Asymptotic normality of the L_1- error for Geffroy's estimate of Poisson point process boundaries

Stéphane Girard[1] & Pierre Jacob[2]

[1] SMS/LMC, Université Grenoble 1,

BP 53, 38041 Grenoble cedex 9, France.

Stephane.Girard@imag.fr

[2] Laboratoire de Probabilités et Statistique, Université Montpellier 2,

Place Eugène Bataillon, 34095 Montpellier Cedex 5, France.

jacob@math.univ-montp2.fr

Abstract

In this Note, the L_1- error of Geffroy's estimate of a Poisson point process boundary is shown to be asymptotically normal. We give conditions such that the asymptotic mean and variance do not depend on the unknown boundary. This result is illustrated by a simulation.

Key Words : Extreme values, Poisson process, central limit theorem.

1 Introduction.

In the literature, many papers address the problem of estimating a set D given a finite random set of points Σ_n drawn from the interior. The first proposal is given by Geffroy [5] in the case when the set can be written

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

and where f is an unknown function. Thus, estimating D reduces to estimating f, and the proposed estimate $\hat{f}_n$ is a kind of histogram based on a set of extreme values of Σ_n (see Section 2 for more details). Geffroy gives convergence conditions and Gumbel limit distributions for the maximal error between $\hat{f}_n$ and f even for non uniform samples. It has been shown in [9], Theorem 4.1.1, that $\hat{f}_n$ can be tuned in order to be minimax for estimating a continuous function f with bounded derivative in the case where Σ_n is a random sample uniform on D. Here, we establish in Section 2 the asymptotic normality of the L_1- error between Geffroy's estimate $\hat{f}_n$ and the true function f in the case where Σ_n is a Poisson point process.

Similar results have been proved in the context of density estimation: The asymptotic distribution of the L_1- distance between $\hat{f}_n$ and $E(\hat{f}_n)$ is established in [1] for univariate densities. This result is precised in [3] by considering the L_1- distance between $\hat{f}_n$ and the unknown density f. The multivariate case is presented in [2]. All the results are established for samples via a Poissonisation technique. Thus, the results presented below should be considered as the first step towards the study of the sample case.

The proof is based on the independence properties of the Poisson process and on precise expansions of the extreme values moments, see Section 3. It will be the starting point for similar works on other estimates obtained by smoothing $\hat{f}_n$ [4, 7, 8, 6].

2 Main result.

For all $n > 0$, let N_n be a stationary Poisson point process on $\mathbb{R}^2$ with intensity rate nc_n, where (c_n) is a positive sequence tending to some $c_\infty > 0$ as n goes to infinity. Actually, suppose that we merely observe the truncated point process $N_n(.\cap D)$ with

$$D = \{(x,y) \in \mathbb{R}^2 \mid 0 \le x \le 1 \; ; \; 0 \le y \le f(x)\}, \tag{2.1}$$

and where f is a α-Lipschitzian $(0 < \alpha \le 1)$ positive function. The set of points associated to N_n is denoted by Σ_n. Let (k_n) be an increasing sequence of integers. Introduce $(I_{n,r})$, $r = 1, \ldots, k_n$ the associated equidistant partition of the unit interval, $(D_{n,r})$, $r = 1, \ldots, k_n$ the corresponding partition of D into k_n cells:

$$D_{n,r} = \{ \ (x,y) \in D \mid x \in I_{n,r} \ \},$$

and $\Sigma_{n,r} = D_{n,r} \cap \Sigma_n$. Let $U_{n,r}$ denote the supremum of the second coordinate of the points associated to the truncated process $N_n(.\cap D_{n,r})$:

$$U_{n,r} = \sup\{Y_i \mid (X_i, Y_i) \in \Sigma_{n,r}\} \text{ if } \Sigma_{n,r} \neq \emptyset, \ U_{n,r} = 0 \text{ otherwise.}$$

Geffroy's estimate [5] is defined by the piecewise constant function:

$$\hat{f}_n(x) = U_{n,r} \text{ for all } x \in I_{n,r}.$$

We denote by

$$\Delta_n = \int_0^1 \left| \hat{f}_n(x) - f(x) \right| dx,$$

the L_1- error between $\hat{f}_n$ and f. The assumptions required for studying the asymptotic behavior of Δ_n are summarized below. Assumptions on the unknown function are collected in **(A1)**, assumptions on the size of the partition are listed in **(A2)** and conditions on the intensity rate of the Poisson point process in **(A3)**.

(A1) f is α-Lipschitzian, $\alpha \in (0,1]$ and strictly positive on $[0,1]$.

