Automation and Remote Control, Vol. 66, No. 12, 2005, pp. 2000–2018. Translated from Avtomatika i Telemekhanika, No. 12, 2005, pp. 143–161.


STOCHASTIC SYSTEMS

# $L_1$-Optimal Nonparametric Frontier Estimation via Linear Programming

**S. Girard*, A. Iouditski*, and A. V. Nazin**[1]**

*LMC-IMAG, University Grenoble-I, Grenoble, France

**Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia

Received January 13, 2005

**Abstract**—A frontier estimation method for a set of points on a plane is proposed, being optimal in $L_1$-norm on a given class of $\beta$-Hölder boundary functions under $\beta \in (0, 1]$. The estimator is defined as sufficiently regular linear combination of kernel functions centered in the sample points, which covers all these points and whose associated support is of minimal surface. The linear combination weights are calculated via solution of the related linear programming problem. The $L_1$-norm of the estimation error is demonstrated to be convergent to zero with probability one, with the optimal rate of convergence.

## 1. INTRODUCTION

Many proposals are given in the literature for stating a problem and defining an estimation method for a set $S \subset \mathbb{R}^2$ based on a finite random set of points drawn from the interior. This problem of edge or support estimation arises in classification [1], clustering problems and discriminant analysis [2, 3], and outliers detection [4]. Applications may be also found in medical diagnosis [5] as well as in condition monitoring of machines [4]. 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 $\mathbb{R}^2$ [6]. We also point out some applications in econometrics (e.g., see [7]). In such and all other cases, when the unknown support can be written in the form

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

where $f : [0, 1] \to (0, +\infty)$ is an unknown bounded function, the problem reduces to estimating the boundary function $f$ (see [8], for instance). The data consist of pairs $(X, Y)$ where $X$ represents, for instance, 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 in the field [9] was written for independent identically distributed observations from a density $\phi$. The proposed estimator is a kind of histogram based on the extreme values of the sample. Subsequently, this work was extended in two main directions.

On the one hand, piecewise polynomial estimators were introduced, which 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 in [6, 10] under weak assumptions on the rate of decrease $\alpha$ of the density $\phi$ towards 0. Extreme values methods are then proposed in [11, 12] to estimate the parameter $\alpha$. Notice, that estimating $f$ can also been considered as a regression problem $Y = f(X) + \varepsilon$ with negative noise $\varepsilon$. In this context, local polynomial estimates are introduced; e.g., see [13, 14].

---

[1] The research has been carried out during the visits to LMC-IMAG (Grenoble–I University) and INRIA Rhône-Alpes, MISTIS project, France, in the year of 2004.

On the other hand, different propositions for smoothing Geffroy's estimator were made in the case of a Poisson point process. In [15], the introduced estimators are based on kernel regressions and orthogonal series method [16, 17]. Similarly, a Faber-Shauder estimator has been proposed in [18]. A general framework for studying estimators of this type is described in [19], as well as the generalizing to sets of the type

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

where $f$ is an unknown function and $E$ an arbitrary, given set.

Note, that the limit distribution of the estimator is established in all these works. We also refer to [20, 21], where similar smoothing approach has been developed for the Poisson process without extreme values application.

The estimator proposed in [22, 23], 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 (see also [24]), we propose essential modifications of the method introduced in [22, 23]. First, a corrected kernel is proposed ensuring a centered bias over all the estimation interval. Second, some regularity constraints are introduced in the optimization problem guaranteeing the necessary estimator regularity. The resulting estimator reaches the optimal in $L_1$-norm minimax convergence rate (up to a logarithmic factor). Note also, that we consider more general w.r.t. [22, 23] class of the Hölder functions which depends not only on Lipschitz constant $L$, but on exponent parameter $\beta \in (0, 1]$; under $\beta = 1$ we arrive at the case studied there in [22, 23].

The paper structure is as follows: the problem and the estimator are described in Section 2, and some preliminary properties of the estimator are formulated in Section 3 and proved in the Appendix; in the subsequent section, the main results are described, and their proof is postponed to the Appendix too.

## 2. PROBLEM STATEMENT AND FRONTIER 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 $Z_N = (X_i, Y_i)_{i=1,\dots,N}$ with independent pairs $(X_i, Y_i)$ being uniformly distributed in the set $S$ defined as in (1). For the sake of simplicity and without generality restriction, we further consider functions $f$ defined over all number axis $\mathbb{R}$, defining $f(x) = 0$ for all $x \notin [0, 1]$. Introduce functional

$$C_f \triangleq \int_0^1 f(u)\, du = \int_{\mathbb{R}} f(u)\, du\,. \tag{2}$$

Thus, random variables $X_i$ are distributed in $[0, 1]$ with the density $f(\cdot)/C_f$, and $Y_i$ have 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 $\beta$-Hölder with constant $L_{f,\beta}$ :

$$|f(x) - f(u)| \le L_{f,\beta}\, |x - u|^{\beta} \quad \forall x, u \in [0, 1]\,. \tag{3}$$

The considered estimator is chosen from the following family of functions:

$$\begin{cases} \widehat{f}_N(x) = \sum_{i=1}^{N} \alpha_i \, K_h(x, X_i) \,, \qquad K_h(x,t) = \dfrac{g(x)}{h} K\left(\dfrac{x-t}{h}\right) \\ \alpha_i \geq 0, \qquad i = 1, \ldots, N, \end{cases} \tag{4}$$

where $K(\cdot)$ is a sufficiently smooth basic kernel function $K : \mathbb{R} \to [0, +\infty)$ meeting normalization condition

$$\int_{\mathbb{R}} K(u) \, du = 1$$

and having the interval $[-1, 1]$ as its support; the bandwidth parameter $h \in (0, 1/2)$ depends on the sample size $N$ such that $h \to 0$ as $N \to \infty$; function

$$g(x) = \left( \int_{(x-1)/h}^{x/h} K(t) \, dt \right)^{-1}, \quad x \in [0, 1] \,, \tag{5}$$

corrects the basic kernel $K$ at the boundaries of $[0, 1]$, 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 $x \in [0, h)$ and/or $x \in (1-h, 1]$.

One may easily observe that

$$\int_0^1 K_h(x, u) \, du = 1 \qquad \forall \, x \in [0, 1] \tag{6}$$

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

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

Note, that Eq. (7) may be verified directly, see [24].

