

CRIME SCENE DO NOT

La formula del *serial killer*

di PAOLO CARESSA

Più vero della finzione. La serie televisiva Numb3rs ha come protagonista un agente dell'FBI, Don Eppes (interpretato da Rob Morrow), che risolve brillantemente i suoi casi con l'aiuto del fratello, il matematico Charlie Eppes (interpretato da David Krumholtz)

Charlie Eppes che è professore presso un fantomatico istituto CalSci (ispirato forse al CalTech, il California Institute of Technology), escogita formule e ragionamenti appropriati per venire a capo delle difficili indagini oggetto di ciascun episodio. Anche se molti aspetti degli episodi della serie possono sembrare inverosimili, e il personaggio del matematico stereotipato, in realtà un fondo di verità c'è, e anzi spesso il ruolo della matematica nelle indagini reali è ben più importante di quel che non sembri dal telefilm.

In particolare, nel primo episodio in assoluto, trasmesso negli USA nel 2005 e in Italia nel 2006, Charlie Eppes scrive alla lavagna una formula, apparentemente un guazzabuglio di simboli senza troppo senso, in base alla quale "stanare" un *serial killer* a partire dai luoghi dove ha colpito le sue vittime. L'idea gli viene osservando un inaffiattoio automatico in funzione e notando come, sebbene sia impossibile predire dove saranno localizzati gli schizzi d'acqua, si possa invece risalire da questi ultimi all'ubicazione dell'innaffiattoio stesso. In modo analogo, a partire dai luoghi del delitto si può risalire, nella finzione del telefilm, all'abitazione del *serial killer*.

Nella realtà, quella formula scritta sulla lavagna è stata inventata da un uomo che in qualche modo incarna entrambi i protagonisti del telefilm: infatti si tratta di un criminologo molto versato in matematica, il canadese Kim Rossmo, che ha ideato una formula utilizzata con successo presso il dipartimento di polizia di Vancouver del quale è stato investigatore prima di intraprendere la carriera accademica. Rossmo ha elaborato la formula, che fra breve andremo a spiegare, collaudandola su casi già risolti, per vedere se il risultato fornito verificava l'esito delle indagini: questi test furono positivi, il che indusse l'investigatore canadese, che nel frattempo aveva conseguito il dottorato in criminologia, ad utilizzare la sua formula "sul campo", spesso scontrandosi con lo scetticismo dei suoi su-



I protagonisti della serie televisiva "Numb3rs"

periori, assai sospettosi (per evidente deformazione professionale!) nei confronti di sistemi poco tradizionali e apparentemente fantascientifici.

Il suo metodo, che si chiama CGT (*Criminal Geographical Targeting*) rientra nella metodologia nota come *definizione del profilo geografico*, locuzione analoga alla definizione del profilo criminale, e che potrebbe essere brevemente chiamato *profilazione geografica* – se il termine *profilazione* (talvolta utilizzato) fosse noto alla lingua italiana. Si basa su alcune idee molto semplici.

Si parte dalle coordinate (su una mappa geografica) dei punti dove il *serial killer* ha colpito, per ricostruire in base alla formula la probabile zona di residenza dell'assassino: a



I luoghi di un delitto

quel punto, utilizzando le normali tecniche investigative di definizione del profilo, si può indagare sulle persone che vivono in quella zona e corrispondono al profilo del killer (per esempio: maschio fra i 20 e i 30 anni, celibe, orfano ecc.) e una volta compilata una rosa di sospettati, si può indagare su di essi fino a “incastrare” l’assassino.

Naturalmente, il procedimento va programmato su un calcolatore, che è in grado sia di memorizzare i dati che di eseguire un software che realizzi la formula e che consenta anche di rappresentare graficamente i risultati dell’elaborazione. Infatti attualmente Rossmo è proprietario di un software chiamato *Rigel* che fa proprio questo e che egli vende a dipartimenti di polizia e ad altri enti interessati al problema di individuare l’origine di una serie di fenomeni localizzati nello spazio (per i più curiosi, il sito web è <http://www.ecricanada.com>).

