Bayesian estimation of the intenisity function for Hawkes processes

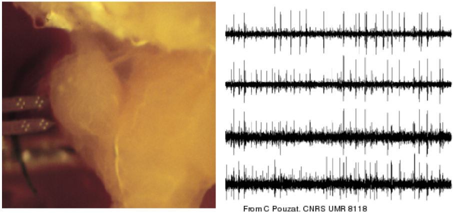
Luminy

S. Donnet, V. Rivoirard, J. Rousseau

July 2017

Spikes and point processes

- neuronal activity : electric impulses = spikes



From C Pouzat. CNRS UMR 8118

- The interesting information from these spikes is related to the time of their appearance and their length(rather than their shape or intensity)

$\Rightarrow$ Point processes On $\mathbb{R}^+$.

Poisson processes and Hawkes processes

On $(\Omega, \mathcal{F}, \mathbb{P})$, **N** is a point process

- Its distribution is caracterized by its conditional intensity process

$$
\begin{aligned}
\lambda(t) &= \lambda(t|\mathcal{F}_t) = \lim_{h \to 0} \frac{1}{h}\mathbb{P}(\text{ an event takes place in } [t, t+h]|\mathcal{F}_t) \\
&= \text{"}\mathbb{E}[dNt|\mathcal{F}_t]\text{"} \quad (\mathcal{F}_t) = \text{ adapted } \quad \text{filtration}
\end{aligned}
$$

- Poisson Process
  - $\lambda(t)$ deterministic
  - intervals between events are independent
- Hawkes Process
  - $\lambda(t) = \left\{\nu + \sum_{T_i < t} h(t - T_i)\right\}_+$
  - When h is positive : self-excitation, when h is negative inhibition

Multidimensional Hawkes processes

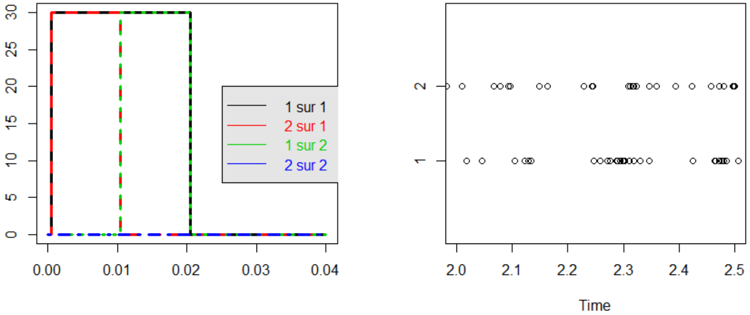
- M neurones interracting : self or inter excitation or inhibition. We observe M point processes non independent.
- Conditional intensity of the process m :

$$\lambda^{(m)}(t) = \left\{ \nu^{(m)} + \sum_{\ell=1}^{M} \sum_{t_i^{(\ell)} < t} h_\ell^{(m)}(t - t_i^{(\ell)}) \right\}_+$$

- Remarks
  - $h_\ell^{(m)}$ is the excitation function of m by ℓ.
  - $h_\ell^{(m)}$ can be negative.
  - Stationary process if the spectral radius of matrix $\boldsymbol{I} = \left(\int_0^\infty |h_\ell^{(m)}(u)| du \right)_{1 \leq \ell, m \leq M}$ is < 1.
  - The functions $h_\ell^{(m)}$ have support : $[0, s_{max}]$

[Reynaud-Bouret et al., 2013]

Example of bi-dimensionnal Hawkes processes



Objectives

- Bayesian estimation $\theta = \left\{ \{\nu^{(m)}\}_{m=1\dots M}, \{h_\ell^{(m)}\}_{(\ell,m)\in\{1,\dots,M\}^2} \right\}$.
- exists : Lasso estimates [Hansen et al., 2014] .
- Hardly anything in the Bayesian framework [Rasmussen, 2013], K. Heller

Bayesian inference

- $\theta = \left\{ \{\nu^{(m)}\}_{m=1\ldots M}, \{h_\ell^{(m)}\}_{(\ell,m)\in\{1,\ldots,M\}^2} \right\}$
- Observations $\boldsymbol{N}$ on M neurones during $[0, T]$
- **Likelihood** : $\mathbf{n}_m$ = number of jumps for neuron m, and $T_1^{(m)}, \ldots, T_{n_m}^{(m)}$ jump times for neurone m, $m \leq M$

$$L(\boldsymbol{N};\theta) = \exp\left[\sum_{m=1}^{M} \sum_{j=1}^{\mathbf{n}_m} \log \lambda^{(m)}(T_j^{(m)}) - \int_0^T \lambda^{(m)}(u)du \right]$$

