Inference of non-linear or imperfectly observed Hawkes processes

Miguel Martinez Herrera

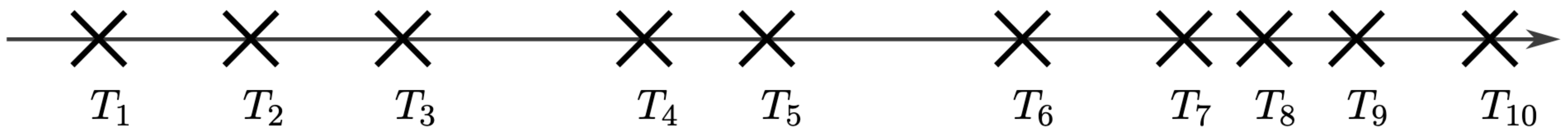
Supervised by Anna Bonnet, Arnaud Guyader and Maxime Sangnier

SORBONNE CENTER FOR ARTIFICIAL INTELLIGENCE

SORBONNE UNIVERSITÉ

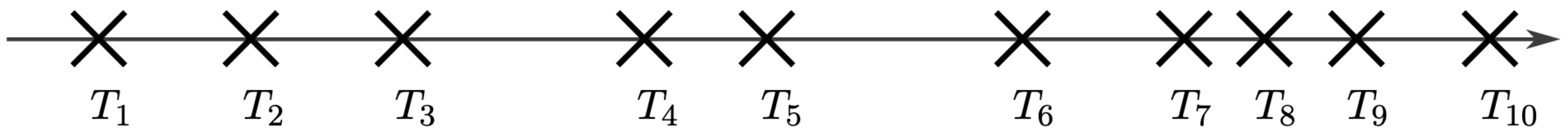
Motivation

- **Goal:** analyse temporal data with dependency on the past through the model of **Hawkes processes.**



Motivation

- **Goal:** analyse temporal data with dependency on the past through the model of **Hawkes processes.**



Earthquake occurrences

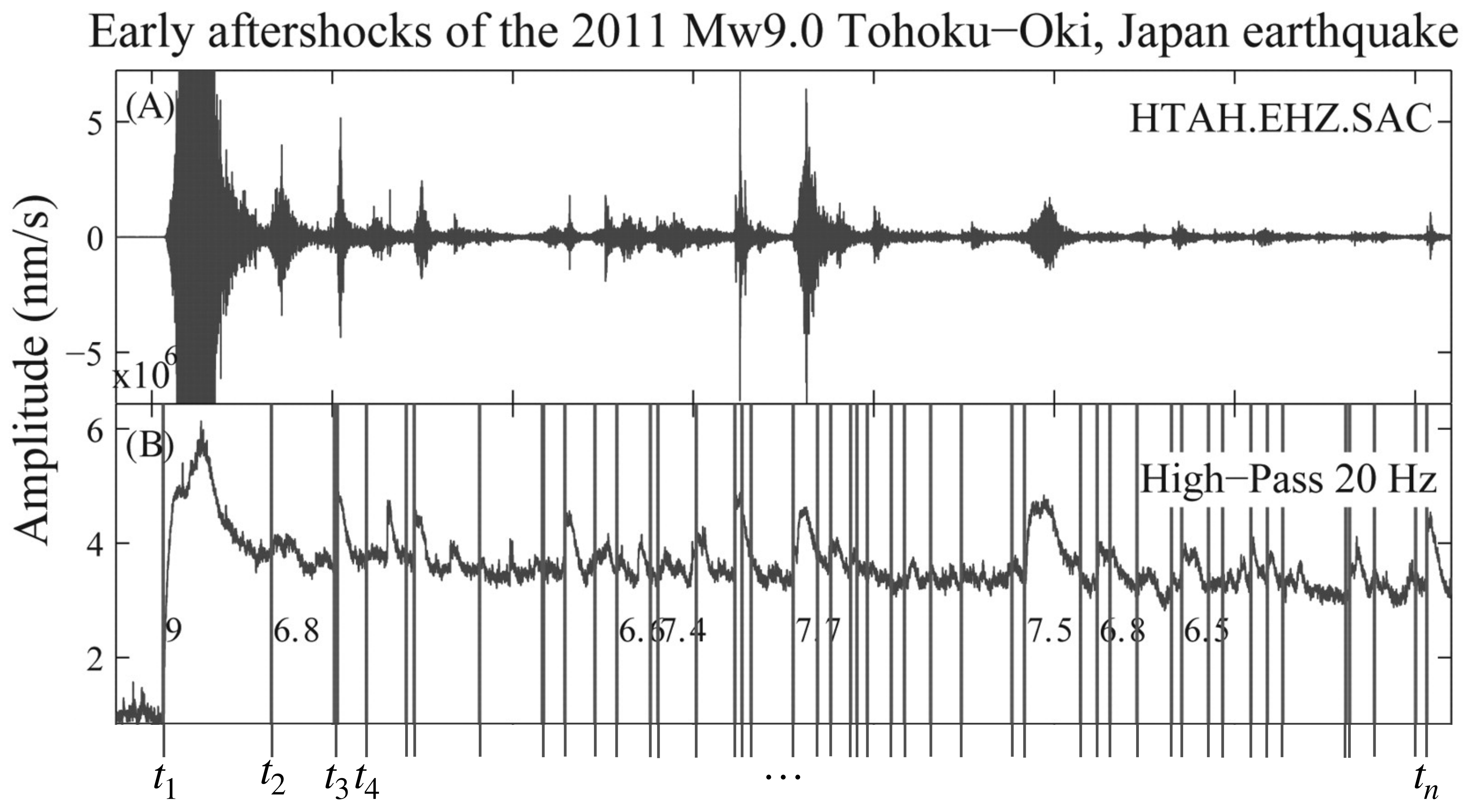
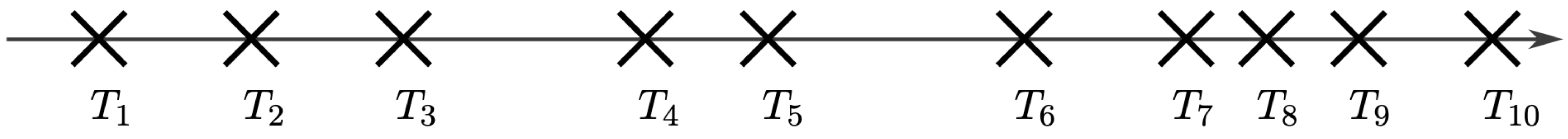


Image from Peng et al. (2012)

Motivation

- **Goal:** analyse temporal data with dependency on the past through the model of **Hawkes processes.**



Earthquake occurrences

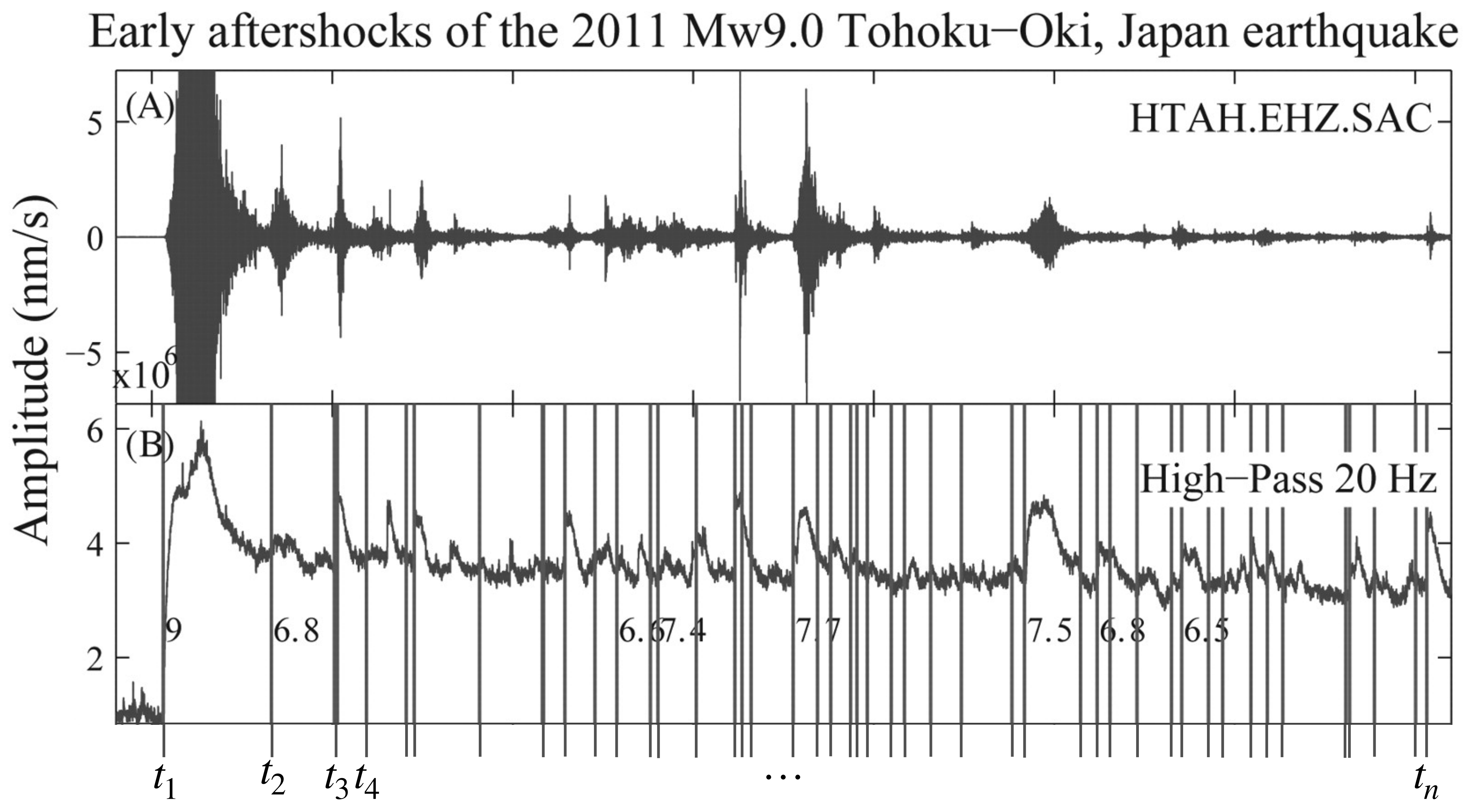


Image from Peng et al. (2012)

Synaptic signalling in neurons

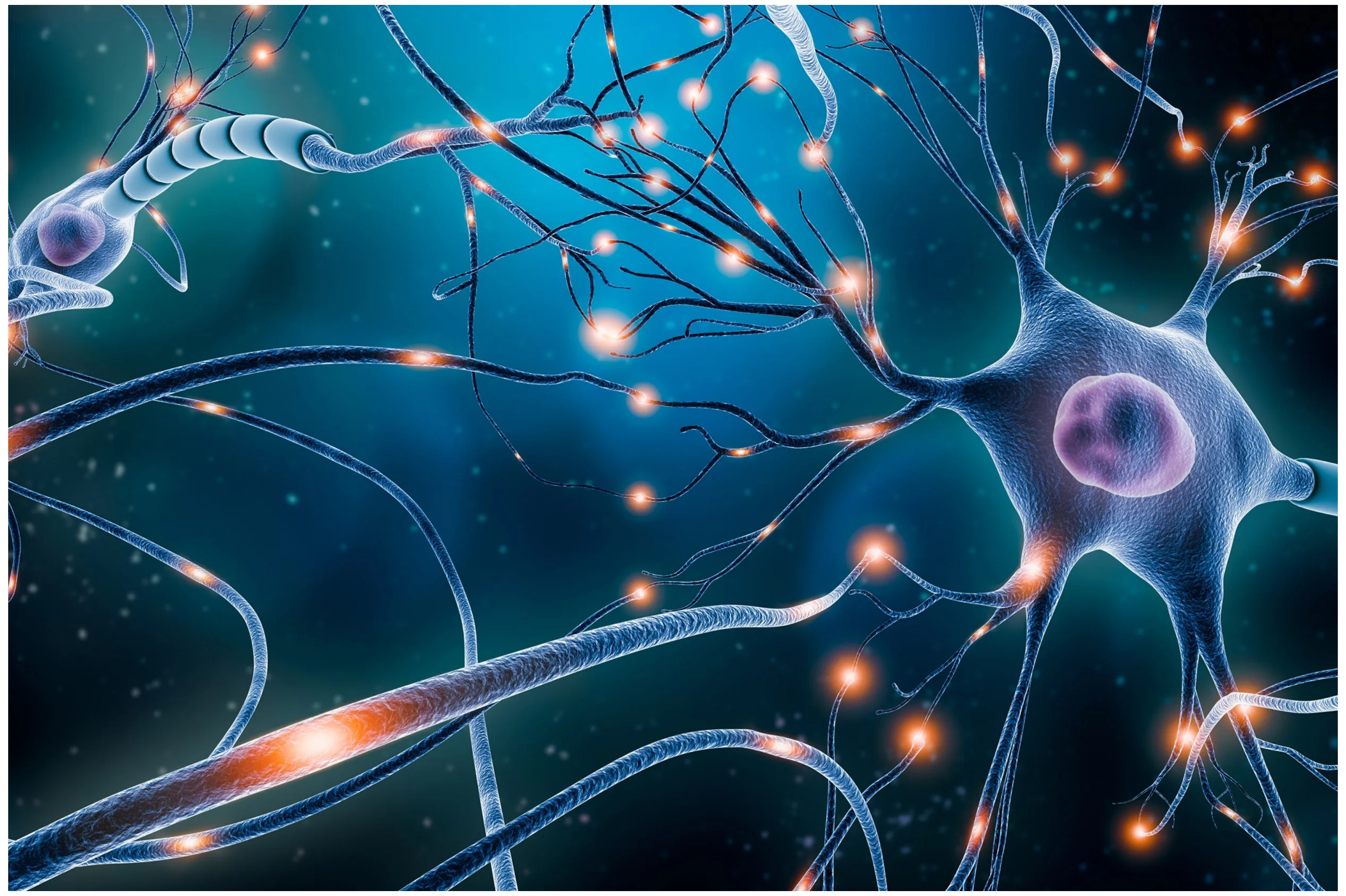
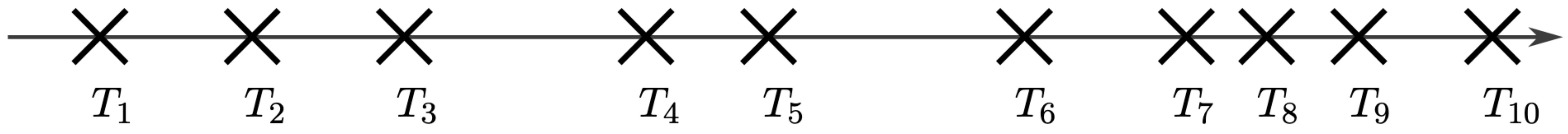


Image from The Harvard Gazette

Point process and conditional intensity function

- Let N be a point process in the real line $\mathbb{R}$ with event times $(T_k)_{k\in\mathbb{Z}}$.
  Let $\mathscr{H}_t = \sigma(T_k, T_k \leq t)$ be the history of N up to time $t \in \mathbb{R}$.



- For any Borel set $B \in \mathscr{B}_{\mathbb{R}}$, $N(B) = \sum_{k\in\mathbb{Z}} \mathbf{1}_{T_k\in B}$ represents the number of points in B.

- The conditional intensity function $\lambda : \mathbb{R} \to \mathbb{R}_{\geq 0}$ of process N is defined as:

$$\lambda(t \mid \mathscr{H}_t) = \lim_{h\to 0} \frac{\mathbb{E}[N([t, t+h)) \mid \mathscr{H}_t]}{h} .$$

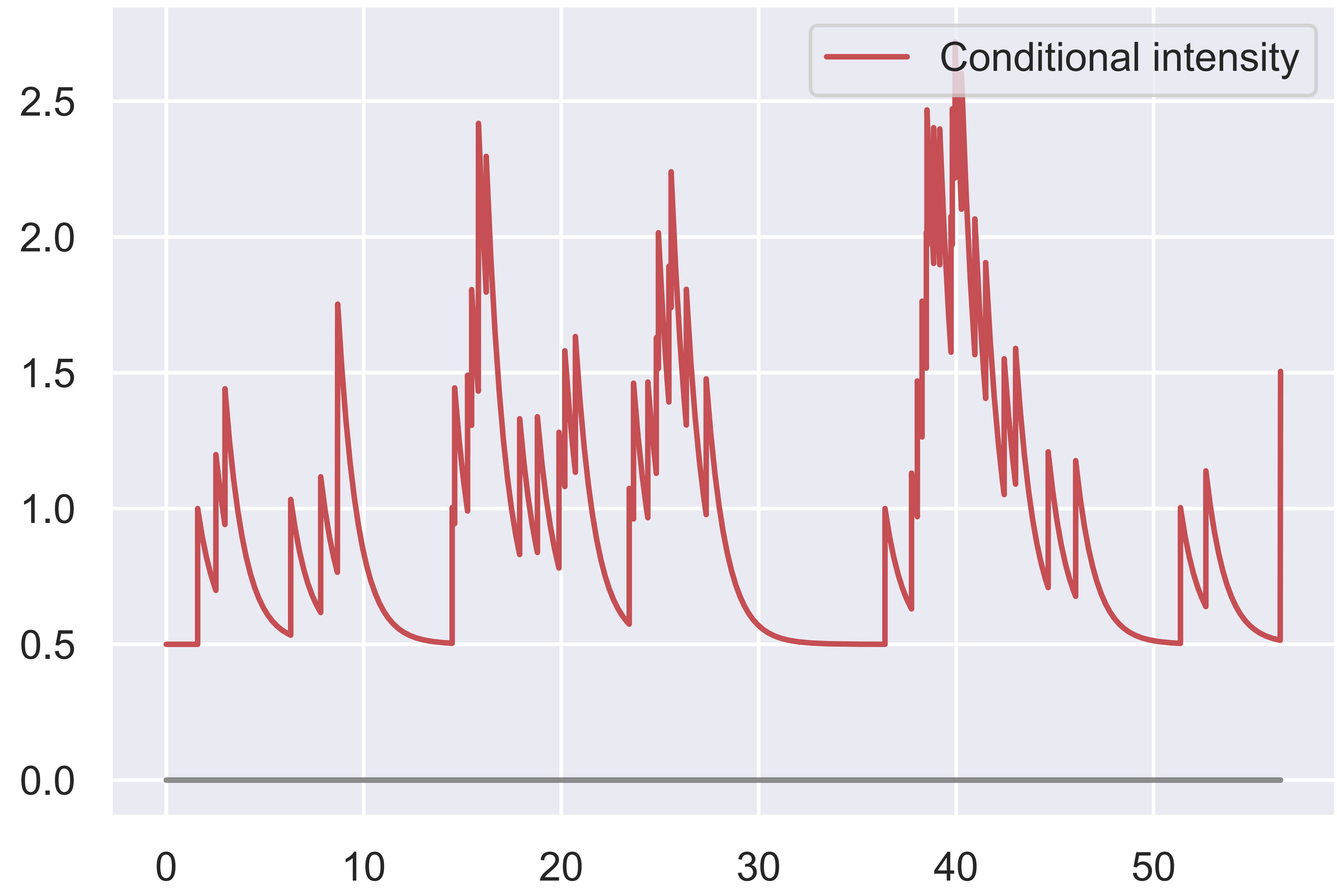
  Intuitively, it quantifies the probability of observing an event at time t.

Hawkes process

- A **linear Hawkes process** H is a univariate point process defined by the conditional intensity function (Hawkes 1971):

$$\lambda(t \mid \mathscr{H}_t) = \mu + \int_{-\infty}^{t} h(t-s)\, N(\mathrm{d}s) = \mu + \sum_{T_k \le t} h(t - T_k)\,,$$

with baseline intensity $\mu > 0$ and kernel function $h\colon \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$ such that $\int_{\mathbb{R}} h(t)\,\mathrm{d}t < 1$.



