# POINT PROCESS AND ITS APPLICATION

## POINT PROCESS III — HAWKES PROCESSES

# review

1. Density, Survival function, Cumulative Distribution Function and Intensity
2. Likelihood
3. Simulation: inversed method and thinning method

# Survival Analysis

- $f$ be the density function of the inter-event time $T$
- Cumulative distribution function

$$F(t) = P(T \leq t) = \int_0^t f(\tau)d\tau$$

- The survival function

$$S(t) = 1 - F(t) = P(T > t) = \int_t^\infty f(\tau)d\tau$$

- Hazard function (intensity)

$$\lambda(t) = \frac{f(t)}{S(t)} = -\frac{1}{dt} d\log S(t)$$

# Example: homogeneous Poisson Process

- homogeneous Poisson Process
- density of inter-event times $t$

$$f(t) = \lambda e^{-\lambda t}$$

- Cumulative distribution function

$$F(t) = \int_0^t f(\tau) d\tau = 1 - e^{-\lambda t}$$

- The survival function

$$S(t) = 1 - F(t) = P(T > t) = e^{-\lambda t}$$

- Hazard function (intensity)

$$\lambda(t) = \frac{f(t)}{S(t)} = \lambda$$

# inhomogeneous Poisson Process

- For inhomogeneous Poisson Process
- Hazard function (intensity) $\lambda(t)$ only depends on $t$
- The survival function

$$\lambda(t) = \frac{f(t)}{S(t)} = -\frac{1}{dt} d \log S(t) \rightarrow S(t) = \exp(-\int_0^t \lambda(\tau)\, d\tau)$$

- Density and Cumulative distribution function

$f(t) = \lambda(t) e^{-\int_0^t \lambda(\tau) d\tau}$ and $F(t) = 1 - \exp(-\int_0^t \lambda(\tau)\, d\tau)$

- Integrated intensity (intensity measure)

$$\Lambda(t) = \int_0^t \lambda(\tau)\, d\tau = EN(t)$$

# Conditional intensity function

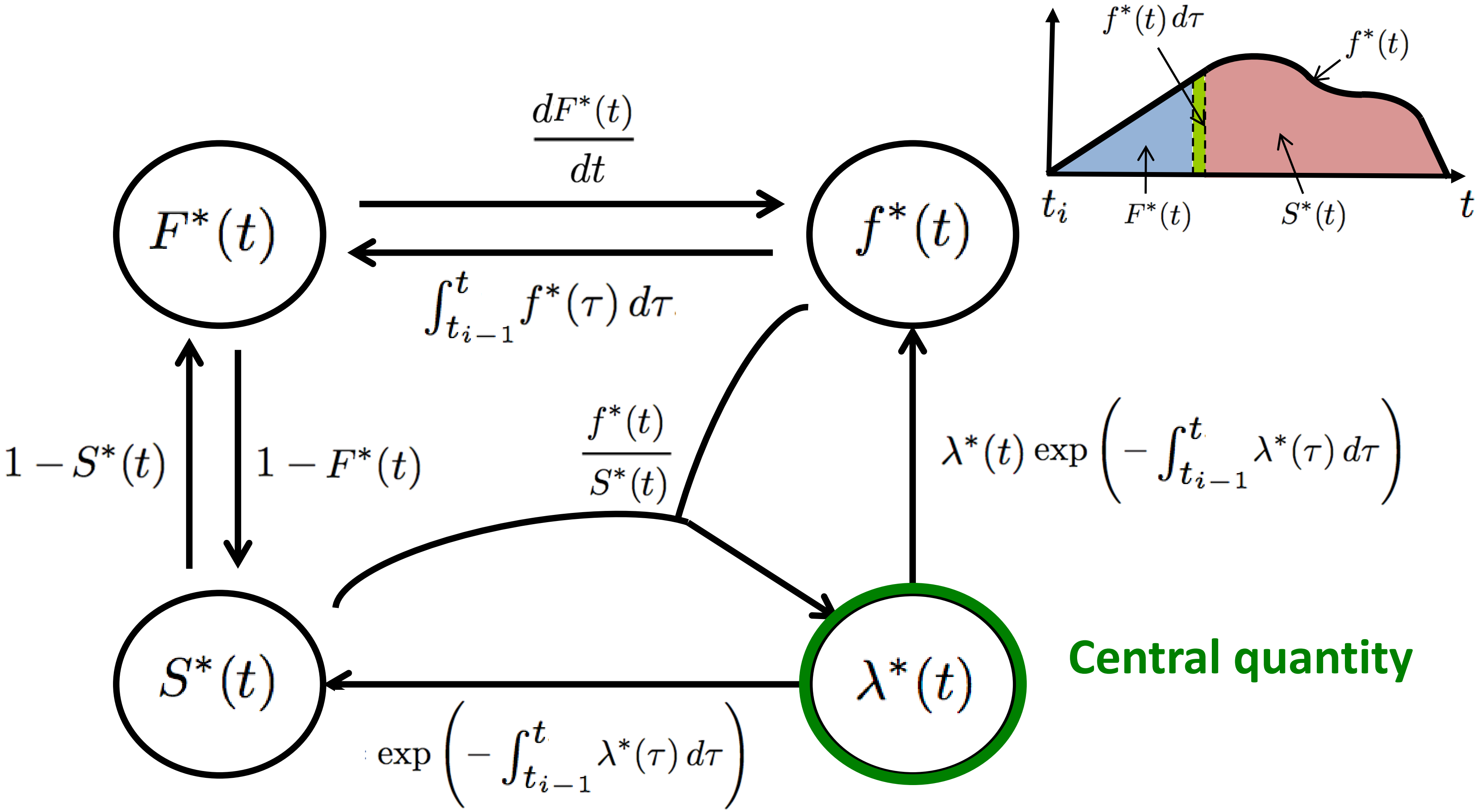
- It turns out the conditional intensity function (hazard function) is more convenient and intuitive (think about exponential distribution)
- conditional intensity function:

The difference with $f(t)$

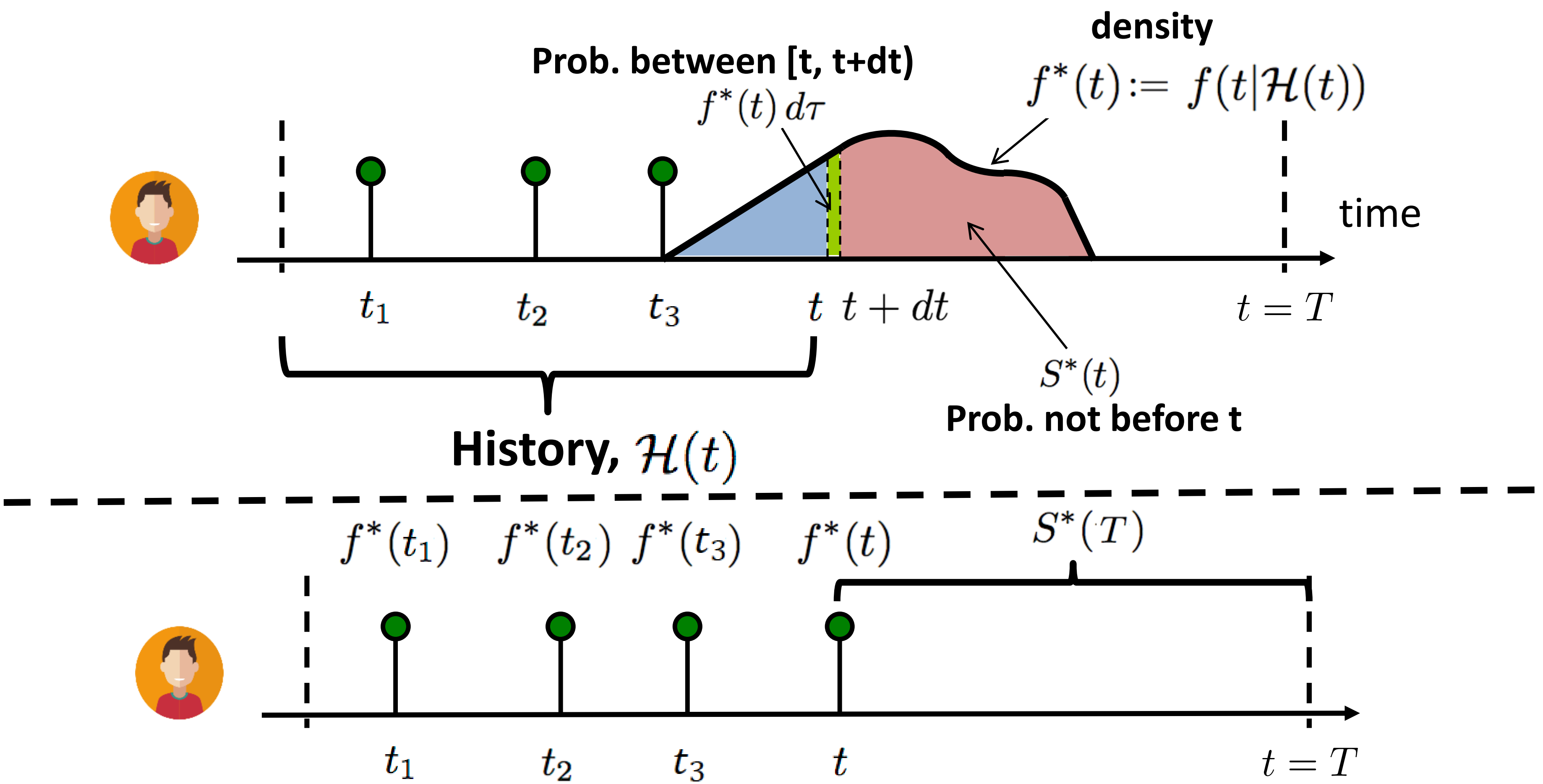
$$\lambda(t|H_{t_n}) = \frac{f(t|H_{t_n})}{1 - F(t|H_{t_n})} = \frac{f^*(t)}{1 - F^*(t)} = \frac{f^*(t)}{S^*(t)}$$

- Notation: $\lambda^*(t) = \lambda(t|H_{t_n})$
- Derivation: $\lambda^*(t) = \frac{f^*(t)}{S^*(t)} = \frac{E[N(t,t+dt)|H_t]}{dt}$ the expectation of event number

# Relation between f*, F*, S*, λ*



# Density and likelihood




**Likelihood of a timeline:** $f^*(t_1)\ f^*(t_2)\ f^*(t_3)\ f^*(t)\ S^*(T)$

# Likelihood

let $t_1 < t_2 < \cdots < t_{n-1} < t_n$, be the event times observed over $[0, T]$, use factorization, we can get the likelihood

$$L = f^*(t_1) \cdot f^*(t_2) \cdot \cdots \cdot f^*(t_n) \cdot S^*(T)$$

$$= \left( \prod_{i=1}^{n} \lambda^*(t_i) \cdot \exp\left( - \int_{t_{i-1}}^{t_i} \lambda^*(s) ds \right) \right) \cdot \exp\left( - \int_{t_n}^{T} \lambda^*(s) ds \right)$$

$$= \left( \prod_{i=1}^{n} \lambda^*(t_i) \right) \cdot \exp\left( - \int_{0}^{T} \lambda^*(s) ds \right)$$

# Simulation—the inversed method

- Algorithm 1. The inverse method algorithm
- 1. set $t = 0$, $t_0 = 0$ , $s_0 = 0$ , $i = 1$
- 2. while $true$:
- (i) generate $U \sim \text{Uniform}([0,1])$
- (ii) calculate $\tau_i = -(\log U)/\lambda$
- (iii) set $s_i = s_{i-1} + \tau_i$
- (iv) calculate $t$ where $t = \Lambda^{*-1}(s_n)$
- (v) if $t < T : i = i + 1, t_i = t$ else break
- Output: Retrieve the simulated process $\{t_n\}$ on $[0, T]$

