

# A Central Limit Theorem for Spatial Observations

István Fazekas<sup>1</sup>, Zsolt Karácsony<sup>2</sup> and Renáta Vas<sup>1</sup>

<sup>1</sup> Faculty of Informatics, University of Debrecen, Hungary

<sup>2</sup> Department of Applied Mathematics, University of Miskolc, Hungary

**Abstract:** The Central Limit Theorem is proved for  $m$ -dependent random fields. The random field is observed in a sequence of irregular domains. The sequence of domains is increasing and at the same time the locations of the observations become more and more dense in the domains.

**Zusammenfassung:** Der zentrale Grenzwertsatz wird für  $m$ -abhängige Zufallsfelder bewiesen. Das Zufallsfeld wird an einer Sequenz von irregulären Punkten beobachtet. Die Sequenz der Gebiete wächst und gleichzeitig werden die Lokationen der Beobachtungen immer weniger dicht.

**Keywords:** Asymptotic Normality, Central Limit Theorem, Random Field, Nearly Infill.

## 1 Introduction

In statistics most asymptotic results concern the increasing domain case. That is, a stochastic process (or a random field) is observed in an increasing sequence of domains  $D_n$  where the size of  $D_n$  goes to infinity as  $n \rightarrow \infty$ . However, in spatial statistics one can observe a random field in a fixed domain such a way that the locations of observations become dense in that domain (infill asymptotics, see Cressie, 1991). The increasing domain and the infill approaches can be combined as well. We call it nearly infill sampling (or infill-increasing setup) when the sequence of domains is increasing and at the same time the locations of observations become dense. Using this setup, Lahiri (1999) studied the asymptotic behavior of the empirical distribution function. Furthermore, Lahiri, Kaiser, Cressie, and Hsu (1999) gave practical application of this approach. Using the infill-increasing approach, Fazekas (2003) obtained a functional limit theorem for the spatial empirical distribution function.

When studying a continuous parameter process, we observe it at discrete points of the domain. This method is called sampling (see Bosq, 1998) and it is closely related to the infill-increasing approach. Moreover, any numerical study (e.g. simulation) of a continuous parameter stochastic process is based on discrete approximations (e.g. an integral is approximated with a sum). Therefore a considerable part of the numerical results on continuous parameter processes can be considered as results in the infill-increasing scheme.

The importance of the infill-increasing setup is underlined by the results of Fazekas and Chuprunov (2006). Their main theorem shows that the limiting distribution of the kernel type density estimator in the infill-increasing model can be a combination of the ones in the discrete parameter and the continuous parameter models. A similar result was obtained for kernel type regression estimator in Karácsony and Filzmoser (2010).

Park, Kim, Park, and Hwang (2009) presented central limit theorems (CLT) for irregular domains assuming nearly infill sampling. They applied the CLTs to show asymptotic normality of the kernel type density estimator. However, they did not consider multivariate normality.

The aim of our paper is to prove a central limit theorem for nearly infill sampling. To obtain results being interesting for applied statisticians the domain of observations is allowed to be irregular.

Consider a domain  $D$  in  $\mathbb{R}^d$ . We observe a random field in certain points of  $D$ . More precisely, the random field  $\xi(\cdot)$  is observed at finitely many locations i.e. at the elements  $\mathbf{s}_{n1}, \dots, \mathbf{s}_{nm}$  lying in the sampling region  $D_n \subset D$ . Since we consider the mixed (or nearly infill or infill-increasing) domain sampling, the sampling region  $D_n$  increases (that is  $D_n \subseteq D_{n+1}$ ) and at the same time, the data sites  $\mathbf{s}_{n1}, \dots, \mathbf{s}_{nm}$  fill in any given sub-region of  $D_n$  increasingly densely as  $n \rightarrow \infty$ . The set of the locations of the observations is denoted by  $\mathcal{R}_n$ :  $\mathcal{R}_n = \{\mathbf{s}_{n1}, \dots, \mathbf{s}_{nm}\}$ . We mention that it is not a real restriction of the generality that the size of the  $n$ -th sample is  $n$ .

To obtain asymptotic normality, we assume that the  $n$ th set of observations is  $\xi_n(\mathbf{s}_{n1}), \dots, \xi_n(\mathbf{s}_{nm})$  where  $\xi_n(\cdot)$ ,  $n = 1, 2, \dots$ , is a sequence of stationary random fields and  $\xi_n(\cdot)$  is weakly dependent for any fixed  $n$ . For the sake of simplicity, we suppose that  $\xi_n(\cdot)$  is  $m$ -dependent. It is a restriction but it has an advantage, namely that we can easily obtain a central limit theorem for irregular domains. We mention that similar results can be obtained for mixing random fields as well (see e.g. Fazekas and Chuprunov, 2004, but there the domain is regular and the conditions are quite difficult to check).

Section 2 contains the notations. The basic result and its proof is presented in Section 3. Our Theorem 1 is a CLT for irregular domains. It is a version of Theorem 2 of Park et al. (2009). To prove Theorem 1 we apply the ideas of Park et al. (2009) that is using the classical theorem of Fréchet and Shohat (1931). We present a detailed proof, most of the technical parts are summarized in Lemma 2. In Section 4 we extend Theorem 1 for multivariate observations. Section 5 contains numerical (simulation) results. A moving average process is observed in a highly irregular two-dimensional domain and the asymptotic normality of the empirical distribution function is shown there.

