Optimal asymptotic quadratic errors of density estimators on random fields

Gérard Biau

Laboratoire de Statistique Théorique et Appliquée, Université Pierre et Marie Curie – Paris VI, Boîte 158, 175 rue du Chevaleret, 75013 Paris, France

Abstract

Let $\mathbb{Z}^N$ denote the integer lattice points in the N-dimensional Euclidean space. Kernel estimation of the multivariate density of a random field indexed by $\mathbb{Z}^N$ is investigated. The loss between the estimator and the unknown density is measured by means of mean squared and mean integrated squared errors. Under mild mixing conditions, we show that the kernel density estimator has exactly the same asymptotic error as in the i.i.d. case.

Key words: Kernel density estimation, Random field, Optimality, Mixing.
1991 MSC: 62G07

1 Introduction

The theory of spatial models, or random fields, has developed significantly over the last few years (Guyon (1995)), stimulated by a growing demand from applied research areas like spatial modeling, image analysis or Geostatistics. Parallel to this development, the existence of high speed, inexpensive computing has made it easy to look at data in ways that were once difficult. One area that has benefited greatly from this new freedom is that of nonparametric statistics. However, most of existing theoretical nonparametric results of dependent random variables pertain to time series (Bosq (1998)) and few generalizations to the spatial domain are available. Our goal in this paper is to study kernel density estimation for random variables which show spatial interaction. Currently the author is working on extensions of the present results

Email address: `biau@ccr.jussieu.fr` (Gérard Biau).

to nonparametric estimation and prediction of random fields continuously indexed. Key references on the subject are Tran (1990), Tran and Yakowitz (1993), Carbon, Hallin and Tran (1996) and Carbon, Tran and Wu (1997).

More precisely, let $\mathbb{Z}^N$, $N \geq 1$, denote the integer lattice points in the N-dimensional Euclidean space. We consider an $\mathbb{R}^d$-valued ($d \geq 1$) random field $(X_{\mathbf{n}})$ indexed by $\mathbf{n}$ in $\mathbb{Z}^N$ and defined on some probability space $(\Omega, \mathcal{A}, \mathbf{P})$. For a point $\mathbf{n} = (n_1, \ldots, n_N)$ in $\mathbb{Z}^N$, we let $\|\mathbf{n}\|$ be a norm of $\mathbf{n}$ and $\hat{\mathbf{n}} = n_1 \ldots n_N$. The spatial dependence of the process will be measured by means of α and β-mixing. We let the α and β-mixing sequences of coefficients of the field $(X_{\mathbf{n}})$ be defined as

$$\alpha_k = \sup_{\substack{\mathbf{i},\mathbf{j} \in \mathbb{Z}^N \\ \|\mathbf{i}-\mathbf{j}\| \geq k}} \alpha(\sigma(X_{\mathbf{i}}), \sigma(X_{\mathbf{j}})), \quad k \in \mathbb{N},$$

and

$$\beta_k = \sup_{\substack{\mathbf{i},\mathbf{j} \in \mathbb{Z}^N \\ \|\mathbf{i}-\mathbf{j}\| \geq k}} \beta(\sigma(X_{\mathbf{i}}), \sigma(X_{\mathbf{j}})), \quad k \in \mathbb{N},$$

where, for $\mathcal{B}$ and $\mathcal{C}$ two sub σ-fields of $\mathcal{A}$,

$$\alpha(\mathcal{B}, \mathcal{C}) = \sup_{\substack{B \in \mathcal{B} \\ C \in \mathcal{C}}} |\mathbf{P}(B \cap C) - \mathbf{P}(B)\mathbf{P}(C)| \tag{1}$$

and

$$\beta(\mathcal{B}, \mathcal{C}) = \frac{1}{2} \sup \Big\{ \sum_{i \in I} \sum_{j \in J} |P(B_i \cap C_j) - P(B_i)P(C_j)| \Big\}, \tag{2}$$

where the last supremum is taken over all finite partitions $(B_i)_{i \in I}$ and $(C_j)_{j \in J}$ of Ω with $B_i \in \mathcal{B}$ and $C_j \in \mathcal{C}$. Obviously, $0 \leq 2\alpha_k \leq \beta_k$. For a complete discussion on mixing, we refer to Doukhan (1994), Rio (2000) and the references therein.

For $\mathbf{n} = (n_1, \ldots, n_N)$ in $\mathbb{Z}^N$, we define the rectangular region $\mathcal{I}_{\mathbf{n}}$ by $\mathcal{I}_{\mathbf{n}} = \{\mathbf{i} \in \mathbb{Z}^N : 1 \leq i_k \leq n_k,\, k = 1, \ldots, N\}$ and we write $\mathbf{n} \to +\infty$ if $\min_{k=1,\ldots,N} n_k \to +\infty$. Assume now that we observe the field $(X_{\mathbf{n}})$ on $\mathcal{I}_{\mathbf{n}}$ and that each $X_{\mathbf{n}}$ has a density f with respect to the Lebesgue measure on $\mathbb{R}^d$. Since the total number of observations in $\mathcal{I}_{\mathbf{n}}$ is $\hat{\mathbf{n}}$, the *kernel estimator* $f_{\mathbf{n}}$ of f (Rosenblatt (1956), Parzen (1962), Devroye (1987)) is given by

$$f_{\mathbf{n}}(x) = \frac{1}{\hat{\mathbf{n}} h_{\mathbf{n}}^d} \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} K\Big(\frac{x - X_{\mathbf{i}}}{h_{\mathbf{n}}}\Big), \quad x \in \mathbb{R}^d,$$

where the *kernel* K is a bounded integrable mapping from $\mathbb{R}^d$ to $\mathbb{R}$ and $(h_{\mathbf{n}})$ is a sequence of positive bandwidths tending to 0 as $\mathbf{n}$ tends to infinity. In a related spatial framework, the asymptotic normality of $f_{\mathbf{n}}(x)$ has been established by

