T E C H N I C A L
R E P O R T

0506

LINEAR PROGRAMMING PROBLEMS FOR L_1- OPTIMAL FRONTIER ESTIMATION

S. GIRARD, A. IOUDITSKI and A. NAZIN

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Abstract

We propose new *optimal* estimators for the Lipschitz frontier of a set of points. They are defined as kernel estimators being sufficiently regular, covering all the points and whose associated support is of smallest surface. The estimators are written as linear combinations of kernel functions applied to the points of the sample. The coefficients of the linear combination are then computed by solving related linear programming problem. The L_1 error between the estimated and the true frontier function with a known Lipschitz constant is shown to be almost surely converging to zero, and the rate of convergence is proved to be optimal.

Contact information

Stéphane Girard and Anatoli Iouditski: SMS/LMC, Université Grenoble I, BP 53, 38041 Grenoble cedex 9, France. {`Stephane.Girard, Anatoli.Iouditski`}`@imag.fr`

Alexander Nazin: Institute of Control Sciences RAS, Profsoyuznaya str., 65, 117997 Moscow, Russia. `nazine@ipu.rssi.ru`

Acknowledgements

Financial support from the IAP research network nr P5/24 of the Belgian Government (Federal Office for Scientific, Technical and Cultural Affairs) is gratefully acknowledged.

The work of A. Nazin has been carried out during his stay in SMS/LMC, Université Grenoble-I, and INRIA Rhône-Alpes hold in June and October–November 2004.

1 Introduction

Many proposals are given in the literature for estimating a set S given a finite random set of points drawn from the interior. This problem of edge or support estimation arises in classification (Hardy & Rasson [20]), clustering problems (Hartigan [21]), discriminant analysis (Baufays & Rasson [2]), and outliers detection (Devroye & Wise [8]). Applications are found in medical diagnosis (Tarassenko *et al* [26]) as well as in condition monitoring of machines (Devroye & Wise [8]). In image analysis, the segmentation problem can be considered under the support estimation point of view, where the support is a convex bounded set in $\mathbb{R}^2$ (Korostelev & Tsybakov [24]). We also point out some applications in econometrics (e.g. Deprins, *et al* [7]). In such cases, the unknown support can be written

$$S \triangleq \{(x, y) : \quad 0 \le \ x \le \ 1 \ \ ; \quad 0 \le \ y \ \le \ f(x)\}, \tag{1}$$

where f is an unknown function. Here, the problem reduces to estimating f, called the production frontier (see for instance Härdle *et al* [18]). The data consist of pair (X, Y) where X represents the input (labor, energy or capital) used to produce an output Y in a given firm. In such a framework, the value $f(x)$ can be interpreted as the maximum level of output which is attainable for the level of input x.

An early paper was written by Geffroy [10] for independent identically distributed observations from a density ϕ. The proposed estimator is a kind of histogram based on the extreme values of the sample. This work was extended in two main directions.

On the one hand, piecewise polynomials estimators were introduced. They are defined locally on a given slice as the lowest polynomial of fixed degree covering all the points in the considered slice. Their optimality in an asymptotic minimax sense is proved under weak assumptions on the rate of decrease α of the density ϕ towards 0 by Korostelev & Tsybakov [24] and by Härdle *et al* [19]. Extreme values methods are then proposed by Hall *et al* [16] and by Gijbels & Peng [11] to estimate the parameter α. Estimating f can also been considered as a regression problem $Y = f(X) + \varepsilon$ with negative noise ε. In this context, local polynomial estimates are introduced, see Knight [23], or Hall *et al* [17] for a similar approach.

On the other hand, different propositions for smoothing Geffroy's estimator were made in the case of a Poisson point process. Girard & Jacob [14] introduced estimators based on kernel regressions and orthogonal series method [12, 13]. In the same spirit, Gardes [9] proposed a Faber-Shauder estimator. Girard & Menneteau [15] introduced a general framework for studying estimators of this type and generalized them to supports writing

$$S = \{(x, y) : \quad x \in E \ \ ; \quad 0 \le \ y \ \le \ f(x)\},$$

where f is an unknown function and E an arbitrary set. In each case, the limit distribution of the estimator is established. We also refer to Abbar [1] and Jacob & Suquet [22] who used a similar smoothing approach, although their estimators are not based on the extreme values of the Poisson process.

The estimator proposed in Bouchard *et al* [6] can be considered to belong to the intersect of these two directions. From the practical point of view, it is defined as a kernel estimator obtained by smoothing some selected points of the sample. These points are chosen automatically by solving a linear programming problem to obtain an estimated support covering all the points and with smallest surface. From the theoretical point of view, this estimator is shown to be consistent for the L_1 norm.

In this paper, we propose several modifications of the above method. First, a bias corrected kernel is proposed. Second, some regularity constraints are introduced in the optimization problem. We show that the resulting estimator reaches the optimal minimax L_1 rate (up to a logarithmic factor). The estimator is defined in Section 2. Some preliminary properties are established in Section 3, and the main result is presented in Section 4. Proofs are postponed to Section 5.

2 Boundary estimator

Let all the random variables be defined on a probability space $(\Omega, \mathcal{F}, P)$. The problem under consideration is to estimate an unknown positive function $f : [0,1] \to (0,\infty)$ on the basis of observations $(X_i, Y_i)_{i=1,\dots,N}$ with independent pairs (X_i, Y_i) being uniformly distributed in the set S defined as

$$S \triangleq \{(x,y) : 0 \le x \le 1 \,,\; 0 \le y \le f(x)\} \,. \tag{2}$$

Letting

$$C_f \triangleq \int_0^1 f(u)\, du \,, \tag{3}$$

each variable X_i is distributed in $[0,1]$ with p.d.f. $f(\cdot)/C_f$ while Y_i has the uniform conditional distribution with respect to X_i in the interval $[0, f(X_i)]$. In what follows we assume $f \in \Sigma_{[0,1]}(\beta, L_{f,\beta})$, $0 < \beta \le 1$ that is function $f : [0,1] \to (0,\infty)$ is β-Lipschitz with constant $L_{f,\beta}$:

$$|f(x) - f(u)| \le L_{f,\beta}\, |x-u|^{\beta} \quad \forall x, u \in [0,1] \,. \tag{4}$$

The considered estimator $\widehat{f}_N : [0,1] \to [0,\infty)$ of the frontier is chosen from the family of functions:

$$\begin{cases} \widehat{f}_N(x) = \displaystyle\sum_{i=1}^{N} \alpha_i\, K_h(x, X_i)\,, & K_h(x,t) = \dfrac{g(x)}{h}\, K\left(\dfrac{x-t}{h}\right), \\ \alpha_i \ge 0, \quad i = 1, \dots, N, & \end{cases} \tag{5}$$

where K is a sufficiently smooth basic kernel function $K : \mathbb{R} \to [0,\infty)$ integrating to one and having the interval $[-1,1]$ as its support; the bandwidth parameter $h \in (0, 1/2)$ depends on N such that $h \to 0$ as $N \to \infty$; and the function

$$g(x) = \left(\int_{(x-1)/h}^{x/h} K(t)\, dt \right)^{-1} , \quad x \in [0,1]\,, \tag{6}$$

corrects the basic kernel K at the boundaries, i.e., when $x \in [0,h)$ or $x \in (1-h, 1]$. Indeed, $g(x) \equiv 1$ on $x \in [h, 1-h]$, while $g(x) > 1$ when $x \in [0,h)$ or $x \in (1-h,1]$.
One may easily observe that

$$\int_0^1 K_h(x,u)\, du = 1 \qquad \forall\, x \in [0,1] \tag{7}$$

and, consequently, due to interplacing the integral and the derivative,

$$\int_0^1 \frac{\partial}{\partial x} K_h(x,u)\, du = 0 \qquad \forall\, x \in [0,1]\,. \tag{8}$$

Note, that equation (8) may be verified directly, as it is demonstrated in the Appendix, Subsection 6.1.

Denote $K_{\max} \triangleq \max K(t)$, $g_{\max} \triangleq \max g(x)$, as well as functionals

$$\begin{aligned} C_\beta(\varphi) &\triangleq \int_{-1}^{1} |t|^\beta |\varphi(t)|\, dt\,, \quad \varphi \in C^0([-1,1])\,, && (9)\\ C_\beta(K,K') &\triangleq g_{\max}\, K_{\max} C_\beta(K) + C_\beta(K')\,. && (10) \end{aligned}$$

We also denote by L_φ a Lipschitz constant for function $\varphi : \mathbb{R} \to \mathbb{R}$, that is

$$|\varphi(s) - \varphi(t)| \le L_\varphi |s-t| \text{ with } L_\varphi < \infty. \tag{11}$$

The indicator function is denoted by $\mathbf{1}\{\cdot\}$ which equals 1 if the argument condition holds true, and 0 otherwise.

As it is proved below in Lemma 1 the surface of the estimated support

$$\widehat{S}_N \triangleq \{(x,y) :\, 0 \le x \le 1\,,\; 0 \le y \le \widehat{f}_N(x)\} \tag{12}$$

may be approximated as follows:

$$\int_0^1 \widehat{f}_N(x)\, dx = \sum_{i=1}^{N} \alpha_i + O(h)\,. \tag{13}$$

This suggests to define the parameter vector $\alpha = (\alpha_1, \ldots, \alpha_N)^T$ as a solution to the following optimization problem

$$J_P^* \;\triangleq\; \min_\alpha \sum_{i=1}^{N} \alpha_i \tag{14}$$

subject to

$$\widehat{f}_N(X_i) \ge Y_i\,, \quad i = 1, \ldots, N\,, \tag{15}$$

$$|\widehat{f}_N'(X_i)| \le L_{f,\beta}\; g_{\max} C_\beta(K,K') \frac{\log N}{Nh^2}\,, \quad i = 0, \ldots, N+1\,, \tag{16}$$

$$\sum_{i=1}^{N} \alpha_i\, \mathbf{1}\{(j-1)/m_h \le X_i < j/m_h\} \le C_\alpha h\,, \quad j = 1, \ldots, m_h\,, \tag{17}$$

$$0 \le \alpha_i\,, \quad i = 1, \ldots, N, \tag{18}$$

where parameter m_h is defined to be the integer part of $1/h$. This optimization problem may be formally written as linear program (LP)

$$J_P^* \;\triangleq\; \min_\alpha\, \mathbf{1}_N^T \alpha \tag{19}$$

subject to

$$\begin{aligned} Y &\le A\alpha\,, && (20)\\ -L_{f,\beta}\; g_{\max} C_\beta(K,K') \frac{\log N}{Nh^2} \mathbf{1}_N &\le B\alpha \;\le\; L_{f,\beta}\; g_{\max} C_\beta(K,K') \frac{\log N}{Nh^2} \mathbf{1}_N\,, && (21)\\ & \quad D^T\alpha \;\le\; C_\alpha h\, \mathbf{1}_{m_h}\,, && (22)\\ 0 &\le \alpha\,. && (23) \end{aligned}$$

There is one positive parameter C_α in the constraints (17) and (22): its value will be discussed in Section 4. Moreover, the following notations have been introduced:

$$\begin{aligned} X_0 &\triangleq 0, \quad X_{N+1} \triangleq 1, && (24)\\ \mathbf{1}_N &\triangleq (1,1,\ldots,1)^T \in \mathbb{R}^N && (25)\\ A &\triangleq \|K_h(X_i\,,X_j)\|_{i,j=1,\ldots,N} && (26)\\ B &\triangleq \left\| \left. \frac{d}{dx} K_h(x,X_j)\right|_{x=X_i} \right\|_{i,j=1,\ldots,N} && (27)\\ D &\triangleq \|\mathbf{1}\{(j-1)/m_h \le X_i < j/m_h\}\|_{i=1,\ldots,N;\, j=1,\ldots,m_h} && (28)\\ Y &\triangleq (Y_1,\ldots,Y_N)^T. && (29) \end{aligned}$$

3 Preliminary results

The basic assumptions on the unknown boundary function are:

A1. $0 < f_{\min} \le f(x) \le f_{\max} < \infty$, for all $x \in [0,1]$,

A2. $|f(x) - f(y)| \le L_{f,\beta}\,|x-y|^\beta$, for all $x,y \in [0,1]$, with $L_{f,\beta} < \infty$ and $0 < \beta \le 1$.