## 2 Notations

Let us consider a zero mean strictly stationary random field  $\{\xi(\mathbf{s}) : \mathbf{s} \in D\}$ ,  $D \subseteq \mathbb{R}^d$ . Here, the strict stationarity of the random field means that for any  $\mathbf{s}_1, \dots, \mathbf{s}_k, \mathbf{t}$ , the distribution of  $(\xi(\mathbf{s}_1), \dots, \xi(\mathbf{s}_k))$  is the same as that of  $(\xi(\mathbf{s}_1 + \mathbf{t}), \dots, \xi(\mathbf{s}_k + \mathbf{t}))$ .

We assume that the random field  $\xi(\cdot)$  is  $m$ -dependent.  $m$ -dependence means that  $m$  is the infimum of the numbers denoted by  $b$  such that if  $\|\mathbf{s}_1 - \mathbf{s}_2\| > b$ , then  $\xi(\mathbf{s}_1)$  and  $\xi(\mathbf{s}_2)$  are independent. Here,  $\|\cdot\|$  denotes the Euclidean norm in  $\mathbb{R}^d$ . For  $\mathbf{u} \in D_n$ , let

$$I_{m,n}(\mathbf{u}) = \{\mathbf{s} \in \mathcal{R}_n : \|\mathbf{s} - \mathbf{u}\| \leq m\}$$

and  $\kappa_n = \max_{\mathbf{u} \in D_n} \# \{I_{m,n}(\mathbf{u})\}$ . Thus,  $\kappa_n$  denotes the number of elements of the set  $I_{m,n}(\mathbf{u})$  with maximal cardinality. Therefore,  $\kappa_n$  is an indicator of the strength of dependence. To avoid the independent case, we assume that  $\kappa_n > 0$  for each  $n$ . We suppose

that the measure  $\kappa_n$  of the density of the locations satisfies

$$\kappa_n \sim n^a \quad \text{with a constant } 0 < a < 1. \quad (1)$$

Here for any two series  $\{t_n\}$  and  $\{v_n\}$  of positive numbers, the notation  $t_n \sim v_n$  means that the relation

$$0 < c_1 \leq \liminf_{n \rightarrow \infty} (t_n/v_n) \leq \limsup_{n \rightarrow \infty} (t_n/v_n) \leq c_2 < \infty$$

holds for positive constants  $c_1$  and  $c_2$ .

For real valued sequences  $\{a_n\}$  and  $\{b_n\}$ , the notation  $a_n = o(b_n)$  (resp.  $a_n = O(b_n)$ ) means that the sequence  $a_n/b_n$  converges to 0 (resp. is bounded).

The sign  $\mathbb{E}$  stands for expectation. Variance and covariance are denoted by  $\text{var}(\cdot)$  and  $\text{cov}(\cdot, \cdot)$ , respectively. The sign " $\Rightarrow$ " denotes convergence in distribution.  $\mathcal{N}(m, \Sigma)$  stands for the (vector) normal distribution with mean (vector)  $m$  and covariance (matrix)  $\Sigma$ .

### 3 Central Limit Theorem for Stationary Random Fields

Now, let us consider a series of strictly stationary  $m$ -dependent random fields  $\{\xi_n(\mathbf{s}) : \mathbf{s} \in D\}$ ,  $D \subseteq \mathbb{R}^d$ ,  $n = 1, 2, \dots$ . For a fixed  $n$ , let us introduce the notations  $\xi_i = \xi_n(\mathbf{s}_{ni})$  and  $S_n = \sum_{i=1}^n \xi_n(\mathbf{s}_{ni}) = \sum_{i=1}^n \xi_i$ . Furthermore, let  $\mathcal{T}_n = \{(i, j) : 0 < \|\mathbf{s}_{ni} - \mathbf{s}_{nj}\| \leq m\}$ ,  $\nu_n = \text{var}(\xi_n(\mathbf{s}))$  and

$$\tau_n = \frac{1}{n\kappa_n} \sum_{(i,j) \in \mathcal{T}_n} \text{cov}(\xi_n(\mathbf{s}_{ni}), \xi_n(\mathbf{s}_{nj})). \quad (2)$$

At this point we notice that  $\text{var}(S_n) = n\nu_n + n\kappa_n\tau_n$  and  $\tau_n$  can be negative as well. The following theorem is a version of Theorem 2 of Park et al. (2009).

**Theorem 1.** *Assume that (1) is satisfied. Consider the following conditions.*

- (i) *Let  $\{\xi_n(\mathbf{s}) : \mathbf{s} \in D\}$ ,  $D \subseteq \mathbb{R}^d$ ,  $n = 1, 2, \dots$  be a sequence of strictly stationary  $m$ -dependent random fields.*
- (ii) *For all  $n$  and  $\mathbf{s}$ , let  $\mathbb{E}\xi_n(\mathbf{s}) = 0$  and  $\nu_n = \text{var}(\xi_n(\mathbf{s}))$ . Assume that  $\nu_n > 0$  if  $n$  is large enough.*
- (iii) *For all  $\{\mathbf{s}'_{n1}, \dots, \mathbf{s}'_{nl}\} \subseteq \{\mathbf{s}_{n1}, \dots, \mathbf{s}_{nn}\}$ , let*

$$\mathbb{E} \left| \prod_{j=1}^l \xi_n(\mathbf{s}'_{nj}) \right| = O\left(\nu_n^{\frac{l}{2}}\right) \quad \text{hold uniformly as } n \rightarrow \infty. \quad (3)$$

