Residuals and goodness-of-fit tests for stationary marked Gibbs point processes

Jean-François Coeurjolly

Université Pierre Mendes France, Grenoble, France

and Frédéric Lavancier

Université de Nantes, France

[Received June 2010. Final revision May 2012]

Summary. The inspection of residuals is a fundamental step for investigating the quality of adjustment of a parametric model to data. For spatial point processes, the concept of residuals has been recently proposed as an empirical counterpart of the Campbell equilibrium equation for marked Gibbs point processes. The paper focuses on stationary marked Gibbs point processes and deals with asymptotic properties of residuals for such processes. In particular, the consistency and the asymptotic normality are obtained for a wide class of residuals including the classical residuals (raw, inverse and Pearson). On the basis of these asymptotic results, we define goodness-of-fit tests with type I error theoretically controlled. One of these tests constitutes an extension of the quadrat counting test that is widely used to test the null hypothesis of a homogeneous Poisson point process.

Keywords: Campbell theorem; Central limit theorem for spatial random fields; Georgii–Nguyen–Zessin formula; Maximum pseudolikelihood estimate; Quadrat counting test

1. Introduction

Recent works on statistical methods for spatial point pattern make parametric inference feasible for a wide range of models; see Møller (2010) for an overview of this topic and more generally Daley and Vere-Jones (1988), Stoyan et al. (1995), Møller and Waagepetersen (2004) or Illian et al. (2008) for a survey on spatial point processes. The question is then whether the model is well fitted to data or not. For classical parametric models, this is usually done via the inspection of residuals. They play a central role in parametric inference; see Atkinson (1985) for instance. This notion is quite complex for spatial point processes and has been recently proposed by Baddeley et al. (2005), following ideas from previous work by Stoyan and Grabarnik (1991).

The definition of residuals for spatial point processes is a natural generalization of the well-known residuals for point processes in one-dimensional time, used in survival analysis (see Fleming and Harrington (1991) or Andersen et al. (1993) for an overview). For example, a simple measure of the adequacy of a one-dimensional point process model consists in computing the difference between the number of events in an interval \([0, t]\) and the conditional intensity (or hazard rate of the lifetime distribution) parametrically estimated and integrated from 0 to \(t\).

The extension in higher dimension requires further developments owing to the lack of natural ordering. It may be done for point processes admitting a conditional density with respect to the
Poisson process. These point processes correspond to the Gibbs measures. The equilibrium in one dimension between the number of events and the integrated hazard rate may be replaced in higher dimension by the _Campbell equilibrium_ equation or _Georgii–Nguyen–Zessin_ formula (see Georgii (1976), Nguyen and Zessin (1979a) and Section 3), which is the basis for defining the class of _h_-residuals where _h_ represents a test function. In particular, Baddeley _et al._ (2005) considered various choices of _h_, leading to the so-called raw residuals, inverse residuals and Pearson residuals, and showed that they share similarities with the residuals that are obtained for generalized linear models.

Thanks to various diagnostic plots that were developed in Baddeley _et al._ (2005) and implementation within the R package _spatstat_ (Baddeley and Turner, 2005), residuals appear to be a very convenient tool in practice. Some properties of the residuals are exhibited in Baddeley _et al._ (2005) and Baddeley _et al._ (2008), including a conditional independence property and variance formulae in particular cases. Baddeley _et al._ (2005, 2008) conjecture that a strong law of large numbers and a central limit theorem should hold for the residuals as the sampling window expands.

Our paper addresses this question for _d_-dimensional stationary marked Gibbs point processes. We obtain strong consistency and asymptotic normality in several contexts for a large class of test functions _h_. The _h_-residuals crucially depend on an estimate of the parameter vector. We consider the natural framework where the estimate is computed with the same data over which the _h_-residuals are assessed. We assume for simplicity that this estimate corresponds to the maximum pseudolikelihood estimate (MPE) (see Baddeley and Turner (2000) for instance), for which asymptotic properties are now well known (see Jensen and Møller (1991), Jensen and Künsch (1994), Billiot _et al._ (2008), Dereudre and Lavancier (2009) and Coeurjolly and Drouilhet (2010)). However, another choice of estimate is possible, provided that some minimal properties are satisfied. The assumptions on the Gibbs models are in particular satisfied by several classical models, such as the area interaction point process, the multi-Strauss marked point process, the Strauss-type disc process and Geyer’s triplet point process. Our asymptotic results require a new multivariate central limit theorem for spatial processes, specifically for some non-stationary conditional centred random fields as in Jensen and Künsch (1994) and Comets and Janzura (1998). But, in contrast with the previous references where the observation domain is assumed to be of the form \([-n, n]^d\) for some integer _n_, we consider domains that may increase continuously up to \(\mathbb{R}^d\). This particularity, which seems more relevant, demands an extension that we display in Appendix A.

Moreover, on the basis of these asymptotic results, we propose goodness-of-fit tests for which the type I error is asymptotically controlled. Two of them are based on the computation of the residuals on different subdomains of the observation window and extend in a very natural way the quadrat counting test for homogeneous Poisson distributions (see Diggle (2003) for instance). Besides, we present a test which combines several different _h_-residuals (associated with different functions _h_), computed on the entire observation window. A short simulation study illustrates our result and assesses the power in some cases. However, a thorough study to compare our tests and to reveal their limits requires extensive simulations and this would deserve a separate paper.

To the best of our knowledge, it is the first attempt to propose goodness-of-fit tests for Gibbs models for which the type I error is asymptotically controlled. The only available theoretical testing procedures concern the null hypothesis of a homogeneous or inhomogeneous Poisson point process; see Guan (2008). On the basis of the residuals, the existing validation methods for marked Gibbs point processes are either graphical (e.g. by using the _QQ_-plot that was proposed by Baddeley _et al._ (2005)) or rely on Monte-Carlo-based simulations; see also Baddeley...
et al. (2011) for some diagnostic scores. Another alternative procedure consists in transforming the point process into a homogeneous Poisson point process, either by thinning (Møller and Schoenberg, 2010) or by superposition (Møller and Berthelsen, 2012), and then applying a standard goodness-of-fit test for homogeneous Poisson processes. However, we stress that these procedures do not integrate the uncertainty from the estimation of the parameters and therefore the type I error is not theoretically controlled.

The rest of the paper is organized as follows. Section 2 gathers the main notation and briefly displays the general background. The definition of marked Gibbs point processes is given and some examples are provided. In Section 3, the Georgii–Nguyen–Zessin formula is recalled, displaying the general background. The definition of marked Gibbs point processes is given and the type I error is not theoretically controlled. Procedures do not integrate the uncertainty from the estimation of the parameters and therefore standard goodness-of-fit test for homogeneous Poisson processes. However, we stress that these procedures do not integrate the uncertainty from the estimation of the parameters and therefore the type I error is not theoretically controlled.

Framework 1 leads to a generalization of the quadrat counting test for homogeneous Poisson point processes. Framework 2 yields a test which combines the information coming from several residuals, such as residuals coming from the estimation of the empty space function at several points. The range of application of our tests is discussed and a short simulation study is proposed. The main material to prove our results is composed of an ergodic theorem obtained by Nguyen and Zessin (1979b) and a new multivariate central limit theorem is presented in Appendix A. The proofs of our main results are gathered in Appendix B. Whereas, for clarity, our results are displayed under simplified assumptions, the proofs are given for a more general setting presented in Appendix B.1.

2. Gibbs point processes

2.1. General notation

We denote by $B(\mathbb{R}^d)$ the space of bounded Borel sets in $\mathbb{R}^d$. For any $\Lambda \in B(\mathbb{R}^d)$, $\Lambda^c$ denotes the complementary set of $\Lambda$ inside $\mathbb{R}^d$. The norm $|\cdot|$ will be used without ambiguity for different kinds of object. For a vector $x$, $|x|$ denotes its uniform norm. For a countable set $\mathcal{J}$, $|\mathcal{J}|$ represents the number of elements belonging to $\mathcal{J}$. For a set $\Delta \in B(\mathbb{R}^d)$, $|\Delta|$ is the volume of $\Delta$. The transpose of a vector $x$ is denoted by $x^T$. The notation $\|x\|$ stands for the Euclidean norm of $x$.

For all $x \in \mathbb{R}^d$ and $\rho > 0$, let $B(x, \rho) := \{y \in \mathbb{R}^d, \|y - x\| \leq \rho\}$. The space $\mathbb{R}^d$ is endowed with the Borel $\sigma$-algebra and the Lebesgue measure $\lambda$. Let $\mathcal{M}$ be a measurable space, which aims at being the mark space, endowed with the $\sigma$-algebra $\mathcal{M}$ and the probability measure $\lambda^m$. The state space of the point processes will be $\mathbb{S} := \mathbb{R}^d \times \mathcal{M}$ measured by $\mu := \lambda \otimes \lambda^m$ where ‘$\otimes$’ stands for the product measure. We shall denote for short $\lambda^m = (\lambda, m)$ an element of $\mathbb{S}$.

The space of point configurations will be denoted by $\Omega$. This is the set of simple locally finite subsets of $\mathbb{S}$. It is endowed with the $\sigma$-algebra $\mathcal{F}$ generated by the sets $\{\varphi \in \Omega, |\varphi \cap (\Lambda \times A)| = n\}$.
for all $n \in \mathbb{N}$, for all $A \in \mathcal{M}$ and for all $\Lambda \in \mathcal{B}(\mathbb{R}^d)$. For any $\varphi \in \Omega$ and any $\Lambda \in \mathcal{B}(\mathbb{R}^d)$, we denote $\varphi_\Lambda := \varphi \cap (\Lambda \times \mathbb{M})$. We shall use without ambiguity the notation $\varphi \cup \{x^m\} = \varphi \cup x^m$ and, for $x^m \in \varphi$, $\varphi \setminus \{x^m\} = \varphi \setminus x^m$.

A marked point process $\Phi$ is an $\Omega$-valued random variable, with probability distribution $P$ on $(\Omega, \mathcal{F})$. Finally, for brevity, we say that ‘$\Phi$ is observed in $\Lambda$’ for $\Lambda \in \mathcal{B}(\mathbb{R}^d)$ when the locations of $\Phi$ are in $\Lambda$ and the marks are in $\mathbb{M}$.

2.2. Definition of Gibbs point processes

The framework of this paper is concerned with marked Gibbs point processes defined on the infinite volume $\mathbb{R}^d$. Since we are interested in asymptotic properties, we restrict our study to stationary point processes, which allows us to use standard limit theorems. Though an extension towards non-stationary point processes is of relevance, this is out of the scope of this paper.

The marked point process that is considered as a reference is the marked Poisson process $\pi^\nu, \nu$ for all $\nu \in \mathbb{M}$.

Let $\theta$ be some parameter belonging to $\mathbb{R}^p$ for some $p \geq 1$. For any $\Lambda \in \mathcal{B}(\mathbb{R}^d)$, let us consider the parametric function $V_\Lambda(\cdot; \theta)$ from $\Omega$ into $\mathbb{R} \cup \{\infty\}$. For fixed $\theta$, $(V_\Lambda(\cdot; \theta))_{\Lambda \in \mathcal{B}(\mathbb{R}^d)}$ constitutes a compatible family of energies if, for every $\Lambda \subset \Lambda'$ in $\mathcal{B}(\mathbb{R}^d)$, there is a measurable function $\psi_{\Lambda, \Lambda'}$ from $\Omega$ into $\mathbb{R} \cup \{\infty\}$ such that

$$V_{\Lambda'}(\varphi; \theta) = V_\Lambda(\varphi; \theta) + \psi_{\Lambda, \Lambda'}(\varphi_{\Lambda'}; \theta) \quad \forall \varphi \in \Omega. \tag{1}$$

From a physical point of view, $V_\Lambda(\varphi_\Lambda; \theta)$ is the energy of $\varphi_\Lambda$ in $\Lambda$ given the outside configuration $\varphi_{\Lambda^c}$. The following definition is the classical way to define Gibbs measures through their conditional specifications (Preston, 1976).

**Definition 1.** A probability measure $P_\theta$ on $\Omega$ is a marked Gibbs measure for the compatible family of energies $(V_\Lambda(\cdot; \theta))_{\Lambda \in \mathcal{B}(\mathbb{R}^d)}$ and the intensity $\nu$ if for every $\Lambda \in \mathcal{B}(\mathbb{R}^d)$, for $P_\theta$ almost every outside configuration $\varphi_{\Lambda^c}$, the law of $P_\theta$ given $\varphi_{\Lambda^c}$ admits the following conditional density with respect to $\pi^\nu$:

$$f_\Lambda(\varphi_\Lambda | \varphi_{\Lambda^c}; \theta) = \frac{1}{Z_\Lambda(\varphi_{\Lambda^c}; \theta)} \exp \{-V_\Lambda(\varphi; \theta)\},$$

where $Z_\Lambda(\varphi_{\Lambda^c}; \theta)$ is a normalization called the partition function.

The existence of a Gibbs measure on $\Omega$ which satisfies these conditional specifications is a difficult issue. We do not want to open this discussion here and we shall assume that the Gibbs measures we consider exist. We refer the interested reader to Ruelle (1969), Preston (1976), Bertin et al. (1999), Dereudre (2005) and Dereudre et al. (2011). We assume moreover that the family of energies is hereditary, which means that, for any $\Lambda \in \mathcal{B}(\mathbb{R}^d)$, for any $\varphi \in \Omega$ and for all $x^m \in \Lambda \times \mathbb{M}$, $V_\Lambda(\varphi; \theta) = \infty$ implies that $V_\Lambda(\varphi \cup \{x^m\}; \theta) = \infty$ (or, equivalently, $f_\Lambda(\varphi_\Lambda | \varphi_{\Lambda^c}; \theta) > 0$ implies that $f_\Lambda(\varphi_\Lambda \setminus \{x^m\} | \varphi_{\Lambda^c}; \theta) > 0$). The setting of this paper can be extended to the non-hereditary case, following similar ideas to those in Dereudre and Lavancier (2009), but we omit this extension for brevity. The main assumption is then as follows.

**Existence assumption.** There exists a non-empty set $\Theta \subseteq \mathbb{R}^d$ such that, for any $\theta \in \Theta$, the compatible family of energies $(V_\Lambda(\cdot; \theta))_{\Lambda \in \mathcal{B}(\mathbb{R}^d)}$ is invariant by translation, and such that an associated Gibbs measure $P_\theta$ exists and is stationary. Our data consist in the realization of a
marked point process with Gibbs measure $P_{\theta^*}$, $\theta^* \in \Theta$ and observed within some bounded domain. The vector $\theta^*$ is thus the true parameter to be estimated.

The local energy that is required to insert a marked point $x_m$ into the configuration $\varphi$ is defined for any $\Lambda$ containing $x_m$ by

$$V(x_m | \varphi; \theta) := V_\Lambda(\varphi \cup \{x_m\}; \theta) - V_\Lambda(\varphi; \theta).$$

From condition (1), this definition does not depend on $\Lambda$. We stress that the local energy function for a Gibbs point process is nothing other than the logarithm of the inverse of the Papangelou conditional intensity (e.g. Möller and Waagepetersen (2004)). We restrict our study to finite range interaction point processes in the sense of the following finite range assumption, which is the main limitation of this paper.

**Finite range assumption.** There exists $D \geq 0$ such that, for all $(m, \varphi) \in M \times \Omega$, $V(0^m | \varphi; \theta) = V(0^m | \varphi_{B(0, D)}; \theta)$.

The range of interaction of the local energy is by definition the smallest $D$ satisfying the above equation. Given a finite range interaction Gibbs model, we assume in this paper that the range is known (in particular we do not estimate it). However, in what follows, $D$ is not necessarily equal to this range but may be any positive number from the above finite range assumption.

### 2.3. Exponential family models

To make the previous setting more explicit, we focus on models whose energies are linear in the parameters: for any $\Lambda \in B(\mathbb{R}^d)$, $V_\Lambda(\varphi; \theta) = \theta^T v_\Lambda(\varphi)$, where $v_\Lambda(\varphi) := (v_{1, \Lambda}(\varphi), \ldots, v_{p, \Lambda}(\varphi))^T$ is the vector of sufficient statistics, or equivalently $V(x^m | \varphi; \theta) = \theta^T v(x^m | \varphi)$, where $v(x^m | \varphi) = (v_1(x^m | \varphi), \ldots, v_p(x^m | \varphi))$. Our asymptotic results will be mainly valid under the following general assumption that was proposed in Billiot et al. (2008).

