The Annals of Statistics
1998, Vol. 26, No. 1, 288–314

NEW GOODNESS-OF-FIT TESTS AND THEIR APPLICATION TO NONPARAMETRIC CONFIDENCE SETS[1]

BY LUTZ DÜMBGEN

Universität Heidelberg and Medizinische Universität zu Lübeck

Suppose one observes a process V on the unit interval, where $dV = f_o + dW$ with an unknown parameter $f_o \in L_1[0,1]$ and standard Brownian motion W. We propose a particular test of one-point hypotheses about f_o which is based on suitably standardized increments of V. This test is shown to have desirable consistency properties if, for instance, f_o is restricted to various Hölder classes of functions. The test is mimicked in the context of nonparametric density estimation, nonparametric regression and interval-censored data. Under shape restrictions on the parameter, such as monotonicity or convexity, we obtain confidence sets for f_o adapting to its unknown smoothness.

1. Introduction. Suppose one observes a stochastic process $V_n = F_n + n^{-1/2}W$ on $[0,1]$, where F_n is an unknown parameter in $\mathscr{C}[0,1]$ with $F_n(0) = 0$, W is standard Brownian motion on $[0,1]$, and $n > 1$ is a known scale parameter. Estimation within this model is closely related to estimation of regression functions or densities based on samples of size n; see Brown and Low (1996) and Nussbaum (1996). Let $C_n(V_n, \alpha)$ be a confidence set for F_n with coverage probability $1 - \alpha \in]0,1[$. Given a model $\mathscr{M} \subset \mathscr{C}[0,1]$ for F_n and any function ϕ on $\mathscr{M}$, the set $\phi(C_n(V_n, \alpha) \cap \mathscr{M})$ is obviously a $(1-\alpha)$-confidence set for $\phi(F_n)$. Numerous applications of this type are described, for instance, by Donoho (1988), Davies (1995) and Hengartner and Stark (1995). The first two authors investigate sets $C_n(V_n, \alpha)$ based on standard goodness-of-fit tests such as the Kolmogorov–Smirnov test. In the context of density estimation, Hengartner and Stark (1995) utilize a special test criterion which may, but need not, give optimal confidence bands. The present paper introduces a new type of goodness-of-fit test such that the resulting confidence sets $\phi(C_n(V_n, \alpha) \cap \mathscr{M})$ have optimal size in terms of rates of convergence simultaneously for various classes $\mathscr{M}$ and functionals ϕ.

Suppose we want to test the hypothesis "$F_n = 0$" versus "$F_n \neq 0$." For fixed numbers $0 \le s < t \le 1$ and $r \in \mathbf{R} \setminus \{0\}$ consider the special alternative "$F_n(\cdot) = \pm r \,\mathrm{Leb}([s,t] \cap [0, \cdot])$." Then an optimal test rejects for large values of $nV_n(s,t)^2/(t-s)$; generally $h(s,t)$ stands for the increment $h(t) - h(s)$ of a function h on the line. Now we combine these special test statistics. Suppose that the triplet (r, s, t) above has an improper prior distribution with Lebesgue

Received May 1996; revised October 1997.

[1]Research supported in part by European Union Human Capital and Mobility Program ERB CHRX-CT 940693.

AMS 1991 *subject classifications*. 62G07, 62G15

Key words and phrases. Adaptivity, conditional median, convexity, distribution-free, interval censoring, modality, monotonicity, signs of residuals, spacings.

density $1\{(s,t) \in \Pi(\delta)\}(t-s)^{1/2}$, where $\Pi(\delta) := \{(s,t) \in [0,1]^2\colon 0 < t-s \le \delta\}$ and $0 < \delta \le 1$. Then the corresponding Bayes test statistic is given by $T(n^{1/2}V_n)$, where

$$T(h) := \int_{\Pi(\delta)} \exp\biggl(\frac{h(s,t)^2}{2(t-s)}\biggr)\, ds\, dt.$$

This statistic is reminiscent of a goodness-of-fit statistic proposed by Shorack and Wellner (1982). The main difference is the exponential function in the integrand, which is essential for our results. It is true, but not obvious, that $T(W)$ is finite almost surely with continuous distribution; see Section 5. (Note that $\mathbb{E}T(W) = \infty$.) Thus we reject the hypothesis "$F_n = 0$" at level α if $T(n^{1/2}V_n)$ exceeds $c(\alpha)$, the $(1-\alpha)$-quantile of $\mathscr{L}(T(W))$. The corresponding $(1-\alpha)$-confidence set for F_n is given by

$$C_n(V_n, \alpha) := \bigl\{F \in \mathscr{C}[0,1]\colon F(0) = 0, T(n^{1/2}(V_n - F)) \le c(\alpha)\bigr\}.$$

As for the power of this test and the size of $C_n(V_n, \alpha)$, note that $T(\cdot)$ is convex on $\mathscr{C}[0,1]$. Hence

$$2T(n^{1/2}(G-F)/2) \le T(n^{1/2}(V_n - G)) + T(n^{1/2}(V_n - F)) \tag{1}$$

for arbitrary $F, G \in \mathscr{C}[0,1]$. In particular, letting $G = 0$ and $F = F_n$ shows that $T(n^{1/2}V_n) \to \infty$ in probability whenever $T(n^{1/2}F_n/2)$ tends to infinity. For any fixed $F_o \ne 0$, it follows from Fatou's lemma that $T(n^{1/2}F_o/2) \to \infty$. Hence $T(\cdot)$ implies an omnibus test. Unless stated otherwise, asymptotic statements refer to $n \to \infty$.

In order to investigate the power of $T(\cdot)$ more thoroughly, we consider the set

$$C_n^{(1)}(V_n, \alpha) := \biggl\{f \in L_1[0,1]\colon \int_{[0,\cdot]} f(x)\, dx \in C_n(V_n, \alpha)\biggr\}.$$

This is certainly a $(1-\alpha)$-confidence set for the L_1-derivative f_n of F_n (if existent). We consider the intersection of $C_n^{(1)}(V_n, \alpha)$ with Hölder smoothness classes: let $I \subset \mathbf{R}$ be an interval and $\beta = k + \gamma$ with a nonnegative integer k and $0 < \gamma \le 1$. Then $\mathscr{F}_{(\beta, L)}(I)$ stands for the set of all real functions f that are k times differentiable on I such that

$$|f^{(k)}(x) - f^{(k)}(y)| \le L|x-y|^\gamma \quad \text{for all } x, y \in I;$$

here $f^{(k)}$ denotes the kth derivative of f (where $f^{(0)} := f$). Further we define the supremum norm $\|f\|_I := \sup_{x \in I} |f(x)|$. All subsequent consistency results are formulated in terms of

$$\rho_n := \log(n)/n.$$

THEOREM 1.1. *For arbitrary fixed* $\beta, L > 0$, *let* $I_n \subset [0,1]$ *be an interval with length*

$$\mathrm{Leb}(I_n) \ge \rho_n^{1/(2\beta+1)}.$$

Then there exists a constant $R = R(\beta, L)$ *such that*

$$\sup\{\|f-g\|_{I(n)}\colon f, g \in C_n^{(1)}(V_n, \alpha), f-g \in \mathscr{F}_{(\beta, L)}(I_n)\} \le R\rho_n^{\beta/(2\beta+1)}$$

for any fixed $\alpha \in \,]0, 1[$ *and sufficiently large* n.

This result has two straightforward consequences. Suppose that F_n is differentiable with derivative $f_n \in \mathscr{F}_{(\beta, L)}[0, 1]$. When testing "$f_n = 0$" versus

$$\text{“}f_n \in \{f \in \mathscr{F}_{(\beta, L)}[0, 1]\colon \|f\|_{[0,1]} \ge \varepsilon_n\}\text{”}$$

at fixed level α, it was shown by Ingster (1993) that the maximin power converges to one or α as $\rho_n^{-\beta/(2\beta+1)}\varepsilon_n$ tends to infinity or zero, respectively. In fact,

$$T(n^{1/2}V_n) \to_p \infty \quad \text{provided that } \frac{\|f_n\|_{[0,1]}}{\rho_n^{\beta/(2\beta+1)}} \to \infty.$$

Thus our test $1\{T(n^{1/2}V_n) \ge c(\alpha)\}$ is asymptotically optimal in terms of rates of consistency for arbitrary Hölder classes. Another interesting reference in the context of nonparametric testing is Spokoiny (1996).

A second implication is that

$$\sup\{\|f - f_n\|_{[0,1]}\colon f \in C_n^{(1)}(V_n, \alpha) \cap \mathscr{F}_{(\beta, L)}[0, 1]\} \le O_p(\rho_n^{\beta/(2\beta+1)}).$$

Note that the confidence set $C_n^{(1)}(V_n, \alpha) \cap \mathscr{F}_{(\beta, L)}[0, 1]$ may be empty, where $\sup(\varnothing) := -\infty$. In that case $\mathscr{F}_{(\beta, L)}[0, 1]$ is regarded as a questionable model for f_n. The rate $O_p(\rho_n^{\beta/(2\beta+1)})$ was shown by Khas'minskii (1978) to be optimal for estimating $f_n \in \mathscr{F}_{(\beta, L)}[0, 1]$ under sup-norm loss.

Smoothness assumptions such as "$f_n \in \mathscr{F}_{(\beta, L)}[0, 1]$" are difficult to justify in practice. It would be desirable to have a $(1-\alpha)$-confidence set for f_n whose size is automatically of the right order of magnitude, depending on the unknown smoothness of f_n. As pointed out by Low (1997), this is essentially impossible. However, some adaptivity is possible if f_n satisfies shape restrictions such as monotonicity. Restrictions of this type are indeed plausible in many applications. Precisely, we shall investigate the classes

$$\mathscr{F}_{\uparrow}(I) := \{f \text{ nondecreasing on } I\} \quad \text{and} \quad \mathscr{F}_{\downarrow}(I) := -\mathscr{F}_{\uparrow}(I),$$
$$\mathscr{F}_{\text{conv}}(I) := \{f \text{ convex on } I\},$$
$$\mathscr{F}_{\text{cc}}(I) := \{f \text{ convex–concave or concave–convex on } I\}.$$

Rather than doing so in the present white noise model, we propose and analyze modifications of T for three different models.

Section 2 investigates tests for distribution functions on the line and their application to density estimation. Let $X_n = (X_{1n}, X_{2n}, \ldots, X_{nn})$ be the order statistic of n independent random variables with unknown distribution function F_n in $\mathscr{P}$, the set of all continuous distribution functions on the line.

Recall that $(F_n(X_{in}))_{1\le i\le n}$ is distributed as the order statistic of n independent random variables with uniform distribution on $[0,1]$ [cf. Shorack and Wellner (1986), Chapter 1]. Thus

$$\{F \in \mathscr{P}: (F(X_{in}))_{1\le i\le n} \in B_n\}$$

defines a confidence set for F_n whose coverage probability depends only on the set $B_n \subset [0,1]^n$. Hengartner and Stark (1995) constructed confidence bands for shape-restricted densities (monotonicity or unimodality) with the help of simultaneous confidence bounds for $F_n[X_{(i-1)K,n}, X_{iK,n}]$, $1 \le i \le n/K$, where $X_{0n} := -\infty$, $X_{n+1,n} := \infty$ and $K = K_n$ is a bandwidth parameter. One can get rid of the tuning parameter K by considering (essentially) all spacings $[X_{jn}, X_{kn}]$, $0 \le j < k \le n+1$, in a suitable way. Our particular modification results in greater computational complexity involving convex rather than linear progamming, the reward being (almost) optimal rates of convergence for several functions of F_n.

Section 3 is concerned with nonparametric regression. Suppose that one observes $Y_{in} = f_n(t_{in}) + E_{in}$, $1 \le i \le n$, with an unknown function f_n on $\mathbf{R}^d$, fixed design points $t_{in} \in \mathbf{R}^d$ and independent errors E_{in} with median zero. Davies (1995) obtained tests and confidence sets for (functions of) f_n via inversion of the runs test, applied to the random vector

$$\operatorname{sign}(Y_n, f) := (\operatorname{sign}(Y_{in} - f(t_{in})))_{1\le i\le n},$$

where f is a candidate for f_n. For a different application of sign tests in nonparametric regression, see Müller (1991). We propose a test criterion, also based on $\operatorname{sign}(Y_n, f)$, that yields adaptively optimal confidence bands for f_n. These results complement the literature on point estimation under shape restrictions [cf. Mammen (1991) and the references therein]. Some numerical examples for our confidence bands are given.

A possible application of the present methods to interval-censored observations is discussed briefly in Section 4. For a detailed treatment of efficient estimation within this model, see Groeneboom and Wellner (1992).