- (iii') *For all pairwise different elements  $\mathbf{s}'_{n1}, \dots, \mathbf{s}'_{nl}$  of the set  $\{\mathbf{s}_{n1}, \dots, \mathbf{s}_{nn}\}$ , relation (3) is satisfied uniformly,  $\kappa_n \nu_n^{\frac{1}{2}} \geq 1$  and  $|\xi_n(\mathbf{s})| \leq 1$  holds for all  $n$  and  $\mathbf{s}$ .*

(iv) For a  $\delta > 0$ , let

$$\nu_n + \kappa_n \tau_n \geq \delta \kappa_n \nu_n \dots$$

Then, either conditions (i)–(ii)–(iii)–(iv) or conditions (i)–(ii)–(iii')–(iv) imply that

$$\frac{S_n}{\sqrt{\text{var}(S_n)}} \Rightarrow \mathcal{N}(0, 1).$$

To prove Theorem 1, we need some preliminary considerations. For a fixed and positive integer  $n$  let us determine the value of the expression

$$\mathbb{E}(S_n^r) = \mathbb{E}(\xi_1 + \dots + \xi_n)^r = \sum \mathbb{E}(\xi_{i_1} \xi_{i_2} \dots \xi_{i_r}), \quad (4)$$

where the sum is to be executed for all the ordered elements  $\{i_1, \dots, i_r\} \subset \{1, \dots, n\}$ . (Among  $i_1, \dots, i_r$  there can be identical indices.)

For each factor of the sum in (4), we give an undirected graph  $\{i_1, \dots, i_r\}$  with vertices  $i_1, \dots, i_r$ . Note here that some vertices can be equal in the graph  $\{i_1, \dots, i_r\}$ . We say that in  $\{i_1, \dots, i_r\}$  the pair  $(i_p, i_q)$  is connected if  $\|\mathbf{s}_{ni_p} - \mathbf{s}_{ni_q}\| \leq m$  or there exists a path  $\{j_1, \dots, j_l\} \subset \{i_1, \dots, i_r\}$  such that

$$\|\mathbf{s}_{ni_p} - \mathbf{s}_{nj_1}\| \leq m, \quad \|\mathbf{s}_{nj_1} - \mathbf{s}_{nj_2}\| \leq m, \quad \dots, \quad \|\mathbf{s}_{nj_l} - \mathbf{s}_{ni_q}\| \leq m,$$

or, in other words, we can get from  $i_p$  to  $i_q$  through vertices which have a distance less than or equal to  $m$  from each other. It means that in the graph there is an edge if and only if the distance of two vertices is less than or equal to  $m$ . Obviously, this graph falls into components which are denoted by  $\mathcal{I}_1, \dots, \mathcal{I}_k$ . Here,  $\mathcal{I}_u \cap \mathcal{I}_v = \emptyset$  if  $u \neq v$  and  $\bigcup_{u=1}^k \mathcal{I}_u = \{i_1, \dots, i_r\}$ . A component  $\mathcal{I}_u$  of our graph is a subset  $\{i_1, \dots, i_l\}$  where all the vertices are connected to each other. Let us recall that  $\kappa_n$  denotes the maximal number of points that can be located at a distance less than or equal to  $m$  from a given point.

For a graph  $G = \{i_1, \dots, i_r\}$ , let its score function (see Park et al., 2009) be defined by

$$s(G) = \mathbb{E}(\xi_{i_1} \dots \xi_{i_r}).$$

Similarly, define the score of a component  $\mathcal{I}$  of a graph by  $s(\mathcal{I})$ . Let  $\mathcal{G}_k$  be the collection of all graphs which have exactly  $k$  components. Then  $\mathbb{E}(S_n^r)$  can be written as

$$\mathbb{E}(S_n^r) = \sum_{k=1}^r \sum_{G \in \mathcal{G}_k} s(G).$$

**Lemma 2.** Assume that (1) is satisfied. Consider the following conditions.

- (i) Let  $\{\xi_n(\mathbf{s}) : \mathbf{s} \in D\}$ ,  $D \subseteq \mathbb{R}^d$ ,  $n = 1, 2, \dots$  be a sequence of strictly stationary  $m$ -dependent random fields.
- (ii) For all  $n$  and  $\mathbf{s}$ , let  $\mathbb{E}\xi_n(\mathbf{s}) = 0$  and  $\nu_n = \text{var}(\xi_n(\mathbf{s}))$ . Assume that  $\nu_n > 0$  if  $n$  is large enough.

(iii) For all  $\{\mathbf{s}'_{n1}, \dots, \mathbf{s}'_{nl}\} \subseteq \{\mathbf{s}_{n1}, \dots, \mathbf{s}_{nn}\}$ , let

$$\mathbb{E} \left| \prod_{j=1}^l \xi_n(\mathbf{s}'_{nj}) \right| = O \left( \nu_n^{\frac{l}{2}} \right) \quad \text{hold uniformly as } n \rightarrow \infty. \quad (5)$$

(iv) For a  $\delta > 0$  let

$$\nu_n + \kappa_n \tau_n \geq \delta \kappa_n \nu_n.$$

(iii') For all pairwise different elements  $\mathbf{s}'_{n1}, \dots, \mathbf{s}'_{nl}$  of the set  $\{\mathbf{s}_{n1}, \dots, \mathbf{s}_{nn}\}$ , relation (5) is satisfied uniformly,  $\kappa_n \nu_n^{1/2} \geq 1$  and  $|\xi_n(\mathbf{s})| \leq 1$  holds for all  $n$  and  $\mathbf{s}$ .