(A2) $k_n \to \infty$ as $n \to \infty$, $k_n = o(n/\ln n)$ and $n = O\left(k_n^{1+\alpha}\right)$.

(A3) $c_n > 0$ and $c_n \to c_\infty$ as $n \to \infty$, with $0 < c_\infty < \infty$.

In most situations, the sequence (c_n) is constant. Nevertheless, **(A3)** allows more general intensity rates, which can prove to be useful for tackling the sample problem. In the sequel, we note $u_n \asymp v_n$ if (u_n) and (v_n) are two positive sequences such that

$$0 < \liminf u_n/v_n \leq \limsup u_n/v_n < \infty.$$

Our result is the following:

Theorem 1 *Under* **(A1)**–**(A3)**,

$$\frac{nc_n}{s_n k_n^{1/2}}(\Delta_n - E(\Delta_n)) \xrightarrow{d} N(0,1),$$

where $s_n \asymp 1$ *if* $n \asymp k_n^{1+\alpha}$ *and* $s_n = 1$ *if* $n = o\left(k_n^{1+\alpha}\right)$.

Corollary 1 *Under the conditions of the above theorem,*

$$\left| E(\Delta_n) - \frac{k_n}{nc_n} \right| = O\left(\frac{n}{k_n^{1+2\alpha}}\right). \tag{2.2}$$

Thus, if moreover $n = o\left(k_n^{3/4+\alpha}\right)$, *then*

$$\frac{nc_n}{k_n^{1/2}}(\Delta_n - \frac{k_n}{nc_n}) \xrightarrow{d} N(0,1), \tag{2.3}$$

when $n \to \infty$.

In view of (2.2) and under **(A2)**, the optimal speed for $\mathrm{E}(\Delta_n)$ is obtained for sequences (k_n) such that $k_n^{1+\alpha} \asymp n$. For such a choice, the result of Theorem 1 is difficult to use in practice since the centering sequence $\mathrm{E}(\Delta_n)$ is not explicit. On the contrary, the choice $n = o\left(k_n^{3/4+\alpha}\right)$ is suboptimal but it leads to an explicit limiting distribution (2.3). Moreover, this distribution is independent of the unknown function f. Then, it can be used to test if the boundary function f has a specified form f_0. To this end, consider the hypotheses $\mathcal{H}_0 : \{f = f_0\}$ and $\mathcal{H}_1 : \{f \neq f_0\}$. Under the hypotheses of Corollary 1, the test which rejects the null hypothesis when

$$\int_0^1 \left| \hat{f}_n(x) - f_0(x) \right| dx > \frac{k_n}{nc_n} + \frac{k_n^{1/2}}{nc_n}\Phi^{-1}(1-\gamma)$$

has asymptotic significance level $\gamma \in (0,1)$, where Φ denotes the standard normal distribution function.

Finally, we simulated 500 replications of a Poisson process with intensity rate $nc_n = 200$ on the set D defined as in (2.1) with

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

On each of the 500 simulations, Geffroy's estimate is computed (with $k_n = 35$), see Figure 1 for an example of result. The normalized L_1- error is then estimated on each simulation and its empirical distribution is compared with the $N(0,1)$ density on Figure 2. The two distributions looks similar.

3 Proofs.

For the sake of simplicity, let us introduce some supplementary notation:

$$m_{n,r} = \inf\{f(x) \mid x \in I_{n,r}\},$$

$$M_{n,r} = \sup\{f(x) \mid x \in I_{n,r}\},$$

$$\lambda_{n,r} = \int_{I_{n,r}} f(x)dx.$$

Let us denote by $G_{n,r}$ the cumulative distribution function of $U_{n,r}$, for each $r = 1, \ldots, k_n$. If $0 \leq y \leq m_{n,r}$, it can be written:

$$G_{n,r}(y) = P(U_{n,r} \leq y) = P(N_n(D_{n,r} \setminus (I_{n,r} \times [0,y])) = 0) = \exp\left(\frac{nc_n}{k_n}(y - k_n \lambda_{n,r})\right). \quad (3.1)$$

We also introduce $g_{n,r}$ the function defined by

$$g_{n,r}(y) = \frac{nc_n}{k_n} G_{n,r}(y) = \frac{nc_n}{k_n} \exp\left(\frac{nc_n}{k_n}(y - k_n \lambda_{n,r})\right). \quad (3.2)$$

The distribution of $U_{n,r}$ involves a Dirac distribution at the origin:

$$G_{n,r}(0) = P(U_{n,r} = 0) = \exp(-nc_n \lambda_{n,r}), \quad (3.3)$$

and a density distribution on the interval $(0, m_{n,r})$:

$$G'_{n,r}(y) = g_{n,r}(y).$$

The distribution of $U_{n,r}$ on the interval $(m_{n,r}, M_{n,r}]$ is not precisely known but it can be controlled with the Lipschitz condition **(A1)** which yields:

$$\max_{1\le r\le k_n} (M_{n,r} - m_{n,r}) = O\left(1/k_n^{\alpha}\right). \tag{3.4}$$

To prove Theorem 1, we need three lemmas. First, we quote in Lemma 3.1 some results on the largest values of the Poisson process.

