Nonparametric Bayesian inference for Hawkes processes

Vincent Rivoirard

Université Paris-Dauphine

Joint work with

- DEBORAH SULEM
  Oxford University

- JUDITH ROUSSEAU
  Oxford University

Intensity of a point process

Definition (Point process)

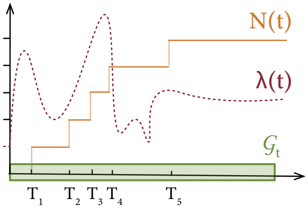
A point process $N = (N_t)_t$ is a random countable set of points of $\mathbb{R}$ or equivalently a non-decreasing integer-valued process.

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:

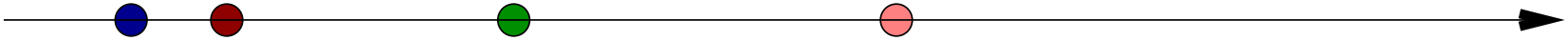
$$\lambda_t = \nu + \int_{-\infty}^{t-} h(t-u)\mathrm{d}N_u = \nu + \sum_{T\in N,\, T<t} h(t-T)$$

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

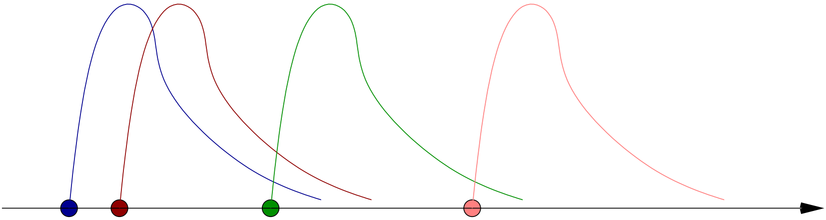
A univariate linear Hawkes process can be viewed as a branching process over an homogeneous Poisson process. Let $\nu > 0$ and $h \geq 0$ supported by $\mathbb{R}_+$.



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

Cluster representation for linear Hawkes processes

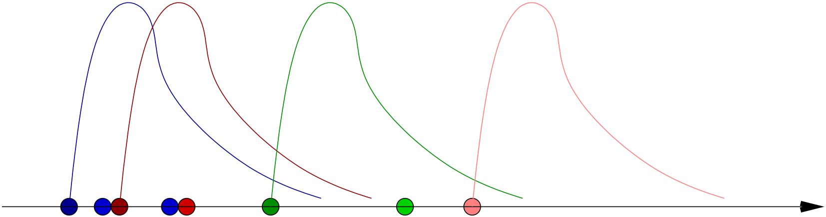
A univariate linear Hawkes process can be viewed as a branching process over an homogeneous Poisson process. Let $\nu > 0$ and $h \geq 0$ supported by $\mathbb{R}_+$.



- 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

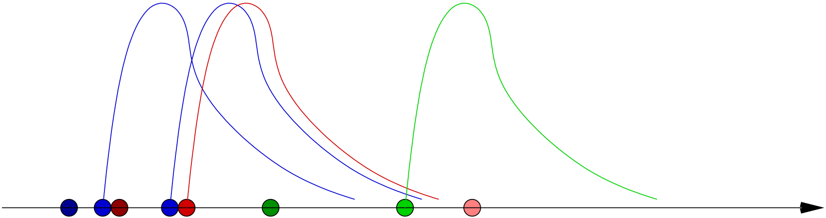
A univariate linear Hawkes process can be viewed as a branching process over an homogeneous Poisson process. Let $\nu > 0$ and $h \geq 0$ supported by $\mathbb{R}_+$.



- 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

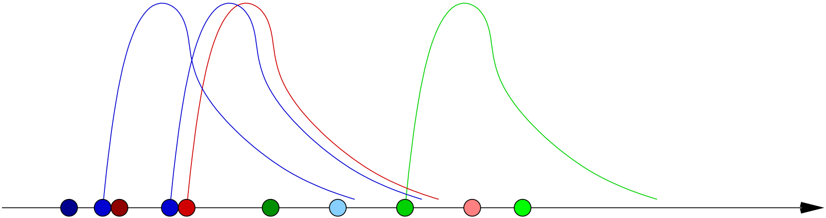
A univariate linear Hawkes process can be viewed as a branching process over an homogeneous Poisson process. Let $\nu > 0$ and $h \geq 0$ supported by $\mathbb{R}_+$.



