Spatial Kernel Density Estimation

Gérard BIAU

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

biau@ccr.jussieu.fr

Abstract

We study kernel estimation of the density of a multidimensional spatial process $(Z_{\mathbf{x}}, \mathbf{x} \in \mathbb{R}^N)$, where the spatial index $\mathbf{x}$ varies continuously throughout the space $\mathbb{R}^N$ $(N \geq 1)$. The loss between the estimator and the unknown density is measured by means of mean square error. Under appropriate conditions, we prove the existence of optimal and parametric rates, provided the process is observed on an hyperrectangle of $\mathbb{R}^N$. Since such an observation is impossible in practice, it is shown that the parametric rate is preserved by using a suitable spatial sampling scheme. Our results are a natural extension of the existing results in nonparametric density estimation for continuous time series $(N = 1)$.

Index Terms **— Kernel density estimation, Spatial process, Mixing, Rate of convergence, Sampling.**

***2000 Mathematics Subject Classification*: 62G07.**

1 Introduction

In recent years, spatial models have attracted a lot of attention in the statistical literature (see Ripley [28], Cressie [17] or Guyon [20] for exposition, methods and applications). Geology, soil science, image processing, epidemiology, crop science, ecology, forestry, astronomy, atmospheric science, or simply any discipline that works with data collected from different spatial locations, need to develop models taking into account dependence between measurements at different positions. When dealing with data where space dependencies are likely, the models need to be more flexible than their temporal counterparts, because past, present, and future have no analogy in space. Thus, stimulated by a growing demand from applied research, the theory of spatial models, or *random fields*, has developed significantly over the last few years.

Parallel to this development, the existence of high speed, inexpensive computing methods 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 (see Bosq [10] for a complete survey), where the unidirectional flow of time underlies the construction of the models, and relatively few generalizations to the spatial domain are available. Key references on this latter topic are Tran [30], Tran and Yakowitz [31], Carbon, Hallin and Tran [13], Carbon, Tran and Wu [14], Hallin [21], [22], and Biau [2], who have investigated nonparametric density estimation for random fields indexed by $\mathbb{Z}^N$, with $N \geq 1$. Roughly speaking, data sets whose spatial locations are regular lattices in $\mathbb{R}^N$ are the closest analogue to time series observed at equally spaced time points. Such data sets occur frequently in practice, either because the process under study is essentially discrete (think to examples in forestry) or just because it is the result of a sampling scheme. To illustrate the latter point, think to satellites orbiting the Earth. By various sampling and integration methods, the Earth's surface is divided into small rectangles (*e.g.* 56 m $\times$ 56 m) called *pixels* (short for *picture elements*). An agriculture scene of interest (around, say, 34,000 km^2) has certain proportions devoted to wheat, corn, soybeans, and so forth that need to be estimated. These various crops have their reflectance properties that, together with noise, are remotely sensored. Thus the data are received as a regular lattice in $\mathbb{R}^2$ and are identified with the centers of their respective pixels.

On the other hand, there are some physical issues where it is simply not reasonable to assume that the stochastic process under study is varying in space in a discrete way. This is typically the case in oil engineering, where one is interested in reserves that are continuously distributed into three-dimensional pockets. Such phenomena are basically studied by *Geostatistics*, where data are thought of as realizations of a random process $(Z_{\mathbf{x}}, \mathbf{x} \in \mathcal{D})$, $\mathcal{D}$ being a fixed subset of $\mathbb{R}^N$ with positive volume. The geostatistical methods have found favor among soil scientists who seek to map soil properties of a field from a small number of soil samples at known locations throughout the field (see Wackernagel [32] for an introduction). To our knowledge, no theoretical work has been done so far to study such continuously indexed spatial fields using nonparametric techniques. As a first modest step towards this direction, we will be concerned in this paper with kernel estimation of the density of a *spatial process* $(Z_{\mathbf{x}}, \mathbf{x} \in \mathbb{R}^N)$ taking values in $\mathbb{R}^d$ $(N, d \geq 1)$. Notations and assumptions will be specified in Section 2. In other words, we let the spatial index $\mathbf{x}$ vary *continuously* throughout the space $\mathbb{R}^N$. We sense a practical need for nonparametric spatial estimation for situations in which parametric families cannot be adopted with confidence. In the time domain $(N = 1)$, nonparametric estimation of the density f of a continuous time process $(Z_t, t \in \mathbb{R})$ observed over the time interval $[0, T]$ has been considered by many authors. Under some mixing conditions and assuming that f belongs to some appropriate Hölder class, the mean square error of a kernel estimator of f turns out to be $\mathrm{O}\big(T^{-4/(4+d)}\big)$. This rate is usually called *optimal*. However, the optimal rate can be sharpened provided that $(Z_t, t \in \mathbb{R})$ should satisfy a *local irregularity condition*, which basically provides additional information to the statistician. In that case, the kernel estimator reaches the *parametric* (or *superoptimal*) rate $1/T$. This rate is called parametric since it corresponds to the rate of the MLE in classical parametric continuous time models such as diffusion processes. The first parametric rate was pointed by Castellana and Leadbetter [15], who noticed that the variance of a density estimator is $\mathrm{O}(1/T)$ for a real "locally irregular" process observed over the time interval $[0, T]$. For other results of this kind, we refer the reader to Banon [1], Delecroix [18], Cheze-Payaud [16], Blanke [4], Bosq [8], [9], [10], Kutoyants [23], Leblanc [24], and Bosq, Merlevède and Peligrad [12], among others.

In the sequel, we show that the existence of optimal and parametric rates transfer, under appropriate assumptions, to spatial processes. This will be the subject of Section 3 (optimal rates) and Section 4 (superoptimal rates). One of the basic assumptions used in all the paper is that the process is completely observable over any given hyperrectangle $[0, T_1] \times \ldots \times [0, T_N]$. Kernel estimators of the density based on the complete paths are constructed. In practice, however, it is clearly impossible to observe the spatial process continuously for various reasons, in particular, precision of measuring instruments, unavailability of observations at all spatial locations as in mining, etc. In other words, the problem of density estimation for *sampled data* of the field is of extreme interest in practical issues. This problem will be discussed in Section 5. Proofs are gathered in Section 6.

2 Notations and assumptions

Spatial density estimation. We let $(Z_{\mathbf{x}}, \mathbf{x} \in \mathbb{R}^N)$ be a measurable $\mathbb{R}^d$-valued spatial process $(N, d \geq 1)$ defined on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$. We suppose that the $Z_{\mathbf{x}}$'s have a common distribution with an unknown density f with respect to the Lebesgue measure λ over $\mathbb{R}^d$. Throughout the paper, the symbol $\|.\|$ will denote any norm over $\mathbb{R}^N$ and any mathematical symbol referring to a spatial location $\mathbf{x} = (x_1, \ldots, x_N) \in \mathbb{R}^N$ (such as "d$\mathbf{x}$" in the integrals) will be in bold print.

Concerning the properties of f we introduce the space $\mathcal{C}_r^d(l)$ $(r = k + c$, $k \in \mathbb{N}^*$, $0 < c \leq 1)$ of real valued functions, defined on $\mathbb{R}^d$, which are k times continuously differentiable and such that

$$\left| \frac{\partial f^{(k)}}{\partial z_1^{j_1} \ldots \partial z_d^{j_d}}(z') - \frac{\partial f^{(k)}}{\partial z_1^{j_1} \ldots \partial z_d^{j_d}}(z) \right| \leq l \, \|z' - z\|^c \,, \tag{2.1}$$

$z, z' \in \mathbb{R}^d$; $j_1 + \ldots + j_d = k$. Here and subsequently, $\|.\|$ denotes any norm over $\mathbb{R}^d$ (possibly different from the norm over $\mathbb{R}^N$). For $\mathbf{T} = (T_1, \ldots, T_N)$ in $\mathbb{R}_+^{*N}$, we define the rectangular region $\mathcal{I}_{\mathbf{T}}$ by $\mathcal{I}_{\mathbf{T}} = \{\mathbf{x} \in \mathbb{R}_+^N \, : \, 0 \leq x_i \leq T_i, \, i = 1, \ldots, N\}$. We write $\mathbf{T} \to +\infty$ if $\min_{i=1,\ldots,N} T_i \to +\infty$ and we set $\hat{\mathbf{T}} = T_1 \ldots T_N$. Our goal is to estimate $f \in \mathcal{C}_r^d(l)$ when a continuous spatial sampling on $\mathcal{I}_{\mathbf{T}}$ is available.

The kernel estimator. The kernel estimator based on $(Z_{\mathbf{x}}, \mathbf{x} \in \mathcal{I}_{\mathbf{T}})$ is defined as

$$f_{\mathbf{T}}(z) = \frac{1}{\hat{\mathbf{T}} h_{\mathbf{T}}^d} \int_{\mathcal{I}_{\mathbf{T}}} K\left(\frac{z - Z_{\mathbf{x}}}{h_{\mathbf{T}}}\right) \mathrm{d}\mathbf{x} \,, \quad z \in \mathbb{R}^d,$$

where $\lim_{\mathbf{T} \to +\infty} h_{\mathbf{T}} = 0(+)$. The *kernel* $K : \mathbb{R}^d \to \mathbb{R}$ is a *bounded integrable* function such that

$$\int_{\mathbb{R}^d} K(u) \mathrm{d}u = 1 \quad \text{and} \quad \lim_{\|u\| \to +\infty} \|u\|^d K(u) = 0.$$

In some situations, we will assume that K belongs to $\mathcal{H}(k, c)$ $(k \in \mathbb{N}^*, 0 < c \leq 1)$ *i.e.*, that it satisfies the additional conditions

$$\int_{\mathbb{R}^d} \|u\|^c \, |u_1|^{\alpha_1} \ldots |u_d|^{\alpha_d} \big|K(u)\big| \mathrm{d}u < +\infty$$

$$\text{and} \tag{2.2}$$

$$\int_{\mathbb{R}^d} u_1^{\alpha_1} \ldots u_d^{\alpha_d} K(u) \mathrm{d}u = 0 \,,$$

$\alpha_1, \dots, \alpha_d \in \mathbb{N}$; $\alpha_1 + \ldots + \alpha_d = j$, $1 \le j \le k$. Such a kernel is usually called a kernel of order $(k,\, c)$. Note that traditional kernels are of order $(1, 1)$.

