$ and is finite almost surely, according to Theorems 2.1 and 2.2. Moreover, if $f = f_o \in H_o$, then

$$T(n^{1/2}Y) \;=\; T(n^{1/2}f_o + W) \le\; T(W).$$

In case of a non-positive function f_o this is obvious, because $\bar{f}_h \le 0$. If $H_o = H_{\uparrow}$ and hence f_o is monotonously increasing, then the assertions follow from the fact that for $0 \le s < t \le 1$ one can write $\bar{f}_h(s) - \bar{f}_h(t) = \int \delta(x) f_o(x)\,dx$ with a function δ such that

$$\int \delta(x)\,dx \;=\; 0 \quad \text{and} \quad \delta(x) \begin{cases} \ge 0 \text{ if } x \le a, \\ \le 0 \text{ if } x \ge a, \end{cases}$$

where $0 < a < 1$. Hence $\bar{f}_h(s) - \bar{f}_h(t)$ equals $\int \delta(x)(f_o(x) - f_o(a))\,dx \le 0$.

Thus we reject the hypothesis H_o if $T(n^{1/2}Y)$ is larger than $d(\alpha)$, the $(1-\alpha)$-quantile of $\mathcal{L}(T(W))$. This test has level

$$\sup_{f_o\in H_o} \mathbb{P}_{f_o}\Big\{T(n^{1/2}Y) > d(\alpha)\Big\} \;=\; \mathbb{P}\Big\{T(W) > d(\alpha)\Big\} \;\le\; \alpha,$$

where $\mathbb{P}_{f_o}$ denotes probability in case of $f = f_o$.

2.2 Tests for $\widetilde{H}_A(h, \tilde{c})$

Here we take $k := \psi$, assuming that Conditions (2.1, 2.2) hold. Then $\sup_{t\in[h,1-h]} \widehat{f}_h(t)$ estimates $\Psi_h(f)$, and we define

$$\widetilde{T}(n^{1/2}Y) \ := \ \sup_{h\in\,]0,1/2],\ t\in[h,1-h]} D(h)^{-1}\Big((nh)^{1/2}\widehat{f}_h(t) - (2\tau^2\log(1/h))^{1/2}\Big).$$

Under Condition (1.1),

$$\widetilde{T}(n^{1/2}Y) \ = \ \widetilde{T}(n^{1/2}f + W) \begin{cases} \leq \ \widetilde{T}(W) & \text{if } f \in H_o, \\ = \ \widetilde{T}(W) & \begin{cases} \text{if } f = 0, \\ \text{if } f \text{ is constant and } H_o = H_\uparrow, \\ \text{if } f \text{ is affine linear and } H_o = H_c. \end{cases} \end{cases}$$

Moreover, $\widetilde{T}(W) < \infty$ almost surely, by Theorem 2.1. Now we reject the hypothesis H_o if $\widetilde{T}(n^{1/2}Y)$ is larger than $\tilde{d}(\alpha)$, the $(1-\alpha)$-quantile of $\mathcal{L}(\widetilde{T}(W))$. Again this test has level α.

In case of $H_o = H_\leq$, Condition (1.1) is satisfied if, and only if, ψ is nonnegative almost everywhere. Two specific examples are

$$\psi_1(x) \ := \ (1-|x|)^+ \quad \text{and} \quad \psi_2(x) \ := \ (1-x^2)^+.$$

If $\psi = \psi_1$, then the resulting test statistic $\widetilde{T}(n^{1/2}Y)$ is just the test statistic $T_\leq(n^{1/2}Y)$ introduced previously.

For $H_o = H_\uparrow$ one may take any odd function ψ such that $\psi \geq 0$ on $[0,1]$. Specific examples are

$$\psi_j(2x+1) - \psi_j(2x-1) \quad \text{for } j = 1, 2.$$

In case of $H_o = H_c$ one may take any even function ψ such that $\psi \geq 0$ on $[0,1/2]$, $\psi \leq 0$ on $[1/2,1]$ and $\int \psi(x)\,dx = 0$. Specific examples are

$$\psi_j(2x) - \psi_j(4x+3)/2 - \psi_j(4x-3)/2 \quad \text{for } j = 1, 2.$$

3 Optimality of the tests

Let $\phi_n(Y)$ be some test, that is, ϕ_n is a function of the observations Y with values in $[0,1]$ and $\phi_n(Y)$ means the probability of rejecting the null hypothesis given Y. We say that the test ϕ_n is of level α with a prescribed $\alpha \in]0,1[$ if

$$\sup_{f_o\in H_o} \mathbb{E}_{f_o}\,\phi_n(Y) \ \leq \ \alpha.$$

The power of this test at a point g from the alternative set H_A is $\mathbb{E}_g\, \phi_n(Y)$. We measure the quality of this test by the minimum of the corresponding power function on the alternative set. More precisely, for the alternative set of the form $H_A(L,c)$ and for each test $\phi_n(Y)$ of level α we characterize its quality by the minimal value $c = c(\phi_n)$ such that

$$\inf_{g \in H_A(L,c)} \mathbb{E}_g\, \phi_n(Y) \;\ge\; \beta$$

for some prescribed $\beta \in]\alpha, 1[$. The next theorem entails that our tests are asymptotically optimal in the sense of minimizing $c(\phi_n)$ over all tests of level α.

Theorem 3.1 *Let H_o be $H_{\le}$ or $H_{\uparrow}$, and let*

$$c_n \;=\; c_n(\lambda) \;:=\; \lambda\Big(L\frac{\log n}{n}\Big)^{1/3}$$

for some $\lambda > 0$.

