Projection estimates of point processes boundaries

Stéphane Girard & Pierre Jacob

Laboratoire de Probabilités et Statistique, Université Montpellier 2,
place Eugène Bataillon, 34095 Montpellier cedex 5, France.
{girard, jacob}@stat.math.univ-montp2.fr

Abstract

We present a 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 projections on C^1 bases and on extreme points of the point process. We give conditions on the Dirichlet's kernel associated to the C^1 bases for various kinds of convergence and asymptotic normality. We propose a method for reducing the negative bias and illustrate it by a simulation.

Keywords: Projection on C^1 bases, Extreme values, Poisson process, Shape estimation.

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

1 Introduction

We address the problem of estimating a bounded set S of $\mathbf{R}^2$ given a finite random set Σ of points drawn from the interior. This kind of problem arises in various frameworks such as classification [9], image processing [13] or econometrics problems [3]. A lot of different solutions were proposed since [6] and [15] depending on the properties of the observed random set Σ and of the unknown set S. In this paper, we focus on the special case where Σ is the set of points of an homogeneous Poisson process whose support is $S = \{(x,y) \in \mathbf{R}^2 \mid 0 \le x \le 1 \; ; \; 0 \le y \le f(x)\}$, where f is an unknown function. Thus, the estimation of subset S reduces to the estimation of function f. Let us note that this kind of support was already considered in [6]. In the wide range of nonparametric functional estimators [2], piecewise polynomials have been especially studied [12, 13, 14, 5, 8] and their asymptotic optimality was established under different regularity conditions on f.
The first support estimator based on orthogonal series appears in [1]. Its properties are extensively studied in [11] in the case of Haar and C^1 bases. The expansion coefficients estimation requires the knowledge of the process intensity. This is a limitation which is avoided in [7] in the case of the Haar basis by considering a coefficient estimation based on the extreme points of the sample. In this paper, a similar study is carried out in the case of C^1 bases. The estimator can be written as a linear combination of extreme values involving the Dirichlet's kernel of the C^1 basis. A close study of the extreme values stochastic properties as well as a precise control of the Dirichlet's kernel behavior allow one to establish general conditions for various convergences and asymptotic normality of the estimator. Our results are illustrated for the trigonometric basis case. Note that the model proposed here is well-adapted to the estimation of a bounded star-shaped subset of the plane. Let D be such a domain. Then, there exists a convex subset C of this domain, called the kernel of D, from which the whole boundary ∂D of D can be seen. If we assume an interior point of C to be known, the use of polar coordinates allows to reduce the problem of estimating ∂D to the problem considered here, that is the estimation of a function

f, with the particularity that $f(0) = f(1)$. In such a situation, extreme points observed in the neighborhood of $x = 0$ bring information on the behavior of f in the neighborhood of $x = 1$ and *vice versa*. Thus, an estimation relying on the trigonometric basis is specially well adapted.
This paper is organized as follows. Section 2 is devoted to the definition of the estimator and Section 3 presents some basic results on extreme values and Dirichlet's kernels. The mean integrated square convergence of the estimate is briefly studied in Section 4 and the asymptotic normality is established in Section 5. In Section 6 a very simple bias correction is proposed, illustrated in Section 7 by a simulation.

2 Definition of the estimator

2.1 Preliminaries

Let N be a Poisson process with a mean measure $\mu = c\lambda$, where the intensity parameter c is unknown, λ is the Lebesgue measure, and the support of N is given by:

$$S = \{(x, y) \in \mathbf{R}^2 \mid 0 \leq x \leq 1 \; ; \; 0 \leq y \leq f(x)\}. \tag{2.1}$$

We assume that f is measurable and satisfies

$$0 < m = \inf_{[0,1]} f \leq M = \sup_{[0,1]} f < +\infty, \tag{2.2}$$

which entails the square integrability of f on $[0, 1]$. In the sequel, we will introduce extra hypothesis on f as needed.
Let $(e_i)_{i\in\mathbf{N}}$ be an orthonormal basis of $L^2([0, 1])$. The expansion of f with respect to the basis is supposed to be both L^2 and pointwise convergent to f on $[0, 1]$:

$$\forall x \in [0, 1], \;\; f(x) = \sum_{i=0}^{+\infty} a_i e_i(x), \tag{2.3}$$

with

$$\forall i \geq 0, \;\; a_i = \int_0^1 e_i(t) f(t)\, dt. \tag{2.4}$$

We denote by K_n the Dirichlet's kernel associated to the orthonormal basis $(e_i)_{i\in\mathbf{N}}$ defined by

$$K_n(x, y) = \sum_{i=0}^{h_n} e_i(x) e_i(y), \qquad (x, y) \in [0, 1]^2, \tag{2.5}$$

where (h_n) is an increasing sequence of integers. The trigonometric basis will provide us with an important example to illustrate our convergence results. It is defined by

$$e_0(x) = 1, \qquad e_{2k-1}(x) = \sqrt{2} \cos 2k\pi x, \qquad e_{2k}(x) = \sqrt{2} \sin 2k\pi x, \quad k \geq 1, \tag{2.6}$$

and we shall suppose for convenience that h_n is even. This leads to

$$K_n(x, y) = \left| \begin{array}{ll} \dfrac{\sin\,(1 + h_n)\pi(x - y)}{\sin \pi(x - y)} & x \neq y, \\ 1 + h_n & x = y. \end{array} \right. \tag{2.7}$$

The speed at which the sequence (a_k) decreases to 0 is linked to the regularity of f. In the case of the trigonometric basis, if f is a function of class C^2 then

$$a_k = O\left(k^{-2}\right), \tag{2.8}$$

(see [4]). The estimator is built in two steps. First, in subsection 2.2, f is approximated by a sequence (f_n) obtained from its expansion with respect to the orthogonal basis. Then, in subsection 2.3, an estimator $\hat{f}_n$ of f_n is proposed.

