Hawkes processes and application in neuroscience and genomic

Patricia Reynaud-Bouret

CNRS, Université de Nice Sophia Antipolis

Journées Aussois 2015, Modélisation Mathématique et Biodiversité

Hawkes processes and application in neuroscience and genomic

Patricia Reynaud-Bouret

CNRS, Université de Nice Sophia Antipolis

Journées Aussois 2015, Modélisation Mathématique et Biodiversité

Table of Contents

Table of Contents

Definition

Point process

N= random countable set of points of $\mathbb{X}$ (usually $\mathbb{R}$).

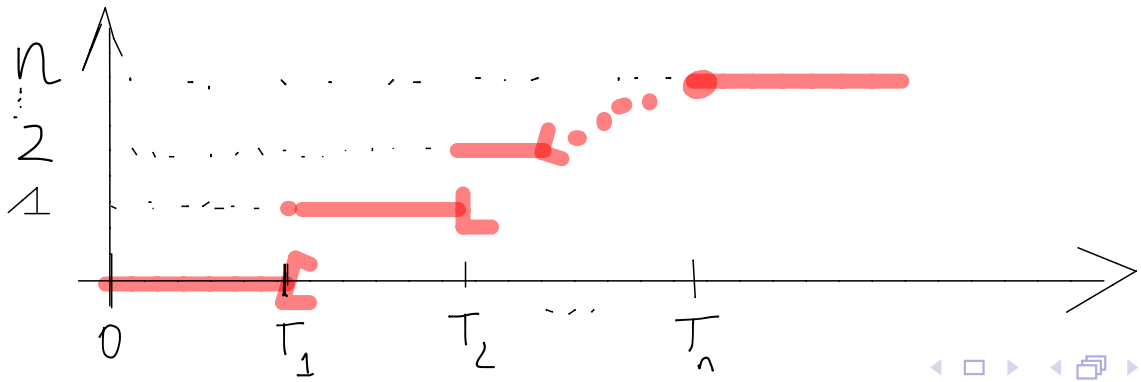
Definition

Point process

N= random countable set of points of $\mathbb{X}$ (usually $\mathbb{R}$).

N_A number of points of N in A, $dN_t = \sum_{T \text{ point de } N} \delta_T$.
$\int f(t) dN_t = \sum_{T \in N} f(T)$

If $\mathbb{X} = \mathbb{R}^+$ and no accumulation, the counting process is $N_t = N_{[0,t]}$.



Examples

- disintegration of radioactive atoms
- date / duration / position of phone calls
- date / position/ magnitude of earthquakes

Examples

- disintegration of radioactive atoms
- date / duration / position of phone calls
- date / position/ magnitude of earthquakes
- position / size / age of trees in a forest
- discovery time / position/ size of petroleum fields
- breakdowns

Examples

- disintegration of radioactive atoms
- date / duration / position of phone calls
- date / position/ magnitude of earthquakes
- position / size / age of trees in a forest
- discovery time / position/ size of petroleum fields
- breakdowns
- time of the event "a bee came out of its hive", "a monkey climbed in the tree", "an agent buys a financial asset", "someone clicks on the website" ....

Examples

- disintegration of radioactive atoms
- date / duration / position of phone calls
- date / position/ magnitude of earthquakes
- position / size / age of trees in a forest
- discovery time / position/ size of petroleum fields
- breakdowns
- time of the event "a bee came out of its hive", "a monkey climbed in the tree", "an agent buys a financial asset", "someone clicks on the website" ....

Everything has been noted, probably link via a model between those quantities. No reason at all that they may be identically distributed and/or independent (not i.i.d.).

Life times

Historically, first evident records of statistics start with Graunt in 1662 and its life table.
NB : at the same time, Pascal and Euler started Probability ... Halley, Huyghens and many others also recorded life tables

"Why are people dying ?" $\rightarrow$ record of the time of deaths in London at that time.

Life times

Historically, first evident records of statistics start with Graunt in 1662 and its life table.

"Why are people dying ?" $\rightarrow$ record of the time of deaths in London at that time.

- Even if considered i.i.d., the interesting quantity is the hazard rate $q(t)$, with

$$q(t)dt \simeq \mathbb{P}(\text{ die just after } t \text{ given that alive in } t)$$

If f density, $q(t) = f(t)/S(t)$, with

$$S(t) = \int_t^{+\infty} f(s)ds.$$

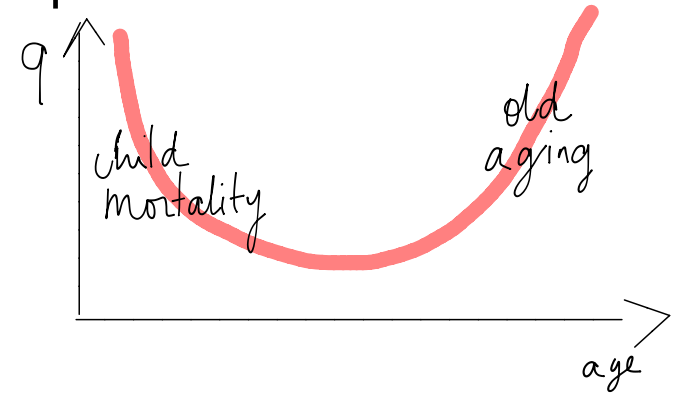
Life times

Historically, first evident records of statistics start with Graunt in 1662 and its life table.

"Why are people dying ?" $\rightarrow$ record of the time of deaths in London at that time.

- Even if considered i.i.d., the interesting quantity is the hazard rate $q(t)$,
- $q = cte$: do not "age", no memory $\rightarrow$ exponential variable
- q increases : better young than old
- q decreases : better old than young.

classical shape for human



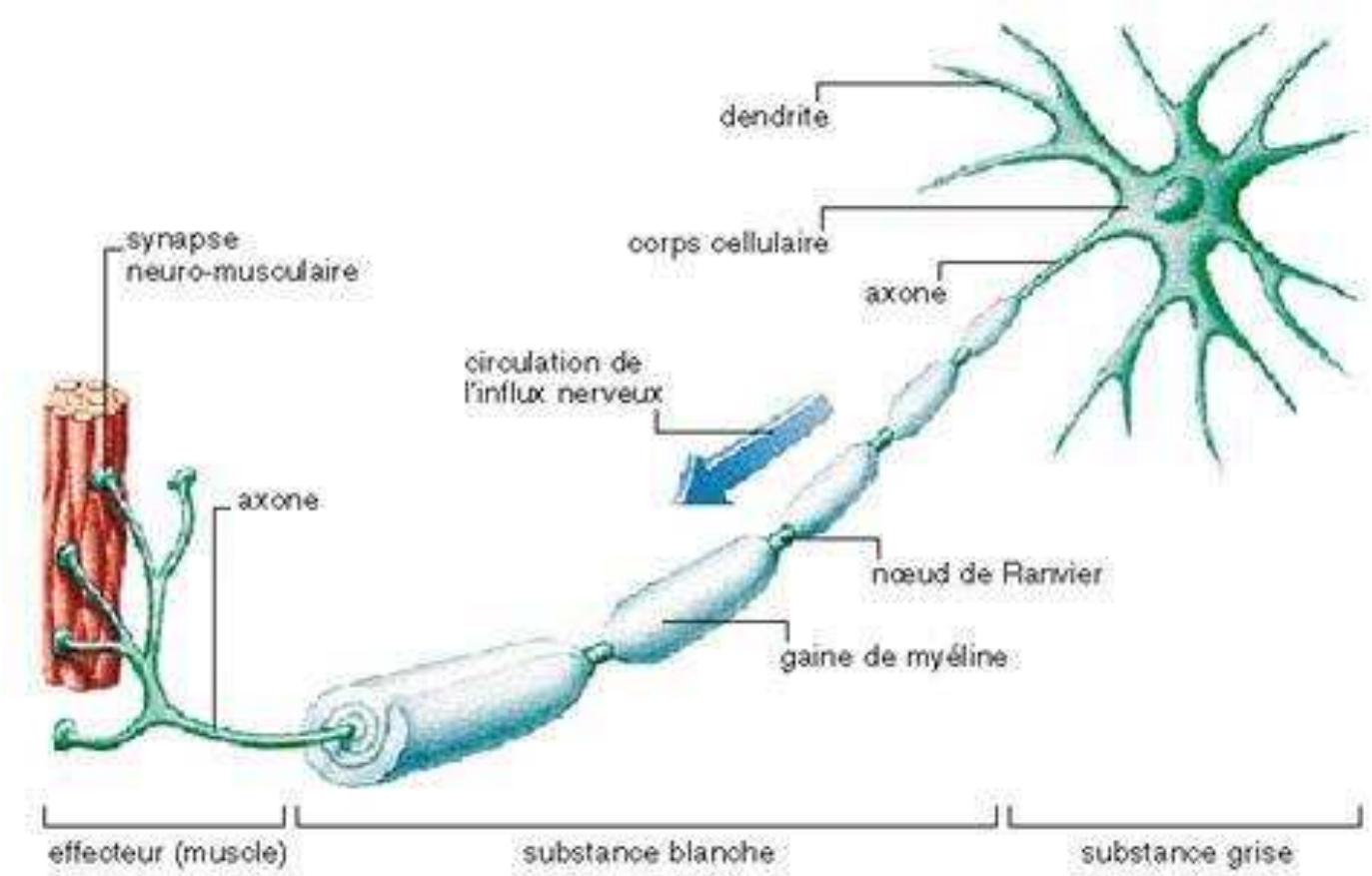
Life times

Historically, first evident records of statistics start with Graunt in 1662 and its life table.

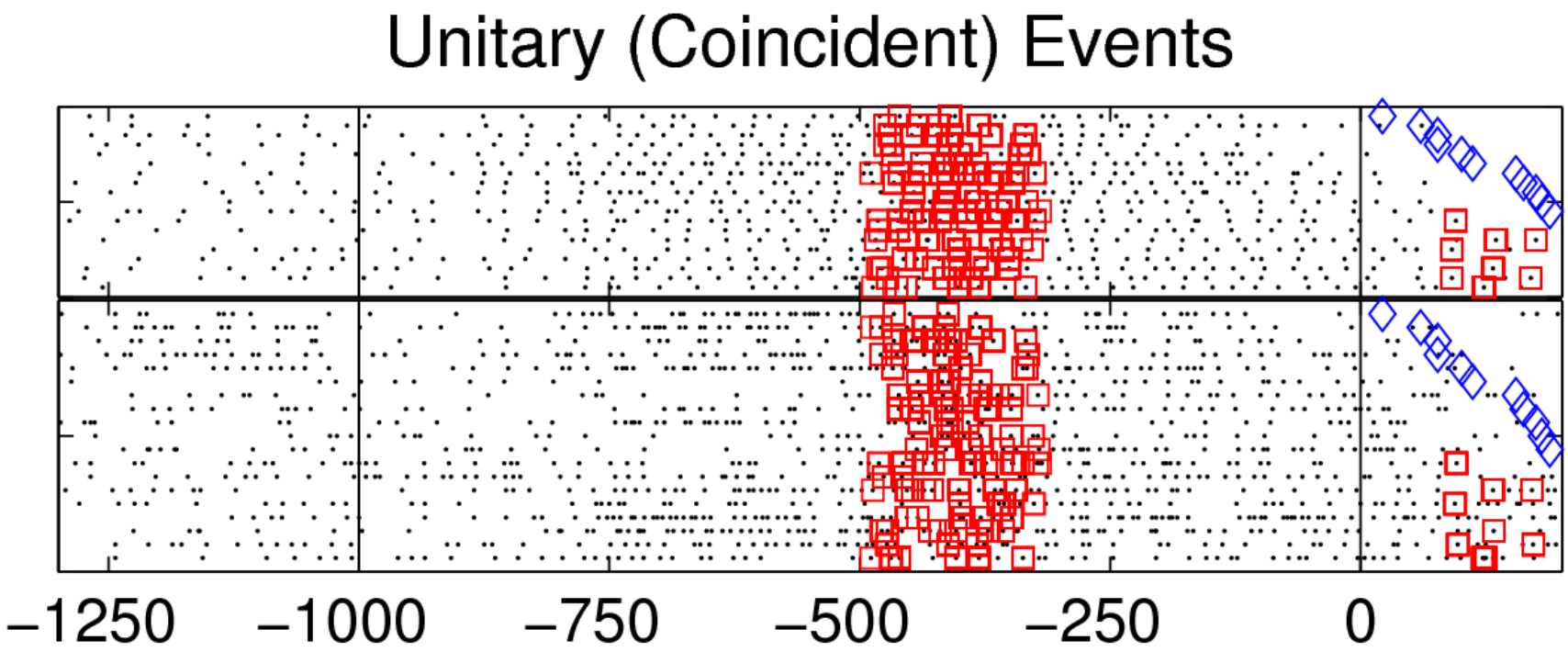
"Why are people dying ?" $\rightarrow$ record of the time of deaths in London at that time.

- Even if considered i.i.d., the interesting quantity is the hazard rate $q(t)$,
- not even clearly i.i.d. : people may move out before disease, may contaminate each other etc... $\rightarrow$ what is the interesting quantity ?

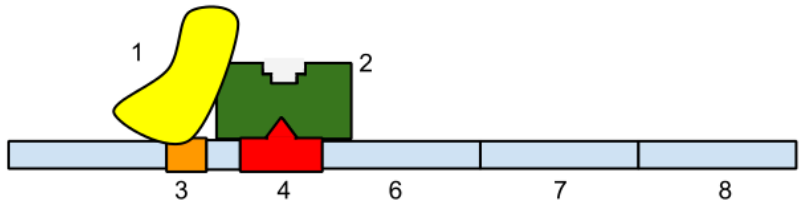
Neuroscience and neuronal unitary activity

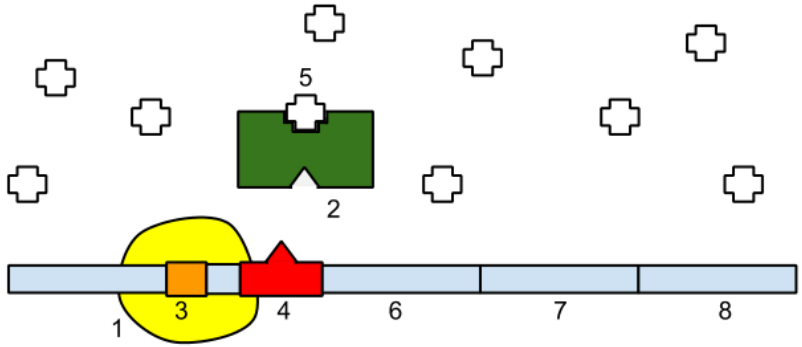


Neuronal data and Unitary Events



Genomics and Transcription Regulatory Elements



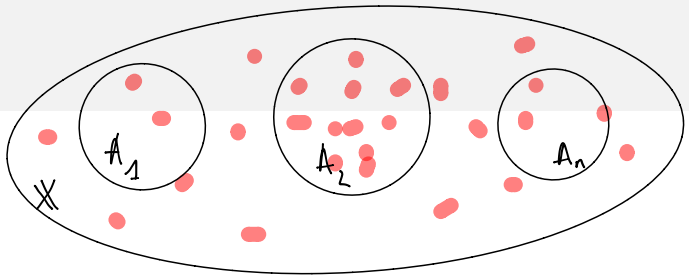


Definition

Poisson processes

- for all integer n, for all $A_1, \ldots, A_n$ disjoint measurable subsets of $\mathbb{X}$, $N_{A_1}, \ldots, N_{A_n}$ are independent random variables.

Definition



Poisson processes

- for all integer n, for all $A_1, \ldots, A_n$ disjoint measurable subsets of $\mathbb{X}$, $N_{A_1}, \ldots, N_{A_n}$ are independent random variables.
- for all measurable subset A of $\mathbb{X}$, N_A obeys a Poisson law with parameter depending on A and denoted $\ell(A)$.

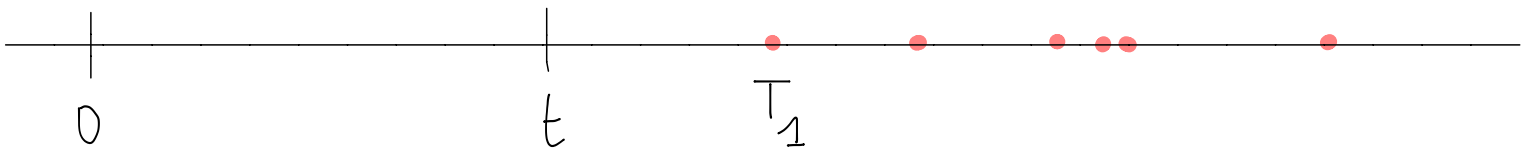
Definition

Poisson processes

- for all integer n, for all $A_1, \ldots, A_n$ disjoint measurable subsets of $\mathbb{X}$, $N_{A_1}, \ldots, N_{A_n}$ are independent random variables.
- for all measurable subset A of $\mathbb{X}$, N_A obeys a Poisson law with parameter depending on A and denoted $\ell(A)$.

If $\mathbb{X} = \mathbb{R}$ and $\ell([0, t]) = \lambda t$, then

$$\mathbb{P}(N_t = 0) = \mathbb{P}(T_1 > t)$$



Definition

Poisson processes

- for all integer n, for all $A_1, \ldots, A_n$ disjoint measurable subsets of $\mathbb{X}$, $N_{A_1}, \ldots, N_{A_n}$ are independent random variables.
- for all measurable subset A of $\mathbb{X}$, N_A obeys a Poisson law with parameter depending on A and denoted $\ell(A)$.

If $\mathbb{X} = \mathbb{R}$ and $\ell([0,t]) = \lambda t$, then

$$\mathbb{P}(N_t = 0) = \mathbb{P}(T_1 > t) = e^{-\lambda t}$$

↑ Poisson distribution with parameter $\lambda t = \theta$

$$P(N_k = k) = \frac{\theta^k}{k!} e^{-\theta}.$$

Definition

Poisson processes

- for all integer n, for all $A_1, \ldots, A_n$ disjoint measurable subsets of $\mathbb{X}$, $N_{A_1}, \ldots, N_{A_n}$ are independent random variables.
- for all measurable subset A of $\mathbb{X}$, N_A obeys a Poisson law with parameter depending on A and denoted $\ell(A)$.

If $\mathbb{X} = \mathbb{R}$ and $\ell([0, t]) = \lambda t$, then

$$\mathbb{P}(N_t = 0) = \mathbb{P}(T_1 > t) = e^{-\lambda t}$$

$\rightarrow$ first time is an exponential variable

Definition

Poisson processes

- for all integer n, for all $A_1, \ldots, A_n$ disjoint measurable subsets of $\mathbb{X}$, $N_{A_1}, \ldots, N_{A_n}$ are independent random variables.
- for all measurable subset A of $\mathbb{X}$, N_A obeys a Poisson law with parameter depending on A and denoted $\ell(A)$.

If $\mathbb{X} = \mathbb{R}$ and $\ell([0, t]) = \lambda t$, then

$$\mathbb{P}(N_t = 0) = \mathbb{P}(T_1 > t) = e^{-\lambda t}$$

$\rightarrow$ first time is an exponential variable
All "intervals" are i.i.d and exponentially distributed.

Definition

Poisson processes

- for all integer n, for all $A_1, \ldots, A_n$ disjoint measurable subsets of $\mathbb{X}$, $N_{A_1}, \ldots, N_{A_n}$ are independent random variables.
- for all measurable subset A of $\mathbb{X}$, N_A obeys a Poisson law with parameter depending on A and denoted $\ell(A)$.

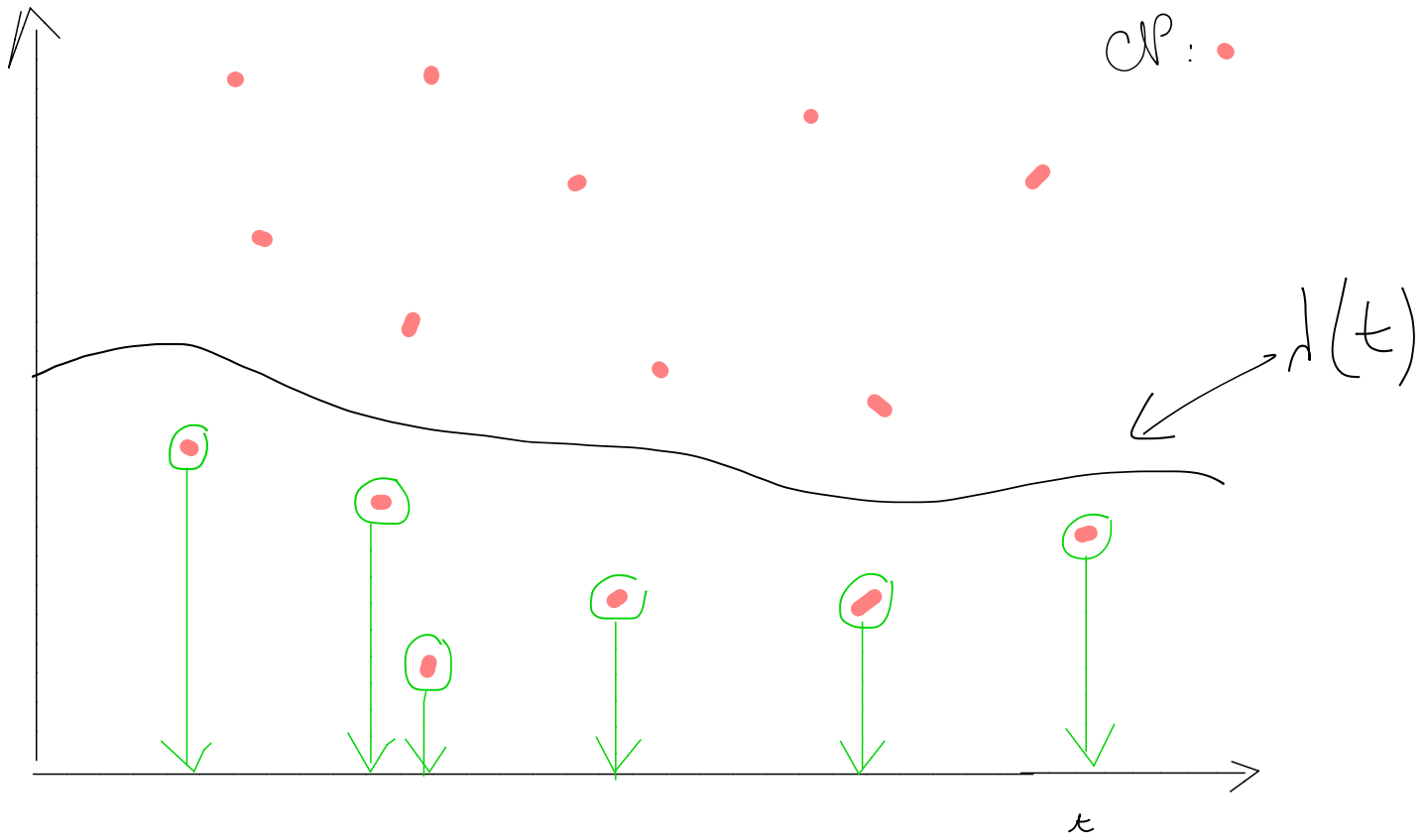
If $\mathbb{X} = \mathbb{R}$ and $\ell([0,t]) = \lambda t$, then

$$\mathbb{P}(N_t = 0) = \mathbb{P}(T_1 > t) = e^{-\lambda t}$$

$\rightarrow$ first time is an exponential variable
All "intervals" are i.i.d and exponentially distributed.
No memory $\Longleftrightarrow$ independence to the past

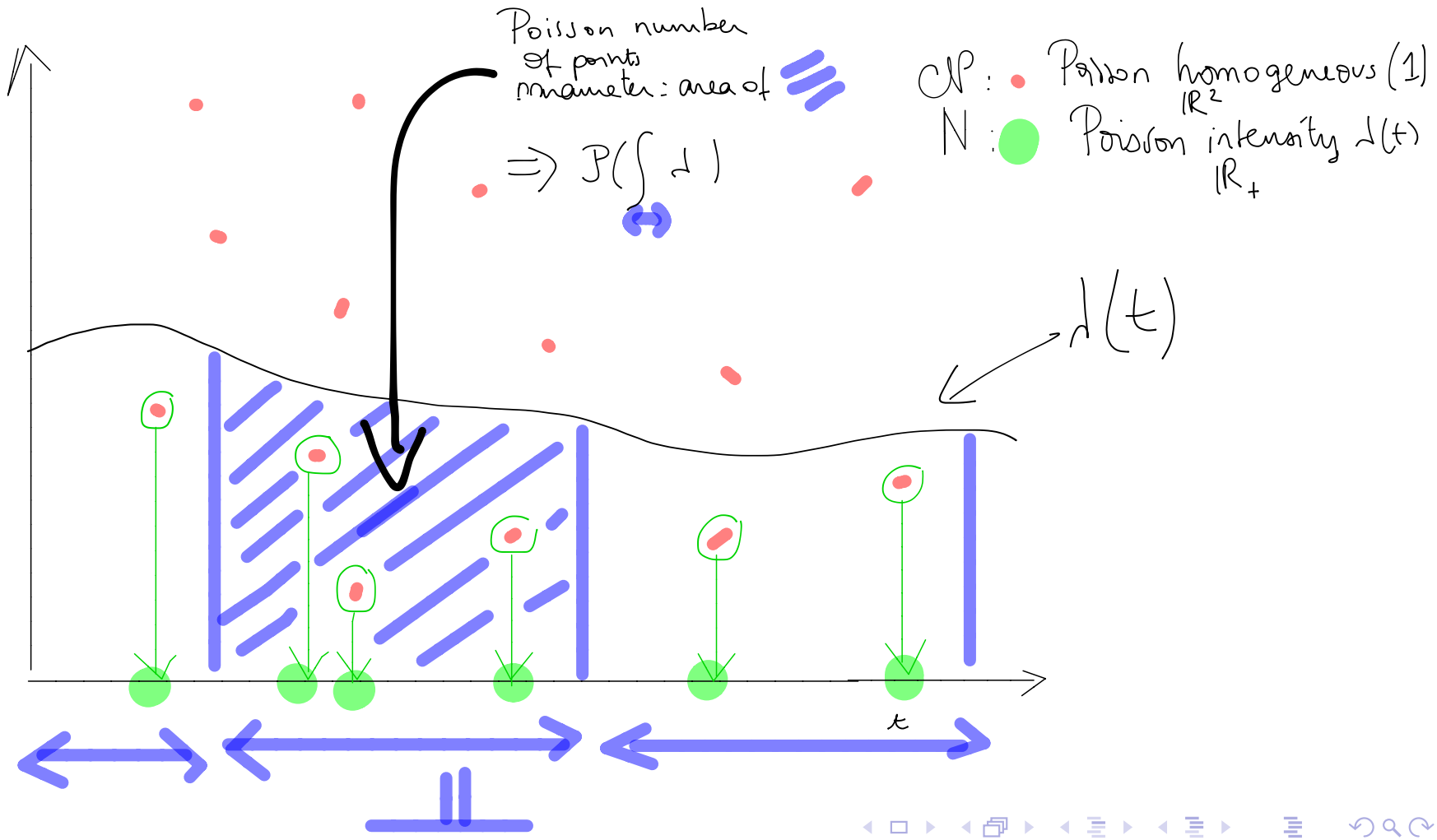
If λ not constant

If $\ell([0,t]) = \int_0^t \lambda(s)ds$, let $\mathcal{N}$ be a Poisson process in $\mathbb{R}_+^2$ and



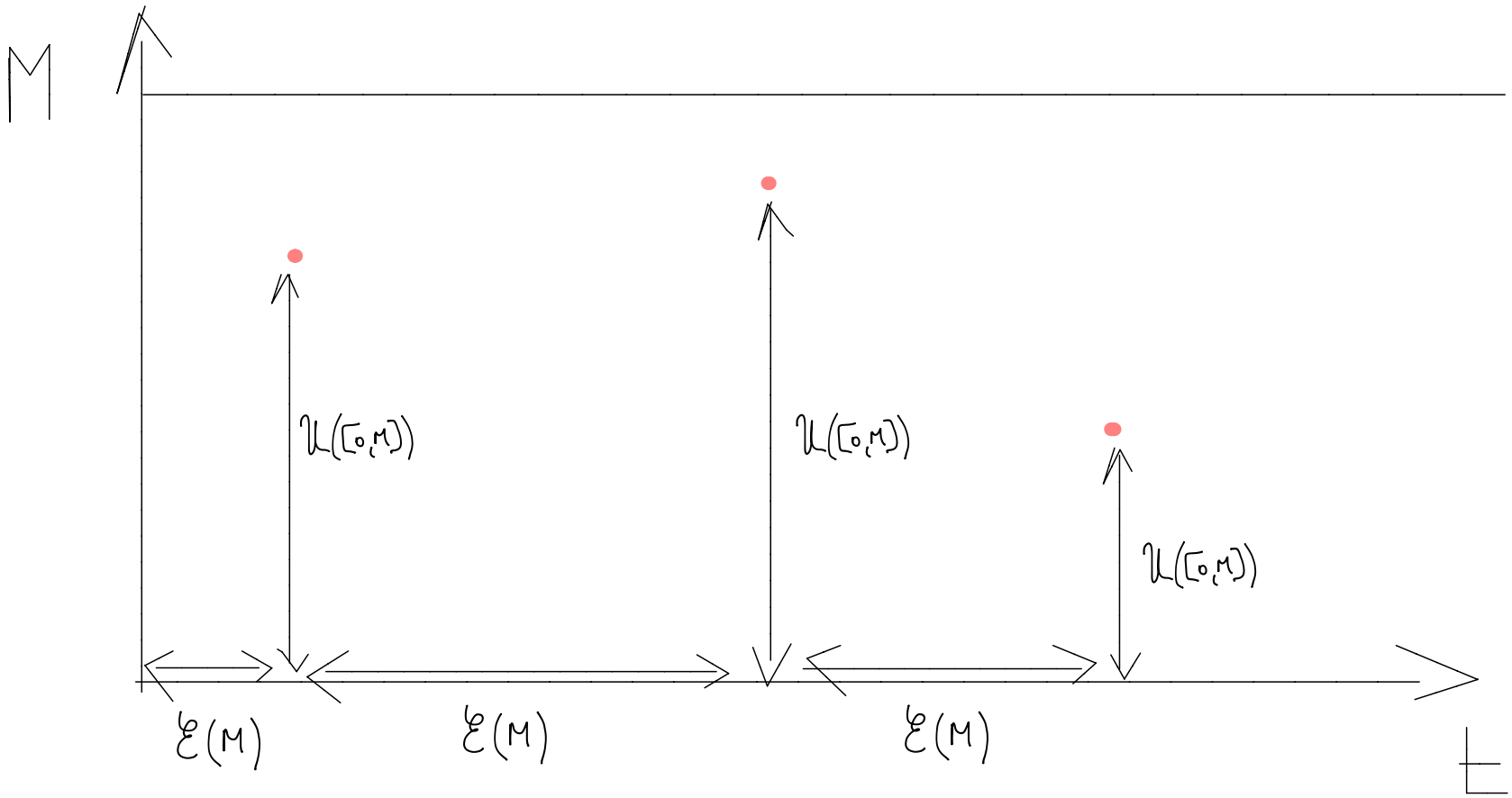
If λ not constant

If $\ell([0,t]) = \int_0^t \lambda(s)ds$, let $\mathcal{N}$ be a Poisson process in $\mathbb{R}_+^2$ and



If λ not constant

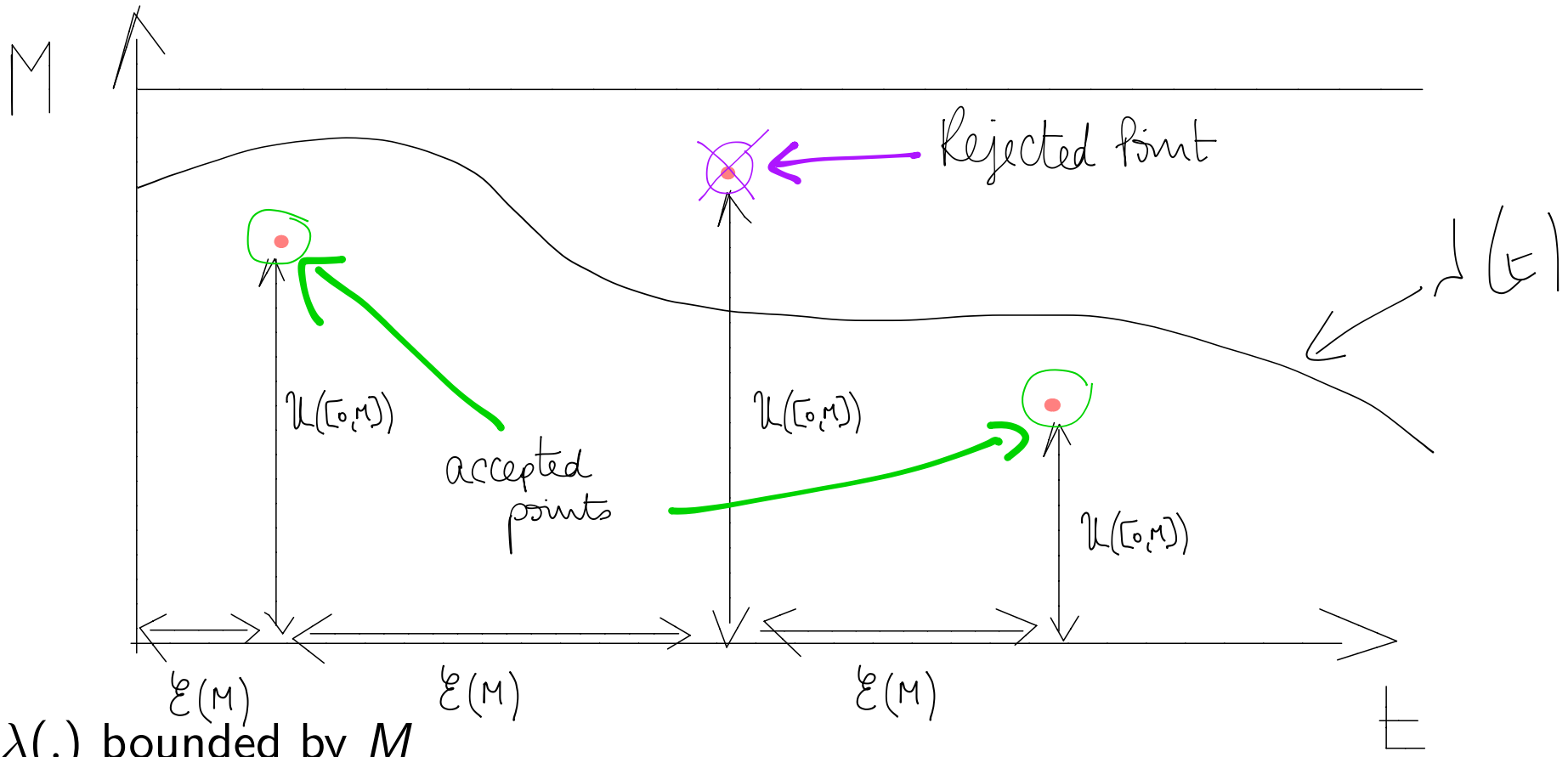
If $\ell([0,t]) = \int_0^t \lambda(s)ds$, let $\mathcal{N}$ be a Poisson process in $\mathbb{R}_+^2$ and



If $\lambda(.)$ bounded by M

If λ not constant

If $\ell([0,t]) = \int_0^t \lambda(s)ds$, let $\mathcal{N}$ be a Poisson process in $\mathbb{R}_+^2$ and



If $\lambda(.)$ bounded by M

Conditional intensity

Need of the same sort of interpretation as hazard rate but given "the Past" !

Conditional intensity

Need of the same sort of interpretation as hazard rate but given "the Past" !

$$\underbrace{dN_t}_{\text{Nbr observed points in } [t, t+dt]} = \overbrace{\underbrace{\lambda(t)\, dt}_{\text{Expected Number given the past before } t}}^{\textbf{intensity}} + \underbrace{\text{noise}}_{\textit{Martingales differences}}$$

Conditional intensity

Need of the same sort of interpretation as hazard rate but given "the Past" !

$$\underbrace{dN_t}_{\text{Nbr observed points in } [t, t+dt]} = \overbrace{\underbrace{\lambda(t)\, dt}_{\text{Expected Number given the past before } t}}^{\textbf{intensity}} + \underbrace{\text{noise}}_{\textit{Martingales differences}}$$

$\lambda(t)$ = instantaneous frequency

Conditional intensity

Need of the same sort of interpretation as hazard rate but given "the Past" !

$$\underbrace{dN_t}_{\text{Nbr observed points in } [t, t+dt]} = \overbrace{\underbrace{\lambda(t)\, dt}_{\text{Expected Number given the past before } t}}^{\textbf{intensity}} + \underbrace{\text{noise}}_{\textit{Martingales differences}}$$

$$\begin{aligned} \lambda(t) &= \text{instantaneous frequency} \\ &= \textbf{random}, \text{ depends on the Past (previous points ...)} \end{aligned}$$

Conditional intensity

Need of the same sort of interpretation as hazard rate but given "the Past" !

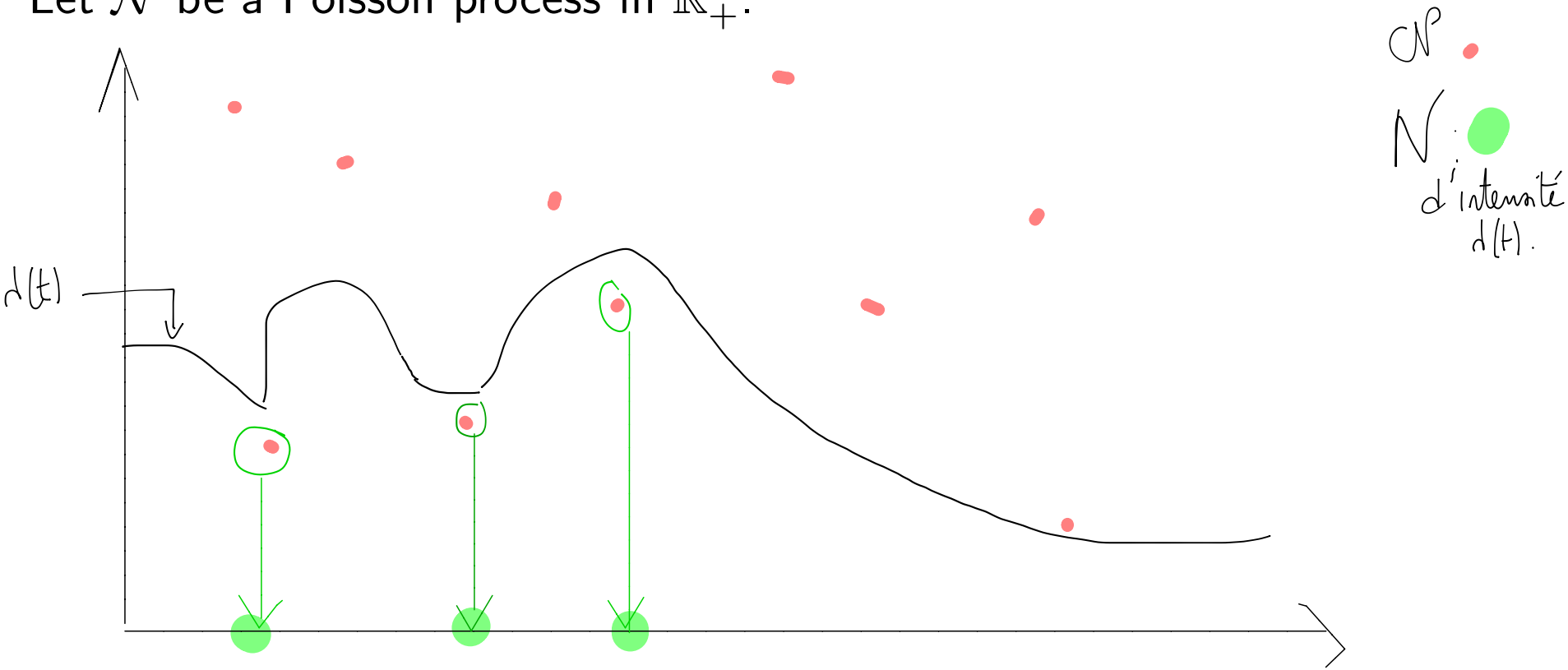
$$\underbrace{dN_t}_{\text{Nbr observed points in } [t, t+dt]} = \overbrace{\underbrace{\lambda(t)\, dt}_{\text{Expected Number given the past before } t}}^{\textbf{intensity}} + \underbrace{\text{noise}}_{\textit{Martingales differences}}$$

$$\begin{aligned} \lambda(t) &= \text{instantaneous frequency} \\ &= \textbf{random}, \text{ depends on the Past (previous points ...)} \\ &= \text{characterizes the distribution when exists} \end{aligned}$$

NB : if $\lambda(t)$ deterministic $\rightarrow$ Poisson process

Thinning

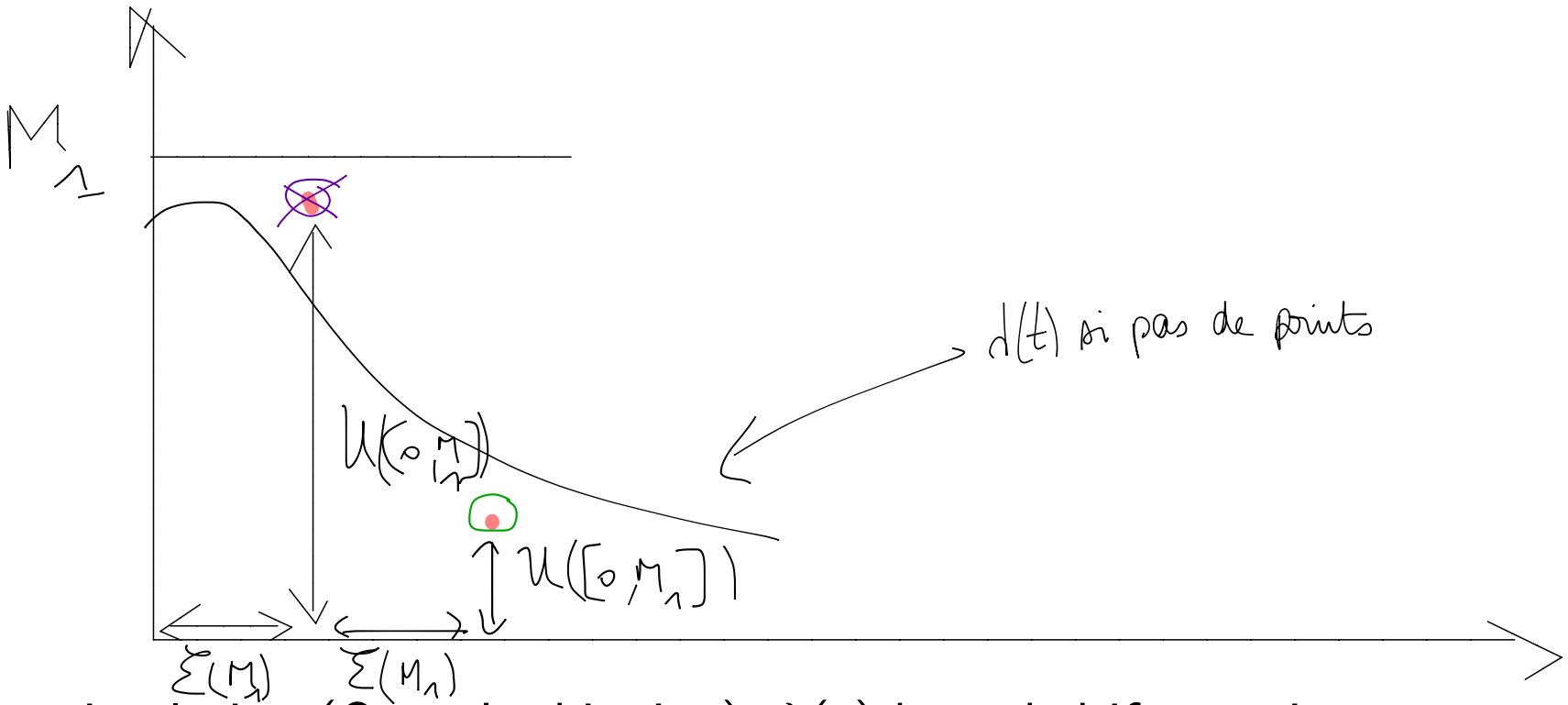
Let $\mathcal{N}$ be a Poisson process in $\mathbb{R}_+^2$.



$\lambda(t)$ is predictable, i.e. depends on the past only, can be drawn from left to right if one discovers one point at a time. Usually "c-à-g" for "continu à gauche" (left continuous)

Thinning

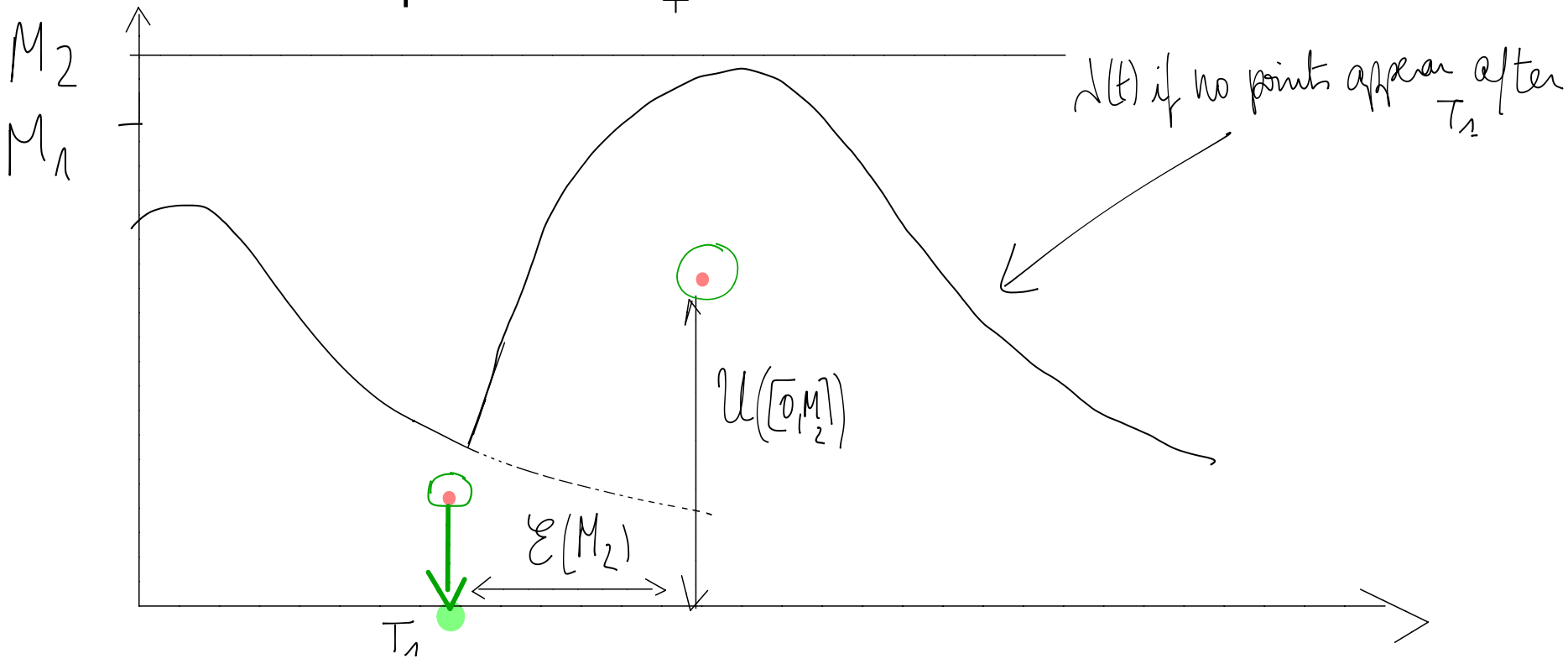
Let $\mathcal{N}$ be a Poisson process in $\mathbb{R}_+^2$.



For simulation (Ogata's thinning), $\lambda(t)$ bounded if no points appear

Thinning

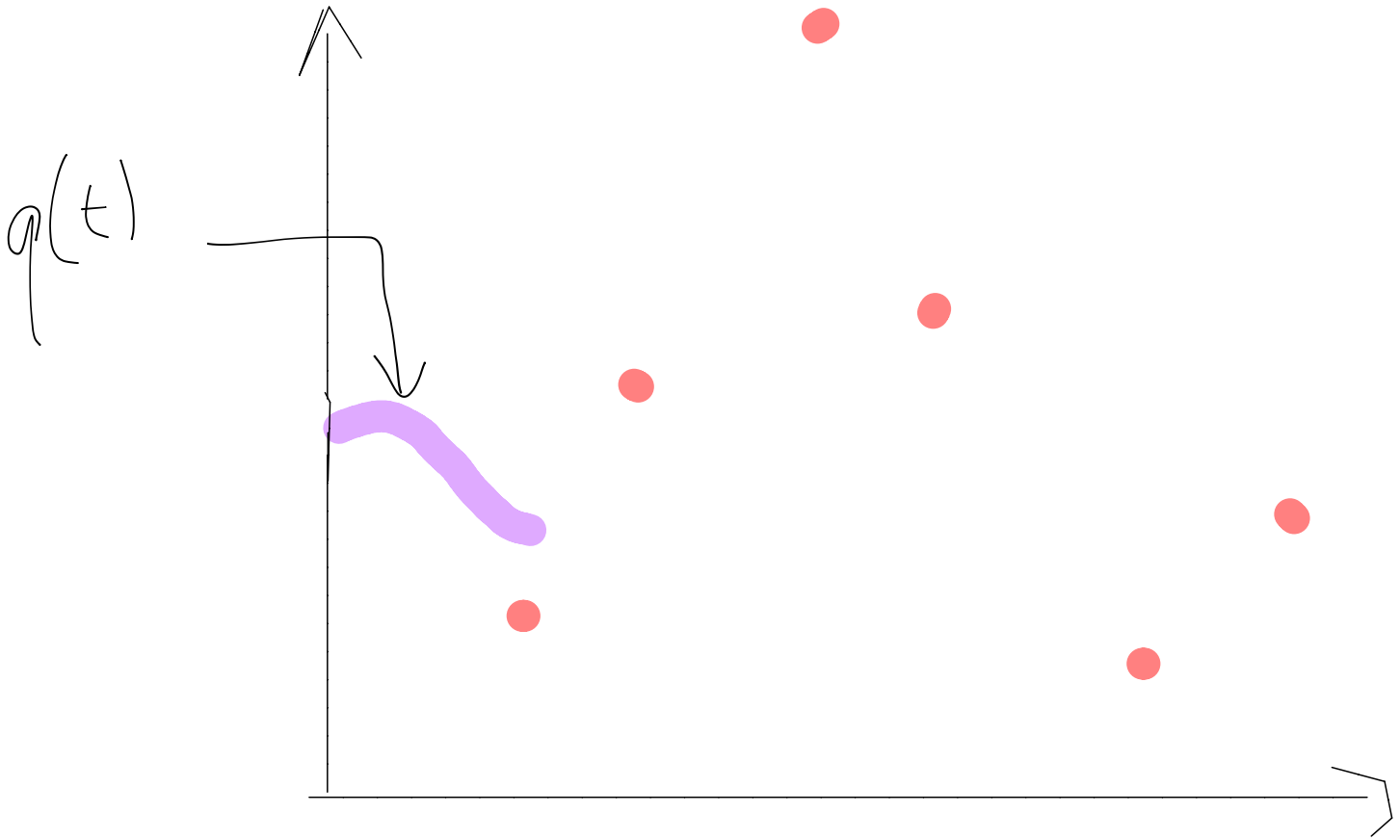
Let $\mathcal{N}$ be a Poisson process in $\mathbb{R}_+^2$.



For simulation (Ogata's thinning), $\lambda(t)$ bounded if no points appear

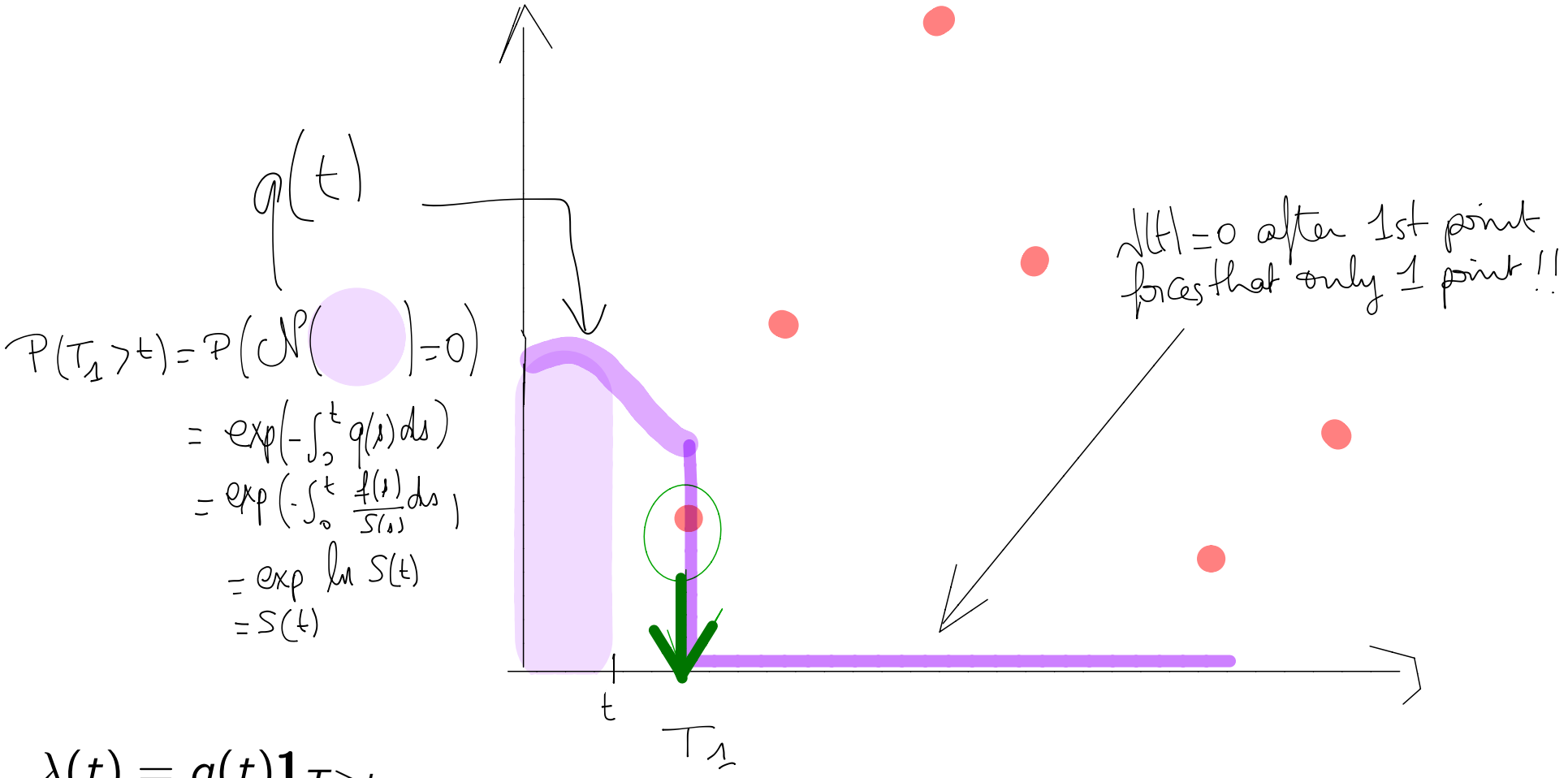
One life time

$N = \{T\}$, T of hazard rate q.



One life time

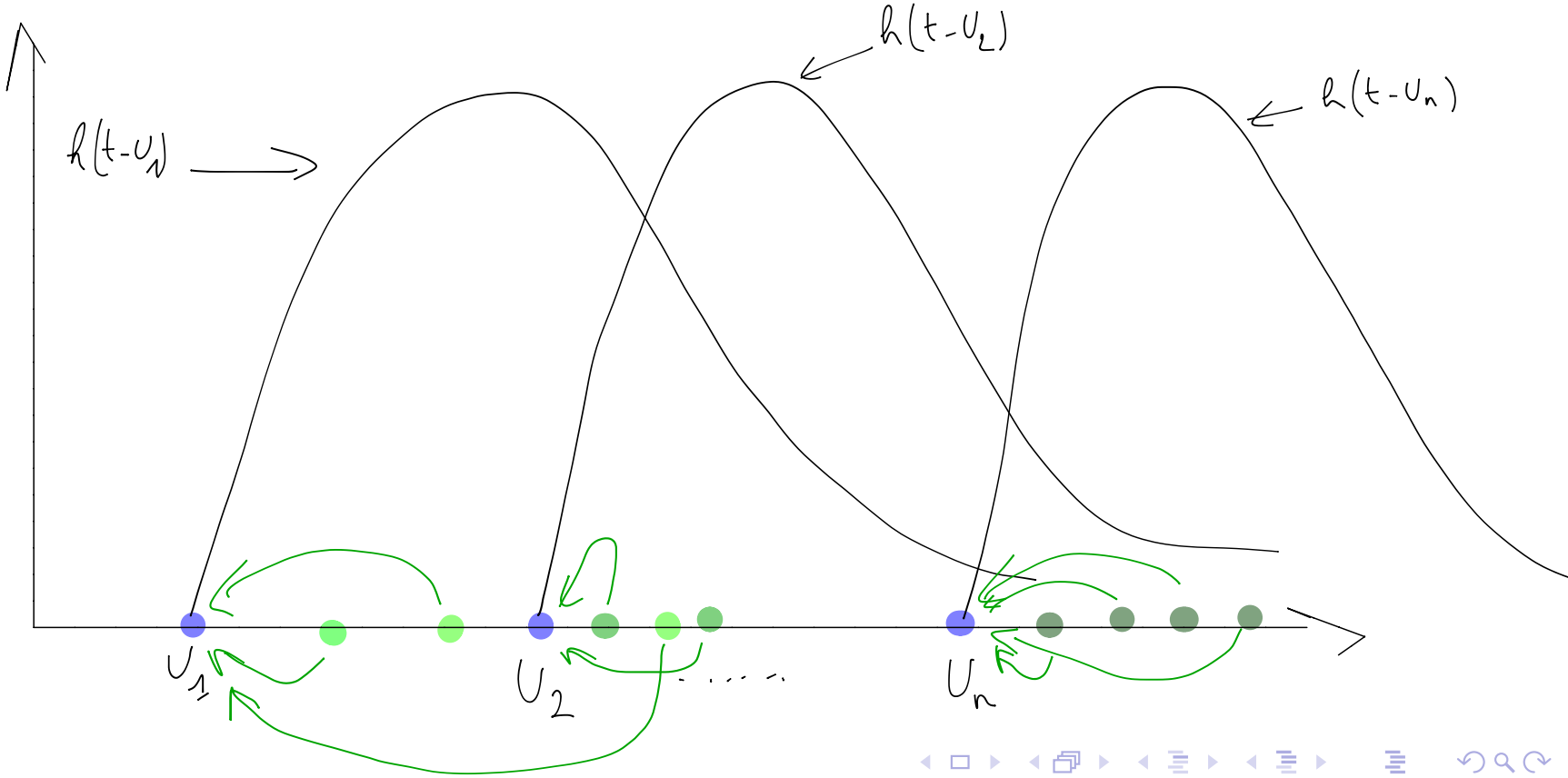
$N = \{T\}$, T of hazard rate q.



$$\lambda(t) = q(t)\mathbf{1}_{T \geq t}$$

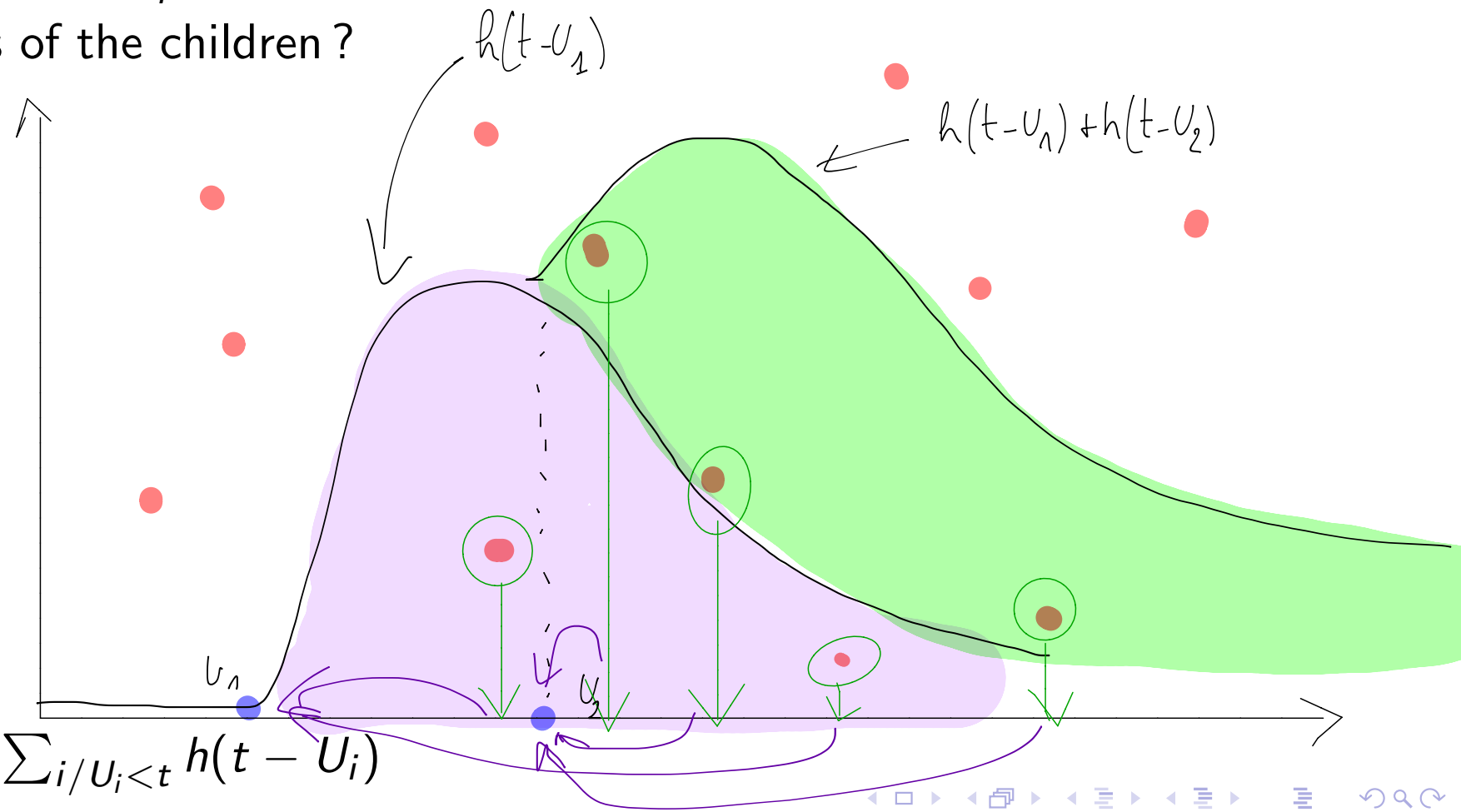
Poissonian interaction

Parents : U_i either (uniform) i.i.d. or (homogeneous) Poisson process
Each parent gives birth according to a Poisson process of intensity h originated in U_i.
Process of the children ?



