Some concentration inequalities that are useful in statistics on point processes.

P. Reynaud-Bouret

CNRS - LJAD University of Nice

Journées MAS, Clermont-Ferrand, 2012
Contents

1 Practical examples and Definitions
Contents

1 Practical examples and Definitions

2 Test
Contents

1 Practical examples and Definitions

2 Test

3 Estimation
   - Model selection, Talagrand inequality and Poisson processes
   - Model selection, Talagrand and other processes
   - Thresholding and Poisson processes
   - Lasso and other counting processes
Neuroscience and neuronal unitary activity
Neuronal data and Unitary Events

Unitary (Coincident) Events

-1250  -1000  -750  -500  -250  0
Genomics and Transcription Regulatory Elements

[Diagram showing genomic elements and transcription regulatory elements]

[Diagram showing genomic elements and transcription regulatory elements]
Point processes and Poisson processes

**Point process**

\[ N = \text{random countable set of points of } \mathbb{R} \text{ (here).} \]
# Point processes and Poisson processes

## Point process

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

$N_A$ number of points of $\mathcal{N}$ in $A$, $N_t = N_{[0,t]}$, 

$$dN_t = \sum_T \text{point de } \mathcal{N} \delta_T \cdot \int f(t) dN_t = \sum_{T \in \mathcal{N}} f(T)$$
**Point processes and Poisson processes**

**Point process**

\[
N = \text{random countable set of points of } \mathbb{R} \text{ (here)}.
\]

\[
N_A \text{ number of points of } N \text{ in } A, \ N_t = N_{[0,t]},
\]

\[
dN_t = \sum_T \text{ point de } \delta_T.
\]

**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.
Point processes and Poisson processes

Point process

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

$N_A$ number of points of $N$ in $A$, $N_t = N_{[0,t]}$, $dN_t = \sum_T \text{point de } N \, \delta_T$.

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 $\ell(A)$.
Point processes and Poisson processes

Point process

\(N=\) random countable set of points of \(\mathbb{R}\) (here).

\(N_A\) number of points of \(N\) in \(A\), \(N_t = N_{[0,t]}\),

\[dN_t = \sum_T \text{point de } N \delta_T.\]

Poisson processes

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

**Point process**

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

$N_A$ number of points of $N$ in $A$, $N_t = N_{[0,t]}$,

$$dN_t = \sum_T \text{point de } N \delta_T.$$

**Poisson processes**

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

Usually $d\ell = \lambda(t)dt$, $\lambda(t)$ is the intensity, if constant $\rightarrow$ homogeneous
Basic questions for Poisson processes

- Is $\lambda(t)$ constant? i.e., is the process stationary?
Basic questions for Poisson processes

- Is $\lambda(t)$ constant? ie is the process stationary? → it highly depends on the experiment! → Test of homogeneity
Basic questions for Poisson processes

- Is $\lambda(t)$ constant? i.e., is the process stationary?
  → it highly depends on the experiment! → Test of homogeneity
- Are the processes identically distributed?
Basic questions for Poisson processes

- Is $\lambda(t)$ constant? ie is the process stationary? → it highly depends on the experiment! → Test of homogeneity
- Are the processes identically distributed? → Two-sample tests
Basic questions for Poisson processes

- Is $\lambda(t)$ constant? i.e., is the process stationary? → it highly depends on the experiment! → Test of homogeneity
- Are the processes identically distributed? → Two-sample tests
- Are they dependent? → Independence tests
Basic questions for Poisson processes

- Is $\lambda(t)$ constant? i.e., is the process stationary? → it highly depends on the experiment! → Test of homogeneity
- Are the processes identically distributed? → Two-sample tests
- Are they dependent? → Independence tests
- Can we detect it locally? → multiple "adaptive" testing problems ...
- Where are the poor or rich regions? → Non parametric estimation
### Synergy and Hawkes processes

<table>
<thead>
<tr>
<th>Genomics</th>
<th>Neuroscience</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;events&quot; on the DNA</td>
<td>&quot;work&quot; together in synergy (TRE)</td>
</tr>
</tbody>
</table>

## Synergy and Hawkes processes

<table>
<thead>
<tr>
<th>Genomics</th>
<th>Neuroscience</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;events&quot; on the DNA</td>
<td>Of course</td>
</tr>
<tr>
<td>&quot;work&quot; together in synergy (TRE)</td>
<td>&quot;neurons&quot; work together.</td>
</tr>
</tbody>
</table>

Of course, "neurons" work together.
### Synergy and Hawkes processes

<table>
<thead>
<tr>
<th>Genomics</th>
<th>Neuroscience</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;events&quot; on the DNA&lt;br&gt;&quot;work&quot; together in synergy (TRE)</td>
<td>Of course&lt;br&gt;&quot;neurons&quot; work together.</td>
</tr>
<tr>
<td>If two motifs are part of a common biological process, the distance $\approx$ fixed&lt;br&gt;→ favored or avoided distances (Gusto, Schbath (2005))</td>
<td></td>
</tr>
</tbody>
</table>
# Synergy and Hawkes processes

<table>
<thead>
<tr>
<th>Genomics</th>
<th>Neuroscience</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;events&quot; on the DNA</td>
<td>Of course &quot;neurons&quot; work together.</td>
</tr>
<tr>
<td>&quot;work&quot; together in synergy (TRE)</td>
<td></td>
</tr>
<tr>
<td>If two motifs are part of a common biological process, the distance $\simeq$ fixed $\rightarrow$ favored or avoided distances (Gusto, Schbath (2005))</td>
<td>When recorded, a fixed delay between spikes hints for a functional/physical link.</td>
</tr>
</tbody>
</table>
Intensity

Usually $\mathbb{R}$ is thought as time

**Intensity**

$t \rightarrow \lambda(t)$ where $\lambda(t)dt$ represents the probability to have a point at time $t$ conditionally to the past before $t$ ($s < t$)
Intensity

Usually $\mathbb{R}$ is thought as time

<table>
<thead>
<tr>
<th>Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t \to \lambda(t)$ where $\lambda(t)dt$ represents the probability to have a point at time $t$ conditionnally to the past before $t$ ($s &lt; t$)</td>
</tr>
</tbody>
</table>

"Past" contains in particular the previous occurrences of points.
Intensity

Usually $\mathbb{R}$ is thought as time

$$t \rightarrow \lambda(t) \text{ where } \lambda(t)dt \text{ represents the probability to have a point at time } t \text{ conditionally to the past before } t \ (s < t)$$

"Past" contains in particular the previous occurrences of points. NB : for Genomics, $\mathbb{R}$ is the DNA strand. The "past" may be interpreted as what has already been read in a prescribed direction (e.g. 5'-3' or 3'-5').
Intensity

Usually $\mathbb{R}$ is thought as time

**Intensity**

$t \rightarrow \lambda(t)$ where $\lambda(t)dt$ represents the probability to have a point at time $t$ conditionnally to the past before $t$ ($s < t$)

”Past” contains in particular the previous occurrences of points. NB : for Genomics, $\mathbb{R}$ is the DNA strand. The ”past” may be interpreted as what has already been read in a prescribed direction (e.g. 5’-3’ or 3’-5’).

NB2 : $(N_t - \int_0^t \lambda(s)ds)_t$ is a martingale.
The simple Hawkes process

The intensity $\lambda(t)$ is given by
The simple Hawkes process

The intensity $\lambda(t)$ is given by

$$\nu$$

Spontaneous
The simple Hawkes process

The intensity $\lambda(t)$ is given by

$$\nu + \sum_{T \in \mathbb{N}} h(t - T)$$

Spontaneous        Self-exciting
The simple Hawkes process

The intensity $\lambda(t)$ is given by

$$\nu + \sum_{T \in \mathbb{N}} h(t - T)$$

Spontaneous  Self-exciting
The simple Hawkes process

The intensity $\lambda(t)$ is given by

$$\nu + \sum_{T \in \mathbb{N}} h(t - T)$$

**Spontaneous** **Self-exciting**

The most classical case corresponds to $h > 0$ (see Hawkes (1971)).
The simple Hawkes process

The intensity \( \lambda(t) \) is given by

\[
\nu + \sum_{T \in \mathbb{N}} h(t - T)
\]

\[+\]

Spontaneous  Self-exciting

The most classical case corresponds to \( h > 0 \) (see Hawkes (1971)).
The Hawkes process interaction with itself + an additional interaction

\[ \lambda(t) = \]
The Hawkes process interaction with itself + an additional interaction

\[ \lambda(t) = \nu \]

Spontaneous
The Hawkes process interaction with itself + an additional interaction

\[ \lambda(t) = \nu + \sum_{T \in N} h(t - T) \]

Spontaneous          Self-interaction
The Hawkes process interaction with itself + an additional interaction

\[ \lambda(t) = \nu + \sum_{T \in \mathcal{N}} h(t - T) + \sum_{X \in \mathcal{N}_2} h_2(t - X) \]

<table>
<thead>
<tr>
<th>Spontaneous</th>
<th>Self-interaction</th>
<th>Interaction with other type</th>
</tr>
</thead>
</table>
The Hawkes process interaction with itself + an additional interaction

\[ \lambda(t) = \left( \nu + \sum_{T \in N} h(t - T) + \sum_{X \in N_2} h_2(t - X) \right) + \]

Spontaneous  Self-interaction  Interaction with other type

If \( h \) is null and if \( N_2 \) is fixed (no reciprocal interaction), then \( N \) is a Poisson process given \( N_2 \).
The multivariate Hawkes process

