Spatial Statistics 18 (2016) 474–488

Contents lists available at ScienceDirect

Spatial Statistics

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

Independent component models for replicated point processes

Daniel Gervini

Department of Mathematical Sciences, University of Wisconsin–Milwaukee, United States

ARTICLE INFO

Article history:
Received 31 March 2016
Accepted 30 September 2016
Available online 17 October 2016

Keywords:
Doubly-stochastic process
Functional data analysis
Latent-variable model
Poisson process
Spline smoothing

ABSTRACT

We propose a semiparametric independent-component model for the intensity functions of replicated point processes. We show that the maximum likelihood estimators of the model parameters are consistent and asymptotically normal when the number of replications goes to infinity. The finite-sample behavior of the estimators is studied by simulation. As an example of application, we analyze the temporal variation in the spatial distribution of street robberies in Chicago.

1. Introduction

Point processes in time and space have a broad range of applications, in diverse areas such as neuroscience, ecology, finance, astronomy, seismology, and many others. Examples are given in classic textbooks like Cox and Isham (1980), Diggle (2013), Møller and Waagepetersen (2004), Streit (2010), and Snyder and Miller (1991), and in the papers cited below. However, the point-process literature has mostly focused on single-realization cases, such as the distribution of trees in a single forest (Jalilian et al., 2013) or the distribution of cells in a single tissue sample (Diggle et al., 2006). Situations where several replications of a process are available are increasingly common. Among the few papers proposing statistical methods for replicated point processes we can cite Diggle et al. (1991), Baddeley et al. (1993), Diggle et al. (2000), Bell and Grunwald (2004), Landau et al. (2004), Wager et al. (2004), and Pawlas (2011). However, these papers propose estimators for summary statistics of the

E-mail address: gervini@uwm.edu.

http://dx.doi.org/10.1016/j.spasta.2016.09.006
2211-6753/

processes (the so-called F, G and K statistics) rather than the intensity functions, which would be more informative.

When several replications of a process are available, it is possible to "borrow strength" across replications to estimate the intensity functions. This is the basic idea behind many functional data methods that are applied to stochastic processes which, individually, are only sparsely sampled (James et al., 2000; Yao et al., 2005; Gervini, 2009). Functional Data Analysis has mostly focused on continuous temporal processes; little work has been done on discrete or spatial point processes. We can mention Bouzas et al. (2006, 2007) and Fernández-Alcalá et al. (2012), which have rather limited scopes since they only estimate the mean of temporal Cox processes, and Wu et al. (2013), who estimate the mean and the principal components of independent and identically distributed realizations of a temporal Cox process, but their method is based on kernel estimators of covariance functions that cannot be easily extended beyond i.i.d. replications of temporal processes. The semiparametric model we propose in this paper, on the other hand, can be applied just as easily to temporal or spatial processes, and it can accommodate covariates and be extended to non-i.i.d. situations like ANOVA models or more complex data structures like marked point processes and multivariate processes, although we will not consider those extensions in this paper.

As an example of application we will analyze the spatial distribution of street robberies in Chicago in the year 2014, where each day is seen as a replication of the spatial process. Our model is presented in Section 2 and the estimation process in Section 3. We establish the consistency and asymptotic normality of the component estimators in Section 4, study their finite-sample behavior by simulation in Section 5. The Chicago crime data is analyzed in Section 6.

2. The model

A point process X is a random countable set in a space $\mathscr{S}$, where $\mathscr{S}$ is usually $\mathbb{R}$ for temporal processes and $\mathbb{R}^2$ or $\mathbb{R}^3$ for spatial processes (Møller and Waagepetersen, 2004, ch. 2; Streit, 2010, ch. 2). Locally finite processes are those for which $\#(X \cap B) < \infty$ with probability one for any bounded $B \subseteq \mathscr{S}$. For such processes we can define the count function $N(B) = \#(X \cap B)$. Given a locally integrable function $\lambda : \mathscr{S} \to [0, \infty)$, i.e. a function such that $\int_B \lambda(t)dt < \infty$ for any bounded $B \subseteq \mathscr{S}$, the process X is a Poisson process with intensity function $\lambda(t)$ if (i) $N(B)$ follows a Poisson distribution with rate $\int_B \lambda(t)dt$ for any bounded $B \subseteq \mathscr{S}$, and (ii) conditionally on $N(B) = m$, the m points in $X \cap B$ are independent and identically distributed with density $\tilde{\lambda}(t) = \lambda(t)/\int_B \lambda$ for any bounded $B \subseteq \mathscr{S}$.

Let X be a Poisson process with intensity function $\lambda(t)$ and $X_B = X \cap B$ for a given B. Then the density function of X_B at $x_B = \{t_1, \ldots, t_m\}$ is

$$\begin{aligned} f(x_B) &= f(m)f(t_1, \ldots, t_m|m) \\ &= \exp\left\{-\int_B \lambda(t)dt\right\} \frac{\left\{\int_B \lambda(t)dt\right\}^m}{m!} \times \prod_{j=1}^m \tilde{\lambda}(t_j). \end{aligned} \tag{1}$$

Since the realizations of X_B are sets, not vectors, what we mean by 'density' is the following: if $\mathscr{N}$ is the family of locally finite subsets of $\mathscr{S}$, i.e. $\mathscr{N} = \{A \subseteq \mathscr{S} : \#(A \cap B) < \infty \text{ for all bounded } B \subseteq \mathscr{S}\}$, then for any $F \subseteq \mathscr{N}$,

$$\begin{aligned} P(X_B \in F) &= \sum_{m=0}^{\infty} \int_B \cdots \int_B \mathbb{I}(\{t_1, \ldots, t_m\} \in F) f(\{t_1, \ldots, t_m\}) dt_1 \cdots dt_m \\ &= \sum_{m=0}^{\infty} \frac{\exp\left\{-\int_B \lambda(t)dt\right\}}{m!} \int_B \cdots \int_B \mathbb{I}(\{t_1, \ldots, t_m\} \in F) \\ &\quad \times \left\{\prod_{j=1}^m \lambda(t_j)\right\} dt_1 \cdots dt_m, \end{aligned}$$