Revista de Matemática: Teoría y Aplicaciones 2007 **14**(2) : 105–122
cimpa – ucr issn: 1409-2433

Frequency polygons for random fields (density estimation for random fields)

Michel Carbon*

Recibido/Received: 22 Feb 2006; Aceptado/Accepted: 22 Jan 2007

Abstract

The purpose of this paper is to investigate the frequency polygon as a density estimator for stationary random fields indexed by multidimensional lattice points space. Optimal bin widths which asymptotically minimize integrated errors (IMSE) are derived. Under weak conditions, frequency polygons achieve the same rate of convergence to zero of the IMSE as kernel estimators. They can also attain the optimal uniform rate of convergence under general conditions. Frequency polygons thus appear to be very good density estimators with respect to both criteria of IMSE and uniform convergence.

Keywords: random field, frequency polygons, bandwidth.

Resumen

El propósito de este artículo es el de investigar el polígono de frecuencias como estimador de densidad para campos aleatorios indexados por un espacio de puntos en un retículo multidimensional. Se deriva la anchura óptima del compartimiento que asintóticamente minimiza los errores integrados (IMSE, por sus siglas en inglés). Bajo condiciones débiles, los polígonos de frecuencia alcanzan la misma tasa de convergencia hacia cero del IMSE como estimadores de núcleo. También pueden alcanzar la tasa óptima de convergencia uniforme bajo condiciones generales. Luego, los polígonos de frecuencia parecen ser entonces muy buenos estimadores de densidad con respecto a ambos criterios, de IMSE y convergencia uniforme.

Palabras clave: campos aleatorios, polígono de frecuencias, ancho de banda.

Mathematics Subject Classification: primary: 62G05; secondary: 62M40, 62M30.

*Université de Rennes 2, France; and Directeur du Département de Statistique, E.N.S.A.I., Rue Blaise Pascal, 35172 Bruz, France. Tel: +(33) 2.99.05.33.16. E-Mail: carbon@ensai.fr

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 $\mathbf{i}$, 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 $\mathrm{dist}\,(S, S')$. 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 $S, S' \subset Z^N$,

$$\alpha(\mathcal{B}(S), \mathcal{B}(S')) = \sup\{|P(AB) - P(A)P(B)|, A \in \mathcal{B}(S), B \in \mathcal{B}(S')\} \tag{1}$$

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

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

$$h(n, m) \leq \min\{m, n\} \tag{2}$$

or

$$h(n, m) \leq C(n + m + 1)^{\tilde{k}} \tag{3}$$

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).

Denote by $I_{\mathbf{n}}$ a rectangular region defined by

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

Assume that we observe $\{X_{\mathbf{n}}\}$ on $I_{\mathbf{n}}$.

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

$$\min\{n_k\} \to \infty \quad \text{and} \quad |n_j/n_k| < C \tag{4}$$

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

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

Under weak dependence conditions, frequency polygons are shown to achieve the rate of convergence to zero of order $\hat{\mathbf{n}}^{-4/5}$ with respect to the criterion of IMSE. In the case $N = 1$, histograms can only achieve the slower rate of convergence of the IMSE of order $n^{-2/3}$. It is also established that frequency polygons attain the uniform rate of convergence $(n^{-1} \log n)^{1/3}$ under appropriate smoothness conditions. This is the optimal rate of convergence for nonparametric estimators of a density function in the i.i.d. case (see Stone (1983)). We here obtain similar results for random fields. Frequency polygons thus appear to be very good density estimators with respect to both criteria of IMSE and uniform convergence. For background material on frequency polygons, see Scott (1985).

Our paper is organized as follows: Section 2 provides some preliminaries and background material. The optimal choice of the bin width which asymptotically minimizes the integrated mean square error is derived in Section 3. Theorem 3.1 generalizes results of Scott (1985), and Carbon, Garel and Tran (1996). In Section 4, the asymptotic variance of the frequency polygon $f_{\mathbf{n}}$ is obtained. Weak conditions for the uniform convergence of $f_{\mathbf{n}}$ on R are obtained in Section 5. Finally, sharp rates of uniform convergence are established in Section 6.

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 set in R.

Denote $\Psi_{\mathbf{n}} = \max(b; (\log \hat{\mathbf{n}}(\hat{\mathbf{n}}b)^{-1})^{1/2})$.

2 Preliminaries

Consider a partition $\cdots < x_{-2} < x_{-1} < x_0 < x_1 < x_2 < \cdots$ 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. Without loss of generality, we assume that there is a mesh node at zero. Consider the two adjacent histogram bins $I_0 = [-b, 0)$ and $I_1 = [0, b)$. Denote the number of observations falling in these intervals respectively by ν_0 and ν_1. The values of the histogram in these previous bins are given by $f_0 = \nu_0 \hat{\mathbf{n}}^{-1} b^{-1}$ and $f_1 = \nu_1 \hat{\mathbf{n}}^{-1} b^{-1}$. The frequency polygon $f_{\mathbf{n}}(x)$ is given by

$$f_{\mathbf{n}}(x) = \left(\frac{1}{2} - \frac{x}{b}\right) f_0 + \left(\frac{1}{2} + \frac{x}{b}\right) f_1, \quad \text{for} \quad -\frac{b}{2} \leq x < \frac{b}{2}. \tag{5}$$

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

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

Then,

$$\nu_0 = \sum_{\mathbf{i} \in \mathbf{I_n}} Y_{\mathbf{i},0}(x) \quad \text{and} \quad \nu_1 = \sum_{\mathbf{i} \in \mathbf{I_n}} Y_{\mathbf{i},1}(x).$$

Let $U = u(X_{\mathbf{i}})$ and $V = v(X_{\mathbf{j}})$, where u and v be are real-valued measurable functions.

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

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

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

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

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 $\eta_{\mathbf{i},k}(x) = Y_{\mathbf{i},k}(x) - EY_{\mathbf{i},k}(x)$.