# Simulation—thinning method

- Algorithm 2. Ogata's modified thinning algorithm
- 1. set $t = 0$, $i = 1$
- 2. while $t \leq T$:
- (i) calculate $m(t), l(t)$
- (ii) generate $U \sim \mathrm{Unif}([0,1])$ then set $s = -(\log U)/\lambda$
  and generate $U' \sim \mathrm{Unif}([0,1])$
- (iii) if: $s > l(t)$, set $t = t + l(t)$
- (iv) elif: $t + s > T$ or $U' > \lambda^*(t+s)/m(t)$, set $t = t + s$
- (v) else: set $n = n + 1$, $t_n = t + s$, $t = t + s$
- Output: Retrieve the simulated process $\{t_n\}$ on $[0, T]$

# outline

1. Introduction
2. One-dimensional Hawkes Process
3. Multi-dimensional Hawkes Process
4. Marked Hawkes Processes

# Event data for point process

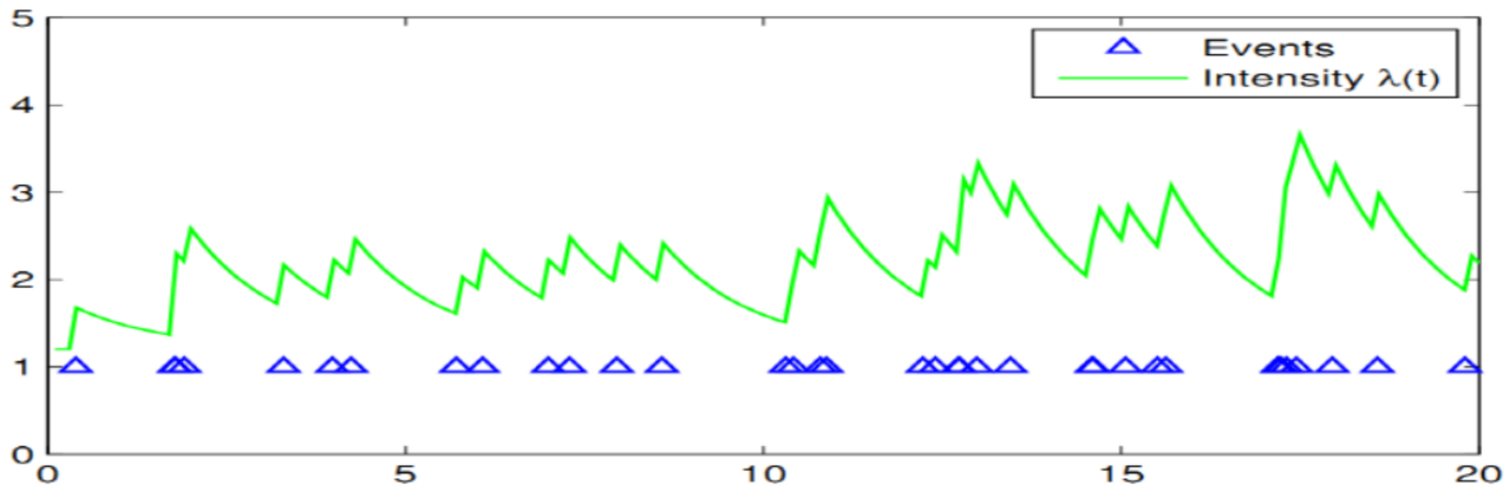
- **Temporal events**: occurrences of events over time

$$t_1 < t_2 < \cdots < t_{n-1} < t_n$$

usually recorded over an observation window $[0, T]$

- **Conditional Intensity function**:

$$\lambda^*(t) = \lambda\left(t|H_{t_n}\right) = \frac{f^*(t)}{1 - F^*(t)} = \frac{d\mathrm{E}[\mathrm{N}(t+h) - N(t)|H_t]}{dh}$$



# temporal point process

- **Counting process**: Given $T_1 < T_2 < \cdots T_{n-1} < T_n$, define counting process

$$N(t) = \sum 1_{T_i < t}$$

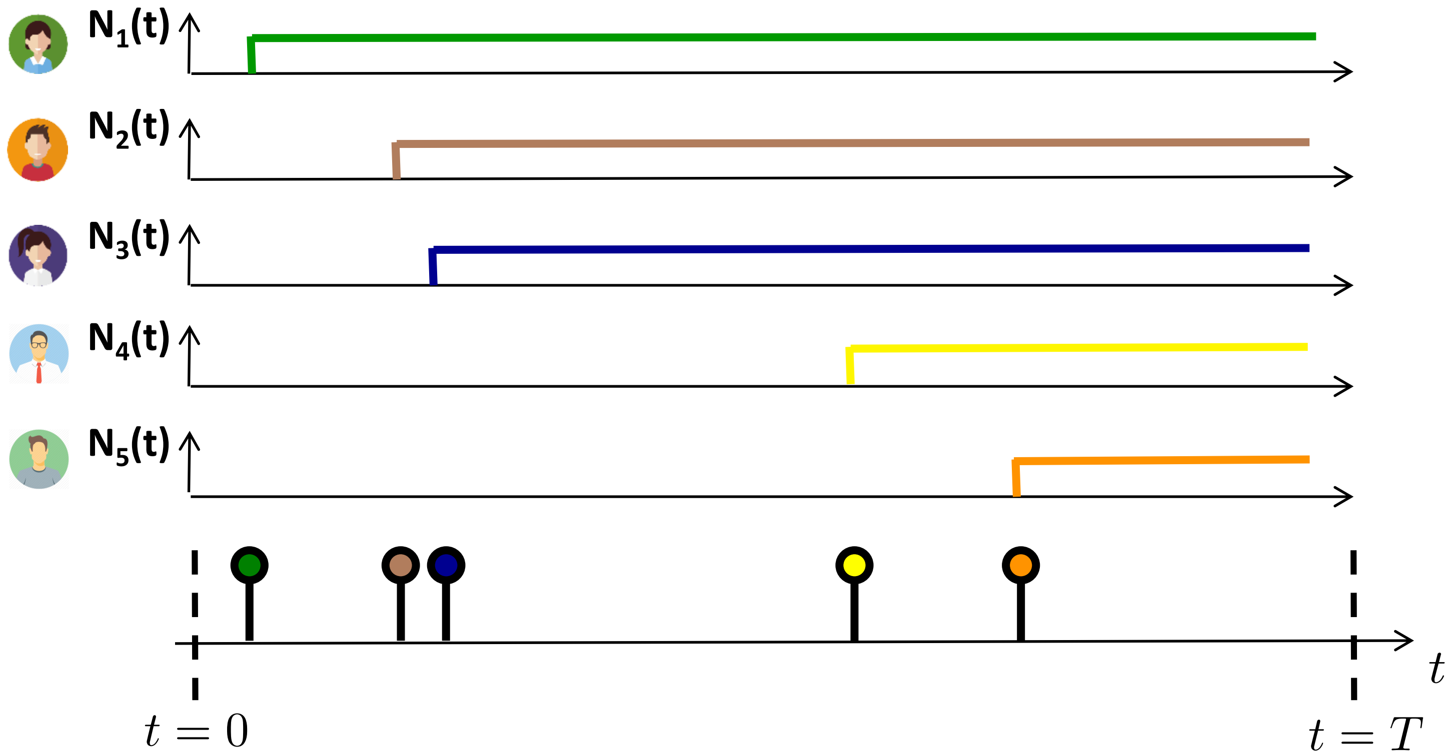
  Right continuous and monotonous without decrement

- **Multi-dimensional point process**

  event data: $\{(t_i^m)_i\}_{m=1}^M$ or $\{(t_1, u_1), (t_2, u_2), \ldots, (t_n, u_n)\}$

  associated counting process $N(t) = (N_1(t), N_2(t), \ldots, N_m(t))$ as M-dimensional point process

# Multi-dimensional point process



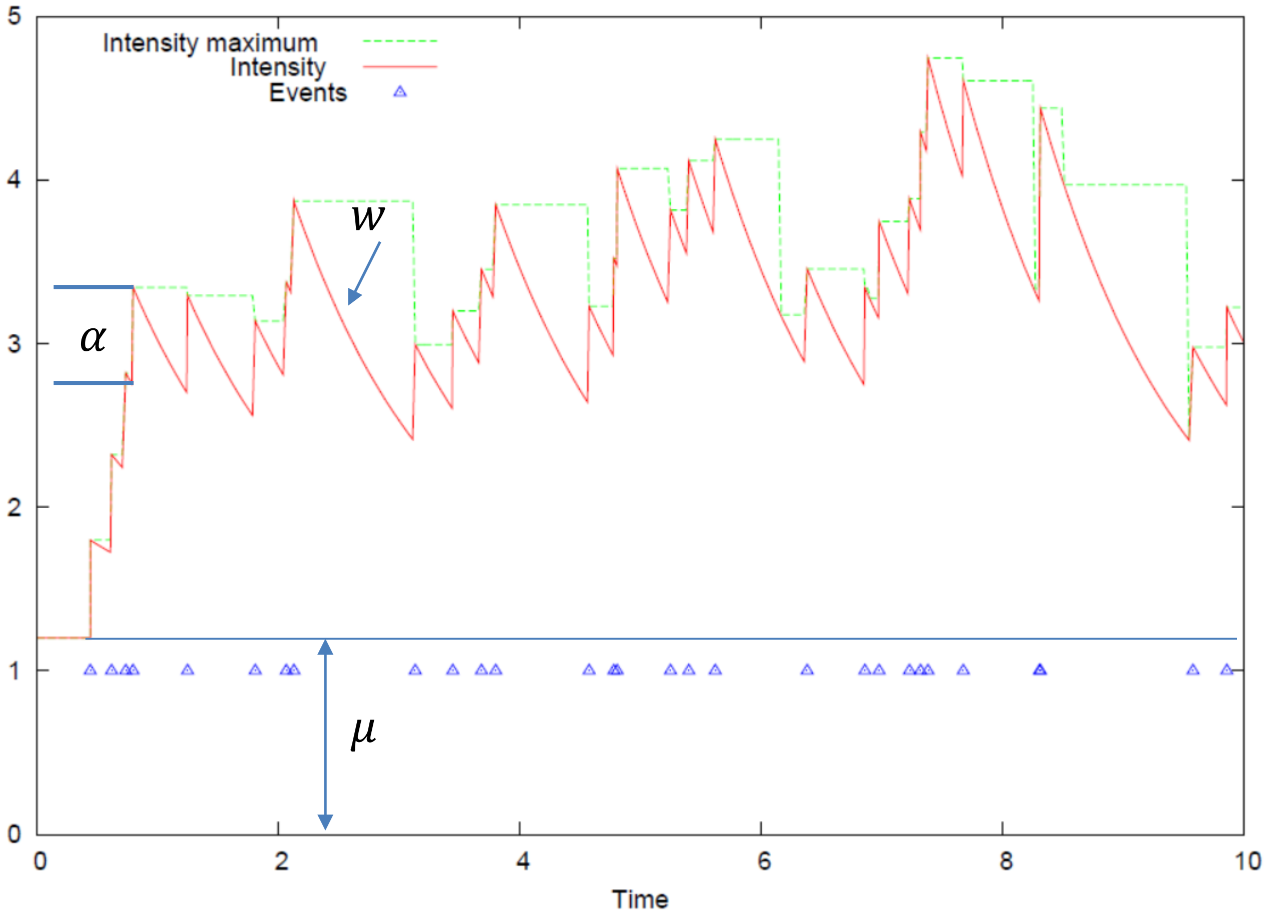
# Hawkes process

- conditional intensity function of Hawkes process: given $t_1 < t_2 < \cdots < t_n$ in the window $[0, T]$

$$\lambda(t) = \mu + \alpha \sum_{t_i < t} \exp(-w(t - t_i))$$

- $\mu, \alpha$ are positive parameters.
- the conditional intensity grows by $\alpha$ and decreases exponentially back towards $\mu$

# Explanations of the parameters



# Hawkes process

- More general formulation (1):

$$\lambda(t) = \mu + \alpha \sum_{t_i < t} g(t - t_i)$$

- $\mu, \alpha$ are positive parameters and $g(t)$ is the decaying kernel, for example $g(t) = \omega e^{-\omega t}$
- More general formulation (2):

$$\lambda(t) = \mu(t) + \alpha \sum_{t_i < t} g(t - t_i)$$

- $\mu$ is time-dependent and $\mu: R_+ \rightarrow R_+$

# Hawkes process

- More general formulation (3):

$$\lambda(t) = \mu(t) + \sum_{t_i<t} \gamma(t - t_i)$$

- $\mu: R \to R_+$ is a deterministic base intensity, for example, $\mu(t) = \mu$ or $\mu(t) = \mu e^{-\omega t} + \mu_0$
- $\gamma: R_+ \to R_+$ expresses the positive influence of the past events $t_i$ on the current value of the intensity process. For example, $\gamma(t) = \sum_{j=1}^{P} \alpha_j e^{-\beta_j t} \cdot 1_{R^+}$.

