Statistics **26** (1995) 329–343
Reprints available directly from the publisher
Photocopying permitted by license only

ASYMPTOTIC NORMALITY OF L_1-ERROR IN DENSITY ESTIMATION

ALAIN BERLINET, LUC DEVROYE and LÁSZLÓ GYÖRFI

Université de Montpellier, McGill University, Montreal and Technical University of Budapest

(Received 15 November, 1993; in final form 1 November, 1994)

Summary. Let f_n be a histogram estimate constructed from a sample of i.i.d. real-valued random variables with common continuously differentiable density f. In this paper we prove a central limit theorem for the L_1 error $\| f_n - f \|$. We determine a positive constant $0 < \sigma^2 \leqslant 1 - 2/\pi$ in order that, under the usual conditions of consistency, the law of

$$\sqrt{n}(\| f_n - f \| - \mathrm{E} \| f_n - f \|)/\sigma$$

be asymptotically Gaussian $\mathcal{N}(0, 1)$.

AMS 1991 subject classifications: Primary 62H12, Secondary 62G05.

Key words: Nonparametric density estimation, central limit theorem, histogram estimate.

I. INTRODUCTION

Although density estimates have been extensively studied during the last thirty years, many results were only obtained under superfluous assumptions. For example, this is the case in asymptotic normality studies for the global measures of deviation. Either the error between the estimate f_n and its expectation $\mathrm{E} f_n$ is considered instead of the real error $f_n - f$ or strong assumptions are made on the density f. In this paper we consider the asymptotic behavior of the L_1 error $\| f_n - f \|$ where f_n is an histogram constructed from a sample of i.i.d. real-valued random variables with common continuously differentiable density f. Results about histograms still present practical interest, as histograms are more adapted to on-line high data speed signal processing. Also, averaging histograms circumvents the problem of their variability (Scott 1985, Härdle 1991).

Let $\mathbb{N}^*$ denote the set of positive integers and let $(X_i)_{i \in \mathbb{N}^*}$ be a sequence of i.i.d. real valued random variables with common unknown density f with respect to the Lebesgue measure λ on $\mathbb{R}$. We denote by μ the measure with density f. For each $n \in \mathbb{N}^*$, let h_n be a positive number and let $\mathscr{P}_n$ be a partition of $\mathbb{R}$ into intervals A_{nj}, $j \in \mathbb{N}^*$, with equal measure h_n:

$$\forall n \in \mathbb{N}^*, \quad \forall j \in \mathbb{N}^*, \quad \lambda(A_{nj}) = h_n.$$

For $n \in \mathbb{N}^*$, let f_n be the standard histogram estimate of f constructed from $X_1, \ldots, X_n$ and the partition $\mathscr{P}_n$, that is

$$f_n(x) = \frac{\mu_n(A_{nj})}{h_n} \quad \text{if} \quad x \in A_{nj}, \tag{1}$$

where the empirical measure μ_n is defined, for any set A in $\mathscr{B}_{\mathbb{R}}$, the Borel σ-algebra of $\mathbb{R}$, by

$$\mu_n(A) = \frac{\#\{i \mid X_i \in A, 1 \leqslant i \leqslant n\}}{n}.$$

We show that the L_1 error $\|f_n - f\| = \int_{\mathbb{R}} |f_n - f|$, suitably standardized, is asymptotically Gaussian $\mathcal{N}(0, 1)$ under the usual conditions of consistency on (h_n). Our technique relies on a Poissonization argument originating from the fact that a multinomial distribution can be written as a conditional distribution of a set of independent Poisson random variables given their sum. From this, Bartlett's idea of partial inversion for obtaining characteristic functions of conditional distributions can be applied. Using this idea Beirlant, Györfi and Lugosi (1994) proved the following results: if $nh_n \to \infty$ and $h_n \to 0$ as $n \to \infty$, then

$$\sqrt{n}(\|f_n - \mathrm{E}f_n\| - \mathrm{E}\|f_n - \mathrm{E}f_n\|)/\sigma_1 \xrightarrow{\mathscr{L}} \mathcal{N}(0, 1), \tag{2}$$

where $\sigma_1^2 = 1 - 2/\pi$. Beirlant and Mason (1992) extended this to the L_p norm $D_n(p) = \|\omega_n^{1/p}(\tau_n - \mathrm{E}\tau_n)\|_p^p$, where ω_n is a weight function and τ_n is either a histogram or a kernel estimate or a regressogram. Note that these results are limit laws on $(f_n - \mathrm{E} f_n)$ and not on $(f_n - f)$. Obviously this limit law can be extended to the L_1 error if the variation term $\|f_n - \mathrm{E}f_n\|$ dominates the bias $\|\mathrm{E}f_n - f\|$. If one wants to have small expected L_1 error then the variation and the bias terms should be of the same order, so in this case the asymptotic normality does not follow. Csörgő and Horváth (1988) and Horváth (1991) proved for the kernel estimate that if f belongs to a subset of twice differentiable densities and the variation term dominates the bias, then the asymptotics of $\|f_n - f\|$ is independent of f. Under the additional conditions, they obtained a limit law when the variation term and the bias term are of the same order. Devroye (1988, 1991) proved that if f_n is the histogram or the kernel estimate, then

$$\mathrm{Var}\{\sqrt{n}\|f_n - f\|\} < 1 \tag{3}$$

and

$$\mathrm{P}\{\sqrt{n}|\|f_n - f\| - \mathrm{E}\|f_n - f\|| > \varepsilon\} \leqslant 2e^{-\varepsilon^2/2} \tag{4}$$

for all f, n, h_n, ε and nonnegative kernel. (2), (3) and (4) suggested the conjecture that we have asymptotic normality with asymptotic variance less than 1 and maybe independent of the density. In fact the asymptotic variance depends on the smoothness of f, but it is smaller than the asymptotic variance of the variation term. According to this $\|f_n - f\| - \mathrm{E}\|f_n - f\|$ is of order $n^{-1/2}$. This should be compared to the rate of convergence of $\mathrm{E}\|f_n - f\|$, which is at least of order $n^{-1/3}$ for differentiable f, and it can be achieved for $h_n = cn^{-1/3}$. The best choice of c is

$$c_{\mathrm{opt}} = \left(\frac{8}{\pi}\left(\frac{\int\sqrt{f}}{\int|f'|}\right)^2\right)^{1/3}$$

(Devroye and Györfi (1985, section 5.6)). The limit law in this paper shows that essentially all information about $\|f_n - f\|$ is contained in $\mathrm{E}\|f_n - f\|$.

