Extreme values and Haar series estimates of point processes boundaries

Stéphane Girard & Pierre Jacob

Laboratoire de Probabilités et Statistique, Université Montpellier 2

Running headline: Haar estimates of boundaries

Abstract

We present a new method for estimating the edge of a two-dimensional bounded set, given a finite random set of points drawn from the interior. The estimator is based both on Haar series and extreme values of the point process. We give conditions for various kind of convergence and we obtain remarkably different possible limit distributions. We propose a method of reducing the negative bias, illustrated by a simulation.

Keywords: Haar basis, Extreme values, Poisson process, Shape estimation.

AMS Subject Classification: Primary 60G70; Secondary 62M30, 62G05, 62G20.

1 Introduction

Many proposals are given in the literature for estimating a bounded set S of $\mathbb{R}^2$, given a finite random set N of points drawn from the interior. Such a diversity follows from crossing properties of the observed random set N (sample, point process, random field on a grid), properties of the unknown bounded set S (convex sets, star-shaped domains, pieces of support under a given curve, images), with the relevant treatments (convex hulls, extreme values, functional estimates, change-point estimates, segmentation).

Here N is a Poisson point process, with observed points belonging to a subset $S = \{(x, y) : 0 \leq x \leq 1 \; ; \; 0 \leq y \leq f(x)\}$, where f is an unknown function, so that the problem reduces to estimating f. A simulation of such a point process is presented Figure 1, and the corresponding unknown function is drawn on Figure 2. This example will be treated in Section 6.

(Figures 1 and 2 here)

This problem arises for instance in econometrics where the function f is called the production frontier (see Härdle *et al* (1995b)). The data consist of pairs (X_i, Y_i) where X_i represents the input (labor, energy or capital) used to produce an output Y_i in a given firm i. 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 (1964) for samples where the estimator was a kind of histogram based on extreme values. More precisely, taking the equidistant partition of $[0,1]$ into k_n intervals, this estimator $\hat{g}_n$ is defined by choosing on each interval the uppest observed point. Some developments were achieved by Chevalier (1976), Gensbittel (1979), and Jacob (1984). Estimating a smooth star-shaped contour is a closely related matter which was adressed by Jacob and Abbar (1989), Abbar (1990) and de Haan and Resnick (1994) with extreme values methods. In Deprins *et al* (1984) and more recently, in Korostelev (1995), f is from economical considerations supposed to be increasing and convex which suggests an adapted estimator, called the DEA (Data Envelopment Analysis) estimator. Asymptotic minimax optimality of piecewise polynomials estimates is studied by Korostelev and Tsybakov (1994), Mammem and Tsybakov (1995), Gayraud (1997) and Härdle *et al* (1995a). Let us note that a precise knowledge of the regularity of the function to estimate is a necessary condition to obtain this optimality. If it is only known that f is a C^1 function, this estimator does not differ from the estimator $\hat{g}_n$ of Geffroy.

The first attempt in the direction of orthogonal series was done by Abbar and Suquet (1993), and more thoroughly by Jacob and Suquet (1995). The method, due to Cencov (1962) and Tiago de Oliveira (1963), consists of estimating the first coefficients of a truncated L^2- expansion f_n of f. The case of the Haar basis, the general case of C^1 bases, and the example of the trigonometric basis were investigated, in the case of Poisson process with a known intensity c, permitting the use of averaging estimates of the Fourier coefficients. In the present paper, the intensity is supposed to be unknown, and we present extreme-valued estimates of these coefficients in the case of the Haar basis. Then, our estimates $\hat{f}_n$ will appear as a linear combination of a set of k_n extreme values, involving the Dirichlet kernel associated to the $(h_n + 1)$ first Haar functions. It is convenient to consider k_n as an integer multiple of $(h_n + 1)$. In the case $(h_n + 1) = k_n$, $\hat{f}_n$ turns out to be simply the estimator $\hat{g}_n$ of Geffroy. In the general case $k_n = d_n(h_n + 1)$, with $d_n \to \infty$. This is one of the advantages of $\hat{f}_n$ to be simple to compute. Moreover, it does not require neither the knowledge of the intensity of the point process nor information on the regularity or geometry of the function.

Of course, it is understood in the sequel that $h_n \to \infty$ so that $k_n \to \infty$, whatever the choice of d_n. In sections 3 and 4, we produce mild conditions for the convergence of $\hat{f}_n$, in terms of local error, mean integrated square error, and for the uniform norm. In subsection 4.1, the choice of $d_n = 1$ or $d_n \to \infty$ appears mainly as a conflict between the systematic part of the bias, $(f - f_n)$, and the variance $\mathrm{Var}\, \hat{f}_n$, while the statistical part of the bias, $(\mathrm{E}\, \hat{f}_n - f_n)$, remains unaffected. The choice $d_n = 1$ privileges the systematic bias, but the choice $d_n \to \infty$ improves the variance. In the case $d_n = 1$, the limit distribution of $\hat{f}_n$ is extremal, while in the case $d_n \to \infty$, $\hat{f}_n$ is asymptotically normal, with a better speed.

In fact, the most awkward part of the bias, in either method, is a negative component of the statistical part $(\mathrm{E}\, \hat{f}_n - f_n)$, which is of order $k_n/(nc)$. Recall that c is assumed to be unknown. In section 5, we present a very simple and efficient method of eliminating this component. The resulting improvement is quite evident on simulations. What is better, we obtain in this way the asymptotical distribution for $(\hat{f}_n - f)$ instead of $(\hat{f}_n - \mathrm{E}\, \hat{f}_n)$.

2 Preliminaries

Let N be a stationary Poisson point process on $\mathbb{R}^2$ with an unknown intensity rate $c > 0$. Actually, suppose that we merely observe the truncated point process $N(. \cap S)$ where

$$S = \{(x, y) \in \mathbb{R}^2 \mid 0 \le x \le 1 \; ; \; 0 \le y \le f(x)\}, \tag{1}$$

and where f is a measurable function such that

$$0 < m = \inf_{[0,1]} f \leq M = \sup_{[0,1]} f < +\infty. \tag{2}$$

For each positive integer i, let us denote by J_i the interval $J_i = \left[\frac{p_i}{2^{q_i-1}}, \frac{p_i+1}{2^{q_i-1}}\right)$ where p_i and q_i are the integers uniquely determined by $i = 2^{q_i-1} + p_i$ and $0 \leq p_i < 2^{q_i-1}$. For $i = 2^{q_i} - 1$, let us close J_i on the right. In what follows we suppose that $h_n + 1 = 2^{h'_n}$, $h'_n \in \mathbb{N}$, so that $\{J_\ell, h_n+1 \leq \ell \leq 2h_n+1\}$ appears as a partition of the unit interval into (h_n+1) equidistant intervals. For each $x \in [0,1]$, denote $J_\ell(x)$ the interval that contains x. Let (k_n) be a further increasing sequence of integers such that $k_n = d_n(h_n+1)$, where $d_n \in \mathbb{N}$, $d_n > 0$. Take the equidistant partition of the unit interval into k_n intervals $I_{n,r}$, $r = 1, \ldots, k_n$ centered at $x_r = (2r-1)/(2k_n)$, and the associated partition of S into k_n cells $D_{n,r}$ with a basis $I_{n,r}$:

$$D_{n,r} = \{\ (x,y) \in S \mid x \in I_{n,r}\ \}. \tag{3}$$

Then, writing $\lambda_{n,r}$ for the Lebesgue measure $\lambda(D_{n,r})$, the piecewise constant function

$$f_n(x) = (h_n+1) \sum_{x_r \in J_\ell(x)} \lambda_{n,r}, \qquad x \in [0,1] \tag{4}$$

appears as an approximation of f.

