Detection of dependence between neurons and synchronization

P. Reynaud-Bouret

U. Côte d'Azur, CNRS

June 26-30th 2017

patricia.bouret@gmail.com

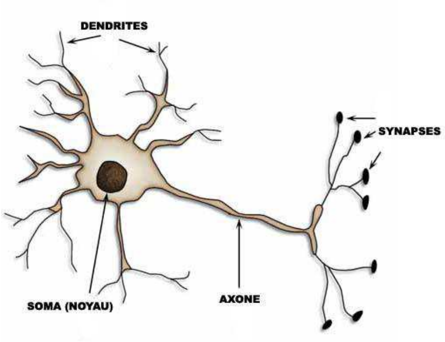
Contents

1. Problem
2. Detection of synchronization
3. Estimation of functional connectivity

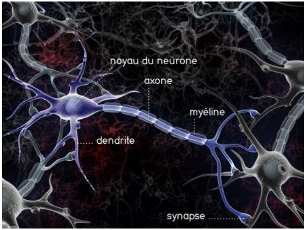
Contents

1. Problem
2. Detection of synchronization
3. Estimation of functional connectivity

Biological framework



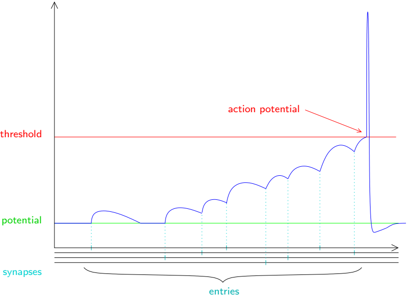
(a) One neuron



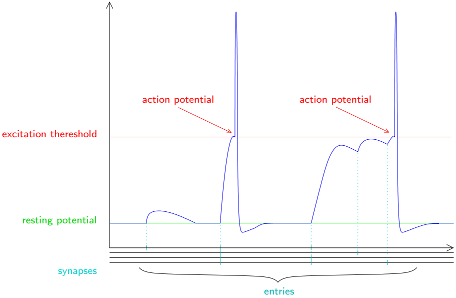
(b) Connected neurons

Action potentials and Synchronization

No synchronization

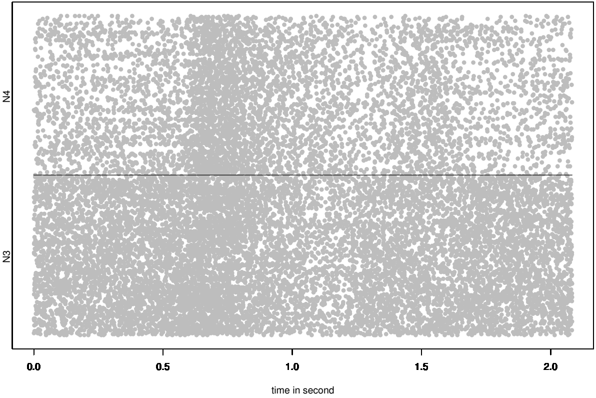


With synchronization



This is a model, nothing as simple is ever recorded !!!

Spike trains data



Contents

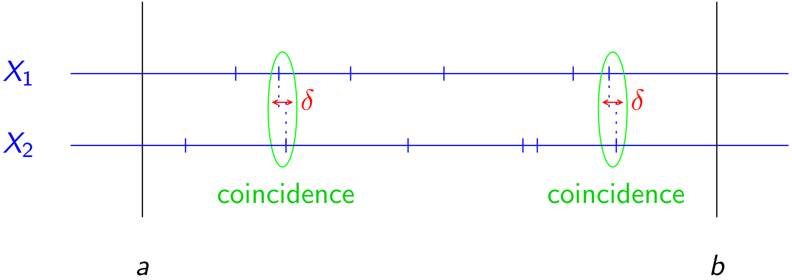
1 Problem

2 Detection of synchronization
- Unitary Events
- Multiple Testing
- Distribution under the null
- Bootstrap
- Back to tests
- Cross-trials non-stationarity

3 Estimation of functional connectivity

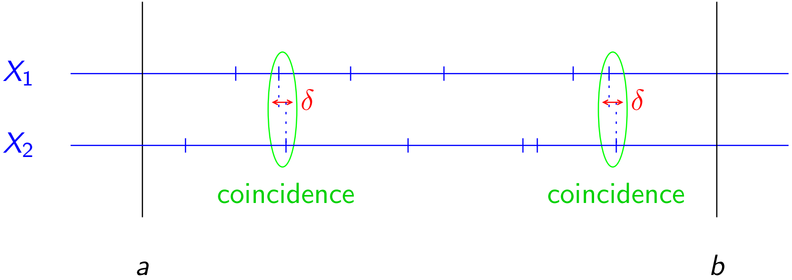
Pure coincidence or synchronization ?

- S. Grün (late 90's) introduced a method based on coincidences



Pure coincidence or synchronization ?

- S. Grün (late 90's) introduced a method based on coincidences



- If $\mathbf{C}^{obs} > c_{exp}$ on a small period of time, then one says that all coincidences are actually synchronizations and not pure coincidences (they do not just happen by chance !)

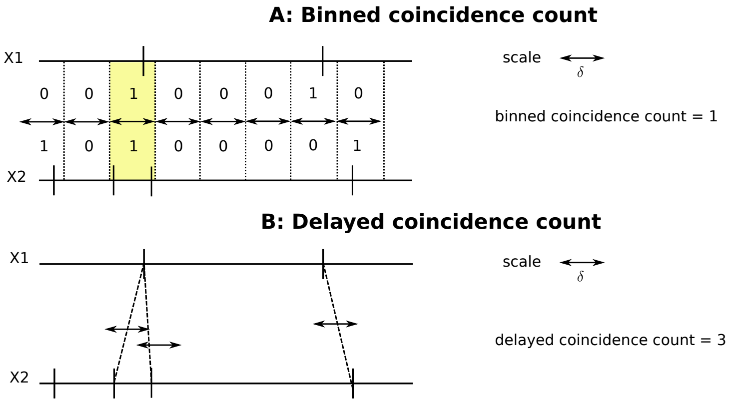
UE methods in a nutshell

Several variants in the past 20 years (mostly Grün and coauthors).

UE methods in a nutshell

Several variants in the past 20 years (mostly Grün and coauthors).

1. Choose a formula for coincidence count, $\varphi(X_1, X_2)$ (eventually more than 2 spike trains)

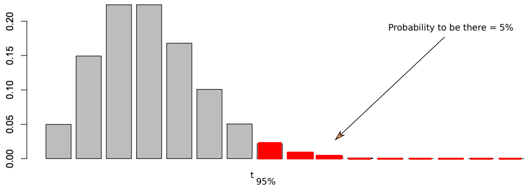


Sum up all the trials → $\mathbf{C}^{obs}$

UE methods in a nutshell

Several variants in the past 20 years (mostly Grün and coauthors).

1. Choose a formula for coincidence count, $\varphi(X_1, X_2)$ (eventually more than 2 spike trains)
   Sum up all the trials $\rightarrow \mathbf{C}^{obs}$
2. Take into account randomness to propose a possible distribution for pure coincidences $\rightarrow c_{exp}$ and p-values



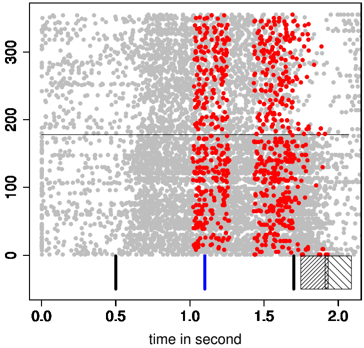
NB : p-value = "How likely is $\mathbf{C}^{obs}$ *as a pure coincidence ?"*

UE methods in a nutshell

Several variants in the past 20 years (mostly Grün and coauthors).

1. Choose a formula for coincidence count, $\varphi(X_1, X_2)$ (eventually more than 2 spike trains)
   Sum up all the trials $\rightarrow \mathbf{C}^{obs}$
2. Take into account randomness to propose a possible distribution for pure coincidences $\rightarrow c_{exp}$ and p-values
   NB : p-value = "How likely is $\mathbf{C}^{obs}$ *as a pure coincidence ?"*
3. Do this for several sliding windows and detect the ones where synchronization

Example



→ discuss all the points with a statistician perspective.

Statistical traduction of UE

- $X = (X_1, X_2)$ is a "typical" couple of spike trains.
  It is random and we are interested in its distribution.

Statistical traduction of UE

- $X = (X_1, X_2)$ is a "typical" couple of spike trains.
- We observe $\mathbb{X}_n$ the collection of n independent X^i that have the same distribution as X.

Statistical traduction of UE

- $X = (X_1, X_2)$ is a "typical" couple of spike trains.
- We observe $\mathbb{X}_n$ the collection of n independent X^i that have the same distribution as X.
- We want to know if X satisfies

$$H_0\text{: " } X_1 \text{ is independent of } X_2\text{".}$$

$\sim$ Pure coincidence

Statistical traduction of UE

- $X = (X_1, X_2)$ is a "typical" couple of spike trains.
- We observe $\mathbb{X}_n$ the collection of n independent X^i that have the same distribution as X.
- We want to know if X satisfies

$$H_0\text{: " } X_1 \text{ is independent of } X_2\text{".}$$

- UE = Statistical test at level $\alpha (= 5\%)$ = control the probability to declare synchronization whereas H_0 is satisfied (False Positive) $\rightarrow$ p-value p
  = the value of α to pass from accept to reject.

Statistical traduction of UE

- $X = (X_1, X_2)$ is a "typical" couple of spike trains.
- We observe $\mathbb{X}_n$ the collection of n independent X^i that have the same distribution as X.
- We want to know if X satisfies

$$H_0\text{: " } X_1 \text{ is independent of } X_2\text{"}.$$

- UE = Statistical test at level $\alpha (= 5\%)$ = control the probability to declare synchronization whereas H_0 is satisfied (False Positive) $\rightarrow$ p-value p
- Test at level $\alpha \Leftrightarrow p \leq \alpha$. Hence, (usually) under H_0,

$$\mathbb{P}(p \leq \alpha) \simeq \alpha,$$

Statistical traduction of UE

- $X = (X_1, X_2)$ is a "typical" couple of spike trains.
- We observe $\mathbb{X}_n$ the collection of n independent X^i that have the same distribution as X.
- We want to know if X satisfies

$$H_0\text{: " } X_1 \text{ is independent of } X_2\text{"}.$$

- UE = Statistical test at level $\alpha(= 5\%)$ = control the probability to declare synchronization whereas H_0 is satisfied (False Positive) $\rightarrow$ p-value p
- Test at level $\alpha \Leftrightarrow p \leq \alpha$.
- several sliding windows $\rightarrow$ multiple testing aspects.

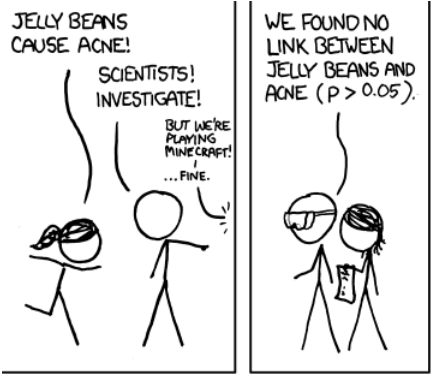
Contents

1 Problem

2 Detection of synchronization
- Unitary Events
- Multiple Testing
- Distribution under the null
- Bootstrap
- Back to tests
- Cross-trials non-stationarity

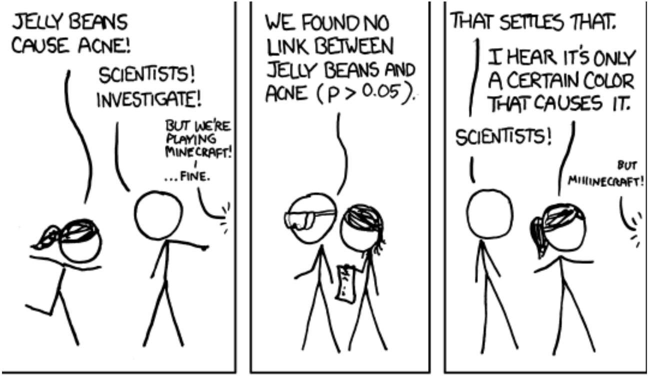
3 Estimation of functional connectivity

Intuitively



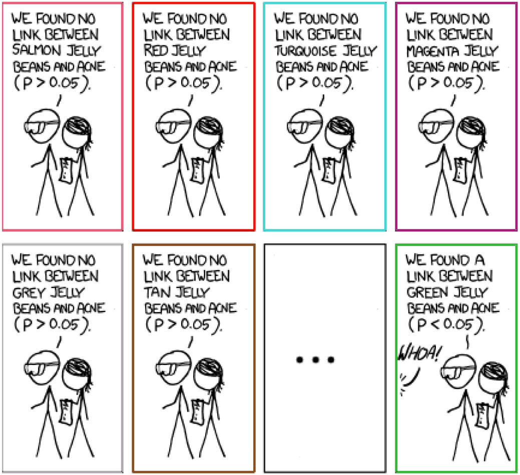
xkcd

Intuitively



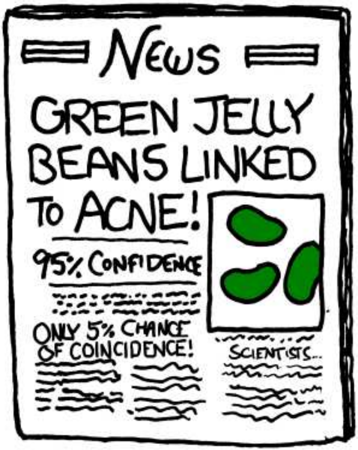
xkcd

Intuitively



xkcd

Intuitively



xkcd

More mathematically

If the K tests are independent, if all the null hypotheses hold and if all the first kind errors are equal to α, then

More mathematically

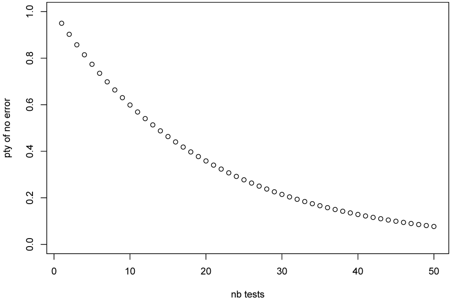
If the K tests are independent, if all the null hypotheses hold and if all the first kind errors are equal to α, then

$$\mathbb{P}(\text{no error}) = (1 - \alpha)^K$$

More mathematically

If the K tests are independent, if all the null hypotheses hold and if all the first kind errors are equal to α, then

$$\mathbb{P}(\text{no error}) = (1-\alpha)^K$$



What can we do ?

