Spatially Adaptive Density Estimation by Localised Haar Projections

Florian Gach*
Richard Nickl*
Vladimir Spokoiny**

* University of Cambridge, United Kingdom
** Weierstrass Institute (WIAS) Berlin, Germany

BERLIN

ECONOMIC RISK

SFB 649

Spatially Adaptive Density Estimation
by Localised Haar Projections

Florian Gach, Richard Nickl, and Vladimir Spokoiny [*]

University of Cambridge [†] and *Weierstrass Institute Berlin* [‡]

March 2011

Abstract

Given a random sample from some unknown density $f_0 : \mathbb{R} \to [0, \infty)$ we devise Haar wavelet estimators for f_0 with variable resolution levels constructed from localised test procedures (as in Lepski, Mammen, and Spokoiny (1997, Ann. Statist.)). We show that these estimators adapt to spatially heterogeneous smoothness of f_0, simultaneously for every point x in a fixed interval, in sup-norm loss. The thresholding constants involved in the test procedures can be chosen in practice under the idealised assumption that the true density is locally constant in a neighborhood of the point x of estimation, and an information theoretic justification of this practice is given.

Keywords: spatially inhomogeneous smoothness, bandwidth choice, propagation approach

JEL-Classification: C14

1 Introduction

One of the most enduring challenges in statistical function estimation is to devise procedures that adapt to the locally variable complexity of the unknown function. For

example, if one observes a random sample $X_1, ..., X_n$ with density $f_0 : \mathbb{R} \to \mathbb{R}$, then f_0 may exhibit *spatially inhomogeneous smoothness*: The density could be infinitely-differentiable on most of its support except for a few points x_m where it behaves locally like $|x - x_m|^{\alpha_m}$ for some distinct numbers α_m. The location of the irregular points x_m will usually not be known, and neither the corresponding degree of smoothness α_m. Moreover f_0 could possess a so-called *multifractal* behavior, changing its Hölder exponents continuously on its domain of definition – in fact, as shown in Jaffard [11], 'typical' functions in the Besov spaces usually considered in nonparametric statistics are always multifractal. Donoho and Johnstone [1] and Donoho, Johnstone, Kerkyacharian, and Picard [2], [3] have suggested that methods based on wavelet shrinkage can, to a certain extent, adapt to spatially inhomogeneous complexity of the unknown function f_0. Moreover, Lepski, Mammen, and Spokoiny [12] showed that this is not intrinsic to wavelet methods, and that similar spatial adaptation results can be proved for kernel methods based on locally variable bandwidth choices.

There are several ways in which one can measure spatial adaptivity of an estimator. A minimal requirement may be to devise a rule $\hat{f}_n(x)$ that estimates $f_0(x)$ in an optimal way at every point x, and the methods suggested in [1] and [12] meet this requirement. These procedures depend on the point x, and the natural question arises as to how a given procedure performs globally as an estimator for f_0. To address this question, Donoho et al. [3] and Lepski et al. [12] considered global L^r-loss, $r < \infty$, and argued that taking L^r-loss over Besov-bodies $B(s, p, q)$ where smoothness is measured in L^p, $r > p$, gives a way to assess the spatial performance of an estimator. A probably more transparent approach to the spatial adaptation problem is to consider sup-norm loss for estimators with locally variable bandwidths: one aims to find an estimator $\hat{f}_n(x)$ that is locally optimal for estimating $f_0(x)$, and simultaneously so for all x. This approach was not considered in the literature so far – the results [6], [7], [8], [9] address the spatially homogeneous setting only.

A first contribution of this article is to show that a dyadic histogram estimator with variable bin size spatially adapts to possibly inhomogeneous local Hölder smoothness of f_0, in global sup-norm loss. More precisely, for $K(x, y)$ the Haar wavelet projection kernel, we shall construct

$$\hat{f}_n(x) = \frac{2^{\hat{j}_n(x)}}{n} \sum_{i=1}^{n} K(2^{\hat{j}_n(x)}x, 2^{\hat{j}_n(x)}X_i),$$

where $\hat{j}_n(x)$ is a variable resolution level that depends both on x and the sample, and

show that the random variable

$$\sup_x \frac{1}{r(n, x, f_0)} \left| \hat{f}_n(x) - f_0(x) \right|$$

is stochastically bounded, where $r(n, x, f_0)$ is the local minimax adaptive rate of estimation of f_0 at the point x. In fact we prove an explicit finite-sample oracle inequality that compares the estimator $\hat{f}_n$ to an optimal (oracle-type) estimator that is described in detail below.

While this result shows that spatial adaptation is indeed possible in a strong theoretical way, a drawback shared by most results in the literature on adaptive estimation remains: The theoretical findings give no indication whatsoever as to how to choose the numerical constants in the thresholds that feature in shrinkage- or Lepski-test-based methods. It has become a common practice that thresholding constants are chosen according to simulation results where simulations are drawn as if the true underlying signal is very simple (say, uniform or piecewise constant). This practice has not had any general theoretical corroboration until recently Spokoiny and Vial [14] gave, in a simple Gaussian regression model, a certain justification based on the idea of 'propagation'. The results in [14] are heavily tied to the simplicity of the model used, in particular to the strong Gaussianity assumption employed, and to the fact that pointwise loss is considered. In the present paper we show how the ideas of [14] generalise, subject to some nontrivial modifications, to nonparametric density estimation. A key idea in the proofs in [14], translated into the density estimation context, is to replace the sampling distribution by a locally constant product measure. The 'transportation cost' of this replacement is easy to control in the Gaussian setting of [14], but in the density estimation case the fluctuations of the likelihood ratios between the unknown sampling distribution and relevant locally constant product measures do not obey a Gaussian regime, but turn out to be of Poisson type, so that the 'Gaussian intuitions' of [14] could be entirely misleading. We show however that the main information theoretic idea of [14] remains sound in this Poissonian setting as well: We use a Lepski-type procedure to construct $\hat{j}_n(x)$, and we show that if we compute sharp thresholds for this procedure as if the true density f_0 belonged to a family $\mathcal{F}$ of locally constant densities, then the resulting estimator is spatially adaptive in sup-norm loss. In contrast to the results in [14], the rates of convergence we obtain for the risk of the final estimator are *exact* rate-adaptive.

While the techniques and results of this paper generalise in principle to more complex estimation problems that involve in particular adaptation to higher degrees of smoothness, we prefer to stay within the simpler setting of Haar wavelets, which allows for a

3

clean exposition of the main ideas.

2 Uniform spatial adaptation using propagation methods

We will use the symbol $\|g\|_T$ to denote the supremum $\sup_{t \in T} |g(t)|$ of a function g over some set T, but we will still use the symbol $\|g\|_\infty$ to denote $\sup_{x \in \mathbb{R}} |g(x)|$ if no confusion can arise.

For any $j \in \mathbb{N}$, we define a dyadic partition of $(0,1]$ into 2^j-many disjoint subintervals by setting $I_{j,k} = (k2^{-j}, (k+1)2^{-j}]$, $k = 0, \ldots, 2^j - 1$; and for $0 < x \le 1$ we denote by $I_{j,k(x)}$ the unique interval containing x. For $j \in \mathbb{N}$, $k = 1, \ldots, 2^j - 1$, let $V_{j,k}$ be the space of all bounded density functions on $\mathbb{R}$ that are constant on $I_{j,k}$. Via the local projections

$$K_{j,x}(f)(z) := \begin{cases} 2^j \int_{I_{j,k(x)}} f(y)dy & \text{if } z \in I_{j,k(x)}, \\ f(z) & \text{otherwise,} \end{cases}$$

we map any bounded density f onto $V_{j,k(x)}$. (Note that $K_{j,x}(f)$ is indeed a density since $K_{j,x}(f)$ and f assign the same probability to the interval $I_{j,k(x)}$.) For $f \in V_{j,k}$ and $j' \ge j$ we clearly have $K_{j',x}(f) = f$.

