Estimation of local independence graphs via Hawkes processes to unravel functional neuronal connectivity

C. Tuleau-Malot (Nice), V. Rivoirard (Dauphine), N.R. Hansen (Copenhagen),P. Reynaud-Bouret(Nice),
F. Grammont (Nice), T. Bessaih, R. Lambert, N. Leresche (Paris 6)

Contents

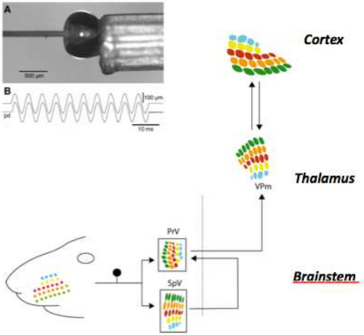
Contents

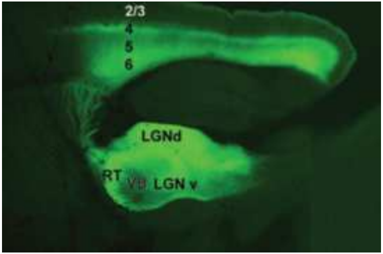
Contents

Contents

Contents

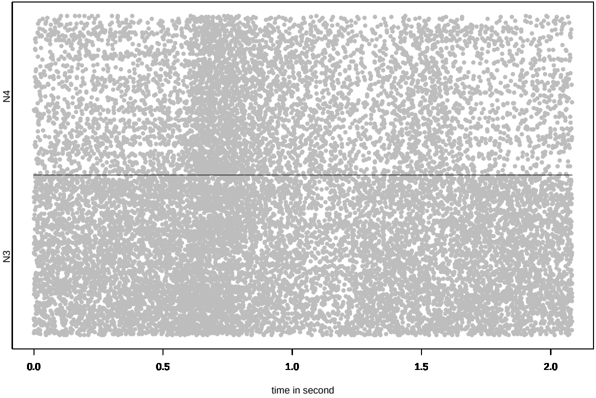
Description of the data (Equipe RNRP de Paris 6)





Data = spike trains

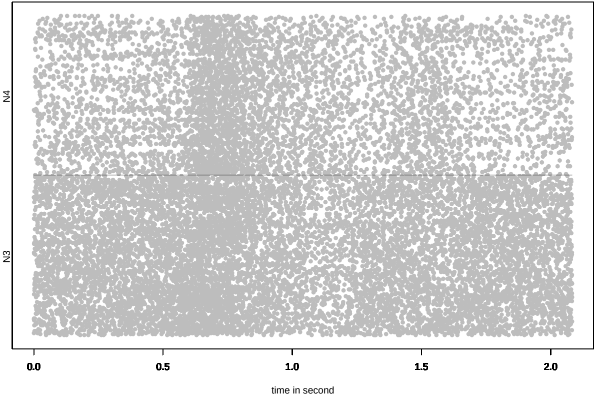
monkey trained to touch the correct target when illuminated



