Asymptotic normality of kernel type regression estimators for random fields

Zsolt Karácsony*

Department of Applied Mathematics, University of Miskolc, H-3515 Miskolc-Egyetemváros, Hungary

Peter Filzmoser

Department of Statistics and Probability Theory, Vienna University of Technology, Wiedner Hauptstraße 8-10, A-1040 Vienna, Austria

Abstract

The asymptotic normality of the Nadaraya-Watson regression estimator is studied for α-mixing random fields. The infill-increasing setting is considered, that is when the locations of observations become dense in an increasing sequence of domains. This setting fills the gap between continuous and discrete models. In the infill-increasing case the asymptotic normality of the Nadaraya-Watson estimator holds, but with an unusual asymptotic covariance structure. It turns out that this covariance structure is a combination of the covariance structures that we observe in the discrete and in the continuous case.

Key words: Central limit theorem, kernel, regression estimator, α-mixing, random field, asymptotic normality of estimators, infill asymptotics,

*Corresponding author

Email addresses: matkzs@uni-miskolc.hu (Zsolt Karácsony), P.Filzmoser@tuwien.ac.at (Peter Filzmoser)

increasing domain asymptotics

2000 MSC: 62G08, 60F05

---

1. Introduction

Kernel type regression estimators have been widely studied in the literature. The original results by Nadaraya (1964) and Watson (1964) have been extended in several papers, and they are summarized for example in Bosq (1998), Devroye and Györfi (1985), and Prakasa Rao (1983). One important issue for kernel type regression estimators is their asymptotic normality, which has been studied in several papers, like in Schuster (1972) and Cai (2001).

In this paper we consider $(X_{\mathbf{t}}, Y_{\mathbf{t}}), \mathbf{t} \in T_\infty$, to be a strictly stationary random field. (Here T_∞ is a domain in $\mathbb{R}^d$, $X_{\mathbf{t}}$ and $Y_{\mathbf{t}}$ are real-valued.) We want to estimate the regression function $r(x) = E\left(\Phi\left(Y_{\mathbf{t}}\right)|X_{\mathbf{t}} = x\right)$, where Φ is a known bounded measurable function. The data set is $(X_{\mathbf{t}}, Y_{\mathbf{t}}), \mathbf{t} \in \mathcal{D}_n$. We consider the well-known kernel type regression estimator

$$r_n(x) = \frac{\sum_{\mathbf{t}\in\mathcal{D}_n} \Phi(Y_{\mathbf{t}}) K\left(\frac{x - X_{\mathbf{t}}}{h}\right)}{\sum_{\mathbf{t}\in\mathcal{D}_n} K\left(\frac{x - X_{\mathbf{t}}}{h}\right)},$$

where K is a kernel function (see Nadaraya, 1964; Watson, 1964). However, our sampling scheme is unusual. The locations of observations become dense in an increasing sequence of domains. It is called the infill-increasing setting, see, for example, Lahiri et al. (1999) and Fazekas (2003). We suppose that the observed random field is weakly dependent, more precisely, the random field satisfies a certain α-mixing condition. The main result of this paper is that $r_n(x)$ is asymptotically normal with an unusual covariance structure.

That is, the asymptotic covariance matrix of $(r_n(x_1), \dots, r_n(x_m))$ is the sum of a diagonal matrix and a matrix containing integrals of the conditional covariances, see Theorem 1. Note that in the classical case for independent observations the joint asymptotic normality of $r_n(x_1), \dots, r_n(x_m)$ is well known (see Schuster, 1972).

The infill-increasing setting can be considered as a compromise of the continuous and the discrete case. However, an essential question is whether the limiting behavior of the continuous model is the same as that of its discrete counterpart. Estimators in the continuous case are mainly defined by integrals, in the discrete case they are defined by sums. However, when integrals are calculated numerically, approximating sums have to be applied. Therefore, considering an asymptotic result, not only the domain of the integration is increasing, but also the subdivision of the domain is more and more dense. In this situation one should check if the limiting behavior is the same as in the increasing domain setting. In this paper we show that these are the same only in special situations, but otherwise they can be different.

Note that the infill-increasing approach is substantially different from the pure infill setting. The infill setting means that the locations of the observations become more and more dense in a fixed domain (Cressie, 1991). In the infill case several well-known estimators are not consistent (Lahiri, 1996). Moreover, in the infill setup one cannot expect asymptotic normality of the estimators (because of the lack of the appropriate central limit theorems).

The infill-increasing approach can be useful in geosciences, meteorology, environmental studies, image processing, etc. In these sciences several processes varying continuously in time or in space are studied. However, in practice, we cannot observe the processes continuously in time or space. So we have to use finite data sets and discrete approximations. Moreover, the theoretical analysis of statistical models often requires simulation studies. In computer simulations discrete approximations are always applied.

Concerning the motivation of our studies we have to refer to the sampling schemes. Continuous time processes can be observed at deterministic or random time. Most of the existing results concern the non infill case (see Masry, 1983; Bosq and Cheze, 1993). In Bosq (1998) the importance of the sampling schemes is expressed, however no explicit result is mentioned for regression. In Bosq (1998, p. 140) only the following hint is given: "regression and density estimators behave alike when sampled data are available". Actually, for kernel type density estimators there are several results for infill-increasing type sampling schemes. We refer to Bosq (1998, pp. 118-127), Blanke-Pumo (2003), and Biau (2004).

This paper is organized as follows. In Section 2 the notation and the main result are presented. Theorem 1 states the asymptotic normality of the regression estimator $r_n(x)$. It is analogous to Theorem 1 of Fazekas and Chuprunov (2006) who proved the asymptotic normality of the kernel type density estimator in the same situation as in our paper. We quote it in Theorem A. In Remark 2 we compare our result with the existing ones. We show that the covariance structure given in Theorem 1 is a combination of the one in the discrete case (see Schuster, 1972) and the one in the continuous case (see Cheze, 1992).

The proof is given in Section 3. We apply the same method as in the proof of Theorem 1 of Fazekas and Chuprunov (2006). We use the notion of direct Riemann integrability presented in Fazekas and Chuprunov (2006). We apply a central limit theorem for random fields (Theorem 2.1 in Fazekas and Chuprunov, 2004, quoted in Theorem B). We also need the Rosenthal inequality for random fields (Theorem in Fazekas et al., 2000, quoted in Theorem C).

In Section 4 simulation results are presented. The numerical examples show the above mentioned unusual covariance structure of the limiting distribution.

2. Notation and the main result

The following notation is used. $\mathbb{Z}$ is the set of all integers, $\mathbb{Z}^d$ is the set of d-dimensional integer lattice points, where d is a fixed positive integer. $\mathbb{R}$ is the real line, $\mathbb{R}^d$ is the d-dimensional space with the usual Euclidean norm $\|\mathbf{x}\|$. In $\mathbb{R}^d$ we shall also consider the distance corresponding to the maximum norm,

$$\varrho(\mathbf{x}, \mathbf{y}) = \max_{1 \le i \le d} |x_i - y_i|$$

where $\mathbf{x} = (x_1, \ldots, x_d)$, $\mathbf{y} = (y_1, \ldots, y_d)$. The distance of two sets in $\mathbb{R}^d$ corresponding to the maximum norm is also denoted by ϱ : $\varrho(A, B) = \inf\{\varrho(\mathbf{x}, \mathbf{y}) : \mathbf{x} \in A, \mathbf{y} \in B\}$.

For real valued sequences $\{a_n\}$ and $\{b_n\}$, $a_n = \mathrm{o}(b_n)$ (resp. $a_n = \mathrm{O}(b_n)$) means that the sequence a_n/b_n converges to 0 (resp. is bounded). We will denote different constants with the same letter c. $\mathrm{I}\{A\}$ denotes the indicator function of the set A. $|\mathcal{D}|$ denotes the cardinality of the finite set $\mathcal{D}$ and at the same time $|T|$ denotes the volume of the domain T.

We shall suppose the existence of an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

The σ-algebra generated by a set of events or by a set of random variables will be denoted by $\sigma\{.\}$. The sign E stands for the expectation. The variance and the covariance are denoted by var(.) and cov(., .), respectively. The L_p-norm of a random (vector) variable η is defined as

$$\|\eta\|_p = \{E\|\eta\|^p\}^{1/p}, \quad 1 \le p < \infty.$$

The sign " $\Rightarrow$ " denotes convergence in distribution. $\mathcal{N}(m, \Sigma)$ stands for the (vector) normal distribution with mean (vector) m and covariance (matrix) Σ.

The scheme of observations is the following. For simplicity we restrict ourselves to rectangles as domains for the observations. Let $\Lambda > 0$ be fixed. By $\left(\frac{\mathbb{Z}}{\Lambda}\right)^d$ we denote the Λ-lattice points in $\mathbb{R}^d$, i.e. lattice points with distance $\frac{1}{\Lambda}$:

$$\left(\frac{\mathbb{Z}}{\Lambda}\right)^d = \left\{ \left(\frac{k_1}{\Lambda}, \dots, \frac{k_d}{\Lambda}\right) : (k_1, \dots, k_d) \in \mathbb{Z}^d \right\}.$$

T will be a bounded, closed rectangle in $\mathbb{R}^d$ with edges parallel to the axes, and $\mathcal{D}$ will denote the Λ-lattice points belonging to T, i.e. $\mathcal{D} = T \cap (\mathbb{Z}/\Lambda)^d$. For describing the limit distribution we consider a sequence of the previous objects. I.e. let $T_1, T_2, \dots$ be bounded, closed rectangles in $\mathbb{R}^d$. Suppose that $T_1 \subset T_2 \subset T_3 \subset \dots, \ \bigcup_{i=1}^{\infty} T_i = T_\infty$.

We assume that the length of each edge of T_n is an integer and converges to ∞, as $n \to \infty$ (e.g. $T_\infty = \mathbb{R}^d$ or $T_\infty = [0, \infty)^d$). Let $\{\Lambda_n\}$ be an increasing sequence of positive integers (the non-integer case is essentially the same) and let $\mathcal{D}_n$ be the Λ_n-lattice points belonging to T_n.

Let $\{\xi_{\mathbf{t}} = (X_{\mathbf{t}}, Y_{\mathbf{t}}), \mathbf{t} \in T_\infty\}$ be a strictly stationary two-dimensional random field. The n-th set of observations involves the values of the random field $(X_{\mathbf{t}}, Y_{\mathbf{t}})$ taken at each point $\mathbf{k} \in \mathcal{D}_n$. We shall construct the estimator from the data $(X_{\mathbf{k}}, Y_{\mathbf{k}}), \mathbf{k} \in \mathcal{D}_n$. Actually, each $\mathbf{k} = \mathbf{k}^{(n)}$ depends on n, but to avoid complicated notation we omit the superscript (n). By our assumptions, $\lim_{n\to\infty} |\mathcal{D}_n| = \infty$.

As the locations of the observations become more and more dense in an increasing sequence of domains, we call our setup infill-increasing (see Cressie, 1991; Lahiri, 1996, for infill asymptotics).