Corollary 1 *For each integer k, there exists some $\xi_k \in I_k$ such that*

(i) $|cov(\eta_{\mathbf{i},k}(x), \eta_{\mathbf{j},k}(x)| \leq C(f(\xi_k)b)^{1/2}(\varphi(\|\mathbf{i} - \mathbf{j}\|))^{1/2}$,

(ii) $|cov(\eta_{\mathbf{i},k-1}(x), \eta_{\mathbf{j},k}(x)| \leq C(f(\xi_k)b)^{1/2}(\varphi(\|\mathbf{i} - \mathbf{j}\|))^{1/2}$

Proof. (i) Taking $r = 2, s = \infty$ and $h = 2$, Lemma 2.2 leads to the following result: if $E|U|^2 < +\infty$ and $P(|V| > 1) = 0$, then

$$|EUV - EUEV| \leq C\|U\|_2\{\varphi(\|\mathbf{j} - \mathbf{i}\|)\}^{1/2}.$$

Taking $U = Y_{\mathbf{i},k}(x)$ and $V = Y_{\mathbf{j},k}(x)$, then

$$\|U\|_2 = (EY_{\mathbf{i},k})^{1/2} = (P[X_{\mathbf{1}} \in I_k])^{1/2} = (f(\xi_k)b)^{1/2}, \text{ where } \xi_k \in I_k.$$

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,

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

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 $|\theta| < 1$. Asssume 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

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

Lemma 2.3 *If Assumption 1 is satisfied, then*

$$\int\int_{I_k \times I_k} |f_{\mathbf{i},\mathbf{j}}(x,y) - f(x)f(y)| \; dx\,dy \leq Mf(\zeta_k)b^2 \quad \textit{with} \quad \zeta_k \in I_k. \tag{6}$$

Proof. Since f is uniformly continuous and integrable,

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

By Assumption 1,

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

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

3 Integrated mean squared error and optimal bin width

For convenience, we define the roughness of the k-th derivative of f by

$$R_k(f) = \int_{-\infty}^{+\infty} \left[f^{(k)}(x)\right]^2 \, dx,$$

and

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

As usual, the IMSE is defined as the sum of two terms: the integrated pointwise squared bias, and the integrated pointwise variance. Define

$$q_{1\mathbf{n}} = \frac{1}{\hat{\mathbf{n}}^2 b^2} \sum_{\mathbf{i} \neq \mathbf{j}} \operatorname{cov}(\eta_{\mathbf{i},0}, \eta_{\mathbf{j},0}),$$

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

$$q_{3\mathbf{n}} = \frac{1}{\hat{\mathbf{n}}^2 b^2} \sum_{\mathbf{i} \neq \mathbf{j}} \operatorname{cov}(\eta_{\mathbf{i},0}, \eta_{j\mathbf{j},1}).$$

Lemma 3.1 *The variance of the frequency polygon $f_{\mathbf{n}}(x)$ defined in (5) is given by*

$$\begin{aligned} var f_{\mathbf{n}}(x) &= \left(\frac{1}{2} - \frac{x}{b}\right)^2 \Big[\frac{1}{\hat{\mathbf{n}} b^2} p_0(1-p_0) + q_{1\mathbf{n}}\Big] \\ &+ \left(\frac{1}{2} + \frac{x}{b}\right)^2 \Big[\frac{1}{\hat{\mathbf{n}} b^2} p_1(1-p_1) + q_{2\mathbf{in}}\Big] + 2\left(\frac{1}{4} - \frac{x^2}{b^2}\right) \Big[-\frac{p_0 p_1}{\hat{\mathbf{n}} b^2} + q_{3\mathbf{n}}\Big]. \end{aligned} \tag{7}$$

Proof. From the expression of the frequency polygon (5),

$$\operatorname{var} f_{\mathbf{n}}(x) = \left(\frac{1}{2} - \frac{x}{b}\right)^2 \operatorname{var} f_0 + \left(\frac{1}{2} + \frac{x}{b}\right)^2 \operatorname{var} f_1 + 2\left(\frac{1}{4} - \frac{x^2}{b^2}\right) \operatorname{cov}(f_0, f_1).$$

Clearly,

$$\operatorname{var} f_0 = \frac{1}{\hat{\mathbf{n}}^2 b^2} \operatorname{var} \left(\sum_{\mathbf{i} \in \mathbf{I_n}} Y_{\mathbf{i},0} \right)$$

with

$$\begin{aligned} \operatorname{var}\Big(\sum_{\mathbf{i} \in \mathbf{I_n}} Y_{\mathbf{i},0}\Big) &= \sum_{\mathbf{i} \in \mathbf{I_n}} \operatorname{var}\Big(Y_{\mathbf{i},0}\Big) + \sum_{\mathbf{i} \neq \mathbf{j}} \operatorname{cov}(Y_{\mathbf{i},0}, Y_{\mathbf{j},0}) \\ &= n p_0 (1-p_0) + \sum_{\mathbf{i} \neq \mathbf{j}} \operatorname{cov}(\eta_{\mathbf{i},0}, \eta_{\mathbf{j},0}). \end{aligned}$$

Similarly,

$$\operatorname{var} f_1 = \frac{1}{\hat{\mathbf{n}}^2 b^2} \operatorname{var} \left(\sum_{\mathbf{i} \in \mathbf{I_n}} Y_{\mathbf{i},1} \right) = n p_1 (1-p_1) + \sum_{\mathbf{i} \neq \mathbf{j}} \operatorname{cov}(\eta_{\mathbf{i},1}, \eta_{\mathbf{j},1}).$$

We get also

$$\operatorname{cov}(f_0,\ f_1) = \frac{1}{\hat{\mathbf{n}}^2 b^2} \operatorname{cov}\Big(\sum_{\mathbf{i} \in \mathbf{I_n}} Y_{\mathbf{i},0},\ \sum_{\mathbf{j} \in \mathbf{I_n}} Y_{\mathbf{j},1}\Big) = \frac{1}{\hat{\mathbf{n}}^2 b^2} \sum_{\mathbf{i} \in \mathbf{I_n}} \operatorname{cov}\left(Y_{\mathbf{i},0},\ Y_{\mathbf{i},1}\right) + q_{3\mathbf{n}}.$$

But,

$$\operatorname{cov}\left(Y_{\mathbf{i},0},\ Y_{\mathbf{i},1}\right) = E\left(Y_{\mathbf{i},0} Y_{\mathbf{i},1}\right) - E\left(Y_{\mathbf{i},0}\right) E\left(Y_{\mathbf{i},1}\right) = -p_0 p_1.$$

Summing up, we get (8). □

Define

$$w_k = \max\Big((f(\xi_k))^{1/2}, Mf(\zeta_k)\Big). \tag{8}$$

Lemma 3.2 *Assume that* $\varphi(k) = O(k^{-\rho})$ *for some* $\rho > 2N + (3/2)$. *Let* $0 < \varepsilon \le (2N-1)(8N-1)^{-1}$. *Then*

$$q_{1\mathbf{n}} + q_{2\mathbf{n}} + q_{3\mathbf{n}} \le C\hat{\mathbf{n}}^{-1}b^{-1+\varepsilon}w_0.$$