II. MAIN RESULTS

Introduce

$$\psi_\alpha(u) = u + \frac{\alpha}{2}\left[\left(1 - \frac{u}{\alpha}\right)^+\right]^2$$

and

$$V(\alpha) = \mathrm{Var}\{\psi_\alpha(|N|)\},$$

where $\alpha > 0$ and N is a standard normal $\mathcal{N}(0, 1)$ random variable.

Theorem 1. *If f is continuously differentiable on $\mathbb{R}$ and if $h_n = cn^{-1/3}$, then*

$$\sqrt{n}(\|f_n - f\| - \mathrm{E}\|f_n - f\|)/\sigma \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1),$$

where

$$\sigma^2 = \int V\left(\frac{c^{3/2}|f'|}{2\sqrt{f}}\right) f \, \mathrm{d}\lambda.$$

Note that we do not have any tail condition: the support of f can be unbounded. Lemma 1 below, giving the behavior of the function V, implies that $\sigma^2 \leqslant 1 - 2/\pi$, and $\sigma^2 \to 1 - 2/\pi$ as $c \downarrow 0$. When c is large, the bias dominates the variation, and σ^2 varies like $1/c^3$ (Lemma 1(c)). Thus if we cannot set $c = c_{\text{opt}}$, then $c > c_{\text{opt}}$ should be preferred over $c < c_{\text{opt}}$.

III. LEMMAS AND PROOFS

Lemma 1.

- (a) *V is monotone decreasing and has infinitely many derivatives.*
- (b) *For α_1 small enough, V is concave on $]0, \alpha_1]$. As $\alpha \downarrow 0$,*

$$V(\alpha) = 1 - \frac{2}{\pi} - \frac{2}{3\pi}\alpha^2 + o(\alpha^2).$$

- (c) *For α_2 large enough, V is convex on $[\alpha_2, +\infty[$. For $1 \leqslant \alpha$,*

$$V(\alpha) \leqslant \frac{1}{2\alpha^2}$$

and

$$\lim_{\alpha \uparrow \infty} \alpha^2 V(\alpha) = \frac{1}{2}.$$

Proof. (a) For $b > 0$ we define the functions $z(x) = \psi_{1/b}(x)$ and $z'(x) = (\partial/\partial b)(z(x))$ and prove that for any positive random variable X with finite variance, $\mathrm{Var}\{z(X)\}$ is a non decreasing function of b. This clearly implies that V is monotone decreasing. We have, if all derivatives are with respect to b,

$$\begin{aligned}(\operatorname{Var}\{z(X)\})' &= (\mathrm{E}\{z^2(X)\})' - (\mathrm{E}^2\{z(X)\})' \\ &= \mathrm{E}\{(z^2)'(X)\} - 2(\mathrm{E}\{z(X)\})'\mathrm{E}\{z(X)\} \\ &= \mathrm{E}\{2z'(X)z(X)\} - 2\mathrm{E}\{z'(X)\}\mathrm{E}\{z(X)\} \\ &\geqslant 0.\end{aligned}$$

Let us explain every step in this chain. The first equality is obvious. In the second one, we only use the fact that with $g(x,b) = z^2(x)$ one has

$$\frac{\partial \mathrm{E}\{g(X,b)\}}{\partial b} = \mathrm{E}\left\{\frac{\partial g}{\partial b}(X)\right\}.$$

The interchange of derivative and expectation is only allowed under certain circumstances: fix $b > 0$ and consider

$$\mathrm{E}\left\{\lim_{u\downarrow 0}\frac{g(X,b+u) - g(X,b)}{u}\right\}.$$

At every $x > 0$, the limit of $(g(x,b+u) - g(x,b))/u$ exists and equals $g'(x) = 2z(x)z'(x)$. Also, as $z'(x) = -(1-b^2x^2)^+/(2b^2)$, it is easy to see that the family $\{(g(x,b+u) - g(x,b))/u\}$ is uniformly integrable in u over a small interval near zero, provided that $\mathrm{E}X^2 < \infty$. Thus, by uniform integrability – the dominated convergence theorem, really, see Chung, 1974, p. 97 –, we note that

$$\begin{aligned}\mathrm{E}\left\{\frac{\partial g(x,b)}{\partial b}\right\} &= \mathrm{E}\left\{\lim_{u\downarrow 0}\frac{g(X,b+u) - g(X,b)}{u}\right\} \\ &= \lim_{u\downarrow 0}\mathrm{E}\left\{\frac{g(X,b+u) - g(X,b)}{u}\right\} \\ &= \frac{\partial(\mathrm{E}g(X,b))}{\partial b}.\end{aligned}$$

The third equation in the original chain follows by taking $g \equiv z$. For fixed $b > 0$, both $z(x)$ and $z'(x)$ are increasing in x. Thus, $z'(X)$ and $z(X)$ are positively associated (Tong, 1980). Hence,

$$\mathrm{E}\{z(X)z'(X)\} \geqslant \mathrm{E}\{z(X)\}\mathrm{E}\{z'(X)\},$$

thus explaining the last step. Now, if ϕ denotes the standard normal density then (b) and the second part of statement (a) follow from the expression:

$$\begin{aligned}V(\alpha) = 1 - \frac{2}{\pi} + \frac{1}{2\alpha^2}\int_0^\alpha (\alpha^2 - y^2)^2\phi(y)\mathrm{d}y \\ - 2\sqrt{\frac{2}{\pi}\frac{1}{\alpha}}\int_0^\alpha (\alpha - y)^2\phi(y)\mathrm{d}y - \left(\frac{1}{\alpha}\int_0^\alpha (\alpha - y)^2\phi(y)\mathrm{d}y\right)^2.\end{aligned}$$