Let N^{*n} denote the sum of independent Poisson processes with the same intensity rate c. N^{*n} can be considered as a Poisson process with mean measure $nc\lambda$. For $r = 1, \ldots, k_n$, let $X^\star_{n,r} = 0$ if $N^{*n}(D_{n,r}) = 0$. If $N^{*n}(D_{n,r}) > 0$, let $X^\star_{n,r}$ denote the supremum of the second coordinate of the points of the truncated process $N^{*n}(. \cap D_{n,r})$. We estimate each $\lambda_{n,r}$ by $X^\star_{n,r}/k_n$, thus obtaining an estimator $\hat{f}_n$ of f_n which is

$$\hat{f}_n(x) = \frac{1}{d_n} \sum_{x_r \in J_\ell(x)} X^\star_{n,r}, \qquad x \in [0,1]. \tag{5}$$

For the sake of brevity, we write, for $r = 1, \ldots, k_n$, $Y_{n,r} = (X^\star_{n,r}/k_n) - \lambda_{n,r}$, obtaining the reduced formula

$$\hat{f}_n(x) - f_n(x) = (h_n+1) \sum_{x_r \in J_\ell(x)} Y_{n,r}, \qquad x \in [0,1]. \tag{6}$$

In a first reading of our paper, this matter-of-fact presentation of $\hat{f}_n$ is sufficient. However, the deep motivation of this definition is that $\hat{f}_n$ belongs to the important family of orthogonal series estimators. Moreover, in most proofs, $\hat{f}_n$ is easy to handle in this setting.

As a square integrable function on $[0,1]$, f admits a L^2-convergent series expansion with respect to every orthonormal basis of $L^2([0,1])$. In particular, the Haar basis is defined by

$$e_0 = \mathbf{1}_{[0,1]}, \qquad e_i = 2^{\frac{q_i-1}{2}} \left(\mathbf{1}_{J_{2i}} - \mathbf{1}_{J_{2i+1}} \right), \quad i \geq 1, \tag{7}$$

see Figure 3 for an illustration.

(Figure 3 here)

The Dirichlet kernel of order h_n associated to the Haar basis is given by:

$$K_n(x,y) = \sum_{i=0}^{h_n} e_i(x) e_i(y) = (h_n+1) \mathbf{1}_{J_\ell(x)}(y). \tag{8}$$

Now, let $\sum_{i=0}^{\infty} a_i e_i$ be the L^2-expansion of f, where $a_i = \int_0^1 e_i(t) f(t) dt$. The truncated expansion $\sum_{i=0}^{h_n} a_i e_i(x)$, $x \in [0,1]$ can be rewritten in terms of the Dirichlet kernel as $\int_0^1 K_n(x,y) f(y) dy$. In view of (4) and (8), we obtain:

$$\sum_{i=0}^{h_n} a_i e_i(x) = \int_0^1 K_n(x,y) f(y) dy = (h_n+1) \int_{J_\ell(x)} f(y) dy = f_n(x), \ x \in [0,1]. \tag{9}$$

Note also that $\hat{f}_n(x)$ can be written on the Haar basis as

$$\hat{f}_n(x) = \sum_{i=0}^{h_n} \hat{a}_{i,k_n} e_i(x), \qquad x \in [0,1], \tag{10}$$

where $\hat{a}_{i,k_n}$ is an estimate of the coefficient a_i defined by

$$\hat{a}_{i,k_n} = \sum_{r=1}^{k_n} e_i(x_r) \frac{X^\star_{n,r}}{k_n}, \qquad 0 \leq i \leq h_n. \tag{11}$$

A straightforward calculation leads to the formula

$$\hat{f}_n(x) = \sum_{r=1}^{k_n} K_n(x_r, x) \frac{X^\star_{n,r}}{k_n}, \qquad x \in [0,1], \tag{12}$$

which provides a general expression for orthogonal series estimates of f. Before proceeding, we collect some further notations to be used in the sequel:

$$\min_{x\in I_{n,r}} f(x) = m_{n,r}, \ \sup_{x\in I_{n,r}} f(x) = M_{n,r}.$$

3 Systematic bias convergence

According to section 2, $\hat{f}_n$ is an estimate of the approximation f_n of the unknown function f. Thus, a preliminary examination of the convergence of the systematic bias $(f_n - f)$ is required. We need the modulus of continuity of f:

$$\omega(f,\delta) = \sup\{|f(x+\eta) - f(x)|\,;\ 0 \leq x \leq x+\eta \leq 1;\ 0 \leq \eta < \delta\}.$$

Then,

$$|f_n(x) - f(x)| \leq (h_n+1)\int_{J_\ell(x)} |f(x) - f(y)|\,dy \leq \omega(f, 1/(h_n+1)). \tag{13}$$

In view of the last inequality, the following proposition is clear:

Proposition 1 *If f is α-Lipschitzian $(0 < \alpha \leq 1)$, then $\| f_n - f \|_\infty = O\left(h_n^{-\alpha}\right)$.*

The convergence of $\| f_n - f \|_2$ is easily deduced from Proposition 1 since $\| f_n - f \|_2^2 \leq \| f_n - f \|_\infty^2$.

Proposition 2 *If f is α-Lipschitzian $(0 < \alpha \leq 1)$, then $\| f_n - f \|_2^2 = O\left(h_n^{-2\alpha}\right)$.*

Remark that, from (9), $(f_n - f)$ is merely the remainder of the partial sum of order h_n of the Haar series of f, so that $h_n = o(n)$ is a sufficient condition for both the L^2-convergence and the uniform convergence when f is continuous.

4 Consistency results

For $r = 1, \ldots, k_n$, the truncated Poisson point processes $N^{*n}(. \cap D_{n,r})$ are independent, and the extreme values $X^\star_{n,r}$ inherit this property. Thus, the distribution of the estimator $\hat{f}_n$ is uniquely determined by the distribution of each $X^\star_{n,r}$. We refer to the Appendix for a series of lemma describing these distributions.

4.1 Local convergence

Theorem 1 *Suppose f satisfies a Lipschitz condition of order α, $(0 < \alpha \leq 1)$. Then, under the condition $k_n = o\left(n/\ln n\right)$, we have for all $x \in [0,1]$:*

$$\left| E\,\hat{f}_n(x) - f_n(x) + \frac{k_n}{nc} \right| = O\left(\frac{1}{k_n^\alpha}\right). \tag{14}$$

If, moreover, $n = o\left(k_n^{1+\alpha}\right)$ then

$$Var\,\hat{f}_n(x) \sim \frac{k_n h_n}{n^2 c^2}. \tag{15}$$

The condition $n = o\left(k_n^{1+\alpha}\right)$ is necessary to control the variance of $X^\star_{n,r}$ (see Lemma 2 (62)). It imposes that, for all $r = 1, \ldots, k_n$, the mean number of points above $m_{n,r}$ in $D_{n,r}$ converges to 0. It is then possible to ignore the contribution of the events $\{X^\star_{n,r} > m_{n,r}\}$ to the variance.

Proof : In view of Lemma 2 (61),

$$\max_{1 \leq r \leq k_n} \left| \mathrm{E}\left(X^\star_{n,r}\right) - m_{n,r} + \frac{k_n}{nc} \right| = O\left(\frac{1}{k_n^\alpha}\right). \tag{16}$$

Thus, since f is α-Lipschitzian,

$$\left| \mathrm{E}\,\hat{f}_n(x) - f_n(x) + \frac{k_n}{nc} \right| = O\left(\frac{1}{k_n^\alpha}\right). \tag{17}$$

Moreover, in view of Lemma 2 (62),

$$\mathrm{Var}\,\hat{f}_n(x) = \mathrm{Var}\left((h_n + 1) \sum_{x_r \in J_\ell(x)} \frac{X^\star_{n,r}}{k_n} \right) = \frac{1}{d_n^2} \sum_{x_r \in J_\ell(x)} \mathrm{Var}\,X^\star_{n,r} \sim \frac{k_n h_n}{n^2 c^2}. \tag{18}$$

■

Of course, for $d_n = 1$, the variance is the worst since $k_n h_n/(n^2c^2) \sim k_n^2/(n^2c^2)$. (See the discussion on the choice of d_n in the introduction).

