The non-linear discrete-time Hawkes process with a Gaussian perturbation

Mahmoud Khabou

Institut de mathématiques de Toulouse

June 2, 2022

Joint work with Lorick Huang

Table of Contents

Motivation for count series

- Many phenomena in nature/economy are observed on regular time intervals: earthquakes, prices, infections etc.

Motivation for count series

- Many phenomena in nature/economy are observed on regular time intervals: earthquakes, prices, infections etc.
- Unobserved variables are considered stochastic.

Motivation for count series

- Many phenomena in nature/economy are observed on regular time intervals: earthquakes, prices, infections etc.
- Unobserved variables are considered stochastic.
- Example: the auto-regressive model $AR(p)$:

$$X_n^h = a^h + \sum_{k=1}^{p} \phi_k^h X_{t-k}^h + \sigma^h \mathcal{N}(0,1),$$

where h is the size of the time interval (e.g $h = 1$m.s)

Motivation for count series

Figure 1: Hourly change of price of the BOBL on a period of 1 month.

Motivation for count series

Figure 2: BOBL futures minute by minute last Friday.

Examples of count series

- A toy example: The Bernoulli process $X_n \sim \mathcal{B}(h\lambda)$.

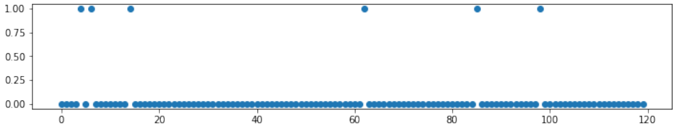

Figure 3: Minute by minute, $\lambda = 4$.

- Discrete time Hawkes process (Seol) $X_n \sim \mathcal{B}\left(h(\lambda + \sum_{k=1}^{n-1} \delta^h_{n-k} X_k)\right)$

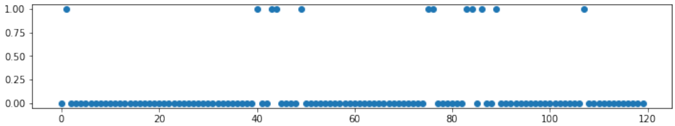

Figure 4: Minute by minute, $\delta^h_k = 1/(2 + 2(hk)^2)$.

INAR(p) process

- Sometimes it is possible to have more than one jump/event per time interval.

INAR(p) process

- Sometimes it is possible to have more than one jump/event per time interval.
- Bernoulli distributions do not allow for a parameter larger than 1.

INAR(p) process

- Sometimes it is possible to have more than one jump/event per time interval.
- Bernoulli distributions do not allow for a parameter larger than 1.
- The INAR(p) process is an integer valued process such that

$$X_n = \varepsilon_n + \sum_{i=1}^{p} \sum_{k=1}^{X_{n-i}} \xi_k^{(n,i)}, \quad n \in \mathbb{Z},$$

where $\varepsilon_n \sim \mathcal{P}(h\lambda)$ and $\xi_k^{(n,i)} \sim \mathcal{P}\left(h\delta_i^h\right)$ independently.

INAR(p) process

- Sometimes it is possible to have more than one jump/event per time interval.
- Bernoulli distributions do not allow for a parameter larger than 1.
- The INAR(p) process is an integer valued process such that

$$X_n = \varepsilon_n + \sum_{i=1}^{p} \sum_{k=1}^{X_{n-i}} \xi_k^{(n,i)}, \quad n \in \mathbb{Z},$$

  where $\varepsilon_n \sim \mathcal{P}(h\lambda)$ and $\xi_k^{(n,i)} \sim \mathcal{P}\left(h\delta_i^h\right)$ independently.
- p can be taken equal to ∞ (Kirchner).

Table of Contents

Simple case

- Let $\beta, \sigma \in \mathbb{R}_+^*$, $\alpha, \mu \in \mathbb{R}$ and $\phi : \mathbb{R} \to \mathbb{R}_+$.

Simple case

- Let $\beta, \sigma \in \mathbb{R}_+^*$, $\alpha, \mu \in \mathbb{R}$ and $\phi : \mathbb{R} \to \mathbb{R}_+$.
- Let $(X_n)_{n \in \mathbb{N}}$ be a count series. Define recursively

$$I_n = \mu(1 - e^{-\beta h}) + e^{-\beta h} I_{n-1} + \alpha X_{n-1} + \sigma \sqrt{h} \mathcal{N}(0,1).$$

such that $X_n \sim \mathcal{P}\left(h\phi(I_n)\right)$.

