Non parametric Bayesian estimation for Hawkes processes

Sophie Donnet, Judith Rousseau, Vincent Rivoirard

Non-parametric Bayesian estimation for intensity function in Hawkes processes

Sophie Donnet, Vincent Rivoirard, Judith Rousseau
UMR AgroParisTech/INRA 518. Team MORSE. INRA. Paris France
`sophie.donnet@agroparistech.fr`

Introduction

Hawkes process is the most basic self-exciting counting process. It as been widely used to modelize earthquakes [1-3], financial or economical data but also in neurobiology. The Bayesian estimation of the conditional intensity function has been tackled by [4] in a parametric context. In this work, we propose to consider a non-parametric Bayesian estimation

Hawkes processes

- The Hawkes process $(N_t)_{t\in\mathbb{R}}$ is defined by its conditional intensity

$$\lambda(t) = \nu + \int_{-\infty}^{t^-} h(t-u)dN_u = \nu + \sum_{T\in N, T<t} h(t-T) \qquad (1)$$

where

→ ν is a positive parameter,
→ h a nonnegative function with support on $\mathbb{R}^+$ and $\int h < 1$
→ dN_u is the point measure associated to the process.

- **Interpretation** : $\lambda(t)dt = \mathbb{P}(\text{point } \in [t; t+dt]$ conditionally to the past before t).
  So from (1) : constant rate ν and also *all the previous occurrences condition the apparition of an occurrence at t.*
- **Stationarity** :

→ We assume that h has a bounded support $[0, s_{max}]$ such that $\int_0^{s_{max}} h(u)du < 1$
→ This condition guarantees the existence of a stationary version of the process
→ And the number of events on $[T^*, T^*+\Delta]$:

$$E[N^*] = E[N_{[T^*,T^*+\Delta]}] = \Delta^* \frac{\nu}{1-\int_0^{s_{max}} h(u)du}$$

Likelihood and prior distributions

- **Parameters of interest** : h and ν
- **Observations and likelihood** : computation of the likelihood on $[T^*-s_{max}, T^*+\Delta]$

$$\log \ell(N; h, \nu) = \sum_{T_i\in[T^*,T^*+\Delta])} \log\lambda(T_i) - \int_{T^*}^{T^*+\Delta} \lambda(t)dt$$

- **Prior distributions**

→ h step function. $h(t) = \sum_{k=0}^{K-1} \alpha_k \mathbb{1}_{[s_k,s_{k+1}]}(t) = l_h \sum_{k=0}^{K-1} \overline{\alpha}_k \mathbb{1}_{[s_k,s_{k+1}]}(t)$ with
  * $s_K = s_{max}$ known and $s_0 = 0$
  * $l_h = \int_0^{s_max} h(t)dt < 1$
  * $\sum_{k=0}^{K-1} \overline{\alpha}_k(s_{k+1}-s_k) = 1$

→ $l_h \sim \mathcal{B}(a_h, b_h)$
→ Prior on s :

$$\left(\frac{s_0}{s_{max}} \dots \frac{s_K}{s_{max}}\right) \sim \mathcal{D}ir(a_1,\dots,a_K) \qquad (2)$$

→ Prior on $\overline{\alpha}_1,\dots,\overline{\alpha}_K$: $\overline{\alpha}_0(s_1-s_0),\dots,\overline{\alpha}_K(s_K-s_{K-1}) \sim \mathcal{D}ir(b_1,\dots,b_K)$
→ $\nu \sim \Gamma(a_\nu, b_\nu)$

Sampling of the posterior distribution : MCMC

At iteration (l),

1. **Sampling of** ν with a classical random walk on $\log\nu$ to guaranty the positivity
2. Sampling of l_h with a random walk on $[0,1]$

$$l_h^c = \Phi^{-1}(\Phi((l_h)^{l-1}) + \varepsilon), \qquad \varepsilon \sim 0.5\mathcal{N}(0,\rho_1) + 0.5\mathcal{N}(0,\rho_2), \quad \rho_1 < \rho_2$$