2.1 Estimation procedure

Let $X, X_1, \ldots, X_n$ be i.i.d. with bounded density $f_0 : \mathbb{R} \to [0, \infty)$, $n > 1$. We wish to construct a single estimator which estimates $f_0(x)$ in an optimal way, uniformly so for points x in the interval $(a, b]$. We shall take without loss of generality $(a, b] = (0, 1]$, and we shall assume throughout that f_0 is bounded away from zero on $(0, 1]$. Let $K(x, y) = \sum_k \phi(x - k)\phi(y - k)$ be the projection kernel based on the Haar wavelet $\phi = 1_{(0,1]}$. We shall write $K_j(x, y) = 2^j K(2^j x, 2^j y)$, and the associated linear density estimator is the dyadic histogram estimator given by

$$f_n(j, x) := \frac{1}{n} \sum_{i=1}^{n} K_j(x, X_i).$$

We make the important observation that $E_f f_n(j, x) = 2^j P_f(I_{j,k(x)})$, which directly follows from the identity $K_j(x, y) = 2^j 1_{I_{j,k(x)}}(y)$. If f is constant on $I_{j,k(x)}$ this in particular implies $E_f f_n(j, x) = f(x)$. (In other words: for any locally (at x) constant density f the bias of $f_n(j, x)$ equals zero if the resolution level is chosen fine enough.)

We finally note that the estimator $f_n(j, x)$ by construction only depends on data

4

points falling into $I_{j,k(x)}$. This amounts to $n2^{-j}$ being the 'effective' sample size for estimating f_0 at x.

2.2 Local choice of the resolution level

We fix $j_{\max} := j_{\max,n} \in \mathbb{N}$ satisfying $2^{-j_{\max}} \geq (\log n)^2/n$ for some $d > 0$. For thresholds ζ_n to be specified below, and for $J \in \mathbb{N}$, $J \leq j_{\max}$ and $0 < x \leq 1$, we define

$$\hat{\hat{j}}_n(J,x) = \min\left\{ j \in \mathbb{N}, \, J \leq j \leq j_{\max} : \right.$$

$$\left. \sqrt{n2^{-j'}} \left| f_n(j',x) - f_n(j,x) \right| \leq \zeta_n \sqrt{f_n(j,x)} \text{ for all } j', \, j < j' \leq j_{\max} \right\} \quad (1)$$

as well as

$$\hat{\hat{j}}_n(x) = \hat{\hat{j}}_n(0,x). \quad (2)$$

(If the condition in (1) is not met for any j, $J \leq j \leq j_{\max}$, we set $\hat{\hat{j}}_n(J,x) = j_{\max}$.) Given the locally variable resolution level $\hat{\hat{j}}_n$, we define the family of nonlinear estimators

$$\hat{f}_n(J,x) := f_n(\hat{\hat{j}}_n(J,x), x), \qquad \hat{f}_n(x) := f_n(\hat{\hat{j}}_n(x), x), \qquad x \in [0,1]. \quad (3)$$

These are estimators for $f_0(x)$ based on a locally variable resolution level depending on x, and they are density-analogues of the estimators introduced in [12] in the context of the Gaussian white noise model. Note that by construction $\hat{\hat{j}}_n(x)$ is a step function in x. Introducing the parameter J will be useful in what follows – effectively, $\hat{f}_n(J,x)$ is a nonlinear estimator based on a search over the resolution levels $j \geq J$ that stops at $j_{\max}$.

2.3 Threshold choice by propagation

One of the main challenges for all adaptive procedures is the choice of the thresholds ζ_n used in the tests defined in (1). Define the standardisation

$$\frac{1}{s_n(j,x)} := \begin{cases} \frac{1}{\sqrt{f_n(j,x)}} & \text{if } f_n(j,x) > 0; \\ 0 & \text{otherwise.} \end{cases}$$

We suggest to choose the thresholds in such a way that the following condition is satisfied:

Condition 1 *Let $\mathcal{F}_{j,k}$ be any triangular array of subsets of $V_{j,k}$, $j \leq j_{\max}$, $k = 0, \ldots, 2^j - 1$, and let $k(m)$ be the unique k such that $I_{j_{\max},m} \subseteq I_{j,k}$. We say that the thresholds ζ_n*

satisfy the uniform propagation condition $\mathrm{UP}(\alpha, \mathcal{F}_{j,k})$ *for some fixed* $\alpha > 0$ *if for every* n, *every* $j \leq j_{\max}$, *every* $m = 0, \ldots, 2^{j_{\max}} - 1$, *and every* $f \in \mathcal{F}_{j,k(m)}$ *we have that*

$$E_f \left(\sup_{x \in I_{j_{\max},m}} \max_{j \leq j' \leq j_{\max}} \sqrt{\frac{n2^{-j'}}{\log n}} \left| \frac{\hat{f}_n(j', x) - f_n(j', x)}{s_n(j', x)} \right| \right)^2 \leq \frac{\alpha}{n2^{2j_{\max}}}. \qquad (4)$$

(Note that since $\hat{j}_n(j') \geq j'$ we have that $f_n(j', x) = 0$ implies $\hat{f}_n(j', x) = 0$ for the fully data-driven estimator $\hat{f}_n(j')$, and so the error $|\hat{f}_n(j', x) - f_n(j', x)|$ is then 0.) An interpretation of this condition can be given along the following lines: For $0 < x \leq 1$ the class $\mathcal{F}_{j,k(x)}$ contains only densities f that can be exactly reconstructed on $I_{j,k(x)}$ by $\int K_j(x, y) f(y) dy$, so that the bias of the linear estimator $f_n(j', x)$ equals zero locally. In particular, any choice of the resolution level finer than j' will only increase the variance without reducing the bias, and we would want $\hat{j}_n(j', x)$ to detect that and equal, with large probability, j'. This property of $\hat{j}_n$ will then be mirrored in the fact that $\hat{f}_n(j', x) - f_n(j', x) = 0$ for every $j' \geq j$ on an event with large probability, in which case the l.h.s. of (4) is exactly equal to zero. The quantity $\alpha/(n2^{2j_{\max}})$ stands for the a priori expected tolerance for a probabilistic error of $\hat{j}_n$ to detect the 'correct' resolution level on each interval $I_{j_{\max},m}$ in this 'no-bias' situation.

The following lemma shows that Condition 1 is not empty and that thresholds ζ_n satisfying the uniform propagation condition exist. It shows furthermore that the thresholds can be taken to be of order $\sqrt{\log n}$ and independent of f, which will be crucial in understanding the adaptive properties of $\hat{f}_n$ below.

Lemma 1 *Let* $\mathcal{F}_{j,k}$ *equal* $V_{j,k}$ *intersected with the set*

$$\left\{ f : 0 < \delta \leq \inf_{0 < x \leq 1} f(x), \|f\|_\infty \leq M \right\}$$

for some fixed $0 < \delta, M < \infty$. *Then for every given* $\alpha > 0$ *there exists a numerical constant* $\kappa > 0$ *that depends only on* α *such that for any threshold choice*

$$\zeta_n \geq \kappa \sqrt{\log n}$$

the uniform propagation condition $\mathrm{UP}(\alpha, \mathcal{F}_{j,k})$ *is at least satisfied for* n *larger than some index that only depends on* δ *and* M.

