The Annals of Statistics
2008, Vol. 36, No. 2, 875–905
DOI: 10.1214/07-AOS500

ADAPTIVE CONFIDENCE BANDS

BY CHRISTOPHER GENOVESE AND LARRY WASSERMAN

Carnegie Mellon University

We show that there do not exist adaptive confidence bands for curve estimation except under very restrictive assumptions. We propose instead to construct adaptive bands that cover a surrogate function $f^\star$ which is close to, but simpler than, f. The surrogate captures the significant features in f. We establish lower bounds on the width for any confidence band for $f^\star$ and construct a procedure that comes within a small constant factor of attaining the lower bound for finite-samples.

1. Introduction.

1.1. *Motivation.* Let $(x_1, Y_1), \dots, (x_n, Y_n)$ be observations from the nonparametric regression model

$$Y_i = f(x_i) + \sigma\varepsilon_i, \tag{1}$$

where $\varepsilon_i \sim N(0, 1)$, $x_i \in (0, 1)$ and f is assumed to lie in some infinite-dimensional class of functions $\mathcal{H}$. We are interested in constructing confidence bands (L, U) for f. Ideally these bands should satisfy

$$\mathbb{P}_f\{L \le f \le U\} = 1 - \alpha \qquad \text{for all } f \in \mathcal{H}, \tag{2}$$

where $L \le f \le U$ means that $L(x) \le f(x) \le U(x)$ for all $x \in \mathcal{X}$, where $\mathcal{X}$ is some subset of $(0, 1)$ such as $\mathcal{X} = \{x\}$, $\mathcal{X} = \{x_1, \dots, x_n\}$ or $\mathcal{X} = (0, 1)$. Throughout this paper, we take $\mathcal{X} = \{x_1, \dots, x_n\}$ but this particular choice is not crucial in what follows.

Attaining (2) is difficult and hence it is common to settle for pointwise asymptotic coverage:

$$\liminf_{n\to\infty} \mathbb{P}_f\{L \le f \le U\} \ge 1 - \alpha \qquad \text{for all } f \in \mathcal{H}. \tag{3}$$

"Pointwise" refers to the fact that the asymptotic limit is taken for each fixed f rather than uniformly over $f \in \mathcal{H}$. Papers on pointwise asymptotic methods include Claeskens and Van Keilegom (2003), Eubank and Speckman (1993), Härdle and Marron (1991), Hall and Titterington (1988), Härdle and Bowman (1988), Neumann and Polzehl (1998) and Xia (1998).

Received January 2007; revised April 2007.
AMS 2000 subject classifications. 62G15, 62G08, 62G20.
Key words and phrases. Nonparametric confidence bands, nonparametric regression, adaptive confidence sets, lower bounds.

Achieving even pointwise asymptotic coverage is nontrivial due to the presence of bias. If $\hat{f}(x)$ is an estimator with mean $\overline{f}(x)$ and standard deviation $s(x)$ then

$$\frac{\hat{f}(x) - f(x)}{s(x)} = \frac{\hat{f}(x) - \overline{f}(x)}{s(x)} + \frac{\text{bias}(x)}{\sqrt{\text{variance}(x)}}.$$

The first term typically satisfies a central limit theorem but the second term does not vanish even asymptotically if the bias and variance are balanced. For discussions on this point, see the papers referenced above as well as Ruppert, Wand and Carroll (2003) and Sun and Loader (1994).

Pointwise asymptotic bands are not uniform, that is, they do not control

$$\inf_{f \in \mathcal{H}} \mathbb{P}_f\{L \le f \le U\}. \tag{4}$$

The sample size $n(f)$ required for the true coverage to approximate the nominal coverage, depends on the unknown function f.

The aim of this paper is to attain uniform coverage over $\mathcal{H}$. We say that $B = (L, U)$ has *uniform coverage* if

$$\inf_{f \in \mathcal{H}} \mathbb{P}_f\{L \le f \le U\} \ge 1 - \alpha. \tag{5}$$

Starting in Section 3, we will insist on coverage over $\mathcal{H} = \{\text{all functions}\}$.

The bound in (5) can be achieved trivially using Bonferroni bands. Set $\ell_i = Y_i - c_n\sigma$ and $u_i = Y_i + c_n\sigma$, where $c_n = \Phi^{-1}(1 - \alpha/2n)$ and Φ is the standard Normal c.d.f. Yet this band is unsatisfactory for several reasons:

1. The width of the band grows with sample size.
2. The band is centered on a poor estimator of the unknown function.
3. The width of the band is independent of the data, and hence cannot adapt to the smoothness of the unknown function.

Problems 1 and 2 are easily remedied by using standard smoothing methods. But the results of Low (1997) suggest that problem 3 is an inevitable consequence of uniform coverage.

The smoother the functions in $\mathcal{H}$, the smaller the width necessary to achieve uniform coverage. Suppose that $\mathcal{F} \subset \mathcal{H}$ contains the "smooth" functions in $\mathcal{H}$ and that $\mathcal{H} - \mathcal{F}$ is nonempty. Uniform coverage over $\mathcal{H}$ requires that the width of fixed-width bands be driven by the "rough" functions in $\mathcal{H} - \mathcal{F}$; the width will thus be large even if $f \in \mathcal{F}$. Ideally, our procedure would adjust automatically to produce narrower bands when the function is smooth ($f \in \mathcal{F}$) and wider bands when the function is rough ($f \notin \mathcal{F}$), but to do that, the width must be determined from the data. Low showed that for density estimation at a single point, fixed-width confidence intervals perform as well as random length intervals; that is, the data do not help reduce the width of the bands for smoother functions. In Section 2, we extend Low's result to nonparametric regression and show that the phenomenon is quite general. Without restrictive assumptions, confidence bands cannot adapt.

These results mean that the width of uniform confidence bands is determined by the greatest roughness we are willing to assume. Because the typical assumptions about $\mathcal{H}$ in the nonparametric regression problem are loosely held and difficult to check, the result is that the confidence band widths are essentially arbitrary. This is not satisfactory in practice.

The contrast with L^2 confidence balls is noteworthy. L^2 confidence sets have been studied by Li (1989), Juditsky and Lambert-Lacroix (2003), Beran and Dümbgen (1998), Genovese and Wasserman (2005), Baraud (2004), Hoffman and Lepski (2002), Cai and Low (2006) and Robins and van der Vaart (2006). Let

$$B = \left\{ f \in \mathbb{R}^n : \frac{1}{n}\sum_{i=1}^{n}(f_i - \hat{f}_i)^2 \le R_n^2 \right\} \tag{6}$$

for some $\hat{f}$ and suppose that

$$\inf_{f \in \mathbb{R}^n} \mathbb{P}_f\{f \in B\} \ge 1 - \alpha. \tag{7}$$

Then

$$\inf_{f \in \mathbb{R}^n} \mathsf{E}_f(R_n) \ge \frac{C_1}{n^{1/4}} \quad \text{and} \quad \sup_{f \in \mathbb{R}^n} \mathsf{E}_f(R_n) \ge C_2, \tag{8}$$

where C_1 and C_2 are positive constants. Moreover, there exist confidence sets that achieve the faster $n^{-1/4}$ rate at some points in $\mathbb{R}^n$. Because fixed-radius confidence sets necessarily have radius of size $O(1)$, the supremum in (8) implies such confidence sets must have random radii. We can construct random-radius confidence balls that improve on fixed-radius confidence sets, for example, by obtaining a smaller radius for subsets of smoother functions f. L^2 confidence balls can therefore adapt to the unknown smoothness of f. Unfortunately, confidence balls can be difficult to work with in high dimensions (large n) and tend to constrain many features of interest rather poorly, for which reasons confidence bands are often desired.

The issues dealt with in this paper are related to the problem of constructing valid confidence intervals for parameters in linear models after model selection. Discussion on this issue as well as some methods for attacking the problem can be found in Kabaila (1995, 1998), Andrews and Guggenberger (2007) and Leeb and Pötscher (2005).

It is also interesting to compare the adaptivity results for estimation and inference. Estimators exist [e.g., Donoho et al. (1995)] that can adapt to unknown smoothness, achieving near optimal rates of convergence over a broad scale of spaces. But since confidence bands cannot adapt, the minimum width bands that achieve uniform coverage over the same scale of spaces have width $O(1)$, overwhelming the differences among reasonable estimators. We are left knowing that we are close to the true function but being unable to demonstrate it inferentially.

The message we take from the nonadaptivity results in Low (1997) and Section 2 of this paper is that the problem of constructing confidence bands for f over nonparametric classes is simply too difficult under the usual definition of coverage. Instead, we introduce a slightly weaker notion—surrogate coverage—under which it is possible to obtain adaptive bands while allowing sharp inferences about the main features of f.

1.2. *Surrogates.* Figure 1 shows two situations where a band fails to capture the true function. The top plot shows a conservative failure: the only place where f is not contained in the band is when the bands are smoother than the truth. The

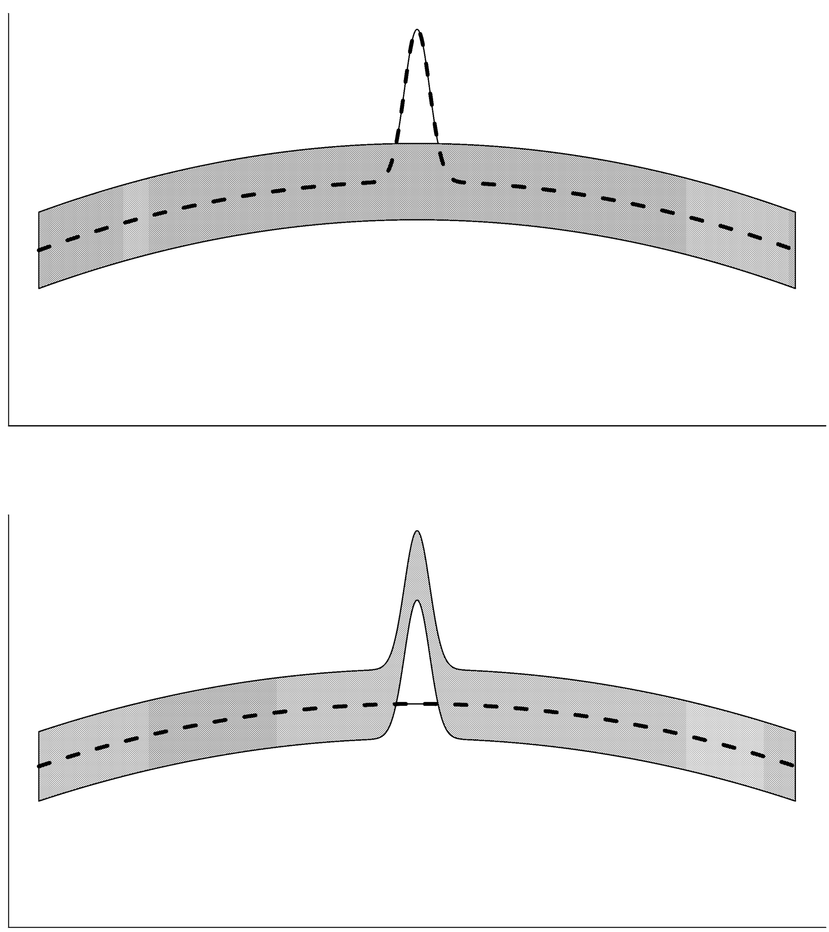

FIG. 1. *The top plot shows a conservative failure: the only place where f is not contained in the band is when the bands are smoother than the truth. The bottom plot shows a liberal failure: the only place where f is not contained in the band is when the bands are less smooth than the truth. The usual notion of coverage treats these failures equally.*

bottom plot shows a liberal failure: the only place where f is not contained in the band is when the bands are less smooth than the truth. The usual notion of coverage treats these failures equally. Yet, in some sense, the second error is more serious than the first since the bands overstate the complexity.

We are thus led to a different approach that treats conservative errors and liberal errors differently. The basic idea is to find a function $f^\star$ that is simpler than f as in Figure 2. We then require that

$$\mathbb{P}_f\{L \le f \le U \text{ or } L \le f^\star \le U\} \ge 1-\alpha \qquad \text{for all functions } f. \tag{9}$$

More generally, we will define a finite set of surrogates $F^\star \equiv F^\star(f) = \{f, f_1^\star, \dots, f_m^\star\}$ and require that a surrogate confidence band (L, U) satisfy

$$\inf_f \mathbb{P}_f\{L \le g \le U \text{ for some } g \in F^\star\} \ge 1-\alpha. \tag{10}$$

We will also consider bands that are adaptive in the following sense: if f lies in some subspace $\mathcal{F}$, then with high probability $\|U-L\|_\infty \le w(\mathcal{F})$, where $w(\mathcal{F})$ is the best width of a uniformly valid confidence band (under the usual definition of coverage) based on the a priori knowledge that $f \in \mathcal{F}$. Among possible surrogates, a surrogate will be optimal if it admits a valid, adaptive procedure and the set $\{f \in \mathcal{F} : F^\star(f) = \{f\}\}$ is as large as possible.

1.3. *Summary of results.* In Section 2, we show that Low's result on density estimation holds in regression as well. Fixed width bands do as well as random width bands, thus ruling out adaptivity. We show this when $\mathcal{H}$ is the set of all functions and when $\mathcal{H}$ is a ball in a Lipschitz, Sobolev or Besov space.