3. **Sampling of** $\overline{\alpha}_k$
   - $k \sim \mathcal{U}_{\{0,\dots,K-1\}}$ and $\overline{\alpha}_k^c = \overline{\alpha}_k^{l-1} + \xi$, with $\xi \sim \sum_{r=1}^4 0.25\mathcal{N}(0,\rho_r)$
   - Correction of the others $\overline{\alpha}$ to keep $\sum_{k=0}^K \overline{\alpha}_k(s_{k+1}-s_k) = 1$:

$$\begin{aligned}\Delta_k &= \overline{\alpha}_k^c(s_{k+1}-s_k) - \overline{\alpha}_k^{l-1}(s_{k+1}-s_k)\\ \overline{\alpha}_j^c(s_{j+1}-s_j) &= \overline{\alpha}_j^{l-1}(s_{j+1}-s_j) - \frac{\Delta_k}{K-1} \quad \forall j \neq k.\end{aligned}$$

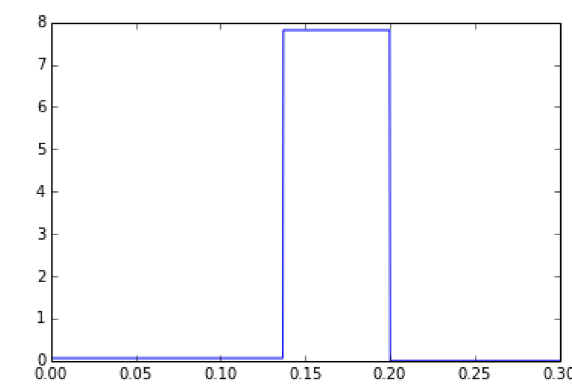
4. **Sampling of** s
   - $k \sim \mathcal{U}_{\{1,\dots,K-1\}}$ and $s_k^c \sim Triang(s_{k-1}^{l-1}, sl-1_k, s_{k+1}^{l-1})$
   - Correction of the $\overline{\alpha}_{k-1}^l$ and $\overline{\alpha}_k^l$ to keep the integral equal to 1:

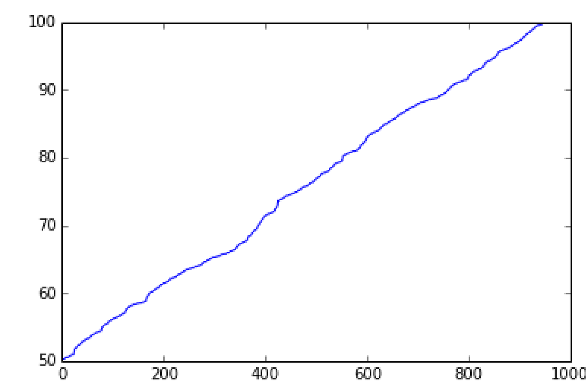
$$\overline{\alpha}_{k-1}^c = \overline{\alpha}_{k-1}^l \frac{s_k^{l-1} - s_{k-1}^{l-1}}{s_k^c - s_{k-1}^{l-1}} \quad \text{and} \quad \overline{\alpha}_k^c = \overline{\alpha}_{k-1}^l * \frac{s_{k+1}^{l-1} - s_k^{l-1}}{s_{k+1}^{l-1} - s_k^c}$$