2.2 Approximation of f

Let (k_n) be an increasing sequence of non-negative integers such that $k_n = o(n)$. Divide S into k_n cells $D_{n,r}$ where:

$$D_{n,r} = \{ (x,y) \in S \mid x \in I_{n,r} \}, \qquad I_{n,r} = \left[\frac{r-1}{k_n}, \frac{r}{k_n}\right[, \quad r = 1, \dots, k_n. \tag{2.9}$$

Each coefficient a_i is approximated by discretizing (2.4) according to:

$$a_{i,k_n} = \sum_{r=1}^{k_n} e_i(x_r)\lambda(D_{n,r}), \qquad x_r = \frac{2r-1}{2k_n}. \tag{2.10}$$

Then, the expansion (2.3) is truncated to the h_n first terms leading to

$$f_n(x) = \sum_{i \le h_n} a_{i,k_n} e_i(x), \qquad x \in [0,1], \tag{2.11}$$

which can be written in terms of the Dirichlet's kernel:

$$f_n(x) = \sum_{r=1}^{k_n} K_n(x_r, x)\lambda(D_{n,r}), \qquad x \in [0,1]. \tag{2.12}$$

Let us emphasize that the approximation f_n of f only depends on the basis $(e_i)_{i\in\mathbf{N}}$ through its Dirichlet's kernel. The next step towards the definition of the estimator consists of estimating $\lambda(D_{n,r})$.

2.3 Estimation of f_n

Let N^{*n} denote the superposition $N_1 + \cdots + N_n$ of n independent copies of the point process N, and Σ_n the set of points generated by N^{*n}. For $r = 1, \dots, k_n$, consider the maximum $X^\star_{n,r}$ of the second coordinates of the set of points $\Sigma_{n,r} = \Sigma_n \cap D_{n,r}$. Of course, if $\Sigma_{n,r} = \emptyset$, set $X^\star_{n,r} = 0$. Then, $\lambda(D_{n,r})$ can be estimated by $X^\star_{n,r}/k_n$. This leads to an estimate $\hat{a}_{i,k_n}$ of a_{i,k_n} defined as:

$$\hat{a}_{i,k_n} = \sum_{r=1}^{k_n} e_i(x_r)\frac{X^\star_{n,r}}{k_n}, \qquad 1 \le i \le h_n, \tag{2.13}$$

and consequently to an estimate $\hat{f}_n(x)$ of $f_n(x)$ via:

$$\hat{f}_n(x) = \sum_{i \le h_n} \hat{a}_{i,k_n} e_i(x) = \sum_{r=1}^{k_n} K_n(x_r, x)\frac{X^\star_{n,r}}{k_n}. \tag{2.14}$$

Two remarks can be made. First, the estimator does not require knowledge of c, which ensures a wide range of applications. Second, the estimator is written as a linear combination of the maxima $X^\star_{n,r}$ involving the Dirichlet's kernel. Thus, the analysis of the behavior of $\hat{f}_n$ will rely on both studies of the Dirichlet's kernel features and of the maxima's stochastic properties. This is the topic of the next section.

3 Basic results

3.1 Bounds on the Dirichlet's kernel

For $x \in [0,1]$ and $j \in \{1,2,3\}$ define

$$B_{n,j}(x) = \left(\sum_{r=1}^{k_n} |K_n(x_r, x)|^j\right)^{1/j}$$

and

$$B_{n,\infty}(x) = \max_{1 \le r \le k_n} |K_n(x_r, x)|.$$

In what follows, some theorems involving conditions on $B_{n,j}(x)$ for $j \in \{1, 2, 3, \infty\}$ are fiven. Some of these conditions follow easily from the properties of K_n, see [11]. In the following lemma two properties which are less straightforward are given for the special case of the trigonometric basis.

Lemma 1 *Suppose K_n is the Dirichlet's kernel associated to the trigonometric basis.*

(i) If $h_n = o(k_n)$ then $\sup_x B_{n,1}(x) = O(k_n \ln h_n)$,

(ii) If $h_n \ln h_n = o(k_n)$, then for all $x \in [0, 1]$, $B_{n,2}(x) \sim (k_n h_n)^{1/2}$.

The proof is postponed to the Appendix.

