Variational Bayesian estimation for multivariate non linear Hawkes processes

Judith Rousseau

University of Oxford and Université Paris-Dauphine

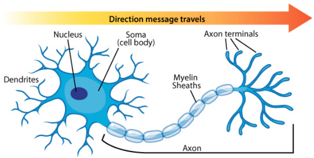
Co-authors

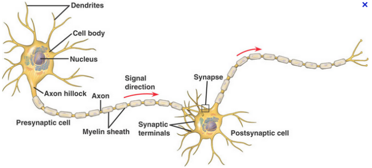
- DÉBORAH SULEM
  Pompeu Fabra University

- VINCENT RIVOIRARD
  I Université Paris Dauphine

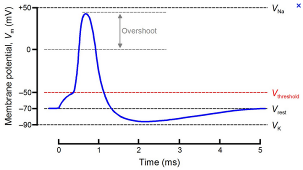
Functional connectivity graph of neurons

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.

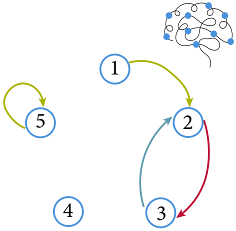


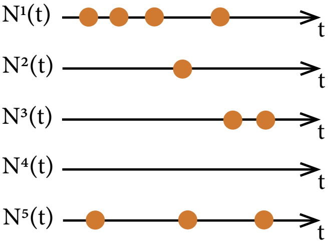
© PhysiologyWeb at www.physiologyweb.com

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.

Functional connectivity graph of neurons

Observations: spike trains on 5 neurons on a time window $[0, T]$





Graph of interactions:

$$\delta = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

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

Point process on $\mathbb{R}$

- N is a point process $\Leftrightarrow$ N is a random set of points on $\mathbb{R}$.

$$N(A) = \textit{number of points in } A$$

- Intensity function

$$\lambda(t)dt = P(\textit{point in } [t, t+dt) | N_s, s < t)$$

- example PP : $\lambda(t)$ is deterministic.
- Univariate linear Hawkes process

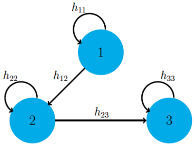
$$\lambda(t) = \nu + \int^{t^-} h(t-s)dN_s = \nu + \sum_{t_i < t} h(t - t_i), \quad h \geq 0$$

h : excitation function, ν background rate

Multidimensional non linear Hawkes processes

- K neurones interacting: K Point Proc **non independent.** $N = (N^{(1)}, \cdots, N^{(k)})$

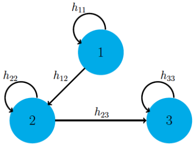
$$\lambda_t^{(k)} = \psi_k\Big(\nu_k + \sum_{\ell=1}^{K} \int_{-\infty}^{t-} h_{\ell k}(t-u)\mathrm{d}N^{(\ell)}(u)\Big)$$
$$= \psi_k\Big(\nu_k + \sum_{\ell=1}^{K} \sum_{T_\ell \in N^{(\ell)},\, T_\ell < t} h_{\ell k}(t - T_\ell)\Big)$$



Multidimensional non linear Hawkes processes

- K neurones interacting: K Point Proc **non independent.** $N = (N^{(1)}, \cdots, N^{(k)})$

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



- We obtain mutually exciting and inhibiting processes:

  $\nu_k > 0$: background rates

  ψ_k: positive and nondecreasing link function

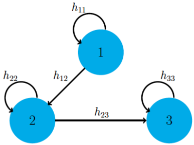
  $h_{\ell k}$: interaction functions

  - If $h_{\ell k} \geq 0$: excitation
  - If $h_{\ell k} \leq 0$: inhibition
  - If $h_{\ell k}$ is signed: both

Multidimensional non linear Hawkes processes

- K neurones interacting: K Point Proc **non independent.** $N = (N^{(1)}, \cdots, N^{(k)})$

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



- 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: both

ψ_k: positive and nondecreasing link function
Typical examples:
- linear: $\psi_k(x) = x$ [requires $h_{\ell k} \geq 0$]
- nonlinear: $\psi_k(x) = x_+$
or $\psi_k(x) = e^x/(1+e^x)$ etc . . .