Proof. By Corollary 1 and Lemma 2.3

$$|\mathrm{cov}(\eta_{\mathbf{i},0}, \eta_{\mathbf{j},0})| \le \min\{C(f(\xi_k))^{1/2}b^{1/2}(\varphi(\|\mathbf{i}-\mathbf{j}\|))^{1/2}, Mf(\zeta_0)b^2\}.$$

Define

$$\begin{aligned} S_1 &= \{\mathbf{i},\mathbf{j} \in I_\mathbf{n} \quad | \quad 0 < \|\mathbf{i}-\mathbf{j}\| \le K_\mathbf{n}\} \\ S_2 &= \{\mathbf{i},\mathbf{j} \in I_\mathbf{n} \quad | \quad \|\mathbf{i}-\mathbf{j}\| > K_\mathbf{n}\} \end{aligned}$$

Split $\sum_{\mathbf{i}\neq\mathbf{j}} \Big| \mathrm{cov}(\eta_{\mathbf{i},0}, \eta_{\mathbf{j},0}) \Big|$ into two separate summations J_1 and J_2 over sites S_1 and S_2. Then

$$\sum_{\mathbf{i}\neq\mathbf{j}} \Big| \mathrm{cov}(\eta_{\mathbf{i},0}, \eta_{\mathbf{j},0}) \Big| \le J_1 + J_2.$$

Now, we have the following majorizations

$$J_1 = \sum_{\mathbf{i},\mathbf{j}\in S_1} \Big| \mathrm{cov}(\eta_{\mathbf{i},0}, \eta_{\mathbf{j},0}) \Big| \le Mf(\zeta_0) \sum_{\mathbf{i},\mathbf{j}\in S_1} b^2 \le Mf(\zeta_0)K_\mathbf{n}^N$$

Let $K_\mathbf{n}^N = b^{(1-\varepsilon)/N}$. Thus

$$J_1 \le Mf(\zeta_0)\,\hat{\mathbf{n}}\; b^{1+\varepsilon} \tag{9}$$

Let now $\nu = \dfrac{N}{2} \cdot \dfrac{1+2\varepsilon}{1-\varepsilon}$. Clearly

$$\begin{aligned} J_2 &= \sum_{\mathbf{i},\mathbf{j}\in S_2} \Big| \mathrm{cov}(\eta_{\mathbf{i},0}, \eta_{\mathbf{j},0}) \Big| \\ &\le C\,(f(\xi_0))^{1/2}\, b^{1/2} \sum_{\mathbf{i},\mathbf{j}\in S_2} (\varphi(\|\mathbf{i}-\mathbf{j}\|))^{1/2} \\ &\le C\,(f(\xi_0))^{1/2}\, b^{1/2}\, \hat{\mathbf{n}} \sum_{\|\mathbf{i}\|>K_\mathbf{n}} (\varphi(\|\mathbf{i}\|))^{1/2} \\ &\le C\,(f(\xi_0))^{1/2}\, b^{1/2}\, \hat{\mathbf{n}}\, K_\mathbf{n}^{-\nu} \sum_{\|\mathbf{i}\|>K_\mathbf{n}} \|\mathbf{i}\|^\nu (\varphi(\|\mathbf{i}\|))^{1/2}. \end{aligned}$$

Since $0 < \varepsilon \le (2N-1)(8N-1)^{-1}$, we have $\nu \le N - \dfrac{1}{4}$. Thus

$$\sum_{\|\mathbf{i}\|>K_\mathbf{n}} \|\mathbf{i}\|^\nu (\varphi(\|\mathbf{i}\|))^{1/2} \le \sum_{i=1}^{\infty} i^{-(2\rho-4N+1)/4} < +\infty$$

since $\rho > 2N + 3/2$. Finally

$$J_2 \leq C\,(f(\xi_0))^{1/2}\, b^{1/2}\, \hat{\mathbf{n}}\, K_{\mathbf{n}}^{-\nu} \leq C\,(f(\xi_0))^{1/2}\, b^{1+\varepsilon}$$

and

$$\sum_{\mathbf{i}\neq\mathbf{j}} \Big|\operatorname{cov}(\eta_{\mathbf{i},0}, \eta_{\mathbf{j},0})\Big| \leq C\, w_0\, \hat{\mathbf{n}}\, b^{1+\varepsilon}$$

Thus $q_{1\mathbf{n}}$ and similarly $q_{2\mathbf{n}}$ are bounded by $C\, w_0\, \hat{\mathbf{n}}^{-1}\, b^{-1+\epsilon}$. Using Corollary 1 (ii) and a slight variation of (6) it can be shown that $q_{3\mathbf{n}}$ is bounded by the same quantity. $\square$

Assumption 2. f is twice continuously differentiable; f'' is absolutely continuous with respect to the Lebesgue measure on R; $f^{1/2} \in L^1$, $f'f^{-1/2} \in L^1$, $f^{(k)} \in L^2$ for $k = 0, 1, 2, 3$. The following Lemma (see Scott (1985, p.350)) will be needed in the sequel

Lemma 3.3 *Suppose that ϕ is absolutely continuous on $(-\infty, \infty)$ with almost everywhere derivative ϕ' and that ϕ, $\phi' \in L_1$. Let c_k be an arbitrary point in bin I_k. Then the following sum exists and may be approximated by an integral*

$$\sum_{k=-\infty}^{\infty} \phi(c_k) b = \int_{-\infty}^{\infty} \phi(x) dx + O(b\|\phi'\|_1).$$

Lemma 3.4 *Suppose Assumptions 1 and 2 are satisfied, and suppose that $\varphi(k) = O(k^{-\rho})$ for some $\rho > 2N + (3/2)$. Let ε be as defined in Lemma 3.2. Then*

$$\left|\int_{-\infty}^{+\infty} var f_{\mathbf{n}}(x)\; dx - \frac{2}{3\hat{\mathbf{n}}b}\right| \leq \frac{1}{\hat{\mathbf{n}}} R_0(f) \tag{10}$$

$$+O(\hat{\mathbf{n}}^{-1} b^{\varepsilon}[\|f^{'}\|_1 + \|(f^{1/2})'\|_1 + b^{1-\varepsilon}\|f\|_2\|f'\|_2] + \hat{\mathbf{n}}^{-1} b^{-1+\varepsilon}[\|f^{1/2}\|_1 + \|f\|_1]).$$

