INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES
Série des Documents de Travail du CREST
(Centre de Recherche en Economie et Statistique)

n° 2008-09

Asymptotic Normality of Frequency Polygons for Random Fields

M. CARBON[1]

[1] Université de Rennes 2, (*Address : Place du Recteur Le Moal, 35043 Rennes Cedex, France),* et CREST-ENSAI.

ASYMPTOTIC NORMALITY OF FREQUENCY POLYGONS FOR RANDOM FIELDS

by

Michel Carbon

Université de Rennes 2 et Ensai

AMS 1980 subject classifications: primary: 62G05; secondary: 62M40, 62M30.

Keywords and phrases: random field, frequency polygons, bandwidth, asymptotic normality.

ABSTRACT

The purpose of this paper is to investigate asymptotic normality of the frequency polygon estimator of a stationary mixing random field indexed by multidimensional lattice points space Z^N. Appropriate choices of the bandwiths are found.

Send Correspondence to:

Carbon Michel

Place du Recteur Le Moal

University of Rennes 2

35043 Rennes Cédex

1. Introduction. Our goal in this paper is to study frequency polygon as a density estimator for random variables which show spatial interaction. We sense a practical need for nonparametric spatial estimation for situations in which parametric families cannot be adopted with confidence. The frequency polygon is constructed by connecting with straight lines the mid-bin values of a histogram. So, the computational effort in constructing the frequency polygon is about equivalent to the histogram.

Denote the integer lattice points in the N-dimensional Euclidean space by $Z^N, N \geq 1$. Consider a strictly stationary random field $\{X_{\mathbf{n}}\}$ indexed by $\mathbf{n}$ in Z^N and defined on some probability space $(\Omega, \mathcal{F}, P)$. A point $\mathbf{n}$ in Z^N will be referred to as a site. For a site $\mathbf{i} = (i_1, \cdots, i_N)$, we denote $\|\mathbf{i}\| = (i_1^2 + \cdots + i_N^2)^{1/2}$. We will write n instead of $\mathbf{n}$ when $N = 1$. For two finite sets of sites S and S', the Borel fields $\mathcal{B}(S) = \mathcal{B}(X_{\mathbf{n}}, \mathbf{n} \in S)$ and $\mathcal{B}(S') = \mathcal{B}(X_{\mathbf{n}}, \mathbf{n} \in S')$ are the σ-fields generated by the random variables $X_{\mathbf{n}}$ with $\mathbf{n}$ ranging over S and S' respectively. Denote the Euclidean distance between S and S' by dist (S, S'). We will assume that $X_{\mathbf{n}}$ satisfies the following mixing condition: there exists a function $\varphi(t) \downarrow 0$ as $t \to \infty$, such that whenever $S, S' \subset Z^N$,

$$(1.1) \qquad \alpha(\mathcal{B}(S), \mathcal{B}(S')) = \sup\{|P(AB) - P(A)P(B)|, A \in \mathcal{B}(S), B \in \mathcal{B}(S')\}$$

$$\leq h(\mathrm{Card}(S), \mathrm{Card}(S'))\varphi(\mathrm{dist}\,(S, S')),$$

where Card(S) denotes the cardinality of S. Here h is a symmetric positive function nondecreasing in each variable. Throughout the paper, assume that h satisfies either

$$(1.2) \qquad h(n, m) \leq \min\{m, n\}$$

or

$$(1.3) \qquad h(n, m) \leq C(n + m + 1)^{\tilde{k}}$$

for some $\tilde{k} \geq 1$ and some $C > 0$. If $h \equiv 1$, then $X_{\mathbf{n}}$ is called strongly mixing. Conditions (1.2) and (1.3) are the same as the mixing conditions used by Neaderhouser (1980) and

Takahata (1983) respectively and are weaker than the uniform mixing condition used by Nahapetian (1980). They are satisfied by many spatial models. Examples can be found in Neaderhouser (1980), Rosenblatt (1985) and Guyon (1987). For relevant works on random fields, see e.g. Neaderhouser (1980), Bolthausen (1982), Guyon and Richardson (1984), Guyon (1987), Nahapetian (1987), Tran (1990), Tran and Yakowitz (1993), Carbon, Hallin and Tran (1996), Carbon, Tran and Wu (1997), Francq, C. and Tran, L.T.(2002), Carbon, Francq and Tran (2007).

Denote by $\mathcal{I}_{\mathbf{n}}$ a rectangular region defined by

$$\mathcal{I}_{\mathbf{n}} = \{\mathbf{i} : \mathbf{i} \in Z^N, 1 \leq i_k \leq n_k, k = 1, \ldots, N\}.$$

Assume that we observe $\{X_{\mathbf{n}}\}$ on $\mathcal{I}_{\mathbf{n}}$.

Suppose $X_{\mathbf{n}}$ takes values in R and has a uniformly continuous density f with a bounded derivative. We write $\mathbf{n} \to \infty$ if

$$\min\{n_k\} \to \infty \quad \text{and} \quad |n_j/n_k| < C \tag{1.4}$$

for some $0 < C < \infty$, $1 \leq j, k \leq N$. All limits are taken as $\mathbf{n} \to \infty$ unless indicated otherwise.

Define $\hat{\mathbf{n}} = n_1 \ldots n_N$.

For background material on frequency polygons, see Scott (1985 and 1992). For rates of convergence of frequency polygons, see Carbon (2006).

Our paper is organized as follows: Section 2 provides some preliminaries and background material. In Section 3, we give the expression of the asymptotic variance of $f_{\mathbf{n}}$, and in section 4, we prove the asymptotic normality of $f_{\mathbf{n}}$, with application to confidence interval.

We use x to denote a fixed point of R. The integer part of a number a is denoted by $[a]$. The letter C will be used to denote constants whose values are unimportant. The letter D denotes an arbitrary compact set in R.

2. Preliminaries. Consider a partition of the real line into equal intervals $I_k = [(k-1)b, kb)$ of length $b = b_{\mathbf{n}}$, where $b_{\mathbf{n}}$ is the bin width. Consider the two adjacent histogram bins $I_{k_0} = [(k_0-1)b, k_0 b)$ and $I_{k_1} = [k_0 b, (k_0+1)b)$ where $k_1 = k_0 + 1$. Denote the number of observations falling in these intervals respectively by ν_{k_0} and ν_{k_1}. The values of the histogram in these previous bins are given by $f_{k_0} = \nu_{k_0}\hat{\mathbf{n}}^{-1}b^{-1}$ and $f_{k_1} = \nu_{k_1}\hat{\mathbf{n}}^{-1}b^{-1}$. The frequency polygon $f_{\mathbf{n}}(x)$ is given by

(2.1)

$$f_{\mathbf{n}}(x) = \left(\frac{1}{2} + k_0 - \frac{x}{b}\right) f_{k_0} + \left(\frac{1}{2} - k_0 + \frac{x}{b}\right) f_{k_1}, \quad \text{for} \quad \left(k_0 - \frac{1}{2}\right) b \le x < \left(k_0 + \frac{1}{2}\right) b.$$

We assume that b tends to zero as $\mathbf{n} \to \infty$. Define

$$Y_{\mathbf{i},k} = \begin{cases} 1, & \text{if } X_{\mathbf{i}} \in I_k; \\ 0, & \text{otherwise.} \end{cases}$$

Then,

$$\nu_{k_0} = \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i},k_0} \quad \text{and} \quad \nu_{k_1} = \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i},k_1}.$$

Let $U = u(S)$ and $V = v(S')$, where u and v are real-valued measurable functions.

Lemma 2.1. *Suppose that $|u| \le C_1$ and $|v| \le C_2$ where C_1 and C_2 are constants. Then*

$$|EUV - EUEV| \le Ch(Card(S), Card(S'))\varphi(dist(S, S')).$$

Lemma 2.2. *Suppose that $\|U\|_r < \infty$ and $\|V\|_s < \infty$ where $\|U\|_r = (E|U|^r)^{1/r}$. If r, s and τ are positive numbers and $r^{-1} + s^{-1} + \tau^{-1} = 1$, then*

$$|EUV - EUEV| \le C\|U\|_r \|V\|_s \{h(Card(S), Card(S'))\varphi(dist(S, S'))\}^{1/\tau}.$$

One or both of r and s can be taken to be ∞ for bounded random variables. For the proof of the Davydov inequality in Lemma 2.2, see Davydov (1970), Deo (1973), Hall and Heyde (1980) or Tran(1990).

Denote $\eta_{\mathbf{i},k} = Y_{\mathbf{i},k} - EY_{\mathbf{i},k}$.