We need the notion of α-mixing (see, e.g. Doukhan, 1994; Guyon, 1995). Let $\mathcal{A}$ and $\mathcal{B}$ be two σ-algebras in $\mathcal{F}$. The α-mixing coefficient of $\mathcal{A}$ and $\mathcal{B}$ is defined as

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup\{|\mathbb{P}(A)\mathbb{P}(B) - \mathbb{P}(AB)| \ : \ A \in \mathcal{A}, B \in \mathcal{B}\}.$$

The α-mixing coefficient of $\{\xi_{\mathbf{t}} \ : \ \mathbf{t} \in T_\infty\}$ is

$$\alpha(r) = \sup\{\alpha(\mathcal{F}_{I_1}, \mathcal{F}_{I_2}) \ : \ \varrho(I_1, I_2) \geq r\},$$

where I_1 and I_2 are finite subsets in T_∞, $\mathcal{F}_{I_i} = \sigma\{\xi_{\mathbf{t}} \ : \ \mathbf{t} \in I_i\}$, $i = 1, 2$.

We list the conditions that will be used in our theorems.

Let

$$\int_0^\infty s^{2d-1} \alpha^{\frac{a-1}{a}}(s) ds < \infty, \quad \text{for some} \quad 1 < a < \infty. \tag{1}$$

A function $K : \mathbb{R} \to [0, \infty)$ will be called a kernel if K is a bounded, continuous, symmetric density function (with respect to the Lebesgue measure),

$$\lim_{|u|\to\infty} |u| K(u) = 0, \quad \int_{-\infty}^{\infty} u^2 K(u) du < \infty. \tag{2}$$

Let g be the (unknown) marginal density function of $X_{\mathbf{t}}$. For the sake of simplicity we assume that $g(x)$ is positive. Let K be a kernel and let $h_n > 0$, then the kernel-type (or Parzen-Rosenblatt-type) estimator of g is

$$g_n(x) = \frac{1}{|\mathcal{D}_n|}\frac{1}{h_n}\sum_{\mathbf{i}\in\mathcal{D}_n} K\left(\frac{x - X_{\mathbf{i}}}{h_n}\right), \quad x \in \mathbb{R}.$$

Our aim is to estimate the regression function

$$r(x) = E\left(\Phi\left(Y_{\mathbf{t}}\right)|X_{\mathbf{t}} = x\right),$$

where Φ is a bounded measurable function.

The usual kernel type estimator of $r(x)$ is

$$r_n(x) = \frac{\frac{1}{|\mathcal{D}_n|}\sum_{\mathbf{t}\in\mathcal{D}_n}\Phi(Y_{\mathbf{t}})\frac{1}{h}K\left(\frac{x-X_{\mathbf{t}}}{h}\right)}{\frac{1}{|\mathcal{D}_n|}\sum_{\mathbf{t}\in\mathcal{D}_n}\frac{1}{h}K\left(\frac{x-X_{\mathbf{t}}}{h}\right)},$$

where $h = h_n > 0$.

Let

$$a(x) = E\left(\Phi^2(Y_{\mathbf{t}})|X_{\mathbf{t}} = x\right).$$

Denote by $\mathbb{R}^d_{\mathbf{0}}$ the set $\mathbb{R}^d \setminus \{\mathbf{0}\}$. Let $g_{\mathbf{u}}(x, y)$ be the joint density function of $X_{\mathbf{0}}$ and $X_{\mathbf{u}}$, if $\mathbf{u} \in \mathbb{R}^d_{\mathbf{0}}$ and $x, y \in \mathbb{R}$. Let

$$a_{\mathbf{u}}(x, y) = E\big\{[\Phi(Y_{\mathbf{0}}) - r(X_{\mathbf{0}})]\,[\Phi(Y_{\mathbf{u}}) - r(X_{\mathbf{u}})]\,\big|X_{\mathbf{0}} = x, X_{\mathbf{u}} = y\big\}.$$

We shall assume that for each fixed $\mathbf{u}$ the functions

$a_{\mathbf{u}}(.,.), g_{\mathbf{u}}(.,.), a(.), r(.), g(.), r'(.), g'(.), r''(.), g''(.)$ are bounded and continuous. (3)

Furthermore we shall suppose that

$$\lim_{n\to\infty}\frac{1}{\Lambda_n^d h_n} = L < \infty, \ \lim_{n\to\infty}\Lambda_n = \infty \text{ and } \lim_{n\to\infty} h_n = 0 \tag{4}$$

and

$$\lim_{n\to\infty}|T_n|h_n^4 = 0. \tag{5}$$

Throughout the paper we concentrate on the case when $\xi_{\mathbf{t}}$ and $\xi_{\mathbf{s}}$ are dependent if $\mathbf{t}$ and $\mathbf{s}$ are close to each other.

First recall the asymptotic normality of the density estimator g_n.

Let $l_{\mathbf{u}}(x,y) = g_{\mathbf{u}}(x,y) - g(x)g(y), \mathbf{u} \in \mathbb{R}^d_{\mathbf{0}}$ and $x, y \in \mathbb{R}$. Let $l_{\mathbf{u}}$ denote $l_{\mathbf{u}}(x,y)$ as a function $l : \mathbb{R}^d_{\mathbf{0}} \to \mathcal{C}(\mathbb{R}^2)$, i.e. a function with values in $\mathcal{C}(\mathbb{R}^2)$ the space of continuous real-valued functions over $\mathbb{R}^2$. Let

$$\|l_{\mathbf{u}}\| = \sup_{(x,y)\in\mathbb{R}^2} |l_{\mathbf{u}}(x,y)| \tag{6}$$

be the norm of $l_{\mathbf{u}}$. Let $x_1, \dots, x_m$ be given distinct real numbers. Let $\Sigma_l = \left(\int_{\mathbb{R}^d_{\mathbf{0}}} l_{\mathbf{u}}(x_i, x_j) d\mathbf{u} \right)_{1\le i,j\le m}$ and let D' be a diagonal matrix with diagonal elements $Lg(x_i) \int_{-\infty}^{\infty} K^2(u)du$, $i = 1, \dots, m$. Let $\Sigma' = \Sigma_l + D'$.

Theorem **A.** *(Theorem 1 in Fazekas and Chuprunov, 2006)*
Assume that $l_{\mathbf{u}}$ *is Riemann integrable (as a function* $l : \mathbb{R}^d_{\mathbf{0}} \to \mathcal{C}(\mathbb{R}^2)$*) on each bounded closed d-dimensional rectangle* $R \subset \mathbb{R}^d_{\mathbf{0}}$*, moreover* $\|l_{\mathbf{u}}\|$ *is directly Riemann integrable (as a function* $\|l\| : \mathbb{R}^d_{\mathbf{0}} \to \mathbb{R}$*). Let* $x_1, \dots, x_m$ *be given distinct real numbers and assume that* Σ' *is positive definite. Suppose that there exists* $1 < a < \infty$ *such that* (1) *is satisfied and*

$$(h_n)^{-1} \le c|T_n|^{\frac{a^2}{(3a-1)(2a-1)}} \quad \text{for each } n. \tag{7}$$

If (4) *and* (5) *are satisfied then*

$$\sqrt{\frac{|\mathcal{D}_n|}{\Lambda_n^d}} \{(g_n(x_i) - g(x_i)), i = 1, \dots, m\} \Rightarrow \mathcal{N}(0, \Sigma') \quad \text{as} \quad n \to \infty. \tag{8}$$

Note that in the recent paper Park, Kim, Park and Hwang (2008) a similar phenomenon was described for a simpler dependence structure (*m*-dependence) but for a more general sampling scheme.

The notion of direct Riemann integrability can be found in Fazekas and Chuprunov (2006). Let $l : \mathbb{R}^d_{\mathbf{0}} \to [0, \infty)$ be given. For a $\delta > 0$ consider a partition of $\mathbb{R}^d$ into (right closed and left open) d-dimensional cubes $\Delta_{\mathbf{i}}$ with edge lengths δ such that the center of $\Delta_{\mathbf{0}}$ is the origin $\mathbf{0} \in \mathbb{R}^d$. The family $\{\Delta_{\mathbf{i}}\}$ is called the subdivision corresponding to δ. If $\mathbf{i} \neq \mathbf{0}$, for $\mathbf{x} \in \Delta_{\mathbf{i}}$ let $\bar{l}_\delta(\mathbf{x}) = \sup\{l(\mathbf{y}) : \mathbf{y} \in \Delta_{\mathbf{i}}\}$, $\underline{l}_\delta(\mathbf{x}) = \inf\{l(\mathbf{y}) : \mathbf{y} \in \Delta_{\mathbf{i}}\}$, while $\bar{l}_\delta(\mathbf{x}) = \underline{l}_\delta(\mathbf{x}) = 0$ if $\mathbf{x} \in \Delta_{\mathbf{0}}$. If

$$\lim_{\delta \to 0} \int_{\mathbb{R}^d} \bar{l}_\delta(\mathbf{x}) d\mathbf{x} = \lim_{\delta \to 0} \int_{\mathbb{R}^d} \underline{l}_\delta(\mathbf{x}) d\mathbf{x} = I$$

and this common value is finite, then l is called directly Riemann integrable (d.R.i.) and I is its direct Riemann integral.

If l is d.R.i., then l is bounded outside each neighborhood of the origin. Moreover, l is continuous almost everywhere (with respect to the Lebesgue measure). Therefore, l is Riemann integrable on each bounded closed d-dimensional rectangle not containing the origin. Call a zone a set $M = R_1 \setminus R_2$, where R_1 is a closed d-dimensional rectangle while R_2 ($\emptyset \neq R_2 \subset R_1$) is an open d-dimensional both rectangles having their centers at the origin. Then one obtains that l is Riemann integrable on each zone.

If $l \geq 0$ is d.R.i. then the improper integral $\int_{\mathbb{R}^d_{\mathbf{0}}} l(\mathbf{x}) d\mathbf{x}$ exists and it is equal to the direct Riemann integral of l. The above statement implies: for any $\varepsilon > 0$ there exists a zone M such that $\int_{\mathbb{R}^d_{\mathbf{0}} \setminus M} l(\mathbf{x}) d\mathbf{x} \leq \varepsilon$.

Finally, we have the following. Let $l \geq 0$ be d.R.i. Let δ_n be positive numbers converging to zero, and let $\{\Delta^{(n)}_{\mathbf{i}}\}$ be the subdivision corresponding to δ_n. Then for any $\varepsilon > 0$ there exist a zone M such that all Riemannian approximating sums (based on the above subdivisions but not containing term $|\Delta_{\mathbf{0}}| l(\mathbf{x_0})$) of the integral $\int_{\mathbb{R}^d_{\mathbf{0}} \setminus M} l(\mathbf{x}) d\mathbf{x}$ are less than ε.

For the definition of the Riemann integrability of a Banach space valued function, see Hille and Phillips (1957, p. 62).

Now we can state our main result. Let $v(x) = a(x) - r^2(x)$.

