

Article

Estimating Smoothness and Optimal Bandwidth for Probability Density Functions

Dimitris N. Politis ^{1,*}, Peter F. Tarassenko ² and Vyacheslav A. Vasiliev ²
¹ Department of Mathematics and Halicioglu Data Science Institute, University of California, San Diego, CA 92093-0112, USA

² Institute of Applied Mathematics and Computer Science, Tomsk State University, 36 Lenin Ave., 634050 Tomsk, Russia

* Correspondence: dpolitis@ucsd.edu

Abstract: The properties of non-parametric kernel estimators for probability density function from two special classes are investigated. Each class is parametrized with distribution smoothness parameter. One of the classes was introduced by Rosenblatt, another one is introduced in this paper. For the case of the known smoothness parameter, the rates of mean square convergence of optimal (on the bandwidth) density estimators are found. For the case of unknown smoothness parameter, the estimation procedure of the parameter is developed and almost surely convergence is proved. The convergence rates in the almost sure sense of these estimators are obtained. Adaptive estimators of densities from the given class on the basis of the constructed smoothness parameter estimators are presented. It is shown in examples how parameters of the adaptive density estimation procedures can be chosen. Non-asymptotic and asymptotic properties of these estimators are investigated. Specifically, the upper bounds for the mean square error of the adaptive density estimators for a fixed sample size are found and their strong consistency is proved. The convergence of these estimators in the almost sure sense is established. Simulation results illustrate the realization of the asymptotic behavior when the sample size grows large.

Keywords: non-parametric kernel density estimators; adaptive density estimators; mean square and almost surely convergence; rate of convergence; smoothness class

PACS: 62G07; 62G20; 62F12; 62G99



Citation: Politis, D.N.; Tarassenko, P.F.; Vasiliev, V.A. Estimating Smoothness and Optimal Bandwidth for Probability Density Functions. *Stats* **2023**, *6*, 30–49. <https://doi.org/10.3390/stats6010003>

Academic Editor: Wei Zhu

Received: 22 November 2022

Revised: 21 December 2022

Accepted: 21 December 2022

Published: 27 December 2022



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Let X_1, \dots, X_n be independent identically distributed random variables (i.i.d. r.v.'s) having a probability density function f . In the typical non-parametric set-up, nothing is assumed about f except that it possesses a certain degree of smoothness, e.g., that it has r continuous derivatives.

Estimating f via kernel smoothing is a sixty year old problem; M. Rosenblatt who was one of its originators discusses the subject's history and evolution in the monograph [1]. For some point x , the kernel smoothed estimator of $f(x)$ is defined by

$$f_{n,h}(x) = \frac{1}{n} \sum_{j=1}^n \frac{1}{h} K\left(\frac{x - X_j}{h}\right) \quad (1)$$

where the kernel $K(\cdot)$ is a bounded function satisfying $\int K(x)dx = 1$ and $\int K^2(x)dx < \infty$, and the positive bandwidth parameter h is a decreasing function of the sample size n .

If $K(\cdot)$ has finite moments up to q -th order, and moments of order up to $q - 1$ equal to zero, then q is called the 'order' of the kernel $K(\cdot)$. Since the unknown function f is assumed to have r continuous derivatives, it typically follows that

$$\text{Var}(f_{n,h}(x)) = \frac{C_{f,K}(x)}{hn} + o\left(\frac{1}{hn}\right),$$

and

$$\text{Bias}(f_{n,h}(x)) = c_{f,K}(x)h^k + o(h^k),$$

where $k = \min(q, r)$, and $C_{f,K}(x), c_{f,K}(x)$ are bounded functions depending on $K(\cdot)$ as well as f and its derivatives, cf. [1] p. 8.

The idea of choosing a kernel of order q bigger (or equal) than r in order to ensure the $\text{Bias}(f_{n,h}(x))$ to be $O(h^r)$ dates back to the early 1960s in work of [2,3]; recent references on higher-order kernels include the following: [4–10]. Note that since r is typically unknown and can be arbitrarily large, it is possible to use kernels of infinite order that achieve the minimal bias condition $\text{Bias}(f_{n,h}(x)) = O(h^r)$ for any r ; Ref. [11] gives many properties of kernels of infinite order. In this paper we will employ a particularly useful class of infinite order kernels namely the *flat-top* family; see [12] for a general definition.

It is a well-known fact that optimal bandwidth selection is perhaps the most crucial issue in such non-parametric smoothing problems; see [13], as well as the book [14]. The goal typically is minimization of the large-sample mean squared error (MSE) of $f_{n,h}(x)$. However, to perform this minimization, the practitioner needs to know the degree of smoothness r , as well as the constants $C_{f,K}(x)$ and $c_{f,K}(x)$. Using an infinite order kernel and focusing just on optimizing the order of magnitude of the large-sample MSE, it is apparent that the optimal bandwidth h must be asymptotically of order $n^{-1/(2r+1)}$; this yields a large-sample MSE of order $n^{-2r/(2r+1)}$.

A generalization of the above scenario is possible using a degree of smoothness r that has another sense, and that is not necessarily an integer. Let $[r]$ denote the integer part of r , and define $\gamma = r - [r]$; then, one may assume that f has $[r]$ continuous derivatives, and that the $[r]$ th derivative satisfies a Lipschitz condition of order γ . Interestingly, even in this case where f is assumed to belong to the Hölder class of degree r (the derivative of the density function of the order r satisfies the Lipschitz condition) the MSE-optimal bandwidth h is still of order $n^{-1/(2r+1)}$ and again yields a large-sample MSE of the order $n^{-2r/(2r+1)}$ (see, e.g., [15–18] among others).

The problem of course is that, as previously mentioned, the underlying degree of smoothness r is typically unknown. In Section 4 of the paper at hand, we develop an estimator r_n of r and prove its strong consistency; this is perhaps the first such result in the literature. In order to construct our estimator r_n , we operate under a class of functions that is slightly more general than the aforementioned Hölder class; this class of functions is formally defined in Section 2 via Equation (3) or (4).

Under such a condition on the tails of the characteristic function we are able to show in Section 3 that the optimized MSE of $\hat{f}_n(x)$ is again of order $n^{-2r/(2r+1)}$ for possibly non-integer r ; this is true, for example, when the characteristic function $\phi(s)$ has tails of order $O(1/|s|^{r+1})$, see Example 2.

Furthermore, in Section 5 we develop an *adaptive* estimator $\hat{f}_n(x)$ that achieves the optimal MSE rate of $n^{-2r/(2r+1)}$ within a logarithmic factor despite the fact that r is unknown, see Examples after Theorem 3. Similar effect arises in the adaptive estimation problem of the densities from the Hölder class; see [18–20]. It should be pointed out that problems of asymptotic adaptive optimal density estimations from another classes have also been considered in the literature; see, e.g., [14,21–23].

The construction of $\hat{f}_n(x)$ is rather technical; it uses the new estimator r_n , and it is inspired from the construction of sequential estimates although we are in a fixed n , non-sequential setting. As the major theoretical result of our paper, we are able to prove a non-asymptotic upper bound for the MSE of $\hat{f}_n(x)$ that satisfies the above mentioned optimal rate. Section 6 contains some simulation results showing the performance of the new estimator $\hat{f}_n(x)$ in practice. All proofs are deferred to Section 7, while Section 8 contains our conclusions and suggestions for future work.

2. Problem Set-Up and Basic Assumptions

Let X_1, \dots, X_n be i.i.d. having a probability density function f . Denote $\phi(s) = \int e^{isx} f(x) dx$ the characteristic function of f and the sample characteristic function $\phi_n(s) = \frac{1}{n} \sum_{k=1}^n e^{isX_k}$. For some finite $r > 0$, define two families \mathcal{F}_r^+ and \mathcal{F}_r of bounded, i.e.,

$$\exists 0 < \bar{f} < \infty : \sup_{y \in \mathcal{R}^1} f(y) \leq \bar{f}, \quad (2)$$

and continuous functions f satisfying one of the following conditions, respectively:

$$\int |s|^r |\phi(s)| ds < \infty, \quad \int |s|^{r+\varepsilon} |\phi(s)| ds = \infty, \quad \text{for all } \varepsilon > 0, \quad (3)$$

$$\int |s|^{r-\varepsilon} |\phi(s)| ds < \infty, \quad \int |s|^r |\phi(s)| ds = \infty, \quad \text{for all } 0 < \varepsilon < r. \quad (4)$$

In other words, \mathcal{F}_r^+ is the family of functions (introduced by M. Rosenblatt) satisfying (2) and (3), while \mathcal{F}_r is the family of functions (introduced in this paper) satisfying (2) and (4). It should be noted that the new class \mathcal{F}_r is a little bit more wide than the classical class \mathcal{F}_r^+ .

In addition, define the family $\mathcal{F}_{r,m}^+$ (respectively, $\mathcal{F}_{r,m}$) as the family of functions f that belong to \mathcal{F}_r^+ (respectively, \mathcal{F}_r) but with f being such that its characteristic function $|\phi(s)|$ has monotonously decreasing tails.

Consider the class Ξ of non-parametric kernel smoothed estimators $f_{n,h}(x)$ of $f(x)$ as given in Equation (1). Note that we can alternatively express $f_{n,h}(x)$ in terms of the Fourier transform of kernel $K(\cdot)$, i.e.,

$$f_{n,h}(x) = \frac{1}{n} \sum_{j=1}^n \frac{1}{h} K\left(\frac{x - X_j}{h}\right) = \frac{1}{2\pi} \int \lambda(s, h) \phi_n(s) e^{-isx} ds \quad (5)$$

where

$$\lambda(s, h) = \int K\left(\frac{x}{h}\right) e^{isx} dx.$$

In this paper, we will employ the family of flat-top infinite order kernels, i.e., we will let the function $\lambda(s, h)$ be of the form

$$\lambda_c(s, h) = \begin{cases} 1 & \text{if } |s| \leq 1/h, \\ g(s, h) & \text{if } 1/h < |s| \leq c/h, \\ 0 & \text{if } |s| \geq c/h, \end{cases}$$

where c is a fixed number in $[1, \infty)$ chosen by the practitioner, and $g(s, h)$ is some properly chosen continuous, real-valued function satisfying $g(s, h) = g(-s, h)$, $g(s, 1) = g(s/h, h)$, and $|g(s, h)| \leq 1$, for any s , with $g(1/h, h) = 1$, and $g(c/h, h) = 0$; see [12,24–26] for more details on the above flat-top family of kernels.

