KYBERNETIKA — VOLUME 47 (2011), NUMBER 5, PAGES 678–695

ESTIMATORS OF THE ASYMPTOTIC VARIANCE OF STATIONARY POINT PROCESSES – A COMPARISON

Michaela Prokešová

We investigate estimators of the asymptotic variance σ^2 of a d–dimensional stationary point process Ψ which can be observed in convex and compact sampling window $W_n = n\,W$. Asymptotic variance of Ψ is defined by the asymptotic relation $Var(\Psi(W_n)) \sim \sigma^2 |W_n|$ (as $n \to \infty$) and its existence is guaranteed whenever the corresponding reduced covariance measure $\gamma^{(2)}_{\mathrm{red}}(\cdot)$ has finite total variation. The three estimators discussed in the paper are the kernel estimator, the estimator based on the second order intesity of the point process and the subsampling estimator. We study the mean square consistency of the estimators. Since the expressions for the variance of the estimators are not available in closed form and depend on higher order moment measures of the point process, only the bias of the estimators can be compared theoretically. The second part of the paper is therefore devoted to a simulation study which compares the efficiency of the estimators by means of the mean squared error and for several clustered and repulsive point processes observed on middle-sized windows.

Keywords: reduced covariance measure, factorial moment and cumulant measures, kernel-type estimator, subsampling, mean squared error, Poisson cluster process, hard-core process

Classification: 60G55,62F12

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 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 two–volume monograph [3] supplemented by the monographs [18] and [11] where special emphasis is put on statistical analysis of point processes.

Throughout this paper we will denote by $\Psi = \sum_{i \geq 1} \delta_{X_i}$ a simple stationary 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 $\{X_1, X_2, \ldots\}$ defined on some common probability space $[\Omega, \mathcal{A}, \mathsf{P}]$. Since we assume the process to be simple – i.e. $\Psi(\{x\}) \leq 1$ for any location x almost surely, we can identify the "atoms" and "points". Therefore we will speak of "points of Ψ" instead of "atoms of Ψ" and write "$x \in \Psi$" instead of "$\Psi(\{x\}) > 0$".

This simplest numerical characteristic associated with Ψ is its intensity λ defined as the mean number of points of Ψ per unit volume $\lambda = \mathsf{E}\Psi([0,1)^d)$. It is standardly estimated by $\widehat{\lambda} = \Psi(W)/|W|$, where $W \subset \mathbb{R}^d$ denotes a bounded sampling window and $|W|$ its volume. We can investigate properties of $\widehat{\lambda}$ under the increasing domain asymptotics when we assume observing the process Ψ on a series of bounded (convex) windows $W_n \subset \mathbb{R}^d$ which are assumed to expand unboundedly in all directions as $n \to \infty$.

Under mild mixing conditions (expressible by the reduced covariance measure of Ψ, see Section 2) the limiting variance of $\widehat{\lambda}_n = \Psi(W_n)/|W_n|$ 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. [6, 10, 12]. 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 [7] and the work [4], such estimators are also needed for testing non–parametric point process hypotheses by using scaled empirical K–functions or integrated squared error of product density estimators, respectively. There are other fields of spatial statistics in which asymptotic variances and their estimation play an important role, see [1, 13].

The problem of estimation of the asymptotic variance was considered in [6], where a class of kernel estimators $\widehat{\sigma}_n^2$ for σ^2 was introduced and their L^2 consistency proved. In [9] the asymptotic results were refined by proving a central limit theorem for $\widehat{\sigma}_n^2$ and by obtaining the optimal convergence rates for $\widehat{\sigma}_n^2$ by a suitable choice of the bandwidth in dependence on the tails of the reduced covariance measure $\gamma^{(2)}_{\mathrm{red}}$. Nevertheless estimation of the asymptotic variance is not a simple task and a sufficiently large amount of data is needed for $\widehat{\sigma}_n^2$ to perform reasonably well. This was affirmed by the simulation study in [9] where the estimation procedure was applied to point processes of different kinds simulated on medium-sized windows. Alternative methods for estimating σ^2 were also mentioned in [9] but they were not investigated further.

The main aim of the present paper is to derive the asymptotic properties of the available estimators of σ^2 ($\widehat{\sigma}_n^2$, an alternative estimator from [9] and a subsampling estimator) and to compare their performance on medium-sized windows by means of a simulation study. The paper is organised as follows. After reviewing the necessary definitions and background knowledge in Section 2 we introduce the discussed estimators of σ^2 in Section 3 and discuss their consistency. After that Section 4 describes the design and the results of the simulation study.

2. BACKGROUND

Let us first recall the definitions and relations between factorial moment and factorial cumulant measures, see [3] 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

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

for any non–negative, Borel measurable function f on $(\mathbb{R}^d)^k$, where the sum $\sum^{\neq}$ runs over k–tuples of distinct points of Ψ. The first-order factorial moment measure $\alpha^{(1)}$ is called intensity measure and for a stationary (i. e. its distribution is invariant with respect to simultaneous translations of its points) point process Ψ it holds $\alpha^{(1)}(A) = \lambda|A|$ for any $A \in \mathcal{B}^d$.

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 $\alpha^{(1)},\dots,\alpha^{(k)}$ by

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

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

$$\gamma^{(2)}(A_1\times A_2) = \alpha^{(2)}(A_1\times A_2) - \alpha^{(1)}(A_1)\,\alpha^{(1)}(A_2) \quad \text{for} \quad A_1, A_2 \in \mathcal{B}^d .$$

The second order factorial cumulant measure $\gamma^{(2)}$ is also called covariance measure because $\mathsf{Cov}(\Psi(A_1),\Psi(A_2)) = \gamma^{(2)}(A_1\times A_2)$ for any disjoint $A_1, A_2 \in \mathcal{B}^d$.

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

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

where f is as in (2). In the same way we may define the reduced kth-order factorial cumulant measure $\gamma^{(k)}_{red}$ which turns out to be a signed measure on $[(\mathbb{R}^d)^{k-1}, \mathcal{B}^{d(k-1)}]$ with the Jordan decomposition $\gamma^{(k)}_{red} = (\gamma^{(k)}_{red})^+ - (\gamma^{(k)}_{red})^-$, see e. g.[19] for details. The corresponding total variation measure $|\gamma^{(k)}_{red}| = (\gamma^{(k)}_{red})^+ + (\gamma^{(k)}_{red})^-$ 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 have $\gamma^{(2)}(\cdot \times \cdot) = \alpha^{(2)}(\cdot \times \cdot) - \lambda^2 \, |\cdot|\,|\cdot|$ and thus we get $\gamma^{(2)}_{red}(\cdot) = \alpha^{(2)}_{red}(\cdot) - \lambda\,|\cdot|$ and call $\gamma^{(2)}_{red}$ the reduced covariance measure of Ψ. The variance $\mathsf{Var}(\Psi(W_n))$ can be expressed by means of this reduced covariance measure which together with (1) leads to

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

whenever W_n increases unboundedly in all directions and $|\gamma^{(2)}_{red}|(\mathbb{R}^d) < \infty$. 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 $\alpha^{(2)}_{red}$ (if it exists) is called the second–order product density of Ψ. Further, if Ψ is also isotropic – i. e. its distribution is invariant under rotations, 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

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

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

There is another popular point process characteristic which is closely related with $\gamma^{(2)}_{\mathrm{red}}$. Namely the K-function

$$K(r) = \frac{1}{\lambda} \left(\gamma^{(2)}_{\mathrm{red}}(B(o,r)) + \kappa_d r^d \right) , \tag{5}$$

where $B(o,r)$ denotes the ball centered in the origin o with radius r. Estimators of the pair-correlation function and the K-function are well investigated (see e. g. [18]) and both the relations (4) and (5) can be used when defining estimators of σ^2 as we will see in the sequel.