Tran (1990). Tran and Yakowitz (1993) have investigated nearest neighbor estimators for random fields. Carbon, Hallin and Tran (1996) showed the L^1 consistency of $f_{\mathbf{n}}$ and Carbon, Tran and Wu (1997) uniform convergence.

In the present paper, we study the consistency of $f_{\mathbf{n}}$ towards f in *mean squared error* (MSE), defined by $\mathbf{E}(f_{\mathbf{n}}(x)-f(x))^2$, as well as in *mean integrated squared error* (MISE), defined by

$$\mathbf{E}\int_{\mathbb{R}^d} (f_{\mathbf{n}}(x) - f(x))^2 \, \mathrm{d}x = \int_{\mathbb{R}^d} \mathbf{E}(f_{\mathbf{n}}(x) - f(x))^2 \, \mathrm{d}x \, .$$

Here and throughout the paper, $\mathbf{E}$ denotes the expectation with respect to $\mathbf{P}$. Under mild mixing assumptions, we show that both errors turn out to be asymptotically the same as in the *i.i.d.* case. These results complete and extend to fields a result of Bosq (1995). The organization of the paper is as follows: covariance and density inequalities are presented in Section 2. We state the MSE results in Section 3 and the MISE results in Section 4. Proofs are given in Section 5.

2 Inequalities

Consider two random variables U and V which take their values in $\mathbb{R}^d$ equipped with its Borel σ-field and the Euclidean norm $\|.\|$. Let the mixing coefficients $\alpha = \alpha(\sigma(U), \sigma(V))$ and $\beta = \beta(\sigma(U), \sigma(V))$ be defined as in (1) and (2), respectively. Set $g_{(U,V)} = f_{(U,V)} - f_U \otimes f_V$, where f_Z denotes the density of a random variable Z with respect to the appropriate Lebesgue measure. The L^p norm in $\mathbb{R}^d$ is denoted by $\|.\|_p$, $1 \le p \le +\infty$. The following inequalities will be useful for our purpose.

Lemma 1 *Let K be a bounded integrable real mapping defined on $\mathbb{R}^d$ and let h be a positive real number. Let $K_h(.) = K(./h)/h^d$.*

(i) *Suppose that $\delta_p = \|g_{(U,V)}\|_p < +\infty$ for some $p \in [1, +\infty]$. Then*

$$\left|\mathrm{Cov}(K_h(U), K_h(V))\right| \le \min\left(\delta_p \, \|K\|_q^2 \, h^{-2d/p}, 4\alpha \, \|K\|_\infty^2 \, h^{-2d}\right),$$

where q satisfies $p^{-1} + q^{-1} = 1$.

(ii) *Suppose that U and V have the same distribution and that the mapping*

$$\psi_{(U,V)} : \begin{cases} \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R} \\ (u, v) \quad \mapsto \displaystyle\int_{\mathbb{R}^d} |g_{(U,V)}(x-u, x-v)| \mathrm{d}x \end{cases}$$

is defined and bounded. Then

$$\int_{\mathbb{R}^d} \Big|\mathrm{Cov}(K_h(x-U), K_h(x-V))\Big| \mathrm{d}x \le \min\left(\|\psi_{(U,V)}\|_\infty \, \|K\|_1^2, 2\beta \, \|K\|_2^2 \, h^{-d}\right).$$

Remark 2 *Note that the assumptions on the kernel K imply that $K \in L^p$ for any $p \ge 1$.*

Lemma 3 *If for some constant l, we have*

$$|g_{(U,V)}(x', y') - g_{(U,V)}(x, y)| \le l(\|x' - x\|^2 + \|y' - y\|^2)^{1/2}, \quad x, x', y, y' \in \mathbb{R}^d,$$

then there exists a constant $\gamma(d, l)$ such that

$$|g_{(U,V)}(x, y)| \le \gamma(d, l) \, \alpha^{1/(2d+1)}, \quad x, y \in \mathbb{R}^d.$$

One may choose $\gamma(d, l) = v_d^{-2} + l\sqrt{2}$, where v_d denotes the volume of the unit ball in $\mathbb{R}^d$.

Lemma 3 is proved in Bosq (1995). Examples are presented below.

3 Optimal rate in mean squared error

In order to state a uniform result, we introduce some constants in the following assumptions.

