Uncovering Causality from Multivariate Hawkes Integrated Cumulants

ICML 2017

Massil Achab[†], Emmanuel Bacry[†], Stéphane Gaïffas[†], Iacopo Mastromatteo[*] and Jean-François Muzy[†,‡]

9th August 2017

[†] Ecole Polytechnique, [*] Capital Fund Management, [‡] Université de Corse

Outline

Introduction

Introduction

- In many applications, one needs to deal with data containing a very large number of irregular timestamped events.

- A **temporal point process** is a collection of points $Z = \{\tau_1, \tau_2, \dots\}$ randomly located on $\mathbb{R}^+$.

- Another representation is based on the counting process $N_t = \sum_{\tau \in Z} \mathbb{1}_{\tau \leq t}$.

Introduction

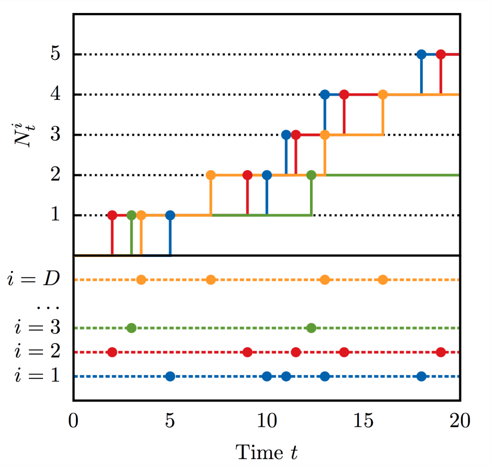

Figure 1: Example of a multivariate point process

How many events of type j are caused by events of type i ?

Hawkes Process

Hawkes Process

- The vector of stochastic intensities $\boldsymbol{\lambda}_t = [\lambda_t^1 \cdots \lambda_t^D]^\top$ associated with the multivariate counting process $\boldsymbol{N}_t$ is defined as

$$\lambda_t^i = \lim_{dt \to 0} \frac{\mathbb{P}(N_{t+dt}^i - N_t^i = 1 | \mathcal{F}_t)}{dt}.$$

- $\boldsymbol{N}_t$ is called a **Hawkes point process** if the stochastic intensities can be written as

$$\lambda_t^i = \mu^i + \sum_{j=1}^{D} \sum_{\tau_j \in Z^j} \phi^{ij}(t - \tau_j),$$

where $\mu^i \in \mathbb{R}^+$ is an exogenous intensity and ϕ^{ij} are positive, integrable and supported in $\mathbb{R}^+$ functions called *kernels* encoding the impact of an action by j on the activity of i.

Hawkes Causality

- Instead of trying to estimate the kernels ϕ^{ij}, we focus on the direct estimation of their *integrals*:

$$g^{ij} = \int_0^{+\infty} \phi^{ij}(u)du \qquad\qquad \boldsymbol{G} = [g^{ij}].$$

- Using Hawkes process representation as a Poisson cluster process, we introduce the counting function $N_t^{i \leftarrow j}$ that counts the number of events of i whose direct ancestor is an event of j. We know that:

$$\mathbb{E}[dN_t^{i \leftarrow j}] = g^{ij}\mathbb{E}[dN_t^j] = g^{ij}\Lambda^j dt.$$

- The integral g^{ij} is **the average number of events of type** i **triggered by one event of type** j.

Integrated Cumulants

Integrated Cumulants

- The integrals of these cumulants are related to $\boldsymbol{R} = (\boldsymbol{I} - \boldsymbol{G})^{-1}$. For the purpose of our method, we only need to consider cumulants up to the third order.

$$\Lambda^i = \sum_{m=1}^{D} R^{im} \mu^m$$

$$C^{ij} = \sum_{m=1}^{D} \Lambda^m R^{im} R^{jm}$$

$$S^{ijk} = \sum_{m=1}^{D} (R^{im} R^{jm} C^{km} + R^{im} C^{jm} R^{km} + C^{im} R^{jm} R^{km} - 2\Lambda^m R^{im} R^{jm} R^{km})$$

- We designed asymptotically unbiased estimators of the integrated cumulants above, see the paper for details.

