Higher Order Estimation at Lebesgue Points

Jan BEIRLANT [a], Alain BERLINET [b] and Gérard BIAU [b,*]

[a] Katholieke Universiteit Leuven, Belgium
[b] Université Montpellier II, France

Abstract

The symmetric derivative of a probability measure at a Lebesgue point can often be specified by an exact relation involving a regularity index. Knowledge of this index is of practical interest, for example to specify the local behavior of the measure under study and to evaluate bandwidths or number of neighbors to take into account in smoothing techniques. This index also determines local rates of convergence of estimators of particular points of curves and surfaces, like minima and maxima. In this paper, we consider the estimation of the d-dimensional regularity index. We introduce an estimator and derive the basic asymptotic results. Our estimator is inspired by an estimator proposed in Drees and Kaufmann [13] in the context of extreme value statistics. Then we show how (estimates of) the regularity index can be used to solve practical problems in nearest neighbor density estimation, such as removing bias or selecting the number of neighbors. Results of simulations are presented.

Index Terms **— Lebesgue point, probability density, regularity index, nearest neighbor density estimation, mode estimation, rate of convergence.**

AMS 2000 Classification**: 62G05.**

*Corresponding author. Email: biau@math.univ-montp2.fr .

1 Introduction

1.1 Motivations

The subject of this paper is related to the general problem of derivation of measures (see Rudin [32], Chapter 7) and finds its motivation in a paper by Berlinet and Levallois [4]. In their paper, Berlinet and Levallois address the problem of the asymptotic normality of the nearest neighbor density estimator (Loftsgaarden and Quesenberry [27], Moore and Yackel [29], see Paragraph 2.1 below) in cases where the density has bad local behavior (*e.g.*, it is not continuous or has infinite derivative). These authors point out that what is important is the local behavior of the probability measure associated with the density, and more exactly the rate at which the local value of the density is approximated by the ratios of ball measures. More precisely, let $\mathcal{B}(\mathbb{R}^d)$ be the Borel σ-field of $\mathbb{R}^d$, $d \geq 1$, and let μ be a probability measure on $\big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$. We denote by λ the Lebesgue measure on $\mathbb{R}^d$ and we equip $\mathbb{R}^d$ with a norm denoted $\|.\|$. Let x be a point in $\mathbb{R}^d$, δ a positive real number and $B(x, \delta)$ the open ball with center at x and radius δ. To appreciate the local behavior of $\mu\big(B(x,\delta)\big)$ with respect to $\lambda\big(B(x,\delta)\big)$ one can consider the ratio of these two quantities. If, for fixed x, the following limit

$$f(x) = \lim_{\delta \downarrow 0} \frac{\mu\big(B(x,\delta)\big)}{\lambda\big(B(x,\delta)\big)} \tag{1.1}$$

does exist, then x is called a *Lebesgue point* of the measure μ. It can be shown (Rudin [32], Chapter 7) that λ-almost all points of $\big(\mathbb{R}^d, \|.\|\big)$ are Lebesgue points of μ. Moreover, if μ is absolutely continuous with respect to λ, then the Radon-Nykodim derivative of μ and f coincide λ-almost everywhere. Thus, in this case, one can select among the versions of the density of μ a particular one, still denoted f, satisfying (1.1) at any point where the limit exists. The notion of Lebesgue point plays a key role in the study of functional estimators and allows to state elegant results with few restrictions on the functions to be estimated. In this context, Berlinet and Levallois [4] define a *ρ-regularity* point of the measure μ as any Lebesgue point x of μ satisfying

$$\left| \frac{\mu\big(B(x,\delta)\big)}{\lambda\big(B(x,\delta)\big)} - f(x) \right| \leq \rho(\delta), \tag{1.2}$$

where ρ is a measurable function such that $\lim_{\delta \downarrow 0} \rho(\delta) = 0$. Roughly, the function ρ is intended to specify the convergence of ball ratios towards $f(x)$ in (1.1). For example, if $d = 1$ and the measure μ has a density f with derivative f' bounded by some constant C_x on a neighborhood of x, then

ρ-regularity holds with $\rho(\delta) = (C_x/2)\,\delta$. It is also clear that if f satisfies a local Hölder condition at the point x with exponent α_x, then we have ρ-regularity with $\rho(\delta) = C_x/(\alpha_x+1)\,\delta^{\alpha_x}$. However, it is easy to exhibit examples of measures with ρ-regularity but bad local behavior of the density such as a discontinuity of second kind, see example f_3 below. It can also be verified that the proposed techniques will continue to work when ρ is in some more general class of regularly varying functions at 0, *i.e.*, which satisfy $\rho(\delta\varepsilon)/\rho(\varepsilon) \to \delta^{\alpha_x}$ as $\varepsilon \downarrow 0$. As observed by Berlinet and Levallois [4], the notion of ρ-regularity actually involves measures rather than densities and is much more appropriate to specify properties of estimators.

Clearly, the function ρ in (1.2) need not be unique and may depend on the underlying norm on $\mathbb{R}^d$. In the present paper, we assume that a more precise relation than (1.2) holds at the Lebesgue point x, namely

$$\frac{\mu\big(B(x,\delta)\big)}{\lambda\big(B(x,\delta)\big)} = f(x) + C_x\,\delta^{\alpha_x} + \mathrm{o}(\delta^{\alpha_x}) \quad \text{as } \delta \downarrow 0, \tag{1.3}$$

where C_x is a non-zero constant and α_x is a positive real number. It is straightforward to show that (1.3) implies ρ-regularity at the point x with $\rho(\delta) \sim \delta^{\alpha_x}$. Note that the constants C_x and α_x of model (1.3) (provided they exist) are uniquely determined. In order to justify the relevance of this model, some examples and an application to mode estimation in the multivariate case are discussed in the next paragraph. The index α_x is a *regularity index* that controls the degree of smoothness of the symmetric derivative of μ with respect to λ. Roughly speaking, the larger the value of α_x, the more regular the derivative of μ is at the point x. The knowledge of this index is of practical interest, for example to specify the local behavior of the measure under study and to evaluate bandwidths or number of neighbors to take into account in smoothing techniques (Bosq and Lecoutre [5]). In fact, the behavior of nonparametric density estimators such as the nearest neighbor estimator at a given point x does strongly depend on this local behavior, and with a smaller index α_x fewer nearest neighbors should be used. See for instance Devroye [11], Lepski, Mammen and Spokoiny [25], and Picard and Tribouley [30] for other references which include the situation where the degree of smoothness is not known in advance. Such consequences are worked out in the second part of the paper. To situate the present work in the related literature it is important to note that *we do not assume any local Hölder property.* The first objective of the paper is to provide a consistent estimator of the index α_x from a sample of multivariate observations. This study is somewhat related to the study of the fractal dimension of μ (see

Cutler and Dawson [7], Davies and Hall [8], and Ferraty and Vieu [17]).

If f is three times continuously differentiable in a neighborhood of $x \in \mathbb{R}^d$, a Taylor series expansion shows that

$$\frac{\mu\big(B(x,\delta)\big)}{\lambda\big(B(x,\delta)\big)} = f(x) + C_x\,\delta^2 + \mathrm{o}(\delta^2) \quad \text{as } \delta \downarrow 0,$$

where

$$C_x = \frac{V_d^{2/d}}{2\pi(d+2)}\,\Gamma^{2/d}\Big(\frac{d}{2}+1\Big)\sum_{i=1}^{d}\frac{\partial^2 f}{\partial^2 x_i}(x)\,, \tag{1.4}$$

with V_d denoting the volume of the unit ball in $\mathbb{R}^d$ and Γ the Gamma function. See Fukunaga and Hostetler [20] or Bosq and Lecoutre [5] for calculation details. Therefore, in this case, formula (1.3) holds with $\alpha_x = 2$ and C_x given by (1.4), provided $\sum_{i=1}^{d}\partial^2 f(x)/\partial^2 x_i \neq 0$. The optimization of the number of nearest neighbors has mostly been studied in this restricted context, see for example Fukunaga and Hostetler [20].

In Section 2 we define an estimator $\hat{\alpha}_{n,x}$ of α_x based on nearest neighbor density estimators and derive weak consistency and asymptotic normality. This estimator is inspired by a proposal by Drees and Kaufmann [13] which itself is a generalization of a Pickands' type estimator [31], a basic tool in extreme value statistics. (For additional references, see Dekkers and de Haan [9], and Embrechts, Klüppelberg and Mikosch [15]). The estimation problem under study has some similarity with the following extreme value problem. Consider for instance the estimation of the second-order parameter $\rho < 0$ in univariate Pareto-type tail models where the quantile function Q satisfies the so-called *Hall condition* (see Hall and Welsh [23]):

$$Q(1-p) = Cp^{-\gamma}\big(1 + Dp^{-\rho} + \mathrm{o}(p^{-\rho})\big) \quad \text{as } p \downarrow 0\,, \tag{1.5}$$

with $\gamma > 0$, $C > 0$, $D \neq 0$. Drees and Kaufmann [13] introduced a Pickands' type estimator for ρ based on extreme order statistics from an univariate *i.i.d.* sample. Recently, Gomes, de Haan and Peng [21] considered the estimation problem of ρ in more detail and proposed more sophisticated estimators. Note also that although our problem is clearly connected to tail index estimation, one can prove that any attempt to reduce it to such a procedure comes up against the non-knowledge of α_x itself. The analogy (already observed in Hall [22]) and differences will be further explored in Section 3.

In Section 3 indeed we provide a regression model for scaled differences of successive nearest neighbor estimators in the spirit of Feuerverger and Hall [18], and Beirlant, Dierckx, Goegebeur and Matthys [3], where a similar representation was introduced in the context of estimation of the Pareto index γ within model (1.5). In this way we hope to clarify the similarities and differences between nonparametric methods in density estimation and extreme value methodology. We show how (estimates of) the regularity index can be used to adapt the nearest neighbor density estimator so that its behavior becomes much more stable as a function of the number of nearest neighbors. This is obtained by estimating the density from this regression model. Alternatively, the regularity index can also be of use when selecting the optimal number of neighbors when using the classical nearest neighbor density estimation in the general model (1.3). Results of simulations are presented in Section 4. Section 5 is devoted to the proofs of some technical results. In subsequent work extensions to kernel, wavelet based techniques and nonparametric regression will be explored.

1.2 Examples

Throughout this paragraph, the underlying norm is the standard Euclidean norm.

• Let f_1 be the probability density defined for $t \in \mathbb{R}$ by

$$f_1(t) = \frac{1}{2}\, e^{-|t|}.$$

This density is continuous but not differentiable at the point 0 and one easily obtains

$$\frac{\mu_1\big(B(0,\delta)\big)}{\lambda\big(B(0,\delta)\big)} = \frac{1}{2} - \frac{\delta}{4} + \mathrm{o}(\delta) \quad \text{as } \delta \downarrow 0.$$

Thus 0 is a Lebesgue point of μ_1 satisfying (1.3) with $C_0 = -1/4$ and $\alpha_0 = 1$.