While Lemma 1 proves the existence of thresholds of the order $\sqrt{\log n}$ under the uniform propagation condition – a fact that will be seen to imply adaptivity of $\hat{f}_n$ below

– it does not suggest a practical choice of ζ_n. Instead, this choice can be made by direct evaluation of (4), as follows: Condition 1 only concerns the local error bounds over small intervals $I_{j_{\max},m}$ on which the function f is constant, which effectively means that it suffices to check this condition only for classes of densities which are constant on the interval of interest. The particular choice of the interval $I_{j_{\max},m}$ is unimportant. Secondly, all quantities in Condition 1 depend on known quantities after f is chosen. By construction of the estimators f_n and $\hat{f}_n$ the random variable featuring in (4) – we call it T – only depends on the number of data points falling into each of the (uniquely determined) j'-fine intervals containing $I_{j_{\max},m}$. This observation allows for an easy computation of the l.h.s. of (4) along the following lines: Fix $0 \le p \le 1$. Then, for any $f \in \mathcal{F}_{j,k(m)}$ satisfying $2^{-j}f = p$ on $I_{j,k(m)}$, the number Z of observations falling into the interval $I_{j,k(m)}$ is binomial $B(n,p)$. Conditionally on $Z = k$, take k-many independent random variables that are uniform on $I_{j,k(m)}$ and count the number of observations $V_{j'}$ in each of the j'-fine intervals. Then compute f_n, $\hat{f}_n$; and T. This shows that T does only depend on $V_{j'}$, $j \le j' \le j_{\max}$, and that the l.h.s. of (4) is therefore equal to $E[E[T(V_j, \ldots, V_{j_{\max}})|Z]]$.

The practical choice of ζ_n can then be obtained via a Monte Carlo simulation of (4) by choosing ζ_n as the smallest threshold for which (4) is satisfied in the simulation for one specific interval $I_{j_{\max},m}$ uniformly over the class of all densities constant on this interval. Given $j_{\max}$ and α, this procedure has to be performed only for one fixed interval $I_{j_{\max},m}$, and then applies for every m simultaneously.

2.4   Local small bias condition

The idea behind Condition 1 is that we take 'idealised' classes of densities $\mathcal{F}$ for which we compute sharp thresholds ζ_n. The danger arises that the true density f_0 may be very different from the elements in $\mathcal{F}$, which may lead to wrong thresholds (and inference). We have to assess the error that comes from replacing f_0 by an element from $\mathcal{F}$, in a neighborhood of a given point x. This can be fundamentally quantified in terms of the log-likelihood ratio between f_0 and its local (at x) approximand in $\mathcal{F}$. As we shall see, one of the deeper reasons behind the fact that propagation methods imply adaptation results is that this error can be related to the usual bias term in linear estimation.

Condition 2 *Given real numbers $\Delta_{j,x}$, $0 < x \le 1$, $j \in \mathbb{N} \cup \{0\}$ satisfying $\Delta_{l',x} \le \Delta_{l,x}$ for every $l' > l$, we say that f_0 satisfies the local small bias condition at $x \in (0,1]$ and*

with $\Delta_{j,x} \equiv \Delta_{j,x}(f_0)$ if

$$\mathrm{Var}_{K_{j,x}(f_0)} \log \frac{f_0}{K_{j,x}(f_0)} \leq \Delta_{j,x}(f_0)$$

for all $j \in \mathbb{N}$.

The local 'cost' of transporting a product measure $\prod_{i=1}^n f_0(x_i)$ to $\prod_{i=1}^n K_{j,x}(f_0)(x_i)$ can be quantified by n times the variance featuring in the above condition, and we shall have to restrict ourselves to resolution levels j for which this transportation cost is at most a fixed constant times the logarithm of the sample size n. The smallest resolution level for which this is still the case will be defined as $j^*(x)$: More precisely, for some fixed positive constant Δ, define the local resolution level

$$j^*(x) := j^*(x, n, \Delta, f_0) = \min \left\{ j \in \mathbb{N} : j \leq j_{\max}, \, n\Delta_{j,x}(f_0) \leq \Delta \log n \right\}. \tag{5}$$

While this is an information-theoretic definition of j^*, a key observation of this subsection is that it has the classical 'bias-variance' tradeoff generically built into it for suitable choices of $\Delta_{j,x}(f_0)$.

Lemma 2 *Suppose f_0 is bounded by some finite number $M > 0$ and that*

$$\inf_{0 < x \leq 1} f_0(x) \geq \delta > 0.$$

Then f_0 satisfies Condition 2 with

$$\Delta_{j,x}(f_0) = \frac{M}{\delta^2} 2^{-j} \| f_0 - K_{j,x}(f_0) \|_\infty^2.$$

Proof. First, observe that $K_{j,x}(f_0)$ is bounded by M and bounded below by $\delta > 0$. Then, using that $K_{j,x}(f_0)$ coincides with f_0 outside of $I_{j,k(x)}$ and the inequality $|\log x - \log y| \leq \max(x^{-1}, y^{-1})|x - y|$, we get

$$\mathrm{Var}_{K_{j,x}(f_0)} \log \frac{f_0}{K_{j,x}(f_0)}$$
$$\leq \int \left(\log \frac{f_0(y)}{K_{j,x}(f_0)(y)} \right)^2 K_{j,x}(f_0)(y) dy$$
$$\leq \int \max \left(f_0(y)^{-2}, K_{j,x}(f_0)(y)^{-2} \right) (f_0(y) - K_{j,x}(f_0)(y))^2 K_{j,x}(f_0)(y) dy$$
$$\leq \frac{M}{\delta^2} \int (f_0(y) - K_{j,x}(f_0)(y))^2 dy \leq \frac{M}{\delta^2} 2^{-j} \| K_{j,x}(f_0) - f_0 \|_\infty^2.$$

■

The lemma shows that the quantity $(n/\log n)\Delta_{j,x}(f_0)$ can be viewed as the square of the 'bias divided by the variance' of linear projection estimators for $f_0(x)$. *Hence, to choose the smallest $j \le j_{\max}$ such that $(n/\log n)\Delta_{j,x}(f_0)$ is still bounded by a fixed constant Δ means to locally balance the 'variance' and 'bias' term in the nonparametric setting.*

To be more concrete, let us briefly discuss what this means in the classical situation where the bias is bounded by local regularity properties of the unknown density f_0. Since we are interested in spatial adaptation, we wish to take locally inhomogeneous smoothness into account by appealing to local Hölder conditions: Let $0 < t \le 1$ and let us say that a function $g : \mathbb{R} \to \mathbb{R}$ is locally t-Hölder at $x \in \mathbb{R}$ if for some $\eta > 0$

$$\sup_{0<|m|\le\eta} \frac{|g(x+m)-g(x)|}{|m|^t} < \infty.$$

Define further a 'local' Hölder ball of bounded functions

$$\mathcal{C}(t,x,L,\eta) := \left\{ g : \mathbb{R} \to \mathbb{R}, \ \max\left(\|g\|_\infty, \sup_{0<|m|\le\eta} \frac{|g(x+m)-g(x)|}{|m|^t} \right) \le L \right\}.$$

Condition 2 then has the following more classical interpretation in terms of local smoothness properties of f_0:

Lemma 3 *If $f_0 \in \mathcal{C}(t,x,L,\eta)$ for some $0 < t \le 1$, then the local bias $\|f_0 - K_{j,x}(f_0)\|_\infty$ is bounded by $c2^{-jt}$ for some constant $c(t,L,\eta)$. Furthermore, if*

$$\inf_{0<x\le 1} f_0(x) \ge \delta > 0,$$

then Condition 2 is satisfied with

$$\Delta_{j,x}(f_0) = c^2 \frac{L}{\delta^2} 2^{-j(2t+1)}. \tag{6}$$

Proof. Let $y \in I_{j,k(x)}$ be arbitrary. Then, using the substitution $2^j z = 2^j y - u$,

$$
\begin{aligned}
|f_0(y) - K_{j,x}(f_0)(y)| &= \left| 2^j \int_{I_{j,k(x)}} (f_0(y) - f_0(z)) dz \right| \\
&\leq \int_{-1}^{1} |f_0(y) - f_0(x) + f_0(x) - f_0(y - 2^{-j}u)| du \\
&\leq 2|f_0(y) - f_0(x)| + \int_{-1}^{1} |f_0(x) - f_0(y - 2^{-j}u)| du
\end{aligned}
$$

By definition of $x, y, I_{j,k(x)}$ we have $|y - x| \leq 2^{-j}$, and also $|y - 2^{-j}u - x| \leq 2^{-j+1}$ by the triangle inequality, so that for $2^{-j+1} \leq \eta$ the last quantity is bounded by $c_0 2^{-jt}$ in view of $f_0 \in \mathcal{C}(t, x, L, \eta)$. If $2^{-j} > \eta/2$, then the quantity in the last display can still be bounded by $6\|f_0\|_\infty \leq 6L$, so that choosing $c_1 = 6L(2/\eta)^t$ establishes the desired bound for $c = \max(c_0, c_1)$. To prove the second claim, apply Lemma 2. ∎

Using the bound from the last lemma to verify Condition 2, we see that, by definition of $j^*(x)$ and for $f_0 \in \mathcal{C}(t, x, L, \eta)$,

$$
\sqrt{\frac{n2^{-j^*(x)}}{\log n}} \sim \left(\frac{n}{\log n} \right)^{\frac{t}{2t+1}} \tag{7}
$$

is the locally (at x) optimal adaptive rate of convergence, so that the local small bias condition constructs a minimax optimal resolution level $j^*(x)$ at every $x \in [0, 1]$.

2.5 Main results

We now state the main results, starting with the following 'oracle' inequality.

Theorem 1 *Let* $\hat{f}_n(\cdot)$ *be the density estimator defined in* (3) *with thresholds* ζ_n *that satisfy the uniform propagation condition* $\mathrm{UP}(\alpha, \mathcal{F}_{j,k})$ *for some* $\mathcal{F}_{j,k}$. *Suppose* f_0 *satisfies Condition 2 for every* $0 < x \leq 1$, *and let* $j^*(x)$ *be as in* (5). *Then we have*

$$
\begin{aligned}
E_{f_0} \sup_{0 < x \leq 1} &\sqrt{\frac{n2^{-j^*(x)}}{\log n}} \left| \frac{\hat{f}_n(x) - f_n(j^*(x), x)}{s_n(j^*(x), x)} \right| \\
&\leq \frac{\zeta_n}{\sqrt{\log n}} + \sqrt{\frac{\alpha}{n}} n^{\Delta e^{4U}}
\end{aligned} \tag{8}
$$

for any U *satisfying*

$$
U \geq \sup_{0 < x \leq 1} \left\| \log \frac{f_0}{K_{j^*(x),x}(f_0)} \right\|_\infty. \tag{9}
$$

10

If $\zeta_n = O(\sqrt{\log n})$ – as follows under the conditions of Lemma 1 – and if one chooses $\Delta < 1/2$, U as in the remark below, then the r.h.s. of (8) is $O(1)$ as n tends to infinity. Theorem 1 thus implies that the estimator $\hat{f}_n$ with resolution levels chosen by the propagation approach is close to the linear 'oracle' estimator evaluated at the locally optimal resolution level $j^*(x)$, and this uniformly so on $(0, 1]$.

Remark 1 *If $\mathcal{F}_{j,k}$ is as in Lemma 1 and f_0 is bounded by M and bounded below by δ, we may apply Lemma 2 (using $2^{-j_{\max}} \geq d(\log n)^2/n$) to get the bound*

$$\log \frac{f_0}{K_{j^*(x),x}(f_0)} = \log \left(1 + \frac{f_0 - K_{j^*(x),x}(f_0)}{K_{j^*(x),x}(f_0)} \right) \leq \log \left(1 + \frac{\Delta}{dM \log n} \right),$$

which tends to zero as n tends to infinity.

Our results then imply the following uniform spatial adaptation result:

Theorem 2 *Assume that f_0 is bounded by M and satisfies $\inf_{0 < x \leq 1} f_0(x) \geq \delta > 0$. Let $\hat{f}_n(\cdot)$ be the density estimator from (3) with thresholds $\zeta_n = O(\sqrt{\log n})$ that satisfy the uniform propagation condition $\mathrm{UP}(\alpha, \mathcal{F}_{j,k})$ for $\mathcal{F}_{j,k}$ as in Lemma 1. Let $j^*(x)$ be as in (5) with $\Delta < 1/2$ and with $\Delta_{j,x}$ as in Lemma 2. Then*

$$\sup_{0 < x \leq 1} \sqrt{\frac{n2^{-j^*(x)}}{\log n}} \left| \hat{f}_n(x) - f_0(x) \right| = O_{\mathrm{Pr}_{f_0}}(1). \tag{10}$$

If $j^*(x)$ – with $\Delta < 1/2$ – is based on $\Delta_{j,x}$ as in Lemma 3, then (10) holds true and the rate is the adaptive locally optimal rate of convergence at every $0 < x \leq 1$ where f_0 is locally t-Hölder with $0 < t \leq 1$, see the discussion in Section 2.4 surrounding (7). A more classical way to formulate Theorem 2 is hence as follows.

Theorem 3 *Suppose the assumptions of Theorem 2 are satisfied and that the true density f_0 lies in $\mathcal{C}(t(x), x, L(x), \eta(x))$, $0 < x \leq 1$, and $t(\cdot), L(\cdot), \eta(\cdot)$ are bounded and uniformly bounded away from zero on $(0, 1]$. Let $j^*(x)$ be as in (5) with $\Delta < 1/2$ and with $\Delta_{j,x}$ as in Lemma 3. Then*

$$\sup_{0 < x \leq 1} \left(\frac{n}{\log n} \right)^{t(x)/(2t(x)+1)} \left| \hat{f}_n(x) - f_0(x) \right| = O_{\mathrm{Pr}_{f_0}}(1).$$

3 Proofs

3.1 Proof of Theorem 1

A first idea is to use a moment bound, localised at any point x of estimation, on the log-likelihood ratio between f_0 and its approximand in $V_{j,k}$.