The theoretical values of σ^2 can be easily obtained for a large class of models, particularly for any models for which the pair-correlation function or the K-function is known. Examples of such processes can be found among the processes used for the simulation study in Section 4. The benchmark value of $\sigma^2 = \lambda$ is obtained for the Poisson process for which the complete spatial randomness of this model implies that $\gamma^{(k)} \equiv 0$ for any k larger than 1.

3. ESTIMATORS OF THE ASYMPTOTIC VARIANCE

3.1. Kernel type estimator

The first estimator of asymptotic variance was defined in [6]. Let $w : \mathbb{R}^d \to [0, \infty)$ be a kernel function which is symmetric, bounded, continuous at the origin $o \in \mathbb{R}^d$ and satisfies $w(o) = 1$, and let $b > 0$ be a bandwidth. The kernel estimator of σ^2 is defined by

$$\widehat{\sigma^2} = \widehat{\lambda} + \sum_{x,y \in \Psi}^{\neq} \frac{w((y-x)/b) \mathbf{1}_W(x) \mathbf{1}_W(y)}{|(W-x) \cap (W-y)|} \; - \; \omega\,(b)^d \, \widehat{\lambda^2} \, , \tag{6}$$

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

L^2-consistency of the estimator was proved in [6] under an increasing domain asymptotics. Thus let us assume that the point process Ψ is observed on a series of windows $\{W_n\}_{n\in\mathbb{N}}$ satisfying assumption:

(A1) The sequence of sampling windows satisfies $W_n = n\,W$ for $n \geq 1$, for some convex and compact $W \in \mathcal{B}^d$ fulfilling $B(o,\epsilon) \subset W$ for some $\epsilon > 0$.

Further we need a sequence of bandwidths which increase to infinity as well but they also should become incresingly small with respect to the size of the observation window. Thus the second assumption is:

(A2) The (positive) sequence of bandwidths (b_n) satisfies $1 \geq b_n \underset{n\to\infty}{\longrightarrow} 0$ and $b_n n \underset{n\to\infty}{\longrightarrow} \infty$.

As we already mentioned in Section 1 following condition ensures the existence of σ^2:

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

Finally a moment assumption is needed for the L^2-consistency:

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

Theorem 3.1. (Heinrich [6]) Let $\widehat{\sigma}^2_n$ be the estimator defined in (6) with $W = W_n$, $b = b_n\, n$ and
$$\widehat{\lambda^2} = (\widehat{\lambda^2})_n = \frac{\Psi(W_n)(\Psi(W_n)-1)}{|W_n|^2}. \tag{7}$$
Under the assumptions (A1) – (A3), the sequence of estimators $(\widehat{\sigma}^2_n)$ is asymptotically unbiased for σ^2, that is $\mathsf{E}\,\widehat{\sigma}^2_n \underset{n\to\infty}{\longrightarrow} \sigma^2$.
Under the additional assumptions (A4) and $b_n^2 n \underset{n\to\infty}{\longrightarrow} 0$ the sequence $(\widehat{\sigma}^2_n)$ is mean square consistent, that is $\mathrm{MSE}(\widehat{\sigma}^2_n) := \mathsf{E}\,(\widehat{\sigma}^2_n - \sigma^2)^2 \underset{n\to\infty}{\longrightarrow} 0$.

From the proof of the theorem in [6] we can moreover get bounds on the bias of $\widehat{\sigma}^2_n$ since
$$\mathsf{E}\,\widehat{\sigma}^2_n - \sigma^2 = \lambda \int_{\mathbb{R}^d} \Big(w\Big(\frac{x}{b_n\, n}\Big) - 1\Big)\gamma^{(2)}_{\mathrm{red}}(\mathrm{d}x) - \frac{\omega\,(b_n\, n)^d\,\lambda}{|W_n|^2}\int_{\mathbb{R}^d} |\,W_n \cap (W_n - y)\,|\,\gamma^{(2)}_{\mathrm{red}}(\mathrm{d}y)\,.$$

Thus the bias can be bounded from above as follows:
$$|\,\mathsf{E}\,\widehat{\sigma}^2_n - \sigma^2\,| \leq \lambda \Big| \int_{\mathbb{R}^d} \Big(w\Big(\frac{x}{b_n\, n}\Big) - 1\Big)\,\gamma^{(2)}_{\mathrm{red}}(\mathrm{d}x)\Big| + \Big(\frac{b_n}{2}\Big)^d \omega\,\lambda\,\|\gamma^{(2)}_{\mathrm{red}}\|_{\mathrm{var}}\,. \tag{8}$$

Let us remark here that for the special case of the cylinder kernel $w(x) = \mathbf{1}_{B(o,1)}(x)$ in $\mathbb{R}^2$ we get that $\widehat{\sigma}_n^2 = \widehat{\lambda_n} + \widehat{\lambda^2 K(n\,b_n)} - \pi(n\,b_n)^2(\widehat{\lambda^2})_n$, where $\widehat{\lambda^2 K(n\,b_n)}$ is the classical edge-corrected estimator of λ^2 times the K-function (see e.g. [18]). Thus we have an analogy of the relation (5). The leading term in the bias bound (8) is then equal to $\lambda\,\gamma_{\mathrm{red}}^{(2)}(B^c(o,n\,b_n))$, i.e. it is proportional to the tails of $\gamma_{\mathrm{red}}^{(2)}$.

3.2. Estimator based on second order intensity of the point process

The starting point for defining the next estimator of σ^2 is the formula (4) which yields

$$\sigma^2 = \lambda + \int_{\mathbb{R}^d} (\,\lambda\,\varrho^{(2)}(x) - \lambda^2\,)\,\mathrm{d}x. \tag{9}$$

Let $\widehat{(\lambda\,\varrho^{(2)})}_n(x)$ denote an appropriate edge–corrected kernel–type estimator for $\lambda\,\varrho^{(2)}(x)$ defined on the sampling window $W_n = nW$. Then we can plug this estimator into the formula (9) together with the estimators $\widehat{\lambda}_n$ and $(\widehat{\lambda^2})_n$ for λ and λ^2, respectively, to obtain the estimator

$$\widehat{\sigma}_{n,I}^2 = \widehat{\lambda}_n + \int_{B(o,n\,b_n)} \Big(\,\widehat{(\lambda\,\varrho^{(2)})}_n(x)\; - \;(\widehat{\lambda^2})_n\,\Big)\,\mathrm{d}x\;, \tag{10}$$

which was introduced in [9]. Here $B(o,n\,b_n)$ is the ball with center in the origin o and radius $n\,b_n$.

