

LOCALLY ADAPTIVE CONFIDENCE BANDS¹

BY TIM PATSCHKOWSKI AND ANGELIKA ROHDE

Ruhr-Universität Bochum and Albert-Ludwigs-Universität Freiburg

We develop honest and locally adaptive confidence bands for probability densities. They provide substantially improved confidence statements in case of inhomogeneous smoothness, and are easily implemented and visualized. The article contributes conceptual work on locally adaptive inference as a straightforward modification of the global setting imposes severe obstacles for statistical purposes. Among others, we introduce a statistical notion of local Hölder regularity and prove a correspondingly strong version of local adaptivity. We substantially relax the straightforward localization of the self-similarity condition in order not to rule out prototypical densities. The set of densities permanently excluded from the consideration is shown to be pathological in a mathematically rigorous sense. On a technical level, the crucial component for the verification of honesty is the identification of an asymptotically least favorable stationary case by means of Slepian's comparison inequality.

1. Introduction. Let X_1, \dots, X_n be independent real-valued random variables which are identically distributed according to some unknown probability measure \mathbb{P}_p with Lebesgue density p . Assume that p belongs to a nonparametric function class \mathcal{P} . For any interval $[a, b]$ and any significance level $\alpha \in (0, 1)$, a confidence band for p , described by a family of random intervals $C_{n,\alpha}(t)$, $t \in [a, b]$, is said to be (asymptotically) honest with respect to \mathcal{P} if the coverage inequality

$$(1.1) \quad \liminf_{n \rightarrow \infty} \inf_{p \in \mathcal{P}} \mathbb{P}_p^{\otimes n}(p(t) \in C_{n,\alpha}(t) \text{ for all } t \in [a, b]) \geq 1 - \alpha$$

is satisfied. Adaptive confidence sets maintain specific coverage probabilities over a large union of models while shrinking at the fastest possible nonparametric rate simultaneously over all submodels. If \mathcal{P} is some class of densities within a union of Hölder balls $\mathcal{H}(\beta, L)$ with fixed radius $L > 0$, the confidence band is called globally adaptive over $\bigcup_{\beta \in [\beta_*, \beta^*]} (\mathcal{P} \cap \mathcal{H}(\beta, L))$ within a range $[\beta_*, \beta^*] \subset (0, \infty)$ [cf. Cai and Low (2004)], if for every $\beta \in [\beta_*, \beta^*]$ and for every $\varepsilon > 0$ there exists

Received August 2017; revised January 2018.

¹Supported by the DFG Collaborative Research Center 823, Subproject C1, and DFG Research Grant RO 3766/4-1.

MSC2010 subject classifications. 62G07, 62G15.

Key words and phrases. Local adaptivity, confidence bands in density estimation.

some constant $c > 0$, such that

$$(1.2) \quad \limsup_{n \rightarrow \infty} \sup_{\substack{p \in \mathcal{P}: \\ p \in \mathcal{H}(\beta, L)}} \mathbb{P}_p^{\otimes n} \left(\sup_{t \in (a, b)} |C_{n, \alpha}(t)| \geq c \cdot r_n(\beta) \right) < \varepsilon.$$

Here, $|C_{n, \alpha}(t)|$ denotes the length of $C_{n, \alpha}(t)$, and $r_n(\beta)$ the minimax-optimal speed of convergence

$$\inf_{\hat{T}_n \text{ estimator}} \sup_{p \in \mathcal{H}(\beta, L) \cap \mathcal{P}} \mathbb{E}_p^{\otimes n} \left[\sup_{t \in \mathbb{R}} |\hat{T}_n(t) - p(t)| \right]$$

for estimation under supremum norm loss over $\mathcal{H}(\beta, L) \cap \mathcal{P}$, possibly inflated by additional logarithmic factors. Note that a logarithmic payment for adaptation is neither avoidable for pointwise confidence intervals nor for pointwise estimation; see Lepskiĭ (1990). Under the so-called self-similarity condition on \mathcal{P} , Giné and Nickl (2010) succeeded to construct confidence bands satisfying both (1.1) and (1.2). Here, the minimax-optimal speed of convergence over $\mathcal{H}(\beta, L) \cap \mathcal{P}$ coincides with the classical rate

$$\left(\frac{\log n}{n} \right)^{\frac{\beta}{2\beta+1}}.$$

They are of the form

$$(1.3) \quad [\hat{p}_n(t) - \sqrt{\hat{p}_n(t)} \cdot \hat{\Delta}_n(\alpha), \hat{p}_n(t) + \sqrt{\hat{p}_n(t)} \cdot \hat{\Delta}_n(\alpha)], \quad t \in [a, b],$$

with an estimator \hat{p}_n of the density p , and a data-driven width parameter $\hat{\Delta}_n(\alpha)$ depending on the significance level α . Although the confidence band’s width depends on t via $\sqrt{\hat{p}_n(t)}$, the stochastic order of the width is *independent of t* as the densities under consideration are assumed to be uniformly bounded away from zero and infinity. However, even one small wiggly part of the density inhibits stronger performance of the procedure in smooth segments. Ideally, a confidence band is automatically thinner in regions where the unknown density is smooth and wider in less smooth parts. Although a plethora of articles dealing with the central problem of local adaptation in the estimation framework has been published over the last decades, the substantially harder problem of locally adaptive confidence bands has not been addressed in the literature. We call a confidence band locally adaptive if for every $\varepsilon > 0$ there exists some constant $c > 0$, such that the confidence band satisfies the stronger performance guarantee

$$(1.4) \quad \sup_{\substack{U \subset [a, b] \\ U \text{ open interval}}} \limsup_{n \rightarrow \infty} \sup_{\substack{p \in \mathcal{P} \\ p|_{U_\delta} \in \mathcal{H}_{U_\delta}(\beta, L)}} \mathbb{P}_p^{\otimes n} \left(\sup_{t \in U} |C_{n, \alpha}(t)| \geq c \cdot r_n(\beta) \right) < \varepsilon,$$

for any $\delta > 0$, ideally for any β in the range of adaptation. Here, U_δ denotes the open δ -enlargement of U , $p|_{U_\delta}$ the restriction of p on U_δ and $\mathcal{H}_{U_\delta}(\beta, L)$ the Hölder ball with radius L of functions from U_δ to \mathbb{R} which are Hölder continuous to the

exponent β . The new contribution of this article is the construction and theoretical investigation of such locally adaptive confidence bands, that is, honest confidence bands with locally adaptive rather than globally adaptive width, which incorporate potentially inhomogeneous regularity of the target function. Typically, $r_n(\beta')/r_n(\beta)$ decays to zero whenever $\beta' > \beta$, implying that (1.4) guarantees significantly tighter confidence bands in case of inhomogeneous smoothness as compared to (1.2). In this case, any confidence band with (possibly) random but t -independent width cannot satisfy (1.4), whenever \mathcal{P} contains functions with inhomogeneous smoothness.

Our new confidence band appealingly relies on a discretized evaluation of a modified Lepski-type kernel density estimator, including an additional supremum in the empirical bias term in the bandwidth selection criterion. A suitable discretization of the interval $[a, b]$ and a locally constant approximation of both the density estimator and the (random) bandwidth allow to piece the segmentwise confidence statements together to obtain a continuum of confidence statements over $[a, b]$. Due to the discretization, the band is computable and feasible from a practical point of view without losing optimality between the mesh points. The t -dependence of $|C_{n,\alpha}(t)|$, $t \in [a, b]$, reflected in the t -dependence of the density estimator's bandwidth, makes the asymptotic calibration of the confidence band to the level α highly nontrivial. Whereas the analysis of the related globally adaptive procedure of Giné and Nickl (2010) reduces to the limiting distribution of the supremum of a stationary Gaussian process, our locally adaptive approach leads to a highly nonstationary situation. A crucial component is therefore the identification of a stationary process as a least favorable case by means of Slepian's comparison inequality, subsequent to a Gaussian reduction using recent techniques of Chernozhukov, Chetverikov and Kato (2014b).

In view of a series of negative results starting with Low (1997), the class of densities has to be restricted for the purpose of honest and adaptive inference. Giné and Nickl (2010) succeeded to construct honest and globally adaptive confidence bands under the so-called self-similarity condition; see Picard and Tribouley (2000). A corresponding condition does not exist for the purpose of local adaptation, and a straightforward localization of the global self-similarity condition imposes severe obstacles for statistical purposes as it rules out prototypical densities. Consequently, we develop a suitable condition under which honest and locally adaptive confidence bands provably exist while representative densities remain included. The set of permanently excluded densities is shown to be pathological in a mathematically rigorous sense.

The main contributions of this article are the following:

- (i) We first develop honest confidence bands which are locally adaptive in the sense of (1.4). Additionally, an even stronger notion of local adaptivity is introduced and proved to be satisfied. These explicitly constructed confidence bands provide substantially improved confidence statements in case of inhomogeneous smoothness.

(ii) Our confidence bands are computable and computationally feasible. The performance is demonstrated in a simulation study.

(iii) The design of a suitably restricted class of densities tailored to *local* adaptation is a challenging task. On the one hand, the self-similarity condition, suited for global adaptation, is too weak for the purpose of honesty and *local* adaptivity. On the other hand, an adequate local condition is supposed not to rule out too many densities. We design a new restricted class of densities \mathcal{P} for which both honesty and local adaptivity are achievable. We prove that the class is massive and, therefore, suitable for statistical purposes in two senses. First, the pointwise minimax rate of convergence remains unchanged when passing from the class $\mathcal{H}(\beta, L)$ to $\mathcal{P} \cap \mathcal{H}(\beta, L)$. Second, the set of permanently excluded densities is shown to be pathological in a mathematically rigorous sense.

(iv) On a technical level, the calibration of the confidence band leads to the distributional approximation by the supremum of a highly nonstationary Gaussian process depending on the unknown density p . Therefore, the crucial ingredient is the identification of a least favorable stationary case by means of Slepian's comparison inequality, which does not depend on p anymore.

Our results are exemplarily formulated in the density estimation framework but can be mimicked in other nonparametric models. To keep the representation concise, we restrict the theory to locally adaptive kernel density estimators. The ideas can be transferred to wavelet estimators to a large extent as has been done for globally adaptive confidence bands in [Giné and Nickl \(2010\)](#).

The article is organized as follows. Basic notation are introduced in Section 2. Section 3 presents the main contributions, that is a substantially relaxed localized self-similarity condition in Section 3.1, the construction and in particular the asymptotic calibration of the confidence band in Section 3.2 as well as its strong local adaptivity properties in Section 3.3. Important supplementary results are postponed to Section 4, whereas Section 5 presents some of the proofs of the main results. The Supplementary Material [[Patschkowski and Rohde \(2019\)](#)] contains the remaining proofs, technical tools for the main proofs, as well as an extended simulation study.

2. Preliminaries and notation. Let X_1, \dots, X_n , $n \geq 4$, be independent random variables identically distributed according to some unknown probability measure \mathbb{P}_p on \mathbb{R} with continuous Lebesgue density p . Subsequently, we consider kernel density estimators

$$\hat{p}_n(\cdot, h) = \frac{1}{n} \sum_{i=1}^n K_h(X_i - \cdot)$$

with bandwidth $h > 0$ and rescaled kernel $K_h(\cdot) = h^{-1}K(\cdot/h)$. If not stated otherwise, K is measurable and symmetric with support contained in $[-1, 1]$, integrating to one, and of bounded variation. Furthermore, a kernel K is said to be of

order $l \in \mathbb{N}$ if

$$\int x^j K(x) dx = 0 \quad \text{for } 1 \leq j \leq l \quad \text{and} \quad \int x^{l+1} K(x) dx \neq 0.$$

For some measure Q , we denote by $\|\cdot\|_{L^p(Q)}$ the L^p -norm with respect to Q . Is Q the Lebesgue measure, we just write $\|\cdot\|_p$. For any interval $U \subset \mathbb{R}$ and any bounded function $f : U \rightarrow \mathbb{R}$, we denote by

$$\|f\|_U = \sup_{x \in U} |f(x)|$$

the supremum norm of f over U . If $U = \mathbb{R}$, we simply write $\|\cdot\|_{\text{sup}}$ for $\|\cdot\|_{\mathbb{R}}$. If well defined,

$$(f_1 * f_2)(\cdot) = \int f_1(u) f_2(\cdot - u) du$$

denotes the convolution of two functions $f_1, f_2 : \mathbb{R} \rightarrow \mathbb{R}$. With

$$\lfloor \beta \rfloor = \max\{n \in \mathbb{N} \cup \{0\} : n < \beta\},$$

the Hölder class $\mathcal{H}_U(\beta)$ to the parameter $\beta > 0$ on the open interval $U \subset \mathbb{R}$ is defined as the set of functions $f : U \rightarrow \mathbb{R}$ admitting derivatives up to the order $\lfloor \beta \rfloor$ and having finite Hölder norm

$$\|f\|_{\beta,U} = \sum_{k=0}^{\lfloor \beta \rfloor} \|f^{(k)}\|_U + \sup_{\substack{x,y \in U \\ x \neq y}} \frac{|f^{(\lfloor \beta \rfloor)}(x) - f^{(\lfloor \beta \rfloor)}(y)|}{|x - y|^{\beta - \lfloor \beta \rfloor}} < \infty.$$

The corresponding Hölder ball with radius $L > 0$ is denoted by

$$\mathcal{H}_U(\beta, L) = \{f \in \mathcal{H}_U(\beta) : \|f\|_{\beta,U} \leq L\}.$$

With the definition of $\|\cdot\|_{\beta,U}$, the Hölder balls are nested, that is, $\mathcal{H}_U(\beta_2, L) \subset \mathcal{H}_U(\beta_1, L)$ for $0 < \beta_1 \leq \beta_2 < \infty$ and $|U| < 1$. Finally, $\mathcal{H}_U(\infty, L) = \bigcap_{\beta > 0} \mathcal{H}_U(\beta, L)$ and $\mathcal{H}_U(\infty) = \bigcap_{\beta > 0} \mathcal{H}_U(\beta)$. Subsequently, for any real function $f(\beta)$, the expression $f(\infty)$ is to be read as $\lim_{\beta \rightarrow \infty} f(\beta)$, provided that this limit exists. Additionally, the class of probability densities p , such that $p|_U$ is contained in the Hölder class $\mathcal{H}_U(\beta, L)$ is denoted by $\mathcal{P}_U(\beta, L)$. The indication of U is omitted when $U = \mathbb{R}$.