Denote $K_{\max} \triangleq \max K(t)$, $g_{\max} \triangleq \sup_{x,h} g(x)$ and introduce functionals

$$C_\beta(\varphi) \triangleq \int_{-1}^{1} |t|^\beta |\varphi(t)| \, dt \,, \quad \varphi \in C^0([-1, 1]) \,, \tag{8}$$

$$C_\beta(K, K') \triangleq g_{\max} \, K_{\max} C_\beta(K) + C_\beta(K') \,. \tag{9}$$

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

$$|\varphi(s) - \varphi(t)| \leq L_\varphi |s - t| \,, \tag{10}$$

with $L_\varphi < \infty$. 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

$$\widehat{S}_N \triangleq \{(x, y) : \, 0 \leq x \leq 1 \,, \; 0 \leq y \leq \widehat{f}_N(x)\} \tag{11}$$

may be approximated as follows:

$$\int_0^1 \widehat{f}_N(x) \, dx = \sum_{i=1}^{N} \alpha_i + O(h) \,. \tag{12}$$

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

$$J_P^* \triangleq \min_{\alpha} \sum_{i=1}^{N} \alpha_i \tag{13}$$

subject to

$$\widehat{f}_N(X_i) \geq Y_i\,, \quad i = 1, \dots, N\,, \tag{14}$$

$$|\widehat{f}_N'(X_i)| \leq L_{f,\beta}\, g_{\max} C_\beta(K, K') \frac{\log N}{Nh^2}\,, \quad i = 0, \dots, N+1\,, \tag{15}$$

$$C_\alpha h \geq \sum_{i=1}^{N} \alpha_i \, \mathbf{1}\{(j-1)/m_h \leq X_i < j/m_h\}\,, \quad j = 1, \dots, m_h\,, \tag{16}$$

$$0 \leq \alpha_i\,, \quad i = 1, \dots, N, \tag{17}$$

where parameter $m_h = \lfloor h^{-1} \rfloor$ stands for the integer part of $1/h$. This optimization problem may be formally written as linear program (LP)

$$J_P^* \triangleq \min_{\alpha} \mathbf{1}_N^T \alpha \tag{18}$$

subject to

$$Y \leq A\alpha\,, \tag{19}$$

$$\mathrm{abs}\,(B\alpha) \leq L_{f,\beta}\, g_{\max} C_\beta(K, K')\, \frac{\log N}{Nh^2}\, \mathbf{1}_N\,, \tag{20}$$

$$D^T \alpha \;\leq\; C_\alpha h\, \mathbf{1}_{m_h}\,, \tag{21}$$

$$0 \leq \alpha\,. \tag{22}$$

Here the vector inequalities and function abs$(\cdot)$ of a vector variable are treated componentwise, the $i$-th component of vector abs$(x)$ equals $|x_i|$. There is a positive parameter $C_\alpha$ in the constraints (16) and (21): its value will be discussed in Section 4. Moreover, the following notations have been introduced:

$$X_0 \triangleq 0, \quad X_{N+1} \triangleq 1, \tag{23}$$

$$\mathbf{1}_N \triangleq (1, 1, \dots, 1)^T \in \mathbb{R}^N, \tag{24}$$

$$A \triangleq \|K_h(X_i\,, X_j)\|_{i,j=1,\dots,N}\;, \tag{25}$$

$$B \triangleq \left\| \frac{d}{dx} K_h(x, X_j) \Big|_{x=X_i} \right\|_{i,j=1,\dots,N}\,, \tag{26}$$

$$D \triangleq \|\mathbf{1}\{(j-1)/m_h \leq X_i < j/m_h\}\|_{i=1,\dots,N;\, j=1,\dots,m_h}\;, \tag{27}$$

$$Y \triangleq (Y_1, \dots, Y_N)^T. \tag{28}$$

The main point of constraints (14)–(17) and/or (19)–(22) becomes clear in the following section after introducing assumptions and establishing the preliminary results.

## 3. ASSUMPTIONS AND PRELIMINARY RESULTS

The basic assumptions on the boundary function are as follows:

(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 a fixed constants $L_{f,\beta} < \infty$ and $\beta \in (0,1]$.

The following assumptions on the kernel function are introduced:

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

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

(B3) Function $K$ is three times continuously differentiable.

Note, that $g_{\max} = \left(\min\left\{\int_0^1 K(t)\,dt, \int_{-1}^0 K(t)\,dt\right\}\right)^{-1}$; in particular, $g_{\max} = 2$ for any even kernel $K(\cdot)$ meeting conditions (B1), (B2). We quote two preliminary results on the estimator $\widehat{f}_N$. First, the surface $\widehat{S}_N$ of the related support estimate is approximately $\sum_{i=1}^N \alpha_i$, see (12). Second, the function $\widehat{f}_N$ is Lispchitzian. The proofs are postponed to the Appendix.

**Lemma 1.** *Suppose assumptions* (B1), (B2) *are verified, and* $0 < h < 1/4$. *Moreover, let conditions* (16) *and* (17) *hold true for* $m_h = \lfloor h^{-1} \rfloor$. *Then the surface of the estimated support* (11) *meets the following inequalities:*

$$-2C_\alpha K_{\max} h \le \widehat{S}_N - \sum_{i=1}^{N} \alpha_i \le 4C_\alpha (g_{\max} - 1) K_{\max} h\,. \tag{29}$$

*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 (29).

*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 assumptions* (A1) *and* (B1)–(B3) *are verified. Let estimator* $\widehat{f}_N$ *be defined by LP* (18)–(22). *Moreover, let* $h \to 0$ *as* $N \to \infty$ *such that*

$$\lim_{N\to\infty} \frac{\log N}{Nh} = 0\,. \tag{30}$$

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

$$L_{\widehat{f}_N} \triangleq \max_{x\in[0,1]} |\widehat{f}_N'(x)| \tag{31}$$

$$\le 2L_{f,\beta}\, g_{\max} C_\beta(K, K') \frac{\log N}{Nh^2}\,. \tag{32}$$

*Remark 3.* As it can be seen from the proof of Lemma 2, namely from (A.11), (A.12), one might slightly decrease the number of constraints (15) on the estimator derivative (13)–(17). 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 (13)–(17) or to its equivalent LP version (18)–(22) defines the kernel estimator of the support covering all the points $(X_i, Y_i)_{i=1,\dots,N}$ and, approximately, having the smallest surface, up to the term $O(h)$ specified in Lemma 1. Moreover, constraints (15), (16) or (20), (21) 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 (4), (18)–(22) may be treated as the approximation to Maximum Likelihood Estimate related to the estimation family (4); see [23] for details.

The following two lemmas prove the upper bound for the functional $J_P^*$ and the pointwise lower bound for the estimator $\widehat{f}_N$ itself.

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

$$\gamma > L_{f,\beta}\, g_{\max} C_\beta(K) \tag{33}$$

*and almost all $\omega \in \Omega$ there exist finite numbers $N_1 = N_1(\omega, \gamma)$ such that for all $N \geq N_1$ the LP* (18)–(22) *is solvable, and*

$$J_P^* \leq C_f + \gamma h^\beta\,. \tag{34}$$

**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 $N \geq N_2(\omega)$*

$$\widehat{f}_N(x) \geq f(x) - \frac{C_4(\beta)}{h^2}\left(\frac{\log N}{N}\right)^{\frac{2+\beta}{1+\beta}} \tag{35}$$

*with constant $C_4(\beta)$ defined in* (39).