3.2 Maxima stochastic properties

In the sequel, we write:

$$\lambda(D_{n,r}) = \lambda_{n,r}, \ \min_{x \in I_{n,r}} f(x) = m_{n,r}, \ \max_{x \in I_{n,r}} f(x) = M_{n,r}.$$

Recall that $X^\star_{n,r}$ is the maximum of the second coordinates of the set of points $\Sigma_{n,r}$.
Noticing that, for $0 \le x \le m_{n,r}$,

$$P(X^\star_{n,r} \le x) = P(N^{\star n}(D_{n,r} \setminus (I_{n,r} \times [0, x])) = 0), \tag{3.1}$$

we easily obtain the distribution function $F_{n,r}(x) = P(X^\star_{n,r} \le x)$ on $[0, m_{n,r}]$:

$$F_{n,r}(x) = \exp\left[\frac{nc}{k_n}(x - k_n\lambda_{n,r})\right]. \tag{3.2}$$

Straight forward calculations lead to the following expansions for the mathematical expectation and the variance of $X^\star_{n,r}$, where the knowledge of the C^1-regularity of f compensates for the lack of a precise expression for $F_{n,r}$ on $]m_{n,r}, M_{n,r}[$.

Lemma 2 *Suppose f is a function of class C^1, $n = o(k_n^2)$ and $k_n = o(n/\ln n)$. Then,*

(i) $E(X^\star_{n,r}) = k_n\lambda_{n,r} - \dfrac{k_n}{nc} + o\left(\dfrac{n}{k_n^3}\right)$,

(ii) $Var(X^\star_{n,r}) \sim \dfrac{k_n^2}{n^2c^2}$.

The proof of the following lemma, which is more difficult, is postponed to the appendix.

Lemma 3 *Suppose f is a function of class C^1 and $k_n = o(n/\ln n)$. Let $(t_{n,r})$ be a sequence such that $t_{n,r} = o(n/k_n)$ and $t_{n,r} = o(k_n^3/n)$. Then, the characteristic function of $(X^\star_{n,r} - m_{n,r})$ can be written at point $t_{n,r}$ as:*

$$\phi_{n,r}(t_{n,r}) = \frac{1 + it_{n,r}\frac{k_n}{nc}\bar{F}_{n,r}(m_{n,r}) + o\left(|t_{n,r}|\,\frac{n}{k_n^3}\right) + o(n^{-s})}{1 + it_{n,r}\frac{k_n}{nc}},$$

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

4 Estimate convergences

We refer to [11] for a careful study of the bias convergences $\| f_n - f \|_2 \to 0$ and $\| f_n - f \|_\infty \to 0$. Then, we just have to consider $(\hat{f}_n - f_n)$. A complete investigation for mean integrated convergence, mean uniform convergence, L^2-almost complete convergence and uniform almost complete convergence is available at our website. In order to avoid lengthy developments, we just propose the following basic result.

Theorem 1 *Suppose f is a C^1 function. If $\| f_n - f \|_2 = o(1)$ and if $\sup_x B_{n,1}(x) = o\left(n^2/k_n\right)$, then $E(\| \hat{f}_n - f \|_2) = o(1)$.*

Proof : Introducing the random variable $Y_{n,r} = (X^\star_{n,r}/k_n) - \lambda_{n,r}$, we have

$$\begin{aligned} \mathrm{E}\left(\| \hat{f}_n - f_n \|_2^2\right) &= \mathrm{E}\left(\sum_{r,s} Y_{n,r} K_n(x_r, x_s) Y_{n,s}\right), & (4.1)\\ &\leq 2\sum_{r,s} |K_n(x_r, x_s)| \left(\mathrm{E}(Y^2_{n,r}) + \mathrm{E}(Y^2_{n,s})\right) & (4.2)\\ &\leq 4 \sup_x B_{n,1}(x) \sum_{r=1}^{k_n} \mathrm{E}(Y^2_{n,r}), & (4.3)\end{aligned}$$

and the result follows from Lemma 2. ■

Corollary 1 *If K_n is the Dirichlet's kernel associated to the trigonometric basis and f is a C^1 function then $h_n \ln^{1/2} h_n = o(k_n)$ and $k_n(\ln h_n)^{1/2} = o(n)$ are sufficient conditions for $E(\| \hat{f}_n - f \|_2) = o(1)$.*

Proof : From [11], Proposition 3, $h_n(\ln h_n)^{1/2} = o(k_n)$ entails $\| f_n - f \|_2 = o(1)$. Moreover, from Lemma 1(i), if $k_n(\ln h_n)^{1/2} = o(n)$, then

$$\sup_n B_{n,1}(x) = O(k_n \ln h_n) = o\left(n^2/k_n\right), \tag{4.4}$$

and the conclusion follows. ■

5 Asymptotic distribution

In this section, we present a limit theorem for the distribution of $(\hat{f}_n - \mathrm{E}\,\hat{f}_n)$. A similar result is not available for $(\hat{f}_n - f)$ without reducing the bias, which is done in the next section.

