Optimal Confidence Bands for Shape-Restricted Curves

Lutz Dümbgen

Department of Mathematical Statistics and Actuarial Science
University of Bern
Sidlerstrasse 5, CH-3012 Bern, Switzerland

E-mail: duembgen@stat.unibe.ch, URL: www.imsv.unibe.ch

August 2002, updated December 2013

This paper has been published in *Bernoulli* ***9***, 423–449 (2003).
The present version corrects two typos in the published version
and contains updated references.

Abstract

Let Y be a stochastic process on $[0,1]$ satisfying $dY(t) = n^{1/2}f(t)dt + dW(t)$, where $n \geq 1$ is a given scale parameter ("sample size"), W is standard Brownian motion and f is an unknown function. Utilizing suitable multiscale tests we construct confidence bands for f with guaranteed given coverage probability, assuming that f is isotonic or convex. These confidence bands are computationally feasible and shown to be asymptotically sharp optimal in an appropriate sense.

Running title. Confidence Bands for Shape-Restricted Curves

Keywords and phrases. adaptivity, concave, convex, isotonic, kernel estimator, local smoothness, minimax bounds, multiscale testing

1 Introduction

Nonparametric statistical models often involve some unknown function f defined on a real interval J. For instance f might be the probability density of some distribution or a regression function. Nonparametric point estimators for such a curve f are abundant. The available methods are based on kernels, splines, local polynomials, or orthogonal series, including wavelets; see Hart (1997) and references cited therein. In order to quantify the precision of estimation, one often wants to replace a point estimator with a confidence band $(\hat{\ell}, \hat{u})$ for f. The latter consists of two functions $\hat{\ell} = \hat{\ell}(\cdot, \text{data})$ and $\hat{u} = \hat{u}(\cdot, \text{data})$ on J with values in $[-\infty, \infty]$ such that, hopefully, $\hat{\ell} \leq f \leq \hat{u}$ pointwise. More precisely, one is aiming at a confidence band such that

$$\mathbf{P}\{\hat{\ell} \leq f \leq \hat{u}\} \ \geq \ 1 - \alpha \tag{1}$$

for a given level $\alpha \in \,]0, 1[$, while $\hat{\ell}$ and $\hat{u}$ should be as close to each other as possible.

Unfortunately, curve estimation is an ill-posed problem, and usually there are no nontrivial bands $(\hat{\ell}, \hat{u})$ satisfying (1) for arbitrary f; see Donoho (1988). Therefore one has to impose some additional restrictions on f. One possibility are smoothness constraints on f, for instance an upper bound on a certain derivative of f. Under such restrictions, (1) can be achieved approximately for large sample sizes; see for example Bickel and Rosenblatt (1973), Knafl et al. (1985), Hall and Titterington (1988), Härdle and Marron (1991), Eubank and Speckman (1993), Fan and Zhang (2000), and the references cited therein.

A problem with the aforementioned methods is that smoothness constraints are hard to justify in practical situations. More precisely, even if the underlying curve f is infinitely often differentiable, the actual coverage probabilities of the confidence bands mentioned above depend on quantitative properties of certain derivatives of f which are difficult to obtain from the data.

In many applications qualitative assumptions about f such as monotonicity, unimodality or concavity/convexity are plausible. One example are growth curves in medicine, e.g. where $f(x)$ is the mean body height of newborns at age x. Here isotonicity of f is a plausible assumption. Another example are so-called Engel curves in econometrics, where $f(x)$ is the mean expenditure for certain consumer goods of households with annual income x. Here one expects f to be isotonic and sometimes concave as well. Under such qualitative assumptions it is possible to construct $(1 - \alpha)$–confidence sets for f based on certain goodness-of-fit tests without relying on asymptotic arguments. Examples for such procedures can be found in Davies (1995), Hengartner and Stark (1995) and Dümbgen (1998). In particular, these papers present confidence bands $(\hat{\ell}, \hat{u})$ for

f such that

$$\mathrm{P}\{\hat{\ell} \le f \le \hat{u}\} \ \ge \ 1 - \alpha \quad \text{whenever } f \in \mathcal{F}. \tag{2}$$

Here $\mathcal{F}$ denotes the specified class of functions. Given a suitable distance measure $D(\cdot,\cdot)$ for functions, the goal is to find a band $(\hat{\ell}, \hat{u})$ satisfying (2) such that either $D(\hat{u}, \hat{\ell})$ or $D(\hat{\ell}, f)$ and $D(\hat{u}, f)$ are as small as possible. The phrase "as small as possible" can be interpreted in the sense of optimal rates of convergence to zero as the sample size n tends to infinity. The papers of Hengartner and Stark (1995) and Dümbgen (1998) contain such optimality results.

In the present paper we investigate optimality of confidence bands in more detail. In addition to optimal rates of convergence we obtain optimal constants and discuss the impact of local smoothness properties of f. Compared to the general confidence sets of Dümbgen (1998), the methods developed here are more stringent and computationally simpler. They are based on multiscale tests as developed by Dümbgen and Spokoiny (2001), who considered tests of qualitative assumptions rather than confidence bands. For further results on testing in nonparametric curve estimation see Hart (1997), Fan et al. (2001), and the references cited there.

2 Basic setting and overview

For mathematical convenience we focus on a continuous white noise model: Suppose that one observes a stochastic process Y on the unit interval $[0, 1]$, where

$$Y(t) \ = \ n^{1/2} \int_0^t f(x)\, dx + W(t).$$

Here f is an unknown function in $L^2[0, 1]$, $n \ge 1$ is a given scale parameter ("sample size"), and W is standard Brownian motion. In this context the bounding functions $\hat{\ell}, \hat{u}$ are defined on $[0, 1]$, but for notational convenience the function f is tacitly assumed to be defined on the whole real line with values in $[-\infty, \infty]$. From now on we assume that

$$f \ \in \ \mathcal{G} \cap L^2[0, 1],$$

where $\mathcal{G}$ denotes one of the following two function classes:

$$\begin{aligned} \mathcal{G}_\uparrow \ &:= \ \Big\{\text{non-decreasing functions } g : \mathbb{R} \to [-\infty, \infty]\Big\}, \\ \mathcal{G}_{\mathrm{conv}} \ &:= \ \Big\{\text{convex functions } g : \mathbb{R} \to \,]-\infty, \infty]\Big\}. \end{aligned}$$

The paper is organized as follows. In Section 3 we treat the case $\mathcal{G} = \mathcal{G}_\uparrow$ and measure the quality of a confidence band $(\hat{\ell}, \hat{u})$ by quantities related to the Levy distance $d_{\mathrm{L}}(\hat{\ell}, \hat{u})$. Generally,

$$d_{\mathrm{L}}(g, h) \ := \ \inf\Big\{\epsilon > 0 : g \le h(\cdot + \epsilon) + \epsilon \text{ and } h \le g(\cdot + \epsilon) + \epsilon \text{ on } [0, 1 - \epsilon]\Big\}$$

for isotonic functions $g, h : [0,1] \to [-\infty, \infty]$. It turns out that a confidence band which is based on a suitable multiscale test as introduced by Dümbgen and Spokoiny (2001) is asymptotically optimal in a strong sense. Throughout this paper asymptotic statements refer to $n \to \infty$, unless stated otherwise.

In Section 4 we treat both classes $\mathcal{G}_\uparrow$ and $\mathcal{G}_{\rm conv}$ simultaneously. We discuss the construction of confidence bands $(\hat{\ell}, \hat{u})$ satisfying (2) such that $D(\hat{\ell}, f)$ and $D(f, \hat{u})$ are as small as possible whenever f satisfies some additional smoothness constraints. Here $D(g, h)$ is a distance measure of the form

$$D(g,h) \ := \ \sup_{x \in [0,1]} w(x,f)(h(x) - g(x))$$

for some weight function $w(\cdot, f) \geq 0$ reflecting local smoothness properties of f. Again it turns out that suitable multiscale procedures yield nearly optimal procedures without additional prior information on f.

In Section 5 we present some numerical examples for the procedures of Section 4. The proofs are deferred to Sections 6, 7 and 8. In particular, Section 7 contains a new minimax bound for confidence rectangles in a gaussian shift model, which may be of independent interest.

As for the white noise model, the results of Brown and Low (1996), Nussbaum (1996) and Grama and Nussbaum (1998) on asymptotic equivalence can be used to transfer the lower bounds of the present paper to other models. Moreover, one can mimick the confidence bands developed here in traditional regression models under minimal assumptions; see Dümbgen and Johns (2004) and Dümbgen (2007).

3 Optimality for isotonic functions in terms of Lévy type distances

In this section we consider the class $\mathcal{G}_\uparrow$. For isotonic functions $g, h : [0,1] \to [-\infty, \infty]$ and $\epsilon > 0$ let

$$D_\epsilon(g,h) \ := \ \inf \Big\{ \lambda \geq 0 : g \leq h(\cdot + \epsilon) + \lambda \text{ and } h \leq g(\cdot + \epsilon) + \lambda \text{ on } [0, 1-\epsilon] \Big\}.$$

Then the Lévy distance $d_{\rm L}(g,h)$ is the infimum of all $\epsilon > 0$ such that $D_\epsilon(g,h) \leq \epsilon$. We use these functionals $D_\epsilon(\cdot,\cdot)$ in order to quantify differences between isotonic functions. Figure 1 depicts one such function g, and the shaded areas represent the set of all functions h with $D_{0.05}(g,h) \leq 0.1$ and $D_{0.05}(g,h) \leq 0.025$, respectively.

The next theorem provides lower bounds for $D_\epsilon(\hat{\ell}, \hat{u})$, $0 < \epsilon \leq 1$. Here and throughout the sequel the dependence of probabilities, expectations and distributions on the functional parameter

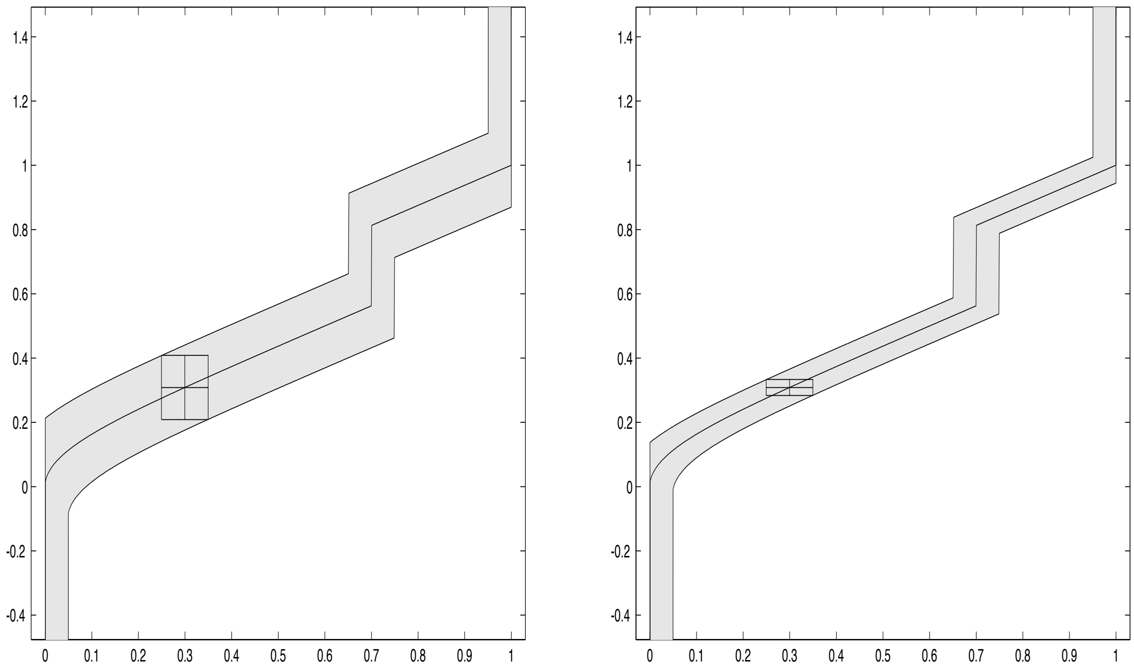

Figure 1: Two $D_{0.05}(\cdot,\cdot)$–neighborhoods of some function g.

f is sometimes indicated by a subscript f.