## 4. MAIN RESULTS

In the following theorem, the consistency and the convergence rate of the described above estimator is established in the sense of the $L_1$-norm for the error on the $[0,1]$ interval.

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

$$\liminf_{N\to\infty} \frac{\log N}{N h^{1+\beta}} > \rho > 0\,, \quad \lim_{N\to\infty} \frac{\log N}{N h^{1+\beta/2}} = 0\,. \tag{36}$$

*Then the estimator $\widehat{f}_N$ being defined by* (4), (5) *via solution to the LP-problem* (18)–(22) *has the following asymptotic properties:*

$$\|\widehat{f}_N - f\|_1 \leq \left(C_{12}(\beta) h^\beta + 2C_4(\beta) h^{-2} (\log N/N)^{\frac{2+\beta}{1+\beta}}\right)(1 + o(1)) \quad \text{a.s.} \tag{37}$$

*with constants*

$$C_{12}(\beta) \triangleq 2L_{f,\beta}\; g_{\max}\, C_\beta(K, K') + 4C_\alpha (g_{\max} - 1) K_{\max} \mathbf{1}\{\beta = 1\} \tag{38}$$

*and*

$$C_4(\beta) \triangleq 2L_{f,\beta}\left[\left(\frac{2C_f}{L_{f,\beta}}\right)^{\frac{\beta}{1+\beta}}\left(\frac{1}{\rho}\right)^{\frac{2}{1+\beta}} + g_{\max} C_\beta(K, K')\left(\frac{2C_f}{L_{f,\beta}}\right)^{\frac{1}{1+\beta}}\right]. \tag{39}$$

**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:*

$$h \sim \widetilde{\rho}\left(\frac{\log N}{N}\right)^{\frac{1}{1+\beta}}, \quad 0 < \widetilde{\rho} < \rho^{-\frac{1}{1+\beta}}, \tag{40}$$

*which reduces the upper bound* (37) *to*

$$\limsup_{N\to\infty}\left(\frac{\log N}{N}\right)^{-\frac{\beta}{1+\beta}} \|\widehat{f}_N - f\|_1 \le C_{12}(\beta)\widetilde{\rho}^{\beta} + 2C_4(\beta)\widetilde{\rho}^{-2} \quad \text{a.s.} \tag{41}$$

Let us highlight that (41) shows that $\widehat{f}_N$ reaches (up to a logarithmic factor) the optimal minimax $L_1$ rate for $\beta$-Hölder frontier $f$, see [7, Theorem 4.1.1].

*Remark 4.* The second condition in (36) may be extended to

$$\lim_{N\to\infty}\frac{\log N}{Nh^{1+\beta/2}} < \infty \tag{42}$$

which leads to another, more general formula for constants in (37)–(39).

The proof of Theorem 1 which is given in the Appendix is based on both upper and lower bounds derived in Lemmas 3 and 4 respectively.

## 5. CONCLUSION

The proposed method of frontier estimation generates estimate as a linear combination of smooth kernel functions, centered in the sample points, which covers all the points and gives a support of a minimal surface. Essentially, method attains $L_1$-optimal rate of convergence $O\left((\log N/N)^{\beta/(1+\beta)}\right)$ on the class of $\beta$-Hölder boundary functions, when the bandwidth parameter $h$ is of order $(\log N/N)^{1/(1+\beta)}$. The optimality became possible due to improvement of the method from [22, 23] by introducing the linear constraints (15), (20) on the estimate regularity into the related LP-problem. Thus, from the practical point of view, the method did not undergo substantial complications, since it reduces to linear program as before.

## *APPENDIX*

**Proof of Lemma 1.** Observe, that definitions (4)–(5) imply the following decomposition:

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

$$= \int_0^1 \sum_{i=1}^N \alpha_i \frac{1}{h} K\left(\frac{x - X_i}{h}\right) dx \tag{A.2}$$

$$+ \sum_{i=1}^N \alpha_i \left(\int_0^h + \int_{1-h}^1\right) \frac{g(x) - 1}{h} K\left(\frac{x - X_i}{h}\right) dx\,. \tag{A.3}$$

Since all $\alpha_i$ and kernel $K$ are non-negative, it follows that

$$\int_0^1 \sum_{i=1}^N \alpha_i \frac{1}{h} K\left(\frac{x - X_i}{h}\right) dx \le \sum_{i=1}^N \alpha_i \int_{\mathbb{R}} \frac{1}{h} K\left(\frac{x - X_i}{h}\right) dx = \sum_{i=1}^N \alpha_i \tag{A.4}$$

and therefore,

$$\int_0^1 \widehat{f}_N(x)\, dx - \sum_{i=1}^N \alpha_i \le \frac{g_{\max} - 1}{h} K_{\max} \sum_{i=1}^N \alpha_i \left(\int_0^h + \int_{1-h}^1\right) \mathbf{1}\{|x - X_i| \le h\} dx \tag{A.5}$$

$$\le (g_{\max} - 1) K_{\max} \left(\sum_{i=1}^N \alpha_i\, \mathbf{1}\{0 \le X_i \le 2h\}\right. \tag{A.6}$$

$$\left. + \sum_{i=1}^N \alpha_i\, \mathbf{1}\{1 - 2h \le X_i \le 1\}\right) \tag{A.7}$$

$$\le (g_{\max} - 1) K_{\max} 4 C_\alpha h\,. \tag{A.8}$$

The inequality (A.8) follows from (16) since both intervals in (A.6), (A.7) 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 (16). Consequently, we have proved the upper bound for the difference in the left hand side (A.5). The lower bound is proved in the same manner. Indeed, decomposition (A.1)–(A.3) implies, since the term (A.3) is non-negative,

$$\int_0^1 \widehat{f}_N(x)\, dx - \sum_{i=1}^N \alpha_i \ge - \sum_{i=1}^N \alpha_i \left(\int_{-h}^0 + \int_1^{1+h}\right) \frac{1}{h} K\left(\frac{x - X_i}{h}\right) dx$$

$$\ge -K_{\max} \left(\sum_{i=1}^N \alpha_i\, \mathbf{1}\{0 \le X_i \le h\}\right.$$

$$\left. + \sum_{i=1}^N \alpha_i\, \mathbf{1}\{1 - h \le X_i \le 1\}\right)$$

$$\ge -K_{\max}\, 2 C_\alpha h\,.$$

This completes the proof of Lemma 1. ■

When proving the following two lemmas, we assumed that sequence of the sample $X$-points $(X_i)_{i=1,\ldots,N}$ is already increase ordered, without changing notation from $X_i$ to $X_{(i)}$ for the sake of simplicity, that is

$$X_i \le X_{i+1}\,, \quad \forall\, i\,. \tag{A.9}$$

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 A.2.