Lemma 4 *If, for fixed $0 < x \leq 1$,*

$$\text{Var}_{K_{j,x}(f_0)} \log \frac{f_0}{K_{j,x}(f_0)} \leq \frac{D \log n}{n} \tag{11}$$

for some $0 < D < \infty$ and every $n \in \mathbb{N}$, then, for every $n \in \mathbb{N}$,

$$E_{K_{j,x}(f_0)} \left(\prod_{i=1}^{n} \frac{f_0(X_i)}{K_{j,x}(f_0)(X_i)} \right)^2 \leq n^{2De^{4U}}$$

holds for any U satisfying

$$U \geq \left\| \log \frac{f_0}{K_{j,x}(f_0)} \right\|_{\infty}.$$

Proof. Since the Kullback-Leibler distance

$$\mathcal{K}(f_0, K_{j,x}(f_0)) = -E_{K_{j,x}(f_0)} \log \frac{f_0}{K_{j,x}(f_0)} \geq 0$$

is non-negative, we have

$$E_{K_{j,x}(f_0)} \left(\prod_{i=1}^{n} \frac{f_0(X_i)}{K_{j,x}(f_0)(X_i)} \right)^2 \leq \left(E_{K_{j,x}(f_0)} e^{2 \left(\log \frac{f_0}{K_{j,x}(f_0)} - E_{K_{j,x}(f_0)} \log \frac{f_0}{K_{j,x}(f_0)} \right)} \right)^n$$

by the i.i.d. assumption. Using the power series expansion of the exponential function and that the variables in the exponent are centered, one easily bounds the previous display by

$$\left(1 + \frac{2De^{4U} \log n}{n} \right)^n \leq e^{2De^{4U} \log n} = n^{2De^{4U}}.$$

∎

Here is the proof of Theorem 1: We first note that Condition 2 allows us to take $\Delta_{j,x}(f_0)$ to be constant on the intervals $I_{j,k}$. Consequently, $j^*(\cdot)$ from (5) is then constant on every interval $I_{j_{\max},m}$, and we set

$$j_m^* = \sup_{x \in I_{j_{\max},m}} j^*(x).$$

To prove the theorem, we split

$$
\begin{aligned}
E_{f_0} \sup_{0 < x \leq 1} & \sqrt{\frac{n2^{-j^*(x)}}{\log n}} \left| \frac{\hat{f}_n(x) - f_n(j^*(x), x)}{s_n(j^*(x), x)} \right| \\
\leq E_{f_0} \sup_{0 < x \leq 1} & \sqrt{\frac{n2^{-j^*(x)}}{\log n}} \left| \frac{\hat{f}_n(x) - f_n(j^*(x), x)}{s_n(j^*(x), x)} \right| 1_{\{\hat{j}_n(x) < j^*(x)\}} \\
+ E_{f_0} \sup_{0 < x \leq 1} & \sqrt{\frac{n2^{-j^*(x)}}{\log n}} \left| \frac{\hat{f}_n(x) - f_n(j^*(x), x)}{s_n(j^*(x), x)} \right| 1_{\{\hat{j}_n(x) \geq j^*(x)\}} \\
=: I + II &
\end{aligned}
$$

according to whether $\hat{j}_n(x)$ comes to lie below the local resolution level $j^*(x)$ or not. By definition of $\hat{j}_n(x)$ in (1) one immediately has

$$
I \leq \frac{\zeta_n}{\sqrt{\log n}}.
$$

About II: Define

$$
S_m = \sup_{x \in I_{j_{\max}, m}} \max_{j_m^* \leq j \leq j_{\max}} \sqrt{\frac{n2^{-j}}{\log n}} \left| \frac{\hat{f}_n(j, x) - f_n(j, x)}{s_n(j, x)} \right|. \tag{12}
$$

Using that on the event $\hat{j}_n(x) \geq j^*(x)$ we necessarily have $\hat{f}_n(x) = \hat{f}_n(j^*(x), x)$, we see that

$$
\begin{aligned}
II & \leq E_{f_0} \sup_{0 < x \leq 1} \sqrt{\frac{n2^{-j^*(x)}}{\log n}} \left| \frac{\hat{f}_n(j^*(x), x) - f_n(j^*(x), x)}{s_n(j^*(x), x)} \right| \\
& \leq E_{f_0} \max_m \sup_{x \in I_{j_{\max}, m}} \max_{j_m^* \leq j \leq j_{\max}} \sqrt{\frac{n2^{-j}}{\log n}} \left| \frac{\hat{f}_n(j, x) - f_n(j, x)}{s_n(j, x)} \right| \\
& \leq 2^{j_{\max}} \max_m E_{f_0} S_m. \tag{13}
\end{aligned}
$$

We use the Cauchy-Schwarz inequality to bound

$$
\begin{aligned}
& E_{f_0} S_m \\
&= \int \cdots \int S_m(x_1, \ldots, x_n) \prod_{i=1}^n f_0(x_i) dx_1 \cdots dx_n \\
&= \int \cdots \int S_m(x_1, \ldots, x_n) \prod_{i=1}^n \frac{f_0(x_i)}{K_{j_m^*, x}(f_0)(x_i)} \prod_{i=1}^n K_{j_m^*, x}(f_0)(x_i) dx_1 \cdots dx_n \\
&\leq \sqrt{E_{K_{j_m^*, x}(f_0)} S_m^2} \sqrt{E_{K_{j_m^*, x}(f_0)} \left(\prod_{i=1}^n \frac{f_0(X_i)}{K_{j_m^*, x}(f_0)(X_i)} \right)^2}
\end{aligned}
$$

by the square-root of the second moment of S_m under the 'idealised' density $K_{j_m^*, x}(f_0)$ times the square-root of the second moment of the likelihood ratio. (Here, x is any point in $I_{j_{\max}, m}$.) Using Condition 1 and Lemma 4, we obtain a bound for the last term in (13) of order

$$
2^{j_{\max}} \max_m E_{f_0} S_m \leq \sqrt{\frac{\alpha}{n}} n^{\Delta e^{4U}},
$$

which concludes the proof of the theorem.

3.2 Proof of Theorems 2 and 3

We first prove Theorem 2: Clearly,

$$
\begin{aligned}
& \sup_{0 < x \leq 1} \sqrt{\frac{n 2^{-j^*(x)}}{\log n}} \left| \hat{f}_n(x) - f_0(x) \right| \\
& \leq \sup_{0 < x \leq 1} \sqrt{\frac{n 2^{-j^*(x)}}{\log n}} \left| \frac{\hat{f}_n(x) - f_n(j^*(x), x)}{s_n(j^*(x), x)} \right| \sqrt{f_n(j^*(x), x)} \\
& \quad + \sup_{0 < x \leq 1} \sqrt{\frac{n 2^{-j^*(x)}}{\log n}} \left| f_n(j^*(x), x) - f_0(x) \right|.
\end{aligned}
$$

The first factor of the first summand is bounded in probability in view of Theorem 1 and of Lemma 1 and the hypothesis $\zeta_n = O(\sqrt{\log n})$. The second factor of the first summand is also bounded in probability since

$$
\sup_{0 < x \leq 1} \max_{j \leq j_{\max}} \left| f_n(j, x) - E_{f_0} f_n(j, x) \right| = o_{P_{f_0}}(1)
$$

by Proposition 2, using $2^{-j_{\max}} \geq d(\log n)^2/n$, and since $\sup_{x,j} |E_{f_0} f_n(j,x)| \leq \|f_0\|_\infty < \infty$. It remains to prove that the second summand is bounded in probability, and we achieve this by bounding the moment

$$
E_{f_0} \sup_{0<x\leq 1} \sqrt{\frac{n2^{-j^*(x)}}{\log n}} |f_n(j^*(x),x) - E_{f_0} f_n(j^*(x),x)|
$$

$$
+ \sup_{0<x\leq 1} \sqrt{\frac{n2^{-j^*(x)}}{\log n}} |E_{f_0} f_n(j^*(x),x) - f_0(x)|
$$

$$
\leq E_{f_0} \sup_{0<x\leq 1} \max_{j\leq j_{\max}} \sqrt{\frac{n2^{-j}}{\log n}} |f_n(j,x) - E_{f_0} f_n(j,x)|
$$

$$
+ \max_{m} \sup_{x\in I_{j_{\max},m}} \sqrt{\frac{n2^{-j^*(x)}}{\log n}} |E_{f_0} f_n(j^*(x),x) - f_0(x)|.
$$

The first term is bounded by a fixed constant using Proposition 2 below. Recalling the definition of j_m^* from the beginning of the proof of Theorem 1 and choosing $\Delta_{j,x}(f_0)$ from Lemma 2, the second term is bounded by

$$
\max_{m} \sqrt{\frac{n2^{-j_m^*}}{\log n}} \sup_{x\in I_{j_{\max},m}} |E_{f_0} f_n(j_m^*,x) - f_0(x)|
$$

$$
\leq \max_{m} \sqrt{\frac{n2^{-j_m^*}}{\log n}} \|K_{j_m^*,x}(f_0) - f_0\|_\infty
$$

$$
\leq \delta\sqrt{\frac{\Delta}{M}},
$$

where x is any point in $I_{j_{\max},m}$, and this completes the proof.

