THIELE CENTRE
FOR APPLIED MATHEMATICS IN NATURAL SCIENCE

On estimating the asymptotic variance of stationary point processes

Lothar Heinrich and Michaela Prokešová

Research Report No. 17 | October 2006

On estimating the asymptotic variance of stationary point processes

This Thiele Research Report is also Research Report number 486 in the Stochastics Series at Department of Mathematical Sciences, University of Aarhus, Denmark.

On estimating the asymptotic variance of stationary point processes

Lothar Heinrich*, Michaela Prokešová**

Abstract

We investigate a class of kernel estimators $\widehat{\sigma}_n^2$ of the asymptotic variance σ^2 of a d–dimensional stationary point process $\Psi = \sum_{i\geq 1} \delta_{X_i}$ which can be observed in a cubic sampling window $W_n = [-n,n]^d$. σ^2 is defined by the asymptotic relation $\mathsf{Var}(\Psi(W_n)) \sim \sigma^2\,(2n)^d$ (as $n \to \infty$) and its existence is guaranteed whenever the corresponding reduced covariance measure $\gamma_{\rm red}^{(2)}(\cdot)$ has finite total variation. Depending on the rate of decay (polynomially or exponentially) of the total variation of $\gamma_{\rm red}^{(2)}(\cdot)$ outside of an expanding ball centered at the origin, we determine optimal bandwidths b_n (up to a constant) minimizing the mean squared error of $\widehat{\sigma}_n^2$. The case when $\gamma_{\rm red}^{(2)}(\cdot)$ has bounded support is of particular interest. Further we suggest an isotropised estimator $\widetilde{\sigma}_n^2$ suitable for motion-invariant point processes and compare its properties with $\widehat{\sigma}^2$. Our theoretical results are illustrated and supported by a simulation study which compares the (relative) mean squared errors of $\widehat{\sigma}^2$ for planar Poisson, Poisson cluster, and hard–core point processes and for various values of $n\,b_n$.

Keywords: reduced covariance measure, factorial moment and cumulant measures, Poisson cluster process, hard–core process, kernel–type estimator, mean squared error, optimal bandwidth, pair correlation function, central limit theorem, Brillinger–mixing

AMS 2000 Subject Classification: Primary 60G55, 62F12; Secondary 62G05, 62G20

1. Introduction

In various fields of application statisticians are faced with irregular but in some sense homogeneous patterns consisting of randomly distributed points or at least point-like objects which can be observed in a more or less large planar or spatial

*Institute of Mathematics, University of Augsburg, D-86135 Augsburg, Germany

**The T.N. Thiele Centre, Department of Mathematical Sciences, University of Aarhus, Ny Munkegade, 8000 Aarhus C, Denmark

sampling window. Stationary point processes provide appropriate models to describe such phenomena. For a rigorous and detailed introduction in this field we refer the reader to the monograph [2] supplemented by the monograph [15] in which special emphasis is put on statistical analysis of point processes and their application in stochastic geometry. Throughout this paper, let $\Psi = \sum_{i\geq 1} \delta_{X_i}$ denote a simple stationary second–order point process on the d–dimensional Euclidean space $\mathbb{R}^d$ (equipped with the Euclidean norm $\|\cdot\|$ and the corresponding Borel σ–field $\mathcal{B}^d$). Mathematically spoken, Ψ is a locally finite random counting measure with the discrete random closed set of atoms $\{X_1, X_2, \dots\}$ defined on some common probability space $[\Omega, \mathcal{A}, \mathsf{P}]$. We will speak of "points of Ψ" instead of "atoms of Ψ" and write "$x \in \Psi$" instead of "$\Psi(\{x\}) > 0$". The mean number of points of Ψ per unit volume $\lambda = \mathsf{E}\,\Psi([0,1)^d)$ is called the *intensity* or *point density* of Ψ. This simplest numerical characteristic associated with Ψ is standardly estimated by $\widehat{\lambda}_n = \Psi(W_n)/|W_n|$, where $W_n \subset \mathbb{R}^d$ denotes a bounded (convex) sampling window with volume $|W_n|$ which is assumed to expand unboundedly in all directions as $n \to \infty$. Under mild mixing conditions (expressible by the reduced covariance measure of Ψ, see Sect. 2) the limiting variance of $\widehat{\lambda}_n$ exists:

$$\sigma^2 := \lim_{n\to\infty} |W_n| \;\mathsf{E}(\widehat{\lambda}_n - \lambda)^2 = \lim_{n\to\infty} \frac{\mathsf{Var}(\Psi(W_n))}{|W_n|}\,. \tag{1}$$

The limit (1) is briefly called *asymptotic variance of* Ψ. Under somewhat stronger mixing assumptions one can show that $\sqrt{|W_n|}\,(\widehat{\lambda}_n - \lambda)$ converges in distribution to a Gaussian random variable $\mathcal{N}(0,\sigma^2)$ with mean zero and variance σ^2 (if $\sigma^2 > 0$), see e.g. [4], [3], [7]. This result suggests an asymptotic significance test to check the hypothetical intensity λ provided that a (weakly) consistent estimator $\widehat{\sigma}_n^2$ for σ^2 is available. In a recent paper [6], such estimators are also needed for testing non–parametric point process hypothesis by using scaled empirical K–functions. There are other fields of spatial statistics in which asymptotic variances and their estimation play an important role, see [1], [9]. The main aim of this paper is a quantitative asymptotic analysis of a class of estimators for σ^2 introduced in Sect. 2. The main results are formulated in Sect. 3, the proofs of which are given in Sect. 4. In Sect. 5, we study a modified estimator for motion–invariant point processes and describe its asymptotic properties. In Sect. 6 we briefly mention two alternative methods to estimate σ^2. We close this paper with a simulation study to compare different estimators of σ^2 for moderate–size windows W_n.

2. Estimating the asymptotic variance

First we recall the definitions and relations between factorial moment and factorial cumulant measures, see [2] for details. The kth-order factorial moment measure $\alpha^{(k)}$ of Ψ is a locally finite measure on $[(\mathbb{R}^d)^k, \mathcal{B}^{dk}]$ defined by

$$\int_{(\mathbb{R}^d)^k} f(x_1,\dots,x_k)\,\alpha^{(k)}(\mathrm{d}(x_1,\dots,x_k)) = \mathsf{E}\Big(\sum_{x_1,\dots,x_k\in\Psi}^{\neq} f(x_1,\dots,x_k)\Big) \tag{2}$$

for any non–negative, Borel measurable function f on $(\mathbb{R}^d)^k$, where the sum $\sum\limits^{\neq}$ runs over k–tuples of distinct points of Ψ . The kth-order factorial cumulant measure $\gamma^{(k)}$ of Ψ is a locally finite signed measure on $[(\mathbb{R}^d)^k, \mathcal{B}^{dk}]$ which is formally connected with the measures $\alpha^{(1)}, \ldots, \alpha^{(k)}$ by

$$\gamma^{(k)}\big(\mathop{\times}_{i=1}^{k} A_i\big) = \sum_{j=1}^{k} (-1)^{j-1}(j-1)! \sum_{K_1\cup\cdots\cup K_j=\{1,\ldots,k\}} \prod_{i=1}^{j} \alpha^{(\#K_i)}\big(\mathop{\times}_{k_i\in K_i} A_{k_i}\big)$$

for bounded $A_1, \ldots, A_k \in \mathcal{B}^d$, where the inner sum is taken over all partitions of the set $\{1, \ldots, k\}$ in disjoint non–empty subsets $K_1, \ldots, K_j$. In particular, we have $\alpha^{(1)}(A) = \gamma^{(1)}(A) = \lambda\,|A|$ for $A \in \mathcal{B}^d$ and

$$\gamma^{(2)}(A_1 \times A_2) = \alpha^{(2)}(A_1 \times A_2) - \lambda^2\,|A_1|\,|A_2| \quad \text{for} \quad A_1, A_2 \in \mathcal{B}^d\,.$$

Since, for any $k \geq 2$, $\alpha^{(k)}$ is invariant under diagonal shifts there exists a corresponding reduced kth-order factorial moment measure $\alpha^{(k)}_{\text{red}}$ on $[(\mathbb{R}^d)^{k-1}, \mathcal{B}^{d(k-1)}]$ which is uniquely determined by the disintegration formula

$$\begin{aligned}\int_{(\mathbb{R}^d)^k} & f(x_1, \ldots, x_k)\alpha^{(k)}(\,\mathrm{d}(x_1, \ldots, x_k)) \\ &= \lambda \int_{\mathbb{R}^d} \int_{(\mathbb{R}^d)^{k-1}} f(x_1, x_2 + x_1, \ldots, x_k + x_1)\, \alpha^{(k)}_{\text{red}}(\,\mathrm{d}(x_2, \ldots, x_k))\,\mathrm{d}x_1 \qquad (3)\end{aligned}$$

where f is as in (2). In the same way we may define the reduced kth-order factorial cumulant measure $\gamma^{(k)}_{\text{red}}$ which turns out to be a signed measure on $[(\mathbb{R}^d)^{k-1}, \mathcal{B}^{d(k-1)}]$ with the Jordan decomposition $\gamma^{(k)}_{\text{red}} = (\gamma^{(k)}_{\text{red}})^+ - (\gamma^{(k)}_{\text{red}})^-$, see e.g. [16] for details. The corresponding total variation measure $|\gamma^{(k)}_{\text{red}}| = (\gamma^{(k)}_{\text{red}})^+ + (\gamma^{(k)}_{\text{red}})^-$ on $[(\mathbb{R}^d)^{k-1}, \mathcal{B}^{d(k-1)}]$ is locally finite, but in general not finite.