• Let f_2 be the bivariate generalization of f_1 given by

$$f_2(t_1, t_2) = \frac{1}{2\pi}\, e^{-\sqrt{t_1^2 + t_2^2}}, \quad (t_1, t_2) \in \mathbb{R}^2\,.$$

This density is continuous but not differentiable at the point $(0, 0)$ and a similar calculation as in the case f_1 leads to

$$\frac{\mu_2\big(B(0,\delta)\big)}{\lambda\big(B(0,\delta)\big)} = \frac{1}{2\pi} - \frac{\delta}{3\pi} + \mathrm{o}(\delta) \quad \text{as } \delta \downarrow 0.$$

• This third example proves that the existence of α_x in (1.3) does not imply any pointwise Hölder property.

The following density model, defined on $[-1,1]$, has neither a left or a right limit at the point 0, but satisfies condition (1.3) with $C_0 = 2/c$ and $\alpha_0 = 1/2$.

$$f_3(t) = \begin{cases} [\sqrt{|t|} - \cos(1/t) + 2t\sin(1/t) + 2]/c & \text{for } t \in [-1,1], t \neq 0 \\ 2/c & \text{for } t = 0\,, \end{cases}$$

where $c = 4 + 4/3 + 2\sin(1)$. The term $\cos(1/t)$ is responsible for the discontinuity of second kind at 0. The part $-\cos(1/t) + 2t\sin(1/t)$ leads to a development of the measure at 0 of the form $\delta\sin(1/\delta)$. On the other hand, the term $\sqrt{|t|}$ yields the regularity index $\alpha_0 = 1/2$:

$$\begin{aligned} \frac{\mu_3\big(B(0,\delta)\big)}{\lambda\big(B(0,\delta)\big)} &= \frac{2}{c} + \frac{2}{3\,c}\,\delta^{1/2} + \frac{\delta}{c}\sin\Big(\frac{1}{\delta}\Big) \\ &= \frac{2}{c} + \frac{2}{3\,c}\,\delta^{1/2} + \mathrm{o}(\delta^{1/2}) \quad \text{as } \delta \downarrow 0\,. \end{aligned}$$

1.3 Application to mode estimation in non-smooth case

In two recent papers Abraham, Biau and Cadre [1], [2] study a simple, but quite efficient, estimator of the mode of a multivariate density. Supposing that the unknown density f has a mode at θ, they consider an estimator θ_n of θ introduced by Devroye [10]. This estimator is obtained by maximizing a kernel estimate (based on an *i.i.d.* sample) not over $\mathbb{R}^d$ but only *over the observed sample*. In the smooth case (f twice continuously differentiable in a neighborhood of θ), they prove that this estimator behaves asymptotically as well as any maximizer of the kernel estimate over $\mathbb{R}^d$ and derive its asymptotic normality. In the non-smooth case (no differentiability condition imposed on f around the mode), they prove strong consistency and give asymptotic bounds for the probability of deviation of θ_n. However, as they say themselves, their result is useless in practice in the non-smooth case. Indeed, the rate of convergence of the estimator of the mode is shown to depend on some constant κ measuring the sharpness of the density around the mode θ. Abraham *et al* [1] call this constant the *peak index* of the density, defined as follows: for any $\varepsilon > 0$, consider the level set

$$A(\varepsilon) = \big\{x \in \mathbb{R}^d : f(x) > f(\theta) - \varepsilon\big\}$$

and its diameter

$$D(\varepsilon) = \sup\big\{\,\|x-y\| : x \in A(\varepsilon), y \in A(\varepsilon)\big\}.$$

The constant $\kappa > 0$ is called the peak index of the density at its mode if we have

$$0 < \liminf_{\varepsilon \downarrow 0} \frac{D(\varepsilon)}{\varepsilon^\kappa} \leq \limsup_{\varepsilon \downarrow 0} \frac{D(\varepsilon)}{\varepsilon^\kappa} < \infty.$$

When κ does exist it is unique. In the non-smooth case (no differentiability condition or even no local Hölder property assumed) any attempt to build confidence intervals for the mode from the results of Abraham *et al* requires estimation of the peak index. Up to now no estimator has been proposed for κ. Actually the parameter κ appears to be in most cases the *inverse of the regularity index* of the measure associated with the unknown density f, leading to another application of the estimation procedure considered in this paper. To be more precise, let us particularize to the easy case $d = 1$. Then we have the following:

If $D(\varepsilon)$ tends to 0 with ε, and the density f satisfies

$$f(x) = f(\theta) + a_p|x - \theta|^p + \mathrm{o}\big(|x - \theta|^p\big) \quad \text{as } |x - \theta| \downarrow 0,$$

where $p > 0$ and $a_p < 0$, then, at the point θ, the density f has a peak index κ, the associated measure has a regularity index α and we have

$$\kappa = \frac{1}{\alpha} = \frac{1}{p}.$$

To understand this, note that if the measure associated with f has regularity index α at θ then $f(\theta+\delta)-f(\theta)$ is of order δ^α. It follows that $f(\theta+\varepsilon^{1/\alpha})-f(\theta)$ is of order ε. Therefore $D(\varepsilon)$ is of order $\varepsilon^{1/\alpha}$ and f has a peak index $1/\alpha$ at its mode θ.

2 Estimation of the regularity index

2.1 Definition of the estimator

In order to estimate the regularity index α_x in model (1.3), we consider a sequence $(X_n)_{n\geq 1}$ of multivariate independent random variables defined on some probability space $(\Omega, \mathcal{A}, \mathbf{P})$ with common probability distribution μ. The probability μ is assumed to have a density f with respect to the Lebesgue measure λ on $\mathbb{R}^d$. Let $(k_n)_{n\geq 1}$ be a sequence of positive integers. The *nearest neighbor estimator* of f at the point x is defined by

$$f_{k_n}(x) = \frac{k_n}{n\lambda\big(\overline{B}_{k_n}(x)\big)}\,, \tag{2.1}$$

where $\overline{B}_{k_n}(x)$ is the smallest closed ball with center at x containing at least k_n sample points. $\overline{B}_{k_n}(x)$ is random through $X_1, \dots, X_n$. The integer k_n plays the role of a smoothing parameter (oversmoothing when k_n is chosen too large, and undersmoothing in the opposite case). In discriminatory analysis, Fix and Hodges [19] introduced the classification rule based on nearest neighbor (see also Devroye, Györfi and Lugosi [12]). As to the nearest neighbor density estimator, it was studied by Loftsgaarden and Quesenberry [27], and Moore and Yackel [29]. For additional results and references, see Bosq and Lecoutre [5], Collomb, Hassani, Sarda and Vieu [6], and Levallois [26].

The estimator (2.1) may be rewritten as follows:

$$f_{k_n}(x) = \frac{k_n}{nV_d \, \|X_{(k_n)}(x) - x\|^d} \, ,$$

where $X_{(k_n)}(x)$ is the k_n-nearest neighbor of x. In case of a distance tie, the candidate with the smaller subscript is said to be closer to x. The key to the estimation of the index α_x is contained in the following proposition. It provides a way to isolate α_x from (1.3) based on a combination of the relative measures (μ/λ) for three open balls with radius δ, $\tau\delta$ and $\tau^2\delta$ for some $\tau > 1$, and $\delta > 0$. This idea can already be traced in Pickands [31].

Proposition 2.1 *Let $x \in \mathbb{R}^d$ be a Lebesgue point of μ satisfying condition (1.3). Then, for any $\tau > 1$,*

$$\lim_{\delta \downarrow 0} \frac{\varphi_{\tau^2\delta}(x) - \varphi_{\tau\delta}(x)}{\varphi_{\tau\delta}(x) - \varphi_{\delta}(x)} = \tau^{\alpha_x},$$

where we denote

$$\varphi_\delta(x) = \frac{\mu\big(B(x,\delta)\big)}{\lambda\big(B(x,\delta)\big)} \, .$$

Proof of Proposition 2.1 In accordance with model (1.3), we can write

$$\varphi_{\tau^2\delta}(x) - \varphi_{\tau\delta}(x) = C_x \left(\tau^{2\alpha_x} - \tau^{\alpha_x}\right) \delta^{\alpha_x} + \mathrm{o}(\delta^{\alpha_x}) \quad \text{as } \delta \downarrow 0. \tag{2.2}$$

Similarly,

$$\varphi_{\tau\delta}(x) - \varphi_{\delta}(x) = C_x \left(\tau^{\alpha_x} - 1\right) \delta^{\alpha_x} + \mathrm{o}(\delta^{\alpha_x}) \quad \text{as } \delta \downarrow 0. \tag{2.3}$$

Since $C_x \neq 0$, we deduce from (2.2) and (2.3) that

$$\lim_{\delta \downarrow 0} \frac{\varphi_{\tau^2\delta}(x) - \varphi_{\tau\delta}(x)}{\varphi_{\tau\delta}(x) - \varphi_{\delta}(x)} = \frac{C_x \left(\tau^{2\alpha_x} - \tau^{\alpha_x}\right)}{C_x \left(\tau^{\alpha_x} - 1\right)} = \tau^{\alpha_x}.$$

$\blacksquare$

Motivated by the estimator (2.1) and Proposition 2.1 we now define an estimator $\hat{\alpha}_{n,x}$ of α_x as follows:

$$\hat{\alpha}_{n,x} = \frac{d}{\log \tau} \log \frac{f_{\lfloor \tau^2 k_n \rfloor}(x) - f_{\lfloor \tau k_n \rfloor}(x)}{f_{\lfloor \tau k_n \rfloor}(x) - f_{k_n}(x)} \tag{2.4}$$

if $\left[f_{\lfloor \tau^2 k_n \rfloor}(x) - f_{\lfloor \tau k_n \rfloor}(x)\right] / \left[f_{\lfloor \tau k_n \rfloor}(x) - f_{k_n}(x)\right] > 0$ and $\hat{\alpha}_{n,x} = 0$ otherwise. The notation $\lfloor . \rfloor$ stands for the integer part function. This estimator is analogous to the estimator proposed in Drees and Kaufmann [13] estimating ρ in (1.5), which itself is reminiscent of Pickands' estimator [31], well known in the theory of extremal events (Dekkers and de Haan [9], Embrechts, Klüppelberg and Mikosch [15]). In fact Pickands considered $\tau = 2$. Recently, more sophisticated estimators of ρ were proposed in Gomes, de Haan and Peng [21]. It is not clear however how those estimators can be transformed to the setting of density estimation.

2.2 Asymptotic results

The first theorem implies the weak consistency of $\hat{\alpha}_{n,x}$ towards α_x. We recall that, throughout the paper, the probability measure μ is assumed to have a density f with respect to the Lebesgue measure λ on $\mathbb{R}^d$ and that f is a version of the density that coincides with (1.1) if the limit exists.