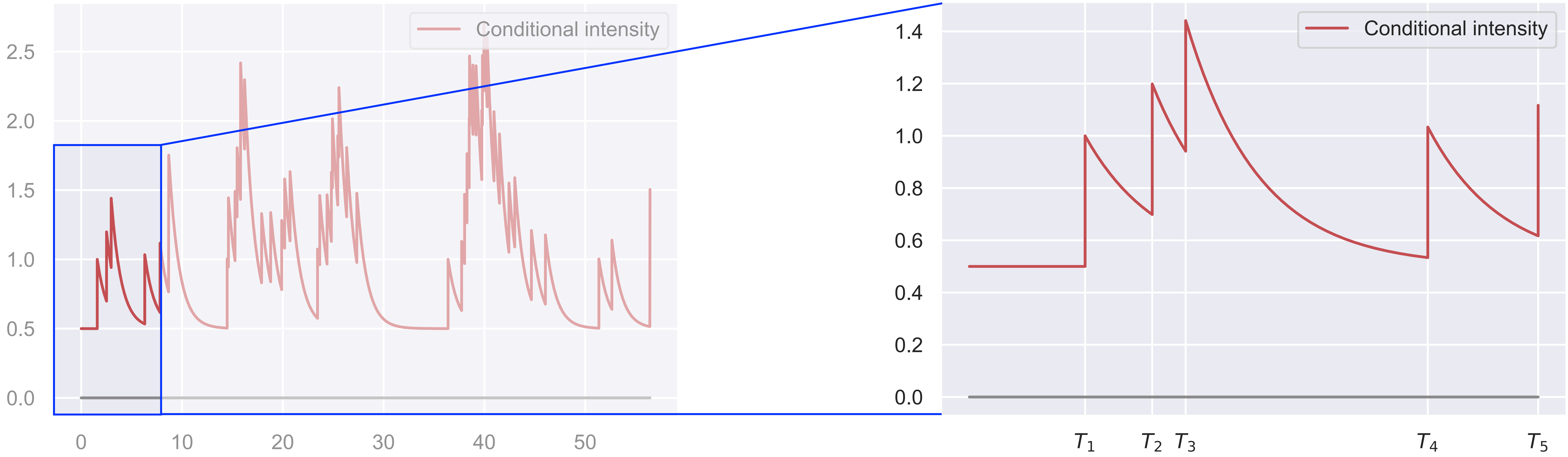
Intensity function of a self-exciting Hawkes process

Hawkes process

- A **linear Hawkes process** H is a univariate point process defined by the conditional intensity function (Hawkes 1971):

$$\lambda(t \mid \mathscr{H}_t) = \mu + \int_{-\infty}^{t} h(t-s)\, N(\mathrm{d}s) = \mu + \sum_{T_k \leq t} h(t - T_k)\,,$$

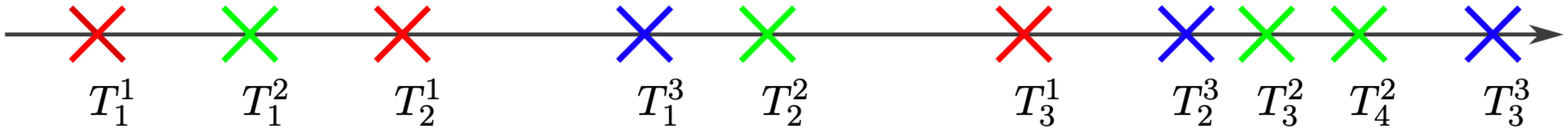
with baseline intensity $\mu > 0$ and kernel function $h\colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ such that $\int_{\mathbb{R}} h(t)\,\mathrm{d}t < 1$.



Intensity function of a self-exciting Hawkes process

Multivariate Hawkes process

- A **multivariate Hawkes process** $H = (H_1, \ldots, H_d)$ with dimension d is a collection of d univariate point processes with respective event times $(T^i_k)_k$,



with intensity functions:

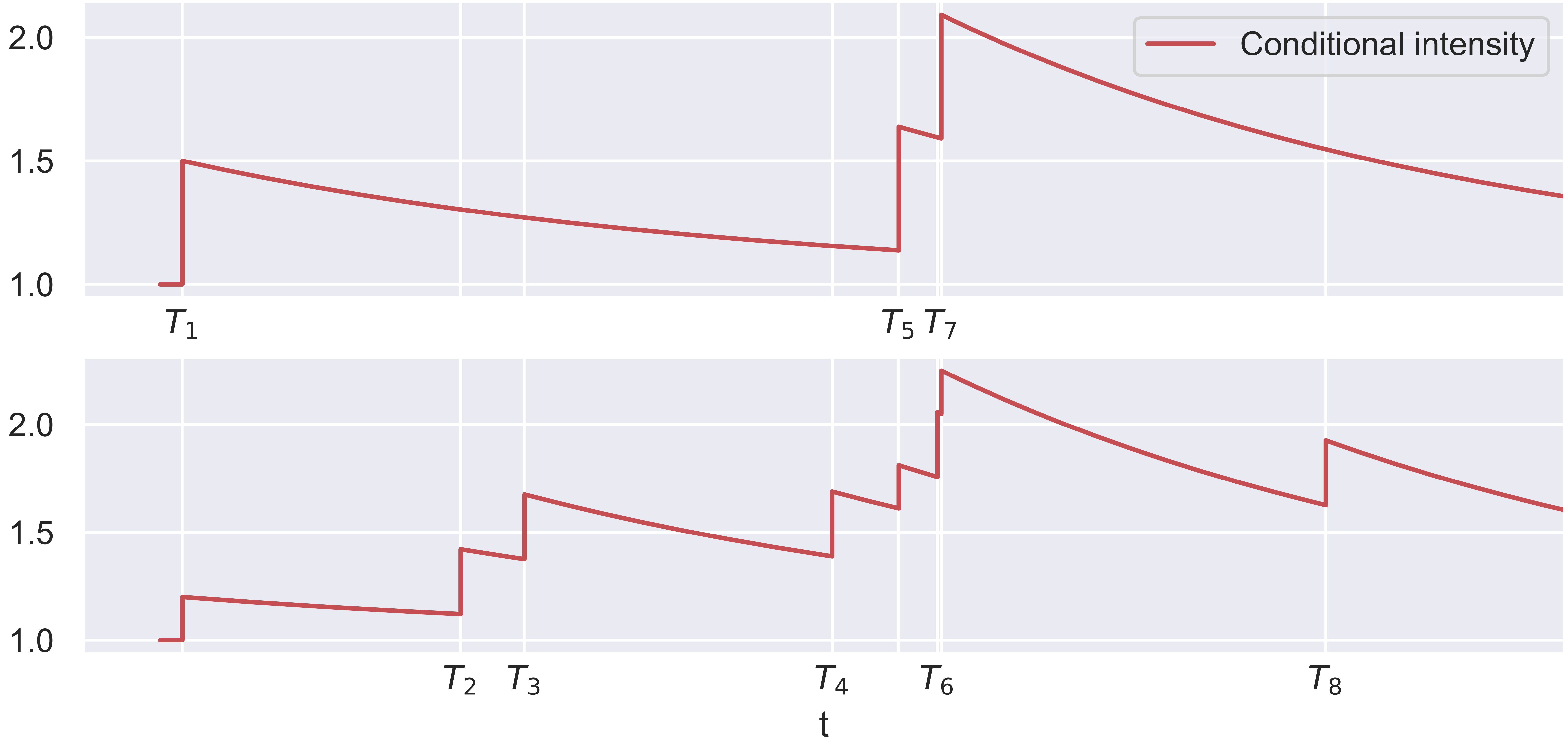
$$\lambda^i(t) = \mu_i + \sum_{j=1}^{d} \sum_{T^j_k \leq t} h_{ij}(t - T^j_k)\,,$$

where $\mu^i > 0$ and $h_{ij} \colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$. The ordered union of events $(T^i_k)_k$ form the events of H noted $(T_{(k)})_{k \in \mathbb{Z}}$.

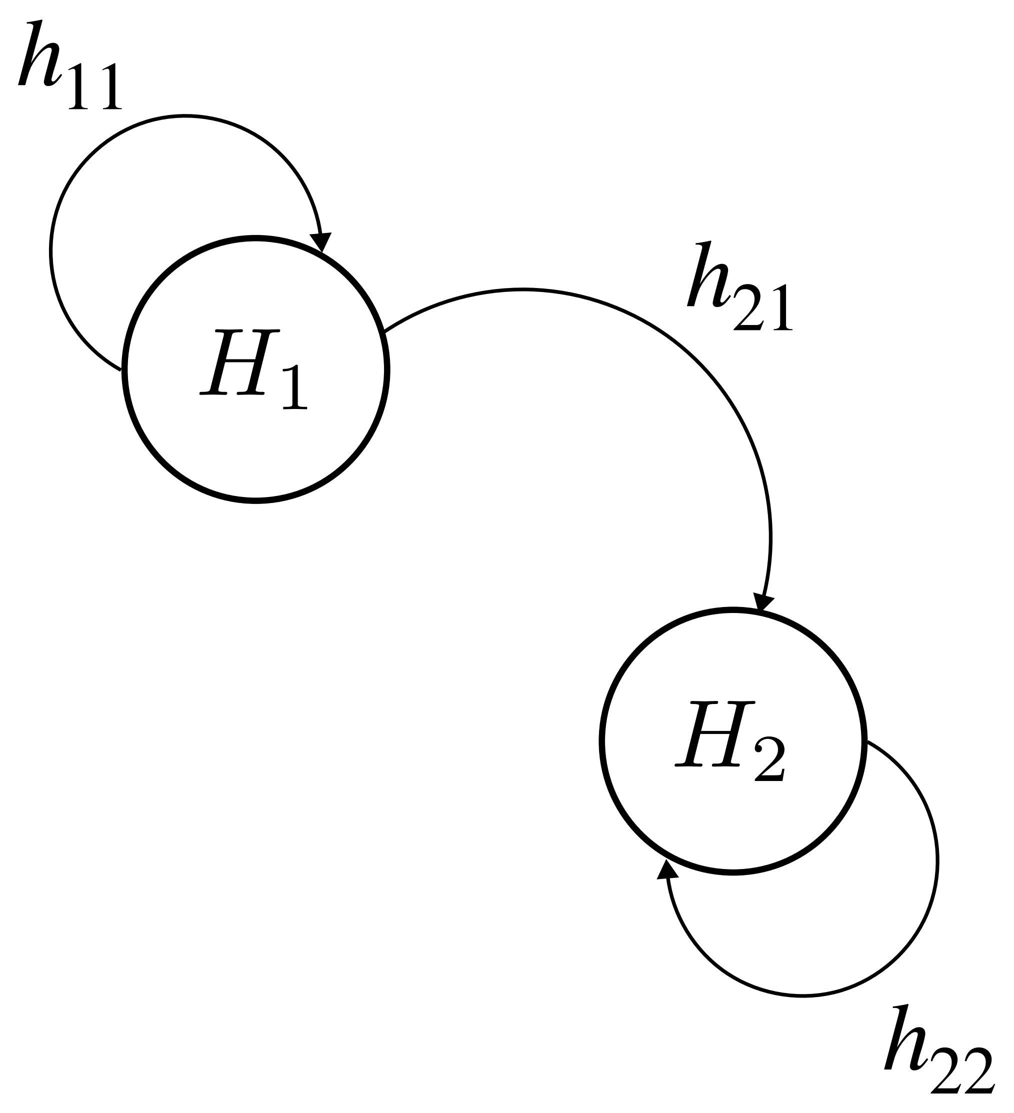
- Let $\|h_{ij}\|_1 = \int_{\mathbb{R}} |\,h_{ij}(t)\,|\,\mathrm{d}t$ and $S = (\|h_{ij}\|_1)_{ij}$, then process H is a **stationary** point process if $\rho(S) < 1$ (Brémaud et al. 1996).

Bivariate Hawkes process

- h_{ij} encodes the influence of process N_j on process N_i.
  In particular, $h_{ij} = 0$ represents an absence of interaction.



Intensity functions of a bivariate Hawkes process



Interaction graph

Statistical framework

- We define a **parametric model** of a Hawkes process (where h is parametrised by a vector γ):

$$\mathcal{Q} = \left\{\lambda_\theta \colon \mathbb{R} \to \mathbb{R}_{\geq 0}, \quad \theta = (\mu, \gamma) \in \Theta\right\} .$$

- Let $(T_k)_{1 \leq k \leq N([0,T])}$ be an observation of a Hawkes process in a time window $[0, T]$.

- **Goal:** propose an inference estimation method to account for two scenarios inspired by neuronal data.

First axis: factoring in inhibition for Hawkes processes.

Second axis: spectral methods for imperfect data

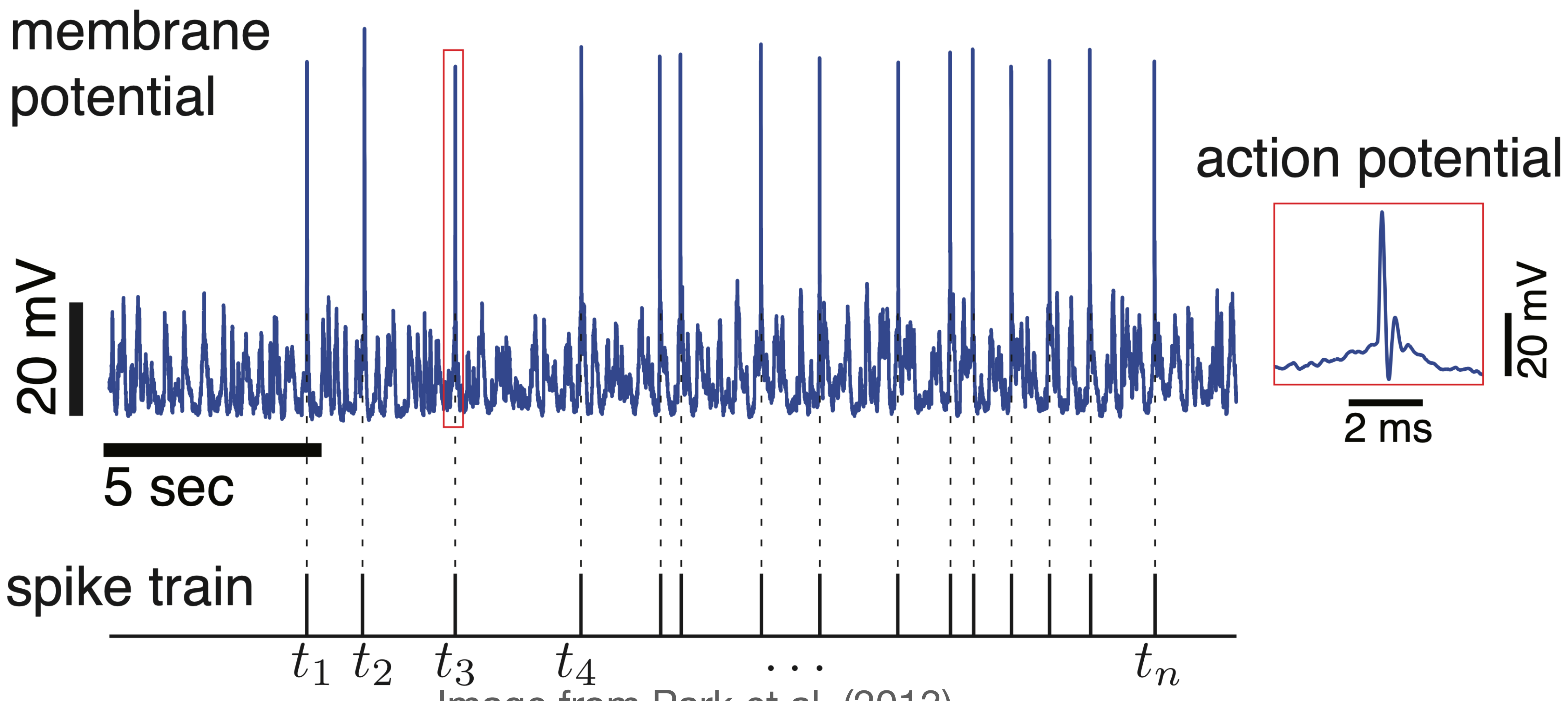


Image from Park et al. (2013)

First axis:

Factoring in inhibition for Hawkes processes

How to model inhibition ?

- **Inhibition** is the opposite effect of excitation → lowering the chances of further events occurring.
- Additive inhibition = allowing h_{ij} to be a signed function (take negative values).
- **Problem:** the intensity functions λ^i has to be non-negative!
- We work then with the **non-linear Hawkes process** (Brémaud et al. 1996) defined by the intensity functions:

$$\lambda^i(t) = \Phi\left(\mu_i + \sum_{j=1}^{d} \sum_{T_k^j \leq t} h_{ij}(t - T_k^j)\right),$$

where $\Phi : \mathbb{R} \to \mathbb{R}_{\geq 0}$ is an L-lipschitz function (for $0 < L < \infty$) such that $\rho(S^+) < 1$, with $S^+ = (L\|h_{ij}^+\|_1)_{ij}$ (Sulem et al. 2024).

Literature review

- **What is in the literature for estimating Hawkes models with additive inhibition?**

  Frequentist settings:

  - Only done through **approximations** in parametric frequentist settings as in Lemonnier et al. (2014) and Bacry et al. (2020), in non-parametric as in Reynaud-Bouret et al. (2014) and Bacry et al. (2016):

$$\lambda^i(t) \approx \mu_i + \sum_{j=1}^{d} \sum_{T_k^j \leq t} h_{ij}(t - T_k^j) \,.$$

  Bayesian settings:

  - Parametric estimation as in Deutsch et al. (2024) with time-varying baselines and efficient prior choice.

  - Non-parametric estimation as in Sulem et al. (2024) for finite-memory kernels.

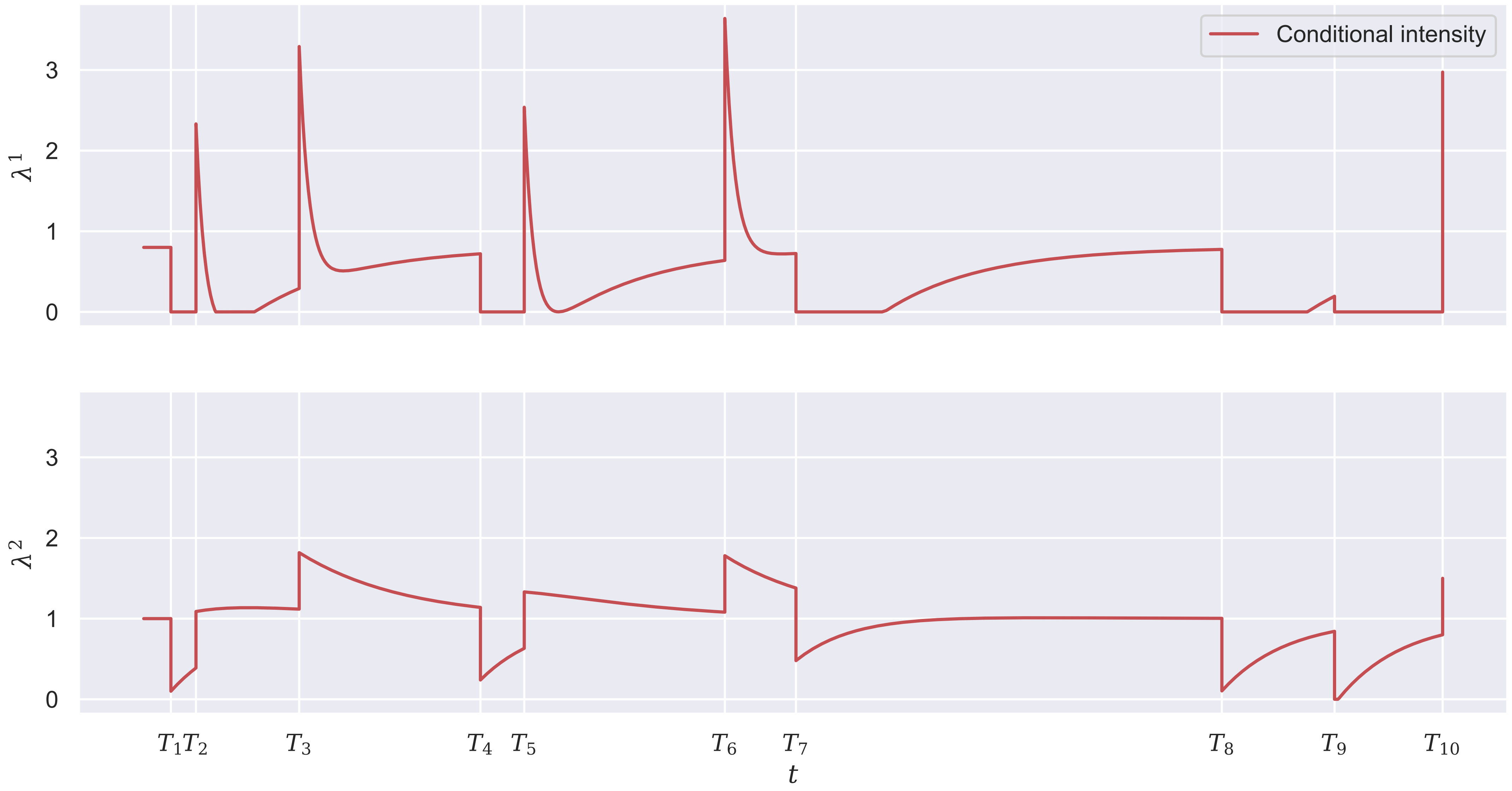
- Missing works in **frequentist parametric** frameworks.