One observes $N^{(1)}, \ldots, N^{(r)}, \ldots, N^{(M)}$ processes such that
The multivariate Hawkes process

One observes \( N^{(1)}, \ldots, N^{(r)}, \ldots, N^{(M)} \) processes such that

\[
\lambda^{(1)}(t) = \lambda^{(2)}(t) = \lambda^{(r)}(t) = \ldots
\]
The multivariate Hawkes process

One observes $N^{(1)}, \ldots, N^{(r)}, \ldots, N^{(M)}$ processes such that
\[
\begin{align*}
\lambda^{(1)}(t) &= \nu_1 \\
\lambda^{(2)}(t) &= \\
\lambda^{(r)}(t) &= 
\end{align*}
\]
The multivariate Hawkes process

One observes $N^{(1)}, \ldots, N^{(r)}, \ldots, N^{(M)}$ processes such that

$$\lambda^{(1)}(t) = \nu_1 + \sum_{T \in N^{(1)}} h^{(1)}_1(t - T)$$

$$\lambda^{(2)}(t) = \lambda^{(r)}(t) = t$$
The multivariate Hawkes process

One observes \( N^{(1)}, \ldots, N^{(r)}, \ldots, N^{(M)} \) processes such that

\[
\lambda^{(1)}(t) = \nu_1 + \sum_{T \in N^{(1)}} h^{(1)}_1(t - T)
\]

\[
\lambda^{(2)}(t) =
\]

\[
\lambda^{(r)}(t) =
\]
The multivariate Hawkes process

One observes $N^{(1)}, \ldots, N^{(r)}, \ldots, N^{(M)}$ processes such that

$$\lambda^{(1)}(t) = \nu_1 + \sum_{T \in N^{(1)}} h_1^{(1)}(t - T)$$

$$\lambda^{(2)}(t) = \lambda^{(r)}(t) = \frac{t^{12/45}}{45}$$
The multivariate Hawkes process

One observes $N^{(1)}, \ldots, N^{(r)}, \ldots, N^{(M)}$ processes such that

$$
\lambda^{(1)}(t) = \nu_1 + \sum_{T \in N^{(1)}} h^{(1)}_1(t - T)
$$

$$
\lambda^{(2)}(t) =
$$

$$
\lambda^{(r)}(t) =
$$
The multivariate Hawkes process

One observes $N^{(1)}, \ldots, N^{(r)}, \ldots, N^{(M)}$ processes such that

$$
\lambda^{(1)}(t) = \nu_1 + \sum_{T \in N^{(1)}} h^{(1)}_1(t - T) + \sum_{\ell \neq 1} \sum_{T \in N^{(\ell)}} h^{(1)}_{\ell}(t - T)
$$

$$
\lambda^{(2)}(t) = \lambda^{(r)}(t) =
$$
The multivariate Hawkes process

One observes $N^{(1)}, \ldots, N^{(r)}, \ldots, N^{(M)}$ processes such that

$$\lambda^{(1)}(t) = \nu_1 + \sum_{T \in N^{(1)}} h^{(1)}_1(t - T) + \sum_{\ell \neq 1} \sum_{T \in N^{(\ell)}} h^{(1)}_{\ell}(t - T)$$

$$\lambda^{(2)}(t) = \nu_2$$

$$\lambda^{(r)}(t) = \nu_2$$
The multivariate Hawkes process

One observes $N^{(1)}, \ldots, N^{(r)}, \ldots, N^{(M)}$ processes such that

$$
\lambda^{(1)}(t) = \nu_1 + \sum_{T \in N^{(1)}} h^{(1)}(t - T) + \sum_{\ell \neq 1} \sum_{T \in N^{(\ell)}} h^{(1)}(t - T)
$$

$$
\lambda^{(2)}(t) = \nu_2
$$

$$
\lambda^{(r)}(t) =
$$
The multivariate Hawkes process

One observes $N^{(1)}, \ldots, N^{(r)}, \ldots, N^{(M)}$ processes such that

$$\lambda^{(1)}(t) = \nu_1 + \sum_{T \in N^{(1)}} h_1^{(1)}(t - T) + \sum_{\ell \neq 1} \sum_{T \in N^{(\ell)}} h_\ell^{(1)}(t - T)$$

$$\lambda^{(2)}(t) = \nu_2 + \sum_{T \in N^{(2)}} h_2^{(2)}(t - T)$$

$$\lambda^{(r)}(t) = \nu_r + \sum_{T \in N^{(r)}} h_r^{(r)}(t - T)$$
The multivariate Hawkes process

One observes \( N^{(1)}, \ldots, N^{(r)}, \ldots, N^{(M)} \) processes such that

\[
\lambda^{(1)}(t) = \nu_1 + \sum_{T \in N^{(1)}} h^{(1)}_1(t - T) + \sum_{\ell \neq 1} \sum_{T \in N^{(\ell)}} h^{(1)}_\ell(t - T)
\]

\[
\lambda^{(2)}(t) = \nu_2 + \sum_{T \in N^{(2)}} h^{(2)}_2(t - T)
\]

\[
\lambda^{(r)}(t) =
\]
The multivariate Hawkes process

One observes $N^{(1)}, \ldots, N^{(r)}, \ldots, N^{(M)}$ processes such that

$$\lambda^{(1)}(t) = \nu_1 + \sum_{T \in N^{(1)}} h^{(1)}_1(t - T) + \sum_{\ell \neq 1} \sum_{T \in N^{(\ell)}} h^{(1)}_{\ell}(t - T)$$

$$\lambda^{(2)}(t) = \nu_2 + \sum_{T \in N^{(2)}} h^{(2)}_2(t - T)$$

$$\lambda^{(r)}(t) =$$
The multivariate Hawkes process

One observes \( N^{(1)}, \ldots, N^{(r)}, \ldots, N^{(M)} \) processes such that

\[
\lambda^{(1)}(t) = \nu_1 + \sum_{T \in N^{(1)}} h^{(1)}_1(t - T) + \sum_{\ell \neq 1} \sum_{T \in N^{(\ell)}} h^{(1)}_\ell(t - T)
\]

\[
\lambda^{(2)}(t) = \nu_2 + \sum_{T \in N^{(2)}} h^{(2)}_2(t - T) + \sum_{\ell \neq 2} \sum_{T \in N^{(\ell)}} h^{(2)}_\ell(t - T)
\]

\[
\lambda^{(r)}(t) = \nu_r + \sum_{T \in N^{(r)}} h^{(r)}_1(t - T) + \sum_{\ell \neq r} \sum_{T \in N^{(\ell)}} h^{(r)}_\ell(t - T)
\]
The multivariate Hawkes process (2)

Link with graphical model of local independence (see Didelez (2008))
The multivariate Hawkes process (2)

Link with graphical model of local independence (see Didelez (2008))
The multivariate Hawkes process (2)

Link with graphical model of local independence (see Didelez (2008))

Hence we need a sparse adaptive estimation (functions, support of the functions)!
Test and level

In the Poisson process framework, observe $N$ with intensity $\lambda$ and find a test $\Delta$ of

$$H_0: \ " \lambda \text{ is constant } \" \text{ against } H_1: \ \" \text{it is not}\"$$

The test is of level $\alpha$ if $\mathbb{P}_{H_0}(\Delta = 1) \leq \alpha$
Test and level

In the Poisson process framework, observe $N$ with intensity $\lambda$ and find a test $\Delta$ of

$$H_0: \ "\lambda \text{ is constant} \" \text{ against } H_1: \ "\text{it is not}"$$

The test is of level $\alpha$ if $\mathbb{P}_{H_0}(\Delta = 1) \leq \alpha$
Power and practice

The power is $\lambda \in H_1 \rightarrow \mathbb{P}_\lambda(\Delta = 1)$.

- when $\lambda$ is almost constant, power $\simeq \mathbb{P}_{H_0}(\Delta = 1)$. 
Power and practice

The power is $\lambda \in H_1 \rightarrow \mathbb{P}_\lambda(\Delta = 1)$.
- when $\lambda$ is almost constant, power $\simeq \mathbb{P}_{H_0}(\Delta = 1)$.
- best to have $\mathbb{P}_{H_0}(\Delta = 1) = \alpha$
Power and practice

The power is $\lambda \in H_1 \rightarrow P_\lambda(\Delta = 1)$.

- when $\lambda$ is almost constant, power $\simeq P_{H_0}(\Delta = 1)$.
- best to have $P_{H_0}(\Delta = 1) = \alpha$
- Morever gives in practice access to meaningful p-values (value of $\alpha$, depending on the observed $N$ where the test changes its decision)
Power and practice

The power is $\lambda \in H_1 \rightarrow P_\lambda(\Delta = 1)$.

- when $\lambda$ is almost constant, power $\simeq P_{H_0}(\Delta = 1)$.
- best to have $P_{H_0}(\Delta = 1) = \alpha$
- Moreover gives in practice access to meaningful p-values (value of $\alpha$, depending on the observed $N$ where the test changes its decision)
- Also p-values involved in multiple testing procedures ...
Power and practice

The power is $\lambda \in H_1 \rightarrow P_\lambda(\Delta = 1)$.

- when $\lambda$ is almost constant, power $\simeq P_{H_0}(\Delta = 1)$.
- best to have $P_{H_0}(\Delta = 1) = \alpha$
- Moreover gives in practice access to meaningful p-values (value of $\alpha$, depending on the observed $N$ where the test changes its decision)
- Also p-values involved in multiple testing procedures ...
- To guarantee $P_{H_0}(\Delta = 1) = \alpha$, best to have some statistics whose law known under $H_0$. 
Power and practice

The power is $\lambda \in H_1 \rightarrow P_{\lambda}(\Delta = 1)$.