Proof. Define $m_k \equiv f(\xi_k)$. By integration, we obtain from (8) the following result

$$\int_{-b/2}^{+b/2} \operatorname{var}\hat{f}_{\mathbf{n}}(x)\; dx = \frac{b}{3}\Big(\frac{m_0}{\hat{\mathbf{n}}b} - \frac{m_0^2}{\hat{\mathbf{n}}} + q_{1\mathbf{n}}\Big) + \frac{b}{3}\Big(\frac{m_1}{\hat{\mathbf{n}}b} - \frac{m_1^2}{\hat{\mathbf{n}}} + q_{2\mathbf{n}}\Big) + \frac{b}{3}\Big(- \frac{m_0 m_1}{\hat{\mathbf{n}}} + q_{3\mathbf{n}}\Big)$$

$$= \frac{1}{3\hat{\mathbf{n}}}(m_0 + m_1) - \frac{b}{3\hat{\mathbf{n}}}(m_0^2 + m_1^2 + m_0 m_1) + \frac{b}{3}(q_{1\mathbf{n}} + q_{2\mathbf{n}} + q_{3\mathbf{n}}).$$

By Lemma 3.2,

$$(b/3)(q_{1\mathbf{n}} + q_{2\mathbf{n}} + q_{3\mathbf{n}}) \leq C\hat{\mathbf{n}}^{-1} b^{\varepsilon} w_0.$$

Now, sum over all bins. Observe that

$$\sum_k m_k b = \sum_k p_k = \int_{-\infty}^{+\infty} f(x)\; dx = 1.$$

With the help of Lemma 3.3 with φ replaced by f^2, we get

$$\sum_k m_k^2 b = \sum_k f^2(\xi_k) b = R_0(f) + O(b\|(f^2)'\|_1) = R_0(f) + O(b\|f\|_2\|f'\|_2).$$

Using the definition of w_k in (8) and Lemma 3.3,

$$\sum_k w_k b \leq \|f^{1/2}\|_1 + \|Mf\|_1 + O\left(b\left(\|f'\|_1 + \|(f^{1/2})'\|_1\right)\right)$$

and (10) follows. □

Lemma 3.5 *If Assumptions 1 and 2 hold, then*

$$\int_{-\infty}^{+\infty} bias(x)^2 \; dx = \frac{49}{2,880} b^4 R_2(f) + O\left(b^5 \|f''\|_2 \|f'''\|_2\right). \tag{11}$$

Proof. Let $J_k = [((k-(1/2)b), ((k+(1/2))b)$ be the k-th frequency polygon bin. If x is a fixed point in J_0, then

$$\text{bias}(x) = \frac{1}{6} b^2 \left(\frac{1}{2} - \frac{x}{b}\right) f''(\xi_0) + \frac{1}{6} b^2 \left(\frac{1}{2} + \frac{x}{b}\right) f''(\xi_1) - \frac{1}{2} x^2 f''(\xi_x).$$

where $\xi_0,\ \xi_1,\ \xi_x$ are points in J_0. Lemma (11) follows by summing up the integrals of the square of the biases over each bin as done in Scott (1985). □

Theorem 3.1 *If Assumptions 1 and 2 hold and $\varphi(k) = O(k^{-\rho})$ for some $\rho > 2N + (3/2)$, the value of the bin width that asymptotically minimizes the IMSE of the frequency polygon is*

$$b = b_{\mathbf{n}} = 2\left(\frac{15}{49 R_2(f)}\right)^{1/5} \hat{\mathbf{n}}^{-1/5}.$$

with corresponding

$$IMSE = \frac{5}{12}\left(\frac{49 R_2(f)}{15}\right)^{1/5} \hat{\mathbf{n}}^{-4/5} + O(\hat{\mathbf{n}}^{-1}).$$

Proof. The IMSE is found by combining the results of Lemmas 10 and 11. The leading terms of the IMSE are found to be

$$\frac{2}{3\hat{\mathbf{n}} b} + \frac{49}{2,880} b^4 R_2(f).$$

It is thus sufficient to minimize this function with respect to b. □

It follows from Theorem 3.1 that frequency polygons can achieve the rate of convergence to 0 of order $\hat{\mathbf{n}}^{-4/5}$ with respect to the criterion of integrated mean square error. This is the same rate of convergence to zero of the IMSE of non-negative kernel estimators in the case $N = 1$.

Remark 3.1. It is possible to obtain Theorem 3.1 without the assumption of the third derivative of f in Assumption 2. Let $I_k = [(k-1)b, kb]$ and $J_k = [(k-.5)b, (k+.5)b]$, $k = 0, \pm 1, \pm 2, \ldots$. Let λ be the probability measure associated with f. For any $x \in J_k$,

$$\begin{aligned} E[f_{\mathbf{n}}(X) - f(X)] &= (k + .5 - x/b)\lambda(I_k)/b + (-k + .5 + x/b)\lambda(I_{k+1})/b - f(x) \\ &= T_{1k}(x) + T_{2k}(x) + T_{3k}(x), \end{aligned}$$

where we denote

$$T_{1k}(x) = (k + .5 - x/b)\{\lambda(I_k)/b - f((k - .5)b)\},$$

$$T_{2k}(x) = (-k + .5 + x/b)\{\lambda(I_{k+1})/b - f((k + .5)b)\},$$

$$T_{3k}(x) = (k + .5 - x/b)f((k - .5)b) + (-k + .5 + x/b)f((k + .5)b) - f(x).$$

Note that

$$\int_{J_k} [T_{1k}(x)]^2 dx \leq Cb^4 \int_{I_k} \int_0^1 [f''(tu + (1-t)(k - .5)b)]^2 du\, dt.$$

Define

$$\Omega(f'', u, \epsilon) = \sup\{|f''(u + h)| : |h| \leq \epsilon\}.$$

If we assume that either

$$\lim_{\epsilon \to 0} \int_{-\infty}^{\infty} \Omega^2(f'', u, \epsilon) du < \infty,$$

or

$$\sup_{0<\epsilon<1} \int_{-\infty}^{\infty} \Omega^2(f'', u, \epsilon) du < \infty.$$

Then it follows that $\sum_{-\infty<k<\infty} \int_{J_k} [T_{1k}(x)]^2 dx < \infty$ by noting that

$$\sum_{-\infty<k<\infty} \int_{I_k} \int_0^1 [f''(tu + (1-t)(k - .5)b)]^2 du\, dt < \infty.$$

This argument takes care of the term $T_{1k}(x)$ and $T_{2k}(x)$. Similar arguments can be used to handle $T_{3k}(x)$.

