Nonparametric Bayesian inference for nonlinear Hawkes processes

Vincent Rivoirard

Université Paris-Dauphine

Nonparametric Bayesian inference for nonlinear Hawkes processes

Joint work with

- Deborah Sulem
  Oxford University

- Judith Rousseau
  Oxford University

Intensity of a point process

Definition (Point process)

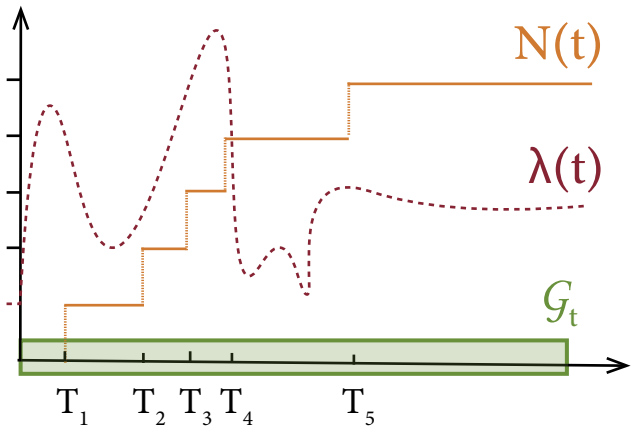
A point process $N = (N_t)_t$ is a random countable set of points of $\mathbb{R}$ or equivalently a non-decreasing integer-valued process.

Definition (Intensity of a point process)

The intensity λ_t of N represents the probability to observe a point at the time t conditionally on the past before t:

$$\lambda_t \mathrm{d}t = \mathbb{P}(N \text{ has a jump} \in [t, t + \mathrm{d}t] \text{ conditionally on the past before } t)$$

Example: Poisson processes correspond to the case where $(\lambda_t)_t$ is not random. And the Poisson process is homogeneous if, in addition, λ_t does not depend on t.



Univariate Hawkes processes

$$\lambda_t \mathrm{d}t = \mathbb{P}(N \text{ has a jump} \in [t, t+\mathrm{d}t] \text{ conditionally on the past before } t)$$

Definition (univariate Hawkes process)

Let $\Phi : \mathbb{R} \mapsto \mathbb{R}_+$ and $h : \mathbb{R}_+ \mapsto \mathbb{R}$ such that $\|h\|_1 < 1$. Then any point process N whose intensity is

$$\lambda_t = \Phi\left(\int_{-\infty}^{t-} h(t-u)\mathrm{d}N_u\right) = \Phi\Big(\sum_{T \in N,\, T<t} h(t-T)\Big)$$

is called a univariate Hawkes process.

See Hawkes (1971), Hawkes and Oakes (1974), Brémaud and Massoulié (1996, 2001).

Univariate Hawkes processes

$$\lambda_t \mathrm{d}t = \mathbb{P}(N \text{ has a jump} \in [t, t+\mathrm{d}t] \text{ conditionally on the past before } t)$$

Definition (univariate Hawkes process)

Let $\Phi : \mathbb{R} \mapsto \mathbb{R}_+$ and $h : \mathbb{R}_+ \mapsto \mathbb{R}$ such that $\|h\|_1 < 1$. Then any point process N whose intensity is

$$\lambda_t = \Phi\left(\int_{-\infty}^{t-} h(t-u)\mathrm{d}N_u\right) = \Phi\Big(\sum_{T \in N,\, T<t} h(t-T)\Big)$$

is called a univariate Hawkes process.

See Hawkes (1971), Hawkes and Oakes (1974), Brémaud and Massoulié (1996, 2001).

Definition (linear univariate Hawkes process)

If $\Phi(x) = x + \nu$ with $\nu > 0$ and $h \geq 0$, the Hawkes process is linear:

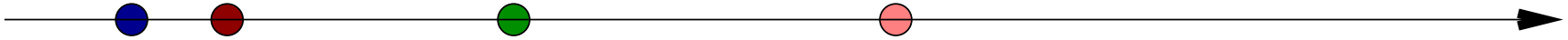
$$\lambda_t = \nu + \int_{-\infty}^{t-} h(t-u)\mathrm{d}N_u = \nu + \sum_{T \in N,\, T<t} h(t-T)$$

with ν called the spontaneous rate and h the self-exciting function.

The study of linear Hawkes processes is much easier thanks to the cluster representation

Cluster representation for linear Hawkes processes

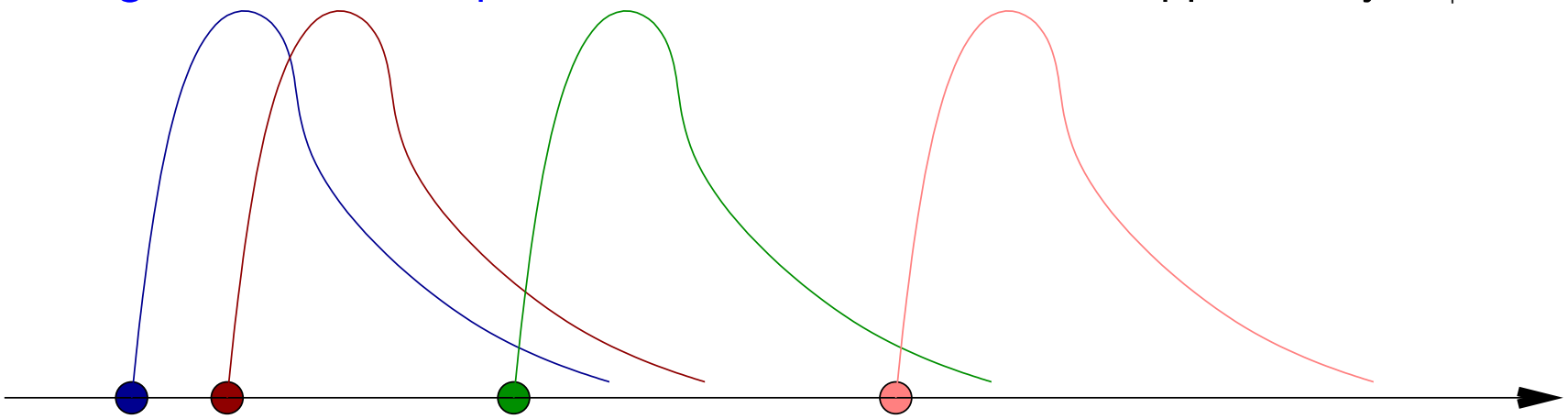
A univariate linear Hawkes process can be viewed as a branching process over an homogeneous Poisson process. Let $\nu > 0$ and $h \geq 0$ supported by $\mathbb{R}_+$.



