CENTRE FOR **STOCHASTIC GEOMETRY**
AND ADVANCED **BIOIMAGING**

Christophe Ange Napoléon Biscio and Rasmus Waagepetersen

A general central limit theorem and subsampling variance estimator for α-mixing point processes

No. 05, April 2018

A general central limit theorem and subsampling variance estimator for α-mixing point processes

Christophe Ange Napoléon Biscio and Rasmus Waagepetersen

Department of Mathematical Sciences, Aalborg University, Denmark

Abstract

We establish a central limit theorem for multivariate summary statistics of non-stationary α-mixing spatial point processes and a subsampling estimator of the covariance matrix of such statistics. The central limit theorem is crucial for establishing asymptotic properties of estimators in statistics for spatial point processes. The covariance matrix subsampling estimator is flexible and model free. It is needed e.g. to construct confidence intervals and ellipsoids based on asymptotic normality of estimators. We also provide a simulation study investigating an application of our results to estimating functions.

Keywords: α-mixing, central limit theorem, estimating function, random field, spatial point process, subsampling, summary statistics.

1 Introduction

Let $\mathbf{X}$ denote a spatial point process on $\mathbb{R}^d$ observed on some bounded window $W \subset \mathbb{R}^d$. In statistics for spatial point processes, much interest is focused on possibly multivariate summary statistics or estimating functions $T_W(\mathbf{X})$ of the form

$$T_W(\mathbf{X}) = \sum_{u_1,\dots,u_p \in \mathbf{X}\cap W}^{\neq} h(u_1, \dots, u_p) \tag{1.1}$$

where $h : \mathbb{R}^p \to \mathbb{R}^q$, $p, q \geq 1$, and the $\neq$ signifies that summation is over pairwise distinct points. Central limit theorems for such statistics have usually been developed using either of the two following approaches, both based on assumptions of α-mixing. One approach uses Bernstein's blocking technique and a telescoping argument that goes back to Ibragimov and Linnik (1971, Chapter 18, Section 4). This approach has been used in a number of papers like Guan and Sherman (2007), Guan and Loh (2007), Prokešová and Jensen (2013), Guan et al. (2015), and Xu et al. (2018). The other approach is due to Bolthausen (1982) who considered stationary random fields and whose proof was later generalised to non-stationary random fields by Guyon (1995) and Karácsony (2006). This approach is e.g. used in Waagepetersen and Guan (2009), Coeurjolly and Møller (2014), Biscio and Coeurjolly (2016), Coeurjolly (2017), and Poinas et al. (2017). Regarding the point process references mentioned above, it is characteristic that essentially the same central limit theorems are (re-)invented again and again for each specific setting and statistic considered. We therefore find it useful to provide a unified framework to state, once and for all, a central limit theorem under general non-stationary settings for multivariate point process statistics $T_W(\mathbf{X})$ admitting certain additive decompositions. We believe this can save a lot of work and tedious repetitions in future applications of α-mixing point processes. The framework of α-mixing is general and easily applicable to e.g. Cox and cluster point processes and a wide class of determinantal point processes (DPPs) (Poinas et al., 2017). For certain model classes other approaches may be more relevant. For Gibbs processes it is often convenient to apply central limits for conditionally centered random fields (Jensen and Künsch, 1994; Coeurjolly and Lavancier, 2017) while Heinrich (1992) developed a central limit theorem specifically for the case of Poisson cluster point processes using their strong independence properties.

Consider for example (1.1) and assume that $\{C(\mathbf{l})\}_{\mathbf{l}\in\mathbb{L}}$ forms a disjoint partitioning of $\mathbb{R}^d$. Then we can decompose $T_W(\mathbf{X})$ as

$$T_W(\mathbf{X}) = \sum_{\mathbf{l}\in\mathbb{L}} f_{\mathbf{l},W}(\mathbf{X}) \tag{1.2}$$

with

$$f_{\mathbf{l},W}(\mathbf{X}) = \sum_{u_1\in\mathbf{X}\cap C(\mathbf{l})\cap W} \sum_{u_2,\dots,u_p\in(\mathbf{X}\cap W)\setminus\{u_1\}}^{\neq} h(u_1,\dots,u_p).$$

Thus $T_W(\mathbf{X})$ can be viewed as a sum of the variables in a discrete index set random field $\{f_{\mathbf{l},W}(\mathbf{X})\}_{\mathbf{l}\in\mathbb{L}}$. This is covered by our set-up provided h satisfies certain finite range conditions, see the following sections for details. In connection to the Bolthausen approach, we remark that Guyon (1995) does not cover the case where the function f in (1.2) depends on the observation window. This kind of generalisation is e.g. needed in Jalilian et al. (2017). By considering triangular arrays, Karácsony (2006) is more general than Guyon (1995), but Karácsony (2006) on the other hand considers a combination of increasing domain and infill asymptotics that is not so natural in a spatial point process framework. Moreover, the results in Guyon (1995) and Karácsony (2006) are not applicable to non-parametric kernel estimators depending on a band width converging to zero. Using Bolthausen's approach, we establish a central limit theorem that does not have these limitations. For completeness we also provide in the supplementary material a central limit theorem based on Bernstein's blocking technique and we discuss why its conditions may be more restrictive than those for our central limit theorem.

A common problem regarding application of central limit theorems is that the variance of the asymptotic distribution is intractable or difficult to compute. However, knowledge of the variance is needed for instance to assess the efficiency of an estimator or to construct confidence intervals and ellipsoids. Bootstrap and subsampling methods for estimation of the variance of statistics of random fields have been studied in e.g. Politis and Romano (1994) and Lahiri (2003). For statistics of points processes, these methods have been considered in e.g. Guan and Sherman (2007), Guan and Loh (2007), Loh (2010) and Mattfeldt et al. (2013) but they have been limited to stationary or second-order intensity reweighted stationary point processes in $\mathbb{R}^2$ and only for estimators of the intensity and Ripley's K-function. For general statistics of the form (1.2), we adapt results from Sherman (1996) and Ekström (2008) to propose a subsampling estimator of the variance. We establish its asymptotic properties in the framework of a possibly non-stationary α-mixing point process and discuss its application to estimate the variance of point process estimating functions. The good performance of our subsampling estimator is illustrated in a simulation study considering coverage of approximate confidence intervals when estimates of intensity function parameters are obtained by composite likelihood.

In Section 2 we define notation and the different α-mixing conditions used in our paper. Section 3 states the central limit theorem based on Bolthausen's technique and the subsampling estimator is described in Section 4. Application of our subsampling estimator to estimating functions is discussed in Section 5 and is illustrated in a simulation study in Section 6. Finally, our subsampling estimator is discussed in relation to other approaches in Section 7. The proofs of our results are presented in the Appendix. A discussion on Bernstein's blocking technique approach, technical lemmas, and some extensive technical derivations are provided in the supplementary material.

2 Mixing spatial point processes and random fields

For $d \in \mathbb{N} = \{1, 2, \ldots\}$, we define a random point process $\mathbf{X}$ on $\mathbb{R}^d$ as a random locally finite subset of $\mathbb{R}^d$ and refer to Daley and Vere-Jones (2003) and Daley and Vere-Jones (2008) for measure theoretical details. We define a lattice $\mathbb{L}$ as a countable subset of $\mathbb{Z}^d$ where $\mathbb{Z} = \mathbb{N} \cup \{0, -1, -2, \ldots\}$. When considering vertices of a lattice, we use bold letter, for instance $\mathbf{i} \in \mathbb{Z}^d$. We define

$$\mathrm{d}(x, y) = \max\{|x_i - y_i| : 1 \leq i \leq d\}, \ x, y \in \mathbb{R}^d.$$

Reusing notation we also define

$$\mathrm{d}(A, B) = \inf\{\mathrm{d}(x, y) : x \in A, \ y \in B\}, \ A, B \subset \mathbb{R}^d.$$

For a subset $A \subseteq \mathbb{R}^d$ we denote by $|A|$ the cardinality or Lebesgue measure of A. The meaning of $|\cdot|$ and $\mathrm{d}(\cdot, \cdot)$ will be clear from the context. Moreover, for $R \geq 0$, we define $A \oplus R = \{x \in \mathbb{R}^d : \inf_{y \in A} d(x, y) \leq R\}$.

The α-mixing coefficient of two random variables X and Y is

$$\begin{aligned}\alpha(X, Y) &= \alpha(\sigma(X), \sigma(Y)) \\ &= \sup\{|P(A \cap B) - P(A)P(B)| : A \in \sigma(X), B \in \sigma(Y)\},\end{aligned}$$

where $\sigma(X)$ and $\sigma(Y)$ are the σ-algebras generated by X and Y, respectively. This definition extends to random fields on a lattice and point processes as follows. The α-mixing coefficient of a random field $\{Z(\mathbf{l})\}_{\mathbf{l} \in \mathbb{L}}$ on a lattice $\mathbb{L}$ and a point process $\mathbf{X}$ are given for $m, c_1, c_2 \geq 0$ by

$$\begin{aligned}\alpha^Z_{c_1,c_2}(m) = \sup\{&\alpha(\sigma((Z(\mathbf{l}) : \mathbf{l} \in I_1)), \sigma((Z(\mathbf{k}) : \mathbf{k} \in I_2))) : \\ &I_1 \subset \mathbb{L}, I_2 \subset \mathbb{L}, |I_1| \leq c_1, |I_2| \leq c_2, \mathrm{d}(I_1, I_2) \geq m\}\end{aligned}$$

and

$$\alpha^{\mathbf{X}}_{c_1,c_2}(m) = \sup\{\alpha(\sigma(\mathbf{X}\cap E_1), \sigma(\mathbf{X}\cap E_2)) : \\ E_1 \subset \mathbb{R}^d, E_2 \subset \mathbb{R}^d, |E_1| \leq c_1, |E_2| \leq c_2, \mathrm{d}(E_1, E_2) \geq m\}.$$

Note that the definition of $\alpha^{\mathbf{X}}_{c_1,c_2}$ differs from the usual definition in spatial statistics, see e.g. Waagepetersen and Guan (2009), by the use of $\mathrm{d}(\cdot,\cdot)$ in place of the Euclidean norm. This choice has been made to ease the proofs and makes no substantial difference since all the norms in $\mathbb{R}^d$ are equivalent. For a matrix M we use the Frobenius norm $|M| = (\sum_{i,j} M^2_{i,j})^{1/2}$.

3 Central limit theorem based on Bolthausen's approach

We consider a sequence of statistics $T_{W_n}(\mathbf{X})$ where $\{W_n\}_{n\in\mathbb{N}}$ is a sequence of increasing compact observation windows that verify

($\mathcal{H}1$) $W_1 \subset W_2 \subset \dots$ and $|\bigcup_{l=1}^{\infty} W_l| = \infty$.

Note that we do not assume that each W_i is convex and that $\cup_{i=1}^{\infty} W_i = \mathbb{R}^d$ as it is usually the case in spatial statistics, see e.g. Waagepetersen and Guan (2009) or Biscio and Lavancier (2017). We assume that $T_{W_n}(\mathbf{X})$ can be additively decomposed as

$$T_{W_n}(\mathbf{X}) = \sum_{\mathbf{l}\in\mathcal{D}_n(W_n)} f_{n,\mathbf{l},W_n}(\mathbf{X}) \tag{3.1}$$

where for $n, q \in \mathbb{N}$, $\mathcal{D}_n$ is a finite index set defined below, and $f_{n,\mathbf{l},W_n}$ is a function on the sample space of $\mathbf{X}$ to $\mathbb{R}^q$. We assume that $f_{n,\mathbf{l},W_n}(\mathbf{X})$ depends on $\mathbf{X}$ only through $\mathbf{X}\cap W_n \cap C^{\oplus R}_n(\mathbf{l})$ for some $R \geq 0$, where $C_n(\mathbf{l})$ is a hyper cube of side length $s_n > 0$,

$$C_n(\mathbf{l}) = \prod_{j=1}^{d} (l_j - s_n/2, l_j + s_n/2], \qquad \mathbf{l} \in s_n\mathbb{Z}^d, \tag{3.2}$$

and $C^{\oplus R}_n(\mathbf{l}) = C_n(\mathbf{l}) \oplus R$. Thus the $C_n(\mathbf{l})$, $\mathbf{l} \in s_n\mathbb{Z}^d$, form a disjoint partition of $\mathbb{R}^d$. We denote by $v_n = |C^{\oplus R}_n(\mathbf{l})|$ the common volume of the $C^{\oplus R}_n(\mathbf{l})$ and $\mathcal{D}_n(A)$ is defined for any $A \subset \mathbb{R}^d$ by

$$\mathcal{D}_n(A) = \{\mathbf{l} \in s_n\mathbb{Z}^d : C_n(\mathbf{l}) \cap A \neq \emptyset\}. \tag{3.3}$$

For brevity, we write $\mathcal{D}_n$ in place of $\mathcal{D}_n(W_n)$. Then W_n is the disjoint union of $C_n(\mathbf{l}) \cap W_n$, $\mathbf{l} \in \mathcal{D}_n$.

For $n \in \mathbb{N}$ and $\mathbf{l} \in \mathbb{Z}^d$ let for ease of notation $Z_n(\mathbf{l}) = f_{n,\mathbf{l},W_n}(\mathbf{X})$ and consider the following assumptions.

($\mathcal{H}2$) There exists $0 \leq \eta < 1$ such that $s_n = |W_n|^{\eta/d}$, and if $\eta > 0$, $|\mathcal{D}_n| = O(|W_n|/s^d_n)$. Further, there exists $\epsilon > 0$ such that $\sup_{n\in\mathbb{N}} \alpha^{\mathbf{X}}_{2v_n,\infty}(s) = O(1/s^{d+\epsilon})$.

($\mathcal{H}3$) There exists $\tau > 2d/\epsilon$ such that $\sup_{n\in\mathbb{N}} \sup_{\mathbf{l}\in\mathcal{D}_n} \mathbb{E}|Z_n(\mathbf{l}) - \mathbb{E}Z_n(\mathbf{l})|^{2+\tau} < \infty$.

($\mathcal{H}4$) We have $0 < \liminf_{n\to\infty} \lambda_{\min}\big(\frac{\Sigma_n}{|\mathcal{D}_n|}\big)$, where $\Sigma_n = \operatorname{Var} T_{W_n}(\mathbf{X})$ and $\lambda_{\min}(M)$ denotes the smallest eigen value of a symmetrix matrix M.

We then obtain the following theorem.

Theorem 3.1. *Let $\{T_{W_n}(\mathbf{X})\}_{n\in\mathbb{N}}$ be a sequence of q-dimensional statistics of the form* (3.1)*. If ($\mathcal{H}1$)–($\mathcal{H}4$) hold, then we have the convergence*

$$\Sigma_n^{-\frac{1}{2}}\big(T_{W_n}(\mathbf{X}) - \mathbb{E}T_{W_n}(\mathbf{X})\big) \xrightarrow[n\to\infty]{distr.} \mathcal{N}(0, I_q)$$

where $\Sigma_n = \operatorname{Var} T_{W_n}(\mathbf{X})$, and I_q is the identity matrix.

Remark 3.2. The existence of $\Sigma_n^{-\frac{1}{2}}$ for n large enough is ensured by ($\mathcal{H}4$).

Remark 3.3. In many applications we can simply take $\eta = 0$ so that $s_n = 1$. In that case, we do not require further assumptions on $\mathcal{D}_n$. However, in applications dealing with kernel estimators depending on a bandwidth h_n tending towards 0, we may have $\operatorname{Var} T_{W_n}(\mathbf{X})$ of the order $|W_n|h_n^d$ (e.g. Heinrich and Klein, 2014). Then, ($\mathcal{H}4$) can be fulfilled if $s_n = 1/h_n$ and $\eta > 0$ so that by ($\mathcal{H}2$), $|\mathcal{D}_n|$ is also of the order $|W_n|/s_n^d = |W_n|h_n^d$.

Remark 3.4. For a point process, moments are calculated using so-called joint intensity functions. To verify ($\mathcal{H}3$) it often suffices to assume boundedness of the joint intensities up to order $2(2 + \lceil\tau\rceil)$.

Remark 3.5. As presented in Section S1, the convergence in Theorem 3.1 can be proved under different assumptions using Bernstein's blocking technique. However, as explained in Section S2, assumptions on the observation windows and on the asymptotic variance of $T_{W_n}(\mathbf{X})$ are more restrictive when working with Bernstein's blocking technique than with Bolthausen's approach.

4 Subsampling variance estimator

By Theorem 3.1, for $\alpha \in (0,1)$, we may establish an asymptotic $1-\alpha$ confidence ellipsoid for $\mathbb{E}T_{W_n}(\mathbf{X})$ using the $1-\alpha$ quantile $q_{1-\alpha}$ of the $\chi^2(q)$ distribution, i.e.

$$\mathrm{P}\big(\mathcal{E}(\mathbf{X}) \le q_{1-\alpha}\big) \xrightarrow[n\to\infty]{} 1-\alpha \tag{4.1}$$

where

$$\mathcal{E}(\mathbf{X}) = |D_n|^{-1}\big(T_{W_n}(\mathbf{X}) - \mathbb{E}T_{W_n}(\mathbf{X})\big)^T \left(\frac{\Sigma_n}{|\mathcal{D}_n|}\right)^{-1} \big(T_{W_n}(\mathbf{X}) - \mathbb{E}T_{W_n}(\mathbf{X})\big).$$

The matrix Σ_n is usually not known in practice. Thus we suggest to replace $\Sigma_n/|\mathcal{D}_n|$ by a subsampling estimate, adapting results from Sherman (1996) and Ekström (2008) to establish the consistency of the subsampling estimator.

The setting and notation are as in Section 3 except that we only consider rectangular windows W_n so that ($\mathcal{H}1$) is replaced with the following assumption.

($\mathcal{S}0$) We let $\{\mathbf{m}_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathbb{N}^d$ such that the rectangles defined by $W_n = \prod_{j=1}^d (-m_{n,j}/2, m_{n,j}/2)$ verifies $W_1 \subset W_2 \subset \cdots$ and $|\bigcup_{n=1}^\infty W_n| = \infty$.