The appropriate estimator for $\lambda\,\varrho^{(2)}(x)$ can be defined by

$$\widehat{(\lambda\,\varrho^{(2)})}_n(x) = \frac{1}{b_n^d}\sum_{u,v\in\Psi}^{\neq} \frac{\mathbf{1}_{W_n}(u)\,\mathbf{1}_{W_n}(v)}{|\,(W_n-u)\cap(W_n-v)\,|}\;k\Big(\frac{v-u-x}{b_n}\Big)\;, \tag{11}$$

(see [18]) where the kernel function $k\,:\,\mathbb{R}^d \mapsto \mathbb{R}^1$ is assumed to be bounded with bounded support such that $\int_{\mathbb{R}^d} k(x)\,\mathrm{d}x = 1$ and the sequence of bandwidths (b_n) satisfies (A2).

Asymptotic properties of $\widehat{\sigma}_{n,I}^2$ are of course in close relations with asymptotic properties of $(\lambda\,\varrho^{(2)})_n(x)$. These were studied in [5] where a central limit theorem was proved for the case of Poisson cluster processes, and in [8] where almost sure convergence of $(\lambda\,\varrho^{(2)})_n(x)$ was proved in the setting of β-mixing. In the recent work [4] the central limit theorem for finite dimensional vectors $((\lambda\,\varrho^{(2)})_n(x_i))_{i=1,\dots k}$ was proved for the general class of Brillinger mixing processes under mild mixing conditions. Nevertheless these results for the pointwise convergence of $(\lambda\,\varrho^{(2)})_n(x)$ are not directly applicable for deriving the properties of the integral $\int_{B(o,b_n\,n)} \widehat{(\lambda\,\varrho^{(2)})}_n(x)\,\mathrm{d}x$. Even the uniform rates of convergence from [8] are not directly useful because they are obtained for a fixed set $K \subset \mathbb{R}^d$ and we have an expanding set $B(o,b_n\,n)$. Thus the results for $\widehat{\sigma}_{n,I}^2$ must be obtained directly. Let us prove here the asymptotic unbiasedness of $\widehat{\sigma}_{n,I}^2$.

Theorem 3.2. Let the kernel k be bounded with bounded support, symmetric and such that $\int_{\mathbb{R}^d} k(x)\,\mathrm{d}x = 1$. Under the assumptions (A1) – (A3), the sequence of estimators $(\widehat{\sigma}_{n,I}^2)$ is asymptotically unbiased for σ^2.

Proof. Let us rewrite

$$\mathsf{E}\,\widehat{\sigma}^2_{n,I} = \lambda + \mathsf{E}\int_{B(o,n\,b_n)} \left((\widehat{\lambda\,\varrho^{(2)}})_n(x) + \lambda^2 \right) \mathrm{d}x - \mathsf{E}\int_{B(o,n\,b_n)} \left(\lambda^2 - (\widehat{\lambda^2})_n \right) \mathrm{d}x\,. \quad (12)$$

By Fubini's theorem and the formula $\mathsf{E}\,(\Psi(W_n))^2 = \alpha^{(2)}(W_n \times W_n) + \lambda|W_n|$ the third term can be rewritten as

$$\int_{B(o,n\,b_n)} \frac{\gamma^{(2)}(W_n \times W_n)}{|W_n|^2}\,\mathrm{d}x \le \lambda \int_{B(o,n\,b_n)} \frac{\gamma^{(2)}_{\mathrm{red}}(\mathbb{R}^d)}{|W_n|}\,\mathrm{d}x = \mathcal{O}(b_n^d),$$

which goes to 0 for $n \to \infty$. Using Fubini's theorem on the second term from (12) we can rewrite it as

$$\begin{aligned}
&\int_{B(o,n\,b_n)} \frac{1}{b_n^d}\left[\int_{W_n}\int_{W_n} \frac{k(\frac{v-u-x}{b_n})}{|(W_n-u)\cap(W_n-v)|}\gamma^{(2)}(\mathrm{d}u,\mathrm{d}v)\right]\mathrm{d}x \\
+&\int_{B(o,n\,b_n)} \frac{1}{b_n^d}\left[\int_{W_n}\int_{W_n} \frac{k(\frac{v-u-x}{b_n})}{|(W_n-u)\cap(W_n-v)|}\lambda^2\,\mathrm{d}u\,\mathrm{d}v - \lambda^2 b_n^d\right]\mathrm{d}x = T_1 + T_2.
\end{aligned} \quad (13)$$

By change of variables we get for T_2

$$\begin{aligned}
T_2 &= \int_{B(o,n\,b_n)} \frac{1}{b_n^d}\left[\int_{W_n\oplus -W_n}\int_{(W_n-\delta)\cap W_n} \frac{k(\frac{\delta-x}{b_n})}{|(W_n-\delta)\cap W_n|}\lambda^2\,\mathrm{d}u\,\mathrm{d}\delta - \lambda^2 b_n^d\right]\mathrm{d}x \\
&= \lambda^2\int_{B(o,n\,b_n)}\left[\int_{W_n\oplus -W_n}\frac{1}{b_n^d}k\left(\frac{\delta-x}{b_n}\right)\mathrm{d}\delta - 1\right]\mathrm{d}x \\
&= \lambda^2\int_{B(o,n\,b_n)}\left[\int_{W_{\frac{n}{b_n}}\oplus -W_{\frac{n}{b_n}}} k\left(z-\frac{x}{b_n}\right)\mathrm{d}z - 1\right]\mathrm{d}x \\
&\le \lambda^2\int_{B(o,n\,b_n)}\left[\int_{B(o,\frac{n}{b_n}(2\rho-b_n))} k(z)\,\mathrm{d}z - 1\right]\mathrm{d}x
\end{aligned}$$

where the symbol $\oplus$ denotes the Minkovski addition and ρ the inradius of W. Since we assumed k to have bounded support there exists n large enough such that $\int_{B(o,\frac{n}{b_n}(2\rho-b_n))} k(z)\,\mathrm{d}z = 1$ and $T_2 = 0$.

For T_1 we can disintegrate $\gamma^{(2)}$ and write

$$\begin{aligned}
T_1 &= \int_{B(o,n\,b_n)} \frac{1}{b_n^d}\left[\int_{W_n\oplus -W_n}\int_{(W_n-\delta)\cap W_n} \frac{k(\frac{\delta-x}{b_n})}{|(W_n-\delta)\cap W_n|}\lambda\,\mathrm{d}u\,\gamma^{(2)}_{\mathrm{red}}(\mathrm{d}\delta)\right]\mathrm{d}x \\
&= \int_{B(o,n\,b_n)} \frac{1}{b_n^d}\left[\int_{W_n\oplus -W_n} k\left(\frac{\delta-x}{b_n}\right)\lambda\,\gamma^{(2)}_{\mathrm{red}}(\mathrm{d}\delta)\right]\mathrm{d}x \\
&= \lambda\int_{W_n\oplus -W_n}\int_{B(o,n)} k\left(\frac{\delta}{b_n}-z\right)\mathrm{d}z\,\gamma^{(2)}_{\mathrm{red}}(\mathrm{d}\delta).
\end{aligned} \quad (14)$$

The integral $\int_{B(o,n)} k\left(\frac{\delta}{b_n} - z\right) \mathrm{d}z$ goes monotonically to 0 when n goes to infinity for any δ and $W_n \oplus -W_n$ converges to $\mathbb{R}^d$ as $n \to \infty$. From (A3) it follows that the term T_1 converges to $\lambda\, \gamma^{(2)}_{\mathrm{red}}(\mathbb{R}^d)$ and thus (12) converges to σ^2.