- Ancestors: Realizations of a Poisson Process with $\lambda_t = \nu$

Cluster representation for linear Hawkes processes

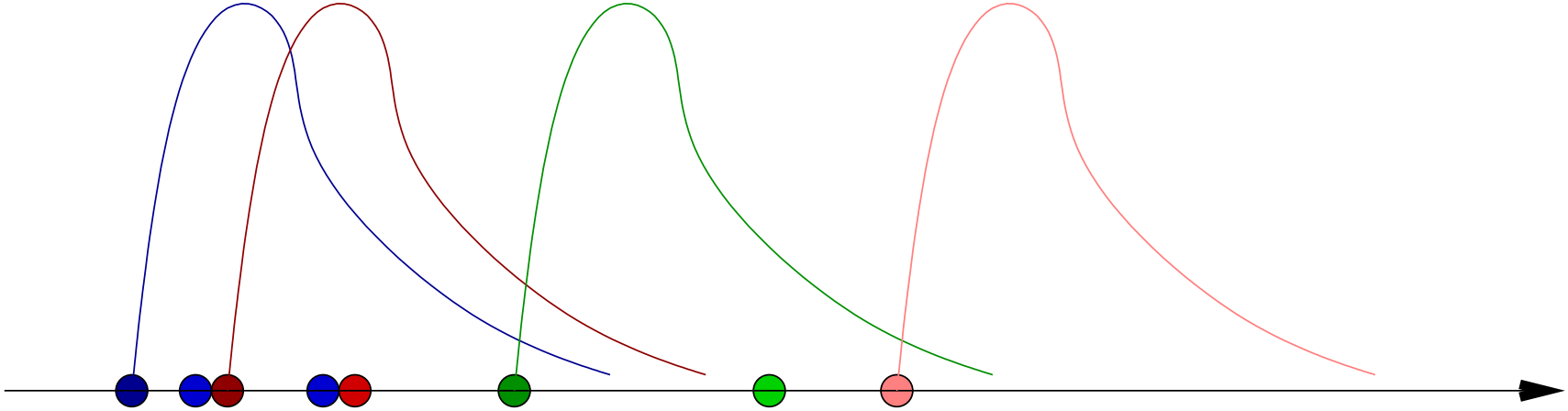
A univariate linear Hawkes process can be viewed as a branching process over an homogeneous Poisson process. Let $\nu > 0$ and $h \geq 0$ supported by $\mathbb{R}_+$.



- Ancestors: Realizations of a Poisson Process with $\lambda_t = \nu$
- Each ancestor can give birth to children according to a P.P. with $\lambda_t = h(t)$

Cluster representation for linear Hawkes processes

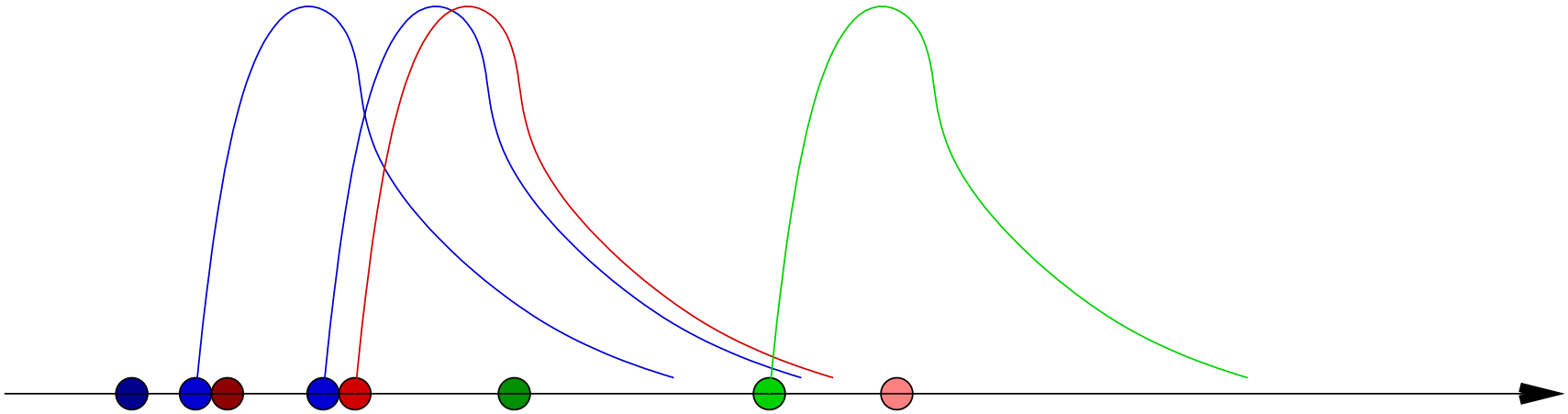
A univariate linear Hawkes process can be viewed as a branching process over an homogeneous Poisson process. Let $\nu > 0$ and $h \geq 0$ supported by $\mathbb{R}_+$.



- Ancestors: Realizations of a Poisson Process with $\lambda_t = \nu$
- Each ancestor can give birth to children according to a P.P. with $\lambda_t = h(t)$

Cluster representation for linear Hawkes processes

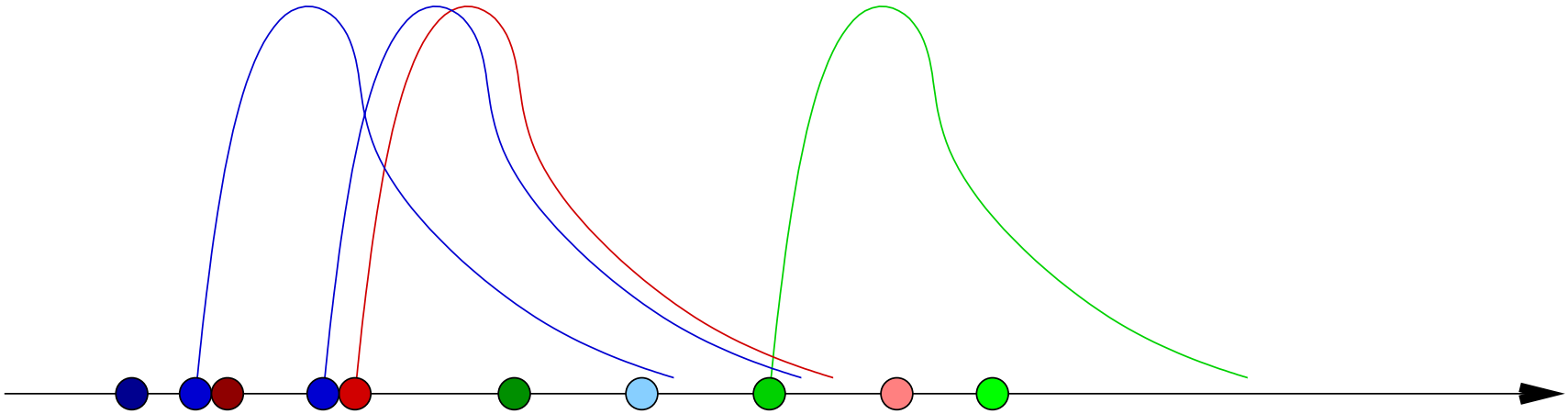
A univariate linear Hawkes process can be viewed as a branching process over an homogeneous Poisson process. Let $\nu > 0$ and $h \geq 0$ supported by $\mathbb{R}_+$.