For a fixed positive integer m and fixed distinct real numbers $x_1, x_2, \ldots, x_m$ we introduce the notation

$$\sigma(x_t, x_s) = \int_{\mathbb{R}_{\mathbf{0}}^d} a_{\mathbf{u}}(x_t, x_s) g_{\mathbf{u}}(x_t, x_s) d\mathbf{u}, \quad t, s = 1, \ldots, m, \tag{9}$$

$$\Sigma^{(m)} = \left(\frac{\sigma(x_t, x_s)}{g(x_t)g(x_s)} \right)_{1 \le t, s \le m}. \tag{10}$$

We assume that

$$\lim_{|z| \to \infty} z^3 |K(z)| = 0. \tag{11}$$

THEOREM **1.** *Let* $(X_{\mathbf{t}}, Y_{\mathbf{t}})$, $\mathbf{t} \in T_\infty$, *be a strictly stationary two-dimensional random field and let* $r(x) = E(\Phi(Y_{\mathbf{t}}) | X_{\mathbf{t}} = x)$ *be the regression function, where* Φ *is a bounded measurable function. Let* K *be a kernel. Assume that the conditions of Theorem A on the function* $l_{\mathbf{u}}$ *are satisfied, and that* Σ' *is positive definite. Furthermore, assume that the marginal density function of* $X_{\mathbf{t}}$ *is positive, and that* $a_{\mathbf{u}} g_{\mathbf{u}}$ *is Riemann integrable (as a function* $a \cdot g : \mathbb{R}_{\mathbf{0}}^d \to \mathcal{C}(\mathbb{R}^2)$*) on each bounded closed d-dimensional rectangle* $R \subset \mathbb{R}_{\mathbf{0}}^d$*. Moreover,* $\|a_{\mathbf{u}} g_{\mathbf{u}}\|$ *where the norm is a similar in* (6)*) is directly Riemann integrable (as a function* $\|a \cdot g\| : \mathbb{R}_{\mathbf{0}}^d \to \mathbb{R}$*). Suppose there exists* $1 < a < \infty$ *such that* (1) *and* (7) *are satisfied. Assume that the matrix* $\Sigma^{(m)} + D$ *is positive definite where* D *is a diagonal matrix with diagonal elements* $Lv(x_i) \int_{-\infty}^{\infty} K^2(t) dt / g(x_i)$, $i = 1, \ldots, m$*. If the conditions* (3), (4), (5) *and* (11) *hold then*

$$\sqrt{\frac{|\mathcal{D}_n|}{\Lambda_n^d}} \{ (r_n(x_i) - r(x_i)), i = 1, \ldots, m \} \Rightarrow \mathcal{N}(0, \Sigma), \quad \text{as} \quad n \to \infty,$$

where

$$\Sigma = \Sigma^{(m)} + D.$$

Remark **2.** We show that the asymptotic covariance matrix Σ in Theorem 1 is a combination of the asymptotic covariance matrices in the discrete and the continuous cases. In Schuster (1972) it is shown that (for independent identically distributed observations) $r_n(x_1), \dots, r_n(x_m)$ is asymptotically normal with diagonal covariance matrix. In particular, $\sqrt{nh_n}(r_n(x_i) - r(x_i)) \Rightarrow \mathcal{N}(0, c_i)$, where $c_i = v(x_i)\int_{-\infty}^{\infty} K^2(t)dt/g(x_i)$. Therefore, in Theorem 1 the diagonal part D corresponds to the limiting covariance matrix in the discrete case.

Now calculate the elements of $\sigma(x_t, x_s)$. Denote by $f_{X_{\mathbf{0}},X_{\mathbf{u}},Y_{\mathbf{0}},Y_{\mathbf{u}}}(x_1, x_2, y_1, y_2)$ the joint density function of $X_{\mathbf{0}}, X_{\mathbf{u}}, Y_{\mathbf{0}}, Y_{\mathbf{u}}, (\mathbf{u} \neq \mathbf{0})$. Then we obtain

$$\begin{aligned} a_{\mathbf{u}}(x_1, x_2) &= \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [\Phi(y_1) - r(x_1)]\,[\Phi(y_2) - r(x_2)]\, f_{X_{\mathbf{0}},X_{\mathbf{u}},Y_{\mathbf{0}},Y_{\mathbf{u}}}(x_1, x_2, y_1, y_2)dy_1dy_2}{g_{\mathbf{u}}(x_1, x_2)} \\ &= \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} M_{\mathbf{u}}(x_1, x_2, y_1, y_2)dy_1dy_2}{g_{\mathbf{u}}(x_1, x_2)}. \end{aligned}$$

Therefore (considering the case $d = 1$)

$$\sigma(x_t, x_s) = \int_{\mathbb{R}_{\mathbf{0}}} \left[\int_{\infty}^{-\infty} \int_{-\infty}^{\infty} M_{\mathbf{u}}(x_1, x_2, y_1, y_2)dy_1dy_2 \right] d\mathbf{u}.$$

In Cheze (1992) and Bosq (1998), p. 138, the kernel regression estimator was considered if $(X_t, Y_t), t \in [0, T]$ is a continuous time stochastic process (with certain α-mixing condition). The estimator of the regression function $r(x) = E(\Phi(Y)|X = x)$ is given by

$$\widetilde{r}_T(x) = \frac{\widetilde{\varphi}_T(x)}{\widetilde{g}_T(x)} \tag{12}$$

where

$$\widetilde{\varphi}_T(x) = \frac{1}{T}\int_0^T \Phi(Y_t)\frac{1}{h_T}K\left(\frac{x-X_t}{h_T}\right)dt, \ \ \widetilde{g}_T(x) = \frac{1}{T}\int_0^T \frac{1}{h_T}K\left(\frac{x-X_t}{h_T}\right)dt.$$

Under some conditions, if $T \to \infty$ and $h_T \to 0$, then $\widetilde{r}_T$ is asymptotically normal. More precisely,

$$\frac{\widetilde{r}_T(x) - r(x)}{\sqrt{d_T(x)}} \Rightarrow \mathcal{N}(0,1)$$

where

$$g^2(x)d_T(x) = (1, -r(x))\,\mathrm{var}\begin{pmatrix} \widetilde{\varphi}_T(x) \\ \widetilde{g}_T(x) \end{pmatrix}\begin{pmatrix} 1 \\ -r(x) \end{pmatrix}.$$

Using the above expression (and assuming certain analytical conditions) we can see that the limit of $Td_T(x)$ is $\sigma(x,x)/g^2(x)$. Thus the result in Cheze (1992) and Bosq (1998) can be formulated as

$$\sqrt{T}\left(\widetilde{r}_T(x) - r(x)\right) \Rightarrow \mathcal{N}\left(0, \sigma(x,x)/g^2(x)\right).$$

Therefore the diagonal elements of our matrix $\Sigma^{(m)}$ correspond to the limiting variances in the continuous case. (In Cheze (1992) and Bosq (1998) joint asymptotic normality of $(\widetilde{r}_T(x_1), \ldots, \widetilde{r}_T(x_m))$ is not studied.)

Remark **3.** If condition (5), i.e. $\lim_{n\to\infty}|T_n|h_n^4 = 0$ is not satisfied, we can prove that

$$\sqrt{\frac{|\mathcal{D}_n|}{\Lambda_n^d}}\{(r_n(x_i) - \widehat{r_n}(x_i)), i = 1, \ldots, m\} \Rightarrow \mathcal{N}(0, \Sigma), \quad \text{as} \quad n \to \infty,$$

where $\widehat{r_n}(x) = \frac{1}{|\mathcal{D}_n|}\sum_{\mathbf{t}\in\mathcal{D}_n} r(X_{\mathbf{t}})\frac{1}{h}K\left(\frac{x-X_{\mathbf{t}}}{h}\right)/g_n(x)$, and
$g_n(x) = \frac{1}{|\mathcal{D}_n|}\sum_{\mathbf{t}\in\mathcal{D}_n}\frac{1}{h}K\left(\frac{x-X_{\mathbf{t}}}{h}\right) \to g(x)$ in probability. It is a consequence of the proof of Theorem 1.

REMARK **4.** Note that in Bosq (1998), p. 140, and in Bosq (1997) the problem of sampling is also studied, i.e., the behavior of the approximation of $\widetilde{r}_T$ in (12) with the corresponding discrete expression, if the process is observed at time instants $\delta_n, 2\delta_n, \dots, n\delta_n$. However, the asymptotic normality is not considered in the above mentioned papers.

3. Proof of the Main Theorem

To prove the main theorem we need the following central limit theorem and a version of the Rosenthal inequality for mixing fields.

First, define the discrete parameter (vector valued) random field $Y_n(\mathbf{k})$ as follows. For each $n = 1, 2, \dots,$ and for each $\mathbf{k} = \mathbf{k}^{(n)} \in \mathcal{D}_n$,

$$\text{let } \mathrm{Y}_n(\mathbf{k}) \text{ be a Borel measurable function of } \xi_{\mathbf{k}^{(n)}} \tag{13}$$

where $\{\xi_\mathbf{t}, \mathbf{t} \in T_\infty\}$ is the underlying random field.

THEOREM **B.** *(Theorem 2.1 in Fazekas and Chuprunov, 2004)*
Let $\xi_\mathbf{t}$ be a random field and let $Y_n(\mathbf{k}) = (Y_n^{(1)}(\mathbf{k}), \dots, Y_n^{(m)}(\mathbf{k}))$ be an m-dimensional random field defined by (13)*. Let $S_n = \sum_{\mathbf{k}\in\mathcal{D}_n} Y_n(\mathbf{k})$, $n = 1, 2, \dots$. Suppose that for each fixed n the field $Y_n(\mathbf{k}), \mathbf{k} \in \mathcal{D}_n$, is strictly stationary with $EY_n(\mathbf{k}) = 0$. Assume that*

$$\|Y_n(\mathbf{k})\| \le M_n, \tag{14}$$

where M_n depends only on n;

$$\sup_{n,\mathbf{k},t} E(Y_n^{(t)}(\mathbf{k}))^2 < \infty; \tag{15}$$

for any increasing, unbounded sequence of rectangles G_n with $G_n \subseteq T_n$

$$\lim_{n\to\infty} \frac{1}{\Lambda_n^d|\mathcal{G}_n|} E\left[\sum_{\mathbf{k}\in\mathcal{G}_n} Y_n^{(t)}(\mathbf{k}) \cdot \sum_{\mathbf{l}\in\mathcal{G}_n} Y_n^{(s)}(\mathbf{l})\right] = \sigma_{ts}, \quad t,s = 1,\ldots,m, \tag{16}$$

where $\mathcal{G}_n = G_n \cap (\mathbb{Z}/\Lambda_n)^d$; the matrix $\Sigma = (\sigma_{ts})_{t,s=1}^m$ is positive definite; there exists $1 < a < \infty$ such that (1) is satisfied; and

$$M_n \le c|T_n|^{\frac{a^2}{(3a-1)(2a-1)}} \quad \text{for each } n. \tag{17}$$

Then

$$\frac{1}{\sqrt{\Lambda_n^d|\mathcal{D}_n|}} S_n \Rightarrow \mathcal{N}(0,\Sigma), \quad \text{as} \quad n \to \infty. \tag{18}$$

In the proof of the main theorem we also use the following form of the Rosenthal inequality for mixing fields.