First numerical results

$T^* = 50 \quad \Delta = 50, \quad l_h = 0.5, \quad \nu = 10$

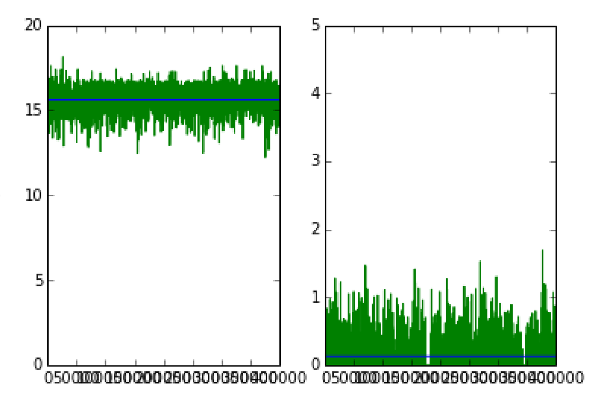
- **Data**



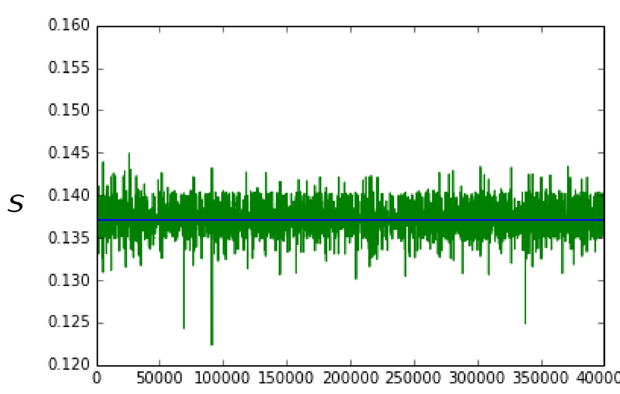


- **MCMC** - 200 000 iterations

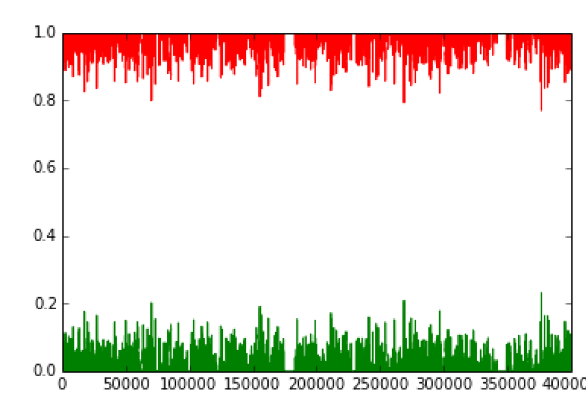
→ α



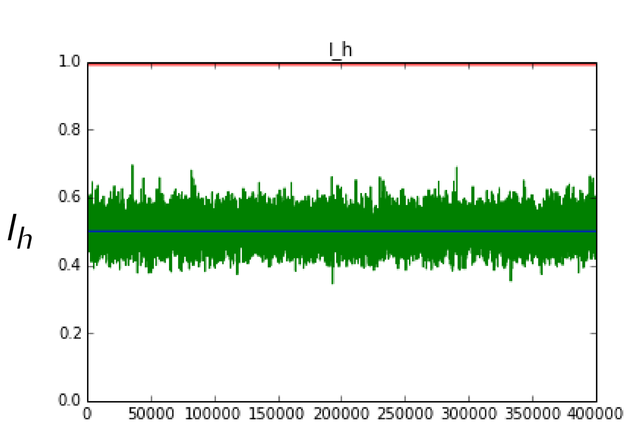
s



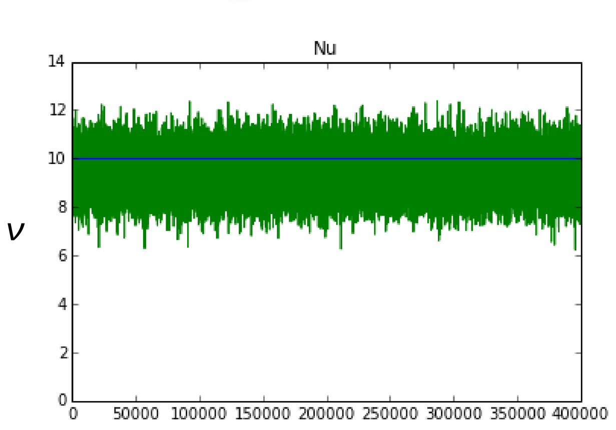
→ $\overline{\alpha}_k^{(l)}(s_{k+1}^{(l)} - s_k^{(l)})$



→ l_h



ν



Perspectives of work

- Unknown K : reversible jump
- $\pi(\overline{\alpha}_k = 0) > 0$
- $\pi(\overline{\alpha}_k < 0) > 0$ to modelize inhibition
- Multi-dimensional Hawkes processes to modelize interactions

Bibliography

[1] Hawkes AG, Adamopoulos L (1973) Cluster models for earthquakesÑregional comparisons. Bull Int Statist Inst 45:454–461

[2] Ogata Y (1988) Statistical models for earthquake occurrences and residual analysis for point processes. J Am Stat Assoc 83(401):9–27

[3] Ogata Y (1998) Space-time point-process models for earthquake occurrences. Ann Inst Stat Math 50(2):379–402

[4] Rasmussen J. (2013) Bayesian Inference for Hawkes Processes. Methodol Comput Appl Probab (2013) 15:623Ð642

[5] Daley, D. J. and Vere-Jones, D. (2005). An Introduction to the Theory of Point Processes. Springer Series in Statistics I. Springer, New York. MR0950166

[6] Reynaud Bouret P. and Schbath, S. (2010). Adaptive estimation for Hawkes processes; Application to genome analysis. The Annals of Statistics 2010, Vol. 38, No. 5, 2781Ð2822