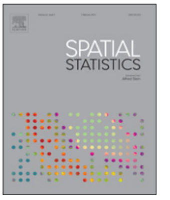
Spatial Statistics 9 (2014) 20–37

Contents lists available at ScienceDirect

Spatial Statistics

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



On the local odds ratio between points and marks in marked point processes

Tonglin Zhang [a,*], Qianlai Zhuang [b,c]

[a] Department of Statistics, Purdue University, 250 North University Street, West Lafayette, IN 47907-2066, USA
[b] Department of Earth, Atmospheric, and Planetary Sciences, Purdue University, West Lafayette, IN, 47907, USA
[c] Department of Agronomy, Purdue University, West Lafayette, IN, 47907, USA

A R T I C L E I N F O

Article history:
Received 26 August 2013
Accepted 9 December 2013
Available online 24 December 2013

Keywords:
Consistency
Contingency table
Intensity functions
Local odds ratios
Marked point processes
Relative risks

A B S T R A C T

Marked point processes are widely used stochastic models for representing a finite number of natural hazard events located in space and time and their data often associate event measurements (i.e. marks) with event locations (i.e. points). An interesting statistical problem of marked point processes is to measure and estimate the localized dependence between points and marks. To solve this problem, an approach of local odds ratio is proposed, where the local odds ratio is defined by the localized ratio of the relative risk for an event to have a small mark to the relative risk to have a large mark. To establish the approach, the article presents definition, estimation, and statistical properties. To justify the usefulness of the approach, the article presents two particular examples in natural hazards: a forest wildfire study and an earthquake study. It finds that values of local odds ratios are mostly likely low in one subarea but high in another subarea, which indicates that events with large mark values are mostly likely to appear in the former subarea but less likely to appear in the latter subarea. It is expected that the proposed approach will be widely applicable in natural hazard studies.

1. Introduction

Marked point processes (MPPs) are commonly used stochastic models, which have been widely applied to data involved both the spatial (including spatiotemporal) coordinates of events and their corresponding measurements. Methods of MPPs are often used to model a number of natural hazards

* Corresponding author. Tel.: +1 765 496 2096; fax: +1 765 494 0558.
E-mail addresses: tlzhang@purdue.edu, tlzhang@stat.purdue.edu (T. Zhang), qzhuang@purdue.edu (Q. Zhuang).

2211-6753/$ – see front matter
http://dx.doi.org/10.1016/j.spasta.2013.12.002

located in space and time. There are many successful applications of MPPs in literature. These include MPP modeling and prediction of earthquakes (Holden et al., 2003; Ogata, 1988; Ogata and Katsura, 1993; Vere-Jones, 1995), where each earthquake is represented by a magnitude and a space–time coordinate. The three-dimensional space coordinate contains the longitude, latitude and depth of earthquake occurrences. MPPs for forest wildfires have been discussed by Peng et al. (2005), where each wildfire is represented by its area burned and space–time coordinate. The two-dimensional space coordinate contains the longitude and latitude of wildfire occurrences.

The approach of the localized dependence between points and marks relies on a strict mathematical definition of MPPs, which will be introduced in the next section. In a simplified version, one can describe an MPP with an artificial order such that data can be represented by $N = \{(S_i, M_i) : i = 1, \ldots, n\}$, where n is the total number of events, S_i are the locations of points, and M_i are the corresponding marks. Specific geostatistical methods including variogram analysis, various kinds of kriging, and geostatistical simulation techniques may be used to model an MPP (Cressie, 1993), but these methods rely on the fundamental assumption that point locations appear independently of marks because the definition of correlation functions used in a geostatistical method often ignores point distributions (Diggle et al., 2003). Since the independence assumption between points and marks is often unrealistic in applications, these methods may not be used if points and marks are highly correlated. For instance, the relative positions of trees in a forest have repercussions on their size owing to their competition for light or nutrient (Schlather et al., 2004), indicating that tree sizes and locations of trees may not be independent. Forest wildfire activities exhibit power-law relationships between frequency and burned area (Malamud et al., 2005), indicating that the burned area and the locations of forest wildfires may not be independent either.

It is especially convenient in modeling, estimation, and prediction in an MPP if marks and points are independent. Many commonly used Hawkes models, such as the epidemic-type aftershock sequences (EATS) model (Ogata, 1998), may exhibit the independence between marks and points (Schoenberg, 2004). In the **spatstat** (Baddeley and Turner, 2005) and **PtProcess** (Harte, 2010) packages in **R** several useful methods based on MPPs under the assumption of independence are available (McElroy and Politis, 2007; Poliltis and Sherman, 2001). If the independence assumption is violated, then intensity-dependent models may also be useful (Ho and Stoyan, 2008; Malinowski et al., 2012; Myllymäki and Penttinen, 2009). However, these methods cannot be used to describe the localized dependence between points and marks because the relationship between points and marks is often modeled globally. For example, the mark (magnitude) distribution in earthquake activities locally depends on their geographical locations which cannot be accounted for by an intensity-dependent model. The mark (area burned) distribution in forest wildfire activities locally depends on forest densities which cannot either be accounted for by an intensity-dependent model. Therefore, it is important to develop a statistical approach to modeling the local dependence between points and marks.

To account for the dependence between marks and points, we modify the approach of the odds ratio for contingency tables (Agresti, 2002) to an approach of local dependence between points and marks of MPPs. We call it the approach of the local odds ratio, where the local odds ratio is a measure of the strength of the local dependence between points and marks. In the approach, we note that the odds ratio is one of the most important measures of row–column dependence in a contingency table and it is also an important index in binomial or Poisson regression. Unlike other measures of dependence in the contingency table (such as the relative risk), the odds ratio treats rows and columns symmetrically. Its value does not change when the orientation of the table reverses so that the rows become the columns and the columns become the rows. Therefore, the odds ratio is invariant under the transpose transformation. To define the localized odds ratio, we modify the classical definition of the odds ratio. We expect that the modified local odds ratio can be theoretically derived at any given location in the whole study area. Based on values of the local odds ratio, one can compare the local risks with the global risks for large mark events. In addition, one can compare local risks between two specific locations. Values of local odds ratio are useful to identify a subarea with high risks of large events. Examples include a method to identify a subarea of high risk of large earthquakes in earthquake studies or a subarea of high risks of large fires in forest wildfire studies. Because large events are more important than small events in the natural hazard studies, the local odds ratio may be used as a standard measure for the risk analysis of large events in these studies.