Poissonian interaction

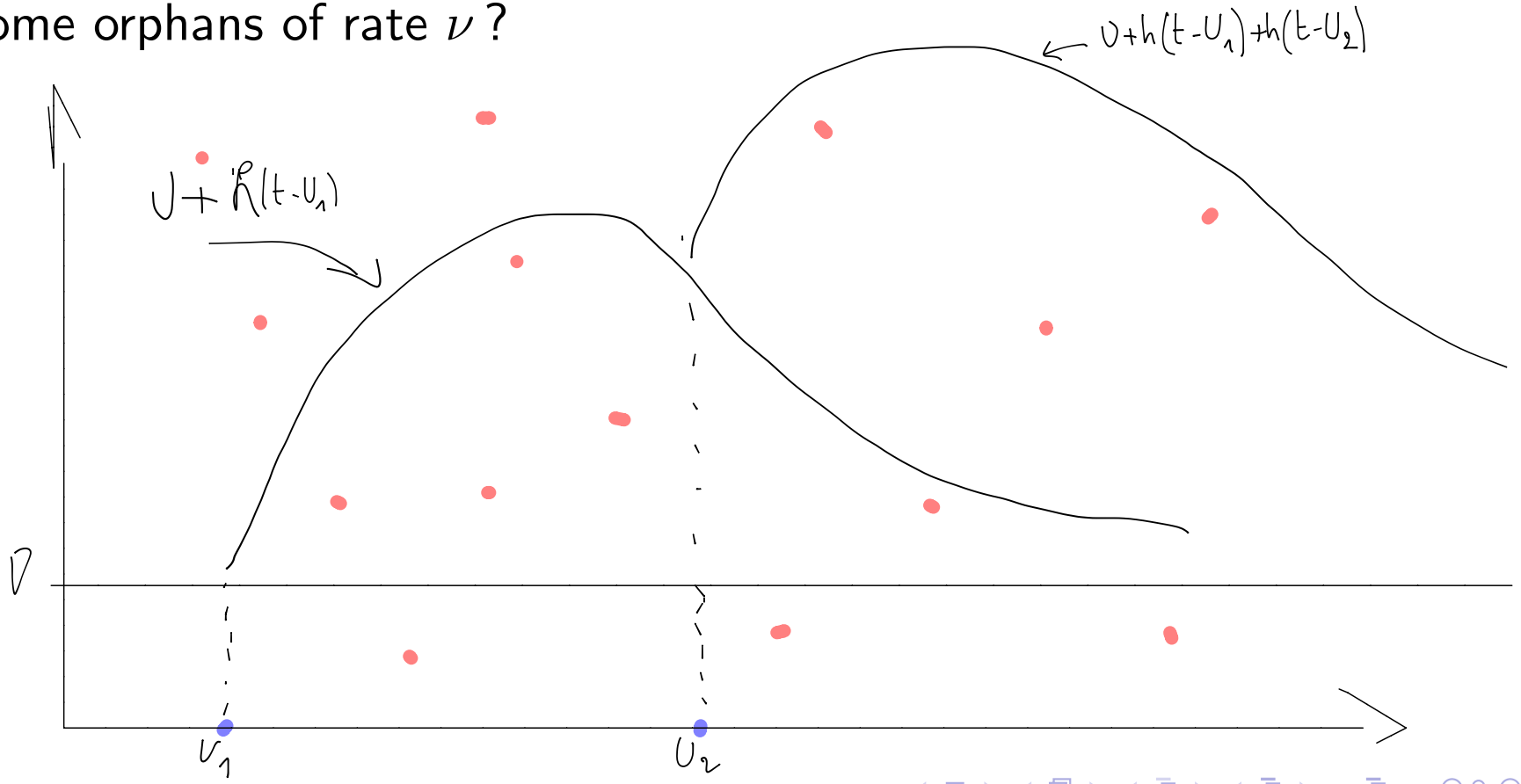
Parents : U_i either (uniform) i.i.d. or (homogeneous) Poisson process
Each parent gives birth according to a Poisson process of intensity h originated in U_i.
Process of the children ?



$$\lambda(t) = \sum_{i/U_i<t} h(t - U_i)$$

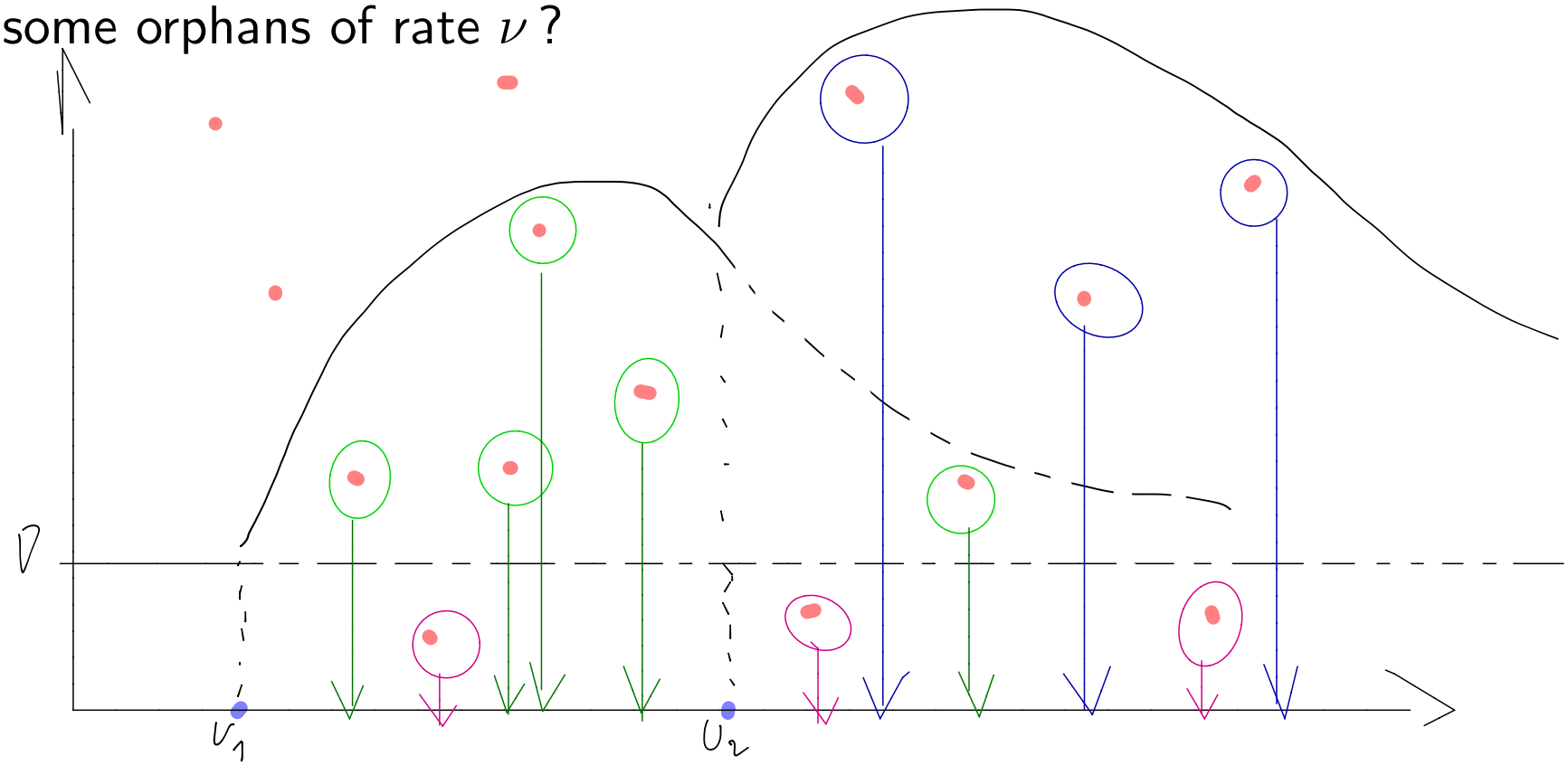
Poissonian interaction

Parents : U_i either (uniform) i.i.d. or (homogeneous) Poisson process
Each parent gives birth according to a Poisson process of intensity h originated in U_i.
With some orphans of rate ν ?



Poissonian interaction

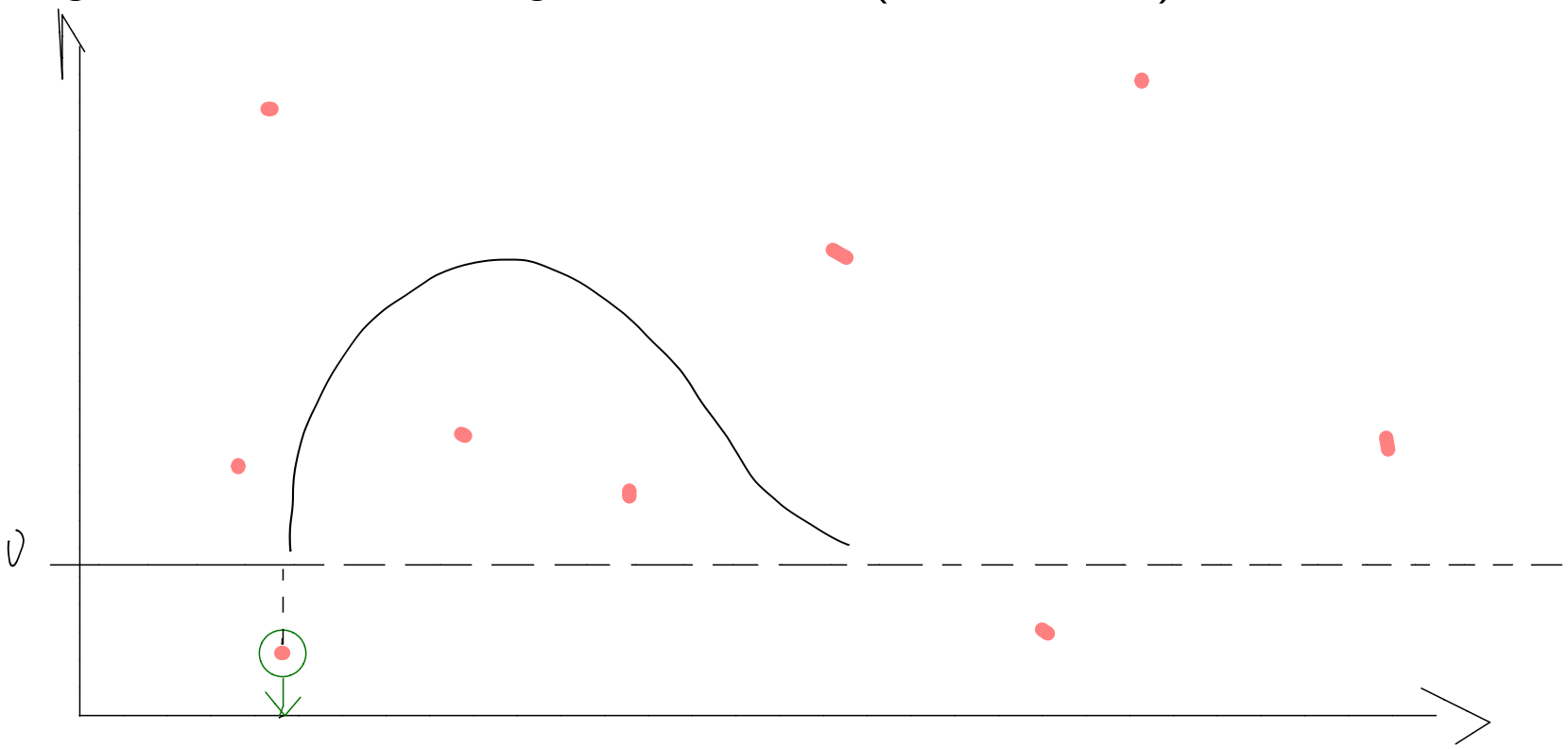
Parents : U_i either (uniform) i.i.d. or (homogeneous) Poisson process
Each parent gives birth according to a Poisson process of intensity h originated in U_i.
With some orphans of rate ν ?



$\lambda(t) = \nu + \sum_{i/U_i<t} h(t - U_i)$

Hawkes process : linear case

A self-exciting process introduced by Hawkes in the 70's to model earthquakes.
Ancestors appear at rate ν.
Each point gives birth according to h etc etc (aftershocks)

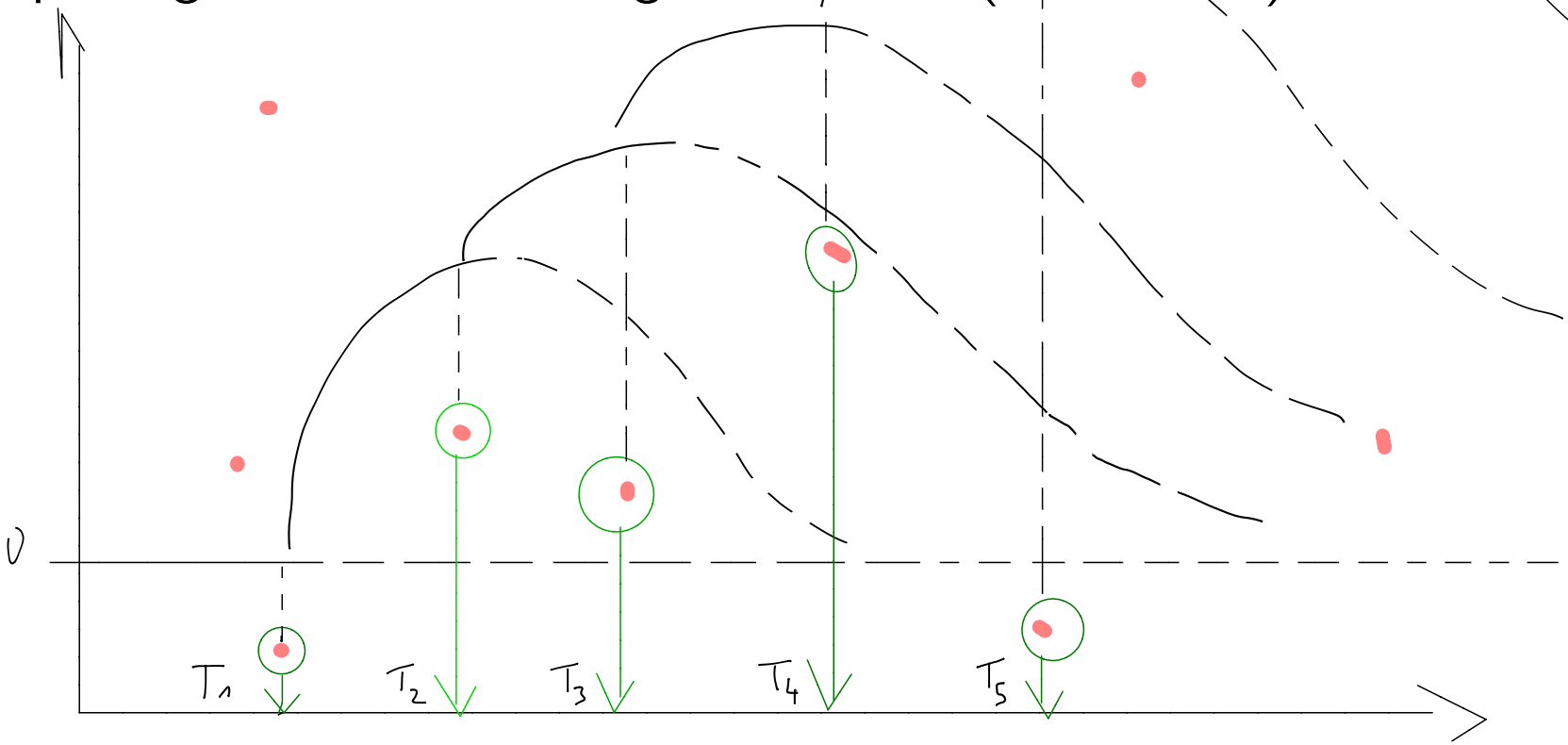


Hawkes process : linear case

A self-exciting process introduced by Hawkes in the 70's to model earthquakes.

Ancestors appear at rate ν.

Each point gives birth according to h etc etc (aftershocks)



$$\lambda(t) = \nu + \sum_{T<t} h(t-T) = \nu + \int_{-\infty}^{t} h(t-u)dN_u$$

Hawkes and Modelisation

- models any sort of self contamination (earthquakes, clicks, etc)

Hawkes and Modelisation

- models any sort of self contamination (earthquakes, clicks, etc)
- can be marked : position or magnitude of earthquakes, neuron on which the spike happens (see later)

Hawkes and Modelisation

- models any sort of self contamination (earthquakes, clicks, etc)
- can be marked : position or magnitude of earthquakes, neuron on which the spike happens (see later)
- can therefore model interaction also (see later)

Hawkes and Modelisation

- models any sort of self contamination (earthquakes, clicks, etc)
- can be marked : position or magnitude of earthquakes, neuron on which the spike happens (see later)
- can therefore model interaction also (see later)
- generally, modelling people "want" stationnarity : if $\int h < 1$, the branching procedure ends $\Longleftrightarrow$ existence of stationnary version.

Hawkes and Modelisation

- models any sort of self contamination (earthquakes, clicks, etc)
- can be marked : position or magnitude of earthquakes, neuron on which the spike happens (see later)
- can therefore model interaction also (see later)
- generally, modelling people "want" stationnarity : if $\int h < 1$, the branching procedure ends $\Longleftrightarrow$ existence of stationnary version.
- for genomics and neuroscience, need of inhibition : possible to use

$$\lambda(t) = \Phi\Big(\int_{-\infty}^{t} h(t-u)dN_u\Big),$$

with Φ 1-Lipschitz positive and h of any sign. Stationnarity condition $\int |h| < 1$.

- In particular, $\Phi = (.)_+$ but loss of the branching structure (?)

Likelihood

- Informally the likelihood of the observation N should give the probability to see (something near) N.

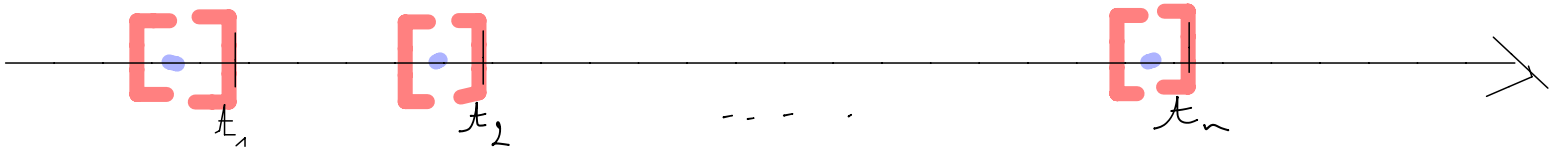
Likelihood

- Informally the likelihood of the observation N should give the probability to see (something near) N. Formally, it is somehow the density wrt (here) Poisson, but without depending on the intensity of the reference Poisson.

Likelihood

- Informally the likelihood of the observation N should give the probability to see (something near) N.
- If $N = \{t_1, ..., t_n\}$, we want for N' of intensity $\lambda(.)$ observed on $[0, T]$,

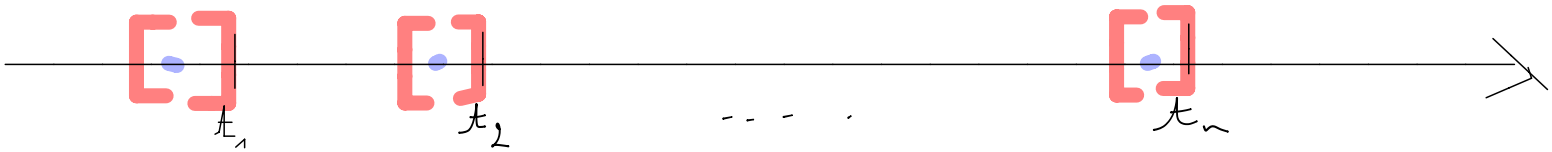
$$\mathbb{P}(N' \text{ has } n \text{ points, each } T_i \text{ in } [t_i - dt_i, t_i])$$



Likelihood

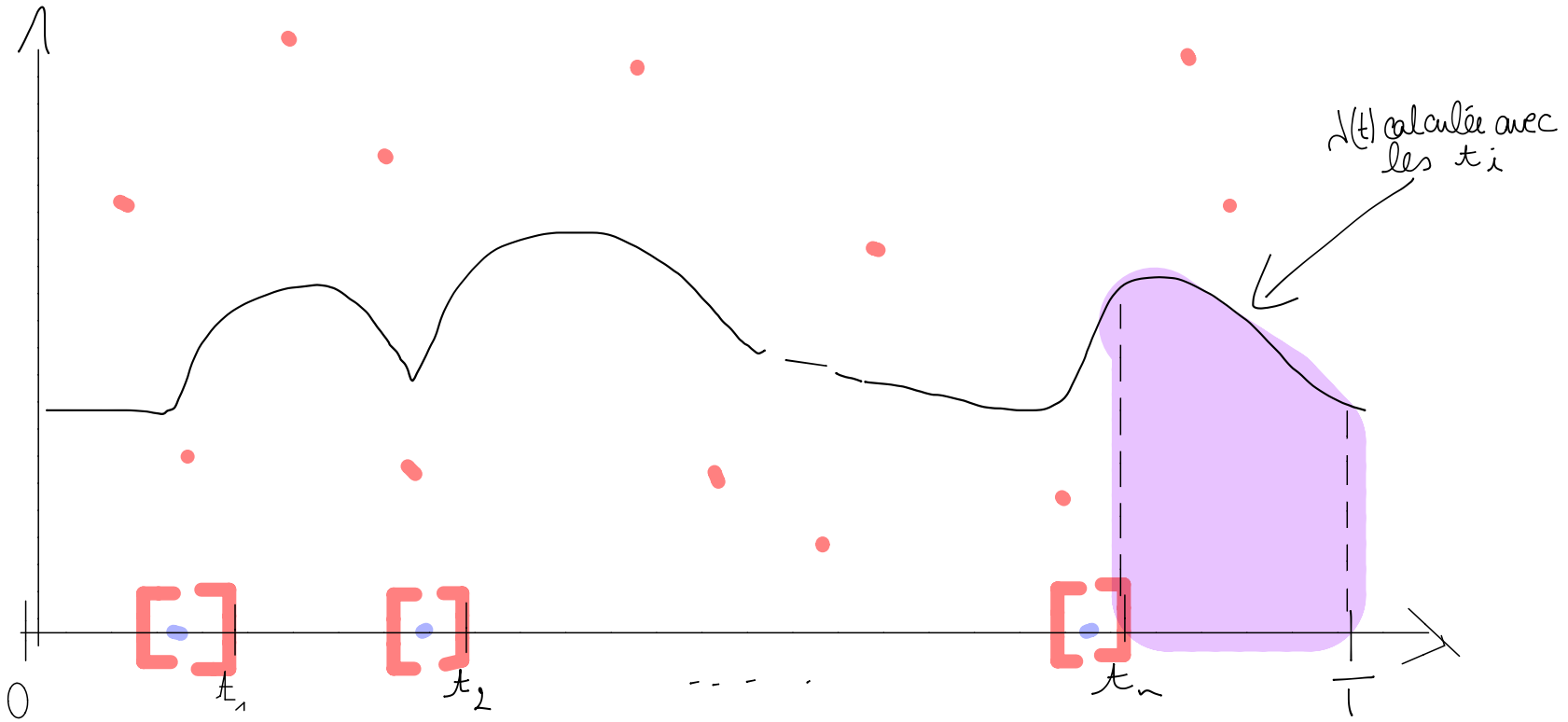
- Informally the likelihood of the observation N should give the probability to see (something near) N.
- If $N = \{t_1, ..., t_n\}$, we want for N' of intensity $\lambda(.)$ observed on $[0, T]$,

$$\mathbb{P}(N' \text{ has } n \text{ points, each } T_i \text{ in } [t_i - dt_i, t_i])$$



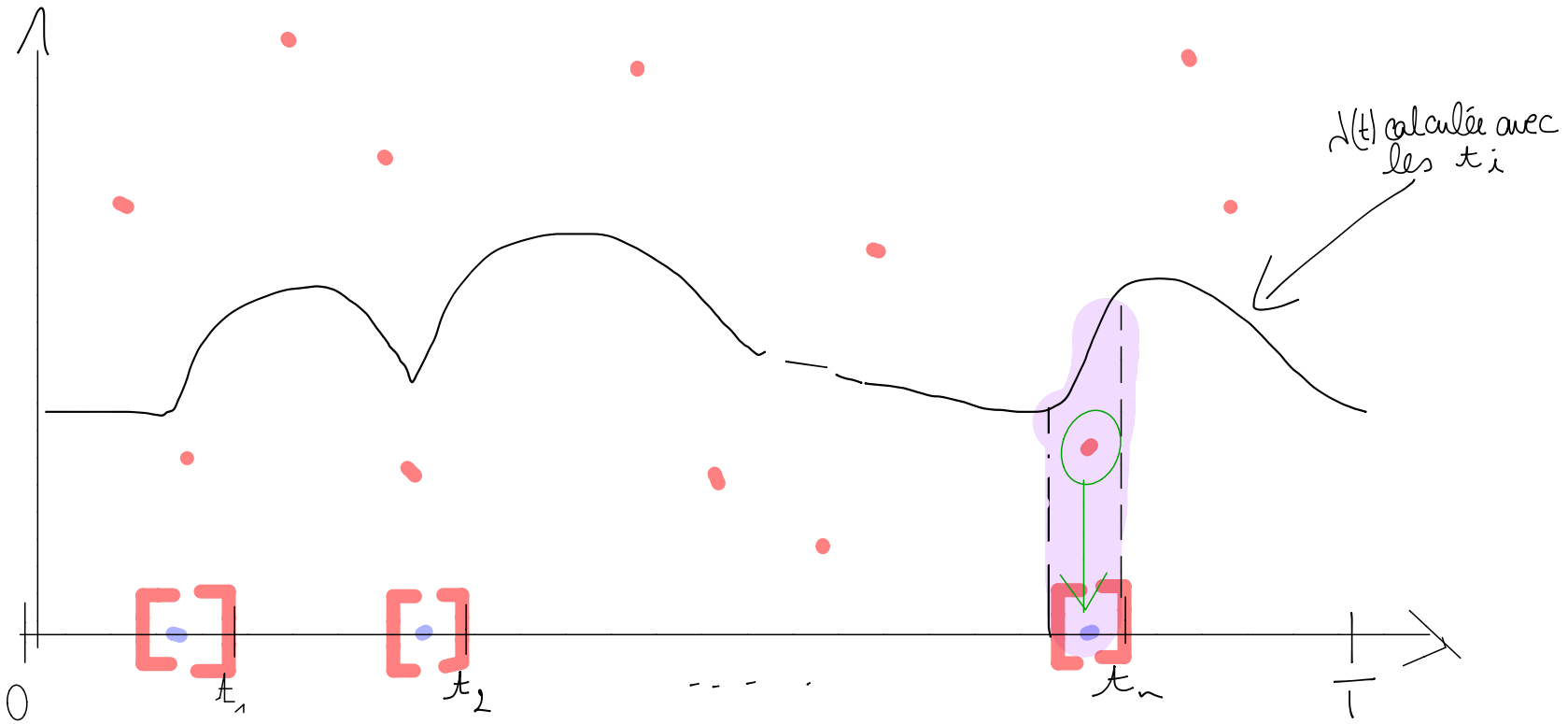
$$\mathbb{P}(\text{no point in } [0, t_1 - dt_1], 1 \text{ point in } [t_1 - dt_1, t_1], ..., \text{no point in} [t_n, T])$$

Likelihood(2)



$$\mathbb{P}(\text{no point in}[t_n, T] \Big| \text{past at } t_n) = e^{-\int_{t_n}^{T} \lambda(s)ds}$$

Likelihood(3)



$$\mathbb{P}(1 \text{ point in} [t_n - dt_n, t_n] \Big| \text{past at } t_n - dt_n) \simeq \lambda(t_n) dt_n e^{-\int_{t_n}^{t_n + dt_n} \lambda(s) ds}$$

Likelihood(4)

$$\mathbb{P}(\text{no point in } [0, t_1 - dt_1], 1 \text{ point in } [t_1 - dt_1, t_1], ..., \text{no point in} [t_n, T])$$

$$= \mathbb{E}\left(\mathbf{1}_{\text{no point in } [0, t_1 - dt_1]} \cdots \mathbf{1}_{1 \text{ point in } [t_n - dt_n, t_n]} \times \mathbb{P}(\text{no point in} [t_n, T] \Big| \text{past at } t_n)\right)$$

$$= \mathbb{E}\left(\mathbf{1}_{\text{no point in } [0, t_1 - dt_1]} \cdots \mathbf{1}_{1 \text{ point in } [t_n - dt_n, t_n]}\right) e^{-\int_{t_n}^{T} \lambda(s) ds}$$

Likelihood(4)

$$\mathbb{P}(\text{no point in } [0, t_1 - dt_1], 1 \text{ point in } [t_1 - dt_1, t_1], ..., \text{no point in} [t_n, T])$$

$$= \mathbb{E}\left(\mathbf{1}_{\text{no point in } [0, t_1 - dt_1]} \cdots \mathbf{1}_{1 \text{ point in } [t_n - dt_n, t_n]} \times \mathbb{P}(\text{no point in} [t_n, T] \Big| \text{past at } t_n)\right)$$

$$= \mathbb{E}\left(\mathbf{1}_{\text{no point in } [0, t_1 - dt_1]} \cdots \mathbf{1}_{1 \text{ point in } [t_n - dt_n, t_n]}\right) e^{-\int_{t_n}^{T} \lambda(s) ds}$$

Likelihood(4)

$$\mathbb{P}(\text{no point in } [0, t_1 - dt_1], 1 \text{ point in } [t_1 - dt_1, t_1], ..., \text{no point in} [t_n, T])$$

$$= \mathbb{E}\left(\mathbf{1}_{\text{no point in } [0,t_1-dt_1]} \cdots \mathbf{1}_{1 \text{ point in } [t_n-dt_n,t_n]} \times \mathbb{P}(\text{no point in} [t_n, T] \Big| \text{past at } t_n)\right)$$

$$= \mathbb{E}\left(\mathbf{1}_{\text{no point in } [0,t_1-dt_1]} \cdots \mathbf{1}_{1 \text{ point in } [t_n-dt_n,t_n]}\right) e^{-\int_{t_n}^{T} \lambda(s)ds}$$

$$= \mathbb{E}\left(\mathbf{1}_{\text{no point in } [0,t_1-dt_1]} \cdots \mathbb{P}(1 \text{ point in} [t_n - dt_n, t_n] \Big| \text{past at } t_n - dt_n)\right) \times e^{-\int_{t_n}^{T} \lambda(s)ds}$$

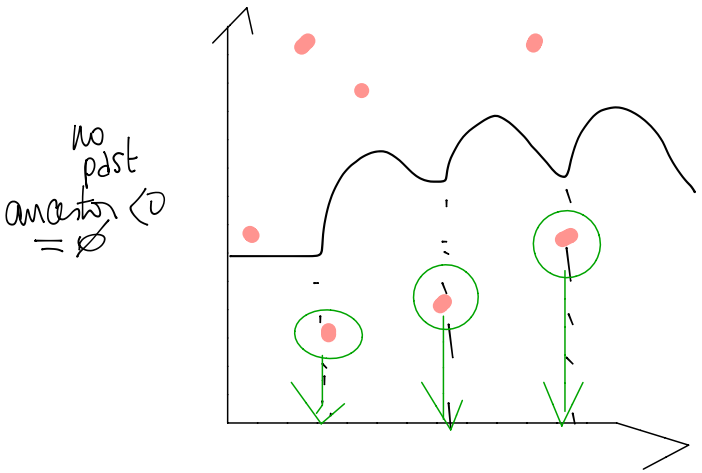
$$\simeq \mathbb{E}\left(\mathbf{1}_{\text{no point in } [0,t_1-dt_1]} \cdots \mathbf{1}_{\text{no point in } [t_{n-1},t_n-dt_n]}\right) \times \lambda(t_n)dt_n e^{-\int_{t_n-dt_n}^{T} \lambda(s)ds}$$

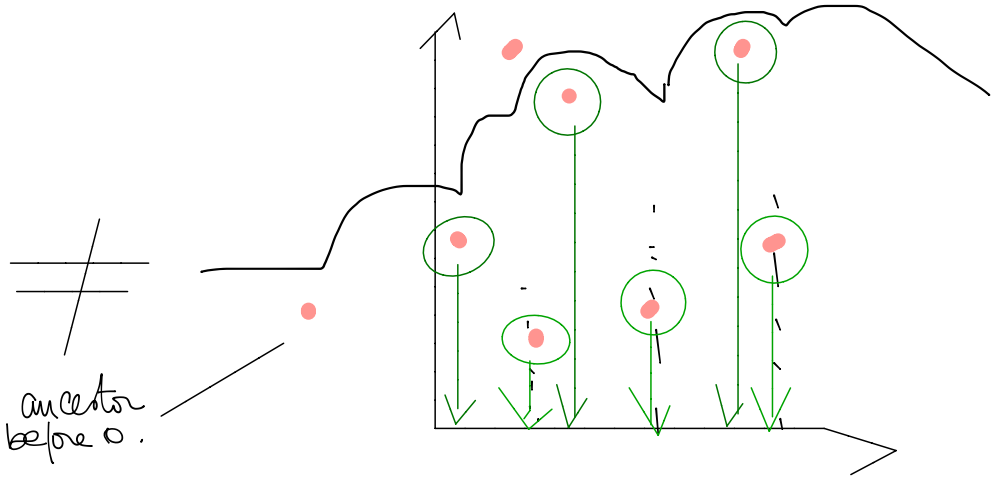
Likelihood and log-likelihood

(Pseudo)Likelihood

$$\prod_i \lambda(t_i) e^{-\int_0^T \lambda(s)ds}$$

dependance on the past not explicit (ex Past before 0 for Hawkes)





Likelihood and log-likelihood

(Pseudo)Likelihood

$$\prod_i \lambda(t_i) e^{-\int_0^T \lambda(s)ds}$$

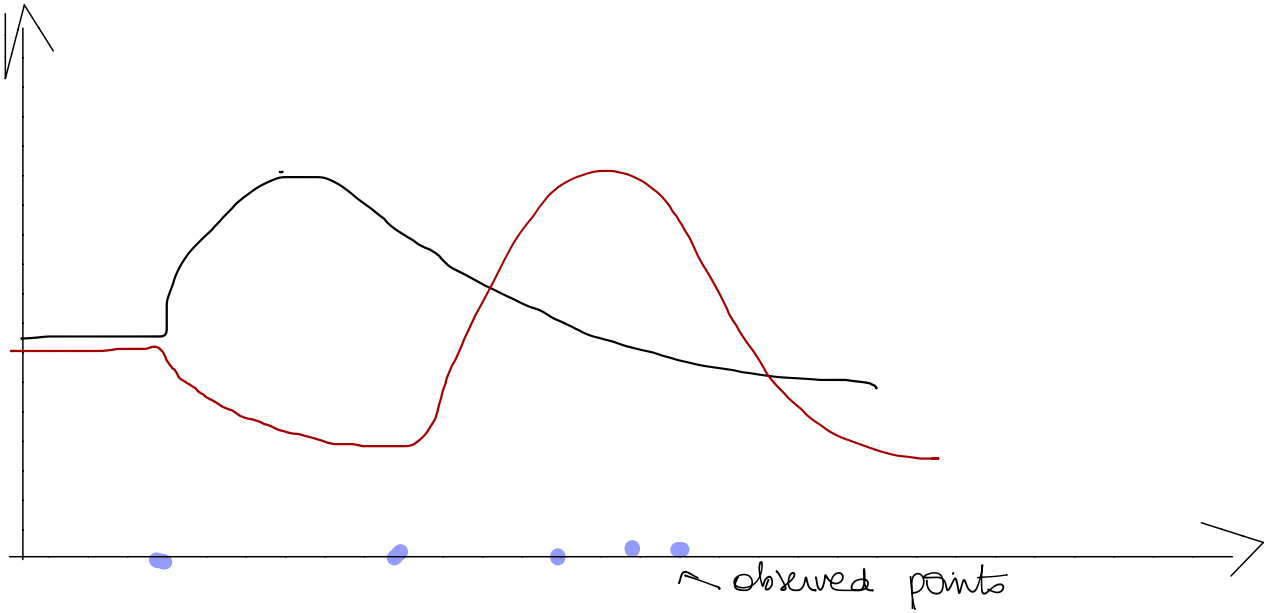
dependance on the past not explicit (ex Past before 0 for Hawkes)

LogLikelihood

$$\ell_\lambda(N) = \int_0^T \log[\lambda(s)] dN_s - \int_0^T \lambda(s)ds$$

Maximum likelihood estimator

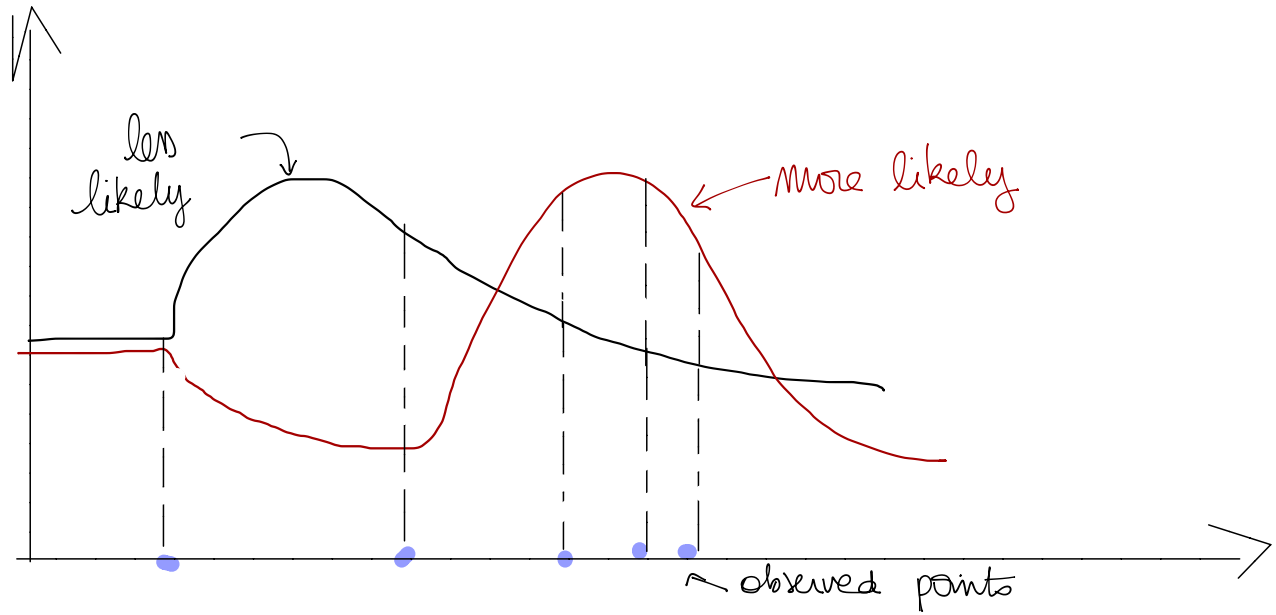
If model for $\lambda \to \lambda_\theta$, $\theta \in \mathbb{R}^d$, how to choose the best θ according to the observed data N on $[0, T]$?



Maximum likelihood estimator

If model for $\lambda \to \lambda_\theta$, $\theta \in \mathbb{R}^d$, how to choose the best θ according to the observed data N on $[0, T]$?

$$\ell_\lambda = \sum_{T_i \in (0,T)} \log\left(\lambda(T_i)\right) - \int_0^T \lambda(s)\, ds .$$



then $MLE = \hat{\theta} = \arg\max_\theta \ell_{\lambda_\theta}(N)$.

Maximum likelihood estimator(2)

If d, number of unknown parameter, fixed one can show usually nice asymptotic properties (in T ...) :

- consistency ($\hat{\theta} \to \theta$)
- asymptotic normality
- efficiency (smallest asymptotic variance)

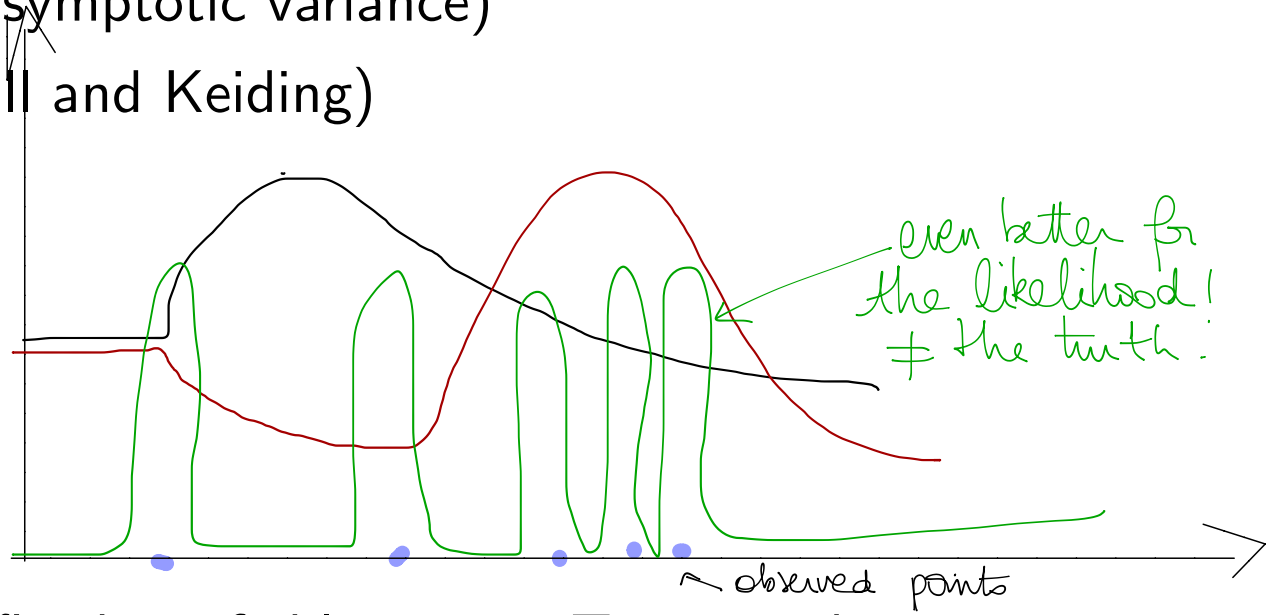
(cf. Andersen, Borgan, Gill and Keiding)

Maximum likelihood estimator(2)

If d, number of unknown parameter, fixed one can show usually nice asymptotic properties (in T ...) :

- consistency ($\hat{\theta} \to \theta$)
- asymptotic normality
- efficiency (smallest asymptotic variance)

(cf. Andersen, Borgan, Gill and Keiding)

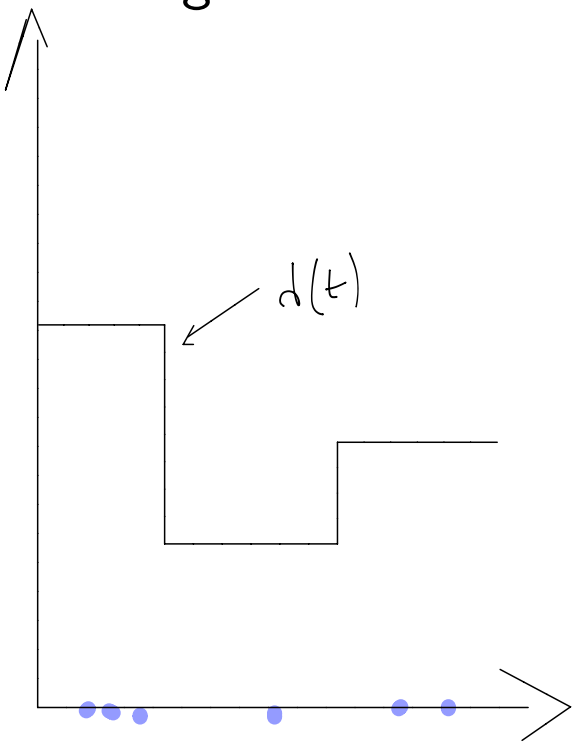


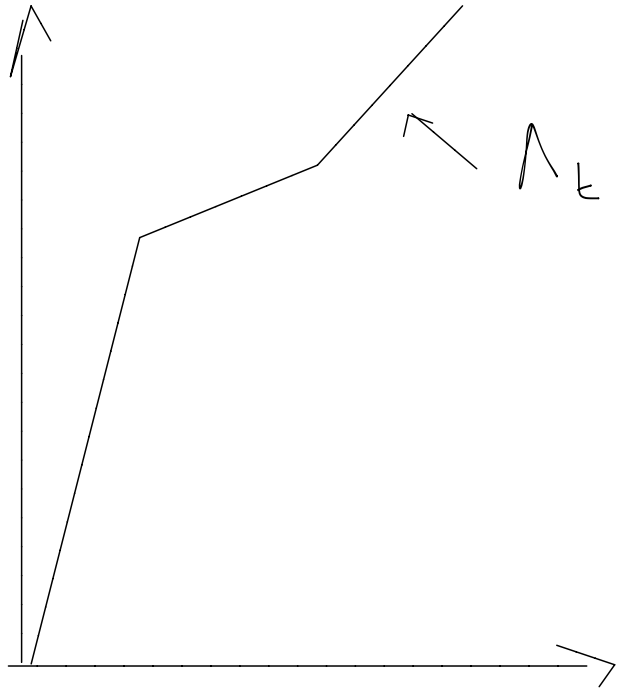
In practice, T and d are fixed $\to$ if d large wrt T, i.e. in the non asymptotic regime, over fitting.

Goodness-of-fit test

If $\Lambda_t = \int_0^t \lambda(s)ds$, then

- Λ_t is a nondecreasing process, predictable and its (pseudo)inverse exists.
- It is the compensator of the counting process N_t, i.e. $(N_t - \Lambda_t)_t$ martingale.

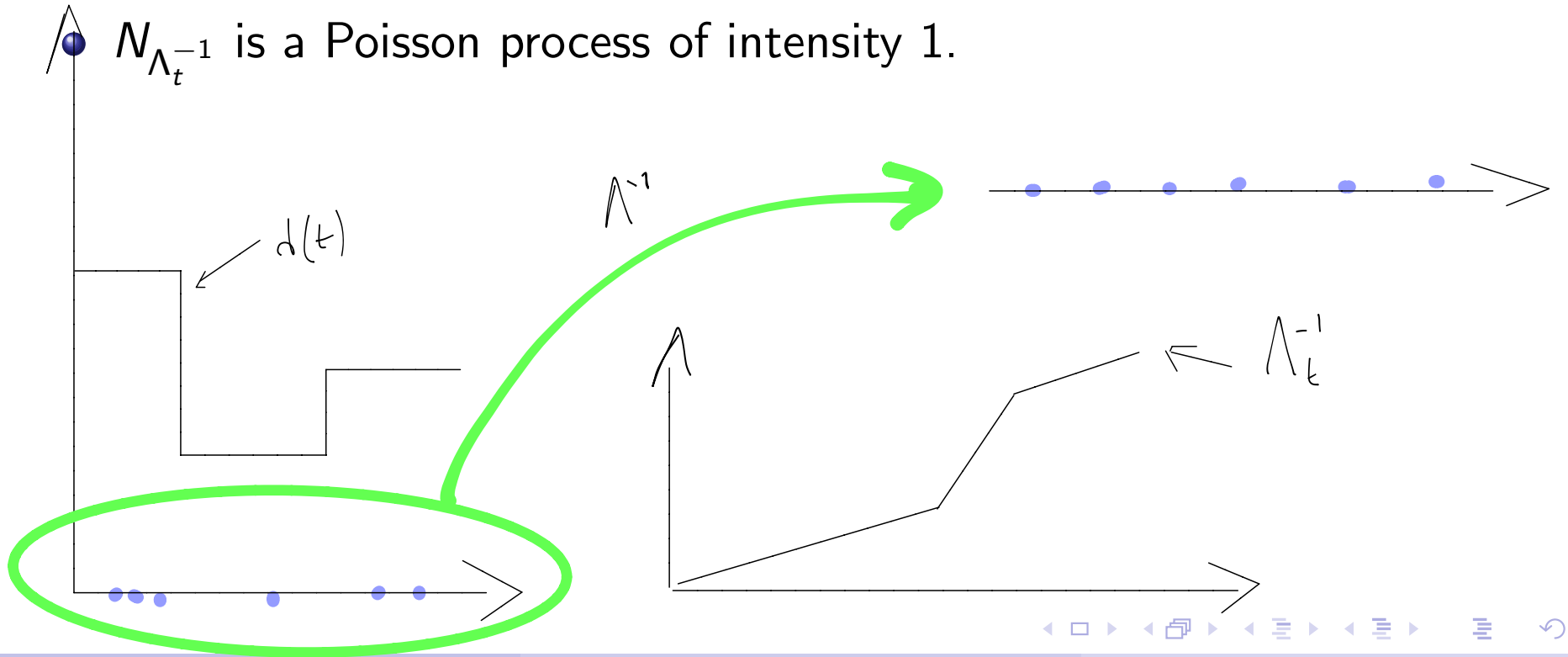




Goodness-of-fit test

If $\Lambda_t = \int_0^t \lambda(s)ds$, then

- Λ_t is a nondecreasing process, predictable and its (pseudo)inverse exists.
- It is the compensator of the counting process N_t, i.e. $(N_t - \Lambda_t)_t$ martingale.
- $N_{\Lambda_t^{-1}}$ is a Poisson process of intensity 1.



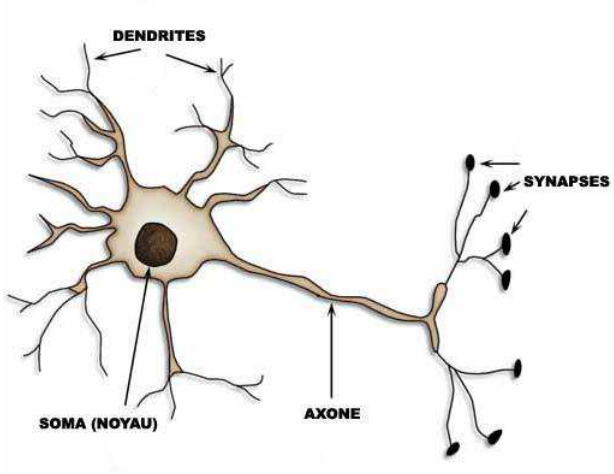
Goodness-of-fit test

If $\Lambda_t = \int_0^t \lambda(s)ds$, then

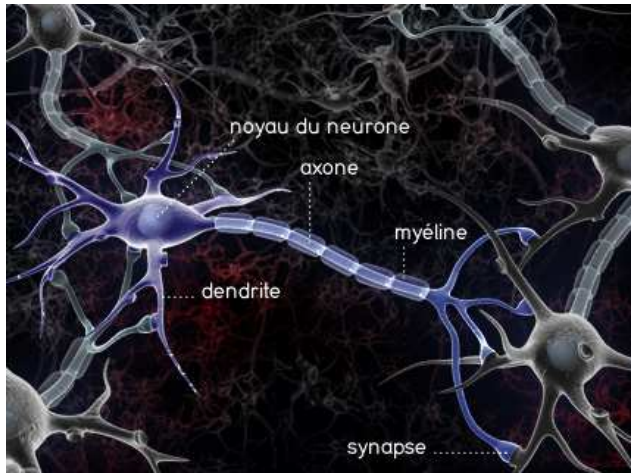
- Λ_t is a nondecreasing process, predictable and its (pseudo)inverse exists.
- It is the compensator of the counting process N_t, i.e. $(N_t - \Lambda_t)_t$ martingale.
- $N_{\Lambda_t^{-1}}$ is a Poisson process of intensity 1.
- $\rightarrow$ test that N (observed) has the desired intensity λ.