spike = point

Data = spike trains

monkey trained to touch the correct target when illuminated

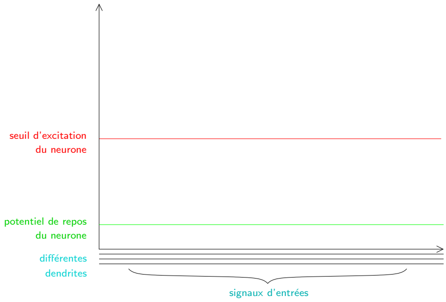


spike = point

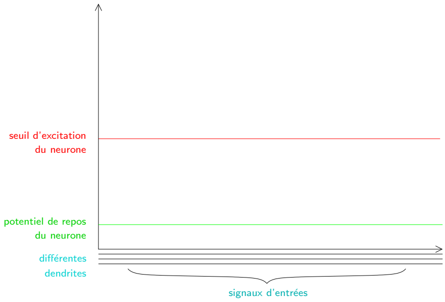
$\rightarrow$ point processes

Synaptic integration

without synchronisation

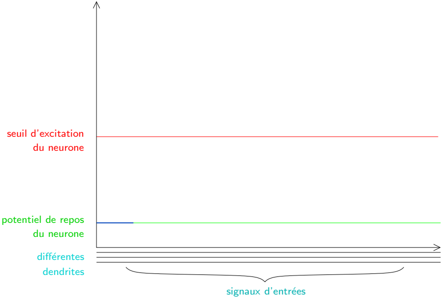


with synchronisation

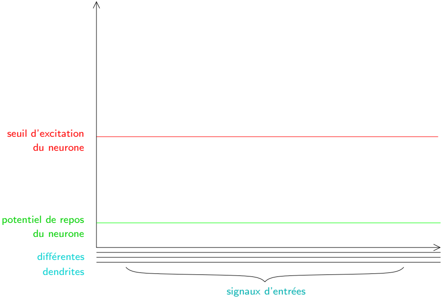


Synaptic integration

without synchronisation

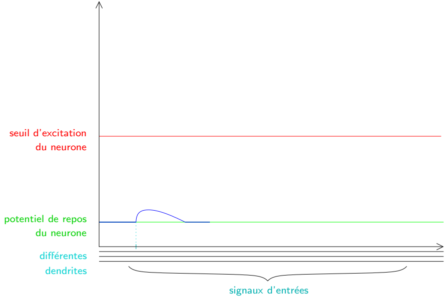


with synchronisation

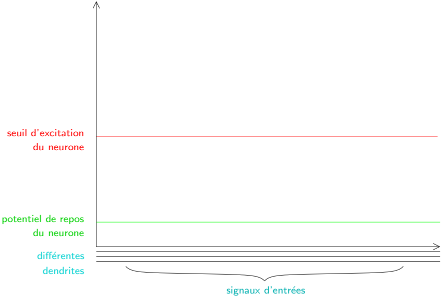


Synaptic integration

without synchronisation

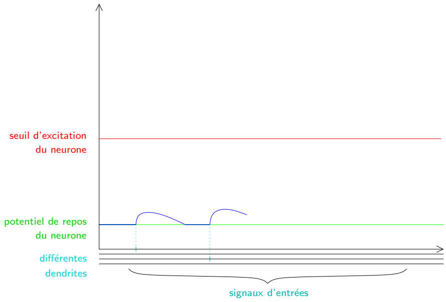


with synchronisation

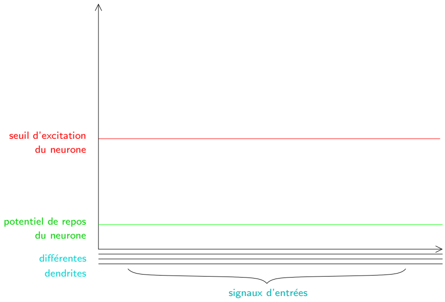


Synaptic integration

without synchronisation

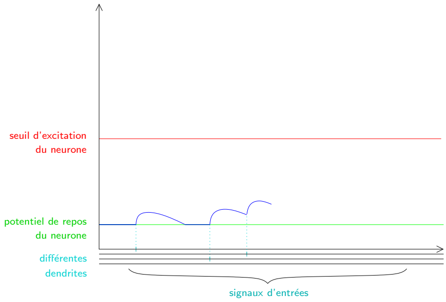


with synchronisation

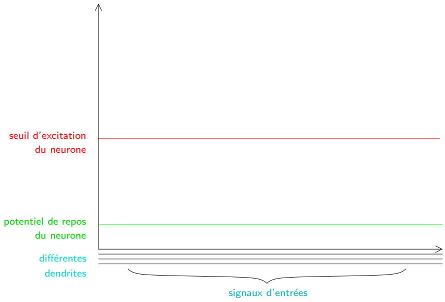


Synaptic integration

without synchronisation

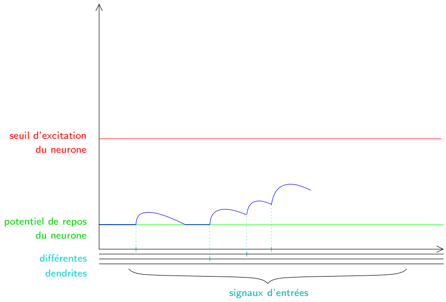


with synchronisation

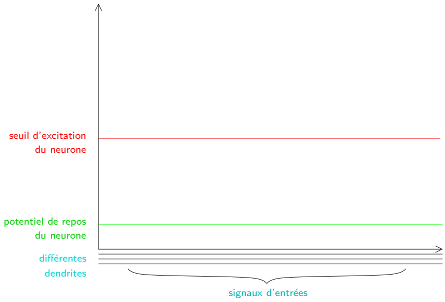


Synaptic integration

without synchronisation

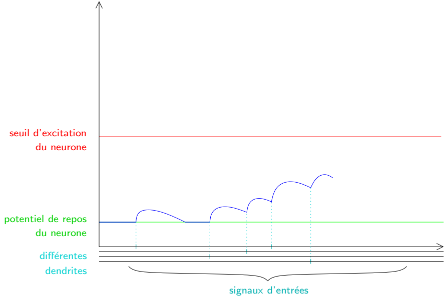


with synchronisation

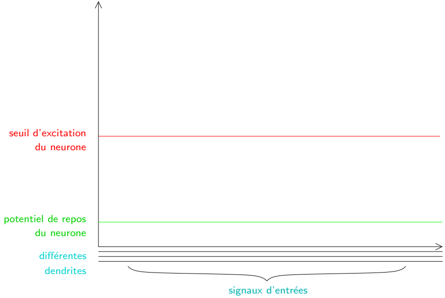


Synaptic integration

without synchronisation

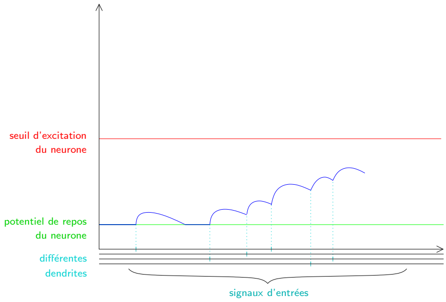


Avec synchronisation

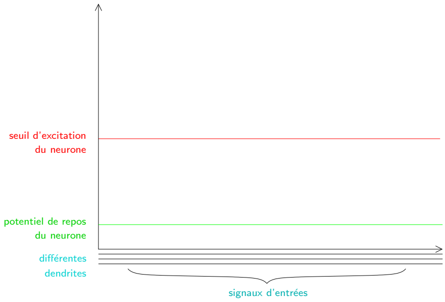


Synaptic integration

without synchronisation

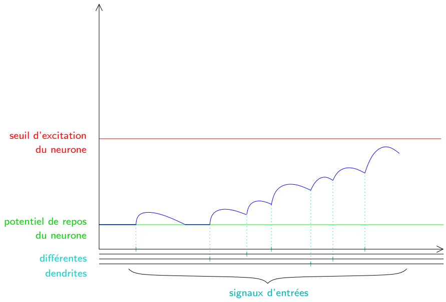


with synchronisation

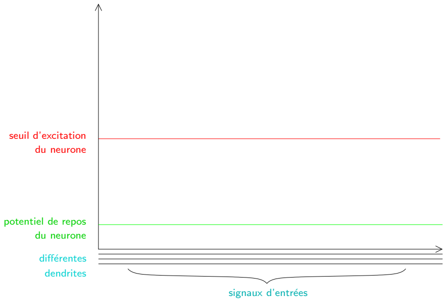


Synaptic integration

without synchronisation

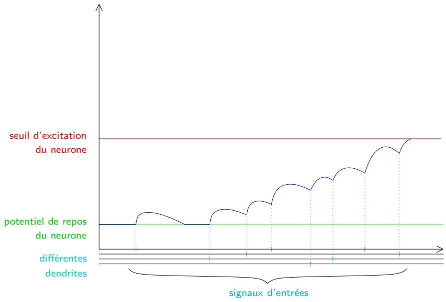


Avec synchronisation

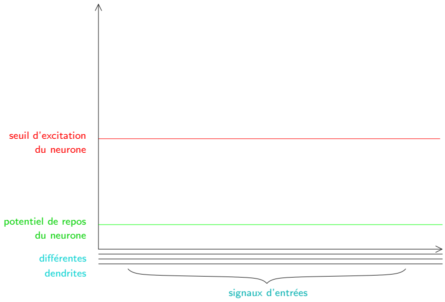


Synaptic integration

without synchronisation

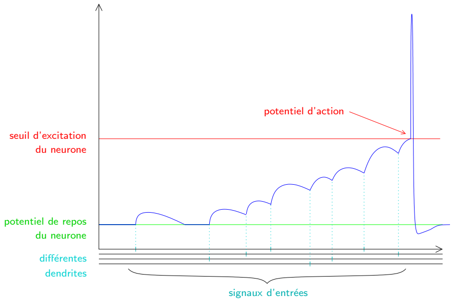


with synchronisation

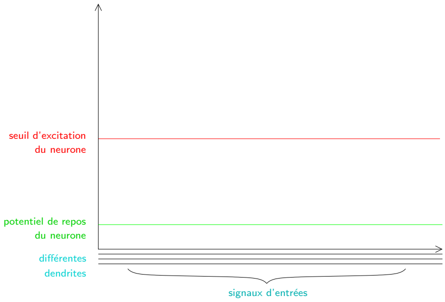


Synaptic integration

without synchronisation

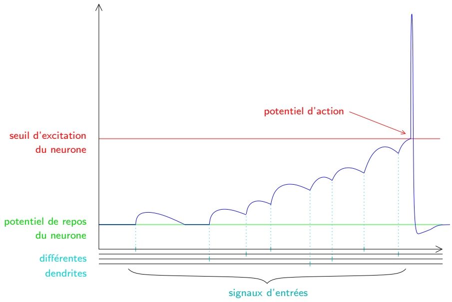


with synchronisation

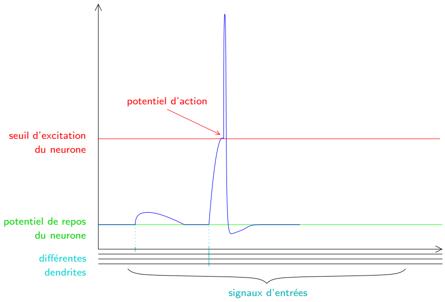


Synaptic integration

without synchronisation

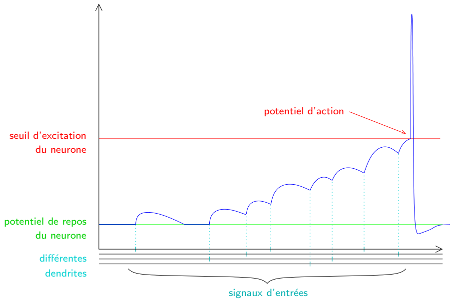


with synchronisation

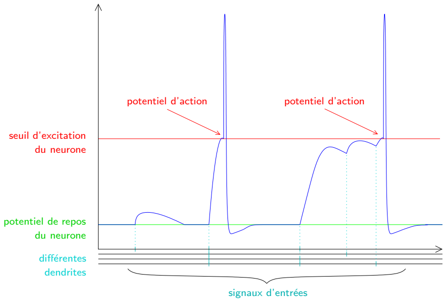


Synaptic integration

without synchronisation



with synchronisation

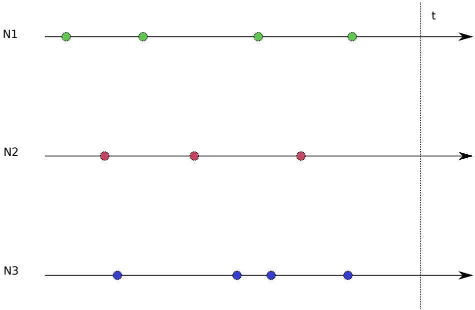


Point processes and conditional intensity

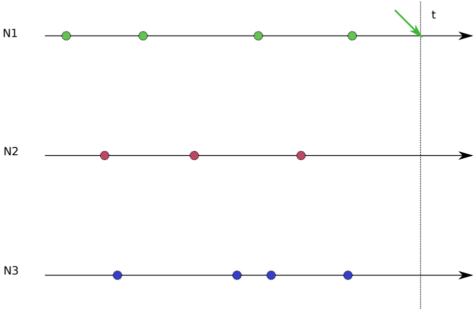
$$\underbrace{dN_t}_{\text{Nbr observed points in } [t, t+dt]} = \underbrace{\overbrace{\lambda(t)}^{\textbf{intensity}} \; dt}_{\text{Expected Number given the past before } t} + \underbrace{\text{noise}}_{\textit{Martingales differences}}$$

$$\begin{aligned} \lambda(t) &= \text{instantaneous frequency} \\ &= \textbf{random}\text{, depends on previous points} \end{aligned}$$