- Ancestors: Realizations of a Poisson Process with $\lambda_t = \nu$
- Each ancestor can give birth to children according to a P.P. with $\lambda_t = h(t)$
- Each child can give birth to children according to a P.P. with $\lambda_t = h(t)$

Cluster representation for linear Hawkes processes

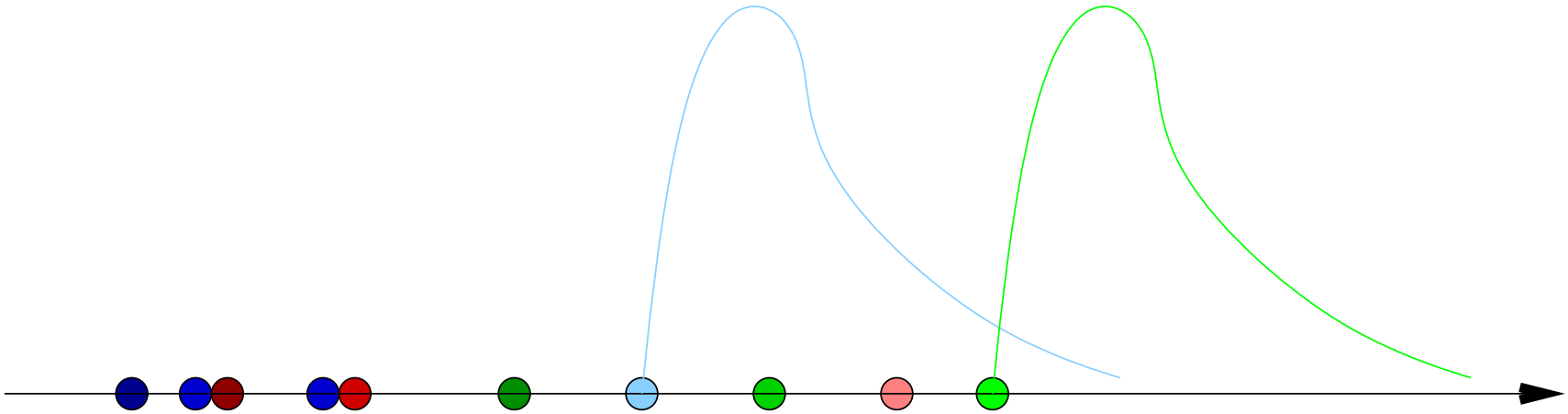
A univariate linear Hawkes process can be viewed as a branching process over an homogeneous Poisson process. Let $\nu > 0$ and $h \geq 0$ supported by $\mathbb{R}_+$.



- Ancestors: Realizations of a Poisson Process with $\lambda_t = \nu$
- Each ancestor can give birth to children according to a P.P. with $\lambda_t = h(t)$
- Each child can give birth to children according to a P.P. with $\lambda_t = h(t)$

Cluster representation for linear Hawkes processes

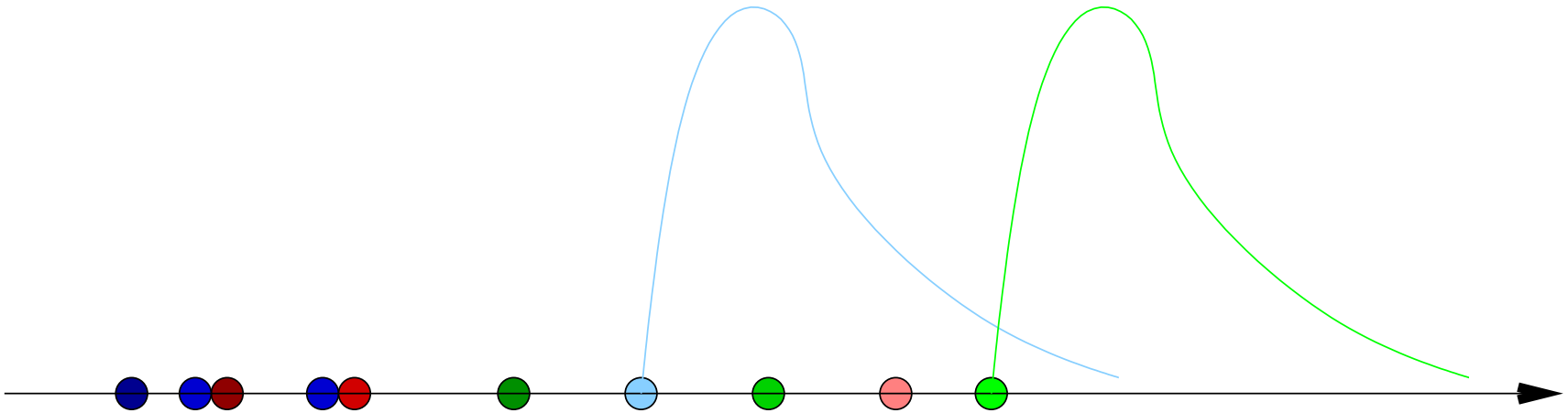
A univariate linear Hawkes process can be viewed as a branching process over an homogeneous Poisson process. Let $\nu > 0$ and $h \geq 0$ supported by $\mathbb{R}_+$.



- Ancestors: Realizations of a Poisson Process with $\lambda_t = \nu$
- Each ancestor can give birth to children according to a P.P. with $\lambda_t = h(t)$
- Each child can give birth to children according to a P.P. with $\lambda_t = h(t)$

Cluster representation for linear Hawkes processes

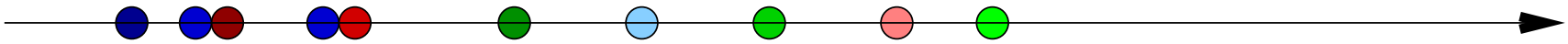
A univariate linear Hawkes process can be viewed as a branching process over an homogeneous Poisson process. Let $\nu > 0$ and $h \geq 0$ supported by $\mathbb{R}_+$.



- Ancestors: Realizations of a Poisson Process with $\lambda_t = \nu$
- Each ancestor can give birth to children according to a P.P. with $\lambda_t = h(t)$
- Each child can give birth to children according to a P.P. with $\lambda_t = h(t)$
- Extinction if $\int_0^{+\infty} h(t)\mathrm{d}t < 1$

Cluster representation for linear Hawkes processes

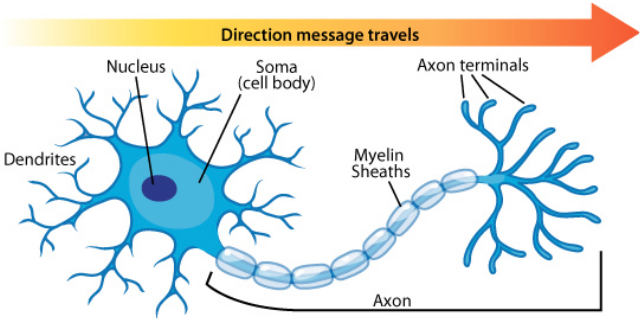
A univariate linear Hawkes process can be viewed as a branching process over an homogeneous Poisson process. Let $\nu > 0$ and $h \geq 0$ supported by $\mathbb{R}_+$.