□

When we compare the estimators $\widehat{\sigma}^2_n$ and $\widehat{\sigma}^2_{n,I}$ heuristically we see the main difference in the fact, that in case of $\widehat{\sigma}^2_{n,I}$ we first estimate the density of $\gamma^{(2)}_{\mathrm{red}}$ and then integrate it over a domain increasing to $\mathbb{R}^d$ whereas by $\widehat{\sigma}^2_n$ we try to estimate the integral $\gamma^{(2)}_{\mathrm{red}}(\mathbb{R}^d)$ directly. Thus the procedure for $\widehat{\sigma}^2_{n,I}$ introduces some extra smoothing which can potentially lead to some extra bias – compare the leading terms in the formulas (8) and (14). From the practical point of view the question is if this extra bias is noticeable or negligeable – and this is hard to decide theoretically. The simulation results in Section 4 show that the bias is negligeable for a good choice of the bandwidth $n\, b_n$.

3.3. Subsampling estimator

Subsampling is a popular approach for estimating variances of various statistics of interest (see [14]). However in spatial statistics this approach is not used very often. Among the reasons are the facts that typically we have only one realization of the point process at our disposal and (with the exception of the Poisson process) a complicated dependence structure in the data. Nevertheless since we are interested in the variance of a very simple statistics $\widehat{\lambda}$ and both the above mentioned estimators $\widehat{\sigma}^2_n$ and $\widehat{\sigma}^2_{n,I}$ need anyway a reasonably large amount of data to perform well, the subsampling approach could lead to a competitive alternative for estimating σ^2. In the sequel we define a moving-block variance estimator $\widehat{\sigma}^2_{n,V}$ inspired by the approach from [15].

Let (b_n) be again a sequence of bandwidths fulfilling (A2) and denote $V_n = W_{n\, b_n} = n\, b_n W$ a scaled version of the observation window W. For the estimation of σ^2 we will use estimates of the intensity $\widehat{\lambda}_{[V_n+y]} = \frac{\Psi(V_n+y)}{|V_n|}$ computed on the subwindows $V_n + y$, where $y \in W_n^{1-b_n} = \{y \in \mathbb{R}^d : V_n + y \subset W_n\}$. The subsampling estimator of σ^2 is defined as follows:

$$\widehat{\sigma}^2_{n,V} = \int_{W_n^{1-b_n}} |V_n| \left(\widehat{\lambda}_{[V_n+y]} - \overline{\widehat{\lambda}_{V_n}}\right)^2 \mathrm{d}y / |W_n^{1-b_n}|, \tag{15}$$

where $\overline{\widehat{\lambda}_{V_n}} = \int_{W_n^{1-b_n}} \widehat{\lambda}_{[V_n+y]}\, \mathrm{d}y / |W_n^{1-b_n}|$.

First let us show that the estimator is asymptotically unbiased.

Theorem 3.3. Under the assumptions (A1) – (A3) the subsampling estimator (15) is asymptotically unbiased for σ^2.

Proof. Obviously

$$\mathsf{E}\,\widehat{\sigma}^2_{n,V} = \frac{\mathsf{Var}(\Psi(V_n))}{|V_n|} + \int_{W_n^{1-b_n}} \int_{W_n^{1-b_n}} \frac{|V_n|}{|W_n^{1-b_n}|^2} \mathsf{Cov}(\widehat{\lambda}_{V_n+y}, \widehat{\lambda}_{V_n+z})\, \mathrm{d}y\, \mathrm{d}z.$$

The first term converges to σ^2 from our assumptions since the convex averaging sequence $V_n \to \mathbb{R}^d$. The second term is equal to

$$\int_{W_n^{1-b_n}} \int_{W_n^{1-b_n}} \frac{1}{|W_n^{1-b_n}|^2|V_n|} (\gamma^{(2)}((V_n+y)\times(V_n+z)) + \lambda|(V_n+y)\times(V_n+z)|)\,\mathrm{d}y\,\mathrm{d}z,$$

and the second summand can be bounded by $\frac{2^d|V_n|}{|W_n^{(1-b_n)}|}$ which goes to 0 by (A2) and the convexity of W. We can rewrite the first summand as

$$\frac{1}{|W_n^{1-b_n}|^2|V_n|} \int_{W_n^{1-b_n}} \int_{W_n^{1-b_n}} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \lambda \mathbf{1}_{(u,u+v)\in V_n\times(V_n-y+z)}\,\mathrm{d}u\gamma_{\mathrm{red}}^{(2)}(\mathrm{d}v)\,\mathrm{d}y\,\mathrm{d}z.$$

By changing the order of integration and starting with z we get the upper bound $\frac{2^d|V_n|}{|W_n^{(1-b_n)}|}\|\gamma_{\mathrm{red}}^{(2)}\|_{\mathrm{var}}$ which goes to 0 by (A2) and (A3).

□

If we want to compare the bias of $\widehat{\sigma}_n^2$ and the subsampling estimator $\widehat{\sigma}_{n,V}^2$ we have to take into account the shape of the kernel w for $\widehat{\sigma}_n^2$. To simplify the situation let us assume the cylinder kernel $w(x) = \mathbf{1}_{B(o,1)}(x)$ for $\widehat{\sigma}_n^2$. Then the leading term in the bias for $\widehat{\sigma}_n^2$ will be equal to $\lambda\,\gamma_{\mathrm{red}}^{(2)}(B^c(o,n\,b_n)$ whereas for $\widehat{\sigma}_{n,V}^2$ it will be $\lambda\,\gamma_{\mathrm{red}}^{(2)}(\mathbb{R}^d) - \frac{\gamma^{(2)}(V_n\times V_n)}{|V_n|}$. Thus the subsampling estimator covers less of the mass of $\gamma_{\mathrm{red}}^{(2)}$ than $\widehat{\sigma}_n^2$ and should have larger bias for the same value of $n\,b_n$. This observation is confirmed by the results of the simulation study in Section 4.

Concerning the mean square consistency of $\widehat{\sigma}_{n,V}^2$ this is harder to prove. It is possible to proceed by analogy to the methods in [15] and prove L^2-consistency under the assumption of strong mixing.

Recall that for two σ-algebras $\mathcal{F}_1$, $\mathcal{F}_2$ defined on the same probability space the strong mixing coefficient is defined by

$$\alpha(\mathcal{F}_1,\mathcal{F}_2) = \sup\{|P(A_1\cap A_2) - P(A_1)P(A_2)| : A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\}. \qquad (16)$$

For a stationary point process Ψ the strong mixing coefficient $\alpha(p;k)$ quantifies the dependence between the behaviour of the point process on sets of volume at most p separated by a distance larger than or equal to k. Thus for a point process we define

$$\alpha(p;k) = \sup\{\alpha(\mathcal{F}^X(A),\mathcal{F}^X(B)) : d(A,B) \geq k, |A| \leq p, |B| \leq p\}, \qquad p \geq 0$$

where $d(A,B) = \inf\{|x-y| : x\in A, y\in B\}$ is the distance between the sets A and B, $\mathcal{F}^X(A)$ denotes the σ-algebra generated by $\Psi\cap A$ and the supremum is taken over all measurable subsets A, B in $\mathcal{B}^d$.

The assumption of strong mixing enables to find bounds on the integrals of the covariances, needed in the proof of the following theorem.

