



HAL
open science

Median-based estimation of the intensity of a spatial point process

Jean-François Coeurjolly

► **To cite this version:**

Jean-François Coeurjolly. Median-based estimation of the intensity of a spatial point process. 2014. hal-01071605v1

HAL Id: hal-01071605

<https://hal.science/hal-01071605v1>

Submitted on 6 Oct 2014 (v1), last revised 24 May 2015 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Median-based estimation of the intensity of a spatial point process

Jean-François Coeurjolly¹

¹Univ. Grenoble Alpes, F-38000 Grenoble, France,
Jean-Francois.Coeurjolly@upmf-grenoble.fr.

Abstract

This paper is concerned with a robust estimator of the intensity of a stationary spatial point process. The estimator corresponds to the median of a jittered sample of the number of points, computed from a tessellation of the observation domain. We show that this median-based estimator satisfies a Bahadur representation from which we deduce its consistency and asymptotic normality under mild assumptions on the spatial point process. Through a simulation study, we compare the new estimator with the standard one counting the mean number of points per unit volume. The empirical study verifies the asymptotic properties established and shows that the median-based estimator is more robust to outliers than the standard estimator.

Keywords: Spatial point processes; Robust statistics; Sample quantiles; Bahadur representation.

1 Introduction

Spatial point patterns are datasets containing the random locations of some event of interest. These datasets arise in many scientific fields such as biology, epidemiology, seismology, hydrology. Spatial point processes are the stochastic models generating such data. We refer to Stoyan et al. (1995), Illian et al. (2008) or Møller and Waagepetersen (2004) for an overview on spatial point processes. These references cover practical as well as theoretical aspects. A point process \mathbf{X} in \mathbb{R}^d is a locally finite random subset of \mathbb{R}^d meaning that the restriction to any bounded Borel set is finite. The point process \mathbf{X} takes values in Ω consisting in all locally finite subsets of \mathbb{R}^d . Thus the distribution of \mathbf{X} is a probability measure on an appropriate σ -algebra consisting of subsets of Ω . The Poisson point process is the reference process which models random locations of points without interaction. Many alternative models such as Cox point processes, determinantal point processes, Gibbs point processes allow to introduce clustering effects or to produce regular pattern (see again e.g. Møller and Waagepetersen (2004) for an overview). If the distribution of \mathbf{X} is invariant by translation, we say that \mathbf{X} is stationary. We are interested in this paper in

first-order characteristics of \mathbf{X} , which under the assumption of stationarity, reduce to a single real parameter denoted by λ . This intensity parameter λ measures the mean number of points per unit volume.

Estimating λ has become a well-known problem and the source of a large literature. Based on a single realization of the point process \mathbf{X} in a bounded domain W of \mathbb{R}^d , the natural way of estimating λ is to compute the mean number of points observed by unit volume, i.e. to evaluate $n(\mathbf{X} \cap W)/|W|$ where $n(\mathbf{y})$ represents the number of a finite set of points \mathbf{y} and where $|W|$ is the volume of W . We refer to as $\hat{\lambda}^{\text{std}}$ for this estimator. If the point process is a homogeneous Poisson point process, $\hat{\lambda}^{\text{std}}$ is also the maximum likelihood estimator. Asymptotic properties of $\hat{\lambda}^{\text{std}}$ are now well established for a large class of models. In particular, as the window of observation expands to \mathbb{R}^d , it can be shown under mild assumptions on \mathbf{X} (mainly mixing conditions) that $\hat{\lambda}^{\text{std}}$ is consistent and satisfies a central limit theorem with asymptotic variance which can be consistently estimated (see Heinrich and Prokešová (2010) and the references therein for more details). In some applications, it may be too time-consuming to count all points. In such situations, distance based methods, where mainly nearest distances between points are used, have been developed (see e.g. Särkkä (1992); Diggle (2003)). Unlike the estimator $\hat{\lambda}^{\text{std}}$, those methods are very sensitive to the model which explains that the only case where it may be realistically applied is the Poisson process (Illian et al., 2008). Other moment-based methods include the adapted estimator proposed by Mrkvička and Molchanov (2005) or the recent Stein estimator (in the Poisson case) proposed by Clausel et al. (2014).

As outlined in particular in the book written by Illian et al. (2008), an important step in the statistical analysis of point patterns is the search for unusual points or unusual point configurations, i.e. the search of outliers. Two kind of outliers make sense when dealing with point pattern: first points may appear at locations where they are not expected. This situation could appear for instance when two species of plants or trees cannot be distinguished at the time of data collection. Second, it is possible that there are unusual missing points in the pattern, i.e. areas of the observation domain, where, according to the general structure of the pattern, points would have been expected. Illian et al. (2008) or Baddeley et al. (2005) have proposed several diagnostic tools to detect outliers and more generally to judge the quality of fit of a model. Assunção and Guttorp (1999) is the only work where robustness estimation procedures are tackled. The authors developed an M-estimator to estimate the intensity of an inhomogeneous Poisson point process. To the best of our knowledge, no free-model robust techniques have been developed in the framework of spatial point processes. The present paper seems to be the first advance in that direction and aims at developing a median-based estimator of λ . It is not so straightforward to see what a median means for a spatial point process but we may remark that if W can be decomposed as a disjoint union of K cells C_k with similar volume, then $n(\mathbf{X} \cap W) = \sum_k n(\mathbf{X} \cap C_k)$ which yields that $\hat{\lambda}^{\text{std}}$ can be actually rewritten as the empirical mean of the normalized counts variables $n(\mathbf{X} \cap C_k)/|C_k|$. We have the cornerstone of defining a more robust estimator by simply replacing the empirical mean by the sample median.

The classical definition of sample quantiles and their asymptotic properties for

continuous distributions are nowadays well-known, see e.g. David and Nagaraja (2003). In particular, sample medians in the i.i.d. setting, computed from an absolutely continuous distribution positive, f , at the true median, Me , are consistent and satisfy a central limit theorem with asymptotic variance $1/4f(\text{Me})^2$. Such a result obviously fails for discrete distributions. In this paper, we follow a strategy introduced by Stevens (1950) and applied to count data by Machado and Santos Silva (2005) which consists in artificially imposing smoothness in the problem through jittering: i.e. we add to each count variable $n(\mathbf{X} \cap C_k)$ a random variable U_k following a uniform distribution on $(0, 1)$. The random variable $n(\mathbf{X} \cap C_k) + U_k$ admits now a density and asymptotic results can be expected. To get around the problem of large sample behavior for discrete distributions, another approach could be to consider the median based on the mid-distribution, see Ma et al. (2011). The authors prove that such sample quantiles behave more favourably than the classical one and satisfy, in the i.i.d. setting, a central limit theorem even if the distribution is discrete. We leave to a future work the question of deriving asymptotic properties for the sample median based on the mid-distribution in our framework of dependent spatial point processes models.

The rest of the paper is organized as follows. Section 2 gives the background on spatial point processes necessary for the present paper, proposes the general strategy to estimate the intensity and presents general notation. The median-based estimator and results establishing a control of the difference between the true median of a jittered count and the intensity parameter λ are described in Section 3. Section 4 contains our main asymptotic results. General assumptions are discussed and a particular focus on Cox point processes is investigated. The main difficulty here is to establish a Bahadur representation for the jittered sample median which can be applied to a large class of point process models. Section 5 presents the results of a simulation study where we compare our procedure with the standard estimator $\hat{\lambda}^{\text{std}}$. Proofs of the results and additional figures and comments are postponed to Appendices A and B

2 Background and strategy

2.1 Spatial point processes

Let \mathbf{X} be a spatial point process defined on \mathbb{R}^d , which we view as a random locally finite subset of \mathbb{R}^d . Let $\mathbf{X}_W = \mathbf{X} \cap W$ where $W \subset \mathbb{R}^d$ is a compact set of positive Lebesgue measure $|W|$. Then, the number of points in \mathbf{X}_W , denoted by $n(\mathbf{X}_W)$, is finite, and a realization of \mathbf{X}_W is of the form $\mathbf{x} = \{x_1, \dots, x_m\} \subset W$ for some nonnegative finite integer m . If $m = 0$, then $\mathbf{x} = \emptyset$ is an empty point pattern in W . For further background and measure theory on spatial point processes, see e.g. Daley and Vere-Jones (2003) and Møller and Waagepetersen (2004). We assume that \mathbf{X} is a stationary point process with intensity parameter λ , characterized by Campbell's theorem (see e.g. Møller and Waagepetersen (2004)), by the fact that for any real Borel function h defined on \mathbb{R}^d and absolutely integrable (with respect

to the Lebesgue measure on \mathbb{R}^d)

$$\mathbb{E} \sum_{u \in \mathbf{X}} h(u) = \lambda \int h(u) du. \quad (2.1)$$

Furthermore, for any integer $l \geq 1$, \mathbf{X} is said to have an l th-order product density ρ_l if ρ_l is a non-negative Borel function on \mathbb{R}^{dl} such that for all non-negative Borel functions h defined on \mathbb{R}^{dl} ,

$$\mathbb{E} \sum_{\substack{\neq \\ u_1, \dots, u_l \in \mathbf{X}}} h(u_1, \dots, u_l) = \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} h(u_1, \dots, u_l) \rho_l(u_1, \dots, u_l) du_1 \cdots du_l, \quad (2.2)$$

where the sign \neq over the summation means that u_1, \dots, u_l are pairwise distinct. Note that $\lambda = \rho_1$ and that for the homogeneous Poisson point process $\rho_l(u_1, \dots, u_l) = \lambda^l$. If $\rho^{(2)}$ exists, then by stationarity of \mathbf{X} , $\rho^{(2)}(u, v)$ depends only on $u - v$. In that case, we define the pair correlation function g as a function from \mathbb{R}^d to \mathbb{R}^+ by $g(u - v) = \rho^{(2)}(u, v) = \lambda^2$. In this paper, we sometimes pay attention on Cox point processes, which are defined as follows.

Definition 2.1. *Let $(\xi(s), s \in \mathbb{R}^d)$ be a non-negative locally integrable random field. Then, \mathbf{X} is a Cox point process if the distribution of \mathbf{X} given ξ is an inhomogeneous Poisson point process with intensity function ξ . If ξ is stationary, so is \mathbf{X} and $\lambda = \mathbb{E}(\xi(s))$ for any s .*

Among often used models of stationary Cox point processes, we can cite

- Log-Gaussian Cox processes (e.g. Møller and Waagepetersen (2004)): Let Y be a stationary Gaussian process on \mathbb{R}^d with mean μ and stationary covariance function $c(u) = \sigma^2 r(u)$, $u \in \mathbb{R}^d$, where $\sigma^2 > 0$ is the variance and r the correlation function. If \mathbf{X} conditional on Y is a Poisson point process with intensity function $\xi = \exp(Y)$, then \mathbf{X} is a (homogeneous) log-Gaussian Cox process. One example of correlation function is the Matérn correlation function (which includes the exponential correlation function) given by $r(u) = (\sqrt{2\nu}\|u\|/\phi)^\nu K_\nu(\sqrt{2\nu}\|u\|/\phi)/(2^{\nu-1}\Gamma(\nu))$ where Γ is the gamma function, K_ν is the modified Bessel function of the second kind, and ϕ and ν are non-negative parameters. In particular, the intensity of \mathbf{X} equals $\lambda = e^{\mu+\sigma^2/2}$.
- Neyman-Scott processes (e.g. Møller and Waagepetersen (2004)): Let \mathbf{C} be a stationary Poisson point process with intensity $\kappa > 0$, and f_σ a density function on \mathbb{R}^d . If \mathbf{X} conditional on \mathbf{C} is a Poisson point process with intensity

$$\alpha \sum_{c \in \mathbf{C}} f_\sigma(u - c)/\kappa, \quad u \in \mathbb{R}^d, \quad (2.3)$$

for some $\alpha > 0$, then \mathbf{X} is a (homogeneous) Neyman-Scott process. When f_σ corresponds to the density of a uniform distribution on $B(0, \sigma^2)$ (resp. a Gaussian random variable with mean 0 and variance σ^2), we refer to \mathbf{X} as the (homogeneous) Matérn Cluster (resp. Thomas) point process. In particular, the intensity of \mathbf{X} equals $\lambda = \alpha\kappa$.

2.2 Notation and strategy

For any real-valued random variable Y , we denote by $F_Y(\cdot)$ its cdf, by $F_Y^{-1}(p)$ its quantile of order $p \in (0, 1)$, by $\text{Me}_Y = F_Y^{-1}(1/2)$ its theoretical median. Based on a sample $\mathbf{Y} = (Y_1, \dots, Y_n)$ of n identically distributed random variables we denote by $\widehat{F}(\cdot; \mathbf{Y})$ the empirical cdf, by $\widehat{F}^{-1}(p; \mathbf{Y})$ the sample quantile of order p . The sample median is simply denoted by $\widehat{\text{Me}}(\mathbf{Y}) = \widehat{F}^{-1}(1/2; \mathbf{Y})$.

We will study the large-sample behavior of estimators of the intensity λ . Specifically, we consider a region W_n assumed to increase to \mathbb{R}^d as $n \rightarrow \infty$. We assume that the domain of observation W_n can be decomposed as $W_n = \cup_{k \in \mathcal{K}_n} C_{n,k}$ where the cells $C_{n,k}$ are disjoint cells with similar volume $c_n = |C_{n,k}|$ and where \mathcal{K}_n is a subset of \mathbb{Z}^d with cardinality $k_n = |\mathcal{K}_n|$. More details on W_n, c_n and k_n will be provided in the appropriate Section 4 when we present asymptotic results. For any $k \in \mathcal{K}_n$, we denote by $N_{n,k} = n(\mathbf{X}_{C_{n,k}})$ and by $\mathbf{N} = (N_{n,k}, k \in \mathcal{K}_n)$. Finally, for any random variable Y or any random vector \mathbf{Y} , we denote by $\check{Y} = Y/c_n$ and $\check{\mathbf{Y}} = \mathbf{Y}/c_n$.