Simple case

- Let $\beta, \sigma \in \mathbb{R}_+^*$, $\alpha, \mu \in \mathbb{R}$ and $\phi : \mathbb{R} \to \mathbb{R}_+$.
- Let $(X_n)_{n \in \mathbb{N}}$ be a count series. Define recursively

$$I_n = \mu(1 - e^{-\beta h}) + e^{-\beta h} I_{n-1} + \alpha X_{n-1} + \sigma\sqrt{h}\mathcal{N}(0,1).$$

  such that $X_n \sim \mathcal{P}\left(h\phi(I_n)\right)$.
- The cumulative number of events is

$$H_n = \sum_{k=0}^{n} X_k.$$

Positive α

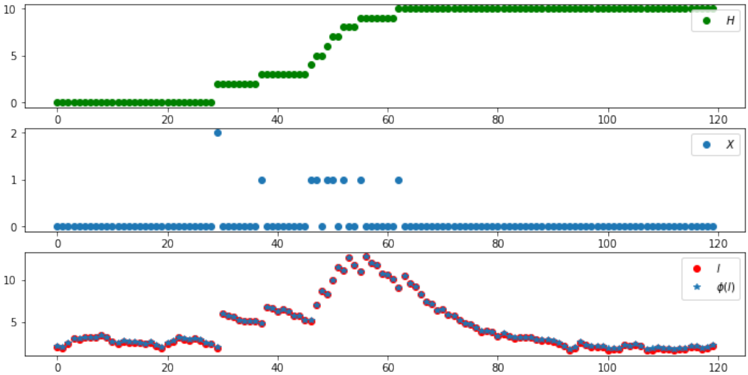

Figure 5: One simulation with $\alpha = 2, \beta = 5, \mu = 2, \sigma = 2$ and $\phi(x) = \ln(1 + e^x)$.

Negative α

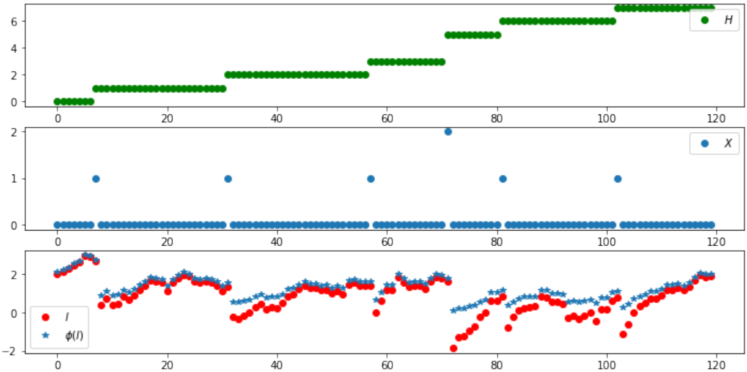

Figure 6: Same parameters, except $\alpha = -2$.

General multivariate case

- Assume that we observe d count series $\boldsymbol{X}_n = (X_n^1, \cdots, X_n^d)$ in discrete time, possibly correlated.
- We say that the $\boldsymbol{X}$ is a DTHP of order $p+1$ if

$$\boldsymbol{X}_n \sim \mathcal{P}\left(h\phi(\boldsymbol{a}_n)\right)$$

where

$$a_n^i = \tilde{a}^i(\boldsymbol{U}_n, \boldsymbol{Y}_n) = \mu_i + \sigma_i Y_n^i + \sum_{j=1}^{d} \langle \boldsymbol{\alpha}^{ij}, \boldsymbol{U}_n^{ij} \rangle,$$

and $\boldsymbol{U}_n$, $\boldsymbol{Y}_n$ are defined by

$$\begin{cases} \boldsymbol{U}_{n+1}^{ij} & = M_{ij}^h \boldsymbol{U}_n^{ij} + X_n^j \boldsymbol{e}_{p+1} \\ \boldsymbol{Y}_{n+1} & = \Omega^h \boldsymbol{Y}_n + \sqrt{h}\mathcal{N}_d \end{cases}$$