- when $\lambda$ is almost constant, power $\simeq P_{H_0}(\Delta = 1)$.
- best to have $P_{H_0}(\Delta = 1) = \alpha$
- Moreover gives in practice access to meaningful p-values (value of $\alpha$, depending on the observed $N$ where the test changes its decision)
- Also p-values involved in multiple testing procedures ...
- To guarantee $P_{H_0}(\Delta = 1) = \alpha$, best to have some statistics whose law known under $H_0$.
- Here, conditionally to the total number of points is $n$, points behave under $H_0$ as a $n$ uniform iid sample $\rightarrow$ easy access to quantile
Alternatives and choice of the test statistics

But here, the alternatives are

- **NOT**: parametric, smooth, detectable by Kolmogorov Smirnov
Alternatives and choice of the test statistics

But here, the alternatives are

- NOT: parametric, smooth, detectable by Kolmogorov Smirnov
- more likely to have spiky distributions with unknown support
Alternatives and choice of the test statistics

But here, the alternatives are

- NOT: parametric, smooth, detectable by Kolmogorov Smirnov
- more likely to have spiky distributions with unknown support

Best to project on a wavelet (Haar) basis and reject when, say, one/few coefficients too high.
Alternatives and choice of the test statistics

But here, the alternatives are

- NOT: parametric, smooth, detectable by Kolmogorov Smirnov
- more likely to have spiky distributions with unknown support

Best to project on a wavelet (Haar) basis and reject when, say, one/few coefficients too high.
"High" = quantile under $H_0$. 
Alternatives and choice of the test statistics

But here, the alternatives are

- NOT: parametric, smooth, detectable by Kolmogorov Smirnov
- more likely to have spiky distributions with unknown support

Best to project on a wavelet (Haar) basis and reject when, say, one/few coefficients too high.

"High" = quantile under $H_0$.

Problem = we don’t know which coefficients $\rightarrow$ aggregation of tests.
Notations

Let $\lambda(t) = Ls(t)$ with $L$ known ($\to \infty$) and $s$ unknown such that

$$s = \alpha_0 \phi_0 + \sum_{j \in \mathbb{N}} \sum_{k=0}^{2^j-1} \alpha(j,k) \phi(j,k),$$
Notations

Let $\lambda(t) = Ls(t)$ with $L$ known ($\to \infty$) and $s$ unknown such that

$$s = \alpha_0 \phi_0 + \sum_{j \in \mathbb{N}} \sum_{k=0}^{2^j-1} \alpha(j,k) \phi(j,k),$$

with $\phi_0(x) = 1_{[0,1]}(x)$ and $\phi(j,k)(x) = 2^{j/2} \psi(2^j x - k)$ where

$$\psi(x) = 1_{[0,1/2]}(x) - 1_{[1/2,1]}(x).$$
Notations

Let $\lambda(t) = Ls(t)$ with $L$ known ($\to \infty$) and $s$ unknown such that

$$s = \alpha_0 \phi_0 + \sum_{j \in \mathbb{N}} \sum_{k=0}^{2^j-1} \alpha_{(j,k)} \phi_{(j,k)},$$

with $\phi_0(x) = 1_{[0,1]}(x)$ and $\phi_{(j,k)}(x) = 2^{j/2} \psi(2^j x - k)$ where $\psi(x) = 1_{[0,1/2]}(x) - 1_{[1/2,1]}(x)$.

We want to reject when the distance between $s$ and $S_0 = \text{Span}(\phi_0)$ is too large.

- Approximate $d(s, S_0)^2$ by $\sum_{(j,k) \in m} \alpha_{(j,k)}^2$. 
Notations

Let $\lambda(t) = Ls(t)$ with $L$ known ($\to \infty$) and $s$ unknown such that

$$s = \alpha_0 \phi_0 + \sum_{j \in \mathbb{N}} \sum_{k=0}^{2^j-1} \alpha(j,k) \phi(j,k),$$

with $\phi_0(x) = 1_{[0,1]}(x)$ and $\phi(j,k)(x) = 2^{j/2} \psi(2^j x - k)$ where $\psi(x) = 1_{[0,1/2]}(x) - 1_{[1/2,1]}(x)$.

We want to reject when the distance between $s$ and $S_0 = \operatorname{Span}(\phi_0)$ is too large.

- Approximate $d(s, S_0)^2$ by $\sum_{(j,k) \in m} \alpha_{(j,k)}^2$.
- Estimate it unbiasly by $T_m = \sum_{(j,k) \in m} T(j,k)$ with $m$ finite and

$$T(j,k) = \hat{\alpha}_{(j,k)}^2 - \frac{1}{L^2} \int \phi_{(j,k)}^2 dN$$
Notations

Let \( \lambda(t) = Ls(t) \) with \( L \) known \((\rightarrow \infty)\) and \( s \) unknown such that

\[
    s = \alpha_0 \phi_0 + \sum_{j \in \mathbb{N}} \sum_{k=0}^{2^j-1} \alpha(j,k) \phi(j,k),
\]

with \( \phi_0(x) = 1_{[0,1]}(x) \) and \( \phi(j,k)(x) = 2^j/2 \psi(2^j x - k) \) where \( \psi(x) = 1_{[0,1/2]}(x) - 1_{[1/2,1]}(x) \).

We want to reject when the distance between \( s \) and \( S_0 = \text{Span}(\phi_0) \) is too large.

- Approximate \( d(s, S_0)^2 \) by \( \sum_{(j,k) \in m} \alpha^2(j,k) \).
- Estimate it unbiasesly by \( T_m = \sum_{(j,k) \in m} T(j,k) \) with \( m \) finite and

\[
    T(j,k) = \hat{\alpha}^2(j,k) - \frac{1}{L^2} \int \phi^2(j,k) dN = \sum_{l \neq l'} \phi(j,k)(X_l) \phi(j,k)(X_{l'})
\]

where \( N \) is the set of points \( X_l \)'s.
Notations

Let \( \lambda(t) = Ls(t) \) with \( L \) known (\( \to \infty \)) and \( s \) unknown such that

\[
s = \alpha_0 \phi_0 + \sum_{j \in \mathbb{N}} \sum_{k=0}^{2^j-1} \alpha(j,k) \phi(j,k),
\]

with \( \phi_0(x) = 1_{[0,1]}(x) \) and \( \phi(j,k)(x) = 2^{j/2} \psi(2^j x - k) \) where \( \psi(x) = 1_{[0,1/2]}(x) - 1_{[1/2,1]}(x) \).

We want to reject when the distance between \( s \) and \( S_0 = \text{Span}(\phi_0) \) is too large.

- Approximate \( d(s, S_0)^2 \) by \( \sum_{(j,k) \in m} \alpha^2(j,k) \).
- Estimate it unbiasedly by \( T_m = \sum_{(j,k) \in m} T(j,k) \) with \( m \) finite and
  \[
  T(j,k) = \hat{\alpha}(j,k)^2 - \frac{1}{L^2} \int \phi^2(j,k) dN = \sum_{l \neq l'} \phi(j,k)(X_l) \phi(j,k)(X_{l'})
  \]
  where \( N \) is the set of points \( X_l \)'s.
- we reject when \( T_m > t_{m,\alpha}(N_{\text{tot}}) \).
Notations

Let $\lambda(t) = Ls(t)$ with $L$ known ($\to \infty$) and $s$ unknown such that

$$s = \alpha_0 \phi_0 + \sum_{j \in \mathbb{N}} \sum_{k=0}^{2^j-1} \alpha(j,k) \phi(j,k),$$

with $\phi_0(x) = 1_{[0,1]}(x)$ and $\phi(j,k)(x) = 2^{j/2} \psi(2^j x - k)$ where $\psi(x) = 1_{[0,1/2]}(x) - 1_{[1/2,1]}(x)$.

We want to reject when the distance between $s$ and $S_0 = \text{Span}(\phi_0)$ is too large.

- Approximate $d(s, S_0)^2$ by $\sum_{(j,k) \in m} \alpha^2(j,k)$.
- Estimate it unbiasly by $T_m = \sum_{(j,k) \in m} T(j,k)$ with $m$ finite and

$$T(j,k) = \hat{\alpha}^2(j,k) - \frac{1}{L^2} \int \phi^2(j,k) dN = \sum_{l \neq l'} \phi(j,k)(X_l)\phi(j,k)(X_{l'})$$

where $N$ is the set of points $X_l$'s.

- we reject when $T_m > t_m^{(N_{tot})}$.
- $t_m^{(n)}$ the $1 - \alpha$ quantile of the conditional distribution.
Aggregation

Let $\mathcal{M}$ be a family of subsets of indices.

**Reject rule**

there exists one $m \in \mathcal{M}$ such that $T_m > t_{m,\alpha_m}^{(N)}$. 


Let $\mathcal{M}$ be a family of subsets of indices.

**Reject rule**

there exists one $m \in \mathcal{M}$ such that $T_m > t_{m, \alpha_m}^{(N)}$, where under $H_0$, $\mathbb{P}(\exists m \in \mathcal{M}, \ T_m > t_{m, \alpha_m}^{(N)}) \leq \alpha$. 
## Aggregation

Let $\mathcal{M}$ be a family of subsets of indices.

### Reject rule

| there exists one $m \in \mathcal{M}$ such that $T_m > t_{m,\alpha_m}^{(N)}$, where under $H_0$, $\mathbb{P}(\exists m \in \mathcal{M}, T_m > t_{m,\alpha_m}^{(N)}) \leq \alpha$. |

- Basic choice: Bonferroni $\alpha_m = \frac{\alpha}{|\mathcal{M}|}$.  