Theorem 2 *Suppose that f is a function of class C^1. If $k_n = o(n/\ln n)$, $n = o\left(k_n^{3/2}\right)$ and $B_{n,\infty}(x) = o(B_{n,2}(x))$, then $(nc/B_{n,2}(x))(\hat{f}_n(x) - E(\hat{f}_n(x)))$ converges in distribution to a standard Gaussian variable for all $x \in [0,1]$.*

Proof : Denote $\alpha_{n,r} = m_{n,r} - k_n/(nc)$ and for all $x \in [0,1]$ introduce $\psi_{n,x}$ the characteristic function of $(nc/B_{n,2}(x))(\hat{f}_n(x) - \mathrm{E}(\hat{f}_n(x)))$. It expands as:

$$\begin{aligned} \psi_{n,x}(t) &= \exp it\left(\frac{nc}{k_n B_{n,2}(x)} \sum_{r=1}^{k_n} K_n(x_r, x)(\alpha_{n,r} - \mathrm{E}(X^\star_{n,r}))\right) & (5.1)\\ &\times\ \mathrm{E}\left[\exp it\left(\frac{nc}{k_n B_{n,2}(x)} \sum_{r=1}^{k_n} K_n(x_r, x)(X^\star_{n,r} - \alpha_{n,r})\right)\right]. & (5.2)\end{aligned}$$

Consider first the argument of (5.1):

$$T_n(x) = \frac{nc}{k_n B_{n,2}(x)} \sum_{r=1}^{k_n} K_n(x_r, x)(\alpha_{n,r} - \mathrm{E}(X^\star_{n,r})), \tag{5.3}$$

and show it converges to 0. By the Cauchy-Schwartz inequality:

$$|T_n(x)| \le \frac{nc}{k_n} \frac{B_{n,1}(x)}{B_{n,2}(x)} \max_r \left|\alpha_{n,r} - \mathrm{E}(X^\star_{n,r})\right| \le \frac{nc}{k_n^{1/2}} \max_r \left|\alpha_{n,r} - \mathrm{E}(X^\star_{n,r})\right|. \tag{5.4}$$

Now Lemma 2 entails

$$|T_n(x)| \le \frac{n}{k_n^{1/2}} \left(\max_r (k_n \lambda_{n,r} - m_{n,r}) + o\left(\frac{n}{k_n^3}\right)\right) = o\left(\frac{n}{k_n^{3/2}}\right), \tag{5.5}$$

which converges to 0. Thus, introducing $t_{n,r} = \frac{ncK_n(x_r,x)t}{k_n B_{n,2}(x)}$, we obtain:

$$\begin{aligned}\psi_{n,x}(t) &\sim \exp it \left(\frac{1}{B_{n,2}(x)} \sum_{r=1}^{k_n} K_n(x_r, x)\right) \mathrm{E}\left[\exp it \left(\frac{nc}{k_n B_{n,2}(x)} \sum_{r=1}^{k_n} K_n(x_r, x)(X^\star_{n,r} - m_{n,r})\right)\right] \\ &= \exp it \left(\frac{1}{B_{n,2}(x)} \sum_{r=1}^{k_n} K_n(x_r, x)\right) \prod_{r=1}^{k_n} \phi_{n,r}(t_{n,r}). \end{aligned} \tag{5.6}$$

To apply Lemma 3, we have to verify that $t_{n,r} = o(n/k_n)$ and $t_{n,r} = o(k_n^3/n)$. The first condition is satisfied since

$$\left|t_{n,r} \frac{k_n}{nc}\right| \le |t| \frac{B_{n,\infty}(x)}{B_{n,2}(x)} = o(1). \tag{5.7}$$

The second condition is satisfied as well:

$$\left|t_{n,r} \frac{n}{k_n^3}\right| = \left|t_{n,r} \frac{k_n}{n}\right| \frac{n^2}{k_n^4} = o(1). \tag{5.8}$$

The characteristic function can be seen to be of the order:

$$\psi_{n,x}(t) \sim \frac{\exp it \left(\dfrac{1}{B_{n,2}(x)} \displaystyle\sum_{r=1}^{k_n} K_n(x_r, x)\right)}{\displaystyle\prod_{r=1}^{k_n} \left(1 + it_{n,r} \frac{k_n}{nc}\right)} \tag{5.9}$$

$$\times \prod_{r=1}^{k_n} \left(1 + it_{n,r} \frac{k_n}{nc} \bar{F}_{n,r}(m_{n,r}) + o\left(|t_{n,r}| \frac{n}{k_n^3}\right) + o(n^{-s})\right). \tag{5.10}$$

Consider first the logarithm of term (5.9). A second-order Taylor expansion yields

$$J_n^{(1)}(x) = \sum_{r=1}^{k_n} \left[it_{n,r} \frac{k_n}{nc} - \ln\left(1 + it_{n,r} \frac{k_n}{nc}\right)\right] = -\frac{t^2}{2} + O\left(\frac{B^3_{n,3}(x)}{B^3_{n,2}(x)}\right). \tag{5.11}$$

Since $B^3_{n,3}(x)/B^3_{n,2}(x) \le B_{n,\infty}(x)/B_{n,2}(x) = o(1)$, it follows that $J_n^{(1)}(x) \to -t^2/2$ as $n \to \infty$. Finally, consider the logarithm of (5.10):

$$J_n^{(2)} = \sum_{r=1}^{k_n} \ln(1 + u_{n,r}), \text{ with } u_{n,r} = it_{n,r} \frac{k_n}{nc} \bar{F}_{n,r}(m_{n,r}) + o\left(|t_{n,r}| \frac{n}{k_n^3}\right) + o(n^{-s}). \tag{5.12}$$

Observe that $\max_r |u_{n,r}|$ converges to 0 from (5.7) and (5.8). Thus, for n large enough $|u_{n,r}| < 1/2$ uniformly in r and the classical identity $|\ln(1+u_{n,r})| < |u_{n,r}|$ yields

$$\left|J_n^{(2)}\right| \leq |t| \frac{B_{n,1}(x)}{B_{n,2}(x)} \frac{nc}{k_n^2} + o\left(\frac{B_{n,1}(x)}{B_{n,2}(x)}\right) \frac{n^2}{k_n^4} + o\left(k_n n^{-s}\right). \tag{5.13}$$