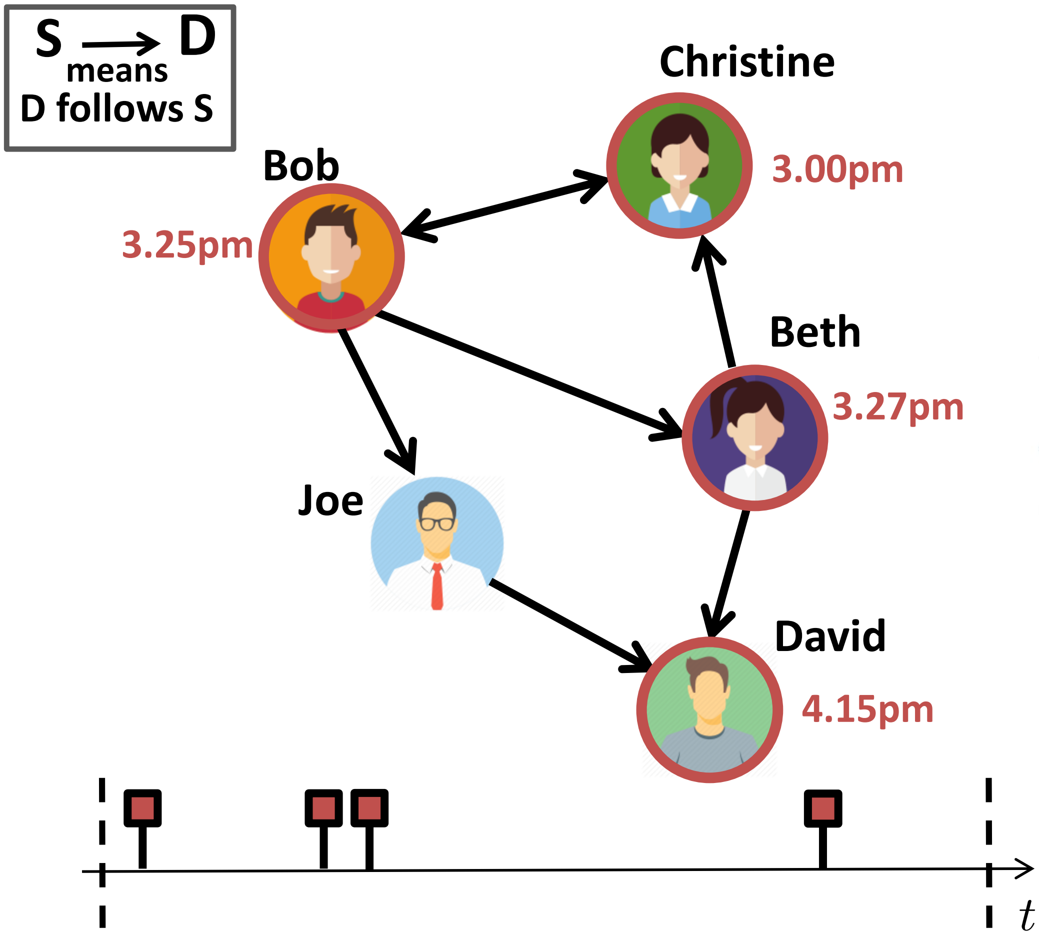
# Multi-dimensional Hawkes process

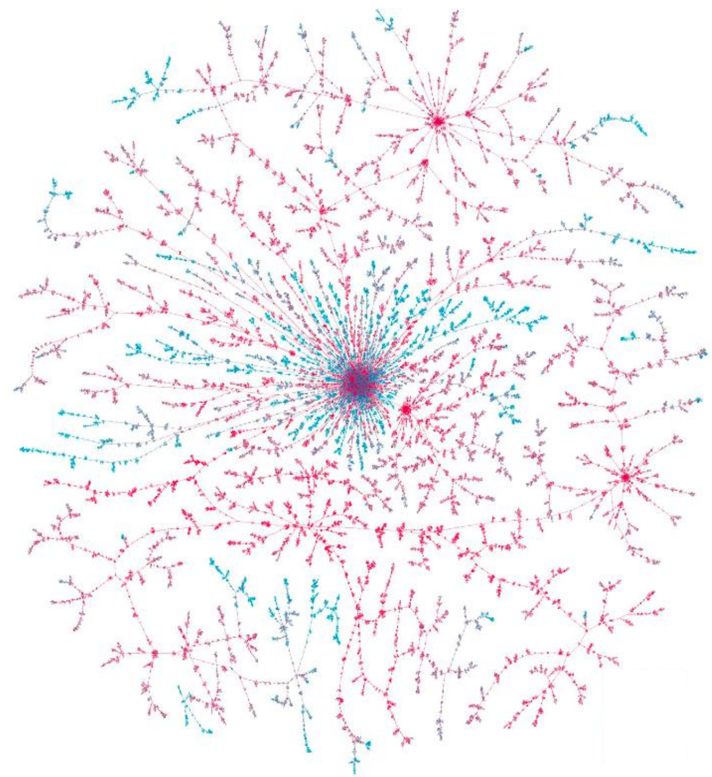
- Intensity of multi-dimensional Hawkes process: given event data $\{(t_i^m)_i\}_{m=1}^M$

$$\lambda_u(t) = \mu_u + \sum_{n=1}^{M} \sum_{t_i^n < t} \alpha_{uu_i} g(t - t_i^n)$$

- where $\mu_u \geq 0$ is the base intensity for the $u$-th Hawkes process
- The coefficient $\alpha_{uu'}$ captures the mutually exciting property between the $u$-th and the $u'$-th dimension. It shows how much influence the events in $u'$-th process have on future events in the $u$-th process.

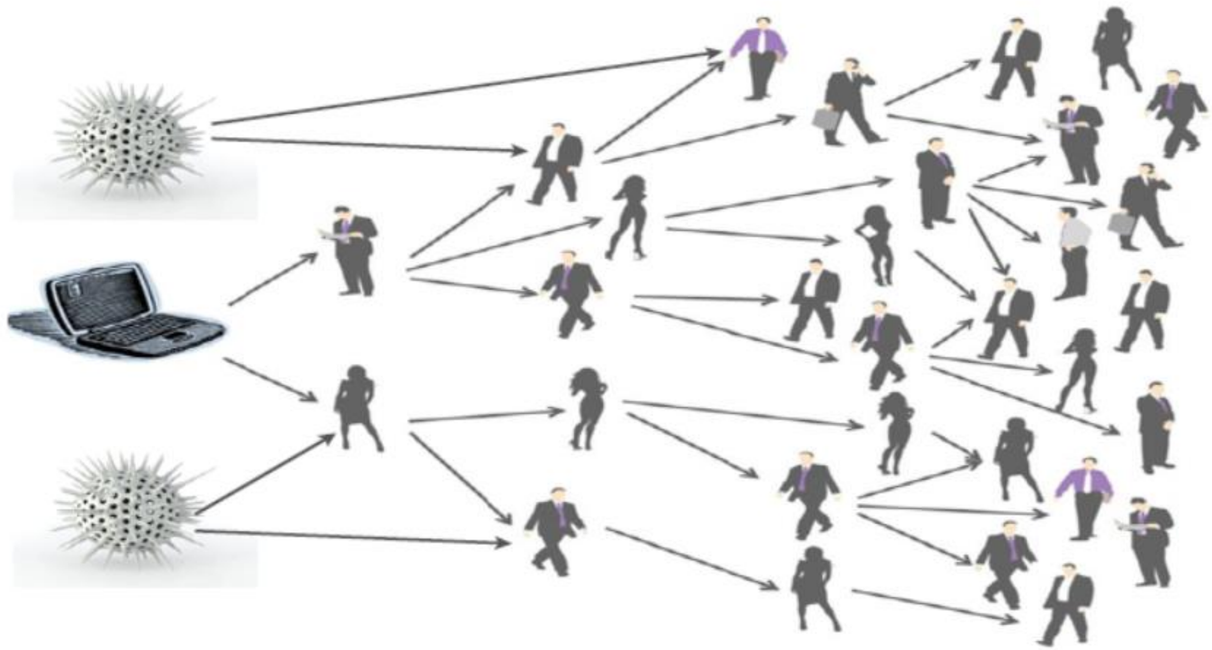
# Application: Information propagation in Social Networks



# Application : Information propagation in Social Networks

- **Multiple memes are evolving and spreading through the same network**
- **Explore the content of the information diffusing through a network**
- **Simultaneous diffusion network inference and meme tracking**



# outline

1. Introduction
2. One-dimensional Hawkes Process
3. Multi-dimensional Hawkes Process
4. Marked Hawkes Processes

# One-dimensional Hawkes Process

- Definition of One-dimensional Hawkes processes
- (also called self-exciting process)
- Intensity:

$$\lambda(t) = \mu(t) + \int_{-\infty}^{t} \gamma(t-s)dN_s = \mu(t) + \sum_{t_i<t} \gamma(t-t_i)$$

where $\mu: R \to R_+$ is a deterministic base intensity and $\gamma: R_+ \to R_+$ expresses the positive influence of the past events $t_i$ on the current value of the intensity process.

# One-dimensional Hawkes Process

- given an exponential kernel $\gamma(t) = \sum_{j=1}^{P} \alpha_j e^{-\beta_j t} \cdot 1_{R^+}$

$$\lambda(t) = \mu(t) + \int_{-\infty}^{t} \sum_{j=1}^{P} \alpha_j e^{-\beta_j (t-s)} \, dN_s$$

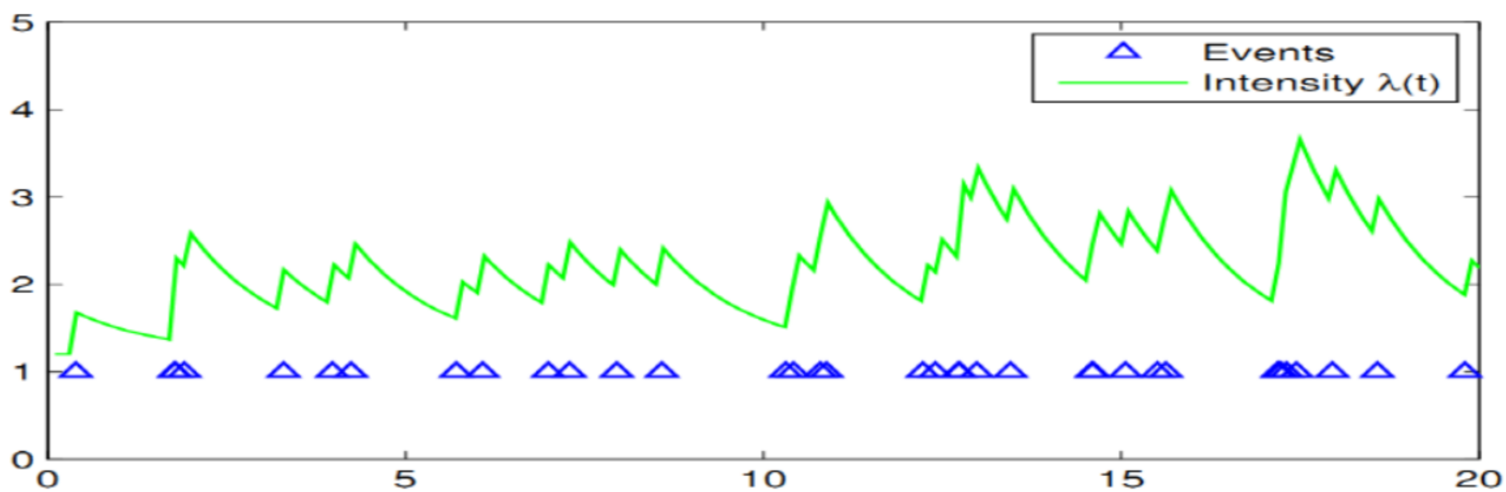
$$= \mu(t) + \sum_{t_i < t} \sum_{j=1}^{P} \alpha_j e^{-\beta_j (t-t_i)}$$

where $\alpha_j, \beta_j$ are positive parameters

# One-dimensional Hawkes Process

- The simplest version with $P = 1$ and $\mu(t) = \mu$ constant

$$\lambda(t) = \mu + \int_{-\infty}^{t} \alpha e^{-\beta(t-s)} dN_s$$
$$= \mu + \sum_{t_i < t} \alpha e^{-\beta(t-t_i)}$$



# Stationarity of One-dimensional Hawkes Process

Assuming stationarity gives $E[\lambda(t)] = m$ constant.