Multivariate Hawkes processes- II

- Existence and uniqueness of a stationary distribution for N established by Brémaud and Massoulié (1996, 2001).
- See also Delattre, Fournier and Hoffmann (2016) and Costa, Graham, Marsalle and Tran (2020) : probabilistic results.
- 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 $(\lambda^{(k)})_{k \in [\![1;K]\!]}$.

State of the art for theoretical results

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

- Linear case in the nonparametric setting: $\psi_k(x) = x$
  - 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, Rivoirard and Rousseau (2020)
- Nonlinear case:
  - Chen, Shojaie, Shea-Brown and Witten (2019) provided an asymptotic analysis of second order statistics (cross-covariance)
  - Parametric approaches: Bonnet, Martinez Herrera and Sangnier (2021a,b) : MLE for exponential kernel functions. Lemonnier and Vayatis (2014): Linear approximation with exponential kernels. Deutsch and Ross (2022) : Bayesian modelling
- Our contribution: (scalable) nonparametric estimation of f in multivariate nonlinear Hawkes

Inference for nonlinear Hawkes models

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

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

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

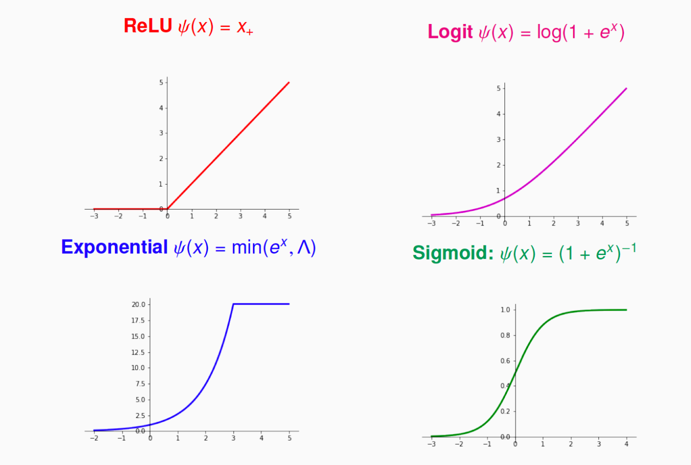
- Assumptions:
  - $\nu_k \in \mathbb{R}$, $k \leq K$.
  - the supp$(h_{\ell k}) \subset [0, A]$, $A > 0$ fixed. $h_{\ell k}$' can be negative, so inhibition is possible.

- 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 $[-A, T]$ with in mind $T \to +\infty$

Typical link functions



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

$$h^+ = \max(h,0), \quad h^- = \max(-h,0), \quad S^+ = (S^+_{lk})_{l,k\le K}, \quad S^+_{lk} = \int h^+_{lk}(x)dx$$

Proposition

- *If $\psi \le \Lambda$ or ψ is L-Lipschitz and $\|S^+\| < 1$*
  then $\exists$ unique stationary version of N with finite average
- *If ψ is bijective on I s.t. $\forall k$*

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

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

Examples : - softmax $\psi(x) = \log(1+e^x)$ and sigmoid $\psi(x) = (1+e^{-x})^{-1}$
- ReLU function, $\psi(x) = \max(x,0)$, if $\forall k$, $\max_\ell \|h^-_{\ell k}\|_\infty < \nu_k$
- clipped exponential , $\psi(x) = \min(e^x, \Lambda)$, if $\forall k$, $\max_\ell \|h^+_{\ell k}\|_\infty + \nu_k < \log \Lambda$.