The multivariate non-linear Hawkes process

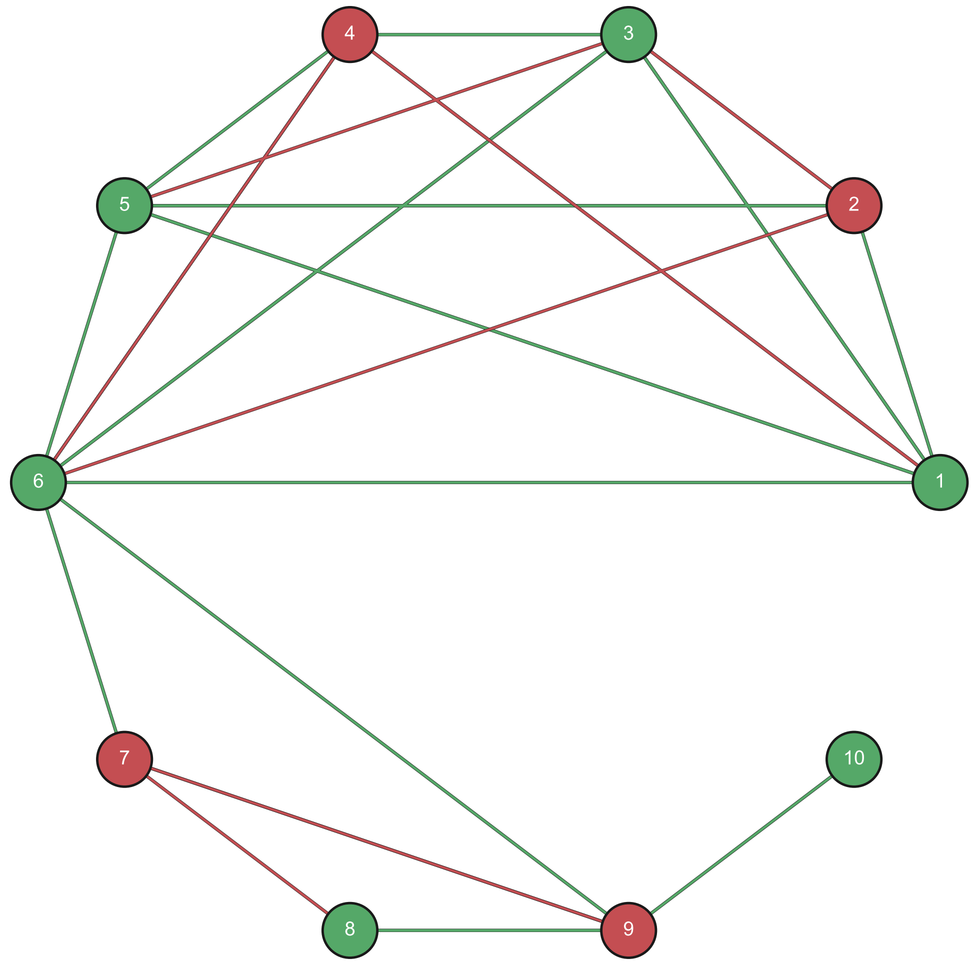
- Let $\Phi(\cdot) = (\cdot)^+ = \max(0, \cdot)$ be the positive part function. For any integer i, the intensity function λ^i of process H^i reads:

$$\lambda^i(t) = \left(\mu_i + \sum_{j=1}^{d} \sum_{T_k^j \le t} h_{ij}(t - T_k^j) \right)^+ .$$

- **Advantage:** if $h_{ij} \ge 0$, for all integers i, j, we retrieve the same intensity function of a linear Hawkes process.



Intensity function of a bivariate exponential Hawkes process



Example of interaction graph of a 10 dimensional Hawkes process

Estimation procedure

- **Goal:** implement the Maximum Likelihood Estimation (MLE) procedure.

- We define the parametric model:

$$\mathcal{Q} = \left\{\lambda_\theta^i \colon \mathbb{R} \to \mathbb{R}_{\geq 0}, \quad \theta = (\mu, \gamma) \in \Theta\right\} .$$

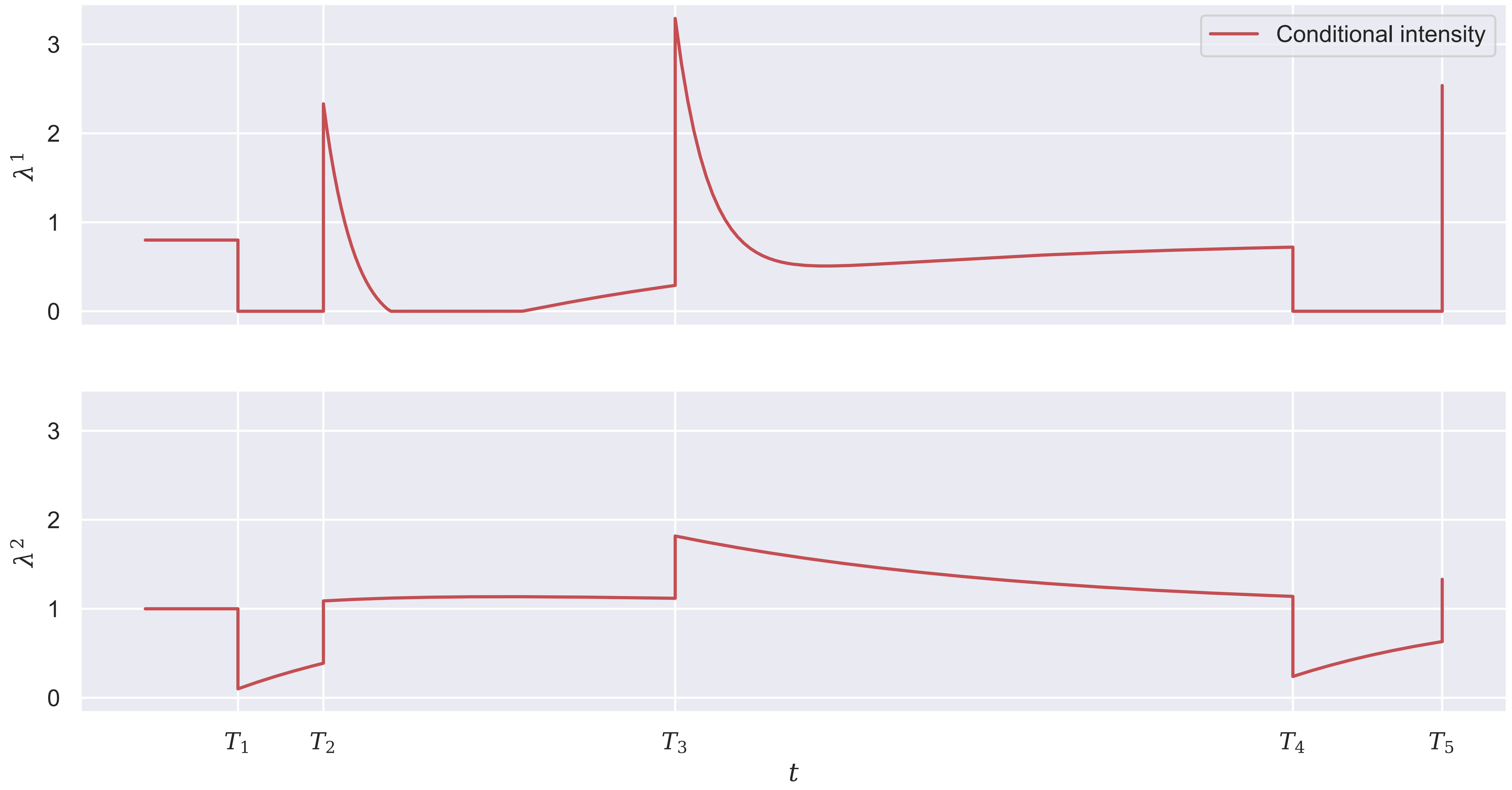
- For an observation of $H = (H_1, \ldots, H_d)$ in the time window $[0, T]$, the log-likelihood of multivariate point process reads:

$$\ell_T(\theta) = \sum_{i=1}^{d} \ell_T^i(\theta) = \sum_{i=1}^{d} \left(\sum_{k=1}^{N^i([0,T])} \log \lambda_\theta^i(T_k^{i-}) - \Lambda_\theta^i(T) \right), \qquad \text{with } \Lambda_\theta^i(T) = \int_0^T \lambda_\theta^i(t)\,\mathrm{d}t\,.$$

How can we compute exactly the log-likelihood when inhibition is present?

Challenge

- **Challenge:** to compute the compensator Λ^i, we need to integrate the function λ^i in the intervals where the intensity is positive. As we can see, even in the intervals $[T_{(k)}, T_{(k+1)})$, the functions are not easy to study (not smooth even with smooth kernels).



Challenge

- We introduce the concept of the **underlying intensity function** $\lambda^{i\star}$ such that $\lambda^i = \Phi \circ \lambda^{i\star}$. For any $t \in \mathbb{R}$:

$$\lambda^{i\star}(t) = \mu_i + \sum_{j=1}^{d} \sum_{T_k^j \leq t} h_{ij}(t - T_k^j) \,,$$

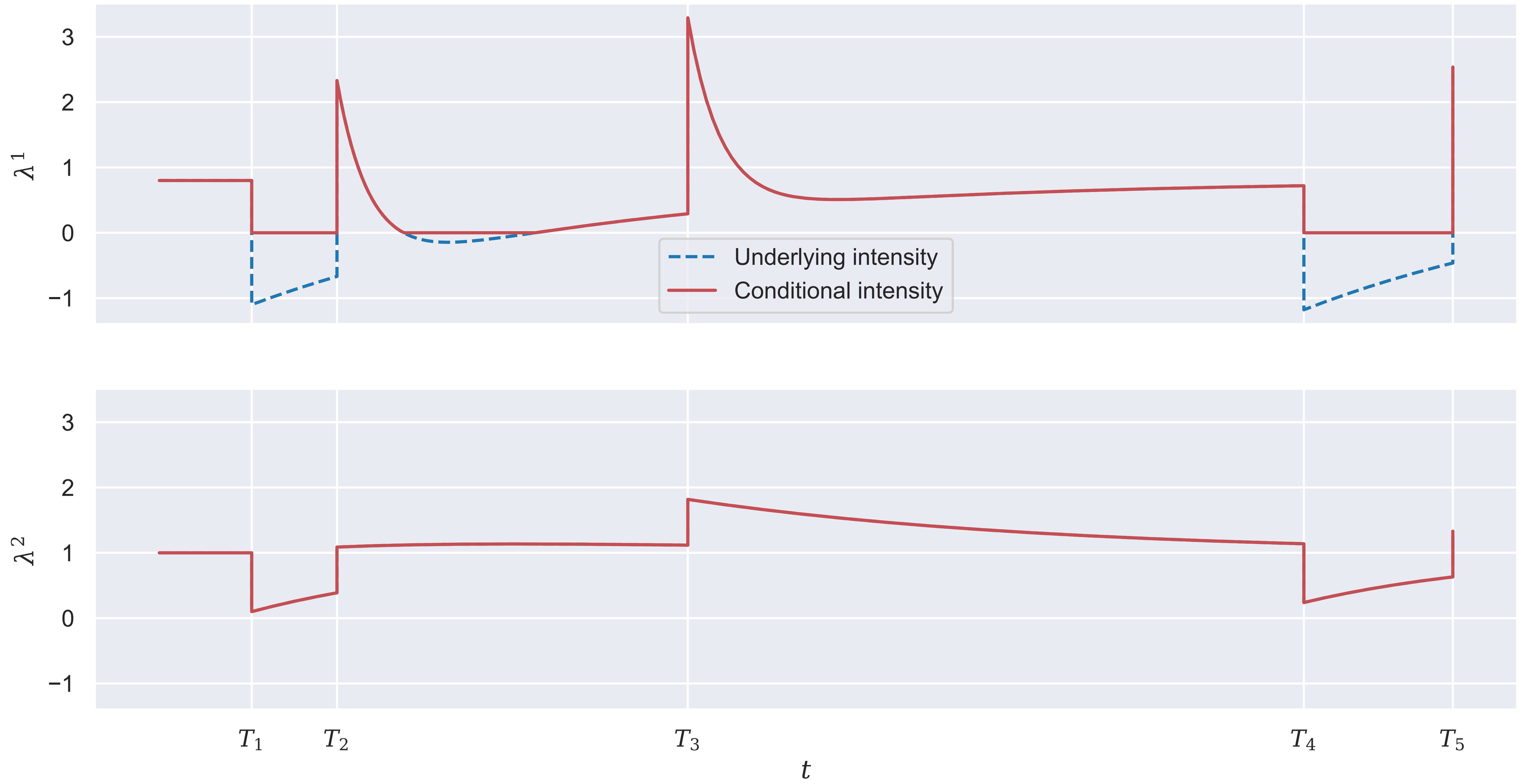
- **Advantage:** under certain conditions, the function $\lambda^{i\star}$ is piecewise smooth in the intervals $[T_{(k)}, T_{(k+1)})$.

Challenge

- We introduce the concept of the **underlying intensity function** $\lambda^{i\star}$ such that $\lambda^i = \Phi \circ \lambda^{i\star}$. For any $t \in \mathbb{R}$:

$$\lambda^{i\star}(t) = \mu_i + \sum_{j=1}^{d} \sum_{T_k^j \leq t} h_{ij}(t - T_k^j)\,,$$

- **Advantage:** under certain conditions, the function $\lambda^{i\star}$ is piecewise smooth in the intervals $[T_{(k)}, T_{(k+1)})$.



The univariate self-inhibiting Hawkes process

- Let us begin by working in the univariate case $d = 1$. The underlying intensity function reads:

$$\lambda^\star(t) = \mu + \sum_{T_k \le t} h(t - T_k)\,.$$

Conditional and underlying intensity functions of a univariate exponential Hawkes process

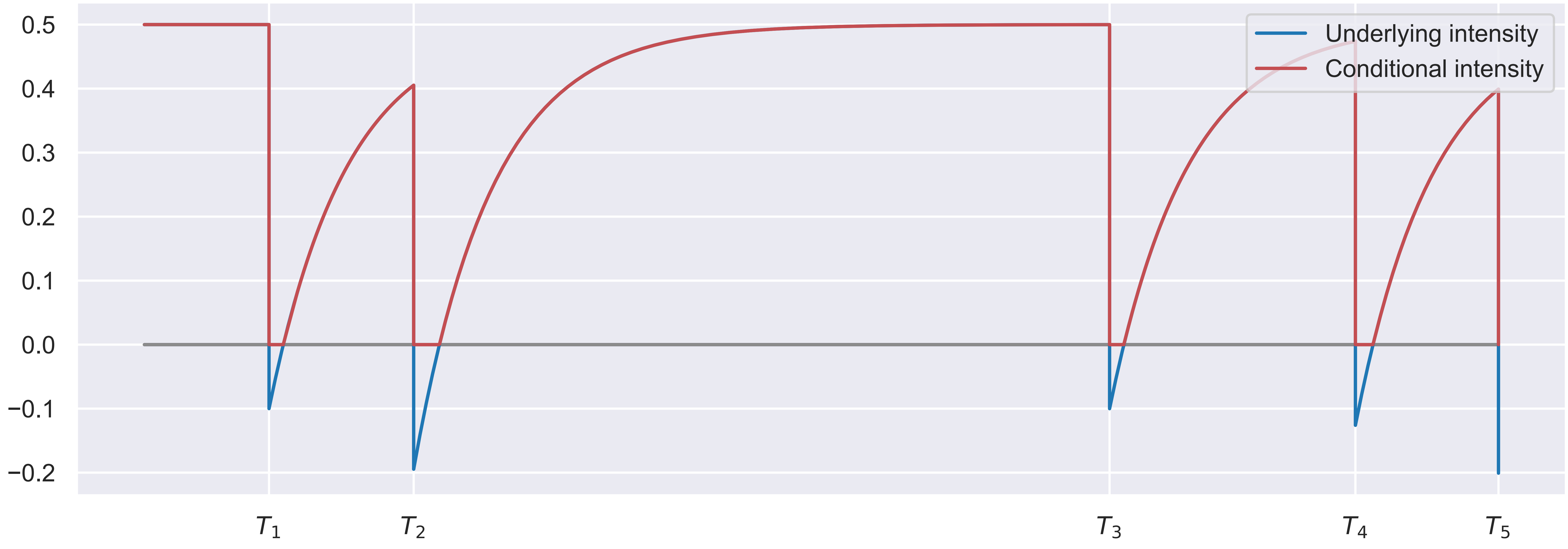
The univariate self-inhibiting Hawkes process

- Let us begin by working in the univariate case $d = 1$. The underlying intensity function reads:

$$\lambda^\star(t) = \mu + \sum_{T_k \le t} h(t - T_k)\,.$$

- We introduce the **restart times** $(T_k^\star)_{k\in\mathbb{Z}}$:

$$T_k^\star = \inf\{t \ge T_k : \lambda^\star(t) \ge 0\}\,.$$



Conditional and underlying intensity functions of a univariate exponential Hawkes process

Compensator

Lemma

If h is a monotonous function, then, for any integer $k \in \mathbb{Z}$, the function $\lambda^\star : [T_k, T_{k+1}) \to \mathbb{R}$ is monotonous and $T_k^\star$ is the only solution to:

$$\lambda^\star(t) = 0, \text{ for } t \in [T_k, T_{k+1}).$$

Furthermore the compensator Λ of H can be expressed as:

$$\Lambda(t) = \begin{cases} \mu t & \text{if } t < T_1 \\ \mu T_1 + \sum_{k=1}^{N([0,t])-1} \int_{T_k^\star}^{T_{k+1}} \lambda^\star(u)\,\mathrm{d}u + \int_{T^\star_{N([0,t])}}^{t} \lambda^\star(u)\,\mathrm{d}u & \text{if } t \geq T_1\,. \end{cases}$$

- Closed-form expression of the compensator → **closed-form expression of the log-likelihood.**
- We can implement the MLE method of estimation with the **same complexity** as in the purely self-exciting scenario.

Exponential kernel

- In the case of an **exponential kernel function** $h(t) = \alpha e^{-\beta t}$, for any $t > 0$, the same method as in Ozaki (1979) can be implemented:

Proposition

Let us assume that h is the exponential kernel function.

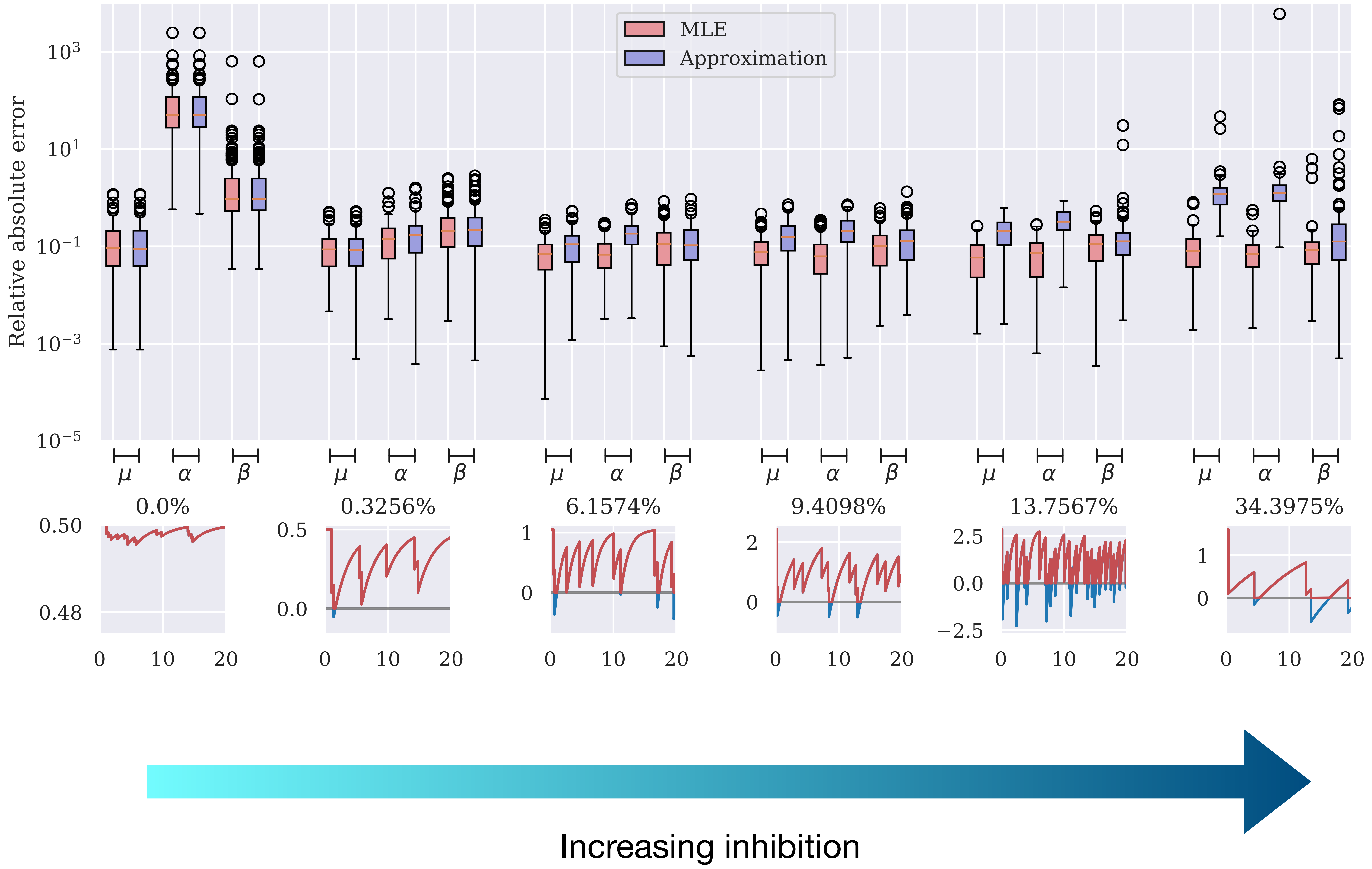
For any integer k, the restart times $T_k^\star$ reads:

$$T_k^\star = T_k + \beta^{-1} \log\left(\frac{\mu - \lambda^\star(T_k)}{\mu}\right) \mathbf{1}_{\lambda^\star(T_k) < 0}\,,$$

and, for any $\tau \in [T_k^\star, T_{k+1}]$:

$$\int_{T_k^\star}^{\tau} \lambda^\star(u)\,\mathrm{d}u = \mu(\tau - T_k^\star) + \beta^{-1}(\lambda^\star(T_k) - \mu)(e^{-\beta(T_k^\star - T_k)} - e^{-\beta(\tau - T_k)})\,.$$