All proofs are deferred to Section 5.

2. Distribution functions and density estimation. The idea is to replace the process $n^{1/2}(V_n - F)$ in Section 1 with the process

$$t \mapsto n^{1/2}(F(X_{\lfloor (n+1)t\rfloor, n}) - t).$$

Let $\bar{D}_n$ denote the set of pairs (j,k) of integers such that $0 \le j < k \le n+1$. Note that $F_n[X_{jn}, X_{kn}]$ has a Beta-distribution with parameters $k-j$ and $n+1-k+j$ [cf. Shorack and Wellner (1986), Chapter 3.1]. We utilize the following bounds for tail probabilities of the Beta-distribution.

PROPOSITION 2.1. *For* $0 < p < 1$, *define*

$$\Psi(x,p) := p\log\frac{p}{x} + (1-p)\log\frac{1-p}{1-x}$$

if $x \in]0,1[$, *and* $\Psi(x,p) := \infty$ *otherwise. Let* B *be a random variable with distribution* $beta(mp, m(1-p))$, $m > 0$. *Then*

$$\mathbb{P}\{B \geq x\} \leq \exp(-m\,\Psi(x,p)) \quad \textit{for } x \geq p,$$
$$\mathbb{P}\{B \leq x\} \leq \exp(-m\,\Psi(x,p)) \quad \textit{for } x \leq p.$$

The function $\Psi(\cdot, p)$ *is strictly convex on* $[0,1]$ *with minimum* $\Psi(p,p) = 0$. *For any* $c \geq 0$, $\Psi(x,p) \leq c$ *implies that*

$$-\sqrt{2p(1-p)c} - (1-2p)^{-}c \leq x - p \leq \sqrt{2p(1-p)c} + (1-2p)^{+}c.$$

With $\delta_{jkn} := (k-j)/(n+1)$ the precise definition of our test statistic is

$$T_n(X_n, F) := n^{-2} \sum_{(j,k)\in D_n} \exp(n\Psi(F[X_{jn}, X_{kn}], \delta_{jkn})),$$

where $D_n := \{(j,k) \in \bar{D}_n \colon \delta_{\min,n} \leq \delta_{jkn} \leq \delta_n\}$ is a nonvoid subset of $\bar{D}_n$ determined by numbers $0 < \delta_{\min,n} \leq \delta_n \leq 1$. A possible reason for using a lower bound $\delta_{\min,n} > 1/(n+1)$ for δ_{jkn} are discretization errors in the data X_{in}. It is assumed throughout that

$$\delta_{\min,n} = O(\rho_n) \quad \text{and} \quad \delta_n \to \delta.$$

Using an upper bound $\delta_n < 1$ reduces computational complexity and emphasizes smaller intervals. With the $(1-\alpha)$-quantile $b_n(\alpha)$ of $T_n(X_n, F_n)$, the set

$$C_n(X_n, \alpha) := \{F \in \mathscr{P} \colon T(X_n, F) \leq b_n(\alpha)\}$$

is a $(1-\alpha)$-confidence set for F_n. The next proposition summarizes some properties of $T_n(X_n, F_n)$ and $C_n(X_n, \alpha)$.

PROPOSITION 2.2.

(a) $$T_n(X_n, F_n) \to_{\mathscr{L}} \int_{\Pi(\delta)} \exp\left(\frac{B(s,t)^2}{2(t-s)(1-t+s)}\right) ds\,dt,$$

where B *is a Brownian bridge on* $[0,1]$.

(b) *There is a constant* K_o *depending only on* $(D_n)_n$ *such that the following inequalities hold for any* $\alpha \in]0,1[$ *and* n *greater than some integer* $n_o(\alpha) \geq 2$:

$$|F(J) - G(J)| \leq \sqrt{K_o \rho_n F(J)} + K_o \rho_n$$

for arbitrary $F, G \in C_n(X_n, \alpha)$ *and intervals* $J \subset \mathbf{R}$.

Part (b) is the key to various consistency results. One particular application are confidence bands for monotone densities. Similarly to Section 1, we define $C_n^{(1)}(X_n, \alpha)$ to be the set of all probability densities on the line whose distribution function belongs to $C_n(X_n, \alpha)$. A possible notion of consistency is in terms of Hausdorff distance between graphs [cf. Marron and Tsybakov (1995)]. In case of monotone functions this is essentially equivalent to considering a Lévy distance: for functions $f, g \in \mathscr{F}_{\downarrow}(I)$, define

$$d(f, g \mid I) := \inf\Big\{\varepsilon > 0\colon (f(x+\varepsilon) - g(x)) \vee (g(x+\varepsilon) - f(x)) \le \varepsilon \\ \text{whenever } x, x+\varepsilon \in I\Big\}.$$

THEOREM 2.3 (Monotone densities). *Let f and g be arbitrary probability densities in $C_n^{(1)}(X_n, \alpha) \cap \mathscr{F}_{\downarrow}(I)$ for some interval $I \subset \mathbf{R}$. With K_o and $n_o(\alpha)$ as in Proposition* 2.2(b),

$$d(f, g \mid I \cap [a, \infty[) \le (K_o f(a)\rho_n)^{1/3} + (K_o\rho_n)^{1/2} \quad \textit{for all } a \in I,$$

provided that $n \ge n_o(\alpha)$.

In addition suppose that $f \in \mathscr{F}_{(\beta, L)}(I)$ for some $\beta \in\,]0, 1]$. Then there exists a constant $K_1 = K_1(K_o, \beta, L)$ such that

$$g(x) - f(x) \le (K_1 f(x)\rho_n)^{\beta/(2\beta+1)} + (K_1\rho_n)^{\beta/(\beta+1)} \\ \textit{for } x \in I \textit{ with } x - (f(x)\rho_n)^{1/(2\beta+1)} - \rho_n^{1/(\beta+1)} \in I,$$

$$f(x) - g(x) \le (K_1 f(x)\rho_n)^{\beta/(2\beta+1)} + (K_1\rho_n)^{\beta/(\beta+1)} \\ \textit{for } x \in I \textit{ with } x + \Big(\inf_{y\in I} f(y)\rho_n\Big)^{1/(2\beta+1)} \in I,$$

provided that $n \ge n_o(\alpha)$.

Analogous inequalities hold in case of f, g being nondecreasing on some interval. Theorem 2.3 also applies to unimodal or piecewise monotone densities. For instance, let $\mathscr{P}_{\text{uni}}$ be the class of unimodal distributions. That means, $F \in \mathscr{P}_{\text{uni}}$ if it has a density which is nondecreasing on $]-\infty, m(F)[$ and nonincreasing on $]m(F), \infty[$ for some real number $m(F)$, a mode of F. Let $F_n = F_o \in \mathscr{P}_{\text{uni}}$ with unique mode $m(F_o)$ and density f_o. Theorem 2.4 below shows that for any fixed neighborhood $]s, t[$ of $m(F_o)$, with high asymptotic probability the mode $m(F)$ of any $F \in C_n(X_n, \alpha) \cap \mathscr{P}_{\text{uni}}$ is contained in $]s, t[$. In that case Theorem 2.3 applies to the two intervals $]-\infty, s]$ and $[t, \infty[$, respectively, so that $C_n(X_n, \alpha) \cap \mathscr{P}_{\text{uni}}$ gives nontrivial confidence bands for f_o.

THEOREM 2.4 (Inference about the mode). *Suppose that $F_n = F_o \in \mathscr{P}_{\text{uni}}$ with unique mode $m(F_o)$. Then for any $\alpha \in\,]0, 1[$,*

$$\sup\big\{|m(F) - m(F_o)|\colon F \in C_n(X_n, \alpha) \cap \mathscr{P}_{\text{uni}}\big\} \to_p 0.$$

In particular, suppose that the density f_o of F_o satisfies

$$\lim_{x \to m(F_o)} \frac{f_o(m(F_o)) - f_o(x)}{(m(F_o) - x)^2} = \gamma > 0. \tag{2}$$

Then

$$\sup\big\{|m(F) - m(F_o)|\colon F \in C_n(X_n, \alpha) \cap \mathscr{P}_{\text{uni}}\big\} = O_p(\rho_n^{1/5}).$$

The rate $O(\rho_n^{1/5})$ for estimating $m(F_o)$ is close to the optimal rate $O(n^{-1/5})$ [cf. Khas'minskii (1979) and Romano (1988)].

3. Confidence sets for regression functions. Given an index set $\mathscr{T}_n = \{t_{1n}, t_{2n}, \dots, t_{nn}\}$ of n points in $\mathbf{R}^d$, let Y_n be a random vector in $\mathbf{R}^n$ with components

$$Y_{in} = f_n(t_{in}) + E_{in}$$

for some unknown function f_n on $\mathbf{R}^d$ and a random error $E_n \in \mathbf{R}^n$ having independent components E_{in} with median zero. That means $\mathbb{P}\{E_{in} \geq 0\} \wedge \mathbb{P}\{E_{in} \leq 0\} \geq 1/2$. For a function f on $\mathbf{R}^d$, define

$$\operatorname{sign}(Y_n, f) := \big\{s \in \{-1, 1\}^n\colon \operatorname{sign}(Y_{in} - f(t_{in})) \in \{0, s_i\} \text{ for } 1 \leq i \leq n\big\}.$$

This somewhat unusual definition is made in order to deal with possibly discrete error distributions. Let Σ_n be uniformly distributed on $\{-1, 1\}^n$. If all components of Y_n have a continuous distribution, then $\operatorname{sign}(Y_n, f_n) = \operatorname{sign}(E_n, 0)$ consists almost surely of one point whose distribution is $\mathscr{L}(\Sigma_n)$. In general, one can easily couple the random vectors Σ_n and E_n in such a way that

$$\Sigma_n \in \operatorname{sign}(Y_n, f_n) \quad \text{almost surely.} \tag{3}$$

Now we define the test statistic

$$T_n(Y_n, f) := \min\big\{\tau_n(s)\colon s \in \operatorname{sign}(Y_n, f)\big\},$$

$$\tau_n(s) := (\#\mathscr{A}_n)^{-1} \sum_{A \in \mathscr{A}_n} \exp\left(\frac{\big(\sum_{i=1}^n 1\{t_{in} \in A\} s_i\big)^2}{2\#A}\right),$$

where $\mathscr{A}_n$ is a family of nonvoid subsets of $\mathscr{T}_n$. A corresponding $(1-\alpha)$-confidence set for f_n is given by

$$C_n^{(1)}(Y_n, \alpha) := \big\{f\colon T_n(Y_n, f) \leq c_n(\alpha)\big\}$$

with $c_n(\alpha)$ denoting the $(1-\alpha)$-quantile of $\mathscr{L}(\tau_n(\Sigma_n))$. Note that $T(Y_n, f_n) \leq \tau_n(\Sigma_n)$ almost surely if (3) holds.

EXAMPLE. Let $n = m^d$ for some integers $m, d > 0$. Then define

$$\mathscr{T}_n^{[d]} := \{1/m, 2/m, \dots, 1\}^d,$$

$$\mathscr{A}_n^{[d]}(\delta_n) := \big\{[x, y] \cap \mathscr{T}_n^{[d]}\colon x, y \in \mathscr{T}_n^{[d]} \text{ with } 0 < y_i - x_i \leq \delta_n \text{ for all } i\big\},$$

where $[x, y] := \prod_{i=1}^d [x_i, y_i]$ and $\delta_n \to \delta$. Here $\#\mathscr{A}_n \leq (m(m-1)/2)^d \leq n^2$. Table 1 gives some Monte Carlo estimates for $c_n(\alpha)$ in dimension one.

Here is the analogue to Proposition 2.2.

PROPOSITION 3.1. *Suppose that* $\#\mathscr{A}_n \to \infty$.