In the special case $k = 2$ we get $\gamma^{(2)}_{\text{red}}(\cdot) = \alpha^{(2)}_{\text{red}}(\cdot) - \lambda\,|\cdot|$ and call $\gamma^{(2)}_{\text{red}}$ the *reduced covariance measure* (briefly: r.c.m.) of Ψ . The variance $\mathsf{Var}(\Psi(W_n))$ can be expressed by means of this r.c.m. which together with (1) leads to

$$\sigma^2 = \lambda + \lambda \lim_{n\to\infty} \int_{\mathbb{R}^d} \frac{|W_n \cap (W_n - x)|}{|W_n|}\, \gamma^{(2)}_{\text{red}}(\,\mathrm{d}x) = \lambda\,\Big(\,1 + \gamma^{(2)}_{\text{red}}(\mathbb{R}^d)\,\Big)\,,$$

whenever W_n increases unboundedly in all directions and $|\gamma^{(2)}_{\text{red}}|(\mathbb{R}^d) < \infty$. Note that the latter condition is sufficient but in some exceptional cases not necessary to ensure the existence of the limit.

The Lebesgue density $\varrho^{(2)}$ of $\alpha^{(2)}_{\text{red}}$ (if it exists) is called the *second–order product density* of Ψ. Further, if Ψ is motion–invariant then $\varrho^{(2)}(x)$ depends only on $\|x\|$ and the function $g(r) := \varrho^{(2)}(x)/\lambda$ for $r = \|x\|$ is called the *pair–correlation function* of Ψ . In this case

$$\gamma^{(2)}_{\text{red}}(\mathbb{R}^d) = \int_{\mathbb{R}^d} (\,\varrho^{(2)}(x) - \lambda\,)\,\mathrm{d}x = \lambda\, d\,\kappa_d \int_0^\infty (\,g(r) - 1\,)r^{d-1}\,\mathrm{d}r \qquad (4)$$

provided the integrals exist, where κ_d denotes the volume of the unit ball in $\mathbb{R}^d$.

To study the asymptotic behaviour of estimates of σ^2 we need some regularity assumptions:

(A0) The sampling windows are cubes $W_n = n\,W$ for $n \geq 1$, where $W = [-1,1]^d$.

(A1) The *kernel function* $w : W \to [0,\infty)$ is symmetric, bounded, and continuous at the origin $o \in \mathbb{R}^d$ such that $w(o) = 1$.

(A2) The (positive) sequence of *bandwidths* (b_n) satisfies $1 \geq b_n \xrightarrow[n\to\infty]{} 0$ and $b_n n \xrightarrow[n\to\infty]{} \infty$.

(A3) The r.c.m. of Ψ has finite total variation, i.e., $\|\gamma^{(2)}_{\mathrm{red}}\|_{\mathrm{var}} := |\gamma^{(2)}_{\mathrm{red}}|(\mathbb{R}^d) < \infty$.

(A4) The third– and fourth–order reduced factorial cumulant measures of Ψ have finite total variation, i.e., $\|\gamma^{(k)}_{\mathrm{red}}\|_{\mathrm{var}} := |\gamma^{(k)}_{\mathrm{red}}|((\mathbb{R}^d)^{k-1}) < \infty$ for $k = 3, 4$.

Now, we are in a position to define the kernel estimators $\widehat{\sigma}^2_n$ of σ^2 by

$$\widehat{\sigma}^2_n = \widehat{\lambda}_n + \sum_{x,y\in\Psi}^{\neq} \frac{w((y-x)/b_n\,n)\mathbf{1}_{W_n}(x)\mathbf{1}_{W_n}(y)}{|(W_n - x)\cap(W_n - y)|} - \omega\,(b_n\,n)^d\,(\widehat{\lambda^2})_n\,, \tag{5}$$

where

$$\omega = \int_W w(x)\,\mathrm{d}x \quad \text{and} \quad (\widehat{\lambda^2})_n = \frac{\Psi(W_n)(\Psi(W_n)-1)}{|W_n|^2}\,.$$

3. Asymptotic behaviour of $\widehat{\sigma}^2_n$ – main results

To begin with we quote two results from [3] stating the qualitative behaviour of the mean and variance of $\widehat{\sigma}^2_n$ when W_n grows large.

Theorem 1 (Heinrich, 1994) *Under the assumptions* (A0)–(A3)*, the sequence of estimators* $(\widehat{\sigma}^2_n)$ *is asymptotically unbiased for* σ^2 *, i.e.*

$$\mathsf{E}\,\widehat{\sigma}^2_n \xrightarrow[n\to\infty]{} \sigma^2\,. \tag{6}$$

and, under the additional assumptions (A4) *and* $b_n^2 n \xrightarrow[n\to\infty]{} 0$*, the sequence* $(\widehat{\sigma}^2_n)$ *is mean square consistent, i.e.*

$$\mathrm{MSE}(\,\widehat{\sigma}^2_n\,) := \mathsf{E}(\,\widehat{\sigma}^2_n - \sigma^2\,)^2 \xrightarrow[n\to\infty]{} 0\,. \tag{7}$$

The rates of convergence in (6) and (7) depend on the chosen kernel function $w(\cdot)$ (in particular on its behaviour near the the origin o), the sequence of bandwidths (b_n), and the rate of decay of $|\gamma^{(2)}_{\mathrm{red}}|(B^c(o,r))$ as $r \to \infty$, where $B^c(o,r)$ denotes the complement of the ball $B(o,r) = \{x \in \mathbb{R}^d : \|x\| \leq r\}$. Our first aim is to determine an asymptotically optimal sequence of bandwidths (b_n) minimizing the *mean squared error* of $\widehat{\sigma}^2_n$ defined by (7), which can be expressed as the sum of variance and squared bias of $\widehat{\sigma}^2_n$,

$$\mathrm{MSE}(\,\widehat{\sigma}^2_n\,) = \mathsf{Var}(\,\widehat{\sigma}^2_n\,) + (\,\mathsf{E}\,\widehat{\sigma}^2_n - \sigma^2)^2\,.$$

We mention that, if in (5) the product $b_n\, n$ is replaced by $c_n := b_n\,|W_n|^{1/d}$, relation (6) holds for any sequence (W_n) of sampling windows satisfying $c_n \xrightarrow[n\to\infty]{} \infty$. On the other hand, relation (7) remains valid for any increasing sequence (W_n) of *convex* windows satisfying additionally $b_n\, c_n \xrightarrow[n\to\infty]{} 0$ and $c_n/r(W_n) \xrightarrow[n\to\infty]{} 0$, where $r(W_n)$ stands for the inball radius of W_n , see [6].

The proof of (6) relies on multiple application of the formula

$$\mathsf{E} \sum_{x,y\in\Psi}^{\neq} f(x,y) = \lambda \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(x,x+y)\, \alpha^{(2)}_{\rm red}(\,\mathrm{d}y)\, \mathrm{d}x$$

(which is easily seen by combining (2) and (3)) to functions $f:\mathbb{R}^{2d}\mapsto \mathbb{R}^1$ occurring on the right–hand side of (5). After some rearrangements we finally arrive at

$$\mathsf{E}\,\widehat{\sigma}^2_n - \sigma^2 = \lambda \int_{\mathbb{R}^d} \Big(w\Big(\frac{x}{b_n\, n}\Big) - 1\Big) \gamma^{(2)}_{\rm red}(\,\mathrm{d}x) - \frac{\omega\,(b_n\, n)^d\,\lambda}{|W_n|^2} \int_{\mathbb{R}^d} |\, W_n \cap (W_n - y)\,|\, \gamma^{(2)}_{\rm red}(\,\mathrm{d}y)\,.$$

Thus, the bias of $\widehat{\sigma}^2_n$ can be estimated as follows:

$$|\,\mathsf{E}\,\widehat{\sigma}^2_n - \sigma^2\,| \le \lambda \Big| \int_{\mathbb{R}^d} \Big(w\Big(\frac{x}{b_n\, n}\Big) - 1\Big)\, \gamma^{(2)}_{\rm red}(\,\mathrm{d}x)\,\Big| + \Big(\frac{b_n}{2}\Big)^d \omega\,\lambda\, \|\gamma^{(2)}_{\rm red}\|_{\rm var}\,. \tag{8}$$

The proof of (7) given in [3] is based on the calculation of the variances of each of the three summands $S^{(n)}_i$, $i = 1,2,3$, in the decomposition

$$\widehat{\sigma}^2_n = \widehat{\lambda}_n + \sum_{x,\,y\in\Psi}^{\neq} f_1(x,y) - \omega\,(b_n\, n)^d \sum_{x,\,y\in\Psi}^{\neq} f_2(x,y)\,,$$

where

$$f_1(x,y) = \frac{w((y-x)/b_n\, n)\, \mathbf{1}_{W_n}(x)\, \mathbf{1}_{W_n}(y)}{|(W_n - x)\cap(W_n - y)|} \qquad \text{and} \qquad f_2(x,y) = \frac{\mathbf{1}_{W_n}(x)\, \mathbf{1}_{W_n}(y)}{|W_n|^2}\,.$$

The variance $\mathsf{Var}(\widehat{\sigma}^2_n)$ is then bounded by $3\,(\,\mathsf{Var}(S^{(n)}_1) + \mathsf{Var}(S^{(n)}_2) + \mathsf{Var}(S^{(n)}_3)\,)$, which gives

$$\mathsf{Var}(\widehat{\sigma}^2_n) = \mathcal{O}((b_n^2 n)^d) \quad \text{as} \quad n\to\infty \quad , \quad \text{see [3]}\,.$$

There are further terms like $\mathcal{O}(b_n^d)$ and $\mathcal{O}(n^{-d})$ hidden behind the $\mathcal{O}$-symbol which can be neglegted due to the assumption (A2).

