Journal of Statistical Planning and Inference 155 (2014) 79–92

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Covariance of empirical functionals for inhomogeneous spatial point processes when the intensity has a parametric form

Jean-François Coeurjolly [a,*], Yongtao Guan [b]

[a] LJK, 51 Rue des Mathématiques, BP 53 - 38041 Grenoble cedex 09, France

[b] Department of Management Science, University of Miami, Coral Gables, FL 33124, USA

ARTICLE INFO

Article history:
Received 9 August 2013
Received in revised form 9 March 2014
Accepted 1 July 2014
Available online 10 July 2014

Keywords:
Spatial statistics
Point processes
Covariance estimation
Kernel estimation

ABSTRACT

This paper is concerned with the problem of estimating covariances of inhomogeneous second-order reweighted stationary spatial point processes when the intensity of the spatial point process has a parametric form. The proposed estimator is based on kernel techniques. It is a very simple and fast estimator which in addition does not require one to model second and higher moments of the spatial point process. Under very mild assumptions, mainly on characteristics of the point process, we prove the mean squared consistency of our estimator. Finally, we show in a simulation study that the kernel-based covariance estimator outperforms existing methods when it is applied to build confidence intervals of the intensity.

1. Introduction

Variance estimation can be a challenging task for statistics derived from spatial point processes, because the dependence structure of the process may be highly complex. For example, in an ecological setting, a spatial point process may consist of the spatial locations of a particular tree species in a plot. These spatial locations may be clustered (i.e., positively correlated) at a relatively larger scale due to seed dispersal, but are inhibitive (i.e., negatively correlated) at a much smaller scale because of competitions among neighboring trees for common resources. It is often difficult to derive a flexible class of parametric models to fully describe the complex dependence structure. As a result, nonparametric variance estimation procedures such as subsampling and block bootstrap have been developed, because they rely on fewer assumptions on the dependence structure of the spatial point process (e.g., Politis and Sherman, 2001; Guan and Loh, 2007; Guan, 2009). To apply either subsampling or block bootstrap, one would have to first form (overlapping) subblocks from the observation window. This can be challenging and computationally intensive, because there are many different ways to select such subblocks.

Heinrich and Prokešová (2010) recently developed a kernel estimator for the variance of an intensity estimator derived from a stationary point process. Their approach does not require forming any subblocks and is closely related to kernel estimation of the spectral density at 0 for stationary time series (e.g. Politis, 2003). In this paper, we extend the approach of Heinrich and Prokešová (2010) to the nonstationary setting, where the intensity of the process varies across space. Specifically, we aim to estimate the covariance between the following two statistics:

$$\sum_{u\in \mathbf{X}_W} f(u) \quad \text{and} \quad \sum_{u\in \mathbf{X}_W} h(u), \tag{1}$$

* Corresponding author.
E-mail addresses: Jean-Francois.Coeurjolly@upmf-grenoble.fr (J.-F. Coeurjolly), yguan@miami.edu (Y. Guan).

http://dx.doi.org/10.1016/j.jspi.2014.07.003
0378-3758/

where $f(\cdot)$ and $h(\cdot)$ are real-valued Borel measurable parametric functions defined on $\mathbb{R}^d$ and $\mathbf{X}_W$ is a spatial point process observed over $W \subset \mathbb{R}^d$. Statistics in the forms of (1) are particularly useful in both nonparametric and parametric estimations of the intensity; see Section 2 for example.

To derive our covariance estimator, we first show that the covariance between $f(\cdot)$ and $h(\cdot)$ can be written as the sum of two integrals, the first of which is a single integral defined in terms of $f(\cdot)$, $h(\cdot)$ and the intensity of the point process while the second is a double integral also involving the pair correlation function. We then develop a plug-in estimator for the single integral and a kernel estimator for the double integral. To avoid numerical integrations which can be computationally intensive, we develop a novel approach by augmenting the original point process with a separate simulated homogeneous Poisson process (see Section 3 for details).