4.2 Mean integrated square convergence

Theorem 2 *Suppose f satisfies a Lipschitz condition of order α, $(0 < \alpha \leq 1)$. Then, for the mean integrated square convergence of $\hat{f}_n$ to f it is sufficient that $k_n = o\,(n/\ln n)$.*

Let us note that the condition $k_n = o\,(n/\ln n)$ is necessary for the application of Lemma 2 (60).

Proof : Utilizing Fubini's theorem, the mean integrated square error can be written as

$$J(\hat{f}_n) = \int_0^1 \mathrm{E}(\hat{f}_n(x) - f(x))^2\,dx = \mathrm{E}\left(\|\ \hat{f}_n - f\ \|_2^2\right). \tag{19}$$

Taking account of the L^2-orthogonality of $(\hat{f}_n - f_n)$ and $(f_n - f)$, we obtain

$$J(\hat{f}_n) = \mathrm{E}\left(\|\ \hat{f}_n - f_n\ \|_2^2\right) + \|\ f_n - f\ \|_2^2, \tag{20}$$

where the second term goes to zero by Proposition 2. In view of (6), the first term can be expanded as

$$\mathrm{E}\left(\|\ \hat{f}_n - f_n\ \|_2^2\right) = (h_n+1)\sum_{\ell=h_n+1}^{2h_n+1} \mathrm{E}\left[\left(\sum_{x_r\in J_\ell(x)} Y_{n,r}\right)^2\right]. \tag{21}$$

From the inequality

$$\mathrm{E}(Y_{n,r}Y_{n,s}) \leq \frac{1}{2}(\mathrm{E}(Y_{n,r}^2) + \mathrm{E}(Y_{n,s}^2)) \leq \max_{1\leq t\leq k_n} \mathrm{E}(Y_{n,t}^2), \tag{22}$$

we obtain

$$\mathrm{E}\left(\|\ \hat{f}_n - f_n\ \|_2^2\right) \leq k_n^2 \max_{1\leq t\leq k_n} \mathrm{E}(Y_{n,t}^2), \tag{23}$$

and the result follows from Lemma 2 (60):

$$\mathrm{E}\left(\|\ \hat{f}_n - f_n\ \|_2^2\right) = O\left(\frac{k_n^2}{n^2}\right) + O\left(\frac{1}{k_n^{2\alpha}}\right). \tag{24}$$

■

4.3 Almost complete uniform convergence

We shall give sufficient conditions for the convergence of the series

$$\forall \varepsilon > 0, \qquad \sum_{n=1}^{+\infty} P\left(\left\{\| \hat{f}_n - f \|_\infty > \varepsilon\right\}\right) < +\infty. \tag{25}$$

Theorem 3 *If f is continuous, then under the condition $k_n = o\,(n/\ln n)$, $\hat{f}_n$ converges to f almost completely uniformly.*

Proof : The condition $k_n = o\,(n/\ln n)$ implies $h_n = o\,(n)$ and consequently the continuity of f yields the uniform convergence of f_n to f. Thus, it remains to consider $(\hat{f}_n - f_n)$, which is given by (6). Then, we have

$$\| \hat{f}_n - f_n \|_\infty = (h_n + 1) \max_{h_n+1 \le \ell \le 2h_n+1} \left| \sum_{x_r \in J_\ell} Y_{n,r} \right|, \tag{26}$$

which implies

$$\begin{aligned} P\left(\left\{\| \hat{f}_n - f_n \|_\infty > \varepsilon\right\}\right) &\le \sum_{\ell=h_n+1}^{2h_n+1} P\left(\left\{\left| \sum_{x_r \in J_\ell} Y_{n,r} \right| > \frac{\varepsilon}{h_n + 1}\right\}\right), \\ &\le \sum_{\ell=h_n+1}^{2h_n+1} P\left(\left\{\bigcup_{x_r \in J_\ell} \left\{|Y_{n,r}| > \frac{\varepsilon}{d_n (h_n + 1)}\right\}\right\}\right), \end{aligned} \tag{27}$$

where $d_n = k_n/(h_n + 1)$ represents the cardinality of $\{x_r \in J_\ell\}$. Then,

$$P\left(\left\{\| \hat{f}_n - f_n \|_\infty > \varepsilon\right\}\right) \le \sum_{\ell=h_n+1}^{2h_n+1} \sum_{x_r \in J_\ell} P\left(\left\{\left|X^\star_{n,r} - k_n \lambda_{n,r}\right| > \varepsilon\right\}\right). \tag{28}$$

Since f is continuous, for n large enough $M_{n,r} - m_{n,r} < \varepsilon/2$ uniformly in r, and therefore $\left\{\left|X^\star_{n,r} - k_n \lambda_{n,r}\right| > \varepsilon\right\} \subset \left\{X^\star_{n,r} < m_{n,r} - \varepsilon/2\right\}$. This yields

$$P\left(\left\{\| \hat{f}_n - f_n \|_\infty > \varepsilon\right\}\right) \le \sum_{\ell=h_n+1}^{2h_n+1} \sum_{x_r \in J_\ell} F_{n,r}(m_{n,r} - \varepsilon/2). \tag{29}$$

In view of Lemma 1, $F_{n,r}(m_{n,r} - \varepsilon/2)$ can be computed, leading to

$$P\left(\left\{\| \hat{f}_n - f_n \|_\infty > \varepsilon\right\}\right) \le \sum_{\ell=h_n+1}^{2h_n+1} \sum_{x_r \in J_\ell} \exp\left(-\frac{nc\varepsilon}{2k_n}\right) \le k_n \exp\left(-\frac{nc\varepsilon}{2k_n}\right), \tag{30}$$

which is the general term of a convergent series provided $k_n = o\,(n/\ln n)$. ■

4.4 Asymptotic distribution

It is worth noticing that in convergence Theorems 1, 2, and 3, conditions $k_n = o(n/\ln n)$ and $n = o\left(k_n^{1+\alpha}\right)$ do not imply that k_n should go to infinity faster than h_n. On the contrary, the asymptotic distribution of $\hat{f}_n$ depends strongly on the relative rates of h_n and k_n. We present two entirely different cases. In the first one ($d_n = 1$), the partition $\{J_\ell,\ h_n + 1 \leq \ell \leq 2h_n + 1\}$, is arranged to be exactly the same as the partition $\{I_{n,r},\ 1 \leq r \leq k_n\}$, so that the estimator $\hat{f}_n$ merely reduces to $\hat{g}_n$, the Geffroy's estimate, a kind of histogram based on the uppermost values of the point process on each $I_{n,r}$. Such a case has already been extensively studied since the generating paper of Geffroy (1964) and gives extremal asymptotic distributions. The second case ($d_n \to \infty$) is more in the spirit of the orthogonal series method, with a condition $h_n = o(k_n)$ suggesting an estimation of the coordinates $a_i = \int_0^1 e_i(t) f(t) dt$ by the random Riemann sums $\hat{a}_{i,k_n}$, see (11). Due to the underlying properties of the Poisson process N^{*n}, these $\hat{a}_{i,k_n}$ are sums of independent terms, so that a Gaussian asymptotic distribution is expected. However, in both theorems, a closer control of k_n is required, which is essentially of the same kind as $n = o\left(k_n^{1+\alpha}\right)$.