Thus,

$$
\begin{aligned}
m = E[\lambda(t)] &= E\left[\mu + \int_{-\infty}^{t} \gamma(t-s)dN_s\right] \\
&= \mu + E\left[\int_{-\infty}^{t} \gamma(t-s)\lambda(s)ds\right] \\
&= \mu + \int_{-\infty}^{t} \gamma(t-s)\cdot m \cdot ds \\
&= \mu + m\int_{0}^{\infty} \gamma(v)dv
\end{aligned}
$$

# Stationarity of One-dimensional Hawkes Process

So we can get the stationarity that

$$m = \frac{\mu}{1 - \int_0^\infty \gamma(v)dv}$$

When $\gamma(t) = \sum_{j=1}^{P} \alpha_j e^{-\beta_j t} \cdot 1_{R^+}$, so that

$$\lambda(t) = \mu(t) + \sum_{t_i<t} \sum_{j=1}^{P} \alpha_j e^{-\beta_j (t-t_i)}$$

So the stationarity is

$$m = \frac{\mu}{1 - \int_0^\infty \gamma(v)dv} = \frac{\mu}{1 - \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}} > 0$$

# Stationarity of One-dimensional Hawkes Process

Stationarity condition for a 1D-Hawkes process

$$\sum_{j=1}^{P} \frac{\alpha_j}{\beta_j} < 1$$

for the one-dimensional Hawkes process with $P = 1$ the unconditional expected value of the intensity process is :

$$\mathrm{E}[\lambda(t)] = \frac{\mu}{1 - \alpha/\beta}$$

# Ogata's modified thinning method for 1D-Hawkes process

- a thinning algorithm based on simulating homogeneous Poisson processes with too high intensities and then thin out the points
- at time $t$, find next point $t_i > t$
- Simulate a homogeneous Poisson process on interval $[t, t + l(t)]$ for some chosen function $l(t)$
- high constant intensity on $[t, t + l(t)]$ that fullfills

$$m(t) \geq sup_{s \epsilon [t, t + l(t)]}(\lambda^*(s))$$

# Ogata's modified thinning method for 1D-Hawkes process

- 1D-Hawkes process, for example

$$\lambda(t) = \mu + \sum_{t_i<t} \alpha e^{-\beta(t-t_i)}$$

- $\lambda(t)$ monotonous with decrement before next point
- so $l(t) = +\infty$
- $m(t)$ right continuous and piecewise function

# Simulation algorithm

- ①Initialization: $t = 0$, $n = 0$, $m(t) = \lambda(0)$
- ②First event:
- (1)generate $V \sim \mathrm{Uniform}([0,1])$, $\mathrm{t} \sim \mathrm{Exp}(\mathrm{m(t)})$ (i.e. $\mathrm{t} \leftarrow -\frac{1}{\mathrm{m(t)}} logV$)
- (2)if: $\mathrm{t} < \mathrm{T}$: $\mathrm{t}_1 \leftarrow \mathrm{t}$, $\mathrm{n} += 1$
- (3)else: go to last step (empty)
- ③While $\mathrm{t} \leq \mathrm{T}$: (General routine)
- (1)Update maximum intensity $\mathrm{m(t)} = \lambda(\mathrm{t_n}) + \alpha$, α a jump size
- (2) generate $\mathrm{s} \sim \mathrm{Exp}(\mathrm{m(t)})$和$\mathrm{U} \sim \mathrm{Uniform}([0,1])$
- (3)if: $\mathrm{t} + \mathrm{s} > \mathrm{T}$ or $\mathrm{U} > \lambda^*(\mathrm{t} + \mathrm{s})/\mathrm{m(t)}$, let $\mathrm{t} = \mathrm{t} + \mathrm{s}$
- (5)else: 令$\mathrm{n} = \mathrm{n} + 1$, $\mathrm{t_n} = \mathrm{t} + \mathrm{s}$, $\mathrm{t} = \mathrm{t} + \mathrm{s}$
- ④Retrieve the simulated process $\{\mathrm{t}_1, \mathrm{t}_2, \mathrm{t}_3, \ldots, \mathrm{t}_n\}$ on $[0, T]$

# Examples of simulations

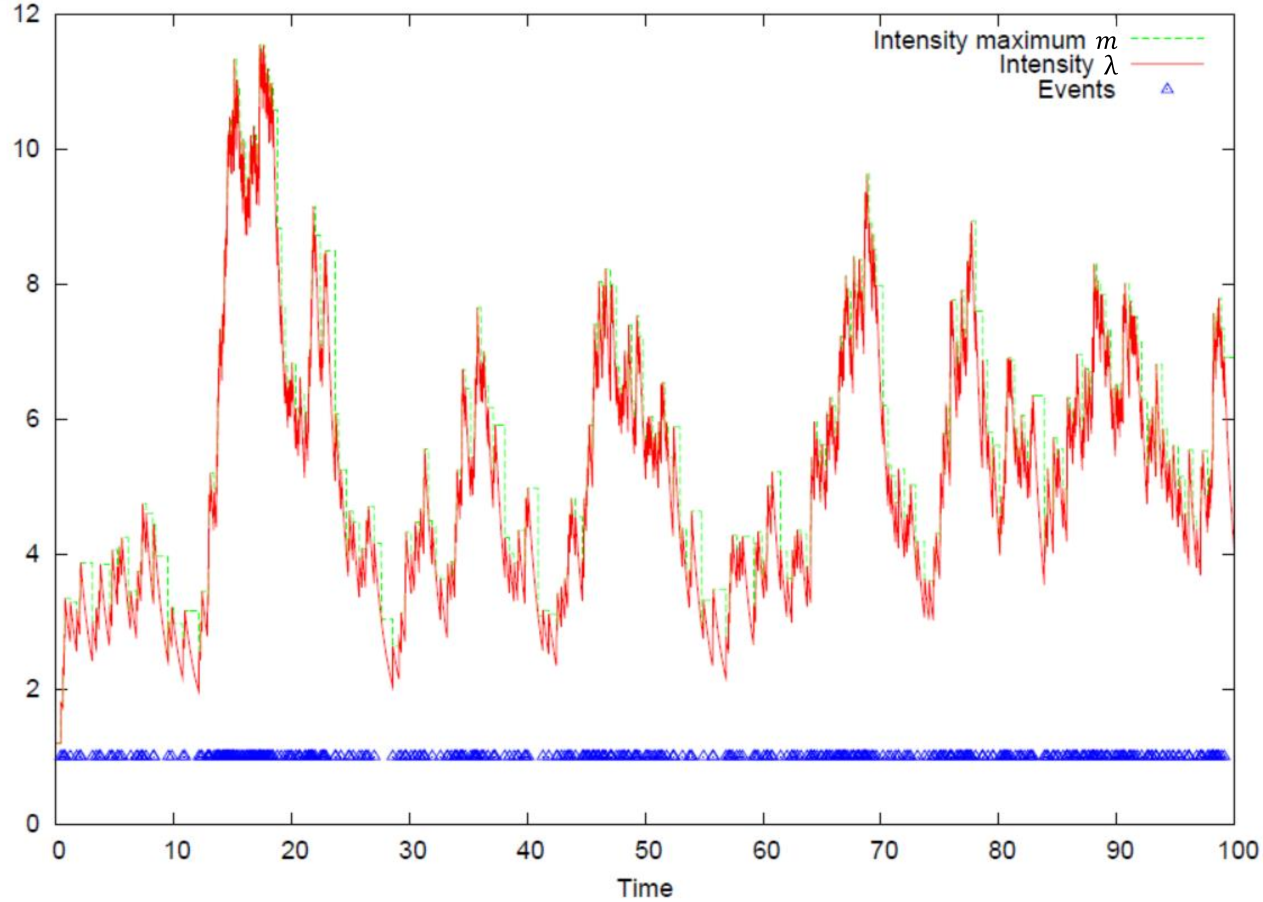



Figure: Simulation of a one-dimensional Hawkes process with parameters $P = 1, \lambda_0 = 1.2, \alpha_1 = 0.6, \beta_1 = 0.8$.

# Examples of simulations

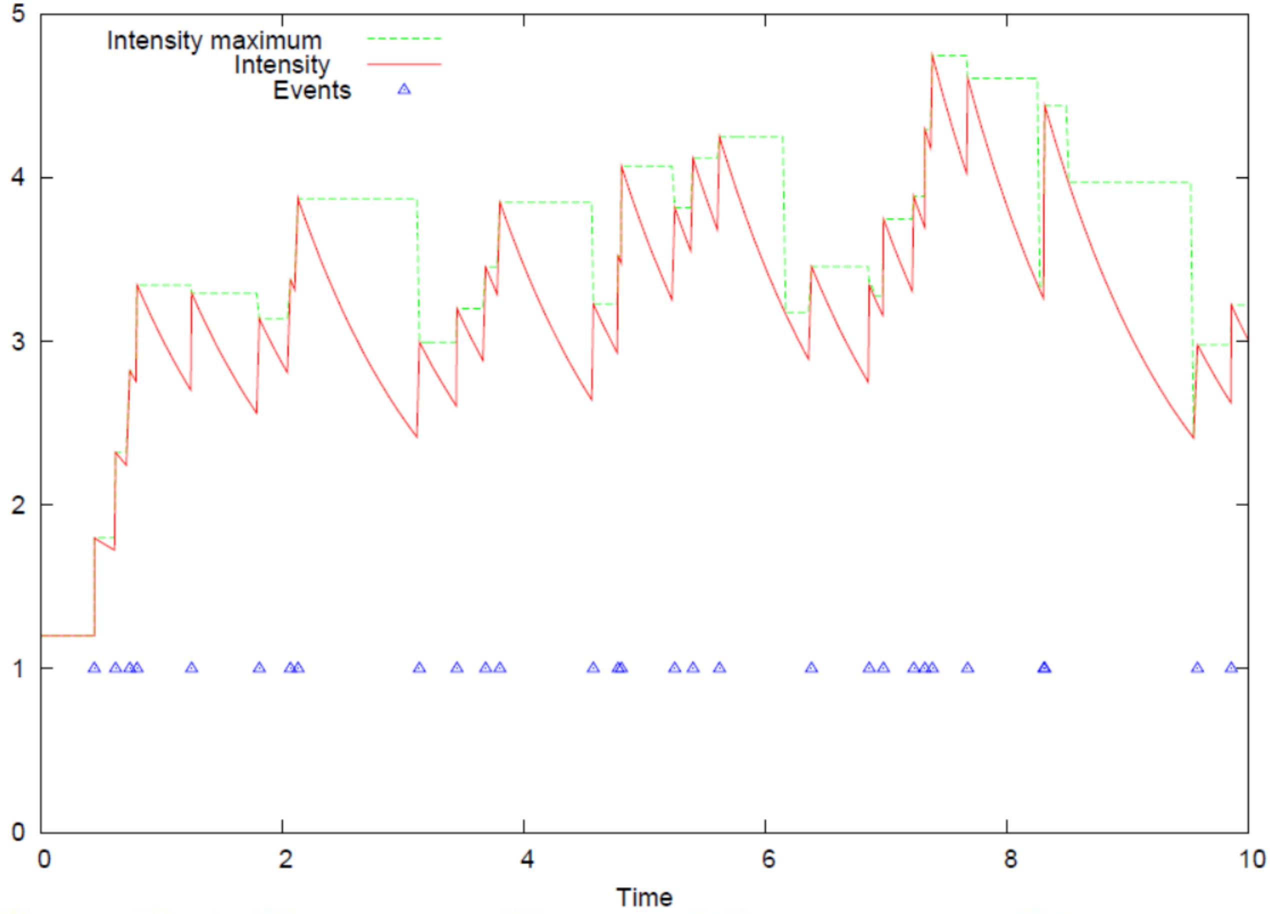



Figure: Simulation of a one-dimensional Hawkes process with parameters $P = 1, \lambda_0 = 1.2, \alpha_1 = 0.6, \beta_1 = 0.8$. (Zoom of the previous figure).

