Journal of Multivariate Analysis 155 (2017) 72–82

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Testing proportionality between the first-order intensity functions of spatial point processes

Tonglin Zhang [a,*], Run Zhuang [b]

[a] Department of Statistics, Purdue University, 250 North University Street, IN 47907-2066, USA

[b] Computation, Statistic-Oncology Eli Lilly and Company Lilly Corporate Center, Indianapolis, IN 46285, USA

ARTICLE INFO

Article history:
Received 14 November 2015
Available online 7 December 2016

AMS 2010 subject classifications:
primary 62M30
secondary 62G05

Keywords:
Pinned Brownian sheet
Converges in distribution
Intensity functions
Point processes
Proportionality
Strong mixing condition

ABSTRACT

This article proposes a Kolmogorov–Smirnov type test for proportionality between the first-order intensity functions of two independent spatial point processes. After appropriate scaling, the test statistic is constructed by maximizing the absolute difference between their point densities over a π-system. By treating non-stationary point processes as transformed from stationary point processes such that all questions of asymptotics related to the tightness can be answered, the article shows that the resulting test statistic converges weakly to the absolute maximum of a pinned Brownian sheet. This may be reduced to the standard Brownian bridge in a special case. A simulation study shows that the type I error probability of the test is close to the significance level and the power function increases to 1 as the magnitude of non-proportionality increases. In applications to two typical natural hazard data, the article concludes that the first-order intensity functions might be proportional in one case and not in the other.

1. Introduction

This research is motivated by two typical problems in natural hazards. The first is about the change of wildfire patterns in forest ecosystems before and after the occurrence of a few extremely large fires. The second is about the change of earthquake patterns in geographic regions before and after a few extremely large earthquakes. Examples include forest wildfires in North America [30], seismic gaps caused by great earthquakes in the San Francisco Bay Area [5] and in Chile [18], as well as landslides triggered by great earthquakes in Japan [31].

Motivated by the above two problems, we study a hypothesis testing problem between two spatial point processes. We assume that dependence between the two processes can be ignored but the dependence within them cannot. We model the dependence within the processes by clusters or inhibitions. Both are important in practice. For instance, earthquake occurrences are often close to each other in space and time because main-shock earthquakes are often accompanied by aftershock earthquakes [21,23,34]. Forest wildfire occurrences may also be close to each other in both space and time because of similarity of local climates and environments [15]. Therefore, it is necessary to address clusters or inhibitions in our approach.

Spatial point processes are important in a variety of scientific phenomena. Examples include human or animal diseases [1,8], forest wildfires [24,28], regional earthquakes [22], tree locations [32], and many others. In the literature, point distributions and dependence structures are often modeled by intensity functions [7]. Examples include estimation

* Corresponding author.
E-mail addresses: tlzhang@purdue.edu (T. Zhang), run@lilly.edu (R. Zhuang).

http://dx.doi.org/10.1016/j.jmva.2016.11.013

of the first-order [6] or second-order intensity function [2], and tests for stationarity [11,33] or isotropy [12]. Popular measurements for interpoint dependence such as the usual K-function [25,26], L-function [3], and pair correlation function [29] have been proposed. These measurements have been extensively used previously [13,27]. However, approaches for the relationship between a few spatial point patterns have not been investigated in detail.

To study this relationship, we propose a formal test about whether the first-order intensity functions of two independent spatial point processes are proportional. We propose a Kolmogorov–Smirnov type statistic to test proportionality. The p-value is calculated using the statistic's asymptotic null distribution. After appropriate scaling, the test statistic is constructed by maximizing the absolute difference between their point densities over a π-system. With a particular choice of π-system, the asymptotic null distribution may be reduced to the distribution of the absolute maximum of the standard Brownian bridge, which has a closed form Taylor expansion [16].

We evaluate the properties of our test by simulations and applications. In simulations, we study type I error probabilities and power. We conclude that the type I error probabilities are always close to the significance level and the power function always increases to 1 as the magnitude of non-proportionality increases. For illustration, we consider a forest wildfire and an earthquake example. In the first case, we compare wildfire patterns in Alberta, Canada before and after an extremely large wildfire occurred in 2002. We conclude that the wildfire patterns are almost proportional. In the second example, we compare earthquake patterns of a certain area in the Indian Ocean before and after the Great Sumatra–Andaman Earthquake, which occurred on December 26, 2004 and killed more than 230,000 people [17]. We conclude that the earthquake patterns are not proportional, implying that a change occurred.

