

Stima della densità su strutture a rete: analisi di distribuzioni di punti nello spazio urbano

Giuseppe BORRUSO (*)

(*) Dipartimento di Scienze Geografiche e Storiche, Sezione di Geografia Economica e Politica del Territorio,
Università degli Studi di Trieste, Piazzale Europa, 1 34127 - Trieste,
Tel. +39 040 558 7008, fax. +39 040 558 7009, e-mail. giuseppe.borruso@econ.units.it

Riassunto

Le analisi sulla distribuzione di elementi puntuali nello spazio, riguardanti ad esempio gli schemi di insediamento basati sul numero civico piuttosto che la posizione di centri commerciali o altri oggetti generalizzabili come punti, fanno generalmente ricorso ad algoritmi e procedure che tendono a calcolare le distanze in modo euclideo, confidando quindi su di una generale struttura omogenea e isotropica dello spazio. Analisi statistiche spaziali quali ad esempio la *Kernel Density Estimation* o le funzioni *K*, utilizzate anche per valutazioni di *geomarketing*, pur consentendo di stimare in modo efficace la densità nella distribuzione di fenomeni puntuali nello spazio, molto spesso presentano limiti quali quelli appena accennati. La presente ricerca si basa sull'esame della distribuzione di elementi puntuali nello spazio a partire da una struttura a rete, abbandonando pertanto le usuali ipotesi di omogeneità dello spazio e considerando spazi di rete quali ambiti privilegiati per l'osservazione della distribuzione di punti. Molti fenomeni umani, infatti, sono distribuiti su spazi non omogenei, quali ad esempio quelli urbani un cui la loro posizione segue una configurazione guidata da strutture a rete. La ricerca si focalizza sulla *point pattern analysis*, che sia nell'analisi spaziale che nella scienza dell'informazione geografica risulta adottata e utilizzata comunemente. Nel presente lavoro viene proposta un'estensione della *Kernel Density Estimation* (KDE), chiamata *Network Density Estimation* (NDE). L'idea consiste nel considerare una funzione *kernel* su distanze di rete invece che su quelle euclidee. Ciò consente l'evidenziazione di *cluster* lineari di elementi puntuali lungo le reti e l'identificazione di una più precisa superficie di densità relativa a fenomeni legati alle rete.

Abstract

Analyses on point distributions on space, usually concerning residents' locations based on address point as well as other features as malls and bank ATMs, generally use algorithms and procedures calculating Euclidean distances over space, thus considering space as homogeneous and isotropic. Spatial statistical techniques as Kernel Density Estimation or K functions, that are also used for spatial marketing analyses, usually allow to estimate efficiently the densities of point distributions over space, however present limitations due to such an initial assumption.

This research is focused on examining point pattern distributions over a network, therefore abandoning the usual hypotheses of homogeneity and isotropy of space and considering network spaces as frameworks for the distribution of point patterns.

Several human-related phenomena are in fact distributed over non-homogeneous spaces, as urban activities where their locations follow a network-led structure. The research is focused on point pattern analysis, now commonly used in both spatial analysis and Geographic Information Science. Here an extension of Kernel Density Estimation (KDE) called Network Density Estimation (NDE) is proposed and tested. The idea is considering a kernel function over network distances rather than on Euclidean ones. That allows highlighting linear clusters of point features along a network.

Introduzione

La ricerca qui presentata intende focalizzarsi sulla possibilità di evidenziare schemi di distribuzione di elementi puntuali in ambito urbano collegati alla struttura reticolare del sistema viario. Si tratta pertanto di modificare le più tradizionali analisi di densità, basate su metodologie di studio che semplificano lo spazio quale omogeneo e isotropico, considerando quindi alcune delle limitazioni di carattere fisico date dal territorio, e tra queste appunto la struttura del sistema viario di una città. L'analisi presentata considera quindi la distribuzione di alcune tipologie di servizi a livello urbano, nella fattispecie gli sportelli bancari e le agenzie assicurative, mentre precedentemente la metodologia qui proposta è stata sperimentata per evidenziare la densità delle infrastrutture viarie all'interno di una città (Borruso, 2005).