(a) *If $\lambda > 1$, then*

$$\lim_{n\to\infty} \inf_{g \in H_A(L,c_n)} \mathbb{P}_g\Big\{T(n^{1/2}Y) > d(\alpha)\Big\} \;=\; 1.$$

(b) *Suppose that $\lambda < 1$, and let $(\phi_n)_{n\ge 1}$ be any family of tests on $\mathcal{C}[0,1]$ of level α. Then*

$$\limsup_{n\to\infty} \inf_{g \in H_A(L,c_n)} \mathbb{E}_g\, \phi_n(Y) \;\le\; \alpha.$$

We see that if $c_n = c_n(\lambda)$ with $\lambda < 1$, then any sequence of tests ϕ_n has trivial minimal power α asymptotically. This means that the minimal distance between the null hypothesis and the alternative allowing consistent testing is not less than $(Ln^{-1}\log n)^{1/3}$.

On the other hand, for the case $c_n = c_n(\lambda)$ with $\lambda > 1$, our tests based on $T(n^{1/2}Y)$ have power one. Following Ingster (1993) the sequence $c_n^* = (Ln^{-1}\log n)^{1/3}$ is called the *optimal rate of testing* and the tests $\phi_n^* = \mathbf{1}(T(n^{1/2}Y) > d(\alpha))$ asymptotically *sharp-optimal.*

Next we consider the alternatives $\widetilde{H}_A(h,\widetilde{c})$.

Theorem 3.2 *For some fixed $\lambda > 0$ define*

$$\widetilde{c}_h = \widetilde{c}_h(\lambda) \;:=\; \lambda\Big(\frac{2\log(1/h)}{\tau^2 h}\Big)^{1/2}$$

The following conclusions hold uniformly in $n \geq 1$:

(a) *If $\lambda > 1$, then*

$$\lim_{h \downarrow 0} \inf_{g \in \widetilde{H}_A(h, n^{-1/2}\widetilde{c}_h)} \mathbb{P}_g\Big\{\widetilde{T}(n^{1/2}Y) > \widetilde{d}(\alpha)\Big\} \ = \ 1.$$

(b) *Suppose that $\lambda < 1$, and let $(\phi_h)_{h \in \,]0,1/2]}$ be any family of tests on $\mathcal{C}[0,1]$. Then*

$$\limsup_{h \downarrow 0} \Big(\inf_{g \in \widetilde{H}_A(h, n^{-1/2}\widetilde{c}_h)} \mathbb{E}_g\, \phi_h(n^{1/2}Y) - \mathbb{E}_0\, \phi_h(n^{1/2}Y) \Big) \ \leq \ 0.$$

Some additional remarks on the tests.

One can easily verify that the preceding optimality results remain valid if the test statistics $T(n^{1/2}Y)$ and $\widetilde{T}(n^{1/2}Y)$ are replaced with the simpler quantities

$$\sup_{h \in \,]0,1/2]} \Big((nh)^{1/2} \Delta(\widehat{f}_h) - (2\tau^2 \log(1/h))^{1/2} \Big)$$

and

$$\sup_{h \in \,]0,1/2],\ t \in [h, 1-h]} \Big((nh)^{1/2} \widehat{f}_h(t) - (2\tau^2 \log(1/h))^{1/2} \Big),$$

respectively. The denominator $D(h) = (\log(e/h))^{-1/2} \log\log(e^e/h)$ puts more weight on smaller scales h.

In both cases Theorem 3.1 remains valid if replaces the whole interval $]0,1/2]$ with some finite subset $\mathcal{H}_n = \{h_{n1}, h_{n2}, \ldots h_{nk(n)}\}$ with $0 =: h_{n0} < h_{n1} < h_{n2} < \cdots < h_{nk(n)}$ provided that

$$(\log(n)/n)^{-1/3} h_{nk(n)} \ \to \ \infty \quad \text{and} \quad \min_{1 \leq \ell \leq k(n)} (\log(n)/n)^{-1/3} (h_{n\ell} - h_{n,\ell-1}) \ \to \ 0.$$

4 An extension of Lévy's modulus of continuity

It is a well-known result of Lévy that

$$\lim_{\delta \downarrow 0} \sup_{s,t \in [0,1] : t-s=\delta} \frac{W(t) - W(s)}{(2\delta \log(e/\delta))^{1/2}} \ = \ 1 \quad \text{almost surely}$$

(cf. Shorack and Wellner 1986, Theorem 14.1.1). This implies that the test statistic

$$\check{S}(W) \ := \ \sup_{0 \leq s < t \leq 1} \frac{(W(t) - W(s))^2}{2(t-s)\log(e/(t-s))}$$

is finite almost surely. However, for statistical purposes $\check{S}(W)$ is suboptimal. The reason is that the distribution of

$$\sup_{s,t\in[0,1]:t-s=\delta} \frac{(W(t)-W(s))^2}{2\delta\log(e/\delta)}$$

becomes degenerate at one as $\delta \downarrow 0$, whereas some quantiles of $\check{S}(W)$ are greater than one because of larger values of δ. Now we propose a better normalization of W's increments: It follows from Theorem 4.1 below that

$$S(W) \ := \ \sup_{0\le s<t\le 1} \Bigl(\frac{(W(t)-W(s))^2}{t-s} - 2\log\Bigl(\frac{1}{t-s}\Bigr)\Bigr) \Big/ \log\log\Bigl(\frac{e^e}{t-s}\Bigr)$$

is finite almost surely. Note that

$$\frac{(W(t)-W(s))^2}{t-s} \ \le \ 2\log\Bigl(\frac{1}{t-s}\Bigr) + S(W)\log\log\Bigl(\frac{e^e}{t-s}\Bigr) \quad \text{for } 0\le s<t\le 1,$$

and the first summand on the right hand side dominates the second one as $t-s\downarrow 0$. Finiteness of $S(W)$ is related to, but not a consequence of, well-known results on the limiting distribution of

$$S_\delta(W) \ := \ \sup_{s,t\in[0,1]:t-s\ge\delta} (t-s)^{-1/2}(W(t)-W(s))$$

(suitably normalized) as $\delta \downarrow 0$.

