Simulation et Estimation des Processus de Hawkes

E.Bacry
CNRS, CMAP, Ecole Polytechnique,
France

http ://www.cmap.polytechnique.fr/ ˜bacry

Hawkes processes - general definition in dimension D

- N_t : a D-dimensional jump process (jumps are all of size 1)
- λ_t : D-dimensional stochastic intensity
- μ : D-dimensional exogenous intensity
- $\Phi(t)$: $D \times D$ square matrix of kernel functions $\Phi^{ij}(t)$ which are positive and causal (i.e., supported by R^+).
- Moreover $||\Phi^{ij}||_{L^1} < +\infty, \quad 1 \leq i,j \leq D$

"Auto-regressive" relation

$$\lambda_\mathbf{t} = \mu + \mathbf{\Phi} \star \mathbf{dN_t},$$

where by definition

$$(\Phi \star dN_t)^{ij} = \sum_{k=1}^{D} \int_{-\infty}^{+\infty} \Phi^{ik}(t-s) dN^k(s)$$

Simulation

- Inverse Methods $\longrightarrow$ based on time change
- Cluster Methods $\longrightarrow$ hierarchical method
- Thinning methods $\longrightarrow$ based on a bound of intensity

In 1d ($D = 1$), it gives

$$F(t) = \int_0^t \lambda(u)du$$

$N(F^{-1}(t))$ **is an homogeneous Poisson process of intensity 1.**

$\Longrightarrow$ One needs to know how to simulate $F^{-1}(t_{i+1}) - F^{-1}(t_i)$ where $t_{i+1} - t_i$ is exponentially distributed

Not that easy !

Dassios, Zhao, 2013

Exponential case $\phi(t) = \alpha e^{-\beta t}$

- $F^{-1}(t_{i+1}) - F^{-1}(t_i)$ is simulated using 2 iid exponential variables $\longrightarrow$ Total is $2N$
- update of the intensity at each jump : multiplication by an exponential $\longrightarrow$ Total is $N(D)$

where N be the total number of jumps

Complexity $\simeq O(ND)$

The “basic“ method in 1d ($D = 1$)

i. Simulate the "immigrants" $\{t_k^0\}_k$ on $[0, T]$ (using Poisson μ)

ii. Set $n = 0$

iii. For each point t_k^n of generation n, we generate a Poisson process on $t \in [t_k^n, T]$ of intensity $\phi(t - t_k^n)$

iv. The so-obtained points form generation $n + 1$: $\{t_l^{n+1}\}_l$.

v $n \leftarrow n + 1$ and go back to iii.

The “basic“ method Problems

- not "causal"
- Edge problems : clusters generated by immigrants before time 0 may contain offspring in $[0, T]$.

$\Longrightarrow$ **Use the "perfect" (!) algorithm by Moller, Rasmussen 2014**

Ogata 1981

Key point : $\phi(t)$ is bounded by a decreasing function $\bar{\phi}(t)$

i. Let t_0 be a time of jump. $n \leftarrow 0$

ii. Compute $\bar{\lambda}_t = \lambda_{t_n^-|\mathcal{F}_{t^-}} + \bar{\phi}(t - t_n)$

iii. $s \leftarrow t_n$

iv. $s' \leftarrow s + d$, where d is exponential of parameter $\bar{\lambda}_s$

vi. If $u < \lambda_{s'}/\bar{\lambda}_s$ with $u \sim$ Uniform in [0,1]
Record a new jump $t_{n+1} = s'$, $n \leftarrow n + 1$ and go to ii.

vii. Thinning : $s \leftarrow s'$ and go to iv.

Simulation : Thinning algorithm

If N' is the total number of jumps (including the thinned ones)

- Any kernel $\Rightarrow$ Complexity is $O({N'}^2 D)$
- Exponential kernel $\Rightarrow$ Complexity is $O(N' D)$

Estimation

- **Parametric estimation (Maximum likelihood)**
  First work : Ogata 78

- **Non parametric estimation**
  - Marsan Lengliné (2008), generalized by Lewis, Mohler (2010)
    → Expected Maximization (EM) procedure of a (penalized) likelihood function
    → Monovariate Hawkes processes, Small amount of data, No theoretical results
  - Reynaud-Bouret and Schbath (2010)
    → Developed for small amount of data (Sparse penalization)
  - E.B. and J.F.Muzy (2014)
    → Developed for large amount of data

In the following, I shall suppose that

$$\rho(||\Phi||_{L^1}) < 1,$$

where $||\Phi||_{L^1} = \{||\Phi^{ij}||_{L^1}\}_{i,j}$.