Mixing coefficients and inequalities. The spatial dependence of the process will be measured by means of α*-mixing*. We let the α-mixing coefficients of the field $(Z_{\mathbf{x}}, \mathbf{x} \in \mathbb{R}^N)$ be defined as

$$\alpha(u) = \sup_{\substack{\mathbf{x},\mathbf{x}' \in \mathbb{R}^N \\ \|\mathbf{x}-\mathbf{x}'\|=u}} \alpha\big(\sigma(Z_{\mathbf{x}}), \sigma(Z_{\mathbf{x}'})\big), \quad u \ge 0,$$

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

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

For a complete discussion on mixing and examples, we refer the reader to Doukhan [19], Rio [27] and the references therein.

Consider now two random variables U and V which take their values in $\mathbb{R}^d$ equipped with its Borel σ-field. Set $g_{(U,V)} = f_{(U,V)} - f_U \otimes f_V$, where $f_{(U,V)}$ (*resp.*, f_U and f_V) denotes the density of the pair (U, V) (*resp.*, of U and V) with respect to the Lebesgue measure. The L^p norm in $\mathbb{R}^d$ is denoted by $\|.\|_p$, $1 \le p \le +\infty$. The following inequality, which is a straightforward application of Hölder's inequality and an inequality from Billingsley ([3], page 376) will be useful for our purpose.

Lemma 2.1 *Let K be a bounded integrable real mapping defined on $\mathbb{R}^d$ and let h be a positive real number. Let $K_h(.) = K(./h)/h^d$. Suppose that $\delta_p = \|g_{(U,V)}\|_p < +\infty$ for some $p \in [1, +\infty]$. Then, with $\alpha = \alpha\big(\sigma(U), \sigma(V)\big)$,*

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

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

3 Optimal quadratic error

In the current section, we will assume that the $Z_{\mathbf{x}}$'s have the same distribution but not a stronger condition like stationarity or isotropy. Under mild mixing conditions, we show that the kernel estimator has at less the same rate as in the *i.i.d.* case. This rate will be called *optimal*. Our results extend and complete the results of Bosq [9] who studied the time domain $(N = 1)$. Let us first state our main assumptions, which are the spatial counterparts of the conditions listed in Bosq [9].

- $A_1(\Gamma, p)$ – There exists a Borel set Γ of $\mathbb{R}^N \times \mathbb{R}^N$ containing
$$\mathcal{D}_N = \big\{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^N \times \mathbb{R}^N \ : \ \mathbf{x} = \mathbf{y}\big\}$$
and $p \in]2, +\infty]$ such that
  (a) $g_{\mathbf{x},\mathbf{y}} = f_{(Z_\mathbf{x}, Z_\mathbf{y})} - f \otimes f$ exists for $(\mathbf{x}, \mathbf{y}) \notin \Gamma$, where $f_{(Z_\mathbf{x}, Z_\mathbf{y})}$ denotes the density of $(Z_\mathbf{x}, Z_\mathbf{y})$;
  (b) $\delta_p(\Gamma) = \sup_{(\mathbf{x},\mathbf{y}) \notin \Gamma} \|g_{\mathbf{x},\mathbf{y}}\|_p < +\infty$;
  (c) There exists a positive constant l_Γ such that
$$\limsup_{\hat{\mathbf{T}} \to +\infty} \frac{1}{\hat{\mathbf{T}}} \int_{(\mathcal{I}_\mathbf{T} \times \mathcal{I}_\mathbf{T}) \cap \Gamma} \mathrm{d}\mathbf{x} \mathrm{d}\mathbf{y} = l_\Gamma < +\infty \,.$$
- $A_2(\gamma, \beta)$ – $\alpha\big(\|\mathbf{x} - \mathbf{y}\|\big) \leq \gamma \, \|\mathbf{x} - \mathbf{y}\|^{-\beta}$ for $(\mathbf{x}, \mathbf{y}) \notin \Gamma$, where $\gamma > 0$ and $\beta > 0$.

$A_2(\gamma, \beta)$ is a mild assumption which is satisfied by a large class of processes. For example, stationary diffusion processes satisfy this condition for each positive beta (see Doukhan [19], pages 113-119). With respect to $A_1(\Gamma, p)$, define the set Γ_ε for every $\varepsilon > 0$ as
$$\Gamma_\varepsilon = \big\{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^N \times \mathbb{R}^N \ : \ \|\mathbf{x} - \mathbf{y}\| < \varepsilon\big\}.$$
Clearly, (c) holds with $l_{\Gamma_\varepsilon} \leq V(\mathcal{B}_N)\, \varepsilon^N$, where $V(\mathcal{B}_N)$ denotes the volume of the *unit ball* $\mathcal{B}_N$ in $\mathbb{R}^N$, defined by $\mathcal{B}_N = \{\mathbf{x} \in \mathbb{R}^N : \|\mathbf{x}\| \leq 1\}$. Thus, if there exists $\varepsilon > 0$ such that $f_{(Z_\mathbf{x}, Z_\mathbf{y})}$ exists for $(\mathbf{x}, \mathbf{y}) \notin \Gamma_\varepsilon$ and if $\|f\|_\infty < +\infty$ and $\sup_{(\mathbf{x},\mathbf{y}) \notin \Gamma_\varepsilon} \|f_{(Z_\mathbf{x}, Z_\mathbf{y})}\|_\infty < +\infty$ then Z satisfies $A_1(\Gamma_\varepsilon, +\infty)$.

We are now in a position to state the main result of this section.

Theorem 3.1 (OPTIMAL RATE) *Assume that $A_1(\Gamma, p)$ and $A_2(\gamma, \beta)$ hold for some Γ, p, γ, β with $\beta > 2N(p-1)/(p-2)$ and that f is bounded. If in addition $f \in \mathcal{C}_r^d(l)$, $K \in \mathcal{H}_{k,c}$ $(r = k + c)$ and $h_\mathbf{T} \sim \hat{\mathbf{T}}^{-1/(2r+d)}$, then*
$$\limsup_{\mathbf{T} \to +\infty} \sup_{z \in \mathbb{R}^d} \hat{\mathbf{T}}^{2r/(2r+d)} \mathbf{E}\big(f_\mathbf{T}(z) - f(z)\big)^2 \leq C \,, \tag{3.1}$$
where C is a positive constant.

Let us note that if $\beta = 2N(p-1)/(p-2)$ (in particular if $p = +\infty$ and $\beta = 2N$) the same rates are valid but with a constant greater than C.

In order to show that the above rates are achieved for some processes, let us consider the family $\mathcal{Z}$ of spatial processes $Z = (Z_\mathbf{x}, \mathbf{x} \in \mathbb{R}^N)$ which satisfy the above hypotheses uniformly, in the following sense: there exist positive constants $f_0, L_0, \delta_0, \gamma_0, \beta_0$ and p_0 such that, for each $Z \in \mathcal{Z}$,

- $\|f_Z\|_\infty \le f_0$;
- $\limsup_{\mathbf{T}\to+\infty} \frac{1}{\hat{\mathbf{T}}} \int_{(\mathcal{I}_\mathbf{T}\times\mathcal{I}_\mathbf{T})\cap\Gamma_Z} \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} \le L_0$;
- $p_Z = p_0 \in]2,+\infty]$ and $\delta_{p_0}(\Gamma_Z) \le \delta_0$;
- $\gamma_Z \le \gamma_0$ and $\beta_Z \ge \beta_0 > 2N\,\frac{p_0-1}{p_0-2}$,

where $f_Z, \Gamma_Z, p_Z, \gamma_Z, \beta_Z$ stand for $f, \Gamma, p, \gamma, \beta$ associated with Z. Then we have

Corollary 3.1 *Assume that $T_i h_\mathbf{T}^d \to +\infty$ for $i = 1, \ldots, N$. Let $\mathcal{Z}_1 = \{Z : Z \in \mathcal{Z}, f_Z \in C_r^d(l)\}$ $(r = k + c)$. If $K \in \mathcal{H}_{k,c}$, the choice $h_\mathbf{T} \sim \hat{\mathbf{T}}^{-1/(2r+d)}$ implies*

$$\limsup_{\mathbf{T}\to+\infty} \sup_{Z\in\mathcal{Z}_1} \sup_{z\in\mathbb{R}^d} \hat{\mathbf{T}}^{2r/(2r+d)} \mathbf{E}_Z\big(f_\mathbf{T}(z) - f(z)\big)^2 \le C\,,$$

where C is a positive constant.

The next theorem provides a *minimax* result, showing that the optimal rate $\hat{\mathbf{T}}^{-2r/(2r+d)}$ is minimax over the class of spatial processes belonging to $\mathcal{Z}$ with a density in $\mathcal{C}_r^d(l)$. Moreover, according to Corollary 3.1, the kernel estimator based on observations on $\mathcal{I}_\mathbf{T}$ achieves this rate.

Theorem 3.2 *Let $\mathcal{Z}_1 = \{Z : Z \in \mathcal{Z}, f_Z \in C_r^d(l)\}$ and $K \in \mathcal{H}_{k,c}$ $(r = k+c)$. Let $\mathcal{F}_\mathbf{T}$ be the class of all measurable estimators of the density based on the data $(Z_\mathbf{x}, \mathbf{x} \in \mathcal{I}_\mathbf{T})$. Then*

$$\liminf_{\mathbf{T}\to+\infty} \inf_{\tilde{f}_\mathbf{T}\in\mathcal{F}_\mathbf{T}} \sup_{Z\in\mathcal{Z}_1} \hat{\mathbf{T}}^{2r/(2r+d)} \mathbf{E}_Z\big(\tilde{f}_\mathbf{T}(z) - f_Z(z)\big)^2 > 0, \quad z \in \mathbb{R}^d.$$

Proof of Theorem 3.2 is based on the observation that each $\tilde{f}_\mathbf{T}$ in $\mathcal{F}_\mathbf{T}$ induces an estimator $\tilde{f}_{[\mathbf{T}]}$ belonging to the family $\mathcal{F}_{[\mathbf{T}]}$ of measurable density estimators based on a suitable *i.i.d.* discretization of the process Z. The proof mimics the proof for $N = 1$ (see Bosq [9]) and details are therefore omitted.

