On the Asymptotic Behavior in Random Fields: The Central Limit Theorem

Mohammad Mehdi Saber¹, Zohreh Shishebor²*, Behnam Amiri²

¹Department of Statistics, Higher Education Center of Eghlid, Eghlid, Iran
²Department of Statistics, Shiraz University, Shiraz, Iran

ABSTRACT

The aim of this paper is to provide an applicable version of Central Limit Theorem for strictly stationary m-dependent random fields on a lattice. The type of sampling is considered increasing domain sampling.

ARTICLE INFO

Article History
Received XX XXX XXXX
Accepted XX XXX XXXX

Keywords
Central Limit Theorem
Increasing domain sampling
m-dependent

1. INTRODUCTION

A random field (RF), \( Z = \{Z(s), s \in S\}, S \subseteq \mathbb{R}^k \), is a collection of random variables indexed by \( s \). RFs are used in a variety of fields including geology, geography, biology, earth science, health, medicine, image processing, etc. A lot of research has been done in this field during the past few years; nonetheless many key statistical topics have not been answered yet. The aim of this study is to investigate the asymptotic properties of some statistics in RFs under increasing domain sampling. Two important statistics are the sample mean and sample variance, which are used to estimate the mean and variance of a population, respectively. Usually the exact distributions of these two statistics is unknown. In this study, we are interested in asymptotic behavior of sample mean \( \bar{Z}_n \) and its limiting distribution. So, different versions of spatial sampling are needed to be explained. “When the sampling region remains bounded as \( n \) tends to infinity but the sample size \( n \) grows to infinity it is called infill domains sampling” [3]. Lahiri [8] proved that in this case, the Central Limit Theorem (CLT) for \( \sum_{i=1}^{n} Z_i \) cannot be hold. “If sampling region becomes unbounded with \( n \) and the structure of \( S_n \) is such that there is a minimum distance between any two sites for all \( n \), it is called increasing domain sampling” [3] which is our focus in this work. “The third one is mixed (or nearly infill) domain sampling, where the infill domain sampling is mixed with the increasing domain sampling.” Lahiri [9] established a spatial CLT for the nearly infill domain sampling. He considered both fixed and random designs. Following the work of Lahiri, Byeong U. Park et al. [2] earned a CLT for m-dependent and stationary RFs with more convenient conditions satisfied.

Various types of CLT for spatial processes have been considered by some authors. Among them, Stein [12] proved consistency and CLT for weighted sum of RFs under stochastic sampling. Jenish and Prucha [6] proved CLTs and uniform law of large numbers for arrays of RFs. They [7] also, proved a CLT and law of large numbers under near-epoch dependence. A Lindeberg CLT for strictly stationary RFs was provided by Cristina Tone in 2013 [13]. Maltz and Samur [10] worked on a uniform CLT on rectangles for functions of mixing RFs. Berkes et al. [1] have achieved some asymptotic results for the empirical process of stationary sequences. Machkouri et al. [11] have proved a CLT for stationary and m-dependent RFs. In this work a CLT for strictly stationary m-dependent RFs is provided under increasing domain sampling. The interesting finding in our study is that the proposed theorem only needs to have finite variance.

The paper is organized as follows. In Section 2, some requirements for CLT such as, the asymptotic behavior of \( \frac{\sum_{i=1}^{n} Z_i}{\alpha} \) for some appropriate values of \( \alpha \) for two cases of m-dependency and ergodicity are studied. Section 3 will give the limiting distribution of \( \bar{Z}_n \) in the case that \( Z \) is a strictly stationary m-dependent RF (Theorem 2). A conclusion is given in Section 4.