Section 3 gives our main results. Theorem 18 establishes lower bounds on the width for any valid surrogate confidence band. Let $\mathcal{F}$ be a subspace of dimension d in $\mathbb{R}^n$. The functions that prevent adaptation are those that are close to $\mathcal{F}$ in L^2 but far in L^∞. Loosely speaking, such functions are close to $\mathcal{F}$ except for isolated, spiky features. If $\|f - \Pi f\|_2 < \varepsilon_2$ and $\|f - \Pi f\|_\infty > \varepsilon_\infty$, for tuning constants $\varepsilon_2, \varepsilon_\infty$, define the surrogate $f^\star$ to be the projection of f onto $\mathcal{F}$, Πf. Otherwise, define $f^\star = f$. We show that if $\mathbb{P}_f\{\|U-L\|_\infty < w\} \ge 1-\gamma$ for all $f \in \mathcal{F}$, then

$$w \ge \max(w_{\mathcal{F}}(\alpha,\gamma,\sigma), v(\varepsilon_2,\varepsilon_\infty,n,d,\alpha,\gamma,\sigma)), \tag{11}$$

where $w_{\mathcal{F}}$ is the minimum width for a uniform confidence band knowing a priori that $f \in \mathcal{F}$ and $v(\varepsilon_2,\varepsilon_\infty,n,d,\alpha,\gamma)$ is described later.

Corollary 30 shows that for proper choice of ε_2 and ε_∞, the v term in the previous equation can be made smaller than $w_{\mathcal{F}}$. Figure 3 represents the functions involved; the gray shaded area are those functions that are replaced by surrogates in the coverage statement, denoted later by $\mathcal{S}(\varepsilon_2,\varepsilon_\infty)$. These are the functions that are both hard to distinguish from $\mathcal{F}$ (because they are close to it) and hard to cover (because they are "spiky"). The optimal choice of ε_2 and ε_∞ minimizes the volume of this set while making the right-hand side in inequality (11) equal to $w_{\mathcal{F}}$.

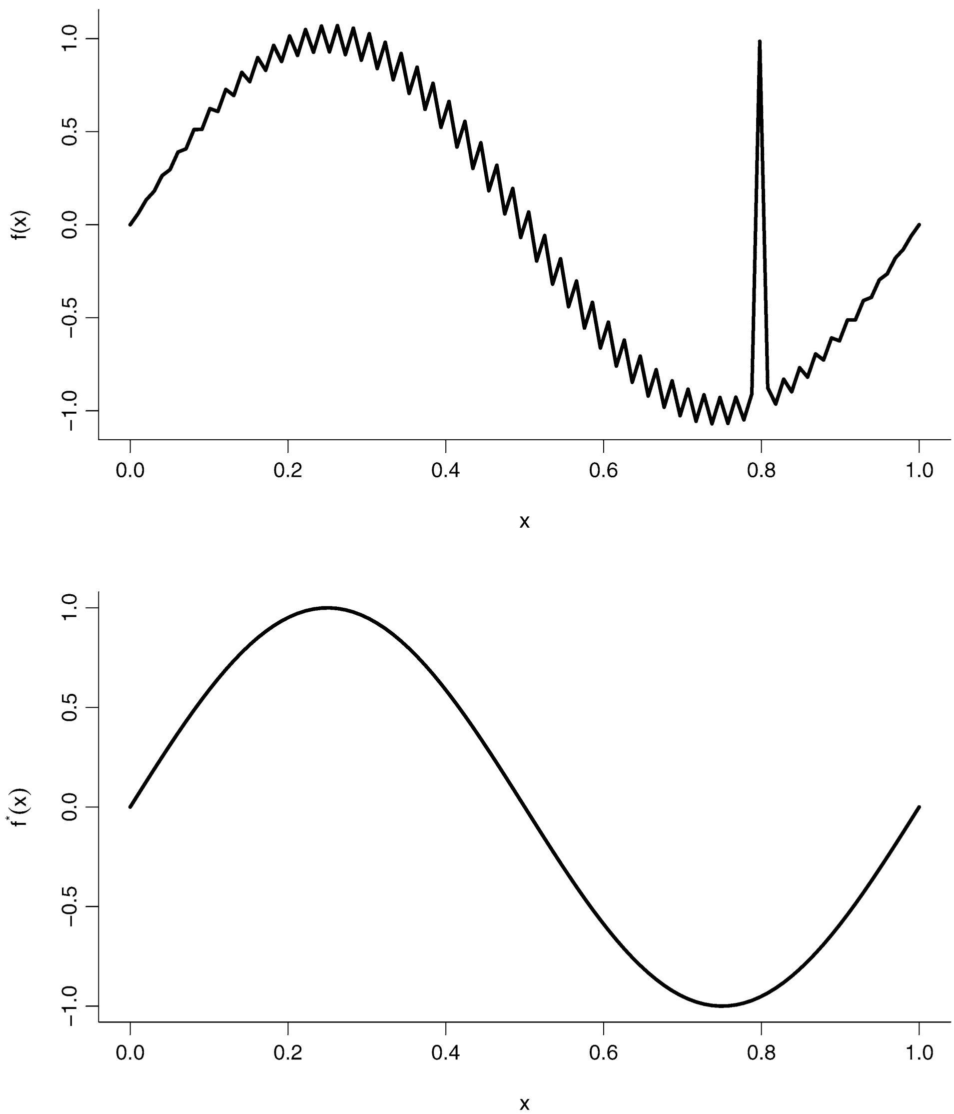

FIG. 2. *The top plot shows a complicated function f. The bottom shows a surrogate $f^\star$ which is simpler than f but retains the main, estimable features of f. Adaptation is possible if we cover $f^\star$ instead of f.*

Put another way, the richest model that permits adaptive confidence bands under the usual notion of coverage is $\mathcal{F} = \mathbb{R}^n - \mathcal{S}(\varepsilon_2, \varepsilon_\infty)$.

Theorem 29 gives a procedure that comes within a factor of 2 of attaining the lower bound for finite-samples. The procedure conducts goodness of fit tests for subspaces and constructs bands centered on the estimator of the lowest dimensional nonrejected subspace. Such a procedure actually reflects common practice.

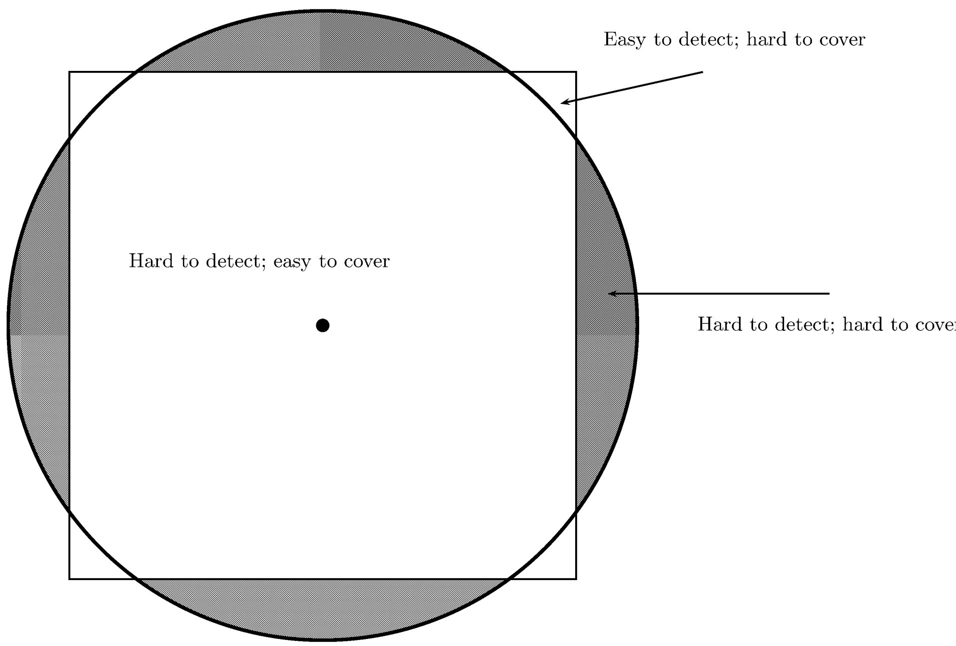

FIG. 3. *The dot at the center represents the subspace $\mathcal{F}$. The shaded area is the set of spoilers $\mathcal{S}(\varepsilon_2, \varepsilon_\infty)$ of vectors for which $f^\star \neq f$. If these vectors were not surrogated, adaptation is not possible. The nonshaded area is the invariant set $\mathcal{I}(\varepsilon_2, \varepsilon_\infty) = \{f : f^\star = f\}$.*

It is not uncommon to fit a model, check the fit, and if the model does not fit then we fit a more complex model. In this sense, we view our results as providing a rigorous basis for common practice. It is known that pretesting followed by inference does not lead to valid inferences for f [Leeb and Pötscher (2005)]. But if we can accept that sometimes we cover a surrogate $f^\star$ rather than f, then validity is restored.

These results are proved in Section 4.

1.4. *Related work.* The idea of estimating the detectable part of f is present, at least implicitly, in other approaches. Davies and Kovac (2001) separate the data into a simple piece plus a noise piece which is similar in spirit to our approach. Another related idea is scale-space inference due to Chaudhuri and Marron (2000) who focus on inference for all smoothed versions of f rather than f itself. Also related is the idea of oversmoothing as described in Terrell (1990) and Terrell and Scott (1985). Terrell argues that "By using the most smoothing that is compatible with the scale of the problem, we tend to eliminate accidental features." The idea of one-sided inference in Donoho (1988) has a similar spirit. Here, one constructs confidence intervals of the form $[L, \infty)$ for functionals such as the number of modes of a density. Bickel and Ritov (2000) make what they call a "radical proposal" to "...determine how much bias can be tolerated without [interesting] features being obscured.” We view our approach as a way of implementing their suggestion. Another related idea is contained in Donoho (1995) who showed that if $\hat{f}$ is the soft threshold estimator of a function and $f(x) = \sum_j \theta_j \psi_j(x)$ is an expansion in an unconditional basis, then $\mathbb{P}_f\{\hat{f} \preceq f\} \geq 1 - \alpha$ where $\hat{f} = \sum_j \hat{\theta}_j \psi_j$ and $\hat{f} \preceq f$ means that $|\hat{\theta}_j| \leq |\theta_j|$ for all j. Finally, we remind the reader that there is a plethora of work on adaptive estimation; see, for example, Cai and Low (2004) and references therein.

1.5. *Notation.* If L and U are random functions on $\mathcal{X} = \{x_1, \ldots, x_n\}$ such that $L \leq U$, we define $B = (L, U)$ to be the (random) set of all functions g on $\mathcal{X}$ for which $L \leq g \leq U$. We call B (or equivalently, the pair L, U) a band; the band covers a function f if $f \in B$ (or equivalently, if $L \leq f \leq U$). Define its width to be the random variable

$$W = \|U - L\|_\infty = \max_{1 \leq i \leq n} \big(U(x_i) - L(x_i)\big). \tag{12}$$

Because we are constructing bands on $\mathcal{X} = \{x_1, \ldots, x_n\}$, we most often refer to functions in terms of their evaluations $f = (f(x_1), \ldots, f(x_n)) \in \mathbb{R}^n$. When we need to refer to a space of functions to which f belongs, we use a $\tilde{}$ to denote the function space and no $\tilde{}$ to denote the vector space of evaluations. Thus, if $\tilde{\mathcal{A}}$ is the space of all functions, then $\mathcal{A} = \mathbb{R}^n$. In both cases, we use the same symbol for the function and let the meaning be clear from context; for example, $f \in \tilde{\mathcal{A}}$ is the function and $f \in \mathcal{A}$ is the vector $(f(x_1), \ldots, f(x_n))$. Define the following norms on $\mathbb{R}^n$:

$$\|f\| = \|f\|_2 = \sqrt{\frac{1}{n}\sum_{i=1}^{n} f_i^2}, \qquad \|f\|_\infty = \max_i |f_i|.$$

We use $\langle \cdot, \cdot \rangle$ to denote the inner product $\langle f, g \rangle = \frac{1}{n}\sum_{i=1}^{n} f_i g_i$ corresponding to $\| \cdot \|$.

If $\mathcal{F}$ is a subspace of $\mathbb{R}^n$, we define $\Pi_{\mathcal{F}}$ to be the Euclidean projection onto $\mathcal{F}$, using just Π if the subspace is clear from context. We use

$$e_i = (\underbrace{0, \ldots, 0}_{i-1 \text{ times}}, 1, \underbrace{0, \ldots, 0}_{n-i \text{ times}})^T \tag{13}$$

to denote the standard basis on $\mathbb{R}^n$.

If F_θ is a family of c.d.f.’s indexed by θ, we write $F_\theta^{-1}(\alpha)$ to denote the lower-tail α-quantile of F_θ. For the standard normal distribution, however, we use z_α to denote the upper-tail α-quantile, and we denote the c.d.f. and p.d.f., respectively, by Φ and ϕ.

Throughout the paper we assume that σ is a known constant; in some cases, we simply set $\sigma = 1$. But see Remark 22 about the unknown σ case.

2. Nonadaptivity of bands. In this section we construct lower bounds on the width of valid confidence bands analogous to (8) and we show that the lower bound is achieved by fixed-width bands.

Low (1997) considered estimating a density f in the class

$$\mathcal{F}(a,k,M)=\left\{f: f\geq 0, \int f=1, f(x_0)\leq a, \|f^{(k)}(x)\|_\infty \leq M\right\}.$$

He shows that if C_n is a confidence interval for $f(0)$, that is,

$$\inf_{f\in\mathcal{F}(a,k,M)} \mathbb{P}_f\{f(x_0)\in C_n\}\geq 1-\alpha,$$

then, for every $\varepsilon>0$, there exists $N=N(\varepsilon,M)$ and $c>0$ such that, for all $n\geq N$,

$$\mathsf{E}_f(\text{length}(C_n))\geq cn^{-k/(2k+1)} \tag{14}$$

for all $f\in\mathcal{F}(a,k,M)$ such that $f(0)>\varepsilon$. Moreover, there exists a fixed-width confidence interval C_n and a constant c_1 such that $\mathsf{E}_f(\text{length}(C_n))\leq c_1 n^{-k/(2k+1)}$ for all $f\in\mathcal{F}(a,k,M)$. Thus, the data play no role in constructing a rate-optimal band, except in determining the center of the interval.

For example, if we use kernel density estimation, we could construct an optimal bandwidth $h=h(n,k)$ depending only on n and k—but not the data—and construct the interval from that kernel estimator. This makes the interval highly dependent on the minimal amount of smoothness k that is assumed. And it rules out the usual data-dependent bandwidth methods such as cross-validation.