Define $g'_h(s, h)$ the partial derivative of the function $g(s, h)$ with respect to the bandwidth h . We will also assume that for some $c_0 > 0$

$$\lim_{h \rightarrow 0} \sup_{1/h < |s| < c/h} |g'_h(s, h)| / |s| < c_0. \quad (6)$$

Denote for every $0 \leq \gamma < r$ the functions

$$\delta_\gamma(h) = \int_{1/h < |s| < c/h} |s|^{r-\gamma} |\phi(s)| ds, \quad \text{when } h > 0, \quad \text{and } \delta_\gamma(0) = 0.$$

From (3) and (5) it follows that $\delta_\gamma(h) = o(1)$ as $h \rightarrow 0$ for $f \in \mathcal{F}_r^+$ and $\gamma = 0$, as well as for $f \in \mathcal{F}_r$ and $0 < \gamma < r$. In other cases $\delta_\gamma(h) = \infty$.

Define the following classes $\overline{\mathcal{F}}_r = \mathcal{F}_r^+ \cup \mathcal{F}_r$ and $\overline{\mathcal{F}}_{r,m} = \mathcal{F}_{r,m}^+ \cup \mathcal{F}_{r,m}$.

The main aim of the paper is the estimation of the parameter r of these classes and adaptive estimation of densities from the class $\overline{\mathcal{F}}_r$ with the unknown parameter r .

3. Asymptotic Mean Square Optimal Estimation

The mean square error (MSE) $u_f^2(f_{n,h}) = E_f(f_{n,h}(x) - f(x))^2$ of the estimators $f_{n,h}(x) \in \Xi$, $f \in \overline{\mathcal{F}}_r$ has the following form:

$$u_f^2(f_{n,h}) = U_f^2(h, c) - \frac{1}{n} \left(\int K(v) f(x - hv) dv \right)^2, \quad (7)$$

where $U_f^2(h, c)$ is the principal term of the MSE,

$$U_f^2(h, c) = \frac{L_1 f(x)}{nh} + \left[\frac{1}{2\pi} \int_{1/h < |s| < c/h} (1 - g(s, h)) \phi(s) e^{-isx} ds \right]^2,$$

$L_1 = \int K^2(v) dv$. Thus, in particular, $\sup_{f \in \overline{\mathcal{F}}_r} \left| \int K(v) f(x - hv) dv \right| < \infty$.

To minimize the principal term $U_f^2(h, c)$ by h we set its first derivative with respect to h to zero which gives the following equality for the optimal (in the mean square sense) value $h^0 = h^0(n)$:

$$\begin{aligned} & \int_{1/h^0 < |s| < c/h^0} (g(s, h^0) - 1) \phi(s) e^{-isx} ds \cdot \{c\phi(c/h^0) e^{-\frac{icx}{h^0}} + c\phi(-c/h^0) e^{\frac{icx}{h^0}}\} \\ & + (h^0)^2 \int_{1/h^0 < |s| < c/h^0} g'_h(s, h^0) \phi(s) e^{-isx} ds = \frac{2\pi^2 L_1 f(x)}{n}. \end{aligned} \quad (8)$$

From the definition of the class of kernels for cases $\delta_\gamma(h) < \infty$ we have

$$\left| \int_{1/h < |s| < c/h} (g(s, h) - 1) \phi(s) e^{-isx} ds \right| \leq 2h^{r-\gamma} \delta_\gamma(h)$$

and for h small enough, according to (6)

$$\left| \int_{1/h < |s| < c/h} g'_h(s, h) \phi(s) e^{-isx} ds \right| \leq c_0 h^{r-1-\gamma} \delta_\gamma(h).$$

Then, by the definition of the class $\overline{\mathcal{F}}_{r,m}$, as h small enough, denoting $c_1(\gamma) = \frac{r+1-\gamma}{2(c^{r+1-\gamma}-1)}$, we have

$$\begin{aligned} \delta_\gamma(h) & \geq \int_{1/h}^{c/h} |s|^{r-\gamma} ds \cdot \left[\inf_{1/h < s < c/h} |\phi(s)| + \inf_{1/h < s < c/h} |\phi(-s)| \right] \\ & \geq (c_1(\gamma))^{-1} h^{-(r+1-\gamma)} [|\phi(c/h)| + |\phi(-c/h)|]. \end{aligned}$$

Thus, for $h \ll 1$,

$$|\phi(c/h)| + |\phi(-c/h)| \leq c_1(\gamma)h^{r+1-\gamma}\delta_\gamma(h)$$

and from (8) it follows

$$(h^0)^{2r+1-2\gamma}\delta_\gamma^2(h^0) \geq \frac{\pi^2 L_1 f(x)}{(c_0 + c_1(\gamma))n}.$$

Define the number $h_1^0 = h_1^0(n)$ from the equality

$$(h_1^0)^{2r+1-2\gamma}\delta_\gamma^2(h_1^0) = \frac{\pi^2 L_1 f(x)}{(c_0 + c_1(\gamma))n}. \quad (9)$$

It is obvious, that $0 < h_1^0 \leq h^0$ and $(h_1^0)^{2r+1-2\gamma}\delta_\gamma^2(h_1^0) \leq (h^0)^{2r+1-2\gamma}\delta_\gamma^2(h^0)$.

Then, from (7) and (9), for every $f \in \overline{\mathcal{F}}_{r,m}$ and $f_n^0 = f_{n,h_1^0}$ as $n \rightarrow \infty$, we have

$$\begin{aligned} u_f^2(f_n^0) &\leq u_f^2(f_{n,h_1^0}) \leq \frac{L_1 f(x)}{nh_1^0} + \frac{1}{\pi^2} (h_1^0)^{2r-2\gamma} \delta_\gamma^2(h_1^0) \\ &= C_\gamma \cdot \frac{\delta_\gamma^{\frac{2}{2r+1-2\gamma}}(h_1^0)}{n^{\frac{2r-2\gamma}{2r+1-2\gamma}}}, \end{aligned} \quad (10)$$

where

$$C_\gamma = L_1^{\frac{2r-2\gamma}{2r+1-2\gamma}}(x) \left(1 + \frac{\pi^2}{c_0 + c_1(\gamma)}\right) \left(\frac{c_0 + c_1(\gamma)}{\pi^2}\right)^{\frac{1}{2r+1-2\gamma}}.$$

In such a way we have proved the following theorem, which gives the rates of convergence of the random quantities $f_n^0(x)$ and $f_{n,h_1^0}(x)$. We can loosely call $f_n^0(x)$ and $f_{n,h_1^0}(x)$ ‘estimators’ although it is clear that these functions can not be considered as estimators in the usual sense in view of the dependence of the bandwidths h^0 and h_1^0 on unknown parameters r and $f(x)$. Nevertheless, this theorem can be used for the construction of bona fide adaptive estimators with the optimal and suboptimal converges rates; see Examples 1 and 2, as well as Section 5.3 in what follows.

Theorem 1. Let $f(x) > 0$. Then, for the asymptotically optimal (with respect to bandwidth h) in the MSE sense ‘estimator’ $f_n^0(x)$ of the function $f \in \overline{\mathcal{F}}_r$ and for the ‘estimator’ $f_{n,h_1^0}(x)$ of $f \in \overline{\mathcal{F}}_{r,m}$ the following limit relations, as $n \rightarrow \infty$, hold

$$\begin{aligned} 1^\circ. & \sup_{f \in \overline{\mathcal{F}}_r} \left| \inf_h u_f^2(f_{n,h}) - U_f^2(h^0, c) \right| = O\left(\frac{1}{n}\right); \\ 2^\circ. & \text{for every } f \in \overline{\mathcal{F}}_{r,m} \text{ with } \gamma = 0 \text{ if } f \in \mathcal{F}_{r,m}^+ \text{ and every } 0 < \gamma < r \text{ if } f \in \mathcal{F}_{r,m} \\ & u_f^2(f_n^0) \leq u_f^2(f_{n,h_1^0}) \leq C_\gamma \cdot \frac{\delta_\gamma^{\frac{2}{2r+1-2\gamma}}(h_1^0)}{n^{\frac{2r-2\gamma}{2r+1-2\gamma}}}, \quad n \geq 1. \end{aligned}$$

Remark 1. The definition (9) of h_1^0 is essentially simpler than the definition (8) of the optimal bandwidth h^0 . From Theorem 1 it follows that the (slightly) suboptimal ‘estimator’ f_{n,h_1^0} can be successfully used instead.

It should be noted that the parameter γ is chosen by the practitioner here and that $\gamma = 0$ if $f \in \mathcal{F}_{r,m}^+$ but $0 < \gamma < r$ if $f \in \mathcal{F}_{r,m}$ in which case we want to choose γ close to 0.

We shall write in the sequel $\varphi(s) \approx \psi(s)$ as $s \rightarrow \infty$ instead of the limit relations

$$0 < \lim_{s \rightarrow \infty} \frac{\varphi(s)}{\psi(s)} \leq \lim_{s \rightarrow \infty} \frac{\varphi(s)}{\psi(s)} < \infty.$$

Example 1. Consider an estimation problem of the function $f \in \mathcal{F}_{r,m}^+$, satisfying the following additional condition

$$|\phi(s)| \approx \frac{1}{|s|^{r+1} \ln^{1+\varphi} |s|} \text{ as } |s| \rightarrow \infty, \quad \varphi > 0,$$

using the kernel estimator $(f_{n,h}(x)) \in \Xi$.