* Corresponding author. Email: shisheb@shirazu.ac.ir
2. BASIC CONCEPTS AND PRELIMINARY RESULTS

Here and after we suppose that the type of sampling is increasing domain sampling. In linear algebra, the size of a vector \( x \) is named the norm of \( x \). There are different kinds of norms on \( \mathbb{R}^k \). In order to have less complicated computations the max norm is applied here which is defined as follows:

\[
\|x\|_\infty = \max \left(|x_1|, \ldots, |x_k|\right) \quad \text{for } x = (x_1, \ldots, x_k) \in \mathbb{R}^k.
\]

We define two types of sets

\[
B(x_0, r) = \{x \in \mathbb{R}^k \mid \|x - x_0\|_\infty \leq r\} \quad \text{(closed Ball)},
\]

\[
S(x_0, r) = \{x \in \mathbb{R}^k \mid \|x - x_0\|_\infty = r\} \quad \text{(sphere)}.
\]

In the two cases, \( r \) is called the radius and \( x_0 \) is the center. For clarification, we emphasize that \( B(x_0, r) \) is a solid square and \( S(x_0, r) \) is a sphere. Also, it is worth to mention that max and Euclidian norms are equivalent and induce the same topology on \( \mathbb{R}^k \). So, all results are valid with the usual Euclidian topology.

Let us define the concepts of stationarity, strict stationarity and \( m \)-dependence.

**Definition 2.1.** The RF \( Z = \{Z(s), s \in S \subseteq \mathbb{R}^k\} \) is said to be stationary if \( \forall s, t \in S \)

(i) \( E(Z(s)) = a \),

(ii) \( E(|Z(s)|^2) < \infty \) and

(iii) \( \text{cov}(Z(s), Z(t)) = \gamma \left(||s - t||_\infty\right) \)

**Definition 2.2.** The RF \( Z = \{Z(s), s \in S \subseteq \mathbb{R}^k\} \) is said to be strictly stationary if \( (Z(s_1 + h), \ldots, Z(s_m + h)) \overset{D}{=} (Z(s_1), \ldots, Z(s_m)), \) for all \( s_1, s_2, \ldots, s_m, h \in S \), where \( D \) stands for identically distributed.

**Definition 2.3.** The RF \( Z = \{Z(s), s \in S \subseteq \mathbb{R}^k\} \) is said to be \( m \)-dependent if \( Z(s) \) and \( Z(t) \) are independent for every \( s, t \in S \) such that \( ||s - t||_\infty > m \).

Whenever we have weak stationarity, two symbols \( \sigma_0^2 = \text{var}(Z(s)) \) and \( \sigma_{|s-t|_\infty}^2 = \text{cov}(Z(s), Z(t)) \) are used. For simplicity of notation we write \( Z_i \) instead of \( Z(s_i) \), when no confusion can arise.

The following result is used to prove the main issue. It also has some direct result concerning the limiting distribution of \( \overline{Z}_n \). So we state it as a theorem.

**Theorem 1.** Suppose that \( \Lambda \) is an \( N^k \) lattice in \( \mathbb{R}^k \) with coordinates \( s_1, \ldots, s_n \) and cell width \( \delta = 1 \) where \( n = N^k \). Let \( Z \) be a mean zero RF on \( \Lambda \) such that \( \text{Var}(Z_i) \leq c < \infty \) for all \( i = 1, \ldots, n \).

i. If \( \rho(Z_i, Z_j) \to 0 \), as \( ||s_i - s_j||_\infty \) tends to infinity and if \( \alpha \geq 1 \), then \( \frac{\sum_{i=1}^{n} Z_i}{n^\alpha} \) tends to zero.

ii. If \( Z \) is an \( m \)-dependent RF, then \( \frac{\sum_{s \in A} Z(s)}{n^\alpha} \) tends to zero for all \( \alpha > \frac{\ln(\#(A))}{2 \ln(n)} \), where \( \#(A) \) shows the number of members in set \( A \).

iii. If \( Z \) is a stationary \( m \)-dependent RF, then \( \lim_{n \to +\infty} \frac{\text{var} \left( \frac{\sum_{i=1}^{n} Z_i}{n} \right)}{\text{var} \left( \frac{\sum_{i=1}^{m} Z_i}{m} \right)} = \sigma_0^2 + \sum_{j=1}^{m} \left( 2j + 1 \right)^k \left( 2j - 1 \right)^k \sigma_j^2. \)