(a) *In general*,

$$\tau_n(\Sigma_n) = O_p(\log(\#\mathscr{A}_n)).$$

TABLE 1
Estimated quantiles $c_n(\alpha)$ of $\tau_n(\Sigma_n)$ for $(\mathcal{T}_n^{[1]}, \mathcal{A}_n^{[1]}(\delta_n))$ (20000 simulations)

	$\delta_n = 0.5$			$\delta_n = 0.25$		
n	$c_n(0.5)$	$c_n(0.1)$	$c_n(0.05)$	$c_n(0.5)$	$c_n(0.1)$	$c_n(0.05)$
50	1.85	3.85	5.56	1.97	3.51	4.47
100	1.87	4.13	6.17	2.02	3.96	5.43
150	1.92	4.29	6.53	2.05	4.12	5.77
200	1.93	4.44	6.67	2.08	4.22	5.82
250	1.95	4.50	6.73	2.08	4.34	6.37
300	1.96	4.53	7.03	2.10	4.35	6.43
400	1.96	4.64	7.15	2.12	4.55	6.60
500	1.98	4.65	7.19	2.13	4.55	6.66

Suppose that $\mathcal{A}_n \subset \{D \cap \mathcal{T}_n : D \in \mathcal{D}\}$ for some Vapnik–Cervonenkis class $\mathcal{D}$ of subsets of $\mathbf{R}^d$, and let

$$\limsup_{n\to\infty} (\#\mathcal{A}_n)^{-1} \sum_{A\in\mathcal{A}(n)} \log(n/\#A) < \infty.$$

Then $\tau_n(\Sigma_n) = O_p(1)$. In particular, for $(\mathcal{T}_n, \mathcal{A}_n) = (\mathcal{T}_n^{[1]}, \mathcal{A}_n^{[1]}(\delta_n))$,

$$(\delta - \delta^2/2)\tau_n(\Sigma_n) \to_{\mathcal{L}} T(W).$$

(b) *Let $H: [0,1] \to [0,\infty]$ be a fixed nondecreasing function such that all random vectors E_n satisfy*

$$\mathbb{P}\{E_{in} \le H(u)\} \wedge \mathbb{P}\{E_{in} \ge -H(u)\} \ge \frac{1+u}{2} \quad \text{for } 1 \le i \le n,\ u \in [0,1].$$

Then for any fixed $\alpha \in]0,1[$, the probability of

$$\bigcup_{A\in\mathcal{A}(n)} \left\{ \sup_{f\in C_n^{(1)}(Y_n,\alpha)} \min_{t\in A}(f(t) - f_n(t)) \vee \min_{t\in A}(f_n(t) - f(t)) > H\left(3\sqrt{\frac{\log(\#\mathcal{A}_n)}{\#A}}\right)\right\}$$

tends to zero, where $H(u) := \infty$ for $u > 1$.

As for part (b), if all components of E_n are Gaussian with variance not greater than τ, one may take $H(u) := \tau\Phi^{-1}((1+u)/2)$ with the standard normal quantile function Φ^{-1}. A key condition on H is

$$\limsup_{u\downarrow 0} \frac{H(u)}{u} < \infty. \tag{4}$$

THEOREM 3.2 (Isotonic regression). *Let all f_n belong to the class $\mathcal{F}_\uparrow([0,1]^d)$ of functions f on $[0,1]^d$ such that $f(s) \le f(t)$ whenever $s \le t$ componentwise.*

For $f, g \in \mathscr{F}_{\uparrow}([0,1]^d)$, define a Lévy distance

$$d(f,g) := \inf\Big\{\varepsilon > 0\colon l(f(s) - g(s+\varepsilon\mathbf{1})) \vee (g(s) - f(s+\varepsilon\mathbf{1})) \le \varepsilon \text{ whenever } s, s+\varepsilon\mathbf{1} \in [0,1]^d\Big\},$$

where $\mathbf{1} := (1,1,\dots,1)' \in \mathbf{R}^d$. Let $(\mathscr{T}_n, \mathscr{A}_n) = (\mathscr{T}_n^{[d]}, \mathscr{A}_n^{[d]}(\delta_n))$ as above. Suppose that the assumption of Proposition 3.1(b) *holds with a function H satisfying* (4). *Then*

$$\sup\big\{d(f, f_n)\colon f \in C_n^{(1)}(Y_n, \alpha) \cap \mathscr{F}_{\uparrow}([0,1]^d)\big\} \le O_p\big(\rho_n^{1/(2+d)}\big).$$

THEOREM 3.3 (Convex–concave regression). *Let $(\mathscr{T}_n, \mathscr{A}_n) = (\mathscr{T}_{!n}^{[1]}, \mathscr{A}_n^{[1]}(\delta_n))$. Suppose that all f_n belong to $\mathscr{F}_{\mathrm{cc}} \cap \mathscr{F}_{(\beta,L)}[a,b]$ for some $0 \le a < b \le 1$, $\beta \in]0,2]$ and $L > 0$. Further suppose that the assumption of Proposition* 3.1(b) *holds with a function H satisfying* (4). *Then*

$$\sup\Big\{\|f - f_n\|_{\big[a+\rho_n^{1/(2\beta+1)},\, b-\rho_n^{1/(2\beta+1)}\big]}\colon f \in C_n^{(1)}(Y_n,\alpha) \cap \mathscr{F}_{\mathrm{cc}}\Big\} = O_p\big(\rho_n^{\beta/(2\beta+1)}\big).$$

Numerical examples. In all subsequent examples we consider the pair $(\mathscr{T}_n^{[1]}, \mathscr{A}_n^{[1]}(0.25))$. We simulated data Y_{in} (shown as dots) having logistic distribution with mean $f_n(i/n)$ (shown as dotted line) and standard deviation v_{in}. Point estimators and confidence bands are shown as solid lines.

Figure 1 shows two data vectors Y_n with $n = 200$ and $v_{in} = 0.4$. We minimized $T_n(Y_n, f)$ over all $f \in \mathscr{F}_{\mathrm{conv}}[0,1]$. In both examples the minimum turned out to be unique, although this is not necessarily the case. This led to a sign vector $s_{\min}$, and the solid lines represent the functions

$$t \mapsto \begin{cases} \min\big\{f(t)\colon f \in \mathscr{F}_{\mathrm{conv}}[0,1],\ s_{\min} \in \mathrm{sign}(Y_n, f)\big\}, \\ \max\big\{f(t)\colon f \in \mathscr{F}_{\mathrm{conv}}[0,1],\ s_{\min} \in \mathrm{sign}(Y_n, f)\big\}. \end{cases}$$

The regression function f_n was taken to be

$$f_n(t) := (1 - 5t/2) \vee (5t/3 - 2/3)^2$$

and

$$f_n(t) := (1 - 5t/2) \vee (5t/3 - 2/3)^2 + 1\{2/5 < t < 4/5\}\sin(5\pi t)/2,$$

respectively. The corresponding observed p-value $\mathbb{P}\{\tau_n(\Sigma_n) \ge \tau_n(s_{\min})\}$ was estimated in 40000 Monte Carlo simulations. In the first example it turned out greater than 0.99. In fact, the distance between estimator and f_n is small in comparison with the noise level v_{in}. In the second example the Monte Carlo p-value was 0.027, so that the nonconvexity of f_n is detected at level 0.05.

Figures 2, 3 and 4 depict examples of confidence bands, that is, the envelope functions

$$t \mapsto \begin{cases} \min\big\{f(t)\colon f \in C_n^{(1)}(Y_n, 0.1) \cap \mathscr{F}\big\}, \\ \max\big\{f(t)\colon f \in C_n^{(1)}(Y_n, 0.1) \cap \mathscr{F}\big\}. \end{cases}$$

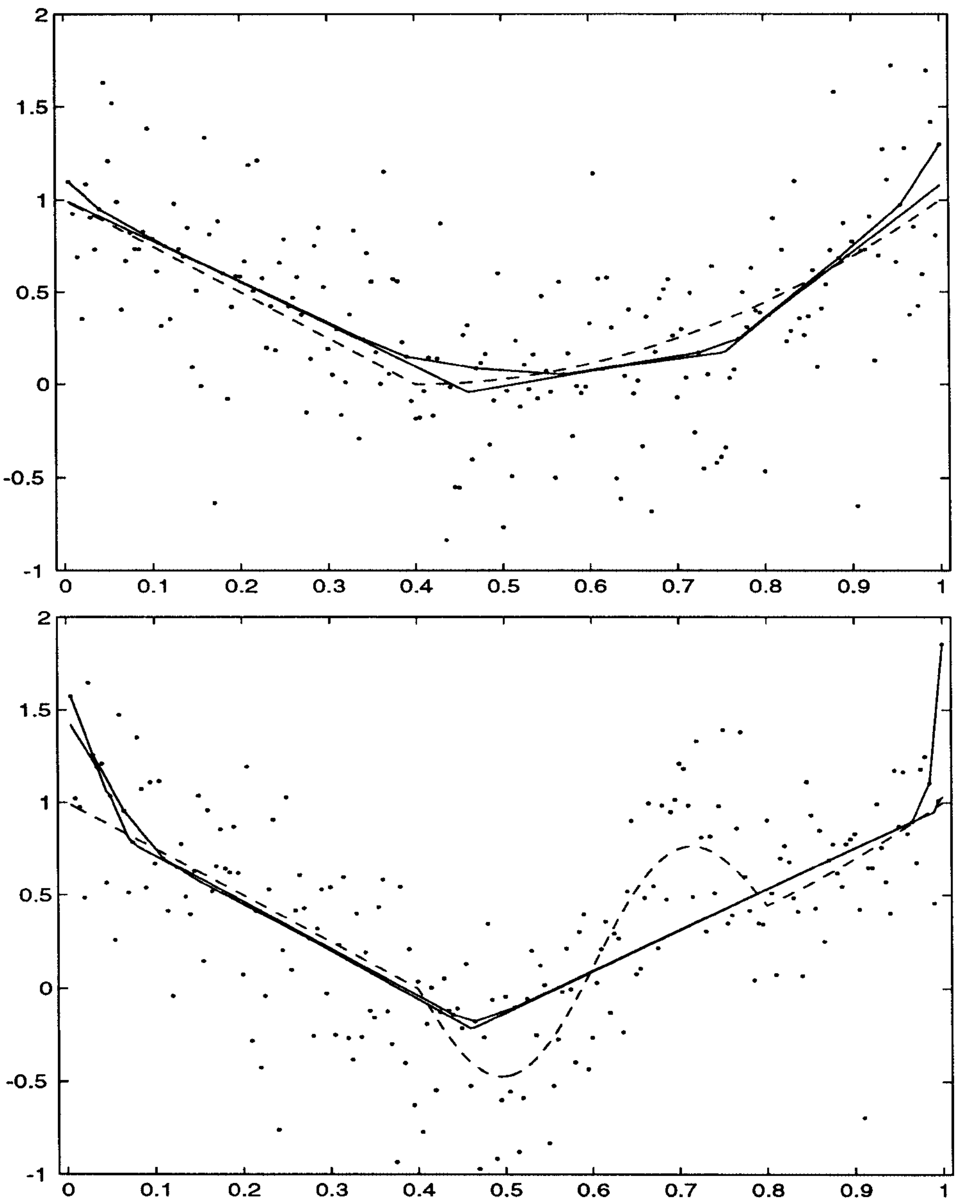

FIG. 1. *Point estimators for* $f_n \in \mathscr{F}_{\text{conv}}[0, 1]$.

Precisely, in Figure 2 the parameters are $n = 250$, $v_{in} = 0.5$ and

$$f_n(t) = 1\{t \geq 1/2\}, \qquad \mathscr{F} = \mathscr{F}_{\uparrow}[0, 1].$$

In Figure 3 we have $n = 250$, $v_{in} = 0.3$ and

$$f_n(t) = (1 - 3t) \vee (3t/2 - 1/2)^2, \qquad \mathscr{F} = \mathscr{F}_{\text{conv}}[0, 1].$$

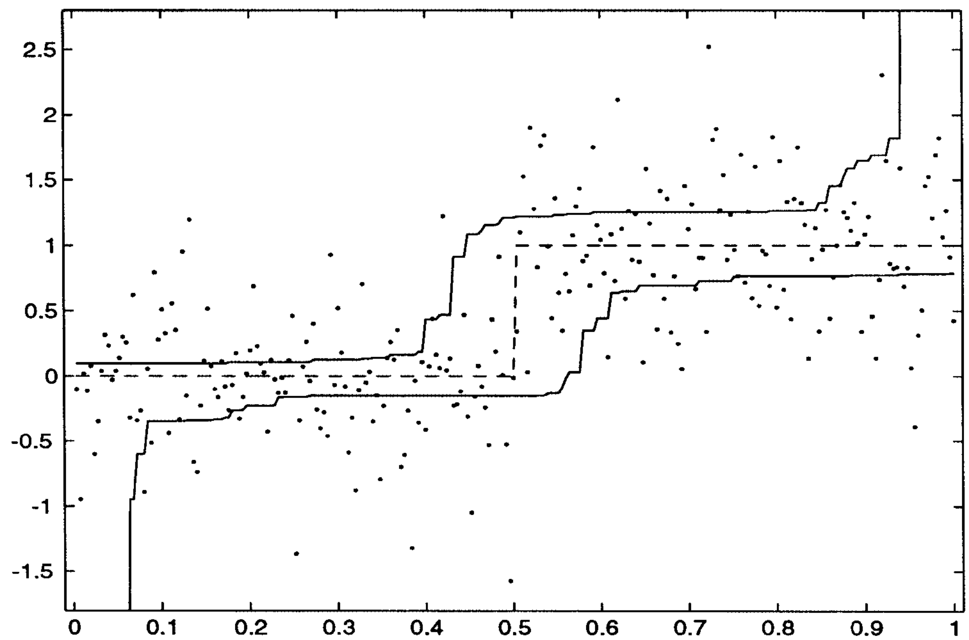

FIG. 2. *Envelope of* $C_n^{(1)}(Y_n, 0.1) \cap \mathscr{F}_{\uparrow}[0,1]$.