Let us finally indicate that, like in the case of spatial processes indexed by lattices (see Biau [2] for details), similar results may be obtained replacing $A_1(\Gamma, p)$ by

$A'(\Gamma, p)$ – $g_{\mathbf{x},\mathbf{y}}$ exists for $(\mathbf{x}, \mathbf{y}) \notin \Gamma$ and is Lipschitzian uniformly with respect to $(\mathbf{x}, \mathbf{y})$, where Γ satisfies the condition

$$\limsup_{\mathbf{T}\to+\infty} \frac{1}{\hat{\mathbf{T}}} \int_{(\mathcal{I}_\mathbf{T}\times\mathcal{I}_\mathbf{T})\cap\Gamma} \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} < +\infty\,.$$

In that case, the condition $\beta > 2N(p-1)/(p-2)$ is replaced by the weaker condition $\beta > N(2d+1)/(d+1)$.

4 Parametric rates

In the context of stationary time series, Castellana and Leadbetter [15] proved that the mean square error of kernel estimators can reach the *parametric* (or *superoptimal*) rate T^{-1}, provided some specific assumptions on the joint density (Z_0, Z_t) are satisfied. Precisely, these authors enlightened the fact that if the distribution of (Z_0, Z_t) is not too close from a singular distribution for t small, which means that local irregularities of sample paths provide additional information to the statistician, the superoptimal rate T^{-1} holds. The pioneering work of Castellana and Leadbetter led to a lot of works in this direction. We refer the reader to Bosq [8], [10] for further results and references. The next theorem provides an extension of Castellana and Leadbetter results to the spatial case. Let us assume that the bivariate distributions of the process $(Z_{\mathbf{x}}, \mathbf{x} \in \mathbb{R}^N)$ are *stationary* and *isotropic*. In particular, this implies that $g_{\mathbf{x},\mathbf{y}} = f_{(Z_{\mathbf{x}}, Z_{\mathbf{y}})} - f \otimes f$, when it exists, depends only on $\|\mathbf{x} - \mathbf{y}\|$.

Theorem 4.1 (PARAMETRIC RATE) *Assume that $g_{\mathbf{x},\mathbf{y}} = g_{\|\mathbf{x}-\mathbf{y}\|}$ exists for $\mathbf{x} \neq \mathbf{y}$ and that the mapping $u \mapsto u^{N-1}\|g_u\|_\infty$ is integrable over $]0, +\infty[$.*

(1) *If $f \in \mathcal{C}_r^d(l)$ and $K \in \mathcal{H}_{k,c}$ $(r = k + c)$, then the choice $h_{\mathbf{T}} = \mathrm{o}(\hat{\mathbf{T}}^{-1/(2r)})$ leads to*

$$\limsup_{\mathbf{T}\to+\infty} \sup_{z\in\mathbb{R}^d} \hat{\mathbf{T}}\,\mathbf{E}\big(f_{\mathbf{T}}(z) - f(z)\big)^2 \le C\,, \tag{4.1}$$

where C is a positive constant.

(2) *If f is twice continuously differentiable with bounded second derivatives and K is a kernel of order $(1,1)$, then the choice $h_{\mathbf{T}} = \mathrm{o}(\hat{\mathbf{T}}^{-1/4})$ leads to (4.1).*

(3) *If f satisfies an Hölder condition of order c $(0 < c \le 1)$ and K is a kernel of order $(1,1)$, then the choice $h_{\mathbf{T}} = \mathrm{O}(e^{-\hat{\mathbf{T}}})$ implies (4.1).*

The integrability condition of the mapping $u \mapsto u^{N-1}\|g_u\|_\infty$ is the spatial equivalent of Assumption A. in Bosq [8]. It contains an asymptotic independence condition (take u large) and a local irregularity condition (take u small). This local condition means that the information provided by $(Z_{\mathbf{x}}, Z_{\mathbf{x}+\mathbf{h}})$ and $Z_{\mathbf{x}}$, respectively, differ significantly even if $\|\mathbf{h}\|$ is small. Roughly speaking, it also means that the sample trajectories are not smooth.

Indeed, for a real stationary continuous time Gaussian process $(Z_t, t \in \mathbb{R})$ having differentiable sample paths, Bosq [8] shows that $T\,\mathbf{E}\big(f_T(z)-f(z)\big)^2 \to +\infty$ as $T \to +\infty$. Finally, local irregularity of the observed paths provides more information than discrete *i.i.d.* data. This partly explains parametric rates which appear above.

Example. Some insight into the applicability of the above results may be obtained by looking at a class of real stationary and isotropic Gaussian fields for which they hold. Specifically, let $(Z_{\mathbf{x}}, \mathbf{x} \in \mathbb{R}^N)$ be a stationary and isotropic spatial Gaussian process with zero mean and autocorrelation function

$$\rho(u) = 1 - au^\theta + \mathrm{o}(u^\theta) \quad \text{as } u \to 0(+), \tag{4.2}$$

where $a, \theta > 0$. Then it is easy to verify that

$$\big|g_u(z_1, z_2)\big| \le a\big|\rho(u)\big|\mathbf{1}_{u>b} + (c + d\,u^{-\theta/2})\mathbf{1}_{u\le b}\,, \quad u > 0,$$

where a, b, c and d are suitable constants. Therefore, conditions in Theorem 4.1 are satisfied as soon as $u \mapsto u^{N-1}\rho(u)$ is integrable on $[b, +\infty[$ and $\theta < 2N$. Gaussian fields with $\theta < 2N$ in (4.2) have "irregular" sample paths in contrast to the more regular case $\theta \ge 2N$. The irregular nature of the spatial paths correspond to less correlation and hence more information in the measurement of $Z_{\mathbf{x}}$, leading to the maximal rate of convergence of the variance to 0. In the time domain, it should be noted that the class with $\theta < 2$ contains interesting cases, such as the *Ornstein-Uhlenbeck* process ($\theta = 1$).

Remark 1. It is possible to derive sufficient conditions for *intermediate* rates between $\hat{\mathbf{T}}^{-1}$ and $\hat{\mathbf{T}}^{-r/(r+d)}$ to occur, but we will not develop this point here. In the time domain, a detailed study can be found in Blanke and Bosq [5].

5 Sampling

In practice, a spatial process is observed only at a *finite* number of appropriately chosen points, and based on these observations an estimate or a statistic is formed for use in the problem at hand. How to select the most appropriate choice of sampling points is the problem of *sampling design*. The statistician may be free to choose any desirable points, or else part of the sampling mechanism may be imposed on the statistician who then controls certain parameters, *e.g.* periodic sampling is imposed where the period is controlled by the statistician, or Poisson sampling (at the times of occurrence of a Poisson stream of events) is imposed but the statistician has control over its rate. Usually, two kinds of sampling schemes are considered: *deterministic* and *random*. In deterministic sampling, the sampling points are chosen according to a deterministic rule, such as periodic sampling. In random sampling, the sampling points are chosen according to a randomized rule, such as Poisson sampling. It is noteworthy that imperfections in deterministic sampling schemes (*e.g.* periodic sampling with jitter) may lead to random sampling.

In this section, we will be concerned with a spatial process $(Z_{\mathbf{x}}, \mathbf{x} \in \mathbb{R}^N)$ observed at deterministic sampling locations. Let us first assume that the process is observed on a grid $\mathcal{G}_{\mathbf{n}} = \{t_{1,1}, \ldots, t_{1,n_1}\} \times \ldots \times \{t_{N,1}, \ldots, t_{N,n_N}\}$ with $\mathbf{n} = (n_1, \ldots, n_N)$. Suppose moreover that there exists a constant $m > 0$ such that, for all $k = 1, \ldots, N$,

$$\min_{1 \leq j \leq n_k - 1} (t_{k,j+1} - t_{k,j}) \geq m.$$

Setting $\hat{\mathbf{n}} = n_1 \ldots n_N$, the corresponding kernel estimator at the point $z \in \mathbb{R}^d$ reads

$$\overline{f}_{\mathbf{n}}(z) = \frac{1}{\hat{\mathbf{n}} h_{\mathbf{n}}^d} \sum_{\mathbf{i} \in \mathcal{G}_{\mathbf{n}}} K\left(\frac{z - Z_{\mathbf{i}}}{h_{\mathbf{n}}}\right).$$

It is not difficult to see that the asymptotic behavior of $\overline{f}_{\mathbf{n}}$ is the same as that of the estimator studied in Biau [2], where density estimation of a process $(Z_{\mathbf{n}}, \mathbf{n} \in \mathbb{Z}^N)$ is considered. All the results of this paper remains therefore valid with slight modifications.

Now, in order to model the fact that the observations are frequent on a large region of the space, assume that the random field $(Z_{\mathbf{x}}, \mathbf{x} \in \mathbb{R}^N)$ is observed on a discrete grid $\mathcal{I}_{\underline{\mathbf{n}}} = \{\delta_n, 2\delta_n, \ldots, n\delta_n\}^N$, where $\underline{\mathbf{n}} = (n, \ldots, n)$ (N times), $\delta_n \to 0(+)$ and $T_n = n\delta_n \to +\infty$ as $n \to +\infty$. Our objective is to select (δ_n) (the *grid step*) so that the rate of kernel estimation should be parametric, *i.e.*, $\mathrm{O}(1/T_n^N)$. Moreover, we are looking for a sequence (δ_n) which is optimal in the following sense: if (δ'_n) is such that the mean square error of the kernel estimator based on the discretized data is $\mathrm{O}(1/T_n^N)$, then $\delta'_n = \mathrm{O}(\delta_n)$. An optimal sampling (δ_n^*) is usually called *admissible*. Typically, the following two situations may occur. Consider first the case where the domain of observation is $[0, T_n]^N$ with a fixed T_n. Since $T_n = n\delta_n$, the knowledge of the optimal δ_n^* permits to select the minimal number n^N of points in $[0, T_n]^N$ which will play a part in the estimation. On the other hand, suppose now that a maximal sample size n^N is given (*e.g.* one wants to minimize costs for taking observations from the hidden process), then it is possible to reduce the total size $T_n^N = n^N \delta_n^N$ of the experiment without altering rate of estimation. In the time domain, this problem has been examined by Bosq [8] and more recently by Blanke and Pumo [6]. For further references, we refer the reader to Masry [25] and Prakasa Rao [26].