**Proof of Lemma 2.** Thus, we assume (A.9). By applying auxiliary Lemma A.2 and Lemma A.4 we first arrive at

$$\max_{x \in [0,1]} |\widehat{f}_N'(x)| = \max_{1 \le i \le N+1} \max_{x \in [X_{i-1}, X_i]} |\widehat{f}_N'(x)| \tag{A.10}$$

$$\le L_{f,\beta}\, g_{\max} C_\beta(K, K') \frac{\log N}{N h^2} + \frac{1}{8} \max_{1 \le i \le N+1} \left[(X_i - X_{i-1})^2 \max_{x \in [X_{i-1}, X_i]} |\widehat{f}_N'''(x)|\right] \tag{A.11}$$

$$\le L_{f,\beta}\, g_{\max} C_\beta(K, K') \frac{\log N}{N h^2} + \frac{1}{8} \left(C_X \frac{\log N}{N}\right)^2 \max_{x \in [0,1]} |\widehat{f}_N'''(x)|\,, \tag{A.12}$$

with $C_X > 4C_f/f_{\min}$. The maximum term in (A.12) is bounded as follows: for any $x \in [0,1]$

$$|\widehat{f}_N'''(x)| \le \sum_{i=1}^{N} \alpha_i \left| \frac{d^3}{dx^3} K_h(x, X_i) \right| \tag{A.13}$$

$$\le \sup_{u,v} \left| \frac{\partial^3}{\partial v^3} K_h(v, u) \right| \sum_{i=1}^{N} \alpha_i \, \mathbf{1}\{|x - X_i| \le h\} \tag{A.14}$$

$$\le g_{\max} L_{\widetilde{K}''} h^{-4} \, 3C_\alpha h \,, \tag{A.15}$$

since (see Lemma A.1)

$$\sup_{u,v} \left| \frac{\partial^3}{\partial v^3} K_h(v, u) \right| \le g_{\max} L_{\widetilde{K}''} h^{-4} \,, \tag{A.16}$$

where

$$L_{\widetilde{K}''} \triangleq L_{K''} + 3L_{K'} K_{\max} g_{\max} + L_K g_{\max} \left( 3L_K + 10K_{\max}^2 g_{\max} \right) + 6K_{\max}^4 g_{\max}^3 \,. \tag{A.17}$$

Substituting (A.13)–(A.15) into (A.12) yields

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

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

$$h \ge \frac{3C_X^2 C_\alpha \, L_{\widetilde{K}''} \, \log N}{8L_{f,\beta} \, C_\beta(K, K') \, N} \,. \tag{A.18}$$

The desired result follows. ∎

**Proof of Lemma 3.** Consider arbitrary $N \ge N_0(\omega)$ with $N_0(\omega)$ from Lemma A.2. Introduce function $f_\gamma(u) = f(u) + \gamma h^\beta$ and pseudo-estimators

$$\widetilde{\alpha}_i = \frac{1 + \delta_{i1}}{2} \int_{X_{i-1}}^{X_i} f_\gamma(u) \, du + \frac{1 + \delta_{iN}}{2} \int_{X_i}^{X_{i+1}} f_\gamma(u) \, du \,, \quad i = 1, \ldots, N \,, \tag{A.19}$$

where $\delta_{ij}$ stands for Kronecker symbol. Below we demonstrate that condition (33) ensures the vector of pseudo-estimators $\widetilde{\alpha} = (\widetilde{\alpha}_1 \ldots, \widetilde{\alpha}_N)^T$ to be an admissible point for the LP (18)–(22), for any sufficiently large $N$. This implies solvability of the LP (18)–(22) and

$$J_P^* \le \sum_{i=1}^{N} \widetilde{\alpha}_i = \int_0^1 (f(u) + \gamma h^\beta) \, du = C_f + \gamma h^\beta \,. \tag{A.20}$$

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

$$h^\beta \le \frac{\log N}{\rho N h} \le \min\left\{\frac{f_{\max}}{\gamma}, \frac{1}{\rho C_X}\right\}, \tag{A.21}$$

which hold true for all $N$ large enough.

(1) First, we prove constraints (14) under $\alpha_i = \widetilde{\alpha}_i$, $i = 1, \ldots, N$. For arbitrary $x \in [0, 1]$,

$$\widetilde{f}_N(x) \triangleq \sum_{i=1}^{N} \widetilde{\alpha}_i K_h(x, X_i) \tag{A.22}$$

$$= \sum_{i=1}^{N+1} \int_{X_{i-1}}^{X_i} f_\gamma(u)\, du\, \frac{K_h(x, X_i) + K_h(x, X_{i-1})}{2} \tag{A.23}$$

$$+ \frac{1}{2} \int_0^{X_1} f_\gamma(u)\, du\, (K_h(x, X_1) - K_h(x, 0)) \tag{A.24}$$

$$+ \frac{1}{2} \int_{X_N}^{1} f_\gamma(u)\, du\, (K_h(x, X_N) - K_h(x, 1)) \tag{A.25}$$

$$= \int_0^1 f_\gamma(u)\, K_h(x, u)\, du \tag{A.26}$$

$$+ \sum_{i=1}^{N+1} \int_{X_{i-1}}^{X_i} f_\gamma(u) \left( \frac{K_h(x, X_i) + K_h(x, X_{i-1})}{2} - K_h(x, u) \right) du \tag{A.27}$$

$$+ \frac{1}{2} \int_0^{X_1} f_\gamma(u)\, du\, (K_h(x, X_1) - K_h(x, 0)) \tag{A.28}$$

$$+ \frac{1}{2} \int_{X_N}^{1} f_\gamma(u)\, du\, (K_h(x, X_N) - K_h(x, 1))\ . \tag{A.29}$$

Now we separately bound each of the summands (A.26)–(A.29) from below. Due to (6), the main term (A.26) is bounded as follows:

$$\int_0^1 f_\gamma(u)\, K_h(x, u)\, du = f(x) + \gamma h^\beta + \int_0^1 (f(u) - f(x))\, K_h(x, u)\, du \tag{A.30}$$

$$\ge f(x) + (\gamma - L_{f,\beta}\, g_{\max} C_\beta(K)) h^\beta\ . \tag{A.31}$$

Furthermore, the $i$-th summand from (A.27) is decomposed and then bounded basing on trapezium formula error:

$$\int_{X_{i-1}}^{X_i} f_\gamma(u)\left(\frac{K_h(x,X_i)+K_h(x,X_{i-1})}{2}-K_h(x,u)\right)du \tag{A.32}$$

