Statistics and Probability Letters 111 (2016) 26–31

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Modelling cluster detection in spatial scan statistics: Formation of a spatial Poisson scanning window and an ADHD case study

S.M. Aboukhamseen [a,*], A.R. Soltani [a], M. Najafi [b]

[a] Department of Statistics and Operations Research, Faculty of Science, Kuwait University, P.O. Box 5969, Safat 13060, Kuwait
[b] Soor Center Kuwait, P.O. Box 4707, Safat 13048, Kuwait

ARTICLE INFO

Article history:
Received 26 July 2015
Received in revised form 16 December 2015
Accepted 21 December 2015
Available online 11 January 2016

MSC:
primary 62G10
60F03
secondary 60F15

Keywords:
Spatial scan statistics
Cluster detection
Prior distribution
Poisson test
ADHD

ABSTRACT

In this article we present a testing procedure for spatial scan statistics when the underlying population characteristics are not known. Specifically, the test procedure is designed for the situation when the number of affected cases in the population is random. We further assume that the number of contaminated case in the geographic region of interest follows a Poisson distribution. Then, under the null assumption of no cluster, we prove that the scanning window detecting contaminated cases is indeed a specific homogeneous spatial Poisson point process on the zones that constitute the region of interest. We then proceed to formulate an effective cluster detection testing procedure together with confidence intervals for the parameters of interests. We apply our procedure to the interesting and intensive real case study of detecting clusters of school-aged children diagnosed with Attention Deficit Hyperactivity Disorder (ADHD) in the State of Kuwait. We observe that geographic boundaries defining ethno-social groups are significant in determining ADHD prevalence among school-aged children in the State of Kuwait.
© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Spatial scan statistics refers to a cluster detection technique whereby a zone Z in a region G is tested as a possible cluster of individuals that carry a characteristic of interest in the underlying binomial or Poisson distributions. The scanning window, Z, is defined as an expandable sub-region that runs over the area composing G. At each expansion of the scan window, a test is conducted to determine if it is a significant cluster of *points*. Several authors have contributed to the development of this procedure including Naus (1965), Kulldorff and Nagarwalla (1995), Tango and Takahashi (2005), and Zhang and Lin (2009). However, cluster detection in spatial scan statistics is often attributed to the fundamental article by Kulldorff (1997).

Although spatial scan statistics provides crucial dynamic analysis for identifying severely affected regions, it has many limitations. One such limitation is the computational intensity of the problem that emanates from the integration of repeated geographic scanning with inferential testing. The Kulldorff (1997) test procedure uses a likelihood ratio test constructed for both the Binomial and the Poisson models to identify clusters. The critical values of the test statistics are computed via simulation which further compounds the required computation. An equivalent and effective test statistic is developed in Soltani and Aboukhamseen (2015) where the distribution can be readily identified in both small and large

* Corresponding author.
E-mail addresses: suja.aboukhamseen@ku.edu.kw (S.M. Aboukhamseen), asoltanir@yahoo.com (A.R. Soltani).

http://dx.doi.org/10.1016/j.spl.2015.12.025
0167-7152/© 2015 Elsevier B.V. All rights reserved.

sampling problems. By foregoing the simulation, the required computations are limited to the geographic aspect of the scan procedure.

Another major limitation of scan statistics is the data required for the analysis. Prior to the examination of a region G for cluster(s), a census of the region must be executed and all individuals must be classified as either *points* of a characteristic of interest or *nonpoints*. In many situations such information is not available. Consider, for instance, the study of a multinational pandemic. In such circumstances, identification of disease clusters is of the upmost importance and urgency, yet the availability of reliable up-to-date population data and case reporting is often times limited and in some instances restricted.

In this article, we address the issue of data limitation. We follow the basic notions developed in Soltani and Aboukhamseen (2015) but also assume a Poisson distribution on the observed number of points in the whole of the region G. This indeed changes the underlying counts on contaminated cases from census based to sample based. We prove that the point counting scan window, Z, as it expands over G under the null assumption of no cluster, follows a certain homogeneous spatial Poisson point process. Based on this, a new testing procedure for cluster detection in spatial scan statistics when the number of points follows a Poisson distribution is presented.

