A zero-inflated Poisson-based spatial scan statistic

André Cançado*, Cibele da-Silva and Michel da Silva

Universidade de Brasilia, Brasilia, Distrito Federal, Brazil

Objective

The scan statistic is widely used in spatial cluster detection applications of inhomogeneous Poisson processes. However, real data may present substantial departure from the underlying Poisson process. One of the possible departures has to do with zero excess. Some studies point out that when applied to zero-inflated data the spatial scan statistic may produce biased inferences. Particularly, Gómez-Rubio and López-Quílez (1) argue that Kulldorff's scan statistic (2) may not be suitable for very rare diseases problems. In this work we develop a closed-form scan statistic for cluster detection of spatial count data with zero excess.

Introduction

The spatial scan statistic proposed by Kulldorff (2) has been widely used in spatial disease surveillance and other spatial cluster detection applications. In one of its versions, such scan statistic was developed for inhomogeneous Poisson process. However, the underlying Poisson process may not be suitable to properly model the data. Particularly, for diseases with very low prevalence, the number of cases may be very low and zero excess may cause bias in the inferences.

Lambert (3) introduced the zero-inflated Poisson (ZIP) regression model to account for excess zeros in counts of manufacturing defects. The use of such model has been applied to innumerous situations. Count data, like contingency tables, often contain cells having zero counts. If a given cell has a positive probability associated to it, a zero count is called a sampling zero. However, a zero for a cell in which it is theoretically impossible to have observations is called structural zero.

Methods

We assume that the case-counts in the regions follow independent ZIP random variables with the same probability p of a structural zero. The ZIP model allows for additional flexibility when compared to the Poisson. When structural zeros occur, the ZIP model accounts, in average, for a reduction in the case-counts. The ZIP model allows for superdispersion or extra-Poisson variation, while the Poisson model often understimates the observed dispersion. Regarding the likelihood ratio test formulation for the ZIP model we describe our Scan-ZIP statistic considering (a) we know when a zero count is a structural one and (b) we do not know, for sure, whether or not a zero count is a structural one. For the latter case, the Scan-ZIP statistic is obtained through an EM procedure.

Table 1. Power, sensitivity and positive predictive value obtained by the three methods for the artificial cluster

Method	Power	Sensitivity	PPV
Poisson	0.6361	0.4828	0.6915
ZIP	0.9500	0.8671	0.8796
ZIP+EM	0.8612	0.7875	0.7513

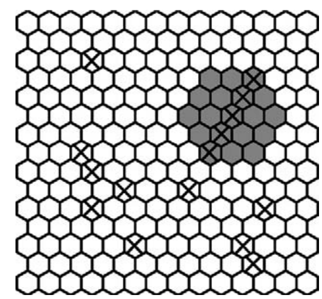

Fig. 1. Example of artificial cluster used to compare the three different scans. The ×'s indicate structural zeros.

Results

Our methodology was evaluated by means of a numerical case study. We constructed artificial clusters using a map consisting of 203 hexagonal cells arranged in a regular grid, 15 of which are structural zeros. An example can be seen in Fig. 1. Gray regions indicate the "true" cluster while the ×'s indicate structural zeros. We compare the Poisson, ZIP and ZIP+EM scans in terms of power, sensitivity and positive predictive value (PPV). The Scan-ZIP and Scan-ZIP+EM methods presented systematically and significantly better results when compared to the Scan-Poisson, as shown in Table 1 for the given cluster of Fig. 1. More examples and a real data application will also be presented.

Conclusions

The Scan-ZIP statistic has shown to be more suitable for the detection and inference of spatial clusters for data with zero excess as it outperforms the Scan-Poisson statistic in terms of power of detection, sensitivity and PPV.

Keywords

Spatial clusters; spatial scan statistic; zero-inflated Poisson

Acknowledgments

CQS acknowledges the support of brazilian agency CNPq.

References

1. Gómez-Rubio V, López-Quílez A. Statistical methods for the geographical analysis of rare diseases. Adv Exper Med Biol. 2010;686:151–71.
2. Kulldorff M. A spatial scan statistic. Commun Stat Theor Meth. 1997;26:1481–96.
3. Lambert D. Zero-inflated poisson regression, with an application to defects in manufacturing. Technometrics. 1992;34:1–14.

*André Cançado
E-mail: cancado@gmail.com

Citation: Emerging Health Threats Journal 2011, **4**: 11116 - DOI: 10.3402/ehtj.v4i0.11116