Corollary 2.1. *For $\gamma : 0 < \gamma < 1$, for each integer k, there exists some $\xi_k \in I_k$ such that* *(i)* $|cov(\eta_{\mathbf{i},k}, \eta_{\mathbf{j},k}| \le C(f(\xi_k)b)^\gamma(\varphi(\|\mathbf{i}-\mathbf{j}\|))^{1-\gamma}$,

(ii) $|cov(\eta_{\mathbf{i},k-1}, \eta_{\mathbf{j},k}| \le C(f(\xi_k)b)^\gamma(\varphi(\|\mathbf{i}-\mathbf{j}\|))^{1-\gamma}$.

Proof. (i) Choosing $r = s = 2\gamma^{-1}, \tau = (1-\gamma)^{-1}$, Lemma 2.2 leads to the following result: if $E|U|^{2/\gamma} < +\infty$ and $E|V|^{2/\gamma} < +\infty$, then

$$|EUV - EUEV| \le C\|U\|_{2/\gamma}\|V\|_{2/\gamma}\{\varphi(\|\mathbf{j}-\mathbf{i}\|)\}^{1-\gamma}.$$

Taking $U = Y_{\mathbf{i},k}$ and $V = Y_{\mathbf{j},k}$, and using the mean-value theorem we have

$$\|U\|_{2/\gamma} = (EY_{\mathbf{i},k})^{\gamma/2} = (P[X_{\mathbf{1}} \in I_k])^{\gamma/2} = (f(\xi_k)b)^{\gamma/2}, \text{ where } \xi_k \in I_k.$$

The proof of (i) thus follows.

(ii) The proof can be handled in the same way. Note that ξ_k is independent of $\mathbf{i}$ and $\mathbf{j}$. □

Denote the conditional density of $X_{\mathbf{j}}$ given $X_{\mathbf{i}}$ by $f_{\mathbf{j}|\mathbf{i}}$ for simplicity.

Assumption 1. For all $\mathbf{i}$, $\mathbf{j}$ and some constant M_1,

$$\sup_{(x,y)\in R\times R} f_{\mathbf{j}|\mathbf{i}}(y|x) \le M_1.$$

Example. In the case $N = 1$, let X_t be a stationary autoregressive process of order 1, for example, $X_t = \theta X_{t-1} + e_t$ where $|\theta| < 1$. Assume the e_t's are i.i.d. random variables and each e_t has a standard Cauchy density. Then

$$X_j = \theta^{j-i}X_i + Z,$$

where Z is a Cauchy r.v. independent of X_i (see Example 2.1 in Tran (1989)) with characteristic function

$$\exp(-|u|(1-\theta^{j-i})/(1-\theta)).$$

The conditional density of X_j given X_i is equal to

$$f_{j|i}(x_j|x_i)) = f_Z(x_j - \theta^{j-i}x_i).$$

A Cauchy density symmetric about zero takes on its maximum value at zero. Thus we can take

$$M_1 = \frac{1}{\pi(1-|\theta|)}.$$

If e_t is assumed to be $\mathcal{N}(0, \sigma^2)$ distributed, instead of Cauchy distributed, then one can take

$$M_1 = \frac{1}{\sigma\sqrt{2\pi}}.$$

Lemma 2.3. *If Assumption 1 is satisfied, then*

$$\int\int_{I_k \times I_k} |f_{\mathbf{i},\mathbf{j}}(x,y) - f(x)f(y)| \; dx\, dy \le M f(\zeta_k) b^2 \quad \textit{with} \quad \zeta_k \in I_k. \tag{2.2}$$

Proof. Since f is uniformly continuous and integrable,

$$\sup_{x \in R} f(x) \equiv \|f\| < \infty.$$

By Assumption 1,

$$\begin{aligned} &\int\int_{I_k \times I_k} |f_{\mathbf{i},\mathbf{j}}(x,y) - f(x)f(y)| \; dx\, dy \\ \le &\int\int_{I_k \times I_k} f(x) \left| f_{\mathbf{j}|\mathbf{i}}(y|x) - f(y) \right| \; dx\, dy \\ &\le Mb \int_{I_k} f(x) dx, \end{aligned}$$

where M can be taken to be $2\max\{M_1, \|f\|\}$. The lemma follows by the mean-value theorem. □

For convenience, we define

$$p_k = \int_{I_k} f(x) \; dx$$

and

$$q_{1\mathbf{n}} = \frac{1}{\hat{\mathbf{n}}^2 b^2} \sum_{\mathbf{i} \neq \mathbf{j}} \mathrm{cov}(\eta_{\mathbf{i},k_0}, \eta_{\mathbf{j},k_0}),$$

$$q_{2\mathbf{n}} = \frac{1}{\hat{\mathbf{n}}^2 b^2} \sum_{\mathbf{i} \neq \mathbf{j}} \mathrm{cov}(\eta_{\mathbf{i},k_1}, \eta_{\mathbf{j},k_1}),$$

$$q_{3\mathbf{n}} = \frac{1}{\hat{\mathbf{n}}^2 b^2} \sum_{\mathbf{i} \neq \mathbf{j}} \mathrm{cov}(\eta_{\mathbf{i},k_0}, \eta_{\mathbf{j},k_1}).$$

Lemma 2.4. *The variance of the frequency polygon $f_{\mathbf{n}}(x)$ defined in (2.1) is given by*

$$var f_{\mathbf{n}}(x) = \left(\frac{1}{2} + k_0 - \frac{x}{b}\right)^2 \Big[\frac{1}{\hat{\mathbf{n}} b^2} p_{k_0}(1 - p_{k_0}) + q_{1\mathbf{n}}\Big] \tag{2.3}$$

$$+\left(\frac{1}{2} - k_0 + \frac{x}{b}\right)^2 \Big[\frac{1}{\hat{\mathbf{n}} b^2} p_{k_1}(1 - p_{k_1}) + q_{2\mathbf{n}}\Big] + 2\left(\frac{1}{4} - {k_0}^2 + \frac{2xk_0}{b} - \frac{x^2}{b^2}\right)\Big[-\frac{p_{k_0} p_{k_1}}{\hat{\mathbf{n}} b^2} + q_{3\mathbf{n}}\Big].$$

Proof. From the expression of the frequency polygon (2.1),

$$\mathrm{var} f_{\mathbf{n}}(x) = \left(\frac{1}{2} + k_0 - \frac{x}{b}\right)^2 \mathrm{var} f_{k_0} + \left(\frac{1}{2} - k_0 + \frac{x}{b}\right)^2 \mathrm{var} f_{k_1}$$

$$+2\left(\frac{1}{4} - {k_0}^2 + \frac{2xk_0}{b} - \frac{x^2}{b^2}\right) \mathrm{cov}(f_{k_0}, f_{k_1}).$$

Clearly,

$$\mathrm{var} f_{k_0} = \frac{1}{\hat{\mathbf{n}}^2 b^2} \mathrm{var}\left(\sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i},k_0}\right)$$

with

$$\mathrm{var}\Big(\sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i},k_0}\Big) = \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} \mathrm{var}\Big(Y_{\mathbf{i},k_0}\Big) + \sum_{\mathbf{i} \neq \mathbf{j}} \mathrm{cov}(Y_{\mathbf{i},k_0}, Y_{\mathbf{j},k_0})$$

$$= \hat{\mathbf{n}} p_{k_0}(1 - p_{k_0}) + \sum_{\mathbf{i} \neq \mathbf{j}} \mathrm{cov}(\eta_{\mathbf{i},k_0}, \eta_{\mathbf{j},k_0}).$$

Similarly,

$$\mathrm{var} f_{k_1} = \frac{1}{\hat{\mathbf{n}}^2 b^2} \mathrm{var}\left(\sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i},k_1}\right) = \frac{p_{k_1}(1 - p_{k_1})}{\hat{\mathbf{n}} b^2} + q_{2\mathbf{n}}.$$

We get also

$$\mathrm{cov}(f_{k_0},\ f_{k_1}) = \frac{1}{\hat{\mathbf{n}}^2 b^2} \mathrm{cov}\Big(\sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i},k_0},\ \sum_{\mathbf{j} \in \mathcal{I}_{\mathbf{n}}} Y_{\mathbf{j},k_1}\Big) = \frac{1}{\hat{\mathbf{n}}^2 b^2} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} \mathrm{cov}\left(Y_{\mathbf{i},k_0},\ Y_{\mathbf{i},k_1}\right) + q_{3\mathbf{n}}.$$