**Exponentiality assumption.** For any $\theta \in \Theta$, $V(x^m | \varphi; \theta) = \theta^T v(x^m | \varphi)$ and

(a) (invariance by translation) $v(x^m | \varphi) = v(0^m | \tau_x \varphi)$ where $\tau_x$ is the translation by $-x$,

(b) (locality) there exists $D \geq 0$ such that $v(x^m | \varphi) = v(x^m | \varphi_{B(x, D)})$,

(c) (stability and integrability) for $i = 1, \ldots, p$, there exist $c_i \geq 0$ and $k_i \in \mathbb{N}$ such that, for all $(m, \varphi) \in M \times \Omega$, either $\theta_i \geq 0$ and $-c_i \leq v_i(0^m | \varphi) \leq c_i \varphi_{B(0, D)} | \varphi_{B(0, D)} | k_i$ or $v_i(0^m | \varphi) \leq c_i$, and

(d) (identifiability) there exist $A_1, \ldots, A_l$, $l \geq p$, disjoint events of $M \times \mathcal{F}$ such that, for $i = 1, \ldots, l$, the product measure $\lambda^\infty \otimes P_{\theta_i}(A_i)$ is positive and, for $(m_i, \varphi_i) \in A_i$, the $(l, p)$ matrix with entries $v_j(0^m_i | \varphi_i)$ is injective.

The interest of such an assumption is threefold. First, Bertin et al. (1999) proved (in particular) that the above assumption is a sufficient condition to ensure the existence and finite range assumptions. Moreover, another pleasant feature is that the MPE is strongly consistent and asymptotically normal under the above assumption (see Billiot et al. (2008)). Finally, the above assumption is fulfilled for a large class of classical Gibbs models (see Billiot et al. (2008) for more details). Let us specify some examples, where references can be found in for example Möller and Waagepetersen (2004).

(i) The homogeneous Poisson point process can be recovered by setting $\mathbb{M} = \{0\}$ and, for any $x \in \mathbb{R}^d$, $V(x | \varphi; \theta) = \theta_1$, for $\theta_1 \in \mathbb{R}$. Its intensity is then $\exp(-\theta_1)$.

(ii) Strauss point process: let $\mathbb{M} = \{0\}$ and $R > 0$. For any $x \in \mathbb{R}^d$, its local energy function is defined by $V(x | \varphi; \theta) = \theta_1 + \theta_2 n_{[0, R]}(x, \varphi)$ where
represents the number of \( R \)-closed neighbours of \( x \) in \( \varphi \). The Strauss model satisfies the above assumption when \( \theta_1 \in \mathbb{R} \) and \( \theta_2 \geq 0 \), with any \( D \geq R \) (the restriction \( \theta_2 \geq 0 \) can be discarded by assuming a hard-core condition). We also mention a standard generalization where the indicator function is replaced by a step function, leading to the so-called piecewise constant pairwise interaction point process.

(iii) The area interaction point process constitutes the main example of a point process with interaction of any order: let \( \mathbb{M} = \{ 0 \} \) and \( R > 0 \); then, for any \( x \in \mathbb{R}^d \), \( V(x|\varphi; \theta) = \theta_1 + \theta_2 \{ A(\varphi \cup x) - A(\varphi) \} \) where \( A(\varphi) = \bigcup_{y \in \varphi \cup B(x,R)} B(y,R) \). This model satisfies the above assumption for any \( \theta_1 \in \mathbb{R} \), \( \theta_2 \in \mathbb{R} \) and \( D \geq 2R \).

(iv) Multitype Strauss point process: this model is a discrete marked extension of the Strauss model. Let \( \mathbb{M} = \{ 1, 2 \} \) (the extension to more marks is straightforward), \( R_1,1 > 0 \), \( R_1,2 > 0 \), \( R_2,2 > 0 \) and \( \theta = (\theta_1^1, \theta_1^2, \theta_2^1, \theta_2^2, \theta_2^2) \); then, for any \( x \in \mathbb{R}^d \) and \( m \in \mathbb{M} \), the local energy function is defined by

\[
V(x^m|\varphi;\theta) = \theta_1^m + \sum_{m'=1}^2 \theta_2^{m,m'} \sum_{y \in \varphi} 1_{[0,R_2,0]}(\|y-x\|)
\]

(with the convention \( \theta_2^{2,1} = \theta_1^1 \) and \( R_2,1 = R_1,2 \)). The above assumption is fulfilled when \( \theta_1^m \in \mathbb{R} \) and \( \theta_2^{m,m'} \geq 0 \) for all \( m \) and \( m' \) and with any \( D \geq \max(R_1,1, R_1,2, R_2,2) \).

3. Innovations and residuals: definition and examples

The basic ingredient for the definition of residuals is the so-called Georgii–Nguyen–Zessin formula stated below.

**Theorem 1** (Georgii–Nguyen–Zessin formula). Under the existence assumption, for any measurable function \( h(\cdot,\cdot;\theta) : \mathcal{E} \times \Omega \to \mathbb{R} \) (depending on the parameter \( \theta \)) such that the following mean values are finite, we have

\[
E \left[ \int_{\mathbb{R}^d \times \mathbb{M}} h(x^m, \Phi; \theta) \exp\{-V(x^m|\Phi; \theta^*)\} \mu(dx^m) \right] = E \left[ \sum_{x^m \in \Phi} h(x^m, \Phi \setminus x^m; \theta^*) \right],
\]

where \( E \) denotes expectation with respect to \( P_{\theta^*} \).

The concepts of innovations and residuals that were proposed by Baddeley et al. (2005) are simply based on empirical versions of both terms appearing in the Campbell equilibrium equation (2).

**Definition 2.** For any bounded domain \( \Lambda \), we define the \( h \)-innovations (denoted by \( I_\Lambda \)) and the \( h \)-residuals (denoted by \( R_\Lambda \) and depending on an estimate \( \hat{\theta} \) of \( \theta^* \)) by

\[
I_\Lambda(\varphi;h,\theta^*) := \int_{\Lambda \times \mathbb{M}} h(x^m, \varphi; \theta^*) \exp\{-V(x^m|\varphi; \theta^*)\} \mu(dx^m) - \sum_{x^m \in \varphi \setminus \Lambda} h(x^m, \varphi \setminus x^m; \theta^*),
\]

\[
R_\Lambda(\varphi;h,\hat{\theta}) := \int_{\Lambda \times \mathbb{M}} h(x^m, \varphi; \hat{\theta}) \exp\{-V(x^m|\varphi; \hat{\theta})\} \mu(dx^m) - \sum_{x^m \in \varphi \setminus \Lambda} h(x^m, \varphi \setminus x^m; \hat{\theta}).
\]

From a practical point of view, the last notion is the most interesting since it provides a data-driven measure. The idea behind the definition of \( h \)-residuals is very simple.
(a) Consider a parametric model for the stationary marked point process $\Phi$ observed within $\Lambda$.
(b) Estimate the parameters by a classical method (maximum likelihood, MPE, . . .).
(c) Suspect an inadequacy of the model if $R_\Lambda$ is far from 0.

The main goal of our paper is to focus on the third step. First we prove that $|\Lambda|^{-1}R_\Lambda$ actually tends to 0 and we study its fluctuations. On the basis of these asymptotic results, we propose to implement the third step by some testing procedures for which the type I error is theoretically controlled.

Let us explain which kind of testing procedures may naturally be constructed. Throughout Baddeley et al. (2005, 2008), the main examples considered by the authors are

(i) the raw residuals with $h(x^m, \varphi; \theta) = 1$,
(ii) the inverse residuals with $h(x^m, \varphi; \theta) = \exp\{V(x^m|\varphi; \theta)\}$ and
(iii) the Pearson residuals with $h(x^m, \varphi; \theta) = \exp\{V(x^m|\varphi; \theta)/2\}$.

The raw residuals constitute a difference of two estimates of the intensity of the point process (up to a normalization by $|\Lambda|$): the first term depends on the model and so is a parametric estimate, whereas the second is a non-parametric estimate (since it is equal to $|\varphi_\Lambda|$). But this choice may be inappropriate to assess a homogeneous Poisson point process with intensity $\rho := \exp(-\theta)$. The raw residuals reduce to $\hat{\rho}|\Lambda| - |\varphi_\Lambda|$, which vanishes for the classical choice $\hat{\rho} = |\varphi_\Lambda|/|\Lambda|$. So in this case the raw residuals are irrelevant to validate the Poisson assumption. A better strategy to test the Poisson model is the quadrat counting test; see Diggle (2003) for instance. It basically consists of a comparison between the total number of points with the number of points in subdomains of $\Lambda$. Following the same idea, it is natural to consider residuals from subdomains of $\Lambda$, whereas the estimate $\hat{\theta}$ is computed from the full domain. This will be our first framework when studying the asymptotic behaviour of residuals. It is precisely specified in Section 4.1. This framework becomes relevant to assess a Poisson model with the raw residuals and actually leads to a generalization of the quadrat counting test to Gibbs processes, as shown in Section 5.1.1.

We can imagine more evolved test functions than those related to raw, inverse or Pearson residuals and which can reflect more complex characteristics of the point process. For example, consider the following test function parameterized by $r > 0$:

$$h_r(x^m, \varphi; \theta) := 1_{[0,r]}\{d(x^m, \varphi)\} \exp\{V(x^m|\varphi; \theta)\}$$  \(3\)

where $d(x^m, \varphi) = \inf_{x^m', \varphi'} \|y - x\|$. This choice leads to

$$R_\Lambda(\varphi; h_r, \hat{\theta}) = \int_{\Lambda \times \Phi} 1_{[0,r]}\{d(x^m, \varphi)\} \mu(dx^m) - \sum_{x^m \in \varphi_\Lambda} h_r(x^m, \varphi; x^m, \hat{\theta}).$$  \(4\)

Therefore $R_\Lambda(\varphi; h_r, \hat{\theta})/|\Lambda|$ corresponds to a balance between two estimates of the well-known empty space function $F$ at distance $r$ (recall that $F(r) := P\{d(0^M, \Phi) \leq r\}$; see Møller and Waagepetersen (2004) for instance). The first term on the right-hand side of equation (4) corresponds to the natural non-parametric estimate of $F(r)$ whereas the second is a parametric estimate of $F(r)$. In what follows, $h$-residuals defined by expression (3) will be called with a slight abuse empty space residuals. We may compute them for various values of $r$ and construct a test combining all of them. This constitutes our second framework when studying the asymptotic behaviour of residuals. It leads to an alternative testing procedure, which is described in Section 5.1.3.
4. Asymptotic properties of the residuals

4.1. Frameworks and assumptions

Following the existence condition, we assume that a realization of $\Phi_{\Lambda_n^+}$ is observed where $\Phi \sim P_{\theta^*}$ and where $\Lambda_n^+$ is a cube in $\mathbb{R}^d$.

According to the finite range locality assumption, the $h$-innovations and $h$-residuals can be computed for $\Lambda = \Lambda_n$, where $\Lambda_n = \Lambda_n^+ \ominus D$ is the erosion of $\Lambda_n^+$ by $D$, namely

$$\Lambda_n = \{ x \in \Lambda_n^+, B(x, D) \subseteq \Lambda_n^+ \}. \quad (5)$$

The aim of this section is to present several asymptotic properties for $I_{\Lambda_n}$ and $R_{\Lambda_n}$ when $\Lambda_n \to \mathbb{R}^d$ as $n \to \infty$. We prove their consistency and we propose asymptotic normality results within the two frameworks motivated earlier.

(a) **Framework 1**: for a fixed test function $h$, $\Lambda_n$ is a cube, divided into a fixed finite number of subcubes (which will increase with $\Lambda_n$). The purpose is then to obtain the asymptotic normality for the vector composed of the $h$-residuals computed in each subcube.

(b) **Framework 2**: we consider $h_1, \ldots, h_s$ different test functions and the aim is to obtain the asymptotic normality of the vector composed of the $h_j$-residuals computed on $\Lambda_n$.

In both frameworks, an estimate of $\theta^*$ is involved. We assume that it is computed from $\Phi_{\Lambda_n^+}$, i.e. with the same data as used to evaluate the $h$-residuals. Moreover, in contrast with previous works dealing with asymptotic properties on Gibbs point processes, where $\Lambda_n$ is assumed to be of the discrete form $[-n, n]^d$, we consider general domains that may grow continuously up to $\mathbb{R}^d$. This seems to be the most natural setting when dealing with point processes.

The asymptotic results depend on three different types of assumption. The first is related to the model, the second to the test function(s) $h$ and the third to the estimate $\hat{\theta}$ that is used to compute the $h$-residuals. To simplify the presentation, we consider exponential family models through the exponentiality assumption and we make the following assumptions.

$h$-assumption. The test function(s) correspond either to the raw residuals, the inverse residuals, the Pearson residuals or the empty space residuals.

Maximum pseudolikelihood estimate assumption. We have that $\theta^* \in \Xi$ for some bounded open set $\Xi \subseteq \theta$. The estimate $\hat{\theta}_n$ corresponds to the MPE defined as $\hat{\theta}_n = \hat{\theta}_n(\varphi) = \arg \max_{\theta \in \Xi} LPL_{\Lambda_n}(\varphi; \theta)$, where

$$LPL_{\Lambda_n}(\varphi; \theta) = - \int_{\Lambda_n \times \mathbb{M}} \exp\{ -\theta^T \varphi(x^m|\varphi) \} \mu(dx^m) - \theta^T \sum_{x^m \in \varphi_{\Lambda_n}} \varphi(x^m|\varphi \setminus x^m). \quad (6)$$

The main results of this paper are presented under these assumptions for clarity. But they are true in a larger setting. We refer to Appendix B.1 for a description of general assumptions required for the asymptotic results that are presented hereafter.

4.2. Consistency and asymptotic normality

We first present consistency results for $I_{\Lambda_n}(\Phi; h, \theta^*)$ and $R_{\Lambda_n}(\Phi; h, \hat{\theta}_n(\Phi))$, where, for all $n \geq 1$, $(\Lambda_n)_{n \geq 1}$ are cubes and $\Lambda_n \to \mathbb{R}^d$ as $n \to \infty$.

**Proposition 1.** Under the exponentiality, $h$- and MPE assumptions $|\Lambda_n|^{-1} I_{\Lambda_n}(\Phi; h, \theta^*)$ and $|\Lambda_n|^{-1} R_{\Lambda_n}(\Phi; h, \hat{\theta}_n(\Phi))$ converge almost surely towards 0 as $n \to \infty$.

To state the asymptotic normality under the two preceding frameworks, we introduce the notation
where, for any \( \Lambda \), \( R_{\infty, \Lambda}(\varphi; h, \theta) := I_{\Lambda}(\varphi; h, \theta) - U_{\Lambda}(\varphi; \theta)^{T} E(h; \theta). \) (7)

Here, \( E(h; \theta) \) is the vector defined by

\[
E(h; \theta) := E[h(0^M, \Phi; \theta) V^{(1)}(0^M|\Phi; \theta) \exp\{-V(0^M|\Phi; \theta)\}],
\]

where, for any \( x^m \) and \( \varphi \), \( V^{(1)}(x^m|\varphi; \cdot) \) denotes the gradient vector function of \( \theta \mapsto V(x^m|\varphi; \theta) \), which is equal to \( v(x^m|\varphi) \) under the exponentiality assumption. Moreover, the vector \( U_{\Lambda} \) is related to the asymptotic expansion of the estimate \( \hat{\theta}_n \)

\[
\hat{\theta}_n(\Phi) - \theta^* = \frac{1}{|\Lambda_n|} U_{\Lambda_n}(\Phi; \theta^*) + o_P(|\Lambda_n|^{-1/2}).
\]

In particular, under the MPE assumption, we may prove (see proposition 6 in Appendix B.1) that

\[
U_{\Lambda}(\varphi; \theta^*) = H(\theta^*)^{-1} LPL_{\Lambda}^{(1)}(\varphi; \theta^*),
\]

where \( LPL_{\Lambda}^{(1)}(\varphi; \theta^*) \) is the gradient vector of the log-pseudolikelihood, i.e.

\[
LPL_{\Lambda}^{(1)}(\varphi; \theta^*) := \int_{\Lambda \times |\Lambda|} v(x^m|\varphi) \exp\{-\theta^*^{T}v(x^m|\varphi)\} \mu(dx^m) - \sum_{x^m \in \Lambda} v(x^m|\varphi \setminus x^m)
\]

and where \( H(\theta^*) \) is the symmetric matrix

\[
H(\theta^*) := E[v(0^M|\Phi) v(0^M|\Phi)^{T} \exp\{-\theta^*^{T}v(0^M|\Phi)\}].
\]

### 4.2.1. Asymptotic normality of the \( h \)-residuals computed on subdomains of \( \Lambda_n \) (framework 1)

In framework 1, we are given a test function \( h \) and compute the \( h \)-residuals on disjoint subdomains of \( \Lambda_n \). Further, we assume that the domain \( \Lambda_n \) is a cube and is divided into a fixed number of subdomains as follows:

\[
\Lambda_n := \bigcup_{j \in J} \Lambda_{j,n}
\]

where \( J \) is a finite set and all \( \Lambda_{j,n} \) are disjoint cubes with the same volume \( |\Lambda_{0,n}| \) increasing up to \( \infty \). Let us denote by \( R_{\infty, J,n}(\varphi; h, \hat{\theta}_n) = (R_{\Lambda_{j,n}}(\varphi; h, \hat{\theta}_n))_{j \in J} \) the vector of the residuals computed on each subdomain and by \( \Delta_k(D) \) for \( k \in \mathbb{Z}^d \) the cube centred at \( kD \) with side length \( D \). Assuming that \( P_{\theta^*} \) is ergodic (see remark 1 below for a discussion), we have the asymptotic normality of \( R_{\infty, J,n}(\varphi; h, \hat{\theta}_n) \).