Bayesian inference framework

- We assume that we observe over $[-A, T]$ a stationary Hawkes process $N = (N^{(1)}, \cdots, N^{(K)})$.
- log-likelihood at $f = (\nu, h) \in \mathcal{F} \subset \mathbb{R}^K \times \mathcal{H}^{K^2}$:

$$L_T(f) := \sum_{k=1}^{K} L_T^k(f), \quad L_T^k(f) = \left[\int_0^T \log(\lambda_t^k(f)) dN_t^k - \int_0^T \lambda_t^k(f) dt \right].$$

- Π prior distribution on $\mathcal{F}$. Then posterior:

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

Remark: Posterior is doubly intractable.

Posterior concentration rates

- true Parameter $f_0 = (\nu^0, \mathbf{h}^0)$
- Conditions for stationarity and identifiability hold and $\inf_x \psi(x) > 0$.
- $\|f - f_0\|_1 = \sum_k |\nu_k - \nu_k^0| + \sum_{l,k} \|h_{lk} - h_{lk}^0\|_1$

$$B_2(\epsilon_T, B) = \left\{ f \in \mathcal{F}; \quad \|\nu - \nu^0\|_{\ell_\infty} \leq \epsilon_T, \ \max_{\ell,k} \|h_{\ell k} - h_{\ell k}^0\|_2 \leq \epsilon_T, \ \max_{\ell,k} \|h_{\ell k}\|_\infty < B \right\}.$$

Theorem

If Π and f_0 verify :

(A0) $\exists c_1 > 0$ *s.t.* $\Pi(B_2(\epsilon_T, B)) \geq e^{-c_1 T \epsilon_T^2}$

(A1) $\exists \mathcal{F}_T \subset \mathcal{F}$ *and* $\zeta_0, x_0 > 0$ *s.t.*

$$\Pi(\mathcal{F}_T^c) = o(e^{-c_1 T \epsilon_T^2}), \quad \log \mathcal{N}(\zeta_0 \epsilon_T, \mathcal{F}_T, \| \cdot \|_1) \leq x_0 T \epsilon_T^2.$$

Then, for $M > 0$ large enough, we have

$$\mathbb{E}_0 \left[\Pi \left(\|f - f_0\|_1 > M \epsilon_T \sqrt{\log T} \middle| N \right) \right] = o(1).$$

Examples of priors: spike and slab parametrization

$$d\Pi(f) = d\Pi_h(\mathbf{h}) \prod_k d\Pi_\nu(\nu_k), \quad f = (\nu_k, (h_{\ell k})_{\ell \in [\![1;K]\!]})_{k \in [\![1;K]\!]}$$

with

1. Π_ν having a positive and continuous density on $\mathbb{R}_+^*$, e.g. a Gamma distribution.
2. Spike and slab parametrization: $\mathbf{h} = (h_{\ell k})_{\ell,k}$,

$$h_{\ell k} = \delta_{\ell k} \bar{h}_{\ell k}, \quad \delta_{\ell k} \in \{0,1\}, \quad \bar{h}_{l,k} \neq 0 \quad \Leftrightarrow \delta_{l,k} \neq 0$$

$\delta = (\delta_{\ell k})_{\ell k}$= connectivity graph.

We then consider

(a) $\delta \sim \Pi_\delta$, where Π_δ is a prior on $\{0,1\}^{K^2}$, e.g. $\delta_{\ell k} \overset{i.i.d.}{\sim} \mathcal{B}er(p)$

(b) Given δ,

$$d\Pi_h(h|\delta) \propto \left(\prod_{\ell,k} d\tilde{\Pi}_{h|\delta}(h_{\ell k}) \right) [\times 1_{\|S^+\| < 1}(h),]$$

with $\quad \tilde{\Pi}_{h|\delta}(h_{\ell k}) = \delta_{\ell k} \tilde{\Pi}_h(\bar{h}_{\ell k}) + (1 - \delta_{\ell k}) \delta_{\{0\}}(\bar{h}_{\ell k}),$