Theorem 2.1 *Let $x \in \mathbb{R}^d$ be a Lebesgue point of μ satisfying condition (1.3). Assume that $f(x) > 0$. Then, under the conditions*

$$\lim_{n\to\infty} k_n = \infty, \quad \lim_{n\to\infty} \frac{k_n}{n} = 0 \quad \textit{and} \quad \lim_{n\to\infty} \frac{k_n^{\alpha_x + d/2}}{n^{\alpha_x}} = \infty\,,$$

we have

$$\frac{f_{\lfloor \tau^2 k_n \rfloor}(x) - f_{\lfloor \tau k_n \rfloor}(x)}{f_{\lfloor \tau k_n \rfloor}(x) - f_{k_n}(x)} \to \tau^{\alpha_x/d} \quad \textit{in probability.}$$

Remark 1 The unknown parameter α_x appears in the asymptotic condition $\lim_{n\to\infty} k_n^{\alpha_x+d/2}/n^{\alpha_x} = \infty$. This is to give minimal conditions. Of course, another condition on the model could be provided, for example $\lim_{n\to\infty} k_n \log n / n = \infty$. Note also that the condition $\lim_{n\to\infty} k_n^{\alpha_x+d/2}/n^{\alpha_x} = \infty$ is comparable to condition (2.11) in Gomes, de Haan and Peng [21]. It states that the number of nearest neighbors to be used in the estimation of α_x should not be too small. This will be confirmed by the simulations in the final section.

Let us now state some technical results that are used in the proof of consistency. For clarity, proofs of these results have been postponed to Section 5. Throughout, we denote by supp μ the support of μ.

Proposition 2.2 *Let $x \in \mathbb{R}^d$.*

(**A**) *Assume that x belongs to* supp μ*. If*

$$\lim_{n\to\infty} \frac{k_n}{n} = 0\,,$$

then

$$\|X_{(k_n)}(x) - x\| \to 0 \quad \mathbf{P}\text{-}a.s.$$

(**B**) *Assume that x is a Lebesgue point of μ and that*

$$\lim_{n\to\infty} k_n = \infty \quad \text{and} \quad \lim_{n\to\infty} \frac{k_n}{n} = 0\,.$$

Then

(i) *$f_{k_n}(x)$ is a weak consistent estimator of $f(x)$,* i.e.,

$$f_{k_n}(x) \to f(x) \quad \text{in probability.}$$

(ii) *Denoting by $\stackrel{\mathcal{D}}{\to}$ the convergence in distribution and $\mathcal{N}$ the Gaussian distribution, we have*

$$\frac{n}{\sqrt{k_n}}\left(\frac{k_n}{n} - \mu\big(\overline{B}_{k_n}(x)\big)\right) \stackrel{\mathcal{D}}{\to} \mathcal{N}(0,1)\,.$$

(iii) *Suppose that $f(x) > 0$. Then, for any $\tau_1, \tau_2 > 1$,*

$$\frac{\|X_{(\lfloor \tau_1 k_n \rfloor)}(x) - x\|}{\|X_{(\lfloor \tau_2 k_n \rfloor)}(x) - x\|} \to \left(\frac{\tau_1}{\tau_2}\right)^{1/d} \quad \text{in probability.}$$

(iv) *Under the additional condition that*

$$\lim_{n\to\infty} \frac{k_n^{\alpha_x + d/2}}{n^{\alpha_x}} = \infty\,,$$

we have

$$\frac{k_n/n - \mu\big(\overline{B}_{k_n}(x)\big)}{V_d\,\|X_{(k_n)}(x) - x\|^d} = \mathrm{o}_{\mathbf{P}}\big(\|X_{(k_n)}(x) - x\|^{\alpha_x}\big) \quad \text{as } n \to \infty\,.$$

Proof of Theorem 2.1 Note first that the assumption $f(x) > 0$ forces the Lebesgue point x to belong to supp μ. We shall assume, without loss of generality, that $x = 0$. Note also that since μ is absolutely continuous with respect to λ, one has

$$\mu\big(\overline{B}_{k_n}(0)\big) = \mu\big(B_{k_n}(0)\big),$$

where $B_{k_n}(0)$ is the (random) open ball with center at 0 and radius $\|X_{(k_n)}(0)\|$. We can write, using condition (1.3) and Proposition 2.2 (**A**),

$$\begin{aligned}
&f_{k_n}(0)\\
&\quad= \frac{k_n}{nV_d\,\|X_{(k_n)}(0)\|^d}\\
&\quad= \frac{k_n/n - \mu\big(\overline{B}_{k_n}(0)\big) + \mu\big(\overline{B}_{k_n}(0)\big)}{V_d\,\|X_{(k_n)}(0)\|^d}\\
&\quad= \frac{k_n/n - \mu\big(\overline{B}_{k_n}(0)\big)}{V_d\,\|X_{(k_n)}(0)\|^d} + f(0) + C_0\,\|X_{(k_n)}(0)\|^{\alpha_0} + \mathrm{o}\big(\|X_{(k_n)}(0)\|^{\alpha_0}\big) \quad \mathbf{P}\text{-a.s.}
\end{aligned}$$

as $n \to \infty$. Applying Proposition 2.2 (**B**)(iv) leads to

$$f_{k_n}(0) = f(0) + C_0\,\|X_{(k_n)}(0)\|^{\alpha_0} + \mathrm{o}_{\mathbf{P}}\big(\|X_{(k_n)}(0)\|^{\alpha_0}\big)$$

as $n \to \infty$. We finally obtain

$$\begin{aligned}
&\frac{f_{\lfloor\tau^2 k_n\rfloor}(0) - f_{\lfloor\tau k_n\rfloor}(0)}{f_{\lfloor\tau k_n\rfloor}(0) - f_{k_n}(0)}\\
&\quad= \frac{\|X_{(\lfloor\tau^2 k_n\rfloor)}(0)\|^{\alpha_0} - \|X_{(\lfloor\tau k_n\rfloor)}(0)\|^{\alpha_0} + \mathrm{o}_{\mathbf{P}}\big(\|X_{(\lfloor\tau^2 k_n\rfloor)}(0)\|^{\alpha_0}\big) + \mathrm{o}_{\mathbf{P}}\big(\|X_{(\lfloor\tau k_n\rfloor)}(0)\|^{\alpha_0}\big)}{\|X_{(\lfloor\tau k_n\rfloor)}(0)\|^{\alpha_0} - \|X_{(k_n)}(0)\|^{\alpha_0} + \mathrm{o}_{\mathbf{P}}\big(\|X_{(\lfloor\tau k_n\rfloor)}(0)\|^{\alpha_0}\big) + \mathrm{o}_{\mathbf{P}}\big(\|X_{(k_n)}(0)\|^{\alpha_0}\big)}
\end{aligned}$$

as $n \to \infty$. Proposition 2.2 (**B**)(iii) leads to the desired conclusion. ■

We now turn to the discussion of the asymptotic normality of the estimator $\hat{\alpha}_{n,x}$. To this end, we will make use of the following proposition which completes Proposition 2.2 (**A**) and (**B**)(ii). The first part of this proposition follows from the combination of the corresponding approximation result for the tail uniform empirical process (see for instance Mason [28]) with Vervaat's lemma (see Shorack and Wellner [33], page 659). This then leads to an approximation for the tail uniform quantile process $\big\{U_{\lfloor tk_n\rfloor,n}\,;\, 0 \le t \le 1\big\}$ with $U_{i,n}$ denoting the i-th order statistic from a uniform $(0,1)$ sample of size n. Since the process $\big\{\mu\big(\overline{B}_{\lfloor tk_n\rfloor}(x)\big)\,;\, 0 \le t \le 1\big\}$ is equal in distribution to this tail uniform quantile process, the first statement follows via a special construction. The second part of Proposition 2.3 follows from a direct application of Theorem 1.2 in Einmahl and Mason [14] on generalized quantiles being defined here as the Lebesgue measure $V_d\,\|X_{(\lfloor tn\rfloor)}(x)-x\|^d$ of the smallest ball with center at x which contains at least $\lfloor tn\rfloor$ observations. Under (1.3) the theoretical quantile function is then given by $U_x(t)=\big(t/f(x)\big)\big[1-C_x\,V_d^{-\alpha_x/d}f^{-1-\alpha_x/d}(x)t^{\alpha_x/d}\big]$.

Proposition 2.3 *There exists, on an appropriate probability space, an* i.i.d. *sequence* $(X_n)_{n\geq 1}$ *with density* f *and*

(*i*) *a sequence of standard Wiener processes* $(W_x^{(n)})_{n\geq1}$, *such that*

$$\sup_{0\leq t\leq 1}\left|\sqrt{k_n}\left[t-\frac{n\mu\big(\overline{B}_{\lfloor tk_n\rfloor}(x)\big)}{k_n}\right]+W_x^{(n)}(t)\right|\to 0\quad \textit{in probability}$$

as

$$k_n\to\infty\quad \textit{and}\quad \frac{k_n}{n}\to 0\,;$$

(*ii*) *a sequence of Brownian bridges* $(B_x^{(n)})_{n\geq1}$, *such that*

$$\sup_{0<t<1}\left|\sqrt{n}f(x)\Big[V_d\,\|X_{(\lfloor tn\rfloor)}(x)-x\|^d-U_x(t)\Big]+B_x^{(n)}(t)\right|\to 0$$

in probability as $n\to\infty$.

In the sequel, we write

$$Q_{n,k_n}(t):=\sqrt{k_n}\left[t-\frac{n\mu\big(\overline{B}_{\lfloor tk_n\rfloor}(x)\big)}{k_n}\right]. \tag{2.5}$$

Next we specify condition (1.3) to

$$\frac{\mu\big(B(x,\delta)\big)}{\lambda\big(B(x,\delta)\big)}=f(x)+C_x\,\delta^{\alpha_x}+D_x\,\delta^{\beta_x}+\mathrm{o}(\delta^{\beta_x})\quad \text{as } \delta\downarrow 0, \tag{2.6}$$

where C_x, D_x are non-zero constants, and α_x, β_x are positive real numbers satisfying $\alpha_x<\beta_x$.