Il grande “successo” ottenuto da Rossmo, che gli ha fornito una sorta di trampolino di lancio mediatico come “cacciatore matematico di criminali”, è stata la spettacolare individuazione nel 1998 di uno stupratore seriale a Lafayette, in Louisiana, un maniaco sfuggito per dieci anni agli investigatori, e stanato grazie alla sua formula.

In effetti, il suo programma riuscì a circoscrivere la zona di provenienza dello stupratore, nella quale una dozzina di sospetti vennero messi sotto controllo, prelevando campioni del loro DNA (a loro insaputa ovviamente, per esempio da capelli o altre tracce organiche come quelle che ci lasciamo dietro ovunque andiamo). Ma l’esame del DNA scagionò i dodici sospetti: tuttavia lo stupratore seriale viveva esattamente nella zona indicata dalla formula di Rossmo, solo che era un vice sceriffo di un dipartimento di polizia!!!

Naturalmente, non sempre questi metodi hanno ottenuto successi simili: le assunzioni sulle quali si basano hanno un fondamento probabilistico, non esatto e dunque possono talvolta produrre risultati inattendibili. Ma vediamo meglio di cosa si tratta.

COSÌ VICINO, COSÌ LONTANO: LA FORMULA DI ROSSMO

L’idea della formula è semplice: si considera la mappa della zona dove il *serial killer* colpisce, fissando un sistema di coordinate cartesiane e quindi associando a ciascun punto una coppia di coordinate (x, y) . Questi sono i dati di *input* della formula, e quindi del programma che la realizza su calcolatore, mentre l’*output*, il risultato, è quello che in matematica si chiama un *campo scalare*, cioè un numero associato a ogni punto della regione considerata: questo numero è tanto più alto quanto più la formula ritiene probabile che l’assassino viva nel punto che ha quelle coordinate.

Supponiamo quindi di partire da N osservazioni, cioè N punti nei quali l’assassino ha colpito: avremo dunque N coppie di coordinate $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$. Quello che vogliamo è una funzione che associ a ciascun punto (x, y) un numero tanto più alto quanto maggiore è la probabilità che lì viva l’assassino: si noti che prendendo il massimo fra questi valori e dividendo ciascun numero per questo massimo, otteniamo dei valori fra 0 e 1, che somigliano a delle probabilità, ed è per questo che spesso questi numeri si indicano con p_{xy} .

Rossmo ha considerato, nello scrivere la sua formula, alcune caratteristiche del comportamento dei serial killer, che ovviamente sono anch’esse di natura probabilistica e non necessariamente corrispondenti al profilo di qualsiasi criminale. Caratteristiche che possiamo riassumere in due osservazioni:

1. il serial killer tende ad agire in una zona sostanzialmente centrata intorno al proprio luogo di residenza, un po’ come un predatore che abbia un suo territorio di caccia; in particolare la probabilità che colpisca in un certo punto (x, y) decresce con l’allontanarsi dalla sua “tana”;
2. il serial killer non colpirà nelle immediate vicinanze della sua “tana”, così la probabilità che uno dei luoghi del delitto sia anche il luogo dove vive è nulla, e cresce con l’allontanarsi da esso. La zona intorno a un luogo del delitto dove è improbabile che il *killer* abbia la sua dimora si chiama *buffer*.

Queste due osservazioni sembrano in qualche modo in contraddizione fra loro: l’idea della formula è di dare il giusto peso a ciascuna di esse usando dei parametri da determinarsi empiricamente, per esempio applicando la formula a casi noti, dei quali si conoscono sia l’*input* che l’*output*, in modo da “tarare” questi parametri su dei valori accettabili (un’operazione che in altri contesti si chiama “calibrazione” dei parametri). È interessante notare come il concetto di distanza usato da Rossmo non sia la solita distanza euclidea, che segue dal teorema di Pitagora, e che definisce la distanza fra due punti $P = (x_P, y_P)$ e $Q = (x_Q, y_Q)$ del piano cartesiano con la formula

$$d(P, Q) = \sqrt{(x_P - x_Q)^2 + (y_P - y_Q)^2}.$$

Al suo posto viene utilizzata la distanza che in analisi matematica si chiama L^1 , detta anche “distanza del taxi” in quanto può servire a calcolare le distanze tenendo conto degli isolati nelle strade cittadine, come pure le distanze su una scacchiera, ecc:

$$d_1(P, Q) = |x_P - x_Q| + |y_P - y_Q|.$$

Questa è una distanza nel senso che soddisfa i tre assiomi per le distanze:

- $d(P, P) = 0$ cioè ogni punto ha distanza nulla da se stesso;
- $d(P, Q) = d(Q, P)$ cioè si tratta di una funzione simmetrica;
- $d(P, Q) \leq d(P, R) + d(R, Q)$ cioè vale la disuguaglianza triangolare.

In effetti si può costruire una geometria usando questa nozione di distanza: per dirne una, i cerchi, cioè le figure determinate dai punti che hanno la stessa distanza da un punto fissato, in questo caso sono dei quadrati. In altri termini siamo di fronte a una distanza con le palle quadrate (!).

Il motivo per cui si usa questa distanza è che le città, di solito il territorio di caccia dei *serial killer*, sono fatte di strade che si incrociano fra loro, e, quindi, per passare da un punto a un altro non si usa la distanza euclidea, a meno di

essere un uccello e poter volare in linea retta, ma si percorrono linee spezzate che seguono il profilo delle strade.

Tornando alla formula del *serial killer*, chiamiamo B il raggio della zona *buffer* in cui il *serial killer* non colpirà in quanto troppo vicina alla sua abitazione (per semplicità identifichiamo la zona con un cerchio ovvero, usando la distanza d_1 un quadrato), e definiamo, per ciascun indice $i = 1, 2, \dots, n$, la funzione $\varphi_i(x, y)$ come 1 se (x, y) dista da (x_i, y_i) per meno del raggio B del *buffer*, e 0 altrimenti, vale a dire:

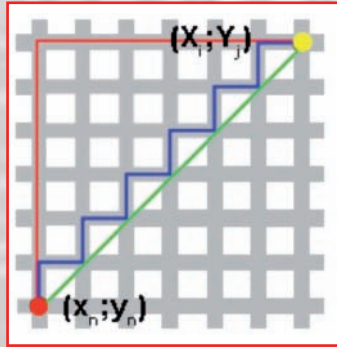
$$\varphi_i(x, y) = \begin{cases} 1 & \text{se } |x - x_i| + |y - y_i| > B, \\ 0 & \text{altrimenti.} \end{cases}$$

Con queste notazioni, e avvertendo che k, f, g sono parametri da determinarsi sperimentalmente, possiamo presentare la formula di Rossmo come segue:

$$F(x, y) = k \sum_{i=1}^n \left(\frac{\varphi_i(x, y)}{(|x - x_i| + |y - y_i|)^f} + \frac{1 - \varphi_i(x, y)}{B^{f-g} (2B - |x - x_i| - |y - y_i|)^g} \right)$$

Il valore $F(x, y)$, del campo scalare è, come abbiamo detto, un valore tanto più alto quanto più il punto di coordinate (x, y) è ritenuto la "tana" del *serial killer*.

Ovviamente, la formula richiede che (x, y) non assuma i valori di nessuna delle coppie $(x_1, y_1), \dots, (x_N, y_N)$ di coordinate dei punti nei quali si sono rilevati i delitti, altrimenti il primo denominatore sarebbe nullo rendendo priva di senso la formula.



In verde un percorso euclideo, in blu e rosso il percorso nella metrica dei taxi: le due linee blu e rossa hanno secondo questa metrica la stessa lunghezza.