It implies that λ_t is (asymptotically) stationary and

$$E(\lambda_t) = \Lambda = (\mathbb{I} - ||\Phi||_{L^1})^{-1}\mu$$

where

$$\Psi(t) = \sum_{k=1}^{+\infty} \Phi^{(*k)}(t).$$

Non parametric estimation

E.B. and J.F. Muzy (2014)

- **The true values of Φ and μ verify a Wiener-Hopf equation** (Hawkes 71 - E.B., J.F. Muzy, 2013)

$$g(t) = \Phi(t) + \Phi * g(t), \quad \forall t > 0$$

where $g^{ij}(t)dt = E(dN_t^i | dN_0^j = 1)$

- Given a "good" g, it has a **unique solution in Φ**

$\Longrightarrow$ **The second-order statistics characterize a Hawkes process**

Non parametric estimation

E.B. and J.F. Muzy (2014)

- **Wiener-Hopf equation**

$$g(t) = \Phi(t) + \int_0^{+\infty} \phi(s) g(t-s) ds, \quad \forall t > 0$$

- Solution is computed using **Nyström method** (Gaussian quadrature)
- Warning : $g(t)$ is discontinuous at $t = 0$
- **Use of Gaussian quadrature $\Longleftrightarrow$ regularization**
- Can be generalized in the case of a marked Hawkes process

$$\lambda_t = \mu + \Phi \star f(m_t) dN_t,$$

where the marks m_t (iid) are known and $f(.)$ is a new parameter

E.B, J.F.Muzy (2013)

4 dimension Hawkes process : $N_t = \begin{pmatrix} T_t \\ P_t \end{pmatrix}$,

- Upward/Downward price jumps : $P_t = \begin{pmatrix} P_t^+ \\ P_t^- \end{pmatrix}$
- Buying/Selling trades : $T_t = \begin{pmatrix} T_t^+ \\ T_t^- \end{pmatrix}$

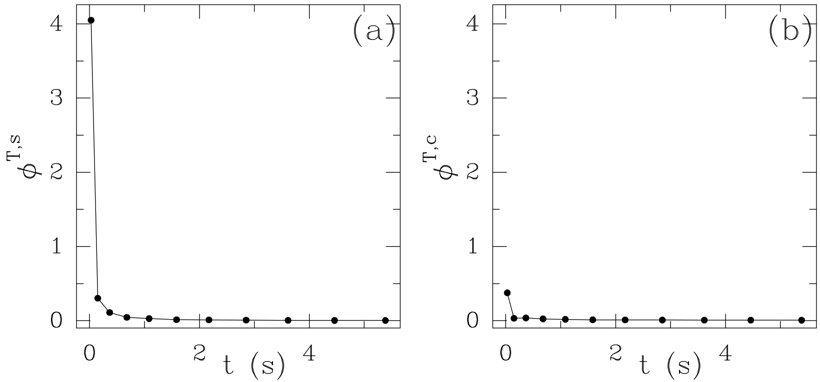
E.B, J.F.Muzy (2013)

The kernels

$$\Phi(t) = \begin{pmatrix} \Phi^T(t) & \Phi^R(t) \\ \Phi^I(t) & \Phi^N(t) \end{pmatrix}$$