(c) is a consequence of

$$\alpha^2 V(\alpha) = \frac{1}{2} - \frac{1}{2}\int_\alpha^\infty (\alpha^2 - y^2)^2 \phi(y)\mathrm{d}y$$
$$+ (\alpha^2+1)\int_\alpha^\infty (\alpha - y)^2\phi(y)\mathrm{d}y - \left(\int_\alpha^\infty (\alpha-y)^2\phi(y)\mathrm{d}y^2\right).$$

To prove Theorem 1, we use Poissonization (for L_p norms, this was also used by Horvâth (1991)). For any positive integer i, let N_i be a Poisson (i) random variable independent of the sequence $(X_j)_{j\in\mathbb{N}^*}$. Define, for $n\in\mathbb{N}^*$

$$\mu_{N_n}(A) = \frac{\#\{i \mid X_i \in A, 1 \leqslant i \leqslant N_n\}}{n}$$

and

$$f_{N_n}(x) = \frac{\mu_{N_n}(A_{nj})}{h_n} \quad \text{if} \quad x \in A_{nj}.$$

Assume that the $\{A_{nj}\}$ are ordered according to non-decreasing distances of their centers from the origin. For $\gamma\in(0,1)$ choose the integer m_n such that

$$1-\gamma_n := \sum_{j=1}^{m_n} \mu(A_{nj}) \leqslant 1-\gamma < \sum_{j=1}^{m_n+1} \mu(A_{nj}).$$

Roughly speaking, $S_{\gamma_n} = \bigcup_{j=1}^{m_n} A_{nj}$ is approximately an interval centered at the origin with $1-\gamma \sim \mu(S_{\gamma_n})$. Obviously $nh_n \to \infty$ implies that $m_n/n \to 0$. Moreover,

$$0 \leqslant \gamma_n - \gamma \leqslant \mu(A_{n,m_n+1}) \leqslant \max_j \mu(A_{nj}) \to 0$$

as $n\to\infty$. Also define S_γ as the interval centered at the origin with the property $1-\gamma = \mu(S_\gamma)$.

Beirlant, Györfi and Lugosi (1994) allow one to extend central limit theorems for Poissonized functions to the original ones (see Lemma 2). To prove our theorem with a centering constant equal to $\mathrm{E}\| f_{N_n} - f\|$, we have to choose suitable functions g_{nj} and to verify the conditions of this Lemma. Then we will get the final result by making use of Lemma 10 which implies that

$$\mathrm{E}\| f_{N_n} - f\| - \mathrm{E}\| f_n - f\| = o\left(\frac{1}{\sqrt{n}}\right).$$

Lemma 2. *Let g_{nj} be real measurable functions with*

$$\mathrm{E}\{g_{nj}(\mu_{N_n}(A_{nj}))\} = 0 \quad (n, j \geqslant 1).$$

Assume that for all t, v and γ

$$\Phi_{n,\gamma}(t,v) = \mathrm{E}\left\{\exp\left(it\sum_{j=1}^{m_n} g_{nj}(\mu_{N_n}(A_{nj})) + iv\frac{N_n - n}{\sqrt{n}}\right)\right\} \to e^{-t^2\rho_\gamma^2/2}e^{-v^2/2},$$

with $\rho_\gamma^2 = \int_{S_\gamma} h(x)dx$, *where* $h(x)$ *is some measurable function such that* $\rho_0^2 = \int_{\mathbb{R}} h(x)\mathrm{d}x < \infty$. *Then*

$$\sum_{j=1}^{\infty} g_{nj}(\mu_n(A_{nj}))/\rho_0 \xrightarrow{\mathscr{L}} \mathcal{N}(0,1).$$

Lemma 3. *Let* f *satisfy the condition of Theorem 1. If* a *is the center of* $A \in \mathscr{P}_n$, *put*

$$g_n(x) = \mathrm{E}\, f_n(a) + f'(a)(x-a) \quad (x \in A)$$

and

$$I_n = \sqrt{n}\int_{S_{\gamma n}} (|f_{N_n} - f| - \mathrm{E}|f_{N_n} - f| - (|f_{N_n} - g_n| - \mathrm{E}|f_{N_n} - g_n|)).$$

Then $\mathrm{E}\{I_n^2\} \to 0$.

Proof.

$$\begin{aligned}
\mathrm{E}\{I_n^2\} &= n\mathrm{E}\left\{\sum_{j=1}^{m_n}\int_{A_{nj}} (|f_{N_n} - f| - \mathrm{E}|f_{N_n} - f| - (|f_{N_n} - g_n| - \mathrm{E}|f_{N_n} - g_n|))\right\}^2 \\
&= n\sum_{j=1}^{m_n} \mathrm{E}\left\{\int_{A_{nj}} (|f_{N_n} - f| - \mathrm{E}|f_{N_n} - f| - (|f_{N_n} - g_n| - \mathrm{E}|f_{N_n} - g_n|))\right\}^2 \\
&= n\sum_{j=1}^{m_n} \mathrm{Var}\left\{\int_{A_{nj}} (|f_{N_n} - f| - |f_{N_n} - g_n|)\right\} \\
&\leqslant n\sum_{j=1}^{m_n} \mathrm{E}\left\{\int_{A_{nj}} (|f_{N_n} - f| - |f_{N_n} - g_n|)\right\}^2 \\
&\leqslant n\sum_{j=1}^{m_n} \left(\int_{A_{nj}} |f - g_n|\right)^2.
\end{aligned}$$

Consider the Taylor expansion of f, $f(x) = f(a) + f'(a)(x-a) + o(h_n)$. Then, as $\int_A f = \int_A \mathrm{E} f_n$ we see that $\mathrm{E} f_n(a) = f(a) + o(h_n)$. Thus,

$$\int_A |f - g_n| = \int_A |f(a) - \mathrm{E}\, f_n(a) + o(h_n)| = \int_A o(h_n) = o(h_n^2),$$

and therefore,

$$\mathrm{E}\{I_n^2\} \leqslant n\sum_{j=1}^{m_n} o(h_n^4) = nm_n o(h_n^4) \leqslant n(\mathrm{diam}(S_\gamma) + h_n)o(1/n) = o(1).$$