The following assumptions on the kernel function are introduced:

B1. $K : \mathbb{R} \to [0,\infty)$ has a compact support: $\operatorname{supp}_{t\in\mathbb{R}} K(t) = [-1,1]$,

B2. $\displaystyle\int_{-1}^{1} K(t)\,dt = 1$,

B3. K is three times continuously differentiable.

Note, that $g_{\max} = 2$ for any unimodal even kernel $K(\cdot)$ meeting conditions B1–B2. We quote two preliminary results on the estimator $\widehat{f}_N$. First, the surface of the related estimated support $\widehat{S}_N$ is approximatively $\sum_{i=1}^N \alpha_i$. Second, the function $\widehat{f}_N$ is Lispchitzian. Proofs are postponed to Subsection 5.1.

Lemma 1 *Suppose B1, B2 are verified and* $0 < h < 1/4$*. Moreover, let conditions (17) and (18) hold true for* $m_h = \lfloor h^{-1} \rfloor$*. Then the surface of the estimated support (12) meets the following inequality:*

$$-2C_\alpha K_{\max} h \le \int_0^1 \widehat{f}_N(x)\,dx - \sum_{i=1}^N \alpha_i \le 4C_\alpha (g_{\max}-1) K_{\max} h\,. \tag{30}$$

Remark 1 *In fact, only one part of Lemma 1 is used in what follows, that is the upper bound for the estimator surface given by the right hand side (30).*

Remark 2 *Lemma 1 as well as the further results may be easily extended to basic kernels* $K(\cdot)$ *having also negative values: then* $K_{\max} \triangleq \max|K(t)|$*, and* $g(x) > 0\ \forall\, x \in [0,1]$ *should be additionally assumed.*

Lemma 2 *Suppose A1 and B1–B3 are verified. Let estimator $\widehat{f}_N$ be defined by LP (19)–(23). Moreover, let $h \to 0$ as $N \to \infty$ such that*

$$\lim_{N\to\infty} \frac{\log N}{Nh} = 0\,. \tag{31}$$

Then, there exists almost surely finite $N_4 = N_4(\omega)$ such that for any $N \geq N_4$ the Lipschitz constant for the estimator $\widehat{f}_N$ over the interval $[0,1]$ is bounded as follows:

$$L_{\widehat{f}_N} \triangleq \max_{x\in[0,1]} |\widehat{f}_N'(x)| \tag{32}$$

$$\leq 2L_{f,\beta}\, g_{\max} C_\beta(K,K') \frac{\log N}{Nh^2}. \tag{33}$$

Remark 3 *As it can be seen from the proof of Lemma 2, namely from (56)–(57), one might slightly decrease the number of constraints (16) on the estimator derivative (14)–(18). In fact, one could impose those type of constraints not at each point X_i, $i = 1, \dots, N$: It would be enough to do at the points with the distance $O\left((h \log N/N)^{1/2}\right)$ between them, or at least $o\left((h \log N/N)^{1/2}\right)$ in order to keep the same Lipschitz constant for $\widehat{f}_N$ as is given by Lemma 2.*

It appears that the estimator $\widehat{f}_N$ being the solution to the optimization problem (14)–(18) or to its equivalent LP version (19)–(23) defines the kernel estimator of the support covering all the points (X_i, Y_i) and, approximately, having the smallest surface, up to the term $O(h)$ specified in Lemma 1. Moreover, constraints (16)–(17) or (21)–(22) ensure $\widehat{f}_N \in \Sigma_{[0,1]}(1, L_{\widehat{f}_N})$ with a particular Lipschitz constant $L_{\widehat{f}_N}$ given in Lemma 2. The constraint $\alpha_i \geq 0$ for all $i = 1, \dots, N$ ensures that $\widehat{f}_N(x) \geq 0$ for all $x \in [0,1]$ since the basic kernel K is chosen to be non-negative; this seems to be natural for function $f(\cdot)$ is positive. Finally, note that the above described estimator (5), (19)–(23) may be treated as the approximation to Maximum Likelihood Estimate related to the estimation family (5); see Bouchard *et al* [6, 5] for the demonstration.

4 Main results

In the following theorem, the consistency and the convergence rate of the estimator towards the true frontier is established with respect to the L_1 norm on the $[0,1]$ interval.

Theorem 1 *Let the above mentioned assumptions A1, A2 and B1–B3 hold true and the estimator parameter $C_\alpha > 6f_{\max}$. Moreover, let $h \to 0$ as $N \to \infty$ such that*

$$\liminf_{N\to\infty} \frac{\log N}{Nh^{1+\beta}} > \rho > 0\,, \quad \lim_{N\to\infty} \frac{\log N}{Nh^{1+\beta/2}} = 0\,. \tag{34}$$

Then estimator (5)–(23) has the following asymptotic properties:

$$\|\widehat{f}_N - f\|_1 \leq \left(C_{12}(\beta)h^\beta + 2C_4(\beta)h^{-2}(\log N/N)^{\frac{2+\beta}{1+\beta}}\right)(1+o(1)) \quad \text{a.s.} \tag{35}$$

with

$$C_{12}(\beta) \triangleq 2L_{f,\beta}\, g_{\max}\, C_\beta(K,K') + 4C_\alpha(g_{\max}-1)K_{\max}\mathbf{1}\{\beta=1\} \tag{36}$$

and

$$C_4(\beta) \triangleq 2L_{f,\beta}\left[\left(\frac{2C_f}{L_{f,\beta}}\right)^{\frac{\beta}{1+\beta}}\left(\frac{1}{\rho}\right)^{\frac{2}{1+\beta}} + g_{\max}C_\beta(K,K')\left(\frac{2C_f}{L_{f,\beta}}\right)^{\frac{1}{1+\beta}}\right]. \tag{37}$$

Corollary 1 *The maximum rate of convergence which is guaranteed by Theorem 1*

$$\|\widehat{f}_N - f\|_1 = O\left((\log N/N)^{\frac{\beta}{1+\beta}}\right) \quad \text{a.s.}$$

is attained for h meeting the following asymptotics:

$$h \sim \widetilde{\rho}\left(\frac{\log N}{N}\right)^{\frac{1}{1+\beta}}, \quad 0 < \widetilde{\rho} < \rho^{-\frac{1}{1+\beta}}\,, \tag{38}$$

which reduces the upper bound (35) to

$$\limsup_{N\to\infty}\left(\frac{\log N}{N}\right)^{-\frac{\beta}{1+\beta}} \|\widehat{f}_N - f\|_1 \le C_{12}(\beta)\widetilde{\rho}^{\beta} + 2C_4(\beta)\widetilde{\rho}^{-2} \quad \text{a.s.} \tag{39}$$

Let us highlight that (39) shows that $\widehat{f}_N$ reaches (up to a logarithmic factor) the minimax L_1 rate for Lipschitz frontier f, see Korostelev & Tsybakov [24], Theorem 4.1.1.

Remark 4 *The second condition in (34) may be extended to*

$$\lim_{N\to\infty}\frac{\log N}{Nh^{1+\beta/2}} < \infty \tag{40}$$

which leads to another, more general formula for constants in (35)–(37).

5 Proofs

The proof of Theorem 1 which is given in Subsection 5.4 is based on both upper and lower bounds derived in Subsection 5.2 and Subsection 5.3 respectively. When proving these bounds, we assume that the sequence of the sample X-points $(X_i)_{i=1,\dots,N}$ is already increase ordered, without changing notation from X_i to $X_{(i)}$ for the sake of simplicity, that is

$$X_i \le X_{i+1}\,, \quad \forall\, i\,. \tag{41}$$

We essentially apply the uniform asymptotic bound $O(\log N/N)$ on $\Delta X_i \triangleq X_i - X_{i-1}$ proved in auxiliary Lemma 6. Before that, we prove in Subsection 5.1 the two preliminary results.

5.1 Proof of preliminary results

Proof of Lemma 1. Note that definitions (5)–(6) imply the following decomposition.

$$\int_0^1 \widehat{f}_N(x)\,dx = \left(\int_0^h + \int_h^{1-h} + \int_{1-h}^1\right)\widehat{f}_N(x)\,dx \tag{42}$$

$$= \int_0^1 \sum_{i=1}^N \alpha_i \frac{1}{h} K\left(\frac{x-X_i}{h}\right) dx \tag{43}$$

$$+ \sum_{i=1}^N \alpha_i \left(\int_0^h + \int_{1-h}^1\right) \frac{g(x)-1}{h} K\left(\frac{x-X_i}{h}\right) dx\,. \tag{44}$$

Since α_i and kernel K are non-negative, it follows that

$$\int_0^1 \sum_{i=1}^N \alpha_i \frac{1}{h} K\left(\frac{x-X_i}{h}\right) dx \leq \sum_{i=1}^N \alpha_i \int_{\mathbb{R}} \frac{1}{h} K\left(\frac{x-X_i}{h}\right) dx = \sum_{i=1}^N \alpha_i \tag{45}$$

and therefore,

$$\begin{aligned}
\int_0^1 \widehat{f}_N(x)\, dx - \sum_{i=1}^N \alpha_i \;&\leq\; \frac{g_{\max}-1}{h} K_{\max} \sum_{i=1}^N \alpha_i \left(\int_0^h + \int_{1-h}^1\right) \mathbf{1}\{|x-X_i| \leq h\} dx & (46)\\
&\leq\; (g_{\max}-1) K_{\max} \left(\sum_{i=1}^N \alpha_i\, \mathbf{1}\{0 \leq X_i \leq 2h\}\right. & (47)\\
&\qquad\qquad \left. + \sum_{i=1}^N \alpha_i\, \mathbf{1}\{1-2h \leq X_i \leq 1\}\right) & (48)\\
&\leq\; (g_{\max}-1) K_{\max} 4 C_\alpha h\,. & (49)
\end{aligned}$$