Theorem 2.2 *Let* $x\in\mathbb{R}^d$ *be a Lebesgue point of* μ *satisfying condition (2.6). Assume that* $f(x)>0$. *Then, under the conditions*

$$\lim_{n\to\infty}k_n=\infty,\quad \lim_{n\to\infty}\frac{k_n}{n}=0,\quad \lim_{n\to\infty}\frac{k_n^{\alpha_x+d/2}}{n^{\alpha_x}}=\infty$$

and

$$\lim_{n\to\infty}\frac{k_n^{\beta_x+d/2}}{n^{\beta_x}}=0,\quad \lim_{n\to\infty}\frac{k_n^{2\alpha_x+d/2}}{n^{2\alpha_x}}=0,$$

we have

$$\frac{k_n^{1/2+\alpha_x/d}}{n^{\alpha_x/d}}\left(\hat{\alpha}_{n,x}-\alpha_x\right)\xrightarrow{\mathcal{D}}\mathcal{N}(0,\sigma_x^2)\,,$$

where

$$\sigma_x^2=\Big(\frac{d}{C_x\log\tau}\,\frac{V_d^{\alpha_x/d}}{\tau^{1+\alpha_x/d}(\tau^{\alpha_x/d}-1)}\,f^{1+\alpha_x/d}(x)\Big)^2\,(1+\tau^{1+2\alpha_x/d})\,(\tau-1).$$

Proof of Theorem 2.2 We consider for simplicity $x=0$. The proof runs in two steps. First we derive the asymptotic normality of

$$T^{(1)}_{n,k_n}=\frac{k_n^{1/2+\alpha_0/d}}{n^{\alpha_0/d}}\left(\hat{\alpha}_{n,0}-\tilde{\alpha}_{n,0}\right),$$

where

$$\tilde{\alpha}_{n,0}=\frac{d}{\log\tau}\log\frac{\phi_{\lfloor\tau^2k_n\rfloor}(0)-\phi_{\lfloor\tau k_n\rfloor}(0)}{\phi_{\lfloor\tau k_n\rfloor}(0)-\phi_{k_n}(0)}$$

with

$$\phi_{k_n}(0)=\frac{\mu\big(\overline{B}_{k_n}(0)\big)}{\lambda\big(\overline{B}_{k_n}(0)\big)}\,.$$

Secondly we will show that under the given conditions

$$T^{(2)}_{n,k_n}=\frac{k_n^{1/2+\alpha_0/d}}{n^{\alpha_0/d}}\left(\tilde{\alpha}_{n,0}-\alpha_0\right)\to 0\quad\text{in probability.}\tag{2.7}$$

In order to apply Proposition 2.3 we consider $T^{(1)}_{n,\lfloor k_n/\tau^2\rfloor}$. Using the delta method (second derivative terms can be shown to be asymptotically negligible when $k_n^{1/2+\alpha_0/d}/n^{\alpha_0/d}\to\infty$) we find that

$$\frac{\log\tau}{d}\,\tau^{1+2\alpha_0/d}\,T^{(1)}_{n,\lfloor k_n/\tau^2\rfloor}$$

has the same asymptotic distribution as

$$\begin{aligned}\frac{k_n^{1/2+\alpha_0/d}}{n^{\alpha_0/d}}\Bigg[&\frac{f_{k_n}(0)-\phi_{k_n}(0)}{\phi_{k_n}(0)-\phi_{\lfloor k_n/\tau\rfloor}(0)}\\ &-\big(f_{\lfloor k_n/\tau\rfloor}(0)-\phi_{\lfloor k_n/\tau\rfloor}(0)\big)\left(\frac{1}{\phi_{k_n}(0)-\phi_{\lfloor k_n/\tau\rfloor}(0)}+\frac{1}{\phi_{\lfloor k_n/\tau\rfloor}(0)-\phi_{\lfloor k_n/\tau^2\rfloor}(0)}\right)\\ &+\frac{f_{\lfloor k_n/\tau^2\rfloor}(0)-\phi_{\lfloor k_n/\tau^2\rfloor}(0)}{\phi_{\lfloor k_n/\tau\rfloor}(0)-\phi_{\lfloor k_n/\tau^2\rfloor}(0)}\Bigg],\end{aligned}$$

or, equivalently, as

$$\frac{V_d^{\alpha_0/d}}{C_0}\Bigg[\frac{Q_{n,k_n}(1)\,k_n^{1+\alpha_0/d}}{n^{1+\alpha_0/d}V_d^{1+\alpha_0/d}\,\|X_{(k_n)}(0)\|^{d+\alpha_0}}\cdot\frac{1}{\Big(1-\frac{\|X_{(\lfloor k_n/\tau\rfloor)}(0)\|^{\alpha_0}}{\|X_{(k_n)}(0)\|^{\alpha_0}}\Big)}$$

$$-\frac{Q_{n,k_n}(1/\tau)\,k_n^{1+\alpha_0/d}}{n^{1+\alpha_0/d}V_d^{1+\alpha_0/d}\,\|X_{(\lfloor k_n/\tau\rfloor)}(0)\|^{d+\alpha_0}}\cdot\Bigg(\frac{1}{\frac{\|X_{(k_n)}(0)\|^{\alpha_0}}{\|X_{(\lfloor k_n/\tau\rfloor)}(0)\|^{\alpha_0}}-1}+\frac{1}{1-\frac{\|X_{(\lfloor k_n/\tau^2\rfloor)}(0)\|^{\alpha_0}}{\|X_{(\lfloor k_n/\tau\rfloor)}(0)\|^{\alpha_0}}}\Bigg)$$

$$+\frac{Q_{n,k_n}(1/\tau^2)\,k_n^{1+\alpha_0/d}}{n^{1+\alpha_0/d}V_d^{1+\alpha_0/d}\,\|X_{(\lfloor k_n/\tau^2\rfloor)}(0)\|^{d+\alpha_0}}\cdot\frac{1}{\Big(\frac{\|X_{(\lfloor k_n/\tau\rfloor)}(0)\|^{\alpha_0}}{\|X_{(\lfloor k_n/\tau^2\rfloor)}(0)\|^{\alpha_0}}-1\Big)}\Bigg].$$

Using Proposition 2.3 (i), Proposition 2.2 $(\mathbf{B})(iii)$, and the consistency of the nearest neighbor density estimator, we can approximate this by

$$\frac{V_d^{\alpha_0/d}\,f^{1+\alpha_0/d}(0)}{C_0}\Bigg[W_0^{(n)}(1)\frac{\tau^{\alpha_0/d}}{\tau^{\alpha_0/d}-1}-W_0^{(n)}(1/\tau)\,\tau^{1+\alpha_0/d}\left(\frac{\tau^{\alpha_0/d}+1}{\tau^{\alpha_0/d}-1}\right)$$
$$+W_0^{(n)}(1/\tau^2)\,\tau^2\,\frac{\tau^{2\alpha_0/d}}{\tau^{\alpha_0/d}-1}\Bigg].$$

This leads to the asymptotic Gaussian distribution given in the statement of Theorem 2.2.

Using Proposition 2.3 (ii) we now find that for some constant M

$$\|X_{(k_n)}(0)\|=\left(\frac{k_n}{nV_d\,f(0)}\right)^{1/d}\left[1+M\Big(\frac{k_n}{n}\Big)^{\alpha_0/d}\big(1+o_{\mathbf{P}}(1)\big)+o_{\mathbf{P}}(1)\right].$$

Hence, using (2.6), we have that (with M_1, M_2, $\tilde M_1$, $\tilde M_2$ denoting some constants that can change values from line to line)

$$\begin{aligned}
&\tilde\alpha_{n,0}\\
&=\frac{d}{\log\tau}\log\Bigg[\frac{(\tau^2k_n/n)^{\alpha_0/d}-(\tau k_n/n)^{\alpha_0/d}+M_1(\tau^2k_n/n)^{2\alpha_0/d}-M_1(\tau k_n/n)^{2\alpha_0/d}}{(\tau k_n/n)^{\alpha_0/d}-(k_n/n)^{\alpha_0/d}+M_1(\tau k_n/n)^{2\alpha_0/d}-M_1(k_n/n)^{2\alpha_0/d}}\\
&\qquad\qquad\frac{+M_2(\tau^2k_n/n)^{\beta_0/d}-M_2(\tau k_n/n)^{\beta_0/d}+o_{\mathbf{P}}\big((k_n/n)^{\beta_0/d}\big)}{+M_2(\tau k_n/n)^{\beta_0/d}-M_2(k_n/n)^{\beta_0/d}+o_{\mathbf{P}}\big((k_n/n)^{\beta_0/d}\big)}\Bigg]\\
&=\frac{d}{\log\tau}\log\Bigg[\tau^{\alpha_0/d}\Bigg(\frac{1+M_1(k_n/n)^{\alpha_0/d}+M_2(k_n/n)^{(\beta_0-\alpha_0)/d}+o_{\mathbf{P}}\big((k_n/n)^{(\beta_0-\alpha_0)/d}\big)}{1+\tilde M_1(k_n/n)^{\alpha_0/d}+\tilde M_2(k_n/n)^{(\beta_0-\alpha_0)/d}+o_{\mathbf{P}}\big((k_n/n)^{(\beta_0-\alpha_0)/d}\big)}\Bigg)\Bigg]\\
&=\alpha_0+M_1\Big(\frac{k_n}{n}\Big)^{\alpha_0/d}+M_2\Big(\frac{k_n}{n}\Big)^{(\beta_0-\alpha_0)/d}+o_{\mathbf{P}}\big((k_n/n)^{(\beta_0-\alpha_0)/d}\big).
\end{aligned}$$

So,

$$\frac{k_n^{1/2+\alpha_0/d}}{n^{\alpha_0/d}}\,\big(\tilde\alpha_{n,0}-\alpha_0\big)$$
$$= M_1\Big(\frac{k_n}{n}\Big)^{2\alpha_0/d}k_n^{1/2}+M_2\Big(\frac{k_n}{n}\Big)^{\beta_0/d}k_n^{1/2}+\mathrm{o}_{\mathbf{P}}\bigg(\Big(\frac{k_n}{n}\Big)^{\beta_0/d}k_n^{1/2}\bigg),$$

from which (2.7) follows. ■

Remark 2 Specifying M_1 and M_2 in the last formula of the above proof with the help of Proposition 2.3 (*ii*) one finds that

$$\tilde\alpha_{n,0}-\alpha_0$$
$$=\frac{d}{\log\tau}\bigg[\Big(\frac{k_n}{n}\Big)^{\alpha_0/d}\frac{\alpha_0\,C_0}{d\,V_d^{\alpha_0/d}f^{1+2\alpha_0/d}(0)}(1-\tau^{2\alpha_0/d})$$
$$+\frac{D_0}{C_0}\big(f(0)V_d\big)^{(\alpha_0-\beta_0)/d}\Big(\frac{k_n}{n}\Big)^{(\beta_0-\alpha_0)/d}\frac{(\tau^{\beta_0/d}-1)(\tau^{(\beta_0-\alpha_0)/d}-1)}{\tau^{\alpha_0/d}-1}\bigg]\big(1+\mathrm{o}_{\mathbf{P}}(1)\big)$$

as $n, k_n\to\infty$, and $k_n/n\to 0$. In the important case where $\beta=2\alpha$ we find that the asymptotic bias equals

$$\frac{d}{\log\tau}\Big(\frac{k_n}{n}\Big)^{\alpha_0/d}(\tau^{2\alpha_0/d}-1)\big(f(0)V_d\big)^{-\alpha_0/d}\bigg[\frac{D_0}{C_0}-\frac{\alpha_0C_0}{d}f^{-1-\alpha_0/d}(0)\bigg],$$

whose absolute value increases with $\tau\in(1,2]$. Consequently, the bias is smaller when τ is smaller than the original value $\tau=2$ proposed by Pickands [31]. Note that however then the asymptotic variance increases.