The classical estimator of the intensity λ is given by $\widehat{\lambda}^{\text{std}} = n(\mathbf{X}_{W_n})/|W_n|$. In order to define a more robust estimator, we can remark that

$$\widehat{\lambda}^{\text{std}} = \frac{1}{k_n} \sum_{k \in \mathcal{K}_n} \frac{n(\mathbf{X}_{C_{n,k}})}{c_n} = \frac{1}{k_n} \sum_{k \in \mathcal{K}_n} \check{N}_{n,k} \quad (2.4)$$

since $|W_n| = k_n c_n$, i.e. $\widehat{\lambda}^{\text{std}}$ is nothing else than the sample mean of intensity estimators computed in cells $C_{n,k}$. The strategy we adopt in this paper is to replace the sample mean by the sample median which is known to be more robust to outliers. As underlined in the introduction, estimators based on count data or more generally on discrete data can cause some troubles in the asymptotic theory. The problems come from the fact that, in the continuous case, the asymptotic variance of the sample median involves the probability distribution function at the true median. We get around the difficulty in the next section by considering an estimator based on jittered data.

3 Median-based estimator of λ

To overcome the problem of discontinuity of the counts variables $N_{n,k}$, we follow a well-known technique (e.g. Machado and Santos Silva (2005)) which introduces smoothness in the problem. Let $(U_k, k \in \mathcal{K}_n)$ be a collection of independent and identically distributed random variables, distributed as $U \sim \mathcal{U}([0, 1])$. Then for any $k \in \mathcal{K}_n$, we define

$$Z_{n,k} = n(\mathbf{X}_{C_{n,k}}) + U_k = N_{n,k} + U_k \quad (3.1)$$

\mathbf{Z} will stand for $(Z_{n,k}, k \in \mathcal{K}_n)$. Since \mathbf{X} is stationary, the variables $Z_{n,k}$ are identically distributed and we let $Z \sim Z_{n,k}$. The jittering effect is straightforwardly seen: the cdf of Z is given for any $t \geq 0$ by

$$F_Z(t) = F_N(\lfloor t \rfloor - 1) + P(N = \lfloor t \rfloor) (t - \lfloor t \rfloor)$$

and for any t , Z admits a density f_Z at t given by

$$f_Z(t) = P(N = \lfloor t \rfloor). \quad (3.2)$$

We define the jittered estimator of the intensity λ by

$$\widehat{\lambda}^J = \widehat{\text{Me}}(\check{\mathbf{Z}}). \quad (3.3)$$

Since it is expected that $\widehat{\text{Me}}(\check{\mathbf{Z}})$ is close to $\text{Me}_{\check{Z}} = \text{Me}_Z/c_n$, we need to understand how far $\text{Me}_{\check{Z}}$ is from λ . Using the definition of the median we can prove the following.

Proposition 3.1. *Assume that the pair correlation function of the stationary point process \mathbf{X} exists for $u, v \in \mathbb{R}^d$ and satisfies $\int_{\mathbb{R}^d} |g(w) - 1| dw < \infty$, then for any $\varepsilon > 0$ we have for n sufficiently large*

$$|\text{Me}_{\check{Z}} - \lambda| \leq \frac{1}{c_n} \left(\frac{1}{2} + \sqrt{\frac{1}{12}} \right) + (1 + \varepsilon) \sqrt{\frac{\sigma}{c_n}} = \mathcal{O}(c_n^{-1/2}) \quad (3.4)$$

where $\sigma^2 = \lambda + \lambda^2 \int_{\mathbb{R}^d} (g(w) - 1) dw$.

The assumption is quite standard when we deal with asymptotic for spatial point processes, see e.g. Guan and Loh (2007) or Heinrich and Prokešová (2010). It ensures that for any regular domain Δ_n , $|\Delta_n|^{-1} \text{Var}(n(\mathbf{X}_{\Delta_n})) \rightarrow \sigma^2$ as $n \rightarrow \infty$. We refer the reader to these papers and to Section 4.1 for a discussion of this assumption.

Proof. By the previous remark $c_n^{-1} \text{Var}(N_{n,k}) \rightarrow \sigma^2$ as $n \rightarrow \infty$. Since for any continuous random variable Y having two moments $|\text{Me}_Y - \text{E}(Y)| \leq \sqrt{\text{Var}(Y)}$ and since $\text{E}(Z) = \lambda c_n + 1/2$, then for any $\varepsilon > 0$ we have for n sufficiently large

$$|\text{Me}_Z - \lambda c_n| \leq \frac{1}{2} + \sqrt{\frac{1}{12} + (1 + \varepsilon)^2 \sigma^2 c_n}$$

which leads to the result. □

Several results are known for the theoretical median of a Poisson distribution, see e.g. Adell and Jodrá (2005). For instance, when ν is an integer $\text{Me}_{\mathcal{P}(\nu)} = \nu$ and for non integer ν , $-\log 2 \leq \text{Me}_{\mathcal{P}(\nu)} \leq 1/3$ (see Figure 1). Based on this, we can obtain a sharper inequality than (3.4) for Poisson and Cox point processes.

Proposition 3.2. *Let \mathbf{X} be a stationary Cox point process with latent random field ξ , then*

$$\lambda c_n - \log 2 \leq \text{Me}_N \leq \lambda c_n + \frac{1}{3} \quad \text{and} \quad |\text{Me}_Z - \lambda c_n| \leq \frac{4}{3}. \quad (3.5)$$

A reformulation of (3.5) is of course $\text{Me}_{\check{Z}} - \lambda = \mathcal{O}(c_n^{-1})$.

Proof. For any $k \in \mathcal{K}_n$, given ξ , $N_{n,k}$ follows a Poisson distribution with intensity $\int_{C_{n,k}} \xi(s) ds$. Denote by $\text{Me}_{N_{n,k}|\xi}$ the median of $N_{n,k}$ given ξ defined by

$$\text{Me}_{N_{n,k}|\xi} = \inf \{ z \in \mathbb{R} : F_{N_{n,k}|\xi}(z) \geq 1/2 \}$$

where $F_{N_{n,k}|\xi}$ is the cumulative distribution function of $N_{n,k}$ given ξ . From the property of the median of a Poisson distribution, we have for any $k \in \mathcal{K}_n$

$$\int_{C_{n,k}} \xi(s) ds - \log 2 \leq \text{Me}_{N_{n,k}|\xi} \leq \int_{C_{n,k}} \xi(s) ds + \frac{1}{3}.$$

Since $E \int_{C_{n,k}} \xi(s) ds = \lambda c_n$, the first result is deduced by taking the expectation of each term of the previous inequality. Since $N \leq Z \leq N+1$, $\text{Me}_N \leq \text{Me}_Z \leq \text{Me}_N + 1$ which leads to the second result. \square

4 Asymptotic results

We state in this section our main results and the general assumptions required to obtain them. Proofs of the different results presented here, as well as auxiliary results, are presented in Appendix A.

4.1 General assumptions and discussion

We recall the classical definition of mixing coefficients (see e.g. Politis et al. (1998)): for $j, k \in \mathbb{N} \cup \{\infty\}$ and $m \geq 1$, define

$$\alpha_{j,k}(m) = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \mathcal{F}(\Lambda_1), B \in \mathcal{F}(\Lambda_2), \\ \Lambda_1 \in \mathcal{B}(\mathbb{R}^d), \Lambda_2 \in \mathcal{B}(\mathbb{R}^d), |\Lambda_1| \leq j, |\Lambda_2| \leq k, d(\Lambda_1, \Lambda_2) \geq m\}$$

where $\mathcal{F}(\Lambda_i)$ is the σ -algebra generated by $\mathbf{X} \cap \Lambda_i$, $i = 1, 2$, $d(\Lambda_1, \Lambda_2)$ is the minimal distance between the sets Λ_1 and Λ_2 , and $\mathcal{B}(\mathbb{R}^d)$ denotes the class of Borel sets in \mathbb{R}^d .

We require the following assumptions to derive our asymptotic results.

(i) For any $n \geq 1$, we assume that $W_n = \cup_{k \in \mathcal{K}_n} C_{n,k}$ where \mathcal{K}_n is a subset of \mathbb{Z}^d with cardinality $k_n = |\mathcal{K}_n|$ and where the cells $C_{n,k}$ are disjoint cubes with volume c_n defined by

$$C_{n,k} = \{u = (u_1, \dots, u_d)^\top \in \mathbb{R}^d : c_n^{1/d}(k_l - 1/2) \leq u_l \leq c_n^{1/d}(k_l + 1/2), l = 1, \dots, d\}.$$

We assume that $0 \in \mathcal{K}_n$ and that there exists $0 < \eta' < \eta$ such that as $n \rightarrow \infty$

$$k_n \rightarrow \infty, \quad c_n \rightarrow \infty, \quad \frac{k_n}{c_n^{\eta'/2 \wedge (1-2\ell)}} \rightarrow 0$$

where ℓ is given by Assumption (ii) and η by Assumption (iv).

(ii)

(ii-1) $\text{Me}_Z - \lambda c_n = \mathcal{O}(c_n^\ell)$ with $0 \leq \ell < 1/2$.

(ii-2) $\forall t_n = \lambda c_n + \mathcal{O}(\sqrt{c_n/k_n})$, $P(N = \lfloor t_n \rfloor)/P(N = \lfloor \lambda c_n \rfloor) \rightarrow 1$.

(ii-3) There exists $\underline{\kappa}, \bar{\kappa} > 0$ such that for n large enough, $\underline{\kappa} \leq \sqrt{c_n} f_Z(\text{Me}_Z) \leq \bar{\kappa}$.

(iii) \mathbf{X} has a pair correlation function g satisfying $\int_{\mathbb{R}^d} |g(w) - 1| dw < \infty$.

(iv) There exists $\eta > 0$ such that

$$\alpha(m) = \sup_{p \geq 1} \frac{\alpha_{p,p}(m)}{p} = \mathcal{O}(m^{-d(1+\eta)}) \quad \text{and} \quad \alpha_{2,\infty}(m) = \mathcal{O}(m^{-d(1+\eta)}).$$

We now discuss the different assumptions. The decomposition of W_n as a disjoint union of cubes is really related to our estimation procedure. Other shapes of domains W_n should be possible. The last statement of Assumption (i) is required to control the dependence between the variables $Z_{n,k}$ via a control of mixing coefficients and to ensure that asymptotically $|W_n|^{1/2}(\widehat{\lambda}^J - \lambda)$ behaves as $|W_n|^{1/2}(\widehat{\lambda}^J - \text{Me}_Z)$. We note that if \mathbf{X} is a stationary Cox point process, Proposition 3.2 yields that (ii-1) is satisfied for $\ell = 0$. And so if $\eta > 2$, Assumption (i) can be rewritten as $c_n \rightarrow \infty$, $k_n \rightarrow \infty$ and $k_n/c_n \rightarrow 0$ as $n \rightarrow \infty$. Regarding Assumption (ii), Proposition 4.1 shows it can be simplified for a large class of Cox point processes. We underline that Assumptions (i), (ii-1)-(ii-2) imply the existence of $\bar{\kappa} < \infty$ such that $\sqrt{c_n}f_Z(\text{Me}_Z) \leq \bar{\kappa}$, so (ii-3) could actually be simplified. Assumption (iii) is very classical when dealing with asymptotics of intensity parameter estimates, see e.g. Heinrich and Prokešová (2010). For isotropic pair correlation functions, i.e. $g(w) = \bar{g}(\|w\|)$ for $\bar{g} : \mathbb{R}^+ \rightarrow \mathbb{R}$, Assumption (iii) is fulfilled when $\bar{g}(r) = 0$ for $r \geq R$ or when $\bar{g}(r) = \mathcal{O}(r^{-d-\gamma})$ for some $\gamma > d$. This includes the Matérn cluster and Thomas processes and the log-Gaussian Cox process with Matérn-Whittle covariance functions. Assumption (iv) is also quite standard and has been discussed a lot in the literature: Guan and Loh (2007); Guan et al. (2007); Prokešová and Jensen (2013) discussed the first part of (iv) while the second one has been commented in Waagepetersen and Guan (2009); Coeurjolly and Møller (2014). Both of them are satisfied for Cox point processes mentioned above. We point out that it is not so common to use both the mixing coefficients $\alpha(m)$ and $\alpha_{2,\infty}(m)$. As detailed in the proof of Theorem 4.2, the first one is used to control the dependence between the random variables $Z_{n,k}$ for $k \in \mathcal{K}_n$ and derive a central limit theorem using the blocking technique developed by Ibragimov and Linnik (1971) which is pertinent and well-suited here since the cells $C_{n,k}$ are increasing. The second mixing coefficient is used to apply a multivariate central limit theorem inside the cell $C_{n,0}$ to prove, in particular, that $\text{P}(n(\mathbf{X} \cap C_{n,0}) \leq \lambda c_n, n(\mathbf{X} \cap C_{n,0}^-) \leq \lambda c_n) \rightarrow 1/2$ as $n \rightarrow \infty$ where $C_{n,0}^-$ is a "small" erosion of $C_{n,0}$ (see the proof of Step 1 of Theorem 4.2 for more details).

We now present how some of the assumptions can be simplified for Cox point processes.