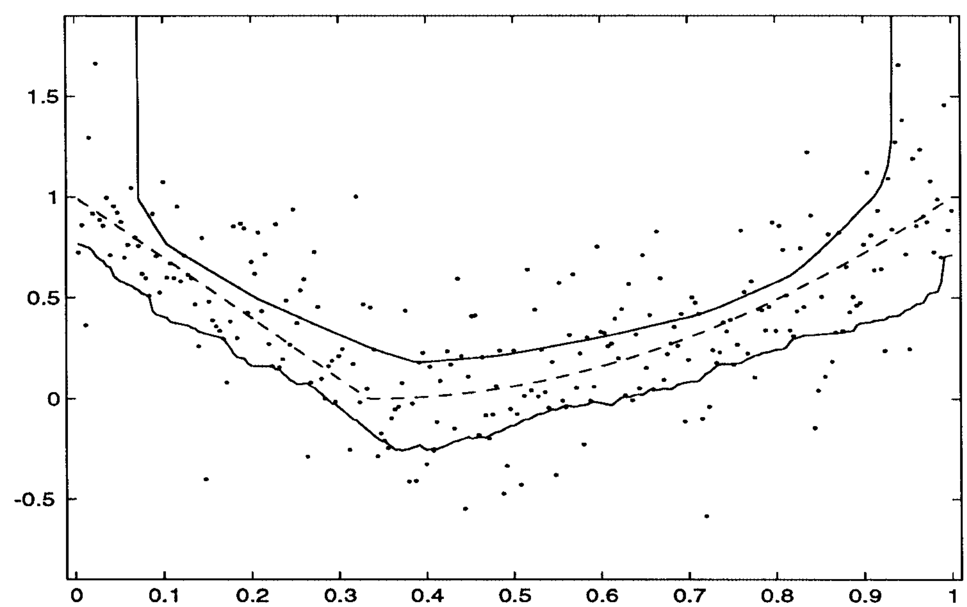

FIG. 3. *Envelope of* $C_n^{(1)}(Y_n, 0.1) \cap \mathscr{F}_{\mathrm{conv}}[0,1]$.

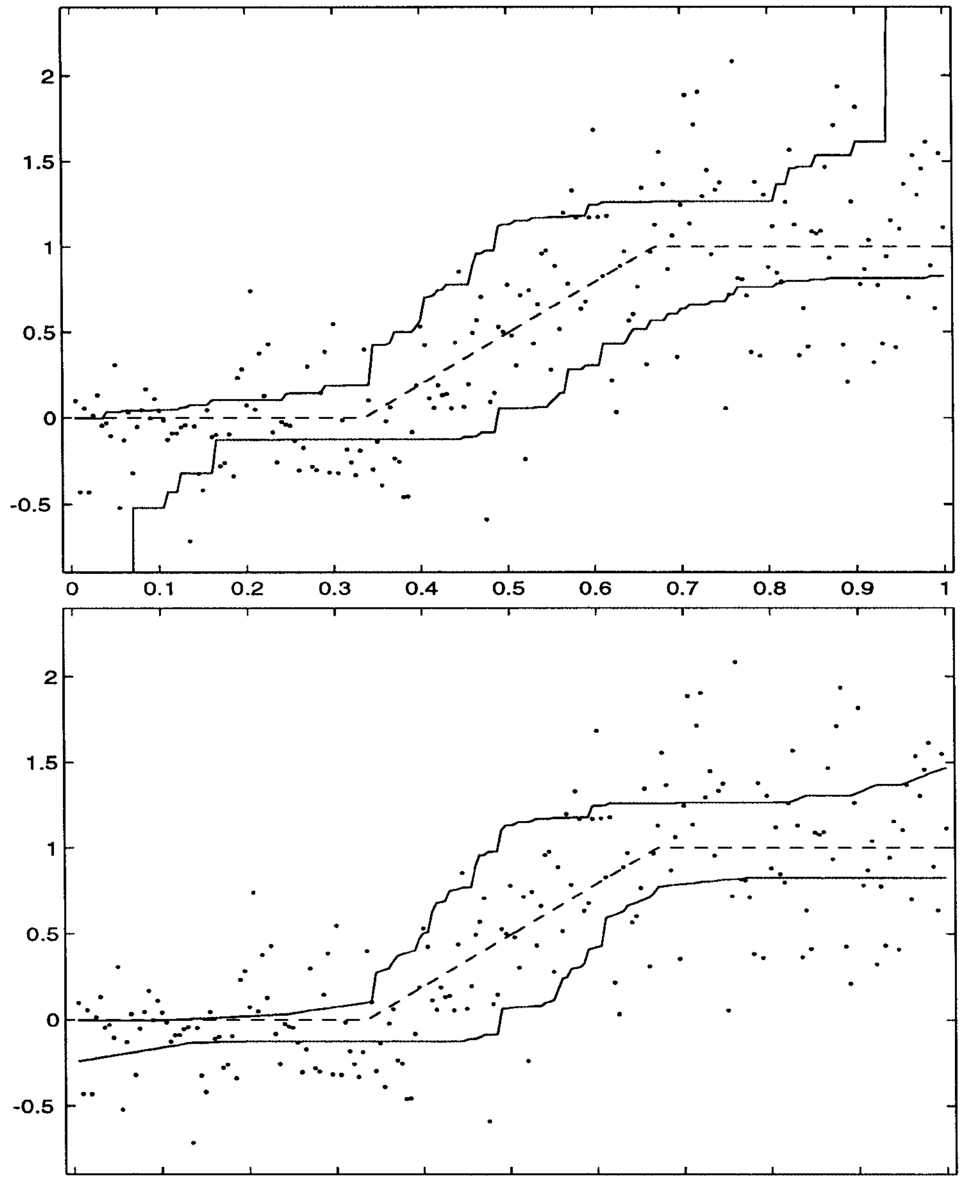

FIG. 4. *Envelopes of* $C_n^{(1)}(Y_n, 0.1) \cap \mathscr{F}_{\uparrow}[0, 1]$ *and* $C_n^{(1)}(Y_n, 0.1) \cap \mathscr{F}_{\uparrow}[0, 1] \cap \mathscr{F}_{cc}[0, 1]$.

Figure 4 depicts heteroscedastic data Y_{in} with $n = 200$ and

$$f_n(t) = (3x - 1)^+ \wedge 1, \qquad v_{\text{in}} = (1 + f_n(i/n))/4.$$

The two plots show the envelopes of $C_n^{(1)}(Y_n, 0.1) \cap \mathscr{F}_{\uparrow}[0, 1]$ and $C_n^{(1)}(Y_n, 0.1) \cap \mathscr{F}_{\uparrow}[0, 1] \cap \mathscr{F}_{cc}[0, 1]$, respectively. Here the additional constraint "$f_n \in \mathscr{F}_{cc}[0, 1]$" led to considerably smaller confidence bands.

4. Interval censoring. Let $\tilde{X}_{1n}, \tilde{X}_{2n}, \dots, \tilde{X}_{nn}$ be independent, identically distributed random variables with distribution function F_n. Rather than $\tilde{X}_{in}$, one only observes $Z_{in} := 1\{\tilde{X}_{in} \le r_{in}\}$, $1 \le i \le n$, where $r_{1n} \le r_{2n} \le \cdots \le r_{nn}$ are given censoring times (viewed as fixed). Given a hypothetical distribution function F, let

$$Z_n(t \mid F) := 2n^{-1/2} \sum_{i \le nt} (Z_{in} - F(r_{in})), \qquad t \in [0,1].$$

Then our test statistic for "$F_n = F$" is $T_n(Z_n(\cdot \mid F))$, where

$$T_n(h) := n^{-2} \sum_{(s,t) \in \Pi_n(\delta_n)} \exp\left(\frac{h(s,t)^2}{2(t-s)}\right)$$

with $\Pi_n(\delta_n) := \Pi(\delta_n) \cap \{0, 1/n, 2/n, \dots 1\}^2$ and $\delta_n \to \delta$. Unfortunately the $(1-\alpha)$-quantile $d_n(\alpha \mid F_n)$ of the distribution of $T_n(Z_n(\cdot \mid F_n))$ depends on the unknown function F_n. However, the case $F_n(r_{1n}) = F_n(r_{nn}) = 1/2$ is the worst case asymptotically. The corresponding quantile is denoted by $d_n(\alpha)$. We define $C_n(Z_n, \alpha)$ to be the set of all distribution functions F such that $T_n(Z_n(\cdot \mid F)) \le d_n(\alpha)$.

PROPOSITION 4.1. (a) *For any fixed* $\alpha \in]0,1[$,

$$\lim_{n\to\infty} d_n(\alpha) = c(\alpha) \quad \textit{and} \quad \limsup_{n\to\infty} \mathbb{P}\{T_n(Z_n(\cdot \mid F_n)) \ge d_n(\alpha)\} \le \alpha.$$

(b) *For any fixed* $\alpha \in]0,1[$,

$$\lim_{n\to\infty} \mathbb{P}\left\{ \sup_{F \in C_n(Z_n, \alpha)} (F(s) - F_n(t)) \vee (F_n(s) - F(t)) \ge \sqrt{\frac{3\delta^{-1}\rho_n}{\mu_n[s,t]}} \textit{ for some } s < t \right\} = 0,$$

where $\mu_n(\cdot) := n^{-1} \sum_{i=1}^n 1\{r_{in} \in \cdot\}$.

Part (b) implies consistency of $C_n(Z_n, \alpha)$ in various senses under certain conditions on the sequence of distributions μ_n. We mention only two simple consequences:

THEOREM 4.2. (a) *Suppose that* $F_n = F_o$ *is continuous and that* μ_n *converges weakly to a probability measure* μ *such that* support$(\mu) \supset$ support(F_o). *Then*

$$\sup\{\|F - F_o\|_{\mathbf{R}} \colon F \in C_n(Z_n, \alpha)\} \to_p 0.$$

(b) *Suppose that that* $F_n(0) = 0$ *and* $F_n \in \mathscr{F}_{(\beta, L)}[0, \infty[$ *for some* $\beta \in]0,1]$ *and* $L > 0$. *Further let* $(r_{1n}, r_{2n}, \dots, r_{nn})$ *be the order statistic of independent, identically distributed random variables* $R_1, R_2, \dots, R_n$ *with distribution* μ *having continuous density* $\mu^{(1)}$ *on* $[0, \infty[$. *Then*

$$\sup\{\|F - F_n\|_I \colon F \in C_n(Z_n, \alpha)\} = O_p\left(\rho_n^{\beta/(2\beta+1)}\right)$$

for any compact subset I *of* $\{t \ge 0 \colon \mu^{(1)}(t) > 0\}$.

5. Proofs. In order to verify the finiteness of $T(W)$, let

$$\check{T}(W) := \sup_{(s,t)\in\Pi(1)} \frac{W(s,t)^2}{2(t-s)\log(e/(t-s))}.$$

According to Lévy's theorem on W's modulus of continuity, $1 \le \check{T}(W) < \infty$ almost surely [cf. Shorack and Wellner (1986), Theorem 14.1.1]. Now the key point is that

$$\begin{aligned}
&\mathbb{E}\, 1\{\check{T}(W) \le M\}\, T(W)\\
&\quad \le \int_{\Pi(\delta)} \mathbb{E}\, 1\left\{\frac{W(s,t)^2}{2(t-s)} \le M \log\left(\frac{e}{(t-s)}\right)\right\} \exp\left(\frac{W(s,t)^2}{2(t-s)}\right) ds\, dt\\
&\quad \le \int_{\Pi(\delta)} \left(1 + 2M \log\left(\frac{e}{(t-s)}\right)\right) ds\, dt\\
&\quad < \infty
\end{aligned}$$

for any constant $M > 1$; see Lemma 5.1. Continuity of $\mathscr{L}(T(W))$ follows from the fact that T is strictly convex on the set $\{G \in \mathscr{C}[0,1]\colon G(0) = 0\}$. For $W(t) = B(t) + \xi t$, $0 \le t \le 1$, with a Brownian bridge B and a standard Gaussian variable ξ such that B and ξ are independent. Thus conditional on B, the test statistic $T(W)$ is a strictly convex function of ξ, whence continuously distributed. This consideration shows in addition that the support of $\mathscr{L}(T(W))$ is connected.