Numerical results

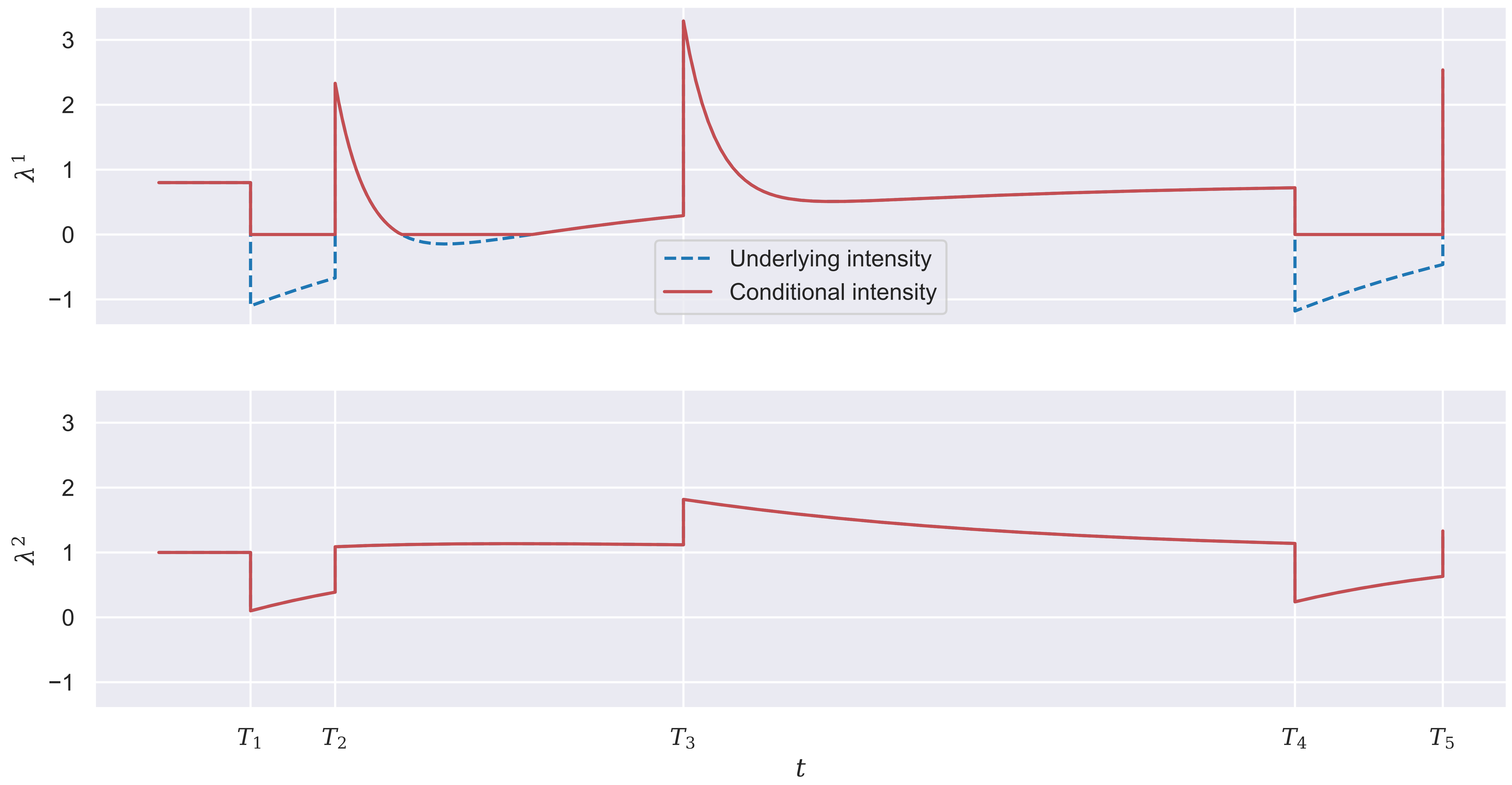


The multivariate framework

- Even when the functions are exponential (or monotonous), the functions $\lambda^{i\star}$ are not monotonous between two consecutive event times $T_{(k)}, T_{(k+1)}$.

- We can define the restart times $T^{i\star}_{(k)} = \min(\inf\{t \geq T_{(k)} : \lambda^{i\star}(t) \geq 0\}, T_{(k+1)})$.

- We adapt our methodology to the exponential kernel functions $h_{ij}(t) = \alpha_{ij} e^{-\beta_{ij} t}$, for $t \geq 0$.



The exponential kernel function

Lemma

Let N be a multivariate Hawkes process defined by its conditional intensities λ^i with exponential kernel functions. Let us assume that for each i, there exists $\beta_i = \beta_{ij}$ for all j.

Then $\lambda^{i\star}$ is piecewise monotonous and, for any $k > 1$:

$$T^{i\star}_{(k)} = \min\left(T_{(k)} + \beta_i^{-1} \log\left(\frac{\mu^i - \lambda^{i\star}(T_{(k)})}{\mu^i} \right) \mathbf{1}_{\{\lambda^{i\star}(T_{(k)})<0\}},\ T_{(k+1)} \right).$$

Furthermore, if $T^{i\star}_{(k)} < T_{(k+1)}$, then

$$\lambda^i(t) = \lambda^{i\star}(t) > 0 \quad \text{for any } t \in (T^{i\star}_{(k)}, T_{(k+1)}).$$

- We can then again implement the MLE procedure.
- **New question:** how can we estimate the **null interactions**?

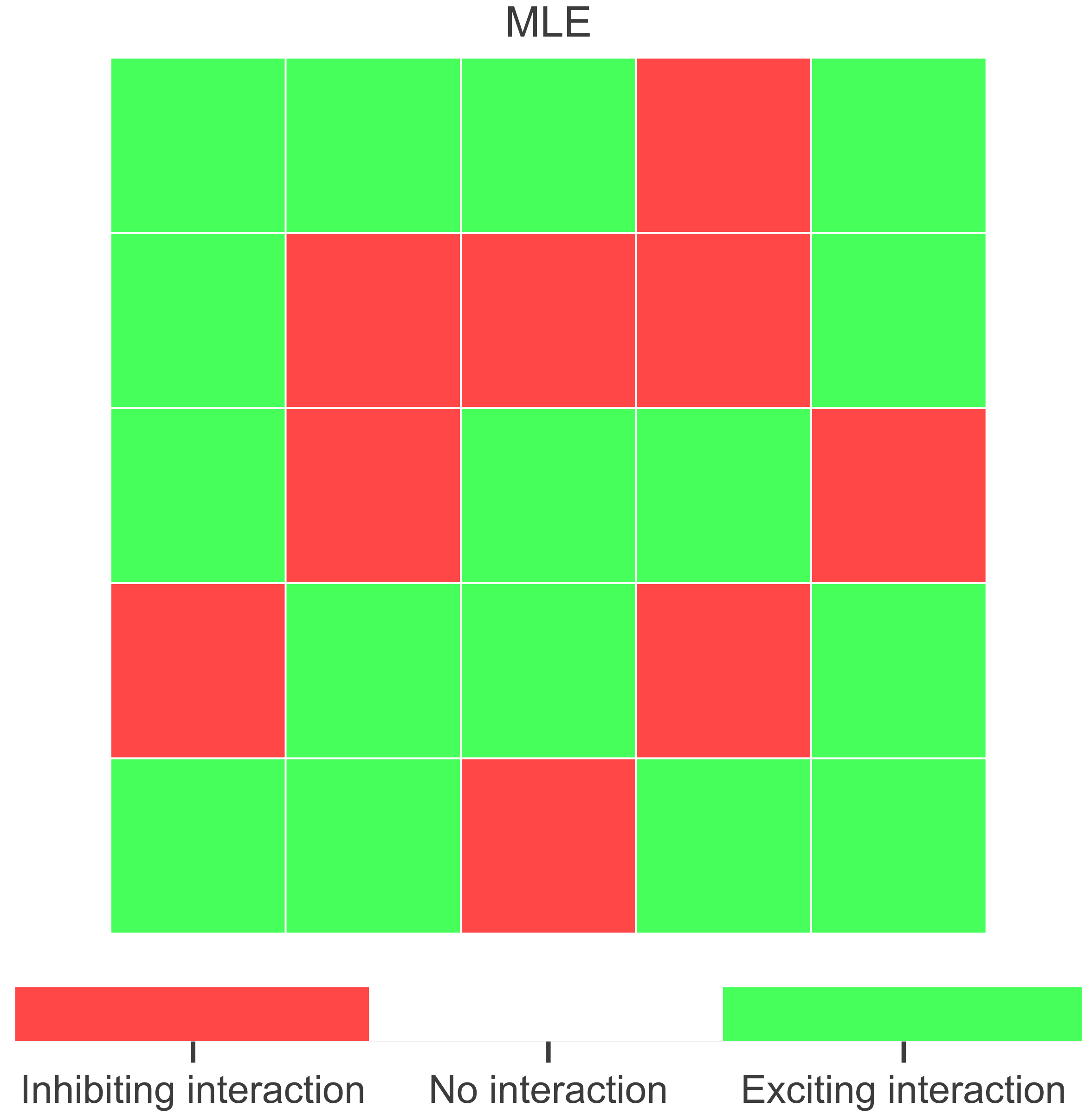
On estimating the support of interactions

- **Challenge:** unlike the univariate scenario, some interactions may be inexistent ($\alpha_{ij} = 0$), which a classical MLE estimation may not capture.
- **Solution:** Do a three-step estimation:

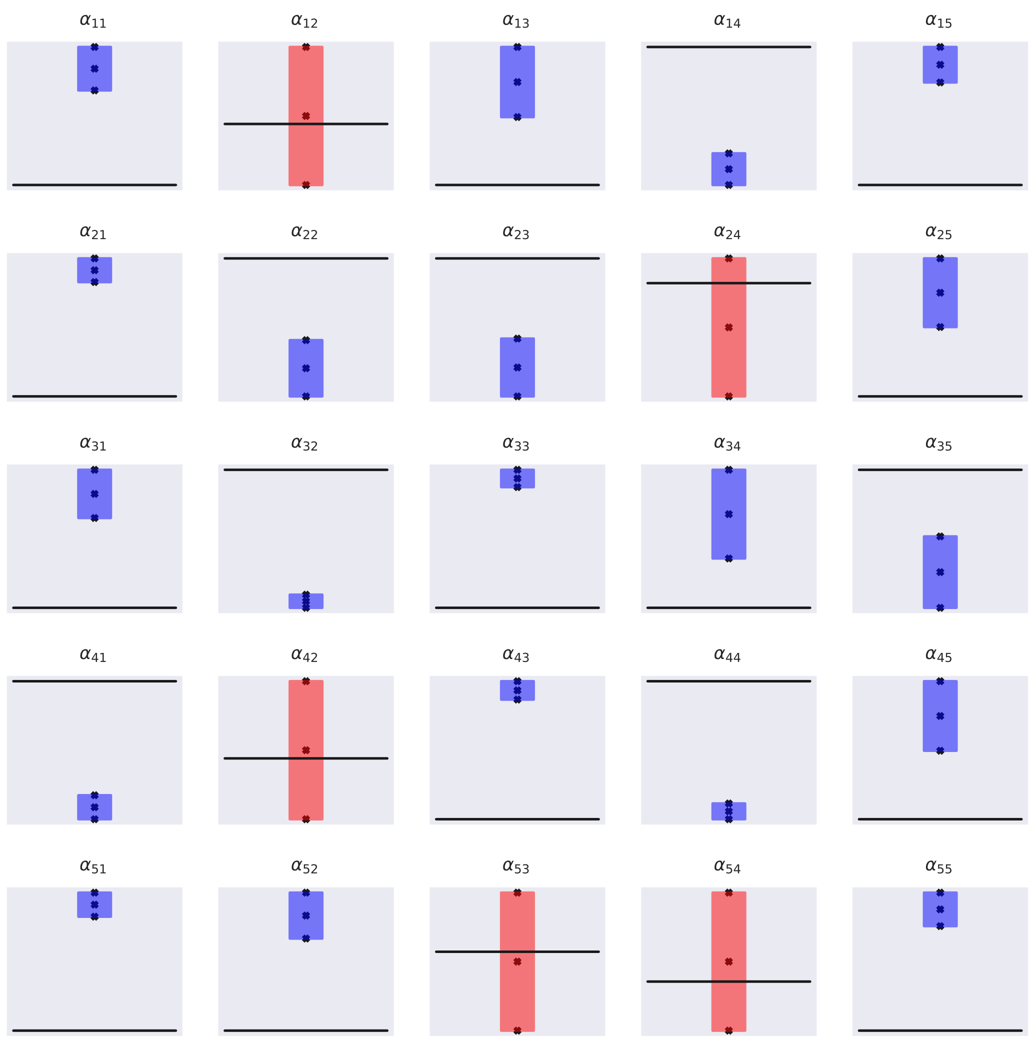
1. Estimation of matrix α through MLE

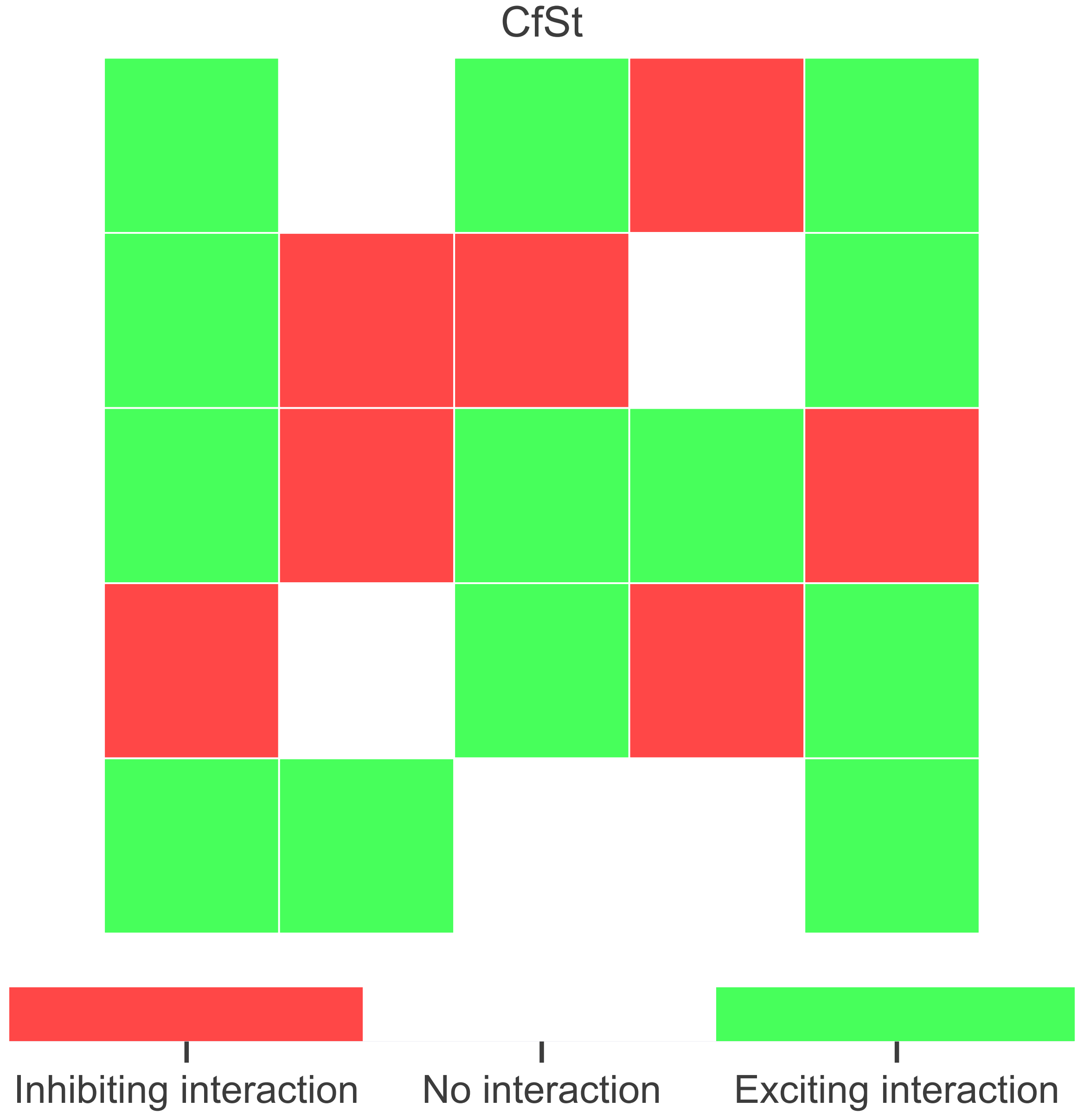
2. Support estimation by thresholding or confidence interval

3. Re-estimation through MLE in reduced model



Sign of estimations





Sign of estimations

2. Support estimation by thresholding or confidence interval

- **MLE-ε:** consider the ordered absolute values of the estimated entries of matrix α, noted $(\tilde{\alpha}_{(l)})_l$.

  Compute the cumulative sums $s_k = \sum_{l=1}^{k} \tilde{\alpha}_{(l)}$ and let $S = s_{d^2}$.

  Set $\tilde{\alpha}_{(k)} = 0$, for all k such that $s_k < \varepsilon S$.

On estimating the support of interactions

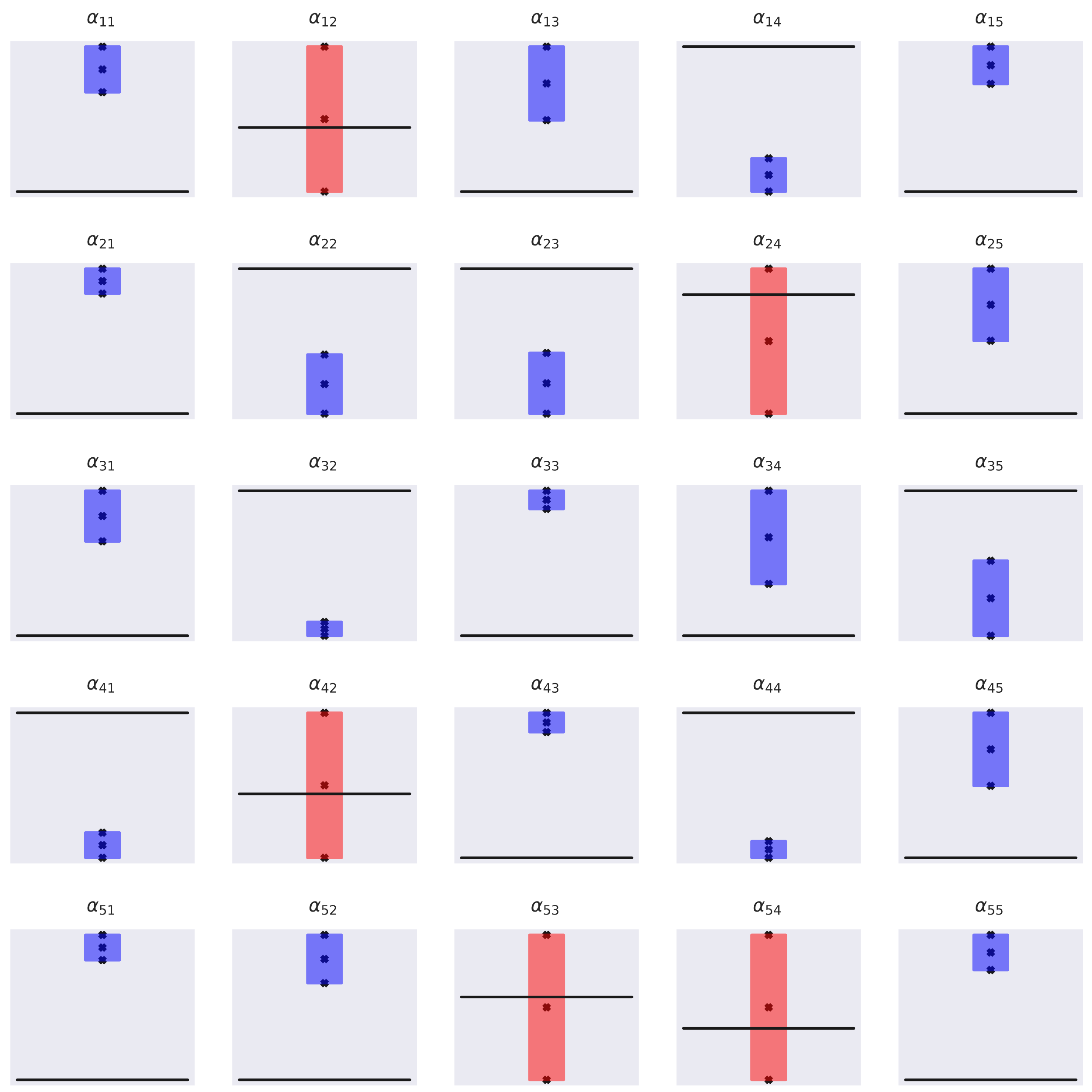
2. Support estimation by thresholding or confidence interval

- **MLE-ε:** consider the ordered absolute values of the estimated entries of matrix α, noted $(\tilde{\alpha}_{(l)})_l$.