Theorem 3.1. *There exists a universal function b on $]0,1]$ with $\lim_{\epsilon\downarrow 0} b(\epsilon) = 0$ such that*

$$\inf_{f\in\mathcal{G}_\uparrow\cap L^2[0,1]} \mathrm{P}_f \left\{ \hat{\ell} \le f \le \hat{u} \text{ and } D_\epsilon(\hat{\ell},\hat{u}) < \frac{(8\log(e/\epsilon))^{1/2} - b(\epsilon)}{(n\epsilon)^{1/2}} \right\} \;\le\; b(\epsilon)$$

for any confidence band $(\hat{\ell},\hat{u})$ and arbitrary $\epsilon \in]0,1]$.

Theorem 3.1 entails a lower bound for $d_{\mathrm{L}}(\hat{\ell},\hat{u})$. For let $\epsilon = \epsilon_n := c\,(\log(n)/n)^{1/3} - \delta n^{-1/3}$ with any fixed $c,\delta > 0$. Then one can show that for sufficiently large n,

$$\frac{(8\log(e/\epsilon))^{1/2} - b(\epsilon)}{(n\epsilon)^{1/2}} \;=\; \Bigl(\frac{8}{3c}\Bigr)^{1/2}\Bigl(\frac{\log n}{n}\Bigr)^{1/3} + o(n^{-1/3}) \;\ge\; \epsilon,$$

provided that c equals $(8/3)^{1/3} \approx 1.387$.

Corollary 3.2. *For each $n \ge 1$ there exists a universal constant β_n such that $\beta_n \to 0$ and*

$$\inf_{f\in\mathcal{G}_\uparrow\cap L^2[0,1]} \mathrm{P}_f \left\{ \hat{\ell} \le f \le \hat{u} \text{ and } d_{\mathrm{L}}(\hat{\ell},\hat{u}) < \Bigl(\frac{8}{3}\Bigr)^{1/3}\Bigl(\frac{\log n}{n}\Bigr)^{1/3} - \beta_n n^{-1/3} \right\} \;\le\; \beta_n$$

for any confidence band $(\hat{\ell},\hat{u})$. $\square$

It is possible to get close to these lower bounds for $D_\epsilon(\hat{\ell},\hat{u})$ *simultaneously* for all $\epsilon \in]0,1]$

while (2) is satisfied. For let κ_α be a real number such that

$$\mathrm{P}\left\{\frac{|W(t)-W(s)|}{(t-s)^{1/2}} \le \Gamma(t-s)+\kappa_\alpha \text{ for } 0 \le s < t \le 1\right\} \ \le\ \alpha,$$

where

$$\Gamma(u) \ := \ (2\log(e/u))^{1/2} \quad \text{for } 0 < u \le 1.$$

The existence of such a critical value κ_α follows from Dümbgen and Spokoiny (2001, Theorem 2.1). With the local averages

$$F_f(s,t) \ := \ \frac{1}{t-s}\int_s^t f(x)\,dx$$

of f and their natural estimators

$$\hat{F}(s,t) \ := \ \frac{Y(t)-Y(s)}{n^{1/2}(t-s)}$$

it follows that

$$\mathrm{P}_f\left\{\left|\hat{F}(s,t)-F_f(s,t)\right| \le \frac{\Gamma(t-s)+\kappa_\alpha}{(n(t-s))^{1/2}} \text{ for } 0 \le s < t \le 1\right\} \ \ge\ 1-\alpha.$$

But for $0 \le s < t \le 1$,

$$f(s) \ \le\ F_f(s,t) \ \le\ f(t) \quad \text{whenever } f \in \mathcal{G}_\uparrow.$$

This implies the first assertion of the following theorem.

Theorem 3.3. *With the critical value κ_α above let*

$$\begin{aligned}
\hat{\ell}(x) \ &:= \ \sup_{0 \le s < t \le x}\Big(\hat{F}(s,t) - \frac{\Gamma(t-s)+\kappa_\alpha}{\sqrt{n(t-s)}}\Big),\\
\hat{u}(x) \ &:= \ \inf_{x \le s < t \le 1}\Big(\hat{F}(s,t) + \frac{\Gamma(t-s)+\kappa_\alpha}{\sqrt{n(t-s)}}\Big).
\end{aligned}$$

This defines a confidence band $(\hat{\ell},\hat{u})$ for f satisfying (2) with $\mathcal{F} = \mathcal{G}_\uparrow \cap L^2[0,1]$. Moreover, in case of $\hat{\ell} \le \hat{u}$,

$$\begin{aligned}
D_\epsilon(\hat{\ell},\hat{u}) \ &\le\ \frac{(8\log(e/\epsilon))^{1/2}+2\kappa_\alpha}{(n\epsilon)^{1/2}} \quad \text{for } 0 < \epsilon \le 1,\\
d_{\mathrm{L}}(\hat{\ell},\hat{u}) \ &\le\ \Big(\frac{8}{3}\Big)^{1/3}\Big(\frac{\log n}{n}\Big)^{1/3} + o(n^{-1/3}).
\end{aligned}$$

Proof. The preceding upper bound for $D_\epsilon(\hat{\ell},\hat{u})$ follows from the fact that for any $x \in [0,1-\epsilon]$,

$$\begin{aligned}
\hat{u}(x)-\hat{\ell}(x+\epsilon) \ &\le\ \Big(\hat{F}(x,x+\epsilon)+\frac{\Gamma(\epsilon)+\kappa_\alpha}{(n\epsilon)^{1/2}}\Big) - \Big(\hat{F}(x,x+\epsilon)-\frac{\Gamma(\epsilon)+\kappa_\alpha}{(n\epsilon)^{1/2}}\Big)\\
&=\ \frac{2\Gamma(\epsilon)+2\kappa_\alpha}{(n\epsilon)^{1/2}}\\
&=\ \frac{(8\log(e/\epsilon))^{1/2}+2\kappa_\alpha}{(n\epsilon)^{1/2}}.
\end{aligned}$$

Letting $\epsilon = \epsilon_n = (8/3)^{1/3}(\log(n)/n)^{1/3}$ yields the upper bound for $d_{\mathrm{L}}(\hat{\ell},\hat{u})$. $\square$

4 Bands for potentially smooth functions

A possible criticism of the preceding results is the fact that the minimax bounds are attained at special step functions. On the other hand one often expects the underlying curve f to be smooth in some vague sense. Therefore we aim now at confidence bands satisfying (2) with $\mathcal{F} = \mathcal{G} \cap L^2[0,1]$, which are as small as possible whenever f satisfies some additional smoothness conditions. Throughout $\mathcal{G}$ stands for $\mathcal{G}_\uparrow$ or $\mathcal{G}_{\rm conv}$.

In the sequel let $\langle g, h\rangle := \int_{-\infty}^{\infty} g(x)h(x)\,dx$ and $\|g\| := \langle g, g\rangle^{1/2}$ for measurable functions g, h on the real line such that these integrals are defined. The confidence bands to be presented here can be described either in terms of kernel estimators for f or in terms of tests. Both viewpoints have their own merits.

4.1 Kernel estimators for f

Let ψ be some kernel function in $L^2(\mathbb{R})$. For technical reasons we assume that ψ satisfies the following three regularity conditions:

$$
\left\{ \begin{array}{l} \psi \text{ has bounded total variation;} \\ \psi \text{ is supported by } [-a, b], \text{ where } a, b \geq 0; \\ \langle 1, \psi\rangle \ > \ 0. \end{array} \right. \tag{3}
$$

For any bandwidth $h > 0$ and location parameter $t \in \mathbb{R}$ let

$$
\psi_{h,t}(x) \ := \ \psi\Big(\frac{x-t}{h}\Big).
$$

Then $\langle g, \psi_{h,t}\rangle = h\,\langle g(t + h\,\cdot), \psi\rangle$ and $\|\psi_{h,t}\| = h^{1/2}\|\psi\|$. A kernel estimator for $f(t)$ with kernel function ψ and bandwidth h is given by

$$
\hat{f}_h(t) \ := \ \frac{\psi Y(h,t)}{n^{1/2} h\,\langle 1, \psi\rangle},
$$

where

$$
\psi Y(h,t) \ := \ \int_0^1 \psi_{h,t}(x)\,dY(x).
$$

From now on suppose that $ah \leq t \leq 1 - bh$. Then $\psi_{h,t}$ is supported by $[0,1]$ and one may write

$$
\begin{aligned}
\mathrm{E}\hat{f}_h(t) \ &= \ \frac{\langle f, \psi_{t,h}\rangle}{h\,\langle 1, \psi\rangle} \ = \ \frac{\langle f(t + h\,\cdot), \psi\rangle}{\langle 1, \psi\rangle}, \\
\mathrm{Var}(\hat{f}_h(t)) \ &= \ \frac{\|\psi_{t,h}\|^2\rangle}{nh^2\langle 1, \psi\rangle^2} \ = \ \frac{\|\psi\|^2}{nh\,\langle 1, \psi\rangle^2}.
\end{aligned}
$$

The random fluctuations of these kernel estimators can be bounded uniformly in $h > 0$. For that purpose we define the multiscale statistic

$$\begin{aligned} T(\pm\psi) &:= \sup_{h>0} \sup_{t\in[ah,1-bh]} \Big(\frac{\pm\psi W(h,t)}{h^{1/2}\|\psi\|} - \Gamma((a+b)h) \Big) \\ &= \sup_{h>0} \sup_{t\in[ah,1-bh]} \Big(\pm \frac{\hat f_h(t) - \mathrm{E}\hat f_h(t)}{\mathrm{Var}(\hat f_h(t))^{1/2}} - \Gamma((a+b)h) \Big), \end{aligned}$$

similarly as in Dümbgen and Spokoiny (2001). It follows from Theorem 2.1 in the latter paper, that $0 \le T(\pm\psi) < \infty$ almost surely. In particular, $|\hat f_h(t) - \mathrm{E}\hat f_h(t)| \le (nh)^{-1/2}\log(e/h)^{1/2}O_p(1)$, uniformly in $h > 0$ and $ah \le t \le 1 - bh$.

It is well-known that kernel estimators are biased in general. But our shape restrictions may be used to construct two kernel estimators whose bias is always non-positive or non-negative, respectively. Precisely, let $\psi^{(\ell)}$ and $\psi^{(u)}$ be two kernel functions satisfying (3) with respective supports $[-a^{(\ell)}, b^{(\ell)}]$ and $[-a^{(u)}, b^{(u)}]$. In addition suppose that

$$\langle g, \psi^{(\ell)}\rangle \le g(0)\langle 1, \psi^{(\ell)}\rangle \quad \text{for all } g \in \mathcal{G} \cap L^2[-a^{(\ell)}, b^{(\ell)}], \tag{4}$$

$$\langle g, \psi^{(u)}\rangle \ge g(0)\langle 1, \psi^{(u)}\rangle \quad \text{for all } g \in \mathcal{G} \cap L^2[-a^{(u)}, b^{(u)}]. \tag{5}$$

These inequalities imply that the corresponding kernel estimators satisfy the inequalities $\mathrm{E}\hat f_h^{(\ell)}(t) \le f(t) \le \mathrm{E}\hat f_h^{(u)}(t)$, and the definition of $T(\pm\psi)$ yields that

$$f(t) \ge \hat f_h^{(\ell)}(t) - \frac{\|\psi^{(\ell)}\|\Big(\Gamma(d^{(\ell)}h) + T(\psi^{(\ell)})\Big)}{\langle 1, \psi^{(\ell)}\rangle (nh)^{1/2}}, \tag{6}$$

$$f(t) \le \hat f_h^{(u)}(t) + \frac{\|\psi^{(u)}\|\Big(\Gamma(d^{(u)}h) + T(-\psi^{(u)})\Big)}{\langle 1, \psi^{(u)}\rangle (nh)^{1/2}}. \tag{7}$$

Here $d^{(z)} := a^{(z)} + b^{(z)}$. Now let κ_α be the $(1-\alpha)$–quantile of the combined statistic $T^* := \max\Big(T(\psi^{(\ell)}), T(-\psi^{(u)})\Big)$, i.e. the smallest real number such that $\mathrm{P}\{T^* \le \kappa_\alpha\} \ge 1 - \alpha$. Then

$$\begin{aligned} \hat\ell(t) &:= \sup_{h>0\,:\,t\in[a^{(\ell)}h,1-b^{(\ell)}h]} \left(\hat f_h^{(\ell)}(t) - \frac{\|\psi^{(\ell)}\|(\Gamma(d^{(\ell)}h) + \kappa_\alpha)}{\langle 1, \psi^{(\ell)}\rangle (nh)^{1/2}} \right), \\ \hat u(t) &:= \inf_{h>0\,:\,t\in[a^{(u)}h,1-b^{(u)}h]} \left(\hat f_h^{(u)}(t) + \frac{\|\psi^{(u)}\|(\Gamma(d^{(u)}h) + \kappa_\alpha)}{\langle 1, \psi^{(u)}\rangle (nh)^{1/2}} \right) \end{aligned}$$

defines a confidence band $(\hat\ell, \hat u)$ for f satisfying (2).