Theorem 4 *Let $x \in [0,1]$ and suppose f is α-Lipschitzian ($0 < \alpha \leq 1$). Suppose $d_n = 1$ and $n = o\left(k_n^{1+\alpha}\right)$.*

(i) If $k_n = o(n)$, then $T_n(x) = \frac{nc}{k_n}(\hat{f}_n(x) - k_n \lambda_{n,r})$ converges in distribution to the Weibull Extremal Value Distribution.

(ii) Suppose $k_n \ln k_n = o(n)$ and denote $Z_n = \max_{1 \leq r \leq k_n} \sup_{x \in I_{n,r}} |\hat{f}_n(x) - M_{n,r}|$, then $\frac{nc}{k_n} Z_n - \ln k_n$ converges in distribution to the Gumbel Extremal Value Distribution.

Let us note that in (ii), the condition $k_n \ln k_n = o(n)$ is just a weaker version of the condition $k_n = o(n/\ln n)$.

Proof :

(i) Denote by $G_{n,x}$ the distribution function of $T_n(x)$, and consider r such that $x \in I_{n,r}$. Then for all $u \in \mathbb{R}$,

$$G_{n,x}(u) = P\left(\left\{\hat{f}_n(x) \leq k_n\lambda_{n,r} + uk_n/nc\right\}\right) = F_{n,r}(k_n\lambda_{n,r} + uk_n/nc), \tag{31}$$

since $\hat{f}_n(x)$ reduces to $X^\star_{n,r}$ on $I_{n,r}$ when $d_n = 1$. Suppose for instance that $u \leq 0$. Then $k_n = o(n)$ implies $n\lambda_{n,r} \to \infty$ and thus for n large enough,

$$u \in \left[-nc\lambda_{n,r}, nc\left(\frac{M_{n,r}}{k_n} - \lambda_{n,r}\right)\right].$$

Lemma 1 yields $G_{n,x}(u) = \exp(u - nc\xi_{n,r}(u))$ with

$$\xi_{n,r}(u) = \varepsilon_{n,r}(k_n\lambda_{n,r} + uk_n/nc) - \lambda_{n,r} + M_{n,r}/k_n, \tag{32}$$

and $0 \leq nc\xi_{n,r}(u) \leq nc(M_{n,r} - m_{n,r})/k_n$. Then $\xi_{n,r} \to 0$ as $n \to \infty$ under condition $n = o\left(k_n^{1+\alpha}\right)$. As a conclusion, $G_{n,x}(u) \to \exp(u)$ if $u \leq 0$. The case $u > 0$ is similar, the result being $G_{n,x}(u) \to 1$.

(ii) When $d_n = 1$, $\hat{f}_n(x)$ reduces to $X^\star_{n,r}$ on $I_{n,r}$, and thus $Z_n = \max\limits_{1\leq r\leq k_n}(M_{n,r} - X^\star_{n,r})$. Let G_n stand for the distribution function of $\frac{nc}{k_n}Z_n - \ln k_n$. Consider $u \in \mathbb{R}$. For n large enough

$$u \in \left[-\ln k_n, \frac{nc}{k_n}\left(M_{n,r} - \frac{k_n \ln k_n}{nc}\right)\right],$$

$\forall r \in \{1, \ldots, k_n\}$ and by a straightforward computation

$$G_n(u) = \prod_{r=1}^{k_n} \bar{F}_{n,r}\left(M_{n,r} - \frac{k_n}{nc}(u + \ln k_n)\right) = \prod_{r=1}^{k_n}\left(1 - e^{-\eta_{n,r}}\frac{e^{-u}}{k_n}\right), \tag{33}$$

with $nc(\lambda_{n,r} - M_{n,r}/k_n) \leq \eta_{n,r} \leq 0$. Now, $\prod\limits_{r=1}^{k_n}\left(1 - \frac{e^{-u}}{k_n}\right) \to e^{-e^{-u}}$, $\eta_{n,r}$ is a null array in view of $n = o\left(k_n^{1+\alpha}\right)$, and Lemma 4 (see Appendix) gives the result. ■

Now, let $x \in [0,1]$ be fixed. Denote $\sigma_n = k_n/(ncd_n^{1/2})$.

Theorem 5 *If f is α-Lipschitzian ($0 < \alpha \leq 1$), under the three conditions $h_n = o\,(k_n)$, $k_n = o\,(n/\ln n)$ and $n = o\left(k_n^{1/2+\alpha} h_n^{1/2}\right)$, $V_n(x) = \sigma_n^{-1}(\hat{f}_n(x) - E(\hat{f}_n(x)))$ converges in distribution to a standard Gaussian variable.*

Proof : Let us denote by $\psi_{n,x}$ the characteristic function of $V_n(x)$. It expands as

$$\begin{aligned}
\psi_{n,x}(t) &= \exp it \left(\frac{nc}{k_n d_n^{1/2}} \sum_{x_r \in J_\ell(x)} (m_{n,r} - \mathrm{E}(X^\star_{n,r})) - d_n^{1/2} \right) \\
&\times \; \mathrm{E}\left[\exp it \left(\frac{nc}{k_n d_n^{1/2}} \sum_{x_r \in J_\ell(x)} (X^\star_{n,r} - m_{n,r}) + d_n^{1/2} \right) \right]. \qquad (34)
\end{aligned}$$

Introducing $a_{n,r} = m_{n,r} - k_n/nc$, (34) can be rewritten as

$$\begin{aligned}
\psi_{n,x}(t) &= \exp it \left(\frac{nc}{k_n d_n^{1/2}} \sum_{x_r \in J_\ell(x)} (a_{n,r} - \mathrm{E}(X^\star_{n,r})) \right) \\
&\times \; \mathrm{E}\left[\exp it \left(\frac{nc}{k_n d_n^{1/2}} \sum_{x_r \in J_\ell(x)} (X^\star_{n,r} - m_{n,r}) + d_n^{1/2} \right) \right]. \qquad (35)
\end{aligned}$$

In view of Lemma 2 (61), the first term in the product converges to 1 under conditions $h_n = o\,(k_n)$, $k_n = o\,(n/\ln n)$ and $n = o\left(k_n^{1/2+\alpha} h_n^{1/2}\right)$. After simplification, and in view of the independence of the $X^\star_{n,r}$,

$$\psi_{n,x}(t) \sim e^{itd_n^{1/2}} \prod_{x_r \in J_\ell(x)} \mathrm{E}\left[\exp it \left(\frac{nc}{k_n d_n^{1/2}} (X^\star_{n,r} - m_{n,r}) \right) \right]. \qquad (36)$$

Introducing the characteristic function of $X^\star_{n,r} - m_{n,r}$, it follows that

$$\psi_{n,x}(t) \sim e^{itd_n^{1/2}} \prod_{x_r \in J_\ell(x)} \Phi_{n,r}(t_n) \qquad (37)$$

where $t_n = tncd_n^{-1/2}/k_n$. Before applying Lemma 3, it remains to verify that $t_n = o\,(n/k_n)$ and $t_n = o\left(k_n^{1+2\alpha}/n\right)$. Clearly,

$$t_n \frac{k_n}{nc} = td_n^{-1/2} = o\,(1) \qquad (38)$$

since $h_n = o\,(k_n)$, and

$$t_n \frac{n}{k_n^{1+2\alpha}} = tcd_n^{-1/2} \left(\frac{n}{k_n^{1+\alpha}} \right)^2 = o\,(1)\,, \qquad (39)$$

since $h_n = o(k_n)$ and $n = o\left(k_n^{1+\alpha}\right)$. The characteristic function is of the order:

$$\psi_{n,x}(t) \sim \exp\left(itd_n^{1/2}\right)\left(1+itd_n^{-1/2}\right)^{-d_n} \tag{40}$$

$$\times \prod_{x_r\in J_\ell(x)}\left(1+itd_n^{-1/2}\bar{F}_{n,r}(m_{n,r})+O\left(|t_n|\frac{n}{k_n^{1+2\alpha}}\right)+o\left(n^{-s}\right)\right). \tag{41}$$