However, a thorough check of these calculations shows that a slightly sharper bound of $\mathsf{Var}(\widehat{\sigma}^2_n)$ and even its exact asymptotic order can be obtained. For this purpose rewrite $\widehat{\sigma}^2_n$ as follows

$$\widehat{\sigma}^2_n = \widehat{\lambda}_n + \sum_{x,\,y\in\Psi}^{\neq} f(x,y) \quad \text{with} \quad f(x,y) = f_1(x,y) - \omega\,(b_n\, n)^d\, f_2(x,y)\,. \tag{9}$$

Theorem 2 *Under the assumptions* (A0) – (A4) *we have*

$$\mathsf{Var}(\widehat{\sigma}_n^2) = \mathcal{O}(n^{-d}) + \mathcal{O}(b_n^{2d+1}\, n^d) + \mathcal{O}(b_n^d) \quad \text{as} \quad n \to \infty\,. \tag{10}$$

If in addition $b_n^{d+1}\, n^d \xrightarrow[n\to\infty]{} 0$ *, then*

$$\frac{\mathsf{Var}(\widehat{\sigma}_n^2)}{b_n^d} \xrightarrow[n\to\infty]{} \tau^2 := \frac{\sigma^4}{2^{d-1}} \int_W w^2(x)\, \mathrm{d}x + \frac{\lambda^2\, \omega}{2^{d-2}} \Big(2\, \gamma_{\mathrm{red}}^{(2)}(\mathbb{R}^d) + \gamma_{\mathrm{red}}^{(3)}(\mathbb{R}^{2d}) \Big)\,. \tag{11}$$

Moreover, if $b_n^{d+1}\, n^d \xrightarrow[n\to\infty]{} 0$ *and the stationary point process* Ψ *is Brillinger-mixing, i.e.* $|\,\gamma_{\mathrm{red}}^{(k)}\,|(\mathbb{R}^{d(k-1)}) < \infty$ *for all* $k \geq 2$*, then* $b_n^{-d/2}\,(\,\widehat{\sigma}_n^2 - \mathsf{E}\,\widehat{\sigma}_n^2\,)$ *converges in distribution (as* $n \to \infty$*) to a Gaussian random variable* $\mathcal{N}(0, \tau^2)$ *.*

Remark *Relation* (10) *remains also valid if* $b_n\, n \xrightarrow[n\to\infty]{} c > 0$ *instead of* $b_n\, n \xrightarrow[n\to\infty]{} \infty$ *.*

As seen from (8) and (10) the choice of the (asymptotically) MSE–optimal bandwidth b_n^* (minimizing $\mathrm{MSE}(\widehat{\sigma}_n^2)$ for large enough n up to a multiplicative constant) is strongly influenced by the behaviour of the integral term on the right–hand side of (8) which in turn depends on the particular shape of w and $\gamma_{\mathrm{red}}^{(2)}$.

To facilitate a more detailed analysis of the MSE–asymptotics we choose kernel functions w being equal to 1 in some neighbourhood of the origin, e.g. the cylinder kernel

$$w(x) = \mathbf{1}_{B(o,1)}(x) \quad \text{with} \quad \omega = \int_W w^2(x)\, \mathrm{d}x = \kappa_d\,. \tag{12}$$

We distinguish three different types of the tail behaviour of $\gamma_{\mathrm{red}}^{(2)}$ – polynomial decay, (sub)exponential decay and $\gamma_{\mathrm{red}}^{(2)}$ having bounded support.

Theorem 3 *Let* (5) *be defined with the cylinder kernel* (12) *and assume that* (A0) *and* (A2)–(A4) *are satisfied. If*

$$|\gamma_{\mathrm{red}}^{(2)}(B^c(o,r))| = \mathcal{O}(r^{-\beta}) \quad \text{for some} \quad \beta > 0\,, \tag{13}$$

then

$$\mathrm{MSE}(\widehat{\sigma}_n^2) = \mathcal{O}\Big(n^{-\min\left\{ \frac{2\beta\, d}{2\beta+d}, \frac{2\beta\,(d+1)}{2\beta+2d+1)} \right\}} \Big) \quad \text{and} \quad b_n^* = c\, n^{-\max\left\{ \frac{2\beta}{2\beta+d}, \frac{d+2\beta}{1+2d+2\beta)} \right\}}$$

with some constant $c > 0$ *. If*

$$|\,\gamma_{\mathrm{red}}^{(2)}(B^c(o,r))\,| = \mathcal{O}(\,\exp\{-h(r)\}\,) \quad , \quad \text{where} \quad \frac{h(r)}{\log r} \xrightarrow[r\to\infty]{} \infty\,, \tag{14}$$

then

$$\mathrm{MSE}(\widehat{\sigma}_n^2) = \mathcal{O}\big(n^{-(1-\varepsilon)d} \big) \quad \text{for} \quad b_n = c\, n^{-(1-\varepsilon)}$$

with some constant $c > 0$ *, where* $\varepsilon > 0$ *can be chosen arbitrarily small.*
If the r.c.m. $\gamma_{\mathrm{red}}^{(2)}$ *has bounded support in* $\mathbb{R}^d$*, i.e.,*

$$c^* = \inf\{r > 0 : |\gamma_{\mathrm{red}}^{(2)}|(B^c(o,r)) = 0\} < \infty\,, \tag{15}$$

then

$$\mathrm{MSE}(\widehat{\sigma}_n^2) = \mathcal{O}(\, n^{-d}\,) \quad \text{and} \quad b_n^* = \frac{c}{n} \quad \text{for some} \quad c > c^*\,.$$

If the product density $\varrho^{(2)}$ exists and Ψ is even isotropic it is often more convenient to express the conditions (13)–(15) in terms of the corresponding pair–correlation function g. Indeed, by the definitions of $\gamma^{(2)}_{\mathrm{red}}$ and g we have

$$\gamma^{(2)}_{\mathrm{red}}(B(o,r)) = \alpha^{(2)}_{\mathrm{red}}(B(o,r)) - \lambda\,|\,B(o,r)\,| = \lambda\, d\, \kappa_d \int_0^r (\,g(u)-1\,)\, u^{d-1}\, \mathrm{d}u\ ,$$

so that

$$|\,\gamma^{(2)}_{\mathrm{red}}\,|(B^c(o,r)) = \lambda\, d\, \kappa_d \int_r^\infty |\,g(u)-1\,|\, u^{d-1}\, \mathrm{d}u\ .$$

Thus, we arrive at

Corollary 4 *Let Ψ be a motion–invariant second–order point process on $\mathbb{R}^d$ with pair–correlation function $g(r)$ $r>0$. Then*

(i) $|\,g(r)-1\,| = \mathcal{O}(r^{-(d+\beta)})$ *(as $r\to\infty$) for some $\beta>0$ implies* (13)*,*

(ii) $|\,g(r)-1\,| = \mathcal{O}(\exp\{-(1+\delta)\,h(r)\})$ *(as $r\to\infty$) for some $\delta>0$ implies* (14)*,*

(iii) $c^* = \sup\{\,r\ge 0 : g(r)\ne 1\,\} < \infty$ *is equivalent to* (15)*.*

4. Proofs of the Theorems 2 and 3

The proof of Theorem 2 is essentially based on the following

Lemma 5 *Let Ψ be a stationary fourth–order point process on $\mathbb{R}^d$ with intensity λ. Further, let $f:\mathbb{R}^d\times\mathbb{R}^d\mapsto\mathbb{R}^1$ be a bounded, symmetric, and Borel measurable function with bounded support. Then*

$$\begin{aligned}
\mathsf{Var}\Big(\sum_{x,y\,\in\Psi}^{\neq} f(x,y)\Big) &= 2\lambda \iint f^2(x,x+y)\,\gamma^{(2)}_{\mathrm{red}}(\mathrm{d}y)\,\mathrm{d}x + 2\,\lambda^2 \iint f^2(x,y)\,\mathrm{d}y\,\mathrm{d}x \\
&+\ 4\lambda \iiint f(x,x+y)f(x+y,x+u)\gamma^{(3)}_{\mathrm{red}}(\,\mathrm{d}(y,u))\,\mathrm{d}x \\
&+\ 4\lambda^2 \iiint f(x,y)[2f(y,y+u)+f(y,x+u)]\gamma^{(2)}_{\mathrm{red}}(\,\mathrm{d}u)\,\mathrm{d}y\,\mathrm{d}x \\
&+\ 4\lambda^3 \iiint f(x,y)f(y,u)\,\mathrm{d}u\,\mathrm{d}y\,\mathrm{d}x \\
&+\ \lambda \iiiint f(x,x+y)f(x+u,x+v)\gamma^{(4)}_{\mathrm{red}}(\,\mathrm{d}(y,u,v))\,\mathrm{d}x \\
&+\ 4\lambda^2 \iiiint f(x,y)f(y+u,y+v)\gamma^{(3)}_{\mathrm{red}}(\,\mathrm{d}(u,v))\,\mathrm{d}y\,\mathrm{d}x \\
&+\ 2\lambda^2 \iiiint f(x,y)f(x+u,y+v)\gamma^{(2)}_{\mathrm{red}}(\,\mathrm{d}u)\gamma^{(2)}_{\mathrm{red}}(\,\mathrm{d}v)\,\mathrm{d}y\,\mathrm{d}x \\
&+\ 4\lambda^3 \iiiint f(x,y)f(x+u,v)\gamma^{(2)}_{\mathrm{red}}(\,\mathrm{d}u)\,\mathrm{d}v\,\mathrm{d}y\,\mathrm{d}x.
\end{aligned}$$

