Adaptive Estimation of Density Function Derivative

DIMITRIS N. POLITIS[1], VYACHESLAV A. VASILIEV[2] AND PETER F. TARASSENKO[2]
[1] *Department of Mathematics, University of California, San Diego, USA*
[2] *Department of Applied Mathematics and Cybernetics, Tomsk State University, Tomsk, Russia*
e-mail: dpolitis@ucsd.edu, vas@mail.tsu.ru, ptara@mail.tsu.ru

Abstract

The properties of non-parametric kernel estimators for the first order derivative of probability density function from special parameterized classes are investigated. In particular, in the case of known smooth classes parameter, rates of mean square convergency of density and its derivative estimators of smooth parameter estimators are found. Adaptive estimators of densities and their first derivatives from the given class with the unknown smooth parameter are constructed. Non-asymptotic and asymptotic properties of these estimators are established.

Keywords: Non-parametric kernel density estimators, smooth parameter estimation; adaptive density derivative estimators, mean square convergence, rate of convergence, smoothness class.

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 nonparametric 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 by [13]. For some point x, the kernel smoothed estimator of $f(x)$ is defined by

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

where the kernel K 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. In this paper we will employ a particularly useful class of infinite order kernels namely the *flat-top* family; see [7] for a general definition.

It is a well-known fact that optimal bandwidth selection is perhaps the most crucial issue in such nonparametric smoothing problems; see [3] and the references therein. The goal typically is minimization of the large-sample Mean Squared Error (MSE) of $f_{n,h}(x)$. However, to do this minimization, the practitioner needs to know the degree of smoothness r. 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 $n^{-1/(2r+1)}$ that yields a large-sample MSE of order $n^{-2r/(2r+1)}$ (see, e.g., [2]).

The problem of course is that, as previously mentioned, the underlying degree of smoothness r is typically unknown. In Section 3 of the paper at hand, we develop an estimator r_n of r and prove its strong consistency. In order to construct our estimator r_n, we operate under a class of functions that is slightly more general than, e.g., the Hölder class; this class of functions is formally defined in Section 1 via eq. (3) or (4). Under such a condition on the tails of the characteristic function we are able to show in Section 2 that the optimized MSE of $\hat{f}_n(x)$ is again of order $n^{-2r/(2r+1)}$ for possibly noninteger r.

Furthermore, in Section 4 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, in particular, from the Hölder class, see [1, 4, 5].

The estimaton problem of the density derivatives is actual as well; in particular for estimation of the logarithmic derivative. As the major theoretical result of our paper, we are able to prove a non-asymptotic upper bound for the MSE of the adaptive estimator of the density f and f'. The rate of convergency in the mean square sense satisfies (for the estimators of f in examples) the abovementioned optimal rate.

Section 5 contains some simulation results showing the performance of the estimator $\hat{f}_n(x)$ in practice.

Full investigation of the density function estimators will be presented in the paper [12].

1 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.,

$$\exists\ 0 < \overline{f} < \infty : \ \sup_{y \in \mathcal{R}^1} f(y) \le \overline{f}, \tag{2}$$

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

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

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

It is easy to verify that the derivative f' satisfies the relations (3) and (4) if $f \in \mathcal{F}_{r+1}^+$ and $f \in \mathcal{F}_{r+1}$ respectively.

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

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

$$f_{n,h}^{(l)}(x) = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{h^{1+l}}K^{(l)}\left(\frac{x-X_j}{h}\right) = \frac{1}{2\pi}\int \lambda^{(l)}(s,h)\phi_n(s)e^{-isx}ds, \quad l=0;1, \quad (5)$$

where $\lambda^{(0)}(s,h) = \int K\left(\frac{x}{h}\right)e^{isx}dx$ and $\lambda^{(1)}(s,h) = \int K'\left(\frac{x}{h}\right)e^{isx}dx = -ish\lambda^{(0)}(s,h)$. In this paper, we will employ the family of flat-top infinite order kernels, i.e., we will let the function $\lambda^{(0)}(s,h)$ be of the form

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

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 $|g(s,h)| \le 1$, for any s, with $g(1/h,h) = 1$, and $g(c/h,h) = 0$; see [7]-[10] for more details on the above flat-top family of kernels.

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

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

From (3) and (5) it follows that $\delta_\gamma(h) = o(1)$ as $h \to 0$ 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 adaptive estimation of densities and their first derivatives from the class $\overline{\mathcal{F}}_r$ with the unknown r.