Lemma 3.1 *Under* **(A1)**–**(A3)**, *we have*

(i) $\max_{1\le r\le k_n} (1 - G_{n,r}(m_{n,r})) = O\left(\frac{n}{k_n^{1+\alpha}}\right)$,

(ii) $\max_{1\le r\le k_n} \left|E(U_{n,r}) - k_n\lambda_{n,r} + \frac{k_n}{nc_n}\right| = O\left(\frac{n}{k_n^{1+2\alpha}}\right)$,

(iii) $\max_{1\le r\le k_n} \left|E\left([U_{n,r} - k_n\lambda_{n,r}]^2\right) - 2\frac{k_n^2}{n^2c_n^2}\right| = O\left(\frac{n}{k_n^{1+3\alpha}}\right)$,

(iv) $\max_{1\le r\le k_n} E([U_{n,r} - k_n\lambda_{n,r}]^3) = -6\frac{k_n^3}{n^3c_n^3} + O\left(\frac{n}{k_n^{1+4\alpha}}\right)$,

(v) $\max_{1\le r\le k_n} \left|Var(U_{n,r} \mid U_{n,r} \le m_{n,r}) - \frac{k_n^2}{n^2c_n^2}\right| = o\left(\frac{k_n^2}{n^2}\right)$.

Proof : (i) is a consequence of (3.1) and (3.4).

Proof of (ii). The mathematical expectation can be expanded in four terms:

$$\begin{aligned} E(U_{n,r} - k_n\lambda_{n,r}) &= -k_n\lambda_{n,r}G_{n,r}(0) + \int_0^{k_n\lambda_{n,r}} (y - k_n\lambda_{n,r})g_{n,r}(y)dy \\ &- \int_{m_{n,r}}^{k_n\lambda_{n,r}} (y - k_n\lambda_{n,r})g_{n,r}(y)dy + \int_{m_{n,r}}^{M_{n,r}} (y - k_n\lambda_{n,r})G_{n,r}(dy) \\ &\stackrel{def}{=} E_{n,r,1} + E_{n,r,2} + E_{n,r,3} + E_{n,r,4}. \end{aligned}$$

Partial integration yields:

$$E_{n,r,2} = -E_{n,r,1} + \frac{k_n}{nc_n}\exp(-nc_n\lambda_{n,r}) - \frac{k_n}{nc_n}. \tag{3.5}$$

We thus obtain

$$\left|\mathrm{E}(U_{n,r}) - k_n\lambda_{n,r} + \frac{k_n}{nc_n}\right| = \left|\frac{k_n}{nc_n}\exp(-nc_n\lambda_{n,r}) + E_{n,r,3} + E_{n,r,4}\right|.$$

The first term is controlled by the condition $k_n = o\left(n/\ln n\right)$:

$$\max_{1\le r\le k_n} \frac{k_n}{nc_n}\exp(-nc_n\lambda_{n,r}) = o\left(n^{-s}\right), \tag{3.6}$$

for all $s>0$. Besides, in view of (3.2), $g_{n,r}(y)\le nc_n/k_n$ for $y\le k_n\lambda_{n,r}$. It follows that

$$E_{n,r,3} \le \frac{nc_n}{k_n}\int_{m_{n,r}}^{k_n\lambda_{n,r}} (k_n\lambda_{n,r}-y)dy \le \frac{nc_n}{k_n}(M_{n,r}-m_{n,r})^2 = O\left(\frac{n}{k_n^{1+2\alpha}}\right), \tag{3.7}$$

uniformly in r with (3.4). The third term is bounded above by

$$|E_{n,r,4}| \le (M_{n,r}-m_{n,r})\int_{m_{n,r}}^{M_{n,r}} G_{n,r}(dy) = (M_{n,r}-m_{n,r})(1-G_{n,r}(m_{n,r})) = O\left(\frac{n}{k_n^{1+2\alpha}}\right), \tag{3.8}$$

uniformly in r with (i) and (3.4). Collecting (3.6)–(3.8) concludes the proof.