Example of priors for $\tilde{\Pi}_h(\bar{h}_{\ell k})$

- Gaussian process or hierarchical Gaussian process priors
- Splines/ bases expansions

$$\bar{h}_{l,k} = \sum_{j \leq J_{lk}} w_{j;lk} e_j, \quad J_{lk} \sim \Pi_J, \quad \theta_{j;lk} \overset{iid}{\sim} \pi_w$$

- mixture types models

Corollary

Holder interaction functions: $h^0_{\ell k} \in \mathcal{H}(\beta, L_0), \quad (\ell, k) \in [K]^2, \beta, L_0 > 0.$ *Then*

$$\mathbb{E}_0 \left[\Pi\left(\|f - f_0\|_1 > \epsilon_T \middle| N\right)\right] = o(1), \quad \epsilon_T = MT^{-\frac{\beta}{2\beta+1}} (\log T)^{\square},$$

optimal up to the logarithmic term and $\hat{f} = (\hat{\nu}, \hat{h}) = \mathbb{E}^{\Pi}[f|N]$

$$\|\hat{f} - f_0\|_1 = O_{P_0}(\epsilon_T)$$

Numerical results - Histogram case

- We sample one observation of a Hawkes process with K neurons, link function ψ and parameter $f_0 = (\nu^0, h^0)$ on $[0, T]$. We take $A = 0.1$.
- We assume $h^0 \in \mathcal{H}^D_{histo}$ for some $D \geq 1$, with

$$\mathcal{H}^D_{histo} = \left\{ h = (h_{\ell k})_{\ell,k};\ h_{\ell k}(x) = \sum_{j=1}^{2^D} w^j_{\ell k} e_j(x),\ x \in [0, A] \right\}, \quad e_j = \frac{2^D}{A} 1_{\left[\frac{A(j-1)}{2^D}, \frac{Aj}{2^D}\right)}$$

- The tested link functions are approximations of ReLU : $\max(x, 0)$:

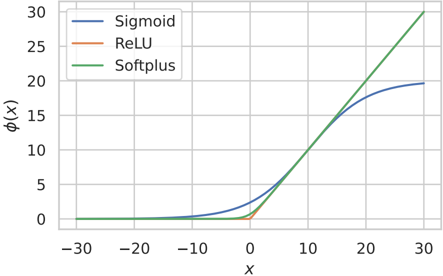


Figure: 3 tested link functions: $x \longmapsto \psi(x)$

MCMC in the univariate nonlinear Hawkes model

- We set $K = 1$, $\nu^0 = 6$ and test three scenarios:
  - Excitation only : h_{11} is positive on its support
  - Mixed effect: h_{11} takes positive and negative values
  - Inhibition only: h_{11} is positive on its support

  h^0 is piecewise constant. The number of pieces is known: $2^D = 4$.
- $T = 500$. Computational time of the algorithm: $\approx$ 40 minutes.

Scenario	Sigmoid	ReLU	Softplus
Excitation only	5250	5352	4953
Mixed effect	3876	3684	3418
Inhibition only	3047	2724	2596

Table: Number of observations for each scenario

- We consider a normal prior on $\mathcal{H}^D_{histo}$:

$$\nu_1 \sim \mathcal{N}(0, 25), \quad w^j_{11} \overset{i.i.d.}{\sim} \mathcal{N}(0, 25), \quad j = 1, 2, 3, 4$$

combined with a Metropolis-Hasting sampler.

MCMC in the univariate nonlinear Hawkes model

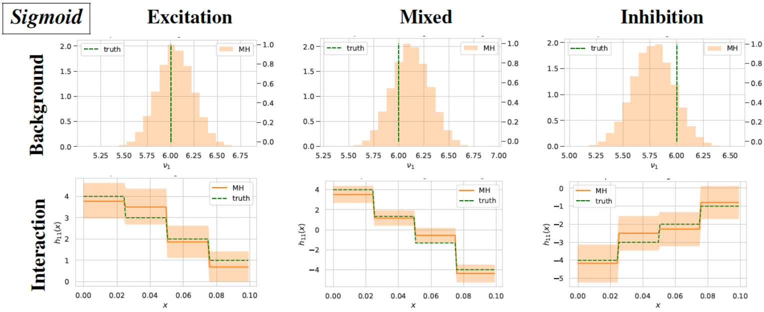


Figure: Posterior distribution on $f = (\nu_1, h_{11})$ obtained with the MH sampler in the sigmoid model. The first row contains the marginal distribution on the background rate ν_1, and the second row represents the posterior mean (solid line) and 95% credible sets (colored areas) on the interaction function h_{11}. The true parameter f^0 is plotted in dotted green line.