- Control of the probability to make even just one false positive by α $\to$ Bonferonni control (divide level by number of tests) [...]
- Control the mean proportion of false positive among the positive (FDR) $\to$ Benjamini-Hochberg method [...] $\to$ UE method proposed in Tuleau-Malot et al. (2014)

Contents

1. Problem
2. Detection of synchronization
   - Unitary Events
   - Multiple Testing
   - Distribution under the null
   - Bootstrap
   - Back to tests
   - Cross-trials non-stationarity
3. Estimation of functional connectivity

Expectation of the number of coincidences

If X_i stationary Poisson processes with intensity λ_i and if $X_1 \perp\!\!\!\perp X_2$

- for a fixed spike on X_1, there are about $2\delta\lambda_2$ spikes of X_2 in the vicinity
- there are about $(b-a)\lambda_1$ spikes of X_1 in a segment $[a,b]$.

$\rightsquigarrow \mathbb{E}_{\perp\!\!\!\perp}(\varphi(X_1,X_2)) \simeq 2\delta(b-a)\lambda_1\lambda_2$

The first studies mainly consisted in assuming that the coincidence count was Poisson with this mean.

Expectation of the number of coincidences

If X_i stationary Poisson processes with intensity λ_i and if $X_1 \perp\!\!\!\perp X_2$

- for a fixed spike on X_1, there are about $2\delta\lambda_2$ spikes of X_2 in the vicinity
- there are about $(b-a)\lambda_1$ spikes of X_1 in a segment $[a,b]$.

$\rightsquigarrow \mathbb{E}_{\perp\!\!\!\perp}(\varphi(X_1, X_2)) \simeq 2\delta(b-a)\lambda_1\lambda_2$

The first studies mainly consisted in assuming that the coincidence count was Poisson with this mean.

But, depending on the exact count, there are some edge or binning effects, e.g. with delayed coincidence count

$$\mathbb{E}_{\perp\!\!\!\perp}(\varphi(X_1, X_2)) = \lambda_1\lambda_2(2\delta(b-a) - \delta^2).$$

Expectation of the number of coincidences

If X_i stationary Poisson processes with intensity λ_i and if $X_1 \perp\!\!\!\perp X_2$

- for a fixed spike on X_1, there are about $2\delta\lambda_2$ spikes of X_2 in the vicinity
- there are about $(b-a)\lambda_1$ spikes of X_1 in a segment $[a,b]$.

$\rightsquigarrow \mathbb{E}_{\perp\!\!\!\perp}(\varphi(X_1, X_2)) \simeq 2\delta(b-a)\lambda_1\lambda_2$

The first studies mainly consisted in assuming that the coincidence count was Poisson with this mean.

But, depending on the exact count, there are some edge or binning effects, e.g. with delayed coincidence count

$$\mathbb{E}_{\perp\!\!\!\perp}(\varphi(X_1, X_2)) = \lambda_1\lambda_2(2\delta(b-a) - \delta^2).$$

$\mathrm{Var}_{\perp\!\!\!\perp}(\varphi(X^1, X^2)) = \lambda_1\lambda_2\left[2\delta(b-a) - \delta^2\right] + \left[\lambda_1^2\lambda_2 + \lambda_1\lambda_2^2\right]\left[4\delta^2(b-a) - \frac{10}{3}\delta^3\right].$

$\Longrightarrow$ Not a Poisson variable

Plug-in

- Find an (approximate) distribution of **C** under H_0
  For instance, Gaussian approximation via CLT

Plug-in

- Find an (approximate) distribution of **C** under H_0
  $\rightarrow$ dependency in unknown parameters such as firing rates etc ...
- Parameters can be estimated by using the trials.
- Plug them in.

Plug-in

- Find an (approximate) distribution of **C** under H_0
  $\rightarrow$ dependency in unknown parameters such as firing rates etc ...
- Parameters can be estimated by using the trials.
- Plug them in.

This may change (for instance) the variance : under independence, if delayed coincidence count [...], this

$$\frac{\sqrt{n}\left(C/n - (2\delta(b-a)-\delta^2)\hat{\lambda}_1\hat{\lambda}_2\right)}{\sqrt{\hat{\lambda}_1\hat{\lambda}_2[2\delta(b-a)-\delta^2+(\frac{2}{3}\delta^3-\frac{\delta^4}{b-a})(\hat{\lambda}_1+\hat{\lambda}_2)]}}$$

is approximately $\mathcal{N}(0,1)$.

Plug-in

- Find an (approximate) distribution of **C** under H_0
  $\rightarrow$ dependency in unknown parameters such as firing rates etc ...
- Parameters can be estimated by using the trials.
- Plug them in.

$\Rightarrow$ Plug-in can be really misleading.
$\Rightarrow$ correction by conditioning (+ ass.) Gütig et al. (2001)
$\Rightarrow$ correction under Poisson assumptions Tuleau et al. (2014)

Small illustration/centering issue

- $\mathbf{C} = \sum_{i=1}^{n} \varphi(X_1^i, X_2^i)$ is a sum of i.i.d. variables
  Central Limit Theorem $\Rightarrow$

$$\frac{\mathbf{C}(\mathbb{X}_n^{\perp\!\!\!\perp}) - c_{0,n}}{\sqrt{v_{0,n}}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

with $c_{0,n} = \mathbb{E}_{\perp\!\!\!\perp}[\mathbf{C}]$ and $v_{0,n} = \mathbb{E}_{\perp\!\!\!\perp}\left[(\mathbf{C} - c_{0,n})^2\right]$

Small illustration/centering issue

- $\mathbf{C} = \sum_{i=1}^{n} \varphi(X_1^i, X_2^i)$ is a sum of i.i.d. variables
  Central Limit Theorem $\Rightarrow$

$$\frac{\mathbf{C}(\mathbb{X}_n^{\perp\!\!\!\perp}) - c_{0,n}}{\sqrt{v_{0,n}}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

with $c_{0,n} = \mathbb{E}_{\perp\!\!\!\perp}[\mathbf{C}]$ and $v_{0,n} = \mathbb{E}_{\perp\!\!\!\perp}\left[(\mathbf{C} - c_{0,n})^2\right] \Rightarrow$

  - distrib. of $(\mathbf{C} - c_{0,n})/\sqrt{v_{0,n}}$ close to $\mathcal{N}(0,1)$
  - distrib. $\mathbf{C} - c_{0,n}$ close to $\mathcal{N}(0, v_{0,n})$
  - distrib. $\mathbf{C}$ distribution under independence is close to $\mathcal{N}(c_{0,n}, v_{0,n})$

Small illustration/centering issue

- $\mathbf{C} = \sum_{i=1}^{n} \varphi(X_1^i, X_2^i)$ is a sum of i.i.d. variables
  Central Limit Theorem $\Rightarrow$

$$\frac{\mathbf{C}(\mathbb{X}_n^{\perp\!\!\!\perp}) - c_{0,n}}{\sqrt{v_{0,n}}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

  with $c_{0,n} = \mathbb{E}_{\perp\!\!\!\perp}[\mathbf{C}]$ and $v_{0,n} = \mathbb{E}_{\perp\!\!\!\perp}\left[(\mathbf{C} - c_{0,n})^2\right] \Rightarrow$
  - distrib. of $(\mathbf{C} - c_{0,n})/\sqrt{v_{0,n}}$ close to $\mathcal{N}(0,1)$
  - distrib. $\mathbf{C} - c_{0,n}$ close to $\mathcal{N}(0, v_{0,n})$
  - distrib. $\mathbf{C}$ distribution under independence is close to $\mathcal{N}(c_{0,n}, v_{0,n})$
- $c_{0,n}$ is unknown but can be estimated by

$$\hat{\mathbf{C}}_0(\mathbb{X}_n) = \frac{1}{n-1} \sum_{i \neq i'} \varphi(X_1^i, X_2^{i'}).$$

Small illustration/centering issue

- $\mathbf{C} = \sum_{i=1}^{n} \varphi(X_1^i, X_2^i)$ is a sum of i.i.d. variables
  Central Limit Theorem $\Rightarrow$

$$\frac{\mathbf{C}(\mathbb{X}_n^{\perp\!\!\perp}) - c_{0,n}}{\sqrt{v_{0,n}}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

  with $c_{0,n} = \mathbb{E}_{\perp\!\!\perp}[\mathbf{C}]$ and $v_{0,n} = \mathbb{E}_{\perp\!\!\perp}\left[(\mathbf{C} - c_{0,n})^2\right] \Rightarrow$
  - distrib. of $(\mathbf{C} - c_{0,n})/\sqrt{v_{0,n}}$ close to $\mathcal{N}(0,1)$
  - distrib. $\mathbf{C} - c_{0,n}$ close to $\mathcal{N}(0, v_{0,n})$
  - distrib. $\mathbf{C}$ distribution under independence is close to $\mathcal{N}(c_{0,n}, v_{0,n})$
- $c_{0,n}$ is unknown but can be estimated by

$$\hat{\mathbf{C}}_0(\mathbb{X}_n) = \frac{1}{n-1} \sum_{i \neq i'} \varphi(X_1^i, X_2^{i'}).$$

- The fluctuations of $\hat{\mathbf{C}}_0(\mathbb{X}_n)$ are as large as the ones of $\mathbf{C}$!

Consequences

- Plug-in of the mean estimator changes the variance !

$$\frac{\mathbf{C}(\mathbb{X}_n^{\perp\!\!\!\perp}) - \hat{\mathbf{C}}_0(\mathbb{X}_n^{\perp\!\!\!\perp})}{\sqrt{w_{0,n}}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1)$$

and $w_{0,n}$ is NOT $v_{0,n}$!

Consequences

- Plug-in of the mean estimator changes the variance !

$$\frac{\mathbf{C}(\mathbb{X}_n^{\perp\!\!\!\perp}) - \hat{\mathbf{C}}_0(\mathbb{X}_n^{\perp\!\!\!\perp})}{\sqrt{w_{0,n}}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1)$$

  and $w_{0,n}$ is NOT $v_{0,n}$!

- Plug-in of variance estimators are usually ok $\rightsquigarrow$

$$\mathbf{Z}(\mathbb{X}_n^{\perp\!\!\!\perp}) = \frac{\mathbf{C}(\mathbb{X}_n^{\perp\!\!\!\perp}) - \hat{\mathbf{C}}_0(\mathbb{X}_n^{\perp\!\!\!\perp})}{\sqrt{n}\hat{\sigma}(\mathbb{X}_n^{\perp\!\!\!\perp})} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

Consequences

- Plug-in of the mean estimator changes the variance !

$$\frac{\mathbf{C}(\mathbb{X}_n^{\perp\!\!\!\perp}) - \hat{\mathbf{C}}_0(\mathbb{X}_n^{\perp\!\!\!\perp})}{\sqrt{w_{0,n}}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1)$$

  and $w_{0,n}$ is NOT $v_{0,n}$!

- Plug-in of variance estimators are usually ok $\rightsquigarrow$

$$\mathbf{Z}(\mathbb{X}_n^{\perp\!\!\!\perp}) = \frac{\mathbf{C}(\mathbb{X}_n^{\perp\!\!\!\perp}) - \hat{\mathbf{C}}_0(\mathbb{X}_n^{\perp\!\!\!\perp})}{\sqrt{n}\hat{\sigma}(\mathbb{X}_n^{\perp\!\!\!\perp})} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

- distrib. of $\mathbf{U}^{\perp\!\!\!\perp} = \mathbf{U}(\mathbb{X}_n^{\perp\!\!\!\perp}) = \mathbf{C}(\mathbb{X}_n^{\perp\!\!\!\perp}) - \hat{\mathbf{C}}_0(\mathbb{X}_n^{\perp\!\!\!\perp})$ close to $\mathcal{N}(0, n\hat{\sigma}^2(\mathbb{X}_n^{\perp\!\!\!\perp}))$.

Consequences

- Plug-in of the mean estimator changes the variance !