Proof of (iii). Similarly, we have

$$\begin{aligned} E([U_{n,r}-k_n\lambda_{n,r}]^2) &= k_n^2\lambda_{n,r}^2 G_{n,r}(0) + \int_0^{k_n\lambda_{n,r}} (y-k_n\lambda_{n,r})^2 g_{n,r}(y)dy \\ &- \int_{m_{n,r}}^{k_n\lambda_{n,r}} (y-k_n\lambda_{n,r})^2 g_{n,r}(y)dy + \int_{m_{n,r}}^{M_{n,r}} (y-k_n\lambda_{n,r})^2 G_{n,r}(dy) \\ &\stackrel{def}{=} E'_{n,r,1}+E'_{n,r,2}+E'_{n,r,3}+E'_{n,r,4}. \end{aligned}$$

A new partial integration yields:

$$E'_{n,r,2} = -E'_{n,r,1} - 2\frac{k_n}{nc_n}E_{n,r,2}. \tag{3.9}$$

We thus obtain

$$\mathrm{E}([U_{n,r})-k_n\lambda_{n,r}]^2) = -2\frac{k_n}{nc_n}E_{n,r,2}+E'_{n,r,3}+E'_{n,r,4}.$$

The first term is evaluated with (3.5) and (3.3):

$$\begin{aligned} -2\frac{k_n}{nc_n}E_{n,r,2} &= 2\frac{k_n^2}{n^2c_n^2} + 2\frac{k_n}{nc_n}\left(E_{n,r,1} - \frac{k_n}{nc_n}\exp(-nc_n\lambda_{n,r})\right) \\ &= 2\frac{k_n^2}{n^2c_n^2} - 2\frac{k_n^2}{nc_n}\exp(-nc_n\lambda_{n,r})\left(\lambda_{n,r}+\frac{1}{nc_n}\right) \\ &= 2\frac{k_n^2}{n^2c_n^2} + o\left(n^{-s}\right), \end{aligned} \tag{3.10}$$

for all $s>0$. The second term is bounded as in (3.7):

$$E'_{n,r,3} \le \frac{nc_n}{k_n}(M_{n,r}-m_{n,r})^3 = O\left(\frac{n}{k_n^{1+3\alpha}}\right), \tag{3.11}$$

uniformly in r with (3.4). The third term is bounded as in (3.8):

$$E'_{n,r,4} \leq (M_{n,r} - m_{n,r})^2 (1 - G_{n,r}(m_{n,r})) = O\left(\frac{n}{k_n^{1+3\alpha}}\right), \tag{3.12}$$

uniformly in r with (i) and (3.4). Collecting (3.10)–(3.12) concludes the proof.

Proof of (iv). The proof follows the same lines as the previous ones. We have

$$\begin{aligned}
E([U_{n,r} - k_n\lambda_{n,r}]^3) &= k_n^3 \lambda_{n,r}^3 G_{n,r}(0) + \int_0^{k_n\lambda_{n,r}} (y - k_n\lambda_{n,r})^3 g_{n,r}(y) dy \\
&- \int_{m_{n,r}}^{k_n\lambda_{n,r}} (y - k_n\lambda_{n,r})^3 g_{n,r}(y) dy + \int_{m_{n,r}}^{M_{n,r}} (y - k_n\lambda_{n,r})^3 G_{n,r}(dy) \\
&\stackrel{def}{=} E''_{n,r,1} + E''_{n,r,2} + E''_{n,r,3} + E''_{n,r,4}.
\end{aligned}$$

A new partial integration yields:

$$E''_{n,r,2} = -E''_{n,r,1} - 3\frac{k_n}{nc_n} E'_{n,r,2}.$$

We thus obtain

$$\mathrm{E}([U_{n,r}) - k_n\lambda_{n,r}]^3) = -3\frac{k_n}{nc_n} E'_{n,r,2} + E''_{n,r,3} + E''_{n,r,4}.$$

The first term is evaluated with (3.9), (3.10) and (3.3) leading to:

$$-3\frac{k_n}{nc_n} E'_{n,r,2} = 6\frac{k_n^3}{n^3 c_n^3} + o\left(n^{-s}\right), \tag{3.13}$$

for all $s > 0$. The second term is bounded as in (3.11):

$$E''_{n,r,3} \leq \frac{nc_n}{k_n} (M_{n,r} - m_{n,r})^4 = O\left(\frac{n}{k_n^{1+4\alpha}}\right), \tag{3.14}$$

uniformly in r with (3.4). The third term is bounded as in (3.12):

$$E''_{n,r,4} \leq (M_{n,r} - m_{n,r})^3 (1 - G_{n,r}(m_{n,r})) = O\left(\frac{n}{k_n^{1+4\alpha}}\right), \tag{3.15}$$

uniformly in r with (i) and (3.4). Collecting (3.13)–(3.15) concludes the proof.

