Lattice Models for the analysis of Urban Crime

Enrico di Bella, Lucia Leporatti, Luca Persico[1]

Abstract
The spatial analysis of Urban Crime is generally made using crime mapping techniques, which are mainly representations of crime dispersion over a specific urban area without any statistical modelling of its dependence from the urban structure of the city or any group of socio-demographic variables. Using a set of measurements (called Space Syntax) based on a segmentation of all the roads of the topographic map of a city into axes and a set of socio-demographic variables, in this work a highly populated district of the City of Genoa has been analyzed, studying crime in the context in which it happens, interpreting the roads network as a graph or a lattice.

Key words. Crime Mapping, Lattice Models, Space Syntax, Spatial Models.

1. Introduction

In recent years analyzing crime data through spatial statistical methods has become a more and more common strategy used by the police and by the local governments to prevent and reduce crime events. Despite this increasing interest in crime prevention, most of the literature focuses the analysis on a graphical study of crime events: through the use of modern crime mapping techniques it is easy to visualize, map and analyze crime incidents patterns [2]. The aim of crime mapping is to identify crime hot spots in order to target police responses to these crime concentrated areas. The limit of such an analysis is the fact that it doesn't take into account important features responsible for urban security, particularly the urban and spatial characteristics of streets and the demographic and socio-economic features of the region analyzed. The purpose of this paper is to overcome the limits of a mere representation of crime data, creating a spatial lattice model to study the relationship between the number of crimes committed in a district of Genoa as function of the spatial features of streets and of the socio-economic characteristics of the district.

2. An integrated database on crimes

Data available derive from three databases owned by different Institutions and their converging to a unique dataset has been possible thanks to an official collaboration between the Municipality of Genoa and the University of Genoa. The first dataset,

[1] University of Genoa, Department of Economics and Quantitative Methods,
edibella@economia.unige.it, lucia.leporatti@unige.it, persico@economia.unige.it

provided by the Territorial Office of the Municipality of Genoa, is the urban graph of a highly populated district of the City of Genoa: each street can be identified using a coordinate system of initial and final points. Each street, called axis, is therefore part of a connected urban network, or graph, which has been analyzed according to a technique named Space Syntax Analysis (SSA). SSA, deeply inspired by graph theory, provides a method for partitioning a spatial system into relatively independent but connected subspaces so that the importance of these subspaces can be measured in terms of their relative nearness or accessibility [5]. A vast number of variables can be measured on each single axis of an Urban Graph through the means of the SSA and, among them, we chose a set which proved itself to be particularly effective in this context [4]: Integration (a measure of the centrality of an axis in the urban graph), Choice (the ratio of geodetic paths in the urban graph including a specific axis), Control (the capability of a street to give access to other streets) and Connectivity (the number of axes connected to a specific axis). The second dataset contains the 1,681 complaints collected by the local Carabinieri Station from 01/01/2009 up to 27/07/2010. These data are classified according to a ministerial standard definition of types of crime in three main categories: violent crimes (e.g. attacks and murders; 49 complaints), crime against property (thefts, robberies; 916) and damages and fires (716). The third and last dataset, provided by the Demographic Office of the Municipality of Genoa, refers to demographic and economic characteristics of the district at street level (e.g. residents per axis, number and types of commercial activities, number of recreation activities). These variables have gone through additional re-elaboration, and, in particular, in this work we will introduce the variables Residential Vocation, obtained as the ratio between the number of shops with a mainly residential vocation (i.e. whose main business is for local residents) and shops with a mainly commercial vocation (i.e. whose main business attracts residents from other areas) and Attractivity, a roughly estimated size of the clientele attracted by the various categories of commercial venues.
Although data are available at single-street level, in order to protect the privacy of the information, data are herein presented in an aggregate form, without any particular mention of the names of streets.

3. Spatial analysis of crime using a lattice model

A typical situation in the analysis of spatial data is the presence of spatial autocorrelation as close regions, areas or, in this case, axes, tend to have a similar behaviour of certain variables (e.g. crimes, socio-demographic characteristics). In this paragraph we propose the use of a spatial lattice model to express the number of crimes occurred in each street as function of both socio-economic and demographic variables and of the spatial configuration of the roads measured through the means of the aforesaid Space Syntax Variables, taking into account the spatial neighbours structure of data. Modelling spatial interactions that arise in spatial data is usually performed by incorporating the spatial dependence into the covariance structure through a spatial autoregressive model [3]: the two approaches that are commonly used in practice are the Conditional Autoregressive model (CAR) and Simultaneous Autoregressive model (SAR) [1]. In order to define the relationships between observations in the lattice model, we created a spatial neighbour structure and the consequent weights matrix to assign spatial weights to each relationship. After evaluating different types of neighbour matrix we have chosen a distance-based matrix defining as neighbours of an axis all the axes contained into a prefixed distance from the midpoint of the road considered.