Let $\{\mathbf{k}_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathbb{N}^d$, consider for $\mathbf{t} \in \mathbb{Z}^d$ the (overlapping) sub-rectangles

$$B_{\mathbf{k}_n,\mathbf{t}} = \prod_{j=1}^d (t_j - k_{n,j}/2, t_j + k_{n,j}/2), \tag{4.2}$$

and define $\mathcal{T}_{\mathbf{k}_n,n} = \{\mathbf{t} \in \mathbb{Z}^d : B_{\mathbf{k}_n,\mathbf{t}} \subset W_n\}$. We want to estimate

$$\varsigma_n = \frac{\mathrm{Var}(T_{W_n}(\mathbf{X}))}{|\mathcal{D}_n|} = \frac{\Sigma_n}{|\mathcal{D}_n|}$$

where $T_{W_n}(\mathbf{X})$ is as in (3.1). We suggest the subsampling estimator

$$\hat{\varsigma}_n = \frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|} \sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \left(\frac{T_{B_{\mathbf{k}_n,\mathbf{t}}}(\mathbf{X})}{\sqrt{|\mathcal{D}_n(B_{\mathbf{k}_n,\mathbf{t}})|}} - \frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|} \sum_{\mathbf{s}\in\mathcal{T}_{\mathbf{k}_n,n}} \frac{T_{B_{\mathbf{k}_n,\mathbf{s}}}(\mathbf{X})}{\sqrt{|\mathcal{D}_n(B_{\mathbf{k}_n,\mathbf{s}})|}} \right)^2. \tag{4.3}$$

To establish consistency of $\hat{\varsigma}_n$ we consider the following assumptions.

($\mathcal{S}1$) For $j = 1, \ldots, d$, $k_{n,j} < m_{n,j}$. There is at least one j such that $m_{n,j}$ goes to infinity. If $m_{n,j} \to \infty$ as $n \to \infty$, so does $k_{n,j}$ and $k_{n,j}/m_{n,j} \to 0$ as $n \to \infty$. If $m_{n,j}$ converges to a constant, then $k_{n,j}$ converges to a constant less than or equal to the previous constant. Moreover, $(\max_i k_{n,i}^d)/\prod_{i=1}^d (m_{n,i} - k_{n,i})$ converges towards 0 as n tends to infinity.

($\mathcal{S}2$) For some $\epsilon' > 0$, $\sup_{n\in\mathbb{N}} \sup_{\mathbf{l}\in\mathcal{T}_{\mathbf{k}_n,n}} \mathbb{E}(|Z_n(\mathbf{l}) - \mathbb{E}(Z_n(\mathbf{l}))|^{4+\epsilon'}) < +\infty$.

($\mathcal{S}3$) We have $|\mathcal{D}_n|^{-1}\Sigma_n - |\mathcal{T}_{\mathbf{k}_n,n}|^{-1} \sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} |\mathcal{D}_n(B_{\mathbf{k}_n,\mathbf{t}})|^{-1} \mathrm{Var}(T_{B_{\mathbf{k}_n,\mathbf{t}}}(\mathbf{X})) \to 0$ as $n \to \infty$ and $\limsup_{n\to\infty} \lambda_{\max}(\Sigma_n) < \infty$ where $\lambda_{\max}(M)$ denotes the maximal eigen value of a symmetric matrix M.

($\mathcal{S}4$) $|\mathcal{T}_{\mathbf{k}_n,n}|^{-1} \sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \{\mathbb{E}(T_{B_{\mathbf{k}_n,\mathbf{t}}}(\mathbf{X})) - \mathbb{E}(|\mathcal{T}_{\mathbf{k}_n,n}|^{-1} \sum_{\mathbf{s}\in\mathcal{T}_{\mathbf{k}_n,n}} T_{B_{\mathbf{k}_n,\mathbf{s}}}(\mathbf{X}))\}^2 \to 0$ as $n \to \infty$.

($\mathcal{S}5$) There exists $c > 0$ and $\delta > 0$, such that $\sup_{p\in\mathbb{N}} \alpha^{\mathbf{X}}_{p,p}(m)/p \le c/m^{d+\delta}$ and, for v_n as below (3.2), $v_n \prod_{j=1}^d (2k_{n,j}+1)/(\max_i k_{n,i} - s_n)^{d+\delta}$ converges towards 0 as n tends to infinity.

($\mathcal{S}6$) There exists $c, \delta' > 0$ and $\epsilon' > \epsilon > 0$ such that $\alpha^{\mathbf{X}}_{5v_n,5v_n}(r) \le c\, r^{-5d\frac{6+\epsilon}{\epsilon} - \delta'}$.

Theorem 4.1. *Let $\{T_{W_n}(\mathbf{X})\}_{n\in\mathbb{N}}$ be a sequence of q-dimensional statistics of the form* (3.1)*. Let further $\hat{\varsigma}_n$ be defined as in* (4.3) *and assume that ($\mathcal{S}0$)–($\mathcal{S}6$) hold. Then we have the convergence*

$$\mathbb{E}\left(\left| \hat{\varsigma}_n - \frac{\Sigma_n}{|\mathcal{D}_n|} \right|^2 \right) \xrightarrow[n\to\infty]{} 0.$$

For practical application, it is enough to state Theorem 4.1 with convergence in probability but the proof is easier when considering mean square convergence. Assumption ($\mathcal{S}1$) ensures that the sub-rectangles are large enough to mimic the behaviour of the point process on W_n while at the same time their number grows to infinity. Assumption ($\mathcal{S}2$) looks stronger than ($\mathcal{H}3$) for Theorem 3.1. However, in ($\mathcal{H}3$) note that τ depends on the mixing properties of the process controlled by ($\mathcal{H}2$). Thus, depending on the mixing properties, ($\mathcal{S}2$) is not much stronger than ($\mathcal{H}3$). Assumption ($\mathcal{S}3$) should hold for any process that is not too exotic and ensures that the variance of the studied statistic on each sub-rectangle is not too different from Σ_n. In particular, it holds naturally if there exists a matrix $\widetilde{\Sigma}$ such that $\lim_{n\to\infty}|\mathcal{D}_n(A)|^{-1}\operatorname{Var}(T_A(\mathbf{X})) = \widetilde{\Sigma}$, $A = W_n$ or $A = B_{\mathbf{k}_n,\mathbf{t}}$, with $\mathbf{t} \in \mathcal{T}_{\mathbf{k}_n,n}$. The condition ($\mathcal{S}4$) is needed to control that the expectations over sub-rectangles $B_{\mathbf{k}_n,\mathbf{t}}$ do not vary too much. For instance, this assumption is automatically verified if the point process $\mathbf{X}$ is stationary or if the statistics (1.2) are centred so that $\mathbb{E}T_{W_n}(\mathbf{X}) = \mathbb{E}T_{B_{\mathbf{k}_n,\mathbf{t}}}(\mathbf{X}) = 0$, for $\mathbf{t} \in \mathcal{T}_{\mathbf{k}_n,n}$, see Section 5. Moreover, depending on the statistic (1.2), this assumption may also be verified if $\mathbf{X}$ is second-order intensity reweighted stationary as assumed for the bootstrap method developed by Loh (2010). Note that ($\mathcal{S}5$) includes a condition on the size of the $B_{\mathbf{k}_n,\mathbf{t}}$ that holds trivially if $s_n = 1$, which is usually the case if we do not consider non-parametric kernel estimators. We use two different α-mixing conditions ($\mathcal{S}5$)–($\mathcal{S}6$) to apply Theorem 4.1. Moreover, the decreasing rate in ($\mathcal{S}6$) is restrictive due to the constant $5d$. Hence, mixing conditions are stronger than for Theorem 3.1. However, in the proof of Theorem 4.1, assumption ($\mathcal{S}6$) is used only to verify (C.4) which ensures the validity of the assumption (i) of Theorem C.1. Depending on the problem considered, (C.4) may be verified without additional constraints on the α-mixing coefficient. For example, if we are in the setting of Biscio and Lavancier (2016, Section 4.1) where in particular $\mathbf{X}$ is a stationary determinantal point process, then (C.4) is an immediate consequence of Biscio and Lavancier (2016, Proposition 4.2).

Remark 4.2. By Theorems 3.1 and 4.1, we may replace the confidence ellipsoid in (4.1) by a subsampling confidence ellipsoid $\hat{\mathcal{E}}_n$, i.e.

$$\mathrm{P}\big(\hat{\mathcal{E}}_n(\mathbf{X}) \le q_{1-\alpha}\big) \xrightarrow[n\to\infty]{} 1-\alpha \tag{4.4}$$

where

$$\hat{\mathcal{E}}_n(\mathbf{X}) = |D_n|^{-1}\big(T_{W_n}(\mathbf{X}) - \mathbb{E}T_{W_n}(\mathbf{X})\big)^T \hat{\varsigma}_n^{-1}\big(T_{W_n}(\mathbf{X}) - \mathbb{E}T_{W_n}(\mathbf{X})\big).$$

Remark 4.3. Although the size of the $B_{\mathbf{k}_n,\mathbf{t}}$ is controlled by assumptions ($\mathcal{S}1$) and ($\mathcal{S}5$), we have not addressed the issue of finding their optimal size, i.e. the one ensuring the fastest convergence rate in Theorem 4.1. Concerning that problem, there are several recommendations in the literature, see for instance Lahiri (2003).

Remark 4.4. Note that the centers of the $B_{\mathbf{k}_n,\mathbf{t}}$ are chosen to be a subset of $\mathbb{Z}^d$ but any other fixed lattice could be used as well. Further, similarly to Loh (2010) and Ekström (2008), it is possible to extend Theorem 4.1 by relaxing the assumption in ($\mathcal{S}0$) that the windows are rectangular.

5 Variance estimation for estimating functions

Consider a parametric family of point processes $\{\mathbf{X}_\theta : \theta \in \Theta\}$ for a non-empty subset $\Theta \subseteq \mathbb{R}^q$, $q \in \mathbb{N}$. We further assume that we observe a realisation of $\mathbf{X}_{\theta_0}$ for a $\theta_0 \in \Theta$.

To estimate θ it is common to use estimating functions of the form

$$e_n(\theta) = \sum_{u_1,\ldots,u_p \in \mathbf{X}_{\theta_0} \cap W_n}^{\neq} h_\theta(u_1,\ldots,u_p) - \int_{W_n^p} h_\theta(u_1,\ldots,u_p)\rho_\theta^{(p)}(u_1,\ldots,u_p)\mathrm{d}u_1 \ldots \mathrm{d}u_p, \tag{5.1}$$

where h_θ is a function from $\mathbb{R}^{dp}$ into $\mathbb{R}^q$, and $\rho_\theta^{(p)}$ denotes the p-th order joint intensities of $\mathbf{X}_\theta$. Then an estimate $\widehat{\theta}_n$ of θ_0 is obtained by solving $e_n(\theta) = 0$. The case $p = 1$ is relevant if interest is focused on estimation of the intensity function $\lambda_\theta(u) = \rho_\theta^{(1)}$. Several papers have discussed choices of h and studied asymptotic properties of $\widehat{\theta}_n$ for the case $p = 1$, see for instance Waagepetersen (2007), Guan and Shen (2010), and Guan et al. (2015). A popular and simple choice is $h_\theta(u) = \nabla_\theta \lambda_\theta(u)/\lambda_\theta(u)$ where ∇_θ denotes the gradient with respect to θ. In this case e_n can be viewed as the score of a composite likelihood.

In the references aforementioned, the asymptotic results are of the form

$$|W_n|^{1/2}\Big(S_n^{-1}(\theta_0)\frac{\Sigma_{n,\theta_0}}{|W_n|}S_n^{-1}(\theta_0)\Big)^{-1/2}(\widehat{\theta}_n - \theta_0) \xrightarrow[n\to\infty]{distr.} \mathcal{N}(0, I_q), \tag{5.2}$$

where $S_n(\theta_0) = |W_n|^{-1}\mathbb{E}(-\mathrm{d}e_n(\theta_0)/\mathrm{d}\theta^T)$ and $\Sigma_{n,\theta_0} = \operatorname{Var} e_n(\theta_0)$. The matrix Σ_{n,θ_0} is crucial but usually unknown. To estimate Σ_{n,θ_0}, a bootstrap method was proposed in Guan and Loh (2007) under several mild mixing and moment conditions. However, their method has been established only for second-order intensity reweighted stationary point processes on $\mathbb{R}^2$ when $p = 1$ and for a specific function h. Using the theory established in Section 4, we propose a subsampling estimator of Σ_{n,θ_0} that may be used in a more general setting but under slightly stronger mixing conditions. Following the notation in Sections 3 and 4, for $\theta \in \Theta$, we let $T_{W_n,\theta}(\mathbf{X}_{\theta_0}) = e_n(\theta)$ and

$$\begin{aligned} Z_{n,\theta}(\mathbf{l}) = & \sum_{u_1 \in \mathbf{X}_{\theta_0} \cap C(\mathbf{l}) \cap W_n} \ \sum_{u_2,\ldots,u_p \in (\mathbf{X}_{\theta_0} \cap W_n)\setminus\{u_1\}}^{\neq} h_\theta(u_1,\ldots,u_p) \\ & - \int_{C(\mathbf{l})\cap W_n} \int_{W_n^{p-1}} h_\theta(u_1,\ldots,u_p)\rho_\theta^{(p)}(u_1,\ldots,u_p)\mathrm{d}u_1 \ldots \mathrm{d}u_p. \end{aligned}$$

Further $\hat{\varsigma}_n(\theta)$ is defined as in (4.3) but now stressing the dependence on θ. In practice, if $|W_n|/|\mathcal{D}_n| \to 1$, we estimate $\Sigma_{n,\theta_0}/|W_n|$ by $\hat{\varsigma}_n(\widehat{\theta}_n)$. The validity of this relies in a standard way on a Taylor expansion

$$\hat{\varsigma}_n(\widehat{\theta}_n) = \hat{\varsigma}_n(\theta_0) + \frac{\mathrm{d}}{\mathrm{d}\theta}\hat{\varsigma}_n(\theta^*)(\widehat{\theta}_n - \theta_0)$$

where $\|\theta^* - \theta_0\| \le \|\widehat{\theta}_n - \theta_0\|$ and one needs to check that $\mathrm{d}\hat{\varsigma}_n(\theta^*)/\mathrm{d}\theta$ is bounded in probability. We illustrate with our simulation study in the next section, the applicability of $\hat{\varsigma}_n(\widehat{\theta}_n)$ to estimate $\Sigma_{n,\theta_0}/|W_n|$.

6 Simulation study

To assess the performance of our subsampling estimator, we estimate by simulation the coverage achieved by asymptotic 95% confidence intervals when considering intensity estimation by composite likelihood as discussed in the previous section. The confidence intervals are obtained in the standard way using the asymptotic normality (5.2) and replacing $\Sigma_{n,\theta_0}/|W_n|$ by our subsampling estimator.

When computing $\hat{\varsigma}_n$, the user must specify the shape, the size, and the possible overlapping of the sub-rectangles (blocks) used for the subsampling estimator. For simplicity we assume that $W_n = [0,n]^2 \subset \mathbb{R}^2$ and use square blocks. We denote by b_l the sidelength of the blocks and by κ the maximal proportion of overlap possible between two blocks. The block centres are located on a grid $(W_n \cap h_{n,\kappa}\mathbb{Z}^2) + h_{n,\kappa}(1/2,1/2)$ where $h_{n,\kappa}$ is chosen such that κ is the ratio between the area of the overlap of two contiguous blocks located at $h_{n,\kappa}(1/2,1/2)$ and $h_{n,\kappa}(1/2,1/2+1)$, and the area of one block. For instance, for $W_1 = [0,1]^2$, $b_l = 0.5$ and $\kappa = 0.5$, the centers of the blocks completely included in W_1 are $(0.25,0.25); (0.5,0.25); (0.75,0.25); (0.25,0.5), \dots$ and so on until $(0.75,0.75)$. The simulations have been done for every possible combination between $n = 1,2,3$, $b_l = 0.2, 0.5$, $\kappa = 0, 0.5, 0.75, 0.875$, and the four following point process models: a non-stationary Poisson point process, two different non-stationary log-Gaussian Cox processes (LGCPs), and a non-stationary determinantal point process (DPP). For a presentation of these models, we refer to Baddeley et al. (2015).

For each point process simulation, the intensity is driven by a realisation $\mathcal{Z}_1$ of a zero mean Gaussian random field with exponential covariance function (scale parameter 0.5 and variance 0.1). As specified below, we have for each realisation of $\mathcal{Z}_1$ chosen the parameters so that the average number of points on $|W_n|$ is $100|W_n|$ (100, 400 or 900).