Now return to the regression model

$$Y_i=f_i+\sigma\varepsilon_i,\qquad i=1,\dots,n, \tag{15}$$

where $\varepsilon_1,\dots,\varepsilon_n$ are independent, Normal$(0,1)$ random variables, and $f=(f_1,\dots,f_n)\in\mathbb{R}^n$.

THEOREM 1. *Let $B=(L,U)$ be a $1-\alpha$ confidence band over Θ, where $0<\alpha<1/2$ and let $g\in\Theta$. Suppose that Θ contains a finite set of vectors Ω, such that*:

1. *for every distinct pair $f,\nu\in\Omega$, we have $\langle f-g,\nu-g\rangle=0$ and*
2. *for some $0<\varepsilon<(1/2)-\alpha$,*

$$\max_{f\in\Omega}\frac{e^{n\|f-g\|^2/\sigma^2}}{|\Omega|}\leq\varepsilon^2. \tag{16}$$

Then,

$$\mathsf{E}_g(W)\geq(1-2\alpha-2\varepsilon)\min_{f\in\Omega}\|g-f\|_\infty. \tag{17}$$

We begin with the case where $\Theta = \mathbb{R}^n$. We will obtain a lower bound on the width of any confidence band and then show that a fixed-width procedure attains that width. The results hinge on finding a least favorable configuration of mean vectors that are as far away from each as possible in L^∞ while staying a fixed distance ε in total-variation distance.

THEOREM 2. *Let $\mathcal{H} = \mathbb{R}^n$ and fix $0 < \alpha < 1/2$. Let $B = (L, U)$ be a $1 - \alpha$ confidence band over $\mathcal{H}$. Then, for every $0 < \varepsilon < (1/2) - \alpha$,*

$$\inf_{f \in \mathbb{R}^n} \mathsf{E}_f(W) \geq (1 - 2\alpha - 2\varepsilon)\sigma\sqrt{\log(n\varepsilon^2)}. \tag{18}$$

The bound is achieved (up to constants) by the fixed-width Bonferroni bands: $\ell_i = Y_i - \sigma z_{\alpha/(2n)}$, $u_i = Y_i + \sigma z_{\alpha/(2n)}$.

THEOREM 3 (Lipshschitz balls). *Define $x_i = i/n$ for $1 \leq i \leq n$. Let*

$$\tilde{\mathcal{H}}(L) = \{f : |f(x) - f(y)| \leq L|x - y|, x, y \in [0, 1]\}, \tag{19}$$

be a ball in Lipschitz space, and let

$$\mathcal{H}(L) = \{(f(x_1), \ldots, f(x_n)) : f \in \tilde{\mathcal{H}}(L)\} \tag{20}$$

be the vector of evaluations on $\mathcal{X}$. Fix $0 < \alpha < 1/2$ and let $B = (L, U)$ be a $1 - \alpha$ confidence band over $\mathcal{H}(L)$. Then, for every $0 < \varepsilon < (1/2) - \alpha$, $\inf_{f \in \mathcal{H}(L)} \mathsf{E}_f(W) \geq a_n$ where

$$\begin{aligned} a_n = {} & \left(\frac{\log n}{n}\right)^{1/3} \times \left(\frac{L\sigma^2}{2}\right)^{1/3} \\ & \times \left(1 + \frac{3\log(1+\varepsilon^2)}{\log n} + \frac{2\log(L/(2\sigma))}{\log n} \right. \\ & \qquad \left. - \frac{\log((1/3)\log n + \log(1+\varepsilon^2) + (2/3)\log(L/(2\sigma)))}{\log n}\right). \end{aligned}$$

The lower bound is achieved (up to logarithmic factors) by a fixed-width procedure.

THEOREM 4 (Sobolev balls). *Let $\tilde{\mathcal{H}}(p, c)$ be a Sobolev ball of order p and radius c and let $B = (L, U)$ be a $1 - \alpha$ confidence band over $\mathcal{H}(p, c)$. For every $0 < \varepsilon < (1/2) - \alpha$, for every $\delta > 0$, and all large n,*

$$\inf_{F \in \mathcal{H}(p, c-\delta)} \mathsf{E}_F(W) \geq (1 - 2\alpha - 2\varepsilon)\left(\frac{c_n}{n^{p/(2p+1)}}\right) \tag{21}$$

for some c_n that increases at most logarithmically. The bound is achieved (up to logarithmic factors) by a fixed-width band procedure.

THEOREM 5 (Besov balls). *Let $\tilde{\mathcal{H}}(p,q,\xi,c)$ be ball of size c in the Besov space $B^{\xi}_{p,q}$ and et $B=(L,U)$ be a $1-\alpha$ confidence band over $\mathcal{H}(p,q,\xi,c)$. For every $0<\varepsilon<(1/2)-\alpha$, and every $\delta>0$,*

$$\inf_{f\in\mathcal{H}(p,q,\xi,c-\delta)} \mathsf{E}_f(W) \ge c_n(1-2\alpha-2\varepsilon)n^{-1/(1/p-\xi-1/2)}. \tag{22}$$

The bound is achieved (*up to logarithmic factors*) *by a fixed-width procedure.*

3. Adaptive bands. Let $\{\mathcal{F}_T : T\in\mathcal{T}\}$ be a scale of linear subspaces. Let w_T denote the smallest width of any confidence band when it is known that $f\in\mathcal{F}_T$ (defined more precisely below). We would like to define an appropriate surrogate and a procedure that gets as close as possible to the target width w_T when $f\in\mathcal{F}_T$. To clarify the ideas, Section 3.2 develops our results in the special case where the subspaces are $\{\mathcal{F},\mathbb{R}^n\}$ for a fixed $\mathcal{F}$ of dimension $d<n$. Section 3.3 handles the more general case of a sequence of nested subspaces.

3.1. *Preliminaries.* We begin by defining several quantities that will be used throughout. Let $\tau(\varepsilon)$ denote the total variation distance between a $N(0,1)$ and a $N(\varepsilon,1)$ distribution. Thus, $\tau(\varepsilon)=\Phi(\varepsilon/2)-\Phi(-\varepsilon/2)$. Then, $\varepsilon\phi(\varepsilon/2)\le\tau(\varepsilon)\le\varepsilon\phi(0)$ and $\tau(\varepsilon)\sim\varepsilon\phi(0)$ as $\varepsilon\to 0$.

LEMMA 6. *If $P=N(f,\sigma^2 I)$ and $Q=N(g,\sigma^2 I)$ are multivariate Normals with $f,g\in\mathbb{R}^n$ then*

$$d_{\mathrm{TV}}(P,Q)=\tau\left(\frac{\sqrt{n}\|f-g\|}{\sigma}\right). \tag{23}$$

We will need several constants. For $0<\alpha<1$ and $0<\gamma<1-2\alpha$ define

$$\kappa(\alpha,\gamma)=\bigl(2\log\bigl(1+4(1-\gamma-2\alpha)^2\bigr)\bigr)^{1/4}. \tag{24}$$

For $0<\beta<1-\xi<1$ and integer $m\ge 1$ define $Q=Q(m,\beta,\xi)$ to be the solution of

$$\xi = 1-F_{0,m}\bigl(F^{-1}_{Q\sqrt{m},m}(\beta)\bigr), \tag{25}$$

where $F_{a,d}$ denotes the c.d.f. of a χ^2 random variable with d degrees of freedom and noncentrality parameter a.

LEMMA 7. *There is a universal constant $\Lambda(\beta,\xi)$ such that $Q(m,\beta,\xi)\le\Lambda(\beta,\xi)$ for all $m\ge 1$. For example, $\Lambda(0.05,0.05)\le 6.25$. Suppose now that $m=m_n$, $\beta=\beta_n$ and $\xi=\xi_n$ are all functions of n. As long as $-\log\beta_n\le\log n$ and $-\log\xi_n\le\sqrt{\log n}$, then $Q(m_n,\beta_n,\xi_n)=O(\sqrt{\log n})$.*

Next, define

$$E(m,\alpha,\gamma) = \max(Q(m,\alpha,\gamma), 2\kappa(\alpha,\gamma)), \tag{26}$$

for $0 < \alpha < 1$ and $0 < \gamma < 1 - 2\alpha$.

Finally, if $\mathcal{F}$ is a subspace of dimension d, define

$$\Omega_{\mathcal{F}} = \max_{1\le i\le n} \frac{\|\Pi_{\mathcal{F}} e_i\|}{\|e_i\|}, \tag{27}$$

where e_i is defined in equation (13). Note that $0 \le \Omega_{\mathcal{F}} \le 1$. The value of $\Omega_{\mathcal{F}}$ relates to the geometry of $\mathcal{F}$ as a hyperplane embedded in $\mathbb{R}^n$, as seen through the following results.

LEMMA 8. *Let $\mathcal{F}$ be a subspace of $\mathbb{R}^n$. Then*

$$\min\{\|v\| : v \in \mathcal{F}, \|v\|_\infty = \varepsilon\} = \frac{\varepsilon}{\sqrt{n}\Omega_{\mathcal{F}}}, \tag{28}$$

$$\max\{\|v\|_\infty : v \in \mathcal{F}, \|v\| = \varepsilon\} = \varepsilon\sqrt{n}\Omega_{\mathcal{F}}. \tag{29}$$

REMARK 9. In the case $\mathcal{X} = [0,1]$, the norm would be $\|f\|^2 = \int_0^1 f^2(x)\,dx$ and $\Omega_{\mathcal{F}}$ would be defined by way of (29).

LEMMA 10. *Let $\{\phi_1,\dots,\phi_d\}$ be orthonormal vectors with respect to $\|\cdot\|$ in $\mathbb{R}^n$ and let $\mathcal{F}$ be the linear span of these vectors. Then $\Omega_{\mathcal{F}} = \max_{1\le i\le n}\sqrt{\sum_{j=1}^d \phi_{ji}^2/n}$. In particular, if $\max_j \max_i \phi_j(i) \le c$ then $\Omega_{\mathcal{F}} \le c\sqrt{d/n}$.*

LEMMA 11. *Let $\{\phi_1,\dots,\phi_d\}$ be orthonormal functions on $[0,1]$. Define $\mathcal{H}_j$ to be the linear span of $\{\phi_1,\dots,\phi_j\}$. Let $x_i = i/n$, $i = 1,\dots,n$, and $\mathcal{F}_j = \{f = (h(x_1),\dots,h(x_n)) : h \in \mathcal{H}_j\}$. Then, $\Omega_{\mathcal{F}}^2 = n^{-1}\max_i \sum_{j=1}^d \phi_j^2(x_i) + O(1/n)$ and if $\max_j \sup_x \phi_j(x) \le c$ then $\Omega_{\mathcal{F}}^2 \le c^2 d/n + O(1/n)$.*

In addition, we need the following lemma first proved, in a related form, in Baraud (2002).

LEMMA 12. *Let $\mathcal{F}$ be a subspace of dimension d. Let $0 < \delta < 1 - \xi$ and*

$$\varepsilon = \frac{(n-d)^{1/4}}{\sqrt{n}}\bigl(2\log(1+4\delta^2)\bigr)^{1/4}. \tag{30}$$

Define $A = \{f : \|f - \Pi_{\mathcal{F}} f\| > \varepsilon\}$. Then,

$$\beta \equiv \inf_{\phi_\alpha\in\Phi_\xi} \sup_{f\in A} \mathbb{P}_f\{\phi_\xi = 0\} \ge 1 - \xi - \delta, \tag{31}$$

where

$$\Phi_\xi = \left\{\phi_\xi : \sup_{f\in\mathcal{F}} \mathbb{P}_f\{\phi_\xi = 0\} \le \xi\right\} \tag{32}$$

is the set of level ξ tests.

3.2. *Single subspace.* To begin, we start with a single subspace $\mathcal{F}$ of dimension d.

DEFINITION 13. For given $\varepsilon_2, \varepsilon_\infty > 0$, define the *surrogate* $f^\star$ of f by

$$f^\star = \begin{cases} \Pi f, & \text{if } \|f - \Pi f\|_2 \le \varepsilon_2 \text{ and } \|f - \Pi f\|_\infty > \varepsilon_\infty, \\ f, & \text{otherwise.} \end{cases} \tag{33}$$

Define the *surrogate set* of f, $F^\star(f) = \{f, f^\star\}$, which will be a singleton when $f^\star = f$. Define the *spoiler set* $\mathcal{S}(\varepsilon_2, \varepsilon_\infty) = \{f \in \mathbb{R}^n : f^\star \ne f\}$ and the *invariant set* $\mathcal{I}(\varepsilon_2, \varepsilon_\infty) = \{f : f^\star = f\}$.

We give a schematic diagram in Figure 3. The gray area represents $\mathcal{S}(\varepsilon_2, \varepsilon_\infty)$. These are the functions that preclude adaptivity. Being close to $\mathcal{F}$ in L^2 makes them hard to detect but being far from $\mathcal{F}$ in L^∞ makes them hard to cover. To achieve adaptivity we must settle for sometimes covering $\Pi_{\mathcal{F}} f$.