LEMMA 5.1. *Let X be a nonnegative random variable such that $\mathbb{P}\{X \ge r\} \le 2\exp(-r)$ for all $r \ge 0$ (e.g., $X = Z^2/2$ with $Z \sim \mathscr{N}(0,1)$). Then for all $\gamma, l > 0$,*

$$\mathbb{E}1\{X \le l\}\exp(\gamma X) \le \begin{cases} 1 + 2l, & \text{if } \gamma = 1,\\ 1 + 2(\exp((\gamma-1)l) - 1)/(1 - 1/\gamma), & \text{if } \gamma \ne 1.\end{cases}$$

PROOF. The expectation of $1\{X \le l\}\exp(\gamma X)$ equals

$$\begin{aligned}
&\int_0^\infty \mathbb{P}\{X \le l \text{ and } \exp(\gamma X) > r\}\, dr\\
&\qquad \le 1 + \int_1^{\exp(\gamma l)} \mathbb{P}\{\exp(\gamma X) > r\}\, dr\\
&\qquad \le 1 + 2\int_1^{\exp(\gamma l)} r^{-1/\gamma}\, dr\\
&\qquad = \begin{cases} 1 + 2l, & \text{if } \gamma = 1,\\ 1 + 2(\exp((\gamma-1)l) - 1)/(1 - 1/\gamma), & \text{if } \gamma \ne 1.\end{cases} \qquad \square
\end{aligned}$$

The proofs of Theorems 1.1 and 3.3 are based on a lemma on Hölder classes of functions.

LEMMA 5.2. *For $\beta, L > 0$ there is a universal constant $K_{(\beta,L)} > 0$ such that for arbitrary compact intervals $I \subset \mathbf{R}$ and any $f \in \mathcal{F}_{(\beta,L)}(I)$ the following hold.*

(a) *There is an interval $J_f \subset I$ such that*

$$|f| \geq K_{(\beta,L)} \|f\|_I \quad \text{on } J_f,$$

$$\mathrm{Leb}(J_f) \geq K_{(\beta,L)}\big(\|f\|_I^{1/\beta} \wedge \mathrm{Leb}(I)\big).$$

(b) *If $\beta \leq 2$, then for arbitrary $g \in \mathcal{F}_{\mathrm{cc}}(I)$ there is an interval $J_{fg} \subset I$ such that*

$$|g - f| \geq |g(x_o) - f(x_o)|/4 \quad \text{on } J_{fg},$$

$$\mathrm{Leb}(J_{fg}) \geq K_{(\beta,L)}\big(|g(x_o) - f(x_o)|^{1/\beta} \wedge \mathrm{Leb}(I)\big),$$

where x_o denotes the midpoint of I.

PROOF OF LEMMA 5.2(a). Let $x_1 \in I$ with $|f(x_1)| = \|f\|_I$, and define $\gamma := \|f\|_I^{1/\beta} \wedge \mathrm{Leb}(I)$.

If $0 < \beta \leq 1$, then $|f(x)| \geq \|f\|_I - L|x - x_1|^\beta \geq \|f\|_I/2$ for any point x in $J_f := [x_1 - (2L)^{-1/\beta}\gamma, x_1 + (2L)^{-1/\beta}\gamma] \cap I$, where $\mathrm{Leb}(J_f) \geq ((2L)^{-1/\beta} \wedge 2^{-1})\gamma$.

For $\beta > 1$ we use induction on $k = k(\beta)$. Suppose the assertion is true for $(\beta - 1, L)$ in place of (β, L). If $|f(x)| \geq \|f\|_I/2$ for all $x \in J'_f := [x_1 - \gamma/2, x_1 + \gamma/2] \cap I$, the assertion would be true with $J_f := J'_f$ and $K_{(\beta,L)} := 1/2$. Otherwise let $x_2 \in J'_f$ with $|f(x_2)| \leq \|f\|_I/2$. Then

$$\|f^{(1)}\|_I \geq |f(x_1) - f(x_2)|/|x_1 - x_2| \geq \|f\|_I/\gamma.$$

By assumption, since $f^{(1)} \in \mathcal{F}_{(\beta-1,L)}(I)$, there is an interval $J''_f \subset I$ such that

$$|f^{(1)}| \geq K_{(\beta-1,L)}\|f\|_I/\gamma \text{ on } J''_f,$$

$$\mathrm{Leb}(J''_f) \geq K_{(\beta-1,L)}((\|f\|_I/\gamma)^{1/(\beta-1)} \wedge \mathrm{Leb}(I)) = K_{(\beta-1,L)}\gamma.$$

Hence, if $a_0 := \inf(J''_f)$ and $a_i := a_0 + (i/4)\,\mathrm{Leb}(J''_f)$, then

$$|f(a_i) - f(a_{i-1})| \geq 4^{-1}K^2_{(\beta-1,L)}\|f\|_I \quad \text{for } 1 \leq i \leq 4.$$

In addition, f is strictly monotone on J''_f by continuity of $f^{(1)}$. Hence one easily verifies that $|f| \geq 4^{-1}K^2_{(\beta-1,L)}\|f\|_I$ on $[a_0, a_1]$ or $[a_3, a_4]$. □

PROOF OF LEMMA 5.2(b). At first we consider the special case, where $I = [-4, 4]$, $f \equiv 0$, $g(0) = 1$ and g is convex–concave on $[-4, 4]$. Under these assumptions there is an interval $J \subset [-4, 4]$ such that

$$|g| \geq 1/2 \text{ on } J \quad \text{and} \quad \mathrm{Leb}(J) \geq 1.$$

Obviously, this is true if $g > 1/2$ on $[-1, 0]$ or on $[0, 1]$. Otherwise, let $-1 \leq x_1 < 0 < x_2 \leq 1$ with $g(x_1) \vee g(x_2) \leq 1/2$. Convex-concavity of g implies that

g is concave on $[x_*, 4]$ and convex on $[-4, x_*]$ for some $x_* \in [-4, 1[$, because otherwise g would be convex on $[-1, 1]$. If $x_* \le 0$, then the L_1-derivative $g^{(1)}$ of g satisfies

$$g^{(1)}(x) \le g^{(1)}(x_2) \le (g(x_2) - 1)/x_2 \le -1/2 \quad \text{for } x_2 \le x < 4.$$

If $x_* > 0$, then convexity of g on $[-4, x_*]$ and $g(x_1) < 1$ together imply that $g(x_*) > 1$, whence

$$g^{(1)}(x) \le g^{(1)}(x_2) \le (g(x_2) - g(x_*))/(x_2 - x_*) < -1/2 \quad \text{for } x_2 \le x < 4.$$

Thus

$$g(x) \le g(x_2) + \int_{x_2}^{x} g^{(1)}(r)\, dr \le 1/2 - (x - x_2)/2 \le -1/2 \quad \text{for } 3 \le x < 4.$$

With the help of affine transformations, one can deduce that in the general case for any $0 < \gamma \le \mathrm{Leb}(I)/2$ there is an interval $J_{fg\gamma} \subset [x_o - \gamma, x_o + \gamma] \subset I$ such that

$$|G| \ge |g(x_o) - f(x_o)|/2 \text{ on } J_{fg\gamma} \quad \text{and} \quad \mathrm{Leb}(J_{fg\gamma}) \ge \gamma/4,$$

where

$$G(x) := \begin{cases} g(x) - f(x_o), & \text{if } 0 < \beta \le 1, \\ g(x) - f(x_o) - f^{(1)}(x_o)(x - x_o), & \text{if } 1 < \beta \le 2, \end{cases}$$

is also in $\mathscr{F}_{\mathrm{cc}}(I)$. But for $x \in [x_o - \gamma, x_o + \gamma]$,

$$|G(x) - (g - f)(x)| = \begin{cases} |f(x) - f(x_o)|, & \text{if } 0 < \beta \le 1, \\ |\int_{x_o}^{x} (f^{(1)}(r) - f^{(1)}(x_o))\, dr|, & \text{if } 1 < \beta \le 2, \end{cases}$$
$$\le L\gamma^{\beta} \le |g(x_o) - f(x_o)|/4,$$

provided that $\gamma \le (4L)^{-1/\beta} |g(x_o) - f(x_o)|^{1/\beta}$. □

PROOF OF THEOREM 1.1. Let F, G be continuous functions on $[0, 1]$ with L_1-derivatives $f, g \in C_n^{(1)}(V_n, \alpha)$, respectively, such that $f - g \in \mathscr{F}_{(\beta, L)}(I_n)$. Then by (1),

$$T(n^{1/2}(F - G)/2) \le 2^{-1}(T(n^{1/2}(V_n - F)) + T(n^{1/2}(V_n - G))) \le c(\alpha).$$

Given any fixed number $R \ge 1$, suppose that $f(x) - g(x) \ge R\rho_n^{\beta/(2\beta+1)}$ for some $x \in I_n$. According to Lemma 5.2 there exists an interval $J = J_{fgn} \subset I_n$ such that

$$f - g \ge K_* R \rho_n^{\beta/(2\beta+1)} \quad \text{on } J,$$
$$\mathrm{Leb}(J) \ge K_* \big(\|f - g\|_{I(n)}^{1/\beta} \wedge \mathrm{Leb}(I_n) \big) \ge K_* \rho_n^{1/(2\beta+1)},$$

where K_* denotes a generic positive constant depending only on (β, L) but possibly different in various places. If $J^{(1)}$ and $J^{(3)}$ denote the left and right third of J, respectively, then for $s \in J^{(1)}$ and $t \in J^{(3)}$,

$$t - s \geq K_* \rho_n^{1/(2\beta+1)},$$
$$\frac{(F-G)(s,t)^2}{2(t-s)} \geq K_* R^2 \rho_n^{2\beta/(2\beta+1)}(t-s) \geq K_* R^2 \rho_n.$$

Thus

$$T(n^{1/2}(F-G)/2) \geq \int 1\{s \in J^{(1)}\} 1\{t \in J^{(3)}\} \exp\big(K_* R^2 n \rho_n\big)\, ds\, dt$$
$$\geq K_* n^{K_* R^2 - 2/(2\beta+1)}.$$

For n and R sufficiently large, the latter bound exceeds $c(\alpha)$. In that case $\|f - g\|_{I(n)}$ is necessarily smaller than $R\rho_n^{\beta/(2\beta+1)}$. □

PROOF OF PROPOSITION 2.1. Let G and G' be independent Gamma-distributed random variables with mean mp and $m(1-p)$, respectively. Then $B := G/(G+G')$ has the desired Beta-distribution, and for $p < x < 1$,

$$\begin{aligned}
\mathbb{P}\{B \geq x\} &= \mathbb{P}\{(1-x)G - xG' \geq 0\} \\
&\leq \inf_{r>0} \mathbb{E} \exp(r(1-x)G - rxG') \\
&= \inf_{0<r<1/(1-x)} \exp(-mp \log(1 - r(1-x)) - m(1-p)\log(1+rx)) \\
&= \exp(-m\,\Psi(x, p)),
\end{aligned}$$

where $r_{\min} = (x-p)/(x(1-x))$. With $\kappa := p(1-p)$ and $\gamma := 1 - 2p$ one can write

$$\begin{aligned}
\Psi(x, p) &= \int_0^{x-p} \frac{r}{\kappa + \gamma r - r^2}\, dr \\
&\geq \int_0^{x-p} \frac{r}{\kappa + \gamma r}\, dr \begin{cases} \geq \kappa^{-1}(x-p)^2/2, & \text{if } p \geq 1/2, \\ = \kappa\gamma^{-2} H(\kappa^{-1}\gamma(x-p)), & \text{if } p < 1/2, \end{cases}
\end{aligned}$$

where $H(y) := y - \log(1+y)$ is strictly increasing in $y \geq 0$. It follows easily from the series expansion of $\exp(\cdot)$ that $H^{-1}(y) \leq (2y)^{1/2} + y$. Thus $\Psi(x, p) \leq c$ implies that

$$x - p \leq \begin{cases} (2\kappa c)^{1/2}, & \text{if } p \geq 1/2, \\ \kappa\gamma^{-1} H^{-1}(\kappa^{-1}\gamma^2 c) \leq (2\kappa c)^{1/2} + \gamma c, & \text{if } p < 1/2. \end{cases}$$

For $0 < x < p$ the assertions follow from the fact that $1 - B \sim \mathrm{Beta}(m(1-p), mp)$ and $\Psi(x, p) = \Psi(1-x, 1-p)$. □

Here is a modified version of Lemma VII.9 of Pollard (1984), which is convenient for our purposes. The proof is essentially the same.