Lemma 4. *Let* g *be a function such that for* $A \in \mathscr{P}_n$ $\int_A \mathrm{E}\, f_n = \int_A f = \int_A g$. *Then*

$$\begin{aligned}
\int_A |g - f_{N_n}| = \int_A |\mathrm{E} f_n - f_{N_n}| \\
&+ 2I_{[\mu_{N_n}(A) > \mu(A)]}\int_A (g - \mathrm{E}\, f_n - |\mathrm{E}\, f_n - f_{N_n}|)^+ \\
&+ 2I_{[\mu_{N_n}(A) \leqslant \mu(A)]}\int_A (-(g - \mathrm{E}\, f_n) - |\mathrm{E}\, f_n - f_{N_n}|)^+.
\end{aligned}$$

Proof.

$$\int_A |g - f_{N_n}| = \int_A (f_{N_n} - g)^+ + \int_A (g - f_{N_n})^+ = \int_A (f_{N_n} - g) + 2\int_A (g - f_{N_n})^+$$

and

$$I_{[\mu_{N_n}(A) > \mu(A)]} \int_A (f_{N_n} - g) = I_{[\mu_{N_n}(A) > \mu(A)]} \int_A (f_{N_n} - \mathrm{E} f_n)$$

$$= I_{[\mu_{N_n}(A) > \mu(A)]} \int_A |f_{N_n} - \mathrm{E} f_n|.$$

Thus,

$$\int_A |g - f_{N_n}| I_{[\mu_{N_n}(A) > \mu(A)]} = I_{[\mu_{N_n}(A) > \mu(A)]} \int_A |f_{N_n} - \mathrm{E} f_n|$$

$$+ 2I_{[\mu_{N_n}(A) > \mu(A)]} \int_A (g - \mathrm{E} f_n - |\mathrm{E} f_n - f_{N_n}|)^+.$$

Similarly.

$$\int_A |g - f_{N_n}| I_{[\mu_{N_n}(A) \leq \mu(A)]} = I_{[\mu_{N_n}(A) \leq \mu(A)]} \int_A |f_{N_n} - \mathrm{E} f_n|$$

$$+ 2I_{[\mu_{N_n}(A) \leq \mu(A)]} \int_A (-(g - \mathrm{E} f_n) - |\mathrm{E} f_n - f_{N_n}|)^+.$$

Remark 2. Lemma 4 holds when f_{N_n} is replaced by f_n as well. Taking $f = g$, this means that for the histogram, the L_1 error is larger than the variation term.

Lemma 5. *Let g be a function such that for $A \in \mathscr{P}_n$, $\int_A \mathrm{E} f_n = \int_A f = \int_A g$, and let g be linear on A with slope C. Then*

$$\int_A |g - f_{N_n}| = \psi_\beta(|\mu_{N_n}(A) - \mu(A)|),$$

where

$$\beta = \frac{|C| h_n^2}{2}.$$

Defining

$$M_{N_n,A} = \frac{\sqrt{n}[\mu_{N_n}(A) - \mu(A)]}{\sqrt{\mu(A)}},$$

we have

$$\sqrt{n} \int_A |g - f_{N_n}| = \sqrt{\mu(A)}\, \psi_{\alpha(A)}(|M_{N_n,A}|),$$

where

$$\alpha(A) = \frac{|C|\sqrt{n}h_n^2}{2\sqrt{\mu(A)}}.$$

Proof. Let $A = [a, b]$, $h_n = b - a$, and $g(x) = Cx + D$, $x \in A$. Because of the condition on g, $\mathrm{E} f_n(a) = C(a + b)/2 + D$. Taking into account

$$\int_A |\mathrm{E} f_n - f_{N_n}| = |\mu_{N_n}(A) - \mu(A)|,$$

and Lemma 4, for $\mu_{N_n}(A) > \mu(A)$ we have

$$\begin{aligned}\int_A (g - \mathrm{E} f_n - |\mathrm{E} f_n - f_{N_n}|)^+ &= \int_A (g - \mathrm{E} f_n - |\mu_{N_n}(A) - \mu(A)|/h_n)^+ \\ &= \int_A (C(x - (a+b)/2) - |\mu_{N_n}(A) - \mu(A)|/h_n)^+ \,\mathrm{d}x \\ &= \frac{\beta}{2}\left(\left(1 - \frac{|\mu_{N_n}(A) - \mu(A)|}{\beta}\right)^+\right)^2.\end{aligned}$$

For $\mu_{N_n}(A) \leqslant \mu(A)$ we obtain a similar result.

Lemma 6. *Let* $H: \mathbb{R}^+ \longrightarrow [c, d)$, $0 \leqslant c < d \leqslant +\infty$ *be an increasing, differentiable and invertible function such that*

$$H'(u) \leqslant c_1 u^{\delta_1}, \quad \delta_1 < 2.$$

Let $(\sqrt{\lambda}M + \lambda)$ *and* N *be respectively a Poisson* (λ) *and a normal* $\mathcal{N}(0, 1)$ *random variable. Then there is a constant* C_0 *such that*

$$|\mathrm{E}\{H(|M|)\} - \mathrm{E}\{H(|N|)\}| \leqslant \frac{C_0}{\sqrt{\lambda}}$$

and

$$|\mathrm{E}\{H(M)I_{[M \geqslant 0]}\} - \mathrm{E}\{H(N)I_{[N \geqslant 0]}\}| \leqslant \frac{C_0}{\sqrt{\lambda}}.$$

Proof. Without loss of generality we may assume that $\lambda > 0$ is integer, thus

$$\sqrt{\lambda}M + \lambda = \sum_{i=1}^{\lambda} M_i,$$

where $M_1, \ldots, M_\lambda$ are i.i.d. Poisson (1). Therefore, by the Berry-Esseen inequality, there is a constant C_1 such that for $u \in \mathbb{R}$

$$|\mathrm{P}\{M \geqslant u\} - \mathrm{P}\{N \geqslant u\}| \leqslant \frac{C_1}{\sqrt{\lambda}} \frac{1}{1 + |u|^3}.$$