  Compute the cumulative sums $s_k = \sum_{l=1}^{k} \tilde{\alpha}_{(l)}$ and let $S = s_{d^2}$.

  Set $\tilde{\alpha}_{(k)} = 0$, for all k such that $s_k < \varepsilon S$.

- **Confidence interval:** construct a confidence interval I_{ij} for each parameter α_{ij} through a set of estimations $(\tilde{\alpha}^k_{ij})_k$.
  Set $\tilde{\alpha}_{ij} = 0$ if and only if $0 \in I_{ij}$.



Confidence intervals for a 5 dimensional Hawkes process.
In red the intervals containing the value 0

Goodness-of-fit test

- **Challenge:** how can we compare different estimations in order to choose the best model in a real data scenario without access to the true parameters?

- **Solution:** Implement a hypothesis testing procedure through the time change theorem.

- For any parameter θ_0, we define the null hypothesis:

  - For every integer i, $\mathscr{H}_i$: $\left(\Lambda^i_{\theta_0}\left(T^i_{k+1}\right) - \Lambda^i_{\theta_0}\left(T^i_k\right)\right)_k$ is an i.i.d sample from a unit rate exponential distribution.

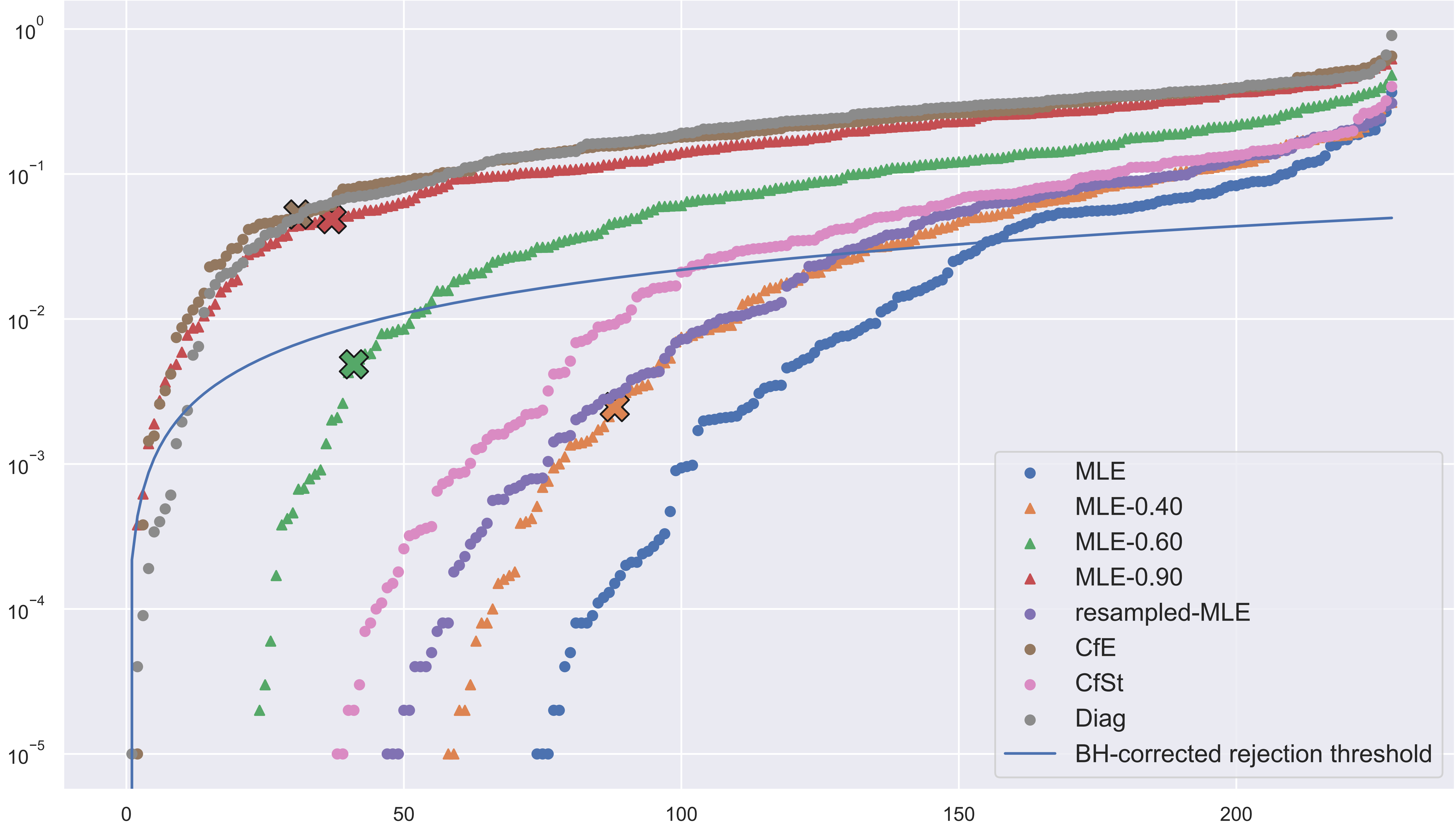
  - Additionally $\mathscr{H}_{tot}$: $\left(\Lambda_{\theta_0}\left(T_{(k+1)}\right) - \Lambda_{\theta_0}\left(T_{(k)}\right)\right)_k$ is an i.i.d sample from a unit rate exponential distribution.

- $\mathscr{H}_i$ tests the **goodness-of-fit** between θ_0 and the observations of process $N_i \to p$-value: p_i.

- $\mathscr{H}_{tot}$ tests the **goodness-of-fit** between θ_0 and the process N seen as a whole $\to p$-value: p_{tot}.

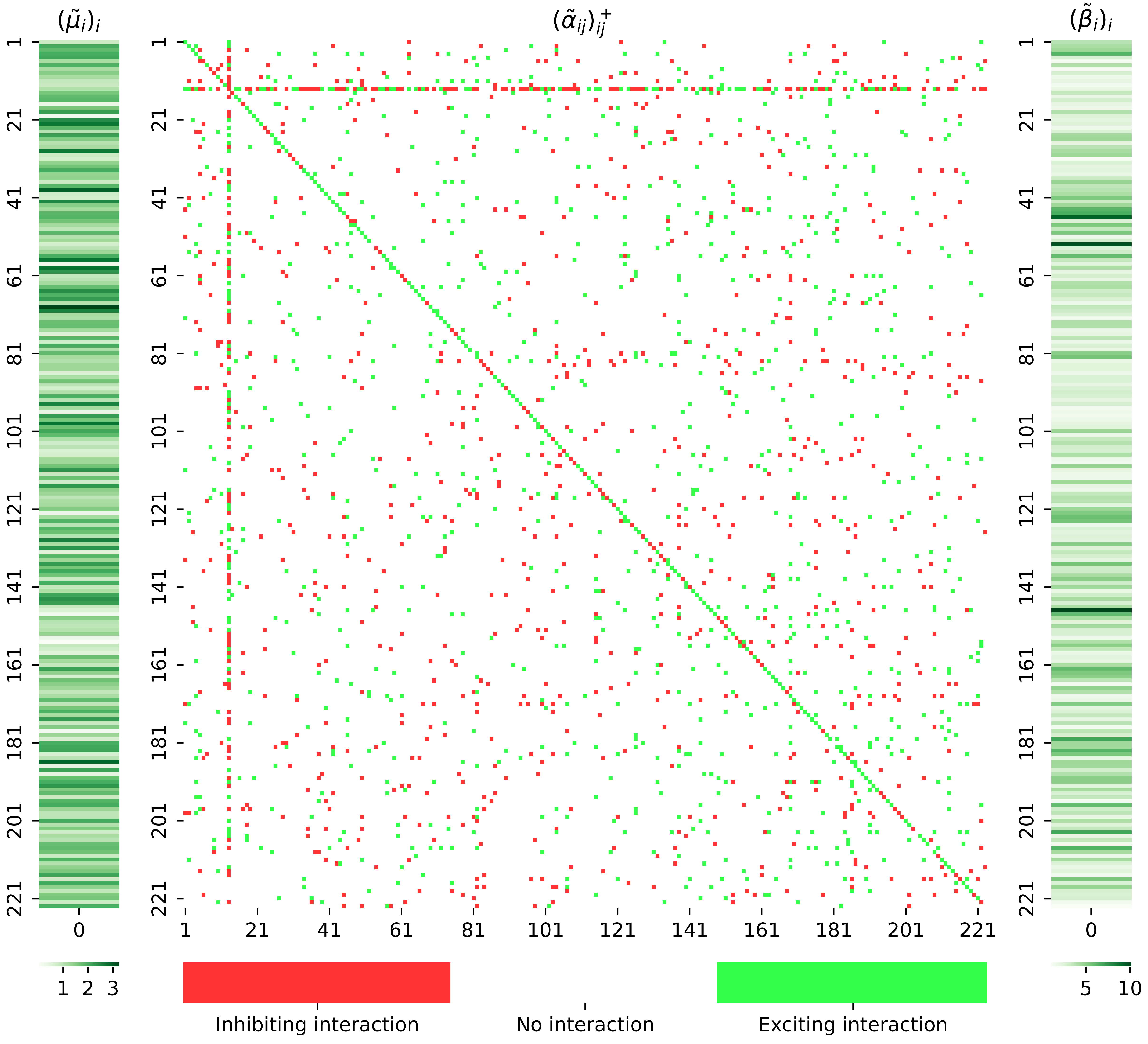
Numerical results: neuronal data

- Our data consists of 10 realisations of neuronal activations from a red-eared turtle. 223 neurons in total → add a multiple testing procedure.



Ordered p-values for all hypothesis tests $\mathscr{H}_i$ and $\mathscr{H}_{tot}$.
p_{tot} appears as a cross for each model.

Numerical results: neuronal data



Support estimation (Empirical confidence interval)

Motivation

- How is temporal data **collected** and **converted** to a point process realisation?

Earthquake occurrences

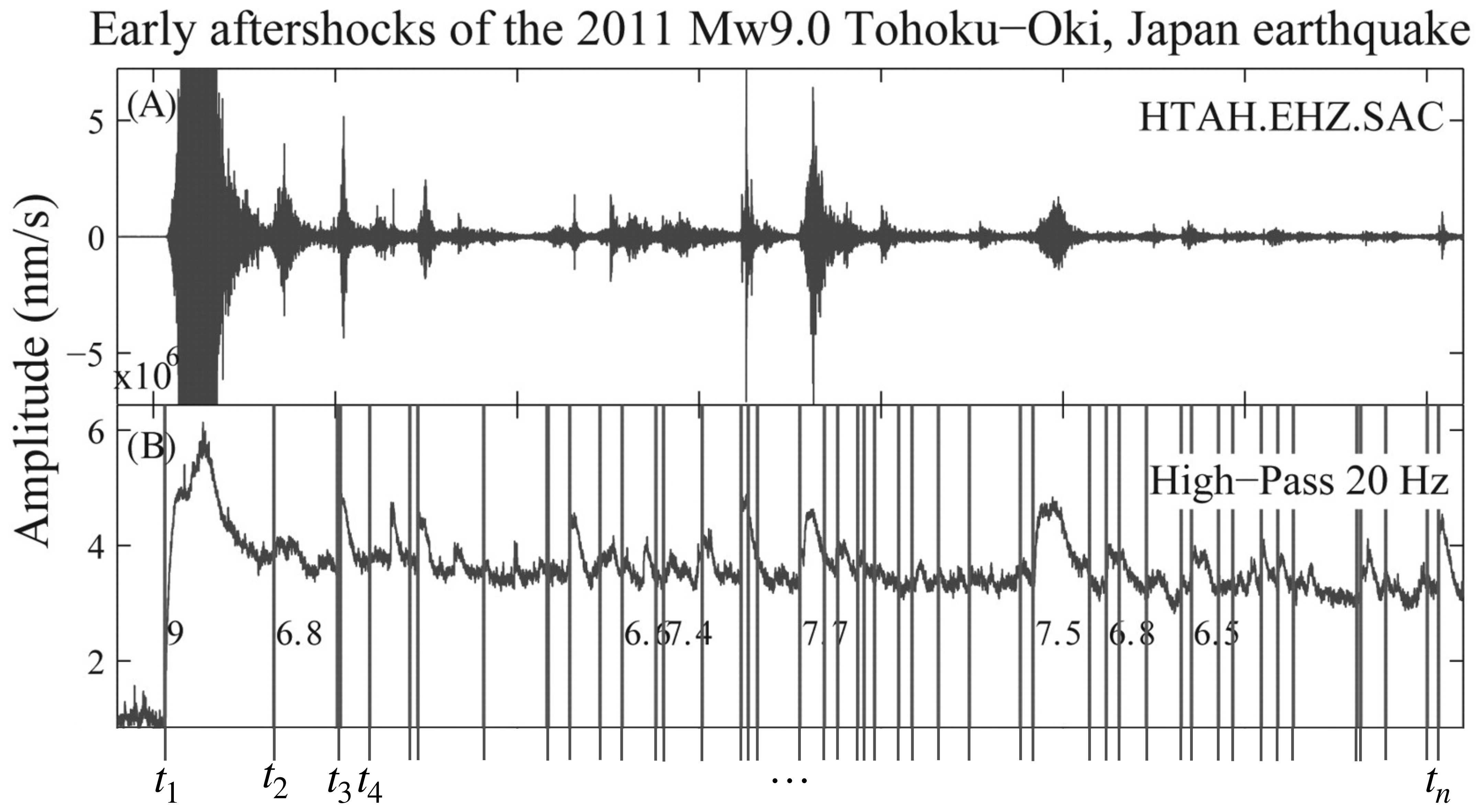


Image from Peng et al. (2012)

Synaptic signalling in neurons

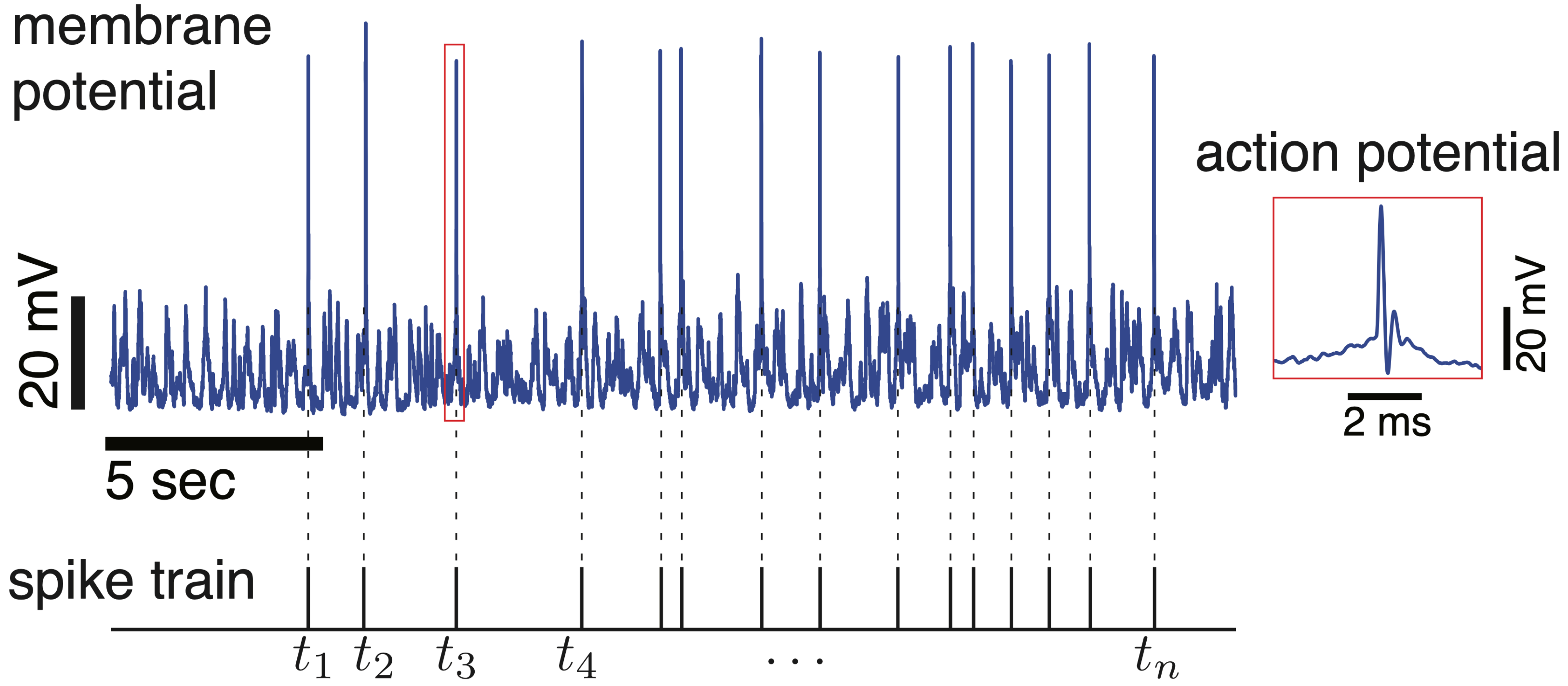


Image from Park et al. (2013)

Second axis:

Spectral methods for imperfect data

Errors and literature overview

- Data may contain **measurement errors** which may introduce bias to any standard inference procedure.

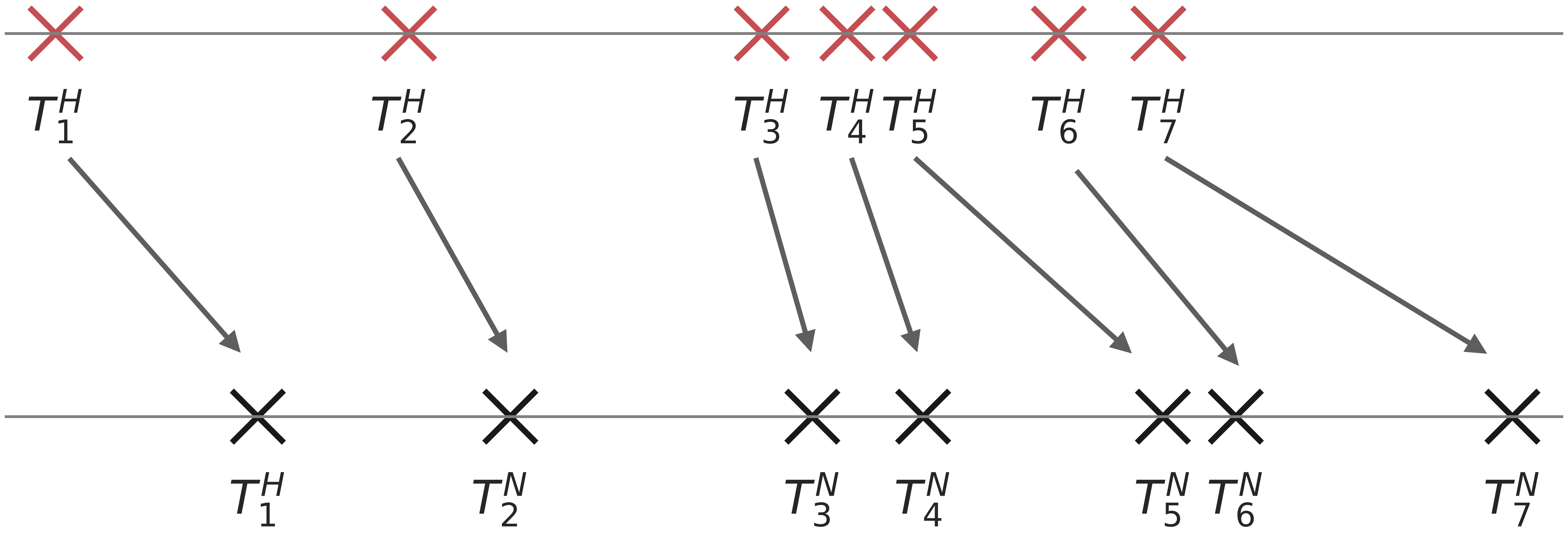
What is there in the literature for imperfect observations of point processes?

- Data may contain **measurement errors** which may introduce bias to any standard inference procedure.

What is there in the literature for imperfect observations of point processes?

- The most common case is jittering or random displacement of points as in Antoniadis et al. (2006), Trouleau et al. (2019) and Bonnet et al. (2022).