Theorem 4.1 *Let X be a stochastic process on a subset Π of $]0,1]\times[0,1]$ such that the following inequalities hold for certain constants $K, L, M > 0$ and arbitrary $(h,t),(h',t')\in\Pi$, $\eta\ge 0$:*

$$\text{(4.1)} \qquad \mathbb{P}\Bigl\{|X(h,t)| > h^{1/2}\eta\Bigr\} \ \le \ K\exp(-\eta^2/2),$$

$$\text{(4.2)} \qquad \mathbb{P}\Bigl\{|X(h,t)-X(h',t')| > (|h-h'|+|t-t'|)^{1/2}\,\eta\Bigr\} \ \le \ L\exp\Bigl(-M\eta^2\Bigr).$$

If we define

$$S(X) \ := \ \sup_{(h,t)\in\Pi} \frac{X(h,t)^2/h - 2\log(1/h)}{\log\log(e^e/h)},$$

then $S(X)<\infty$ almost surely. More precisely, $\mathbb{P}\{S(X)>R\}\le \pi(R\,|\,K,L,M)$ for some universal function $\pi(\cdot\,|\,K,L,M)$ such that $\lim_{R\to\infty}\pi(R\,|\,K,L,M)=0$.

Remark. By definition, $X(h,t)^2/h \le 2\log(1/h) + S(X)\log\log(e^e/h)$ for arbitrary $(h,t)\in\Pi$. Since $(A+B)^{1/2}\le A^{1/2}+B^{1/2}$ and $(A+B)^{1/2}\le A^{1/2}+A^{-1/2}B/2$ for arbitrary positive numbers A, B, one can easily show that Theorem 4.1 implies finiteness of

$$\sup_{(h,t)\in\Pi} D(h)^{-1}\Big(h^{-1/2}|X(h,t)| - (2\log(1/h))^{1/2}\Big)$$

almost surely, where $D(h) = (\log(e/h))^{-1/2}\log\log(e^e/h)$.

As mentioned above, Theorem 4.1 applies to increments of Brownian motion, if we define $X(h,t) := W(t+h/2) - W(t-h/2)$. More generally, combined with the previous remark it implies Theorem 2.1 in Section 2, as shown in Section 6.

5 Some further developments

5.1 Other nonparametric models

In this paper we restrict ourselves to the ideal "signal + white noise" model which can be viewed as a prototype for more realistic statistical models involving regression functions or distribution densities. The results by Brown and Low (1996), and Nussbaum (1996) and Grama and Nussbaum (1997) on asymptotic equivalence of these models can be helpful in this context.

It is also worth mentioning that our procedures are formulated in terms of kernel smoothers and thus apply directly to Gaussian regression models with an equispaced or regular design. One possible modification in order to treat regression models with arbitrary error distributions uses simultaneous rank tests. This will be the subject of another paper.

5.2 Other type of smoothness constraints

We assume that the underlying function f belongs to a Lipschitz function class $\mathcal{F}(L)$. More generally the case of a Hölder class $\mathcal{F}_\beta(L)$ can be considered. We refer to Lepski and Tsybakov (1996) where sharp-optimal tests under Hölder constraints are described. The procedure is similar to the case of Lipschitz classes and it is also based on kernel smoothers but they apply special kernels arising from an optimal recovery problem. The problem of adaptive testing when both β and L are unknown seems to be more involved.

However, it follows from Ingster (1986) that the our tests with the triangular kernel are *rate optimal* over arbitrary Hölder classes $\mathcal{F}_\beta(L)$.

6 Proofs

Proof of Equality (2.4). In case of $H_o = H_\leq$ the assertion is easily verified. As for $H_\uparrow$, let $\Delta_o = \sup_{0\leq s<t\leq 1}(g(s)-g(t))/2$ be the asserted expression for $\Delta(g)$. Let $0 \leq s < t \leq 1$ be such that $\delta := (g(s) - g(t))/2 > 0$. Then for any $f \in H_\uparrow$ either $f(s) \leq g(s) - \delta$ or $f(t) \geq f(s) > g(s) - \delta = g(t) + \delta$. Thus $\|g - f\|_\infty \geq \delta$, whence $\Delta(g) \geq \Delta_o$. On the other hand, if $\Delta_o < \infty$, then

$$f(t) \ := \ \sup_{s\in[0,t]} g(s) - \Delta_o$$

defines a function in $H_\uparrow$ such that $f(t) \geq g(t) - \Delta_o$ and

$$f(t) \ \leq \ g(t) + \sup_{s\in[0,t]} (g(s) - g(t)) - \Delta_o \ \leq \ g(t) + \Delta_o. \qquad \square$$

Now we prove Theorems 3.1 and 3.2. When proving Theorem 3.1 we only consider the hypothesis $H_o = H_\uparrow$, because the arguments for $H_o = H_\leq$ are similar or even simpler. The main arguments for the latter case are also contained in the proof of Theorem 3.2.