3 Optimization of the k_n-nearest neighbor density estimator

3.1 The asymptotic mean squared error

Here we are interested in a functional form of k_n in (2.1) that will essentially indicate how k_n should depend on the number of sample points n and the regularity index α_x. To obtain this, we optimize k_n with respect to the mean squared error criterion

$$\Delta_n(x)=\mathbf{E}\big(f_{k_n}(x)-f(x)\big)^2,$$

where the expectation is over the sample set $X_1,\ldots,X_n$. This theorem, proved in Section 5, generalizes the results presented in Fukunaga and Hostetler [20].

Theorem 3.1 *Let $x \in \mathbb{R}^d$ be a Lebesgue point of μ satisfying condition (1.3). Assume that supp μ is compact and $f(x) > 0$. Then, under the conditions*

$$\lim_{n\to\infty} k_n = \infty \quad \text{and} \quad \lim_{n\to\infty} \frac{k_n}{n} = 0,$$

the following asymptotic developments hold:

$$\mathbf{E} f_{k_n}(x) - f(x) = \frac{f(x)}{k_n} + \frac{C_x}{V_d^{\alpha_x/d}} f^{-\alpha_x/d}(x) \Big(\frac{k_n}{n}\Big)^{\alpha_x/d} + \mathrm{o}\Big(\frac{1}{k_n} + \Big(\frac{k_n}{n}\Big)^{\alpha_x/d}\Big)$$

and

$$\mathbf{Var} f_{k_n}(x) = \frac{f^2(x)}{k_n} + \mathrm{o}\Big(\frac{1}{k_n}\Big) \quad \text{as } n \to \infty\,.$$

Consequently,

$$\begin{aligned}\Delta_n(x) &= \mathbf{E}\big(f_{k_n}(x) - f(x)\big)^2 \\ &= \frac{f^2(x)}{k_n} + \frac{C_x^2}{V_d^{2\alpha_x/d}} f^{-2\alpha_x/d}(x) \Big(\frac{k_n}{n}\Big)^{2\alpha_x/d} + \mathrm{o}\Big(\frac{1}{k_n} + \Big(\frac{k_n}{n}\Big)^{2\alpha_x/d}\Big).\end{aligned}$$

Corollary 3.1 *Under the assumptions of Theorem 3.1, the value of k_n that minimizes the asymptotic mean squared error is*

$$k_n^* = \left(\frac{dV_d^{2\alpha_x/d}}{2\alpha_x C_x^2} f^{2+2\alpha_x/d}(x)\right)^{d/(d+2\alpha_x)} n^{2\alpha_x/(d+2\alpha_x)},$$

and the associated mean squared error is

$$\Delta_n^*(x) = \frac{d+2\alpha_x}{2\alpha_x}\left(\frac{2\alpha_x C_x^2}{dV_d^{2\alpha_x/d}} f^{2\alpha_x/d}(x)\right)^{d/(d+2\alpha_x)} n^{-2\alpha_x/(d+2\alpha_x)} + \mathrm{o}\big(n^{-2\alpha_x/(d+2\alpha_x)}\big).$$

Remark 3 The optimal k_n^* is seen to depend upon the dimension of the observation space and upon the characteristics of the symmetric derivative of μ at the point x, namely $f(x)$, C_x and α_x. The coefficient C_x can be interpreted as a measure of the variation of the underlying distribution. Thus, if C_x is large (indicating $f(x)$ has large second derivatives when f is sufficiently smooth, and is therefore changing rapidly in the region around x), we see that the optimal k_n is made smaller to compensate for this fact. We also observe that the optimal number of nearest neighbors is increasing with α_x, as expected.

Remark 4 The choice of a bandwidth in non-smooth cases has been considered for instance in van Es [16], who explored cross-validation bandwidths for kernel estimators. Below we present a new approach based on the regularity index discussed here.

3.2 A regression model and selection of the number k_n of nearest neighbors

In this paragraph we construct a regression model which allows to reduce the bias of a k_n-nearest neighbor density estimator. This enables to construct diagnostics for selecting the number k_n of nearest neighbors. We consider the following scaled differences of consecutive nearest neighbor density estimators at a given point:

$$Z_j := jf_j(x) - (j-1)f_{j-1}(x), \quad j \geq 1,$$

setting $0f_0(x) = 0$. From Proposition 2.3 and the basic condition (1.3), the following expansion can be derived for the variables Z_j (one can justify this in a formal way as will be seen in the proof of Theorem 3.2):

$$\begin{aligned}
Z_j &= \sqrt{k_n}\left[Q_{n,k_n}\Big(\frac{j}{k_n}\Big)f_j(x) - Q_{n,k_n}\Big(\frac{j-1}{k_n}\Big)f_{j-1}(x)\right] \\
&\quad + \left[j\,\frac{\mu\big(\overline{B}_j(x)\big)}{\lambda\big(\overline{B}_j(x)\big)} - (j-1)\,\frac{\mu\big(\overline{B}_{j-1}(x)\big)}{\lambda\big(\overline{B}_{j-1}(x)\big)}\right] \\
&\simeq \sqrt{k_n}\,f(x)\left[W_x\Big(\frac{j}{k_n}\Big) - W_x\Big(\frac{j-1}{k_n}\Big)\right] \\
&\quad + j\big[f(x) + C_x\,\|X_{(j)}(x) - x\|^{\alpha_x}\big] - (j-1)\big[f(x) + C_x\,\|X_{(j-1)}(x) - x\|^{\alpha_x}\big],
\end{aligned}$$

where Q_{n,k_n} is defined in (2.5). On the other hand, using the mean value theorem, we obtain that

$$\begin{aligned}
& j\,\|X_{(j)}(x) - x\|^{\alpha_x} - (j-1)\,\|X_{(j-1)}(x) - x\|^{\alpha_x} \\
&\quad = V_d^{-\alpha_x/d}\left[jf_j^{-\alpha_x/d}(x)\Big(\frac{j}{n}\Big)^{\alpha_x/d} - (j-1)f_{j-1}^{-\alpha_x/d}(x)\Big(\frac{j-1}{n}\Big)^{\alpha_x/d}\right] \\
&\quad \simeq V_d^{-\alpha_x/d}f^{-\alpha_x/d}(x)\Big(1 + \frac{\alpha_x}{d}\Big)\Big(\frac{j}{n}\Big)^{\alpha_x/d}.
\end{aligned}$$

Hence we are led to approximate Z_j by

$$Z_j \simeq f(x) + b_{k_n}\Big(\frac{j}{k_n}\Big)^{\alpha_x/d} + f(x)\,\varepsilon_j, \quad j = 1, \ldots, k_n, \tag{3.1}$$

where

$$b_{k_n} = C_x V_d^{-\alpha_x/d} f^{-\alpha_x/d}(x)\Big(1 + \frac{\alpha_x}{d}\Big)\Big(\frac{k_n}{n}\Big)^{\alpha_x/d}$$

and

$$\varepsilon_j = \sqrt{k_n}\left[W_x\Big(\frac{j}{k_n}\Big) - W_x\Big(\frac{j-1}{k_n}\Big)\right] \overset{i.i.d.}{\sim} \mathcal{N}(0,1).$$

Representation (3.1) yields a regression model with covariates j/k_n, $j = 1, \ldots, k_n$. Note also that $b_{k_n}/(1+\alpha_x/d)$ is the factor which dominates the bias of the nearest neighbor estimator as given in Theorem 3.1. Further, the regression model (3.1) can be exploited directly to propose an estimator for b_{k_n} using a least squares method, after replacing α_x for instance by the estimator $\hat{\alpha}_{n,x}$. In Feuerverger and Hall [18], and Beirlant, Dierckx, Goegebeur and Matthys [3] a representation similar to (3.1) was introduced in the context of estimation of a Pareto index γ within model (1.5). More precisely, denoting the order statistics of a sample by $X_{1,n} \leq \ldots \leq X_{n,n}$ and $\tilde{Z}_j := j(\log X_{n-j+1,n} - \log X_{n-j,n})$, $j = 1, \ldots, k_n$, with k_n/n small, these authors considered the model:

$$\tilde{Z}_j = \gamma + b_{k_n}\left(\frac{j}{k_n}\right)^{-\rho} + \gamma\,\tilde{\varepsilon}_j\,, \tag{3.2}$$

where b_{k_n} is of order $(n/k_n)^\rho$. Comparing (3.1) and (3.2) we observe the similarities between the estimation of a density f at a point x and of the extreme value index γ. Observe further that when estimating $f(x)$ (for instance by least squares) from the simple location model

$$Z_j \simeq f(x) + f(x)\,\varepsilon_j, \quad j = 1, \ldots, k_n,$$

(obtained by setting b_{k_n} equal to 0 in (3.1)), we are led to

$$\frac{1}{k_n}\sum_{j=1}^{k_n} Z_j = f_{k_n}(x).$$

Comparison of (3.1) and (3.2) is quite instructive in understanding the similarities between estimation of Pareto-type tails and density estimation:

- the role of $f(x)$ is taken over by the tail index γ in extreme value methodology;
- the nearest neighbor estimator $k_n^{-1}\sum_{j=1}^{k_n} Z_j = f_{k_n}(x)$ of $f(x)$ is the analogue of Hill's estimator [24] $k_n^{-1}\sum_{j=1}^{k_n}\tilde{Z}_j = k_n^{-1}\sum_{j=1}^{k_n}\big(\log X_{n-j+1,n} - \log X_{n-k_n,n}\big)$ of the tail index γ;
- the bias of the Hill's estimator is strongly influenced by the second-order parameter ρ, while for $f_{k_n}(x)$ the regularity index α_x is predominant.

By virtue of this analogy, methods that have been worked out recently in extreme value statistics can be carried over to density estimation. Here we illustrate this with two techniques: the adaptive choice of k_n when using the nearest neighbor estimator, and secondly, the proposal of a density estimator which reduces bias. Both techniques are based on a consistent estimator $\hat{\alpha}_{n,x}$. In accordance with Theorem 2.1, we let the number of nearest neighbors $\tilde{k}_n$ used in $\hat{\alpha}_{n,x}$ satisfy $\tilde{k}_n^{\alpha_x+d/2}/n^{\alpha_x} \to \infty$.