Markov property

Proposition

The cascade $(\boldsymbol{U}, \boldsymbol{Y}) \in \left((\mathbb{R}^{p+1})^{d\times d} \times \mathbb{R}^d\right)$ is a Markov chain with the one step generator

$$
\begin{aligned}
\mathcal{T}^h f(\boldsymbol{u},\boldsymbol{y}) =& \mathbb{E}\left[f(\boldsymbol{U}_{n+1}, \boldsymbol{Y}_{n+1}) \middle| (\boldsymbol{U}_n, \boldsymbol{Y}_n) = (\boldsymbol{u},\boldsymbol{y})\right], \\
=& e^{-h\sum_{i=1}^d \phi_i\left(\tilde{a}^i(\boldsymbol{u},\boldsymbol{y})\right)} \mathbb{E}^{\mathcal{N}}\left[f\left(\left(M_{ij}^h \boldsymbol{u}^{ij}\right)_{i,j=1,\cdots,d}, \Omega^h \boldsymbol{y} + \sqrt{h}\mathcal{N}_d\right)\right] \\
& + \sum_{m=1}^d \mathbb{E}^{\mathcal{N}}\left[f\left(\left(M_{ij}^h \boldsymbol{u}^{ij} + \mathbb{1}_{m=j}\boldsymbol{e}_{p+1}\right)_{i,j=1,\cdots,d}, \Omega^h \boldsymbol{y} + \sqrt{h}\mathcal{N}_d\right)\right] \\
& h\phi_m(\tilde{a}^m(\boldsymbol{u},\boldsymbol{y})) e^{-h\sum_{i=1}^d \phi_i\left(\tilde{a}^i(\boldsymbol{u},\boldsymbol{y})\right)} \\
& + R_h(\boldsymbol{u},\boldsymbol{y}).
\end{aligned}
$$

Table of Contents

Thinning

Definition

Let $\Gamma^1, \cdots, \Gamma^d$ be d Poisson measures on $\mathbb{R}_+^2$ with compensators $\mathrm{d}t\mathrm{d}\theta$. Let $\delta_{ij} : \mathbb{R}_+ \to \mathbb{R}$ and $\phi_i : \mathbb{R} \to \mathbb{R}_+$. The multivariate Hawkes process is a solution to the system

$$\begin{cases} N_t^i = \displaystyle\int_{(0,t]\times\mathbb{R}_+} \mathbb{1}_{\{\theta \le \lambda_s^i\}} \Gamma^i(\mathrm{d}s, \mathrm{d}\theta), \quad t \ge 0, \quad i = 1, \cdots, d \\ \lambda_t^i = \phi_i \left(\mu_i + \int_{[0,t)} \left(\sum_{j=1}^d \delta_{ij}(t-s)\mathrm{d}N_s^j \right) + \sigma_i e^{-\omega_i(t-s)} \mathrm{d}W_s^i \right). \end{cases} \tag{1}$$

Thinning

Definition

Let $\Gamma^1, \cdots, \Gamma^d$ be d Poisson measures on $\mathbb{R}_+^2$ with compensators $\mathrm{d}t\mathrm{d}\theta$. Let $\delta_{ij} : \mathbb{R}_+ \to \mathbb{R}$ and $\phi_i : \mathbb{R} \to \mathbb{R}_+$. The multivariate Hawkes process is a solution to the system

$$\begin{cases} N_t^i = \displaystyle\int_{(0,t]\times\mathbb{R}_+} \mathbb{1}_{\{\theta \leq \lambda_s^i\}} \Gamma^i(\mathrm{d}s, \mathrm{d}\theta), \quad t \geq 0, \quad i = 1, \cdots, d \\ \lambda_t^i = \phi_i \left(\mu_i + \int_{[0,t)} \left(\sum_{j=1}^d \delta_{ij}(t-s)\mathrm{d}N_s^j \right) + \sigma_i e^{-\omega_i(t-s)} \mathrm{d}W_s^i \right). \end{cases} \tag{1}$$

Remark

Knowing $A_t^i := \mu_i + \int_{[0,t)} \left(\sum_{j=1}^d \delta_{ij}(t-s)\mathrm{d}N_s^j \right) + \sigma_i e^{-\omega_i(t-s)} \mathrm{d}W_s^i$ or $\lambda_t^i = \phi_i(A_t^i)$ on $[0, T]$ is sufficient to construct N_t^i on the same interval.