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

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

$$\geq -(f_{\max}+\gamma h^\beta)\frac{(X_i-X_{i-1})^3}{12}\max_{u\in[0,1]}\left|\frac{\partial^2 K_h(x,u)}{\partial u^2}\right|\mathbf{1}\{|x-X_i|\leq 2h\} \tag{A.33}$$

$$-L_{f,\beta}\int_{X_{i-1}}^{X_i}|u-x|^\beta\,\mathbf{1}\{|x-X_i|\leq 2h\}\frac{g_{\max}L_K}{2h^2}\left[(u-X_{i-1})+(X_i-u)\right]du. \tag{A.34}$$

By applying Lemma A.2, the first term is bounded as follows

$$(A.33)\geq -\left(C_X\frac{\log N}{N}\right)^2\frac{f_{\max}\,g_{\max}L_{K'}}{6h^3}(X_i-X_{i-1})\,\mathbf{1}\{|x-X_i|\leq 2h\}; \tag{A.35}$$

and the second one is bounded by:

$$(A.34)\geq -\frac{g_{\max}L_{f,\beta}\,L_K}{2h^2}C_X\frac{\log N}{N}\mathbf{1}\{|x-X_i|\leq 2h\}\int_{X_{i-1}}^{X_i}|u-x|^\beta du\,. \tag{A.36}$$

Moreover, from Lemma A.2, 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

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

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

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

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

Thus, we arrive at the bound for the sum (A.27) as follows:

$$\sum_{i=1}^{N+1}\int\limits_{X_{i-1}}^{X_i} f_\gamma(u)\left(\frac{K_h(x,X_i)+K_h(x,X_{i-1})}{2}-K_h(x,u)\right)du$$
$$\ge -g_{\max}C_X\frac{\log N}{Nh}\left(4+C_X\frac{\log N}{Nh}\right)\left(\frac{C_X f_{\max}L_{K'}}{6}\frac{\log N}{Nh}\right.$$
$$\left.+\frac{L_{f,\beta}\,L_K}{2}h^\beta\left(2+C_X\frac{\log N}{Nh}\right)^\beta\right)$$
$$\ge -\frac{5}{6}g_{\max}C_X\left(\frac{\log N}{Nh}\right)^2\left(C_X f_{\max}L_{K'}+3^{\beta+1}\rho^{-1}L_{f,\beta}\,L_K\right).$$

At last, it is similarly demonstrated that both summands (A.28) and (A.29) are bounded above by $O((\log N/(Nh))^2)$. For instance, for (A.28) we obtain

$$\left|\int\limits_0^{X_1} f_\gamma(u)\,du\,(K_h(x,X_1)-K_h(x,0))\right| \le (f_{\max}+\gamma h^\beta)X_1\,|K_h(x,X_1)-K_h(x,0)|$$
$$\le 2f_{\max}\,g_{\max}L_K\left(C_X\frac{\log N}{Nh}\right)^2. \tag{A.37}$$

Thus, it follows from (A.22)–(A.37) for each $j=1,\ldots,N$, that

$$\widetilde{f}_N(X_j)\ge f(X_j)+(\gamma-L_{f,\beta}\,g_{\max}C_\beta(K))h^\beta+O\left(\left(\frac{\log N}{Nh}\right)^2\right)\ge Y_j \tag{A.38}$$

for sufficiently large $N\ge N_0(\omega)$ when both inequalities (A.21) and the following one hold true:

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

(2) Similarly, constraints (15) hold true under $\alpha_i=\widetilde{\alpha}_i$, $i=1,\ldots,N$. Indeed, for arbitrary $x\in[0,1]$, we now have to bound the absolute value of

$$\widetilde{f}'_N(x)=\sum_{i=1}^N\widetilde{\alpha}_i\,\frac{d}{dx}\,K_h(x,X_i)=\sum_{i=1}^N\widetilde{\alpha}_i\,\widetilde{K}_h(x,X_i) \tag{A.39}$$

instead of (A.22). Here

$$\widetilde{K}_h(x,u)\triangleq\frac{\partial}{\partial x}\,K_h(x,u) \tag{A.40}$$

with the following upper bound

$$\left|\widetilde{K}_h(x,u)\right|\le h^{-2}g_{\max}\left\{\left|K'\left(\frac{x-u}{h}\right)\right|+g_{\max}\,K_{\max}\left|K\left(\frac{x-u}{h}\right)\right|\right\}. \tag{A.41}$$

Hence, one may repeat the arguments of (A.23)–(A.29) by changing $K_h$ for $\widetilde{K}_h$. Therefore, all the rates from (A.32)–(A.38) should be divided by $h$, while the absolute value of the main term of decomposition, due to (7), is bounded as follows:

$$\left|\int\limits_0^1 f_\gamma(u)\,\widetilde{K}_h(x,u)\,du\right|=\left|\int\limits_0^1(f(u)-f(x))\,\frac{\partial}{\partial x}\,K_h(x,u)\,du\right|$$
$$\le L_{f,\beta}\,g_{\max}C_\beta(K,K')h^{\beta-1},$$

instead of (A.30), (A.31). Remind the definition (8), (9) for $C_\beta(K,K')$ which follows from (A.41). Thus, for sufficiently large $N \geq N_0(\omega)$ and for each $X_j$ we arrive at

$$\left|\widetilde{f}_N'(X_j)\right| \leq L_{f,\beta}\, g_{\max} C_\beta(K,K') h^{\beta-1} + O\left(\frac{\log^2 N}{N^2 h^3}\right) \leq L_{f,\beta}\, g_{\max} C_\beta(K,K') \frac{\log N}{\rho N h^2}. \quad \text{(A.42)}$$

Namely, inequality (A.42) holds true almost surely for all those $N \geq N_0(\omega)$ such that (A.21) is verified and

$$L_{f,\beta}\, C_\beta(K,K')\left(\frac{\log N}{h^{\beta+1}N} - 1\right) \geq \frac{5}{6}\, g_{\max} C_X \left(\frac{\log N}{N h^{1+\beta/2}}\right)^2 \quad \text{(A.43)}$$

$$\times \left(C_X f_{\max}\left(L_{\widetilde{K}'} + \frac{12 L_{\widetilde{K}}}{5}\right) + 3^{\beta+1}\frac{L_{f,\beta}\, L_{\widetilde{K}}}{\rho}\right), \quad \text{(A.44)}$$

where (see proof of Lemma A.1 for the details)

$$L_{\widetilde{K}} \triangleq L_{K'} + L_K\, g_{\max} K_{\max}\,, \quad L_{\widetilde{K}'} \triangleq L_{K''} + L_{K'}\, g_{\max} K_{\max}\,. \quad \text{(A.45)}$$

(3) Finally, the constraints (16) with

$$C_\alpha \geq 6 f_{\max} \quad \text{(A.46)}$$

also hold true under $\alpha_i = \widetilde{\alpha}_i$, $i = 1, \ldots, N$. Indeed, by Lemma A.2 the following inequalities hold a.s. for all $N \geq N_0(\omega)$ and for each $j = 1, \ldots, m_h$, where $m_h = \lfloor h^{-1} \rfloor$ :

$$\sum_{i=1}^{N} \widetilde{\alpha}_i\, \mathbf{1}\{(j-1)/m_h \leq X_i < j/m_h\} \;\leq\; (f_{\max} + \gamma h^\beta)\left(1/m_h + 2C_X \frac{\log N}{N}\right) \quad \text{(A.47)}$$

$$\leq\; 6 f_{\max} h, \quad \text{(A.48)}$$

under additional assumptions (A.21). Thus, constraints (16) are fulfilled under (A.46) almost surely, for any sufficiently large $N$.