We next prove Theorem 3: Using the hypotheses on $t(\cdot), L(\cdot), \eta(\cdot)$, the proof of Lemma 3 shows that f_0 satisfies Condition 2 with

$$
\Delta_{j,x}(f_0) = c'2^{-j(2t(x)+1)}, \tag{14}
$$

$0 < c' < \infty$, where c' does not depend on x. Using that $t(\cdot)$ is bounded below by some positive number implies that

$$
\Delta_{j_{\max},x}(f_0) = c'2^{-j_{\max}(2t(x)+1)} \leq \frac{\Delta \log n}{n}.
$$

holds for n large enough (independent of $x \in (0,1]$), so that $j^*(x)$, when based on $\Delta_{j,x}(f_0)$ as in (14), is asymptotically equivalent to the minimax optimal locally adaptive

15

rate, uniformly so for all x.

3.3 Proof of Lemma 1

The proof relies on Propositions 1 and 2 which are given below. Recall first from Section 2.1 that for $f \in V_{j,k}$ and $j' \geq j$ we necessarily have $E_f f_n(j', x) - f(x) = 0$ for every $x \in I_{j,k}$, so the bias at $x \in I_{j,k}$ is exactly zero, a fact we shall use repeatedly below without separate mentioning. Write

$$\sup_{x \in I_{j_{\max}, m}} \max_{j' \geq j} \sqrt{\frac{n 2^{-j'}}{\log n}} \left| \frac{\hat{f}_n(j', x) - f_n(j', x)}{s_n(j', x)} \right|$$

$$= \sup_{x \in I_{j_{\max}, m}} \max_{j' \geq j} \sqrt{\frac{n 2^{-j'}}{\log n}} \sum_{l > j'} \left| \frac{f_n(l, x) - f_n(j', x)}{s_n(j', x)} \right| 1_{\{\hat{j}_n(j', x) = l\}}. \tag{15}$$

To treat the indicator, observe that

$$\{\hat{j}_n(j', x) = l\}$$
$$\subseteq \left\{ \sqrt{n 2^{-l'}} |f_n(l', x) - f_n(l-1, x)| > \zeta_n \sqrt{f_n(l-1, x)} \text{ for some } l' \geq l \right\}$$
$$\subseteq \left\{ \sqrt{n 2^{-l'}} |f_n(l', x) - E_f f_n(l', x) + E_f f_n(l-1, x) - f_n(l-1, x)| \right.$$
$$\left. > \zeta_n \frac{\sqrt{f(x)}}{2} \text{ for some } l' \geq l \right\}$$
$$\cup \left\{ \min_{\ell \geq j} \sqrt{f_n(\ell, x)} \leq \frac{\sqrt{f(x)}}{2} \right\}$$

Observe that the first set is a subset of

$$\left\{ \sqrt{n 2^{-l'}} \left| f_n(l', x) - E_f f_n(l', x) \right| \geq \frac{\zeta_n \sqrt{f(x)}}{4} \text{ for some } l' \geq l \right\}$$
$$\cup \left\{ \sqrt{n 2^{-(l-1)}} \left| f_n(l-1, x) - E_f f_n(l-1, x) \right| > \frac{\zeta_n \sqrt{f(x)}}{4} \right\}$$
$$\subseteq \left\{ \max_{\ell \geq j} \sqrt{n 2^{-\ell}} \| f_n(\ell) - E_f f_n(\ell) \|_{I_{j_{\max}, m}} > \frac{\zeta_n \sqrt{\|f\|_{I_{j,k(m)}}}}{4} \right\} =: B_1$$

16

and that, using $y \geq \sqrt{\delta y}$ for $y \geq \delta$, the second set is contained in

$$
\left\{ \max_{\ell \geq j} |f_n(\ell, x) - E_f f_n(\ell, x)| > \frac{f(x)}{2} \right\}
$$

$$
\subseteq \left\{ \max_{\ell \geq j} |f_n(\ell, x) - E_f f_n(\ell, x)| > \frac{\sqrt{\delta f(x)}}{2} \right\}
$$

$$
\subseteq \left\{ \sup_{x \in I_{j_{\max}, m}, \ell \geq j} |f_n(\ell, x) - E_f f_n(\ell, x)| > \frac{\sqrt{\delta \|f\|_{I_{j,k(m)}}}}{2} \right\} := B_2;
$$

so that $\{\hat{j}_n(j', x) = l\} \subseteq B_1 \cup B_2 =: B$, a set which does not depend on j', x or l. Hence, $1_{\{\hat{j}_n(j', x) = l\}} \leq 1_B$ uniformly in j', x, l, so that the quantity in (15) is bounded from above by

$$
1_B \sup_{x \in I_{j_{\max}, m}} \max_{j' \geq j} \sqrt{\frac{n 2^{-j'}}{\log n}} \sum_{l > j'} \left| \frac{f_n(l, x) - f_n(j', x)}{s_n(j', x)} \right|,
$$

and therefore the second moment of (15) is bounded, using the Cauchy-Schwarz inequality, by

$$
\mathrm{Pr}_f(B)^{1/2} \left\| \sup_{x \in I_{j_{\max}, m}} \max_{j' \geq j} \sqrt{\frac{n 2^{-j'}}{\log n}} \sum_{l > j'} \left| \frac{f_n(l, x) - f_n(j', x)}{s_n(j', x)} \right| \right\|^2_{4, \mathrm{Pr}_f} =: I \times II.
$$

We first bound II: By the triangle inequality and since the bias is exactly zero, this term is less than or equal to

$$
2 \left\| \sup_{x \in I_{j_{\max}, m}} \max_{j' \geq j} \sqrt{\frac{n 2^{-j'}}{\log n}} \sum_{l > j'} \left| \frac{f_n(l, x) - E_f f_n(l, x)}{s_n(j', x)} \right| \right\|^2_{4, \mathrm{Pr}_f}
$$

$$
+ 2 \left\| \sup_{x \in I_{j_{\max}, m}} \max_{j' \geq j} \sqrt{\frac{n 2^{-j'}}{\log n}} \sum_{l > j'} \left| \frac{f_n(j', x) - E_f f_n(j', x)}{s_n(j', x)} \right| \right\|^2_{4, \mathrm{Pr}_f} . \tag{16}
$$