Basic choice: Bonferroni $\alpha_m = \frac{\alpha}{|\mathcal{M}|}$. 
Aggregation

Let $\mathcal{M}$ be a family of subsets of indices.

**Reject rule**

there exists one $m \in \mathcal{M}$ such that $T_m > t^{(N)}_{m,\alpha_m}$.

where under $H_0$, $\mathbb{P}(\exists m \in \mathcal{M}, \ T_m > t^{(N)}_{m,\alpha_m}) \leq \alpha$.

- Basic choice : Bonferroni $\alpha_m = \frac{\alpha}{|\mathcal{M}|}$.  
- with weights : $\alpha_m = \alpha e^{-W_m}$ such that $\sum e^{-W_m} \leq 1$.  

Aggregation

Let $\mathcal{M}$ be a family of subsets of indices.

**Reject rule**

there exists one $m \in \mathcal{M}$ such that $T_m > t^{(N)}_{m, \alpha_m}$,

where under $H_0$, $\mathbb{P}(\exists m \in \mathcal{M}, \ T_m > t^{(N)}_{m, \alpha_m}) \leq \alpha$.

- Basic choice: Bonferroni $\alpha_m = \frac{\alpha}{|\mathcal{M}|}$.
- with weights: $\alpha_m = \alpha e^{-W_m}$ such that $\sum e^{-W_m} \leq 1$
- refined .... for simulation (possible to guarantee equality in the level)
Need of concentration?

For $\lambda$ in $H_1$, Error of 2nd kind = 
$\mathbb{P}_\lambda(\forall m \in \mathcal{M}, T_m \leq t^{(N)}_{m,\alpha_m}) \leq \mathbb{P}_\lambda(T_m \leq t^{(N)}_{m,\alpha_m})$ for all $m$ in $\mathcal{M}$. 
Need of concentration?

For $\lambda$ in $H_1$, Error of 2nd kind = 
\[ P_\lambda(\forall m \in \mathcal{M}, T_m \leq t_{m,\alpha_m}^{(N)}) \leq P_\lambda(T_m \leq t_{m,\alpha_m}^{(N)}) \text{ for all } m \text{ in } \mathcal{M}. \]

How $t_{m,\alpha_m}^{(N)} = t_{m,\frac{\alpha}{|\mathcal{M}|}}^{(N)}$ deteriorates with respect $|\mathcal{M}|$?
Need of concentration?

For $\lambda$ in $H_1$, Error of 2nd kind =
$\mathbb{P}_\lambda(\forall m \in \mathcal{M}, T_m \leq t_m^{(N)}(\alpha)) \leq \mathbb{P}_\lambda(T_m \leq t_m^{(N)}(\alpha))$ for all $m$ in $\mathcal{M}$.

How $t_m^{(N)}(\alpha) = t_m^{(N)}(\alpha/|\mathcal{M}|)$ deteriorates with respect $|\mathcal{M}|$?

$\rightarrow$ how $t_m^{(N)}(\alpha)$ depends on $\alpha$?

- if there is exponential decay, possible to aggregate $|\mathcal{M}|$
  without losing much more than a logarithmic term

- Hence methods powerful against "ugly" alternatives (such as
  weak Besov spaces) and usually minimax if well done ...
Concentration of U-statistics

$T_m$ is a degenerate U-statistics of order 2 under $H_0$ conditionnally to $N_{tot} = n$, ie it’s a

$$U_n = \sum_{i \neq j} g(X_i, X_j),$$

with $g$ symmetric $\mathbb{E}(g(X_i, X_j)|X_j) = 0$.

**Theorem**

*If $\|g\|_{\infty} \leq A$ then for all $u, \varepsilon > 0$*

$$\mathbb{P}(U_n \geq 2(1 + \varepsilon)^{3/2} C\sqrt{u} + \square_{\varepsilon} Du + \square_{\varepsilon} B u^{3/2} + \square_{\varepsilon} A u^2) \leq \square e^{-u}$$

with $C^2 = \sum_{i \neq j} \mathbb{E}(g(X_i, X_j)^2)$ and $B$ and $D$ other functions of $g$. 
Concentration of U-statistics

$T_m$ is a degenerate U-statistics of order 2 under $H_0$ conditionnally to $N_{tot} = n$, ie it’s a

$$U_n = \sum_{i \neq j} g(X_i, X_j),$$

with $g$ symmetric $\mathbb{E}(g(X_i, X_j)|X_j) = 0$.

**Theorem**

*If $\|g\|_\infty \leq A$ then for all $u, \varepsilon > 0$*

$$\mathbb{P}(U_n \geq 2(1 + \varepsilon)^{3/2} C\sqrt{u} + \Box_\varepsilon Du + \Box_\varepsilon Bu^{3/2} + \Box_\varepsilon Au^2) \leq \Box e^{-u}$$

With $C^2 = \sum_{i \neq j} \mathbb{E}(g(X_i, X_j)^2)$ and $B$ and $D$ other functions of $g$.

- with constant Houdré, RB (2003) - also Poisson processes
- higher order Adamczak (2006)
Conclusions for testing

- Concentration inequalities are a tool to evaluate the dependency in $\alpha$ of the $1 - \alpha$ quantile.
- In the upper bound, no need for precise constants or observable quantities.
- But dependency of for instance, $A, B, C, D$ in $m$ crucial... Best if dimension free or dependency in $m$ as small as possible $\rightarrow$ choice of the test statistics and the $M$’s.
Poisson case

Here again $\lambda(t) = Ls(t)$ with $L$ known ($\to \infty$), $s$ unknown.

Least square contrast

$$\gamma(f) = -\frac{2}{L} \int f(t) dN_t + \int f^2(t) dt$$
Poisson case

Here again $\lambda(t) = Ls(t)$ with $L$ known ($\rightarrow \infty$), $s$ unknown.

Least square contrast

$$\gamma(f) = -\frac{2}{L} \int f(t) dN_t + \int f^2(t) dt$$

$$E(\gamma(f)) = -2 < f, s > + \|f\|^2 = \|f - s\|^2 - \|s\|^2 \text{ minimal when } f = s.$$  

- Let $S_m$ be any finite vectorial subspace with ONB $(\varphi_\lambda, \lambda \in \Lambda_m)$.
- $\hat{s}_m = \text{argmin}_{f \in S_m} \gamma(f)$
Poisson case

Here again $\lambda(t) = Ls(t)$ with $L$ known ($\to \infty$), $s$ unknown.

Least square contrast

$$
\gamma(f) = -\frac{2}{L} \int f(t) dN_t + \int f^2(t) dt
$$

$$
E(\gamma(f)) = -2 < f, s > + \|f\|^2 = \|f - s\|^2 - \|s\|^2 \text{ minimal when } f = s.
$$

- Let $S_m$ be any finite vectorial subspace with ONB $(\varphi_\lambda, \lambda \in \Lambda_m)$.
- $\hat{s}_m = \text{argmin}_{f \in S_m} \gamma(f)$
- $E(\|s - \hat{s}_m\|^2) = \|s - s_m\|^2 + \frac{1}{L} \sum_{\lambda \in \Lambda_m} \int \varphi_\lambda^2(t)s(t)dt \leq \|s - s_m\|^2 + \frac{|m|}{L}\|s\|_\infty.$ \to \text{ penalisation}
Poisson case

Here again $\lambda(t) = Ls(t)$ with $L$ known ($\to \infty$), $s$ unknown.

Least square contrast

$$\gamma(f) = -\frac{2}{L} \int f(t) dN_t + \int f^2(t) dt$$

$$\mathbb{E}(\gamma(f)) = -2 < f, s > + \|f\|^2 = \|f - s\|^2 - \|s\|^2$$

minimal when $f = s$.

- Let $S_m$ be any finite vectorial subspace with ONB $(\varphi_\lambda, \lambda \in \Lambda_m)$.
- $\hat{s}_m = \text{argmin}_{f \in S_m} \gamma(f)$
- $\mathbb{E}(\|s - \hat{s}_m\|^2) = \|s - s_m\|^2 + \frac{1}{L} \sum_{\lambda \in \Lambda_m} \int \varphi_\lambda^2(t)s(t)dt \leq \|s - s_m\|^2 + \frac{|m|}{L} \|s\|_{\infty}$. → penalisation

Penalized model selection

$$\hat{m} = \text{argmin}_{m \in M} \{ \gamma(\hat{s}_m) + \text{pen}(m) \}$$
An easy calculus (1)

\[ \gamma(f) = -\frac{2}{L} \int f(t)(dN_t - s(t)dt) + \|f - s\|^2 - \|s\|^2. \]

Let \( \delta(f) = \frac{1}{L} \int f(t)(dN_t - Ls(t)dt) \) (zero mean)
An easy calculus (1)

\[ \gamma(f) = -\frac{2}{L} \int f(t)(dN_t - s(t)dt) + \|f - s\|^2 - \|s\|^2. \]

Let \( \delta(f) = \frac{1}{L} \int f(t)(dN_t - Ls(t)dt) \) (zero mean)

\[ \gamma(f) = -2\delta(f) + \|f - s\|^2 - \|s\|^2. \]