(4) Since all $\widetilde{\alpha}_i \geq 0$, constraints (17) hold true, and Lemma 3 is proved. ■

**Proof of Lemma 4.** Let us take use of Lemma A.3 and its Corollary A.1 introducing

$$\delta_y = L_{f,\beta}\, \delta_x^\beta, \quad \delta_x \triangleq \left(\frac{2C_f \log N}{f_{\min} L_{f,\beta}\, N}\right)^{\frac{1}{1+\beta}}. \quad \text{(A.49)}$$

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

$$|x - X_{i_k}| \leq \delta_x \quad \text{(A.50)}$$

and

$$Y_{i_k} \geq f(X_{i_k}) - \delta_y\,. \quad \text{(A.51)}$$

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

$$f(x) - \widehat{f}_N(x) = [f(x) - f(X_{i_k})] \quad \text{(A.52)}$$

$$+ \left[f(X_{i_k}) - \widehat{f}_N(X_{i_k})\right] \quad \text{(A.53)}$$

$$+ \left[\widehat{f}_N(X_{i_k}) - \widehat{f}_N(x)\right]. \quad \text{(A.54)}$$

The term in the right hand side (A.52) may be bounded as follows

$$|f(x) - f(X_{i_k})| \le L_{f,\beta}\, |x - X_{i_k}|^\beta \le L_{f,\beta}\, \delta_x^\beta\,, \tag{A.55}$$

as well as the term (A.54)

$$\left|\widehat{f}_N(X_{i_k}) - \widehat{f}_N(x)\right| \le L_{\widehat{f}_N}\, |x - X_{i_k}| \le L_{\widehat{f}_N}\, \delta_x, \tag{A.56}$$

with a Lipschitz constant $L_{\widehat{f}_N}$ for the function estimator $\widehat{f}_N(x)$. Remind that $\widehat{f}_N(X_{i_k}) \ge Y_{i_k}$ due to (14) or (19). Thus, (A.51) implies

$$f(X_{i_k}) - \widehat{f}_N(X_{i_k}) \le (Y_{i_k} + \delta_y) - Y_{i_k} = \delta_y\,. \tag{A.57}$$

Combining all these bounds we obtain from (A.52) that for all $N \ge N_6(\omega)$,

$$f(x) - \widehat{f}_N(x) \le \delta_y + L_{f,\beta}\, \delta_x^\beta + L_{\widehat{f}_N}\delta_x\,. \tag{A.58}$$

Therefore, applying Lemma 2 and substituting expressions (A.49) for $\delta_x$ and $\delta_y$ into (A.58) lead to the lower bound

$$\widehat{f}_N(x) \ge f(x) - \left(2L_{f,\beta}\, \delta_x^\beta + L_{\widehat{f}_N}\delta_x\right) \tag{A.59}$$

$$\ge f(x) - \frac{C_4(\beta)}{h^2}\left(\frac{\log N}{N}\right)^{\frac{2+\beta}{1+\beta}} \tag{A.60}$$

for any sufficiently large $N$ (starting from random a.s. finite integer, which does not depend on $x$). The first inequality in (36) has been applied here in order to simplify the lower bound. Lemma 4 is proved. ■

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

$$\|\widehat{f}_N - f\|_1 = \int_0^1 \left[\widehat{f}_N(x) - f(x)\right] dx \tag{A.61}$$

$$+2\int_0^1 \left[f(x) - \widehat{f}_N(x)\right] \mathbf{1}\left\{\widehat{f}_N(x) < f(x)\right\} dx. \tag{A.62}$$

Applying Lemmas 1 and 2 to the right hand side (A.61) yields

$$\limsup_{N\to\infty} h^{-\beta}\left(\int_0^1 \left[\widehat{f}_N(x) - f(x)\right] dx\right) \le \gamma + 4C_\alpha(g_{\max} - 1)K_{\max}\mathbf{1}\{\beta = 1\} \quad \text{a.s.} \tag{A.63}$$

Note, that one may fix $\gamma = 2L_{f,\beta}\, g_{\max}\, C_\beta(K)$, for instance. In order to obtain a similar result for the term (A.62), 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

$$\varepsilon_{LB}(N) \triangleq \frac{1}{h^2}\left(\frac{\log N}{N}\right)^{\frac{2+\beta}{1+\beta}}. \tag{A.64}$$

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

$$\limsup_{N\to\infty} \varepsilon_{LB}^{-1}(N) \int\limits_0^1 \left[f(x) - \widehat{f}_N(x)\right] \mathbf{1}\left\{\widehat{f}_N(x) < f(x)\right\} dx$$

$$\leq \int\limits_0^1 \limsup_{N\to\infty} \zeta_N(x,\omega)\, \mathbf{1}\{\zeta_N(x,\omega) > 0\}\, dx$$

$$\leq C_4(\beta) < \infty \quad \text{a.s.}$$

Thus, the obtained relations together with (A.61) and (A.62) imply (37). Theorem 1 is proved. ■

The following lemma establishes Lipschitz constants in (A.45) and (A.16), (A.17). We do not provide its proof which is of technical nature and specified in [24].

**Lemma A.1.** *Let kernel $K_h$ defined in* (4), (5) *meets the assumptions* (B1)–(B3), *and the bandwidth $h \in (0, 1/2)$. Let $\widetilde{K}_n$ be defined by* (A.40). *Then the following upper bounds hold true:*

$$\left|\widetilde{K}_h(x,u)\right| \leq g_{\max} h^{-2} \left(L_K + g_{\max} K_{\max}^2\right), \tag{A.65}$$

$$\left|\frac{\partial}{\partial u}\,\widetilde{K}_h(x,u)\right| \leq g_{\max} h^{-3} L_{\widetilde{K}}, \qquad \left|\frac{\partial^2}{\partial u^2}\,\widetilde{K}_h(x,u)\right| \leq g_{\max} h^{-4} L_{\widetilde{K}'}, \tag{A.66}$$

*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,*

$$\left|\frac{\partial^3}{\partial x^3}\, K_h(x,u)\right| \leq g_{\max} L_{\widetilde{K}''} h^{-4}, \tag{A.67}$$

*where*

$$L_{\widetilde{K}''} = g_{\max} \left[L_{K''} + 3L_{K'} K_{\max} g_{\max} + 3L_K g_{\max} K_{\max}^2 (1 + 3g_{\max})\right. \tag{A.68}$$

$$\left. + (L_K^2 + 2g_{\max}^2 K_{\max}^4)(1 + 2g_{\max})\right]. \tag{A.69}$$

The following auxiliary results are briefly proved below for the sake of completeness.