- Posterior (pseudo) :

$$\pi(d\theta|\boldsymbol{N}) = \frac{\pi(d\theta)L(\boldsymbol{N};\theta)}{p(\boldsymbol{N})}$$

- Depends on $\mathbf{N}_{[-s_{max}:T]}$

Posterior concentration on θ : $T \to +\infty$

- Posterior concentration If θ^* = true parameter, $\epsilon_T = o(1)$

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

General Theorem on posterior concentration rates G& VdV (07)

- **Kullback- Leibler Condition**

$$\Pi\left(\{K(p_{\theta^*,T}, p_{\theta,T}) \leq T\epsilon_T^2; V(p_{\theta^*,T}, p_{\theta,T}) \lesssim T\epsilon_T^2\}\right) > e^{-cT\epsilon_T^2}$$

- **Tests** $\exists \Theta_T \ \pi(\Theta_T^c) \leq e^{-(c+2)T\epsilon_T^2}$ s.t.

$$\exists \phi_T; \mathbb{E}_{\theta^*}(\phi_T) = o(1); \quad \sup_{\theta \in \Theta_T^c; d(\theta,\theta^*) > B\epsilon_T} \mathbb{E}_\theta (1 - \phi_T) \leq e^{-(c+2)T\epsilon_T^2}$$

Then

$$\Pi\left(d(\theta, \theta^*) > M\epsilon_T | \boldsymbol{N}\right)$$

Application to Hawkes processes

- Parameter space :

$$\Theta = \{(\nu_\ell, h_{k,\ell}, k, \ell \leq M);\ h_{k,\ell} \geq 0,\ \nu_\ell > 0,\ \underbrace{\|\rho\| < 1}_{\text{For stationarity}}\ \}$$

$$\rho = (\rho_{k,\ell})_{k,\ell \leq M}, \quad \rho_{k,\ell} = \int h_{k,\ell}(x)dx$$

- Metrics : L_1 and L_1 stochastic

$$d_{1,T}(\theta, \theta^*) = \int_0^T \sum_\ell |\lambda_{\ell,\theta} - \lambda_{\ell,\theta^*}|(t)dt,$$

$$\|\theta - \theta^*\|_1 = \sum_\ell |\nu_\ell - \nu_\ell^*| + \sum_{\ell,k} \|h_{k,\ell} - h_{k,\ell}^*\|_1$$

$$\lambda_{\ell,\theta}(t) = \nu_\ell + \sum_{k=1}^M \int_{t-s_{max}}^{t^-} h_{k,\ell}(t-u)dN_u^k$$

- Observations $\mathbf{N}_{[-s_{max}, T]} = (N_t^\ell, t \leq T, \ell \leq M)$.

Concentration in $d_{1,T}$

- True $\theta^* = (\nu^*, h^*)$ with $\nu^* > 0$ and $\|\rho^*\| < 1$. If
  - **KL** : $\exists c > 0$ such that

$$\Pi(|\nu - \nu^*| < \epsilon_T; \max_{\ell,k} \|h_{\ell,k} - h^*_{\ell,k}\|_2 \leq \epsilon_T (\log\log T)^{-1/2}) > e^{-cT\epsilon_T^2}$$

  - **Sieve** : $\exists \Theta_T \subset \Theta$, $\Pi(\Theta_T^c) = o(e^{-CT\epsilon_T^2})$
  - **Entropy** :

$$\log N(\epsilon_T, \Theta_T, \|.\|_1) \lesssim T\epsilon_T^2$$

Then

$$\mathbb{E}^* \left[\pi \left(d_{1,T}(\theta^*, \theta) > M\epsilon_T | \mathbf{N}\right)\right] \lesssim \frac{1}{T\epsilon_T^2}$$

How does it work : (1) concentration in $d_{1,T}$

- Prove that
  $\forall \theta \in \{|\nu - \nu^*| < \epsilon_T; \max_{\ell,k} \|h_{\ell,k} - h^*_{\ell,k}\|_2 \leq \epsilon_T / \log\log T\}$

  $$KL_T(\theta^*, \theta) \leq \kappa T \epsilon_T^2, \quad \mathbb{P}^* \left(\ell_T(\theta) - \ell_T(\theta^*) < -(\kappa+1) T \epsilon_T^2\right) \leq \frac{\log T}{T \epsilon_T^2}$$