Theorem **C.** *(Theorem in Fazekas et al., 2000)*
Let $1 < l \le 2$ and $\tau > 0$. Let $Y_{\mathbf{k}}$, $\mathbf{k} \in \mathbb{Z}^d$, be centered random variables with $E|Y_{\mathbf{k}}|^{l+\tau} < \infty$, $\mathbf{k} \in \mathbb{Z}^d$. Introduce the notation

$$L(l,\tau,\mathcal{D}) = \sum_{\mathbf{k}\in\mathcal{D}} \left(E|Y_{\mathbf{k}}|^{l+\tau}\right)^{\frac{l}{l+\tau}},$$

if $\mathcal{D}$ is a finite set in $\mathbb{Z}^d$. Let

$$c_{1,1}^{(\tau)} = 1 + \sum_{s=1}^{\infty} s^{d-1}[\alpha_Y(s,1,1)]^{\frac{\tau}{2+\tau}},$$

where $\alpha_Y(s,1,1)$ is the α-mixing coefficient of the field $\{Y_{\mathbf{k}}\}$, i.e. $\alpha_Y(s,1,1) = \sup\{\alpha(Y_{\mathbf{u}},Y_{\mathbf{v}}) : \varrho(\mathbf{u},\mathbf{v}) \ge s\}$. Assume that $c_{1,1}^{(\tau)} < \infty$. Then there is a constant c such that

$$E\left|\sum\nolimits_{\mathbf{k}\in\mathcal{D}} Y_{\mathbf{k}}\right|^l \le c \cdot c_{1,1}^{(\tau)} L(l,\tau,\mathcal{D}), \tag{19}$$

for any finite subset $\mathcal{D}$ of $\mathbb{Z}^d$.

Details and the general form of the Rosenthal inequality can be found e.g. in Fazekas et al. (2000).

In the proof of the main theorem we will use the next theorem several times. This is a particular case of Theorem 2.1.1 in Prakasa Rao (1983).

THEOREM **D.** *(Theorem 2.1.1 in Prakasa Rao, 1983)*
Let $K : \mathbb{R} \to \mathbb{R}$ *be measurable such that*

$$|K(z)| \leq M, \quad z \in \mathbb{R},$$
$$\int_{-\infty}^{\infty} |K(z)| dz < \infty,$$

and

$$|z||K(z)| \to 0 \ \text{ as } \ |z| \to \infty.$$

Furthermore, let $g : \mathbb{R} \to \mathbb{R}$ *be measurable such that*

$$\int_{-\infty}^{\infty} |g(z)| dz < \infty.$$

Define

$$g_n(x) = \frac{1}{h_n} \int_{-\infty}^{\infty} K\left(\frac{z}{h_n}\right) g(x-z) dz$$

where $0 < h_n \to 0$ *as* $n \to \infty$*. Then, if* g *is continuous,*

$$\lim_{n\to\infty} g_n(x) = g(x) \int_{-\infty}^{\infty} K(z) dz, \tag{20}$$

and if g *is uniformly continuous, then the convergence in* (20) *is uniform.*

REMARK **5.** We shall often use the next limit relations (see, e.g., Fazekas and Chuprunov, 2006). Assume that the density function g is continuous, K is a kernel, then as $h_n \to 0 \ \ (h_n > 0)$ we have the following.

$$E\left(\frac{1}{h_n} K\left(\frac{x - X_{\mathbf{t}}}{h_n}\right)\right) = \int_{-\infty}^{\infty} \frac{1}{h_n} K\left(\frac{x-u}{h_n}\right) g(u) du \to g(x), \tag{21}$$

$$E\frac{1}{h_n}K^2\left(\frac{x-X_{\mathbf{t}}}{h_n}\right)=\int_{-\infty}^{\infty}\frac{1}{h_n}K^2\left(\frac{x-u}{h_n}\right)g(u)du\to g(x)\int_{-\infty}^{\infty}K^2\left(u\right)du, \tag{22}$$

$$E\frac{1}{h_n^2}K\left(\frac{x_r-X_{\mathbf{t}}}{h_n}\right)K\left(\frac{x_s-X_{\mathbf{t}}}{h_n}\right)=\int_{-\infty}^{\infty}\frac{1}{h_n^2}K\left(\frac{x_r-u}{h_n}\right)K\left(\frac{x_s-u}{h_n}\right)g(u)du\to 0, \tag{23}$$

if $x_r \neq x_s$.

PROOF OF THEOREM 1. Consider the following decomposition

$$\sqrt{\frac{|\mathcal{D}_n|}{\Lambda_n^d}}\left(r_n(x)-r(x)\right)=\sqrt{\frac{|\mathcal{D}_n|}{\Lambda_n^d}}\frac{\frac{1}{|\mathcal{D}_n|}\sum_{\mathbf{t}\in\mathcal{D}_n}\left[\Phi(Y_{\mathbf{t}})-r(x)\right]\frac{1}{h}K\left(\frac{x-X_{\mathbf{t}}}{h}\right)}{\frac{1}{|\mathcal{D}_n|}\sum_{\mathbf{t}\in\mathcal{D}_n}\frac{1}{h}K\left(\frac{x-X_{\mathbf{t}}}{h}\right)}$$

$$=\frac{\frac{1}{\sqrt{|\mathcal{D}_n|\Lambda_n^d}}\frac{1}{h}\left[\sum_{\mathbf{t}\in\mathcal{D}_n}\left[\Phi(Y_{\mathbf{t}})-r(X_{\mathbf{t}})\right]K\left(\frac{x-X_{\mathbf{t}}}{h}\right)+\sum_{\mathbf{t}\in\mathcal{D}_n}\left[r(X_{\mathbf{t}})-r(x)\right]K\left(\frac{x-X_{\mathbf{t}}}{h}\right)\right]}{\frac{1}{|\mathcal{D}_n|}\sum_{\mathbf{t}\in\mathcal{D}_n}\frac{1}{h}K\left(\frac{x-X_{\mathbf{t}}}{h}\right)}$$

$$=\frac{J_1(x)+J_2(x)}{J_3(x)},$$

where

$$J_1(x)=\frac{1}{\sqrt{|\mathcal{D}_n|\Lambda_n^d}}\sum_{\mathbf{t}\in\mathcal{D}_n}\frac{1}{h}\left[\Phi(Y_{\mathbf{t}})-r(X_{\mathbf{t}})\right]K\left(\frac{x-X_{\mathbf{t}}}{h}\right),$$

$$J_2(x)=\frac{1}{\sqrt{|\mathcal{D}_n|\Lambda_n^d}}\sum_{\mathbf{t}\in\mathcal{D}_n}\frac{1}{h}\left[r(X_{\mathbf{t}})-r(x)\right]K\left(\frac{x-X_{\mathbf{t}}}{h}\right),$$

and

$$J_3(x)=\frac{1}{|\mathcal{D}_n|}\sum_{\mathbf{t}\in\mathcal{D}_n}\frac{1}{h}K\left(\frac{x-X_{\mathbf{t}}}{h}\right).$$

First we prove the asymptotic normality of J_1. We have to check the conditions of Theorem B.

Let $x_1, x_2, \dots, x_m$ be fixed distinct real numbers. We need to prove the joint asymptotic normality of $J_1 = (J_1(x_1), J_1(x_2), \dots, J_1(x_m))^\top$. Define the m-dimensional random vector $Z_n(\mathbf{i})$ with the following coordinates:

$$Z_n^{(s)}(\mathbf{i}) = \frac{1}{h}\left[\Phi(Y_\mathbf{i}) - r(X_\mathbf{i})\right] K\left(\frac{x_s - X_\mathbf{i}}{h}\right),$$

for $s = 1, \dots, m$ and $\mathbf{i} \in \mathcal{D}_n$.

Divide T_n into d-dimensional unit cubes (having Λ_n^d points of $\mathcal{D}_n$ in each of them). Denote $\mathcal{D}_n^{'}$ the set of these cubes. Let $V_n(\mathbf{k}) = (V_n^{(1)}(\mathbf{k}), \dots, V_n^{(m)}(\mathbf{k}))$ be the arithmetical mean of the variables $Z_n(\mathbf{i})$ having indices $\mathbf{i}$ in the $\mathbf{k}$-th unit cube. Then for each fixed n the field $V_n(\mathbf{k})$, $\mathbf{k} \in \mathcal{D}_n^{'}$, is strictly stationary. We shall apply Theorem B to $V_n(\mathbf{k})$, $\mathbf{k} \in \mathcal{D}_n^{'}$, i.e. we shall use a non-infill form of that theorem. We have

$$J_1(x_s) = \frac{1}{\sqrt{|\mathcal{D}_n|\Lambda_n^d}} \Lambda_n^d \sum_{\mathbf{i}\in\mathcal{D}_n^{'}} V_n^{(s)}(\mathbf{i}) = \sqrt{\frac{\Lambda_n^d}{|\mathcal{D}_n|}} \sum_{\mathbf{i}\in\mathcal{D}_n^{'}} V_n^{(s)}(\mathbf{i}).$$

To see that $EV_n(\mathbf{k}) = 0$ consider

$$EZ_n^{(s)}(\mathbf{i}) = E\left(\frac{1}{h}\left[\Phi(Y_\mathbf{i}) - r(X_\mathbf{i})\right] K\left(\frac{x_s - X_\mathbf{i}}{h}\right)\right) = 0,$$

because $E\left(\Phi(Y)K\left(\frac{x-X}{h}\right)\right) = E\Big[\underbrace{E\left\{\Phi(Y)|X\right\}}_{r(X)} K\left(\frac{x-X}{h}\right)\Big] = E\left(r(X)K\left(\frac{x-X}{h}\right)\right)$.

Since Φ, r and K are bounded, equation (7) implies (14) and (17).

To prove (15), we have to consider

$$E\left(V_n^{(s)}(\mathbf{k})\right)^2 = E\left(\frac{1}{\Lambda_n^d}\sum_{\mathbf{i}} \frac{1}{h}\left[\Phi(Y_\mathbf{i}) - r(X_\mathbf{i})\right] K\left(\frac{x_s - X_\mathbf{i}}{h}\right)\right)^2$$

where $\sum_{\mathbf{i}}$ means that $\mathbf{i}$ belongs to the $\mathbf{k}$-th unit cube. The boundedness of this expression can be checked similarly to the next proof (showing that condition (16) is satisfied).