**Lemma A.2.** *Let function $f : [0,1] \to (0,+\infty)$ 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* (A.9), *where $C_f$ is defined by* (2). *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*

$$\max_{i=1,\ldots,N+1} \Delta X_i \leq C_X \frac{\log N}{N} \quad \forall\, N \geq N_0 \tag{A.70}$$

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

$$C_X = 5 f_{\max}/f_{\min}. \tag{A.71}$$

**Proof of Lemma A.2.** Introduce an equidistant partition of the interval $[0,1]$ onto $m_N$ subintervals $\Delta_k$ with equal Lebesgue measures

$$\ell(\Delta_k) \triangleq 1/m_N \leq C_X \log N/(2N), \quad k = 1, \ldots, m_N, \tag{A.72}$$

where size of partition

$$m_N \triangleq \min\{\text{integer } m \;:\; m \geq 2N/(C_X \log N)\} \tag{A.73}$$

$$\leq 1 + \frac{2N}{C_X \log N} \leq \frac{(2+\varepsilon)N}{C_X \log N} \tag{A.74}$$

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

$$A_N \;\triangleq\; \{\omega \;:\; \max_{i=1,\ldots,N+1} \Delta X_i \leq C_X \log N/N\} \;\supseteq\; \bigcap_{k=1}^{m_N} \left[ \bigcup_{i=1}^{N} \{X_i \in \Delta_k\} \right].$$

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,

$$P\left(A_N^c\right) \leq \sum_{k=1}^{m_N} P\left( \bigcap_{i=1}^{N} \{X_i \notin \Delta_k\} \right) \tag{A.75}$$

$$= \sum_{k=1}^{m_N} \prod_{i=1}^{N} \left( 1 - \int\limits_{\Delta_k} C_f^{-1} f(u)\, du \right) \tag{A.76}$$

$$\leq m_N \left( 1 - \frac{f_{\min}}{C_f}\, \ell(\Delta_1) \right)^N \tag{A.77}$$

$$\leq \frac{(2+\varepsilon)N}{C_X \log N} \exp\left\{ -\frac{f_{\min} C_X}{(2+\varepsilon) C_f} \log N \right\} \tag{A.78}$$

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

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

$$\sum_{N=1}^{\infty} P\left(A_N^c\right) < \infty\,, \tag{A.79}$$

and the Borel–Cantelli lemma applies. Note, that events $\bigcap_{i=1}^{N} \{X_i \notin \Delta_k\}$ do not depend on renumbering of $(X_i)_{i=1,\ldots,N}$ which lead to (A.76) from (A.75); moreover, we have used both definition (A.73) and inequality $1 - x \leq e^{-x}$ there in (A.77), (A.78). Lemma A.2 is proved. ∎

**Lemma A.3.** *Let random sample $\{(X_i, Y_i) \,|\, i = 1, \ldots, N\}$ be defined as in Section 2. Let sequence $\delta_x = \delta_x(N)$ be positive, and for some $\varepsilon > 0$*

$$\liminf_{N \to \infty} N^{1-\varepsilon} \delta_x > 0\,. \tag{A.80}$$

*Define*

$$m_\delta \triangleq \min\{\textit{integer}\, m \;:\; m \geq \delta_x^{-1}\} \tag{A.81}$$

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

$$\delta_y \geq \kappa\, m_\delta \frac{\log N}{N}\,, \quad \textit{where} \quad \kappa > \frac{(2-\varepsilon)C_f}{f_{\min}}\,. \tag{A.82}$$

*Then, under the assumptions of Lemma A.2, there exists almost surely finite number $N_6(\omega)$ such that for any $N \geq N_6(\omega)$ there is such a subset of points $\{(X_{i_k}, Y_{i_k})\,,\, k = 1, \ldots, m_\delta\}$ in the sample $\{(X_i, Y_i)\,,\;\; i = 1, \ldots, N\}$, that the following inequalities hold true:*

$$(k-1)/m_\delta \leq X_{i_k} < k/m_\delta\,, \quad f(X_{i_k}) - \delta_y \leq Y_{i_k} \leq f(X_{i_k})\,. \tag{A.83}$$

**Proof of Lemma A.3** is similar to that of Lemma A.2. Introduce an equidistant partition of the interval $[0, 1]$ onto subintervals $[(k-1)/m_\delta, k/m_\delta]$, $k = 1, \dots, m_\delta$. Moreover, introduce the related subsets in $\mathbb{R}^2$

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

Consider the event

$$A_N \triangleq \{\omega\,:\, \forall\, k = 1, \dots, m_\delta\ \exists\, i = 1, \dots, N : (X_i, Y_i) \in \Delta_k\} = \bigcap_{k=1}^{m_\delta}\left[\bigcup_{i=1}^{N}\{(X_i, Y_i) \in \Delta_k\}\right].$$

Bound the probability of the complementary event:

$$\begin{aligned} P\left(A_N^c\right) &\le \sum_{k=1}^{m_\delta} P\left(\bigcap_{i=1}^{N}\{(X_i, Y_i) \notin \Delta_k\}\right) = \sum_{k=1}^{m_\delta}\prod_{i=1}^{N}\left(1 - \int\limits_{\Delta_k} C_f^{-1} f(u)\,du\,dv\right) \\ &\le m_\delta\left(1 - \frac{f_{\min}\,\delta_y}{C_f\,m_\delta}\right)^N \le \left(1 + \delta_x^{-1}\right)\,\exp\left\{-\frac{f_{\min}\kappa}{C_f}\,\log N\right\} \\ &= O\left(N^{1-\varepsilon-f_{\min}\kappa/C_f}\right). \end{aligned}$$

Hence, condition $\kappa > (2-\varepsilon)C_f/f_{\min}$ implies $\sum_{N=1}^{\infty} P\left(A_N^c\right) < \infty$, and one may apply Borel–Cantelli lemma. Lemma A.3 is proved. ■

**Corollary A.1.** *Let $\delta_x$ and $\delta_y$ meet the conditions of Lemma A.3. Then, with probability 1, for any $N \ge N_6(\omega)$ and any $x \in [0, 1]$ there exists integer $i_k \in \{1, \dots, N\}$ 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 A.4.** *Let function $g : [0, \Delta] \to \mathbb{R}$ be twice continuous differentiable, $\Delta > 0$. Then*

$$\max_{x\in[0,\Delta]} |g(x)| \le \max\{|g(0)|, |g(\Delta)|\} + \frac{\Delta^2}{8}\max_{x\in[0,\Delta]} |g''(x)|\,. \tag{A.84}$$

