ISDS 2012 Conference Abstracts

Spatial Scan Statistics for Models with Excess Zeros and Overdispersion

Max Sousa de Lima[2], Luiz H. Duczmal*[1] and Letícia P. Pinto[1]

[1]Universidade Federal de Minas Gerais, Belo Horizonte, Brazil; [2]Universidade Federal do Amazonas, Manaus, Brazil

Objective

To propose a more realistic model for disease cluster detection, through a modification of the spatial scan statistic to account simultaneously for inflated zeros and overdispersion.

Introduction

Spatial Scan Statistics [1] usually assume Poisson or Binomial distributed data, which is not adequate in many disease surveillance scenarios. For example, small areas distant from hospitals may exhibit a smaller number of cases than expected in those simple models. Also, underreporting may occur in underdeveloped regions, due to inefficient data collection or the difficulty to access remote sites. Those factors generate excess zero case counts or overdispersion, inducing a violation of the statistical model and also increasing the type I error (false alarms). Overdispersion occurs when data variance is greater than the predicted by the used model. To accommodate it, an extra parameter must be included; in the Poisson model, one makes the variance equal to the mean.

Methods

Tools like the Generalized Poisson (GP) and the Double Poisson [2] may be a better option for this kind of problem, modeling separately the mean and variance, which could be easily adjusted by covariates. When excess zeros occur, the Zero Inflated Poisson (ZIP) model is used, although ZIP's estimated parameters may be severely biased if nonzero counts are too dispersed, compared to the Poisson distribution. In this case the Inflated Zero models for the Generalized Poisson (ZIGP), Double Poisson (ZIDP) and Negative Binomial (ZINB) could be good alternatives to the joint modeling of excess zeros and overdispersion. By one hand, Zero Inflated Poisson (ZIP) models were proposed using the spatial scan statistic to deal with the excess zeros [3]. By the other hand, another spatial scan statistic was based on a Poisson-Gamma mixture model for overdispersion [4]. In this work we present a model which includes inflated zeros and overdispersion simultaneously, based on the ZIDP model. Let the parameter p indicate the zero inflation. As the the remaining parameters of the observed cases map and the parameter p are not independent, the likelihood maximization process is not straightforward; it becomes even more complicated when we include covariates in the analysis. To solve this problem we introduce a vector of latent variables in order to factorize the likelihood, and obtain a facilitator for the maximization process using the E-M (Expectation-Maximization) algorithm. We derive the formulas to maximize iteratively the likelihood, and implement a computer program using the E-M algorithm to estimate the parameters under null and alternative hypothesis. The p-value is obtained via the Fast Double Bootstrap Test [5].

Results

Numerical simulations are conducted to assess the effectiveness of the method. We present results for Hanseniasis surveillance in the Brazilian Amazon in 2010 using this technique. We obtain the most likely spatial clusters for the Poisson, ZIP, Poisson-Gamma mixture and ZIDP models and compare the results.

Conclusions

The Zero Inflated Double Poisson Spatial Scan Statistic for disease cluster detection incorporates the flexibility of previous models, accounting for inflated zeros and overdispersion simultaneously.

The Hanseniasis study case map, due to excess of zero cases counts in many municipalities of the Brazilian Amazon and the presence of overdispersion, was a good benchmark to test the ZIDP model. The results obtained are easier to understand compared to each of the previous spatial scan statistic models, the Zero Inflated Poisson (ZIP) model and the Poisson-Gamma mixture model for overdispersion, taken separetely. The E-M algorithm and the Fast Double Bootstrap test are computationally efficient for this type of problem.

Keywords

Scan statistics; Zero inflated; Overdispersion; Expectation-Maximization algorithm

Acknowledgments

The authors acknowledge the grants provided by FAPEAM and CNPq.

References

[1] Kulldorff, M. (1999). Spatial scan statistics: Models, calculations and applications, in J. Glaz & N. Balakrishnan (eds), Scan Statistics and Applications, Springer Netherlands, pp. 303–322.

[2] Efron, B. (1986) Double Exponential Families and Their Use in Generalized Linear Regression, Journal of the American Statistical Association, 81, pp. 709-721.

[3] Cançado A., da Silva C. and da Silva M.(2011) A zero-inflated Poisson-based spatial scan statistic. Emerging Health Threats Journal, 2011;4:

[4] Zhang T., Zhang Z.; Lin G.(2012) Spatial scan statistics with overdispersion. Statistics in Medicine,31(8):762-774.

[5] Davidson, R. and J. G. MacKinnon (2001). "Improving the reliability of bootstrap tests", Queen's University Institute for Economic Research Discussion Paper No. 995, revised.

*Luiz H. Duczmal
E-mail: duczmal@ufmg.br