Multivariate Hawkes processes

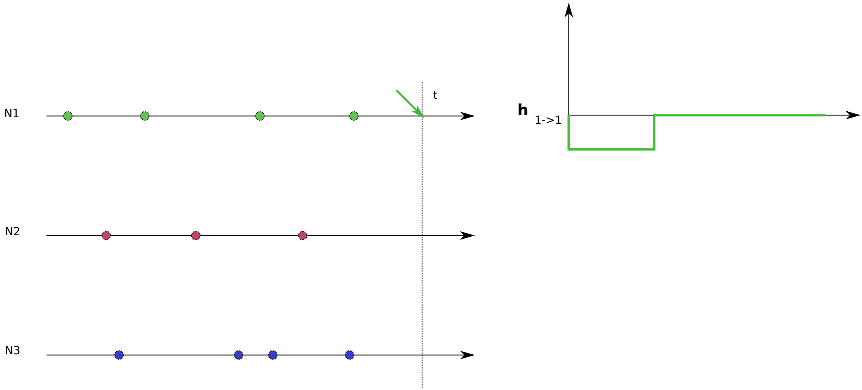


Multivariate Hawkes processes



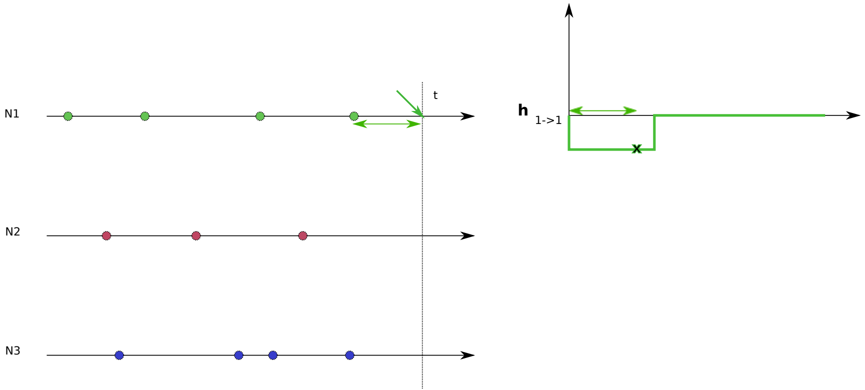
$$\lambda^{(1)}(t) = \underbrace{\nu}_{\text{spontaneous}}$$

Multivariate Hawkes processes



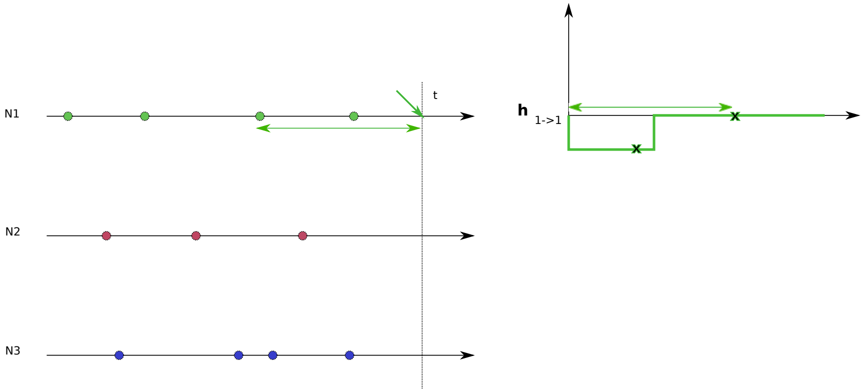
$$\lambda^{(1)}(t) = \underbrace{\nu}_{\text{spontaneous}} + \underbrace{\sum_{\substack{T<t \\ \text{pt de N1}}} h_{1\to 1}(t-T)}_{\text{self-interaction}}$$

Multivariate Hawkes processes



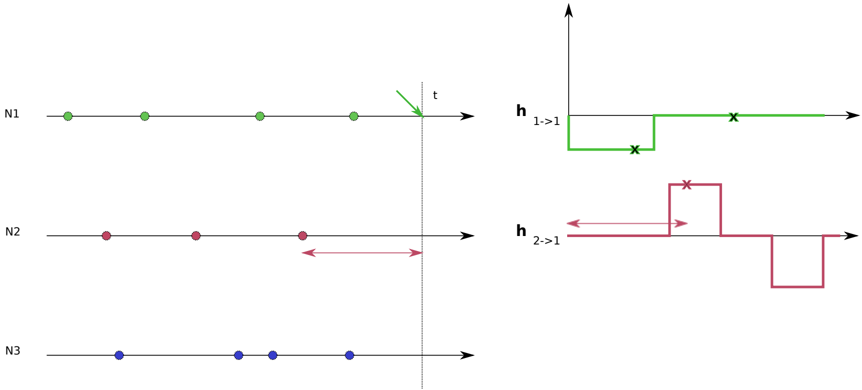
$$\lambda^{(1)}(t) = \underbrace{\nu}_{\text{spontaneous}} + \underbrace{\sum_{\substack{T<t \\ \text{pt de N1}}} h_{1\to 1}(t-T)}_{\text{self-interaction}}$$

Multivariate Hawkes processes



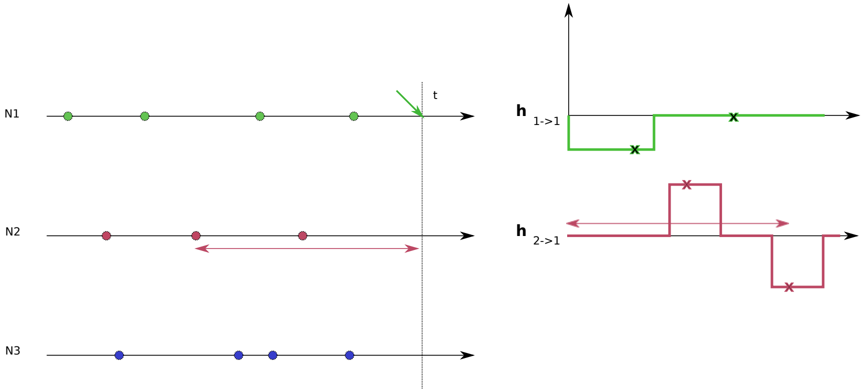
$$\lambda^{(1)}(t) \quad = \quad \underbrace{\nu}_{\text{spontaneous}} \quad + \quad \underbrace{\sum_{\substack{T<t \\ \text{pt de N1}}} \quad h_{1\to 1}(t-T)}_{\text{self-interaction}}$$

Multivariate Hawkes processes



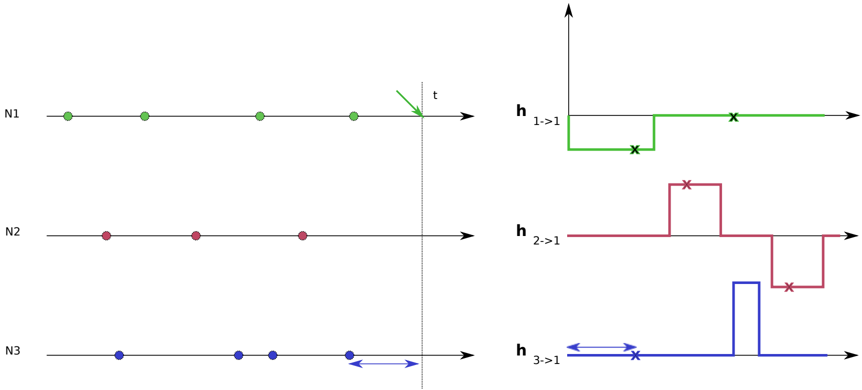
$$\lambda^{(1)}(t) = \underbrace{\nu}_{\text{spontaneous}} + \underbrace{\sum_{\substack{T<t \\ \text{pt de N1}}} h_{1\to1}(t-T)}_{\text{self-interaction}} + \underbrace{\sum_{\substack{m\neq 1, \\ T<t, \\ \text{pt de Nm}}} h_{m\to1}(t-T)}_{\text{interaction}}$$