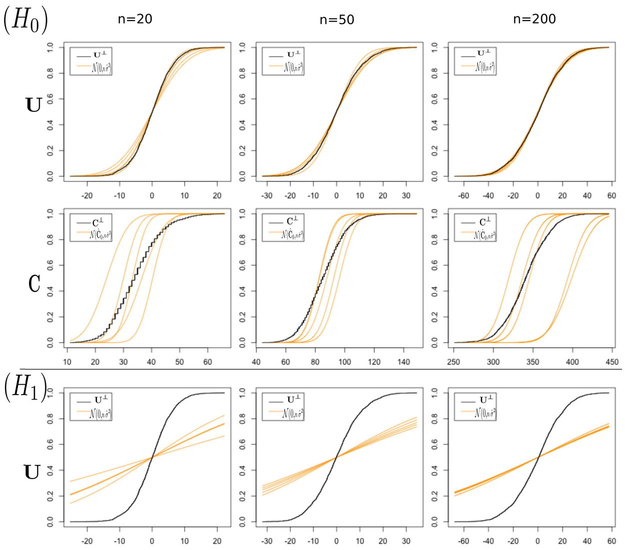
$$\frac{\mathbf{C}(\mathbb{X}_n^{\perp\!\!\perp}) - \hat{\mathbf{C}}_0(\mathbb{X}_n^{\perp\!\!\perp})}{\sqrt{w_{0,n}}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1)$$

  and $w_{0,n}$ is NOT $v_{0,n}$!
- Plug-in of variance estimators are usually ok $\rightsquigarrow$

$$\mathbf{Z}(\mathbb{X}_n^{\perp\!\!\perp}) = \frac{\mathbf{C}(\mathbb{X}_n^{\perp\!\!\perp}) - \hat{\mathbf{C}}_0(\mathbb{X}_n^{\perp\!\!\perp})}{\sqrt{n}\hat{\sigma}(\mathbb{X}_n^{\perp\!\!\perp})} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

- distrib. of $\mathbf{U}^{\perp\!\!\perp} = \mathbf{U}(\mathbb{X}_n^{\perp\!\!\perp}) = \mathbf{C}(\mathbb{X}_n^{\perp\!\!\perp}) - \hat{\mathbf{C}}_0(\mathbb{X}_n^{\perp\!\!\perp})$ close to $\mathcal{N}(0, n\hat{\sigma}^2(\mathbb{X}_n^{\perp\!\!\perp}))$.
- But STOP there! the distrib. of $\mathbf{C}(\mathbb{X}_n^{\perp\!\!\perp})$ is NOT close to $\mathcal{N}(\hat{\mathbf{C}}_0(\mathbb{X}_n^{\perp\!\!\perp}), \sqrt{n}\hat{\sigma}(\mathbb{X}_n^{\perp\!\!\perp}))$.

Illustration / the centering issue



Contents

1 Problem

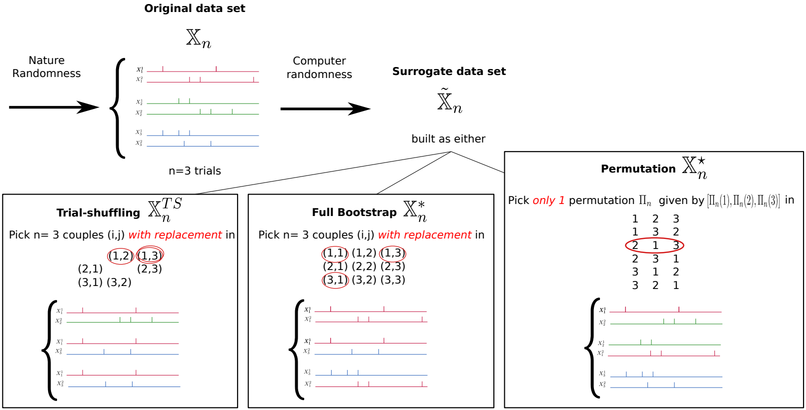
2 Detection of synchronization
- Unitary Events
- Multiple Testing
- Distribution under the null
- Bootstrap
- Back to tests
- Cross-trials non-stationarity

3 Estimation of functional connectivity

Why bootstrap ?

- Classical CLT $\Rightarrow$ approximations valid only large n (in practice $n = 20$ usual)
- Bootstrap or resampling (Efron, 79 [...]) are known technics in statistics to lead to much better approximations (valid for smaller values of n)
- In neuroscience, they are usually known under the name "Trial-Shuffling". In the sequel, TS refers to Pipa and Grün version (2003).

Principle



Unconditional distribution: *all* possible *choices* of both *Nature and Computer* randomness

Conditional distribution: *1 fixed* original data set (*Nature* randomness), *all* possible *choices* of *Computer* randomness

Discussion on the unconditional distribution

- Trial-Shuffling : $\mathbb{E}[\mathbf{C}^{TS}] = c_{0,n} = \mathbb{E}_{\perp\!\!\!\perp}[\mathbf{C}]$

Discussion on the unconditional distribution

- Trial-Shuffling : $\mathbb{E}[\mathbf{C}^{TS}] = c_{0,n} = \mathbb{E}_{\perp\!\!\!\perp}[\mathbf{C}]$ but because couples may be picked twice, distrib. of $\mathbf{C}^{TS}$ and $\mathbf{C}^{\perp\!\!\!\perp}$ are different

Discussion on the unconditional distribution

- Trial-Shuffling : $\mathbb{E}[\mathbf{C}^{TS}] = c_{0,n} = \mathbb{E}_{\perp\!\!\!\perp}[\mathbf{C}]$ but because couples may be picked twice, distrib. of $\mathbf{C}^{TS}$ and $\mathbf{C}^{\perp\!\!\!\perp}$ are different
- Full Boostrap : $\mathbb{E}_{\perp\!\!\!\perp}[\mathbf{C}^*] = \mathbb{E}_{\perp\!\!\!\perp}[\mathbf{C}]$ but not true if not H_0.

Discussion on the unconditional distribution

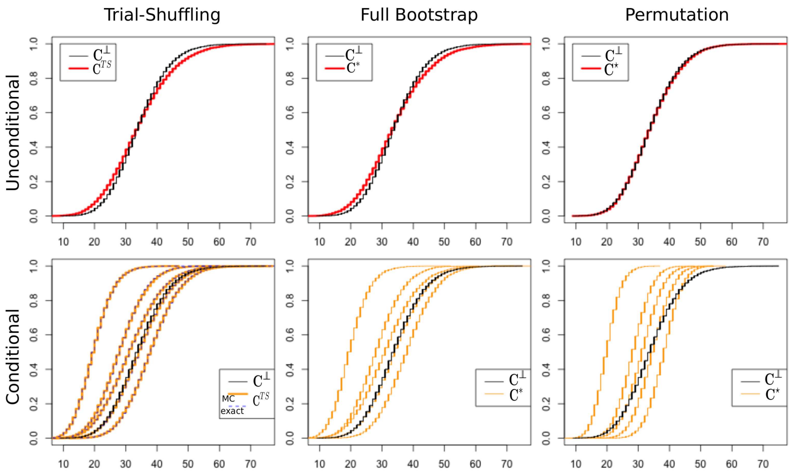
- Trial-Shuffling : $\mathbb{E}[\mathbf{C}^{TS}] = c_{0,n} = \mathbb{E}_{\perp\!\!\!\perp}[\mathbf{C}]$ but because couples may be picked twice, distrib. of $\mathbf{C}^{TS}$ and $\mathbf{C}^{\perp\!\!\!\perp}$ are different
- Full Boostrap : $\mathbb{E}_{\perp\!\!\!\perp}[\mathbf{C}^{*}] = \mathbb{E}_{\perp\!\!\!\perp}[\mathbf{C}]$ but not true if not H_0. distrib. are different.
- Permutation : $\mathbb{E}[\mathbf{C}^{\star}] = \mathbb{E}[\mathbf{C}^{*}]$.

Discussion on the unconditional distribution

- Trial-Shuffling : $\mathbb{E}[\mathbf{C}^{TS}] = c_{0,n} = \mathbb{E}_{\perp\!\!\!\perp}[\mathbf{C}]$ but because couples may be picked twice, distrib. of $\mathbf{C}^{TS}$ and $\mathbf{C}^{\perp\!\!\!\perp}$ are different
- Full Boostrap : $\mathbb{E}_{\perp\!\!\!\perp}[\mathbf{C}^*] = \mathbb{E}_{\perp\!\!\!\perp}[\mathbf{C}]$ but not true if not H_0.
  distrib. are different.
- Permutation : $\mathbb{E}[\mathbf{C}^\star] = \mathbb{E}[\mathbf{C}^*]$.
  distrib. are equal if under H_0.

But because in practice, only 1 "Nature" observation, in fact we only have access to Conditional Distribution

Illustration under H_0



Monte-Carlo with $B = 10000$ replicates for the bootstrap distributions.

Why conditional distributions are so wrong?

Proofs of convergence of bootstrap procedures mainly exists for centered quantities.

Why conditional distributions are so wrong?

Proofs of convergence of bootstrap procedures mainly exists for centered quantities.

The bootstrap of the mean in a nutshell

$\mathbb{Y}_n = (Y_1, ..., Y_n)$ are i.i.d. variables.
The distrib. of $\bar{Y} - \mathbb{E}(\bar{Y})$ (empirical mean - expectation) is correctly approximated by

Why conditional distributions are so wrong?

Proofs of convergence of bootstrap procedures mainly exists for centered quantities.

The bootstrap of the mean in a nutshell

$\mathbb{Y}_n = (Y_1, ..., Y_n)$ are i.i.d. variables.
The distrib. of $\bar{Y} - \mathbb{E}(\bar{Y})$ (empirical mean - expectation) is correctly approximated by

- replace "empirical mean" by "empirical bootstrap mean"
- replace "expectation" by "conditional expectation"

$\Rightarrow$ the distrib. of $\bar{Y}^* - \mathbb{E}(\bar{Y}^* | \mathbb{Y}_n)$.

Center !

If one wants to use bootstrap distrib. for testing, it is better if one knows the expectation under H_0

$\rightsquigarrow$ Centered quantities, i.e. $\mathbf{U} = \mathbf{C} - \hat{\mathbf{C}}_0(\mathbb{X}_n)$

- Full Bootstrap and Permutation : $\mathbb{E}(\mathbf{U}(\mathbb{X}_n^*)|\mathbb{X}_n) = \mathbb{E}(\mathbf{U}(\mathbb{X}_n^\star)|\mathbb{X}_n) = 0$.
  $\rightsquigarrow$ conditional distrib. of $\mathbf{U}(\mathbb{X}_n^*)$ and $\mathbf{U}(\mathbb{X}_n^\star)$ should be good approx. of the distrib. of $\mathbf{U}^{\perp\!\!\!\perp}$.

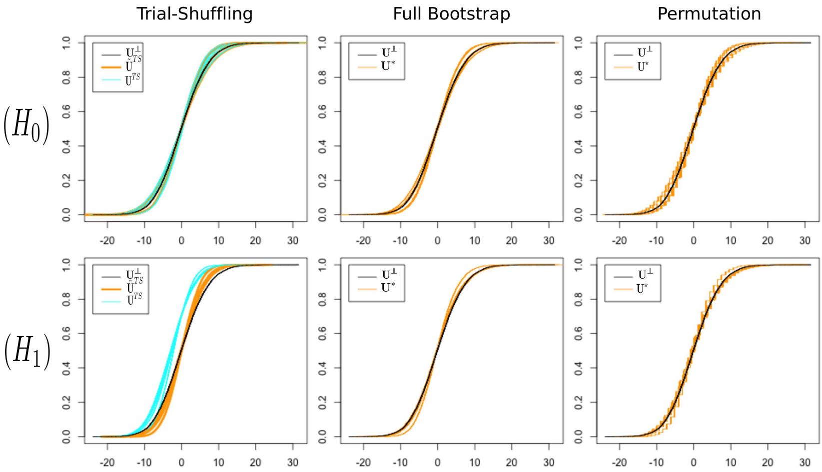
Center !

If one wants to use bootstrap distrib. for testing, it is better if one knows the expectation under H_0

$\rightsquigarrow$ Centered quantities, i.e. $\mathbf{U} = \mathbf{C} - \hat{\mathbf{C}}_0(\mathbb{X}_n)$

- Full Bootstrap and Permutation : $\mathbb{E}(\mathbf{U}(\mathbb{X}_n^*)|\mathbb{X}_n) = \mathbb{E}(\mathbf{U}(\mathbb{X}_n^\star)|\mathbb{X}_n) = 0$.
  $\rightsquigarrow$ conditional distrib. of $\mathbf{U}(\mathbb{X}_n^*)$ and $\mathbf{U}(\mathbb{X}_n^\star)$ should be good approx. of the distrib. of $\mathbf{U}^{\perp\!\!\!\perp}$.
- Trial-Shuffling : $\mathbb{E}(\mathbf{U}(\mathbb{X}_n^{TS})|\mathbb{X}_n) = -\mathbf{U}(\mathbb{X}_n)/n$.
  $\rightsquigarrow$ conditional distrib. of $\tilde{\mathbf{U}}^{TS} = \mathbf{U}(\mathbb{X}_n^{TS}) + \mathbf{U}(\mathbb{X}_n)/n$ should be good approx. of the distrib. of $\mathbf{U}^{\perp\!\!\!\perp}$.