For the non-stationary Poisson point processes we use the intensity function

$$\lambda_n(x) = \exp\big(\theta_{0,n} + \mathcal{Z}_1(x)\big) \tag{6.1}$$

where $\theta_{0,n} = \log(100|W_n|) - \log \int_{W_n} \exp(\mathcal{Z}_1(x))\mathrm{d}x$. For the two LGCPs, the random intensity functions are of the form

$$\Lambda_n(x) = \exp\big(\theta_{0,n} + \mathcal{Z}_1(x) + \mathcal{Z}_2(x)\big) \tag{6.2}$$

where $\theta_{0,n} = \log(100|W_n|) - \mathrm{Var}(\mathcal{Z}_2(0))/2 - \log \int_{W_n} \exp(\mathcal{Z}_1(x))\mathrm{d}x$, and where $\mathcal{Z}_2$ is a zero mean Gaussian random field independent of $\mathcal{Z}_1$ and also with exponential covariance function (scale parameter 0.05, and variance 0.25 for one LGCP and 1 for the other). The non-stationary DPP has been simulated by: First simulating a stationary DPP using the Gaussian kernel

$$C_n(x,y) = \lambda_{n,\mathrm{dom}} \exp\left(-\frac{|x-y|^2}{\beta}\right),$$

where $\beta \simeq 0.04$ and $\lambda_{n,\mathrm{dom}} = 100|W_n| / \int_{W_n} \exp(\mathcal{Z}_1(x) - \max_{x\in W_n} \mathcal{Z}_1(x))\mathrm{d}x$; Second, applying an independent thinning with probability $\lambda'_n(x) = \exp\big(\mathcal{Z}_1(x) - \max_{x\in W_n} \mathcal{Z}_1(x)\big)$, $x \in W_n$, of retaining af point. Specifically, β actually equals $1/\sqrt{\pi\lambda_{n,\mathrm{dom}}}$ and corresponds to the most repulsive Gaussian DPP according to Lavancier et al. (2014). Following Appendix A in Lavancier et al. (2014), the result is then a realisation of a non-stationary DPP with kernel

$$C'_n(x,y) = \sqrt{\lambda'_n(x)\lambda'_n(y)}\lambda_{n,\mathrm{dom}} \exp\left(-\frac{|x-y|^2}{\beta}\right). \tag{6.3}$$

The intensity of the DPP is given by $\lambda_n(x) = C'_n(x,x) = \lambda_{n,\mathrm{dom}}\lambda'_n(x)$. Realisations of the DPP and each LGCP are plotted in Figure 1 along with the corresponding pair correlation functions defined by $g(r) = \lambda^{(2)}(u,v)/\lambda(u)\lambda(v)$ where $r = |u-v|$, and λ, $\lambda^{(2)}$ denote the intensity and second order product density, respectively. Note that in the DPP case, the pair correlation function depends on the realisation of $\mathcal{Z}_1$ via β. In Figure 1 the DPP pair correlation function is plotted with $\beta = 0.04$.

For each of the models, the intensity function is of log-linear form $\lambda_\theta(x) = \exp(\theta_0+\theta_1\mathcal{Z}_1(x))$, $x \in \mathbb{R}^2$ where $\mathcal{Z}_1$ is considered as a known covariate. The parameter $\theta = (\theta_0, \theta_1)$ is estimated using composite likelihood which is implemented in the **R**-package **spatstat** (Baddeley et al., 2015) procedure `ppm`. The true value of θ_1 is one while the true value of θ_0 depends on the window W_n and the realization of $\mathcal{Z}_1$. For each combination of n, b_l, κ and point process type we apply the estimation procedure to 5000 simulations and compute the estimated coverage of the confidence interval for the parameter θ_1. The results are plotted in Figure 2. The Monte Carlo standard error for the estimated coverages is approximately 0.003. For each combination of b_l and n, the corresponding plot shows the estimated coverage for combinations of $\kappa = 0, 0.5, 0.75, 0.875$ and the four point process models. To aid the visual interpretation, points are connected by line segments.

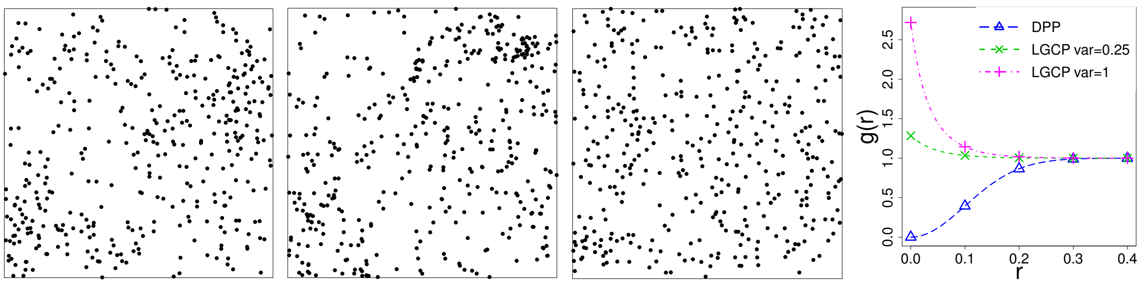

Figure 1: From left to right, the first three panels show a realisation on $[0,2]^2$ of a LGCP with variance parameter 0.25, a LGCP with variance parameter 1, and a DPP. Last panel: a plot of the corresponding theoretical pair correlation functions.

Except for the lower left plot, the results seem rather insensitive to the choice of κ (in the lower left plot, $b_l = 0.5$ seems to be too large relative to the window W_1). From a computational point of view $\kappa = 0$ is advantageous and is never outperformed in terms of coverage by other choices of κ. The results are more sensitive to the choice of b_l. For the LGCPs we see the anticipated convergence of the estimated coverages to 95% when $b_l = 0.5$ and n is increased but not when $b_l = 0.2$. This suggests that $b_l = 0.2$ is too small for the statistics on blocks to represent the statistic on the windows $W_1 - W_3$ in case of the LGCPs. Among the LGCPs, the coverages are closer to 95% for the LGCP with the lowest variance. For the Poisson process and the DPP, the estimated coverages are very close to 95% both for $b_l = 0.2$ and $b_l = 0.5$ except for the small window W_1. The general impression from the simulation study is that the subsampling method works well when the point patterns are of reasonable size (hundreds of points), and the blocks are of appropriate size relative to the observation window.

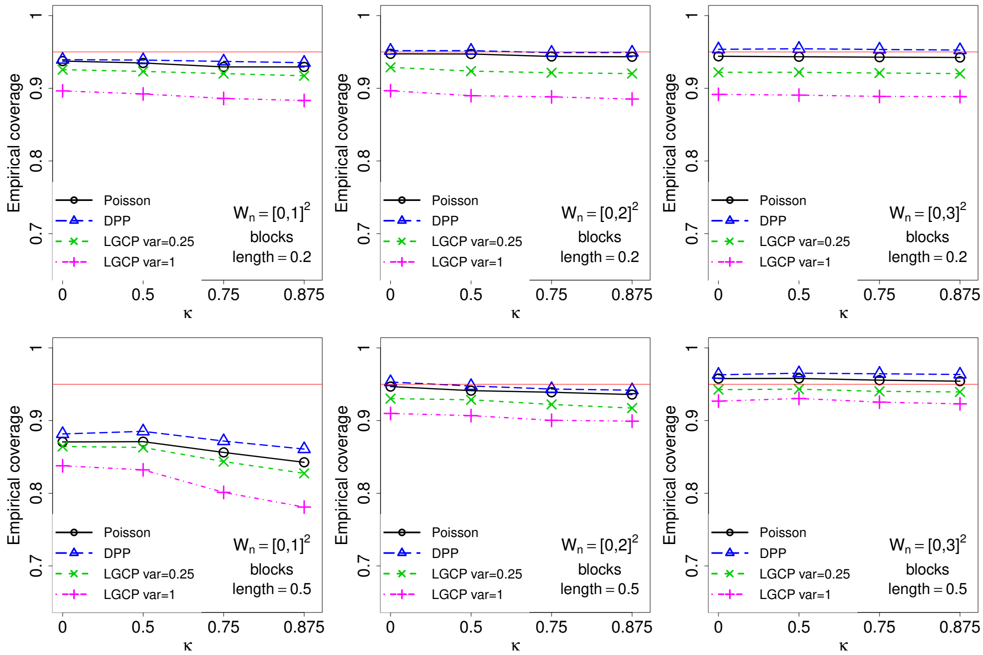

Figure 2: Estimated coverages of the confidence intervals for θ_1 when using the subsampling estimator (4.3). Upper row to lower row: $b_l = 0.2, 0.5$. Left column to right column: $n = 1, 2, 3$. In each plot the estimated coverage is computed for four point process models: non-stationary Poisson point process, DPP, and two LGCPs; and $\kappa = 0, 0.5, 0.75, 0.875$. The lines joining the points just serve to aid visual interpretation. The straight horizontal red line indicates the value 0.95.

7 Discussion

Our simulations have shown that the subsampling estimator may be used to obtain confidence intervals in the framework of intensity estimation by composite likelihood. The results obtained were satisfying with estimated coverages close to the nominal level 95% except for small point patterns and provided a suitable block size was used.

These results may be compared with the estimated coverage obtained when using the variance estimate provided by the function `vcov.kppm` of the **R**-package **spatstat**. This function computes an estimate of the asymptotic variance of the composite likelihood estimators by plugging in a parametric estimate of the pair correlation function into the theoretical expression for the covariance matrix following Waagepetersen (2007). Using `vcov.kppm` for the simulated realisations of LGCPs from Section 6, the estimated coverages of the resulting approximate confidence intervals for the parameter θ_1 ranges from 93% to 96% (including results for $n = 1$ and cases with a misspecified parametric model for the pair correlation function). Thus, the results are closer to the nominal level of the confidence interval than for the subsampling estimator. On the other hand, the subsampling estimator is much more flexible as it is model free and may be applied to any statistic of the form (3.1), in any dimension.

We have also compared our subsampling estimator with the thinned block bootstrap estimator proposed in Guan and Loh (2007) and got very similar results within the simulation study settings of that paper. This is to be expected given the similarities of the methods. However, the method in Guan and Loh (2007) requires that it is possible to thin the point process into a second-order stationary point process.

Acknowledgements

Christophe A.N. Biscio and Rasmus Waagepetersen are supported by The Danish Council for Independent Research | Natural Sciences, grant DFF – 7014-00074 "Statistics for point processes in space and beyond", and by the "Centre for Stochastic Geometry and Advanced Bioimaging", funded by grant 8721 from the Villum Foundation.

References

Adrian Baddeley, Ege Rubak, and Rolf Turner. *Spatial Point Patterns: Methodology and Applications with R.* Chapman and Hall/CRC Press, 2015. URL `https://www.crcpress.com/Spatial-Point-Patterns-Methodology-and-Applications-with-R/Baddeley-Rubak-Turner/p/book/9781482210200/`.

Patrick Billingsley. *Probability and measure.* John Wiley & Sons, New York-Chichester-Brisbane, 1995. Wiley Series in Probability and Mathematical Statistics.

Christophe Ange Napoléon Biscio and Jean-François Coeurjolly. Standard and robust intensity parameter estimation for stationary determinantal point processes. *Spatial Statistics*, 18:24–39, 2016.

Christophe Ange Napoléon Biscio and Frédéric Lavancier. Brillinger mixing of determinantal point processes and statistical applications. *Electronic Journal of Statistics*, pages 582–607, 2016.

Christophe Ange Napoléon Biscio and Frédéric Lavancier. Contrast estimation for parametric stationary determinantal point processes. *Scandinavian Journal of Statistics*, 44:204–229, 2017. 10.1111/sjos.12249.

Christophe Ange Napoléon Biscio, Arnaud Poinas, and Rasmus Waagepetersen. A note on gaps in proofs of central limit theorems. *Statistics and Probability Letters*, 2017.

Erwin Bolthausen. On the central limit theorem for stationary mixing random fields. *The Annals of Probability*, 10(4):1047–1050, 1982.

Kai Lai Chung. *A Course in Probability Theory.* Academic press, 2001.

Jean-François Coeurjolly. Median-based estimation of the intensity of a spatial point process. *Annals of the Institute of Statistical Mathematics*, 69:303–331, 2017.

Jean-François Coeurjolly and Frédéric Lavancier. Parametric estimation of pairwise Gibbs point processes with infinite range interaction. *Bernoulli*, 23:1299–1334, 2017.

Jean-François Coeurjolly and Jesper Møller. Variational approach for spatial point process intensity estimation. *Bernoulli*, 20(3):1097–1125, 08 2014. doi: 10.3150/13-BEJ516. URL `http://dx.doi.org/10.3150/13-BEJ516`.

Daryl J. Daley and David Vere-Jones. *An Introduction to the Theory of Point Processes. Volume I, Elementary Theory and Methods*. Springer-Verlag, New York, second edition, 2003.

Daryl J. Daley and David Vere-Jones. *An Introduction to the Theory of Point Processes, Vol. II, General Theory and Structure*. Springer-Verlag, New York, second edition, 2008.

Paul Doukhan. *Mixing, Properties and Examples*, volume 85 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1994.

Magnus Ekström. Subsampling variance estimation for non-stationary spatial lattice data. *Scandinavian Journal of Statistics*, 35:38–63, 2008. ISSN 1467-9469. doi: 10.1111/j.1467-9469.2007.00572.x. URL `http://dx.doi.org/10.1111/j.1467-9469.2007.00572.x`.

István Fazekas, Alexander Kukush, and Tibor Tómács. On the Rosenthal inequality for mixing fields. *Ukrainian Mathematical Journal*, 52(2):305–318, 2000.

Yongtao Guan and Ji Meng Loh. A thinned block bootstrap variance estimation procedure for inhomogeneous spatial point patterns. *Journal of the American Statistical Association*, 102(480):1377–1386, 2007. doi: 10.1198/016214507000000879. URL `http://dx.doi.org/10.1198/016214507000000879`.

Yongtao Guan and Ye Shen. A weighted estimating equation approach for inhomogeneous spatial point processes. *Biometrika*, 97:867–880, 2010. doi: 10.1093/biomet/asq043.

Yongtao Guan and Michael Sherman. On least squares fitting for stationary spatial point processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(1):31–49, 2007. ISSN 1467-9868. doi: 10.1111/j.1467-9868.2007.00575.x. URL `http://dx.doi.org/10.1111/j.1467-9868.2007.00575.x`.

Yongtao Guan, Abdollah Jalilian, and Rasmus Waagepetersen. Quasi-likelihood for spatial point processes. *Journal of the Royal Statistical Society, Series B*, 77:677–697, 2015.

Xavier Guyon. *Random fields on a network*. Probability and its Applications. Springer, New York, 1995.

Lothar Heinrich. Minimum contrast estimates for parameters of spatial ergodic point processes. In *Transactions of the 11th Prague Conference on Random Processes, Information Theory and Statistical Decision Functions*, volume 479, page 492, 1992.

Lothar Heinrich and Stella Klein. Central limit theorems for empirical product densities of stationary point processes. *Statistical Inference for Stochastic Processes*, 17(2): 121–138, Jul 2014.

Ildar Abdulovich Ibragimov and Yuri Vladimirovich Linnik. *Independent and stationary sequences of random variables*. Wolters-Noordhoff, 1971.

Abdollah Jalilian, Yongtao Guan, and Rasmus Waagepetersen. Orthogonal series estimation of the pair correlation function for spatial point processes. *Statistica Sinica*, 2017. accepted.

Jens Ledet Jensen and Hans R. Künsch. On asymptotic normality of pseudo likelihood estimates for pairwise interaction processes. *Annals of the Institute of Statistical Mathematics*, 46:475–486, 1994. ISSN 1572-9052. doi: 10.1007/BF00773511. URL http://dx.doi.org/10.1007/BF00773511.

Zsolt Karácsony. A central limit theorem for mixing random fields. *Miskolc Mathematical Notes. A Publication of the University of Miskolc*, 7(2):147–160, 2006.

Soumendra Nath Lahiri. *Resampling Methods for Dependent Data*. Springer-Verlag, New York, 2003.

Frédéric Lavancier, Jesper Møller, and Ege Rubak. Determinantal point process models and statistical inference (extended version). Technical report, Available on arXiv:1205.4818, 2014.

Ji Meng Loh. Bootstrapping an inhomogeneous point process. *Journal of Statistical Planning and Inference*, 140:734–749, 2010. ISSN 0378-3758. doi: https://doi.org/10.1016/j.jspi.2009.09.003.

Eugene Lukacs. *Characteristic Functions*. Griffin, 1970.

Torsten Mattfeldt, Henrike Häbel, and Frank Fleischer. Block bootstrap methods for the estimation of the intensity of a spatial point process with confidence bounds. *Journal of Microscopy*, 251:84–98, 2013. ISSN 1365-2818. doi: 10.1111/jmi.12048. URL http://dx.doi.org/10.1111/jmi.12048.

Arnaud Poinas, Bernard Delyon, and Frédéric Lavancier. Mixing properties and central limit theorem for associated point processes. *arXiv:1705.02276*, 2017.

Dimitris N. Politis and Joseph P. Romano. Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics*, 22:2031–2050, 1994. doi: 10.1214/aos/1176325770.

Michaela Prokešová and Eva Bjørn Vedel Jensen. Asymptotic Palm likelihood theory for stationary point processes. *Annals of the Institute of Statistical Mathematics*, 65:387–412, 2013.

Michael Sherman. Variance estimation for statistics computed from spatial lattice data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58:509–523, 1996. ISSN 00359246. URL http://www.jstor.org/stable/2345890.

Rasmus Waagepetersen and Yongtao Guan. Two-step estimation for inhomogeneous spatial point processes. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 71(3):685–702, 2009.

Rasmus Plenge Waagepetersen. An estimating function approach to inference for inhomogeneous Neyman–Scott processes. *Biometrics*, 63:252–258, 2007. doi: 10.1111/j.1541-0420.2006.00667.x.

Ganggang Xu, Rasmus Waagepetersen, and Yongtao Guan. Stochastic quasi-likelihood for case-control point pattern data. *Journal of the American Statistical Association*, 0 (ja):0–0, 2018. doi: 10.1080/01621459.2017.1421543.

The following appendix contains the proofs of Theorems 3.1 and 4.1. The last section of the appendix contains a number of technical lemmas used in the proofs of the main results. Proofs of technical lemmas and some lengthy technical derivations are available in the supplementary material.

A Proof of Theorem 3.1

Suppose first that we have verified Theorem 3.1 in the univariate case $q = 1$. Then, by $(\mathcal{H}4)$ and Lemma F.3, we may use the extension of the Cramér-Wold device in Lemma F.6 to verify Theorem 3.1 also for $q > 1$. We thus focus on the case $q = 1$.

The proof of Theorem 3.1 for $q = 1$ follows quite closely Karácsony (2006) and is based on the following theorem which is proved in Section B.

Theorem A.1. *Let the situation be as in Theorem 3.1 with $q = 1$, and assume in addition*

$(\mathcal{H}_b)$ $Z_n(\mathbf{l})$ is uniformly bounded with respect to $n \in \mathbb{N}$ and $\mathbf{l} \in \mathcal{D}_n$.