LEMMA 5.3 (Chaining). *Let $S = (S(t))_{t\in\mathscr{T}}$ be a stochastic process on a totally bounded metric space $(\mathscr{T}, \rho)$ having continuous sample paths. Let Q be a measurable, nonnegative function on $]0,\infty[^2$ such that for all $\eta, \delta > 0$ and $s, t \in \mathscr{T}$,*

$$\mathbb{P}\{|S(s) - S(t)| \geq \rho(s,t)Q(\eta,\delta)\} \leq 2\exp(-\eta) \quad \text{if } \rho(s,t) \geq \delta.$$

Then

$$\mathbb{P}\{|S(s) - S(t)| > 12J(\rho(s,t), a) \text{ for some } s, t \in \mathscr{T} \text{ with } \rho(s,t) \leq \delta\} \leq 2\delta/a$$

for arbitrary $a, \delta > 0$, where

$$J(\varepsilon, a) := \int_0^\varepsilon Q(\log(aD(u)^2/u), u)\, du,$$
$$D(u) := \sup\{\#\mathscr{T}_o : \mathscr{T}_o \subset \mathscr{T}, \rho(s,t) > u \text{ for different } s, t \in \mathscr{T}_o\}. \qquad \square$$

PROOF OF PROPOSITION 2.2(a). At first it is shown that

$$\tilde{T}_n := \max_{(j,k)\in\bar{D}_n} \frac{n\Psi(F_n[X_{jn}, X_{kn}], \delta_{jkn})}{\log(\delta_{jkn}^{-1}(1-\delta_{jkn})^{-1})} = O_p(1).$$

It follows from Proposition 2.1 that for $\eta_n > 0$,

$$\mathbb{P}\left\{\max_{(j,k)\in\bar{D}_n} n\Psi(F_n[X_{jn}, X_{kn}], \delta_{jkn}) \geq \eta_n\right\} \leq 2\binom{n+2}{2}\exp(-\eta_n).$$

If $\eta_n := 3\log(n)$, the latter bound tends to zero. Thus for arbitrary fixed $0 < \gamma < 1/2$,

$$\max_{(j,k)\in\bar{D}_n:\, \delta_{jkn}(1-\delta_{jkn})\leq n^{-\gamma}} \frac{n\Psi(F_n[X_{jn}, X_{kn}], \delta_{jkn})}{\log(\delta_{jkn}^{-1}(1-\delta_{jkn})^{-1})} = O_p(1).$$

On the other hand,

$$\max_{(j,k)\in\bar{D}_n:\, \delta_{jkn}(1-\delta_{jkn})\geq n^{-\gamma}} \frac{n\tilde{\Psi}(F_n[X_{jn}, X_{kn}], \delta_{jkn})}{\log(\delta_{jkn}^{-1}(1-\delta_{jkn})^{-1})} = O_p(1), \tag{5}$$

where $\tilde{\Psi}(x, p) := (2p(1-p))^{-1}(x-p)^2$. This follows, for instance, from the Chaining Lemma 5.3 applied to the uniform quantile process $S(j/(n+1)) := (n+1)^{1/2}F_n(X_{jn})$ on $\mathscr{T} := \{j/(n+1)\colon 0 \leq j \leq n+1\}$ equipped with $\rho(s,t) := \mathrm{Var}(B(s,t))^{1/2}$. For elementary calculations, show that $D(u) \leq 2/u^2$ for $0 < u \leq 1$. Further, one can easily deduce from Proposition 2.1 that

$$Q(\eta, \delta) := (2\eta)^{1/2} + \max\{(n+1)^{1/2}\delta, 1\}^{-1}\eta$$

satisfies the assertion of Lemma 5.3. Then elementary calculations show that

$$J(\varepsilon, a) \leq K(a)\big(\varepsilon\log(1/\varepsilon)^{1/2} + n^{-1/2}\log(n)^2\big)$$

for all $\varepsilon \in]0, 1/2]$ and some constant $K(\alpha)$ not depending on n. Alternatively one may deduce (5) from the Hungarian approximation [cf. Shorack and Wellner (1986), Chapter 12.2]. But (5) implies that

$$\max_{(j,k)\in\bar{D}_n:\,\delta_{jkn}(1-\delta_{jkn})\geq n^{-\gamma}} |\mathrm{logit}(F_n[X_{jn}, X_{kn}]) - \mathrm{logit}(\delta_{jkn})| \to_p 0,$$

where $\mathrm{logit}(x) := \log(x/(1-x))$. Elementary calculations show that $\Psi(x, p)/\tilde{\Psi}(x, p) \to 1$ as $\mathrm{logit}(x) - \mathrm{logit}(p) \to 0$. Thus one may replace $\tilde{\Psi}$ in (5) with Ψ and obtains that $\tilde{T}_n = O_p(1)$.

Analogously one can show that

$$\tilde{T}(B) := \sup_{(s,t)\in\Pi(1)} \frac{B(s,t)^2}{2\rho(s,t)^2 \log(\rho(s,t)^{-2})}$$

is finite almost surely.

Now it follows from Lemma 5.1 that for arbitrary $\varepsilon, M > 0$,

$$\begin{aligned}
&\mathbb{E}\, 1\{\tilde{T}_n \leq M\} n^{-2} \sum_{(j,k)\in D_n:\,\delta_{jkn}(1-\delta_{jkn})\leq\varepsilon} \exp(n\Psi(F_n[X_{jn}, X_{kn}], \delta_{jkn})) \\
&\quad \leq n^{-2} \sum_{(j,k)\in D_n:\,\delta_{jkn}(1-\delta_{jkn})\leq\varepsilon} (1 + 2M\log(\delta_{jkn}^{-1}(1-\delta_{jkn})^{-1})) \\
&\quad \to \int_{\{(s,t)\in\Pi(\delta):\,\rho(s,t)^2\leq\varepsilon\}} (1 + 2M\log(\rho(s,t)^{-2}))\,ds\,dt, \\
&\mathbb{E}\, 1\{\tilde{T}(B) \leq M\} \int_{\{(s,t)\in\Pi(\delta):\,\rho(s,t)^2\leq\varepsilon\}} \exp\biggl(\frac{B(s,t)^2}{2\rho(s,t)^2}\biggr)\,ds\,dt \\
&\quad \leq \int_{\{(s,t)\in\Pi(\delta):\,\rho(s,t)^2\leq\varepsilon\}} (1 + 2M\log(\rho(s,t)^{-2}))\,ds\,dt.
\end{aligned}$$

This bound tends to zero as $\varepsilon \downarrow 0$. Moreover, $S(\cdot) = S_n(\cdot)$ converges in distribution to B if it is suitably extended to $S_n \in \mathscr{C}[0, 1]$, whence

$$\begin{aligned}
&n^{-2} \sum_{(j,k)\in D_n:\,\delta_{jkn}(1-\delta_{jkn})\geq\varepsilon} \exp(n\Psi(F_n[X_{jn}, X_{kn}], \delta_{jkn})) \\
&\qquad \to_{\mathscr{L}} \int_{\{(s,t)\in\Pi(\delta):\,\rho(s,t)^2\geq\varepsilon\}} \exp\biggl(\frac{B(s,t)^2}{2\rho(s,t)^2}\biggr)\,ds\,dt.
\end{aligned}$$

[Here we applied Rubin's extended continuous mapping theorem; see Billingsley (1968), Theorem 5.5.] These two facts together imply the asserted convergence in distribution of $T_n(X_n, F_n)$. □

PROOF OF PROPOSITION 2.2(b). Let K_* be a generic real constant depending only on $(D_n)_n$ and possibly different in various (in)equalities. Let F be an arbitrary element of $C_n(X_n, \alpha)$. It follows straightforwardly from part (a) that

$$\Psi(F[X_{jn}, X_{kn}], \delta_{jkn}) \leq 3\rho_n \quad \text{for all } (j,k) \in D_n, \tag{6}$$

provided that $n \ge n_o(\alpha) \ge 2$. This implies that

$$|F[X_{jn}, X_{kn}] - \delta_{jkn}| \le (K_* \delta_{jkn} \rho_n)^{1/2} + K_* \rho_n \quad \text{for all } (j, k) \in \bar{D}_n. \tag{7}$$

It follows from Proposition 2.1 and (6) that (7) holds for D_n in place of $\bar{D}_n$ with $K_* = 6$. Then elementary considerations show that D_n can be replaced with $\bar{D}_n$ if K_* is adjusted properly.

Now let G be another element of $C_n(X_n, \alpha)$ and $J \subset \mathbf{R}$ an interval with $F(J) < G(J)$. Define

$$j = j(J) := \max\{l\colon X_{ln} \le \inf(J)\},$$
$$k = k(J) := \min\{l\colon X_{ln} \ge \sup(J)\}.$$

Then (7) implies that

$$G(J) \le G[X_{jn}, X_{kn}] \le \delta_{jkn} + (K_* \delta_{jkn} \rho_n)^{1/2} + K_* \rho_n, \tag{8}$$

$$F(J) \ge F[X_{j+1,n}, X_{k-1,n}] \ge \delta_{jkn} - 2/n - (K_* \delta_{jkn} \rho_n)^{1/2} - K_* \rho_n, \tag{9}$$

and one easily deduces from (9) that

$$\delta_{jkn} \le 2F(J) + K_* \rho_n. \tag{10}$$

Now subtracting (9) from (8) and plugging in (10) yields

$$G(J) - F(J) \le (K_* F(J) \rho_n)^{1/2} + K_* \rho_n. \qquad \square$$

PROOF OF THEOREM 2.3. Let F and G be the distribution function of f and g, respectively. For arbitrary $a, x, y \in I$ with $a \le x < y$, the monotonicity of f and g, together with Proposition 2.2(b), implies that

$$\begin{aligned}(g(y) - f(x)) \vee (f(y) - g(x)) &\le |G[x, y] - F[x, y]|/(y - x) \\ &\le \big(K_o F[x, y] \rho_n/(y - x)^2\big)^{1/2} + K_o \rho_n/(y - x) \\ &\le (K_o f(a) \rho_n/(y - x))^{1/2} + K_o \rho_n/(y - x),\end{aligned}$$

where we assume throughout that $n \ge n_o(\alpha)$. If $y - x$ is greater than

$$\kappa_n(a) := (K_o f(a) \rho_n)^{1/3} + (K_o \rho_n)^{1/2},$$

then $(K_o f(a) \rho_n/(y - x))^{1/2} + K_o \rho_n/(y - x) \le \kappa_n(a)$. Hence $d(f, g \,|\, I \cap [a, \infty[) \le \kappa_n(a)$.