Illustration



Monte-Carlo with $B = 10000$ replicates for the bootstrap distributions. H_1 is $X_1 = X_2$

Contents

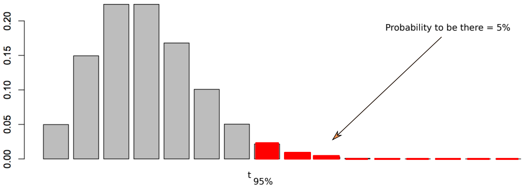
1 Problem

2 Detection of synchronization
- Unitary Events
- Multiple Testing
- Distribution under the null
- Bootstrap
- Back to tests
- Cross-trials non-stationarity

3 Estimation of functional connectivity

What about the resulting tests ?

- distrib. under the null → a critical value for rejection → a p-value p



- If critical value = quantile of level α, then the p-value satisfies

$$\mathbb{P}_{\perp\!\!\!\perp}(p \leq \alpha) \simeq \alpha.$$

$\rightsquigarrow$ p-values are uniform under the null. Their c.d.f should be the diagonal of $[0,1]^2$ under H_0

What about the resulting tests ?

- distrib. under the null $\rightarrow$ a critical value for rejection $\rightarrow$ a p-value p
- If critical value = quantile of level α, then the p-value satisfies

$$\mathbb{P}_{\perp\!\!\!\perp}(p \leq \alpha) \simeq \alpha.$$

$\rightsquigarrow$ p-values are uniform under the null. Their c.d.f should be the diagonal of $[0,1]^2$ under H_0

- If not under H_0, p-values should be small, hence c.d.f. above the diagonal.

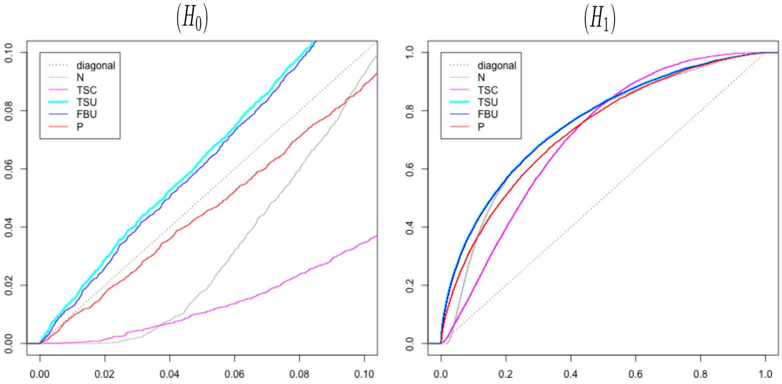
What about the resulting tests ?

- distrib. under the null $\rightarrow$ a critical value for rejection $\rightarrow$ a p-value p
- If critical value = quantile of level α, then the p-value satisfies

$$\mathbb{P}_{\perp\!\!\!\perp}(p \leq \alpha) \simeq \alpha.$$

  $\rightsquigarrow$ p-values are uniform under the null. Their c.d.f should be the diagonal of $[0,1]^2$ under H_0
- If not under H_0, p-values should be small, hence c.d.f. above the diagonal.
- NB: some tests (the ones based on CLT or Permutation) have their **U** or **C** versions equivalent. This is due to some very particular properties of invariance (by shift for Gaussian approx. or by permutation). No equivalence in general.

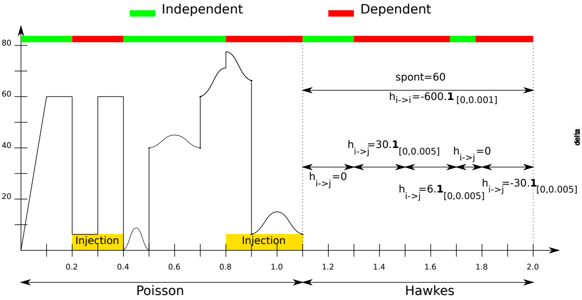
Illustration



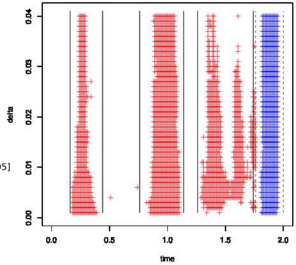
$n = 20$ trials . H_1 is injection models.

Permutation UE

Permutation tests + BH on sliding windows
(`https://github.com/ybouret/neuro-stat`)



A: Description of Experiment 1



B: Result of the permutation method

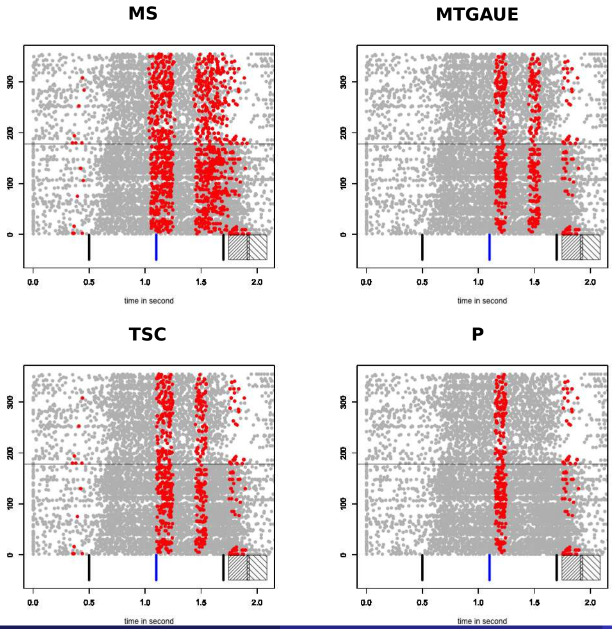
Permutation UE

Permutation tests + BH on sliding windows
(https://github.com/ybouret/neuro-stat)

	Independ.	Depend.	Total
Rejected	V	S	R
Accepted	U	T	$m-R$
Total	m_0	$m-m_0$	m

	Experiment 1		Experiment 2	
	$FDR = \mathbb{E}(V/R)$	$FNDR = \mathbb{E}(T/(m-R))$	FDR	FNDR
MTGAUE	0.10	0.17	0.04	0
TSC	0.01	0.26	0.25	0
TSC + BH	0	0.32	0	0
P	0.01	0.23	0.02	0

And on real data...



Contents

1. Problem

2. Detection of synchronization
   - Unitary Events
   - Multiple Testing
   - Distribution under the null
   - Bootstrap
   - Back to tests
   - Cross-trials non-stationarity

3. Estimation of functional connectivity

Mathematical assumptions

- Full Bootstrap and Permutation are particular cases of Albert et al. (2015), where the convergence of the renormalized bootstrap distribution of $\mathbf{U}$ is proved. Typically, one can show that

$$d_{W_2}\left(\mathcal{L}\left(\frac{\mathbf{U}^\star}{\sqrt{n}}\middle|\mathbb{X}_n\right), \mathcal{L}\left(\frac{\mathbf{U}^{\perp\!\!\!\perp}}{\sqrt{n}}\right)\right) \xrightarrow[n\to+\infty]{\mathbb{P}} 0$$

.

- Assumptions = i.i.d. trials and some moment assumptions

Mathematical assumptions

- Full Bootstrap and Permutation are particular cases of Albert et al. (2015), where the convergence of the renormalized bootstrap distribution of $\mathbf{U}$ is proved.
- Assumptions = i.i.d. trials and some moment assumptions
- Moment assumptions $\simeq$ no explosion of the spike trains $\rightsquigarrow$ biologically reasonable.
- independent trials : usually the animal rests between trials and experiments are past training phase
  $\rightarrow$ biologically reasonable
- identically distributed trials $\rightarrow$ maybe less biologically reasonable w.r.t. cross-trial non-stationarity ?

Statistical definition of cross-trial non-stationarity ?

- Number of spikes varies $\rightarrow$ random

Statistical definition of cross-trial non-stationarity ?

- Number of spikes varies $\rightarrow$ random
- Fano factor $FF = Var(N)/\mathbb{E}(N) \neq 1 \rightarrow$ not Poisson

Statistical definition of cross-trial non-stationarity ?

- Number of spikes varies $\rightarrow$ random
- Fano factor $FF = Var(N)/\mathbb{E}(N) \neq 1$ $\rightarrow$ not Poisson
- $FF \neq Var(ISI)$ (ISI = InterSpike Interval) $\rightarrow$ correlated ISI

Statistical definition of cross-trial non-stationarity ?

- Number of spikes varies $\rightarrow$ random
- Fano factor $FF = Var(N)/\mathbb{E}(N) \neq 1 \rightarrow$ not Poisson
- $FF \neq Var(ISI)$ (ISI = InterSpike Interval) $\rightarrow$ correlated ISI
- Turn to the models that are qualified of "cross-trial non-stationarity" : hidden command (see the "intensity command" of Churchland et al, 2011)

Hidden command

The distribution of (X_1, X_2) is not given intrinsically but given another variable Y (the hidden command, usually not observed)

- change of firing rates (intensities)
- anesthesia variation, degree of decision making
- stimulus variability, oscillatory potential influencing both neurons

Common or non common hidden command

- hidden command Y common to both neurons $\Rightarrow$ globally dependent (tests reject with high proba) even if $X_1 \perp\!\!\!\perp X_2$ given Y.

Common or non common hidden command

- hidden command Y common to both neurons $\Rightarrow$ globally dependent (tests reject with high proba) even if $X_1 \perp\!\!\!\perp X_2$ given Y.
- hidden command $Y = (Y_1, Y_2)$ non common, $Y_1 \perp\!\!\!\perp Y_2$ and $X_1 \perp\!\!\!\perp X_2$ given Y $\Rightarrow$ independent (tests accept with high proba).

If you need to be convinced

Let A, B be two measurable sets

$$\begin{aligned}\mathbb{P}\left(X^1 \in A, X^2 \in B\right) &= \mathbb{E}\left[\mathbb{P}\left(X^1 \in A, X^2 \in B | Y\right)\right] \\ &= \mathbb{E}\left[\mathbb{P}\left(X^1 \in A | Y\right)\mathbb{P}\left(X^2 \in B | Y\right)\right] \text{ since } X_1 \perp\!\!\!\perp X_2 \text{ given } Y\end{aligned}$$

If you need to be convinced

Let A, B be two measurable sets

$$\mathbb{P}\left(X^1 \in A, X^2 \in B\right) = \mathbb{E}\left[\mathbb{P}\left(X^1 \in A, X^2 \in B | Y\right)\right]$$
$$= \mathbb{E}\left[\mathbb{P}\left(X^1 \in A | Y\right)\mathbb{P}\left(X^2 \in B | Y\right)\right] \text{ since } X_1 \perp\!\!\!\perp X_2 \text{ given } Y$$

$$= \mathbb{E}\left[\mathbb{P}\left(X^1 \in A | Y^1\right)\mathbb{P}\left(X^2 \in B | Y^2\right)\right]$$

since there is no common command variable

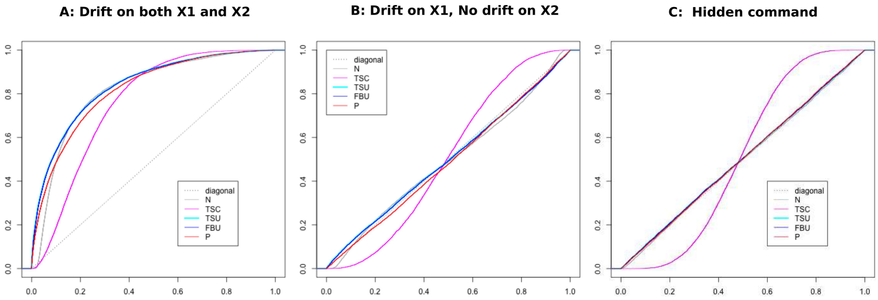
If you need to be convinced

Let A, B be two measurable sets

$$\mathbb{P}\left(X^1 \in A, X^2 \in B\right) = \mathbb{E}\left[\mathbb{P}\left(X^1 \in A, X^2 \in B | Y\right)\right]$$
$$= \mathbb{E}\left[\mathbb{P}\left(X^1 \in A | Y\right)\mathbb{P}\left(X^2 \in B | Y\right)\right] \text{ since } X_1 \perp\!\!\!\perp X_2 \text{ given } Y$$
$$= \mathbb{E}\left[\mathbb{P}\left(X^1 \in A | Y^1\right)\mathbb{P}\left(X^2 \in B | Y^2\right)\right]$$
$$\text{since there is no common command variable}$$
$$= \mathbb{E}\left[\mathbb{P}\left(X^1 \in A | Y^1\right)\right]\mathbb{E}\left[\mathbb{P}\left(X^2 \in B | Y^2\right)\right] \text{ since } Y^1 \perp\!\!\!\perp Y^2$$
$$= \mathbb{P}\left(X^1 \in A\right)\mathbb{P}\left(X^2 \in B\right).$$

Hence X_1 and X_2 are independent !

Illustration



Partial Conclusions

- Possible to detect global dependency between point processes (spike trains) in a complete distribution free manners.
- Global dependency does not really mean synchronization even if interest per se.
- More local structures usually need models : Hawkes (Hansen et al.) , loglinear (Kass et al.) [...]
- Maybe dithering can help (Louis et al. 2010, procedure implemented in Human Brain Project) ....