# Testing the simulated process

For any consecutive events $t_{i-1}$ and $t_i$

$$\begin{aligned}
\Lambda(t_{i-1}, t_i) &= \int_{t_{i-1}}^{t_i} \lambda(s) ds \\
&= \int_{t_{i-1}}^{t_i} \mu(s) ds + \int_{t_{i-1}}^{t_i} \sum_{t_k < s} \sum_{j=1}^{P} \alpha_j e^{-\beta_j (s - t_k)} ds \\
&= \int_{t_{i-1}}^{t_i} \mu(s) ds + \int_{t_{i-1}}^{t_i} \sum_{t_k \leq t_{i-1}} \sum_{j=1}^{P} \alpha_j e^{-\beta_j (s - t_k)} ds \\
&= \int_{t_{i-1}}^{t_i} \mu(s) ds + \sum_{t_k \leq t_{i-1}} \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j} (e^{-\beta_j (t_{i-1} - t_k)} - e^{-\beta_j (t_i - t_k)})
\end{aligned}$$

# Testing the simulated process

This computation can be simplified with a recursive element.

Let us denote

$$A_j(i-1) = \sum_{t_k \le t_{i-1}} e^{-\beta_j (t_{i-1} - t_k)}$$

We observe that

$$\begin{aligned} A_j(i-1) &= \sum_{t_k \le t_{i-1}} e^{-\beta_j (t_{i-1} - t_k)} \\ &= 1 + e^{-\beta_j (t_{i-1} - t_{i-2})} \sum_{t_k \le t_{i-1}} e^{-\beta_j (t_{i-1} - t_k)} \\ &= 1 + e^{-\beta_j \left(t_{i-1} - t_{i-2}\right)} A_j(i-2) \end{aligned}$$

# Testing the simulated process

Finally, the integrated density can be written $i \in N$

$$\Lambda(t_{i-1}, t_i) = \int_{t_{i-1}}^{t_i} \mu(s) ds + \sum_{t_k \le t_{i-1}} \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j} (e^{-\beta_j (t_{i-1} - t_k)} - e^{-\beta_j (t_i - t_k)})$$

$$= \int_{t_{i-1}}^{t_i} \mu(s) ds + \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j} (1 - e^{-\beta_j (t_i - t_{i-1})}) A_j(i-1)$$

where $A_j(i-1) = \sum_{t_k \le t_{i-1}} e^{-\beta_j (t_{i-1} - t_k)}$

# Testing the simulated process

Defining $\{\tau_i\}$ as

$$\tau_0 = \int_0^{t_0} \lambda(s)ds = \Lambda(0, t_0)$$

$$\tau_i = \tau_{i-1} + \int_{t_{i-1}}^{t_i} \lambda(s)ds = \tau_{i-1} + \Lambda(t_{i-1}, t_i)$$

the durations $\tau_i - \tau_{i-1} = \Lambda(t_{i-1}, t_i)$ are exponentially distributed

# Maximum-likelihood estimation

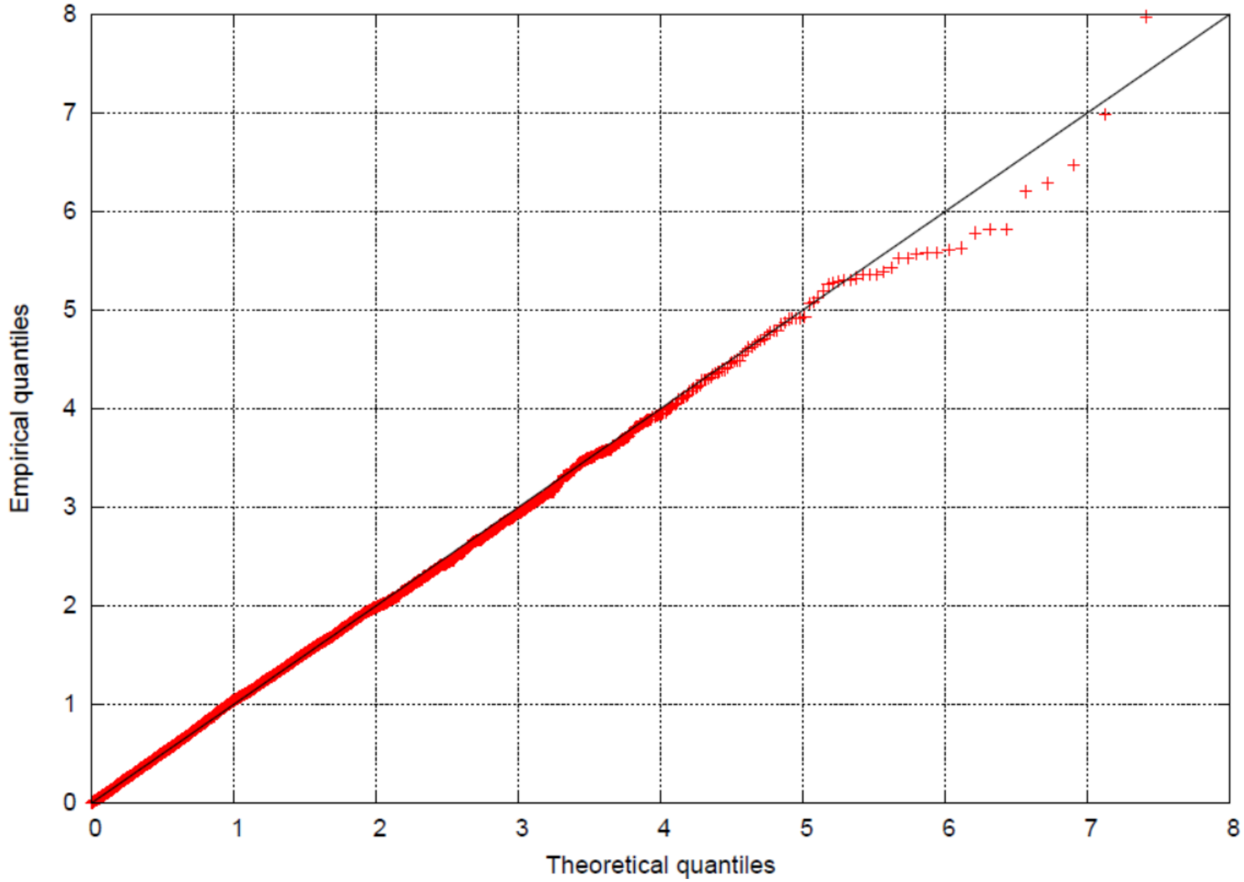



Figure: Quantile plot for one sample of simulated data of a one-dimensional Hawkes process with parameters $P = 1, \lambda_0 = 1.2, \alpha_1 = 0.6, \beta_1 = 0.8$, on an interval $[0, 10000]$.

# Maximum-likelihood estimation

1D-Hawkes process, for example

$$\lambda(t) = \mu + \sum_{t_i<t} \alpha e^{-\beta(t-t_i)}$$

The likelihood is

$$L = \exp\left(-\int_0^T \lambda(s)ds\right)\prod_{i=1}^{n}\lambda(t_i)$$

The log-likelihood is

$$\log L = \sum_{i=1}^{n} \log \lambda(t_i) - \int_0^T \mu + \sum_{t_i<t} \alpha e^{-\beta(t-t_i)}\, dt$$

$$= \sum_{i=1}^{n} \log \lambda(t_i) - \left(\mu T + \sum_{j=1}^{n} \alpha G(T-t_j)\right)$$

# Maximum-likelihood estimation

The log-likelihood for three parameters $\mu, \alpha, \beta$

$$\log L = \sum_{i=1}^{n} \log \lambda(t_i) - \int_0^T \mu + \sum_{t_i<t} \alpha e^{-\beta(t-t_i)}\, dt$$

$$= \sum_{i=1}^{n} \log \lambda(t_i) - \left( \mu T + \sum_{j=1}^{n} \alpha G\big(T - t_j\big) \right)$$

$$= \sum_{i=1}^{n} \log \left( \mu + \sum_{j=1}^{i-1} \alpha e^{-\beta\left(t_i - t_j\right)} \right) - \left( \mu T + \sum_{j=1}^{n} \alpha G\big(T - t_j\big) \right)$$

where $G(t) = \int_0^t e^{-\beta\tau} d\tau = -\frac{1}{\beta}(e^{-\beta t} - 1)$

# EM algorithm

Then use EM algorithm to estimate the parameter

Jensen's inequality:

$$\log L = \sum_{i=1}^{n} \log\left(\mu + \sum_{j=1}^{i-1} \alpha e^{-\beta(t_i - t_j)}\right) - \left(\mu T + \sum_{j=1}^{n} \alpha G(T - t_j)\right)$$

$$\geq \underbrace{\sum_{i=1}^{n}\left(p_{ii}\log\frac{\mu}{p_{ii}} + \sum_{j=1}^{i-1} p_{ij}\log\frac{\alpha e^{-\beta(t_i - t_j)}}{p_{ij}}\right) - \left(\mu T + \sum_{j=1}^{n} \alpha G(T - t_j)\right)}_{\text{lower bound}}$$

# EM algorithm

## For E-step

$$p_{ii}^{(k+1)} = \frac{\mu^{(k)}}{\mu^k + \sum_{j=1}^{i-1} \alpha^{(k)} e^{-\beta^{(k)}(t_i - t_j)}}$$

$$p_{ij}^{(k+1)} = \frac{\alpha^{(k)} e^{-\beta^{(k)}(t_i - t_j)}}{\mu^k + \sum_{j=1}^{i-1} \alpha^{(k)} e^{-\beta^{(k)}(t_i - t_j)}}$$

The probability that the event $i$ is triggered by the base intensity $\mu$

The probability that the event $i$ is triggered by the event $j$

# EM algorithm

For M-step (do partial differential equation for $\mu$ and $\alpha$)

$$\mu^{(k+1)} = \frac{1}{T}\sum_{i=1}^{n} p_{ii}^{(k+1)}$$

$$\alpha^{(k+1)} = \frac{\sum_{i>j} p_{ij}^{(k)}}{\sum_{j=1}^{n} G\left(T - t_j\right)}$$

For $\beta$, if $e^{-\beta(T-t_i)} \approx 0$

$$\beta^{(k+1)} = \frac{\sum_{i>j} p_{ij}^{(k+1)}}{\sum_{i>j}(t_i - t_j)p_{ij}^{(k+1)}}$$

# Inference

- Why not other optimal algorithm based on gradient?
- For example: SGD, ADAM
- For log-likelihood of event data

$$\log L = \sum_{i=1}^{n} \log\left(\mu + \sum_{j=1}^{i-1} \alpha e^{-\beta(t_i - t_j)}\right) - \left(\mu T + \sum_{j=1}^{n} \alpha G(T - t_j)\right)$$

- the objective is very long for (most) event datasets.