Variational Bayesian estimation

- Difficulty of computing the nonparametric posterior distribution since

$$\Pi(B|N) = \frac{\int_B e^{L_T(f)} d\Pi(f)}{\int_{\mathcal{F}} e^{L_T(f)} d\Pi(f)} \quad e^{L_T(f)} = \prod_{k=1}^{K} \left[e^{-\int_0^T \lambda_t^k(f) dt} \prod_{\substack{T_k \in N^{(k)} \\ T_k \leq T}} \lambda_{T_k}^{(k)}(f) \right]$$

and

$$\lambda_t^{(k)}(f) = \psi\left(\nu_k + \sum_{\ell=1}^{K} \sum_{\substack{T_\ell \in N^{(\ell)} \\ T_\ell < t}} h_{\ell k}(t - T_\ell) \right)$$

Variational Bayes methods consist in approximating the posterior distribution.

- Let $\mathcal{V}$ be an approximating family of distributions on $\mathcal{F}$.

$$\hat{Q} := \arg\min_{Q \in \mathcal{V}} KL\left(Q || \Pi(.|N)\right), \quad KL(Q||Q') := \begin{cases} \int \log\left(\frac{dQ}{dQ'}\right) dQ & \text{if } Q \ll Q' \\ +\infty & \text{otherwise} \end{cases}$$

- variational class:mean-field family

$$\mathcal{V}_{MF} = \left\{ Q;\ dQ(\vartheta) = \prod_{d=1}^{D} dQ_d(\vartheta_d) \right\}.$$

Sigmoid link function: augmented scheme of Zhou et al. (2021) and Malem-Shinitski et al. (2021)

$$\psi(x) = (1+e^{-x})^{-1} = \mathbb{E}_{\omega \sim p_{PG}}\left[e^{g(\omega,x)}\right], \quad g(\omega,x) = -\frac{\omega x^2}{2} + \frac{x}{2} - \log 2$$

with p_{PG}: the Polya-Gamma density

$$L_T(f,\omega;N) = \sum_{k\in[K]}\left\{\sum_{i\in[N_k]}\left(g(\omega_i^k, \tilde{\lambda}_{T_i^k}(f)) + \log p_{PG}(\omega_i^k)\right) - \int_0^T \psi(\tilde{\lambda}_t^k(f))dt\right\},$$

$$\exp\left\{-\int_0^T \psi(\tilde{\lambda}_t^k(f))dt\right\} = \mathbb{E}\prod_{(\bar{T}_j^k,\bar{\omega}_j^k)\in\bar{N}^k} e^{g(\bar{\omega}_k, -\tilde{\lambda}^k_{\bar{T}_j}(f))}$$

Augmented mean field VB

$$\mathcal{V}_{AMF} = \{Q \; Q(f,z) = Q_1(f)Q_2(z)\},$$
$$z = (\omega_i^k, i \le N_k, ; \bar{N}^k, k \le K) \quad \text{augmented variables}$$