To the best of our knowledge, our approach is the first formal test for proportionality between two spatial point processes. As the test statistic is nonparametric and independent of the shape of the region, our approach can be easily implemented to study natural hazards in any given geographical region. As the computation of the test statistic does not contain estimates of intensity functions, our approach avoids the complicated nonparametric estimation problem.

The article is organized as follows. In Section 2, we propose our test statistic and derive its asymptotic null distribution. In Section 3, we evaluate the performance of our test statistic by Monte Carlo simulations. In Section 4, we apply our approach to the Alberta forest wildfire data and the Indian Ocean earthquake data. In Section 5, we provide a discussion.

2. Method

We review a few important concepts of spatial point processes in Section 2.1, propose our test statistic in Section 2.2, and derive its asymptotic null distribution in Section 2.3. All of those are important in our approach.

2.1. Spatial point process

Let $\mathcal{S} \subseteq \mathbb{R}^d$ be measurable. A spatial point process $\mathcal{N}$ on $\mathcal{S}$ is composed of random points in $\mathcal{S}$. Let $\mathcal{B}(\mathcal{S})$ be the collection of all Borel sets of $\mathcal{S}$ and $\mathcal{N}(A)$ be the number of points in $A \in \mathcal{B}(\mathcal{S})$. If A is bounded, then $\mathcal{N}(A)$ is finite. The distribution of $\mathcal{N}$ can be theoretically defined through the counting measure approach which is available in many textbooks. The kth order intensity function of $\mathcal{N}$ (if it exists) is defined as

$$\lambda_k(\mathbf{s}_1, \ldots, \mathbf{s}_k) = \lim_{d(U_{\mathbf{s}_i}) \to 0, i=1,\ldots,k} \frac{\mathrm{E}\left\{\prod_{i=1}^{k} \mathcal{N}(U_{\mathbf{s}_i})\right\}}{\prod_{i=1}^{k} |U_{\mathbf{s}_i}|},$$

where $\mathbf{s}_1, \ldots, \mathbf{s}_k \in \mathcal{S}$ are distinct, $U_{\mathbf{s}}$ is a neighbor of $\mathbf{s}$, $|U_{\mathbf{s}}|$ is its Lebesgue measure, and $d(U_{\mathbf{s}})$ is the diameter of $U_{\mathbf{s}}$. The spatial point process $\mathcal{N}$ is strongly stationary if for any $A_1, \ldots, A_k \in \mathcal{B}(\mathcal{S})$ the joint distribution of $\mathcal{N}(A_1+\mathbf{s}), \ldots, \mathcal{N}(A_k+\mathbf{s})$ does not depend on $\mathbf{s}$.

The mean structure of $\mathcal{N}$ is

$$\mu(A) = \mathrm{E}\{\mathcal{N}(A)\} = \int_A \lambda(\mathbf{s})d\mathbf{s},$$

where $\lambda(\mathbf{s}) = \lambda_1(\mathbf{s})$ is the first-order intensity function. The covariance structure of $\mathcal{N}$ is

$$\mathrm{cov}\{\mathcal{N}(A_1), \mathcal{N}(A_2)\} = \int_{A_1}\int_{A_2} \{g(\mathbf{s}_1, \mathbf{s}_2) - 1\}\lambda(\mathbf{s}_1)\lambda(\mathbf{s}_2)d\mathbf{s}_2 d\mathbf{s}_1 + \mu(A_1 \cap A_2),$$

where $g(\mathbf{s}_1, \mathbf{s}_2) = \lambda_2(\mathbf{s}_1, \mathbf{s}_2)/\{\lambda(\mathbf{s}_1)\lambda(\mathbf{s}_2)\}$ is the pair correlation function. The covariance function of $\mathcal{N}$ is

$$\Gamma(\mathbf{s}_1, \mathbf{s}_2) = \{g(\mathbf{s}_1, \mathbf{s}_2) - 1\}\lambda(\mathbf{s}_1)\lambda(\mathbf{s}_2) + \lambda(\mathbf{s}_1)\delta_{\mathbf{s}_1,\mathbf{s}_1}(\mathbf{s}_2, \mathbf{s}_2),$$

where $\delta_{\mathbf{s},\mathbf{s}}$ represents the point measure at $(\mathbf{s}, \mathbf{s}) \in \mathcal{S} \times \mathcal{S}$.

Let G be the probability generating function of $\mathcal{N}$ defined as

$$G(\varphi) = \mathrm{E}\left\{\exp \int_{\mathcal{S}} \ln \varphi(\mathbf{s}) d\mathcal{N}(\mathbf{s})\right\},$$