(iii'') Let  $\{\xi_n(\mathbf{s}_{ni}), i = 1, 2, \dots, n\}$  be a Gaussian system for each  $n \in \mathbb{N}$ .

Then, either from conditions (i)–(ii)–(iii)–(iv), (i)–(ii)–(iii')–(iv) or (i)–(ii)–(iii'')–(iv) it follows that

$$\begin{aligned} \text{for } k > \frac{r}{2}: \quad & \sum_{G \in \mathcal{G}_k} s(G) = 0, \\ \text{for } k < \frac{r}{2}: \quad & \frac{\sum_{G \in \mathcal{G}_k} s(G)}{(\text{var } S_n)^{\frac{r}{2}}} = o(1), \end{aligned}$$

and for  $k = \frac{r}{2}$  (if  $r$  is even):  $\sum_{G \in \mathcal{G}_k} s(G)$  gives the sum of products of the scores belonging to the components of size 2.

*Proof.* If  $\mathcal{I}_1, \dots, \mathcal{I}_k$  are the components of  $G$ , then  $s(G) = \mathbb{E}(\xi_{i_1} \cdots \xi_{i_r}) = \prod_{u=1}^k s(\mathcal{I}_u)$ , because for any two different components  $\mathcal{I}_u$  and  $\mathcal{I}_v$  the relation

$$\min_{i \in \mathcal{I}_u, j \in \mathcal{I}_v} \|\mathbf{s}_{ni} - \mathbf{s}_{nj}\| > m$$

is satisfied. (So the expected value falls into products because of the  $m$ -dependence.)

Now, let us determine the value of  $\sum_{G \in \mathcal{G}_k} s(G)$  depending on  $k$ .

Let  $k > r/2$ . Then each graph  $G$  in  $\mathcal{G}_k$  contains at least one component of size 1. Since  $\mathbb{E}(\xi_i) = 0$  follows from the conditions, therefore

$$\sum_{G \in \mathcal{G}_k} s(G) = 0.$$

It is obviously true in the case of either (iii), (iii') or (iii'').

Let  $k < r/2$ . Assume that (iii') is satisfied. We say that a component has degree  $l$  if it has  $l$  distinct vertices. Let  $G$  be a graph in  $\mathcal{G}_k$  with degrees  $l_1, \dots, l_k$  of components. Then, from (iii') it follows that

$$\begin{aligned} s(G) &= s(\mathcal{I}_1) \cdots s(\mathcal{I}_k) \\ &= O \left( \nu_n^{\frac{l_1}{2}} \right) \cdots O \left( \nu_n^{\frac{l_k}{2}} \right) \\ &= O \left( \nu_n^{\frac{1}{2} \sum_{u=1}^k l_u} \right). \end{aligned} \quad (6)$$

Let  $N_k$  be the number of the graphs in  $\mathcal{G}_k$  which have degrees  $l_1, \dots, l_k$  of components. Then, since in the  $u$ -th component there are  $l_u$  points of which we have maximum  $n$  possibilities to choose the first and maximum  $\kappa_n$  possibilities to choose the other points, we obtain that

$$\begin{aligned} N_k &\leq c_r \prod_{u=1}^k n \kappa_n^{l_u-1} \\ &= c_r n^k \kappa_n^{\sum_{u=1}^k (l_u-1)} \\ &= c_r n^k \kappa_n^{\sum_{u=1}^k l_u} \kappa_n^{-k}. \end{aligned} \quad (7)$$

Furthermore, based on (6) and (7), the sum of the scores of these graphs  $G$  satisfies the relation

$$\begin{aligned} \sum_{G \in \mathcal{G}_k} s(G) &\leq N_k \max s(G) \\ &= O \left( n^k \kappa_n^{-k} \kappa_n^{\sum_{u=1}^k l_u} \nu_n^{\frac{1}{2} \sum_{u=1}^k l_u} \right). \end{aligned} \quad (8)$$

By condition (iv) of the theorem, we know that  $\text{var}(S_n) \geq \delta n \kappa_n \nu_n$  for  $\delta > 0$  and considering also (8), we get that

$$\begin{aligned} \frac{\sum_{G \in \mathcal{G}_k} s(G)}{(\text{var} S_n)^{\frac{r}{2}}} &= O \left( \frac{n^k \kappa_n^{\sum_{u=1}^k l_u} \nu_n^{\frac{1}{2} \sum_{u=1}^k l_u}}{(\delta n \kappa_n \nu_n)^{\frac{r}{2}}} \right) \\ &= O \left( n^{k-\frac{r}{2}} \left( \kappa_n \nu_n^{\frac{1}{2}} \right)^{\sum_{u=1}^k l_u} \kappa_n^{-k-\frac{r}{2}} \nu_n^{-\frac{r}{2}} \right) \\ &= O \left( n^{k-\frac{r}{2}} \left( \kappa_n \nu_n^{\frac{1}{2}} \right)^r \kappa_n^{-k-\frac{r}{2}} \nu_n^{-\frac{r}{2}} \right) \\ &= O \left( \left( \frac{n}{\kappa_n} \right)^{k-\frac{r}{2}} \right) \\ &= o(1). \end{aligned} \quad (9)$$

In the last step we used condition (1).