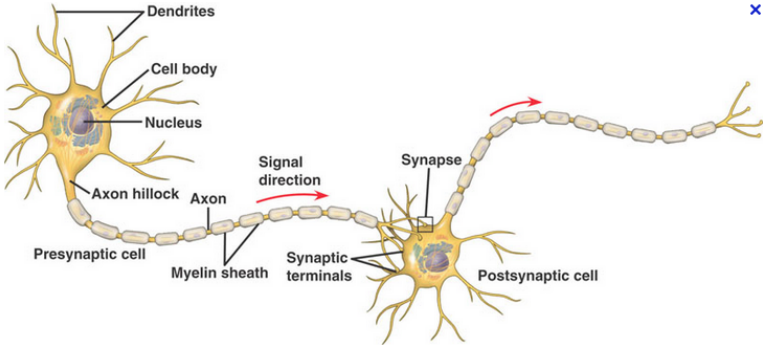


- Ancestors: Realizations of a Poisson Process with $\lambda_t = \nu$
- Each ancestor can give birth to children according to a P.P. with $\lambda_t = h(t)$
- Each child can give birth to children according to a P.P. with $\lambda_t = h(t)$
- Extinction if $\int_0^{+\infty} h(t)\mathrm{d}t < 1$
- Hawkes process = all the points where colors are not distinguished
- See Hawkes and Oakes (1974)

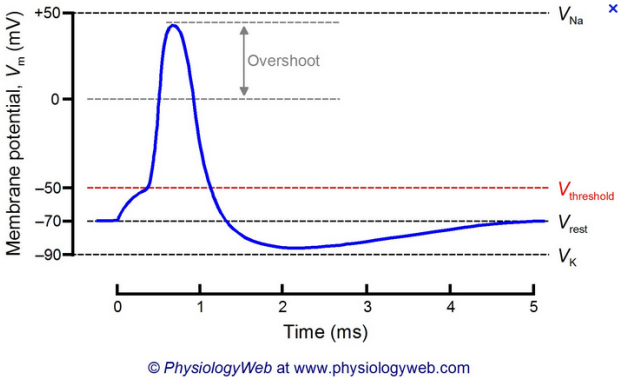
Multivariate Hawkes process: Neurobiological motivations

A neuron is an electrically excitable cell that processes and transmits information through electrical signals





If upstream signal is strong enough, this cell produces an action potential (also called spike), which is a spiky (electric) signal. Then, this signal is propagated to downstream neurons.



Action potentials can be recorded and the excitations times can be seen as a point process, each point corresponding to the peak of one action potential of this neuron.

Goal: Using the recorded activity of K neurons, we wish to infer the graph between them. For this purpose, we use models based on multivariate Hawkes processes.

Multivariate Hawkes processes

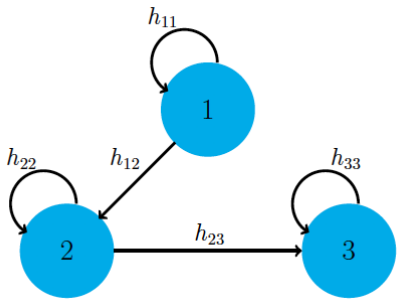
- We naturally modify the intensity of a univariate Hawkes process given by

$$\lambda_t = \psi\Big(\nu + \int_{-\infty}^{t-} h(t-u)\mathrm{d}N_u\Big) = \psi\Big(\nu + \sum_{T\in N,\, T<t} h(t-T)\Big),$$

to model interactions between K neurons: For a given neuron $k \in [\![1;K]\!]$, we model its activity by a point process $N^{(k)}$ whose intensity is

$$\lambda_t^{(k)} = \psi_k\Big(\nu_k + \sum_{\ell=1}^{K} \int_{-\infty}^{t-} h_{\ell k}(t-u)\mathrm{d}N^{(\ell)}(u)\Big)$$

$$= \psi_k\Big(\nu_k + \sum_{\ell=1}^{K} \sum_{T_\ell \in N^{(\ell)},\, T_\ell<t} h_{\ell k}(t-T_\ell)\Big)$$



- We obtain mutually exciting and inhibiting processes:

$\nu_k > 0$: background rates
$h_{\ell k}$: interaction functions
- If $h_{\ell k} \geq 0$: excitation
- If $h_{\ell k} \leq 0$: inhibition
- If $h_{\ell k}$ is signed: excitation and inhibition

ψ_k: positive and nondecreasing link function
Typical examples:
- linear: $\psi_k(x) = x$ [requires $h_{\ell k} \geq 0$]
- nonlinear: $\psi_k(x) = x_+ = \max(x, 0)$
- nonlinear: $\psi_k(x) = \exp(x)$

Multivariate Hawkes processes

Definition

A K-dimensional continuous time process $N = (N_t)_t = (N_t^{(1)}, \dots, N_t^{(K)})_t$ is a multivariate nonlinear Hawkes process if

(i) almost surely, for $k \neq \ell$, $(N_t^{(k)})_t$ and $(N_t^{(\ell)})_t$ never jump simultaneously

(ii) for all k, the intensity of $(N_t^{(k)})_t$ is given by

$$\lambda_t^{(k)} = \psi_k\Big(\nu_k + \sum_{\ell=1}^{K} \int_{-\infty}^{t-} h_{\ell k}(t-u)\mathrm{d}N^{(\ell)}(u)\Big).$$

Theorem (Brémaud and Massoulié (1996))

Existence and uniqueness of a stationary distribution for N:
- if $\forall k \in [\![1;K]\!]$ $\|\psi_k\|_\infty < \infty$ or
- if $\forall k \in [\![1;K]\!]$ ψ_k is α_k-Lipschitz and the matrix Γ with entries $\Gamma_{\ell k} = \alpha_k \|h_{\ell k}\|_1$ has a spectral radius $\rho(\Gamma) < 1$.

Applications of Hawkes processes

Hawkes processes are useful to model many situations where excitation or inhibition phenomena play a crucial role.

- to model earthquakes: Ozaki (1979), Ogata and Akaike (1982), Vere-Jones and Ozaki (1982) and Zhuang, Ogata and Vere-Jones (2002)
- to neuroscience: Chornoboy, Schramm and Karr (1988) combined Hawkes processes with maximum likelihood in the parametric setting.
- to genome analysis: Gusto and Schbath (2005), Carstensen, Sandelin, Winther and Hansen (2010) and Reynaud-Bouret and Schbath (2010)
- to financial data: Embrechts, Liniger and Lin (2011), Bacry and Muzy (2013, 2014) and Bacry, Delattre, Hoffmann and Muzy (2012)
- to study diffusion across social networks: Crane and Sornette (2008) and Yang and Zha (2013)
- to analyze and predict the diffusion of COVID-19: Mengersen, Paraha, R., Rousseau and Sulem (2020)
- etc.