Augmented mean-field VB posterior:

$$\hat{Q}_{AMF}(f,z) := \arg\min_{Q \in \mathcal{V}_{AMF}} KL\left(Q(f,z) || \Pi_A(f,z|N)\right) =: \hat{Q}_1(f)\hat{Q}_2(z)$$

$$\hat{Q}_1(f) \propto \exp\left\{\mathbb{E}_{\hat{Q}_2}[\log p(f,z,N)]\right\},$$
$$\hat{Q}_2(z) \propto \exp\left\{\mathbb{E}_{\hat{Q}_1}[\log p(f,z,N)]\right\},$$

where $p(f,z,N)$ is the joint density of the f,z,N

sparse large connectivity graph : model selection VB

- Parameter $f = (\nu_k, \delta_{lk}\bar{h}_{lk}, l, k \leq K)$
- Model selection for δ : 2^{K^2} models
- model selection for $[h_{lk}|\delta_{lk} = 1] = \sum_{j=1}^{J_k} w_{j;lk} e_j, \ w_{j;lk} \sim \mathcal{N}(0, \sigma^2)$.

Model selection VB $\forall s = (J_k, k \leq K; \delta)$

- Compute VB posterior $\hat{Q}_s$ and

$$\mathrm{ELBO}(\hat{Q}_s) = E_{\hat{Q}_s}\left[\log \frac{p(f, z, N)}{\hat{q}_s(f, z)}\right]$$

- Choose

$$\hat{s} = \mathrm{argmax}_s \mathrm{ELBO}(\hat{Q}_s)$$

- Can also do model averaging

$$\hat{Q}(f) = \frac{\pi(s) e^{\mathrm{ELBO}(\hat{Q}_s)}(f)}{\sum_{s'} \pi(s) e^{\mathrm{ELBO}(\hat{Q}_s)}(f)}$$

But K large : too many models

Two step procedure: thresholding = cheap and dirty

Step 1 (i) Fix $\delta_{lk} = 1$ (complete graph) and do model selection VB on $s = (J_k, k \leq K)$:

$$\hat{Q}^0 = \hat{Q}^{\text{full}}_{\hat{s}}(f) \Rightarrow \hat{S}_{lk} = \|\hat{h}^{\text{full}}_{lk}\|_1 \quad \hat{h}^{\text{full}}_{lk} = E_{\hat{Q}^0}(h_{lk})$$

- (ii) order and plot $\hat{S}_{lk}$ and threshold at the first jump : η_0

$$\hat{\delta}_{lk} = 0 \quad \Leftrightarrow \hat{S}_{lk} \leq \eta_0$$

Step 2 Compute $\hat{Q}^{\hat{\delta}}_{\hat{s}}$ model selection VB with graph $\hat{\delta}$.

illustration of the thresholding heuristic

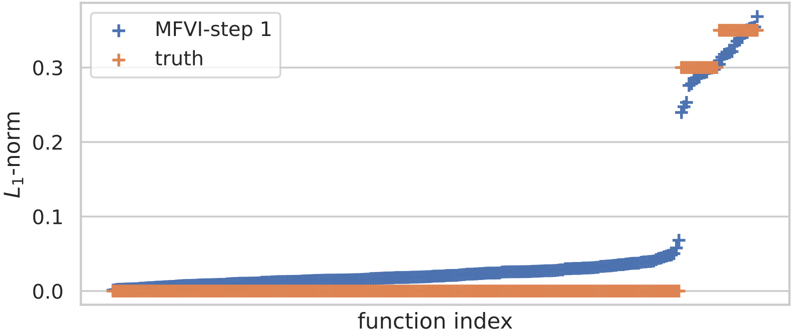


Figure: $\hat{S}_{lk}$ - based on the full graph VB, plotted in increasing order with $K = 16$.

Theoretical justification: K fixed (but possibly large) $T \to \infty$ using Zhang and Gao, 20; Ridgeway, Alquier

- mean field VB and model selection/model averaging VB
- same assumptions on Π and f_0 as in posterior concentration rates leading to

$$\Pi(\|f - f_0\|_1 > \epsilon_T | N) = o_{P_0}(1)$$

$$\text{Then} \quad \hat{Q}_{\hat{s}}^{\text{full}}(\|f - f_0\|_1 > \epsilon_T) = o_{P_0}(1)$$