If instead of condition (iii') we have condition (iii), then (9) can be obtained in the same way. The only difference is that  $l_u$  should be the number of all vertices of  $\mathcal{I}_u$  instead of the number of different vertices. Then in (8) we have  $O \left( \nu_n^{r/2} n^k \kappa_n^{r-k} \right)$ .

Now, if condition (iii'') is given instead of (iii), then we obtain another situation because of the normality. Here, using Hölder inequality, we have

$$\mathbb{E}|a_1 \cdots a_l| \leq \left( \prod_{i=1}^l \mathbb{E}|a_i|^l \right)^{\frac{1}{l}}$$

for  $a_1, \dots, a_l \in \mathbb{R}$ . It is known that

$$\begin{aligned} \mathbb{E}|\mathcal{N}(0, 1)|^p &= C_p \\ &= \frac{2^{\frac{p}{2}}}{\sqrt{\pi}} \Gamma \left( \frac{p+1}{2} \right). \end{aligned}$$

Therefore, we have

$$\begin{aligned} \mathbb{E} \left( \prod_{i=1}^l |\xi_i|^{p_i} \right) &\leq \prod_{i=1}^l (\mathbb{E} |\xi_i|^{p_i l})^{\frac{1}{l}} \\ &= \prod_{i=1}^l C_{p_i l} \nu_n^{\frac{p_i}{2}} \\ &\leq c_r \nu_n^{\frac{\sum_{i=1}^l p_i}{2}}. \end{aligned} \tag{10}$$

In this case,

$$\begin{aligned} s(G) &= O \left( \nu_n^{\frac{1}{2} \sum_{u=1}^k l_u} \right) \\ &= O \left( \nu_n^{\frac{r}{2}} \right), \end{aligned}$$

where  $l_u$  is the number of vertices in  $I_u$ . So the upper bound of the scores is the same as in (6). Hence, (9) is true.

If  $k = r/2$  (where  $r$  is even) then, since there are  $r$  factors and  $k = r/2$  components, each component with nonzero score contains exactly 2 factors. (The expectation of a component of one factor is 0 and if a component contains at least 3 factors, then another component contains 1 factor so the expectation is also 0 here.) Since the distance among the components is greater than  $m$ , these components of size 2 are independent from each other and hence the expectation of their product is equal to the product of their expectations. So the sum of products of the expectations of the components of size 2 is  $\sum_{G \in \mathcal{G}_{\frac{r}{2}}} s(G)$ .  $\square$

*Proof of Theorem 1.* First we remark that  $\text{var}(S_n) > 0$  because, by condition (iv),

$$\text{var}(S_n) \geq n \delta \kappa_n \nu_n > 0.$$

It is enough to show that the moments of  $S_n / \sqrt{\text{var}(S_n)}$  converge to the appropriate moments of the standard normal distribution since here the convergence in distribution holds by the theorem of Fréchet and Shohat (1931).

Let  $\{Y_1, \dots, Y_n\}$  be observations of a strictly stationary Gaussian random field where  $\mathbb{E}(Y_i) = 0$ ,  $\text{var}(Y_i) = \text{var}(\xi_i)$  and  $\text{cov}(Y_i, Y_j) = \text{cov}(\xi_i, \xi_j)$  hold for all  $i \neq j$ . Let  $Y_1, \dots, Y_n$  be located in the same way as  $\xi_1, \dots, \xi_n$ . As in the Gaussian case the independence and the zero correlation coincide, then because of the same covariance structure and the same appropriate locations we can state that  $\{Y_1, \dots, Y_n\}$  is  $m$ -dependent. Let us introduce the notation  $T_n = Y_1 + \dots + Y_n$ . Hence

$$\text{var}(S_n) = \text{var}(T_n). \tag{11}$$

Decomposing  $\{Y_1, \dots, Y_n\}$  into components can be accomplished in the same way as decomposing  $\{\xi_1, \dots, \xi_n\}$ .

Let  $s_Y(G)$  be the score function of  $\{Y_1, \dots, Y_n\}$ . By Lemma 2, for an odd  $r$

$$\begin{aligned} \mathbb{E} \left( \frac{S_n}{\sqrt{\text{var} S_n}} \right)^r &= \frac{\sum_{k=1}^r \sum_{G \in \mathcal{G}_k} s(G)}{(\text{var} S_n)^{\frac{r}{2}}} \\ &= o(1) \rightarrow 0 \\ &= \mathbb{E}(\mathcal{N}(0, 1))^r. \end{aligned}$$

Obviously,  $T_n/\sqrt{\text{var}(T_n)}$  has standard normal distribution. By Lemma 2, for an even  $r$

$$\begin{aligned} \mathbb{E}(\mathcal{N}(0, 1))^r &= \mathbb{E} \left( \frac{T_n}{\sqrt{\text{var}(T_n)}} \right)^r \\ &= \frac{\sum_{G \in \mathcal{G}_{\frac{r}{2}}} s_Y(G)}{(\text{var}(T_n))^{\frac{r}{2}}} + o(1). \end{aligned}$$

For  $G \in \mathcal{G}_{\frac{r}{2}}$  we have  $s(G) = s_Y(G)$ . Therefore, by Lemma 2 and equation (11), we obtain that

$$\begin{aligned} \mathbb{E} \left( \frac{S_n}{\sqrt{\text{var}(S_n)}} \right)^r &= \frac{\sum_{G \in \mathcal{G}_{\frac{r}{2}}} s(G)}{(\text{var}(S_n))^{\frac{r}{2}}} + o(1) \\ &= \frac{\sum_{G \in \mathcal{G}_{\frac{r}{2}}} s_Y(G)}{(\text{var}(T_n))^{\frac{r}{2}}} + o(1). \end{aligned}$$