Analisi di densità sulla distribuzione dei servizi sono state proposte per osservare la forma della città e individuare aree funzionali. Autori quali Burgess e Hoyt (1939) e Harris e Ullman (1945) hanno studiato il CBD – *Central Business District* nelle aree centrali delle città, mentre Alonso (1960) ha evidenziato la decrescenza nella densità dell'uso del suolo urbano; in seguito nuovi modelli per indicare la natura monocentrica o policentrica delle città sono stati proposti (Hoch, Waddel, 1993). Più di recente, strumenti GIS (Batty, Longley, 1994) e di analisi spaziale (Cuthbert, Anderson, 2002) sono stati proposti per l'analisi dello sviluppo delle aree e delle funzioni urbane, anche con l'utilizzo di appositi indicatori per legare la dotazione di servizi urbani alla forma della città (Borruso, Schoier, 2004).

Analisi di densità di distribuzioni di punti

Diverse metodologie sono state proposte da autori provenienti dal campo geografico e statistico per analizzare distribuzioni di elementi puntuali nello spazio. Bailey e Gatrell (1995) considerano *eventi*, riferiti a fenomeni osservati su una distribuzione di punti, e *punti*, riferiti a tutti gli altri luoghi nell'area di studio. La semplice osservazione della distribuzione di eventi nello spazio può fornire alcune informazioni iniziali sulla struttura della distribuzione, ma strumenti analitici più raffinati sono necessari per analisi più approfondite, e particolarmente per identificare la presenza di *clusters* o di regolarità nella distribuzione.

L'analisi dei quadrati costituisce una delle metodologie di ordinamento dello schema distributivo di eventi all'interno di una regione R . La procedura comporta la suddivisione dell'area di studio in sotto-regioni aventi superfici uguali o equivalenti, o quadrati (anche se sono utilizzabili altre forme, quali triangoli, esagoni o altre strutture poligonali). Il passo seguente consiste nel conteggio del numero di eventi che cadono all'interno di ogni sotto-regione (quadrato) al fine di semplificare e raggruppare la distribuzione di punti. In questo modo la numerosità di eventi diventa un attributo di ogni quadrato. Diventa così possibile rappresentare la distribuzione per mezzo di area omogenee e facilmente confrontabili, rendendo possibili analisi di densità per mezzo di metodi computazionali facili da utilizzare. Analisi di questo tipo possono essere svolte inoltre utilizzando le funzioni standard presenti all'interno dei pacchetti GIS sul mercato.

Il metodo presenta tuttavia alcuni svantaggi, quali una certa perdita di informazioni rispetto ai dati di partenza, unita a diversi livelli di arbitrarietà che derivano dalla scelta della dimensione dei quadrati, dall'orientamento e dalla scelta del punto di origine della griglia. Analisi differenti possono essere perciò effettuate variando l'origine della griglia o la dimensione dei quadrati. Una soluzione può essere quella di considerare gli eventi per ogni area entro una 'finestra' mobile: fissato un raggio adeguato, si ipotizza di centrarlo su un certo numero di luoghi nello spazio, dove gli eventi sono organizzati nell'ambito di una griglia sovrapposta alla regione di studio R . In questo modo viene fornita una stima dell'intensità in ogni punto. La stima della variazione di intensità è pertanto più 'arrotondata' rispetto alla semplice rappresentazione dell'analisi dei quadrati.

Questa ultima metodologia si avvicina alla *Kernel Density Estimation* (KDE), una 'funzione mobile a tre dimensioni che pesa gli eventi entro la sua sfera di influenza, a seconda della loro distanza dal punto dal quale viene stimata l'intensità' (Gatrell, 1994).