Denoting by $\mathbf{i} = (i_1, \ldots, i_N)$ a generic point of $\mathcal{I}_{\underline{\mathbf{n}}}$, the kernel estimator based on the observations on $\mathcal{I}_{\underline{\mathbf{n}}}$ is defined by

$$f^*_{\underline{\mathbf{n}}}(z) = \frac{1}{n^N h_n^d} \sum_{\mathbf{i} \in \mathcal{I}_{\underline{\mathbf{n}}}} K\left(\frac{z - Z_{\mathbf{i}}}{h_n}\right), \quad z \in \mathbb{R}^d.$$

In order to specify an admissible sampling we need the following assumptions.

(1) $g_{\mathbf{x},\mathbf{y}} = g_{\|\mathbf{x}-\mathbf{y}\|}$ exists for $\mathbf{x} \neq \mathbf{y}$ and $\|g_u\|_\infty \leq \pi(u), u > 0$, where $(1 + u + \ldots + u^{2N-1})\pi(u)$ is integrable over $]0, +\infty[$, $u^{2N-1}\pi(u)$ is ultimately decreasing and $u^k\pi(u)$ is bounded for $k = N, \ldots, 2N-1$. Furthermore, g_u is continuous over the diagonal of $\mathbb{R}^{2d}$ and,

(2) for $k = 0, \ldots, N-1$,

$$\sup_{(z_1,z_2)\in\mathbb{R}^{2d}} \left| \int_0^{+\infty} u^k g_u(z_1, z_2) - \sum_{i=1}^{+\infty} \delta_n (i\delta_n)^k g_{i\delta_n}(z_1, z_2) \right| \to 0$$

as $\delta_n \to 0$.

These assumptions are satisfied if, for example, $(Z_{\mathbf{x}}, \mathbf{x} \in \mathbb{R}^N)$ is a stationary and isotropic Gaussian field with zero mean and autocorrelation $\exp(-au^\theta), a > 0$ and $0 < \theta < 2$ $\big($see Example (4.2)$\big)$. From now on, it is assumed for simplicity that the underlying norm over $\mathbb{R}^N$ is the supremum norm, *i.e.*, for $\mathbf{x} = (x_1, \ldots, x_N) \in \mathbb{R}^N$, $\|\mathbf{x}\| = \max_{1\leq i\leq N} |x_i|$.

Theorem 5.1 *If assumptions (1) and (2) hold, if $f \in \mathcal{C}_r^d(l)$ and $K \in \mathcal{H}_{k,c}$ $(r = k + c)$, then the sampling $\delta_n = T_n^{-d/(2r)}$ is admissible provided $h_n \sim n^{-N/(2r+d)}$. More precisely, for $z \in \mathbb{R}^d$,*

$$\mathbf{E}\big(f_n^*(z) - f(z)\big)^2 = \mathrm{O}\left(\frac{1}{T_n^N}\right)$$

and the sequence (δ_n) is optimal.

Note that the choice of δ_n in Theorem 5.1 implies $T_n = n^{2r/(2r+d)}$; consequently, although the obtained rates are parametric with respect to T_n, they are, in fact, nonparametric with respect to n. However, it is interesting to note that the n^N discrete observations are sampled over an hypercube of size $T_n^N << n^N$. Proof of Theorem 5.1 will strongly rely on the following combinatorial lemma, which is interesting by itself.

Lemma 5.1 *Consider the grid $\{\delta_n, 2\delta_n, \ldots, n\delta_n\}^N$. Then the total number of pairs of points on this grid whose components are separated by a distance of $i\delta_n$ ($i = 1, \ldots, n-1$), say $P_{n,i}$, is a polynomial of the form*

$$P_{n,i} = n^N \left[\sum_{k=0}^{N-2} i^k \mathrm{O}_k(1) + i^{N-1} \left(\Delta + \mathrm{O}_{N-1}\Big(\frac{1}{n}\Big) \right) + \sum_{k=N}^{2N-1} i^k \mathrm{O}_k \Big(\frac{1}{n^{k-N+1}} \Big) \right],$$

where $\Delta = N 2^N$.

The following corollary provides finally the exact asymptotic quadratic error associated with an admissible sampling.

Corollary 5.1 *If assumptions (1) and (2) hold, if f is twice continuously differentiable with bounded second derivatives and K is a kernel of order $(1,1)$, then the choice $\delta_n = T_n^{-d/4}$ and $h_n \sim n^{-N/(4+d)}$ leads, for $z \in \mathbb{R}^d$, to*

$$\lim_{n \to +\infty} T_n^N \, \mathbf{E} \big(f_n^*(z) - f(z) \big)^2 = C \,,$$

where C is a positive constant.

6 Proofs

Given a kernel K and a smoothing parameter h, we set $K_h(.) = K(./h)/h^d$. Throughout the section, we denote by $V(\mathcal{S}_{N-1})$ the volume of the *unit sphere* $\mathcal{S}_{N-1}$ in $\mathbb{R}^N$, defined by $\mathcal{S}_{N-1} = \{\mathbf{x} \in \mathbb{R}^N : \|\mathbf{x}\| = 1\}$. Recall that, when the underlying norm over $\mathbb{R}^N$ is the Euclidean one, for $p \geq 0$,

$$V(\mathcal{S}_{2p}) = \frac{2^{p+1}\pi^p}{1.3\ldots(2p-1)} \quad \text{and} \quad V(\mathcal{S}_{2p+1}) = \frac{2\pi^{p+1}}{p!} \,.$$

Moreover, for $N \geq 1$,

$$V(\mathcal{B}_N) = \frac{V(\mathcal{S}_{N-1})}{N} \,.$$

Proof of Theorem 3.1. Proof of Theorem 3.1 will strongly rely on the following lemma.

Lemma 6.1

(1) *Let $z \in \mathbb{R}^d$. If $A_1(\Gamma, p)$ and $A_2(\gamma, \beta)$ hold for some Γ, p, γ, β with $\beta \geq 2N(p-1)/(p-2)$, then*

$$\begin{aligned}\text{Var } f_{\mathbf{T}}(z) \leq & \frac{1}{\hat{\mathbf{T}} h_{\mathbf{T}}^d} \mathbf{E}\left[\frac{1}{h_{\mathbf{T}}^d} K^2\left(\frac{z - Z_0}{h_{\mathbf{T}}}\right)\right] . \frac{1}{\hat{\mathbf{T}}} \int_{(\mathcal{I}_{\mathbf{T}} \times \mathcal{I}_{\mathbf{T}}) \cap \Gamma} \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} \\ & + \left(V(\mathcal{B}_N)\, \delta_p(\Gamma)\, \|K\|_q^2 + \frac{4V(\mathcal{S}_{N-1})\, \gamma\, \|K\|_\infty^2}{\beta - N}\right) \frac{h_{\mathbf{T}}^\eta}{\hat{\mathbf{T}} h_{\mathbf{T}}^d},\end{aligned} \tag{6.1}$$

where $q = p/(p-1)$ and $\eta = 2d(1 - N/\beta)/q - d \geq 0$.

(2) *Let $z \in \mathbb{R}^d$. If $A_1(\Gamma, p)$ and $A_2(\gamma, \beta)$ hold for some Γ, p, γ, β with $\beta > 2N(p-1)/(p-2)$, f is continuous at z and $\hat{\mathbf{T}} h_{\mathbf{T}}^d \to +\infty$, then*

$$\limsup_{\mathbf{T} \to +\infty} \hat{\mathbf{T}} h_{\mathbf{T}}^d \text{Var } f_{\mathbf{T}}(z) \leq l_\Gamma\, f(z)\, \|K\|_2^2\,; \tag{6.2}$$

if f is bounded, then

$$\limsup_{\mathbf{T} \to +\infty} \sup_{z \in \mathbb{R}^d} \hat{\mathbf{T}} h_{\mathbf{T}}^d \text{Var } f_{\mathbf{T}}(z) \leq l_\Gamma\, \|f\|_\infty\, \|K\|_2^2\,. \tag{6.3}$$

Proof of (1). Let us consider the decomposition

$$\begin{aligned}\hat{\mathbf{T}} h_{\mathbf{T}}^d \text{Var } f_{\mathbf{T}}(z) = & \int_{(\mathcal{I}_{\mathbf{T}} \times \mathcal{I}_{\mathbf{T}}) \cap \Gamma} \text{Cov}\left[K\left(\frac{z - Z_{\mathbf{x}}}{h_{\mathbf{T}}}\right), K\left(\frac{z - Z_{\mathbf{y}}}{h_{\mathbf{T}}}\right)\right] \frac{\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}}{\hat{\mathbf{T}} h_{\mathbf{T}}^d} \\ & + \int_{(\mathcal{I}_{\mathbf{T}} \times \mathcal{I}_{\mathbf{T}}) \cap \Gamma^c} \text{Cov}\left[K\left(\frac{z - Z_{\mathbf{x}}}{h_{\mathbf{T}}}\right), K\left(\frac{z - Z_{\mathbf{y}}}{h_{\mathbf{T}}}\right)\right] \frac{\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}}{\hat{\mathbf{T}} h_{\mathbf{T}}^d}.\end{aligned}$$

The first integral can be bounded above by

$$I_{\mathbf{T}} := \frac{1}{h_{\mathbf{T}}^d} \mathbf{E} K^2\left(\frac{z - Z_0}{h_{\mathbf{T}}}\right) . \frac{1}{\hat{\mathbf{T}}} \int_{(\mathcal{I}_{\mathbf{T}} \times \mathcal{I}_{\mathbf{T}}) \cap \Gamma} \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}\,. \tag{6.4}$$