Table of Contents

Biological framework



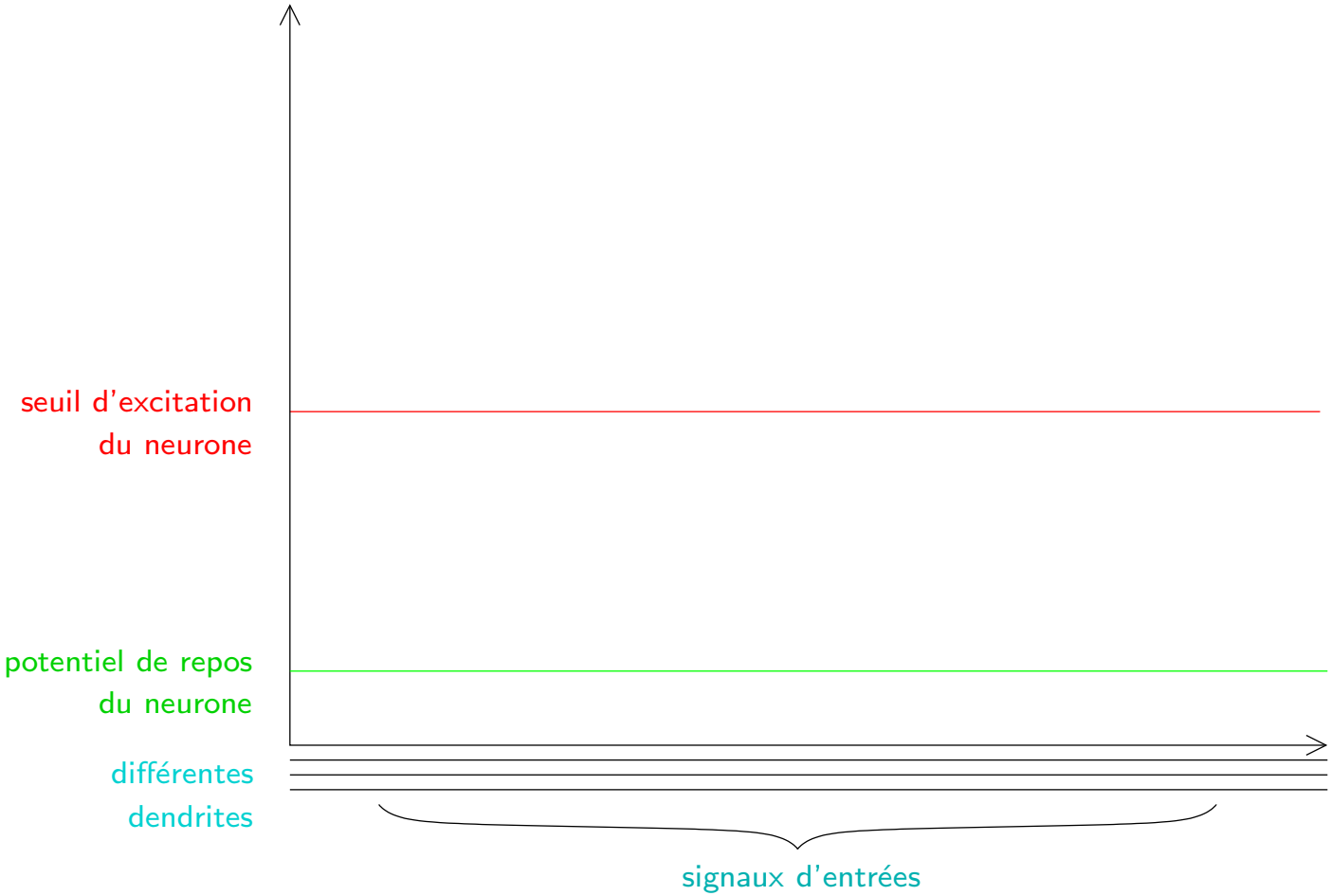
(a) One neuron



(b) Connected neurons

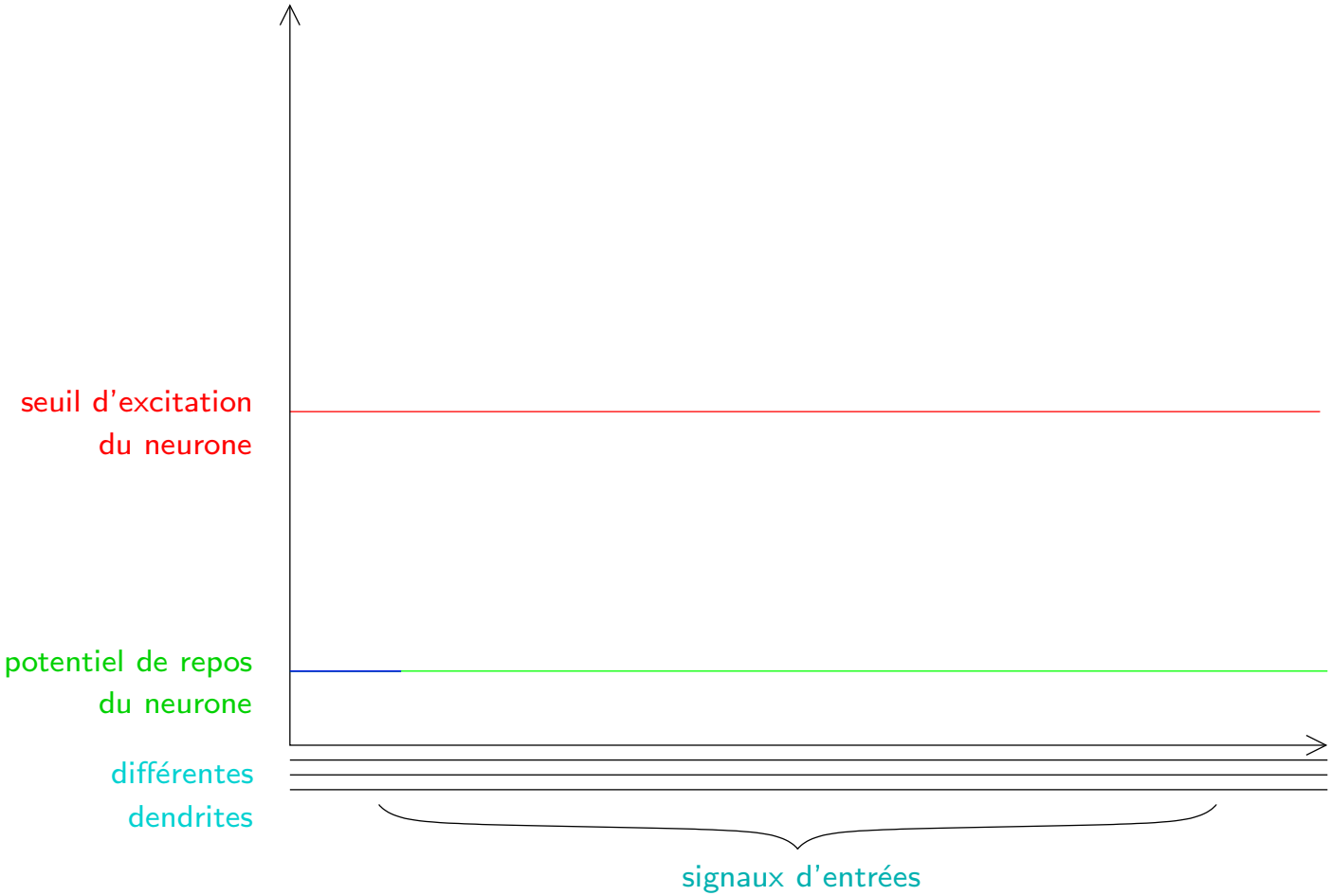
Synaptic integration

without synchronization



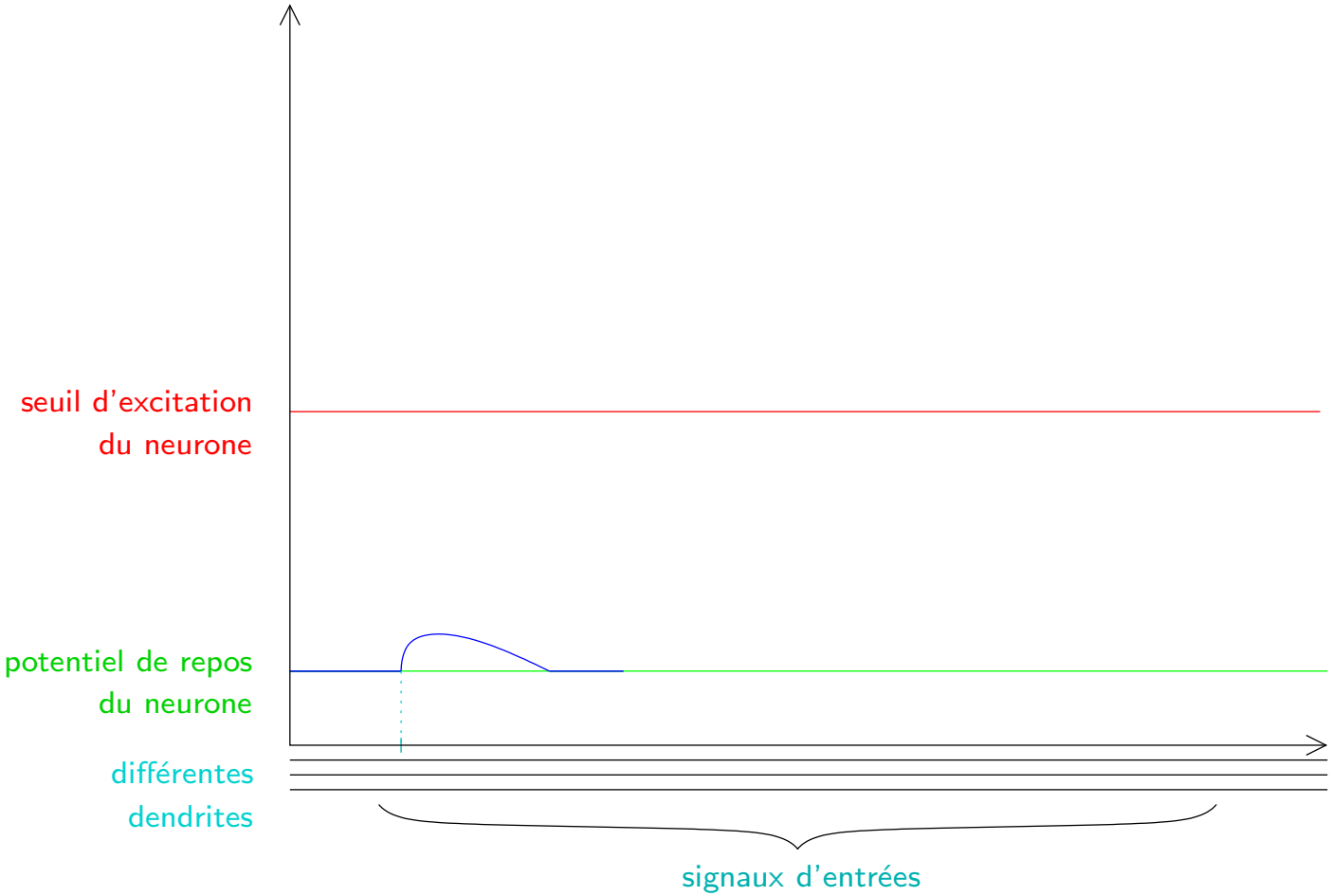
Synaptic integration

without synchronization



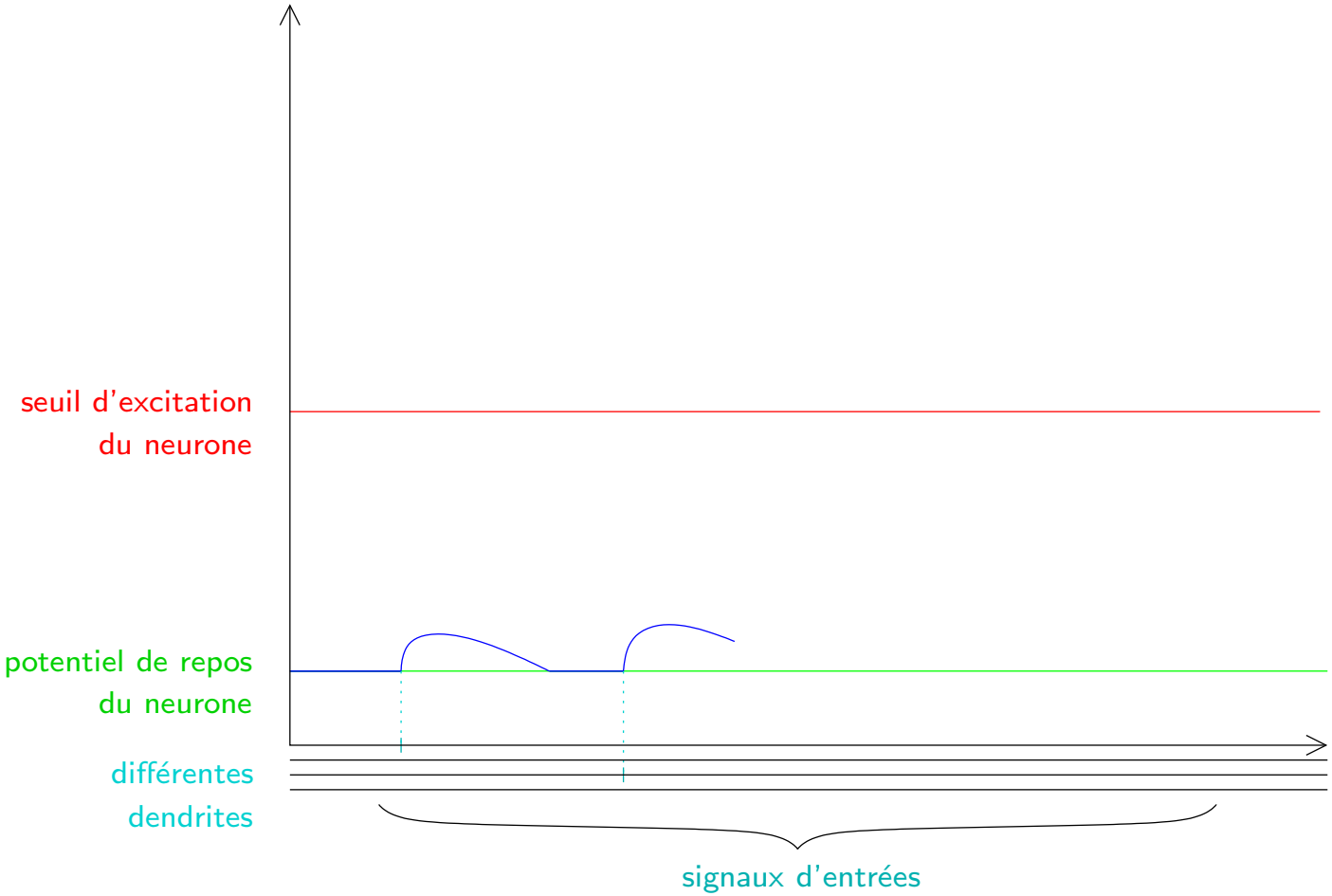
Synaptic integration

without synchronization



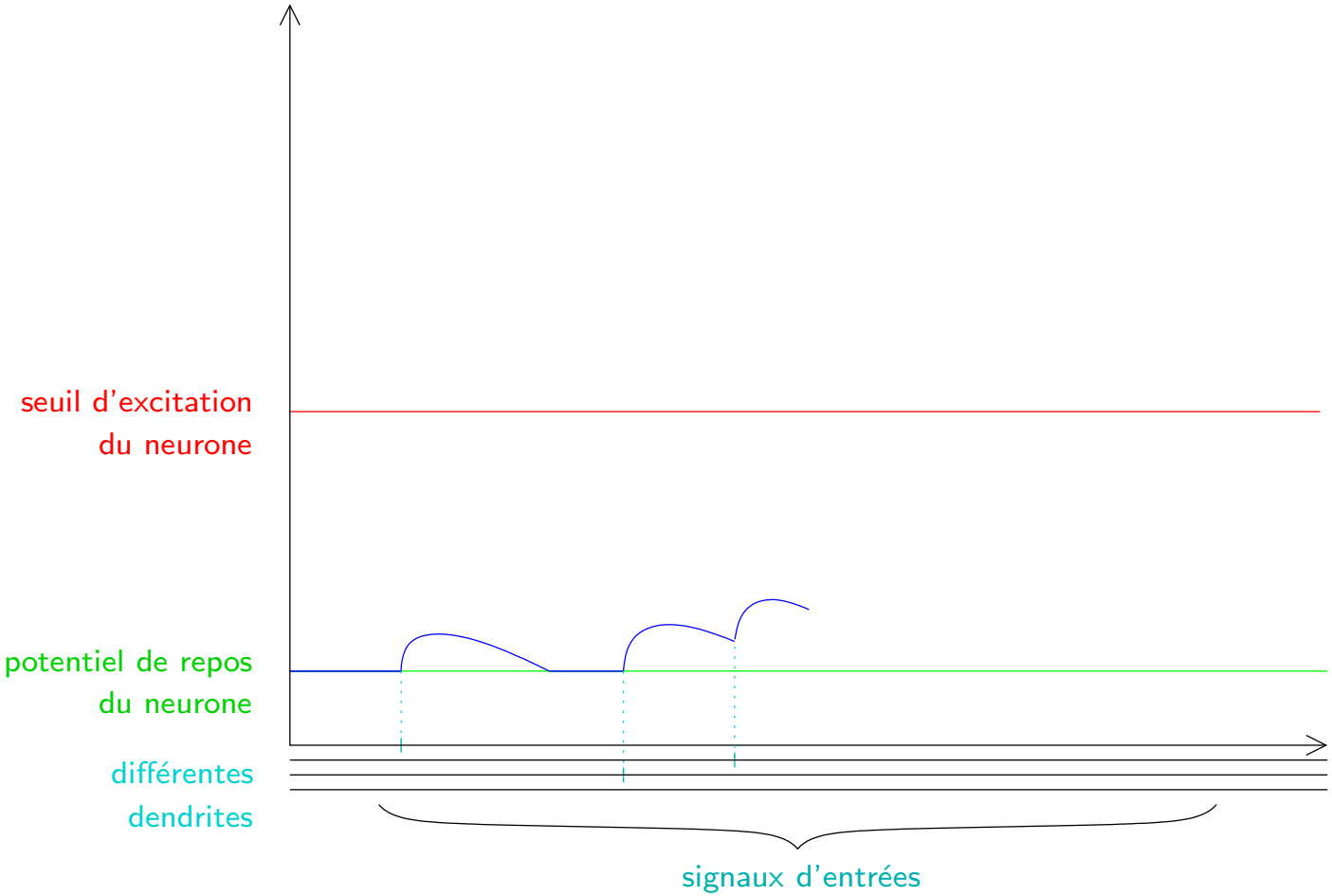
Synaptic integration

without synchronization



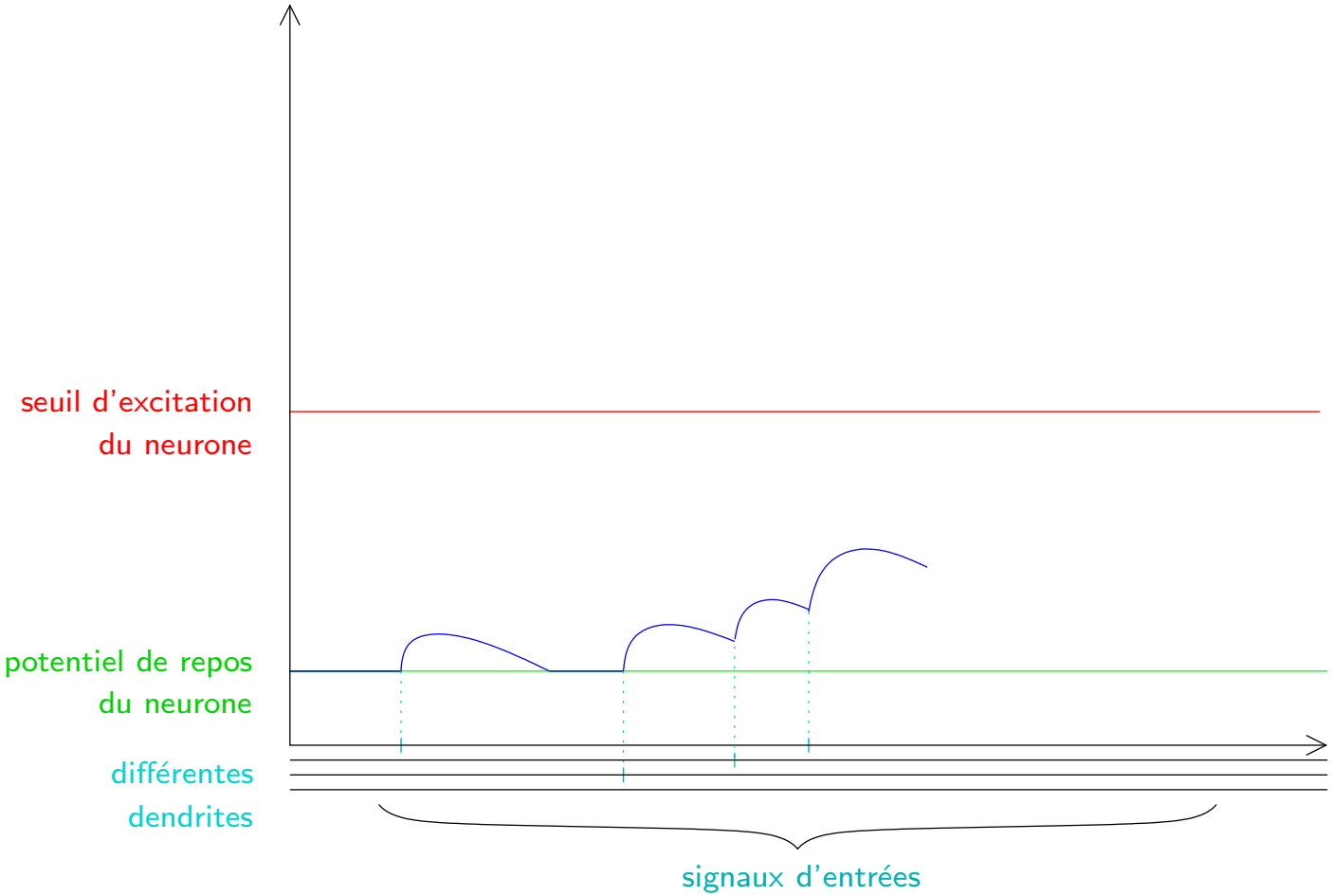
Synaptic integration

without synchronization



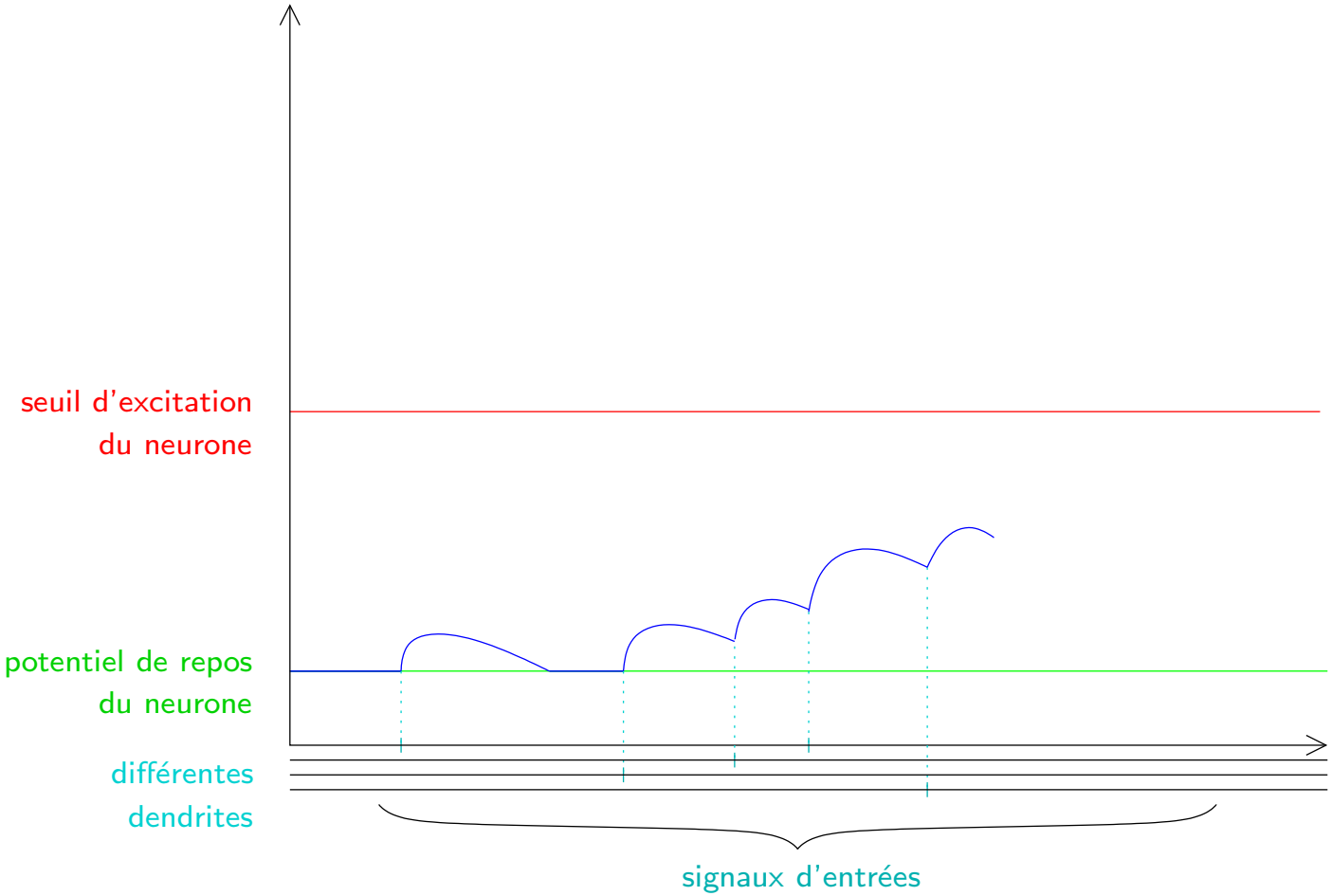
Synaptic integration

without synchronization



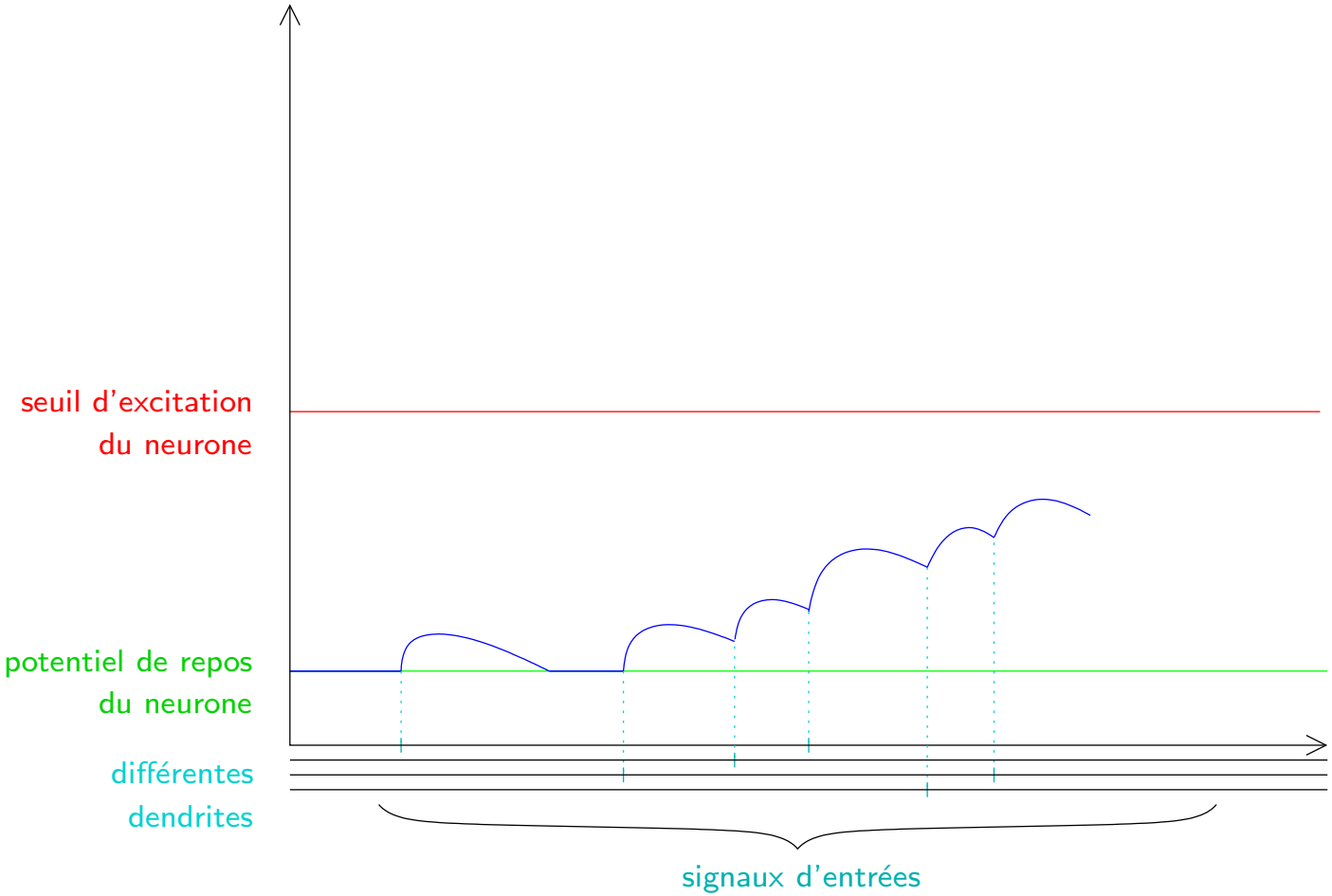
Synaptic integration

without synchronization



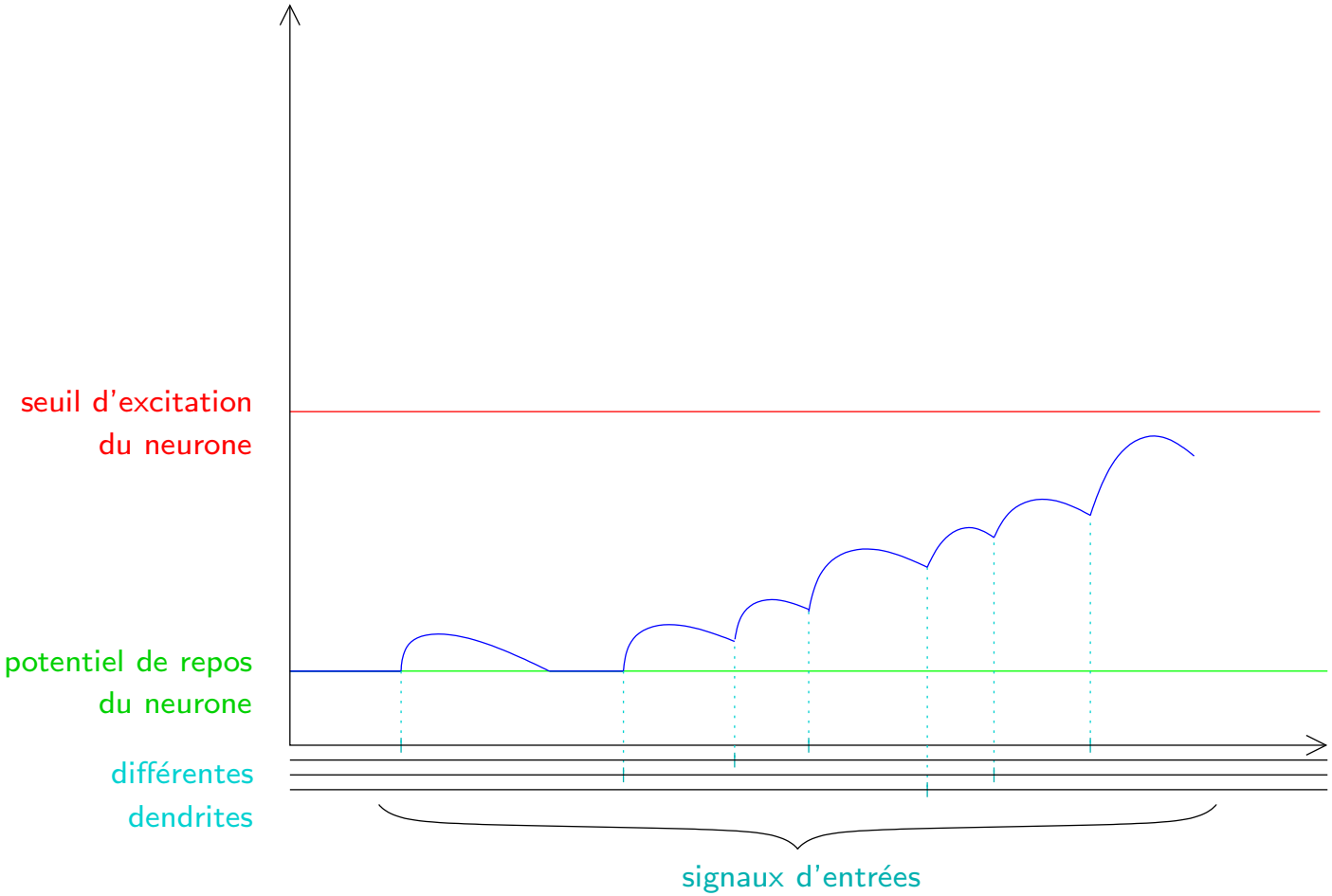
Synaptic integration

without synchronization



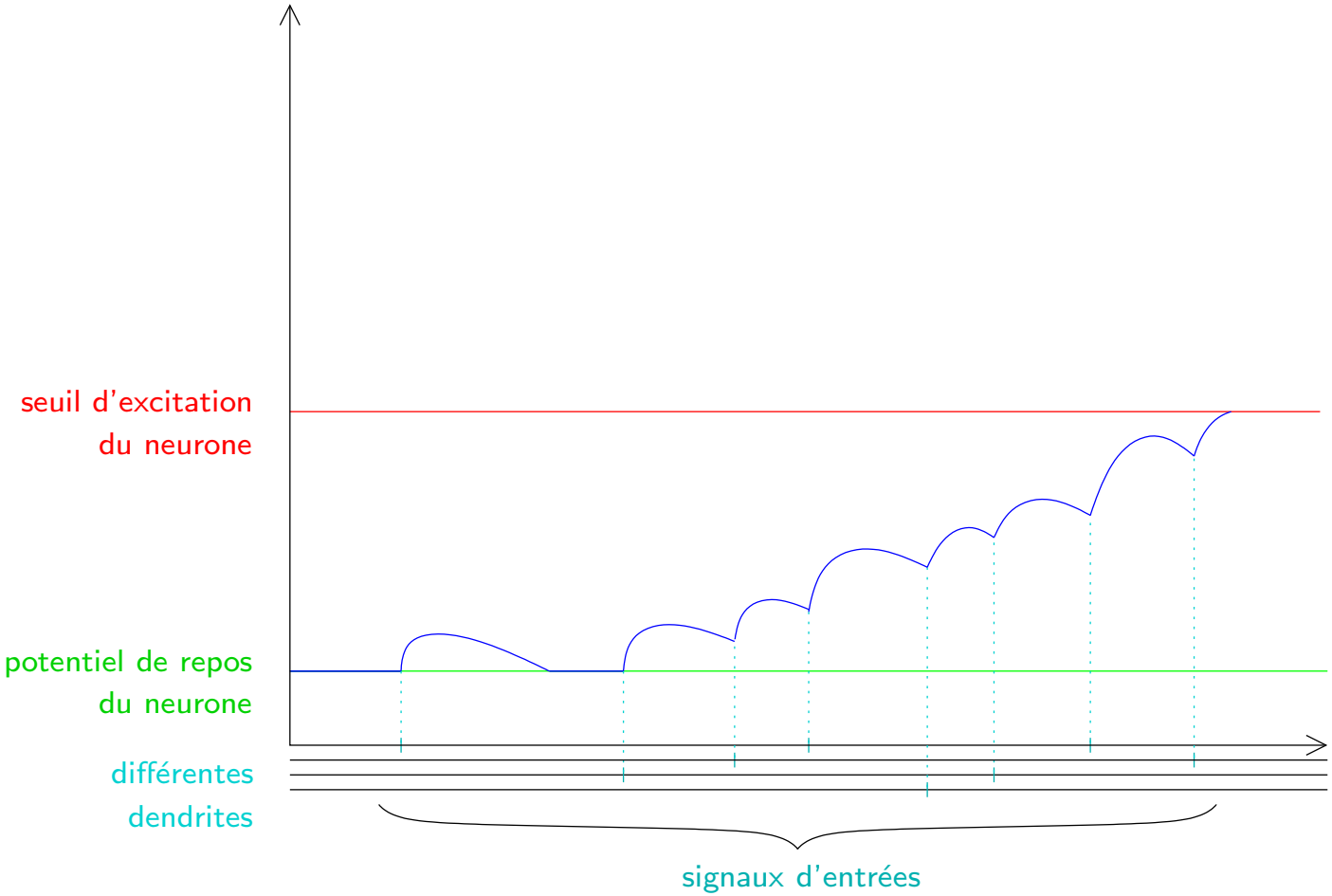
Synaptic integration

without synchronization



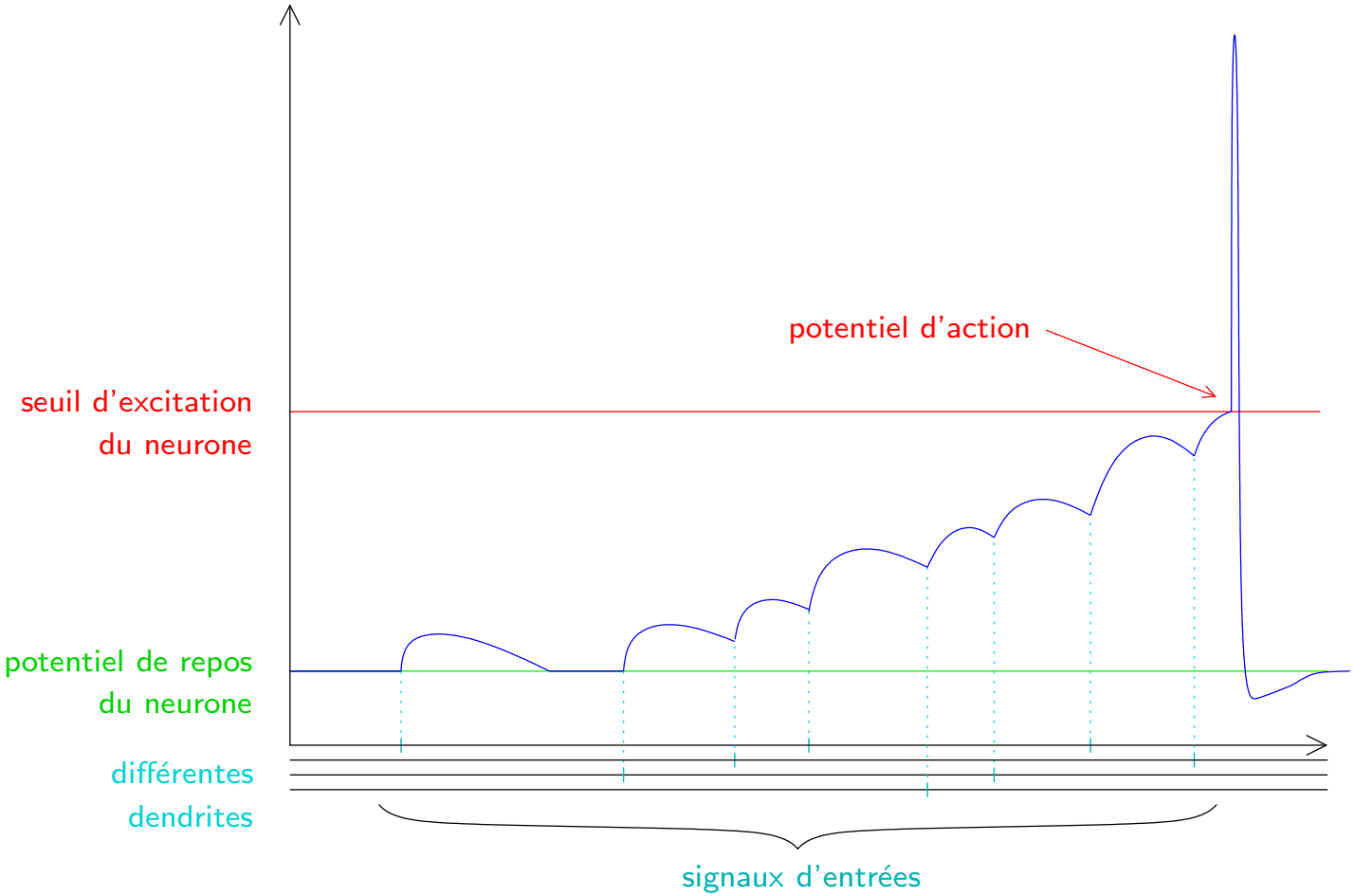
Synaptic integration

without synchronization



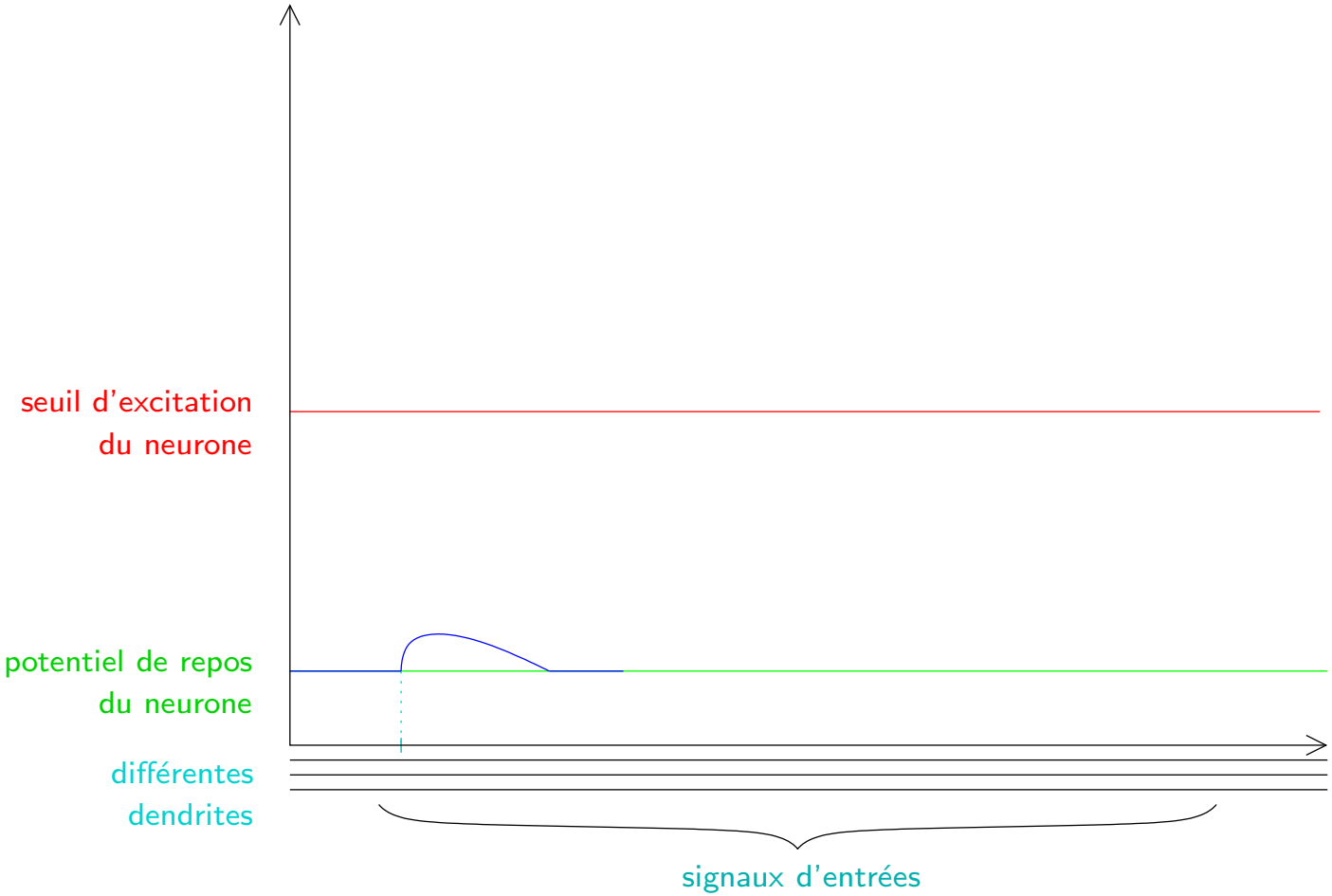
Synaptic integration

without synchronization



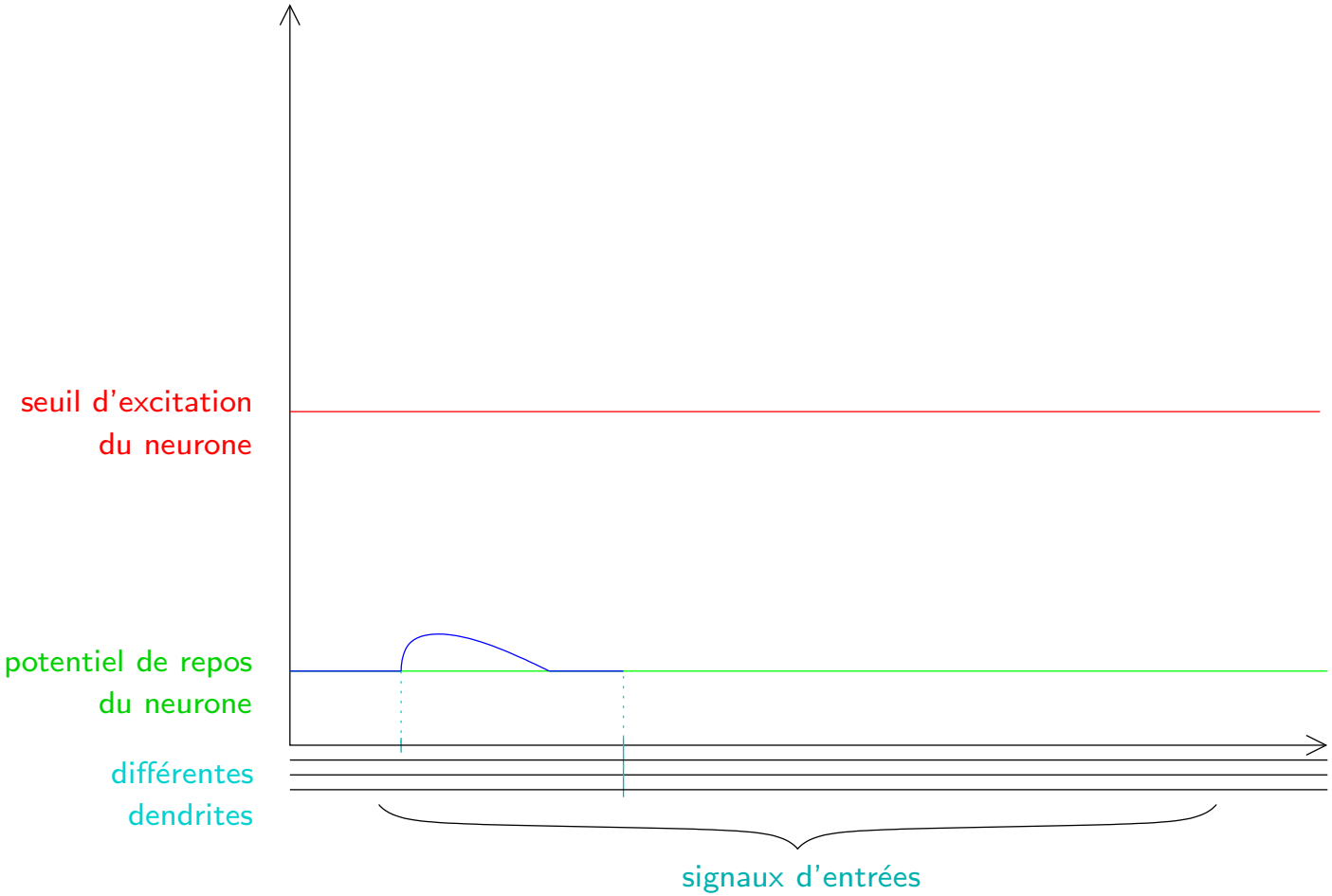
Synaptic integration

with synchronization



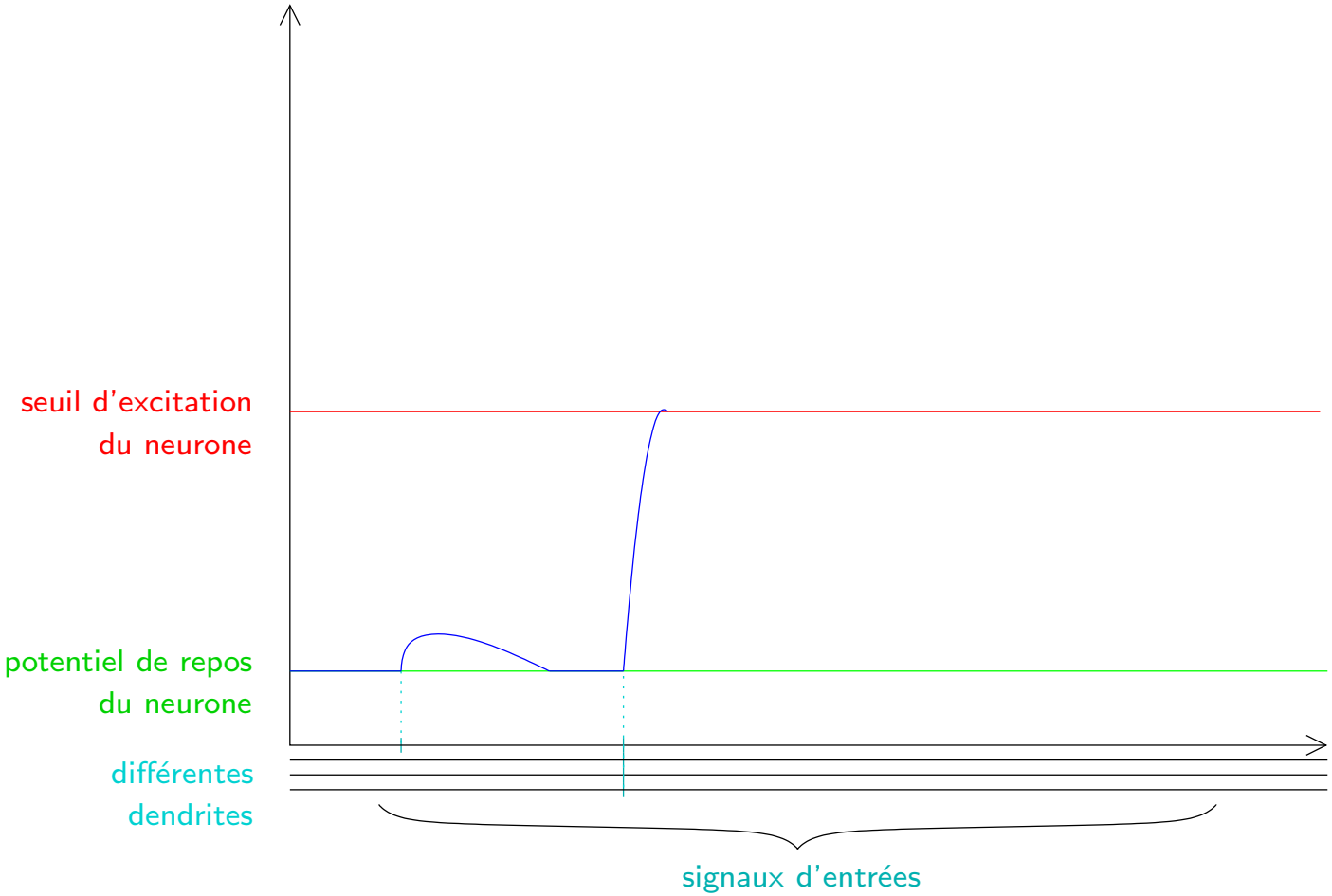
Synaptic integration

with synchronization



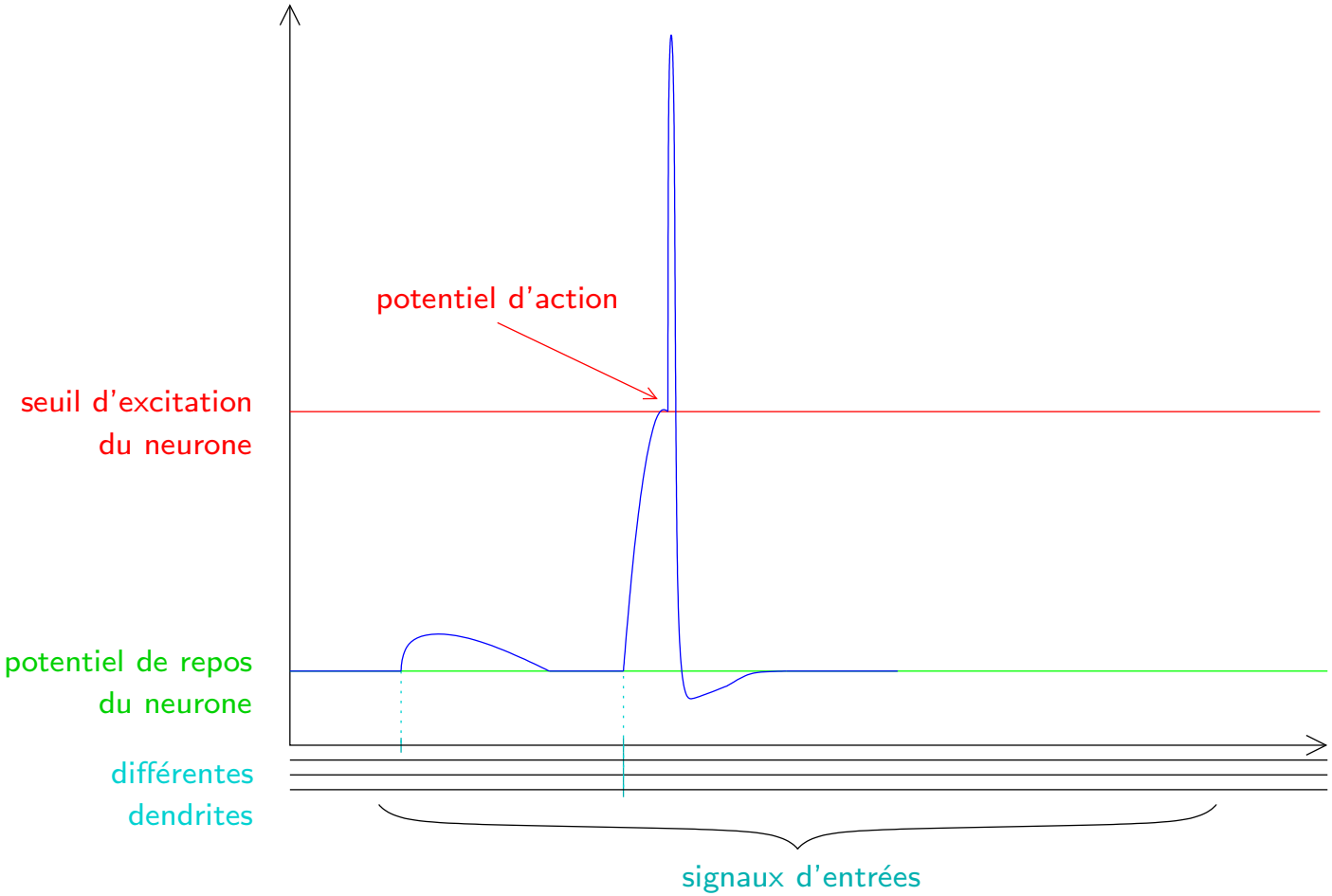
Synaptic integration

with synchronization



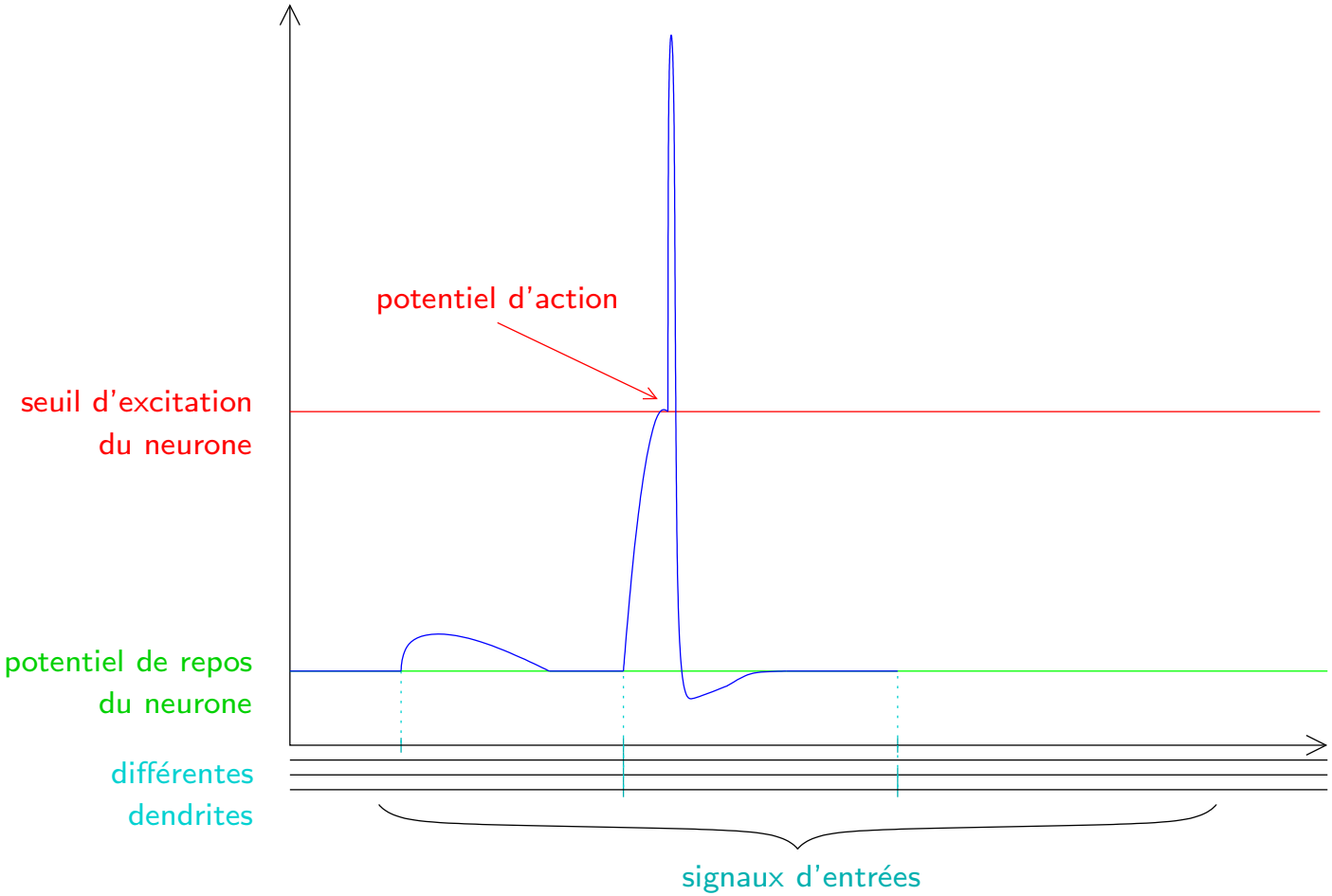
Synaptic integration

with synchronization



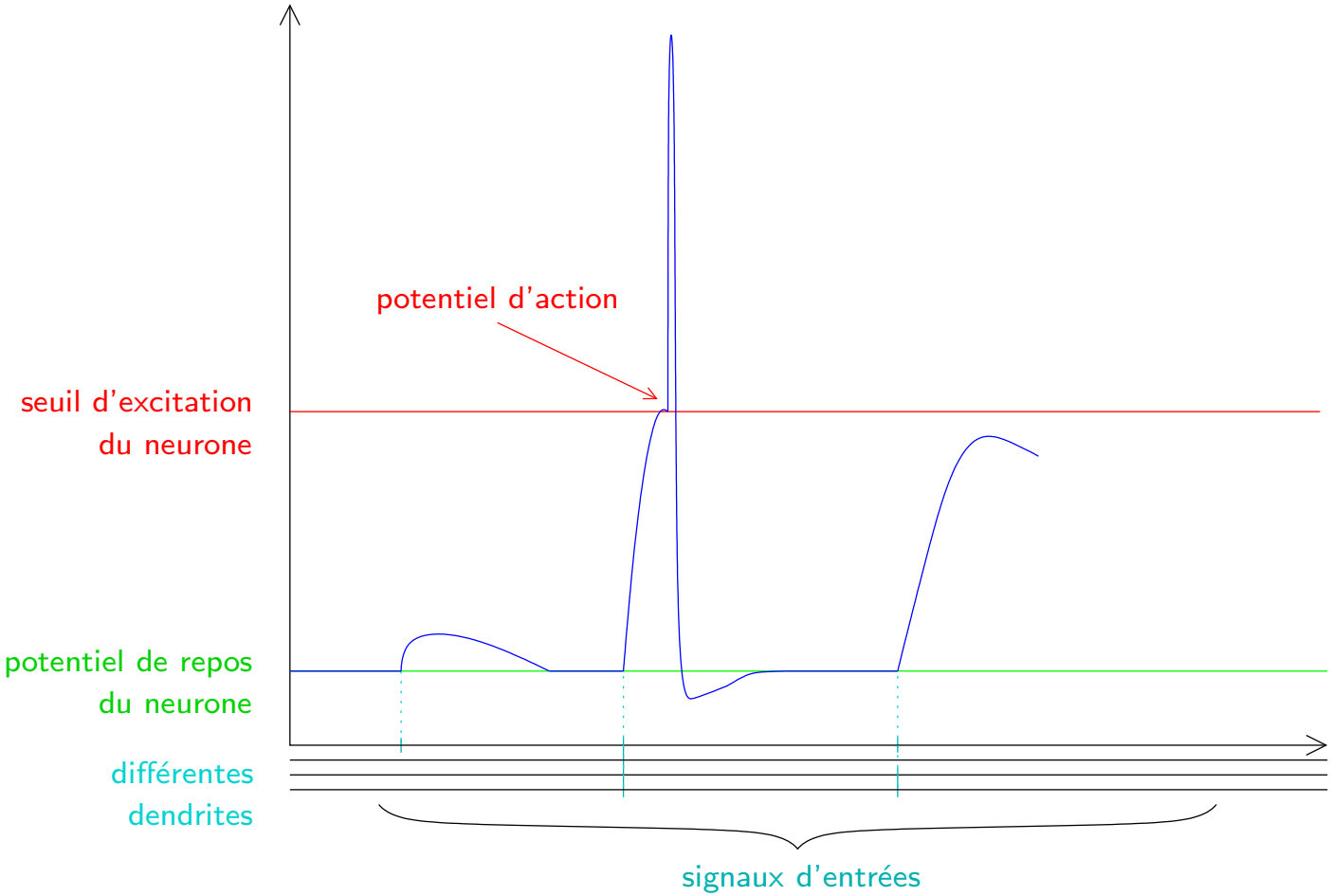
Synaptic integration

with synchronization



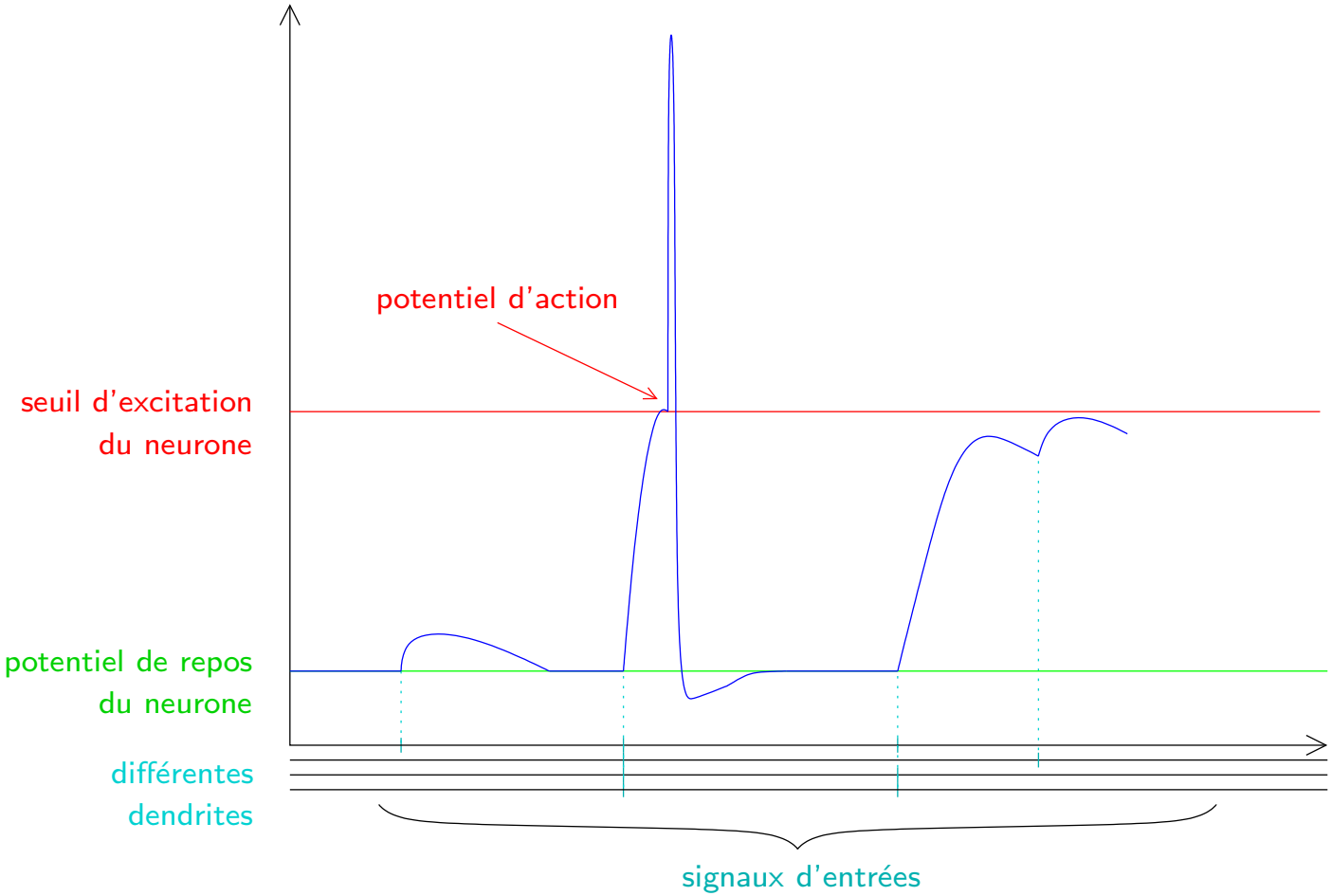
Synaptic integration

with synchronization



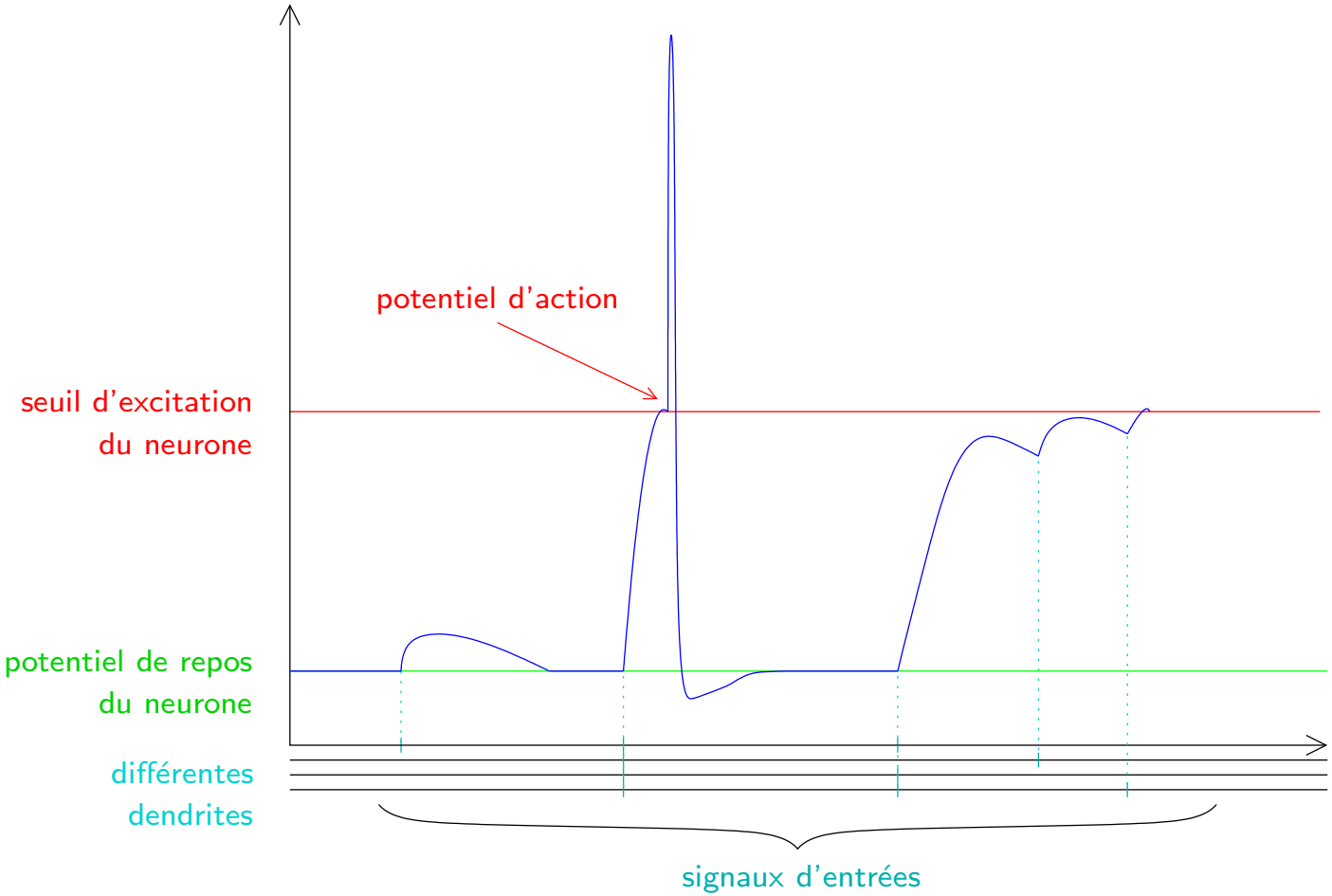
Synaptic integration

with synchronization



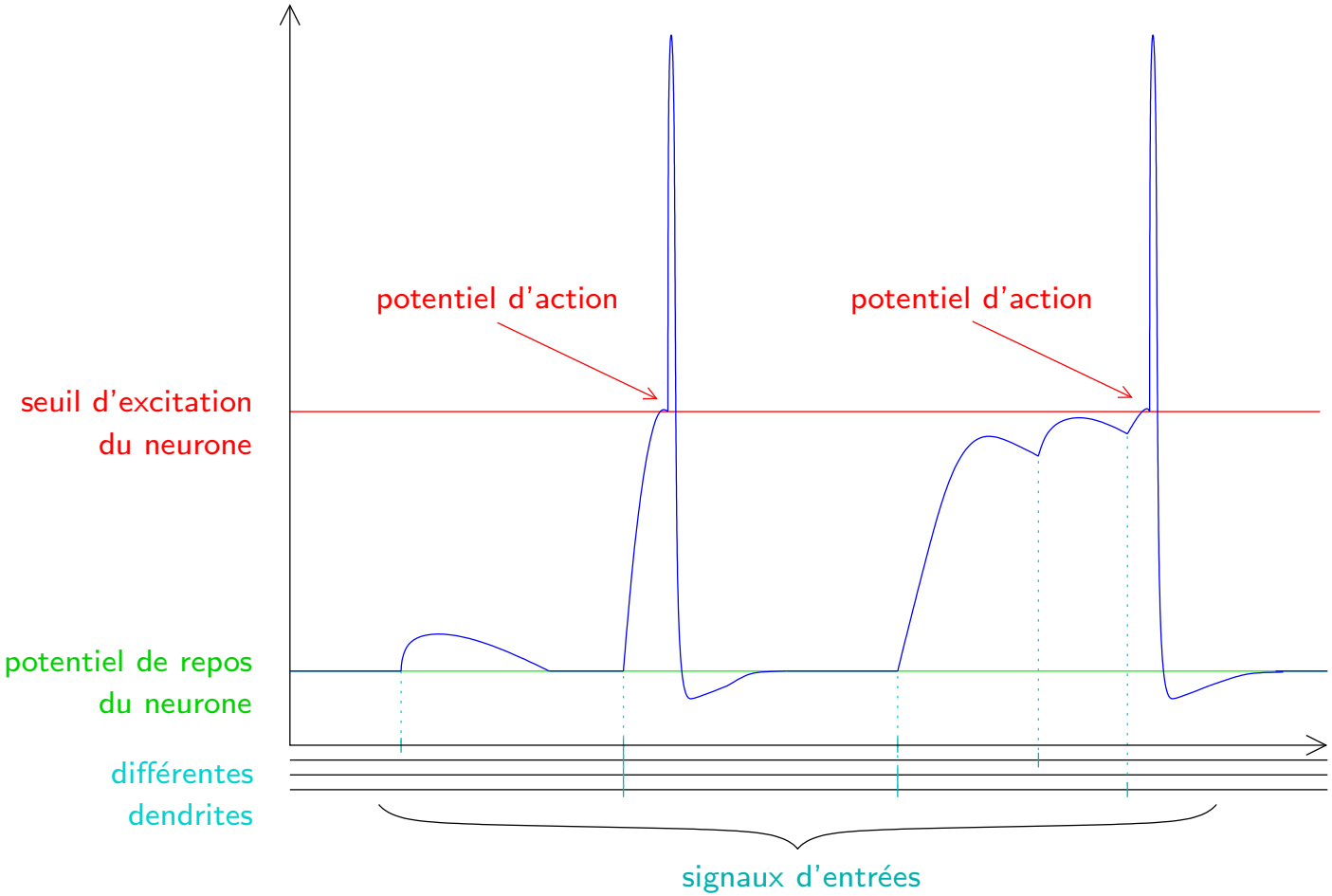
Synaptic integration

with synchronization



Synaptic integration

with synchronization



Point process and genomics

There are several "events" of different types on the DNA that may "work" together in synergy.

Point process and genomics

There are several "events" of different types on the DNA that may "work" together in synergy.

Motifs

= words in the DNA-alphabet $\{\texttt{actg}\}$.

Point process and genomics

There are several "events" of different types on the DNA that may "work" together in synergy.

Motifs

= words in the DNA-alphabet $\{\mathtt{actg}\}$.

How can statistician suggest functional motifs based on the statistical properties of their occurrences ?

- Unexpected frequency $\rightarrow$ Markov models (see for a review Reinert, Schbath, Waterman (2000))

Point process and genomics

There are several "events" of different types on the DNA that may "work" together in synergy.

Motifs

= words in the DNA-alphabet $\{\mathtt{actg}\}$.

How can statistician suggest functional motifs based on the statistical properties of their occurrences ?

- Unexpected frequency $\rightarrow$ Markov models (see for a review Reinert, Schbath, Waterman (2000))
- Poor or rich regions $\rightarrow$ scan statistics (see, for instance, Robin Daudin (1999) or Stefanov (2003))

Point process and genomics

There are several "events" of different types on the DNA that may "work" together in synergy.

Motifs

= words in the DNA-alphabet {actg}.

How can statistician suggest functional motifs based on the statistical properties of their occurrences ?

- Unexpected frequency $\rightarrow$ Markov models (see for a review Reinert, Schbath, Waterman (2000))
- Poor or rich regions $\rightarrow$ scan statistics (see, for instance, Robin Daudin (1999) or Stefanov (2003))
- If two motifs are part of a common biological process, the space between their occurrences (not necessarily consecutive) should be somehow fixed $\rightarrow$ favored or avoided distances (Gusto, Schbath (2005))

Point process and genomics

There are several "events" of different types on the DNA that may "work" together in synergy.

Motifs

= words in the DNA-alphabet $\{\mathtt{actg}\}$.

How can statistician suggest functional motifs based on the statistical properties of their occurrences ?

- Unexpected frequency $\rightarrow$ Markov models (see for a review Reinert, Schbath, Waterman (2000))
- Poor or rich regions $\rightarrow$ scan statistics (see, for instance, Robin Daudin (1999) or Stefanov (2003))
- If two motifs are part of a common biological process, the space between their occurrences (not necessarily consecutive) should be somehow fixed $\rightarrow$ favored or avoided distances (Gusto, Schbath (2005)) pairwise study.

Why real multivariate point processes in genomics ?

TRE

Transcription Regulatory Elements = "everything" that may enhance or repress gene expression

Why real multivariate point processes in genomics ?

TRE

Transcription Regulatory Elements = "everything" that may enhance or repress gene expression

- promoter, enhancer, silencer, histone modifications, replication origin on the DNA.... They should interact but how ? Can we have a statistical guess ?

Why real multivariate point processes in genomics ?

TRE

Transcription Regulatory Elements = "everything" that may enhance or repress gene expression

- promoter, enhancer, silencer, histone modifications, replication origin on the DNA.... They should interact but how ? Can we have a statistical guess ?
- There are methods (ChIP-chip experiments, ChIP-seq experiments) where after preprocessing the data one has access to the (almost exact) positions of several type of TREs at one time, and this under different experimental conditions. (ENCODE)

Why real multivariate point processes in genomics ?

TRE

Transcription Regulatory Elements = "everything" that may enhance or repress gene expression

- promoter, enhancer, silencer, histone modifications, replication origin on the DNA.... They should interact but how ? Can we have a statistical guess ?
- There are methods (ChIP-chip experiments, ChIP-seq experiments) where after preprocessing the data one has access to the (almost exact) positions of several type of TREs at one time, and this under different experimental conditions. (ENCODE)
- On the real line = DNA if the 3D structure of the DNA is negligible (typically interaction range between points $\leq$ 10 kB)

Why real multivariate point processes in genomics ?

