Modeling earthquake data using spatial statistics techniques

S. ZIMERAS
University of the Aegean
Department of Statistics and Actuarial-Financial Mathematics
G.R.832 00, Karlovassi, Samos, Greece

Abstract: Spatial statistics is concerned with statistical methods that explicitly analyses spatial arrangement of the data. When analyzing these kinds of data, we might be interested investigating the temporal data process that generated the data. Typically spatial point patterns are data that made up of the location of point events. We are interested in whether or not their relative locations represent a significant pattern. For example, seismologists have been showing the distribution of earthquakes in a region. They would like to known if there is any pattern that might help them make predictions about future earthquakes. In this work, an investigation of earthquakes spatial data is analyzed, using specific statistical modeling techniques.
Keywords: Spatial point analysis, earthquakes data, spatial statistics.

1. Introduction

Spatial statistics is concerned with the study of spatially referenced data combined with appropriate statistical models and process. For many researchers, point pattern analysis has been a powerful tool, to investigate the relationship between different locations of a particular event (earthquake) applying specific statistical technique (estimation of the intensities using measures like nearest neighbors distance). Few works have been have been introduced spatial statistical methods to evaluate earthquake data and reproduce predictions for the future [1, 2].

Spatially data consist of measurements for a spatially 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 variability. 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).

A stochastic process that takes place over space gives rise to events, which are represented by their locational coordinates. The epicenters of earthquakes, for example, can be considered as events in space. The stochastic process that gives rise to these events is governed by a variety of complex geological phenomena. A collection of events in a study region constitutes a point pattern. The object of point pattern analysis is to study the stochastic process that gives rise to the observed point pattern. To this end, it is important to determine in a statistical sense whether the events in the point pattern under study are regularly spaced or clustered. The former implies repulsion between events while the latter implies attraction. If none of those cases applies then we say that the pattern is random.

Most recently, statistical techniques applying to point patterns are considered because of the developments of geographical information systems (GIS). Under these systems, it could be generated predictions of the spatial location for predictions about the future phenomena. Finally they provide a variety of tools 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 of earthquakes spatial data is analyzed, using specific statistical modeling techniques. Statistical models could be useful for the prediction of the earthquake locations in the future.

2. Point Processes

A point process is a process of locations of events taking place in some space *X*. Each event has associated with a mark, taking place in some mark space *Y* [3]. The term event (points) can refer to any spatial phenomena that occur at a point location. The epicenters of earthquakes, for example, can be considered as events in space

A point pattern is a set of point locations $\mathbf{s}=\{s_1, s_2,\ldots,s_n\}$ in a specific sub-region A. [4]. The random variable N(A) is the number of events in the set $A\subset X$ introducing a random process. Since a random process have been considered, the behave in terms of first-order (related to expected values) and second-order properties (related to the covariance) could be analyzed [5]. In the case of point pattern analysis, second-order properties are perhaps less informative [6] but are still important factor for the analysis of point patterns [4, 7].

First-order properties are described in terms of intensity $\lambda(\mathbf{s})$, of the process, which is the mean number of events per unit at the point **s** [4]. This is defined as $\lambda(s)=\lim_{ds\to 0}\left\{\frac{E(N(ds))}{ds}\right\}$, where d(**s**) is a small region around the point **s**, E(.) is the expected value, and N(d(**s**)) is the number of events in the small region. The second-order properties, of a spatial point process introduce the relationship between numbers of events in pairs of sub-regions as $\gamma(s_i,s_j)=\lim_{ds_i,ds_j\to 0}\left\{\frac{E(N(ds_i)N(ds_j))}{ds_i ds_j}\right\}$, with similar notations. A point process is stationary if the intensity is constant over A, so $\lambda(\mathbf{s}) = \lambda$ and $\gamma(s_i,s_j)= \gamma(s_i-s_j)= \gamma(d)$ (depending only on direction and distance). Finally, the process is defined as isotropic if the second-order intensity depends only on the distance between s_i and s_j (does not depend on the direction).

3. Visualizing spatial point process.

The most effective way to visualise spatial pattern data is to plot them as a dot map. A dot map is a region over which the events are observed as points. The dot map has long been one of the most popular cartographic tools of geographers. It is popular since it displays spatial distributions with both clarity and simplicity. Fig. 1 represents the locations of the epicenters of the earthquakes where observations in this data frame are 2049 earthquake epicenters in the San Francisco Bay area for the time period 1962-1981. Obviously the pattern appears clusters.

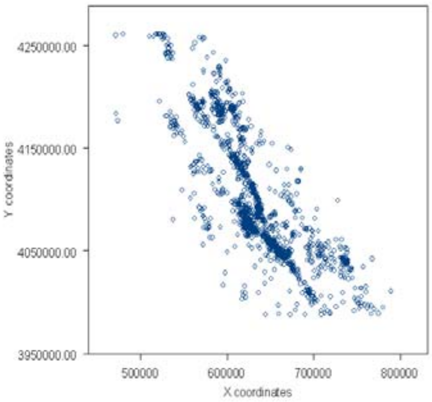

Fig. 1. Spatial locations of the epicenters of earthquakes

Once we have obtained a dot map, the next step is to see whether the resulting spatial distribution exhibits any recognizable patterns. The search for patterns should be directed ideally by hypotheses proposed in previous studies, particularly studies that have developed theories and models. In this way, seismologists can employ specific techniques in order to search for specific types of patterns. In searching for patterns it must never be forgotten that every pattern is the result of some process at a given point in time and space, and the analysis of point processes ought to incorporate some idea on how the pattern evolved. Thus, the patterns we observe can be thought of as abstractions because in reality, time never stops. It is a continuous on-going process. There are no patterns just processes. Thus, it is important that we view our point process as simply the visual expression at one point in time of the process operating continuously over space.