To deal with the second integral, we introduce the set $\mathcal{J}_{\mathbf{T}} = \{(\mathbf{x}, \mathbf{y}) \in \mathcal{I}_{\mathbf{T}} \times \mathcal{I}_{\mathbf{T}} \,:\, \|\mathbf{x} - \mathbf{y}\| \leq h_{\mathbf{T}}^{-2d/(q\beta)}\}$. We get, using Lemma 2.1 and assumption $A_1(\Gamma, p)$,

$$\begin{aligned}J_{\mathbf{T}} := & \frac{1}{\hat{\mathbf{T}} h_{\mathbf{T}}^d} \int_{(\mathcal{I}_{\mathbf{T}} \times \mathcal{I}_{\mathbf{T}}) \cap \Gamma^c \cap \mathcal{J}_{\mathbf{T}}} \left|\text{Cov}\left[K\left(\frac{z - Z_{\mathbf{x}}}{h_{\mathbf{T}}}\right), K\left(\frac{z - Z_{\mathbf{y}}}{h_{\mathbf{T}}}\right)\right]\right| \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} \\ \leq & \frac{1}{\hat{\mathbf{T}} h_{\mathbf{T}}^d}\, \delta_p(\Gamma)\, \|K\|_q^2\, \lambda(\mathcal{J}_{\mathbf{T}})\, h_{\mathbf{T}}^{2d/q}.\end{aligned}$$

Since $\lambda(\mathcal{J}_{\mathbf{T}}) \leq V(\mathcal{B}_N)\, \hat{\mathbf{T}}\, h_{\mathbf{T}}^{-2dN/(q\beta)}$, we obtain

$$J_{\mathbf{T}} \leq V(\mathcal{B}_N)\, \delta_p(\Gamma)\, \|K\|_q^2\, h_{\mathbf{T}}^{2d(1-N/\beta)/q - d}\,. \tag{6.5}$$

On the other hand, Lemma 2.1 with $A_2(\gamma,\beta)$ yields

$$\begin{aligned}
J'_{\mathbf{T}} &:= \frac{1}{\hat{\mathbf{T}} h_{\mathbf{T}}^d} \int_{(\mathcal{I}_{\mathbf{T}}\times\mathcal{I}_{\mathbf{T}})\cap\Gamma^c\cap\mathcal{J}_{\mathbf{T}}^c} \left| \mathrm{Cov}\left[K\left(\frac{z-Z_{\mathbf{x}}}{h_{\mathbf{T}}}\right), K\left(\frac{z-Z_{\mathbf{y}}}{h_{\mathbf{T}}}\right)\right]\right| \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} \\
&\leq \frac{4\gamma \left\|K\right\|_\infty^2}{\hat{\mathbf{T}} h_{\mathbf{T}}^d} \int_{(\mathcal{I}_{\mathbf{T}}\times\mathcal{I}_{\mathbf{T}})\cap\Gamma^c\cap\mathcal{J}_{\mathbf{T}}^c} \left\|\mathbf{x}-\mathbf{y}\right\|^{-\beta} \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} \\
&\leq \frac{4V(\mathcal{S}_{N-1})\,\gamma\left\|K\right\|_\infty^2}{h_{\mathbf{T}}^d} \int_{h_{\mathbf{T}}^{-2d/(q\beta)}}^{+\infty} u^{N-1} u^{-\beta} \mathrm{d}u \\
&\qquad \text{(see Rudin [29], Chapter 8, page 175),}
\end{aligned}$$

hence

$$J'_{\mathbf{T}} \leq \frac{4V(\mathcal{S}_{N-1})\,\gamma\left\|K\right\|_\infty^2}{\beta - N} h_{\mathbf{T}}^{2d(1-N/\beta)/q-d}, \tag{6.6}$$

and finally (6.4), (6.5) and (6.6) imply (6.1).

Proof of (2). Using (6.1), noting that here η is (strictly) positive and, by the classical Bochner's lemma (Bosq and Lecoutre [11]), that

$$\lim_{\mathbf{T}\to+\infty} \mathbf{E} K_{h_{\mathbf{T}}}^2(z - Z_0) = f(z) \int_{\mathbb{R}^d} K^2(u)\mathrm{d}u\,,$$

we get (6.2). Concerning (6.3), it suffices to note that

$$\mathbf{E} K_{h_{\mathbf{T}}}^2(z - Z_0) \leq \|f\|_\infty \int_{\mathbb{R}^d} K^2(u)\mathrm{d}u\,.$$

■

We are now ready to prove Theorem 3.1. Recall that

$$\mathbf{E}\big(f_{\mathbf{T}}(z) - f(z)\big)^2 = \mathrm{Var}\; f_{\mathbf{T}}(z) + \big(\mathbf{E} f_{\mathbf{T}}(z) - f(z)\big)^2.$$

Taylor's formula and (2.2) entail

$$\begin{aligned}
\mathbf{E} f_{\mathbf{T}}(z) - f(z) &= \int_{\mathbb{R}^d} K(u)\big[f(z - h_{\mathbf{T}} u) - f(z)\big]\mathrm{d}u \\
&= h_{\mathbf{T}}^k \int_{\mathbb{R}^d} K(u) \sum_{j_1+\ldots+j_d=k} \frac{u_1^{j_1}\ldots u_d^{j_d}}{j_1!\ldots j_d!} \frac{\partial^{(k)} f}{\partial z_1^{j_1}\ldots\partial z_d^{j_d}} (z - \theta h_{\mathbf{T}} u)\mathrm{d}u\,,
\end{aligned}$$

where $0<\theta<1$. Using again (2.2) we obtain

$$\begin{aligned}
\mathbf{E} f_{\mathbf{T}}(z) - f(z) = h_{\mathbf{T}}^k \int_{\mathbb{R}^d} K(u) \sum_{j_1+\ldots+j_d=k} \frac{u_1^{j_1}\ldots u_d^{j_d}}{j_1!\ldots j_d!} \Bigg[&\frac{\partial^{(k)} f}{\partial z_1^{j_1}\ldots\partial z_d^{j_d}} (z - \theta h_{\mathbf{T}} u) \\
&- \frac{\partial^{(k)} f}{\partial z_1^{j_1}\ldots\partial z_d^{j_d}} (z)\mathrm{d}u\Bigg],
\end{aligned}$$

and (2.1) implies

$$\big|\mathbf{E}f_{\mathbf{T}}(z)-f(z)\big|\le c_{(r)}h_{\mathbf{T}}^{k+c}\,, \tag{6.7}$$

where

$$c_{(r)}=\sum_{j_1+\ldots+j_d=k}\frac{l}{j_1!\ldots j_d!}\int_{\mathbb{R}^d}\|u\|^c\,|u_1|^{j_1}\ldots|u_d|^{j_d}\big|K(u)\big|\mathrm{d}u\,. \tag{6.8}$$

Now, assume that $h_{\mathbf{T}}=c_{\mathbf{T}}\hat{\mathbf{T}}^{-1/(2r+d)}$, where $c_{\mathbf{T}}\to c'>0$. Inequality (3.1) is a straightforward consequence of (6.3) and (6.7), with

$$\begin{aligned}C=&\ \frac{l_\Gamma\,\|f\|_\infty\,\|K\|_2^2}{c'^d}\\&+c'^{2r}\Big(\sum_{j_1+\ldots+j_d=k}\frac{l}{j_1!\ldots j_d!}\int_{\mathbb{R}^d}\|u\|^c\,|u_1|^{j_1}\ldots|u_d|^{j_d}\,\big|K(u)\big|\mathrm{d}u\Big)^2.\end{aligned}$$

■

Proof of Corollary 3.1. We start with a proposition.

Proposition 6.1 *Assume that $T_ih_{\mathbf{T}}^d\to+\infty$ for $i=1,\ldots,N$. Then*

$$\lim_{\mathbf{T}\to+\infty}\max_{Z\in\mathcal{Z}}\sup_{z\in\mathbb{R}^d}\hat{\mathbf{T}}h_{\mathbf{T}}^d\mathrm{Var}_Z\,f_{\mathbf{T}}(z)=L_0\,f_0\,\|K\|_2^2\,, \tag{6.9}$$

where Var_Z *denotes the variance when the underlying process is Z.*

Proof. An easy consequence of (6.1) is

$$\limsup_{\mathbf{T}\to+\infty}\sup_{Z\in\mathcal{Z}}\sup_{z\in\mathbb{R}^d}\hat{\mathbf{T}}h_{\mathbf{T}}^d\mathrm{Var}_Z\,f_{\mathbf{T}}(z)\le L_0f_0\int_{\mathbb{R}^d}K^2(u)\mathrm{d}u\,.$$

It remains to exhibit a process $\tilde{Z}$ in $\mathcal{Z}$ such that

$$\sup_{z\in\mathbb{R}^d}\hat{\mathbf{T}}h_{\mathbf{T}}^d\mathrm{Var}_{\tilde{Z}}\,f_{\mathbf{T}}(z)\to L_0f_0\int_{\mathbb{R}^d}K^2(u)\mathrm{d}u\quad\text{as }\mathbf{T}\to+\infty\,.$$

To this aim we consider a sequence $(Y_{\mathbf{n}},\mathbf{n}\in\mathbb{Z}^N)$ of *i.i.d.* $\mathbb{R}^d$-valued random variables with a uniformly continuous density f such that $\|f\|_\infty=f(z_0)=f_0$ for some z_0. Set

$$\tilde{Z}_{\mathbf{x}}=Y_{[\mathbf{x}/L_0^{1/N}]},\quad \mathbf{x}\in\mathbb{R}^N,$$

where, for $\mathbf{x}=(x_1,\ldots,x_N)\in\mathbb{R}^N$, the notation $[\mathbf{x}]$ stands for the integer part of $\mathbf{x}$ *coordinatewise, i.e.*, $[\mathbf{x}]=\big([x_1],\ldots,[x_N]\big)$. It is easy to check that the spatial process $\tilde{Z}$ belongs to $\mathcal{Z}$ with

$$\Gamma=\bigcup_{\mathbf{n}\in\mathbb{Z}^N}\big\{(\mathbf{x},\mathbf{y})\in\mathbb{R}^N\times\mathbb{R}^N\ :\ \big[\mathbf{x}/L_0^{1/N}\big]=\big[\mathbf{y}/L_0^{1/N}\big]=\mathbf{n}\big\},$$

$\delta_{p_0}(\Gamma) = 0$ and $\alpha\big(\|\mathbf{x}-\mathbf{y}\|\big) = 0$ if $(\mathbf{x},\mathbf{y}) \in \Gamma^c$.