**Proposition 2.** Under framework 1, assume the exponentiality, \( h \)- and MPE conditions; then for any ergodic measure \( P_{\theta^*} \) the following convergence in distribution holds, as \( n \to \infty \):

\[
|\Lambda_{0,n}|^{-1/2} R_{\infty, J,n}(\varphi; h, \hat{\theta}_n) \overset{d}{\to} N(0, \Sigma_{J}(\theta^*)) ,
\]

where \( \Sigma_{J}(\theta^*) = \lambda_{\text{inn}} I_{|J|} + |J|^{-1}(\lambda_{\text{res}} - \lambda_{\text{inn}}) J \) where \( I_{|J|} \) is the \( |J| \times |J| \) identity matrix and where \( J = ee^{T} \) and \( e = (1, \ldots, 1)^{T} \). The constants \( \lambda_{\text{inn}} \) and \( \lambda_{\text{res}} \) are respectively defined by

\[
\lambda_{\text{inn}} = D^{-d} \sum_{|k| \leq 1} E[I_{\Delta_0(D)}(\Phi; h, \theta^*) I_{\Delta_k(D)}(\Phi; h, \theta^*)],
\]

\[
\lambda_{\text{res}} = D^{-d} \sum_{|k| \leq 1} E[R_{\infty, \Delta_0(D)}(\Phi; h, \theta^*) R_{\infty, \Delta_k(D)}(\Phi; h, \theta^*)].
\]
4.2.2. Asymptotic normality of the \((h_j)_{j=1,...,s}\) residuals computed on \(\Lambda_n\) (framework 2)

In framework 2, we consider \(s\) different test functions and we compute all \(h_j\)-residuals on the same domain \(\Lambda_n\), which is assumed to be a cube growing up to \(\mathbb{R}^d\) when \(n \to \infty\).

We present an asymptotic normality result for the random vector \((R_{\Lambda_n}(\hat{\Phi}; h_j, \hat{\theta}_n))_{j=1,...,s}\).

Proposition 3. Under framework 2, assume the exponentiality, \(h\)- and MPE conditions; then for any ergodic measure \(P_{\theta^*}\) the following convergence in distribution holds, as \(n \to \infty\):

\[
|\Lambda_n|^{-1/2}(R_{\Lambda_n}(\Phi; h_j, \hat{\theta}_n))_{j=1,...,s} \xrightarrow{d} \mathcal{N}(0, \Sigma_2(\theta^*)),
\]

where \(\Sigma_2(\theta^*)\) is the \((s,s)\) matrix given by

\[
\Sigma_2(\theta^*) = D^{-d} \sum_{|k| \leq 1} E[R_{\infty,\Delta_0(D)}(\Phi; h, \theta^*) R_{\infty,\Delta_k(D)}(\Phi; h, \theta^*)^T],
\]

and \(R_{\infty,\Lambda}(\varphi; h, \theta^*) := (R_{\infty,\Lambda}(\varphi; h_j, \theta^*))_{j=1,...,s}\).

Remark 1. The asymptotic normality stated in propositions 2 and 3 holds only if \(P_{\theta^*}\) is an ergodic measure. We refer to Preston (1976) for a definition and further properties. If the measure \(P_{\theta^*}\) is not ergodic, then it can be represented as a mixture of ergodic measures. In this case, the asymptotic distributions in expressions (11) and (14) become a mixture of Gaussian distributions. Following Jensen and Künsch (1994), if we renormalize the left-hand sides of expressions (11) and (14) by some empirical estimate of their respective asymptotic covariance matrix, then this mixture becomes a standard Gaussian distribution. Hence the assumption that \(P_{\theta^*}\) is an ergodic measure will no longer be required for the testing procedures displayed in corollaries 1–3.

5. Goodness-of-fit tests for stationary marked Gibbs point processes

We present three goodness-of-fit tests, based on the residuals computed according to the two different frameworks that were considered in Section 4. Then we discuss their range of validity. It mainly depends on whether the asymptotic covariance matrices that are involved in the previous section are invertible. This requirement is always satisfied for the first test (the generalization of the quadrat counting test), but it is a restriction for the two other tests, depending on the model and the test function. A practical criterion is given. One must also estimate these matrices to implement the tests. We briefly present some estimates though other choices are possible. A short simulation study concludes this section.

5.1. Testing procedures

5.1.1. Quadrat-type test with \(|J| - 1\) degrees of freedom

According to the first framework, we divide the domain \(\Lambda_n\) into a fixed number of subdomains, namely \(\Lambda_n := \bigcup_{j \in J} \Lambda_{j,n}\) where \(J\) is a finite set and all the \(\Lambda_{j,n}\) are disjoint cubes with the same volume \(|\Lambda_{0,n}|\).

Corollary 1. Assume the exponentiality, \(h\)- and MPE conditions. Then \(\lambda_{\text{Inn}} > 0\). Moreover, if \(\hat{\lambda}_{n,\text{Inn}}\) is a consistent estimate of \(\lambda_{\text{Inn}}\), then

\[
T_{1,n} := |\Lambda_{0,n}|^{-1} \hat{\lambda}_{n,\text{Inn}}^{-1} \|R_{\mathcal{J},n}(\Phi; h, \hat{\theta}_n) - \tilde{R}_{\mathcal{J},n}(\Phi; h, \hat{\theta}_n)\|^2 \xrightarrow{d} \chi^2(|\mathcal{J}| - 1),
\]

where \(\tilde{R}_{\mathcal{J},n}(\varphi; h, \hat{\theta}_n) = |\mathcal{J}|^{-1} \sum_{j \in \mathcal{J}} R_{\Lambda_{j,n}}(\varphi; h, \hat{\theta}_n)\).
This result leads to a goodness-of-fit test for $H_0 : \Phi \sim P_{\theta^*}$ versus $H_1 : \Phi \sim P_{\theta^*}$. We briefly summarize the different steps to implement the test for a given asymptotic level $\alpha \in (0, 1)$.

**Step 1**: let $\varphi$ be a marked point pattern observed on a cube $\Lambda_n^+$. Assume that $\varphi$ is the realization of a marked Gibbs point process satisfying the exponentiality assumption. Compute the MPE $\theta_n$ from expression (6) where $\Lambda_n = \Lambda_n^+ \ominus D$ and where $D$ is a real number not lower than the range of interaction.

**Step 2** (framework 1):

(a) consider a test function $h$ satisfying the $h$-assumption, divide $\Lambda_n$ into $|J|$ cubes and compute the $h$-residuals on each different cube;
(b) estimate $\hat{\lambda}_{\text{inn}}$ and compute the test statistic $T_{1,n}$ that is defined in expression (16).

**Step 3**: reject the model if $T_{1,n}(\varphi) > \chi^2_{1-\alpha}(|J| - 1)$.

Note that, in the particular case of a homogeneous Poisson point process with intensity $z$ and when considering the raw residuals ($h = 1$), this test is exactly the Poisson dispersion test applied to the $|J|$ quadrat counts, which is also called the quadrat counting test. Indeed, in this case, $R_{\varphi/h, n}(\varphi; h, \theta_n) - R_{\varphi/h, n}(\varphi; h, \theta_n)$ is the vector of quadrat counts and $\hat{\lambda}_{\text{inn}} = z$. Considering $|\Lambda_{0,n}|\hat{\lambda}_{\text{inn}}$ as an estimator of the intensity on $\Lambda_{0,n}$, the statistic $T_{1,n}$ reduces to the ratio of the sum of squares of the quadrat counts over their estimated mean.

5.1.2. **Quadrat-type test with $|J|$ degrees of freedom**

**Corollary 2.** Assume the exponentiality, $h$- and MPE conditions and $\lambda_{\text{Res}} > 0$. Let $\hat{\lambda}_{\text{inn}}$ and $\hat{\lambda}_{\text{Res}}$ be consistent estimators of $\lambda_{\text{inn}}$ and $\lambda_{\text{Res}}$ and let

$$\tilde{R}_{1,n}(\varphi; h, \hat{\theta}_n) := \hat{\lambda}_{\text{inn}}^{-1/2} R_{\varphi/h, n}(\varphi; h, \hat{\theta}_n) + (\hat{\lambda}_{\text{Res}} - \hat{\lambda}_{\text{inn}}^{-1/2}) R_{\varphi/h, n}(\varphi; h, \hat{\theta}_n).$$

Then $\tilde{R}_{1,n}(\varphi; h, \hat{\theta}_n) = \sum_{J=1}^{n-1} R_{\varphi/h, n}(\varphi; h, \hat{\theta}_n)$ and, as $n \to \infty$,

$$T'_{1,n}(\Phi) := |\Lambda_{0,n}|^{-1} \|\tilde{R}_{1,n}(\Phi; h, \hat{\theta}_n)\|^2 \overset{d}{\to} \chi^2(|J|). \quad (17)$$

A goodness-of-fit test with asymptotic level $\alpha \in (0, 1)$ is deduced similarly to the previous section. The steps to follow in practice are the same except that, in step 2(b), we must estimate both $\hat{\lambda}_{\text{inn}}$ and $\hat{\lambda}_{\text{Res}}$, and in step 3 we reject the model if $T'_{1,n}(\varphi) > \chi^2_{1-\alpha}(|J|)$.

5.1.3. **Empty space function type test**

Let us consider the setting of proposition 3, where $s$ different residuals are computed on $\Lambda_n$.

**Corollary 3.** Assume the exponentiality, $h$- and MPE conditions. Assume also that $\Sigma_2(\theta^*)$ is positive definite. Let $\hat{\Sigma}_{2,n}$ be a consistent estimate of $\Sigma_2(\theta^*)$ and let

$$\tilde{R}_{2,n}(\varphi; h, \hat{\theta}_n) := \hat{\Sigma}_{2,n}^{-1/2} (R_{\varphi/h, n}(\varphi; h, \hat{\theta}_n))_{j=1,\ldots,s}.$$

Then as $n \to \infty$

$$T_{2,n}(\Phi) := |\Lambda_n|^{-1} \|\tilde{R}_{2,n}(\Phi; h, \hat{\theta}_n)\|^2 \overset{d}{\to} \chi^2(s). \quad (18)$$

A goodness-of-fit test for $H_0 : \Phi \sim P_{\theta^*}$ versus $H_1 : \Phi \sim P_{\theta^*}$, with asymptotic level $\alpha \in (0, 1)$, is deduced as before. From a practical point of view, the steps detailed above are modified as follows.
Step 1: let \( \varphi \) be a marked point pattern observed on a cube \( \Lambda_n^+ \). Assume that \( \varphi \) is the realization of a marked Gibbs point process satisfying the exponentiality assumption. Compute the MPE \( \hat{\theta}_n \) from expression (6) where \( \Lambda_n = \Lambda_n^+ \ominus D \) and where \( D \) is a real number not lower than the range of interaction.

Step 2 (framework 2):

(a) consider \( s \) different test functions satisfying the \( h \)-assumption and compute the \( s \) different \( h_j \)-residuals on \( \Lambda_n \);
(b) estimate \( \sum_2(\theta^*) \) and compute \( \sum_2^{-1/2} \) by any numerical routine (e.g. Choleski decomposition or singular value decomposition).

Step 3: fix \( \alpha \in (0, 1) \) and reject the model if \( T_{2,n}(\varphi) > \chi^2_{1-\alpha}(s) \).

5.2. Discussion
5.2.1. Positive definiteness of covariance matrices
Whereas \( \lambda_{\text{inn}} > 0 \) is implied by the exponentiality and \( h \)-assumptions (see lemma 4 in Appendix B.9, both \( A_{\text{Res}} \) and \( \sum_2(\theta^*) \) may fail to be positive definite for an inappropriate choice of test function, as shown by the following proposition.

Proposition 4. Under the exponentiality and MPE assumptions, if \( h(x^m, \varphi; \theta) = \omega^T v(x^m | \varphi) \) for some \( \omega \in \mathbb{R}^p \setminus \{0\} \), then \( \lambda_{\text{Res}} = 0 \) and the matrices \( \sum_1(\theta) \) and \( \sum_2(\theta^*) \) in propositions 2 and 3 are only positive semidefinite matrices.

As a consequence of proposition 4, the two goodness-of-fit tests based on \( T'_{1,n} \) and \( T_{2,n} \) in sections 5.1.2 and 5.1.3 are not available if the test function \( h \) is a linear combination of the sufficient statistics \( v(x^m | \varphi) \). Since for most models that are used in practice, including examples that were presented in Section 2.3, the value 1 can be obtained from a linear combination of \( v(x^m | \varphi) \), the raw residuals (\( h = 1 \)) are not an appropriate choice for these two tests.

We present below a general criterion which ensures that \( \sum_2(\theta^*) \) is positive definite (or \( \lambda_{\text{Res}} > 0 \)).

Positive definiteness assumption. For some \( \bar{\Delta} := \bigcup_{|k| \leq 1} \Delta_k(D) \), there exists \( B \in \mathcal{F} \) and \( A_0, \ldots, A_l, l \geq 1 \), disjoint events of \( \bar{\Omega}_B := \{ \varphi \in \Omega : \varphi_{\Delta_k(D)} \in B, 1 \leq |k| \leq 2 \} \) such that, for \( i = 0, \ldots, l \), \( P_{\theta^*}(A_j) > 0 \) and for any \( (\varphi_0, \ldots, \varphi_l) \in A_0 \times \cdots \times A_l \), the \( (l, s) \) matrix with entries \( (R_{\infty,\bar{\Delta}}(\varphi_i; h_j, \theta^*))_{i=1,\ldots,l, j=1,\ldots,s} \), is injective, i.e., for any \( y \in \mathbb{R}^s \),

\[
(\forall i \in \{1, \ldots, s\}, y^T(R_{\infty,\bar{\Delta}}(\varphi_i; h, \theta^*) - R_{\infty,\bar{\Delta}}(\varphi_0; h, \theta^*)) = 0) \Rightarrow y = 0,
\]

where \( R_{\infty,\bar{\Delta}}(\varphi; h, \theta^*) = (R_{\infty,\bar{\Delta}}(\varphi; h_j, \theta^*))_{j=1,\ldots,s} \).

Proposition 5. Under the positive definiteness assumption, the matrix \( \sum_2(\theta^*) \) is positive definite. In particular, for \( s = 1 \), \( \lambda_{\text{Res}} > 0 \).

The positive definiteness assumption is associated with some characteristics of the point process \( \Phi \). Given a model, the event \( B \) is chosen to let the different configurations sets \( A_0, A_1, \ldots, A_l \) be as simple as possible. For most models, a convenient choice is \( B = \{\emptyset\} \) where \( \emptyset \) denotes the empty marked point configuration. A typical application of the positive definiteness assumption is implemented in the proof of lemma 4. Nevertheless, checking the positive definiteness assumption in practice depends on the model and/or the test function. As an example, we have for the 2-type marked Strauss point process and the area interaction point process that
(a) the matrix $\Sigma_1(\theta^*)$ that is obtained in framework 1 from the inverse residuals $h = \exp(V)$ is positive definite and
(b) the matrix $\Sigma_2(\theta^*)$ that is obtained in framework 2 from the empty space residuals is positive definite.

These results are proved in the on-line supporting information for this paper. Note that, by proposition 4, these results are not valid when considering the raw residuals for these models.