Proposition 4.1. *Let \mathbf{X} be a stationary Cox point process with latent random field $(\xi(s), s \in \mathbb{R}^d)$ satisfying the Assumptions (iii)-(iv). Assume there exists $\delta > 2/\eta$, where η is given by Assumption (iv), such that $\text{E}(|\xi(0)|^{2+\delta}) < \infty$. Let $t_n = \lambda c_n + \mathcal{O}(\sqrt{c_n/k_n})$ and $T_n = \lfloor t_n \rfloor^{-1} \int_{C_{n,0}} \xi(s) ds$. We also assume that the sequence of random variables $(B_n)_n$ defined by $\log(B_n) = \lfloor t_n \rfloor (\log(T_n) - (T_n - 1) + (T_n - 1)^2/2)$ is uniformly integrable. Then Assumption (ii) holds (with $\ell = 0$) and as $n \rightarrow \infty$*

$$\sqrt{c_n} \text{P}(N = \lfloor \lambda c_n \rfloor) \rightarrow (2\pi\sigma^2)^{-1/2} \quad (4.1)$$

where N is a random variable distributed as $N_{n,0} = n(\mathbf{X}_{C_{n,0}})$ and where $\sigma^2 = \lambda + \lambda^2 \int_{\mathbb{R}^d} (g(w) - 1) dw$.

4.2 Results

In this section, we present, in particular, the asymptotic results we obtained for the median-based estimator $\widehat{\lambda}^J$.

Theorem 4.2. *Under the Assumptions (i)-(iv), we have the two following statements.*

(a) *Let $(a_n)_{n \geq 1}$ be a sequence of real numbers satisfying $\lambda c_n = a_n + o(\sqrt{c_n})$ then as $n \rightarrow \infty$*

$$\sqrt{k_n} \left(\widehat{F}(\lambda c_n + a_n; \mathbf{Z}) - F_Z(\lambda c_n + a_n) \right) \rightarrow \mathcal{N}(0, 1/4)$$

in distribution.

(b) *As $n \rightarrow \infty$*

$$\sqrt{k_n} \left(\widehat{F}(\text{Me}_Z; \mathbf{Z}) - 1/2 \right) \rightarrow \mathcal{N}(0, 1/4)$$

in distribution.

The next result establishes a Bahadur representation for the sample median which leads to its asymptotic normality. The notation $X_n = o_{\mathbb{P}}(v_n^{-1})$ for a sequence of random variables X_n and a sequence of positive real numbers v_n means that $v_n X_n$ tends to 0 in probability as $n \rightarrow \infty$.

Theorem 4.3. *Under the assumptions (i)-(iv), we have the two following statements.*

(a) *As $n \rightarrow \infty$*

$$\widehat{\text{Me}}(\mathbf{Z}) - \text{Me}_Z = \frac{1/2 - \widehat{F}(\text{Me}_Z; \mathbf{Z})}{f_Z(\text{Me}_Z)} + o_{\mathbb{P}} \left(\sqrt{\frac{c_n}{k_n}} \right). \quad (4.2)$$

(b) *Let $s_n = \sqrt{c_n} \mathbb{P}(N = \lfloor \lambda c_n \rfloor)$ where N is a random variable distributed as $N_{n,0} = n(\mathbf{X}_{C_{n,0}})$, then as $n \rightarrow \infty$*

$$|W_n|^{1/2} s_n \left(\widehat{\lambda}^J - \lambda \right) \rightarrow \mathcal{N}(0, 1/4) \quad (4.3)$$

in distribution.

We deduce the following Corollary given without proof for Cox point processes.

Corollary 4.4. *Under the Assumption (i) and the Assumptions of Proposition 4.1, we have*

$$|W_n|^{1/2} \left(\widehat{\lambda}^J - \lambda \right) \rightarrow \mathcal{N} \left(0, \pi \sigma^2 / 2 \right)$$

where $\sigma^2 = \lambda + \lambda^2 \int_{\mathbb{R}^d} (g(w) - 1) dw$.

As detailed after Proposition 3.1, σ^2 corresponds to the asymptotic variance of $|W_n|^{-1} n(\mathbf{X}_{W_n})$. Actually, if we denote by $\widehat{\lambda}^{\text{std}}$ the standard estimator of λ given by $\widehat{\lambda}^{\text{std}} = |W_n|^{-1} n(\mathbf{X}_{W_n})$ then with quite similar assumptions, it has been proved, see e.g. Heinrich and Prokešová (2010), that $|W_n|^{1/2} (\widehat{\lambda}^{\text{std}} - \lambda) \rightarrow \mathcal{N}(0, \sigma^2)$. It is worth interesting to note that the two estimators $\widehat{\lambda}^{\text{std}}$ and $\widehat{\lambda}^J$ only differ on their asymptotic variance and that the ratio of the asymptotic variances is equal to $\pi/2$. When we estimate the location of a Gaussian sample using the sample mean or the sample median, it is remarkable that the ratio of the asymptotic variances is also $\pi/2$.

Finally, let us add that on the basis of Corollary 4.4, an asymptotic confidence interval of λ can be constructed using a consistent estimator of σ^2 . By the previous remark, we can use the kernel-based estimator proposed by Heinrich and Prokešová (2010) (or any other estimator presented in the mentioned paper), which precisely estimates the asymptotic variance of $\widehat{\lambda}^{\text{std}}$, i.e. σ^2 .

5 Simulation study

We present in this section a simulation study where in particular we intend to compare the median-based estimator defined by (3.3) with the standard moment-based estimator $\widehat{\lambda}^{\text{std}} = n(\mathbf{X}_W)/|W|$. We focus on the planar case $d = 2$. In all this section, we fixed the intensity parameter to the value $\lambda = 50$. Three models of spatial point processes are considered (see Section 2.1 for details):

- Poisson point processes (referred to as POISSON) with intensity λ
- Thomas point processes (referred to as THOMAS) with the parameters $\kappa = 10$ and $\sigma = 0.05$. The parameter α is fixed by the relation $\alpha = \kappa/\lambda$.
- Log-Gaussian Cox Processes (referred to as LGCP) point processes with exponential covariance function. We fixed the variance to 1 and ϕ to 0.05. The parameter μ is fixed by the relation $\mu = \log \lambda - \sigma^2/2$.

Figure 2 illustrates these three models and in particular the clustering effect inherent to the two last ones. The simulations have been performed using the R package `spatstat` (Baddeley and Turner, 2005). To illustrate the performances of (3.3) we generated the point processes on the domain of observation $W_n = [-n, n]^2$ for different values of n and considered the three following settings: let \mathbf{y} be a replication of one of the three models above generated on W_n

- (A) Pure case: no modification is considered, $\mathbf{x}_{W_n} = \mathbf{y}$.
- (B) A few points are added: in a sub-square Δ_n with side-length $n/5$ included in W_n and randomly chosen, we have generated a point process \mathbf{y}^{add} of $n^{\text{add}} = \rho n(\mathbf{y})$ uniform points in Δ_n . We chose $\rho = 10\%$ or 20% . Then we defined $\mathbf{x}_{W_n} = \mathbf{y} \cup \mathbf{y}^{\text{add}}$.
- (C) A few points are deleted: let $\Delta_n = \cup_{m=1}^4 \Delta_n^m$ where the Δ_n^m 's are the four squares included in W_n , located in each corner of W_n and with similar volume. The volume $|\Delta_n|$ is chosen such that $E(n(\mathbf{Y}_{\Delta_n})) = \rho E(n(\mathbf{Y})) = \rho \lambda |W_n|$ and we chose either $\rho = 10\%$ or 20% . Then, we define $\mathbf{x}_{W_n} = \mathbf{y} \setminus \mathbf{y}_{\Delta_n}$, .i.e. \mathbf{x}_{W_n} is the initial configuration thinned by 10% or 20% of its points.

We conducted a Monte-Carlo simulation and generated 1000 replications of the models POISSON, THOMAS, LGCP and for the three different settings (A)-(C). The observation windows for which we report the empirical results hereafter are $n = 1, 2$ when we consider the setting (A) or (B) and $n = 2, 3$ when we consider the setting (C). The last setting requires more observable points to clearly see the advantages of the median-based estimator. Regarding that setting (C), we placed the squares where points are thinned at the corners of W_n . By stationarity, the empirical results are the same if we had decided to choose them randomly.

For each replication, we evaluated $\widehat{\lambda}^{\text{std}}$ and $\widehat{\lambda}^J$ for different number of cells k_n with same volume. More precisely, we chose $k_n = 9, 16, 25, 36, 49$.

Table 1 and Figure 3 in Appendix B are related to the case (A). Empirical means and standard deviations are reported. We can check, as expected, that the standard

estimator is of course unbiased and that the variance decreases by a factor close to 4 which is equal to $|W_2|/|W_1|$. The median-based estimator is not theoretically unbiased but the bias is clearly not important and tend to decrease when the observation window grows up. Similarly, the rate of convergence of the empirical variance is not too far from the expected value 4. We also computed separately $\widehat{\text{Var}}(\widehat{\lambda}^{\text{std}})/\widehat{\text{Var}}(\widehat{\lambda}^J)$ for each value of k_n and n and found interesting that these ratios lie in the interval $[1.31, 1.67]$, i.e. not too far from $\pi/2$. Finally, we underline that the choice of the number of cells k_n has a little influence on the performances. When $n = 1$, a too large value of k_n seems to increase the bias, especially for the THOMAS model. The differences are however erased when $n = 2$. We also remark that the empirical variance is almost the same whatever the value of k_n .

Tables 2 and 3 are respectively related to the settings (B) and (C) described above and are summaries of Figures 4 and 5 postponed to Appendix B. In both tables, we reported only the gain (in percent) in terms of mean squared error of the median based-estimator with respect to the standard one, i.e. for each model and each value of ρ, n, k_n , we computed

$$\widehat{\text{Gain}} = \left(\frac{\widehat{\text{MSE}}(\widehat{\lambda}^{\text{std}}) - \widehat{\text{MSE}}(\widehat{\lambda}^J)}{\widehat{\text{MSE}}(\widehat{\lambda}^{\text{std}})} \right) \times 100\% \quad (5.1)$$

where $\widehat{\text{MSE}}$ is the empirical mean squared error based on the 1000 replications. Thus a positive (resp. negative) empirical gain means that the median-based is more efficient (resp. less efficient) than the standard procedure.

The standard estimator, based only on the global number of points, is of course not robust to perturbations. It is clearly seen in Figures 4 and 5 where we can observe that $\widehat{\lambda}^{\text{std}}$ is more and more biased as ρ (the ratio of points added or deleted) increases. Unlike this, the median-based estimator shows its advantages. When points are added (setting (B)), the estimator $\widehat{\lambda}^J$ remains much more stable and is more efficient in terms of MSE except when $\rho = 10\%$, $n = 1$ for the THOMAS and LGCP models. But when the window of observation expands or when ρ takes a higher value, the conclusions are unambiguous. The empirical gains are more important under the POISSON and LGCP models than for the very clustered THOMAS process. Still, when $\rho = 20\%$, the gain is at least 9% for the worst choice of k_n and at least 30% if k_n is appropriately chosen. We also mention that in a separate simulation study not show here, we also tried to add a clustered point process or repulsive point process, instead of adding $\rho n(\mathbf{x}_{W_n})$ uniform points. The empirical results remained almost unchanged.

Comments regarding Table 3 (setting (C)) are very similar. It seems however that the choice of k_n is more sensitive. For example for the LGCP model and when $\rho = 20\%$ and $n = 2$ the empirical gain is 29% when $k_n = 9$ and can reach 75% when $k_n = 16$.

As a conclusion of the simulation study, it turns out that the estimator $\widehat{\lambda}^J$ clearly satisfies expected asymptotic properties and improves the robustness property of the standard procedure. It is the topic of a future work to propose a data-driven procedure to select the number of cells k_n .

	$\widehat{\lambda}^{\text{std}}$	$\widehat{\lambda}^J$				
		$k_n = 9$	16	25	36	49
POISSON, $n = 1$	49.9 (3.5)	50.7 (4.3)	51.2 (4.2)	51.9 (4.2)	53.0 (4.2)	54.0 (4.5)
$n = 2$	50.0 (1.7)	50.2 (2.1)	50.4 (2.2)	50.5 (2.2)	50.8 (2.1)	51.1 (2.1)
THOMAS, $n = 1$	49.8 (9.0)	49.1 (10.4)	48.5 (10.3)	47.2 (10.6)	45.9 (10.4)	44.0 (11.2)
$n = 2$	49.9 (4.3)	49.7 (5.1)	49.5 (5.1)	49.1 (5.2)	48.8 (5.3)	48.4 (5.3)
LGCP, $n = 1$	49.8 (4.6)	50.2 (5.5)	50.3 (5.4)	50.4 (5.4)	51.0 (5.3)	51.6 (5.5)
$n = 2$	50.0 (2.5)	50.0 (3.0)	50.1 (3.1)	50.1 (3.0)	50.0 (3.0)	50.1 (3.0)

Table 1: Empirical means and standard deviations between brackets of estimates of the intensity $\lambda = 50$ for different models of spatial point processes (POISSON, THOMAS, LGCP). The empirical results are based on 1000 replications simulated on $[-n, n]^2$ for $n = 1, 2$. The second and third columns corresponds to the standard estimator $\widehat{\lambda}^{\text{std}} = n(\mathbf{X}_{W_n})/|W_n|$, while the following ones correspond to the median-based estimator (3.3) for different number of cells k_n .

	Gain of MSE (%)				
	$k_n = 9$	16	25	36	49
$\rho = 10\%$, POISSON, $n = 1$	3	0	-18	-48	-76
$n = 2$	62	65	63	61	56
THOMAS, $n = 1$	-26	-16	-13	-17	-30
$n = 2$	3	19	19	28	33
LGCP, $n = 1$	-5	-5	-5	-7	-17
$n = 2$	45	54	54	56	52
$\rho = 20\%$, POISSON, $n = 1$	58	60	53	46	31
$n = 2$	86	89	88	88	86
THOMAS, $n = 1$	9	21	24	31	26
$n = 2$	59	60	63	68	70
LGCP, $n = 1$	47	53	52	52	46
$n = 2$	78	82	83	84	84

Table 2: Empirical gains in percent, see (5.1), of the median based estimator for different values of k_n . The empirical results are based on 1000 replications generated on $[-n, n]^2$ for $n = 1, 2$ for the models POISSON, THOMAS, LGCP where $\rho = 10\%$ or 20% of points are added to each configuration. This corresponds to the case (B) described in details above.