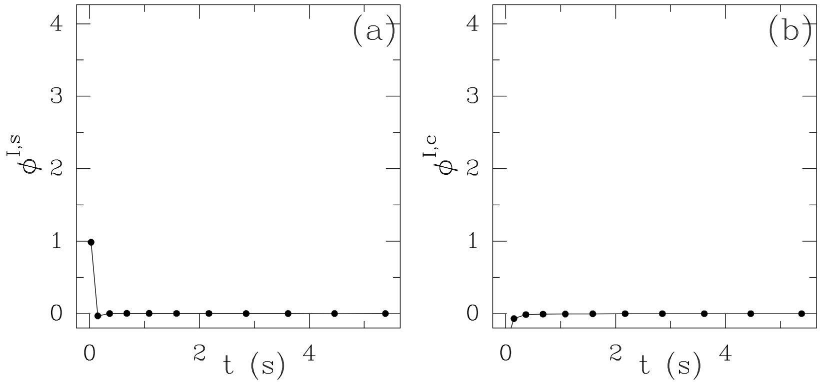
- each Φ and Φ are 2×2 symmetric matrices : $\begin{pmatrix} \phi^s & \phi^c \\ \phi^c & \phi^s \end{pmatrix}$
- $\Phi^T(t)$: Auto-correlation of trades
- $\Phi^I(t)$: Impact of trades on the price
- $\Phi^N(t)$: Influence of past price moves on future price moves
- $\Phi^R(t)$: Retro-influence of price moves on trades

Non parametric estimation of Φ^T for Eurostoxx Futures 10h-12h, 2009-2012 (800 days)



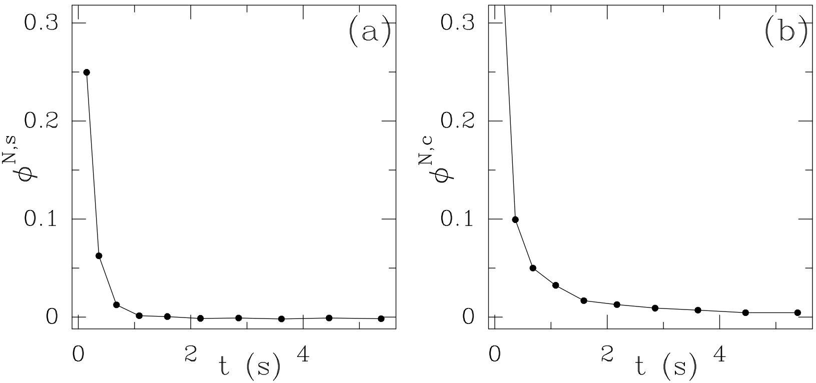
Trade auto-correlation $\longrightarrow$ "Positive" correlation : Order splitting

Non parametric estimation of Φ^I for Eurostoxx Futures 10h-12h, 2009-2012 (800 days)



Trade "instantaneous" impact

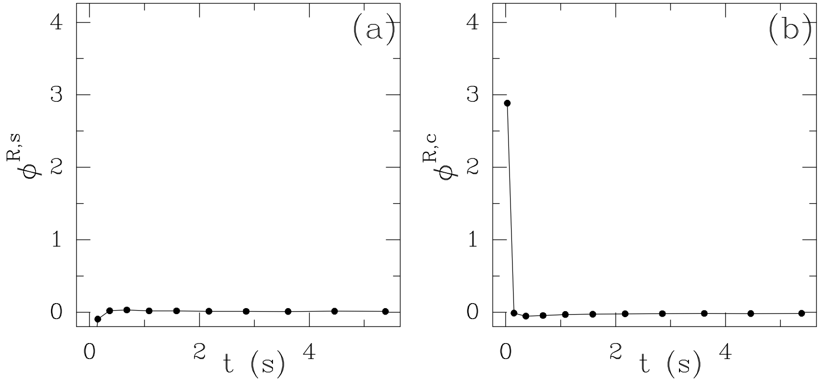
Non parametric estimation of Φ^N for Eurostoxx Futures 10h-12h, 2009-2012 (800 days)



Influence of past price moves on future price moves

$\longrightarrow$ Mostly mean reverting (micro-structure)

Non parametric estimation of Φ^R for Eurostoxx Futures 10h-12h, 2009-2012 (800 days)



Retro-influence of price moves on anonymous trades :

Price goes up $\Longrightarrow$ more sell market orders

Non parametric estimation : Application to ETAS model

ETAS 1d model (Kagan and Knopoff 81, 87 - Ogata, 88)