(A1) The kernel $K : \mathbb{R}^d \to \mathbb{R}$ is a *symmetric integrable bounded* function such that

$$\int_{\mathbb{R}^d} K(u) \, \mathrm{d}u = 1, \qquad \lim_{\|u\| \to +\infty} \|u\|^d K(u) = 0$$

and, for some constant A,

$$\|K\|_\infty \le A \quad \text{and} \quad \int_{\mathbb{R}^d} \|u\|^2 K(u) \, \mathrm{d}u \le A.$$

(A2) $h_{\mathbf{n}} \to 0$ and $\hat{\mathbf{n}} h_{\mathbf{n}} \to +\infty$ as $\mathbf{n} \to +\infty$.

(A'2) $h_{\mathbf{n}} = b_{\mathbf{n}} \hat{\mathbf{n}}^{-1/(d+4)}$, where $b_{\mathbf{n}} \to B > 0$ as $\mathbf{n} \to +\infty$.

(A3) The $X_{\mathbf{n}}$'s have a common density f twice continuously differentiable and, for some constant C, $\|f^{(i)}\|_\infty \le C$, $i = 0, 1, 2$, where $f^{(i)}$ denotes any partial derivative of order i.

(A4) For $\mathbf{i} \neq \mathbf{j}$, $(X_{\mathbf{i}}, X_{\mathbf{j}})$ has a density and, for some $p \in]2, +\infty]$ and some constant D,

$$\sup_{\mathbf{i} \neq \mathbf{j}} \|g_{(X_{\mathbf{i}}, X_{\mathbf{j}})}\|_p \leq D.$$

(A'4) For $\mathbf{i} \neq \mathbf{j}$, $(X_{\mathbf{i}}, X_{\mathbf{j}})$ has a density and, for some constant L,

$$\sup_{\mathbf{i} \neq \mathbf{j}} \|g_{(X_{\mathbf{i}}, X_{\mathbf{j}})}(x', y') - g_{(X_{\mathbf{i}}, X_{\mathbf{j}})}(x, y)\| \leq L(\|x' - x\|^2 + \|y' - y\|^2)^{1/2},$$

$x, y, x', y' \in \mathbb{R}^d$.

(A5) $\alpha_k \leq R k^{-r}$ for $k \geq 1$, where R and r are positive constants.

Example 4 *Consider a stationary Gaussian field* $(X_{\mathbf{n}})$ (i.e., *with* $\mathrm{Cov}(X_{\mathbf{i}}, X_{\mathbf{j}})$ *function of* $\mathbf{i} - \mathbf{j}$ *only) with a spatial autocorrelation function* $\rho(\mathbf{i})$ *satisfying* $\sup_{\mathbf{i} \neq 0} |\rho(\mathbf{i})| < 1$, *then* **(A3)**, **(A4)**, **(A'4)** *hold. If in addition* $(X_{\mathbf{n}})$ *is Markovian, then* **(A5)** *holds. More generally, necessary and sufficient conditions for* **(A5)** *to hold can be found in Rosenblatt (1956) and Doukhan (1994).*

We are now ready to state the result.

Theorem 5 (i) *If* **(A1)**, **(A2)**, **(A3)**, **(A4)** (resp., **(A'4)**) *and* **(A5)** *are satisfied, and if* $r > 2N(p-1)/(p-2)$ (resp., $r > N(2d+1)/(d+1)$) *then, for* $x \in \mathbb{R}^d$,

$$\mathbf{E}(f_{\mathbf{n}}(x) - f(x))^2 = \frac{h_{\mathbf{n}}^4}{4} \Delta_d^2(x) + \frac{f(x)}{\hat{\mathbf{n}} h_{\mathbf{n}}^d} \int_{\mathbb{R}^d} K^2(u) \mathrm{d}u + \mathrm{o}\Big(h_{\mathbf{n}}^4 + \frac{1}{\hat{\mathbf{n}} h_{\mathbf{n}}^d}\Big),$$

where

$$\Delta_d(x) = \sum_{1 \leq i,j \leq d} \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \int_{\mathbb{R}^d} u_i u_j K(u) \mathrm{d}u \, .$$

(ii) *Under the conditions above, where* **(A2)** *is replaced by* **(A'2)**, *we have*

$$\hat{\mathbf{n}}^{4/(d+4)} \mathbf{E}(f_{\mathbf{n}}(x) - f(x))^2 \to \chi(K, f),$$

where

$$\chi(K, f) = \frac{B^4}{4} \Delta_d^2(x) + \frac{f(x)}{B^d} \int_{\mathbb{R}^d} K^2(u) \mathrm{d}u \, .$$

Let us define $\mathcal{F}(A, B)$ as the family of density estimators $(f_{\mathbf{n}})$ which satisfy **(A1)** and **(A'2)**, and $\mathcal{X} = \mathcal{X}(C, p_0, D, R, r)$ (*resp.*, $\mathcal{X} = \mathcal{X}(C, L, R, r)$) as the family of $\mathbb{R}^d$-valued random fields $(X_{\mathbf{n}})$ which satisfy **(A3)**, **(A4)** and **(A5)** (*resp.*, **(A3)**, **(A'4)** and **(A5)**) for fixed $p_0 \in]2, +\infty]$ and $r > 2N(p_0 - 1)/(p_0 - 2)$ (*resp.*, $r > N(2d+1)/(d+1)$). Then our uniformity result can be formulated as follows.