Equality holds in (2) if $\mathcal{G} = \mathcal{G}_\uparrow$ and f is constant, or if $\mathcal{G} = \mathcal{G}_{\mathrm{conv}}$ and f is linear, provided that $\kappa_\alpha > 0$. For then it follows from (4) and (5) with $g(x) = \pm 1$ or $g(x) = \pm x$ that the kernel estimators are unbiased. Thus $\hat\ell \le f \le \hat u$ is equivalent to $T^* > \kappa_\alpha$. Moreover, using general theory for gaussian measures on Banach spaces one can show that the distribution of T^* is continuous on $]0, \infty[$.

Sufficient conditions for requirements (4) and (5) in general are provided by Lemma 8.1 in Section 8. The confidence band presented in Section 3 is a special case of the one derived here, if we define $\psi^{(\ell)}(x) := 1\{x \in [-1, 0]\}$ and $\psi^{(u)}(x) := 1\{x \in [0, 1]\}$ and apply postprocessing as described below.

4.2 Postprocessing of confidence bands

Any confidence band $(\hat{\ell}, \hat{u})$ for f can be enhanced, if we replace $\hat{\ell}(x)$ and $\hat{u}(x)$ with

$$\hat{\hat{\ell}}(x) \ := \ \inf\Big\{g(x) : g \in \mathcal{G}, \hat{\ell} \le g \le \hat{u}\Big\} \quad \text{and} \quad \hat{\hat{u}}(x) \ := \ \sup\Big\{g(x) : g \in \mathcal{G}, \hat{\ell} \le g \le \hat{u}\Big\},$$

respectively. Here we assume tacitly that the set $\{g \in \mathcal{G} : \hat{\ell} \le g \le \hat{u}\}$ is nonempty.

In case of $\mathcal{G} = \mathcal{G}_{\uparrow}$ one can easily show that

$$\hat{\hat{\ell}}(x) \ = \ \sup_{t \in [0,x]} \hat{\ell}(t) \quad \text{and} \quad \hat{\hat{u}}(x) \ = \ \inf_{s \in [x,1]} \hat{u}(s).$$

Note also that $\hat{\hat{\ell}}$ and $\hat{\hat{u}}$ are isotonic, whereas the raw functions $\hat{\ell}$ and $\hat{u}$ need not be.

In case of $\mathcal{G} = \mathcal{G}_{\text{conv}}$ the modified upper bound $\hat{\hat{u}}$ is the greatest convex minorant of $\hat{u}$ and can be computed (in discrete models) by means of the pool-adjacent-violators algorithm (cf. Robertson et al. 1988). The modified lower bound $\hat{\hat{\ell}}(x)$ can be shown to be

$$\hat{\hat{\ell}}(x) \ = \ \max\left\{ \sup_{0 \le s < t \le x} \Big(\hat{\hat{u}}(s) + \frac{\hat{\ell}(t) - \hat{\hat{u}}(s)}{t - s}\,(x - s) \Big), \sup_{x \le s < t \le 1} \Big(\hat{\hat{u}}(t) - \frac{\hat{\hat{u}}(t) - \hat{\ell}(s)}{t - s}\,(t - x) \Big) \right\}.$$

This improved bound $\hat{\hat{\ell}}$ is not a convex function, though more regular than the raw function $\hat{\ell}$. Figure 2 depicts some hypothetical confidence band $(\hat{\ell}, \hat{u})$ for a function $f \in \mathcal{G}_{\text{conv}}$ and its improvement $(\hat{\hat{\ell}}, \hat{\hat{u}})$.

4.3 Adaptivity in terms of rates

Whenever we construct a band following the recipe above we end up with a confidence band adapting to the unknown smoothness of f in terms of rates of convergence. For $\beta, L > 0$ the Hölder smoothness class $\mathcal{H}_{\beta,L}$ is defined as follows: In case of $0 < \beta \le 1$ let

$$\mathcal{H}_{\beta,L} \ := \ \Big\{ g : |g(x) - g(y)| \le L|x - y|^{\beta} \text{ for all } x, y \Big\}.$$

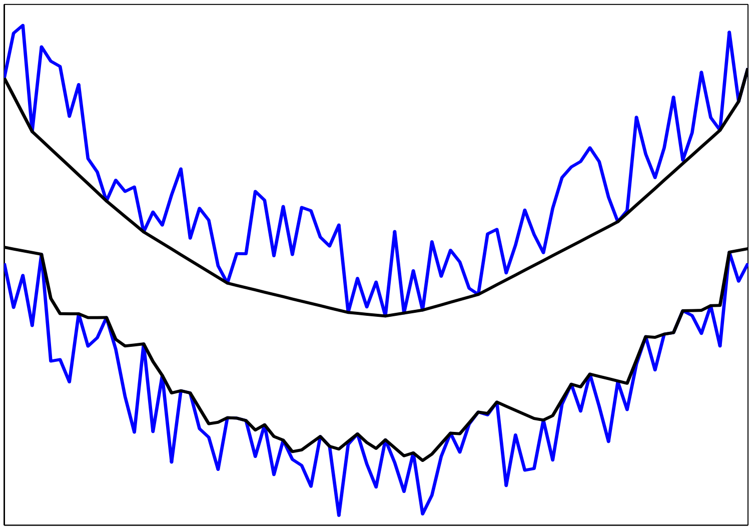

Figure 2: Improvement $(\hat{\hat{\ell}}, \hat{u})$ of a band $(\hat{\ell}, \hat{u})$ if $\mathcal{G} = \mathcal{G}_{\mathrm{conv}}$.

In case of $1 < \beta \le 2$ let

$$\mathcal{H}_{\beta,L} \ := \ \Bigl\{ g \in \mathcal{C}^1 : g' \in \mathcal{H}_{\beta-1,L} \Bigr\}.$$

Theorem 4.1. *Suppose that $f \in \mathcal{G} \cap \mathcal{H}_{\beta,L}$, where either $\mathcal{G} = \mathcal{G}_{\uparrow}$ and $\beta \le 1$, or $\mathcal{G} = \mathcal{G}_{\mathrm{conv}}$ and $1 \le \beta \le 2$. Let $(\hat{\ell}, \hat{u})$ be the confidence band for f based on test functions $\psi^{(\ell)}, \psi^{(u)}$ as described previously. Then there exists a constant Δ depending only on (β, L) and $(\psi^{(\ell)}, \psi^{(u)})$ such that*

$$\sup_{t \in [\epsilon_n, 1-\epsilon_n]} \Bigl(\hat{u}(t) - \hat{\ell}(t) \Bigr) \ \le \ \Delta \rho_n \Bigl(1 + \frac{\kappa_\alpha + T(\psi^{(u)}) + T(-\psi^{(\ell)})}{\log(en)^{1/2}} \Bigr),$$

where $\epsilon_n := \rho_n^{1/\beta}$ and

$$\rho_n \ := \ \Bigl(\frac{\log(en)}{n} \Bigr)^{\beta/(2\beta+1)}.$$

Using the same arguments as Khas'minskii (1978) one can show that for any $0 \le r < s \le 1$,

$$\inf_{f \in \mathcal{G} \cap \mathcal{H}_{\beta,L}} \mathrm{P}_f \left\{ \sup_{t \in [r,s]} (\hat{u}(t) - \hat{\ell}(t)) \le \Delta \rho_n \right\} \ \to \ 0,$$

provided that $\Delta > 0$ is sufficiently small. Thus our confidence bands adapt to the unknown smoothness of f.

4.4 Testing hypotheses about $f(t)$

In order to find suitable kernel functions $\psi^{(\ell)}, \psi^{(u)}$ we proceed similarly as Dümbgen and Spokoiny (2001, Section 3.2). That means we consider temporarily tests of the null hypothesis

$$\mathcal{F}_o \ := \ \Bigl\{ f \in \mathcal{G} \cap L^2[0,1] : f(t) \le r - \delta \Bigr\}$$

versus the alternative hypothesis

$$\mathcal{F}_A \;:=\; \Big\{ f \in \mathcal{G} \cap \mathcal{H}_{k,L} : f(t) \ge r \Big\}.$$

Here $t \in [0,1]$, $r \in \mathbb{R}$ and $L, \delta > 0$ are arbitrary fixed numbers, while

$$(\mathcal{G}, k) \;=\; (\mathcal{G}_\uparrow, 1) \quad \text{or} \quad (\mathcal{G}, k) \;=\; (\mathcal{G}_{\rm conv}, 2). \tag{8}$$

Note that $\mathcal{F}_o$ and $\mathcal{F}_A$ are closed, convex subsets of $L^2[0,1]$. Suppose that there are functions $f_o \in \mathcal{F}_o$ and $f_A \in \mathcal{F}_A$ such that

$$\int_0^1 (f_o - f_A)(x)^2 \, dx \;=\; \min_{g_o \in \mathcal{F}_o,\, g_A \in \mathcal{F}_A} \int_0^1 (g_o - g_A)(x)^2 \, dx.$$

Then optimal tests of $\mathcal{F}_o$ versus $\mathcal{F}_A$ are based on the linear test statistic $\int_0^1 (f_A - f_o)\, dY$, where critical values have to be computed under the assumption $f = f_o$. The problem of finding such functions f_o, f_A is treated in Section 8. Here is the conclusion: Let

$$\psi^{(\ell)}(x) \;:=\; \begin{cases} 1\{x \in [-1,0]\}(1+x) & \text{if } \mathcal{G} = \mathcal{G}_\uparrow, \\ 1\{x \in [-2,2]\}\Big(1 - (3/2)|x| + x^2/2\Big) & \text{if } \mathcal{G} = \mathcal{G}_{\rm conv}. \end{cases} \tag{9}$$

Then the functions

$$f_A(s) \;:=\; \begin{cases} r + L(s-t) & \text{if } \mathcal{G} = \mathcal{G}_\uparrow \\ r + L(s-t)^2/2 & \text{if } \mathcal{G} = \mathcal{G}_{\rm conv} \end{cases} \tag{10}$$

and

$$f_o \;:=\; f_A - \delta \psi_{h,t}^{(\ell)} \quad \text{with } h := (\delta/L)^{1/k}$$

solve our minimzation problem, provided that $a^{(\ell)} h \le t \le 1 - b^{(\ell)} h$. Thus the optimal linear test statistic may be written as $\int_0^1 \psi_{h,t}\, dY = \psi Y(h,t)$. Elementary considerations show that the inequality

$$\hat{f}_h^{(\ell)}(t) - \frac{\|\psi^{(\ell)}\|(\Gamma(d^{(\ell)}h) + \kappa_\alpha)}{\langle 1, \psi^{(\ell)}\rangle (nh)^{1/2}} \;\le\; r_o$$

is equivalent to

$$\begin{aligned} \psi Y(h,t) \;&\le\; n^{1/2} h r_o \langle 1, \psi^{(\ell)}\rangle + h^{1/2} \|\psi^{(\ell)}\| (\Gamma(d^{(\ell)}h) + \kappa_\alpha) \\ &=\; \mathrm{E}_{f_o}(\psi Y(h,t)) + \mathrm{Var}(\psi Y(h,t))^{1/2} (\Gamma(d^{(\ell)}h) + \kappa_\alpha). \end{aligned}$$

Thus our lower confidence bound $\hat{\ell}$ may be interpreted as a multiple test of all null hypotheses $\{f \in \mathcal{G} : f(t) \le r_o\}$ with $t \in [0,1]$ and $r_o \in \mathbb{R}$.