By a third-order Taylor expansion, one sees that (40) converges to $e^{-t^2/2}$ as $d_n \to +\infty$. Finally, consider the logarithm of (41):

$$W_n = \sum_{x_r\in J_\ell(x)} \ln(1+u_{n,r}), \text{ with } u_{n,r} = itd_n^{-1/2}\bar{F}_{n,r}(m_{n,r})+O\left(|t_n|\frac{n}{k_n^{1+2\alpha}}\right)+o\left(n^{-s}\right). \tag{42}$$

Observe that $\max_{1\le r\le k_n}|u_{n,r}|$ converges to 0 with (38) and (39). Thus, for n large enough $|u_{n,r}| < 1/2$ uniformly in r and the classical inequality $|\ln(1+u_{n,r})| < 2|u_{n,r}|$ yields

$$|W_n| \le 2|t|\,d_n^{1/2}\bar{F}_{n,r}(m_{n,r})+O\left(|t_n|\,d_n\frac{n}{k_n^{1+2\alpha}}\right)+o\left(d_n n^{-s}\right). \tag{43}$$

Using Lemma 1, we have $\bar{F}_{n,r}(m_{n,r}) = O\left(n/k_n^{1+\alpha}\right)$ leading to

$$|W_n| = O\left(\frac{n}{h_n^{1/2}k_n^{1/2+\alpha}}\right)+O\left(\frac{n^2}{h_n^{1/2}k_n^{3/2+2\alpha}}\right)+o\left(\frac{k_n}{h_n n^s}\right). \tag{44}$$

Therefore, choosing $s = 2$ entails $W_n \to 0$ and $\psi_{n,x}(t) \to e^{-t^2/2}$ as $n \to \infty$. ■

5 Bias reduction

Theorem 5 is not convenient for providing confidence intervals for $f(x)$, since it only gives the asymptotic distribution of $\sigma_n^{-1}(\hat{f}_n(x) - \mathrm{E}(\hat{f}_n(x)))$. A quick calculation shows that even in the case $f(x) \equiv a > 0$, there is no hope of a similar result for $\sigma_n^{-1}(\hat{f}_n(x) - a)$. Indeed,

$$\mathrm{E}\left(a - X_{n,r}^\star\right) = \frac{k_n}{nc}\left(1-\exp\left(-\frac{nca}{k_n}\right)\right) \ge \frac{1}{2}\frac{k_n}{nc} \tag{45}$$

for n large enough. Thus, for every $x \in [0,1]$, under the hypotheses of Theorem 5:

$$\sigma_n^{-1}(a-\mathrm{E}(\hat{f}_n(x))) = \sigma_n^{-1}\frac{1}{d_n}\sum_{x_r\in J_\ell(x)}(a-\mathrm{E}(X_{n,r}^\star)) \ge \frac{1}{2}\frac{ncd_n^{1/2}}{k_n}\frac{k_n}{nc} = \frac{1}{2}d_n^{1/2} \to \infty. \tag{46}$$

Define $\check{f}_n(x) = \hat{f}_n(x) + k_n/(nc)$. Since c is not supposed to be known, $\check{f}_n$ is of no use in practice. Nevertheless, it is a first theoretical step to introduce an estimator with a smaller bias and to study its asymptotic distribution (see Theorem 7).

Theorem 6 *If f is α-Lipschitzian $(0 < \alpha \leq 1)$, under the three conditions $h_n = o(k_n)$, $k_n = o(n/\ln n)$, and $n = o\left(k_n^{1/2} h_n^{1/2+\alpha}\right)$, $\sigma_n^{-1}(\check{f}_n(x) - f(x))$ converges in distribution to a standard Gaussian variable.*

Proof : The proof is based on the expansion

$$\sigma_n^{-1}(\check{f}_n - f) = \sigma_n^{-1}(\check{f}_n - \mathrm{E}(\check{f}_n)) + \sigma_n^{-1}(\mathrm{E}(\check{f}_n) - f_n) + \sigma_n^{-1}(f_n - f). \tag{47}$$

Theorem 5 shows that the first term converges in distribution to a standard Gaussian variable. Now, from Lemma 2 (61),

$$\max_{1 \leq r \leq k_n} \left| \mathrm{E}\left(X_{n,r}^\star\right) + \frac{k_n}{nc} - k_n \lambda_{n,r} \right| = O\left(\frac{1}{k_n^\alpha}\right), \tag{48}$$

and we obtain,

$$\sigma_n^{-1}\left(\mathrm{E}(\check{f}_n(x) - f_n(x)\right) = \sigma_n^{-1}(h_n + 1) \sum_{x_r \in J_\ell(x)} \left(\frac{\mathrm{E}\left(X_{n,r}^\star\right)}{k_n} + \frac{1}{nc} - \lambda_{n,r}\right) = O\left(\frac{n}{k_n^{1/2+\alpha} h_n^{1/2}}\right). \tag{49}$$

Finally, let us consider the third term. In view of Proposition 1,

$$\sigma_n^{-1} \left| f_n(x) - f(x) \right| = \frac{ncd_n^{1/2}}{k_n} O\left(h_n^{-\alpha}\right) = O\left(\frac{n}{k_n^{1/2} h_n^{1/2+\alpha}}\right) \to 0. \tag{50}$$

■

We are now ready to introduce a very simple and ad hoc estimate of $k_n/(nc)$:

$$Z_n = \frac{1}{k_n} \sum_{r=1}^{k_n} Z_{n,r}^\star, \tag{51}$$

where $Z_{n,r}^\star$ is defined by $Z_{n,r}^\star = 0$ if $N^{*n}(D_{n,r}) = 0$ and $Z_{n,r}^\star$ is the infimum of the second coordinates of the points of the truncated process $N^{*n}(. \cap D_{n,r})$ if $N^{*n}(D_{n,r}) > 0$. Define

$$\tilde{f}_n(x) = \sum_{r=1}^{k_n} K_n(x_r, x) \left(\frac{X_{n,r}^\star + Z_n}{k_n}\right). \tag{52}$$

The moments of $Z^\star_{n,r}$ can be expanded similarly to that of $X^\star_{n,r}$, leading to

$$\mathrm{E}(Z_n) = \frac{k_n}{nc} + o\left(n^{-s}\right), \; \forall s > 0 \text{ and } \mathrm{Var}(Z_n) \sim \frac{k_n}{n^2c^2}. \tag{53}$$

We now prove the following theorem which states that $\tilde{f}_n$ inherits its asymptotic behavior from $\check{f}_n$.