3. Main results. In this section, we pursue the new approach of locally adaptive confidence bands and present the main contribution of this article.

For the new challenge of locally adaptive confidence bands, a condition of admissibility necessarily has to be introduced. Although this condition is tailored to the construction of the confidence band, this is the logical first step because the calibration of the band to the level α explicitly involves the class of admissible functions. In Section 3.1, we define and motivate the class of admissible densities \mathcal{P}_n (containing densities with smaller and smaller Lipschitz constants for growing n).

While not claiming the admissibility condition to be weakest possible, we prove in view of statistical purposes that \mathcal{P}_n is massive in two senses. First, the pointwise minimax-rates do not change when passing from $\mathcal{P}(\beta, L)$ to $\mathcal{P}(\beta, L) \cap \mathcal{P}_n$ within the aspired range of adaptation, and second, the set of permanently excluded densities is shown to be pathological in a sense of Baire categories. Proving these results, we have gained new insight into analytical properties of the Weierstraß function, which are of independent interest while providing deeper understanding of the admissibility condition. They are deferred to the Supplementary Material [Patschkowski and Rohde (2019)].

In Section 3.2, we develop the new confidence band $(C_{n,\alpha}(t))_{t \in [a,b]}$. For simplicity, $[a, b] = [0, 1]$ from now on. Here, we are facing two challenges. First, the construction has to be computable and visualizable, and to perform well in practice. As local adaptation is generically carried out separately at every point $t \in [0, 1]$, a suitable procedure is far from being straightforward. Second, the construction has to be calibrated to a prespecified significance level, uniformly over the class of admissible densities. The calibration turns out to be complex because the distributional approximation of the statistic involves the supremum over a highly nonstationary Gaussian process even depending on the unknown density. The innovative point for the calibration is the identification of a least favorable stationary case, which does not depend on the unknown density anymore.

Finally, in Section 3.3, we analyze the performance of our confidence band. Besides verifying property (1.4), we introduce an even stronger notion of local adaptivity, which is statistically even more informative. We prove that the confidence band also possesses this strong local adaptivity property.

3.1. *Admissible functions.* If \mathcal{P} equals the set of *all* densities contained in

$$\bigcup_{0 < \beta \leq \beta^*} \mathcal{H}(\beta, L),$$

honest and globally adaptive confidence bands provably do not exist although adaptive estimation is possible; see the pioneering contribution of Low (1997). Numerous attempts have been made to tackle this adaptation problem in alternative formulations. Whereas Genovese and Wasserman (2008) relax the coverage property and do not require the confidence band to cover the function itself but a simpler surrogate function capturing the original function's significant features, most of the approaches are based on a restriction of the parameter space. Under qualitative shape constraints, Hengartner and Stark (1995), Dümbgen (1998), Dümbgen (2003) and Davies, Kovac and Meise (2009) achieve adaptive inference. Within the models of nonparametric regression and Gaussian white noise, Picard and Tribouley (2000) investigate on pointwise adaptive confidence intervals under a self-similarity condition on the parameter space; see also Kueh (2012) for thresholded needlet estimators. Under a similar condition, Giné and Nickl (2010) even develop asymptotically honest confidence bands for probability densities whose

width is adaptive to the global Hölder exponent. Bull (2012) works under a slightly weakened version of the self-similarity condition. Kerkycharian, Nickl and Picard (2012) develop corresponding results in the context of needlet density estimators on compact homogeneous manifolds. Under the same type of self-similarity condition, adaptive confidence bands are developed under a considerably generalized Smirnov–Bickel–Rosenblatt assumption based on Gaussian multiplier bootstrap; see Chernozhukov, Chetverikov and Kato (2014a). Hoffmann and Nickl (2011) introduce a nonparametric distinguishability condition, under which adaptive confidence bands exist for finitely many models under consideration. Their condition is shown to be necessary and sufficient.

Similar important conclusions concerning adaptivity in terms of confidence statements are obtained under Hilbert space geometry with corresponding L^2 -loss; see Juditsky and Lambert-Lacroix (2003), Baraud (2004), Genovese and Wasserman (2005), Cai and Low (2006), Robins and van der Vaart (2006), Bull and Nickl (2013) and Nickl and Szabó (2016). Concerning L^p -loss, we also draw attention to Carpentier (2013).

Our subsequently introduced notion of admissibility aligns to the (global) self-similarity condition. Recall that $f_1 * f_2$ denotes the convolution of two functions f_1 and f_2 , and $K_h(\cdot) = h^{-1}K(\cdot/h)$ is the rescaled kernel corresponding to the bandwidth $h > 0$.

CONDITION 3.1 [Global self-similarity condition; Picard and Tribouley (2000), Giné and Nickl (2010)]. Suppose $p \in \mathcal{H}(\beta, L^*)$ for some $\beta \in [\beta_*, \beta^*]$ with $\beta^* = l + 1$ and l the order of the kernel K , and assume that there exist a positive real constant b_1 and a positive integer j_0 such that for every integer $j \geq j_0$,

$$\frac{b_1}{2^j \beta} \leq \|K_{2^{-j}} * p - p\|_{\text{sup}}.$$

Giné and Nickl (2010) construct globally adaptive confidence bands over the set

$$(3.1) \quad \bigcup_{\beta_* \leq \beta \leq \beta^*} \left\{ p \in \mathcal{P}(\beta, L) : p \geq \delta \text{ on } [-\varepsilon, 1 + \varepsilon], \right. \\ \left. \frac{c}{2^j \beta} \leq \|K_{2^{-j}} * p - p\|_{\text{sup}} \text{ for all } j \geq j_0 \right\}$$

for some constant $c > 0$ and $0 < \varepsilon < 1$. They work on the scale of Hölder–Zygmund rather than Hölder classes. For this reason, they include the corresponding bias upper bound condition which is not automatically satisfied for $\beta = \beta^*$ in that case.

REMARK 1. As mentioned in Giné and Nickl (2010), if $K(\cdot) = \frac{1}{2}\mathbb{1}\{\cdot \in [-1, 1]\}$ is the rectangular kernel, all twice differentiable densities p that are sup-

ported in a fixed compact interval satisfy the lower bound constraint

$$(3.2) \quad \|K_{2^{-j}} * p - p\|_{\text{sup}} \geq c \cdot 2^{-2j} + o(2^{-2j})$$

with a constant $c > 0$. The reason is that due to the constraint of being a probability density, $\|p''\|_{\text{sup}}$ is bounded away from zero uniformly over this class, in particular p'' cannot vanish everywhere. That is, Condition 3.1 does not appear to be restrictive.

From Condition 3.1, we can straightforwardly deduce a sufficient condition on the class of densities under consideration for the new problem of honest and locally adaptive confidence bands as follows.

CONDITION 3.2 (Local self-similarity condition). There exist a positive real constant b_1 and a positive integer j_0 such that for any nondegenerate interval $(v, w) \subset [0, 1]$, there exists some $\beta \in [\beta_*, \beta^*]$ with $\beta^* = l + 1$ and l the order of the kernel K , such that

$$(3.3) \quad p|_{(v,w)} \in \mathcal{H}_{(v,w)}(\beta, L^*)$$

and

$$(3.4) \quad \frac{b_1}{2^{j\beta}} \leq \|K_{2^{-j}} * p - p\|_{(v+2^{-j}, w-2^{-j})}$$

are satisfied for all $j \geq j_0 \vee \log_2(1/(w - v))$.

However, a condition like (3.4) rules out examples which seem to be typical to statisticians:

(i) In contrast to the observation in Remark 1, for any density p , $\|p''\|_U$ may vanish for subintervals U within the support of p . As a consequence, the lower bound condition (3.4) is violated on such subintervals U for every $\beta \in (0, \beta^*]$. [Recall that the kernel K is symmetric (see Section 2) and hence of order $l \geq 1$.]

EXAMPLE 3.3. Assume that the kernel K is of order $l \geq 1$, and recall $\beta^* = l + 1$. Then (3.4) excludes for instance the triangular density

$$(3.5) \quad p(t) = \max\{1 - |t - 1/2|, 0\}, \quad t \in \mathbb{R},$$

because the second derivative exists and vanishes when restricted to any open interval $U \subset [0, 1/2) \cup (1/2, 1]$.

For the same reason, densities with a constant piece are excluded. In general, if p restricted to U is a polynomial of order at most l , (3.4) is violated as the left-hand side is not equal to zero. At the same time, the kernel density estimator is bias-free in these regions, for which reason it cannot be necessary to exclude these examples from consideration.

(ii) For $p \in \mathcal{P}(\beta_*, L)$ and any fixed $h > 0$, the map

$$t \mapsto \|K_{2^{-j}} * p - p\|_{(t-h+2^{-j}, t+h-2^{-j})}$$

is continuous for any natural number j with $2^{-j} < h$. At the same time, the map

$$(3.6) \quad t \mapsto \sup\{\beta \leq \beta^* : p|_{(t-h, t+h)} \in \mathcal{H}_{(t-h, t+h)}(\beta, L)\}$$

may be discontinuous, in which case the local self-similarity condition is violated.

EXAMPLE 3.3 (Continued). We consider again the triangular density in (3.5). Then

$$\begin{aligned} & \sup\{\beta \leq \beta^* : p|_{(t-h, t+h)} \in \mathcal{H}_{(t-h, t+h)}(\beta, 1)\} \\ &= \begin{cases} 1 & \text{if } t \in \left(\frac{1}{2} - h, \frac{1}{2} + h\right), \\ \beta^* & \text{if } t \in [0, 1] \setminus \left(\frac{1}{2} - h, \frac{1}{2} + h\right). \end{cases} \end{aligned}$$

In view of the deficiencies described in (i) and (ii), it is insufficient just to replace the global self-similarity condition by the local self-similarity condition for the purpose of locally adaptive confidence bands.

Instead, we introduce Condition 3.5. Before, to unify notation, we define the β^* -capped Hölder norm.

DEFINITION 3.4 (β^* -capped Hölder norm). For $\beta > 0$, for some bounded open interval $U \subset \mathbb{R}$, and $p : U \rightarrow \mathbb{R}$ with $p \in \mathcal{H}_U(\beta)$, define the β^* -capped Hölder norm

$$\|p\|_{\beta, \beta^*, U} = \sum_{k=0}^{\lfloor \beta \wedge \beta^* \rfloor} \|p^{(k)}\|_U + \sup_{\substack{x, y \in U \\ x \neq y}} \frac{|p^{(\lfloor \beta \wedge \beta^* \rfloor)}(x) - p^{(\lfloor \beta \wedge \beta^* \rfloor)}(y)|}{|x - y|^{\beta - \lfloor \beta \wedge \beta^* \rfloor}},$$

whenever the expression is finite.

Note that if $\beta - \lfloor \beta \wedge \beta^* \rfloor > 1$, then $\|p\|_{\beta, \beta^*, U}$ can only be finite if $p|_U^{(\lfloor \beta \rfloor)}$ is constant, in which case

$$p|_U^{(\beta^*)} \equiv 0.$$

If for some open interval $U \subset [0, 1]$, the derivative $p|_U^{(\beta^*)}$ exists and equals zero restricted to U , then $\|p\|_{\beta, \beta^*, U}$ is finite uniformly over all $\beta > 0$. If it exists and is not identical to the zero function on U , then $\|p\|_{\beta, \beta^*, U}$ is finite if and only if $\beta \leq \beta^*$ as a consequence of the mean value theorem. That is,

$$\sup\{\beta \in (0, \infty] : p|_U \in \mathcal{H}_{\beta^*, U}(\beta, L^*)\} \in (0, \beta^*] \cup \{\infty\}.$$

Correspondingly, define the β^* -capped Hölder ball and β^* -capped Hölder class by

$$(3.7) \quad \mathcal{H}_{\beta^*,U}(\beta, L) = \{p \in \mathcal{H}_U(\beta) : \|p\|_{\beta,\beta^*,U} \leq L\}$$

and

$$(3.8) \quad \mathcal{H}_{\beta^*,U}(\beta) = \{p \in \mathcal{H}_U(\beta) : \|p\|_{\beta,\beta^*,U} < \infty\},$$

respectively. As verified in Lemma A.11 in the Supplementary Material [Patschkowski and Rohde (2019)], $\|p\|_{\beta_1,\beta^*,U} \leq \|p\|_{\beta_2,\beta^*,U}$ for $0 < \beta_1 \leq \beta_2 < \infty$ and $|U| \leq 1$. Finally, denote $\mathcal{H}_{\beta^*,U}(\infty, L) = \bigcap_{\beta>0} \mathcal{H}_{\beta^*,U}(\beta, L)$ and $\mathcal{H}_{\beta^*,U}(\infty) = \bigcap_{\beta>0} \mathcal{H}_{\beta^*,U}(\beta)$.