Proof of Lemma 5 : Applying the defining relation (2) to the decomposition

$$\Big(\sum_{x,y\in\Psi}^{\neq} f(x,y)\Big)^2 = 2\sum_{x,y\in\Psi}^{\neq} f^2(x,y) + 4\sum_{x,y,u\in\Psi}^{\neq} f(x,y)\,f(y,u) + \sum_{x,y,u,v\in\Psi}^{\neq} f(x,y)f(u,v)$$

we find that

$$\begin{aligned}
\mathsf{Var}\Big(\sum_{x,y\in\Psi}^{\neq} f(x,y)\Big) &= \mathsf{E}\Big(\sum_{x,y\in\Psi}^{\neq} f(x,y)\Big)^2 - \Big(\mathsf{E}\sum_{x,y\in\Psi}^{\neq} f(x,y)\Big)^2\\
&= 2\iint f^2(x,y)\,\alpha^{(2)}(\,\mathrm{d}(x,y)) + 4\iiint f(x,y)f(y,u)\,\alpha^{(3)}(\,\mathrm{d}(x,y,u))\\
&\quad + \iiiint f(x,y)\,f(u,v)\,\big[\alpha^{(4)}(\,\mathrm{d}(x,y,u,v)) - \alpha^{(2)}(\,\mathrm{d}(x,y))\alpha^{(2)}(\,\mathrm{d}(u,v))\big]\\
&= 2\iint f^2(x,y)\,\big[\gamma^{(2)}(\,\mathrm{d}(x,y)) + \lambda^2\,\mathrm{d}x\,\mathrm{d}y\big] + 4\iiint f(x,y)f(y,u)\\
&\quad \times\big[\,\gamma^{(3)}(\,\mathrm{d}(x,y,u)) + 2\lambda\,\gamma^{(2)}(\,\mathrm{d}(y,u))\,dx + \lambda\,\gamma^{(2)}(\,\mathrm{d}(x,u))\,\mathrm{d}y + \lambda^3\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}u\,\big]\\
&\quad + \iiiint f(x,y)\,f(u,v)\,\big[\gamma^{(4)}(\,\mathrm{d}(x,y,u,v)) + 4\,\lambda\,\gamma^{(3)}(\,\mathrm{d}(y,\mathrm{d}u,\mathrm{d}v)\,\mathrm{d}x\\
&\qquad\qquad + 2\gamma^{(2)}(\,\mathrm{d}(x,u))\,\gamma^{(2)}(\,\mathrm{d}(y,v)) + 4\,\lambda^2\,\gamma^{(2)}(\,\mathrm{d}(x,u))\,\mathrm{d}y\,\mathrm{d}v\,\big].
\end{aligned}$$

The second equality is seen by expressing the third–order factorial cumulant measure $\gamma^{(3)}$ in terms of $\alpha^{(3)}$, $\alpha^{(2)}$, and $\alpha^{(1)}$ and applying the identity

$$\begin{aligned}
\gamma^{(4)}\big(\mathop{\times}_{i=1}^{4} A_i\,\big) &= \alpha^{(4)}(A_1\times A_2\times A_3\times A_4) - \alpha^{(2)}(A_1\times A_2)\alpha^{(2)}(A_3\times A_4)\\
&\quad - \lambda\,\Big(\,|A_1|\,\gamma^{(3)}(A_2\times A_3\times A_4) + |A_2|\,\gamma^{(3)}(A_1\times A_3\times A_4)\\
&\qquad + |A_3|\,\gamma^{(3)}(A_1\times A_2\times A_4) + |A_4|\,\gamma^{(3)}(A_1\times A_2\times A_3)\Big)\\
&\quad - \gamma^{(2)}(A_1\times A_3)\,\gamma^{(2)}(A_2\times A_4) - \gamma^{(2)}(A_1\times A_4)\,\gamma^{(2)}(A_2\times A_3)\\
&\quad - \lambda^2\,\Big(\,|A_2|\,|A_4|\,\gamma^{(2)}(A_1\times A_3) + |A_2|\,|A_3|\,\gamma^{(2)}(A_1\times A_4)\\
&\qquad + |A_1|\,|A_4|\,\gamma^{(2)}(A_2\times A_3) + |A_1|\,|A_3|\,\gamma^{(2)}(A_2\times A_4)\Big)
\end{aligned}$$

for any $A_1,\ldots,A_4\in\mathcal{B}^d$. Finally, the assertion of Lemma 5 follows by disintegrating the factorial cumulant measures in analogy to (3). □

Proof of Theorem 2 : From (9) and (1) it is easily seen that

$$\begin{aligned}
\mathsf{Var}(\widehat{\sigma}_n^2) &\le 2\ \mathsf{Var}(\widehat{\lambda}_n) + 2\ \mathsf{Var}(S_n)\\
&\le 2\,\lambda\,(1+\|\gamma^{(2)}_{\mathrm{red}}\|_{\mathrm{var}})\,|W_n|^{-1} + 2\ \mathsf{Var}(S_n)\ ,
\end{aligned}\tag{16}$$

where $S_n = \sum^{\neq}_{x,y\in\Psi} f(x,y)$ with $f(x,y)$ defined in (9). Applying Lemma 5 to this function $f(x,y)$ we may write $\mathsf{Var}(S_n) = \sum_{i=1}^{9} T_i^{(n)}(f)$, where the summands $T_i^{(n)}(f)$

appear in the same order as in the variance representation of Lemma 5. By the rough estimate

$$|f(x,y)| \leq \frac{c_1}{|W_n|}\,\mathbf{1}_{W_n}(x)\,\mathbf{1}_{W_n}(y)$$

and the assumptions (A3) and (A4) it is readily seen that

$$|\,T_i^{(n)}(f)\,| \leq c_2\,|W_n|^{-1} \quad \text{for} \quad i = 1, 3, 6\ .$$

Next we consider those terms in which exactly two integrals w.r.t. the Lebesgue measure on $\mathbb{R}^d$ appear. It turns out that the asymptotic behaviour of each of these terms is determined by the function $f_1(x,y)$. We demonstrate this in detail with the typical term $T_8^{(n)}(f)$. A short calculation shows that

$$\iint f(x,y)\, f(x+u, y+v)\; \mathrm{d}y\, \mathrm{d}x = \iint f_1(x,y)\, f_1(x+u, y+v)\; \mathrm{d}y\, \mathrm{d}x + \mathcal{O}(b_n^{2d})$$

and, together with (A2),

$$\begin{aligned} &\frac{1}{b_n^d} \iint f_1(x,y)\, f_1(x+u,y+v)\; \mathrm{d}y\, \mathrm{d}x \\ &\quad = n^d \int \frac{w(y)\, w(y + \frac{v-u}{n\,b_n})\, |W_n \cap (W_n - nb_n\, y) \cap (W_n - u) \cap (W_n - nb_n\, y - v)|}{|W_n \cap (W_n - nb_n\, y)|\, |(W_n - u) \cap (W_n - nb_n\, y - v)|}\; \mathrm{d}y \\ &\quad \xrightarrow[n\to\infty]{} \frac{1}{2^d} \int w^2(y)\; \mathrm{d}y \end{aligned}$$

for any fixed $u, v \in \mathbb{R}^d$. Here, we have used among others that, for any $\varepsilon > 0$, there is a continuous function $w_\varepsilon : W \to \mathbb{R}^1$ such that $\int |w(y) - w_\varepsilon(y)|\,\mathrm{d}y \leq \varepsilon$. Likewise, one can verify that

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{b_n^d} \iint f(x,y)\, f(y+u, y+v)\; \mathrm{d}y\, \mathrm{d}x \\ &\quad = \lim_{n\to\infty} n^d \int \frac{w(y)\, w(\frac{v-u}{n\,b_n})\, |W_n \cap (W_n - nb_n\, y) \cap (W_n - u) \cap (W_n - v)|}{|W_n \cap (W_n - nb_n\, y)|\, |(W_n - u) \cap (W_n - v)|}\; \mathrm{d}y \\ &\quad = \frac{1}{2^d} \int w(y)\; \mathrm{d}y \end{aligned}$$

for any fixed $u, v \in \mathbb{R}^d$. Hence, in view of the assumptions (A3) and (A4), the dominated convergence theorem yields the asymptotic relations

$$\begin{aligned} T_2^{(n)}(f) &\sim 2\,\lambda^2\, \frac{b_n^d}{2^d} \int w^2(y)\; \mathrm{d}y\ , \\ T_4^{(n)}(f) &\sim 4\,\lambda^2\, \frac{b_n^d}{2^d}\, \gamma_{\text{red}}^{(2)}(\mathbb{R}^d) \Big(2 \int w(y)\; \mathrm{d}y + \int w^2(y)\; \mathrm{d}y \Big)\ , \\ T_7^{(n)}(f) &\sim 4\,\lambda^2\, \frac{b_n^d}{2^d}\, \gamma_{\text{red}}^{(3)}(\mathbb{R}^d \times \mathbb{R}^d) \int w(y)\; \mathrm{d}y\ , \\ T_8^{(n)}(f) &\sim 2\,\lambda^2\, \frac{b_n^d}{2^d}\, (\gamma_{\text{red}}^{(2)}(\mathbb{R}^d))^2 \int w^2(y)\; \mathrm{d}y\ , \end{aligned}$$

where $a_n \sim a'_n$ means $a_n/a'_n \underset{n\to\infty}{\longrightarrow} 1$. To study the contribution of the remaining terms $T_5^{(n)}(f)$ and $T_9^{(n)}(f)$ to $\mathsf{Var}(S_n)$ we have to evaluate the corresponding integrals explicitly. After a lengthy calculation we get

$$T_5^{(n)}(f) = 4\,\lambda^3 \Big(\frac{b_n^2\,n}{2}\Big)^d \iint \Big(\frac{|W_n|\,|W_n \cap (W_n - nb_n\,x) \cap (W_n - nb_n\,y)|}{|W_n \cap (W_n - nb_n\,x)|\,|W_n \cap (W_n - nb_n\,y)|} - 1 \Big) \\ \times\; w(x)\,w(y)\;\mathrm{d}x\,\mathrm{d}y$$

and

$$T_9^{(n)}(f) = 4\,\lambda^3 \Big(\frac{b_n^2\,n}{2}\Big)^d \iiint \frac{w(x)\,w(y)\,|W_n|}{|W_n \cap (W_n - nb_n\,x)|} \\ \times \Bigg(\frac{|W_n^{(u)} \cap (W_n - nb_n\,x) \cap (W_n - nb_n\,y - u)|}{|W_n \cap (W_n - nb_n\,y)|} - \frac{|W_n^{(u)} \cap (W_n - nb_n\,x)|}{|W_n|} \\ + \frac{|W_n^{(u)}|}{|W_n|} \Big(\frac{|W_n \cap (W_n - nb_n\,x)|}{|W_n|} - 1 \Big) \Bigg)\,\mathrm{d}x\;\mathrm{d}y\,\gamma_{\mathrm{red}}^{(2)}(\mathrm{d}u)\,, \tag{17}$$

where we have used the abbreviation $W_n^{(u)} = W_n \cap (W_n - u)$ for $u \in \mathbb{R}^d$.