Corollary 6

$$\sup_{(f_{\mathbf{n}})\in\mathcal{F}(A,B)} \limsup_{\mathbf{n}\to+\infty} \sup_{(X_{\mathbf{n}})\in\mathcal{X}} \hat{\mathbf{n}}^{4/(d+4)} \sup_{x\in\mathbb{R}^d} \mathbf{E}(f_{\mathbf{n}}(x)-f(x))^2 < +\infty.$$

4 Optimal rate in mean integrated squared error

As in the non-integrated case, we need some technical assumptions.

(A1), **(A2)** and **(A'2)** as in Section 3.

(A3) The $X_{\mathbf{n}}$'s have a common density f twice continuously differentiable with, for some constant C, $\|f\|_2$ and $\|f^{(2)}\|_2 \le C$.

(A4) For $\mathbf{i}\neq\mathbf{j}$, $(X_{\mathbf{i}},X_{\mathbf{j}})$ has a density, the mapping

$$\psi_{(\mathbf{i},\mathbf{j})} : \begin{cases} \mathbb{R}^d\times\mathbb{R}^d \to \mathbb{R} \\ (u,v) \mapsto \displaystyle\int_{\mathbb{R}^d} |g_{(X_{\mathbf{i}},X_{\mathbf{j}})}(x-u,x-v)|\,\mathrm{d}x \end{cases}$$

is defined and, for some constant D, $\sup_{\mathbf{i}\neq\mathbf{j}} \|\psi_{(\mathbf{i},\mathbf{j})}\|_\infty \le D$.

(A5) The series $\sum_{k\ge1} k^{N-1}\beta_k$ is convergent.

A close examination shows that the assumptions **(A3)** and **(A4)** hold for example for a stationary Gaussian field $(X_{\mathbf{n}})$ with an autocorrelation function $\rho(\mathbf{i})$ satisfying $\sup_{\mathbf{i}\neq0}|\rho(\mathbf{i})|<1$. Sufficient conditions for **(A5)** to hold can be found in Doukhan (1994).

Theorem 7 (i) *If* **(A1)**, **(A2)**, **(A3)**, **(A4)** *and* **(A5)** *are satisfied, then*

$$\mathbf{E}\int_{\mathbb{R}^d}(f_{\mathbf{n}}(x)-f(x))^2\mathrm{d}x = \frac{h_{\mathbf{n}}^4}{4}\Gamma_d^2 + \frac{\int_{\mathbb{R}^d} K^2(u)\mathrm{d}u}{\hat{\mathbf{n}}h_{\mathbf{n}}^d} + \mathrm{o}\Big(h_{\mathbf{n}}^4 + \frac{1}{\hat{\mathbf{n}}h_{\mathbf{n}}^d}\Big),$$

where

$$\Gamma_d^2 = \int_{\mathbb{R}^d}\Big[\sum_{1\le i,j\le d}\frac{\partial^2 f}{\partial x_i\partial x_j}(x)\int_{\mathbb{R}^d}u_iu_jK(u)\mathrm{d}u\Big]^2\mathrm{d}x\,.$$

(ii) *Under the conditions above, with* **(A2)** *being replaced by* **(A'2)**, *we have*

$$\hat{\mathbf{n}}^{4/(d+4)}\mathbf{E}\int_{\mathbb{R}^d}(f_{\mathbf{n}}(x)-f(x))^2\mathrm{d}x \to \kappa(K,f),$$

where

$$\kappa(K,f) = \frac{B^4}{4}\,\Gamma_d^2 + \frac{\int_{\mathbb{R}^d} K^2(u)\mathrm{d}u}{B^d}\,.$$

Denoting by $\mathcal{X} = \mathcal{X}(C,D)$ the family of $\mathbb{R}^d$-valued random fields $(X_{\mathbf{n}})$ which satisfy **(A3)**, **(A4)** and **(A5)** with $\sum_{k\geq 1} k^{N-1} \sup_{X\in\mathcal{X}} \beta_k^X < +\infty$ (with clear notations), we get the following corollary.

Corollary 8

$$\sup_{(f_{\mathbf{n}})\in\mathcal{F}(A,B)} \limsup_{\mathbf{n}\to+\infty} \sup_{(X_{\mathbf{n}})\in\mathcal{X}} \hat{\mathbf{n}}^{4/(d+4)}\mathbf{E}\int_{\mathbb{R}^d} (f_{\mathbf{n}}(x) - f(x))^2\mathrm{d}x < +\infty.$$

5 Proofs

5.1 Proof of Lemma 1

Proof of (i) We have

$$\mathrm{Cov}(K_h(U),K_h(V)) = \frac{1}{h^{2d}}\int_{\mathbb{R}^d\times\mathbb{R}^d} K\Big(\frac{u}{h}\Big)K\Big(\frac{v}{h}\Big)g_{(U,V)}(u,v)\,\mathrm{d}u\,\mathrm{d}v\,,$$

and then the Hölder inequality yields

$$\Big|\mathrm{Cov}(K_h(U),K_h(V))\Big| \leq \delta_p\,\|K\|_q^2\,h^{-2d/p}.$$

The second part of the inequality is a straightforward application of an inequality from Billingsley (Billingsley (1986), p. 376).