3.2.1. *Lower bounds.* We begin with two lemmas. The first controls the minimum width of a band and the second controls the maximum. The second is of more interest for our purposes; the first lemma is included for completeness. For any $1 \le p \le \infty$, $\varepsilon > 0$, and $A \subset \mathbb{R}^n$ define

$$M_p(\varepsilon, A) = \sup\{d_{\mathrm{TV}}(P_f, P_g) : f, g \in A, \|f - g\|_p \le \varepsilon\} \tag{34}$$

and

$$m_\infty(\varepsilon, A_0, A_1) = \inf\{d_{\mathrm{TV}}(P_f, P_g) : f \in A_0, g \in A_1, \|f - g\|_\infty \ge \varepsilon\}. \tag{35}$$

LEMMA 14. *Suppose that* $\inf_{f \in A} \mathbb{P}_f\{L \le f \le U\} \ge 1 - \alpha$. *Let* $1 \le p \le \infty$ *and* $\varepsilon > 0$. *For* $f \in A$, *define*

$$\varepsilon(f, q) = \sup\{\|f - h\|_q : h \in A, \|f - h\|_p \le \varepsilon\},$$

where $1 \le q \le \infty$. *Then, for any* $A_0 \subset A$,

$$\inf_{f \in A_0} \mathbb{P}_f\{W > \varepsilon(f, \infty)\} \ge 1 - 2\alpha - \sup_{f \in A_0} M_p(\varepsilon(f, p), A), \tag{36}$$

where $W = \|U - L\|_\infty$. *If every point in* A *is contained in a subset of* A *of* ℓ^p-*diameter* ε, *then* $\varepsilon(f, p) \equiv \varepsilon$, *and*

$$\inf_{f \in A_0} \mathbb{P}_f\{W > \varepsilon\} \ge 1 - 2\alpha - M_p(\varepsilon, A). \tag{37}$$

LEMMA 15. *Suppose that* $\inf_{f \in A} \mathbb{P}_f\{L \le f \le U\} \ge 1 - \alpha$. *Suppose that* $A = A_0 \cup A_1$ (*not necessarily disjoint*). *Let* $\varepsilon > 0$ *be such that for each* $f \in A_0$ *there exists* $g \in A_1$ *for which* $\|f - g\|_\infty = \varepsilon$. *Then*,

$$\sup_{f \in A_0} \mathbb{P}_f\{W > \varepsilon\} \ge 1 - 2\alpha - m_\infty(\varepsilon, A_0, A_1), \tag{38}$$

where $W = \|U - L\|_\infty$.

Now we establish the target rate, the smallest width of a band if we knew a priori that $f \in \mathcal{F}$. Define

$$w_{\mathcal{F}} \equiv w_{\mathcal{F}}(\alpha, \gamma, \sigma) = \Omega_{\mathcal{F}}\, \sigma\, \tau^{-1}(1 - 2\alpha - \gamma). \tag{39}$$

THEOREM 16. *Let* $0 < \gamma < 1 - \alpha$. *Suppose that*

$$\inf_{f \in \mathcal{F}} \mathbb{P}_f\{L \le f \le U\} \ge 1 - \alpha. \tag{40}$$

If $\inf_{f \in \mathcal{F}} \mathbb{P}_f\{W \le w\} \ge 1 - \gamma$ *then* $w \ge w_{\mathcal{F}}$.

A band that achieves this width, up to logarithmic factors, is $(L, U) = \hat{f} \pm c$ *where* $\hat{f} = \Pi Y$ *and* $c = \max_i \sigma \Pi_{ii} z_{\alpha/2n}$.

REMARK 17. Using an argument similar to that in Theorem 1, it is possible to improve this lower bound by an additional $\sqrt{\log d}$ factor, but this is inconsequential to the rest of the paper.

Next, we give the main result for this case.

$$v_0(\varepsilon_2, \varepsilon_\infty, n, \alpha, \gamma, \sigma) = \min\{\sqrt{n}\varepsilon_2,\ \varepsilon_\infty,\ \sigma\tau^{-1}(1 - 2\alpha - \gamma)\}, \tag{41}$$

$$v_1(\varepsilon_2, n, d, \alpha, \gamma, \sigma) = \begin{cases} 0, & \text{if } \varepsilon_2 \ge 2v_2(n, d, \alpha, \gamma), \\ v_2(n, d, \alpha, \gamma), & \text{if } \varepsilon_2 < 2v_2(n, d, \alpha, \gamma), \end{cases} \tag{42}$$

$$v_2(n, d, \alpha, \gamma) = \kappa(\alpha, \gamma)(n - d)^{1/4} n^{-1/2} \tag{43}$$

and define

$$v(\varepsilon_2, \varepsilon_\infty, n, d, \alpha, \gamma, \sigma) = \max(v_0, v_1). \tag{44}$$

THEOREM 18 (Lower bound for surrogate confidence band width). *Fix* $0 < \alpha < 1$ *and* $0 < \gamma < 1 - 2\alpha$. *Suppose that for bands* $B = (L, U)$

$$\inf_{f \in \mathbb{R}^n} \mathbb{P}_f\{F^\star(f) \cap B \ne \varnothing\} \ge 1 - \alpha. \tag{45}$$

Then,

$$\inf_{f \in \mathcal{F}} \mathbb{P}_f\{W \le w\} \ge 1 - \gamma \tag{46}$$

implies

$$\begin{aligned} w &\ge \underline{w}(\mathcal{F}, \varepsilon_2, \varepsilon_\infty, n, d, \alpha, \gamma, \sigma) \\ &\equiv \max\{w_{\mathcal{F}}(\alpha, \gamma, \sigma), v(\varepsilon_2, \varepsilon_\infty, n, d, \alpha, \gamma, \sigma)\}. \end{aligned} \tag{47}$$

The inequality (45) ensures that B is a valid surrogate confidence band: for every function, either the function or its surrogate is covered with at least the target probability. The result gives a probabilistic lower bound on the width of the band that is at least as big as the best a priori width for the subspace. As we will see, with proper choice of ε_2 and ε_∞, the v term can be made small, giving the subspace width $w_{\mathcal{F}}$ for the lower bound.

Next, we address the question of optimality. Consider, for example, the trivial surrogate that maps all functions to 0. We can cover the surrogate using 0 width bands with probability 1, but this would not be too interesting. There is a tradeoff between the width of the bands on low-dimensional subspaces and the volume of the spoiler set, the functions that are surrogated. We characterize optimality here as minimizing the volume of the spoiler set $\mathcal{S}(\varepsilon_2, \varepsilon_\infty)$ while still attaining the target width with high probability when f truly lies in the subspace. In this sense, the surrogate defined above is optimal.

THEOREM 19 (Optimality). *Let $\underline{w}$ denote the right-hand side of inequality* (47). *Then $\underline{w} \geq w_{\mathcal{F}}$, where $w_{\mathcal{F}}$ is defined in* (39). *Setting*

$$\varepsilon_2 = 2\kappa(\alpha, \gamma)(n-d)^{1/4} n^{-1/2}, \quad \varepsilon_\infty = w_{\mathcal{F}}$$

minimizes Volume$(\mathcal{S}(\varepsilon_2, \varepsilon_\infty))$ *subject to achieving the lower bound on $\underline{w}$.*

3.2.2. *Achievability.* Having established a lower bound, we need to show that the lower bound is sharp. We do this by constructing a finite-sample procedure that achieves the bound within a factor of 2. Let $F_{a,d}$ denote the c.d.f. of a χ^2 random variable with d degrees of freedom and noncentrality parameter a and let $\chi^2_{\alpha,d} = F^{-1}_{0,d}(1-\alpha)$. Let $T = \|Y - \Pi Y\|^2$ and define

$$B = (L, U) = \hat{f} \pm c\sigma, \tag{48}$$

where

$$\hat{f} = \begin{cases} Y, & \text{if } T > \chi^2_{\gamma,n-d}, \\ \Pi Y, & \text{if } T \leq \chi^2_{\gamma,n-d} \end{cases} \tag{49}$$

and

$$c = z_{\alpha/2n} \times \begin{cases} \omega_{\mathcal{F}} + \varepsilon_\infty, & \text{if } T \leq \chi^2_{\gamma,n-d}, \\ 1, & \text{if } T > \chi^2_{\gamma,n-d}. \end{cases} \tag{50}$$

THEOREM 20. *If $\gamma \geq 1 - F_{0,n-d}(F^{-1}_{n\varepsilon_2^2,n-d}(\alpha/2))$ then*

$$\inf_{f \in \mathbb{R}^n} \mathbb{P}_f\{F^\star(f) \cap B \neq \varnothing\} \geq 1 - \alpha \tag{51}$$

and

$$\inf_{f \in \mathcal{F}} \mathbb{P}_f\{W \leq w_{\mathcal{F}} + \varepsilon_\infty\} \geq 1 - \gamma. \tag{52}$$

If $\varepsilon_2 \geq E(n-d, \alpha/2, \gamma)(n-d)^{1/4} n^{-1/2}$, *where* $E(m, \alpha, \gamma)$ *is defined in* (26), *then*

$$\inf_{f \in \mathcal{F}} \mathbb{P}_f\{W \leq 2\underline{w}(\mathcal{F}, \varepsilon_2, \varepsilon_\infty, \alpha, \gamma, n, d)\} \geq 1 - \gamma, \tag{53}$$

where $\underline{w}(\mathcal{F}, \varepsilon_2, \varepsilon_\infty, \alpha, \gamma, n, d)$ *is defined* (47). *Hence*, *the procedure adapts to within a logarithmic factor of the lower bound* $\underline{w}$ *given in Theorem* 18.

COROLLARY 21. *Setting*

$$\varepsilon_2 = E(n-d, \alpha/2, \gamma)(n-d)^{1/4} n^{-1/2}, \qquad \varepsilon_\infty = w_{\mathcal{F}}$$

in the above procedure, *minimizes* $\mathrm{Volume}(\mathcal{S}(\varepsilon_2, \varepsilon_\infty))$ *subject to satisfying* (53).

REMARK 22. The results can be extended to unknown σ by replacing σ with a nonparametric estimate $\hat{\sigma}$. However, the results are then asymptotic rather than finite sample. Moreover, a minimal amount of smoothness is required to ensure that $\hat{\sigma}$ uniformly consistently estimates σ; see Genovese and Wasserman (2005). So as not to detract from our main points, we continue to take σ known.

3.2.3. *Remarks on estimation and the modulus of continuity.* It is interesting to note that the bands defined above cover the true f over a set V that is larger than $\mathcal{F}$. In this section we take a brief look at the properties of V.

Define

$$C(\alpha, a, b) = \sup_{u>0}(au + b)\big(1 - \alpha - \tfrac{1}{4} + \tfrac{1}{2}\Phi(-u/2)\big), \tag{54}$$

and let $C(\alpha) \equiv C(\alpha, 1, 0)$. Let $\mathcal{F}^\perp$ be the orthogonal complement of $\mathcal{F}$. Let $B_k^\perp(0, \varepsilon)$ be a ℓ^k-ball around 0 in $\mathcal{F}^\perp$ ($k = 2, \infty$). For $f \in \mathbb{R}^n$, let $B_k^\perp(f, \varepsilon) = f + B_k^\perp(0, \varepsilon)$. Define

$$V \equiv V(\mathcal{F}, \varepsilon_2, \varepsilon_\infty) = \bigcup_{f \in \mathcal{F}} \big(B_2^\perp(f, \varepsilon_2) \cap B_\infty^\perp(f, \varepsilon_\infty)\big). \tag{55}$$

LEMMA 23. *Let* $B = (L, U)$ *be defined as in* (48). *Then*

$$\inf_{f \in V} \mathbb{P}_f\{L \leq f \leq U\} \geq 1 - \alpha. \tag{56}$$

Let $Tf = f_1$. The next lemma gives the modulus of continuity [Donoho and Liu (1991)] of T over V which measures the difficulty of estimation over V. The modulus of continuity of T over a set $\mathcal{A}$ is

$$\omega(u, \mathcal{A}) = \sup\{|Tf - Tg| : \|f - g\|_2 \leq u;\ f, g \in \mathcal{A}\}. \tag{57}$$

Donoho and Liu showed that the difficulty of estimation over $\mathcal{A}$ is often characterized by $\omega(1/\sqrt{n}, \mathcal{A})$ in the sense that this quantity defines a lower bound on estimation rates.

LEMMA 24 (Modulus of continuity). *We have*

$$\omega(u, V) = \left(u\Omega\sqrt{n}\sqrt{\frac{\Omega^2}{1+\Omega^2}} + \min\left(\frac{u\sqrt{n}}{\sqrt{1+\Omega^2}}, \varepsilon_2 \wedge (\varepsilon_\infty/\sqrt{n})\right)\right). \tag{58}$$

Note that when $\varepsilon_2 = \varepsilon_\infty = 0$ and $\Omega \sim \sqrt{d/n}$, we have $\omega(1/\sqrt{n}, \mathcal{A}) \sim \sqrt{d/n}$ as expected. However, when $\varepsilon \equiv \varepsilon_2 = \varepsilon_\infty/\sqrt{n}$ is large we will have that $\omega(1/\sqrt{n}, \mathcal{A}) \sim \sqrt{d/n} + \varepsilon/\sqrt{1 + d^2/n}$. The extra term $\varepsilon/\sqrt{1 + d^2/n}$ reflects the "ball-like" behavior of V in addition to the subspace-like behavior of V. The bands need to cover over this extra set to maintain valid coverage and this leads to larger lower bounds than just covering over $\mathcal{F}$.

3.3. *Nested subspaces.* Now suppose that we have nested subspaces $\mathcal{F}_1 \subset \cdots \subset \mathcal{F}_m \subset \mathcal{F}_{m+1} \equiv \mathbb{R}^n$. Let Π_j denote the projector onto $\mathcal{F}_j$. We define the surrogate as follows.

DEFINITION 25. For given $\varepsilon_2 = (\varepsilon_{2,1}, \dots, \varepsilon_{2,m})$ and $\varepsilon_\infty = (\varepsilon_{\infty,1}, \dots, \varepsilon_{\infty,m})$ define

$$\mathcal{J}(f) = \{1 \le j \le m : \|f - \Pi_j f\|_2 \le \varepsilon_{2,j} \text{ and } \|f - \Pi_j f\|_\infty > \varepsilon_{\infty,j}\}. \tag{59}$$

Then define the surrogate set

$$F^\star(f) = \{\Pi_j f : j \in \mathcal{J}(f)\} \cup \{f\}. \tag{60}$$

DEFINITION 26. We say that $B = \{g : L \le g \le U\} \equiv (L, U)$ has coverage $1 - \alpha$ if

$$\inf_{f \in \mathbb{R}^n} \mathbb{P}_f\{F^\star \cap B \ne \varnothing\} \ge 1 - \alpha. \tag{61}$$

3.3.1. *Lower bounds.*

THEOREM 27 (Lower bound for surrogate confidence band width). *Fix $0 < \alpha < 1$ and $0 < \gamma < 1 - 2\alpha$. Suppose that for bands $B = (L, U)$*

$$\inf_{f \in \mathbb{R}^n} \mathbb{P}_f\{F^\star(f) \cap B \ne \varnothing\} \ge 1 - \alpha. \tag{62}$$

Then

$$\inf_{f \in \mathcal{F}_j} \mathbb{P}_f\{W \le w\} \ge 1 - \gamma, \tag{63}$$

implies

$$w \ge \underline{w}(\mathcal{F}_j, \varepsilon_{2,j}, \varepsilon_{\infty,j}, n, d_j, \alpha, \gamma, \sigma), \tag{64}$$

where $\underline{w}$ is given in Theorem 18.