Stability and stationarity

Does $\boldsymbol{N}$ have a (asymptotically) stationary distribution?
Yes if either of the following is true:

1. ϕ are bounded: in this case

$$\lambda_t^i = \phi_i\left(A_t^i\right) \leq \|\phi\|_\infty,$$

hence $\boldsymbol{N}$ is bounded by a Poisson process.

Stability and stationarity

Does $\boldsymbol{N}$ have a (asymptotically) stationary distribution?
Yes if either of the following is true:

1. ϕ are bounded: in this case

$$\lambda_t^i = \phi_i \left(A_t^i \right) \leq \|\phi\|_\infty,$$

   hence $\boldsymbol{N}$ is bounded by a Poisson process.

2. ϕ are Lipschitz and the matrix

$$\left(\|\delta_{ij}\|_1 \|\phi_i\|_L \right)_{i,j=1,\cdots,d}$$

   has a spectral radius strictly less than one. In this case

$$\lambda_t^i = \phi_i \left(A_t^i \right) \leq C + \|\phi_i\|_L |A_t^i|$$

Markov property

- In general, neither the Hawkes process nor its intensity are Markov processes.

Markov property

- In general, neither the Hawkes process nor its intensity are Markov processes.
- If $\delta_{ij}(t) = \sum_{q=0}^{p} \frac{\alpha_q^{ij} t^q}{q!} e^{-\beta_{ij} t}$ (Erlang function) set

$$v_t^{ij,q} := \int_0^{t-} \frac{(t-s)^q}{q!} e^{-\beta_{ij}(t-s)} \mathrm{d}H_s^j, q = 0, \cdots, p$$

and

$$z_t^i := \int_0^{t-} e^{-\omega_i(t-s)} \mathrm{d}W_s^i.$$

so that $\lambda_t^i = \phi_i\left(\tilde{a}^i(\boldsymbol{V}_t, \boldsymbol{Z}_t)\right) = \phi_i\left(\mu_i + \sigma_i z_t^i + \sum_{j=1}^{d} \langle \boldsymbol{\alpha}^{ij}, \boldsymbol{V}_t^{ij} \rangle\right)$.

Markov property

Proposition

The cascade $(\boldsymbol{V}, \boldsymbol{Z}) = \left((\boldsymbol{V}^{ij})_{i,j=1,\cdots,d}, (z^i)_{i=1,\cdots,d}\right)$ is a Markov process that follows the SDE:

$$\begin{cases} \mathrm{d}\boldsymbol{V}_t^{ij} & = (K - \beta_{ij} I)\boldsymbol{V}_t^{ij}\mathrm{d}t + \mathrm{d}N_t^j \boldsymbol{e}_{p+1}, \\ \mathrm{d}\boldsymbol{Z}_t & = -diag(\omega_1, \cdots, \omega_d)\boldsymbol{Z}_t\mathrm{d}t + \mathrm{d}\boldsymbol{W}_t \end{cases}$$

where $\boldsymbol{e}_{p+1} = (0, \cdots, 0, 1)$ and $\boldsymbol{W} = (W^1, \cdots, W^d)$.

Ideas for the proof:

- Multiply by $e^{\beta_{ij} t}$ and use Newton's binomial formula to get

$$v_t^{ij,q} e^{\beta_{ij} t} = \frac{1}{q!} \sum_{l=0}^{q} C_q^l t^l \int_0^t (-s)^{q-l} e^{\beta_{ij} s} \mathrm{d}N_s^j,$$

- Differentiate and use $lC_q^l = qC_{q-1}^{l-1}$.

The infinitesimal generator

Proposition

Let f be a $\mathcal{C}^1$ function in the first variable and $\mathcal{C}^2$ in the second variable. Then the infinitesimal generator $\mathcal{A}$ of the process $(\boldsymbol{V}, \boldsymbol{Z})$ at f is