**Proof:** (i) Since \( \frac{\sum_{i=1}^{n} Z_i}{n^\alpha} \) is an unbiased estimator of zero, it is enough to show that \( \text{Var} \left( \frac{\sum_{i=1}^{n} Z_i}{n^\alpha} \right) \) goes to zero as \( n \) goes to infinity. So, let us compute the variance of \( \frac{\sum_{i=1}^{n} Z_i}{n^\alpha} \).

\[
\text{Var} \left( \frac{\sum_{i=1}^{n} Z_i}{n^\alpha} \right) = \frac{1}{n^{2\alpha}} \sum_{i=1}^{n} \sum_{j=1}^{n} \text{cov}(Z_i, Z_j) = \frac{1}{n^{2\alpha}} \sum_{i=1}^{n} \sum_{j=1}^{n} \sqrt{\text{Var}(Z_i) \text{Var}(Z_j)} \rho(Z_i, Z_j) \
\leq \frac{c}{n^{2\alpha}} \sum_{i=1}^{n} \sum_{j=1}^{n} |\rho(Z_i, Z_j)|.
\]

The uniform convergence to zero of the sequence \( \left\{ \rho(Z_i, Z_j) \right\} \), implies that for all \( \varepsilon > 0 \), there exists an integer \( N_{\varepsilon} \) large enough such that \( |\rho(Z_i, Z_j)| \leq \varepsilon \) for all \( Z_i \) and \( Z_j \) whenever \( ||s_i - s_j||_\infty > N_{\varepsilon} \). The number of points in \( B(s_i, N_{\varepsilon}) \) is at most \( (2N_{\varepsilon})^k \). Thus the total number of points with condition \( ||s_i - s_j||_\infty \leq N_{\varepsilon} \), \( i, j = 1, \ldots, n \), are at last \( n (2N_{\varepsilon})^k \). The computation of the later inequality can be followed as
The last term tends to zero as \( n \) tends to infinity if \( \alpha \geq 1 \), and this completes the proof of part (i).

(ii) Following the same method as in part (i), we have

\[
\text{Var} \left( \frac{\sum_{s \in A} Z(s)}{n^\alpha} \right) \leq \frac{c}{n^{2\alpha}} \sum_{s \notin A} \sum_{t \in A} |\rho(Z(s), Z(t))| \leq \frac{c}{n^{2\alpha}} \left(n (2N_\epsilon)^k + \left(n^2 - n (2N_\epsilon)^k\right) \epsilon \right)
\]

Let \( \#(A) \propto n^\delta \), then if \( 2\alpha > q \) results in \( \text{Var} \left( \frac{\sum_{s \in A} Z(s)}{n^\alpha} \right) \) tends to 0 as \( n \) tends to infinity. Note that \( \#(A) \propto n^\delta \) is equivalent to \( q \propto \frac{\ln(\#(A))}{\ln(n)} \). This completes the proof.

(iii) By \( m \)-dependence and stationarity we have

\[
\text{var} \left( \sum_{i=1}^n Z_i \right) = \sum_{i=1}^n \sum_{j=1}^n \text{cov}(Z_i, Z_j) = \sum_{i=1}^n \sum_{s \in B(i, m)} \sigma^2_{\|s-i\|_\infty} = \sum_{s \in Q} \sum_{i \in B(s, m)} \sigma^2_{\|s-i\|_\infty} + \sum_{j \in Q} \sum_{i \in B(s, m)} \sigma^2_{\|s-i\|_\infty}
\]

where \( Q \) is the set of all indexes for which \( \#(B(s, m)) = (2m + 1)^k \). It is easy to see that \( \#(Q) = \left( \frac{n^k - 2m}{n^k - 2m} \right)^k \).