Analogous considerations yield a candidate for $\psi^{(u)}$: Let

$$\mathcal{F}_o \;:=\; \Big\{ f \in \mathcal{G} \cap L^2[0,1] : f(t) \ge r + \delta \Big\}$$

and

$$\mathcal{F}_A \;:=\; \Big\{ f \in \mathcal{G} \cap \mathcal{H}_{k,L} : f(t) \le r \Big\}.$$

Then the function f_A in (10) and

$$f_o \;:=\; f_A + \delta \psi_{h,t}^{(u)} \quad \text{with } h := (\delta/L)^{1/k}$$

form a least favorable pair (f_o, f_A) in $\mathcal{F}_o \times \mathcal{F}_A$, where

$$\psi^{(u)}(x) \;:=\; \begin{cases} 1\{x \in [0,1]\}(1-x) & \text{if } \mathcal{G} = \mathcal{G}_{\uparrow}, \\ 1\{x \in [-2^{1/2}, 2^{1/2}]\}(1-x^2/2) & \text{if } \mathcal{G} = \mathcal{G}_{\mathrm{conv}}. \end{cases} \tag{11}$$

Figures 3 and 4 depict the functions $\psi^{(\ell)}$ in (9) and $\psi^{(u)}$ in (11).

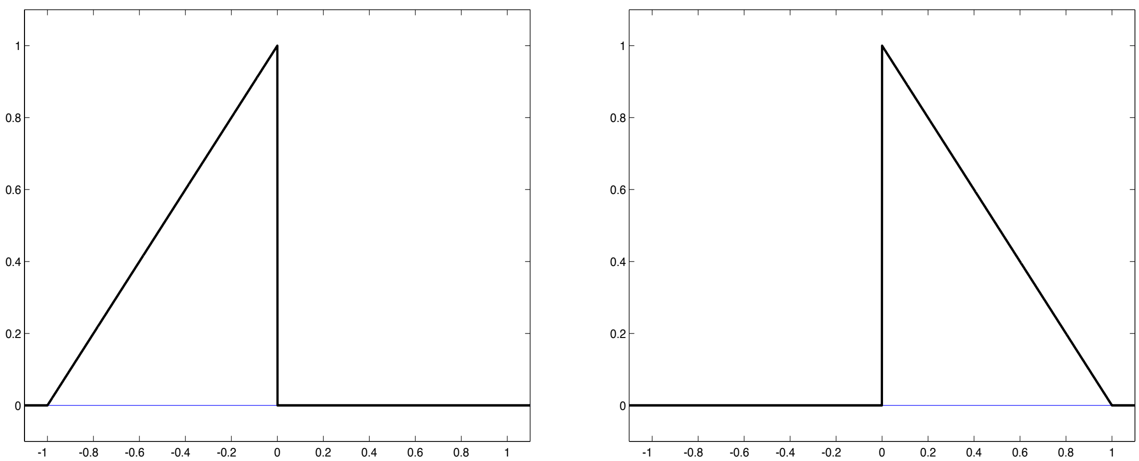

Figure 3: Kernel functions $\psi^{(\ell)}, \psi^{(u)}$ for $\mathcal{G}_{\uparrow}$.

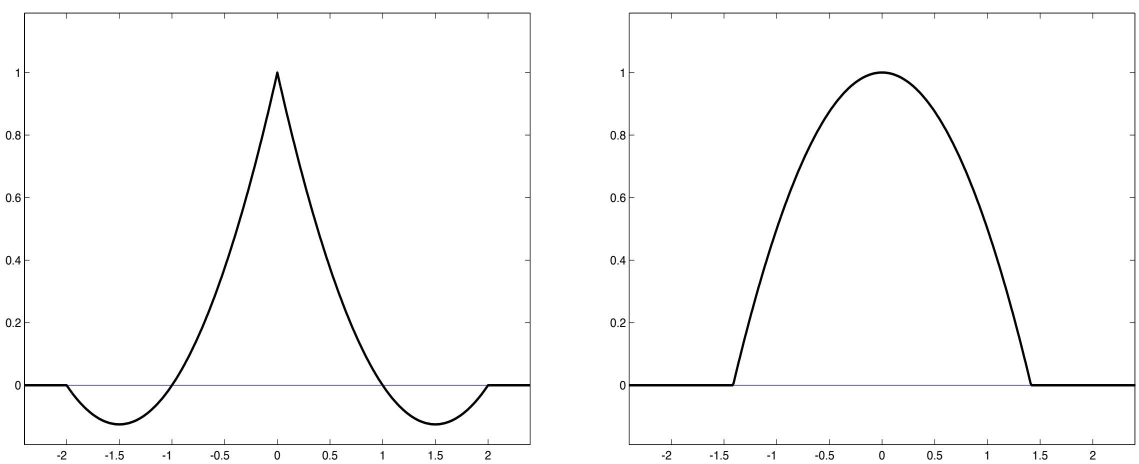

Figure 4: Kernel functions $\psi^{(\ell)}, \psi^{(u)}$ for $\mathcal{G}_{\mathrm{conv}}$.

4.5 Optimal constants and local adaptivity

Now we are going to show that our multiscale confidence band $(\hat{\ell}, \hat{u})$, if constructed with the kernel functions in (9) and (11), is locally adaptive in a certain sense. Precisely, we consider an arbitrary fixed function $f_o \in \mathcal{G} \cap \mathcal{C}^k$ with $(\mathcal{G}, k)$ as specified in (8). We analyze quantities such as

$$\|(\hat{u} - f_o)w\|_{r,s}^+ \quad \text{and} \quad \|(f_o - \hat{\ell})w\|_{r,s}^+,$$

where w is some positive weight function on the unit interval and

$$\|g\|_{r,s}^+ \ := \ \sup_{t \in [r,s]} g(t).$$

The function w should reflect local smoothness properties of f_o in an appropriate way. The following theorem demonstrates that the k–th derivative of f_o, denoted by $\nabla^k f_o$, plays a crucial role.

Theorem 4.2. *For arbitrary fixed numbers* $0 \le r < s \le 1$ *let*

$$L \ := \ \max_{t \in [r,s]} \nabla^k f_o(t).$$

Then for any $\gamma \in \,]0, 1[$,

$$\begin{aligned} \inf_{(\hat{\ell},\hat{u})} \mathrm{P}_{f_o}\Big\{\|f - \hat{\ell}\|_{r,s}^+ \ge \gamma \Delta^{(\ell)} L^{1/(2k+1)} \rho_n\Big\} &\ \ge \ 1 - \alpha + o(1), \\ \inf_{(\hat{\ell},\hat{u})} \mathrm{P}_{f_o}\Big\{\|\hat{u} - f\|_{r,s}^+ \ge \gamma \Delta^{(u)} L^{1/(2k+1)} \rho_n\Big\} &\ \ge \ 1 - \alpha + o(1), \end{aligned}$$

where both infima are taken over all confidence bands $(\hat{\ell}, \hat{u})$ *satisfying (2), and*

$$\begin{aligned} \Delta^{(z)} &\ := \ \Big((k + 1/2)\|\psi^{(z)}\|^2\Big)^{-k/(2k+1)}, \\ \rho_n &\ := \ \Big(\frac{\log(en)}{n}\Big)^{k/(2k+1)}. \end{aligned}$$

In case of $\mathcal{G} = \mathcal{G}_\uparrow$, the critical constants are $\Delta^{(\ell)} = \Delta^{(u)} = 2^{1/3} \approx 1.260$. In case of $\mathcal{G} = \mathcal{G}_{\mathrm{conv}}$,

$$\Delta^{(\ell)} \ = \ (3/4)^{2/5} \ \approx \ 0.891 \quad \text{and} \quad \Delta^{(u)} \ = \ 3^{2/5}/128^{1/5} \ \approx \ 0.588.$$

This indicates that bounding a convex function from below is more difficult than finding an upper bound.

In view of Theorem 4.2 we introduce for arbitrary fixed $\epsilon > 0$ the weight function

$$w_\epsilon \ := \ \Big(\max(\nabla^k f_o, \epsilon)\Big)^{-1/(2k+1)}$$

reflecting the local smoothness of f_o. The next theorem shows that our particular confidence band $(\hat{\ell}, \hat{u})$ attains the lower bounds of Theorem 4.2 pointwise. Suprema such as $\|(f_o - \hat{\ell})w_\epsilon\|_{r,s}^+$ and $\|(\hat{u} - f_o)w_\epsilon\|_{r,s}^+$ attain their respective lower bounds $\Delta^{(\ell)}$, $\Delta^{(u)}$ up to a multiplicative factor $2^{k/(k+1/2)} + o_p(1)$.

Theorem 4.3. *Let $(\hat{\ell}, \hat{u})$ be the confidence band based on the kernel functions in (9) and (11). If $f = f_o$, then for arbitrary $\epsilon > 0$ and any $t \in]0, 1[$,*

$$\begin{aligned} (f_o - \hat{\ell})(t)w_\epsilon(t) &\leq \Big(\Delta^{(\ell)} + o_p(1)\Big)\rho_n, \\ (\hat{u} - f_o)(t)w_\epsilon(t) &\leq \Big(\Delta^{(u)} + o_p(1)\Big)\rho_n. \end{aligned}$$

Moreover,

$$\begin{aligned} \|(f_o - \hat{\ell})w_\epsilon\|_{\epsilon,1-\epsilon}^+ &\leq \Big(2^{k/(k+1/2)}\Delta^{(\ell)} + o_p(1)\Big)\rho_n, \\ \|(\hat{u} - f_o)w_\epsilon\|_{\epsilon,1-\epsilon}^+ &\leq \Big(2^{k/(k+1/2)}\Delta^{(u)} + o_p(1)\Big)\rho_n. \end{aligned}$$

If we used kernel functions differing from (9) and (11), then pointwise optimality would be lost, and the constants for the supremum distances would get worse.

5 Simulations and numerical examples

Here we demonstrate the performance of the procedures in Section 4. We replace the continuous white noise model with a discrete one: Suppose that one observes a random vector $\vec{Y} \in \mathbb{R}^n$ with components

$$Y_i = f(x_i) + \epsilon_i, \tag{12}$$

where $x_i := (i - 1/2)/n$, and the random errors ϵ_i are independent with Gaussian distribution $\mathcal{N}(0, \sigma^2)$. Our kernel functions $\psi^{(\ell)}$ and $\psi^{(u)}$ are rescaled as follows:

$$\begin{aligned} \psi^{(\ell)}(x) &:= \begin{cases} 1\{x \in [-1, 0]\}(1 + x) & \text{if } \mathcal{G} = \mathcal{G}_\uparrow, \\ 1\{x \in [-1, 1]\}\Big(1 - 3|x| + 2x^2\Big) & \text{if } \mathcal{G} = \mathcal{G}_{\text{conv}}, \end{cases} \\ \psi^{(u)}(x) &:= \begin{cases} 1\{x \in [0, 1]\}(1 - x) & \text{if } \mathcal{G} = \mathcal{G}_\uparrow, \\ 1\{x \in [-1, 1]\}(1 - x^2) & \text{if } \mathcal{G} = \mathcal{G}_{\text{conv}}. \end{cases} \end{aligned}$$

Note that now $a^{(\ell)}, a^{(u)}, b^{(\ell)}, b^{(u)} \in \{0, 1\}$. For convenience we compute kernel estimators and confidence bounds for f only on the grid $\mathcal{T}_n := \{1/n, 2/n, \ldots, 1 - 1/n\}$, while the bandwidth parameter h is restricted to

$$H_n := \begin{cases} \{1/n, 2/n, \ldots, 1\} & \text{if } \mathcal{G} = \mathcal{G}_\uparrow, \\ \{1/n, 2/n, \ldots, \lfloor n/2 \rfloor/n\} & \text{if } \mathcal{G} = \mathcal{G}_{\text{conv}}. \end{cases}$$

Let ψ stand for $\psi^{(\ell)}$ or $\psi^{(u)}$ with support $[-a, b]$. Then for $h \in H_n$ and $t \in \mathcal{T}_n$ with $ah \le t \le 1 - bh$ we define

$$\psi\vec{Y}(h,t) \ := \ \sum_{i=1}^{n} \psi\Big(\frac{x_i - t}{h}\Big) Y_i \ = \ \sum_{j=1-anh}^{bnh} \psi\Big(\frac{j-1/2}{nh}\Big) Y_{nt+j}$$

and

$$\hat{f}_h(t) \ := \ \frac{\psi\vec{Y}(h,t)}{S_{nh}},$$

where S_d stands for $\sum_{j=1-d}^{d} \psi((j-1/2)/d)$. The standard deviation of $\hat{f}_h(t)$ equals $\sigma_h := \sigma R_{nh}^{1/2}/S_{nh}$, where $R_d := \sum_{j=1-d}^{d} \psi((j-1/2)/d)^2$. Tedious but elementary calculations show that in case of $\mathcal{G} = \mathcal{G}_{\uparrow}$,

$$S_d \ = \ d/2 \quad \text{and} \quad R_d \ = \ d/3 - 1/(12d).$$

In case of $\mathcal{G} = \mathcal{G}_{\mathrm{conv}}$,

$$\begin{array}{rclcrcl} S_d^{(\ell)} & = & d/3 - 1/(3d) & \text{and} & R_d^{(\ell)} & = & 4d/15 - 1/(2d) + 7/(30d^3), \\ S_d^{(u)} & = & 4d/3 + 1/(6d) & \text{and} & R_d^{(u)} & = & 16d/15 + 7/(120d^3). \end{array}$$

Note that here $S_1^{(\ell)} = 0 = \psi^{(\ell)}\vec{Y}(1/n, \cdot)$, whence the bandwidth $1/n$ is excluded from any computation involving $\psi^{(\ell)}$.