Moreover for all \( m \in \mathcal{M} \)

\[ \gamma(\hat{s}_m) + \text{pen}(\hat{m}) \leq \gamma(\hat{s}_m) + \text{pen}(m) \leq \gamma(s_m) + \text{pen}(m). \]
An easy calculus (1)

$$\gamma(f) = -\frac{2}{L} \int f(t)(dN_t - s(t)dt) + \|f - s\|^2 - \|s\|^2.$$

Let $\delta(f) = \frac{1}{L} \int f(t)(dN_t - Ls(t)dt)$ (zero mean)

$$\gamma(f) = -2\delta(f) + \|f - s\|^2 - \|s\|^2.$$

Moreover for all $m \in M$

$$\gamma(\hat{s}_\hat{m}) + \text{pen}(\hat{m}) \leq \gamma(\hat{s}_m) + \text{pen}(m) \leq \gamma(s_m) + \text{pen}(m).$$

$$\|\hat{s}_\hat{m} - s\|^2 \leq \|s - s_m\|^2 + \text{pen}(m) - 2\delta(s_m) + 2\delta(\hat{s}_\hat{m}) - \text{pen}(\hat{m})$$
An easy calculus (2)

Starting point

\[ \| \hat{s}_m - s \|^2 \leq \| s - s_m \|^2 + \text{pen}(m) - 2\delta(s_m - s_{\hat{m}}) + 2\delta(\hat{s}_m - s_m) - \text{pen}(\hat{m}) \]

\[ \delta(s_m) \rightarrow \text{neglicable (also } \delta(s_{\hat{m}})) \]
An easy calculus (2)

Starting point

\[ \| \hat{s}_m - s \|^2 \leq \| s - s_m \|^2 + \text{pen}(m) - 2\delta(s_m - s) + 2\delta(\hat{s}_m - s) - \text{pen}(\hat{m}) \]

- \( \delta(s_m) \rightarrow \) negligible (also \( \delta(s) \))
- \( \delta(\hat{s}_m - s) = \sum_{\lambda \in \Lambda} \left( \frac{1}{L} \int \varphi_\lambda(t)(dN_t - Ls(t)dt) \right)^2 \)
An easy calculus (2)

Starting point

\[ \| \hat{s}_m - s \|^2 \leq \| s - s_m \|^2 + \text{pen}(m) - 2\delta(s_m - \hat{s}_m) + 2\delta(\hat{s}_m - s_m) - \text{pen}(\hat{m}) \]

- \( \delta(s_m) \to \) negligible (also \( \delta(s_{\hat{m}}) \))
- \( \delta(\hat{s}_m - s_m) = \sum_{\lambda \in \Lambda_{\hat{m}}} (\frac{1}{L} \int \varphi\lambda(t)(dN_t - Ls(t)dt))^2 = \chi^2(\hat{m}) \)
An easy calculus (2)

Starting point

\[
\|\hat{s}_m - s\|^2 \leq \|s - s_m\|^2 + \text{pen}(m) - 2\delta(s_m - s) + 2\delta(\hat{s}_m - s_m) - \text{pen}(\hat{m})
\]

- \(\delta(s_m) \rightarrow\) negligible (also \(\delta(s_m)\))
- \(\delta(\hat{s}_m - s_m) = \sum_{\lambda \in \Lambda} \left( \frac{1}{L} \int \varphi_\lambda(t)(dN_t - Ls(t)dt) \right)^2 = \chi^2(\hat{m})\)
- \(\mathbb{E}(\chi^2(m)) = \frac{1}{L} \sum_{\lambda \in \Lambda_m} \int \varphi_\lambda^2(t)s(t)dt\) ie variance
**An easy calculus (2)**

**Starting point**

\[
\|\hat{s}_m - s\|^2 \leq \|s - s_m\|^2 + \text{pen}(m) - 2\delta(s_m - s) + 2\delta(\hat{s}_m - s_m) - \text{pen}(\hat{m})
\]

- \(\delta(s_m) \to \text{negligeable (also } \delta(s_m)\)
- \(\delta(\hat{s}_m - s_m) = \sum_{\lambda \in \Lambda_m} \left(\frac{1}{L} \int \varphi_\lambda(t)(dN_t - Ls(t)dt)\right)^2 = \chi^2(\hat{m})\)
- \(\mathbb{E}(\chi^2(m)) = \frac{1}{L} \sum_{\lambda \in \Lambda_m} \int \varphi_\lambda^2(t)s(t)dt \text{ ie variance}\)
- Hence if \(\text{pen}(m) \simeq 2 \times \text{variance} \to \text{oracle inequality}\)
An easy calculus (2)

Starting point

\[ \| \hat{s}_m - s \|^2 \leq \| s - s_m \|^2 + \text{pen}(m) - 2\delta(s_m - s_{\hat{m}}) + 2\delta(\hat{s}_m - s_{\hat{m}}) - \text{pen}(\hat{m}) \]

- \( \delta(s_m) \rightarrow \) negligible (also \( \delta(s_{\hat{m}}) \))
- \( \delta(\hat{s}_m - s_{\hat{m}}) = \sum_{\lambda \in \Lambda_{\hat{m}}} (\frac{1}{L} \int \varphi_{\lambda}(t)(dN_t - Ls(t)dt))^2 = \chi^2(\hat{m}) \)
- \( \mathbb{E}(\chi^2(m)) = \frac{1}{L} \sum_{\lambda \in \Lambda_m} \int \varphi_{\lambda}^2(t)s(t)dt \) ie variance
- Hence if \( \text{pen}(m) \simeq 2 \times \text{variance} \rightarrow \) oracle inequality
- But \( \chi^2(\hat{m}) \rightarrow \) control of all the \( \chi^2(m) \)
An easy calculus (2)

Starting point

\[ \| \hat{s}_m - s \|^2 \leq \| s - s_m \|^2 + \text{pen}(m) - 2\delta(s_m - s_m) + 2\delta(\hat{s}_m - s_m) - \text{pen}(\hat{m}) \]

- \( \delta(s_m) \rightarrow \) negligible (also \( \delta(s_{\hat{m}}) \))
- \( \delta(\hat{s}_m - s_m) = \sum_{\lambda \in \Lambda_{\hat{m}}} \left( \frac{1}{L} \int \varphi_{\lambda}(t)(dN_t - Ls(t)dt) \right)^2 = \chi^2(\hat{m}) \)
- \( \mathbb{E}(\chi^2(m)) = \frac{1}{L} \sum_{\lambda \in \Lambda_m} \int \varphi_{\lambda}^2(t)s(t)dt \) ie variance
- Hence if \( \text{pen}(m) \simeq 2 \times \text{variance} \rightarrow \) oracle inequality
- But \( \chi^2(\hat{m}) \rightarrow \) control of all the \( \chi^2(m) \)
- Exponential inequality
Talagrand type inequality for Poisson processes

\[ \chi(m) = \frac{1}{T} \sup_{\|f\| = 1, f \in S_m} \int f(t)(dN_t - Ls(t)dt). \]
**Talagrand type inequality for Poisson processes**

\[ \chi(m) = \frac{1}{L} \sup_{\|f\|=1, f \in S_m} \int f(t)(dN_t - Ls(t)dt). \]

**Theorem (RB 2003)**

Let \( \{\psi_a, a \in A\} \) a countable family of functions with values in \([-b; b]\).
Talagrand type inequality for Poisson processes

\[ \chi(m) = \frac{1}{L} \sup_{f, \|f\|=1, f \in S_m} \int f(t)(dN_t - Ls(t)dt). \]

**Theorem (RB 2003)**

Let \( \{\psi_a, a \in A\} \) a countable family of functions with values in \([-b; b]\).

If \( Z = \sup_{a \in A} \int_X \psi_a(x)(dN_x - d\ell_x), \)
Talagrand type inequality for Poisson processes

\[ \chi(m) = \frac{1}{L} \sup_{\|f\|=1, f \in S_m} \int f(t)(dN_t - Ls(t)dt). \]

**Theorem (RB 2003)**

Let \( \{\psi_a, a \in A\} \) a countable family of functions with values in \([-b; b]\).

If \( Z = \sup_{a \in A} \int_X \psi_a(x)(dN_x - d\ell_x) \), then for all \( u, \varepsilon > 0 \),

\[ \mathbb{P}(Z \geq (1 + \varepsilon)\mathbb{E}(Z) + 2\sqrt{\kappa v u} + \kappa(\varepsilon)bu) \leq e^{-u}, \]

with \( v = \sup_{a \in A} \int_X \psi_a^2(x)d\ell_x \)

and \( \kappa = 6, \kappa(\varepsilon) = 1.25 + 32\varepsilon^{-1} \).
Talagrand type inequality for Poisson processes

\[ \chi(m) = \frac{1}{L} \sup_{\|f\|=1, f \in S_m} \int f(t)(dN_t - Ls(t)dt). \]

**Theorem (RB 2003)**

Let \( \{\psi_a, a \in A\} \) a countable family of functions with values in \([-b; b]\).

If \( Z = \sup_{a \in A} \int_X \psi_a(x)(dN_x - d\ell_x) \), then for all \( u, \varepsilon > 0 \),

\[ \mathbb{P}(Z \geq (1 + \varepsilon)\mathbb{E}(Z) + 2\sqrt{\kappa v u + \kappa(\varepsilon)bu}) \leq e^{-u}, \]

with \( v = \sup_{a \in A} \int_X \psi_a^2(x)d\ell_x \)

and \( \kappa = 6, \kappa(\varepsilon) = 1.25 + 32\varepsilon^{-1} \).
Application to $\chi(m)$

**Corollary (RB 2003)**

Let

$$M_m = \sup_{f \in S_m, \|f\| = 1} \int_X f^2(x)s(x)dx \quad \text{et} \quad B_m = \sup_{f \in S_m, \|f\| = 1} \|f\|_{\infty}. $$

then for all $u, \varepsilon > 0$,

$$\mathbb{P} \left( \chi(m) \geq (1 + \varepsilon) \sqrt{\frac{1}{L} \sum_{\lambda} \varphi_\lambda^2(x)s(x)dx} + \sqrt{\frac{2\kappa M_m u}{L} + \kappa(\varepsilon) \frac{B_m u}{L}} \right) \leq e^{-u}. $$
Oracle inequality for Poisson processes
simplified in the case of piecewise constant models on a fine grid $\Gamma$. 