Thus,

$$\begin{aligned}|\mathrm{E}\{H(|M|)\}-\mathrm{E}\{H(|N|)\}| &= \left|\int_0^\infty \mathrm{P}\{H(|M|)>t\}\,\mathrm{d}t-\int_0^\infty \mathrm{P}\{H(|N|)>t\}\,\mathrm{d}t\right| \\ &= \left|\int_c^d \mathrm{P}\{H(|M|)>t\}\,\mathrm{d}t-\int_c^d \mathrm{P}\{H(|N|)>t\}\,\mathrm{d}t\right| \\ &\leqslant \int_c^d |\mathrm{P}\{|M|>H^{-1}(t)\}-\mathrm{P}\{|N|>H^{-1}(t)\}|\,\mathrm{d}t \\ &= \int_0^\infty |\mathrm{P}\{|M|>u\}-\mathrm{P}\{|N|>u\}|H'(u)\,\mathrm{d}u \\ &\leqslant \frac{C_1}{\sqrt{\lambda}}\int_0^\infty \frac{H'(u)}{1+u^3}\,\mathrm{d}u \\ &\leqslant \frac{C_1c_1}{\sqrt{\lambda}}\int_0^\infty \frac{u^{\delta_1}}{1+u^3}\,\mathrm{d}u \\ &= \frac{C_0}{\sqrt{\lambda}}.\end{aligned}$$

The proof of the second statement is analogous.

Lemma 7. *Properties of* $\psi_\alpha(u)$, *for* $\alpha>0$ *and* $u\geqslant 0$:

$$0\leqslant \psi_\alpha(u)'\leqslant 1,\quad 0\leqslant \psi_\alpha(u)\psi_\alpha(u)'\leqslant u,$$

$$|\psi_\alpha(u)-\psi_\alpha(v)|\leqslant |u-v|,\quad \frac{\alpha}{2}\leqslant \psi_\alpha(u)\leqslant \frac{\alpha}{2}+u.$$

Lemma 8. *Let* g_n *be defined as in Lemma* 3. *Then*

$$\lim_{n\to\infty}\mathrm{Var}\left\{\sqrt{n}\int_{S_{c,n}}|f_{N_n}-g_n|\right\}=\int_{S_c}V\left(\frac{c^{3/2}|f'|}{2\sqrt{f}}\right)f=\sigma_\gamma^2.$$

Proof. Let $\sqrt{\lambda}M+\lambda$ be Poisson (λ), then first we show that there is a universal constant C_2 such that

$$|\mathrm{Var}\{\psi_\alpha(|M|)\}-\mathrm{Var}\{\psi_\alpha(|N|)\}|\leqslant C_2\left(\frac{1+\mathrm{E}\{\psi_\alpha(|N|)\}}{\sqrt{\lambda}}+\frac{1}{\lambda}\right)$$

Defining, for $r>0$,

$$\Delta_r:=|\mathrm{E}\{\psi_\alpha(|M|)^r\}-\mathrm{E}\{\psi_\alpha(|N|)^r\}|,$$

we have

$$|\mathrm{Var}\{\psi_\alpha(|M|)\}-\mathrm{Var}\{\psi_\alpha(|N|)\}|\leqslant \Delta_2+\Delta_1^2+2\Delta_1\mathrm{E}\{\psi_\alpha(|N|)\}.$$

Thus we get the result by applying Lemma 6 with $H=\psi_\alpha$ and $H=\psi_\alpha^2$. Therefore denoting by a_{nj} the center of A_{nj} and setting

$$\alpha(A_{nj})=\frac{|f'(a_{nj})|\sqrt{n h_n^2}}{2\sqrt{\mu(A_{nj})}},$$

we have

$$\left|\mathrm{Var}\left\{\sqrt{n}\int_{S_{\cdot n}}|f_{N_n}-g_n|\right\}-\sum_{j=1}^{m_n}\mu(A_{nj})V\left(\frac{c^{3/2}|f'(a_{nj})|}{2\sqrt{\mu(A_{nj})/h_n}}\right)\right|$$

$$\leqslant\sum_{j=1}^{m_n}\mu(A_{nj})|\mathrm{Var}\{\psi_{\alpha(A_{nj})}(|M_{N_n,A_{nj}}|)\}-\mathrm{Var}\{\psi_{\alpha(A_{nj})}(|N|)\}|$$

$$\leqslant C_2\sum_{j=1}^{m_n}\mu(A_{nj})\left(\frac{1+\mathrm{E}\{\psi_{\alpha(A_{nj})}(|N|)\}}{\sqrt{n\mu(A_{nj})}}+\frac{1}{n\mu(A_{nj})}\right)$$

$$=C_2\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{m_n}\sqrt{\mu(A_{nj})}(1+\mathrm{E}\{\psi_{\alpha(A_{nj})}(|N|)\}+\frac{m_n}{n}\right)$$

$$\leqslant C_2\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{m_n}\sqrt{\mu(A_{nj})}\left(1+\frac{\alpha(A_{nj})}{2}+\mathrm{E}\{|N|\}\right)+\frac{m_n}{n}\right)$$

$$\leqslant C_3\left(\frac{1}{\sqrt{nh_n}}\sum_{j=1}^{m_n}\sqrt{\frac{\mu(A_{nj})}{h_n}}h_n+h_n\sum_{j=1}^{m_n}|f'(a_{nj})|h_n+\frac{m_n}{n}\right)$$

$$=C_3\left(\frac{1}{\sqrt{nh_n}}\int_{S_{\cdot}}\sqrt{f}(1+o(1))+h_n\int_{S_{\cdot}}|f'|(1+o(1))+\frac{m_n}{n}\right)\to 0.$$

Since

$$\sum_{j=1}^{m_n}\mu(A_{nj})V\left(\frac{c^{3/2}|f'(a_{nj})|}{2\sqrt{\mu(A_{nj})/h_n}}\right)\to\sigma_\gamma^2,$$

we get

$$\mathrm{Var}\left\{\sqrt{n}\int_{S_{\cdot n}}|f_{N_n}-g_n|\right\}\to\sigma_\gamma^2.$$

Lemma 9. (*Beirlant, Mason* (1994)): *If* $\sqrt{\lambda}M+\lambda$ *is Poisson* (λ) *then for each* $r\geqslant 1$ *there is* $K_r>0$ *such that*

$$\mathrm{E}\{|\sqrt{\lambda}M|^{2r}\}\leqslant K_r(\lambda^r+\lambda I_{[\lambda\leqslant 1]}).$$

Lemma 10. *If* $\sup_j\mu(A_{nj})\leqslant 1/4$, *then*

$$\left|\mathrm{E}\int|f_{N_n}-f|-\mathrm{E}\int|f_n-f|\right|\leqslant\frac{8\sup_j\sqrt{\mu(A_{nj})}}{\sqrt{n}}+2n\exp[-n(1-\log 2)/2].$$

Proof. By $f_{k,n}$ we denote the following density estimate:

$$f_{k,n}(x)=\frac{1}{nh}\sum_{j=1}^{k}I_{X_j\in A_n(x)}$$

where $A_n(x)$ is the set of $\mathscr{P}_n$ containing x.