By making use of (9) and (10) we find the rate of convergence of the MSE $u_f^2(f_n^0)$ and $u_f^2(f_{n,h_1^0})$. To this end we calculate

$$\delta_0(h) = \int_{1/h < |s| < c/h} |s|^r |\phi(s)| ds \approx \frac{1}{(\ln h^{-1})^\varphi} - \frac{1}{(\ln h^{-1} + \ln c)^\varphi} \approx \frac{1}{(\ln^2 h^{-1})^{1+\varphi}}.$$

It is easy to verify that $f_n^0, f_{n,h_1^0} \in \Xi$. Thus, from (9), as $n \rightarrow \infty$,

$$h_1^0 \approx (n\gamma_n)^{-\frac{1}{2r+1}},$$

where $\gamma_n \approx \ln^{-2(1+\varphi)} n$, $n \rightarrow \infty$ is a solution of the equation

$$\gamma_n [\ln n + \ln \gamma_n]^{2(1+\varphi)} = 1.$$

Therefore, as $n \rightarrow \infty$, we have

$$h_1^0 \approx \left(\frac{\ln^{2(1+\varphi)} n}{n} \right)^{\frac{1}{2r+1}} \text{ and } u_f^2(f_{n,h_1^0}) = O\left(\frac{1}{n^{2r} \ln^{2(1+\varphi)} n} \right)^{\frac{1}{2r+1}}.$$

Consider the piecewise linear flat-top kernel $\lambda_c^{LIN}(s, h)$, introduced by [25] (see [26] as well):

$$\lambda_c^{LIN}(s, h) = \frac{c}{c-1} \left(1 - \frac{h}{c} |s| \right)^+ - \frac{1}{c-1} (1 - h|s|)^+,$$

where $(x)^+ = \max(x, 0)$ is the positive part function.

Then, from (8) we obtain

$$\{\phi(c/h^0) e^{-\frac{icx}{h^0}} + \phi(-c/h^0) e^{\frac{icx}{h^0}}\}^2 \approx \frac{1}{n}$$

and, for n large enough

$$|\phi(c/h^0)|^2 + |\phi(-c/h^0)|^2 \geq \frac{C}{n}.$$

Thus, similarly to h_1^0 , as $n \rightarrow \infty$, for $f \in \mathcal{F}_r$ we find

$$h^0 \approx \left(\frac{\ln^{2(1+\varphi)} n}{n} \right)^{\frac{1}{2(r+1)}}$$

and

$$u_f^2(f_n^0) = O\left(\frac{1}{n^{2r+1} \ln^{2(1+\varphi)} n} \right)^{\frac{1}{2(r+1)}} = o\left(u_f^2(f_{n,h_1^0}) \right).$$

Example 2. Consider an estimation problem of the function $f \in \mathcal{F}_{r,m}$, satisfying the following additional condition:

$$|\phi(s)| \approx \frac{1}{|s|^{r+1}} \text{ as } |s| \rightarrow \infty,$$

using the kernel estimator $(f_{n,h}(x)) \in \Xi$.

Using (9) and (10) we will find the rate of convergence of the MSE $u_f^2(f_n^0)$ and $u_f^2(f_{n,h_1^0})$. To this end, we calculate

$$\delta_\gamma(h) = \int_{1/h < |s| < c/h} |s|^{r-\gamma} |\phi(s)| ds \approx h^\gamma.$$

It is easy to verify that $f_n^0, f_{n,h_1^0} \in \Xi$. Thus, from (9), as $n \rightarrow \infty$,

$$h_1^0 \approx n^{-\frac{1}{2r+1}}.$$

Therefore, we have

$$u_f^2(f_{n,h_1^0}) = O\left(\frac{1}{n^{\frac{2r}{2r+1}}}\right), \quad n \rightarrow \infty.$$

Similarly to Example 1 as $n \rightarrow \infty$, for $f \in \mathcal{F}_r$ we find

$$h^0 \approx \frac{1}{n^{\frac{1}{2(r+1)}}}$$

and

$$u_f^2(f_n^0) = O\left(\frac{1}{n^{\frac{2r}{2(r+1)}}}\right) = o\left(u_f^2(f_{n,h_1^0})\right).$$

4. Estimation of the Degree of Smoothness r

Define the functions

$$\Phi_\alpha(A, B) = \int_{A < |s| < B} |s|^\alpha |\phi(s)| ds, \quad \Phi_\alpha = \Phi_\alpha(0, \infty),$$

$$\Phi_{n,\alpha}(A, B) = \int_{A < |s| < B} |s|^\alpha |\phi_n(s)| ds, \quad \Phi_{n,\alpha} = \Phi_{n,\alpha}(0, \infty).$$

Let $(\delta_n)_{n \geq 1}$ and $(\rho_n)_{n \geq 1}$ be two given sequences of positive numbers chosen by the practitioner such that $\delta_n \rightarrow 0$ and $\rho_n \rightarrow \infty$ as $n \rightarrow \infty$. The sequence (δ_n) represents the 'grid'-size in our search of the correct exponent r , while (ρ_n) represents an upper bound that limits this search.

Define the following sets of non-random sequences

$$\mathcal{C}_+ = \{(A_n, B_n, \delta_n)_{n \geq 1} : A_n \rightarrow \infty, 0 < A_n < B_n \rightarrow \infty, \delta_n \rightarrow 0 \text{ as } n \rightarrow \infty;$$

$$\text{for some } m_0 \geq 2, \sum_{n \geq 1} \frac{B_n^{2m_0(\rho_n+1+\delta_n)}}{n^{m_0}} < \infty; \Phi_{r+\varepsilon}(A_n, B_n) \rightarrow \infty, \forall \varepsilon > 0\},$$

$$\mathcal{C} = \{(A_n, B_n, \delta_n)_{n \geq 1} : A_n \rightarrow \infty, 0 < A_n < B_n \rightarrow \infty, \delta_n \rightarrow 0 \text{ as } n \rightarrow \infty;$$

$$\text{for some } m_0 \geq 2, \sum_{n \geq 1} \frac{B_n^{2m_0(\rho_n+1+\delta_n)}}{n^{m_0}} < \infty; \Phi_r(A_n, B_n) \rightarrow \infty\}.$$

Remark 2. Formally, the definition of sets \mathcal{C}_+ , \mathcal{C} and, as follows of estimators r_n^+ and r_n , as well of sets \mathcal{C}_+^* , \mathcal{C}_*^* defined below depend on the unknown function $\Phi_\alpha(A, B)$. At the same time, the set \mathcal{C}_+ (and, as follows, the estimator r_n^+ and the set \mathcal{C}_+^*) can be defined independently of $\Phi_\alpha(A, B)$.

Indeed, denote $\alpha_s = |s|^{r+1} |\phi(s)|$ and

– let $f \in \mathcal{F}_r^+$.

Then for every $\varepsilon > 0$

$$\lim_{s \rightarrow \infty} s^{\varepsilon/2} \alpha_s = \infty \quad \text{and} \quad \overline{\lim}_{s \rightarrow \infty} \alpha_s \cdot \log s < \infty.$$

Thus $(A_n, B_n, \delta_n) \in \mathcal{C}_+$ for appropriate chosen (δ_n) and $A_n = O(B_n^{1/2})$ because (consider for simplification the case $A_n > 0$)

$$\begin{aligned}\Phi_{r+\varepsilon}(A_n, B_n) &= \int_{A_n}^{B_n} |s|^{r+\varepsilon} |\phi(s)| ds = \int_{A_n}^{B_n} s^{\varepsilon-1} \alpha_s ds = \int_0^{B_n} s^{\varepsilon-1} \alpha_s ds - \int_0^{A_n} s^{\varepsilon-1} \alpha_s ds \\ &\geq C_1 \int_0^{B_n} s^{\varepsilon/2-1} ds - C_2 \int_0^{A_n} s^{\varepsilon-1} \log^{-1} s ds \sim B_n^{\varepsilon/2} - A_n^{\varepsilon} \log^{-1} A_n \\ &\quad + \int_0^{A_n} s^{\varepsilon-1} \log^{-2} s ds \rightarrow \infty.\end{aligned}$$

According to the definition of the class \mathcal{F}_r it is impossible to find elements of the set \mathcal{C} independently of the function to be estimated without usage of an a priori information about f . Consider one simple example.

– Let $f \in \mathcal{F}_r$.

Suppose, e.g., in addition that

$$\lim_{s \rightarrow \infty} \log s \cdot \frac{1}{s} \int_0^s |u|^{r+1} |\phi(u)| du > 0.$$

Then $(A_n, B_n, \delta_n) \in \mathcal{C}$ for appropriate chosen (δ_n) and $A_n = o(B_n)$ because

$$\begin{aligned}\Phi_r(A_n, B_n) &= \int_{A_n}^{B_n} s^r |\phi(s)| ds = \int_{A_n}^{B_n} s^{-1} d \int_0^s \alpha_u du ds = \frac{1}{s} \int_0^s \alpha_u du \Big|_{A_n}^{B_n} \\ &\quad + \int_{A_n}^{B_n} s^{-2} \int_0^s \alpha_u du ds \geq C \int_{A_n}^{B_n} \frac{1}{s \log s} ds \geq C \log \log B_n \rightarrow \infty.\end{aligned}$$

Another examples are in Example 3 (see also Remark 3 and Example 4).

For an arbitrary given $H > 0$ chosen by the practitioner, define the estimators $(r_n^+)_{n \geq 1}$ and $(r_n)_{n \geq 1}$ of the parameter r in (3) and (4) as follows

$$r_n^+ = \min[\varrho_n, (\delta_n \cdot \inf\{k \geq 1 : \Phi_{n,(k+1)\delta_n}(A_n, B_n) \geq H, (A_n, B_n, \delta_n) \in \mathcal{C}_+\})]. \quad (11)$$

$$r_n = \min[\varrho_n, (\delta_n \cdot \inf\{k \geq 1 : \Phi_{n,k\delta_n}(A_n, B_n) \geq H, (A_n, B_n, \delta_n) \in \mathcal{C}\})]. \quad (12)$$

Example 3. For the functions $\phi(\cdot)$ from Examples 1 and 2, we can use the definitions (11) and (12) with the following choices:

$$B_n = \ln n, \quad \varrho_n = \rho \frac{\ln n}{\ln \ln n}, \quad \rho \in (0, (2m_0)^{-1}),$$

arbitrary $\delta_n \rightarrow 0$ and $A_n = o(B_n)$, as $n \rightarrow \infty$. Indeed, for $f \in \mathcal{F}_{r,m}^+$ and every $\varepsilon > 0$ (Example 1),

$$\Phi_{r+\varepsilon}(A_n, B_n) = \int_{A_n < |s| < B_n} |s|^{r+\varepsilon} |\phi(s)| ds \approx \frac{B_n^\varepsilon}{\ln^\varphi B_n} - \frac{A_n^\varepsilon}{\ln^\varphi A_n} \approx \frac{\ln^\varepsilon n}{\ln^\varphi \ln n} \rightarrow \infty$$

and for $f \in \mathcal{F}_{r,m}$ (Example 2),

$$\Phi_r(A_n, B_n) = \int_{A_n < |s| < B_n} |s|^r |\phi(s)| ds \approx \ln \frac{B_n}{A_n} \rightarrow \infty$$

and, as follows, the classes \mathcal{C}_+ and \mathcal{C} are not empty.