The inequality (49) follows from (17) since both intervals in (47)–(48) are of the length $2h$ and thus may be covered by two related intervals of the form $[(j-1)/m_h, j/m_h]$ in (17). Consequently, we have proved the upper bound for the difference in the left hand side (46). The lower bound is proved in the same manner. Indeed, decomposition (42)–(44) implies, since the term (44) is non-negative,

$$\begin{aligned}
\int_0^1 \widehat{f}_N(x)\, dx - \sum_{i=1}^N \alpha_i \;&\geq\; -\sum_{i=1}^N \alpha_i \left(\int_{-h}^0 + \int_1^{1+h}\right) \frac{1}{h} K\left(\frac{x-X_i}{h}\right) dx & (50)\\
&\geq\; -K_{\max} \left(\sum_{i=1}^N \alpha_i\, \mathbf{1}\{0 \leq X_i \leq h\}\right. & (51)\\
&\qquad\qquad \left. + \sum_{i=1}^N \alpha_i\, \mathbf{1}\{1-h \leq X_i \leq 1\}\right) & (52)\\
&\geq\; -K_{\max}\, 2 C_\alpha h\,. & (53)
\end{aligned}$$

This completes the proof of Lemma 1. ■

Proof of Lemma 2. Remind that we assume (41). By applying auxiliary Lemma 6 and Lemma 8 we first arrive at

$$\begin{aligned}
& \max_{x\in[0,1]} |\widehat{f}_N'(x)| & (54)\\
=\; & \max_{1\leq i\leq N+1} \max_{x\in[X_{i-1},X_i]} |\widehat{f}_N'(x)| & (55)\\
\leq\; & L_{f,\beta}\, g_{\max} C_\beta(K,K') \frac{\log N}{Nh^2} + \frac{1}{8} \max_{1\leq i\leq N+1} \left[(X_i - X_{i-1})^2 \max_{x\in[X_{i-1},X_i]} |\widehat{f}_N'''(x)|\right] & (56)\\
\leq\; & L_{f,\beta}\, g_{\max} C_\beta(K,K') \frac{\log N}{Nh^2} + \frac{1}{8}\left(C_X \frac{\log N}{N}\right)^2 \max_{x\in[0,1]} |\widehat{f}_N'''(x)|\,, & (57)
\end{aligned}$$

with $C_X > 4C_f/f_{\min}$. The maximum term in (57) is bounded as follows: for any $x \in [0,1]$

$$
\begin{aligned}
|\widehat{f}_N'''(x)| &\le \sum_{i=1}^{N} \alpha_i \left| \frac{d^3}{dx^3} K_h(x, X_i) \right| &(58)\\
&\le \sup_{u,v} \left| \frac{\partial^3}{\partial v^3} K_h(v,u) \right| \cdot \sum_{i=1}^{N} \alpha_i \, \mathbf{1}\{|x - X_i| \le h\} &(59)\\
&\le g_{\max} L_{\widetilde{K}''} h^{-4} \cdot 3C_\alpha h \,, &(60)
\end{aligned}
$$

since (see Lemma 5 for the detailed demonstration)

$$
\sup_{u,v} \left| \frac{\partial^3}{\partial v^3} K_h(v,u) \right| \le g_{\max} L_{\widetilde{K}''} h^{-4} \tag{61}
$$

where

$$
L_{\widetilde{K}''} \triangleq L_{K''} + 3L_{K'} K_{\max} g_{\max} + L_K g_{\max} \left(3L_K + 10K_{\max}^2 g_{\max}\right) + 6K_{\max}^4 g_{\max}^3 \,. \tag{62}
$$

Substituting (58), (60) into (57) yields

$$
\begin{aligned}
\max_{x\in[0,1]} |\widehat{f}_N'(x)| &\le L_{f,\beta}\, g_{\max} C_\beta(K,K') \frac{\log N}{Nh^2} + \frac{3}{8}\, g_{\max} L_{\widetilde{K}''} C_\alpha \left(C_X \frac{\log N}{Nh^2} \right)^2 h &(63)\\
&\le 2L_{f,\beta}\, g_{\max} C_\beta(K,K') \frac{\log N}{Nh^2} &(64)
\end{aligned}
$$

under additional assumption (which hold true for all sufficiently large N):

$$
h \ge \frac{3C_X^2 C_\alpha\, L_{\widetilde{K}''} \, \log N}{8L_{f,\beta}\, C_\beta(K,K')\, N} \,. \tag{65}
$$

The result follows. ∎

5.2 Upper bound for $\widehat{f}_N$ in terms of J_P^*

Lemma 3 *Let the assumptions of Theorem 1 hold true. Then for any finite*

$$
\gamma > L_{f,\beta}\, g_{\max} C_\beta(K) \tag{66}
$$

and almost all $\omega \in \Omega$ there exist finite numbers $N_1 = N_1(\omega,\gamma)$ such that for all $N \ge N_1$ the LP (19)–(23) is solvable and

$$
J_P^* \le C_f + \gamma h^\beta \,. \tag{67}
$$

Proof of Lemma 3. Consider arbitrary $N \ge N_0(\omega)$ with $N_0(\omega)$ from Lemma 6. Introduce function $f_\gamma(u) = f(u) + \gamma h^\beta$ and pseudo-estimators

$$
\widetilde{\alpha}_i = \frac{1+\delta_{i1}}{2} \int_{X_{i-1}}^{X_i} f_\gamma(u)\, du + \frac{1+\delta_{iN}}{2} \int_{X_i}^{X_{i+1}} f_\gamma(u)\, du \,, \quad i = 1, \ldots, N \tag{68}
$$

where δ_{ij} stands for Kronecker symbol. Below we demonstrate that condition (66) ensures the vector of pseudo-estimators $\widetilde{\alpha} = (\widetilde{\alpha}_1 \ldots, \widetilde{\alpha}_N)^T$ to be an admissible point for the LP (19)–(23), for any sufficiently large N. This implies solvability of the LP (19)– (23) and

$$J_P^* \leq \sum_{i=1}^N \widetilde{\alpha}_i = \int_0^1 (f(u) + \gamma h^\beta)\, du = C_f + \gamma h^\beta \,. \tag{69}$$

Let $C_X > 4C_f/f_{\min}$. For the sake of simplicity, we impose the additional assumptions

$$h^\beta \leq \frac{\log N}{\rho N h} \leq \min\left\{\frac{f_{\max}}{\gamma}, \frac{1}{\rho C_X}\right\}, \tag{70}$$

which hold true for all N large enough.

1. First, we prove constraints (15) under $\alpha_i = \widetilde{\alpha}_i$, $i = 1, \ldots, N$. For arbitrary $x \in [0,1]$,

$$\begin{aligned}
\widetilde{f}_N(x) \;&\triangleq\; \sum_{i=1}^N \widetilde{\alpha}_i K_h(x, X_i) && (71)\\
&=\; \sum_{i=1}^{N+1} \int_{X_{i-1}}^{X_i} f_\gamma(u)\, du \, \frac{K_h(x,X_i) + K_h(x, X_{i-1})}{2} && (72)\\
&\quad + \frac{1}{2}\int_0^{X_1} f_\gamma(u)\, du\, (K_h(x, X_1) - K_h(x,0)) && (73)\\
&\quad + \frac{1}{2}\int_{X_N}^1 f_\gamma(u)\, du\, (K_h(x, X_N) - K_h(x,1)) && (74)\\
&=\; \int_0^1 f_\gamma(u)\, K_h(x,u)\, du && (75)\\
&\quad + \sum_{i=1}^{N+1} \int_{X_{i-1}}^{X_i} f_\gamma(u) \left(\frac{K_h(x,X_i) + K_h(x,X_{i-1})}{2} - K_h(x,u)\right) du && (76)\\
&\quad + \frac{1}{2}\int_0^{X_1} f_\gamma(u)\, du\, (K_h(x, X_1) - K_h(x,0)) && (77)\\
&\quad + \frac{1}{2}\int_{X_N}^1 f_\gamma(u)\, du\, (K_h(x, X_N) - K_h(x,1)) \,. && (78)
\end{aligned}$$

Now we separately bound each of the summands (75)–(78) from below. Due to (7), the main term (75) is bounded as follows:

$$\begin{aligned}
\int_0^1 f_\gamma(u)\, K_h(x,u)\, du \;&=\; f(x) + \gamma h^\beta + \int_0^1 (f(u) - f(x))\, K_h(x,u)\, du && (79)\\
&\geq\; f(x) + (\gamma - L_{f,\beta}\, g_{\max} C_\beta(K)) h^\beta \,. && (80)
\end{aligned}$$

The i-th summand from (76) is decomposed and then bounded basing on trapezium formula error as follows:

$$\int_{X_{i-1}}^{X_i} f_\gamma(u)\left(\frac{K_h(x,X_i)+K_h(x,X_{i-1})}{2}-K_h(x,u)\right)du \tag{81}$$

$$\geq\ f_\gamma(x)\int_{X_{i-1}}^{X_i}\left(\frac{K_h(x,X_i)+K_h(x,X_{i-1})}{2}-K_h(x,u)\right)du \tag{82}$$

$$-\int\limits_{X_{i-1}}^{X_i}|f_\gamma(u)-f_\gamma(x)|\left|\frac{K_h(x,X_i)+K_h(x,X_{i-1})}{2}-K_h(x,u)\right|du \tag{83}$$

$$\geq\ -(f_{\max}+\gamma h^\beta)\,\frac{(X_i-X_{i-1})^3}{12}\max_{u\in[0,1]}\left|\frac{\partial^2 K_h(x,u)}{\partial u^2}\right|\mathbf{1}\{|x-X_i|\leq 2h\} \tag{84}$$

$$-L_{f,\beta}\int_{X_{i-1}}^{X_i}|u-x|^\beta\,\mathbf{1}\{|x-X_i|\leq 2h\}\frac{g_{\max}L_K}{2h^2}\left[(u-X_{i-1})+(X_i-u)\right]du. \tag{85}$$

By applying Lemma 6, the first term is bounded as follows

$$(84)\geq-\left(C_X\frac{\log N}{N}\right)^2\frac{f_{\max}\,g_{\max}L_{K'}}{6h^3}\,(X_i-X_{i-1})\,\mathbf{1}\{|x-X_i|\leq 2h\}, \tag{86}$$

and the second one is bounded by:

$$(85)\geq-\frac{g_{\max}L_{f,\beta}\,L_K}{2h^2}C_X\frac{\log N}{N}\mathbf{1}\{|x-X_i|\leq 2h\}\int_{X_{i-1}}^{X_i}|u-x|^\beta du\,. \tag{87}$$