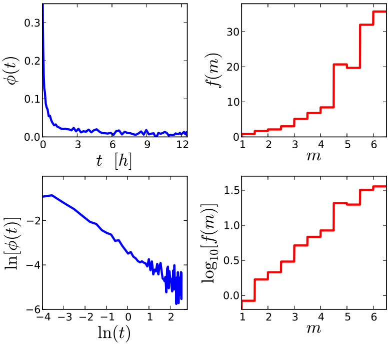
$$\lambda_t = \mu + \int_0^t \phi(t-s) f(m_t) dN_t$$

- the marks (magnitude) m_t are supposed to be iid
- mark : $f(m) = Ae^{\alpha m}$
- kernel $\phi(t) = \frac{C}{(1+t/c)^p}$
- Many parametric estimation
- From our knowledge, single previous non parametric estimation for earthquakes : Marsan and Lengliné (2008),

E.B. and J.F. Muzy (2014)

estimation from North-Carolina Earthquake Catalog

$p \simeq 0.7$ and $c \simeq 1$ mn



Parametric estimation : Maximum likelihood

Likelihood of a Hawkes process $\{N_t\}_{t\in[0,T]}$ (Ogata, 78) :

$$L(\Phi, \mu, N_t) = \sum_{i=1}^{D} \int_0^T \ln(\lambda_t^i) dN_t^i - \sum_{i=1}^{D} \int_0^T \lambda_t^i dt$$

Parametric estimation :

- General kernel case :
  $\longrightarrow$ complexity $O(N^2 D)$ (N : total number of jumps)
- Exponential kernel case : $\Phi^{ij}(t) = a^{ij} e^{-\beta^{ij} t}$
  $\longrightarrow$ complexity $O(ND)$

Parametric estimation in large dimensions

E.B, S.Gaiffas, J.F.Muzy (in prep.)
Application : Social Network, Finance, ...

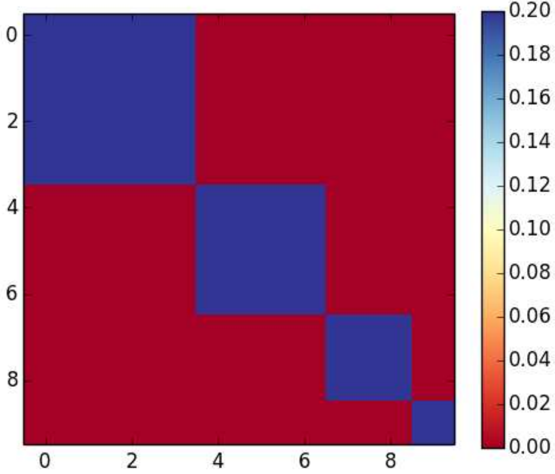
- **Exponential kernels** : $\Phi^{ij}(t) = a^{ij} e^{-\beta^{ij} t}$
  - $A = \{a^{ij}\}_{1 \leq i,j \leq D}$: "adjacency" matrix
  - $\{\beta^{ij}\}_{1 \leq i,j \leq D}$: "decay" matrix
- **Likelihood convex parametrization** : $\Phi^{ij}(t) = a^{ij} e^{-\beta^{ij}/a^{ij} t}$
  - $B = \{\beta^{ij}/a^{ij}\}_{1 \leq i,j \leq D}$
- Regularization (group of similar agents)
  - **Sparsity** of $A \longrightarrow L^1$ penalization on the a^{ij}
  - A should be **low rank** $\longrightarrow$ sparsity on the spectrum of A, Trace-norm penalization of A