To calculate the limit in (16), let $\{G_n\}$ be an increasing sequence of d-dimensional rectangles, each G_n being a union of d-dimensional unit cubes. Then

$$\frac{1}{|G_n|}E\left[\sum_{\mathbf{k}\in G_n\cap\mathbb{Z}^d} V_n^{(t)}(\mathbf{k})\cdot\sum_{\mathbf{l}\in G_n\cap\mathbb{Z}^d} V_n^{(s)}(\mathbf{l})\right]$$

$$=\frac{1}{\Lambda_n^d|\mathcal{G}_n|}\sum_{\mathbf{i}\in\mathcal{G}_n}\sum_{\mathbf{j}\in\mathcal{G}_n}\frac{1}{h^2}E\left[\left[\Phi(Y_{\mathbf{i}})-r(X_{\mathbf{i}})\right]K\left(\frac{x_t-X_{\mathbf{i}}}{h}\right)\left[\Phi(Y_{\mathbf{j}})-r(X_{\mathbf{j}})\right]K\left(\frac{x_s-X_{\mathbf{j}}}{h}\right)\right]$$

$$=A+B,$$

where $\mathcal{G}_n=G_n\cap(\mathbb{Z}/\Lambda_n)^d$, and A denotes the part of the sum with $\mathbf{i}=\mathbf{j}$, while B denotes the part of the sum with $\mathbf{i}\neq\mathbf{j}$.

For A we have

$$A=\frac{1}{\Lambda_n^d|\mathcal{G}_n|}\frac{1}{h}\sum_{\mathbf{i}\in\mathcal{G}_n}E\left[\frac{1}{h}\left[\Phi(Y_{\mathbf{i}})-r(X_{\mathbf{i}})\right]^2K\left(\frac{x_t-X_{\mathbf{i}}}{h}\right)K\left(\frac{x_s-X_{\mathbf{i}}}{h}\right)\right].$$

If $t=s$ we obtain

$$A=\frac{1}{\Lambda_n^d|\mathcal{G}_n|}\frac{1}{h}\sum_{\mathbf{i}\in\mathcal{G}_n}E\left[\frac{1}{h}\left[\Phi(Y_{\mathbf{i}})-r(X_{\mathbf{i}})\right]^2K^2\left(\frac{x_s-X_{\mathbf{i}}}{h}\right)\right].$$

We have

$$E\left[\frac{1}{h}\left[\Phi(Y_{\mathbf{i}})-r(X_{\mathbf{i}})\right]^2K^2\left(\frac{x_s-X_{\mathbf{i}}}{h}\right)\right]$$

$$=\underbrace{E\left(\frac{1}{h}\Phi^2(Y_{\mathbf{i}})K^2\left(\frac{x_s-X_{\mathbf{i}}}{h}\right)\right)}_{*}-\underbrace{E\left(\frac{1}{h}r^2(X_{\mathbf{i}})K^2\left(\frac{x_s-X_{\mathbf{i}}}{h}\right)\right)}_{**}.$$

We calculate * and ** one after another:

$$*=E\left[E\left(\frac{1}{h}\Phi^2(Y_{\mathbf{i}})\,K^2\left(\frac{x_s-X_{\mathbf{i}}}{h}\right)\middle|\,X_{\mathbf{i}}\right)\right]=E\left(\frac{1}{h}K^2\left(\frac{x_s-X_{\mathbf{i}}}{h}\right)a(X_{\mathbf{i}})\right),$$

where $a(x) = E\left(\Phi^2(Y)|X = x\right)$. Therefore, by (22),

$$* = \int_{-\infty}^{\infty} a(u)\frac{1}{h}K^2\left(\frac{x_s - u}{h}\right)g(u)du = \int_{-\infty}^{\infty} a(x_s - ht)g(x_s - ht)K^2(t)\,dt$$

$$\to a(x_s)g(x_s)\int_{-\infty}^{\infty} K^2(t)\,dt, \text{ if } h \to 0,$$

because a and g are bounded and continuous and K^2 is integrable.

Similarly we get

$$** = E\left(\frac{1}{h}r^2(X_{\mathbf{i}})K^2\left(\frac{x_s - X_{\mathbf{i}}}{h}\right)\right) = \int_{-\infty}^{\infty}\frac{1}{h}r^2(u)K^2\left(\frac{x_s - u}{h}\right)g(u)du$$

$$= \int_{-\infty}^{\infty} r^2(x_s - ht)K^2(t)\,g(x_s - ht)dt \to r^2(x_s)g(x_s)\int_{-\infty}^{\infty} K^2(t)\,dt, \text{ if } h \to 0,$$

because r and g are bounded and continuous and K^2 is integrable.

Applying (4), we have

$$A \simeq \frac{1}{\Lambda_n^d h_n}\frac{1}{|\mathcal{G}_n|}\sum_{\mathbf{i}\in\mathcal{G}_n}\left[a(x_s) - r^2(x_s)\right]g(x_s)\int_{-\infty}^{\infty} K^2(t)\,dt$$

$$\simeq Lv(x_s)g(x_s)\int_{-\infty}^{\infty} K^2(t)\,dt,$$

where $v(x_s) = a(x_s) - r^2(x_s)$.

We remind that $v(x)$ is the conditional variance of $\Phi(Y)$, that is

$$v(x) = E\left(\Phi^2(Y)|X = x\right) - \left[E\left(\Phi(Y)|X = x\right)\right]^2$$

$$= E\left\{\left[\Phi(Y) - E\left(\Phi(Y)|X = x\right)\right]^2|X = x\right\}.$$

If $t \neq s$ we obtain

$$A = \frac{1}{\Lambda_n^d}\frac{1}{h_n^2}E\left(\left[\Phi(Y_{\mathbf{i}}) - r(X_{\mathbf{i}})\right]^2 K\left(\frac{x_t - X_{\mathbf{i}}}{h}\right)K\left(\frac{x_s - X_{\mathbf{i}}}{h}\right)\right)$$

$$= \frac{1}{\Lambda_n^d}\left[E\left(\frac{1}{h^2}a(X_{\mathbf{i}})K\left(\frac{x_t - X_{\mathbf{i}}}{h}\right)K\left(\frac{x_s - X_{\mathbf{i}}}{h}\right)\right)\right.$$

$$-E\left(\frac{1}{h^2}r^2(X_{\mathbf{i}})K\left(\frac{x_t-X_{\mathbf{i}}}{h}\right)K\left(\frac{x_s-X_{\mathbf{i}}}{h}\right)\right)\Bigg] = \frac{1}{\Lambda_n^d}(A_1+A_2).$$

The boundedness of $a(x)$ and $r(x)$ and (23) imply that

$$|A_1|, |A_2| \leq c\int_{-\infty}^{\infty}\frac{1}{h^2}K\left(\frac{x_t-u}{h}\right)K\left(\frac{x_s-u}{h}\right)g(u)du \to 0.$$

Thus, for $t \neq s$ we obtain that $A \to 0$, as $\Lambda_n^d \to \infty$.

Now turn to B.

$$B = \frac{1}{\Lambda_n^d}\frac{1}{|\mathcal{G}_n|}\sum_{\mathbf{i}\neq\mathbf{j}} E\left(a(X_{\mathbf{i}}, X_{\mathbf{j}})\frac{1}{h^2}K\left(\frac{x_t-X_{\mathbf{i}}}{h}\right)K\left(\frac{x_s-X_{\mathbf{j}}}{h}\right)\right),$$

where

$$a(X_{\mathbf{i}}, X_{\mathbf{j}}) = a_{\mathbf{i}-\mathbf{j}}(X_{\mathbf{i}}, X_{\mathbf{j}}) = E\left\{[\Phi(Y_{\mathbf{i}})-r(X_{\mathbf{i}})]\,[\Phi(Y_{\mathbf{j}})-r(X_{\mathbf{j}})]\,|X_{\mathbf{i}}, X_{\mathbf{j}}\right\}.$$

Therefore

$$B = \frac{1}{\Lambda_n^d}\frac{1}{|\mathcal{G}_n|}\sum_{\mathbf{i}\neq\mathbf{j}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} a_{\mathbf{i}-\mathbf{j}}(u,v)\frac{1}{h^2}K\left(\frac{x_t-u}{h}\right)K\left(\frac{x_s-v}{h}\right)g_{\mathbf{i}-\mathbf{j}}(u,v)dudv,$$

where $g_{\mathbf{i}-\mathbf{j}}(u,v)$ is the joint density function of $X_{\mathbf{i}}$ and $X_{\mathbf{j}}$.

As the random field is strictly stationary, we can assume that the center of the rectangle G_n is the origin. Then the set of vectors of the form $\mathbf{i}-\mathbf{j}$ with $\mathbf{i},\mathbf{j} \in \mathcal{G}_n$ is $2\mathcal{G}_n$, where $2\mathcal{G}_n$ is defined as $(2G_n)\cap(\mathbb{Z}/\Lambda_n)^d$. If $\mathbf{u} \in 2\mathcal{G}_n$ is fixed, then denote by $|\mathcal{G}_{n,\mathbf{u}}|$ the number of pairs $(\mathbf{i},\mathbf{j}) \in \mathcal{G}_n \times \mathcal{G}_n$ with $\mathbf{i}-\mathbf{j} = \mathbf{u}$. Then

$$B = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left\{\frac{1}{h^2}K\left(\frac{x_t-u}{h}\right)K\left(\frac{x_s-v}{h}\right)\right.$$
$$\left.\times\left(\frac{1}{\Lambda_n^d}\sum_{\mathbf{u}\in 2\mathcal{G}_n^{\mathbf{0}}}\frac{|\mathcal{G}_{n,\mathbf{u}}|}{|\mathcal{G}_n|}a_{\mathbf{u}}(u,v)g_{\mathbf{u}}(u,v)\right)\right\}dudv, \qquad (24)$$

where $2\mathcal{G}_n^{\mathbf{0}} = 2\mathcal{G}_n \setminus \{\mathbf{0}\}$. Now fix an $\varepsilon > 0$. As $||a_{\mathbf{u}}g_{\mathbf{u}}||$ is directly Riemann integrable, one can find a zone $M_\varepsilon \subset \mathbb{R}^d$ (with center in the origin) such that

$$\int_{\mathbb{R}_{\mathbf{0}}^d \setminus M_\varepsilon} ||a_{\mathbf{u}}g_{\mathbf{u}}||d\mathbf{u} \leq \varepsilon \tag{25}$$

and at the same time the Riemannian approximating sums of this integral do not exceed ε if the diagonals of the subdivision are small enough (see Fazekas and Chuprunov, 2006). Therefore, as $|\mathcal{G}_{n,\mathbf{u}}|/|\mathcal{G}_n| \leq 1$,

$$\frac{1}{\Lambda_n^d} \sum_{\mathbf{u} \in 2\mathcal{G}_n^{\mathbf{0}} \setminus M_\varepsilon} \frac{|\mathcal{G}_{n,\mathbf{u}}|}{|\mathcal{G}_n|} ||a_{\mathbf{u}}g_{\mathbf{u}}|| \leq \varepsilon, \tag{26}$$