Define

$$J_{n,\alpha} = \int_{A_n < |s| < B_n} |s|^\alpha |\phi_n(s) - \phi(s)| ds, \quad \alpha > 0, \quad n \geq 1.$$

Lemma 1. Let $(A_n, B_n, \delta_n) \in \mathcal{C}_+ \cup \mathcal{C}$. Then, for every $\alpha > 0$, $m \geq 1$ and $n \geq 1$ there exist positive numbers $C_{\alpha,m}$ such that

$$\sup_{f \in \overline{\mathcal{F}}_r} E_f J_{n,\alpha}^{2m} \leq C_{\alpha,m} \frac{B_n^{2m(\alpha+1)}}{n^m} \quad (13)$$

and for every $f \in \overline{\mathcal{F}}_r$

$$J_{n,\varrho_n+\delta_n} = o(1) \quad P_f - a.s.$$

Define the sets \mathcal{C}_+^* and \mathcal{C}^* of non-random sequences $(A_n, B_n, \delta_n)_{n \geq 1}$

$$\mathcal{C}_+^* = \{(A_n, B_n, \delta_n) \in \mathcal{C}_+ : \lim_{n \rightarrow \infty} \Phi_{r+\delta_n}(A_n, B_n) = \infty\},$$

$$\mathcal{C}^* = \{(A_n, B_n, \delta_n) \in \mathcal{C} : \lim_{n \rightarrow \infty} \Phi_{r-\delta_n}(A_n, B_n) = 0\}.$$

Remark 3. It can be directly verified that under the conditions of Remark 2 the sequences $(A_n, B_n, \delta_n) \in \mathcal{C}_+^*$ if $A_n = O(B_n^{1/2})$ and $\delta_n^{-1} = o(\log B_n)$, as well as $(A_n, B_n, \delta_n) \in \mathcal{C}^*$ if $A_n = o(B_n)$ and

$$\delta_n = o\left(\frac{\log \log \log B_n}{\log B_n}\right).$$

Moreover, under the conditions of Example 3.1, $(A_n, B_n, \delta_n) \in \mathcal{C}_+^*$ if we put

$$\delta_n = \delta \cdot \frac{\ln \ln \ln(n+1)}{\ln \ln(n+1)}, \quad \delta > \varphi.$$

Example 4. Consider the functions $\phi(\cdot)$ from Examples 1, 2 and suppose, that the smooth parameter $r \leq R$ for some known number R . Then the sequences $(A_n, B_n, \delta_n) \in \mathcal{C}_+^* \cup \mathcal{C}^*$ if we put

$$B_n = n^b, \quad 0 < b < \frac{m_0 - 1}{4(R+1)}, \quad A_n = o(B_n), \quad \varrho_n = R,$$

$$\delta_n = \delta \cdot \frac{\ln \ln(n+1)}{\ln(n+1)}, \quad \delta b > \varphi.$$

Theorem 2. The estimators r_n^+ and r_n , defined in (11) and (12), respectively, with $\rho_n \rightarrow \infty$ have the following properties

1°

(a) if $f \in \mathcal{F}_r^+$ and $(A_n, B_n, \delta_n) \in \mathcal{C}_+$, then

$$\lim_{n \rightarrow \infty} r_n^+ = r \quad P_f - a.s.$$

(b) if $f \in \mathcal{F}_r$ and $(A_n, B_n, \delta_n) \in \mathcal{C}$, then

$$\lim_{n \rightarrow \infty} r_n = r \quad P_f - a.s.$$

2°

(a) if $f \in \mathcal{F}_r^+$ and for some $\delta_n \rightarrow 0$ the sequences $(A_n, B_n, \delta_n) \in \mathcal{C}_+^*$, then

$$\lim_{n \rightarrow \infty} \delta_n^{-1}(r_n^+ - r) = 0 \quad P_f - a.s.$$

(b) if $f \in \mathcal{F}_r$ and for some $\delta_n \rightarrow 0$ the sequences $(A_n, B_n, \delta_n) \in \mathcal{C}_*$, then

$$\lim_{n \rightarrow \infty} \delta_n^{-1}(r_n - r) = 0 \quad P_f - a.s.$$

5. Adaptive Estimation of the Functions $f \in \overline{\mathcal{F}}_r$

The purpose of this section is the construction and investigation of an adaptive estimator of the function $f \in \overline{\mathcal{F}}_r$ with unknown r , which can either serve as the main estimator (since it achieves the optimal rate of convergence within $\overline{\mathcal{F}}_r$) or can serve as a ‘pilot’ estimator to be used in (8) and (9) for the construction of an adaptive optimal and suboptimal bandwidths \hat{h}^0 and \hat{h}_1^0 .

5.1. Adaptive MSE–Optimal Estimation

We define an adaptive estimator of $f \in \overline{\mathcal{F}}_r$ as follows

$$\hat{f}_n(x) = \frac{1}{n} \sum_{j=1}^n \Lambda_{j-1}(x - X_j) = \frac{1}{2\pi n} \sum_{j=1}^n \int \lambda_{j-1}(s) e^{-is(x-X_j)} ds, \quad (14)$$

where $\Lambda_{j-1}(z) = \frac{1}{\hat{h}_{j-1}} K\left(\frac{z}{\hat{h}_{j-1}}\right) = \frac{1}{2\pi} \int \lambda_{j-1}(s) e^{-isz} ds$ is the smoothing kernel, and $\lambda_{j-1}(s) = \lambda_c(s, \hat{h}_{j-1})$; the required bandwidths are defined by

$$\hat{h}_j = (j+1)^{-\frac{1}{1+2r(j)}}, \quad j \geq 1,$$

where $r(j) = r_j^+$ if $f \in \mathcal{F}_r^+$ and $r(j) = r_j$ if $f \in \mathcal{F}_r$; recall that the estimators r_j^+ and r_j are defined in (11) and (12), respectively.

From the definition of $r(j)$ it follows, that $\hat{h}_j \leq \bar{h}_j$, $j \geq 1$, where $\bar{h}_j = (j+1)^{-\frac{1}{1+2q_j}}$. Note, that $\bar{h}_j \leq 1$ and $\bar{h}_j \rightarrow 0$ if the following additional condition

$$\lim_{j \rightarrow \infty} \frac{q_j}{\ln j} = 0 \quad (15)$$

on the sequence (q_j) defined in the beginning of Section 4 holds.

Denote

$$n_1 = \begin{cases} \sup\{n \geq 1 : \Phi_r(A_n, B_n) > H - 1\} & \text{if } f \in \mathcal{F}_r^+, \\ \sup\{n \geq 1 : \Phi_{r-\delta_n}(A_n, B_n) > H - 1\} & \text{if } f \in \mathcal{F}_r, \end{cases}$$

$$n_2 = \begin{cases} \sup\{n \geq 1 : \Phi_{r+\delta_n}(A_n, B_n) < H + 1\} & \text{if } f \in \mathcal{F}_r^+, \\ \sup\{n \geq 1 : \Phi_r(A_n, B_n) < H + 1\} & \text{if } f \in \mathcal{F}_r, \end{cases}$$

where the constant H first used in (11) and (12). Define the following sequences for $j \geq 0$, $0 \leq \gamma \leq r$,

$$h_j = (j+1)^{-\frac{1}{1+2r-2\gamma}}, \quad h_j^* = (j+1)^{-\frac{1}{1+2(r-\gamma-\delta_{j+1})}},$$

$$\tilde{h}_j = (j+1)^{-\frac{1}{1+2(r-\gamma+\delta_{j+1})}} \quad \text{and} \quad \Delta_\gamma(h) = \int_{|s|>h^{-1}} |s|^{r-\gamma} |\phi(s)| ds,$$

as well as the constants

$$C_1 = \bar{f} \cdot \int K^2(u) du, \quad \tilde{C}_1 = C_1 \left(\sum_{j=1}^{n_1} j + C_{r,2m_0} \sum_{j>n_1} \frac{B_j^{4m_0(r+1)}}{j^{2m_0-1}} \right), \quad \tilde{C}_2(\gamma) = \bar{f}^2 + C_2(\gamma),$$

$$C_2(\gamma) = \frac{1}{4\pi^2} \left[\sum_{j=1}^{n_2} \bar{h}_{j-1}^{-2r-2\gamma} \Delta_\gamma^2(\bar{h}_{j-1}) + C_{r+1, m_0} \sum_{j>n_2} \bar{h}_{j-1}^{-2r-2\gamma} \Delta_\gamma^2(\bar{h}_{j-1}) \frac{B_j^{2m_0(r+1+\delta_j)}}{n^{m_0}} \right]$$

and the function

$$\Psi_\gamma(n) = \frac{C_1}{n^2} \sum_{j=1}^n \frac{1}{h_{j-1}^*} + \frac{1}{4\pi^2 n} \sum_{j=1}^n \tilde{h}_{j-1}^{2r-2\gamma} \Delta_\gamma^2(\tilde{h}_{j-1}).$$

Note that the summability of the series in the definitions of the constants \tilde{C}_1 and $C_2(\gamma)$ follows from the corresponding demand in the definition of the classes \mathcal{C}^+ and \mathcal{C} .

Main properties of constructed estimators are stated in the following theorem.

Theorem 3. Let the sequences (A_n, B_n, δ_n) in the definition of the estimator r_n^+ belong to the set \mathcal{C}_+^* and in the definition of the estimator r_n to the set \mathcal{C}_* and the condition (15) is fulfilled.