From the previous equalities we get that

$$\mathbb{E} \left( \frac{S_n}{\sqrt{\text{var}(S_n)}} \right)^r = \mathbb{E}(\mathcal{N}(0, 1))^r + o(1)$$

holds. So the moments of  $S_n/\sqrt{\text{var}(S_n)}$  converge to the moments of the standard normal distribution. We also know that the distribution function determined by the moments of the standard normal law is unique. Then, by a theorem of Fréchet and Shohat (1931), we can state that there is a convergence in distribution.  $\square$

## 4 Multidimensional Case

The following theorem is the multivariate version of Theorem 1.

**Theorem 3.** *Assume that (1) is satisfied. Consider the following conditions:*

- (I) *Let  $\{\xi_n(\mathbf{s}) : \mathbf{s} \in D\}$ ,  $D \subseteq \mathbb{R}^d$ ,  $n = 1, 2, \dots$  be a sequence of strictly stationary  $m$ -dependent vector valued random fields of dimension  $q$ .*
- (II) *For all  $n$  and  $\mathbf{s}$ , let  $\mathbb{E}\xi_n(\mathbf{s}) = 0$  and let  $V_n = \text{var}(\xi_n(\mathbf{s}))$ . Assume that  $V_n$  is positive definite if  $n$  is large enough.*



(III) For all  $\{\mathbf{s}'_{n1}, \dots, \mathbf{s}'_{nl}\} \subseteq \{\mathbf{s}_{n1}, \dots, \mathbf{s}_{nn}\}$ , let

$$\mathbb{E} \left( \prod_{j=1}^l \|\xi_n(\mathbf{s}'_{nj})\| \right) = O \left( (\lambda_{\min}(V_n))^{\frac{l}{2}} \right) \quad \text{hold uniformly as } n \rightarrow \infty, \quad (12)$$

where  $\lambda_{\min}(V_n)$  denotes the smallest eigenvalue of  $V_n$ .

(III') For all pairwise different elements  $\mathbf{s}'_{n1}, \dots, \mathbf{s}'_{nl}$  of the set  $\{\mathbf{s}_{n1}, \dots, \mathbf{s}_{nn}\}$ , relation (12) is satisfied uniformly,  $\kappa_n (\lambda_{\min}(V_n))^{\frac{1}{2}} \geq 1$  and  $\|\xi_n(\mathbf{s})\| \leq 1$  for all  $n$  and  $\mathbf{s}$ .

(IV) For  $\delta > 0$ , let

$$V_n + \kappa_n T_n - \delta \kappa_n V_n$$

be a positive semidefinite matrix where

$$T_n = \frac{1}{n\kappa_n} \sum_{(i,j) \in \mathcal{T}_n} \text{cov}(\xi_n(\mathbf{s}_{ni}), \xi_n(\mathbf{s}_{nj})).$$

Let  $\xi_i = \xi_n(\mathbf{s}_{ni})$  and  $S_n = \xi_1 + \dots + \xi_n$ . Then either conditions (I)–(II)–(III)–(IV) or conditions (I)–(II)–(III')–(IV) imply that

$$S_n \text{ is asymptotically } \mathcal{N}(0, \text{var}(S_n)).$$

*Proof.* Let  $a \in \mathbb{R}^q$  be a fixed vector. Let us consider the scalar valued random fields  $a^\top \xi_n(\mathbf{s})$ . By the Cramér-Wold device (see Rao, 1965, pp. 103), we have to prove the convergence of  $a^\top S_n$ . Therefore, we have to check the conditions of Theorem 1 for  $a^\top \xi_n(\mathbf{s})$ . It is enough to consider  $a$  with the condition  $\|a\| = 1$ .

We see that  $\{a^\top \xi_n(\mathbf{s}), \mathbf{s} \in D\}$ ,  $n = 1, 2, \dots$  is a sequence of strictly stationary  $m$ -dependent random fields,  $\mathbb{E}(a^\top \xi_n(\mathbf{s})) = 0$  and  $\nu_n = \text{var}(a^\top \xi_n(\mathbf{s})) = a^\top V_n a$ ,  $\nu_n > 0$  for  $n$  large enough.

Now, it is true by (III) that

$$\begin{aligned} \mathbb{E} \left| \prod_{j=1}^l a^\top \xi_n(\mathbf{s}'_{nj}) \right| &\leq \|a\|^l \mathbb{E} \left( \prod_{j=1}^l \|\xi_n(\mathbf{s}'_{nj})\| \right) \\ &= \|a\|^l O \left( (\lambda_{\min}(V_n))^{\frac{l}{2}} \right) \\ &= O \left( (a^\top V_n a)^{\frac{l}{2}} \right) \\ &= O \left( (\nu_n)^{\frac{l}{2}} \right). \end{aligned} \quad (13)$$

Therefore (III) implies (iii).

Now let  $\mathbf{s}'_{n1}, \dots, \mathbf{s}'_{nl}$  be different points. Then (III') implies (13). Moreover,  $|a^\top \xi_n(\mathbf{s})| \leq \|a\| \|\xi_n(\mathbf{s})\| \leq 1$  if  $\|a\| = 1$  and  $\|\xi_n(\mathbf{s})\| \leq 1$ . So (III') implies (iii').