Multivariate Hawkes processes



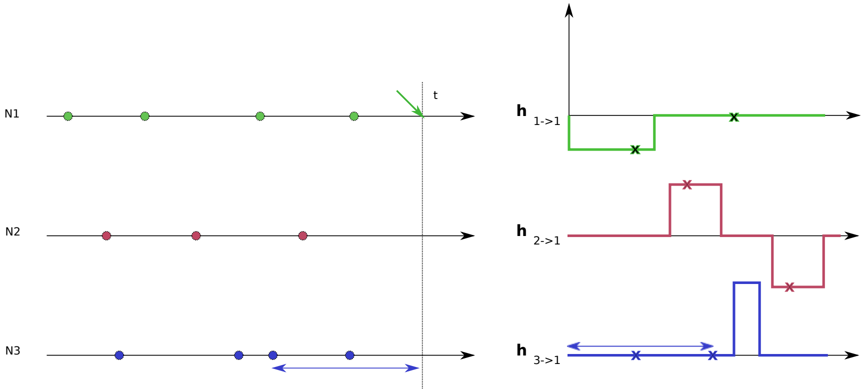
$$\lambda^{(1)}(t) = \underbrace{\nu}_{\text{spontaneous}} + \underbrace{\sum_{\substack{T<t \\ \text{pt de N1}}} h_{1\to 1}(t-T)}_{\text{self-interaction}} + \underbrace{\sum_{\substack{m\neq 1, \\ T<t, \\ \text{pt de Nm}}} h_{m\to 1}(t-T)}_{\text{interaction}}$$

Multivariate Hawkes processes



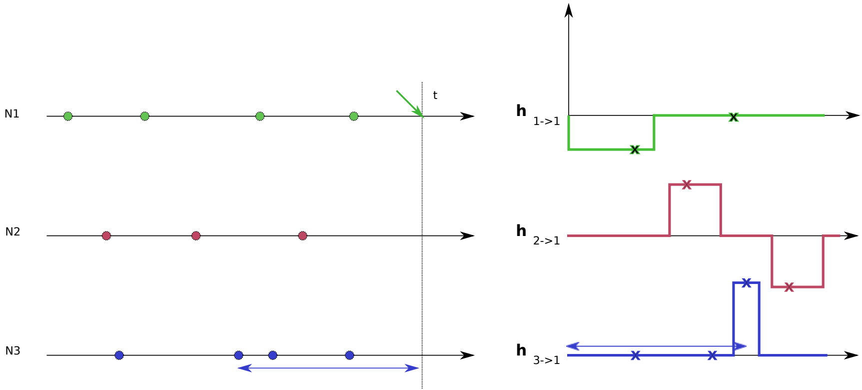
$$\lambda^{(1)}(t) = \underbrace{\nu}_{\text{spontaneous}} + \underbrace{\sum_{\substack{T<t \\ \text{pt de N1}}} h_{1\to 1}(t-T)}_{\text{self-interaction}} + \underbrace{\sum_{\substack{m\neq 1, \\ T<t, \\ \text{pt de Nm}}} h_{m\to 1}(t-T)}_{\text{interaction}}$$

Multivariate Hawkes processes



$$\lambda^{(1)}(t) = \underbrace{\nu}_{\text{spontaneous}} + \underbrace{\sum_{\substack{T<t \\ \text{pt de N1}}} h_{1\to 1}(t-T)}_{\text{self-interaction}} + \underbrace{\sum_{\substack{m\neq 1, \\ T<t, \\ \text{pt de Nm}}} h_{m\to 1}(t-T)}_{\text{interaction}}$$

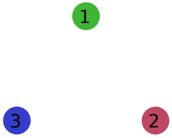
Multivariate Hawkes processes

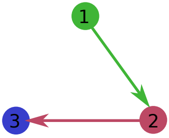


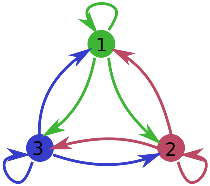
$$\lambda^{(1)}(t) = \underbrace{\nu}_{\text{spontaneous}} + \underbrace{\sum_{\substack{T<t \\ \text{pt de N1}}} h_{1\to1}(t-T)}_{\text{self-interaction}} + \underbrace{\sum_{\substack{m\neq 1, \\ T<t, \\ \text{pt de Nm}}} h_{m\to1}(t-T)}_{\text{interaction}}$$

Link Graphical model of local independence

(Didelez (2008))

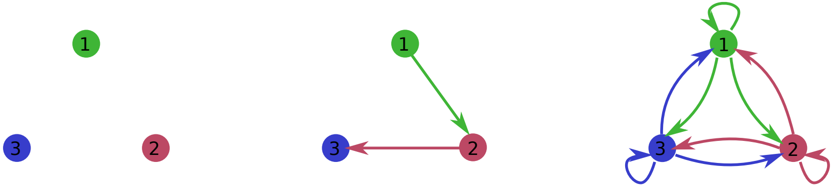






Link Graphical model of local independence

(Didelez (2008))



→ (?) functional connectivity graphs in Neurosciences, but also useful in sismology, genomics, marketing, finance, epidemiology ....

More formally

- Only excitation (all the $h_\ell^{(r)}$ are positive): for all r,

$$\lambda^{(r)}(t) = \nu_r + \sum_{\ell=1}^{M} \int_{-\infty}^{t-} h_\ell^{(r)}(t-u)dN_u^{(\ell)}.$$

Branching / Cluster representation, stationary process if the spectral radius of $\left(\int h_\ell^{(r)}(t)dt\right)$ is < 1.

More formally

- Only excitation (all the $h_\ell^{(r)}$ are positive): for all r,

$$\lambda^{(r)}(t) = \nu_r + \sum_{\ell=1}^{M} \int_{-\infty}^{t-} h_\ell^{(r)}(t-u)dN_u^{(\ell)}.$$

  Branching / Cluster representation, stationary process if the spectral radius of $\left(\int h_\ell^{(r)}(t)dt\right)$ is < 1.

- Interaction, for instance, (in general any 1-Lipschitz function, Brémaud Massoulié 1996)

$$\lambda^{(r)}(t) = \left(\nu_r + \sum_{\ell=1}^{M} \int_{-\infty}^{t-} h_\ell^{(r)}(t-u)dN_u^{(\ell)}\right)_+.$$

More formally

- Only excitation (all the $h_\ell^{(r)}$ are positive): for all r,

$$\lambda^{(r)}(t) = \nu_r + \sum_{\ell=1}^{M} \int_{-\infty}^{t-} h_\ell^{(r)}(t-u)dN_u^{(\ell)}.$$

Branching / Cluster representation, stationary process if the spectral radius of $\left(\int h_\ell^{(r)}(t)dt\right)$ is < 1.

- Interaction, for instance, (in general any 1-Lipschitz function, Brémaud Massoulié 1996)

$$\lambda^{(r)}(t) = \left(\nu_r + \sum_{\ell=1}^{M} \int_{-\infty}^{t-} h_\ell^{(r)}(t-u)dN_u^{(\ell)}\right)_+.$$

- Exponential (Multiplicative shape but no guarantee of a stationary version ...)

$$\lambda^{(r)}(t) = \exp\left(\nu_r + \sum_{\ell=1}^{M} \int_{-\infty}^{t-} h_\ell^{(r)}(t-u)dN_u^{(\ell)}\right).$$

Previous works

- Maximum likelihood estimates eventually + AIC (Ogata, Vere-Jones etc mainly for sismology, Chornoboy et al., for neuroscience, Gusto and Schbath for genomics)
- Parametric tests for the detection of edge + Maximum likelihood + exponential formula + spline estimation (Carstensen et al., in genomics)
- Univariate processes + ℓ_0 penalty, oracle inequalities (RB and Schbath)
- Maximum likelihood + exponential formula + ℓ_1 "group Lasso" penalty (Pillow et al. in neuroscience)
- Thresholding + tests for very particular bivariate models, oracle inequality (Sansonnet)

Another parametric method : Least-squares

- A good estimate of the parameters should correspond to an intensity candidate close to the true one.

Another parametric method : Least-squares

- A good estimate of the parameters should correspond to an intensity candidate close to the true one.
- Hence one would like to minimize $\|\eta - \lambda\|^2 = \int_0^T [\eta(t) - \lambda(t)]^2 dt$ in η

Another parametric method : Least-squares

- A good estimate of the parameters should correspond to an intensity candidate close to the true one.
- Hence one would like to minimize $\|\eta - \lambda\|^2 = \int_0^T [\eta(t) - \lambda(t)]^2 dt$ in η
- or equivalently $-2\int_0^T \eta(t)\lambda(t)dt + \int_0^T \eta(t)^2 dt$.

Another parametric method : Least-squares

- A good estimate of the parameters should correspond to an intensity candidate close to the true one.
- Hence one would like to minimize $\|\eta - \lambda\|^2 = \int_0^T [\eta(t) - \lambda(t)]^2 dt$ in η
- or equivalently $-2\int_0^T \eta(t)\lambda(t)dt + \int_0^T \eta(t)^2 dt$.
- But dN_t randomly fluctuates around $\lambda(t)dt$ and is observable.