	Gain of MSE (%)				
	$k_n = 9$	16	25	36	49
$\rho = 10\%$, POISSON, $n = 2$	3	52	69	78	74
$n = 3$	31	76	84	87	83
THOMAS, $n = 2$	-29	-29	-41	-22	-45
$n = 3$	-20	-0	9	33	13
LGCP, $n = 2$	-9	22	41	58	46
$n = 3$	1	56	69	76	66
$\rho = 20\%$, POISSON, $n = 2$	56	86	78	59	59
$n = 3$	78	93	83	69	76
THOMAS, $n = 2$	-20	25	19	-9	-46
$n = 3$	7	64	53	23	5
LGCP, $n = 2$	29	75	64	41	33
$n = 3$	63	88	76	56	57

Table 3: Empirical gains in percent, see (5.1), of the median based estimator for different values of k_n . The empirical results are based on 1000 replications generated on $[-n, n]^2$ for $n = 2, 3$ for the models POISSON, THOMAS, LGCP where $\rho = 10\%$ or 20% of points are deleted to each configuration. This corresponds to the case (C) described in details above.

Acknowledgements

The author would like to thank sincerely Michaela Prokešová for discussing the initial idea of this paper and Frédéric Lavancier for discussions on an earlier version of the manuscript.

A Proofs

In all the proofs, κ denotes a generic constant which may be different from line to line. For $k = (k_1, \dots, k_d)^\top \in \mathbb{Z}^d$, we denote by $|k|$ the norm, $|k| = \max(|k_1|, \dots, |k_d|)$.

A.1 Proof of Proposition 4.1

Proof. Assumption (ii-1) corresponds to Proposition 3.2.

Assumptions (ii-2) and (ii-3). By definition of \mathbf{X} ,

$$\sqrt{2\pi\lambda c_n} \mathbb{P}(N_{n,0} = \lfloor t_n \rfloor \mid \xi) = \frac{\left(\int_{C_{n,0}} \xi(s) ds \right)^{\lfloor t_n \rfloor} e^{-\int_{C_{n,0}} \xi(s) ds}}{\lfloor t_n \rfloor! e^{-\lfloor t_n \rfloor}} v_n$$

where

$$v_n = \sqrt{\frac{\lambda c_n}{\lfloor t_n \rfloor} \frac{\sqrt{2\pi} \lfloor t_n \rfloor^{\lfloor t_n \rfloor + 1/2} e^{-\lfloor t_n \rfloor}}{\lfloor t_n \rfloor!}}.$$

Since $t_n/(\lambda c_n) \rightarrow 1$ as $n \rightarrow \infty$, then using Stirling's Formula we obviously have $v_n \rightarrow 1$ as $n \rightarrow \infty$. Now using the notation $T_n = \lfloor t_n \rfloor^{-1} \int_{C_{n,0}} \xi(s) ds$, we rewrite the

first equation as follows

$$(v_n)^{-1} \sqrt{2\pi\lambda c_n} \mathbb{P}(N_{n,0} = \lfloor t_n \rfloor \mid \xi) = T_n^{\lfloor t_n \rfloor} e^{\lfloor t_n \rfloor (1-T_n)} = A_n B_n$$

where A_n and B_n are defined by

$$A_n = e^{-\lfloor t_n \rfloor (T_n - 1)^2 / 2} \quad \text{and} \quad B_n = e^{\lfloor t_n \rfloor (\log T_n - (T_n - 1) + (T_n - 1)^2 / 2)}.$$

Since $\mathbb{E} |\xi(0)|^{2+\delta} < \infty$ for some $\delta > 2/\eta$ where η is given by Assumption (iv) we ensure that $\alpha_{2,\infty} = \mathcal{O}(m^{-\nu})$ for some $\nu > d(2+\delta)/\delta$. Therefore, we can apply Guyon (1991, Theorem 3.3.1) and show that there exists $\tau > 0$ such that $\sqrt{\lambda c_n} (I_n - 1) \rightarrow \mathcal{N}(0, \tau^2)$ in distribution where $I_n = (\lambda c_n)^{-1} \int_{C_{n,0}} \xi(s) ds$. To compute τ^2 , we observe that using the definition of a Cox point process

$$\begin{aligned} \text{Var}(N_{n,0}) &= \text{Var}(n(\mathbf{X}_{C_{n,0}})) = \mathbb{E} \left(\text{Var}(n(\mathbf{X}_{C_{n,0}}) \mid \xi) \right) + \text{Var} \left(\mathbb{E} (n(\mathbf{X}_{C_{n,0}})) \mid \xi \right) \\ &= \lambda c_n + \text{Var} \int_{C_{n,0}} \xi(s) ds. \end{aligned}$$

We use Assumption (iii) and Lemma A.1 to deduce that as $n \rightarrow \infty$

$$\text{Var} \int_{C_{n,0}} \xi(s) ds \sim \lambda^2 c_n \int_{\mathbb{R}^d} (g(w) - 1) dw$$

which leads to $\text{Var}(\sqrt{\lambda c_n} I_n) \sim \lambda \int_{\mathbb{R}^d} (g(w) - 1) dw$ as $n \rightarrow \infty$. From the definition of t_n and Slutsky's Lemma, it can be shown that $\sqrt{\lfloor t_n \rfloor} (T_n - I_n) \rightarrow 0$ in probability which leads to $T_n \rightarrow 1$ in probability and $\sqrt{\lfloor t_n \rfloor} (T_n - 1) \rightarrow \mathcal{N}(0, \tau^2)$ in distribution. We deduce that $A_n \rightarrow A = e^{-\tau^2 L^2 / 2}$ in distribution, where $L \sim \mathcal{N}(0, 1)$, which, by the uniform integrability of the sequence $(A_n)_n$, leads to $A_n \rightarrow A$ in L^1 . Now a Taylor expansion shows that there exists $\tilde{T}_n \in (0 \wedge (T_n - 1), 0 \vee (T_n - 1))$ such that

$$|\log(B_n)| = \lfloor t_n \rfloor |T_n - 1| \frac{\tilde{T}_n^2}{1 + \tilde{T}_n} \leq \lfloor t_n \rfloor (T_n - 1)^2 \frac{|T_n - 1|}{\tilde{T}_n + 1}.$$

It is clear that \tilde{T}_n tends to 0 in probability as $n \rightarrow \infty$, which yields that $\log(B_n) \rightarrow 0$ and $B_n \rightarrow 1$ in probability by Slutsky's Lemma. Again, the uniform integrability assumption of the sequence $(B_n)_n$ implies that $B_n \rightarrow 1$ in L^1 . Since $|A_n B_n - A| \leq |A_n - A| + |A| |B_n - 1|$, we conclude that $A_n B_n \rightarrow A$ in L^1 as $n \rightarrow \infty$. In other words as $n \rightarrow \infty$

$$\sqrt{2\pi\lambda c_n} \mathbb{P}(N = \lfloor t_n \rfloor) \sim v_n^{-1} \mathbb{E} \left(\sqrt{2\pi\lambda c_n} \mathbb{P}(N_{n,0} = \lfloor t_n \rfloor \mid \xi) \right) \rightarrow \mathbb{E}(A).$$

Using the definition of the moment generating function of a χ_1^2 distribution, we have $\mathbb{E}(A) = (1 + \tau^2)^{-1/2}$ whereby we deduce that

$$\sqrt{c_n} \mathbb{P}(N = \lfloor t_n \rfloor) \rightarrow (2\pi\lambda(1 + \tau^2))^{-1/2} = (2\pi\sigma^2)^{-1/2}$$

with $\sigma^2 = \lambda + \lambda^2 \int_{\mathbb{R}^d} (g(w) - 1) dw$. □

A.2 Proof of Theorem 4.2

Proof. We focus only on (a) as (b) follows from (a), Slutsky's Lemma and Assumption (ii-2). Let $t_n = \lambda c_n + a_n$. By definition

$$\widehat{F}(t_n; \mathbf{Z}) - F_Z(t_n) = \frac{1}{k_n} \sum_{k \in \mathcal{K}_n} (\mathbf{1}(Z_{n,k} \leq t_n) - \mathbb{P}(Z_{n,k} \leq t_n)).$$

We let $(\varepsilon_n)_{n \geq 1}$ be a sequence of real numbers such that $\varepsilon_n \rightarrow 0$ and $\varepsilon_n c_n^{1/d} \rightarrow \infty$ as $n \rightarrow \infty$. We denote by $Z_{n,k}^- = n(\mathbf{X} \cap C_{n,k}^-) + U_k$ where $C_{n,k}^-$ is the erosion of the cell $C_{n,k}$ by a closed ball centered at k and with radius $\varepsilon_n c_n^{1/d}$. Two cells $C_{n,k}^-$ and $C_{n,k'}^-$ for $k, k' \in \mathcal{K}_n$ ($k \neq k'$) are therefore at distance greater than $2\varepsilon_n c_n^{1/d}$. To prove Theorem 4.2 (a), we use the blocking technique introduced by Ibragimov and Linnik (1971) and applied to spatial point processes by Guan and Loh (2007); Guan et al. (2007) and Prokešová and Jensen (2013). To this end, we need additional notation. For any $n \geq 1$ and $k \in \mathcal{K}_n$, let $t_n^- = \lambda|C_{n,k}^-| + 1/2 = \lambda(1 - \varepsilon_n)^d c_n + 1/2$ and let $(\widetilde{Z}_{n,k}^-, k \in \mathcal{K}_n)$ be a collection of independent random variables such that $\widetilde{Z}_{n,k}^- \stackrel{d}{=} Z_{n,k}^-$. Finally we define the random variables $D_{n,k}$ by

$$D_{n,k} = \mathbf{1}(Z_{n,k} \leq t_n) - \mathbb{P}(Z_{n,k} \leq t_n) - \mathbf{1}(Z_{n,k}^- \leq t_n^-) + \mathbb{P}(Z_{n,k}^- \leq t_n^-).$$

We decompose the proof into three steps. As $n \rightarrow \infty$, we prove that

Step 1. $D_n/\sqrt{k_n} \rightarrow 0$ in probability where $D_n = \sum_{k \in \mathcal{K}_n} D_{n,k}$.

Step 2. for any $u \in \mathbb{R}$, $\phi_n^-(u) - \widetilde{\phi}_n^-(u) \rightarrow 0$ as $n \rightarrow \infty$ where $i = \sqrt{-1}$, $\phi_n^-(u) = \mathbb{E}(e^{iuS_n^-/\sqrt{k_n}})$ and $\widetilde{\phi}_n^-(u) = \mathbb{E}(e^{iu\widetilde{S}_n^-/\sqrt{k_n}})$. This will imply that $(S_n^- - \widetilde{S}_n^-)/\sqrt{k_n} \rightarrow 0$ in probability where

$$S_n^- = \sum_{k \in \mathcal{K}_n} \mathbf{1}(Z_{n,k}^- \leq t_n^-) - \mathbb{P}(Z_{n,k}^- \leq t_n^-) \text{ and } \widetilde{S}_n^- = \sum_{k \in \mathcal{K}_n} \mathbf{1}(\widetilde{Z}_{n,k}^- \leq t_n^-) - \mathbb{P}(\widetilde{Z}_{n,k}^- \leq t_n^-).$$

Step 3. $\widetilde{S}_n^-/\sqrt{k_n} \rightarrow \mathcal{N}(0, 1/4)$ in distribution.

The conclusion will follow directly from Steps 1-3 and Slutsky's Lemma.

Step 1. To achieve this step, we prove that $k_n^{-1} \text{Var}(D_n) \rightarrow 0$ as $n \rightarrow \infty$. We have

$$\frac{1}{k_n} \text{Var}(D_n) = \frac{1}{k_n} \sum_{\substack{k, k' \in \mathcal{K}_n \\ |k-k'| \leq 1}} \text{Cov}(D_{n,k}, D_{n,k'}) + \frac{1}{k_n} \sum_{\substack{k, k' \in \mathcal{K}_n \\ |k-k'| > 1}} \text{Cov}(D_{n,k}, D_{n,k'}).$$

Let $k, k' \in \mathcal{K}_n$ $k \neq k'$, Assumption (i) asserts that $d(C_{n,k}, C_{n,k'}) = |k - k' - 1|c_n^{1/d}$. Since $D_{n,k} \in \mathcal{F}(C_{n,k})$ and $D_{n,k'} \in \mathcal{F}(C_{n,k'})$, we have from Zhengyan and Chuanrong (1996, Lemma 2.1)

$$\begin{aligned} \text{Cov}(D_{n,k}, D_{n,k'}) &\leq 4\alpha_{c_n, c_n} (|k - k' - 1|c_n^{1/d}) \\ &\leq 4c_n \alpha (|k - k' - 1|c_n^{1/d}) = \mathcal{O}(|k - k' - 1|^{-d(1+\eta)} c_n^{-\eta}). \end{aligned}$$

Since the series $\sum_{k \in \mathbb{Z}^d \setminus \{0\}} |k|^{-d(1+\eta)}$ is convergent, it is clear that

$$\frac{1}{k_n} \sum_{\substack{k, k' \in \mathcal{K}_n \\ |k-k'| > 1}} \text{Cov}(D_{n,k}, D_{n,k'}) = \mathcal{O}(c_n^{-\eta}) \quad (\text{A.1})$$

which tends to 0 as $n \rightarrow \infty$. From Cauchy-Schwarz's inequality and since the variables $D_{n,k}$ are identically distributed

$$\begin{aligned} \left| \frac{1}{k_n} \sum_{\substack{k, k' \in \mathcal{K}_n \\ |k-k'| \leq 1}} \text{Cov}(D_{n,k}, D_{n,k'}) \right| &\leq \frac{1}{k_n} \sum_{\substack{k, k' \in \mathcal{K}_n \\ |k-k'| \leq 1}} \sqrt{\text{Var}(D_{n,k}) \text{Var}(D_{n,k'})} \\ &\leq \text{Var}(D_{n,0}) \frac{1}{k_n} \sum_{\substack{k, k' \in \mathcal{K}_n \\ |k-k'| \leq 1}} 1 \\ &\leq 3^d \text{Var}(D_{n,0}). \end{aligned}$$