Oracle inequality for Poisson processes
simplified in the case of piecewise constant models on a fine grid $\Gamma$.

**Proposition (RB 2003)**

Let $\{L_m, m \in \mathcal{M}\}$ such that $\sum_{m \in \mathcal{M}} e^{-L_m|m|} \leq \Sigma$ with $|\Gamma| \leq L(\ln L)^{-2}$. 
Oracle inequality for Poisson processes
simplified in the case of piecewise constant models on a fine grid $\Gamma$.

**Proposition (RB 2003)**

Let $\{L_m, m \in \mathcal{M}\}$ tq $\sum_{m \in \mathcal{M}} e^{-L_m|m|} \leq \Sigma$ with $|\Gamma| \leq L(\ln L)^{-2}$.

For all $c > 1$, if

$$\text{pen}(m) = \frac{c\tilde{M}|m|}{L}(1 + \sqrt{2\kappa L_m})^2 \text{ avec } \tilde{M} = \sup_{I \in \Gamma} \frac{N_I}{\mu(I)},$$
Oracle inequality for Poisson processes
simplified in the case of piecewise constant models on a fine grid $\Gamma$.

**Proposition (RB 2003)**

Let $\{L_m, m \in M\}$ tq $\sum_{m \in M} e^{-L_m|m|} \leq \Sigma$ with $|\Gamma| \leq L(\ln L)^{-2}$.

For all $c > 1$, if $\text{pen}(m) = \frac{c\tilde{M}|m|}{L}(1 + \sqrt{2\kappa L_m})^2$ avec $\tilde{M} = \sup_{I \in \Gamma} \frac{N_I}{\mu(I)}$,
then

$$
\mathbb{E}(\|s - \hat{s}_{\tilde{m}}\|^2) \leq \square_c \inf_{m \in M} \left[ \|s - s_m\|^2 + \frac{M|m|}{L}(1 + L_m) \right] + \square_{c,\Sigma,M} \frac{1}{L},
$$
Oracle inequality for Poisson processes
simplified in the case of piecewise constant models on a fine grid $\Gamma$.

**Proposition (RB 2003)**

\[
\text{Let } \{L_m, m \in \mathcal{M}\} \text{ tq } \sum_{m \in \mathcal{M}} e^{-L_m|m|} \leq \Sigma \text{ with } |\Gamma| \leq L(\ln L)^{-2}.
\]

For all $c > 1$, if

\[
\text{pen}(m) = \frac{c\tilde{M}|m|}{L}(1 + \sqrt{2\kappa L_m})^2 \text{ avec } \tilde{M} = \sup_{I \in \Gamma} \frac{N_I}{\mu(I)},
\]

then

\[
\mathbb{E}(\|s - \hat{s}_m\|^2) \leq \square_c \inf_{m \in \mathcal{M}} \left[ \|s - s_m\|^2 + \frac{M|m|}{L}(1 + L_m) \right] + \square_{c,\Sigma,M} \frac{1}{L},
\]

where $M = \sup_{I \in \Gamma} \frac{\int_I s(x)dx}{\mu(I)}$. 

Oracle inequality for Poisson processes
simplified in the case of piecewise constant models on a fine grid $\Gamma$.

**Proposition (RB 2003)**

Let $\{L_m, m \in \mathcal{M}\}$ tq $\sum_{m \in \mathcal{M}} e^{-L_m|m|} \leq \Sigma$ with $|\Gamma| \leq L(\ln L)^{-2}$.

For all $c > 1$, if

$$\text{pen}(m) = \frac{c\tilde{M}|m|}{L}(1 + \sqrt{2\kappa L_m})^2$$

avec $\tilde{M} = \sup_{I \in \Gamma} \frac{N_I}{\mu(I)}$,
then

$$\mathbb{E}(\|s - \hat{s}_{\tilde{m}}\|^2) \leq \square_c \inf_{m \in \mathcal{M}} \Big[ \|s - s_m\|^2 + \frac{M|m|}{L}(1 + L_m) \Big] + \square_{c, \Sigma, M} \frac{1}{L},$$

where $M = \sup_{I \in \Gamma} \frac{\int_I s(x)dx}{\mu(I)}$.

Here constants in the concentration inequalities are crucial $\rightarrow$ penalty.
Counting processes with linear intensities

\[ \lambda(t) = \Psi_s(t) \]

where \( \Psi(.) \) known predictable linear transformation. Functional parameter \( s \) unknown.
Counting processes with linear intensities

\[ \lambda(t) = \Psi_s(t) \]

where \( \Psi(.) \) known predictable linear transformation. Functional parameter \( s \) unknown.

- Poisson process on \( \mathbb{R} : \Psi_s(.) = Ls(.) \) with unknown function \( s \).
- Processus de Hawkes :

\[
\Psi_s(t)^{(r)} = \lambda^{(r)}(t) = \nu_r + \sum_{\ell=1}^{M} \int_{-\infty}^{t-} h^{(r)}_{\ell}(t - u) dN^{(\ell)}_u.
\]

with \( s = (\nu_r, h^{(r)}_{\ell})_{\ell, r} \)
Counting processes with linear intensities

\[ \lambda(t) = \Psi_s(t) \]

where \( \Psi(.) \) known predictable linear transformation. Functional parameter \( s \) unknown.

- Poisson process on \( \mathbb{R} \) : \( \Psi_s(.) = Ls(.) \) with unknown function \( s \).
- Processus de Hawkes :

\[
\Psi_s(t)^{(r)} = \lambda^{(r)}(t) = \nu_r + \sum_{\ell=1}^{M} \int_{-\infty}^{t} h^{(r)}_{\ell}(t - u) dN^{(\ell)}_u.
\]

with \( s = (\nu_r, h^{(r)}_{\ell})_{\ell, r} \)

Observation on \([0, T]\).
Least square contrast

\[ \gamma(f) = -\frac{2}{T} \int_{0}^{T} \psi_f(t)dN_t + \frac{1}{T} \int_{0}^{T} \psi_f(t)^2 dt. \]
Least square contrast

\[ \gamma(f) = -\frac{2}{T} \int_0^T \psi_f(t) dN_t + \frac{1}{T} \int_0^T \psi_f(t)^2 dt. \]

- taking the compensator,
  \[ \gamma(f) \simeq -\frac{2}{T} \int_0^T \psi_f(t) \psi_s(t) dt + \frac{1}{T} \int_0^T \psi_f(t)^2 dt \]
Least square contrast

\[ \gamma(f) = -\frac{2}{T} \int_{0}^{T} \Psi_f(t) dN_t + \frac{1}{T} \int_{0}^{T} \Psi_f(t)^2 dt. \]

taking the compensator,

\[ \gamma(f) \simeq -\frac{2}{T} \int_{0}^{T} \Psi_f(t) \Psi_s(t) dt + \frac{1}{T} \int_{0}^{T} \Psi_f(t)^2 dt = \frac{1}{T} \int_{0}^{T} \Psi_f - s(t)^2 dt - \frac{1}{T} \int_{0}^{T} \Psi_s(t)^2 dt. \]
Least square contrast

\[ \gamma(f) = -\frac{2}{T} \int_0^T \psi_f(t) dN_t + \frac{1}{T} \int_0^T \psi_f(t)^2 dt. \]

taking the compensator,

\[ \gamma(f) \simeq -\frac{2}{T} \int_0^T \psi_f(t) \psi_s(t) dt + \frac{1}{T} \int_0^T \psi_f(t)^2 dt = \]

\[ \frac{1}{T} \int_0^T \psi_{f-s}(t)^2 dt - \frac{1}{T} \int_0^T \psi_s(t)^2 dt. \]

minimal when \( \psi_{f-s}(t) = 0 \) a.s., a.e. \( \rightarrow f = s. \)
Least square contrast

$$\gamma(f) = -\frac{2}{T} \int_0^T \Psi_f(t) dN_t + \frac{1}{T} \int_0^T \Psi_f(t)^2 dt.$$ 

taking the compensator,

$$\gamma(f) \simeq -\frac{2}{T} \int_0^T \Psi_f(t) \Psi_s(t) dt + \frac{1}{T} \int_0^T \Psi_f(t)^2 dt = \frac{1}{T} \int_0^T \Psi_{f-s}(t)^2 dt - \frac{1}{T} \int_0^T \Psi_s(t)^2 dt.$$ 

minimal when $$\Psi_{f-s}(t) = 0$$ a.s., a.e. $$\rightarrow f = s$$.