Theorem 3.4. Let (A1) – (A3) be fullfiled and assume

(A5) there exist $\epsilon > 0$ and $C_\epsilon > 0$ such that $\mathsf{E}\left(\frac{\Psi(V_n)-\lambda|V_n|}{\sqrt{|V_n|}}\right)^{4+\epsilon} < C_\epsilon$ for any n.

Moreover assume that the point process Ψ is strongly mixing – i.e. $\alpha(p, p^d) \to 0$ as $p \to \infty$. Then the subsampling estimator $\widehat{\sigma}^2_{n,V}$ is mean square consistent.

P r o o f. We can rewrite $\widehat{\sigma}^2_{n,V}$ as

$$\begin{aligned}
\widehat{\sigma}^2_{n,V} &= \frac{1}{|W_n^{1-b_n}|} \int_{|W_n^{1-b_n}|} \left[\sqrt{|V_n|}\left(\widehat{\lambda}_{V_n+y} - \mathsf{E}\widehat{\lambda}_{V_n+y}\right)\right]^2 \mathrm{d}y \\
&\quad - \frac{1}{|W_n^{1-b_n}|^2} \left[\int_{|W_n^{1-b_n}|} \sqrt{|V_n|}\left(\widehat{\lambda}_{V_n+y} - \mathsf{E}\widehat{\lambda}_{V_n+y}\right) \mathrm{d}y\right]^2 \\
&= \frac{1}{|W_n^{1-b_n}|} \int_{|W_n^{1-b_n}|} [h(V_n+y)]^2 \,\mathrm{d}y - \frac{1}{|W_n^{1-b_n}|^2} \left[\int_{|W_n^{1-b_n}|} h(V_n+y)\,\mathrm{d}y\right]^2,
\end{aligned}$$

where we denote $h(V_n+y) = \sqrt{|V_n|}\left(\widehat{\lambda}_{V_n+y} - \mathsf{E}\widehat{\lambda}_{V_n+y}\right)$.

We will proceed analogously to the general proof of Theorem 2 in [15]. Let us moreover denote $H(V_n+y) = |V_n|\left(\widehat{\lambda}_{V_n+y} - \mathsf{E}\widehat{\lambda}_{V_n+y}\right)^2$. We will start by showing the following two convergencies:

$$Y_n = \frac{\int_{|W_n^{1-b_n}|} h(V_n+y)\,\mathrm{d}y}{|W_n^{1-b_n}|} \xrightarrow{L^2} \lim_{n\to\infty} \mathsf{E}\,h(W_n) = 0, \tag{17}$$

and

$$U_n = \frac{\int_{|W_n^{1-b_n}|} H(V_n+y)\,\mathrm{d}y}{|W_n^{1-b_n}|} \xrightarrow{L^2} \lim_{n\to\infty} \mathsf{E}\,H(W_n) = \lim_{n\to\infty} \frac{\mathsf{Var}\,(\Psi(W_n))}{|W_n|} = \sigma^2. \tag{18}$$

For both Y_n and U_n we have that $\mathsf{E}\,Y_n = \mathsf{E}\,h(V_n)$ and $\mathsf{E}\,U_n = \mathsf{E}\,H(V_n)$ by Fubini and the stationarity of Ψ. Thus it is enough to show that both $\mathsf{Var}\,Y_n$ and $\mathsf{Var}\,U_n$ go to zero to obtain (17) and (18). By using Fubini's theorem again we get

$$\mathsf{Var}\,Y_n = \int_{W_n^{1-b_n}} \int_{W_n^{1-b_n}} \frac{1}{|W_n^{1-b_n}|^2} \mathsf{Cov}(h(V_n+y), h(V_n+z))\,\mathrm{d}y\,\mathrm{d}z,$$

and analogous expression for U_n and H. For h we have proved that this expression goes to 0 with $n \to \infty$ in the proof of Theorem 3.3. For the case of H we will proceed analogously to the proof of Theorem 1 in [15].

We divide the integral into two parts:

$$\begin{aligned}
\mathsf{Var}\,U_n &= \int\limits_{W_n^{1-b_n}} \int\limits_{\substack{W_n^{1-b_n} \\ d(V_n+y, V_n+z) \le p}} \frac{1}{|W_n^{1-b_n}|^2} \mathsf{Cov}(H(V_n+y), H(V_n+z))\,\mathrm{d}y\,\mathrm{d}z \\
&\quad + \int\limits_{W_n^{1-b_n}} \int\limits_{\substack{W_n^{1-b_n} \\ d(V_n+y, V_n+z) > p}} \frac{1}{|W_n^{1-b_n}|^2} \mathsf{Cov}(H(V_n+y), H(V_n+z))\,\mathrm{d}y\,\mathrm{d}z \\
&= T_1 + T_2.
\end{aligned}$$

Now, let $p = |V_n|^{1/d}$ and denote $r = \sup\{|x-y|, x, y \in W\}$, the size of W. Then $|\{z : d(V_n+y, V_n+z) \le p\}|$ is for any fixed y bounded from above by $\kappa_d(3n\,b_n r+2l)^d$ and thus

$$T_1 \le \int_{W_n^{1-b_n}} \frac{\mathsf{Var}(H(V_n))\kappa_d(3n\,b_n r+2l)^d}{|W_n^{1-b_n}|^2}$$
$$\le \frac{C_\epsilon^{\frac{\epsilon}{4+\epsilon}}\kappa_d 2^d((3r\,n\,b_n)^d + 2^d(n\,b_n)^d|W|)}{|W_n^{1-b_n}|} = \mathcal{O}(b_n^d), \quad (19)$$

where we used Jensen's inequality and the assumption (A5).

For T_2 first observe that under the assumption $d(V_n+y, V_n+z) > p$ we have

$$\mathsf{Cov}(H(V_n+y), H(V_n+z)) = \mathsf{Var}(H(V_n))\mathsf{Corr}(H(V_n+y), H(V_n+z)) \le C_\epsilon^{\frac{\epsilon}{4+\epsilon}}\,4\,\alpha(p, |V_n|),$$

using again Jensen's inequality and assumption (A5) and by Lemma 1.2.1 in [20]. Thus

$$T_2 \le 4C_\epsilon^{\frac{\epsilon}{4+\epsilon}}\,\alpha(p, |V_n|) = 4C_\epsilon^{\frac{\epsilon}{4+\epsilon}}\,\alpha(p, p^d),$$

which converges to 0 by our mixing assumption.

Now comming back to the expression for $\widehat{\sigma}^2_{n,V}$ we see that $\widehat{\sigma}^2_{n,V} = U_n + Y_n^2$. U_n converges to σ^2 in L^2 and thus we only need to show that $Y_n^2 \xrightarrow{L^2} 0$, or equivalently $Y_n \xrightarrow{L^4} 0$.

This is proved exactly like in the proof of Theorem 2 in [15] from $Y_n \xrightarrow{L^2} 0$, $U_n \xrightarrow{L^2} \sigma^2$, $\mathsf{E}Y_n^4 \le \mathsf{E}U_n^2$ by a lemma about uniform integrability and L^r convergence from [2]. □

The assumption of strong mixing is actually not such a strong assumption and for example all the processes we used in the simulation study satisfy it (see e. g. [16] chapter 5 for a detailed discussion). Moreover it seems plausible (see [4] chapter 7) that the L^2 consistency of the subsampling estimator $\widehat{\sigma}^2_{n,V}$ can be proved without strong mixing only under suitable integrability assumptions on the higher order reduced cumulant measures. Nevertheless the proof of such a proposition would be highly technical (compare [4] chapter 7) and out of the scope of the current paper.