- tests : $\theta_1 \in \Theta_T$, $S_j = \{\theta, d_{1,T}(\theta^*, \theta) \in (j\epsilon_T, (j+1)\epsilon_T)\}$.
  $A_1 = \{t;\ \lambda_{\theta_1} \geq \lambda_{\theta^*}(t)\}$,

  $$\phi_{\theta_1} = \max_{\ell} \left(\mathbb{I}\{N^\ell(A_1) - \Lambda^\ell(A_1; \theta^*) \geq jT\epsilon_T/8\} \vee \mathbb{I}\{N^\ell(A_1^c) - \Lambda^\ell(A_1^c; \theta^*) \geq j\right.$$

  then

  $$\mathbb{E}^*\left(\mathbb{I}_{\Omega_T}\phi_{\theta_1}\right) + \sup_{\|\theta-\theta_1\|_1 \leq j c_0 \epsilon_T} \mathbb{E}^*\left[\mathbb{E}_\theta\left(\mathbb{I}_{\Omega_T}\mathbb{I}_{\theta\in S_j}(1-\phi_{\theta_1}) | \mathcal{G}_0\right)\right] \leq 2e^{-x_0 T j \epsilon_T [1 \wedge j\epsilon_T}$$

weak conditions - examples of priors

$$\rho \sim \pi_\rho; \quad \bar{h}_{k,\ell} = h_{k,\ell}/\rho_{k,\ell} \overset{iid}{\sim} \pi_h, \quad \nu_\ell \overset{iid}{\sim} \pi_\nu$$

- Histograms

$$\bar{h}_{k,\ell} = \delta_{k,\ell} g_{k,\ell}, \quad g_{k,\ell} \overset{iid,\mathcal{D}}{=} g, \quad \delta_{k,\ell} \overset{iid}{\sim} Be(p)$$

and g has distribution given by

$$g = \sum_{j=1}^{J} \frac{z_j w_j}{|I_j|}, \quad I_j = (s_j, s_{j+1}), \quad z_j \overset{iid}{\sim} Be(q), \quad \sum_{j:z_j=1} w_j = 1$$

$$J - 1 \sim \mathcal{P}(a), \quad (w_j, j \in \{z_j = 1\}) \sim \mathcal{D}(\alpha_J, \cdots, \alpha_J),$$

$$(s_1, \cdots, s_J) \sim \Pi_s$$

If $\bar{h}^*$ are Hölder β then

$$\epsilon_T \lesssim (T/\sqrt{\log\log T}\log T)^{-\beta/(2\beta+1)}, \quad \beta \leq 1$$

Mixtures of Betas : Adaptive extimation over Hölder classes

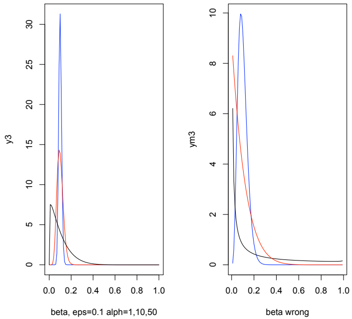
- Prior model

$$h_{k,\ell} = \left(\int_0^1 g_{\alpha,\epsilon} dM(\epsilon) \right)_+ , \quad g_{\alpha,\epsilon}(x) = \text{Beta}\left(\frac{\alpha}{1-\epsilon}, \frac{\alpha}{\epsilon} \right)$$

M : signed bounded measure on $[0,1]$.

- Why ?

$$E\left[\text{Beta}\left(\frac{\alpha}{1-\epsilon}, \frac{\alpha}{\epsilon} \right) \right] = \epsilon, \quad var\left[\text{Beta}\left(\frac{\alpha}{1-\epsilon}, \frac{\alpha}{\epsilon} \right) \right] = \frac{\epsilon(1-\epsilon)}{\alpha}$$



Why postitive part of mixtures of Betas ?