By assumption (A0) it is easily shown that

$$1 - \frac{|W_n^{(n\,b_n\,y)} \cap (W_n - n\,b_n\,x)|}{|W_n^{(n\,b_n\,y)}|} = 1 - \frac{|W \cap (W - b_n\,y) \cap (W - b_n\,x)|}{|W \cap (W - b_n\,y)|} \le c_3\,b_n$$

with some constant $c_3 > 0$ not depending on n and $x, y \in W$.

Hence,

$$T_5^{(n)}(f) = \mathcal{O}(b_n^{2d+1}\,n^d) \quad \text{as} \quad n \to \infty\,.$$

Using the inequality

$$\begin{aligned} &|W_n^{(u)} \cap (W_n - nb_n\,x)| - |W_n^{(u)} \cap (W_n - nb_n\,x) \cap (W_n - nb_n\,y - u)| \\ &\quad = |W_n \cap (W_n - nb_n\,x) \cap \big((W_n - u) \setminus (W_n - nb_n\,y - u)\big)| \\ &\quad \le |W_n| - |W_n \cap (W_n - nb_n\,y)| \\ &\quad \le c_3\,|W_n|\,b_n\,, \end{aligned}$$

we see that the integrand in (17) is uniformly bounded by $c_4\,b_n$ for all $(x, y, u) \in W \times W \times \mathbb{R}^d$ and $n \ge 1$ so that, together with (A3),

$$T_9^{(n)}(f) = \mathcal{O}(b_n^{2d+1}\,n^d) \quad \text{as} \quad n \to \infty\,.$$

Finally, summarizing the obtained bounds of $T_1^{(n)}(f), \ldots, T_9^{(n)}(f)$ yields $\mathsf{Var}(S_n) = \mathcal{O}(n^{-d}) + \mathcal{O}(b_n^{2d+1}\,n^d) + \mathcal{O}(b_n^d)$ and then (10) follows from (16).

If the sequence $b_n\,n$ has a finite and positive limit, then the above estimates remains almost unchanged resulting in $\mathsf{Var}(\,\widehat{\sigma}_n^2\,) = \mathcal{O}(n^{-d})$. This justifies the Remark immediately after Theorem 2.

Further, provided that $b_n\, n \xrightarrow[n\to\infty]{} \infty$ and $b_n^{d+1}\, n^d \xrightarrow[n\to\infty]{} 0$, the above asymptotic relations for $T_i^{(n)}(f)$, $i = 2, 4, 7, 8$, imply the existence of the limit

$$\lim_{n\to\infty} \frac{\mathsf{Var}(\widehat{\sigma}_n^2)}{b_n^d} = \lim_{n\to\infty} b_n^{-d} \Big(T_2^{(n)}(f) + T_4^{(n)}(f) + T_7^{(n)}(f) + T_8^{(n)}(f) \Big) \,,$$

which gives (11) after some obvious rearrangements.

The essential step to prove the third assertion of Theorem 2 consists in showing that the k-th order cumulant of the normalized sum $b_n^{-d/2}\, S_n$ converges to 0 for any $k \geq 3$. For this purpose the assumption $|\,\gamma_{\mathrm{red}}^{(j)}\,|(\mathbb{R}^{d(k-1)}) < \infty$ for $j = 2, \ldots, k$ is needed. The treatment of the higher–order cumulants of S_n is quite similar to the estimation procedure carried out in [7] to prove a central limit theorem for functionals of the form $\sum_{x,y\in\Psi}^{\neq} \mathbf{1}_{W_n}(x) g(y-x)$, where the function $g : \mathbb{R}^d \mapsto \mathbb{R}^1$ does not depend on n. We omit the details of these rather lengthy computations. The remaining term $b_n^{-d/2}\,(\lambda_n - \lambda)$ tends to 0 in probability. Thus, together with Slutsky's lemma, see [16], and (11) we obtain the desired normal convergence. This completes the proof of Theorem 2. □

Proof of Theorem 3 : The relations (10) and (8) with the cylinder kernel (12) imply

$$\mathrm{MSE}(\,\widehat{\sigma}_n^2\,) = \mathcal{O}\Big(\big(\gamma_{\mathrm{red}}^{(2)}(B^c(o, nb_n))\big)^2 \Big) + \mathcal{O}(b_n^d) + \mathcal{O}(b_n^{2d+1}\, n^d) + \mathcal{O}(n^{-d}) \qquad (18)$$

Suppose now that (13) holds. By the equations

$$(\,b_n\, n\,)^{-2\beta} = b_n^d \quad \text{and} \quad (\,b_n\, n\,)^{-2\beta} = b_n^{2d+1}\, n^d$$

we get two canditates of bandwidths $b_n^{(1)} = n^{-2\beta/(2\beta+d)}$ and $b_n^{(2)} = n^{-(2\beta+d)/(2\beta+2d+1)}$ (up to a multiplicative constant $c > 0$) which minimize the right–hand side of (18). Therefore,

$$\mathrm{MSE}(\,\widehat{\sigma}_n^2\,) = \mathcal{O}\Big(\max\{(b_n^{(1)})^d, (b_n^{(2)})^{2d+1}\, n^d)\} \Big)$$

with the optimal bandwidth $b_n^* = c\, \min\{b_n^{(1)}, b_n^{(2)}\}$. Note that $b_n^{(1)} \leq b_n^{(2)}$ iff $\beta \geq d^2/2$ and this is equivalent to $(b_n^{(1)})^d \geq (b_n^{(2)})^{2d+1}\, n^d$.

If (14) holds we may take the parameter β in the previous case arbitrarily large. In particular, letting $2\beta = d(1-\varepsilon)/\varepsilon$ for $0 < \varepsilon \leq 1/(d+1)$ gives the optimal bandwidth $b_n^* = c\, n^{-(1-\varepsilon)}$.

The third part of Theorem 3 is an obvious consequence of (18) combined with the Remark after Theorem 2. □

5. The isotropised estimator $\widetilde{\sigma}_n^2$

Inserting the kernel function (12) in (5) we may write the estimator $\widehat{\sigma}_n^2$ as follows:

$$\widehat{\sigma}_n^2 = \widehat{\lambda}_n + \widehat{(\lambda^2 K)}_n(n\, b_n) - \kappa_d\, (n\, b_n)^d\, \widehat{(\lambda^2)}_n \,. \qquad (19)$$

The middle term is obtained by taking $r = n\, b_n$ in

$$\widehat{(\lambda^2 K)}_n(r) = \sum_{x,\, y \in \Phi}^{\neq} \frac{\mathbf{1}_{B(o,r)}(x-y)\, \mathbf{1}_{W_n}(x)\mathbf{1}_{W_n}(y)}{|(W_n - x) \cap \ (W_n - y)|}\ , \tag{20}$$

which is nothing else but the standard edge–corrected Horvitz–Thompson type estimator of $\lambda^2\, K(r)$, where $K(r) (= \lambda^{-1}\, \alpha^{(2)}_{\rm red}(B(o,r)))$ denotes Ripley's K-function, see [12], [15] and [10, Sect.9.1.2] for its relevance in statistical analysis of planar and spatial point patterns. If the stationary point process Ψ is additionally isotropic (that is, Ψ is motion–invariant), the following isotropised estimator of $\lambda^2\, K(r)$ is preferably used due to certain optimality properties, see [12, Chapt.3], [15, p.135/136] and references given there or [13]:

$$\widetilde{(\lambda^2 K)}_n(r) = \sum_{x,y\, \in \Phi}^{\neq} \frac{\mathbf{1}_{B(o,r)}(x-y)\, \mathbf{1}_{W_n}(x)\mathbf{1}_{W_n}(y)}{|W_n^{[\|x-y\|]}|}\, k(x,y) \quad \text{for} \quad 0 \le r \le r_n^*\ , \tag{21}$$

where $W_n^{[r]} = \{x \in W_n : \partial B(x,r) \cap W_n \neq \emptyset\}$, $r_n^* = \sup\{r : |W_n^{(r)}| > 0\}$, and

$$k(x,y) = \frac{|\partial B(x, \|x-y\|)|_{d-1}}{|\partial B(x, \|x-y\|) \cap W_n|_{d-1}} \quad \text{for} \quad x, y \in W_n\ ,\ x \neq y\, .$$

Here $|\cdot|_{d-1}$ designates the $(d-1)$–dimensional Hausdorff measure in $\mathbb{R}^d$ and $\partial B(x,r)$ denotes the surface of the ball $B(x,r)$. In the special case $d = 2$, we get $k(x,y) = 2\pi/\alpha(x,y)$, where $\alpha(x,y)$ equals the sum of all angles of the arcs in W_n induced by a circle with radius $r = \|x-y\|$ centered at $x \in W_n$.

The isotropised estimator (21) is believed to have similar properties like the Horvitz–Thompson type estimator (20) for r being small in comparison with the length of the edges of W_n, and for large r it should have smaller variance than (20).

Hence, for a motion–invariant point process Ψ we introduce a new estimator of σ^2 defined by

$$\widetilde{\sigma}_n^2 = \widehat{\lambda}_n \, + \, \widetilde{(\lambda^2 K)}_n(b_n\, n) \, - \, \kappa_d\, (n\, b_n)^d\, \widehat{(\lambda^2)}_n\ , \tag{22}$$

where $|W_n^{[b_n\, n]}| = |W_n|$ for $n \ge 1$, since $b_n \le 1$ by (A2).