As for the bias of these kernel estimators, one can deduce from Lemma 8.1 that $\mathrm{E}\hat{f}_h^{(\ell)}(t) \le f(t)$ and $\mathrm{E}\hat{f}_h^{(u)}(t) \ge f(t)$ whenever $f \in \mathcal{G}$. Here is a discrete version of our multiscale test statistic: $T_n^* := \max\Big(T_n(\psi^{(\ell)}), T_n(-\psi^{(u)})\Big)$, where

$$T_n(\pm\psi) \ := \ \max_{h \in H_n} \ \max_{t \in \mathcal{T}_n \cap [ah, 1-bh]} \Big(\pm\sigma^{-1} R_{nh}^{-1/2} \psi\vec{E}(h,t) - \Gamma((a+b)h) \Big)$$

with $\vec{E} := (\epsilon_i)_{i=1}^n$. Let $\kappa_{\alpha,n}$ be the $(1-\alpha)$–quantile of T_n^*. Then

$$\begin{array}{rcl} \hat{\ell}(t) & := & \displaystyle\max_{h \in H_n \,:\, t \in [a^{(\ell)}h, 1-b^{(\ell)}h]} \Big(\hat{f}_h^{(\ell)}(t) - \sigma_h^{(\ell)} (\Gamma(d^{(\ell)}h) + \kappa_{\alpha,n}) \Big), \\ \hat{u}(t) & := & \displaystyle\min_{h \in H_n \,:\, t \in [a^{(u)}h, 1-b^{(u)}h]} \Big(\hat{f}_h^{(u)}(t) + \sigma_h^{(u)} (\Gamma(d^{(u)}h) + \kappa_{\alpha,n}) \Big), \end{array}$$

defines a confidence band for f such that

$$\mathrm{P}\Big\{ \hat{\ell} \le f \le \hat{u} \text{ on } \mathcal{T}_n \Big\} \ \ge \ 1 - \alpha \quad \text{whenever } f \in \mathcal{G}.$$

Equality holds if $\mathcal{G} = \mathcal{G}_{\uparrow}$ and f is constant, or if $\mathcal{G} = \mathcal{G}_{\mathrm{conv}}$ and f is linear. If the noise variance σ^2 is unknown, it may be estimated as described in Dümbgen and Spokoiny (2001). Then, under moderate regularity assumptions on f, our confidence bands have *asymptotic* coverage probability at least $1 - \alpha$ as n tends to infinity.

Critical values. For various values of n we estimated several quantiles $\kappa_{\alpha,n}$ in 9999 Monte-Carlo simulations; see Table 1. One can easily show that the critical value $\kappa_{\alpha,n}$ converges to the corresponding quantile κ_α for the continuous white noise model as $n \to \infty$. Software for the computation of critical values as well as confidence bands may be obtained from the author's URL.

	$\mathcal{G}_\uparrow$			$\mathcal{G}_{\mathrm{conv}}$		
n	$\kappa_{0.5,n}$	$\kappa_{0.1,n}$	$\kappa_{0.05,n}$	$\kappa_{0.5,n}$	$\kappa_{0.1,n}$	$\kappa_{0.05,n}$
100	0.330	1.092	1.349	0.350	1.053	1.283
200	0.433	1.146	1.392	0.430	1.121	1.342
300	0.475	1.169	1.416	0.470	1.126	1.342
400	0.507	1.204	1.446	0.489	1.128	1.340
500	0.526	1.222	1.450	0.512	1.143	1.358
700	0.570	1.252	1.492	0.536	1.162	1.380
1000	0.585	1.250	1.483	0.552	1.178	1.393

Table 1: Some critical values for the discrete white noise model

Two numerical examples. Figure 5 shows a simulated data vector $\vec{Y}$ with $n = 500$ components together with the corresponding 95%–confidence band $(\hat{\ell}, \hat{u})$ after postprocessing, where f is assumed to be isotonic. The latter function is depicted as well. Note that the band is comparatively narrow in the middle of $]0, 1/3[$, on which f is constant. On $]1/3, 1]$ the width $\hat{u} - \hat{\ell}$ tends to inrease, as does ∇f. These findings are in accordance with Theorem 4.3.

An analogous plot for a convex function f can be seen in Figure 6. Note that the deviation $f - \hat{\ell}$ is mostly greater than $\hat{u} - f$, as predicted by Theorem 4.3.

6 Proofs

Proof of Theorem 3.1. In order to prove lower bounds we construct unfavorable subfamilies of $\mathcal{G}_\uparrow$ similarly as Khasminski (1978). For a given integer $m > 0$ we define $I_1 := [0, 1/m]$ and $I_j :=](j-1)/m, j/m]$ for $1 < j \le m$. Then we define step functions g and h_ξ for $\xi \in \mathbb{R}^m$ via

$$g(t) \ := \ 2j - 1 \quad \text{and} \quad h_\xi(t) \ := \ \xi_j \quad \text{for } t \in I_j, 1 \le j \le m.$$

For any $\delta > 0$ and $\xi \in [-\delta, \delta]^m$ the function $\delta g + h_\xi$ is isotonic on $[0, 1]$. Now we restrict our attention to the parametric submodel $\mathcal{F}_o = \Big\{\delta g + h_\xi : \xi \in [-\delta, \delta]^m\Big\}$ of $\mathcal{G}_\uparrow \cap L^2[0, 1]$. Any confidence band $(\hat{\ell}, \hat{u})$ for $f = \delta g + h_\xi$ defines a confidence set $S = S_1 \times S_2 \times \cdots \times S_m$ for ξ via

$$S_j \ := \ \Big[\sup_{t \in I_j} \hat{\ell}(t) - \delta(2j-1), \inf_{t \in I_j} \hat{u}(t) - \delta(2j-1)\Big].$$

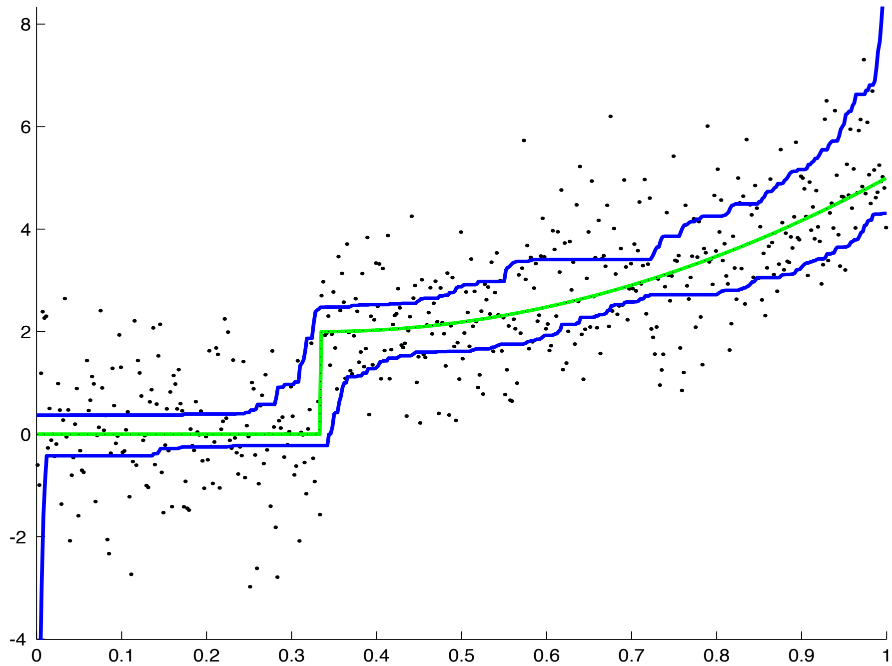

Figure 5: Data $\vec{Y}$ and 95%–confidence band for $f \in \mathcal{G}_{\uparrow}$.

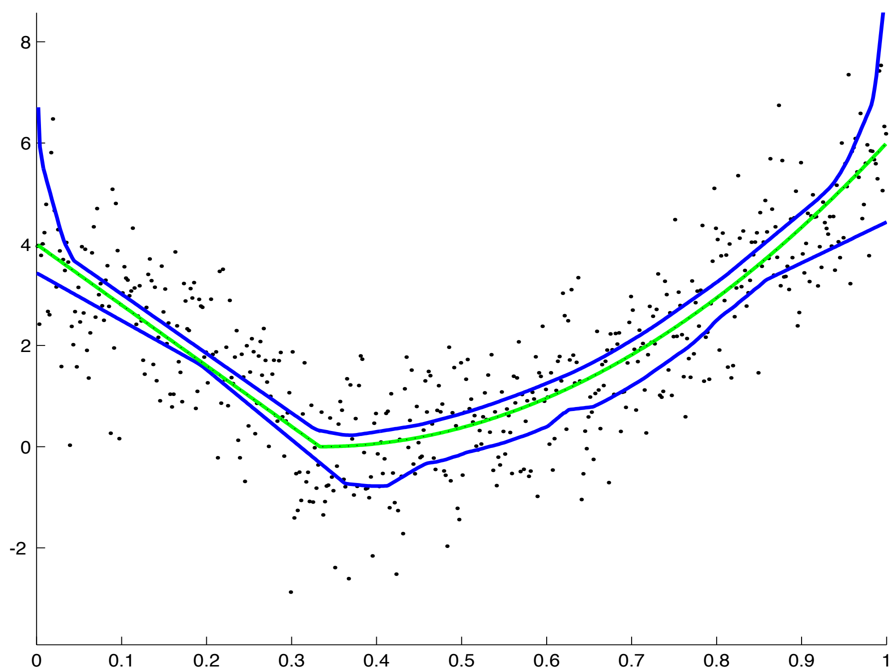

Figure 6: Data $\vec{Y}$ and 95%–confidence band for $f \in \mathcal{G}_{\mathrm{conv}}$.