Let $\gamma = 0$ if $f \in \mathcal{F}_r^+$ and $\gamma \in (0, r)$ if $f \in \mathcal{F}_r$. Then, for every $f \in \mathcal{F}_r$ and $n \geq 1$ the estimator (14) has the following properties:

- 1°. $u_f^2(\hat{f}_n) \leq \Psi_\gamma(n) + \frac{\tilde{C}_1}{n^2} + \frac{\tilde{C}_2(\gamma)}{n}$;
- 2°. the estimator \hat{f}_n is strongly consistent: $\lim_{n \rightarrow \infty} \hat{f}_n(x) = f(x) \quad P_f - a.s.$

Example 1. (Examples 1 and 4 revisited, $f \in \mathcal{F}_r^+$) In this case

$$\frac{1}{n^2} \sum_{j=1}^n \frac{1}{h_{j-1}^*} = \frac{1}{n^2} \sum_{j=1}^n \frac{1}{h_{j-1}} \cdot (\ln j)^{\frac{2\delta}{(1+2r)^2}} \approx \frac{1}{nh_n} \cdot (\ln n)^{\frac{2\delta}{(1+2r)^2}}$$

and

$$\frac{1}{n} \sum_{j=1}^n \tilde{h}_{j-1}^{2r} \Delta_0^2(\tilde{h}_{j-1}) \approx \frac{1}{n} \sum_{j=1}^n h_{j-1}^{2r} \cdot (\ln j)^{\frac{4r\delta}{(1+2r)^2} - 2\varphi}.$$

Thus, under the following conditions

$$m_0 > 16R(R+1) + 1, \quad 4R < b < \frac{m_0 - 1}{4(R+1)}, \quad \frac{\varphi}{b} < \delta < \frac{\varphi}{4R}$$

we have, as $n \rightarrow \infty$,

$$\frac{1}{n} \sum_{j=1}^n \tilde{h}_{j-1}^{2r} \Delta_0^2(\tilde{h}_{j-1}) = o\left(\frac{1}{n^2} \sum_{j=1}^n \frac{1}{h_{j-1}^*}\right)$$

and, as follows,

$$\Psi_0(n) \approx \frac{1}{nh_n} \cdot (\ln n)^{\frac{2\delta}{(1+2r)^2}} \approx \frac{1}{n^{\frac{2r}{1+2r}}} \cdot (\ln n)^{\frac{2\delta}{(1+2r)^2}}.$$

Then, according to Theorem 2, in this case the rate of convergence of adaptive density estimators of $f \in \mathcal{F}_r^+$ differs from the rate of non-adaptive estimators in [26] on the extra log-factor only.

For the functions $f \in \mathcal{F}_r$ and $\gamma \in (0, \min(r, 1))$ from Examples 2 and 4, it is easy to verify, that

$$\Psi_\gamma(n) \approx \frac{1}{n^{\frac{2(r-\gamma)}{1+2(r-\gamma)}}} (\ln n)^{\frac{\delta}{1+2(r-\gamma)}} \quad \text{as } n \rightarrow \infty.$$

5.2. A Symmetric Estimator

Noting that the construction of the estimator $\hat{f}_n(x)$ depends on the order by which the data X_1, \dots, X_n are employed, a simple improvement is immediately available. Let

$X_{[1]} \leq X_{[2]} \leq \dots \leq X_{[n]}$ be the order statistics that are a sufficient statistic in the case of our i.i.d. sample X_1, \dots, X_n . Hence, by the Rao–Blackwell theorem, the estimator

$$E(\hat{f}_n(x) | X_{[1]}, \dots, X_{[n]}) \quad (16)$$

will have smaller (or, at least, not larger) MSE than $\hat{f}_n(x)$.

Unfortunately, the estimator (16) is difficult to compute. However, it is possible to construct a simple estimator that captures the same idea. To do this, consider all distinct permutations of the data X_1, \dots, X_n , and order them in some fashion so that $X_1^{(k)}, \dots, X_n^{(k)}$ is the k th permutation. For unifying presentation, the 1st permutation will be the original data X_1, \dots, X_n . Because of the continuity of the r.v.s X_1, \dots, X_n , the number of such permutations is $n!$ with probability one.

So let $\hat{f}_n^{(k)}(x)$ be the estimator $\hat{f}_n(x)$ as computed from the k th permutation $X_1^{(k)}, \dots, X_n^{(k)}$, i.e., Equation (14) with $X_1^{(k)}, \dots, X_n^{(k)}$ instead of X_1, \dots, X_n .

Finally, let $b \leq n!$ be a positive integer (possibly depending on n), and let

$$\bar{f}_{n,b}(x) = \frac{1}{b} \sum_{k=1}^b \hat{f}_n^{(k)}(x). \quad (17)$$

Theorem 4. For any choice of $b (\leq n!)$, we have $MSE(\bar{f}_{n,b}(x)) \leq MSE(\hat{f}_n(x))$.

Ideally, the practitioner would use a high value of b —even $b = n!$ if the latter is computationally feasible. However, even moderate values of b would give some improvement; in this case, the b permutations to be included in the construction of $\bar{f}_{n,b}(x)$ might be picked randomly as in resampling/subsampling methods—see e.g., [27].

5.3. Adaptive Optimal Bandwidth

Define

$$\begin{aligned} \mathcal{L}(h, \phi) = & \int_{1/h < |s| < c/h} (g(s, h) - 1) \phi(s) e^{-isx} ds \cdot \{c\phi(c/h) e^{-\frac{icx}{h}} + c\phi(-c/h) e^{\frac{icx}{h}} \\ & + h^2 \int_{1/h < |s| < c/h} g'_h(s, h) \phi(s) e^{-isx} ds\}. \end{aligned}$$

According to (8) the optimal bandwidth h^0 is defined from the equality

$$\mathcal{L}(h^0, \phi) = \frac{2\pi^2 L_1 f(x)}{n}.$$

Thus, it is natural to define the adaptive (to the unknown parameter r and the function $f(x)$) optimal bandwidth \hat{h}^0 from the equality

$$\mathcal{L}(\hat{h}^0, \phi_n) = \frac{2\pi^2 L_1 \bar{f}_{n,b}(x)}{n},$$

where the adaptive estimator $\bar{f}_{n,b}(x)$ is defined in (17).

It is hoped that the bandwidths h^0 and \hat{h}^0 have similar asymptotic properties in view of the fact that, according to Theorems 3 and 4 the function

$$n\Psi^{-1/2}(n)[\mathcal{L}(h^0, \phi) - \mathcal{L}(\hat{h}^0, \phi_n)]$$

is bounded in probability.

6. Simulation Results

In this section we provide results of a simulation study regarding the estimators introduced in Section 3.

Two flat-top kernels have been used in the simulation. The first one has the piecewise linear kernel characteristic function introduced in [26], i.e.,

$$\lambda(s) = \begin{cases} 1, & |s| \leq 1, \\ (c - |s|)/(c - 1), & 1 < |s| < c, \\ 0, & |s| \geq c. \end{cases}$$

The piecewise linear characteristic function and corresponding kernel are shown in Figure 1.

The second case refers to the infinitely differentiable flat-top kernel characteristic function defined in [28], i.e.,

$$\lambda(s) = \begin{cases} 1, & |s| \leq c, \\ \exp[-b \exp[-b/(|s| - c)^2]/(|s| - 1)^2], & c < |s| < 1, \\ 0, & |s| \geq 1. \end{cases}$$

The characteristic function and kernel of the second case are shown in Figure 2.

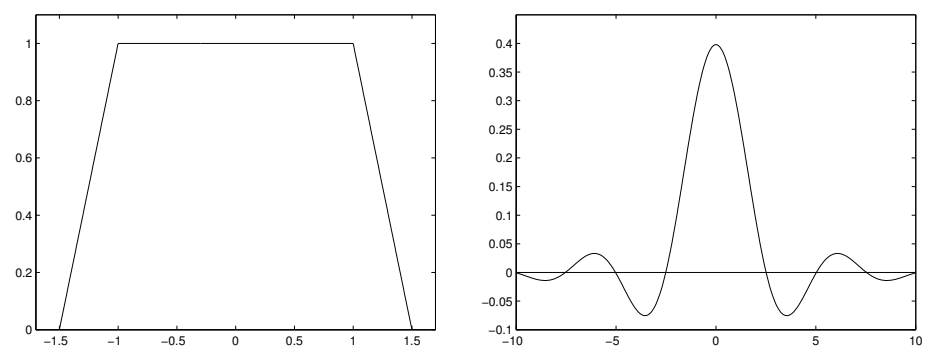


Figure 1. Piecewise linear characteristic function (left) and corresponding kernel (right), $c = 1.5$.

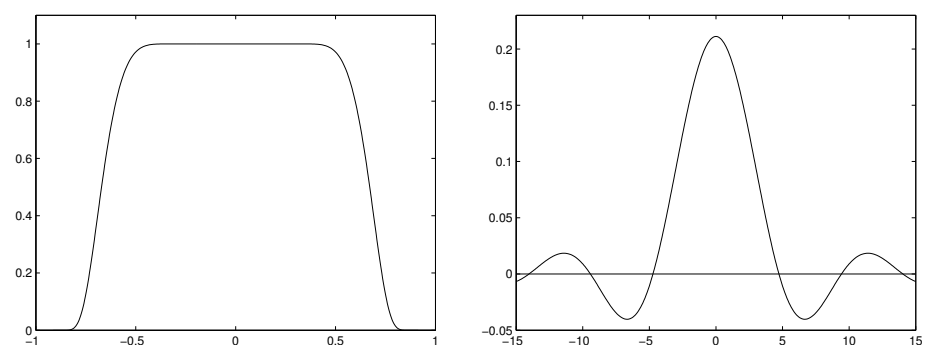


Figure 2. Infinitely differentiable flat-top characteristic function (left) and corresponding kernel (right), $c = 0.05$, $b = 1$.

We examine kernel density estimators of triangular, exponential, Laplace, and gamma (with various shape parameter) distributions. Figures 3–5 illustrate the estimator MSE as a function of the sample size.

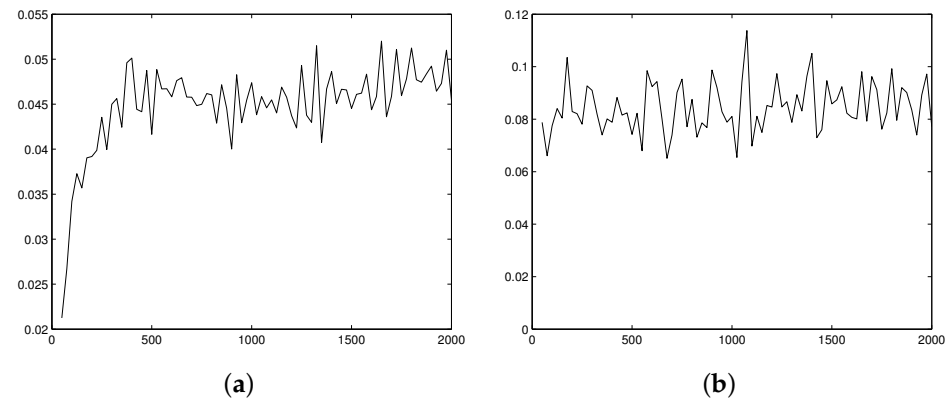


Figure 3. MSE of kernel estimators multiplied by $n^{3/4}$ as a function of the sample size n for the triangle density function. (a) MSE of estimator with piecewise linear kernel characteristic function. (b) MSE of estimator with infinitely differentiable flat-top kernel characteristic function.