Recall the definition $\|f\|_U = \sup_{t \in U} |f(t)|$ for any subset $U \subset \mathbb{R}$ and bounded $f : U \rightarrow \mathbb{R}$.

ADMISSIBILITY CONDITION 3.5. For sample size $n \in \mathbb{N}$, some $0 < \varepsilon < 1$, $0 < \beta_* < 1$, and $L^* > 0$, a density p is said to be admissible if $p \in \mathcal{P}_{(-\varepsilon, 1+\varepsilon)}(\beta_*, L^*)$ and the following holds true: for any $t \in [0, 1]$ and for any $h \in \mathcal{G}_\infty$ with

$$(3.9) \quad \mathcal{G}_\infty = \{2^{-j} : j \in \mathbb{N}, j \geq j_{\min} = \lceil 2 \vee \log_2(2/\varepsilon) \rceil\},$$

there exists some $\beta \in [\beta_*, \beta^*] \cup \{\infty\}$ such that the following conditions are satisfied for $u = h$ or $u = 2h$:

$$(3.10) \quad p|_{(t-u, t+u)} \in \mathcal{H}_{\beta^*,(t-u, t+u)}(\beta, L^*)$$

and

$$(3.11) \quad \|K_g * p - p\|_{(t-(u-g), t+(u-g))} \geq \frac{g^\beta}{\log n}$$

for all $g \in \mathcal{G}_\infty$ with $g \leq u/8$.

The set of admissible densities is denoted by $\mathcal{P}_n^{\text{adm}} = \mathcal{P}_n^{\text{adm}}(K, \beta_*, L^*, \varepsilon)$.

The new problem of locally adaptive confidence bands requires a new type of restriction for the class of densities under consideration. On the one hand, our formulated local self-similarity Condition 3.2 is sufficient, but limits the statistical usability dramatically on the other hand. Contrarily, the weaker Condition 3.5 incorporates the following three crucial aspects.

(i) Passing from the Hölder norm to the β^* -capped Hölder norm enlarges the set of densities under consideration. First of all, densities which restricted to $[0, 1]$ are described by a polynomial of order at most l are now included. Here, the order l is a natural limit because a kernel of order l is bias-free for polynomials up to the order l , that is, for any $0 < h < 1/2$,

$$\mathbb{E}_p^{\otimes n} \hat{p}_n(t, h) = p(t), \quad t \in [h, 1 - h].$$

(ii) We relax the requirement of (3.3) and (3.4) to hold for every interval (v, w) by requiring (3.10) and (3.11) to be satisfied for $u = h$ or $u = 2h$. It turns out to be essential for incorporating densities with abrupt changes in the smoothness behavior.

(iii) The collection of admissible densities is increasing with the number of observations, that is $\mathcal{P}_n^{\text{adm}} \subset \mathcal{P}_{n+1}^{\text{adm}}$, $n \in \mathbb{N}$. The logarithmic denominator even weakens the assumption for growing sample size, permitting smaller and smaller Lipschitz constants. Note that a generic lower bound as (3.2) in Remark 1 is locally not natural.

The benefit of (i)–(iii) is demonstrated in the following example.

EXAMPLE 3.3 (Continued). If K is the rectangular kernel and L^* is sufficiently large, the triangular density $p(t) = \max\{1 - |t - 1/2|, 0\}$, $t \in \mathbb{R}$, is (eventually—for sufficiently large n) admissible. It is globally not smoother than Lipschitz, and the bias lower bound condition (3.11) is (eventually) satisfied for $\beta = 1$ and pairs (t, h) with $|t - 1/2| < (7/8)h$. Although the bias lower bound condition to the exponent $\beta^* = 2$ is not satisfied for any (t, h) with $t \in [0, 1] \setminus (1/2 - h, 1/2 + h)$, these tuples (t, h) fulfill (3.10) and (3.11) for $\beta = \infty$, which is not excluded anymore by Condition 3.5. Finally, if the conditions (3.10) and (3.11) are not simultaneously satisfied for some pair (t, h) with

$$\frac{7}{8}h < \left|t - \frac{1}{2}\right| < h,$$

then they are fulfilled for the pair $(t, 2h)$ and $\beta = 1$, because $|t - 1/2| < (7/8)2h$.

We now denote by

$$\mathcal{P}_n = \mathcal{P}_n(K, \beta_*, L^*, \varepsilon, M) = \left\{p \in \mathcal{P}_n^{\text{adm}}(K, \beta_*, L^*, \varepsilon) : \inf_{x \in [-\varepsilon, 1 + \varepsilon]} p(x) \geq M\right\}$$

the set of admissible densities being bounded below by $M > 0$ on $[-\varepsilon, 1 + \varepsilon]$. We restrict our considerations to combinations of parameters for which the class \mathcal{P}_n is nonempty.

The remaining results of this subsection are about the massiveness of the function classes \mathcal{P}_n . They are stated for the particular case of the rectangular kernel. Other kernels may be treated with the same idea; verification of (3.11) however appears to require a case-by-case analysis for different kernels. The following proposition demonstrates that the pointwise minimax rate of convergence remains unchanged when passing from the class $\mathcal{H}(\beta, L^*)$ to $\mathcal{P}_n \cap \mathcal{H}(\beta, L^*)$.

PROPOSITION 3.6 (Lower pointwise risk bound). *For the rectangular kernel K_R there exists some constant $M > 0$, such that for any $t \in [0, 1]$, for any $\beta \in$*

$[\beta_*, 1]$, for any $0 < \varepsilon < 1$, and for any $k \geq k_0(\beta_*)$ there exists some $x > 0$ and some $L(\beta) > 0$ with

$$\inf_{T_n} \sup_{\substack{p \in \mathcal{P}_k: \\ P_{|(-\varepsilon, 1+\varepsilon)} \in \mathcal{H}_{(-\varepsilon, 1+\varepsilon)}(\beta, L)}} \mathbb{P}_p^{\otimes n} (n^{\frac{\beta}{2\beta+1}} |T_n(t) - p(t)| \geq x) > 0$$

for all $L \geq L(\beta)$ and for all $n \geq n_0$, for the class $\mathcal{P}_k = \mathcal{P}_k(K_R, \beta_*, L^*, \varepsilon, M)$, where the infimum is running over all estimators T_n based on X_1, \dots, X_n .

Note that the classical construction for the sequence of hypotheses in order to prove minimax lower bounds consists of a smooth density distorted by small β -smooth perturbations, properly scaled with the sample size n . However, not all of its members satisfy both (3.10) and (3.11). Thus, the constructed hypotheses in our proof are substantially more complex, for which reason we restrict attention to $\beta \leq 1$.

Although Condition 3.5 is getting weaker for growing sample size, some densities are permanently excluded from consideration. The following proposition states that the exceptional set of permanently excluded densities is pathological.

PROPOSITION 3.7. *For the rectangular kernel $K_R(\cdot) = \frac{1}{2}\mathbb{1}\{\cdot \in [-1, 1]\}$ and $n \in \mathbb{N}$, let*

$$\mathcal{R}_n^{\text{adm}}(K_R, \beta_*, L^*, \varepsilon) = \{p \in \mathcal{H}_{(-\varepsilon, 1+\varepsilon)}(\beta_*, L^*) : p \text{ satisfies (3.10) and (3.11)}\}$$

and

$$\mathcal{R} = \bigcup_{n \in \mathbb{N}} \mathcal{R}_n^{\text{adm}}(K_R, \beta_*, L^*, \varepsilon).$$

Then, for any $t \in [0, 1]$, for any $h \in \mathcal{G}_\infty$ and for any $\beta \in [\beta_*, 1)$, the set

$$\mathcal{H}_{(t-h, t+h)}(\beta, L^*) \setminus \mathcal{R}_{|(t-h, t+h)}$$

is nowhere dense in $\mathcal{H}_{(t-h, t+h)}(\beta, L^*)$ with respect to $\|\cdot\|_{\beta, (t-h, t+h)}$.

The whole scale of parameters $\beta \in [\beta_*, 1]$ in Proposition 3.7 can be covered by passing over from Hölder classes to Hölder–Zygmund classes in the definition of \mathcal{P}_n ; see Remark A.5 in the Supplementary Material [Patschkowski and Rohde (2019)]. The local adaptivity theory can be likewise developed on the scale of Hölder–Zygmund rather than Hölder classes—here, we restrict attention to Hölder classes because they are commonly considered in the theory of kernel density estimation.

3.2. *Construction of the confidence band.* The new confidence band is based on a kernel density estimator with variable bandwidth incorporating a localized but not the fully pointwise Lepskiĭ (1990) bandwidth selection procedure. A suitable discretization and a locally constant approximation allow to piece the pointwise constructions together in order to obtain a continuum of confidence statements. The complex construction makes the asymptotic calibration of the confidence band to the level α nontrivial. Whereas the analysis of the related globally adaptive procedure of Giné and Nickl (2010) reduces to the limiting distribution of the supremum of a stationary Gaussian process, our locally adaptive approach leads to a highly nonstationary situation, which even depends on the unknown density. An essential component is therefore the identification of a stationary process as a least favorable case by means of Slepian’s comparison inequality, this stationary approximation not involving the unknown density p anymore.

We now describe the procedure. First, the sample is split into two subsamples. For simplicity, we divide the sample into two parts of equal size $\tilde{n} = \lfloor n/2 \rfloor$, leaving possibly out the last observation. Let

$$\chi_1 = \{X_1, \dots, X_{\tilde{n}}\}, \quad \chi_2 = \{X_{\tilde{n}+1}, \dots, X_{2\tilde{n}}\}$$

be the distinct subsamples and denote by $\hat{p}_n^{(1)}(\cdot, h)$ and $\hat{p}_n^{(2)}(\cdot, h)$ the kernel density estimators with bandwidth h based on χ_1 and χ_2 , respectively. $\mathbb{E}_p^{\chi_1}$ and $\mathbb{E}_p^{\chi_2}$ denote the expectations with respect to the product measures

$$\begin{aligned} \mathbb{P}_p^{\chi_1} &= \text{joint distribution of } X_1, \dots, X_{\tilde{n}}, \\ \mathbb{P}_p^{\chi_2} &= \text{joint distribution of } X_{\tilde{n}+1}, \dots, X_{2\tilde{n}}. \end{aligned}$$

Next, the interval $[0, 1]$ is discretized into equally spaced grid points, which serve as evaluation points for the locally adaptive estimator. We discretize by a mesh of width

$$\delta_n = \left[2^{j_{\min}} \left(\frac{\log \tilde{n}}{\tilde{n}} \right)^{-\kappa_1} (\log \tilde{n})^{\frac{2}{\beta_*}} \right]^{-1}$$

with $\kappa_1 \geq 1/(2\beta_*)$ and set

$$(3.12) \quad \mathcal{H}_n = \{k\delta_n : k \in \mathbb{Z}\}.$$

Fix now constants

$$(3.13) \quad c_1 > \frac{2}{\beta_* \log 2} \quad \text{and} \quad \kappa_2 > c_1 \log 2 + 7.$$

With j_{\min} specified in (3.9), consider the set of bandwidth exponents

$$\mathcal{J}_n = \left\{ j \in \mathbb{N} : j_{\min} \leq j \leq j_{\max} = \left\lfloor \log_2 \left(\frac{\tilde{n}}{(\log \tilde{n})^{\kappa_2}} \right) \right\rfloor \right\},$$

and the corresponding dyadic grid of bandwidths

$$(3.14) \quad \mathcal{G}_n = \{2^{-j} : j \in \mathcal{J}_n\}.$$

The bound j_{\max} is standard and particularly guarantees pointwise consistency of the kernel density estimator with every bandwidth within \mathcal{G}_n . The constraint on κ_2 in (3.13) can be relaxed by an inflation of the confidence band’s width by logarithmic factors, as discussed in the simulation study in the Supplementary Material [Patschkowski and Rohde (2019)]. To keep the formulation of the following results as concise as possible, we refrain from this issue at this point. We define the set of admissible bandwidths for $t \in [0, 1]$ as

$$(3.15) \quad \mathcal{A}_n(t) = \left\{ j \in \mathcal{J}_n : \max_{s \in (t-\frac{7}{8} \cdot 2^{-j}, t+\frac{7}{8} \cdot 2^{-j}) \cap \mathcal{H}_n} |\hat{p}_n^{(2)}(s, m) - \hat{p}_n^{(2)}(s, m')| \leq c_2 \sqrt{\frac{\log \tilde{n}}{\tilde{n} 2^{-m}}} \right. \\ \left. \text{for all } m, m' \in \mathcal{J}_n \text{ with } m > m' > j + 2 \right\},$$

with constant $c_2 = c_2(A, \nu, \beta_*, L^*, K, \varepsilon)$ specified in the proof of Proposition 4.1. Furthermore, let

$$(3.16) \quad \hat{j}_n(t) = \min \mathcal{A}_n(t), \quad t \in [0, 1],$$

and $\hat{h}_n(t) = 2^{-\hat{j}_n(t)}$. Note that a slight difference to the classical Lepski procedure is the additional maximum in (3.15), which reflects the idea of adapting localized but not completely pointwise for fixed sample size n . The bandwidth (3.16) is determined for all mesh points $k\delta_n, k \in T_n = \{1, \dots, \delta_n^{-1}\}$ in $[0, 1]$, and set piecewise constant in between. Accordingly, with

$$\hat{h}_{n,1}^{\text{loc}}(k) = 2^{-\hat{j}_n((k-1)\delta_n) - u_n}, \quad \hat{h}_{n,2}^{\text{loc}}(k) = 2^{-\hat{j}_n(k\delta_n) - u_n},$$

where $u_n = c_1 \log \log \tilde{n}$ is some sequence implementing the undersmoothing, the estimators are defined as

$$(3.17) \quad \hat{h}_n^{\text{loc}}(t) = \hat{h}_{n,k}^{\text{loc}} = \min\{\hat{h}_{n,1}^{\text{loc}}(k), \hat{h}_{n,2}^{\text{loc}}(k)\} \quad \text{and} \\ \hat{p}_n^{\text{loc}}(t, h) = \hat{p}_n^{(1)}(k\delta_n, h)$$

for $t \in I_k = [(k - 1)\delta_n, k\delta_n], k \in T_n \setminus \{\delta_n^{-1}\}, I_{\delta_n^{-1}} = [1 - \delta_n, 1]$. The following theorem lays the foundation for the construction of honest and locally adaptive confidence bands.