THEOREM 28 (Optimality). *Let $\underline{w}$ denote the right-hand side of inequality* (64). *Then $\underline{w} \ge w_{\mathcal{F}_j}$, where $w_{\mathcal{F}_j}$ is defined in* (39). *Setting*

$$\varepsilon_{2_j} = 2\kappa(\alpha, \gamma)(n - d_j)^{1/4} n^{-1/2}, \qquad \varepsilon_{\infty, j} = w_{\mathcal{F}_j}$$

minimizes the volume of the set

$$\{f : \|f - \Pi_j f\| \le \varepsilon_{2,j} \text{ and } \|f - \Pi_j f\|_\infty > \varepsilon_{2,\infty}\} \tag{65}$$

subject to achieving the lower bound on $\underline{w}$.

3.3.2. *Achievability.* Define $T_j = \|Y - \Pi_j Y\|^2$ and $\hat{f} = \Pi_{\hat{J}} Y$, where

$$\hat{J} = \min\{1 \le j \le m : T_j \le \chi^2_{\gamma, n-d_j}\}, \tag{66}$$

where $\hat{J} = m + 1$ if the set is empty, and define

$$c_j = z_{\alpha_j/2n} \times \begin{cases} \omega_{\mathcal{F}_j}(\alpha_j) + \varepsilon_{\infty,j}, & \text{if } 1 \le j \le m, \\ 1, & \text{if } j = m + 1. \end{cases} \tag{67}$$

Finally, let $B = (L, U) = \hat{f} \pm c_{\hat{J}} \sigma$ where $\sum_j \alpha_j \le \alpha$.

THEOREM 29. *If*

$$\gamma \ge 1 - \min_j F_{0,n-d_j}(F^{-1}_{n\varepsilon^2_{2,j}, n-d_j}(\alpha_j)) \tag{68}$$

then

$$\inf_{f \in \mathbb{R}^n} \mathbb{P}_f\{F^\star \cap B \ne \varnothing\} \ge 1 - \alpha. \tag{69}$$

Let $w_j = w_{\mathcal{F}_j}(\alpha_j) + \varepsilon_{\infty,j}$. If $w_1 \le \cdots \le w_{m+1}$ then

$$\inf_{f \in \mathcal{F}_j} \mathbb{P}_f\{W \le w_j\} \ge 1 - \gamma. \tag{70}$$

If in addition $\varepsilon_{2,j} \ge E(n - d_j, \alpha_j, \gamma)(n - d_j)^{1/4} n^{-1/2}$ and $\varepsilon_{\infty,j} \le w_{\mathcal{F}_j}$ then

$$\inf_{f \in \mathcal{F}_j} \mathbb{P}_f\{W \le 2\underline{w}(\varepsilon_{2,j}, \varepsilon_{\infty,j}, \alpha_j, \gamma, n, d_j)\} \ge 1 - \gamma, \tag{71}$$

where $\underline{w}(\varepsilon_{2,j}, \varepsilon_{\infty,j}, \alpha_j, \gamma, n, d_j)$ is defined (47). *Hence, the procedure adapts to within a logarithmic factor of the lower bound $\underline{w}$ given in Theorem* 18.

COROLLARY 30. *Suppose $\alpha_1 = \cdots = \alpha_{m+1} = \alpha/(m + 1)$. Then $w_1 \le \cdots \le w_{m+1}$ so* (70) *holds. Moreover, setting*

$$\varepsilon_{2,j} = E(n - d_j, \alpha_j, \gamma)(n - d_j)^{1/4} n^{-1/2} \tag{72}$$

and

$$\varepsilon_{\infty,j} = w_{\mathcal{F}_j} \tag{73}$$

in the above procedure, minimizes the volume of the set (65) *satisfying* (64).

EXAMPLE 31. Suppose that $x_i = i/n$ and let $B_1 = [0, 1/d]$, $B_2 = (1/d, 2/d]$, $\ldots, B_d = ((d-1)/d, 1]$. Write $f = (f(x_i) : i = 1, \ldots, n)$ and let $\mathcal{F}$ denote the subspace of vectors f that are constant over each B_j. Then $\Omega_{\mathcal{F}} = \sqrt{d/n}$. The above procedure then produces a band with width no more that $O(\sqrt{d/n})$ with probability at least $1-\gamma$.

4. Proofs. In this section, we prove the main results. We omit proofs for a few of the simpler lemmas. Throughout this section, we write $x_n = O^*(b_n)$ to mean that $x_n = O(c_n b_n)$ where c_n increases at most logarithmically with n.

The following lemma is essentially from Section 3.3 of Ingster and Suslina (2003).

LEMMA 32. *Let M be a probability measure on $\mathbb{R}^n$ and let $Q(\cdot) = \int P_f(\cdot)\, dM(f)$ where $P_f(\cdot)$ denotes the measure for a multivariate Normal with mean $f = (f_1, \ldots, f_n)$ and covariance $\sigma^2 I$. Then*

$$L_1(Q, P_g) \le \sqrt{\int\int \exp\left\{\frac{n\langle f-g, \nu-g\rangle}{\sigma^2}\right\} dM(f)\, dM(\nu) - 1}. \tag{74}$$

In particular, if Q is uniform on a finite set Ω, then

$$L_1(Q, P_g) \le \sqrt{\left(\frac{1}{|\Omega|}\right)^2 \sum_{f,\nu\in\Omega} \exp\left\{\frac{n\langle f-g, \nu-g\rangle}{\sigma^2}\right\} - 1}. \tag{75}$$

PROOF. Let p_f denote the density of a multivariate Normal with mean f and covariance $\sigma^2 I$ where I is the identity matrix. Let q be the density of $Q : q(y) = \int p_f(y)\, dM(f)$. Then,

$$\begin{aligned}
\int |p_g(x) - q(x)|\, dx &= \int \frac{|p_g(x) - q(x)|}{\sqrt{p_g(x)}} \sqrt{p_g(x)}\, dx \\
&\le \sqrt{\int \frac{(p_g(x) - q(x))^2}{p_g(x)}\, dx} \qquad (76) \\
&= \sqrt{\int \frac{q^2(x)}{p_g(x)}\, dx - 1}.
\end{aligned}$$

Now,

$$\begin{aligned}
\int \frac{q^2(x)}{p_g(x)}\, dx &= \int \left(\frac{q(x)}{p_g(x)}\right)^2 p_g(x)\, dx = \mathsf{E}_g\left(\frac{q(x)}{p_g(x)}\right)^2 \\
&= \int\int \mathsf{E}_g\left(\frac{p_f(x) p_\nu(x)}{p_g^2(x)}\right) dM(f)\, dM(\nu)
\end{aligned}$$

$$= \iint \exp\left\{-\frac{n}{2\sigma^2}(\|f-g\|^2 + \|\nu - g\|^2)\right\}$$
$$\times \mathsf{E}_g\bigl(\exp\{\varepsilon^T(f+\nu-2g)/\sigma^2\}\bigr)\,dM(f)\,dM(\nu)$$
$$= \iint \exp\left\{-\frac{n}{2\sigma^2}(\|f-g\|^2 + \|\nu - g\|^2)\right\}$$
$$\times \exp\left\{\sum_{i=1}^{n}(f_i - g_i + \nu_i - g_i)^2/(2\sigma^2)\right\} dM(f)\,dM(\nu)$$
$$= \iint \exp\left\{\frac{n\langle f-g, \nu-g\rangle}{\sigma^2}\right\} dM(f)\,dM(\nu)$$

and the result follows from (76). □

PROOF OF THEOREM 1. Let $N = |\Omega|$ and let $b^2 = n\max_{f\in\Omega}\|f-g\|^2$. Let p_f denote the density of a multivariate normal with mean f and covariance $\sigma^2 I$ where I is the identity matrix. Define the mixture $q(y) = N^{-1}\sum_{f\in\Omega} p_f(y)$. By Lemma 32,

$$\int |p_g(x) - q(x)|\,dx \le \sqrt{\left(\frac{1}{N}\right)^2 \sum_{f,\nu\in\Omega} \exp\left\{\frac{n\langle f-g,\nu-g\rangle}{\sigma^2}\right\} - 1}$$
$$= \sqrt{\left(\frac{1}{N}\right)^2 [Ne^{b^2/\sigma^2} + N(N-1)] - 1}$$
$$\le \sqrt{e^{b^2/\sigma^2}/N} = \varepsilon.$$

Define two events, $A = \{\ell \le g \le u\}$ and $B = \{\ell \le f \le u, \text{ for some } f \in \Omega\}$. Then, $A \cap B \subset \{w_n \ge a\}$, where $a = \min_{f\in\Omega}\|g - f\|_\infty$. Since $\mathbb{P}_f\{\ell \le f \le u\} \ge 1-\alpha$ for all f, it follows that $\mathbb{P}_f\{B\} \ge 1-\alpha$ for all $f \in \Omega$. Hence, $Q(B) \ge 1-\alpha$. So,

$$\begin{aligned}\mathbb{P}_g\{w_n \ge a\} &\ge \mathbb{P}_g\{A\cap B\} \ge Q(A\cap B) - \varepsilon\\ &= Q(A) + Q(B) - Q(A\cup B) - \varepsilon\\ &\ge Q(A) + Q(B) - 1 - \varepsilon \ge Q(A) + (1-\alpha) - 1 - \varepsilon\\ &\ge \mathbb{P}_g\{A\} + (1-\alpha) - 1 - 2\varepsilon\\ &\ge (1-\alpha) + (1-\alpha) - 1 - 2\varepsilon = 1 - 2\alpha - 2\varepsilon.\end{aligned}$$

So, $\mathsf{E}_g(w_n) \ge (1-2\alpha-2\varepsilon)a$. □

Theorems 2, 3, 4 and 5 follow easily from Theorem 1 and so the proofs are omitted.

PROOF OF LEMMA 7. Q is the solution, with respect to c, to $\xi = 1 - F_{0,m}(r(c))$ where the function $r(c) = F^{-1}_{c\sqrt{m},m}(\beta))$ is monotonically increasing in c. Also, $F_{0,m}(r(0)) = \beta$ and $F_{0,m}(r(\infty)) = 1$ so a solution exists since $0 < \beta < 1 - \xi < 1$. Now we bound Q from above.

To upper bound Q it suffices to find c such that $F^{-1}_{c\sqrt{m},m}(\beta) \geq F^{-1}_{0,m}(1-\xi)$. From Birgé (2001) we have

$$F^{-1}_{z,d}(u) \leq z + d + 2\sqrt{(2z+d)\log(1/(1-u))} + 2\log(1/(1-u)), \tag{77}$$

$$F^{-1}_{z,d}(u) \geq z + d - 2\sqrt{(2z+d)\log(1/u)}. \tag{78}$$

Hence,

$$F^{-1}_{c\sqrt{m},m}(\beta) \geq m + c\sqrt{m} - 2\sqrt{(2c\sqrt{m}+m)\log\frac{1}{\beta}}, \tag{79}$$

$$F^{-1}_{0,m}(1-\gamma) \leq m + 2\sqrt{m\log\frac{1}{\gamma}} + 2\log\frac{1}{\gamma}. \tag{80}$$

It suffices to find c that satisfies

$$m + c\sqrt{m} - 2\sqrt{(2c\sqrt{m}+m)\log\frac{1}{\beta}} \geq m + 2\sqrt{m\log\frac{1}{\gamma}} + 2\log\frac{1}{\gamma}, \tag{81}$$

or equivalently,

$$c \geq 2\sqrt{\left(\frac{c}{\sqrt{m}}+1\right)\log\frac{1}{\beta}} + 2\left(\sqrt{\log\frac{1}{\gamma}} + \log\frac{1}{\gamma}\right). \tag{82}$$

The right-hand side of the last inequality is largest when $m = 1$, and equality can be achieved when $m = 1$ at some $\Lambda(\beta, \xi)$ for any β, ξ satisfying the stated conditions. Equality can be achieved then for any m at some $Q(m, \beta, \xi) \leq \Lambda(\beta, \xi)$. This proves the first claim. The second claim follows immediately by inspection. □

PROOF OF LEMMA 12. We find a $P_0 \in \mathcal{F}_j$ and a measure μ supported on A such that $d_{\mathrm{TV}}(P_0, P_\mu) \leq 2\delta$. We then have, following Ingster (1993),

$$\beta \geq \inf_{\phi_\xi \in \Phi_\xi} P_\mu\{\phi_\xi = 0\} \tag{83}$$

$$\geq 1 - \xi - \sup_{R:\, P_0(R) \leq \xi} |P_0(R) - P_\mu(R)| \tag{84}$$

$$\geq 1 - \xi - \sup_R |P_0(R) - P_\mu(R)| \tag{85}$$

$$= 1 - \xi - \tfrac{1}{2} d_{\mathrm{TV}}(P_0, P_\mu) \geq 1 - \xi - \delta. \tag{86}$$

Let $\psi_1, \psi_2, \dots, \psi_n$ be an orthonormal basis for $\mathbb{R}^n$ such that $\psi_1, \dots, \psi_d$ form an orthonormal basis for $\mathcal{F}$. Fix $\tau > 0$ small and let $\lambda^2 = n\varepsilon^2/(n-d) + \tau^2/(n-d)$. Define

$$f_E = \lambda \sum_{s=d+1}^{m} E_s \psi_s, \tag{87}$$

where $(E_s : s = d+1, \dots, n)$ are independent Rademacher random variables, that is, $\mathbb{P}\{E_s = 1\} = \mathbb{P}\{E_s = -1\} = 1/2$. Now, $\Pi_{\mathcal{F}} f_E = 0$ and hence $\|f_E - \Pi_{\mathcal{F}} f_E\|^2 = \lambda^2 > \varepsilon^2$, and hence $f_E \in A$ for each choice of the Rademachers.