4 Asymptotic variance of $f_{\mathbf{n}}$

Assumption 2^***.** There exists a constant $C > 0$ such that

$$\left|f(x) - f(x')\right| \leq C \left|x - x'\right| \quad \text{for} \quad x, x' \in R.$$

Theorem 4.1 *Let x be a point of the interval $[-b/2, +b/2)$. If Assumptions 1 and 2^* are satisfied and $\varphi(k) = O(k^{-\rho})$ for some $\rho > 2N + (3/2)$, then*

$$\hat{\mathbf{n}}\, b\, var f_{\mathbf{n}}(x) - \left(\frac{1}{2} + \frac{2x^2}{b^2}\right) f(x) \to 0.$$

Proof. Without loss of generality, take $x \in [-b/2, b/2)$. By (8)

$$\begin{aligned} \hat{\mathbf{n}}\, b\, \mathrm{var} f_{\mathbf{n}}(x) &= \left(\frac{1}{2} - \frac{x}{b}\right)^2 \left[\frac{p_0(1-p_0)}{b} + \hat{\mathbf{n}}\, b\, q_{1\mathbf{n}}\right] \\ &+ \left(\frac{1}{2} + \frac{x}{b}\right)^2 \left[\frac{p_1(1-p_1)}{b} + \hat{\mathbf{n}}\, b\, q_{2\mathbf{n}}\right] + 2\left(\frac{1}{4} - \frac{x^2}{b^2}\right)\left[-\frac{p_0 p_1}{b} + \hat{\mathbf{n}}\, b\, q_{3\mathbf{n}}\right]. \end{aligned} \tag{12}$$

By Lemma 3.2,

$$\hat{\mathbf{n}}\, b \left| \left(\frac{1}{2} - \frac{x}{b}\right)^2 q_{1\mathbf{n}} + \left(\frac{1}{2} + \frac{x}{b}\right)^2 q_{2\mathbf{n}} + 2\left(\frac{1}{4} - \frac{x^2}{b^2}\right) q_{3\mathbf{n}} \right| \le C\, b^{\varepsilon} w_0. \tag{13}$$

Since $p_0 = \int_{-b}^{0} f(u)\, du$, Assumption 2 implies that

$$\max\{0, f(x)b - Cb^2\} \le p_0 \le f(x)b + Cb^2.$$

Thus

$$\max\{0, f(x) - (C + f^2(x))b + C^2 b^3\} \le p_0(1-p_0)/b \le f(x) + (C - f^2(x))b + C^2 b^3. \tag{14}$$

Notice that (14) holds also when p_0 is replaced by p_1. Using (12)-(14),

$$\max\left\{0, \left(\frac{1}{2} + \frac{2x^2}{b^2}\right) f(x) - A\right\} \le \hat{\mathbf{n}}\, b\, \mathrm{var} f_{\mathbf{n}}(x) \le \left(\frac{1}{2} + \frac{2x^2}{b^2}\right) f(x) + B$$

with

$$A = \left[C\left(\frac{1}{2} + \frac{2x^2}{b^2}\right) + \frac{4x^2}{b^2} f^2(x)\right] b - \left(1 - \frac{4x^2}{b^2}\right) C f(x) b^2 - C^2 b^3 - C b^{\varepsilon} \omega_0$$

and

$$B = \left[C\left(\frac{1}{2} + \frac{2x^2}{b^2}\right) - \frac{4x^2}{b^2} f^2(x)\right] b - \left(1 - \frac{4x^2}{b^2}\right) C f(x) b^2 + C^2 b^3 + C b^{\varepsilon} \omega_0.$$

The lemma follows easily, because A and B tend to 0 as $n \to \infty$. □

5 Uniform convergence of the frequency polygon estimator

Define

$$\Delta_{\mathbf{i}}(x) = (\hat{\mathbf{n}}b)^{-1}\left[\Big(\frac{1}{2}-\frac{x}{b}\Big)\eta_{\mathbf{i},0}(x) + \Big(\frac{1}{2}+\frac{x}{b}\Big)\eta_{\mathbf{i},1}(x)\right], \tag{15}$$

$$I_{\mathbf{n}}(x) = \sum_{\mathbf{i}\in I_{\mathbf{n}}} E(\Delta_{\mathbf{i}}(x))^2 \quad \text{and} \quad R_{\mathbf{n}}(x) = \sum_{\mathbf{j}\in I_{\mathbf{n}}}\sum_{\substack{\mathbf{i}\in I_{\mathbf{n}} \\ i_k \neq j_k \text{ for some } k}} |\mathrm{Cov}\{\Delta_{\mathbf{i}}(x), \Delta_{\mathbf{j}}(x)\}|$$

Lemma 5.1 *If $\varphi(k) = O(k^{-\rho})$ for some $\rho > 2N + (3/2)$, then*

$$\lim \hat{\mathbf{n}}b(I_{\mathbf{n}}(x) + R_{\mathbf{n}}(x)) < C,$$

where C is a constant independent of x.

Proof. Lemma 5.1 follows by a careful analysis of the proof of Lemma 3.2.

The following lemma of Rio (1993) will be needed in the sequel. Its proof is found in Rio (1995) (see Theorem 4).

Lemma 5.2 *Suppose $\mathcal{A}$ is a σ-field of $(\Omega, \mathcal{F}, P)$ and X is a real-valued random variable taking a.s. its values in $[a, b]$. Suppose furthermore that there exists a random variable U with uniform distribution over $[0, 1]$, independent of $\mathcal{A} \vee \sigma(X)$. Then there exists some random variable X^* independent of $\mathcal{A}$ and with the same distribution as X such that*

$$E|X - X^*| \leq 2(b-a)\alpha(\mathcal{A}, \sigma(X)).$$

Moreover, X^* is a $\mathcal{A} \vee \sigma(X) \vee \sigma(U)$-measurable random variable.

The approximation of strongly mixing r.v.'s by independent ones used later is presented below.

Lemma 5.3 *Suppose S_1, S_2,...,S_r be sets containing m sites each with dist$(S_i, S_j) \geq \delta$ for all $i \neq j$ where $1 \leq i \leq r$ and $1 \leq j \leq r$. Suppose Y_1, Y_2,...,Y_r is a sequence of real-valued r.v.'s measurable with respect to $\mathcal{B}(S_1)$, $\mathcal{B}(S_2)$,...,$\mathcal{B}(S_r)$ respectively and Y_i takes values in $[a, b]$. Then there exists a sequence of independent r.v.'s Y_1^*, Y_2^*,...,Y_r^* independent of Y_1, Y_2,...,Y_r such that Y_i^* has the same distribution as Y_i and satisfies*

$$\sum_{i=1}^{r} E|Y_i - Y_i^*| \leq 2r(b-a)h((r-1)m, m)\varphi(\delta). \tag{16}$$

Proof. Suppose $\delta_j, j \geq 1$ is a sequence of i.i.d. uniform $[0, 1]$ r.v.'s independent of $Y_j, j \geq 1$. Define $Y_1^* = Y_1$. By Lemma 5.2, for every $i \geq 2$, there exists a measurable function f_i such that $Y_i^* = f_i(Y_1, ..., Y_i, \delta_i)$. In addition, each Y_i^* is independent of $Y_1, ..., Y_{i-1}$, has the same distribution as Y_i and satisfies

$$E|Y_i - Y_i^*| \leq 2(b-a)\alpha(\sigma(Y_\ell : \ell < i-1), \sigma(Y_i)) \leq 2(b-a)h((i-1)m, m)\varphi(\delta).$$

The last inequality follows by using (1). For $1 \le i \le r$, we have $h((i-1)m, m) \le h(rm, m)$ since h is nondecreasing in each variable as stated in the introduction and (16) follows by summing up on $1 \le i \le r$.