**Proof of Lemma A.4.** 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

$$|g(x_1)| = \max_{x\in[0,\Delta]} |g(x)| > \bar{g}_b\,. \tag{A.85}$$

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

$$g(x_1) = g(x) - \int\limits_{x_1}^{x} dt \int\limits_{x_1}^{t} g''(u)\,du\,. \tag{A.86}$$

Therefore, putting $x = \Delta$ one obtains from (A.86)

$$|g(x_1)| \le |g(\Delta)| + \int\limits_{x_1}^{\Delta} dt \int\limits_{x_1}^{t} |g''(u)|\,du \le \bar{g}_b + \frac{(\Delta - x_1)^2}{2}\max_{x\in[0,\Delta]} |g''(x)|\,. \tag{A.87}$$

Similarly, fixing $x = 0$ there in (A.86) leads to

$$|g(x_1)| \le |g(0)| + \int\limits_{0}^{x_1} dt \int\limits_{x_1}^{t} |g''(u)|\,du \le \bar{g}_b + \frac{x_1^2}{2}\max_{x\in[0,\Delta]} |g''(x)|\,. \tag{A.88}$$

Thus, combining (A.87) and (A.88) we arrive at

$$|g(x_1)| \leq \bar{g}_b + \frac{1}{2}\min\{(\Delta - x_1)^2, x_1^2\} \max_{x\in[0,\Delta]} |g''(x)|. \tag{A.89}$$

Since

$$\max_{x\in[0,\Delta]} \min\{(\Delta - x)^2, x^2\} = \frac{\Delta^2}{4}, \tag{A.90}$$

the desired inequality (A.84) follows immediately from (A.85), (A.89), (A.90). ■

## REFERENCES


1. Hardy, A. and Rasson, J.P., Une nouvelle approche des problèmes de classification automatique, *Stat. Anal. Données*, 1982, vol. 7, pp. 41–56.
2. Hartigan, J.A., *Clustering Algorithm*, Chichester: Wiley, 1975.
3. Baufays, P. and Rasson, J.P., A New Geometric Discriminant Rule, *Comput. Stat. Quaterly*, 1985, vol. 2, pp. 15–30.
4. Devroye, L.P. and Wise, G.L., Detection of Abnormal Behavior via Non Parametric Estimation of the Support, *SIAM J. Appl. Math.*, 1980, vol. 38, pp. 448–480.
5. Tarassenko, L., Hayton, P., Cerneaz, N., and Brady, M., Novelty Detection for the Identification of Masses in Mammograms, *Proc. 4th IEE Int. Conf. Artificial Neural Networks*, Cambridge, 1995, pp. 442–447.
6. Korostelev, A.P. and Tsybakov, A.B., *Minimax Theory of Image Reconstruction, Lect. Notes Statist.*, New York: Springer-Verlag, 1993, vol. 82.
7. Deprins, D., Simar, L., and Tulkens, H., Measuring Labor Efficiency in Post Offices, in *The Performance of Public Enterprises: Concepts and Measurements*, Marchand, M., Pestieau, P., and Tulkens, H., Eds., Amsterdam: North-Holland, 1984, pp. 243-267.
8. Härdle, W., Hall, P., and Simar, L., Iterated Bootstrap with Application to Frontier Models, *J. Productiv. Anal.*, 1995, vol. 6, pp. 63–76.
9. Geffroy, J., Sur un problème d'estimation géométrique, in *Publications de l'Institut de Stat., Université de Paris*, 1964, vol. XIII, pp. 191–200.
10. Härdle, W., Park, B.U., and Tsybakov, A.B., Estimation of a Non Sharp Support Boundaries, *J. Multivariate Anal.*, 1995, vol. 43, pp. 205–218.
11. Hall, P., Nussbaum, M., and Stern, S.E., On the Estimation of a Support Curve of Indeterminate Sharpness, *J. Multiv. Anal.*, 1997, vol. 62, pp. 204–232.
12. Gijbels, I. and Peng, L., Estimation of a Support Curve via Order Statistics, *Extremes*, 2000, vol. 3, pp. 251–277.
13. Knight, K., Limiting Distributions of Linear Programming Estimators, *Extremes*, 2001, vol. 4, no. 2, pp. 87–103.
14. Hall, P., Park, B.U., and Stern, S.E., On Polynomial Estimators of Frontiers and Boundaries, *J. Multiv. Anal.*, 1998, vol. 66, pp. 71–98.
15. Girard, S. and Jacob, P., Extreme Values and Kernel Estimates of Point Processes Boundaries, *ESAIM: Prob. Stat.*, 2004, vol. 8, pp. 150–168.
16. Girard, S. and Jacob, P., Extreme Values and Haar Series Estimates of Point Processes Boundaries, *Scandinav. J. Stat.*, 2003, vol. 30, no. 2, pp. 369–384.
17. Girard, S. and Jacob, P., Projection Estimates of Point Processes Boundaries, *J. Stat. Planning Inference*, 2003, vol. 116, no. 1, pp. 1–15.

18. Gardes, L., Estimating the Support of a Poisson Process via the Faber-Shauder Basis and Extreme Values, in *Publications de l'Institut de Stat., Université de Paris*, 2002, vol. XXXXVI, pp. 43–72.

19. Girard, S. and Menneteau, L., Central Limit Theorems for Smoothed Extreme Value Estimates of Poisson Point Processes Boundaries, *J. Stat. Planning Inference*, 2005, vol. 135, no. 2, pp. 433–460.

20. Abbar, H., Un estimateur spline du contour d'une répartition ponctuelle aléatoire, *Stat. Anal. Données*, 1990, vol. 15, no. 3, pp. 1–19.

21. Jacob, P. and Suquet, P., Estimating the Edge of a Poisson Process by Orthogonal Series, *J. Stat. Planning Inference*, 1995, vol. 46, pp. 215–234.

22. Bouchard, G., Girard, S., Iouditski, A., and Nazin, A., Linear Programming Problems for Frontier Estimation, *Technical Report INRIA RR-4717*, 2003 (`http://www.inria.fr/rrrt/rr-4717.html`); *Technical Report IAP RT-0304* ( `http://www.stat.ucl.ac.be/Iapdp/tr2003/TR0304.ps`).

23. Bouchard, G., Girard, S., Iouditski, A.B., and Nazin, A.V., Nonparametric Frontier Estimation by Linear Programming, *Avtom. Telemekh.*, 2004, no. 1, pp. 66–73.

24. Girard, S., Iouditski, A., and Nazin, A., Linear Programming Problems for $L_1$-Frontier Estimation, *Technical Report INRIA RR-5466,* 2005 (`http://www.inria.fr/rrrt/rr-5466.html`); *Technical Report IAP RT-0506* (`http://www.stat.ucl.ac.be/ISpub/tr/2005/TR0506.pdf`).