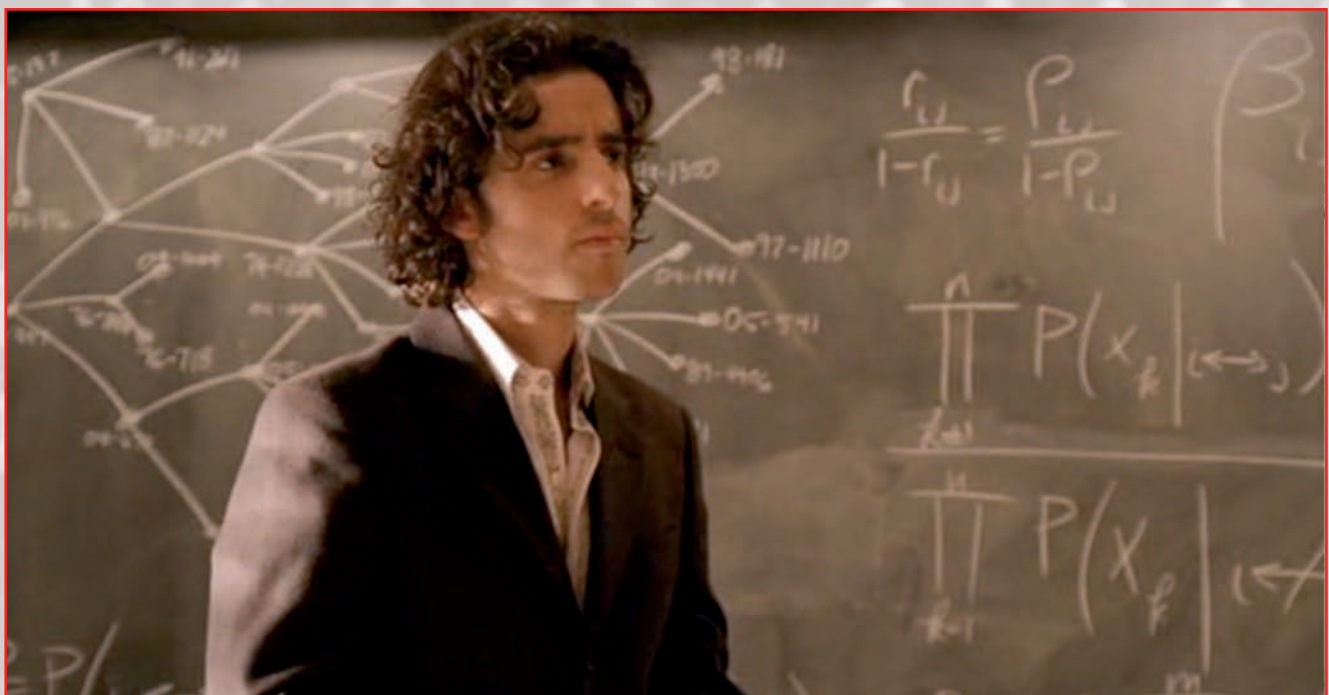
La formula di Rossmo è più semplice di quanto non sembri a prima vista: intanto notiamo che, lasciando da parte la costante k che serve semplicemente a normalizzare i valori in modo che il massimo sia 1, abbiamo una somma di tanti termini quante sono le osservazioni dei crimini.

Per ciascuna di esse consideriamo poi la somma di due termini: il primo, quello a sinistra, è inversamente proporzionale alla distanza fra il punto sotto esame e un luogo del delitto, tenendo così conto della prima delle nostre due osservazioni, cioè che la "densità dei delitti" decresce col crescere della distanza dall'abitazione del *serial killer*. Notiamo, inoltre, che il numeratore $\varphi_i(x, y)$ rende nullo il contributo di questo termine se la distanza è molto piccola, tenendo dunque conto della seconda osservazione, cioè che il killer non agisce all'interno della sua zona *buffer*.

Il secondo addendo completa la ricezione della seconda osservazione, andando a far pesare maggiormente i punti vicini alla zona *buffer* ma esterni ad essa: infatti a denominatore abbiamo una sottrazione che produce un risultato tanto più piccolo quanto più la distanza dal *buffer* cresce. A numeratore abbiamo un fattore che è nullo se $\varphi_i(x, y)$ vale 1 ed è 1 se $\varphi_i(x, y)$ vale 0, e un ulteriore fattore dato da una potenza di B , che serve ad accentuare il valore del primo o del secondo addendo a seconda che si abbia $g < f$ o viceversa.