$$\begin{aligned}\mathcal{A}f(\boldsymbol{v}, \boldsymbol{z}) = &\sum_{i=1}^{d} \frac{\partial^2 f}{\partial z_i^2}(\boldsymbol{v}, \boldsymbol{z}) - \langle diag(\omega_1, \cdots, \omega_d)\boldsymbol{z}, \nabla_{\boldsymbol{z}} f(\boldsymbol{v}, \boldsymbol{z})\rangle \\ &+ \sum_{i,j=1}^{d} \left\langle (K - \beta_{ij} I)\boldsymbol{v}_{ij}, \nabla_{\boldsymbol{v}_{ij}} f(\boldsymbol{v}, \boldsymbol{z})\right\rangle \\ &+ \sum_{m=1}^{d} \phi_m\left(\tilde{a}^m(\boldsymbol{v}, \boldsymbol{z})\right)\left(f\left(\left(\boldsymbol{v}_{ij} + \mathbb{1}_{j=m}\boldsymbol{e}_{p+1}\right)_{i,j=1,\cdots,d}, \boldsymbol{z}\right) - f\left(\boldsymbol{v}, \boldsymbol{z}\right)\right)\end{aligned}$$

where $\tilde{a}^m(\boldsymbol{v}, \boldsymbol{z}) = \mu_m + \sigma_m z_m + \sum_{j=1}^{d} \langle \boldsymbol{\alpha}^{mj}, \boldsymbol{v}_{mj}\rangle$.

Table of Contents

The space $\mathcal{D}$

- We call $\mathcal{D}$ the space of *càdlàg* functions on $\mathbb{R}_+$.
- The uniform metric is not adapted in this case. E.g: $x_n = \mathbb{1}_{[0,1/n)}$ converges to $x = \mathbb{1}_0$, yet $\|x_n - x\|_\infty = 1$.
- A metric that tolerates small movements in the t- axis is needed. In $\mathcal{D}[0,1]$ we define

$$d(x,y) = \inf_{\xi \in \Xi} \{\|\xi - I\|_\infty \vee \|x - y\xi\|_\infty\},$$

where Ξ is the set of increasing continuous functions that map $[0,1]$ to itself.

Back to the DTHP

- We recall the auxiliary series

$$\begin{cases} \boldsymbol{U}_{n+1}^{ij} & = M_{ij}^{h} \boldsymbol{U}_{n}^{ij} + X_{n}^{j} \boldsymbol{e}_{p+1} \\ \boldsymbol{Y}_{n+1} & = \Omega^{h} \boldsymbol{Y}_{n} + \sqrt{h} \mathcal{N}_{d} \end{cases}$$

  and the cumulative sum $\boldsymbol{H}_n = \sum_{k=0}^{n} \boldsymbol{X}_k$.
- Define $(\tilde{\boldsymbol{V}}_t^h, \tilde{\boldsymbol{Z}}_t^h, \tilde{\boldsymbol{N}}_t^h) = (\boldsymbol{U}_{\lfloor t/h \rfloor}, \boldsymbol{Y}_{\lfloor t/h \rfloor}, \boldsymbol{H}_{\lfloor t/h \rfloor})$.
- We are interested in its behaviour when the time step h goes to zero.

Convergence of the generators

Proposition

Let $\mathcal{T}_h f(\boldsymbol{v}, \boldsymbol{z}) = \mathbb{E}[f(\boldsymbol{U}_{n+1}, \boldsymbol{Y}_{n+1}) | (\boldsymbol{U}_n, \boldsymbol{Y}_n) = (\boldsymbol{v}, \boldsymbol{z})]$ be the transition operator of the DTHP and define

$$\mathcal{A}_h f = \frac{\mathcal{T}_h f - I}{h}.$$

If the jump-rate ϕ is <u>bounded</u>, then

$$\|\mathcal{A}_h f - \mathcal{A} f\|_\infty \longrightarrow 0,$$

when h goes to zero, for any f in the Schwartz space $\mathcal{S}$ of functions.

Proof sketch: We prove that $R_h = O(h^2)$ then we proceed carefully with Taylor expansions.

Main result

Theorem

The process $(\tilde{\boldsymbol{V}}^h_t, \tilde{\boldsymbol{Z}}^h_t, \tilde{\boldsymbol{N}}^h_t)_{t\in\mathbb{R}_+}$ converges weakly in the Skorokhod topology to the continuous time Hawkes process $(\boldsymbol{V}_t, \boldsymbol{Z}_t, \boldsymbol{N}_t)_{t\in\mathbb{R}_+}$ when the time step h goes to zero.