But,

$$\mathrm{cov}\left(Y_{\mathbf{i},k_0},\ Y_{\mathbf{i},k_1}\right) = E\left(Y_{\mathbf{i},k_0} Y_{\mathbf{i},k_1}\right) - E\left(Y_{\mathbf{i},k_0}\right) E\left(Y_{\mathbf{i},k_1}\right) = -p_{k_0} p_{k_1}.$$

Then (2.3) follows. □

3. Asymptotic variance of $f_{\mathbf{n}}$.

Lemma 3.1. *Assume Assumption 1 holds and* $X_{\mathbf{n}}$ *satisfies (1.1) and (1.2) or (1.3) with* $\sum_{i=1}^{\infty} i^{N-1}(\varphi(i))^{\beta} < +\infty$ *for some* $0 < \beta < 1/2$. *For* x *in* $J_{k_0} = [k_0 b - b/2, k_0 b + b/2)$, *then we have*

$$\hat{\mathbf{n}} b \left(\frac{1}{2} + k_0 - \frac{x}{b}\right)^2 |q_{1\mathbf{n}}| + \hat{\mathbf{n}} b \left(\frac{1}{2} - k_0 + \frac{x}{b}\right)^2 |q_{2\mathbf{n}}|$$

$$+2\hat{\mathbf{n}} b \left(\frac{1}{4} - {k_0}^2 + \frac{2xk_0}{b} - \frac{x^2}{b^2}\right)|q_{3\mathbf{n}}| \longrightarrow 0 \ \textit{as} \ \mathbf{n} \to +\infty. \tag{3.1}$$

Proof. By Corollary 2.1 and Lemma 2.3, we have

$$|\mathrm{cov}(\eta_{\mathbf{i},k_0}, \eta_{\mathbf{j},k_0})| \le \min\left(C(f(\xi_{k_0})b)^{\gamma}(\varphi(\|\mathbf{i}-\mathbf{j}\|))^{1-\gamma}; Mf(\zeta_{k_0})b^2\right) \tag{3.2}$$

Let $K_{\mathbf{n}} = b_{\mathbf{n}}^{-(1-\gamma)\nu^{-1}}$, where $\nu = -N - \varepsilon + (1-\gamma)N\beta^{-1}$ with γ and ε being small positive numbers such that $\beta^{-1} - (N+\varepsilon)(N(1-\gamma))^{-1} \ge 1$. This can be done since $0 < \beta < 1/2$. Note also that $\nu > N(1-\gamma)$.

Define

$$S_1 = \{\mathbf{i}, \mathbf{j} \in \mathcal{I}_{\mathbf{n}} \mid 0 < \|\mathbf{i}-\mathbf{j}\| \le K_{\mathbf{n}}\},$$
$$S_2 = \{\mathbf{i}, \mathbf{j} \in \mathcal{I}_{\mathbf{n}} \mid \|\mathbf{i}-\mathbf{j}\| > K_{\mathbf{n}}\}.$$

Split $\sum_{\mathbf{i}\neq\mathbf{j}} \left|\mathrm{cov}(\eta_{\mathbf{i},k_0}, \eta_{\mathbf{j},k_0})\right|$ into two separate summations A_1 and A_2 over sites S_1 and S_2. Then

$$\sum_{\mathbf{i}\neq\mathbf{j}} \left|\mathrm{cov}(\eta_{\mathbf{i},k_0}, \eta_{\mathbf{j},k_0})\right| \le A_1 + A_2.$$

Now, using (3.2), we have the following upper bound

$$A_1 = \sum_{\mathbf{i},\mathbf{j}\in S_1} \left|\mathrm{cov}(\eta_{\mathbf{i},k_0}, \eta_{\mathbf{j},k_0})\right| \le Mf(\zeta_{k_0})b^2\hat{\mathbf{n}}K_{\mathbf{n}}^N .$$

Thus, with $\theta = \dfrac{\nu - N(1-\gamma)}{\nu}$ $(0 < \theta < 1)$, we have

$$A_1 \le Mf(\zeta_{k_0})\hat{\mathbf{n}}b^{2-(1-\gamma)\nu^{-1}} \le Mf(\zeta_{k_0})\hat{\mathbf{n}}b^{1+\theta} .$$

Using (3.2), we have successively

$$\begin{aligned}A_2 =& \sum_{\mathbf{i},\mathbf{j}\in S_2} \left|\operatorname{cov}(\eta_{\mathbf{i},k_0}, \eta_{\mathbf{j},k_0})\right| \\ \leq& C(f(\xi_{k_0}))^\gamma b^\gamma \sum_{\mathbf{i},\mathbf{j}\in S_2} (\varphi(\|\mathbf{i}-\mathbf{j}\|))^{1-\gamma} \\ \leq& C(f(\xi_{k_0}))^\gamma b^\gamma \hat{\mathbf{n}} \sum_{\|\mathbf{i}\|>K_\mathbf{n}} (\varphi(\|\mathbf{i}\|))^{1-\gamma} \\ \leq& C(f(\xi_{k_0}))^\gamma b^\gamma \hat{\mathbf{n}} K_\mathbf{n}^{-\nu} \sum_{\|\mathbf{i}\|>K_\mathbf{n}} \|\mathbf{i}\|^\nu (\varphi(\|\mathbf{i}\|))^{1-\gamma}\end{aligned}$$

By assumption $\sum_{i=1}^{\infty} i^{N-1}(\varphi(i))^\beta < +\infty$. Thus, $i^{N-1}(\varphi(i))^\beta = o(1/i)$ or $\varphi(i) = o(i^{-N/\beta})$ as $i \to +\infty$. Since φ is a nonincreasing function, we have $\varphi(x) = o(x^{-N/\beta})$ as $x \to +\infty$. Therefore

$$\begin{aligned}\|\mathbf{i}\|^\nu \left(\varphi(\|\mathbf{i}\|)\right)^{1-\gamma} =& \|\mathbf{i}\|^\nu \; o\left(\|\mathbf{i}\|^{-N(1-\gamma)\beta^{-1}}\right) \\ =& \, o\left(\|\mathbf{i}\|^{-N-\varepsilon}\right),\end{aligned}$$

since $\nu = -N - \varepsilon + (1-\gamma)N\beta^{-1}$. Thus

$$\sum_{\substack{i_h=1 \\ h=1,\cdots,N}}^{+\infty} \|\mathbf{i}\|^\nu (\varphi(\|\mathbf{i}\|))^{1-\gamma} < +\infty. \tag{3.3}$$

Using the majorization of A_2 and (3.3), and noting that $b^\gamma K_\mathbf{n}^{-\nu} = b$, we obtain

$$A_2 \leq C(f(\xi_{k_0}))^\gamma \, b \, \hat{\mathbf{n}} \sum_{\|\mathbf{i}\|>K_\mathbf{n}} \|\mathbf{i}\|^\nu (\varphi(\|\mathbf{i}\|))^{1-\gamma}.$$

Finally

$$|q_{1\mathbf{n}}| \leq M f(\zeta_{k_0}) \hat{\mathbf{n}}^{-1} b^{-1+\theta} + C(f(\xi_{k_0}))^\gamma \, b^{-1} \, \hat{\mathbf{n}}^{-1} \sum_{\|\mathbf{i}\|>K_\mathbf{n}} \|\mathbf{i}\|^\nu (\varphi(\|\mathbf{i}\|))^{1-\gamma}.$$

and

$$\hat{\mathbf{n}}\, b \left(\frac{1}{2} + k_0 - \frac{x}{b}\right)^2 |q_{1\mathbf{n}}| \leq M f(\zeta_{k_0}) b^\theta + C(f(\xi_{k_0}))^\gamma \sum_{\|\mathbf{i}\|>K_\mathbf{n}} \|\mathbf{i}\|^\nu (\varphi(\|\mathbf{i}\|))^{1-\gamma},$$

which tends to 0 since $0 \leq \left(\frac{1}{2} + k_0 - \frac{x}{b}\right) \leq 1$, $b \to 0$ and $K_\mathbf{n} \to +\infty$. Using the same arguments, it can be shown that $\hat{\mathbf{n}}\, b \left(\frac{1}{2} - k_0 + \frac{x}{b}\right)^2 |q_{2\mathbf{n}}|$ and $2\hat{\mathbf{n}}\, b \left(\frac{1}{4} - {k_0}^2 + \frac{2xk_0}{b} - \frac{x^2}{b^2}\right)|q_{3\mathbf{n}}|$ tend to 0 as $\mathbf{n} \to +\infty$. □

Define

$$\Lambda = \left(\frac{1}{2} + k_0 - \frac{x}{b}\right)^2 \frac{p_{k_0}(1-p_{k_0})}{b} + \left(\frac{1}{2} - k_0 + \frac{x}{b}\right)^2 \frac{p_{k_1}(1-p_{k_1})}{b} + 2\left(\frac{1}{4} - {k_0}^2 + \frac{2xk_0}{b} - \frac{x^2}{b^2}\right)\left(-\frac{p_{k_0}p_{k_1}}{b}\right). \tag{3.4}$$

Lemma 3.2. *Let x be a point of the interval $J_{k_0} = [k_0 b - b/2, k_0 b + b/2)$, then*

$$\lim_{\mathbf{n}\to+\infty}\left(\Lambda - \left[\frac{1}{2} + \left(2k_0 - \frac{x}{b}\right)^2\right] f(x)\right) = 0 \tag{3.5}$$

Proof. Taylor expansion give

$$p_{k_0} = bf(x) + \frac{1}{2}b[2(kb-x)-b]f^{'}(\xi_{k_0})$$

and

$$p_{k_1} = bf(x) + \frac{1}{2}b[2(kb-x)+b]f^{'}(\xi_{k_1}),$$

where $\xi_{k_0} \in J_{k_0}$ and $\xi_{k_1} \in J_{k_1}$.