Then

$$\frac{1}{\sigma_n} \sum_{\mathbf{l} \in \mathcal{D}_n} \big(Z_n(\mathbf{l}) - \mathbb{E}Z_n(\mathbf{l})\big) \xrightarrow[n\to\infty]{distr.} \mathcal{N}(0,1)$$

where $\sigma_n^2 = \operatorname{Var} \sum_{\mathbf{l} \in \mathcal{D}_n} Z_n(\mathbf{l})$.

Proof of Theorem 3.1. Define for $L > 0$ and $n \in \mathbb{N}$:

- for $\mathbf{l} \in \mathbb{Z}^d$, $Z_n^{(L)}(\mathbf{l}) = Z_n(\mathbf{l})\, \mathbf{1}\big(Z_n(\mathbf{l}) \in [-L, L]\big)$,
- for $\mathbf{l} \in \mathbb{Z}^d$, $\breve{Z}_n^{(L)}(\mathbf{l}) = Z_n(\mathbf{l}) - Z_n^{(L)}(\mathbf{l})$,
- $X_n = \frac{1}{\sigma_n} \sum_{\mathbf{l} \in \mathcal{D}_n} \big(Z_n(\mathbf{l}) - \mathbb{E}Z_n(\mathbf{l})\big)$,
- $X_n^{(L)} = \frac{1}{\sigma_n} \sum_{\mathbf{l} \in \mathcal{D}_n} \big(Z_n^{(L)}(\mathbf{l}) - \mathbb{E}Z_n^{(L)}(\mathbf{l})\big)$,
- $\breve{X}_n^{(L)} = \frac{1}{\sigma_n} \sum_{\mathbf{l} \in \mathcal{D}_n} \big(\breve{Z}_n^{(L)}(\mathbf{l}) - \mathbb{E}\breve{Z}_n^{(L)}(\mathbf{l})\big)$.

By Lemma F.2, we have for $s \geq 0$,

$$\alpha_{1,1}^{\breve{Z}_n^L}(s_n r) \leq \alpha_{v_n, v_n}^{\mathbf{X}}(s_n r - s_n - 2R).$$

Further, by $(\mathcal{H}1)$–$(\mathcal{H}2)$, s_n is not decreasing with respect to n so we may find $r_0 \geq 1$ such that for all $r \geq r_0$ and $n \in \mathbb{N}$, $s_n(1 - 1/r) - 2R/r > 0$. Combining this with $(\mathcal{H}2)$, there exist constants $c_0, c_1 > 0$ so that

$$\begin{aligned}
&\sup_{n\in\mathbb{N}} \sum_{r=1}^{\infty} r^{d-1} \alpha_{1,1}^{\breve{Z}_n^{(L)}}(r s_n)^{\frac{\tau}{2+\tau}} \\
&\quad \leq c_0 + c_1 \sup_{n\in\mathbb{N}} \sum_{r=r_0}^{\infty} r^{d-1} (r s_n - s_n - 2R)^{\frac{-(d+\epsilon)\tau}{2+\tau}} \\
&\quad \leq c_0 + c_1 \sup_{n\in\mathbb{N}} \sum_{r=r_0}^{\infty} r^{d-1-\frac{(d+\epsilon)\tau}{2+\tau}} \left(s_n\Big(1 - \tfrac{1}{r}\Big) - \frac{2R}{r} \right)^{\frac{-(d+\epsilon)\tau}{2+\tau}} \\
&\quad \leq c_0 + c_1 \sup_{n\in\mathbb{N}} \left(s_n\Big(1 - \tfrac{1}{r_0}\Big) - \frac{2R}{r_0} \right)^{\frac{-(d+\epsilon)\tau}{2+\tau}} \sum_{r=r_0}^{\infty} r^{d-1-\frac{(d+\epsilon)\tau}{2+\tau}}.
\end{aligned}$$

By $(\mathcal{H}3)$, the last expression in the inequality is bounded. We may then adapt Theorem 1 in Fazekas et al. (2000) to the lattice $s_n\mathbb{Z}^d$ and so there exist a constant $c_2 > 0$ such that

$$\begin{aligned}\mathbb{E}(\breve{X}_n^{(L)})^2 &= \mathbb{E}\Big|\frac{1}{\sigma_n}\sum_{\mathbf{l}\in\mathcal{D}_n}\Big(\breve{Z}_n^{(L)}(\mathbf{l}) - \mathbb{E}\breve{Z}_n^{(L)}(\mathbf{l})\Big)\Big|^2 \\ &\le \frac{1}{\sigma_n^2}\Big(1 + 16d\sum_{r=1}^{\infty}(2r+1)^{d-1}\alpha_{1,1}^{\breve{Z}_n^{(L)}}(rs_n)^{\frac{\tau}{2+\tau}}\Big)\sum_{\mathbf{l}\in\mathcal{D}_n}\Big(\mathbb{E}|\breve{Z}_n^{(L)}(\mathbf{l})|^{2+\tau}\Big)^{\frac{2}{2+\tau}} \\ &\le \frac{c_2|\mathcal{D}_n|}{\sigma_n^2}\sup_{n\in\mathbb{N}}\sup_{\mathbf{l}\in\mathcal{D}_n}\Big(\mathbb{E}|\breve{Z}_n^{(L)}(\mathbf{l})|^{2+\tau}\Big)^{\frac{2}{2+\tau}}.\end{aligned}$$

By $(\mathcal{H}3)$,

$$\sup_{n\in\mathbb{N}}\sup_{\mathbf{l}\in\mathcal{D}_n}\Big(\mathbb{E}|\breve{Z}_n^{(L)}(\mathbf{l})|^{2+\tau}\Big)^{\frac{2}{2+\tau}} \xrightarrow[L\to\infty]{} 0.$$

Hence, it follows from the two last equations and $(\mathcal{H}4)$ that

$$\sup_{n\in\mathbb{N}}\mathbb{E}(\breve{X}_n^{(L)})^2 \xrightarrow[L\to\infty]{} 0. \tag{A.1}$$

We denote by $\sigma_n^2(L)$ the variance of $\sigma_n X_n^{(L)}$. Noticing that $\mathbb{E}X_n^2 = 1$, we have

$$\begin{aligned}\frac{\sigma_n^2(L)}{\sigma_n^2} - 1 &= \mathbb{E}(X_n^{(L)})^2 - \mathbb{E}X_n^2 \\ &= \mathbb{E}(X_n - \breve{X}_n^{(L)})^2 - \mathbb{E}X_n^2 \\ &= \mathbb{E}(\breve{X}_n^{(L)})^2 - 2\mathbb{E}(X_n\breve{X}_n^{(L)}).\end{aligned}$$

Then, by the Cauchy-Schwarz inequality and (A.1),

$$\sup_{n\in\mathbb{N}}\Big|\frac{\sigma_n^2(L)}{\sigma_n^2} - 1\Big| \xrightarrow[L\to\infty]{} 0. \tag{A.2}$$

For $n \in \mathbb{N}$, we have

$$\begin{aligned}&\big|\mathbb{E}e^{itX_n} - e^{-\frac{t^2}{2}}\big| \\ &\quad= \big|\mathbb{E}\big[\big(e^{it\breve{X}_n^{(L)}} - 1\big)e^{itX_n^{(L)}} + e^{itX_n^{(L)}} - e^{-\frac{t^2}{2}}\big]\big| \\ &\quad\le \mathbb{E}\big|e^{it\breve{X}_n^{(L)}} - 1\big| + \Big|\mathbb{E}e^{itX_n^{(L)}} - e^{-\frac{\sigma_n^2(L)}{\sigma_n^2}\frac{t^2}{2}}\Big| + \Big|e^{-\frac{\sigma_n^2(L)}{\sigma_n^2}\frac{t^2}{2}} - e^{-\frac{t^2}{2}}\Big|.\end{aligned} \tag{A.3}$$

Since for all $x \in \mathbb{R}$, $|e^{ix} - 1| \le |x|$,

$$\mathbb{E}|e^{it\breve{X}_n^{(L)}} - 1| \le \mathbb{E}|t\breve{X}_n^{(L)}| \le |t|\sup_{n\in\mathbb{N}}\sqrt{\mathbb{E}(\breve{X}_n^{(L)})^2}. \tag{A.4}$$

Writing $\delta_L = \sup_{n\in\mathbb{N}}|\frac{\sigma_n^2(L)}{\sigma_n^2} - 1|$ and $U_n = \frac{\sigma_n}{\sigma_n(L)}X_n^{(L)}$, we have

$$\begin{aligned}&\Big|\mathbb{E}e^{itX_n^{(L)}} - e^{-\frac{\sigma_n^2(L)}{\sigma_n^2}\frac{t^2}{2}}\Big| \\ &\quad= \Big|\mathbb{E}e^{it\frac{\sigma_n(L)}{\sigma_n}U_n} - e^{-\frac{\sigma_n^2(L)}{\sigma_n^2}\frac{t^2}{2}}\Big| \le \sup_{v\in[1-\delta_L,1+\delta_L]}\Big|\mathbb{E}e^{itvU_n} - e^{-\frac{(tv)^2}{2}}\Big|\end{aligned}$$

so by Theorem A.1 and Corollary 1 to Theorem 3.6.1 in Lukacs (1970), for $L \geq 0$,

$$\left|\mathbb{E}e^{itX_n^{(L)}} - e^{-\frac{\sigma_n^2(L)}{\sigma_n^2}\frac{t^2}{2}}\right| \xrightarrow[n\to\infty]{} 0. \tag{A.5}$$

Moreover, by a first order Taylor expansion with remainder,

$$\begin{aligned}\sup_{n\in\mathbb{N}}&\left|e^{-\frac{\sigma_n^2(L)}{\sigma_n^2}\frac{t^2}{2}} - e^{-\frac{t^2}{2}}\right| \\ &= e^{-\frac{t^2}{2}}\sup_{n\in\mathbb{N}}\left|e^{-(\frac{\sigma_n^2(L)}{\sigma_n^2}-1)\frac{t^2}{2}} - 1\right| \leq \frac{t^2}{2}\delta_L + \frac{t^4}{8}\exp(\delta_L\frac{t^2}{2})\delta_L^2.\end{aligned} \tag{A.6}$$

Therefore, by (A.3), (A.4), (A.5), and (A.6),

$$\limsup_{n\to\infty}\left|\mathbb{E}e^{itX_n} - e^{-\frac{t^2}{2}}\right| \leq |t|\sup_{n\in\mathbb{N}}\sqrt{\mathbb{E}(\breve{X}_n^{(L)})^2} + \frac{t^2}{2}\delta_L,$$

which by (A.1) and (A.2) tends to 0 as L tends to infinity.

□

B Proof of Theorem A.1

For ease of presentation, we assume that the bound in ($\mathcal{H}_b$) is 1. Define

- $Y_n(\mathbf{l}) = Z_n(\mathbf{l}) - \mathbb{E}Z_n(\mathbf{l})$,
- $S_n = \sum_{\mathbf{l}\in\mathcal{D}_n} Y_n(\mathbf{l})$,
- $a_n = \sum_{\mathbf{i},\mathbf{j}\in\mathcal{D}_n,\ \mathrm{d}(\mathbf{i},\mathbf{j})\leq m_n} \mathbb{E}[Y_n(\mathbf{i})Y_n(\mathbf{j})]$,
- $\bar{S}_n = \frac{1}{\sqrt{a_n}}\sum_{\mathbf{l}\in\mathcal{D}_n} Y_n(\mathbf{l})$,
- $\bar{S}_n(\mathbf{i}) = \frac{1}{\sqrt{a_n}}\sum_{\mathbf{j}\in\mathcal{D}_n,\ \mathrm{d}(\mathbf{i},\mathbf{j})\leq m_n} Y_n(\mathbf{j})$,

where if $\eta = 0$, $m_n = |D_n|^{1/(2d+\epsilon/2)}$ and if $\eta > 0$, $m_n = |W_n|^{\xi/d}$ with ξ verifying $\max\{\eta, d(1-\eta)/(2(d+\epsilon))\} < \xi < (1+\eta)/2$. Note that such ξ always exists. Then, by ($\mathcal{H}1$)–($\mathcal{H}2$),

$$m_n \to \infty, \qquad m_n/s_n \to \infty, \tag{B.1}$$

$$\sqrt{|\mathcal{D}_n|}m_n^{-d-\epsilon} \xrightarrow[n\to\infty]{} 0, \tag{B.2}$$

and

$$\sqrt{|\mathcal{D}_n|}(m_n/s_n)^{-d} \xrightarrow[n\to\infty]{} \infty. \tag{B.3}$$

By Lemma F.5, $\sup_{n\in\mathbb{N}}\mathbb{E}\bar{S}_n^2 < \infty$. Thus by Lemma 2 in Bolthausen (1982) (see also the discussion in Biscio et al., 2017), Theorem A.1 is proved if

$$\mathbb{E}[(it - \bar{S}_n)e^{it\bar{S}_n}] \xrightarrow[n\to\infty]{} 0. \tag{B.4}$$

Notice that

$$(it - \bar{S}_n)e^{it\bar{S}_n} = A_1 - A_2 - A_3,$$

where

$$A_1 = ite^{it\bar{S}_n}\Big(1 - \frac{1}{a_n} \sum_{\substack{\mathbf{i},\mathbf{j} \in \mathcal{D}_n \\ \mathrm{d}(\mathbf{i},\mathbf{j}) \le m_n}} Y_n(\mathbf{i})Y_n(\mathbf{j})\Big), \tag{B.5}$$

$$A_2 = \frac{e^{it\bar{S}_n}}{\sqrt{a_n}} \sum_{\mathbf{i} \in \mathcal{D}_n} Y_n(\mathbf{i})(1 - it\bar{S}_n(\mathbf{i}) - e^{-it\bar{S}_n(\mathbf{i})}), \tag{B.6}$$

$$A_3 = \frac{1}{\sqrt{a_n}} \sum_{\mathbf{i} \in \mathcal{D}_n} Y_n(\mathbf{i}) e^{it(\bar{S}_n - \bar{S}_n(\mathbf{i}))}. \tag{B.7}$$

Hence, (B.4) follows from the convergences to zero of A_1, A_2, and A_3 as established in Section S4 in the supplementary material.

C Proof of Theorem 4.1

The proof is based on the following result for a random field on a lattice that is proved in Section D.

Theorem C.1. *For $n \in \mathbb{N}$, let R_n be a random field on $\mathbb{Z}^d$, $\{W_n\}_{n\in\mathbb{N}}$ be a sequence of compact sets verifying ($\mathcal{S}0$) and, for $n \in \mathbb{N}$, let $\{B_{\mathbf{k}_n,\mathbf{t}} : \mathbf{k}_n \in \mathbb{N}^d, \mathbf{t} \in \mathcal{T}_{\mathbf{k}_n,n}\}$ be sub-rectangles defined as in (4.2) and such that ($\mathcal{S}1$) holds. For $q, n \in \mathbb{N}$ and $\mathbf{t} \in \mathbb{Z}^d$, let further Ψ be a function defined on subsets of the sample space of R_n and taking values in $\mathbb{R}^q$ and let $\Psi_A = \Psi((R_n(\mathbf{l}) : \mathbf{l} \in \mathbb{Z}^d \cap A))$, for $A = B_{\mathbf{k}_n,\mathbf{t}}$ or $A = W_n$. We assume that the following assumptions hold:*

(i) $\{|\Psi_{B_{\mathbf{k}_n,\mathbf{t}}} - \mathbb{E}\Psi_{B_{\mathbf{k}_n,\mathbf{t}}}|^4 : \mathbf{t} \in \mathcal{T}_{\mathbf{k}_n,n}, n \in \mathbb{N}\}$ *is uniformly integrable,*

(ii) $\alpha^{R_n}_{b_n,b_n}(\max_i k_{n,i}) \to 0$ *as* $n \to \infty$, *where* $b_n = \prod_{j=1}^d (2k_{n,j} + 1)$,

(iii) $\mathrm{Var}(\Psi_{W_n}) - \frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|} \sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \mathrm{Var}(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}}) \to 0$ *as* $n \to \infty$,

(iv) $\frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|} \sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \big(\mathbb{E}(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}}) - \mathbb{E}(\sum_{\mathbf{s}\in\mathcal{T}_{\mathbf{k}_n,n}} \frac{\Psi_{B_{\mathbf{k}_n,\mathbf{s}}}}{|\mathcal{T}_{\mathbf{k}_n,n}|})\big)^2 \to 0$ *as* $n \to \infty$.

Let further

$$\hat{\varsigma}^{R_n}_n = \frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|} \sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \Big(\Psi(B_{\mathbf{k}_n,\mathbf{t}}) - \frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|} \sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \Psi(B_{\mathbf{k}_n,\mathbf{t}})\Big)^2.$$

Then, we have the convergence,

$$\lim_{n\to\infty} \mathbb{E}(|\hat{\varsigma}^{R_n}_n - K_n|^2) = 0$$

where $K_n = \mathrm{Var}(\Psi_{W_n})$.

Below, we check the assumptions (i)–(iii) in Theorem C.1 with $R_n(\mathbf{l}) = Z_n(\mathbf{l})$ and $\Psi_A = T_A(\mathbf{X})/\sqrt{|\mathcal{D}_n(A)|}$, for $A \subset \mathbb{R}^d$. Then, Theorem 4.1 is proved directly by Theorem C.1.

Assumption (i).