Here $\hat{\ell} \le f \le \hat{u}$ if, and only if, $\xi \in S$. Moreover,

$$D_\epsilon(\hat{\ell}, \hat{u}) \ \ge \ \max_{j=1,\ldots,m} \mathrm{length}(S_j) \quad \text{for } 1/(m+1) \le \epsilon < 1/m.$$

However,

$$\begin{aligned}
\log \frac{d\mathrm{P}_{\delta g + h_\xi}}{d\mathrm{P}_{\delta g}}(Y) &= n^{1/2} \int_0^1 h_\xi \, d\tilde{Y} - n \int_0^1 h_\xi(t)^2 \, dt/2 \\
&= \sum_{j=1}^m \Big((n/m)^{1/2} \xi_j X_j - (n/m) \xi_j^2/2 \Big) \\
&= \log \frac{d\mathcal{N}((n/m)^{1/2}\xi, I)}{d\mathcal{N}(0, I)}(X),
\end{aligned}$$

where $\tilde{Y}(t) := Y(t) - n^{1/2} \int_0^t \delta g(s) \, ds$ and $X := (X_j)_{j=1}^m$ with components

$$X_j \ := \ m^{1/2} \Big(\tilde{Y}(j/m) - \tilde{Y}((j-1)/m) \Big).$$

In case of $f = \delta g$ these random variables are independent and standard normal. Consequently, X is a sufficient statistic for the parametric submodel $\mathcal{F}_o$ with distribution $\mathcal{N}_m((n/m)^{1/2}\xi, I)$ in case of $f = \delta g + h_\xi$. In particular, the conditional distribution of S given X does not depend on ξ. Hence letting $\delta = (n/m)^{-1/2} c_m$ with $c_m := (2 \log m)^{1/2}$ it follows from Theorem 7.1 (b) in Section 7 that for $1/(m+1) \le \epsilon < 1/m$,

$$\begin{aligned}
&\inf_{f \in \mathcal{G}_\uparrow \cap L^2[0,1]} \mathrm{P}_f \Big\{ \hat{\ell} \le f \le \hat{u} \text{ and } D_\epsilon(\hat{\ell}, \hat{u}) \le 2 \frac{c_m - b_m}{(n/m)^{1/2}} \Big\} \\
&\le \ \min_{\xi \in [-\delta, \delta]^m} \mathrm{P}_\xi \Big\{ \xi \in S \text{ and } \max_{j=1,\ldots,m} \mathrm{length}(S_j) \le 2 \frac{c_m - b_m}{(n/m)^{1/2}} \Big\} \ \le \ b_m,
\end{aligned}$$

where $b_1, b_2, b_3, \ldots$ are universal positive numbers such that $\lim_{m \to \infty} b_m = 0$. This entails the assertion of Theorem 3.1 with $\log(1/\epsilon)$ in place of $\log(e/\epsilon)$ and

$$b(\epsilon) \ := \ (2 \log(1/\epsilon))^{1/2} - (m\epsilon)^{1/2} (c_m - b_m) \quad \text{for } 1/(m+1) \le \epsilon < 1/m.$$

Finally note that $\log(e/\epsilon)^{1/2} = \log(1/\epsilon)^{1/2} + o(1)$ as $\epsilon \downarrow 0$. $\square$

Proof of Theorem 4.1. Instead of an upper bound for $\hat{u} - \hat{\ell}$ we prove an upper bound for $\hat{u} - f$, because analogous arguments apply to $f - \hat{\ell}$. In what follows let $\psi = \psi^{(u)}$ with support $[-a, b]$. For $t \in [0, 1]$ and $h > 0$ with $ah \le t \le 1 - bh$,

$$\begin{aligned}
\hat{u}(t) - f(t) &\le \hat{f}_h(t) - f(t) + \frac{\|\psi\| (\Gamma((a+b)h) + \kappa_\alpha)}{\langle 1, \psi \rangle (nh)^{1/2}} \\
&= \frac{\Big\langle f(t + h\,\cdot) - f(t), \psi \Big\rangle}{\langle 1, \psi \rangle} + \frac{\psi W(h, t)}{n^{1/2} h \langle 1, \psi \rangle} + \frac{\|\psi\| (\Gamma((a+b)h) + \kappa_\alpha)}{\langle 1, \psi \rangle (nh)^{1/2}} \\
&\le \frac{\Big\langle f(t + h\,\cdot) - f(t), \psi \Big\rangle}{\langle 1, \psi \rangle} + \frac{\|\psi\| \Big(2\Gamma((a+b)h) + \kappa_\alpha + T(\psi) \Big)}{\langle 1, \psi \rangle (nh)^{1/2}}. \qquad (13)
\end{aligned}$$

For any function $g \in \mathcal{H}_{\beta,L}$,

$$\begin{aligned} \Big| g(x) - g(0) \Big| \ &\le \ L|x|^\beta \quad \text{if } \beta \le 1, \\ \Big| g(x) - g(0) - g'(0)x \Big| \ &\le \ L|x|^\beta \quad \text{if } 1 < \beta \le 2. \end{aligned}$$

Since $f(t + h\,\cdot) \in \mathcal{H}_{\beta,Lh^\beta}$ if $f \in \mathcal{H}_{\beta,L}$, this implies that

$$\frac{\Big\langle f(t+h\,\cdot) - f(t), \psi \Big\rangle}{\langle 1, \psi \rangle} \ \le \ \frac{Lh^\beta \int_{-a}^b |x|^\beta |\psi(x)| \, dx}{\langle 1, \psi \rangle} \ \le \ \Delta h^\beta.$$

Here and subsequently Δ denotes a generic constant depending only on (β, L) and ψ. Its value may vary from one place to another. In case of $t \in [\epsilon_n, 1 - \epsilon_n]$ and $h = \epsilon_n / \max(a, b)$ the right-hand side of (13) is not greater than

$$\Delta \epsilon_n^\beta + \frac{\Delta\Big(\log(en)^{1/2} + \kappa_\alpha + T(\psi)\Big)}{(n\epsilon_n)^{1/2}} \ = \ \Delta \rho_n \Big(1 + \frac{\kappa_\alpha + T(\psi)}{\log(en)^{1/2}}\Big). \qquad \Box$$

Proof of Theorem 4.2. We prove only the lower bound for $f_o - \hat{\ell}$, because $\hat{u} - f_o$ can be treated analogously. It suffices to consider the case $L > 0$ and to show that for any fixed number $\gamma \in \,]0, 1[$,

$$\mathbf{P}_{f_o}\Big\{ \|f_o - \hat{\ell}\|_{r,s}^+ \ge \gamma \Delta^{(\ell)} L^{1/(2k+1)} \rho_n \Big\} \ \ge \ 1 - \alpha + o(1)$$

for arbitrary confidence bands $(\hat{\ell}, \hat{u}) = (\hat{\ell}_n, \hat{u}_n)$ satisfying (2). Without loss of generality one may assume that

$$\nabla^k f_o \ \ge \ L \quad \text{on } [r, s].$$

Otherwise one could increase γ and decrease L without changing $\gamma L^{1/(2k+1)}$, and replace $[r, s]$ with some nondegenerate subinterval. Let ψ stand for $\psi^{(\ell)}$ with support $[-a, b]$. For $0 < h \le (s - r)/(a + b)$ and positive integers $j \le m := \lfloor (s - r)/((a + b)h \rfloor$ let

$$t_j \ := \ s + ah + (j - 1)(a + b)h \quad \text{and} \quad f_j \ := \ f_o - Lh^k \psi_{h,t_j}.$$

It follows from Lemma 8.4 that these functions f_j belong to $\mathcal{G} \cap L^2[0, 1]$. Thus (2) implies that the event

$$A \ := \ \Big\{ \hat{\ell} \le f_j \text{ for some } j \le m \Big\}$$

satisfies the inequality $\mathbf{P}_{f_j}(A) \ge 1 - \alpha$ for all $j \le m$. Since $\|f_o - f_j\|_{r,s}^+ \ge \delta$, this entails the inequality

$$\mathbf{P}_{f_o}\Big\{ \|f_o - \hat{\ell}\|_{r,s}^+ \ge Lh^k \Big\} \ \ge \ \mathbf{P}_{f_o}(A) \ \ge \ 1 - \alpha - \min_{j \le m} \Big(\mathbf{P}_{f_j}(A) - \mathbf{P}_{f_o}(A) \Big).$$

Now let $h := (c\rho_n)^{1/k}$ so that $Lh^k = Lc\rho_n$, where $c > 0$ is some number to be specified later. For sufficiently large n this bandwidth h is smaller than $(s-r)/(a+b)$. Then

$$\log \frac{d\mathrm{P}_{f_j}}{d\mathrm{P}_{f_o}}(Y) \ = \ n^{1/2}h^{k+1/2}L\|\psi\|X_j - nh^{2k+1}L^2\|\psi\|^2/2,$$

where $X_j := h^{-1/2}\|\psi\|^{-1}\int_0^1 \psi_{h,t_j}\,d\tilde{Y}$ and $\tilde{Y}(t) := Y(t) - n^{1/2}\int_0^t f_o(x)\,dx$. Thus $X := (X_j)_{j=1}^m$ is a sufficient statistic for the restricted model $\{f_o, f_1, f_2, \ldots, f_m\}$, where $\mathcal{L}_{f_o}(X)$ is a standard normal distribution on $\mathbb{R}^m$. Thus it follows from Theorem 7.1 (a) and a standard sufficiency argument that

$$\lim_{n\to\infty} \min_{1\le j\le m} \Big(\mathrm{P}_{f_j}(A) - \mathrm{P}_{f_o}(A)\Big) \ = \ 0 \quad \text{if} \quad \lim_{n\to\infty} \frac{nh^{2k+1}L^2\|\psi\|^2}{2\log m} \ < \ 1.$$

Since $\log m = (1+o(1))\log(n)/(2k+1)$, the limit on the right hand side is equal to

$$c^{(2k+1)/k}L^2\|\psi\|^2(k+1/2)$$

and smaller than one if c equals $\gamma\Delta^{(\ell)}L^{-2k/(2k+1)}$. In that case, the lower bound $Lh^k = Lc\rho_n$ for $\|f_o - \hat{\ell}\|_{r,s}^+$ equals $\gamma\Delta^{(\ell)}L^{1/(2k+1)}\rho_n$ as desired. $\square$

Proof of Theorem 4.3. Again we restrict our attention to $f_o - \hat{\ell}$ and let $\psi := \psi^{(\ell)}$ with support $[-a,b]$. For any fixed $\epsilon > 0$ and arbitrary $t \in [0,1]$ let $h_t > 0$ and

$$L_t \ := \ \max_{s\in[t-ah_t, t+bh_t]\cap[0,1]} \max(\nabla^k f_o(s), \epsilon).$$

In case of $ah_t \le t \le 1 - bh_t$ the inequality $(f_o - \hat{\ell})(t) \ge L_t h_t^k$ implies that

$$\hat{f}_{h_t}(t) - \frac{\|\psi\|\Big(\Gamma((a+b)h_t) + \kappa_\alpha\Big)}{(nh_t)^{1/2}\langle 1, \psi\rangle} \ \le \ f_o(t) - L_t h_t^k.$$

Since $f = f_o$, this can be rewritten as

$$\begin{aligned}
\frac{\psi W(h_t,t)}{h_t^{1/2}\|\psi\|} \quad &\le \quad -\frac{(nh_t)^{1/2}}{\|\psi\|}\Big\langle f_o(t + h_t\,\cdot) - f_o(t) + L_t h_t^k, \psi\Big\rangle + \Gamma((a+b)h_t) + \kappa_\alpha \\
&\le \quad -n^{1/2}L_t h_t^{k+1/2}\|\psi\| + \Gamma((a+b)h_t) + \kappa_\alpha,
\end{aligned}$$

where the latter inequality follows from Lemma 8.4 (c). Specifically let

$$h_t \ := \ cw_\epsilon(t)^2\rho_n^{1/k}$$

for some positive constant c to be specified later. By continuity of $\nabla^k f_o$, the weight function w_ϵ is bounded away from zero and infinity. Hence $h_t \to 0$ and $L_t \max(\nabla^k f_o(t), \epsilon)^{-1} \to 1$, uniformly in $t \in [0,1]$. In particular,

$$\begin{aligned}
\Gamma((a+b)h_t) &\leq (k+1/2)^{-1/2}\log(en)^{1/2} \quad \text{for } n \geq n_o, \\
n^{1/2}L_t h_t^{k+1/2}\|\psi\| &\geq c^{k+1/2}\|\psi\|\log(en)^{1/2}, \\
L_t h_t^k &\leq w_\epsilon(t)^{-1}c^k(1+b_n)\rho_n,
\end{aligned}$$

where n_o and b_n are positive numbers depending only on f_o, ϵ and c such that $b_n \to 0$. Consequently, for $n \geq n_o$,

$$ah_t \leq t \leq 1 - bh_t \quad \text{and} \quad (f_o - \hat{\ell})(t)w_\epsilon(t) \ \geq \ c^k(1+b_n)\rho_n$$

implies that

$$\frac{\psi W(h_t,t)}{h_t^{1/2}\|\psi\|} \ \leq \ -\Big(c^{k+1/2}\|\psi\| - (k+1/2)^{-1/2}\Big)\log(en)^{1/2} + \kappa_\alpha.$$

Whenever $c > (\Delta^{(\ell)})^{1/k}$, the right-hand side of the preceding inequality tends to minus infinity, while the random variable on the left-hand side has mean zero and variance one. Since the limit of $c^k(1+b_n)$ can be arbitrarily close to $\Delta^{(\ell)}$, these considerations show that $(f_o - \hat{\ell})(t)w_\epsilon(t) \leq (\Delta^{(\ell)} + o_p(1))\rho_n$ for any fixed $t \in \,]0,1[$.