TRE

Transcription Regulatory Elements = "everything" that may enhance or repress gene expression

- promoter, enhancer, silencer, histone modifications, replication origin on the DNA.... They should interact but how ? Can we have a statistical guess ?
- There are methods (ChIP-chip experiments, ChIP-seq experiments) where after preprocessing the data one has access to the (almost exact) positions of several type of TREs at one time, and this under different experimental conditions. (ENCODE)
- On the real line = DNA if the 3D structure of the DNA is negligible (typically interaction range between points $\leq$ 10 kB)
- If the real structure $\rightarrow$ 3D point processes on graphs... (? ?)

Why just DNA ? RNA etc ...

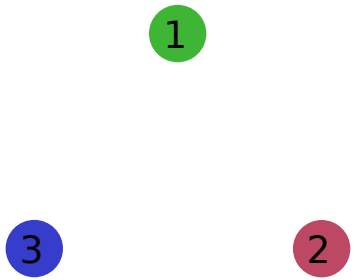
Point processes and conditional intensity

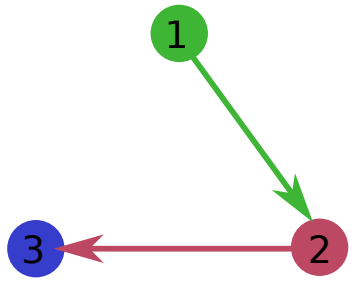
$$\underbrace{dN_t}_{\text{Nbr observed points in } [t,t+dt]} = \overbrace{\underbrace{\lambda(t)\,dt}_{\text{Expected Number given the past before } t}}^{\textbf{intensity}} + \underbrace{\text{noise}}_{\textit{Martingales differences}}$$

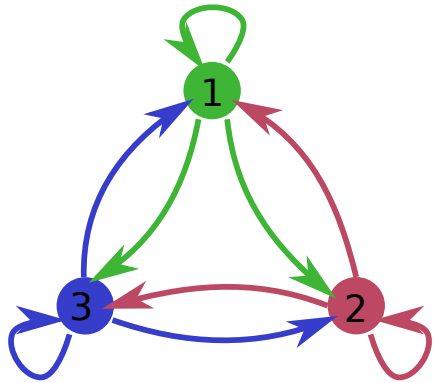
$$\lambda(t) = \text{instantaneous frequency} = \textbf{random}, \text{ depends on previous points}$$

Local independence graphs

(Didelez (2008))

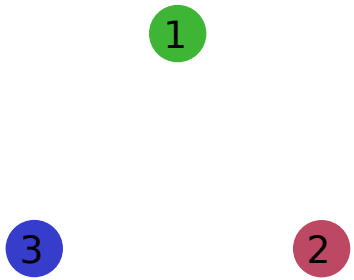


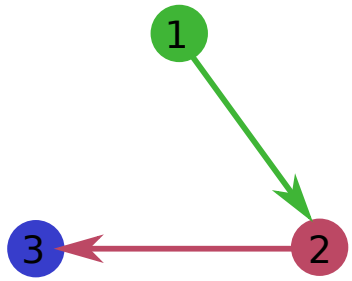


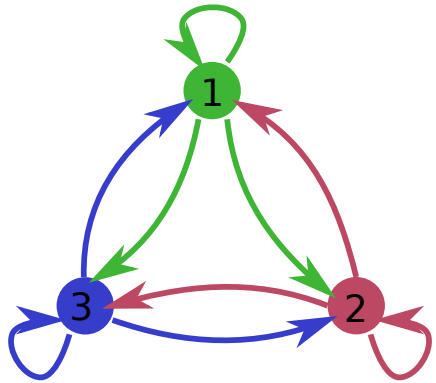


Local independence graphs

(Didelez (2008))

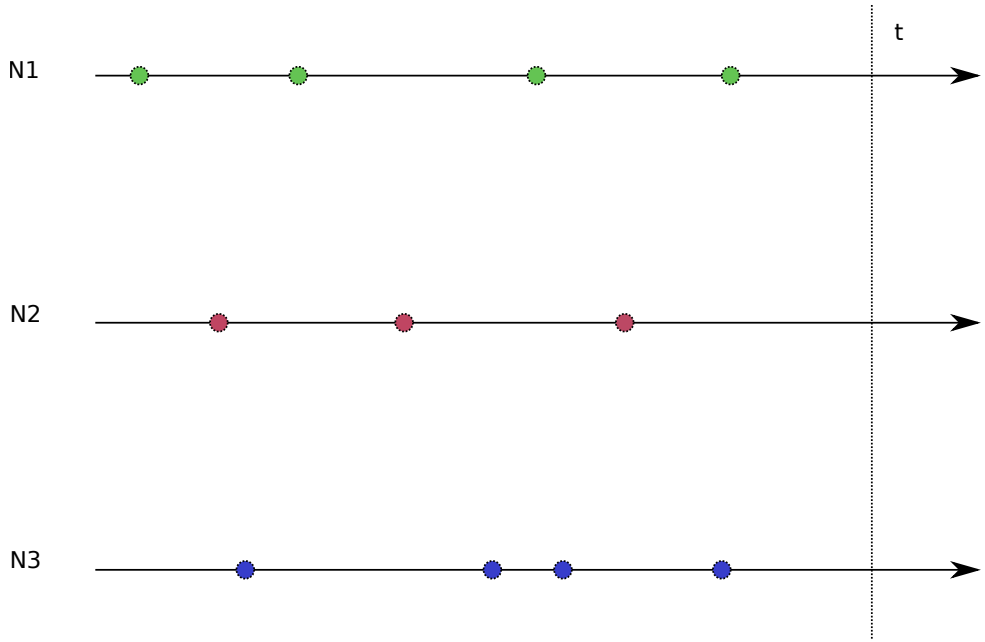




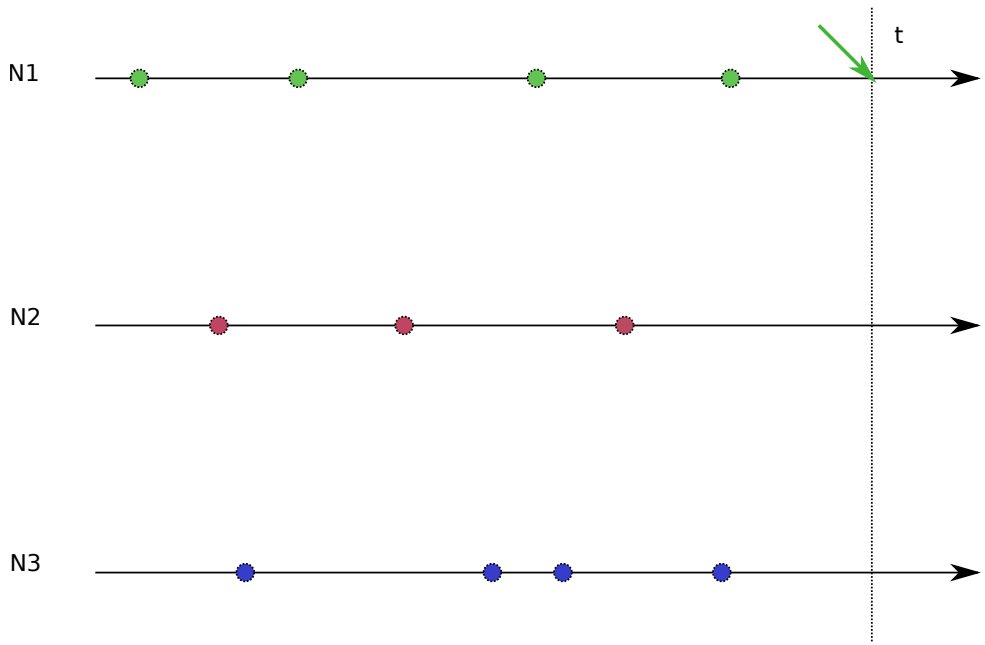


If one is able to infer a local independence graph, then we may have access to "functional connectivity".
$\hookrightarrow$ needs a "model" .

Multivariate Hawkes processes

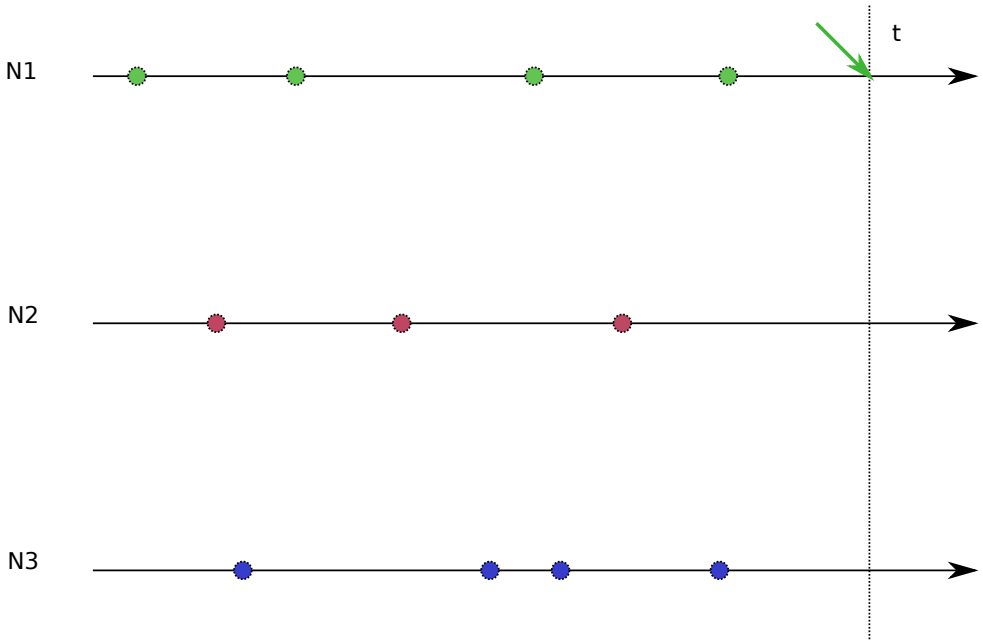


Multivariate Hawkes processes



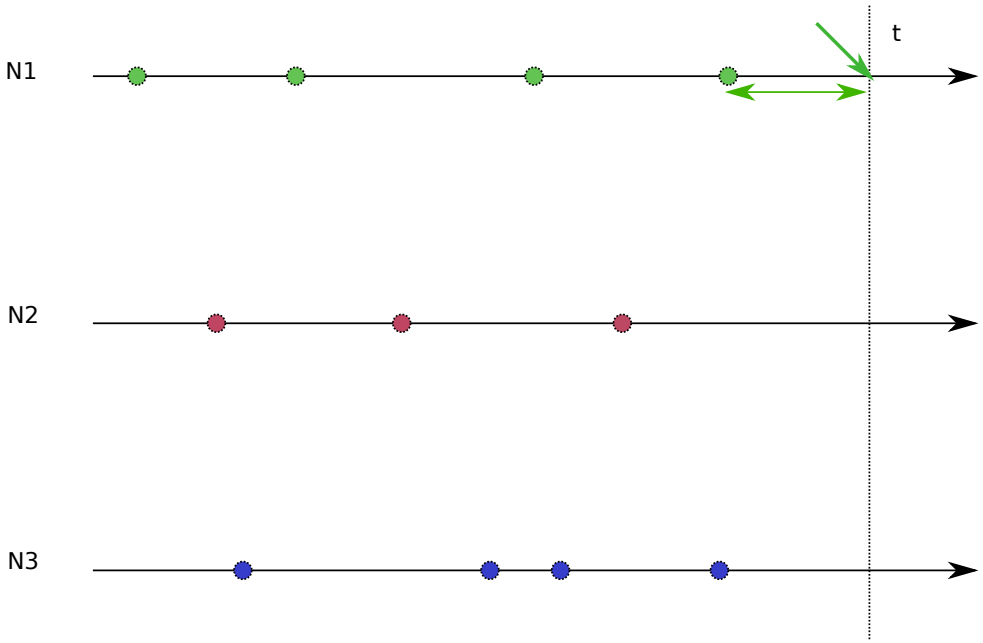
$$\lambda^{(1)}(t) = \underbrace{\nu}_{\text{spontaneous}}$$

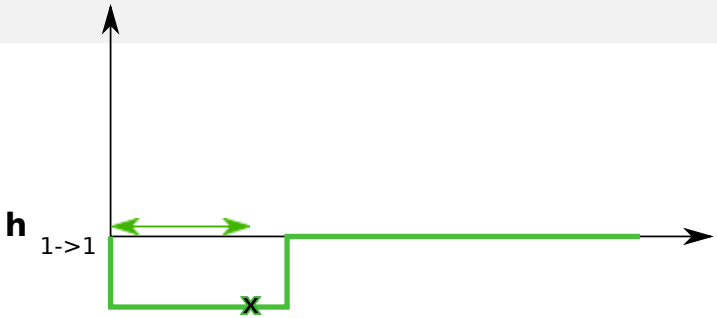
Multivariate Hawkes processes



$$\lambda^{(1)}(t) = \underbrace{\nu}_{\text{spontaneous}} + \underbrace{\sum_{T<t,\ \text{pt de N1}} h_{1\to 1}(t-T)}_{\text{self-interaction}}$$

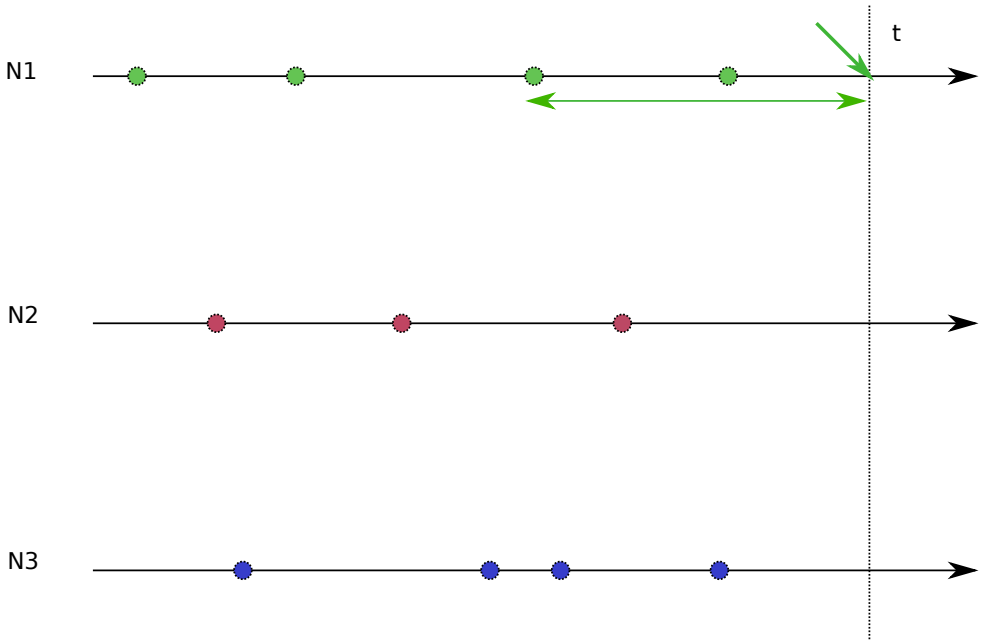
Multivariate Hawkes processes

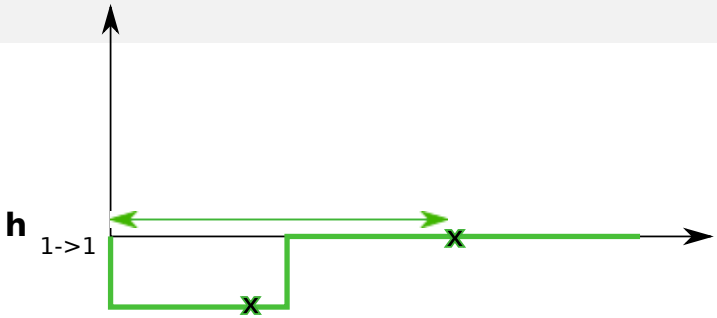




$$\lambda^{(1)}(t) = \underbrace{\nu}_{\text{spontaneous}} + \underbrace{\sum_{\substack{T<t \\ \text{pt de N1}}} h_{1\to 1}(t-T)}_{\text{self-interaction}}$$

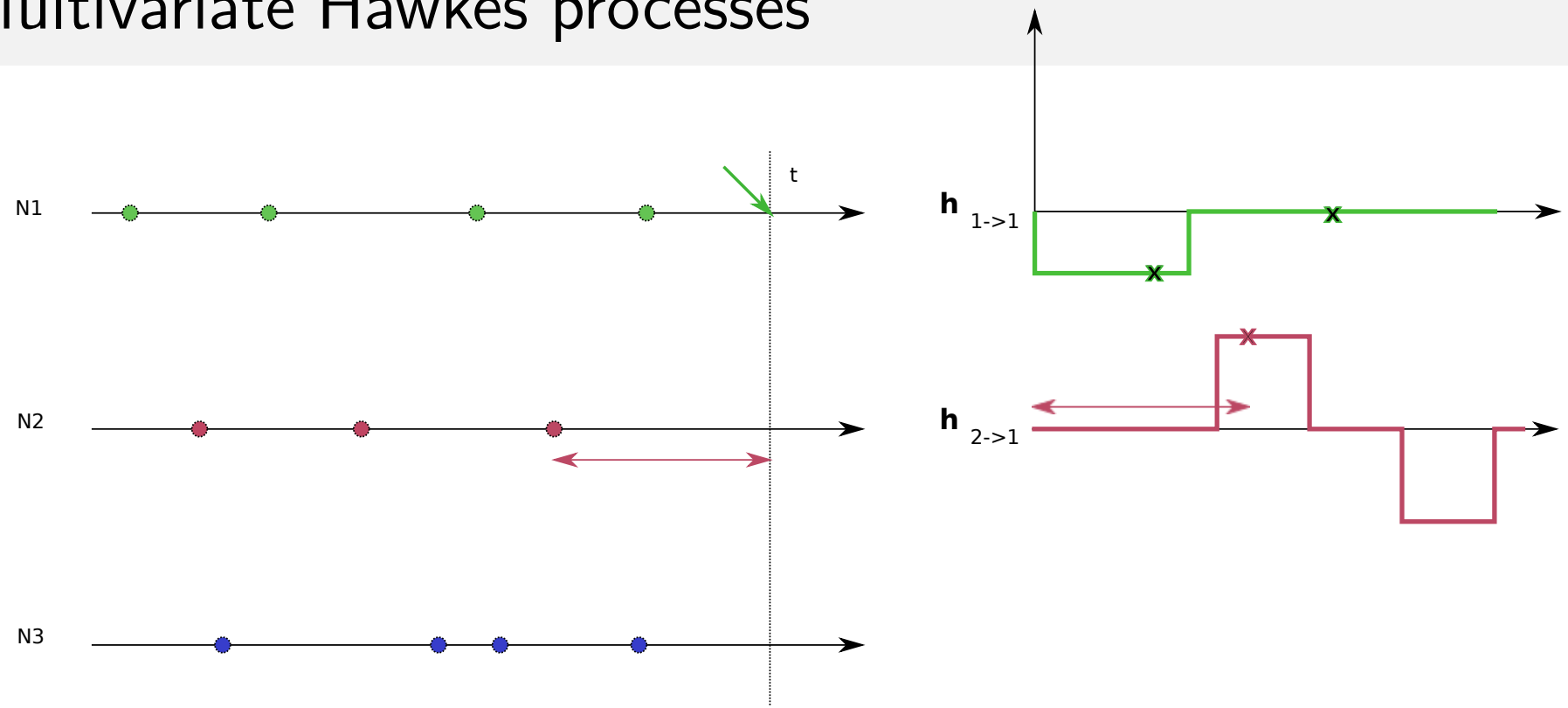
Multivariate Hawkes processes





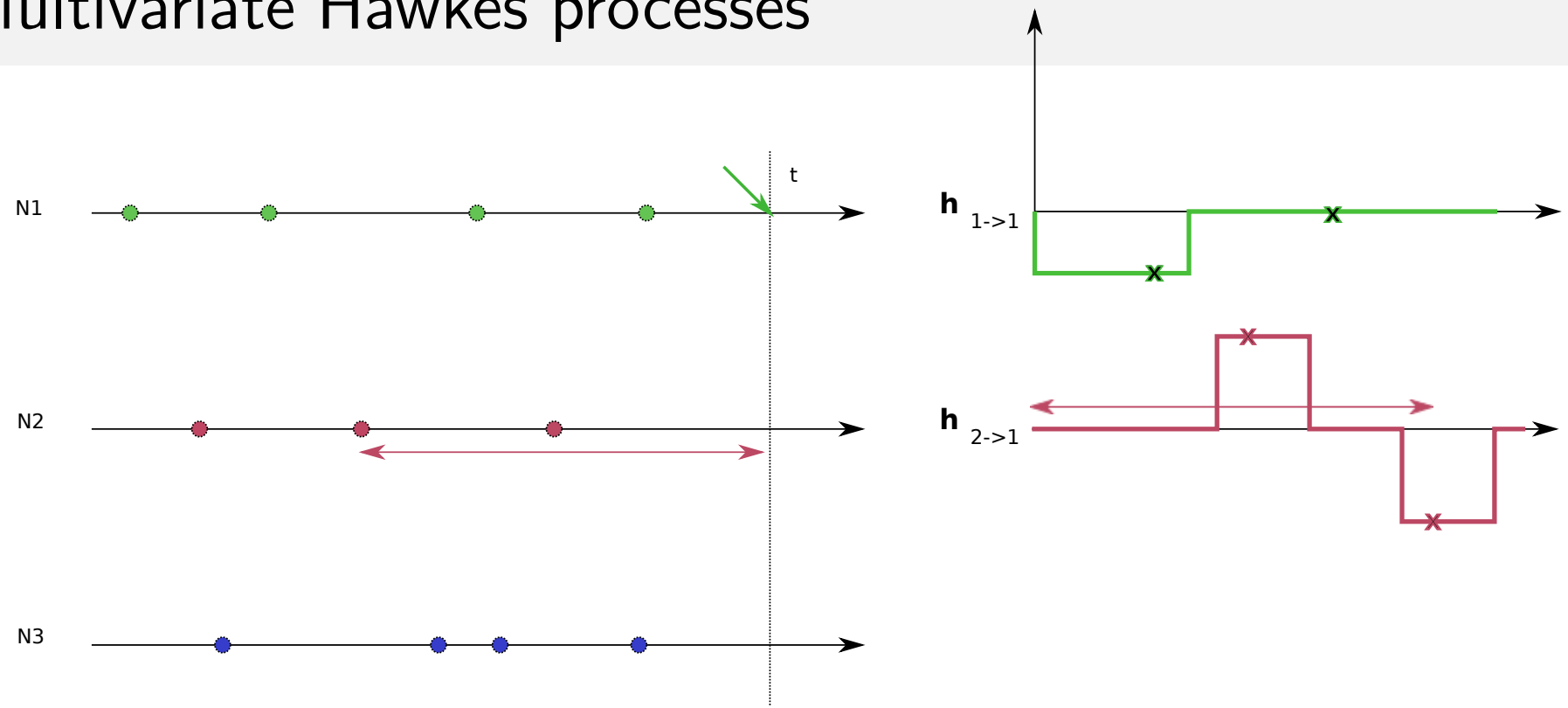
$$\lambda^{(1)}(t) = \underbrace{\nu}_{\text{spontaneous}} + \underbrace{\sum_{\substack{T<t \\ \text{pt de N1}}} h_{1\to 1}(t-T)}_{\text{self-interaction}}$$

Multivariate Hawkes processes



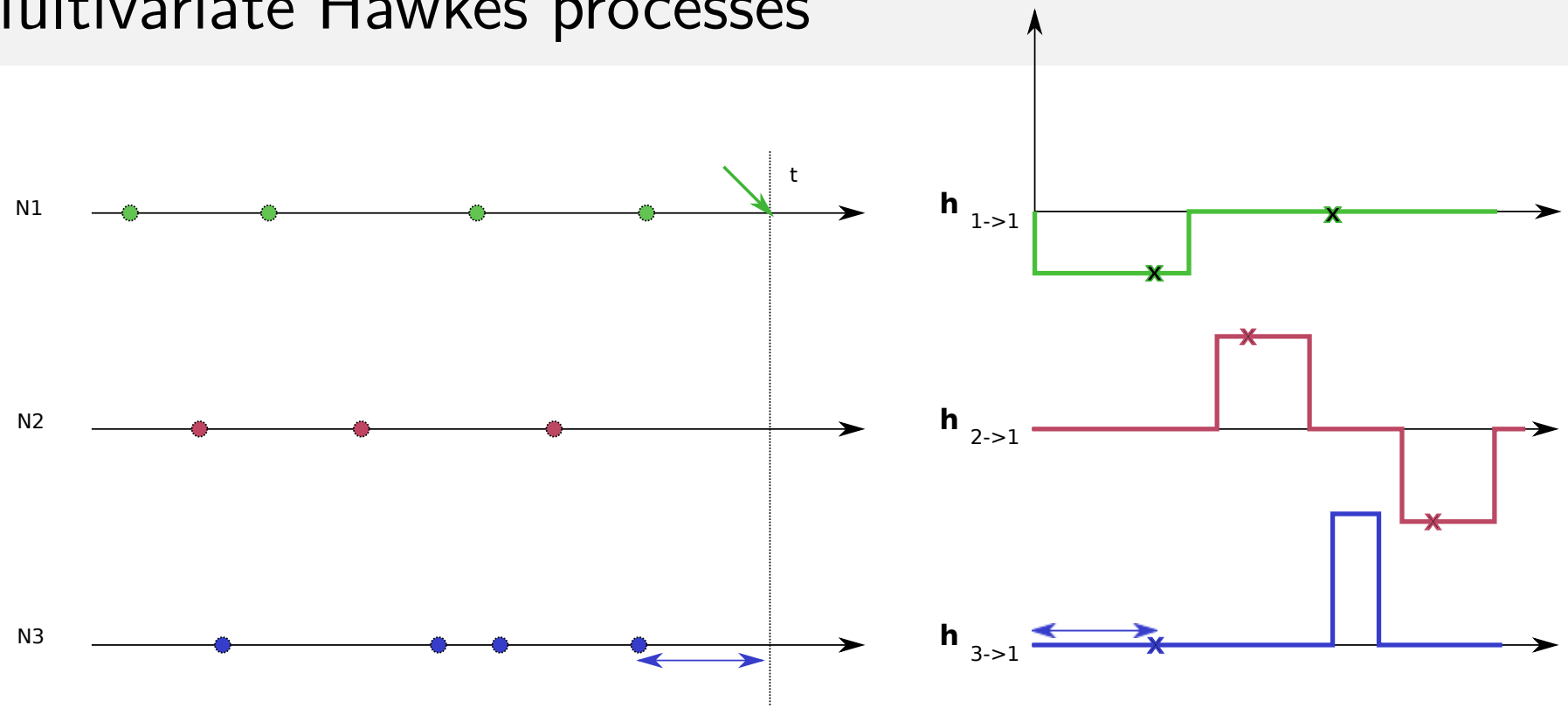
$$\lambda^{(1)}(t) = \underbrace{\nu}_{\text{spontaneous}} + \underbrace{\sum_{\substack{T<t \\ \text{pt de N1}}} h_{1\to 1}(t-T)}_{\text{self-interaction}} + \underbrace{\sum_{\substack{m\neq 1, \\ T<t, \\ \text{pt de Nm}}} h_{m\to 1}(t-T)}_{\text{interaction}}$$

Multivariate Hawkes processes



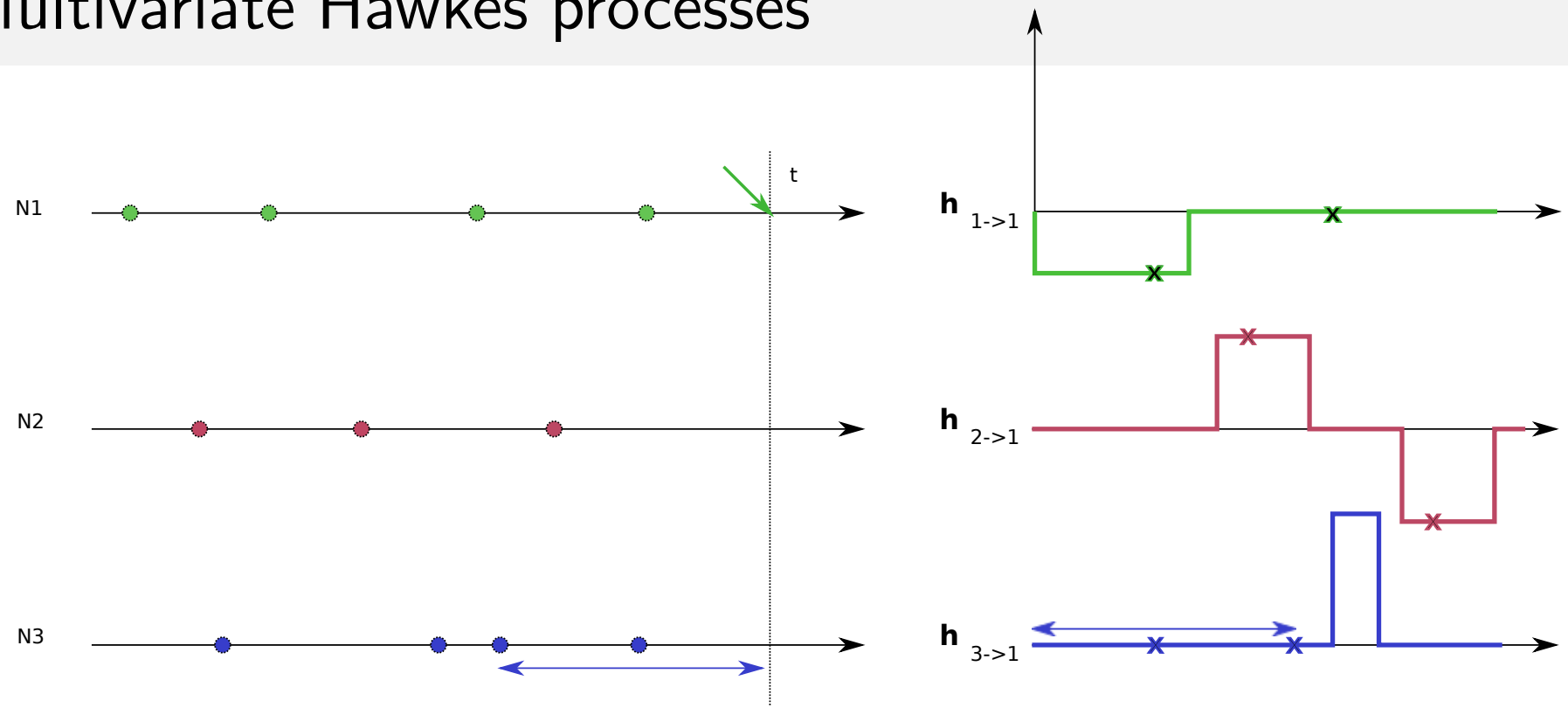
$$\lambda^{(1)}(t) = \underbrace{\nu}_{\text{spontaneous}} + \underbrace{\sum_{\substack{T<t \\ \text{pt de N1}}} h_{1\to 1}(t-T)}_{\text{self-interaction}} + \underbrace{\sum_{\substack{m\neq 1, \\ T<t, \\ \text{pt de Nm}}} h_{m\to 1}(t-T)}_{\text{interaction}}$$

Multivariate Hawkes processes



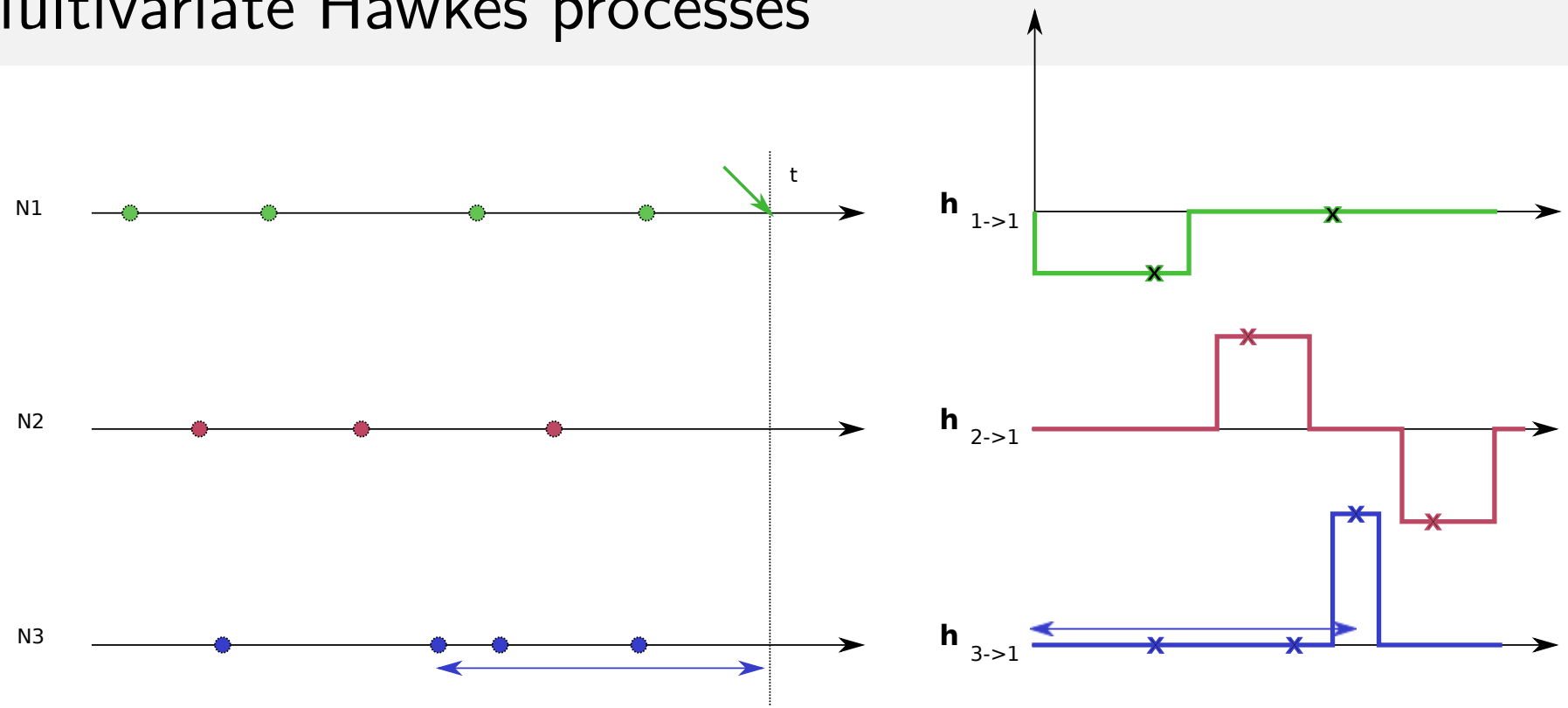
$$\lambda^{(1)}(t) = \underbrace{\nu}_{\text{spontaneous}} + \underbrace{\sum_{\substack{T<t \\ \text{pt de N1}}} h_{1\to 1}(t-T)}_{\text{self-interaction}} + \underbrace{\sum_{\substack{m\neq 1,\\ T<t,\\ \text{pt de Nm}}} h_{m\to 1}(t-T)}_{\text{interaction}}$$

Multivariate Hawkes processes



$$\lambda^{(1)}(t) = \underbrace{\nu}_{\text{spontaneous}} + \underbrace{\sum_{\substack{T<t \\ \text{pt de N1}}} h_{1\to 1}(t-T)}_{\text{self-interaction}} + \underbrace{\sum_{\substack{m\neq 1,\\ T<t,\\ \text{pt de Nm}}} h_{m\to 1}(t-T)}_{\text{interaction}}$$

Multivariate Hawkes processes



$$\lambda^{(1)}(t) = \underbrace{\nu}_{\text{spontaneous}} + \underbrace{\sum_{\substack{T<t \\ \text{pt de N1}}} h_{1\to 1}(t-T)}_{\text{self-interaction}} + \underbrace{\sum_{\substack{m\neq 1,\\ T<t,\\ \text{pt de Nm}}} h_{m\to 1}(t-T)}_{\text{interaction}}$$

More formally

- Only excitation (all the $h_\ell^{(r)}$ are positive) : for all r,

$$\lambda^{(r)}(t) = \nu_r + \sum_{\ell=1}^{M} \int_{-\infty}^{t-} h_\ell^{(r)}(t-u) dN_u^{(\ell)}.$$

Branching / Cluster representation, stationary process if the spectral radius of $\left(\int h_\ell^{(r)}(t) dt \right)$ is < 1.

More formally

- Only excitation (all the $h_\ell^{(r)}$ are positive) : for all r,

$$\lambda^{(r)}(t) = \nu_r + \sum_{\ell=1}^{M} \int_{-\infty}^{t-} h_\ell^{(r)}(t-u) dN_u^{(\ell)}.$$

  Branching / Cluster representation, stationary process if the spectral radius of $\left(\int h_\ell^{(r)}(t)dt \right)$ is < 1.

- Interaction, for instance, (in general any 1-Lipschitz function, Brémaud Massoulié 1996)

$$\lambda^{(r)}(t) = \left(\nu_r + \sum_{\ell=1}^{M} \int_{-\infty}^{t-} h_\ell^{(r)}(t-u) dN_u^{(\ell)} \right)_+ .$$

More formally

- Only excitation (all the $h_\ell^{(r)}$ are positive) : for all r,

$$\lambda^{(r)}(t) = \nu_r + \sum_{\ell=1}^{M} \int_{-\infty}^{t-} h_\ell^{(r)}(t-u) dN_u^{(\ell)}.$$

  Branching / Cluster representation, stationary process if the spectral radius of $\left(\int h_\ell^{(r)}(t)dt \right)$ is < 1.

- Interaction, for instance, (in general any 1-Lipschitz function, Brémaud Massoulié 1996)

$$\lambda^{(r)}(t) = \left(\nu_r + \sum_{\ell=1}^{M} \int_{-\infty}^{t-} h_\ell^{(r)}(t-u) dN_u^{(\ell)} \right)_+ .$$

- Exponential (Multiplicative shape but no guarantee of a stationary version ...) $\lambda^{(r)}(t) = \exp\left(\nu_r + \sum_{\ell=1}^{M} \int_{-\infty}^{t-} h_\ell^{(r)}(t-u) dN_u^{(\ell)} \right)$.

Previous works

- Maximum likelihood estimates eventually (Ogata, Vere-Jones etc mainly for sismology, Chornoboy et al., for neuroscience, Gusto and Schbath for genomics)
- Parametric tests for the detection of edge + Maximum likelihood + exponential formula + spline estimation (Carstensen et al., in genomics)

Previous works

- Maximum likelihood estimates eventually (Ogata, Vere-Jones etc mainly for sismology, Chornoboy et al., for neuroscience, Gusto and Schbath for genomics)
- Parametric tests for the detection of edge + Maximum likelihood + exponential formula + spline estimation (Carstensen et al., in genomics)
- Main problem not enough spikes ! so either over-fitting because d (number of parameters) too large or bad estimation if d too small (see later).
- Also MLE are not that easy to compute (EM algorithm etc)
- $\neq$ least-squares if linear parametric model.

Another parametric method : Least-squares

- Observation on $[0, T]$
- A good estimate of the parameters should correspond to an intensity candidate close to the true one.

Another parametric method : Least-squares

- Observation on $[0, T]$
- A good estimate of the parameters should correspond to an intensity candidate close to the true one.
- Hence one would like to minimize $\|\eta - \lambda\|^2 = \int_0^T [\eta(t) - \lambda(t)]^2 dt$ in η

Another parametric method : Least-squares

- Observation on $[0, T]$
- A good estimate of the parameters should correspond to an intensity candidate close to the true one.
- Hence one would like to minimize $\|\eta - \lambda\|^2 = \int_0^T [\eta(t) - \lambda(t)]^2 dt$ in η
- or equivalently $-2\int_0^T \eta(t)\lambda(t)dt + \int_0^T \eta(t)^2 dt$.

Another parametric method : Least-squares

- Observation on $[0, T]$
- A good estimate of the parameters should correspond to an intensity candidate close to the true one.
- Hence one would like to minimize $\|\eta - \lambda\|^2 = \int_0^T [\eta(t) - \lambda(t)]^2 dt$ in η
- or equivalently $-2\int_0^T \eta(t)\lambda(t)dt + \int_0^T \eta(t)^2 dt$.
- But dN_t randomly fluctuates around $\lambda(t)dt$ and is observable.

Another parametric method : Least-squares

- Observation on $[0, T]$
- A good estimate of the parameters should correspond to an intensity candidate close to the true one.
- Hence one would like to minimize $\|\eta - \lambda\|^2 = \int_0^T [\eta(t) - \lambda(t)]^2 dt$ in η
- or equivalently $-2\int_0^T \eta(t)\lambda(t)dt + \int_0^T \eta(t)^2 dt$.
- But dN_t randomly fluctuates around $\lambda(t)dt$ and is observable.
- Hence minimize

$$\gamma(\eta) = -2\int_0^T \eta(t)dN_t + \int_0^T \eta(t)^2 dt,$$

  for a model $\eta = \lambda_{\mathbf{a}}(t)$,
- If multivariate minimize $\sum_m \gamma_m(\eta^{(m)})$.

Least-square contrast on a toy model

Let $[a, b]$ an interval of $\mathbb{R}^+$.

Least-square contrast on a toy model

Let $[a, b]$ an interval of $\mathbb{R}^+$.

$$\eta(t) = \sum_{T \in N} \alpha \mathbf{1}_{a \leq t-T \leq b} = \alpha N_{[t-b,t-a]}$$

There is only one parameter $\alpha \rightarrow$ minimizing γ

Least-square contrast on a toy model

Let $[a, b]$ an interval of $\mathbb{R}^+$.

$$\eta(t) = \sum_{T \in N} \alpha \mathbf{1}_{a \leq t-T \leq b} = \alpha N_{[t-b,t-a]}$$

There is only one parameter $\alpha \rightarrow$ minimizing γ

$$\gamma(\eta) = -2\alpha \int_0^T N_{[t-b,t-a]} dN_t + \alpha^2 \int_0^T N^2_{[t-b,t-a]} dt$$

Least-square contrast on a toy model

Let $[a, b]$ an interval of $\mathbb{R}^+$.

$$\eta(t) = \sum_{T \in N} \alpha \mathbf{1}_{a \leq t-T \leq b} = \alpha N_{[t-b,t-a]}$$

There is only one parameter $\alpha \rightarrow$ minimizing γ

$$\hat{\alpha} = \frac{\int_0^T N_{[t-b,t-a]} dN_t}{\int_0^T N_{[t-b,t-a]}^2 dt}$$

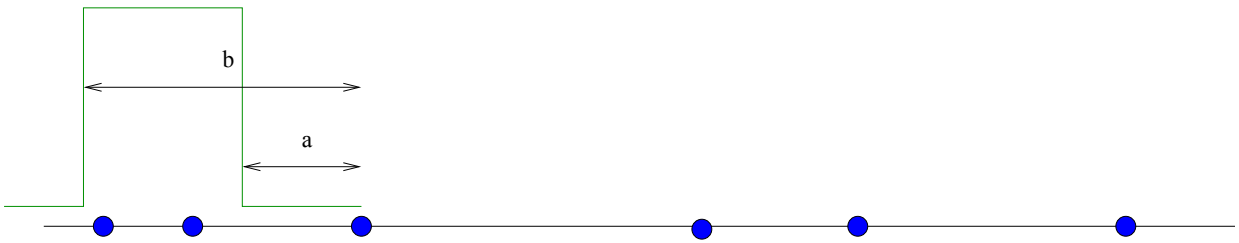
Least-square contrast on a toy model

Let $[a, b]$ an interval of $\mathbb{R}^+$.

$$\eta(t) = \sum_{T \in N} \alpha \mathbf{1}_{a \leq t-T \leq b} = \alpha N_{[t-b,t-a]}$$

There is only one parameter $\alpha \rightarrow$ minimizing γ

$$\hat{\alpha} = \frac{\int_0^T N_{[t-b,t-a]} dN_t}{\int_0^T N^2_{[t-b,t-a]} dt}$$



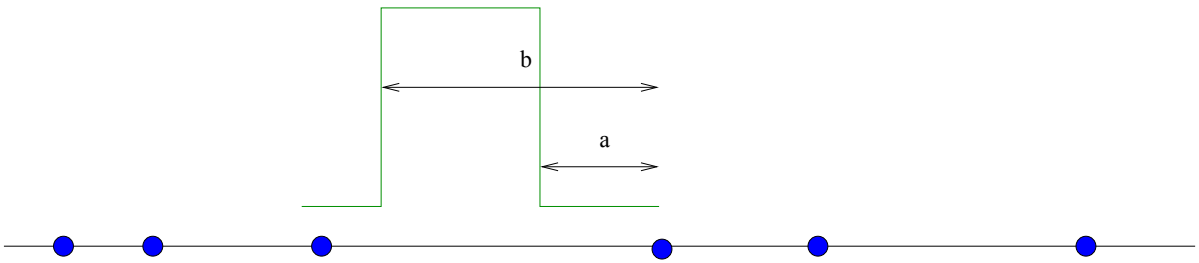
Least-square contrast on a toy model

Let $[a, b]$ an interval of $\mathbb{R}^+$.

$$\eta(t) = \sum_{T \in N} \alpha \mathbf{1}_{a \leq t-T \leq b} = \alpha N_{[t-b,t-a]}$$

There is only one parameter $\alpha \rightarrow$ minimizing γ

$$\hat{\alpha} = \frac{\int_0^T N_{[t-b,t-a]} dN_t}{\int_0^T N_{[t-b,t-a]}^2 dt}$$



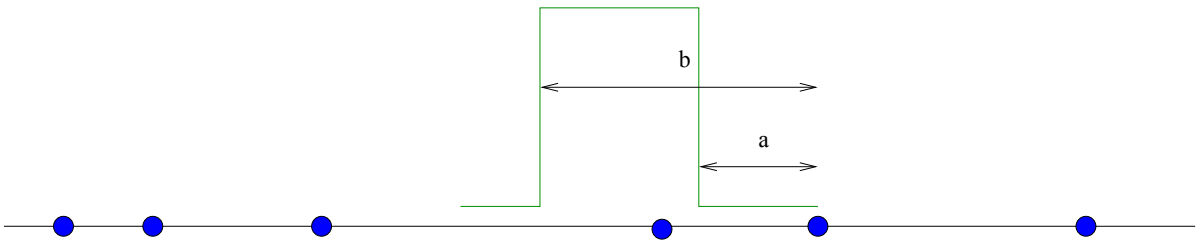
Least-square contrast on a toy model

Let $[a, b]$ an interval of $\mathbb{R}^+$.

$$\eta(t) = \sum_{T \in N} \alpha \mathbf{1}_{a \leq t-T \leq b} = \alpha N_{[t-b,t-a]}$$

There is only one parameter $\alpha \rightarrow$ minimizing γ

$$\hat{\alpha} = \frac{\int_0^T N_{[t-b,t-a]} dN_t}{\int_0^T N^2_{[t-b,t-a]} dt}$$



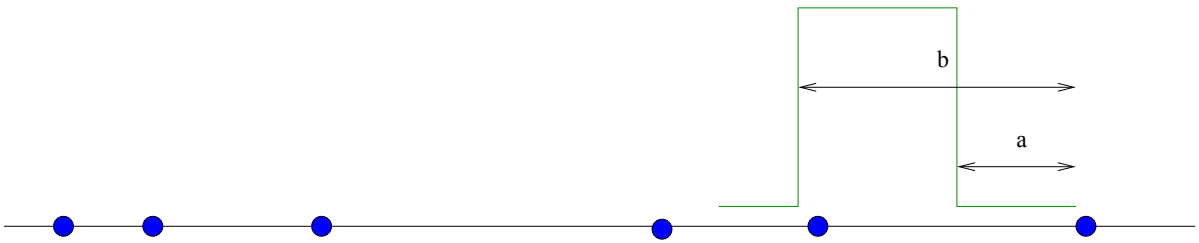
Least-square contrast on a toy model

Let $[a, b]$ an interval of $\mathbb{R}^+$.

$$\eta(t) = \sum_{T \in N} \alpha \mathbf{1}_{a \leq t-T \leq b} = \alpha N_{[t-b,t-a]}$$

There is only one parameter $\alpha \rightarrow$ minimizing γ

$$\hat{\alpha} = \frac{\int_0^T N_{[t-b,t-a]} dN_t}{\int_0^T N_{[t-b,t-a]}^2 dt}$$



Least-square contrast on a toy model

Let $[a, b]$ an interval of $\mathbb{R}^+$.

$$\eta(t) = \sum_{T \in N} \alpha \mathbf{1}_{a \leq t-T \leq b} = \alpha N_{[t-b,t-a]}$$

There is only one parameter $\alpha \rightarrow$ minimizing γ

$$\hat{\alpha} = \frac{\int_0^T N_{[t-b,t-a]} dN_t}{\int_0^T N^2_{[t-b,t-a]} dt}$$

The numerator is the number of pairs of points with delay in $[a, b]$

Full multivariate processes and linear parametrisation

$$\lambda^{(r)}(t) \stackrel{?}{=} \nu^{(r)} + \sum_{\ell=1}^{M} \sum_{T<t, T \text{ in } N^{(\ell)}} h_{\ell}^{(r)}(t-T).$$

Full multivariate processes and linear parametrisation

$$\lambda^{(r)}(t) \stackrel{?}{=} \nu^{(r)} + \sum_{\ell=1}^{M} \sum_{T<t, T \text{ in } N^{(\ell)}} h_\ell^{(r)}(t-T).$$

+ Piecewise constant model with parameter $\mathbf{a}$

Full multivariate processes and linear parametrisation

$$\lambda^{(r)}(t) \stackrel{?}{=} \nu^{(r)} + \sum_{\ell=1}^{M} \sum_{T<t, T \text{ in } N^{(\ell)}} h_\ell^{(r)}(t-T).$$

$+$ Piecewise constant model with parameter $\mathbf{a}$

By linearity,

$$\lambda^{(r)}(t) \stackrel{?}{=} (\mathbf{R}\mathbf{c}_t)'\mathbf{a},$$

Full multivariate processes and linear parametrisation

$$\lambda^{(r)}(t) \stackrel{?}{=} \nu^{(r)} + \sum_{\ell=1}^{M} \sum_{T<t, T \text{ in } N^{(\ell)}} h_\ell^{(r)}(t-T).$$

$+$ Piecewise constant model with parameter $\mathbf{a}$

By linearity,

$$\lambda^{(r)}(t) \stackrel{?}{=} (\mathbf{Rc}_t)'\mathbf{a},$$

with $\mathbf{Rc}_t$ being the renormalized instantaneous count given by

$$(\mathbf{Rc}_t)' = \left(1, \delta^{-1/2}(\mathbf{c}_t^{(1)})', \ldots, \delta^{-1/2}(\mathbf{c}_t^{(M)})'\right),$$

Full multivariate processes and linear parametrisation

$$\lambda^{(r)}(t) \stackrel{?}{=} \nu^{(r)} + \sum_{\ell=1}^{M} \sum_{T<t, T \text{ in } N^{(\ell)}} h_\ell^{(r)}(t-T).$$

$+$ Piecewise constant model with parameter $\mathbf{a}$

By linearity,

$$\lambda^{(r)}(t) \stackrel{?}{=} (\mathbf{Rc}_t)'\mathbf{a},$$

with $\mathbf{Rc}_t$ being the renormalized instantaneous count given by

$$(\mathbf{Rc}_t)' = \left(1, \delta^{-1/2}(\mathbf{c}_t^{(1)})', \ldots, \delta^{-1/2}(\mathbf{c}_t^{(M)})'\right),$$

and with $\mathbf{c}_t^{(\ell)}$ being the vector of instantaneous count with delay of N_ℓ i.e.

$$(\mathbf{c}_t^{(\ell)})' = \left(N^{(\ell)}_{[t-\delta,t)}, \ldots, N^{(\ell)}_{[t-K\delta, t-(K-1)\delta)}\right).$$

An heuristic for the least-square estimator

Informally, the link between the point process and its intensity can be written as

$$dN^{(r)}(t) \simeq (\mathbf{R}\mathbf{c}_t)'\mathbf{a}_*^{(r)}dt + \text{ noise.}$$

An heuristic for the least-square estimator

Informally, the link between the point process and its intensity can be written as

$$dN^{(r)}(t) \simeq (\mathbf{Rc}_t)' \mathbf{a}_*^{(r)} dt + \text{ noise.}$$

Let

$$\mathbf{G} = \int_0^T \mathbf{Rc}_t (\mathbf{Rc}_t)' dt,$$

the integrated covariation of the renormalized instantaneous count.

An heuristic for the least-square estimator

Informally, the link between the point process and its intensity can be written as

$$dN^{(r)}(t) \simeq (\mathbf{Rc}_t)' \mathbf{a}_*^{(r)} dt + \text{ noise.}$$

Let

$$\mathbf{G} = \int_0^T \mathbf{Rc}_t (\mathbf{Rc}_t)' dt,$$

the integrated covariation of the renormalized instantaneous count.

$$\mathbf{b}^{(r)} := \int_0^T \mathbf{Rc}_t dN^{(r)}(t) \simeq \mathbf{G}\mathbf{a}_*^{(r)} + \text{ noise.}$$

An heuristic for the least-square estimator

Informally, the link between the point process and its intensity can be written as

$$dN^{(r)}(t) \simeq (\mathbf{Rc}_t)' \mathbf{a}_*^{(r)} dt + \text{ noise.}$$

Let

$$\mathbf{G} = \int_0^T \mathbf{Rc}_t (\mathbf{Rc}_t)' dt,$$

the integrated covariation of the renormalized instantaneous count.

$$\mathbf{b}^{(r)} := \int_0^T \mathbf{Rc}_t dN^{(r)}(t) \simeq \mathbf{Ga}_*^{(r)} + \text{ noise.}$$

where in b lies again the number of couples with a certain delay (cross-correlogram).

Least-square estimate

$$\mathbf{G} = \int_0^T \mathbf{Rc}_t (\mathbf{Rc}_t)' dt,$$

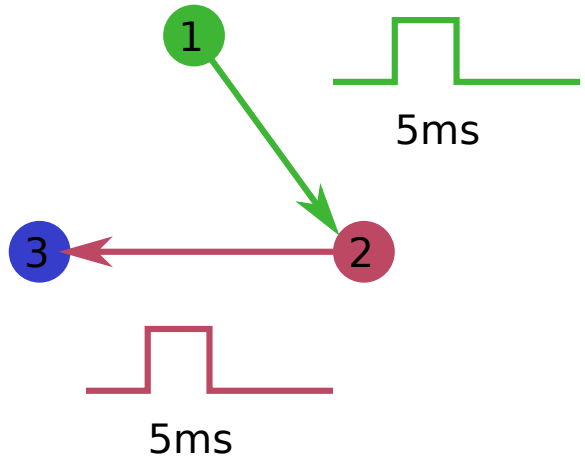
$$\mathbf{b}^{(r)} := \int_0^T \mathbf{Rc}_t dN^{(r)}(t)$$

Least-square estimate

$$\hat{\mathbf{a}}^{(r)} = \mathbf{G}^{-1}\mathbf{b}^{(r)},$$

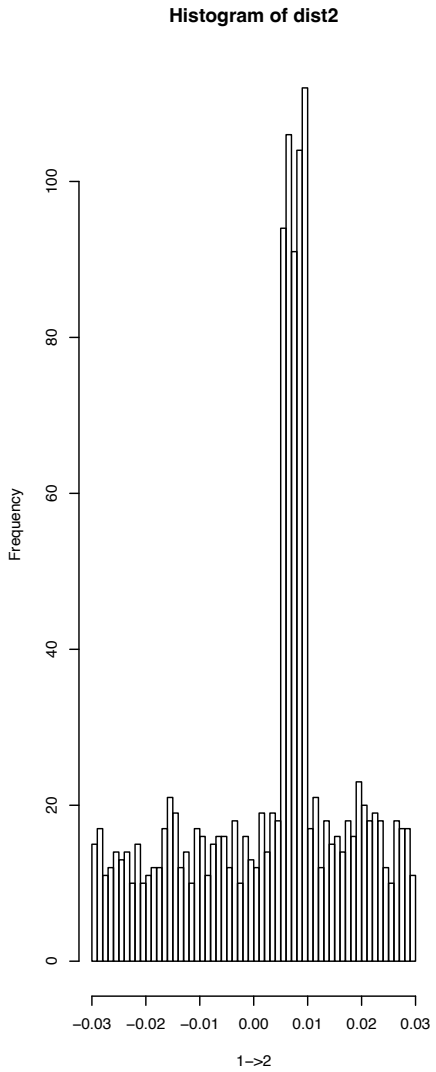
$\rightarrow$ simpler formula than Maximum Likelihood Estimators for similar properties (except efficiency).

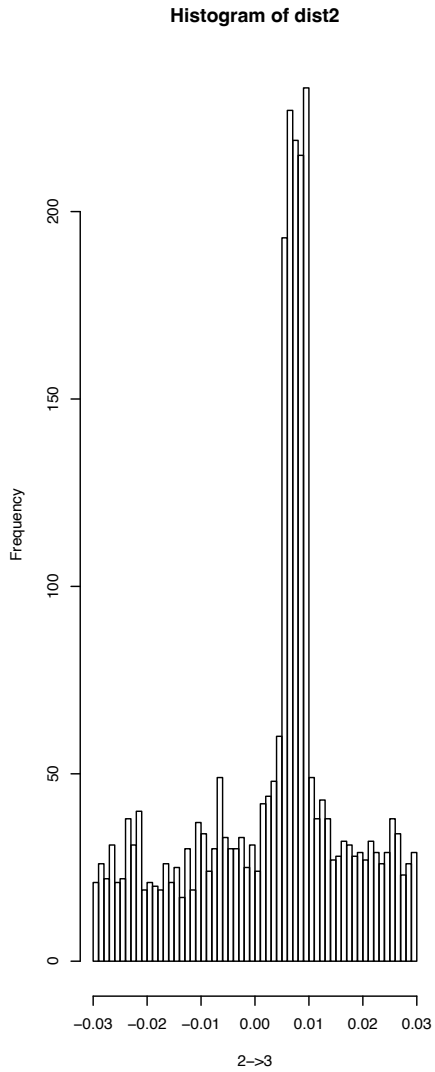
What gain wrt cross correlogramm ?

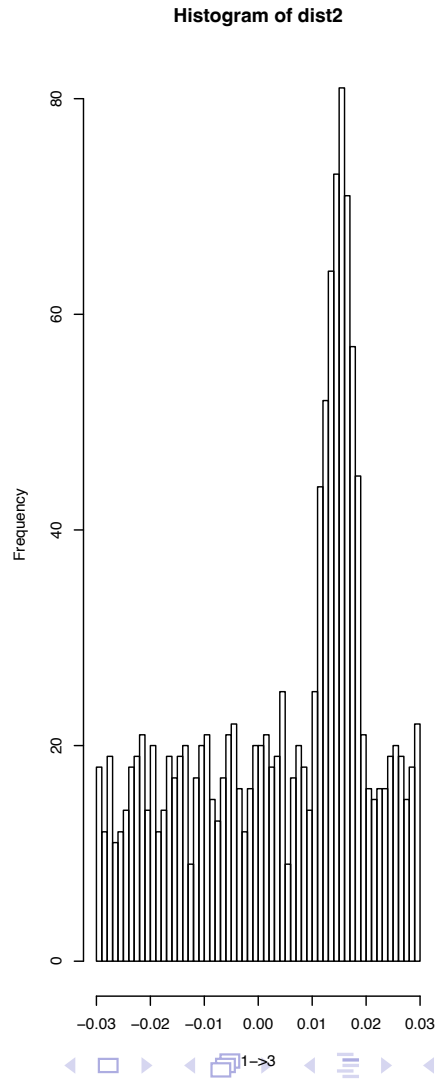


only 2 non zero interaction functions over 9

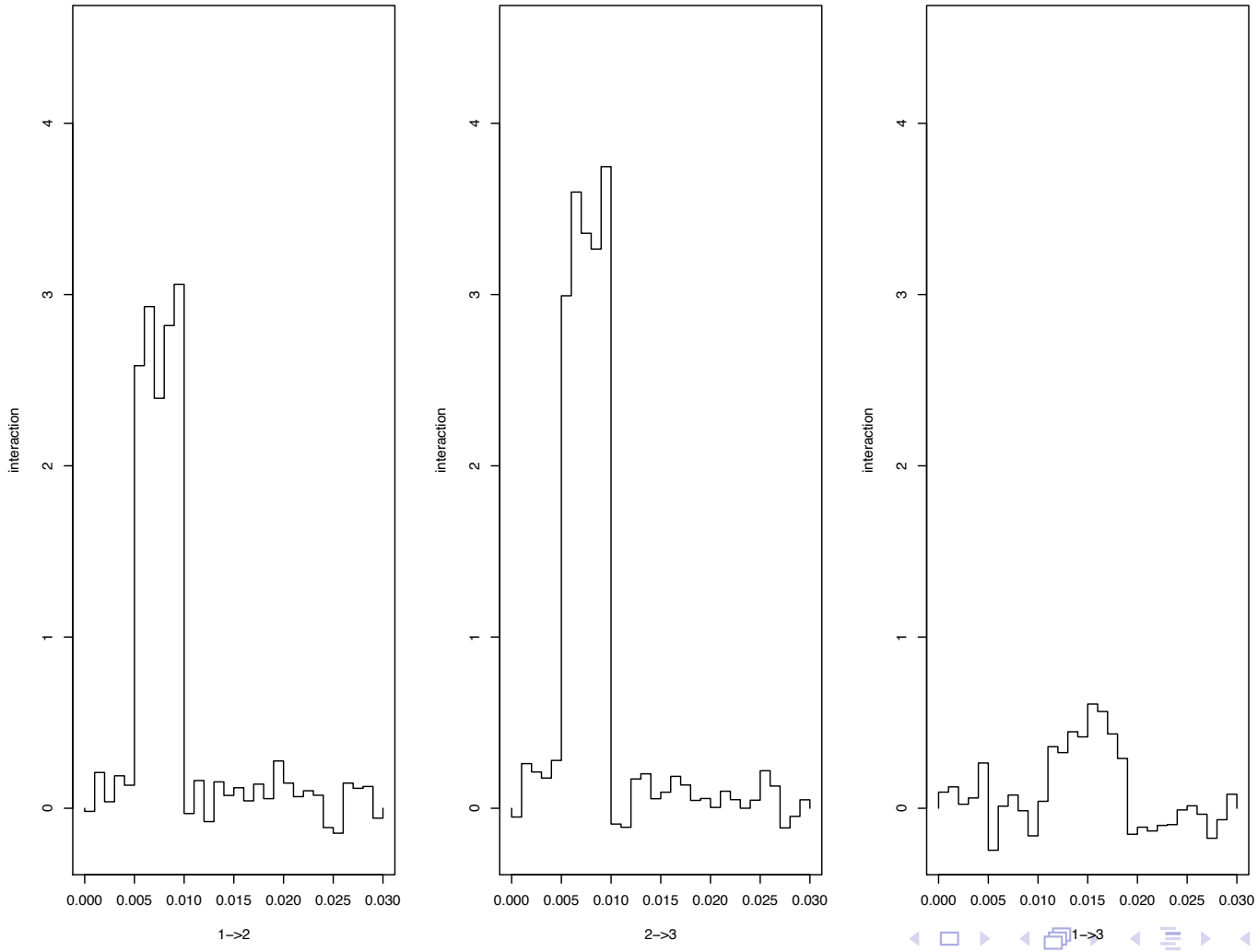
What gain wrt cross correlogramm ?







What gain wrt cross correlogramm ?



