Self-Exciting Point Processes for Crime

Craig Gilmour
Prof. D. Higham

University of Strathclyde

craig.gilmour@strath.ac.uk

Monday 17th March

Why Self-exciting Point Processes?

- Studies have shown certain types of crime, e.g. burglaries and gang violence are more likely to occur soon after a previous crime has been committed.
- ‘Repeat victimisation’ and ‘near repeat victimisation’ - offences target past victims or occur in nearby locations.
- Self-exciting point processes used in earthquake modelling (e.g. Vere-Jones 1970) - aftershocks occur in the vicinity of the primary earthquake.
- These types of models have been more recently used in the context of crimes (e.g. Mohler 2011).

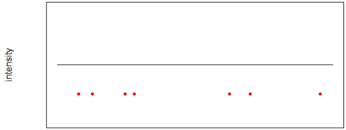
Self-exciting Point Processes

Self-exciting point processes can be defined by the *conditional intensity function*

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{\mathbb{E}(N(t, t + \Delta t) | \mathcal{H}_t)}{\Delta t}.$$

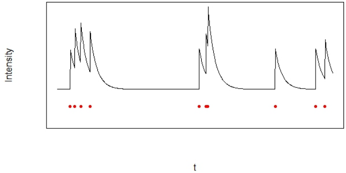
Poisson process

$$\lambda(t) = \mu.$$



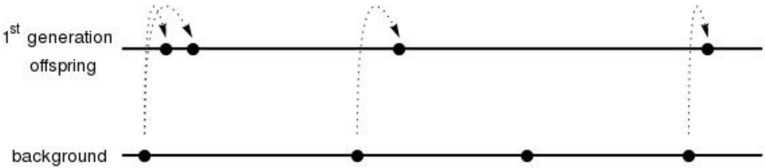
Hawkes process (Hawkes 1971)

$$\lambda(t) = \mu + \sum_{t_i < t} g(t - t_i).$$

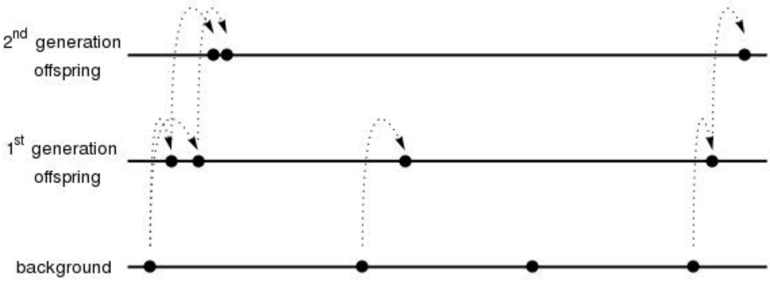


Self-Exciting Point Process

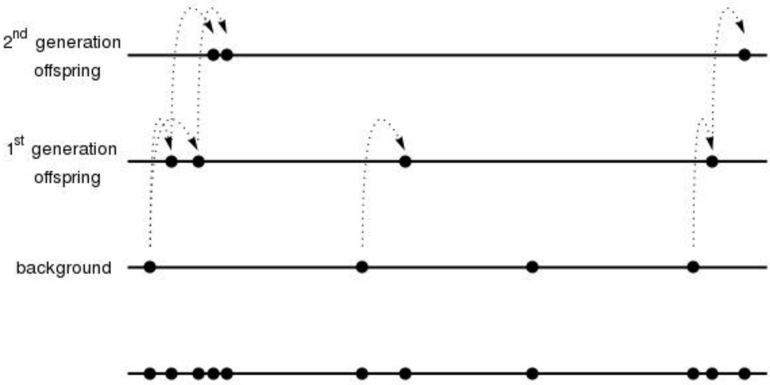
Self-Exciting Point Process



Self-Exciting Point Process



Self-Exciting Point Process



Parametric Formulation

Often a parametric form is assumed for the triggering function, e.g.

$$g(t - t_i) = \alpha\omega e^{-\omega(t-t_i)}$$

$$\lambda(t) = \mu + \sum_{t_i<t} \alpha\omega e^{-\omega(t-t_i)}$$

Given a point process $(t_1, t_2, \ldots, t_n)$ on an interval $[0, T)$, the likelihood function is given by

$$L = \left(\prod_{i=1}^{n} \lambda(t_i)\right) \exp\left(-\int_0^T \lambda(t)\, dt\right).$$

For the above ETAS model the log-likelihood can be given as

$$\log L = \sum_{i=1}^{n} \log\left(\mu + \alpha\omega \sum_{t_j<t_i} e^{-\omega(t_i-t_j)}\right) - \mu T + \sum_{i=1}^{n} \alpha(e^{-\omega(T-t_i)} - 1).$$

Model Calibration

Often the log-likelihood function is very flat, meaning standard iterative methods may not be suitable.

The model can be viewed instead as an incomplete data problem where we don't know whether an event is a background event or has been triggered by a previous event.