# outline

1. introduction
2. One-dimensional Hawkes Process
3. Multi-dimensional Hawkes Process
4. Marked Hawkes Processes

# Multi-dimensional Hawkes Process

Multi-dimensional Hawkes process $N(t) = (N_1(t), N_2(t), \dots, N_M(t))$ is defined with intensities $\lambda_m$ for $m = 1,2,\dots,M$ given by

$$\lambda_m(t) = \mu_m(t) + \sum_{n=1}^{M} \int_0^t \sum_{j=1}^{P} \alpha_{mn}^j e^{-\beta_{mn}^j (t-s)}\, dN_s$$

in its simplest version with $P = 1$ and $\mu_m(t)$ constant

$$\lambda_m(t) = \mu_m + \sum_{n=1}^{M} \int_0^t \alpha_{mn} e^{-\beta_{mn}\,(t-s)} dN_s$$
$$= \mu_m + \sum_{n=1}^{M} \sum_{t_i^n < t} \alpha_{mn} e^{-\beta_{mn}(t-t_i^n)}$$

Multi-dimensional Hawkes process is also called multivariate Hawkes process

# Multi-dimensional Hawkes Process

Take $P = 1$ here to simplify the notations. Rewriting the intensities of multi-dimensional Hawkes processes using vectorial notation, we have :

$$\boldsymbol{\lambda}(t) = \boldsymbol{\mu} + \int_0^t \boldsymbol{G}(t-s)d\boldsymbol{N}_s$$

where

$$\boldsymbol{G}(t) = \left(\alpha_{mn}e^{-\beta_{mn}t}\right)_{m,n=1,\dots,M}$$

Assuming stationarity gives $\boldsymbol{\nu} = E[\boldsymbol{\lambda}(t)]$ constant vector, and thus stationary intensities must satisfy :

$$\boldsymbol{\nu} = \left(\boldsymbol{I} - \int_0^\infty \boldsymbol{G}(u)du\right)^{-1} \boldsymbol{\mu}$$

# Stationarity of Multivariate Hawkes process

A sufficient condition for a multivariate Hawkes process to be stationary is that the spectral radius of the matrix

$$\Gamma = \int_0^\infty \boldsymbol{G}(u)du = \left(\frac{\alpha_{mn}}{\beta_{mn}}\right)_{m,n=1,\dots,M}$$

be strictly smaller than 1.

Recall the spectral radius of the matrix G is defined as :

$$\rho(G) = \max_{a \in S(G)} |a|$$

where $S(G)$ denotes the set of all eigenvalues of $G$

# Simulation of multivariate Hawkes process

Define the sum of the intensities of the first *K* components of the multivariate process as

$$I^K(t) = \sum_{n=1}^{K} \lambda_n(t)$$

$I^M(t) = \sum_{n=1}^{M} \lambda_n(t)$ is thus the total intensity of the multivariate process and we set $I^0 = 0$, so

$$I^M(0) = \sum_{n=1}^{M} \mu_n(0)$$

# Simulation algorithm

- ①Initialization： set $i = 1, i^1 = 1, i^2 = 1, \dots, i^M = 1, I^* = I^M(0)$
- ②First event：
- (1) generate $V \sim \mathrm{Uniform}([0,1])$， $t \sim \mathrm{Exp}(I^*)$ (i.e. $t \leftarrow -\frac{1}{m(t)} logV$)
- (2) if $t > T$: go to last step (empty)
- (3) Attribution Test: generate $D \sim \mathrm{Uniform}([0,1])$ and set $t_1^{n_0} = t$

where $n_0$ satisfies that $\frac{I^{n_0-1}}{I^*} \leq D \leq \frac{I^{n_0}}{I^*}$

# Simulation algorithm

- ③While $\mathrm{true}$: (General routine) set $i^{n_0} = i^{n_0} + 1$ and $i = i + 1$
- (1)Update maximum intensity: $I^* = I^M(t) + \sum_{n=1}^{M} \sum_{j=0}^{P} \alpha_{nn_0}^{j}$
- (2)New event: generate $s \sim \mathrm{Exp}(I^*), \mathrm{t} += s$ and $U \sim \mathrm{Uniform}([0,1])$
- if: $t > T$, then go to last step (break )
- (3)Attribution-Rejection test : generate $D \sim \mathrm{Uniform}([0,1])$
- if: $D \leq \frac{I^{n_0-1}}{I^*}$,
- then set $t_{i^{n_0}}^{n_0} = t$ where $n_0$ satisfy $\frac{I^{n_0-1}}{I^*} \leq D \leq \frac{I^{n_0}}{I^*}$
- ④Retrieve the simulated process $\{t_i^n\}$ on $[0, T]$

# Simulation

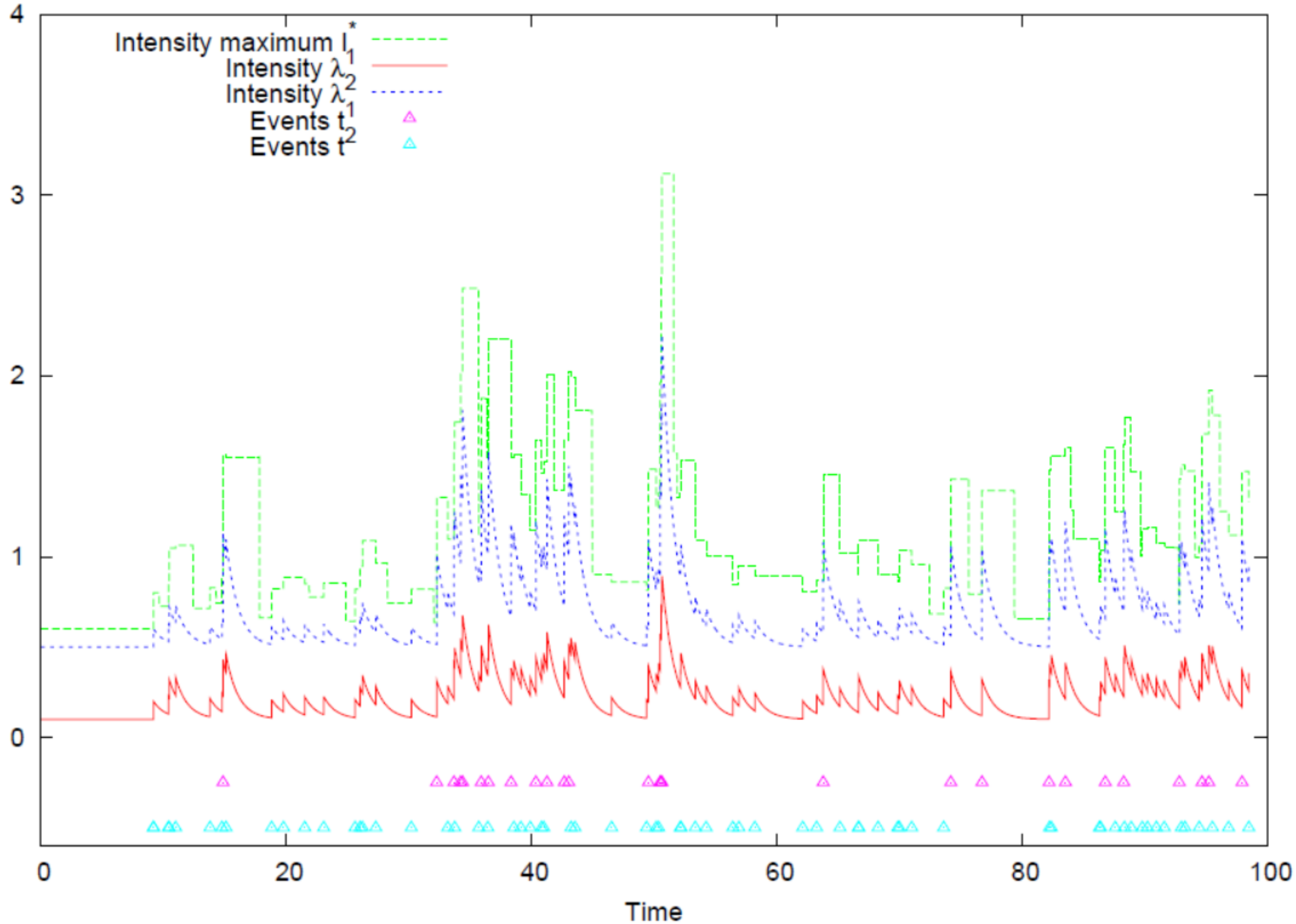



Figure: Simulation of a two-dimensional Hawkes process with P = 1

# Simulation

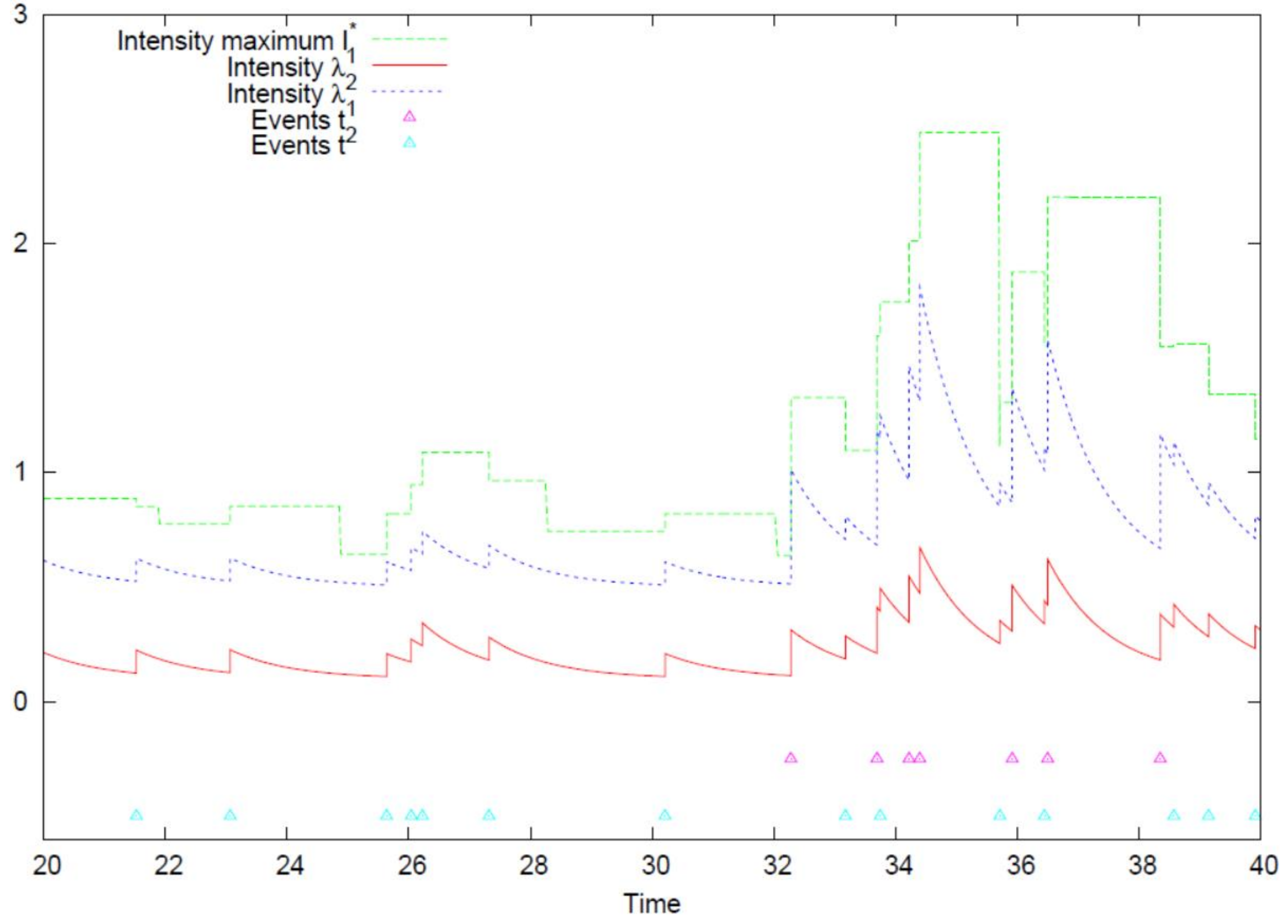



Figure: Simulation of a two-dimensional Hawkes process with P = 1 (Zoom of the previous figure).