- It implies the convergence of $\hat{S}$

$$\Rightarrow \quad \max_{lk} |\hat{S}_{lk} - S_{lk}^0| = O_{P_0}(\epsilon_T)$$

- If $\min\{S_{lk}^0; \delta_{lk}^0 = 1\} >> \epsilon_T$ then

$$\max\{\hat{S}_{lk}; \delta_{lk}^0 = 0\} = o_{P_0}(\min\{\hat{S}_{lk}; \delta_{lk}^0 = 1\}) \quad \Rightarrow \mathbb{P}_0(\hat{\delta} = \delta_0) = 1 + o(1)$$

- Then

$$\hat{Q}_{\hat{s}}^{\hat{\delta}}(\|f - f_0\|_1 > M\epsilon_T | N) = o_{P_0}(1).$$

sparse graph scenario, well specified link function; $K \in \{2, 4, 8, 16, 32, 64\}$

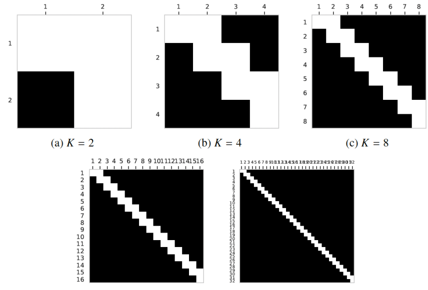
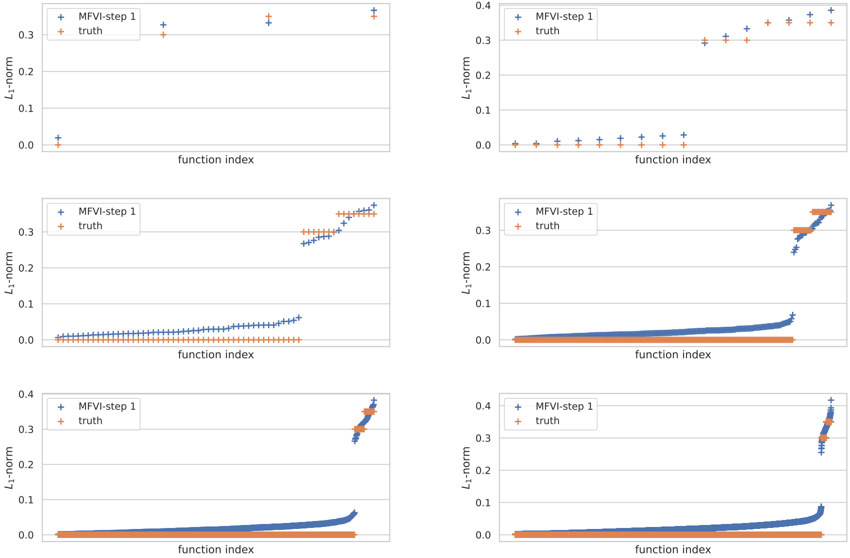

(a) $K = 2$ (b) $K = 4$ (c) $K = 8$ (d) $K = 16$ (e) $K = 32$

Figure: True (sparse) graph parameter for different dimensions : $2K - 1$ non-zero interaction functions for $K \in \{2, 4, 8, 16, 32\}$ (correspond to white squares)

K	Scenario	# observations
2	Excitation inhibition	5680 4800
4	Excitation inhibition	11338 9895
8	Excitation inhibition	22514 19746
16	Excitation Self-inhibition	51246 37000
32	Excitation inhibition	96803 76106
64	Excitation inhibition	117800 133000

Step 1: thresholding in the complete graph, $k = 2, 4, 8, 16, 32, 64$



Variational Bayesian estimation

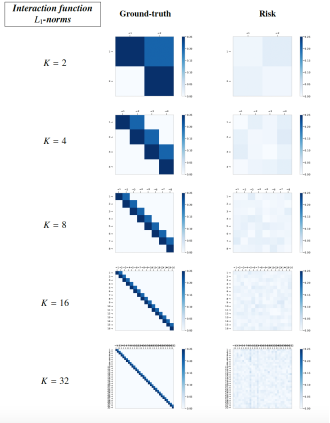


Figure: Heatmaps of the entries of the matrix $(\|h^0_{\ell k}\|_1)_{\ell,k}$ (left column) and $\mathbb{L}_1$-risk, i.e., $(\mathbb{E}[\|h^0_{\ell k} - h_{\ell k}\|_1])_{\ell,k}$ (right column) in the excitation scenario.