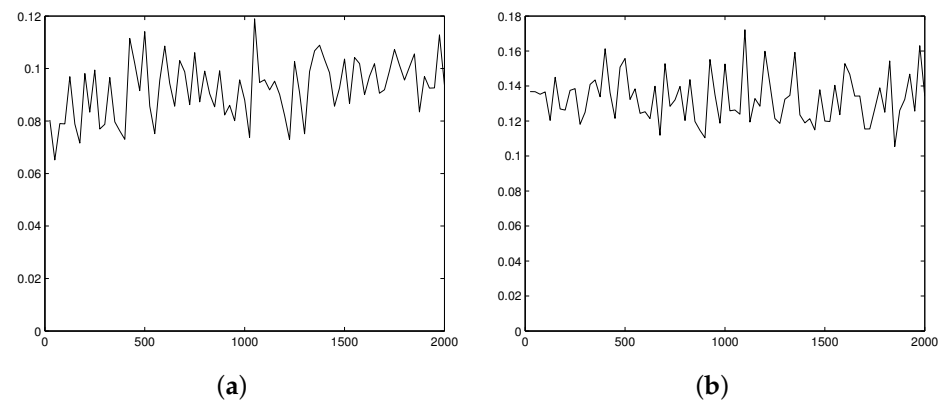


Figure 4. MSE of kernel estimators (with piecewise linear kernel characteristic function) as a function of the sample size n . (a) Laplace density function ($r = 1$, MSE multiplied by $n^{3/4}$). (b) Gamma density function ($k = 3$, $r = 2$, MSE multiplied by $n^{5/6}$).

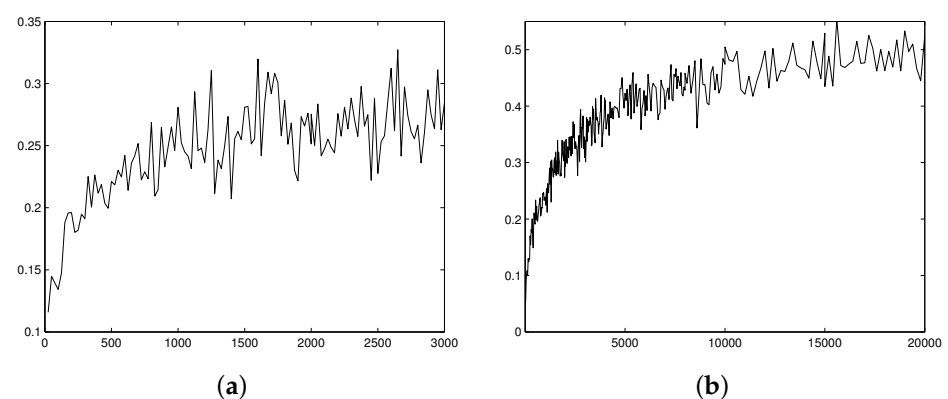


Figure 5. MSE of kernel estimators (with piecewise linear kernel characteristic function) as a function of the sample size n . (a) Gamma distribution shape parameter $k = 4$, $r = 3$ (MSE multiplied by $n^{7/8}$). (b) Gamma distribution shape parameter $k = 6$, $r = 5$ (MSE multiplied by $n^{11/12}$).

Using notation $C(x) = \{0, x < 0; 1, x \geq 0\}$ for Heaviside step function, the triangular density function is defined as $f(x) = ((\lambda - |x|)/\lambda^2)C(\lambda - x)C(\lambda + x)$ having characteristic function $\phi(s) = 2(1 - \cos(\lambda s))/(\lambda s)^2$. Laplace density $f(x) = \lambda/2 \exp(-\lambda|x|)$ has characteristic function $\phi(s) = \lambda^2/(\lambda^2 + s^2)$, gamma density $f(x) = \lambda^k x^{k-1} e^{-\lambda x}/\Gamma(k)$ has characteristic function $\phi(s) = \lambda^k/(\lambda - is)^k$.

In all cases we choose scale parameter λ to have variation equals to 1, and consider estimation of density function $f(x)$ at point $x = 1$.

All the above-mentioned characteristic functions $\phi(s)$ satisfy condition (4) for $r = 1$ (triangular and Laplace), and $r = k - 1$ (gamma, $k > 1$); therefore, all distributions belong to the family \mathcal{F}_r with corresponding value of r . In addition, all $\phi(s)$ meet the requirements of Example 2. Thus, the bandwidth can be taken in the form $h = O(n^{-1/(2(r+1))})$ and the expected convergence rate of the kernel estimator MSE is $n^{-(2r+1)/(2(r+1))}$.

The main goal of the simulation study is investigation of the MSE behavior for the kernel estimator with the growth of sample size. We generate sequences of 150 samples for sample size from 25 to 2000 with step 25, and for some distributions for sample size from 2000 to 20,000 with step 100 or 200. Then, for each sample size we calculate the estimator MSE multiplied by $n^{-(2r+1)/(2(r+1))}$ and expect visual stabilization of the sequence of resulting values with growth of n .

Typical examples of the simulation results are presented at Figure 3 (for $r = 1$), Figure 4 (for $r = 1$ and $r = 2$), and Figure 5 (for $r = 3$ and $r = 5$). The expected stabilization of the scaled MSE is observed in all cases. Moreover, increasing r causes enlargement of sample size that is needed to achieve limiting asymptotic behavior. For $r = 1$ and $r = 2$ we can see stabilization starting from $n \approx 500$, for $r = 3$ it starts from $n \approx 1500$, while for $r = 5$ the asymptotic behavior is observed to start from sample size $n \approx 15,000$.

7. Technical Proofs

7.1. Proof of Lemma 1

First we note that for every $m \geq 1$ and $n \geq 1$ there exist positive numbers κ_m such that

$$\sup_{f \in \overline{\mathcal{F}}_r} E_f |\phi_n(s) - \phi(s)|^{2m} \leq \frac{\kappa_m}{n^m}.$$

These inequalities follow from the Burkholder inequality (see, for example, [29]) for the martingale $(\sum_{k=1}^n (e^{isX_k} - \phi(s)), \mathcal{F}_n^X), \mathcal{F}_n^X = \sigma\{X_1, \dots, X_n\}$ and finiteness of the function $\phi(\cdot)$.

Using this and Hölder's inequalities we can estimate

$$\begin{aligned} \sup_{f \in \overline{\mathcal{F}}_r} E_f J_{n,\alpha}^{2m} &= \sup_{f \in \overline{\mathcal{F}}_r} E_f \left[\int_{A_n < |s| < B_n} |s|^\alpha |\phi_n(s) - \phi(s)| ds \right]^{2m} \\ &\leq \left(\int_{A_n < |s| < B_n} |s|^{\frac{2m\alpha}{2m-1}} ds \right)^{2m-1} \cdot \int_{A_n < |s| < B_n} \sup_{f \in \overline{\mathcal{F}}_r} E_f |\phi_n(s) - \phi(s)|^{2m} ds \leq \frac{C_{\alpha,m} B_n^{2m(\alpha+1)}}{n^m}. \end{aligned}$$

From the Borel–Cantelli lemma and the assumed summability of the right-hand side of (13) for $m = m_0$, $\alpha = \varrho_n + \delta_n$ and $f \in \overline{\mathcal{F}}_r$ follows the second assertion of Lemma 1.

7.2. Proof of Theorem 2

We prove now the statements 1° (a) and 2° (a) of Theorem 1. First, we show for n large enough the inequalities

$$r_n^+ < \varrho_n \quad P_f - \text{a.s.} \quad (18)$$

To this end, according to the definition of the estimator r_n^+ , it is enough to establish for some $\alpha > r$ the limiting relation

$$\lim_{n \rightarrow \infty} \Phi_{n,\alpha}(A_n, B_n) = \infty \quad P_f - \text{a.s.},$$

which follows from the definition of the class \mathcal{C}_+ and Lemma 1:

$$\begin{aligned}\Phi_{n,\alpha}(A_n, B_n) &= \Phi_{n,\alpha}(A_n, B_n) - \Phi_\alpha(A_n, B_n) + \Phi_\alpha(A_n, B_n) \\ &\geq \Phi_\alpha(A_n, B_n) - J_{n,\alpha} \rightarrow \infty \quad P_f - \text{a.s.}\end{aligned}$$

From (18) and by the definition of the estimator r_n^+ , for n large enough, we have

$$\Phi_{n,r_n^+ + \delta_n}(A_n, B_n) \geq H.$$

Thus,

$$\begin{aligned}\lim_{n \rightarrow \infty} \Phi_{r_n^+ + \delta_n}(A_n, B_n) &\geq \lim_{n \rightarrow \infty} \Phi_{n,r_n^+ + \delta_n}(A_n, B_n) \\ &- \overline{\lim}_{n \rightarrow \infty} J_{n,r_n^+ + \delta_n} \geq H \quad P_f - \text{a.s.}\end{aligned} \quad (19)$$

Analogously,

$$\overline{\lim}_{n \rightarrow \infty} \Phi_{r_n^+}(A_n, B_n) \leq H \quad P_f - \text{a.s.} \quad (20)$$

From (19) and (20) it follows, that for any $\varepsilon > 0$ and $\delta > 0$ for n large enough

$$r - \varepsilon - \delta < r_n^+ < r + \varepsilon \quad P_f - \text{a.s.}$$

and the assertion 1(a) of Theorem 2 is proved.