Concerning the practical implementation of the subsampling estimator the integrals in the definition of $\widehat{\sigma}^2_{n,V}$ must be of course approximated by finite Riemann sums resulting in

$$\widehat{\sigma}^2_{n,V} = \sum_{y_k \in W_n^{1-b_n}\cap G} \delta^d\,|V_n|\left(\widehat{\lambda}_{[V_n+y_k]} - \overline{\widehat{\lambda}_{V_n}}\right)^2 / |W_n^{1-b_n}|, \quad (20)$$

where $\overline{\widehat{\lambda}_{V_n}}$ is also approximated as follows

$$\overline{\widehat{\lambda}_{V_n}} = \left(\sum_{y_k \in W_n^{1-b_n}\cap G} \delta^d\,\widehat{\lambda}_{[V_n+y_k]}\right) / |W_n^{1-b_n}|.$$

Here G denotes a rectangular grid of points $\frac{\delta}{2}\mathbf{1} + \delta\mathbb{Z}^d$. Under the assumptions of Theorem 3.4 and if $n, \frac{1}{\delta} \to \infty$ the estimator (20) is also an L^2-consistent estimator of σ^2.

In practice we must of course decide the value of δ which may influence the behaviour of the approximated estimator $\widehat{\sigma}^2_{n,V}$. Nevertheless the simulation experiments indicate that if the value of δ is reasonably small (see the next section for details) then the efficiency of $\widehat{\sigma}^2_{n,V}$ is practically not influenced by the exact choice of δ – at least for the range of reasonable choices of the tuning parameter b.

4. SIMULATION STUDY

To compare the efficiency of the three above introduced estimators on medium-sized windows a simulation study was carried out. We considered several specific stationary point processes in $\mathbb{R}^2$ some of them exibiting clustering among the points some of them repulsion (they will be described in detail later on). Since the simulation study in [9] showed qualitatively similar behaviour of the estimator $\widehat{\sigma}^2_n$ on windows of various sizes and a reasonable performance only for large enough windows we decided to use the window $W_{20} = [-20, 20]^2$ for processes of (approximately) unit intensity.

For the simulated processes values of the three different estimators of σ^2 were computed for the same generated point patterns:

— $\widehat{\sigma}^2_n$ with the cylinder kernel $w(x) = \mathbf{1}_{B(o,1)}(x)$ (black colour in the graphs)

— $\widehat{\sigma}^2_{n,I}$ with the cylinder kernel $k(x) = \frac{1}{\pi}\mathbf{1}_{B(o,1)}(x)$ (green colour in the graphs)

— $\widehat{\sigma}^2_{n,V}$ where the integral was approximated by the Riemann sum (20) over subsquares with centers in a regular grid with δ equal to 0.2 (red colour in the graphs).

We have chosen the cylinder kernel $\mathbf{1}_{B(o,1)}$ for computational simplicity and because it proved to be optimal for $\widehat{\sigma}^2_n$ according to the simulation study in [9].

All estimators were computed for a series of bandwidths $b = n\,b_n$ the range of which depended on the particular point processes used for the estimation.

The squared bias, the variance and the MSE were estimated for each of the three estimators from 1000 realizations of the simulated point processes.

Before coming to the results we briefly describe the point process models used and compute their σ^2 as well as their reduced covariance mesure $\gamma^{(2)}_{\mathrm{red}}$; more information on these models can be found e. g. in [18].

1. Poisson process with intensity $\lambda = 1$ $\quad(\gamma^{(2)}_{\mathrm{red}} \equiv 0$ and $\sigma^2 = \lambda = 1)$.

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

   This Poisson cluster process is generated by a stationary Poisson process of parent points with intensity μ^{-1}; the typical cluster consists of a Poisson distributed number of daughter points with locations independently and uniformly distributed on the disk $B(o,r)$.

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

   In comparison with the point process 2 we have less concentrated clusters.

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

   This Poisson cluster process has the same parent point process and cluster size distribution as in 2 and 3, but each member in the typical cluster has independent $\mathcal{N}(0,v)$–distributed components. Thus, the clusters are unbounded (and the support of $\gamma^{(2)}_{\text{red}}$ is the whole $\mathbb{R}^1$) in contrast to the Matérn cluster process 3 and 4.

5. Thomas cluster process with intensity $\lambda = 1$, mean cluster size $\mu = 5$ and variability parameter $v = 1$.

   The clusters of this process are less concentrated than those of 4.

6. 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

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

The point processes (1.) – (5.) have intensity $\lambda = 1$ and the remaining point process (6.) has a slightly smaller intensity.

The Figures 1 and 2 present the obtained results of our simulation study. The solid lines in the below graphics show the (empirical) relative MSE