5.2.2. Estimating the covariance matrices

Assume that the marked point process is observed in the cube $\Lambda_n^+$. Let $\Lambda_n = \Lambda_n^+ \ominus D$ where $\Lambda_n = \bigcup_{k \in \mathcal{K}_n} \Delta_k(D)$ for some finite set $\mathcal{K}_n \subset \mathbb{Z}^d$ and the $\Delta_k(D)$s are disjoint cubes with side length $D$ and centred, without loss of generality, at $kD$. According to the definitions of $\lambda_{\text{Inn}}$, $\lambda_{\text{Res}}$ and $\Sigma_2(\theta^*)$, we may propose the following (natural) empirical estimates:

\[
\hat{\lambda}_{n,\text{Inn}}(\phi; h, \hat{\Theta}_n) = [\Lambda_n]^{-1} \sum_{i \in \mathcal{K}_n} \sum_{|j-i| \leq 1, j \in \mathcal{K}_n} I_{\Delta_k(D)}(\phi; h, \hat{\Theta}_n) I_{\Delta_j(D)}(\phi; h, \hat{\Theta}_n),
\]

\[
\hat{\lambda}_{n,\text{Res}}(\phi; h, \hat{\Theta}_n) = [\Lambda_n]^{-1} \sum_{i \in \mathcal{K}_n} \sum_{|j-i| \leq 1, j \in \mathcal{K}_n} \hat{R}_{\infty,\Delta_k(D)}(\phi; h, \hat{\Theta}_n) \hat{R}_{\infty,\Delta_j(D)}(\phi; h, \hat{\Theta}_n),
\]

\[
\hat{\Sigma}_2(n)(\phi; h, \hat{\Theta}_n) = [\Lambda_n]^{-1} \sum_{i \in \mathcal{K}_n} \sum_{|j-i| \leq 1, j \in \mathcal{K}_n} \hat{R}_{\infty,\Delta_k(D)}(\phi; h, \hat{\Theta}_n) \hat{R}_{\infty,\Delta_j(D)}(\phi; h, \hat{\Theta}_n) T,
\]

where $\hat{R}_{\infty,\Delta}(\phi; h, \theta) = I_{\Lambda}(\phi; h, \theta) - \mathbf{LPL}_n^{(1)}(\phi; \theta) \hat{H}_n^{-1}(\phi; \theta) \hat{\mathcal{E}}(\phi; h, \theta)$ and

\[
\hat{H}_n(\phi; \theta) = [\Lambda_n]^{-1} \int_{\Lambda_n \times \mathbb{R}} v(x^m|\phi) v(x^m|\phi)^T \exp\{-\theta^T v(x^m|\phi)\} \mu(dx^m),
\]

\[
\hat{\mathcal{E}}_n(\phi; h, \theta) = [\Lambda_n]^{-1} \int_{\Lambda_n \times \mathbb{R}} h(x^m, \phi, \theta) v(x^m|\phi) \exp\{-\theta^T v(x^m|\phi)\} \mu(dx^m).
\]

Using an analytic calculation of the covariances of the innovations, Coeurjolly and Rubak (2012) have proposed covariance estimates which are easily calculated. We mention here, for instance, how $\lambda_{\text{Inn}}$ can be consistently estimated when using the raw residuals ($h = 1$), which will be the test function in our simulation:

\[
\hat{\lambda}_{n,\text{Inn}}(\phi; h, \hat{\Theta}_n) = \frac{[\phi \Delta_n]}{[\Lambda_n]} + \frac{1}{[\Lambda_n]} \sum_{x^m, y^m \in \phi \Delta_n} \{\exp(-\hat{\Theta}_n^T v(x^m|\phi \setminus \{x^m, y^m\}) - v(x^m|\phi \setminus x^m)) - 1\}.
\]

We do not want to enter more into detail but we claim that, under the exponentiality, $h$- and MPE assumptions the various estimates that were introduced above are consistent. Then corollaries 1–3 can be applied with these choices of estimate.

5.3. A simulation study

We illustrate the use of the quadrat-type test that was introduced in Section 5.1.1, for the raw residuals ($h = 1$). A thorough simulation study would be necessary to assess properly the three testing procedures and to compare them. This will be the subject of a separate paper and is outside the intended scope of this section.

We consider the null hypothesis of a Strauss model with range $R > 0$ (see Section 2.3 for the definition). The estimation of the eigenvalue $\lambda_{\text{Inn}}$ in expression (16) is implemented according
Fig. 1. QQ-plots of the quadrat test statistics with (a) \(|\mathcal{J}| = 4\) and (b) \(|\mathcal{J}| = 9\), based on 500 realizations of Strauss point processes with parameters \(R = 0.05\) and \(\beta := \exp(-\theta_1) = 200\) and interaction parameter \(\gamma := \exp(-\theta_2)\) equal to (a) 0.7 and (b) 0.3: the simulated realizations are generated in \([-R, L + R]^2\) for \(L = 1\) (○), 2 (△), 3 (×); the theoretical quantiles come from a \(\chi^2(|\mathcal{J}| - 1)\) distribution.
to equation (19), which gives

$$\lambda_{n,\text{Inn}}(\varphi; h, \hat{\theta}_n) = \frac{|\varphi \Lambda_n|}{|\Lambda_n|} + \frac{\exp(\hat{\theta}_2) - 1}{|\Lambda_n|} \sum_{x \in \varphi \Lambda_n} n_{[0, R]}(x, \varphi \backslash x),$$

where \(n_{[0, R]}(x, \varphi)\) is the number of \(R\)-closed neighbours of \(x\) in \(\varphi\).

First, to illustrate the convergence that is stated in corollary 1, we generated 500 Strauss point processes on the domains \([-R, L + R]^2\) for \(L = 1, 2, 3\), for various values of the parameters \(\theta_1\) and \(\theta_2\) and for \(R = 0.05\). The simulations have been carried out by using the R package spatstat (Baddeley and Turner, 2005). The parameters are estimated by the pseudolikelihood method (6) where \(\Lambda_n = [0, L]^2\). For each point pattern, the quadstat test statistic (on the left-hand side of expression (16)) is computed from two partitions: \(|\mathcal{J}| = 4\) or \(|\mathcal{J}| = 9\). For each value of the parameters, \(L\) and \(|\mathcal{J}|\), a QQ-plot of the 500 quadstat test statistics versus the theoretical quantiles of a \(\chi^2(|\mathcal{J}| - 1)\) distribution is plotted. Fig. 1 displays some results. We can observe that the asymptotic result in expression (16) is confirmed, even for domains with moderate volumes.

Second, the quadstat-type test (with \(h = 1\)) is implemented with type I error \(\alpha = 5\%\) to test for the null hypothesis of a homogeneous Strauss model. Table 1 presents the percentage rejection of the null hypothesis from two partitions (\(|\mathcal{J}| = 4\) or \(|\mathcal{J}| = 9\)) and \(L = 1, 2, 3\) in the presence of

(a) Strauss point processes with different sets of parameters \(\theta_1\) and \(\theta_2\),
(b) piecewise constant pairwise interaction point processes, namely Gibbs point processes with \(V(x|\varphi; \theta) = \theta_1 + \theta_2 n_{[0, R]}(x, \varphi) + \theta_3 n_{[R_1, R_2]}(x, \varphi),\)
(c) an inhomogeneous Strauss point process, with first-order interaction function \(\beta(\cdot)\) given by

| Table 1. Percentage of rejection of the Strauss model hypothesis, with the quadstat-type test with \(h = 1\) and \(|\mathcal{J}| = 4\) or \(|\mathcal{J}| = 9\)† |
|---------------------------------------------------------------|
| **Results (%) for |\(\mathcal{J}| = 4\) | **Results (%) for |\(\mathcal{J}| = 9\) |
| \(L = 1\) | \(L = 2\) | \(L = 3\) | \(L = 1\) | \(L = 2\) | \(L = 3\) |
| Strauss model, \(R = 0.05\) | | | | | |
| \(\exp(-\theta_1) = 200\), \(\exp(-\theta_2) = 1\) | 6.3 | 6 | 5.5 | 5.8 | 5.8 | 5.6 |
| \(\exp(-\theta_1) = 300\), \(\exp(-\theta_2) = 0.3\) | 4.8 | 5.4 | 5.0 | 4.0 | 4.8 | 5.4 |
| \(\exp(-\theta_1) = 300\), \(\exp(-\theta_2) = 0.7\) | 4.8 | 5.0 | 5.4 | 4.4 | 4.5 | 5.5 |
| Pairwise step function, \((R_1, R_2) = (0.06, 0.1)\) | | | | | |
| \(\exp(-\theta_1) = 500\), \(\exp(-\theta_2) = 0.3\), \(\exp(-\theta_3) = 0.7\) | 1.6 | 2.4 | 3.0 | 1.0 | 2.8 | 2.8 |
| \(\exp(-\theta_1) = 500\), \(\exp(-\theta_2) = 0.7\), \(\exp(-\theta_3) = 0.3\) | 3.2 | 4.2 | 5.2 | 2.8 | 5 | 4.4 |
| Inhomogeneous Strauss model, \(R = 0.05\) | | | | | |
| \(\exp(-\theta_1) = 300\beta(\cdot)\), \(\exp(-\theta_2) = 0.3\) | 12.0 | 68.4 | 100.0 | 9.0 | 68.2 | 100.0 |
| Neyman–Scott process | | | | | |
| \(\rho = 30\), \(r = 0.6\) | 20.2 | 36.2 | 38.2 | 19.6 | 45 | 51.8 |
| \(\rho = 30\), \(r = 0.2\) | 53.8 | 56.2 | 57.8 | 65.4 | 83.8 | 86.2 |

†The simulations are based on 500 realizations of several models of point processes in the domain \([-D, L + D]^2\) for \(L = 1, 2, 3\), where \(D\) equals the (known) finite range of interaction. The parameters of the Strauss model are estimated by using the pseudolikelihood method (6) where \(\Lambda_n = [0, L]^2\). The realizations are obtained by using the R package spatstat.
\[ \beta(x_1, x_2) = \exp \left( \frac{6x_1 + 5x_2 - 7x_1^2 + 5x_1x_2 - 3x_2^2}{10L} \right) , \]

\((x_1, x_2) \in \mathbb{R}^2\), and interaction parameter \(\theta_2\) and

(d) Neyman–Scott processes (see Möller and Waagepetersen (2004) for a definition): the clusters are centred with respect to a Poisson process with intensity \(\rho\), and each cluster consists of five points uniformly sampled in a disc with radius \(r\).

We set \(D\) equal to the range of interaction, which is assumed to be known for all models. For the Neyman–Scott model, which has no finite range, we set \(D\) equal to the estimate of the range that is obtained by the profile pseudolikelihood method over the interval [0.02, \(L/5\)].

The first rows of Table 1 show that the 5% level is well respected, which confirms the results in Fig. 1. The testing procedure detects also very well alternatives such as the inhomogeneous Strauss model or the Neyman–Scott model, which provide point patterns that are significantly different from homogeneous Strauss models. However, this test cannot reject significantly a piecewise constant pairwise interaction point process. Other simulations, which are not shown here, tend to the same conclusion for other pairwise interaction point processes including the Diggle–Gratton model (see Diggle and Gratton (1984)) and Geyer’s saturation model (see for example Möller and Waagepetersen (2004) for a definition). Our opinion is that, to reject these kind of alternatives, a test based on several residuals (like in corollary 3) should perform better. This perspective will be explored in future work.

Acknowledgements

The authors are sincerely grateful to both referees for their numerous insightful comments and suggestions which considerably helped them in improving the presentation of a previous version. This research was supported by Joseph Fourier University of Grenoble funded by the grant ‘SpaComp’.

Appendix A: Central limit theorem

The following result is a central limit theorem for conditionally centred random fields. It generalizes theorem 2.1 in Jensen and Künsch (1994) to a non-stationary setting. A general result has been proved by Comets and Janzura (1998) for self-normalized sums, assuming a fourth-moment condition. Our result is in the same spirit but it is proved for triangular arrays and without self-normalization, which is well adapted to the residuals process framework. This allows us in particular to avoid the fourth-moment assumption.

We need the following additional notation. Let \(\mathbf{M}\) be a matrix. Denote by \(\|\mathbf{M}\|\) the Frobenius norm of \(\mathbf{M}\) defined by \(\|\mathbf{M}\|^2 = \text{tr}(\mathbf{M}^T \mathbf{M})\), where \(\text{tr}\) is the trace operator and \(\mathbf{M}^T\) is the transpose of \(\mathbf{M}\).

Theorem 2. Let \(X_{n,j}, n \in \mathbb{N}, j \in \mathbb{Z}^d\), be a triangular array field in a measurable space \(\mathcal{S}\). For \(n \in \mathbb{N}\), let \(\mathcal{K}_n \subset \mathbb{Z}^d\) and, for \(k \in \mathcal{K}_n\), assume that

\[ Z_{n,k} = f_{n,k}(X_{n,k+1}, j \in \mathcal{I}_0), \]

where \(\mathcal{I}_0 = \{ j \in \mathbb{Z}^d, |j| \leq 1 \}\) and \(f_{n,k} : \mathcal{S}_{\mathcal{I}_0}^T \rightarrow \mathbb{R}^p\). Let \(S_n = \sum_{k \in \mathcal{K}_n} Z_{n,k}\). If

(a) \(c_3 := \sup_{n \in \mathbb{N}} \sup_{k \in \mathcal{K}_n} E|Z_{n,k}|^3 < \infty\),

(b) \(\forall n \in \mathbb{N}, \forall k \in \mathcal{K}_n, E|Z_{n,k}|X_{n,j}, j \neq k| = 0\),

(c) \(|\mathcal{K}_n| \rightarrow \infty\) and \(|\partial \mathcal{K}_n|/|\mathcal{K}_n| \rightarrow 0\) as \(n \rightarrow \infty\), and

(d) there is a symmetric matrix \(\Sigma \succeq 0\) such that

\[ E \left\| \mathcal{K}_n^{-1} \sum_{k \in \mathcal{K}_n} \sum_{j \in \mathcal{K}_n} Z_{n,k} Z_{n,j}^T - \Sigma \right\| \rightarrow 0, \]
then $|K_n|^{-1/2}S_n \xrightarrow{d} \mathcal{N}(0, \Sigma)$ as $n \to \infty$.

**Proof.** Let us first assume that $\Sigma$ is a positive definite matrix (i.e. $\Sigma > 0$). According to Stein’s method (see also Bolthausen (1982)), it suffices to prove that, for all $e \in \mathbb{R}^p$ such that $\|e\| = 1$ and for all $\lambda \in \mathbb{R}$,

$$E[(i\lambda - e^T|K_n|^{-1/2}\Sigma^{-1/2}S_n) \exp(i\lambda e^T|K_n|^{-1/2}\Sigma^{-1/2}S_n)] \to 0,$$

where $i = \sqrt{-1}$. Denoting $u = i\lambda e$, this is equivalent to proving that, for all $u \in \mathbb{R}^p$,

$$E\left[\left(u - |K_n|^{-1/2}\Sigma^{-1/2}S_n\right)\exp(iu^T|K_n|^{-1/2}\Sigma^{-1/2}S_n)\right] \to 0.$$

We decompose the term $A$ in the same spirit as in Bolthausen (1982), Jensen and Künsch (1994) and Comets and Janzura (1998). We denote by $I_p$ the identity matrix of size $p$ and $S_n^2 = \sum_{j \in \mathcal{K}_n, |j-k| \leq 1} Z_{n,j}$. Noting that $u^T\Sigma^{-1/2}S_n = S_n^2u^T\Sigma^{-1/2}u$, the decomposition is $E[A] = E[A_1 - A_2 - A_3]$ where

$$A_1 = i\exp(iu^T|K_n|^{-1/2}\Sigma^{-1/2}S_n)(I_p - |K_n|^{-1/2}\Sigma^{-1/2}\sum_{k \in \mathcal{K}_n} Z_{n,k}S_n^2)^{1/2}u,$$

$$A_2 = \exp(iu^T|K_n|^{-1/2}\Sigma^{-1/2}S_n)|K_n|^{-1/2}\Sigma^{-1/2}\sum_{k \in \mathcal{K}_n} Z_{n,k}\{1 - \exp(-iu^T|K_n|^{-1/2}\Sigma^{-1/2}S_n)\} - iu^T|K_n|^{-1/2}\Sigma^{-1/2}Z_n^S.$$

The two last terms $A_2$ and $A_3$ can be handled as in Jensen and Künsch (1994): $E[A_2] = 0$ by assumption (b) and $|E[A_3]| \to 0$ from the same inequalities therein and the submultiplicative property of the Frobenius norm. These inequalities rely on two facts: for all $x \in \mathbb{R}$, $|1 - \exp(-ix) - ix| \leq x^2/2$, and, for all $n$ and for all $(k_1, k_2, k_3) \in \mathcal{K}_n$, $E[|Z_{n,k_1}Z_{n,k_2}Z_{n,k_3}|] \leq \sqrt{E[|Z_{n,k_1}|^3]E[|Z_{n,k_2}|^3]E[|Z_{n,k_3}|^3]}/3$ which is less than $c_3$ by assumption (a).