For $\mathbf{l} \in \mathbb{Z}^d$ and $n \in \mathbb{N}$, let $Y_n(\mathbf{l}) = Z_n(\mathbf{l}) - \mathbb{E}(Z_n(\mathbf{l}))$ such that

$$T_{B_{\mathbf{k}_n,\mathbf{t}}}(\mathbf{X}) - \mathbb{E}(T_{B_{\mathbf{k}_n,\mathbf{t}}}(\mathbf{X})) = \sum_{\mathbf{l} \in \mathcal{D}_n(B_{\mathbf{k}_n,\mathbf{t}})} Y_n(\mathbf{l}). \tag{C.1}$$

Let ϵ' be as in ($\mathcal{S}2$), then by Lemma F.2 and ($\mathcal{S}6$), we have for $\epsilon < \epsilon'$,

$$\sum_{r=1}^{\infty} \big(\alpha_{5,5}^{Z_n}(r)\big)^{\frac{\epsilon}{6+\epsilon}} r^{5d-1} \leq \sum_{r=1}^{\infty} \big(\alpha_{5v_n,5v_n}^{\mathbf{X}}(r-1-2R)\big)^{\frac{\epsilon}{6+\epsilon}} r^{5d-1} < \infty. \tag{C.2}$$

Then, by ($\mathcal{S}2$) and (C.2) we may apply Theorem 1 in Fazekas et al. (2000) which states the existence of $c_1 > 0$ such that

$$\mathbb{E}\Bigg|\sum_{\mathbf{l} \in \mathcal{D}_n(B_{\mathbf{k}_n,\mathbf{t}})} \frac{Y_n(\mathbf{l})}{\sqrt{|\mathcal{D}_n(B_{\mathbf{k}_n,\mathbf{t}})|}}\Bigg|^{4+\epsilon'/2} \leq c_1 \max\{U, V\} \tag{C.3}$$

where

$$U = \sum_{\mathbf{l} \in \mathcal{D}_n(B_{\mathbf{k}_n,\mathbf{t}})} \Bigg(\mathbb{E}\Bigg|\frac{Y_n(\mathbf{l})}{\sqrt{|\mathcal{D}_n(B_{\mathbf{k}_n,\mathbf{t}})|}}\Bigg|^{4+\epsilon'}\Bigg)^{\frac{4+\epsilon'/2}{4+\epsilon'}},$$

$$V = \Bigg\{\sum_{\mathbf{l} \in \mathcal{D}_n(B_{\mathbf{k}_n,\mathbf{t}})} \Bigg(\mathbb{E}\Bigg|\frac{Y_n(\mathbf{l})}{\sqrt{|\mathcal{D}_n(B_{\mathbf{k}_n,\mathbf{t}})|}}\Bigg|^{2+\epsilon'/2}\Bigg)^{\frac{2}{2+\epsilon'/2}}\Bigg\}^{2+\epsilon'/4}.$$

Further, by ($\mathcal{S}2$), there exists a constant c_2 such that

$$U \leq c_2 \frac{|\mathcal{D}_n(B_{\mathbf{k}_n,\mathbf{t}})|}{|\mathcal{D}_n(B_{\mathbf{k}_n,\mathbf{t}})|^{2+\epsilon'/4}} = \frac{c_2}{|\mathcal{D}_n(B_{\mathbf{k}_n,\mathbf{t}})|^{1+\epsilon'/4}}$$

and

$$V \leq \left(c_2 \frac{|\mathcal{D}_n(B_{\mathbf{k}_n,\mathbf{t}})|}{|\mathcal{D}_n(B_{\mathbf{k}_n,\mathbf{t}})|}\right)^{2+\epsilon'/4} = c_2^{2+\epsilon'/4}.$$

Both of the last upper bounds on U and V are bounded. Therefore, by (C.1) and (C.3),

$$\sup_{n \in \mathbb{N}} \sup_{t \in \mathcal{T}_{\mathbf{k}_n,n}} \mathbb{E}\Bigg|\frac{T_{B_{\mathbf{k}_n,\mathbf{t}}}(\mathbf{X}) - \mathbb{E}\big(T_{B_{\mathbf{k}_n,\mathbf{t}}}(\mathbf{X})\big)}{\sqrt{|\mathcal{D}_n(B_{\mathbf{k}_n,\mathbf{t}})|}}\Bigg|^{4+\epsilon'/2} < \infty \tag{C.4}$$

which implies (i) by (25.13) in Billingsley (1995).

Assumption (ii).

For c, δ as in ($\mathcal{S}5$) and v_n as below (3.2), we have by Lemma F.2,

$$\alpha_{b_n,b_n}^{Z_n}(\max_i k_{n,i}) \leq \alpha_{b_n v_n, b_n v_n}^{\mathbf{X}}(\max_i k_{n,i} - s_n - 2R) \leq c \frac{b_n v_n}{(\max_i k_{n,i} - s_n)^{d+\delta}}$$

which by ($\mathcal{S}5$) converges towards 0 as n tends to infinity.

Assumptions (iii)–(iv).

Assumptions (iii)–(iv) are the same as ($\mathcal{S}3$)–($\mathcal{S}4$).

D Proof of Theorem C.1

The proof of Theorem C.1 is based on several applications of the following Theorem D.1 which states an intermediate result and is proved in Section E. Consequently, these theorems may looks similar at first sight.

Theorem D.1. *For $n \in \mathbb{N}$, let R_n be a random field on $\mathbb{Z}^d$, $\{W_n\}_{n\in\mathbb{N}}$ be a sequence of compact sets verifying ($\mathcal{S}0$) and, for $n \in \mathbb{N}$, let $\{B_{\mathbf{k}_n,\mathbf{t}} : \mathbf{k}_n \in \mathbb{N}^d, \mathbf{t} \in \mathcal{T}_{\mathbf{k}_n,n}\}$ be sub-rectangles defined as in* (4.2) *and verifying ($\mathcal{S}1$). For $q, n \in \mathbb{N}$ and $\mathbf{t} \in \mathbb{Z}^d$, let further h be a function defined on subsets of the sample space of R_n, taking values into $\mathbb{R}^q$ and let $h_A = h((R_n(\mathbf{l}) : \mathbf{l} \in \mathbb{Z}^d \cap A))$ for $A = B_{\mathbf{k}_n,\mathbf{t}}$ or $A = W_n$. We assume that the following assumptions hold:*

(i') $\{h^2_{B_{\mathbf{k}_n,\mathbf{t}}} : \mathbf{t} \in \mathcal{T}_{\mathbf{k}_n,n}, n \in \mathbb{N}\}$ is uniformly integrable,

(ii') $\alpha^{R_n}_{b_n,b_n}(\max_i k_{n,i}) \to 0$ as $n \to \infty$, where $b_n = \prod_{j=1}^d (2k_{n,j} + 1)$,

(iii') $\mathbb{E}\big(\sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \frac{h_{B_{\mathbf{k}_n,\mathbf{t}}}}{|\mathcal{T}_{\mathbf{k}_n,n}|}\big) - \mathbb{E}(h_{W_n}) \to 0$ as $n \to \infty$.

Then, we have the convergence

$$\frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|} \sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \big(h_{B_{\mathbf{k}_n,\mathbf{t}}} - \mathbb{E}(h_{W_n})\big) \xrightarrow[n\to\infty]{L^2} 0.$$

We now give the proof of Theorem C.1 and to shorten, we define $\overline{\Psi}_{B_{\mathbf{k}_n}} = \sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \Psi_{B_{\mathbf{k}_n,\mathbf{t}}}/|\mathcal{T}_{\mathbf{k}_n,n}|$. For $x = (x_1, \dots, x_d)^T \in \mathbb{R}^d$ and M a square matrix in $\mathbb{R}^d \times \mathbb{R}^d$, we further denote by $|x| = \sqrt{\sum_{i=1}^d x_i^2}$ and $|M| = \sqrt{\sum_{i,j} M_{ij}^2}$ the Euclidean norms of x and M, respectively, and by x^2 the matrix xx^T.

From the statement of Theorem C.1, we have

$$\hat{\varsigma}^{R_n}_n = \frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|} \sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \big(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}} - \mathbb{E}(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}}) + \mathbb{E}(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}}) - \mathbb{E}(\overline{\Psi}_{B_{\mathbf{k}_n}}) + \mathbb{E}(\overline{\Psi}_{B_{\mathbf{k}_n}}) - \overline{\Psi}_{B_{\mathbf{k}_n}}\big)^2.$$

Hence,

$$\hat{\varsigma}^{R_n}_n = C_1 + C_2 + C_3 + C_4 + C_5 + C_6, \tag{D.1}$$

where the terms C_1–C_6 are all $q \times q$ matrices given below:

$$C_1 = \frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|} \sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \big(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}} - \mathbb{E}(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}})\big)^2,$$

$$C_2 = \frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|} \sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \big(\mathbb{E}(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}}) - \mathbb{E}(\overline{\Psi}_{B_{\mathbf{k}_n}})\big)^2,$$

$$C_3 = \big(\mathbb{E}(\overline{\Psi}_{B_{\mathbf{k}_n}}) - \overline{\Psi}_{B_{\mathbf{k}_n}}\big)^2,$$

$$
\begin{aligned}
C_4 = & \frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|} \sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \big(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}} - \mathbb{E}(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}})\big)\big(\mathbb{E}(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}}) - \mathbb{E}(\overline{\Psi}_{B_{\mathbf{k}_n}})\big)^T \\
& + \frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|} \sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \big(\mathbb{E}(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}}) - \mathbb{E}(\overline{\Psi}_{B_{\mathbf{k}_n}})\big)\big(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}} - \mathbb{E}(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}})\big)^T, \\
C_5 = & \frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|} \sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \big(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}} - \mathbb{E}(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}})\big)\big(\mathbb{E}(\overline{\Psi}_{B_{\mathbf{k}_n}}) - \overline{\Psi}_{B_{\mathbf{k}_n}}\big)^T \\
& + \frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|} \sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \big(\mathbb{E}(\overline{\Psi}_{B_{\mathbf{k}_n}}) - \overline{\Psi}_{B_{\mathbf{k}_n}}\big)\big(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}} - \mathbb{E}(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}})\big)^T, \\
C_6 = & \frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|} \sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \big(\mathbb{E}(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}}) - \mathbb{E}(\overline{\Psi}_{B_{\mathbf{k}_n}})\big)\big(\mathbb{E}(\overline{\Psi}_{B_{\mathbf{k}_n}}) - \overline{\Psi}_{B_{\mathbf{k}_n}}\big)^T \\
& + \frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|} \sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \big(\mathbb{E}(\overline{\Psi}_{B_{\mathbf{k}_n}}) - \overline{\Psi}_{B_{\mathbf{k}_n}}\big)\big(\mathbb{E}(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}}) - \mathbb{E}(\overline{\Psi}_{B_{\mathbf{k}_n}})\big)^T.
\end{aligned}
$$

The assumption (iv), implies directly that

$$
C_2 \xrightarrow[n\to\infty]{} 0. \tag{D.2}
$$

By applying the Cauchy-Schwarz inequality for each sum in C_4, we have $\mathbb{E}(|C_4|^2) \leq 4\mathbb{E}(|C_1|)|C_2|$. Further, by (i) and (25.11) in Billingsley (1995), $\mathbb{E}(|C_1|)$ is uniformly bounded with respect to $n \in \mathbb{N}$ and $\mathbf{t} \in \mathbb{Z}^d$. Thus, by (D.2), it follows that

$$
\mathbb{E}(|C_4|^2) \xrightarrow[n\to\infty]{} 0. \tag{D.3}
$$

We have

$$
\begin{aligned}
C_5 + C_6 = & \frac{\mathbb{E}(\overline{\Psi}_{B_{\mathbf{k}_n}}) - \overline{\Psi}_{B_{\mathbf{k}_n}}}{|\mathcal{T}_{\mathbf{k}_n,n}|} \sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \big(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}} - \mathbb{E}(\overline{\Psi}_{B_{\mathbf{k}_n}})\big)^T \\
& + \frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|} \sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \big(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}} - \mathbb{E}(\overline{\Psi}_{B_{\mathbf{k}_n}})\big)\big(\mathbb{E}(\overline{\Psi}_{B_{\mathbf{k}_n}}) - \overline{\Psi}_{B_{\mathbf{k}_n}}\big)^T \\
= & -2(\mathbb{E}(\overline{\Psi}_{B_{\mathbf{k}_n}}) - \overline{\Psi}_{B_{\mathbf{k}_n}})^2
\end{aligned}
$$

so

$$
C_3 + C_5 + C_6 = -\big(\mathbb{E}(\overline{\Psi}_{B_{\mathbf{k}_n}}) - \overline{\Psi}_{B_{\mathbf{k}_n}}\big)^2. \tag{D.4}
$$

Let $Y_n = \frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|}\sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}}(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}} - \mathbb{E}(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}}))$ and notice that $C_3 + C_5 + C_6 = -Y_n^2$. Using (i)–(ii), we may apply Theorem D.1 with $h_\cdot = \Psi_\cdot - \mathbb{E}(\Psi_\cdot)$ so that

$$
\mathbb{E}(|Y_n|^2) \xrightarrow[n\to\infty]{} 0. \tag{D.5}
$$

Using (i)–(ii) and (iii), we may apply Theorem D.1 with $h_\cdot = (\Psi_\cdot - \mathbb{E}(\Psi_\cdot))^2$. Hence,

$$
\frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|} \sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \big(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}} - \mathbb{E}(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}})\big)^2 - \mathbb{E}\Big(\big(\Psi_{W_n} - \mathbb{E}(\Psi_{W_n})\big)^2\Big) \xrightarrow[n\to\infty]{L^2} 0
$$

which may be written as the convergence

$$
C_1 - K_n \xrightarrow[n\to\infty]{L^2} 0. \tag{D.6}
$$

By Theorem 4.5.4 in Chung (2001), (D.6) implies that $|C_1 - K_n|^2$ is uniformly integrable with respect to n. Moreover, $\mathbb{E}(|C_1|^2) = \mathbb{E}(|C_1 - K_n + K_n|^2) \le \mathbb{E}(|C_1 - K_n|^2) + |K_n|^2$ and it follows from $(\mathcal{S}3)$ that K_n is uniformly bounded. Thus, $|C_1|^2$ is uniformly integrable so that by Lemma F.7, Y_n^4 is also uniformly integrable. Further, (D.5) implies that $Y_n^2 \xrightarrow[n\to\infty]{\mathbb{P}} 0$ so by Theorem 4.5.4 in Chung (2001), we have $Y_n \xrightarrow[n\to\infty]{L^2} 0$. By (D.4), the last implies that

$$C_3 + C_5 + C_6 \xrightarrow[n\to\infty]{L^2} 0. \tag{D.7}$$

Finally, Theorem C.1 is proved by combining (D.1), (D.2), (D.3), (D.6), and (D.7).

E Proof of Theorem D.1

$$\begin{aligned}
&\mathbb{E}\Big|\sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \frac{h_{B_{\mathbf{k}_n,\mathbf{t}}}}{|\mathcal{T}_{\mathbf{k}_n,n}|} - \mathbb{E}(h_{W_n})\Big|^2 \\
&\quad = \mathbb{E}\Big|\sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \frac{h_{B_{\mathbf{k}_n,\mathbf{t}}} - \mathbb{E}(h_{B_{\mathbf{k}_n,\mathbf{t}}})}{|\mathcal{T}_{\mathbf{k}_n,n}|}\Big|^2 + \Big|\sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \frac{\mathbb{E}(h_{B_{\mathbf{k}_n,\mathbf{t}}})}{|\mathcal{T}_{\mathbf{k}_n,n}|} - \mathbb{E}(h_{W_n})\Big|^2
\end{aligned} \tag{E.1}$$

Hence, if in (E.1) the first expectation on the right-hand side converges to 0 as n tends to infinity, Theorem D.1 is proved by (iii') and (E.1). We have

$$\begin{aligned}
\mathbb{E}\Big|\sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \frac{h_{B_{\mathbf{k}_n,\mathbf{t}}} - \mathbb{E}(h_{B_{\mathbf{k}_n,\mathbf{t}}})}{|\mathcal{T}_{\mathbf{k}_n,n}|}\Big|^2 &\le \frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|^2} \sum_{\mathbf{t}_1,\mathbf{t}_2\in\mathcal{T}_{\mathbf{k}_n,n}} \operatorname{Cov}(|h_{B_{\mathbf{k}_n,\mathbf{t}_1}}|, |h_{B_{\mathbf{k}_n,\mathbf{t}_2}}|) \\
&= M_1 + M_2,
\end{aligned}$$

where

$$\begin{aligned}
M_1 &= \frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|^2} \sum_{\substack{\mathbf{t}_1,\mathbf{t}_2\in\mathcal{T}_{\mathbf{k}_n,n},\\ \mathrm{d}(\mathbb{Z}^d\cap B_{\mathbf{k}_n,\mathbf{t}_1},\mathbb{Z}^d\cap B_{\mathbf{k}_n,\mathbf{t}_2})\le\max\mathbf{k}_{n,i}}} \operatorname{Cov}(|h_{B_{\mathbf{k}_n,\mathbf{t}_1}}|, |h_{B_{\mathbf{k}_n,\mathbf{t}_2}}|), \\
M_2 &= \frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|^2} \sum_{\substack{\mathbf{t}_1,\mathbf{t}_2\in\mathcal{T}_{\mathbf{k}_n,n},\\ \mathrm{d}(\mathbb{Z}^d\cap B_{\mathbf{k}_n,\mathbf{t}_1},\mathbb{Z}^d\cap B_{\mathbf{k}_n,\mathbf{t}_2})>\max\mathbf{k}_{n,i}}} \operatorname{Cov}(|h_{B_{\mathbf{k}_n,\mathbf{t}_1}}|, |h_{B_{\mathbf{k}_n,\mathbf{t}_2}}|).
\end{aligned}$$