Another parametric method : Least-squares

- A good estimate of the parameters should correspond to an intensity candidate close to the true one.
- Hence one would like to minimize $\|\eta - \lambda\|^2 = \int_0^T [\eta(t) - \lambda(t)]^2 dt$ in η
- or equivalently $-2\int_0^T \eta(t)\lambda(t)dt + \int_0^T \eta(t)^2 dt$.
- But dN_t randomly fluctuates around $\lambda(t)dt$ and is observable.
- Hence one minimizes

$$-2\int_0^T \eta(t)dN_t + \int_0^T \eta(t)^2 dt,$$

for a model $\eta = \lambda_{\mathbf{a}}(t)$

On Hawkes processes

Recall that for the process $N^{(r)}$

$$\lambda^{(r)}(t) = \nu^{(r)} + \sum_{\ell=1}^{M} \sum_{T<t, T \text{ in } N^{(\ell)}} h_\ell^{(r)}(t - T).$$

On Hawkes processes

Recall that for the process $N^{(r)}$

$$\lambda^{(r)}(t) = \nu^{(r)} + \sum_{\ell=1}^{M} \sum_{T<t, T \ in \ N^{(\ell)}} h_\ell^{(r)}(t-T).$$

$+$ Piecewise constant model with parameter $\mathbf{a}$

On Hawkes processes

Recall that for the process $N^{(r)}$

$$\lambda^{(r)}(t) = \nu^{(r)} + \sum_{\ell=1}^{M} \sum_{T<t, T \text{ in } N^{(\ell)}} h_\ell^{(r)}(t - T).$$

+ Piecewise constant model with parameter **a**

By linearity,

$$\lambda^{(r)}(t) = (\mathbf{R}\mathbf{c}_t)'\mathbf{a}_*^{(r)},$$

On Hawkes processes

Recall that for the process $N^{(r)}$

$$\lambda^{(r)}(t) = \nu^{(r)} + \sum_{\ell=1}^{M} \sum_{T<t, T \text{ in } N^{(\ell)}} h_\ell^{(r)}(t-T).$$

+ Piecewise constant model with parameter **a**

By linearity,

$$\lambda^{(r)}(t) = (\mathbf{Rc}_t)' \mathbf{a}_*^{(r)},$$

with $\mathbf{Rc}_t$ being the renormalized instantaneous count given by

$$(\mathbf{Rc}_t)' = \left(1, \delta^{-1/2}(\mathbf{c}_t^{(1)})', \ldots, \delta^{-1/2}(\mathbf{c}_t^{(M)})'\right),$$

On Hawkes processes

Recall that for the process $N^{(r)}$

$$\lambda^{(r)}(t) = \nu^{(r)} + \sum_{\ell=1}^{M} \sum_{T<t, T \text{ in } N^{(\ell)}} h_\ell^{(r)}(t-T).$$

+ Piecewise constant model with parameter **a**

By linearity,

$$\lambda^{(r)}(t) = (\mathbf{Rc}_t)'\mathbf{a}_*^{(r)},$$

with $\mathbf{Rc}_t$ being the renormalized instantaneous count given by

$$(\mathbf{Rc}_t)' = \left(1, \delta^{-1/2}(\mathbf{c}_t^{(1)})', \dots, \delta^{-1/2}(\mathbf{c}_t^{(M)})'\right),$$

and with $\mathbf{c}_t^{(\ell)}$ being the vector of instantaneous count with delay of N_ℓ i.e.

$$(\mathbf{c}_t^{(\ell)})' = \left(N_{[t-\delta,t)}^{(\ell)}, \dots, N_{[t-K\delta,t-(K-1)\delta)}^{(\ell)}\right).$$

An heuristic for the least-square estimator

Informally, the link between the point process and its intensity can be written as

$$dN^{(r)}(t) \simeq (\mathbf{R}\mathbf{c}_t)'\mathbf{a}_*^{(r)} dt + \text{ noise.}$$

An heuristic for the least-square estimator

Informally, the link between the point process and its intensity can be written as

$$dN^{(r)}(t) \simeq (\mathbf{Rc}_t)'\mathbf{a}_*^{(r)}dt + \text{ noise.}$$

Let

$$\mathbf{G} = \int_0^T \mathbf{Rc}_t(\mathbf{Rc}_t)'dt,$$

the integrated covariation of the renormalized instantaneous count.

An heuristic for the least-square estimator

Informally, the link between the point process and its intensity can be written as

$$dN^{(r)}(t) \simeq (\mathbf{R}\mathbf{c}_t)'\mathbf{a}_*^{(r)}dt + \text{ noise.}$$

Let

$$\mathbf{G} = \int_0^T \mathbf{R}\mathbf{c}_t(\mathbf{R}\mathbf{c}_t)'dt,$$

the integrated covariation of the renormalized instantaneous count.

$$\mathbf{b}^{(r)} := \int_0^T \mathbf{R}\mathbf{c}_t dN^{(r)}(t) \simeq \mathbf{G}\mathbf{a}_*^{(r)} + \text{ noise.}$$

An heuristic for the least-square estimator

Informally, the link between the point process and its intensity can be written as

$$dN^{(r)}(t) \simeq (\mathbf{R}\mathbf{c}_t)'\mathbf{a}_*^{(r)}dt + \text{ noise.}$$

Let

$$\mathbf{G} = \int_0^T \mathbf{R}\mathbf{c}_t(\mathbf{R}\mathbf{c}_t)'dt,$$

the integrated covariation of the renormalized instantaneous count.

$$\mathbf{b}^{(r)} := \int_0^T \mathbf{R}\mathbf{c}_t dN^{(r)}(t) \simeq \mathbf{G}\mathbf{a}_*^{(r)} + \text{ noise.}$$

The least-square estimate is

$$\hat{\mathbf{a}}^{(r)} = \mathbf{G}^{-1}\mathbf{b}^{(r)},$$

where in b lies again the number of couples with a certain delay.

An heuristic for the least-square estimator

Informally, the link between the point process and its intensity can be written as

$$dN^{(r)}(t) \simeq (\mathbf{R}\mathbf{c}_t)'\mathbf{a}_*^{(r)}dt + \text{ noise.}$$

Let

$$\mathbf{G} = \int_0^T \mathbf{R}\mathbf{c}_t(\mathbf{R}\mathbf{c}_t)'dt,$$

the integrated covariation of the renormalized instantaneous count.

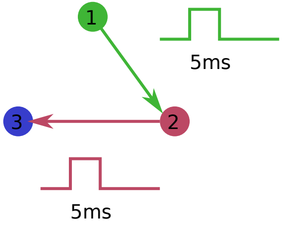
$$\mathbf{b}^{(r)} := \int_0^T \mathbf{R}\mathbf{c}_t dN^{(r)}(t) \simeq \mathbf{G}\mathbf{a}_*^{(r)} + \text{ noise.}$$

The least-square estimate is

$$\hat{\mathbf{a}}^{(r)} = \mathbf{G}^{-1}\mathbf{b}^{(r)},$$

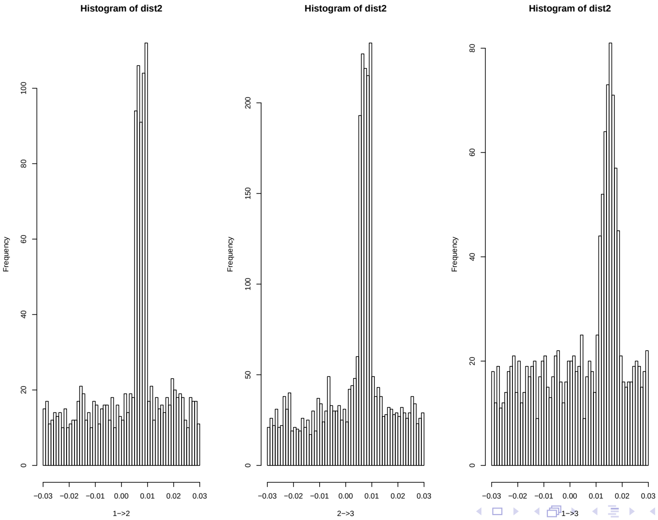
where in b lies again the number of couples with a certain delay.
→ simpler formula than Maximum Likelihood Estimators for similar properties

What gain wrt cross correlogramm ?