Thus, Step 1 is achieved once we prove that $\text{Var} D_{n,0} \rightarrow 0$ as $n \rightarrow \infty$. A straightforward calculation yields that

$$\begin{aligned} \text{Var}(D_{n,0}) &= \text{P}(Z_{n,0} \leq t_n)(1 - \text{P}(Z_{n,0} \leq t_n)) + \text{P}(Z_{n,0}^- \leq t_n^-)(1 - \text{P}(Z_{n,0}^- \leq t_n^-)) \\ &\quad + 2\text{P}(Z_{n,0} \leq t_n)\text{P}(Z_{n,0}^- \leq t_n^-) - 2\text{P}(Z_{n,0} \leq t_n, Z_{n,0}^- \leq t_n^-). \end{aligned}$$

Let Δ_j the unit cube centered at $j \in \mathbb{Z}^d$, let \mathcal{J}_n be the subset of \mathbb{Z}^d of cubes Δ_j which intersect $C_{n,0}$. We denote by $Y_{n,j}$ the following random vector

$$Y_{n,j} = \left(\frac{U_{n,0}}{j_n} + \mathbf{1}(u \in C_{n,0} \cap \Delta_j), \frac{U_{n,0}}{j_n} + \mathbf{1}(u \in C_{n,0}^- \cap \Delta_j) \right)^\top$$

where $j_n = |\mathcal{J}_n|$ satisfies $j_n \sim c_n$ as $n \rightarrow \infty$. We have $(Z_{n,0}, Z_{n,0}^-)^\top = \sum_{j \in \mathcal{J}_n} Y_{n,j}$ and we remark that $\sup_{n \geq 1} \sup_{j \in \mathcal{J}_n} \|Y_{n,j}\|_\infty < \infty$. Since $\varepsilon_n \rightarrow 0$ and $c_n^{1/d} \varepsilon_n \rightarrow \infty$ as $n \rightarrow \infty$, we can apply Lemma A.1 (c) to derive

$$\text{Var}(Z_{n,0}) \sim \text{Var}(Z_{n,0}^-) \sim \text{Cov}(Z_{n,0}, Z_{n,0}^-) \sim \frac{1}{12} + \sigma^2 c_n$$

where $\sigma^2 = \lambda + \lambda^2 \int_{\mathbb{R}^d} (g(w) - 1) dw$. In other words,

$$j_n^{-1} \text{Var}((Z_{n,0}, Z_{n,0}^-)^\top) \rightarrow \Sigma = \sigma^2 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad (\text{A.2})$$

which is a matrix with rank 1. By combining this with Assumption (iv), we can apply Theorem A.2 to get as $n \rightarrow \infty$

$$c_n^{-1/2} (Z_{n,0} - \text{E}(Z_{n,0}), Z_{n,0}^- - \text{E}(Z_{n,0}^-))^\top \rightarrow \mathcal{N}(0, \Sigma)$$

in distribution. Since $t_n^- = \text{E}(Z_{n,0}^-)$ and $\text{E}(Z_{n,0}) - t_n = 1/2 - a_n = o(c_n^{1/2})$ by definition of t_n , an application of Slutsky's Lemma yields that

$$c_n^{-1/2} (Z_{n,0} - t_n, Z_{n,0}^- - t_n^-)^\top \rightarrow \mathcal{N}(0, \Sigma)$$

in distribution as $n \rightarrow \infty$ whereby we deduce that

$$\mathbb{P}(Z_{n,0} \leq t_n) \rightarrow 1/2 \quad \text{and} \quad \mathbb{P}(Z_{n,0}^- \leq t_n^-) \rightarrow 1/2. \quad (\text{A.3})$$

Rose and Smith (1996) proved that if $U = (U_1, U_2)^\top$ follows a bivariate normal distribution with mean 0, variance 1 and correlation ρ , $\mathbb{P}(U_1 \leq 0, U_2 \leq 0) = 1/4 + \sin^{-1}(\rho)/2\pi$ which equals to $1/2$ when $\rho = 1$. From (A.2), this shows that $\mathbb{P}(Z_{n,0} \leq t_n, Z_{n,0}^- \leq t_n^-) \rightarrow 1/2$ as $n \rightarrow \infty$. As a consequence, $\text{Var}(D_{n,0}) \rightarrow 0$ which combined with (A.1) leads to $k_n^{-1} \text{Var}(D_n) \rightarrow 0$ as $n \rightarrow \infty$.

Step 2. This step is the core of the blocking technique. We mimic the proof here. We label the cells $C_{n,k}^-$ by $\mathcal{C}_1, \dots, \mathcal{C}_{k_n}$ such that for any $j = 1, \dots, k_n$, the sequence (\mathcal{C}_j) forms an increasing sequence of domains. Let f denote the bijection $f : \mathcal{K}_n \rightarrow \{1, \dots, k_n\}$. Let $V_j = e^{iu(\mathbf{1}(Z_{n,f^{-1}(j)} \leq t_n^-) - \mathbb{P}(Z_{n,f^{-1}(j)} \leq t_n^-)) / \sqrt{k_n}}$. Then

$$\phi_n^-(u) = \mathbb{E} \prod_{j=1}^{k_n} V_j \quad \text{and} \quad \tilde{\phi}_n^-(u) = \prod_{j=1}^{k_n} \mathbb{E}(V_j).$$

and

$$|\phi_n^-(u) - \tilde{\phi}_n^-(u)| \leq \sum_{j=1}^{k_n-1} \left| \mathbb{E} \left(\prod_{s=1}^{j+1} V_s \right) - \mathbb{E} \left(\prod_{s=1}^j V_s \right) \mathbb{E}(V_{j+1}) \right|.$$

Let $A_j = \prod_{s=1}^j V_s$. Clearly, $A_j \in \mathcal{F}(\cup_{s=1}^j C_{n,f^{-1}(s)}^-)$ and $V_{j+1} \in \mathcal{F}(C_{n,f^{-1}(j+1)}^-)$, $|\cup_{s=1}^j C_{n,f^{-1}(s)}^-| = j(1-\varepsilon_n)^d c_n$, $|C_{n,f^{-1}(j+1)}^-| = (1-\varepsilon_n)^d c_n$ and $d(\cup_{s=1}^j C_{n,f^{-1}(s)}^-, C_{n,f^{-1}(j+1)}^-) \geq 2\varepsilon_n c_n^{1/d}$. Since A_j and V_{j+1} are bounded random variables, we have the following upper-bound on their covariance by means of the strong mixing coefficient, see Zhengyan and Chuanrong (1996, Lemma 2.1)

$$\begin{aligned} \text{Cov}(A_j, V_{j+1}) &\leq 4\alpha_{j(1-\varepsilon_n)^d c_n, (1-\varepsilon_n)^d c_n} (2\varepsilon_n c_n^{1/d}) \\ &\leq 4j c_n \sup_p \frac{\alpha_{p,p}(2\varepsilon_n c_n^{1/d})}{p} \\ &\leq 4c_n k_n \mathcal{O}(\varepsilon_n^{-d(1+\eta)} c_n^{-(1+\eta)}) = \mathcal{O}(k_n \varepsilon_n^{-d(1+\eta)} c_n^{-\eta}) \end{aligned}$$

whereby we deduce that $|\phi_n^-(u) - \tilde{\phi}_n^-(u)| = \mathcal{O}(k_n^2 c_n^{-\eta} \varepsilon_n^{-d(1+\eta)})$. Now we can fix the sequence $(\varepsilon_n)_{n \geq 1}$. Specifically, we set $\varepsilon_n = c_n^{(\eta' - \eta)/d(1+\eta)}$ for some $0 < \eta' < \eta$. This choice ensures that $\varepsilon_n \rightarrow 0$, $c_n^{1/d} \varepsilon_n = c_n^{(1+\eta')/d(1+\eta)} \rightarrow \infty$ and yields that $|\phi_n^-(u) - \tilde{\phi}_n^-(u)| = \mathcal{O}(k_n^2/c_n^{\eta'})$ which tends to 0 as $n \rightarrow \infty$ by Assumption (i).

Step 3. Since $\tilde{Z}_{n,k}^- \stackrel{d}{=} Z_{n,k}^-$ and since $\mathbb{P}(Z_{n,k}^- \leq t_n^-) \rightarrow 1/2$ as $n \rightarrow \infty$ from Step 2, we deduce that

$$\begin{aligned} \text{Var}(\tilde{S}_n^-) &= \sum_{k \in \mathcal{K}_n} \mathbb{P}(\tilde{Z}_{n,k}^- \leq t_n^-) (1 - \mathbb{P}(\tilde{Z}_{n,k}^- \leq t_n^-)) \\ &= k_n \mathbb{P}(\tilde{Z}_{n,0}^- \leq t_n^-) (1 - \mathbb{P}(\tilde{Z}_{n,0}^- \leq t_n^-)) \sim k_n/4 \end{aligned}$$

as $n \rightarrow \infty$. Since $(\mathbf{1}(\tilde{Z}_{n,k}^- \leq t_n^-), k \in \mathcal{K}_n)$ is a collection of bounded and independent random variables, Step 3 follows from an application of Lyapounov Theorem. \square

A.3 Proof of Theorem 4.3

Proof. (a) Let us define for any $t \geq 0$

$$A_n = \sqrt{\frac{k_n}{c_n}} \left(\widehat{\text{Me}}(\mathbf{Z}) - \text{Me}_Z \right) \quad \text{and} \quad B_n(t) = \sqrt{\frac{k_n}{c_n}} \left(\frac{F_Z(t) - \widehat{F}(t; \mathbf{Z})}{f_Z(\text{Me}_Z)} \right).$$

We have to prove that $A_n - B_n(\text{Me}_Z)$ converges in probability to 0 as $n \rightarrow \infty$. The proof is based on the application of Ghosh (1971, Lemma 1) which consists in satisfying the two following conditions:

(I) for all $\delta > 0$, there exists $\varepsilon = \varepsilon(\delta)$ such that $\text{P}(|B_n(\text{Me}_Z)| > \varepsilon) < \delta$.

(II) for all $y \in \mathbb{R}$ and $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} \text{P}(A_n \leq y, B_n(\text{Me}_Z) \geq y + \varepsilon) = \lim_{n \rightarrow \infty} \text{P}(A_n \geq y + \varepsilon, B_n(\text{Me}_Z) \leq y) = 0.$$

(I) is in particular fulfilled if we prove that $\text{Var} B_n(\text{Me}_Z) = \mathcal{O}(1)$. The proof of Theorem 4.2 shows in particular that $\text{Var} \widehat{F}(\text{Me}_Z; \mathbf{Z}) = \mathcal{O}(k_n^{-1})$ as $n \rightarrow \infty$. By Assumption (ii), we are led to

$$\text{Var} B_n(\text{Me}_Z) = \frac{1}{c_n f_Z(\text{Me}_Z)^2} \text{Var}(\sqrt{k_n} \widehat{F}(\text{Me}_Z; \mathbf{Z})) = \mathcal{O}(1).$$

(II) Let $y \in \mathbb{R}$ (and without loss of generality assume $y \geq 0$). By definition of the sample median, we have

$$\begin{aligned} \{A_n \leq y\} &= \left\{ \widehat{\text{Me}}(\mathbf{Z}) \leq \text{Me}_Z + y \sqrt{c_n/k_n} \right\} \\ &= \left\{ \frac{1}{2} \leq \widehat{F} \left(\text{Me}_Z + y \sqrt{c_n/k_n} \right) \right\} \\ &= \left\{ B_n \left(\text{Me}_Z + y \sqrt{c_n/k_n} \right) \leq y_n \right\} \end{aligned}$$

where

$$y_n = \sqrt{k_n/c_n} \frac{1}{f_Z(\text{Me}_Z)} \left(F_Z \left(\text{Me}_Z + y \sqrt{c_n/k_n} \right) - F_Z(\text{Me}_Z) \right).$$

We now intend to prove that as $n \rightarrow \infty$, $y_n \rightarrow y$ and $\widetilde{B}_n = B_n(\text{Me}_Z + y \sqrt{c_n/k_n}) - B_n(\text{Me}_Z) \rightarrow 0$ in probability. First, since Z admits a density everywhere, there exists $\tau_n \in (\text{Me}_Z, \text{Me}_Z + y \sqrt{c_n/k_n})$ such that $y_n = y f_Z(\tau_n)/f_Z(\text{Me}_Z)$. From (3.2)

$$\frac{f_Z(\tau_n)}{f_Z(\text{Me}_Z)} = \frac{\text{P}(N = \lfloor \tau_n \rfloor)}{\text{P}(N = \lfloor \text{Me}_Z \rfloor)}$$

which tends to 1 by Assumption (ii-2) and implies the convergence of y_n towards y . Second, we show that $\text{Var}(\widetilde{B}_n) \rightarrow 0$ as $n \rightarrow \infty$ by decomposing the variance as follows. Let $\widetilde{B}_{n,k} = \mathbf{1}(\text{Me}_Z \leq Z_{n,k} \leq \text{Me}_Z + y \sqrt{c_n/k_n}) - \text{P}(\text{Me}_Z \leq Z_{n,k} \leq \text{Me}_Z + y \sqrt{c_n/k_n})$

$$\begin{aligned}
\text{Var}(\tilde{B}_n) &= \frac{1}{c_n f_Z(\text{Me}_Z)^2} \frac{1}{k_n} \sum_{k, k' \in \mathcal{K}_n} \text{Cov}(\tilde{B}_{n,k}, \tilde{B}_{n,k'}) \\
&\leq \frac{\kappa}{k_n} \sum_{\substack{k, k' \in \mathcal{K}_n \\ |k-k'| \leq 1}} |\text{Cov}(\tilde{B}_{n,k}, \tilde{B}_{n,k'})| + \frac{\kappa}{k_n} \sum_{\substack{k, k' \in \mathcal{K}_n \\ |k-k'| > 1}} |\text{Cov}(\tilde{B}_{n,k}, \tilde{B}_{n,k'})|. \quad (\text{A.4})
\end{aligned}$$

