Aalborg Universitet

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

Biscio, Christophe; Waagepetersen, Rasmus

Published in:
Scandinavian Journal of Statistics

DOI (link to publication from Publisher):
10.1111/sjos.12389

Publication date:
2019

Document Version
Accepted author manuscript, peer reviewed version

Citation for published version (APA):
Biscio, C., & Waagepetersen, R. (2019). A general central limit theorem and a subsampling variance estimator for mixing point processes. *Scandinavian Journal of Statistics*, *46*(4), 1168-1190. https://doi.org/10.1111/sjos.12389

Received: 20 March 2018 | Revised: 25 October 2018 | Accepted: 12 January 2019

DOI: 10.1111/sjos.12389

ORIGINAL ARTICLE

Scandinavian Journal of Statistics

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

Christophe Ange Napoléon Biscio | Rasmus Waagepetersen

Department of Mathematical Sciences, Aalborg University, Aalborg, Denmark

Correspondence
Christophe Ange Napoléon Biscio, Department of Mathematical Sciences, Aalborg University, 9220 Aalborg, Denmark.
Email: christophe@math.aau.dk

Funding information
The Danish Council for Independent Research | Natural Sciences, Grant/Award Number: DFF 7014-00074; Villum Foundation, Grant/Award Number: 8721

Abstract

We establish a central limit theorem for multivariate summary statistics of nonstationary 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, for example, 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, marked point process, spatial point process, subsampling

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

$$T_W(\mathbf{X}) = \sum_{u_1,\ldots,u_p \in \mathbf{X}\cap W}^{\neq} h(u_1,\ldots,u_p), \tag{1}$$

where $h : \mathbb{R}^{dp} \to \mathbb{R}^q$, $p, q \geq 1$, 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 such as Guan and Sherman (2007); Guan and Loh (2007); Prokešová and Jensen (2013); Guan, Jalilian, and Waagepetersen (2015); and Xu, Waagepetersen, and Guan (2018). The other approach is due to Bolthausen (1982) who considered stationary random fields and whose proof was later generalised to nonstationary random fields by Guyon (1995) and Karácsony (2006). This approach is, for example, used in Waagepetersen and Guan (2009), Coeurjolly and Møller (2014), Biscio and Coeurjolly (2016), Coeurjolly (2017), and Poinas, Delyon, and Lavancier (2017). Regarding the point process references mentioned above, 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 nonstationary 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, for example, 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 centred random fields (Coeurjolly & Lavancier, 2017; Jensen & Künsch, 1994); however, 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) 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

$$T_W(\mathbf{X}) = \sum_{\mathbf{l}\in\mathbb{L}} f_{\mathbf{l},W}(\mathbf{X}) \tag{2}$$

with

$$f_{\mathbf{l},W}(\mathbf{X}) = \sum_{u_1\in\mathbf{X}\cap C(\mathbf{l})\cap W} \quad \sum_{u_2,\dots,u_p\in(\mathbf{X}\cap W)\setminus\{u_1\}}^{\neq} h(u_1,\dots,u_p). \tag{3}$$

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 setup provided h satisfies a certain finite range condition; see the following sections for details. The finite range condition is satisfied for the majority of summary statistics considered in spatial statistics and, hence, does not seem restrictive in practical applications. In connection to the Bolthausen approach, we remark that Guyon (1995) does not cover the case where the function f in (2) depends on the observation window. This kind of generalisation is, for example, 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 nonparametric kernel estimators depending on a bandwidth 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, for example, Politis and Romano (1994) and Lahiri (2003). For statistics of point processes, these methods have been considered in, for example, 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 (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 nonstationary 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 5. The application of our subsampling estimator to estimating functions is discussed in Section 6 and is illustrated in a simulation study in Section 7. Finally, our subsampling estimator is discussed in relation to other approaches in Section 8. The proofs of our results are presented in the Appendix. The Appendix contains the proofs of Theorems 1 and 2. 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 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 $d \in \mathbb{N} = \{1, 2, \dots\}$, 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, 2008) for measure theoretical details. We define a lattice $\mathbb{L}$ as a countable subset of $\mathbb{Z}^d$ where $\mathbb{Z} = \mathbb{N} \cup \{0, -1, -2, \dots\}$. When considering vertices of a lattice, we use bold letter, for instance, $\mathbf{i} \in \mathbb{Z}^d$. We define

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

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 $A \subset \mathbb{R}^d$, 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 $R \ge 0$, we define

$$A \oplus R = \left\{ x \in \mathbb{R}^d : \inf_{y \in A} \mathrm{d}(x, y) \le R \right\}. \tag{4}$$

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

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

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

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

and

$$\alpha^{\mathbf{X}}_{c_1,c_2}(m) = \sup\{\alpha(\sigma(\mathbf{X}\cap E_1), \sigma(\mathbf{X}\cap E_2)) : E_1 \subset \mathbb{R}^d, E_2 \subset \mathbb{R}^d, |E_1| \leq c_1, |E_2| \leq c_2, \mathrm{d}(E_1, E_2) \geq m\Big\}. \tag{5}$$

Note that the definition of $\alpha^{\mathbf{X}}_{c_1,c_2}$ differs from the usual definition in spatial statistics (see, e.g., Waagepetersen & 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 because 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

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

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

$$T_{W_n}(\mathbf{X}) = \sum_{\mathbf{l}\in\mathcal{D}_n(W_n)} f_{n,\mathbf{l},W_n}(\mathbf{X}), \tag{6}$$

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 $\mathbf{X}\cap W_n \cap C_n^{\oplus R}(\mathbf{l})$ for some $R \geq 0$, where $C_n(\mathbf{l})$ is a hyper cube of side length $s_n > 0$,

$$C_n(\mathbf{l}) = \prod_{j=1}^{d} \left(l_j - s_n/2, l_j + s_n/2\right], \quad \mathbf{l} \in s_n\mathbb{Z}^d, \tag{7}$$

and $C_n^{\oplus R}(\mathbf{l}) = C_n(\mathbf{l}) \oplus R$; see (4). 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 $v_n = |C_n^{\oplus R}(\mathbf{l})|$ the common volume of the $C_n^{\oplus R}(\mathbf{l})$, and $\mathcal{D}_n(A)$ is defined for any $A \subset \mathbb{R}^d$ by

$$\mathcal{D}_n(A) = \left\{\mathbf{l} \in s_n\mathbb{Z}^d : C_n(\mathbf{l}) \cap A \neq \emptyset\right\}. \tag{8}$$

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 $\eta > 0$, $|\mathcal{D}_n| = O(|W_n|/s_n^d)$. Furthermore, 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 $0 < \liminf_{n\to\infty}\lambda_{\min}\left(\frac{\Sigma_n}{|\mathcal{D}_n|}\right)$, where $\Sigma_n = \operatorname{Var} T_{W_n}(\mathbf{X})$ and $\lambda_{\min}(M)$ denotes the smallest eigenvalue of a symmetric matrix M.

We then obtain the following theorem.