It remains to show that $Y_1^*, \dots, Y_r^*$ are independent. To prove this it is sufficient to show that Y_i^* and $(Y_1^*, \dots, Y_{i-1}^*)$ are independent for $i > 1$. Note that $(Y_1, \dots, Y_i)$, $\delta_1, \dots, \delta_i$ are independent. Thus $(Y_1, \dots, Y_i, \delta_i)$, $\delta_1, \dots, \delta_{i-1}$ are independent. Since Y_i^* is a measurable function of $Y_1, \dots, Y_i, \delta_i$, it follows that $(Y_i^*, Y_1, \dots, Y_{i-1}), \delta_1, \dots, \delta_{i-1}$ are independent. Now Y_i^* is independent of $Y_1, \dots, Y_{i-1}$. Hence $Y_i^*, (Y_1, \dots, Y_{i-1}), \delta_1, \dots, \delta_{i-1}$ are independent. Finally Y_i^* and $(Y_1^*, \dots, Y_{i-1}^*)$ are independent since $(Y_1^*, \dots, Y_{i-1}^*)$ is measurable with respect to the σ-field generated by $Y_1, \dots, Y_{i-1}, \delta_1, \dots, \delta_{i-1}$. □

Define

$$S_{\mathbf{n}}(x) = \sum_{\substack{i_k=1\\k=1,\dots,N}}^{n_k} \Delta_{\mathbf{i}}(x). \tag{17}$$

Then

$$S_{\mathbf{n}}(x) = f_{\mathbf{n}}(x) - E f_{\mathbf{n}}(x). \tag{18}$$

Without loss of generality assume that $n_i = 2pq_i$ for $1 \le i \le N$. The random variables $\Delta_{\mathbf{i}}(x)$ can be grouped into $2^N q_1 \times q_2 \times \ldots \times q_N$ cubic blocks of side p. Denote

$$U(1, \mathbf{n}, \mathbf{j}, x) = \sum_{\substack{i_k=2j_kp+1\\k=1,\dots,N}}^{(2j_k+1)p} \Delta_{\mathbf{i}}(x), \tag{19}$$

$$U(2, \mathbf{n}, \mathbf{j}, x) = \sum_{\substack{i_k=2j_kp+1\\k=1,\dots,N-1}}^{(2j_k+1)p} \sum_{i_N=(2j_N+1)p+1}^{2(j_N+1)p} \Delta_{\mathbf{i}}(x),$$

$$U(3, \mathbf{n}, \mathbf{j}, x) = \sum_{\substack{i_k=2j_kp+1\\k=1,\dots,N-2}}^{(2j_k+1)p} \sum_{i_{N-1}=(2j_{N-1}+1)p+1}^{2(j_{N-1}+1)p} \sum_{i_N=2j_Np+1}^{(2j_N+1)p} \Delta_{\mathbf{i}}(x),$$

$$U(4, \mathbf{n}, \mathbf{j}, x) = \sum_{\substack{i_k=2j_kp+1\\k=1,\dots,N-2}}^{(2j_k+1)p} \sum_{i_{N-1}=(2j_{N-1}+1)p+1}^{2(j_{N-1}+1)p} \sum_{i_N=(2j_N+1)p+1}^{2(j_N+1)p} \Delta_{\mathbf{i}}(x),$$

and so on. Note that

$$U(2^{N-1}, \mathbf{n}, \mathbf{j}, x) = \sum_{\substack{i_k=(2j_k+1)p+1\\k=1,\dots,N-1}}^{2(j_k+1)p} \sum_{i_N=2j_Np+1}^{(2j_N+1)p} \Delta_{\mathbf{i}}(x).$$

Finally

$$U(2^N, \mathbf{n}, \mathbf{j}, x) = \sum_{\substack{i_k=(2j_k+1)p+1\\k=1,\dots,N}}^{2(j_k+1)p} \Delta_{\mathbf{i}}(x).$$

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

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

Clearly

$$S_{\mathbf{n}}(x) = \sum_{i=1}^{2^N} T(\mathbf{n}, i, x). \tag{20}$$

The blocking idea here is reminiscient of the blocking scheme in Tran (1990) and Politis and Romano (1993).

Without loss of generality we will write all for $i = 1$. Now, $T(\mathbf{n}, 1, x)$ is the sum of

$$r = q_1 \times q_2 \times \ldots \times q_N \tag{21}$$

of the $U(1, \mathbf{n}, \mathbf{j}, x)$'s. Note that $U(1, \mathbf{n}, \mathbf{j}, x)$ is measurable with the σ-field generated by $X_{\mathbf{i}}$ with $\mathbf{i}$ belonging to the set of sites

$$\{\mathbf{i} : 2j_k p + 1 \le i_k \le (2j_k + 1)p, \quad k = 1, \ldots, N\}.$$

These sets of sites are separated by a distance of at least p. Enumerate the r.v.'s $U(1, \mathbf{n}, \mathbf{j}, x)$ and the corresponding σ-fields with which they are measurable in an arbitrary manner and refer to them respectively as $Y_1, Y_2, \ldots, Y_r$ and $S_1, S_2, \ldots, S_r$. Approximate $Y_1, Y_2, \ldots, Y_r$ by the r.v.'s $Y_1^*, Y_2^*, \ldots, Y_r^*$ as was done in Lemma 5.3. Clearly,

$$|Y_i| < Cp^N (\hat{\mathbf{n}} b)^{-1}. \tag{22}$$

Denote

$$\varepsilon_{\mathbf{n}} = \eta(\log \hat{\mathbf{n}} (\hat{\mathbf{n}} b)^{-1})^{1/2},$$

where η is a constant to be chosen later.