We see that  $\nu_n = a^\top V_n a$  and  $\tau_n = \frac{1}{n\kappa_n} \sum_{(i,j) \in \mathcal{T}_n} \text{cov}(a^\top \xi_n(\mathbf{s}_{ni}), a^\top \xi_n(\mathbf{s}_{nj})) = a^\top T_n a$ . Therefore, by (IV),

$$\nu_n + \kappa_n \tau_n - \delta \kappa_n \nu_n = a^\top (V_n + \kappa_n T_n - \delta \kappa_n V_n) a \geq 0,$$

i.e. (IV) implies (iv).

We see that  $\text{var}(S_n) = nV_n + n\kappa_n T_n$  is positive definite if  $n$  is large enough.  $\square$

## 5 Simulation Results

We present an example that gives numerical evidence for the phenomena described in the paper.

**Example: Two-dimensional moving average.** The locations are the  $l$ -lattice points of the domain  $D = [0, t]^2$  with  $l = 0.1$  and  $t = 10$ . It means that the distance between two neighbors is  $l = 0.1$ .

Then we consider the empirical distribution function  $F_n^*$  of  $\xi(\mathbf{s})$  (the process is stationary, so  $\xi(1), \dots, \xi(n)$  are identically distributed). Thus, the random field is  $z_{(i,j)} = \xi_{(i/10, j/10)}$ ,  $i, j = 1, \dots, 100$ . Let  $y_{k,l}$ ,  $k, l = 1, \dots, 102$ , be independent, uniformly distributed random variables on  $[0, 1]$  and let

$$z_{(i,j)} = \frac{1}{9} \sum_{k=i}^{i+2} \sum_{l=j}^{j+2} y_{k,l}, \quad i, j = 1, \dots, 100.$$

Therefore, the random field is  $m$ -dependent with  $m = 3$ .

Then some points from the locations are omitted. In Figure 2 we mark with dark the small squares where the locations are deleted. We see that in each white small square we have 16 sites of observations. Denote the set of the remaining locations by  $D$ . So the observations are  $z_{(i,j)}$ ,  $i, j \in D$ . Therefore the actual sample size is 7056.

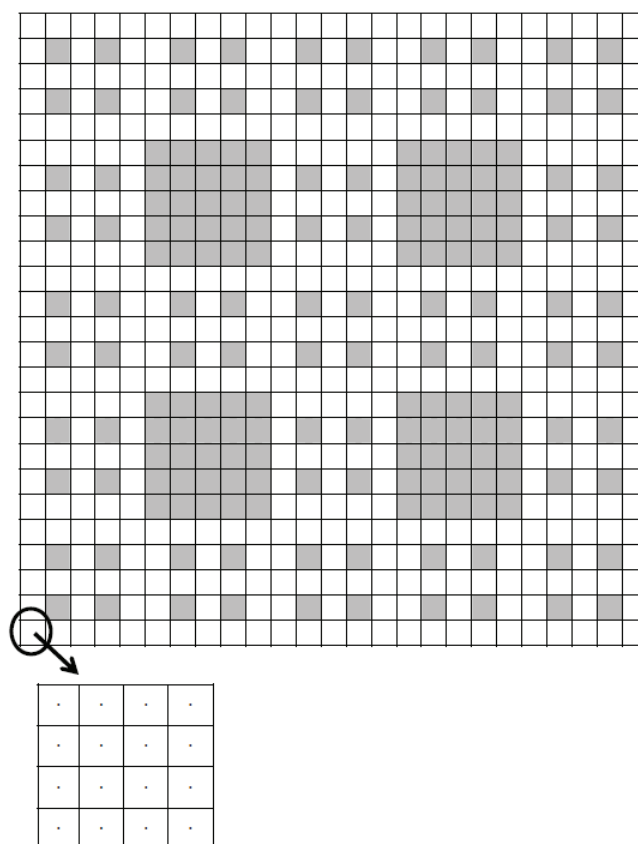


Figure 1: Sampling sites.

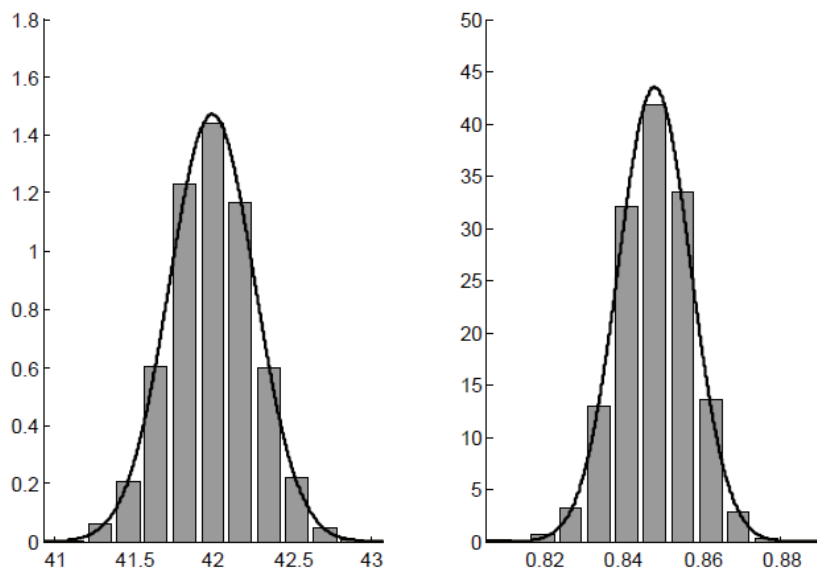


Figure 2: Histograms of  $S_n/\sqrt{n}$  and  $F_n^*(0.6)$  in the example.