Moreover, from Lemma 6, one can show first that

$$\sum_{i=1}^{N+1}\mathbf{1}\{|x-X_i|\leq 2h\}(X_i-X_{i-1})\leq 4h+\frac{C_X\log N}{N},$$

and second that

$$\sum_{i=1}^{N+1}\mathbf{1}\{|x-X_i|\leq 2h\}\int_{X_{i-1}}^{X_i}|u-x|^\beta du \tag{88}$$

$$\leq\ \int_{x-2h-C_X(\log N)/N}^{x+2h}|u-x|^\beta du \tag{89}$$

$$\leq\ \left(4h+C_X\frac{\log N}{N}\right)\max_{v\in[-2h-C_X(\log N)/N,2h]}|v|^\beta \tag{90}$$

$$\leq\ \left(4h+C_X\frac{\log N}{N}\right)\left(2h+C_X\frac{\log N}{N}\right)^\beta. \tag{91}$$

Thus, we arrive at the bound for the sum (76) as follows:

$$\sum_{i=1}^{N+1}\int_{X_{i-1}}^{X_i}f_\gamma(u)\left(\frac{K_h(x,X_i)+K_h(x,X_{i-1})}{2}-K_h(x,u)\right)du \tag{92}$$

$$\geq\ -g_{\max}C_X\frac{\log N}{Nh}\left(4+C_X\frac{\log N}{Nh}\right)\left(\frac{C_Xf_{\max}L_{K'}}{6}\frac{\log N}{Nh}\right. \tag{93}$$

$$\left.+\frac{L_{f,\beta}\,L_K}{2}h^\beta\left(2+C_X\frac{\log N}{Nh}\right)^\beta\right) \tag{94}$$

$$\geq\ -\frac{5}{6}g_{\max}C_X\left(\frac{\log N}{Nh}\right)^2\left(C_Xf_{\max}L_{K'}+3^{\beta+1}\rho^{-1}L_{f,\beta}\,L_K\right). \tag{95}$$

At last, it is similarly demonstrated that both summands (77) and (78) are bounded above by $O((\log N/(Nh))^2)$. For instance, for (77), one obtains

$$\begin{aligned}\left|\int_0^{X_1} f_\gamma(u)\,du\,(K_h(x,X_1)-K_h(x,0))\right| &\le (f_{\max}+\gamma h^\beta)X_1\,|K_h(x,X_1)-K_h(x,0)| \\ &\le 2f_{\max}\,g_{\max}L_K\left(C_X\frac{\log N}{Nh}\right)^2. \qquad (96)\end{aligned}$$

Thus, from (71)–(96) it follows for each $j=1,\dots,N$ that

$$\widetilde{f}_N(X_j) \ge f(X_j) + (\gamma - L_{f,\beta}\,g_{\max}C_\beta(K))h^\beta + O\left(\left(\frac{\log N}{Nh}\right)^2\right) \ge Y_j \qquad (97)$$

for sufficiently large $N \ge N_0(\omega)$ when both inequalities (70) and the following one hold true:

$$\gamma - L_{f,\beta}\,g_{\max}C_\beta(K) \ge \frac{5}{6}\,g_{\max}C_X\left(\frac{\log N}{Nh^{1+\beta/2}}\right)^2\left(C_X f_{\max}\left(L_{K'}+\frac{12L_K}{5}\right)+3^{\beta+1}\frac{L_{f,\beta}\,L_K}{\rho}\right). \qquad (98)$$

2. Similarly, constraints (16) hold true under $\alpha_i=\widetilde{\alpha}_i$, $i=1,\dots,N$. Indeed, for arbitrary $x\in[0,1]$, we now have to bound the absolute value of

$$\widetilde{f}'_N(x) = \sum_{i=1}^N \widetilde{\alpha}_i\,\frac{d}{dx}\,K_h(x,X_i) = \sum_{i=1}^N \widetilde{\alpha}_i\,\widetilde{K}_h(x,X_i) \qquad (99)$$

instead of (71). Here

$$\widetilde{K}_h(x,u) \triangleq \frac{\partial}{\partial x}\,K_h(x,u) \qquad (100)$$

(see Subsection 6.1) with the following upper bound

$$\left|\widetilde{K}_h(x,u)\right| \le h^{-2}g_{\max}\left\{\left|K'\left(\frac{x-u}{h}\right)\right| + g_{\max}\,K_{\max}\left|K\left(\frac{x-u}{h}\right)\right|\right\}. \qquad (101)$$

deduced form (145), (146). Hence, one may repeat the arguments of (72)–(78) by changing K_h for $\widetilde{K}_h$. Therefore, all the rates from (81)–(97) should be divided by h, while the absolute value of the main term of decomposition, due to (8), is bounded as follows:

$$\begin{aligned}\left|\int_0^1 f_\gamma(u)\,\widetilde{K}_h(x,u)\,du\right| &= \left|\int_0^1 (f(u)-f(x))\,\frac{\partial}{\partial x}\,K_h(x,u)\,du\right| \qquad (102)\\ &\le L_{f,\beta}\,g_{\max}C_\beta(K,K')h^{\beta-1}, \qquad (103)\end{aligned}$$

instead of (79)–(80). Remind the definition (9)–(10) for $C_\beta(K,K')$ which follows from (101). Thus, for sufficiently large $N\ge N_0(\omega)$ and for each X_j we arrive at

$$\left|\widetilde{f}'_N(X_j)\right| \le L_{f,\beta}\,g_{\max}C_\beta(K,K')h^{\beta-1} + O\left(\frac{\log^2 N}{N^2h^3}\right) \le L_{f,\beta}\,g_{\max}C_\beta(K,K')\frac{\log N}{\rho Nh^2}. \qquad (104)$$

Namely, inequality (104) holds true almost surely for all those $N \ge N_0(\omega)$ such that (70) is verified and

$$L_{f,\beta}\, C_\beta(K,K')\left(\frac{\log N}{h^{\beta+1}N}-1\right) \;\ge\; \frac{5}{6}\, g_{\max} C_X \left(\frac{\log N}{N h^{1+\beta/2}}\right)^2 \tag{105}$$

$$\cdot\left(C_X f_{\max}\left(L_{\widetilde{K}'}+\frac{12 L_{\widetilde{K}}}{5}\right)+3^{\beta+1}\frac{L_{f,\beta}\, L_{\widetilde{K}}}{\rho}\right) \tag{106}$$

where (see Lemma 5 for the detailed demonstration)

$$L_{\widetilde{K}} \triangleq L_{K'} + L_K\, g_{\max} K_{\max}\,, \quad L_{\widetilde{K}'} \triangleq L_{K''} + L_{K'}\, g_{\max} K_{\max}\,. \tag{107}$$

3. Finally, the constraints (17) with

$$C_\alpha \ge 6 f_{\max} \tag{108}$$

also hold true under $\alpha_i = \widetilde{\alpha}_i$, $i = 1,\ldots,N$. Indeed, by Lemma 6 the following inequalities hold a.s. for all $N \ge N_0(\omega)$ and for each $j = 1,\ldots,m_h$, where $m_h = \lfloor h^{-1} \rfloor$:

$$\sum_{i=1}^{N} \widetilde{\alpha}_i\, \mathbf{1}\{(j-1)/m_h \le X_i < j/m_h\} \;\le\; (f_{\max}+\gamma h^\beta)\left(1/m_h + 2C_X\frac{\log N}{N}\right) \tag{109}$$

$$\le\; 6 f_{\max} h, \tag{110}$$

under additional assumptions (70). Thus, constraints (17) are fulfilled under (108) almost sure, for any sufficiently large N.

4. Since all $\widetilde{\alpha}_i \ge 0$, constraints (18) hold true, and Lemma 3 is proved. ■

Remark 5 *By applying Lemma 9, under additional assumptions (219) on h, one may ameliorate the related bounds in (92)–(98) and (105)–(106). Indeed, Lemma 9, being applied with its parameter $\nu \in (1,2)$, states that*

$$\sum_{i=1}^{N+1} \mathbf{1}\{|x - X_i| \le 2h\}(X_i - X_{i-1})^3 = o\left(\frac{h\log^\nu N}{N^2}\right) \tag{111}$$

hence, the sum of the term (84) is negligible with respect to that of (85). It means, roughly speaking, that we may remove the term $L_{K'}$ from (95), (98) as well as $L_{\widetilde{K}'}$ from (106). However, it does not change much in the main result of the Theorem. That is why we restrict ourselves to the pointing out this possibility here.

5.3 Lower bound for $\widehat{f}_N$

Lemma 4 *Under the assumptions of Theorem 1, for almost all $\omega \in \Omega$ there exist finite numbers $N_2(\omega)$ such that for any $x \in [0,1]$ and for all $N \ge N_2(\omega)$*

$$\widehat{f}_N(x) \ge f(x) - \frac{C_4(\beta)}{h^2}\left(\frac{\log N}{N}\right)^{\frac{2+\beta}{1+\beta}} \tag{112}$$

with constant $C_4(\beta)$ defined in (37).

Proof of Lemma 4. Let us take use of Lemma 7 and its Corollary 2 introducing

$$\delta_y = L_{f,\beta}\,\delta_x^{\beta}, \quad \delta_x \triangleq \left(\frac{2C_f \log N}{f_{\min} L_{f,\beta}\, N}\right)^{\frac{1}{1+\beta}} . \tag{113}$$

Thus, for any $N \geq N_6(\omega)$ and any $x \in [0,1]$ there exists (with probability one) an integer $i_k \in \{1, \ldots, N\}$ such that

$$|x - X_{i_k}| \leq \delta_x \tag{114}$$

and

$$Y_{i_k} \geq f(X_{i_k}) - \delta_y\,. \tag{115}$$

Now, the estimation error at a point x can be expanded as

$$\begin{aligned} f(x) - \widehat{f}_N(x) &= [f(x) - f(X_{i_k})] && (116)\\ &+ \left[f(X_{i_k}) - \widehat{f}_N(X_{i_k})\right] && (117)\\ &+ \left[\widehat{f}_N(X_{i_k}) - \widehat{f}_N(x)\right]. && (118) \end{aligned}$$

The term in the right hand side (116) may be bounded as follows

$$|f(x) - f(X_{i_k})| \leq L_{f,\beta}\, |x - X_{i_k}|^{\beta} \leq L_{f,\beta}\,\delta_x^{\beta}, \tag{119}$$

as well as the term (118)

$$\left|\widehat{f}_N(X_{i_k}) - \widehat{f}_N(x)\right| \leq L_{\widehat{f}_N}\, |x - X_{i_k}| \leq L_{\widehat{f}_N}\,\delta_x, \tag{120}$$

with a Lipschitz constant $L_{\widehat{f}_N}$ for the function estimator $\widehat{f}_N(x)$. Remind that $\widehat{f}_N(X_{i_k}) \geq Y_{i_k}$ due to (15) or (20). Thus, (115) implies

$$f(X_{i_k}) - \widehat{f}_N(X_{i_k}) \leq (Y_{i_k} + \delta_y) - Y_{i_k} = \delta_y\,. \tag{121}$$