Adaptive statistics

- If no model known, one usually wants to consider the largest possible model (piecewise constant with hundreds of parameters etc)
- If use MLE or OLS, each parameter estimate has a variance $\simeq 1/T$

Adaptive statistics

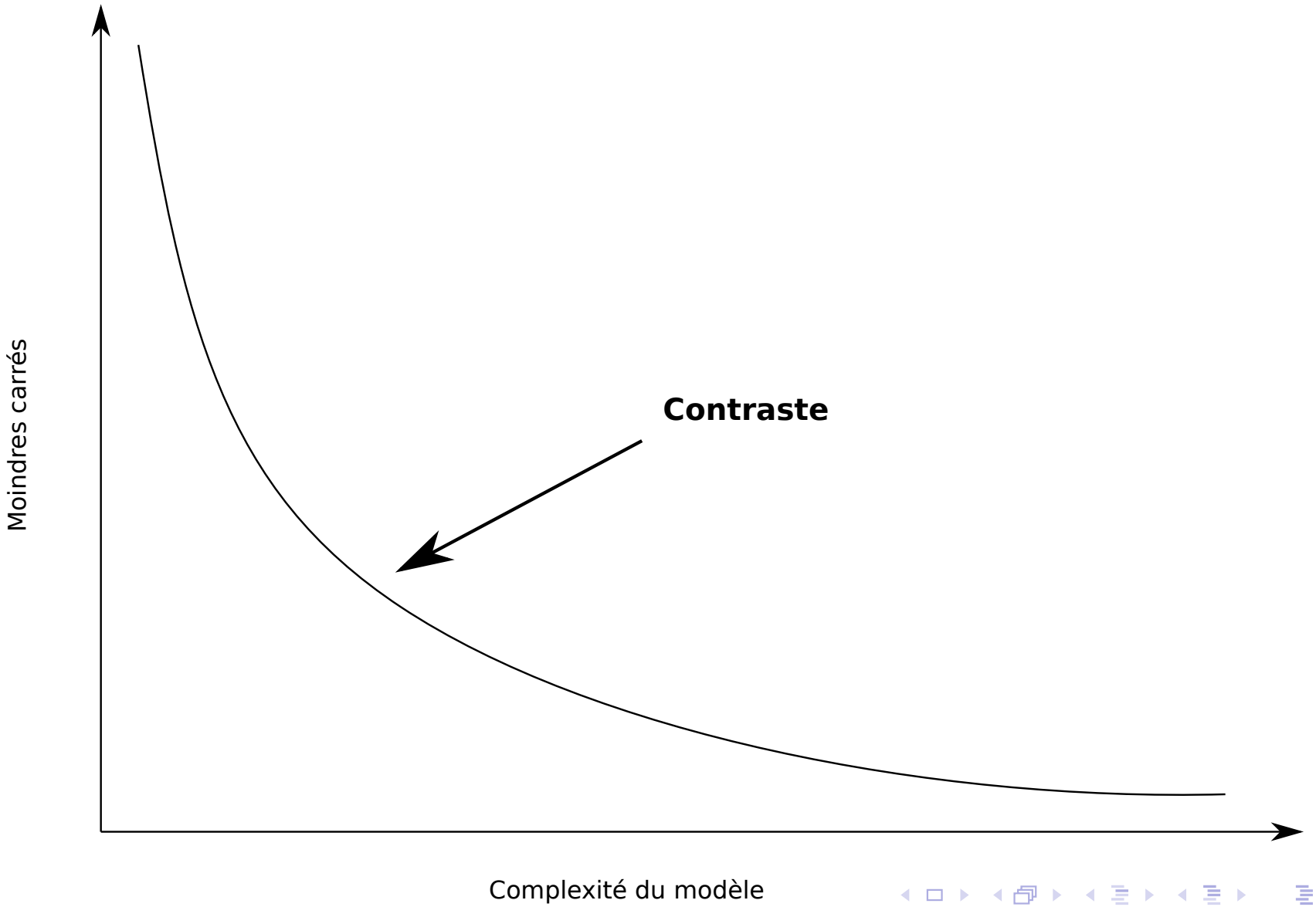
- If no model known, one usually wants to consider the largest possible model (piecewise constant with hundreds of parameters etc)
- If use MLE or OLS, each parameter estimate has a variance $\simeq 1/T$
- If there are d parameters, global variance $\simeq d/T$

Adaptive statistics

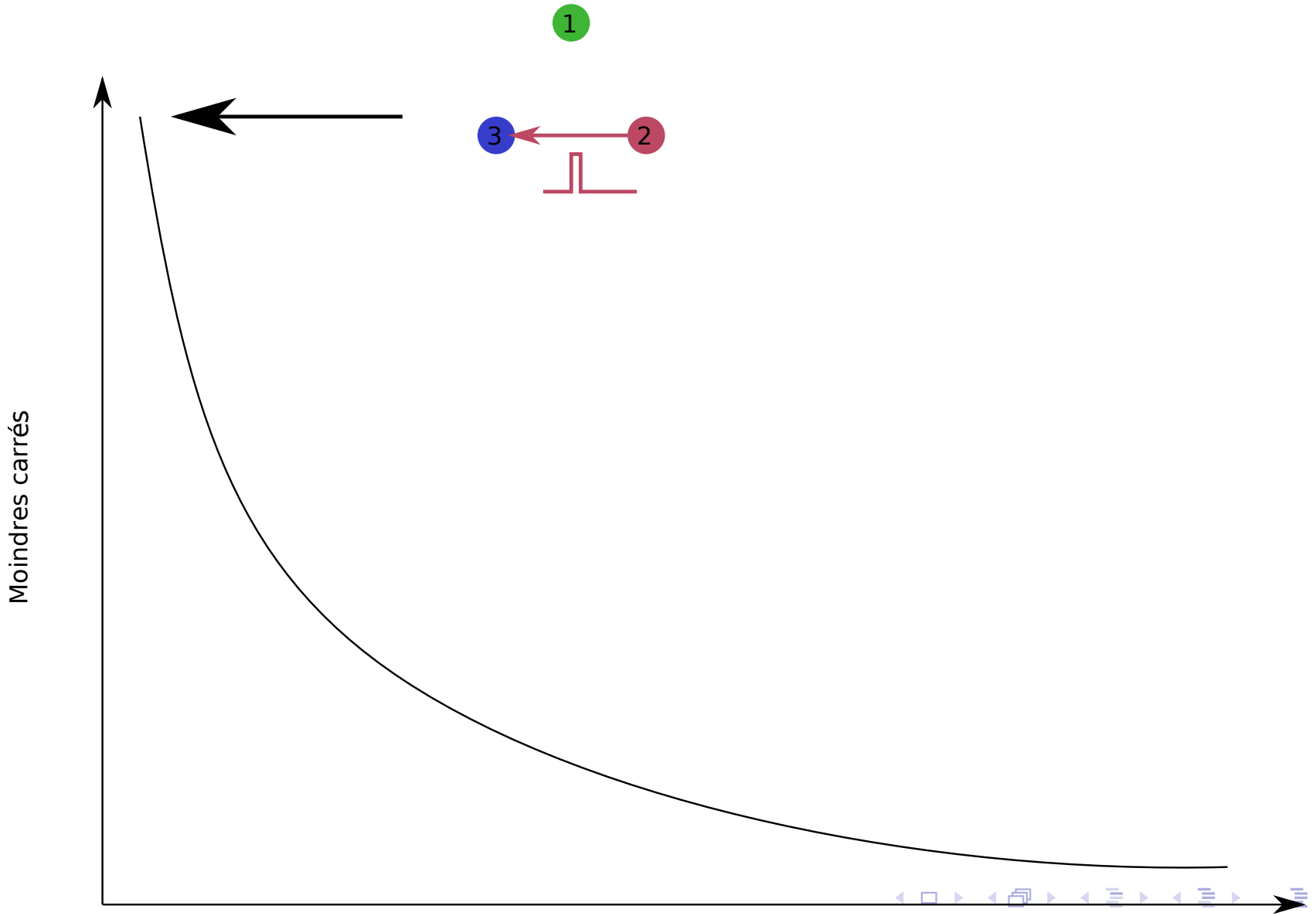
- If no model known, one usually wants to consider the largest possible model (piecewise constant with hundreds of parameters etc)
- If use MLE or OLS, each parameter estimate has a variance $\simeq 1/T$
- If there are d parameters, global variance $\simeq d/T$

Over-fitting ! ! ! !

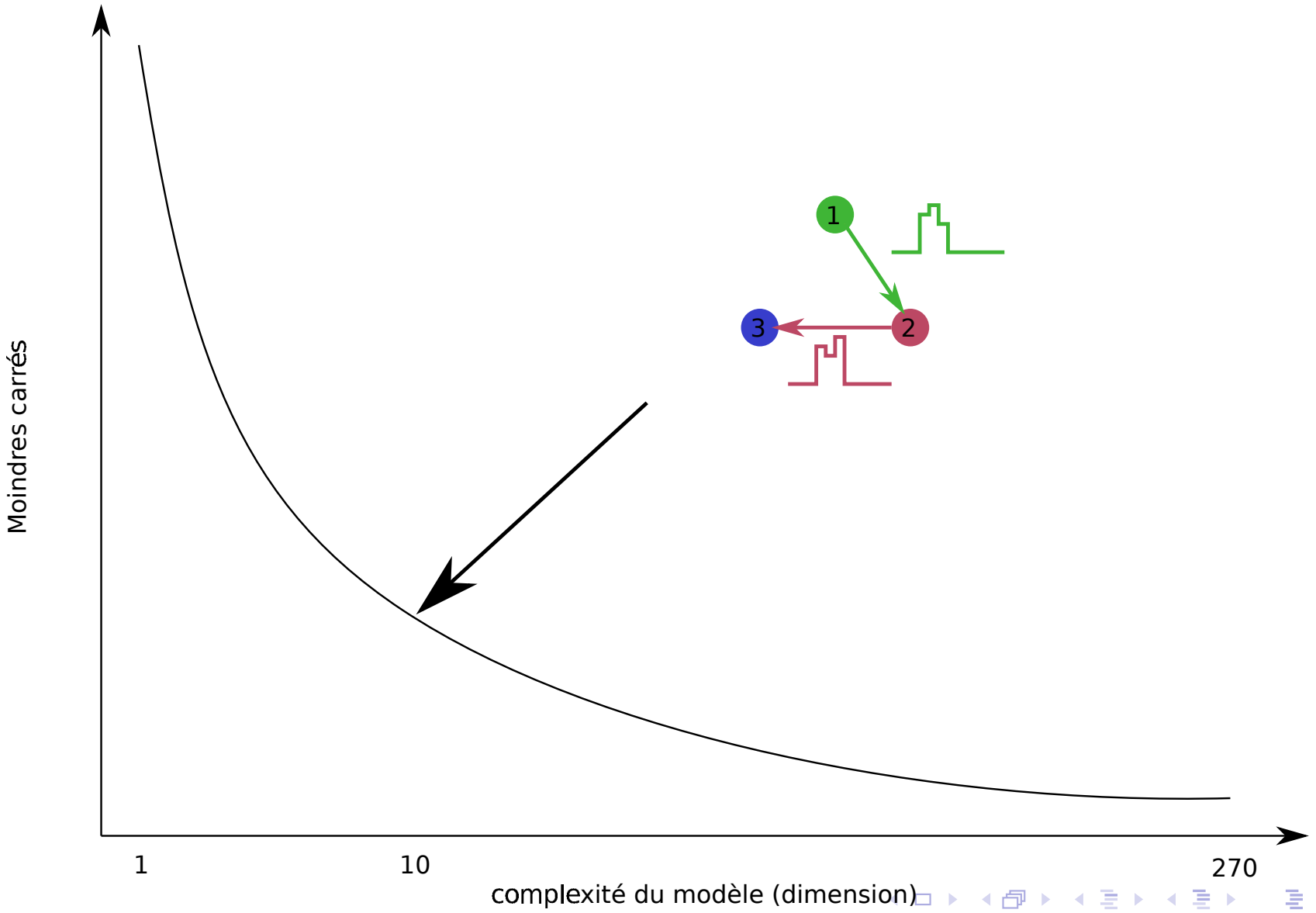
Penalty and parameter choices



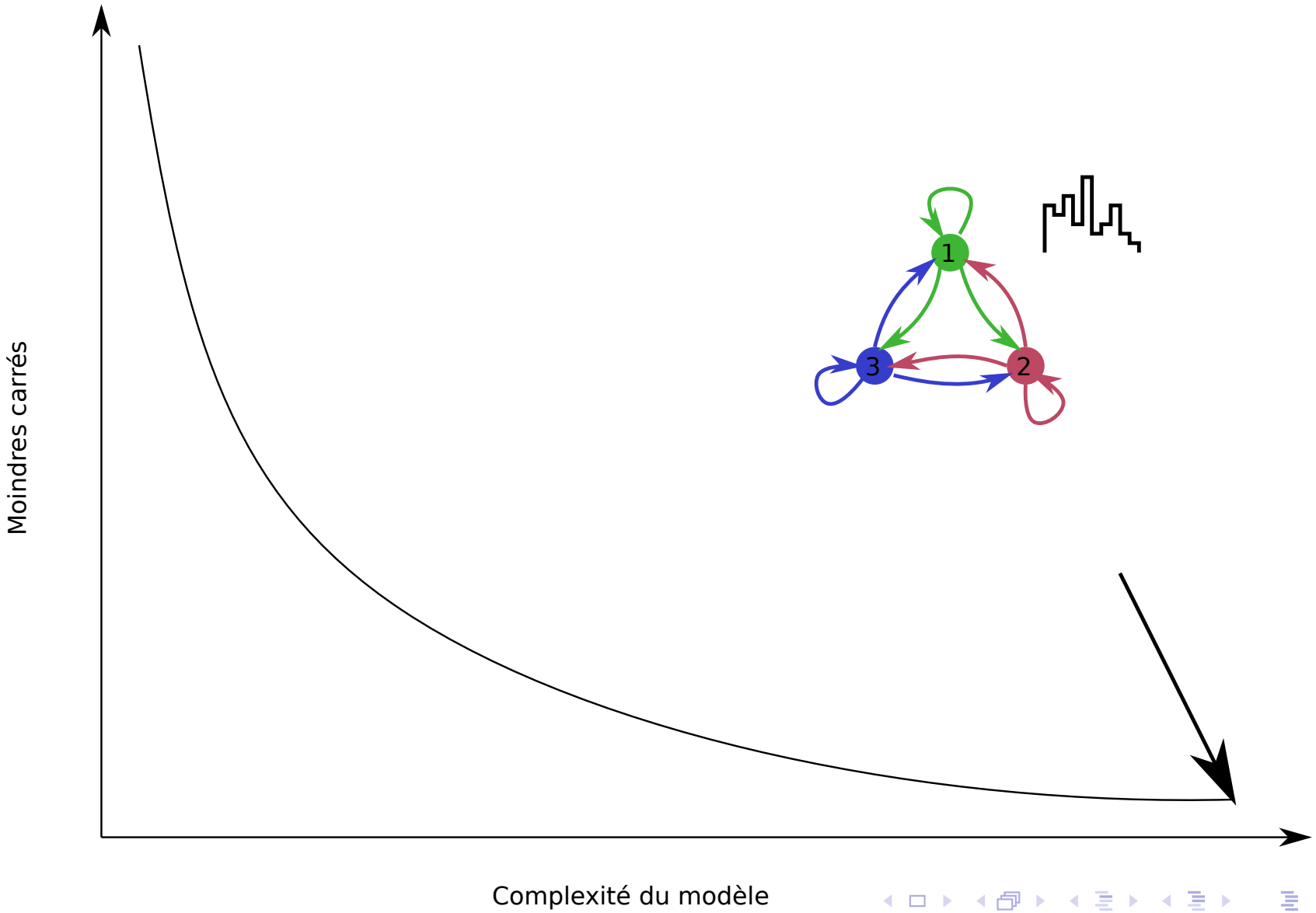
Penalty and parameter choices



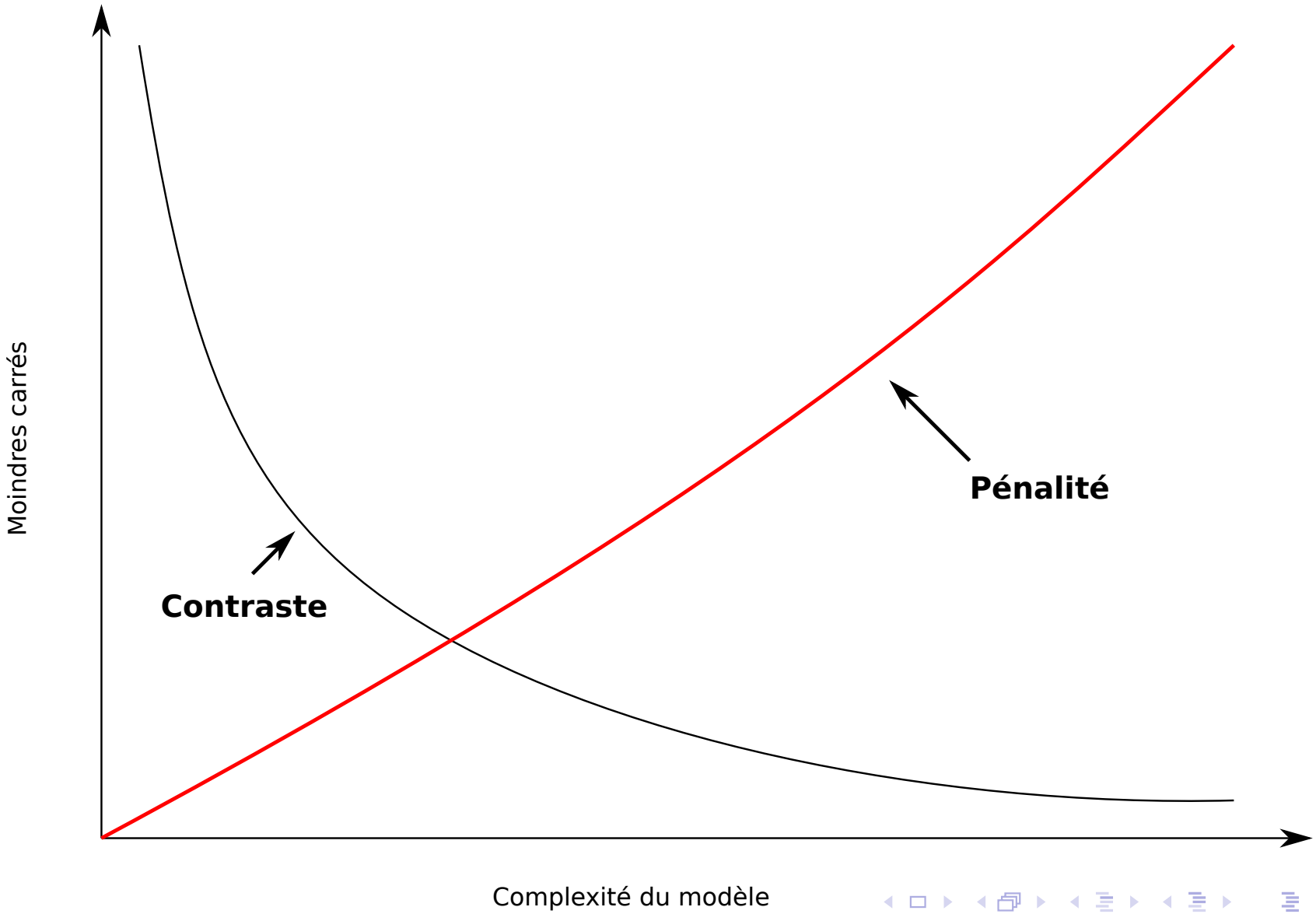
Penalty and parameter choices



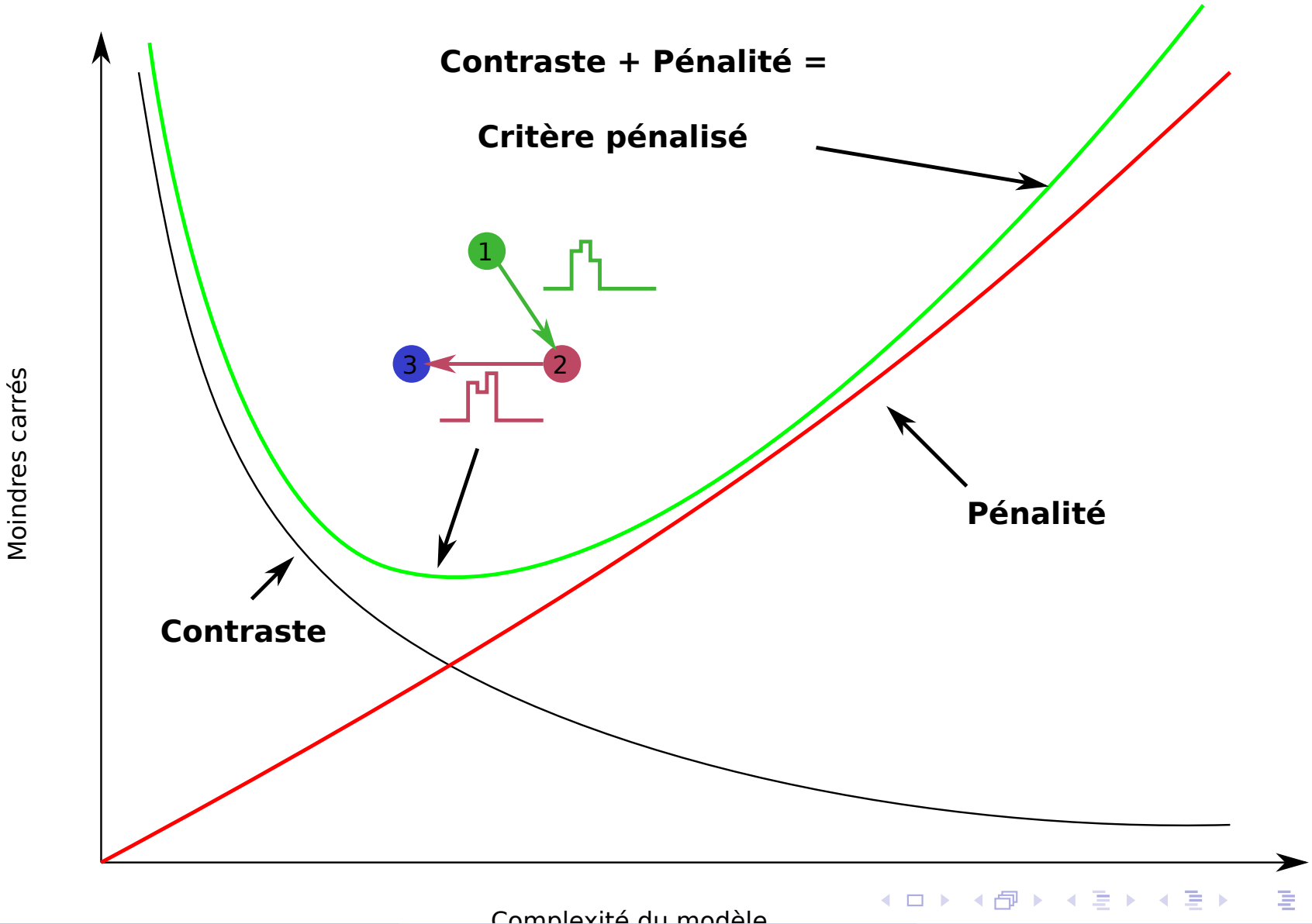
Penalty and parameter choices



Penalty and parameter choices



Penalty and parameter choices



Previous works

- log-likelihood penalized by AIC (Akaike Criterion) (Ogata, Vere-Jones, Gusto and Schbath) : choose the model with d parameters such that

$$\ell(\lambda_{\hat{\theta}_d}) + d$$

Generally works well if few models and largest dimension fixed whereas $T \to \infty$

Previous works

- log-likelihood penalized by AIC (Akaike Criterion) (Ogata, Vere-Jones, Gusto and Schbath) : choose the model with d parameters such that

$$\ell(\lambda_{\hat{\theta}_d}) + d$$

  Generally works well if few models and largest dimension fixed whereas $T \to \infty$
- least-squares+ ℓ_0 penalty, (RB and Schbath)

$$\gamma(\lambda_{\hat{\theta}_d}) + \hat{c}d$$

  with $\hat{c}$ data-driven. Proof that if $\hat{c} > \kappa$, it works even if d grows with T (moderately, $\log(T)$ with d)

Previous works

- log-likelihood penalized by AIC (Akaike Criterion) (Ogata, Vere-Jones, Gusto and Schbath) : choose the model with d parameters such that

$$\ell(\lambda_{\hat{\theta}_d}) + d$$

  Generally works well if few models and largest dimension fixed whereas $T \to \infty$
- least-squares+ ℓ_0 penalty, (RB and Schbath)

$$\gamma(\lambda_{\hat{\theta}_d}) + \hat{c}d$$

  with $\hat{c}$ data-driven. Proof that if $\hat{c} > \kappa$, it works even if d grows with T (moderately, $\log(T)$ with d)
- $\to$ possible mathematically to search for the "zeros"

Previous works

- log-likelihood penalized by AIC (Akaike Criterion) (Ogata, Vere-Jones, Gusto and Schbath) : choose the model with d parameters such that

$$\ell(\lambda_{\hat{\theta}_d}) + d$$

  Generally works well if few models and largest dimension fixed whereas $T \to \infty$
- least-squares+ ℓ_0 penalty, (RB and Schbath)

$$\gamma(\lambda_{\hat{\theta}_d}) + \hat{c}d$$

  with $\hat{c}$ data-driven. Proof that if $\hat{c} > \kappa$, it works even if d grows with T (moderately, $\log(T)$ with d)
- $\to$ possible mathematically to search for the "zeros"
- Problem only for univariate because computational time and memory size awfully large !

Previous works

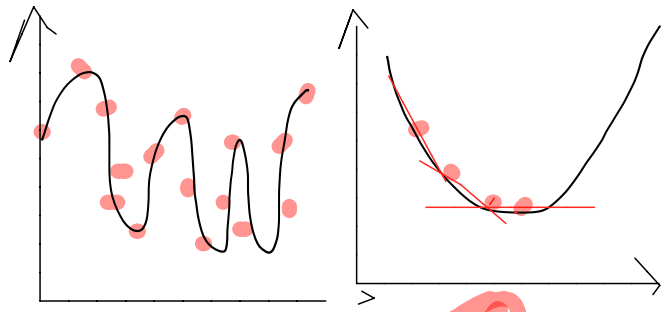
- log-likelihood penalized by AIC (Akaike Criterion) (Ogata, Vere-Jones, Gusto and Schbath) : choose the model with d parameters such that

$$\ell(\lambda_{\hat{\theta}_d}) + d$$

Generally works well if few models and largest dimension fixed whereas $T \to \infty$

- least-squares+ ℓ_0 penalty, (RB and Schbath)

$$\gamma(\lambda_{\hat{\theta}_d}) + \hat{c}d$$



with $\hat{c}$ data-driven. Proof that if $\hat{c} > \kappa$, it works even if d grows with T (moderately, $\log(T)$ with d)

- $\to$ possible mathematically to search for the "zeros"
- Problem only for univariate because computational time and memory size awfully large !
- $\to$ convex criterion

Other works

- Maximum likelihood + exponential formula + ℓ_1 "group Lasso" penalty (Pillow et al. in neuroscience) but no mathematical proof
- Thresholding + tests for very particular bivariate models, oracle inequality (Sansonnet)

ℓ_1 penalty

with N.R. Hansen (Copenhagen), and V. Rivoirard (Dauphine) (2012)

The Lasso criterion can be expressed independently for each sub-process by :

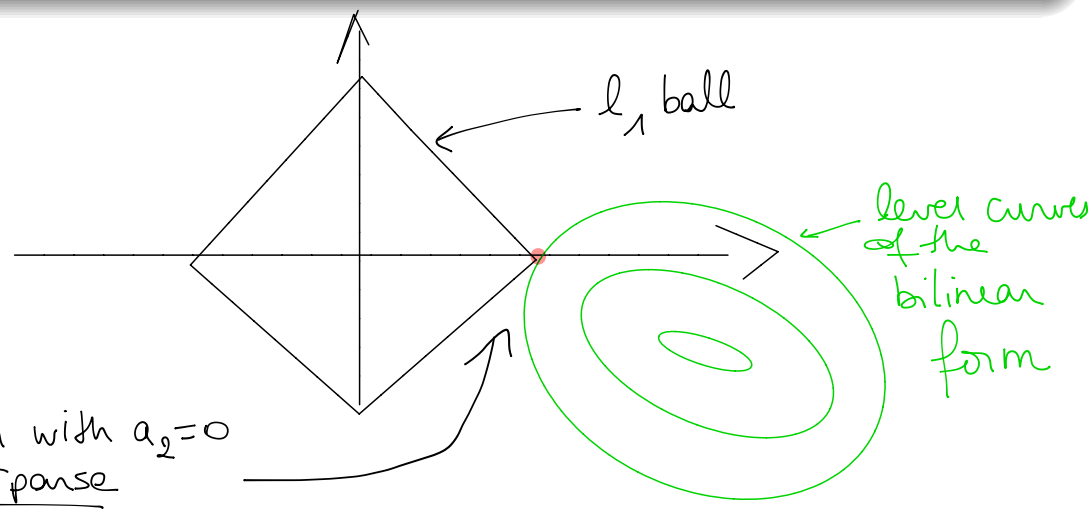
introduced by Tibshirani (1996) = Least-Absolute Shrinkage and selection Operator

Lasso criterion

$$\begin{aligned}\hat{\mathbf{a}}^{(r)} &= \text{argmin}_{\mathbf{a}}\{\gamma_T(\lambda_{\mathbf{a}}^{(r)}) + \text{pen}(\mathbf{a})\} \\ &= \text{argmin}_{\mathbf{a}}\{-2\mathbf{a}'\mathbf{b}_r + \mathbf{a}'\mathbf{G}\mathbf{a} + 2(\mathbf{d}^{(r)})'|\mathbf{a}|\}\end{aligned}$$

with $\mathbf{d}'|\mathbf{a}| = \sum_k d_k|a_k|$

Minimizing Lasso
$\Leftrightarrow$ minimizing bilinear form under constraint on l_1 norm.



ℓ_1 penalty

with N.R. Hansen (Copenhagen), and V. Rivoirard (Dauphine) (2012)

The Lasso criterion can be expressed independently for each sub-process by :

Lasso criterion

$$\begin{aligned}\hat{\mathbf{a}}^{(r)} &= \operatorname{argmin}_{\mathbf{a}}\{\gamma_T(\lambda_{\mathbf{a}}^{(r)}) + \operatorname{pen}(\mathbf{a})\} \\ &= \operatorname{argmin}_{\mathbf{a}}\{-2\mathbf{a}'\mathbf{b}_r + \mathbf{a}'\mathbf{G}\mathbf{a} + 2(\mathbf{d}^{(r)})'|\mathbf{a}|\}\end{aligned}$$

- The crucial choice is the $\mathbf{d}^{(r)}$, should be data-driven !
- The theoretical validation : be able to state that our choice is the best possible choice.
- The practical validation : on simulated Hawkes processes (done), on simulated neuronal networks, on real data (RNRP Paris 6 work in progress)...

Theoretical Validation

Recall that

$$\mathbf{b}^{(r)} = \int_0^T \mathbf{Rc}_t dN^{(r)}(t)$$

and

$$\mathbf{G} = \int_0^T \mathbf{Rc}_t (\mathbf{Rc}_t)' dt.$$

Hansen,Rivoirard,RB

If $\mathbf{G} \geq cI$ with $c > 0$ and if

$$\left| \int_0^T \mathbf{Rc}_t \left(dN^{(r)}(t) - \lambda^{(r)}(t) dt \right) \right| \leq \mathbf{d}^{(r)}, \quad \forall r$$

then

$$\sum_r \|\lambda^{(r)} - \mathbf{Rc}_t \hat{\mathbf{a}}^{(r)}\|^2 \leq \square \inf_{\mathbf{a}} \left\{ \sum_r \|\lambda^{(r)} - \mathbf{Rc}_t \mathbf{a}\|^2 + \frac{1}{c} \sum_{i \in supp(\mathbf{a})} (d_i^{(r)})^2 \right\}.$$

Small proof (Univariate case)

Let us fix some $\mathbf{a}$ and let $\eta_{\mathbf{a}}(t) = \mathbf{Rc}_t'\mathbf{a}$, our candidate intensity.
Since $\mathbf{G} = \int_0^T \mathbf{Rc}_t(\mathbf{Rc}_t)'dt, \quad \mathbf{a}'\mathbf{Ga} = \|\eta_{\mathbf{a}}\|^2$

Small proof (Univariate case)

Let us fix some $\mathbf{a}$ and let $\eta_{\mathbf{a}}(t) = \mathbf{R}\mathbf{c}_t'\mathbf{a}$, our candidate intensity.
Since $\mathbf{G} = \int_0^T \mathbf{R}\mathbf{c}_t(\mathbf{R}\mathbf{c}_t)'dt, \quad \mathbf{a}'\mathbf{G}\mathbf{a} = \|\eta_{\mathbf{a}}\|^2$

$$\|\lambda - \eta_{\hat{\mathbf{a}}}\|^2 = \|\lambda\|^2 + \|\eta_{\hat{\mathbf{a}}}\| - 2\int_0^T \eta_{\hat{\mathbf{a}}}(t)\lambda(t)dt$$

$$= \|\lambda\|^2 + \gamma(\eta_{\hat{\mathbf{a}}}) + 2\int_0^T \eta_{\hat{\mathbf{a}}}(t)(dN_t - \lambda(t)dt)$$

Small proof (Univariate case)

Let us fix some $\mathbf{a}$ and let $\eta_{\mathbf{a}}(t) = \mathbf{Rc}_t'\mathbf{a}$, our candidate intensity.
Since $\mathbf{G} = \int_0^T \mathbf{Rc}_t(\mathbf{Rc}_t)'dt, \quad \mathbf{a}'\mathbf{Ga} = \|\eta_{\mathbf{a}}\|^2$

$$\|\lambda - \eta_{\hat{\mathbf{a}}}\|^2 = \|\lambda\|^2 + \|\eta_{\hat{\mathbf{a}}}\| - 2\int_0^T \eta_{\hat{\mathbf{a}}}(t)\lambda(t)dt$$

$$= \|\lambda\|^2 + \gamma(\eta_{\hat{\mathbf{a}}}) + 2\int_0^T \eta_{\hat{\mathbf{a}}}(t)(dN_t - \lambda(t)dt)$$

$$\leq \|\lambda\|^2 + \gamma(\eta_{\mathbf{a}}) + 2\mathbf{d}'(|\mathbf{a}| - |\hat{\mathbf{a}}|) + 2\int_0^T \eta_{\hat{\mathbf{a}}}(t)(dN_t - \lambda(t)dt)$$

Small proof (Univariate case)

Let us fix some $\mathbf{a}$ and let $\eta_{\mathbf{a}}(t) = \mathbf{Rc}_t'\mathbf{a}$, our candidate intensity.
Since $\mathbf{G} = \int_0^T \mathbf{Rc}_t(\mathbf{Rc}_t)'dt, \quad \mathbf{a}'\mathbf{Ga} = \|\eta_{\mathbf{a}}\|^2$

$$\|\lambda - \eta_{\hat{\mathbf{a}}}\|^2 = \|\lambda\|^2 + \|\eta_{\hat{\mathbf{a}}}\| - 2\int_0^T \eta_{\hat{\mathbf{a}}}(t)\lambda(t)dt$$

$$= \|\lambda\|^2 + \gamma(\eta_{\hat{\mathbf{a}}}) + 2\int_0^T \eta_{\hat{\mathbf{a}}}(t)(dN_t - \lambda(t)dt)$$

$$\leq \|\lambda\|^2 + \gamma(\eta_{\mathbf{a}}) + 2\mathbf{d}'(|\mathbf{a}| - |\hat{\mathbf{a}}|) + 2\int_0^T \eta_{\hat{\mathbf{a}}}(t)(dN_t - \lambda(t)dt)$$

$$\leq \|\lambda - \eta_{\mathbf{a}}\|^2 + 2\mathbf{d}'(|\mathbf{a}| - |\hat{\mathbf{a}}|) + 2\int_0^T [\eta_{\hat{\mathbf{a}}}(t) - \eta_{\mathbf{a}}(t)](dN_t - \lambda(t)dt)$$

Small proof (Univariate case)

Let us fix some $\mathbf{a}$ and let $\eta_{\mathbf{a}}(t) = \mathbf{R}\mathbf{c}_t'\mathbf{a}$, our candidate intensity.
Since $\mathbf{G} = \int_0^T \mathbf{R}\mathbf{c}_t(\mathbf{R}\mathbf{c}_t)'dt, \quad \mathbf{a}'\mathbf{G}\mathbf{a} = \|\eta_{\mathbf{a}}\|^2$

$$\|\lambda - \eta_{\hat{\mathbf{a}}}\|^2 = \|\lambda\|^2 + \|\eta_{\hat{\mathbf{a}}}\| - 2\int_0^T \eta_{\hat{\mathbf{a}}}(t)\lambda(t)dt$$

$$= \|\lambda\|^2 + \gamma(\eta_{\hat{\mathbf{a}}}) + 2\int_0^T \eta_{\hat{\mathbf{a}}}(t)(dN_t - \lambda(t)dt)$$

$$\leq \|\lambda\|^2 + \gamma(\eta_{\mathbf{a}}) + 2\mathbf{d}'(|\mathbf{a}| - |\hat{\mathbf{a}}|) + 2\int_0^T \eta_{\hat{\mathbf{a}}}(t)(dN_t - \lambda(t)dt)$$

$$\leq \|\lambda - \eta_{\mathbf{a}}\|^2 + 2\mathbf{d}'(|\mathbf{a}| - |\hat{\mathbf{a}}|) + 2\left(\int_0^T \mathbf{R}\mathbf{c}_t'(dN_t - \lambda(t)dt\right)(\hat{\mathbf{a}} - \mathbf{a})$$

Small proof (Univariate case)

Let us fix some $\mathbf{a}$ and let $\eta_{\mathbf{a}}(t) = \mathbf{Rc}_t'\mathbf{a}$, our candidate intensity.
Since $\mathbf{G} = \int_0^T \mathbf{Rc}_t(\mathbf{Rc}_t)'dt, \quad \mathbf{a}'\mathbf{Ga} = \|\eta_{\mathbf{a}}\|^2$

$$\|\lambda - \eta_{\hat{\mathbf{a}}}\|^2 = \|\lambda\|^2 + \|\eta_{\hat{\mathbf{a}}}\| - 2\int_0^T \eta_{\hat{\mathbf{a}}}(t)\lambda(t)dt$$

$$= \|\lambda\|^2 + \gamma(\eta_{\hat{\mathbf{a}}}) + 2\int_0^T \eta_{\hat{\mathbf{a}}}(t)(dN_t - \lambda(t)dt)$$

$$\leq \|\lambda\|^2 + \gamma(\eta_{\mathbf{a}}) + 2\mathbf{d}'(|\mathbf{a}| - |\hat{\mathbf{a}}|) + 2\int_0^T \eta_{\hat{\mathbf{a}}}(t)(dN_t - \lambda(t)dt)$$

$$\leq \|\lambda - \eta_{\mathbf{a}}\|^2 + 2\mathbf{d}'(|\mathbf{a}| - |\hat{\mathbf{a}}|) + 2\mathbf{d}'|\mathbf{a} - \hat{\mathbf{a}}|$$

Small proof (Univariate case)(2)

$$\|\lambda - \eta_{\hat{\mathbf{a}}}\|^2 \leq \|\lambda - \eta_{\mathbf{a}}\|^2 + 2\mathbf{d}'(|\mathbf{a}| - |\hat{\mathbf{a}}|) + 2\mathbf{d}'|\mathbf{a} - \hat{\mathbf{a}}|$$

Small proof (Univariate case)(2)

$$\|\lambda - \eta_{\hat{\mathbf{a}}}\|^2 \leq \|\lambda - \eta_{\mathbf{a}}\|^2 + 2\mathbf{d}'(|\mathbf{a}| - |\hat{\mathbf{a}}|) + 2\mathbf{d}'|\mathbf{a} - \hat{\mathbf{a}}|$$
$$\leq \|\lambda - \eta_{\mathbf{a}}\|^2 + 4 \sum_{i \in supp(a)} \mathbf{d}_i |\mathbf{a}_i - \hat{\mathbf{a}}_i|$$

Small proof (Univariate case)(2)

$$\|\lambda - \eta_{\hat{\mathbf{a}}}\|^2 \leq \|\lambda - \eta_{\mathbf{a}}\|^2 + 2\mathbf{d}'(|\mathbf{a}| - |\hat{\mathbf{a}}|) + 2\mathbf{d}'|\mathbf{a} - \hat{\mathbf{a}}|$$

$$\leq \|\lambda - \eta_{\mathbf{a}}\|^2 + 4 \sum_{i \in supp(a)} \mathbf{d}_i |\mathbf{a}_i - \hat{\mathbf{a}}_i|$$

$$\leq \|\lambda - \eta_{\mathbf{a}}\|^2 + 4\|\mathbf{a} - \hat{\mathbf{a}}\| \left(\sum_{i \in supp(a)} \mathbf{d}_i^2 \right)^{1/2}$$

Small proof (Univariate case)(2)

$$\|\lambda - \eta_{\hat{\mathbf{a}}}\|^2 \leq \|\lambda - \eta_{\mathbf{a}}\|^2 + 2\mathbf{d}'(|\mathbf{a}| - |\hat{\mathbf{a}}|) + 2\mathbf{d}'|\mathbf{a} - \hat{\mathbf{a}}|$$
$$\leq \|\lambda - \eta_{\mathbf{a}}\|^2 + 4 \sum_{i \in supp(a)} \mathbf{d}_i |\mathbf{a}_i - \hat{\mathbf{a}}_i|$$
$$\leq \|\lambda - \eta_{\mathbf{a}}\|^2 + 4\|\mathbf{a} - \hat{\mathbf{a}}\| \left(\sum_{i \in supp(a)} \mathbf{d}_i^2 \right)^{1/2}$$

But

$$\|\mathbf{a} - \hat{\mathbf{a}}\|^2 \leq \frac{1}{c}(\mathbf{a} - \hat{\mathbf{a}})'\mathbf{G}(\mathbf{a} - \hat{\mathbf{a}}) = \frac{1}{c}\|\eta_{\mathbf{a}} - \eta_{\hat{\mathbf{a}}}\|^2$$
$$\leq \frac{2}{c}\left[\|\eta_{\mathbf{a}} - \lambda\|^2 + \|\lambda - \eta_{\hat{\mathbf{a}}}\|^2\right]$$

Small proof (Univariate case)(2)

$$\|\lambda - \eta_{\hat{\mathbf{a}}}\|^2 \leq \|\lambda - \eta_{\mathbf{a}}\|^2 + 2\mathbf{d}'(|\mathbf{a}| - |\hat{\mathbf{a}}|) + 2\mathbf{d}'|\mathbf{a} - \hat{\mathbf{a}}|$$

$$\leq \|\lambda - \eta_{\mathbf{a}}\|^2 + 4 \sum_{i \in supp(a)} \mathbf{d}_i |\mathbf{a}_i - \hat{\mathbf{a}}_i|$$

NB: $uv \leq \varepsilon u^2 + \frac{1}{\varepsilon} v^2 \quad \forall \varepsilon > 0$

$\Leftarrow \quad \left(\sqrt{\varepsilon} u + \frac{v}{\sqrt{\varepsilon}}\right)^2 = \varepsilon u^2 + \frac{v^2}{\varepsilon} + 2uv.$

$$\leq \|\lambda - \eta_{\mathbf{a}}\|^2 + 4\|\mathbf{a} - \hat{\mathbf{a}}\| \left(\sum_{i \in supp(a)} \mathbf{d}_i^2 \right)^{1/2}$$

Hence for all $\alpha > 0$

$$\|\lambda - \eta_{\hat{\mathbf{a}}}\|^2 \leq \|\lambda - \eta_{\mathbf{a}}\|^2 + \alpha \left[\|\eta_{\mathbf{a}} - \lambda\|^2 + \|\lambda - \eta_{\hat{\mathbf{a}}}\|^2 \right] + \frac{8\alpha}{c} \sum_{i \in supp(a)} \mathbf{d}_i^2$$

What did we proved ? (Univariate)

Recall that

$$\mathbf{b} = \int_0^T \mathbf{Rc}_t dN(t)$$

and

$$\mathbf{G} = \int_0^T \mathbf{Rc}_t (\mathbf{Rc}_t)' dt.$$

Hansen,Rivoirard,RB

If $\mathbf{G} \geq cI$ with $c > 0$ and if

$$\left| \int_0^T \mathbf{Rc}_t \left(dN(t) - \lambda(t) dt \right) \right| \leq \mathbf{d},$$

then

$$\|\lambda - \mathbf{Rc}_t \hat{\mathbf{a}}\|^2 \leq \square \inf_{\mathbf{a}} \left\{ \|\lambda - \mathbf{Rc}_t \mathbf{a}\|^2 + \frac{1}{c} \sum_{i \in supp(\mathbf{a})} (d_i)^2 \right\}.$$

Known support

If there is a true parameter $\mathbf{a}^*$, then $\lambda(t) = \mathbf{R}\mathbf{c}_t'\mathbf{a}^*$ and

$$\bar{\mathbf{b}} = \int_0^T \mathbf{R}\mathbf{c}_t \lambda(t) dt = \mathbf{G}\mathbf{a}^*$$

Known support

If there is a true parameter $\mathbf{a}^*$, then $\lambda(t) = \mathbf{R}\mathbf{c}_t'\mathbf{a}^*$ and

$$\bar{\mathbf{b}} = \int_0^T \mathbf{R}\mathbf{c}_t \lambda(t) dt = \mathbf{G}\mathbf{a}^*$$

If support of $\mathbf{a}^*$, S, known, the least-square estimate on $S \to \hat{\mathbf{a}}_S$

Known support

If there is a true parameter $\mathbf{a}^*$, then $\lambda(t) = \mathbf{R}\mathbf{c}_t'\mathbf{a}^*$ and

$$\bar{\mathbf{b}} = \int_0^T \mathbf{R}\mathbf{c}_t \lambda(t) dt = \mathbf{G}\mathbf{a}^*$$

If support of $\mathbf{a}^*$, S, known, the least-square estimate on $S \to \hat{\mathbf{a}}_S$

$$\begin{aligned} \|\lambda - \mathbf{R}\mathbf{c}_t \hat{\mathbf{a}}_S\|^2 &= (\mathbf{a}^* - \hat{\mathbf{a}}_S)'\mathbf{G}(\mathbf{a}^* - \hat{\mathbf{a}}_S) \\ &= (\bar{\mathbf{b}} - \mathbf{b})'\mathbf{G}^{-1}(\bar{\mathbf{b}} - \mathbf{b}) \\ &\leq \frac{1}{c}\|\bar{\mathbf{b}} - \mathbf{b}\|^2 \end{aligned}$$

Oracle inequality

Lasso criterion

$$\begin{aligned}\hat{\mathbf{a}}^{(r)} &= \operatorname{argmin}_{\mathbf{a}}\{\gamma_T(\lambda_{\mathbf{a}}^{(r)}) + \mathrm{pen}(\mathbf{a})\} \\ &= \operatorname{argmin}_{\mathbf{a}}\{-2\mathbf{a}'\mathbf{b}_r + \mathbf{a}'\mathbf{G}\mathbf{a} + 2(\mathbf{d}^{(r)})'|\mathbf{a}|\}\end{aligned}$$

Hansen,Rivoirard,RB

If $\mathbf{G} \geq cI$ with $c > 0$ and if

$$\left|\mathbf{b}^{(r)} - \bar{\mathbf{b}}^{(r)}\right| \leq \mathbf{d}^{(r)}, \quad \forall r$$

then

$$\sum_r \|\lambda^{(r)} - \mathbf{R}\mathbf{c}_t\hat{\mathbf{a}}^{(r)}\|^2 \leq \square \inf_{\mathbf{a}} \left\{ \sum_r \|\lambda^{(r)} - \mathbf{R}\mathbf{c}_t\mathbf{a}\|^2 + \frac{1}{c} \sum_{i \in supp(\mathbf{a})} (d_i^{(r)})^2 \right\}.$$

Comments

- If we can find $\mathbf{d}$ sharp and data-driven, then this is an oracle inequality ! ! :
  - Only an oracle could know the support in advance
  - Up to constant, we can do as well as the oracle

Comments

- If we can find **d** sharp and data-driven, then this is an oracle inequality ! ! :
  - Only an oracle could know the support in advance
  - Up to constant, we can do as well as the oracle
- We do not pay anything here for the size, except **G** invertible
  - it's a "cheap" oracle inequality ... In fact could be done under more relaxed assumption...
  - c large (and we can observe it !) $\rightarrow$ quite confident for good reconstruction =quality indicator.

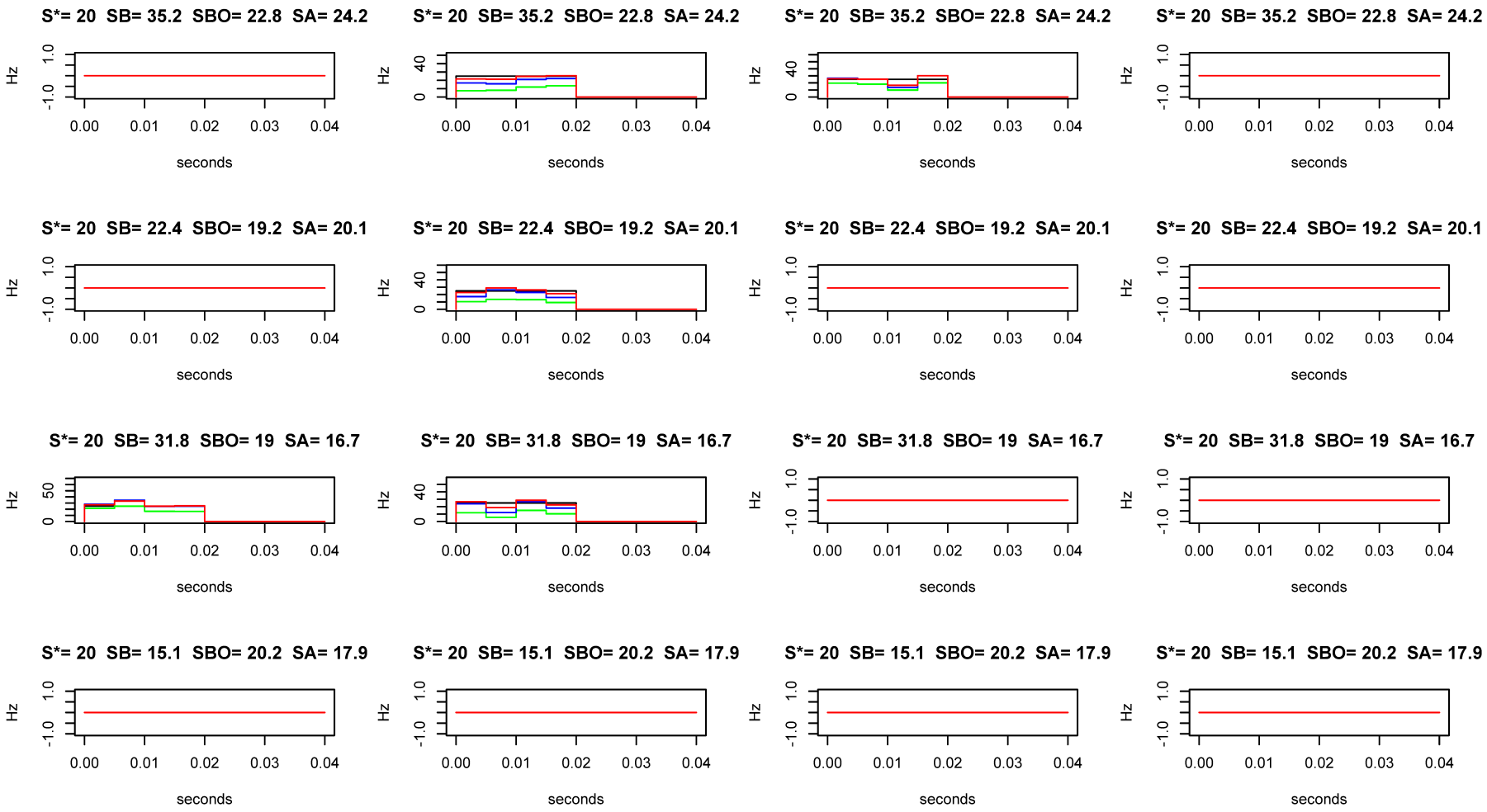
Comments

- If we can find **d** sharp and data-driven, then this is an oracle inequality ! ! :
  - Only an oracle could know the support in advance
  - Up to constant, we can do as well as the oracle
- We do not pay anything here for the size, except **G** invertible
  - it's a "cheap" oracle inequality ... In fact could be done under more relaxed assumption...
  - c large (and we can observe it !) $\rightarrow$ quite confident for good reconstruction =quality indicator.
- True even if the Hawkes linear model not true ! ! ! and we just pay "quality of approximation"
- quality of approximation + unavoidable price due to estimation.

Mathematical "debts"

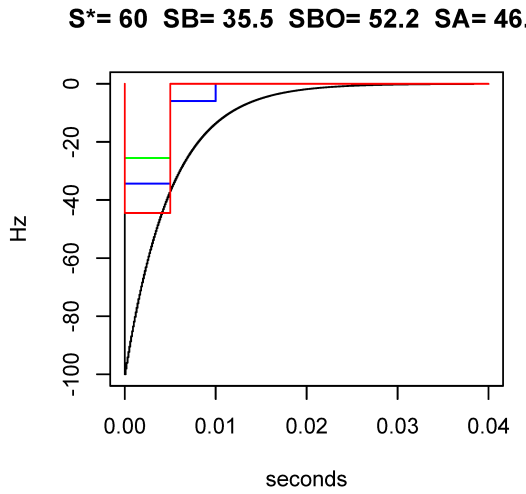
- works for other models (anything that is linear), works for other basis and dictionary ...
- Choice of **d** : $\rightarrow$ martingale calculus
- **G** invertible in some cases ? ? ? $\rightarrow$ branching structure

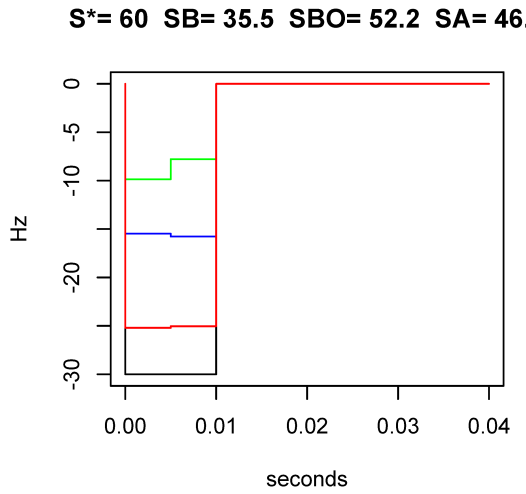
Simulation study - Estimation

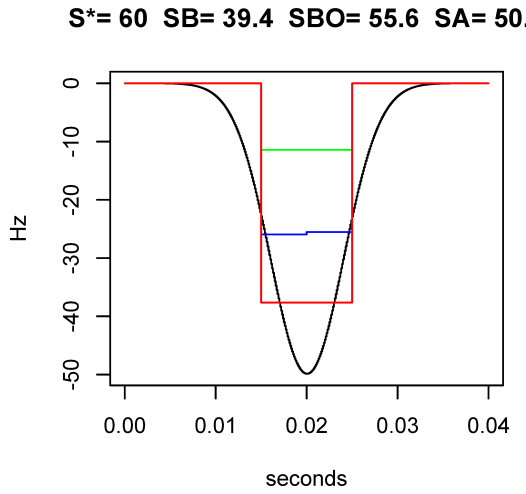


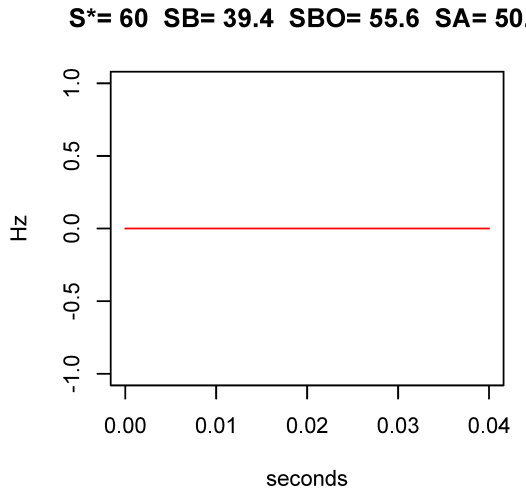
Interactions reconstructed with 'Adaptive Lasso', 'Bernstein Lasso' and 'Bernstein Lasso+OLS'. Above graphs, estimation of spontaneous rates

Another example with inhibition

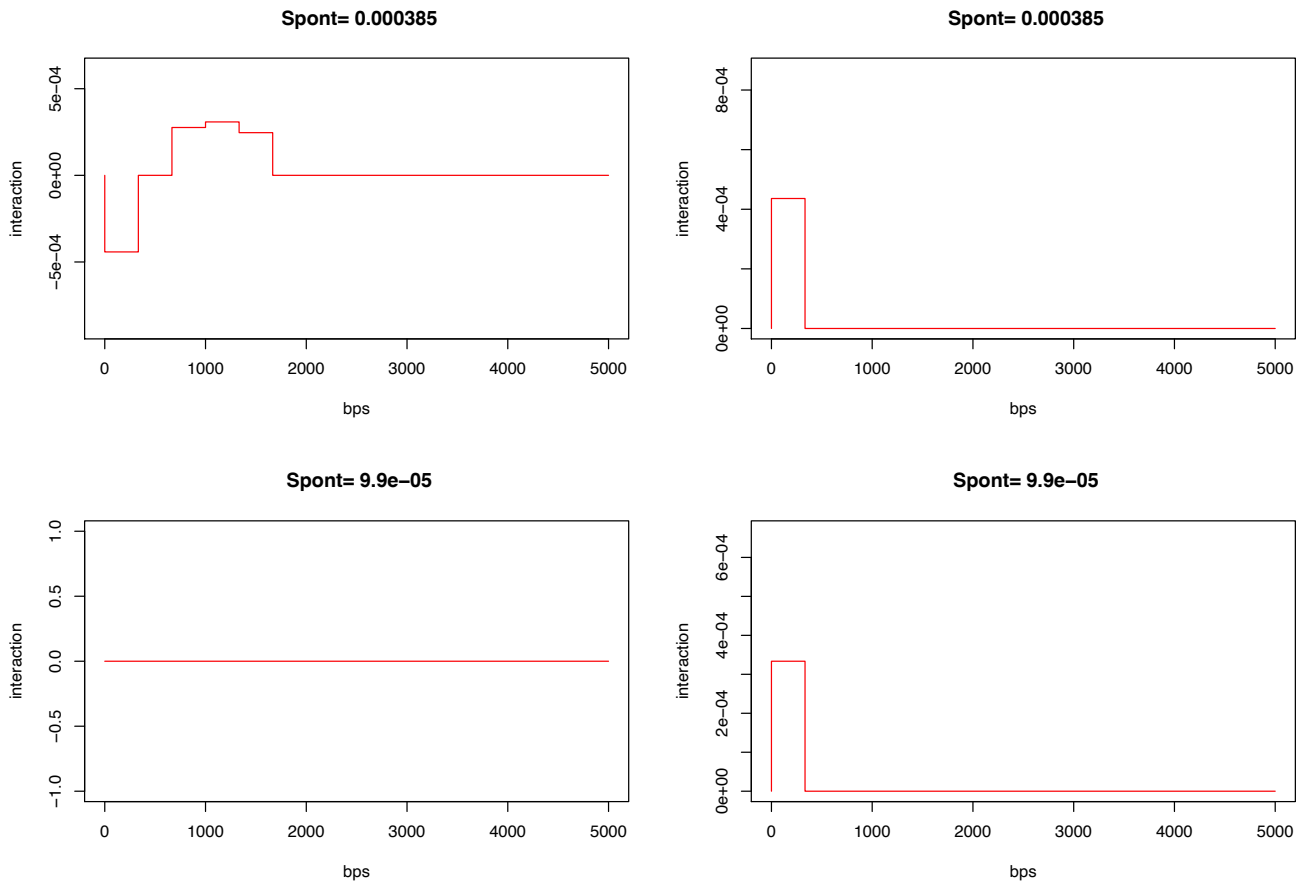








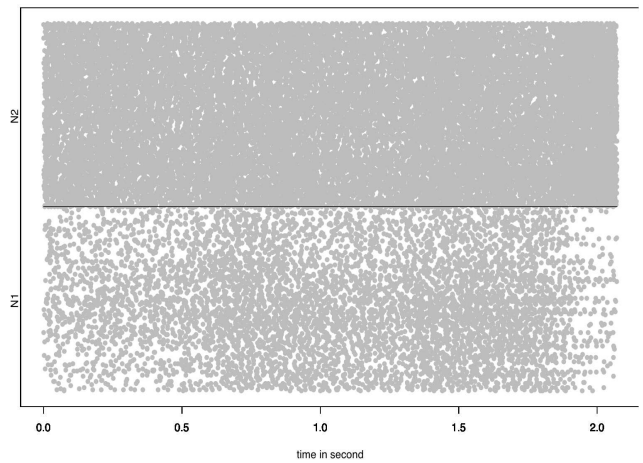
On real (genomic) data

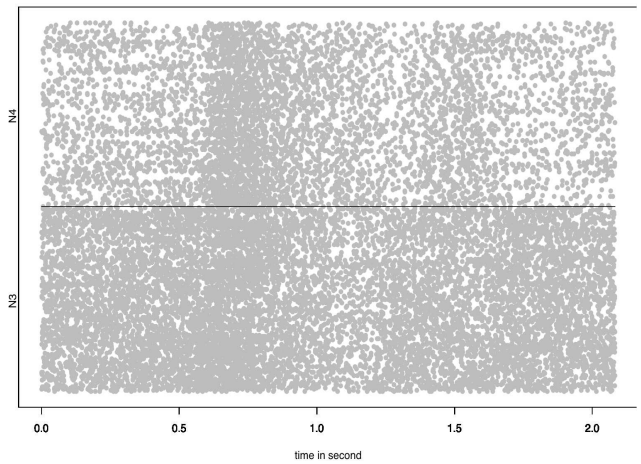


4290 genes and 1036 `tataat` of E. coli ($T = 9288442$, $A = 10000$)

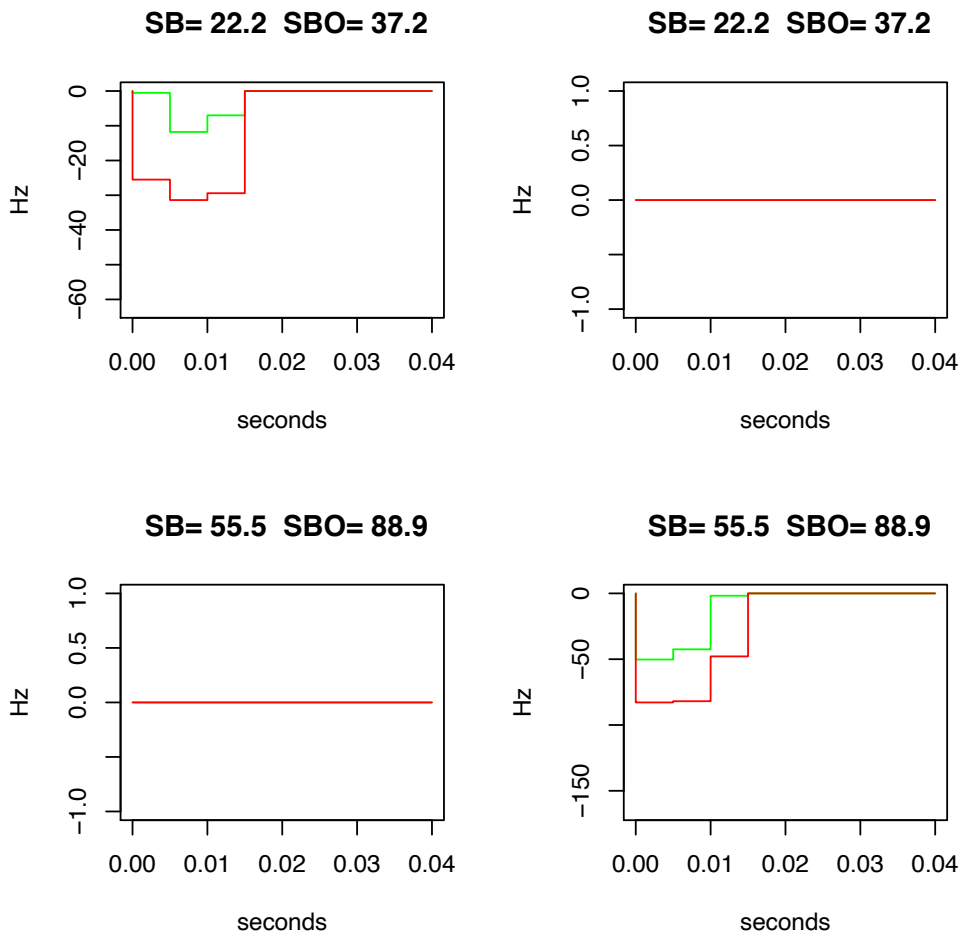
Sensory-motor task

(F. Grammont (Nice), A. Riehle (Marseille))





On neuronal data (sensorimotor task)



30 trials : monkey trained to touch the correct target when illuminated.