Theorem 1. *Let $\{T_{W_n}(\mathbf{X})\}_{n\in\mathbb{N}}$ be a sequence of q-dimensional statistics of the form (6). If (H1)–(H4) hold, then we have the convergence*

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

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

Remark 1. In the case where $f_{n,\mathbf{l},W_n}$ is defined in terms of a function h as in (3), we need h to satisfy a finite range condition, that is, for $u_1,\dots,u_p\in\mathbb{R}^d$, $h(u_1,\dots,u_p) = 0$ if $\mathrm{d}(u_i,u_j) > R$ for some $1 \le i < j \le p$.

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

Remark 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 & 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 4. For a point process, moments are calculated using the 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 5. As presented in Section 1, the convergence in Theorem 1 can be proved under different assumptions using Bernstein's blocking technique. However, as explained in Section 2, 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 | EXTENSION TO THE CASE OF MARKED SPATIAL POINT PROCESSES

By suitable modification of the α-mixing coefficient (5), Theorem 1 can be straightforwardly extended to the case of a marked point process. In this case, $\mathbf{X}$ is a point process on $\mathbb{R}^d\times M$ for some mark space M. A typical example is $M = \{1,\dots,L\}$, $L > 1$, in which case $\mathbf{X}$ is called a multitype or multivariate point process. Another example is $M = [0,\infty)$ so that a mark could describe the size of an object represented by a marked point in $\mathbf{X}$. For sets $A\subset\mathbb{R}^d$ and $B\subset M$, we let

$$\mathbf{X}_{A,B} = X\cap A\times B$$

denote the set of marked points in $\mathbf{X}$ whose "point parts" fall in A and whose marks fall in B. Then, we propose the following α-mixing coefficient for a marked point process $\mathbf{X}$:

$$\alpha^{\mathbf{X}}_{c_1,c_2}(m) = \sup \left\{ \alpha \left(\sigma(\mathbf{X}_{E_1,M}), \sigma(\mathbf{X}_{E_2,M}) \right) : E_1 \subset \mathbb{R}^d, E_2 \subset \mathbb{R}^d, |E_1| \le c_1, |E_2| \le c_2, \mathrm{d}(E_1, E_2) \ge m \right\}. \tag{9}$$

This is a natural extension of the previous α-mixing coefficient (5) because mixing is as before essentially a property related to spatial distance between "point parts" of $\mathbf{X}$ (other types of mixing coefficients in the same spirit are given in Definition 12.3.I in Daley & Vere-Jones, 2008). Regarding the settings in Section 3, we similarly just need to modify the requirement on f_n so that $f_{n,\mathbf{l},W_n}(\mathbf{X})$ depends on $\mathbf{X}$ only through $\mathbf{X}_{W_n \cap C_n^{\oplus R}(\mathbf{l}),M}$. Upon replacing (5) with (9) and modifying the requirement of f_n, the proof of Theorem 1 carries directly over the marked case. The crucial observation here is that the proof only involves the derived random fields Z_n and their mixing properties, which are the same in the marked case as in the unmarked case.

Consider as an example a multivariate Cox process $\mathbf{X}$ driven by a multivariate random intensity function $\Lambda = \{(\Lambda_1(u), \dots, \Lambda_p(u))\}_{u \in \mathbb{R}^d}$. Suppose Λ is m-dependent, meaning that $\Lambda(u)$ and $\Lambda(v)$ are independent whenever $\mathrm{d}(u, v) \ge m$. Then, the part of $(\mathcal{H}2)$ regarding α-mixing trivially holds for any $\epsilon > 0$.

5 | SUBSAMPLING VARIANCE ESTIMATOR

By Theorem 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, that is,

$$\lim_{n \to \infty} \mathrm{P}\left(\mathcal{E}(\mathbf{X}) \le q_{1-\alpha}\right) = 1 - \alpha, \tag{10}$$

where

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

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 $W_1 \subset W_2 \subset \dots$ 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) subrectangles

$$B_{\mathbf{k}_n,\mathbf{t}} = \prod_{j=1}^d \left(t_j - k_{n,j}/2, t_j + k_{n,j}/2 \right), \tag{11}$$

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 (6). We suggest the subsampling estimator

$$\hat{\varsigma}_n = \frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|} \sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \left(\frac{T_{B_{\mathbf{k}_n,\mathbf{t}}}(\mathbf{X})}{\sqrt{|\mathcal{D}_n(B_{\mathbf{k}_n,\mathbf{t}})|}} - \frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|} \sum_{\mathbf{s}\in\mathcal{T}_{\mathbf{k}_n,n}} \frac{T_{B_{\mathbf{k}_n,\mathbf{s}}}(\mathbf{X})}{\sqrt{|\mathcal{D}_n(B_{\mathbf{k}_n,\mathbf{s}})|}} \right)^2. \tag{12}$$

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

- ($\mathcal{S}1$) For $j = 1, \dots, d$, $k_{n,j} < m_{n,j}$. 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 $\epsilon' > 0$, $\sup_{n\in\mathbb{N}} \sup_{\mathbf{l}\in\mathcal{T}_{\mathbf{k}_n,n}} \mathbb{E}\left(|Z_n(\mathbf{l}) - \mathbb{E}(Z_n(\mathbf{l}))|^{4+\epsilon'}\right) < \infty$.
- ($\mathcal{S}3$) 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 $\lim\sup_{n\to\infty} \lambda_{\max}(\Sigma_n) < \infty$, where $\lambda_{\max}(M)$ denotes the maximal eigenvalue of a symmetric matrix M.
- ($\mathcal{S}4$) $|\mathcal{T}_{\mathbf{k}_n,n}|^{-1}\sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \left(\mathbb{E}(T_{B_{\mathbf{k}_n,\mathbf{t}}}(\mathbf{X})) - \mathbb{E}\left(|\mathcal{T}_{\mathbf{k}_n,n}|^{-1}\sum_{\mathbf{s}\in\mathcal{T}_{\mathbf{k}_n,n}} T_{B_{\mathbf{k}_n,\mathbf{s}}}(\mathbf{X})\right)\right)^2 \to 0$ as $n \to \infty$.
- ($\mathcal{S}5$) There exist $c > 0$ and $\delta > 0$ such that $\sup_{p\in\mathbb{N}} \frac{\alpha_{p,p}^{\mathbf{X}}(m)}{p} \leq \frac{c}{m^{d+\delta}}$ and, for v_n as below (7), $v_n \prod_{j=1}^d (2k_{n,j}+1)/(\max_i k_{n,i} - s_n)^{d+\delta}$ converges towards 0 as n tends to infinity.
- ($\mathcal{S}6$) There exist $c, \delta' > 0$ and $\epsilon' > \epsilon > 0$ such that $\alpha_{5v_n,5v_n}^{\mathbf{X}}(r) \leq c r^{-5d\frac{6+\epsilon}{\epsilon}-\delta'}$.