THEOREM 3.8 (Least favorable case). *For the estimators defined in (3.17) and normalizing sequences*

$$a_n = c_3(-2 \log \delta_n)^{1/2}, \quad b_n = \frac{3}{c_3} \left\{ (-2 \log \delta_n)^{1/2} - \frac{\log(-\log \delta_n) + \log 4\pi}{2(-2 \log \delta_n)^{1/2}} \right\},$$

with $c_3 = \sqrt{2}/TV(K)$, it holds

$$\begin{aligned} \liminf_{n \rightarrow \infty} \inf_{p \in \mathcal{P}_n} \mathbb{P}_p^{\otimes n} \left(a_n \left(\sup_{t \in [0,1]} \sqrt{\tilde{n} \hat{h}_n^{\text{loc}}(t)} |\hat{p}_n^{\text{loc}}(t, \hat{h}_n^{\text{loc}}(t)) - p(t)| - b_n \right) \leq x \right) \\ \geq 2\mathbb{P}(\sqrt{L^*}G \leq x) - 1 \end{aligned}$$

for some standard Gumbel distributed random variable G .

The proof of Theorem 3.8 is based on several completely nonasymptotic approximation techniques. The asymptotic Komlós–Major–Tusnády-approximation technique, used in Giné and Nickl (2010), has been evaded using nonasymptotic Gaussian approximation results recently developed in Chernozhukov, Chetverikov and Kato (2014b). The essential component of the proof of Theorem 3.8 is the application of Slepian’s comparison inequality to reduce considerations from a nonstationary Gaussian process to the least favorable case of a maximum of δ_n^{-1} independent and identical standard normal random variables.

With $q_{1-\alpha/2}$ denoting the $(1 - \alpha/2)$ -quantile of the standard Gumbel distribution, we define the confidence band as the family of piecewise constant random intervals $C_{n,\alpha}^{\text{loc}} = (C_{n,\alpha}^{\text{loc}}(t))_{t \in [0,1]}$ with

$$(3.18) \quad C_{n,\alpha}^{\text{loc}}(t) = \left[\hat{p}_n^{\text{loc}}(t, \hat{h}_n^{\text{loc}}(t)) - \frac{q_n(\alpha)}{\sqrt{\tilde{n} \hat{h}_n^{\text{loc}}(t)}}, \hat{p}_n^{\text{loc}}(t, \hat{h}_n^{\text{loc}}(t)) + \frac{q_n(\alpha)}{\sqrt{\tilde{n} \hat{h}_n^{\text{loc}}(t)}} \right]$$

and

$$(3.19) \quad q_n(\alpha) = \frac{\sqrt{L^*} \cdot q_{1-\alpha/2}}{a_n} + b_n.$$

For fixed $\alpha > 0$, $q_n(\alpha) = O(\sqrt{\log n})$ as n goes to infinity.

COROLLARY 3.9 (Honesty). *The confidence band as defined in (3.18) satisfies*

$$\liminf_{n \rightarrow \infty} \inf_{p \in \mathcal{P}_n} \mathbb{P}_p^{\otimes n} (p(t) \in C_{n,\alpha}^{\text{loc}}(t) \text{ for every } t \in [0, 1]) \geq 1 - \alpha.$$

3.3. *Local Hölder regularity and local adaptivity.* We demonstrate that the new confidence band is locally adaptive in the sense of (1.4). Recall that by Proposition 3.6 the pointwise minimax-rate of convergence over $\mathcal{P}_{n|U_\delta} \cap \mathcal{H}_{\beta^*, U_\delta}(\beta, L^*)$ remains $n^{-\beta/(2\beta+1)}$, and that $|C_{n,\alpha}^{\text{loc}}(t)|$ denotes the length of the interval $C_{n,\alpha}^{\text{loc}}(t)$.

THEOREM 3.10 (Local adaptivity). *For every open interval $U \subset [0, 1]$, and for any $\delta > 0$,*

$$\limsup_{n \rightarrow \infty} \sup_{\substack{p \in \mathcal{P}_n: \\ p|_{U_\delta} \in \mathcal{H}_{U_\delta}(\beta, L^*)}} \mathbb{P}_p^{\otimes n} \left(\sup_{t \in U} |C_{n,\alpha}^{\text{loc}}(t)| \geq \left(\frac{\log n}{n} \right)^{\frac{\beta}{2\beta+1}} (\log n)^\gamma \right) = 0$$

for every $\beta \in [\beta_*, \beta^*]$ and $\gamma = \gamma(c_1)$, where U_δ is the open δ -enlargement of U .

If $p \in \mathcal{H}(\beta, L)$ and $p|_U \in \mathcal{H}_U(\beta', L)$ for some $\beta' > \beta$ and some open interval $U \subset [0, 1]$, then the maximal width over U of our new confidence band is of the stochastic order

$$\mathcal{O}_{\mathbb{P}_p} \left(\left(\frac{\log n}{n} \right)^{\frac{\beta'}{2\beta'+1}} (\log n)^\gamma \right),$$

whereas globally but not locally adaptive confidence bands guarantee a width of stochastic order $\mathcal{O}_{\mathbb{P}_p}(n^{-\beta/(2\beta+1)})$ (up to logarithmic factors) only.

In the remaining part of this section, we develop an even stronger notion of local adaptivity, which is of particular interest for the statistician. Here, the asymptotic statement is not formulated for an arbitrary but fixed interval U only. Indeed, the more observations are available, the more localized and smaller are regions the statistician would like to learn about. Precisely, the goal would be to adapt even to some pointwise or local Hölder regularity, two well-established notions from analysis.

DEFINITION 3.11 [Pointwise Hölder exponent, [Seuret and Véhel \(2002\)](#)]. Let $p : \mathbb{R} \rightarrow \mathbb{R}$ be a function, $\beta > 0$, $\beta \notin \mathbb{N}$, and $t \in \mathbb{R}$. Then $p \in C^\beta(t)$ if and only if there exists a real $R > 0$, a polynomial P with degree less than $\lfloor \beta \rfloor$, and a constant c such that

$$|p(x) - P(x - t)| \leq c|x - t|^\beta$$

for all $x \in (t - R, t + R)$. The pointwise Hölder exponent is denoted by

$$\beta_p(t) = \sup\{\beta : p \in C^\beta(t)\}.$$

DEFINITION 3.12 [Local Hölder exponent, [Seuret and Véhel \(2002\)](#)]. Let $p : \Omega \rightarrow \mathbb{R}$ be a function and $\Omega \subset \mathbb{R}$ an open set. One classically says that $p \in C_l^\beta(\Omega)$, where $0 < \beta < 1$, if there exists a constant c such that

$$|p(x) - p(y)| \leq c|x - y|^\beta$$

for all $x, y \in \Omega$. If $m < \beta < m + 1$ for some $m \in \mathbb{N}$, then $p \in C_l^\beta(\Omega)$ means that there exists a constant c such that

$$|\partial^m p(x) - \partial^m p(y)| \leq c|x - y|^{\beta-m}$$

for all $x, y \in \Omega$. Set now

$$\beta_p(\Omega) = \sup\{\beta : p \in C_l^\beta(\Omega)\}.$$

Finally, the local Hölder exponent in t is defined as

$$\beta_p^{\text{loc}}(t) = \sup\{\beta_p(O_i) : i \in I\},$$

where $(O_i)_{i \in I}$ is a decreasing family of open sets with $\bigcap_{i \in I} O_i = \{t\}$. [By Lemma 2.1 in [Seuret and Véhel \(2002\)](#), this notion is well defined, that is, it does not depend on the particular choice of the decreasing sequence of open sets.]

The next proposition shows that attaining the minimax rates of convergence corresponding to the pointwise or local Hölder exponent (possibly inflated by some logarithmic factor) uniformly over \mathcal{P}_n is an unachievable goal.

PROPOSITION 3.13. *For the rectangular kernel K_R , there exists some constant $M > 0$, such that for any $t \in [0, 1]$, for any $\beta \in [\beta_*, 1]$, for any $0 < \varepsilon < 1$, and for any $k \geq k_0(\beta_*)$, there exists some $x > 0$ and constants $L = L(\beta) > 0$ and $c_4 = c_4(\beta) > 0$ with*

$$\inf_{T_n} \sup_{p \in \mathcal{S}_k(\beta)} \mathbb{P}_p^{\otimes n} (n^{\frac{\beta}{2\beta+1}} |T_n(t) - p(t)| \geq x) > 0 \quad \text{for all } k \geq k_0(\beta_*)$$

for all $n \geq n_0$, with

$$\begin{aligned} \mathcal{S}_k(\beta) &= \mathcal{S}_k(L, \beta, \beta_*, M, K_R, \varepsilon) \\ &= \{p \in \mathcal{P}_k(K_R, \beta_*, L, \varepsilon, M) : \exists r \geq c_4 n^{-\frac{1}{2\beta+1}} \\ &\quad \text{such that } p|_{(t-r, t+r)} \in \mathcal{H}_{(t-r, t+r)}(\infty, L)\} \cap \mathcal{H}_{(-\varepsilon, 1+\varepsilon)}(\beta, L), \end{aligned}$$

where the infimum is running over all estimators T_n based on X_1, \dots, X_n .

Therefore, we introduce an n -dependent statistical notion of local regularity for any point t . Roughly speaking, we intend it to be the maximal β such that the density attains this Hölder exponent within $(t - h_{\beta,n}, t + h_{\beta,n})$, where $h_{\beta,n}$ is of the optimal adaptive bandwidth order $(\log n/n)^{1/(2\beta+1)}$. We realize this idea with $\|\cdot\|_{\beta, \beta_*, U}$ as introduced in Definition 3.4 and used in Condition 3.5.

DEFINITION 3.14 (n -dependent local Hölder exponent). With the classical optimal bandwidth within the class $\mathcal{H}(\beta)$

$$h_{\beta,n} = 2^{-j_{\min}} \cdot \left(\frac{\log \tilde{n}}{\tilde{n}}\right)^{\frac{1}{2\beta+1}},$$

define the class $\mathcal{H}_{\beta^*, n, t}(\beta, L)$ as the set of functions $p : (t - h_{\beta,n}, t + h_{\beta,n}) \rightarrow \mathbb{R}$, such that p admits derivatives up to the order $\lfloor \beta \wedge \beta^* \rfloor$ and $\|p\|_{\beta, \beta^*, (t-h_{\beta,n}, t+h_{\beta,n})} \leq L$, and $\mathcal{H}_{\beta^*, n, t}(\beta)$ the class of functions $p : (t - h_{\beta,n}, t + h_{\beta,n}) \rightarrow \mathbb{R}$ for which the norm $\|p\|_{\beta, \beta^*, (t-h_{\beta,n}, t+h_{\beta,n})}$ is well defined and finite. The n -dependent local Hölder exponent for the function p at point t is defined as

$$(3.20) \quad \beta_{n,p}(t) = \sup\{\beta > 0 : p|_{(t-h_{\beta,n}, t+h_{\beta,n})} \in \mathcal{H}_{\beta^*, n, t}(\beta, L^*)\}.$$

If the supremum is running over the empty set, we set $\beta_{n,p}(t) = 0$.

Finally, the next theorem shows that the confidence band adapts to the n -dependent local Hölder exponent.

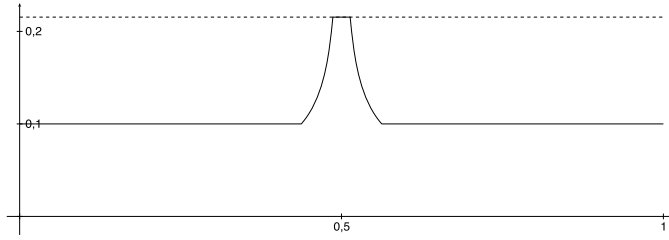


FIG. 1. Stochastic width orders.