Theorem 7 *Under conditions of Theorem 6, $\sigma_n^{-1}(\tilde{f}_n(x) - f(x))$ converges in distribution to a standard Gaussian random variable.*

Proof : In view of Theorem 6, it suffices to verify that $\sigma_n^{-1}(\check{f}_n(x) - \tilde{f}_n(x))$ converges to 0 in probability. We shall show that for all $\varepsilon > 0$,

$$P\left(\left\{\sigma_n^{-1}\left|\check{f}_n(x) - \tilde{f}_n(x)\right| > \varepsilon\right\}\right) \le P\left(\left\{\sigma_n^{-1}\left|Z_n - \frac{k_n}{nc}\right| > \varepsilon\right\}\right) \tag{54}$$

converges to zero as n goes to infinity. To this end, consider the expansion

$$\sigma_n^{-1}\left|Z_n - \frac{k_n}{nc}\right| \le \sigma_n^{-1}\left|Z_n - \mathrm{E}(Z_n)\right| + \sigma_n^{-1}\left|\mathrm{E}(Z_n) - \frac{k_n}{nc}\right|. \tag{55}$$

In view of (53), the second term is bounded above by $\varepsilon/2$ for n large enough, and therefore

$$\begin{aligned} P\left(\left\{\sigma_n^{-1}\left|\check{f}_n(x) - \tilde{f}_n(x)\right| > \varepsilon\right\}\right) &\le P\left(\left\{\sigma_n^{-1}\left|Z_n - \mathrm{E}(Z_n)\right| > \varepsilon/2\right\}\right) & (56) \\ &\le \frac{4}{\varepsilon^2}\mathrm{Var}\left(\sigma_n^{-1}Z_n\right) = O\left(1/h_n\right), & (57) \end{aligned}$$

with (53). ■

For instance, in the case where f is continuously differentiable ($\alpha = 1$), one can choose $h_n = n^{1/2}$ and $k_n = n^{1/2}(\ln n)^\varepsilon$ for all $\varepsilon > 0$ arbitrarily small. Theorem 7 shows that $\tilde{f}_n(x)$ converges to $f(x)$ with the rate $n^{-1/2}(\ln n)^\varepsilon$ for all $x \in [0, 1]$.

6 Illustration on a simulation

In this very simple illustration, the sum of $n = 100$ independent Poisson processes with the same intensity rate $c = 3$ is simulated on the set S defined as in (1) with

$$f(x) = \left| \begin{array}{ll} 1 + \exp\left(-60(x-1/4)^2\right) & \text{if } 0 \le x \le 1/3, \\ 1 + \exp\left(-5/12\right) & \text{if } 1/3 < u \le 2/3, \\ 3x - 1 & \text{if } 2/3 < u \le 1. \end{array} \right. \tag{58}$$

Note that f is not derivable at $x = 1/3$ and not continuous at $x = 2/3$ (see Figure 2). We consider an expansion of order $h_n = 15$ of f in the Haar basis, see (4), where each coefficient is approximated by dividing $[0,1]$ into $k_n = 32$ intervals. In this case $d_n = 2$, then for each $x \in [0,1]$, the estimator $\hat{f}_n$ is simply written $\hat{f}_n(x) = (X^\star_{2p+1} + X^\star_{2p+2})/2$, where p denotes the integer value of $32x$. The estimate is represented on Figure 4 where it is compared with $\tilde{f}_n$ obtained by reducing the bias as described in the previous section.

(Figure 4 here)

7 Conclusion and further developments

In this paper, the underlying Poisson point process is supposed to be homogeneous. It is easily seen that all convergence theorems of section 4 still hold if the intensity is an arbitary function on S, bounded below by a constant $c > 0$. Such an extension for limit distributions is not straightforward. We plan to publish a companion paper devoted to C^1-basis estimate. A first draft of this study, based on formula (12), can be found in Girard and Jacob (2000). This case involves rather different difficulties due to the Dirichlet kernel feature. However, this provides smooth estimators more adapted to regular boundaries.

References

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

Abbar, H. and Suquet, Ch. (1993) Estimation L^2 du contour d'un processus de Poisson homogène sur le plan. *Publ. IRMA Lille*, **31**, **II**.

Cencov, N. N. (1962) Evaluation of an unknown distribution density from observations. *Soviet Math.* **3**, 1559–1562.

Chevalier, J. (1976) Estimation du support et du contour d'une loi de probabilité. *Ann. Inst. H. Poincaré Probab. Statist.*, sect. B, 12, 339–364.

de Haan, L. and Resnick, S. (1994) Estimating the home range. *J. Appl. Probab.*, **31**, 700–720.

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.

Gayraud, G. (1997) Estimation of functionals of density support. *Math. Methods Statist.* **6**(1), 26–46.

Geffroy, J. (1964) Sur un problème d'estimation géométrique. *Publ. Inst. Statist. Univ. Paris*, **XIII**, 191–200.

Gensbittel, M. H. (1979) *Contribution à l'étude statistique de répartitions ponctuelles aléatoires.* Thèse, Université Pierre et Marie Curie, Paris.

Girard, S. and Jacob, P. (2000) Extreme values and C^1 projection estimates of point processes boundaries. *Technical report ENSAM-INRA-UM2*, **00-03**.

Härdle, W., Park, B. U. and Tsybakov, A. B. (1995a) Estimation of a non sharp support boundaries. *J. Multivariate Anal.*, **43**, 205–218.

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

Jacob, P. (1984) Estimation du contour discontinu d'un processus ponctuel sur le plan. *Publ. Inst. Statist. Univ. Paris*, **XXIX**, 1–25.

Jacob, P. and Abbar, H. (1989) Estimating the edge of a Cox process area. *Cahiers du Centre d'Etudes de Recherche Opérationnelle*, **31**, 215–226.

Jacob, P. and Suquet, P. (1995) Estimating the edge of a Poisson process by orthogonal series. *J. Statist. Plann. Inference*, **46**, 215–234.

Korostelev, A., Simar, L. and Tsybakov, A. B. (1995) Efficient estimation of monotone boundaries. *Ann. Statist.*, **23**, 476–489.

Korostelev, A. P. and Tsybakov, A. B. (1993) Minimax linewise algorithm for image reconstruction. in *Computer intensive methods in statistics* by W. Härdle, L. Simar ed., Statistics and Computing, Physica Verlag, Springer.

Mammen, E. and Tsybakov, A. B. (1995) Asymptotical minimax recovery of set with smooth boundaries. *Ann. Statist.*, **23**(2), 502–524.

Tiago de Oliveira, J. (1963) Estatistica de densidades; Resultados assintoticos. *Rev. Fac. Ci. Univ. Lisboa* **A9**, 65–171.

Appendix: Auxilary results

In the following lemma we give the distribution function of $X^{\star}_{n,r}$ $(1 \leq r \leq k_n)$, with an accurate approximation determined by the bounds $m_{n,r}$ and $M_{n,r}$ of f on $I_{n,r}$ (the notations are those introduced in section 2). The proof is straightforward after noticing that, for every measurable set B, $P(N^{*n}(B) = 0) = \exp\left(-nc\lambda(B)\right)$.

Lemma 1 *The distribution function of $X^\star_{n,r}$ $(1 \le r \le k_n)$ is given by*

$$F_{n,r}(u) = \left| \begin{array}{ll} 0 & \text{if } u < 0, \\ \exp\left[\frac{nc}{k_n}(u - M_{n,r}) - nc\varepsilon_{n,r}(u)\right] & \text{if } 0 \le u \le M_{n,r}, \\ 1 & \text{if } M_{n,r} \le u, \end{array} \right. \tag{59}$$

with $\varepsilon_{n,r}(u) = \lambda_{n,r} - \frac{M_{n,r}}{k_n}$, if $0 \le u \le m_{n,r}$, and $\lambda_{n,r} - \frac{M_{n,r}}{k_n} \le \varepsilon_{n,r}(u) \le 0$, if $m_{n,r} \le u \le M_{n,r}$.