Theorem 2. *Let $\{T_{W_n}(\mathbf{X})\}_{n\in\mathbb{N}}$ be a sequence of q-dimensional statistics of the form (6). Let further $\hat{\varsigma}_n$ be defined as in (12) and assume that ($\mathcal{S}0$)–($\mathcal{S}6$) hold. Then, we have the convergence*

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

For practical application, it is enough to state Theorem 2 with convergence in probability, but the proof is easier when considering mean square convergence. Assumption ($\mathcal{S}1$) ensures that the subrectangles 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 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 subrectangle is not too different from Σ_n. In particular, it holds naturally if there exists a matrix $\tilde{\Sigma}$ such that $\lim_{n\to\infty}|\mathcal{D}_n(A)|^{-1}\mathrm{Var}(T_A(\mathbf{X})) = \tilde{\Sigma}$, $A = W_n$ 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 subrectangles $B_{\mathbf{k}_n,\mathbf{t}}$ do not vary too much. For instance, this assumption is automatically verified if the point process $\mathbf{X}$ is stationary or if the statistics (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 6). Moreover, depending on the statistic (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 nonparametric kernel estimators. We use two different α-mixing conditions ($\mathcal{S}5$)–($\mathcal{S}6$) to apply Theorem 2. Moreover, the decreasing rate in ($\mathcal{S}6$) is restrictive due to the constant $5d$. Hence, mixing conditions are stronger than for Theorem 1. However, in the proof of Theorem 2, Assumption ($\mathcal{S}6$) is used only to verify (C4), which ensures the validity of the assumption (i) of Theorem C1. Depending on the problem considered, (C4) 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 DPP, then (C4) is an immediate consequence of (Biscio & Lavancier, 2016, Proposition 4.2).

Remark 6. By Theorems 1 and 2, we may replace the confidence ellipsoid in (10) by a subsampling confidence ellipsoid $\hat{\mathcal{E}}_n$, that is,

$$\lim_{n\to\infty} \mathrm{P}\left(\hat{\mathcal{E}}_n(\mathbf{X}) \le q_{1-\alpha}\right) = 1-\alpha \tag{13}$$

where

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

Remark 7. 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, that is, the one ensuring the fastest convergence rate in Theorem 2. Concerning that problem, there are several recommendations in the literature; see for instance Lahiri (2003).

Remark 8. Note that the centres 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. Furthermore, similarly to Loh (2010) and Ekström (2008), it is possible to extend Theorem 2 by relaxing the assumption in ($\mathcal{S}0$) that the windows are rectangular.

6 | VARIANCE ESTIMATION FOR ESTIMATING FUNCTIONS

Consider a parametric family of point processes $\{\mathbf{X}_\theta : \theta \in \Theta\}$ for a nonempty subset $\Theta \subset \mathbb{R}^q$, $q \in \mathbb{N}$. 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

$$e_n(\theta) = \sum_{u_1,\dots,u_p \in \mathbf{X}_{\theta_0}\cap W_n}^{\neq} h_\theta(u_1,\dots,u_p) - \int_{W_n^p} h_\theta(u_1,\dots,u_p)\rho_\theta^{(p)}(u_1,\dots,u_p)\mathrm{d}u_1\dots\mathrm{d}u_p, \tag{14}$$

where h_θ is a function from $\mathbb{R}^{dp}$ into $\mathbb{R}^q$ and $\rho_\theta^{(p)}$ denotes the pth order joint intensities of $\mathbf{X}_\theta$. Then, an estimate $\hat{\theta}_n$ 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 $\hat{\theta}_n$ for the case $p = 1$ (see, for instance, Guan et al., 2015; Guan & Shen, 2010; Waagepetersen, 2007). 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

$$|W_n|^{1/2}\left(S_n^{-1}(\theta_0)\frac{\Sigma_{n,\theta_0}}{|W_n|}S_n^{-1}(\theta_0)\right)^{-1/2}(\hat{\theta}_n - \theta_0) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}(0, I_q), \tag{15}$$

where $S_n(\theta_0) = |W_n|^{-1}\mathbb{E}(-\mathrm{d}e_n(\theta_0)/\mathrm{d}\theta^T)$ and $\Sigma_{n,\theta_0} = \mathrm{Var}\, e_n(\theta_0)$. 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 5, 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 5, for $\theta \in \Theta$, we let $T_{W_n,\theta}(\mathbf{X}_{\theta_0}) = e_n(\theta)$ and

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

Furthermore, $\hat{\varsigma}_n(\theta)$ is defined as in (12) 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 $\hat{\varsigma}_n(\hat{\theta}_n)$. The validity of this relies in a standard way on a Taylor expansion

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

where $\|\theta^* - \theta_0\| \leq \|\hat{\theta}_n - \theta_0\|$ 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 $\hat{\varsigma}_n(\hat{\theta}_n)$ to estimate $\Sigma_{n,\theta_0}/|W_n|$.

7 | 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 (15) 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 subrectangles (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 side length 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 centres of the blocks completely included in W_1 are $(0.25,0.25)$; $(0.5,0.25)$; $(0.75,0.25)$; $(0.25,0.5)$, $\ldots$, $(0.75,0.75)$. 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 nonstationary Poisson point process, two different nonstationary log-Gaussian Cox processes (LGCPs), and a nonstationary DPP. For a presentation of these models, we refer to Baddeley, Rubak, and Turner (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 chosen the parameters for each realisation of $\mathcal{Z}_1$ so that the average number of points on $|W_n|$ is $100|W_n|$ (100, 400 or 900).

For the nonstationary Poisson point processes, we use the intensity function

$$\lambda_n(x) = \exp\left(\theta_{0,n} + \mathcal{Z}_1(x)\right), \tag{16}$$

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

$$\Lambda_n(x) = \exp\left(\theta_{0,n} + \mathcal{Z}_1(x) + \mathcal{Z}_2(x)\right), \tag{17}$$

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

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

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

$$C'_n(x, y) = \sqrt{\lambda'_n(x)\lambda'_n(y)}\lambda_{n,\text{dom}} \exp\left(-\frac{|x-y|^2}{\beta}\right). \quad (18)$$

The intensity of the DPP is given by $\lambda_n(x) = C'_n(x,x) = \lambda_{n,\text{dom}}\lambda'_n(x)$. 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, whereas the true value of θ_0 depends on the window W_n and the realisation 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.