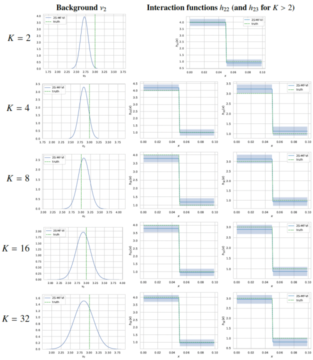
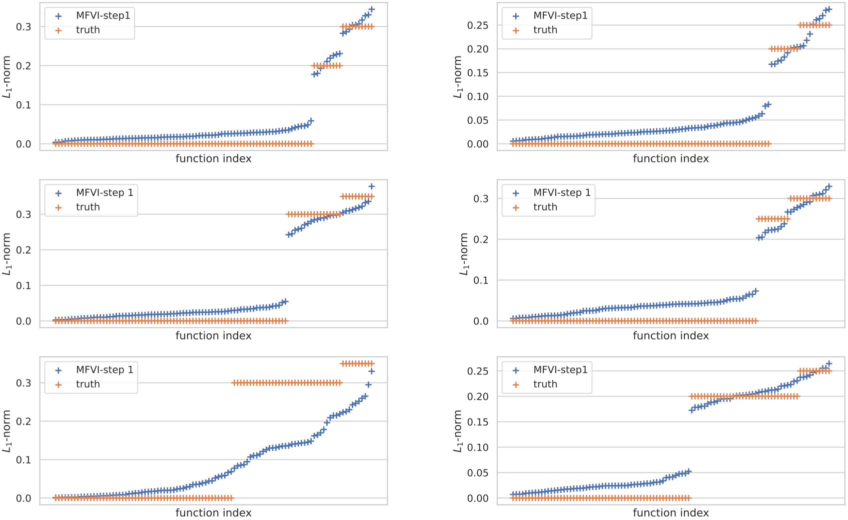


Figure: Mode variational posterior distributions on ν_2 (left column) and interaction functions h_{22} and h_{32} (for $K > 2$)(second and third columns) in the excitation scenario.

sparse-random-dense scenariis



Slide 13

Variational Bayesian estimation - Setting $T = 1000$

K	Scenario	$\hat{s} = s_0$	Risk
2	Excitation Self-inhibition	Yes Yes	0.408 0.277
4	Excitation Self-inhibition	Yes Yes	0.697 0.767
8	Excitation Self-inhibition	Yes Yes	1.672 2.312
16	Excitation Self-inhibition	Yes Yes	4.692 4.688
32	Excitation Self-inhibition	Yes Yes	11.066 12.074

Table: Performance of Algorithm. We report the $\mathbb{L}_1$-risk and if the model with largest marginal probability corresponds to the true one.

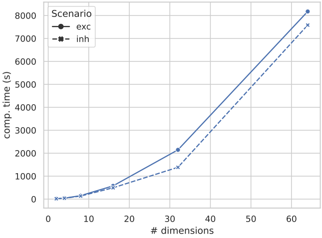


Figure: Computational times of our two-step mean-field variational algorithm in the Excitation (exc) and Self-inhibition (inh) scenarios for $K = 2, 4, 8, 16, 32, 64$.

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

References:

- Sulem D., Rivoirard V. and Rousseau J. (2023) *Bayesian estimation of nonlinear Hawkes processes.* To appear in Bernoulli

- Sulem D., Rivoirard V. and Rousseau J. (2023) *Scalable Variational Bayes methods for Hawkes processes.* In preparation.

European Research Council
Established by the European Commission