Combining all these bounds we obtain from (116) that for all $N \geq N_6(\omega)$,

$$f(x) - \widehat{f}_N(x) \leq \delta_y + L_{f,\beta}\,\delta_x^{\beta} + L_{\widehat{f}_N}\delta_x\,. \tag{122}$$

Therefore, applying Lemma 2 and substituting expressions (113) for δ_x and δ_y into (122) lead to the lower bound

$$\begin{aligned} \widehat{f}_N(x) &\geq f(x) - \left(2L_{f,\beta}\,\delta_x^{\beta} + L_{\widehat{f}_N}\delta_x\right) && (123)\\ &\geq f(x) - \frac{C_4(\beta)}{h^2}\left(\frac{\log N}{N}\right)^{\frac{2+\beta}{1+\beta}} && (124) \end{aligned}$$

for any sufficiently large N (starting from random a.s. finite integer, which does not depend on x). The first inequality in (34) has been applied here in order to simplify the lower bound. Lemma 4 is proved. ■

5.4 Proof of Theorem 1

1. Since $|u| = u - 2u\mathbf{1}\{u < 0\}$, the L_1-norm of estimation error can be expanded as

$$\|\widehat{f}_N - f\|_1 = \int_0^1 \left[\widehat{f}_N(x) - f(x)\right] dx \tag{125}$$

$$+ \ 2\int_0^1 \left[f(x) - \widehat{f}_N(x)\right] \mathbf{1}\left\{\widehat{f}_N(x) < f(x)\right\} dx. \tag{126}$$

2. Applying Lemmas 1 and 3 to the right hand side (125) yields

$$\limsup_{N\to\infty} h^{-\beta}\left(\int_0^1 \left[\widehat{f}_N(x) - f(x)\right] dx\right) \le \gamma + 4C_\alpha(g_{\max} - 1)K_{\max}\mathbf{1}\{\beta = 1\} \quad \text{a.s.} \tag{127}$$

Note, that one may fix $\gamma = 2L_{f,\beta}\, g_{\max}\, C_\beta(K)$, for instance.

3. In order to obtain a similar result for the term (126), note that Lemma 4 implies

$$\zeta_N(x,\omega) \triangleq \varepsilon_{LB}^{-1}(N)\left[f(x) - \widehat{f}_N(x)\right] \le C_4(\beta) < \infty \quad \text{a.s.}$$

uniformly with respect to both $x \in [0,1]$ and $N \ge N_2(\omega)$, with

$$\varepsilon_{LB}(N) \triangleq \frac{1}{h^2}\left(\frac{\log N}{N}\right)^{\frac{2+\beta}{1+\beta}}. \tag{128}$$

Hence, one may apply Fatou lemma, taking into account that $u\mathbf{1}\{u > 0\}$ is a continuous, monotone function:

$$\limsup_{N\to\infty} \varepsilon_{LB}^{-1}(N) \int_0^1 \left[f(x) - \widehat{f}_N(x)\right] \mathbf{1}\left\{\widehat{f}_N(x) < f(x)\right\} dx \tag{129}$$

$$\le \int_0^1 \limsup_{N\to\infty} \zeta_N(x,\omega)\,\mathbf{1}\{\zeta_N(x,\omega) > 0\}\, dx \tag{130}$$

$$\le C_4(\beta) < \infty \quad \text{a.s.} \tag{131}$$

4. Thus, the obtained relations together with (125) and (126) imply (35). Theorem 1 is proved. ∎

6 Appendix

In Subsection 6.1 we establish some properties related to the corrected kernel. Subsection 6.2 presents some auxiliary lemmas which have been used to prove Theorem 1. Finally, we collect in Subsection 6.3 some lemmas dedicated to the proof of Remark 5.

6.1 Corrected kernel

Let the basic kernel function K be defined as in Section 4, and the bandwidth $h \in (0, 1/2)$. Remind the estimator $\widehat{f}_N$ defined in (5) as follows:

$$\begin{cases} \widehat{f}_N(x) = \sum_{i=1}^{N} \alpha_i K_h(x, X_i) \\ \alpha_i \ge 0, \qquad i = 1, \ldots, N, \end{cases} \tag{132}$$

where the kernel function

$$K_h(x,t) = h^{-1}K((x-t)/h) \qquad \forall\, x \in (h, 1-h) \tag{133}$$

while

$$K_h(x,t) = h^{-1}K((x-t)/h) \left(\int_{-1}^{x/h} K(t)\, dt \right)^{-1} \qquad \forall\, x \in [0, h] \tag{134}$$

and

$$K_h(x,t) = h^{-1}K((x-t)/h) \left(\int_{(x-1)/h}^{1} K(t)\, dt \right)^{-1} \qquad \forall\, x \in [1-h, 1]\,. \tag{135}$$

Thus, the kernel function $K_h(x,t)$ is defined for any $(x,t) \in [0,1] \times \mathbb{R}$, and the estimator (132) is defined for any $x \in [0,1]$ via the kernel $K_h(x,t)$ corrected at the "boundaries". One may easily observe that

$$\int_0^1 K_h(x,u)\, du = 1 \qquad \forall\, x \in [0,1] \tag{136}$$

and, consequently, due to exchanging the integral and the derivative,

$$\int_0^1 \frac{\partial}{\partial x} K_h(x,u)\; du = 0 \qquad \forall\, x \in [0,1] \tag{137}$$

Note, that equation (137) may also be verified directly. For instance, on the left boundary, i.e. for $x \in [0,h]$, we have

$$K_h(x,t) = h^{-1}K((x-t)/h) g(x)\,, \qquad g(x) = \left(\int_{-1}^{x/h} K(t)\, dt \right)^{-1}. \tag{138}$$

Denoting

$$\widetilde{K}_h(x,u) = \frac{\partial}{\partial x} K_h(x,u)\,, \tag{139}$$

we thus have

$$g'(x) = -\left(\int_{-1}^{x/h} K(t)\, dt \right)^{-2} h^{-1}K(x/h) = -g^2(x) h^{-1} K(x/h) \tag{140}$$

and

$$\begin{aligned} \widetilde{K}_h(x,u) &= h^{-1}K((x-u)/h) g'(x) + g(x) h^{-2} K'((x-u)/h) & (141)\\ &= g(x) h^{-1} \left(h^{-1} K'\left(\frac{x-u}{h}\right) - K\left(\frac{x-u}{h}\right) K_h(x,0) \right). & (142) \end{aligned}$$

Hence, the integral

$$\int_0^1 \widetilde{K}_h(x,u)\; du = -\frac{g^2(x)}{h^2} K\left(\frac{x}{h}\right) \int_0^1 K\left(\frac{x-u}{h}\right)\, du + \frac{g(x)}{h} \left[-K\left(\frac{x-u}{h}\right) \right]_{u=0}^{u=1} \tag{143}$$

equals zero for $x \in [0,h]$. A similar proof might be repeated for $x \in [1-h,1]$. Finally, equality (136) holds true for all $x \in (h, 1-h)$ too, since $g(x) \equiv 1$ over this interval.

In what follows, we use more general formulas (5)–(6) instead of (138), that is

$$K_h(x,t) = h^{-1}K((x-t)/h)\,g(x)\,, \quad g(x) = \left(\int_{(x-1)/h}^{x/h} K(t)\,dt\right)^{-1}, \quad x \in [0,1]\,. \tag{144}$$

Therefore, as follows from (139), (144) for any $x \in [0,1]$,

$$\begin{aligned}
\widetilde{K}_h(x,u) &= h^{-1}K((x-u)/h)g'(x) + g(x)h^{-2}K'((x-u)/h) && (145)\\
&= \frac{g(x)}{h}\left(\frac{1}{h}K'\left(\frac{x-u}{h}\right) + K\left(\frac{x-u}{h}\right)(K_h(x,1)-K_h(x,0))\right). && (146)
\end{aligned}$$

The following Lemma proves Lipschitz-like constants in (107) and (61)–(62).

Lemma 5 *Let kernel K_h defined in (5)–(6) meets the assumptions B1–B3, and the bandwidth $h \in (0,1/2)$. Let $\widetilde{K}_n$ be defined by (139). Then the following upper bounds hold true:*

$$\left|\widetilde{K}_h(x,u)\right| \le g_{\max}h^{-2}\left(L_K + g_{\max}K_{\max}^2\right)\,, \tag{147}$$

$$\left|\frac{\partial}{\partial u}\,\widetilde{K}_h(x,u)\right| \le g_{\max}h^{-3}L_{\widetilde{K}}\,, \quad \left|\frac{\partial^2}{\partial u^2}\,\widetilde{K}_h(x,u)\right| \le g_{\max}h^{-4}L_{\widetilde{K}'}\,, \tag{148}$$

where $L_{\widetilde{K}} = L_{K'} + L_K\,g_{\max}K_{\max}$ and $L_{\widetilde{K}'} = L_{K''} + L_{K'}\,g_{\max}K_{\max}$. Moreover,

$$\left|\frac{\partial^3}{\partial x^3}\,K_h(x,u)\right| \le g_{\max}L_{\widetilde{K}''}h^{-4} \tag{149}$$

where

$$\begin{aligned}
L_{\widetilde{K}''} &= g_{\max}\left[L_{K''} + 3L_{K'}K_{\max}g_{\max} + 3L_K g_{\max}K_{\max}^2(1+3g_{\max})\right. && (150)\\
&\quad \left.+ (L_K^2 + 2g_{\max}^2K_{\max}^4)(1+2g_{\max})\right]\,. && (151)
\end{aligned}$$

Proof of Lemma 5. The upper bound (147) follows directly from (145)–(146). Furthermore, taking (145), (146) into account, one easily may come to (107) since

$$\left|\frac{\partial}{\partial u}\,\widetilde{K}_h(x,u)\right| \le g_{\max}h^{-3}\left(L_{K'} + L_K\,g_{\max}K_{\max}\right) = g_{\max}h^{-3}L_{\widetilde{K}}\,, \tag{152}$$

and, similarly,

$$\left|\frac{\partial^2}{\partial u^2}\,\widetilde{K}_h(x,u)\right| \le g_{\max}h^{-4}\left(L_{K''} + L_{K'}\,g_{\max}K_{\max}\right) = g_{\max}h^{-4}L_{\widetilde{K}'}\,. \tag{153}$$