Let $P_\mu = \mathsf{E}(P_E)$ where P_E is the distribution under f_E and the expectation is with respect to the Rademachers. Choose $f_0 \in \mathcal{F}$ and let P_0 be the corresponding distribution. As in Baraud, we use the bound

$$d_{\mathrm{TV}}(P_\mu, P_0) \le \sqrt{\mathsf{E}_0\left(\frac{dP_\mu}{dP_0}(Y)\right)^2 - 1}. \tag{88}$$

We take $f_0 = (0, \dots, 0) \in \mathcal{F}$ and so

$$\left(\frac{dP_\mu}{dP_0}(Y)\right) = \mathsf{E}_E\left(\exp\left\{-\frac{1}{2}\lambda^2(n-d) + \lambda \sum_{s=d+1}^{n} E_s \sum_i Y_i \psi_{si}\right\}\right) \tag{89}$$

$$= e^{-\lambda^2/2} \prod_{s=d+1}^{n} \cosh(\lambda(Y \cdot \psi_s)). \tag{90}$$

Since $\mathsf{E}_0 \cosh^2(\lambda(Y \cdot \psi_j)) = e^{\lambda^2} \cosh(\lambda^2)$ and $\cosh(x) \le e^{x^2/2}$ we have

$$\mathsf{E}_0\left(\frac{dP_\mu}{dP_0}(Y)\right)^2 = (\cosh(\lambda^2))^{n-d} \le e^{(n-d)\lambda^4/2} \tag{91}$$

$$= \exp\left(\frac{n^2}{2(n-d)}\varepsilon^4 + \frac{\tau^4}{2(n-d)} + \frac{n}{n-d}\tau^2\varepsilon^2\right). \tag{92}$$

By the definition of ε (in terms of δ), $\beta \ge 1 - \xi - \delta + O(\tau)$, and because this holds for every τ, the result follows. $\square$

PROOF OF LEMMA 14. Let $f, g \in A$ be such that $\|f - g\|_p \le \varepsilon$. Then, $\mathbb{P}_g\{L \le f \le U\} = \mathbb{P}_f\{L \le f \le U\} + \mathbb{P}_g\{L \le f \le U\} - \mathbb{P}_f\{L \le f \le U\} \ge \mathbb{P}_f\{L \le f \le U\} - d_{\mathrm{TV}}(P_f, P_g) \ge 1 - \alpha - M_p(\|f - g\|_p, A) \ge 1 - \alpha - M_p(\varepsilon(f, p), A)$. We also have that $\mathbb{P}_g\{L \le g \le U\} \ge 1 - \alpha$. Hence, $\mathbb{P}_g\{L \le g \le U, L \le f \le U\} \ge \mathbb{P}_g\{L \le g \le U\} + \mathbb{P}_g\{L \le f \le U\} - 1 \ge 1 - \alpha + 1 - \alpha - M_p(\varepsilon(f, p), A) - 1 \ge 1 - 2\alpha - M_p(\varepsilon(f, p), A)$. The event $\{L \le g \le U, L \le f \le U\}$ implies that $W \ge \|g - f\|_\infty$. Hence, $\mathbb{P}_f\{W > \|f - g\|_\infty\} \ge 1 - 2\alpha - M_p(\varepsilon(f, p), A) \ge 1 - 2\alpha - M_p(\varepsilon(f, p), A) \ge 1 - 2\alpha - M_p(\varepsilon, A)$. It follows then that $\mathbb{P}_f\{W > \varepsilon(f, \infty)\} = \inf_g \mathbb{P}_f\{W > \|f - g\|_\infty\}$ and thus

$$\inf_{f \in A_0} \mathbb{P}_f\{W > \varepsilon(f, \infty)\} \ge 1 - 2\alpha - \sup_{f \in A_0} M_p(\varepsilon(f, p), A). \tag{93}$$

This proves the first claim. But $\varepsilon(f,\infty) \geq \varepsilon(f,p)$ for any $1 \leq p \leq \infty$. The final claim follows immediately. □

PROOF OF LEMMA 15. Choose $f \in A_0$. Choose $g \in A_1$ to minimize $d_{\mathrm{TV}}(p_f, p_g)$ such to such that $\|f-g\|_\infty = \varepsilon$. Hence, $d_{\mathrm{TV}}(p_f, p_g) = m_\infty(\varepsilon, A_0, A_1)$. Then, $\mathbb{P}_f\{L \leq g \leq U\} = \mathbb{P}_g\{L \leq g \leq U\} + \mathbb{P}_f\{L \leq g \leq U\} - \mathbb{P}_g\{L \leq g \leq U\} \geq \mathbb{P}_g\{L \leq g \leq U\} - d_{\mathrm{TV}}(P_f, P_g) \geq 1-\alpha - m_\infty(\varepsilon, A_0, A_1)$ because, by assumption. $\mathbb{P}_g\{L \leq g \leq U\} \geq 1-\alpha$. We also have that $\mathbb{P}_f\{L \leq f \leq U\} \geq 1-\alpha$. Hence, $\mathbb{P}_f\{L \leq f \leq U, L \leq g \leq U\} \geq \mathbb{P}_f\{L \leq f \leq U\} + \mathbb{P}_f\{L \leq g \leq U\} - 1 \geq 1-\alpha+1-\alpha - m_\infty(\varepsilon, A_0, A_1) \geq 1-2\alpha - m_\infty(\varepsilon, A_0, A_1)$. The event $\{L \leq f \leq U, L \leq g \leq U\}$ implies that $W \geq \|f-g\|_\infty$. Hence,

$$\mathbb{P}_f\{W > \|f-g\|_\infty\} \geq 1-2\alpha - m_\infty(\varepsilon, A_0, A_1). \tag{94}$$

It follows then that $\sup_{f\in A_0} \mathbb{P}_f\{W > \varepsilon\} \geq 1-2\alpha - m_\infty(\varepsilon, A_0, A_1)$. □

PROOF OF THEOREM 16. First, we compute $m_\infty(\varepsilon, \mathcal{F}, \mathcal{F})$. Note that for all $f \in \mathcal{F}$, $d_{\mathrm{TV}}(\mathbb{P}_f, \mathbb{P}_0) = \tau(\sqrt{n}\|f\|)$. Hence, $m_\infty(\varepsilon, \mathcal{F}, \mathcal{F}) = \tau(\sqrt{n}v)$ where $v = \min\{\|f\| : f \in \mathcal{F}, \|f\|_\infty = \varepsilon\}$. By Lemma 8, $v = \varepsilon/(\sqrt{n}\Omega_{\mathcal{F}})$. It follows by Lemma 15 that $\sup_{f\in\mathcal{F}} \mathbb{P}\{W > w\} \geq 1-2\alpha - \tau(\frac{w}{\sigma\Omega_{\mathcal{F}}})$. Let $w_* = \sigma\Omega\tau^{-1}(1-2\alpha-\gamma)$. It follows that if $w < w_*$ then $\inf_{f\in\mathcal{F}} \mathbb{P}\{W \leq w\} < 1-\gamma$ which is a contradiction.

That the proposed band has correct coverage follows easily. Now, $(\Pi\Pi^T)_{ii} \leq \Omega_{\mathcal{F}}$ and $z_{\alpha/2n} \leq \sqrt{c\log n}$ for some c and the claim follows. □

PROOF OF THEOREM 18. We break the argument up into three parts. Parts I and II taken together contribute the term v_0 from (41) to the bounds. The logic of both parts is the same: find a value w_* such that if $w < w_*$ then $\sup_{f\in\mathcal{F}} \mathbb{P}\{W > w\} > \gamma$ and, equivalently, $\inf_{f\in\mathcal{F}} \mathbb{P}\{W \leq w\} < 1-\gamma$, which gives a contradiction under the assumptions of the theorem. Part III contributes the term v_1 from (42) to the bounds. It is based on using the confidence bands to construct both an estimator and a test. Throughout the proof, we refer to the space $V \supset \mathcal{F}$ defined in (55); this is the set of spoilers that are within ε_2 of $\mathcal{F}$.

Part I. First, we compute $m_\infty(w, \mathcal{F}, \mathcal{F})$. Note that for all $f \in \mathcal{F}$, $d_{\mathrm{TV}}(\mathbb{P}_f, \mathbb{P}_0) = \tau(\sqrt{n}\|f\|/\sigma)$. Hence, $m_\infty(w, \mathcal{F}, \mathcal{F}) = \tau(\sqrt{n}v/\sigma)$ where $v = \min\{\|f\| : f \in \mathcal{F}, \|f\|_\infty = \varepsilon\}$. By Lemma 8, $v = w/(\sqrt{n}\Omega_{\mathcal{F}})$. It follows by Lemma 15 that

$$\sup_{f\in\mathcal{F}} \mathbb{P}\{W > w\} \geq 1-2\alpha - \tau\left(\frac{w}{\sigma\Omega_{\mathcal{F}}}\right). \tag{95}$$

Take $w_* = \sigma\Omega_{\mathcal{F}}\tau^{-1}(1-2\alpha-\gamma)$.

Part II. *Case* (a) $\varepsilon_2 \leq \varepsilon_\infty/\sqrt{n}$. First, note that $m_\infty(w, \mathcal{F}, V) = \tau(\sqrt{n}\frac{w}{\sigma\sqrt{n}}) = \tau(w/\sigma)$ for $w \leq \sqrt{n}\varepsilon_2$, because the minimum two-norm for a given infinity-norm is achieved on the coordinate axis. Second, let $A_0 = \mathcal{F}$ and $A_1 = V$ in Lemma 15. Then, for $w \le \sqrt{n}\varepsilon_2$,

$$\sup_{f\in\mathcal{F}} \mathbb{P}\{W > w\} \ge 1 - 2\alpha - \tau\left(\frac{w}{\sigma}\right). \tag{96}$$

Let $w_* = \sigma \min(\tau^{-1}(1-2\alpha-\gamma), \varepsilon_2\sqrt{n})$, then $\sup_{f\in\mathcal{F}} \mathbb{P}\{W > w_0\} \ge \gamma$.

Case (b) $\varepsilon_2 > \varepsilon_\infty/\sqrt{n}$. First, note that $m_\infty(w, \mathcal{F}, V) = \tau(\sqrt{n}\frac{w}{\sigma\sqrt{n}}) = \tau(w/\sigma)$ for $w \le \varepsilon_\infty$. Second, let $A_0 = \mathcal{F}$ and $A_1 = V$ in Lemma 15. Then, for $w \le \varepsilon_\infty$,

$$\sup_{f\in\mathcal{F}} \mathbb{P}\{W > w\} \ge 1 - 2\alpha - \tau\left(\frac{w}{\sigma}\right). \tag{97}$$

Let $w_* = \sigma \min(\tau^{-1}(1-2\alpha-\gamma), \varepsilon_\infty)$, then $\sup_{f\in\mathcal{F}} \mathbb{P}\{W > w_0\} \ge \gamma$.

Part III. The argument here is based on an argument in Baraud (2004). Let $\hat{f} = (U+L)/2$. Define a rejection region

$$\mathcal{R} = \{W > w\} \cup \left\{\|\hat{f} - \Pi\hat{f}\|_2 > \frac{W}{2}\right\}. \tag{98}$$

Now, for any $f \in \mathcal{F}$, $f^\star = f$, $\|\hat{f} - \Pi\hat{f}\|_2 \le \|\hat{f} - f\|_2$ and

$$\begin{aligned}
\mathbb{P}_f(\mathcal{R}) &\le \mathbb{P}_f\{W > w\} + \mathbb{P}_f\{\|\hat{f} - \Pi\hat{f}\|_2 > W/2\} && (99)\\
&\le \gamma + \mathbb{P}_f\{\|\hat{f} - \Pi\hat{f}\|_2 > W/2\} && (100)\\
&\le \gamma + \mathbb{P}_f\{\|f - \hat{f}\|_2 > W/2\} && (101)\\
&= \gamma + \mathbb{P}_f\{\|f^\star - \hat{f}\|_2 > W/2\} && (102)\\
&\le \gamma + \mathbb{P}_f\{\|f^\star - \hat{f}\|_\infty > W/2\} \le \gamma + \alpha && (103)
\end{aligned}$$

which bounds the type I error of $\mathcal{R}$.