THEOREM 3.15 (Strong local adaptivity). *There exists some $\gamma = \gamma(c_1)$, such that*

$$\limsup_{n \rightarrow \infty} \sup_{p \in \mathcal{D}_n} \mathbb{P}_p^{\otimes n} \left(\sup_{t \in [0,1]} |C_{n,\alpha}^{\text{loc}}(t)| \cdot \left(\frac{\log n}{n} \right)^{-\frac{\beta_{n,p}(t)}{2\beta_{n,p}(t)+1}} \geq (\log n)^\gamma \right) = 0.$$

Note that the case $\beta_{n,p}(t) = \infty$ is not excluded in the formulation of Theorem 3.15. That is, if $p|_U$ can be represented as a polynomial of degree strictly less than β^* , the confidence band attains even adaptively the parametric width $n^{-1/2}$, up to logarithmic factors. In particular, the band can be tighter than $n^{-\beta^*/(2\beta^*+1)}$. In general, as long as $\delta \leq \varepsilon$ and $(t - h_{\beta^*,n}, t + h_{\beta^*,n}) \subset U_\delta$,

$$\beta_{n,p}(t) \geq \beta_p(U_\delta) \quad \text{for all } t \in U.$$

EXAMPLE 3.3 (Continued). Figure 1 and Figure 2 illustrate the strong local adaptivity property of our confidence band for the particular example of the triangular density in (3.5) for $n = 100$. As already discussed in Section 3.1, the triangular density satisfies both the global self-similarity Condition 3.1 as well as our Admissibility Condition 3.5. The quantity

$$(3.21) \quad n^{-\frac{\beta_{n,p}(t)}{2\beta_{n,p}(t)+1}}, \quad t \in [0, 1],$$

is (up to logarithmic factors) the stochastic order of the width

$$\frac{2q_n(\alpha)}{\sqrt{\tilde{n}\hat{h}_n^{\text{loc}}(t)}}, \quad t \in [0, 1],$$

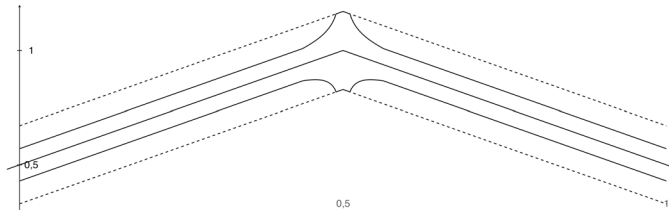


FIG. 2. Bands around the triangular density.

achieved by our new locally adaptive confidence band, which is defined in (3.18) and (3.19), whereas $n^{-1/3}$ is (up to logarithmic factors) the stochastic order of the width of the globally adaptive confidence band (1.3). Figure 1 contrasts our stochastic width order (3.21) (solid line) with $n^{-1/3}$ (dashed line). It shows the substantial benefit of the locally adaptive confidence band outside of a shrinking neighborhood around the maximal point. Our confidence band attains (up to logarithmic factors) the width corresponding to the minimax-optimal rate under Lipschitz smoothness around $t = 1/2$, and the parametric width $n^{-1/2}$ (up to logarithmic factors) outside of the interval $(1/2 - 2^{-j_{\min}}, 1/2 + 2^{-j_{\min}})$.

In Figure 2, we plot the bands

$$([p(t) - n^{-\frac{\beta_{n,p}(t)}{2\beta_{n,p}(t)+1}}, p(t) + n^{-\frac{\beta_{n,p}(t)}{2\beta_{n,p}(t)+1}}])_{t \in [0,1]} \quad (\text{solid lines})$$

and

$$([p(t) - n^{-\frac{1}{3}}, p(t) + n^{-\frac{1}{3}}])_{t \in [0,1]} \quad (\text{dashed lines}).$$

These illustrations are underlined by an extensive simulation study in the Supplementary Material [Patschkowski and Rohde (2019)]. Besides, an algorithm for the computation of the new locally adaptive confidence band is provided.

4. Supplementary notation and results. The following auxiliary results are crucial ingredients for the proofs of Theorem 3.8 and Theorem 3.15.

Recalling the quantity $h_{\beta,n}$ in Definition 3.14, Proposition 4.1 shows that $2^{-\hat{j}_n(\cdot)}$ lies in a band around

$$(4.1) \quad \bar{h}_n(\cdot) = h_{\beta_{n,p}(\cdot),n}$$

uniformly over all admissible densities $p \in \mathcal{P}_n$. Proposition 4.1 furthermore reflects the necessity to undersmooth, which has been already discovered by Bickel and Rosenblatt (1973), leading to a bandwidth deflated by some logarithmic factor. Set now

$$\bar{j}_n(\cdot) = \left\lceil \log_2 \left(\frac{1}{\bar{h}_n(\cdot)} \right) \right\rceil + 1,$$

such that the bandwidth $2^{-\bar{j}_n(\cdot)}$ is an approximation of $\bar{h}_n(\cdot)$ by the next smaller bandwidth on the grid \mathcal{G}_n with

$$\frac{1}{2} \bar{h}_n(\cdot) \leq 2^{-\bar{j}_n(\cdot)} \leq \bar{h}_n(\cdot).$$

PROPOSITION 4.1. *The bandwidth $\hat{j}_n(\cdot)$ defined in (3.16) satisfies*

$$\lim_{n \rightarrow \infty} \sup_{p \in \mathcal{P}_n} \{1 - \mathbb{P}_p^{\chi^2}(\hat{j}_n(k\delta_n) \in [k_n(k\delta_n), \bar{j}_n(k\delta_n) + 1]) \text{ for all } k \in T_n\} = 0,$$

where $k_n(\cdot) = \bar{j}_n(\cdot) - m_n$, and $m_n = \frac{1}{2} c_1 \log \log \tilde{n}$.

LEMMA 4.2. *Let $s, t \in [0, 1]$ be two points with $s < t$, and let $z \in (s, t)$. If*

$$(4.2) \quad |s - t| \leq \frac{1}{8} h_{\beta^*, n}$$

then

$$\frac{1}{3} \bar{h}_n(z) \leq \min\{\bar{h}_n(s), \bar{h}_n(t)\} \leq 3\bar{h}_n(z).$$

LEMMA 4.3. *Recall the definitions of \mathcal{H}_n and \mathcal{G}_n in (3.12) and (3.14), respectively. There exist positive and finite constants $c_5 = c_5(A, \nu, K)$ and $c_6 = c_6(A, \nu, L^*, K)$, and some $\eta_0 = \eta_0(A, \nu, L^*, K) > 0$, such that*

$$\begin{aligned} \sup_{p \in \mathcal{P}_n} \mathbb{P}_p^{X_i} \left(\sup_{s \in \mathcal{H}_n} \max_{h \in \mathcal{G}_n} \sqrt{\frac{\tilde{n}h}{\log \tilde{n}}} |\hat{p}_n^{(i)}(s, h) - \mathbb{E}_p^{X_i} \hat{p}_n^{(i)}(s, h)| > \eta \right) \\ \leq c_5 \tilde{n}^{-c_6 \eta}, \quad i = 1, 2 \end{aligned}$$

for sufficiently large $n \geq n_0(A, \nu, L^*, K)$ and for all $\eta \geq \eta_0$.

The next lemma extends the classical upper bound on the bias for the β^* -capped Hölder ball $\mathcal{H}_{\beta^*, U}(\beta, L)$ as defined in (3.7).

LEMMA 4.4. *Let $t \in \mathbb{R}$ and $g, h > 0$. Any density $p : \mathbb{R} \rightarrow \mathbb{R}$ with*

$$p|_{(t-(g+h), t+(g+h))} \in \mathcal{H}_{\beta^*, (t-(g+h), t+(g+h))}(\beta, L)$$

for some $0 < \beta \leq \infty$ and some $L > 0$ satisfies

$$(4.3) \quad \sup_{s \in (t-g, t+g)} |(K_h * p)(s) - p(s)| \leq b_2 h^\beta$$

for some positive and finite constant $b_2 = b_2(L, K)$.

LEMMA 4.5. *For symmetric kernels K and $\beta = 1$, the bias bound (4.3) continues to hold if the Lipschitz balls are replaced by the corresponding Zygmund balls.*

5. Proofs. Due to space constraints, we restrict to the proofs of Proposition 3.7, Proposition 3.13 and Theorem 3.15. The remaining proofs of Section 3 as well as all proofs of Section 4 are deferred to the Supplementary Material [Patschkowski and Rohde (2019)].

PROOF OF PROPOSITION 3.7. Define

$$\tilde{\mathcal{H}} = \bigcup_{n \in \mathbb{N}} \tilde{\mathcal{H}}_n$$

with

$$\tilde{\mathcal{R}}_n = \left\{ p \in \mathcal{H}_{(-\varepsilon, 1+\varepsilon)}(\beta_*) : \forall t \in [0, 1] \forall h \in \mathcal{G}_\infty \exists \beta \in [\beta_*, \beta^*] \text{ with} \right. \\ \left. p|_{(t-h, t+h)} \in \mathcal{H}_{(t-h, t+h)}(\beta) \text{ and } \|(K_{R,g} * p) - p\|_{(t-(h-g), t+(h-g))} \geq \frac{g^\beta}{\log n} \right. \\ \left. \text{for all } g \in \mathcal{G}_\infty \text{ with } g \leq h/8 \right\}$$

and $K_{R,g}(\cdot) = g^{-1}K_R(\cdot/g)$. Furthermore, let

$$E_n(\beta) = \left\{ p \in \mathcal{H}_{(-\varepsilon, 1+\varepsilon)}(\beta) : \|(K_{R,g} * p) - p\|_{(t-(h-g), t+(h-g))} \geq \frac{2}{\log n} g^\beta \right. \\ \left. \text{for all } t \in [0, 1], \text{ for all } h \in \mathcal{G}_\infty, \text{ and for all } g \in \mathcal{G}_\infty \text{ with } g \leq h/8 \right\}.$$

Note that Lemma A.4 shows that $E_n(\beta)$ is nonempty as soon as

$$\frac{2}{\log n} \leq 1 - \frac{4}{\pi}.$$

Note additionally that $E_n(\beta) \subset \tilde{\mathcal{R}}_n$ for any $\beta \in [\beta_*, \beta^*]$, and

$$\bigcup_{n \in \mathbb{N}} E_n(\beta) \subset \tilde{\mathcal{R}}.$$

With

$$A_n(\beta) = \left\{ \tilde{f} \in \mathcal{H}_{(-1, 2)}(\beta) : \|\tilde{f} - f\|_{\beta, (-\varepsilon, 1+\varepsilon)} < \frac{\|K_R\|_1^{-1}}{\log n} \text{ for some } f \in E_n(\beta) \right\},$$

we get for any $\tilde{f} \in A_n(\beta)$ and a corresponding $f \in E_n(\beta)$ with

$$\|\check{f}\|_{\beta, (-\varepsilon, 1+\varepsilon)} < \|K_R\|_1^{-1} \frac{1}{\log n}$$

and $\check{f} = \tilde{f} - f$, the lower bound

$$\|(K_{R,g} * \tilde{f}) - \tilde{f}\|_{(t-(h-g), t+(h-g))} \\ \geq \|(K_{R,g} * f) - f\|_{(t-(h-g), t+(h-g))} - \|\check{f} - (K_{R,g} * \check{f})\|_{(t-(h-g), t+(h-g))} \\ = \frac{2}{\log n} g^\beta - \sup_{s \in (t-(h-g), t+(h-g))} \left| \int K_R(x) \{ \check{f}(s+gx) - \check{f}(s) \} dx \right| \\ \geq \frac{2}{\log n} g^\beta$$

$$\begin{aligned}
 & -g^\beta \cdot \int |K_R(x)| \sup_{s \in (t-(h-g), t+(h-g))} \sup_{\substack{s' \in (s-g, s+g) \\ s' \neq s}} \frac{|\check{f}(s') - \check{f}(s)|}{|s - s'|^\beta} dx \\
 & \geq \frac{2}{\log n} g^\beta - g^\beta \cdot \|K_R\|_1 \cdot \|\check{f}\|_{\beta, (-\varepsilon, 1+\varepsilon)} \\
 & \geq \frac{1}{\log n} g^\beta
 \end{aligned}$$

for all $g, h \in \mathcal{G}_\infty$ with $g \leq h/8$ and for all $t \in [0, 1]$ and, therefore,

$$A = \bigcup_{n \in \mathbb{N}} A_n(\beta) \subset \check{\mathcal{H}}.$$

Clearly, $A_n(\beta)$ is open in $\mathcal{H}_{(-\varepsilon, 1+\varepsilon)}(\beta)$. Hence, the same holds true for A . Next, we verify that A is dense in $\mathcal{H}_{(-\varepsilon, 1+\varepsilon)}(\beta)$. Let $p \in \mathcal{H}_{(-\varepsilon, 1+\varepsilon)}(\beta)$ and let $\delta > 0$. We now show that there exists some function $\tilde{p}_\delta \in A$ with $\|p - \tilde{p}_\delta\|_{\beta, (-\varepsilon, 1+\varepsilon)} \leq \delta$. For the construction of the function \tilde{p}_δ , set the grid points

$$t_{j,1}(k) = (4j + 1)2^{-k}, \quad t_{j,2}(k) = (4j + 3)2^{-k}$$

for $j \in \{-2^{k-2}, -2^{k-2} + 1, \dots, 2^{k-1} - 1\}$ and $k \geq 2$. The function \tilde{p}_δ shall be defined as the limit of a recursively constructed sequence. The idea is to recursively add appropriately rescaled sine waves at those locations where the bias condition is violated. Let $p_{1,\delta} = p$, and denote

$$J_k = \left\{ j \in \{-2^{k-2}, \dots, 2^{k-1} - 1\} : \right.$$

$$\left. \max_{i=1,2} |(K_{R,2^{-k}} * p_{k-1,\delta})(t_{j,i}(k)) - p_{k-1,\delta}(t_{j,i}(k))| < \frac{1}{2} c_9 \delta \left(1 - \frac{2}{\pi}\right) 2^{-k\beta} \right\}$$

for $k \geq 2$, where

$$c_9 = c_9(\beta) = \left(\frac{3\pi}{2} \cdot \frac{1}{1 - 2^{\beta-1}} + \frac{7}{1 - 2^{-\beta}} \right)^{-1}.$$