State of the art in the nonparametric setting

- Nonparametric inference for multivariate Hawkes processes:

$$\lambda_t^{(k)} = \psi_k\Big(\nu_k^* + \sum_{\ell=1}^{K} \int_{-\infty}^{t-} h_{\ell k}^*(t-u)\mathrm{d}N^{(\ell)}(u)\Big).$$

  Statistical Goal: Estimation of $f^* = (\nu_k^*, (h_{\ell k}^*)_{\ell \in [\![1;K]\!]})_{k \in [\![1;K]\!]}$ based on observations of $N = (N^{(k)})_{k \in [\![1;K]\!]}$ on $[0, T]$ with intensity process $(\lambda^{(k)})_{k \in [\![1;K]\!]}$.
- Linear case: $\psi_k(x) = x$
  - Lasso-type estimation: Hansen, Reynaud-Bouret and R. (2015) extended by Chen, Witten and Shojaie (2017). See also Bacry, Bompaire, Gaïffas and Muzy (2020)
  - Bayesian estimation: Donnet, R. and Rousseau (2020)
- Nonlinear case:
  - Chen, Shojaie, Shea-Brown and Witten (2019) derived bounds on the weak dependence coefficient for the Hawkes process using the coupling technique of Dedecker and Prieur (2014), providing an asymptotic analysis of second order statistics (cross-covariance)
  - Estimation of f^* in full generality remains an open question (to the best of our knowledge)

Our contributions

Inference for nonlinear Hawkes models

- We observe $N = (N^{(k)})_{k \in [\![1;K]\!]}$ on $[0, T]$ with intensity process $(\lambda^{(k)})_{k \in [\![1;K]\!]}$. given by

$$\lambda_t^{(k)} = \psi\Big(\nu_k^* + \sum_{\ell=1}^{K} \int_{-\infty}^{t-} h_{\ell k}^*(t-u) \mathrm{d}N^{(\ell)}(u)\Big)$$

  where $\psi : \mathbb{R} \longmapsto \mathbb{R}_+$ is known and non-decreasing

- Assumptions:
  - the ν_k^*'s are positive
  - the $h_{\ell k}^*$'s are bounded
  - the support of the $h_{\ell k}^*$'s is included into $[0, A]$, with $A < \infty$ known

  We do not assume that the $h_{\ell k}^*$'s are non-negative, so inhibition is possible.

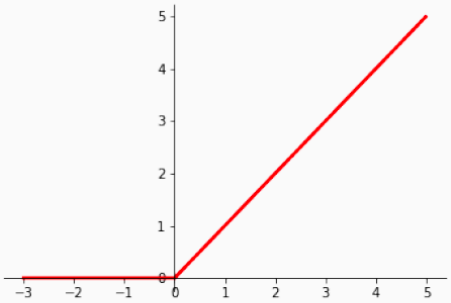
- Statistical goals: (Bayesian) estimation of

$$f^* = (\nu_k^*, (h_{\ell k}^*)_{\ell \in [\![1;K]\!]})_{k \in [\![1;K]\!]}$$

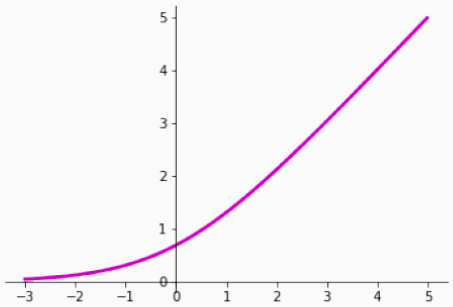
  by using observations of $N = (N^{(k)})_{k \in [\![1;K]\!]}$ on $[0, T]$ with in mind $T \to +\infty$

Typical link functions

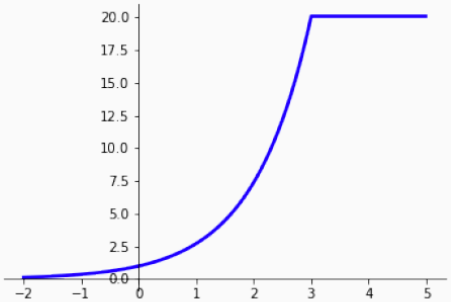
ReLU $\psi(x) = x_+$



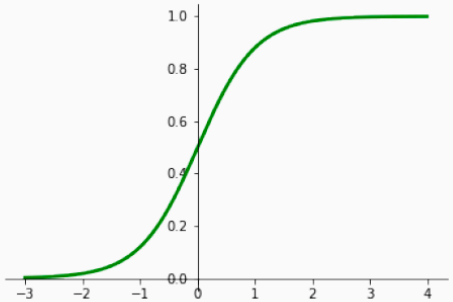
Logit $\psi(x) = \log(1 + e^x)$



Exponential $\psi(x) = \min(e^x, \Lambda)$



Sigmoid: $\psi(x) = (1 + e^x)^{-1}$



Stationarity

Intensity process of $N = (N^{(k)})_{k \in [\![1;K]\!]}$:

$$\lambda_t^{(k)} = \psi\Big(\nu_k^* + \sum_{\ell=1}^{K} \int_{-\infty}^{t-} h_{\ell k}^*(t-u)\mathrm{d}N^{(\ell)}(u)\Big)$$

with $\psi : \mathbb{R} \longmapsto \mathbb{R}_+$ known and non-decreasing

Lemma

If one of the following conditions is satisfied:

(C1) *ψ is bounded:* $\exists \Lambda > 0, \forall x \in \mathbb{R}, \psi(x) \leq \Lambda$

(C2) *ψ is L-Lipschitz, with $L > 0$ and $r(S)$, the spectral radius of the matrix S with entries $S_{lk} = L\|h_{\ell k}^*\|_1$ satisfies $r(S) < 1$*

(C3) *ψ is L-Lipschitz with $L > 0$ and $L \times \max_\ell \sum_k \|h_{\ell k}^{*+}\|_1 < 1$*

then there exists a unique stationary version of the process N with finite average

Notation:

$$h_{\ell k}^{*+}(x) = \max(h_{\ell k}^*(x), 0), \quad h_{\ell k}^{*-}(x) = \max(-h_{\ell k}^*(x), 0)$$

$$|h_{\ell k}^*(x)| = h_{\ell k}^{*+}(x) + h_{\ell k}^{*-}(x), \quad h_{\ell k}^*(x) = h_{\ell k}^{*+}(x) - h_{\ell k}^{*-}(x)$$

Identifiability

Intensity process of $N = (N^{(k)})_{k \in [\![1;K]\!]}$:

$$\lambda_t^{(k)} = \psi\Big(\nu_k^* + \sum_{\ell=1}^{K} \int_{-\infty}^{t-} h_{\ell k}^*(t-u)\mathrm{d}N^{(\ell)}(u)\Big)$$

with $\psi : \mathbb{R} \longmapsto \mathbb{R}_+$ known, non-decreasing and L-Lipschitz.