Then $f_{n,n}(x)=f_n(x)$ and $f_{N_n,n}(x)=f_{N_n}(x)$. We begin with simple discrete sum calculus. Introduce the notation $J_k=\int|f_{k,n}-f|$, and $\Delta J_k=J_{k+1}-J_k$. Iterating this, we have $\Delta^2 J_j=J_{j+2}-2J_{j+1}+J_j$. It is trivial to see that

$$|\Delta^2 J_j|\leqslant 2/n$$

for all j. We show that

$$C_n:=\frac{8\sup_j\sqrt{\mu(A_{nj})}}{n^{3/2}}\geqslant \mathrm{E}\Delta^2 J_i$$

when $i\geqslant n/2$ and $\sup_j\mu(A_{nj})\leqslant 1/4$. Fix $i\geqslant n/2$. If K and L are the (random) indices j of the intervals in $\{A_{nj}\}$ to which X_{i+1} and X_{i+2} belong, and if P_j denotes the number of points among $X_1,\ldots,X_i$ that belong to A_{nj}, then

$$\begin{aligned}
&\mathrm{E}\{\Delta^2 J_i\}\\
&=\mathrm{E}\left\{\sum_{k,l} I_{K=k}I_{L=l}\Delta^2 J_i\right\}\\
&=\sum_{k\neq l}\mathrm{P}\{K=k,L=l\}\,\mathrm{E}\left\{\int_{A_{nl}}\left|\frac{P_l+1}{nh}-f\right|-\int_{A_{nl}}\left|\frac{P_l}{nh}-f\right|\right\}\\
&\quad-\sum_{k\neq l}\mathrm{P}\{K=k,L=l\}\,\mathrm{E}\left\{\int_{A_{nk}}\left|\frac{P_k+1}{nh}-f\right|-\int_{A_{nk}}\left|\frac{P_k}{nh}-f\right|\right\}\\
&\quad+\sum_k \mathrm{P}\{K=L=k\}\times\mathrm{E}\left\{\int_{A_{nk}}\left|\frac{P_k+2}{nh}-f\right|+\int_{A_{nk}}\left|\frac{P_k}{nh}-f\right|-2\int_{A_{nk}}\left|\frac{P_k+1}{nh}-f\right|\right\}\\
&=\sum_k \mu(A_{nk})^2\,\mathrm{E}\left\{\int_{A_{nk}}\left|\frac{P_k+2}{nh}-f\right|+\int_{A_{nk}}\left|\frac{P_k}{nh}-f\right|-2\int_{A_{nk}}\left|\frac{P_k+1}{nh}-f\right|\right\}
\end{aligned}$$

where we made heavy use of symmetry. Also,

$$\begin{aligned}
&\mathrm{E}\left\{\left|\int_{A_{nk}}\left|\frac{P_k+2}{nh}-f\right|+\int_{A_{nk}}\left|\frac{P_k}{nh}-f\right|-2\int_{A_{nk}}\left|\frac{P_k+1}{nh}-f\right|\right|\right\}\\
&\leqslant\frac{2}{nh}\mathrm{E}\lambda(A_{nk}\cap\{x:P_k<nhf(x)<P_k+2\})\\
&=\frac{2}{nh}\sum_{j=0}^{i}\lambda(A_{nk}\cap\{x:j<nhf(x)<j+2\})\mathrm{P}\{P_k=j\}\\
&\leqslant\frac{2}{nh}\max_j \mathrm{P}(P_k=j)\sum_{j=0}^{i}\lambda(A_{nk}\cap\{x:j<nhf(x)<j+2\})\\
&\leqslant\frac{2}{nh}\sqrt{\frac{2}{(i+1)\mu(A_{nk})}}\,2\lambda(A_{nk})\\
&=\frac{4}{n}\sqrt{\frac{2}{(i+1)\mu(A_{nk})}},
\end{aligned}$$

where in the last inequality we used that for a binomial (n, p) random variable B, with $p \leqslant 1/4$, we have

$$\sup_j \mathrm{P}\{B=j\} \leq \sqrt{\frac{2}{(n+1)p}},$$

which follows from standard upper bounds (see, e.g., Mitrinović, 1970, p. 197). Thus for $i \geq n/2$ and $\sup_j \mu(A_{nj}) \leq 1/4$,

$$\begin{aligned}
|\mathrm{E}\{\Delta^2 J_i\}| &\leq \sum_k \mu(A_{nk})^2 \frac{4}{n}\sqrt{\frac{2}{(i+1)\mu(A_{nk})}} \\
&\leq \sum_k \mu(A_{nk})^{3/2} \frac{8}{n^{3/2}} \\
&\leq \frac{8\sup_k \sqrt{\mu(A_{nk})}}{n^{3/2}}.
\end{aligned}$$

It is easy to see that

$$J_k = J_n + (k-n)\Delta J_n + \begin{cases} \sum_{i=n}^{k-1}(k-1-i)\Delta^2 J_i & \text{if} \quad k>n \\ \sum_{i=k-1}^{n-1}(i-k+1)\Delta^2 J_i & \text{if} \quad k<n. \end{cases}$$