References

- Grün, S., Diesmann, M., & Aertsen, A.M.*Unitary Events Analysis. In Analysis of Parallel Spike Trains*, Grün, S., & Rotter, S., Springer Series in Computational Neuroscience (2010).
- Tuleau-Malot, C., Rouis, A., Grammont, F., Reynaud-Bouret, P. *Multiple tests based on a Gaussian Approximation of the Unitary Events*, Neural Computations (2014)
- Albert, M., Bouret, Y. , Fromont, M., Reynaud-Bouret, P. *Bootstrap and permutation tests of independence for point processes*, Annals of Statistics (2015).
- Albert, M., Bouret, Y. , Fromont, M., Reynaud-Bouret, P. *Surrogate data methods on a shuffling of the trials for synchrony detection: the centering issue*, Neural Computation (2016).
- Chevallier, J., Laloë, T. *Detection of dependence patterns with delay*, Biometrical Journal (2014).

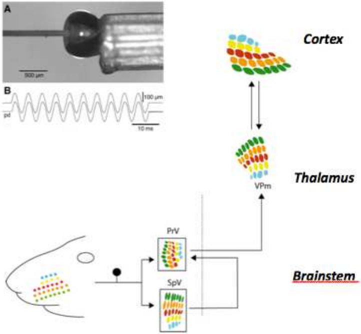
Contents

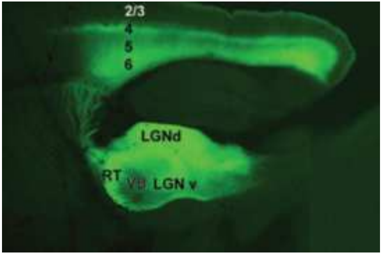
1 Problem

2 Detection of synchronization

3 Estimation of functional connectivity
- A slightly different problem
- Hawkes processes
- Estimation methods
- Simulations
- On real data

Description of the experiment (Team RNRP of Paris 6)

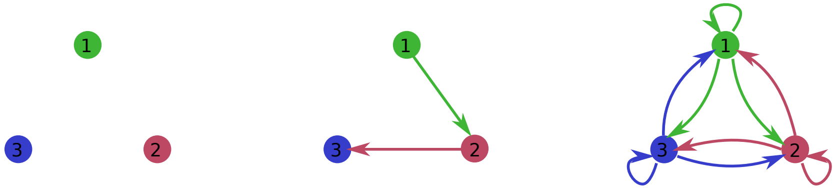




$\leadsto$ move one or two vibrissa (whiskers) at different frequencies and see how it is coded in the cortical barrels (layer 4).

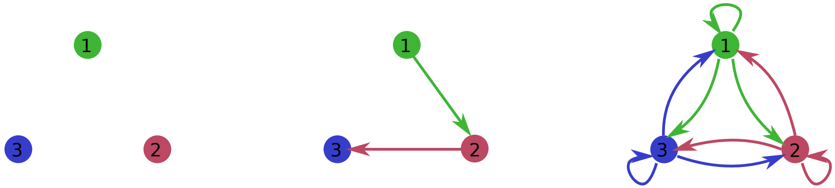
Visual aim

We want to produce that:



Visual aim

We want to produce that:



For a biologist, this is not a graph with a physical meaning but with a functional meaning (functional connectivity).
What is the mathematical meaning of it ?

Contents

1 Problem

2 Detection of synchronization

3 Estimation of functional connectivity
- A slightly different problem
- Hawkes processes
- Estimation methods
- Simulations
- On real data

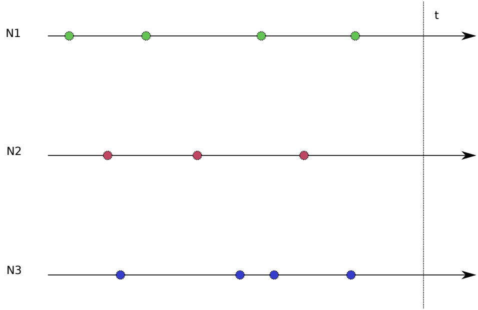
Point processes and conditional intensity

dN point measure $= \sum_{T \in N} \delta_T$

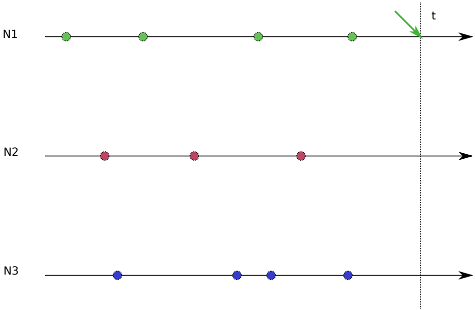
$$\underbrace{dN_t}_{\substack{\text{Nbr observed points} \\ \text{in } [t, t+dt]}} = \underbrace{\overbrace{\lambda(t)}^{\textbf{intensity}} \; dt}_{\substack{\text{Expected Number} \\ \text{given the past before } t}} + \underbrace{\text{noise}}_{\substack{\textit{Martingales} \\ \textit{differences}}}$$

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

Multivariate Hawkes processes



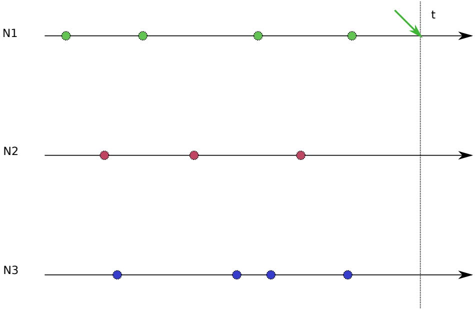
Multivariate Hawkes processes



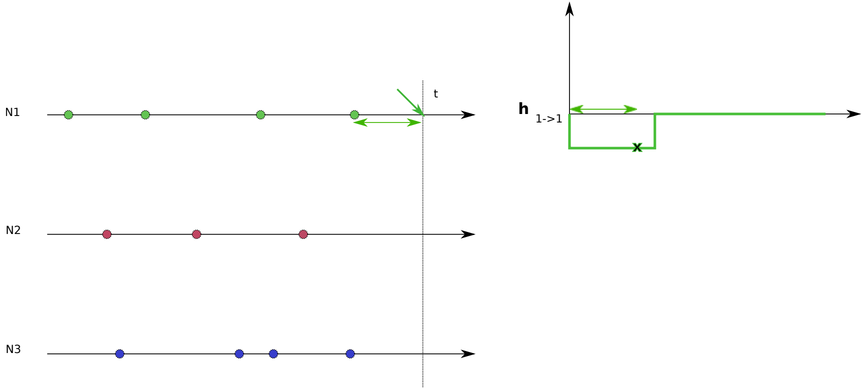
$$\underbrace{\lambda^{(1)}(t) \quad = \quad \nu}_{}$$

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

Multivariate Hawkes processes

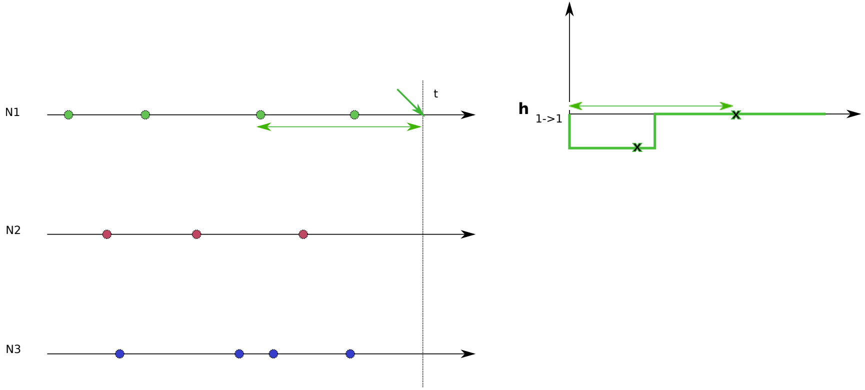


$$\lambda^{(1)}(t) = \underbrace{\nu}_{\text{spontaneous}} + \underbrace{\sum_{\substack{T<t \\ \text{pt de N1}}} h_{1\to 1}(t-T)}_{\text{self-interaction}}$$



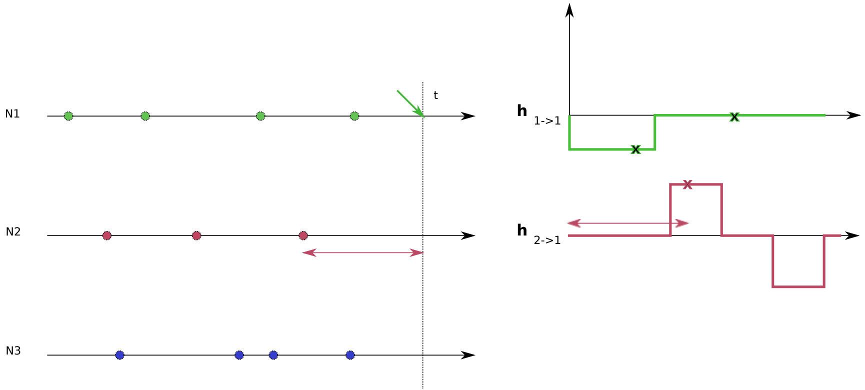
$$\lambda^{(1)}(t) = \underbrace{\nu}_{\text{spontaneous}} + \underbrace{\sum_{\substack{T<t \\ \text{pt de N1}}} h_{1\to 1}(t-T)}_{\text{self-interaction}}$$

Multivariate Hawkes processes



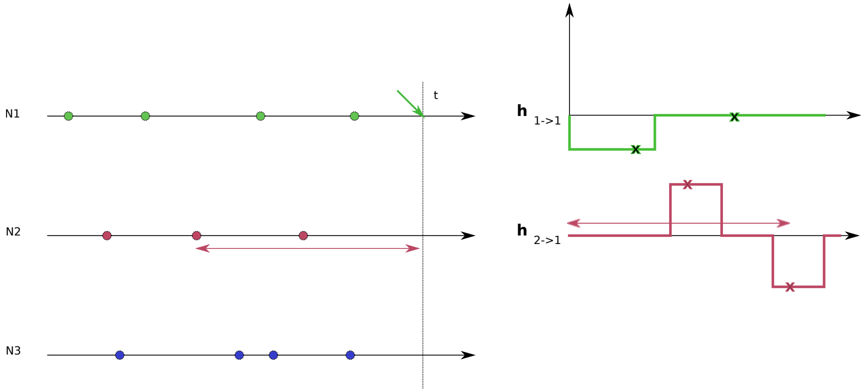
$$\lambda^{(1)}(t) = \underbrace{\nu}_{\text{spontaneous}} + \underbrace{\sum_{\substack{T<t \\ \text{pt de N1}}} h_{1\to 1}(t-T)}_{\text{self-interaction}}$$

Multivariate Hawkes processes



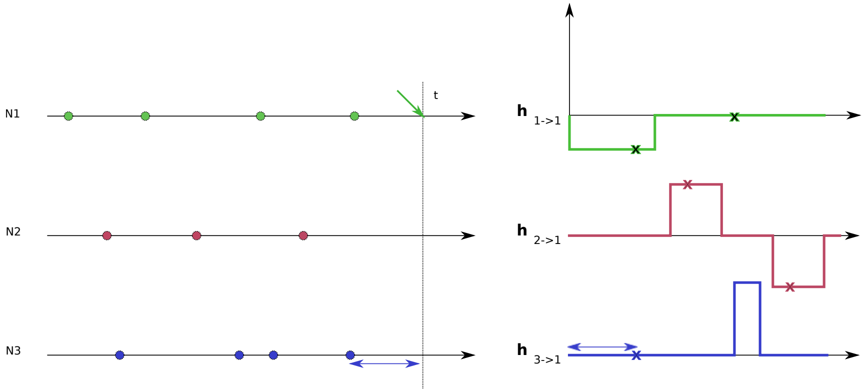
$$\lambda^{(1)}(t) = \underbrace{\nu}_{\text{spontaneous}} + \underbrace{\sum_{\substack{T<t \\ \text{pt de N1}}} h_{1\to 1}(t-T)}_{\text{self-interaction}} + \underbrace{\sum_{\substack{m\neq 1, \\ T<t, \\ \text{pt de Nm}}} h_{m\to 1}(t-T)}_{\text{interaction}}$$

Multivariate Hawkes processes



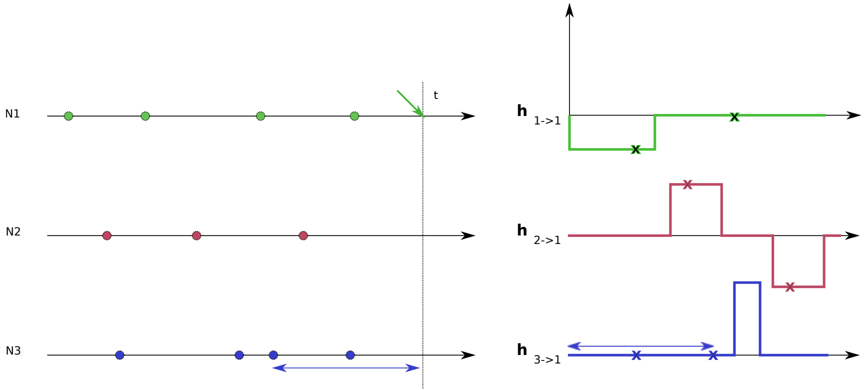
$$\lambda^{(1)}(t) = \underbrace{\nu}_{\text{spontaneous}} + \underbrace{\sum_{\substack{T<t \\ \text{pt de N1}}} h_{1\to 1}(t-T)}_{\text{self-interaction}} + \underbrace{\sum_{\substack{m\neq 1, \\ T<t, \\ \text{pt de Nm}}} h_{m\to 1}(t-T)}_{\text{interaction}}$$