2 Asymptotic mean square optimal estimation of f

According to [10, 11] 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:

$$u_f^2(f_{n,h}) = U_f^2(h,c) - \frac{1}{n}\left(\int K(v)f(x-hv)dv\right)^2, \quad (6)$$

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

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

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

The optimal (in the mean square sense) value $h^0 = h^0(n)$ is defined from minimization of the principal term $U_f^2(h,c)$.

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

$$(h_1^0)^{2r+1-2\gamma}\delta_\gamma^2(h_1^0) = \frac{\pi^2 L_1 f(x)}{(c_0 + c_1(\gamma))n}. \tag{7}$$

In such a way we have proved the following theorem, which gives the rates of convergence of the random quantities $f_n^0(x) = f_{n,h^0}(x)$ 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, e.g., Examples 1 and 2 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 $n \to \infty$, hold*

1°. $\sup\limits_{f\in\overline{\mathcal{F}}_r} \left|\inf\limits_h u_f^2(f_{n,h}) - U_f^2(h^0, c)\right| = O\left(\frac{1}{n}\right)$;

2°. *for every $f \in \overline{\mathcal{F}}_{r,m}$ with $\gamma = 0$ if $f \in \mathcal{F}_{r,m}^+$ and every $0 < \gamma < r$ if $f \in \mathcal{F}_{r,m}$, as well as some constant C_γ, we have*

$$u_f^2(f_n^0) \le u_f^2(f_{n,h_1^0}) \le C_\gamma \cdot \frac{\delta_\gamma^{\frac{2}{2r+1-2\gamma}}(h_1^0)}{n^{\frac{2r-2\gamma}{2r+1-2\gamma}}}, \quad n \ge 1.$$

Remark 1. *The definition (7) of h_1^0 is essentially simpler than the definition 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.*

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

$$|\phi(s)| \approx \frac{1}{|s|^{r+1}\ln^{1+\varphi}|s|} \quad as \ |s| \to \infty, \quad \varphi > 0. \tag{8}$$

We find the rates of convergence of the MSE $u_f^2(f_n^0)$ and $u_f^2(f_{n,h_1^0})$:

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

and using the piecewise linear flat-top kernel $\lambda_c^{LIN}(s,h)$, introduced by [9] (see [10] as well)

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

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

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

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

$$|\phi(s)| \approx \frac{1}{|s|^{r+1}} \quad as \quad |s| \to \infty.$$

We find the rate of convergence of the MSE $u_f^2(f_n^0)$ and $u_f^2(f_{n,h_1^0})$. From (7) we have

$$h_1^0 \approx n^{-\frac{1}{2r+1}} \qquad and \qquad u_f^2(f_{n,h_1^0}) = O\left(n^{-\frac{2r}{2r+1}}\right), \qquad as\ n \to \infty.$$

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

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

Similar results can be obtained for the estimators of f'.

3 Estimation of the degree of smoothness r

Define the functions

$$\Phi_\alpha(A,B) = \int\limits_{A<|s|<B} |s|^\alpha |\phi(s)| ds, \qquad \Phi_{n,\alpha}(A,B) = \int\limits_{A<|s|<B} |s|^\alpha |\phi_n(s)| ds.$$

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 $\delta_n \to 0$ and $\rho_n \to \infty$ as $n \to \infty$. 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

$$C_+ = \{(A_n, B_n, \delta_n)_{n\geq 1} :\ A_n \to \infty,\ 0 < A_n < B_n \to \infty,\ \delta_n \to 0 \text{ as } n \to \infty; \text{ for some } m_0 \geq 2,$$
$$\sum_{n\geq 1} \frac{B_n^{2m_0(\varrho_n+1+\delta_n)}}{n^{m_0}} < \infty;\ \Phi_{r+\varepsilon}(A_n,B_n) \to \infty,\ \forall \varepsilon > 0;\ \Phi_{r+\delta_n}(A_n,B_n) \to \infty\},$$

$$C = \{(A_n, B_n, \delta_n)_{n\geq 1} :\ A_n \to \infty,\ 0 < A_n < B_n \to \infty,\ \delta_n \to 0 \text{ as } n \to \infty; \text{ for some } m_0 \geq 2,$$
$$\sum_{n\geq 1} \frac{B_n^{2m_0(\varrho_n+1+\delta_n)}}{n^{m_0}} < \infty;\ \Phi_{r-\delta_n}(A_n,B_n) \to 0;\ \Phi_r(A_n,B_n) \to \infty\}$$