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

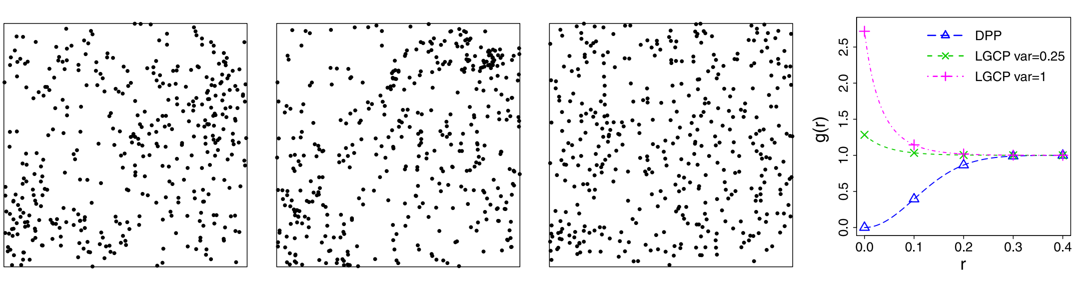

FIGURE 1 From left to right, the first three panels show a realisation on $[0, 2]^2$ of a log-Gaussian Cox process (LGCP) with variance parameter 0.25, an LGCP with variance parameter 1, and a determinantal point process (DPP). Last panel: a plot of the corresponding theoretical pair correlation functions [Colour figure can be viewed at wileyonlinelibrary.com]

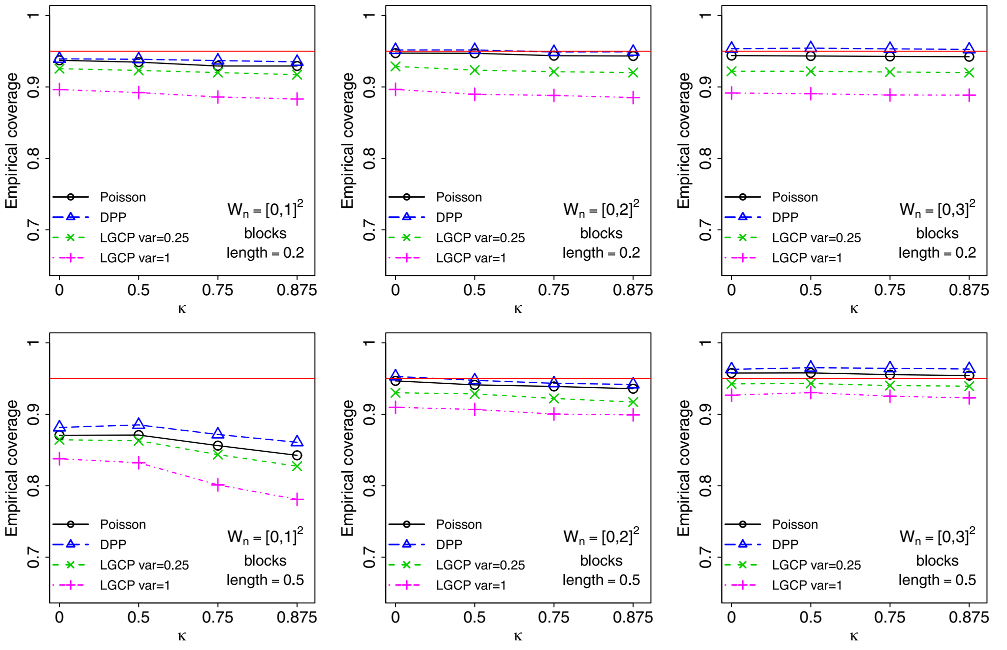

FIGURE 2 Estimated coverages of the confidence intervals for θ_1 when using the subsampling estimator (12). Upper row to lower row: $b_l = 0.2, 0.5$. Left column to right column: $n = 1, 2, 3$. In each plot, the estimated coverage is computed for four point process models, namely, nonstationary Poisson point process, determinantal point process (DPP), and two log-Gaussian Cox processed (LGCPs); $\kappa = 0, 0.5, 0.75, 0.875$. The lines joining the points just serve to aid visual interpretation. The straight horizontal red line indicates the value 0.95 [Colour figure can be viewed at wileyonlinelibrary.com]

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. It may seem odd that overlapping blocks do not outperform nonoverlapping blocks but, as suggested by a referee, this may be due to an over-representation of the middle part of the observation window when using overlapping blocks.

8 | 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 that 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 7, the estimated coverages of the resulting approximate confidence intervals for the parameter θ_1 range 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 (6), in any dimension.

We have also compared our subsampling estimator with the thinned block bootstrap estimator proposed in Guan and Loh (2007) by doing the same simulation study as in Guan and Loh (2007). For each estimator and simulation study setting, we identified the block size that gave the coverage closest to 95% percent. For each simulation setting, these coverages obtained with the two estimators differed by no more than 2%, and none of the estimators was consistently better. 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

We thank the reviewers for their very careful reading of the manuscript and their constructive and helpful comments. 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.

ORCID

Christophe Ange Napoléon Biscio https://orcid.org/0000-0002-2218-6825

REFERENCES

Baddeley, A., Rubak, E., & Turner, R. (2015). *Spatial point patterns: Methodology and applications with R*. Boca Raton, FL: Chapman and Hall/CRC Press. Retrieved from https://www.crcpress.com/Spatial-Point-Patterns-Methodology-and-Applications-with-R/Baddeley-Rubak-Turner/p/book/9781482210200/

Billingsley, P. (1995). *Wiley series in probability and mathematical statistics. Probability and measure*. Hoboken, NJ: John Wiley & Sons.

Biscio, C. A. N., & Coeurjolly, J.-F. (2016). Standard and robust intensity parameter estimation for stationary determinantal point processes. *Spatial Statistics*, *18*, 24–39.

Biscio, C. A. N., & Lavancier, F. (2016). Brillinger mixing of determinantal point processes and statistical applications. *Electronic Journal of Statistics*, *10*, 582–607.

Biscio, C. A. N., & Lavancier, F. (2017). Contrast estimation for parametric stationary determinantal point processes. *Scandinavian Journal of Statistics*, *44*, 204–229. https://doi.org/10.1111/sjos.12249

Biscio, C. A. N., Poinas, A., & Waagepetersen, R. (2018). A note on gaps in proofs of central limit theorems. *Statistics & Probability Letters*, 7–10.

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

Chung, K. L. (2001). *A course in probability theory*. New York, NY: Academic Press.

Coeurjolly, J.-F. (2017). Median-based estimation of the intensity of a spatial point process. *Annals of the Institute of Statistical Mathematics*, *69*, 303–331.

Coeurjolly, J.-F., & Lavancier, F. (2017). Parametric estimation of pairwise Gibbs point processes with infinite range interaction. *Bernoulli*, *23*, 1299–1334.

Coeurjolly, J.-F., & Møller, J. (2014). Variational approach for spatial point process intensity estimation. *Bernoulli*, *20*, 1097–1125. https://doi.org/10.3150/13-BEJ516

Daley, D. J., & Vere-Jones, D. (2003). *An introduction to the theory of point processes: Volume I: Elementary theory and methods* (2nd ed.). New York, NY: Springer-Verlag.

Daley, D. J., & Vere-Jones, D. (2008). *An introduction to the theory of point processes: Vol. II: General theory and structure* (2nd ed.). New York, NY: Springer-Verlag.

Ekström, M. (2008). Subsampling variance estimation for non-stationary spatial lattice data. *Scandinavian Journal of Statistics*, *35*, 38–63. https://doi.org/10.1111/j.1467-9469.2007.00572.x

Fazekas, I., Kukush, A., & Tómács, T. (2000). On the Rosenthal inequality for mixing fields. *Ukrainian Mathematical Journal*, *52*, 305–318.