Proof of (v). The variance can be expanded as

$$\mathrm{Var}(U_{n,r} \mid U_{n,r} \leq m_{n,r}) = \frac{\mathrm{E}(U_{n,r}^2 \mathbb{I}\{U_{n,r} \leq m_{n,r}\})}{G_{n,r}(m_{n,r})} - \left(\frac{\mathrm{E}(U_{n,r} \mathbb{I}\{U_{n,r} \leq m_{n,r}\})}{G_{n,r}(m_{n,r})}\right)^2, \tag{3.16}$$

where $\mathbb{I}\{.\}$ is the indicator function. Now,

$$\mathrm{E}(U_{n,r} \mathbb{I}\{U_{n,r} \leq m_{n,r}\}) = \int_0^{m_{n,r}} y g_{n,r}(y) dy = \left(m_{n,r} - \frac{k_n}{nc_n} + o\left(\frac{k_n^2}{n^2}\right)\right) G_{n,r}(m_{n,r}),$$

where the sequence $o(k_n^2/n^2)$ is independent of r and, similarly,

$$\mathrm{E}(U_{n,r}^2 \mathbb{I}\{U_{n,r} \le m_{n,r}\}) = \left(m_{n,r}^2 - 2\frac{m_{n,r}k_n}{nc_n} + 2\frac{k_n^2}{n^2c_n^2} + o\left(\frac{k_n^2}{n^2}\right)\right) G_{n,r}(m_{n,r}).$$

Replacing in (3.16) gives the result. ∎

Introducing for $r = 1, \dots, k_n$,

$$A_{n,r} = \int_{I_{n,r}} \left|f(x) - \hat{f}_n(x)\right| dx = \int_{I_{n,r}} |f(x) - U_{n,r}|\, dx,$$

the L_1- error can be rewritten

$$\Delta_n = \sum_{r=1}^{k_n} A_{n,r}.$$

The $(A_{n,r})$ are non negative and independent since they are built on extreme values of non-overlapping Poisson processes. The moments of the random variables $(A_{n,r})$ are then controlled thanks to the Lipschitz condition of **(A1)** in Lemma 3.2. These results are complemented by Lemma 3.3 which provides a lower bound for the variance.

Lemma 3.2 *Under* **(A1)**–**(A3)**,

(i) $\max_{1\le r\le k_n} \left|E(A_{n,r}) - \frac{1}{nc_n}\right| = O\left(\frac{n}{k_n^{2+2\alpha}}\right)$,

(ii) $\max_{1\le r\le k_n} \left|E(A_{n,r}^2) - \frac{2}{n^2c_n^2}\right| = O\left(\frac{n}{k_n^{3+3\alpha}}\right)$,

(iii) $\max_{1\le r\le k_n} E(A_{n,r}^3) \le \frac{6}{n^3c_n^3} + O\left(\frac{n}{k_n^{4+4\alpha}}\right)$.

Proof : Recall that, for any function g, we can always write $|g| = g + 2g^-$, where $g^- = -\min(g, 0)$. Thus, $A_{n,r}$ can be expanded as $A_{n,r} = B_{n,r} + C_{n,r}$, for $r = 1, \dots, k_n$ with

$$B_{n,r} = \int_{I_{n,r}} (f(x) - \hat{f}_n(x))dx = \lambda_{n,r} - \frac{U_{n,r}}{k_n}, \tag{3.17}$$

$$C_{n,r} = 2\int_{I_{n,r}} (f(x) - \hat{f}_n(x))^- dx = 2\int_{I_{n,r}} (U_{n,r} - f(x))\mathbb{I}\{U_{n,r} > f(x)\}\, dx. \tag{3.18}$$

Proof of (i). We have the evident inequality

$$\max_{1\le r\le k_n} \left|\mathrm{E}(A_{n,r}) - \frac{1}{nc_n}\right| \le \max_{1\le r\le k_n} \left|\mathrm{E}(B_{n,r}) - \frac{1}{nc_n}\right| + \max_{1\le r\le k_n} \mathrm{E}(C_{n,r}),$$

where the first term is controlled with Lemma 3.1(ii):

$$\max_{1\le r\le k_n}\left|\mathrm{E}(B_{n,r})-\frac{1}{nc_n}\right|\le\max_{1\le r\le k_n}\left|\frac{\mathrm{E}(U_{n,r})}{k_n}-\lambda_{n,r}+\frac{1}{nc_n}\right|=O\left(\frac{n}{k_n^{2+2\alpha}}\right).$$