# Testing the simulated data

The integrated intensity of the $m$-th coordinate of a multidimensional Hawkes process between two consecutive events $t_{i-1}^m$ and $t_i^m$ of type $m$ is computed as:

$$\Lambda_m(t_{i-1}, t_i) = \int_{t_{i-1}^m}^{t_i^m} \lambda_m(s) ds$$

$$= \int_{t_{i-1}^m}^{t_i^m} \mu_m(s) ds + \int_{t_{i-1}^m}^{t_i^m} \sum_{n=1}^{M} \sum_{t_k^n < s} \sum_{j=1}^{P} \alpha_{mn}^j e^{-\beta_{mn}^j (s - t_k^n)} ds$$

$$= \int_{t_{i-1}^m}^{t_i^m} \mu_m(s) ds + \int_{t_{i-1}^m}^{t_i^m} \sum_{n=1}^{M} \sum_{t_k^n < t_{i-1}^m} \sum_{j=1}^{P} \alpha_{mn}^j e^{-\beta_{mn}^j (s - t_k^n)} ds$$

$$+ \int_{t_{i-1}^m}^{t_i^m} \sum_{n=1}^{M} \sum_{t_{i-1}^m < t_k^n < s} \sum_{j=1}^{P} \alpha_{mn}^j e^{-\beta_{mn}^j (s - t_k^n)} ds$$

# Testing the simulated data

$$\Lambda_m(t_{i-1}, t_i) = \int_{t_{i-1}^m}^{t_i^m} \mu_m(s)ds + \int_{t_{i-1}^m}^{t_i^m} \sum_{n=1}^{M} \sum_{t_k^n < t_{i-1}^m} \sum_{j=1}^{P} \alpha_{mn}^j e^{-\beta_{mn}^j (s - t_k^n)}\, ds$$

$$+ \int_{t_{i-1}^m}^{t_i^m} \sum_{n=1}^{M} \sum_{t_{i-1}^m \le t_k^n < s} \sum_{j=1}^{P} \alpha_{mn}^j e^{-\beta_{mn}^j (s - t_k^n)}\, ds$$

$$= \int_{t_{i-1}^m}^{t_i^m} \mu_m(s)ds + \sum_{n=1}^{M} \sum_{t_k^n < t_{i-1}^m} \sum_{j=1}^{P} \frac{\alpha_{mn}^j}{\beta_{mn}^j} \left(e^{-\beta_{mn}^j (t_{i-1}^m - t_k^n)} - e^{-\beta_{mn}^j (t_i^m - t_k^n)}\right)$$

$$+ \sum_{n=1}^{M} \sum_{t_{i-1}^m \le t_k^n < s} \sum_{j=1}^{P} \frac{\alpha_{mn}^j}{\beta_{mn}^j} \left(1 - e^{-\beta_{mn}^j (t_i^m - t_k^n)}\right.$$

# Testing the simulated data

This computation can be simplified with a recursive element.

Let us denote

$$A_j^{mn}(i-1) = \sum_{t_k^n < t_{i-1}^m} e^{-\beta_{mn}^j (t_{i-1}^m - t_k^n)}$$

We observe that

$$\begin{aligned}
A_j^{mn}(i-1) &= \sum_{t_k^n < t_{i-1}^m} e^{-\beta_{mn}^j (t_{i-1}^m - t_k^n)} \\
&= e^{-\beta_{mn}^j (t_{i-1}^m - t_{i-2}^m)} \sum_{t_k^n < t_{i-2}^m} e^{-\beta_{mn}^j (t_{i-1}^m - t_k^n)} + \sum_{t_{i-2}^m \le t_k^n < t_{i-1}^m} e^{-\beta_{mn}^j (t_{i-1}^m - t_k^n)} \\
&= e^{-\beta_{mn}^j (t_{i-1}^m - t_{i-2}^m)} A_j^{mn}(i-2) + \sum_{t_{i-2}^m \le t_k^n < t_{i-1}^m} e^{-\beta_{mn}^j (t_{i-1}^m - t_k^n)}
\end{aligned}$$

# Testing the simulated data

Finally, the integrated density can be written $i \in N$

$$\Lambda_m(t_{i-1}, t_i) = \int_{t_{i-1}^m}^{t_i^m} \lambda_m(s)ds$$

$$= \int_{t_{i-1}^m}^{t_i^m} \mu_m(s)ds + \sum_{n=1}^{M} \sum_{j=1}^{P} \frac{\alpha_{mn}^j}{\beta_{mn}^j} [(1 - e^{-\beta_{mn}^j (t_i^m - t_{i-1}^m)}) \times A_j^{mn}(i-1)$$

$$+ \sum_{t_{i-1}^m \leq t_k^n < s} (1 - e^{-\beta_{mn}^j (t_i^m - t_k^n)})]$$

Time change property:

the durations $\tau_i^m - \tau_{i-1}^m = \Lambda_m(t_{i-1}, t_i)$ are exponentially distributed

# Testing the simulated data

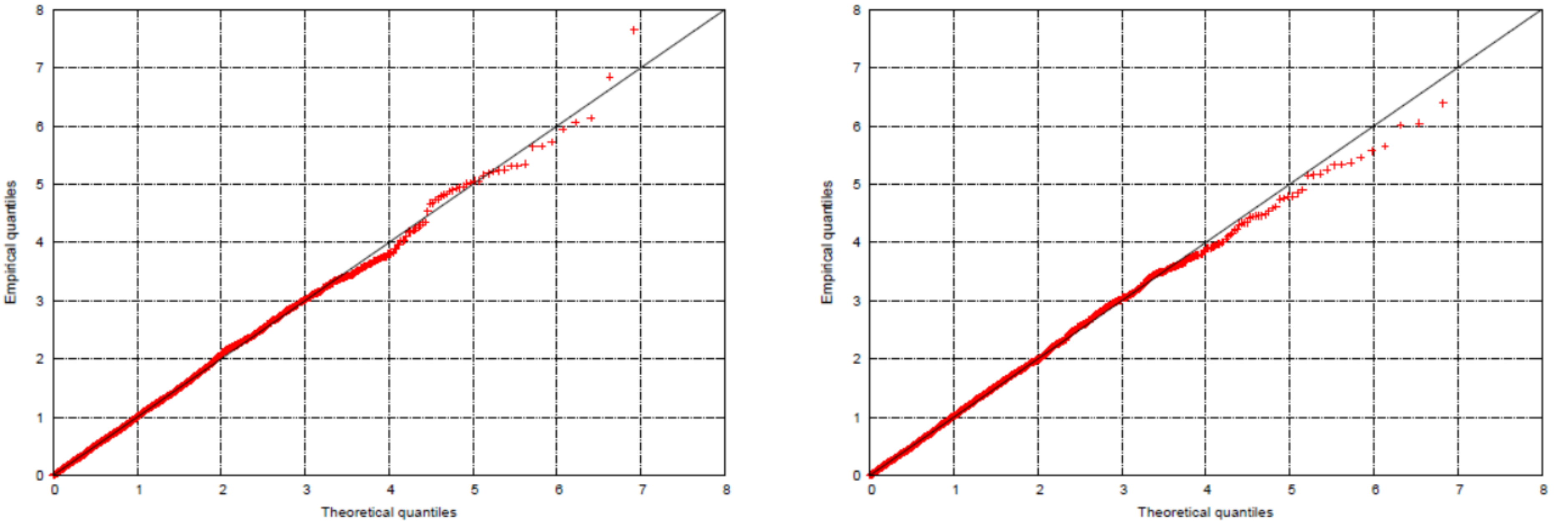



Figure: Quantile plots for one sample of simulated data of a two-dimensional Hawkes process with P = 1. (Left) m = 0.(Right) m = 1.

# Maximum-likelihood estimation

For MLE, we use the data $\{t_i, d_i\}_{i=1}^{N}$ instead of $\{(t_i^m)_i\}_{m=1}^{M}$

For a simple multi-dimensional Hawkes processes:

$$\lambda_d = \mu_d + \sum_{i:t_i<t} \alpha_{dd_i} e^{-\beta(t-t_i)}$$

log-likelihood:

$$\log L = \sum_{d=1}^{M} \left\{ \sum_{(t_i,d_i)|d_i=d} \log \lambda_{d_i}(t_i) - \int_0^T \lambda_d(t)dt \right\}$$

$$= \sum_{i=1}^{n} \log\left( \mu_{d_i} + \sum_{t_j<t_i} \alpha_{d_i d_j} e^{-\beta(t_i-t_j)} \right) - T \sum_{d=1}^{M} \mu_d - \sum_{d=1}^{M} \sum_{j=1}^{n} \alpha_{dd_j} G_{dd_j}(T-t_j)$$

# Maximum-likelihood estimation

Jensen equality:

$$\log L = \sum_{i=1}^{n} \log\left(\mu_{d_i} + \sum_{t_j<t_i} \alpha_{d_i d_j} e^{-\beta(t_i-t_j)}\right) - T\sum_{d=1}^{M}\mu_d - \sum_{d=1}^{M}\sum_{j=1}^{n}\alpha_{dd_j}G_{dd_j}(T-t_j)$$

$$\geq \underbrace{\sum_{i=1}^{n}\left(p_{ii}\log\frac{\mu_{d_i}}{p_{ii}} + \sum_{j=1}^{i-1}p_{ij}\log\frac{\alpha_{d_i d_j}e^{-\beta(t_i-t_j)}}{p_{ij}}\right) - T\sum_{d=1}^{M}\mu_d - \sum_{d=1}^{M}\sum_{j=1}^{n}\alpha_{dd_j}G_{dd_j}(T-t_j)}_{\text{lower bound}}$$

# Maximum-likelihood estimation

So for E-step

$$p_{ii}^{(k+1)} = \frac{\mu_{d_i}^{(k)}}{\mu_{d_i}^{(k)} + \sum_{j=1}^{i-1} \alpha_{d_i d_j}{}^{(k)(k)} e^{-\beta(t_i - t_j)}}$$

$$p_{ij}^{(k+1)} = \frac{\alpha^{(k)} e^{-\beta(t_i - t_j)}}{\mu_{d_i}^{(k)} + \sum_{j=1}^{i-1} \alpha_{d_i d_j}{}^{(k)(k)} e^{-\beta(t_i - t_j)}}$$

The probability that the event $i$ is triggered by the base intensity $\mu$

The probability that the event $i$ is triggered by the event $j$

# Maximum-likelihood estimation

M-step (do partial differential equation for $\mu$ and $\alpha$)

$$\mu_d^{(k+1)} = \frac{1}{T}\sum_{i=1,d_i=d}^{n} p_{ii}^{(k+1)}$$

$$\alpha_{uv}^{(k+1)} = \frac{\sum_{i=1,d_i=u}^{n}\sum_{j=1,d_j=v}^{i-1} p_{ij}^{(k+1)}}{\sum_{j=1,d_j=v}^{n} G\left(T-t_j\right)}$$

For $\beta$, if $e^{-\beta(T-t_i)} \approx 0$

$$\beta^{(k+1)} = \frac{\sum_{i>j} p_{ij}^{(k+1)}}{\sum_{i>j}(t_i - t_j)p_{ij}^{(k+1)}}$$

# Application

- For $\alpha_{ij}$, influence from dimension $i$ to $j$
- Social Infectivity
- If high dimension, overfitting for $\boldsymbol{A} = [\alpha_{ij}]$
- Sparse Low-rank Networks
- regularize the maximum likelihood estimator

$$\min_{A \geq 0, \mu \geq 0} -L(A, \mu) + \lambda_1 \|A\|_* + \lambda_2 \|A\|_1$$

- $\|A\|_*$ is the nuclear norm of matrix A, which is defined to be the sum of its singular value