- (R. 08) : $\forall f \in \mathcal{H}(\beta, L)$, there exists f_1 such that

$$\|g_{\alpha,f_1} - f\|_\infty = O(\alpha^{-\beta/2}), \quad g_{\alpha,f_1}(x) = \int_0^1 g_{\alpha,\epsilon}(x) f_1(\epsilon) d\epsilon$$

and $\exists P = \sum_{j=1}^J w_j \delta_{(\epsilon_j)}$, $J \lesssim \sqrt{\alpha}(\log \alpha)^{3/2}$

$$\|g_{\alpha,f_1} - g_{\alpha,P}\|_2 = O(\alpha^{-\beta/2})$$

- Prior

$$J \sim \mathcal{P}(a), \quad (|w_1|, \cdots, |w_J|) \sim \mathcal{D}(\alpha/J, \cdots, \alpha/J), \quad \sqrt{\alpha} \sim \Gamma(a_0, b_0)$$

with

$$w_j = \zeta_j |w_j|, \quad \zeta_j \in \{-1, 1\} \quad iid, \quad \epsilon_j \overset{iid}{\sim} \text{Beta}(a_1, b_1)$$

- Posterior concentration rate : adative

$$\epsilon_T \lesssim T^{-\beta/(2\beta+1)} (\log T)^q$$

Concentration in L_1 :

$$\|\theta - \theta_0\|_1 = \sum_\ell |\nu_\ell - \nu_\ell^*| + \sum_{\ell,k} \|h_{\ell,k} - h^*_{\ell,k}\|_1$$

$$d_{1,T}(\theta, \theta^*) = \int_0^T |\lambda_\theta(t) - \lambda_{\theta^*}(t)|dt \text{ : non explicit}$$

Theorem

Under the same assumptions and if

$$\Pi(\|\rho\| > 1 - u_0\epsilon_T) \leq e^{-cT\epsilon_T^2}$$

then $\exists M_0 > 0$

$$\mathbb{E}^* \left[\Pi\left(\|\theta - \theta^*\|_1 > M_0\epsilon_T|\mathbf{N}\right)\right] = o(1)$$

From $d_{1,T}$ to L_1

$$\mathbb{E}^* \left[\Pi(\{d_{1,T}(\theta,\theta^*) \leq \epsilon_T\} \cap \{\|\theta - \theta^*\|_1 > M_0 \epsilon_T | \mathbf{N})\right]$$

$$\leq \mathbb{P}^* \left(D_T < e^{-cT\epsilon_T^2} \right) + e^{cT\epsilon_T^2} \int_{\|\theta-\theta^*\|_1 > M_0\epsilon_T} \mathbb{E}^* \left[\mathbb{P}_\theta \left[d_{1,T}(\theta^*, \theta) \leq \epsilon_T | \mathcal{G}_0 \right] \right] d\Pi(\theta$$

Some consequences

- Point estimators $\hat{f} = E[f|\mathbf{N}]$ satisfies

$$\mathbb{P}_{\theta^*}\left(\|f^* - \hat{f}\|_1 \lesssim \epsilon_T\right) \to 1$$

- *Easy conditions* Compared to Hansen et al. – who had a stronger condition on the type dictionnary

Simulation study

- Prior : random histogram

$$h_{\ell,k} = \underbrace{\delta_{\ell,k}}_{\text{Be}(1/2)} \sum_{j=1}^{M^{\ell},k} \mathbb{I}_{I_j^{\ell,k}} \frac{w_j^{\ell,k}}{|I_j^{\ell,k}|}, \quad M^{\ell,k} \overset{iid}{\sim} 1 + \mathcal{P}(a)$$

$$\log \nu_\ell \sim \mathcal{N}(3,1), \quad w_j^{\ell,k} \sim \text{Be}(1/2) \times \ln \mathcal{N}(\mu_\alpha, s_\alpha^2)$$

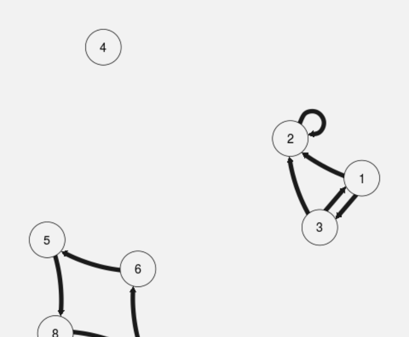
Scenarios

- $\nu_\ell = 20$
- *Scenario 1* : We first consider $K = 2$ neurons and piecewise constant interactions :

$$h_{1,1} = 30 \cdot \mathbb{I}_{(0,0.02]}, \quad h_{2,1} = 30 \cdot \mathbb{I}_{(0,0.01]}, \quad h_{1,2} = 30 \cdot \mathbb{I}_{(0.01,0.02]}, \quad h_{2,2} = 0.$$

- *Scenario 2* : In this scenario, we mimic $K = 8$ neurons belonging to three independent groups. The non-null interactions are the piecewise constant functions defined as :

$$h_{2,1} = h_{3,1} = h_{2,2} = h_{1,3} = h_{2,3} = h_{8,5} = h_{5,6} = h_{6,7} = h_{7,8} = 30 \cdot \mathbb{I}_{(0,0.02]}.$$