For any $k \geq 2$, set

$$p_{k,\delta}(x) = p_{k-1,\delta}(x) + c_9 \delta \sum_{j \in J_k} S_{k,\beta,j}(x)$$

with functions

$$S_{k,\beta,j}(x) = 2^{-k\beta} \sin(2^{k-1} \pi x) \mathbb{1}\{|(4j + 2)2^{-k} - x| \leq 2^{-k+1}\}$$

exemplified in Figure 3. That is,

$$p_{k,\delta}(x) = p(x) + c_9 \delta \sum_{l=2}^k \sum_{j \in J_l} S_{l,\beta,j}(x),$$

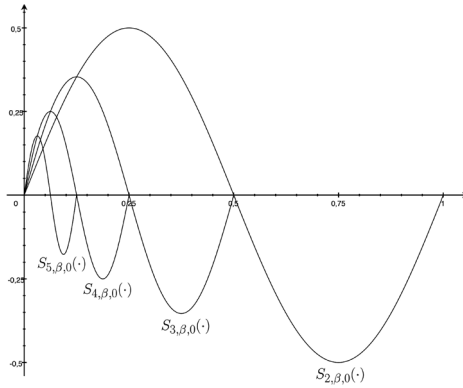


FIG. 3. Functions $S_{k,\beta,0}$ for $k = 2, \dots, 5$ and $\beta = 0.5$.

and we define \tilde{p}_δ as the limit

$$\begin{aligned} \tilde{p}_\delta(x) &= p(x) + c_9\delta \sum_{l=2}^\infty \sum_{j \in J_l} S_{l,\beta,j}(x) \\ &= p_{k,\delta}(x) + c_9\delta \sum_{l=k+1}^\infty \sum_{j \in J_l} S_{l,\beta,j}(x). \end{aligned}$$

The function \tilde{p}_δ is well defined as the series on the right-hand side converges: for fixed $l \in \mathbb{N}$, the indicator functions

$$\mathbb{1}\{|(4j + 2)2^{-k} - x| \leq 2^{-k+1}\}, \quad j \in \{-2^{l-2}, -2^{l-2} + 1, \dots, 2^{l-1} - 1\}$$

have disjoint supports, such that

$$\left\| \sum_{j \in J_l} S_{l,\beta,j} \right\|_{(-\varepsilon, 1+\varepsilon)} \leq 2^{-l\beta}.$$

Hence,

$$\sum_{l=2}^\infty \left\| \sum_{j \in J_l} S_{l,\beta,j} \right\|_{(-\varepsilon, 1+\varepsilon)} \leq \sum_{l=0}^\infty 2^{-l\beta} < \infty,$$

that is the series $\sum_{l=2}^\infty \sum_{j \in J_l} S_{l,\beta,j}$ is normally convergent. In particular, the limit function is continuous.

It remains to verify that $\tilde{p}_\delta \in \bigcup_{n \in \mathbb{N}} E_n(\beta) \subset A$ and also $\|p - \tilde{p}_\delta\|_{\beta, (-\varepsilon, 1+\varepsilon)} \leq \delta$. As concerns the inequality $\|p - \tilde{p}_\delta\|_{\beta, (-\varepsilon, 1+\varepsilon)} \leq \delta$, it remains to show that

$$\left\| \sum_{l=2}^\infty \sum_{j \in J_l} S_{l,\beta,j} \right\|_{\beta, (-\varepsilon, 1+\varepsilon)} \leq \frac{1}{c_9}.$$

For $s, t \in (-\varepsilon, 1 + \varepsilon)$ with $|s - t| \leq 1$, we obtain

$$\begin{aligned}
 & \left| \sum_{l=2}^{\infty} \sum_{j \in J_l} S_{l,\beta,j}(s) - \sum_{l=2}^{\infty} \sum_{j \in J_l} S_{l,\beta,j}(t) \right| \\
 (5.1) \quad & \leq \sum_{l=2}^{\infty} 2^{-l\beta} \left| \sin(2^{l-1}\pi s) \sum_{j \in J_l} \mathbb{1}\{|(4j+2)2^{-l} - s| \leq 2^{-l+1}\} \right. \\
 & \quad \left. - \sin(2^{l-1}\pi t) \sum_{j \in J_l} \mathbb{1}\{|(4j+2)2^{-l} - t| \leq 2^{-l+1}\} \right|.
 \end{aligned}$$

Choose now $k' \in \mathbb{N}$ maximal, such that both

$$(4j+2)2^{-k'} - 2^{-k'+1} \leq s \leq (4j+2)2^{-k'} + 2^{-k'+1}$$

and

$$(4j+2)2^{-k'} - 2^{-k'+1} \leq t \leq (4j+2)2^{-k'} + 2^{-k'+1}$$

for some $j \in \{-2^{k'-2}, \dots, 2^{k'-1} - 1\}$. For $2 \leq l \leq k'$, we have

$$\begin{aligned}
 & \left| \sin(2^{l-1}\pi s) \sum_{j \in J_l} \mathbb{1}\{|(4j+2)2^{-l} - s| \leq 2^{-l+1}\} \right. \\
 (5.2) \quad & \quad \left. - \sin(2^{l-1}\pi t) \sum_{j \in J_l} \mathbb{1}\{|(4j+2)2^{-l} - t| \leq 2^{-l+1}\} \right| \\
 & \leq |\sin(2^{l-1}\pi s) - \sin(2^{l-1}\pi t)| \\
 & \leq \min\{2^{l-1}\pi|s - t|, 2\}
 \end{aligned}$$

by the mean value theorem. For $l \geq k' + 1$,

$$\begin{aligned}
 & \left| \sin(2^{l-1}\pi s) \sum_{j \in J_l} \mathbb{1}\{|(4j+2)2^{-l} - s| \leq 2^{-l+1}\} \right. \\
 & \quad \left. - \sin(2^{l-1}\pi t) \sum_{j \in J_l} \mathbb{1}\{|(4j+2)2^{-l} - t| \leq 2^{-l+1}\} \right| \\
 & \leq \max\{|\sin(2^{l-1}\pi s)|, |\sin(2^{l-1}\pi t)|\}.
 \end{aligned}$$

Furthermore, due to the choice of k' , there exists some $z \in [s, t]$ with

$$\sin(2^{l-1}\pi z) = 0$$

for all $l \geq k' + 1$. Thus, for any $l \geq k' + 1$, by the mean value theorem,

$$\begin{aligned}
 |\sin(2^{l-1}\pi s)| &= |\sin(2^{l-1}\pi s) - \sin(2^{l-1}\pi z)| \\
 &\leq \min\{2^{l-1}\pi|s - z|, 1\} \\
 &\leq \min\{2^{l-1}\pi|s - t|, 1\}.
 \end{aligned}$$

Analogously, we obtain

$$|\sin(2^{l-1}\pi t)| \leq \min\{2^{l-1}\pi |s - t|, 1\}.$$

Consequently, together with inequality (5.1) and (5.2),

$$\left| \sum_{l=2}^{\infty} \sum_{j \in J_l} S_{l,\beta,j}(s) - \sum_{l=2}^{\infty} \sum_{j \in J_l} S_{l,\beta,j}(t) \right| \leq \sum_{l=2}^{\infty} 2^{-l\beta} \min\{2^{l-1}\pi |s - t|, 2\}.$$

Choose now $k \in \mathbb{N} \cup \{0\}$, such that $2^{-(k+1)} < |s - t| \leq 2^{-k}$. If $k \leq 1$,

$$\sum_{l=2}^{\infty} 2^{-l\beta} \min\{2^{l-1}\pi |s - t|, 2\} \leq 2 \frac{2^{-2\beta}}{1 - 2^{-\beta}} \leq \frac{2}{1 - 2^{-\beta}} |s - t|^{\beta}.$$

If $k \geq 2$, we decompose

$$\begin{aligned} \sum_{l=2}^{\infty} 2^{-l\beta} \min\{2^{l-1}\pi |s - t|, 2\} &\leq \frac{\pi}{2} |s - t| \sum_{l=0}^k 2^{l(1-\beta)} + 2 \sum_{l=k+1}^{\infty} 2^{-l\beta} \\ &= \frac{\pi}{2} |s - t| \frac{2^{k(1-\beta)} - 2^{\beta-1}}{1 - 2^{\beta-1}} + 2 \cdot \frac{2^{-(k+1)\beta}}{1 - 2^{-\beta}} \\ &\leq |s - t|^{\beta} \cdot \left(\frac{\pi}{2} \cdot \frac{1}{1 - 2^{\beta-1}} + \frac{2}{1 - 2^{-\beta}} \right). \end{aligned}$$

Since furthermore

$$\left\| \sum_{l=2}^{\infty} \sum_{j \in J_l} S_{l,\beta,j} \right\|_{\sup} \leq \frac{1}{1 - 2^{-\beta}},$$

we have

$$\left\| \sum_{l=2}^{\infty} \sum_{j \in J_l} S_{l,\beta,j} \right\|_{\beta, (-\varepsilon, 1+\varepsilon)} \leq 3 \left(\frac{\pi}{2} \cdot \frac{1}{1 - 2^{\beta-1}} + \frac{2}{1 - 2^{-\beta}} \right) + \frac{1}{1 - 2^{-\beta}} = \frac{1}{c_9}$$

and finally $\|p - \tilde{p}_{\varepsilon}\|_{\beta, (-\varepsilon, 1+\varepsilon)} \leq \delta$. In particular, $\tilde{p}_{\delta} \in \mathcal{H}_{(-\varepsilon, 1+\varepsilon)}(\beta)$.

We now show that the function \tilde{p}_{δ} is contained in $\bigcup_{n \in \mathbb{N}} E_n(\beta) \subset A$. For any bandwidths $g, h \in \mathcal{G}_{\infty}$ with $g \leq h/8$, it holds that $h - g \geq 4g$. Thus, for any $g = 2^{-k}$ with $k \geq 2$ and for any $t \in (-\varepsilon, 1 + \varepsilon)$, there exists some $j = j(t, h, g) \in \{-2^{k-2}, \dots, 2^{k-1} - 1\}$ such that both $t_{j,1}(k)$ and $t_{j,2}(k)$ are contained in $(t - (h - g), t + (h - g))$, which implies

$$\begin{aligned} (5.3) \quad &\sup_{s \in (t - (h - g), t + (h - g))} |(K_{R,g} * \tilde{p}_{\delta})(s) - \tilde{p}_{\delta}(s)| \\ &\geq \max_{i=1,2} |(K_{R,g} * \tilde{p}_{\delta})(t_{j,i}(k)) - \tilde{p}_{\delta}(t_{j,i}(k))|. \end{aligned}$$

By linearity of the convolution and the theorem of dominated convergence,

$$\begin{aligned}
 & (K_{R,g} * \tilde{p}_\delta)(t_{j,i}(k)) - \tilde{p}_\delta(t_{j,i}(k)) \\
 (5.4) \quad &= (K_{R,g} * p_{k,\delta})(t_{j,i}(k)) - p_{k,\delta}(t_{j,i}(k)) \\
 &+ c_9 \delta \sum_{l=k+1}^\infty \sum_{j \in J_l} ((K_{R,g} * S_{l,\beta,j})(t_{j,i}(k)) - S_{l,\beta,j}(t_{j,i}(k))).
 \end{aligned}$$

We analyze the convolution $K_{R,g} * S_{l,\beta,j}$ for $l \geq k + 1$. Here,

$$\sin(2^{l-1}\pi t_{j,1}(k)) = \sin(2^{l-k-1}\pi(4j + 1)) = 0$$

and

$$\sin(2^{l-1}\pi t_{j,2}(k)) = \sin(2^{l-k-1}\pi(4j + 3)) = 0.$$

Hence,

$$\sum_{j \in J_l} S_{l,\beta,j}(t_{j,i}(k)) = 0, \quad i = 1, 2$$

for any $l \geq k + 1$. Furthermore,

$$\begin{aligned}
 (K_{R,g} * S_{l,\beta,j})(t_{j,i}(k)) &= \frac{1}{2g} \int_{-g}^g S_{l,\beta,j}(t_{j,i}(k) - x) \, dx \\
 &= \frac{1}{2g} \int_{t_{j,i}(k)-g}^{t_{j,i}(k)+g} S_{l,\beta,j}(x) \, dx, \quad i = 1, 2.
 \end{aligned}$$