Therefore, the Cauchy-Schwartz inequality leads to $J_n^{(2)} \to 0$ and $\psi_{n,x}(t) \to e^{-t^2/2}$ as $n \to \infty$. ■

Corollary 2 *Suppose K_n is the Dirichlet's kernel associated to the trigonometric basis and f is a C^1 function. If $h_n = o(k_n)$, $k_n = o(n/\ln n)$ and $n = o\left(k_n^{3/2}\right)$, then, for all $x \in [0,1]$, $nc(h_n k_n)^{-1/2}(\hat{f}_n(x) - E(\hat{f}_n(x)))$ converges in distribution to a standard Gaussian variable.*

Proof : From (2.7), $B_{n,\infty}(x) \leq \| K_n \|_\infty = 1 + h_n$, and from Lemma 1(ii), $B_{n,2}(x) \sim (h_n k_n)^{1/2}$ so that $B_{n,\infty}(x)/B_{n,2}(x) = o\left((h_n/k_n)^{1/2}\right) = o(1)$. ■

Possible choices of k_n and h_n sequences in Corollary 2 are $k_n = n^{2/3}(\ln n)^\varepsilon$ and $h_n = (\ln n)^\varepsilon$ for $\varepsilon > 0$ arbitrary small. These choices entail $n(h_n k_n)^{-1/2} = n^{2/3}(\ln n)^{-\varepsilon}$.

6 Reducing the bias

The bias can be decomposed as follows:

$$\mathrm{E}(\hat{f}_n - f) = (\mathrm{E}\,\hat{f}_n - f_n) + (f_n - f), \tag{6.1}$$

where the first term in the sum is the statistical part of the bias, and the second term is the systematic part of the bias. First, consider the irreductible bias $(f_n - f)$. In order to obtain a limit distribution for $(nc/B_{n,2}(x))(\hat{f}_n(x) - f(x))$, we need to satisfy the condition

$$\lim_{n\to\infty} \frac{nc}{B_{n,2}(x)}(f_n(x) - f(x)) = 0. \tag{6.2}$$

Introduce $S_n(f) = \sum_{i=0}^{h_n} a_i e_i$. Equation (3.12) in [11] provides sharp bounds for $(S_n(f) - f_n)$, so that the question reduces to considering $(S_n(f) - f)$, which only depends on the basis. We shall see that the trigonometric basis satisfies (6.2) under reasonable conditions on h_n and k_n. In the case of a general C^1 basis, we shall take (6.2) as a condition.
Now, it follows from Lemma 2(i) that

$$\mathrm{E}\,\hat{f}_n(x) - f_n(x) = \sum_{r=1}^{k_n} K_n(x_r, x)\left(\frac{\mathrm{E}(X_{n,r}^\star)}{k_n} - \lambda_{n,r}\right) \tag{6.3}$$

presents a negative component which should be eliminated. To this end, for $r = 1, \ldots, k_n$, define $Z_{n,r}^\star$ by $Z_{n,r}^\star = 0$ if $\Sigma_{n,r} = \emptyset$ and $Z_{n,r}^\star$ is the infimum of the second coordinates of the points of $\Sigma_{n,r}$ otherwise. Then, the random variable

$$Z_n = \frac{1}{k_n}\sum_{r=1}^{k_n} Z_{n,r}^\star, \tag{6.4}$$

has a mathematical expectation

$$\mathrm{E}(Z_n) = \frac{k_n}{nc} + o\left(\frac{n}{k_n^3}\right), \tag{6.5}$$

and a variance

$$\mathrm{Var}(Z_n) \sim \frac{k_n}{n^2c^2}, \tag{6.6}$$

(use Lemma 2 for $Z^\star_{n,r}$). We define a corrected estimate by

$$\tilde{f}_n(x) = \sum_{r=1}^{k_n} K_n(x_r, x)\left(\frac{X^\star_{n,r} + Z_n}{k_n}\right) = \hat{f}_n(x) + \hat{g}_n(x). \tag{6.7}$$

Lemma 4 *Suppose f is a function of class C^1, $n = o\left(k_n^2\right)$ and $k_n = o\left(n/\ln n\right)$. Then,*

$$\frac{nc}{B_{n,2}(x)}\left|E(\tilde{f}_n(x)) - f_n(x)\right| = o\left(n^2/k_n^{7/2}\right),\ \forall x \in [0,1].$$