The second part of this article is an intensive case study; an application in the health sciences. Specifically, we employ our derivations on spatial scan statistics in the detection of residential districts in the State of Kuwait with a distinctly higher population of Attention Deficit Hyperactivity Disorder, ADHD, cases, among school-aged children. ADHD is a common neuro-behavioral disorder in children that can remain until adulthood. Many developed nations, including the US, have reported an increases in the prevalence of ADHD cases among school age children in the last two decades. It is estimated that ADHD affects 8%–12% of children worldwide (Polanczyk et al., 2007). Interestingly, our analysis reveals a single cluster among Kuwait's residential districts.

This article is organized as follows. In Section 2, the basic notions behind the analysis is introduced and the notation is presented. In Section 3, we give our cluster detection testing procedure, together with certain confidence intervals for the parameters of interest. A cluster detection algorithm is explained in Section 4, and in Section 5 we apply our testing procedure to the real data of ADHD cases. A discussion of the results is given in Section 6.

2. Preliminaries

Assume that G is a region and a number of individuals are spread out over this region. Each individual in G carries one and only one of K possible labels; thus individuals are classified into K classes. The k^{th} class consists of individuals that carry label k, $k = 1, \ldots, K$. Let $Z_1, \ldots, Z_J$ be a finite number of non-overlapping subregions in G, $G = Z_1 \cup Z_2 \cup \cdots \cup Z_J$; $Z_i \cap Z_j = \emptyset$, $i \neq j$. For a subregion Z in G, we let $n_k(Z)$ denote the total number of individuals in Z that carry label k. In addition, let A_Z be the event that an individual is in Z and B_k denote the event that a label is of type k, $k = 1, \ldots, K$. We assume that there is a counting measure μ defined on G giving realized individual frequencies in the zones of G. Hence, the probability of the event A_Z, denoted by $\nu(Z)$, is clearly equal to $\nu(Z) = \frac{\mu(Z)}{\mu(G)}$.

For $K = 2$, the labels are denoted by $+$ and $-$. An individual carrying label $+$ is called a *point*. An individual in Z is a point with probability p, while in Z^c, the complement of Z, an individual is a point with probability q. Kulldorff (1997) considers $H_0: \ p = q$ verses $H_1: \ p > q$. Following the classical approach of the likelihood ratio test, Kulldorff develops a testing procedure for the spatial scan statistic, but its exact and limiting distribution is not, or cannot be, identified. As pointed out earlier in the introduction, the Kulldorff test procedure identifies the critical values for the test statistic by simulation and intensive computation. The authors in Soltani and Aboukhamseen (2015) translate the hypothesis given above into

$$\mathcal{A}: H_0: P_{k|Z} = P_{k|Z^c} \quad \text{vs} \quad H_1: P_{k|Z} > P_{k|Z^c},$$

where $P_{k|Z} = P(B_k|A_Z)$ and $P_{k|Z} = P(B_k|A_{Z^c})$, $k = 1, \ \ldots, \ K$. The authors prove in Soltani and Aboukhamseen (2015) that the hypotheses in (A) is equivalent to

$$\mathcal{B}: H_0: P_{Z|+} = \nu(Z) \quad \text{vs} \quad H_1: P_{Z|+} > \nu(Z).$$

Then we provided a classic test statistic for testing the hypotheses in $\mathcal{B}$.

In this article we assume that $n_+(G)$ is a random variable that follows a specific distribution function. By assuming that $n_+(G)$, the number of points in the whole region G, is random, the applicability of scan statistics is broadened to included situations where a census of the scan region is not available.

As it is customary in scan statistics to use the Binomial and the Poisson models, we assume that $n_+(G)$ has a Poisson distribution with mean $\lambda(G)$. Then, under the Binomial model assumptions of Kulldorff (1997), the unconditional distribution of $X_{Z,+}$ will be Poisson with mean $\lambda(G)\nu(Z)$. Thus the region Z is deemed a cluster of individuals of sign $+$ if in the following test

$$\mathcal{C}: H_0: \ E[X_{Z,+}] = \lambda(G)\nu(Z) \quad \text{vs} \quad H_1: \ E[X_{Z,+}] > \lambda(G)\nu(Z),$$

H_0 is rejected in favor of H_1. This intuitively makes sense, since under H_0 in $\mathcal{C}$,

$$E[X_{Z,+}]/\mu(Z) = \lambda(G)/\mu(G),$$

which states that the relative number of individuals in Z with label $+$ is proportional to the number of individuals in the whole region G with label $+$.

Download English Version:

https://daneshyari.com/en/article/7549024

Download Persian Version:

https://daneshyari.com/article/7549024

Daneshyari.com