Due to the identities

$$(4j + 2)2^{-k} - 2^{-k+1} = t_{j,1}(k) - g,$$

$$(4j + 2)2^{-k} + 2^{-k+1} = t_{j,2}(k) + g,$$

we have either

$$[(4j + 2)2^{-l} - 2^{-l+1}, (4j + 2)2^{-l} + 2^{-l+1}] \subset [t_{j,1}(k) - g, t_{j,2}(k) + g]$$

or

$$[(4j + 2)2^{-l} - 2^{-l+1}, (4j + 2)2^{-l} + 2^{-l+1}] \cap [t_{j,1}(k) - g, t_{j,2}(k) + g] = \emptyset$$

for any $l \geq k + 1$. Therefore, for $i = 1, 2$,

$$\begin{aligned}
 & \sum_{j \in J_l} (K_{R,g} * S_{l,\beta,j})(t_{j,i}(k)) \\
 &= \sum_{j \in J_l} \frac{1}{2g} \int_{t_{j,i}(k)-g}^{t_{j,i}(k)+g} 2^{-l\beta} \sin(2^{l-1}\pi x) \mathbb{1}\{|(4j + 2)2^{-l} - x| \leq 2^{-l+1}\} \, dx \\
 &= 0
 \end{aligned}$$

such that equation (5.4) then simplifies to

$$\begin{aligned} & (K_{R,g} * \tilde{p}_\delta)(t_{j,i}(k)) - \tilde{p}_\delta(t_{j,i}(k)) \\ &= (K_{R,g} * p_{k,\delta})(t_{j,i}(k)) - p_{k,\delta}(t_{j,i}(k)), \quad i = 1, 2. \end{aligned}$$

Together with (5.3), we obtain

$$\begin{aligned} & \sup_{s \in (t-(h-g), t+(h-g))} |(K_{R,g} * \tilde{p}_\delta)(s) - \tilde{p}_\delta(s)| \\ & \geq \max_{i=1,2} |(K_{R,g} * p_{k,\delta})(t_{j,i}(k)) - p_{k,\delta}(t_{j,i}(k))| \end{aligned}$$

for some $j \in \{-2^{k-2}, -2^{k-2} + 1, \dots, 2^{k-2} - 1\}$. If $j \notin J_k$, then

$$\begin{aligned} & \max_{i=1,2} |(K_{R,g} * p_{k,\delta})(t_{j,i}(k)) - p_{k,\delta}(t_{j,i}(k))| \\ &= \max_{i=1,2} |(K_{R,g} * p_{k-1,\delta})(t_{j,i}(k)) - p_{k-1,\delta}(t_{j,i}(k))| \\ & \geq \frac{1}{2} c_9 \delta \left(1 - \frac{2}{\pi}\right) g^\beta. \end{aligned}$$

If $j \in J_k$, then

$$\begin{aligned} & \max_{i=1,2} |(K_{R,g} * p_{k,\delta})(t_{j,i}(k)) - p_{k,\delta}(t_{j,i}(k))| \\ & \geq c_9 \delta \max_{i=1,2} |(K_{R,g} * S_{k,\beta,j})(t_{j,i}(k)) - S_{k,\beta,j}(t_{j,i}(k))| \\ & \quad - \max_{i=1,2} |(K_{R,g} * p_{k-1,\delta})(t_{j,i}(k)) - p_{k-1,\delta}(t_{j,i}(k))| \\ & \geq c_9 \delta \max_{i=1,2} |(K_{R,g} * S_{k,\beta,j})(t_{j,i}(k)) - S_{k,\beta,j}(t_{j,i}(k))| - \frac{1}{2} c_9 \delta \left(1 - \frac{2}{\pi}\right) g^\beta. \end{aligned}$$

Similarly as above we obtain

$$\begin{aligned} & (K_{R,g} * S_{k,\beta,j})(t_{j,1}(k)) - S_{k,\beta,j}(t_{j,1}(k)) \\ &= \frac{1}{2g} \int_{t_{j,1}(k)-g}^{t_{j,1}(k)+g} 2^{-k\beta} \sin(2^{k-1}\pi x) dx - 2^{-k\beta} \\ &= \frac{1}{2g} 2^{-k\beta} \int_0^{2^{-k+1}} \sin(2^{k-1}\pi x) dx - 2^{-k\beta} \\ &= g^\beta \left(\frac{2}{\pi} - 1\right) \end{aligned}$$

as well as

$$(K_{R,g} * S_{k,\beta,j})(t_{j,2}(k)) - S_{k,\beta,j}(t_{j,2}(k)) = g^\beta \left(1 - \frac{2}{\pi}\right),$$

such that

$$\max_{i=1,2} |(K_{R,g} * p_{k,\delta})(t_{j,i}(k)) - p_{k,\delta}(t_{j,i}(k))| \geq \frac{1}{2} c_9 \delta \left(1 - \frac{2}{\pi}\right) g^\beta.$$

Combining the two cases finally gives

$$\sup_{s \in (t-(h-g), t+(h-g))} |(K_{R,g} * \tilde{p}_\delta)(s) - \tilde{p}_\delta(s)| \geq \frac{1}{2} c_9 \delta \left(1 - \frac{2}{\pi}\right) g^\beta.$$

In particular, $\tilde{p}_\delta \in E_n(\beta)$ for sufficiently large $n \geq n_0(\beta, \delta)$, and thus $\tilde{p}_\delta \in A$.

Since A is open and dense in the class $\mathcal{H}_{(-\varepsilon, 1+\varepsilon)}(\beta)$ and $A \subset \tilde{\mathcal{H}}$, the complement $\mathcal{H}_{(-\varepsilon, 1+\varepsilon)}(\beta) \setminus \tilde{\mathcal{H}}$ is nowhere dense in $\mathcal{H}_{(-\varepsilon, 1+\varepsilon)}(\beta)$. Thus, because of

$$\mathcal{H}_{(-\varepsilon, 1+\varepsilon)}(\beta)|_{(t-h, t+h)} = \mathcal{H}_{(t-h, t+h)}(\beta),$$

and the fact that for any $x \in \mathcal{H}_{(-\varepsilon, 1+\varepsilon)}(\beta)$ and any $z' \in \mathcal{H}_{(t-h, t+h)}(\beta)$ with

$$\|x|_{(t-h, t+h)} - z'\|_{\beta, (t-h, t+h)} < \delta$$

there exists an extension $z \in \mathcal{H}_{(-\varepsilon, 1+\varepsilon)}(\beta)$ of z' with

$$\|x - z\|_{\beta, (-\varepsilon, 1+\varepsilon)} < \delta,$$

the set $\mathcal{H}_{(t-h, t+h)}(\beta) \setminus \tilde{\mathcal{H}}|_{(t-h, t+h)}$ is nowhere dense in $\mathcal{H}_{(t-h, t+h)}(\beta)$. \square

PROOF OF PROPOSITION 3.13. The proof is based on a reduction of the supremum over the class to a maximum over two distinct hypotheses.

Part 1. For $\beta \in [\beta_*, 1)$, the construction of the hypotheses is based on the Weierstraß function as defined in (A.3) in the Supplementary Material. As in the proof of Proposition 3.6 (see the Supplementary Material [Patschkowski and Rohde (2019)]), consider the function $p_0 : \mathbb{R} \rightarrow \mathbb{R}$ with

$$p_0(x) = \begin{cases} 0 & \text{if } |x - t| \geq \frac{10}{3}, \\ \frac{1}{4} + \frac{3}{16}(x - t + 2) & \text{if } -\frac{10}{3} < x - t < -2, \\ \frac{1}{6} + \frac{1 - 2^{-\beta}}{12} W_\beta(x - t) & \text{if } |x - t| \leq 2, \\ \frac{1}{4} - \frac{3}{16}(x - t - 2) & \text{if } 2 < x - t < \frac{10}{3}, \end{cases}$$

and the functions $p_{1,n}, p_{2,n} : \mathbb{R} \rightarrow \mathbb{R}$ with

$$p_{1,n}(x) = p_0(x) + q_{t+\frac{9}{4},n}(x; g\beta,n) - q_{t,n}(x; g\beta,n), \quad x \in \mathbb{R},$$

$$p_{2,n}(x) = p_0(x) + q_{t+\frac{9}{4},n}(x; c_{18} \cdot g\beta,n) - q_{t,n}(x; c_{18} \cdot g\beta,n), \quad x \in \mathbb{R}$$

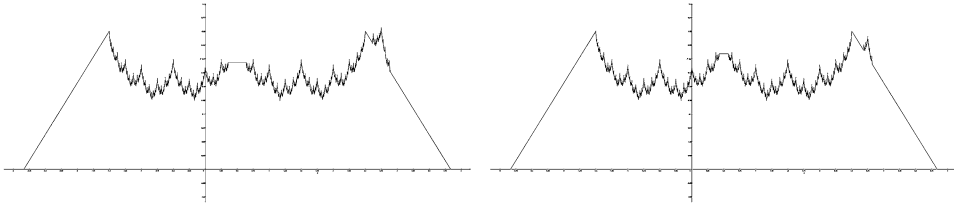


FIG. 4. Functions $p_{1,n}$ and $p_{2,n}$ for $t = 0.5$, $\beta = 0.5$ and $n = 50$.

for $g_{\beta,n} = \frac{1}{4}n^{-1/(2\beta+1)}$ and $c_{18} = c_{18}(\beta) = (2L_W(\beta))^{-1/\beta}$, where

$$q_{a,n}(x; g) = \begin{cases} 0 & \text{if } |x - a| > g, \\ \frac{1 - 2^{-\beta}}{12} (W_\beta(x - a) - W_\beta(g)) & \text{if } |x - a| \leq g, \end{cases}$$

for $a \in \mathbb{R}$ and $g > 0$.

Figure 4 visualizes the densities $p_{1,n}$ and $p_{2,n}$ exemplarily for $t = 0.5$, $\beta = 0.5$ and $n = 50$. Following the lines of the proof of Proposition 3.6, both $p_{1,n}$ and $p_{2,n}$ are contained in the class $\mathcal{P}_k(L, \beta_*, M, K_R, \varepsilon)$ for sufficiently large $k \geq k_0(\beta_*)$. Moreover, both $p_{1,n}$ and $p_{2,n}$ are constant on $(t - c_{18} \cdot g_{\beta,n}, t + c_{18} \cdot g_{\beta,n})$, so that

$$p_{1,n}|_{(t-c_{18} \cdot g_{\beta,n}, t+c_{18} \cdot g_{\beta,n})}, p_{2,n}|_{(t-c_{18} \cdot g_{\beta,n}, t+c_{18} \cdot g_{\beta,n})} \in \mathcal{H}_{(t-c_{18} \cdot g_{\beta,n}, t+c_{18} \cdot g_{\beta,n})}(\infty, L)$$

for some constant $L = L(\beta)$. Using Lemma A.4 and

$$|p_0(t) - p_{1,n}(t)| \geq \frac{1 - 2^{-\beta_*}}{12} g_{\beta,n}^\beta$$

(see the Supplementary Material [Patschkowski and Rohde (2019)]), the absolute distance of the two hypotheses in t is at least

$$\begin{aligned} |p_{1,n}(t) - p_{2,n}(t)| &= |q_{t,n}(t; g_{\beta,n}) - q_{t,n}(t; c_{18} \cdot g_{\beta,n})| \\ &= \frac{1 - 2^{-\beta}}{12} |W_\beta(g_{\beta,n}) - W_\beta(c_{18} \cdot g_{\beta,n})| \\ &\geq \frac{1 - 2^{-\beta_*}}{12} (|W_\beta(g_{\beta,n}) - W_\beta(0)| - |W_\beta(c_{18} \cdot g_{\beta,n}) - W_\beta(0)|) \\ &\geq \frac{1 - 2^{-\beta_*}}{12} (g_{\beta,n}^\beta - L_W(\beta)(c_{18} \cdot g_{\beta,n})^\beta) \\ &\geq 2c_{19}g_{\beta,n}^\beta, \end{aligned}$$

where

$$c_{19} = c_{19}(\beta_*) = \frac{1 - 2^{-\beta_*}}{48}.$$

Since furthermore

$$\int (p_{2,n}(x) - p_{1,n}(x)) \, dx = 0,$$

and $\log(1 + x) \leq x$ for $x > -1$, the Kullback–Leibler divergence between the associated product probability measures $\mathbb{P}_{1,n}^{\otimes n}$ and $\mathbb{P}_{2,n}^{\otimes n}$ is bounded from above by

$$\begin{aligned} K(\mathbb{P}_{2,n}^{\otimes n}, \mathbb{P}_{1,n}^{\otimes n}) &\leq n \int \frac{(p_{2,n}(x) - p_{1,n}(x))^2}{p_{1,n}(x)} \, dx \\ &\leq 12n \int (p_{2,n}(x) - p_{1,n}(x))^2 \, dx \\ &= 24n \int (q_{0,n}(x; g_{\beta,n}) - q_{0,n}(x, c_{18} \cdot g_{\beta,n}))^2 \, dx \\ &= 24n \left(\frac{1 - 2^{-\beta}}{12} \right)^2 \left(2 \int_{c_{18} \cdot g_{\beta,n}}^{g_{\beta,n}} (W_{\beta}(x) - W_{\beta}(g_{\beta,n}))^2 \, dx \right. \\ &\quad \left. + \int_{-c_{18} \cdot g_{\beta,n}}^{c_{18} \cdot g_{\beta,n}} (W_{\beta}(c_{18} \cdot g_{\beta,n}) - W_{\beta}(g_{\beta,n}))^2 \, dx \right) \\ &\leq 24n L_W(\beta)^2 \left(\frac{1 - 2^{-\beta}}{12} \right)^2 \left(2 \int_{c_{18} \cdot g_{\beta,n}}^{g_{\beta,n}} (g_{\beta,n} - x)^{2\beta} \, dx \right. \\ &\quad \left. + 2(1 - c_{18})^2 c_{18} g_{\beta,n}^{2\beta+1} \right) \\ &= c_{20} \end{aligned}$$

with

$$\begin{aligned} c_{20} &= c_{20}(\beta) \\ &= 48 L_W(\beta)^2 4^{-(2\beta+1)} \left(\frac{1 - 2^{-\beta}}{12} \right)^2 \left(\frac{(1 - c_{18})^{2\beta+1}}{2\beta + 1} + (1 - c_{18})^2 c_{18} \right), \end{aligned}$$

where we used Lemma A.4 in the last inequality. Theorem 2.2 in Tsybakov (2009) then yields

$$\begin{aligned} \inf_{T_n} \sup_{p \in \mathcal{S}_k(\beta)} \mathbb{P}_p^{\otimes n} \left(n^{\frac{\beta}{2\beta+1}} |T_n(t) - p(t)| \geq c_{19} \right) \\ \geq \max \left\{ \frac{1}{4} \exp(-c_{20}), \frac{1 - \sqrt{c_{20}/2}}{2} \right\} > 0. \end{aligned}$$