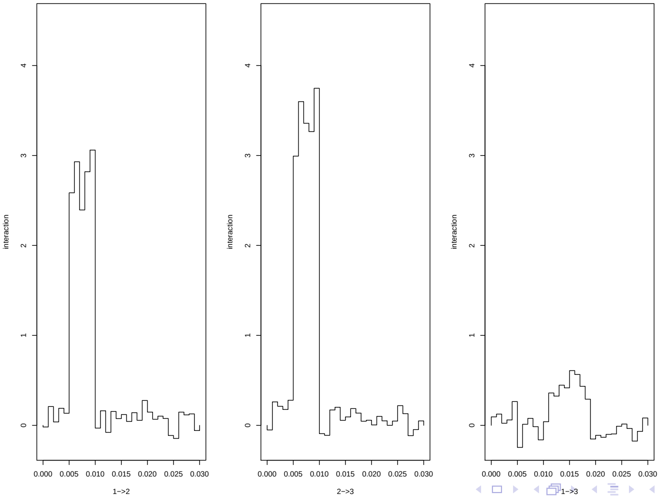


only 2 non zero interaction functions over 9

What gain wrt cross correlogramm ?



What gain wrt cross correlogramm ?



Full Multivariate Hawkes process and ℓ_1 penalty

with N.R. Hansen (Copenhagen), and V. Rivoirard (Dauphine) (2012)

The Lasso criterion can be expressed independently for each sub-process by :

Lasso criterion

$$\hat{\mathbf{a}}^{(r)} = \operatorname{argmin}_{\mathbf{a}}\{\gamma_T(\lambda_{\mathbf{a}}^{(r)}) + \operatorname{pen}(\mathbf{a})\}$$
$$= \operatorname{argmin}_{\mathbf{a}}\{-2\mathbf{a}'\mathbf{b}_r + \mathbf{a}'\mathbf{G}\mathbf{a} + 2(\mathbf{d}^{(r)})'|\mathbf{a}|\}$$

- The crucial choice is the $\mathbf{d}^{(r)}$, should be data-driven !
- The theoretical validation : be able to state that our choice is the best possible choice.
- The practical validation : on simulated Hawkes processes (done), on simulated neuronal networks, on real data (RNRP Paris 6 work in progress)...

Theoretical Validation = Oracle inequality

Recall that

$$\mathbf{b}^{(r)} = \int_0^T \mathbf{Rc}_t dN^{(r)}(t)$$

and

$$\mathbf{G} = \int_0^T \mathbf{Rc}_t (\mathbf{Rc}_t)' dt.$$

Hansen,Rivoirard,RB

If $\mathbf{G} \geq cI$ with $c > 0$ and if

$$\left| \int_0^T \mathbf{Rc}_t \left(dN^{(r)}(t) - \lambda^{(r)}(t)dt \right) \right| \leq \mathbf{d}^{(r)}, \quad \forall r$$

then

$$\sum_r \|\lambda^{(r)} - \mathbf{Rc}_t \hat{\mathbf{a}}^{(r)}\|^2 \leq \square \inf_{\mathbf{a}} \left\{ \sum_r \|\lambda^{(r)} - \mathbf{Rc}_t \mathbf{a}\|^2 + \frac{1}{c} \sum_{i \in supp(\mathbf{a})} (d_i^{(r)})^2 \right\}.$$

Explanations

- Main Point: **d** controls the random fluctuations / noise $\rightarrow$ should be data-driven and sharp !

Explanations

- Main Point: $\mathbf{d}$ controls the random fluctuations / noise $\rightarrow$ should be data-driven and sharp !
- The loss $(1/T)\sum_{i\in supp(\mathbf{a})}(d_i^{(r)})^2$ is unavoidable even for only one choice of set of non-zeros coefficients. Should be read as "capacity of approximation" + unavoidable loss due to the noise.

Explanations

- Main Point: **d** controls the random fluctuations / noise $\rightarrow$ should be data-driven and sharp !
- The loss $(1/T)\sum_{i\in supp(\mathbf{a})}(d_i^{(r)})^2$ is unavoidable even for only one choice of set of non-zeros coefficients. Should be read as "capacity of approximation" + unavoidable loss due to the noise.
- Factor $1/c$ gives a "quality indicator" since G observable in practice. Maybe not the best thing, but with stationnary Hawkes process, one can prove that $c \simeq T$.

Explanations

- Main Point: **d** controls the random fluctuations / noise → should be data-driven and sharp !
- The loss $(1/T)\sum_{i \in supp(\mathbf{a})}(d_i^{(r)})^2$ is unavoidable even for only one choice of set of non-zeros coefficients. Should be read as "capacity of approximation" + unavoidable loss due to the noise.
- Factor $1/c$ gives a "quality indicator" since G observable in practice. Maybe not the best thing, but with stationnary Hawkes process, one can prove that $c \simeq T$.
- Gives the "best" linear approximation and in this sense valid even if λ not of the prescribed shape, in particular in case of inhibition.

Explanations

- Main Point: **d** controls the random fluctuations / noise $\rightarrow$ should be data-driven and sharp !
- The loss $(1/T)\sum_{i\in supp(\mathbf{a})}(d_i^{(r)})^2$ is unavoidable even for only one choice of set of non-zeros coefficients. Should be read as "capacity of approximation" + unavoidable loss due to the noise.
- Factor $1/c$ gives a "quality indicator" since G observable in practice. Maybe not the best thing, but with stationnary Hawkes process, one can prove that $c \simeq T$.
- Gives the "best" linear approximation and in this sense valid even if λ not of the prescribed shape, in particular in case of inhibition.
- In practice small unavoidable bias, possible to correct by two step procedures (OLS on the support of the Lasso estimate).

Explanations

- Main Point: **d** controls the random fluctuations / noise → should be data-driven and sharp !
- The loss $(1/T)\sum_{i\in supp(\mathbf{a})}(d_i^{(r)})^2$ is unavoidable even for only one choice of set of non-zeros coefficients. Should be read as "capacity of approximation" + unavoidable loss due to the noise.
- Factor $1/c$ gives a "quality indicator" since G observable in practice. Maybe not the best thing, but with stationnary Hawkes process, one can prove that $c \simeq T$.
- Gives the "best" linear approximation and in this sense valid even if λ not of the prescribed shape, in particular in case of inhibition.
- In practice small unavoidable bias, possible to correct by two step procedures (OLS on the support of the Lasso estimate).
- Possible to do it with other basis (Fourier) or other linear model (Aalen)

One of the main probabilistic ingredients

Bernstein type inequality for counting processes (H., R.B., R.)

Let $(H_s)_{s\geq 0}$ be a predictable process and
$M_t = \int_0^t H_s(dN_s - \lambda(s)ds)$. Let $b > 0$ and $v > w > 0$.
For all $x, \mu > 0$ such that $\mu > \phi(\mu)$, let
$\hat{V}_\tau^\mu = \frac{\mu}{\mu - \phi(\mu)} \int_0^\tau H_s^2 dN_s + \frac{b^2 x}{\mu - \phi(\mu)}$, where $\phi(u) = \exp(u) - u - 1$.
Then for every stopping time τ and every $\varepsilon > 0$

$$\mathbb{P}\left(M_\tau \geq \sqrt{2(1+\varepsilon)\hat{V}_\tau^\mu x} + bx/3, \quad w \leq \hat{V}_\tau^\mu \leq v \text{ and } \sup_{s\in[0,\tau]} |H_s| \leq b \right)$$
$$\leq 2\frac{\log(v/w)}{\log(1+\varepsilon)} e^{-x}.$$

One of the main probabilistic ingredients

Bernstein type inequality for counting processes (H., R.B., R.)

Let $(H_s)_{s\geq 0}$ be a predictable process and
$M_t = \int_0^t H_s(dN_s - \lambda(s)ds)$. Let $b > 0$ and $v > w > 0$.
For all $x, \mu > 0$ such that $\mu > \phi(\mu)$, let
$\hat{V}^\mu_\tau = \frac{\mu}{\mu-\phi(\mu)} \int_0^\tau H_s^2 dN_s + \frac{b^2 x}{\mu-\phi(\mu)}$, where $\phi(u) = \exp(u) - u - 1$.
Then for every stopping time τ and every $\varepsilon > 0$

$$\mathbb{P}\left(M_\tau \geq \sqrt{2(1+\varepsilon)\hat{V}^\mu_\tau x} + bx/3, \quad w \leq \hat{V}^\mu_\tau \leq v \text{ and } \sup_{s\in[0,\tau]} |H_s| \leq b \right) \leq 2\frac{\log(v/w)}{\log(1+\varepsilon)} e^{-x}.$$

NB : $\hat{V}^\mu_\tau$ is an estimate of the bracket of the martingale, i.e. the martingale equivalent of the variance → Gaussian behavior for the main term.