Now suppose in addition that $f \in \mathscr{F}_{(\beta, L)}(I)$. Then

$$\begin{aligned}&(g(y) - f(y)) \vee (f(x) - g(x)) \\ &\quad \le L(y - x)^\beta + (K_o f(x) \rho_n/(y - x))^{1/2} + K_o \rho_n/(y - x).\end{aligned} \tag{11}$$

Let $x := y - (f(y)\rho_n)^{1/(2\beta+1)} - \rho_n^{1/(\beta+1)}$, assuming that this point is also in I. If $(f(y)\rho_n)^{1/(2\beta+1)} \ge \rho_n^{1/(\beta+1)}$, which is equivalent to $\rho_n \le f(y)^{(\beta+1)/\beta}$, then

$$f(x) \le f(y) + L(y - x)^\beta \le f(y) + L2^\beta (f(y)\rho_n)^{\beta/(2\beta+1)} \le (1 + L2^\beta) f(y),$$

and (11) yields

$$g(y) - f(y) \le \big(L2^\beta + \big(K_o(1 + L2^\beta)\big)^{1/2} + K_o\big)(f(y)\rho_n)^{\beta/(2\beta+1)}.$$

On the other hand, $(f(y)\rho_n)^{1/(2\beta+1)} \le \rho_n^{1/(\beta+1)}$ is equivalent to $f(y) \le \rho_n^{\beta/(\beta+1)}$ and implies that

$$f(x) \le f(y) + L2^\beta \rho_n^{\beta/(\beta+1)} \le (1 + L2^\beta)\rho_n^{\beta/(\beta+1)}.$$

Thus (11) leads to

$$g(y) - f(y) \le \big(L2^\beta + (K_o(1 + L2^\beta))^{1/2} + K_o\big)\rho_n^{\beta/(\beta+1)}.$$

As for the lower bound, let $\gamma := \inf_{y \in I} f(y)$, and suppose that $x + (\gamma\rho_n)^{1/(2\beta+1)} \in I$. Since $f(x) - g(x) \le f(x)$, we may assume that $f(x) \ge K\rho_n^{\beta/(\beta+1)}$ and define $y := x + (f(x)\rho_n/K)^{1/(2\beta+1)}$ for some constant $K \ge 1$ to be specified later. This definition implies that

$$\rho_n^{1/(\beta+1)} \le y - x \le (f(x)/K)^{1/\beta}.$$

If $y \in I$, one can easily deduce from (11) that

$$f(x) - g(x) \le \big(LK^{-\beta/(2\beta+1)} + K_o^{1/2} K^{1/(4\beta+2)}\big)(f(x)\rho_n)^{\beta/(2\beta+1)} + K_o\rho_n^{\beta/(\beta+1)}.$$

It remains to be shown that $y \in I$ for suitable $K = K(\beta, L)$. If $f(x) \le K\gamma$, then $y - x \le (\gamma\rho_n)^{1/(2\beta+1)}$. Otherwise,

$$L(z - x)^\beta \ge (1 - f(z)/f(x))f(x) \ge (1 - 1/K)f(x)$$

for some $z \in I$, $z > x$. Thus if $K := L + 1$, then $z - x \ge (f(x)/K)^{1/\beta} \ge y - x$. □

PROOF OF THEOREM 2.4. According to Proposition 2.2(b), it suffices to show that $m(G_n) \to m(F_o)$ and, in case of (2), $m(G_n) = m(F_o) + O(\rho_n^{1/5})$, where $(G_n)_n$ is an arbitrary sequence of distribution functions in $\mathscr{P}_{\text{uni}}$ with

$$|G_n(J) - F_o(J)| \le (K_o\rho_n F_o(J))^{1/2} + K_o\rho_n$$

for all intervals $J \subset \mathbf{R}$ and any $n > 1$. For fixed $\varepsilon > 0$ there are bounded, nondegenerate intervals $J_1 \le J_2 \le J_3$ (in a pointwise sense) such that

$$\max_{l=1,3} F_o(J_l)/\operatorname{Leb}(J_l) < F_o(J_2)/\operatorname{Leb}(J_2),$$

$$J_1 \cup J_2 \cup J_3 \subset [m(F_o) - \varepsilon, m(F_o) + \varepsilon].$$

It follows from Proposition 2.2(b), that there is an integer n_1 such that

$$\max_{l=1,3} G_n(J_l)/\operatorname{Leb}(J_l) < G_n(J_2)/\operatorname{Leb}(J_2) \quad \text{if } n \ge n_1.$$

But these two inequalities for G_n imply that $m(G_n) \subset J_1 \cup J_2 \cup J_3$.

Suppose that f_o satisfies the regularity condition (2), where $m(F_o) = 0$ without loss of generality. We define

$$J_{n1} := [-2\kappa_n, -\kappa_n], \qquad J_{n2} := [-\kappa_n, \kappa_n], \qquad J_{n3} := [\kappa_n, 2\kappa_n]$$

for a sequence $(\kappa_n)_n$ in $\mathbf{R}^+$ tending to zero. Then

$$\begin{aligned} F_o(J_{nl}) &= f_o(0)\kappa_n - (\gamma + o(1)) \int_{\kappa_n}^{2\kappa_n} x^2\,dx \\ &= f_o(0)\kappa_n - (7/3)\gamma\kappa_n^3 + o(\kappa_n^3) \quad \text{for } l = 1, 3, \\ F_o(J_{n2}) &= 2f_o(0)\kappa_n - 2(\gamma + o(1)) \int_0^{\kappa_n} x^2\,dx \\ &= 2f_o(0)\kappa_n - (2/3)\gamma\kappa_n^3 + o(\kappa_n^3). \end{aligned}$$

Thus for $l = 1, 3$,

$$\begin{aligned} &G_n(J_{n2})/\operatorname{Leb}(J_{n2}) - G_n(J_{nl})/\operatorname{Leb}(J_{nl}) \\ &\quad \geq F_o(J_{n2})/(2\kappa_n) - F_o(J_{nl})/\kappa_n - 2\big((K_o\rho_n F_o(J_{n2}))^{1/2} + K_o\rho_n\big)/\kappa_n \\ &\quad = 2\gamma\kappa_n^2 - 2(2K_of_o(0))^{1/2}\rho_n^{1/2}\kappa_n^{-1/2} + o\big(\kappa_n^2\big) + o\big(\rho_n^{1/2}\kappa_n^{-1/2}\big) \\ &\qquad - 2K_o\rho_n/\kappa_n. \end{aligned} \tag{12}$$

Specifically, let $\kappa_n = K\rho_n^{1/5}$ for some positive number K. Then the bound (12) equals

$$2\big(\gamma K^2 - (2K_of_o(0))^{1/2}K^{-1/2}\big)\rho_n^{2/5} + o\big(\rho_n^{2/5}\big),$$

which is strictly positive if n is greater than some integer n_2, provided that K is greater than $(2K_of_o(0)/\gamma^2)^{1/5}$. Consequently, $M(g)$ is contained in the interval $J_{n1} \cup J_{n2} \cup J_{n3} = [-2K\rho_n^{1/5}, 2K\rho_n^{1/5}]$ for $n \geq n_2$. □

Proof of Proposition 3.1(a). With $\sigma_n(A) := (2\#A)^{-1/2}\sum_{i=1}^n 1\{t_{in} \in A\}\Sigma_{in}$ it follows from Hoeffding's (1963) inequality that $\mathbb{P}\{|\sigma_n(A)| \geq \eta^{1/2}\} \leq 2\exp(-\eta)$ for all $\eta \geq 0$ and $A \in \mathscr{A}_n$. Thus

$$\tilde{\tau}_n := \max_{A \in \mathscr{A}(n)} \sigma_n(A)^2 \leq \log(\#\mathscr{A}_n) + O_p(1).$$

But it follows from Lemma 5.1 that for any positive constant M,

$$\begin{aligned} &\mathbb{E}\,1\{\tilde{\tau}_n \leq \log(\#\mathscr{A}_n) + M\}\tau_n(\Sigma_n) \\ &\quad = \mathbb{E}\,1\{\tilde{\tau}_n \leq \log(\#\mathscr{A}_n) + M\}(\#\mathscr{A}_n)^{-1}\sum_{A \in \mathscr{A}_n}\exp\big(\sigma_n(A)^2\big) \\ &\quad \leq 1 + 2\log(\#\mathscr{A}_n) + 2M. \end{aligned}$$

Under the stronger assumption that $\mathscr{A}_n \subset \mathscr{D} \cap \mathscr{T}_n$ for a fixed VC-class $\mathscr{D}$,

$$\check{\tau}_n := \max_{A \in \mathscr{A}(n)} \frac{\sigma_n(A)^2}{\log(en/\#A)} = O_p(1).$$

This follows from Lemma 5.3 applied to $X(A) := (\#A/n)^{1/2}\sigma_n(A)$, $A \in \{\varnothing\} \cup \mathscr{A}_n$, where $\rho(A, B) := (\#(A \triangle B)/n)^{1/2}$ and $Q(\eta, \delta) := \eta^{1/2}$. The capacity function $D(\varepsilon)$ is bounded by $K(\varepsilon \wedge 1)^{-L}$ for positive constants K, L depending only on $\mathscr{D}$ [cf. Dudley (1978), Lemma 7.13], and $J(\varepsilon, a)$ is not greater than a constant $K(\alpha)$ times $\varepsilon \log(e/\varepsilon)^{1/2}$ for all $\varepsilon \in]0,1]$. Consequently for any fixed $M > 0$,

$$\mathbb{E}\, 1\{\check{\tau}_n \le M\}\tau_n(\Sigma_n) \le (\#\mathscr{A}_n)^{-1} \sum_{A \in \mathscr{A}_n} (1 + 2M \log(en/\#A)),$$

which is bounded by assumption.

Convergence in distribution of $\tau_n(\Sigma_n)$ in case of $(\mathscr{T}_n, \mathscr{A}_n) = (\mathscr{T}_n^{[1]}, \mathscr{A}_n^{[1]}(\delta_n))$ is proved analogously as Proposition 2.2(a), utilizing Donsker's theorem about weak convergence of partial sum processes. □

PROOF OF PROPOSITION 3.1(b). Let

$$Y_n(A, f) = \left((2\#A)^{-1/2} \sum_{i=1}^{n} 1\{t_{in} \in A\}\big(2\{Y_{in} > f(t_{in})\} - 1\big) \right)^{+}.$$

Then $T_n(Y, f) \ge \exp(\max_{A \in \mathscr{A}(n)} Y_n(A, f)^2 - \log(\#\mathscr{A}_n))$, and part (a) entails that for any $\alpha \in]0,1[$ and $n \ge n_o(\alpha)$,

$$Y_n(A, f)^2 \le \kappa \log(\#\mathscr{A}_n) \quad \text{for all } f \in C_n^{(1)}(Y_n, \alpha),\ A \in \mathscr{A}_n, \tag{13}$$

where $\kappa > 1$ is an arbitrary fixed number.

On the other hand, for some sequence $(u_n)_n$ of positive numbers, define $v_n(A) := H((\#A)^{-1/2} u_n)$ and

$$E_n(A) := (2\#A)^{-1/2} \sum_{i=1}^{n} 1\{t_{in} \in A\}\big(2\{E_{in} \ge -v_n(A)\} - 2\mathbb{P}\{E_{in} \ge -v_n(A)\}\big).$$

Then

$$\mathbb{P}\left\{ \max_{A \in \mathscr{A}_n} E_n(A)^2 \ge \kappa \log(\#\mathscr{A}_n) \right\} \to 0,$$

by Hoeffding (1963). Now let $f \in C_n^{(1)}(Y_n, \alpha)$ and $A \in \mathscr{A}_n$ such that $(f_n - f)(t) > v_n(A)$ for all $t \in A$. Then

$$\begin{aligned} Y_n(A, f) &\ge (2\#A)^{-1/2} \sum_{i=1}^{n} 1\{t_{in} \in A\}\big(2\{E_{in} \ge -v_n(A)\} - 1\big) \\ &\ge 2^{-1/2} u_n - \max_{B \in \mathscr{A}_n} |E_n(B)| \\ &\ge 2^{-1/2} u_n - (\kappa \log(\#\mathscr{A}_n))^{1/2} \end{aligned}$$

with asymptotic probability 1. But this is not compatible with (13) unless

$$u_n \le (8\kappa \log(\#\mathscr{A}_n))^{1/2}.$$

Hence $\max_{t \in A}(f - f_n)(t)$ is greater than $H((8\kappa \log(\#\mathscr{A}_n)/\#A)^{1/2})$ for arbitrary $A \in \mathscr{A}_n$ and $f \in C_n^{(1)}(Y_n, \alpha)$ with probability tending to 1.

Analogous arguments apply to $\min_{t \in A}(f - f_n)(t)$. □

PROOF OF THEOREMS 3.2 AND 3.3. Because of Proposition 3.1(b), we consider an arbitrary fixed sequence $(g_n)_{n>1}$ of functions on $\mathbf{R}^d$ such that

$$\min_{t\in A}(g_n - f_n)(t) \vee \min_{t\in A}(f_n - g_n)(t) \\ \leq H((K\log(n)/\#A)^{1/2}) \quad \text{for all } A \in \mathscr{A}_n^{[d]}(\delta_n),$$

where $K := 18$.