Lemma

If ψ is bijective on an open interval I so that for any k

$$[\nu_k^* - \max_\ell \|h_{\ell k}^{*-}\|_\infty ; \nu_k^* + \max_\ell \|h_{\ell k}^{*+}\|_\infty] \subset I,$$

then the distribution of N is identifiable for T large enough.

Remark: Identifiability is satisfied
- for the logit $\psi(x) = \log(1+e^x)$ and sigmoid $\psi(x) = (1+e^{-x})^{-1}$ link functions
- for the ReLU function, $\psi(x) = \max(x,0) = x^+$, we assume for any k,

$$\max_\ell \|h_{\ell k}^{*-}\|_\infty < \nu_k^*$$

- for the exponentiel function, $\psi(x) = \min(e^x, \Lambda)$, we assume for any k,

$$\max_\ell \|h_{\ell k}^{*+}\|_\infty + \nu_k^* < \log \Lambda$$

The Bayesian statistical approach

- Unlike the frequentist approach, the (pure) Bayesian approach does not assume the existence of a true parameter.
- The parameter to infer is assumed to be random: Let Π a prior on $f \in \mathcal{F}$, with

$$f = (\nu_k, (h_{\ell k})_{\ell \in [\![1;K]\!]})_{k \in [\![1;K]\!]} \quad \text{or} \quad f = (\nu_k, (h_{\ell k})_{\ell \in [\![1;K]\!]}, \theta_k)_{k \in [\![1;K]\!]}$$

  and we study the posterior distribution $\Pi(\cdot|N)$, which can be written

$$\Pi(B|N) = \frac{\int_B \exp(\mathcal{L}_T(f))\mathrm{d}\Pi(f)}{\int_{\mathcal{F}} \exp(\mathcal{L}_T(f))\mathrm{d}\Pi(f)}, \quad B \subset \mathcal{F}.$$

- The log-likelihood function of the process observed on the interval $[0, T]$ is

$$\mathcal{L}_T(f) := \sum_{k=1}^{K} \left[\int_0^T \log(\lambda_t^{(k)}(f))\mathrm{d}N_t^{(k)} - \int_0^T \lambda_t^{(k)}(f)\mathrm{d}t \right],$$

  where $\lambda_t^{(k)}(f)$ is the intensity associated with f.
- From the posterior distribution, we can build
  - estimates e.g. the posterior mean $\hat{f} := \mathbb{E}^{\pi}[f|N]$
  - credible regions (uncertainty measure) $\hat{C} = \hat{C}(N)$ with $\Pi(\hat{C}|N) \geq 1 - \alpha$
  - predictors $\widehat{\lambda}_t^{(k)} := \int \lambda_t^{(k)}(f)\mathrm{d}\Pi(f|N)$
  - etc.

The frequentist analysis of the Bayesian approach

- Our observations N are generated from a true parameter f^*
- Assumptions on the true parameters:
  - stationarity condition (C3): ψ is L-Lipschitz with $L > 0$ and

$$L \times \max_{\ell} \sum_{k} \|h_{\ell k}^{*+}\|_1 < 1$$

  - Identifiability: ψ restricted to I is bijective from I to $J = \psi(I)$
  - ψ^{-1} is L'- Lipschitz on J for $L' > 0$
- We fix a prior Π on the set $\mathcal{F}$ of parameters f such that
  - the ν_k's are positive
  - the $h_{\ell k}$'s are bounded and are supported by $[0, A]$
  - $L \times \max_\ell \sum_k \|h_{\ell k}\|_1 < 1$
- We consider the posterior distribution $\Pi(\cdot|N)$

$$\Pi(B|N) = \frac{\int_B \exp(\mathcal{L}_T(f))\mathrm{d}\Pi(f)}{\int_{\mathcal{F}} \exp(\mathcal{L}_T(f))\mathrm{d}\Pi(f)}, \quad B \subset \mathcal{F}.$$

- For a distance d, we derive posterior concentration rates: for $\epsilon_T \to 0$, when $T \to +\infty$,

$$\mathbb{E}_{f^*}\left[\Pi\left(d(f^*, f) > \epsilon_T|N\right)\right] = o(1).$$

Posterior concentration rates for estimating f^*

- For a distance d, posterior concentration means that for $\epsilon_T \to 0$, when $T \to +\infty$,

$$\mathbb{E}_{f^*}\left[\Pi\left(d(f,f^*) > \epsilon_T | N\right)\right] = o(1).$$

We study posterior concentration rates for d the classical $\mathbb{L}_1$-distance:

$$d(f,f^*) := \|f - f^*\|_1 := \sum_{k=1}^{K} |\nu_k - \nu_k^*| + \sum_{k=1}^{K}\sum_{\ell=1}^{K} \|h_{\ell k} - h_{\ell k}^*\|_1$$

- We apply the standard Ghosal, Ghosh and van der Vaart approach and write

$$\Pi\left(B|N\right) = \frac{\int_B \exp(\mathcal{L}_T(f))\mathrm{d}\Pi(f)}{\int_{\mathcal{F}} \exp(\mathcal{L}_T(f))\mathrm{d}\Pi(f)} = \frac{\int_B \exp(\mathcal{L}_T(f) - \mathcal{L}_T(f^*))\mathrm{d}\Pi(f)}{\int_{\mathcal{F}} \exp(\mathcal{L}_T(f) - \mathcal{L}_T(f^*))\mathrm{d}\Pi(f)} =: \frac{N_T}{D_T}.$$

- We deal with the numerator by using $\mathbb{L}_1$-tests, so we need convenient concentration inequalities
- We deal with the denominator by controlling the Kullback-loss on

$$\overline{B}_\infty(\epsilon_T, R) := \{f \in \mathcal{F} : \ |\nu_k - \nu_k^*| \le \epsilon_T, \ \|h_{\ell k} - h_{\ell k}^*\|_\infty \le \epsilon_T, \ \|h_{\ell k}\|_\infty \le R \ \forall \ell, k\}$$

Posterior concentration rates for estimating f^*

Theorem

We assume

$$\inf_x \psi(x) > 0. \tag{1.1}$$

Let Π be a prior distribution and $\epsilon_T \to 0$ such that

$$\log^3(T) = O(T\epsilon_T^2).$$

(i) *There exists $R > 0$ such that*

$$\Pi\left(\overline{B}_\infty(\epsilon_T, R)\right) \geq e^{-\square T\epsilon_T^2}$$