From the definitions of the estimator r_n^+ , class \mathcal{C}_+^* , Chebyshev's inequality and (13), for $m \geq 1$, $n > n_1$ and $f \in \mathcal{F}_r^+$ we have

$$\begin{aligned}P_f(r_{n-1}^+ < r - \delta_{n-1}) &\leq P_f(\Phi_{n,r}(A_n, B_n) \geq H) \leq P_f(J_{n,r} \geq H - \Phi_r(A_n, B_n)) \\ &\leq \frac{E_f J_{n,r}^{2m}}{(H - \Phi_r(A_n, B_n))^{2m}} \leq \frac{C_{r,m} B_n^{2m(r+1)}}{n^m}.\end{aligned} \quad (21)$$

Similar to (21) for $n > n_2$ and $m \geq 1$ we obtain

$$\begin{aligned}P_f(r_{n-1}^+ > r + \delta_{n-1}) &\leq P_f(\Phi_{n,r+\delta_n}(A_n, B_n) < H) = P_f(\Phi_{r+\delta_n}(A_n, B_n) - H \leq J_{n,r+\delta_n}) \\ &\leq \frac{E_f J_{n,r+\delta_n}^{2m}}{(\Phi_{r+\delta_n}(A_n, B_n) - H)^{2m}} \leq \frac{C_{r+1,m} B_n^{2m(r+1+\delta_n)}}{n^m}\end{aligned} \quad (22)$$

and, as follows, for $f \in \mathcal{F}_r^+$,

$$P_f(\delta_n^{-1} |r_n^+ - r| \geq 1) \leq \frac{2C_{r+1,m} B_n^{2m(r+1+\delta_n)}}{n^m}. \quad (23)$$

From the Borel–Cantelli lemma and the assumed summability of the right hand side in (23) for $m \geq m_0$ follows the assertion 2(a) of Theorem 2.

The other statements of Theorem 2 for the estimator r_n can be proved analogically.

7.3. Proof of Theorem 3

Consider the deviation of the estimator (14) in the following form:

$$\hat{f}_n(x) - f(x) = I_1(n) + I_2(n), \quad (24)$$

where

$$\begin{aligned}I_1(n) &= \hat{f}_n(x) - \tilde{f}_n(x), \quad I_2(n) = \tilde{f}_n(x) - f(x), \\ \tilde{f}_n(x) &= \frac{1}{n} \sum_{j=1}^n \int K(z) f(x - \hat{h}_{j-1}z) dz = \frac{1}{2\pi n} \sum_{j=1}^n \int \lambda_{j-1}(s) \phi(s) e^{-isx} ds.\end{aligned}$$

Now we estimate second moments of $I_1(n)$ and $I_2(n)$. Denote $\mathcal{F}_j^X = \sigma\{X_1, \dots, X_j\}$. For $f \in \overline{\mathcal{F}}_r$ we have

$$\begin{aligned} E_f I_1^2(n) &= \frac{1}{n^2} E_f \left(\sum_{j=1}^n \frac{1}{\hat{h}_{j-1}} \left[K\left(\frac{x - X_j}{\hat{h}_{j-1}}\right) - \hat{h}_{j-1} \int K(z) f(x - \hat{h}_{j-1}z) dz \right] \right)^2 \\ &= \frac{1}{n^2} E_f \sum_{j=1}^n \frac{1}{\hat{h}_{j-1}^2} E_f \left(\left[K\left(\frac{x - X_j}{\hat{h}_{j-1}}\right) - \hat{h}_{j-1} \int K(z) f(x - \hat{h}_{j-1}z) dz \right]^2 \middle| \mathcal{F}_{j-1}^X \right) \\ &= \frac{1}{n^2} E_f \sum_{j=1}^n \frac{1}{\hat{h}_{j-1}} \int [K(u) - \hat{h}_{j-1} \int K(z) f(x - \hat{h}_{j-1}z) dz]^2 f(x - \hat{h}_{j-1}u) du \\ &\leq \frac{C_1}{n^2} \sum_{j=1}^n E_f \frac{1}{\hat{h}_{j-1}} + \frac{\bar{f}^2}{n} \leq \frac{C_1}{n^2} \sum_{j=1}^n \frac{1}{h_{j-1}^*} + \frac{C_1}{n^2} \sum_{j=1}^n j P_f(r(j) < r - \delta_j) + \frac{\bar{f}^2}{n}. \end{aligned}$$

From (21) for $m = 2m_0$ we obtain

$$\begin{aligned} E_f I_1^2(n) &\leq \frac{C_1}{n^2} \sum_{j=1}^n \frac{1}{h_{j-1}^*} + \frac{C_1}{n^2} \sum_{j=1}^{n_1} j + \frac{C_1 C_{r, 2m_0}}{n^2} \sum_{j > n_1} \frac{j B_j^{4m_0(r+1)}}{j^{2m_0}} + \frac{\bar{f}^2}{n} \\ &= \frac{C_1}{n^2} \sum_{j=1}^n \frac{1}{h_{j-1}^*} + \frac{\tilde{C}_1}{n^2} + \frac{\bar{f}^2}{n}. \end{aligned} \quad (25)$$

Further, by the definition of the function $\tilde{f}(x)$, the Cauchy–Bunyakovskii–Schwarz inequality and from (22) we have

$$\begin{aligned} E_f I_2^2(n) &= \frac{1}{4\pi^2 n^2} E_f \left[\sum_{j=1}^n \int_{|s| \geq \hat{h}_{j-1}^{-1}} (\lambda_{j-1}(s) - 1) \phi(s) e^{-isx} ds \right]^2 \\ &\leq \frac{1}{4\pi^2 n} \sum_{j \geq 1} E_f \left[\int_{|s| \geq \hat{h}_{j-1}^{-1}} |\phi(s)| ds \right]^2 \leq \frac{1}{4\pi^2 n} \sum_{j=1}^n E_f \hat{h}_{j-1}^{2r-2\gamma} \cdot \left[\int_{|s| \geq \hat{h}_{j-1}^{-1}} |s|^{r-\gamma} |\phi(s)| ds \right]^2 \\ &= \frac{1}{4\pi^2 n} \sum_{j=1}^n E_f \hat{h}_{j-1}^{2r-2\gamma} \Delta_\gamma^2(\hat{h}_{j-1}) \leq \frac{1}{4\pi^2 n} \left[\sum_{j=1}^n \tilde{h}_{j-1}^{2r-2\gamma} \Delta_\gamma^2(\tilde{h}_{j-1}) + \sum_{j=1}^{n_2} \bar{h}_{j-1}^{2r-2\gamma} \Delta_\gamma^2(\bar{h}_{j-1}) \right. \\ &\quad \left. + \sum_{j > n_2} \bar{h}_{j-1}^{2r-2\gamma} \Delta_\gamma^2(\bar{h}_{j-1}) P_f(r(j) > r - \gamma + \delta_j) \right] \leq \frac{1}{4\pi^2 n} \left[\sum_{j=1}^n \tilde{h}_{j-1}^{2r-2\gamma} \Delta_\gamma^2(\tilde{h}_{j-1}) \right. \\ &\quad \left. + \sum_{j=1}^{n_2} \bar{h}_{j-1}^{2r-2\gamma} \Delta_\gamma^2(\bar{h}_{j-1}) + C_{r+1, m_0} \sum_{j > n_2} \bar{h}_{j-1}^{2r-2\gamma} \Delta_\gamma^2(\bar{h}_{j-1}) \frac{B_j^{2m_0(r+1+\delta_j)}}{n^{m_0}} \right] \\ &\leq \frac{1}{4\pi^2 n} \sum_{j=1}^n \tilde{h}_{j-1}^{2r-2\gamma} \Delta_\gamma^2(\tilde{h}_{j-1}) + \frac{C_2}{n}. \end{aligned} \quad (26)$$

From (24)–(26) follows the first assertion of Theorem 3.

For the proof of the second assertion we estimate first, for some integer $m > 1$ the rate of convergence of the moment $E_f I_1^{2m}(n)$.

Let $\alpha_1, \dots, \alpha_m$ be non-negative integers and denote

$$\tilde{K}_j(x) = K\left(\frac{x - X_j}{\hat{h}_{j-1}}\right) - \hat{h}_{j-1} \int K(z) f(x - \hat{h}_{j-1}z) dz,$$

$$\bar{K}_j(x) = \int [K(u) - \hat{h}_{j-1} \int K(z) f(x - \hat{h}_{j-1}z) dz]^2 f(x - \hat{h}_{j-1}u) du.$$

By the Burkholder inequality for the martingale $\sum_{j=1}^n \frac{1}{\hat{h}_{j-1}} \tilde{K}_j(x)$ we have

$$\begin{aligned} E_f I_1^{2m}(n) &= \frac{1}{n^{2m}} E_f \left(\sum_{j=1}^n \frac{1}{\hat{h}_{j-1}} \tilde{K}_j(x) \right)^{2m} \leq \frac{C}{n^{2m}} E_f \left(\sum_{j=1}^n \frac{1}{\hat{h}_{j-1}^2} \tilde{K}_j^2(x) \right)^m \\ &\leq \frac{C}{n^{2m}} \sum_{\alpha_1 + \dots + \alpha_m = m} \sum_{1 \leq j_1 < \dots < j_m \leq n} E_f \frac{1}{\hat{h}_{j_1-1}^2 \dots \hat{h}_{j_m-1}^2} \tilde{K}_{j_1}^2(x) \dots \tilde{K}_{j_m}^2(x) \\ &\leq \frac{C}{n^{2m}} \sum_{\alpha_1 + \dots + \alpha_m = m} \sum_{1 \leq j_1 < \dots < j_m \leq n} E_f \frac{\tilde{K}_{j_1}^2(x) \dots \tilde{K}_{j_{m-1}}^2(x) \cdot \bar{K}_{j_m}(x)}{\hat{h}_{j_1-1}^2 \dots \hat{h}_{j_{m-1}-1}^2 \cdot h_{j_m-1}^*} \\ &\leq \frac{C}{n^{2m}} \sum_{\alpha_1 + \dots + \alpha_m = m} \sum_{1 \leq j_1 < \dots < j_m \leq n} E_f \frac{1}{\hat{h}_{j_1-1}^2 \dots \hat{h}_{j_{m-2}-1}^2 \cdot h_{j_{m-1}-1}^* \cdot h_{j_m-1}^*} \\ &\quad \cdot \tilde{K}_{j_1}^2(x) \dots \tilde{K}_{j_{m-2}}^2(x) \cdot \bar{K}_{j_{m-1}}(x) \leq \dots \leq \frac{C}{n^{2m}} \left(\sum_{j=1}^n \frac{1}{\hat{h}_{j-1}} \right)^m. \end{aligned}$$

By the definition of h_j^* for some $0 < r_* < r - \gamma$ and $j_* < \infty$, $h_j^* \geq n^{-\frac{1}{1+2r_*}}$ and, as follows

$$E_f I_1^{2m}(n) = O\left(n^{-\frac{2mr_*}{1+2r_*}}\right) \text{ as } n \rightarrow \infty.$$

Thus $\frac{2mr_*}{1+2r_*} > 1$ for $m > \frac{1+2r_*}{2r_*}$ and by the Borel–Cantelli lemma, as $n \rightarrow \infty$,

$$I_1(n) \rightarrow 0 \quad P_f - \text{a.s.} \quad (27)$$

Further,

$$|I_2(n)| \leq \frac{1}{4\pi^2 n} \sum_{j=1}^n \hat{h}_{j-1}^r |\Delta_\gamma(\hat{h}_{j-1})|$$

and, as follows, as $n \rightarrow \infty$,

$$I_2(n) \rightarrow 0 \quad P_f - \text{a.s.} \quad (28)$$

From (27) and (28) follows the second assertion of Theorem 3.