We follow here the proof of Step 1 of Theorem 4.2. For any $k, k' \in \mathcal{K}_n$ $k \neq k'$, $\text{Cov}(\tilde{B}_{n,k}, \tilde{B}_{n,k'}) = \mathcal{O}(|k - k' - 1|^{-d(1+\eta)} c_n^{-\eta})$. So

$$\frac{1}{k_n} \sum_{\substack{k, k' \in \mathcal{K}_n \\ |k-k'| > 1}} |\text{Cov}(\tilde{B}_{n,k}, \tilde{B}_{n,k'})| = \mathcal{O}(c_n^{-\eta})$$

which tends to 0 as $n \rightarrow \infty$. The first double sum of (A.4) is upper-bounded by $3^d \kappa \text{Var}(\tilde{B}_{n,0})$ and

$$\text{Var}(\tilde{B}_{n,0}) = \text{P}(\text{Me}_Z \leq Z_{n,0} \leq \text{Me}_Z + y\sqrt{c_n/k_n}) (1 - \text{P}(\text{Me}_Z \leq Z_{n,0} \leq \text{Me}_Z + y\sqrt{c_n/k_n})).$$

Since $\text{Me}_Z = \lambda c_n + o(\sqrt{c_n})$ and $\text{Me}_Z + y\sqrt{c_n/k_n} = \lambda c_n + o(\sqrt{c_n})$ for every $y \in \mathbb{R}$ by Assumption (i)-(ii), then we can apply (A.3) which leads to $\text{P}(Z_{n,0} \geq \text{Me}_Z) \rightarrow 1/2$, $\text{P}(Z_{n,0} \leq \text{Me}_Z + y\sqrt{c_n/k_n}) \rightarrow 1/2$ and finally to $\text{Var}(\tilde{B}_{n,0}) \rightarrow 0$ and $\tilde{B}_n \rightarrow 0$ in probability as $n \rightarrow \infty$.

We can now conclude. For all $\varepsilon > 0$, there exists $n_0(\varepsilon)$ such that for all $n \geq n_0(\varepsilon)$, $y_n \leq y + \varepsilon/2$. Therefore for $n \geq n_0(\varepsilon)$

$$\begin{aligned}
\text{P}(A_n \leq y, B_n(\text{Me}_Z) \geq y + \varepsilon) &= \text{P}(B_n(\text{Me}_Z + y\sqrt{c_n/k_n}) \leq y_n, B_n(\text{Me}_Z) \geq y + \varepsilon) \\
&\leq \text{P}(B_n(\text{Me}_Z + y\sqrt{c_n/k_n}) \leq y + \varepsilon/2, B_n(\text{Me}_Z) \geq y + \varepsilon) \\
&\leq \text{P}\left(\left|B_n(\text{Me}_Z + y\sqrt{c_n/k_n}) - B_n(\text{Me}_Z)\right| \geq \varepsilon/2\right)
\end{aligned}$$

which tends to 0 as $n \rightarrow \infty$ and (II) is proved.

(b) It is sufficient to combine Theorem 4.2 (b) and Theorem 4.3 (a). From Slutsky's Lemma and by Assumptions (ii-2)-(ii-3), the following convergence in distribution holds as $n \rightarrow \infty$

$$\sqrt{k_n/c_n} s_n \left(\widehat{\text{Me}}(\mathbf{Z}) - \text{Me}_Z \right) \rightarrow \mathcal{N}(0, 1/4)$$

where $s_n = \sqrt{c_n} \text{P}(N = \lfloor \lambda c_n \rfloor)$. Since $\widehat{\text{Me}}(\mathbf{Z}) = c_n \widehat{\text{Me}}(\check{\mathbf{Z}})$, $\text{Me}_Z = c_n \text{Me}_{\check{Z}}$ and $|W_n| = k_n c_n$, this can be rewritten as

$$|W_n|^{1/2} s_n \left(\widehat{\text{Me}}(\check{\mathbf{Z}}) - \text{Me}_{\check{Z}} \right) \rightarrow \mathcal{N}(0, 1/4).$$

From (B.1) and by Assumptions (i)-(ii), $\text{Me}_{\check{Z}} = \lambda + \mathcal{O}(c_n^{\ell-1})$ and $\sqrt{k_n c_n} c_n^{\ell-1} \rightarrow 0$ as $n \rightarrow \infty$. Hence, a last application of Slutsky's Lemma concludes the proof. \square

A.4 Auxiliary result

We present here an auxiliary result on the control of the covariance of counting variables. We define \mathcal{C}_τ the cube centered at 0 with volume $\tau^d c_n$, i.e.

$$\mathcal{C}_\tau = \{u = (u_1, \dots, u_d)^\top \in \mathbb{R}^d : |u_l| \leq \tau c_n^{1/d}/2, l = 1, \dots, d\}.$$

Lemma A.1. *Under the Assumption (iii), we have the two following statements.*

(a) For any $\tau \in (0, 1]$

$$\text{Var}(n(\mathbf{X}_{\mathcal{C}_\tau})) \sim |\mathcal{C}_\tau| \left(\lambda + \lambda^2 \int_{\mathbb{R}^d} (g(w) - 1) dw \right)$$

as $n \rightarrow \infty$.

(b) Let $\varepsilon \in (0, 1)$ then

$$\text{Cov}(n(\mathbf{X}_{\mathcal{C}_{1-\varepsilon}}), n(\mathbf{X}_{\mathcal{C}_1})) \sim \lambda |\mathcal{C}_{1-\varepsilon}| + \lambda^2 |\mathcal{C}_{1-\varepsilon/2}| \int_{\mathbb{R}^d} (g(w) - 1) dw$$

as $n \rightarrow \infty$.

(c) Let $(\varepsilon_n)_{n \geq 1}$ be a sequence of real numbers such that $\varepsilon_n \rightarrow 0$ and $c_n^{1/d} \varepsilon_n \rightarrow \infty$ as $n \rightarrow \infty$, then

$$\begin{aligned} \text{Var}(n(\mathbf{X}_{\mathcal{C}_{1-\varepsilon_n}})) &\sim \text{Var}(n(\mathbf{X}_{\mathcal{C}_1})) \sim \text{Cov}(n(\mathbf{X}_{\mathcal{C}_{1-\varepsilon_n}}), n(\mathbf{X}_{\mathcal{C}_1})) \\ &\sim c_n \left(\lambda + \lambda^2 \int_{\mathbb{R}^d} (g(w) - 1) dw \right) \end{aligned}$$

as $n \rightarrow \infty$.

Proof. (a) is a classical result, see e.g. Heinrich and Prokešová (2010). As we need to refer to specific equations, we report the proof here. Using Campbell Theorem and since \mathbf{X} admits a pair correlation function

$$\begin{aligned} \text{Var}(n(\mathbf{X}_{\mathcal{C}_\tau})) &= \lambda |\mathcal{C}_\tau| + \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbf{1}(u \in \mathcal{C}_\tau) \mathbf{1}(v \in \mathcal{C}_\tau) (g(u-v) - 1) dudv \\ &= \lambda |\mathcal{C}_\tau| + \lambda^2 \int_{\mathbb{R}^d} |\mathcal{C}_\tau \cap (\mathcal{C}_\tau)_{-w}| (g(w) - 1) dw \\ &= \lambda |\mathcal{C}_\tau| + \lambda^2 \int_{\mathcal{C}_{2\tau}} |\mathcal{C}_\tau \cap (\mathcal{C}_\tau)_{-w}| (g(w) - 1) dw \end{aligned} \tag{A.5}$$

$$= \lambda |\mathcal{C}_\tau| + \lambda^2 \int_{\mathcal{C}_{2\tau}} \prod_{l=1}^d (\tau c_n^{1/d} - |w_l|) (g((w_1, \dots, w_d)^\top) - 1) dw_1 \dots dw_d \tag{A.6}$$

$$\sim |\mathcal{C}_\tau| \left(\lambda + \lambda^2 \int_{\mathbb{R}^d} (g(w) - 1) dw \right)$$

by Assumption (iii).

(b) For brevity, let K_ε denote the covariance to evaluate. Following (a) we have

$$\begin{aligned} K_\varepsilon &= \lambda |\mathcal{C}_{1-\varepsilon} \cap \mathcal{C}_1| + \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbf{1}(u \in \mathcal{C}_{1-\varepsilon}) \mathbf{1}(v \in \mathcal{C}_1) (g(u-v) - 1) dudv \\ &= \lambda |\mathcal{C}_{1-\varepsilon}| + \lambda^2 \int_{\mathbb{R}^d} |\mathcal{C}_{1-\varepsilon} \cap (\mathcal{C}_1)_{-w}| (g(w) - 1) dw. \end{aligned}$$

Let $w = (w_1, \dots, w_d)^\top$. We can check that

$$|\mathcal{C}_{1-\varepsilon} \cap (\mathcal{C}_1)_{-w}| = \begin{cases} 0 & \text{if } w \in \mathbb{R}^d \setminus \mathcal{C}_{2-\varepsilon} \\ \prod_{l=1}^d \left((1 - \frac{\varepsilon}{2}) c_n^{1/d} - |w_l| \right) & \text{if } w \in \mathcal{C}_{2-\varepsilon} \end{cases}$$

whereby we deduce using (A.5)-(A.6) and Assumption (iii) that

$$\begin{aligned} K_\varepsilon &= \lambda |\mathcal{C}_{1-\varepsilon}| + \lambda^2 \int_{\mathcal{C}_{2-\varepsilon}} \prod_{l=1}^d \left((1 - \varepsilon/2) c_n^{1/d} - |w_l| \right) \left(g((w_1, \dots, w_d)^\top) - 1 \right) dw_1 \dots dw_d \\ &= \lambda |\mathcal{C}_{1-\varepsilon}| + \lambda^2 \int_{\mathcal{C}_{2-\varepsilon}} |\mathcal{C}_{1-\varepsilon/2} \cap (\mathcal{C}_{1-\varepsilon/2})_{-w}| (g(w) - 1) dw \\ &\sim \lambda |\mathcal{C}_{1-\varepsilon}| + \lambda^2 |\mathcal{C}_{1-\varepsilon/2}| \int_{\mathbb{R}^d} (g(w) - 1) dw \end{aligned}$$

as $n \rightarrow \infty$.

(c) The assumptions on the sequence (ε_n) allow us to apply (a)-(b) which leads to the result since $|\mathcal{C}_1| \sim |\mathcal{C}_{1-\varepsilon_n}| \sim |\mathcal{C}_{1-\varepsilon_n/2}| \sim c_n$ as $n \rightarrow \infty$. \square

A.5 A central limit theorem

We present here a central limit theorem for stationary random fields with asymptotic covariance matrix not necessarily positive definite. It is very close to Guyon (1991, Theorem 3.3.1) and to Karáczony (2006, Theorem 1) but we were not able to find it in the following form in the literature.

Theorem A.2. *Let $(X_k, k \in \mathbb{Z}^d)$ be a stationary random field in a measurable space S . Let $\mathcal{K}_n \subset \mathbb{Z}^d$ with $k_n = |\mathcal{K}_n| \rightarrow \infty$ as $n \rightarrow \infty$. For any $n \geq 1$ and $k \in \mathcal{K}_n$, we define $Y_{n,k} = f_{n,k}(X_k)$ where $f_{n,k} : S \rightarrow \mathbb{R}^p$ for some $p \geq 1$ is a measurable function. We denote by $S_n = \sum_{k \in \mathcal{K}_n} Y_{n,k}$ and by $\Sigma_n = \text{Var}(S_n)$ and assume that for any $n \geq 1$, $k \in \mathcal{K}_n$, $\mathbb{E}Y_{n,k} = 0$. We also assume that*

- (I) $\sup_{n \geq 1} \sup_{k \in \mathcal{K}_n} \|Y_{n,k}\|_\infty < \infty$.
- (II) There exists $\eta > 0$ such that $\alpha_{2,\infty}(m) = \mathcal{O}(m^{-d(1+\eta)})$.
- (III) There exists $\Sigma \geq 0$ a (p, p) matrix with rank $1 \leq r \leq p$ such that $k_n^{-1} \Sigma_n \rightarrow \Sigma$ as $n \rightarrow \infty$.

Then, $k_n^{-1/2} S_n \rightarrow \mathcal{N}(0, \Sigma)$ in distribution as $n \rightarrow \infty$.

We present Theorem A.2 for bounded random vectors and with only one mixing coefficient, namely $\alpha_{2,\infty}$. It can obviously be generalized along similar lines as in Guyon (1991, Theorem 3.3.1).

Proof. Assume $\Sigma > 0$, then for n large enough $k_n^{-1} \Sigma_n \geq \Sigma/2 > 0$, which combined with Assumptions (I)-(II) allows us to apply Karáczony (2006, Theorem 1) to conclude the result.

The end of the proof follows the same arguments as the proof of a central limit theorem for triangular arrays of conditionally centered random fields obtained by Coeurjolly and Lavancier (2013, Theorem 2). If Σ is not positive definite, we can find an orthonormal basis (h_1, \dots, h_p) of \mathbb{R}^p where the f_i 's are eigenvectors of Σ .