Regarding M_1, for a given $\mathbf{t}_1 \in \mathcal{T}_{\mathbf{k}_n,n}$, there is at most $(2\max_i \mathbf{k}_{n,i} + 1)^d$ choices for t_2. Thus,

$$\begin{aligned}
M_1 &\le \frac{(2\max_i \mathbf{k}_{n,i} + 1)^d}{|\mathcal{T}_{\mathbf{k}_n,n}|} \sup_{\substack{\mathbf{t}_1,\mathbf{t}_2\in\mathcal{T}_{\mathbf{k}_n,n},\\ \mathrm{d}(\mathbb{Z}^d\cap B_{\mathbf{k}_n,\mathbf{t}_1},\mathbb{Z}^d\cap B_{\mathbf{k}_n,\mathbf{t}_2})\le\max\mathbf{k}_{n,i}}} \operatorname{Cov}(|h_{B_{\mathbf{k}_n,\mathbf{t}_1}}|, |h_{B_{\mathbf{k}_n,\mathbf{t}_2}}|) \\
&\le \frac{(2\max_i \mathbf{k}_{n,i} + 1)^d}{|\mathcal{T}_{\mathbf{k}_n,n}|} \sup_{\substack{\mathbf{t}_1,\mathbf{t}_2\in\mathcal{T}_{\mathbf{k}_n,n},\\ \mathrm{d}(\mathbb{Z}^d\cap B_{\mathbf{k}_n,\mathbf{t}_1},\mathbb{Z}^d\cap B_{\mathbf{k}_n,\mathbf{t}_2})\le\max\mathbf{k}_{n,i}}} \sqrt{\mathbb{E}(|h_{B_{\mathbf{k}_n,\mathbf{t}_1}}|^2)\mathbb{E}(|h_{B_{\mathbf{k}_n,\mathbf{t}_2}}|^2)}.
\end{aligned}$$

By (i'), there exists a constant $c_1 > 0$ such that

$$M_1 \le \frac{(2\max_i k_{n,i} + 1)^d}{|\mathcal{T}_{\mathbf{k}_n,n}|} c_1 = \frac{(2\max_i k_{n,i} + 1)^d}{\prod_{i=1}^d (m_{n,i} - k_{n,i} + 1)} c_1.$$

which by ($\mathcal{S}1$) implies that M_1 tends to 0 as n tends to infinity. We have

$$M_2 \leq \sup_{\substack{\mathbf{t}_1,\mathbf{t}_2\in\mathcal{T}_{\mathbf{k}_n,n},\\ \mathrm{d}(\mathbb{Z}^d\cap B_{\mathbf{k}_n,\mathbf{t}_1},\mathbb{Z}^d\cap B_{\mathbf{k}_n,\mathbf{t}_2})>\max \mathbf{k}_{n,i}}} \mathrm{Cov}(|h_{B_{\mathbf{k}_n,\mathbf{t}_1}}|,|h_{B_{\mathbf{k}_n,\mathbf{t}_2}}|).$$

Further, by (4.2) for all $t\in\mathbb{Z}^d$, $|\mathbb{Z}^d_{B_{\mathbf{k}_n,\mathbf{t}}}|\leq b_n$, where $b_n=\prod_{j=1}^d(2k_{n,j}+1)$. Then, by Lemma 1 in Sherman (1996), for any $\eta>0$, we have

$$M_2 \leq 4\eta^2\alpha^{Z_n}_{b_n,b_n}(\max \mathbf{k}_{n,i}) + 3\sqrt{c_2}\sqrt{\mathbb{E}(X_1^{(\eta)})^2} + 3\sqrt{c_2}\sqrt{\mathbb{E}(X_2^{(\eta)})^2} \tag{E.2}$$

where for $i=1,2$, $X_i=|h_{B_{\mathbf{k}_n,\mathbf{t}_i}}|$, $X_i^\eta = X_i\,\mathbf{1}(X_i\geq\eta)$, and $c_2=\sup_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}}\mathbb{E}(|h_{B_{\mathbf{k}_n,\mathbf{t}}}|^2)$ which by (i') is finite. Hence, by first taking lim sup as n tends to infinity and second as η tends to infinity, it follows by (E.2), (i'), and (ii'), that M_2 tends to 0 as n tends to infinity. Therefore, $\mathbb{E}\left|\sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}}(h_{B_{\mathbf{k}_n,\mathbf{t}}}-\mathbb{E}(h_{B_{\mathbf{k}_n,\mathbf{t}}}))/|\mathcal{T}_{\mathbf{k}_n,n}|\right|^2$ converges to 0 as n tends to infinity.

F Lemmas

This section contains a number of technical lemmas used in the proofs of the main results. Proofs of the lemmas are given in the supplementary material.

Lemma F.1. *For all* $\mathbf{l},\mathbf{k}\in s_n\mathbb{Z}^d$, *we have*

$$\mathrm{d}(\mathbf{l},\mathbf{k})-s_n-2R\leq \mathrm{d}\big(C_n^{\oplus R}(\mathbf{l}),C_n^{\oplus R}(\mathbf{k})\big)\leq \mathrm{d}(\mathbf{l},\mathbf{k})+s_n+2R.$$

Lemma F.2. *For* $c_1,c_2,r\geq 0$, *we have*

$$\alpha^Z_{c_1,c_2}(r)\leq\alpha^{\mathbf{X}}_{c_1v_n,c_2v_n}(r-s_n-2R).$$

Lemma F.3. *We have*

$$\limsup_{n\to\infty}\lambda_{\max}\left(\frac{\Sigma_n}{|\mathcal{D}_n|}\right)<\infty.$$

where $\lambda_{\max}(M)$ *denotes the maximal eigen value of a symmetric matrix* M.

Lemma F.4. *For* $k\in\mathbb{N}$ *and* $\mathbf{i}\in s_n\mathbb{Z}^d$,

$$|\{\mathbf{j}\in s_n\mathbb{Z}^d:\mathrm{d}(\mathbf{i},\mathbf{j})=s_nk\}|\leq 3^dk^{d-1}.$$

Lemma F.5. *Under the assumptions* ($\mathcal{H}_b$), ($\mathcal{H}2$), *and* ($\mathcal{H}4$), *we have the convergence*

$$\left|1-\frac{a_n}{\sigma_n^2}\right|\xrightarrow[n\to\infty]{}0.$$

Lemma F.6 (Biscio et al. (2017))**.** *Let* $\{X_n\}_{n\in\mathbb{N}}$ *be a sequence of random variables in* $\mathbb{R}^p$, *for* $p\in\mathbb{N}$, *such that*

$$0<\liminf_{n\to\infty}\lambda_{\min}\big(\mathrm{Var}(X_n)\big)<\limsup_{n\to\infty}\lambda_{\max}\big(\mathrm{Var}(X_n)\big)<\infty,$$

where for a symmetric matrix M, $\lambda_{\min}(M)$ *and* $\lambda_{\max}(M)$ *denote the minimal and maximal eigen values of* M.

Then, $\mathrm{Var}(X_n)^{-1/2}X_n\xrightarrow[n\to\infty]{distr.}\mathcal{N}(0,I_p)$ *if for all* $a\in\mathbb{R}^p$,

$$\big(a^T\,\mathrm{Var}(X_n)a\big)^{-\frac12}a^TX_n\xrightarrow[n\to\infty]{distr.}\mathcal{N}(0,1).$$

Lemma F.7. *Let the situation be as in Appendix D. We have*

$$|Y_n|^4 \leq q|C_1|^2.$$

Proof of Lemma F.7. For any vector x, let $[x]_i$ denotes its i-th coordinate. By the Cauchy-Schwarz inequality,

$$|Y_n|^2 \leq \frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|} \sum_{i=1}^{q} \sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} [\Psi_{B_{\mathbf{k}_n,\mathbf{t}}} - \mathbb{E}(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}})]_i^2.$$

so that by applying the Cauchy-Schwarz inequality on the first sum,

$$|Y_n|^4 \leq \frac{q}{|\mathcal{T}_{\mathbf{k}_n,n}|^2} \sum_{i=1}^{q} \Big(\sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} [\Psi_{B_{\mathbf{k}_n,\mathbf{t}}} - \mathbb{E}(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}})]_i^2 \Big)^2.$$

On the other hand,

$$|C_1|^2 = \frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|^2} \sum_{i,j=1}^{q} \Big(\sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} [\Psi_{B_{\mathbf{k}_n,\mathbf{t}}} - \mathbb{E}(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}})]_i [\Psi_{B_{\mathbf{k}_n,\mathbf{t}}} - \mathbb{E}(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}})]_j \Big)^2$$

which implies that $|Y_n|^4 \leq q|C_1|^2$. □

Supplemental material

S1 Central limit theorem by blocking technique

For $0 < \beta < \gamma < 1$ and $\mathbf{l}_n = (l_{1,n}, \dots, l_{d,n}) \in n^\gamma \mathbb{Z}$, define blocks

$$B_n(\mathbf{l}_n) = \prod_{j=1}^{d} \big(l_{j,n} - (n^\gamma - n^\beta)/2, l_{j,n} + (n^\gamma - n^\beta)/2\big]. \tag{S1.1}$$

For $n \in \mathbb{N}$, we denote by $w_n = (n^\gamma - n^\beta)^d$ the volume of each $B_n(\mathbf{l}_n)$. Note that contrary to (3.2) the union of blocks defined in (S1.1) over $n^\gamma \mathbb{Z}$ do not cover $\mathbb{R}^d$. To apply the blocking technique central limit theorem we assume that $T_{W_n}(\mathbf{X})$ can be approximated (see ($\mathcal{C}5$) below) by a sum

$$\sum_{\mathbf{k}_n \in \mathcal{E}_n} f_{n,B_n(\mathbf{k}_n)}(\mathbf{X})$$

where

$$\mathcal{E}_n = \{\mathbf{l} \in n^\gamma \mathbb{Z},\ B_n(\mathbf{l}) \subset W_n\} \tag{S1.2}$$

and for each n and $\mathbf{k}_n \in n^\gamma \mathbb{Z}$, $f_{n,B_n(\mathbf{k}_n)}(\mathbf{X})$ is a $q \geq 1$ dimensional statistic depending on $\mathbf{X}$ only through $\mathbf{X} \cap (B_n(\mathbf{k}_n) \oplus R)$. We consider the following conditions.

($\mathcal{C}1$) The cardinality $|\mathcal{E}_n|$ of $\mathcal{E}_n$ verifies $|\mathcal{E}_n| = O(n^{d(1-\gamma)})$.

($\mathcal{C}2$) There exists $\epsilon > 0$ such that $\sup_{m \geq 0} \alpha^{\mathbf{X}}_{m,m}(s)/m = O(\frac{1}{s^{d+\epsilon}})$ and $2d/(2d+\epsilon) < \beta < \gamma < 1$.

($\mathcal{C}3$) There exists a $\tau > 0$ so that

$$\sup_{n \in \mathbb{N}} \sum_{\mathbf{k}_n \in \mathcal{E}_n} \mathbb{E}\Big|(w_n|\mathcal{E}_n|)^{-\frac{1}{2}} [f_{n,B_n(\mathbf{k}_n)}(\mathbf{X}) - \mathbb{E}(f_{n,B_n(\mathbf{k}_n)}(\mathbf{X}))]\Big|^{2+\tau} < \infty.$$

($\mathcal{C}4$) There exists a positive definite matrix Σ, such that for all $\mathbf{k}_n \in n^\gamma \mathbb{Z}$,

$$\lim_{n \to \infty} w_n^{-1} \operatorname{Var} f_{n,B_n(\mathbf{k}_n)}(\mathbf{X}) = \Sigma.$$

A preliminary central limit theorem is then given below.

Theorem S1.1. *If ($\mathcal{C}1$)–($\mathcal{C}4$) holds,*

$$(w_n|\mathcal{E}_n|)^{-\frac{1}{2}} \sum_{\mathbf{k} \in \mathcal{E}_n} \big(f_{n,B_n(\mathbf{k})}(\mathbf{X}) - \mathbb{E}[f_{n,B_n(\mathbf{k})}(\mathbf{X})]\big) \xrightarrow[n \to \infty]{distr.} N(0, \Sigma).$$

($\mathcal{C}5$) $\lim_{n \to \infty} \operatorname{Var}\big[|W_n|^{-\frac{1}{2}} T_{W_n}(\mathbf{X}) - (w_n|\mathcal{E}_n|)^{-\frac{1}{2}} \sum_{\mathbf{k} \in \mathcal{E}_n} f_{n,B_n(\mathbf{k})}(\mathbf{X})\big] = 0.$

A prerequisite for verifying ($\mathcal{C}5$) will typically be $\lim_{n \to \infty} |W_n|/(w_n|\mathcal{E}_n|) = 1$, see for instance Prokešová and Jensen (2013). In other words, W_n can be approximated well by the sum of blocks $B_n(\mathbf{k}_n)$ contained in W_n. This implies that W_n expands in all directions and, in light of ($\mathcal{C}1$), that $|W_n|$ is proportional to n^d.

The next theorem follows easily from Theorem S1.1 and Chebyshev's inequality.

Theorem S1.2. *If ($\mathcal{C}1$)–($\mathcal{C}5$) hold, then*

$$|W_n|^{-\frac{1}{2}}\big(T_{W_n}(\mathbf{X}) - \mathbb{E}T_{W_n}(\mathbf{X})\big) \xrightarrow[N\to\infty]{distr.} (0, \Sigma),$$

where $\Sigma = \lim_{n\to\infty}|W_n|^{-1}\operatorname{Var}T_{W_n}(\mathbf{X})$.

In the literature, Theorem S1.1 usually appears as an intermediary result in the proof of Theorem S1.2. However, in addition to ease the reading of the proof, we believe that it is important to state these theorems separately in order to easily understand the differences with Theorem 3.1.

Conditions ($\mathcal{C}3$)–($\mathcal{C}4$) ensure that we may use Lyapunov's central limit theorem in the proof of S1.1. Other central limit theorems, e.g. Lindeberg's, for triangular arrays could be used instead.

S2 Comparison between Theorems 3.1 and S1.2

Theorems 3.1 and S1.1 are both based on subdivisions of $\mathbb{R}^d$ into blocks or subsquares. However, for Theorem S1.1, we consider a set of blocks that do not cover W_n. For this reason we need condition ($\mathcal{C}5$) for Theorem S1.2, i.e. that contributions to $T_{W_n}(\mathbf{X})$ from the omitted part of $|W_n|$ are negligible asymptotically.

The condition ($\mathcal{C}5$) is usually not stated explicitly in the aforementioned references that used a central limit theorem based on the blocking technique. Instead ($\mathcal{C}5$) is checked by direct calculation for the specific statistics considered. Note that in all these references, statistics of the form (1.1) are considered with $p = 2$ and additional assumptions on $f_{\mathrm{I},W}$ in (1.2) and $\mathbf{X}$ are imposed. For instance, it is often assumed that $f_{\mathrm{I},W}$ is bounded with finite range and that the total variation of the reduced factorial cumulant measures of $\mathbf{X}$ up to order 4 is finite. However, under our general setting we have not been able to make such a calculation. Therefore, ($\mathcal{C}5$) must be checked for each application, using specific properties of $\mathbf{X}$ and $f_{\mathrm{I},W}$.

The remaining assumptions in Theorems 3.1 and S1.2 may be separated in different categories: ($\mathcal{H}1$), ($\mathcal{C}1$) (together with ($\mathcal{C}5$)) deal with the observation windows; ($\mathcal{H}2$), ($\mathcal{C}2$) are mixing conditions of the point process $\mathbf{X}$; ($\mathcal{H}3$), ($\mathcal{C}3$) are moment conditions of the statistics $T_W(\mathbf{X})$ that ensure uniform integrability; and finally ($\mathcal{H}4$), ($\mathcal{C}4$) control the asymptotic variance.

The main difference between the theorems is the conditions on the observation windows. Assumption ($\mathcal{H}1$) is rather weak and does not impose strong restrictions on the shape of the observation window. For instance, if $d = 2$, ($\mathcal{H}1$) holds for a sequence of rectangles W_n of width n and constant length. However, ($\mathcal{C}1$) does not hold for this choice of W_n since $|W_n| = O(n)$. Moreover, it is worth mentioning that we do not assume in Theorems 3.1 and S1.2 that W_n is convex as is usually done in the literature, see for instance Guan and Sherman (2007), Waagepetersen and Guan (2009) and Prokešová and Jensen (2013). The mixing coefficient conditions ($\mathcal{H}2$) and ($\mathcal{C}2$) have both been used extensively in spatial statistics. We are not aware of point process examples where one condition holds and the other does not. So for applications in point process statistics it is not clear that one is more advantageous than the other. In $(\mathcal{H}3)$, we require that $\tau > 2d/\epsilon$, where ϵ is the same as in $(\mathcal{H}2)$, but in $(\mathcal{C}3)$ we only have $\tau > 0$. Hence, it seems that Theorem 3.1 requires stronger moments conditions on $\mathbf{X}$ than Theorem S1.2. However, as discussed above, further assumptions on the moments of $\mathbf{X}$ are usually needed to check $(\mathcal{C}5)$. Assumptions $(\mathcal{H}4)$ is weaker than $(\mathcal{C}4)$ since it does not assume the existence of a limiting variance.

In conclusion, Theorem S1.2 is more restrictive regarding the conditions on the observation windows than Theorem 3.1 and requires $(\mathcal{C}5)$ whose verification may be very challenging in practice and typically requires further assumptions. Thus in general we recommend to use Theorem 3.1 instead of Theorem S1.2.