Thus

$$\max\{0; f(x)b - Cb^2\} \le p_{k_i} \le f(x)b + Cb^2 \qquad i = 0,1 .$$

So we have the following majorizations (for $i = 0,1$)

$$\max\{0; f(x) - (C + f^2(x))b + C^2b^3\} \le \frac{p_{k_i}(1-p_{k_i})}{b} \le f(x) + (C - f^2(x))b + C^2b^3, \tag{3.6}$$

$$\max\{0; bf^2(x) - 2b^2Cf(x) + C^2b^3\} \le \frac{p_{k_0}p_{k_1}}{b} \le bf^2(x) + 2b^2Cf(x) - C^2b^3. \tag{3.7}$$

As $b \to 0$, using (3.6) and (3.7), we easily obtain the result. □

Théorème 3.1 *Assume Assumption 1 holds and $X_\mathbf{n}$ satisfies (1.1) and (1.2) or (1.3) with $\sum_{i=1}^{\infty} i^{N-1}(\varphi(i))^\beta < +\infty$ for some $0 < \beta < 1/2$. For x in $J_{k_0} = [k_0 b - b/2, k_0 b + b/2)$, we have*

$$\lim_{\mathbf{n}\to+\infty}\left(\hat{\mathbf{n}} b\, var f_\mathbf{n}(x) - \left[\frac{1}{2} + \left(2k_0 - \frac{x}{b}\right)^2\right] f(x)\right) = 0 \tag{3.8}$$

Proof. Using (2.3), lemmas 3.1 and 3.2, we obtain (3.8) □

4. Asymptotic normality of $f_{\mathbf{n}}$.

Théorème 4.1 *Assume Assumption 1 holds and* $X_{\mathbf{n}}$ *satisfies (1.1) and (1.2) or (1.3) with* $\varphi(u) = O(u^{-\theta})$ *with* $\theta > 2N$. *For some* $0 < \gamma < (\theta - N)\theta^{-1}$, *suppose the following conditions hold :*

(i) The bandwidth $b_{\mathbf{n}}$ *tends to zero in a manner such that* $\hat{\mathbf{n}} b_{\mathbf{n}}^{1+2(1-\gamma)N} \longrightarrow +\infty$ *as* $\mathbf{n} \to +\infty$.

(ii) There exists a sequence of positive integers $q = q_{\mathbf{n}} \to +\infty$ *with* $q = o(\hat{\mathbf{n}} b_{\mathbf{n}}^{1+2(1-\gamma)N})$ *such that* $b_{\mathbf{n}}^{-(1-\gamma)} q^{N-(1-\gamma)\theta} \longrightarrow 0$ *as* $\mathbf{n} \to +\infty$.

(iii) $b_{\mathbf{n}}$ *tends to zero in such a manner that* $\hat{\mathbf{n}}^{1+\tilde{k}} (\hat{\mathbf{n}} b_{\mathbf{n}})^{-1/2} q^{-\theta} \longrightarrow 0$ *as* $\mathbf{n} \to +\infty$.

For a point x *of the interval* $J_{k_0} = [k_0 b - b/2, k_0 b + b/2)$, *let*

$$\sigma^2(x) = \left[\frac{1}{2} + \left(2k_0 - \frac{x}{b}\right)^2\right] f(x).$$

Then, for x *such that* $f(x) > 0$, $(\hat{\mathbf{n}} b_{\mathbf{n}})^{1/2} [f_{\mathbf{n}}(x) - E f_{\mathbf{n}}(x)] \sigma^{-1}(x)$ *has a standard normal distribution as* $\mathbf{n} \to +\infty$.

Proof. It will done in three steps.

Step 1. For x in $J_{k_0} = [k_0 b - b/2, k_0 b + 1/2)$, define

$$Z_{\mathbf{i},k_0} = b_{\mathbf{n}}^{-1/2} \left[\left(\frac{1}{2} + k_0 - \frac{x}{b}\right) Y_{\mathbf{i},k_0} + \left(\frac{1}{2} - k_0 + \frac{x}{b}\right) Y_{\mathbf{i},k_1}\right]$$

and

$$S_{\mathbf{n}}(x) = \hat{\mathbf{n}}^{-1/2} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} Z_{\mathbf{i},k_0}\,.$$

It's easy to see that $|Z_{\mathbf{i},k_0}| \le b_{\mathbf{n}}^{-1/2}$ and $S_{\mathbf{n}}(x) = \hat{\mathbf{n}}^{1/2} b_{\mathbf{n}}^{1/2} f_{\mathbf{n}}(x)$.

For $\theta > 2N$, if $N\theta^{-1} < \beta < \frac{1}{2}$ and $\sum_{i=1}^{+\infty} i^{N-1} (\varphi(i))^{\beta} < +\infty$, using theorem 3.1, we have

$$\lim_{n \to +\infty} S_{\mathbf{n}}(x) = \sigma^2(x) = \left[\frac{1}{2} + \left(2k_0 - \frac{x}{b}\right)^2\right] f(x)\,. \tag{4.1}$$

By (i) and (ii), there exists a sequence of positive integers $(s_{\mathbf{n}})$ tending to infinity in a manner such that

$$s_{\mathbf{n}} q_{\mathbf{n}} = o((\hat{\mathbf{n}} b_{\mathbf{n}}^{1+2(1-\gamma)N})^{1/2N})\,. \tag{4.2}$$

Choose $p_{\mathbf{n}} = p = [(\hat{\mathbf{n}} b_{\mathbf{n}})^{1/2N} s_{\mathbf{n}}^{-1}]$. By (4.2), $qp^{-1} \le Cb^{1-\gamma}$ which tends to zero as $\mathbf{n} \to +\infty$. Thus $q < Cp$.

Multiplying $s_{\mathbf{n}}$ by a constant if necessary, it can be assumed without loss of generality that $q < p$.

Assume for some integers $r_1, \cdots, r_N$, we have $n_1 = r_1(p+q), \cdots, n_N = r_N(p+q)$, and note $L_{r_{\mathbf{n}}} = \{\mathbf{j} : 0 \le j_k \le r_k - 1 ; \forall k = 1, \cdots, N\}$ with $r_{\mathbf{n}} = \{r_1, \cdots, r_N\}$.