Proof of Theorem 3.1 (a). First we note that

$$\operatorname{Var} \widehat{f}_h(t) := \mathbb{E}\Big[\widehat{f}_h(t) - \bar{f}_h(t)\Big]^2 \ = \ n^{-1}\int k_{h,t}(x)^2\, dx.$$

An important property of the kernel functions $k_{h,t}$ is that

$$h\int k_{h,t}(x)^2\, dx \ \in \ [\tau^2, 4\tau^2] \quad \text{for all } h \in \,]0,1/2] \text{ and } t \in [0,1].$$

Indeed

$$\begin{aligned} h\int k_{h,t}(x)^2\, dx \ &= \ \int 1\Big\{y \leq \frac{t\wedge(1-t)}{h}\Big\}k(y)^2\, dy \Big/ \Big(\int 1\Big\{y \leq \frac{t\wedge(1-t)}{h}\Big\}k(y)\, dy\Big)^2 \\ &= \ (\tau^2 - \gamma^3/3)/(1-\gamma^2/2)^2 \\ &= \ \tau^2(1-\gamma^3/2)/(1-\gamma^2/2)^2 \\ &\in \ [\tau^2, 4\tau^2], \end{aligned}$$

where $\gamma := \Big(1 - (t\wedge(1-t))/h\Big)^+ \in [0,1]$. This implies that for all $0 \leq s < t \leq 1$,

$$\tau^2_{h,s,t} := \operatorname{Var}\Big(2^{-1}(nh)^{1/2}(\widehat{f}_h(s) - \widehat{f}_h(t))\Big) \ \leq \ 4\tau^2.$$

Now let $f \in H_A(L,c)$ with $c/L \le 1/2$. Let $0 \le s < t \le 1$ such that $f(s) - f(t) \ge 2c$. With $\gamma := (f(s) + f(t))/2$ and $h := c/L$,

$$\begin{aligned}
\bar{f}_h(s) - \bar{f}_h(t) &= \int (k_{h,s} - k_{h,t})(x) f(x)\, dx \\
&= \int (k_{h,s} - k_{h,t})(x)(f(x) - \gamma)\, dx \\
&\ge \int \Big(k_{h,s}(x)(c - L|x-s|) - k_{h,t}(x)(-c + L|x-t|) \Big)\, dx \\
&= ch \int \Big(k_{h,s}(x)^2 + k_{h,t}(x)^2 \Big)\, dx \\
&\ge 2c\tau^2
\end{aligned}$$

where $\tau^2 = \int k(x)^2\, dx = 2/3$. Next $T(n^{1/2}Y) > d(\alpha)$ entails that

$$\begin{aligned}
2^{-1}(nh)^{1/2}(\widehat{f}_h(s) - \widehat{f}_h(t)) &> (2\tau^2 \log(1/h))^{1/2} + d(\alpha) \log\log(e^e/h)(\log(e/h))^{-1/2} \\
&= (2\tau^2 \log(1/h))^{1/2}(1 + r_\alpha(h))
\end{aligned}$$

where $\lim_{h\to 0} r_\alpha(h) = 0$. Hence $\mathbb{P}\Big\{T(n^{1/2}Y) > d(\alpha)\Big\}$ is not smaller than

$$\begin{aligned}
&\mathbb{P}\Big\{2^{-1}(nh)^{1/2}(\widehat{f}_h(s) - \widehat{f}_h(t)) > (2\tau^2 \log(1/h))^{1/2}(1 + r_\alpha(h))\Big\} \\
&\ge \mathbb{P}\Big\{2^{-1}(nh)^{1/2}\Big((\widehat{f}_h - \bar{f}_h)(s) - (\widehat{f}_h - \bar{f}_h)(t)\Big) \\
&\qquad\qquad > (2\tau^2 \log(1/h))^{1/2}(1 + r_\alpha(h)) - (nh)^{1/2} c\tau^2\Big\} \\
&= \mathbb{P}\Big\{2^{-1}(nh)^{1/2}\Big((\widehat{f}_h - \bar{f}_h)(s) - (\widehat{f}_h - \bar{f}_h)(t)\Big) \\
&\qquad\qquad > \tau^2(3\log(L/c))^{1/2}(1 + r_\alpha(c/L)) - \tau^2(nc^3/L)^{1/2}\Big\} \\
&= \Phi\Big(\tau^2 \tau_{h,s,t}^{-1}\Big((nc^3/L)^{1/2} - (2\log(L/c))^{1/2}(1 + r_\alpha(c/L))\Big)\Big),
\end{aligned}$$

where Φ is the standard Gaussian distribution function. If $c = c_n = \lambda(L\log(n)/n)^{1/3}$ with $\lambda > 1$, then

$$(nc^3/L)^{1/2} - (3\log(L/c))^{1/2}(1 + r_\alpha(c/L)) = \log(n)^{1/2}\Big(\lambda^{3/2} - 1 + o(1)\Big),$$

whence

$$\Phi\Big(\tau^2 \tau_{h,s,t}^{-1}\Big((nc^3/L)^{1/2} - (3\log(L/c))^{1/2}(1 + r_\alpha(c/L))\Big)\Big) \to 1,$$

uniformly in $0 \le s < t \le 1$. □

Proof of Theorem 3.2 (a). A simple rescaling argument shows that it suffices to consider the case $n = 1$.