7.4. Proof of Theorem 4

Note that the distribution of $\hat{f}_n^{(k)}(x)$ is the same as that of $\hat{f}_n^{(j)}(x)$ for all k, j . Hence, $E\bar{f}_{n,b}(x) = E\hat{f}_n(x)$. Now $|\text{Cov}(\hat{f}_n^{(k)}(x), \hat{f}_n^{(j)}(x))| \leq \text{Var}(\hat{f}_n^{(1)}(x))$ by the Cauchy–Schwarz inequality and the fact that $\text{Var}(\hat{f}_n^{(k)}(x)) = \text{Var}(\hat{f}_n^{(j)}(x))$. Thus,

$$\text{Var}(\bar{f}_{n,b}(x)) = \frac{1}{b^2} \sum_{k=1}^b \sum_{j=1}^b \text{Cov}(\hat{f}_n^{(k)}(x), \hat{f}_n^{(j)}(x)) \leq \text{Var}(\hat{f}_n^{(1)}(x))$$

and the theorem is proven.

8. Conclusions

Non-parametric kernel estimation crucially depends on the bandwidth choice which, in turn, depends on the smoothness of the underlying function. Focusing on estimating a probability density function, we define a smoothness class and propose a data-based estimator of the underlying degree of smoothness. The convergence rates in the almost sure sense of the proposed estimators are obtained. Adaptive estimators of densities from the given class on the basis of the constructed smoothness parameter estimators are also presented, and their consistency is established. Simulation results illustrate the realization of the asymptotic behavior when the sample size grows large.

Recently, there has been an increasing interest in nonparametric estimation with dependent data both in terms of theory as well as applications; see, e.g., [15,30–33]. With respect to probability density estimation, many asymptotic results remain true when moving from i.i.d. data to data that are weakly dependent. For example, the estimator variance, bias and MSE have the same asymptotic expansions as in the i.i.d. case subject to some limitations on the allowed bandwidth rate; fortunately, the optimal bandwidth rate of $n^{-1/5}$ is in the allowed range—see [34,35].

Consequently, it is conjectured that our proposed estimator of smoothness—as well as resulting data-based bandwidth choice and probability density estimator—will retain their validity even when the data are weakly dependent. Future work may confirm this conjecture especially since working with dependent data can be quite intricate. For example, [36] extended the results of [34] from the realm of linear time series to strong-mixing process. In so doing, Remark 5 of [36] pointed to a nontrivial error in the work of [34] which is directly relevant to optimal bandwidth choice.

Author Contributions: All authors contributed equally to this project. All authors have read and agreed to the published version of the manuscript.

Funding: Partial funding by NSF grant DMS 19-14556.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors are very thankful to O. Lepskii for helpful comments and remarks.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rosenblatt, M. *Stochastic Curve Estimation*; NSF-CBMS Regional Conference Series in Probability and Statistics, 3; Institute of Mathematical Statistics: Hayward, CA, USA, 1991.
2. Bartlett, M.S. Statistical estimation of density function. *Sankhya Indian J. Stat.* **1963**, *A25*, 245–254.
3. Parzen, E. On estimation of a probability density function and mode. *Ann. Math. Stat.* **1962**, *33*, 1065–1076. [\[CrossRef\]](#)
4. Devroye, L. *A Course in Density Estimation*; Birkhäuser: Boston, MA, USA; Basel, Switzerland; Stuttgart, Germany, 1987.
5. Gasser, T.; Müller, H.-G.; Marmmützsch, V. Kernels for nonparametric curve estimation. *J. R. Stat. Soc. Ser. B* **1985**, *60*, 238–252. [\[CrossRef\]](#)
6. Granovsky, B.L.; Müller, H.-G. Optimal kernel methods: A unifying variational principle. *Int. Stat. Rev.* **1991**, *59*, 373–388. [\[CrossRef\]](#)
7. Jones, M.C. On higher order kernels. *J. Nonparametric Stat.* **1995**, *5*, 215–221. [\[CrossRef\]](#)
8. Marron, J.S. Visual understanding of higher order kernels. *J. Comput. Graph. Stat.* **1994**, *3*, 447–458.
9. Müller, H.-G. *Nonparametric Regression Analysis of Longitudinal Data*; Springer: Berlin/Heidelberg, Germany, 1988.
10. Scott, D.W. *Multivariate Density Estimation: Theory, Practice and Visualization*; Wiley: New York, NY, USA, 1992.
11. Devroye, L. A note on the usefulness of superkernels in density estimation. *Ann. Stat.* **1992**, *20*, 2037–2056. [\[CrossRef\]](#)
12. Politis, D.N. On nonparametric function estimation with infinite-order flat-top kernels. In *Probability and Statistical Models with Applications*; Charalambides, C., Ed.; Chapman and Hall/CRC: Boca Raton, FL, USA, 2001; pp. 469–483.
13. Jones, M.C.; Marron, J.S.; Sheather, S.J. A brief survey of bandwidth selection for density estimation. *J. Am. Stat. Assoc.* **1996**, *91*, 401–407. [\[CrossRef\]](#)
14. Tsybakov, A. *Introduction to Nonparametric Estimation*; Springer Series in Statistics; Springer: New York, NY, USA, 2009.

15. Dobrovodov, A.V.; Koshkin, G.M.; Vasiliev, V.A. *Non-Parametric State Space Models*; Kendrick Press: Heber City, UT, USA, 2012.
16. Ibragimov, I.A.; Khasminskii, R.Z. *Statistical Estimation: Asymptotic Theory*; Springer: Berlin/Heidelberg, Germany, 1981.
17. Lehmann, E.L.; Romano, J.P. *Testing Statistical Hypotheses*, 4th ed.; Springer Texts in Statistics: New York, NY, USA, 2022.
18. Lepskii, O.V.; Spokoiny, V.G. Optimal pointwise adaptive methods in nonparametric estimation. *Ann. Stat.* **1997**, *25*, 2512–2546. [[CrossRef](#)]
19. Brown, L.D.; Low, M.G. *Superefficiency and Lack of Adaptability in Functional Estimation*; Technical Report; Cornell University: Ithaca, NY, USA, 1992.
20. Lepskii, O.V. On a problem of adaptive estimation in Gaussian white noise. *Theory Probab. Its Appl.* **1990**, *35*, 454–466. [[CrossRef](#)]
21. Butucea, C. Exact adaptive pointwise estimation on Sobolev classes of densities. *ESAIM Probab. Stat.* **2001**, *5*, 1–31. [[CrossRef](#)]
22. Goldenshluger, A.; Lepski, O. Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **2011**, *39*, 1608–1632. [[CrossRef](#)]
23. Lacour, C.; Massart, P.; Rivoirard, V. Estimator selection: A new method with applications to kernel density estimation. *Sankhya A* **2017**, *79*, 298–335. [[CrossRef](#)]
24. Politis, D.N. Adaptive bandwidth choice. *J. Nonparametric Stat.* **2003**, *15*, 517–533. [[CrossRef](#)]
25. Politis, D.N.; Romano, J.P. On a Family of Smoothing Kernels of Infinite Order. In *Computing Science and Statistics, Proceedings of the 25th Symposium on the Interface, San Diego, CA, USA, 14–17 April 1993*; Tarter, M., Lock, M., Eds.; The Interface Foundation of North America: San Diego, CA, USA, 1993; pp. 141–145.
26. Politis, D.N.; Romano, J.P. Multivariate density estimation with general flat-top kernels of infinite order. *J. Multivar. Anal.* **1999**, *68*, 1–25. [[CrossRef](#)]
27. Politis, D.N.; Romano, J.P.; Wolf, M. *Subsampling*; Springer: New York, NY, USA, 1999.
28. McMurtry, T.; Politis, D.N. Nonparametric regression with infinite order flat-top kernels. *J. Nonparametric Stat.* **2004**, *16*, 549–562. [[CrossRef](#)]
29. Liptser, R.; Shiryaev, A. *Theory of Martingales*; Springer: New York, NY, USA, 1988.
30. Bijloos, G.; Meyers, J.A. Fast-Converging Kernel Density Estimator for Dispersion in Horizontally Homogeneous Meteorological Conditions. *Atmosphere* **2021**, *12*, 1343. [[CrossRef](#)]
31. Cortes Lopez, J.C.; Jornet Sanz, M. Improving Kernel Methods for Density Estimation in Random Differential Equations Problems. *Math. Comput. Appl.* **2020**, *25*, 33. [[CrossRef](#)]
32. Correa-Quezada, R.; Cueva-Rodriguez, L.; Alvarez-Garcia, J.; del Rio-Rama, M.C. Application of the Kernel Density Function for the Analysis of Regional Growth and Convergence in the Service Sector through Productivity. *Mathematics* **2020**, *8*, 1234. [[CrossRef](#)]
33. Vasiliev, V.A. A truncated estimation method with guaranteed accuracy. *Ann. Inst. Stat. Math.* **2014**, *66*, 141–163. [[CrossRef](#)]
34. Hallin, M.; Tran, L.T. Kernel density estimation for linear processes: Asymptotic normality and bandwidth selection. *Ann. Inst. Stat. Math.* **1996**, *48*, 429–449. [[CrossRef](#)]
35. Wu, W.-B.; Mielniczuk, J. Kernel density estimation for linear processes. *Ann. Stat.* **2002**, *30*, 1441–1459. [[CrossRef](#)]
36. Lu, Z. Asymptotic normality of kernel density estimators under dependence. *Ann. Inst. Stat. Math.* **2001**, *53*, 447–468. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.