Multivariate Hawkes processes



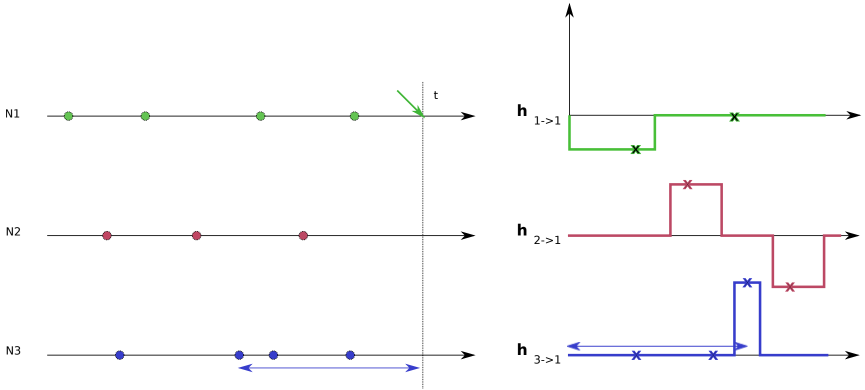
$$\lambda^{(1)}(t) = \underbrace{\nu}_{\text{spontaneous}} + \underbrace{\sum_{\substack{T<t \\ \text{pt de N1}}} h_{1\to 1}(t-T)}_{\text{self-interaction}} + \underbrace{\sum_{\substack{m\neq 1, \\ T<t, \\ \text{pt de Nm}}} h_{m\to 1}(t-T)}_{\text{interaction}}$$

Multivariate Hawkes processes



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

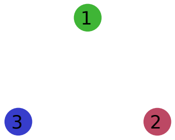
Multivariate Hawkes processes

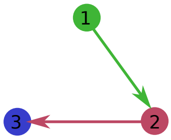


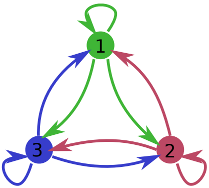
$$\lambda^{(1)}(t) = \underbrace{\nu}_{\text{spontaneous}} + \underbrace{\sum_{\substack{T<t \\ \text{pt de N1}}} h_{1\to 1}(t-T)}_{\text{self-interaction}} + \underbrace{\sum_{\substack{m\neq 1, \\ T<t, \\ \text{pt de Nm}}} h_{m\to 1}(t-T)}_{\text{interaction}}$$

Link Graphical model of local independence

(Didelez (2008))

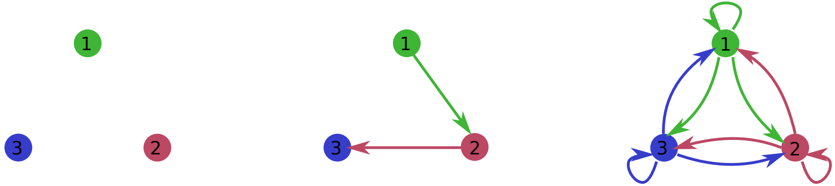






Link Graphical model of local independence

(Didelez (2008))



$\rightarrow$ (?) functional connectivity graphs in Neurosciences,
but also useful in sismology, genomics, marketing, finance, epidemiology ....

More formally

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

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

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

More formally

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

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

  Branching / Cluster representation, stationary process if the spectral radius of $\left(\int h_{\ell\to r}dt\right)$ is < 1.
- Interaction, for instance, (in general any 1-Lipschitz function, Brémaud Massoulié 1996)

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

More formally

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

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

  Branching / Cluster representation, stationary process if the spectral radius of $\left(\int h_{\ell\to r}dt\right)$ is < 1.
- Interaction, for instance, (in general any 1-Lipschitz function, Brémaud Massoulié 1996)

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

- Exponential (Multiplicative shape but no guarantee of a stationary version ...)

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

Contents

1. Problem
2. Detection of synchronization
3. Estimation of functional connectivity
   - A slightly different problem
   - Hawkes processes
   - Estimation methods
   - Simulations
   - On real data

Previous works

- Maximum likelihood estimates eventually + AIC (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)
- Univariate processes + ℓ_0 penalty, oracle inequalities (RB and Schbath)
- Maximum likelihood + exponential formula + ℓ_1 "group Lasso" penalty (Pillow et al. in neuroscience)
- Thresholding + tests for very particular bivariate models, oracle inequality (Sansonnet)

Another parametric method : Least-squares

- A good estimate of the parameters should correspond to an intensity candidate close to the true one.

Another parametric method : Least-squares

- 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

- 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

- 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

- 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 one minimizes

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

for a model $\eta = \lambda_{\mathbf{a}}(t)$

On Hawkes processes

Recall that for the process $N^{(r)}$

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

On Hawkes processes

Recall that for the process $N^{(r)}$

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

$+$ Piecewise constant model with parameter **a**

On Hawkes processes

Recall that for the process $N^{(r)}$

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

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

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

On Hawkes processes

Recall that for the process $N^{(r)}$

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

+ Piecewise constant model with parameter **a**
By linearity,

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

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

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

Recall that for the process $N^{(r)}$

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

$+$ Piecewise constant model with parameter **a**
By linearity,

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

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.

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

The least-square estimate is

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

where in b lies again the number of couples with a certain delay.

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.

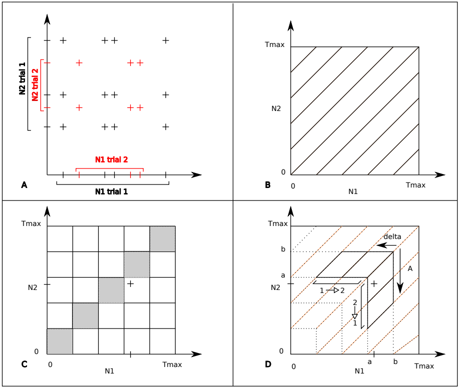
$$\mathbf{b}^{(r)} := \int_0^T \mathbf{Rc}_t dN^{(r)}(t) \simeq \mathbf{G}\mathbf{a}_*^{(r)} + \text{ noise.}$$

The least-square estimate is

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

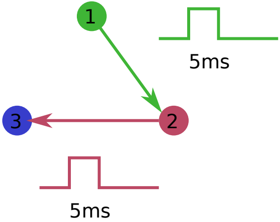
where in b lies again the number of couples with a certain delay.
$\rightarrow$ simpler formula than Maximum Likelihood Estimators for similar properties

Link with cross-correlogram and Joint PeriStimulus Histogram



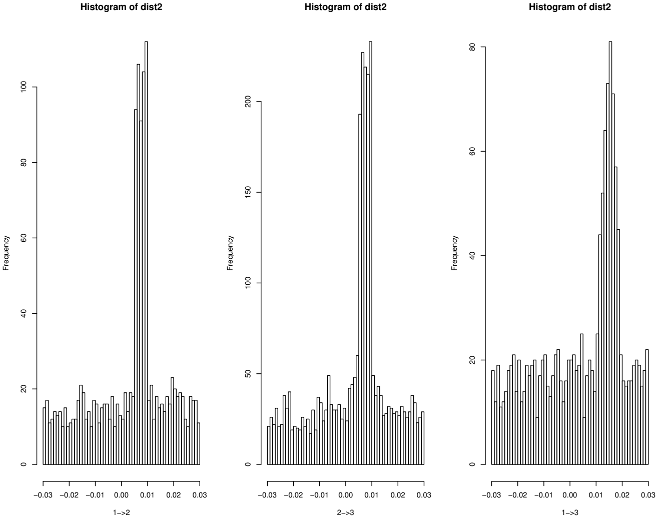
see also Zhang et al (2007) in genomics and Aertsen et al (1989) in Neurosciences

What gain wrt cross correlogramm ?

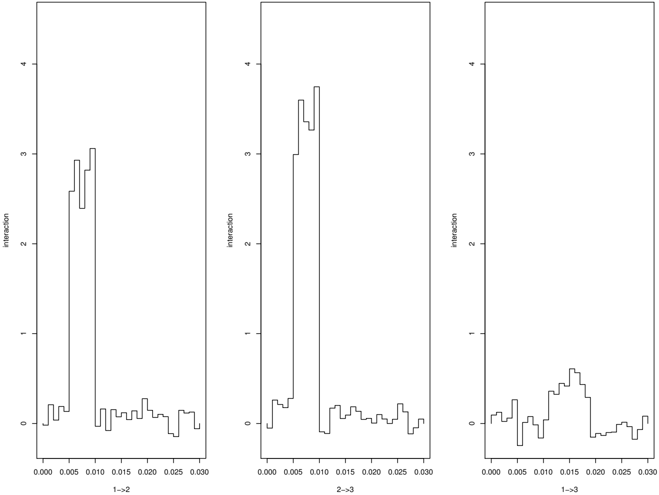


only 2 non zero interaction functions over 9

What gain wrt cross correlogramm ?



What gain wrt cross correlogramm ?



Full Multivariate Hawkes process and ℓ_1 penalty

with N.R. Hansen (Copenhagen), and V. Rivoirard (Dauphine)
The Lasso criterion for each sub-process:

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, on simulated neuronal networks, on real data...

Theoretical Validation = Oracle inequality

Let

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

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

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

then

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

Explanations

- Main Point: **d** controls the random fluctuations / noise $\rightarrow$ should be data-driven and sharp !

Explanations

- Main Point: $\mathbf{d}$ controls the random fluctuations / noise $\rightarrow$ should be data-driven and sharp !
- The loss $(1/T)\sum_{i\in supp(\mathbf{a})}(d_i^{(r)})^2$ is unavoidable even for only one choice of set of non-zeros coefficients. Should be read as "capacity of approximation" + unavoidable loss due to the noise.

Explanations

- Main Point: $\mathbf{d}$ controls the random fluctuations / noise $\rightarrow$ should be data-driven and sharp !
- The loss $(1/T)\sum_{i\in supp(\mathbf{a})}(d_i^{(r)})^2$ is unavoidable even for only one choice of set of non-zeros coefficients. Should be read as "capacity of approximation" + unavoidable loss due to the noise.
- Factor $1/c$ gives a "quality indicator" since G observable in practice. Maybe not the best thing, but with stationnary Hawkes process, one can prove that $c \simeq T$.

Explanations

- Main Point: $\mathbf{d}$ controls the random fluctuations / noise $\rightarrow$ should be data-driven and sharp !
- The loss $(1/T)\sum_{i\in supp(\mathbf{a})}(d_i^{(r)})^2$ is unavoidable even for only one choice of set of non-zeros coefficients. Should be read as "capacity of approximation" + unavoidable loss due to the noise.
- Factor $1/c$ gives a "quality indicator" since G observable in practice. Maybe not the best thing, but with stationnary Hawkes process, one can prove that $c \simeq T$.
- Gives the "best" linear approximation and in this sense valid even if λ not of the prescribed shape, in particular in case of inhibition.

Explanations

- Main Point: **d** controls the random fluctuations / noise → should be data-driven and sharp !
- The loss $(1/T)\sum_{i\in supp(\mathbf{a})}(d_i^{(r)})^2$ is unavoidable even for only one choice of set of non-zeros coefficients. Should be read as "capacity of approximation" + unavoidable loss due to the noise.
- Factor $1/c$ gives a "quality indicator" since G observable in practice. Maybe not the best thing, but with stationnary Hawkes process, one can prove that $c \simeq T$.
- Gives the "best" linear approximation and in this sense valid even if λ not of the prescribed shape, in particular in case of inhibition.
- In practice small unavoidable bias, possible to correct by two step procedures (OLS on the support of the Lasso estimate).

Explanations

- Main Point: $\mathbf{d}$ controls the random fluctuations / noise $\rightarrow$ should be data-driven and sharp !
- The loss $(1/T)\sum_{i\in supp(\mathbf{a})}(d_i^{(r)})^2$ is unavoidable even for only one choice of set of non-zeros coefficients. Should be read as "capacity of approximation" + unavoidable loss due to the noise.
- Factor $1/c$ gives a "quality indicator" since G observable in practice. Maybe not the best thing, but with stationnary Hawkes process, one can prove that $c \simeq T$.
- Gives the "best" linear approximation and in this sense valid even if λ not of the prescribed shape, in particular in case of inhibition.
- In practice small unavoidable bias, possible to correct by two step procedures (OLS on the support of the Lasso estimate).
- Possible to do it with other basis (Fourier) or other linear model (Aalen)

One of the main probabilistic ingredients

Bernstein type inequality for counting processes (H., R.B., R.)

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
$\hat{V}^{\mu}_{\tau} = \frac{\mu}{\mu-\phi(\mu)} \int_0^{\tau} H_s^2 dN_s + \frac{b^2 x}{\mu-\phi(\mu)}$, where $\phi(u) = \exp(u) - u - 1$.
Then for every stopping time τ and every $\varepsilon > 0$

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

One of the main probabilistic ingredients

Bernstein type inequality for counting processes (H., R.B., R.)