For $A_1$, we cannot use a mean ergodic theorem as in Jensen and Künsch (1994), but assumption (d) is sufficient. Indeed,

$$\|E[A_1]\| \leq \|u\| \|E[I_p - |K_n|^{-1/2}\Sigma^{-1/2}\sum_{k \in \mathcal{K}_n, |j-k| \leq 1} Z_{n,k}Z_n^T\Sigma^{-1/2}]\|,$$

which tends to 0 by assumption (d).

Now, if $\Sigma$ is not a positive definite matrix, one can find an orthonormal basis $(f_1, \ldots, f_p)$ of $\mathbb{R}^p$, where the $f_i$ are eigenvectors of $\Sigma$. We agree that, if $r < p$ denotes the rank of $\Sigma$, then $(f_1, \ldots, f_r)$ is a basis of the image of $\Sigma$, whereas $(f_{r+1}, \ldots, f_p)$ is a basis of its kernel.

We denote by $V_{im}$ the matrix whose columns are $(f_1, \ldots, f_r)$ and $V_{Ker}$ the matrix whose columns are $(f_{r+1}, \ldots, f_p)$. Similarly, for any $u \in \mathbb{R}^p$, we denote by $u_j$ its $j$th co-ordinate in the basis $(f_1, \ldots, f_p)$ and $u_{im} = (u_1, \ldots, u_r)$, $u_{Ker} = (u_{r+1}, \ldots, u_p)$. Hence $u = V_{im}u_{im} + V_{Ker}u_{Ker}$.

The convergence in law of $|K_n|^{-1/2}S_n$ to a Gaussian vector reduces to the convergence of $u^T|K_n|^{-1/2}S_n$ for all $u \in \mathbb{R}^p$. We have

$$u^T|K_n|^{-1/2}S_n = V_{im}^TV_{im}|K_n|^{-1/2}S_n + u_{Ker}^TV_{Ker}|K_n|^{-1/2}S_n.$$

From assumption (d) and since $V_{Ker}^T\Sigma V_{Ker} = 0$, we deduce that

$$E\left[\||K_n|^{-1/2}\sum_{j \in \mathcal{K}_n, |j-k| \leq 1} V_{Ker}^T Z_{n,k}Z_n^T V_{Ker}\right] \to 0,$$

which means that $V_{Ker}^T|K_n|^{-1/2}S_n \to 0$ in quadratic mean.
However, the assumptions of theorem 2 imply that conditions (a)–(d) remain true when we replace \( Z_{n,k} \) by \( V_{n,m}^{T} Z_{n,k} \) and \( \Sigma \) by \( V_{n,m}^{T} \Sigma V_{n,m} \). Since \( V_{n,m}^{T} \Sigma V_{n,m} \) is positive definite, the convergence in law of \( V_{n,m}^{T} |K_n|^{-1/2} S_n \) holds for the same reasons as in the first part of the proof.

Therefore, we have proved that, for all \( u \in \mathbb{R}^p \),
\[
 u^{T} |K_n|^{-1/2} S_n \overset{d}{\to} N(0, u^{T} V_{n,m}^{T} \Sigma V_{n,m} u_{n,m}).
\]

It is easy to check that \( u^{T} V_{n,m}^{T} \Sigma V_{n,m} u_{n,m} = u^{T} \Sigma u \), which concludes the proof.

Appendix B: Proofs

B.1. General assumptions

All theoretical results in Sections 4 and 5 are stated for simplicity under the exponentiality, \( h \)- and MPE assumptions. But they remain true under more general assumptions as described below.

First, the consistency of the innovations and residuals (proposition 1) holds if the parametric Gibbs model satisfies the existence and finite range assumptions and the following assumptions are satisfied.

**Assumption 1.** For all \( x^{m} \in \mathcal{S} \), for all \( \varphi \in \Omega \), \( h(x^{m}, \varphi; \theta^{*}) = h(0^{m}, \tau, \varphi; \theta^{*}) \) and
\[
 E[|h(0^{M}, \Phi; \theta^{*})| \exp \{-V(0^{M}|\Phi; \theta^{*})\}] < \infty.
\]

**Assumption 2.** For all \( (m, \varphi) \in \mathcal{M} \times \Omega \), the functions \( h(0^{m}, \varphi; \theta) \) and \( f(0^{m}, \varphi; \theta) := h(0^{m}, \varphi; \theta) \exp \{-V(0^{m}|\varphi; \theta)\} \) are continuously differentiable with respect to \( \theta \) in a neighbourhood \( \mathcal{V}(\theta^{*}) \) of \( \theta^{*} \) and
\[
 E[||f^{(1)}(0^{M}, \Phi; \theta^{*})||] < \infty
\]
and
\[
 E[||h^{(1)}(0^{M}, \Phi; \theta^{*})|| \exp \{-V(0^{M}|\Phi; \theta^{*})\}] < \infty,
\]
where \( f^{(1)} \) denotes the gradient vector of \( f \) with respect to \( \theta \).

**Assumption 3.** \( \hat{\theta}_{n}(\Phi) \) converges almost surely towards \( \theta^{*} \) as \( n \to \infty \).

Asymptotic normality (propositions 2 and 3) holds if we assume in addition the following conditions.

**Assumption 4.** For all \( (m, \varphi) \in \mathcal{M} \times \Omega \), the functions \( h(0^{m}, \varphi; \theta) \) and \( f(0^{m}, \varphi; \theta) \) (defined in assumption 1) are twice continuously differentiable with respect to \( \theta \) in a neighbourhood \( \mathcal{V}(\theta^{*}) \) of \( \theta^{*} \) and
\[
 E[||f^{(2)}(0^{M}, \Phi; \theta^{*})||] < \infty
\]
and
\[
 E[||h^{(2)}(0^{M}, \Phi; \theta^{*})|| \exp \{-V(0^{M}|\Phi; \theta^{*})\}] < \infty,
\]
where
\[
 g^{(2)}(0^{m}, \varphi; \theta^{*}) = \left( \frac{\partial^{2}}{\partial \theta_{j} \partial \theta_{k}} g(0^{m}, \varphi; \theta^{*}) \right)_{1 \leq j, k \leq p}
\]
for \( g = f, h \).

**Assumption 5.** For any bounded domain \( \Lambda \), \( E[|I_{\Lambda}(\Phi; h, \theta^{*})|^2] < \infty \).

**Assumption 6.** For any sequence of bounded domains \( \Gamma_{n} \) such that \( \Gamma_{n} \to 0 \) when \( n \to \infty \),
\[
 E[I_{\Lambda_{n}}(\Phi; h, \theta^{*})^2] \to 0.
\]

**Assumption 7.** For any \( \varphi \in \Omega \) and any bounded domain \( \Lambda \), \( I_{\Lambda}(\varphi; \theta^{*}) \) depends only on \( \varphi_{\Lambda^{+}} \) where \( \Lambda^{+} \) is such that \( \Lambda = \Lambda^{+} \cup D \).

**Assumption 8.** The estimate admits the expansion
\[
 \hat{\theta}_{n}(\Phi) - \theta^{*} = \frac{1}{|\Lambda_{n}|} U_{\Lambda_{n}}(\Phi; \theta^{*}) + o_{P}(|\Lambda_{n}|^{-1/2}),
\]
where
such that \( \Lambda = \Lambda^+ \cap D \).

**Remark 2.** Assumption 8 implies a central limit theorem for \( \hat{\theta}_n(\phi) \). Indeed, under this assumption one may apply theorem 2.1 of Jensen and Künsch (1994) and assert that there exists a matrix \( \Sigma \) such that \( |\Lambda_n|^{-1/2} U_{\Lambda_n}(\phi; \theta^*) = \mathcal{N}(0, \Sigma) \), as \( n \to \infty \).

We now show that the exponentiality, \( h \)- and MPE assumptions are sufficient conditions.

**Proposition 6.** The exponentiality \( h \)- and MPE assumptions imply the existence and finite range assumptions, and assumptions 1–8, where \( U_{\Lambda_n}(\phi; \theta^*) \) is given by expression (9).

**Proof.** The exponentiality assumption implies that the local energy function is local and stable, which, from results of Bertin et al. (1999), proposition 1, implies that the existence and finite range assumptions are fulfilled. A direct consequence of exponentiality is that for every \( \alpha > 0 \), for all \( \theta \) and for all \( i = 1, \ldots, p \)

\[
E[|U_i(0^M|\phi)|^\alpha \exp\{-\theta^T v(0^M|\phi)\}] < \infty, \tag{21}
\]

which ensures the integrability assumptions 1, 2, 4 and 5 for the test functions considered in the \( h \)-assumption. The locality assumption 7 is contained in the exponentiality assumption and an application of the dominated convergence theorem, with the help of expression (21), shows assumption 6.

Assumption 3 is proved by Billiot et al. (2008) under the exponentiality assumption. Moreover, let \( Z_n(\phi; \theta) := -|\Lambda_n|^{-1} \mathcal{L}_{\Lambda_n}(\phi; \theta) \). If \( \hat{\theta}_n(\phi) \) denotes the MPE, and \( Z_n \) denotes the gradient vector of \( Z_n \) with respect to \( \theta \), one derives from the mean value theorem

\[
Z_n^{(1)}(\phi; \hat{\theta}_n) - Z_n^{(1)}(\phi; \theta^*) = 0 - Z_n^{(1)}(\phi; \theta^*) = H_n(\phi; \theta^*, \hat{\theta}_n) \{ \hat{\theta}_n(\phi) - \theta^* \}
\]

with \( H_n(\phi; \theta^*, \hat{\theta}_n) = \int_0^1 Z_n^{(2)}(\phi; \theta^* + t(\hat{\theta}_n(\phi) - \theta^*)) \, dt \). Under the exponentiality assumption, for \( n \) sufficiently large, \( H_n \) is invertible and converges almost surely towards the matrix \( H(\theta^*) \) given by expression (10). Moreover, following the proof of theorem 2 of Billiot et al. (2008) (see condition (iii), pages 257–258), we derive \( \var{Z_n^{(1)}(\phi; \theta^*)} = O(|\Lambda_n|^{-1}) \). So

\[
|\Lambda_n|^{-1/2} \{ \hat{\theta}_n(\phi) - \theta^* + H_n^{-1}(\theta^*) Z_n^{(1)}(\phi; \theta^*) \} = -|\Lambda_n|^{-1/2} \{ H_n^{-1}(\phi; \theta^*, \hat{\theta}_n) - H_n^{-1}(\theta^*) \} Z_n^{(1)}(\phi; \theta^*) \to 0
\]
in probability as \( n \to \infty \). This implies expression (9). Finally, \( U_{\Lambda_n}(\phi; \theta^*) \) fulfills properties (a)–(c) in assumption 8 from expression (21) and, for property (e), from the proof of theorem 2 (step 1, page 257) in Billiot et al. (2008).

**Remark 3.** Our general setting is not restricted to exponential family models. As an example, following ideas of Coeurjolly and Drouilhet (2010), one may prove that the above assumptions are fulfilled for Lennard–Jones-type models (provided that a certain locality assumption holds).

**B.2. Proof of proposition 1**

Since any stationary Gibbs measure can be represented as a mixture of ergodic measures (see Preston (1976)), it is sufficient to prove the convergence for ergodic measures. We therefore assume that \( P_{\theta^*} \) is ergodic.

Under assumption 1, the ergodic theorem of Nguyen and Zessin (1979b) holds for both terms appearing in the definition of \( I_{\Lambda_n}(\phi; h, \theta^*) \). Then, as \( n \to \infty \), we have \( P_{\theta^*} \) almost surely

\[
|\Lambda_n|^{-1} I_{\Lambda_n}(\phi; h, \theta^*) = E[\hat{h}(0^M, \phi; \theta^*) \exp\{-V(0^M|\phi; \theta^*)\}] - E[h(0^M, \phi \setminus 0^M; \theta^*)],
\]

which equals 0 from the Georgii–Nguyen–Zessin formula (2). This proves the first statement. To prove the consistency of the residuals, it suffices to show that \( |\Lambda_n|^{-1} [R_{\Lambda_n}(\phi; h, \theta_n(\phi)) - I_{\Lambda_n}(\phi; h, \theta^*)] \) converges
towards 0 for $P_\theta$, almost every $\varphi$. Let us write

$$R_{\lambda_n}(\varphi; h, \hat{\theta}_n(\varphi)) - I_{\lambda_n}(\varphi; h, \theta^*) := T_1(\varphi) - T_2(\varphi)$$

with

$$T_1(\varphi) := \int_{\lambda_n \times \mathcal{M}} \{ f(x^m, \varphi; \hat{\theta}_n(\varphi)) - f(x^m, \varphi; \theta^*) \} \mu(dx^m),$$

$$T_2(\varphi) := \sum_{x^m \in \varphi \lambda_n} h(x^m, \varphi \setminus x^m; \hat{\theta}_n(\varphi)) - h(x^m, \varphi \setminus x^m; \theta^*),$$

where we recall that $f(x^m, \varphi; \theta) = h(x^m, \varphi; \theta) \exp\{-V(x^m|\varphi; \theta)\}$. Under assumptions 2 and 3, from the ergodic theorem and the Georgii–Nguyen–Zessin formula, there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$

$$|\Lambda_n|^{-1} |T_1(\varphi)| \leq 2\|\hat{\theta}_n(\varphi) - \theta^*\| \frac{1}{|\Lambda_n|} \int_{\lambda_n \times \mathcal{M}} \|f^{(1)}(x^m, \varphi; \theta^*)\| \mu(dx^m)$$

$$\leq 4\|\hat{\theta}_n(\varphi) - \theta^*\| \mathbb{E}[\|f^{(1)}(0^M, \Phi; \theta^*)\|] \| \exp\{-V(0^M|\Phi; \theta^*)\}],$$

and

$$|\Lambda_n|^{-1} |T_2(\varphi)| \leq 4\|\hat{\theta}_n(\varphi) - \theta^*\| \mathbb{E}[\|h^{(1)}(0^M, \Phi; \theta^*)\|] \| \exp\{-V(0^M|\Phi; \theta^*)\}].$$

Equations (24) and (25) lead to

$$|\Lambda_n|^{-1} R_{\lambda_n}(\varphi; h, \hat{\theta}_n(\varphi)) - I_{\lambda_n}(\varphi; h, \theta^*)| \leq c\|\hat{\theta}_n(\varphi) - \theta^*\|,$$

for $n$ sufficiently large, with $c > 0$. The proof is completed by use of assumption 3.

### B.3. Preliminary results

We first provide a lemma concerning the departure of the residuals from the innovations and $\hat{\theta}_n - \theta^*$. This is a crucial step to investigate the asymptotic normality of the residuals.

**Lemma 1.** Under assumptions 1–4 and 8, assuming that $|\tilde{\lambda}_n| = O(|\lambda_n|)$, then as $n \to \infty$,

$$R_{\lambda_n}(\Phi; h, \hat{\theta}_n(\Phi)) = I_{\lambda_n}(\Phi; h, \theta^*) - |\lambda_n| (\hat{\theta}_n(\Phi) - \theta^*)^T \mathcal{E}(h; \theta^*) + o_P(|\tilde{\lambda}_n|^{1/2}),$$

where $\mathcal{E}(h; \theta^*)$ is the vector defined by

$$\mathcal{E}(h; \theta^*) := \mathbb{E}[h(0^M, \Phi; \theta^*) V^{(1)}(0^M|\Phi; \theta^*) \exp\{-V(0^M|\Phi; \theta^*)\}],$$

and, for any $x^m$ and $\varphi$, $V^{(1)}(x^m|\varphi; \cdot)$ denotes the gradient vector function of $\theta \mapsto V(x^m|\varphi; \theta)$.

**Proof.** Recall that

$$R_{\lambda_n}(\varphi; h, \hat{\theta}_n(\varphi)) - I_{\lambda_n}(\varphi; h, \theta^*) = T_1(\varphi) - T_2(\varphi)$$

where $T_1(\varphi)$ and $T_2(\varphi)$ are defined by expressions (22) and (23).