For the above ETAS model the log-likelihood can be given as

$$\log L = \sum_{i=1}^{n} p_{ii} \log(\mu(t_i)) - \int_0^T \mu(t) \mathrm{d}t + \sum_{i=1}^{n} \sum_{t_j < t_i} p_{ij} \log(g(t_i - t_j)) - \sum_{i=1}^{n} \int_{t_j}^{T} g(t - t_i) \mathrm{d}t$$

.

EM-algorithm

For a spatio-temporal point process

$$\lambda(x,y,t) = \mu(x,y) + \sum_{t>t_i} g(t-t_i)f(x-x_i, y-y_i)$$

$$g(t)f(x,y) = \alpha\omega e^{-\omega t} \cdot \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$$

We can create the following EM-algorithm

$$\mu = \frac{\sum_{i=1}^n p_{ii}}{T}, \quad \alpha = \frac{\sum_{i=1}^n \sum_{j=1}^{i-1} p_{ij}}{n},$$

$$\omega = \frac{\sum_{i=1}^n \sum_{j=1}^{i-1} p_{ij}}{\sum_{i=1}^n \sum_{j=1}^{i-1} p_{ij}(t_i - t_j)},$$

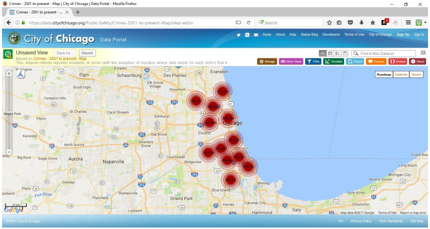
$$\sigma^2 = \frac{\sum_{i=1}^n \sum_{j=1}^{i-1} p_{ij}((x_i - x_j)^2 + (y_i - y_j)^2)}{\sum_{i=1}^n \sum_{j=1}^{i-1} 2p_{ij}(t_i - t_j)}.$$

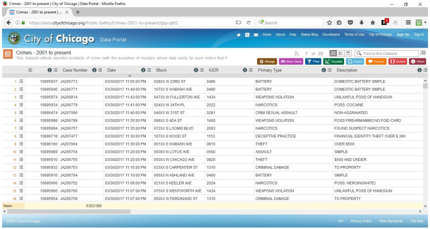
EM-Algorithm

We iterate between this maximisation step, and the following expectation step until convergence

$$p_{ii} = \frac{\mu}{\mu + \sum_{j=1}^{i-1} g(t_i - t_j)},$$
$$p_{ij} = \frac{g(t_i - t_j)}{\mu + \sum_{j=1}^{i-1} g(t_i - t_j)}.$$

Crime Data





Crime data from Chicago is publicly available at https://data.cityofchicago.org/

Data includes locations and details of every crime recorded in Chicago since 2001.

Voronoi Residual Analysis

We can use Voronoi residual analysis to assess the fit of a spatio-temporal point process (Bray 2014).

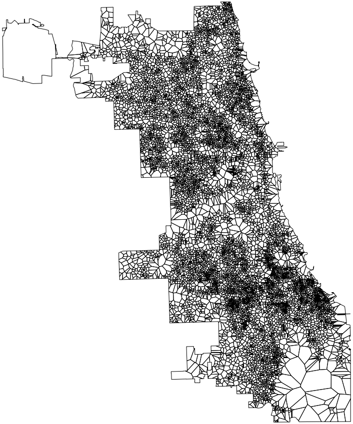
For an event (x_i, y_i), the associated Voronoi cell C_i can be defined as

$$C_i = \{(x, y) \mid d((x, y), (x_i, y_i)) \leq d((x, y), (x_j, y_j)) \quad \forall i \neq j\},$$

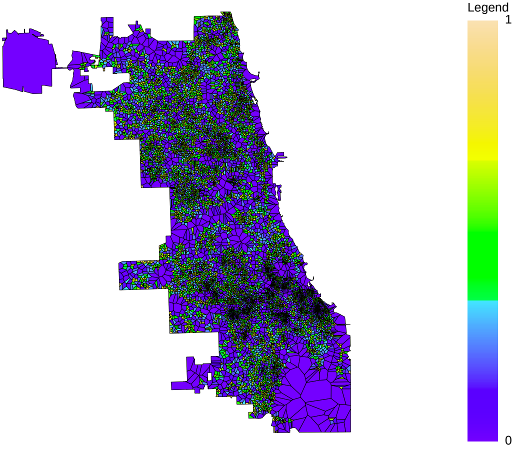
We can evaluate Voronoi residuals R^i_{vor} for each Voronoi cell i, where

$$R^i_{\text{vor}} = 1 - \int_0^T \int \int_{(x,y) \in C_i} \lambda(x, y, t) \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}t.$$

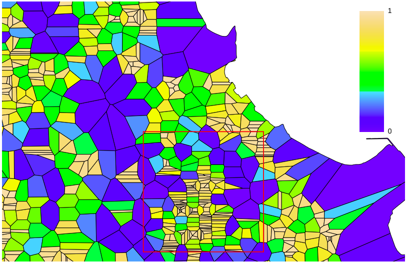
Voronoi Residual Analysis

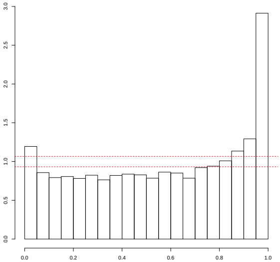


Voronoi Residual Analysis



Voronoi Residual Analysis





Modelling crime with a homogeneous background rate is not suitable.