Proof of (ii) We have

$$\Big|\mathrm{Cov}(K_h(x-U),K_h(x-V))\Big| \leq \frac{1}{h^{2d}}\int_{\mathbb{R}^d\times\mathbb{R}^d} \Big|K\Big(\frac{u}{h}\Big)K\Big(\frac{v}{h}\Big)g_{(U,V)}(x-u,x-v)\Big|\,\mathrm{d}u\,\mathrm{d}v\,.$$

Thus

$$\int_{\mathbb{R}^d} \Big|\mathrm{Cov}(K_h(x-U),K_h(x-V))\Big|\,\mathrm{d}x \leq \|\psi_{(U,V)}\|_\infty\,\|K\|_1^2.$$

For the second part, we use a covariance inequality from Delyon (1990) as made explicit by Viennet (1997) (see also Rio (2000), p. 29). One has

$$\Big|\mathrm{Cov}(K_h(x-U),K_h(x-V))\Big| \leq \frac{1}{h^{2d}}\int_{\mathbb{R}^d} (b_1(u)+b_2(u))K^2\Big(\frac{x-u}{h}\Big)f(u)\,\mathrm{d}u\,,$$

where f is the common density of U and V and b_1 and b_2 are two measurable functions such that

$$\int_{\mathbb{R}^d} b_1(u)f(u)\,\mathrm{d}u = \int_{\mathbb{R}^d} b_2(u)f(u)\,\mathrm{d}u = \beta.$$

Consequently

$$\int_{\mathbb{R}^d} \Big|\mathrm{Cov}(K_h(x-U), K_h(x-V))\Big|\,\mathrm{d}x \le 2\beta\,\|K\|_2^2\,h^{-d}.$$

5.2 Proof of Theorem 5

Proof of (i) Given a kernel K and a smoothing parameter h, we set $K_h(.) = K(./h)/h^d$. For $x \in \mathbb{R}^d$, the following decomposition is valid:

$$\begin{aligned}\mathbf{E}(f_{\mathbf{n}}(x) - f(x))^2 &= (\mathbf{E}f_{\mathbf{n}}(x) - f(x))^2 + \frac{1}{\hat{\mathbf{n}}}\mathrm{Var}\ K_{h_{\mathbf{n}}}(x - X_{\mathbf{1}}) \\ &\quad + \frac{1}{\hat{\mathbf{n}}^2}\sum_{\substack{(\mathbf{i},\mathbf{j})\in\mathcal{I}_{\mathbf{n}}\times\mathcal{I}_{\mathbf{n}}\\ \mathbf{i}\neq\mathbf{j}}} \mathrm{Cov}(K_{h_{\mathbf{n}}}(x - X_{\mathbf{i}}), K_{h_{\mathbf{n}}}(x - X_{\mathbf{j}})) \\ &\stackrel{\text{def}}{=} B_{\mathbf{n}}^2(x) + \tilde{V}f_{\mathbf{n}}(x) + C_{\mathbf{n}}(x). \end{aligned} \tag{3}$$

Note that $\tilde{V}f_{\mathbf{n}}(x)$ is nothing else but the variance of $f_{\mathbf{n}}(x)$ in the *i.i.d.* case. Given **(A1)**, **(A2)** and **(A3)** it is easy to check (see *e.g.* Bosq and Lecoutre (1987)) that

$$B_{\mathbf{n}}^2(x) = \frac{h_{\mathbf{n}}^4}{4}\,\Delta_d^2(x) + \mathrm{o}(h_{\mathbf{n}}^4)$$

and

$$\tilde{V}f_{\mathbf{n}}(x) = \frac{f(x)}{\hat{\mathbf{n}}h_{\mathbf{n}}^d}\int_{\mathbb{R}^d} K^2(u)\,\mathrm{d}u + \mathrm{o}\Big(\frac{1}{\hat{\mathbf{n}}h_{\mathbf{n}}^d}\Big).$$

The covariance term $C_{\mathbf{n}}(x)$ remains to be studied. Let us show that $\hat{\mathbf{n}}h_{\mathbf{n}}^dC_{\mathbf{n}}(x) = \mathrm{o}(1)$. Assume first that **(A4)** is satisfied, and define the sets $\mathcal{S}_1$ and $\mathcal{S}_2$ by

$$\mathcal{S}_1 = \{(\mathbf{i},\mathbf{j}) \in \mathcal{I}_{\mathbf{n}} \times \mathcal{I}_{\mathbf{n}} : 0 < \|\mathbf{i}-\mathbf{j}\|_1 \le \gamma_{\mathbf{n}}\}$$

and

$$\mathcal{S}_2 = \{(\mathbf{i},\mathbf{j}) \in \mathcal{I}_{\mathbf{n}} \times \mathcal{I}_{\mathbf{n}} : \|\mathbf{i}-\mathbf{j}\|_1 > \gamma_{\mathbf{n}}\},$$

where $\gamma_{\mathbf{n}} = \lfloor h_{\mathbf{n}}^{-2d/(qr)}\rfloor$. The notation $\lfloor . \rfloor$ stands for the integer part function and $\|.\|_1$ stands for the L^1 norm, *i.e.*, $\|\mathbf{n}\|_1 = |n_1| + \ldots + |n_N|$ for $\mathbf{n} = (n_1, \ldots, n_N) \in \mathbb{Z}^N$. We have