Moreover, one may continue calculation of further derivatives from (139)–(146) as follows:

$$\begin{aligned}
\frac{\partial^2}{\partial x^2}\,K_h(x,u) &= g'(x)h^{-1}\left(h^{-1}K'\left(\frac{x-u}{h}\right)\right. && (154)\\
&\qquad \left.+K\left(\frac{x-u}{h}\right)(K_h(x,1)-K_h(x,0))\right) && (155)\\
&\quad + g(x)h^{-1}\left(h^{-2}K''\left(\frac{x-u}{h}\right)\right. && (156)\\
&\qquad +h^{-1}K'\left(\frac{x-u}{h}\right)(K_h(x,1)-K_h(x,0)) && (157)\\
&\qquad \left.+K\left(\frac{x-u}{h}\right)\left(\frac{\partial}{\partial x}\,K_h(x,1) - \frac{\partial}{\partial x}\,K_h(x,0)\right)\right) && (158)
\end{aligned}$$

and

$$
\begin{aligned}
\frac{\partial^3}{\partial x^3} K_h(x,u) &= g''(x)h^{-1}\left(h^{-1}K'\left(\frac{x-u}{h}\right)\right. && (159)\\
&\qquad\left. + K\left(\frac{x-u}{h}\right)(K_h(x,1)-K_h(x,0))\right) && (160)\\
&\quad + 2g'(x)h^{-1}\left(h^{-2}K''\left(\frac{x-u}{h}\right)\right. && (161)\\
&\qquad + h^{-1}K'\left(\frac{x-u}{h}\right)(K_h(x,1)-K_h(x,0)) && (162)\\
&\qquad\left. + K\left(\frac{x-u}{h}\right)\left(\frac{\partial}{\partial x}K_h(x,1)-\frac{\partial}{\partial x}K_h(x,0)\right)\right) && (163)\\
&\quad + g(x)h^{-1}\left(h^{-3}K'''\left(\frac{x-u}{h}\right)\right. && (164)\\
&\qquad + h^{-2}K''\left(\frac{x-u}{h}\right)(K_h(x,1)-K_h(x,0)) && (165)\\
&\qquad + 2h^{-1}K'\left(\frac{x-u}{h}\right)\left(\frac{\partial}{\partial x}K_h(x,1)-\frac{\partial}{\partial x}K_h(x,0)\right) && (166)\\
&\qquad\left. + K\left(\frac{x-u}{h}\right)\left(\frac{\partial^2}{\partial x^2}K_h(x,1)-\frac{\partial^2}{\partial x^2}K_h(x,0)\right)\right). && (167)
\end{aligned}
$$

Moreover, from (144) the derivatives follow

$$
\begin{aligned}
g'(x) &= g^2(x)h^{-1}\left(K\left(\frac{x-1}{h}\right)-K\left(\frac{x}{h}\right)\right), && (168)\\
g''(x) &= 2g(x)g'(x)h^{-1}\left(K\left(\frac{x-1}{h}\right)-K\left(\frac{x}{h}\right)\right) && (169)\\
&\quad + g^2(x)h^{-2}\left(K'\left(\frac{x-1}{h}\right)-K'\left(\frac{x}{h}\right)\right), && (170)
\end{aligned}
$$

therefore,

$$
\begin{aligned}
|g'(x)| &\leq g_{\max}^2 K_{\max} h^{-1}, && (171)\\
|g''(x)| &\leq g_{\max}^2\left(L_K + 2g_{\max}K_{\max}^2\right)h^{-2}. && (172)
\end{aligned}
$$

Finally, using the bounds (171)–(172) in (159)–(167) and the definition of $\widetilde{K}_h$ (144) we arrive at the bound (61)–(62):

$$h^4 \left| \frac{\partial^3}{\partial x^3} K_h(x,u) \right| \leq g_{\max}^2 \left(L_K + 2 g_{\max} K_{\max}^2 \right) \left(L_K + g_{\max} K_{\max}^2 \right) \tag{173}$$

$$+ 2 g_{\max}^2 K_{\max} \left(L_{K'} + L_K g_{\max} K_{\max} \right. \tag{174}$$

$$\left. + K_{\max} h^2 \max_{x,u} |\widetilde{K}_h(x,u)| \right) \tag{175}$$

$$+ g_{\max} (L_{K''} + L_{K'} g_{\max} K_{\max} + 2 L_K h^2 \max_{x,u} |\widetilde{K}_h(x,u)| \tag{176}$$

$$+ K_{\max} h^3 \max_{x,u} |\partial \widetilde{K}_h(x,u)/\partial x|) \tag{177}$$

$$\leq g_{\max} [L_{K''} + 3 L_{K'} g_{\max} K_{\max} \tag{178}$$

$$+ 3 L_K g_{\max} \left(L_K + K_{\max}^2 g_{\max} \right) \tag{179}$$

$$+ K_{\max} g_{\max}^2 \left(3 K_{\max} L_K + 2 K_{\max}^3 g_{\max}^2 \right)] \tag{180}$$

$$= g_{\max} \left[L_{K''} + 3 L_{K'} K_{\max} g_{\max} + 3 L_K g_{\max} K_{\max}^2 (1 + 3 g_{\max}) \right. \tag{181}$$

$$\left. + (L_K^2 + 2 g_{\max}^2 K_{\max}^4)(1 + 2 g_{\max}) \right] . \tag{182}$$

Here we applied the upper bound (147) as well as the one, followed from (145)–(146) and (154)–(158):

$$h^3 \max_{x,u} \left| \frac{\partial}{\partial x} \widetilde{K}_h(x,u) \right| \leq g_{\max}^2 K_{\max} \left(L_K + g_{\max} K_{\max}^2 \right) \tag{183}$$

$$+ g_{\max} h^{-3} \left(L_{K'} + L_K g_{\max} K_{\max} \right. \tag{184}$$

$$\left. + K_{\max} g_{\max} \left(L_K + K_{\max}^2 g_{\max} \right) \right) . \tag{185}$$

Lemma 5 is proved. ■

6.2 Auxiliary lemmas. I

The following results are proved here for the sake of completeness.

Lemma 6 *Let function $f : [0,1] \to \mathbb{R}$ meets the assumption A1 and sequence $(X_i)_{i=1,\ldots,N}$ be obtained from an independent sample with p.d.f. $f(x)/C_f$ by increase ordering (41), where C_f is defined by (3). Denote $X_0 = 0$ and $X_{N+1} = 1$. Then for any finite constant $C_X > 4C_f/f_{\min}$ there exist almost surely finite number $N_0 = N_0(\omega)$ such that*

$$\max_{i=1,\ldots,N+1} \Delta X_i \leq C_X \frac{\log N}{N} \quad \forall\, N \geq N_0 \tag{186}$$

with probability 1. For instance, one may fix constant C_X as follows:

$$C_X = 5 f_{\max}/f_{\min} \,. \tag{187}$$

Proof of Lemma 6. Introduce a uniform partition of the interval $[0,1]$ onto m_N subintervals Δ_k with equal Lebesgue measures

$$\ell(\Delta_k) \triangleq 1/m_N \leq C_X \log N/(2N) \,, \quad k = 1, \ldots, m_N \,, \tag{188}$$

where size of partition

$$m_N \triangleq \min\{\text{integer } m \,:\, m \geq 2N/(C_X \log N)\} \tag{189}$$

$$\leq 1 + \frac{2N}{C_X \log N} \leq \frac{(2+\varepsilon)N}{C_X \log N} \tag{190}$$

for an arbitrary $\varepsilon > 0$ and for any sufficiently large N. Hence, the event

$$A_N \triangleq \{\omega : \max_{i=1,\dots,N+1} \Delta X_i \leq C_X \log N/N\} \tag{191}$$

$$\supseteq \bigcap_{k=1}^{m_N} \left[\bigcup_{i=1}^{N} \{X_i \in \Delta_k\} \right] . \tag{192}$$

Basing on Borel–Cantelli lemma we prove that the complementary event $A_N^c \triangleq \Omega \setminus A_N$ may occur only finite number of times (with probability 1). Evidently,

$$P\left(A_N^c\right) \leq \sum_{k=1}^{m_N} P\left(\bigcap_{i=1}^{N} \{X_i \notin \Delta_k\} \right) \tag{193}$$

$$= \sum_{k=1}^{m_N} \prod_{i=1}^{N} \left(1 - \int_{\Delta_k} C_f^{-1} f(u)\, du \right) \tag{194}$$

$$\leq m_N \left(1 - \frac{f_{\min}}{C_f} \cdot \ell(\Delta_1) \right)^N \tag{195}$$

$$\leq \frac{(2+\varepsilon)N}{C_X \log N} \exp\left\{ -\frac{f_{\min} C_X}{(2+\varepsilon) C_f} \cdot \log N \right\} \tag{196}$$

$$= O\left(N^{1 - f_{\min} C_X/((2+\varepsilon)C_f)} \right) . \tag{197}$$

Hence, condition $C_X > 4C_f/f_{\min}$ implies the existence of positive ε ensuring the convergence of series

$$\sum_{N=1}^{\infty} P\left(A_N^c\right) < \infty , \tag{198}$$

and the Borel–Cantelli lemma applies. Note, that events $\bigcap_{i=1}^{N} \{X_i \notin \Delta_k\}$ do not depend on renumbering of $(X_i)_{i=1,\dots,N}$ which lead to (194) from (193); moreover, we have used both definition (189) and inequality $1 - x \leq e^{-x}$ there in (195)–(196). Lemma 6 is proved.

■

Lemma 7 *Let random sample $\{(X_i, Y_i) \mid i = 1, \dots, N\}$ be defined as in Section 2. Let sequence $\delta_x = \delta_x(N)$ be positive, and for some $\varepsilon > 0$*

$$\liminf_{N \to \infty} N^{1-\varepsilon} \delta_x > 0 . \tag{199}$$

Define

$$m_\delta \triangleq \min\{\text{integer } m : m \geq \delta_x^{-1}\} \tag{200}$$

and assume a positive sequence $\delta_y = \delta_y(N) < f_{\min}$ meeting for all sufficiently large N the inequality

$$\delta_y \geq \kappa\, m_\delta \frac{\log N}{N} , \quad \text{with} \quad \kappa > \frac{(2-\varepsilon)C_f}{f_{\min}} . \tag{201}$$

Then, under the assumptions of Lemma 6, with probability 1, there exists finite number $N_6(\omega)$ such that for any $N \geq N_6(\omega)$ there is such a subset of points $\{(X_{i_k}, Y_{i_k}),\ k = 1, \dots, m_\delta\}$ in the sample $\{(X_i, Y_i),\ i = 1, \dots, N\}$, that the following inequalities hold:

$$(k-1)/m_\delta \leq X_{i_k} < k/m_\delta , \quad f(X_{i_k}) - \delta_y \leq Y_{i_k} \leq f(X_{i_k}) . \tag{202}$$

Proof of Lemma 7. It is similar to that of Lemma 6. Introduce an equidistant partition of the interval $[0,1]$ onto subintervals $[(k-1)/m_\delta, k/m_\delta]$, $k = 1,\ldots,m_\delta$. Moreover, introduce the related subsets in $\mathbb{R}^2$

$$\Delta_k \triangleq \{(u,v) : (k-1)/m_\delta \le u \le k/m_\delta\,,\ f(u) - \delta_y \le v \le f(u)\}\,, \quad k = 1,\ldots,m_\delta\,. \tag{203}$$