Our proposed covariance estimator is more general than the variance estimator given in Heinrich and Prokešová (2010) for two reasons. First, we consider nonstationary point processes with varying and typically unknown intensities instead of stationary point processes as considered in Heinrich and Prokešová (2010). The unknown intensity must be estimated and its effect needs to be properly accounted for especially when studying the theoretical properties of our estimator. Second, we consider more general forms of $f(\cdot)$ and $h(\cdot)$ whereas Heinrich and Prokešová (2010) focused on the case $f(\cdot) = h(\cdot) = 1$. Compared to estimators based on subsampling or block bootstrap, our estimator is computationally more efficient. Moreover, it does not require forming any subblocks and can therefore be applied to data observed over highly irregularly shaped windows. It should be noted that subsampling and block bootstrap can be applied to estimate other distributional features of these statistics (e.g. Politis et al., 1999; Lahiri, 2003). However, we focus only on covariance estimation here.

The rest of the paper is organized in the following way. In Section 2, we provide background on spatial point processes and introduce notation. In Section 3, we derive our proposed estimator and study its asymptotic properties. We illustrate the methodology through simulation in Section 4 and an application to tropical forestry ecology in Section 5. All technical proofs are contained in the Appendix.

2. Background and notation

Let $\mathbf{X}$ be a spatial point process defined on $\mathbb{R}^d$, which we view as a random locally finite subset of $\mathbb{R}^d$. Let $\mathbf{X}_W = \mathbf{X} \cap W$ where $W \subset \mathbb{R}^d$ is a compact set of positive Lebesgue measures $|W|$. Then, the number of points in $\mathbf{X}_W$, denoted by $n(\mathbf{X}_W)$ or by $N(W)$ in this paper, is finite, and a realization of $\mathbf{X}_W$ is of the form $\mathbf{x} = \{x_1, \dots, x_m\} \subset W$ for some nonnegative finite integer m. If $m = 0$, then $\mathbf{x} = \emptyset$ is an empty point pattern in W. For further background and measure theory on spatial point processes, see e.g. Daley and Vere-Jones (2003) and Møller and Waagepetersen (2004). For $u \in \mathbb{R}^d$ and $E \subset \mathbb{R}^d$, $\|u\|$ stands for the Euclidean norm of u and $d(\cdot, E)$ denotes the distance given by $d(u, E) = \inf_{v \in E}\{\|v - u\|\}$. Finally, $\mathcal{B}(u, R)$ is the Euclidean ball centered at u with radius R.

We assume that $\mathbf{X}$ has a locally integrable intensity function ρ. By Campbell's theorem (e.g. Møller and Waagepetersen, 2004), for any real Borel function k defined on $\mathbb{R}^d$ such that $k\rho$ is absolutely integrable with respect to the Lebesgue measure on $\mathbb{R}^d$,

$$\mathrm{E} \sum_{u \in \mathbf{X}} k(u) = \int_{\mathbb{R}^d} k(u)\rho(u)\,\mathrm{d}u. \tag{2}$$

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

$$\mathrm{E} \sum_{u_1, \dots, u_l \in \mathbf{X}}^{\neq} k(u_1, \dots, u_l) = \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} k(u_1, \dots, u_l)\rho_l(u_1, \dots, u_l)\,\mathrm{d}u_1 \cdots \mathrm{d}u_l, \tag{3}$$

where the sign $\neq$ over the summation means that $u_1, \dots, u_l$ are pairwise distinct. Note that $\rho = \rho_1$.

We assume that ρ has a parametric form, i.e. $\rho(u) = \rho(u; \theta)$ for $\theta \in \Theta \subseteq \mathbb{R}^p$ and for all $u \subset \mathbb{R}^d$. We consider in this paper the class of second-order intensity reweighted stationary (SOIRWS) point processes (e.g. Møller and Waagepetersen, 2004), for which the pair-correlation function g, given by

$$g(u, v) = \frac{\rho_2(u, v)}{\rho(u)\rho(v)},$$

for any $u, v \in \mathbb{R}^d$ with the convention $a/0 = 0$ for $a \geq 0$, is translation invariant, i.e. $g(u, v) = g(u - v, o)$ where o is the origin point. For ease of presentation, we denote $g(u - v) = g(u, v)$.

Let $\theta^\star$ be the true value of the parameter and let $(W_n)_{n \geq 1}$ be a sequence of bounded regions with volume $|W_n|$ such that W_n expands unboundedly in all directions as $n \to \infty$. The aim of this paper is to estimate the quantity:

$$c_n = c_n(\theta^\star) := |W_n|^{-1} \operatorname{Cov}\left[\sum_{u \in X_{W_n}} f(u; \theta^\star), \sum_{u \in X_{W_n}} h(u; \theta^\star)\right], \tag{4}$$