Accept the test of Hawkes hypothesis. Work with F. Grammont. V.

Tests

with C. Tuleau-Malot, F. Grammont (Nice) and V. Rivoirard (Dauphine) (2013)

- It is possible to test whether a process is a Hawkes process with prescribed interaction functions by using the time-rescaling theorem. (known since Ogata)

Tests

with C. Tuleau-Malot, F. Grammont (Nice) and V. Rivoirard (Dauphine) (2013)

- It is possible to test whether a process is a Hawkes process with prescribed interaction functions by using the time-rescaling theorem. (known since Ogata)
- By using subsampling, it is possible to plug an estimate of the functions and still to test with controlled asymptotic level.

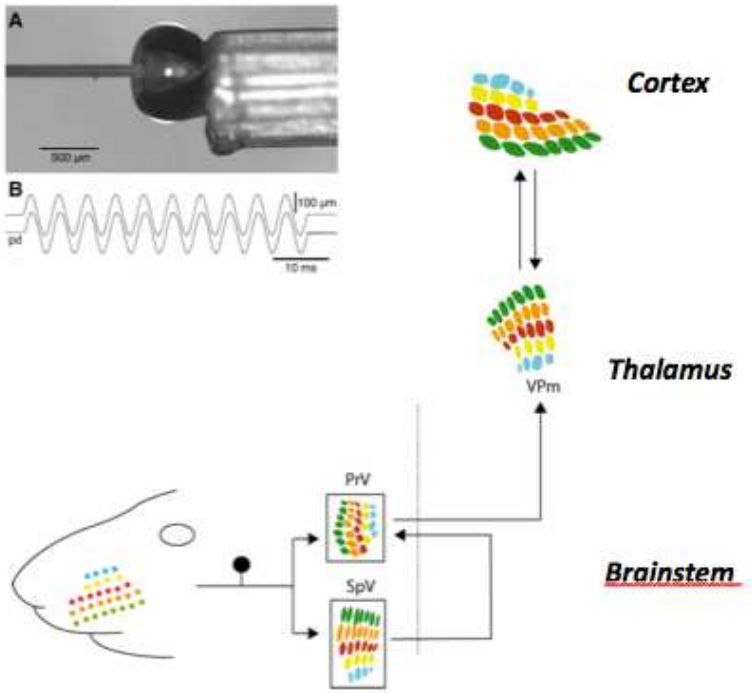
Tests

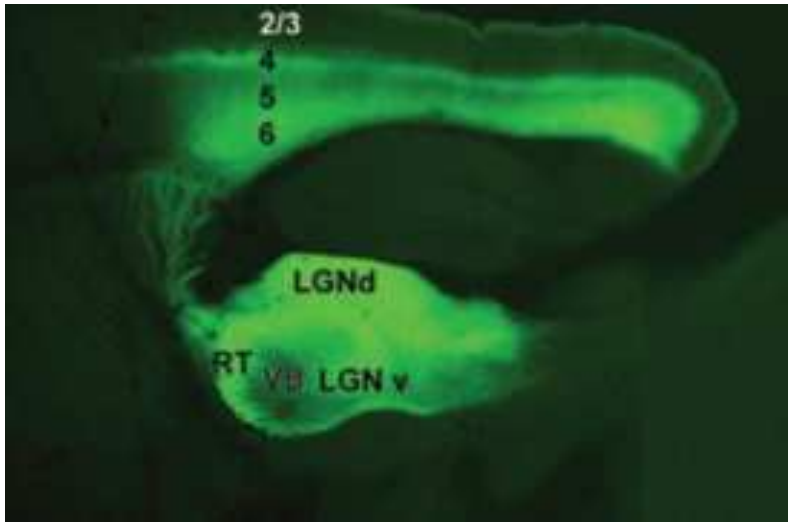
with C. Tuleau-Malot, F. Grammont (Nice) and V. Rivoirard (Dauphine) (2013)

- It is possible to test whether a process is a Hawkes process with prescribed interaction functions by using the time-rescaling theorem. (known since Ogata)
- By using subsampling, it is possible to plug an estimate of the functions and still to test with controlled asymptotic level.
- On both previous data sets, the Hawkes hypothesis is accepted (p-values depends on the sub-sample, usually between 20 and 80 %), whereas the homogeneous Poisson hypothesis (i.e. no interactions) is rejected (p-values in $10^{-4}, 10^{-16}$ for the neuronal data)

Functional Connectivity ?

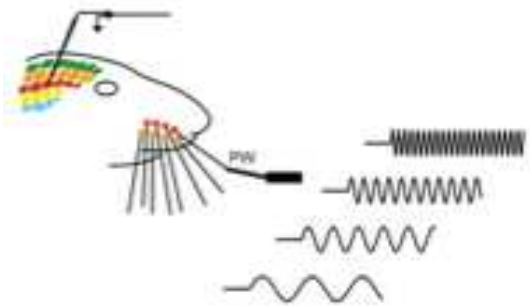
(Equipe RNRP de Paris 6)

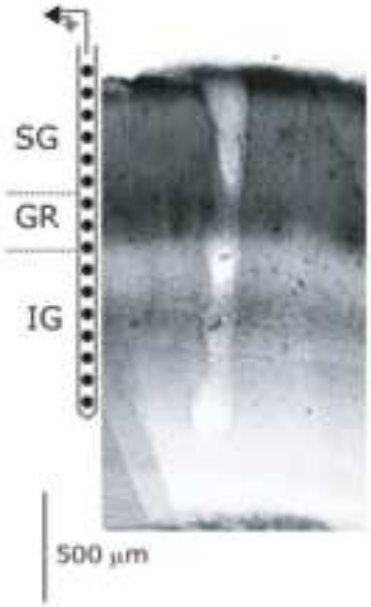


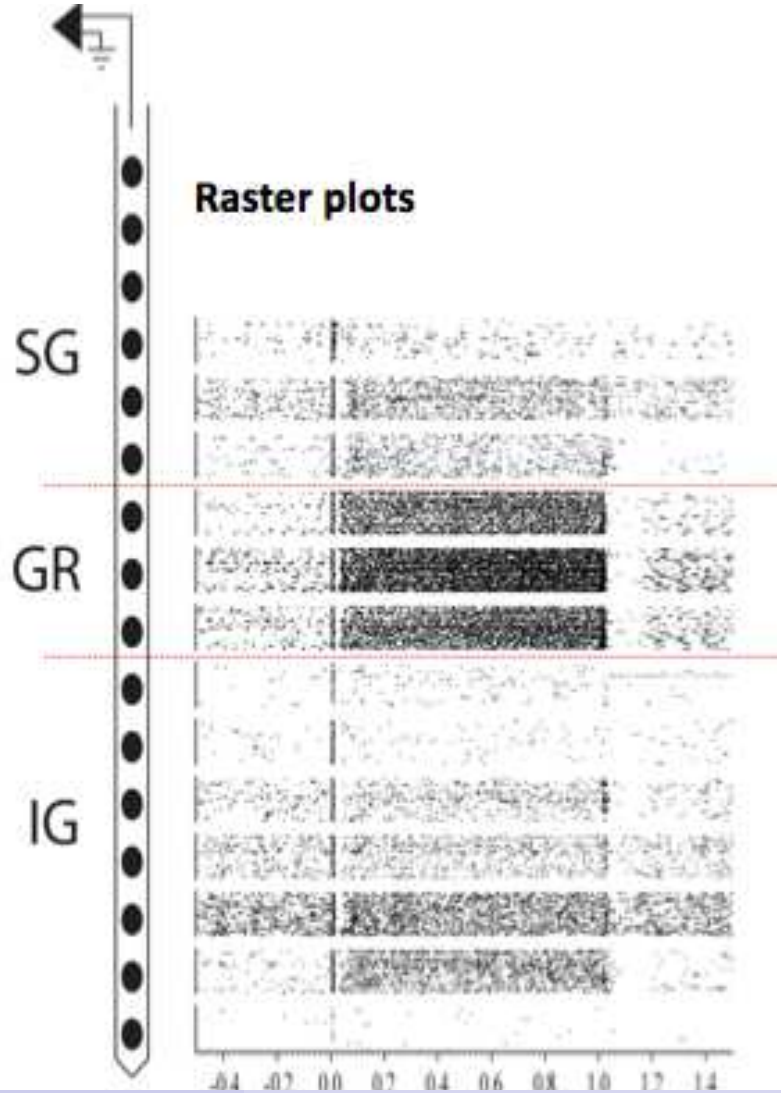


Tetrode data on the rat

(Equipe RNRP de Paris 6)



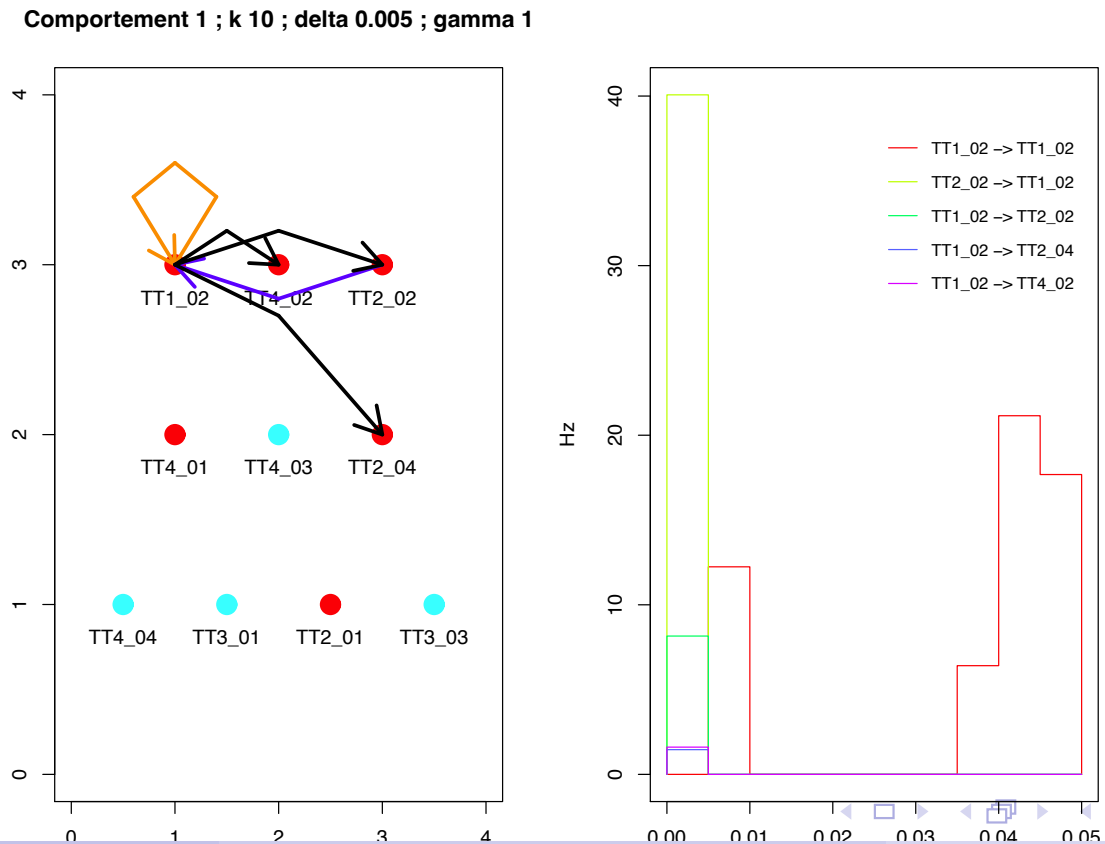




On neuronal data (vibrissa excitation)

Joint work with RNRP (Paris 6). Behavior : vibrissa excitation at low frequency. $T = 90.5$, $M = 10$

Neuron	$TT1_{02}$	$TT2_{01}$	$TT2_{02}$	$TT2_{04}$	$TT3_{01}$	$TT3_{03}$	$TT4_{01}$	$TT4_{02}$	$TT4_{03}$	$TT4_{04}$
Spikes	9191	99	544	149	15	18	136	282	8	6



Application on real data

Data :

Neuron	$TT1_{02}$	$TT2_{01}$	$TT2_{02}$	$TT2_{04}$	$TT3_{01}$	$TT3_{03}$	$TT4_{01}$	$TT4_{02}$	$TT4_{03}$	$TT4_{04}$
Spikes	9191	99	544	149	15	18	136	282	8	6

Simulation :

Neuron	$TT1_{02}$	$TT2_{01}$	$TT2_{02}$	$TT2_{04}$	$TT3_{01}$	$TT3_{03}$	$TT4_{01}$	$TT4_{02}$	$TT4_{03}$	$TT4_{04}$
Spikes	9327	92	585	148	13	23	133	271	8	3

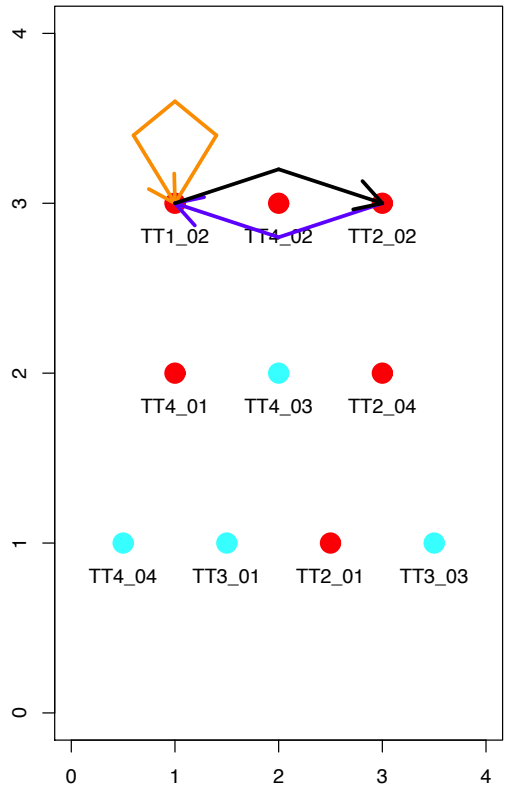
Application on real data

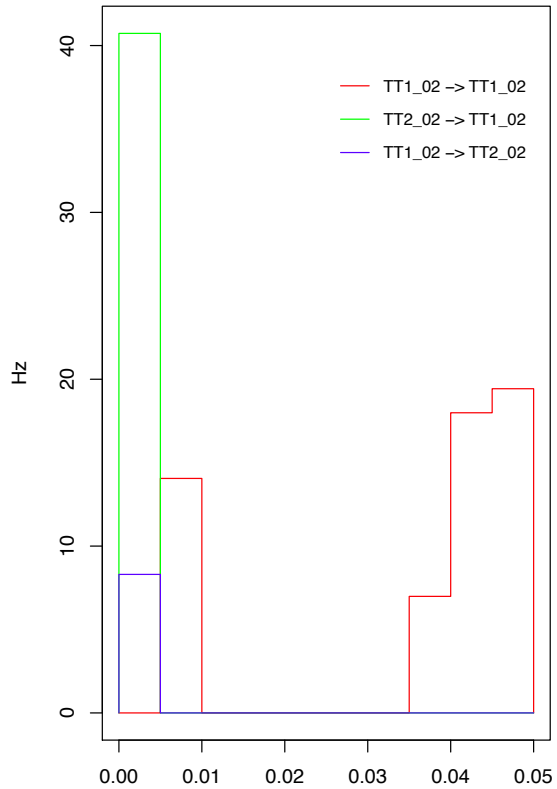
Data :

Neuron	$TT1_{02}$	$TT2_{01}$	$TT2_{02}$	$TT2_{04}$	$TT3_{01}$	$TT3_{03}$	$TT4_{01}$	$TT4_{02}$	$TT4_{03}$	$TT4_{04}$
Spikes	9191	99	544	149	15	18	136	282	8	6

Simulation :

Neuron	$TT1_{02}$	$TT2_{01}$	$TT2_{02}$	$TT2_{04}$	$TT3_{01}$	$TT3_{03}$	$TT4_{01}$	$TT4_{02}$	$TT4_{03}$	$TT4_{04}$
Spikes	9327	92	585	148	13	23	133	271	8	3





Evolution of the dependance graph as a fonction of the vibrissa excitation

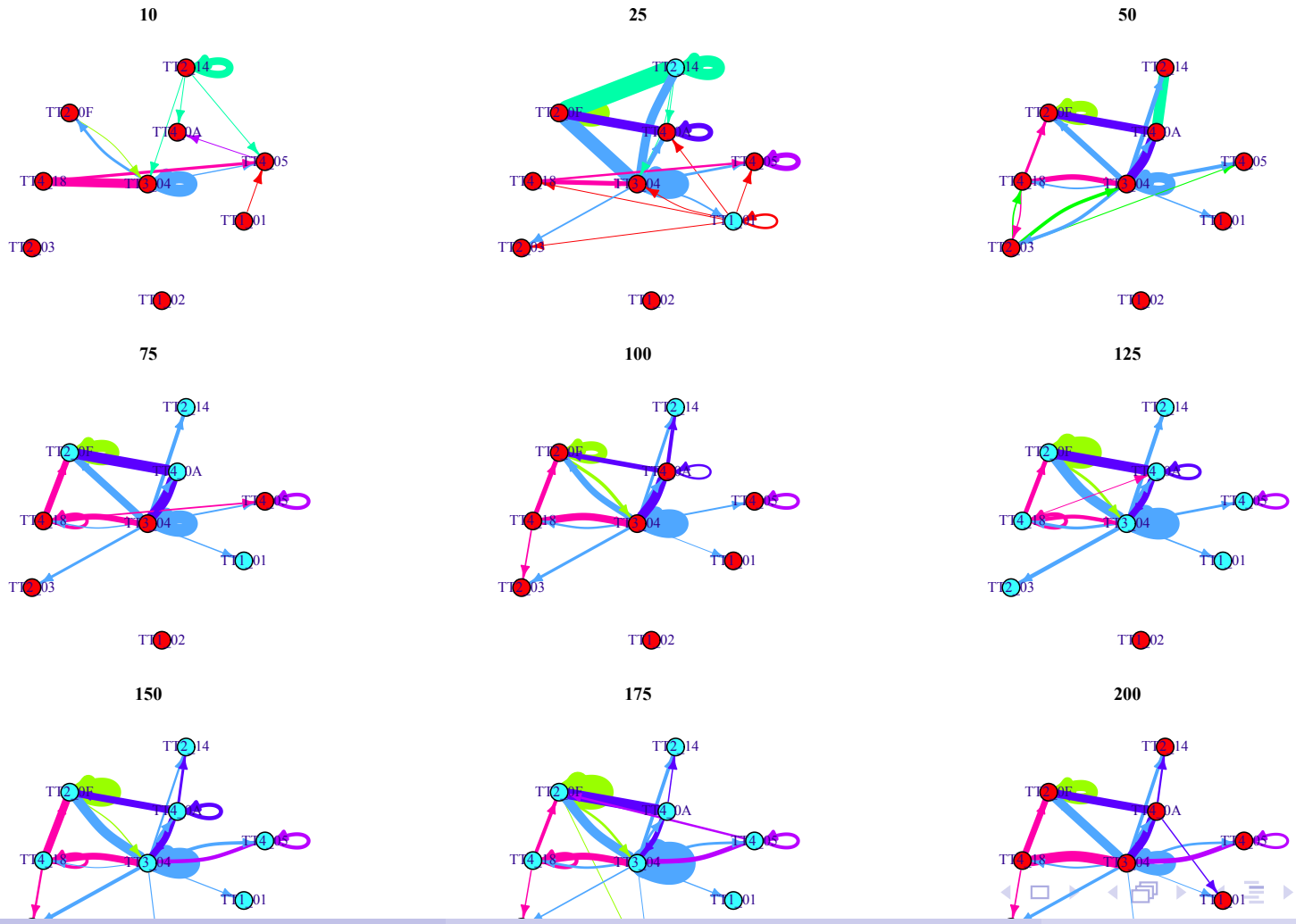


Table of Contents

Table of Contents

Probability generating functional and Laplace transform

p.g.fl

For any point process N, the functional which associates to any positive function h

$$\mathbb{E}(\prod_{x \in N} h(x))$$

Laplace functional

For any point process N, the functional which associates to any function f

$$\mathbb{E}\left(\exp\left[\int f(x) dN_x\right]\right).$$

equivalence with $f = \log(h)$.

Campbell's theorem

Let N be a Poisson process with mean measure μ,

Poisson processes

- for all integer n, for all $A_1, \ldots, A_n$ disjoint measurable subsets of $\mathbb{X}$, $N_{A_1}, \ldots, N_{A_n}$ are independent random variables.
- for all measurable subset A of $\mathbb{X}$, N_A obeys a Poisson law with parameter depending on A and denoted $\mu(A)$.

Campbell's theorem

Let N be a Poisson process with mean measure μ,
typically, $d\mu_t = \lambda(t)dt$, with λ deterministic,

Campbell's theorem

Let N be a Poisson process with mean measure μ,
typically, $d\mu_t = \lambda(t)dt$, with λ deterministic,

Campbell's theorem

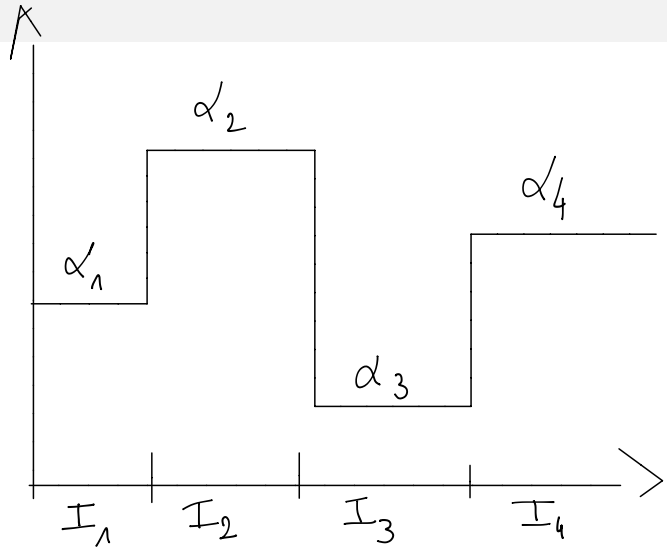
If f is such that $\int \min(|f(x)|, 1) d\mu_x < \infty$, then

$$\mathbb{E}\left(\exp\left[\int f(x) dN_x\right]\right) = \exp\left(\int \left[e^{f(x)} - 1\right] d\mu_x\right).$$

Small proof

Take f piecewise constant,

$$f = \sum_I \alpha_I \mathbf{1}_I.$$



$$\begin{aligned}
\mathbb{E}\left(\exp\left[\int f(x) dN_x\right]\right) &= \mathbb{E}\left(\prod_I e^{\alpha_I N_I}\right) \\
&= \prod_I \mathbb{E}\left(e^{\alpha_I N_I}\right) \text{ (by independence)} \\
&= \prod_I \exp\left((e^{\alpha_I} - 1)\mu_I\right) \text{ (Laplace of a Poisson)} \\
&= \exp\left(\int (e^{f(x)} - 1) d\mu_x\right)
\end{aligned}$$

Exponential inequality for Poisson process

Aim

For all $\xi > 0$,

$$\mathbb{P}\left(\int f(x)[dN_x - \lambda(x)dx] \geq ???(y)\right) \leq e^{-y}$$

NB : easier this way for interpretation in statistics... (quantile)

Let f be some fixed function (deterministic) and let $\theta > 0$.
Apply Campbell's theorem to θf :

$$\mathbb{E}\left(\exp\left[\int \theta f(x)[dN_x - \lambda(x)dx]\right]\right) = \exp\left(\int (e^{\theta f(x)} - \theta f(x) - 1)d\mu_x\right).$$

Exponential inequality for Poisson process

Aim

For all $\xi > 0$,

$$\mathbb{P}\left(\int f(x)[dN_x - \lambda(x)dx] \geq ???(y)\right) \leq e^{-y}$$

NB : easier this way for interpretation in statistics... (quantile)

Let f be some fixed function (deterministic) and let $\theta > 0$.
Apply Campbell's theorem to θf :

$$\mathbb{E}\left(\exp\left[\int \theta f(x)[dN_x - \lambda(x)dx]\right]\right) = \exp\left(\int (e^{\theta f(x)} - \theta f(x) - 1)d\mu_x\right).$$

Hence if $E = \exp\left[\int \theta f(x)[dN_x - \lambda(x)dx] - \int (e^{\theta f(x)} - \theta f(x) - 1)d\mu_x\right]$,

$$\mathbb{E}(E) = 1.$$

Towards a (weak) Bernstein inequality...

Therefore for all $y > 0$

$$\mathbb{P}\left(\int \theta f(x)[dN_x - \lambda(x)dx] \geq \int (e^{\theta f(x)} - \theta f(x) - 1)d\mu_x + y\right) = \mathbb{P}(E \geq e^y)$$

Towards a (weak) Bernstein inequality...

Therefore for all $y > 0$

$$\mathbb{P}\left(\int \theta f(x)[dN_x - \lambda(x)dx] \geq \int (e^{\theta f(x)} - \theta f(x) - 1)d\mu_x + y\right) = \mathbb{P}(E \geq e^y)$$

By Markov,

$$\mathbb{P}\left(\int \theta f(x)[dN_x - \lambda(x)dx] \geq \int (e^{\theta f(x)} - \theta f(x) - 1)d\mu_x + y\right) \leq \mathbb{E}(E)e^{-y}.$$

Towards a (weak) Bernstein inequality...

Therefore for all $y > 0$

$$\mathbb{P}\left(\int \theta f(x)[dN_x - \lambda(x)dx] \geq \int (e^{\theta f(x)} - \theta f(x) - 1)d\mu_x + y\right) = \mathbb{P}(E \geq e^y)$$

Hence for all $\theta, y > 0$

$$\mathbb{P}\left(\int f(x)[dN_x - \lambda(x)dx] \geq \frac{1}{\theta}\int (e^{\theta f(x)} - \theta f(x) - 1)d\mu_x + y/\theta\right) \leq e^{-y}.$$

Towards a (weak) Bernstein inequality...

Therefore for all $y > 0$

$$\mathbb{P}\left(\int \theta f(x)[dN_x - \lambda(x)dx] \geq \int (e^{\theta f(x)} - \theta f(x) - 1)d\mu_x + y\right) = \mathbb{P}(E \geq e^y)$$

Hence for all $\theta, y > 0$

$$\mathbb{P}\left(\int f(x)[dN_x - \lambda(x)dx] \geq \frac{1}{\theta}\int (e^{\theta f(x)} - \theta f(x) - 1)d\mu_x + y/\theta\right) \leq e^{-y}.$$

But

$$\begin{aligned} e^{\theta f(x)} - \theta f(x) - 1 &\leq \sum_{k=2}^{\infty} \frac{\theta^k}{k!}\|f\|_\infty^{k-2} f(x)^2 \\ &\leq \frac{\theta^2 f(x)^2}{2\left(1 - \frac{\theta\|f\|_\infty}{3}\right)} \end{aligned}$$

Let $v = \int f(x)^2 d\mu_x$.

Towards a (weak) Bernstein inequality...(2)

Let $v = \int f(x)^2 \lambda(x) dx$.
Hence for all $\theta, y > 0$

$$\mathbb{P}\left(\int f(x)[dN_x - \lambda(x)dx] \geq \frac{\theta v}{2\left(1 - \frac{\theta \|f\|_\infty}{3}\right)} + y/\theta \right) \leq e^{-y}$$

Towards a (weak) Bernstein inequality...(2)

Let $v = \int f(x)^2 \lambda(x) dx$.
Hence for all $\theta, y > 0$

$$\mathbb{P}\left(\int f(x)[dN_x - \lambda(x)dx] \geq \frac{\theta v}{2\left(1 - \frac{\theta \|f\|_\infty}{3}\right)} + y/\theta \right) \leq e^{-y}$$

Optimum in $\theta = g(v, \|f\|_\infty, y)$ and

NB : si $X \sim \mathcal{N}(m, \sigma^2)$
$\mathbb{P}(X - m \geq t) \leq e^{-\frac{t^2}{2\sigma^2}}$
ie $\mathbb{P}(X - m \geq \sqrt{2\sigma^2 y}) \leq e^{-y}$

Exponential inequality "à la" Bernstein

For all $y > 0$,

$$\mathbb{P}\left(\int f(x)[dN_x - \lambda(x)dx] \geq \sqrt{2vy} + \frac{\|f\|_\infty y}{3} \right) \leq e^{-y}$$

NB : $v = \mathbb{E}\left(\left[\int f(x)[dN_x - \lambda(x)dx] \right]^2 \right) = Var\left(\int f(x) dN_x \right)$. Hence first order Gaussian tail with variance v

Aim

- same kind of exponential inequality for Hawkes (or other general counting process)
- $f(.)$ deterministic has to be replaced by H_t (ex : $\mathbf{Rc}_t$) predictable
- The "variance" term v should follow $\int H_s dN_s \to$ expectation given the past.

Aim

- same kind of exponential inequality for Hawkes (or other general counting process)
- $f(.)$ deterministic has to be replaced by H_t (ex : $\mathbf{Rc}_t$) predictable
- The "variance" term v should follow $\int H_s dN_s \to$ expectation given the past.
- optional : v still depends on λ, unknown $\to$ as to be replaced by observable quantity.

Martingale

Let N counting process with (predictable) intensity λ. Let H_s be any predictable process and $t > u$, then

$$\mathbb{E}(H_t dN_t \Big| \text{ past at t}) = H_t \mathbb{E}(dN_t | \text{ past at t}) = H_t \lambda(t) dt$$

Martingale

Let N counting process with (predictable) intensity λ. Let H_s be any predictable process and $t > u$, then

$$\mathbb{E}(H_t dN_t \Big| \text{ past at t}) = H_t \mathbb{E}(dN_t | \text{ past at t}) = H_t \lambda(t) dt$$

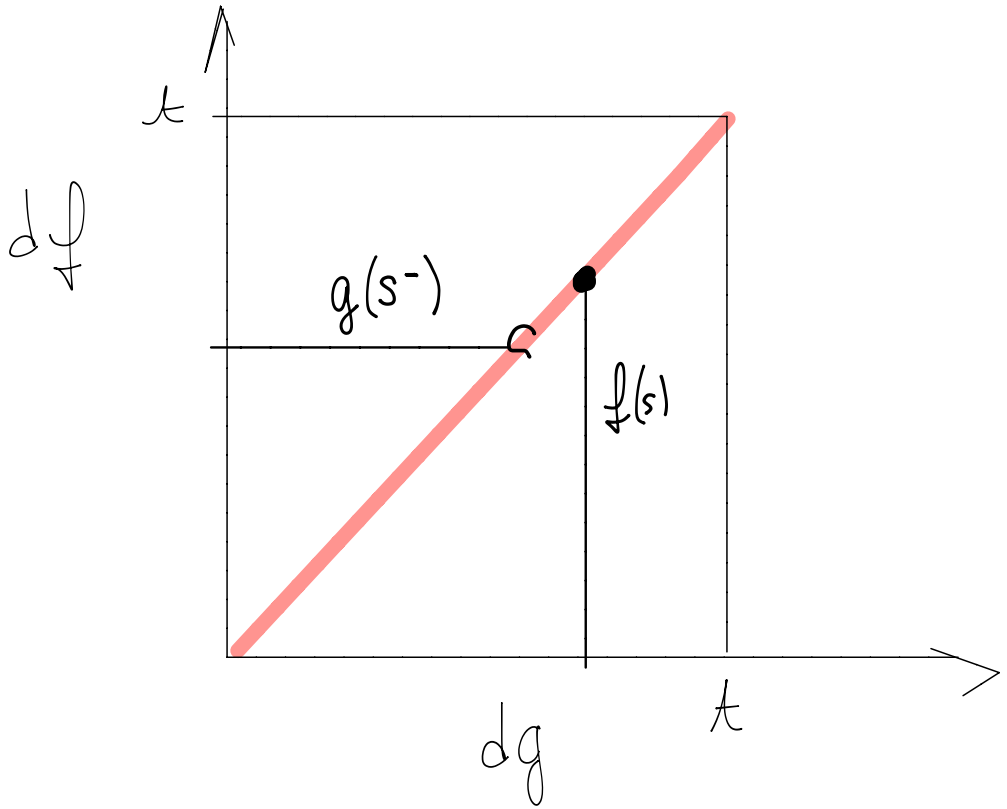
$$\mathbb{E}\left(\int_0^t H_s [dN_s - \lambda(s) ds] \Big| \text{ past at u} \right) = \int_0^u H_s [dN_s - \lambda(s) ds]$$

$\rightarrow$ martingale.

Bracket

A Stieljes integration by part : whatever the bounded variation càd-làg functions (i.e. works for dN_t, $\lambda(t)dt$ or $dN_t - \lambda((t)dt)$

$$f(t)g(t) = f(0)g(0) + \int_0^t f(s)dg_s + \int_0^t g(s^-)df_s$$



Bracket

A Stieljes integration by part : whatever the bounded variation càd-làg functions (i.e. works for dN_t, $\lambda(t)dt$ or $dN_t - \lambda((t)dt)$

$$f(t)g(t) = f(0)g(0) + \int_0^t f(s)dg_s + \int_0^t g(s^-)df_s$$

Hence, with $f(t) = g(t) = \int_0^t H_s[dN_s - \lambda(s)ds]$,

$$\left(\int_0^t H_s[dN_s - \lambda(s)ds]\right)^2 =$$

$$\int_0^t \left(\int_0^s H_u[dN_u - \lambda(u)du]\right) H_s[dN_s - \lambda(s)ds] + \int_0^t \left(\int_0^{s-} H_u[dN_u - \lambda(u)du]\right) H_s[dN_s - \lambda(s)ds]$$

Bracket

A Stieljes integration by part : whatever the bounded variation càd-làg functions (i.e. works for dN_t, $\lambda(t)dt$ or $dN_t - \lambda((t)dt)$

$$f(t)g(t) = f(0)g(0) + \int_0^t f(s)dg_s + \int_0^t g(s^-)df_s$$

Hence, with $f(t) = g(t) = \int_0^t H_s[dN_s - \lambda(s)ds]$,

$$\left(\int_0^t H_s[dN_s - \lambda(s)ds]\right)^2 =$$

$$\int_0^t \left(\int_0^s H_u[dN_u - \lambda(u)du]\right) H_s[dN_s - \lambda(s)ds] + \int_0^t \left(\int_0^{s-} H_u[dN_u - \lambda(u)du]\right) H_s[dN_s - \lambda(s)ds]$$

$$= \int_0^t H_s^2 dN_s + \text{ martingale}$$

$\rightarrow$ compensator ($\simeq$ variance given the past) $V_t = \int_0^t H_s^2 \lambda(s)ds$

predictable and $\left(\int_0^t H_s[dN_s - \lambda(s)ds]\right)^2 - V_t$ martingale

Exponential martingale

Aim : the martingale equivalent of Campbell's theorem.

Exponential martingale

Aim : the martingale equivalent of Campbell's theorem.

$$E_t = \exp\left(\int_0^t H_s[dN_s - \lambda(s)ds] - \int_0^t (e^{H_s} - H_s - 1)\lambda(s)ds\right)$$

is the unique solution of

$$E_t = E_0 + \int_0^t E_{s-}(e^{H_s} - 1)[dN_s - \lambda(s)ds].$$

Exponential martingale

Aim : the martingale equivalent of Campbell's theorem.

$$E_t = \exp\left(\int_0^t H_s[dN_s - \lambda(s)ds] - \int_0^t (e^{H_s} - H_s - 1)\lambda(s)ds\right)$$

is the unique solution of

$$E_t = E_0 + \int_0^t E_{s-}(e^{H_s} - 1)[dN_s - \lambda(s)ds].$$

Hence martingale and $\mathbb{E}(E_t) = E_0 = 1$.
NB : eventually, integrability problems, so $\mathbb{E}(E_t) \leq 1$...

Exponential inequality à la Bernstein

$H_s \to \theta H_s$ and the corresponding $E_t \to E$ in the Poisson Bernstein proof :

for all $\theta, y > 0$

$$\mathbb{P}\left(\int_0^t H_s [dN_s - \lambda(s)ds] \geq \frac{\theta V_t}{2\left(1 - \frac{\theta \|H\|_\infty}{3}\right)} + y/\theta \right) \leq e^{-y}$$

Exponential inequality à la Bernstein

$H_s \to \theta H_s$ and the corresponding $E_t \to E$ in the Poisson Bernstein proof :

for all $\theta, y > 0$

$$\mathbb{P}\left(\int_0^t H_s[dN_s - \lambda(s)ds] \geq \frac{\theta V_t}{2\left(1 - \frac{\theta\|H\|_\infty}{3}\right)} + y/\theta\right) \leq e^{-y}$$

- Assuming $\|H\|_\infty \leq b$ deterministic and known, not a big deal since second order term,
- but assuming $V_t \leq v$ and replacing V_t by v deterministic is not sharp at all !
- However optimising in θ needs a θ deterministic (because ultimately E_t depends on θ)

Slicing

Let $\mathcal{S} = \{w \leq V_t \leq v = (1+\epsilon)^K w\}$ and consider the slice

$$\mathcal{S}_k = \{(1+\epsilon)^k w \leq V_t \leq (1+\epsilon)^{k+1} w)\}$$

Slicing

Let $\mathcal{S} = \{w \leq V_t \leq v = (1+\epsilon)^K w\}$ and consider the slice

$$\mathcal{S}_k = \{(1+\epsilon)^k w \leq V_t \leq (1+\epsilon)^{k+1} w)\}$$

Then

$$\mathbb{P}\left(\int_0^t H_s[dN_s - \lambda(s)ds] \geq \frac{\theta V_t}{2\left(1 - \frac{\theta b}{3}\right)} + y/\theta \text{ and } \mathcal{S}_k \right) \leq e^{-y}$$

Slicing

Let $\mathcal{S} = \{w \leq V_t \leq v = (1+\epsilon)^K w\}$ and consider the slice

$$\mathcal{S}_k = \{(1+\epsilon)^k w \leq V_t \leq (1+\epsilon)^{k+1} w)\}$$

Then

$$\mathbb{P}\left(\int_0^t H_s[dN_s - \lambda(s)ds] \geq \frac{\theta V_t}{2\left(1-\frac{\theta b}{3}\right)} + y/\theta \text{ and } \mathcal{S}_k\right) \leq e^{-y}$$

$$\mathbb{P}\left(\int_0^t H_s[dN_s - \lambda(s)ds] \geq \frac{\theta(1+\epsilon)^{k+1} w}{2\left(1-\frac{\theta b}{3}\right)} + y/\theta \text{ and } \mathcal{S}_k\right) \leq e^{-y}$$

Slicing

Let $\mathcal{S} = \{w \leq V_t \leq v = (1+\epsilon)^K w\}$ and consider the slice

$$\mathcal{S}_k = \{(1+\epsilon)^k w \leq V_t \leq (1+\epsilon)^{k+1} w)\}$$

Then

$$\mathbb{P}\left(\int_0^t H_s[dN_s - \lambda(s)ds] \geq \frac{\theta V_t}{2\left(1-\frac{\theta b}{3}\right)} + y/\theta \text{ and } \mathcal{S}_k\right) \leq e^{-y}$$

$$\mathbb{P}\left(\int_0^t H_s[dN_s - \lambda(s)ds] \geq \frac{\theta(1+\epsilon)^{k+1} w}{2\left(1-\frac{\theta b}{3}\right)} + y/\theta \text{ and } \mathcal{S}_k\right) \leq e^{-y}$$

Here choose $\theta = g((1+\epsilon)^{k+1} w, b, y)$ optimising ...

$$\mathbb{P}\left(\int_0^t H_s[dN_s - \lambda(s)ds] \geq \sqrt{2(1+\epsilon)^{k+1} wy} + \frac{by}{3} \text{ and } \mathcal{S}_k\right) \leq e^{-y}$$

Slicing

Let $\mathcal{S} = \{w \leq V_t \leq v = (1+\epsilon)^K w\}$ and consider the slice

$$\mathcal{S}_k = \{(1+\epsilon)^k w \leq V_t \leq (1+\epsilon)^{k+1} w)\}$$

Then

$$\mathbb{P}\left(\int_0^t H_s[dN_s - \lambda(s)ds] \geq \frac{\theta V_t}{2\left(1-\frac{\theta b}{3}\right)} + y/\theta \text{ and } \mathcal{S}_k\right) \leq e^{-y}$$

$$\mathbb{P}\left(\int_0^t H_s[dN_s - \lambda(s)ds] \geq \frac{\theta(1+\epsilon)^{k+1} w}{2\left(1-\frac{\theta b}{3}\right)} + y/\theta \text{ and } \mathcal{S}_k\right) \leq e^{-y}$$

Here choose $\theta = g((1+\epsilon)^{k+1} w, b, y)$ optimising ...

$$\mathbb{P}\left(\int_0^t H_s[dN_s - \lambda(s)ds] \geq \sqrt{2(1+\epsilon)^{k+1} wy} + \frac{by}{3} \text{ and } \mathcal{S}_k\right) \leq e^{-y}$$

$$\mathbb{P}\left(\int_0^t H_s[dN_s - \lambda(s)ds] \geq \sqrt{2(1+\epsilon) V_t y} + \frac{by}{3} \text{ and } \mathcal{S}_k\right) \leq e^{-y}$$

Almost there

Grouping the slices

$$\mathbb{P}\left(\int_0^t H_s[dN_s - \lambda(s)ds] \geq \sqrt{2(1+\epsilon)V_t y} + \frac{by}{3} \text{ and } \mathcal{S}\right) \leq Ke^{-y}$$

Almost there

Grouping the slices

$$\mathbb{P}\left(\int_0^t H_s[dN_s - \lambda(s)ds] \geq \sqrt{2(1+\epsilon)V_t y} + \frac{by}{3} \text{ and } \mathcal{S}\right) \leq Ke^{-y}$$

But $V_t = \int_0^t H_s^2 \lambda(s)ds$, still depends on the unknown λ

Almost there

Grouping the slices

$$\mathbb{P}\left(\int_0^t H_s[dN_s - \lambda(s)ds] \geq \sqrt{2(1+\epsilon)V_t y} + \frac{by}{3} \text{ and } \mathcal{S}\right) \leq Ke^{-y}$$

But $V_t = \int_0^t H_s^2 \lambda(s)ds$, still depends on the unknown λ
$\rightarrow$ one more turn using the fact that

$$\hat{V}_t = \int_0^t H_s^2 dN_s$$

is observable and

$$\hat{V}_t - V_t = \text{martingale}$$

Bernstein-type inequality for general counting process (Hansen, RB, Rivoirard)

Let $(H_s)_{s\geq 0}$ be a predictable process and $M_t = \int_0^t H_s(dN_s - \lambda(s)ds)$.
Let $b > 0$ and $v > w > 0$. For all $x, \mu > 0$ such that $\mu > \phi(\mu)$, let

$$\hat{V}_\tau^\mu = \frac{\mu}{\mu - \phi(\mu)} \int_0^\tau H_s^2 dN_s + \frac{b^2 x}{\mu - \phi(\mu)},$$

where $\phi(u) = \exp(u) - u - 1$.
Then for every stopping time τ and every $\varepsilon > 0$

$$\mathbb{P}\left(M_\tau \geq \sqrt{2(1+\varepsilon)\hat{V}_\tau^\mu x} + bx/3, \quad w \leq \hat{V}_\tau^\mu \leq v \text{ and } \sup_{s\in[0,\tau]} |H_s| \leq b \right) \leq 2\frac{\log(v/w)}{\log(1+\varepsilon)} e^{-x}.$$

$\rightsquigarrow$ a generic choice of $\mathbf{d}$ for the Lasso whatever the underlying process....

Back to Lasso for Hawkes

In the Hawkes case, $H_s \to \mathbf{Rc}_s$, renormalized counts in small intervals.

- Hence there is no absolute v or b.

Back to Lasso for Hawkes

In the Hawkes case, $H_s \to \mathbf{Rc}_s$, renormalized counts in small intervals.

- Hence there is no absolute v or b.
- If controlled number of points in small intervals via exponential inequality, then can find v and b such that $\mathbb{P}(V_t \geq v \text{ or } \|H\|_\infty \geq b)$ exponentially small

Back to Lasso for Hawkes

In the Hawkes case, $H_s \to \mathbf{Rc}_s$, renormalized counts in small intervals.

- Hence there is no absolute v or b.
- If controlled number of points in small intervals via exponential inequality, then can find v and b such that $\mathbb{P}(\hat{V}_t \geq v \text{ or } \|H\|_\infty \geq b)$ exponentially small
- Use to stop the martingale before $\hat{V}_t$ reaches level v (same for b)
- one can always choose $w = \frac{b^2 x}{\mu - \phi(\mu)}$, nice since cannot stop the martingale on this side.

Back to Lasso for Hawkes

In the Hawkes case, $H_s \to \mathbf{Rc}_s$, renormalized counts in small intervals.

- Hence there is no absolute v or b.
- If controlled number of points in small intervals via exponential inequality, then can find v and b such that $\mathbb{P}(\hat{V}_{\mu t} \geq v \text{ or } \|H\|_\infty \geq b)$ exponentially small
- Use to stop the martingale before $\hat{V}_{\mu t}$ reaches level v (same for b)
- one can always choose $w = \frac{b^2 x}{\mu - \phi(\mu)}$, nice since cannot stop the martingale on this side.

Lasso criterion

$$\begin{aligned}\hat{\mathbf{a}}^{(r)} &= \operatorname{argmin}_{\mathbf{a}}\{\gamma_T(\lambda_{\mathbf{a}}^{(r)}) + \mathrm{pen}(\mathbf{a})\} \\ &= \operatorname{argmin}_{\mathbf{a}}\{-2\mathbf{a}'\mathbf{b}_r + \mathbf{a}'\mathbf{G}\mathbf{a} + 2(\mathbf{d}^{(r)})'|\mathbf{a}|\}\end{aligned}$$

M number of interacting processes, K number of bins, A maximal interaction delay.

Oracle inequality in probability

If $\mathbf{d} = \sqrt{2(1+\varepsilon)\hat{V}_T^{\mu} x} + bx/3$ then with probability larger than

$$1 - \square M^2 K \log(T)^2 e^{-x} - \mathbb{P}(\forall t < T, i, N^i_{[t-A,t]} > \square \log(T)^2) - \mathbb{P}(\mathbf{G} \not\geq cI)$$

Lasso criterion

$$\begin{aligned}\hat{\mathbf{a}}^{(r)} &= \operatorname{argmin}_{\mathbf{a}}\{\gamma_T(\lambda_{\mathbf{a}}^{(r)}) + \mathrm{pen}(\mathbf{a})\} \\ &= \operatorname{argmin}_{\mathbf{a}}\{-2\mathbf{a}'\mathbf{b}_r + \mathbf{a}'\mathbf{G}\mathbf{a} + 2(\mathbf{d}^{(r)})'|\mathbf{a}|\}\end{aligned}$$

M number of interacting processes, K number of bins, A maximal interaction delay.

Oracle inequality in probability

If $\mathbf{d} = \sqrt{2(1+\varepsilon)\hat{V}_T^{\mu} x} + bx/3$ then with probability larger than

$$1 - \square M^2 K \log(T)^2 e^{-x} - \mathbb{P}(\forall t < T, i, N^i_{[t-A,t]} > \square \log(T)^2) - \mathbb{P}(\mathbf{G} \not\geq cI)$$

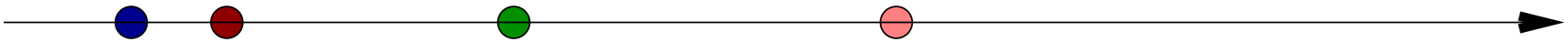
$$\sum_r \|\lambda^{(r)} - \mathbf{R}\mathbf{c}_t\hat{\mathbf{a}}^{(r)}\|^2 \leq \square \inf_{\mathbf{a}} \left\{ \sum_r \|\lambda^{(r)} - \mathbf{R}\mathbf{c}_t\mathbf{a}\|^2 + \frac{1}{c} \sum_{i \in supp(\mathbf{a})} (d_i^{(r)})^2 \right\}.$$

Mathematical debts

- Control of the number of points per interval
- Control of c the smallest eigenvalue of $\mathbf{G}$
- Choice of x.

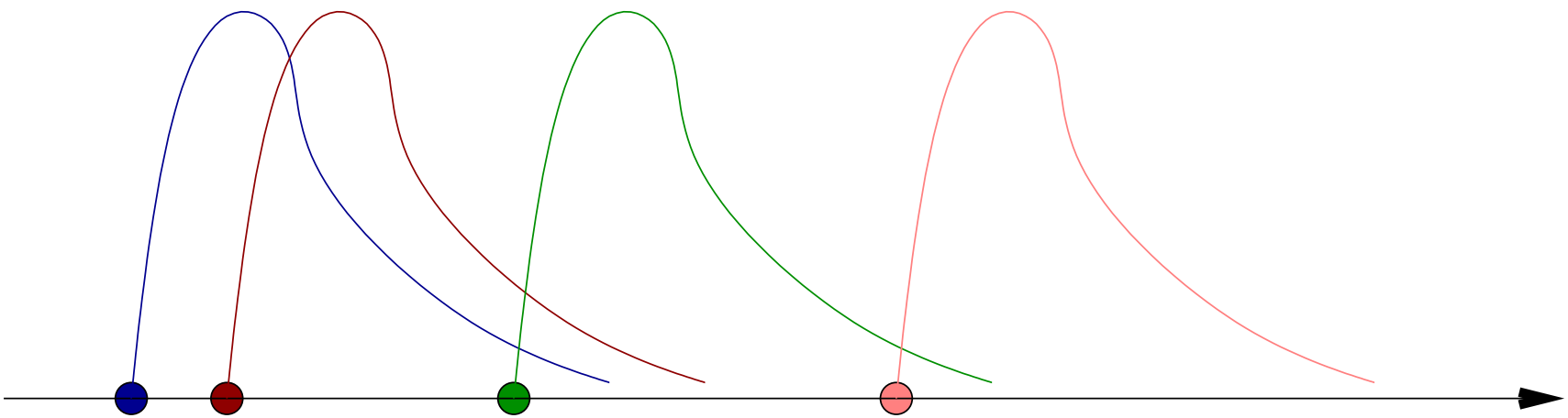
Table of Contents

A branching representation of the linear Hawkes model (univariate)



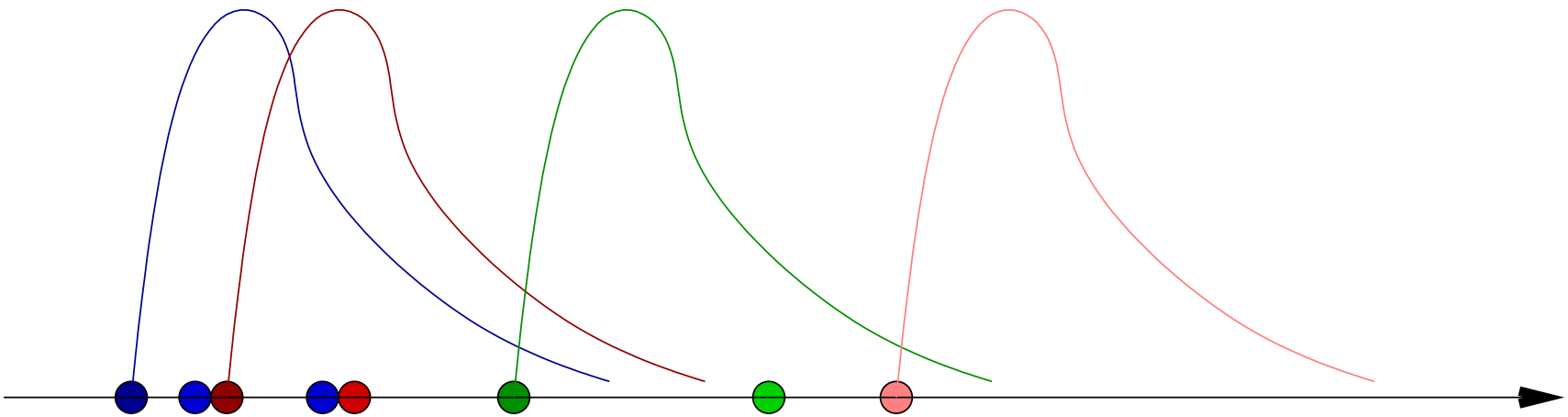
- Linear Hawkes process= Branching process on a Poisson process
- Start = homogeneous Poisson (ν) = ancestors

A branching representation of the linear Hawkes model (univariate)



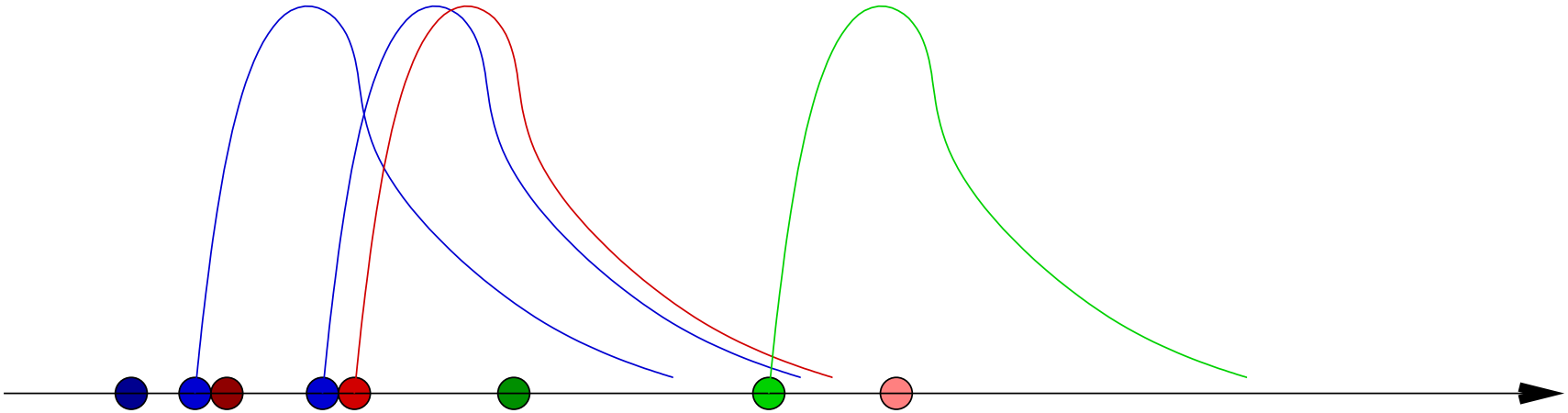
- Linear Hawkes process= Branching process on a Poisson process
- Start = homogeneous Poisson (ν) = ancestors
- Each point generates children according to a Poisson process of intensity h

A branching representation of the linear Hawkes model (univariate)



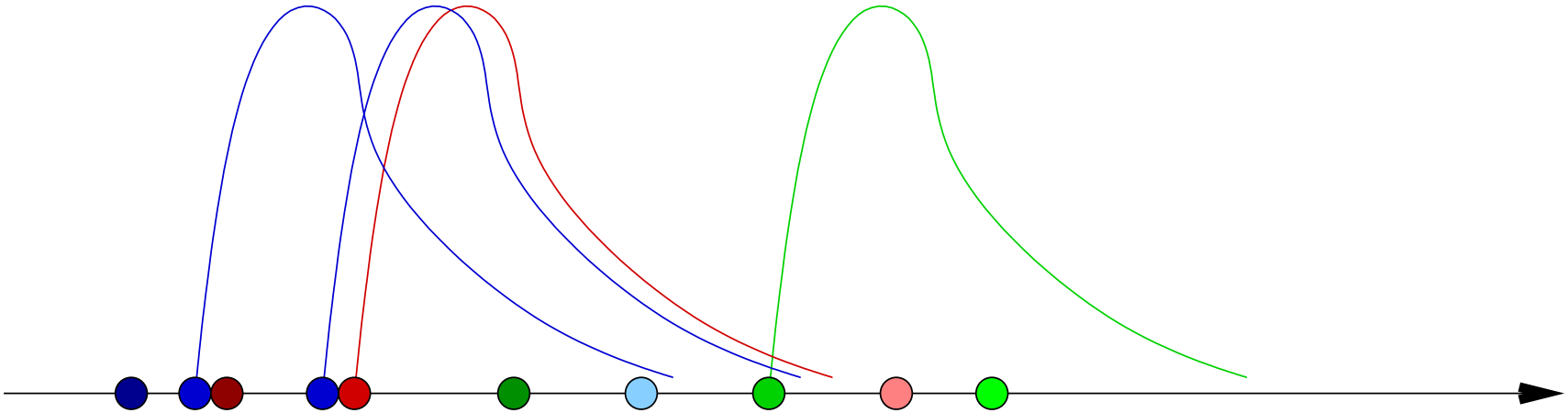
- Linear Hawkes process= Branching process on a Poisson process
- Start = homogeneous Poisson (ν) = ancestors
- Each point generates children according to a Poisson process of intensity h

A branching representation of the linear Hawkes model (univariate)



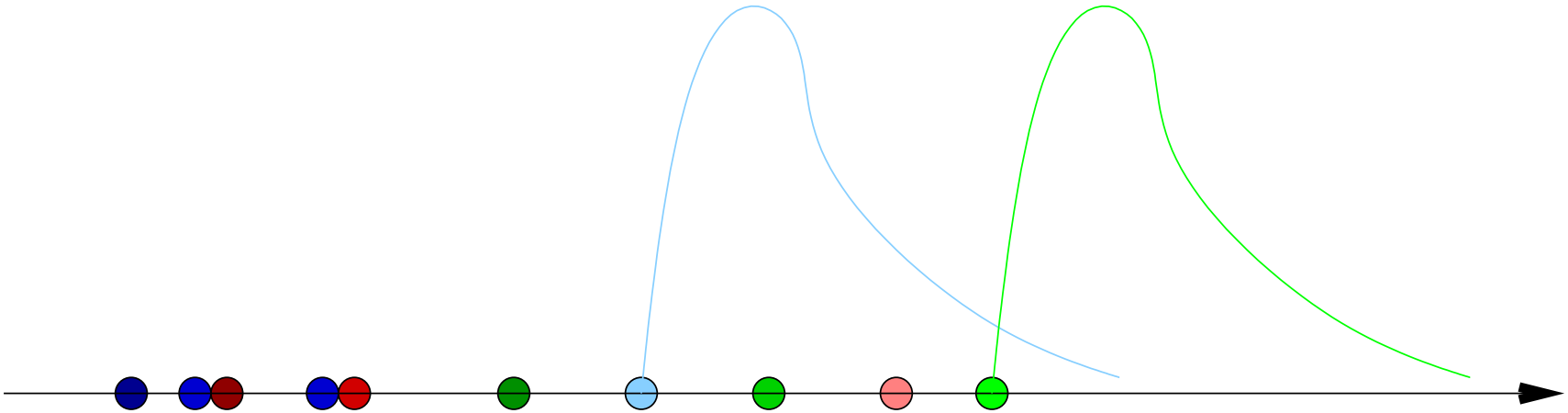
- Linear Hawkes process= Branching process on a Poisson process
- Start = homogeneous Poisson (ν) = ancestors
- Each point generates children according to a Poisson process of intensity h
- Again and again until extinction (almost sure when $\int h < 1$)

A branching representation of the linear Hawkes model (univariate)



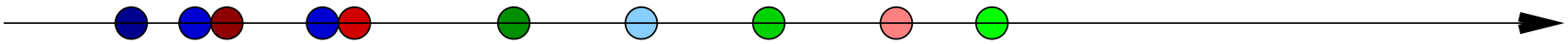
- Linear Hawkes process= Branching process on a Poisson process
- Start = homogeneous Poisson (ν) = ancestors
- Each point generates children according to a Poisson process of intensity h
- Again and again until extinction (almost sure when $\int h < 1$)

A branching representation of the linear Hawkes model (univariate)



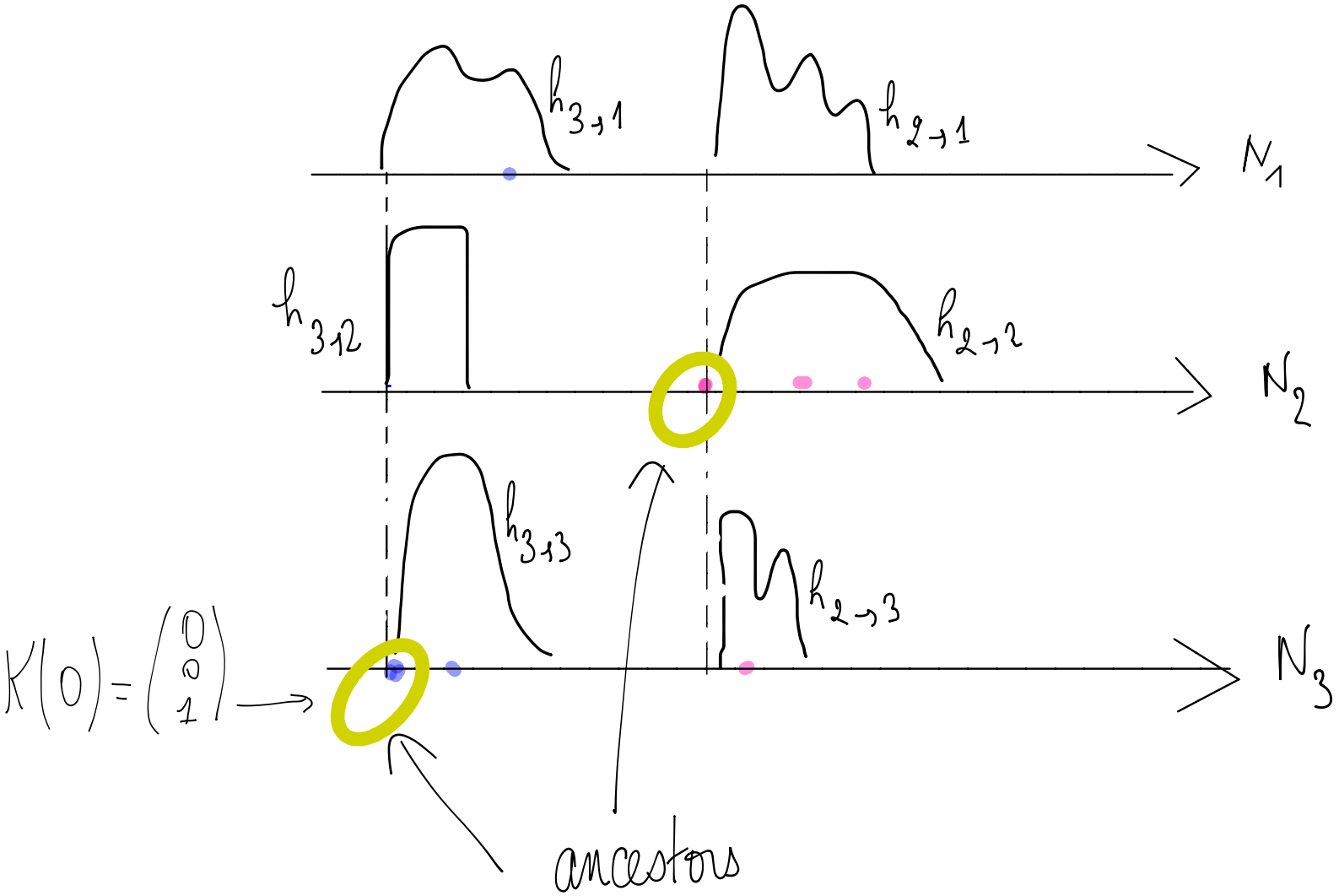
- Linear Hawkes process= Branching process on a Poisson process
- Start = homogeneous Poisson (ν) = ancestors
- Each point generates children according to a Poisson process of intensity h
- Again and again until extinction (almost sure when $\int h < 1$)

A branching representation of the linear Hawkes model (univariate)