Suppose that $f \in \widetilde{H}_A(h, \widetilde{c})$ i.e. $\sup_{t \in [h,1-h]} \bar{f}_h(t) \geq \tilde{c}\tau^2$. For $0 < \kappa < 1$ let $t \in [h, 1-h]$ such that $\bar{f}_h(t) > \kappa\widetilde{c}\tau^2$. Then $\mathbb{P}\Big\{\widetilde{T}(Y) > \widetilde{d}(\alpha)\Big\}$ is not smaller than

$$\begin{aligned}
&\mathbb{P}\Big\{(nh)^{1/2}\widehat{f}_h(t) > (2\tau^2\log(1/h))^{1/2}(1+\widetilde{r}_\alpha(h))\Big\} \\
&\geq \mathbb{P}\Big\{(nh)^{1/2}(\widehat{f}_h - \bar{f}_h)(t) > (2\tau^2\log(1/h))^{1/2} - \kappa\widetilde{c}\tau^2(nh)^{1/2}\Big\} \\
&= \Phi\Big(\tau^{-1}\Big(\kappa\widetilde{c}\tau^2(nh)^{1/2} - (2\tau^2\log(1/h))^{1/2}(1+\widetilde{r}_\alpha(h))\Big)\Big),
\end{aligned}$$

where $\lim_{h\to 0}\widetilde{r}_\alpha(h) = 0$. If $\widetilde{c} = \widetilde{c}_h = \lambda\tau^{-1}(2(nh)^{-1}\log(1/h))^{1/2}$ and $\kappa\lambda > 1$, then

$$\begin{aligned}
\kappa\widetilde{c}\tau^2(nh)^{1/2} - (2\tau^2\log(1/h))^{1/2}(1+\widetilde{r}_\alpha(h)) &= (2\tau^2\log(1/h))^{1/2}(\kappa\lambda - 1 - \widetilde{r}_\alpha(h)) \\
&\to \infty \quad \text{as } h \downarrow 0. \qquad \Box
\end{aligned}$$

The proofs of Theorems 3.1 (b) and 3.2 (b) rely on the following result (cf. Ingster 1993, Lepski and Tsybakov 1996).

Lemma 6.1 *Let $\Gamma_1, \Gamma_2, \Gamma_3, \ldots$ be independent random variables with standard Gaussian distribution. Then*

$$\lim_{m\to\infty} \mathbb{E}\Big|\frac{1}{m}\sum_{i=1}^m \exp(w_m\Gamma_i - w_m^2/2) - 1\Big| = 0,$$

if $w_m = (2\log m - \kappa_m)^{1/2}$ with $\lim_{m\to\infty}(\log m)^{-1/2}\kappa_m = \infty$.

For the reader's convenience a proof is given here.

Proof of Lemma 6.1. Let $Z_m := \exp(w_m\Gamma_1 - w_m^2/2)$. Since $\mathbb{E}\, Z_m = 1$, the assertion follows from the weak law of large numbers for triangular arrays, provided that

$$\lim_{m\to\infty} \mathbb{E}\,|Z_m - 1|\{|Z_m - 1| \geq \epsilon m\} = 0 \quad \text{for any } \epsilon > 0.$$

But for $m \geq 1/\epsilon$, the expectation of $|Z_m - 1|\{|Z_m - 1| \geq \epsilon m\}$ is not greater than

$$\begin{aligned}
\mathbb{E}\, Z_m\{Z_m \geq \epsilon m\} &\leq \mathbb{E}\, Z_m^{1+\delta}(\epsilon m)^{-\delta} \quad \text{(for any } \delta > 0), \\
&= \exp\Big(\delta(1+\delta)w_m^2/2 - \delta\log(\epsilon m)\Big) \\
&= \exp\Big(\delta^2 w_m^2/2 - \delta(\log(\epsilon m) - w_m^2/2)\Big) \\
&= \exp\Big(-\frac{(\log(\epsilon m) - w_m^2/2)^2}{2w_m^2}\Big) \quad (\text{if } \delta = w_m^{-2}(\log(\epsilon m) - w_m^2/2)) \\
&\leq \exp\Big(-\frac{(\kappa_m/2 + \log\epsilon)^2}{4\log m}\Big) \\
&\to 0 \quad \text{as } m \to \infty. \qquad \Box
\end{aligned}$$

Proof of Theorem 3.1 (b). Let $c = c_n = \lambda(L\log(n)/n)^{1/3}$ with $\lambda < 1$, and $h := c/L$. Now define $m := \lfloor 1/(2h) \rfloor$, and for $1 \le j \le m$ let

$$\begin{aligned} t_j &:= (2j-1)h, \\ g_j &:= chk_{h,t_j}, \\ \Gamma_j &:= (nh/\tau^2)^{1/2}\widehat{f}_h(t_j). \end{aligned}$$

Then $\Gamma_1, \Gamma_2, \ldots, \Gamma_m$ are independent Gaussian random variables with $\mathrm{Var}(\Gamma_j) = 1$ and

$$\begin{aligned} \mathbb{E}_{g_j}\,\Gamma_\ell &= \quad \text{if } j \neq \ell, \\ \mathbb{E}_{g_j}\,\Gamma_j &= (c^2\tau^2 nh)^{1/2} \;=\; \lambda^{3/2}((2/3)\log n)^{1/2} \;=\; \gamma_n(2\log m)^{1/2} \;=:\; w_n, \end{aligned}$$

where $\gamma_n \to \lambda^{3/2} < 1$. All functions $g_j - g_\ell$, $1 \le j < \ell \le m$, belong to $H_A(L,c)$, and

$$\frac{d\,\mathbb{P}_{g_j - g_\ell}}{d\,\mathbb{P}_0} \;=\; Z_{jn}\bar{Z}_{\ell n}$$

with $Z_{jn} := \exp(w_n\Gamma_j - w_n^2/2)$ and $\bar{Z}_{\ell n} := \exp(-w_n\Gamma_\ell - w_n^2/2)$. Defining the sums $S_n := 2m^{-1}\sum_{j \le m/2}(Z_{jn} - 1)$ and $\bar{S}_n := 2m^{-1}\sum_{\ell > m/2}(\bar{Z}_{\ell n} - 1)$ it follows from Lemma 6.1 that $\mathbb{E}_0\,|S_n| + \mathbb{E}_0\,|\bar{S}_n| \to 0$. Thus for arbitrary tests ϕ_n on $\mathcal{C}[0,1]$,