If, moreover, for all $x \in [0,1]$, $k_n = o\left(B^2_{n,2}(x)\right)$ and (6.2) holds, then

$$\frac{nc}{B_{n,2}(x)}(\hat{g}_n(x) - E(\hat{g}_n(x))) = o_P(1).$$

Proof : We have

$$\begin{aligned}\frac{nc}{B_{n,2}(x)}\left|\mathrm{E}(\tilde{f}_n(x)) - f_n(x)\right| &\leq \frac{nc}{k_n}\frac{B_{n,1}(x)}{B_{n,2}(x)}\max_{1\leq r\leq k_n}\left|\mathrm{E}(X^\star_{n,r}) + \mathrm{E}(Z_n) - k_n\lambda_{n,r}\right| && (6.8)\\ &= \frac{nc}{k_n}\frac{B_{n,1}(x)}{B_{n,2}(x)} O\left(\frac{n}{k_n^3}\right) = O\left(\frac{n^2}{k_n^{7/4}}\right), && (6.9)\end{aligned}$$

from (6.5), Lemma 2 and the Cauchy-Schwartz inequality.
Now, applying (6.2) to the constant function $f = 1$ yields

$$\frac{1}{k_n}\sum_{r=1}^{k_n} K_n(x_r, x) \to 1, \tag{6.10}$$

as $n \to \infty$. Therefore, $\mathrm{Var}(\hat{g}_n(x)) \sim \mathrm{Var}(Z_n) \sim k_n/(n^2c^2)$ and then

$$\mathrm{Var}\left(\frac{nc}{B_{n,2}(x)}(\hat{g}_n(x) - \mathrm{E}(\hat{g}_n(x)))\right) \sim \frac{k_n}{B^2_{n,2}(x)} \tag{6.11}$$

which converges to 0. ∎

Theorem 3 *Suppose that f is a function of class C^1. If the following conditions are verified*

(i) $k_n = o\left(n/\ln n\right)$, $n = o\left(k_n^{3/2}\right)$,

(ii) for all $x \in [0,1]$, $\max(k_n^{1/2}, B_{n,\infty}(x)) = o\left(B_{n,2}(x)\right)$,

(iii) for all $x \in [0,1]$, $\dfrac{nc}{B_{n,2}(x)}\left|f_n(x) - f(x)\right| = o(1)$,

then, for all $x \in [0,1]$, $(nc/B_{n,2}(x))(\tilde{f}_n(x) - f(x))$ converges in distribution to a standard Gaussian variable.

The proof is a simple consequence of the expansion

$$(\tilde{f}_n - f) = (\hat{f}_n - \mathrm{E}(\hat{f}_n)) + (\hat{g}_n - \mathrm{E}(\hat{g}_n)) + (E(\tilde{f}_n) - f_n) + (f_n - f), \tag{6.12}$$

and of Theorem 2 and Lemma 4.

Corollary 3 *Suppose K_n is the Dirichlet's kernel associated to the trigonometric basis and f is a C^2 function. If $h_n \ln h_n = o(k_n)$, $n = o\left(h_n^{3/2} k_n^{1/2}\right)$, $n h_n^{1/2} \ln h_n = o\left(k_n^{3/2}\right)$ and $k_n = o(n/\ln n)$ then, for all $x \in [0,1]$, $nc(h_n k_n)^{-1/2}(\tilde{f}_n(x) - f(x))$ converges in distribution to a standard Gaussian variable.*

Proof : Conditions (i), (ii) of Theorem 3 are verified. Consider (iii). On using (2.8) we have

$$\begin{aligned} nc(h_n k_n)^{-1/2} \left| S_n(f)(x) - f(x) \right| &\leq nc(h_n k_n)^{-1/2} \sum_{i \geq h_n} |a_i||e_i(x)| \\ &\leq \sqrt{2} nc(h_n k_n)^{-1/2} \sum_{i \geq h_n} |a_i| \\ &\leq \sqrt{2} nc(h_n k_n)^{-1/2} \sum_{i \geq h_n} i^{-2}. \end{aligned} \tag{6.13}$$

A straightforward calculation yields

$$\frac{nc}{B_{n,2}(x)} \left| S_n(f)(x) - f(x) \right| = O\left(n h_n^{-3/2} k_n^{-1/2}\right). \tag{6.14}$$

From [11], equations (3.11) and (3.12),

$$\left| S_n(f)(y) - f_n(y) \right| = O\left(\frac{1}{k_n} \int_0^1 \left| \frac{\partial K_n}{\partial x}(v, y) \right| dv\right) \tag{6.15}$$

and

$$\int_0^1 \left| \frac{\partial K_n}{\partial x}(v, y) \right| dv = O\left(h_n \ln h_n\right). \tag{6.16}$$

Therefore,

$$\frac{nc}{B_{n,2}(y)} \left| S_n(f)(y) - f_n(y) \right| = O\left(n k_n^{-3/2} h_n^{1/2} \ln h_n\right) \tag{6.17}$$

and (6.14) with (6.17) conclude the proof. ■

Possible choices of k_n and h_n sequences in Corollary 3 are $k_n = n^{4/5}(\ln n)^{3/5}(\ln\ln n)^{\varepsilon}$ and $h_n = n^{2/5}(\ln n)^{-1/5}(\ln\ln n)^{\varepsilon}$ for $\varepsilon > 0$ arbitrary small. These choices entail $n(h_n k_n)^{-1/2} = n^{2/5}(\ln n)^{-1/5}(\ln\ln n)^{-\varepsilon}$.

7 Illustration on a simulation

We consider the following 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 (2.1) with

$$f(x) = 3 + \frac{\sin(2\pi x)}{1.1 + \cos(2\pi x)}. \tag{7.1}$$

The function is represented on Figure 1. We consider an expansion of order $h_n = 16$ of f in the trigonometric basis, where each coefficient is approximated by dividing $[0,1]$ into $k_n = 32$ intervals. The estimate $\hat{f}_n$ is represented on Figure 2 where it is compared with $\tilde{f}_n$ obtained by reducing the bias as described in the previous section.