Remark that in (3.18), $U_{n,r}\le M_{n,r}$ and $f(x)\ge m_{n,r}$. Thus,

$$0\le C_{n,r}\le 2(M_{n,r}-m_{n,r})\int_{I_{n,r}}\mathbb{I}\left\{U_{n,r}>m_{n,r}\right\}dx\le 2\frac{M_{n,r}-m_{n,r}}{k_n}\mathbb{I}\left\{U_{n,r}>m_{n,r}\right\},\tag{3.19}$$

and therefore

$$\mathrm{E}(C_{n,r})\le 2\frac{M_{n,r}-m_{n,r}}{k_n}(1-G_{n,r}(m_{n,r}))=O\left(\frac{n}{k_n^{2\alpha+2}}\right),$$

uniformly in r with Lemma 3.1(i), and the result follows.

Proof of (ii). We have the inequality

$$\left|\mathrm{E}(A_{n,r}^2)-\frac{2}{n^2c_n^2}\right|\le\left|\mathrm{E}(B_{n,r}^2)-\frac{2}{n^2c_n^2}\right|+2\,\mathrm{E}(|B_{n,r}|\,C_{n,r})+\mathrm{E}(C_{n,r}^2).$$

Let us consider the three terms separately. In view of Lemma 3.1(iii), we have

$$\left|\mathrm{E}(B_{n,r}^2)-\frac{2}{n^2c_n^2}\right|=\left|\mathrm{E}\left(\lambda_{n,r}-\frac{U_{n,r}}{k_n}\right)^2-\frac{2}{n^2c_n^2}\right|=O\left(\frac{n}{k_n^{3+3\alpha}}\right),\tag{3.20}$$

uniformly in r. Now, (3.19) implies for all $r=1,\dots,k_n$,

$$0\le C_{n,r}^2\le 4\frac{(M_{n,r}-m_{n,r})^2}{k_n^2}\mathbb{I}\left\{U_{n,r}>m_{n,r}\right\},$$

and thus

$$\mathrm{E}(C_{n,r}^2)\le 4\frac{(M_{n,r}-m_{n,r})^2}{k_n^2}(1-G_{n,r}(m_{n,r}))=O\left(\frac{n}{k_n^{3+3\alpha}}\right)\tag{3.21}$$

uniformly in r with Lemma 3.1 (i). Finally, in view of (3.19),

$$\begin{aligned}\mathrm{E}(|B_{n,r}|\,C_{n,r})&\le 2\frac{M_{n,r}-m_{n,r}}{k_n}\,\mathrm{E}\left(\left|\lambda_{n,r}-\frac{U_{n,r}}{k_n}\right|\mathbb{I}\left\{U_{n,r}>m_{n,r}\right\}\right)\\&\le 2\frac{(M_{n,r}-m_{n,r})^2}{k_n^2}(1-G_{n,r}(m_{n,r}))\\&= O\left(\frac{n}{k_n^{3+3\alpha}}\right),\end{aligned}\tag{3.22}$$

uniformly in r with Lemma 3.1 (i). Collecting (3.20)–(3.22) gives the result.

Proof of (iii). Using the expansion $A_{n,r}=B_{n,r}+C_{n,r}$, we have

$$\mathrm{E}(A_{n,r}^3)\le\left|\mathrm{E}(B_{n,r}^3)\right|+3\,\mathrm{E}(B_{n,r}^2C_{n,r})+3\,\mathrm{E}(|B_{n,r}|\,C_{n,r}^2)+\mathrm{E}(C_{n,r}^3).$$

First, Lemma 3.1(iv) yields

$$\left|E(B_{n,r}^3)\right| = \frac{1}{k_n^3}\left|E([U_{n,r} - k_n\lambda_{n,r}]^3)\right| \leq \frac{6}{n^3 c_n^3} + O\left(\frac{n}{k_n^{4+4\alpha}}\right), \tag{3.23}$$

uniformly in r. Second, in view of (3.19),

$$\begin{aligned}
\mathrm{E}(B_{n,r}^2 C_{n,r}) &\leq 2\frac{M_{n,r} - m_{n,r}}{k_n}\,\mathrm{E}\left(\left[\lambda_{n,r} - \frac{U_{n,r}}{k_n}\right]^2 \mathbb{I}\left\{U_{n,r} > m_{n,r}\right\}\right) \\
&\leq 2\frac{(M_{n,r} - m_{n,r})^3}{k_n^3}(1 - G_{n,r}(m_{n,r})) \\
&= O\left(\frac{n}{k_n^{4+4\alpha}}\right), \qquad (3.24)
\end{aligned}$$