Let us compute the value of the first term in the above expression. By stationarity \( \sum_{s \in B(s, m)} \sigma^2_{\|s-i\|_\infty} = \frac{1}{n^k - 2m} \) \( \sum_{s \in B(s, m)} \sigma^2_{\|s-i\|_\infty} \). The set \( B(s, m) \) will be partitioned into \( m + 1 \) disjoint sets \( S(s, j), j = 0, ..., m \). Since for all \( s \in S(s, j) \) we have \( \|s_i - s\|_\infty = j \), the value of \( \#S(s, j) \) must be computed. Clearly \( \#S(s, 0) = \#s \) and \( \#S(s, 0) = 1 \). Also, by construction we have \( \#S(s, j) = \#B(s, j) = (2j + 1)^k - (2j - 1)^k \), \( j = 1, ..., m \).

To see more clearly see Fig. 1 for \( k = 2 \) and \( j = 1 \). Spurred by this finding we have \( \sum_{s \in B(s, m)} \sigma^2_{\|s-i\|_\infty} = \sigma^2_0 + \sum_{j=1}^m (2j + 1)^k - (2j - 1)^k \sigma^2_j \). To calculate the second term note that \( |\sum_{s \in B(s, m)} \sigma^2_{\|s-i\|_\infty} \leq |\sigma^2_0| + \sum_{j=1}^m (2j + 1)^k - (2j - 1)^k \sigma^2_j = M < +\infty \) for \( i \notin Q \). Thus, \( \sum_{s \in B(s, m)} \sigma^2_{\|s-i\|_\infty} \leq \left(n - \left( \frac{1}{n^k - 2m} \right)^k \right) \sigma^2_0 + \sum_{j=0}^{k-1} \left( j \right)(2m)^{k-j} \sigma^2_j \). Finally, \( \lim_{n \to +\infty} \sigma^2_0 \sum_{j=0}^{k-1} \left( j \right)(2m)^{k-j} M \) and \( \lim_{n \to +\infty} \sigma^2_0 \sum_{j=0}^{k-1} \left( j \right)(2m)^{k-j} M = 0 \) which complete the proof.

We finish this section with a direct result of Theorem 1.

**Corollary 1.** (Consistency of \( \overline{Z}_n \)) If \( Z \) is a RF with non-zero mean \( \overline{Z}_n \to \mu \).

**Remark 1.** We point out that the obtained results are maintained under the Euclidian topology, with \( \#(B(s, r)) \leq (2r + 1)^k \).
3. CLT FOR RFS ON A REGULAR LATTICE

In this section a CLT for strictly stationary m-dependent RFs on a regular lattice under increasing domain sampling has been established. First, let us state the basic lemma from Ferguson ([4], p. 77) in advanced probability.

Lemma 1. Suppose \( Y_n = Z_{mq} + X_{mq} \) for \( n = 1, 2, \ldots \) and \( n = 1, 2, \ldots \). If

i. \( X_{mq} \xrightarrow{D} 0 \) uniformly in \( n \) as \( q \) tends to infinity,

ii. \( Z_{mq} \xrightarrow{D} Z_q \) as \( n \) tends to infinity for all \( q \),

iii. \( Z_q \xrightarrow{D} Z_j \) as \( q \) tends to infinity,

then, \( Y_n \rightarrow Z \) as \( n \) tends to infinity.

Now, we state the CLT and we will prove it using Lemma 1.

Theorem 2. Let \( \Lambda \) be a square lattice in \( \mathbb{R}^k \) with coordinates, \( s_1, \ldots, s_n \) and cell width \( \delta = 1 \). Let \( Z \) be a strictly stationary m-dependent RF with finite variance \( \sigma_x^2 \) on this lattice. If \( T_n = \sum_{n=1}^{\infty} Z(s) \), then \( \frac{T_n - \mu n}{\sigma \sqrt{n}} \xrightarrow{D} N(0, 1) \) where \( \sigma^2 = \left( \sigma_x^2 + \sum_{j=1}^{m} \left( (2j + 1)^k - (2j - 1)^k \right) \sigma^2 \right) \) and \( \sigma_{\perp s_0} = \text{cov}(Z(s_0), Z(s)) \).