$$\begin{aligned} &\inf_{g \in H_A(L,c)} \mathbb{E}_g\,\phi_n(Y) - \mathbb{E}_0\,\phi_n(Y) \\ &\le \inf_{j \le m/2, \ell > m/2} \mathbb{E}_{g_j - g_\ell}\,\phi_n(Y) - \mathbb{E}_0\,\phi_n(Y) \\ &= \inf_{j \le m/2, \ell > m/2} \mathbb{E}_0(Z_{jn}\bar{Z}_{\ell n} - 1)\phi_n(Y) \\ &\le \mathbb{E}_0\,\frac{4}{m^2} \sum_{j \le m/2, \ell > m/2} (Z_{jn}\bar{Z}_{\ell n} - 1)\phi_n(Y) \\ &= \mathbb{E}_0(S_n\bar{S}_n + S_n + \bar{S}_n)\phi_n(Y) \\ &\le \mathbb{E}_0\,|S_n||\bar{S}_n| + \mathbb{E}_0\,|S_n| + \mathbb{E}_0\,|\bar{S}_n| \\ &= \mathbb{E}_0\,|S_n|\,\mathbb{E}_0\,|\bar{S}_n| + \mathbb{E}_0\,|S_n| + \mathbb{E}_0\,|\bar{S}_n| \\ &= o(1). \qquad \Box \end{aligned}$$

Proof of Theorem 3.2 (b). Again one may assume that $n = 1$. For $h \in]0, 1/2]$ let $\tilde{c} = \tilde{c}_h = \lambda\tau^{-1}(2(nh)^{-1}\log(1/h))^{1/2}$ with $\lambda < 1$. Define m, t_j, g_j and Γ_j as in the proof of Theorem 3.1 (b) with $\tilde{c}$ in place of c. Again $\Gamma_1, \Gamma_2, \ldots, \Gamma_m$ are independent Gaussian random variables with $\mathrm{Var}(\Gamma_j) = 1$ and

$$\mathbb{E}_{g_j}\,\Gamma_j \;=\; \tilde{c}\tau(nh)^{1/2} \;=\; \lambda(2\log(1/h))^{1/2} \;=\; \lambda_h(2\log m)^{1/2} \;=:\; w_h,$$

where $\lim_{h\downarrow 0}\lambda_h = \lambda$. All functions $g_1, g_2, \ldots, g_m$ belong to $\widetilde{H}_A(L, \widetilde{c})$, and

$$\frac{d\,\mathbb{P}_{g_j}}{d\,\mathbb{P}_0} \;=\; Z_{h,j} := \exp(w_h\Gamma_j - w_h^2/2).$$

Defining $S_h := m^{-1}\sum_{j=1}^m (Z_{h,j} - 1)$ it follows from Lemma 6.1 that $\mathbb{E}_0\,|S_h| \to 0$ as $h \downarrow 0$. Thus for arbitrary tests ϕ_h on $\mathcal{C}[0,1]$,

$$\begin{aligned}
&\inf_{g\in \widetilde{H}_A(h,\widetilde{c}_h)} \mathbb{E}_g\,\phi_h(Y) - \mathbb{E}_0\,\phi_h(Y) \\
&\le \inf_{1\le j\le m} \mathbb{E}_{g_j}\,\phi_h(Y) - \mathbb{E}_0\,\phi_h(Y) \;\le\; \mathbb{E}_0\,S_h\phi_h(Y) \;\le\; \mathbb{E}_0\,|S_h| \;\to\; 0 \quad \text{as } h\downarrow 0. \qquad \square
\end{aligned}$$

Proof of Theorem 4.1. We consider the metric

$$\rho((h,t),(h',t')) \;:=\; (|h-h'| + |t-t'|)^{1/2}$$

on Π. For $u > 0$ let Π_u be a maximal subset of Π such that

$$\rho(a,b) \;>\; u \quad \text{for different } a,b \in \Pi_u.$$

One can easily verify that $]0,1] \times [0,1]$ can be covered by at most $(1+2u^{-2})^2$ squares with ρ-diameter at most u. Each such square contains at most one point of Π_u, whence

$$\#\Pi_u \;\le\; (1+2u^{-2})^2 \;\le\; 1 + 8u^{-4}.$$

Hence, defining

$$\omega(X,\delta) \;:=\; \sup_{a,b\in\Pi:\rho(a,b)\le\delta} |X(a) - X(b)|,$$

it follows from (4.2), Theorem 2.2.4 of van der Vaart and Wellner (1996) and elementary calculations that

$$\text{(6.1)} \qquad \mathbb{P}\Big\{\omega(X,\kappa^{1/2}) > \eta\Big\} \;\le\; C\exp\Big(-\frac{\eta^2}{C\kappa\log(e/\kappa)}\Big) \quad \text{for all } \kappa \in\,]0,1], \eta > 0.$$

Here and throughout the sequel C denotes a generic positive constant depending only on K, L, M. Its value may differ from place to place.