Define $\alpha_{\mathbf{n}} = b\, h(\hat{\mathbf{n}}, p^N)\varphi(p)(\log \hat{\mathbf{n}}(\hat{\mathbf{n}} b)^{-1})^{-1/2}$.

Lemma 5.4 *Given an arbitrarily large positive constant a, there exists a positive constant C such that*

$$P\left[\sup_{x \in D} |T(\mathbf{n}, 1, x_k)| > \varepsilon_{\mathbf{n}}\right] \le C(b^{-1}\hat{\mathbf{n}}^{-a} + \alpha_{\mathbf{n}} b^{-1}$$

Proof. Since $T(\mathbf{n}, 1, x)$ is equal to $\sum_{i=1}^r Y_i$, we have

$$P[|T(\mathbf{n}, 1, x)| > \varepsilon_{\mathbf{n}}] \le P\Big[|\sum_{i=1}^r Y_i^*| > \varepsilon_{\mathbf{n}}/2\Big] + P\Big[\sum_{i=1}^r |Y_i - Y_i^*| > \varepsilon_{\mathbf{n}}/2\Big]. \tag{23}$$

We now proceed to obtain bounds for the two terms on the right hand side of (23).

By Markov's inequality and using (16), (22) and recall that the sets of sites with respect to which the Y_i's are measurable are separated by a distance of at least p,

$$P\left[\sum_{i=1}^{r}|Y_i - Y_i^*| > \varepsilon_{\mathbf{n}}\right] \leq Crp^N(\hat{\mathbf{n}}b)^{-1}h(\hat{\mathbf{n}}, p^N)\varphi(p)\varepsilon_{\mathbf{n}}^{-1} \sim \alpha_{\mathbf{n}}. \tag{24}$$

Set

$$\lambda_{\mathbf{n}} = (\hat{\mathbf{n}}b\log\hat{\mathbf{n}})^{1/2}, \tag{25}$$

$$p = \left[\left(\frac{\hat{\mathbf{n}}b}{4\lambda_{\mathbf{n}}}\right)^{1/N}\right] \sim \left(\frac{\hat{\mathbf{n}}b}{\log\hat{\mathbf{n}}}\right)^{\frac{1}{2N}}. \tag{26}$$

A simple computation yields,

$$\lambda_{\mathbf{n}}\varepsilon_{\mathbf{n}} = \eta\log\hat{\mathbf{n}},$$

and by Lemma 5.1

$$\lambda_{\mathbf{n}}^2\sum_{i=0}^{r}E(Y_i^*)^2 \leq C\hat{\mathbf{n}}b(I_{\mathbf{n}}(x) + R_{\mathbf{n}}(x))\log\hat{\mathbf{n}} < C\log\hat{\mathbf{n}}.$$

Using (22), we have $|\lambda_{\mathbf{n}}Y_i^*| < 1/2$ for large $\hat{\mathbf{n}}$. Applying Berstein's inequality,

$$\begin{aligned} P\Big[|\sum_{i=0}^{r}Y_i^*| > \varepsilon_{\mathbf{n}}\Big] &\leq 2\exp\Big(-\lambda_{\mathbf{n}}\varepsilon_{\mathbf{n}} + \lambda_{\mathbf{n}}^2\sum_{i=0}^{r}E(Y_i^*)^2\Big) \\ &\leq 2\exp\left((-\eta + C)\log\hat{\mathbf{n}}\right) \leq \hat{\mathbf{n}}^{-a}, \end{aligned} \tag{27}$$

for sufficiently large $\hat{\mathbf{n}}$.

Combining (23), (24) and (27),

$$P[\sup_{x\in D}|T(\mathbf{n},1,x)| > \varepsilon_{\mathbf{n}}] \leq C(b^{-1}\hat{\mathbf{n}}^{-a} + \alpha_{\mathbf{n}}b^{-1}).$$

□

Denote

$$\theta_1 = \frac{\rho + 3N}{\rho - 3N}, \quad \theta_2 = \frac{N - \rho}{\rho - 3N}$$

$$\theta_3 = \frac{3N + \rho}{\rho - N - 2N\tilde{k}}, \quad \theta_4 = \frac{N - \rho}{\rho - N - 2\tilde{k}N}.$$

Theorem 5.1 *Suppose $\varphi(k) = O(k^{-\rho})$ for some $\rho > 0$ holds.*

(i) If (2) is satisfied and

$$\hat{\mathbf{n}}b_{\mathbf{n}}^{\theta_1}(\log\hat{\mathbf{n}})^{\theta_2} \to \infty, \tag{28}$$

(ii) or if (3) is satisfied and

$$\hat{\mathbf{n}}b_{\mathbf{n}}^{\theta_3}(\log\hat{\mathbf{n}})^{\theta_4} \to \infty. \tag{29}$$

then

$$\sup_{x\in D}|f_{\mathbf{n}}(x)-Ef_{\mathbf{n}}(x)| = O\Big(\big(\log\hat{\mathbf{n}}(\hat{\mathbf{n}}b)^{-1}\big)^{1/2}\Big) \quad \textit{in probability.} \tag{30}$$

Remark 5.1. For (28) to hold it is necessary that $\rho > 3N$ since $b = b_{\mathbf{n}}$ goes to zero. Hence $\theta_1 > 1$ and (28) implies

$$\hat{\mathbf{n}}b \to \infty. \tag{31}$$

which is a condition for $f_{\mathbf{n}}(x)$ to converge to $f(x)$ in the case $N = 1$.

Similarly, it can be shown that (29) implies (31).

Proof of Theorem 5.1.
(i) To complete the proof, we will show that $b^{-1}\hat{\mathbf{n}}^{-a} \to 0$ and $\alpha_{\mathbf{n}}b^{-1} \to 0$.
(31) implies $b^{-1}\hat{\mathbf{n}}^{-a} \to 0$ for $a > 1$.
Moreover

$$b^{-1}\hat{\mathbf{n}}^{-a} \sim b^{-(3N-\rho)/2N}\hat{\mathbf{n}}^{(3N-\rho)/2N}\log\hat{\mathbf{n}}^{-(N-\rho)/2N} \tag{32}$$

and (28) is equivalent to $(\alpha_{\mathbf{n}}b^{-1})^{-1} \to \infty$, which implies $\alpha_{\mathbf{n}}b^{-1} \to 0$.
The proof of (ii) is similar. □

Theorem 5.2 *Suppose $\varphi(k) = O(k^{-\rho})$ for some $\rho > 0$ and Assumption 2^* hold. If (2) and (28) hold, or (3) and (29) hold, then*

$$\sup_{x\in D}|f_{\mathbf{n}}(x)-f(x)| = O(\Psi_{\mathbf{n}}) \quad \textit{in probability.}$$

Proof. By assumption 2^*,

$$\sup_{x\in D}|Ef_{\mathbf{n}}(x)-f(x)| \le C\,b \tag{33}$$

The proof follows easily from Theorem 5.1 and (33). □

Example 5.1.