Nonparametric Estimation

Estimation procedure

- Usual framework: estimate $[\phi^{ij}(t)]$ via maximum likelihood making assumptions on the kernels' shapes.

- Estimation of $\boldsymbol{G}$ without any parametric modeling and estimation of the kernel ϕ^{ij} themselves.

- To recover $\boldsymbol{G}$, we minimize the following criterion using SGD:

$$\mathcal{L}(\boldsymbol{R}) = (1-\kappa)||\boldsymbol{S_c}(\boldsymbol{R}) - \widehat{\boldsymbol{S_c}}||_2^2 + \kappa||\boldsymbol{C}(\boldsymbol{R}) - \widehat{\boldsymbol{C}}||_2^2,$$

- Our estimator is defined as: $\widehat{\boldsymbol{G}} = \boldsymbol{I} - (\arg\min_{\boldsymbol{R}} \mathcal{L}(\boldsymbol{R}))^{-1}$.

Consistency of the estimator

Suppose that $\boldsymbol{N_t}$ is observed on $\mathbb{R}^+$ and assume that

- $g_0(\boldsymbol{R}) = 0$ if and only if $\boldsymbol{R} = \boldsymbol{R}_0$;
- $\boldsymbol{R} \in \Theta$, which is a compact set;
- the spectral radius of the kernel norm matrix satisfies $||\boldsymbol{G}_0|| < 1$;
- $H_T \to \infty$ and $H_T^2/T \to 0$.

Then

$$\widehat{\boldsymbol{G}}_T = \boldsymbol{I} - \left(\arg\min_{\boldsymbol{R}\in\Theta} \mathcal{L}_T(\boldsymbol{R})\right)^{-1} \xrightarrow{\mathbb{P}} \boldsymbol{G}_0.$$

Experimental results

Order book model

We consider the following 8-dimensional point process which models the dynamic of an order book:

$$N_t = (P_t^+, P_t^-, T_t^a, T_t^b, L_t^a, L_t^b, C_t^a, C_t^b)$$

Each dimension counts the number of events before t:

- P^+ (P^-): upwards (downward) mid-price move triggered by any order.
- T^a (T^b): market order at the ask (bid) that does not move the price.
- L^a (L^b): limit order at the ask (bid) that does not move the price.
- C^a (C^b): cancel order at the ask (bid) that does not move the price.

Hawkes Causality map

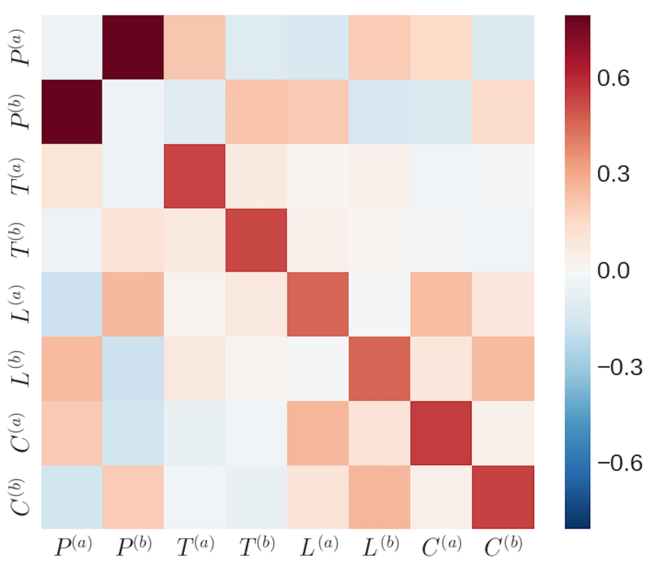

Figure 2: Estimated $\widehat{G}$ on DAX order book data.

Questions ? (Poster # 75)