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

$$r_n^+ = \min[\varrho_n, (\delta_n \cdot \inf\{k \geq 1 :\ \Phi_{n,(k+1)\delta_n}(A_n,B_n) \geq H,\ (A_n,B_n,\delta_n) \in C_+\})], \tag{9}$$

$$r_n = \min[\varrho_n, (\delta_n \cdot \inf\{k \geq 1 :\ \Phi_{n,k\delta_n}(A_n,B_n) \geq H,\ (A_n,B_n,\delta_n) \in C\})]. \tag{10}$$

Theorem 2. *The estimators r_n^+ and r_n, defined in (9) and (10) respectively have the following properties*

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

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

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

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

4 Adaptive estimation of the functions $f, f' \in \overline{\mathcal{F}}_r$

The purpose of this section is the construction and investigation of an adaptive estimator of the functions $f, f' \in \overline{\mathcal{F}}_r$ with unknown r, which can either serve as the main estimator or can serve as a 'pilot' estimator for the construction of an adaptive optimal and suboptimal bandwidths $\hat{h}^0$ and $\hat{h}_1^0$.

We define an adaptive estimators of f and f' from $\overline{\mathcal{F}}_r$ as follows

$$\hat{f}_n^{(l)}(x) = \frac{1}{n}\sum_{j=1}^{n} \Lambda_{j-1}^{(l)}(x - X_j) = \frac{1}{2\pi n}\sum_{j=1}^{n} \int \lambda_{j-1}^{(l)}(s) e^{-is(x-X_j)} ds, \qquad (11)$$

where $\Lambda_{j-1}^{(l)}(z) = \frac{1}{\hat{h}_{j-1}^{1+l}} K^{(l)}\left(\frac{z}{\hat{h}_{j-1}}\right) = \frac{1}{2\pi}\int \lambda_{j-1}^{(l)}(s)e^{-isz}ds$ is the smoothing kernel, and $\lambda_{j-1}^{(0)}(s) = \lambda_c(s, \hat{h}_{j-1})$, $l = 0; 1$. The required bandwidths are defined by

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

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 (9) and (10) respectively.

Below $C(\gamma, l)$ are some constants and $\Psi_{\gamma,l}(n)$ are concrete decreasing to zero functions. 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 C_+ and in the definition of the estimator r_n to the set C. Let $\gamma = 0$ if $f \in \mathcal{F}_r^+$ and $\gamma \in (0, r)$ if $f \in \mathcal{F}_r$, as well as $r > 0$ if $l = 0$ and $r > 1$ if $l = 1$. Then for every $n \geq 1$ the estimators (11) has the following properties:*

$$\sup_{f \in \overline{\mathcal{F}}_r} u_f^2(\hat{f}_n^{(l)}) \leq \Psi_{\gamma,l}(n) + \frac{C(\gamma, l)}{n}, \qquad l = 0; 1.$$

Examples 1 and 2 revisited.

Under appropriate chosen $\delta > 0$ and sequences (A_n, B_n, δ_n) in the definition of sets C_+, C :

In Example 1 (case $(f \in \mathcal{F}_r^+)$)

$$\Psi_{0,0}(n) \approx (nh_n)^{-1} \cdot (\ln n)^{\frac{2\delta}{(1+2r)^2}} \approx n^{-\frac{2r}{1+2r}} \cdot (\ln n)^{\frac{2\delta}{(1+2r)^2}}.$$

Then, according to Theorem 3, 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 [10] on the extra log-factor only.

For the functions $f \in \mathcal{F}_r$ and $\gamma \in (0, \min(r, 1))$ from Example 2 it follows that

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

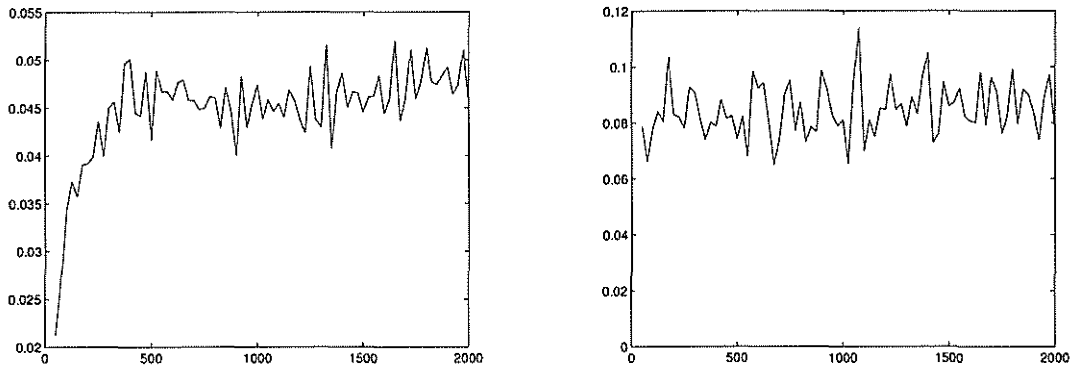

Figure 1: MSE of kernel estimators multiplied by $n^{3/4}$ versus $n \in \{25, 2000\}$. Left chart corresponds to the estimator with piece-wise linear kernel characteristic function. Right chart corresponds to the estimator with infinitely-differentiable flat-top kernel characteristic function.