Ovviamente, dato che il primo addendo è moltiplicato per $\varphi_i(x, y)$ mentre il secondo per $1 - \varphi_i(x, y)$, e dato che questi valori sono 0 o 1, per ciascun termine della sommatoria solo uno dei due addendi è non nullo, secondo che (x, y) si trovi o meno all'interno del *buffer* del punto i -esimo.

Come si è detto, per poter applicare questa formula è necessario determinare sperimentalmente il valore dei coefficienti $f > 0$ e $g > 0$ che, come si vede, giocano un ruolo notevole comparando come esponenti di diversi termini: in particolare, tanto più alto è il valore di f tanto più si accentua il peso della prima osservazione, mentre tanto più alto è il valore di g tanto più si accentua il peso della seconda osservazione.



Il matematico Charlie Eppes (interpretato da David Krumholtz) protagonista della serie televisiva "Numb3rs"

LA CACCIA AI SERIAL KILLER VISTA "SOTTO ANGOLI DIVERSI"

Malgrado la formula di Rossmo sia a tutti gli effetti una formula esatta, non probabilistica, il fatto che dipenda da alcuni parametri da determinarsi sperimentalmente la rende non meno approssimata di un algoritmo iterativo o "adattivo", che normalmente non può formularsi in maniera così chiara e semplice.

Detto in altri termini, la formula non è pronta per essere utilizzata così com'è, ma va in qualche modo tarata, configurata in base a parametri calcolabili applicandola a casi noti, e questa taratura introduce un livello di approssimazione e arbitrarietà: inoltre questa operazione di calibrazione può avvenire solo utilizzando un campione di dati e un programma che realizzi la formula (che la "implementi" come dicono gli informatici, che amano gli anglicismi spesso ignorando l'inglese).

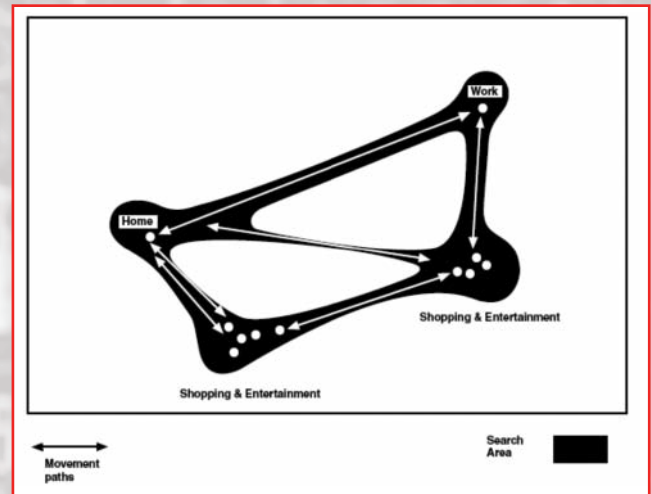
Anche la verifica dell'efficacia della formula viene fatta per via sperimentale, per esempio avvalendosi della tecnica "Monte Carlo", utilizzata a ogni pie' sospinto in fisica e finanza, che consiste nel simulare dei crimini seriali e studiare il comportamento statistico della formula su questi campioni di dati "artificiali".

I risultati di queste analisi, che posseggono ovviamente dei margini di errore, sono confortanti, tanto che Rossmo ha recentemente esteso il campo di applicabilità delle sue teorie alle scienze naturali, presentando nel 2009, assieme a due zoologi marini, uno studio sulla distribuzione degli attacchi di squali, e in generale di predatori. Il parallelo fra *serial killer* e predatori è affascinante, ma anche fuorviante: per esempio, gli squali non hanno una tana, che sarebbe l'obiettivo della formula.