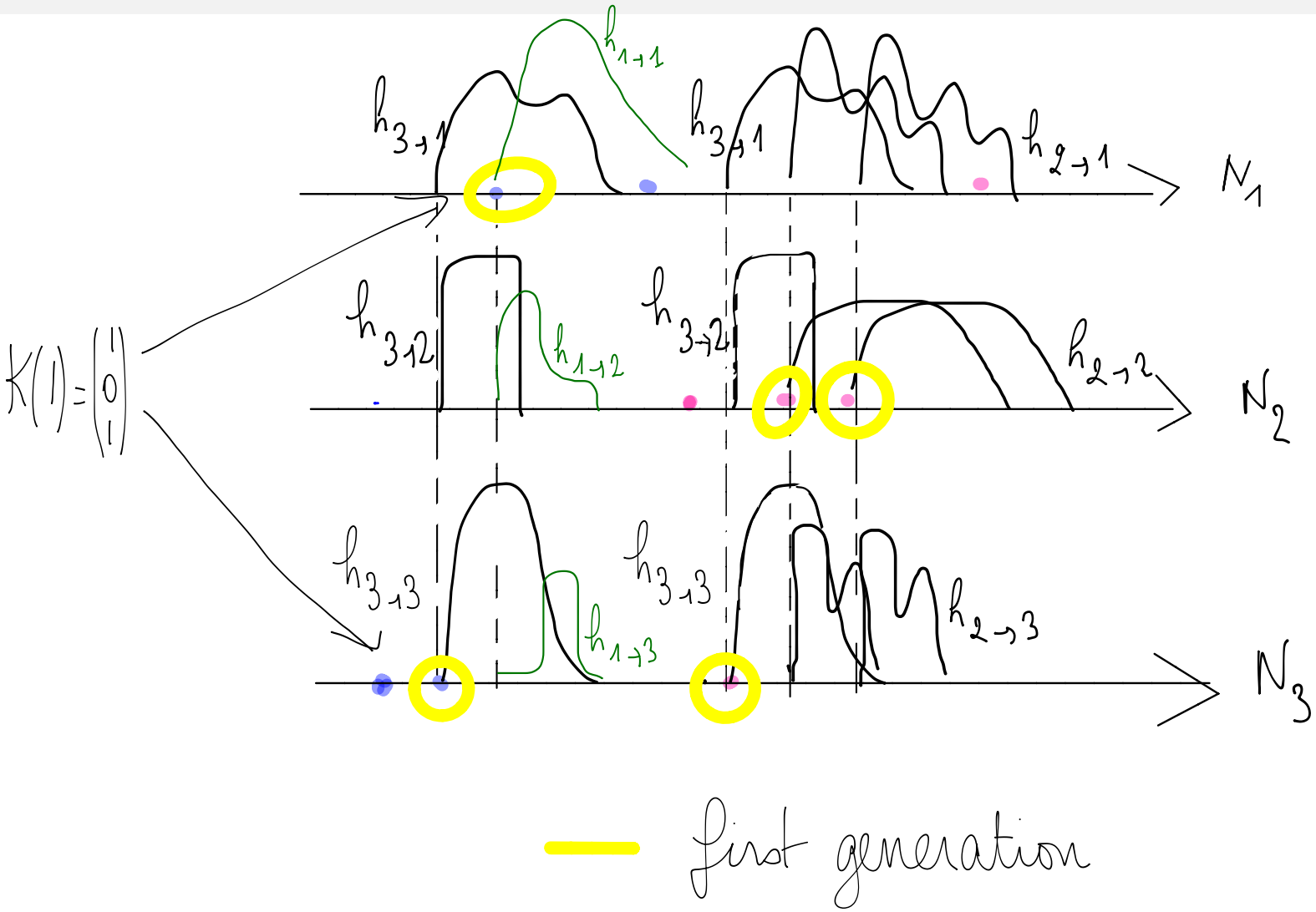


- Linear Hawkes process= Branching process on a Poisson process
- Start = homogeneous Poisson (ν) = ancestors
- Each point generates children according to a Poisson process of intensity h
- Again and again until extinction (almost sure when $\int h < 1$)
- Hawkes= Final process without colors

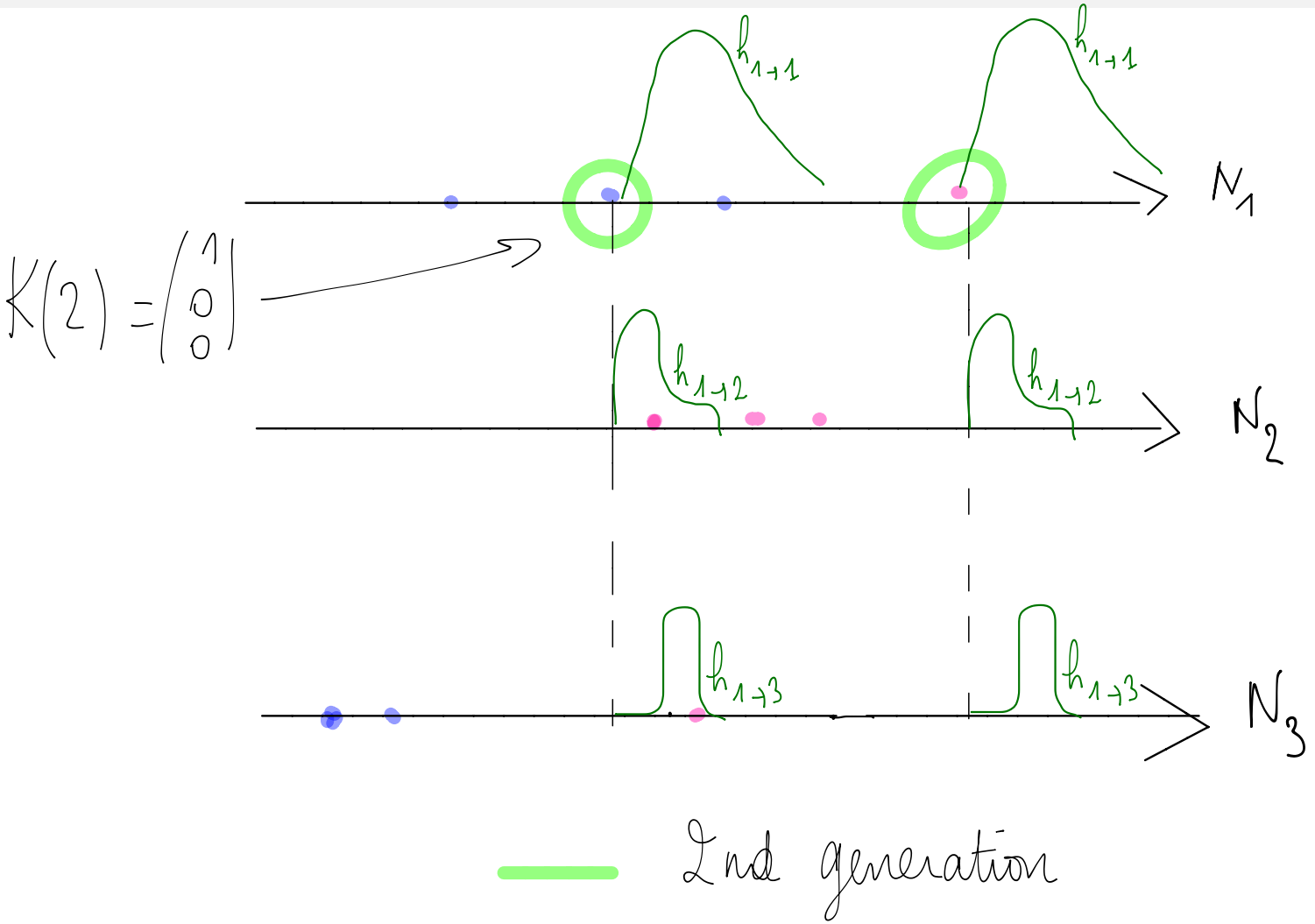
A branching representation of the linear Hawkes model (multivariate)



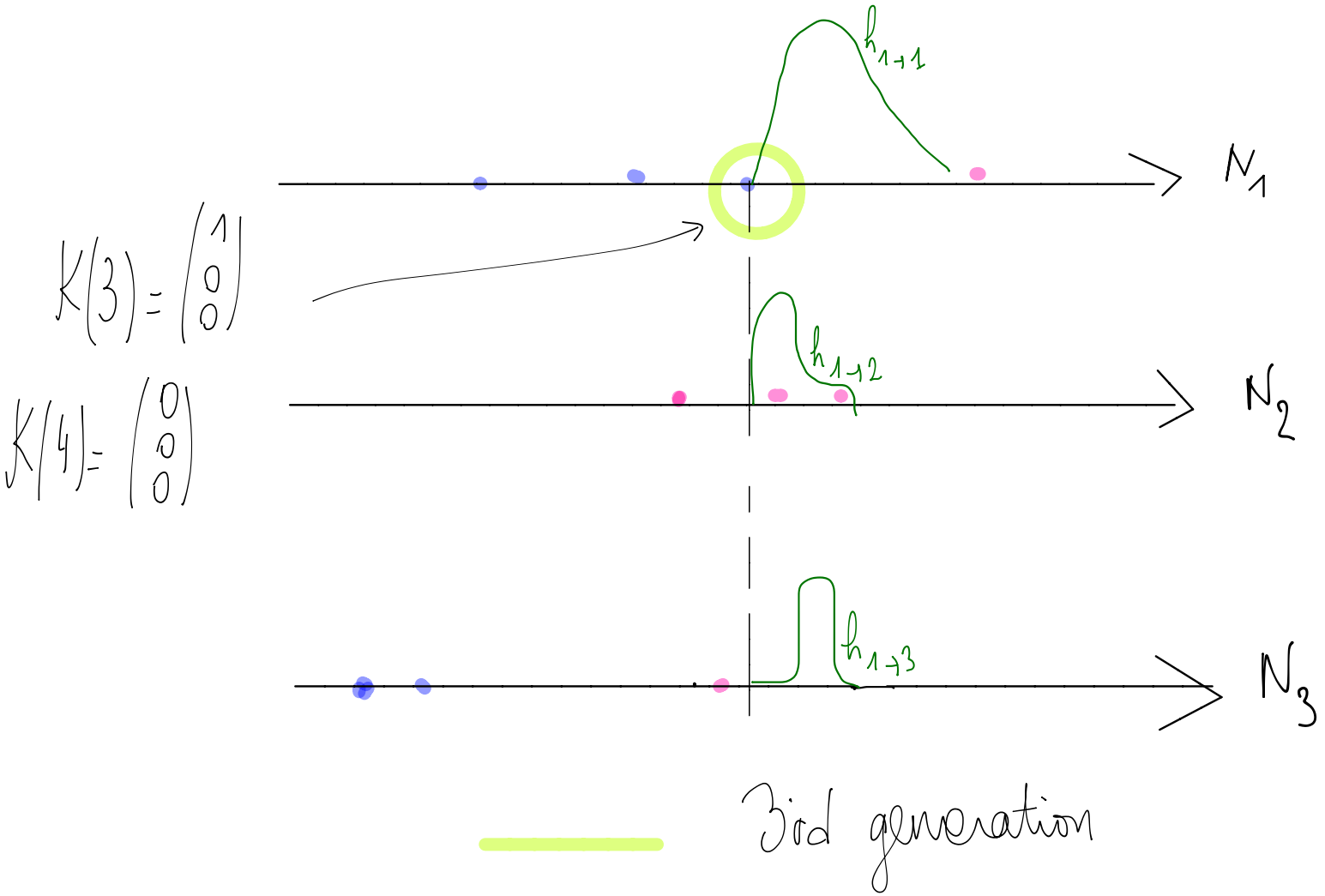
A branching representation of the linear Hawkes model (multivariate)



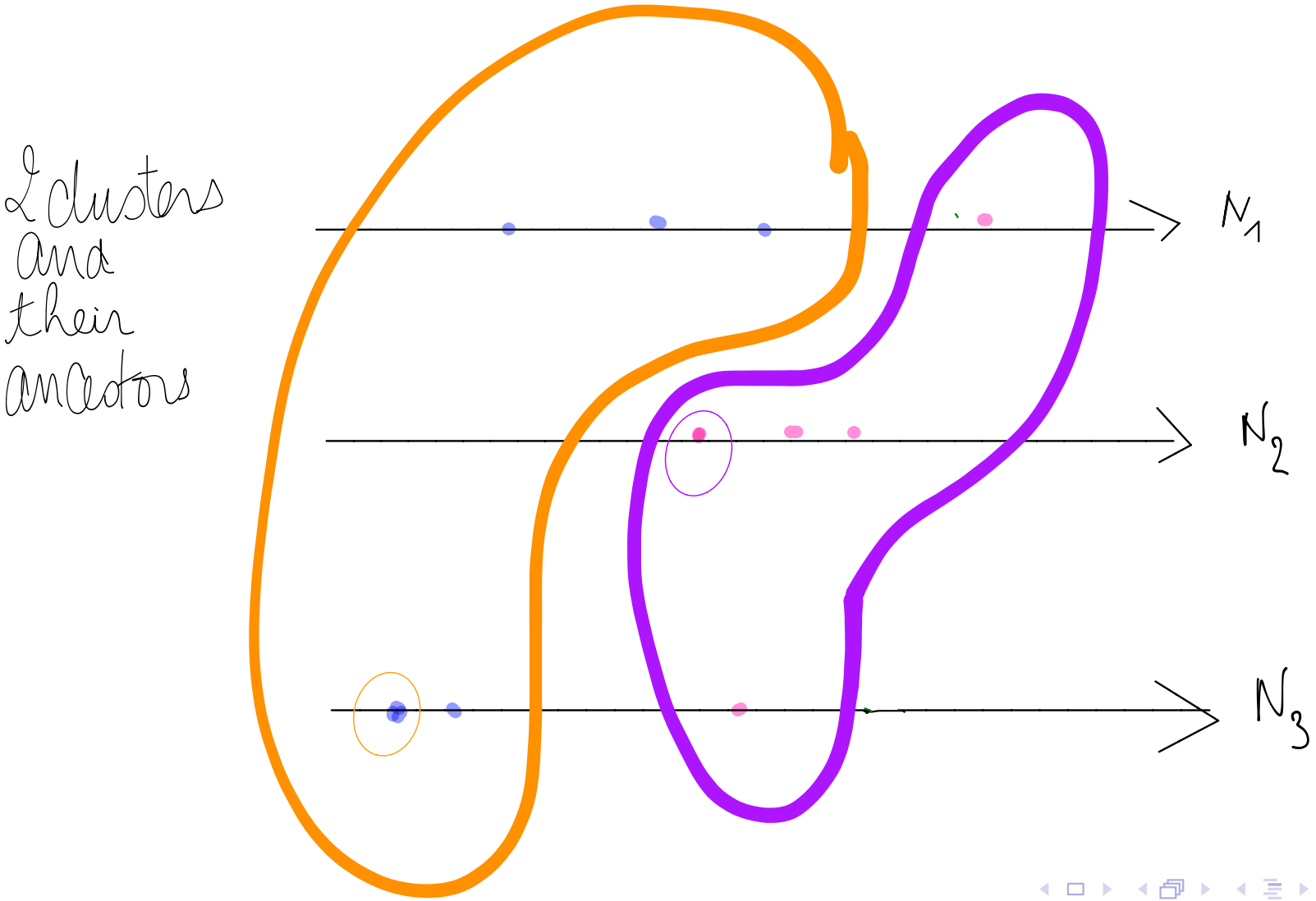
A branching representation of the linear Hawkes model (multivariate)



A branching representation of the linear Hawkes model (multivariate)



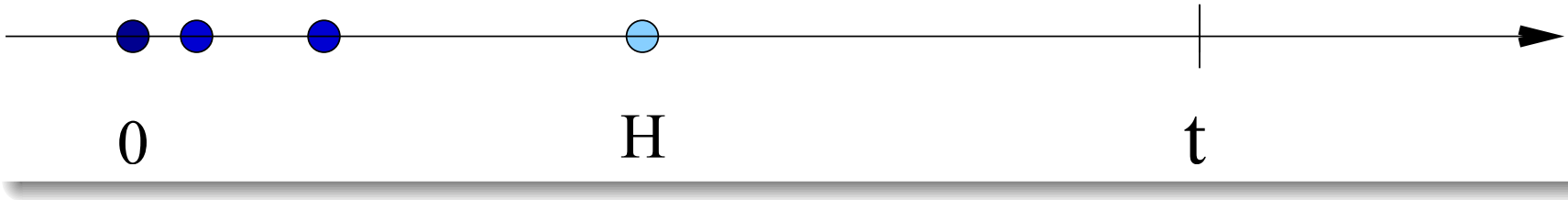
A branching representation of the linear Hawkes model (multivariate)



Univariate and multivariate clusters

Cluster process

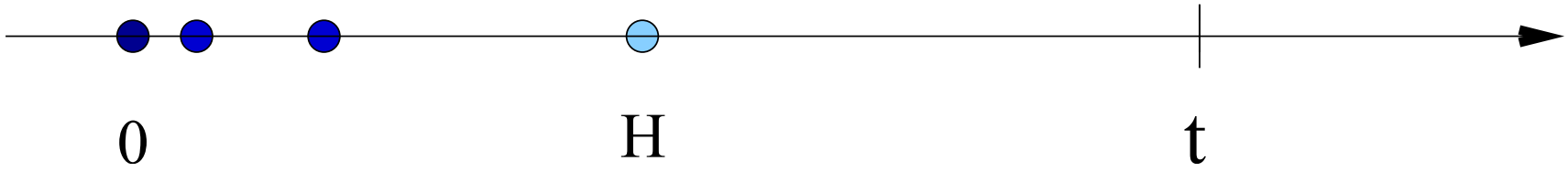
ancestor in 0 + all its family



Univariate and multivariate clusters

Cluster process

ancestor in 0 + all its family



If multivariate, the distribution of the cluster depends on the type of the ancestor.

Number of points of one cluster

If ancestor of type $\ell \to \mathbb{E}_\ell$.

- $K(n)$ vector of the number of points per type in the nth generation
- If ancestor of type ℓ, $K(0) = e_\ell$

Number of points of one cluster

If ancestor of type $\ell \to \mathbb{E}_\ell$.

- $K(n)$ vector of the number of points per type in the nth generation
- If ancestor of type ℓ, $K(0) = e_\ell$
- $W(n) = \sum_0^n K(\,\cdot\,)$: number of points in the cluster per type up to generation n

Number of points of one cluster

If ancestor of type $\ell \to \mathbb{E}_\ell$.

- $K(n)$ vector of the number of points per type in the nth generation
- If ancestor of type ℓ, $K(0) = e_\ell$
- $W(n) = \sum_0^n K(j)$: number of points in the cluster per type up to generation n
- Has $W = W(\infty)$ a Laplace transform ? i.e. can we find a vector θ of positive coordinates such that

$$\mathbb{E}_\ell \left(e^{\theta' W(\infty)} \right) < \infty$$

Number of points of one cluster

If ancestor of type $\ell \to \mathbb{E}_\ell$.

- $K(n)$ vector of the number of points per type in the nth generation
- If ancestor of type ℓ, $K(0) = e_\ell$
- $W(n) = \sum_0^n K(j)$: number of points in the cluster per type up to generation n
- Has $W = W(\infty)$ a Laplace transform ? i.e. can we find a vector θ of positive coordinates such that

$$\mathbb{E}_\ell \left(e^{\theta' W(\infty)} \right) < \infty$$

- If yes,

$$\mathbb{P}_\ell(N_{tot} \geq x) \leq \mathbb{E}_\ell \left(e^{\min_i(\theta_i) N_{tot}} \right) e^{-\min_i(\theta_i)x} \leq \mathbb{E}_\ell \left(e^{\theta' W(\infty)} \right) e^{-\min_i(\theta_i)x}$$

exponentially small

Laplace transform for the cluster

$$\phi(\theta)' = (\phi_1(\theta), ..., \phi_M(\theta)),$$
$$\phi_\ell(\theta) = \log \mathbb{E}_\ell(e^{\theta' K(1)}).$$

Laplace transform for the cluster

$$\phi(\theta)' = (\phi_1(\theta), ..., \phi_M(\theta)),$$
$$\phi_\ell(\theta) = \log \mathbb{E}_\ell(e^{\theta' K(1)}).$$

$$\mathbb{E}_\ell\left(e^{\theta' W(n)}\right) = \mathbb{E}_\ell\left(e^{\theta' W(n-1)} \mathbb{E}\left[e^{\theta' K(n)} \middle| \text{ generations} < n\right]\right)$$

Laplace transform for the cluster

$$\phi(\theta)' = (\phi_1(\theta), ..., \phi_M(\theta)),$$
$$\phi_\ell(\theta) = \log \mathbb{E}_\ell(e^{\theta' K(1)}).$$

$$\begin{aligned}
\mathbb{E}_\ell\left(e^{\theta' W(n)}\right) &= \mathbb{E}_\ell\left(e^{\theta' W(n-1)} \mathbb{E}\left[e^{\theta' K(n)} \middle| \text{ generations} < n\right]\right) \\
&= \mathbb{E}_\ell\left(e^{\theta' W(n-1)} e^{\phi(\theta)' K(n-1)}\right) \\
&= \exp(u_n(\theta)_\ell)
\end{aligned}$$

with $u_n(\theta) = \theta + \phi(u_{n-1}(\theta))$, $u_0(\theta) = \theta$

Laplace transform for the cluster(2)

If ϕ local contraction for a certain norm $\|.\|$:
there exists $r > 0$ and $C < 1$ st if $\|s\| < r$ then

$$\|\phi(s)\| \leq C\|s\|.$$

Laplace transform for the cluster(2)

If ϕ local contraction for a certain norm $\|.\|$:
there exists $r > 0$ and $C < 1$ st if $\|s\| < r$ then

$$\|\phi(s)\| \leq C\|s\|.$$

If $\|\theta\| \leq r(1 - C)$,

$$\begin{aligned} \|u_n(\theta)\| &\leq \|\theta\| + \|\phi(u_{n-1}(\theta))\| \\ &\leq \|\theta\| + C\|u_{n-1}(\theta)\| \end{aligned}$$

Laplace transform for the cluster(2)

If ϕ local contraction for a certain norm $\|.\|$:
there exists $r > 0$ and $C < 1$ st if $\|s\| < r$ then

$$\|\phi(s)\| \leq C\|s\|.$$

If $\|\theta\| \leq r(1-C)$,

$$\begin{aligned}
\|u_n(\theta)\| &\leq \|\theta\| + \|\phi(u_{n-1}(\theta))\| \\
&\leq \|\theta\| + C\|u_{n-1}(\theta)\| \\
&\leq \|\theta\|(1 + C + ... + C^n) \\
&\leq \frac{\|\theta\|}{1-C} \leq r
\end{aligned}$$

Hence each coordinate of $u_n(\theta)$ increases (since $W(n)$ increases) but remains in a compact set. Therefore it converges and for θ small enough, $W(\infty)$ has a Laplace transform.

Contraction ?

$K(1)$ has a Laplace transform (only thing needed $\Longleftarrow$ it's Poisson).

$$\phi_\ell(\theta) = \log \mathbb{E}_\ell(e^{\theta' K(1)}).$$

Hence $\partial\phi(0) = \Gamma = (\int h_\ell^{(m)})_{\ell,m}$ with spectral radius < 1.

Contraction ?

$K(1)$ has a Laplace transform (only thing needed $\Longleftarrow$ it's Poisson).

$$\phi_\ell(\theta) = \log \mathbb{E}_\ell(e^{\theta' K(1)}).$$

Hence $\partial\phi(0) = \Gamma = (\int h_\ell^{(m)})_{\ell,m}$ with spectral radius < 1.

Householder theorem

there exists a norm on $\mathbb{R}^M$ s.t. the associated operator norm of $\Gamma < 1$

Contraction ?

$K(1)$ has a Laplace transform (only thing needed $\Longleftarrow$ it's Poisson).

$$\phi_\ell(\theta) = \log \mathbb{E}_\ell(e^{\theta' K(1)}).$$

Hence $\partial\phi(0) = \Gamma = (\int h_\ell^{(m)})_{\ell,m}$ with spectral radius < 1.

Householder theorem

there exists a norm on $\mathbb{R}^M$ s.t. the associated operator norm of $\Gamma < 1$

By continuity (ϕ infinitely differentiable in a neighborhood of 0),

$$\exists c \in (0,1), \xi > 0, \quad \forall \|s\| < \xi, \quad \|\partial\phi(s)\| \leq c.$$

Contraction ?

$K(1)$ has a Laplace transform (only thing needed $\Longleftarrow$ it's Poisson).

$$\phi_\ell(\theta) = \log \mathbb{E}_\ell(e^{\theta' K(1)}).$$

Hence $\partial\phi(0) = \Gamma = (\int h_\ell^{(m)})_{\ell,m}$ with spectral radius < 1.

Householder theorem

there exists a norm on $\mathbb{R}^M$ s.t. the associated operator norm of $\Gamma < 1$

By continuity (ϕ infinitely differentiable in a neighborhood of 0),

$$\exists c \in (0,1), \xi > 0, \quad \forall \|s\| < \xi, \quad \|\partial\phi(s)\| \leq c.$$

Since $\phi(0) = 0$, by continuity

$$\exists C \in (0,1), r > 0, \quad \forall \|s\| < r, \quad \|\phi(s)\| \leq C\|s\|.$$

Number of points for Hawkes processes per interval

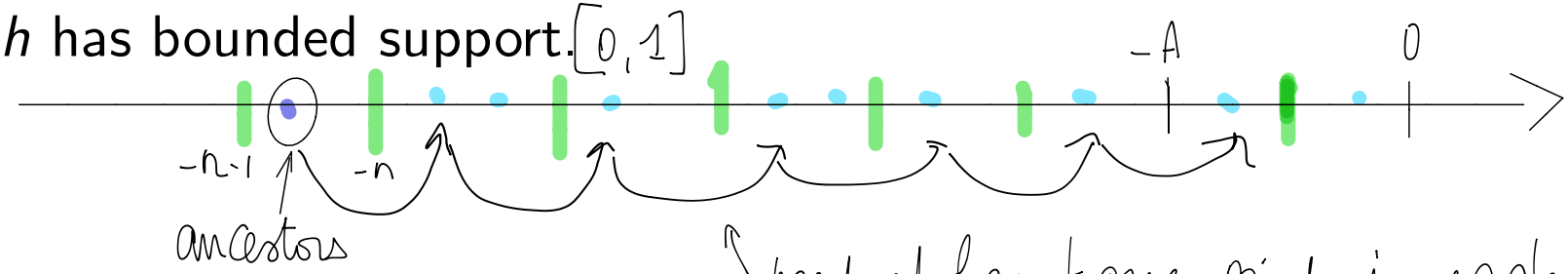
Assume

- stationary (spectral radius of $\Gamma < 1$)
- h has bounded support.

Number of points for Hawkes processes per interval

Assume

- stationary (spectral radius of $\Gamma < 1$)
- h has bounded support. $[0,1]$



$\tilde{N}_{A,\ell}$ number of points in $[-A, 0)$ whose ancestor is of type ℓ

number of ancestors of type ℓ in $[-n-1, n]$

$$\tilde{N}_{A,\ell} \leq \sum_{n} \sum_{k=1}^{A_n} \max\{W_{n,k} - n + \lceil A \rceil, 0\}$$

$W_{n,k}$: number of points in the cluster of ancestor k

Number of points for Hawkes processes per interval(2)

Let $H_n(\theta_\ell) = \mathbb{E}_\ell \left(e^{\theta_\ell \max\{W-n+\lceil A \rceil, 0\}} \right)$, then

$$\begin{aligned} \mathbb{E}\left(e^{\theta_\ell \tilde{N}_{A,\ell}} \right) &\leq \prod_n \mathbb{E}(H_n(\theta_\ell)^{A_n}) \\ &\leq \prod_n \exp\left(\nu_\ell (H_n(\theta_\ell) - 1)\right) \end{aligned}$$

Since W has a Laplace transform, H_n exponentially decreasing with n and it converges.

Number of points for Hawkes processes per interval(2)

Let $H_n(\theta_\ell) = \mathbb{E}_\ell\left(e^{\theta_\ell \max\{W-n+\lceil A\rceil,0\}}\right)$, then

$$\begin{aligned}\mathbb{E}\left(e^{\theta_\ell \tilde{N}_{A,\ell}}\right) &\leq \prod_n \mathbb{E}(H_n(\theta_\ell)^{A_n}) \\ &\leq \prod_n \exp\left(\nu_\ell(H_n(\theta_\ell)-1)\right)\end{aligned}$$

Since W has a Laplace transform, H_n exponentially decreasing with n and it converges.
$\rightarrow$ Laplace transform of $N_{[-A,0)}$ exists
$\rightarrow \mathbb{P}(N_{[-A,0)} > y)$ exponentially small...
NB : true as soon as the branching distribution has a Laplace,
constant unknown... depends on M ! ! !

The expectation of G

Recall that

$$\eta_{\mathbf{a}}(t) = \mathbf{R}\mathbf{c}_t'\mathbf{a} = \nu + \sum_k a_k \int_{-\infty}^{t} \delta^{-1/2} \mathbf{1}_{[t-(k+1)\delta, t-k\delta]} dN_{t-u}$$

$$\mathbf{a}'\mathbf{G}\mathbf{a} = \int_0^T \eta_{\mathbf{a}}(t)^2 dt$$

The expectation of G

Recall that

$$\eta_{\mathbf{a}}(t) = \mathbf{R}\mathbf{c}_t'\mathbf{a} = \nu + \sum_k a_k \int_{-\infty}^{t} \delta^{-1/2} \mathbf{1}_{[t-(k+1)\delta, t-k\delta]} dN_{t-u}$$

$$\mathbf{a}'\mathbf{G}\mathbf{a} = \int_0^T \eta_{\mathbf{a}}(t)^2 dt$$

Hence

$$\mathbb{E}(\mathbf{G}) \geq cI \iff \forall \mathbf{a}, \mathbb{E}\left(\int_0^T \eta_{\mathbf{a}}(t)^2 dt\right) \geq c\|\mathbf{a}\|^2.$$

The expectation of G

Recall that

$$\eta_{\mathbf{a}}(t) = \mathbf{R}\mathbf{c}_t'\mathbf{a} = \nu + \sum_k a_k \int_{-\infty}^{t} \delta^{-1/2}\mathbf{1}_{[t-(k+1)\delta, t-k\delta]}\, dN_{t-u}$$

$$\mathbf{a}'\mathbf{G}\mathbf{a} = \int_0^T \eta_{\mathbf{a}}(t)^2 dt$$

Hence

$$\mathbb{E}(\mathbf{G}) \geq cI \iff \forall \mathbf{a}, \mathbb{E}\Big(\int_0^T \eta_{\mathbf{a}}(t)^2 dt\Big) \geq c\|\mathbf{a}\|^2.$$

If the process is stationnary $\mathbb{E}(\eta_{\mathbf{a}}(t)^2)$ does not depend on t.
Hence we want to show that $c = T\alpha$ with $\mathbb{E}(\eta_{\mathbf{a}}(0)^2) \geq \alpha\|\mathbf{a}\|^2$.
Then we need to concentrate $\frac{1}{T}\mathbf{G}$ around its expectation.

Minoration of the expectation

- If

$$\eta_{\mathbf{a}}(t) = \nu + \sum_k a_k \int_{-\infty}^{t} \delta^{-1/2} \mathbf{1}_{[t-(k+1)\delta, t-k\delta]} dN_{t-u}$$

and N homogeneous Poisson process, then we can find α s.t. $\mathbb{E}(\eta_{\mathbf{a}}(0)^2) \geq \alpha \|\mathbf{a}\|^2$.

Minoration of the expectation

- If

$$\eta_{\mathbf{a}}(t) = \nu + \sum_k a_k \int_{-\infty}^{t} \delta^{-1/2} \mathbf{1}_{[t-(k+1)\delta, t-k\delta]} dN_{t-u}$$

  and N homogeneous Poisson process, then we can find α s.t. $\mathbb{E}(\eta_{\mathbf{a}}(0)^2) \geq \alpha \|\mathbf{a}\|^2$.
- Use the likelihood and Girsanov theorem, to transfer the minoration for Poisson to another minoration for Hawkes....
- because of the likelihood, need again to have a Laplace transform of the number of points...

Clusters and their size

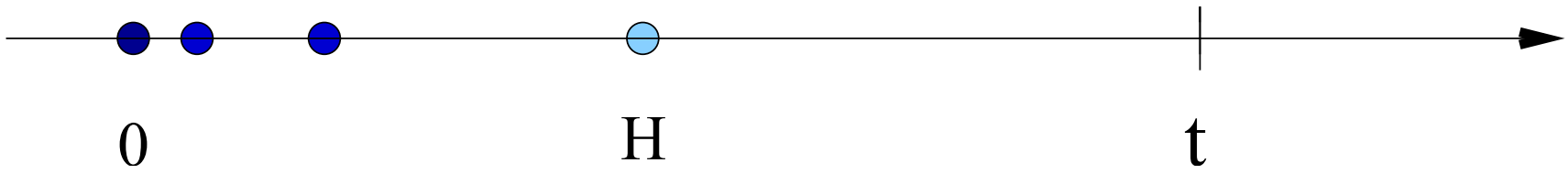
(Univariate)

- Cluster = ancestor in 0 + all its family

Clusters and their size

(Univariate)

- Cluster = ancestor in 0 + all its family

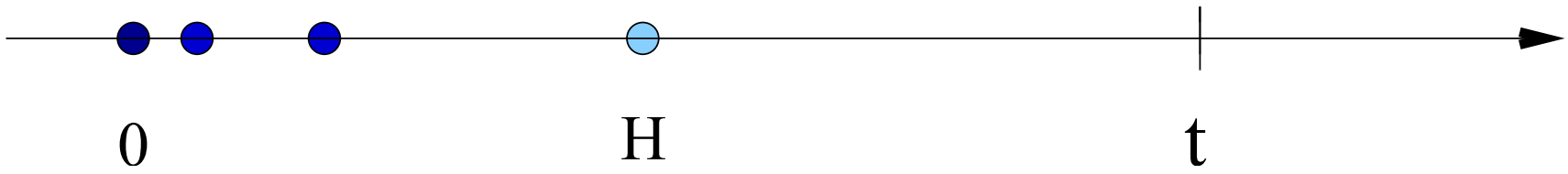


- Size of cluster H $\leq GW * A$ where A maximal support size for h

Clusters and their size

(Univariate)

- Cluster = ancestor in 0 + all its family

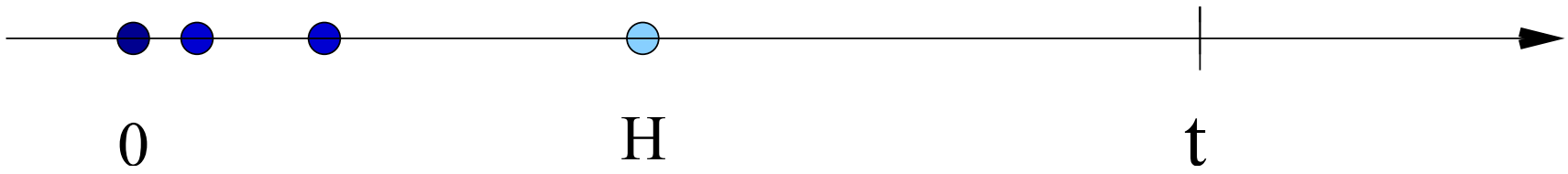


- Size of cluster H $\leq GW * A$ where A maximal support size for h
- and GW total number of births in a Galton-Watson process (Poisson($\int h$)).

Clusters and their size

(Univariate)

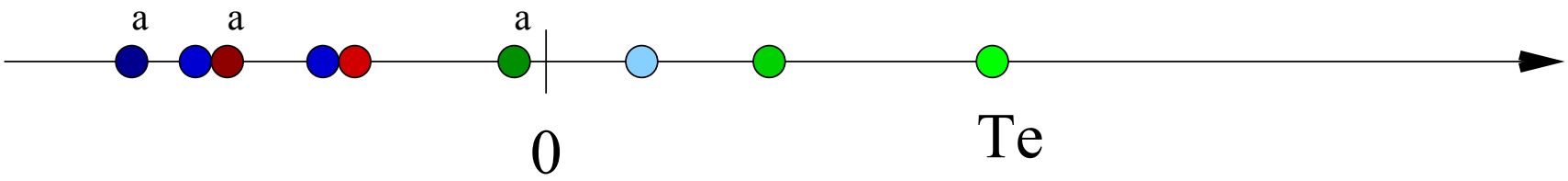
- Cluster = ancestor in 0 + all its family



- Size of cluster H $\leq GW * A$ where A maximal support size for h
- and GW total number of births in a Galton-Watson process (Poisson($\int h$)).
- Hence $\mathbb{P}(\mathrm{H} > t)$ exponentially small

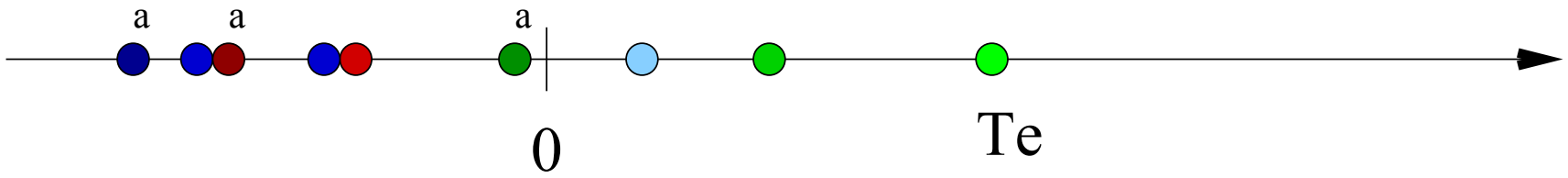
Extinction time

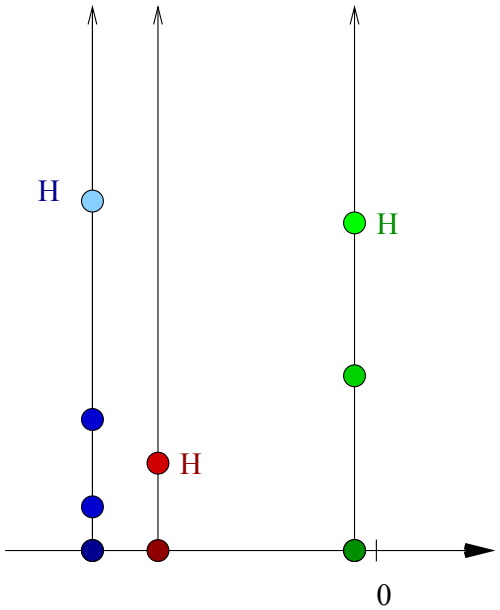
- Extinction time Te = time from 0 until the last birth, if no ancestor after 0



Extinction time

- Extinction time Te = time from 0 until the last birth, if no ancestor after 0





Conditionnally to the ancestor at t,
$P(H > t-a)$ = pty that it reaches a.

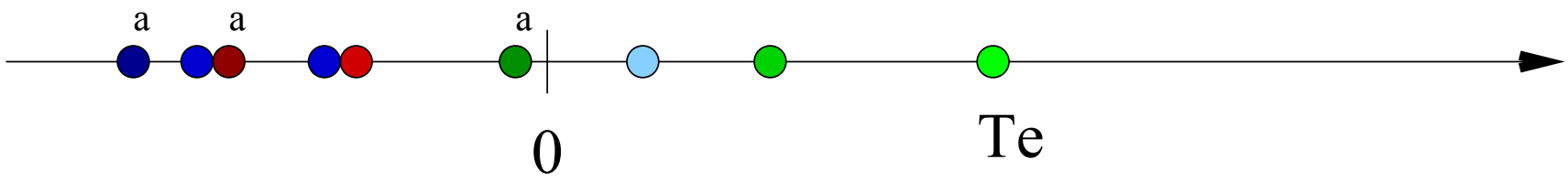
$$\Rightarrow P(T_e \le a \mid \text{ancestors}) = \prod_{anc} P(H > anc - a).$$

$$= \exp \int_{-\infty}^{0} \log P(H > t-a)\, dN_{anc}(t)$$

- Poisson ancestor process before 0 marked by the size of the associated cluster, H

Extinction time

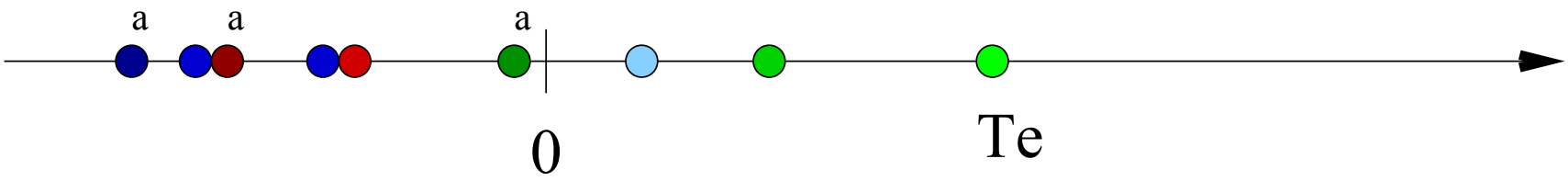
- Extinction time Te = time from 0 until the last birth, if no ancestor after 0



- $\mathbb{P}(\mathsf{Te} \leq a) = \exp\left(-\nu \int_a^{+\infty} \mathbb{P}(\mathsf{H} > t)dt\right)$

Extinction time

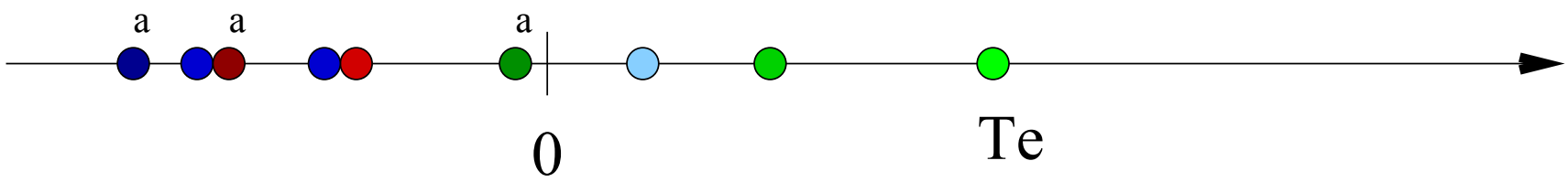
- Extinction time Te = time from 0 until the last birth, if no ancestor after 0



- Hence $\mathbb{P}(\mathrm{Te} > a)$ exponentially small

Extinction time

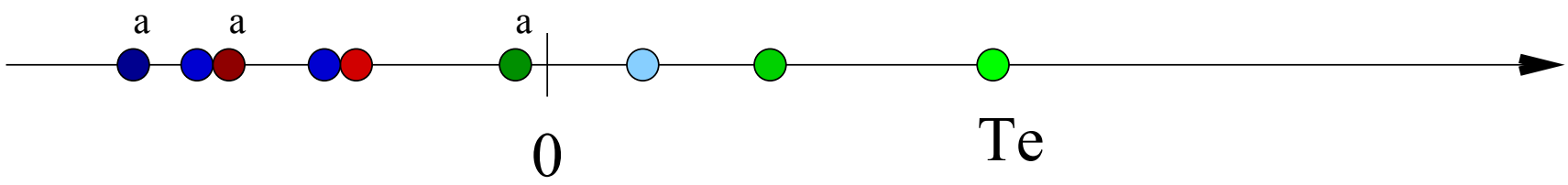
- Extinction time Te = time from 0 until the last birth, if no ancestor after 0



- Hence $\mathbb{P}(\mathrm{Te} > a)$ exponentially small
- Approximated simulation of a stationnary Hawkes process.

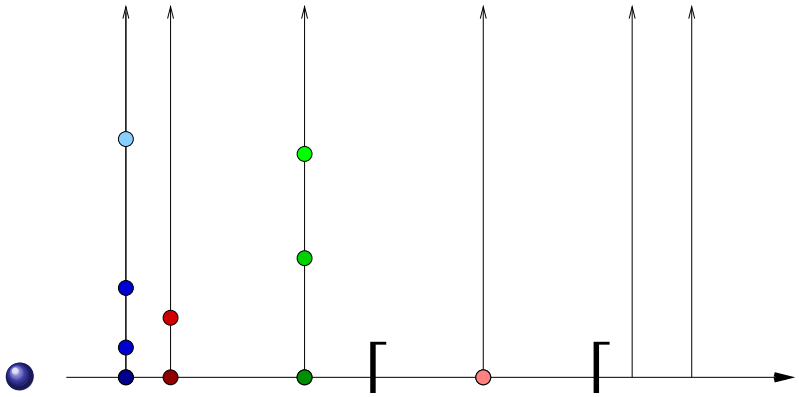
Extinction time

- Extinction time Te = time from 0 until the last birth, if no ancestor after 0

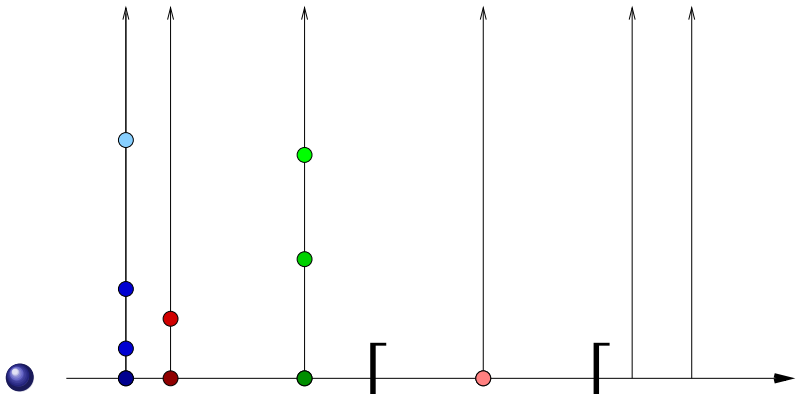


- Hence $\mathbb{P}(\mathrm{Te} > a)$ exponentially small
- Approximated simulation of a stationnary Hawkes process.
- For exact simulation see work of Möller etc

Towards independence

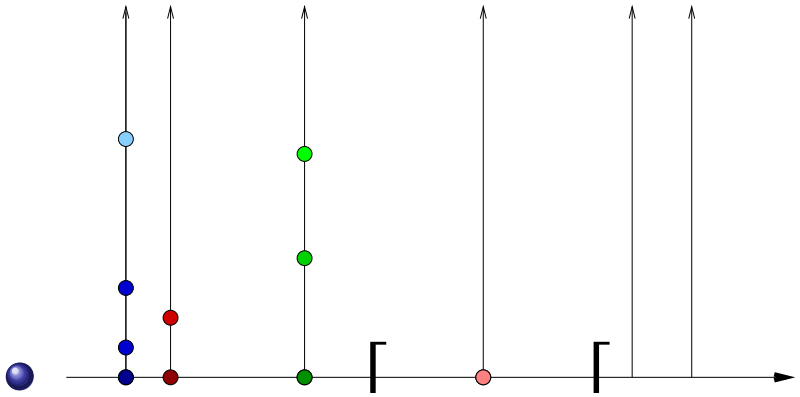


Towards independence



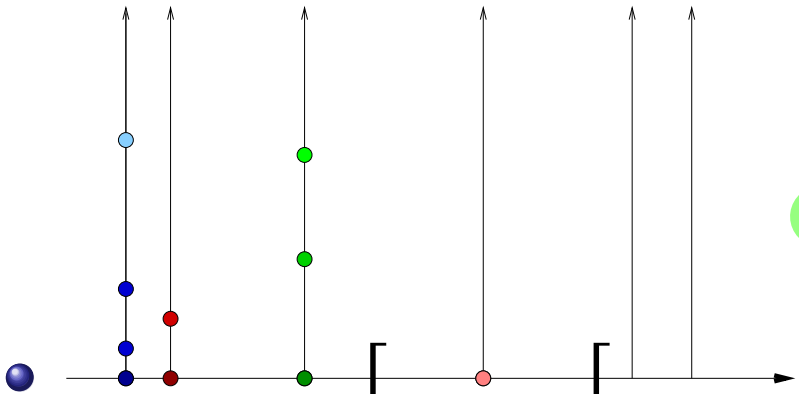
Cut $[0, T]$ in almost independent pieces (cf Berbee's lemma, discrete case (Baraud,Comte et Viennet))

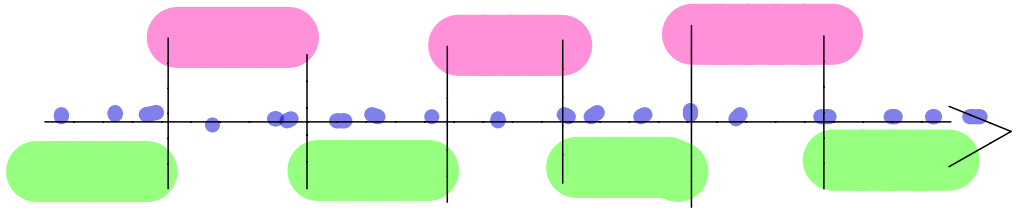
Towards independence

- 

  Cut $[0, T]$ in almost independent pieces (cf Berbee's lemma, discrete case (Baraud,Comte et Viennet))
- Control of the total variation distance.

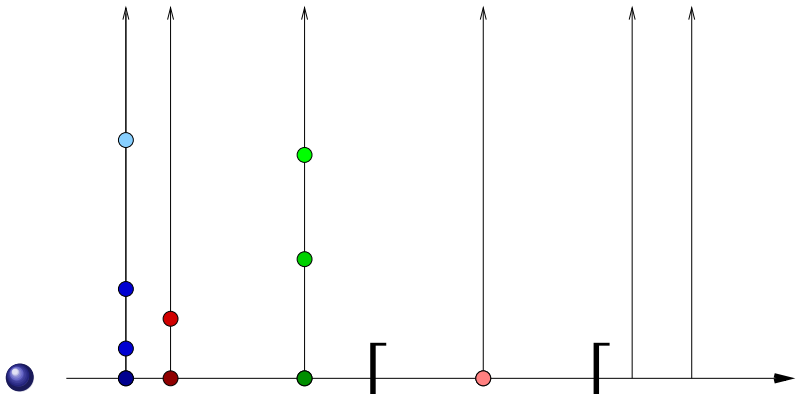
Towards independence





- Cut $[0, T]$ in almost independent pieces (cf Berbee's lemma, discrete case (Baraud,Comte et Viennet))
- Control of the total variation distance.
- Up to this error = sum of two groups of independent variables.

Towards independence



Cut $[0, T]$ in almost independent pieces (cf Berbee's lemma, discrete case (Baraud,Comte et Viennet))

- Control of the total variation distance.
- Up to this error = sum of two groups of independent variables.
- All concentration inequalities for i.i.d. variables apply (Bernstein ...)

Table of Contents

Lasso criterion

$$\begin{aligned}\hat{\mathbf{a}}^{(r)} &= \operatorname{argmin}_{\mathbf{a}}\{\gamma_T(\lambda_{\mathbf{a}}^{(r)}) + \operatorname{pen}(\mathbf{a})\} \\ &= \operatorname{argmin}_{\mathbf{a}}\{-2\mathbf{a}'\mathbf{b}_r + \mathbf{a}'\mathbf{G}\mathbf{a} + 2(\mathbf{d}^{(r)})'|\mathbf{a}|\}\end{aligned}$$

M number of interacting processes, K number of bins, A maximal interaction delay. If linear stationnary Hawkes

Oracle inequality in probability

If $\mathbf{d} = \sqrt{2(1+\varepsilon)\hat{V}_T^{\mu} x} + bx/3$ then with probability larger than

$$1 - \square M^2 K \log(T)^2 e^{-x} - \mathbb{P}(\forall t < T, i, N^i_{[t-A,t]} > \square \log(T)^2) - \mathbb{P}(\mathbf{G} \not\geq \square I/T)$$

Lasso criterion

$$\begin{aligned}\hat{\mathbf{a}}^{(r)} &= \operatorname{argmin}_{\mathbf{a}}\{\gamma_T(\lambda_{\mathbf{a}}^{(r)}) + \mathrm{pen}(\mathbf{a})\} \\ &= \operatorname{argmin}_{\mathbf{a}}\{-2\mathbf{a}'\mathbf{b}_r + \mathbf{a}'\mathbf{G}\mathbf{a} + 2(\mathbf{d}^{(r)})'|\mathbf{a}|\}\end{aligned}$$

M number of interacting processes, K number of bins, A maximal interaction delay. If linear stationnary Hawkes

Oracle inequality in probability

If $\mathbf{d} = \sqrt{2(1+\varepsilon)\hat{V}_T^{\mu} x} + bx/3$ then with probability larger than

$$1 - \square M^2 K \log(T)^2 e^{-x} - \mathbb{P}(\forall t < T, i, N^i_{[t-A,t]} > \square \log(T)^2) - \mathbb{P}(\mathbf{G} \not\geq \square I/T)$$

$$\sum_r \|\lambda^{(r)} - \mathbf{R}\mathbf{c}_t\hat{\mathbf{a}}^{(r)}\|^2 \leq \square \inf_{\mathbf{a}} \left\{ \sum_r \|\lambda^{(r)} - \mathbf{R}\mathbf{c}_t\mathbf{a}\|^2 + \frac{1}{T} \sum_{i \in supp(\mathbf{a})} (d_i^{(r)})^2 \right\}.$$

Lasso criterion...

$$\begin{aligned}\hat{\mathbf{a}}^{(r)} &= \operatorname{argmin}_{\mathbf{a}}\{\gamma_T(\lambda_{\mathbf{a}}^{(r)}) + \operatorname{pen}(\mathbf{a})\} \\ &= \operatorname{argmin}_{\mathbf{a}}\{-2\mathbf{a}'\mathbf{b}_r + \mathbf{a}'\mathbf{G}\mathbf{a} + 2(\mathbf{d}^{(r)})'|\mathbf{a}|\}\end{aligned}$$

M number of interacting processes, K number of bins, A maximal interaction delay. If linear stationnary Hawkes and K not too large ($K \leq \sqrt{T}/\log(T)^3$)

Oracle inequality in probability

If $\mathbf{d} = \sqrt{2(1+\varepsilon)\hat{V}_T^{\mu}x} + bx/3$ then with probability larger than

$$1 - \square M^2 K \log(T)^2 e^{-x} - \square T^{-1}$$

Lasso criterion...

$$\begin{aligned}\hat{\mathbf{a}}^{(r)} &= \operatorname{argmin}_{\mathbf{a}}\{\gamma_T(\lambda_{\mathbf{a}}^{(r)}) + \operatorname{pen}(\mathbf{a})\} \\ &= \operatorname{argmin}_{\mathbf{a}}\{-2\mathbf{a}'\mathbf{b}_r + \mathbf{a}'\mathbf{G}\mathbf{a} + 2(\mathbf{d}^{(r)})'|\mathbf{a}|\}\end{aligned}$$

M number of interacting processes, K number of bins, A maximal interaction delay. If linear stationnary Hawkes and K not too large ($K \leq \sqrt{T}/\log(T)^3$)

Oracle inequality in probability

If $\mathbf{d} = \sqrt{2(1+\varepsilon)\hat{V}_T^{\mu}x} + bx/3$ then with probability larger than

$$1 - \square M^2 K \log(T)^2 e^{-x} - \square T^{-1}$$

$$\sum_r \|\lambda^{(r)} - \mathbf{R}\mathbf{c}_t\hat{\mathbf{a}}^{(r)}\|^2 \leq \square \inf_{\mathbf{a}} \left\{ \sum_r \|\lambda^{(r)} - \mathbf{R}\mathbf{c}_t\mathbf{a}\|^2 + \frac{1}{T} \sum_{i \in supp(\mathbf{a})} (d_i^{(r)})^2 \right\}.$$

Choice of x

If $x = \gamma \log(T)$, with probability larger than

$$1 - M^2 K \log(T)^2 T^{-\gamma} - \square T^{-1}$$

$$\sum_r \|\lambda^{(r)} - \mathbf{Rc}_t \hat{\mathbf{a}}^{(r)}\|^2 \leq \square \inf_{\mathbf{a}} \left\{ \sum_r \|\lambda^{(r)} - \mathbf{Rc}_t \mathbf{a}\|^2 + \frac{1}{T} \sum_{i \in supp(\mathbf{a})} (d_i^{(r)})^2 \right\}.$$

Choice of x

If $x = \gamma \log(T)$, with probability larger than

$$1 - M^2 K \log(T)^2 T^{-\gamma} - \square T^{-1}$$

$$\sum_r \|\lambda^{(r)} - \mathbf{R}\mathbf{c}_t \hat{\mathbf{a}}^{(r)}\|^2 \leq \square \inf_{\mathbf{a}} \left\{ \sum_r \|\lambda^{(r)} - \mathbf{R}\mathbf{c}_t \mathbf{a}\|^2 + \frac{1}{T} \sum_{i \in supp(\mathbf{a})} (d_i^{(r)})^2 \right\}.$$

- good renormalization by T $\rightsquigarrow$, if piecewise constant true,

$$\text{loss on interaction function} \leq 0 + \square \frac{|\mathbf{a}^*|_0 \log(T)^3}{T}$$

with probability larger than $1 - \square M^2 K \log(T)^2 T^{-\gamma} - \square T^{-1}$

Choice of x

If $x = \gamma \log(T)$, with probability larger than

$$1 - M^2 K \log(T)^2 T^{-\gamma} - \square T^{-1}$$

$$\sum_r \|\lambda^{(r)} - \mathbf{R}\mathbf{c}_t \hat{\mathbf{a}}^{(r)}\|^2 \leq \square \inf_{\mathbf{a}} \left\{ \sum_r \|\lambda^{(r)} - \mathbf{R}\mathbf{c}_t \mathbf{a}\|^2 + \frac{1}{T} \sum_{i \in supp(\mathbf{a})} (d_i^{(r)})^2 \right\}.$$

- If distance bounded (reasonable),

$$\mathbb{E}(\text{ loss }) \leq \square \frac{|\mathbf{a}^*|_0 \log(T)^3}{T} + \square M^2 K \log(T)^2 T^{-\gamma} + \square T^{-1}$$

Choice of x

If $x = \gamma \log(T)$, with probability larger than

$$1 - M^2 K \log(T)^2 T^{-\gamma} - \square T^{-1}$$

$$\sum_r \|\lambda^{(r)} - \mathbf{R}\mathbf{c}_t \hat{\mathbf{a}}^{(r)}\|^2 \leq \square \inf_{\mathbf{a}} \left\{ \sum_r \|\lambda^{(r)} - \mathbf{R}\mathbf{c}_t \mathbf{a}\|^2 + \frac{1}{T} \sum_{i \in supp(\mathbf{a})} (d_i^{(r)})^2 \right\}.$$

- If distance bounded (reasonable),

$$\mathbb{E}(\text{ loss }) \leq \mathbb{I} \frac{|\mathbf{a}^*|_0 \log(T)^3}{T} + \square M^2 K \log(T)^2 T^{-\gamma} + \square T^{-1}$$

- Hence $\gamma > 1$

Choice of x

If $x = \gamma \log(T)$, with probability larger than

$$1 - M^2 K \log(T)^2 T^{-\gamma} - \square T^{-1}$$

$$\sum_r \|\lambda^{(r)} - \mathbf{R}\mathbf{c}_t \hat{\mathbf{a}}^{(r)}\|^2 \leq \square \inf_{\mathbf{a}} \left\{ \sum_r \|\lambda^{(r)} - \mathbf{R}\mathbf{c}_t \mathbf{a}\|^2 + \frac{1}{T} \sum_{i \in supp(\mathbf{a})} (d_i^{(r)})^2 \right\}.$$

- If distance bounded (reasonable),

$$\mathbb{E}(\text{ loss }) \leq \frac{|\mathbf{a}^*|_0 \log(T)^3}{T} + \square M^2 K \log(T)^2 T^{-\gamma} + \square T^{-1}$$

- Hence $\gamma > 1$
- Good reason to believe (proof in Poisson case) that $\gamma \in [1, \square]$, γ smaller changes the rate, γ larger changes the constant ...

Bernstein-type inequality for general counting process (Hansen, RB, Rivoirard)

Let $(H_s)_{s\geq 0}$ be a predictable process and $M_t = \int_0^t H_s(dN_s - \lambda(s)ds)$.
Let $b > 0$ and $v > w > 0$. For all $x, \mu > 0$ such that $\mu > \phi(\mu)$, let

$$\hat{V}_\tau^\mu = \frac{\mu}{\mu - \phi(\mu)} \int_0^\tau H_s^2 dN_s + \frac{b^2 x}{\mu - \phi(\mu)},$$

where $\phi(u) = \exp(u) - u - 1$.
Then for every stopping time τ and every $\varepsilon > 0$

$$\mathbb{P}\left(M_\tau \geq \sqrt{2(1+\varepsilon)\hat{V}_\tau^\mu x} + bx/3, \quad w \leq \hat{V}_\tau^\mu \leq v \text{ and } \sup_{s\in[0,\tau]} |H_s| \leq b \right)$$
$$\leq 2\frac{\log(v/w)}{\log(1+\varepsilon)} e^{-x}.$$

$\rightsquigarrow$ a generic choice of $\mathbf{d}$ for the Lasso whatever the underlying process....

Practical choice of **d**

Recall that the oracle inequality needs

$$\left|\mathbf{b}^{(r)} - \bar{\mathbf{b}}^{(r)}\right| \leq \mathbf{d}^{(r)}, \quad \forall r$$

with

$$\mathbf{b} = \int_0^T \mathbf{Rc}_t dN_t^{(r)} \text{ and } \bar{\mathbf{b}} = \int_0^T \mathbf{Rc}_t \lambda^{(r)}(t) dt.$$

- 'Bernstein Lasso' and 'Bernstein Lasso + OLS'

$$d_i = \sqrt{2\gamma \log(T) \hat{V}_i} + \frac{B_i \gamma \log(T)}{3},$$

$$\hat{V}_i = \int_0^T (\mathbf{Rc}_{t,i})^2 \mathrm{d}N_{t,r_i}, \quad B_i = \sup_{t \in [0,T], m} |\mathbf{Rc}_{t,i}|.$$

Practical choice of $\mathbf{d}$

Recall that the oracle inequality needs

$$\left|\mathbf{b}^{(r)} - \bar{\mathbf{b}}^{(r)}\right| \leq \mathbf{d}^{(r)}, \quad \forall r$$

with

$$\mathbf{b} = \int_0^T \mathbf{Rc}_t dN_t^{(r)} \text{ and } \bar{\mathbf{b}} = \int_0^T \mathbf{Rc}_t \lambda^{(r)}(t) dt.$$

- 'Bernstein Lasso' and 'Bernstein Lasso + OLS'

$$d_i = \sqrt{2\gamma \log(T) \hat{V}_i} + \frac{B_i \gamma \log(T)}{3},$$

$$\hat{V}_i = \int_0^T (\mathbf{Rc}_{t,i})^2 \mathrm{d}N_{t,r_i}, \quad B_i = \sup_{t \in [0,T], m} |\mathbf{Rc}_{t,i}|.$$

- 'Adaptive Lasso' (Zou)

$$d_i = \frac{\gamma}{2|\hat{a}_i^{ols}|}.$$

Why +OLS ?

Lasso : Least-absolute Shrinkage and Selection Operator

If $\mathbf{G} = I$, minimizing

$$-2\mathbf{a}'b + \mathbf{a}'G\mathbf{a} + 2\mathbf{d}'|\mathbf{a}|$$

$\rightsquigarrow$ a soft-thresholding estimator, i.e.

$$\hat{\mathbf{a}}_{Lasso} = (\mathbf{b} - sign(\mathbf{b})\mathbf{d})_+.$$

Why +OLS ?