# Sparse Low-rank

- Optimization with Sparse Low-rank constraint

$$\min_{A\geq 0,\mu\geq 0} -L(A,\mu)+\lambda_1\|A\|_*+\lambda_2\|A\|_1$$

- which are equivalent by introducing two auxiliary variables $Z_1$ and $Z_2$

$$\min_{A\geq 0,\mu\geq 0,Z_1,Z_2} -L(A,\mu)+\lambda_1\|Z_1\|_*+\lambda_2\|Z_2\|_1$$

- In ADMM, we optimize the argumented Lagrangian of the above problem

dual variable

$$L_\rho = -L(A,\mu)+\lambda_1\|Z_1\|_*+\lambda_2\|Z_2\|_1+\rho trace\left(U_1^T(A-Z_1)\right)$$
$$+\rho trace\left(U_2^T(A-Z_2)\right)+\frac{\rho}{2}(\|A-Z_1\|^2+\|A-Z_2\|^2)$$

# Sparse Low-rank

- solving the above augmented Lagrangian problem involves the following key iterative steps

$$
\begin{aligned}
A^{k+1}, \mu^{k+1} &= \operatorname*{argmin}_{A \geq 0, \mu \geq 0} L_\rho(A, \mu, Z_1^k, Z_2^k, U_1^k, U_2^k) \\
Z_1^{k+1} &= \operatorname*{argmin}_{A \geq 0, \mu \geq 0} L_\rho(A^{k+1}, \mu^{k+1}, Z_1, Z_2^k, U_1^k, U_2^k) \\
Z_2^{k+1} &= \operatorname*{argmin}_{A \geq 0, \mu \geq 0} L_\rho(A^{k+1}, \mu^{k+1}, Z_1^k, Z_2, U_1^k, U_2^k) \\
U_1^{k+1} &= U_1^k + (A^{k+1} - Z_1^{k+1}) \\
U_2^{k+1} &= U_2^k + (A^{k+1} - Z_2^{k+1})
\end{aligned}
$$

# Sparse Low-rank — Solving for $Z_1$ and $Z_2$

- When solving for $Z_1$ for equation

$$Z_1^{k+1} = \underset{A\geq 0, \mu \geq 0}{\operatorname{argmin}} L_\rho(A^{k+1}, \mu^{k+1}, Z_1, Z_2^k, U_1^k, U_2^k)$$

- the relevant terms from $L_\rho$ are

$$\lambda_1 \|Z_1\|_* + \rho trace\left((U_1^k)^T\left(A^{k+1} - Z_1\right)\right) + \frac{\rho}{2}(\left\|A^{k+1} - Z_1\right\|^2)$$

- which can be simplified to an equivalent problem

$$\lambda_1 \|Z_1\|_* + \frac{\rho}{2}\left(\left\|A^{k+1} - Z_1 + U_1^k\right\|^2\right)$$

- The above problem has a closed form solution

$$Z_1^{k+1} = S_{\lambda_1/\rho}(A^{k+1} + U_1^k)$$

- where $S_\alpha$ is a soft-thresholding function defined as

# Sparse Low-rank — Solving for $Z_1$ and $Z_2$

$$\lambda_1 \|Z_1\|_* + \frac{\rho}{2}\left(\left\|A^{k+1} - Z_1 + U_1^k\right\|^2\right)$$

- which can be simplified to an equivalent problem

$$Z_1^{k+1} = \underset{Z_1}{\operatorname{argmin}}\, \lambda_1 \|Z_1\|_* + \frac{\rho}{2}\left(\left\|A^{k+1} - Z_1 + U_1^k\right\|^2\right)$$

- The above problem has a closed form solution

$$Z_1^{k+1} = S_{\lambda_1/\rho}(A^{k+1} + U_1^k)$$

- where $S_\alpha$ is a soft-thresholding function defined as

$$S_\alpha(X) = U\big((\sigma_i - \alpha)_+\big)V^T$$

- for all matrix X with singular value decomposition

$$X = U(\sigma_i)V^T$$

# Sparse Low-rank — Solving for $Z_1$ and $Z_2$

- Similarly, the optimization for $Z_2$ can be simplified into the following equivalent form

$$Z_2^{k+1} = \underset{A \geq 0, \mu \geq 0}{\operatorname{argmin}} \lambda_2 \|Z_2\|_1 + \frac{\rho}{2}\left(\left\|A^{k+1} - Z_2 + U_2^k\right\|^2\right)$$

- In this case, depending on the magnitude of the ij-th entry of the matrix $A^{k+1} + U_2^k$, the corresponding $(Z_2^{k+1})_{ij}$ is updated as

$$\begin{cases} (A^{k+1} + U_2^k)_{ij} - \frac{\lambda_2}{\rho} & (A^{k+1} + U_2^k)_{ij} \geq \frac{\lambda_2}{\rho} \\ (A^{k+1} + U_2^k)_{ij} + \frac{\lambda_2}{\rho} & (A^{k+1} + U_2^k)_{ij} \leq -\frac{\lambda_2}{\rho} \\ 0 & else \end{cases}$$

# Sparse Low-rank — Solving for $A, \mu$

- The optimization problem is equivalent to

$$A^{k+1}, \mu^{k+1} = \underset{A \geq 0, \mu \geq 0}{\operatorname{argmin}} f(A, \mu)$$

- Where $f(A, \mu) = -L(A, \mu) + \frac{\rho}{2}\left(\left\|A - Z_1 + U_1^k\right\|^2 + ||A^{k+1} - Z_2 + U_2^k||^2\right)$. Then similar above with EM algorithm

$$\mathrm{Q}\left(A, \mu \middle| A^{(k)}, \mu^{(k)}\right) = \sum_{i=1}^{n}\left(p_{ii}\log\frac{\mu_{d_i}}{p_{ii}} + \sum_{j=1}^{i-1} p_{ij}\log\frac{\alpha_{d_i d_j} e^{-\beta(t_i - t_j)}}{p_{ij}}\right)$$

$$-T\sum_{d=1}^{M}\mu_d - \sum_{d=1}^{M}\sum_{j=1}^{n}\alpha_{dd_j}G_{dd_j}(T - t_j) + \frac{\rho}{2}\left(\left\|A - Z_1 + U_1^k\right\|^2 + ||A^{k+1} - Z_2 + U_2^k||^2\right)$$

where

$$p_{ii}^{(k+1)} = \frac{\mu_{d_i}^{(k)}}{\mu_{d_i}^{(k)} + \sum_{j=1}^{i-1} \alpha_{d_i d_j}{}^{(k)}{}^{(k)} e^{-\beta(t_i - t_j)}}$$

$$p_{ii}^{(k+1)} = \frac{\alpha^{(k)} e^{-\beta(t_i - t_j)}}{\mu_{d_i}^{(k)} + \sum_{j=1}^{i-1} \alpha_{d_i d_j}{}^{(k)}{}^{(k)} e^{-\beta(t_i - t_j)}}$$

and

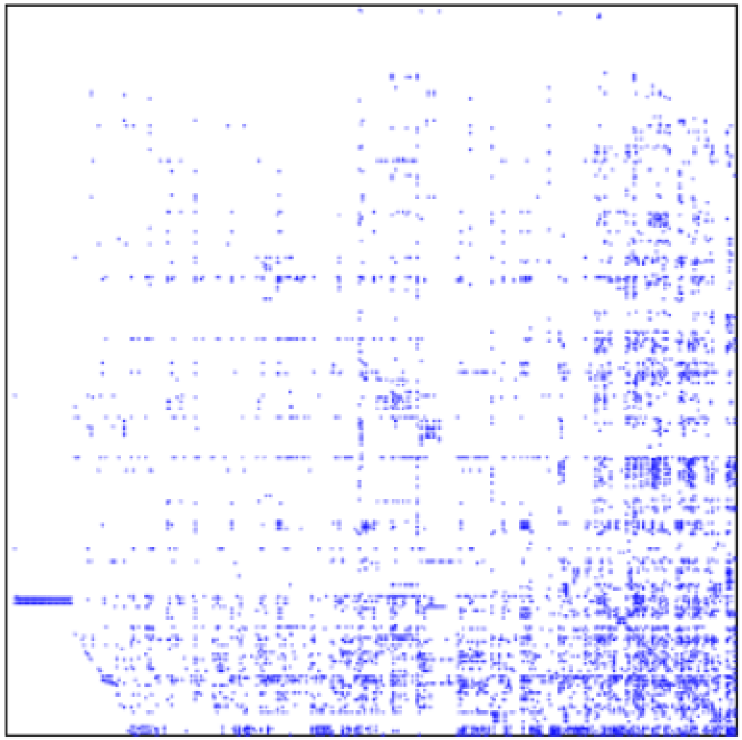
$$\mu_d^{(k+1)} = \frac{1}{T} \sum_{i=1, d_i = d}^{n} p_{ii}^{(k+1)}$$

$$\alpha_{uv}^{(k+1)} = \frac{-B + \sqrt{B^2 + 8\rho C}}{4\rho}$$

where

$$B = \sum_{j=1, d_j=v}^{n} G\left(T - t_j\right) + \rho(-z_{1,uv} + u_{1,uv} - z_{2,uv} + u_{2,uv})$$

$$C = \sum_{i=1, d_i=u}^{n} \sum_{j=1, d_j=v}^{i-1} p_{ij}^{(k+1)}$$



Influence structure estimated from the MemeTracker dataset

# outline

1. introduction
2. One-dimensional Hawkes Process
3. Multi-dimensional Hawkes Process
4. Marked Hawkes Processes

# Conditional Intensity Function: Marked Point Processes

- The conditional intensity function for the marked case

$$\lambda^*(t,k) = \lambda^*(t) f^*(k|t)$$

- $\lambda^*(t)$ ground intensity, history dependent, also may depend on all the past marks as well
- $f^*(k|t) = f\left(k|t, \{(t_i,k_i)\}_{t_i<t}\right)$

# Marked Hawkes Processes

Marked Hawkes Process is given by the conditional density

$$\lambda^*(t,k) = f^*(k|t)(\mu(t) + \sum_{t_i<t} \alpha(k_i) g(t - t_i, k_i))$$

$$= f^*(k|t)(\mu(t) + \int_0^t \alpha(k)\beta(t - s, k) N(ds, dk))$$

# Marked Hawkes Processes

Defined in terms of conditional density functions

- **Immigration intensity**: $\mu(t)$ with parameter $\theta_\mu$
- **Total offspring intensity**: $\alpha(k)$ with parameter $\theta_\alpha$
- **Normalized offspring intensity**:$g(t,k)$ with parameter $\theta_g$
- **Mark density:** $f^*(k|t)$ with parameter $\theta_f$
- The product function $\alpha(k)g(t,k)$ is called offspring intensit (infectivity function, triggering kernel)

# Marked Hawkes Processes

- **Example:** Reproducing population with exponential survival time, and individuals reproduce uniformly throughout their survival time
- $\mu(t)=\mu_1$, $\alpha(k) = \alpha_1 k$, $g(t,k) = 1_{[0,k]}(t)/k$,
- $f^*(k|t) = f_1 \exp(-f_1 k)$
- Then the intensity function

$$\lambda^*(t,k) = f_1 \exp(-f_1 k) \left(\mu_1 + \sum_{t_i<t} \alpha_1 1_{[0,k_i]}(t)\right)$$

# Marked Hawkes Processes

- **Example:**
- A simple earthquake model $k = (m, x, y)$
- $\mu(t)=\mu_1$, $\alpha(k) = \alpha_1 \exp(\alpha_2 k)$,
- $g(t, k) = \beta_2/\beta_1 \cdot (1 + t/\beta_1)^{-\beta_2 - 1}$,
- $f^*(k|t) = f_1 \exp(-f_1 k)\, 1_{(x,y)\in W}/|W|$
- Then the intensity function

$$\lambda^*(t, k) = f_1 \exp(-f_1 k)\, 1_{(x,y)\in W}/|W| \cdot (\mu_1 + \sum_{t_i<t} \alpha_1 \exp(\alpha_2 k)\, \beta_2/\beta_1 \cdot (1 + (t - t_i)/\beta_1)^{-\beta_2-1})$$

# Thanks