$$
\begin{aligned}
|C_{\mathbf{n}}(x)| \;\le\; & \frac{1}{\hat{\mathbf{n}}^2} \sum_{(\mathbf{i},\mathbf{j})\in\mathcal{S}_1} \Big|\mathrm{Cov}(K_{h_{\mathbf{n}}}(x-X_{\mathbf{i}}), K_{h_{\mathbf{n}}}(x-X_{\mathbf{j}}))\Big| \\
& + \frac{1}{\hat{\mathbf{n}}^2} \sum_{(\mathbf{i},\mathbf{j})\in\mathcal{S}_2} \Big|\mathrm{Cov}(K_{h_{\mathbf{n}}}(x-X_{\mathbf{i}}), K_{h_{\mathbf{n}}}(x-X_{\mathbf{j}}))\Big|.
\end{aligned}
$$

Thus, Lemma 1 together with **(A5)** yield

$$
|C_{\mathbf{n}}(x)| \le \frac{D}{\hat{\mathbf{n}}^2} \, \|K\|_q^2 \, h_{\mathbf{n}}^{-2d/p} \,\mathrm{Card}(\mathcal{S}_1) + \frac{4R}{\hat{\mathbf{n}}^2} \, \|K\|_\infty^2 \, h_{\mathbf{n}}^{-2d} \sum_{(\mathbf{i},\mathbf{j})\in\mathcal{S}_2} \lfloor \|\mathbf{i}-\mathbf{j}\| \rfloor^{-r}.
$$

A close examination shows that $\mathrm{Card}(\mathcal{S}_1) \le 2^N \hat{\mathbf{n}} \gamma_{\mathbf{n}}^N$. Moreover

$$
\begin{aligned}
\sum_{(\mathbf{i},\mathbf{j})\in\mathcal{S}_2} \lfloor \|\mathbf{i}-\mathbf{j}\| \rfloor^{-r} \le C_1 \hat{\mathbf{n}} \sum_{\substack{\mathbf{i}\in\mathbb{Z}^N \\ \|\mathbf{i}\|_1 > \gamma_{\mathbf{n}}}} \|\mathbf{i}\|_1^{-r} \le C_2 \hat{\mathbf{n}} \sum_{k>\gamma_{\mathbf{n}}} \frac{k^{N-1}}{k^r} & \le C_2 \hat{\mathbf{n}} \int_{\gamma_{\mathbf{n}}}^{+\infty} \frac{t^{N-1}}{t^r} \,\mathrm{d}t \\
& = \frac{C_2 \hat{\mathbf{n}}}{r-N} \, \gamma_{\mathbf{n}}^{N-r},
\end{aligned}
$$

where C_1 and C_2 are positive constant. Putting all pieces together, we obtain

$$
\hat{\mathbf{n}} h_{\mathbf{n}}^d C_{\mathbf{n}}(x) = \mathrm{O}\Big(h_{\mathbf{n}}^{\frac{2d}{qr}[r(1-q/2)-N]} \Big), \tag{4}
$$

and the bound vanishes at infinity since $r > 2N(p-1)/(p-2) = 2N/(2-q) > N$.

When **(A'4)** is satisfied, the proof is similar. The only difference lies in the estimation of the covariance term. We may use Lemma 3 and write

$$
\begin{aligned}
& \Big|\mathrm{Cov}(K_{h_{\mathbf{n}}}(x-X_{\mathbf{i}}), K_{h_{\mathbf{n}}}(x-X_{\mathbf{j}}))\Big| \\
& \quad \le \frac{1}{h_{\mathbf{n}}^{2d}} \int_{\mathbb{R}^d\times\mathbb{R}^d} \Big| K\Big(\frac{x-u}{h_{\mathbf{n}}}\Big) K\Big(\frac{x-v}{h_{\mathbf{n}}}\Big) g_{(X_{\mathbf{i}},X_{\mathbf{j}})}(u,v) \Big| \,\mathrm{d}u\,\mathrm{d}v \\
& \quad \le \gamma(d,L) \, \|K\|_1^2 \, \alpha_{\lfloor \|\mathbf{i}-\mathbf{j}\| \rfloor}^{1/(2d+1)}.
\end{aligned}
$$

Consequently, choosing $\gamma_{\mathbf{n}} = \lfloor h_{\mathbf{n}}^{-(2d+1)/r} \rfloor$, we obtain as above

$$
\begin{aligned}
|C_{\mathbf{n}}(x)| \;\le\; & \frac{C_1 R \gamma(d,L)}{\hat{\mathbf{n}}} \, \|K\|_1^2 \sum_{\substack{\mathbf{i}\in\mathbb{Z}^N \\ 0<\|\mathbf{i}\|_1 \le \gamma_{\mathbf{n}}}} \|\mathbf{i}\|_1^{-r/(2d+1)} \\
& + \frac{4C_2 R}{\hat{\mathbf{n}}} \, \|K\|_\infty^2 \, h_{\mathbf{n}}^{-2d} \sum_{\substack{\mathbf{i}\in\mathbb{Z}^N \\ \|\mathbf{i}\|_1 > \gamma_{\mathbf{n}}}} \|\mathbf{i}\|_1^{-r},
\end{aligned}
$$