Let us introduce the kernel estimator $\hat{f}_{\mathbf{T}}$ associated with the discrete *i.i.d.* sample $(Y_{\mathbf{n}}, \mathbf{n} \in \hat{\mathcal{I}}_{\mathbf{T}})$, with $\hat{\mathcal{I}}_{\mathbf{T}} = \big\{\mathbf{n} \in \mathbb{N}^N \,:\, 0 \le n_i \le [T_i/L_0^{1/N}] - 1, i = 1, \ldots, N\big\}$. Then for the particular process $\tilde{Z}_{\mathbf{x}}$ observed on $\mathcal{I}_{\mathbf{T}}$, $f_{\mathbf{T}}$ takes a special form, namely

$$f_{\mathbf{T}}(z) = \frac{[T_1/L_0^{1/N}] \ldots [T_N/L_0^{1/N}]}{\hat{\mathbf{T}}/L_0}\, \hat{f}_{\mathbf{T}}(z) + \frac{1}{\hat{\mathbf{T}} h_{\mathbf{T}}^d} \int_{\mathcal{P}_{\mathbf{T}}} K\left(\frac{z - \tilde{Z}_{\mathbf{x}}}{h_{\mathbf{T}}}\right) \mathrm{d}\mathbf{x}\,,$$

where $\mathcal{P}_{\mathbf{T}} = \{\mathbf{x} \in \mathcal{I}_{\mathbf{T}} \,:\, x_i \ge [T_i/L_0^{1/N}]\, L_0^{1/N}, i = 1, \ldots, N\}$. Using a well-known result (Bosq and Lecoutre [11]), we deduce that

$$[T_1/L_0^{1/N}] \ldots [T_N/L_0^{1/N}]\, h_{\mathbf{T}}^d \sup_{z \in \mathbb{R}^d} \mathrm{Var}\ \hat{f}_{\mathbf{T}}(z) \to f_0 \int_{\mathbb{R}^d} K^2(u) \mathrm{d}u\,,$$

and consequently

$$\hat{\mathbf{T}} h_{\mathbf{T}}^d \sup_{z \in \mathbb{R}^d} \mathrm{Var}\ \hat{f}_{\mathbf{T}}(z) \to L_0\, f_0 \int_{\mathbb{R}^d} K^2(u) \mathrm{d}u\,.$$

On the other hand, since $\tilde{Z}_{\mathbf{x}}$ is piecewise constant on $\mathcal{P}_{\mathbf{T}}$, attention shows that

$$\begin{aligned}
&\sup_{z \in \mathbb{R}^d} \mathrm{Var}\left[\frac{1}{\hat{\mathbf{T}} h_{\mathbf{T}}^d} \int_{\mathcal{P}_{\mathbf{T}}} K\left(\frac{z - \tilde{Z}_{\mathbf{x}}}{h_{\mathbf{T}}}\right) \mathrm{d}\mathbf{x}\right] \\
&\quad \le \frac{\|K\|_\infty^2\, L_0^2}{\hat{\mathbf{T}}^2 h_{\mathbf{T}}^{2d}} \sum_{1 \le i_1 < \ldots < i_{N-1} \le N} \left(\left[\frac{T_{i_1}}{L_0^{1/N}}\right] + 1\right) \cdots \left(\left[\frac{T_{i_{N-1}}}{L_0^{1/N}}\right] + 1\right),
\end{aligned}$$

with the convention that the sum is 1 for $N = 1$. Finally, using the assumptions on the T_i's, we obtain

$$\hat{\mathbf{T}} h_{\mathbf{T}}^d \sup_{z \in \mathbb{R}^d} \mathrm{Var}_{\tilde{Z}}\, f_{\mathbf{T}}(z) \to L_0\, f_0 \int_{\mathbb{R}^d} K^2(u) \mathrm{d}u\,,$$

which implies (6.9). It should be noted that the process $\tilde{Z}$ is not stationary. ■

Proof of Corollary 3.1 is now an obvious consequence of Proposition 6.1 together with (6.7) and (6.8). Note that

$$\begin{aligned}
C = {} & \frac{L_0\, f_0\, \|K\|_2^2}{c'^d} \\
& + c'^{2r} \Big(\sum_{j_1 + \ldots + j_d = k} \frac{l}{j_1! \ldots j_d!} \int_{\mathbb{R}^d} \|u\|^c\, |u_1|^{j_1} \ldots |u_d|^{j_d}\, \big|K(u)\big| \mathrm{d}u\Big)^2,
\end{aligned}$$

where $h_{\mathbf{T}} = c_{\mathbf{T}} \hat{\mathbf{T}}^{-1/(2r+d)}$ and $c_{\mathbf{T}} \to c' > 0$. ■

Proof of Theorem 4.1. Theorem 4.1 follows directly from the following proposition.

Proposition 6.2 *Let $z \in \mathbb{R}^d$. If $g_{\mathbf{x},\mathbf{y}} = g_{\|\mathbf{x}-\mathbf{y}\|}$ exists for $\mathbf{x} \neq \mathbf{y}$, if the mapping*

$$(z_1, z_2) \mapsto \int_0^{+\infty} u^{N-1} \big| g_u(z_1, z_2) \big| \mathrm{d}u$$

is defined, bounded and continuous at (z, z), then

$$\limsup_{\mathbf{T}\to+\infty} \hat{\mathbf{T}} \operatorname{Var} f_{\mathbf{T}}(z) \leq V(\mathcal{S}_{N-1}) \, \|K\|_1^2 \int_0^{+\infty} u^{N-1} \big| g_u(z, z) \big| \mathrm{d}u \,. \tag{6.10}$$

Proof. Fubini's theorem entails

$$\operatorname{Var} f_{\mathbf{T}}(z) = \frac{1}{\hat{\mathbf{T}}^2} \int_{\mathcal{I}_{\mathbf{T}} \times \mathcal{I}_{\mathbf{T}}} \operatorname{Cov}\big(K_{h_{\mathbf{T}}}(z - Z_{\mathbf{x}}), K_{h_{\mathbf{T}}}(z - Z_{\mathbf{y}})\big) \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} \,.$$

Using the condition $g_{\mathbf{x},\mathbf{y}} = g_{\|\mathbf{x}-\mathbf{y}\|}$, we can write

$$\operatorname{Cov}\big(K_{h_{\mathbf{T}}}(z-Z_{\mathbf{x}}), K_{h_{\mathbf{T}}}(z-Z_{\mathbf{y}})\big) = \int_{\mathbb{R}^{2d}} K_{h_{\mathbf{T}}}(z-z_1) K_{h_{\mathbf{T}}}(z-z_2) g_{\|\mathbf{x}-\mathbf{y}\|}(z_1, z_2) \mathrm{d}z_1 \mathrm{d}z_2 \,.$$

Therefore

$$\begin{aligned}
&\hat{\mathbf{T}} \operatorname{Var} f_{\mathbf{T}}(z) \\
&\quad = \frac{1}{\hat{\mathbf{T}}} \int_{\mathcal{I}_{\mathbf{T}} \times \mathcal{I}_{\mathbf{T}}} \left[\int_{\mathbb{R}^{2d}} K_{h_{\mathbf{T}}}(z - z_1) K_{h_{\mathbf{T}}}(z - z_2) g_{\|\mathbf{x}-\mathbf{y}\|}(z_1, z_2) \mathrm{d}z_1 \mathrm{d}z_2 \right] \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} \\
&\quad = \frac{1}{\hat{\mathbf{T}}} \int_{\mathbb{R}^{2d}} K_{h_{\mathbf{T}}}(z - z_1) K_{h_{\mathbf{T}}}(z - z_2) \left[\int_{\mathcal{I}_{\mathbf{T}} \times \mathcal{I}_{\mathbf{T}}} g_{\|\mathbf{x}-\mathbf{y}\|}(z_1, z_2) \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} \right] \mathrm{d}z_1 \mathrm{d}z_2 \\
&\quad \leq V(\mathcal{S}_{N-1}) \int_{\mathbb{R}^{2d}} \big| K_{h_{\mathbf{T}}}(z - z_1) K_{h_{\mathbf{T}}}(z - z_2) \big| \left[\int_0^{+\infty} u^{N-1} \big| g_u(z_1, z_2) \big| \mathrm{d}u \right] \mathrm{d}z_1 \mathrm{d}z_2 \,.
\end{aligned}$$

Taking lim sup on both side and applying Bochner's lemma we obtain (6.10). ■

Proof of Lemma 5.1. Let us first evaluate the total number of pairs of points on the grid whose components are separated by a distance of at least $i\delta_n$, $i = 1, \ldots, n-1$. To this aim, take $\mathbf{X} = (X_1, \ldots, X_N)$ and $\mathbf{Y} = (Y_1, \ldots, Y_N)$ on the grid and consider $\mathbf{X}$ and $\mathbf{Y}$ as two independent random variables with uniform distribution on $\{\delta_n, \ldots, n\delta_n\}^N$. Then

$$\begin{aligned}
&\mathbf{P}\big(\|\mathbf{X}-\mathbf{Y}\| \geq i\delta_n\big) \\
&\quad = 1 - \mathbf{P}\big(\|\mathbf{X}-\mathbf{Y}\| < i\delta_n\big) \\
&\quad = 1 - \mathbf{P}\big(|X_1 - Y_1| < i\delta_n, \ldots, |X_N - Y_N| < i\delta_n\big) \\
&\quad = 1 - \Big[\mathbf{P}\big(|X_1 - Y_1| < i\delta_n\big)\Big]^N \\
&\quad = 1 - \Big[\mathbf{P}\big(|X_1 - Y_1| = 0\big) + \mathbf{P}\big(|X_1 - Y_1| = \delta_n\big) \\
&\qquad\qquad + \ldots + \mathbf{P}\big(|X_N - Y_N| = (i-1)\delta_n\big)\Big]^N.
\end{aligned}$$