As for Theorem 3.2, let $f_n,\ g_n \in \mathscr{F}_{\uparrow}([0,1]^d)$. It suffices to show that $d(g_n, f_n) = O(\kappa_n)$, where $\kappa_n := \rho_n^{1/(d+2)}$. For that purpose let $s,\ s+\varepsilon\mathbf{1} \in [0,1]^d$ with $\varepsilon \geq 2\kappa_n$. For n greater than some $n_1((\delta_n)_n)$ there exists an $A \in \mathscr{A}_n^{[d]}(\delta_n)$ with $A \subset [s, s+\varepsilon\mathbf{1}]$ and $\#A/n \geq \kappa_n$. By monotonicity of f_n and g_n,

$$(g_n(s) - f_n(s+\varepsilon\mathbf{1})) \vee (f_n(s) - g_n(s+\varepsilon\mathbf{1})) \\ \leq \min_{t\in A}(g_n - f_n)(t) \vee \min_{t\in A}(f_n - g_n)(t) \leq H((K\rho_n/\kappa_n)^{1/2}) = O(\kappa_n).$$

As for Theorem 3.3, let $f_n,\ g_n \in \mathscr{F}_{\mathrm{cc}}[0,1]$ and $f_n \in \mathscr{F}_{(\beta,L)}[a,b]$. With $I_n := [a+\rho_n^{1/(2\beta+1)}, b-\rho_n^{1/(2\beta+1)}]$ it suffices to show that $\|g_n - f_n\|_{I(n)} = O(\rho_n^{\beta/(2\beta+1)})$. Suppose that $|g_n - f_n|(x_o) \geq R\rho_n^{\beta/(2\beta+1)}$ for some constant $R \geq 2$ and some $x_o \in I_n$. Then Lemma 5.2(b) implies that for n greater than some $n_o(a,b,\beta,L)$ there exists an $A \in \mathscr{A}_n^{[1]}(\delta_n)$ with

$$\#A/n \geq K_{(\beta,L)}\rho_n^{1/(2\beta+1)}/2,$$
$$\min_{t\in A}(g_n - f_n)(t) \vee \min_{t\in A}(f_n - g_n)(t) \geq (R/4)\rho_n^{\beta/(2\beta+1)}.$$

Thus

$$(R/4)\rho_n^{\beta/(2\beta+1)} \leq H\big((32\rho_n^{2\beta/(2\beta+1)}/K_{(\beta,L)})^{1/2}\big),$$

that means, $R \leq R_1(H,\beta,L)$ for $n \geq n_1(a,b,\beta,L,H)$. □

PROOF OF PROPOSITION 4.1. As in the proof of Proposition 3.1, one can apply Hoeffding's (1963) inequality, Lemma 5.3 and Lemma 5.1 in order to show that

$$\max_{(s,t)\in\Pi(\varepsilon)} Z_n(s,t\,|\,F_n)^2 \to_p 0$$

and

$$n^{-2}\sum_{(s,t)\in\Pi_n(\varepsilon)} \exp\big(Z_n(s,t\,|\,F_n)^2/(2(t-s))\big) \to_p 0 \quad \text{as } n\to\infty,\ \varepsilon\downarrow 0.$$

The same is true, if $Z_n(t\,|\,F_n)$ is replaced with $W(t)$ or

$$W_n(t) := 2n^{-1/2}\sum_{i\leq nt} W_{in},$$

where W_{in}, $1 \le i \le n$, are independent Gaussian random variables with mean zero and variance $F_n(r_{in})(1 - F_n(r_{in})) \le 1/4$. Now,

$$\limsup_{n\to\infty} \mathbb{P}\{T_n(Z_n(\cdot \mid F_n)) \ge d_n(\alpha)\} \le \limsup_{n\to\infty} \mathbb{P}\{T_n(W_n) \ge d_n(\alpha) + \varepsilon\}$$
$$\le \limsup_{n\to\infty} \mathbb{P}\{T_n(W) \ge d_n(\alpha) + \varepsilon\}$$

for any fixed $\varepsilon > 0$. The first inequality follows via standard approximation arguments. The second inequality follows from Anderson's (1955) lemma, because $T_n(\cdot)$ is convex and $\mathrm{Var}(\sum_{i=1}^n h(i/n)W_n(i/n)) \le \mathrm{Var}(\sum_{i=1}^n h(i/n)W(i/n))$ for arbitrary functions h on $[0, 1]$. Furthermore, $\mathscr{L}(T_n(W))$ and $\mathscr{L}_*(T_n(Z_n \times (\cdot \mid F_n)))$ converge weakly to $\mathscr{L}(T(W))$, where $\mathscr{L}_*$ denotes the distribution in case of $F_n(r_{1n}) = F_n(r_{nn}) = 1/2$. Since $\mathscr{L}(T(W))$ is continuous and has connected support, $\lim_{n\to\infty} d_n(\alpha) = c(\alpha)$ and

$$\limsup_{n\to\infty} \mathbb{P}\{T_n(Z_n(\cdot \mid F_n)) \ge d_n(\alpha)\} \le \mathbb{P}\{T(W) \ge c(\alpha) + \varepsilon\}$$
$$\to \mathbb{P}\{T(W) \ge c(\alpha)\} = \alpha, \qquad \varepsilon \downarrow 0.$$

It follows from part (a) that for arbitrary fixed $\gamma > 2$ and $\alpha \in]0, 1[$ the event

$$A_n := \Big\{ \sup_{F \in C_n(Z_n, \alpha) \cup \{F_n\}} Z_n(s, t \mid F)^2 > \gamma(t - s)\log(n)$$
$$\text{for some } (s, t) \in \Pi_n(\delta_n)\Big\}$$

has asymptotic probability 0. For $-\infty \le a < b \le \infty$, let $(s, t) \in \Pi_n(\delta_n)$ such that $[r_{ns+1,n}, r_{nt,n}] \subset [a, b]$ and $t - s = \mu_n[a, b] \wedge \delta_n$, where $n\delta_n$ is assumed to be an integer. Then outside from A_n for any $F \in C_n(Z_n, \alpha)$,

$$(F(a) - F_n(b)) \vee (F_n(b) - F(a))$$
$$\le (F(r_{ns+1,n}) - F_n(r_{nt,n})) \vee (F_n(r_{ns+1,n}) - F(r_{nt,n}))$$
$$\le (|Z_n(s, t \mid F)| + |Z_n(s, t \mid F_n)|)/(2n^{1/2}(t - s))$$
$$\le (\gamma\delta_n^{-1}\rho_n/\mu_n[a, b])^{1/2}. \qquad \square$$

PROOF OF THEOREM 4.2. In view of Proposition 4.1(b), we consider an arbitrary fixed sequence of distribution functions G_n such that

$$(G_n(s) - F_n(t)) \vee (F_n(s) - G_n(t)) \le (K\rho_n/\mu_n[s, t])^{1/2} \tag{14}$$

for $s < t$ and arbitrary $n > 1$. Then it suffices to show that $\|G_n - F_o\|_{\mathbf{R}} \to 0$ in part (a), and that $\|G_n - F_n\|_I = O(\rho_n^{\beta/(2\beta+1)})$ in part (b).

As for part (a), by continuity of F_o one has only to verify pointwise convergence of $(G_n)_n$. For any $t \in \mathbf{R}$ and $\varepsilon > 0$ there exists $s < t$ with $0 < F_o[s, t] < \varepsilon$. Weak convergence of $(\mu_n)_n$ to μ and support$(\mu) \supset$ support(F_o) together imply that $\liminf_{n\to\infty} \mu_n(]s, t[) \ge \mu(]s, t[) > 0$. Hence

$$G_n(t) \ge F_o(s) + o(1) \ge F_o(t) - \varepsilon + o(1).$$

Analogously, one shows that $G_n(t) \leq F_o(t) + \varepsilon + o(1)$ for any fixed $\varepsilon > 0$, whence $(G_n(t))_n$ tends to $F_o(t)$.

Under the assumptions of part (b), (14) entails

$$(G_n - F_n)(s) \leq L(t-s)^\beta + (K\rho_n/\mu_n[s,t])^{1/2} \quad \text{and} \\ (G_n - F_n)(t) \geq -L(t-s)^\beta - (K\rho_n/\mu_n[s,t])^{1/2} \tag{15}$$

for $0 \leq s < t$ and arbitrary $n > 1$. Let $\kappa_n := \rho_n^{1/(2\beta+1)}$. Continuity of $\mu^{(1)}$ and compactness of $I \subset \{\mu^{(1)} > 0\}$ together imply that

$$\inf_{s,\,t\geq 0:\,[s,\,t]\cap I\neq\varnothing,\,t-s=\kappa_n} \mu[s,t] \geq (\gamma + o(1))\kappa_n$$

for some $\gamma > 0$. Since $\kappa_n/\rho_n \to \infty$,

$$\sup_{s,\,t\geq 0:\,[s,\,t]\cap I\neq\varnothing,\,t-s=\kappa_n} |\mu_n[s,t]/\mu[s,t] - 1| \to_p 0,$$

which is well known from empirical process theory (see also Proposition 2.2). Hence

$$\sup_{s\in I}(G_n - F_n)(s) \vee \sup_{t\in I:\,t\geq\kappa_n}(F_n - G_n)(t)$$
$$\leq L\kappa_n^\beta + \big(K\rho_n\kappa_n^{-1}/(\gamma + o_p(1))\big)^{1/2} = O_p\big(\rho_n^{\beta/(2\beta+1)}\big),$$

while $(F_n - G_n)(t) \leq F_n(t) \leq L\kappa_n^\beta = L\rho_n^{\beta/(2\beta+1)}$ for all $t \in [0, \kappa_n]$. □

Some final remarks. For the sake of simplicity we confined our attention to one particular type of test statistic. Obviously there is some arbitrariness in this definition. For instance, the results for the white noise model remain valid, if $T(h)$ is replaced with

$$T_{a,\gamma}(h) := \int_{\Pi(\delta)} (t-s)^a \exp\left(\frac{\gamma h(s,t)^2}{2(t-s)}\right) ds\,dt,$$

where $\gamma > 0$ and $a - (\gamma - 1)^+ > -1$, or with the statistic $\check{T}(h)$ defined at the beginning of Section 5. A very early version of the present paper treated only test statistics similar to $\check{T}(h)$, but numerical examples showed that confidence sets based on maximum type test statistics are often larger than those considered here. Using an exponent $\gamma \geq 1$ for $T_{a,\gamma}(h)$ increases sensitivity to deviations on small intervals. The only price to pay for using such an exponent is the extra arguments involving Lemma 5.1.

Acknowledgments. This paper grew out of discussions with Laurie Davies and Nick Hengartner. I am also grateful to Enno Mammen, D. W. Müller and Jon Wellner for stimulating conversations. Section 4 is a tentative answer to Jon's question about applications to interval-censored data. The computations were made in *Oberon* using code of *Voyager* [Sawitzki (1996)]. Many thanks to Felix Friedrich and Günter Sawitzki for answering numerous questions about computing.

REFERENCES

ANDERSON, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proc. Amer. Math. Soc.* **6** 170–176.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

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

DAVIES, P. L. (1995). Data features. *Statist. Neerlandica* **49** 185–245.

DONOHO, D. L. (1988). One-sided inference about functionals of a density. *Ann. Statist.* **16** 1390–1420.

DUDLEY, R. M. (1978). Central limit theorems for empirical measures. *Ann. Probab.* **6** 899–929.

GROENEBOOM, P. and WELLNER, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimators.* Birkhäuser, Basel.

HENGARTNER, N. W. and STARK, P. B. (1995). Finite-sample confidence envelopes for shape-restricted densities. *Ann. Statist.* **23** 525–550.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

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

KHAS'MINSKII, R. Z. (1978). A lower bound on the risks of nonparametric estimates of densities in the uniform metric. *Theory Probab. Appl.* **23** 794–798.

KHAS'MINSKII, R. Z. (1979). Lower bounds for the risk of nonparametric estimates of the mode. In *Contributions to Statistics, Jaroslav Hájek Memorial Volume* (J. Jureckova, ed.) 91–97. Academia, Prague.

LOW, M. G. (1997). Personal communication.

MAMMEN, E. (1991). Nonparametric regression under qualitative smoothness assumptions. *Ann. Statist.* **19** 741–759.

MARRON, J. S. and TSYBAKOV, A. B. (1995). Visual error criteria for qualitative smoothing. *J. Amer. Statist. Assoc.* **90** 499–507.

MÜLLER, D. W. (1991). Quantile regression. Preprint 624, Univ. Heidelberg, Sonderforschungsbereich 123.

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

POLLARD, D. (1984). *Convergence of Stochastic Processes.* Springer, New York.

ROMANO, J. P. (1988). On weak convergence and optimality of kernel density estimates of the mode. *Ann. Statist.* **16** 629–647.

SAWITZKI, G. (1996). Extensible statistical software: on a voyage to Oberon. *J. Comput. Graph. Statist.* **5** 263–283.

SHORACK, G. R. and WELLNER, J. A. (1982). Limit theorems and inequalities for the uniform empirical process indexed by intervals. *Ann. Probab.* **10** 639–652.

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

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

INSTITUT FÜR ANGEWANDTE MATHEMATIK
UNIVERSITÄT HEIDELBERG
IM NEUENHEIMER FELD 294
69120 HEIDELBERG
GERMANY
E-MAIL: lutz@statlab.uni-heidelberg.de

INSTITUT FÜR MATHEMATIK
MEDIZINISCHE UNIVERSITÄT ZU LÜBECK
WALLSTRAßE 40
23560 LÜBECK
GERMANY
E-MAIL: duembgen@informatik.mu-luebeck.de