For $\delta \in\,]0,1]$ let $\Pi(\delta) := \{(h,t) \in \Pi : h \le \delta\}$. Now fix some $D \ge 2$. For each $(h,t) \in \Pi(\delta) \setminus \Pi(\delta/2)$ there exists a point $(h',t') \in \Pi_{(\delta/D)^{1/2}}$ such that $\rho((h,t),(h',t'))^2 \le \delta/D$. For $0 < \lambda < 1$ and $\eta \ge 1$, the inequality

$$X(h,t)^2 \;>\; 2h(\log(1/h) + \eta)$$

implies that

$$
\begin{aligned}
X(h',t')^2 &> 2h(\log(1/h)+\eta)(1-\lambda)^2 \\
&= 2h'(h'/h)^{-1}(\log(1/\delta)+\eta)(1-\lambda)^2 \\
&\geq 2h'(1+(\delta/D)/h)^{-1}(\log(1/\delta)+\eta)(1-\lambda)^2 \\
&\geq 2h'(1+2/D)^{-1}(\log(1/\delta)+\eta)(1-\lambda)^2 \\
&\geq 2h'(\log(1/\delta)+\eta)(1-2/D)(1-\lambda)^2,
\end{aligned}
$$

or

$$
\begin{aligned}
\omega\Big(X,(\delta/D)^{1/2}\Big)^2 &> 2h(\log(1/h)+\eta)\lambda^2 \\
&> \delta(\log(1/\delta)+\eta)\lambda^2.
\end{aligned}
$$

But $h' \leq h+\delta/D \leq (1+1/D)\delta$, and covering $]0,(1+1/D)\delta]\times[0,1]$ with suitable squares reveals that

$$
\#\Big\{(h',t')\in\Pi_{(\delta/D)^{1/2}} : h' \leq (1+1/D)\delta\Big\} \;\leq\; (2D+3)(1+2D/\delta).
$$

Consequently, it follows from (4.1) and (6.1) that

$$
\begin{aligned}
&\mathbb{P}\Big\{X(h,t)^2 > 2h(\log(1/h)+\eta) \text{ for some } (h,t)\in\Pi(\delta)\setminus\Pi(\delta/2)\Big\} \\
&\quad\leq (2D+3)(1+2D/\delta)\,K\exp\Big(-(\log(1/\delta)+\eta)(1-2/D)(1-\lambda)^2\Big) \\
&\qquad + C\exp\Big(-\frac{\delta(\log(1/\delta)+\eta)\lambda^2}{C(\delta/D)\log(eD/\delta)}\Big) \\
&\quad\leq C\exp\Big(2\log D+\log(1/\delta)-(\log(1/\delta)+\eta)(1-2/D)(1-\lambda)^2\Big) \\
&\qquad + C\exp\Big(-\frac{D(\log(1/\delta)+\eta)\lambda^2}{C\log(eD/\delta)}\Big) \\
&\quad\leq C\exp\Big(2\log D+\log(1/\delta)-(\log(1/\delta)+\eta)(1-2/D)(1-2\lambda)\Big) \\
&\qquad + C\exp\Big(-\frac{D\lambda^2}{C\Big(1+\log(D)/\log(e/\delta)\Big)}\Big) \\
&\quad\leq C\exp\Big(2\log D+\log(1/\delta)-(\log(1/\delta)+\eta)(1-2/D)(1-2\lambda)\Big) + C\exp\Big(-\frac{D\lambda^2}{C}\Big).
\end{aligned}
$$

Now we take

$$
\begin{aligned}
\eta = \eta_\delta &:= R\log\log(e^e/\delta) \quad \text{with } R \geq 1, \\
\lambda = \lambda_\delta &:= \frac{\eta/4}{\log(1/\delta)+\eta} \;\geq\; \frac{1}{4\log(e/\delta)}.
\end{aligned}
$$

Then

$$
\begin{aligned}
& C\exp\Big(2\log D + \log(1/\delta) - (\log(1/\delta)+\eta)(1-2/D)(1-2\lambda)\Big) + C\exp\Big(-\frac{D\lambda^2}{C}\Big) \\
\le\ & C\exp\Big(2\log D - \eta/2 + (2\log(1/\delta)+\eta)/D\Big) + C\exp\Big(-\frac{D}{C\log(e/\delta)^2}\Big).
\end{aligned}
$$

Letting $D = R\log(e/\delta)^2 \log\log(e^e/\delta) = \log(e/\delta)^2\eta$ and $R \ge 2$, the latter bound is easily shown to be not greater than

$$
C\exp\Big(-(R/C - C)\log\log(e^e/\delta)\Big).
$$

Now we apply this bound to $\delta = 2^{-k}$, $k \ge 0$. This yields

$$
\begin{aligned}
& \mathbb{P}\Big\{X(h,t)^2/h > 2\log(1/h) + R\log\log(e^e/h) \text{ for some } (h,t) \in \Pi\Big\} \\
\le\ & \sum_{k=0}^{\infty} \mathbb{P}\Big\{X(h,t)^2/h > 2\log(1/h) + R\log\log(e^e/2^{-k}) \\
& \qquad\qquad \text{for some } (h,t) \in \Pi(2^{-k}) \setminus \Pi(2^{-k-1})\Big\} \\
\le\ & C\sum_{k=0}^{\infty} \exp\Big(-(R/C - C)\log\log(e^e 2^k)\Big) \\
=\ & C\sum_{k=0}^{\infty} (e + k\log 2)^{-(R/C-C)} \\
\to\ & 0 \quad \text{as } R \to \infty. \qquad \Box
\end{aligned}
$$