(ii) *There exists a subset $\mathcal{F}_T \subset \mathcal{F}$, such that*

$$\frac{\Pi\left(\mathcal{F}_T^c\right)}{\Pi\left(\overline{B}_\infty(\epsilon_T, R)\right)} \leq e^{-\square T\epsilon_T^2}$$

(iii) *The metric entropy of the space $\mathcal{F}_T$ for the $\mathbb{L}_1$-norm satisfies*

$$\log \mathcal{N}(\epsilon_T, \mathcal{F}_T, \|.\|_1) \leq \square T\epsilon_T^2$$

Then, for C a constant large enough,

$$\mathbb{E}_{f^*}\left[\Pi\left(\|f - f^*\|_1 > C\epsilon_T | N\right)\right] = o(1).$$

Positive consequences of the theorem

- Assumptions are similar to those set for simple models like density estimation or regression. Prior models used for classical models can then be used for nonlinear Hawkes processes
- From previous results we derive the following result for frequentist estimates

Corollary

We assume conditions of the previous theorem are satisfied. If

$$\int \|f\|_1 \mathrm{d}\Pi(f) < +\infty,$$

then the posterior mean $\hat{f} = \mathbb{E}^{\pi}[f|N]$ *is converging to* f^* *at the rate* ϵ_T*: for* C *a constant large enough*

$$\mathbb{P}_{f^*}\left(\|\hat{f} - f^*\|_1 > C\epsilon_T\right) = o(1).$$

- Posterior concentration rates are obtained for random histogram priors based on random partitions. And on Hölder classes $\mathcal{H}(\beta, L)$, with $\beta \leq 1$, we obtain the posterior concentration rate

$$\epsilon_T = \left(\frac{\log T}{T}\right)^{\frac{\beta}{2\beta+1}},$$

which is optimal up to the logarithmic term.

Drawbacks of the theorem

- The use of
$$\overline{B}_\infty(\epsilon_T, R) := \{f \in \mathcal{F} : \ |\nu_k - \nu_k^*| \leq \epsilon_T, \ \|h_{\ell k} - h_{\ell k}^*\|_\infty \leq \epsilon_T, \ \|h_{\ell k}\|_\infty \leq R \ \forall \ell, k\}$$
prevents from using some classical priors such that mixtures of beta distributions. It can be replaced by
$$\overline{B}_2(\epsilon_T, R) := \{f \in \mathcal{F} : \ |\nu_k - \nu_k^*| \leq \epsilon_T, \ \|h_{\ell k} - h_{\ell k}^*\|_2 \leq \epsilon_T, \ \|h_{\ell k}\|_\infty \leq R \ \forall \ell, k\}$$
at the price of a $\sqrt{\log \log T}$-term in rates.
- Assumption (1.1) of the theorem, i.e. $\inf_x \psi(x) > 0$, is quite strong. It plays a key role to control the Kullback-Leibler divergence defined by
$$\mathbb{E}_{f^*}[\mathcal{L}_T(f^*) - \mathcal{L}_T(f)].$$
However, Assumption (1.1) is not fulfilled by the 4 instances of link functions. It can be overcome by replacing ψ with
$$\widetilde{\psi}(x) = \theta + \psi(x), \quad x \in \mathbb{R},$$
for $\theta > 0$ small (known or not). It is not very satisfying.
- The result of the theorem holds if $\inf_x \psi(x) = 0$ by replacing ϵ_T with $\epsilon_T \sqrt{\log T}$ and if we further assume that $\psi(x) > 0$ for any $x \in \mathbb{R}$ and $\sqrt{\psi}$ and $\log(\psi)$ are Lipschitz functions. It is satisfied by logit, sigmoid and exponential functions.

The case of the ReLU function $\psi(x) = \max(x, 0)$

- The result of the theorem holds if

$$\psi(x) = \max(x, 0)$$

by replacing ϵ_T with $\epsilon_T \log T$ and if we further assume that

$$\limsup_{T \to +\infty} \frac{1}{T} \mathbb{E}_{f^*} \left[\int_0^T \frac{1_{\{\lambda_t^{(k)}(f^*) > 0\}}}{\lambda_t^{(k)}(f^*)} dt \right] < +\infty, \quad \forall k \in [\![1; K]\!].$$

This is satisfied if for instance for any ℓ $h_{\ell k}^*$ is an histogram and for all t, $h_{\ell k}^*(t) \in \mathbb{Q}$

- The case

$$\psi(x) = \theta + \max(x, 0), \quad x \in \mathbb{R},$$

with θ unknown with positive components can be dealt with. Under the same assumptions, we also achieve the posterior concentration rate ϵ_T.

- Open question: Can we deal with more general link functions that vanish on a whole interval? With mild assumptions? Work in progress with Déborah Sulem in the frequentist setting.

Difficulties and technical tools

- Since $\mathbb{P}(\mathrm{d}N^{(k)}t = 1|\text{past before } t) = \lambda_t^{(k)}(f^*)$, the first step consists in obtaining rates for the stochastic loss defined through intensities:

$$d_{1,T}(f, f^*) := \frac{1}{T}\sum_{k=1}^{K}\int_0^T \left|\lambda_t^{(k)}(f) - \lambda_t^{(k)}(f^*)\right|\mathrm{d}t$$

  by using
  1. new Bernstein-type concentration inequalities for martingales
  2. a sharp control of the number of points falling in intervals

  See Hansen, Reynaud-Bouret and R. (2015)

- For point 2 and for the linear case $\psi(x) = x$, the cluster representation is the main tool.Crucial assumption: the $h^*_{\ell k}$'s are non negative

New probabilistic tools

- We cannot rely on the cluster representation anymore, which allows the Hawkes process N to be represented as a sum of independent processes,
- But Costa, Graham, Marsalle and Tran (2020) have studied Hawkes processes with signed reproduction functions by using renewal techniques: By setting

$$X_t := N|_{(t-A,t]}$$

and the regeneration times

$$\tau_j = \begin{cases} 0 & \text{if } j = 0 \\ \inf\{t \in (\tau_{j-1}, T] : X_{t^-} \neq \emptyset, X_t = \emptyset\} & \text{if } j \geq 1 \end{cases},$$

we have:

1. the point measure $(X_t)_t$ is a strong Markov process with positive recurrent state the null measure
2. almost surely, the variables $(\tau_j)_j$ are finite stopping times for N
3. if we set, $\tau_{J_T+1} = T$, the intervals $((\tau_j, \tau_{j+1}])_{j=0,\dots,J_T}$ form a partition of $(0, T]$.
4. the random measures $(N|_{[\tau_j,\tau_{j+1}]})_{j\geq 1}$ are i.i.d. (called excursion)
5. Moments properties: for some $\alpha > 0$,

$$\mathbb{E}\left[e^{\alpha(\tau_2-\tau_1)}\right] < \infty$$

6. An ergodic theorem and exponential concentration inequalities were established

Technical tools

- Second step: To move from rates on intensities to rates on parameters: based on controls (with large probability) of the stochastic distance by the deterministic one:

$$d_{1,T}(f,f^*) := \frac{1}{T}\sum_{k=1}^{K}\int_0^T \left|\lambda_t^{(k)}(f) - \lambda_t^{(k)}(f^*)\right| \mathrm{d}t \lesssim \|f - f^*\|_1$$