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 $\varepsilon > 0$

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

NB : $\hat{V}_\tau^\mu$ is an estimate of the bracket of the martingale, i.e. the martingale equivalent of the variance $\rightarrow$ Gaussian behavior for the main term.

One of the main probabilistic ingredients

Bernstein type inequality for counting processes (H., R.B., R.)

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
$\hat{V}^\mu_\tau = \frac{\mu}{\mu-\phi(\mu)} \int_0^\tau H_s^2 dN_s + \frac{b^2 x}{\mu-\phi(\mu)}$, where $\phi(u) = \exp(u) - u - 1$.
Then for every stopping time τ and every $\varepsilon > 0$

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

NB : $\hat{V}^\mu_\tau$ is an estimate of the bracket of the martingale, i.e. the martingale equivalent of the variance → Gaussian behavior for the main term.
Applied to $\int_0^T \mathbf{Rc}_t \left(dN^{(r)}(t) - \lambda^{(r)}(t)dt\right)$: **d** is given by the right hand-side ($x \simeq \gamma \ln(\#param)$) → Bernstein Lasso.

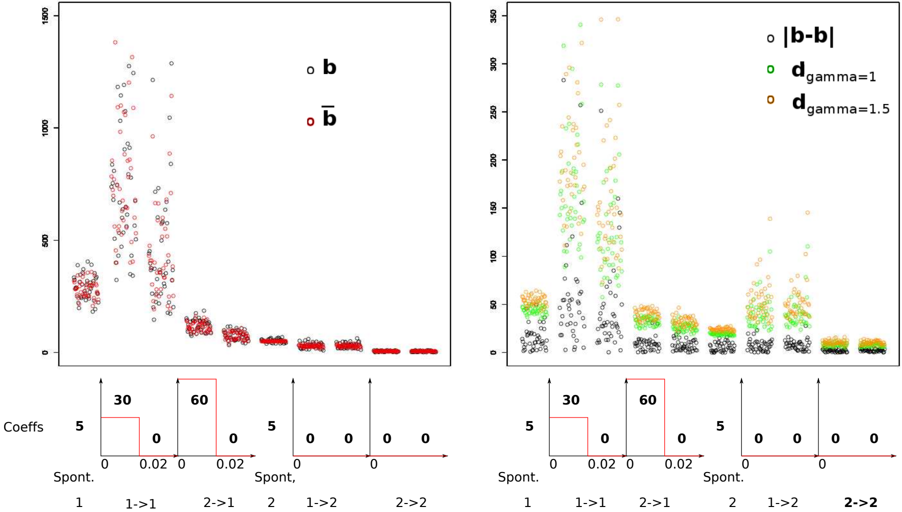
Contents

1 Problem

2 Detection of synchronization

3 Estimation of functional connectivity
- A slightly different problem
- Hawkes processes
- Estimation methods
- Simulations
- On real data

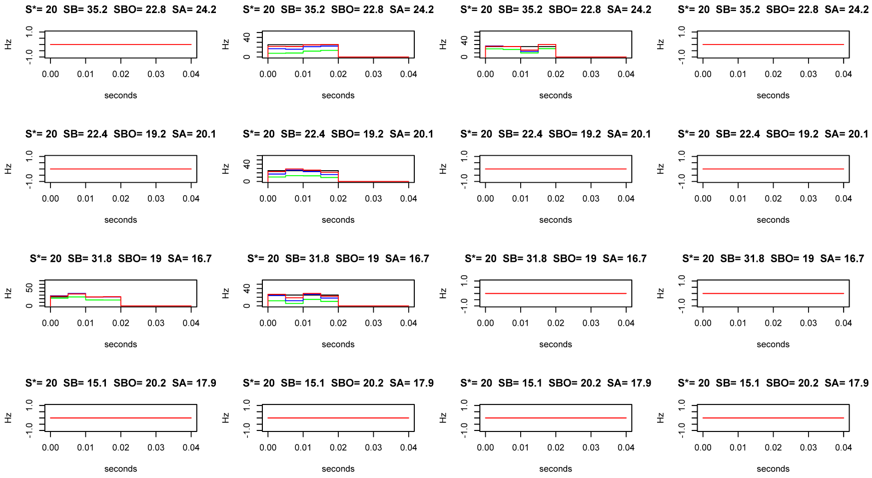
Choices of the weights d_λ



NB : γ larger, d larger and fewer non zero coefficients

NB2: Adaptive Lasso (Zou) $d_\lambda = \gamma/|\hat{a}_\lambda|$

Simulation study - Estimation ($T = 20$, $M = 8$, $K = 8$)



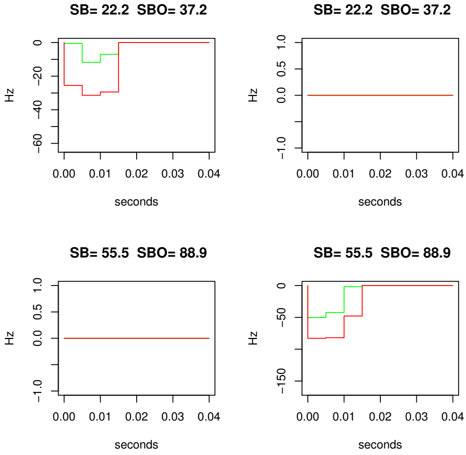
Interactions reconstructed with 'Adaptive Lasso', 'Bernstein Lasso' and 'Bernstein Lasso+OLS'. Above graphs, estimation of spontaneous rates

How to convince biologists to use Hawkes ?

- show by goodness-of-fit test that data are Hawkes. Only statistical test where one would like to accept, which is a bit problematic.
- show how the reconstruction works when something different from Hawkes is simulated, typically leaky integrate and fire.

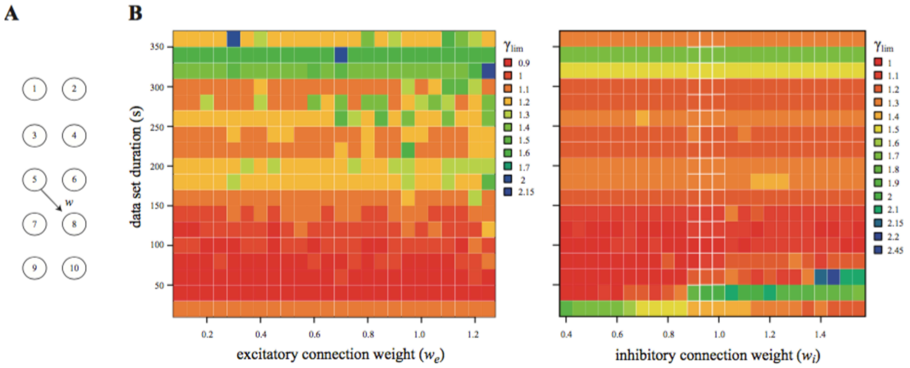
On neuronal data (sensorimotor task)

Joint work with F. Grammont, V. Rivoirard and C. Tuleau-Malot.
30 trials : monkey trained to touch the correct target when illuminated.

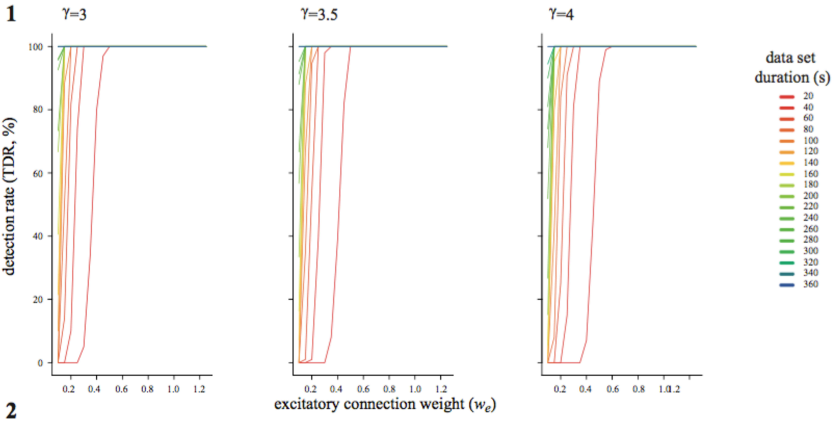


Accept the test of Hawkes hypothesis.

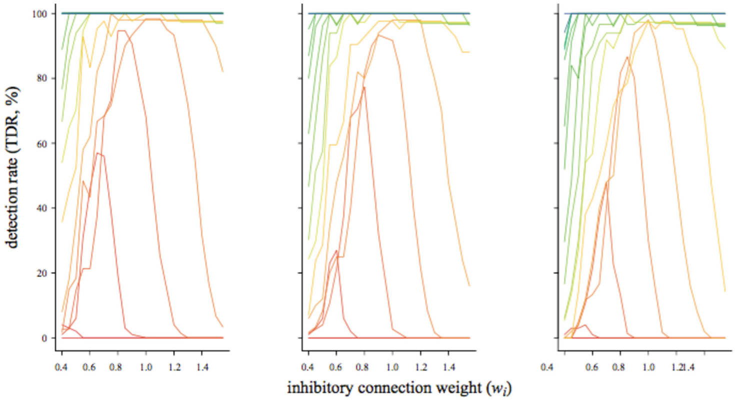
Simulated IF



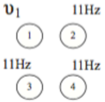
Simulated IF

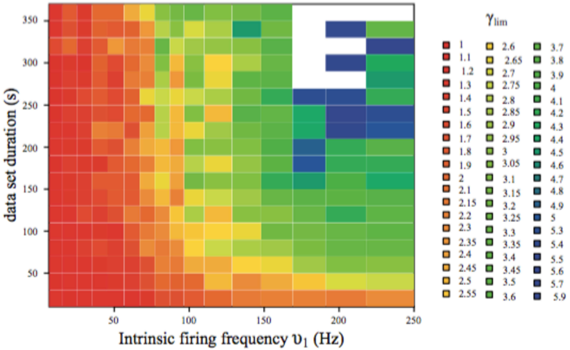


Simulated IF

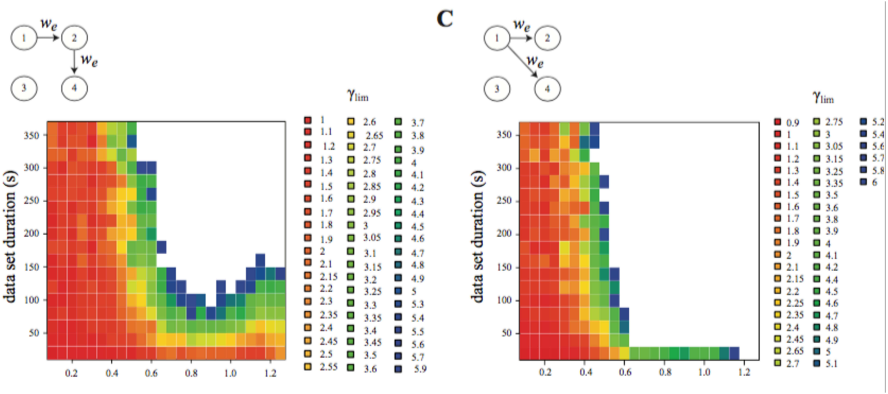


Simulated IF



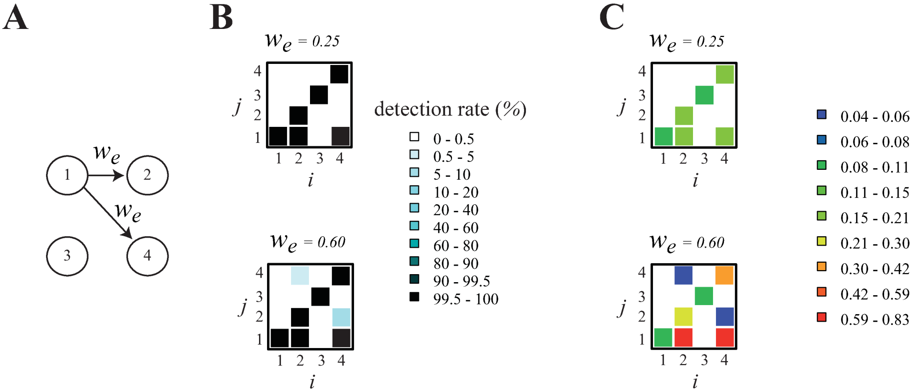


Simulated IF

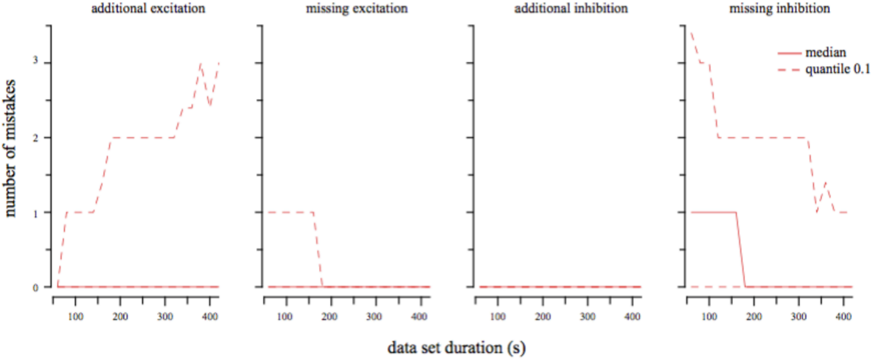


Simulated IF

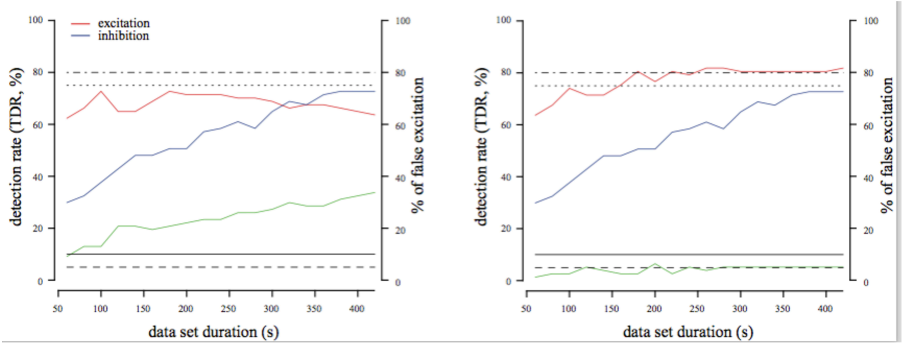
Joint work with R. Lambert, T. Bessaih et al.



