Exploratory Point Pattern Analysis for Modeling Biological Data

Stelios Zimeras, Department of Statistics and Actuarial-Financial Mathematics, University of the Aegean, Karlovassi, Samos, Greece

ABSTRACT

Data in the form of sets of points, irregular distributed in a region of space could be identified in varies biological applications for examples the cell nuclei in a microscope section of tissue. These kinds of data sets are defined as spatial point patterns and the presentation of the positions in the space are defined as points. The spatial pattern generated by a biological process, can be affected by the physical scale on which the process is observed. With these spatial maps, the biologists will usually want a detailed description of the observed patterns. One way to achieve this is by forming a parametric stochastic model and fitting it to the data. The estimated values of the parameters could be used to compare similar data sets providing statistical measures for fitting models. Also a fitted model can provide an explanation of the biological processes. Model fitting especially for large data sets is difficult. For that reason, statistical methods can apply with main purpose to formulate a hypothesis for the implementation of biological process. Spatial statistics could be implemented using advance statistical techniques that explicitly analyses and simulates point structures data sets. Typically spatial point patterns are data that explain the location of point events. The author's interest is the investigation of the significance of these patterns. In this work, an investigation of biological spatial data is analyzed, using advance statistical modeling techniques like kriging.

Keywords: Biological Point Data, Point Pattern Analysis, Point Process, Spatial Point Analysis, Spatial Statistics

INTRODUCTION

Spatial statistics is concerned with the study of spatially referenced data combined with appropriate statistical models and process. In spatial analysis, point pattern process has been a powerful method, to investigate the relationship between different locations of a particular event (locations of the nuclei cells) applying advance statistical technique (estimation of the intensities using measures like nearest neighbors distance for the investigation of the homogeneity). Some important works have been introduced spatial statistical methods to evaluate biological data and reproduce predictions for the spatial structure of the homogeneous regions (Cressie, 1993; Diggle, 1983; 2003; Ripley, 1981).

Figure 1. General spatial model
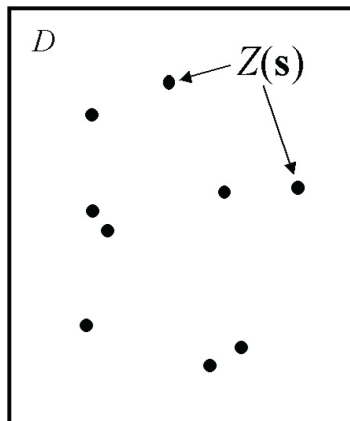
In this application, investigation of spatial locations estimates is proposed based on kernel smoothing (on the position of the centers of nuclei) using the K-Ripley modified functions (Dale, 1999; Diggle, 1983; 2003).

Spatially data consist of measurements for a stochastic process at a set of sampled locations. The objective of the analysis is to model the variability of the associated random variable over space and to examine statistically factors that explain this behaves. Models are also used to predict the values of the random variable at locations in the study area that were not sampled (this is called spatial interpolation) (Diggle, 1983; 2003; Cressie, 1993)

A stochastic process that takes place over space could be represented by the epicenters of locations leading us to a variety of complex biological phenomena. This variability of the stochastic process due the time and space, introduces point patterns.

The object of point pattern analysis is to study the stochastic process given by the observed point pattern and especially to investigate if the pattern is regular or clustered. If none of those cases applies then we say that the pattern is random.

Most recently, statistical techniques applying to point patterns are considered generating estimations of the spatial location for predictions about the future phenomena. Finally they provide a variety of methods for the visualization of point data. They allow us simultaneously to view the point patterns, explore structure in data by estimations appropriate models and test hypothesis relating to the process considering the observed event distribution.

In this work, an investigation for biological data is analyzed, using advance statistical modeling techniques. Statistical models could be useful for the prediction of the locations in the future.

SPATIAL STATISTICS MODELING

Spatial modelling is analysing various data of spatial process: **1**. geostatistical data, **2**. lattice data, and **3**. point patterns (Figure 1).

The notations that will be used are based on (Cressie, 1993; Diggle, 1983; 2003; Ripley, 1981). The study region (domain) is given by D. Usually D is a subset of 2-dimensional space, but it could be 1-dimensional or 3-dimensional or even beyond so $D \subset \Re^d$, where d is the dimension we choose. The vector **s** denotes the data location. Locations in D are denoted by the vector **s**. For example, in 2-dimensional space, **s** will have 2 components containing the coordinates (x, y), such as latitude and longitude. At location **s**, we obtain some value z. So z(**s**)

Related Content

Genomatrices and the Genetic Octet Yin-Yang-Algebras

Sergey Petoukhovand Matthew He (2010). *Symmetrical Analysis Techniques for Genetic Systems and Bioinformatics: Advanced Patterns and Applications (pp. 132-167).*

www.irma-international.org/chapter/genomatrices-genetic-octet-yin-yang/37900

Computational Analysis and Characterization of Marfan Syndrome Associated Human Proteins

K. Sivakumar (2010). *Biocomputation and Biomedical Informatics: Case Studies and Applications (pp. 143-157).*

www.irma-international.org/chapter/computational-analysis-characterization-marfan-syndrome/39609

Applying Bayesian Networks in the Early Diagnosis of Bulimia and Anorexia Nervosa in Adolescents: Applying Bayesian Networks in Early Diagnosis in Adolescents

Placido Rogerio Pinheiro, Mirian Caliope Dantas Pinheiro, Victor Câmera Damasceno, Marley Costa Marques, Raquel Souza Bino Araújoand Layane Mayara Gomes Castelo Branco (2018). *Applying Big Data Analytics in Bioinformatics and Medicine (pp. 364-379).*

www.irma-international.org/chapter/applying-bayesian-networks-in-the-early-diagnosis-of-bulimia-and-anorexia-nervosa-in-adolescents/182955

SPCCTDM, a Catalogue for Analysis of Therapeutic Drug Monitoring Related Contents in the Drug Prescription Information

Sven Ulrich, Pierre Baumann, Andreas Conca, Hans-Joachim Kuss, Viktoria Stieffenhoferand Christoph Hiemke (2010). *International Journal of Knowledge Discovery in Bioinformatics (pp. 1-11).*

www.irma-international.org/article/spcctdm-catalogue-analysis-therapeutic-drug/45162

In Silico Prediction of Blood Brain Barrier Permeability: A Support Vector Machine Model

Zhi Wang, Aixia Yanand Jiaxuan Li (2011). *Interdisciplinary Research and Applications in Bioinformatics, Computational Biology, and Environmental Sciences (pp. 155-171).*

www.irma-international.org/chapter/silico-prediction-blood-brain-barrier/48373