Non-parametric Methods

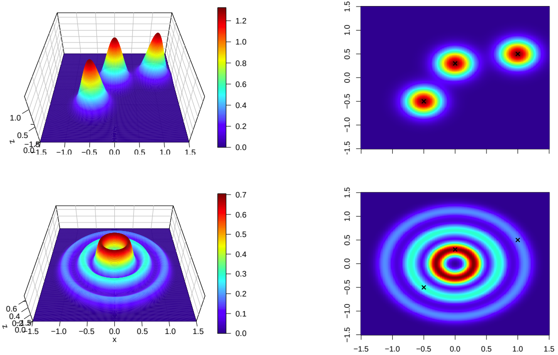
Marsan and Lengliné (2008) introduced Model Independent Stochastic Declustering (MISD) in order to find a non-parametric form for the triggering function without any prior assumptions.

This has been used to model crime events (e.g. Mohler et al. 2011).

We use different adaptations of this to model crime in Chicago.

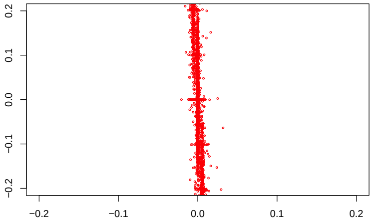
Non-parametric Methods

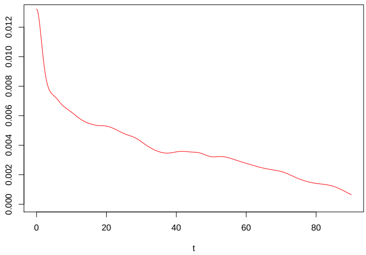
We model the crime assuming isotropy in the triggering component.



Results using model from Mohler et al. 2011.

$$\lambda(x, y, t|\mathcal{H}_t) = \nu(t)\mu(x, y) + \sum_{t>t_i} g(x - x_i, y - y_i, t - t_i).$$

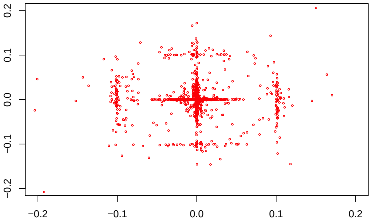


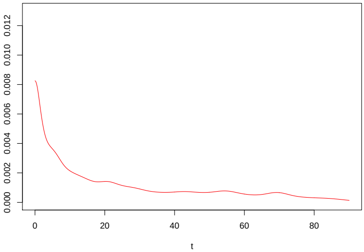


Results on Chicago Data

Results when assuming isotropy.

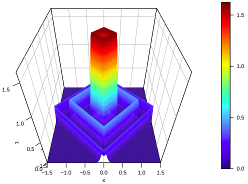
$$\lambda(x, y, t|\mathcal{H}_t) = \nu(t)\mu(x, y) + \sum_{t>t_i} g(r - r_i, t - t_i).$$

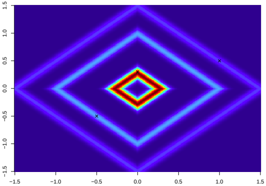




Non-parametric Methods

We also model the triggering function using Manhattan distance to measure the distance between events.

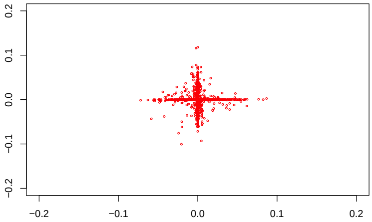


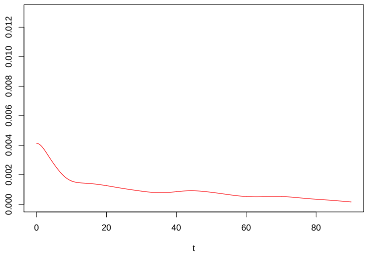


Results on Chicago Data

Results when using Manhattan distance.

$$\lambda(x, y, t|\mathcal{H}_t) = \nu(t)\mu(x, y) + \sum_{t>t_i} g(d - d_i, t - t_i).$$





Summary

- Tested self-exciting point process models on Chicago crime data
- Tested variants of the standard approaches
- Could maintain predictive power with reduced dimensionality

The End