Proof of Theorem 2.1. Without loss of generality let $\tau^2 = 1$. With $X(h,t) := h\int k_{h,t}(x)\,dW(x)$ it suffices to show that Conditions (4.1) and (4.2) of Theorem 4.1 are satisfied. It is well known that X is a centered Gaussian process with

$$
\mathrm{Var}(X(h,t)) \;=\; h^2\int k_{h,t}(x)^2\,dx \;=\; h\tau^2 \;=\; h.
$$

Thus Condition (4.1) holds with $K = 1$. As for Condition (4.2), one can write

$$
k(x) \;=\; \int 1\{y \le x\} f(y)\,P(dy)
$$

for all but countably many $x \in \mathbf{R}$, where P is some probability measure on $[-1/2, 1/2]$ and $f \in \mathcal{L}^1(P)$ with $|f| \equiv \mathrm{TV}(k)$. Thus the Cauchy-Schwarz inequality yields

$$
\begin{aligned}
\mathrm{Var}(X(h,t) - X(h',t')) \;&=\; \int \Big(k\Big(\frac{u-t}{h}\Big) - k\Big(\frac{u-t'}{h'}\Big)\Big)^2\,du \\
&=\; \int \Big(\int \Big(1\Big\{y \le \frac{u-t}{h}\Big\} - 1\Big\{y \le \frac{u-t'}{h'}\Big\}\Big) f(y)\,P(dy)\Big)^2\,du
\end{aligned}
$$

$$
\begin{aligned}
&\leq \mathrm{TV}(k)^2 \int\int \Big| 1\Big\{y \leq \frac{u-t}{h}\Big\} - 1\Big\{y \leq \frac{u-t'}{h'}\Big\}\Big| \, P(dy)\, du \\
&= \mathrm{TV}(k)^2 \int\int \Big| 1\{u \geq t + hy\} - 1\{u \geq t' + h'y\}\Big| \, du \, P(dy) \\
&\leq \mathrm{TV}(k)^2 \int \Big(|t-t'| + |y||h-h'|\Big) P(dy) \\
&\leq \mathrm{TV}(k)^2 \Big(|t-t'| + |h-h'|\Big).
\end{aligned}
$$

Thus Condition (4.2) holds with $L = 1$ and $M = \mathrm{TV}(k)^{-2}/2$. □

Proof of Theorem 2.2. By assumption, for $h \in]0, 1/2]$ and $t \in [0, h[\cup]1-h, 1]$ there exist a probability measure $P_{h,t}$ on $[t-h, t+h] \cap [0,1]$ and a measurable function $b_{h,t}$ with $|b_{h,t}| \leq M$ such that

$$
h \int k_{h,t}(x)\, dW(x) = \begin{cases} \int_{[0,2h]} W(s) b_{h,t}(s)\, P_{h,t}(ds) & \text{if } t < h, \\ \int_{[1-2h,h]} (W(1) - W(s)) b_{h,t}(s)\, P_{h,t}(ds) & \text{if } t > 1-h, \end{cases}
$$

whence

$$
\begin{aligned}
&\sup_{t \in [0,h[\cup]1-h,1]} (nh)^{1/2} |\widehat{f}_h(t) - \bar{f}_h(t)| \\
&\leq M\Big(\sup_{s \in [0,2h]} h^{-1/2} |W(s)| \vee \sup_{t \in [1-2h,1]} h^{-1/2} |W(1) - W(s)| \Big).
\end{aligned}
$$

Now the assertion follows from the Law of the Iterated Logarithm for Brownian motion. □

References

L.D. Brown and M.G. Low (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24**, 2384–2398.

Chaudhuri, P. and J.S. Marron (1997). SiZer for exploration of Structures in curves. Preprint

Dümbgen, L. (1998). New goodness-of-fit tests and their application to nonparametric confidence sets. to appear in *Ann. Statist.* **26**

Fan, J. (1996) Test of significance based on wavelet thresholding and Neyman's truncation. *J. Amer. Statist. Assos.* **91**, 674–688.

Grama, I. and Nussbaum, M. (1997) Asymptotic equivalence for nonparametric generalized linear models. Preprint 289. Weierstrass-Institute, Berlin.

Yu.I. Ingster (1986). Minimax Testing of Nonparametric Hypothesis on a Distribution Density in L_p-metrics. *Theory Probab. Appl.* **32**, 333–337.

Ingster, Y.I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives, I-III. *Math. Methods Statist.* **2**; 85-114, 171-189, 249-268.

Ingster, Y.I. and Suslina, I. (1997). Minimax nonparametric hypothesis testing for ellipsoids and Besov bodies. Report 12. Weierstrass-Institute, Berlin.

Ledwina, T. and Kallenberg, W.C.M. (1995) Consistency and Monte Carlo simulation of a data driven version of smooth goodness-of-fit tests. *Ann. Statist.* **23**, 1594–1608.

Kallenberg W.C.M. and Ledwina, T. (1995) On data driven Neyman's tests. *Probab. Math. Statist.* **15**, 409–426.

Lepski, O.V. and A.B. Tsybakov (1996). Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. Preprint. Humboldt University, Berlin.

Nussbaum, M. (1996) Asymptotic equivalence of density estimation and white noise. *Ann. Statist.* **24**, 2399–2430.

Shorack, G.R. and J.A. Wellner (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

Spokoiny, V. (1996) Adaptive hypothesis testing using wavelets. *Ann. Statist.* **24**, 2477–2498.

van der Vaart, A.W. and J.A. Wellner (1996). *Weak Convergence and Empirical Processes with Applications to Statistics.* Springer, New York.

Lutz Dümbgen

Institut für Mathematik

Medizinische Universität zu Lübeck

Wallstraße 40

D-23560 Lübeck

Germany

E-mail: `duembgen@math.mu-luebeck.de`

Vladimir G. Spokoiny

Weierstraß-Institut für Angewandte Analysis und Stochastik

Mohrenstraße 39

D-10117 Berlin

Germany

E-mail: `spokoiny@wias-berlin.de`