Now let f be such that $\|f - \Pi f\| > \max\{w, \varepsilon_2\}$. Because $\|f - \Pi\hat{f}\| > \|f - \Pi f\|$, $\|f - \Pi f\| > \varepsilon_2$ implies that $f^\star = f$. And thus,

$$\|\hat{f} - \Pi\hat{f}\|_2 \ge \|f - \Pi\hat{f}\|_2 - \|f - \hat{f}\|_2 \ge w - \|f - \hat{f}\|_2. \tag{104}$$

Hence,

$$\begin{aligned}
\mathbb{P}_f(\mathcal{R}^c) &= \mathbb{P}_f\{\|\hat{f} - \Pi\hat{f}\|_2 \le W/2, W/2 \le w/2\} && (105)\\
&\le \mathbb{P}_f\{\|\hat{f} - \Pi\hat{f}\|_2 \le w/2, W \le w\} && (106)\\
&\le \mathbb{P}_f\{\|f - \hat{f}\|_2 \ge w/2, w \ge W\} && (107)\\
&\le \mathbb{P}_f\{\|f - \hat{f}\|_2 \ge W/2\} && (108)\\
&= \mathbb{P}_f\{\|f^\star - \hat{f}\|_2 \ge W/2\} && (109)\\
&\le \mathbb{P}_f\{\|f^\star - \hat{f}\|_\infty \ge W/2\} \le \alpha. && (110)
\end{aligned}$$

Thus, $\mathcal{R}$ defines a test for $H_0: f \in \mathcal{F}$ with level $\alpha + \gamma$ whose power more than a distance $\max\{w, \varepsilon_2\}$ from $\mathcal{F}$ is at least $1 - \alpha$. Using Lemma 12 with $\xi = \alpha + \gamma$ and $\delta = 1 - \gamma - 2\alpha$, this implies that $\max\{w, \varepsilon_2\} \geq 2\kappa(\alpha, \gamma)(n-d)^{1/4} n^{-1/2}$. The result follows. □

PROOF OF THEOREM 19. The volume is minimized by making ε_∞ as large as possible and ε_2 as small as possible. To achieve the lower bound on the width requires $\varepsilon_\infty \leq w_{\mathcal{F}}$ and $\varepsilon_2 \geq 2\kappa(\alpha, \gamma)(n-d)^{1/4} n^{-1/2}$. □

PROOF OF THEOREM 20. Let $A = \{T \leq \chi^2_{\gamma, n-d}\}$. Then, $\mathbb{P}_f\{f^\star \notin B\} = \mathbb{P}_f\{f^\star \notin B, A\} + \mathbb{P}_f\{f^\star \notin B, A^c\}$. We claim that $\mathbb{P}_f\{f^\star \notin B, A\} \leq \alpha/2$ and $\mathbb{P}_f\{f^\star \notin B, A^c\} \leq \alpha/2$. There are four cases.

Case I. $f \in \mathcal{F}$. Then $f = f^\star$ and $\mathbb{P}_f\{f \notin B, A^c\} \leq \mathbb{P}_f\{A^c\} \leq \alpha/2$. $\mathbb{P}_f\{f \notin B, A\} \leq \mathbb{P}_f\{f \notin B\} = \mathbb{P}_{\Pi f}\{\Pi f \notin B\} \leq \mathbb{P}_{\Pi f}\{\|\hat{f} - \Pi f\|_\infty > w_{\mathcal{F}}\} \leq \alpha/2$.

Case II. $f \in V - \mathcal{F}$ where $V = \{f : \|f - \Pi f\| \leq \varepsilon_2, \|f - \Pi f\|_\infty \leq \varepsilon_\varepsilon\}$. Again, $f = f^\star$. First, $\mathbb{P}_f\{f \notin B, A^c\} \leq \mathbb{P}_f\{\|Y - f\|_\infty > z_{\alpha/2n}\} \leq \alpha/2$. Next, we bound $\mathbb{P}_f\{f \notin B, A\}$. Note that $\hat{f} = \Pi Y \sim N(g, \sigma^2 \Pi \Pi^T)$, where $g = \Pi f$. Then $\hat{f}_i \sim N(g_i, \Omega_i^2)$. Let $B_0 = (L + \varepsilon_\infty, U - \varepsilon_\infty)$. Then, $\Pi f \in B_0$ implies $f \in B$ and $\mathbb{P}_f\{\notin B, A\} \leq \mathbb{P}_f\{\Pi f \notin B_0\} \leq \alpha/2$.

Case III. $f \notin V$, $\|f - \Pi f\| \leq \varepsilon_2$ and $\|f - \Pi f\|_\infty > \varepsilon_\infty$. In this case, $f^\star = \Pi f$. Then $\mathbb{P}_f\{f^\star, f \in B^c, A^c\} \leq \mathbb{P}_f\{f \in B^c, A^c\} \leq \alpha/2$. Also, $\mathbb{P}_f\{f^\star, f \in B^c, A\} \leq \mathbb{P}_f\{f^\star \notin B\} = \mathbb{P}_{\Pi f}\{\Pi f \notin B\} \leq \mathbb{P}_{\Pi f}\{\|\hat{f} - \Pi f\|_\infty > w_{\mathcal{F}}\} \leq \alpha/2$.

Case IV. $f \notin V$ and $\|f - \Pi f\| > \varepsilon_2$. In this case, $f^\star = f$. But

$$\mathbb{P}_f\{f \notin B, A\} \leq \mathbb{P}_f\{A\} \leq F_{f - \Pi f, n-d}(\chi^2_{\gamma, n-d}) \leq F_{\varepsilon_2, n-d}(\chi^2_{\gamma, n-d}) \leq \alpha/2$$

and

$$\mathbb{P}_f\{f \notin B, A^c\} \leq \mathbb{P}_f\{f \notin B, A^c\} \leq \alpha/2.$$

Thus, $\mathbb{P}_f\{f^\star \notin B\} \leq \alpha$. Equation (52) follows since $\mathbb{P}_f\{T \leq \chi^2_{\gamma, n-d}\} \geq 1 - \gamma$ for all $f \in \mathcal{F}$. □

PROOF OF LEMMA 24. First note that if B is a ball in $\mathbb{R}^n$ in any norm, then $B - B = 2B$. Second, we have that

$$\omega(u) = \sup\{|Tg| : \|g\|_2 \leq u, g \in V - V\} \tag{111}$$

$$= \sup\{|Tg| : \|g\|_2 \leq u, g \in V(2\varepsilon_2, 2\varepsilon_\infty)\}. \tag{112}$$

To see the latter equality, note that if $g, h \in V$, then we can write $g - h = f + \delta_1 - \delta_2$ where $f \in \mathcal{F}$ and δ_i are in $B_k^\perp(0, \varepsilon_k)$ for $k = 2, \infty$. Thus, $\delta_1 - \delta_2$ is in $2B_2^\perp(0, \varepsilon_2) \cap 2B_\infty^\perp(0, \varepsilon_\infty)$.

Set $B^*(f) = B_2^\perp(f, 2\varepsilon_2) \cap B_\infty^\perp(f, 2\varepsilon_\infty)$. We have that

$$\omega(\eta, \mathcal{F}) = \sup\{f_1 : \|f\|_2 \leq \eta, f \in \mathcal{F}\}, \tag{113}$$

$$\omega(\eta, B^*(0)) = \sup\{f_1 : \|f\|_2 \leq \eta, f \in B^*(0)\}. \tag{114}$$

For any $g \in V(2\varepsilon_2, 2\varepsilon_\infty)$, we can write $g = g_1 + g_2$ where $g_1 \in \mathcal{F}$ and $g_2 \in B^*(0)$ and the two functions are orthogonal. Let $A = \{g_1, g_2 : \|g_1\|_2 \le \sqrt{u^2 - c^2}, \|g_2\|_2 \le c^2, g_1 \in \mathcal{F}, g_2 \in B^*(0)\}$. Then,

$$w(u, V) = \sup\{T(g) : g \in V(2\varepsilon_2, 2\varepsilon_\infty), \|g\|_2 \le u\} \tag{115}$$

$$= \sup_{0 \le c \le u} \{T(g_1 + g_2) : g_1, g_2 \in A\} \tag{116}$$

$$\le \sup_{0 \le c \le u} \Bigg[\sup_{\substack{g_1 \in \mathcal{F} \\ \|g_1\|_2 \le \sqrt{u^2 - c^2}}} T(g_1) + \sup_{\substack{g_2 \in B^*(0) \\ \|g_2\|_2 \le c}} T(g_2) \Bigg] \tag{117}$$

$$= \sup_{0 \le c \le u} \big[\omega\big(\sqrt{u^2 - c^2}, \mathcal{F}\big) + \omega(c, B^*(0))\big]. \tag{118}$$

Moreover, equality can be attained for each c by choosing g_1 and g_2 to be the maximizers (or suitably close approximants thereof) of each term in the last equation. Consequently,

$$\omega(u) = \sup_{0 \le c \le u} \omega\big(\sqrt{u^2 - c^2}, \mathcal{F}\big) + \omega(c, B^*(0)). \tag{119}$$

To derive $\omega(\eta, B^*(0))$, note that $f = ((\eta \wedge \varepsilon_2)\sqrt{n} \wedge \varepsilon_\infty, 0, 0, \ldots, 0)$ maximizes f_1 subject to the norm constraint. Hence, $\omega(\eta, B^*(0)) = \min((\eta \wedge \varepsilon_2)\sqrt{n}, \varepsilon_\infty)$. For $\omega(\eta, \mathcal{F})$, let $e = (1, 0, \ldots, 0) \in \mathbb{R}^n$. Recall that $\Omega_{\mathcal{F}} = \frac{\langle e, \Pi_{\mathcal{F}} e\rangle}{\|e\| \, \|\Pi_{\mathcal{F}} e\|} = \frac{\|\Pi_{\mathcal{F}} e\|}{\|e\|}$, which is between 0 and 1. Maximizing $e^T f$ for $f \in \mathcal{F}$ and $\|f\|_2 \le \eta$ is equivalent to maximizing $n\langle e, f\rangle = n\langle \Pi_{\mathcal{F}} e, f\rangle$. The maximum subject to the constraint occurs at $f^\star = \eta \Pi e / \|\Pi e\|$ Hence, $\omega(\eta, \mathcal{F}) = \eta\sqrt{n}\Omega_{\mathcal{F}}$. Note that η is in terms of the normalized two norm; in the "natural" (root sum of squares) norm, the modulus would be $\omega_\natural(u, \mathcal{F}) = u\Omega_{\mathcal{F}}$.

It follows that

$$\omega(u, V) = \sup_{0 \le c \le u} \big[\omega\big(\sqrt{u^2 - c^2}, \mathcal{F}\big) + \omega(c, B^*(0))\big] \tag{120}$$

$$= \sup_{0 \le c \le u} \big[\sqrt{n}\Omega_{\mathcal{F}}\sqrt{u^2 - c^2} + \min\big((c \wedge \varepsilon_2)\sqrt{n}, \varepsilon_\infty\big)\big] \tag{121}$$

$$= \sqrt{n} \sup_{0 \le c \le u} \big[\Omega_{\mathcal{F}}\sqrt{u^2 - c^2} + \min\big(c, \varepsilon_2 \wedge \big(\varepsilon_\infty/\sqrt{n}\big)\big)\big] \tag{122}$$

$$= \sqrt{n}\bigg(u\Omega\sqrt{\frac{\Omega^2}{1 + \Omega^2}} + \min\bigg(\frac{u}{\sqrt{1 + \Omega^2}}, \varepsilon_2 \wedge \big(\varepsilon_\infty/\sqrt{n}\big)\bigg)\bigg) \tag{123}$$

$$= \bigg(u\sqrt{n}\Omega\sqrt{\frac{\Omega^2}{1 + \Omega^2}} + \min\bigg(\frac{u\sqrt{n}}{\sqrt{1 + \Omega^2}}, \varepsilon_2\sqrt{n}, \varepsilon_\infty\bigg)\bigg) \tag{124}$$

because the supremum over c is maximized at $c = u/(1+\Omega^2)$. In the natural two norm, we have

$$\omega_\natural(u, V) = \left(u\Omega\sqrt{\frac{\Omega^2}{1+\Omega^2}} + \min\left(\frac{u}{\Omega}\sqrt{\frac{\Omega^2}{1+\Omega^2}}, \varepsilon_{2,\natural}, \varepsilon_\infty\right)\right). \tag{125}$$ □

Next, we prove the lower bound result generalized to a nested sequence of subspaces. To do so, we need to prove several auxiliary lemmas. Define for each $1 \le j \le m$,

$$U_j = \{f \in \mathbb{R}^n : F^\star(f) = \{\Pi_j f, f\} \text{ or } F^\star(f) = \{f\}\}. \tag{126}$$

Referring to the definition of V in equation (55), define here $V_j = V(\mathcal{F}_j, \varepsilon_{2,j}, \varepsilon_{\infty,j})$.

LEMMA 33. *Let $w > 0$. Then,*

$$m_\infty(w, \mathcal{F}_j \cap U_j, \mathcal{F}_j \cap U_j) = m_\infty(w, \mathcal{F}_j, \mathcal{F}_j), \tag{127}$$

$$m_\infty(w, \mathcal{F}_j \cap U_j, V_j \cap U_j) = m_\infty(w, \mathcal{F}_j, V_j). \tag{128}$$

PROOF. First, let $f, g \in \mathcal{F}_j$ be the minimal pair for $m_\infty(w, \mathcal{F}_j, \mathcal{F}_j)$. Let ψ be a unit-2-norm vector in $\mathcal{F}_j \cap \mathcal{F}_{j-1}^\perp$. Let $\lambda > \varepsilon_{2,1}$ and define $\tilde{f} = \lambda\psi + f$, $\tilde{g} = \lambda\psi + g$. Then, $\tilde{f}, \tilde{g} \in \mathcal{F}_j \cap U_j$ because if either f or g were in $\mathcal{F}_j \cap U_j^c$ then adding $\lambda\psi$ makes the distance from the projection on one of the lower spaces larger than the corresponding ε_2. Also $d_{\mathrm{TV}}(P_{\tilde{f}}, P_{\tilde{g}}) = d_{\mathrm{TV}}(P_f, P_g)$ and $\|\tilde{f} - \tilde{g}\|_\infty = \|f - g\|_\infty$. Hence, $m_\infty(w, \mathcal{F}_j \cap U_j, \mathcal{F}_j \cap U_j) \le m_\infty(w, \mathcal{F}_j, \mathcal{F}_j)$. But $\mathcal{F}_j \cap U_j \subset \mathcal{F}_j$, so $m_\infty(w, \mathcal{F}_j \cap U_j, \mathcal{F}_j \cap U_j) = m_\infty(w, \mathcal{F}_j, \mathcal{F}_j)$ as was to be proved.