One of the main probabilistic ingredients

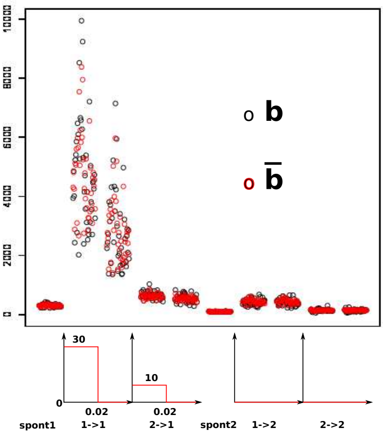
Bernstein type inequality for counting processes (H., R.B., R.)

Let $(H_s)_{s\geq 0}$ be a predictable process and $M_t = \int_0^t H_s(dN_s - \lambda(s)ds)$. Let $b > 0$ and $v > w > 0$.
For all $x, \mu > 0$ such that $\mu > \phi(\mu)$, let
$\hat{V}_\tau^\mu = \frac{\mu}{\mu-\phi(\mu)} \int_0^\tau H_s^2 dN_s + \frac{b^2 x}{\mu-\phi(\mu)}$, where $\phi(u) = \exp(u) - u - 1$.
Then for every stopping time τ and every $\varepsilon > 0$

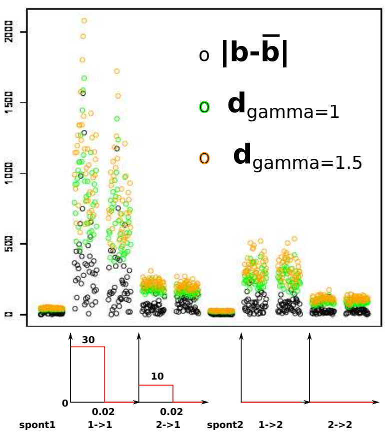
$$\mathbb{P}\left(M_\tau \geq \sqrt{2(1+\varepsilon)\hat{V}_\tau^\mu x} + bx/3, \quad w \leq \hat{V}_\tau^\mu \leq v \text{ and } \sup_{s\in[0,\tau]} |H_s| \leq b \right)$$
$$\leq 2\frac{\log(v/w)}{\log(1+\varepsilon)} e^{-x}.$$

NB : $\hat{V}_\tau^\mu$ is an estimate of the bracket of the martingale, i.e. the martingale equivalent of the variance $\rightarrow$ Gaussian behavior for the main term.
Applied to $\int_0^T \mathbf{Rc}_t \left(dN^{(r)}(t) - \lambda^{(r)}(t)dt\right)$: $\mathbf{d}$ is given by the right hand-side ($x \simeq \gamma \ln(T)$) $\rightarrow$ Bernstein Lasso.

Choices of the weights d_λ



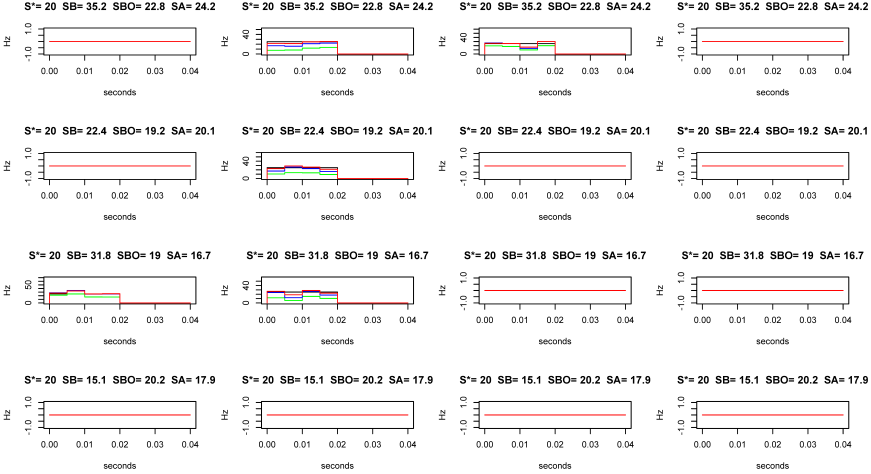
Coeff: 10 4.2 0 1.4 0 10 0 0 0 0



Coeff: 10 4.2 0 1.4 0 10 0 0 0 0

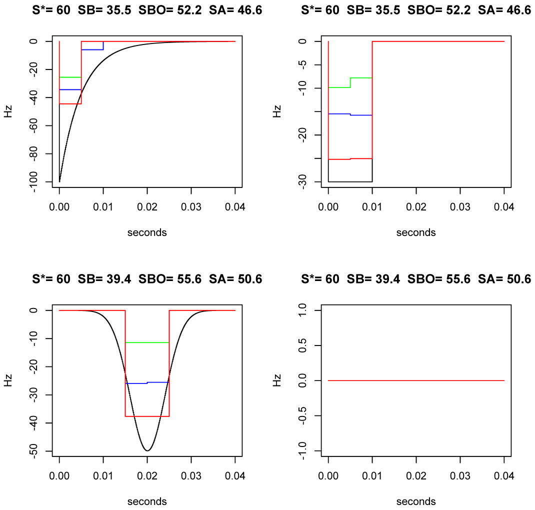
NB : Adaptive Lasso (Zou) $d_\lambda = \gamma / |\hat{a}_\lambda|$

Simulation study - Estimation ($T = 20$, $M = 8$, $K = 8$)



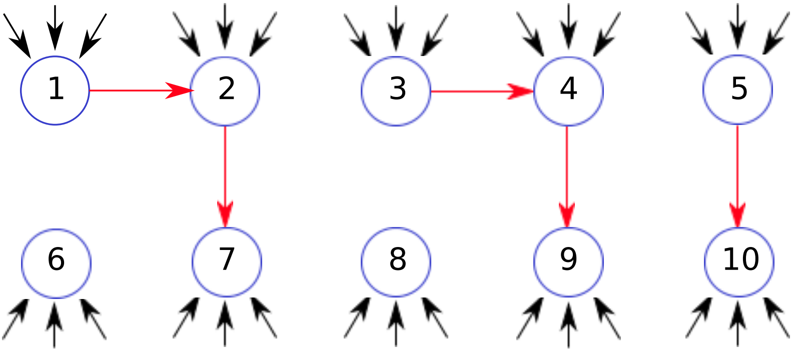
Interactions reconstructed with 'Adaptive Lasso', 'Bernstein Lasso' and 'Bernstein Lasso+OLS'. Above graphs, estimation of spontaneous rates