Let us write

$$T_1(\varphi) := \int_{\lambda_n \times \mathcal{M}} (\hat{\theta}_n(\varphi) - \theta^*)^T f^{(1)}(x^m, \varphi; \theta^*) \mu(dx^m) + T'_1(\varphi),$$

$$T_2(\varphi) := \sum_{x^m \in \varphi \lambda_n} (\hat{\theta}_n(\varphi) - \theta^*)^T h^{(1)}(x^m, \varphi \setminus x^m; \theta^*) + T'_2(\varphi),$$

with

$$T'_1(\varphi) := \int_{\lambda_n \times \mathcal{M}} A_1(x^m, \varphi; \hat{\theta}_n(\varphi)) \mu(dx^m),$$

$$T'_2(\varphi) := \sum_{x^m \in \varphi \lambda_n} A_2(x^m, \varphi \setminus x^m; \hat{\theta}_n(\varphi))$$

and
Let $\hat{\theta}_n(\varphi)=(\hat{\theta}_1, \ldots, \hat{\theta}_p)^T$ and $\theta^*=(\theta_1^*, \ldots, \theta_p^*)^T$. From the mean value theorem, for $l=1, 2$, there exist $\xi_l=(1-c_l)\hat{\theta}_l(\varphi)+c_l\theta^*$ with $c_l \in (0, 1)$, such that

$$A_1(x^m, \varphi; \hat{\theta}_n(\varphi)) = \sum_{j=1}^p (\hat{\theta}_j - \theta_j^*) \{ f_j^{(1)}(x^m, \varphi; \xi_l) - f_j^{(1)}(x^m, \varphi; \theta^*) \},$$

(28)

$$A_2(x^m, \varphi; \hat{\theta}_n(\varphi)) = \sum_{j=1}^p (\hat{\theta}_j - \theta_j^*) \{ h_j^{(1)}(x^m, \varphi \setminus x^m; \xi_l) - h_j^{(1)}(x^m, \varphi \setminus x^m; \theta^*) \}.$$  

(29)

Again from the mean value theorem, for $l=1, 2$ and $j=1, \ldots, p$, there exist $\eta_{l,j}=(1-c_{l,j})\xi_l+c_{l,j}\theta^*$ with $c_{l,j} \in (0, 1)$ such that

$$f_j^{(1)}(x^m, \varphi; \xi_l) - f_j^{(1)}(x^m, \varphi; \theta^*) = \sum_{k=1}^p (\xi_{l,k} - \theta_k^*) \{ f_{jk}^{(2)}(x^m, \varphi; \eta_{l,j}) \},$$

(30)

$$h_j^{(1)}(x^m, \varphi \setminus x^m; \xi_l) - h_j^{(1)}(x^m, \varphi \setminus x^m; \theta^*) = \sum_{k=1}^p (\xi_{l,k} - \theta_k^*) \{ h_{jk}^{(2)}(x^m, \varphi \setminus x^m; \eta_{l,j}) \}.$$  

(31)

By combining equations (28)–(31) and under condition 4, we can deduce the existence of $n_0 \in \mathbb{N}$ such that, for all $n \geq n_0$, we have for $P_\Phi$-almost every $\varphi$

$$|\tilde{\Lambda}_n|^{-1} |T_1(\varphi)| \leq \frac{2}{|\tilde{\Lambda}_n|} \int_{\lambda_n \times \mathcal{M}} \sum_{j,k} |\hat{\theta}_j - \theta_j^*| |\hat{\theta}_k - \theta_k^*| f_{jk}^{(2)}(x^m, \varphi; \theta^*) | \mu(dx^m)$$

$$\leq 2\|\hat{\theta}_n(\varphi) - \theta^*\|^2 |\tilde{\Lambda}_n|^{-1} \int_{\lambda_n \times \mathcal{M}} \|f_{jk}^{(2)}(x^m, \varphi; \theta^*)\| \mu(dx^m)$$

$$\leq 4\|\hat{\theta}_n(\varphi) - \theta^*\|^2 \mathbb{E}[\|f_{jk}^{(2)}(0^M, \Phi; \theta^*)\|]$$

and

$$|\tilde{\Lambda}_n|^{-1} |T_2(\varphi)| \leq \frac{2}{|\tilde{\Lambda}_n|} \sum_{x^m \in \varphi_{\lambda_n}} \sum_{j,k} |\hat{\theta}_j - \theta_j^*| |\hat{\theta}_k - \theta_k^*| h_{jk}^{(2)}(x^m, \varphi \setminus x^m; \theta^*)$$

$$\leq 2\|\hat{\theta}_n(\varphi) - \theta^*\|^2 \frac{1}{|\tilde{\Lambda}_n|} \sum_{x^m \in \varphi_{\lambda_n}} \|h_{jk}^{(2)}(x^m, \varphi \setminus x^m; \theta^*)\|$$

$$\leq 4\|\hat{\theta}_n(\varphi) - \theta^*\|^2 \mathbb{E}[\|h_{jk}^{(2)}(0^M, \Phi; \theta^*)\| \exp\{-V(0^M|\Phi; \theta^*)\}].$$

Since

$$|\tilde{\Lambda}_n|^{1/2} \|\hat{\theta}_n(\varphi) - \theta^*\|^2 = \left( \frac{|\tilde{\Lambda}_n|}{|\Lambda_n|} \right)^{1/2} \|\Lambda_n|^{1/2} \{\hat{\theta}_n(\varphi) - \theta^*\} \| \Lambda_n \{\hat{\theta}_n(\varphi) - \theta^*\} \|$$

$$\text{then, under the assumptions of lemma 1, we have, from Slutsky’s theorem, the following convergence in probability as } n \to \infty:\n
$$

$$|\tilde{\Lambda}_n|^{1/2} \|\hat{\theta}_n(\Phi) - \theta^*\|^2 \overset{P}{\to} 0.$$

By combining all these results, we obtain the following convergence in probability, as $n \to \infty$

$$|\tilde{\Lambda}_n|^{-1/2} \{ T_1(\Phi) - T_1(\Phi) \} - |\tilde{\Lambda}_n| \{\hat{\theta}_n(\Phi) - \theta^*\}^T X_{\lambda_n}(\Phi) = |\tilde{\Lambda}_n|^{-1/2} \{ T_2(\Phi) - T_2(\Phi) \} \overset{P}{\to} 0$$

where $X_{\lambda_n}(\Phi)$ is the random vector defined for all $j=1, \ldots, p$ by

$$(X_{\lambda_n}(\Phi))_j := \frac{1}{|\tilde{\Lambda}_n|} \int_{\lambda_n \times \mathcal{M}} f_j^{(1)}(x^m, \Phi; \theta^*) \mu(dx^m) - \frac{1}{|\tilde{\Lambda}_n|} \sum_{x^m \in \varphi_{\lambda_n}} h_j^{(1)}(x^m, \Phi \setminus x^m; \theta^*).$$
By using the ergodic theorem and the Georgii–Nguyen–Zessin formula (2), we have $P_{\theta^*}$ almost surely as $n \to \infty$

$$(X_{\Lambda_n}(\Phi))_j \to E[ f_j^{(1)}(0^M, \Phi; \theta^*) - h_j^{(1)}(0^M, \Phi; \theta^*) \exp\{ - V(0^M | \Phi; \theta^*) \}].$$

Finally, note that for all $(m, \varphi) \in \mathbb{M} \times \Omega$ and for all $j = 1, \ldots, p$

$$f_j^{(1)}(0^m, \varphi; \theta^*) = \frac{\partial}{\partial \theta_j} [h(0^m, \varphi; \theta) \exp\{ - V(0^m | \varphi; \theta) \}]|_{\theta = \theta^*} = h_j^{(1)}(0^m, \varphi; \theta^*) \exp\{ - V(0^m | \varphi; \theta^*) \} - h(0^m, \varphi; \theta) v_j^{(1)}(0^m | \varphi; \theta^*) \exp\{ - V(0^m | \varphi; \theta^*) \}.$$ 

Therefore $(X_{\Lambda_n}(\Phi))_j \to -E_j(h, \theta^*) P_{\theta^*}$ almost surely as $n \to \infty$. This finally leads to the following convergence in probability, as $n \to \infty$:

$$\frac{1}{D_n} |\tilde{\Delta}_n|^{-1/2} [T_1(\Phi) - T_2(\Phi) + |\tilde{\Delta}_n| (\hat{\theta}_n(\Phi) - \theta^*)^T E(h; \theta^*)] \to 0.$$

We give a last preliminary result which will be crucial in the computation of the covariance structure of the residuals.

**Lemma 2.** For any bounded domain $\Lambda$ and for any test function $h$

$$E[I_\Lambda(\Phi; h; \theta^*) | \Phi]\Lambda.] = 0. 

(32)$$

The proof of lemma 2 is omitted since it corresponds to the proof of theorem 2 (step 1, page 257) of Billiot et al. (2008) by substituting $v_j(x^n | \varphi)$ with the test function $h(x^n, \varphi; \theta^*)$.

### B.4. Proof of proposition 2

For all $n \in \mathbb{N}$, the domain $\Lambda_n$ is assumed to be a cube divided as $\Lambda_n = \bigcup_{j \in \mathcal{J}_n} \Lambda_{j,n}$ where, for all $j \in \mathcal{J}_n$, the $\Lambda_{j,n}$s are disjoint cubes. So $|\Lambda_n| = |\mathcal{J}| |\Lambda_{j,n}| = |\mathcal{J}| |\Lambda_{\Phi,n}|$. Moreover, for all $j \in \mathcal{J}_n$, we can decompose each $\Lambda_{j,n}$ in the following way:

$$\Lambda_{j,n} := \bigcup_{k \in \mathcal{K}_{j,n}} \Delta_k(D_n)$$

where the $\Delta_k(D_n)$s are disjoint cubes with side length $D_n$ and $\mathcal{K}_{j,n} \subset \mathbb{Z}^d$. The side length $D_n$ is chosen greater than $D$ and as close as possible to $D$, leading to

$$D_n = \frac{|\Lambda_n|^{1/d}}{|\mathcal{J}|^{1/d} (|\Lambda_n|^{1/d} / (|\mathcal{J}|^{1/d} D))}.$$ 

This choice implies that $D_n \to D$ when $n \to \infty$ and guarantees $D \leq D_n \leq 2D$ as soon as $|\Lambda_n| \geq |\mathcal{J}| D^d$. The cubes $\Lambda_{j,n}$ are therefore divided into $|\mathcal{K}_{j,n}| = |\Lambda_{\Phi,n}| D_n^{-d}$ cubes whose volumes are close to $D^d$. Denoting $\mathcal{K}_n = \bigcup_{j \in \mathcal{J}_n} \mathcal{K}_{j,n}$, we have $|\mathcal{K}_n| = |\Lambda_n| D_n^{-d} = |\mathcal{J}| |\mathcal{K}_{j,n}|$ and finally

$$\Lambda_n = \bigcup_{j \in \mathcal{J}_n} \bigcup_{k \in \mathcal{K}_{j,n}} \Delta_k(D_n) = \bigcup_{k \in \mathcal{K}_n} \Delta_k(D_n).$$

(34)

From lemma 1 in Appendix B.3 and under assumption 8, we have for any $j \in \mathcal{J}_n$

$$|\Lambda_{\Phi,n}|^{-1/2} R_{\Lambda_{\Phi,n}}(\Phi; h, \theta^*) = |\Lambda_{\Phi,n}|^{-1/2} R_{\infty, \Lambda_{\Phi,n}}(\Phi; h, \theta^*) + o_P(1),$$

where $R_{\infty, \Lambda_{\Phi,n}}(\Phi; h, \theta^*)$ is defined in expression (7).

Therefore the proof of proposition 2 reduces to the proof of the asymptotic normality of the vector $|\Lambda_{j,n}|^{-1/2} R_{\infty, \Lambda_{j,n}}(\Phi; h, \theta^*)$ for $j \in \mathcal{J}_n$. Now

$$|\Lambda_{j,n}|^{-1/2} R_{\infty, \Lambda_{j,n}}(\Phi; h, \theta^*) = |\Lambda_{\Phi,n}|^{-1/2} \{ I_{\Lambda_{j,n}}(\Phi; h, \theta^*) - u_{\Lambda_{\Phi,n}}(\Phi; \theta^*) E(h; \theta^*) \}$$

$$= \frac{|\Lambda_{\Phi,n}|^{-1/2}}{|\Lambda_n|} \{ |\mathcal{J}| I_{\Lambda_{j,n}}(\Phi; h, \theta^*) - u_{\Lambda_{\Phi,n}}(\Phi; \theta^*) E(h; \theta^*) \}$$

$$= \frac{1}{D_n^{3/2}} |\mathcal{J}|^{1/2} \sum_{k \in \mathcal{K}_n} W_{j,n, \Lambda_k(D_n)}(\Phi; \theta^*),$$

(35)
where for any $\varphi \in \Omega$

\[
W_{j,n,\Delta_k(D_n)}(\varphi; \theta^*) = \begin{cases} 
W_{\Delta_k(D_n)}^{(1)}(\varphi; \theta^*) := |\mathcal{J}| I_{\Delta_k(D_n)}(\varphi; h, \theta^*) - U_{\Delta_k(D_n)}(\varphi; \theta^*)^T \mathcal{E}(h; \theta^*) & \text{if } k \in K_{j,n}, \\
W_{\Delta_k(D_n)}^{(2)}(\varphi; \theta^*) := -U_{\Delta_k(D_n)}(\varphi; \theta^*)^T \mathcal{E}(h; \theta^*) & \text{if } k \in K_n \setminus K_{j,n}. 
\end{cases}
\]

(36)

To prove a central limit theorem for the vector $(|\Delta_{j,n}|^{-1/2} R_{\infty,\Lambda_{j,n}}(\Phi; h, \theta^*))_{i \in \mathcal{J}}$, it suffices to apply theorem 2 (see Appendix A), where in its statement we choose $Z_{n,k} = (W_{j,n,\Delta_k(D_n)}(\Phi; \theta^*))_{i \in \mathcal{J}}$, $X_{n,i} = \Phi_{\Delta_k(D_n)}$ and $p = |\mathcal{J}|$. For this, we first must specify the asymptotic variance matrix $\Sigma_\Delta$ and then check the assumptions of theorem 2.

\[|K_n'| = |\mathcal{J}|[|\Lambda_n|^{1/2}/(|\mathcal{J}|^{1/2} D)]^d\]

elements.

Set, for all $k, k' \in \mathbb{R}^d$,

\[
E_{k,k'}^{(l)}(D_n) := E[W_{\Delta_k(D_n)}^{(l)}(\Phi; \theta^*) W_{\Delta_{k'}(D_n)}^{(j)}(\Phi; \theta^*)], \quad l = 1, 2, \quad j = 1, 2.
\]

From the stationarity of the point process, we have $E_{k,k'}^{(l)}(D_n) = E_{0,k-k'}^{(l)}(D_n)$, for $l = 1, 2, 3$. Moreover, under assumptions 7 and 8, for any $k \in K_n$ and for any configuration $\varphi$, since $D_n \geq D$, $W_{\Delta_k(D_n)}(\varphi; \theta^*)$, $i = 1, 2$, depends only on $\varphi_{\Delta_k(D_n)}$. As a consequence, if $k'$ is such that $|k' - k| > 1$, $W_{\Delta_{k'}(D_n)}(\Phi; \theta^*)$ is a measurable function of $\Phi_{\Delta_{k'}(D_n)}$. This leads, for $i, j = 1, 2$,

\[
E[W_{\Delta_k(D_n)}^{(l)}(\Phi; \theta^*) W_{\Delta_{k'}(D_n)}^{(j)}(\Phi; \theta^*)] = E[E[W_{\Delta_k(D_n)}^{(l)}(\Phi; \theta^*) W_{\Delta_{k'}(D_n)}^{(j)}(\Phi; \theta^*)]|\Phi_{\Delta_{k'}(D_n)}] = E[W_{\Delta_{k'}(D_n)}^{(l)}(\Phi; \theta^*) E[W_{\Delta_k(D_n)}^{(j)}(\Phi; \theta^*)]|\Phi_{\Delta_{k'}(D_n)}].
\]

(37)

From lemma 2 in Appendix B.3 and under condition 8, then, for any $k \in \mathbb{Z}^d$ and for $i = 1, 2$,

\[
E[W_{\Delta_k(D_n)}^{(l)}(\Phi; \theta^*)|\Phi_{\Delta_{k'}(D_n)}] = 0.
\]

(38)

From equations (37) and (38), we deduce that, for $l = 1, 2, 3$,

\[
|k' - k| > 1 \Rightarrow E_{k,k'}^{(l)}(D_n) = 0.
\]

(39)

We can now compute the covariance. For any $i$ and $j$ in $\mathcal{J}$, from equation (35),

\[
\text{cov}\{|\Delta_{j,n}|^{-1/2} R_{\infty,\Lambda_{j,n}}(\Phi^*; h, \theta^*), |\Delta_{j,n}|^{-1/2} R_{\infty,\Lambda_{j,n}}(\Phi; h, \theta^*)\} = \frac{1}{D_n^{|\mathcal{J}|}} E \left[ \frac{1}{|K_n'|} \sum_{k \in K_n} \sum_{k' \in K_n} W_{i,n,\Delta_k(D_n)}(\Phi; \theta^*) W_{j,n,\Delta_{k'}(D_n)}(\Phi; \theta^*) \right].
\]

(40)

Let us first consider the case $i = j$. We may write
Therefore,

\[
\begin{align*}
E \left[ \frac{1}{|K_n|} \sum_{k \in K_{i,n}} \sum_{k' \in K_{i,n}} W_{i,n, \Delta_k(D_n)}(\Phi; \theta^*) \right]^2 &= \frac{1}{|K_n|} \left\{ \sum_{k, k' \in K_{i,n}} E^{(1)}_{k, k'}(D_n) + 2 \sum_{k, k' \in K_{i,n}, k' \in K_{i,n}} E^{(1)}_{k, k'}(D_n) \right\} = S_1 \\
&+ \frac{1}{|K_n|} \sum_{k, k' \in K_{i,n}, k' \in K_{i,n}} E^{(2)}_{k, k'}(D_n) \\
&= S_2 + S_3.
\end{align*}
\]

The following lemma will be useful to drop the dependence on $D_n$ in each term $S_1$, $S_2$ and $S_3$ above.