If $\mathbf{G} = I$, minimizing

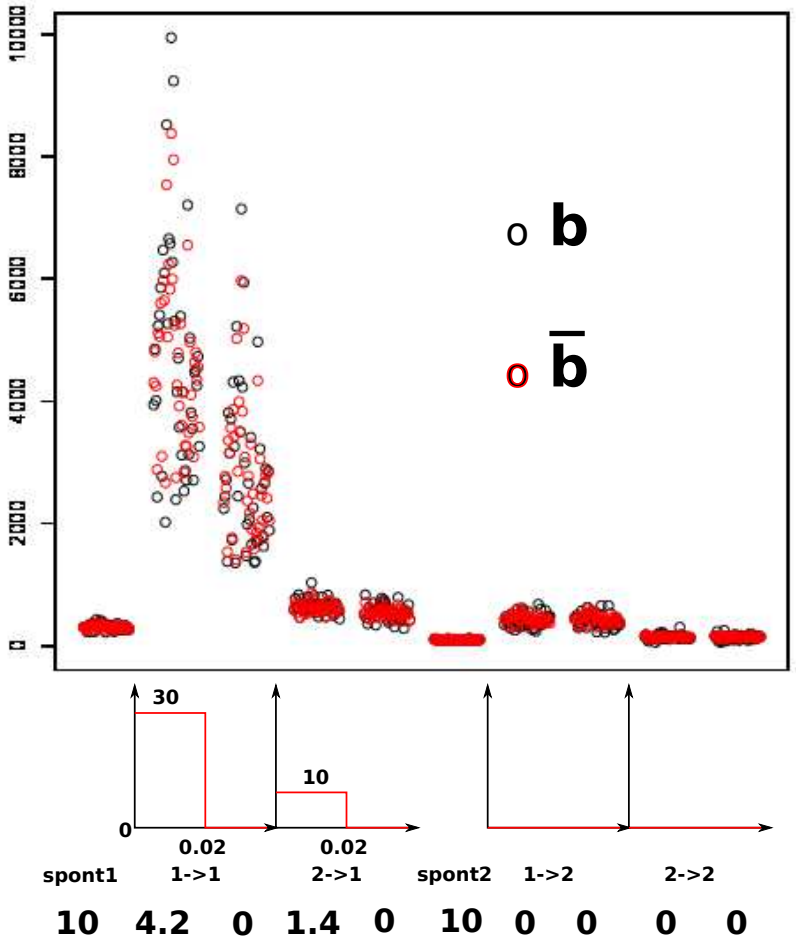
$$-2\mathbf{a}'b + \mathbf{a}'G\mathbf{a} + 2\mathbf{d}'|\mathbf{a}|$$

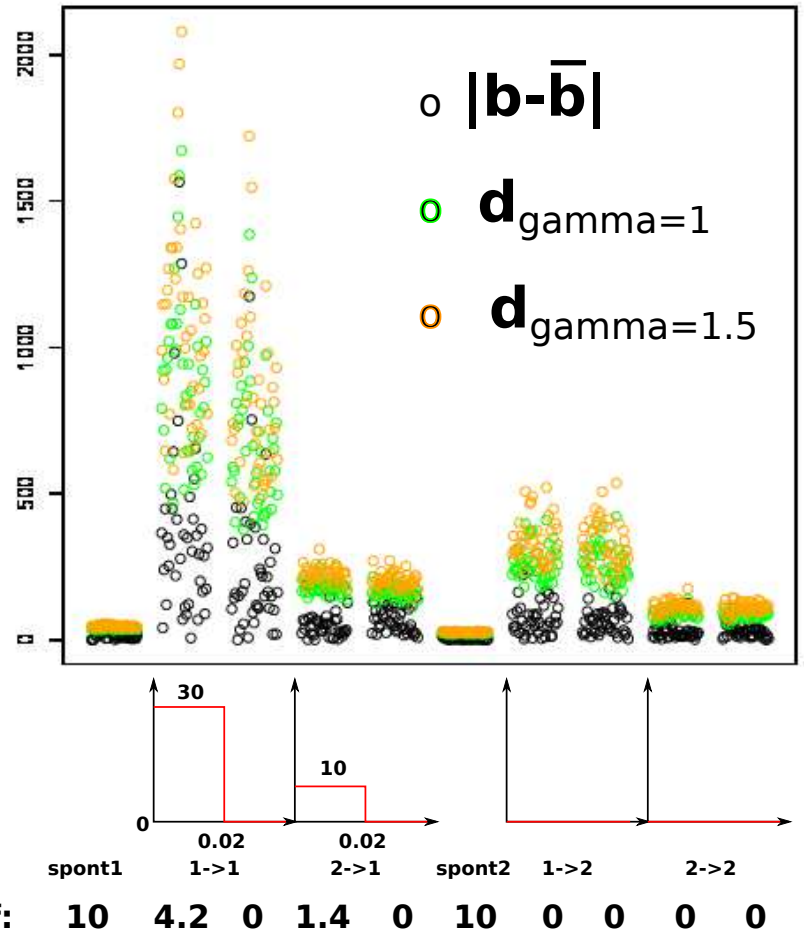
$\rightsquigarrow$ a soft-thresholding estimator, i.e.

$$\hat{\mathbf{a}}_{Lasso} = (\mathbf{b} - sign(\mathbf{b})\mathbf{d})_+.$$

Hence small bias anytime $\rightsquigarrow$
once the support is estimated by Lasso, compute the ordinary least-square estimate (OLS) on the resulting support.

Choices of the weights d_λ





NB : Adaptive Lasso (Zou) $d_\lambda = \gamma/(2|\hat{a}_\lambda|)$

Simulation study - Support recovery

We perform 100 runs with $T = 2$, $M = 8$, $K = 4$ (264 coefficients to be estimated by using 636 observations)

	Bernstein Lasso			Adaptive Lasso		
Tuning parameter γ	0.5	1	2	2	200	1000
Correct clusters identif. $\in [0, 100]$	0	**32**	24	0	0	**32**
False non-zero interactions $\in [0, 55]$	17	6	**1**	55	13	**1**
False zero interactions $\in [0, 9]$	**0**	**0**	2	**0**	**0**	2
False zero spontaneous rates $\in [0, 8]$	**0**	**0**	**0**	**0**	1	3
False non-zero coeff. $\in [0, 238]$	22	7	**1**	199	17	**1**
False zero coeff. $\in [0, 18]$	1	2	7	**0**	2	7

Simulation study - Support recovery

We perform 100 runs with $T = 20$, $M = 8$, $K = 4$

Tuning parameter γ	Bernstein Lasso 0.5	Bernstein Lasso 1	Bernstein Lasso 2	Adaptive Lasso 2	Adaptive Lasso 200	Adaptive Lasso 1000
Correct clusters identif. $\in [0, 100]$	63	99	**100**	0	0	90
False non-zero interactions $\in [0, 55]$	3	1	**0**	55	10	**0**
False zero interactions $\in [0, 9]$	**0**	**0**	**0**	**0**	**0**	**0**
False zero spontaneous rates $\in [0, 8]$	**0**	**0**	**0**	**0**	**0**	**0**
False non-zero coeff. $\in [0, 238]$	4	1	**0**	197	13	**0**
False zero coeff. $\in [0, 18]$	**0**	**0**	**0**	**0**	**0**	**0**

For $T = 20$,

- $\gamma = 1$ or $\gamma = 2$ is convenient for Bernstein Lasso. It was also the case for $T = 2$.
- $\gamma = 1000$ is convenient for Adaptive Lasso. It was not the case for $T = 2$.

Simulation study - Influence of the OLS step

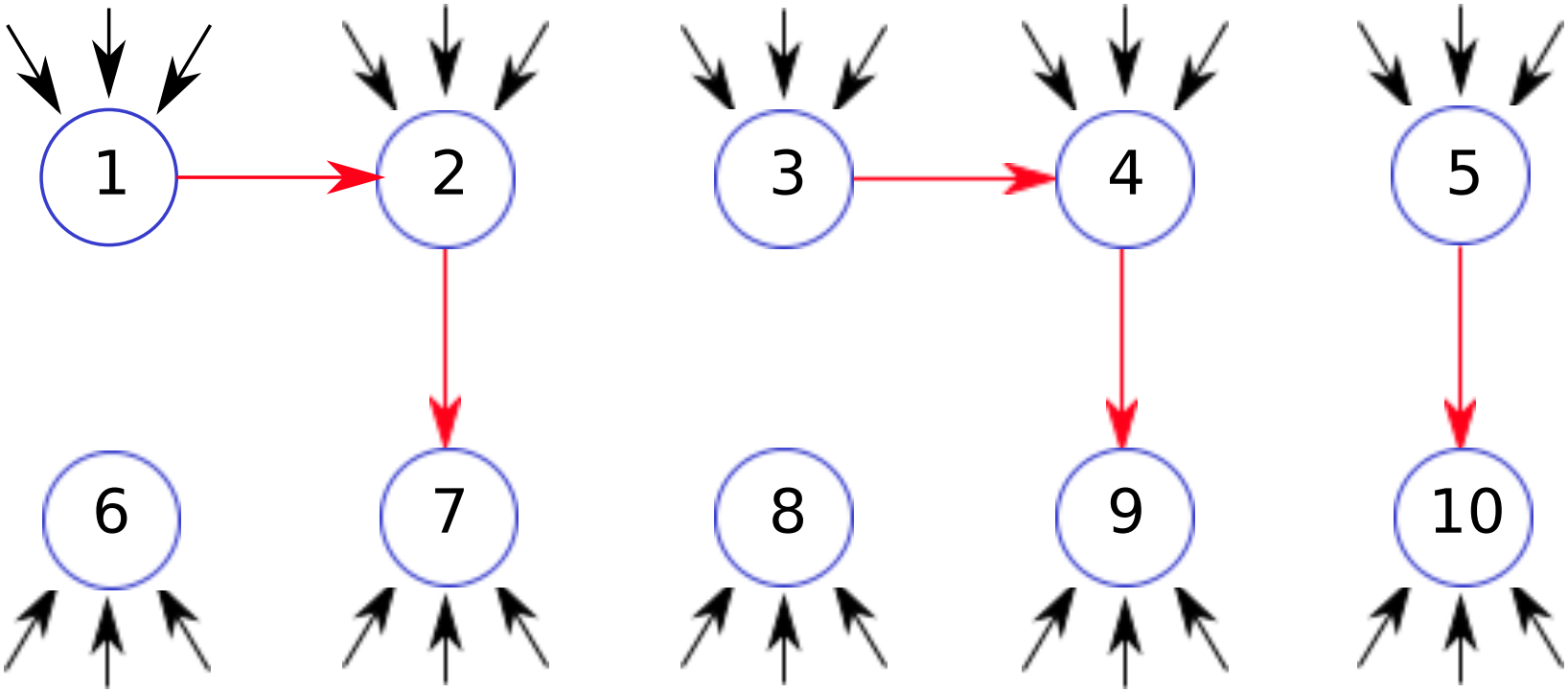
We perform 100 runs with $T = 20$, $M = 8$, $K = 4$.

Tuning parameter γ	Bernstein Lasso $\gamma = 1$	Bernstein Lasso $\gamma = 2$	Adaptive Lasso $\gamma = 1000$
MSE for spontaneous rates	37	69	27
MSE for spontaneous rates after OLS	10	9	10
MSE for interaction functions	3	6	0.5
MSE for interaction functions after OLS	0.5	0.4	0.4

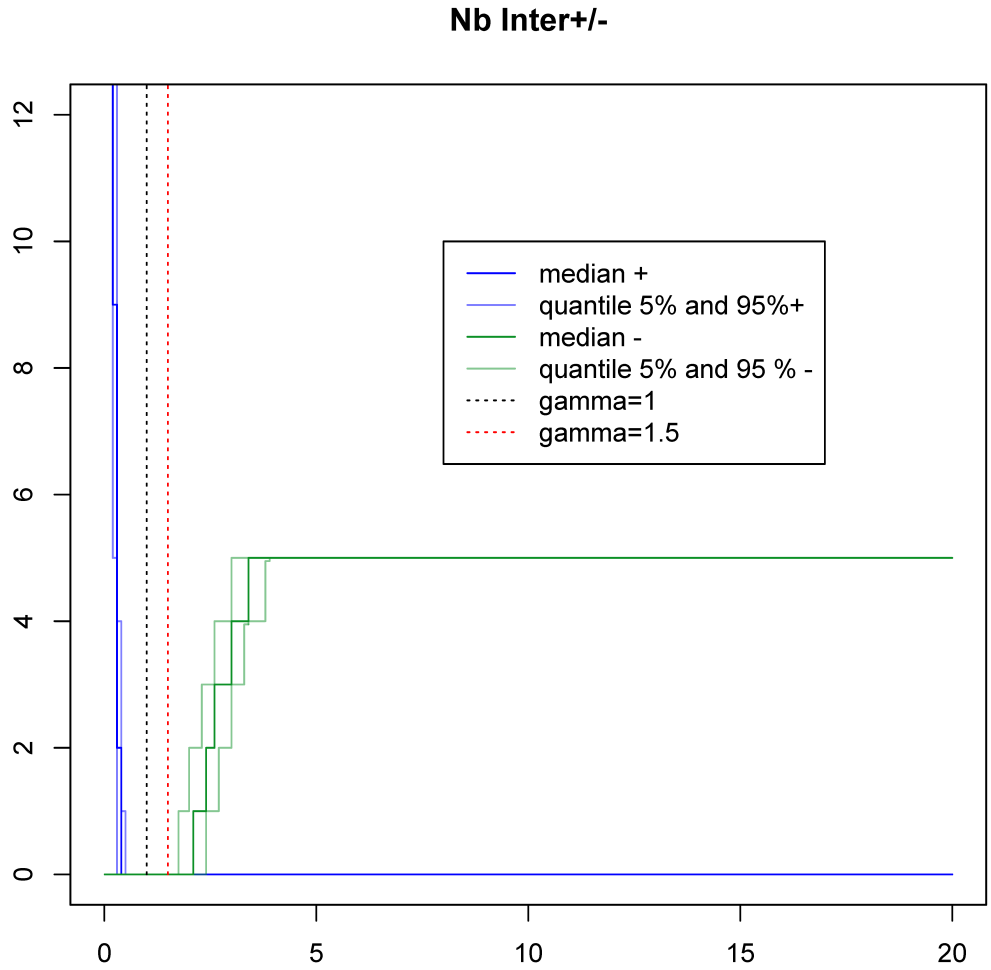
$$\text{MSE for spontaneous rates } = \sum_{m=1}^{M} (\hat{\nu}^{(m)} - \nu^{(m)})^2$$

$$\text{MSE for interaction functions } = \sum_{m=1}^{M} \sum_{\ell=1}^{M} \int (\hat{h}_\ell^{(m)}(t) - h_\ell^{(m)}(t))^2 \mathrm{d}t$$

Another (more realistic ?) neuronal network : Integrate and Fire

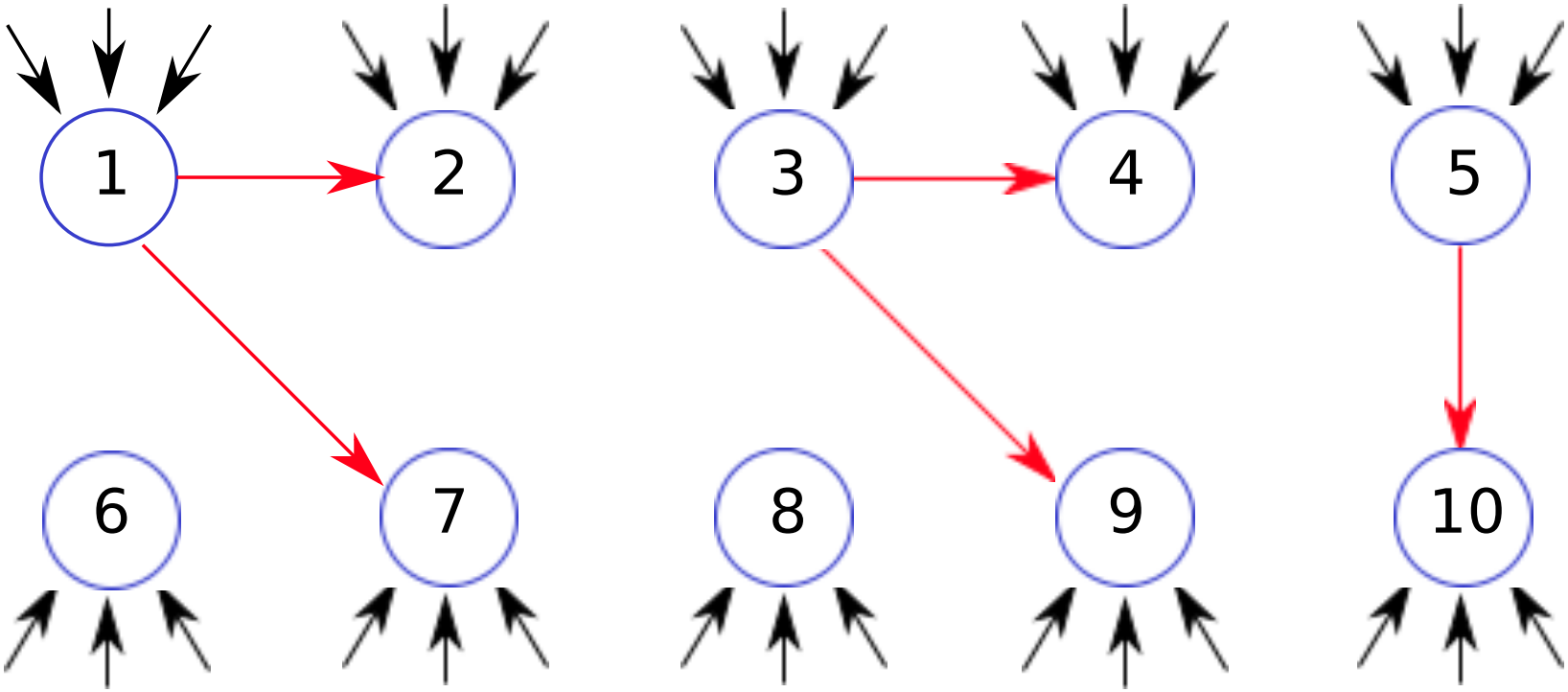


Another (more realistic ?) neuronal network : Integrate and Fire



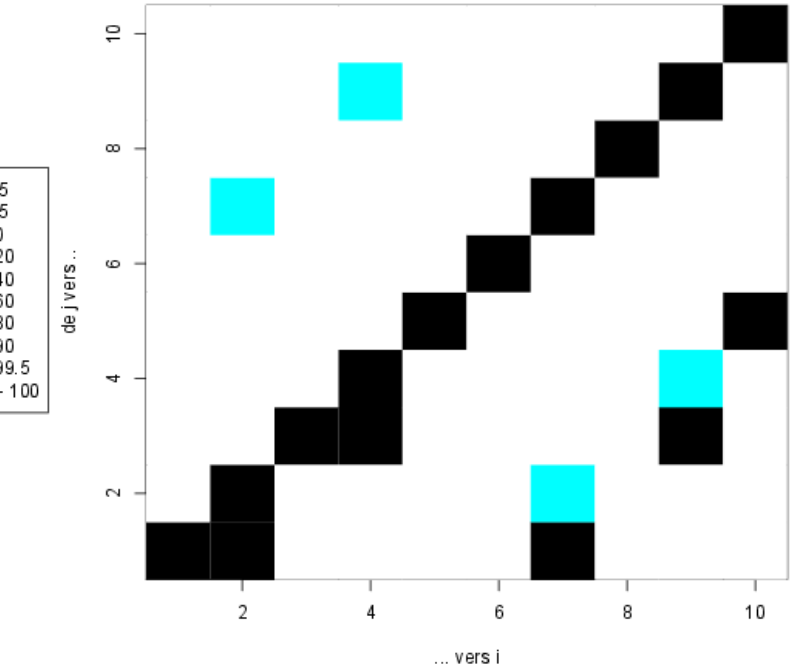
$T = 60s$

Another (more realistic ?) neuronal network : Integrate and Fire

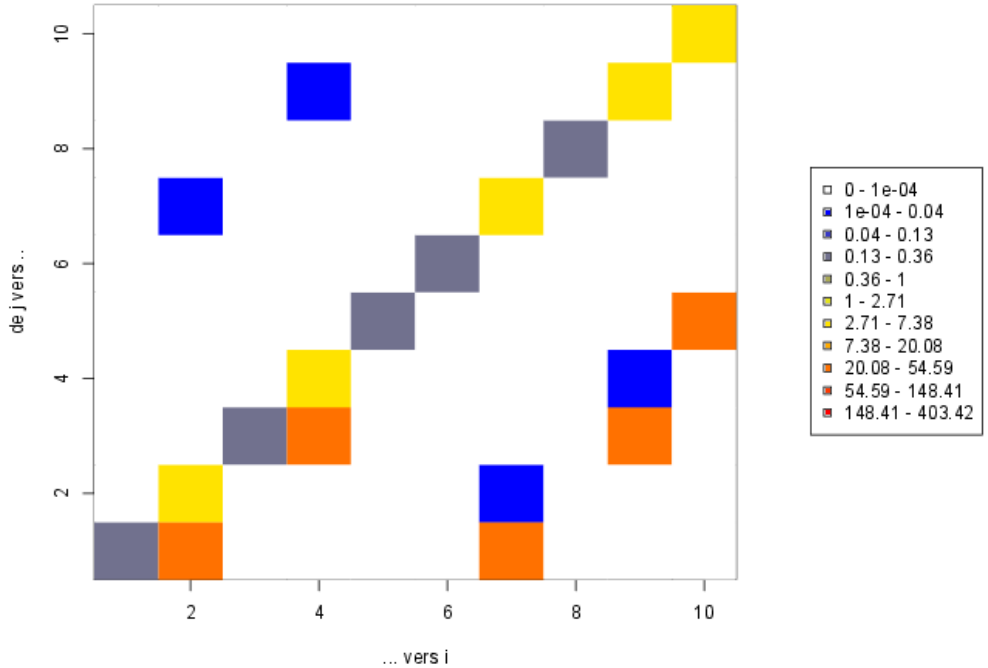


Another (more realistic ?) neuronal network : Integrate and Fire

Number of times that a given interaction is detected



Mean energy ($\int h^2$) when detected



Open questions for Lasso

- Branching arguments ok for linear Hawkes
- Hence number of points controlled if intensity smaller than stationnary Hawkes (thinning) and in particular for $(.)_+$ and inhibition.

Open questions for Lasso

- Branching arguments ok for linear Hawkes
- Hence number of points controlled if intensity smaller than stationnary Hawkes (thinning) and in particular for $(.)_+$ and inhibition.
- Minoration of $\mathbb{E}(\mathbf{G})$ can be done if intensity lower bounded and upper bounded by Hawkes

Open questions for Lasso

- Branching arguments ok for linear Hawkes
- Hence number of points controlled if intensity smaller than stationnary Hawkes (thinning) and in particular for $(.)_+$ and inhibition.
- Minoration of $\mathbb{E}(\mathbf{G})$ can be done if intensity lower bounded and upper bounded by Hawkes
- The "Berbee's lemma" : no idea how to get it without linear Hawkes. Hence what about $(.)_+$? ? ?

Open questions for Lasso

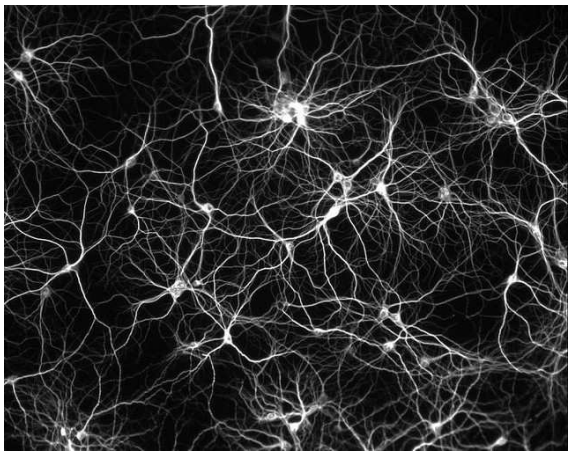
- Branching arguments ok for linear Hawkes
- Hence number of points controlled if intensity smaller than stationnary Hawkes (thinning) and in particular for $(.)_+$ and inhibition.
- Minoration of $\mathbb{E}(\mathbf{G})$ can be done if intensity lower bounded and upper bounded by Hawkes
- The "Berbee's lemma" : no idea how to get it without linear Hawkes. Hence what about $(.)_+$? ? ?
- Still on simulation works even if not Hawkes and $\gamma = 1.5$ works !

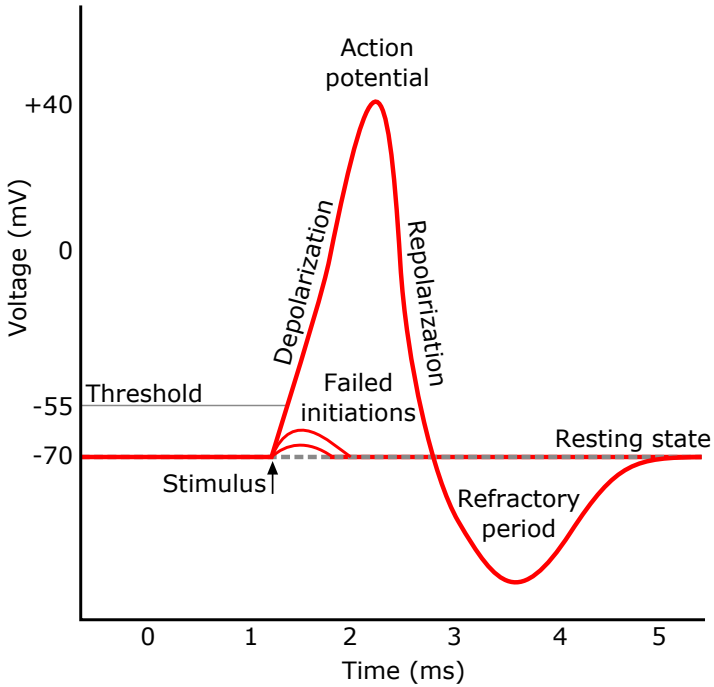
Open questions for Lasso

- Branching arguments ok for linear Hawkes
- Hence number of points controlled if intensity smaller than stationnary Hawkes (thinning) and in particular for $(.)_+$ and inhibition.
- Minoration of $\mathbb{E}(\mathbf{G})$ can be done if intensity lower bounded and upper bounded by Hawkes
- The "Berbee's lemma" : no idea how to get it without linear Hawkes. Hence what about $(.)_+$? ? ?
- Still on simulation works even if not Hawkes and $\gamma = 1.5$ works !
- What about group Lasso ? (no sharp enough concentration inequality yet)
- What about non stationnarity ? Segmentation, clustering (F. Picard, C. Tuleau-Malot)

Table of Contents

Biological context





Physiological constraint : refractory period.

Age structured equations (Pakdaman, Perthame, Salort, 2010)

Age = delay since last spike.

$$n(t,s) = \begin{cases} \text{probability density of finding a neuron with age } s \text{ at time } t. \\ \text{ratio of the population with age } s \text{ at time } t. \end{cases}$$

Age structured equations (Pakdaman, Perthame, Salort, 2010)

Age = delay since last spike.

$$n(t,s) = \begin{cases} \text{probability density of finding a neuron with age } s \text{ at time } t. \\ \text{ratio of the population with age } s \text{ at time } t. \end{cases}$$

$$\begin{cases} \dfrac{\partial n(t,s)}{\partial t} + \dfrac{\partial n(t,s)}{\partial s} + p(s, X(t))\, n(t,s) = 0 \\ m(t) := n(t,0) = \displaystyle\int_0^{+\infty} p(s, X(t))\, n(t,s)\, ds \end{cases} \qquad \text{(PPS)}$$

Age structured equations (Pakdaman, Perthame, Salort, 2010)

Age = delay since last spike.

$$n(t,s) = \begin{cases} \text{probability density of finding a neuron with age } s \text{ at time } t. \\ \text{ratio of the population with age } s \text{ at time } t. \end{cases}$$

$$\begin{cases} \dfrac{\partial n(t,s)}{\partial t} + \dfrac{\partial n(t,s)}{\partial s} + p(s, X(t))\, n(t,s) = 0 \\ m(t) := n(t,0) = \displaystyle\int_0^{+\infty} p(s, X(t))\, n(t,s)\, ds \end{cases} \qquad \text{(PPS)}$$

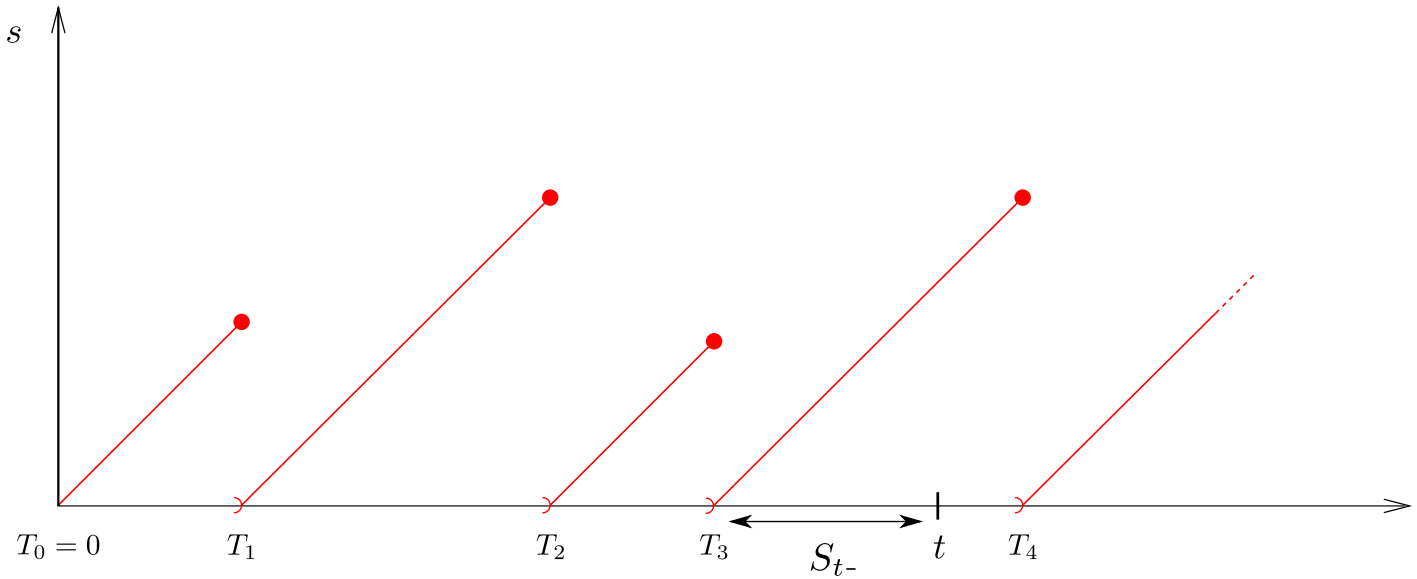
p represents the firing rate. For example, $p(s,X) = \mathbb{1}_{\{s>\sigma(X)\}}$.

$$X(t) = \int_0^t d(x) m(t-v) dv \quad \text{(global neural activity)}$$

d = delay function. For example, $d(v) = e^{-\tau v}$.

One neuron = point process $\rightsquigarrow$ age ?

- Age = delay since last spike.



Microscopic age

- We consider the continuous to the left (hence predictable) version of the age.
- The age at time 0 depends on the spiking times before time 0.
- The dynamic is characterized by the spiking times after time 0.

Framework

$\cdots < T_{-1} < T_0 \leq 0 < T_1 < \ldots$ the ordered sequence of points of N.
Dichotomy of the behaviour of N with respect to time 0 :

Framework

$\cdots < T_{-1} < T_0 \leq 0 < T_1 < \ldots$ the ordered sequence of points of N.
Dichotomy of the behaviour of N with respect to time 0 :

- $N_- = N \cap (-\infty, 0]$ is a point process with distribution $\mathbb{P}_0$ (initial condition).
  The age at time 0 is finite $\Leftrightarrow N_- \neq \emptyset$.
- $N_+ = N \cap (0, +\infty)$ is a point process admitting some intensity $\lambda(t, \mathcal{F}_{t-}^N) \rightsquigarrow$ "probability to find a new point at time t given the past"

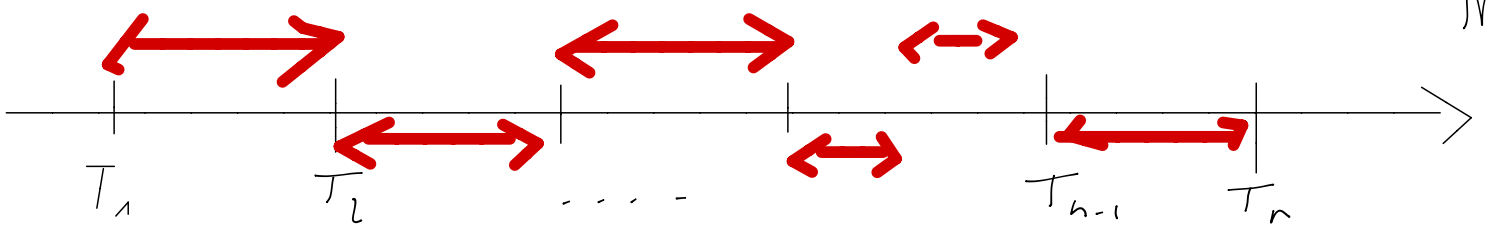
Framework

$\cdots < T_{-1} < T_0 \leq 0 < T_1 < \ldots$ the ordered sequence of points of N.
Dichotomy of the behaviour of N with respect to time 0 :

- $N_- = N \cap (-\infty, 0]$ is a point process with distribution $\mathbb{P}_0$ (initial condition).
  The age at time 0 is finite $\Leftrightarrow N_- \neq \emptyset$.
- $N_+ = N \cap (0, +\infty)$ is a point process admitting some intensity $\lambda(t, \mathcal{F}_{t-}^N) \rightsquigarrow$ "probability to find a new point at time t given the past"
- $p(s, X(t))$ and $\lambda(t, \mathcal{F}_{t-}^N)$ are analogous.

Some classical point processes in neuroscience

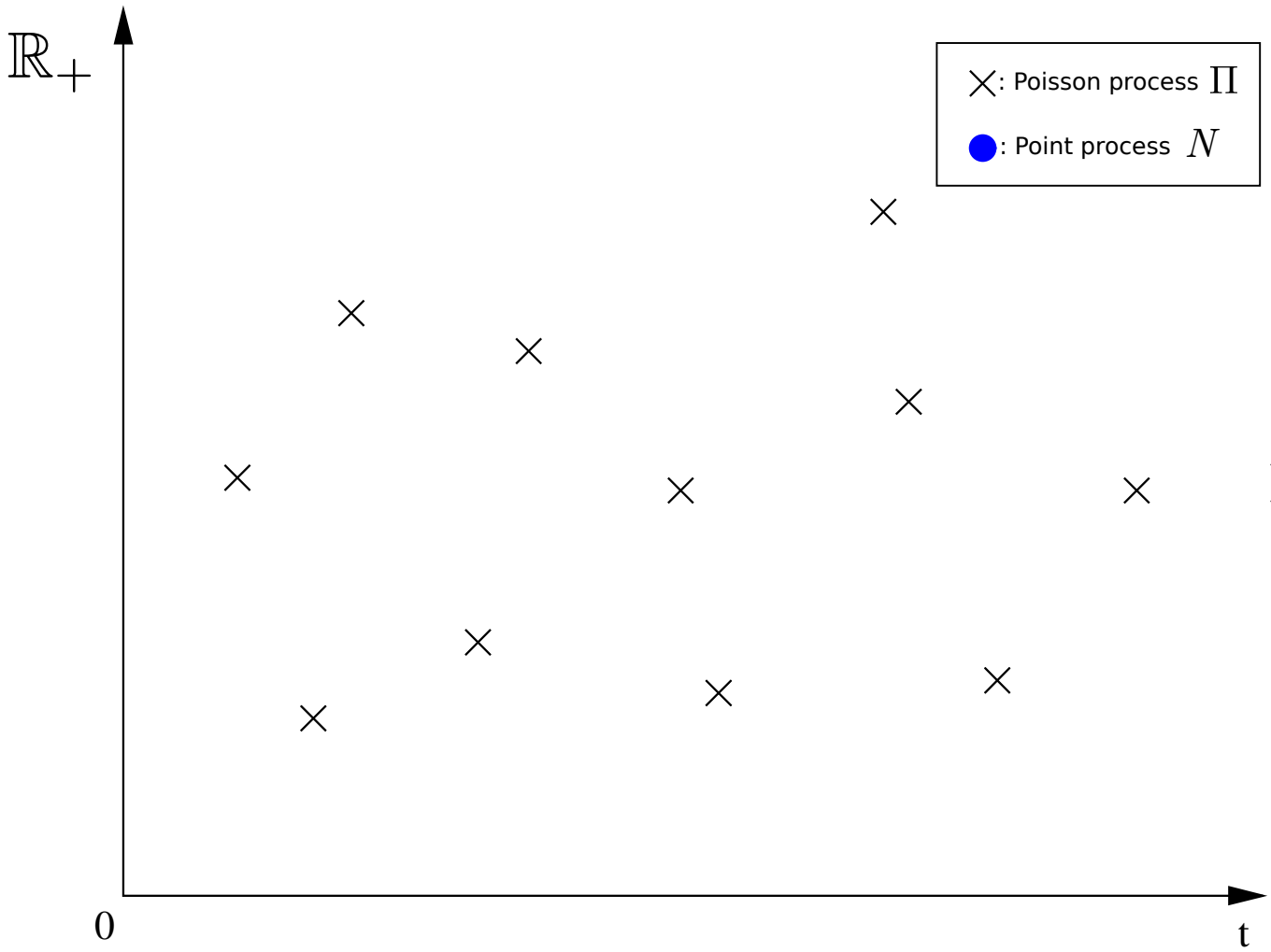
- Poisson process : $\lambda(t, \mathcal{F}_{t-}^N) = \lambda(t) =$ deterministic function.
- Renewal process : $\lambda(t, \mathcal{F}_{t-}^N) = f(S_{t-}) \Leftrightarrow$ i.i.d. ISIs.



- Hawkes process : $\lambda(t, \mathcal{F}_{t-}^N) = \mu + \int_{-\infty}^{t-} h(t-v)N(dv)$

$$\int_{-\infty}^{t-} h(t-v)N(dv) \quad \longleftrightarrow \quad \int_0^t d(v)m(t-v)dv = X(t).$$

Dynamic = Ogata's thinning



Π is a Poisson process with rate 1.

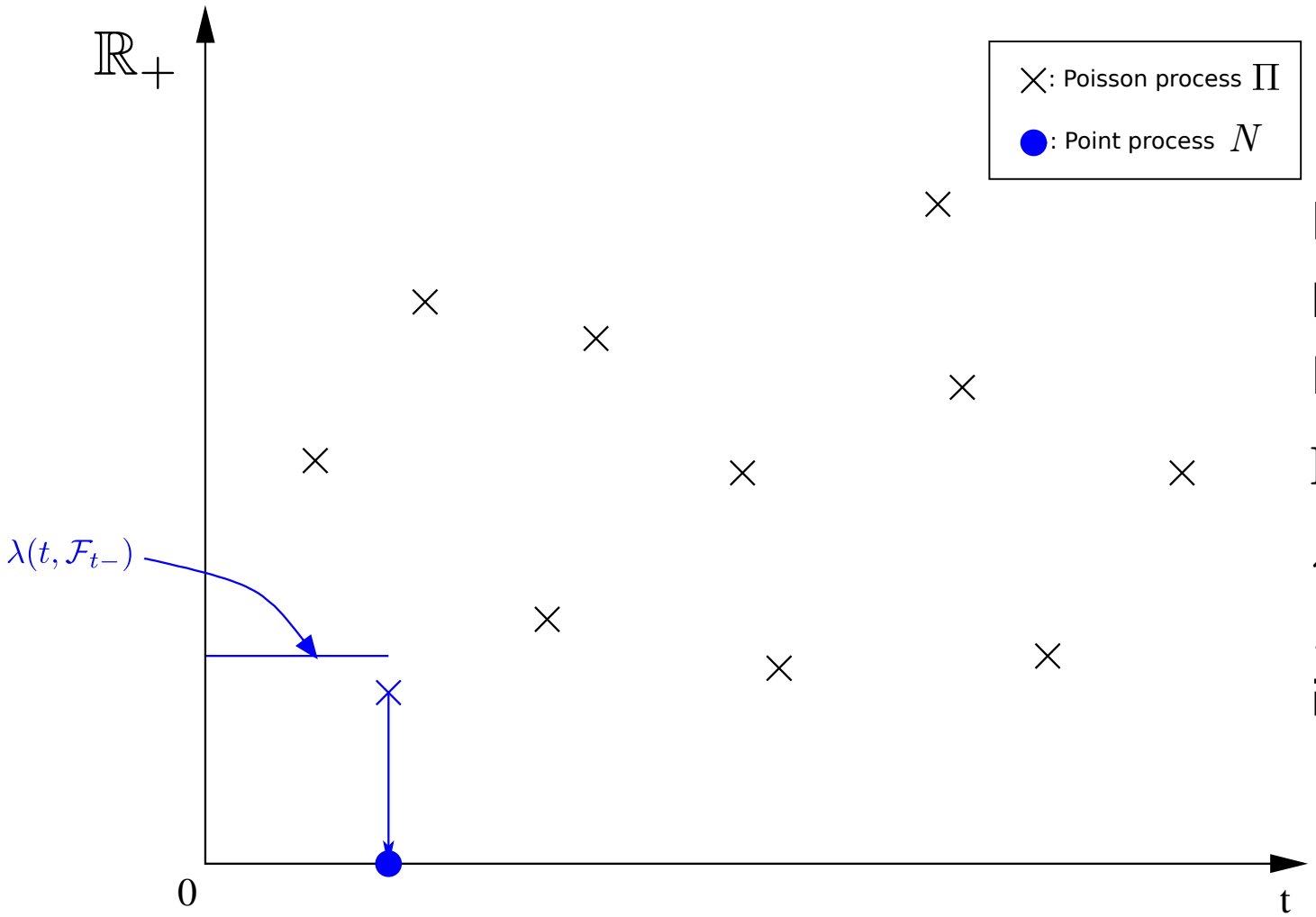
$\Pi(dt, dx) = \sum \delta_{\mathsf{X}}$.

$\mathbb{E}[\Pi(dt, dx)] = dtdx$.

λ is random.

N admits λ as an intensity.

Dynamic = Ogata's thinning



Π is a Poisson process with rate 1.

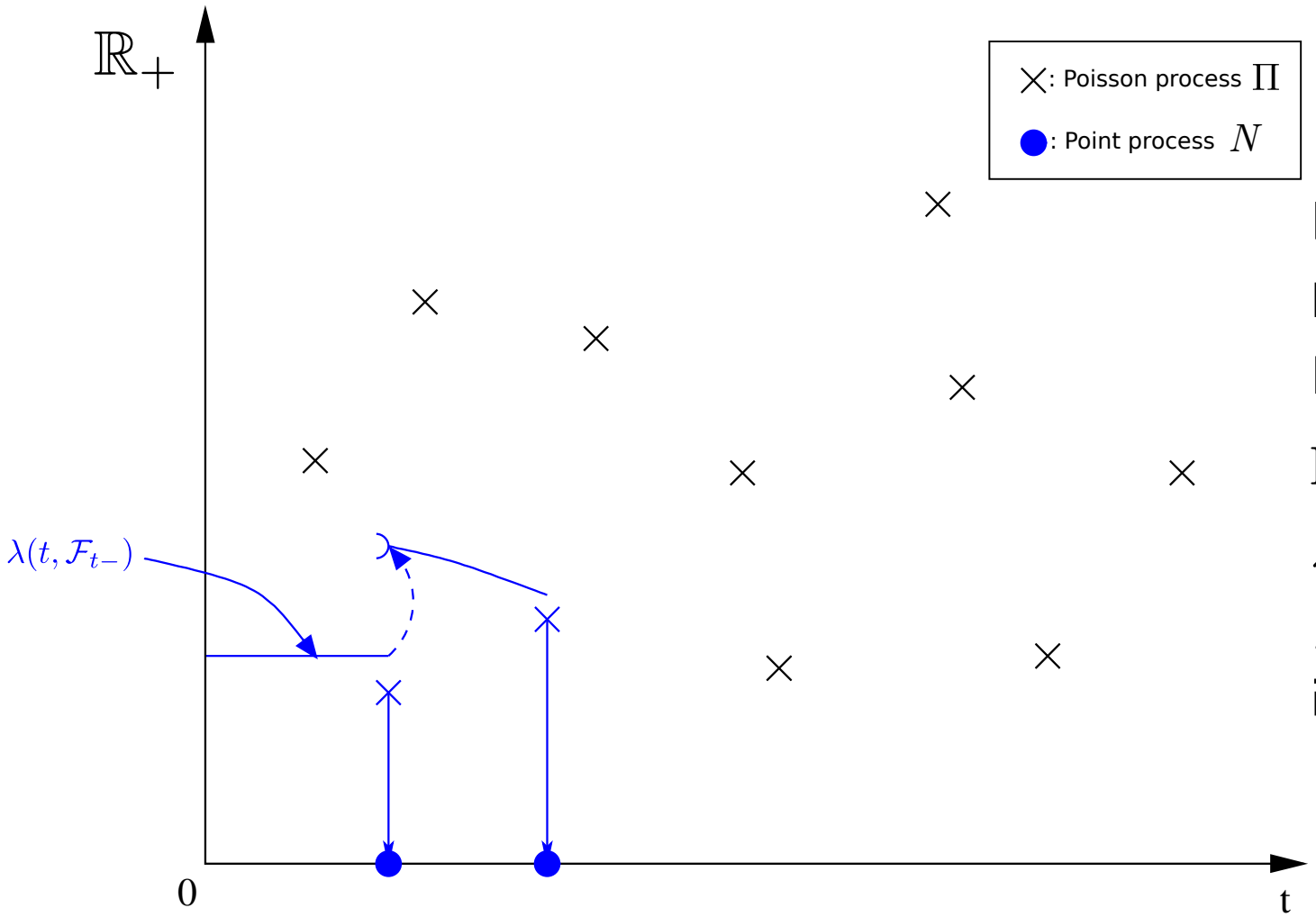
$\Pi(dt, dx) = \sum \delta_{\mathsf{x}}$.

$\mathbb{E}[\Pi(dt, dx)] = dtdx$.

λ is random.

N admits λ as an intensity.

Dynamic = Ogata's thinning



Π is a Poisson process with rate 1.

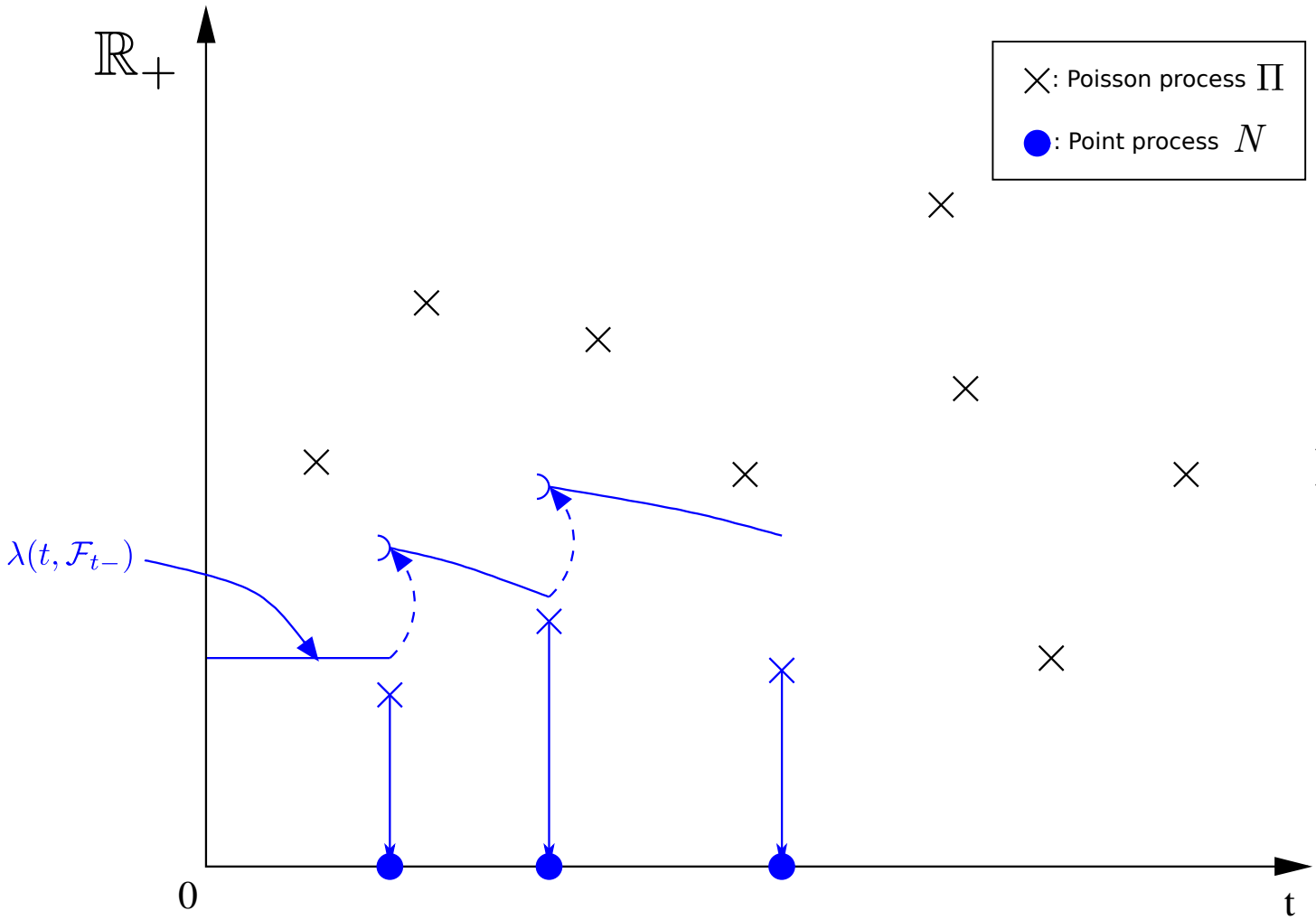
$\Pi(dt, dx) = \sum \delta_{\mathsf{x}}$.

$\mathbb{E}[\Pi(dt, dx)] = dtdx$.

λ is random.

N admits λ as an intensity.

Dynamic = Ogata's thinning



Π is a Poisson process with rate 1.

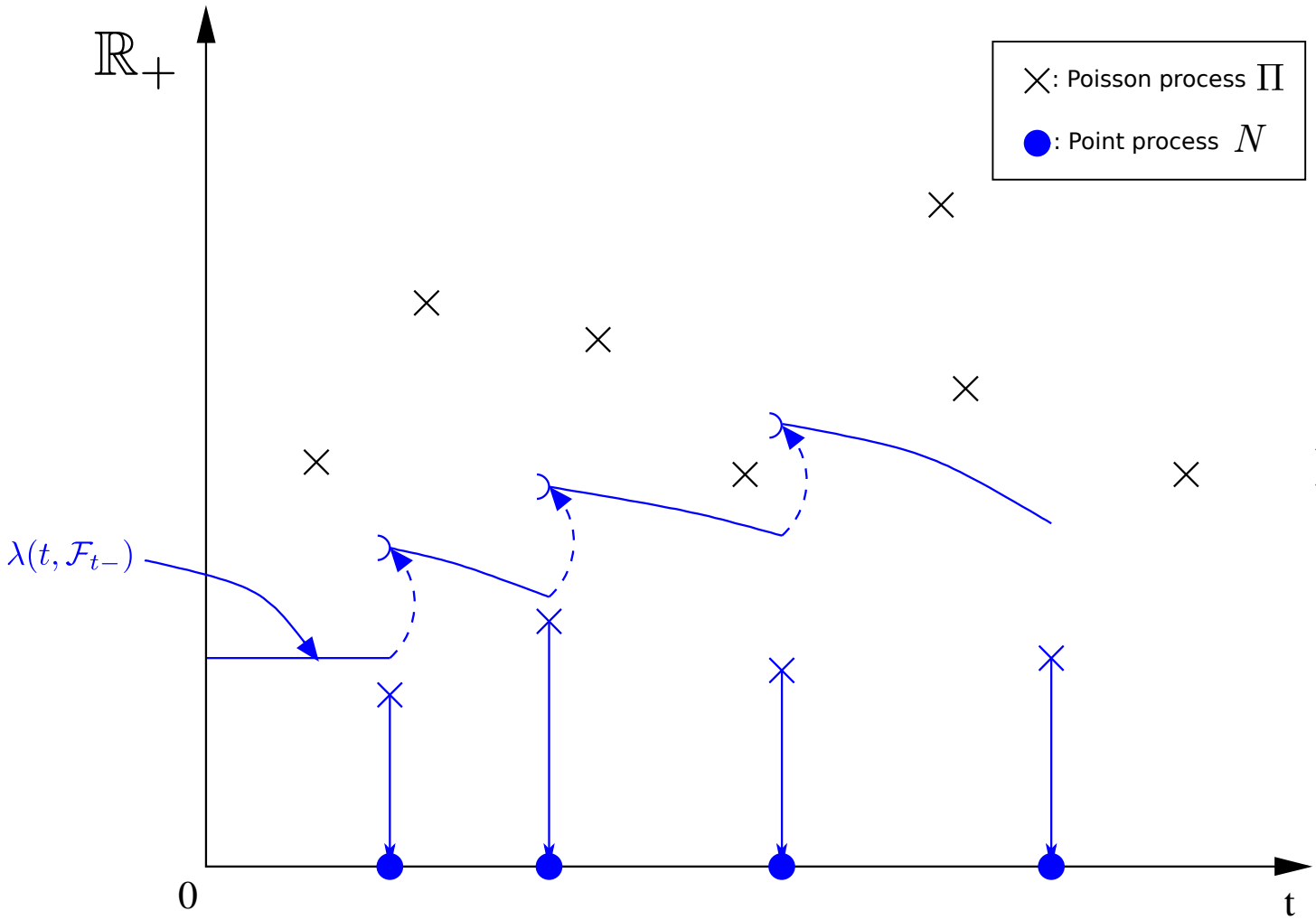
$\Pi(dt, dx) = \sum \delta_{\mathsf{x}}$.

$\mathbb{E}[\Pi(dt, dx)] = dtdx$.

λ is random.

N admits λ as an intensity.

Dynamic = Ogata's thinning



Π is a Poisson process with rate 1.

$\Pi(dt, dx) = \sum \delta_{\mathsf{x}}$.

$\mathbb{E}[\Pi(dt, dx)] = dtdx$.

λ is random.

N admits λ as an intensity.

A microscopic analogous to n

- $n(t,.)$ is the probability density of the age at time t.
- At fixed time t, we are looking at a Dirac mass at S_{t-}.

A microscopic analogous to n

- $n(t,.)$ is the probability density of the age at time t.
- At fixed time t, we are looking at a Dirac mass at S_{t-}.

What we need

- Random measure U on $\mathbb{R}^2$.
- Action over test functions : $\forall \varphi \in C^{\infty}_{c,b}(\mathbb{R}^2_+)$,

$$\int \varphi(t,s)U(dt,ds) = \int \varphi(t,S_{t-})dt.$$

What we define

We construct an ad hoc random measure U which satisfies a system of stochastic differential equations similar to (PPS).

Microscopic equation (Chevallier, Cáceres, Doumic, RB)

Let Π be a Poisson measure. Let $\left(\lambda(t, \mathcal{F}_{t-}^N)\right)_{t>0}$ be some non negative predictable process which is L^1_{loc} a.s.
The measure U satisfies the following system a.s.

$$\begin{cases} (\partial_t + \partial_s)\{U(dt, ds)\} + \left(\int_{x=0}^{\lambda\left(t, \mathcal{F}_{t-}^N\right)} \Pi(dt, dx) \right) U(t, ds) = 0, \\ U(dt, 0) = \int_{s \in \mathbb{R}} \left(\int_{x=0}^{\lambda\left(t, \mathcal{F}_{t-}^N\right)} \Pi(dt, dx) \right) U(t, ds), \end{cases}$$

in the weak sense with initial condition $\lim_{t \to 0^+} U(t, \cdot) = \delta_{-T_0}$. ($-T_0$ is the age at time 0)

Microscopic equation (Chevallier, Cáceres, Doumic, RB)

Let Π be a Poisson measure. Let $\left(\lambda(t, \mathcal{F}_{t-}^N)\right)_{t>0}$ be some non negative predictable process which is L^1_{loc} a.s.
The measure U satisfies the following system a.s.

$$\begin{cases} (\partial_t + \partial_s)\{U(dt,ds)\} + \left(\int_{x=0}^{\lambda\left(t,\mathcal{F}_{t-}^N\right)} \Pi(dt,dx)\right) U(t,ds) = 0, \\ U(dt,0) = \int_{s\in\mathbb{R}} \left(\int_{x=0}^{\lambda\left(t,\mathcal{F}_{t-}^N\right)} \Pi(dt,dx)\right) U(t,ds), \end{cases}$$

in the weak sense with initial condition $\lim_{t\to 0^+} U(t,\cdot) = \delta_{-T_0}$. ($-T_0$ is the age at time 0)

$\rightsquigarrow$ $p(s, X(t))$ is replaced by $\int_{x=0}^{\lambda\left(t,\mathcal{F}_{t-}^N\right)} \Pi(dt,dx)$.

$$\mathbb{E}\left[\int_{x=0}^{\lambda\left(t,\mathcal{F}_{t-}^N\right)} \Pi(dt,dx) \middle| \mathcal{F}_{t-}^N\right] = \lambda\left(t,\mathcal{F}_{t-}^N\right) dt.$$

Taking the expectation

[...] defining $u = "\mathbb{E}(U)"$, $u(t,.)$ distribution of S_{t-}.
Let $\left(\lambda(t, \mathcal{F}_{t-}^N)\right)_{t>0}$ be some non negative predictable process which is L^1_{loc} a.s.
The measure U satisfies the following system,

$$\begin{cases} (\partial_t + \partial_s)\{U(dt,ds)\} + \left(\int_{x=0}^{\lambda\left(t,\mathcal{F}_{t-}^N\right)} \Pi(dt,dx)\right) U(t,ds) = 0, \\ U(dt,0) = \int_{s\in\mathbb{R}} \left(\int_{x=0}^{\lambda\left(t,\mathcal{F}_{t-}^N\right)} \Pi(dt,dx)\right) U(t,ds), \end{cases}$$

in the weak sense with initial condition $\lim_{t\to 0^+} U(t,\cdot) = \delta_{-T_0}$.

Taking the expectation

[...] defining $u = "\mathbb{E}(U)"$, $u(t,.)$ distribution of S_{t-}.
Let $\left(\lambda(t, \mathcal{F}_{t-}^N)\right)_{t>0}$ be some non negative predictable process which is L^1_{loc} in expectation, and which admits a finite mean.
The measure $u = "\mathbb{E}(U)"$ satisfies the following system,

$$\begin{cases} (\partial_t + \partial_s) u(dt, ds) + \rho_{\lambda, \mathbb{P}_0}(t,s) u(dt, ds) = 0, \\ u(dt, 0) = \displaystyle\int_{s \in \mathbb{R}} \rho_{\lambda, \mathbb{P}_0}(t,s) u(t, ds)\, dt, \end{cases}$$

in the weak sense where $\rho_{\lambda, \mathbb{P}_0}(t,s) = \mathbb{E}\left[\lambda\left(t, \mathcal{F}_{t-}^N\right) \middle| S_{t-} = s\right]$ for almost every t. The initial condition $\lim_{t \to 0^+} u(t, \cdot)$ is given by the distribution of $-T_0$.

Law of large numbers

- Law of large numbers.
- Population-based approach.

Theorem

Let $(N^i)_{i\geq 1}$ be some i.i.d. point processes on $\mathbb{R}$ with L^1_{loc} intensity in expectation. For each i, let $\left(S^i_{t-}\right)_{t>0}$ denote the age process associated to N^i. Then, for every test function φ,

$$\int \varphi(t,s)\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{S^i_{t-}}(ds)\right)dt \xrightarrow[n\to\infty]{a.s.} \int \varphi(t,s)u(dt,ds),$$

with u satisfying the deterministic system.

Review of the examples

The system in expectation

$$\begin{cases} (\partial_t + \partial_s) u(dt, ds) + \rho_{\lambda, \mathbb{P}_0}(t, s)\, u(dt, ds) = 0, \\ u(dt, 0) = \displaystyle\int_{s \in \mathbb{R}} \rho_{\lambda, \mathbb{P}_0}(t, s)\, u(t, ds)\, dt. \end{cases}$$

where $\rho_{\lambda, \mathbb{P}_0}(t, s) = \mathbb{E}\left[\lambda\left(t, \mathcal{F}_{t-}^N\right) \middle| S_{t-} = s\right]$.

- This result may seem OK to a probabilist,
- But analysts need some explicit expression for ρ.
- In particular, this system may seem linear, but it is non-linear in general.

Review of the examples

The system in expectation

$$\begin{cases} (\partial_t + \partial_s) u(dt, ds) + \rho_{\lambda, \mathbb{P}_0}(t, s)\, u(dt, ds) = 0, \\ u(dt, 0) = \displaystyle\int_{s \in \mathbb{R}} \rho_{\lambda, \mathbb{P}_0}(t, s)\, u(t, ds)\, dt. \end{cases}$$

where $\rho_{\lambda, \mathbb{P}_0}(t, s) = \mathbb{E}\left[\lambda\left(t, \mathcal{F}_{t-}^N\right) \middle| S_{t-} = s\right]$.

- This result may seem OK to a probabilist,
- But analysts need some explicit expression for ρ.
- In particular, this system may seem linear, but it is non-linear in general.

- Poisson process. $\rightarrow \rho_{\lambda, \mathbb{P}_0}(t, s) = f(t)$.
- Renewal process. $\rightarrow \rho_{\lambda, \mathbb{P}_0}(t, s) = f(s)$.

Review of the examples

The system in expectation

$$\begin{cases} (\partial_t + \partial_s) u(dt, ds) + \rho_{\lambda, \mathbb{P}_0}(t, s)\, u(dt, ds) = 0, \\ u(dt, 0) = \displaystyle\int_{s \in \mathbb{R}} \rho_{\lambda, \mathbb{P}_0}(t, s)\, u(t, ds)\ dt. \end{cases}$$

where $\rho_{\lambda, \mathbb{P}_0}(t, s) = \mathbb{E}\left[\lambda\left(t, \mathcal{F}_{t-}^N\right) \middle| S_{t-} = s\right]$.

- This result may seem OK to a probabilist,
- But analysts need some explicit expression for ρ.
- In particular, this system may seem linear, but it is non-linear in general.

- Poisson process. $\rightarrow \rho_{\lambda, \mathbb{P}_0}(t, s) = f(t)$.
- Renewal process. $\rightarrow \rho_{\lambda, \mathbb{P}_0}(t, s) = f(s)$.
- Hawkes process. $\rightarrow \rho_{\lambda, \mathbb{P}_0}$ is much more complex.

For Hawkes

Recall that

$$\int_{-\infty}^{t-} h(t-x)N(dx) \quad \longleftrightarrow \quad \int_0^t d(x)m(t-x)dx = X(t).$$

For Hawkes

Recall that

$$\int_{-\infty}^{t-} h(t-x)N(dx) \quad \longleftrightarrow \quad \int_0^t d(x)m(t-x)dx = X(t).$$

What we expected

Replacement of $p(s, X(t))$ by

$$\mathbb{E}\left[\lambda(t, \mathcal{F}_{t-}^N)\right] = \mu + \int_0^t h(t-x)\, u(dx, 0) \longleftrightarrow X(t)$$

What we find

$p(s, X(t))$ is replaced by $\rho_{\lambda, \mathbb{P}_0}(t, s)$ which is the conditional expectation, not the full expectation.

Conclusions : work in progress of J. Chevallier

- Univariate Hawkes processes do NOT lead to a PPS of the form $p(s, X_t) = \nu + X_t$, but to another completely explicit closed PDE system (NB : Non Markovian process)

Conclusions : work in progress of J. Chevallier

- Univariate Hawkes processes do NOT lead to a PPS of the form $p(s, X_t) = \nu + X_t$, but to another completely explicit closed PDE system (NB : Non Markovian process)
- Interacting Hawkes processes without refractory periods DO lead to this intuition (mean field)
- Even true for more realistic models with refractory periods and special $p(s, X_t)$.
- Propagation of chaos, CLT ...

Open questions

- But then when we pick some neurons and measure their activity via electrodes, why do we measure interactions ?
- What kind of models would allow both : independence with most of the neurons, sparse dependence with some (functional connectivity) ?
- What can we exactly infer in the reconstructed graphs ?
- Can it be couple to a more global measure of activity (LFP) to infer more information ?

Many thanks to :

M. Albert, Y. Bouret, J. Chevallier, F. Delarue,
F. Grammont, T. Laloë, A. Rouis, C. Tuleau-Malot (Nice),
N.R Hansen (Copenhagen), V. Rivoirard (Dauphine),
M. Fromont (Rennes 2),
M. Doumic (Inria Rocquencourt), M. Cáceres (Granada),
L. Sansonnet (AgroParisTech),
S. Schbath (INRA Jouy-en-Josas), F. Picard (LBBE Lyon),
T. Bessaïh, R. Lambert, N. Leresche (Neuronal Networks and Physiopathological Rhythms, NPA, Paris 6)

and to Sylvie Méléard and Vincent Bansaye for this very nice opportunity !

Slides and references on `http://math.unice.fr/~reynaudb/`

Many thanks to :

M. Albert, Y. Bouret, J. Chevallier, F. Delarue,
F. Grammont, T. Laloë, A. Rouis, C. Tuleau-Malot (Nice),
N.R Hansen (Copenhagen), V. Rivoirard (Dauphine),
M. Fromont (Rennes 2),
M. Doumic (Inria Rocquencourt), M. Cáceres (Granada),
L. Sansonnet (AgroParisTech),
S. Schbath (INRA Jouy-en-Josas), F. Picard (LBBE Lyon),
T. Bessaïh, R. Lambert, N. Leresche (Neuronal Networks and Physiopathological Rhythms, NPA, Paris 6)

and to Sylvie Méléard and Vincent Bansaye for this very nice opportunity !

Slides and references on `http://math.unice.fr/ reynaudb/`