The r.v.'s $(Z_{\mathbf{i},k_0} - EZ_{\mathbf{i},k_0})$ are now set into large blocks and small blocks. Let

$$U(1,\mathbf{n},x,\mathbf{j}) = \hat{\mathbf{n}}^{-1/2} \sum_{\substack{i_k=j_k(p+q)+1 \\ k=1,\cdots,N}}^{j_k(p+q)+p} (Z_{\mathbf{i},k_0} - EZ_{\mathbf{i},k_0})$$

$$U(2,\mathbf{n},x,\mathbf{j}) = \hat{\mathbf{n}}^{-1/2} \sum_{\substack{i_k=j_k(p+q)+1 \\ k=1,\cdots,N-1}}^{j_k(p+q)+p} \sum_{i_N=j_N(p+q)+p+1}^{(j_N+1)(p+q)} (Z_{\mathbf{i},k_0} - EZ_{\mathbf{i},k_0})$$

$$U(3,\mathbf{n},x,\mathbf{j}) = \hat{\mathbf{n}}^{-1/2} \sum_{\substack{i_k=j_k(p+q)+1 \\ k=1,\cdots,N-2}}^{j_k(p+q)+p} \sum_{i_{N-1}=j_{N-1}(p+q)+p+1}^{(j_{N-1}+1)(p+q)} \sum_{i_N=j_N(p+q)+1}^{j_N(p+q)+p} (Z_{\mathbf{i},k_0} - EZ_{\mathbf{i},k_0})$$

$$U(4,\mathbf{n},x,\mathbf{j}) = \hat{\mathbf{n}}^{-1/2} \sum_{\substack{i_k=j_k(p+q)+1 \\ k=1,\cdots,N-2}}^{j_k(p+q)+p} \sum_{i_{N-1}=j_{N-1}(p+q)+p+1}^{(j_{N-1}+1)(p+q)} \sum_{i_N=j_N(p+q)+p+1}^{(j_N+1)(p+q)} (Z_{\mathbf{i},k_0} - EZ_{\mathbf{i},k_0})$$

and so on. Note that

$$U(2^{N-1},\mathbf{n},x,\mathbf{j}) = \hat{\mathbf{n}}^{-1/2} \sum_{\substack{i_k=j_k(p+q)+p+1 \\ k=1,\cdots,N-1}}^{(j_k+1)(p+q)} \sum_{i_N=j_N(p+q)+1}^{j_N(p+q)+p} (Z_{\mathbf{i},k_0} - EZ_{\mathbf{i},k_0}) .$$

Finally

$$U(2^N,\mathbf{n},x,\mathbf{j}) = \hat{\mathbf{n}}^{-1/2} \sum_{\substack{i_k=j_k(p+q)+p+1 \\ k=1,\cdots,N}}^{(j_k+1)(p+q)} (Z_{\mathbf{i},k_0} - EZ_{\mathbf{i},k_0}) .$$

For each integer $1 \le i \le 2^N$, define

$$T(\mathbf{n},x,i) = \sum_{\substack{j_k=0 \\ k=1,\cdots,N}}^{r_k-1} U(i,\mathbf{n},x,\mathbf{j}) .$$

Clearly

$$(4.3)\qquad S_{\mathbf{n}}(x) - E(S_{\mathbf{n}}(x)) = \sum_{i=1}^{2^N} T(\mathbf{n}, x, i)\,.$$

Note that $T(\mathbf{n}, x, 1)$ is the sum of the random variables $(Z_{\mathbf{i},k_0} - EZ_{\mathbf{i},k_0})$ in large blocks. The $T(\mathbf{n}, x, i),\ 2 \le i \le 2N$ are sums of random variables in small blocks.

If it is not the case that $n_1 = r_1(p+q), \cdots, n_N = r_N(p+q)$ for some integers $r_1, \cdots, r_N$, then a term, say, $T(\mathbf{n}, x, 2^N+1)$, containing all the $(Z_{\mathbf{i},k_0} - EZ_{\mathbf{i},k_0})$ at the ends not included in big or small blocks can be added. This term will not change the proof.

In this step, we have decomposed $S_{\mathbf{n}}(x) - E(S_{\mathbf{n}}(x))$ in two terms

$$\begin{aligned} S_{\mathbf{n}}(x) - E(S_{\mathbf{n}}(x)) &= T(\mathbf{n}, x, 1) + \sum_{i=2}^{2^N} T(\mathbf{n}, x, i) \\ &= A_{\mathbf{n}} + B_{\mathbf{n}}\,. \end{aligned}$$

Step 2. We will prove here that $\lim_{\mathbf{n}\to+\infty} B_{\mathbf{n}}^2 = 0$.

It(s sufficient to prove that, for $i = 2, \cdots, 2^N$, we have

$$\lim_{\mathbf{n}\to+\infty} E(T^2(\mathbf{n}, x, i)) = 0$$

For simplicity, we take $i = 2$. For each $\mathbf{j} \in L_{r_{\mathbf{n}}}$, define

$$\begin{aligned} \mathcal{I}(2, \mathbf{n}, x, \mathbf{j}) = \{\mathbf{i} \in \mathcal{I}_{\mathbf{n}} \,:\, & j_k(p+q) + 1 \le i_k \le j_k(p+q) + p\,,\ 1 \le k \le N-1\,, \\ & j_N(p+q) + p + 1 \le i_N \le (j_N + 1)(p+q)\}\,. \end{aligned}$$

Distinct sets of sites $\mathcal{I}(2, \mathbf{n}, x, \mathbf{j})$ for $\mathbf{j} \neq \mathbf{j}'$ are far apart by a distance of at least q.

We have

$$\begin{aligned} E(T^2(\mathbf{n}, x, 2)) &= \sum_{\mathbf{j}\in L_{r_{\mathbf{n}}}} var[U(2, \mathbf{n}, x, \mathbf{j})] + \sum_{\substack{\mathbf{j}\,,\mathbf{j}'\in L_{r_{\mathbf{n}}} \\ \mathbf{j}\neq\mathbf{j}'}} cov(U(2, \mathbf{n}, x, \mathbf{j}), U(2, \mathbf{n}, x, \mathbf{j}')) \\ &= \Lambda_1 + \Lambda_2\,. \end{aligned}$$

The r.v.'s $U(2, \mathbf{n}, x, \mathbf{j})$ for $j \in L_{r_{\mathbf{n}}}$ have the same law and

$$U(2, \mathbf{n}, x, \mathbf{j}) = \hat{\mathbf{n}}^{-1/2} \sum_{\mathbf{i}\in\mathcal{I}(2,\mathbf{n},x,\mathbf{0})} (Z_{\mathbf{i},k_0} - EZ_{\mathbf{i},k_0})\,.$$

Thus

$$varU(2,\mathbf{n},x,\mathbf{j}) = \hat{\mathbf{n}}^{-1}p^{N-1}q\, var(Z_{\mathbf{i},k_0}) + \hat{\mathbf{n}}^{-1} \sum_{\substack{\mathbf{i}\,,\,\mathbf{i}'\in\mathcal{I}(2,\mathbf{n},x,\mathbf{0})\\ \mathbf{i}\neq\mathbf{i}'}} cov(Z_{\mathbf{i},k_0}, Z_{\mathbf{i}',k_0})\,.$$

It's easy to see that

$$E(Z^2_{\mathbf{i},k_0}) = b^{-1}\left(\frac{1}{2}+k_0-\frac{x}{b}\right)^2 p_{k_0} + b^{-1}\left(\frac{1}{2}-k_0+\frac{x}{b}\right)^2 p_{k_1}\,.$$

Using Taylor expansion as in Lemma 3.2, we have

$$\begin{aligned} E(Z^2_{\mathbf{i},k_0}) &\le (f(x)+C_1 b)\left(\frac{1}{2}+k_0-\frac{x}{b}\right)^2 + (f(x)+C_2 b)\left(\frac{1}{2}-k_0+\frac{x}{b}\right)^2 \\ &\le C\,, \end{aligned}$$

because $b_{\mathbf{n}} \longrightarrow_0$ as $\mathbf{n}\to+\infty$, and f is bounded.

Similarly, we have for $\delta>0$

$$E|Z_{\mathbf{i},k}|^{2+\delta} \le C b^{-\delta/2}\,.$$

Let $\tilde{S}_1 = \{(\mathbf{i},\mathbf{i}')\in\mathcal{I}(2,\mathbf{n},x,\mathbf{0})^2\ :\ \|\mathbf{i}-\mathbf{i}'\|\ge 1\}\,.$

Then, using Lemma 2.2 with $s=r=2+\delta$

$$\begin{aligned} \sum_{\substack{\mathbf{i}\,,\,\mathbf{i}'\in\mathcal{I}(2,\mathbf{n},x,\mathbf{0})\\ \mathbf{i}\neq\mathbf{i}'}} |\,cov(Z_{\mathbf{i},k_0}, Z_{\mathbf{i}',k_0})| &\le \sum_{\mathbf{i}\,,\,\mathbf{i}'\in\tilde{S}_1} |\,cov(Z_{\mathbf{i},k_0}, Z_{\mathbf{i}',k_0})| \\ &\le Cp^{N-1}q\|Z_{\mathbf{i},k_0}\|^2_{2+\delta} \\ &\le Cp^{N-1}q\, b^{-\delta/(\delta+2)} \\ &\le Cp^{N-1}q\, b^{-(1+\gamma)}\,, \end{aligned}$$

avec $\delta = 2(1-\gamma)\gamma^{-1}$.