Define now $S = \{\sup_{x \in I_{j_{\max},m}} \min_{j' \geq j} f_n(j',x) \geq \delta/2\}$. Note that, by definition of $f_n(j')$, $f_n(j',x) > 0$ implies $f_n(j',x) \geq 2^{j'}/n$. Then, for every $1 \leq p < \infty$,

$$E_f \left(\sup_{x \in I_{j_{\max},m}} \max_{j' \geq j} \frac{1}{s_n(j',x)} \right)^p = E_f \left(\sup_{x \in I_{j_{\max},m}} \max_{j' \geq j} \frac{1}{s_n(j',x)} (1_S + 1_{S^c}) \right)^p$$

$$\leq \frac{2^{3p/2-1}}{\delta^{p/2}}$$

$$+ 2^{p-1} n^{p/2} E_f 1 \left\{ \sup_{x \in I_{j_{\max},m}} \min_{j' \geq j} |f_n(j',x) - E_f f_n(j',x) + f(x)| < \frac{\delta}{2} \right\}$$

$$\leq \frac{2^{3p/2-1}}{\delta^{p/2}} + 2^{p-1} n^{p/2} \mathrm{Pr}_f \left\{ \sup_{x \in I_{j_{\max},m}, \, j' \geq j} |f_n(j',x) - E_f f_n(j',x)| > \frac{\delta}{2} \right\}$$

$$\leq \frac{2^{3p/2-1}}{\delta^{p/2}}$$

$$+ 2^{p-1} n^{p/2} \mathrm{Pr}_f \left\{ \sup_{x \in I_{j_{\max},m}, \, j' \geq j} \sqrt{n 2^{-j'}} |f_n(j',x) - E_f f_n(j',x)| > \frac{\delta \sqrt{d} \log n}{2} \right\}$$

$$\leq \frac{2^{3p/2-1}}{\delta^{p/2}} + 2^{p-1} n^{p/2} c n^{-\frac{\delta^2 d}{4c} \log n}$$

for large n in view of Proposition 1 (using that $2^{-j_{\max}} \geq d(\log n)^2/n$), so that this expectation is bounded uniformly in n by some constant $c_1(p,\delta,M)$. Using this, the Cauchy-Schwarz inequality and Proposition 2, the square of the first term in (16) is less than or equal to

$$c_2 2^{2j_{\max}} j_{\max}^4$$

$$\times E_f \left(\sup_{x \in I_{j_{\max},m}} \max_{l \geq j} \sqrt{\frac{n 2^{-l}}{\log n}} |f_n(l,x) - E_f f_n(l,x)| \sup_{x \in I_{j_{\max},m}} \max_{l \geq j} \frac{1}{s_n(j',x)} \right)^4$$

$$\leq c_2 2^{2j_{\max}} j_{\max}^4 \left(E_f \left(\sup_{x \in I_{j_{\max},m}} \max_{l \geq j} \sqrt{\frac{n 2^{-l}}{\log n}} |f_n(l,x) - E_f f_n(l,x)| \right)^8 \right)^{1/2}$$

$$\times \left(E_f \left(\sup_{x \in I_{j_{\max},m}} \max_{l \geq j} \frac{1}{s_n(l,x)} \right)^8 \right)^{1/2} \leq c_3 2^{2j_{\max}} j_{\max}^4;$$

and the same reasoning also implies that the second term in (16) is less than or equal to some constant, so that we can conclude, using the lower bound of $2^{-j_{\max}}$, that

$$II \leq c_4 n \tag{17}$$

18

for some fixed constant c_4 that depends only on δ and M.

To bound I, we have the following: First, using Proposition 1 below, we see

$$\begin{aligned}
\Pr_f(B_1) &= \Pr_f\left\{\max_{\ell \geq j} \sqrt{n2^{-\ell}}\|f_n(\ell) - E_f f_n(\ell)\|_{I_{j_{\max},m}} > \frac{\zeta_n\sqrt{\|f\|_{I_{j,k(m)}}}}{4}\right\} \\
&\leq Dn^{-\frac{\kappa^2\delta}{4D}}
\end{aligned} \tag{18}$$

for large n, with D only depending on M. Furthermore, using $2^{-j_{\max}} \geq d(\log n)^2/n$ and Proposition 1 below,

$$\begin{aligned}
&\Pr_f(B_2) \\
&\leq \Pr_f\left\{\sup_{x \in I_{j_{\max},m}, \ell \geq j} \sqrt{n2^{-\ell}}|f_n(\ell,x) - E_f f_n(\ell,x)| > \frac{\sqrt{d\delta\|f\|_{I_{j,k(m)}}}\log n}{2}\right\} \\
&\leq Dn^{-\frac{d\delta^2}{4D}\log n}
\end{aligned}$$

for large n. Thus, choosing κ large enough but finite depending on the choice of α, we obtain for n large enough

$$I \times II \leq c_4 Dn\left(n^{-\frac{\kappa^2\delta}{4D}} + n^{-\frac{d\delta^2}{4D}\log n}\right) \leq \frac{\alpha}{n2^{2j_{\max}}}.$$

This completes the proof.

3.4 Uniform-in-bandwidth bounds for Haar wavelet density estimators and some consequences

The following exponential inequality was used repeatedly in the proofs.

Proposition 1 *Let $j_{\max} \in \mathbb{N}$ such that $2^{-j_{\max}} \geq d(\log n)^2/n$. Let $I = (2^{-j}k, 2^{-j}(k+1)]$ for some $j \leq j_{\max}$ and $k \in \mathbb{Z}$, and suppose $f : \mathbb{R} \to [0,\infty)$ is a density that satisfies $\|f\|_I \leq M$ and*

$$\inf_{x \in I} f(x) \geq \delta > 0.$$

There exist constants $C_1(d)$, $C_2(d)$ and an index $n(\delta, M)$ such that for all $n \geq n(\delta, M)$ and all $C_3 \geq C_2(d)$, if

$$C_1(d)\sqrt{\|f\|_I \log n} \leq u \leq C_3\|f\|_I\sqrt{n2^{-j_{\max}}}, \tag{19}$$

then

$$\Pr{}_f\left\{\sup_{x\in I}\max_{j\le j'\le j_{\max}}\sqrt{n2^{-j'}}|f_n(j',x)-E_f f_n(j',x)|\ge u\right\}\le De^{-\frac{u^2}{D}},$$

where D only depends on C_3 and M.

Proof. Writing

$$\sqrt{n2^{-j'}}\left|f_n(j',x)-E_f f_n(j',x)\right|$$
$$=2\sqrt{\frac{2^{j_{\max}}}{n}}\frac{\sqrt{2^{j'-j_{\max}}}}{2}\left|\sum_{i=1}^{n}(K(2^{j'}x,2^{j'}X_i)-E_f K(2^{j'}x,2^{j'}X_i))\right|,$$

we have to consider the supremum

$$2\sqrt{\frac{2^{j_{\max}}}{n}}\sup_{h\in\mathcal{H}}\left|\sum_{i=1}^{n}(h(X_i)-E_f h(X_i))\right|$$

of the (scaled) empirical processes indexed by the class of functions

$$\mathcal{H}:=\left\{\frac{\sqrt{2^{j'-j_{\max}}}}{2}K(2^{j'}x,2^{j'}(\cdot)):x\in I,\ j'\ge j\right\}.$$

This class has constant envelope $1/2$ since $j'\le j_{\max}$ and since $\sup_{x,y}|K(x,y)|=1$. Furthermore, noting that $K^2(x,y)=K(x,y)$ for every x,y, we have for $h\in\mathcal{H}$ that

$$\begin{aligned}E_f h^2(X) &= \frac{2^{j'-j_{\max}}}{4}\int K^2(2^{j'}x,2^{j'}y)f(y)dy\\ &= \frac{2^{j'-j_{\max}}}{4}\int_{2^{-j'}k(x)}^{2^{-j'}(k(x)+1)}f(y)dy\le\frac{2^{-j_{\max}}}{4}\|f\|_I.\end{aligned}$$