where $1/\Lambda_n^d$ is small enough, i.e. when n is large enough: $n \geq n_\varepsilon$. Fix $\varepsilon, M_\varepsilon$ and assume that $n \geq n_\varepsilon$. Because $a_{\mathbf{u}}g_{\mathbf{u}}$ is Riemann integrable as a function $a \cdot g : \mathbb{R}_{\mathbf{0}}^d \to \mathcal{C}(\mathbb{R}^2)$ on R for each bounded closed d-dimensional rectangle R in $\mathbb{R}_{\mathbf{0}}^d$, we have

$$\left\| \frac{1}{\Lambda_n^d} \sum_{\mathbf{u} \in 2\mathcal{G}_n^{\mathbf{0}} \cap M_\varepsilon} a_{\mathbf{u}}g_{\mathbf{u}} - \int_{M_\varepsilon} a_{\mathbf{u}}g_{\mathbf{u}}d\mathbf{u} \right\| \leq \varepsilon \tag{27}$$

in the space $\mathcal{C}(\mathbb{R}^2)$, if n is large enough. This relation and (25) imply that

$$\int_{\mathbb{R}_{\mathbf{0}}^d} a_{\mathbf{u}}(x,y)g_{\mathbf{u}}(x,y)d\mathbf{u}$$

exists and is continuous in (x, y). As each edge of G_n converges to ∞, $\frac{|\mathcal{G}_{n,\mathbf{u}}|}{|\mathcal{G}_n|} \to 1$ uniformly according to $\mathbf{u} \in M_\varepsilon$. Therefore, using that $\|a_{\mathbf{u}}g_{\mathbf{u}}\|$ is directly Riemann integrable, we obtain that

$$\left\| \frac{1}{\Lambda_n^d} \sum_{\mathbf{u} \in 2\mathcal{G}_n^{\mathbf{0}} \cap M_\varepsilon} \frac{|\mathcal{G}_{n,\mathbf{u}}|}{|\mathcal{G}_n|} a_{\mathbf{u}}g_{\mathbf{u}} - \frac{1}{\Lambda_n^d} \sum_{\mathbf{u} \in 2\mathcal{G}_n^{\mathbf{0}} \cap M_\varepsilon} a_{\mathbf{u}}g_{\mathbf{u}} \right\| \leq \varepsilon \tag{28}$$

if n is large enough.

Relations (25)-(28) imply that

$$\left\| \frac{1}{\Lambda_n^d} \sum_{\mathbf{u} \in 2\mathcal{G}_n^{\mathbf{0}}} \frac{|\mathcal{G}_{n,\mathbf{u}}|}{|\mathcal{G}_n|} a_{\mathbf{u}} g_{\mathbf{u}} - \int_{\mathbb{R}_{\mathbf{0}}^d} a_{\mathbf{u}} g_{\mathbf{u}} d\mathbf{u} \right\| \leq 4\varepsilon \tag{29}$$

if n is large enough.

Therefore, using that $\frac{1}{h} K\left(\frac{x_t - u}{h}\right)$ is a density function, we have

$$\left| B - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ \frac{1}{h^2} K\left(\frac{x_t - u}{h}\right) K\left(\frac{x_s - v}{h}\right) \int_{\mathbb{R}_{\mathbf{0}}^d} a_{\mathbf{u}}(u,v) g_{\mathbf{u}}(u,v) d\mathbf{u} \right\} du dv \right| \leq 4\varepsilon \tag{30}$$

if n is large enough. As $\int_{\mathbb{R}_{\mathbf{0}}^d} a_{\mathbf{u}}(u,v) g_{\mathbf{u}}(u,v) d\mathbf{u}$ is continuous according to (u,v), the limit of the double integral in expression (30) is $\int_{\mathbb{R}_{\mathbf{0}}^d} a_{\mathbf{u}}(x_t, x_s) g_{\mathbf{u}}(x_t, x_s) d\mathbf{u} = \sigma(x_t, x_s)$ (see Theorem 2.1.1 in Prakasa Rao, 1983). Therefore

$$B \to \int_{\mathbb{R}_{\mathbf{0}}^d} a_{\mathbf{u}}(x_t, x_s) g_{\mathbf{u}}(x_t, x_s) d\mathbf{u} = \sigma(x_t, x_s).$$

Therefore we obtain the asymptotic covariance of J_1 in the following form

$$L \int_{-\infty}^{\infty} K^2(t) dt \cdot diag\left(v(x_t) g(x_t)\right) + \left(\sigma(x_t, x_s)\right)_{t,s=1}^m .$$

Now turn to J_2.

$$J_2(x) = \frac{1}{\sqrt{|\mathcal{D}_n| \Lambda_n^d}} \sum_{\mathbf{t} \in \mathcal{D}_n} \frac{1}{h} \left[r(X_{\mathbf{t}}) - r(x) \right] K\left(\frac{x - X_{\mathbf{t}}}{h}\right).$$

Then, using Taylor's expansion ($r(u) = r(x) + r'(x)(u-x) + \frac{1}{2} r''(\widetilde{x})(u-x)^2$), we get

$$E\left(\frac{1}{h}\left[r(X_{\mathbf{t}}) - r(x)\right] K\left(\frac{x - X_{\mathbf{t}}}{h}\right)\right) = \int_{-\infty}^{\infty} \frac{1}{h}\left[r(u) - r(x)\right] K\left(\frac{x-u}{h}\right) g(u) du$$

$$= \int_{-\infty}^{\infty} \frac{1}{h} \left[r'(x)(u-x) + \frac{1}{2} r''(\widetilde{x})(u-x)^2 \right] K\left(\frac{x-u}{h}\right) g(u) du$$

$$= \int_{-\infty}^{\infty} \frac{1}{h} \left[r'(x)z - \frac{1}{2} r''(\widetilde{x}) z^2 \right] K\left(\frac{z}{h}\right) g(x-z)dz$$

$$= r'(x)h \int_{-\infty}^{\infty} \frac{1}{h}\frac{z}{h} K\left(\frac{z}{h}\right) g(x-z)dz - \frac{1}{2}\int_{-\infty}^{\infty} \frac{1}{h} r''(\widetilde{x}) z^2 K\left(\frac{z}{h}\right) g(x-z)dz = A_{11} + A_{12}.$$

We show that

$$|A_{11}|, |A_{12}| \leq h^2 C. \tag{31}$$

To obtain this relation, first consider A_{11}. Using substitution $t = \frac{z}{h}$, Taylor's expansion $g(x-th) = g(x) + g'(\widetilde{x})(-th)$, the boundedness of g'', and the symmetry of K, we obtain:

$$\int_{-\infty}^{\infty} \frac{1}{h}\frac{z}{h} K\left(\frac{z}{h}\right) g(x-z)dz = \int_{-\infty}^{\infty} tK(t)g(x-th)dt =$$

$$\int_{-\infty}^{\infty} tK(t)g(x)dt + \int_{-\infty}^{\infty} tK(t)g'(\widetilde{x})(-th)dt = -hc\int_{-\infty}^{\infty} t^2 K(t)dt.$$

Therefore $|A_{11}| \leq h^2 C$.

For $|A_{12}|$ we have $|A_{12}| \leq Ch^2 \int_{-\infty}^{\infty} \frac{1}{h} \left(\frac{z}{h}\right)^2 K\left(\frac{z}{h}\right) g(x-z)dz$. By (11) and Theorem D we have $\int_{-\infty}^{\infty} \frac{1}{h}\left(\frac{z}{h}\right)^2 K\left(\frac{z}{h}\right) g(x-z)dz \to g(x)\int_{-\infty}^{\infty} z^2 K(z)dz$. Therefore $|A_{12}| \leq h^2 C$.

Now, by (31), $|E(J_2(x))| \lesssim \sqrt{\frac{|\mathcal{D}_n|}{\Lambda_n^d}} h^2 C = \sqrt{|T_n|} h^2 C$. Therefore $E(J_2(x)) \to 0$, because by (5), $|T_n| h_n^4 \to 0$.

We want to show that $E|J_2|^l \to 0$ where $1 < l < 2$.

As $E|J_2|^l \leq C\left(E|J_2 - E(J_2)|^l + |E(J_2)|^l\right)$ and $E(J_2) \to 0$, it is sufficient to show that $E|J_2 - E(J_2)|^l \to 0$.

We have

$$E|J_2 - E(J_2)|^l = \left(\frac{1}{\sqrt{|\mathcal{D}_n|\Lambda_n^d}}\right)^l \left(\frac{1}{h}\right)^l E\left|\sum_{\mathbf{t}\in\mathcal{D}_n} \eta_{\mathbf{t}} - E\eta_{\mathbf{t}}\right|^l,$$

where $\eta_{\mathbf{t}} = (r(X_{\mathbf{t}}) - r(x))\, K\left(\frac{x - X_{\mathbf{t}}}{h}\right)$.

Now joining $\eta_{\mathbf{t}}$ into the unit cubes (denote a cube with $\mathcal{K}$) and applying the Rosenthal inequality (19) for these, we obtain

$$E|J_2 - E(J_2)|^l \leq C\left(\frac{1}{|\mathcal{D}_n|\Lambda_n^d}\right)^{\frac{l}{2}}\left(\frac{1}{h}\right)^l \sum_{\mathcal{K}\in\mathcal{D}_n'}\left(E\left|\sum_{\mathbf{t}\in\mathcal{K}}\eta_{\mathbf{t}} - E\eta_{\mathbf{t}}\right|^{l+\varepsilon}\right)^{\frac{l}{l+\varepsilon}}$$

$$\leq C\left(\frac{1}{|\mathcal{D}_n|\Lambda_n^d}\right)^{\frac{l}{2}}\left(\frac{1}{h}\right)^l \sum_{\mathcal{K}\in\mathcal{D}_n'}\left(E\left|\sum_{\mathbf{t}\in\mathcal{K}}\eta_{\mathbf{t}}\right|^{l+\varepsilon}\right)^{\frac{l}{l+\varepsilon}}, \tag{32}$$

as $E|\eta - E\eta|^k \leq CE|\eta|^k$ for $k > 1$. (We see that $c_{1,1}^{(\varepsilon)} < \infty$ follows from (1).) Applying the Jensen inequality, we get

$$\left|\sum_{\mathbf{t}\in\mathcal{K}}\eta_{\mathbf{t}}\right|^{l+\varepsilon} = \Lambda_n^{d(l+\varepsilon)}\left|\sum_{\mathbf{t}\in\mathcal{K}}\frac{1}{\Lambda_n^d}\eta_{\mathbf{t}}\right|^{l+\varepsilon} \leq \Lambda_n^{d(l+\varepsilon)}\sum_{\mathbf{t}\in\mathcal{K}}\frac{1}{\Lambda_n^d}\left|\eta_{\mathbf{t}}\right|^{l+\varepsilon},$$

which implies that

$$E\left|\sum_{\mathbf{t}\in\mathcal{K}}\eta_{\mathbf{t}}\right|^{l+\varepsilon} \leq \Lambda_n^{d(l+\varepsilon)}\sum_{\mathbf{t}\in\mathcal{K}}\frac{1}{\Lambda_n^d}E\left|\eta_{\mathbf{t}}\right|^{l+\varepsilon} = \Lambda_n^{d(l+\varepsilon)}E\left|\eta_{\mathbf{t}}\right|^{l+\varepsilon}. \tag{33}$$