(i) Take $b = C\hat{\mathbf{n}}^{-1/5}$ where b is the optimal bin width derived in Section 3. Then (28) is satisfied if $\rho > 9N/2$, and (29) is satisfied if $\rho > 2N + (5N\tilde{k})/2$.

(ii) Take $b = \left(\hat{\mathbf{n}}^{-1}\log\hat{\mathbf{n}}\right)^{1/3}$. Then $\Psi_{\mathbf{n}} = \left(\hat{\mathbf{n}}^{-1}\log\hat{\mathbf{n}}\right)^{1/3}$, which is the optimal rate for the *i.i.d.* case for $N = 1$. Then (28) is satisfied if $\rho > 6N$, and (29) is satisfied if $\rho > 3N + N\tilde{k}$.

6 Rate of the a.s. convergence of $f_{\mathbf{n}}$

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

Define

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

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

Theorem 6.1 *Suppose $\varphi(k) = O(k^{-\rho})$ for some $\rho > 0$ and Assumption 2^* hold.*

(i) If (2) is satisfied and

$$\hat{\mathbf{n}} b_{\mathbf{n}}^{\theta_1^*} (\log \hat{\mathbf{n}})^{\theta_2^*} (g(\mathbf{n}))^{-2N/(\rho-5N)} \to \infty, \tag{34}$$

(ii) or if (3) is satisfied and

$$\hat{\mathbf{n}} b_{\mathbf{n}}^{\theta_3^*} (\log \hat{\mathbf{n}})^{\theta_4^*} (g(\mathbf{n}))^{-2N/(\rho-(2\tilde{k}+3)N} \to \infty. \tag{35}$$

then

$$\sup_{x \in D} |f_{\mathbf{n}}(x) - f(x)| = O(\Psi_{\mathbf{n}}) \quad a.s. \tag{36}$$

Proof.
(i) Condition (34) is equivalent to

$$b^{-1} \alpha_{\mathbf{n}} \hat{\mathbf{n}} g(\mathbf{n}) \to 0,$$

which entails

$$\sum_{\mathbf{n} \in Z^N} b^{-1} \alpha_{\mathbf{n}} < \infty.$$

The theorem folllows easily by the Borel-Cantelli lemma and (33).
(ii) The proof of (ii) is similar to that of (i) and is omitted. □

References

[1] Bolthausen, E. (1982) "On the central limit theorem for stationary random fields", *Ann. Probab.* **10**: 1047–1050.

[2] Bradley, R.C. (1986) "Basic properties of strong mixing conditions. In : *Dependence in Probability and Statistics*, vol 11, Birkhäuser: 165–192.

[3] 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.

[4] 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.

[5] Carbon, M. (2006) "Polygones des fréquences pour des champs aléatoires", *Comptes Rendus de l'Académie des Sciences de Paris*, A paraître.

[6] Davydov, Yu A. (1970) "The invariant principle for stationary processes", *Theor. Probab. Appl.* **14**: 487–498.

[7] Gorodetskii, V.V. (1977) "On the strong mixing property for linear sequences", *Theory Probability Appl.* **22**: 411–413.

[8] Guyon, X.; Richardson, S. (1984) "Vitesse de convergence du théorème de la limite centrale pour des champs faiblement dépendants", *Z. Wahrsch. Verw. Gebiete* **66**: 297–314.

[9] Guyon, X. (1987) "Estimation d'un champ par pseudo-vraisemblance conditionnelle: Étude asymptotique et application au cas Markovien", *Proc. 6th Franco-Belgian Meeting of Statisticians.*

[10] Hall, P.; Hannan, E.J. (1988) "On stochastic complexity and nonparametric density estimation", *Biometrika* **75**: 705–714.

[11] Ibragimov, I. A.; Linnik, Yu. V. (1971) *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhoff, Groningen.

[12] Masry, E.; Györfi (1987) "Strong consistency and rates for recursive density estimators for stationary mixing processes", *J. Multivariate Anal.* **22**: 79–93.

[13] Nahapetian, B. S. (1980) "The central limit theorem for random fields with mixing conditions", in: R. L. Dobrushin & Ya G. Sinai (Eds.) *Adv. in Probability 6, Multicomponent Systems*: 531–548.

[14] Nahapetian, B. S. (1987) "An approach to proving limit theorems for dependent random variables", *Theory Prob. Appl.* **32**: 535–539.

[15] Neaderhouser, C. C. (1980) "Convergence of block spins defined on random fields", *J. Statist. Phys.* **22**: 673–684.

[16] Politis, D. N.; Romano, J. P. (1993). "Nonparametric resampling for homogeneous strong mixing random fields", *J. Multivariate Anal.* **47**: 301–328.

[17] Rio, E. (1995) "The functional law of the iterated logarithm for stationary strongly mixing sequences", *Ann. Prob.* **23**: 1188–1203.

[18] Robinson, P.M. (1983) "Nonparametric estimators for time series", *J. Time Series Anal.* **4**: 185–207.

[19] Rosenblatt, M. (1985) *Stationary Sequences and Random Fields.* Birkhäuser, Boston.

[20] Roussas, G.G. (1969) "Nonparametric estimation of the transition distribution of a Markov process", *Ann. Inst. Statist. Math.* **21**: 73–87.

[21] Roussas, G.G. (1988) "Nonparametric estimation in mixing sequences of random variables", *Jour. Statist. Plann. Inference* **18**: 135–149.

[22] Scott, D.W. (1985) "Frequency polygons, theory and applications", *J. Amer. Stat. Assoc.* **80**: 348–354.

[23] Stone, C.J. (1983) "Optimal unifom rate of convergence for non parametric estimators of a density function and its derivative", in: M.H. Revzi, J.S. Rustagi & D. Siegmund (Eds.) *Recent Advances in Statistics: Papers in Honor of H. Chernoff*: 393–406.

[24] Takahata, H. (1983) "On the rates in the central limit theorem for weakly dependent random fields", *Z. Wahrsch. Verw. Gebiete* **62**: 477–480.

[25] Tran, L.T. (1990) "Kernel density estimation on random fields", *J. Multivariate Anal.* **34**: 37–53.

[26] Tran, L.T.; Yakowitz, S. (1993) "Nearest neighbor estimators for random fields", *J. Multivariate Anal.* **44**: 23–46.