Guan, Y., Jalilian, A., & Waagepetersen, R. (2015). Quasi-likelihood for spatial point processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *77*, 677–697.

Guan, Y., & Loh, J. M. (2007). A thinned block bootstrap variance estimation procedure for inhomogeneous spatial point patterns. *Journal of the American Statistical Association*, *102*, 1377–1386. https://doi.org/10.1198/016214507000000879

Guan, Y., & Shen, Y. (2010). A weighted estimating equation approach for inhomogeneous spatial point processes. *Biometrika*, *97*, 867–880. https://doi.org/10.1093/biomet/asq043

Guan, Y., & Sherman, M. (2007). On least squares fitting for stationary spatial point processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *69*, 31–49. https://doi.org/10.1111/j.1467-9868.2007.00575.x

Guyon, X. (1995). *Probability and its applications*. *Random fields on a network*. New York, NY: Springer.

Heinrich, L. (1992). Minimum contrast estimates for parameters of ergodic spatial point processes. Paper presented at Transactions of the 11th Prague Conference on Random Processes, Information Theory and Statistical Decision Functions, Prague, Czech Republic.

Heinrich, L., & Klein, S. (2014). Central limit theorems for empirical product densities of stationary point processes. *Statistical Inference for Stochastic Processes*, *17*, 121–138.

Ibragimov, I. A., & Linnik, Y. V. (1971). *Independent and stationary sequences of random variables*. Groningen, The Netherlands: Wolters-Noordhoff.

Jalilian, A., Guan, Y., & Waagepetersen, R. (2017). Orthogonal series estimation of the pair correlation function for spatial point processes. *Statistica Sinica*. Accepted.

Jensen, J. L., & Künsch, H. R. (1994). On asymptotic normality of pseudo likelihood estimates for pairwise interaction processes. *Annals of the Institute of Statistical Mathematics*, *46*, 475–486. https://doi.org/10.1007/BF00773511

Karácsony, Z. (2006). A central limit theorem for mixing random fields. *Miskolc mathematical notes. A publication of the University of Miskolc*, *7*, 147–160.

Lahiri, S. N. (2003). *Resampling methods for dependent data*. New York, NY: Springer-Verlag.

Lavancier, F., Møller, J., & Rubak, E. (2014). Determinantal point process models and statistical inference (extended version). *Royal Statistical Society*, *77*, 853–877. Retrieved from arXiv:1205.4818.

Loh, J. M. (2010). Bootstrapping an inhomogeneous point process. *Journal of Statistical Planning and Inference*, *140*, 734–749. https://doi.org/10.1016/j.jspi.2009.09.003

Lukacs, E. (1970). *Characteristic functions*. London, UK: Griffin.

Mattfeldt, T., Häbel, H., & Fleischer, F. (2013). Block bootstrap methods for the estimation of the intensity of a spatial point process with confidence bounds. *Journal of Microscopy*, *251*, 84–98. https://doi.org/10.1111/jmi.12048

Poinas, A., Delyon, B., & Lavancier, F. (2017). Mixing properties and central limit theorem for associated point processes. Retrieved from arXiv:1705.02276.

Politis, D. N., & Romano, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics*, *22*, 2031–2050. https://doi.org/10.1214/aos/1176325770

Prokešová, M., & Jensen, E. B. V. (2013). Asymptotic Palm likelihood theory for stationary point processes. *Annals of the Institute of Statistical Mathematics*, *65*, 387–412.

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

Waagepetersen, R. (2007). An estimating function approach to inference for inhomogeneous Neyman–Scott processes. *Biometrics*, *63*, 252–258.

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

Xu, G., Waagepetersen, R., & Guan, Y. (2018). Stochastic quasi-likelihood for case-control point pattern data. *Journal of the American Statistical Association*.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Biscio CAN, Waagepetersen R. A general central limit theorem and a subsampling variance estimator for α-mixing point processes. *Scand J Statist*. 2019;1–23. https://doi.org/10.1111/sjos.12389

APPENDIX A

PROOF OF THEOREM 1

Suppose first that we have verified Theorem 1 in the univariate case $q = 1$. Then, by $(\mathcal{H}4)$ and Lemma F3, we may use the extension of the Cramér–Wold device in Lemma F6 to verify Theorem 1 also for $q > 1$. We thus focus on the case $q = 1$.

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

Theorem A1. *Let the situation be as in Theorem 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,

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

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

Proof of Theorem A1. Define for $L > 0$ and $n \in \mathbb{N}$:

- for $\mathbf{l} \in \mathbb{Z}^d$, $Z_n^{(L)}(\mathbf{l}) = (Z_n(\mathbf{l}) - \mathbb{E}(Z_n))\mathbf{1}\left(|Z_n(\mathbf{l}) - \mathbb{E}(Z_n(\mathbf{l}))| \le L\right)$,
- for $\mathbf{l} \in \mathbb{Z}^d$, $\breve{Z}_n^{(L)}(\mathbf{l}) = (Z_n(\mathbf{l}) - \mathbb{E}(Z_n))\mathbf{1}\left(|Z_n(\mathbf{l}) - \mathbb{E}(Z_n(\mathbf{l}))| > L\right)$,
- $X_n = \frac{1}{\sigma_n} \sum_{\mathbf{l} \in \mathcal{D}_n} (Z_n(\mathbf{l}) - \mathbb{E}Z_n(\mathbf{l}))$,
- $X_n^{(L)} = \frac{1}{\sigma_n} \sum_{\mathbf{l} \in \mathcal{D}_n} Z_n^{(L)}(\mathbf{l})$,
- $\breve{X}_n^{(L)} = \frac{1}{\sigma_n} \sum_{\mathbf{l} \in \mathcal{D}_n} \breve{Z}_n^{(L)}(\mathbf{l})$.

By Lemma F2, we have for $r \ge 0$

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

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

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

Note that such τ' exists by $(\mathcal{H}3)$. In addition, by $(\mathcal{H}3)$, the last expression in the inequality is bounded. We may then adapt Theorem 1 in Fazekas, Kukush, and Tómács (2000) to the lattice $s_n\mathbb{Z}^d$ and so there exists a constant $c_2 > 0$ such that

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

By $(\mathcal{H}3)$ and (25.18) in Billingsley (1995), the collection of random variables $\{|Z_n(\mathbf{l}) - \mathbb{E}Z_n(\mathbf{l})|^{2+\tau'}, n \in \mathbb{N}, \mathbf{l} \in \mathcal{D}_n\}$ is uniformly integrable so that

$$\lim_{L\to\infty} \sup_{n\in\mathbb{N}} \sup_{\mathbf{l}\in\mathcal{D}_n} \left(\mathbb{E}\left|\check{Z}_n^{(L)}(\mathbf{l})\right|^{2+\tau'} \right)^{\frac{2}{2+\tau'}} = 0.$$