An approximation of the first moments can be derived from the distribution function given in Lemma 1:

Lemma 2 *Suppose f is α-Lipschitzian $(0 < \alpha \le 1)$ and $k_n = o\,(n/\ln n)$. Then,*

$$\max_{1 \le r \le k_n} E\left(Y^2_{n,r}\right) = O\left(\frac{1}{n^2}\right) + O\left(\frac{1}{k_n^{2(1+\alpha)}}\right), \tag{60}$$

and

$$\max_{1 \le r \le k_n} \left|E\left(X^\star_{n,r}\right) - a_{n,r}\right| = O\left(\frac{1}{k_n^\alpha}\right), \tag{61}$$

where $a_{n,r} = m_{n,r} - \frac{k_n}{nc}$. If moreover $n = o\left(k_n^{1+\alpha}\right)$,

$$Var\left(X^\star_{n,r}\right) \sim \frac{k_n^2}{n^2c^2}, \;\; 1 \le r \le k_n. \tag{62}$$

Proof :

- Observe that

$$\begin{aligned} \mathrm{E}\left[\left(X^\star_{n,r} - k_n\lambda_{n,r}\right)^2\right] = k_n^2\lambda_{n,r}^2 P(X^\star_{n,r} = 0) \;&+\; \int_0^{k_n\lambda_{n,r}} (u - k_n\lambda_{n,r})^2 F'_{n,r}(u)du \\ &+\; \int_{k_n\lambda_{n,r}}^{M_{n,r}} (u - k_n\lambda_{n,r})^2 F_{n,r}(du). \end{aligned} \tag{63}$$

  The distribution function $F_{n,r}$ is given in Lemma 1 using $\varepsilon_{n,r}$, a non-decreasing function on $[0, M_{n,r}]$. Then, the derivative $\varepsilon'_{n,r}$ exists almost everywhere and is non-negative. We thus obtain

$$F'_{n,r}(u) \le \frac{nc}{k_n} \exp\left[\frac{nc}{k_n}\left(u - k_n\lambda_{n,r}\right)\right] \text{ a. e.} \tag{64}$$

with equality for $0 \le u \le m_{n,r}$. The following bound may be found quite easily:

$$\mathrm{E}\left[\left(X^\star_{n,r} - k_n\lambda_{n,r}\right)^2\right] \le k_n^2\lambda_{n,r}^2 e^{-nc\lambda_{n,r}} + \frac{2k_n^2}{n^2c^2} + (M_{n,r} - m_{n,r})^2, \tag{65}$$

for $r = 1, \ldots, k_n$, and (60) is proved.

- From Lemma 1,

$$\mathrm{E}\left(X^\star_{n,r}\right) = \int_0^\infty \bar{F}_{n,r}(u)du = \int_0^{m_{n,r}} \left[1 - \exp\left(\frac{nc}{k_n}\left(u - k_n\lambda_{n,r}\right)\right)\right] du + \int_{m_{n,r}}^{M_{n,r}} \bar{F}_{n,r}(u)du,$$

where the last integral is positive and less than $(M_{n,r} - m_{n,r})$. Therefore, for $r = 1, \ldots, k_n$,

$$m_{n,r} + \frac{k_n}{nc}e^{-nc\lambda_{n,r}}\left(1 - e^{nc\frac{m_{n,r}}{k_n}}\right) \le \mathrm{E}(X^\star_{n,r}) \le M_{n,r} + \frac{k_n}{nc}e^{-nc\lambda_{n,r}}\left(1 - e^{nc\frac{m_{n,r}}{k_n}}\right), \tag{66}$$

and consequently

$$\left|\mathrm{E}\left(X^\star_{n,r}\right) - a_{n,r}\right| \le (M_{n,r} - m_{n,r}) + \frac{k_n}{nc}\left|1 + e^{-nc\lambda_{n,r}} - e^{-nc(\lambda_{n,r} - m_{n,r}/k_n)}\right|. \tag{67}$$

The first term of the sum is obviously $O\left(1/k_n^\alpha\right)$, uniformly in r. Furthermore

$$\begin{aligned}\left|1 + e^{-nc\lambda_{n,r}} - e^{-nc(\lambda_{n,r} - m_{n,r}/k_n)}\right| &\le e^{-ncm_{n,r}/k_n} + 1 - e^{-nc(M_{n,r} - m_{n,r})/k_n} \\ &\le e^{-ncm_{n,r}/k_n} + \frac{nc(M_{n,r} - m_{n,r})}{k_n} = O\left(\frac{n}{k_n^{1+\alpha}}\right),\end{aligned} \tag{68}$$

uniformly in r and (61) is proved.

- (62) is established with similar techniques. ■

The expression of the characteristic function can also be deduced from the distribution function given in Lemma 1:

Lemma 3 *Suppose f is α-Lipschitzian ($0 < \alpha \le 1$) and $k_n = o(n/\ln n)$. Let $(t_n) \in \mathbb{R}$ be a sequence such that $t_n = o(n/k_n)$ and $t_n = o\left(k_n^{1+2\alpha}/n\right)$. Then, the characteristic function of $(X^\star_{n,r} - m_{n,r})$ can be written at point t_n as:*

$$\Phi_{n,r}(t_n) = \frac{1 + it_n\frac{k_n}{nc}\bar{F}_{n,r}(m_{n,r}) + O\left(|t_n|\frac{n}{k_n^{1+2\alpha}}\right) + o\left(n^{-s}\right)}{1 + it_n\frac{k_n}{nc}},$$

with s arbitrary large, and $\bar{F}_{n,r} = 1 - F_{n,r}$ ($1 \le r \le k_n$).

Proof : Consider the expansion

$$\mathrm{E}(e^{it_n X^\star_{n,r}}) = P(X^\star_{n,r}=0) + \int_0^{m_{n,r}} e^{ixt_n} F'_{n,r}(x)dx + \int_{m_{n,r}}^{M_{n,r}} e^{ixt_n} F_{n,r}(dx). \qquad (69)$$

The first and second term can be computed explicitely since Lemma 1 provides an expression of $F_{n,r}$ on $[0, m_{n,r}]$:

$$\mathrm{E}(e^{it_n X^\star_{n,r}}) = e^{-nc\lambda_{n,r}} + \frac{e^{it_n m_{n,r}} F_{n,r}(m_{n,r}) - e^{-nc\lambda_{n,r}}}{1+it_n \frac{k_n}{nc}} + \int_{m_{n,r}}^{M_{n,r}} e^{ixt_n} F_{n,r}(dx). \qquad (70)$$

The third term expands as

$$\int_{m_{n,r}}^{M_{n,r}} e^{ixt_n} F_{n,r}(dx) = e^{it_n m_{n,r}} \bar{F}_{n,r}(m_{n,r}) + \int_{m_{n,r}}^{M_{n,r}} (e^{ixt_n} - e^{it_n m_{n,r}}) F_{n,r}(dx), \qquad (71)$$

with $\left|e^{ixt_n} - e^{it_n m_{n,r}}\right| \le (M_{n,r} - m_{n,r})\,|t_n|$. Thus

$$\begin{aligned} \left| \int_{m_{n,r}}^{M_{n,r}} (e^{ixt_n} - e^{it_n m_{n,r}}) F_{n,r}(dx) \right| &\le (M_{n,r} - m_{n,r})\,|t_n|\, \bar{F}_{n,r}(m_{n,r}) \\ &\le (M_{n,r} - m_{n,r})\,|t_n|\, \frac{nc}{k_n}(k_n\lambda_{n,r} - m_{n,r}) \\ &= O\left(|t_n| \frac{n}{k_n^{1+2\alpha}}\right). \end{aligned} \qquad (72)$$

Now, take the common denominator in (70) and replace (71) and (72) in the numerator. Remark that $k_n = o(n/\ln n)$ yields $e^{-nc\lambda_{n,r}} = o(n^{-s})$, $\forall s > 0$ and multiply everything by $e^{-it_n m_{n,r}}$ to conclude the proof. ■

In order to be more self-contained, we quote here Lemma 4.1 of Jacob and Abbar (1989) which is used to prove Theorem 4(ii). Let us recall that a triangular array $\{\eta_{n,r},\ r = 1, \dots, k_n\}$ is called a null array if $\max_{1\le r\le k_n} |\eta_{n,r}| \to 0$ as $n \to \infty$.