First, based on (3.1), one can propose a procedure to estimate k_n^*. Indeed, one easily checks that

$$k_n^* = (b_{k_0})^{-2d/(d+2\alpha_x)} k_0^{2\alpha_x/(d+2\alpha_x)} \left[f^2(x) \Big(1 + \frac{\alpha_x}{d}\Big)^2 \frac{d}{2\alpha_x} \right]^{d/(d+2\alpha_x)}$$

for any preliminary value k_0. An approximate least squares estimator of b_{k_0} is then given by

$$\hat{b}_{k_0} = \frac{(1+\hat{\alpha}_{n,x}/d)^2\,(1+2\hat{\alpha}_{n,x}/d)}{(\hat{\alpha}_{n,x}/d)^2} \times \frac{1}{k_0} \sum_{j=1}^{k_0} Z_j \bigg(\Big(\frac{j}{k_0}\Big)^{\hat{\alpha}_{n,x}/d} - \frac{1}{1+\hat{\alpha}_{n,x}/d} \bigg).$$

We find an estimator for k_n^* replacing b_{k_0}, α_x and $f(x)$ by their respective estimators $\hat{b}_{k_0}$, $\hat{\alpha}_{n,x}$ and $f_{k_0}(x)$:

$$\hat{k}_{n,k_0}^* = (\hat{b}_{k_0})^{-2d/(d+2\hat{\alpha}_{n,x})} k_0^{2\hat{\alpha}_{n,x}/(d+2\hat{\alpha}_{n,x})} \left[f_{k_0}^2(x) \Big(1 + \frac{\hat{\alpha}_{n,x}}{d}\Big)^2 \frac{d}{2\hat{\alpha}_{n,x}} \right]^{d/(d+2\hat{\alpha}_{n,x})}. \tag{3.3}$$

Let us now prove the consistency of this method.

Theorem 3.2 *Let $x \in \mathbb{R}^d$ be a Lebesgue point of μ satisfying condition (1.3). Assume that $f(x) > 0$. Then as $k_0 \to \infty$, $k_0/n \to 0$ and $k_0^{\alpha_x+d/2}/(n^{\alpha_x} \log k_0) \to \infty$, we have*

$$\frac{\hat{k}_{n,k_0}^*}{k_n^*} = 1 + \mathrm{o}_{\mathbf{P}}(1).$$

In particular, $f_{\hat{k}_{n,k_0}^*}(x)$ has the same asymptotic efficiency as $f_{k_n^*}(x)$.

Proof of Theorem 3.2 The result will follow if we show that, for $k_0^{\alpha_x+d/2}/(n^{\alpha_x} \log k_0) \to \infty$, we have

$$\frac{\hat{b}_{k_0}}{b_{k_0}} \to 1 \quad \text{in probability.} \tag{3.4}$$

We consider $x = 0$. To derive (3.4) note that

$$\hat{b}_{k_0} = \frac{(1+\hat{\alpha}_{n,0}/d)^2\,(1+2\hat{\alpha}_{n,0}/d)}{(\hat{\alpha}_{n,0}/d)^2} \bigg[\frac{\hat{\alpha}_{n,0}/d}{1+\hat{\alpha}_{n,0}/d}\, f_{k_0}(0) + \frac{1}{k_0} \sum_{j=1}^{k_0-1} j \bigg(\Big(\frac{j}{k_0}\Big)^{\hat{\alpha}_{n,0}/d} - \Big(\frac{j+1}{k_0}\Big)^{\hat{\alpha}_{n,0}/d} \bigg) f_j(0) \bigg],$$

which is asymptotically equivalent to

$$\frac{(1+\hat{\alpha}_{n,0}/d)^2\,(1+2\hat{\alpha}_{n,0}/d)}{\hat{\alpha}_{n,0}/d}\left[\frac{f_{k_0}(0)}{1+\hat{\alpha}_{n,0}/d}-\frac{1}{k_0}\sum_{j=1}^{k_0-1}\left(\frac{j}{k_0}\right)^{\hat{\alpha}_{n,0}/d} f_j(0)\right].$$

As in the proof of Theorem 3.1, we find that

$$\begin{aligned}
\frac{\hat{b}_{k_0}}{b_{k_0}} = {} & \frac{(1+\hat{\alpha}_{n,0}/d)^2\,(1+2\hat{\alpha}_{n,0}/d)}{(1+\alpha_0/d)\,\hat{\alpha}_{n,0}/d} \\
& \times\left[\frac{(n/k_0)^{\alpha_0/d}\Big(\mu\big(\overline{B}_{k_0}(0)\big)\Big)^{\alpha_0/d}}{1+\hat{\alpha}_{n,0}/d}\right. \\
& \left. -\frac{1}{k_0}\sum_{j=1}^{k_0-1}\left(\frac{j}{k_0}\right)^{\hat{\alpha}_{n,0}/d}\Big(\mu\big(\overline{B}_j(0)\big)\Big)^{\alpha_0/d-1}\left(\frac{n}{k_0}\right)^{\alpha_0/d}\frac{j}{n}\right] \\
& + o_{\mathbf{P}}(1) \\
= {} & \frac{(1+\hat{\alpha}_{n,0}/d)^2\,(1+2\hat{\alpha}_{n,0}/d)}{(1+\alpha_0/d)\,\hat{\alpha}_{n,0}/d}\left(\frac{n}{k_0}\mu\big(\overline{B}_{k_0}(0)\big)\right)^{\alpha_0/d} \\
& \times\left[\frac{1}{1+\hat{\alpha}_{n,0}/d}-\frac{k_0}{n\mu\big(\overline{B}_{k_0}(0)\big)}\frac{1}{k_0}\sum_{j=1}^{k_0-1}\left(\frac{j}{k_0}\right)^{\hat{\alpha}_{n,0}/d+1}\left(\frac{\mu\big(\overline{B}_j(0)\big)}{\mu\big(\overline{B}_{k_0}(0)\big)}\right)^{\alpha_0/d-1}\right] \\
& + o_{\mathbf{P}}(1).
\end{aligned}$$

Denote by $\mathcal{B}(a,b)$ the beta distribution with parameters a and b. Using the consistency of $\hat{\alpha}_{n,0}$, $n\mu\big(\overline{B}_{k_0}(0)\big)/k_0 = 1 + o_{\mathbf{P}}(1)$, and the fact that

$$\frac{1}{k_0}\sum_{j=1}^{k_0-1}\left(\frac{j}{k_0}\right)^{\hat{\alpha}_{n,0}/d+1}\left(\frac{\mu\big(\overline{B}_j(0)\big)}{\mu\big(\overline{B}_{k_0}(0)\big)}\right)^{\alpha_0/d-1} = \frac{1}{1+2\alpha_0/d} + o_{\mathbf{P}}(1)$$

since

$$\frac{\mu\big(\overline{B}_j(0)\big)}{\mu\big(\overline{B}_{k_0}(0)\big)} \sim \mathcal{B}(j, k_0-j+1), \quad j = 1,\ldots,k_0-1,$$

(Wilks [34], page 239), the first result now follows.

The second assertion in the statement of Theorem 3.2 follows from Hall and Welsh [23], Theorem 4.1. ■

Let us now discuss the least squares estimator of $f(x)$ based on (3.1) after substitution of α_x by the consistent estimator $\hat{\alpha}_{n,x}$ based on an appropriate number of neighbors as indicated in Theorem 2.1. This estimator takes the form

$$f_{k_n}^{(a)}(x) = f_{k_n}(x) - \hat{b}_{k_n} \frac{d}{d + \hat{\alpha}_{n,x}} .$$

Theorem 3.3 *Let $x \in \mathbb{R}^d$ be a Lebesgue point of μ satisfying condition (1.3). Assume that $f(x) > 0$. Then as $k_n \to \infty$, $k_n/n \to 0$ and ${k_n}^{\alpha_x + d/2}/n^{\alpha_x} = \mathrm{O}(1)$, we have*

$$\sqrt{k_n}\big(f_{k_n}^{(a)}(x) - f(x)\big) \xrightarrow{\mathcal{D}} \mathcal{N}\Big(0, f^2(x)\Big(\frac{d + \alpha_x}{\alpha_x}\Big)^2\Big).$$

Proof of Theorem 3.3 Again let $x = 0$. First one shows using Theorem 2.1 and the mean value theorem that

$$f_{k_n}^{(a)}(0) - \Big(f_{k_n}(0) - \tilde{b}_{k_n} \frac{d}{d + \alpha_0}\Big) = \mathrm{o}_{\mathbf{P}}(1) ,$$

where

$$\tilde{b}_{k_n} = \frac{(1 + \alpha_0/d)^2 \, (1 + 2\alpha_0/d)}{(\alpha_0/d)^2} \times \frac{1}{k_n} \sum_{j=1}^{k_n} Z_j \bigg(\Big(\frac{j}{k_n}\Big)^{\alpha_0/d} - \frac{1}{1 + \alpha_0/d}\bigg).$$

Then

$$\begin{aligned}
& f_{k_n}(0) - \tilde{b}_{k_n} \frac{d}{d + \alpha_0} \\
& = -f_{k_n}(0)\Big(\frac{d + \alpha_0}{\alpha_0}\Big) \\
& \quad + \frac{(d + \alpha_0)(d + 2\alpha_0)}{d\,\alpha_0} \times \frac{1}{k_n} \sum_{j=1}^{k_n - 1} \Big(\frac{j}{k_n + 1}\Big)^{\alpha_0/d} \frac{j}{n} \big(V_d \, \|X_{(j)}(0)\|^d\big)^{-1} \\
& \quad + \mathrm{o}_{\mathbf{P}}(1).
\end{aligned}$$

Using the method of proof of Theorem 2.2 with the help of Proposition 2.3, one shows that

$$\begin{aligned}
& \sqrt{k_n}\Big[\Big(f_{k_n}(0) - \tilde{b}_{k_n} \frac{d}{d + \alpha_0}\Big) - f(0)\Big] \\
& = \Big(\frac{d}{d + \alpha_0}\Big) f(0) \Big[W_0^{(n)}(1) - \Big(\frac{d + 2\alpha_0}{d}\Big) \frac{1}{k_n} \sum_{j=1}^{k_n - 1} \Big(\frac{j}{k_n + 1}\Big)^{\alpha_0/d - 1} W_0^{(n)}\Big(\frac{j}{k_n}\Big)\Big] \\
& \quad + \mathrm{o}_{\mathbf{P}}\bigg(\Big(\frac{k_n}{n}\Big)^{\alpha_0/d}\bigg).
\end{aligned}$$

The limit distribution is now obtained from this linear combination of $W_0^{(n)}(j/k_n)$ $(j = 1, \ldots, k_n)$ taking $k_n \to \infty$. ■

4 Simulations

To illustrate the results, we present now some simulations. For the probability density f_1 defined in the first section, we estimated the regularity index α_0 at the point 0 using the estimator $\hat{\alpha}_{n,0}$.