Another example with inhibition

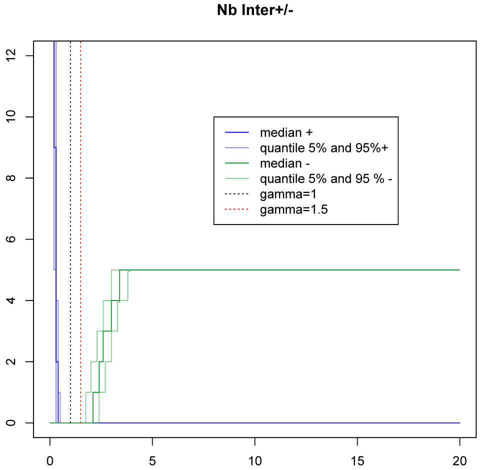


$T = 20s$

Another (more realistic?) neuronal network: Integrate and Fire

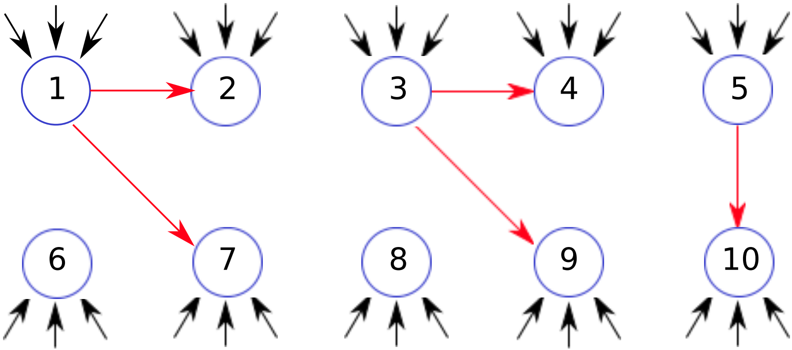


Another (more realistic?) neuronal network: Integrate and Fire

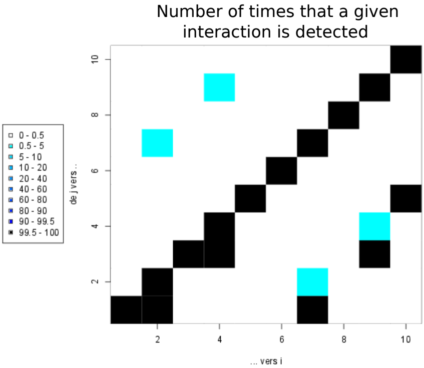


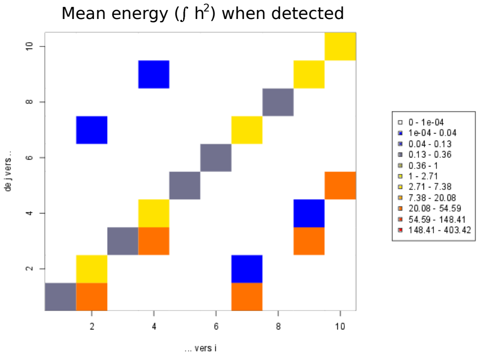
$T = 60s$

Another (more realistic?) neuronal network: Integrate and Fire

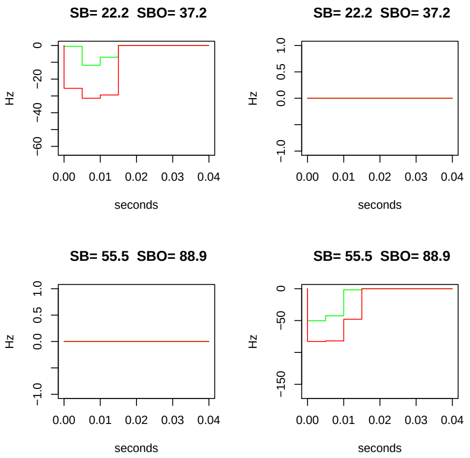


Another (more realistic?) neuronal network: Integrate and Fire





On neuronal data (sensorimotor task)

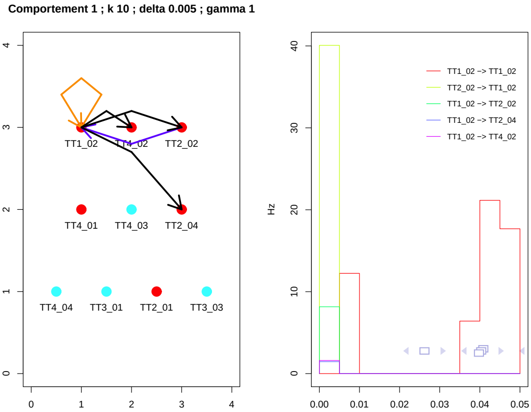


30 trials : monkey trained to touch the correct target when illuminated. Accept the test of Hawkes hypothesis. Work with F. Grammont, V. Rivoirard and C. Tuleau-Malot

On neuronal data (vibrissa excitation)

Joint work with RNRP (Paris 6). Behavior: vibrissa excitation at low frequency. $T = 90.5$, $M = 10$

Neuron	$TT1_{02}$	$TT2_{01}$	$TT2_{02}$	$TT2_{04}$	$TT3_{01}$	$TT3_{03}$	$TT4_{01}$	$TT4_{02}$	$TT4_{03}$	$TT4_{04}$
Spikes	9191	99	544	149	15	18	136	282	8	6



Application on real data

Data:

Neuron	$TT1_{02}$	$TT2_{01}$	$TT2_{02}$	$TT2_{04}$	$TT3_{01}$	$TT3_{03}$	$TT4_{01}$	$TT4_{02}$	$TT4_{03}$	$TT4_{04}$
Spikes	9191	99	544	149	15	18	136	282	8	6

Simulation:

Neuron	$TT1_{02}$	$TT2_{01}$	$TT2_{02}$	$TT2_{04}$	$TT3_{01}$	$TT3_{03}$	$TT4_{01}$	$TT4_{02}$	$TT4_{03}$	$TT4_{04}$
Spikes	9327	92	585	148	13	23	133	271	8	3

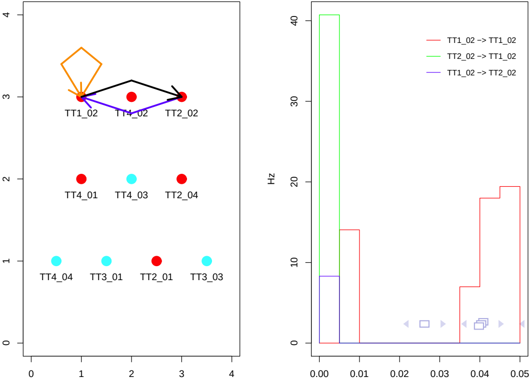
Application on real data

Data:

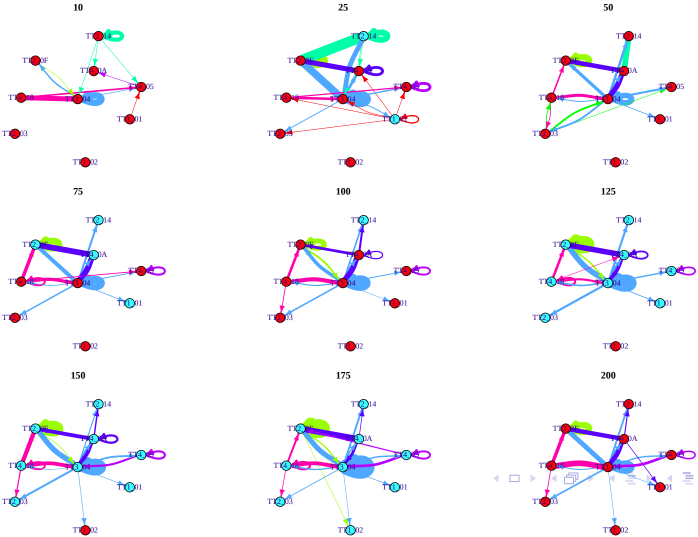
Neuron	$TT1_{02}$	$TT2_{01}$	$TT2_{02}$	$TT2_{04}$	$TT3_{01}$	$TT3_{03}$	$TT4_{01}$	$TT4_{02}$	$TT4_{03}$	$TT4_{04}$
Spikes	9191	99	544	149	15	18	136	282	8	6

Simulation:

Neuron	$TT1_{02}$	$TT2_{01}$	$TT2_{02}$	$TT2_{04}$	$TT3_{01}$	$TT3_{03}$	$TT4_{01}$	$TT4_{02}$	$TT4_{03}$	$TT4_{04}$
Spikes	9327	92	585	148	13	23	133	271	8	3



Evolution of the dependance graph as a fonction of the vibrissa excitation



Conclusions and Perspectives

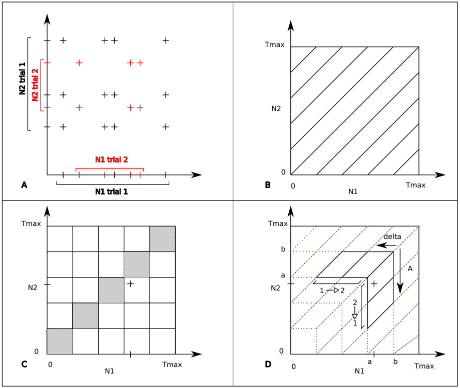
- It is possible to estimate interaction functions and even a graph of possible interactions, to have some mathematical guarantee on it and this even if the Hawkes linear model is not completely true (robustness).
- It is even possible to (non parametrically) test that such interactions exists. However the power of such tests (not studied yet) should be linked to a "distance" to the Hawkes model.
- One of the main issue is the lack of stationarity $\rightarrow$ a work in progress with F. Picard (Lyon) and C. Tuleau-Malot (Nice) based on segmentation and clustering.

Thank you !

References

- Hansen, N.R., Reynaud-Bouret, P. and Rivoirard, V. *Lasso and probabilistic inequalities for multivariate point processes.* to appear in Bernoulli (2012).
- Tuleau-Malot, C., Rouis, A., Grammont, F., and Reynaud-Bouret, P. *Multiple Tests based on a Gaussian Approximation of the Unitary Events method.* in revision (2013)
- Reynaud-Bouret, P., Tuleau-Malot, C., Rivoirard, V. and Grammont, F. *Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis* to appear in Journal of Mathematical Neuroscience (2013)

Link with cross-correlogram and Joint PeriStimulus Histogram



see also Zhang et al (2007) in genomics and Aertsen et al (1989) in Neurosciences