Simulated IF



Simulated IF



Contents

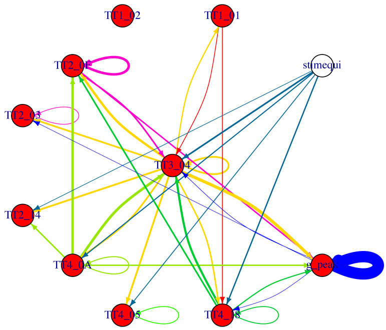
1. Problem
2. Detection of synchronization
3. Estimation of functional connectivity
   - A slightly different problem
   - Hawkes processes
   - Estimation methods
   - Simulations
   - On real data

On neuronal data (vibrissa excitation)

Joint work with RNRP (Paris 6). Behavior: stim. at 25-50Hz.
$T = 72s$, 9 Neurons + stimulation + gamma peaks ($M = 11$)

"Neuron"	TT1_01	TT1_02	TT2_0F	TT2_03	TT2_14	TT3_04
Obs. Spikes	241	136	839	951	446	1310

"Neuron"	TT4_0A	TT4_05	TT4_18	stimequi	g_peak	
Obs. Spikes	734	925	1021	2695	1559	

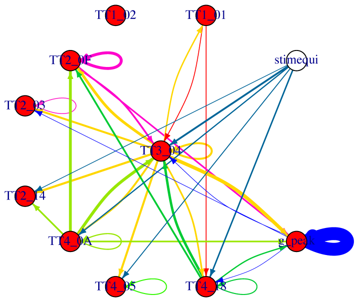


Stability

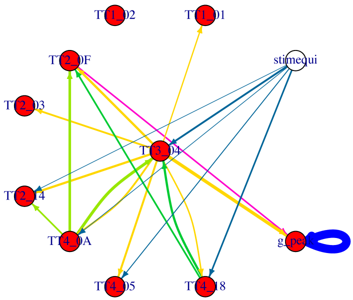
"Neuron"	TT1_01	TT1_02	TT2_0F	TT2_03	TT2_14	TT3_04
Obs. Spikes	241	136	839	951	446	1310
Sim. Spikes	238	135	966	1005	507	1548

"Neuron"	TT4_0A	TT4_05	TT4_18	stimequi	g_peak	
Obs. Spikes	734	925	1021	2695	1559	
Sim. Spikes	816	993	1124	3469	1785	

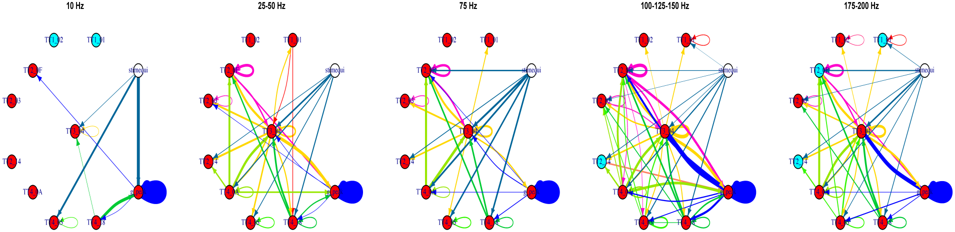
On real data



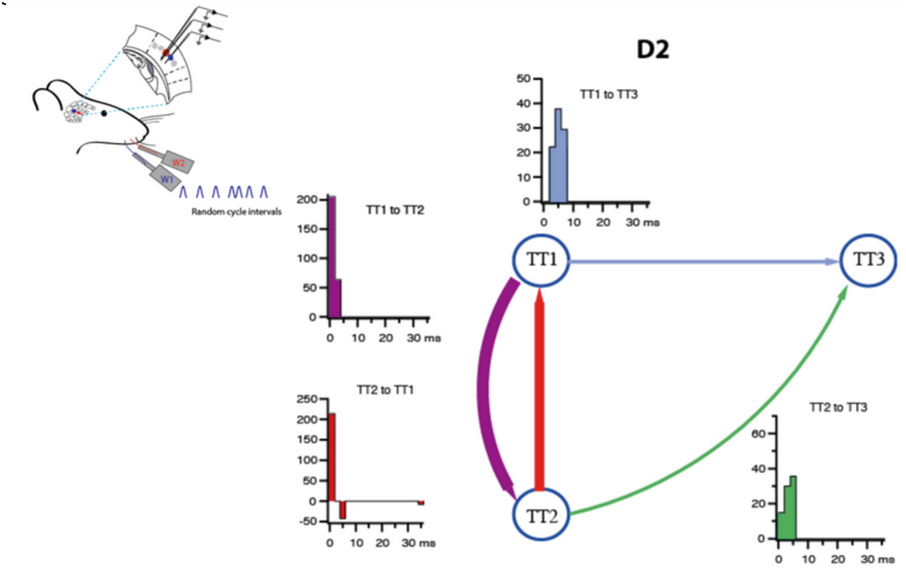
On simulated data

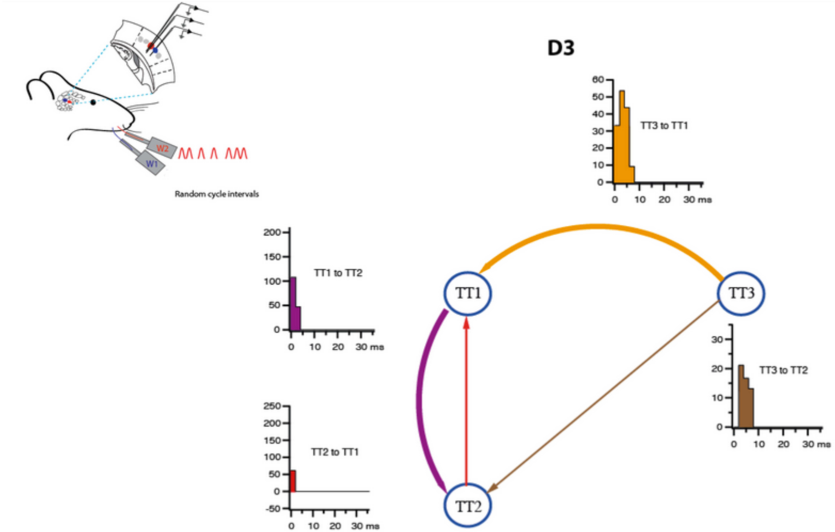


Evolution of the dependance graph as a fonction of the vibrissa excitation

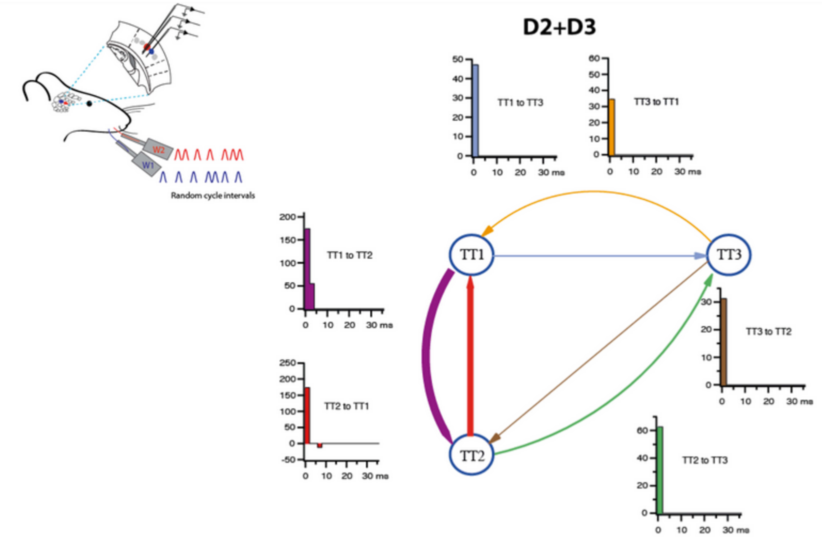


A neural code ?

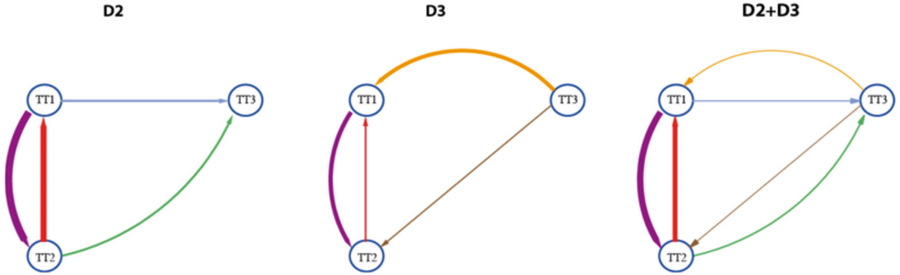



W2
W1
Random cycle intervals
D3
TT3 to TT1
TT1 to TT2
TT2 to TT1
TT3 to TT2
TT1
TT2
TT3
0 10 20 30 ms

A neural code ?



A neural code ?



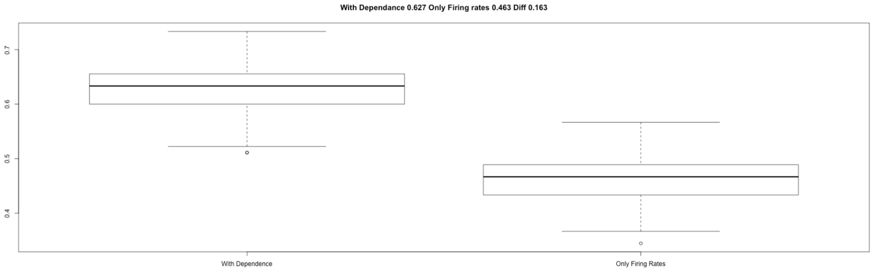
A neural code ?

Joint work with R. Lambert, T. Bessaih [...] and M. Quiquempoix

How much does this code ?

- one rat/site, 3 tetrodes, 9 type of stimulation (D2, D3 or D2+D3 and 3 frequencies), 100 times the same stimulation for 1 second.
- Estimate the Hawkes model for each stimulus (='label') with 50 seconds
- Leave 9 * 50 seconds unlabeled.
- Find the labels by minimizing the least-square.
- Do this several times by picking the 50 "boxes" at random among the 100.

A neural code ?



Conclusions and Perspectives

- It is possible to estimate interaction functions and even a graph of possible interactions, to have some mathematical guarantee on it and this even if the Hawkes linear model is not completely true (robustness).
- It is even possible to (non parametrically) test that such interactions exists. However the interpretation of such tests (not studied yet) should be linked to a "distance" to the Hawkes model.

Conclusions and Perspectives

- It is possible to estimate interaction functions and even a graph of possible interactions, to have some mathematical guarantee on it and this even if the Hawkes linear model is not completely true (robustness).
- It is even possible to (non parametrically) test that such interactions exists. However the interpretation of such tests (not studied yet) should be linked to a "distance" to the Hawkes model.
- Several issues remain
  - lack of stationarity (segmentation/ clustering ?).
  - the recording is scarce (or not spike sorted). What are the effects of both ? mean field theory ?

Coworkers

- Probability : J. Chevallier, N. R. Hansen.
- Statistics : M. Albert, T. Laloë, M. Fromont, V. Rivoirard, C. Tuleau-Malot, A. Rouis.
- Computer Sc : Y. Bouret, C. Mascart, G. Scarella
  C. Bahbah, I. Chaarana, H. Issarane, I. Mouline, K. Zahri
- Neurophysiology : T. Bessaih, F. Grammont, R. Lambert, N. Leresche, M. Quiquempoix.

Thank you !

References

- Hansen, N.R., Reynaud-Bouret, P. and Rivoirard, V. *Lasso and probabilistic inequalities for multivariate point processes.* Bernoulli (2015).
- Reynaud-Bouret, P., Tuleau-Malot, C., Rivoirard, V. and Grammont, F. *Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis* Journal of Mathematical Neuroscience (2014)
- Reynaud-Bouret, P., Rivoirard, V., Tuleau-Malot, C. *Inference of functional connectivity in Neurosciences via Hawkes processes*, 1st IEEE Global Conference on Signal and Information Processing, Austin, Texas (2013).