In general, $$\frac{1}{T} \int_0^T \Psi_f(t)^2 dt$$ is random, true norm only with high probability.
Model selection and $\chi^2$

For each $S_m$, $\hat{s}_m = \arg\min_{f \in S_m} \gamma(f)$
Model selection and $\chi^2$

- For each $S_m$, $\hat{s}_m = \arg\min_{f \in S_m} \gamma(f)$
- Family $\mathcal{M}$ + penalty and

$$\hat{m} = \arg\min_{m \in \mathcal{M}} \{ \gamma(\hat{s}_m) + \text{pen}(m) \}.$$
Model selection and $\chi^2$

- For each $S_m$, $\hat{s}_m = \arg\min_{f \in S_m} \gamma(f)$
- Family $\mathcal{M} +$ penalty and
  $$\hat{m} = \arg\min_{m \in \mathcal{M}} \{ \gamma(\hat{s}_m) + \text{pen}(m) \}.$$  

- The statistics to control is
  $$\chi^2(m) = \sum_{\lambda \in \Lambda_m} \left( \frac{1}{T} \int_0^T \psi_\varphi(t)(dN_t - \Psi_s(t)dt) \right)^2.$$
Model selection and $\chi^2$

- For each $S_m$, $\hat{s}_m = \arg\min_{f \in S_m} \gamma(f)$
- Family $\mathcal{M} +$ penalty and

$$\hat{m} = \arg\min_{m \in \mathcal{M}} \{ \gamma(\hat{s}_m) + \text{pen}(m) \}.$$  

- The statistics to control is

$$\chi^2(m) = \sum_{\lambda \in \Lambda_m} \left( \frac{1}{T} \int_0^T \psi_{\varphi \lambda}(t)(dN_t - \psi_s(t)dt) \right)^2.$$  

- Once again

$$\chi(m) = \sup_{\|f\| = 1, f \in S_m} \frac{1}{T} \int \psi_f(t)(dN_t - \psi_s(t)dt).$$
"Talagrand" type inequality for general counting processes

Theorem (RB 2006)

Let $\lambda(t)$ be a.s integrable on $[0, T]$. 


"Talagrand" type inequality for general counting processes

Theorem (RB 2006)

Let $\lambda(t)$ be a.s integrable on $[0, T]$. Let $\{(H_{a,t})_{t\geq 0}, a \in A\}$ be a countable family of predictable process

$$\forall t \geq 0, \quad Z_t = \sup_{a \in A} \int_0^t H_{a,s}(dN_s - \lambda(s)ds).$$
"Talagrand" type inequality for general counting processes

**Theorem (RB 2006)**

Let $\lambda(t)$ be a.s integrable on $[0, T]$. Let $\{(H_{a,t})_{t\geq 0}, a \in A\}$ be a countable family of predictable process

$$\forall t \geq 0, \quad Z_t = \sup_{a \in A} \int_0^t H_{a,s}(dN_s - \lambda(s)ds).$$

Then its compensator exists $(A_t)_{t \geq 0}$, it is positive and non decreasing and

$$\forall 0 \leq t \leq T, \quad Z_t - A_t = \int_0^t \Delta Z(s)(dN_s - \lambda(s)ds),$$

for a predictable $\Delta Z(s)$ st $\Delta Z(s) \leq \sup_{a \in A} H_{a,s}$. 
"Talagrand" type inequality for general counting processes

Theorem (RB 2006)

Let $\lambda(t)$ be a.s integrable on $[0, T]$.
Let $\{(H_a,t)_{t \geq 0}, a \in A\}$ be a countable family of predictable processes

$$\forall t \geq 0, \quad Z_t = \sup_{a \in A} \int_0^t H_{a,s}(dN_s - \lambda(s)ds).$$

If the $H_a$ have values in $[-b, b]$ and if $\int_0^T \sup_{a \in A} H_{a,s}^2 \lambda(s)ds \leq v$ as, then for all $u > 0$,

$$\mathbb{P} \left( \sup_{[0,T]} (Z_t - A_t) \geq \sqrt{2vu} + \frac{bu}{3} \right) \leq e^{-u}.$$
And for the $\chi^2$ ...

Let

$$C = \sum_\lambda \int_0^T \frac{\psi_{\varphi\lambda}(x)^2}{T^2} \lambda(x) dx,$$

with $C \leq v$ et $\sum_\lambda \psi_{\varphi\lambda}(x)^2 \leq b$ for all $x \in [0, T]$. Then for all $u > 0$,

$$\mathbb{P} \left( \chi(m) \geq \sqrt{C} + 3\sqrt{2}vu + bu \right) \leq 2e^{-u}.$$
And for the $\chi^2$ ...

Let

$$C = \sum_{\lambda} \int_{0}^{T} \frac{\psi_{\varphi,\lambda}(x)^2}{T^2} \lambda(x) dx,$$

with $C \leq v$ et $\sum_{\lambda} \psi_{\varphi,\lambda}(x)^2 \leq b$ for all $x \in [0, T]$. Then for all $u > 0$,

$$\mathbb{P} \left( \chi(m) \geq \sqrt{C} + 3\sqrt{2}vu + bu \right) \leq 2e^{-u}.$$

$v$ is of the order of $D_m \neq$ Poisson case → a ”worse” oracle inequality (family of models to be handle are smaller)
And for the $\chi^2$ ...

Let

$$C = \sum_\lambda \int_0^T \frac{\psi_{\varphi_\lambda}(x)^2}{T^2} \lambda(x)dx,$$

with $C \leq \nu$ et $\sum_\lambda \psi_{\varphi_\lambda}(x)^2 \leq b$ for all $x \in [0, T]$. Then for all $u > 0$,

$$\mathbb{P}\left(\chi(m) \geq \sqrt{C} + 3\sqrt{2\nu u} + bu\right) \leq 2e^{-u}.$$

- $\nu$ is of the order of $D_m \neq$ Poisson case $\rightarrow$ a "worse" oracle inequality (family of models to be handle are smaller)
- Improvement sometimes possible Baraud (2010) but need of an upper bound on $\sqrt{C}$. 
And for the $\chi^2$ ...

Let

$$C = \sum_{\lambda} \int_0^T \frac{\psi_{\varphi}(x)^2}{T^2} \lambda(x) dx,$$

with $C \leq \nu$ et $\sum_{\lambda} \psi_{\varphi}(x)^2 \leq b$ for all $x \in [0, T]$. Then for all $u > 0$,

$$\mathbb{P} \left( \chi(m) \geq \sqrt{C} + 3\sqrt{2\nu u} + bu \right) \leq 2e^{-u}.$$

- $\nu$ is of the order of $D_m \neq$ Poisson case → a "worse" oracle inequality (family of models to be handle are smaller)
- Improvement sometimes possible Baraud (2010) but need of an upper bound on $\sqrt{C}$.
- Still $\lambda$ inside, which is in general difficult to estimate → usually assume known upper bound.
Concrete Problems due to the concentration...

- No theoretical access to a fully data-driven penalty.
Concrete Problems due to the concentration...

- No theoretical access to a fully data-driven penalty.
- Even in the Poisson case, variance upper bounded and then overestimation ... of the upper bound.
Concrete Problems due to the concentration...

- No theoretical access to a fully data-driven penalty.
- Even in the Poisson case, variance upper bounded and then overestimation ... of the upper bound.
- We would like to be closer to the true variance of $\hat{s}_m$ and estimate it without bias.
Concrete Problems due to the concentration...

- No theoretical access to a fully data-driven penalty.
- Even in the Poisson case, variance upper bounded and then overestimation ... of the upper bound.
- We would like to be closer to the true variance of $\hat{s}_m$ and estimate it without bias.
- Talagrand type inequalities lead us to estimate the supremum of the variances (Poisson) or the variance of the supremum
Poisson process and Thresholding

\[ \| \hat{s}_m - s \|^2 \leq \| s - s_m \|^2 + \text{pen}(m) - 2\delta(s_m - \hat{s}_m) + 2\delta(\hat{s}_m - s_m) - \text{pen}(\hat{m}) \]

Here there exists a large ONB \((\varphi_\lambda, \lambda \in \Lambda)\) and for \(m \subset \Lambda\),
\[ S_m = \text{Span}(\varphi_\lambda, \lambda \in m) \]
Poisson process and Thresholding

\[ \| \hat{s}_m - s \|^2 \leq \| s - s_m \|^2 + \text{pen}(m) - 2\delta(s_m - s_m) + 2\delta(\hat{s}_m - s_m) - \text{pen}(\hat{m}) \]

- Here there exists a large ONB \((\varphi_\lambda, \lambda \in \Lambda)\) and for \(m \subset \Lambda\),
  \[ S_m = \text{Span}(\varphi_\lambda, \lambda \in m) \]
- \[ \beta_\lambda = \int \varphi_\lambda s, \quad \hat{\beta}_\lambda = (1/L) \int \varphi_\lambda dN \]
Poisson process and Thresholding

\[ \| \hat{s}_m - s \|^2 \leq \| s - s_m \|^2 + \text{pen}(m) - 2\delta(s_m - \hat{s}_m) + 2\delta(\hat{s}_m - s_m) - \text{pen}(\hat{m}) \]

- Here there exists a large ONB \((\varphi_\lambda, \lambda \in \Lambda)\) and for \(m \subset \Lambda\),

\[ S_m = \text{Span}(\varphi_\lambda, \lambda \in m) \]

- \(\beta_\lambda = \int \varphi_\lambda s, \quad \hat{\beta}_\lambda = (1/L) \int \varphi_\lambda dN\)

- \(\delta(\hat{s}_m - s_m) = \chi^2(\hat{m}) = \sum_{\lambda \in \hat{m}} \left( \frac{1}{L} \int \varphi_\lambda(t)(dN_t - s(t)dt) \right)^2 \)
Poisson process and Thresholding

\[ \| \hat{s}_m - s \|^2 \leq \| s - s_m \|^2 + \text{pen}(m) - 2\delta(s_m - s) + 2\delta(\hat{s}_m - s) - \text{pen}(\hat{m}) \]