uniformly in r with Lemma 3.1 (i). Similary,

$$\begin{aligned}
\mathrm{E}(|B_{n,r}|\, C_{n,r}^2) &\leq 4\frac{(M_{n,r} - m_{n,r})^2}{k_n^2}\,\mathrm{E}\left(\left|\lambda_{n,r} - \frac{U_{n,r}}{k_n}\right| \mathbb{I}\left\{U_{n,r} > m_{n,r}\right\}\right) \\
&\leq 2\frac{(M_{n,r} - m_{n,r})^3}{k_n^3}(1 - G_{n,r}(m_{n,r})) \\
&= O\left(\frac{n}{k_n^{4+4\alpha}}\right), \qquad (3.25)
\end{aligned}$$

uniformly in r. Finally, (3.19) yields

$$\mathrm{E}(C_{n,r}^3) \leq 8\frac{(M_{n,r} - m_{n,r})^3}{k_n^3}(1 - G_{n,r}(m_{n,r})) = O\left(\frac{n}{k_n^{4+4\alpha}}\right), \tag{3.26}$$

uniformly in r. Collecting (3.23) – (3.26) concludes the proof. ■

Lemma 3.3 *Under* **(A1)**–**(A3)**, *there exists* $K > 0$ *such that*

$$\min_{1 \leq r \leq k_n} \mathit{Var}(A_{n,r}) \geq \frac{K}{n^2 c_n^2}(1 + o(1)).$$

Proof : As a consequence of the variance decomposition formula:

$$\begin{aligned}
\mathrm{Var}\,(A_{n,r}) &\geq P(U_{n,r} \leq m_{n,r})\,\mathrm{Var}\,(A_{n,r} \mid U_{n,r} \leq m_{n,r}) \\
&= G_{n,r}(m_{n,r})\frac{1}{k_n^2}\,\mathrm{Var}\,(U_{n,r} \mid U_{n,r} \leq m_{n,r}),
\end{aligned}$$

since $A_{n,r} = B_{n,r} = \lambda_{n,r} - U_{n,r}/k_n$ when $U_{n,r} \leq m_{n,r}$ (see (3.17)). Remarking that $G_{n,r}(m_{n,r})$ is uniformly bounded from below and taking into account of Lemma 3.1(v) conclude the proof. ■

Proof of Theorem 1. The quantity of interest can be expanded as the sum of a triangular array of independent and centered random variables $(Y_{n,r})$:

$$S_n = \frac{nc_n}{k_n^{1/2}}(\Delta_n - \mathrm{E}(\Delta_n)) = \sum_{r=1}^{k_n} \frac{nc_n}{k_n^{1/2}}(A_{n,r} - \mathrm{E}(A_{n,r})) = \sum_{r=1}^{k_n} Y_{n,r},$$

where $Y_{n,r}$ is defined by, for $r = 1, \ldots, k_n$:

$$Y_{n,r} = \frac{nc_n}{k_n^{1/2}}(A_{n,r} - \mathrm{E}(A_{n,r})).$$

Let $s_n^2 = \operatorname{Var} S_n$. A sufficient condition for $S_n/s_n \xrightarrow{d} N(0,1)$ is provided by the Lyapounov condition

$$\lim_{n\to\infty} \frac{1}{s_n^3} \sum_{r=1}^{k_n} \mathrm{E}\left(|Y_{n,r}|^3\right) = 0. \tag{3.27}$$

Lemma 3.2(i)–(ii) yields

$$\begin{aligned} s_n^2 &= \frac{n^2 c_n^2}{k_n} \sum_{r=1}^{k_n} [\mathrm{E}(A_{n,r}^2) - \mathrm{E}^2(A_{n,r})] \\ &= n^2 c_n^2 \left[\frac{2}{n^2 c_n^2} + O\left(\frac{n}{k_n^{3+3\alpha}}\right) - \left(\frac{1}{nc_n} + O\left(\frac{n}{k_n^{2+2\alpha}}\right)\right)^2 \right] \\ &= 1 + O\left(\frac{n^2}{k_n^{2+2\alpha}}\right). \end{aligned} \tag{3.28}$$