Results Scenario 1 : K=2 – ν

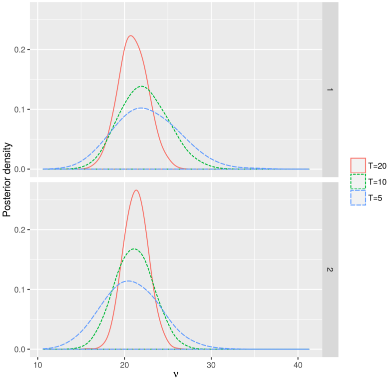


Figure: **Results for scenario 1**. *On the left*, posterior distribution of (ν_1, ν_2) with $T = 5$, $T = 10$ and $T = 20$ for one dataset. *On the right*, distribution of the posterior mean of (ν_1, ν_2) $\left(\widehat{\mathbb{E}}\left[\nu_k | (N_t^{sim})_{t\ in[0,T]}\right]\right)_{sim=1\ldots25}$ over the 25 simulated datasets.

Results Scenario 1 : K=2 – $h_{k,\ell}$

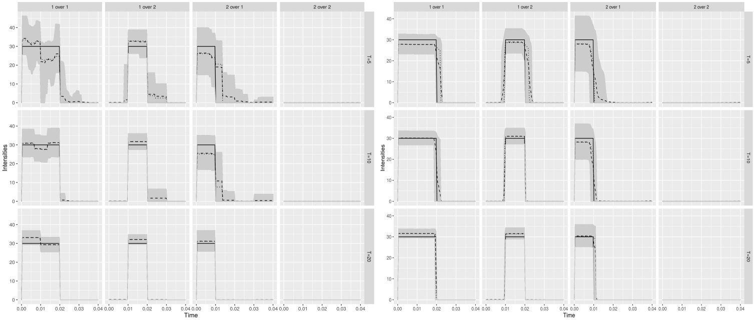
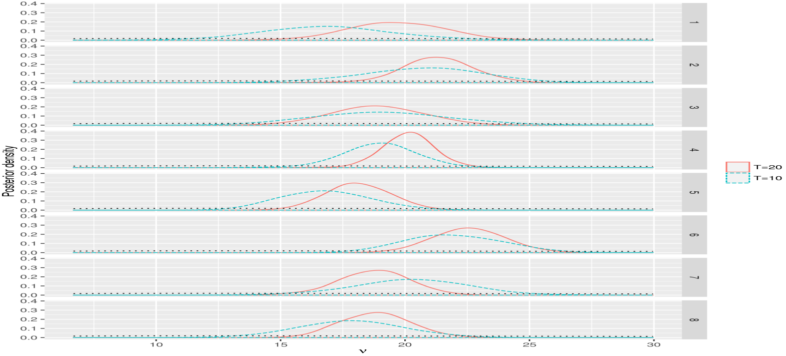


Figure: **Scenario 1, K=2**. Estimation of the $(h_{\ell,k})_{\ell,k=1,2}$ using the regular prior (left) continuous prior (right). The gray region indicates the credible region for $h_{\ell,k}(t)$ (delimited by the 5% and 95% percentiles of the posterior distribution). The true $h_{\ell,k}$ is in plain line, the posterior expectation and posterior median for $h_{\ell,k}(t)$ are in dotted and dashed lines respectively.

Results for Scenario 2 : K=8–ν_ℓ



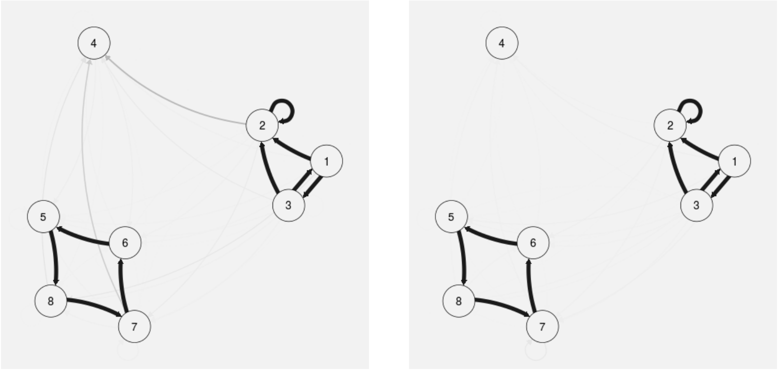


Figure: **Results for scenario 2 for one given dataset**. Posterior estimation of the interaction graph for $T = 10$ *on the left* and $T = 20$ *on the right*, for one randomly chosen dataset. Level of grey and width of the edges proportional to the posterior estimated probability of $\widehat{\mathbb{P}}(\delta_{\ell,k} = 1 | (N_t^{sim})_{t \ in[0,T]})$.