Elementary calculations show that

$$\mathbf{P}\big(|X_1 - Y_1| = 0\big) = \frac{1}{n}$$

and

$$\mathbf{P}\big(|X_1 - Y_1| = j\delta_n\big) = \frac{2(n-j)}{n^2}, \quad j = 1, \ldots, i-1.$$

Therefore

$$\mathbf{P}\big(\|\mathbf{X}-\mathbf{Y}\| \geq i\delta_n\big) = 1 - \bigg(-\frac{1}{n} + \frac{2i}{n} + \frac{i}{n^2} - \frac{i^2}{n^2}\bigg)^N, \quad i = 1, \ldots, n-1.$$

Similarly, we obtain

$$\mathbf{P}\big(\|\mathbf{X}-\mathbf{Y}\| \geq (i+1)\delta_n\big) = 1 - \bigg(\frac{1}{n} + \frac{2i}{n} - \frac{i}{n^2} - \frac{i^2}{n^2}\bigg)^N, \quad i = 1, \ldots, n-1.$$

Thus, for $i = 1, \ldots, n-1$,

$$\begin{aligned}
&\mathbf{P}\big(\|\mathbf{X}-\mathbf{Y}\| = i\delta_n\big) \\
&\quad = \mathbf{P}\big(\|\mathbf{X}-\mathbf{Y}\| \geq i\delta_n\big) - \mathbf{P}\big(\|\mathbf{X}-\mathbf{Y}\| \geq (i+1)\delta_n\big) \\
&\quad = \bigg(\frac{1}{n} + \frac{2i}{n} - \frac{i}{n^2} - \frac{i^2}{n^2}\bigg)^N - \bigg(-\frac{1}{n} + \frac{2i}{n} + \frac{i}{n^2} - \frac{i^2}{n^2}\bigg)^N.
\end{aligned}$$

Noting that $P_{n,i} = n^{2N}\,\mathbf{P}\big(\|\mathbf{X}-\mathbf{Y}\| = i\delta_n\big)$, we obtain

$$P_{n,i} = n^N\Bigg[\bigg(1 + \Big(2 - \frac{1}{n}\Big)i - \frac{i^2}{n}\bigg)^N - \bigg(-1 + \Big(2 + \frac{1}{n}\Big)i - \frac{i^2}{n}\bigg)^N\Bigg].$$

A multinomial expansion leads to

$$P_{n,i} = n^N \sum_{\substack{0\le\alpha_1,\alpha_2,\alpha_3\le N\\ \alpha_1+\alpha_2+\alpha_3=N}} \frac{N!}{\alpha_1!\alpha_2!\alpha_3!}\Bigg[\Big(2-\frac{1}{n}\Big)^{\alpha_2} i^{\alpha_2}(-1)^{\alpha_3}\frac{i^{2\alpha_3}}{n^{\alpha_3}}$$
$$- (-1)^{\alpha_1}\Big(2+\frac{1}{n}\Big)^{\alpha_2} i^{\alpha_2}(-1)^{\alpha_3}\frac{i^{2\alpha_3}}{n^{\alpha_3}}\Bigg]$$
$$= n^N \sum_{\substack{0\le\alpha_1,\alpha_2,\alpha_3\le N\\ \alpha_1+\alpha_2+\alpha_3=N}} \frac{N!}{\alpha_1!\alpha_2!\alpha_3!} i^{\alpha_2}(-1)^{\alpha_3}\frac{i^{2\alpha_3}}{n^{\alpha_3}}\Bigg[\Big(2-\frac{1}{n}\Big)^{\alpha_2} - (-1)^{\alpha_1}\Big(2+\frac{1}{n}\Big)^{\alpha_2}\Bigg].$$

In conclusion,

$$P_{n,i} = n^N\Bigg[\sum_{k=0}^{N-2} i^k \mathrm{O}_k(1) + i^{N-1}\Big(\Delta + \mathrm{O}_{N-1}\Big(\frac{1}{n}\Big)\Big) + \sum_{k=N}^{2N-1} i^k \mathrm{O}_k\Big(\frac{1}{n^{k-N+1}}\Big)\Bigg],$$

with $\Delta = N2^N > 0$ ■.

Remark 2. Note that the constant Δ equals $V(\mathcal{S}_{N-1})$ with our choice of the supremum norm.

Before proving Theorem 5.1, we still need a technical result.

Lemma 6.2

$$\lim_{n\to+\infty} \|H_n - G_n\|_\infty = 0, \tag{6.11}$$

where

$$H_n(z_1, z_2) = \sum_{i=1}^{+\infty} \delta_n^N \Bigg[\sum_{k=0}^{N-2} i^k \mathrm{O}_k(1) + i^{N-1}\Big(\Delta + \mathrm{O}_{N-1}\Big(\frac{1}{n}\Big)\Big)\Bigg] g_{i\delta_n}(z_1, z_2)$$

and

$$G_n(z_1, z_2) = \frac{1}{n^N}\sum_{i=1}^{n-1} \delta_n^N\, P_{n,i}\, g_{i\delta_n}(z_1, z_2).$$

Proof. In order to prove (6.11), note first that the mappings $\pi(u), \ldots, u^{2N-1}\pi(u)$ are decreasing for u large enough, $u > u_0$ say. Therefore, for $k = 0, \ldots, 2N-1$,

$$\sum_{i>u_0/\delta_n+1} \big|\delta_n (i\delta_n)^k g_{i\delta_n}(z_1, z_2)\big| \le \sum_{i>u_0/\delta_n+1} \delta_n (i\delta_n)^k \pi(i\delta_n)$$
$$\le \int_{u_0}^{+\infty} u^k \pi(u)\mathrm{d}u < +\infty\,.$$

Now we have

$$H_n - G_n = \sum_{i=n}^{+\infty} \delta_n^N \Big[\sum_{k=0}^{N-2} i^k \mathrm{O}_k(1) + i^{N-1} \Big(\Delta + \mathrm{O}_{N-1}\Big(\frac{1}{n}\Big)\Big)\Big] g_{i\delta_n}$$
$$- \sum_{i=1}^{n-1} \delta_n^N \sum_{k=N}^{2N-1} i^k \mathrm{O}_k\Big(\frac{1}{n^{k-N+1}}\Big) g_{i\delta_n}.$$

Hence

$$\|H_n - G_n\|_\infty \leq A \sum_{i=n}^{+\infty} \delta_n^N (1 + i + \ldots + i^{N-1}) \pi(i\delta_n)$$
$$+ \sum_{i=1}^{n-1} \delta_n^N \sum_{k=N}^{2N-1} i^k \Big| \mathrm{O}_k\Big(\frac{1}{n^{k-N+1}}\Big)\Big| \pi(i\delta_n),$$

where A is a positive constant. Therefore, for n large enough,

$$\|H_n - G_n\|_\infty \leq A \Big(\delta_n^{N-1} \int_{(n-1)\delta_n}^{+\infty} \pi(u)\mathrm{d}u + \ldots + \int_{(n-1)\delta_n}^{+\infty} u^{N-1}\pi(u)\mathrm{d}u \Big)$$
$$+ \sum_{i \leq u_0/\delta_n + 1} \delta_n^N \sum_{k=N}^{2N-1} i^k \Big| \mathrm{O}_k\Big(\frac{1}{n^{k-N+1}}\Big)\Big| \pi(i\delta_n)$$
$$+ \sum_{i > u_0/\delta_n + 1} \delta_n^N \sum_{k=N}^{2N-1} i^k \Big| \mathrm{O}_k\Big(\frac{1}{n^{k-N+1}}\Big)\Big| \pi(i\delta_n). \tag{6.12}$$

The first of the three terms above goes to 0 as $n \to +\infty$ since $\delta_n \to 0$, $n\delta_n \to +\infty$ and $\pi(u) + \ldots + u^{N-1}\pi(u)$ is integrable. With respect to the second term, we have

$$\sum_{i \leq u_0/\delta_n + 1} \delta_n^N \sum_{k=N}^{2N-1} i^k \Big| \mathrm{O}_k\Big(\frac{1}{n^{k-N+1}}\Big)\Big| \pi(i\delta_n)$$
$$\leq B \Big(\frac{1}{n\delta_n} \|u^N \pi(u)\|_\infty + \ldots + \frac{1}{n^N \delta_n^N} \|u^{2N-1}\pi(u)\|_\infty \Big),$$

where B is a positive constant. This expression tends to 0 since $n\delta_n \to +\infty$ as $n \to +\infty$. With respect to the third term (6.12), write

$$\sum_{i > u_0/\delta_n + 1} \delta_n^N \sum_{k=N}^{2N-1} i^k \Big| \mathrm{O}_k\Big(\frac{1}{n^{k-N+1}}\Big)\Big| \pi(i\delta_n)$$
$$\leq C \Big(\frac{1}{n\delta_n} \int_{u_0}^{+\infty} u^N \pi(u)\mathrm{d}u + \ldots + \frac{1}{n^N \delta_n^N} \int_{u_0}^{+\infty} u^{2N-1}\pi(u)\mathrm{d}u \Big),$$

where C is a positive constant. Hence (6.11) since $u^N\pi(u)+\ldots+u^{2N-1}\pi(u)$ is integrable and $n\delta_n \to +\infty$. ■

Proof of Theorem 5.1. The variance of f_n^* has the classical decomposition

$$\text{Var } f_n^*(z) = \tilde{V}_n + C_n,$$

where $\tilde{V}_n$ stands for the sum of variance and where

$$C_n = \frac{1}{n^N\delta_n^N}\int_{\mathbb{R}^{2d}} K_{h_n}(z-z_1)K_{h_n}(z-z_2)G_n(z_1,z_2)\mathrm{d}z_1\mathrm{d}z_2\,.$$