If n is sufficiently large, then $ah_t \leq t \leq 1 - bh_t$ and

$$\frac{\psi W(h_t,t)}{h_t^{1/2}\|\psi\|} \ \geq \ -T(-\psi) - \Gamma((a+b)h_t)$$

for all $t \in [\epsilon, 1-\epsilon]$. Consequently,

$$\sup_{t\in[\epsilon,1-\epsilon]} (f_o - \hat{\ell})(t(w_\epsilon(t) \ \geq \ c^k(1+b_n)$$

implies that

$$\begin{aligned}
T(-\psi) &\geq n^{1/2}L_t h_t^{k+1/2}\|\psi\| - 2\Gamma((a+b)h_t) - \kappa_\alpha \\
&\geq \Big(c^{k+1/2}\|\psi\| - 2(k+1/2)^{-1/2}\Big)\log(en)^{1/2} - \kappa_\alpha.
\end{aligned}$$

Whenever $c > 2^{1/(k+1/2)}(\Delta^{(\ell)})^{1/k}$, the right hand side of the preceding inequality tends to infinity. Since the limit of $c^k(1+b_n)$ can be arbitrarily close to $2^{k/(k+1/2)}\Delta^{(\ell)}$, these considerations reveal that $\|(f_o - \hat{\ell})w_\epsilon\|_{\epsilon,1-\epsilon}^+$ is not greater than $\Big(2^{k/(k+1/2)}\Delta^{(\ell)} + o_p(1)\Big)\rho_n$. $\Box$

7 Some decision theory

Let $X = (X_i)_{i=1}^m$ be a random vector with distribution $\mathcal{N}_m(\theta, I)$. In what follows we consider tests $\phi : \mathbb{R}^m \to [0,1]$ and confidence sets

$$S \ = \ S_1 \times S_2 \times \cdots \times S_m$$

for θ with random intervals $S_j \subset \mathbb{R}$. The conditional distribution of S, given X, does not depend on θ. The possibility of randomized confidence sets S, i.e. confidence sets not just being a function of X, has to be included for technical reasons. Unless specified differently, asymptotic statements in this section refer to $m \to \infty$.

Theorem 7.1. *Let $c_m := (2 \log m)^{1/2}$. There are universal positive numbers b_m with $b_m \to 0$ such that the following two inequalities are satisfied:*

(a) *For arbitrary tests ϕ,*

$$\min_{j=1,\ldots,m} \mathrm{E}_{(c_m - b_m)e_j} \phi(X) - \mathrm{E}_0 \phi(X) \ \le \ b_m,$$

where $e_1, e_2, \ldots, e_m$ denotes the standard basis of $\mathbb{R}^m$.

(b) *For arbitrary confidence sets S as above,*

$$\min_{\theta \in [-c_m, c_m]^m} \mathrm{P}_\theta \Big\{ \theta \in S \text{ and } \max_{j=1,\ldots,m} \mathrm{length}(S_j) < 2(c_m - b_m) \Big\} \ \le \ b_m.$$

Proof of Theorem 7.1. Part (a) is classical and can be proved by a Bayesian argument; see for instance Ingster (1993) or Dümbgen and Spokoiny (2001). In order to prove part (b) we also consider a Bayesian model: Let θ have independent components each of which is uniformly distributed on the three-point set $K_m := \{-\kappa_m, 0, \kappa_m\}$, where $\kappa_m := c_m - b_m$ with constants $b_m \in [0, c_m]$ to be specified later on. Let $\mathcal{L}(X \,|\, \theta) = \mathcal{N}_m(\theta, I)$. Let $\mathrm{P}(\cdot), \mathrm{E}(\cdot)$ denote probabilities and expectations in this Bayesian context, whereas $\mathrm{P}_\theta(\cdot), \mathrm{E}_\theta(\cdot)$ are used in case of a fixed parameter θ. For any confidence set S,

$$\begin{aligned} &\min_{\theta \in [-c_m, c_m]^m} \mathrm{P}_\theta \Big\{ \theta \in S \text{ and } \max_{j=1,\ldots,m} \mathrm{length}(S_j) < 2\kappa_m \Big\} \\ &\le \ \ \mathrm{P} \Big\{ \theta \in S \text{ and } \max_{j=1,\ldots,m} \mathrm{length}(S_j) < 2\kappa_m \Big\} \ \le \ \mathrm{P}\{\theta \in \tilde{S}\}, \end{aligned}$$

where

$$\tilde{S} \ := \ \begin{cases} S & \text{if } \max_{j=1,\ldots,m} \mathrm{length}(S_j) < 2\kappa_m, \\ \{0\} \times \cdots \times \{0\} & \text{else.} \end{cases}$$

The conditional distribution of θ given (X, S) is also a product of m probability measures: For any $\eta \in K_m^m$,

$$\mathrm{P}(\theta = \eta \,|\, X, S) \ = \ \prod_{i=1}^m g(\eta_i \,|\, X_i) \quad \text{with} \quad g(z \,|\, x) \ := \ \frac{\exp(-(x-z)^2/2)}{\sum_{y \in K_m} \exp(-(x-y)^2/2)}.$$

Since each factor $\tilde{S}_j$ of $\tilde{S}$ contains at most two points from K_m,

$$\begin{aligned}
\mathrm{P}\{\theta \in \tilde{S}\} &= \mathrm{E}\mathrm{P}(\theta \in \tilde{S} \,|\, X, S) \\
&\leq \mathrm{E} \max_{\eta \in K_m^m} \mathrm{P}(\theta_i \neq \eta_i \text{ for } i = 1, \ldots, m \,|\, X, S) \\
&= \mathrm{E} \prod_{i=1}^{m} \Big(1 - \min_{z \in K_m} g(z \,|\, X_i)\Big) \\
&= \Big(1 - \mathrm{E} \min_{z \in K_m} g(z \,|\, X_1)\Big)^m \\
&\leq \Big(1 - 3^{-1}\mathrm{E} \min_{z \in K_m} \exp(-(X_1 - z)^2/2)\Big)^m.
\end{aligned}$$

The latter expectation can be bounded from below as follows:

$$\begin{aligned}
&3^{-1}\mathrm{E} \min_{z \in K_m} \exp(-(X_1 - z)^2/2) \\
&\quad\geq 3^{-1}\mathrm{P}\{|X_1| \leq b_m/2\} \exp(-(\kappa_m + b_m/2)^2/2) \\
&\quad\geq 3^{-1}\mathrm{P}\{|\theta_1| = 0, |X_1| \leq b_m/2\} \exp(-(c_m - b_m/2)^2/2) \\
&\quad= 9^{-1}(2\pi)^{-1/2}(b_m + O(b_m^2)) \exp(c_m b_m/2 - b_m^2/8) m^{-1}.
\end{aligned}$$

In case of $b_m := 1\{m > 1\} c_m^{-1/2} = o(1)$ the latter bound is easily seen to be $a_m m^{-1}$ with $a_m = a_m(b_m) \to \infty$. Thus

$$\mathrm{P}\{\theta \in \tilde{S}\} \;\leq\; (1 - a_m m^{-1})^m \;\to\; 0.$$

Replacing b_m with $\max\{b_m, (1 - a_m m^{-1})^m\}$ yields the assertion of part (b). $\square$

8 Related optimization problems

As in Section 4 let $(\mathcal{G}, k)$ be either $(\mathcal{G}_\uparrow, 1)$ or $(\mathcal{G}_{\mathrm{conv}}, 2)$. In view of future applications to other regression models we extend our framework slightly and consider $\langle g, h\rangle := \int gh \, d\mu$, $\|g\| := \langle g, g\rangle^{1/2}$ for some measure μ on the real line such that $\mu(C) < \infty$ for bounded intervals $C \subset \mathbb{R}$.

Let ψ be some bounded function on the real line with $\psi(x) = 0$ for $x \notin [-a, b]$ and $\langle 1, \psi\rangle \geq 0$, where $a, b \geq 0$. The next lemma provides sufficient conditions for one of the following two requirements:

$$\langle g, \psi\rangle \;\leq\; g(0)\langle 1, \psi\rangle \quad \text{whenever } g \in \mathcal{G}, 1_{[-a,b]} g \in L^1(\mu), \tag{14}$$

$$\langle g, \psi\rangle \;\geq\; g(0)\langle 1, \psi\rangle \quad \text{whenever } g \in \mathcal{G}, 1_{[-a,b]} g \in L^1(\mu). \tag{15}$$

Lemma 8.1. *Let $\mathcal{G} = \mathcal{G}_\uparrow$ and $\psi \geq 0$. Then $b = 0$ entails condition (14), while $a = 0$ implies condition (15).*

Let $\mathcal{G} = \mathcal{G}_{\text{conv}}$ and $\int_{-\infty}^{\infty} x\psi(x)\,\mu(dx) = 0$. Condition (15) is satisfied if $\psi \geq 0$. On the other hand, condition (14) is a consequence of the following two requirements: $\int x^{\pm}\psi(x)\,\mu(dx) = 0$ and

$$\psi \begin{cases} \geq 0 & \text{on } [c,d] \\ \leq 0 & \text{on } \mathbb{R}\setminus[c,d] \end{cases}$$

for some numbers $c < 0 < d$, where $\mu([-a,c]), \mu([d,b]) > 0$. (Here $y^+ := \max(y,0)$ and $y^- := \max(-y,0)$.)

With Lemma 8.1 at hand one can solve two mimimization problems leading to the special kernels in (9) and (11). In both cases we consider two disjoint convex sets $\mathcal{G}_o, \mathcal{G}_A \subset \mathcal{G}$ and construct functions $G_o \in \mathcal{G}_o$, $G_A \in \mathcal{G}_A$ such that

$$\|G_o - G_A\| \;=\; \min_{g_o\in\mathcal{G}_o,\, g_A\in\mathcal{G}_A} \|g_o - g_A\|. \tag{16}$$

Theorem 8.2. *Let $\mathcal{G}_o := \Big\{g \in \mathcal{G} : g(0) \leq -1\Big\}$ and $\mathcal{G}_A := \Big\{g \in \mathcal{G}\cap\mathcal{H}_{k,1} : g(0) \geq 0\Big\}$. In case of $\mathcal{G} = \mathcal{G}_\uparrow$ let $G_A(x) := x$ and*

$$G_o(x) \;:=\; \begin{cases} -1 & \text{if } x \in [-1,0], \\ G_A(x) & \text{else.} \end{cases}$$

In case of $\mathcal{G} = \mathcal{G}_{\text{conv}}$ let $G_A(x) := x^2/2$ and

$$G_o(x) \;:=\; \begin{cases} -1 + (a/2 + 1/a)x^- + (b/2 + 1/b)x^+ & \text{if } x \in [-a,b], \\ G_A(x) & \text{else,} \end{cases}$$

where $a, b \geq 2^{1/2}$ are chosen such that $\int x^{\pm}(G_A - G_o)(x)\,\mu(dx) = 0$.

Then equation (16) holds in both cases. More precisely, the function $\psi := G_A - G_o$ satisfies the inequalities $\langle 1,\psi\rangle \geq \|\psi\|^2$, (14) and

$$\langle g,\psi\rangle \;\geq\; \|\psi\|^2 - \langle 1,\psi\rangle \quad \text{whenever } g \in \mathcal{H}_{k,1}, g(0) \geq 0. \tag{17}$$

In case of μ being Lebesgue measure, $\psi = G_A - G_o$ coincides with the function $\psi^{(\ell)}$ in (9), where $a = b = 2$.