- 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

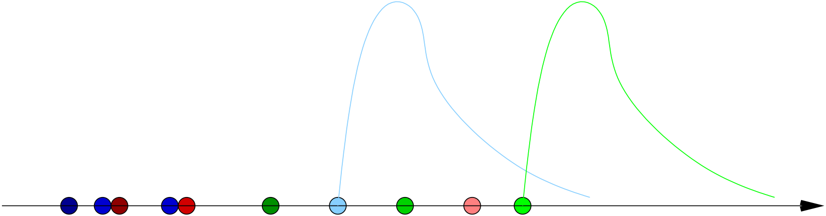
A univariate linear Hawkes process can be viewed as a branching process over an homogeneous Poisson process. Let $\nu > 0$ and $h \geq 0$ supported by $\mathbb{R}_+$.



- 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

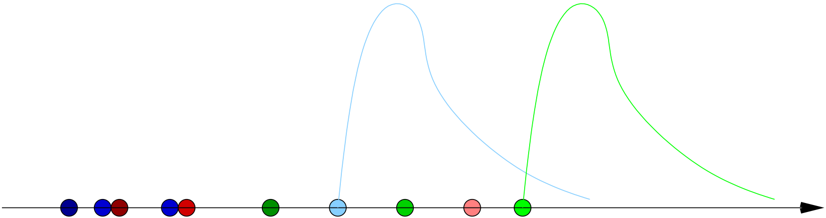
A univariate linear Hawkes process can be viewed as a branching process over an homogeneous Poisson process. Let $\nu > 0$ and $h \geq 0$ supported by $\mathbb{R}_+$.



- 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

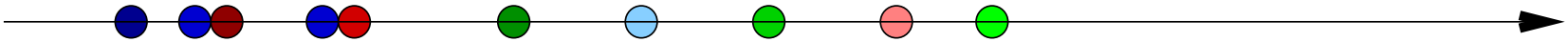
A univariate linear Hawkes process can be viewed as a branching process over an homogeneous Poisson process. Let $\nu > 0$ and $h \geq 0$ supported by $\mathbb{R}_+$.



- 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

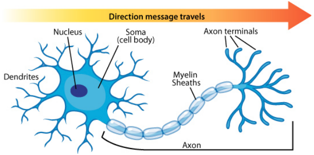
A univariate linear Hawkes process can be viewed as a branching process over an homogeneous Poisson process. Let $\nu > 0$ and $h \geq 0$ supported by $\mathbb{R}_+$.

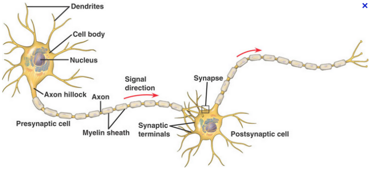


- Ancestors: Realizations of a Poisson Process with $\lambda_t = \nu$
- Each ancestor can give birth to children according to a P.P. with $\lambda_t = h(t)$
- Each child can give birth to children according to a P.P. with $\lambda_t = h(t)$
- Extinction if $\int_0^{+\infty} h(t)\mathrm{d}t < 1$
- Hawkes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

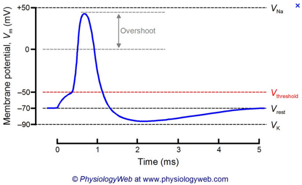
Multivariate Hawkes process: Neurobiological motivations

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.