Hence, the event

$$A_N \triangleq \{\omega : \forall\, k = 1,\ldots,m_\delta\ \exists\, i = 1,\ldots,N : (X_i,Y_i) \in \Delta_k\} \tag{204}$$

$$= \bigcap_{k=1}^{m_\delta}\left[\bigcup_{i=1}^{N}\{(X_i,Y_i) \in \Delta_k\}\right]\,. \tag{205}$$

Basing on Borel–Cantelli lemma we prove that the complementary event $A_N^c \triangleq \Omega \setminus A_N$ may occur only finite number of times (with probability 1). Evidently,

$$P\left(A_N^c\right) \le \sum_{k=1}^{m_\delta} P\left(\bigcap_{i=1}^{N}\{(X_i,Y_i) \notin \Delta_k\}\right) \tag{206}$$

$$= \sum_{k=1}^{m_\delta}\prod_{i=1}^{N}\left(1 - \int_{\Delta_k} C_f^{-1} f(u)\,du\,dv\right) \tag{207}$$

$$\le m_\delta\left(1 - \frac{f_{\min}}{C_f}\cdot\frac{\delta_y}{m_\delta}\right)^N \tag{208}$$

$$\le \left(1 + \delta_x^{-1}\right)\,\exp\left\{-\frac{f_{\min}\kappa}{C_f}\cdot\log N\right\} \tag{209}$$

$$= O\left(N^{1-\varepsilon-f_{\min}\kappa/C_f}\right)\,. \tag{210}$$

Hence, condition $\kappa > (2-\varepsilon)C_f/f_{\min}$ implies

$$\sum_{N=1}^{\infty} P\left(A_N^c\right) < \infty\,, \tag{211}$$

and one may apply Borel–Cantelli lemma. Lemma 7 is proved. ■

Corollary 2 *Let δ_x and δ_y meet the conditions of Lemma 7. Then, with probability 1, for any $N \ge N_6(\omega)$ and any $x \in [0,1]$ there exists integer $i_k \in \{1,\ldots,N\}$ such that $|x - X_{i_k}| \le \delta_x$ and $f(X_{i_k}) - \delta_y \le Y_{i_k} \le f(X_{i_k})$.*

Lemma 8 *Let function $g : [0,\Delta] \to \mathbb{R}$ be twice continuous differentiable, $\Delta > 0$. Then*

$$\max_{x\in[0,\Delta]}|g(x)| \le \max\{|g(0)|, |g(\Delta)|\} + \frac{\Delta^2}{8}\max_{x\in[0,\Delta]}|g''(x)|\,. \tag{212}$$

Proof of Lemma 8. Denote $\bar g_b = \max\{|g(0)|, |g(\Delta)|\}$. It suffices to prove the case where a point $x_1 \in (0,\Delta)$ exists with

$$|g(x_1)| = \max_{x\in[0,\Delta]}|g(x)| > \bar g_b\,. \tag{213}$$

Then $g'(x_1) = 0$, and for any $x \in [0, \Delta]$

$$g(x_1) = g(x) - \int_{x_1}^{x} dt \int_{x_1}^{t} g''(u)\, du\,. \tag{214}$$

Therefore, putting $x = \Delta$ one obtains from (214)

$$|g(x_1)| \le |g(\Delta)| + \int_{x_1}^{\Delta} dt \int_{x_1}^{t} |g''(u)|\, du \le \bar{g}_b + \frac{(\Delta - x_1)^2}{2} \max_{x\in[0,\Delta]} |g''(x)|\,. \tag{215}$$

Similarly, fixing $x = 0$ there in (214) leads to

$$|g(x_1)| \le |g(0)| + \int_{0}^{x_1} dt \int_{x_1}^{t} |g''(u)|\, du \le \bar{g}_b + \frac{x_1^2}{2} \max_{x\in[0,\Delta]} |g''(x)|\,. \tag{216}$$

Thus, combining (215) and (216) we arrive at

$$|g(x_1)| \le \bar{g}_b + \frac{1}{2} \min\{(\Delta - x_1)^2, x_1^2\} \max_{x\in[0,\Delta]} |g''(x)|\,. \tag{217}$$

Since

$$\max_{x\in[0,\Delta]} \min\{(\Delta - x)^2, x^2\} = \frac{\Delta^2}{4}\,, \tag{218}$$

the desired inequality (212) follows immediately from (213), (217)–(218). ■

6.3 Auxiliary lemmas. II

Lemma 9 states the results announced in the Remark 5, Subsection 5.2.

Lemma 9 *Let numbers h_N form a positive sequence, non-increasing for $N \ge N_1$ and meeting condition*

$$\frac{h_{N-1}}{h_N} \le 1 + \frac{\kappa}{N} \quad \forall\, N \ge N_1 \tag{219}$$

with finite, positive constants κ and N_1 and such that

$$\lim_{N\to\infty} \frac{\log N}{N h_N} = 0\,. \tag{220}$$

Then, under the assumptions of Lemma 6, as $N \to \infty$, for an arbitrary $\nu > 1$ and for any $x \in [0, 1]$

$$S_N \triangleq \sum_{i=1}^{N+1} (\Delta X_i)^3\, \mathbf{1}\{|x - X_i| \le 2h_N\} = o\left(\frac{h_N \log^{\nu} N}{N^2}\right) \quad \text{a.s.} \tag{221}$$

where $o(\cdot)$ does not depend on x.

Proof of Lemma 9.

1. Remind that the sequence of random points (X_i) is obtained from that of i.i.d. with the p.d.f. $f(\cdot)/C_f$ for X_i by their increase ordering. Furthermore, $\Delta X_i \triangleq X_i - X_{i-1}$, $X_0 = 0$, and $X_N \equiv 1$. Introduce σ-algebras $\mathcal{F}_N \triangleq \sigma\{X_1, \ldots, X_N\}$. Thus, $(S_N, \mathcal{F}_N)$ is a non-negative stochastic sequence. Let us denote X the new point (hence, independent

of $\mathcal{F}_N$) when passing from S_N to S_{N+1}. With these notations and due to the evident inequality

$$\mathbf{1}\{|x-X_i| \le 2h_{N+1}\} \le \mathbf{1}\{|x-X_i| \le 2h_N\}$$

one may write

$$\mathbb{E}(S_{N+1}|\mathcal{F}_N) \le \mathbb{E}\left\{\sum_{j=1}^{N+1} \mathbf{1}\{X \in [X_{j-1}, X_j)\} \cdot \left[\sum_{i \ne j}^{N+1} (\Delta X_i)^3 \mathbf{1}\{|x-X_i| \le 2h_N\}\right.\right. \quad (222)$$

$$+ \big((X - X_{j-1})^3 \mathbf{1}\{|x-X| \le 2h_N\} \quad (223)$$

$$\left.\left. +(X_j - X)^3 \mathbf{1}\{|x-X_j| \le 2h_N\}\big)\right] |\mathcal{F}_N\right\} \quad (224)$$

$$\le S_N - \sum_{j=1}^{N+1} \mathbf{1}\{|x-X_j| \le 2h_N\}\, \mathbb{E}\left\{\mathbf{1}\{X \in [X_{j-1}, X_j)\}\right. \quad (225)$$

$$\left. \cdot \left[(\Delta X_j)^3 - (X - X_{j-1})^3 - (X_j - X)^3\right] |\mathcal{F}_N\right\} \quad (226)$$

$$+ \sum_{j=1}^{N+1} \mathbb{E}\left\{(X - X_{j-1})^3 \mathbf{1}\{X \in [X_{j-1}, X_j)\} \left(\mathbf{1}\{|x-X| \le 2h_N\}\right.\right. \quad (227)$$

$$\left.\left. -\mathbf{1}\{|x-X_j| \le 2h_N\}\right) |\mathcal{F}_N\right\} . \quad (228)$$

2. The first need now is to evaluate the conditional expectation in (225)–(226). A simple algebras imply

$$(\Delta X_j)^3 - (X - X_{j-1})^3 - (X_j - X)^3 = 3(X(X_j + X_{j-1}) - X^2 - X_j X_{j-1})\Delta X_j$$

which is non-negative for any $X \in [X_{j-1}, X_j)$. Therefore, the bounding from below leads to

$$\mathbb{E}\left\{\mathbf{1}\{X \in [X_{j-1}, X_j)\} \left[(\Delta X_j)^3 - (X - X_{j-1})^3 - (X_j - X)^3\right] |\mathcal{F}_N\right\} \quad (229)$$

$$= 3\Delta X_j \int_{X_{j-1}}^{X_j} \frac{f(x)}{C_f} \left(x(X_j + X_{j-1}) - x^2 - X_j X_{j-1}\right) dx \quad (230)$$

$$\ge \frac{f_{\min}}{2C_f} (\Delta X_j)^4 . \quad (231)$$

Substituting to (225)–(226) and applying Iensen's inequality for the convex function $\psi(s) \triangleq s^{4/3}$,

$$\left(\frac{S_N}{\#\{|x-X_j| \le 2h_N\}}\right)^{4/3} \le \frac{\sum_{j=1}^{N+1} \mathbf{1}\{|x-X_j| \le 2h_N\}\ (\Delta X_j)^4}{\#\{|x-X_j| \le 2h_N\}} , \quad (232)$$

lead to

$$\mathbb{E}(S_N|\mathcal{F}_{N-1}) \le S_{N-1} - \frac{f_{\min}}{2C_f} \frac{S_{N-1}^{4/3}}{(Nq_N)^{1/3}} + r_N \quad (233)$$

where

$$q_N \triangleq \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}\{|x-X_j| \le 2h_{N-1}\} = O(h_{N-1}) , \quad (234)$$

and r_N denotes the related sum in (227)–(228), that is

$$r_N \triangleq \sum_{j=1}^{N} \mathbb{E}\left\{(X-X_{j-1})^3 \mathbf{1}\{X \in [X_{j-1}, X_j)\} \left(\mathbf{1}\{|x-X| \le 2h_{N-1}\}\right.\right. \tag{235}$$

$$\left.\left. -\mathbf{1}\{|x-X_j| \le 2h_{N-1}\}\right) |\mathcal{F}_{N-1}\right\} . \tag{236}$$

Note, that one may define $0/0 \triangleq 0$ to treat the case of zero denominators there in (232), for instance.