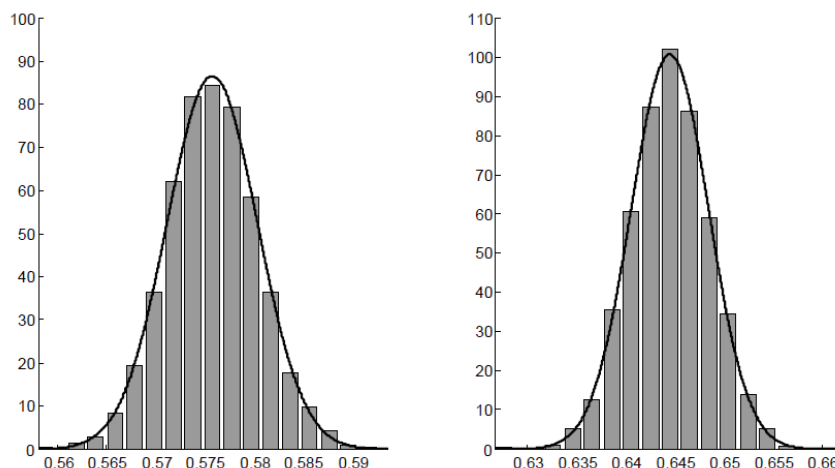


Figure 3: Histograms of  $(F_n^*(0.4) + F_n^*(0.8))/2$  and  $0.3F_n^*(0.3) + 0.4F_n^*(0.6) + 0.3F_n^*(0.8)$  in the example.

It can be seen that the resulted domain is not convex. In the above proposition the asymptotic properties of the estimator remain true. It is clearly shown by the following numerical results.

The simulation was performed with MATLAB, 5000 repetitions of the procedure were made.

In the left side of Figure 2 the histogram of  $S_n/\sqrt{n}$ ,  $n = 7056$ , is shown with the theoretical normal densities with mean and variance estimated from the data of the example. In the right side of Figure 2 the density histogram of  $F_n^*(0.6)$  can be seen with the theoretical normal densities with mean and variance estimated from the data of the

example.

In the left side of Figure 3 the density histogram of  $(F_n^*(0.4) + F_n^*(0.8))/2$  is shown with the theoretical normal densities where the mean and the variance are estimated from the data of the example. In the right side of Figure 3 the density histogram of  $0.3F_n^*(0.3) + 0.4F_n^*(0.6) + 0.3F_n^*(0.8)$  can be seen with the theoretical normal densities where the mean and the variance estimated from the data of the example. These figures show the joint asymptotic normality.

## Acknowledgement

The second-named author's research was carried out as part of the TAMOP-4.2.1.B-10/2/KONV-2010-0001 project with support by the European Union, co-financed by the European Social Fund.

## References

- Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes*. New York: Springer-Verlag.
- Cressie, N. A. C. (1991). *Statistics for Spatial Data*. New York: Wiley.
- Fazekas, I. (2003). Limit theorems for the empirical distribution function in the spatial case. *Statistics and Probability Letters*, 62, 251-262.
- Fazekas, I., and Chuprunov, A. (2004). A central limit theorem for random fields. *Acta Mathematica Academiae Paedagogicae Nyiregyhaziensis*, 20, 93-104.
- Fazekas, I., and Chuprunov, A. (2006). Asymptotic normality of kernel type density estimators for random fields. *Statistical Inference for Stochastic Processes*, 9, 161-178.
- Fréchet, M., and Shohat, J. (1931). A proof of the generalized second limit theorem in the theory of probability. *Transactions of the American Mathematical Society*, 33, 533-543.
- Karácsony, Z., and Filzmoser, P. (2010). Asymptotic normality of kernel type regression estimators for random fields. *Journal of Statistical Planning and Inference*, 140, 872-886.
- Lahiri, S. N. (1999). Asymptotic distribution of the empirical spatial cumulative distribution function predictor and prediction bands based on a subsampling method. *Probability Theory and Related Fields*, 114, 55-84.
- Lahiri, S. N., Kaiser, M. S., Cressie, N. A. C., and Hsu, N. J. (1999). Prediction of spatial cumulative distribution functions using subsampling. *Journal of the American Statistical Association*, 94, 86-110.
- Park, B. U., Kim, T. Y., Park, T.-S., and Hwang, S. Y. (2009). Practically applicable central limit theorem for spatial statistics. *Mathematical Geosciences*, 41, 555-569.
- Rao, C. R. (1965). *Linear Statistical Inference and Its Applications* (2nd ed.). New York: Wiley-Interscience.

Authors' addresses:

István Fazekas, Renáta Vas  
Faculty of Informatics  
University of Debrecen  
Kassai út 26  
H-4032 Debrecen  
Hungary  
E-mail: fazekasi@inf.unideb.hu, vas.renata@inf.unideb.hu

Zsolt Karácsony  
Department of Applied Mathematics  
University of Miskolc  
H-3515 Miskolc-Egyetemváros  
Hungary  
E-mail: matkzs@uni-miskolc.hu