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

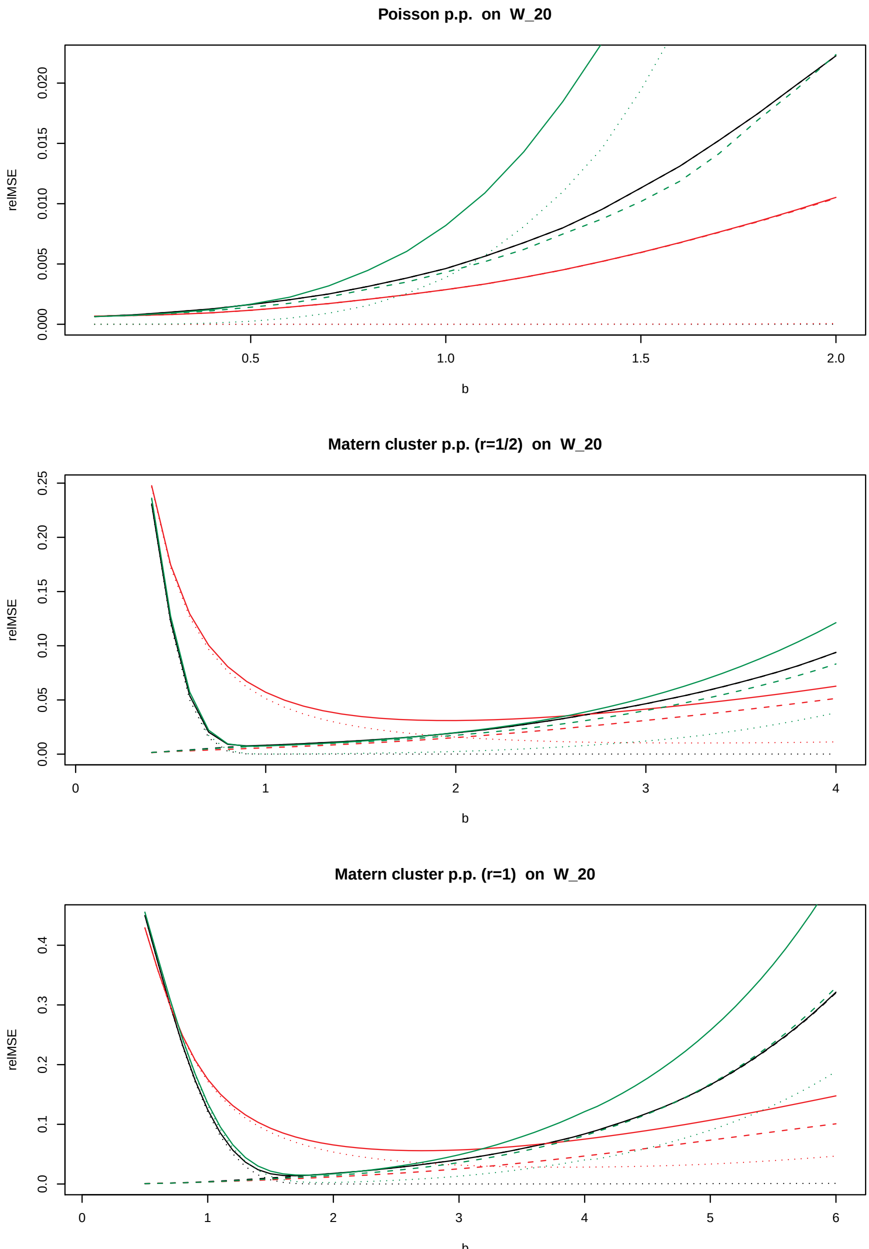

Fig. 1. Performance of the estimators of σ^2 on W_{20}.

of the different estimators of σ^2 as function of the quantity $b = n\, b_n$. To get the idea of the bias-variance trade off in dependence on the choice of b, values of the empirical

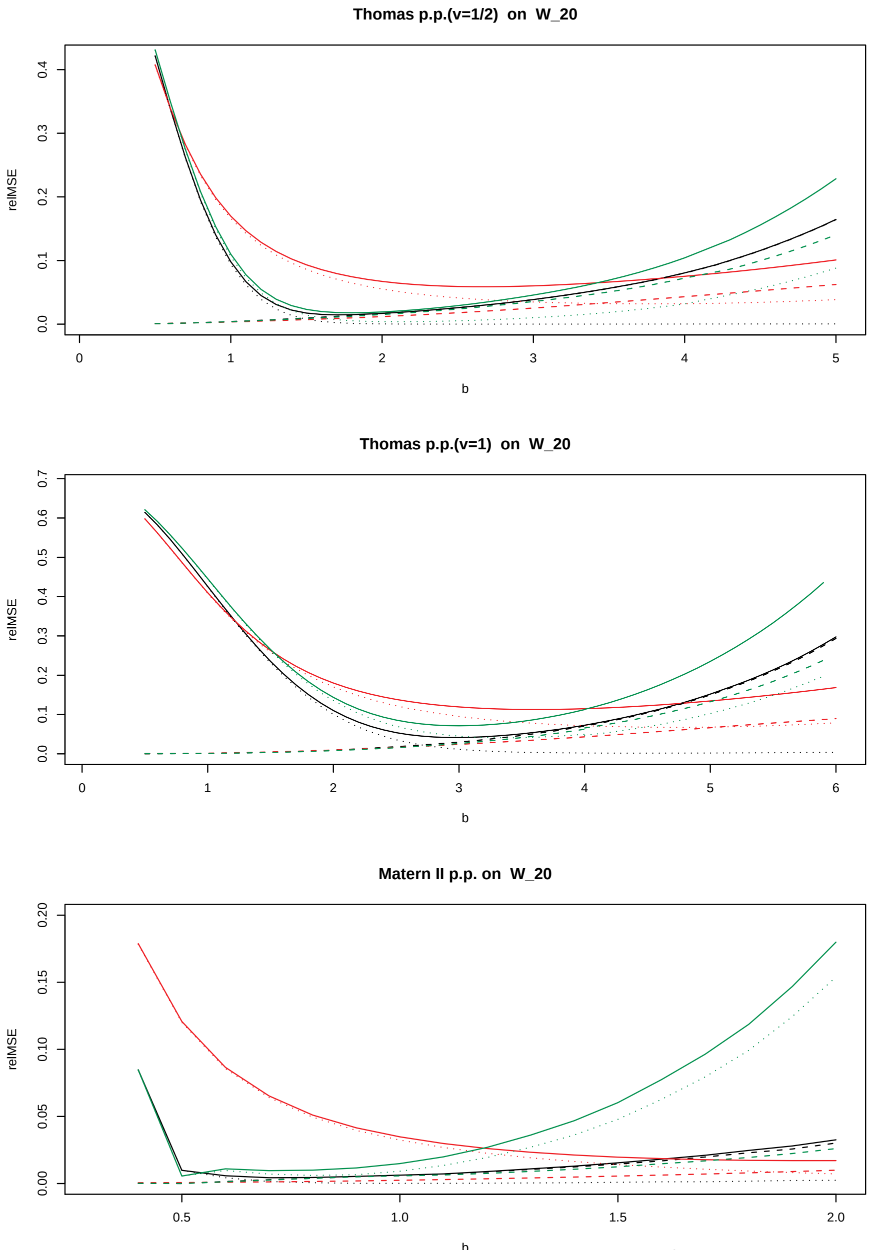

Fig. 2. Performance of the estimators of σ^2 on W_{20}.

variance of the estimators (dashes lines) and the squared bias (dotted lines) are also shown in the figures.

We can say that the kernel estimator $\widehat{\sigma}^2_n$ has overall the best performance closely followed by the estimator $\widehat{\sigma}^2_{n,I}$. The smaller efficiency of $\widehat{\sigma}^2_{n,I}$ is caused by the bias of the estimator, the variances of $\widehat{\sigma}^2_n$ and $\widehat{\sigma}^2_{n,I}$ are similar especially for the cluster processes. This is in accordance with the theoretical properties since $\widehat{\sigma}^2_n$ with the cylinder kernel is unbiased for b such that $\gamma^{(2)}_{\mathrm{red}}(B^c(o,b)) = 0$, not just asymptotically unbiased like the other estimators. The bias of $\widehat{\sigma}^2_{n,I}$ can be explained by "oversmoothing", since in $\widehat{\sigma}^2_{n,I}$ we first estimate the density of $\gamma^{(2)}_{\mathrm{red}}$ and then integrate it to get the estimate of $\gamma^{(2)}_{\mathrm{red}}(\mathbb{R}^2)$ whereas $\widehat{\sigma}^2_n$ estimates the integral $\gamma^{(2)}_{\mathrm{red}}(\mathbb{R}^2)$ directly. The "oversmoothing" effect can be also seen in case of the Matérn hard-core process and larger values of the bandwidth b – the kernel estimate of the $\gamma^{(2)}_{\mathrm{red}}$ density is too coarse to identify the fluctuations and thus produces an extra negative bias in the estimate of σ^2. Figure 3 shows the bias separately for the two representative cases of the clustered Thomas process (4.) and the Matérn hard-core process.