$$\hat{\lambda}(s) = \sum_{i=1}^n \frac{1}{\tau^2} k\left(\frac{s-s_i}{\tau}\right) \quad [1]$$

con $\hat{\lambda}(s)$ la stima dell'intensità della distribuzione di punti, misurata nel punto s ; s_i è l' i th evento, e $k(\cdot)$ rappresenta la funzione di kernel e τ la soglia. Tale soglia consente di ottenere superfici più o meno arrotondate del fenomeno, che rappresentano la variazione del fenomeno su di una certa area, consentendo di analizzare la sua distribuzione alle diverse scale. Tre risultano essere i passaggi necessari per applicare la procedura (Chainey e altri, 2002):

- una griglia viene stesa sulla regione di indagine e sulla distribuzione di punti;
- una funzione mobile a tre dimensioni visita ogni cella e calcola il peso per ogni punto nell'ambito di un certo raggio (soglia);
- vengono calcolati dei valori di cella sommando i valori ottenuti per tutte le superfici circolari calcolate su ogni 'visita'.

Per quanto riguarda le funzioni tridimensionali, diversi metodi di interpolazione possono essere utilizzati per stimare l'intensità del fenomeno. I tipi più utilizzati comprendono la distribuzione normale, che genera una stima della densità sull'intera regione, e distribuzioni che stimano la densità soltanto all'interno del raggio (soglia) selezionato: le distribuzioni sono la uniforme, la quartica, la triangolare e la negativa esponenziale. Di queste, le prime due tendono ad arrotondare maggiormente le superfici di densità, mentre le ultime due producono strutture con 'picchi' e 'vallate' più pronunciati (Levine, 2004).

Network density estimation: l'algoritmo

Le analisi basate sul KDE tendono a evidenziare soprattutto *cluster* circolari all'interno di una particolare regione dello spazio. I *cluster* possono tuttavia formarsi seguendo diversi orientamenti distributivi, quali strutture spaziali reticolari. Di seguito viene presentata una procedura per considerare tale schema reticolare per lo studio di distribuzioni di punti. La funzione *kernel* viene modificata, in modo tale da considerare delle funzioni di ricerca 'a rete', che calcolino cioè la distanza di soglia seguendo il reticolo stradale, piuttosto che considerare aree circolari, basate sulla distanza euclidea. Il confronto con lo stimatore *kernel* verrà quindi effettuato considerando l'interpolazione uniforme, più vicina concettualmente alla presente proposta.

Le fasi dell'algoritmo:

1. selezione di una distribuzione di punti nello spazio all'interno di un'area (es.: popolazione per numero civico; posizione di sportelli bancari; uffici pubblici);
2. costruzione di una griglia regolare sull'area di studio;
3. costruzione dei centroidi delle celle appartenenti alla griglia regolare sovrimposta all'area di studio;
4. definizione di un raggio di ricerca (soglia di osservazione);
5. costruzione di una rete (es. rete stradale);
6. calcolo delle 'aree di servizio' di ogni centroide sulla rete e secondo il raggio di ricerca definito (figura 1);
7. overlay (=intersezione spaziale) delle aree di servizio con la distribuzione di punti;
8. conteggio dei punti (eventi) all'interno di ogni area di servizio;
9. attribuzione del conteggio degli eventi (eventualmente pesato: densità relativa) al centroide di ogni cella pertinente un'area di servizio, e quindi a ogni cella

9a. [se necessario, prima di visualizzare la superficie di densità può essere effettuata un'interpolazione tra i centroidi delle celle per smussare la stessa]

10. visualizzazione della superficie di densità, ottenuta rappresentando tematicamente i valori di 'peso' (=conteggio di eventi) per ogni cella.

La funzione di densità che si ottiene riflette la struttura dello spazio. La distribuzione di punti viene valutata allentando l'ipotesi di omogeneità e isotropia dello spazio, prendendo in considerazione la struttura della rete. La funzione di densità non è quindi il risultato di un raggio di ricerca circolare, bensì basato sulla distanza di rete. Ciò consente quindi di identificare *cluster* di fenomeni puntuali, quali ad esempio distribuzione di servizi in ambito urbano, distribuiti lungo vie di comunicazione, non sempre evidenti utilizzando analisi di densità più tradizionali di tipo euclideo.