H: Hawkes process (μ, h)



Errors and literature overview

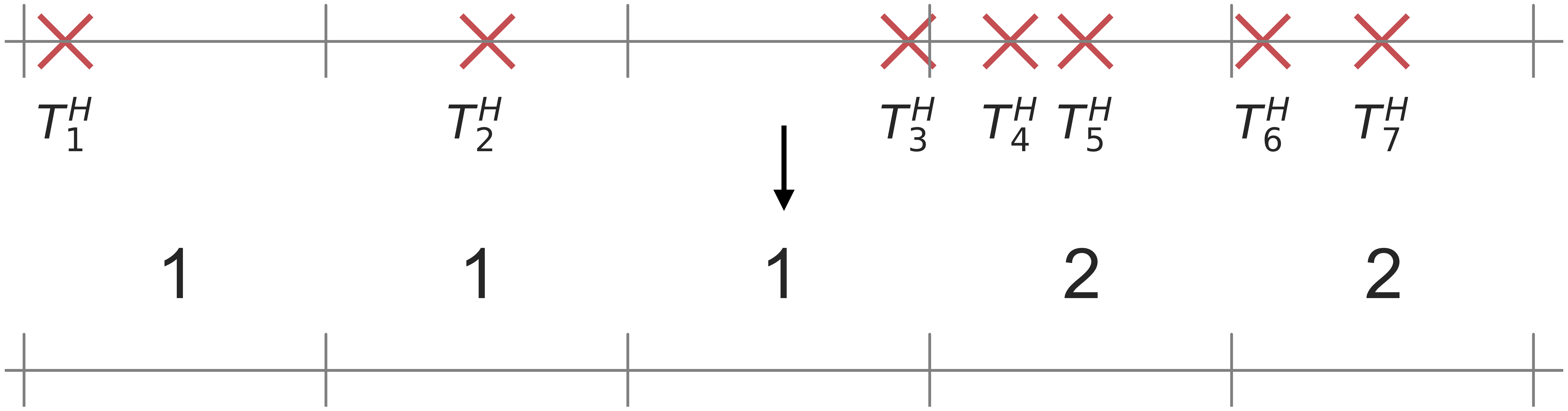
- Data may contain **measurement errors** which may introduce bias to any standard inference procedure.

What is there in the literature for imperfect observations of point processes?

- The most common case is jittering or random displacement of points as in Antoniadis et al. (2006), Trouleau et al. (2019) and Bonnet et al. (2022).

- In Cheysson et al. (2022), binned observations of H are studied through spectral theory.

H: Hawkes process (μ, h)



Noise by superposition

First scenario: additional external points (superposition)

Errors and literature overview

- Data may contain **measurement errors** which may introduce bias to any standard inference procedure.

What is there in the literature for imperfect observations of point processes?

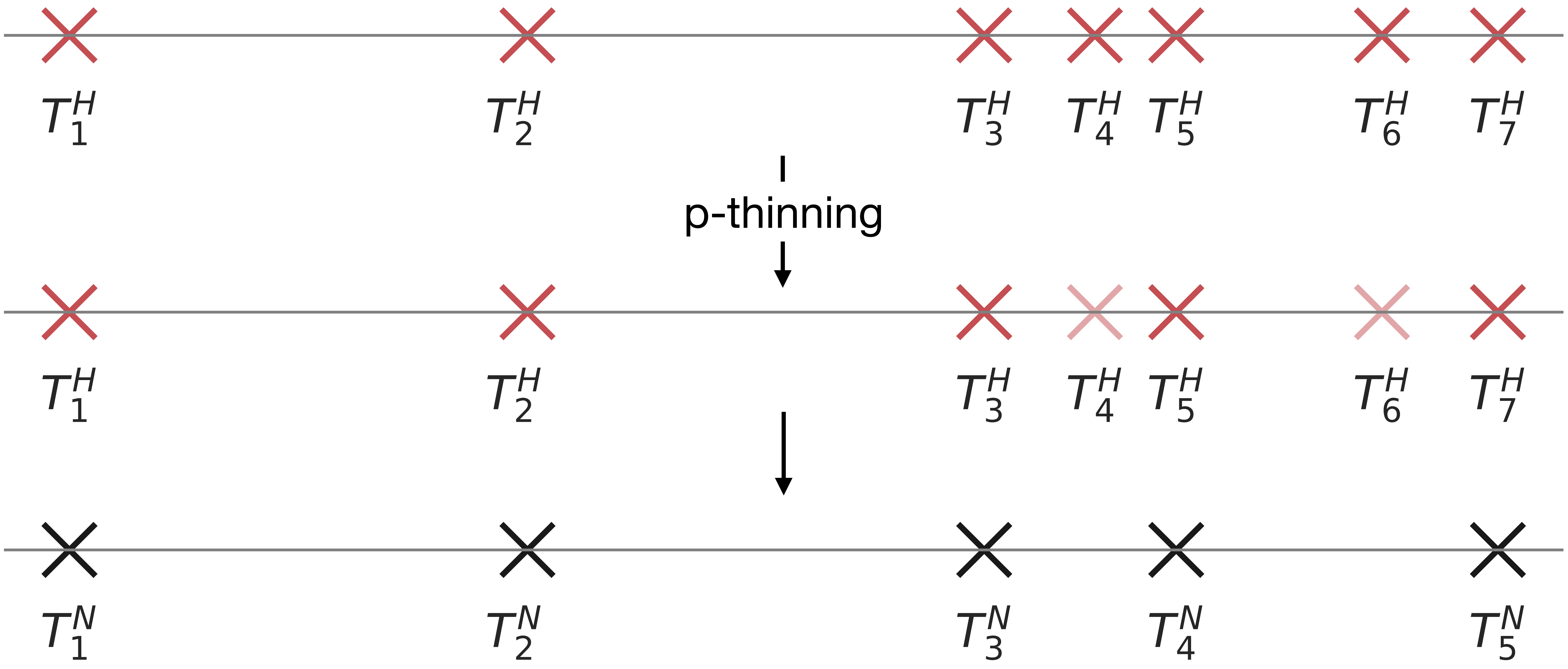
- The most common case is jittering or random displacement of points as in Antoniadis et al. (2006), Trouleau et al. (2019) and Bonnet et al. (2022).

- In Cheysson et al. (2022), binned observations of H are studied through spectral theory.

- Lund et al. (2000) focuses on processes noised by **superposition**, thinning and jittering by studying the conditional log-likelihood of N given H.
  Staerman et al. (2024) for marked Hawkes processes by latent factor estimation through an EM algorithm.

Second scenario: missing points (thinning)

H: Hawkes process (μ, h)



N: Thinned Hawkes process (μ, h, p)

Errors and literature overview

- Data may contain **measurement errors** which may introduce bias to any standard inference procedure.

What is there in the literature for imperfect observations of point processes?

- The most common case is jittering or random displacement of points as in Antoniadis et al. (2006), Trouleau et al. (2019) and Bonnet et al. (2022).

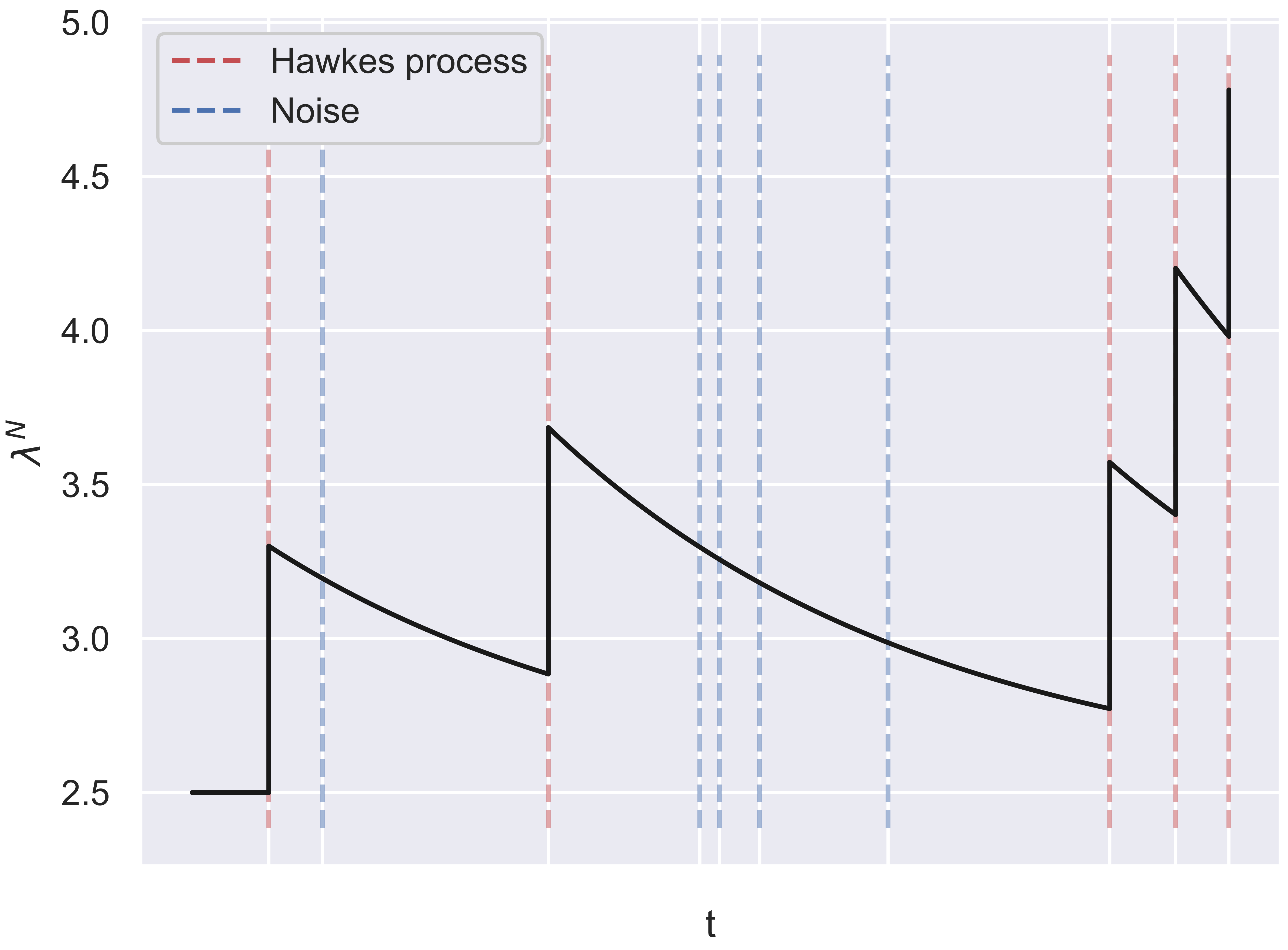
- In Cheysson et al. (2022), binned observations of H are studied through spectral theory.

- Lund et al. (2000) focuses on processes noised by **superposition**, thinning and jittering by studying the conditional log-likelihood of N given H.
  Staerman et al. (2024) for marked Hawkes processes by latent factor estimation through an EM algorithm.

- The case of missing points, obtained by **thinning**, is studied in Mei et al. (2019) and Deutsch et al. (2020).

Challenge

- Let H be a univariate (or multivariate) self-exciting Hawkes process.
- Let N be the resulting point process obtained from H by either superposition or by thinning.
- **Goal:** provide a parametric inference procedure of the parameters of H and of the noise.
- **Challenge:** we do not have access to the distribution of N.



Intensity function of a univariate Hawkes process superposed to a homogeneous Poisson process

Spectral theory of point processes

- Spectral analysis from time series theory, introduced for point processes in Bartlett (1963).
  - Growing interest, in particular in spatial contexts (Rajala et al. 2023) with advancements in asymptotic results in $\mathbb{R}^d$ (Yang et al. 2024).
  - Application to locally stationary Hawkes processes (Roueff et al. 2015).

- Based on the the **spectral density** f and the **periodogram** I^T.

- The spectral density f characterises the second-order moment of a point process instead of focusing on its distribution.

- For an observation $(T_k)_{1 \leq k \leq N([0,T])}$ of N, its periodogram reads, for any $\omega \in \mathbb{R}$:

$$I^T(\omega) = \frac{1}{T} \sum_{k=1}^{N(T)} \sum_{l=1}^{N(T)} \mathrm{e}^{-2\pi \mathrm{i} \omega (T_k^N - T_l^N)} .$$

- This quantity can be computed even in the presence of measurement errors.

The spectral log-likelihood

- The periodogram and the spectral density are asymptotically linked:

$$I^T(\omega) \xrightarrow[T\to+\infty]{d} \text{Exp}\left(\frac{1}{f(\omega)}\right).$$

And for any $(\omega_k)_{1\leq k\leq M}$ such that $\omega_i \neq \omega_j$ for all $i \neq j$, the r.v. $(I^T(\omega_k))_k$ are asymptotically independent.

- The **spectral log-likelihood** (Whittle 1952) an be defined as:

$$\ell_T = -\frac{1}{T}\sum_{k=1}^{M}\left(\log\left(f(\omega_k)\right) + \frac{I^T(\omega_k)}{f(\omega_k)}\right), \quad \text{for } \omega_k = k/T\,.$$

- We obtain an estimation of the parameters by maximising ℓ_T.

- **Challenge:** obtain an expression of the spectral density f for our noised processes N.

The spectral density of noised processes

Proposition

Let X, Y be two independent stationary point processes with respective spectral densities f^X, f^Y and $p \in (0,1)$.

1. The superposition $X + Y$ admits a spectral density such as:

$$f^{X+Y} = f^X + f^Y .$$

2. The p-thinning X_p of X admits a spectral density such as:

$$f^{X_p} = p^2 f^X + p(1-p)\,\mathbb{E}[\lambda^H(0)] .$$

The superposition noise

Superposition noise

- Let, for any $t \in \mathbb{R}$, $h(t) = \alpha\beta \mathrm{e}^{-\beta t}$ for $0 < \alpha < 1$ and $\beta > 0$.
- The spectral density of a Hawkes process is known (Hawkes, 1971) and so the spectral density of the superposition $N = H + P$ is given by:

$$f^N(\omega) = \frac{\mu}{1-\alpha}\left(1 + \frac{\beta\alpha(2-\alpha)}{(\beta(1-\alpha))^2 + 4\pi^2\omega^2}\right) + \lambda_0 \,.$$

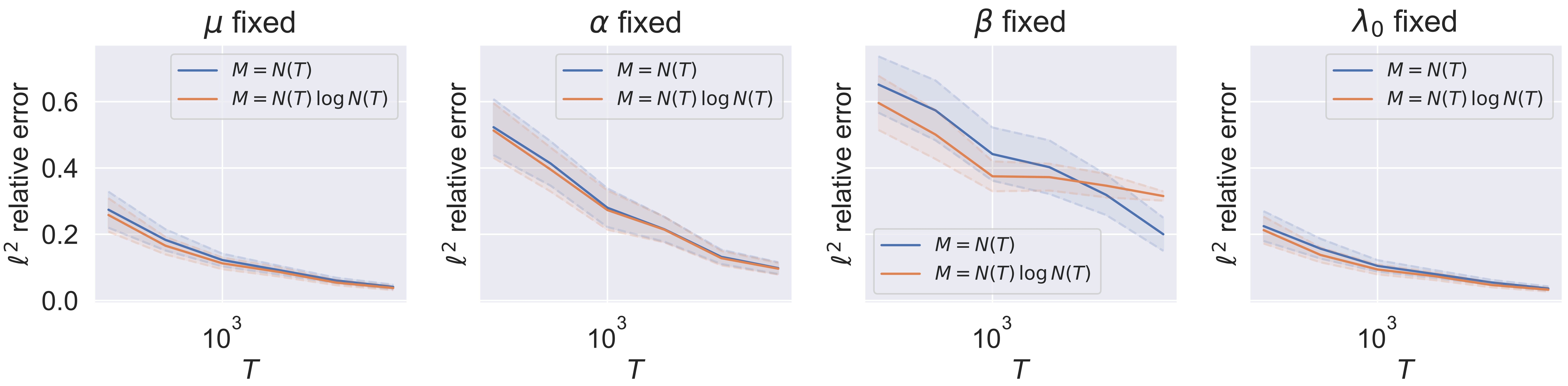
- We define the parametric model:

$$\mathcal{Q} = \{f_\theta^N : \mathbb{R} \to \mathbb{C}, \quad \theta = (\mu, \alpha, \beta, \lambda_0) \in \Theta\}$$

Proposition

The model $\mathcal{Q}$ is **identifiable** if and only if one of the parameters in the 4-uplet $\theta = (\mu, \alpha, \beta, \lambda_0)$ is fixed.

Convergence of the estimator



ℓ_2 relative error with respect to observation horizon T

The bivariate case

- Let H be a bivariate exponential Hawkes process ($h_{ij} = \alpha_{ij}\beta_i e^{-\beta_i t}$) and P a homogeneous bivariate Poisson process with constant rate λ_0.

- We define the parametric model:

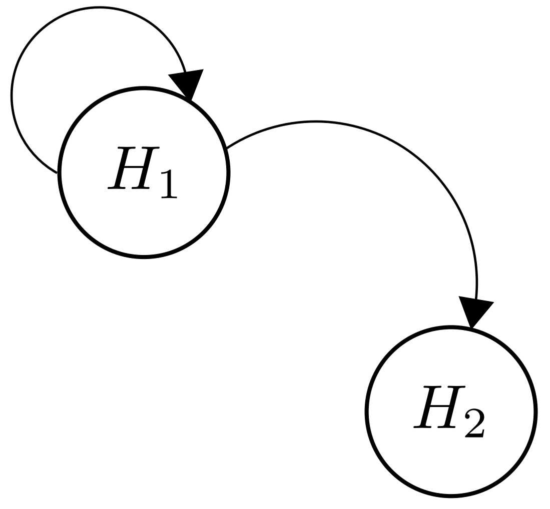
$$\mathcal{Q}_\Lambda = \left\{ f_\theta^N : \mathbb{R} \to \mathbb{C}, \quad \theta = (\mu, \alpha, \beta, \lambda_0) \in \mathbb{R}^2_{>0} \times \Lambda \times \mathbb{R}^2_{>0} \times \mathbb{R}_{>0} \right\}$$

Proposition

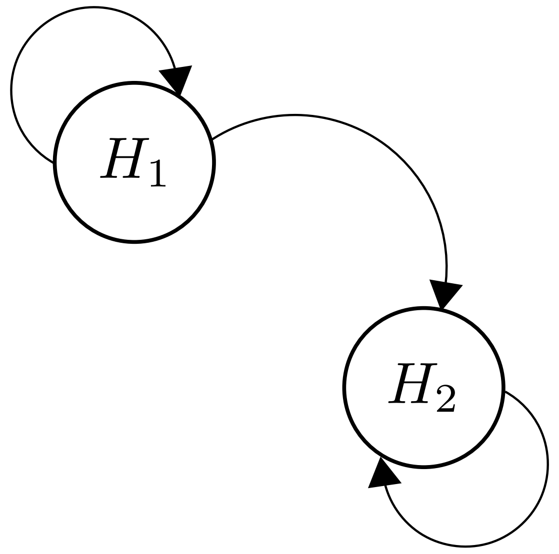
The model $\mathcal{Q}_\Lambda$ is **identifiable** if:

1. $\Lambda = \left\{ \begin{pmatrix} \alpha_{11} & 0 \\ \alpha_{21} & 0 \end{pmatrix}, \quad 0 \leq \alpha_{11} < 1, \alpha_{21} > 0 \right\}.$

2. $\Lambda = \left\{ \begin{pmatrix} \alpha_{11} & 0 \\ \alpha_{21} & \alpha_{22} \end{pmatrix}, \quad 0 < \alpha_{11} < 1, \alpha_{21} > 0, 0 \leq \alpha_{22} < 1 \right\}.$



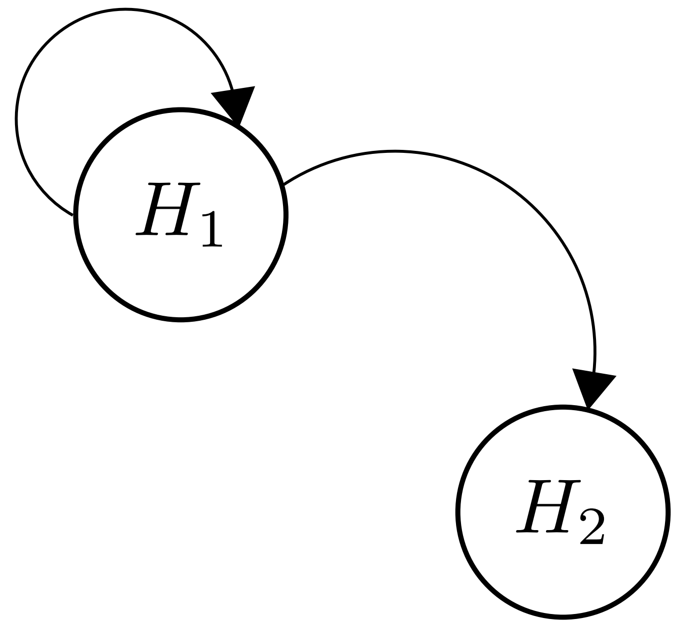
Scenario 1 of identifiability

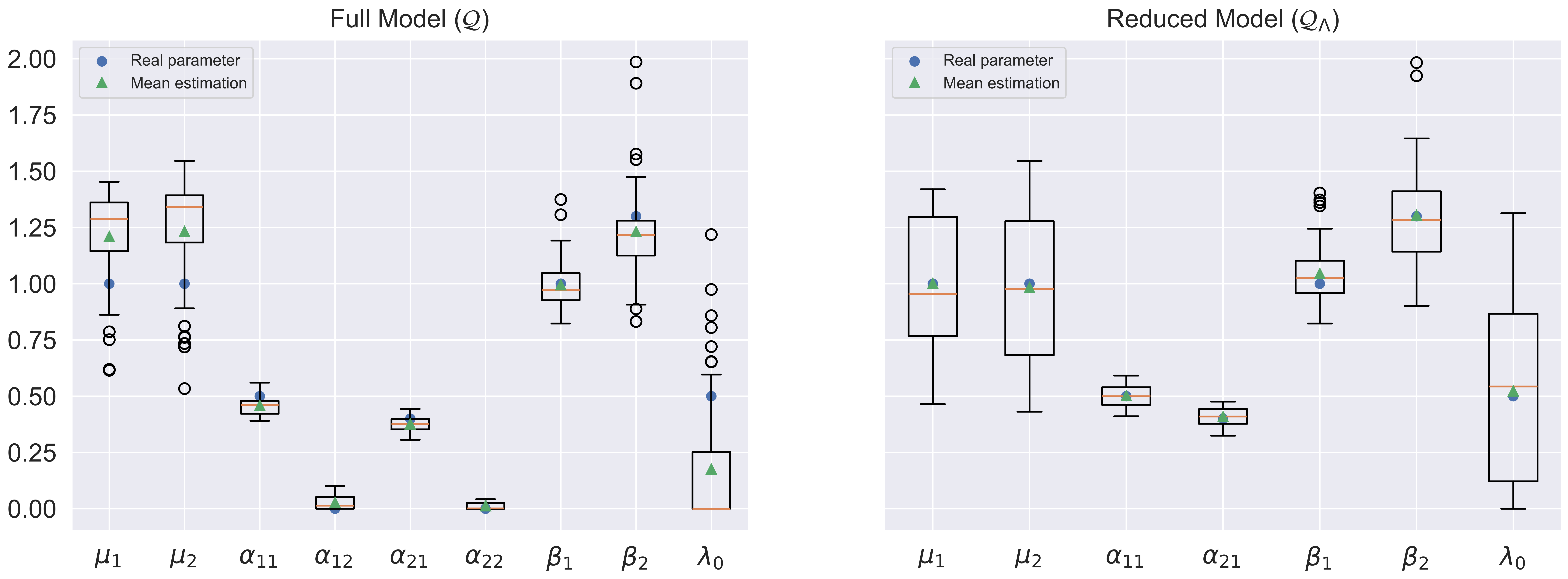


Scenario 2 of identifiability

Model

Bivariate scenario: $\begin{pmatrix} \alpha_{11} & 0 \\ \alpha_{21} & 0 \end{pmatrix}$





Boxplot of estimations for each parameter

Thinning noise

- Let, for any $t \in \mathbb{R}$, $h(t) = \alpha\beta \mathrm{e}^{-\beta t}$ for $0 < \alpha < 1$ and $\beta > 0$.