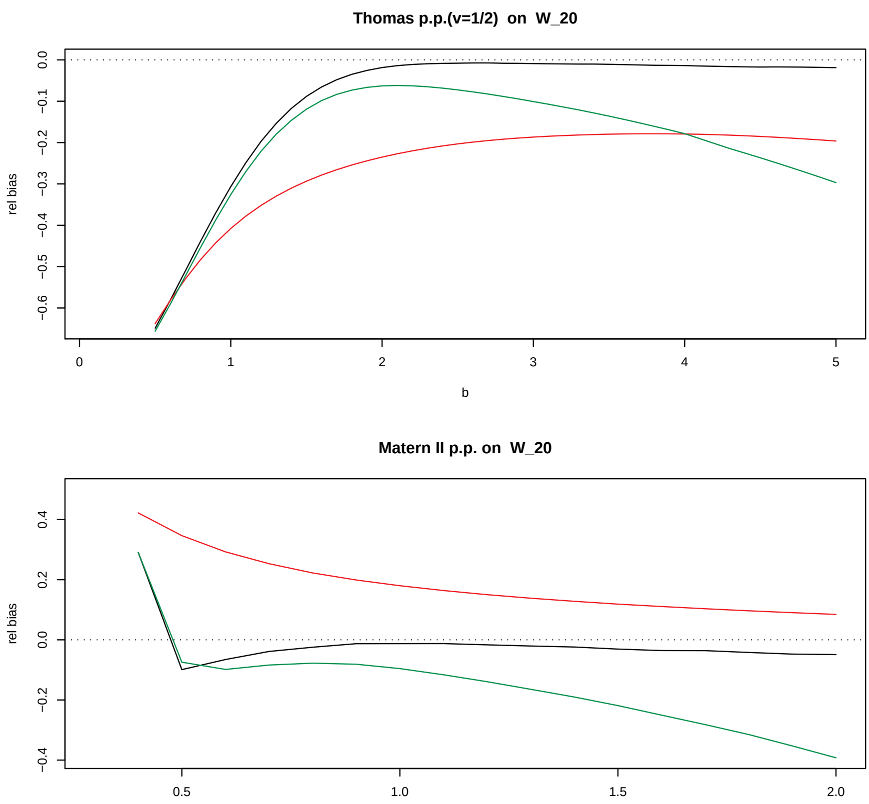

Fig. 3. Relative bias of the estimators of σ^2 for Thomas and Matérn II processes.

Concerning the subsampling estimator $\widehat{\sigma}^2_{n,V}$ it has smaller variance than the other two nevertheless its performance is mainly determined by the bias which is substantially larger than by the other two estimators. This can be nicely seen in case of the Poisson process where the $\gamma^{(2)}_{\text{red}}$ vanishes thus we estimate $\sigma^2 = \lambda$ and $\widehat{\sigma}^2_{n,V}$ exhibits the best performance.

By the subsampling estimator there is moreover one extra tuning parameter – namely the value of the grid cell size δ used for the discretization. The results shown in the figures used the choice $\delta = 0.2$. Nevertheless to check the influence of this parameter we also computed the estimates $\widehat{\sigma}^2_{n,V}$ with other values of δ, namely 0.1 and 0.5. The results were almost identical (therefore we do not show the graphs here) some difference could be observed only for the larger values of the parameter b where the efficiency of the estimator is not so good anyway.

Let us also remark here that even though we have the common scale for the b parameter on the x-axis it has different meaning for the subsampling and the other two estimators. Thus the minimal MSE (which of course depends on the choice of b) is attained in different values of b for $\widehat{\sigma}^2_{n,V}$ and for the other two estimators. $\widehat{\sigma}^2_n$ and $\widehat{\sigma}^2_{n,I}$ have virtually the same argument of minima of the MSE and their performance is the same in the neighbourhood of this optimal value. As expected this optimal value is close to the value r for which $\gamma^{(2)}_{\text{red}}(B^c(o,r)) = 0$ or $\gamma^{(2)}_{\text{red}}(B^c(o,r)) \approx 0$ (for processes with $\gamma^{(2)}_{\text{red}}$ with infinite support) since for $b \approx r$ the estimator σ^2 becomes unbiased and any further increase of the bandwidth b produces only increase of the variance of the estimator, but does not bring any further information (compare Figure 2 and Figure 3). The optimal value of b for the subsampling estimator is significantly larger than for $\widehat{\sigma}^2_n$.

ACKNOWLEDGEMENT

This work was supported by projects GA ČR 201/08/P100 from the Czech Science Foundation and MSM0021620839 financed by the Ministry of Education, Youth and Sports of the Czech Republic.

(Received January 22, 2011)

REFERENCES

[1] S. Böhm, L. Heinrich, and V. Schmidt: Kernel estimation of the spectral density of stationary random closed sets. Austral. & New Zealand J. Statist. *46* (2004), 41–52.

[2] K. L. Chung: A Course in Probability Theory. Second edition. Harcourt Brace Jovanovich, New York 1974.

[3] D. J. Daley and D. Vere-Jones: An Introduction to the Theory of Point Processes. Second edition. Vol I and II, Springer, New York 2003, 2008.

[4] S. David: Central Limit Theorems for Empirical Product Densities of Stationary Point Processes. Phd. Thesis, Augsburg Universität 2008.

[5] L. Heinrich: Asymptotic gaussianity of some estimators for reduced factorial moment measures and product densities of stationary Poisson cluster processes. Statistics *19* (1988), 87–106.

[6] L. Heinrich: Normal approximation for some mean–value estimates of absolutely regular tesselations. Math. Methods Statist. *3* (1994), 1–24.

[7] L. Heinrich: Asymptotic goodness–of–fit tests for point processes based on scaled empirical K–functions. Submitted.

[8] L. Heinrich and E. Liebscher: Strong convergence of kernel estimators for product densities of absolutely regular point processes. J. Nonparametric Statist. *8* (1997), 65–96.

[9] L. Heinrich and M. Prokešová: On estimating the asymptotic variance of stationary point processes. Methodology Comput. in Appl. Probab. *12* (2010), 451–471.

[10] L. Heinrich and V. Schmidt: Normal convergence of multidimensional shot noise and rates of this convergence. Adv. in Appl. Probab. *17* (1985), 709–730.

[11] J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan: Statistical Analysis and Modelling of Spatial Point Patterns. John Wiley and Sons, Chichester 2008.

[12] E. Jolivet: Central limit theorem and convergence of empirical processes of stationary point processes. In: Point Processes and Queueing Problems (P. Bartfai and J. Tomko, eds.), North–Holland, New York 1980, pp. 117–161.

[13] S. Mase: Asymptotic properties of stereological estimators for stationary random sets. J. Appl. Probab. *19* (1982), 111–126.

[14] D. N. Politis: Subsampling. Springer, New York 1999.

[15] D. N. Politis and M. Sherman: Moment estimation for statistics from marked point processes. J. Roy. Statist. Soc. Ser. B *63* (2001), 261–275.

[16] M. Prokešová and E. B. Vedel-Jensen: Asymptotic Palm likelihood theory for stationary point processes. Submitted.

[17] B. D. Ripley: Statistical Inference for Spatial Processes. Cambridge University Press, Cambridge 1988.

[18] D. Stoyan, W. S. Kendall, and J. Mecke: Stochastic Geometry and its Applications. Second edition. J. Wiley & Sons, Chichester 1995.

[19] J. C. Taylor: An Introduction to Measure and Probability. Springer, New York 1997.

[20] L. Zhengyan and L. Chuanrong: Limit Theory for Mixing Dependent Random Variables. Kluwer Academic Publishers, Dordrecht 1996.

Michaela Prokešová, Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University, Sokolovská 83, 186 75 Praha 8. Czech Republic.

e-mail: prokesov@karlin.mff.cuni.cz