So

$$\begin{aligned} varU(2,\mathbf{n},x,\mathbf{j}) &\le C\hat{\mathbf{n}}^{-1}p^{N-1}q + Cp^{N-1}q\, b^{-(1-\gamma)} \\ &\le C\hat{\mathbf{n}}^{-1}p^{N-1}q\, b^{-(1-\gamma)}\,. \end{aligned}$$

Noting that $\hat{\mathbf{r}} = \prod_{k=1}^{N} r_k = \dfrac{\hat{\mathbf{n}}}{(p+q)^N}$ and $p = \left[\dfrac{(\hat{\mathbf{n}}b)^{1/2N}}{s_{\mathbf{n}}}\right]$, we have

$$
\begin{aligned}
\Lambda_1 &\leq C\hat{\mathbf{r}}p^{N-1}q\, b^{-(1-\gamma)} \\
&\leq Cp^{-1}q\, b^{-(1-\gamma)} \\
&\leq Cs_{\mathbf{n}}q(\hat{\mathbf{n}}b)^{-1/2N}\, b^{-(1-\gamma)} \\
&\leq Cs_{\mathbf{n}}q(\hat{\mathbf{n}}b^{1+2(1-\gamma)N})^{1/2N} .
\end{aligned}
$$

From (4.2), we immediatly have $\lim_{\mathbf{n}\to+\infty} \Lambda_1 = 0$.

Now, let $\tilde{S}_1 = \{(\mathbf{i}, \mathbf{i}') \in \mathcal{I}_{\mathbf{n}}^2 \,:\, \|\mathbf{i} - \mathbf{i}'\| \geq q\}$.

Notice that $d(\mathcal{I}(2, \mathbf{n}, x, \mathbf{j}), \mathcal{I}(2, \mathbf{n}, x, \mathbf{j}')) \geq q$ for $\mathbf{j} \neq \mathbf{j}'$.

Using Lemma 2.2, with $r = s = \delta + 2$, $\tau = (1-\gamma)^{-1}$, we have

$$
\begin{aligned}
\Lambda_2 &\leq \hat{\mathbf{n}}^{-1} \sum_{\substack{\mathbf{j}\,,\,\mathbf{j}'\in L_{r_{\mathbf{n}}}\\ \mathbf{j}\neq\mathbf{j}'}} \left(\sum_{\mathbf{i}\in\mathcal{I}(2,\mathbf{n},x,\mathbf{j})} \sum_{\mathbf{i}'\mathcal{I}(2,\mathbf{n},x,\mathbf{j}')} |\, cov(Z_{\mathbf{i},k_0}, Z_{\mathbf{i}',k_0})| \right) \\
&\leq C\,\hat{\mathbf{n}}^{-1} \sum_{\substack{\mathbf{i}\,,\,\mathbf{i}'\in L_{r_{\mathbf{n}}}\\ \|\mathbf{i}-\mathbf{i}'\|\geq q}} |\, cov(Z_{\mathbf{i},k_0}, Z_{\mathbf{i}',k_0})| \\
&\leq \hat{\mathbf{n}}^{-1} \sum_{(\mathbf{i},\mathbf{i}')\in\tilde{S}_q} |\, cov(Z_{\mathbf{i},k_0}, Z_{\mathbf{i}',k_0})| \\
&\leq C\,\|Z_{\mathbf{i},k_0}\|_{2+\delta}^2\; q^{N-(\theta/\tau)} \\
&\leq C\, b^{-\delta/(\delta+2)}\; q^{N-\theta(1-\gamma)} \\
&= C\, b^{1-\gamma}\; q^{N-\theta-(1-\theta)} .
\end{aligned}
$$

Thus (ii) implies that $\lim_{\mathbf{n}\to+\infty} \Lambda_2 = 0$.

Step 3. We will prove that $A_{\mathbf{n}} = T(\mathbf{n}, x, 1)$ is such that $\dfrac{A_{\mathbf{n}}}{\sigma(x)} = \dfrac{T(\mathbf{n}, x, 1)}{\sigma(x)}$ has a standard normal distribution as $\mathbf{n} \to +\infty$.

For each $\mathbf{j} \in L_{r_{\mathbf{n}}}$, define

$$
\mathcal{I}(1, \mathbf{n}, x, \mathbf{j}) = \{\mathbf{i} \in \mathcal{I}_{\mathbf{n}} \,:\, j_k(p+q) + 1 \leq i_k \leq j_k(p+q) + p\,,\; k = 1, \cdots, N\}\,.
$$

Then, for $\mathbf{j} \in L_{r_{\mathbf{n}}}$, $card(\mathcal{I}(1, \mathbf{n}, x, \mathbf{j})) = p^N$ and, if $\mathbf{j} \neq \mathbf{j}'$

$$
d(\mathcal{I}(1, \mathbf{n}, x, \mathbf{j}), \mathcal{I}(1, \mathbf{n}, x, \mathbf{j}')) > q\,.
$$

Enumerate the r.v.'s $U(1,\mathbf{n},x,\mathbf{j})$ in an arbitrary manner and refer to them as $V_1,\cdots,V_n$. Note that $M=\prod_{k=1}^{N} r_k=\hat{\mathbf{r}}=\hat{\mathbf{n}}(p+q)\le \hat{\mathbf{n}}\,p^{-N}$.

Now $\sup_{\mathbf{i}\in\mathcal{I}_{\mathbf{n}}}|Z_{\mathbf{i},k_0}|\le C\,b_{\mathbf{n}}^{-1/2}$ implies that $\sup_{1\le i\le\hat{\mathbf{r}}}|V_i|\le C\,\hat{\mathbf{n}}^{-1/2}\,p^N\,b_{\mathbf{n}}^{-1/2}$.

Using Lemma 1.3 in Carbon and al (97), there exists $\tilde{V}_1,\cdots,\tilde{V}_{\hat{\mathbf{r}}}$, independent random variables, independent of $V_1,\cdots,V_{\hat{\mathbf{r}}}$ with the same law verifying

$$\sum_{i=1}^{\hat{\mathbf{r}}} E|V_i-\tilde{V}_i|\le C\,\hat{\mathbf{r}}\,\hat{\mathbf{n}}^{-1/2}\,p^N\,b^{-1/2}h((\hat{\mathbf{r}}-1)p^N,p^N)\varphi(q)\,.$$

Now we have

$$var(A_{\mathbf{n}})=\hat{\mathbf{r}}\,var(V_1)+\sum_{i=1}^{\hat{\mathbf{r}}}\sum_{j=1}^{\hat{\mathbf{r}}}cov(V_i,V_j)\mathbf{1}_{\{i\neq j\}}\,.$$

Using similar arguments as in step 2, we have

$$\begin{aligned}\left|\sum_{i=1}^{\hat{\mathbf{r}}}\sum_{j=1}^{\hat{\mathbf{r}}}cov(V_i,V_j)\mathbf{1}_{\{i\neq j\}}\right| &\le \sum_{\substack{\mathbf{j}\,,\,\mathbf{j}'\in L_{r_{\mathbf{n}}}\\ \mathbf{j}\neq\mathbf{j}'}}\left|cov(U(1,\mathbf{n},x,\mathbf{j}),U(1,\mathbf{n},x,\mathbf{j}'))\right|\\ &\le \hat{\mathbf{n}}^{-1}\sum_{\substack{\mathbf{j}\,,\,\mathbf{j}'\in L_{r_{\mathbf{n}}}\\ \mathbf{j}\neq\mathbf{j}'}}\ \sum_{\mathbf{i}\in\mathcal{I}(1,\mathbf{n},x,\mathbf{j})}\ \sum_{\mathbf{i}'\mathcal{I}(1,\mathbf{n},x,\mathbf{j}')}|\,cov(Z_{\mathbf{i},k_0},Z_{\mathbf{i}',k_0})|\\ &\le C\,\hat{\mathbf{n}}^{-1}\sum_{\substack{\mathbf{i}\,,\,\mathbf{i}'\in\mathcal{I}_{\mathbf{n}}\\ \|\mathbf{i}-\mathbf{i}'\|\ge q}}|\,cov(Z_{\mathbf{i},k_0},Z_{\mathbf{i}',k_0})|\\ &\le C\,\hat{\mathbf{n}}^{-1}\sum_{(\mathbf{i},\mathbf{i}')\in\tilde{S}_q}|\,cov(Z_{\mathbf{i},k_0},Z_{\mathbf{i}',k_0})|\\ &\le C\,b^{-(1-\gamma)}\,q^{N-\theta(1-\gamma)}\,.\end{aligned}$$

From (ii), thus we have

$$\lim_{\mathbf{n}\to+\infty}\left|\sum_{i=1}^{\hat{\mathbf{r}}}\sum_{j=1}^{\hat{\mathbf{r}}}cov(V_i,V_j)\mathbf{1}_{\{i\neq j\}}\right|=0\,.$$