Two situations appear. If $n = o\left(k_n^{1+\alpha}\right)$ then $s_n = 1 + o(1)$. If $n \asymp k_n^{1+\alpha}$, Lemma 3.3 entails $s_n^2 \geq K(1 + o(1))$ which, together with (3.28), yield $s_n \asymp 1$. In both cases, condition (3.27) reduces to

$$\lim_{n\to\infty} \sum_{r=1}^{k_n} \mathrm{E}\left(|Y_{n,r}|^3\right) = 0.$$

We have

$$\begin{aligned} \sum_{r=1}^{k_n} \mathrm{E}\left(|Y_{n,r}|^3\right) &= \frac{n^3 c_n^3}{k_n^{3/2}} \sum_{r=1}^{k_n} \mathrm{E}\left(|A_{n,r} - \mathrm{E}(A_{n,r})|^3\right) \\ &\leq \frac{8n^3 c_n^3}{k_n^{3/2}} \sum_{r=1}^{k_n} \mathrm{E}\left(|A_{n,r}|^3\right) \\ &\leq \frac{48}{k_n^{1/2}} \left[1 + O\left(\frac{n^4}{k_n^{4+4\alpha}}\right)\right] \end{aligned}$$

with Lemma 3.2(iii) and thus (3.27) is verified. As a conclusion, $S_n/s_n \xrightarrow{d} N(0,1)$ as $n \to \infty$.

■

Proof of Corollary 1. As a consequence of Lemma 3.2(i):

$$\left|\mathrm{E}(\Delta_n) - \frac{k_n}{nc_n}\right| \le k_n \max_{1\le r\le k_n} \left|\mathrm{E}(A_{n,r}) - \frac{1}{nc_n}\right| = O\left(\frac{n}{k_n^{1+2\alpha}}\right),$$

and (2.2) is proved. Then,

$$\frac{nc_n}{k_n^{1/2}}\left(\mathrm{E}(\Delta_n) - \frac{k_n}{nc_n}\right) = O\left(\frac{n^2}{k_n^{3/2+2\alpha}}\right)$$

which converges to 0 under $n = o\left(k_n^{3/4+\alpha}\right)$. ■

References

[1] Beirlant, J., Györfi, L. and Lugosi, G. (1994) On the asymptotic normality of the L_1- and L_2-errors in histogram density estimation. *Canad. J. Statist.*, **22**(3), 309–318.

[2] Beirlant, J. and Györfi, L. (1998) On the L_1-error in histogram density estimation: the multidimensional case. *J. Nonparametr. Statist.*, **9**(2), 197–216.

[3] Berlinet, A., Devroye, L. and Györfi, L. (1995) Asymptotic normality of L_1-error in density estimation. *Statistics*, **26**(4), 329–343.

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

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

[6] Girard, S. and Jacob, P. (2001) Extreme values and kernel estimates of point processes boundaries. *Technical report ENSAM-INRA-UM2*, **01-02**.

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

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

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

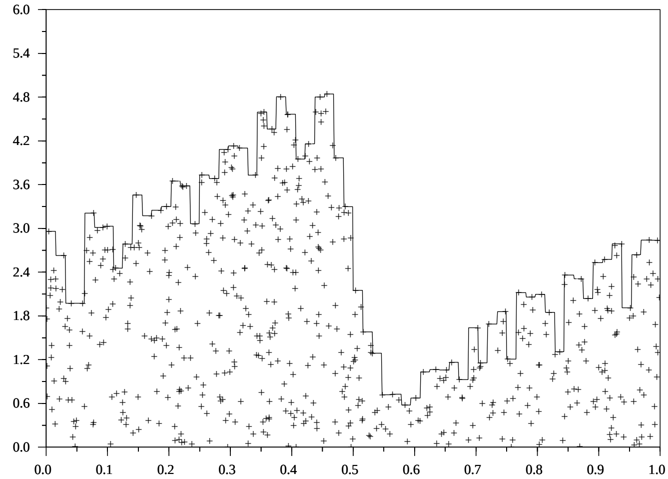

Figure 1: Example of result obtained on a simulation with Geffroy's estimate.

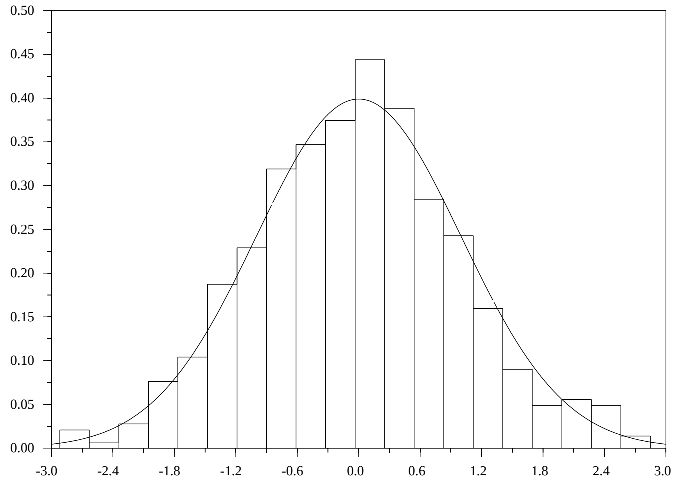

Figure 2: Comparison between the empirical distribution of the normalized L_1- error and the standard normal density.