**Lemma 3.** For any $i, j = 1, 2$, denoting $\tilde{\Delta}(\tau) = \bigcup_{|k| \leq 1} \Delta_k(\tau)$ (for some $\tau > 0$), we have

\[
W^{(i)}_{\Delta_0(D_n)}(\Phi; \theta^*) W^{(j)}_{\Delta_0(D_n)}(\Phi; \theta^*) \xrightarrow{L^2} W^{(i)}_{\Delta_0(D)}(\Phi; \theta^*) W^{(j)}_{\Delta_0(D)}(\Phi; \theta^*).
\]

**Proof.** For any $i = 1, 2$, $W^{(i)}_{\Delta_0(D_n)}$ is a linear combination of $I^{(i)}_{\Delta_0(D_n)}$ and $U^{(i)}_{\Delta_0(D_n)}$, which converge respectively in $L^2$ to $I^{(i)}_{\Delta_0(D)}$ and $U^{(i)}_{\Delta_0(D)}$ by conditions 6 and 8 since $D_n \to D$. Thus $W^{(i)}_{\Delta_0(D_n)}$ converges in $L^2$ to $W^{(i)}_{\Delta_0(D)}$ as $n \to \infty$. Similarly, for any $j = 1, 2$, $W^{(j)}_{\Delta_0(D_n)}$ tends in $L^2$ to $W^{(j)}_{\Delta_0(D)}$. The convergence stated in lemma 3 then follows.

Let us focus on the asymptotic of each term $S_1$, $S_2$ and $S_3$. From expression (39),

\[
S_1 = \sum_{k \in K_{i,n}} \left\{ \sum_{k' \in K_{i,n}, |k' - k| \leq 1} E^{(1)}_{k, k'}(D_n) + \sum_{k' \in K_{i,n}, |k' - k| > 1} E^{(1)}_{k, k'}(D_n) \right\} = \sum_{k \in K_{i,n}} \sum_{k' \in K_{i,n}, |k' - k| \leq 1} E^{(1)}_{k, k'}(D_n).
\]

Let $\tilde{K}_{i,n} := \{ k \in K_{i,n}, \exists j \in K_{i,n} | j - k | \leq 2 \}$ and note that $|\tilde{K}_{i,n}|/|K_{i,n}| \to 0$ as $n \to \infty$. Then,

\[
S_1 = \sum_{k \in K_{i,n}} \sum_{k' \in K_{i,n}, |k' - k| \leq 1} E^{(1)}_{k, k'}(D_n) = \sum_{k \in K_{i,n}} \sum_{k' \in K_{i,n}, |k' - k| \leq 1} E^{(1)}_{k, k'}(D_n).
\]

Since

\[
\frac{1}{|K_n|} |A_1| \leq \frac{|\tilde{K}_{i,n}|}{|K_n|} \sum_{|k| \leq 1} |E^{(1)}_{0, k}(D_n)| \xrightarrow{n \to \infty} 0
\]

(because $D \leq D_n \leq 2D$ and $|\tilde{K}_{i,n}|/|K_{i,n}| \to 0$), we obtain, as $n \to \infty$,

\[
\frac{1}{|K_n|} S_1 \sim \frac{|K_{i,n}|}{|K_n|} \sum_{|k| \leq 1} E^{(1)}_{0, k}(D_n) \sim \frac{|K_{i,n}|}{|K_n|} \sum_{|k| \leq 1} E^{(1)}_{0, k}(D_n).
\]

From lemma 3,

\[
\sum_{|k| \leq 1} E^{(1)}_{0, k}(D_n) = E[W^{(1)}_{\Delta_0(D_n)}(\Phi; \theta^*) W^{(1)}_{\Delta_0(D_n)}(\Phi; \theta^*)] \to \sum_{|k| \leq 1} E^{(1)}_{0, k}(D).
\]

Therefore,

\[
\frac{1}{|K_n|} S_1 \sim \frac{1}{|J|} \sum_{|k| \leq 1} E^{(1)}_{0, k}(D).
\]

With decompositions and arguments similar to those above (omitted here), we obtain easily as $n \to \infty$

\[
\frac{1}{|K_n|} S_2 \to 0,
\]

\[
\frac{1}{|K_n|} S_3 \sim \frac{|J| - 1}{|J|} \sum_{|k| \leq 1} E^{(2)}_{0, k}(D).
\]
Combining the three terms $S_1$, $S_2$ and $S_3$, we have, as $n \to \infty$,

$$
E \left[ \frac{1}{|K_n|} \sum_{k \in K_n} \sum_{k' \in K_n} W_{i,n,k} \Delta_k(D_n)(\Phi; \theta^*) \right]^2 
\sim \sum_{|\mathcal{I}| \leq 1} \left\{ \frac{1}{|J|} E_{0,k}^{(i)}(D) + \frac{|J| - 1}{|J|} E^{(2)}_{0,k}(D) \right\}.
$$

(41)

When $i \neq j$, there are three main cases in equation (40), according to $k, k' \in K_{i,n}, k, k' \in K_{j,n}$ or $k, k' \in K_n \setminus (K_{i,n} \cup K_{j,n})$. As for the case $i = j$ treated before, the other situations involve non-zero correlations on edge sets like $K_{i,n}$, which are negligible with respect to $|K_n|$. The covariance is therefore equivalent, up to $D_3|\mathcal{J}|$, to

$$
\frac{1}{|K_n|} \sum_{k, k' \in K_{i,n}} E^{(12)}_{k,k'}(D_n) + \frac{1}{|K_n|} \sum_{k, k' \in K_{j,n}} E^{(12)}_{k,k'}(D_n) + \frac{1}{|K_n|} \sum_{k, k' \in K_n \setminus (K_{i,n} \cup K_{j,n})} E^{(2)}_{k,k'}(D_n).
$$

The simplification occurs as for the case $i = j$ and, since $|K_{i,n}| = |K_{j,n}|$, we obtain the asymptotic equivalent for the covariance (40)

$$
\frac{1}{D_3|\mathcal{J}|} \sum_{|\mathcal{I}| \leq 1} \left\{ \frac{2}{|J|} E^{(12)}_{0,k}(D) + \frac{|J| - 2}{|J|} E^{(2)}_{0,k}(D) \right\}.
$$

(42)

Finally, from expressions (41) and (42), we deduce that $\Sigma_1(\theta^*)$, defined in proposition 2, corresponds to the asymptotic variance of $|\Delta_{i,j}|^{-1/2} R_{\infty,\Delta_{i,j}}(\Phi; h, \theta^*)$ for $\theta^* \in \mathcal{J}$.

B.4.2 Second step: application of theorem 2

We apply theorem 2 with $Z_{n,k} = (W_{j,n,k} \Delta_k(D_n))_{j \in \mathcal{J}}$, $X_{n,i} = \Phi_{\Delta_i(D_n)}$, $p = |\mathcal{J}|$ and $\Sigma = \Sigma_1(\theta^*) \times D_3|\mathcal{J}|$, which is a symmetric positive semidefinite matrix as the limit of a covariance matrix (from the first step of the proof).

Assumption (20) holds from conditions 7 and 8 and because $D_n \geq D$. Assumptions (a)–(c) are direct consequences of conditions 8 and 5 and lemma 2. It remains to prove part (d). Assuming $\Sigma = (\Sigma_{i,j})$ for $1 \leq i, j \leq p$, from the definition of the Frobenius norm, we have

$$
E \left[ |K_n|^{-1} \sum_{k \in K_n} \sum_{k' \in K_n \setminus [k', \mathcal{I}]} Z_{n,k}^T Z_{n,k'} - \Sigma \right] \leq \sum_{i=1}^p \sum_{j=1}^p E \left[ |K_n|^{-1} \sum_{k \in K_n} \sum_{k' \in K_n \setminus [k', \mathcal{I}]} W_{i,n,k} \Delta_k(D_n) W_{j,n,k'}(D_n) - \Sigma_{i,j} \right].
$$

(43)

We first assume that $i \neq j$ are fixed and denote

$$
Y_{n,k}(D_n) = W_{i,n,k} \Delta_k(D_n),
$$

$$
S_n^k(D_n) = \sum_{k' \in K_n \setminus [k', \mathcal{I}]} W_{j,n,k'}(D_n).
$$

We have

$$
E \left[ |K_n|^{-1} \sum_{k \in K_n} \sum_{k' \in K_n \setminus [k', \mathcal{I}]} W_{i,n,k} \Delta_k(D_n) W_{j,n,k'}(D_n) - \Sigma_{i,j} \right] = |K_n|^{-1} E \left[ \sum_{k \in K_n} \left\{ Y_{n,k}(D_n) S_n^k(D_n) - \Sigma_{i,j} \right\} \right]
\leq E_1 + E_2 + E_3 + E_4,
$$

where

$$
E_1 = \frac{|K_{i,n}|}{|K_n|} E \left[ |K_{i,n}|^{-1} \sum_{k \in K_{i,n}} (Y_{n,k}(D_n) S_n^k(D_n) - E[Y_{n,k}(D_n) S_n^k(D_n)]) \right],
$$

$$
E_2 = \frac{|K_{j,n}|}{|K_n|} E \left[ |K_{j,n}|^{-1} \sum_{k \in K_{j,n}} (Y_{n,k}(D_n) S_n^k(D_n) - E[Y_{n,k}(D_n) S_n^k(D_n)]) \right],
$$

$$
E_3 = \frac{|K_n \setminus (K_{i,n} \cup K_{j,n})|}{|K_n|} E \left[ |K_n \setminus (K_{i,n} \cup K_{j,n})|^{-1} \sum_{k \in K_n \setminus (K_{i,n} \cup K_{j,n})} (Y_{n,k}(D_n) S_n^k(D_n) - E[Y_{n,k}(D_n) S_n^k(D_n)]) \right],
$$

and

$$
E_4 = \frac{|K_{i,n} \setminus K_{j,n}|}{|K_n|} E \left[ |K_{i,n} \setminus K_{j,n}|^{-1} \sum_{k \in K_{i,n} \setminus K_{j,n}} (Y_{n,k}(D_n) S_n^k(D_n) - E[Y_{n,k}(D_n) S_n^k(D_n)]) \right].
$$
\[
E_k = \frac{|\mathcal{K}_{i,n}|}{|\mathcal{K}_n|} \sum_{k \in \mathcal{K}_{i,n}} \mathbb{E}[Y_{n,k}(D_n) S_n^k(D_n)] + \frac{|\mathcal{K}_{i,n}|}{|\mathcal{K}_n|} \sum_{k \in \mathcal{K}_{j,n}} \mathbb{E}[Y_{n,k}(D_n) S_n^j(D_n)] + \frac{|\mathcal{K}_n \setminus (\mathcal{K}_{i,n} \cup \mathcal{K}_{j,n})|}{|\mathcal{K}_n|} \times \sum_{k \in \mathcal{K}_n \setminus (\mathcal{K}_{i,n} \cup \mathcal{K}_{j,n})} \mathbb{E}[Y_{n,k}(D_n) S_n^k(D_n)] - \Sigma_{ij}.
\]

The first three terms \( E_1, E_2 \) and \( E_3 \) can be handled similarly. Let us focus on \( E_1 \) for instance:

\[
\frac{|\mathcal{K}_n|}{|\mathcal{K}_{i,n}|} E_1 \leq |\mathcal{K}_{i,n}|^{-1} \sum_{k \in \mathcal{K}_{i,n}} \mathbb{E}[Y_{n,k}(D_n) S_n^k(D_n) - Y_{n,k}(D) S_n^k(D)] + |\mathcal{K}_{i,n}|^{-1} \mathbb{E} \sum_{k \in \mathcal{K}_{i,n}} (Y_{n,k}(D) S_n^k(D) - \mathbb{E}[Y_{n,k}(D) S_n^k(D)]) - \mathbb{E}[Y_{n,k}(D) S_n^k(D)] + |\mathcal{K}_{i,n}|^{-1} \sum_{k \in \mathcal{K}_{i,n}} |\mathbb{E}[Y_{n,k}(D) S_n^k(D)] - \mathbb{E}[Y_{n,k}(D) S_n^k(D)]|.
\]

(44)

Up to the edge effects which are negligible with respect to \( |\mathcal{K}_{i,n}| \), \( (Y_{n,k}(D) S_n^k(D))_k \) is stationary when \( k \in \mathcal{K}_{i,n} \), since in this case, from expression (36), \( W_{i,n,\Delta_\lambda(D)} = W_{\Delta_\lambda(D)}^{(j)} \) does not depend on \( n \). Therefore the second term in inequality (44) tends to 0 by the mean ergodic theorem. For a fixed \( n \), we have also by stationarity (up to the edge effects)

\[
\mathbb{E}|Y_{n,k}(D_n) S_n^k(D_n) - Y_{n,k}(D) S_n^k(D)| = \mathbb{E}|Y_{n,0}(D_n) S_n^0(D_n) - Y_{n,0}(D) S_n^0(D)| = \mathbb{E}|W_{\Delta_0(D_n)}^{(1)} W_{\Delta_0(D)}^{(2)} - W_{\Delta_0(D_n)}^{(1)} W_{\Delta_0(D)}^{(2)}|,
\]

where \( \Delta_\lambda(D) = \bigcup_{|j| \leq 1} \Delta_j(D_n) \). From lemma 3, this term tends to 0; therefore the first term in inequality (44) asymptotically vanishes. The same argument shows that the third term in inequality (44) also tends to 0 as \( n \to \infty \). As a consequence, \( E_1 \to 0 \).

The same decomposition as in inequality (44) may be done for \( E_2 \) and \( E_3 \), which leads by similar arguments to \( E_2 \to 0 \) and \( E_3 \to 0 \). The last term \( E_4 \) involves the difference between \( \Sigma_{ij} \) and its empirical counterpart. The same calculations as in the first step of the proof show that \( E_4 \to 0 \).

Therefore, we have proved that the terms in the double sum (43) corresponding to \( i \neq j \) asymptotically vanish. The same result can be proved similarly when \( i = j \). Thus assumption (d) in theorem 2 holds and the convergence in law is deduced.

B.5. Proof of proposition 3

We can decompose \( \Lambda_n \) in the following way:

\[
\Lambda_n := \bigcup_{k \in \mathcal{K}_n} \Delta_k(D_n)
\]

where the \( \Delta_k \)s are disjoint cubes with side length \( D_n \) and \( \mathcal{K}_n \subset \mathbb{Z}^d \) satisfies \( |\mathcal{K}_n| = |\Lambda_n| D_n^{-d} \). Similarly to the proof of proposition 2, we choose

\[
D_n = \frac{|\Lambda_n|^{1/d}}{||\Lambda_n|^{1/d} / D|},
\]

which implies that \( D_n \to D \) when \( n \to \infty \) and guarantees \( D \leq D_n \leq 2D \) as soon as \( |\Lambda_n| \geq D^d \).