(Figures 1 and 2 here)

8 Conclusion and further developments

In this paper, we showed how the convergence results established in [7] in the case of the Haar basis can be adapted for any C^1 basis under some assumptions on the Dirichlet's kernel behavior. We have emphasized that the estimator and these assumptions only depend on the Dirichlet's kernel of the basis. This suggests to define a new estimator based on a Parzen-Rosenblatt kernel.

Appendix

Proof of Lemma 1

(i) In the sequel, we shall use the inequality found in [11], equation (2.11):

$$\left|\frac{\sin(pu)}{\sin(u)}\right| \le p\mathbf{1}_{[0,\delta]}(|u|) + \frac{\pi}{2|u|}\mathbf{1}_{[\delta,\pi/2]}(|u|), \tag{8.1}$$

for all $p > 0$, $0 < \delta < \pi/2$ and $|u| < \pi/2$. Taking account of the periodicity and symmetry properties of the trigonometric kernel, it suffices to study

$$\sup_{x\in[0,1/k_n]} \frac{2}{k_n} \sum_{r=1}^{[k_n/2]+1} |K_n(x_r, x)|, \tag{8.2}$$

where $[u]$ denotes the integer part of u. Let us write

$$\frac{1}{k_n}\sum_{r=1}^{[k_n/2]+1} K_n(x_r, x) = \frac{1}{k_n}\sum_{r=1}^{[\gamma_n]} K_n(x_r, x) + \frac{1}{k_n}\sum_{r=[\gamma_n]+1}^{[k_n/2]+1} K_n(x_r, x), \tag{8.3}$$

with $\gamma_n = k_n/(h_n+1)$, and consider the two terms separately.

- Introduce $\delta = (\pi/k_n)([\gamma_n] - 1/2)$. For $r = 1, \dots, [\gamma_n]$, we have $\pi(x_r - x) \le \delta$ and thus (8.1) yields $|K_n(x_r, x)| \le 1 + h_n$ which gives in turn

$$\frac{1}{k_n}\sum_{r=1}^{[\gamma_n]} |K_n(x_r, x)| \le 1. \tag{8.4}$$

- For $r = [\gamma_n]+1, \dots, [k_n/2]+1$, we have $\pi(x_r - x) \ge \delta$ and consequently (8.1) yields

$$\frac{1}{k_n}\sum_{r=[\gamma_n]+1}^{[k_n/2]+1} |K_n(x_r, x)| \le \frac{1}{k_n}\sum_{r=[\gamma_n]+1}^{[k_n/2]+1} \frac{1}{2(x_r - x)} \le \frac{1}{2}\frac{1}{k_n}\sum_{r=[\gamma_n]}^{[k_n/2]} \frac{1}{\frac{1}{k_n}\left(r - \frac{1}{2}\right)}. \tag{8.5}$$

  Therefore,

$$\frac{1}{k_n}\sum_{r=[\gamma_n]+1}^{[k_n/2]+1} |K_n(x_r, x)| \le \frac{1}{2}\int_{\frac{\delta}{\pi}}^{\frac{1}{2}+\frac{1}{2k_n}} \frac{du}{u} \le \frac{1}{2}\ln(4(h_n+1)), \tag{8.6}$$

  for $k_n > 2(h_n+1)$.

Finally, collecting (8.4) and (8.6), we obtain

$$\sup_{x\in[0,1]} \frac{1}{k_n}\sum_{r=1}^{k_n} |K_n(x_r, x)| \le 2 + \ln(4(h_n+1)). \tag{8.7}$$

(ii) From [11], equation (4.14),

$$\left|\frac{B_{n,2}^2(x)}{k_n K_n(x,x)} - 1\right| \le \frac{\| K_n \|_\infty}{K_n(x,x)} \frac{1}{k_n} \sup_x \int_0^1 \left|\frac{\partial K_n}{\partial y}(x,v)\right| dv. \tag{8.8}$$

In the case of the trigonometric basis $K_n(x,x) = \| K_n \|_\infty = 1 + h_n$ (see (2.7)) and from [11], equation (3.12) we have

$$\int_0^1 \left|\frac{\partial K_n}{\partial y}(x,v)\right| dv = O\left(h_n \ln h_n\right) = o\left(k_n\right). \tag{8.9}$$

The conclusion follows from (8.8) and (8.9).

■

Proof of Lemma 3

Consider the expansion

$$\mathrm{E}(e^{it_{n,r}X^\star_{n,r}}) = P(X^\star_{n,r}=0) + \int_0^{m_{n,r}} e^{ixt_{n,r}} F'_{n,r}(x)dx + \int_{m_{n,r}}^{M_{n,r}} e^{ixt_{n,r}} F_{n,r}(dx). \tag{8.10}$$

The first and second term can be computed explicitely since (3.2) provides an expression of $F_{n,r}$ on $[0, m_{n,r}]$:

$$\mathrm{E}(e^{it_{n,r}X^\star_{n,r}}) = e^{-nc\lambda_{n,r}} + \frac{e^{it_{n,r}m_{n,r}}F_{n,r}(m_{n,r}) - e^{-nc\lambda_{n,r}}}{1+it_{n,r}\frac{k_n}{nc}} + \int_{m_{n,r}}^{M_{n,r}} e^{ixt_{n,r}} F_{n,r}(dx). \tag{8.11}$$