In ogni caso, Rossmo non è stato né il primo né l'ultimo a pensare a modelli matematici relativi alla localizzazione geografica dei delitti seriali e alla possibile correlazione con la residenza del criminale. In effetti il suo lavoro prende le mosse da quello dei suoi "maestri" Patricia e Paul Brantingham, iniziatori, negli anni '70 e '80 del secolo scorso, degli studi di "criminologia ambientale", ai quali Rossmo riconosce il suo debito nell'introduzione del suo volume *Geographic Profiling* (2000). Ma va notato che la teoria dei Brantingham è orientata a predire i luoghi dei delitti futuri piuttosto che a ricercare il luogo di residenza del *serial killer*, utilizzando un modello basato sulla rilevazione degli "spostamenti" del *killer*, modello che in qualche modo fornisce l'idea del primo principio sul quale Rossmo ha basato la sua formula.

C'è da dire che i Brantingham hanno sottoposto a revisione queste idee in tempi più recenti, riconoscendo anche la tendenza di alcuni *serial killer* a spostarsi per motivi non necessariamente legati alla loro attività criminale, e quindi a commettere dei crimini anche in aree apparentemente discoste dalla loro zona abituale di "caccia". Ciò corrisponde a

non circoscrivere l'area dei crimini a un "cerchio" ma a più cerchi collegati fra loro, e corrispondenti alle zone dove il criminale spende parte del suo tempo nell'arco della giornata (la casa, il luogo di lavoro, i luoghi di svago ecc).



L'area battuta per la "caccia" al killer

Tuttavia Rossmo non ha solo dei precursori, ma anche degli antagonisti: in particolare, il concetto di *buffer*, e in generale i principi alla base della sua formula, sono stati criticati da un altro ricercatore che si occupa di definizione del profilo geografico, Maurice Godwin, che, manco a dirlo, ha blindato i risultati delle sue teorie in un programma per computer, *Predator*, che vende, assieme alle consulenze necessarie per configurarlo, a dipartimenti di polizia di mezzo mondo: sul sito di Godwin (<http://www.drmauricegodwin.com>) si possono trovare informazioni sulle sue ricerche e sul suo software.

L'idea di Godwin è di considerare le reciproche relazioni spaziali fra i luoghi dei delitti del *serial killer*. Per farlo, viene prima costruita una mappa per questi luoghi in termini delle loro coordinate geografiche, e poi vengono calcolati gli angoli fra le rette che congiungono queste località. Il motivo sta nel non voler considerare le zone di azione di un *serial killer* come semplicemente dei cerchi centrati nel luogo dove vive e di raggio opportuno, ma regioni la cui geometria è dettata dalle relazioni spaziali fra i luoghi dove il *serial killer* ha effettivamente colpito. Questi calcoli di angoli vengono poi elaborati dal programma che, esattamente come quello di Rossmo, produce un campo di probabilità per la dimora del *serial killer*.

Ovviamente anche Godwin ha dei casi nei quali può vantare d'aver in questo modo individuato l'abitazione di uno o più assassini, ma non sembra, dalla letteratura, che un confronto serio fra le diverse tecniche di definizione del profilo geografico sia mai stato tentato sugli stessi insiemi di dati, cioè sugli stessi casi investigativi.

Insomma, i matematici cacciatori di *serial killer* sono solitari, per il momento, esattamente come le loro prede.

Paolo Caressa

È nato a Roma, dove vive e lavora: dopo aver conseguito la laurea e il dottorato di ricerca in matematica, ha svolto attività di insegnamento e ricerca universitaria nel settore della geometria differenziale, per poi passare a svolgere consulenze per aziende di *software*. In seguito ha lavorato come analista quantitativo per un importante istituto di credito, sviluppando e implementando modelli matematici per la finanza. Attualmente si occupa di gestione di progetti *software* nell'ambito della sicurezza. Oltre ad alcuni lavori scientifici ha pubblicato articoli divulgativi (di matematica, informatica, letteratura), una *Piccola storia della matematica* (2010) per Alphatest e *Matemática escolar desde un punto de vista superior* (2011) per le edizioni UAM di Madrid. www.caressa.it