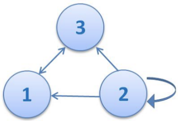
Multivariate Hawkes processes

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

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

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

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



- 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

Φ_k: link function. Typical examples:
- Ex (linear): $\Phi_k(x) = x$ [requires $h_{\ell k} \geq 0$]
- Ex (nonlinear): $\Phi_k(x) = \max(x, 0)$
- Ex (nonlinear): $\Phi_k(x) = \exp(x)$

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

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

Theorem (Brémaud and Massoulié (1996))

Existence and uniqueness of a stationary distribution for N:

- if $\forall k \in [\![1; K]\!]$ $\|\Phi_k\|_\infty < \infty$ or

- if $\forall k \in [\![1; K]\!]$ Φ_k is 1-Lipschitz and the matrix Γ with entries $\Gamma_{\ell k} = \|h_{\ell k}\|_1$ 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:

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

  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: $\Phi_k(x) = x$
  - 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

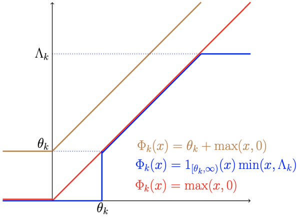
Inference for nonlinear Hawkes models

- We now consider

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

with in mind 3 examples:
- Model 1: $\Phi_k(x) = \theta_k + \max(x,0)$
- Model 2: $\Phi_k(x) = 1_{[\theta_k,\infty)}(x)\min(x,\Lambda_k)$
- Model 3: $\Phi_k(x) = \max(x,0)$

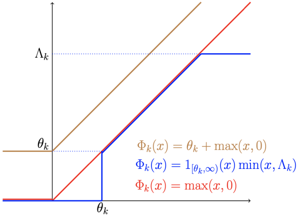
with $0 < \theta_k < \Lambda_k$



Inference for nonlinear Hawkes models

- We now consider

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



with in mind 3 examples:
- Model 1: $\Phi_k(x) = \theta_k + \max(x, 0)$
- Model 2: $\Phi_k(x) = 1_{[\theta_k,\infty)}(x) \min(x, \Lambda_k)$
- Model 3: $\Phi_k(x) = \max(x, 0)$

with $0 < \theta_k < \Lambda_k$

- Statistical Goals:
  1. 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 in mind $T \to +\infty$). The Φ_k's are known
  2. The θ_k's are unknown. With $\theta = (\theta_k)_{k \in [\![1;K]\!]}$, Estimation of (f^*, θ). The Λ_k's are known
  3. We consider the Bayesian approach
- Assumptions:
  1. The $h_{\ell k}^*$'s are bounded and have support $[0, A]$, with $A < \infty$ known
  2. The matrix Γ^* with entries $\Gamma_{\ell k}^* = \|h_{\ell k}^*\|_1$ has a spectral norm < 1. This implies existence and uniqueness of a stationary distribution

Identifiability

- Remember

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

with 3 possible models:
- Model 1: $\Phi_k(x) = \theta_k + \max(x,0)$
- Model 2: $\Phi_k(x) = 1_{[\theta_k,\infty)}(x)\min(x,\Lambda_k)$
- Model 3: $\Phi_k(x) = \max(x,0)$

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

- Identifiability:
  - Models 1, 2 and 3: Estimation of f^*: we need $\forall\,(k,\ell)\in[\![1;K]\!]$

$$\|(h_{\ell k}^*)^-\|_\infty < \begin{cases}\nu_k^* & \text{(Models 1 and 3)}\\ \nu_k^* - \theta_k & \text{(Model 2)}\end{cases}$$

  - Models 1 and 2: Estimation of (f^*,θ): We also need: $\forall k\ \exists\ell,\ (h_{\ell k}^*)^-(0) = 0$ and $(h_{\ell k}^*)^-$ is continuous and increasing on $[0, v_0)$ for some $v_0 > 0$

The Bayesian statistical approach

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

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

where $\lambda_t^{(k)}(f)$ is the intensity associated with $f = (\nu_k, (h_{\ell k})_{\ell \in [\![1;K]\!]})_{k \in [\![1;K]\!]}$

- We fix a prior Π on the set $\mathcal{F}$[1] of parameters f such that (for Models 1 and 3)
  - the $h_{\ell k}$'s are bounded and are supported by $[0, A]$
  - the matrix with entries $\|h_{\ell k}\|_1$ has a spectral radius < 1
  - $\forall (k, \ell) \in [\![1;K]\!]$, $\|(h_{\ell k})^-\|_\infty < \nu_k$
- We study the posterior distribution $\Pi(\cdot|N)$, with for any $B \subset \mathcal{F}$,

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

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

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

From the posterior distribution, we can build estimates, credible sets, etc.