The third term can be expanded as

$$\int_{m_{n,r}}^{M_{n,r}} e^{ixt_{n,r}} F_{n,r}(dx) = e^{it_{n,r}m_{n,r}}\bar{F}_{n,r}(m_{n,r}) + \int_{m_{n,r}}^{M_{n,r}} (e^{ixt_{n,r}} - e^{it_{n,r}m_{n,r}})F_{n,r}(dx), \tag{8.12}$$

with $\left|e^{ixt_{n,r}} - e^{it_{n,r}m_{n,r}}\right| \leq (M_{n,r}-m_{n,r})\,|t_{n,r}|$. Thus

$$\begin{aligned}\left|\int_{m_{n,r}}^{M_{n,r}} (e^{ixt_{n,r}} - e^{it_{n,r}m_{n,r}})F_{n,r}(dx)\right| &\leq (M_{n,r}-m_{n,r})\,|t_{n,r}|\,\bar{F}_{n,r}(m_{n,r}) \\ &\leq (M_{n,r}-m_{n,r})\,|t_{n,r}|\,\frac{nc}{k_n}(k_n\lambda_{n,r}-m_{n,r}) \\ &= O\left(|t_{n,r}|\,\frac{n}{k_n^3}\right). \end{aligned} \tag{8.13}$$

Collecting (8.11)–(8.13), and remarking that $k_n = o\left(n/\ln n\right)$ yields $\left|\frac{k_n}{nc}e^{-nc\lambda_{n,r}}\right| = O\left(n^{-s}\right)$ concludes the proof. ■

References

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

[2] Bosq, D. (1977) Contribution à la théorie de l'estimation fonctionnelle. *Publications de l'Institut de Statistique de l'Université de Paris*, **XIX**, 2, 1–96.

[3] 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.

[4] Gasquet, C., Witomski, P. (1990) *Analyse de Fourier et Applications*, Masson.

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

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

[7] Girard, S., Jacob, P. (1999) Extreme values and Haar series of point processes boundaries. *Technical report ENSAM-INRA-UM2*, **99-08**.

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

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

[10] 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.

[11] 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.

[12] Korostelev, A., Simar, L. and Tsybakov, A. B. (1995) Efficient estimation of monotone boundaries. *The Annals of Statistics*, **23**, 476–489.

[13] 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.

[14] Mammen, E. and Tsybakov, A. B. (1995) Asymptotical minimax recovery of set with smooth boundaries. *The Annals of Statistics*, **23**(2), 502–524.

[15] Renyi, A. and Sulanke, R. (1963) Uber die konvexe Hülle von n zufälligen gewählten Punkten. *Z. Wahrscheinlichkeitstheorie verw. Geb.* **2**, 75–84.

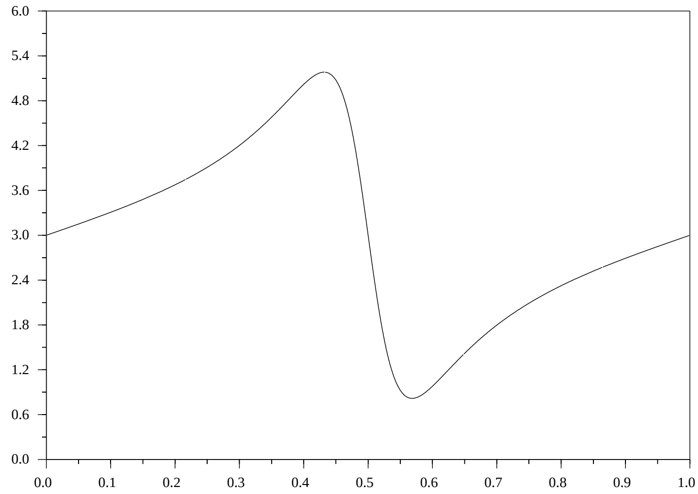

Figure 1: Function f to estimate.

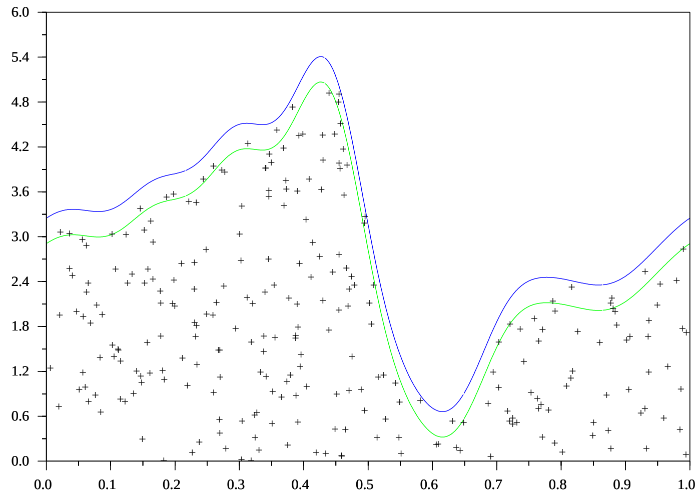

Figure 2: Comparison between the two estimates $\hat{f}_n$ (gray line) and $\tilde{f}_n$ (dark line).