First, for $n = 1000$, in Figure 1 we plot single typical trajectories of $f_{k_n}(0)$, $f_{k_n}^{(a)}(0)$, $\hat{\alpha}_{n,0}$ (with $\tau = \sqrt{2}$) and $\log \hat{k}^*_{n,k_0}$ on the same k_n-scale. Note the stability of the plots of $\hat{\alpha}_{n,0}$ and $\log \hat{k}^*_{n,k_0}$ in the region k_n, k_0 beyond 150, while the density estimates $f_{k_n}(0)$ and $f_{k_n}^{(a)}(0)$ coincide closely up to $k_n = 150$ beyond which the stability of $f_{k_n}^{(a)}(0)$ is striking. Here we took $\hat{\alpha}_{n,0}$ as the median of α-estimates obtained for k_n between 150 and 500.

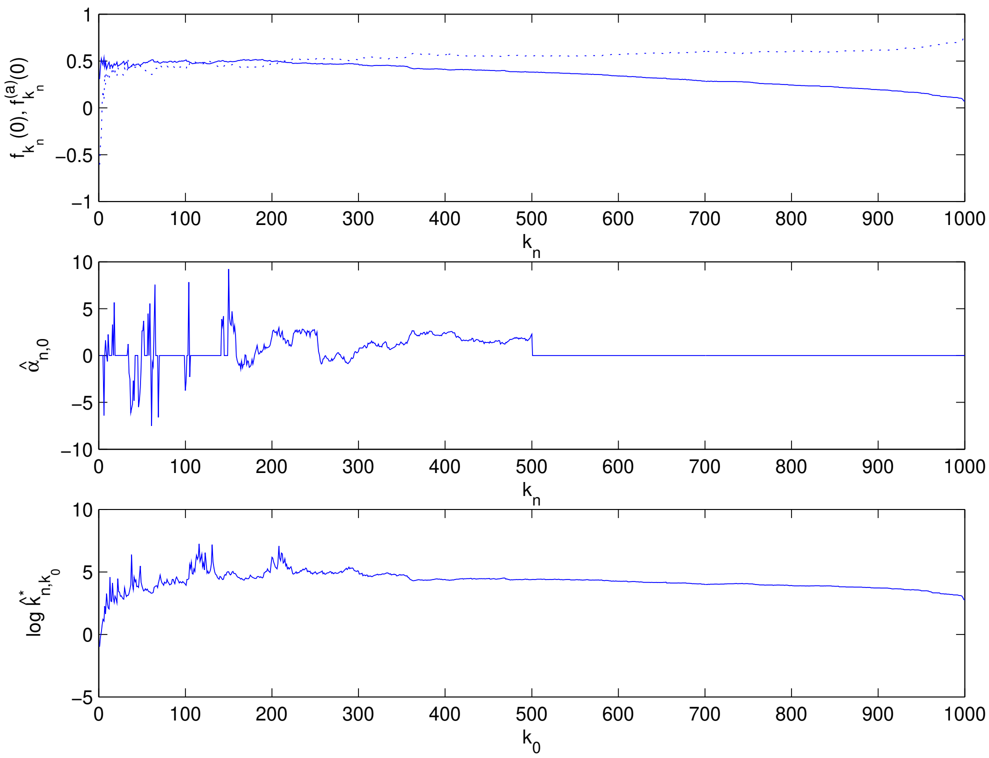

Figure 1: Results for $\mathbf{f_{k_n}(0)}$ (top, solid line), $\mathbf{f^{(a)}_{k_n}(0)}$ (top, dotted line), $\hat{\alpha}_{\mathbf{n,0}}$ (middle, with $\tau = \sqrt{\mathbf{2}}$), and $\log \hat{\mathbf{k}}^*_{\mathbf{n,k_0}}$ (bottom) for a sample of size $\mathbf{1000}$ from $\mathbf{f_1}$.

For fixed n, the value of k_n in the estimator (2.4) should depend on n so that $k_n \to \infty$, $k_n/n \to 0$ and $k_n^{\alpha_x + d/2}/n^{\alpha_x} \to \infty$ as $n \to \infty$. In Figure 2 we consider, for each k_n separately, the sample distribution of 100 independent results of the estimator $\hat{\alpha}_{n,0}$ using $\tau = 2$. These different sample distributions are summarized using the mean, median, first and third quartiles of the 100 outcomes.

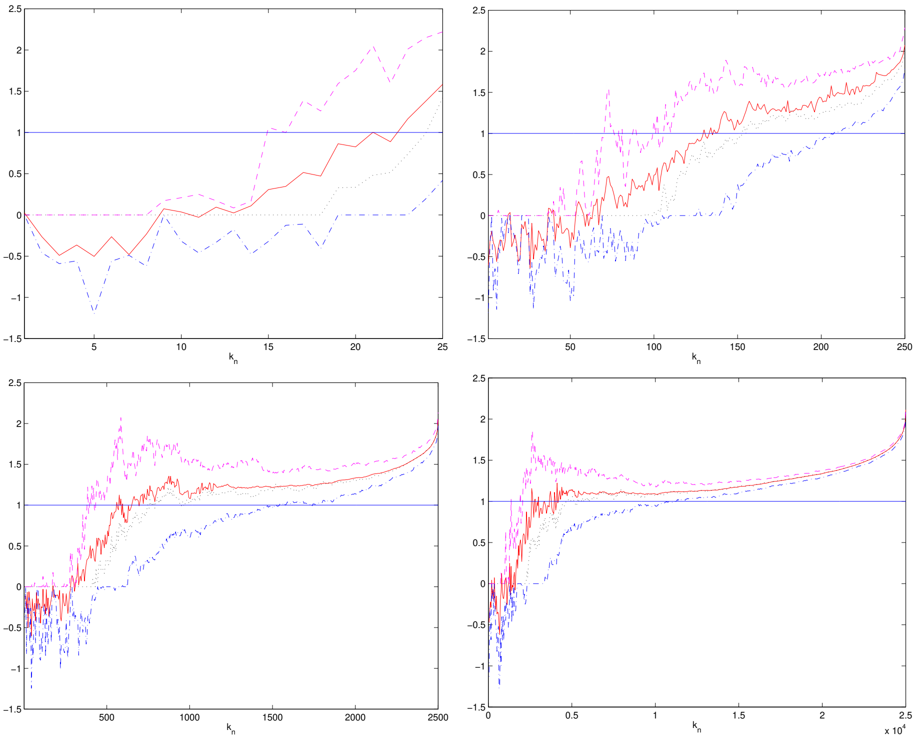

Figure 2: First quartile (dashdotted lines), median (dotted lines), third quartile (dashed lines) and mean (solid lines) of the estimations of α_0 obtained from $\mathbf{100}$ repetitions for each $\mathbf{k_n}$ ranging from $\mathbf{1}$ to $\lfloor \mathbf{n/4} \rfloor$ $\big(\mathbf{n = 100}$ (top left), $\mathbf{n = 1000}$ (top right), $\mathbf{n = 10000}$ (bottom left) and $\mathbf{n = 10^5}$ (bottom right)$\big)$ for the density $\mathbf{f_1}$. We also show the true value of α_0 (horizontal solid line).

The obtained results essentially enlighten the consistency of the estimator under study, as well as the importance of a good choice of k_n with respect to n. In this example consistency is obtained when $k_n/n^{2/3} \to \infty$. The results in case of f_2 turned out to be of similar nature.

A drawback of the selection criterion based on $\hat{k}^*_{n,k_0}$ is that it involves the use of a primary guess k_0 that has to satisfy the conditions outlined in Theorem 3.2 for the selection criterion to be asymptotically efficient. However, as can be derived from (3.3), $\hat{k}^*_{n,k_0}/k^*_n$ approximately behaves as a realization from a normal distribution centered at 1 for values of k_0 smaller than the values indicated in Theorem 3.2. Hence also for smaller k_0 values, $\hat{k}^*_{n,k_0}$ is still median unbiased. Thus, in order to set up an automatic method, from a practical point of view we propose to use the median of the first $n/2$ $\hat{k}$ values as an estimate for k^*_n:

$$\tilde{k} = \mathrm{med}\{\hat{k}^*_{n,k_0}; k_0 = 5, \ldots, n/2\}.$$

Figure 3 summarizes the results of the nearest neighbor estimates of $f_1(0)$ based on $\tilde{k}$ neighbors through a boxplot based on 100 independent repetitions of this adaptive technique, for samples of size 500, 1000 and 5000. Of course the typical under-estimation of the nearest neighbor density estimator of $f_1(0)$ remains present.

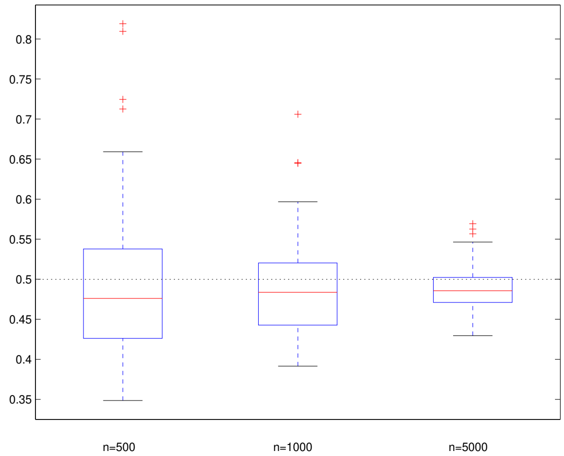

Figure 3: Boxplots based on $\mathbf{100}$ estimates $\mathbf{f}_{\tilde{\mathbf{k}}}(\mathbf{0})$ when estimating $\mathbf{f_1}(\mathbf{0}) = \mathbf{0.5}$ from samples of size $\mathbf{n} = \mathbf{500}$, respectively $\mathbf{n} = \mathbf{1000}$ and $\mathbf{n} = \mathbf{5000}$. We also show the true value of $\mathbf{f_1}(\mathbf{0})$ (horizontal dashed line).