We let (f_1, \dots, f_r) be the basis of the image of Σ and (f_{r+1}, \dots, f_p) be the basis of its kernel. Let also H_{Im} (resp. H_{Ker}) be the matrix formed by the column vectors of (f_1, \dots, f_r) (resp. (f_{r+1}, \dots, f_p)). Similarly for $v \in \mathbb{R}^p$, we denote by v_j its j th coordinate in the basis of (f_1, \dots, f_p) , $v_{Im} = (v_1, \dots, v_r)$ and $v_{Ker} = (v_{r+1}, \dots, v_p)$. Using the Cramer-Wold device, we need to prove that for any $v \in \mathbb{R}^p$, $v^\top k_n^{-1/2} S_n$ converges towards a Gaussian random variable. We have

$$v^\top k_n^{-1/2} S_n = v_{Im}^\top H_{Im}^\top k_n^{-1/2} S_n + v_{Ker}^\top H_{Ker}^\top k_n^{-1/2} S_n.$$

Let $S'_n = \sum_k Y'_{n,k}$ where $Y'_{n,k} = H_{Im}^\top Y_{n,k}$. The random variables $Y'_{n,k}$ are bounded variables for any $n \geq 1$ and $k \in \mathcal{K}_n$. By assumption (III), $k_n^{-1} \text{Var}(S'_n) \rightarrow H_{Im}^\top \Sigma H_{Im}$ which is a positive definite matrix since $r \geq 1$. Therefore from the first part of the proof, $v_{Im}^\top H_{Im}^\top k_n^{-1/2} S_n$ tends to a Gaussian random variable in distribution as $n \rightarrow \infty$. By Slutsky's Lemma, the proof will be done if $v_{Ker}^\top H_{Ker}^\top k_n^{-1/2} S_n$ tends to 0 in probability as $n \rightarrow \infty$. Since, $H_{Ker}^\top \Sigma H_{Ker} = 0$, the expected convergence follows from

$$\begin{aligned} \text{Var}(v_{Ker}^\top H_{Ker}^\top k_n^{-1/2} S_n) &= v_{Ker}^\top H_{Ker}^\top k_n^{-1} \Sigma_n H_{Ker} v_{Ker} \\ &= v_{Ker}^\top H_{Ker}^\top (k_n^{-1} \Sigma_n - \Sigma) H_{Ker} v_{Ker} \\ &\leq \|v_{Ker}\| \|H_{Ker}\| \|k_n^{-1} \Sigma_n - \Sigma\| \end{aligned}$$

which tends to 0 by Assumption (III). \square

B Additional comments and figures

This section contains additional comments and several figures related to Section 5.

B.1 The way of jittering a sample of counts

We could think about generalizing (3.1) slightly and introduce a function of a uniform random variable, i.e. define for any $k \in \mathcal{K}_n$

$$Z_{n,k} = n(\mathbf{X}_{C_{n,k}}) + \varphi^{-1}(U_k) = N_{n,k} + \varphi^{-1}(U_k)$$

where $\varphi : [0, 1] \rightarrow [0, 1]$ is a continuously differentiable increasing function. The cdf of Z would be in that case

$$F_Z(t) = F_N(\lfloor t \rfloor - 1) + P(N = \lfloor t \rfloor) \varphi(t - \lfloor t \rfloor)$$

and for any $t \notin \mathbb{N}$, Z would admit a density f_Z at t given by

$$f_Z(t) = P(N = \lfloor t \rfloor) \varphi'(t - \lfloor t \rfloor).$$

When $t \in \mathbb{N}$, since $(F_Z(t+h) - F_Z(t))/h$ tends to $P(N = \lfloor t \rfloor) \varphi'(0)$ when $h \rightarrow 0^+$ and to $P(N = \lfloor t \rfloor) \varphi'(1)$ when $h \rightarrow 0^-$, Z would also admit a density at t if we add the condition $\varphi'(0) = \varphi'(1)$. Our Theorem 4.3 requires however another assumption. Namely, we need to assume that for any $t_n = \lambda c_n + \mathcal{O}(\sqrt{c_n/k_n})$, $f_Z(t_n)/f_Z(\lambda c_n)$ tends to 1. To this end, we would have to combine Assumption (ii-2), with an assumption like $\inf_t \varphi'(t) = \sup_t \varphi'(t)$. This explains why we focused on the case $\varphi(t) = t$ in Section 3 and in the presentation of our asymptotic results in Section 4.

B.2 Rule of thumb under the Poisson case

In this section, we want to examine the value of the true median of Z under the Poisson case. Even if this is useless we also had a look at different functions φ . Figure 1 presents the true median of N and $Z = N + \varphi^{-1}(U)$ where N follows a Poisson distribution with parameter ν and where U is a uniform random variable on $[0, 1]$. We considered the cases $\varphi(t) = \sqrt{t}, t, t^2$ and examined the true median minus ν in terms of ν . First, we recover a result obtained by Adell and Jodrá (2005): when ν is an integer, the median of N equals ν and for other values of ν , it lies in the interval $[\nu - \log(2), \nu + 1/3]$. It is worth interesting to observe that for the three functions investigated the choice $\varphi(t) = t$ leads us to conjecture that when ν is large $\text{Me}_{\mathcal{P}(\nu)}$ is very close to $\nu + 1/3$.

So, we could use the rule of thumb derived under the Poisson case and modify the jittered estimator (3.3) as follows

$$\widehat{\lambda}^{J,2} = \widehat{\lambda}^J - \frac{1}{3c_n} = \text{Me}_{\mathbf{Z}} - \frac{1}{3c_n}. \quad (\text{B.1})$$

Since $|W_n|^{1/2}/c_n = \sqrt{k_n/c_n} \rightarrow 0$ by Assumption (i), this produces no differences asymptotically: $\widehat{\lambda}^{J,2}$ has the same behaviour as $\widehat{\lambda}^J$ and satisfies the central limit theorem given by (4.3) or Corollary 4.4. We compared $\widehat{\lambda}^J$ and $\widehat{\lambda}^{J,2}$ in the framework of the simulation study presented in Section 5. The evidence of better empirical results was unclear which explains why we did not present $\widehat{\lambda}^{J,2}$ before and kept $\widehat{\lambda}^J$ in the simulation study.

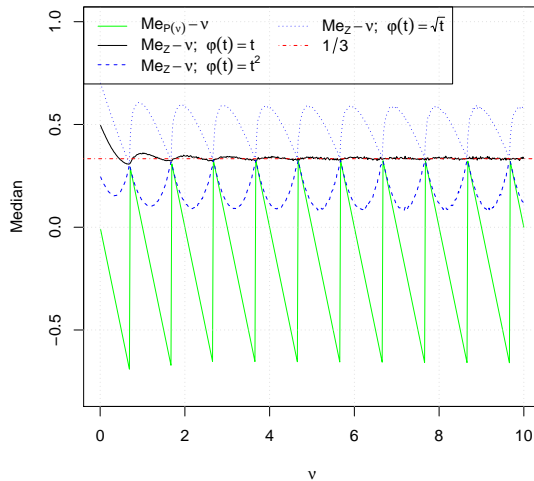


Figure 1: Plot of sample medians in terms of ν of 10^6 replications of random variables following a $\mathcal{P}(\nu)$ and $\mathcal{P}(\nu) + \varphi^{-1}(U)$ for $\varphi(t) = t, t^2$ and \sqrt{t} .

B.3 Examples of pattern produced in the simulation study

Figure 2 illustrates spatial point processes described in Section 5.

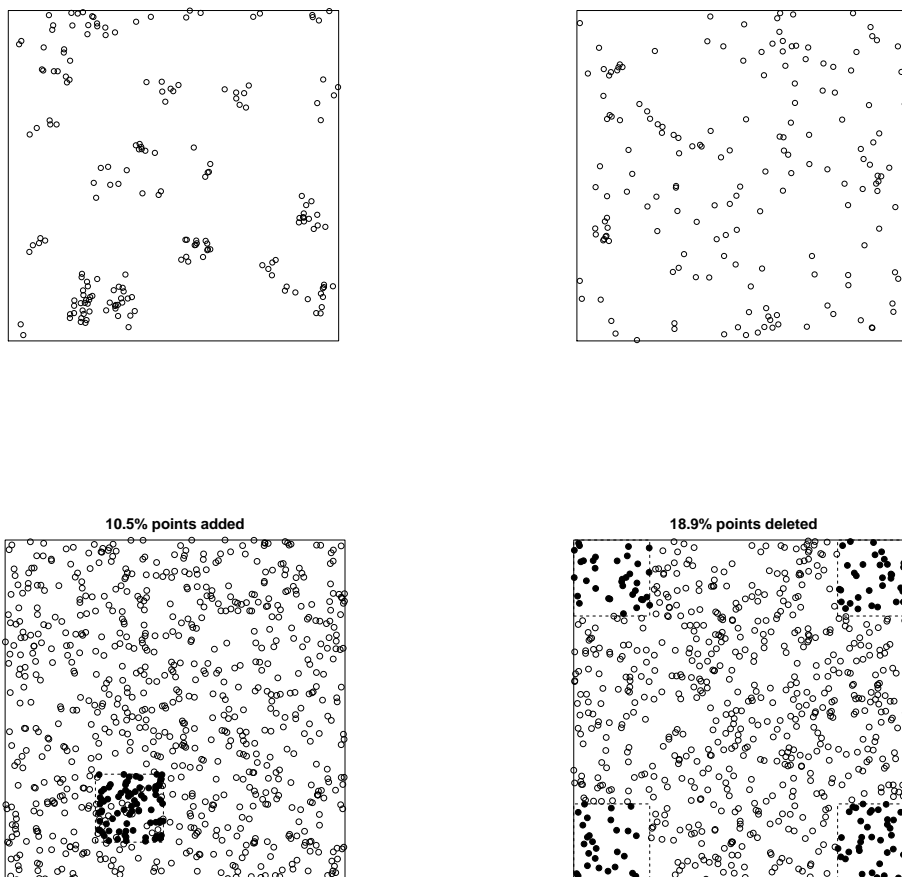


Figure 2: Simulation of point processes models with intensity $\lambda = 50$ generated in $[-1, 1]^2$ for the first row and in $[-2, 2]^2$ for the second row. Top Left: THOMAS process with parameter $\kappa = 10$ and $\sigma = 0.05$; Top Right: LGCP with exponential covariance function with variance $\sigma^2 = 1$ and scale parameter $\phi = 0.05$; Bottom Left: illustrates the contamination (B). To a realization of a POISSON point process (empty circles) is added a configuration of (here) 10.5% (filled circles) in a small square domain included in the initial domain; Bottom Right: illustrates the contamination (C). The resulting point pattern corresponds to the configuration of empty circles which is obtained by thinning the initial pattern (here a realization of a POISSON point process). 18.9% of points are deleted in this example.

B.4 Boxplots of standard and median-based estimates of λ

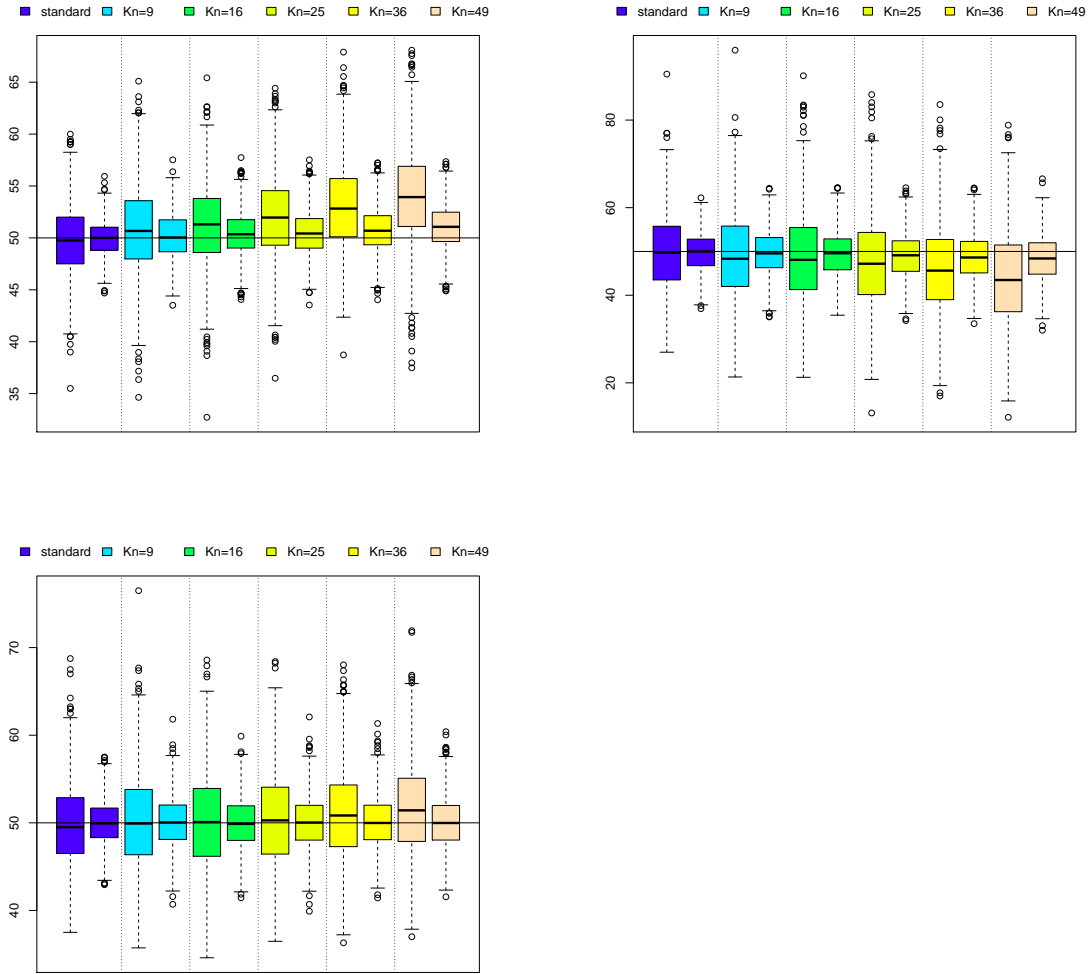


Figure 3: Boxplots of 1000 estimates of the intensity $\lambda = 50$ from different models of spatial point processes generated in $[-n, n]^2$. The point patterns are pure simulations (case (A) described in Section 5) of POISSON (top left), THOMAS (top right) and LGCP (bottom left). In each cell of two boxplots the left one (resp. the right one) corresponds to the case $n = 1$ (resp. $n = 2$).

