R Notebook

Simulation, Estimation and Applications of Hawkes Processes

Hawkes processes are a sub-class of stochastic processes that were introduced in 1971 by Hawkes to model the occurrence of seismic events. The processes themselves are characterised by a stochastic intensity vector, which represents the conditional probability density of the occurrence of an event in the immediate future. Hawkes Processes are a type of point process that are uniquely characterised by being self-exciting. That is, the arrival of a given event increases the probability that the given event will occur again for a given period of time. Their flexibility and real-world relevancy has resulted in a host of applications. For example, in the case of financial data, this allows propagation of stock crashes and surges to be modelled. In social media, they are used for modelling 'twitter cascades'. Further examples include the modelling of earthquake events, neuron firing times, and terrorist activity.

A vital aspect of model fitting and parameter estimation is one of identifiability. Identifiability can roughly be taken to mean that there exists is a unique maximum likelihood estimator for any arbitrary dataset. When models become unidentifiable, there are multiple parameter values that are equally likely to have given rise to the data. This can cause significant problems as the fitted model can diverge significantly from the true data generating process, rendering inference and prediction useless. Understanding when Hawkes processes become unidentifiable has yet to be fully explored, particularly in the multivariate setting and when event observations are aggregated into bins to form a time series of counts, as is common with many datasets.

This project will investigate under what circumstances Hawkes processes become unidentifiable. In doing so you will make a vital contribution to an ongoing project in applying multivariate Hawkes processes to computer network data for cyber-security purposes.

Useful Papers in General

Note that in general there is a lot of good application papers for siesmic data sets and financial data sets but very few good theoretical avenues have been explored.

Dirichlet-Hawkes Processes with Applications to Clustering Continuous-Time Document Streams

code for Dirichlet-Hawkes Processes

Neural Hawkes Process

Reconfigurable Predictive Systems for Event Streams

Online Learning for Multivariate Hawkes Processes

Self-Exciting Spatio-Temporal Point Processes

Maximum Likelihood Estimation of Poisson and Hawkes Processes

Likelihood based inference for the multivariate renewal Hawkes process

Likelihood function for multivariate Hawkes

Indian Buffet

Bayesian Inference for Hawkes

Fake news and Hawkes Processes

Identifiability Issues

twitter cascades

twitter information diffusion

What is not proven yet

Non-parametric estimation

Thoughts

A new criterion could be invested in that allows more flexibility in model comparison than AIC or BIC that doesn't solely penalise on the number of parameters but also the simplicity and importance of said parameter. (How like backgeound knowledge is something?)

Useful Code

In **R** there is the **hawkes** library and the **THAP** toolbox in MATLAB. Lots of papers discussing code can be found on arxiv.