Part 2. For $\beta = 1$, consider the function $p_0 : \mathbb{R} \rightarrow \mathbb{R}$ with

$$p_0(x) = \begin{cases} 0 & \text{if } |x - t| > 4, \\ \frac{1}{4} - \frac{1}{16}|x - t| & \text{if } |x - t| \leq 4, \end{cases}$$

and the functions $p_{1,n}, p_{2,n} : \mathbb{R} \rightarrow \mathbb{R}$ with

$$p_{1,n}(x) = p_0(x) + q_{t+\frac{9}{4},n}(x; g_{1,n}) - q_{t,n}(x; g_{1,n}),$$

$$p_{2,n}(x) = p_0(x) + q_{t+\frac{9}{4},n}(x; g_{1,n}/2) - q_{t,n}(x; g_{1,n}/2)$$

for $g_{1,n} = \frac{1}{4}n^{-1/3}$, where

$$q_{a,n}(x; g) = \begin{cases} 0 & \text{if } |x - a| > g, \\ \frac{1}{16}(g - |x - a|) & \text{if } |x - a| \leq g, \end{cases}$$

for $a \in \mathbb{R}$ and $g > 0$. Following the lines of the proof of Proposition 3.6, both $p_{1,n}$ and $p_{2,n}$ are contained in the class \mathcal{P}_k for sufficiently large $k \geq k_0(\beta_*)$. Moreover, both $p_{1,n}$ and $p_{2,n}$ are constant on $(t - g_{1,n}/2, t + g_{1,n}/2)$, so that

$$P_{1,n|(t-g_{1,n}/2, t+g_{1,n}/2)}, P_{2,n|(t-g_{1,n}/2, t+g_{1,n}/2)} \in \mathcal{H}_{(t-g_{1,n}/2, t+g_{1,n}/2)}(\infty, 1/4).$$

The absolute distance of $p_{1,n}$ and $p_{2,n}$ in t is given by

$$|p_{1,n}(t) - p_{2,n}(t)| = \frac{1}{32}g_{1,n},$$

whereas the Kullback–Leibler divergence between the associated product probability measures $\mathbb{P}_{1,n}^{\otimes n}$ and $\mathbb{P}_{2,n}^{\otimes n}$ is upper bounded by

$$\begin{aligned} K(\mathbb{P}_{2,n}^{\otimes n}, \mathbb{P}_{1,n}^{\otimes n}) &\leq n \int \frac{(p_{2,n}(x) - p_{1,n}(x))^2}{p_{1,n}(x)} dx \\ &\leq 16n \int (p_{2,n}(x) - p_{1,n}(x))^2 dx \\ &= 32n \int (q_{0,n}(x; g_{1,n}) - q_{0,n}(x, g_{1,n}/2))^2 dx \\ &= 32n \left(2 \int_{g_{1,n}/2}^{g_{1,n}} \left(\frac{1}{16}(g_{1,n} - x) \right)^2 dx + \int_{-g_{1,n}/2}^{g_{1,n}/2} \left(\frac{g_{1,n}}{32} \right)^2 dx \right) \\ &= \frac{2}{3 \cdot 32^2} + \frac{1}{32}. \end{aligned}$$

Together with Theorem 2.2 in Tsybakov (2009) the result follows. \square

PROOF OF THEOREM 3.15. Recall the notation of Section 3.2, in particular the definitions of $\hat{h}_n^{\text{loc}}(t)$ in (3.17), of $q_n(\alpha)$ in (3.19), of $\beta_{n,p}(t)$ in (3.20), and of $\bar{h}_n(t)$ in (4.1). Furthermore, set $\tilde{\gamma} = \tilde{\gamma}(c_1) = \frac{1}{2}(c_1 \log 2 - 1)$. To show that the confidence band is adaptive, note that according to Proposition 4.1 and Lemma 4.2 for any $\delta > 0$ there exists some $n_0(\delta)$, such that

$$\sup_{p \in \mathcal{P}_n} \mathbb{P}_p^{\chi^2} \left(\sup_{t \in [0,1]} |C_{n,\alpha}^{\text{loc}}(t)| \cdot \left(\frac{\log \tilde{n}}{\tilde{n}} \right)^{\frac{-\beta_{n,p}(t)}{2\beta_{n,p}(t)+1}} \geq \sqrt{6} \cdot 2^{1-\frac{j_{\min}}{2}} q_n(\alpha) (\log \tilde{n})^{\tilde{\gamma}} \right)$$

$$\begin{aligned}
 &= \sup_{p \in \mathcal{P}_n} \mathbb{P}_p^{\chi^2} \left(\sup_{t \in [0,1]} \frac{\bar{h}_n(t)}{\hat{h}_n^{\text{loc}}(t)} \cdot 2^{-u_n} \geq 6 \right) \\
 &= \sup_{p \in \mathcal{P}_n} \mathbb{P}_p^{\chi^2} \left(\max_{k \in T_n} \sup_{t \in I_k} \frac{\bar{h}_n(t)}{\min\{2^{-\hat{j}_n((k-1)\delta_n)}, 2^{-\hat{j}_n(k\delta_n)}\}} \geq 6 \right) \\
 &\leq \sup_{p \in \mathcal{P}_n} \mathbb{P}_p^{\chi^2} \left(\max_{k \in T_n} \frac{\min\{\bar{h}_n((k-1)\delta_n), \bar{h}_n(k\delta_n)\}}{\min\{2^{-\hat{j}_n((k-1)\delta_n)}, 2^{-\hat{j}_n(k\delta_n)}\}} \geq 2 \right) \\
 &\leq \sup_{p \in \mathcal{P}_n} \mathbb{P}_p^{\chi^2} \left(\exists k \in T_n : \frac{\min\{2^{-\bar{j}_n((k-1)\delta_n)}, 2^{-\bar{j}_n(k\delta_n)}\}}{\min\{2^{-\hat{j}_n((k-1)\delta_n)}, 2^{-\hat{j}_n(k\delta_n)}\}} \geq 1 \right) \\
 &= \sup_{p \in \mathcal{P}_n} \left\{ 1 - \mathbb{P}_p^{\chi^2} \left(\forall k \in T_n : \frac{\min\{2^{-\bar{j}_n((k-1)\delta_n)}, 2^{-\bar{j}_n(k\delta_n)}\}}{\min\{2^{-\hat{j}_n((k-1)\delta_n)}, 2^{-\hat{j}_n(k\delta_n)}\}} < 1 \right) \right\} \\
 &\leq \sup_{p \in \mathcal{P}_n} \left\{ 1 - \mathbb{P}_p^{\chi^2} (\hat{j}_n(k\delta_n) < \bar{j}_n(k\delta_n) \text{ for all } k \in T_n) \right\} \\
 &\leq \delta
 \end{aligned}$$

for all $n \geq n_0(\delta)$. \square

SUPPLEMENTARY MATERIAL

Supplement to “Locally adaptive confidence bands” (DOI: [10.1214/18-AOS1690SUPP](https://doi.org/10.1214/18-AOS1690SUPP); .pdf). Supplement A is organized as follows. Section A.1 develops connections between the Weierstraß function and the Admissibility Condition 3.5. Further notation and auxiliary results from empirical process theory are provided in Section A.2, whereas Section A.3 provides a simulation study together with an algorithm for the calculation of the locally adaptive confidence band. Section A.4 presents the remaining proofs of the results of Section 3. We proceed with the proofs of the results of Section 4 in Section A.5. Auxiliary results are stated and proved in Section A.6.

REFERENCES

BARAUD, Y. (2004). Confidence balls in Gaussian regression. *Ann. Statist.* **32** 528–551. [MR2060168](#)
 BICKEL, P. J. and ROSENBLATT, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1** 1071–1095. [MR0348906](#)
 BULL, A. D. (2012). Honest adaptive confidence bands and self-similar functions. *Electron. J. Stat.* **6** 1490–1516. [MR2988456](#)
 BULL, A. D. and NICKL, R. (2013). Adaptive confidence sets in L^2 . *Probab. Theory Related Fields* **156** 889–919. [MR3078289](#)
 CAI, T. T. and LOW, M. G. (2004). An adaptation theory for nonparametric confidence intervals. *Ann. Statist.* **32** 1805–1840. [MR2102494](#)
 CAI, T. T. and LOW, M. G. (2006). Adaptive confidence balls. *Ann. Statist.* **34** 202–228. [MR2275240](#)

- CARPENTIER, A. (2013). Honest and adaptive confidence sets in L_p . *Electron. J. Stat.* **7** 2875–2923. [MR3148371](#)
- CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014a). Anti-concentration and honest, adaptive confidence bands. *Ann. Statist.* **42** 1787–1818. [MR3262468](#)
- CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014b). Gaussian approximation of suprema of empirical processes. *Ann. Statist.* **42** 1564–1597. [MR3262461](#)
- DAVIES, P. L., KOVAC, A. and MEISE, M. (2009). Nonparametric regression, confidence regions and regularization. *Ann. Statist.* **37** 2597–2625. [MR2541440](#)
- DÜMBGEN, L. (1998). New goodness-of-fit tests and their application to nonparametric confidence sets. *Ann. Statist.* **26** 288–314. [MR1611768](#)
- DÜMBGEN, L. (2003). Optimal confidence bands for shape-restricted curves. *Bernoulli* **9** 423–449. [MR1997491](#)
- GENOVESE, C. R. and WASSERMAN, L. (2005). Confidence sets for nonparametric wavelet regression. *Ann. Statist.* **33** 698–729. [MR2163157](#)
- GENOVESE, C. and WASSERMAN, L. (2008). Adaptive confidence bands. *Ann. Statist.* **36** 875–905. [MR2396818](#)
- GINÉ, E. and NICKL, R. (2010). Confidence bands in density estimation. *Ann. Statist.* **38** 1122–1170. [MR2604707](#)
- HENGARTNER, N. W. and STARK, P. B. (1995). Finite-sample confidence envelopes for shape-restricted densities. *Ann. Statist.* **23** 525–550. [MR1332580](#)
- HOFFMANN, M. and NICKL, R. (2011). On adaptive inference and confidence bands. *Ann. Statist.* **39** 2383–2409. [MR2906872](#)
- JUDITSKY, A. and LAMBERT-LACROIX, S. (2003). Nonparametric confidence set estimation. *Math. Methods Statist.* **12** 410–428. [MR2054156](#)
- KERKYACHARIAN, G., NICKL, R. and PICARD, D. (2012). Concentration inequalities and confidence bands for needlet density estimators on compact homogeneous manifolds. *Probab. Theory Related Fields* **153** 363–404. [MR2925578](#)
- KUEH, A. (2012). Locally adaptive density estimation on the unit sphere using needlets. *Constr. Approx.* **36** 433–458. [MR2996439](#)
- LEPSKIĪ, O. V. (1990). A problem of adaptive estimation in Gaussian white noise. *Teor. Veroyatn. Primen.* **35** 459–470. [MR1091202](#)
- LOW, M. G. (1997). On nonparametric confidence intervals. *Ann. Statist.* **25** 2547–2554. [MR1604412](#)
- NICKL, R. and SZABÓ, B. (2016). A sharp adaptive confidence ball for self-similar functions. *Stochastic Process. Appl.* **126** 3913–3934. [MR3565485](#)
- PATSCHKOWSKI, T. and ROHDE, A. (2019). Supplement to “Locally adaptive confidence bands.” DOI:10.1214/18-AOS1690SUPP.
- PICARD, D. and TRIBOULEY, K. (2000). Adaptive confidence interval for pointwise curve estimation. *Ann. Statist.* **28** 298–335. [MR1762913](#)
- ROBINS, J. and VAN DER VAART, A. (2006). Adaptive nonparametric confidence sets. *Ann. Statist.* **34** 229–253. [MR2275241](#)
- SEURET, S. and VÉHEL, J. L. (2002). The local Hölder function of a continuous function. *Appl. Comput. Harmon. Anal.* **13** 263–276. [MR1942744](#)
- TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer, New York. Revised and extended from the 2004 French original, translated by Vladimir Zaiats. [MR2724359](#)

FAKULTÄT FÜR MATHEMATIK
 RUHR-UNIVERSITÄT BOCHUM
 44780 BOCHUM
 GERMANY
 E-MAIL: tim.patschkowski@ruhr-uni-bochum.de

MATHEMATISCHES INSTITUT
 ALBERT-LUDWIGS-UNIVERSITÄT FREIBURG
 ERNST-ZERMELO-STRASSE 1
 79104 FREIBURG IM BREISGAU
 GERMANY
 E-MAIL: angelika.rohde@stochastik.uni-freiburg.de