5 Some proofs

Proof of Proposition 2.2 Proof of $(\mathbf{A})$ can be found in Devroye, Györfi and Lugosi [12], Chapter 5. Proof of $(\mathbf{B})(i)$ and $(\mathbf{B})(ii)$ is due to Loftsgaarden and Quesenberry [27] and Moore and Yackel [29]. With respect to $(\mathbf{B})(iii)$, we know from $(\mathbf{B})(i)$ that

$$\frac{\lfloor \tau_1 k_n \rfloor}{nV_d \left\| X_{(\lfloor \tau_1 k_n \rfloor)}(x) - x \right\|^d} \to f(x) \quad \text{in probability,}$$

and, similarly, that

$$\frac{\lfloor \tau_2\, k_n \rfloor}{nV_d\, \|X_{(\lfloor \tau_2 k_n \rfloor)}(x) - x\|^d} \to f(x) \quad \text{in probability.}$$

Since $f(x) > 0$, it follows that

$$\frac{\|X_{(\lfloor \tau_1 k_n \rfloor)}(x) - x\|}{\|X_{(\lfloor \tau_2 k_n \rfloor)}(x) - x\|} \to \Big(\frac{\tau_1}{\tau_2}\Big)^{1/d} \quad \text{in probability.}$$

To prove $(\mathbf{B})(iv)$, observe that, according to $(\mathbf{B})(ii)$,

$$\frac{k_n}{n} - \mu\big(\overline{B}_{k_n}(x)\big) = \mathrm{O}_{\mathbf{P}}\Big(\frac{\sqrt{k_n}}{n}\Big) \quad \text{as } n \to \infty\,.$$

Further,

$$\frac{n}{\sqrt{k_n}} V_d\, \|X_{(k_n)}(x) - x\|^{d+\alpha_x} = \frac{k_n^{1/2+\alpha_x/d}}{(nV_d)^{\alpha_x/d}} \bigg(\frac{nV_d\, \|X_{(k_n)}(x) - x\|^d}{k_n}\bigg)^{1+\alpha_x/d},$$

which tends to ∞ in probability according to the conditions on k_n and $(\mathbf{B})(i)$.
■

Proof of Theorem 3.1 Again, without loss of generality, we assume that $x = 0$. By definition

$$f_{k_n}(0) = \frac{k_n}{n\mu\big(B_{k_n}(0)\big)}\, \frac{\mu\big(B_{k_n}(0)\big)}{V_d\, \|X_{(k_n)}(0)\|^d}\,. \tag{5.1}$$

Using (5.1) and Proposition 2.3 (ii), we obtain

$$f_{k_n}(0) = \frac{k_n}{n\mu\big(B_{k_n}(0)\big)} f(0) + \frac{k_n}{n} \frac{C_0}{V_d^{\alpha_0/d}} f^{-\alpha_0/d}(0) \bigg(\frac{1}{\mu\big(B_{k_n}(0)\big)}\bigg)^{1-\alpha_0/d} \big(1+\mathrm{o}(1)\big) \quad \mathbf{P}\text{-a.s.}$$

as $n \to \infty$. But $\mu\big(B_{k_n}(0)\big)$ is known to have a beta distribution $\mathcal{B}(k_n, n - k_n + 1)$ (Wilks [34], page 239). Therefore,

$$\mathbf{E}\Big(\frac{1}{\mu\big(B_{k_n}(0)\big)}\Big) = \frac{n}{k_n - 1}$$

and

$$\mathbf{E}\Big(\frac{1}{\mu\big(B_{k_n}(0)\big)}\Big)^{1-\alpha_0/d} = \Big(\frac{n}{k_n - 1}\Big)^{1-\alpha_0/d} \big(1 + \mathrm{o}(1)\big)$$

(see Fukunaga and Hostetler [20] for details). Clearly, the o(1) function is bounded on compact sets. Moreover, by assumption, supp μ is compact. Consequently, the Lebesgue's dominated convergence theorem entails

$$\mathbf{E}f_{k_n}(0) = f(0)\Big(1+\frac{1}{k_n-1}\Big) + \frac{C_0}{V_d^{\alpha_0/d}} f^{-\alpha_0/d}(0)\Big(\frac{k_n}{n}\Big)^{\alpha_0/d}\big(1+\mathrm{o}(1)\big),$$

and thus

$$\mathbf{E}f_{k_n}(0) - f(0) = \frac{f(0)}{k_n} + \frac{C_0}{V_d^{\alpha_0/d}} f^{-\alpha_0/d}(0)\Big(\frac{k_n}{n}\Big)^{\alpha_0/d} + \mathrm{o}\bigg(\frac{1}{k_n} + \Big(\frac{k_n}{n}\Big)^{\alpha_0/d}\bigg).$$

With respect to the variance term, we use (5.1) and Proposition 2.3 (*ii*) to obtain

$$f_{k_n}(0) = \frac{k_n}{n\mu\big(B_{k_n}(0)\big)} f(0)\big(1+\mathrm{o}(1)\big) \quad \mathbf{P}\text{-a.s.}$$

as $n \to \infty$. According to the distribution of $\mu\big(B_{k_n}(0)\big)$,

$$\mathbf{E}f_{k_n}(0) = \frac{k_n}{n}\frac{n}{k_n-1} f(0)\big(1+\mathrm{o}(1)\big)$$

and

$$\mathbf{E}f^2_{k_n}(0) = \frac{k_n^2}{n^2}\frac{n(n-1)}{(k_n-1)(k_n-2)} f^2(0)\big(1+\mathrm{o}(1)\big).$$

As a consequence,

$$\mathbf{Var}f_{k_n}(0) = \frac{f^2(0)}{k_n}\big(1+\mathrm{o}(1)\big).$$

The end of the proof of Theorem 3.1 as well as the proof of Corollary 3.1 are straightforward. ■

Acknowledgments. The authors would like to thank David Mason for helping out in Proposition 2.3. They are indebted to a referee for a very careful reading of the paper and stimulating questions and remarks.

References

[1] Abraham, C., Biau, G. and Cadre, B. (2003). Simple estimation of the mode of a multivariate density, *The Canadian Journal of Statistics*, **Vol. 31**, pp. 23–34.

[2] Abraham, C., Biau, G. and Cadre, B. (2004). On the asymptotic properties of a simple estimate of the mode, *ESAIM: Probability and Statistics*, **Vol. 8**, pp. 1–11.

[3] Beirlant, J., Dierckx, G., Goegebeur, Y. and Matthys, G. (1999). Tail index estimation and an exponential regression model, *Extremes*, **Vol. 2**, pp. 177–200.

[4] Berlinet, A. and Levallois, S. (2000). Higher order analysis at Lebesgue points, in *G.G. Roussas Festschrift – Asymptotics in Statistics and Probability*, ed. Puri, M.L., pp. 1–16.

[5] Bosq, D. and Lecoutre, J.P. (1987). *Théorie de l'Estimation Fonctionnelle*, Economica, Paris.

[6] Collomb, G., Hassani, S., Sarda, P. and Vieu, P. (1985). Convergence uniforme d'estimateurs de la fonction de hasard pour des observations dépendantes : méthode du noyau et des k-points les plus proches, *Comptes Rendus de l'Académie des Sciences de Paris*, **Vol. 301**, pp. 653–656.

[7] Cutler, C.D. and Dawson, D.A. (1990). Nearest-neighbor analysis of a family of fractal distributions, *The Annals of Probability*, **Vol. 18**, pp. 256–271.

[8] Davies, S. and Hall, P. (1999). Fractal analysis of surface roughness by using spatial data, *Journal of the Royal Statistical Society, Series B*, **Vol. 61**, pp. 3–37.

[9] Dekkers, A.L.M. and de Haan, L. (1989). On the estimation of the extreme-value index and large quantile estimation, *The Annals of Statistics*, **Vol. 17**, pp. 1795–1832.

[10] Devroye, L. (1979). Recursive estimation of the mode of a multivariate density, *The Canadian Journal of Statistics*, **Vol. 7**, pp. 159–167.

[11] Devroye, L. (1997). Universal smoothing factor selection in density estimation, *Test*, **Vol. 6**, pp. 223–320.

[12] Devroye, L., Györfi, L. and Lugosi, G. (1996). *A Probabilistic Theory of Pattern Recognition*, Springer–Verlag, New York.

[13] Drees, H. and Kaufmann, E. (1998). Selecting the optimal sample fraction in univariate extreme value estimation, *Stochastic Processes and their Applications*, **Vol. 75**, pp. 149–172.

[14] Einmahl, J.H.J. and Mason, D.M. (1992). Generalized quantile processes, *The Annals of Statistics*, **Vol. 20**, pp. 1062–1078.

[15] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events*, Springer–Verlag, Berlin.

[16] van Es, B. (1992). Asymptotics for least squares cross-validation bandwidths in nonsmooth cases, *The Annals of Statistics*, **Vol. 20**, pp. 1647–1657.

[17] Ferraty, F. and Vieu, P. (2000). Dimension fractale et estimation de la régression dans des espaces vectoriels semi-normés, *Comptes Rendus de l'Académie des Sciences de Paris*, **Vol. 330**, pp. 139–142.

[18] Feuerverger, A. and Hall, P. (1999). Estimating a tail exponent by modelling departure from a Pareto distribution, *The Annals of Statistics*, **Vol. 27**, pp. 760–781.

[19] Fix, E. and Hodges, J.L.Jr. (1951). Discriminatory analysis, nonparametric discrimination: consistency properties, *Report number 4, USAF School of Aviation Medicine, Randolph Field, Texas.*

[20] Fukunaga, K. and Hostetler, L.D. (1973). Optimization of k-nearest neighbor density estimates, *IEEE Transactions on Information Theory*, **Vol. 19**, pp. 320-326.

[21] Gomes, M.I., de Haan, L. and Peng, L. (2002). Semi-parametric estimation of the second order parameter in statistics of extremes, *Extremes*, **Vol. 5**, pp. 387–414.

[22] Hall, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems, *Journal of Multivariate Analysis*, **Vol. 32**, pp. 177–203.

[23] Hall, P. and Welsh, A. (1985). Adaptive estimates of parameters of regular variation, *The Annals of Statistics*, **Vol. 13**, pp. 331–341.

[24] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution, *The Annals of Statistics*, **Vol. 3**, pp. 1163–1174.

[25] Lepski, O.V., Mammen, E. and Spokoiny, V.G. (1997). Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors, *The Annals of Statistics*, **Vol. 25**, pp. 929–947.

[26] Levallois, S. (1998). *Convergence et Normalité Asymptotique d'Estimateurs de la Densité* (*Cas* i.i.d. *et Ergodique*), PhD Thesis, Université Montpellier II, France.

[27] Loftsgaarden, D.O. and Quesenberry, C.P. (1965). A nonparametric estimate of a multivariate density function, *The Annals of Mathematical Statistics*, **Vol. 36**, pp. 1049–1051.

[28] Mason, D.M. (1988). A strong invariance theorem for the tail empirical process, *Annales de l'Institut Henri Poincaré (B)*, **Vol. 24**, pp. 491–506.

[29] Moore, D.S. and Yackel, J.W. (1977). Large sample properties of nearest neighbour density function estimates, in *Statistical Decision Theory and Related Topics II*, ed. Gupta, S.S. and Moore, D.S., Academic Press, New York.

[30] Picard, D. and Tribouley, K. (2000). Adaptive confidence interval for pointwise curve estimation, *The Annals of Statistics*, **Vol. 28**, pp. 298–335.

[31] Pickands, J. III (1975). Statistical inference using extreme order statistics, *The Annals of Statistics*, **Vol. 3**, pp. 119–131.

[32] Rudin, W. (1987). *Real and Complex Analysis – 3rd Edition*, McGraw–Hill, New York.

[33] Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics*, Wiley, New York.

[34] Wilks, S.S. (1962). *Mathematical Statistics*, Wiley, New York.