Then

$$\lim_{\mathbf{n}\to+\infty}(varA_{\mathbf{n}}-\hat{\mathbf{r}}var(V_1))=0\,.$$

So

$$\sigma^2=\lim_{\mathbf{n}\to+\infty}varS_{\mathbf{n}}=\lim_{\mathbf{n}\to+\infty}var(A_{\mathbf{n}}+B_{\mathbf{n}})=\lim_{\mathbf{n}\to+\infty}varB_{\mathbf{n}}=\lim_{\mathbf{n}\to+\infty}\hat{\mathbf{r}}\,var(V_1)\,,$$

which implies

$$\begin{aligned}\lim_{\mathbf{n}\to+\infty} var\sum_{i=1}^{\hat{\mathbf{r}}}\tilde{V}_i &= \lim_{\mathbf{n}\to+\infty}\sum_{i=1}^{\hat{\mathbf{r}}} var\tilde{V}_i = \lim_{\mathbf{n}\to+\infty}\hat{\mathbf{r}}\, var(\tilde{V}_1) \\ &= \lim_{\mathbf{n}\to+\infty}\hat{\mathbf{r}}\, var(V_1) = \sigma^2\end{aligned} .$$

We establish now that the r.v.'s $\tilde{V}_i$ satisfy the Lindeberg-Feller condition, that is, for each $\varepsilon > 0$,

$$\lim_{\mathbf{n}\to+\infty}\sum_{i=1}^{\hat{\mathbf{r}}} E\tilde{V}_i^2\,\mathbf{1}_{\{|\tilde{V}_i|>\varepsilon\sigma\}} = 0\,.$$

For each $i = 1,\cdots,\hat{\mathbf{r}}$, we have

$$|\tilde{V}_i| \le C\,\hat{\mathbf{n}}^{-1}\,p^N\,b_{\mathbf{n}}^{-1/2}\,.$$

So

$$\begin{aligned}\sum_{i=1}^{\hat{\mathbf{r}}} E\tilde{V}_i^2\,\mathbf{1}_{\{|\tilde{V}_i|>\varepsilon\sigma\}} &\le C\,\hat{\mathbf{n}}^{-1}\,p^N\,b_{\mathbf{n}}^{-1}\,p^{2N}\sum_{i=1}^{\hat{\mathbf{r}}} P(|\tilde{V}_i| > \varepsilon\sigma\ ) \\ &\le C\,s_{\mathbf{n}}^{-2N}\sum_{i=1}^{\hat{\mathbf{r}}} P(|\tilde{V}_i| > \varepsilon\sigma\ )\,,\end{aligned}$$

because $p = \left[\dfrac{(\hat{\mathbf{n}}b)^{1/2N}}{s_{\mathbf{n}}}\right]$.
We also have, for each $i = 1,\cdots,\hat{\mathbf{r}}$

$$\sigma^{-1}|\tilde{V}_i| \le C\,\hat{\mathbf{n}}^{-1/2}\,b_{\mathbf{n}}^{-1/2}\,p^N \le C\,s_{\mathbf{n}}^{-N}\,.$$

Then $P(|\tilde{V}_i| > \varepsilon\sigma\ ) = 0$ for $\min_{1\le j\le N} n_j$ sufficiently large, that is possible because $\lim_{\mathbf{n}\to+\infty} s_{\mathbf{n}} = +\infty$.

We obtain that $\dfrac{\sum_{i=1}^{\hat{\mathbf{r}}}\tilde{V}_i}{\sigma(x)}$ has a standard normal distribution as $\mathbf{n}\to+\infty$.
We just now have to prove that

$$\lim_{\mathbf{n}\to+\infty} P\biggl(\sum_{i=1}^{\hat{\mathbf{r}}} P(|V_i - \tilde{V}_i| > \varepsilon\sigma\ )\biggr) = 0\,.$$

Using Markov inequality, we have

$$P\Bigg(\sum_{i=1}^{\hat{\mathbf{r}}} P(|V_i - \tilde{V}_i| > \varepsilon\sigma\ )\Bigg) \leq C\sum_{i=1}^{\hat{\mathbf{r}}} E|V_i - \tilde{V}_i|$$

$$\leq C\,\hat{\mathbf{r}}\,\hat{\mathbf{n}}^{-1/2}\,p^N\,b_{\mathbf{n}}^{-1/2}\,h((\hat{\mathbf{r}}-1)p^N, p^N)\varphi(q)$$

$$\leq C\,\hat{\mathbf{r}}^{1+\tilde{k}}\,\hat{\mathbf{n}}^{-1/2}\,p^{N(1+\tilde{k})}\,b_{\mathbf{n}}^{-1/2}\,q^{-\theta}$$

$$\leq C\,\hat{\mathbf{n}}^{1+\tilde{k}}\left(\frac{p}{p+q}\right)^{N(1+\tilde{k})}\hat{\mathbf{n}}^{-1/2}\,b_{\mathbf{n}}^{-1/2}\,q^{-\theta}$$

$$\leq C\,\hat{\mathbf{n}}^{1+\tilde{k}}\,(\hat{\mathbf{n}}b_{\mathbf{n}})^{-1/2}\,q^{-\theta}\,,$$

and using (iii) the last term tends to zero. So

$$\frac{\sum_{i=1}^{\hat{\mathbf{r}}}\tilde{V}_i}{\sigma(x)} = o_P(1) + \frac{\sum_{i=1}^{\hat{\mathbf{r}}}V_i}{\sigma(x)}\,,$$

where $o_P(1)$ is a r.v. going to zero in probability. Finally, we obtain that

$$\frac{A_{\mathbf{n}}}{\sigma(x)} = \frac{T(\mathbf{n},x,1)}{\sigma(x)} = \frac{\sum_{i=1}^{\hat{\mathbf{r}}}V_i}{\sigma(x)}$$

has a standard normal distribution as $\mathbf{n}\to+\infty$, and

$$\frac{S_{\mathbf{n}}(x) - ES_{\mathbf{n}}(x)}{\sqrt{\hat{\mathbf{n}}}} = (\hat{\mathbf{n}}b_{\mathbf{n}})^{1/2}\,(f_{\mathbf{n}}(x) - Ef_{\mathbf{n}}(x))$$

has a normal distribution $\mathcal{N}(0,\sigma^2(x))$ as $\mathbf{n}\to+\infty$. □

Corollary 4.1 *Assume the conditions of the theorem 4.1 are satisfied. Suppose that f has bounded seconde derivative. If $\hat{\mathbf{n}}b_{\mathbf{n}}^5\to 0$, then, for x such that $f(x)>0$,*

$$(\hat{\mathbf{n}}b_{\mathbf{n}})^{1/2}[f_{\mathbf{n}}(x) - f(x)]\sigma^{-1}(x)$$

has a standard normal distribution as $\mathbf{n}\to+\infty$.

Proof. Using a Taylor expansion, there exists $\xi_0\in J_{k_0}$ such that

$$p_{k_0} = \int_{(k_0-1)b}^{k_0 b} f(y)dy = bf(k_0 b) - \frac{1}{2}f'(k_0 b) + \frac{1}{6}b^3 f''(\xi_0)\,,$$

and there exists $\xi_1 \in J_{k_1}$ such that

$$p_{k_1} = \int_{k_0 b}^{(k_0+1)b} f(y)dy = bf(k_0 b) + \frac{1}{2} f'(k_0 b) - \frac{1}{6} b^3 f''(\xi_1) \, .$$

For $x \in J_{k_0}$, we also have

$$f(x) = f(k_0 b) + x f'(k_0 b) + \frac{1}{2} x^2 f''(\xi_x) \quad \text{with} \quad \xi_x \in J_{k_0} \, .$$

Then, using the two preceding equations, the bias is

$$Ef_{\mathbf{n}}(x) - f(x) = \left(\frac{1}{2} + k_0 - \frac{x}{b}\right) \frac{p_{k_0}}{b} + \left(\frac{1}{2} - k_0 + \frac{x}{b}\right) \frac{p_{k_1}}{b} - f(x)$$
$$= \left(\frac{1}{2} + k_0 - \frac{x}{b}\right) \frac{b^2}{6} f''(\xi_0) + \left(\frac{1}{2} - k_0 + \frac{x}{b}\right) \frac{b^2}{6} f''(\xi_1) - \frac{(x - k_0 b)^2}{2} f''(\xi_x) \, .$$

So we have

$$|Ef_{\mathbf{n}}(x) - f(x)| \leq C \, b_{\mathbf{n}}^2 \, ,$$

and

$$\left| \, (\hat{\mathbf{n}} b_{\mathbf{n}})^{1/2} [Ef_{\mathbf{n}}(x) - f(x)] \, \right| \leq C \, (\hat{\mathbf{n}} b_{\mathbf{n}}^5)^{1/2} \, .$$

We conclude using the theorem 4.1. □

Remark 4.1. The condition $\hat{\mathbf{n}} b_{\mathbf{n}}^5 \to 0$ is approximately the optimal bin width we have found for the mean quadratic convergence (see Carbon (2006) for example).