Proof: Without loss of generality we can assume that \( \mu = 0 \). The method of the proof involves splitting the sum \( T_n \) into two parts, one being a sum of independent random variables for which the CLT can be applied, and the other sum containing negligible terms. Let \( n = (b (m + q) + e)^k \), \( 0 \leq e < m + q \) and \( q > m \) where \( m, q \) and \( e \) are positive integers. It is worth to mention that, \( m, q \) and \( e \) do not depend on \( n \) but \( b \) increases as \( n \) increases. We break the lattice \( \Lambda \) into \( b^k \) sections \( V_n^{(u_1, \ldots, u_{b-k})} \), \( u_i = 1, \ldots, b \) of size \( (m + q)^k \) and a remainder \( R_n \). By construction, \( R_n \) can be divided into sections that at least one of their dimensions is equal with \( e \). The set \( T_n \) can be decomposed as

\[
T_n = \sum_{u_1=1}^{b} \ldots \sum_{u_{b-k}=1}^{b} \sum_{j \in W_n^{(u_1, \ldots, u_{b-k})}} Z(s) + \sum_{j \in R_n} Z(s). \tag{1}
\]

The section \( W_n^{(u_1, \ldots, u_{b-k})} \) is broken down into \( 2^k \) subsections \( V_{t_1,t_2,\ldots,t_k}^{(u_1, \ldots, u_{b-k})} \), \( t_i = 1, 2 \).

\[
W_n^{(u_1, \ldots, u_{b-k})} = \bigcup_{t_1=1}^{2} \ldots \bigcup_{t_k=1}^{2} V_{t_1,t_2,\ldots,t_k}^{(u_1, \ldots, u_{b-k})}. \tag{2}
\]

The \( i \)th dimension of \( V_{t_1,t_2,\ldots,t_k}^{(u_1, \ldots, u_{b-k})} \) in (2) is equal with \( q \) and \( m \) for \( t_i = 1 \) and \( t_i = 2 \), respectively. Let \( T_{t_1,t_2,\ldots,t_k} = \bigcup_{u_i=1}^{b} \ldots \bigcup_{u_{b-k}=1}^{b} V_{t_1,t_2,\ldots,t_k}^{(u_1, \ldots, u_{b-k})} \). Therefore (1) can be rewritten as

\[
T_n = \sum_{t_1=1}^{2} \ldots \sum_{t_k=1}^{2} \sum_{j \in T_{t_1,t_2,\ldots,t_k}} Z(s) + \sum_{j \in R_n} Z(s). \tag{3}
\]
To have a better understanding of definitions see Fig. 2 for the special case \(k = 2\) and \(b = 2\). To complete the proof, it suffice to apply Lemma 1 for \(Y_n = \frac{T_n}{\sqrt{n}}\), \(Z_{nq} = \frac{\sum_{s \in T_{1,1},...,1} Z(s) + \sum_{s \in R_n} Z(s)}{\sqrt{n}}\) and \(X_nq = \frac{1}{\sqrt{n}} \sum_{t_{1} = 1}^{k} \ldots \sum_{t_{k} = 1}^{k} \sum_{s \in T_{t_{1},...,t_{k}}} Z(s)\).

The CLT along with the strict stationarity assumption guarantees that for any fixed \(q\),
\[
\frac{\sum_{s \in T_{1,1},...,1} Z(s)}{\sqrt{n}} \xrightarrow{D} N \left( 0, \text{var} \left( \sum_{s \in V_{1,1}^{1,1}} Z(s) \right) \right).
\]

Since \(\lim_{n \to \infty} \frac{b^k}{n} = \frac{1}{(m + q)^k}\), it follows from Slutsky theorem that
\[
\sum_{s \in T_{1,1},...,1} Z(s) \sqrt{n} \xrightarrow{D} N \left( 0, \frac{\text{var} \left( \sum_{s \in V_{1,1}^{1,1}} Z(s) \right)}{(m + q)^k} \right).
\]