[1] modified in a natural way for Model 2 and for estimating (f^*, θ)

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

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

Posterior concentration rates for estimating f^*

Theorem

Assume

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

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*

$$\Pi\left(B(\epsilon_T, R)\right) \geq e^{-\Box T\epsilon_T^2}$$

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

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

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

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

Then, for C *a constant large enough,*

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

Discussion - Rates for Bayesian estimators

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

- Model 1: $\Phi_k(x) = \theta_k + \max(x,0)$
- Model 2: $\Phi_k(x) = 1_{[\theta_k,\infty)}(x)\min(x,\Lambda_k)$
- Model 3: $\Phi_k(x) = \max(x,0)$

So, the previous condition

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

is satisfied for Models 1 and 2. For Model 3, it is satisfied if for instance for any ℓ $h_{\ell k}^*$ is an histogram and for all t, $h_{\ell k}^*(t) \in \mathbb{Q}$.

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

Posterior rates for estimating (f^*, θ^*) - Prior models

- We consider (change notations $\theta_k \to \theta_k^*$):
  - Model 1: $\Phi_k(x) = \theta_k^* + \max(x, 0)$
  - Model 2: $\Phi_k(x) = 1_{[\theta_k^*, \infty)}(x) \min(x, \Lambda_k)$

  We estimate $\theta^* = (\theta_k^*)_{k \in [\![;K]\!]}$
- Adapting naturally the setting to the problem of estimating (f^*, θ^*), we obtain:

$$\mathbb{E}_{f^*}\left[\Pi\left(\|\theta - \theta^*\|_1 + \|f - f^*\|_1 > C\epsilon_T \middle| N\right)\right] = o(1) \quad \text{(Model 1)}$$

$$\mathbb{E}_{f^*}\left[\Pi\left(\|\theta - \theta^*\|_1 > C\sqrt{\epsilon_T} \middle| N\right) + \Pi\left(\|f - f^*\|_1 > C\epsilon_T \middle| N\right)\right] = o(1) \quad \text{(Model 2)}$$

- We also derive a Bayesian estimate $\hat{\theta}$ such that under mild assumptions

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

$$\mathbb{P}_{f^*}\left(\|\hat{\theta} - \theta^*\|_1 > C\sqrt{\epsilon_T}\right) = o(1) \quad \text{(Model 2)}$$

Posterior rates for estimating (f^*, θ^*) - Prior models

- We consider (change notations $\theta_k \to \theta_k^*$):
  - Model 1: $\Phi_k(x) = \theta_k^* + \max(x, 0)$
  - Model 2: $\Phi_k(x) = 1_{[\theta_k^*, \infty)}(x) \min(x, \Lambda_k)$

  We estimate $\theta^* = (\theta_k^*)_{k \in [\![;K]\!]}$
- Adapting naturally the setting to the problem of estimating (f^*, θ^*), we obtain:

$$\mathbb{E}_{f^*}\left[\Pi\left(\|\theta - \theta^*\|_1 + \|f - f^*\|_1 > C\epsilon_T \middle| N\right)\right] = o(1) \quad \text{(Model 1)}$$

$$\mathbb{E}_{f^*}\left[\Pi\left(\|\theta - \theta^*\|_1 > C\sqrt{\epsilon_T} \middle| N\right) + \Pi\left(\|f - f^*\|_1 > C\epsilon_T \middle| N\right)\right] = o(1) \quad \text{(Model 2)}$$

- We also derive a Bayesian estimate $\hat{\theta}$ such that under mild assumptions

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

$$\mathbb{P}_{f^*}\left(\|\hat{\theta} - \theta^*\|_1 > C\sqrt{\epsilon_T}\right) = o(1) \quad \text{(Model 2)}$$