With the theorem 4.1 or the corollary 4.1, we can't directly obtain confidence interval for the density $f(x)$, because $\sigma(x)$ is depending on $f(x)$. Obviously, we can estimate $\sigma^2(x)$ by $\hat{\sigma}^2(x) = \left[\frac{1}{2} + \left(2k_0 - \frac{x}{b}\right)^2\right] f(x)$.

Let ε be an arbitrary small positive number and denote $g(\mathbf{n}) = \prod_{i=1}^N (\log n_i)(\log \log n_i)^{1+\varepsilon}$.
Clearly, $\sum \frac{1}{\hat{\mathbf{n}} g(\mathbf{n})} < \infty$, where the summation is over all $\mathbf{n}$ in Z^N.
Define

$$\theta_1^* = \frac{\rho + 3N}{\rho - 5N} \; , \quad \theta_2^* = \frac{3N - \rho}{\rho - 5N}$$

$$\theta_3^* = \frac{\rho + 3N}{\rho - (2\tilde{k} + 3)N} \; , \quad \theta_4^* = \frac{N - \rho}{\rho - (2\tilde{k} + 3)N} .$$

Théorème 4.2 *Assume the conditions of the theorem 4.1 are satisfied. Suppose that f has bounded seconde derivative.*

(i) If (1.2) is satisfied, $\rho > 3N$ and

$$\hat{\mathbf{n}} b_{\mathbf{n}}^{\theta_1^*} (\log \hat{\mathbf{n}})^{\theta_2^*} (g(\mathbf{n}))^{-2N/(\rho - 5N)} \to \infty, \tag{4.4}$$

(ii) or if (1.3) is satisfied and

$$\hat{\mathbf{n}} b_{\mathbf{n}}^{\theta_3^*} (\log \hat{\mathbf{n}})^{\theta_4^*} (g(\mathbf{n}))^{-2N/(\rho - (2\tilde{k}+3)N} \to \infty. \tag{4.5}$$

If $\hat{\mathbf{n}} b_{\mathbf{n}}^5 \to 0$, then, for x such that $f(x) > 0$,

$$(\hat{\mathbf{n}} b_{\mathbf{n}})^{1/2} [f_{\mathbf{n}}(x) - f(x)] \hat{\sigma}^{-1}(x)$$

has a standard normal distribution as $\mathbf{n} \to +\infty$.

Proof.

Using theorem 4.1, it is sufficient to prove that for all $x \in R$,

$$\hat{\sigma}^2(x) \longrightarrow \sigma^2(x) \qquad a.s. \ ,$$

that is

$$f_{\mathbf{n}}(x) \longrightarrow f(x) \qquad a.s. \ ,$$

which is a direct consequence of theorem 6.1 of Carbon and al. (2008).

Thus, we obtain a confidence interval at a $(1 - \alpha)$ rate for all $x \in R$:

$$f(x) \in \left[f_{\mathbf{n}}(x) - \hat{\sigma}(x) (\hat{\mathbf{n}} b_{\mathbf{n}})^{-1/2} U_{\alpha/2} \,; \, f_{\mathbf{n}}(x) + \hat{\sigma}(x) (\hat{\mathbf{n}} b_{\mathbf{n}})^{-1/2} U_{\alpha/2} \right] \ ,$$

where $U_{\alpha/2}$ is the $(1 - \alpha/2)$ quantile of the standard normal distribution.

REFERENCES

Biau G.(2003). Spatial kernel density estimation, Mathematical Methods of Statistics, Vol. 12, pp. 371–390.

Bolthausen, E. (1982). On the central limit theorem for stationary random fields. Ann. Probab. 10, 1047-1050.

Bradley, R.C. (1986). Basic properties of strong mixing conditions. In : Dependence in Probability and Statistics. vol 11, Birkhuser, 165-192.

Carbon, M., Tran, L.T. and Wu, B. (1997). Kernel density estimation for random fields (Density estimation for random fields). Statistics and Probability Letters 36, 115-125.

Carbon, M., Hallin, M. and Tran, L.T. (1996). Kernel density estimation for random fields: The L_1 theory. Journal of Non Parametric Statistics 6, 157-170.

Carbon, M., Francq, C. and Tran, L.T. (2007) Kernel Regression Estimation for Random Fields. Journal of Statistical Planning and Inference, 137 (issue3), 778-798

Carbon, M., Francq, C. and Tran, L.T. (2008) Frequency polygon for random fields (submitted)

Carbon, M. (2006) Polygone des fréquences pour des champs aléatoires, Note au Comptes Rendus de l'Académie des Sciences, Série I, 342, 693-696.

Davydov, Yu A. (1970). The invariant principle for stationary processes Theor. Probab. Appl. 14, 487-498.

Francq, C. and Tran, L.T. (2002). Nonparametric Estimation of Density, Regression And Dependence Coefficients. Nonparametric Statistics 14, 729-747.

Gorodetskii, V.V. (1977). On the strong mixing property for linear sequences. Theory Probability Appl. 22, 411-413.

Guyon, X. and Richardson, S. (1984). Vitesse de convergence du théorème de la limite centrale pour des champs faiblement dépendants. Z. Wahrsch. Verw. Gebiete 66, 297-314.

Guyon, X. (1987). Estimation d'un champ par pseudo-vraisemblance conditionnelle: Etude asymptotique et application au cas Markovien. Proc. 6^{th} Franco-Belgian Meeting of Statisticians.

Hall, P. and Hannan, E.J. (1988). On stochastic complexity and nonparametric density estimation. Biometrika 75, 705-714.

Ibragimov, I. A. and Linnik, Yu. V. (1971). Independent and stationary sequences of random variables. Wolters-Noordhoff, Groningen.

Masry, E. and Györfi (1987). Strong consistency and rates for recursive density estimators for stationary mixing processes. J. Multivariate Anal. 22 (1987), 79-93.

Nahapetian, B. S. (1980). The central limit theorem for random fields with mixing conditions. Adv. in Probability 6, Multicomponent systems. ed. R. L. Dobrushin & Ya G. Sinai, 531-548.

Nahapetian, B. S. (1987). An approach to proving limit theorems for dependent random variables. Theory Prob. Appl. 32, 535-539.

Neaderhouser, C. C. (1980). Convergence of block spins defined on random fields. J. Statist. Phys. 22, 673-684.

Politis, D. N. and Romano, J. P. (1993). Nonparametric resampling for homogeneous strong mixing random fields. J. Multivariate Anal. 47, 301-328.

Rio, E. (1995). The functional law of the iterated logarithm for stationary strongly mixing sequences. Ann. Prob. 23, 1188-1203.

Robinson, P.M. (1983). Nonparametric estimators for time series. J. Time Series Anal. 4, 185-207.

Rosenblatt, M. (1985). Stationary sequences and random fields. Birkhauser, Boston.

Roussas, G. G. (1969). Nonparametric estimation of the transition distribution of a Markov process. Ann. Inst. Statist. Math. 21, 73-87.

Roussas, G. G. (1988). Nonparametric estimation in mixing sequences of random variables. Jour. Statist. Plann. Inference 18, 135-149.

Scott, D.W. (1985). Frequency polygons, theory and applications. J. Amer. Stat. Assoc., 80, p 348-354.

Scott, D.W. (1992). Multivariate density estimation : Theory, Practice and Vizualisation, J. Wiley, New-York.

Stone, C.J. (1983). Optimal unifom rate of convergence for non parametric estimators of a density function and its derivative. Recent advances in Statistics: Papers in Honor of H. Chernoff, ed. by M.H. Revzi, J.S. Rustagi, and D. Siegmund, 393-406.

Takahata, H. (1983). On the rates in the central limit theorem for weakly dependent random fields. Z. Wahrsch. Verw. Gebiete 62, 477-480.

Tran, L. T. (1990). Kernel density estimation on random fields. J. Multivariate Anal. 34, 37-53.

Tran, L. T. and Yakowitz, S. (1993). Nearest neighbor estimators for random fields. J. Multivariate Anal. 44, 23-46.

Younso, A. (2007). Estimation et classification non paramétriques par la méthode du noyau sur les champs aléatoires mélangeants. Thèse de l'Université de Montpellier.