Second,let $f \in \mathcal{F}_j$ and $g \in V_j$ be the minimal pair for $m_\infty(w, \mathcal{F}_j, V_j)$. Now apply the same argument. □

LEMMA 34. *Let $0 < \delta < 1 - \xi$ and*

$$\varepsilon = \frac{(n - d_j)^{1/4}}{\sqrt{n}}\left(2\log(1 + 4\delta^2)\right)^{1/4}. \tag{129}$$

Define $A_j = U_j \cap \{f : \|f - \Pi_j f\| > \varepsilon\}$. Then,

$$\beta \equiv \inf_{\phi_\alpha \in \Phi_\xi} \sup_{f \in A_j} \mathbb{P}_f\{\phi_\xi = 0\} \ge 1 - \xi - \delta, \tag{130}$$

where

$$\Phi_\xi = \left\{\phi_\xi : \sup_{f \in \mathcal{F}_j} \mathbb{P}_f\{\phi_\xi = 0\} \le \xi\right\} \tag{131}$$

is the set of level ξ tests.

PROOF. Let f_E be defined as in equation (87) in the proof of Lemma 12. Let ψ be a unit vector in $\mathcal{F}_{j+1} \cap \mathcal{F}_j^{\perp}$ and let $\lambda > \varepsilon_{2,1}$. Then, define $\tilde{f}_E = \lambda\psi + f_E$. Now apply the proof of Lemma 12 using $f_0 = \lambda\psi$ instead of 0. The total variation distances among corners of the hypercube do not change and the result follows. $\square$

LEMMA 35. *Fix* $0 < \alpha < 1$ *and* $0 < \gamma < 1 - 2\alpha$. *Suppose that for bands* $B = (L, U)$

$$\inf_{f \in U_j} \mathbb{P}_f\{F^\star(f) \cap B \neq \varnothing\} \geq 1 - \alpha. \tag{132}$$

Then

$$\inf_{f \in \mathcal{F}_j} \mathbb{P}_f\{W \leq w\} \geq 1 - \gamma, \tag{133}$$

implies

$$w \geq \underline{w}(\mathcal{F}_j, \varepsilon_{2,j}, \varepsilon_{\infty,j}, n, d_j, \alpha, \gamma, \sigma), \tag{134}$$

where $\underline{w}$ *is given in Theorem* 18.

PROOF. To prove this lemma, we will adapt the proof of Theorem 18 as follows. By Lemma 33, the argument for parts I and II is the same with $\mathcal{F}$ replaced with $\mathcal{F}_j \cap U_j$ and V replaced with $V_j \cap U_j$. By replacing the reference to Lemma 12 with Lemma 34, the argument for Part III also follows exactly. The result follows. $\square$

PROOF OF THEOREM 27. The result follows directly from Lemma 35 because $\inf_{f \in \mathbb{R}^n} \mathbb{P}\{F^\star(f) \cap B \neq \varnothing\} \geq 1 - \alpha$ implies $\inf_{f \in U_j} \mathbb{P}\{F^\star(f) \cap B \neq \varnothing\} \geq 1 - \alpha$. $\square$

PROOF OF THEOREM 29. Note that $\mathbb{P}_f\{F^\star \cap B = \varnothing\} = \sum_j \mathbb{P}_f\{F^\star \cap B = \varnothing, \hat{J} = j\}$. We show that $\mathbb{P}_f\{F^\star \cap B = \varnothing, \hat{J} = j\} \leq \alpha_j$ for each j. There are three cases. Throughout the proof, we take $\sigma = 1$.

Case I. $\|f - \Pi_j f\| > \varepsilon_{2,j}$. Then,

$$\begin{aligned}
\mathbb{P}_f\{F^\star \cap B = \varnothing, \hat{J} = j\} \leq \mathbb{P}_f\{\hat{J} = j\} &\leq F_{f - \Pi_j f, n-d_j}(\chi^2_{\gamma, n-d_j}) \\
&\leq F_{\varepsilon_{2,j}, n-d_j}(\chi^2_{\gamma, n-d_j}) \\
&\leq \alpha_j
\end{aligned}$$

due to (68).

Case II. $\|f - \Pi_j f\| \le \varepsilon_{2,j}$ and $\|f - \Pi_j f\|_\infty \le \varepsilon_{\infty,j}$. So,

$$\begin{aligned}
\mathbb{P}_f\{F^\star \cap B = \varnothing, \hat{J} = j\} &\le \mathbb{P}_f\{f \notin B, \hat{J} = j\} \\
&\le \mathbb{P}_f\{\|f - \hat{f}\|_\infty > w_{\mathcal{F}_j} + \varepsilon_{\infty,j}\} \\
&\le \mathbb{P}_f\{\|f - \Pi_j f\|_\infty + \|\Pi_j f - \Pi_j Y\|_\infty > w_{\mathcal{F}_j} + \varepsilon_{\infty,j}\} \\
&\le \mathbb{P}_f\{\|\Pi_j f - \Pi_j Y\|_\infty > w_{\mathcal{F}_j}\} \\
&= \mathbb{P}_{\Pi_j f}\{\|\Pi_j f - \Pi_j Y\|_\infty > w_{\mathcal{F}_j}\} \\
&\le \alpha_j.
\end{aligned}$$

Case III. $\|f - \Pi_j f\| \le \varepsilon_{2,j}$ and $\|f - \Pi_j f\|_\infty > \varepsilon_{\infty,j}$. Now,

$$\begin{aligned}
\mathbb{P}_f\{F^\star \cap B = \varnothing, \hat{J} = j\} &\le \mathbb{P}_f\{\Pi_j f \notin B, \hat{J} = j\} \\
&= \mathbb{P}_f\{\|\Pi_j Y - \Pi_j f\|_\infty > c_j, \hat{J} = j\} \\
&\le \mathbb{P}_f\{\|\Pi_j Y - \Pi_j f\|_\infty > c_j\} \\
&= \mathbb{P}_{\Pi_j f}\{\|\Pi_j Y - \Pi_j f\|_\infty > c_j\} \\
&\le \alpha_j.
\end{aligned}$$

To prove (70), suppose that $f \in \mathcal{F}_j$. Then, $\mathbb{P}_f\{\hat{J} > j\} \le \gamma$. But, as long as $\hat{J} \le j$, $W = w_{\hat{\mathcal{J}}}(\alpha_{\hat{J}}) + \varepsilon_{\infty,\hat{J}} \le w_j(\alpha_j) + \varepsilon_{\infty,j}$. The last statement follows since, when $\varepsilon_{2,j} \ge Q(n - d_j, \alpha/2, \gamma)(n - d_j)^{1/4} n^{-1/2}$. $\square$

5. Discussion. We have shown that adaptive confidence bands for f are possible if coverage is replaced by surrogate coverage. We focused on projection surrogates but there are other classes of surrogates that could be defined, for example, based on wavelet shrinkage or kernel smoothing.

Our results apply to nested subspaces. The nonnested cases is more complicated, and we suspect that extension to this case will require something akin to the between-class modulus of continuity defined in Cai and Low (2004).

Let us also mention average coverage [Wahba (1983) and Cummins, Filloon and Nychka (2001)]. Bands (L, U) have average coverage if $\mathbb{P}_f\{L(\xi) \le f(\xi) \le U(\xi)\} \ge 1 - \alpha$ where $\xi \sim \text{Uniform}(0, 1)$. A way to combine average coverage with the surrogate idea is to enforce something stronger than average coverage such as

$$\mathbb{P}_f\{L(\xi) \le f(\xi) \le U(\xi) \text{ and } \hat{f} \preceq f\} \ge 1 - \alpha,$$

where $\hat{f} = (L + U)/2$ and $\hat{f} \preceq f$ means that $\hat{f}$ is simpler than f according to a partial order $\preceq$, for example, $f \preceq g$ if $\int (f'')^2 \le \int (g'')^2$.

Acknowledgments. The authors thank the Associate Editor and two referees for their comments.

REFERENCES

ANDREWS, D. W. K. and GUGGENBERGER, P. (2007). Hybrid and size-corrected subsample methods. Discussion Paper 1606, Cowles Foundation, Yale Univ.

BARAUD, Y. (2002). Non-asymptotic minimax rates of testing in signal detection. *Bernoulli* **8** 577–606. MR1935648

BARAUD, Y. (2004). Confidence balls in Gaussian regression. *Ann. Statist.* **32** 528–551. MR2060168

BERAN, R. and DÜMBGEN, L. (1998). Modulation of estimators and confidence sets. *Ann. Statist.* **26** 1826–1856. MR1673280

BICKEL, P. J. and RITOV, Y. (2000). Non- and semiparametric statistics: Compared and contrasted. *J. Statist. Plann. Inference* **91** 209–228. MR1814781

BIRGÉ, L. (2001). An alternative point of view on Lepski's method. In *State of the Art in Probability and Statistics* (M. de Gunst, C. Klaassen and A. van der Vaart, eds.) 113–133. IMS, Beachwood, OH. MR1836557

CAI, T. and LOW, M. (2006). Adaptive confidence balls. *Ann. Statist.* **34** 202–228. MR2275240

CAI, T. and LOW, M. G. (2004). An adaptation theory for nonparametric confidence intervals. *Ann. Statist.* **32** 1805–1840. MR2102494

CHAUDHURI, P. and MARRON, J. S. (2000). Scale space view of curve estimation. *Ann. Statist.* **28** 408–428. MR1790003

CLAESKENS, G. and VAN KEILEGOM, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Ann. Statist.* **31** 1852–1884. MR2036392

CUMMINS, D., FILLOON, T. and NYCHKA, D. (2001). Confidence intervals for nonparametric curve estimates: Toward more uniform pointwise coverage. *J. Amer. Statist. Assoc.* **96** 233–246. MR1952734

DAVIES, P. L. and KOVAC, A. (2001). Local extremes, runs, strings and multiresolution (with discussion). *Ann. Statist.* **29** 1–65. MR1833958

DONOHO, D. (1988). One-sided inference about functionals of a density. *Ann. Statist.* **16** 1390–1420. MR0964930

DONOHO, D. (1995). De-noising by soft-thresholding. *IEEE Trans. Inform. Theory* **41** 613–627. MR1331258

DONOHO, D. and LIU, R. (1991). Geometrizing rates of convergence. II. *Ann. Statist.* **19** 633–667. MR1105839

DONOHO, D., JOHNSTONE, I. M., KERKYACHARIAN, G. and PICARD, D. (1995). Wavelet shrinkage: Asymptopia. *J. Roy. Statist. Soc. Ser. B* **57** 301–369. MR1323344

EUBANK, R. L. and SPECKMAN, P. L. (1993). Confidence bands in nonparametric regression. *J. Amer. Statist. Assoc.* **88** 1287–1301. MR1245362

GENOVESE, C. and WASSERMAN, L. (2005). Nonparametric confidence sets for wavelet regression. *Ann. Statist.* **33** 698–729. MR2163157

HALL, P. and TITTERINGTON, M. (1988). On confidence bands in nonparametric density estimation and regression. *J. Multivariate Anal.* **27** 228–254. MR0971184

HÄRDLE, W. and BOWMAN, A. W. (1988). Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands. *J. Amer. Statist. Assoc.* **83** 102–110. MR0941002

HÄRDLE, W. and MARRON, J. S. (1991). Bootstrap simultaneous error bars for nonparametric regression. *Ann. Statist.* **19** 778–796. MR1105844

HOFFMAN, O. and LEPSKI, M. (2002). Random rates in anisotropic regression (with discussion). *Ann. Statist.* **30** 325–396. MR1902892

INGSTER, Y. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives, I and II. *Math. Methods Statist.* **2** 85–114. MR1257978

INGSTER, Y. and SUSLINA, I. (2003). *Nonparametric Goodness of Fit Testing Under Gaussian Models.* Springer, New York. MR1991446

JUDITSKY, A. and LAMBERT-LACROIX, S. (2003). Nonparametric confidence set estimation. *Math. Methods Statist.* **19** 410–428. MR2054156

KABAILA, P. (1995). The effects of model selection on confidence regions and prediction regions. *Econometric Theory* **11** 537–549. MR1349934

KABAILA, P. (1998). Valid confidence intervals in regression after variable selection. *Econometric Theory* **14** 463–482. MR1650037

LEEB, H. and PÖTSCHER, B. M. (2005). Model selection and inference: Facts and fiction. *Econometric Theory* **21** 21–59. MR2153856

LI, K.-C. (1989). Honest confidence regions for nonparametric regression. *Ann. Statist.* **17** 1001–1008. MR1015135

LOW, M. G. (1997). On nonparametric confidence intervals. *Ann. Statist.* **25** 2547–2554. MR1604412

NEUMANN, M. H. and POLZEHL, J. (1998). Simultaneous bootstrap confidence bands in nonparametric regression. *J. Nonparametr. Statist.* **9** 307–333. MR1646905

ROBINS, J. and VAN DER VAART, A. (2006). Adaptive nonparametric confidence sets. *Ann. Statist.* **34** 229–253. MR2275241

RUPPERT, D., WAND, M. P. and CARROLL, R. J. (2003). *Semiparametric Regression*. Cambridge Univ. Press. MR1998720

SUN, J. and LOADER, C. R. (1994). Simultaneous confidence bands for linear regression and smoothing. *Ann. Statist.* **22** 1328–1345. MR1311978

TERRELL, G. R. and SCOTT, D. W. (1985). Oversmoothed nonparametric density estimates. *J. Amer. Statist. Assoc.* **80** 209–214. MR0786608

TERRELL, G. R. (1990). The maximal smoothing principle in density estimation. *J. Amer. Statist. Assoc.* **85** 470–477. MR1141750

WAHBA, G. (1983). Bayesian "confidence intervals" for the cross-validated smoothing spline. *J. Roy. Statist. Soc. Ser. B Methodol.* **45** 133–150. MR0701084

XIA, Y. (1998). Bias-corrected confidence bands in nonparametric regression. *J. Roy. Statist. Soc. Ser. B* **60** 797–811. MR1649488

DEPARTMENT OF STATISTICS
CARNEGIE MELLON UNIVERSITY
PITTSBURGH, PENNSYLVANIA 15213
USA
E-MAIL: genovese@stat.cmu.edu
larry@stat.cmu.edu