Proof sketch:

- We prove that $\mathcal{T}(t)f(\boldsymbol{v},\boldsymbol{z}) = \mathbb{E}\left[f(\boldsymbol{V}_t, \boldsymbol{Z}_t)|(\boldsymbol{V}_0, \boldsymbol{Z}_0) = (\boldsymbol{v}, \boldsymbol{z})\right]$ is a Feller semigroup.
- We prove that $\mathcal{S}$ is a core for $\mathcal{T}$ (dense and stable).
- Using the previous result $\mathcal{A}_h f \longrightarrow \mathcal{A}f, \forall f \in \mathcal{S}$, we conclude using Theorem 2.6 and 6.5 from EK.

Simulation $h = 0.125$

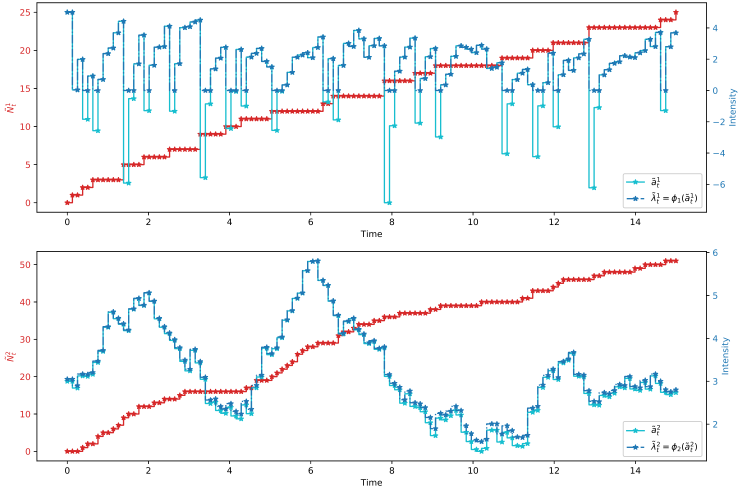

Figure 7: $\phi_1(x) = (x)_+ \wedge 40, \quad \phi_2(x) = \log(1 + e^x) \wedge 40$

Simulation $h = 5 \cdot 10^{-3}$

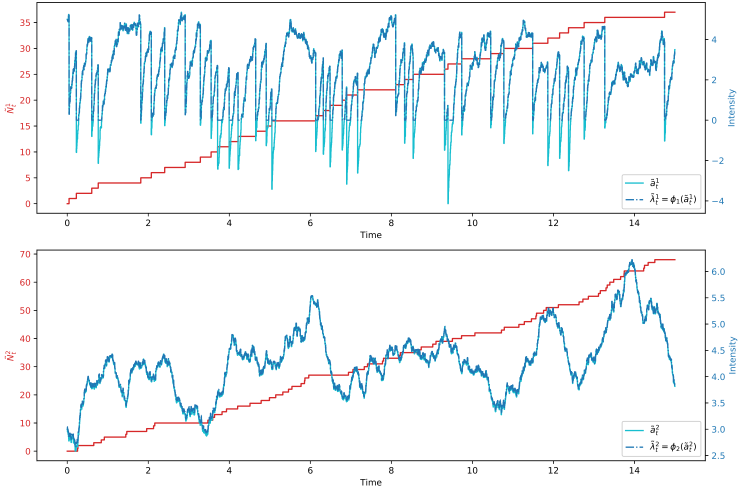

Figure 8: Indistinguishable from a perturbed continuous-time Hawkes process

Table of Contents

Conditional least squares

- We suggest estimating the parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu})$ by minimizing

$$
\begin{aligned}
Q_M &= \sum_{k=1}^{M} \left\| \boldsymbol{X}_k - \mathbb{E}\left[\boldsymbol{X}_k | \mathcal{F}_{k-1}\right] \right\|_2^2 , \\
&= \sum_{k=1}^{M} \sum_{i=1}^{d} \left(X_k^i - h\phi_i \left(\mu_i + \sum_{j=1}^{d} \langle \boldsymbol{\alpha}^{ij}, \boldsymbol{U}_n^{ij} \rangle \right) \right)
\end{aligned}
$$

Conditional least squares

- We suggest estimating the parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu})$ by minimizing

$$Q_M = \sum_{k=1}^{M} \left\| \boldsymbol{X}_k - \mathbb{E}\left[\boldsymbol{X}_k | \mathcal{F}_{k-1}\right] \right\|_2^2,$$
$$= \sum_{k=1}^{M} \sum_{i=1}^{d} \left(X_k^i - h\phi_i \left(\mu_i + \sum_{j=1}^{d} \langle \boldsymbol{\alpha}^{ij}, \boldsymbol{U}_n^{ij} \rangle \right) \right)$$

- We do not seek to estimate the Gaussian perturbation parameters.