Lemma 4 *Let $\{x_{n,r},\ r = 1, \dots, k_n\}$ be a positive triangular array such that $\{z_{n,r}\} = \{e^{-x_{n,r}}\}$ is a null array. If $\prod_{r=1}^{k_n}(1 - z_{n,r})$ has a limit $\ell > 0$, then $\prod_{r=1}^{k_n}(1 - z_{n,r}e^{\eta_{n,r}})$ has the same limit ℓ, whenever $\{\eta_{n,r}\}$ is a null array.*

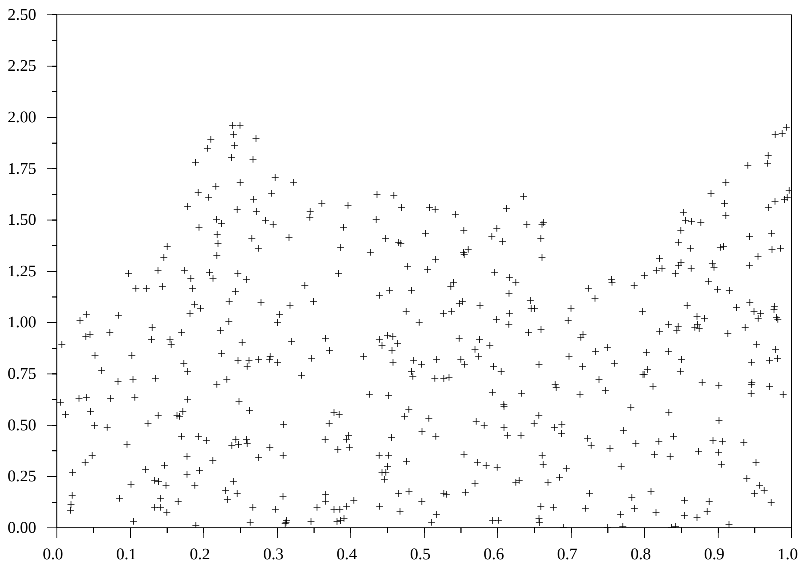

Figure 1: Simulation of a point process with support of the form $S = \{(x, y) : 0 \leq x \leq 1 \;;\; 0 \leq y \leq f(x)\}$.

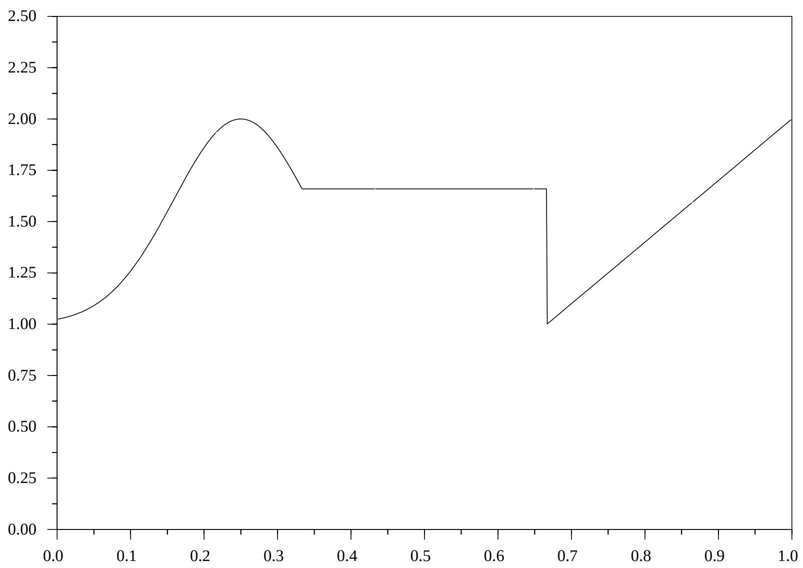

Figure 2: Function f associated to the simulation of Figure 1.

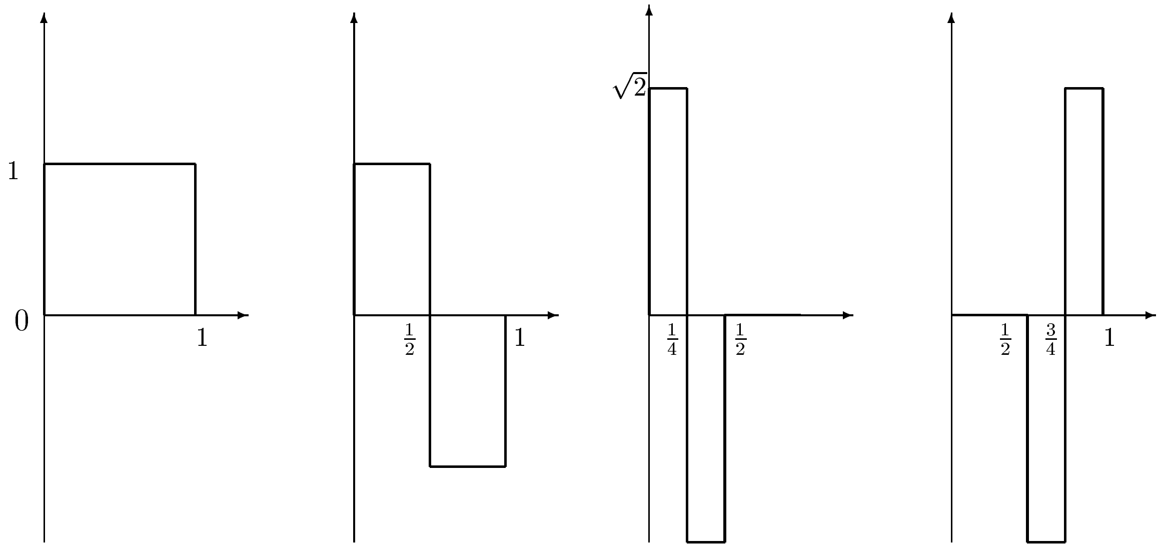

Figure 3: First functions of the Haar basis. From left to right: e_0, e_1, e_2 and e_3.

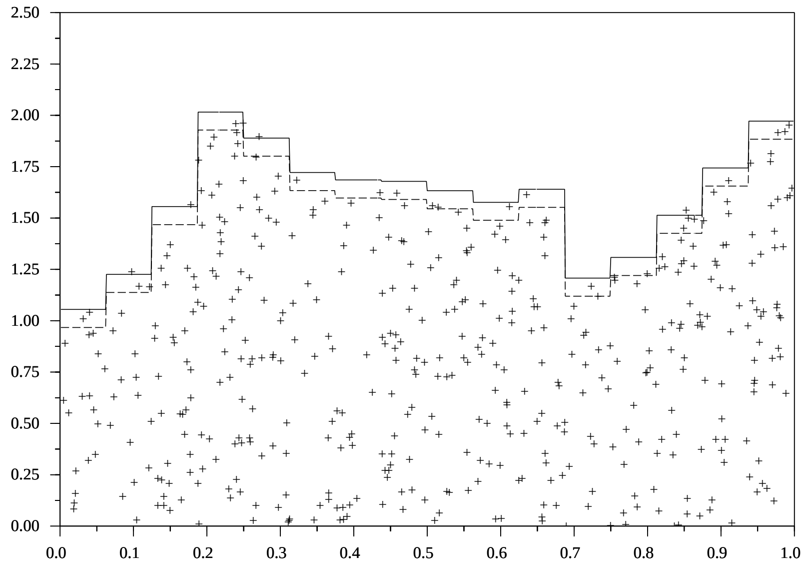

Figure 4: Comparison between the two estimates $\hat{f}_n$ (dashed line) and $\tilde{f}_n$ (continuous line).

Authors' address

Pierre Jacob & Stéphane Girard,

Laboratoire de Probabilités et Statistique, Université Montpellier 2,

place Eugène Bataillon, 34095 Montpellier cedex 5, France.

{jacob, girard}@stat.math.univ-montp2.fr