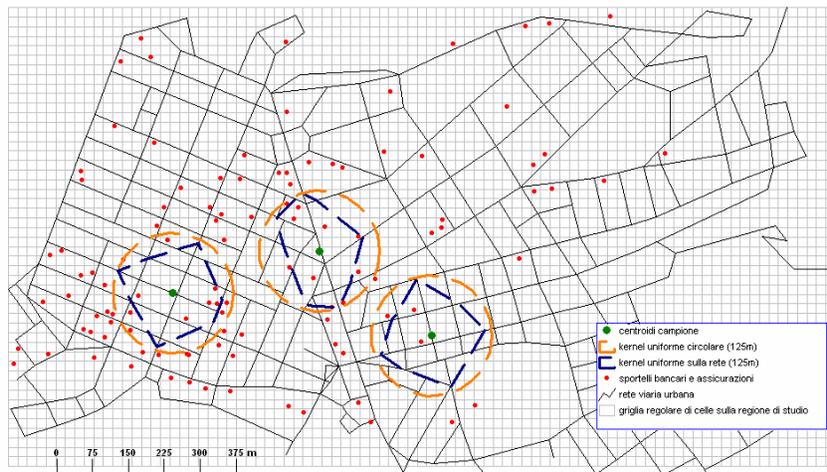


Figura 1 – la regione di studio, con esempi di 'raggi di ricerca' secondo il KDE e il NDE

Applicazione della NDE: analisi sulla distribuzione di sportelli bancari e assicurativi in ambito urbano

La metodologia NDE è stata testata su una serie di dati reali, quali la distribuzione degli sportelli bancari e assicurativi localizzate nell'area centrale del Comune di Trieste, per evidenziare concentrazioni nei luoghi di offerta e l'esistenza di un 'quartiere finanziario'. All'interno dell'area considerata è stata riportato altresì il reticolo della viabilità cittadina, elemento fondamentale per l'analisi di densità proposta. Si è partiti dunque da un dataset iniziale di 109 eventi puntuali rappresentanti sportelli bancari e assicurativi. La regione è stata suddivisa con una griglia regolare di celle quadrate, di 20m di lato, per un totale di 4263 celle (49 righe X 87 colonne) di cui si è calcolato il centroide. Come suggerito dai passaggi dell'algorithmo, si è provveduto a costruire delle aree di servizio a partire da ogni centroide¹ calcolando un 'raggio' di 125 m². Le aree di servizio sono state poi utilizzate per contare il numero di eventi ricadenti all'interno di ognuna di esse (*spatial intersection*), il cui valore è stato poi assegnato come attributo al centroide di ogni cella, e quindi alla cella relativa per la successiva visualizzazione³. Successivamente si è proceduto con

¹ L'analisi di rete è stata effettuata utilizzando il software ESRI ArcGis 8.3. Le altre elaborazioni in ambiente GIS sono state effettuate utilizzando GeoMedia Professional 4.0 e 5.1

² La distanza di 125 m è stata scelta sulla base di simulazioni effettuate durante la sperimentazione della metodologia, vista soprattutto la scala cui si faceva riferimento. In altre analisi, relative a un'estensione territoriale più ampia, il raggio di ricerca sui servizi misurava 300 m, anche sulla scorta di altri studi effettuati sulle aree urbane (Thurstain-Goodwin M., Unwin D. J., 2000).

³ Si è provveduto sia a calcolare delle densità in termini assoluti, quindi calcolando il numero complessivo di eventi all'interno di ogni area di servizio, sia ad osservare densità relative, dividendo quindi il valore di ogni cella per il numero complessivo di eventi entro ogni area di servizio e moltiplicando per la numerosità reale del campione. La

l'interpolazione dei valori al fine di produrre delle superfici continue a tre e due dimensioni. Dall'ambiente GIS quindi i risultati, espressi in termini di coordinate X e Y (esprese nel sistema di riferimento Gauss Boaga Roma 40 fuso Est) e valore di densità sono stati esportati per essere elaborati in un ambiente 3D⁴. Le figure 2 e 3 illustrano i risultati, rispettivamente nelle visualizzazioni a 3D e 2D. Si possono notare delle zone di picco, soprattutto nella parte centro-occidentale dell'area e nel settore sud-occidentale, corrispondente al centro cittadino, dove notiamo una maggiore concentrazione di eventi rappresentati da due picchi nella superficie di densità.