Convex optimization / numerical aspects

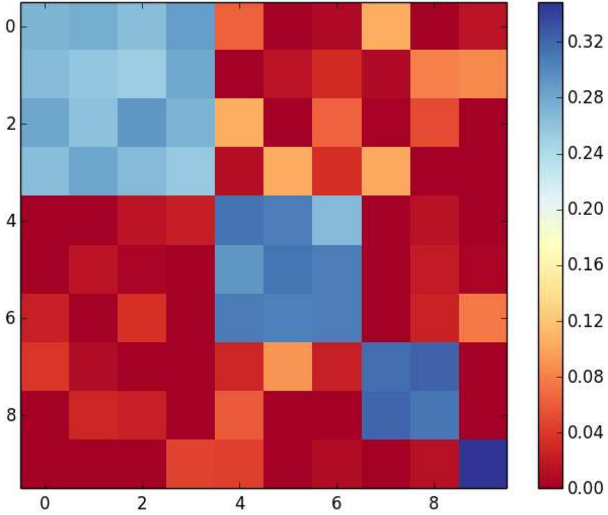
- Accelerated proximal gradient descent such as Fista [Beck Teboulle (2009)] or Prisma [Orabona et al (2012)]
- When carefully done complexity of one gradient is $O(ND)$ (instead of $O(N^2D)$ for the naive approach)
- We have a parallelized code for this : the gradient on each node $j \in \{1, \ldots, D\}$ can be computed **in parallel**
- Computation bottleneck is the heavy use of exp and log ! (can be accelerated using some ugly hacking)
- Proximal operator of ℓ_1-norm and trace-norm are the standard soft-thresholding and spectral soft-thresholding operators
- spectral soft-thresholding operators requires truncated SVD : we use the Lanczos's default implementation of `Python`, it is fast enough

Parametric estimation in large dimensions

Dimension $= 10$ (210 parameters)
Adjacency matrix :

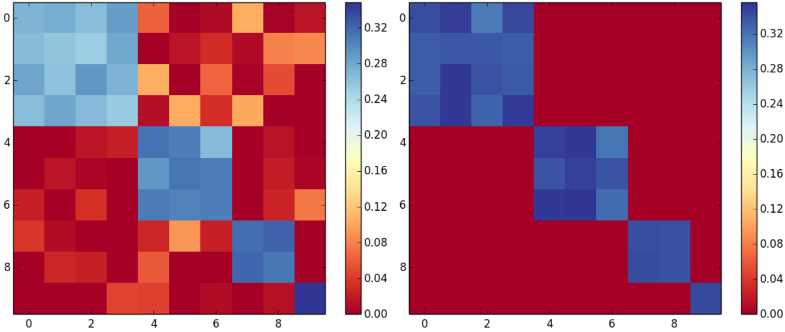


No Penalization

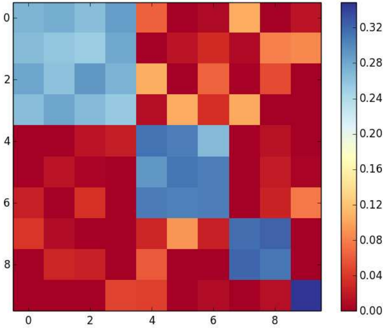


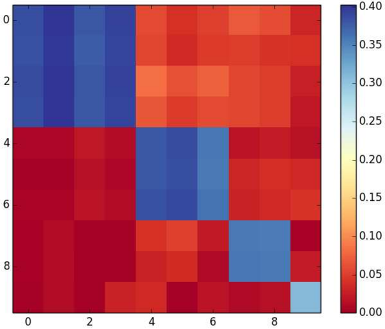
Parametric estimation in large dimensions

No penalization (left) and Lasso Penalization (right)

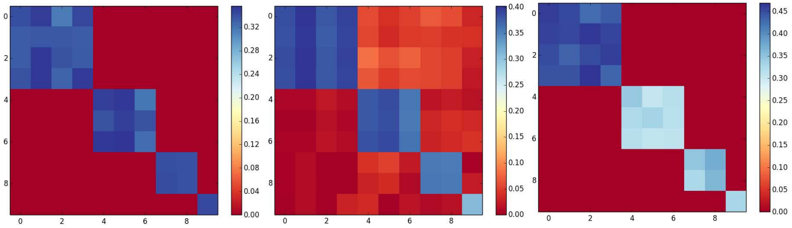


No penalization (left) and Trace norm Penalization (right)

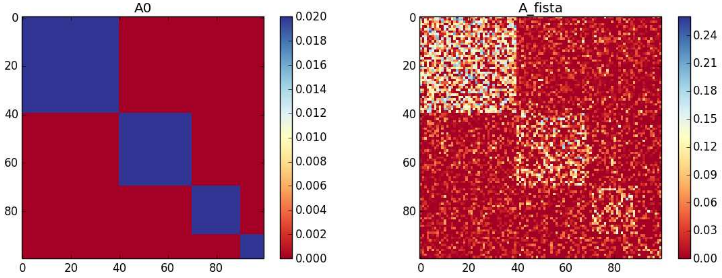




Lasso (left), Trace Norm (middle), both (right)