For $\tilde{V}_n$ we have again the well-known result (Bosq and Lecoutre [11])

$$n^N h_n^d \tilde{V}_n \to f(z)\int_{\mathbb{R}^d} K^2(u)\mathrm{d}u\,. \tag{6.13}$$

Concerning C_n, note that

$$\left| n^N\delta_n^N C_n - \int_{\mathbb{R}^{2d}} K_{h_n}(z-z_1)K_{h_n}(z-z_2)H_n(z_1,z_2)\mathrm{d}z_1\mathrm{d}z_2 \right| \le \|G_n - H_n\|_\infty$$

and

$$\left| \int_{\mathbb{R}^{2d}} K_{h_n}(z-z_1)K_{h_n}(z-z_2)\big[H_n(z_1,z_2)-G(z_1,z_2)\big]\mathrm{d}z_1\mathrm{d}z_2 \right| \le \|H_n - G\|_\infty,$$

where $G(z_1,z_2) = \Delta\displaystyle\int_0^{+\infty} u^{N-1}g_u(z_1,z_2)\mathrm{d}u$. It is easy to check that assumption (2) entails $\|H_n - G\|_\infty \to 0$. Therefore, using (6.11),

$$\left| n^N\delta_n^N C_n - \int_{\mathbb{R}^{2d}} K_{h_n}(z-z_1)K_{h_n}(z-z_2)G(z_1,z_2)\mathrm{d}z_1\mathrm{d}z_2 \right| \to 0.$$

Since G is continuous at (z,z), we find

$$n^N\delta_n^N C_n \to \Delta\int_0^{+\infty} u^{N-1}g_u(z,z)\mathrm{d}u\,. \tag{6.14}$$

Now the bias is given by (6.7) and (6.8). Using (6.13) and (6.14) we obtain

$$\mathbf{E}\big(f_n^*(z)-f(z)\big)^2 = \mathrm{O}\left(\frac{1}{n^N h_n^d}\right) + \mathrm{O}(h_n^{2r}) + \mathrm{O}\left(\frac{1}{n^N\delta_n^N}\right).$$

Hence, with the choice $h_n \sim n^{-N/(2r+d)}$

$$\mathbf{E}\big(f_n^*(z) - f(z)\big)^2 = \mathrm{O}\left(\frac{1}{n^{2rN/(2r+d)}}\right) + \mathrm{O}\left(\frac{1}{n^N \delta_n^N}\right).$$

Therefore, since $n\delta_n = T_n$,

$$\mathbf{E}\big(f_n^*(z) - f(z)\big)^2 = \mathrm{O}\left(\frac{1}{T_n^N}\right) + \mathrm{O}\left(\left(\frac{\delta_n}{T_n}\right)^{2rN/(2r+d)}\right)$$

and the full rate is obtained by choosing $\delta_n = T_n^{-d/(2r)}$.

It remains to prove that δ_n is maximal. To this aim, consider a sequence (δ_n') which generates the full rate and note that there exist $a_1 > 0$ and $a_2 > 0$ such that

$$\frac{a_1}{T_n^N} \geq \mathbf{E}\big(f^{**}(z) - f(z)\big)^2 \geq \frac{a_2}{n^{2rN/(2r+d)}} = a_2 \left(\frac{\delta_n'}{T_n}\right)^{2rN/(2r+d)}, \tag{6.15}$$

where f_n^{**} is associated with the grid step δ_n'. Then (6.15) yields

$$\delta_n' \leq \frac{a_1}{a_2} T_n^{-d/(2r)} = \mathrm{O}(\delta_n)$$

and the proof of Theorem 5.1 is therefore complete.

Proof of Corollary 5.1 is clear using (6.13), (6.14) and the fact that

$$h_n^{-4}\, \mathbf{E}\big(f_n^*(z) - f(z)\big)^2 \to \frac{1}{4}\left(\sum_{1 \leq i,j \leq d} \frac{\partial^2 f}{\partial z_i \partial z_j}(z) \int_{\mathbb{R}^d} u_i u_j K(u) \mathrm{d}u\right)^2.$$

Note that

$$C = \Delta \int_0^{+\infty} u^{N-1} g_u(z,z) \mathrm{d}u + \frac{f(z)}{c'^d} \|K\|_2^2$$
$$+ \frac{c'^2}{4}\left(\sum_{1 \leq i,j \leq d} \frac{\partial^2 f}{\partial z_i \partial z_j}(z) \int_{\mathbb{R}^d} u_i u_j K(u) \mathrm{d}u\right)^2,$$

where $h_n = c_n n^{-N/(4+d)}$ and $c_n \to c' > 0$. ■

Acknowledgments. The author greatly thank an Associate Editor and an anonymous referee for a careful reading of the paper and many valuable comments.

References

[1] Banon, G. and Nguyen, H.T. (1978). Sur l'estimation récurrente de la densité pour un processus de Markov, *Comptes Rendus de l'Académie des Sciences de Paris*, **Vol. 286**, pp. 691–694.

[2] Biau, G. (2002). Optimal asymptotic quadratic errors of density estimators on random fields, *Statistics and Probability Letters*, **Vol. 60**, pp. 297–307.

[3] Billingsley, P. (1986). *Probability and Measure – 2nd Edition*, Wiley, New York.

[4] Blanke, D. (1996). Estimation de la densité pour des trajectoires non directement observables, *Publications de l'Institut de Statistique de l'Université de Paris*, **Vol. 40**, pp. 21–36.

[5] Blanke, D. and Bosq, D. (2000). A family of minimax rates for density estimators in continuous time, *Stochastic Analysis and Applications*, **Vol. 18**, pp. 871–900.

[6] Blanke, D. and Pumo, B. (2003). Optimal sampling for density estimation in continuous time, *Journal of Time Series Analysis*, **Vol. 24**, pp. 1–23.

[7] Bosq, D. (1995). Optimal asymptotic quadratic error of density estimators for strong mixing or chaotic data, *Statistics and Probability Letters*, **Vol. 22**, pp. 339–347.

[8] Bosq, D. (1997). Parametric rates of nonparametric estimators and predictors for continuous time processes, *The Annals of Statistics*, **Vol. 25**, pp. 982–1000.

[9] Bosq, D. (1998). Minimax rates of density estimators for continuous time processes, *Sankhyā*, **Vol. 68**, pp. 18–28.

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

[11] Bosq, D. and Lecoutre, J.P. (1987). *Théorie de l'Estimation Fonctionnelle*, Economica, Paris.

[12] Bosq, D., Merlevède, F. and Peligrad, M. (1999). Asymptotic normality for density kernel estimators in discrete and continuous time, *Journal of Multivariate Analysis*, **Vol. 68**, pp. 78–95.

[13] Carbon, M., Hallin, M. and Tran, L.T. (1996). Kernel density estimation for random fields: the L_1 theory, *Journal of Nonparametric Statistics*, **Vol. 6**, pp. 157–170.

[14] Carbon, M., Tran, L.T. and Wu, B. (1997). Kernel density estimation for random fields, *Statistics and Probability Letters*, **Vol. 36**, pp. 115–125.

[15] Castellana, J.V. and Leadbetter, M.R. (1986). On smoothed probability density estimation for stationary processes, *Stochastic Processes and their Applications*, **Vol. 21**, pp. 179–193.

[16] Cheze-Payaud, N. (1994). Nonparametric regression and prediction for continuous-time processes, *Publications de l'Institut de Statistique de l'Université de Paris*, **Vol. 25**, pp. 17–39.

[17] Cressie, N.A.C. (1991). *Statistics for Spatial Data*, Wiley Series in Probability and Mathematical Statistics, New York.

[18] Delecroix, M. (1980). Sur l'estimation des densités d'un processus stationnaire à temps continu, *Publications de l'Institut de Statistique de l'Université de Paris*, **Vol. 38**, pp. 37–58.

[19] Doukhan, P. (1994). *Mixing – Properties and Examples*, Lecture Notes in Statistics, Springer–Verlag, New York.

[20] Guyon, X. (1995). *Random Fields on a Network – Modeling, Statistics, and Applications*, Springer–Verlag, New York.

[21] Hallin, M., Lu, Z. and Tran, L.T. (2001). Density estimation for spatial linear processes, *Bernoulli*, **Vol. 7**, pp. 657–668.

[22] Hallin, M., Lu, Z. and Tran, L.T. (2003). Kernel density estimation for spatial processes: the L_1 theory, *Journal of Multivariate Analysis*, to appear.

[23] Kutoyants, Y.A. (1997). Some problems of nonparametric estimation by observations of ergodic diffusion processes, *Statistics and Probability Letters*, **Vol. 32**, pp. 311–320.

[24] Leblanc, F. (1997). Density estimation for a class of continuous time processes, *Mathematical Methods of Statistics*, **Vol. 6**, pp. 171–199.

[25] Masry, E. (1983). Probability density estimation from sampled data, *IEEE Transactions on Information Theory*, **Vol. 29**, pp. 696–709.

[26] Prakasa Rao, B.L.S. (1990). Nonparametric density estimation for stochastic processes from sampled data, *Publications de l'Institut de Statistique de l'Université de Paris*, **Vol. 35**, pp. 51–83.

[27] Rio, E. (2000). *Théorie Asymptotique des Processus Aléatoires Faiblement Dépendants*, Springer–Verlag, Berlin.

[28] Ripley, B.D. (1981). *Spatial Statistics*, Wiley Series in Probability and Mathematical Statistics, New York.

[29] Rudin, W. (1987). *Real and Complex Analysis – 3rd Edition*, McGraw-Hill, New York.

[30] Tran, L.T. (1990). Kernel density estimation on random fields, *Journal of Multivariate Analysis*, **Vol. 34**, pp. 37–53.

[31] Tran, L.T. and Yakowitz, S. (1993). Nearest neighbor estimators for random fields, *Journal of Multivariate Analysis*, **Vol. 44**, pp. 23–46.

[32] Wackernagel, H. (1995). *Multivariate Geostatistics*, Springer–Verlag, Berlin.