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

$$\lim_{L\to\infty} \sup_{n\in\mathbb{N}} \mathbb{E}\left(\check{X}_n^{(L)}\right)^2 = 0. \tag{A1}$$

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

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

Then, by the Cauchy–Schwarz inequality and (A1),

$$\lim_{L\to\infty} \sup_{n\in\mathbb{N}} \left| \frac{\sigma_n^2(L)}{\sigma_n^2} - 1 \right| = 0. \tag{A2}$$

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

$$\left|\mathbb{E}e^{itX_n} - e^{-\frac{t^2}{2}}\right| = \left|\mathbb{E}\left[\left(e^{it\check{X}_n^{(L)}} - 1\right)e^{itX_n^{(L)}} + e^{itX_n^{(L)}} - e^{-\frac{t^2}{2}}\right]\right|$$

$$\leq \mathbb{E}\left|e^{it\check{X}_n^{(L)}} - 1\right| + \left|\mathbb{E}e^{itX_n^{(L)}} - e^{-\frac{\sigma_n^2(L)}{\sigma_n^2}\frac{t^2}{2}}\right| + \left|e^{-\frac{\sigma_n^2(L)}{\sigma_n^2}\frac{t^2}{2}} - e^{-\frac{t^2}{2}}\right|. \tag{A3}$$

Because, for all $x \in \mathbb{R}$, $\left|e^{ix} - 1\right| \leq |x|$,

$$\mathbb{E}\left|e^{it\check{X}_n^{(L)}} - 1\right| \leq \mathbb{E}\left|t\check{X}_n^{(L)}\right| \leq |t| \sup_{n\in\mathbb{N}} \sqrt{\mathbb{E}\left(\check{X}_n^{(L)}\right)^2}. \tag{A4}$$

Let $\delta_L = \sup_{n\in\mathbb{N}}\left|\frac{\sigma_n(L)}{\sigma_n} - 1\right|$. By (A2), we may consider L to be large enough so that $\delta_L < 1$. Then, writing $U_n = \frac{\sigma_n}{\sigma_n(L)}X_n^{(L)}$, we have

$$\left|\mathbb{E}e^{itX_n^{(L)}} - e^{-\frac{\sigma_n^2(L)}{\sigma_n^2}\frac{t^2}{2}}\right| = \left|\mathbb{E}e^{it\frac{\sigma_n(L)}{\sigma_n}U_n} - e^{-\frac{\sigma_n^2(L)}{\sigma_n^2}\frac{t^2}{2}}\right| \leq \sup_{v\in[1-\delta_L, 1+\delta_L]}\left|\mathbb{E}e^{itvU_n} - e^{-\frac{(tv)^2}{2}}\right|,$$

so by Theorem A1 and Corollary 1 to Theorem 3.6.1 in Lukacs (1970), for $L \geq 0$,

$$\lim_{n\to\infty}\left|\mathbb{E}e^{itX_n^{(L)}} - e^{-\frac{\sigma_n^2(L)}{\sigma_n^2}\frac{t^2}{2}}\right| = 0. \tag{A5}$$

Moreover, by a first-order Taylor expansion with remainder,

$$\sup_{n\in\mathbb{N}}\left|e^{-\frac{\sigma_n^2(L)}{\sigma_n^2}\frac{t^2}{2}} - e^{-\frac{t^2}{2}}\right| = e^{-\frac{t^2}{2}}\sup_{n\in\mathbb{N}}\left|e^{-\left(\frac{\sigma_n^2(L)}{\sigma_n^2}-1\right)\frac{t^2}{2}} - 1\right| \leq \frac{t^2}{2}\delta_L + \frac{t^4}{8}\exp\left(\delta_L\frac{t^2}{2}\right)\delta_L^2. \tag{A6}$$

Therefore, by (A3), (A4), (A5), and (A6),

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

which by (A1) and (A2) tends to 0 as L tends to infinity. □

APPENDIX B

PROOF OF THEOREM A1

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 $\eta = 0$, $m_n = |\mathcal{D}_n|^{1/(2d+\epsilon/2)}$ 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 $(\mathcal{H}1)$–$(\mathcal{H}2)$,

$$m_n \to \infty, \quad m_n/s_n \to \infty, \tag{B1}$$

$$\lim_{n\to\infty} \sqrt{|\mathcal{D}_n|} m_n^{-d-\epsilon} = 0, \tag{B2}$$

and

$$\lim_{n\to\infty} \sqrt{|\mathcal{D}_n|} (m_n/s_n)^{-d} = \infty. \tag{B3}$$

By Lemma F5, $\sup_{n\in\mathbb{N}} \mathbb{E}\bar{S}_n^2 < \infty$. Thus, by Lemma F2 in Bolthausen (1982; see also the discussion in Biscio, Poinas, & Waagepetersen, 2018), Theorem A1 is proved if

$$\lim_{n\to\infty} \mathbb{E}\left[(it - \bar{S}_n)e^{it\bar{S}_n}\right] = 0. \tag{B4}$$

Notice that

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

where

$$A_1 = ite^{it\bar{S}_n}\left(1 - \frac{1}{a_n}\sum_{\substack{\mathbf{i},\mathbf{j}\in\mathcal{D}_n\\ \mathrm{d}(\mathbf{i},\mathbf{j})\le m_n}} Y_n(\mathbf{i})Y_n(\mathbf{j})\right), \tag{B5}$$

$$A_2 = \frac{e^{it\bar{S}_n}}{\sqrt{a_n}}\sum_{\mathbf{i}\in\mathcal{D}_n} Y_n(\mathbf{i})\left(1 - it\bar{S}_n(\mathbf{i}) - e^{-it\bar{S}_n(\mathbf{i})}\right), \tag{B6}$$

$$A_3 = \frac{1}{\sqrt{a_n}}\sum_{\mathbf{i}\in\mathcal{D}_n} Y_n(\mathbf{i})e^{it\left(\bar{S}_n - \bar{S}_n(\mathbf{i})\right)}. \tag{B7}$$

Hence, (B4) follows from the convergences to zero of A_1, A_2, and A_3, as established in Section 4 in the supplementary material.

APPENDIX C

PROOF OF THEOREM 2

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

Theorem C1. *For $n \in \mathbb{N}$, let R_n be a random field on $\mathbb{Z}^d$; let $\{W_n\}_{n\in\mathbb{N}}$ 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 subrectangles defined as in (11) 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)* $\mathrm{Var}\left(\Psi_{W_n}\right) - \frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|}\sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \mathrm{Var}\left(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}}\right) \to 0$ *as* $n \to \infty$;
- *(iv)* $\frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|}\sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}}\left(\mathbb{E}\left(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}}\right) - \mathbb{E}\left(\sum_{\mathbf{s}\in\mathcal{T}_{\mathbf{k}_n,n}} \frac{\Psi_{B_{\mathbf{k}_n,\mathbf{s}}}}{|\mathcal{T}_{\mathbf{k}_n,n}|}\right)\right)^2 \to 0$ *as* $n \to \infty$.