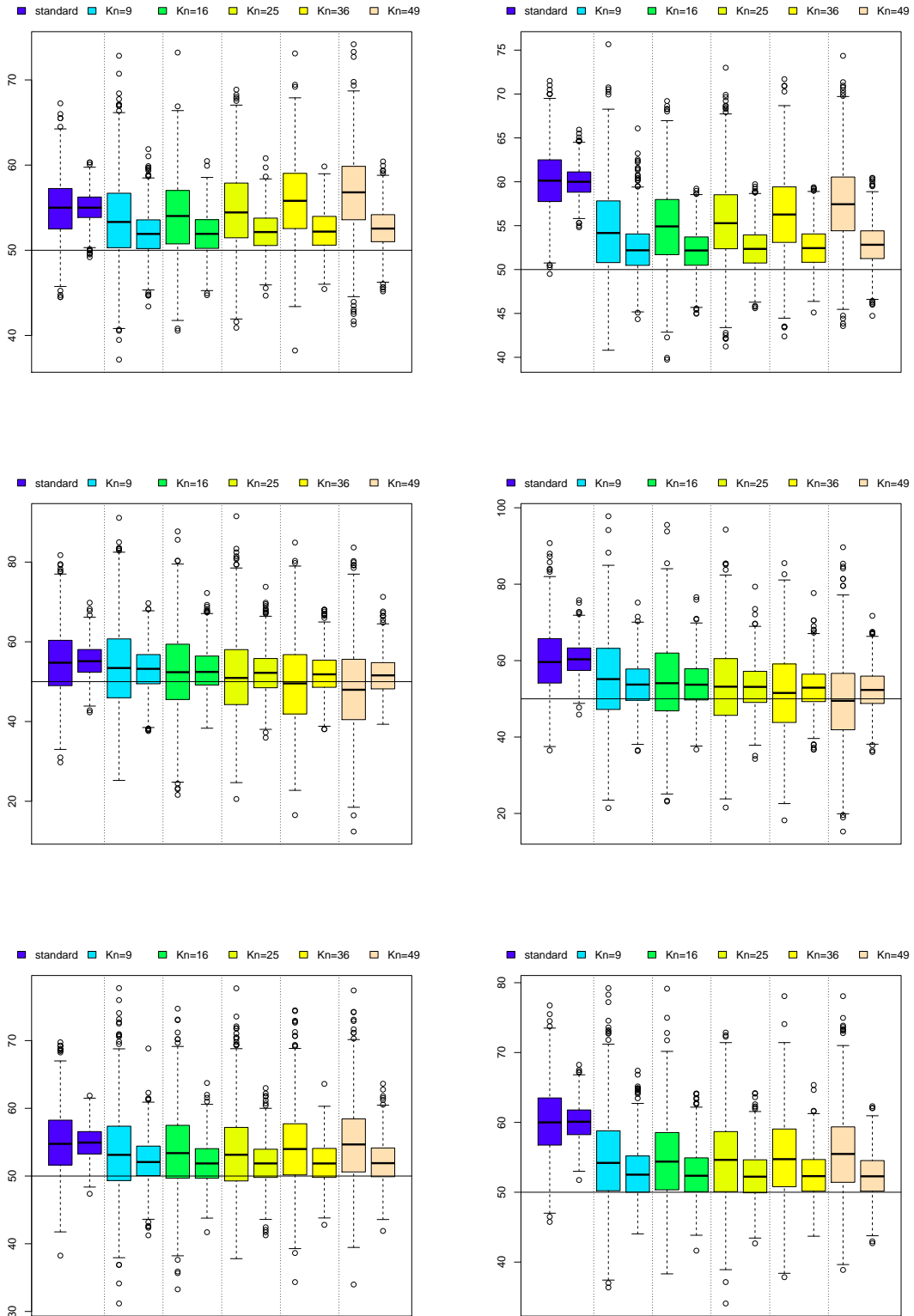


Figure 4: Boxplots of 1000 estimates of the intensity $\lambda = 50$ from different models of spatial point processes generated in $[-n, n]^2$. The point patterns are contaminated simulations of POISSON (first row), THOMAS (second row) and LGCP (third row) point pattern. The contamination corresponds to the case (B): a proportion of $\rho = 10\%$ (first column) or $\rho = 20\%$ of points is added to the initial configuration. In each cell of two boxplots the left one (resp. the right one) corresponds to the case $n = 1$ (resp. $n = 2$).

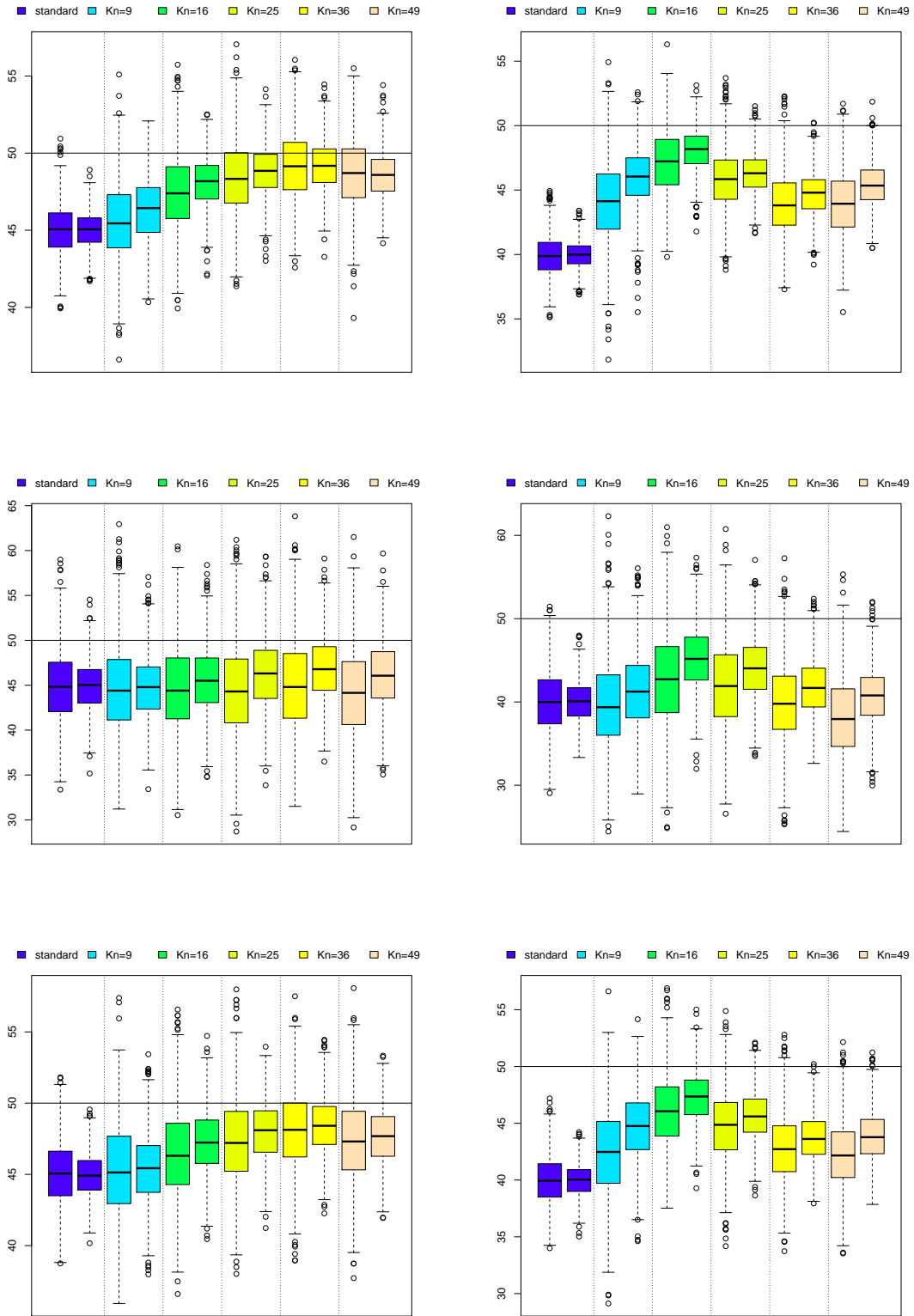


Figure 5: Boxplots of 1000 estimates of the intensity $\lambda = 50$ from different models of spatial point processes generated in $[-n, n]^2$. The point patterns are contaminated simulations of POISSON (first row), THOMAS (second row) and LGCP (third row) point pattern. The contamination corresponds to the case (C): a proportion of $\rho = 10\%$ (first column) or $\rho = 20\%$ of points from the initial configuration is deleted. In each cell of two boxplots the left one (resp. the right one) corresponds to the case $n = 2$ (resp. $n = 3$).

References

- J.A. Adell and P. Jodrá. The median of the Poisson distribution. *Metrika*, 61(3):337–346, 2005.
- R. Assunção and Guttorp. Robustness for inhomogeneous Poisson point processes. *Annals of the Institute of Statistical Mathematics*, 51:657–678, 1999.
- A. Baddeley and R. Turner. Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12:1–42, 2005.
- A. Baddeley, R. Turner, J. Møller, and M. Hazelton. Residual analysis for spatial point processes (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(5):617–666, 2005.
- M. Clausel, J.-F. Coeurjolly, and J. Lelong. Stein estimation of the intensity of a spatial homogeneous Poisson point process. *arXiv preprint arXiv:1407.4372*, 2014.
- J.-F. Coeurjolly and F. Lavancier. Residuals and goodness-of-fit tests for stationary marked Gibbs point processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(2):247–276, 2013.
- J.-F. Coeurjolly and J. Møller. Variational approach to estimate the intensity of spatial point processes. *Bernoulli*, 20(3):1097–1125, 2014.
- D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods*. Springer-Verlag, New York, second edition, 2003.
- H.A. David and H.N. Nagaraja. *Order statistics (3rd ed.)*. Hoboken: Wiley, 2003.
- P. J. Diggle. *Statistical Analysis of Spatial Point Patterns*. Arnold, London, second edition, 2003.
- J.K. Ghosh. A new proof of the Bahadur representation of quantiles and an application. *Annals of Mathematical Statistics*, 42(6):1957–1961, 1971.
- Y. Guan and J. M. Loh. A thinned block bootstrap procedure for modeling inhomogeneous spatial point patterns. *Journal of the American Statistical Association*, 102:1377–1386, 2007.
- Y. Guan, M. Sherman, and J. A. Calvin. On asymptotic properties of the mark variogram estimator of a marked point process. *Journal of statistical planning and inference*, 137(1):148–161, 2007.
- X. Guyon. *Random Fields on a Network*. Springer-Verlag, New York, 1991.
- L. Heinrich and M. Prokešová. On estimating the asymptotic variance of stationary point processes. *Methodology and Computing in Applied Probability*, 12(3):451–471, 2010.
- I. A. Ibragimov and Y. V. Linnik. *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff, Groningen, 1971.
- J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan. *Statistical Analysis and Modelling of Spatial Point Patterns*. Statistics in Practice. Wiley, Chichester, 2008.

- Z. Karáczony. A central limit theorem for mixing random fields. *Miskolc Mathematical Notes*, 7:147–160, 2006.
- Y. Ma, M. Genton, and E. Parzen. Asymptotic properties of sample quantiles of discrete distributions. *Annals of the Institute of Statistical Mathematics*, 63(2):227–243, 2011.
- J.A.F. Machado and J.M.C. Santos Silva. Quantiles for counts. *Journal of the American Statistical Association*, 100(472):1226–1237, 2005.
- J. Møller and R. P. Waagepetersen. *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, Boca Raton, 2004.
- T. Mrkvička and I. Molchanov. Optimisation of linear unbiased intensity estimators for point processes. *Annals of the Institute of Statistical Mathematics*, 57(1):71–81, 2005.
- D.N. Politis, E. Paparoditis, and J.P. Romano. Large sample inference for irregularly spaced dependent observations based on subsampling. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 274–292, 1998.
- M. Prokešová and E.B.V. Jensen. Asymptotic Palm likelihood theory for stationary point processes. *Annals of the Institute of Statistical Mathematics*, 65(2):387–412, 2013.
- C. Rose and M.D. Smith. The multivariate normal distribution. *Mathematica Journal*, 6(1):32–37, 1996.
- WL Stevens. Fiducial limits of the parameter of a discontinuous distribution. *Biometrika*, 37(1-2):117–129, 1950.
- D. Stoyan, W. S. Kendall, and J. Mecke. *Stochastic Geometry and Its Applications*. Wiley, Chichester, second edition, 1995.
- A. Särkkä. A note on robust intensity estimation for point processes. *Biometrical Journal*, 34(6):757–767, 1992.
- R. Waagepetersen and Y. Guan. Two-step estimation for inhomogeneous spatial point processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71:685–702, 2009.
- L. Zhengyan and L. Chuanrong. *Limit theory for mixing dependent random variables*, volume 378. Springer, 1996.