Let \(r_n\) indicate the number of indices used in \(R_n\). Then,
\[
\lim_{n \to \infty} \frac{r_n^{2(k-1)}}{\sqrt{n}} = \lim_{b \to \infty} \frac{\left( k \left( b (m + q) \right)^{k-1} e \right)^{2(k-1)}}{(b (m + q) + e)^{2}} = (ke)^{2(k-1)}.
\]

The first equality comes from the fact that \(r_n = (b (m + q) + e)^k - (b (m + q))^k\) and \(n = (b (m + q) + e)^k\). Since \(\frac{k}{2(k-1)} > \frac{1}{2}\) for every \(k\), from Part (ii) of Theorem 1, we have that
\[
\sum_{s \in R_n} Z(s) \sqrt{\frac{k}{r_n^{2(k-1)}}} \xrightarrow{L^2} 0 \text{ as } n \to \infty.
\]

(5) and (6) imply that
\[
\sum_{s \in R_n} Z(s) \sqrt{\frac{k}{r_n^{2(k-1)}}} \xrightarrow{L^2} 0 \text{ as } n \to \infty.
\]

Now it follows from (4) and (7) that \(Z_{nq} \xrightarrow{D} Z_q \sim N \left( 0, \frac{\text{var} \left( \sum_{s \in V_{1,1}^{1,1}} Z(s) \right)}{(m + q)^k} \right)\). Finally part (iii) of Theorem 1 implies that \(Z_q \xrightarrow{} N \left( 0, \sigma^2 \right)\) as \(q\) tends to infinity.

![Figure 2](image-url) For parameters \(b = 2\) and \(k = 2\), the lattice has been broken. Here, blue color shows \(T_{1,1}\).
It remains to check the first condition of Lemma 1. In part (ii) of Theorem 1 let \( A \) be the set \( T_{t_1, t_2, \ldots, t_k}, (t_1, \ldots, t_k) \neq (1, \ldots, 1) \) and \( \alpha = \frac{1}{2} \).

Since \#(A) \propto q^{k-1} \) and \( n \propto q^k \), we have \( \frac{\sum_{q=1}^{Z_n} Z_n}{\sqrt{n}} \) \( \sim \) \( n \) \( \to 0 \) as \( q \) tends to infinity if \( \frac{1}{2} > \frac{\log q}{\log (q^p)} = k = 1 - \frac{1}{2} k \). Clearly the last condition is satisfied for every \( k \). This means that \( X_{nq} \to 0 \) as \( q \) tends to infinity, uniformly in \( n \).

\[
W_n^{(1,1)} = V_{1,1} + V_{1,2} + V_{2,1} + V_{2,2}, \quad W_n^{(1,2)} = V_{1,1} + V_{1,2} + V_{2,1} + V_{2,2},
\]
\[
W_n^{(2,1)} = V_{1,1} + V_{1,2} + V_{2,1} + V_{2,2}, \quad W_n^{(2,2)} = V_{1,1} + V_{1,2} + V_{2,1} + V_{2,2},
\]
\[
T_{1,1} = V_{1,1} + V_{1,2} + V_{2,1} + V_{2,2}, \quad T_{1,2} = V_{1,1} + V_{1,2} + V_{2,1} + V_{2,2},
\]
\[
T_{2,1} = V_{1,1} + V_{1,2} + V_{2,1} + V_{2,2}, \quad T_{2,2} = V_{1,1} + V_{1,2} + V_{2,1} + V_{2,2}.
\]

4. CONCLUSION

In this work under increasing domain sampling the asymptotic behavior of \( \frac{\sum_{q=1}^{Z_n} Z_n}{n^2} \) for some appropriate value of \( \alpha \) has been given in two cases, the case of ergodicity and the case of \( m \)-dependence. In the latter case a version of CLT was given. Extending this work to other sampling modes is an interesting work which it should be stated in the future.

REFERENCES

6. N. Jenish, P.R. Ingmar, J. Econ. 150 (2009), 86–98.