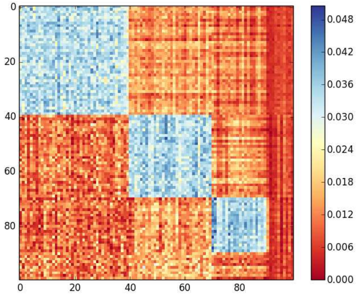


Dimension = 100 (20100 Parameters), Lasso penalization



Parametric estimation in large dimensions

Dimension = 100 (20100 Parameters), Trace norm penalization



Oracle Inequality

- L^2 framework
- Penalization in $\hat{\theta}$ depending on data-driven weights

$$\mathrm{pen}(\theta) = ||\mu||_{1,\hat{w}} + ||\mathbf{A}||_{1,\hat{\mathbf{W}}} + \hat{w}_* ||\mathbf{A}||_*$$

given by empirical variance terms

Parametric estimation in large dimensions

E.B, S.Gaiffas, J.F.Muzy (in prep.)

Theorem : Oracle inequality

We have

$$||\lambda_{\hat{\theta}} - \lambda^*||_T^2 \leq \inf_{\theta} \left\{ ||\lambda_\theta - \lambda^*||_T^2 + \text{complexity}(\theta) \right\}$$

with a probability larger than $1 - ce^{-x}$ with

$$\begin{aligned} \text{complexity}(\theta) \lesssim & \frac{||\mu||_0 (x + \log D)}{T} \max_j N^j([0, T])/T \\ & + \frac{||\mathbf{A}||_0 (x + 2\log D)}{T} \max_{j,k} \hat{\mathbf{V}}^{jk}(T) \\ & + \frac{\text{rank}(A)(x + \log D)}{T} ||\hat{\mathbf{V}}_1(T)||_{\text{op}} \vee ||\hat{\mathbf{V}}_2(T)||_{\text{op}} \end{aligned}$$

- Leading constant is 1

Main tool : **new matrix-martingale concentration inequality** for continuous-time martingale

- Consider the random matrix $\mathbf{Z}(t)$

$$\mathbf{Z}(t) = dM \star \Phi \star dN_t$$

[M = martingale obtained by compensation of N]

- This is the noise term in our problem

Parametric estimation in large dimensions

E.B, S.Gaiffas, J.F.Muzy (in prep.)

Theorem

For any $x > 0$, we have

$$\frac{||\mathbf{Z}(t)||_{\mathrm{op}}}{t}$$

$$\leq 4\sqrt{\frac{(x + \log D + \hat{\ell}_x(t))||\hat{\mathbf{V}}_1(t)||_{\mathrm{op}} \vee ||\hat{\mathbf{V}}_2(t)||_{\mathrm{op}}}{t}}$$

$$+ \frac{(x + \log D + \hat{\ell}_x(t))(10.34 + 2.65\sup_{t\in[0,T]}||\mathbf{H}(t)||_{2,\infty})}{t}$$

with a probability larger than $1 - 84.9e^{-x}$

- First non-commutative version of Bernstein's inequality for countinuous-time martingales
- Extension of [Tropp (2012)] results

Where : $||\mathbf{H}(t)||_{2,\infty}$ = maximum ℓ_2-norm of rows of $\mathbf{H}(t)$, where

$$\mathbf{H}(t) = \Phi \star dN_t$$

$\hat{\mathbf{V}}_1(t)$ diagonal matrix with entries

$$\hat{\mathbf{V}}_1^{jj}(t) = \frac{1}{t}\int_0^t ||\mathbf{H}(s)||^2_{2,\infty} dN^j(s)$$

$\hat{\mathbf{V}}_2(t)$ matrix with entries

$$\hat{\mathbf{V}}_2^{jk}(t) = \frac{1}{t}\int_0^t ||\mathbf{H}(s)||^2_{2,\infty} \sum_{l=1}^{d} \frac{\Phi^{jl}(s)\Phi^{kl}(s)}{||\mathbf{H}_{l,\bullet}(s)||_2^2} dN_l(s),$$