Conditional least squares

- We suggest estimating the parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu})$ by minimizing

$$
\begin{aligned}
Q_M &= \sum_{k=1}^{M} \| \boldsymbol{X}_k - \mathbb{E}\left[\boldsymbol{X}_k | \mathcal{F}_{k-1}\right] \|_2^2 , \\
&= \sum_{k=1}^{M} \sum_{i=1}^{d} \left(X_k^i - h\phi_i \left(\mu_i + \sum_{j=1}^{d} \langle \boldsymbol{\alpha}^{ij}, \boldsymbol{U}_n^{ij} \rangle \right) \right)
\end{aligned}
$$

- We do not seek to estimate the Gaussian perturbation parameters.
- We assume that ϕ is known and differentiable.

Conditional least squares

- We suggest estimating the parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu})$ by minimizing

$$Q_M = \sum_{k=1}^{M} \|\boldsymbol{X}_k - \mathbb{E}[\boldsymbol{X}_k|\mathcal{F}_{k-1}]\|_2^2,$$
$$= \sum_{k=1}^{M}\sum_{i=1}^{d}\left(X_k^i - h\phi_i\left(\mu_i + \sum_{j=1}^{d}\langle\boldsymbol{\alpha}^{ij}, \boldsymbol{U}_n^{ij}\rangle\right)\right)$$

- We do not seek to estimate the Gaussian perturbation parameters.
- We assume that ϕ is known and differentiable.
- The minimization of Q_M is done using the `SLSQP` method of the function `scipy.optimize.minimize` with the Lipschitz assumption as a constraint.

Simulation in the multi-variate setting

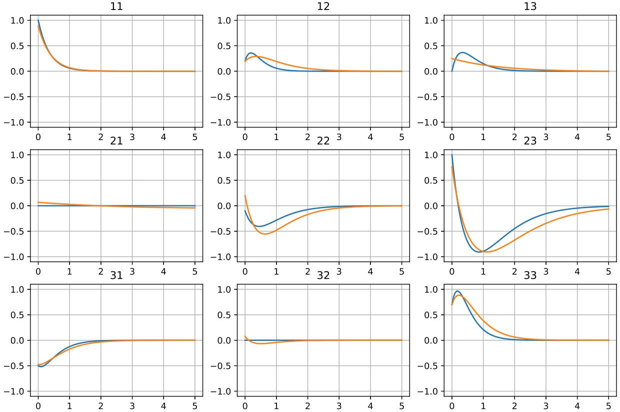

Figure 9: In blue: ground truth kernels. In orange: estimated kernels.

Regression with a lower degree

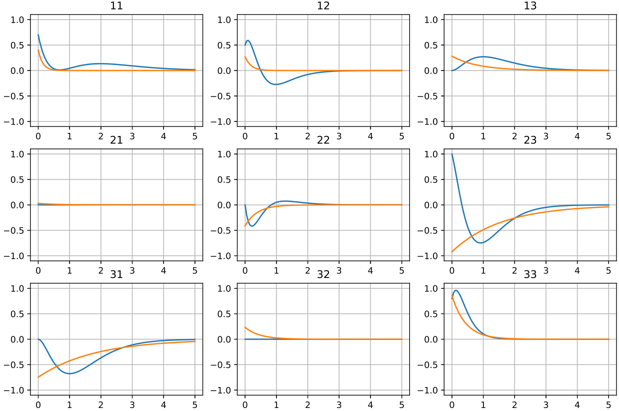

Figure 10: The overall effect (excitation, inhibition, independence) are well captured by the exponential kernels.

Regression for the baseline

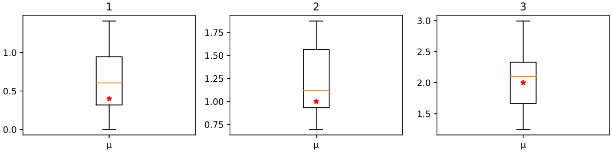

Figure 11: The estimator tends to estimate larger values for the baseline intensities. This is due to a discretization bias, where some self-excitations are ignored

Thank you for your attention.