Note that the estimator $\widetilde{\sigma}_n^2$ does not allow to be written in the form (5) because the function $k(x,y)$ depends on the window W_n. The specific shape of $W_n = [-n,n]^d$ implies that

$$k(x,y) \le 2^d \qquad \text{for all} \quad x, y, \in W_n \quad \text{satisfying} \quad \|x-y\| \le n\, ,$$

and in the special case $d = 2$, a little trigonometry yields

$$\frac{1}{k(x,y)} = 1 - a_n^{(1)}(x,y) - a_n^{(2)}(x,y) - \min\left\{ \frac{1}{4}, a_n^{(1)}(x,y) + a_n^{(2)}(x,y) \right\}\ ,$$

where, for $x = (x_1, x_2)\, , y \in W_n$,

$$a_n^{(i)}(x,y) = \begin{cases} \frac{1}{2\pi}\ \arccos\left(\frac{n - |x_i|}{\|x-y\|}\right) & \text{if} \quad 0 \le n - \|x-y\| \le |x_i| \\ \\ 0 & \text{if} \quad 0 \le |x_i| \le n - \|x-y\| \end{cases}\ , \qquad i = 1, 2\, .$$

A careful analysis of the quantities MSE($\widetilde{\sigma}_n^2$) and $\mathsf{Var}(\widetilde{\sigma}_n^2)$ confirms that the sequence of estimators $\widetilde{\sigma}_n^2$ behaves asymptotically like $\widehat{\sigma}_n^2$ with the cylinder kernel (12). This results are summarized in

Theorem 6 *If the point process Ψ is additionally isotropic, the Theorems 1–3 formulated for the particular case of the cylinder kernel* (12) *remain completely valid when $\widehat{\sigma}_n^2$ is replaced by $\widetilde{\sigma}_n^2$.*

Though $\widetilde{\sigma}_n^2$ has the same asymptotic behaviour as $\widehat{\sigma}_n^2$, the advantage of $\widetilde{\sigma}_n^2$ in comparison with $\widehat{\sigma}_n^2$ is that the correction weights $k(x,y)$ exceed 1 not for all pairs of distant points $x,y \in W_n$, but only when these points are close to the boundary of W_n . Thus the variance of $\widetilde{\sigma}_n^2$ on moderate–size windows is expected to be smaller than that of $\widehat{\sigma}_n^2$ with cylinder kernel. On the other hand, $\widetilde{\sigma}_n^2$ should have approximately the same bias as $\widehat{\sigma}_n^2$ with the cylinder kernel (12). The behaviour of $\widehat{\sigma}_n^2$ and $\widetilde{\sigma}_n^2$ on moderate–size windows will be discussed in detail in Sect. 7.

Theorem 6 will not be proved in detail because its proof differs from the above proofs of the Theorems 1–3 only in a few technical arguments. In the remaining part of Sect. 5 we only outline these slight differences. First note that in case of motion-invariant point processes the estimator $\widetilde{(\lambda^2 K)}_n(r)$ is unbiased for $\lambda\, \alpha^{(2)}_{\rm red}(B(o,r))$, see [11], which implies

$$\mathsf{E}\,\widetilde{\sigma}_n^2 = \lambda + \lambda\, \gamma^{(2)}_{\rm red}(B(o,b_n\, n)) - \frac{\kappa_d (b_n\, n)^d \lambda}{|W_n|^2} \int_{\mathbb{R}^d} |W_n \cap (W_n - y)| \gamma^{(2)}_{\rm red}({\rm d}y)\,.$$

Therefore we get in analogy to (8) that

$$|\,\mathsf{E}\,\widetilde{\sigma}_n^2 - \sigma^2| \le \lambda\, |\gamma^{(2)}_{\rm red}(B^c(o,b_n\, n))| + \left(\frac{b_n}{2}\right)^d \lambda\, \kappa_d\, \|\gamma^{(2)}_{\rm red}\|_{\rm var},$$

which shows the asymptotic unbiasedness of $\widetilde{\sigma}_n^2$ under (A0)–(A3). Under the additional assumption (A4) it follows that MSE($\widetilde{\sigma}_n^2$) $\xrightarrow[n\to\infty]{} 0$ provided that $b_n^{2d+1}\, n^d \xrightarrow[n\to\infty]{} 0$. This is directly seen from the latter inequality in combination with

$$\mathsf{Var}(\widetilde{\sigma}_n^2) = \mathcal{O}(n^{-d}) + \mathcal{O}(b_n^{2d+1}\, n^d) + \mathcal{O}(b_n^d) \quad \text{as} \quad n \to \infty\,. \tag{23}$$

This and the foregoing relation enable us to derive the optimal order of MSE($\widetilde{\sigma}_n^2$) just as in Theorem 3. To prove (23) we start from (22) with the simple inequality

$$\mathsf{Var}(\widetilde{\sigma}_n^2) \le 2\,\lambda\,(1 + \|\gamma^{(2)}_{\rm red}\|_{\rm var})\, |W_n|^{-1} + 2\; \mathsf{Var}(\widetilde{S}_n)\,,$$

where

$$\widetilde{S}_n = \frac{1}{|W_n|} \sum_{x,y \in \Psi}^{\neq} \Big(\widetilde{f}(x,y) - \frac{\kappa_d}{2^d}\, b_n^d \Big)\, \mathbf{1}_{W_n}(x)\, \mathbf{1}_{W_n}(y)$$

with a symmetric function $\widetilde{f} : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}^1$ defined by

$$\tilde{f}(x,y) = \tfrac{1}{2}\,(\, k(x,y) + k(y,x)\,)\, \mathbf{1}_{B(o,n\, b_n)}(x-y)\;.$$

Here we have used the definitions (21) and (22) and that $W_n^{[r]} = W_n$ for $r \in [0, b_n\, n]$. Now, we may proceed as in the proof of Theorem 3. It turns out after some lengthy calculations that the both weights $k(x,y)$ and $k(y,x)$ in the definition of $\widetilde{f}$ can be replaced by 1 without changing the asymptotic order of the variance $\mathsf{Var}(\widetilde{S}_n)$. In this way we finally obtain (23) and that the limit $\lim_{n\to\infty} \mathsf{Var}(\widetilde{\sigma}_n^2)\, b_n^{-d}$ exists and coincides with (11) for $w(x) = \mathbf{1}_{B(o,1)}(x)$. □

6. Other methods of estimating σ^2

For the sake of completeness we briefly discuss two other possibilities to estimate the asymptotic variance (1). The above formula (4) yields

$$\sigma^2 = \lambda + \int_{\mathbb{R}^d} (\, \lambda\, \varrho^{(2)}(x) - \lambda^2\,)\, \mathrm{d}x$$

and this suggests a further estimator $\widehat{\sigma}_{n,1}^2$ for σ^2 by using an appropriate edge–corrected kernel–type estimator for $\lambda\, \varrho^{(2)}(x)$ which can be defined on the sampling window $W_n = [-n, n]^d$ by

$$\widehat{(\lambda\, \varrho^{(2)})}_n(x) = \frac{1}{b_n^d} \sum_{u,v\in\Psi}^{\neq} \frac{\mathbf{1}_{W_n}(u)\, \mathbf{1}_{W_n}(v)}{|\,(W_n - u) \cap (W_n - v)\,|}\, k\Big(\frac{v-u-x}{b_n}\Big)\,, \tag{24}$$

see [15],[10],[14], where the sequence of bandwidths (b_n) satisfies (A2) and the kernel function $k\,:\,\mathbb{R}^d \mapsto \mathbb{R}^1$ is assumed to be bounded with bounded support such that $\int_{\mathbb{R}^d} k(x)\, \mathrm{d}x = 1$. Together with the estimators $\widehat{\lambda}_n$ and $(\widehat{\lambda^2})_n$ for λ and λ^2, respectively, we introduce the estimator

$$\widehat{\sigma}_{n,1}^2 = \widehat{\lambda}_n + \int_{B(o,b_n\, n)} \Big(\, \widehat{(\lambda\, \varrho^{(2)})}_n(x) \;-\; (\widehat{\lambda^2})_n \Big)\, \mathrm{d}x\;,$$

which seems to have similar properties as $\widehat{\sigma}_n^2$. Rates of the strong convergence of $\widehat{(\lambda\, \varrho^{(2)})}_n(x)$ obtained in [5] can be used to derive strong convergence results for $\widehat{\sigma}_{n,1}^2$. One can also define an isotropised variant by using corresponding kernel–type estimators of the pair–correlation function g, see [10, Chapt. 9.1.3] and [14] for several improved versions.

There is a further alternative to estimate σ^2, namely by applying spectral analysis of stationary second–order point processes, see [2, Chapt. 11] for details and [1], [8] for extensions to random closed sets. Let the point process Ψ satisfy (A3) and let $h\,:\,\mathbb{R}^d \mapsto \mathbb{R}^1$ be a nonnegative, integrable function such that $\int_{\mathbb{R}^d} h(x)\, \mathrm{d}x = 1$. The covariance function $K_h(x) = \mathsf{Cov}(\xi_h(x), \xi_h(o))$ of the stationary random field $\xi_h(x) = \sum_{y\in\Psi} h(y-x)$, $x \in \mathbb{R}^d$, (being a shot-noise process, see [4]) possesses the form

$$K_h(x) = \lambda\, h^{(2)}(x) + \lambda \int_{\mathbb{R}^d} h^{(2)}(x-y)\gamma_{\mathrm{red}}^{(2)}(\mathrm{d}y) \;\text{ where }\; h^{(2)}(x) = \int_{\mathbb{R}^d} h(z)\, h(z+x)\, \mathrm{d}z\,,$$

and is absolutely integrable with $\int_{\mathbb{R}^d} K_h(x)\,\mathrm{d}z = \sigma^2$. In view of Bochner's theorem there exists a unique nonnegative, integrable (continuous) function $f_h : \mathbb{R}^d \mapsto \mathbb{R}^1$, the *spectral density* of ξ_h, which coincides with the usual Fourier transform of K_h given by

$$f_h(y) = \int_{\mathbb{R}^d} K_h(x)\, e^{i\langle x,y\rangle}\,\mathrm{d}x = \lambda\,|\,H(y)\,|^2 \Big(1 + \int_{\mathbb{R}^d} e^{i\langle x,y\rangle} \gamma^{(2)}_{\mathrm{red}}(\mathrm{d}y)\Big)\,,$$

where $H(y) = \int_{\mathbb{R}^d} h(z)\, e^{i\langle y,z\rangle}\,\mathrm{d}z$. By the properties of h it is easily seen that $f_h(o) = \sigma^2$.