From a statistical point of view, it is possible to account for correlated observations by considering a structure of the following kind in the model. If the vector of response variables is multivariate normal, we can express the model as follows:

$$Y_i = \mu_i + \delta_i \quad (1)$$

where Y_i is the random process in i, μ_i is the mean in i, and δ_i is the i-th value of a vector of normally distributed random errors with zero mean and covariance matrix:

$$\boldsymbol{\Sigma} = [(\boldsymbol{I} - \rho \boldsymbol{N})^T \boldsymbol{D}^{-1}(I - \rho \boldsymbol{N})^{-1}\sigma^2 \quad (2)$$

where ρ represents the spatial autocorrelation and σ^2 is the variability measure, $\boldsymbol{N}$ is a weighted neighbourhood matrix and $\boldsymbol{D}$ is a diagonal matrix used to account for nonhomogeneous variance of the marginal distributions. The small scale variation due to interactions with neighbours is modelled by fitting an autoregressive covariance model to $\boldsymbol{\Sigma}$. To take into account the covariance structure of data we can express the model with the following equation:

$$Y = \boldsymbol{X}\beta + \rho \boldsymbol{N}(y - \boldsymbol{X}\beta) + \boldsymbol{W}^{\frac{1}{2}}\varepsilon \quad (3)$$

where $\boldsymbol{X}\beta$ is the linear trend, $\rho \boldsymbol{N}(y - \boldsymbol{X}\beta)$ the covariance structure and $\boldsymbol{W}^{\frac{1}{2}}\varepsilon$ the noise of model.

4. Preliminary Results and further work

The preliminary results given in Table 1 refer to model (3) which has been estimated time by time using as dependent variables the total number of reported crimes and the three aforesaid subcategories of crimes. The selected set of independent variables, belonging to each of the thematic areas covered in this research (demography, spatial configuration, commercial fabric), looks explicative enough to produce an R^2 adjusted value ranging from 0.6 up to 0.85 depending on the categories of crime analyzed. The spatial dependence structure seems to be absent as the Moran's I test for spatial autocorrelation of the residuals [6] is weakly significant (p-value = 0.08) and the statistical significance of the ρ coefficient of autocorrelation of the model (see Table 1) is not significant except for the violent crimes model. Among the variables used for the presented models, three present a particularly interesting behaviour. Integration, which is usually associated with intense pedestrian flows in public spaces, looks significantly and positively associated with crime as well as the Number of Residents, suggesting that the most crime-ridden streets in the area are intensely residential, central and full of pedestrians. Although not significant, Attractivity looks to have a deterrent effect on crime confirming that the relationship between pedestrian presence in public spaces and crime is a complex one. We plan to extend the analysis to a wider research area in order to improve the statistical significance and confirm this preliminary result.

These preliminary results should be read carefully for various reasons. Firstly, crime data are count data and the normal assumptions for the error component is not fully realistic. Secondly, the spatial dependence structure is still under focus: we expect that the weakness of spatial autocorrelation is due to the fact that a relevant part of it is implicitly specified inside the Space Syntax variables used in model (3) and further work will be done in this direction. Moreover, most axes recorded zero crimes over the period and the implementation of zero inflated count data models might be fruitfully used. A key role in the next model specifications is then covered by the autocorrelation component in the mentioned terms.

Table 1: SAR model (3): results

	Total crimes			***Crimes against property***		
Coefficients	***Value***	***Std Err***	***Pr(>│t│)***	***Value***	***Std Err***	***Pr(>│t│)***
Intercept	-16.83	6.20	0.00665	-6.36	2.80	0.02336
Choice	-43.77	30.90	0.15654	-25.28	13.99	0.07083
Connectivity	-104.43	63.04	0.09761	-41.83	28.66	0.14448
Control	53.88	36.53	0.14020	13.01	16.65	0.43470
IntegrationR3	60.12	18.35	0.00105	26.67	8.30	0.00131
Number of residents	65.34	11.59	0.00000	23.07	5.32	0.00000
Attractivity	-19.83	13.07	0.12913	-11.55	6.02	0.05514
Recreative activities	77.59	37.74	0.03980	37.29	17.28	0.03096
Number of shops	61.57	41.31	0.13604	54.13	18.91	0.00420
Residential Vocation	-10.27	12.11	0.39623	-4.98	5.58	0.37199
AIC		863.31			700.73	
R^2-adjusted		0.81			0.85	
Rho	0.05		0.21224	0.04		0.4146
	Damages and fires			***Violent crimes***		
Coefficients	***Value***	***Std Err***	***Pr(>│t│)***	***Value***	***Std Err***	***Pr(>│t│)***
Intercept	-9.08	3.12	0.00364	-0.74	0.30	0.01214
Choice	-11.04	15.61	0.47937	-4.83	1.46	0.00098
Connectivity	72.32	31.94	0.02358	-6.92	2.97	0.01987
Control	38.75	18.55	0.03673	4.51	1.71	0.00844
IntegrationR3	31.44	9.26	0.00068	2.37	0.87	0.00632
Number of residents	36.08	5.91	0.00000	1.97	0.54	0.00024
Attractivity	-11.08	6.69	0.09799	-0.24	0.60	0.68547
Recreative activities	54.89	19.24	0.00433	0.65	1.76	0.71090
Number of shops	-30.18	21.05	0.15153	4.23	1.92	0.02795
Residential Vocation	-3.22	6.20	0.60434	-0.41	0.56	0.45522
AIC		723.00			226.12	
R^2-adjusted		0.60			0.64	
Rho	0.04		0.34103	0.08		0.04694

References

1. Besag, J. : Spatial Interaction and the Statistical Analysis of Lattice System, Journal of the Royal Statistical Society, Series B, Vol 36 (2): 192-236, (1974)
2. Boba, R. : Crime Analysis with Crime Mapping, Sage, (2009)
3. Cressie, N.A.C. : Statistics for Spatial Data, Rev Ed., Wiley (1993)
4. di Bella, E., Corsi, M., Persico, L. : A Multivariate Analysis Of The Space Syntax Output For The Definition Of Strata In Street Security Surveys, DISEFIN Series of economic working papers, 5, (2011)
5. Hillier, B., Hanson J. : The Social Logic of Space, Cambridge University Press, (1984)
6. Moran, P.A.P.: Notes on Continuous Stochastic Phenomena, Biometrika 37: 17-33, (1950)