Note further that $\mathcal{H}$ is a VC-type class of functions by using Lemma 2 in [7] and a simple computation on covering numbers (including an obvious covering of the set $[2^{-j_{\max}},1]\subseteq[0,1]$). Rewrite

$$\Pr{}_f\left\{\sup_{x\in I}\max_{j\le j'\le j_{\max}}\sqrt{n2^{-j'}}|f_n(j',x)-E_f f_n(j',x)|\ge u\right\}$$
$$=\Pr{}_f\left\{\sup_{h\in\mathcal{H}}\left|\sum_{i=1}^{n}(h(X_i)-E_f h(X_i))\right|\ge\frac{u\sqrt{n2^{-j_{\max}}}}{2}\right\}$$

and apply expression (21) in [8], with

$$\sigma^2 := \frac{2^{-j_{\max}} \|f\|_I}{4} \wedge \frac{1}{4}$$

and

$$\lambda := \begin{cases} c_1(d) \sqrt{\frac{\log n}{\|f\|_I \log \frac{n}{\|f\|_I}}} & \text{if } \|f\|_I \leq 1, \\ c_2(d) & \text{otherwise;} \end{cases}$$

for appropriate constants $c_1(d)$, $c_2(d)$ that only depend on d. ∎

Proposition 2 *Let $j_{\max}$, I and f be as in Proposition 1. Then there exists a constant $D(d, \delta, M)$ such that for every $1 \leq p < \infty$ we have*

$$E_f \left(\sup_{x \in I} \max_{j \leq j' \leq j_{\max}} \sqrt{\frac{n2^{-j'}}{\log n}} \left| f_n(j', x) - E_f f_n(j', x) \right| \right)^p \leq D^p. \tag{20}$$

Proof. The proof follows from considering the same empirical process as in the proof of Proposition 1, and using bounds for p-th moments of empirical processes indexed by uniformly bounded VC-classes of functions, e.g., the bound in the display following (21) in [8], with σ^2 and λ as in the proof of Proposition 1, together with Proposition 3.1 in [5]. ∎

References

[1] DONOHO, D. L. and JOHNSTONE, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81** 425-455.

[2] DONOHO, D. L.; JOHNSTONE, I. M.; KERKYACHARIAN, G.; and PICARD, D. (1995). Wavelet shrinkage: asymptopia? *J. Roy. Statist. Soc. Ser. B* **57** 301-369.

[3] DONOHO, D. L.; JOHNSTONE, I. M.; KERKYACHARIAN, G.; and PICARD, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.* **24** 508-539.

[4] GINÉ, E. and GUILLOU, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. Inst. H. Poincaré Probab. Statist.* **38** 907-921.

[5] GINÉ, E.; LATALA, R.; and ZINN, J. (2000). Exponential and moment inequalities for U-statistics. In: *High Dimensional Probability II* (eds. E. Giné, D. Mason, J. A. Wellner), 13-38.

[6] GINÉ, E. and NICKL, R. (2009). An exponential inequality for the distribution function of the kernel density estimator, with applications to adaptive estimation. *Probab. Theory Related Fields* **143** 569-596.

[7] GINÉ, E. and NICKL, R. (2009). Uniform limit theorems for wavelet density estimators. *Ann. Probab.* **37** 1605-1646.

[8] GINÉ, E. and NICKL, R. (2010). Adaptive estimation of a distribution function and its density in sup-norm loss by wavelet and spline projections. *Bernoulli*, **16** 1137-1163.

[9] GOLDENSHLUGER, A. and LEPSKI, O. (2009). Structural adaptation via $\mathbb{L}_p$-norm oracle inequalities. *Probab. Theory Related Fields* **143** 41-71.

[10] HÄRDLE, W.; KERKYACHARIAN, G.; PICARD, D.; and TSYBAKOV, A. (1998). *Wavelets, approximation, and statistical applications.* Lecture Notes in Statistics **129**. Springer, New York.

[11] JAFFARD, S. (2000). On the Frisch-Parisi conjecture. *J. Math. Pures Appl.* **79** 525-552.

[12] LEPSKI, I.; MAMMEN, E.; and SPOKOINY, V. (1997). Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. *Ann. Statist.* **25** 929-947.

[13] POLZEHL, J. and SPOKOINY, V. (2006). Propagation-separation approach for local likelihood estimation. *Probab. Theory Related Fields* **135** 335-362.

[14] SPOKOINY, V. and VIAL, C. (2009) Parameter tuning in pointwise adaptation using a propagation approach. *Ann. Statist.* **37** 2783-2807.

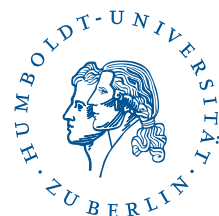# SFB 649 Discussion Paper Series 2011

For a complete list of Discussion Papers published by the SFB 649,
please visit http://sfb649.wiwi.hu-berlin.de.

SFB 649 Discussion Paper Series 2011

For a complete list of Discussion Papers published by the SFB 649,
please visit http://sfb649.wiwi.hu-berlin.de.

SFB 649 Discussion Paper Series 2011

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

044 "Predicting Bid-Ask Spreads Using Long Memory Autoregressive Conditional Poisson Models" by Axel Groß-Klußmann and Nikolaus Hautsch, July 2011.

045 "Bayesian Networks and Sex-related Homicides" by Stephan Stahlschmidt, Helmut Tausendteufel and Wolfgang K. Härdle, July 2011.

046 "The Regulation of Interdependent Markets", by Raffaele Fiocco and Carlo Scarpa, July 2011.

047 "Bargaining and Collusion in a Regulatory Model", by Raffaele Fiocco and Mario Gilli, July 2011.

048 "Large Vector Auto Regressions", by Song Song and Peter J. Bickel, August 2011.

049 "Monetary Policy, Determinacy, and the Natural Rate Hypothesis", by Alexander Meyer-Gohde, August 2011.

050 "The impact of context and promotion on consumer responses and preferences in out-of-stock situations", by Nicole Wiebach and Jana L. Diels, August 2011.

051 "A Network Model of Financial System Resilience", by Kartik Anand, Prasanna Gai, Sujit Kapadia, Simon Brennan and Matthew Willison, August 2011.

052 "Rollover risk, network structure and systemic financial crises", by Kartik Anand, Prasanna Gai and Matteo Marsili, August 2011.

053 "When to Cross the Spread: Curve Following with Singular Control" by Felix Naujokat and Ulrich Horst, August 2011.

054 "TVICA - Time Varying Independent Component Analysis and Its Application to Financial Data" by Ray-Bing Chen, Ying Chen and Wolfgang K. Härdle, August 2011.

055 "Pricing Chinese rain: a multi-site multi-period equilibrium pricing model for rainfall derivatives" by Wolfgang K. Härdle and Maria Osipenko, August 2011.

056 "Limit Order Flow, Market Impact and Optimal Order Sizes: Evidence from NASDAQ TotalView-ITCH Data" by Nikolaus Hautsch and Ruihong Huang, August 2011.

057 "Optimal Display of Iceberg Orders" by Gökhan Cebiroğlu and Ulrich Horst, August 2011.

058 "Optimal liquidation in dark pools" by Peter Kratz and Torsten Schöneborn, September 2011.

059 "The Merit of High-Frequency Data in Portfolio Allocation" by Nikolaus Hautsch, Lada M. Kyj and Peter Malec, September 2011.

060 "On the Continuation of the Great Moderation: New evidence from G7 Countries" by Wenjuan Chen, September 2011.

061 "Forward-backward systems for expected utility maximization" by Ulrich Horst, Ying Hu, Peter Imkeller, Anthony Réveillac and Jianing Zhang.

062 "On heterogeneous latent class models with applications to the analysis of rating scores" by Aurélie Bertrand and Christian M. Hafner, October 2011.

063 "Multivariate Volatility Modeling of Electricity Futures" by Luc Bauwens, Christian Hafner and Diane Pierret, October 2011.

SFB 649 Discussion Paper Series 2011

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.