In the above equations for J_k, we replace k by the Poisson (n) random variable N_n and take expectations. The coefficient of ΔJ_n drops out. Thus,

$$\begin{aligned}
\mathrm{E}\{J_{N_n} - J_n\} &= \mathrm{E}\left\{ I_{N>n} \sum_{i=n}^{N-1} (N-1-i)\mathrm{E}\Delta^2 J_i \right\} \\
&\quad + \mathrm{E}\left\{ I_{N<n} \sum_{i=N-1}^{n-1} (i-N+1)\mathrm{E}\Delta^2 J_i \right\}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
|\mathrm{E}\{J_{N_n} - J_n\}| &\leq C_n \mathrm{E}\left\{ I_{N>n} \sum_{i=n}^{N-1} (N-1-i) \right\} \\
&\quad + C_n \mathrm{E}\left\{ I_{n/2 \leq N<n} \sum_{i=N-1}^{N-1} (i-N+1) \right\} + \mathrm{E}\left\{ I_{N<n/2} \sum_{i=N-1}^{n-1} (i-N+1)\frac{2}{n} \right\} \\
&\leq C_n \mathrm{E}\{I_{N>n}(N-1-n)^2\} + C_n \mathrm{E}\{I_{n/2\leq N<n}(n-N)^2\} + 2n\mathrm{E}\{I_{N<n/2}\} \\
&\leq C_n \mathrm{E}(N-n)^2 + 2n\mathrm{P}\{N<n/2\} \\
&= C_n n + 2nP\{N<n/2\} \\
&\leq \frac{8\sup_k \sqrt{\mu(A_{nk})}}{\sqrt{n}} + 2n\exp[-n(1-\log 2)/2].
\end{aligned}$$

Proof of Theorem 1 First we show that

$$\sqrt{n}(\|f_n - f\| - \mathrm{E}\|f_{N_n} - f\|)/\sigma \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

from which, using Lemma 10, Theorem 1 follows. Now, check the conditions of Lemma 2. Choose the functions g_{nj} as

$$g_{nj}(x) = \sqrt{n}\left(\int_{A_{nj}}\left|\frac{x}{h_n} - f\right| - \mathrm{E}\int_{A_{nj}}\left|\frac{\mu_{N_n}(A_{nj})}{h_n} - f\right|\right) \quad (j = 1, 2, \ldots),$$

Introduce

$$S_n = t\sqrt{n}\sum_{j=1}^{m_n}\left(\int_{A_{nj}}\left|\frac{\mu_{N_n}(A_{nj})}{h_n} - f\right| - \mathrm{E}\int_{A_{nj}}\left|\frac{\mu_{N_n}(A_{nj})}{h_n} - f\right|\right) + v\frac{N_n - n}{\sqrt{n}},$$

for which a central limit result holds as we will show. Note that

$$\mathrm{Var}(S_n) = n\sum_{j=1}^{m_n}\left\{t^2\,\mathrm{Var}\left\{\int_{A_{nj}}\left|\frac{\mu_{N_n}(A_{nj})}{h_n} - f\right|\right\}\right.$$
$$\left. + 2tv\,\mathrm{E}\left\{\int_{A_{nj}}\left|\frac{\mu_{N_n}(A_{nj})}{h_n} - f\right|(\mu_{N_n}(A_{nj}) - \mu(A_{nj}))\right\}\right\} + v^2.$$

By Lemmas 3 and 8

$$n\sum_{j=1}^{m_n}\mathrm{Var}\left\{\int_{A_{nj}}\left|\frac{\mu_{N_n}(A_{nj})}{h_n} - f\right|\right\} \to \int_{S} V\left(\frac{c^{3/2}|f'|}{2\sqrt{f}}\right)f.$$

To finalize the asymptotics for $\mathrm{Var}(S_n)$ it remains to show that

$$n\sum_{j=1}^{m_n}\mathrm{E}\left\{\int_{A_{nj}}\left|\frac{\mu_{N_n}(A_{nj})}{h_n} - f\right|(\mu_{N_n}(A_{nj}) - \mu(A_{nj}))\right\} \to 0 \quad \text{as} \quad n \to \infty.$$

Because of the proof of Lemma 3 it suffices to show that

$$n\sum_{j=1}^{m_n}\mathrm{E}\left\{\int_{A_{nj}}\left|\frac{\mu_{N_n}(A_{nj})}{h_n} - g_n\right|(\mu_{N_n}(A_{nj}) - \mu(A_{nj}))\right\} \to 0 \quad \text{as} \quad n \to \infty.$$

Then

$$n\sum_{j=1}^{m_n}\mathrm{E}\left\{\int_{A_{nj}}|f_{N_n} - g_n|(\mu_{N_n}(A_{nj}) - \mu(A_{nj}))\right\}$$
$$= n\sum_{j=1}^{m_n}\mathrm{E}\{\psi_{\alpha(A_{nj})}(|M_{N_n,A_{nj}}|)M_{N_n,A_{nj}}\}.$$

Applying Lemma 6 with $H = \psi_{\alpha(A_{nj})}$ we get

$$|\mathrm{E}\{\psi_{\alpha(A_{nj})}(|M_{N_n,A_{nj}}|)M^+_{N_n,A_{nj}}\} - \mathrm{E}\{\psi_{\alpha(A_{nj})}(|N|)N^+\}| \leqslant \frac{C_0}{\sqrt{n\mu(A_{nj})}}$$

and similarly for $M^{-}_{N_n,A_{nj}}$ and N^{-}. As $\mathrm{E}\{\psi_{\alpha(A_{nj})}(|N|)N\}=0$ we get

$$n\left|\sum_{j=1}^{m_n}\mathrm{E}\left\{\int_{A_{nj}}|f_{N_n}-g_n|(\mu_{N_n}(A_{nj})-\mu(A_{nj}))\right\}\right|\leqslant\frac{2C_0}{\sqrt{nh_n}}\sum_{j=1}^{m_n}\sqrt{\frac{\mu(A_{nj})}{h_n}}h_n$$

and the sum in the right hand side tends to $\int_S\sqrt{f}$. This completes the calculation of the asymptotic variance. To finish the proof of

$$S_n\xrightarrow{\mathscr{L}}\mathcal{N}\left(0,t^2\int_S V\left(\frac{c^{3/2}|f'|}{2\sqrt{f}}\right)f+v^2\right)\quad\text{as}\quad n\to\infty,$$

we apply Lyapunov's central limit theorem, and note that we only need to show that

$$\sum_{j=1}^{m_n}\mathrm{E}\left\{\left|t\int_{A_{nj}}\left|\frac{\mu_{N_n}(A_{nj})}{h_n}-f\right|-\mathrm{E}\int_{A_{nj}}\left|\frac{\mu_{N_n}(A_{nj})}{h_h}-f\right|\right)+v(\mu_{N_n}(A_{nj})-\mu(A_{nj}))\right|^3\right\}$$