Let further

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

Then, we have the convergence

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

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

Below, we check the assumptions (i)–(iii) in Theorem C1 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 2 is proved directly by Theorem C1.

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

$$T_{B_{\mathbf{k}_n,\mathbf{t}}}(\mathbf{X}) - \mathbb{E}(T_{B_{\mathbf{k}_n,\mathbf{t}}}(\mathbf{X})) = \sum_{\mathbf{l}\in\mathcal{D}_n(B_{\mathbf{k}_n,\mathbf{t}})} Y_n(\mathbf{l}). \tag{C1}$$

Let ϵ' be as in $(\mathcal{S}2)$; then, by Lemma F2 and $(\mathcal{S}6)$, we have, for $\epsilon < \epsilon'$,

$$\sum_{r=1}^{\infty}\left(\alpha_{5,5}^{Z_n}(r)\right)^{\frac{\epsilon}{6+\epsilon}} r^{5d-1} \le \sum_{r=1}^{\infty}\left(\alpha_{5v_n,5v_n}^{\mathbf{X}}(r-1-2R)\right)^{\frac{\epsilon}{6+\epsilon}} r^{5d-1} < \infty. \tag{C2}$$

Then, by $(\mathcal{S}2)$ and (C2), we may apply Theorem 1 in Fazekas et al. (2000), which states the existence of $c_1 > 0$ such that

$$\mathbb{E}\left|\sum_{\mathbf{l}\in\mathcal{D}_n(B_{\mathbf{k}_n,\mathbf{t}})} \frac{Y_n(\mathbf{l})}{\sqrt{|\mathcal{D}_n(B_{\mathbf{k}_n,\mathbf{t}})|}}\right|^{4+\epsilon'/2} \le c_1 \max\{U, V\}, \tag{C3}$$

where

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

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

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

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

and

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

Both of the last upper bounds on U and V are bounded. Therefore, by (C1) and (C3),

$$\sup_{n\in\mathbb{N}} \sup_{t\in\mathcal{T}_{\mathbf{k}_n,n}} \mathbb{E}\left|\frac{T_{B_{\mathbf{k}_n,\mathbf{t}}}(\mathbf{X}) - \mathbb{E}\left(T_{B_{\mathbf{k}_n,\mathbf{t}}}(\mathbf{X})\right)}{\sqrt{|\mathcal{D}_n(B_{\mathbf{k}_n,\mathbf{t}})|}}\right|^{4+\epsilon'/2} < \infty, \tag{C4}$$

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

Assumption (ii).

For c, δ as in $(\mathcal{S}5)$ and v_n as below (7), we have, by Lemma F2,

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

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

Assumptions (iii)–(iv).

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

APPENDIX D

PROOF OF THEOREM C1

The proof of Theorem C1 is based on several applications of the following Theorem D1, which states an intermediate result and is proved in Appendix E. Consequently, these theorems may look similar at first sight.

Theorem D1. *For $n \in \mathbb{N}$, let R_n be a random field on $\mathbb{Z}^d$; let $\{W_n\}_{n\in\mathbb{N}}$ 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 subrectangles defined as in (11) 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') $\mathbb{E}\left(\sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \frac{h_{B_{\mathbf{k}_n,\mathbf{t}}}}{|\mathcal{T}_{\mathbf{k}_n,n}|}\right) - \mathbb{E}(h_{W_n}) \to 0$ as $n \to \infty$.

Then, we have the convergence

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

We now give the proof of Theorem C1, 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 $x = (x_1, \ldots, x_d)^T \in \mathbb{R}^d$ 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 C1, we have

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

Hence,

$$\hat{\varsigma}^{R_n}_n = C_1 + C_2 + C_3 + C_4 + C_5 + C_6, \tag{D1}$$

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

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

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

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

$$C_4 = \frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|} \sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \left(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}} - \mathbb{E}\left(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}}\right)\right)\left(\mathbb{E}\left(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}}\right) - \mathbb{E}\left(\overline{\Psi}_{B_{\mathbf{k}_n}}\right)\right)^T$$
$$+ \frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|} \sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \left(\mathbb{E}\left(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}}\right) - \mathbb{E}\left(\overline{\Psi}_{B_{\mathbf{k}_n}}\right)\right)\left(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}} - \mathbb{E}\left(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}}\right)\right)^T,$$

$$C_5 = \frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|} \sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \left(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}} - \mathbb{E}\left(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}}\right)\right)\left(\mathbb{E}\left(\overline{\Psi}_{B_{\mathbf{k}_n}}\right) - \overline{\Psi}_{B_{\mathbf{k}_n}}\right)^T$$
$$+ \frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|} \sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \left(\mathbb{E}\left(\overline{\Psi}_{B_{\mathbf{k}_n}}\right) - \overline{\Psi}_{B_{\mathbf{k}_n}}\right)\left(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}} - \mathbb{E}\left(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}}\right)\right)^T,$$

$$C_6 = \frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|} \sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \left(\mathbb{E}\left(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}}\right) - \mathbb{E}\left(\overline{\Psi}_{B_{\mathbf{k}_n}}\right)\right)\left(\mathbb{E}\left(\overline{\Psi}_{B_{\mathbf{k}_n}}\right) - \overline{\Psi}_{B_{\mathbf{k}_n}}\right)^T$$
$$+ \frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|} \sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}} \left(\mathbb{E}\left(\overline{\Psi}_{B_{\mathbf{k}_n}}\right) - \overline{\Psi}_{B_{\mathbf{k}_n}}\right)\left(\mathbb{E}\left(\Psi_{B_{\mathbf{k}_n,\mathbf{t}}}\right) - \mathbb{E}\left(\overline{\Psi}_{B_{\mathbf{k}_n}}\right)\right)^T.$$

The assumption (iv) implies directly that

$$\lim_{n\to\infty} C_2 = 0. \tag{D2}$$

By applying the Cauchy–Schwarz inequality for each sum in C_4, we have $\mathbb{E}(|C_4|^2) \le 4\mathbb{E}(|C_1|)|C_2|$. Furthermore, 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 (D2), it follows that

$$\lim_{n\to\infty} \mathbb{E}(|C_4|^2) = 0. \tag{D3}$$

We have

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

so

$$C_3 + C_5 + C_6 = -\left(\mathbb{E}\left(\overline{\Psi}_{B_{\mathbf{k}_n}}\right) - \overline{\Psi}_{B_{\mathbf{k}_n}}\right)^2. \tag{D4}$$

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 D1 with $h_{\cdot} = \Psi_{\cdot} - \mathbb{E}(\Psi_{\cdot})$ so that

$$\lim_{n\to\infty} \mathbb{E}(|Y_n|^2) = 0. \tag{D5}$$

Using (i)–(ii) and (iii), we may apply Theorem D1 with $h_\cdot = (\Psi_\cdot - \mathbb{E}(\Psi_\cdot))^2$. Hence,