5 Simulation results

In this section we provide brief results of simulation study of the estimators introduced in Section 2. We examine kernel estimators of triangular probability density function $f(x) = (a - |x|)/a^2, |x| \le a$ belonging to the family $\mathcal{F}_1$ with characteristic function $\phi(s) = 2(1 - \cos(as))/(as)^2$. Also $\phi(s)$ meets requirements of the Example 2. Thus the bandwidth can be taken in the form $h = O(n^{-1/4})$ and expected convergence rate of the kernel estimator MSE is $n^{-3/4}$.

Two flat-top kernels have been used in the simulation. First one has the piece-wise linear kernel characteristic function introduced in [10]: $\lambda(s) = \{1, |s| \le 1; (c-|s|)/(c-1), 1 < |s| < c; 0, |s| \ge c\}$. Second case refers to the infinitely-differentiable flat-top kernel characteristic function (see [6]) $\lambda(s) = \{1, |s| \le c; exp\left[\frac{-b\exp[-b/(|s|-c)^2]}{(|s|-1)^2}\right],$ $c < |s| < 1; 0, |s| \ge 1\}$.

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 sequence of 150 samples for each sample size from 25 to 2000 with step 25, then calculate the estimator MSE multiplied by $n^{3/4}$ and expect visual stabilization of the sequence of resulting values with growth of n.

Two typical examples are presented at the Figure 1. Both cases refer to estimation of triangle density function $f(x)$ with unit variation (which support is bounded by ± 2.45, $a = 2.45$) at the point $x = 1.0$ by kernel estimators with flat-top kernels. The expected stabilization is observing in both cases.

References

[1] Brown L.D., Low M.G. (1992) Superefficiency and lack of adaptibility in functional estimation. Technical report, Cornell Univ.

[2] Dobrovidov, A.V., Koshkin, G.M., Vasiliev, V.A. (2012) Non-parametric state space models. Heber City, Utah: Kendrick Press.

[3] Jones M.C., Marron J.S., Sheather S.J. (1996) A brief survey of bandwidth selection for density estimation, J. Amer. Statist. Assoc., vol. 91, 401-407.

[4] Lepski O.V. (1990) One problem of adaptive estimation in Gaussian white noise. Theor. Probab. Appl., 35, pp. 459–470 (in Russian).

[5] Lepski O.V., Spokoiny V.G. (1997) Optimal pointwise adaptive methods in nonparametric estimation. Ann. Statist., 25 (6), pp. 2512–2546.

[6] McMurry T., Politis D.N. (2004) Nonparametric regression with infinite order flat-top kernels, J. Nonparam. Statist., vol. 16, no. 3-4, 549–562, 2004.

[7] Politis D.N. (2001) On nonparametric function estimation with infinite-order flat-top kernels, in Probability and Statistical Models with applications, Ch. Charalambides et al. (Eds.), Chapman and Hall/CRC: Boca Raton, pp. 469-483.

[8] Politis D.N. (2003) Adaptive bandwidth choice, J. Nonparam. Statist., vol. 15, no. 4-5, 517-533, 2003.

[9] Politis D.N., Romano J.P. (1993) On a Family of Smoothing Kernels of Infinite Order, in Computing Science and Statistics, Proceedings of the 25th Symposium on the Interface, San Diego, California, April 14-17, 1993, (M. Tarter and M. Lock, eds.), The Interface Foundation of North America, pp. 141-145.

[10] Politis D.N., Romano J.P. (1999) Multivariate density estimation with general flat-top kernels of infinite order. Journal of Multivariate Analysis, 1999, 68, 1-25.

[11] Politis D.N., Romano J.P., Wolf M. (1999) Subsampling. Springer, New York, 1999.

[12] Politis D.N., Vasiliev V.A., Tarassenko P.F. (2015) Estimating smoothness and optimal bandwidth for probability density functions. Metrika, Springer (submitted).

[13] Rosenblatt M. (1991) Stochastic curve estimation. NSF-CBMS Regional Conference Series, 1991, 3, Institute of Mathematical Statistics, Hayward.