- 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

$$\epsilon_T = (\log T)^{\frac{3\beta}{2\beta+1}} T^{-\frac{\beta}{2\beta+1}}$$

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. an ergodic theorem (Reynaud-Bouret and Roy (2003))
  3. a sharp control of the number of points falling in intervals
- For points 2 and 3, the cluster representation is the main tool. In particular:
  Lemma: *Assume* $\|\Gamma\| < 1$ *and consider* ζ *such that* $0 \leq \zeta \leq \frac{1-\|\Gamma\|}{2\sqrt{M}} \log\left(\frac{1+\|\Gamma\|}{2\|\Gamma\|}\right)$. *Then, for any ancestor of type* ℓ, *if* W^ℓ *the number of points in its cluster,*

$$\mathbb{E}[\exp(\zeta W^\ell)] \leq \frac{1+\|\Gamma\|}{2\|\Gamma\|}.$$

- Crucial assumption: $\Phi_k(x) = x$ and the $h_{\ell k}$'s non negative
- See Hansen, Reynaud-Bouret and R (2015)

Difficulties and 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:

$$\lambda_t^{(k)}(f_1) := \nu_{1k} + \sum_{\ell=1}^{K} \int_{-\infty}^{t-} h_{1\ell k}(t-u)\mathrm{d}N^{(\ell)}(u) \quad \lambda_t^{(k)}(f_2) := \nu_{2k} + \sum_{\ell=1}^{K} \int_{-\infty}^{t-} h_{2\ell k}(t-u)\mathrm{d}N^{(\ell)}(u)$$

$$d_{1,T}(f_1,f_2) := \frac{1}{T}\sum_{k=1}^{K}\int_0^T \left|\lambda_t^{(k)}(f_1)-\lambda_t^{(k)}(f_2)\right|\mathrm{d}t \lesssim \|f_1-f_2\|_1 := \sum_{k=1}^{K}|\nu_{1k}-\nu_{2k}| + \sum_{k=1}^{K}\sum_{\ell=1}^{K}\|g_{1\ell k}-h_{2\ell k}\|_1$$

With $J_T \to +\infty$ well chosen (in particular $J_T = o(T)$) and

$$Z_{k,m} := \int_{\frac{2mT}{2J_T}}^{\frac{(2m+1)T}{2J_T}} \left|\lambda_t^{(k)}(f_1)-\lambda_t^{(k)}(f_2)\right|\mathrm{d}t = \int_{\frac{2mT}{2J_T}}^{\frac{(2m+1)T}{2J_T}} \left|\nu_{1k}-\nu_{2k}+\sum_{\ell=1}^{M}\int_{t-A}^{t^-}(g_{1\ell k}-g_{2\ell k})(t-s)dN_s^k\right|\mathrm{d}t$$

The $Z_{k,m}$'s only depend on points of the process of the interval $\mathcal{I}_{m,T} := \left[\frac{2mT}{2J_T} - A; \frac{(2m+1)T}{2J_T}\right]$, so they are "almost independent" (cluster representation), since $\max(\mathcal{I}_{m,T}) \ll \min(\mathcal{I}_{m+1,T})$

$$\begin{aligned} Td_{1,T}(f,f^*) &\geq \max_{1\leq k\leq K}\left\{\sum_{m=1}^{J_T-1}\mathbb{E}[Z_{k,m}] + \sum_{m=1}^{J_T-1}\left[Z_{k,m}-\mathbb{E}[Z_{k,m}]\right]\right\} \\ &\gtrsim T\|f_1-f_2\|_1 + \max_{1\leq k\leq K}\left\{\sum_{m=1}^{J_T-1}\left[\tilde{Z}_{k,m}-\mathbb{E}[\tilde{Z}_{k,m}]\right]\right\} \gtrsim T\|f_1-f_2\|_1 \end{aligned}$$

- Crucial assumption: $\Phi_k(x) = x$ and the $h_{\ell k}$'s non negative
- See Donnet, R and Rousseau (2020)

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 (2018) 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

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

Reference:

SULEM D., RIVOIRARD V. AND ROUSSEAU J. (2020) *Bayesian estimation of nonlinear Hawkes processes.* In preparation