Theorem 8.3. *Let $\mathcal{G}_o := \Big\{g \in \mathcal{G} : g(0) \geq 1\Big\}$, $\mathcal{G}_A := \Big\{g \in \mathcal{G}\cap\mathcal{H}_{k,1} : g(0) \leq 0\Big\}$, and define G_A as in Theorem 8.2. In case of $\mathcal{G} = \mathcal{G}_\uparrow$ let*

$$G_o(x) \;:=\; \begin{cases} 0 & \text{if } x \in [0,1], \\ G_A(x) & \text{else.} \end{cases}$$

In case of $\mathcal{G} = \mathcal{G}_{\text{conv}}$ suppose that $\mu(]-\infty,0[), \mu(]0,\infty[) > 0$ and let

$$G_o(x) \;:=\; \begin{cases} 1 + cx & \text{if } x \in [-a,b], \\ G_A(x) & \text{else,} \end{cases}$$

where $a := -c + (c^2+2)^{1/2}$, $b := c + (c^2+2)^{1/2}$, *and* c *is chosen such that* $\int x(G_o - G_A)(x)\,\mu(dx) = 0$.

Then equation (16) is satisfied in both cases. More precisely, the function $\psi := G_o - G_A$ *satisfies the inequalities* $\langle 1, \psi\rangle \geq \|\psi\|^2$, *(15) and*

$$\langle g, \psi\rangle \ \leq\ \langle 1, \psi\rangle - \|\psi\|^2 \quad \textit{whenever } g \in \mathcal{H}_{k,1}, g(0) \geq 0. \tag{18}$$

In case of μ being Lebesgue measure, $\psi = G_o - G_A$ coincides with the function $\psi^{(u)}$ in (11), where $c = 0$ and $a = b = 2^{1/2}$.

The following lemma summarizes essential properties of the optimal kernels $\psi^{(\ell)}$ and $\psi^{(u)}$.

Lemma 8.4. *Let* $\psi^{(\ell)}$ *and* $\psi^{(u)}$ *be the kernel functions in (9) and (11), and let* $h, L > 0$ *and* $t \in \mathbb{R}$.

(a) *If* $\mathcal{G} = \mathcal{G}_\uparrow$, *then* $\langle 1, \psi^{(\ell)}\rangle = \langle 1, \psi^{(u)}\rangle = 1/2$ *and* $\|\psi^{(\ell)}\|^2 = \|\psi^{(u)}\|^2 = 1/3$. *If* $f : \mathbb{R} \to \mathbb{R}$ *satisfies* $f(y) - f(x) \geq L(y-x)$ *for all* $x < y$, *then*

$$f - Lh^{-1}\psi^{(\ell)}_{h,t}, f + Lh^{-1}\psi^{(u)}_{h,t} \ \in\ \mathcal{G}_\uparrow.$$

(b) *If* $\mathcal{G} = \mathcal{G}_{\mathrm{conv}}$, *then* $\langle 1, \psi^{(\ell)}\rangle = 2/3$, $\|\psi^{(\ell)}\|^2 = 8/15$, $\langle 1, \psi^{(u)}\rangle = 2^{2.5}/3$ *and* $\|\psi^{(u)}\|^2 = 2^{4.5}/15$. *Let* $f : \mathbb{R} \to \mathbb{R}$ *be absolutely continuous with derivative* f' *such that* $f'(y) - f'(x) \geq L(y-x)$ *for all* $x < y$. *Then*

$$f - Lh^{-2}\psi^{(\ell)}_{h,t}, f + Lh^{-2}\psi^{(u)}_{h,t} \ \in\ \mathcal{G}_{\mathrm{conv}}.$$

(c) *In general, for any function* $f \in \mathcal{H}_{k,L}$,

$$\begin{aligned}
\Big\langle f(t + h\,\cdot) - r + Lh^k, \psi^{(\ell)}\Big\rangle &\ \geq\ Lh^k\|\psi^{(\ell)}\|^2 \quad \textit{if } f(t) \geq r,\\
\Big\langle f(t + h\,\cdot) - r - Lh^k, \psi^{(u)}\Big\rangle &\ \leq\ -Lh^k\|\psi^{(u)}\|^2 \quad \textit{if } f(t) \leq r.
\end{aligned}$$

Proof of Lemma 8.1. The assertions for $\mathcal{G} = \mathcal{G}_\uparrow$ are a simple consequence of $g \leq g(0)$ on $]-\infty, 0]$ and $g \geq g(0)$ on $[0, \infty[$.

Now let $\mathcal{G} = \mathcal{G}_{\mathrm{conv}}$. If $\psi \geq 0$ and $\int x\psi(x)\,\mu(dx) = 0$, then Condition (15) follows from Jensen's inequality applied to the probability measure $P(dx) = \langle 1, \psi\rangle^{-1}\psi(x)\,\mu(dx)$.

On the other hand, suppose that $\psi \geq 0$ on $[c, d]$ and $\psi \leq 0$ on $\mathbb{R} \setminus [c, d]$, where $c < 0 < d$ and $\mu([-a, c]), \mu([d, b]) > 0$. For $g \in \mathcal{G}_{\mathrm{conv}}$ with $1_[-a,b]g \in L^1(\mu)$, both $g(c)$ and $g(d)$ have to be finite, and we define

$$\tilde{g}(x) \ :=\ g(x) - \begin{cases} d^{-1}(g(d) - g(0))x & \text{if } x \geq 0,\\ c^{-1}(g(c) - g(0))x & \text{if } x \leq 0.\end{cases}$$

By convexity of g, this auxiliary function $\tilde{g}$ satisfies $\tilde{g} \le g(0)$ on $[c,d]$ and $\tilde{g} \ge g(0)$ on $\mathbb{R} \setminus [c,d]$. Thus $\langle \tilde{g}, \psi \rangle \le g(0)\langle 1, \psi \rangle$. If in addition $\int x^{\pm}\psi(x)\,\mu(dx) = 0$, then $\langle g, \psi \rangle = \langle \tilde{g}, \psi \rangle$. $\square$

Proof of Theorem 8.2. One can easily deduce from Lemma 8.1 that the function $\psi = G_A - G_o$ satisfies inequality (14). But G_A is an extremal point of $\mathcal{G}_A$ in the sense that

$$G_A - g \;\in\; \mathcal{G} \quad \text{for any } g \in \mathcal{H}_{k,1}.$$

For let $x < y$. If $\mathcal{G} = \mathcal{G}_{\uparrow}$, then

$$(G_A - g)(y) - (G_A - g)(x) \;=\; y - x - (g(y) - g(x)) \;\ge\; y - x - |y - x| \;=\; 0,$$

whence $G_A - g$ is non-decreasing. In case of $\mathcal{G} = \mathcal{G}_{\mathrm{conv}}$ the same argument applies to the first derivative of $G_A - g$. Together with (14) this implies that

$$\begin{aligned}
\langle g, \psi \rangle &= \langle G_A, \psi \rangle - \langle G_A - g, \psi \rangle \\
&\ge \langle G_A, \psi \rangle - (G_A - g)(0)\langle 1, \psi \rangle \\
&= \langle G_A, \psi \rangle + g(0)\langle 1, \psi \rangle \\
&= \|\psi\|^2 + \langle G_o, \psi \rangle + g(0)\langle 1, \psi \rangle \\
&= \|\psi\|^2 + (g(0) - 1)\langle 1, \psi \rangle.
\end{aligned}$$

The latter equation follows from $\langle G_o, \psi \rangle = \langle -1, \psi \rangle$, which is easily verified. The special case $g = 0$ yields the inequality $\langle 1, \psi \rangle \ge \|\psi\|^2$. Then inequality (17) becomes obvious.

It remains to be shown that in case of $\mathcal{G} = \mathcal{G}_{\mathrm{conv}}$ there exist numbers $a, b \ge 2^{1/2}$ such that $\psi = \psi(\cdot, a, b)$ satisfies $\int x^{\pm}\psi(x)\,\mu(dx) = 0$. In fact, for any fixed x the number $\psi(x, a, b) \le 1$ can be shown to be continuous and decreasing in a and b. Precisely, $\psi(0, a, b) = 1$ and $\lim_{a\to\infty} \psi(x, a, \cdot) = \lim_{b\to\infty} \psi(y, \cdot, b) = -\infty$ for $x < 0 < y$. Hence the assertion is a consequence of monotone convergence. $\square$

Proof of Theorem 8.3. This proof is analogous to the proof of Theorem 8.2 and thus omitted. $\square$

Proof of Lemma 8.4. The calculations of $\langle 1, \psi \rangle$ and $\|\psi\|^2$ are elementary and thus omitted. Elementary calculations show that $g := -Lh^{-k}\psi_{t,h}^{(\ell)}$ as well as $g := Lh^{-k}\psi^{(u)}$ satisfies

$$\left.\begin{array}{c} g(y) - g(x) \\ g'(y) - g'(x) \end{array}\right\} \ge -L(y - x) \quad \text{if } \mathcal{G} = \left\{\begin{array}{c} \mathcal{G}_{\uparrow}, \\ \mathcal{G}_{\mathrm{conv}}, \end{array}\right.$$

where $g'(x)$ denotes any number between the right- and left-sided derivative of g at x. Thus $f + g$ belongs to $\mathcal{G}$, whenever f satisfies the inequalities stated in parts (a) and (b).

As for part (c), for $f \in \mathcal{H}_{k,L}$ and $t \in \mathbb{R}$, $h, c > 0$ the function $cf(t + h\,\cdot)$ belongs to $\mathcal{H}_{k,cLh^k}$. If we take $c := (Lh^k)^{-1}$, the inequality (17) implies that

$$\begin{aligned} \left\langle f(t+h\,\cdot) - r + Lh^k, \psi^{(\ell)} \right\rangle &= Lh^k \left\langle c(f(t+h\,\cdot) - f(t)) + 1, \psi^{(\ell)} \right\rangle \\ &\geq Lh^k \|\psi^{(\ell)}\|^2. \end{aligned}$$

Analogously one can deduce the lower bound for $\langle f(t+h\,\cdot) - r - Lh^k, \psi^{(u)} \rangle$. $\square$

Acknowledgements. The author is grateful to Lars Hömke for his assistance in Section 5. Constructive comments of a referee and an associate editor helped to improve the presentation. This work has been supported by Deutsche Forschungsgemeinschaft, grant Du 238/5-1.

References

BICKEL, P.J. AND M. ROSENBLATT (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1**, 1071–1095

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

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

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

DÜMBGEN, L. (1998). New goodness-of-fit tests and their application to nonparametric confidence sets. *Ann. Statist.* **26**, 288–314

DÜMBGEN, L. (2007). Confidence bands for convex median functions using sign tests. In: *Asymptotics: Particles, Processes and Inverse Problems (E. Cator, G. Jongbloed, C. Kraaikamp, R. Lopuhaä, J.A. Wellner, eds.)*, pp. 85-100. Lecture Notes - Monograph Series **55**, IMS, Hayward, USA.

DÜMBGEN, L. AND R.B. JOHNS (2004). Confidence bands for isotonic median functions using sign tests. *J. Comp. Graph. Statist.* **13**, 519–533

DÜMBGEN, L. AND V.G. SPOKOINY (2001). Multiscale testing of qualitative hypotheses. *Ann. Statist.* **29**, 124–152

EUBANK, R.L. AND P.L. SPECKMAN (1993). Confidence bands in nonparametric regression. *J. Amer. Statist. Assoc.* **88**, 1287–1301

FAN, J. AND W. ZHANG (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scand. J. Statist.* **27**, 715–731

FAN, J., C. ZHANG AND J. ZHANG (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Statist.* **29**, 153–193

GRAMA, I. AND M. NUSSBAUM (1998). Asymptotic equivalence for nonparametric generalized linear models. *Prob. Theory and Related Fields* **111**, 167–214.

HÄRDLE, W. AND J.S. MARRON (1991). Bootstrap simultaneous error bars for nonparametric regression. *Ann. Statist.* **19**, 778–796

HALL, P. AND D.M. TITTERINGTON (1988). On confidence bands in nonparametric density estimation. *J. Multivar. Anal.* **27**, 228–254

HART, J.D. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests.* Springer, New York

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

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

KNAFL, G., J. SACHS AND D. YLVISAKER (1985). Confidence bands for regression functions. *J. Amer. Statist. Assoc.* **80**, 683–691

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

ROBERTSON, T., F.T. WRIGHT AND R.L. DYKSTRA (1988). *Order Restricted Statistical Inference.* Wiley, New York

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