where C_1 and C_2 are positive constants. Therefore,

$$\begin{aligned}
|C_{\mathbf{n}}(x)| &\le \frac{C_3 R\gamma(d,L)}{\hat{\mathbf{n}}}\,\|K\|_1^2 \sum_{0<k\le\gamma_{\mathbf{n}}} \frac{k^{N-1}}{k^{r/(2d+1)}} + \frac{4C_4R}{\hat{\mathbf{n}}}\,\|K\|_\infty^2\, h_{\mathbf{n}}^{-2d} \sum_{k>\gamma_{\mathbf{n}}} \frac{k^{N-1}}{k^r} \\
&\le \frac{C_3 R\gamma(d,L)}{\hat{\mathbf{n}}}\,\|K\|_1^2 \Big(1 + \int_2^{\gamma_{\mathbf{n}}} \frac{t^{N-1}}{t^{r/(2d+1)}}\,\mathrm{d}t\Big) \\
&\quad + \frac{4C_4R}{\hat{\mathbf{n}}}\,\|K\|_\infty^2\, h_{\mathbf{n}}^{-2d} \int_{\gamma_{\mathbf{n}}}^{+\infty} \frac{t^{N-1}}{t^r}\,\mathrm{d}t\,,
\end{aligned}$$

where C_3 and C_4 are positive constants. If $r \ne N(2d+1)$, $\hat{\mathbf{n}} h_{\mathbf{n}}^d C_{\mathbf{n}}(x)$ is

$$\mathrm{O}(h_{\mathbf{n}}^d) + \mathrm{O}\Big(h_{\mathbf{n}}^{(d+1)-\frac{N(2d+1)}{r}}\Big), \tag{5}$$

hence $\hat{\mathbf{n}} h_{\mathbf{n}}^d C_{\mathbf{n}}(x) \to 0$ since $r > N(2d+1)/(d+1) > N$. If $r = N(2d+1)$, $\hat{\mathbf{n}} h_{\mathbf{n}}^d C_{\mathbf{n}}(x)$ is $\mathrm{O}(h_{\mathbf{n}}^d \log 1/h_{\mathbf{n}}^{1/N})$ and the proof of **(i)** is complete.

Proof of (ii) The optimal rate $\hat{\mathbf{n}}^{4/(d+4)}$ is obtained by balancing the bias and variance terms obtained in **(i)**.

Proof of Corollary 6 Under our assumptions, it is easy to show that one can obtain uniform bounds for $B_{\mathbf{n}}^2(x)$, $\tilde{V}_{\mathbf{n}}(x)$ (see Bosq and Lecoutre (1987), p. 76) and $C_{\mathbf{n}}(x)$ (see (4)-(5)). Details are omitted since they are obvious.

5.3 Proof of Theorem 7

Proof of (i) According to decomposition (3), one has

$$\int_{\mathbb{R}^d} \mathbf{E}(f_{\mathbf{n}}(x) - f(x))^2\,\mathrm{d}x = \int_{\mathbb{R}^d} B_{\mathbf{n}}^2(x)\,\mathrm{d}x + \int_{\mathbb{R}^d} \tilde{V} f_{\mathbf{n}}(x)\,\mathrm{d}x + \int_{\mathbb{R}^d} C_{\mathbf{n}}(x)\,\mathrm{d}x\,.$$

It is easy to check, under **(A1)**, **(A2)** and **(A3)** (see *e.g.* Bosq and Lecoutre (1987)), that

$$\int_{\mathbb{R}^d} B_{\mathbf{n}}^2(x)\,\mathrm{d}x = \frac{h_{\mathbf{n}}^4}{4}\,\Gamma_d^2 + \mathrm{o}(h_{\mathbf{n}}^4)$$

and

$$\int_{\mathbb{R}^d} \tilde{V} f_{\mathbf{n}}(x)\,\mathrm{d}x = \frac{\int_{\mathbb{R}^d} K^2(u)\,\mathrm{d}u}{\hat{\mathbf{n}} h_{\mathbf{n}}^d} + \mathrm{o}\Big(\frac{1}{\hat{\mathbf{n}} h_{\mathbf{n}}^d}\Big).$$

Thus, it suffices to verify that $\hat{\mathbf{n}} h_{\mathbf{n}}^d \int\limits_{\mathbb{R}^d} C_{\mathbf{n}}(x)\,\mathrm{d}x = \mathrm{o}(1)$. As in the previous paragraph, define the sets $\mathcal{S}_1$ and $\mathcal{S}_2$ by

$$\mathcal{S}_1 = \{(\mathbf{i},\mathbf{j}) \in \mathcal{I}_{\mathbf{n}} \times \mathcal{I}_{\mathbf{n}} : 0 < \|\mathbf{i}-\mathbf{j}\|_1 \leq \gamma_{\mathbf{n}}\}$$

and

$$\mathcal{S}_2 = \{(\mathbf{i},\mathbf{j}) \in \mathcal{I}_{\mathbf{n}} \times \mathcal{I}_{\mathbf{n}} : \|\mathbf{i}-\mathbf{j}\|_1 > \gamma_{\mathbf{n}}\},$$