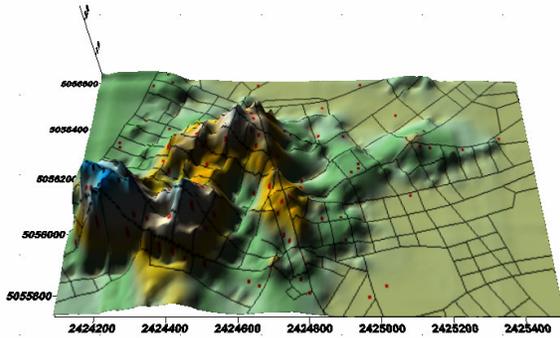


Figura 2 – NDE su servizi, raggio di 125m (3D)

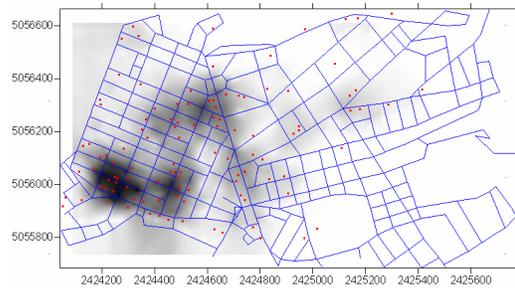


Figura 3 – NDE su servizi, raggio di 125m (2D)

La metodologia NDE è stata confrontata con il più tradizionale KDE da cui origina. La *Kernel Density Estimation* è stata dunque calcolata sul medesimo *dataset* di partenza utilizzando il programma statistico spaziale CrimeStat III. Si è mantenuto lo stesso livello di risoluzione delle celle di 20m di lato e il raggio (soglia) di ricerca di 125 metri. Per mantenere un'omogeneità di confronto con il NDE, si è scelto il tipo di interpolazione 'uniforme', che, pur presentando dei risultati più 'spigolosi' rispetto ad altre funzioni, quali la normale o la 'quartica', si adatta a confrontare sistemi di ricerca circolari con quelli a rete. Le figure 4 e 5 illustrano i risultati del KDE uniforme, rispettivamente nelle visualizzazioni a 3D e 2D.

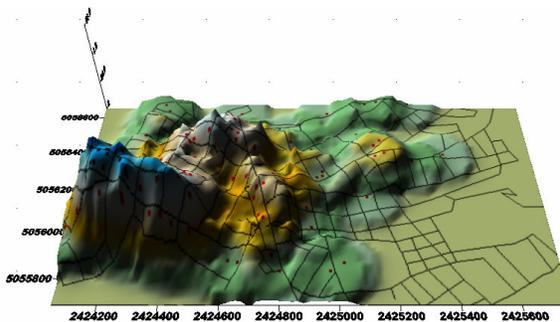


Figura 4 – KDE uniforme su servizi, raggio di 125m (3D)

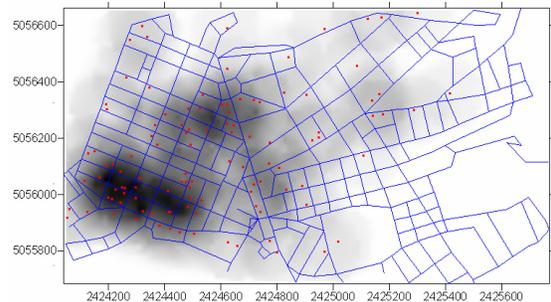


Figura 5 – KDE uniforme su servizi, raggio di 125m (2D)

Anche qui, confrontando con le precedenti immagini, notiamo la presenza di picchi nella distribuzione in aree pressoché analoghe a quelle osservate in precedenza. I picchi nell'area sud-occidentale dell'area di studio tuttavia nel presente caso risultano uniti a formare un unico *cluster* allungato, meno consistente con la forma della struttura viaria, così come altre aree risultano più

densità assoluta infatti causa il computo multiplo degli stessi casi, anche se, ai fini della visualizzazione finale, la forma della superficie di densità non cambia nelle due diverse soluzioni (Levine, 2004).