From lemma 1 (see Appendix B.3) and under assumption 8, for all \( i = 1, \ldots, s \),

\[
|\Lambda_n|^{-1/2} R_{\lambda_n}(\Phi; h_i, \hat{\theta}_n(\Phi)) = |\Lambda_n|^{-1/2} R_{\lambda_n}(\Phi; h_i, \theta^*) + o_p(1)
\]

\[
= |\Lambda_n|^{-1/2} \{ I_{\lambda_n}(\Phi; h_i, \theta^*) - U_{\lambda_n}(\Phi; \theta^*)^T \mathbb{E}(h_i; \theta^*) \} + o_p(1)
\]

\[
= \frac{1}{D_n^{d/2}} \frac{|\mathcal{K}_n|^{1/2}}{|\mathcal{K}_n|^{1/2}} \sum_{k \in \mathcal{K}_n} W_{\lambda_k}(\Phi; h_i, \theta^*) + o_p(1),
\]

where for any \( \varphi \in \Omega \)

\[
W_{\lambda_k}(\varphi; h_i, \theta^*) := I_{\lambda_k}(\varphi; h_i, \theta^*) + U_{\lambda_k}(\varphi; \theta^*)^T \mathbb{E}(h_i; \theta^*)
\]

We apply theorem 2 in the simpler case when \( f_{n,k} = f \) for all \( n \in \mathbb{N} \) and all \( k \in \mathcal{K}_n \). If \( D_n = D \) for all \( n \), this framework would reduce to a stationary setting similar to theorem 2.1 in Jensen and Künsch (1994). But, as \( \Lambda_n \) is allowed to increase continuously up to \( \mathbb{R}^d \), \( D_n \equiv D \) is impossible. We shall therefore apply theorem 2 in Appendix A with \( Z_{n,k} = (W_{\lambda_k}(\Phi; h_j, \theta^*))_{j=1,\ldots,s} \), \( X_{n,i} = \Phi_{\lambda_i}(D_n) \) and \( p = s \).
We first compute the covariance matrix of \( (|A_n|^{-1/2} R_{\infty,A_n}(\Phi; h_i, \theta^*))_{i=1,\ldots,t} \). By the same calculations as for the term \( S_1 \) in the proof of proposition 2, we obtain

\[
\text{cov}\{ |A_n|^{-1/2} R_{\infty,A_n}(\Phi; h_i, \theta^*), |A_n|^{-1/2} R_{\infty,A_n}(\Phi; h_j, \theta^*) \} = \frac{1}{D_n^4} E \left[ \frac{1}{|K_n|} \sum_{k \in K_n} \sum_{k' \in K_n} W_{\Delta_k(D_n)}(\Phi; h_i, \theta^*) W_{\Delta_{k'}(D_n)}(\Phi; h_j, \theta^*) \right]
\]

\[
\sim \frac{1}{D_n^4} \sum_{|k| \leq 1} E[W_{\Delta_k(D)}(\Phi; h_i, \theta^*) W_{\Delta_k(D)}(\Phi; h_j, \theta^*)].
\]  

(45)

The asymptotic covariance matrix is thus \( \Sigma_n(\theta^*) \) defined in proposition 3. We can now apply theorem 2 in Appendix A with \( \Sigma = \Sigma_n(\theta^*) \times D^2 \). Assumption (20) holds from conditions 7 and 8 and because \( D_n \geq D \). Assumptions (a)–(c) follow from conditions 8 and 5 and lemma 2. Assumption (d) may be checked easily as in the second step of the proof of proposition 2, by using expression (45).

### B.6. Proof of corollaries 1–3

We present only the proof of corollary 1, since the others are similarly obtained from proposition 2 and 3.

First assume that \( P_{\theta^*} \) is an ergodic measure. Then proposition 2 applies. An easy computation shows that \( \lambda_{\text{inn}} \) and \( \lambda_{\text{res}} \) are the two eigenvalues of \( \Sigma_n(\theta^*) \) with respective order \(|J| - 1\) and 1. Let \( P_{\text{inn}}^T \) be the matrix of orthonormalized eigenvectors associated with \( \lambda_{\text{inn}} \). This matrix of size \((|J|, |J| - 1)\) satisfies by definition \( P_{\text{inn}}^T P_{\text{inn}} = I_{|J| - 1} \). Moreover, we can check that \( P_{\text{inn}}^T P_{\text{inn}} = I_{|J| - 1} \), which leads to

\[
||R_{J,n}(\varphi; h) - R_{J,n}(\varphi; h)||^2 = ||P_{\text{inn}}^T R_{J,n}(\varphi; h)||^2.
\] 

(46)

From expression (11), and since \( P_{\text{inn}}^T \Sigma_n(\theta^*) P_{\text{inn}} = \lambda_{\text{inn}} P_{\text{inn}} = \lambda_{\text{inn}} I_{|J| - 1} \),

\[
|A_{0,n}|^{-1/2} P_{\text{inn}}^T R_{J,n}(\varphi; h) \xrightarrow{d} N(0, \lambda_{\text{inn}} I_{|J| - 1}).
\]

From lemma 4 which will be stated in Appendix B.9 and Slutsky’s theorem,

\[
|A_{0,n}|^{-1/2} \lambda_{\text{inn}} P_{\text{inn}} R_{J,n}(\varphi; h) \xrightarrow{d} N(0, I_{|J| - 1}).
\] 

(47)

Second, if \( P_{\theta^*} \) is not an ergodic measure, it can be decomposed as a mixture of ergodic measures; see Preston (1976). Since any mixture of standard Gaussian distributions is a standard Gaussian distribution, result (47) remains true. The result of corollary 3 follows directly from expressions (46) and (47).

### B.7. Proof of proposition 4

From the definition (7) of \( R_{\infty,A} \), proposition 4 is proved by noting that

\[
H(\theta^*) \omega = E[v(0^M|\Phi) v(0^M|\Phi)^T \exp\{-V(0^M|\Phi; \theta^*)\}]_{\omega} = E[v(0^M|\Phi) (\omega^T v(0^M|\Phi))^T \exp\{-V(0^M|\Phi; \theta^*)\}] = E[h(0^M, \Phi; \theta^*) v(0^M|\Phi) \exp\{-V(0^M|\Phi; \theta^*)\}] = E(\omega^T v, \theta^*).
\]

Therefore, \( H(\theta^*)^{-1} E(\omega^T v, \theta^*) = \omega \), so for any \( \varphi \in \Omega \) and any bounded domain \( \Lambda \)

\[
R_{\infty,A}(\varphi; \omega^T v, \theta^*) = I_{\Lambda}(\varphi; \omega^T v, \theta^*) - L P L^{(1)}(\varphi; \theta^*)^T \omega = 0.
\]

This means that if, for framework 1, the test function is of the form \( h = \omega^T v \) then \( \lambda_{\text{res}} = 0 \) and if, for framework 2, one of the test functions is of the form \( h = \omega^T v \), then \( \Sigma_n(\theta^*) \) is necessarily singular.

### B.8. Proof of proposition 5

The proof of proposition 5 follows arguments presented in Jensen and Künsch (1994). Let \( C_n(D) = [-nD - D/2, nD + D/2]^d \), so \( C_n(D) = \bigcup_{k \in K_n} \Delta_k(D) \), where \( K_n = [-n, n]^d \cap \mathbb{Z}^d \) and \( \Delta_k(D) \) is the cube centred at \( kD \) with side length \( D \). We have
\[ \text{var}\{C_n(D)^{-1} R_{\infty,C_n(D)}(\Phi; h, \theta^*)\} = |C_n(D)|^{-1} \sum_{i,j \in \mathcal{K}_n} E[R_{\infty,\Delta_i(D)}(\Phi; h, \theta^*) R_{\infty,\Delta_j(D)}(\Phi; h, \theta^*)^T] \]

\[ = |C_n(D)|^{-1} \sum_{i \in \mathcal{K}_n} \sum_{j \in \mathcal{K}_n} E[R_{\infty,\Delta_i(D)}(\Phi; h, \theta^*) R_{\infty,\Delta_j(D)}(\Phi; h, \theta^*)^T]. \]

Since \(|C_n(D)| = D'|\mathcal{K}_n|\), from the ergodic theorem,

\[ \text{var}\{C_n(D)^{-1} R_{\infty,C_n(D)}(\Phi; h, \theta^*)\} \to D^{-d} \sum_{j \in \mathcal{K}_n} E[R_{\infty,\Delta_j(D)}(\Phi; h, \theta^*) R_{\infty,\Delta_j(D)}(\Phi; h, \theta^*)^T] \]

which is nothing other than \(\sum_j (\theta^*).\) Therefore, it is sufficient to prove that the covariance matrix \(\text{var}\{C_n(D)^{-1} R_{\infty,C_n(D)}(\Phi; h, \theta^*)\}\) is positive definite for \(n\) sufficiently large. Let \(x \in \mathbb{R}^1 \setminus \{0\}\); we must show that

\[ V := x^T \text{var}\{C_n(D)^{-1} R_{\infty,C_n(D)}(\Phi; h, \theta^*)\} x > 0. \]

Since, for two random variables \(X\) and \(X'\) with finite variance

\[ \text{var}(X) = E[\text{var}(X|X')] + \text{var}(E[X|X']) \geq E[\text{var}(X|X')], \]

we have, by denoting \(L := 3D^d,\)

\[ V \geq |C_n(D)|^{-1} E[\text{var}\{x^T R_{\infty,C_n(D)}(\Phi; h, \theta^*)|\Phi_{\Delta_l(D)}(\Phi; h, \theta^*)| \neq L\}]
\]

\[ = |C_n(D)|^{-1} x^T \text{var}\left[ \sum_{l \in L \cap \mathcal{K}_n} \sum_{j \in \mathcal{K}_n} R_{\infty,\Delta_j(D)}(\Phi; h, \theta^*) \right] \times \Phi_{\Delta_l(D)},| \neq L \right]\]

\[ := \mathbb{S}_{l,n}(\Phi). \]

Note that, from the locality property, \(\mathbb{S}_{l,n}(\Phi)\) depends only on \(\Phi_{\Delta_l(D)}\) for \(|l| \leq 2\). Therefore, conditionally on \(\Phi_{\Delta_l(D)}, \Phi_{\Delta_l(D)} \neq L\), the variables \(\mathbb{S}_{l,n}(\Phi)\) and \(\mathbb{S}_{l',n}(\Phi)\) for \((l \neq l')\) are independent. Now, let \(\bar{\Delta} := \bigcup_{|l| \leq 2} \Delta_l(D)\); by stationarity we have for \(n\) sufficiently large

\[ V \geq |C_n(D)|^{-1} x^T \sum_{l \in L \cap \mathcal{K}_n} E[\text{var}\{\mathbb{S}_{l,n}(\Phi)|\Phi_{\Delta_l(D)}(\Phi; h, \theta^*)| \neq L\}]
\]

\[ \geq \frac{D^{-d} |L \cap \mathcal{K}_n|}{2} E[\text{var}\{x^T R_{\infty,\bar{\Delta}}(\Phi; h, \theta^*)|\Phi_{\bar{\Delta}_l(D)}, 1 \leq |l| \leq 2\}]
\]

\[ \geq \kappa(D, d) E[\text{var}\{x^T R_{\infty,\bar{\Delta}}(\Phi; h, \theta^*)|\Phi_{\bar{\Delta}_l(D)}, 1 \leq |l| \leq 2\}], \]

where \(\kappa(D, d)\) is a positive constant. Assume that there is some positive constant \(c\) such that \(P_{\theta^*}\) almost surely \(x^T R_{\infty,\bar{\Delta}}(\Phi; h, \theta^*) = c\) when the variables \(\Phi_{\bar{\Delta}_l(D)}, 1 \leq |l| \leq 2\), are fixed to belong to \(B\), where \(B \in \mathcal{F}\) is defined in the positive definiteness assumption. It follows that, for any \(\varphi_i \in A_i, i = 0, \ldots, l\) (with \(l \geq 1\)), where the \(A_i\)'s come from the positive definiteness assumption, \(x^T \{R_{\infty,\bar{\Delta}}(\varphi_i; h, \theta^*) - R_{\infty,\bar{\Delta}}(\varphi_0; h, \theta^*)\} = 0.\) Since the matrix with entries \(R_{\infty,\bar{\Delta}}(\varphi_i; h, \theta^*) - R_{\infty,\bar{\Delta}}(\varphi_0; h, \theta^*), i = 1, \ldots, s, j = 1, \ldots, l,\) is assumed to be injective, this leads to \(x = 0\) and hence to some contradiction. Therefore, when the variables \(\Phi_{\bar{\Delta}_l(D)}, 1 \leq |l| \leq 2\), are for example assumed to belong to \(B\), the variable \(x^T R_{\infty,\bar{\Delta}}(\Phi; h, \theta^*)\) is almost surely not a constant and so \(V > 0\).

**B.9. Positivity of \(\lambda_{\ln n}\)**

**Lemma 4.** Under the exponentiality and \(h\)-assumptions, \(\lambda_{\ln n} > 0\).

Assume, then, the positive definiteness condition where \(R_{\infty,\bar{\Delta}}(\varphi_i; h, \theta^*)\) is replaced by \(I_{\bar{\Delta}}(\varphi; h, \theta^*)\) (here \(s = 1\)). Then, by similar arguments to those for proposition 5, we obtain \(\lambda_{\ln n} > 0\). Therefore the proof reduces to checking that this version of the positive definiteness condition is satisfied under the exponentiality and \(h\)-assumptions.

In the positive definiteness assumption we fix \(B = \emptyset\). Let us write \(\Omega := \Omega_B\) and consider the following events for some \(n \geq 1:\)

\[ A_0 = \{\varphi \in \Omega : |\varphi_{\Delta_l(D)}| = 0\}, \]

\[ A_n = \{\varphi \in \Omega : |\varphi_{\Delta_l(D)}| = n\}. \]
Let $\varphi_0 \in A_0$ and $\varphi_n \in A_n$. Recall that the local stability property (ensured by the exponentiality assumption) asserts that there exists $K \geq 0$ such that $V(x^n; \varphi; \theta^*) \geq -K$ for any $x^n \in \mathbb{S}$ and any $\varphi \in \Omega$. Now, let us consider the different types of residuals in the $h$-assumption.

(a) **Raw residuals** ($h = 1$): from the local stability property

\[ |I_\Delta(\varphi_n; h, \theta^*) - I_\Delta(\varphi_0; h, \theta^*)| \geq n - \left| \int_{\Delta \times M} \exp\{-V(x^n|\varphi_n; \theta^*)\} - \exp\{-V(x^n|\varphi_0; \theta^*)\} \mu(dx^n) \right| \]

for $n$ sufficiently large. And so assuming that the left-hand side is 0 leads to a contradiction, which proves the positive definiteness assumption.

(b) **Inverse residuals** ($h = \exp(V)$): again, from the local stability property

\[ |I_\Delta(\varphi_n; h, \theta^*) - I_\Delta(\varphi_0; h, \theta^*)| = \left| \sum_{x^n \in \varphi_n, \Delta} \exp\{V(x^n|\varphi_n \setminus x^n; \theta^*)\} \right| \geq n \exp(-K) > 0, \]

which proves the positive definiteness assumption similarly to the previous case.

(c) **Pearson residuals** ($h = \exp(V/2)$): from the same argument

\[ |I_\Delta(\varphi_n; h, \theta^*) - I_\Delta(\varphi_0; h, \theta^*)| = \left| \sum_{x^n \in \varphi_n, \Delta} \exp\{V(x^n|\varphi_n \setminus x^n; \theta^*)/2\} - \int_{\Delta \times M} \exp\{-V(x^n|\varphi_0; \theta^*)/2\} \mu(dx^n) \right| \]

\[ \geq n \exp(-K/2 - 2|\Delta| \exp(K/2) > 0, \]

for $n$ sufficiently large.

(d) **Empty space residuals** ($h(x^n, \varphi; \theta) = I_{(0,\varphi)}\{d(x^n, \varphi)\} \exp\{V(x^n|\varphi; \theta)\}$): let $\tilde{A}_n = \{ \varphi \in \Omega : |\varphi_{\Delta_0(0) \cap B(0, r)}| = n \}$ and $\tilde{\varphi}_n \in \tilde{A}_n$. For some $c > 0$,

\[ |I_\Delta(\varphi; h, \theta^*) - I_\Delta(\tilde{\varphi}_n; h, \theta^*)| \geq c(r^d \wedge |\Delta|) + \sum_{x^n \in \tilde{\varphi}_n, \Delta} I_{(0,\varphi)}\{d(x^n, \tilde{\varphi}_n)\} \exp\{V(x^n|\tilde{\varphi}_n \setminus x^n; \theta^*)\} \]

\[ - \int_{\Delta \times M} I_{(0,\varphi)}\{d(x^n, \tilde{\varphi}_n)\} \mu(dx^n) \]

\[ \geq c(r^d \wedge |\Delta| - (r + D\sqrt{d} \wedge |\Delta|)) + n \exp(-K) > 0 \]

for $n$ sufficiently large, which ends the proof.

**References**


Supporting information

Additional ‘supporting information’ may be found in the on-line version of this article:

‘Technical note on the paper “Residuals and goodness-of-fit tests for stationary marked Gibbs point processes”’.

Please note: Wiley–Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the author for correspondence for the article.