where $\gamma_{\mathbf{n}} = \lfloor h_{\mathbf{n}}^{-d/(2N)} \rfloor$. We have

$$\begin{aligned} |C_{\mathbf{n}}(x)| \;\leq\; & \frac{1}{\hat{\mathbf{n}}^2} \sum_{(\mathbf{i},\mathbf{j})\in\mathcal{S}_1} \Big|\mathrm{Cov}(K_{h_{\mathbf{n}}}(x-X_{\mathbf{i}}), K_{h_{\mathbf{n}}}(x-X_{\mathbf{j}}))\Big| \\ & + \frac{1}{\hat{\mathbf{n}}^2} \sum_{(\mathbf{i},\mathbf{j})\in\mathcal{S}_2} \Big|\mathrm{Cov}(K_{h_{\mathbf{n}}}(x-X_{\mathbf{i}}), K_{h_{\mathbf{n}}}(x-X_{\mathbf{j}}))\Big|. \end{aligned}$$

If **(A4)** is satisfied, Lemma 1 yields

$$\Big|\int\limits_{\mathbb{R}^d} C_{\mathbf{n}}(x)\,\mathrm{d}x\Big| \leq \frac{D}{\hat{\mathbf{n}}^2}\,\|K\|_1^2\,\mathrm{Card}(\mathcal{S}_1) + \frac{2}{\hat{\mathbf{n}}^2}\,\|K\|_2^2\,h_{\mathbf{n}}^{-d} \sum_{(\mathbf{i},\mathbf{j})\in\mathcal{S}_2} \beta_{\lfloor\|\mathbf{i}-\mathbf{j}\|\rfloor}.$$

Since $\mathrm{Card}(\mathcal{S}_1) \leq 2^N \hat{\mathbf{n}} \gamma_{\mathbf{n}}^N$, we are led to

$$\Big|\int\limits_{\mathbb{R}^d} C_{\mathbf{n}}(x)\,\mathrm{d}x\Big| \leq \frac{2^N D}{\hat{\mathbf{n}}}\,\|K\|_1^2\,\gamma_{\mathbf{n}}^N + \frac{2C_1}{\hat{\mathbf{n}}}\|K\|_2^2\,h_{\mathbf{n}}^{-d} \sum_{k\geq\lfloor C_2\gamma_{\mathbf{n}}\rfloor} k^{N-1}\beta_k,$$

where C_1 and C_2 are positive constants. Therefore, under assumption **(A5)**, we obtain $\hat{\mathbf{n}} h_{\mathbf{n}}^d \int\limits_{\mathbb{R}^d} C_{\mathbf{n}}(x)\,\mathrm{d}x = \mathrm{o}(1)$. This completes the proof of **(i)**.

Proof of (ii) As in Theorem 5, the optimal rate $\hat{\mathbf{n}}^{4/(d+4)}$ is obtained by balancing the bias and variance terms obtained in **(i)**.

Proof of Corollary 8 is similar to that of Corollary 6.

Acknowledgments

I would like to express my gratitude to the anonymous referee for his/her careful reading of the paper and valuable suggestions. I am indebted to Denis Bosq for many stimulating conversations on nonparametric estimation.

References

Billingsley, P., 1986. Probability and Measure – 2nd Edition. Wiley, New York.

Bosq, D., 1995. Optimal asymptotic quadratic error of density estimators for strong mixing or chaotic data. Statistics and Probability Letters 22, 339–347.

Bosq, D., 1998. Nonparametric Statistics for Stochastic Processes – Estimation and Prediction – 2nd Edition. Lecture Notes in Statistics, Springer–Verlag, New York.

Bosq, D., Lecoutre, J.P., 1987. Théorie de l'Estimation Fonctionnelle. Economica, Paris.

Carbon, M., Hallin, M., Tran, L.T., 1996. Kernel density estimation for random fields: the L_1 theory. Nonparametric Statistics 6, 157–170.

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

Delyon, B., 1990. Limit theorems for mixing processes. Technical Report, IRISA, Rennes I 546.

Devroye, L., 1987. A Course in Density Estimation. Birkhäuser, Boston.

Doukhan, P., 1994. Mixing – Properties and Examples. Lecture Notes in Statistics, Springer–Verlag, New York.

Guyon, X., 1995. Random Fields on a Network – Modeling, Statistics, and Applications. Springer–Verlag, New York.

Parzen, E., 1962. On estimation of a probability density function and mode. The Annals of Mathematical Statistics 33, 1065–1076.

Rio, E., 2000. Théorie Asymptotique des Processus Aléatoires Faiblement Dépendants. Springer–Verlag, Berlin.

Rosenblatt, M., 1956. Remarks on some nonparametric estimates of a density function. The Annals of Mathematical Statistics 27, 832–837.

Tran, L.T., 1990. Kernel density estimation on random fields. Journal of Multivariate Analysis 34, 37–53.

Tran, L.T., Yakowitz, S., 1993. Nearest neighbor estimators for random fields. Journal of Multivariate Analysis 44, 23–46.

Viennet, G., 1997. Inequalities for absolutely regular sequences: application to density estimation. Probability Theory and Related Fields 107, 467–492.