⁴ I risultati sono stati elaborati all'interno del programma Golden Software Surfer 8.0. Le figure 2, 3, 4 e 5 sono state realizzate a partire da questo programma.

dense in termini assoluti. La metodologia NDE presenta degli aspetti interessanti per quanto riguarda la considerazione della struttura reticolare della città quale 'guida' per la distribuzione di certe tipologie di attività. Il NDE non si pone comunque in alternativa al KDE, che già si dimostra utile per analisi di questo tipo, con numerosi vantaggi in termini di facilità di utilizzo e ridotte limitazioni, bensì quale metodologia per affinare la ricerca sulla distribuzione di fenomeni in ambienti urbani. Ulteriori evoluzioni della metodologia potranno riguardare la direzionalità della rete, nonché l'inserimento di altre funzioni di interpolazione quale la normale e la quartica all'interno dell'NDE.

Riferimenti bibliografici

Alonso W. (1960), "A Theory of the Urban Land Market", in *Papers and Proceedings of the Regional Science Association*, 6, 149-157

Bailey T.C. - Gatrell A.C. (1995), *Interactive Spatial Data Analysis*. Harlow: Longman

Borruso G. (2005) "Network Density Estimation: Analysis of Point Patterns over a Network", in Gervasi O. et al. (Eds.): ICCSA 2005, LNCS 3482, pp. 126 – 132, 2005, Springer-Verlag Berlin Heidelberg 2005

Borruso G. e Schoier G. (2004), "Density Analysis on Large Geographical Databases. Search for an Index of Centrality of Services at Urban Scale" in Lagana' A. et al. (eds.): *Computational Science and Its Applications - ICCSA 2004, Lecture Notes in Computer Science (LNCS) 3044*, Springer-Verlag Berlin Heidelberg, pp. 1009-1015.

Chainey S., Reid S and Stuart N., "When is a hotspot a hotspot? A procedure for creating statistically robust hotspot maps of crime", in Kidner D., Higgs G. and White S. (eds.), *Socio-Economic Applications of Geographic Information Science, Innovations in GIS 9*, Taylor and Francis, (2002), pp. 21-36

Cuthbert A. L., Anderson W. P. (2002), "Using Spatial Statistics to Examine the Pattern of Urban Land Development in Halifax-Dartmouth", in *The Professional Geographer*: 54 (4), 521-532.

Gatrell A. (1994), "Density Estimation and the Visualisation of Point Patterns". In: Hearnshaw H. M., Unwin D. (eds), *Visualisation in Geographical Information Systems*. Chichester, Wiley. 65-75.

Harris C. D., Ullman E. L. (1945), *The Nature of Cities*, 242, 7-17.

Hoch I., Waddel P. (1993), "Apartment Rents: Another Challenge to the Monocentric Model". In *Geographical Analysis* 25(1) 20-34.

Hoyt H. (1939), *The Structure and Growth of Residential Neighborhoods in American Cities*, Washington, D.C.: U.S. Government Printing Office.

Levine N. (2004), *CrimeStat III: A Spatial Statistics Program for the Analysis of Crime Incident Locations (version 2.0)*. Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, DC.

Mennis J. (2003), "Generating Surface Models of Population Using Dasymetric Mapping" *The Professional Geographer*, 55: pp. 31-42.

Okabe, A., Yamada, I (2001).: "The K-function method on a network and its computational implementation". *Geographical Analysis*, Vol.33, No.3, pp.271-290.

Thurstain-Goodwin M., Unwin D. J. (2000), "Defining and Delimiting the Central Areas of Towns for Statistical Modelling Using Continuous Surface Representations". In *Transactions in GIS*, 4: 305-317.

Waddel P., Berry B. J. L., Hoch I. (1993), "The Intersection of Space and Built Form". In *Geographical Analysis* 25(1) 5-19.