S3 Proof of Theorem S1.1

For $n \in \mathbb{N}$, $a \in \mathbb{R}^q$ and $\mathbf{k}_n \in \mathcal{E}_n$, define

$$X_n(\mathbf{k}_n) = a^T (w_n|\mathcal{E}_n|)^{-\frac{1}{2}} (f_{n,B_n(\mathbf{k}_n)}(\mathbf{X}) - \mathbb{E}[f_{n,B_n(\mathbf{k}_n)}(\mathbf{X})])$$

and for $t \in \mathbb{R}$, $\phi_n(t) = \mathbb{E}e^{it\sum_{\mathbf{k}_n\in\mathcal{E}_n} X_n(\mathbf{k}_n)}$. Further, let $\{X'_n(\mathbf{k}_n)\}_{\mathbf{k}_n\in\mathcal{E}_n}$ be a sequence of mutually independent random variables such that for $n \in \mathbb{N}$ and $\mathbf{k}_n \in \mathcal{E}_n$, $X'_n(\mathbf{k}_n)$ has the same distribution as $X_n(\mathbf{k}_n)$. Finally, we write ϕ'_n the characteristic function of $\sum_{\mathbf{k}_n\in\mathcal{E}_n} X'_n(\mathbf{k}_n)$. For $t \in \mathbb{R}$, we have

$$\begin{aligned} \phi_n(t) - \phi'_n(t) &= \mathbb{E}\Big(\prod_{\mathbf{k}_n\in\mathcal{E}_n} e^{itX_n(\mathbf{k}_n)}\Big) - \prod_{\mathbf{k}_n\in\mathcal{E}_n} \mathbb{E}(e^{itX'_n(\mathbf{k}_n)}) \\ &= \mathbb{E}\Big(\prod_{\mathbf{k}_n\in\mathcal{E}_n} e^{itX_n(\mathbf{k}_n)}\Big) - \prod_{\mathbf{k}_n\in\mathcal{E}_n} \mathbb{E}(e^{itX_n(\mathbf{k}_n)}). \end{aligned} \tag{S3.1}$$

Then, denoting $\mathbf{j}_1, \dots, \mathbf{j}_{|\mathcal{E}_n|}$ the elements of $\mathcal{E}_n$

$$\begin{aligned} \phi_n(t) - \phi'_n(t) = \mathbb{E}\Big(\prod_{k=1}^{|\mathcal{E}_n|} e^{itX_n(\mathbf{j}_k)}\Big) &- \mathbb{E}(e^{itX_n(\mathbf{j}_1)})\mathbb{E}\Big(\prod_{k=2}^{|\mathcal{E}_n|} e^{itX_n(\mathbf{j}_k)}\Big) \\ &+ \mathbb{E}(e^{itX_n(\mathbf{j}_1)})\Big[\mathbb{E}\Big(\prod_{k=2}^{|\mathcal{E}_n|} e^{itX_n(\mathbf{j}_k)}\Big) - \prod_{k=2}^{|\mathcal{E}_n|} \mathbb{E}(e^{itX_n(\mathbf{j}_k)})\Big]. \end{aligned} \tag{S3.2}$$

Since the term inside the bracket in (S3.2) is as (S3.1) up to the index of the product, we may repeat the operation done between (S3.1) and (S3.2) $|\mathcal{E}_n| - 2$ times. Hence

$$\begin{aligned} &\phi_n(t) - \phi'_n(t) \\ &= \sum_{s=1}^{|\mathcal{E}_n|-1} \Big[\mathbb{E}\Big(\prod_{k=s}^{|\mathcal{E}_n|} e^{itX_n(\mathbf{j}_k)}\Big) - \mathbb{E}(e^{itX_n(\mathbf{j}_s)})\mathbb{E}\Big(\prod_{k=s+1}^{|\mathcal{E}_n|} e^{itX_n(\mathbf{j}_k)}\Big)\Big] \prod_{k=1}^{s-1} \mathbb{E}(e^{itX_n(\mathbf{j}_k)}) \end{aligned}$$

with the convention that a product equals 1 if it runs over the null set. Thus, it follows that

$$\begin{aligned} |\phi_n(t) - \phi'_n(t)| &\le \sum_{s=1}^{|\mathcal{E}_n|-1} \Big|\mathbb{E}\Big(\prod_{k=s}^{|\mathcal{E}_n|} e^{itX_n(\mathbf{j}_k)}\Big) - \mathbb{E}(e^{itX_n(\mathbf{j}_s)})\mathbb{E}\Big(\prod_{k=s+1}^{|\mathcal{E}_n|} e^{itX_n(\mathbf{j}_k)}\Big)\Big| \\ &\le \sum_{s=1}^{|\mathcal{E}_n|-1} \Big|\mathrm{Cov}\Big(e^{itX_n(\mathbf{j}_s)}, \prod_{k=s+1}^{|\mathcal{E}_n|} e^{itX_n(\mathbf{j}_k)}\Big)\Big|. \end{aligned}$$

Then, by Doukhan (1994, Lemma 3, p.10) and (S1.1),

$$|\phi_n(t) - \phi_n'(t)| \leq 4 \sum_{s=1}^{|\mathcal{E}_n|-1} \alpha^{\mathbf{X}}_{w_n,(|\mathcal{E}_n|-s)w_n}(2n^\beta) \leq 4 \sum_{s=1}^{|\mathcal{E}_n|-1} \alpha^{\mathbf{X}}_{(|\mathcal{E}_n|-s)w_n,(|\mathcal{E}_n|-s)w_n}(2n^\beta).$$

Hence,

$$|\phi_n(t) - \phi_n'(t)| \leq 4|\mathcal{E}_n|^2 w_n \sup_{m\geq 0} \frac{\alpha^{\mathbf{X}}_{m,m}(2n^\beta)}{m}. \tag{S3.3}$$

By $(\mathcal{C}1)$, there exists $c > 0$ such that $|\mathcal{E}_n| \leq cn^{d(1-\gamma)}$ so by $(\mathcal{C}2)$ and (S3.3),

$$|\phi_n(t) - \phi_n'(t)| \leq \frac{4c^2}{2^\epsilon} n^{2d-d\gamma-\beta(d+\epsilon)} \leq \frac{4c^2}{2^\epsilon} n^{2d-\beta(2d+\epsilon)} \tag{S3.4}$$

which tends to 0 by $(\mathcal{C}2)$. We verify now (27.16) in Billingsley (1995) to apply Lyapunov's central limit theorem on $\sum_{l\in\mathcal{E}_n} X_n'(\mathbf{k}_n)$. By $(\mathcal{C}3)$ we have

$$\sum_{\mathbf{k}_n\in\mathcal{E}_n} \mathbb{E}|X_n'(\mathbf{k}_n) - \mathbb{E}X_n'(\mathbf{k}_n)|^{2+\tau} = O\big((w_n|\mathcal{E}_n|)^{-(1+\frac{\tau}{2})}\big). \tag{S3.5}$$

Further, it follows by $(\mathcal{C}4)$ that

$$\sum_{\mathbf{k}_n\in\mathcal{E}_n} \operatorname{Var} X_n'(\mathbf{k}_n) \xrightarrow[n\to\infty]{} a^T\Sigma a. \tag{S3.6}$$

Hence by (S3.5)–(S3.6),

$$\sum_{\mathbf{k}_n\in\mathcal{E}_n} \frac{\mathbb{E}|X_n'(\mathbf{k}_n) - \mathbb{E}X_n'(\mathbf{k}_n)|^{2+\tau}}{\big(\sum_{\mathbf{k}_n\in\mathcal{E}_n} \operatorname{Var} X_n'(\mathbf{k}_n)\big)^{2+\tau}} = O\big((w_n|\mathcal{E}_n|)^{-(1+\frac{\tau}{2})}\big) \tag{S3.7}$$

which tends to 0 as n goes to infinity. Therefore, using (S3.7) and the independence of $\{X_n'(\mathbf{k}_n)\}_{\mathbf{k}_n\in\mathcal{E}_n}$ for $n \in \mathbb{N}$, we have by Lyapunov's central limit theorem, see Billingsley (1995, Theorem 27.3, p.362), that

$$\phi_n'(t) \xrightarrow[n\to\infty]{} e^{-\frac{1}{2}a^T\Sigma a t}. \tag{S3.8}$$

Finally, Theorem S1.1 is proved by (S3.3), (S3.4), (S3.8) and the Cramér-Wold device, see for instance Billingsley (1995, Theorem 29.4, p.383).

S4 Convergences of (B.5)–(B.7)

Convergence of A_1

Since $|ie^{it\bar{S}_n}| = 1$ and

$$\mathbb{E}\Big(1 - \frac{1}{a_n} \sum_{\substack{\mathbf{i},\mathbf{j}\in\mathcal{D}_n \\ \mathrm{d}(\mathbf{i},\mathbf{j})\leq m_n}} Y_n(\mathbf{i})Y_n(\mathbf{j})\Big) = 0,$$

we have

$$\mathbb{E}|A_1|^2 = \frac{t^2}{a_n^2}\operatorname{Var}\Big(\sum_{\substack{\mathbf{i},\mathbf{j}\in\mathcal{D}_n\\ \mathrm{d}(\mathbf{i},\mathbf{j})\le m_n}} Y_n(\mathbf{i})Y_n(\mathbf{j})\Big)$$

$$= \frac{t^2}{a_n^2}\sum_{\mathcal{I}_n}\operatorname{Cov}\big(Y_n(\mathbf{i})Y_n(\mathbf{j}), Y_n(\mathbf{i}')Y_n(\mathbf{j}')\big), \tag{S4.1}$$

where $\mathcal{I}_n = \{\mathbf{i},\mathbf{i}',\mathbf{j},\mathbf{j}' \in \mathcal{D}_n : \mathrm{d}(\mathbf{i},\mathbf{j}) \le m_n, \mathrm{d}(\mathbf{i}',\mathbf{j}') \le m_n\}$. We denote by B_1 the terms in the last sum verifying $\mathrm{d}(\mathbf{j},\mathbf{j}') \ge 3m_n$ and by B_2 the others. If $\mathrm{d}(\mathbf{i},\mathbf{j}) \le m_n$, $\mathrm{d}(\mathbf{i}',\mathbf{j}') \le m_n$, and $\mathrm{d}(\mathbf{j},\mathbf{j}') \ge 3m_n$ then

$$\min\{\mathrm{d}(\mathbf{i},\mathbf{j}'),\mathrm{d}(\mathbf{i},\mathbf{i}'),\mathrm{d}(\mathbf{j},\mathbf{j}'),\mathrm{d}(\mathbf{j},\mathbf{i}')\} \ge \mathrm{d}(\mathbf{j},\mathbf{j}') - 2m_n$$

so by Doukhan (1994, Lemma 3, p.10),

$$\operatorname{Cov}\big(Y_n(\mathbf{i})Y_n(\mathbf{j}), Y_n(\mathbf{i}')Y_n(\mathbf{j}')\big) \le 4\alpha_{2,2}^{Z_n}(\mathrm{d}(\mathbf{j},\mathbf{j}') - 2m_n).$$

Let $m_n' = \lceil m_n/s_n\rceil$ be the smallest integer greater than m_n/s_n. Assuming that $\mathrm{d}(\mathbf{i},\mathbf{j}) \le m_n$ and $\mathrm{d}(\mathbf{i}',\mathbf{j}') \le m_n$, for $\mathbf{j},\mathbf{j}' \in \mathcal{D}_n$, there are at most $(2m_n'+1)^d$ choices for $\mathbf{i}$ and the same for $\mathbf{i}'$ when $\mathbf{j}$ respectively $\mathbf{j}'$ is given. Thus from (S4.1) and the last equation,

$$B_1 \le 4\frac{t^2}{a_n^2}|\mathcal{D}_n|(2m_n'+1)^{2d}\sup_{\mathbf{j}\in\mathcal{D}_n}\sum_{\substack{\mathbf{j}'\in\mathcal{D}_n\\ \mathrm{d}(\mathbf{j},\mathbf{j}')\ge 3m_n}} \alpha_{2,2}^{Z_n}(\mathrm{d}(\mathbf{j},\mathbf{j}') - 2m_n).$$

Then by Lemma F.4,

$$B_1 \le 4\frac{3^d t^2}{a_n^2}|\mathcal{D}_n|(2m_n'+1)^{2d}\sup_{\mathbf{j}\in\mathcal{D}_n}\sum_{\substack{r\in s_n\mathbb{Z}\\ r\ge 3m_n}} (r/s_n)^{d-1}\alpha_{2,2}^{Z_n}(r-2m_n),$$

and by Lemma F.2

$$B_1 \le 4\frac{3^d t^2}{a_n^2}|\mathcal{D}_n|(2m_n'+1)^{2d}\sum_{\substack{r\in s_n\mathbb{Z}\\ r\ge 3m_n}} (r/s_n)^{d-1}\alpha_{2v_n,2v_n}^{\mathbf{X}}(r-2m_n-s_n-2R).$$

Invoking $(\mathcal{H}2)$, and noting that $r-2m_n-s_n-2R > 0$ for n large enough when $r \ge 3m_n$,

$$B_1 \le 4\frac{3^d t^2}{a_n^2}|\mathcal{D}_n|(2m_n'+1)^{2d}\sum_{\substack{r\in\mathbb{Z}\\ r\ge 3m_n}} (r/s_n)^{d-1}(r-2m_n-s_n-2R)^{-d-\epsilon}$$

$$= 4\frac{3^d t^2}{a_n^2}|\mathcal{D}_n|(2m_n'+1)^{2d}\sum_{\substack{r\in s_n\mathbb{Z}\\ r\ge 3m_n}} (r/s_n)^{d-1}r^{-d-\epsilon}\left(1-\frac{2m_n+s_n+2R}{r}\right)^{-d-\epsilon}$$

$$= 4\frac{3^d t^2|\mathcal{D}_n|^2}{a_n^2}|\mathcal{D}_n|^{-1}\left(2m_n'+1\right)^{2d}\sum_{\substack{z\in\mathbb{Z}\\ z\ge 3m_n/s_n}} z^{-1-\epsilon}s_n^{-d-\epsilon}\left(1-\frac{2m_n+s_n+2R}{s_n z}\right)^{-d-\epsilon}.$$

When $s_n z \geq 3m_n$, $1-(2m_n+s_n+2R)/(s_n z) \geq (1-(2m_n+s_n+2R)/(3m_n))$ where $1-(2m_n+s_n+2R)/(3m_n)$ converges to 1/3 by (B.1). Also, s_n is increasing. Thus the infinite sum is finite. Moreover, by Lemma F.5 and ($\mathcal{H}4$), $|\mathcal{D}_n|/a_n$ is uniformly bounded with respect to n. Hence, up to a constant, the above expression is bounded by $|\mathcal{D}_n|^{-1}(2m'_n+1)^{2d}$ which by (B.3) converges to 0 so

$$B_1 \xrightarrow[n\to\infty]{} 0. \tag{S4.2}$$

Turning to

$$B_2 = \frac{t^2}{a_n^2}\sum_{\mathcal{I}'_n} \operatorname{Cov}\big(Y_n(\mathbf{i})Y_n(\mathbf{j}), Y_n(\mathbf{i}')Y_n(\mathbf{j}')\big).$$

where $\mathcal{I}'_n = \{\mathbf{i},\mathbf{i}',\mathbf{j},\mathbf{j}' \in \mathcal{D}_n : \mathrm{d}(\mathbf{i},\mathbf{j}) \leq m_n, \mathrm{d}(\mathbf{i}',\mathbf{j}') \leq m_n, \mathrm{d}(\mathbf{j},\mathbf{j}') < 3m_n\}$. Let

$$h = \min\{\mathrm{d}(\mathbf{i},\mathbf{j}'), \mathrm{d}(\mathbf{i},\mathbf{i}'), \mathrm{d}(\mathbf{j},\mathbf{j}'), \mathrm{d}(\mathbf{j},\mathbf{i}')\}.$$

Then by Doukhan (1994, Lemma 3, p. 10),

$$\operatorname{Cov}\big(Y_n(\mathbf{i})Y_n(\mathbf{j}), Y_n(\mathbf{i}')Y_n(\mathbf{j}')\big) \leq 4\alpha_{2,2}^{Z_n}(h). \tag{S4.3}$$

We now bound B_2 by the sum of four terms $B_{2,k}$, $k=1,\dots,4$ according to whether $h=\mathrm{d}(\mathbf{i},\mathbf{j}')$, $h=\mathrm{d}(\mathbf{i},\mathbf{i}')$, $h=\mathrm{d}(\mathbf{j},\mathbf{j}')$, or $h=\mathrm{d}(\mathbf{j},\mathbf{i}')$ and also apply (S4.3). Thus,

$$B_{2,1} = \frac{t^2}{a_n^2}\sum_{\mathcal{I}'_n} 4\alpha_{2,2}^{Z_n}\big(\mathrm{d}(\mathbf{i},\mathbf{j}')\big)$$

and similarly for $B_{2,k}$, $k=2,3,4$.