4. Estimating the intensities

One way to summarised the events in a spatial point pattern is to divide the regions into sub-regions of equal areas (quadrats). By counting the number of events inside each quadrat, we end with a measure (frequency or histogram) that summarizes the spatial pattern. The intensity of a point pattern is the mean number of points per unit area. Intensity plots display a smooth estimate of intensity for a spatial point pattern.

One way to estimate the intensities of a point pattern is the binning process, where it is uses a two-dimensional histogram to form rectangular bins. The counts in these bins are smoothed using a loess smoothing algorithm. Using the binning method for the data as explained in the sequence above yielded the following plot (Fig. 2). Easily, it can be seen that the two main clustered regions are described, with especially high intensities in the middle of the area. Regions with high intensities represent intensive frequencies for the epicenters of the earthquakes. Fig. 3 illustrates the 3D map of the 2D estimate intensities.

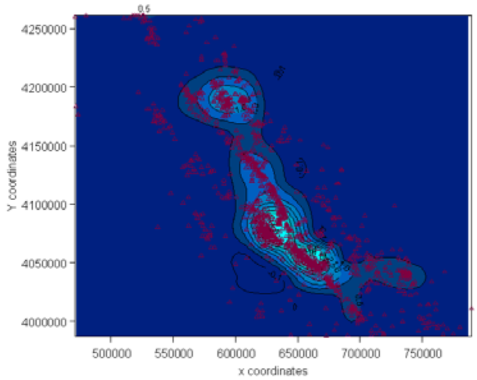

Fig. 2. Intensities using binning process.

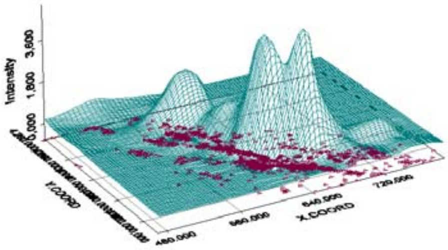

Fig.3. 3D of the Intensities using binning process.

Another sophisticate method could be introduced applying kernel process to get an estimation of the intensity that is smoother than the quadrat method. Based that method, an estimate of the intensity is given by: $\hat{\lambda}_h(s)=\frac{1}{\delta_h(s)}\sum_{i=1}^{n}\frac{1}{h^2}k\left(\frac{s-s_i}{h}\right)$, where k is the kernel and h is the bandwidth.

The choice of the bandwidth (h) is purely subjective. It is up to the researcher to investigate the results using several different bandwidths and selecting the one that best represents the pattern. However, it is important to remember that when h0 is large, the kernel estimator produces a smooth estimate of the density function (i.e. small variance, large bias). If h is small, the kernel estimator produces a rough estimate of the density function (i.e. large variance, small bias). Also, the number of bins can be altered. The numbers of bins that are selected in each direction determine how smooth the kernel looks. Results are given in Fig. 4 (for 2D presentation) and Fig. 5 (for 3D presentation), with bandwidth= 600 m.

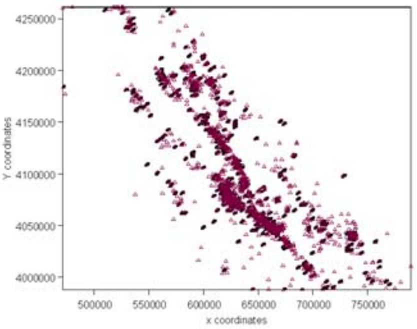

Fig. 4. Intensities using kernel process.

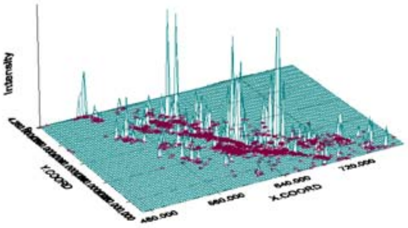

Fig. 5. 3D of the Intensities using binning process.

.

Compared with the previous method, it is clear that the prediction of the earthquakes locations are appared clearly using the kernel method compered with the binnig process. Also the 3D graph is sharper, justifing the highrest spatial locations of the epicenters.

5. Conclusions

Spatial statistics is a powerful tool, either for presentation of existing spatial phenomena like earthquakes, or investigating relationships between different locations of a particular event (earthquake) applying specific statistical technique. In this work, we have described how spatial point patterns can be represented statistically through first and second order properties. Kernel estimation was suggested as an appropriate method for prediction of mean intensities for earthquake data.

Reference:

1. Choi E. and Hall P. (1999): Nonparametric approach to analysis of space-time data on earthquake occurrences, *Journal of Computational and Graphical Statistics, 8: 733-748.*
2. Ogata Y. (1999): Seismicity analysis through point-process modeling: a review, *Pure and Applied Geophysics, 155: 471-507.*
3. Daley and Vere-Jones (1988): An introduction to the Theory of point processes, Springer-Verlag, N.Y.
4. Diggle P. J. (1983): *Statistical Analysis of Spatial Pattern Point*, Academic Press, London
5. Bailey T.C. and Catrell A.C. (1992): Interactive spatial data analysis, London, Longman Scientific and Technical.
6. Baddeley A.J. and Silverman B.W (1984): A cautionary example of the use of second order methods for analyzing point patterns, Biometrics, 40, 1089-1093.
7. Ripley B.D. (1981): Spatial Statistics, N.Y. John Wiley and Sons