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

which may be written as the convergence

$$C_1 - K_n \xrightarrow[n\to\infty]{L^2} 0. \tag{D6}$$

By Theorem 4.5.4 in Chung (2001), (D6) implies that $|C_1 - K_n|^2$ is uniformly integrable with respect to n. Moreover, $\mathbb{E}(|C_1|^2) = \mathbb{E}(|C_1 - K_n + K_n|^2) \le 2\mathbb{E}(|C_1 - K_n|^2) + 2|K_n|^2$ and it follows from ($\mathcal{S}3$) that K_n is uniformly bounded. Thus, $|C_1|^2$ is uniformly integrable so that, by Lemma F7, Y_n^4 is also uniformly integrable. Furthermore, (D5) implies that $Y_n^2 \xrightarrow[n\to\infty]{\mathbb{P}} 0$, so by Theorem 4.5.4 in Chung (2001), we have $Y_n \xrightarrow[n\to\infty]{L^4} 0$. By (D4), the last implies that

$$C_3 + C_5 + C_6 \xrightarrow[n\to\infty]{L^2} 0. \tag{D7}$$

Finally, Theorem C1 is proved by combining (D1), (D2), (D3), (D6), and (D7).

APPENDIX E

PROOF OF THEOREM D1

$$\mathbb{E}\left|\sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}}\frac{h_{B_{\mathbf{k}_n,\mathbf{t}}}}{|\mathcal{T}_{\mathbf{k}_n,n}|}-\mathbb{E}(h_{W_n})\right|^2=\mathbb{E}\left|\sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}}\frac{h_{B_{\mathbf{k}_n,\mathbf{t}}}-\mathbb{E}\left(h_{B_{\mathbf{k}_n,\mathbf{t}}}\right)}{|\mathcal{T}_{\mathbf{k}_n,n}|}\right|^2+\left|\sum_{\mathbf{t}\in\mathcal{T}_{\mathbf{k}_n,n}}\frac{\mathbb{E}\left(h_{B_{\mathbf{k}_n,\mathbf{t}}}\right)}{|\mathcal{T}_{\mathbf{k}_n,n}|}-\mathbb{E}(h_{W_n})\right|^2 \tag{E1}$$

Hence, if in (E1) the first expectation on the right-hand side converges to 0 as n tends to infinity, Theorem D1 is proved by (iii') and (E1). We have

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

where

$$M_1 = \frac{1}{|\mathcal{T}_{\mathbf{k}_n,n}|^2}\sum_{\substack{\mathbf{t}_1,\mathbf{t}_2\in\mathcal{T}_{\mathbf{k}_n,n},\\ \mathrm{d}\left(\mathbb{Z}^d\cap B_{\mathbf{k}_n,\mathbf{t}_1},\mathbb{Z}^d\cap B_{\mathbf{k}_n,\mathbf{t}_2}\right)\le \max \mathbf{k}_{n,i}}}\mathrm{Cov}\left(\left|h_{B_{\mathbf{k}_n,\mathbf{t}_1}}\right|,\left|h_{B_{\mathbf{k}_n,\mathbf{t}_2}}\right|\right),$$

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

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,

$$M_1 \le \frac{(2\max_i \mathbf{k}_{n,i}+1)^d}{|\mathcal{T}_{\mathbf{k}_n,n}|} \sup_{\substack{\mathbf{t}_1,\mathbf{t}_2 \in \mathcal{T}_{\mathbf{k}_n,n},\\ \mathrm{d}\left(\mathbb{Z}^d \cap B_{\mathbf{k}_n,\mathbf{t}_1}, \mathbb{Z}^d \cap B_{\mathbf{k}_n,\mathbf{t}_2}\right) \le \max \mathbf{k}_{n,i}}} \mathrm{Cov}\left(\left|h_{B_{\mathbf{k}_n,\mathbf{t}_1}}\right|, \left|h_{B_{\mathbf{k}_n,\mathbf{t}_2}}\right|\right)$$

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

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

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

Furthermore, by (11) for all $t \in \mathbb{Z}^d$, $|\mathbb{Z}^d \cap B_{\mathbf{k}_n,\mathbf{t}}| \le b_n$, where $b_n = \prod_{j=1}^d (2k_{n,j}+1)$. Then, by Lemma 1 in Sherman (1996), for any $\eta > 0$, we have

$$M_2 \le 4\eta^2 \alpha^{Z_n}_{b_n,b_n}(\max \mathbf{k}_{n,i}) + 3\sqrt{c_2}\sqrt{\mathbb{E}\left(X_1^{(\eta)}\right)^2} + 3\sqrt{c_2}\sqrt{\mathbb{E}\left(X_2^{(\eta)}\right)^2}, \tag{E2}$$

where for $i = 1, 2$, $X_i = |h_{B_{\mathbf{k}_n,\mathbf{t}_i}}|$, $X_i^{\eta} = X_i \mathbf{1}(X_i \ge \eta)$, 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 (E2), (i'), and (ii') that M_2 tends to 0 as n tends to infinity. Therefore, $\mathbb{E}|\sum_{\mathbf{t} \in \mathcal{T}_{\mathbf{k}_n,n}} (h_{B_{\mathbf{k}_n,\mathbf{t}}} - \mathbb{E}(h_{B_{\mathbf{k}_n,\mathbf{t}}}))/|\mathcal{T}_{\mathbf{k}_n,n}||^2$ converges to 0 as n tends to infinity.

APPENDIX 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 F1. *For all* $\mathbf{l}, \mathbf{k} \in s_n \mathbb{Z}^d$, *we have*

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

Lemma F2. *For* $c_1, c_2, r \ge 0$, *we have*

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

Lemma F3. *We have*

$$\limsup_{n\to\infty} \lambda_{\max}\left(\frac{\Sigma_n}{|D_n|}\right) < \infty,$$

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

Lemma F4. *For $k \in \mathbb{N}$ and $\mathbf{i} \in s_n\mathbb{Z}^d$,*

$$\left|\left\{\mathbf{j} \in s_n\mathbb{Z}^d \,:\, \mathrm{d}(\mathbf{i},\mathbf{j}) = s_n k\right\}\right| \le 3^d k^{d-1}.$$

Lemma F5. *Under the assumptions $(\mathcal{H}_b)$, $(\mathcal{H}2)$, and $(\mathcal{H}4)$, we have the convergence*

$$\lim_{n\to\infty}\left|1 - \frac{a_n}{\sigma_n^2}\right| = 0.$$

Lemma F6 (Biscio et al., 2018).
Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables in $\mathbb{R}^p$, for $p \in \mathbb{N}$, such that

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

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

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

$$\left(a^T\mathrm{Var}(X_n)a\right)^{-\frac{1}{2}} a^T X_n \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}(0,1).$$

Lemma F7. *Let the situation be as in Appendix D. We have*

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

Proof of Lemma F1. For any vector x, let $[x]_i$ denote its ith coordinate. By the Cauchy–Schwarz inequality,

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

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

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

On the other hand,

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

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