and

$$\sum_{j=m_n+1}^{\infty}\mathrm{E}\{|v(\mu_{N_n}(A_{nj})-\mu(A_{nj}))|^3\}$$

are both $o(n^{-3/2})$. By invoking the c_r inequality, this would follow from

$$LY_n=n^{3/2}\sum_{j=1}^{m_n}\mathrm{E}\left\{\left|\int_{A_{nj}}\left|\frac{\mu_{N_n}(A_{nj})}{h_n}-f\right|-\mathrm{E}\int_{A_{nj}}\left|\frac{\mu_{N_n}(A_{nj})}{h_n}-f\right|\right|^3\right\}\to 0,$$

and

$$n^{3/2}\sum_{j=1}^{\infty}\mathrm{E}\{|\mu_{N_n}(A_{nj})-\mu(A_{nj})|^3\}\to 0.$$

This last statement is shown in Beirlant, Györfi, Lugosi (1994). In order to show the former, let F_{nj} be the distribution function of $|M_{N_n,A_{nj}}|$. Then

$$\begin{aligned}LY_n&=\sum_{j=1}^{m_n}\mathrm{E}\left\{\left|\sqrt{\mu(A_{nj})}(\psi_{\alpha(A_{nj})}(|M_{N_n,A_{nj}}|)-\mathrm{E}\{\psi_{\alpha(A_{nj})}(|M_{N_n,A_{nj}}|)\}\right|^3\right\}\\&=\sum_{j=1}^{m_n}\mu(A_{nj})^{3/2}\int\left|\psi_{\alpha(A_{nj})}(u)-\int\psi_{\alpha(A_{nj})}(v)\mathrm{d}F_{nj}(v)\right|^3\mathrm{d}F_{nj}(u)\\&\leqslant\sum_{j=1}^{m_n}\mu(A_{nj})^{3/2}\int\int|\psi_{\alpha(A_{nj})}(u)-\psi_{\alpha(A_{nj})}(v)|^3\mathrm{d}F_{nj}(u)\mathrm{d}F_{nj}(v)\\&\leqslant\sum_{j=1}^{m_n}\mu(A_{nj})^{3/2}\int\int|u-v|^3\mathrm{d}F_{nj}(u)\mathrm{d}F_{nj}(v)\\&\leqslant 4\sum_{j=1}^{m_n}\mu(A_{nj})^{3/2}\int\int(u^3+v^3)\mathrm{d}F_{nj}(u)\mathrm{d}F_{nj}(v)\\&=8\sum_{j=1}^{m_n}\mu(A_{nj})^{3/2}\mathrm{E}\{|M_{N_n,A_{nj}}|^3\}.\end{aligned}$$

By Lemma 9,

$$\mathrm{E}\{|M_{N_n,A_{nj}}|^3\} \leqslant K_{1.5}\left(\frac{(n\mu(A_{nj}))^{3/2}}{(n\mu(A_{nj}))^{3/2}} + \frac{n\mu(A_{nj})}{(n\mu(A_{nj}))^{3/2}} I_{[n\mu(A_{nj}) \leqslant 1]}\right)$$
$$\leqslant K_{1.5}\left(1 + \frac{1}{(n\mu(A_{nj}))^{1/2}} I_{[n\mu(A_{nj}) \leqslant 1]}\right).$$

Thus,

$$LY_n \leqslant 8K_{1.5}\left(\sum_{j=1}^{m_n} \mu(A_{nj})^{3/2} + \frac{1}{\sqrt{n}} \sum_{j=1}^{m_n} \mu(A_{nj}) I_{[n\mu(A_{nj}) \leqslant 1]}\right)$$
$$\leqslant 8K_{1.5}\left(\max_{j \leqslant m_n} \sqrt{\mu(A_{nj})} + \frac{1}{\sqrt{n}}\right) \to 0,$$

and we get the first limit relation in Lyapunov's condition.

References

[1] Bartlett, M. S. (1938). The characteristic function of a conditional statistic. *London Math. Soc.* **13**, 62-67.
[2] 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.* **3**, 309-318.
[3] Beirlant, J. and Mason, D. M. (1994). On the asymptotic normality of L_p norms of empirical functionals. *Mathematical Methods of Statistics*. To appear.
[4] Chung, K. L. (1974). *A Course in Probability Theory*. Academic Press, New York.
[5] Csörgő, M. and Horváth, L. (1988). Central limit theorems for L_p-norms of density estimators. *Probab. Theory Related Fields* **80**, 269-291.
[6] Devroye, L. and Györfi, L. (1985). *Nonparametric Density Estimation: The L_1-View*, Wiley, New York.
[7] Devroye, L. (1988). The kernel estimate is relatively stable. *Probab. Theory Related Fields* **77**, 521-536.
[8] Devroye, L. (1991). Exponential inequalities in nonparametric estimation In: *Nonparametric Functional Estimation and Related Topics*, G. Roussas (ed.), Kluwer Academic Publishers, 631-645.
[9] Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivariate Anal.* **14**, 1-16.
[10] Härdle, W. (1991). *Smoothing techniques with implementation in S*, Springer-Verlag, New York.
[11] Horváth, L. (1991). On L_p-norms of multivariate density estimators. *Ann. Statist.* **19**, 1933-1949.
[12] Mitrinovic, D. S. (1970). *Analytic Inequalities*, Springer-Verlag, Berlin.
[13] Scott, D. W. (1985). Averaged shifted histograms: effective nonparametric density estimators in several dimensions. *Ann. Statist.* **13**, 1024-1040.
[14] Tong, J. L. (1980). *Probability Inequalities in Multivariate Distributions*, Academic Press, New York.

A. Berlinet
Laboratoire de Biométrie
ENSAM-INRA, Université Montpellier II
9, Place Viala
34060 Montpellier cedex
France

L. Devroye
School of Computer Science
McGill University
H3A 2A7, Montreal
Canada

L. Györfi
Department of Mathematics and Computer Science
Technical University of Budapest
1521 Stoczek u.2, Budapest
Hungary