Given $\mathbf{j}' \in \mathcal{D}_n$, if $\mathrm{d}(\mathbf{j},\mathbf{j}') < 3m_n$, there are at most $(6m'_n+1)^d$ choices for $\mathbf{j}$, and given $\mathbf{j}' \in \mathcal{D}_n$, there are at most $(2m'_n+1)^d$ choices for $\mathbf{i}'$ if $\mathrm{d}(\mathbf{i}',\mathbf{j}') \leq m_n$. Also $\mathrm{d}(\mathbf{i},\mathbf{j}') \leq 4m_n$ if both $\mathrm{d}(\mathbf{j},\mathbf{j}') < 3m_n$ and $\mathrm{d}(\mathbf{i}',\mathbf{j}') \leq m_n$. Thus,

$$B_{2,1} \leq 4\frac{t^2}{a_n^2}|\mathcal{D}_n|(6m'_n+1)^{2d}\sum_{\substack{\mathbf{i}\in\mathcal{D}_n\\ \mathrm{d}(\mathbf{i},\mathbf{j}')<4m_n}} \alpha_{2,2}^{Z_n}\big(\mathrm{d}(\mathbf{i},\mathbf{j}')\big).$$

Then, by Lemma F.4 and Lemma F.2

$$B_{2,1} \leq 4\frac{3^d t^2}{a_n^2}|\mathcal{D}_n|(6m'_n+1)^{2d}\sum_{\substack{r\in s_n\mathbb{Z}\\ 0<r\leq 4m_n}} (r/s_n)^{d-1}\alpha_{2v_n,2v_n}^{\mathbf{X}}(r-s_n-2R).$$

Following the approach for B_1, and noting that $zs_n - s_n - 2R > 0$ for $z>0$ and n large enough, it follows by ($\mathcal{H}2$) that

$$\begin{aligned} B_{2,1} \leq & 4\frac{3^d t^2}{a_n^2}|\mathcal{D}_n|(6m'_n+1)^{2d}\sum_{\substack{z\in\mathbb{Z}\\ 0<z\leq 4m_n/s_n}} z^{d-1}(zs_n - s_n - 2R)^{-d-\epsilon} \\ = & 4\frac{3^d t^2}{a_n^2}|\mathcal{D}_n|(6m'_n+1)^{2d}\sum_{\substack{z\in\mathbb{Z}\\ 0<z\leq 4m_n/s_n}} z^{-1-\epsilon}\left(s_n - \frac{s_n+2R}{z}\right)^{-d-\epsilon}. \end{aligned}$$

Hence, similarly to (S4.2),

$$B_{2,1} \xrightarrow[n\to\infty]{} 0.$$

By the same reasoning $B_{2,k}$, $k=2,3,4$ also converge to zero so that

$$B_2 \xrightarrow[n\to\infty]{} 0. \tag{S4.4}$$

Therefore, by (S4.1), (S4.2), and (S4.4),

$$\mathbb{E}|A_1| \leq \sqrt{\mathbb{E}|A_1|^2} \xrightarrow[n\to\infty]{} 0. \tag{S4.5}$$

Convergence of A_2

According to the formula of the remainder of Taylor's expansion, there exists a constant c such that

$$|1 - it\bar{S}_n(\mathbf{i}) - e^{-it\bar{S}_n(\mathbf{i})}| \leq ct^2\bar{S}_n^2(\mathbf{i}).$$

Then, since $\sup_{n\in\mathbb{N}}\sup_{\mathbf{i}\in\mathcal{D}_n}|Y_n(\mathbf{i})| \leq 1$ by $(\mathcal{H}_b)$, we have

$$\mathbb{E}|A_2| \leq \frac{1}{\sqrt{a_n}}\sum_{\mathbf{i}\in\mathcal{D}_n}\mathbb{E}|1 - it\bar{S}_n(\mathbf{i}) - e^{-it\bar{S}_n(\mathbf{i})}| \leq \frac{1}{\sqrt{a_n}}|\mathcal{D}_n|\sup_{\mathbf{i}\in\mathcal{D}_n}\mathbb{E}\big(ct^2\bar{S}_n^2(\mathbf{i})\big).$$

Thus, for $\mathcal{J}_n(\mathbf{i}) = \{\mathbf{j},\mathbf{j}' \in \mathcal{D}_n : \mathrm{d}(\mathbf{i},\mathbf{j}) \leq m_n, \mathrm{d}(\mathbf{i},\mathbf{j}') \leq m_n\}$,

$$\mathbb{E}|A_2| \leq \frac{ct^2}{a_n^{\frac{3}{2}}}|\mathcal{D}_n|\sup_{\mathbf{i}\in\mathcal{D}_n}\sum_{\mathcal{J}_n(\mathbf{i})}\mathbb{E}\big(Y_n(\mathbf{j})Y_n(\mathbf{j}')\big)$$

and as the variables Y_n are centred,

$$\mathbb{E}|A_2| \leq \frac{ct^2}{a_n^{\frac{3}{2}}}|\mathcal{D}_n|\sup_{\mathbf{i}\in\mathcal{D}_n}\sum_{\mathcal{J}_n(\mathbf{i})}\mathrm{Cov}\big(Y_n(\mathbf{j}),Y_n(\mathbf{j}')\big).$$

Then, by Doukhan (1994, Lemma 3, p. 10),

$$\mathbb{E}|A_2| \leq \frac{ct^2}{a_n^{\frac{3}{2}}}|\mathcal{D}_n|\sup_{\mathbf{i}\in\mathcal{D}_n}\sum_{\mathcal{J}_n(\mathbf{i})}4\alpha_{1,1}^{Z_n}\big(\mathrm{d}(\mathbf{j},\mathbf{j}')\big).$$

By the triangular inequality, we have in the last sum, $\mathrm{d}(\mathbf{j},\mathbf{j}') \leq 2m_n$. Further, if $\mathrm{d}(\mathbf{i},\mathbf{j}') \leq m_n$, there are at most $(2m_n'+1)^d$ choices possible for $\mathbf{j}'$. Thus,

$$\mathbb{E}|A_2| \leq \frac{4ct^2}{a_n^{\frac{3}{2}}}|\mathcal{D}_n|(2m_n'+1)^d\sup_{\mathbf{i}\in\mathcal{D}_n}\sup_{\mathbf{j}'\in\mathcal{D}_n}\sum_{\substack{\mathbf{j}\in\mathcal{D}_n\\ \mathrm{d}(\mathbf{j},\mathbf{j}')\leq 2m_n}}\alpha_{1,1}^{Z_n}\big(\mathrm{d}(\mathbf{j},\mathbf{j}')\big)$$

and by Lemma F.4

$$\mathbb{E}|A_2| \leq 4\frac{3^d ct^2}{a_n^{\frac{3}{2}}}|\mathcal{D}_n|(2m_n'+1)^d\Big(1 + \sum_{\substack{r\in s_n\mathbb{Z}\\ 0<r\leq 2m_n}}(r/s_n)^{d-1}\alpha_{1,1}^{Z_n}(r)\Big).$$

Then, by Lemma F.2

$$\mathbb{E}|A_2| \le 4\frac{3^d ct^2}{a_n^{\frac{3}{2}}}|\mathcal{D}_n|(2m'_n+1)^d\Big(1+\sum_{\substack{r\in s_n\mathbb{Z}\\ 0<r\le 2m_n}}(r/s_n)^{d-1}\alpha^{\mathbf{X}}_{v_n,v_n}(r-s_n-2R)\Big).$$

Thus, using ($\mathcal{H}2$) we have

$$\begin{aligned}\mathbb{E}|A_2| &\le 4\frac{3^d ct^2}{a_n^{\frac{3}{2}}}|\mathcal{D}_n|(2m'_n+1)^d\Big\{1+\sum_{\substack{z\in\mathbb{Z}\\ 0<z\le 2m_n/s_n}} z^{-1-\epsilon}\left(s_n-\frac{s_n+2R}{z}\right)^{-d-\epsilon}\Big\}\\ &\le 4\frac{3^d ct^2|D_n|\sqrt{|\mathcal{D}_n|}}{a_n a_n^{\frac{1}{2}}}\frac{(2m'_n+1)^d}{\sqrt{|\mathcal{D}_n|}}\Big\{1+\sum_{\substack{z\in\mathbb{Z}\\ 0<z\le 2m_n/s_n}} z^{-1-\epsilon}\left(s_n-\frac{s_n+2R}{z}\right)^{-d-\epsilon}\Big\}.\end{aligned}$$

Hence, by ($\mathcal{H}4$), Lemma F.5, and (B.3) it follows that

$$\mathbb{E}|A_2| \xrightarrow[n\to\infty]{} 0. \tag{S4.6}$$

Convergence of A_3

We have by Doukhan (1994, Lemma 3, p. 10)

$$\begin{aligned}|\mathbb{E}A_3| &\le \frac{1}{\sqrt{a_n}}\sum_{\mathbf{i}\in\mathcal{D}_n}\left|\mathrm{Cov}(Y_n(\mathbf{i}), e^{it(\bar{S}_n-\bar{S}_n(\mathbf{i}))})\right|\\ &\le \frac{4|\mathcal{D}_n|}{\sqrt{a_n}}\alpha^{Z_n}_{1,\infty}(m_n).\end{aligned}$$

Then by Lemma F.2

$$\begin{aligned}|\mathbb{E}A_3| &\le \frac{4|\mathcal{D}_n|}{\sqrt{a_n}}\alpha^{\mathbf{X}}_{v_n,\infty}(m_n-s_n-2R)\\ &= \frac{4\sqrt{|\mathcal{D}_n|}}{\sqrt{a_n}}\sqrt{|D_n|}\alpha^{\mathbf{X}}_{v_n,\infty}(m_n-s_n-2R).\end{aligned} \tag{S4.7}$$

Thus by Lemma F.5, assumption ($\mathcal{H}4$), and (B.2), $|\mathbb{E}A_3|$ tends to zero.

S5 Proofs of lemmas

S5.1 Proof of Lemma F.1

For $x\in C_n^{\oplus R}(\mathbf{l})$ and $y\in C_n^{\oplus R}(\mathbf{k})$,

$$\mathrm{d}(x,y)=\max\{|x_1-y_1|,|x_2-y_2|\dots,|x_d-y_d|\}$$

and by (3.2),

$$|l_i-k_i|-s_n-2R\le|x_i-y_i|\le|l_i-k_i|+s_n+2R.$$

Therefore,

$$\mathrm{d}(\mathbf{l},\mathbf{k})-s_n-2R\le\mathrm{d}(x,y)\le\mathrm{d}(\mathbf{l},\mathbf{k})+s_n+2R$$

so we have

$$\mathrm{d}(\mathbf{l},\mathbf{k})-s_n-2R\le\mathrm{d}\big(C_n^{\oplus R}(\mathbf{l}),C_n^{\oplus R}(\mathbf{k})\big)\le\mathrm{d}(\mathbf{l},\mathbf{k})+s_n+2R.$$

S5.2 Proof of Lemma F.2

Let $I_1 \subset s_n\mathbb{Z}^d$ and $I_2 \subset s_n\mathbb{Z}^d$ be such that $|I_1| \leq c_1$, $|I_2| \leq c_2$ and $\mathrm{d}(I_1, I_2) \geq r$. Then,

$$\alpha\Big(\sigma((Z_n(\mathbf{l}))_{\mathbf{l}\in I_1}), \sigma((Z_n(\mathbf{k}))_{\mathbf{k}\in I_2})\Big) \leq \alpha\Big(\mathbf{X} \cap \bigcup_{\mathbf{l}\in I_1} C_n^{\oplus R}(\mathbf{l}), \mathbf{X} \cap \bigcup_{\mathbf{k}\in I_2} C_n^{\oplus R}(\mathbf{k})\Big). \quad \text{(S5.1)}$$

Further, since $|I_1| \leq c_1$, $|I_2| \leq c_2$,

$$\Big|\bigcup_{\mathbf{l}\in I_1} C_n^{\oplus R}(\mathbf{l})\Big| \leq c_1 v_n, \quad \text{and} \quad \Big|\bigcup_{\mathbf{k}\in I_2} C_n^{\oplus R}(\mathbf{k})\Big| \leq c_2 v_n,$$

and as $\mathrm{d}(I_1, I_2) \geq r$, we have by Lemma F.1,

$$\begin{aligned}
\mathrm{d}\Big(\bigcup_{\mathbf{l}\in I_1} C^{\oplus R}(\mathbf{l}), \bigcup_{\mathbf{k}\in I_2} C^{\oplus R}(\mathbf{k})\Big) &= \inf_{\mathbf{l}\in I_1, \mathbf{k}\in I_2} \mathrm{d}(C^{\oplus R}(\mathbf{l}), C^{\oplus R}(\mathbf{k})) \\
&\geq \inf_{\mathbf{l}\in I_1, \mathbf{k}\in I_2} \mathrm{d}(\mathbf{l}, \mathbf{k}) - s_n - 2R \\
&\geq r - s_n - 2R.
\end{aligned}$$

It follows by (S5.1) that for $\mathbf{l} \in I_1$, $\mathbf{k} \in I_2$,

$$\alpha\Big(\sigma((Z_n(\mathbf{l}))_{\mathbf{l}\in I_1}), \sigma((Z_n(\mathbf{k}))_{\mathbf{k}\in I_2})\Big) \leq \alpha^{\mathbf{X}}_{c_1 v_n, c_2 v_n}(r - s_n - 2R)$$

which concludes the proof.

S5.3 Proof of Lemma F.3

By Lemma F.2, for $c_1, c_2, r \geq 0$ we have

$$\alpha^{Z_n}_{c_1,c_2}(rs_n) \leq \alpha^{\mathbf{X}}_{c_1 v_n, c_2 v_n}(rs_n - s_n - 2R).$$

Thus, for τ as in $(\mathcal{H}3)$, we have by $(\mathcal{H}2)$, $\sup_{n\in\mathbb{N}} \sum_{r=1}^{\infty} r^{d-1} \alpha^{Z_n}_{1,1}(rs_n)^{\frac{\tau}{2+\tau}} < \infty$. Using this last result and $(\mathcal{H}3)$, we may apply Theorem 1 in Fazekas et al. (2000) which states that

$$\mathbb{E}\Big|\sum_{\mathbf{l}\in\mathcal{D}_n} Z_n(\mathbf{l})\Big|^2 \leq \Big(1 + 8\sum_{r=1}^{\infty} \alpha^{Z_n}_{1,1}(rs_n)^{\frac{\tau}{2+\tau}} r^{d-1}\Big) \sum_{\mathbf{l}\in\mathcal{D}_n} (\mathbb{E}|Z_n(\mathbf{l})|^{2+\tau})^{\frac{2}{2+\tau}}.$$

Hence, by $(\mathcal{H}3)$,

$$\sup_{a\in\mathbb{R}^q} \sup_{n\in\mathbb{N}} \frac{a^T \operatorname{Var}\left(\sum_{\mathbf{l}\in\mathcal{D}_n} Z_n(\mathbf{l})\right) a}{|\mathcal{D}_n||a|^2} < \infty \quad \text{(S5.2)}$$

which implies that $\limsup_{n\to\infty} \lambda_{\max}\Big(\operatorname{Var}\left(\sum_{\mathbf{l}\in\mathcal{D}_n} Z_n(\mathbf{l})\right)/|\mathcal{D}_n|\Big) < \infty$.

S5.4 Proof of Lemma F.4

Without loss of generality, we let $\mathbf{i} = \mathbf{0}$. Then,

$$
\begin{aligned}
&|\{\mathbf{j} \in s_n\mathbb{Z}^d : \mathrm{d}(0,\mathbf{j}) = s_n k\}| \\
&\qquad = |\{\mathbf{j} \in s_n\mathbb{Z}^d : \mathrm{d}(0,\mathbf{j}) \le s_n k\}| - |\{\mathbf{j} \in s_n\mathbb{Z}^d : \mathrm{d}(0,\mathbf{j}) \le s_n(k-1)\}|.
\end{aligned}
$$

When $\mathrm{d}(\mathbf{0},\mathbf{j}) \le s_n k$, each coordinate of $\mathbf{j}$ may take $2k+1$ values, so

$$
|\{\mathbf{j} \in s_n\mathbb{Z}^d : \mathrm{d}(0,\mathbf{j}) = s_n k\}| = (2k+1)^d - (2k-1)^d.
$$

Further, by the binomial theorem

$$
\begin{aligned}
|\{\mathbf{j} \in s_n\mathbb{Z}^d : \mathrm{d}(0,\mathbf{j}) = s_n k\}| &= \sum_{i=1}^{d} \binom{d}{i} (2k)^{d-i} \big(1^i - (-1)^i\big) \\
&\le k^{d-1} \sum_{i=1}^{d} \binom{d}{i} 2^{d-i} \big(1^i - (-1)^i\big) \\
&\le 3^d k^{d-1}.
\end{aligned}
$$

S5.5 Proof of Lemma F.5

Since $\sigma_n^2 = \mathrm{Var}\big(\sum_{\mathbf{l}\in\mathcal{D}_n} Z_n(\mathbf{l})\big) = \mathrm{Var}\big(\sum_{\mathbf{l}\in\mathcal{D}_n} Y_n(\mathbf{l})\big)$, we have

$$
\sigma_n^2 = a_n + \sum_{\substack{\mathbf{i},\mathbf{j}\in\mathcal{D}_n \\ \mathrm{d}(\mathbf{i},\mathbf{j}) > m_n}} \mathbb{E}[Y_n(\mathbf{j})Y_n(\mathbf{i})].
$$

Thus,

$$
\begin{aligned}
|\sigma_n^2 - a_n| &\le \sum_{\substack{\mathbf{i},\mathbf{j}\in\mathcal{D}_n \\ \mathrm{d}(\mathbf{i},\mathbf{j}) > m_n}} \big|\mathbb{E}[Y_n(\mathbf{j})Y_n(\mathbf{i})]\big| \\
&\le \sum_{\substack{\mathbf{i},\mathbf{j}\in\mathcal{D}_n \\ \mathrm{d}(\mathbf{i},\mathbf{j}) > m_n}} \big|\mathrm{Cov}\big(Z_n(\mathbf{j}), Z_n(\mathbf{i})\big)\big|.
\end{aligned}
$$

Then, by $(\mathcal{H}_b)$ and Doukhan (1994, Lemma 3, p. 10),

$$
|\sigma_n^2 - a_n| \le 4 \sum_{\substack{\mathbf{i},\mathbf{j}\in\mathcal{D}_n \\ \mathrm{d}(\mathbf{i},\mathbf{j}) > m_n}} \alpha_{1,1}^{Z_n}\big(\mathrm{d}(\mathbf{i},\mathbf{j})\big).
$$

Thus, by Lemma F.4,

$$
|\sigma_n^2 - a_n| \le 4(3^d)|\mathcal{D}_n| \sum_{\substack{r\in s_n\mathbb{Z} \\ r > m_n}} (r/s_n)^{d-1} \alpha_{1,1}^{Z}(r).
$$

Using Lemma F.2, $(\mathcal{H}2)$, and noting that $r - s_n - 2R > 0$ for n large enough when $r > m_n$,

$$
|\sigma_n^2 - a_n| \le 4(3^d)|\mathcal{D}_n| \sum_{\substack{r\in s_n\mathbb{Z} \\ r > m_n}} (r/s_n)^{d-1} (r - s_n - 2R)^{-d-\epsilon}.
$$

Thus

$$\left|1 - \frac{a_n}{\sigma_n^2}\right| \leq \frac{4(3^d)|\mathcal{D}_n|}{\sigma_n^2} \sum_{\substack{z \in \mathbb{Z} \\ z > m_n/s_n}} z^{-1-\epsilon} \left(s_n - \frac{s_n + 2R}{z}\right)^{-d-\epsilon}. \tag{S5.3}$$

By $(\mathcal{H}4)$, $|\mathcal{D}_n|/\sigma_n^2$ is bounded. Further, since m_n/s_n tends to infinity, the infinite sum tends to zero. Thus we obtain

$$\left|1 - \frac{a_n}{\sigma_n^2}\right| \xrightarrow[n \to \infty]{} 0.$$