A quite natural kernel-type estimator for the spectral density f_h, see e.g. [1], is given by

$$\widehat{(f_h)}_n(y) = \frac{1}{|W_n|\, b_n^d} \int_{\mathbb{R}^d} k\Big(\frac{y-x}{b_n}\Big) \Big|\int_{W_n} (\,\xi_h(z) - \lambda\,)\, e^{i\langle x,z\rangle}\,\mathrm{d}z\Big|^2 \mathrm{d}x \;\;, \quad \text{for } y \in \mathbb{R}^d\,,$$

where the kernel function k and the bandwidths (b_n) are chosen as in (24) ensuring the pointwise asymptotic unbaisedness of $\widehat{(f_h)}_n$ for f_h. Hence, $\widehat{(f_h)}_n(o)$ is a suitable estimator for σ^2.

Finally, after replacing h by $h_\varepsilon(x) = (2\varepsilon)^{-d}\,\mathbf{1}_{\varepsilon W}(x)$ we can show that

$$f_{h_\varepsilon}(y) \underset{\varepsilon\to 0}{\longrightarrow} \lambda\Big(1 + \int_{\mathbb{R}^d} e^{i\langle x,y\rangle}\gamma^{(2)}_{\mathrm{red}}(\mathrm{d}x)\Big)$$

and

$$\int_{W_n} \xi_{h_\varepsilon}(z)\, e^{i\langle x,z\rangle}\,\mathrm{d}z \underset{\varepsilon\to 0}{\longrightarrow} \sum_{z\in\Psi} \mathbf{1}_{W_n}(z)\, e^{i\langle x,z\rangle}\,.$$

Thus, we may suggest a further estimator for σ^2,

$$\widehat{\sigma}^2_{n,2} = \frac{1}{|W_n|\, b_n^d} \int_{\mathbb{R}^d} k\Big(\frac{-x}{b_n}\Big) \Big|\sum_{z\in\Psi} \mathbf{1}_{W_n}(z)\, e^{i\langle x,z\rangle} - \lambda \int_{W_n} e^{i\langle x,z\rangle}\,\mathrm{d}z\Big|^2 \mathrm{d}x\,.$$

In practise the use of the Fast Fourier Transform could reduce the complexity of the computation of $\widehat{\sigma}^2_{n,2}$. The asymptotic properties of $\widehat{\sigma}^2_{n,2}$ can be obtained with the methods used in [1].

7. Behaviour of the estimators on sampling windows W_n of moderate size – simulation results

Turning back to the decomposition (19) we can see once again that the middle term is responsible for the main part of the variability of the estimator $\widehat{\sigma}^2_n$, especially, when the value of b_n is close to 1 since the correction weights

$$1 \le \frac{|W_n|}{|(W_n - x)\cap(W_n - y)|} \le \frac{1}{(1-b_n)^d} \qquad \text{for} \quad (x-y) \in [-n\,b_n, n\,b_n]^d, \qquad (25)$$

may exceed 1 considerably. For small windows the estimator may take even negative values. The variability of the middle term depends also on the shape of the chosen kernel w. For kernel functions like

$$w(x) = \max(1 - \|x\|^d, 0) \quad \text{or} \quad w(x) = \max(1 - \|x\|, 0);,$$

the larger correction weights (25) are balanced by smaller values of the kernel function for larger values of $x - y$. On the other hand, for these two kernel functions the bias of $\widehat{\sigma}_n^2$ can be larger than that for the cylinder kernel (12) due to the integral term in (8).

To get an impression of how the estimators $\widehat{\sigma}_n^2$ and $\widetilde{\sigma}_n^2$ behave on moderate–size windows we present a simulation study. For this we generate realizations of four different models of planar motion–invariant point processes satisfying (A4) on three squares of different size,

$$W_5 = [-5, 5]^2, \quad W_{10} = [-10, 10]^2, \quad \text{and} \quad W_{20} = [-20, 20]^2 .$$

For each window size and each of the simulated point patterns we compare the values of four different estimators of σ^2 which are

- $\widehat{\sigma}_n^2$ with the cylinder kernel $w(x) = \mathbf{1}_{B(o,1)}(x)$
- $\widehat{\sigma}_n^2$ with the half-ball kernel $w(x) = \max(1 - \|x\|^2, 0)$
- $\widehat{\sigma}_n^2$ with the cone kernel $w(x) = \max(1 - \|x\|, 0)$
- the isotropised estimator $\widetilde{\sigma}_n^2$

with variable bandwidths b_n satisfying $n\, b_n \in [1, 3]$, $n\, b_n \in [1, 5]$ and $n\, b_n \in [1, 6]$, for W_5, W_{10} and W_{20}, respectively. The squared bias, the variance and the MSE are estimated for each of these estimators from 100 realizations of the simulated point processes.

To be precise we briefly describe the point process models and compute σ^2 as well as their r.c.m. $\gamma_{\text{red}}^{(2)}$; more information on these models the reader can find e.g. in [15].

(i) Poisson process with intensity $\lambda = 1$ ($\gamma_{\text{red}}^{(2)} \equiv 0$ and $\sigma^2 = \lambda = 1$).

(ii) Matérn cluster process with intensity $\lambda = 1$, mean cluster size $\mu = 5$ and cluster radius $r = 1$ ($\gamma_{\text{red}}^{(2)}(B^c(o, 2)) = 0$ and $\sigma^2 = \lambda(1 + \mu) = 6$).

This Poisson cluster process is generated by a stationary Poisson process of parent points with intensity $\mu^{-1} = 0.2$; the typical cluster consists of a Poisson distributed number of daughter points with locations independently and uniformly distributed on the disk $B(o, r)$. This gives $\gamma_{\text{red}}^{(2)}(B(o, r)) = \mu\, |B(o, 1)|^{-2} \int_{B(o,r)} |B(o, 1) \cap B(x, 1)| \, dx$.

(iii) Modified Thomas process with intensity $\lambda = 1$, mean cluster size $\mu = 5$, and variability parameter $v = 1$ ($\gamma_{\text{red}}^{(2)}(B(o, r)) = \mu\,(1 - \exp\{-r^2/4v\})$ and $\sigma^2 = \lambda(1 + \mu) = 6$).

This Poisson cluster process has the same parent point process and cluster size distribution as in (ii), but each member in the typical cluster has independent $\mathcal{N}(0,v)$–distributed coordinates. Thus, the clusters are unbounded in contrast to Matérn's cluster processes (ii).

(iv) Matérn (II) hard–core process with hard–core distance $h = 1/2$ and $\lambda_p = 1$.

This point process, denoted by Ψ_{hc}, is derived from a stationary Poisson process Ψ_p with intensity λ_p by dependent thinning. The points $x \in \Psi_p$ are marked independently by random numbers $m(x)$ distributed uniformly on $(0,1)$. Then Ψ_{hc} consists of those points of Ψ_p which survive the following thinning procedure:

$$x \in \Psi_{hc} \quad \text{iff} \quad x \in \Psi_p \quad \text{and} \quad m(x) < \min\{m(y) : y \in \Psi_p, 0 < \|y - x\| \le h\}.$$

It can be shown that Ψ_{hc} has the intensity $\lambda = (1 - \exp\{-\lambda_p \pi h^2\})/\pi h^2$ and the pair–correlation function

$$g(r) = \begin{cases} 0 & \text{if } r < h\,, \\ \frac{2G_h(r)(1-\exp(-\lambda_p\pi h^2)) - 2\pi h^2(1-\exp(-\lambda_p G_h(r)))}{\pi h^2 G_h(r)(G_h(r)-\pi h^2)\lambda^2} & \text{if } h \le r < 2\,h\,, \\ 1 & \text{if } r > 2\,h\,, \end{cases}$$

where $G_h(r) = 2h^2\big(\pi - \arccos(\frac{r}{2h}) + \frac{h}{2}\sqrt{4h^2 - r^2}\,\big)$.
Thus, $|\gamma^{(2)}_{\text{red}}|(B^c(o,2h)) = 0$ and in our case with $\lambda_p = 1$ and $h = 1/2$ we get

$$\gamma^{(2)}_{\text{red}}(B(o,1)) = -0.494, \quad \lambda = 0.693, \quad \sigma^2 = 0.350\,.$$

The point processes (i) – (iii) have intensity $\lambda = 1$ (i.e. the mean number of observed points in the above three windows equals 100, 400, and 1600, respectively) and the remaining point process (iv) has a slightly smaller intensity.

The Figures 1–4 present the obtained results of our simulation study only for the largest window W_{20}. The plotted curves in the below graphics show the behaviour of the (empirical) relative MSE

$$\text{rel MSE}(\cdot) = \text{MSE}(\cdot)/(\sigma^2)^2,$$

of the different estimators of σ^2 as function of the quantity $n\,b_n$. The solid line corresponds to $\widehat{\sigma}_n^2$ with cylinder kernel, the dashed line to $\widetilde{\sigma}_n^2$, the dotted line to $\widehat{\sigma}_n^2$ with the half-ball kernel, and the dash–dotted line to $\widehat{\sigma}_n^2$ with the cone kernel.

The complete set of plots for the three sizes of observation windows and two further point processes, including display of the variance and squared bias part of rel MSE the reader can find at
http://www.math.uni-augsburg.de/stochastik/heinrich/papers/asymvar.pdf.