- With

$$\tau_j = \inf\{t \in (\tau_{j-1}, T] : X_{t-} \neq \emptyset, X_t = \emptyset\}$$

and $T_j^{(2)}$ the second event after τ_j

$$d_{1,T}(f,f^*) \geq \frac{1}{T}\sum_{k=1}^{K}\sum_{j=1}^{J_T}\int_{\tau_j}^{T_j^{(2)}\wedge\tau_{j+1}} \left|\lambda_t^{(k)}(f) - \lambda_t^{(k)}(f^*)\right| \mathrm{d}t =: \sum_{j=1}^{J_T} Z_{j,T}$$

We establish:
- a concentration inequality on J_T (number of excursions)
- the lower bound

$$T\mathbb{E}_f[Z_{j,T}] \gtrsim \|f - f^*\|_1$$

- a Bernstein-type concentration inequality on

$$S_J := \sum_{j=1}^{J}\left[Z_{j,T} - \mathbb{E}_{f^*}[Z_{j,T}]\right]$$

Consistency on the connectivity graph

- We consider the same setting and we consider the graph adjacency matrix $\delta^* = (\delta^*_{\ell k})_{\ell k} \in \{0,1\}^{K^2}$ by setting

$$\delta^*_{\ell k} = 0 \iff h^*_{\ell k} = 0.$$

- We assume that $h^*_{\ell k} = \delta^*_{\ell k} h^*$, with $h^* \neq 0$.
- We consider a hierarchical prior model Π, writing $h_{\ell k} = \delta_{\ell k} h$ such that
  - the $\delta_{\ell k}$'s are i.i.d. Bernoulli(p)-variables with $p \in (0,1)$
  - the prior Π_h on h satisfies

$$\Pi_h(\|h^* - h\|_\infty \leq \epsilon_T) \geq e^{-\square T \epsilon_T^2}$$

  - the prior on $\nu = (\nu_k)_k$ has a continuous positive density
- Under assumptions of the previous theorem,

$$\mathbb{E}_{f^*}[\Pi(\delta \neq \delta^* | N)] = o(1)$$

- Previous results can be extended to the case where $h^*_{\ell k} = \delta^*_{\ell k} h^*_k$, with $h^*_k \neq 0$. For this purpose, we further need

$$\Pi_\delta(\delta = \delta^*) \geq e^{-\square T \epsilon_T^2}$$

Consistency on the connectivity graph

- The Bayesian estimate, defined by

$$\hat{\delta}_{\ell k}(N) = 1 \iff \Pi(\delta_{\ell k} = 1|N) > \Pi(\delta_{\ell k} = 0|N)$$

satisfies

$$\mathbb{P}_{f^*}(\hat{\delta}(N) \neq \delta^*) = o(1)$$

- The general case $h^*_{\ell k} = \delta^*_{\ell k} h^*_{\ell k}$, with $h^*_{\ell k} \neq 0$ can be considered by using a more general loss function, namely

$$L(\hat{\delta}, f) = \sum_{l,k=1}^{K} 1_{\hat{\delta}_{lk}=0} 1_{\delta_{lk}=1} + 1_{\hat{\delta}_{lk}=1}\left(1_{\delta_{lk}=0} + 1_{\delta_{lk}=1} F(\|h_{lk}\|_1)\right),$$

with $F : \mathbb{R}^+ \to [0,1]$ non-increasing, with $F(0) = 1$. The risk of the estimator $\hat{\delta}$ is

$$\begin{aligned} r(\hat{\delta}, \Pi|N) &= \int_{\mathcal{F}} L(\hat{\delta}, f) d\Pi(f|N) \\ &= \sum_{l,k} 1_{\hat{\delta}_{lk}=0} \Pi(\delta_{lk} = 1|N) + 1_{\hat{\delta}_{lk}=1} \left[\Pi(\delta_{lk} = 0|N) + \mathbb{E}^{\Pi}(1_{\delta_{lk}=1} F(\|h_{lk}\|_1)|N)\right] \end{aligned}$$

Consistency on the connectivity graph

- The Bayesian estimate, defined by

$$\hat{\delta}_{\ell k}(N) = 1 \iff \Pi(\delta_{\ell k} = 1|N) > \Pi(\delta_{\ell k} = 0|N)$$

satisfies

$$\mathbb{P}_{f^*}(\hat{\delta}(N) \neq \delta^*) = o(1)$$

- The general case $h^*_{\ell k} = \delta^*_{\ell k} h^*_{\ell k}$, with $h^*_{\ell k} \neq 0$ can be considered by using a more general loss function, namely

$$L(\hat{\delta}, f) = \sum_{l,k=1}^{K} 1_{\hat{\delta}_{lk}=0} 1_{\delta_{lk}=1} + 1_{\hat{\delta}_{lk}=1}\big(1_{\delta_{lk}=0} + 1_{\delta_{lk}=1} F(\|h_{lk}\|_1)\big),$$

with $F : \mathbb{R}^+ \to [0,1]$ non-increasing, with $F(0) = 1$. The risk of the estimator $\hat{\delta}$ is

$$r(\hat{\delta}, \Pi|N) = \int_{\mathcal{F}} L(\hat{\delta}, f) d\Pi(f|N)$$

Then the Bayesian estimate, $\hat{\delta}^{\Pi,L}(N) = \arg\min_{\delta \in \{0,1\}^{K^2}} r(\delta, \Pi|N)$, verifies

$$\hat{\delta}^{\Pi,L}_{lk}(N) = 1 \iff \mathbb{E}^{\Pi}[(1 - F(\|h_{lk}\|_1))1_{\delta_{lk}=1}|N] \geq \Pi(\delta_{lk} = 0|N)$$

and

$$\mathbb{P}_{f^*}(\hat{\delta}(N) \neq \delta^*) = o(1)$$

Conclusions

- We propose new results for estimating multivariate nonlinear Hawkes processes, to take into account possible inhibition.
- We consider the nonparametric Bayesian setting and we propose a theory to derive $\mathbb{L}_1$-posterior concentration rates under quite mild assumptions if the link function is positive.
- We prove consistency on the connectivity graph.
- Difficulties to deal when the link function ψ vanishes. Partial results for the case $\psi(x) = \max(x, 0)$.
- Can these problems be circumvented in the frequentist setting? Can we deal with more general link functions? Or with infinite-memory linear and nonlinear models? Work in progress.

Thank you for your attention. Questions and remarks are welcomed!

Reference:

Sulem D., Rivoirard V. and Rousseau J. (2021) *Bayesian estimation of nonlinear Hawkes processes.* Submitted