- The spectral density of a p-thinned Hawkes process is given by:

$$f^{N}(\omega) = \frac{\mu p}{1-\alpha}\left(1 + p\frac{\beta\alpha(2-\alpha)}{(\beta(1-\alpha))^2 + 4\pi^2\omega^2}\right).$$

- We define the parametric model:

$$\mathcal{Q} = \{f_\theta^N : \mathbb{R} \to \mathbb{C}, \quad \theta = (\mu, \alpha, \beta, p) \in \Theta\}$$

Proposition

The model $\mathcal{Q}$ is **identifiable** if and only if one of the parameters in the 4-uplet $\theta = (\mu, \alpha, \beta, p)$ is fixed.

The p-thinning of a Hawkes process

- **Advantage:** thinning a point process can be used as a **subsampling** (Biscio, 2019).
- Subsampling can improve estimations when only a few realisations are available, which can be common in real data contexts.

- Let us compare the performance of spectral estimators with 1 observation of H in a small window $[0, T]$ in 4 scenarios:

  1. $\hat{\theta}$: obtained by maximising the non-penalised spectral log-likelihood.
  2. $\hat{\theta}^L$: obtained by maximising the L_2-penalised spectral log-likelihood.
  3. $\hat{\theta}^L_{partition}$: the average estimation by partitioning the window $[0, T]$.
  4. $\hat{\theta}^L_{thinning}$: the average estimation obtained by **p-thinning** l times process H.

Numerical results

- We carry out the estimations in 1000 different estimations.
- The subsampling by thinning scheme provides the best MSRE overall and the best estimator in 70% of the cases.

Estimator	MSRE				% best
	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\theta}$	
$\hat{\theta}$	0.18	0.13	5.14×10^2	2.85×10^2	1%
$\hat{\theta}^L$	0.12	0.09	0.13	0.12	6.7%
$\hat{\theta}^L_{partition}$	0.08	0.07	0.04	0.05	22.3%
$\hat{\theta}^L_{thinning}$	0.03	0.04	0.02	0.02	70%

Mean squared relative error (MSRE) of (μ, α, β) along with the MSRE of θ.
Last column shows the proportion of times that each estimator achieves the lowest relative ℓ_2 error.

Conclusion

- Our contributions concern the study of two extensions of the classical Hawkes process model in parametric frequentist settings:

A. The study of additive inhibition:

- Implemented the MLE for multivariate Hawkes processes.
- Proposed three methods to infer the null interactions.
- Illustrated in neuronal activity data with a multiple testing procedure to compare models and validate our estimations

Chapter 3: Bonnet, A., Martinez Herrera, M., Sangnier, M., “Maximum Likelihood Estimation for Hawkes Processes with self-excitation or inhibition.” (2021) in *Statistics and Probability Letters.*

Chapter 4: Bonnet, A., Martinez Herrera, M., Sangnier, M., “Inference of multivariate exponential Hawkes processes with inhibition and application to neuronal activity.” (2023) in *Statistics and Computing.*

Conclusion

- Our contributions concern the study of two extensions of the classical Hawkes process model in parametric frequentist settings:

B. Accounting for measurement errors in the form of additional or missing points:

  - Studied the spectral density of two models of noised Hawkes processes.
  - Proposed different conditions to retrieve identifiability for the statistical models.
  - Applied the results concerning the thinning of a point process to improve numerical estimations.

Chapter 5: Bonnet, A., Cheysson, F., Martinez Herrera, M., Sangnier, M., "Spectral analysis for the inference of noisy Hawkes processes." (2024) *In revision.*

Chapter 6: Cheysson F., Martinez Herrera, M., "A numerical exploration of thinned Hawkes processes through spectral theory." *Ongoing work.*

Perspectives

- Extend the MLE procedure with inhibition to **other kernel functions and non-linear functions** → model more accurately a wider spectrum of phenomena.

- Add **Ridge and Lasso penalisation** methods with efficient optimisation procedure and parameter selection paradigms → improve estimations in particular to better estimate the interaction matrix.

- Extend the study of **noised** Hawkes processes to **include inhibition** → open up the applications to real world data (like neuronal activity).

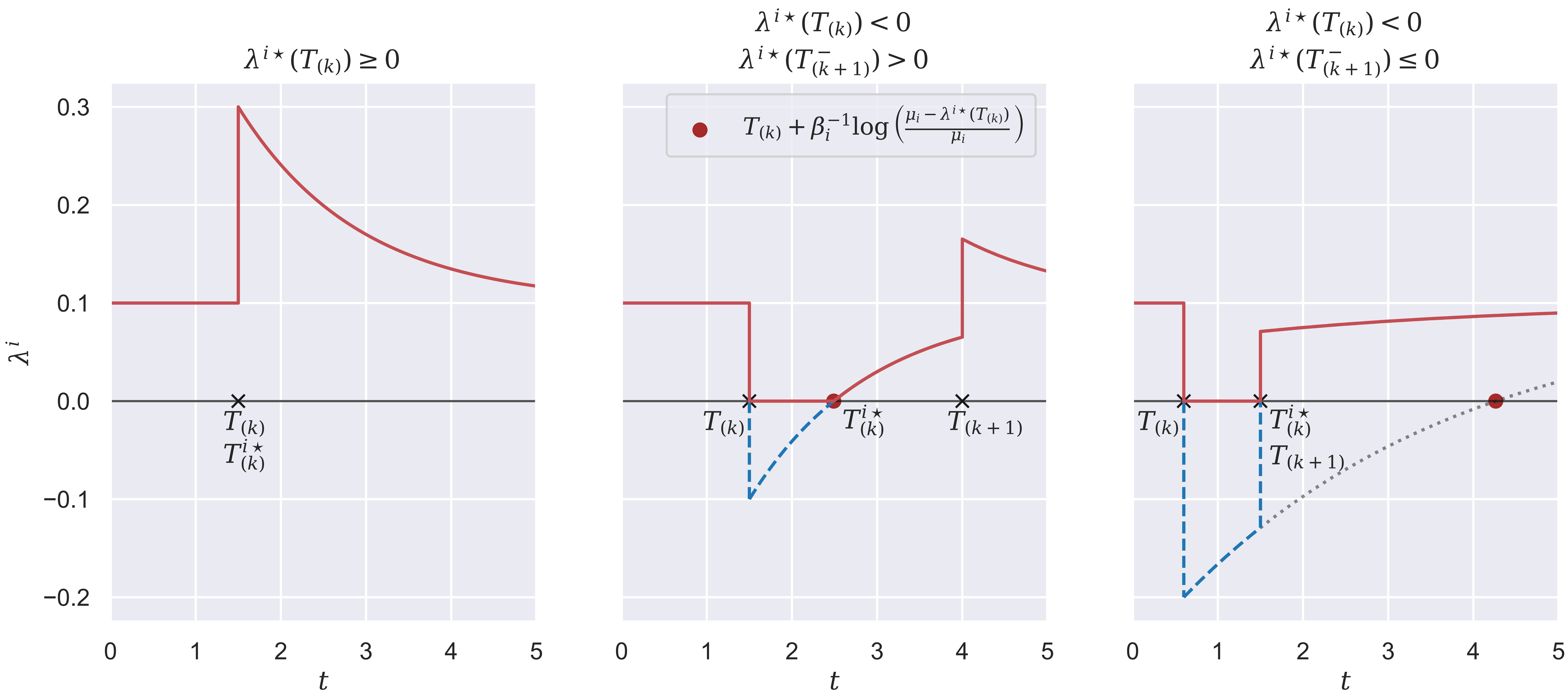
- Establish **asymptotic results** for our estimators → gain access to asymptotic confidence intervals.

Thank you very much

Appendix

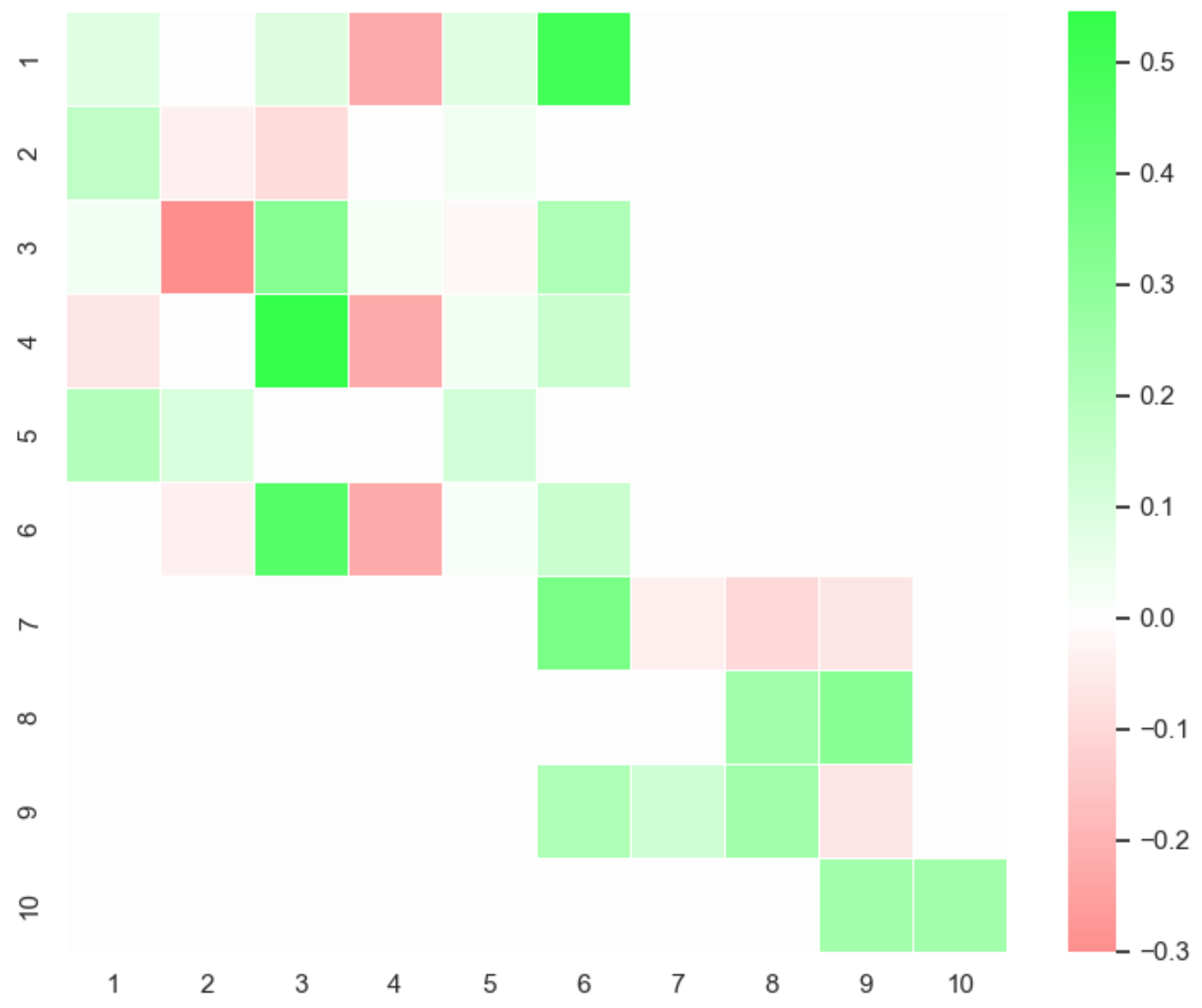
Motivation

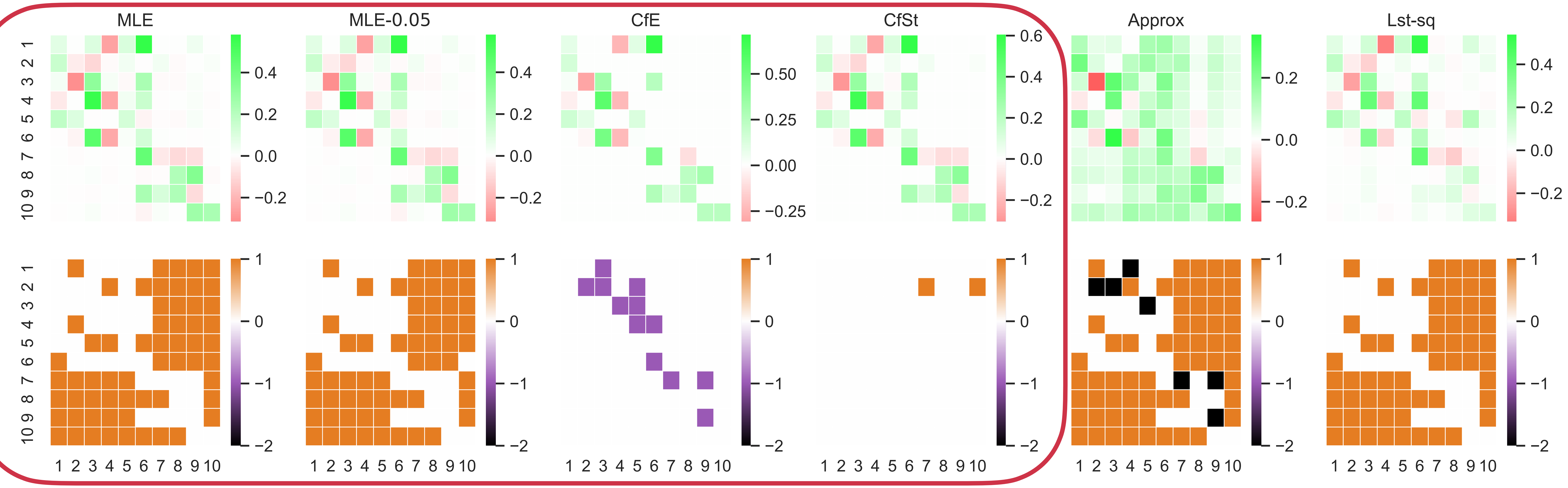
- There are three possible scenarios for the restart times:



Numerical results: simulated data

- Simulations in dimension $d = 10$ compared to:
  - Approximated log-likelihood maximisation procedure (Lemonnier et al. 2014).
  - Least-squares minimisation procedure (Bacry et al. 2020).





Appendix | Algorithm of log-likelihood computation (multivariate)

Algorithm 1: Estimation of log-likelihood $\ell_t(\theta)$ of a multivariate exponential Hawkes process.

Input Parameters μ^i, α_{ij}, β_i for $i, j \in \{1, \dots, d\}$, list of event times and marks $(T_{(k)}, m_k)_{k=1:N(t)}$;

Initialisation Initialize for all i, $\Lambda_k^i = \mu^i T_{(1)}$, $\lambda^{i\star}(T_{(k)}^-) = \mu^i$, $\lambda_k^{i\star} = \mu^i + \alpha_{im_1}$ and $\ell_t(\theta) = \log(\lambda^{m_1\star}(T_{(k)}^-)) - \sum_{i=1}^d \Lambda_k^i$;

for $k = 2$ *to* $N(t)$ **do**

- Compute for all i, $T_{(k-1)}^{i\star} = \min\left(T_{(k-1)} + \beta_i^{-1} \log\left(\frac{\mu^i - \lambda_k^{i\star}}{\mu^i}\right) \mathbb{1}_{\{\lambda_k^{i\star} < 0\}}, T_{(k)}\right)$;
- Compute for all i, $\Lambda_k^i = \mu^i (T_{(k)} - T_{(k-1)}^{i\star}) + \beta_i^{-1} (\lambda_k^{i\star} - \mu^i)(\mathrm{e}^{-\beta_i (T_{(k-1)}^{i\star} - T_{(k-1)})} - \mathrm{e}^{-\beta_i (T_{(k)} - T_{(k-1)})})$;
- Compute for all i, $\lambda^{i\star}(T_{(k)}^-) = \mu^i + (\lambda_k^{i\star} - \mu^i)\mathrm{e}^{-\beta_i (T_{(k)} - T_{(k-1)})}$;
- Update $\ell_t(\theta) = \ell_t(\theta) + \log(\lambda^{m_k\star}(T_{(k)}^-)) - \sum_{i=1}^d \Lambda_k^i$;
- Compute for all i, $\lambda_k^{i\star} = \lambda^{i\star}(T_{(k)}^-) + \alpha_{im_k}$;

end

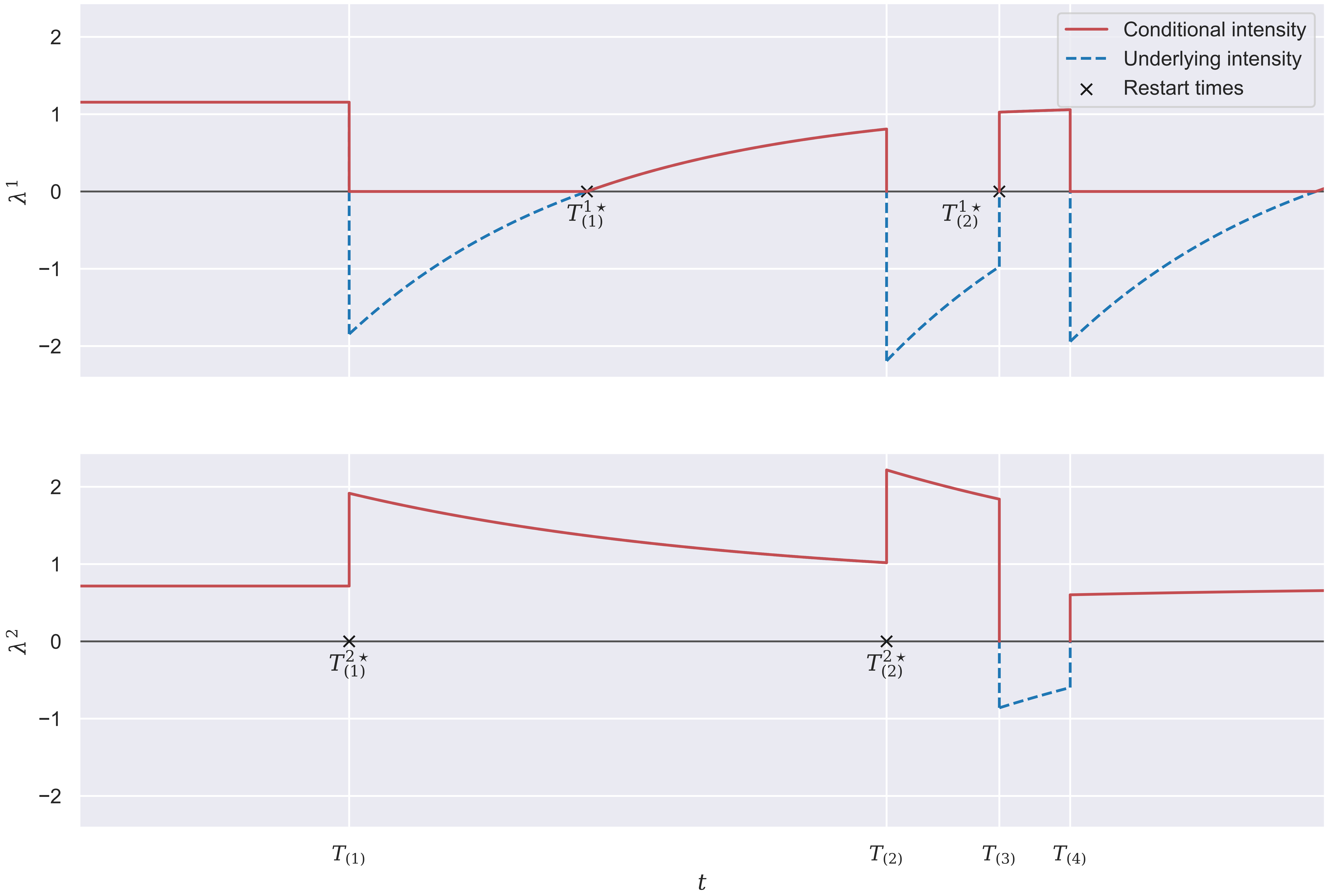
Compute for all i, $T_{(N(t))}^{i\star} = \min\left(T_{(N(t))} + \beta_i^{-1} \log\left(\frac{\mu^i - \lambda_k^{i\star}}{\mu^i}\right) \mathbb{1}_{\{\lambda_k^{i\star} < 0\}}, t\right)$;

Compute for all i,
$\Lambda_k^i = \left[\mu^i (t - T_{(N(t))}^{i\star}) + \beta_i^{-1} (\lambda_k^{i\star} - \mu^i)(\mathrm{e}^{-\beta_i (T_{N(t)}^{i\star} - T_{(N(t))})} - \mathrm{e}^{-\beta_i (t - T_{(N(t))})})\right] \mathbb{1}_{\{t > T_{(N(t))}^{i\star}\}}$;

Update $\ell_t(\theta) = \ell_t(\theta) - \sum_{i=1}^d \Lambda_k^i$;

return Log-likelihood $\ell_t(\theta)$.