To conclude with, all the estimators show good performance for the Poisson process (i) since there is no problem with the bias. For the Matérn (II) hard–core process (iv) we get similar results since $|\gamma^{(2)}_{\text{red}}|(B^c(o,1)) = 0$. In this case the estimators

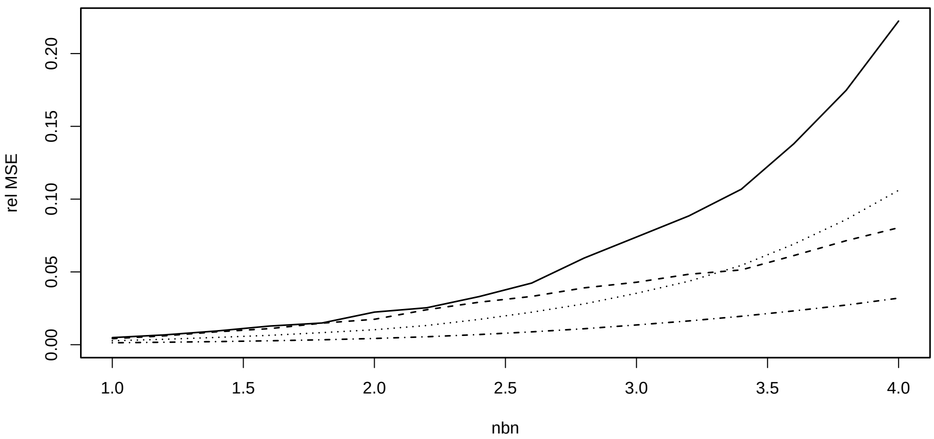

Figure 1: Performance of the estimators of σ^2 for the Poisson process (i) on W_{20}.

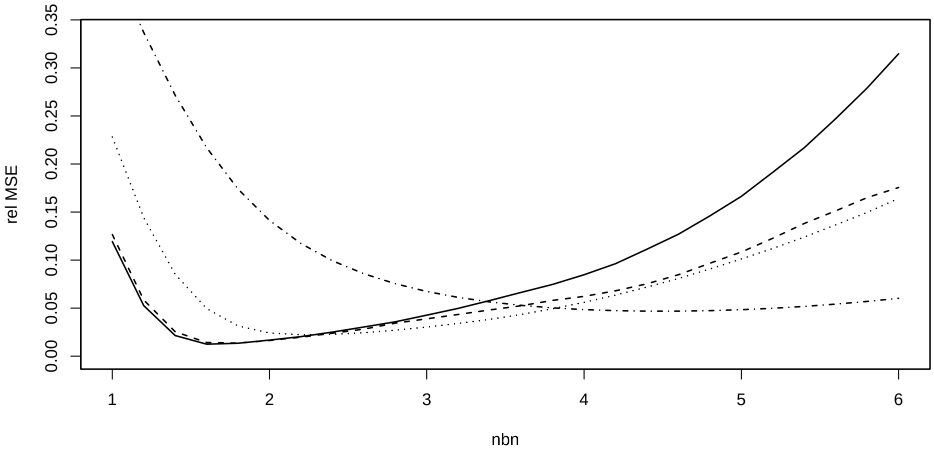

Figure 2: Performance of the estimators of σ^2 for the Matérn cluster process (ii) on W_{20}.

$\widetilde{\sigma}_n^2$ and $\widehat{\sigma}_n^2$ with the cylinder kernel (12) are unbiased even for $n\,b_n = 1$ and the variability of them is smaller because of more regularity within the point pattern.

The situation is quite different for the Poisson cluster processes (ii) and (iii) since $\inf\{r > 0 : |\gamma_{\mathrm{red}}^{(2)}|(B^c(o,r)) = 0\}$ is larger or even infinite and the bias courses more problems. The smallest possible values of rel MSE are obtained (in any of the models (ii)–(iv)) for $\widehat{\sigma}_n^2$ with cylinder kernel (12) and the isotropised estimator $\widetilde{\sigma}_n^2$ with a suitably chosen bandwidth.

However, looking on the most favouring values of rel MSE for the point processes (ii) and (iii) on W_5 and W_{10} , we must recognize that it is hardly possible to estimate σ^2 satisfactorily for small window sizes. Our estimators behave reasonably well only for larger windows like W_{20}.

Acknowledgements

This research was supported by DAAD which financed a three-month-stay of M. Prokešová at the University of Augsburg. M. Prokešová also acknowledges support from the Carlsberg foundation.

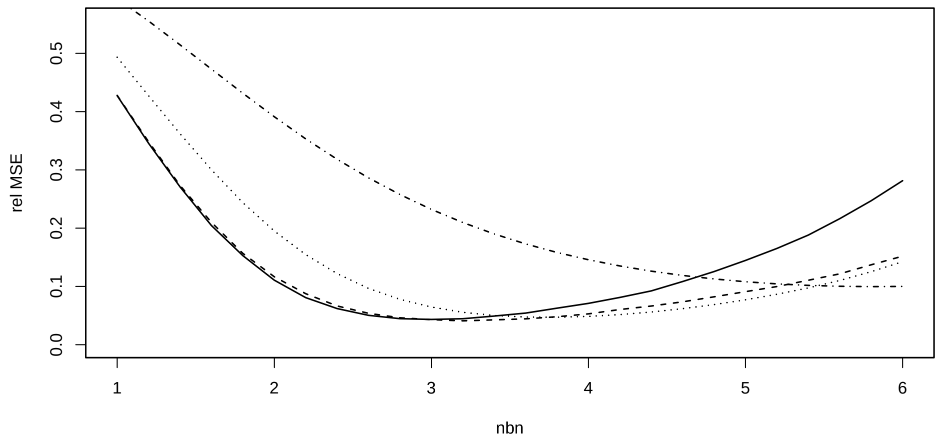

Figure 3: Performance of the estimators of σ^2 for the modified Thomas process (iii) on W_{20}.

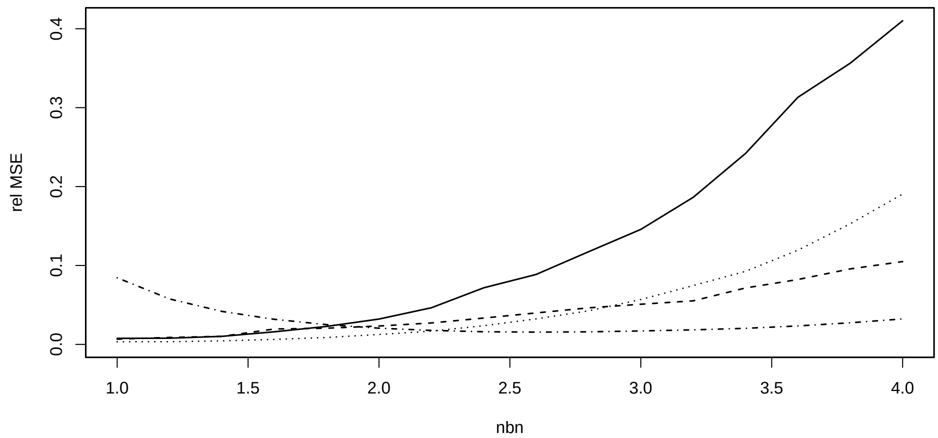

Figure 4: Performance of the estimators of σ^2 for the Matérn (II) hard–core process (iv) on W_{20}.

References

[1] S. Böhm, L. Heinrich and V. Schmidt, "Kernel estimation of the spectral density of stationary random closed sets," *Australian & New Zealand Journal of Statistics* vol. 46 pp. 41–52, 2004.

[2] D. J. Daley and D. Vere-Jones, *An introduction to the theory of point processes*, Springer: New York, 1988.

[3] L. Heinrich, "Normal approximation for some mean–value estimates of absolutely regular tesselations," *Mathematical Methods of Statistics* vol. 3 pp. 1–24, 1994.

[4] L. Heinrich and V. Schmidt, "Normal convergence of multidimensional shot noise and rates of this convergence," *Adv. Appl. Prob.* vol. 17 pp. 709–730, 1985.

[5] L. Heinrich and E. Liebscher, "Strong convergence of kernel estimators for product densities of absolutely regular point processes," *Journal of Nonparametric Statistics* vol. 8 pp. 65–96, 1997.

[6] L. Heinrich, "Asymptotic goodness–of–fit tests for point processes based on scaled empirical K–functions," *submitted*.

[7] E. Jolivet, "Central limit theorem and convergence of empirical processes of stationary point processes," in *Point Processes and Queueing Problems,* (edited by P.Bartfai and J.Tomko), North–Holland, pp. 117–161, 1980.

[8] K. Koch, J. Ohser and K. Schladitz, "Spectral theory for random closed sets and estimating the covariance via frequence space," *Adv. Appl. Prob.* vol. 35 pp. 603–613, 2003.

[9] S. Mase, "Asymptotic properties of stereological estimators for stationary random sets," *J. Appl. Prob.* vol. 19 pp. 111–126, 1982.

[10] J. Ohser and F. Mücklich, *Statistical analysis of microstructures in material science*, J. Wiley & Sons: Chichester, 2000.

[11] J. Ohser and D. Stoyan, " On the second-order and orientation analysis of planar stationary point processes," *Biom. J.*, vol. 23 pp. 523–533, 1981.

[12] B. D. Ripley, *Statistical inference for spatial processes*, Cambridge University Press: Cambridge, 1988.

[13] M. Stein, "Asymptotically optimal estimation for the reduced second moment measure of point processes," *Biometrika* vol. 80 pp. 443–449, 1993.

[14] D. Stoyan and H. Stoyan, "Improving ratio estimators of second order point process characeristics," *Scand. J. Statist.*, vol. 27 pp. 641–656, 2000.

[15] D. Stoyan, W. S. Kendall, J. Mecke, *Stochastic geometry and its applications*, 2^{nd} ed., J. Wiley & Sons: Chichester, 1995.

[16] J. C. Taylor, *An introduction to measure and probability*, Springer: New York, 1997.