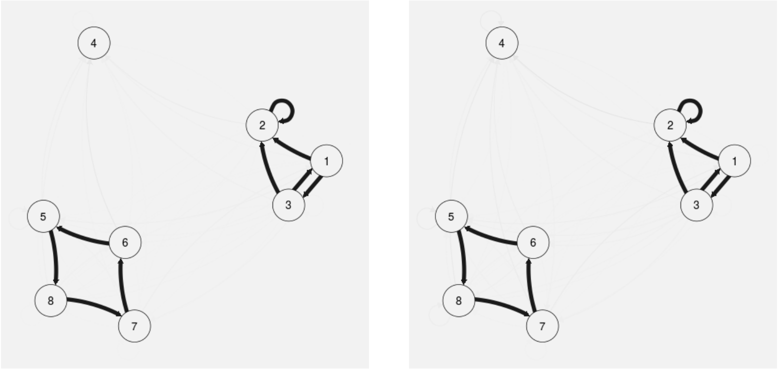


Figure: **Results for scenario 2 over the** 25 **simulated datasets**. Posterior estimation of the interaction graph for $T = 10$ *on the left* and $T = 20$ *on the right*. Level of grey and width of the edges are proportional to the posterior estimated probability of $\frac{1}{25}\sum_{sim=1}^{25}\widehat{\mathbb{P}}(\delta_{\ell,k} = 1|(N_t^{sim})_{t\ in[0,T]})$.

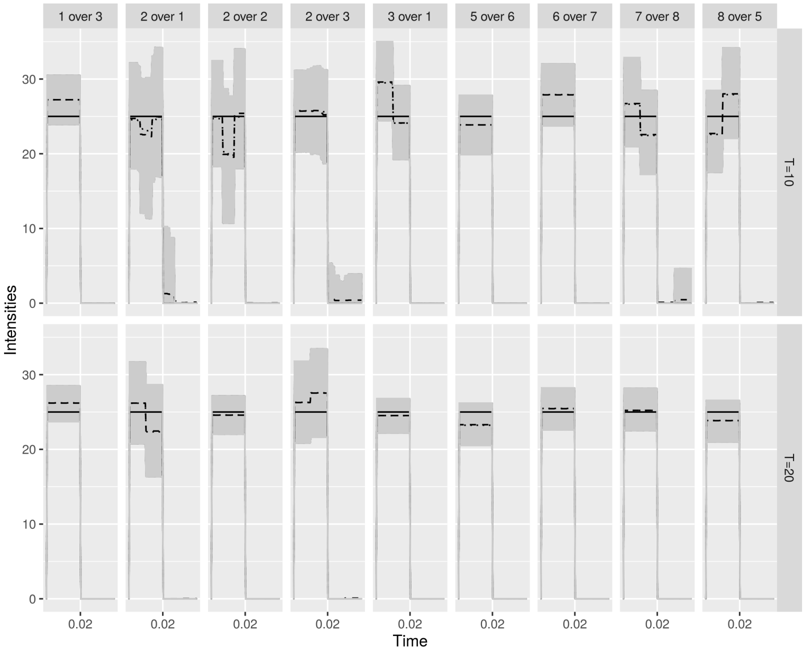


Figure: **Results for scenario 2 for one given dataset**. Estimation of the non null

Conclusion

- Theory for L_1 posterior concentration rates under quite weak assumptions
- Only for non negative $h_{k,\ell}$
- Simulations : too slow for the moment to treat many neurons. Ok for $<= 10$ (just) – How about the mixture of Beta or other priors ?
- Understanding credible regions

Thank You

Hansen, N. R., Reynaud-Bouret, P., and Rivoirard, V. (2014).
Lasso and probabilistic inequalities for multivariate point processes. to appear in, .
Bernoulli, to appear.

Rasmussen, J. (2013).
Bayesian inference for hawkes processes.
Methodology and Computing in Applied Probability, 15(3) :623–642.

Reynaud-Bouret, P., Rivoirard, V., and Tuleau-Malot, C. (2013).
Inference of functional connectivity in neurosciences via hawkes processes.
1st IEEE global Conference on Signal and Information Processing.