The bound $O(h_N)$ for q_N stated in (234) is proved below in Lemma 11. In order to bound r_N from above one may easily see that the difference between the two indicators in (235)–(236) is positive iff the first of them equals 1 while the second does 0. Due to Lemma 6 and the property (220), i.e. $\log N/(Nh_N) \to 0$, this may almost surely arise only for the following event (for any sufficiently large N): $x - 2h_{N-1} \le X_{j-1} < X \le x + 2h_{N-1} < X_j$. Given a sequence (X_i), this event arises only for one j, say $j = j_0$, which depends on x . Thus,

$$r_N \le \int_{X_{j_0-1}}^{X_{j_0}} (u - X_{j_0-1})^3 \frac{f(u)}{C_f} du \le \frac{f_{\max}}{C_f} \max_{i=1,\ldots,N+1} (X_i - X_{i-1})^4 \tag{237}$$

$$= O\left((\log N/N)^4\right), \tag{238}$$

with non-random $O(\cdot)$ being independent of x, from Lemma 6.

3. The next step is to come from the nonlinear inequality (233) to a linear one. The convexity of function $\psi(s) = s^{4/3}$ gives, for an arbitrary $a_N > 0$, the lower bound as follows:

$$\psi(S_N) \ge \psi(a_N) + \psi'(a_N)(S_N - a_N) = \frac{4}{3} a_N^{1/3} S_N - \frac{1}{3} a_N^{4/3} .$$

Thus, inequality (233) and the choice $a_{N-1} \triangleq a^3 q_N / N^2$ with $a > 0$ lead to

$$\mathbb{E}(S_N|\mathcal{F}_{N-1}) \le S_{N-1} - \frac{2f_{\min}}{3C_f} \left(\frac{a_{N-1}}{Nq_N}\right)^{1/3} S_{N-1} + \frac{f_{\min}}{6C_f} \left(\frac{a_{N-1}^4}{Nq_N}\right)^{1/3} + r_N \tag{239}$$

$$\le S_{N-1} - \frac{\mu}{N} S_{N-1} + \frac{\mu a^3}{4N^3} q_N + r_N \tag{240}$$

where

$$\mu \triangleq \frac{2af_{\min}}{3C_f} > 2 + \kappa \tag{241}$$

for sufficiently large a.

4. Finally, using relations (234)–(236) in (239)–(240) and applying Lemma 10 we arrive at the result of Lemma 9. ∎

Lemma 10 *Let $(w_N, \mathcal{F}_N)$ be non-negative stochastic sequence meeting the inequality*

$$\mathbb{E}(w_N|\mathcal{F}_{N-1}) \le \left(1 - \frac{\mu}{N}\right) w_{N-1} + \frac{d_N h_N}{N^{1+p}} \quad \text{a.s.} \quad \forall N \ge N_1 \tag{242}$$

where $\mu > p + \kappa$, h_N and κ meet conditions of Lemma 9, d_N is $\mathcal{F}_{N-1}$-measurable, non-negative and bounded a.s., and $N_1 < \infty$. Then, as $N \to \infty$, for any $\nu > 1$

$$w_N = o\left(\frac{\log^\nu N}{N^p} h_N\right) \quad \text{a.s.} \tag{243}$$

Proof of Lemma 10. Introduce

$$v_N \triangleq \frac{N^p}{h_N \log^\nu N} w_N \quad \text{a.s.} \tag{244}$$

The inequalities $\log N > \log(N-1)$, $h_{N-1} \geq h_N$, and (242) imply

$$\begin{aligned}
\mathbb{E}(v_N|\mathcal{F}_{N-1}) &\leq \left(1-\frac{\mu}{N}\right)\left(\frac{N}{N-1}\right)^p \frac{h_{N-1}}{h_N} v_{N-1} + \frac{d_N}{N\log^\nu N} &(245)\\
&\leq \left(1-\frac{\mu-p-\kappa}{N} + o\left(\frac{1}{N}\right)\right) v_{N-1} + \frac{d}{N\log^\nu N} &(246)
\end{aligned}$$

Since

$$\sum^\infty \frac{\mu-p-\kappa}{N} = \infty \quad \text{and} \quad \sum^\infty \frac{d_N}{N\log^\nu N} < \infty \quad \text{a.s.}\,, \tag{247}$$

one may apply Robbins–Siegmund almost supermartingale convergence theorem ROBBINS & SIEGMUND [25] which implies $v_N \to 0$ as $N \to \infty$. Lemma 10 is proved. ■

Lemma 11 *Let $h = h_N \to 0$ as $N \to \infty$. Then, under the assumptions of Lemma 6 and Lemma 9, the bound (234) holds true for any $x \in [0,1]$, that is*

$$q_N \triangleq \frac{1}{N}\sum_{j=1}^N \mathbf{1}\{|x-X_j| \leq 2h_{N-1}\} = O(h_N) \tag{248}$$

where $O(\cdot)$ does not depend on x.

Proof of Lemma 11. Introduce

$$\zeta_i \triangleq \mathbf{1}\{|x-X_i| \leq 2h_{N-1}\} - P\{|x-X_i| \leq 2h_{N-1}\} \tag{249}$$

leading to the decomposition

$$q_N = P\{|x-X_1| \leq 2h_{N-1}\} + \frac{1}{N}\sum_{j=1}^N \zeta_i\,. \tag{250}$$

Since X_i are i.i.d. with the bounded p.d.f. $f(\cdot)/C_f$, the probability

$$P\{|x-X_i| \leq 2h_{N-1}\} \leq \int_{x-2h_{N-1}}^{x+2h_{N-1}} \frac{f(u)}{C_f}\,du = O(h_{N-1}) \tag{251}$$

with $O(\cdot)$ being independent of x and of i. Furthermore, observe that $|\zeta_i| \leq 1$ a.s., and

$$\mathbb{E}\zeta_i = 0\,, \qquad \mathbb{E}\zeta_i^2 \leq P\{|x-X_i| \leq 2h_{N-1}\} = O(h_{N-1})\,. \tag{252}$$

Thus, in order to bound the stochastic term in the right hand side (250) one may apply the Bernstein inequality (see, e.g., BIRGÉ & MASSART [3] or BOSQ [4], Theorem 2.6) with the standard treatment via Borel–Cantelli lemma (e.g., as in BOUCHARD *et al* [5], Appendix, Lemma 5). This directly yields

$$q_N = O(h_N) + O\left(\left(\frac{h_N \log N}{N}\right)^{1/2}\right) = O(h_N) \quad \text{a.s.} \tag{253}$$

Lemma 11 is proved. ■

References

[1] Abbar, H. (1990) Un estimateur spline du contour d'une répartition ponctuelle aléatoire. *Statistique et analyse des données*, **15**(3), 1–19.

[2] Baufays, P. and Rasson, J.P. (1985) A new geometric discriminant rule. *Computational Statistics Quaterly*, **2**, 15–30.

[3] Birgé, L. and Massart, P. (1995) Minimum contrast estimators on sieves. *Preprint Université Paris Sud, France*, **95-42**.

[4] Bosq, D. (2000) Linear processes in function spaces. Theory and applications. in *Lecture Notes in Statistics*, **149**, Springer-Verlag, New York.

[5] Bouchard, G., Girard, S., Iouditski A.B., and Nazin A.V.(2003) Linear programming problems for frontier estimation. *Technical Report INRIA-4717;* `http://www.inria.fr/rrrt/rr-4717.html` and *Technical Report IAP RT-0304;* `http://www.stat.ucl.ac.be/Iapdp/tr2003/TR0304.ps`.

[6] Bouchard, G., Girard, S., Iouditski A.B., and Nazin A.V.(2004) Nonparametric Frontier Estimation by Linear Programming. *Automation and Remote Control,* **65**(1), 58–64.

[7] Deprins, D., Simar, L. and Tulkens, H. (1984) Measuring Labor Efficiency in Post Offices. in *The Performance of Public Enterprises: Concepts and Measurements* by M. Marchand, P. Pestieau and H. Tulkens, North Holland ed, Amsterdam.

[8] Devroye, L.P. and Wise, G.L. (1980) Detection of abnormal behavior via non parametric estimation of the support. *SIAM J. Applied Math.*, **38**, 448–480.

[9] Gardes, L. (2002) Estimating the support of a Poisson process via the Faber-Shauder basis and extreme values. *Publications de l'Institut de Statistique de l'Université de Paris*, **XXXXVI**, 43–72.

[10] Geffroy, J. (1964) Sur un problème d'estimation géométrique. *Publications de l'Institut de Statistique de l'Université de Paris*, **XIII**, 191–200.

[11] Gijbels, I. and Peng, L. (2000). Estimation of a support curve via order statistics. *Extremes*, **3**, 251–277.

[12] Girard, S. and Jacob, P. (2003a) Extreme values and Haar series estimates of point processes boundaries. *Scandinavian Journal of Statistics*, **30**(2), 369–384.

[13] Girard, S. and Jacob, P. (2003b) Projection estimates of point processes boundaries. *Journal of Statistical Planning and Inference*, **116**(1), 1–15.

[14] Girard, S. and Jacob, P. (2004) Extreme values and kernel estimates of point processes boundaries. *ESAIM: Probability and Statistics*, **8**, 150–168.

[15] Girard, S. and Menneteau, L. (2004) Limit theorems for smoothed extreme values estimates of point processes boundaries. *Journal of Statistical Planning and Inference*, to appear.

[16] Hall, P., Nussbaum, M. and Stern, S.E. (1997) On the estimation of a support curve of indeterminate sharpness. *J. Multivariate Anal.*, **62**, 204–232.

[17] Hall, P., Park, B. U. and Stern, S. E. (1998) On polynomial estimators of frontiers and boundaries. *J. Multiv. Analysis*, **66**, 71–98.

[18] Härdle, W., Hall, P. and Simar, L. (1995) Iterated bootstrap with application to frontier models. *J. Productivity Anal.*, **6**, 63–76.

[19] Härdle, W., Park, B. U. and Tsybakov, A. B. (1995) Estimation of a non sharp support boundaries. *J. Multiv. Analysis*, **43**, 205–218.

[20] Hardy, A. and Rasson, J.P. (1982) Une nouvelle approche des problèmes de classification automatique. *Statistique et Analyse des données*, **7**, 41–56.

[21] Hartigan, J.A. (1975) *Clustering Algorithms,* Wiley, Chichester.

[22] Jacob, P. and Suquet, P. (1995) Estimating the edge of a Poisson process by orthogonal series. *Journal of Statistical Planning and Inference*, **46**, 215–234.

[23] Knight, K. (2001) Limiting distributions of linear programming estimators. *E*xtremes, **4**(2), 87–103.

[24] Korostelev, A.P. and Tsybakov, A.B. (1993) Minimax theory of image reconstruction. in *Lecture Notes in Statistics*, **82**, Springer-Verlag, New York.

[25] Robbins, H. and Siegmund, D. (1971) A convergence theorem for non-negative almost supermartingales and some applications. In *J.S. Rustagi, editor, Optimizing Methods in Statistics,* Academic Press, New York, 233–257.

[26] Tarassenko, L., Hayton, P., Cerneaz, N. and Brady, M. (1995) Novelty detection for the identification of masses in mammograms. In *Proceedings fourth IEE International Conference on Artificial Neural Networks*, 442–447, Cambridge.