Hence, by (32) and (33), we get

$$E|J_2 - E(J_2)|^l \leq C\left(\frac{1}{|\mathcal{D}_n|\Lambda_n^d}\right)^{\frac{l}{2}}\left(\frac{1}{h}\right)^l \frac{|\mathcal{D}_n|}{\Lambda_n^d}\Lambda_n^{d\cdot l}\left(E\left|\eta_{\mathbf{t}}\right|^{l+\varepsilon}\right)^{\frac{l}{l+\varepsilon}}. \tag{34}$$

We can calculate the limit of $E\left|\eta_{\mathbf{t}}\right|^{l+\varepsilon}$ in the following way:

$$E\left|\eta_{\mathbf{t}}\right|^{l+\varepsilon} = E\left|r(X_{\mathbf{t}}) - r(x)\right|^{l+\varepsilon} K^{l+\varepsilon}\left(\frac{x - X_{\mathbf{t}}}{h}\right)$$

$$= \int_{-\infty}^{\infty}\Big|\underbrace{r(u) - r(x)}_{r'(\widetilde{x})(x-u)}\Big|^{l+\varepsilon} K^{l+\varepsilon}\left(\frac{x-u}{h}\right) g(u)du$$

$$\leq ch^{1+l+\varepsilon}\int_{-\infty}^{\infty}\frac{1}{h}\left|\frac{x-u}{h}\right|^{l+\varepsilon} K^{l+\varepsilon}\left(\frac{x-u}{h}\right) g(u)du$$

$$\to h^{1+l+\varepsilon} c g(x) \int_{-\infty}^{\infty} |z|^{l+\varepsilon} K^{l+\varepsilon}(z) dz.$$

(Here we applied Theorem D.)

Therefore, by (34), we have

$$E|J_2 - E(J_2)|^l \leq C \left(\frac{1}{|\mathcal{D}_n| \Lambda_n^d} \right)^{\frac{l}{2}} \left(\frac{1}{h} \right)^l \frac{|\mathcal{D}_n|}{\Lambda_n^d} \Lambda_n^{d \cdot l} h^{\frac{l(1+l+\varepsilon)}{l+\varepsilon}}$$

$$= C \left|\mathcal{D}_n\right|^{1-\frac{l}{2}} \left(\Lambda_n^d\right)^{\frac{l}{2}-1} h^{\frac{l}{l+\varepsilon}} = C \left|T_n\right|^{1-\frac{l}{2}} h^{\frac{l}{l+\varepsilon}}.$$

Choosing appropriate l and ε (e.g. $l = 1.98$, $\varepsilon = 0.01$) relation (5) implies that $\left|T_n\right|^{1-\frac{l}{2}} h^{\frac{l}{l+\varepsilon}} \to 0$.

Therefore $E|J_2|^l \to 0$, so $J_2 \to 0$ in probability.

Finally, we deal with J_3:

$$J_3(x) = \frac{1}{|\mathcal{D}_n|} \sum_{\mathbf{t} \in \mathcal{D}_n} \frac{1}{h} K \left(\frac{x - X_{\mathbf{t}}}{h} \right).$$

By Theorem A, $\sqrt{|T_n|}(J_3(x) - g(x))$ is convergent in distribution, therefore $J_3(x) \to g(x)$ in probability.

Remark **6.** We see that (5) and (7) can be satisfied simultaneously only if $1 < a < \frac{5+\sqrt{17}}{4}$.

4. Examples

In this section we present simple examples that give numerical evidence for the phenomena described in Theorem 1.

Let $X_{\mathbf{u}}, \mathbf{u} \in \mathbb{R}^d$, be a stationary Gaussian random field with mean value function zero and covariance function $\rho_{\mathbf{u}}$. In the following examples we consider the same random fields $X_{\mathbf{u}}$ which were studied in Fazekas and Chuprunov

(2006) and Fazekas (2007). We will choose $\Phi(Y_{\mathbf{u}}) = 10\sin(X_{\mathbf{u}}) + 100 + \delta_{\mathbf{u}}$, where $\delta_{\mathbf{u}} = \widetilde{X_{\mathbf{u}}}$, $\widetilde{X_{\mathbf{u}}}$ is a stationary random field having the same distribution as $X_{\mathbf{u}}$ and being independent of $X_{\mathbf{u}}$.

Example 1: Consider the Gaussian process $X(u), u \in \mathbb{R}$, with mean zero and covariance function $\rho_u = e^{-|u|}, u \in \mathbb{R}$. We consider this process in the $1/\Lambda$-lattice points of the domain $T = [0, t]$ with $\Lambda = 40$ and $t = 60$. That is, the sample is $z_1 = X(1/40), \ldots, z_s = X(2400/40)$ with $s = 2400$. Now the covariance matrix of this data vector is $(\rho^{|i-j|})_{i,j=1}^{s}$, where $\rho = e^{-1/\Lambda}$. Therefore the data generation for the simulation is easy. Let $y_1, \ldots, y_s$ be i.i.d. standard normal and choose $z_i = \rho^{i-1} y_1 + \sqrt{1-\rho^2} \sum_{j=2}^{i} \rho^{i-j} y_j, i = 1, \ldots, s$.

Using these data, we calculated the regression estimator r_n at the points $x_1 = -0.5$, $x_2 = -0.25$, $x_3 = 0$, $x_4 = 0.25$, and $x_5 = 0.5$. We used two values of the bandwidth, $h_1 = 0.025$ and $h_2 = 0.005$, and applied the standard normal density function as kernel K.

The simulations were performed with MATLAB, 5000 repetitions of the procedure were made. The data sets for both bandwidths h_1 and h_2 were the same. The theoretical values of the regression function and the average of their estimators are shown in the Table 1. For both values of the bandwidths we can see a close agreement of the theoretical and empirical values.

We calculated the empirical covariance matrices Σ_1 (corresponding to bandwidth h_1) and Σ_2 (corresponding to bandwidth h_2) for our standardized estimators $\sqrt{\frac{|\mathcal{D}|}{\Lambda}}(r_n(x_1)-r(x_1), \ldots, r_n(x_5)-r(x_5))$ (the standardization factor is $\sqrt{\frac{|\mathcal{D}|}{\Lambda}} = 7.7459$):

Table 1: Theoretical values of the regression function and the average of their estimators for the data of Example 1.

x	-0.5	-0.25	0	0.25	0.5
$r(x)$	95.2057	97.5260	100.0000	102.4740	104.7943
$r_n(x)$ with $h_1 = 0.025$	95.2039	97.5220	99.9953	102.4684	104.7929
$r_n(x)$ with $h_2 = 0.005$	95.1970	97.5229	99.9939	102.4707	104.7976

$$\Sigma_1 = \begin{bmatrix} 3.8773 & 2.6953 & 2.1923 & 1.7857 & 1.5073 \\ 2.6953 & 3.5796 & 2.5499 & 2.1007 & 1.7623 \\ 2.1923 & 2.5499 & 3.4399 & 2.4892 & 2.0995 \\ 1.7857 & 2.1007 & 2.4892 & 3.4852 & 2.6500 \\ 1.5073 & 1.7623 & 2.0995 & 2.6500 & 3.8147 \end{bmatrix};$$

$$\Sigma_2 = \begin{bmatrix} 7.2195 & 2.8212 & 2.1822 & 1.8547 & 1.5020 \\ 2.8212 & 6.5902 & 2.5058 & 2.0756 & 1.7223 \\ 2.1822 & 2.5058 & 6.2153 & 2.4162 & 2.1099 \\ 1.8547 & 2.0756 & 2.4162 & 6.5192 & 2.6732 \\ 1.5020 & 1.7223 & 2.1099 & 2.6732 & 7.1689 \end{bmatrix}.$$

The difference in the diagonals of Σ_1 and Σ_2 is clearly visible. The off-diagonal elements are almost the same.

Now calculate the additional terms in the diagonals of the covariance matrices described in Theorem 1. In our case the elements of the diagonal matrix D_k for bandwidth h_k (for $k = 1, 2$) are

$$\frac{1}{\Lambda}\frac{1}{h_k}v(x_i)\frac{1}{g(x_i)}\int_{-\infty}^{\infty} K^2(u)du = \frac{1}{40}\frac{1}{h_k} \cdot 1 \cdot \frac{1}{g(x_i)}\frac{1}{2\sqrt{\pi}}.$$

Because in the infill-increasing case only the diagonals of the limit covariance matrices can be different for different values of the bandwidth, we show the ratio between the diagonals of the difference of the empirical covariance matrices, $diag(\Sigma_2 - \Sigma_1)$, and of the theoretical covariance matrices, $diag(D_2 - D_1)$, in Table 2.

Table 2: Ratio between the diagonals of the difference of the empirical covariance matrices and of the theoretical covariance matrices for the data of Example 1.

x	-0.5	-0.25	0	0.25	0.5
$\frac{diag(\Sigma_2-\Sigma_1)}{diag(D_2-D_1)}$	1.0428	1.0316	0.9812	1.0397	1.0465

As the ratios are close to 1, the results show that the diagonal matrix D of Theorem 1 explains well the dependence of the limit covariance matrix on the bandwidth.

Finally, Figure 1 shows histograms with the relative frequencies of the estimators of $r(x_3 = 0)$ for the bandwidths $h_1 = 0.025$ (left picture) and $h_2 = 0.005$ (right picture). The histograms are shown together with the theoretical normal densities with mean and variance estimated from the data used for the histograms. The approximate normal distribution of the regression estimator stated in Theorem 1 is reflected in these figures. Different bandwidths lead to a different spread of the normal distribution.

Example 2: In this example we consider the Gaussian process $X(u, v), (u, v) \in \mathbb{R}^2$, with mean zero and covariance function $\rho_{(u,v)} = e^{-(|u|+|v|)}, (u, v) \in \mathbb{R}^2$. As in the previous example, let $\Phi(Y_{\mathbf{u}}) = 10 \sin(X_{\mathbf{u}}) + 100 + \widetilde{X_{\mathbf{u}}}$. This process is observed in the $1/\Lambda$-lattice points of the domain $T = [0, t]^2$ with $\Lambda = 10$ and $t = 30$, and thus the sample is $z_{(i,j)} = X_{(i/10,j/10)}, i, j = 1, \ldots, 300$, with sample size $(30 \cdot 10)^2 = 90000$. Therefore, we generate data $y_{k,l}$, for

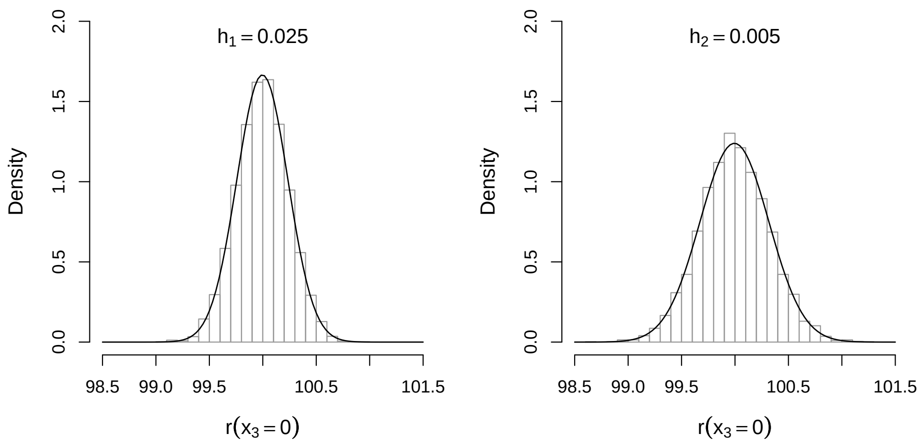

Figure 1: Histograms with the relative frequencies of the estimators of $r(x_3 = 0)$ for the bandwidths $h_1 = 0.025$ (left) and $h_2 = 0.005$ (right), together with the theoretical densities of the normal distribution.