Appendix | Restart times



Identifiability for multivariate Hawkes processes

Theorem

Let $(T_{(k)})_{k>0}$ be a realisation of a multivariate Hawkes process H and $\mathscr{H}_t$ be the corresponding filtration.

Let us assume that a.s. for every $(i,j) \in \{1,\ldots,d\}^2$, $i \neq j$, there exist an event time τ from process N^j, and an event time $\tau_+ > \tau$ from process N^i, such that:

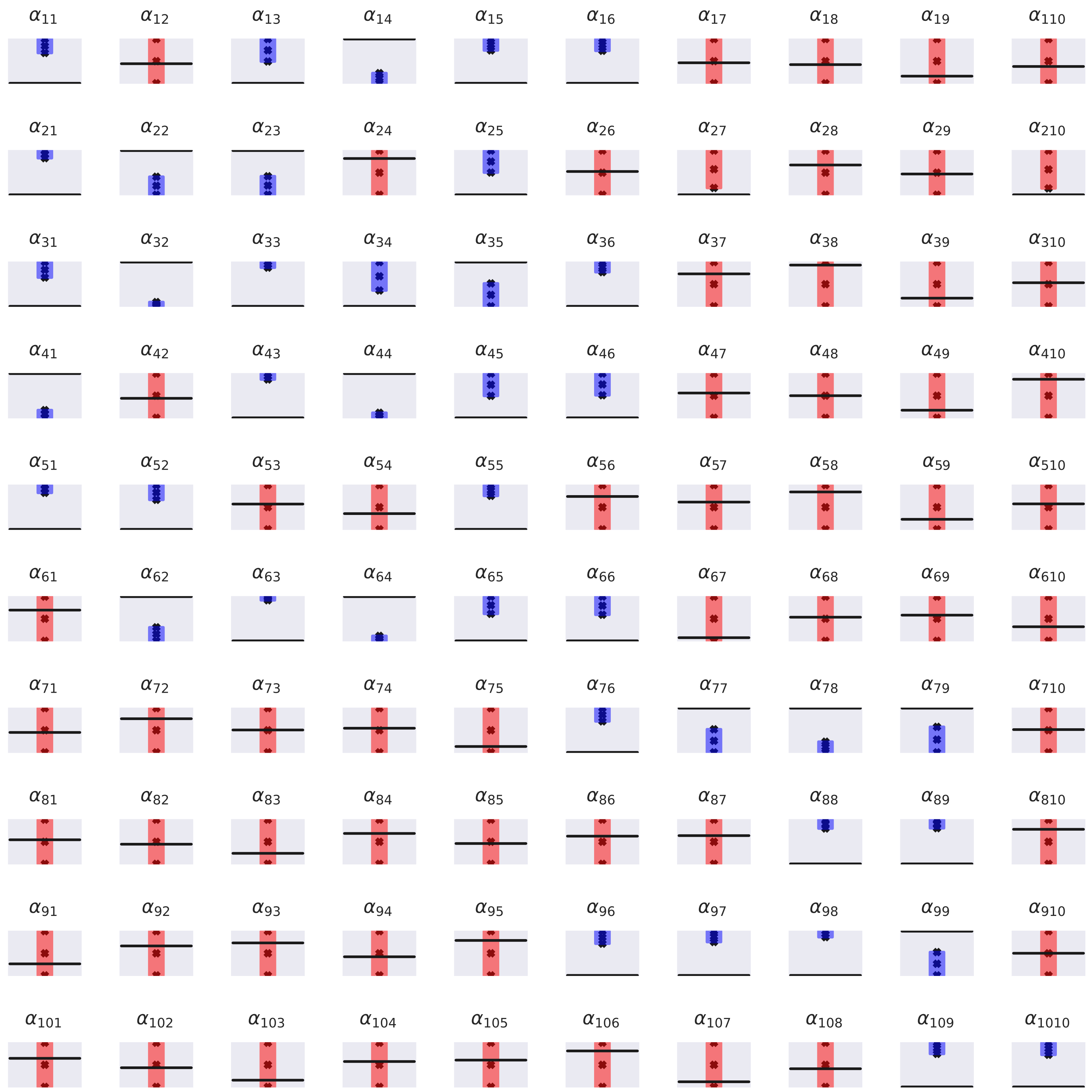
- $\lambda^i_{\theta_i}(\tau^-) > 0.$
- There are only events of process N^j in the interval $[\tau \,.\, \tau_+)$.

Then for any $\theta' \in \Theta$,

$$\forall i \in \{1,\ldots,d\}, \quad \lambda^i_{\theta_i}(t \mid \mathscr{H}_t) = \lambda^i_{\theta'_i}(t \mid \mathscr{H}_t) \text{ a.e.} \iff \theta = \theta' .$$

Appendix | Empirical confidence estimation

2. Support estimation by thresholding or empirical confidence interval



- The **Benjamini-Hochberg** procedure for multiple testing → control the False Discovery Rate (FDR)

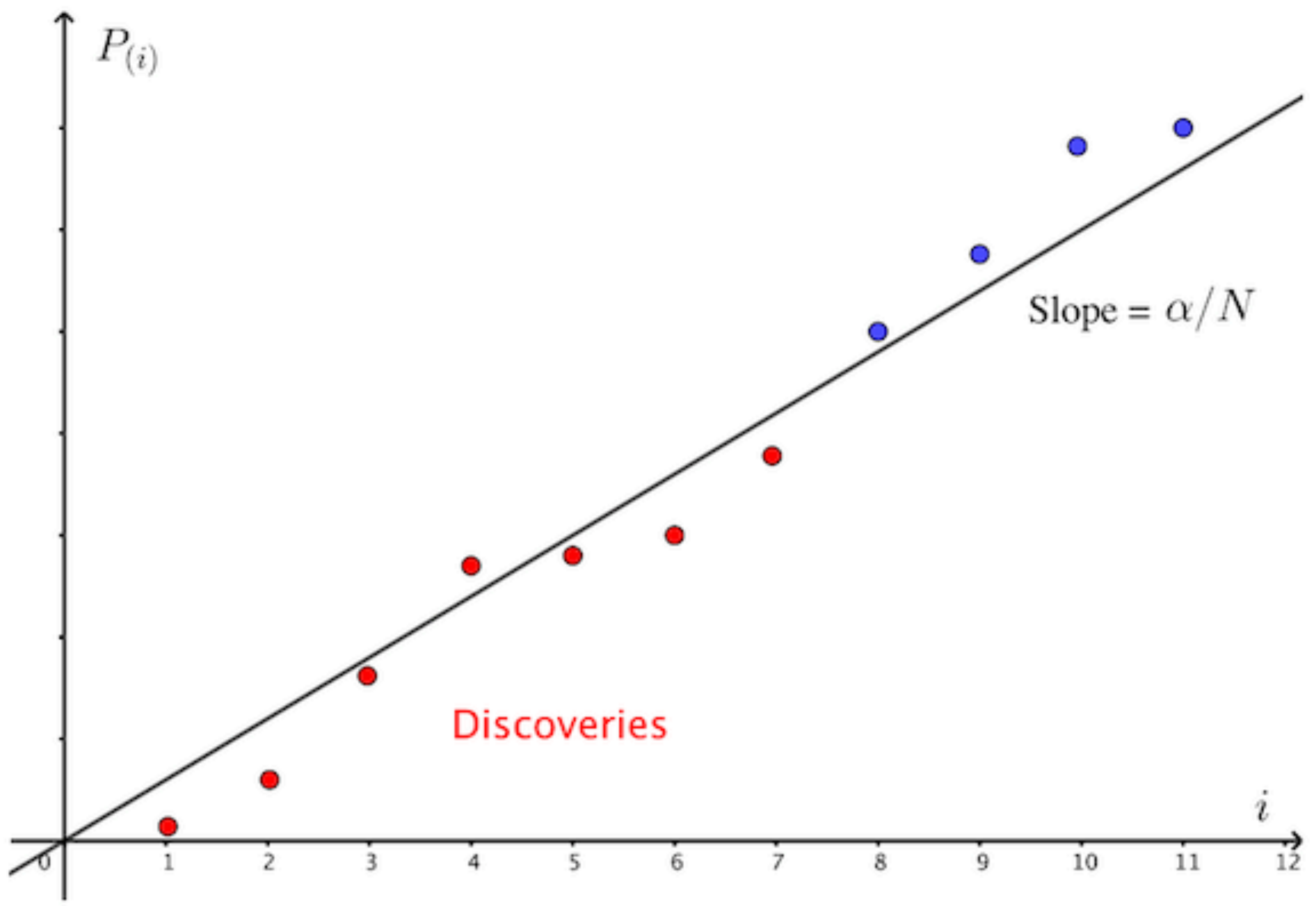
$$FDR = \mathbb{E}\left[\frac{V}{S+V}\right],$$

where V is the number of false discoveries and S is the number of true discoveries.

- For a given confidence level $1-\alpha$ and a collection of ordered p-values $(p_{(k)})_{1\leq k\leq m}$, find the largest k such that:

$$p_{(k)} \leq \frac{k}{m}\alpha\,.$$

- Reject all null hypothesis $\mathscr{H}_{(i)}$ such that $i \leq k$.



Spectral density of a Hawkes process with function h:

$$f(\omega) = m_1^H \frac{1}{|1 - \alpha\tilde{h}(\omega)|^2}$$

Spectral density of a multivariate Hawkes process:

$$f(\omega) = \left(I_d - \tilde{h}(\omega)\right)^{-1} \text{diag}(m^H)\left(I_d - \tilde{h}(-\omega)^T\right)^{-1}$$

Appendix / Bivariate non-identifiable

- Let H be a bivariate exponential Hawkes process ($h_{ij} = \alpha_{ij}\beta_i e^{-\beta_i t}$), P a homogeneous bivariate Poisson process with constant rate λ_0.
- We define the parametric model:

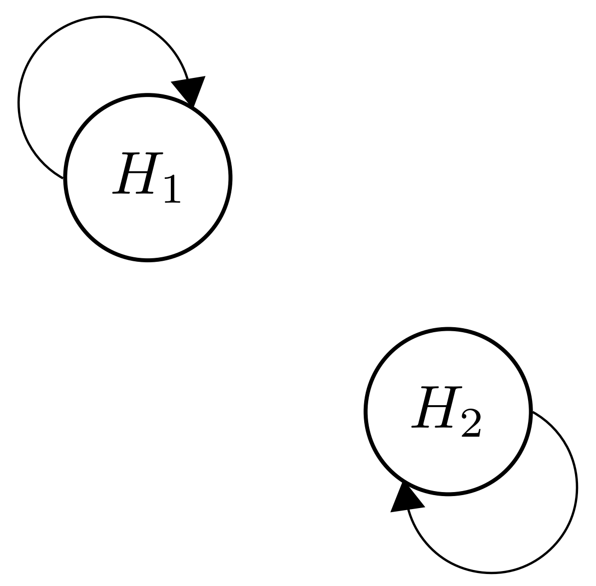
$$\mathcal{Q}_\Lambda = \left\{ f_\theta^N : \mathbb{R} \to \mathbb{C}, \quad \theta = (\mu, \alpha, \beta, \lambda_0) \in \mathbb{R}^2_{>0} \times \Lambda \times \mathbb{R}^2_{>0} \times \mathbb{R}_{>0} \right\}$$

Proposition

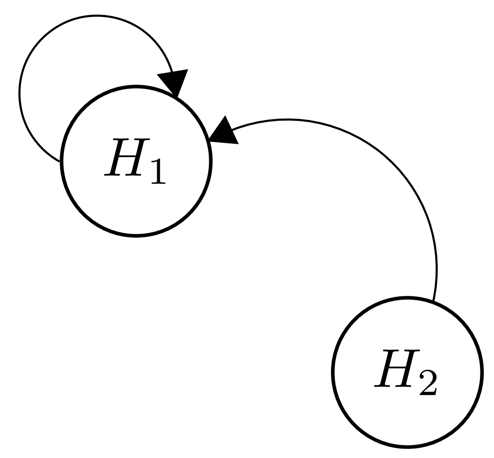
The model $\mathcal{Q}_\Lambda$ is **not** identifiable if:

1. $\Lambda = \left\{ \begin{pmatrix} \alpha_{11} & 0 \\ 0 & \alpha_{22} \end{pmatrix}, \quad 0 \leq \alpha_{11}, \alpha_{22} < 1 \right\}$

2. $\Lambda = \left\{ \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ 0 & 0 \end{pmatrix}, \quad 0 < \alpha_{11} < 1, \alpha_{12} > 0 \right\}$

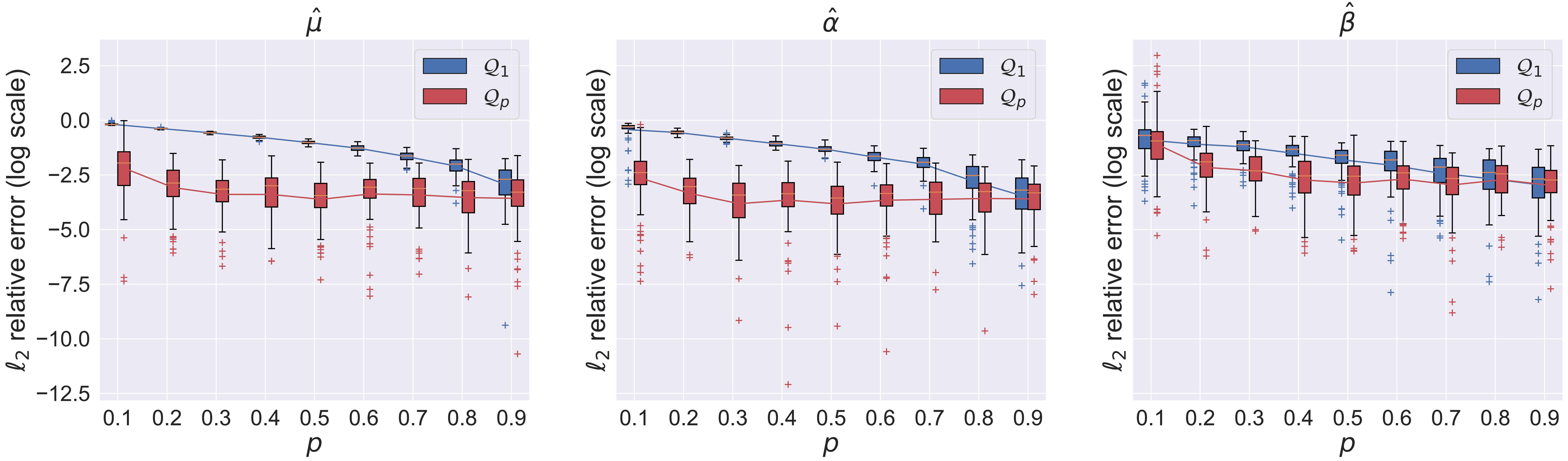


Scenario 1 of non-identifiability

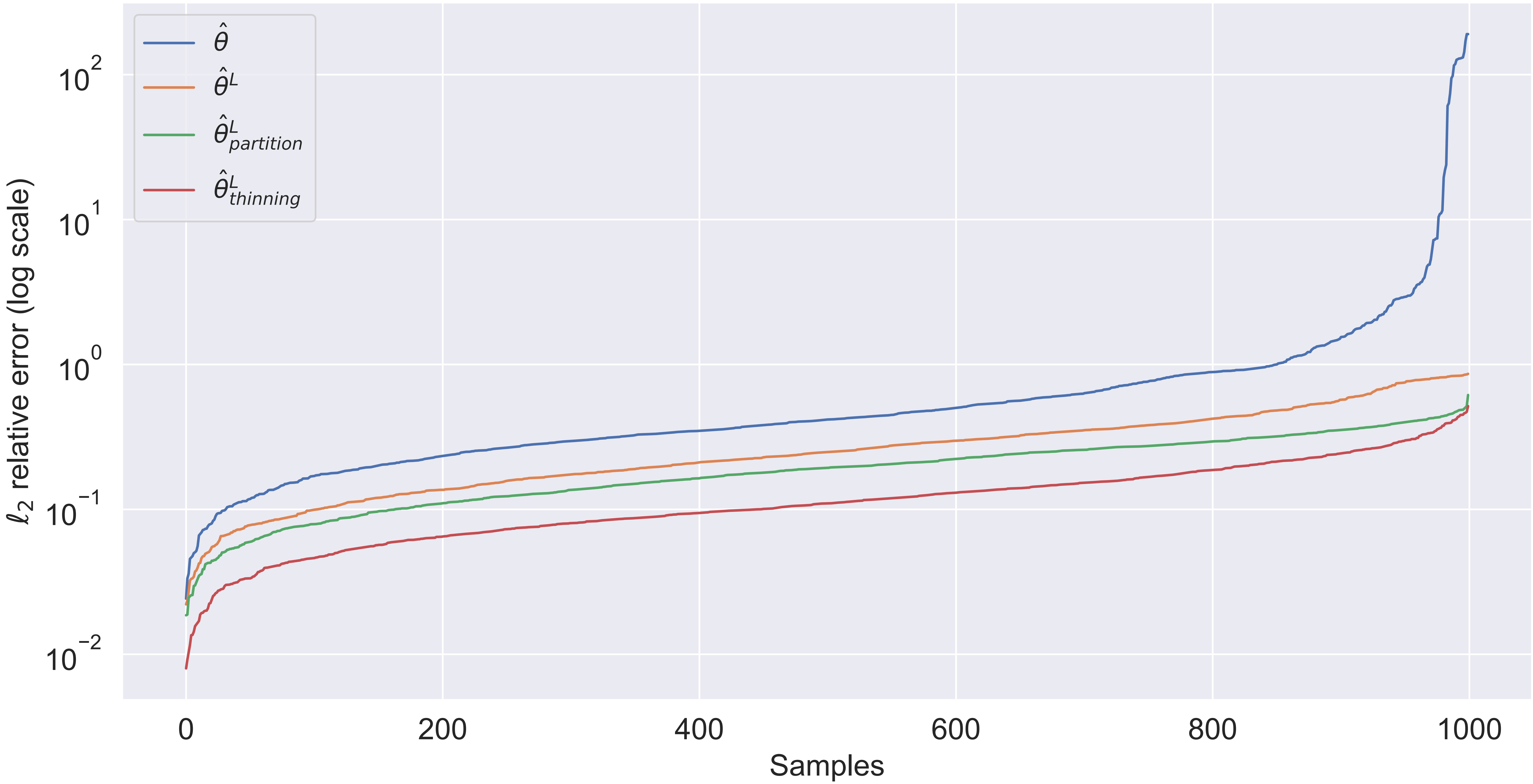


Scenario 2 of non-identifiability

Appendix / Estimation w.r.t. p



Appendix / Distribution of estimations with subsampling



Appendix / Non-parametric estimation of the spectral density of H