$k, l = 1, \ldots, 300$, to be i.i.d. standard normal, and choose

$$z_{(i,j)} = \rho^{i+j-2} y_{1,1} + \sqrt{1-\rho^2}\rho^{j-1} \sum_{k=2}^{i} \rho^{i-k} y_{k,1}$$

$$+\sqrt{1-\rho^2}\rho^{i-1} \sum_{l=2}^{j} \rho^{j-l} y_{1,l} + (1-\rho^2) \sum_{k=2}^{i} \sum_{l=2}^{j} \rho^{i-k} \rho^{j-l} y_{k,l},$$

$i, j = 1, \ldots, 300$, where $\rho = e^{-1/\Lambda}$.

As in the previous example, we calculated the regression estimator r_n at the points $x_1 = -0.5, x_2 = -0.25, x_3 = 0, x_4 = 0.25, x_5 = 0.5$. We used the bandwidth $h_1 = 0.01$ and $h_2 = 0.002$ and applied the standard normal density function as kernel K. The data sets for both bandwidths were the same, and 5000 repetitions were performed. Table 3 shows that the theoretical values of the regression function and the average of their estimators are very similar.

The standardized estimators (the standardization factor is $\sqrt{\frac{|\mathcal{D}|}{\Lambda^2}} = 30$)

Table 3: Theoretical values of the regression function and the average of their estimators for the data of Example 2.

x	-0.5	-0.25	0	0.25	0.5
$r(x)$	95.2057	97.5260	100.0000	102.4740	104.7943
$r_n(x)$ with $h = 0.0100$	95.2069	97.5270	99.9999	102.4735	104.7937
$r_n(x)$ with $h = 0.0020$	95.2074	97.5276	100.0001	102.4724	104.7955

have the empirical covariance matrices

$$\Sigma_1 = \begin{bmatrix} 5.1226 & 4.1911 & 3.9213 & 3.7524 & 3.5575 \\ 4.1911 & 4.9262 & 4.0812 & 3.9838 & 3.8019 \\ 3.9213 & 4.0812 & 4.7751 & 4.0712 & 3.9663 \\ 3.7524 & 3.9838 & 4.0712 & 4.9523 & 4.2129 \\ 3.5575 & 3.8019 & 3.9663 & 4.2129 & 5.1573 \end{bmatrix}$$

$$\Sigma_2 = \begin{bmatrix} 8.2768 & 4.2437 & 3.9402 & 3.7704 & 3.6560 \\ 4.2437 & 7.8458 & 4.1450 & 4.1074 & 3.8184 \\ 3.9402 & 4.1450 & 7.5220 & 4.0948 & 4.0625 \\ 3.7704 & 4.1074 & 4.0948 & 7.9032 & 4.3544 \\ 3.6560 & 3.8184 & 4.0625 & 4.3544 & 8.4931 \end{bmatrix}$$

for the bandwidths h_1 and h_2, respectively. Again, the agreement of the off-diagonal elements and the difference in the diagonal becomes visible.

Similar to the previous example, we show the ratios $\frac{diag(\Sigma_2-\Sigma_1)}{diag(D_2-D_1)}$ in Table 4. These are close to 1 as expected from Theorem 1.

Since, according to Theorem 1, the regression estimator should approach multivariate normality for different values x_i, we present in Figure 2 the resulting estimations of $r(x_1 = -0.5)$ (horizontal axes) and $r(x_2 = -0.25)$

Table 4: Ratio between the diagonals of the difference of the empirical covariance matrices and of the theoretical covariance matrices for the data of Example 2.

x	-0.5	-0.25	0	0.25	0.5
$\frac{diag(\Sigma_2-\Sigma_1)}{diag(D_2-D_1)}$	0.9841	1.0004	0.9711	1.0112	1.0408

(vertical axes) for the bandwidths $h_1 = 0.01$ (left picture) and $h_2 = 0.002$ (right picture). The estimated contour lines for certain levels are drawn with dashed lines. The solid ellipses represent the same levels, but taken from the density functions of the bivariate normal density with mean and covariance taken from the underlying data. The close agreement is clearly visible. Moreover, for both bandwidths the ellipses have the same orientation but different size, which refers to the closeness of the off-diagonal elements and to the disagreement of the diagonal elements of Σ_1 and Σ_2 from above.

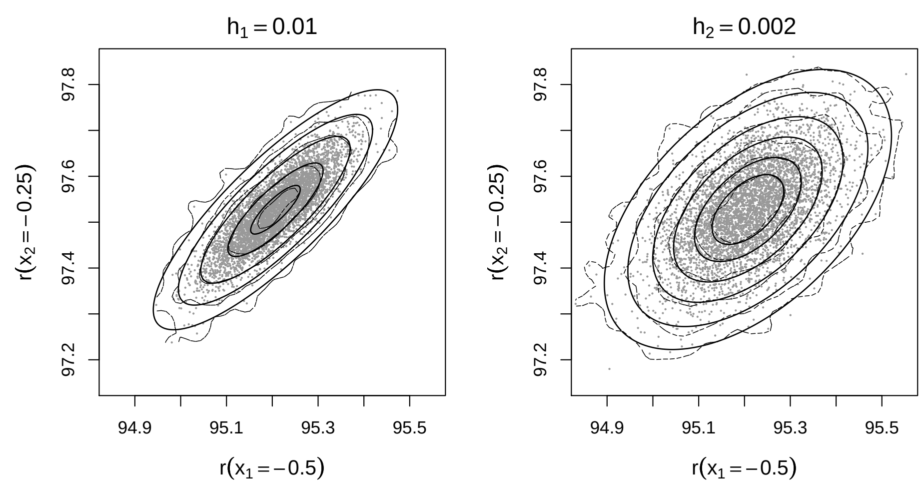

Figure 2: Two-dimensional representations of the estimators of $r(x_1 = -0.5)$ and $r(x_2 = -0.25)$ for the bandwidths $h_1 = 0.01$ (left) and $h_2 = 0.002$ (right), together with contour lines (dashed) and ellipses for theoretical values of the normal densities (solid).

Acknowledgement

The authors are grateful to the referee for helpful comments and suggestions.

References

Biau, G. (2004), Spatial kernel density estimation. *Mathematical Methods of Statistics* **12** (2003), no.4, 371–390 (2004).

Blanke, D., Pumo, B. (2002), Optimal sampling for density estimation in continuous time. *J. Time Ser. Anal.* **24** (2003), no.1, 1–23.

Bosq, D. (1997), Parametric rates of nonparametric estimators and predictors for continuous time processes. *The Annals of Statistics,* **25** (3), 982–1000.

Bosq, D. (1998), *Nonparametric Statistics for Stochastic Processes.* Springer, New York - Berlin - Heidelberg.

Bosq, D., Cheze, N. (1993), Erreur quadratique asymptotique optimale de l'estimateur non paramétrique de la régression pour des observations discrétisées d'un processus stationnaire á temps continu. *C. R. Acad. Sci. Paris* **317** (I), no. 9, 891–894.

Cai, Z. (2001), Weighted Nadaraya-Watson regression estimation. *Statistics & Probability Letters,* **51**, 307–318.

Cheze, N. (1992), Régression non paramétrique pour un processus à temps continu. *C. R. Acad. Sci. Paris* **315** (I), 1009–1012.

Cressie, N.A.C. (1991), *Statistics for Spatial Data.* Wiley, New York.

Devroye, L. and Györfi, L. (1985), *Nonparametric density estimation. The L_1 view.* Wiley, New York.

Doukhan, P. (1994), *Mixing. Properties and Examples.* Lecture Notes in Statistics **85**, Springer, New York.

Fazekas, I. (2003), Limit theorems for the empirical distribution function in the spatial case. *Statistics & Probability Letters,* **62**, 251–262.

Fazekas, I. (2007), Central limit theorems for kernel type density estimators. *Proceedings of the 7th International Conference on Applied Informatics,* Eger, (Vol. 1), 209–219.

Fazekas, I. and Chuprunov, A. (2004), A central limit theorem for random fields. *Acta Mathematica Academiae Paedagogicae Nyiregyhaziensis,* **20** (1), 93–104, www.emis.de/journals/AMAPN.

Fazekas, I. and Chuprunov, A. (2006), Asymptotic normality of kernel type density estimators for random fields. *Stat. Inf. Stoch. Proc.* **9**, 161–178.

Fazekas, I., Kukush, A.G. and Tómács, T. (2000), On the Rosenthal inequality for mixing fields. *Ukrainian Math. J.,* **52***(2), 266–276.*

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

Hille, E. and Phillips, R.S. (1957), *Functional Analysis and Semi-groups.* AMS, Providence.

Lahiri, S.N. (1996), On inconsistency of estimators based on spatial data under infill asymptotics. *Sankhya*, **58** Ser. A, 403–417.

Lahiri, S.N., Kaiser, M.S., Cressie, N. and Hsu, N. J. (1999), Prediction of spatial cumulative distribution functions using subsampling. *J. Amer. Statist. Assoc.* **94** (445), 86–110.

Masry, E. (1983), Probability density estimation from sampled data. *IEEE Transactions in information theory*, vol. IT-29, no. 5, 696–709.

Nadaraya, E.A. (1964), On estimating regression. *Theor. Probability Appl.,* **9**, 141–142.

Park, B., Kim, T.Y.K., Park, J-S., Hwang, S.Y. (2008), Practically applicable central limit theorem for spatial statistics. *Math. Geosci, Online: http://springerlink.metapress.com/content/121014/?Content+Status=Accepted*

Prakasa Rao, B.L.S. (1983), *Nonparametric Functional Estimation.* Academic Press, INC. London.

Schuster, E.F. (1972), Joint asymptotic distribution of the estimated regression function at a finite number of distinct points. *The Annals of Mathematical Statistics,* **43** (1), 84–88.

Watson, G.S. (1964), Smooth regression analysis. *Sankhya. Ser. A* **26**, 359–372.