- Here there exists a large ONB \((\varphi_\lambda, \lambda \in \Lambda)\) and for \(m \subset \Lambda\), \(S_m = \text{Span}(\varphi_\lambda, \lambda \in m)\)
- \(\beta_\lambda = \int \varphi_\lambda s, \quad \hat{\beta}_\lambda = (1/L) \int \varphi_\lambda dN\)
- \(\delta(\hat{s}_m - s) = \chi^2(\hat{m}) = \sum_{\lambda \in \hat{m}} \left( \frac{1}{L} \int \varphi_\lambda(t)(dN_t - s(t)dt) \right)^2\)
- If \(\hat{m}\) better understood, not forced to control all the \(\chi(m)\).
Poisson process and Thresholding

\[ \| \hat{s}_m - s \|^2 \leq \| s - s_m \|^2 + \text{pen}(m) - 2\delta(s_m - s) + 2\delta(\hat{s}_m - s) - \text{pen}(\hat{m}) \]

- Here there exists a large ONB \((\varphi_\lambda, \lambda \in \Lambda)\) and for \(m \subset \Lambda\), 
  \( S_m = \text{Span}(\varphi_\lambda, \lambda \in m) \)

- \( \beta_\lambda = \int \varphi_\lambda s, \ \hat{\beta}_\lambda = (1/L) \int \varphi_\lambda dN \)

- \( \delta(\hat{s}_m - s_m) = \chi^2(\hat{m}) = \sum_{\lambda \in \hat{m}} \left( \frac{1}{L} \int \varphi_\lambda(t)(dN_t - s(t)dt) \right)^2 \)

- If \( \hat{m} \) better understood, not forced to control all the \( \chi(m) \).

- If \( \mathcal{M} = \{ m \subset \Gamma \} \), where \( \Gamma \) finite subset of \( \Lambda \) and if 
  \( \text{pen}(m) = \sum_{\lambda \in m} \eta^2_\lambda \) then
Poisson process and Thresholding

\[ \| \hat{s}_m - s \|^2 \leq \| s - s_m \|^2 + \text{pen}(m) - 2\delta(s_m - s) + 2\delta(\hat{s}_m - \hat{s}_m) - \text{pen}(\hat{m}) \]

- Here there exists a large ONB \((\varphi_\lambda, \lambda \in \Lambda)\) and for \(m \subset \Lambda\),
  \[ S_m = \text{Span}(\varphi_\lambda, \lambda \in m) \]
- \( \beta_\lambda = \int \varphi_\lambda s, \ \hat{\beta}_\lambda = (1/L) \int \varphi_\lambda dN \)
- \( \delta(\hat{s}_m - s) = \chi^2(m) = \sum_{\lambda \in \hat{m}} \left( \frac{1}{L} \int \varphi_\lambda(t)(dN_t - s(t)dt) \right)^2 \)
- If \( \hat{m} \) better understood, not forced to control all the \( \chi(m) \).
- If \( M = \{ m \subset \Gamma \} \), where \( \Gamma \) finite subset of \( \Lambda \) and if
  \[ \text{pen}(m) = \sum_{\lambda \in m} \eta_\lambda^2 \] then
  \[ \hat{m} = \arg\min_{m \in M} (\gamma(\hat{s}_m) + \text{pen}(m)) \].
Poisson process and Thresholding

\[ \| \hat{s}_m - s \|^2 \leq \| s - s_m \|^2 + \text{pen}(m) - 2\delta(s_m - s_{\hat{m}}) + 2\delta(\hat{s}_m - s_{\hat{m}}) - \text{pen}(\hat{m}) \]

- Here there exists a large ONB \((\varphi_\lambda, \lambda \in \Lambda)\) and for \(m \subset \Lambda\),
  \(S_m = \text{Span}(\varphi_\lambda, \lambda \in m)\)
- \(\beta_\lambda = \int \varphi_\lambda s, \hat{\beta}_\lambda = (1/L) \int \varphi_\lambda dN\)
- \(\delta(\hat{s}_m - s_{\hat{m}}) = \chi^2(\hat{m}) = \sum_{\lambda \in \hat{m}} \left( \frac{1}{L} \int \varphi_\lambda(t)(dN_t - s(t)dt) \right)^2\)
- If \(\hat{m}\) better understood, not forced to control all the \(\chi(m)\).
- If \(\mathcal{M} = \{m \subset \Gamma\}\), where \(\Gamma\) finite subset of \(\Lambda\) and if \(\text{pen}(m) = \sum_{\lambda \in m} \eta^2_\lambda\) then

\[ \hat{m} = \{ \lambda \in \Gamma \mid |\hat{\beta}_\lambda| > \eta_\lambda \} \]
Poisson process and Thresholding

\[ \| \hat{s}_m - s \|^2 \leq \| s - s_m \|^2 + \text{pen}(m) - 2\delta(s_m - s) + 2\delta(\hat{s}_m - \hat{s}_m) - \text{pen}(\hat{m}) \]

- Here there exists a large ONB \((\varphi_\lambda, \lambda \in \Lambda)\) and for \(m \subset \Lambda\), \(S_m = \text{Span}(\varphi_\lambda, \lambda \in m)\)
- \(\beta_\lambda = \int \varphi_\lambda s, \ \hat{\beta}_\lambda = (1/L) \int \varphi_\lambda dN\)
- \(\delta(\hat{s}_m - s_m) = \chi^2(\hat{m}) = \sum_{\lambda \in \hat{m}} \left( \frac{1}{L} \int \varphi_\lambda(t)(dN_t - s(t)dt) \right)^2\)
- If \(\hat{m}\) better understood, not forced to control all the \(\chi(m)\).
- If \(\mathcal{M} = \{m \subset \Gamma\}\), where \(\Gamma\) finite subset of \(\Lambda\) and if \(\text{pen}(m) = \sum_{\lambda \in m} \eta_\lambda^2\) then

\[ \hat{m} = \{ \lambda \in \Gamma / |\hat{\beta}_\lambda| > \eta_\lambda \}. \]

\[ \chi^2(\hat{m}) = \sum_{\lambda \in \Gamma} (\hat{\beta}_\lambda - \beta_\lambda)^2 1_{|\hat{\beta}_\lambda| > \eta_\lambda}. \]
A general thresholding theorem

Theorem (RB Rivoirard 2010)

Let $\beta = (\beta_\lambda)_{\lambda \in \Lambda}$ s.t $\|\beta\|_{\ell_2} < \infty$ be unknown. Let us observe $(\hat{\beta}_\lambda)_{\lambda \in \Gamma}$, where $\Gamma \subset \Lambda$ and $(\eta_\lambda)_{\lambda \in \Gamma}$. 
A general thresholding theorem

**Theorem (RB Rivoirard 2010)**

Let \( \beta = (\beta_\lambda)_{\lambda \in \Lambda} \) st \( \|\beta\|_{\ell_2} < \infty \) be unknown. Let us observe \( (\hat{\beta}_\lambda)_{\lambda \in \Gamma} \), where \( \Gamma \subset \Lambda \) and \( (\eta_\lambda)_{\lambda \in \Gamma} \).

Let \( \tilde{\beta} = (\hat{\beta}_\lambda \mathbf{1}_{|\hat{\beta}_\lambda| \geq \eta_\lambda})_{\lambda \in \Lambda} \).
A general thresholding theorem

**Theorem (RB Rivoirard 2010)**

Let $\beta = (\beta_\lambda)_{\lambda \in \Lambda}$ s.t $\|\beta\|_{\ell_2} < \infty$ be unknown. Let us observe $(\hat{\beta}_\lambda)_{\lambda \in \Gamma}$, where $\Gamma \subset \Lambda$ and $(\eta_\lambda)_{\lambda \in \Gamma}$.

Let $\tilde{\beta} = (\hat{\beta}_\lambda \mathbf{1}_{|\hat{\beta}_\lambda| \geq \eta_\lambda})_{\lambda \in \Lambda}$.

Let $\epsilon > 0$ be fixed. If one finds $(F_\lambda)_{\lambda \in \Gamma}$ and $\kappa \in [0, 1[$, $\omega \in [0, 1]$, $\zeta > 0$ s.t
## A general thresholding theorem

**Theorem (RB Rivoirard 2010)**

*Let \( \beta = (\beta_\lambda)_{\lambda \in \Lambda} \) st \( \|\beta\|_2 < \infty \) be unknown. Let us observe \((\hat{\beta}_\lambda)_{\lambda \in \Gamma} \), where \( \Gamma \subset \Lambda \) and \((\eta_\lambda)_{\lambda \in \Gamma} \).

*Let \( \tilde{\beta} = (\hat{\beta}_\lambda \mathbf{1}_{|\hat{\beta}_\lambda| \geq \eta_\lambda} \mathbf{1}_{\lambda \in \Gamma})_{\lambda \in \Lambda} \).*

*Let \( \epsilon > 0 \) be fixed. If one finds \((F_\lambda)_{\lambda \in \Gamma} \) and \( \kappa \in [0, 1[ \), \( \omega \in [0, 1] \), \( \zeta > 0 \) st

\[(A1) \text{ For all } \lambda \text{ in } \Gamma, \ \mathbb{P} \left( |\hat{\beta}_\lambda - \beta_\lambda| > \kappa \eta_\lambda \right) \leq \omega.\]
A general thresholding theorem

Theorem (RB Rivoirard 2010)

Let $\beta = (\beta_\lambda)_{\lambda \in \Lambda}$ st $\|\beta\|_{\ell_2} < \infty$ be unknown. Let us observe $(\hat{\beta}_\lambda)_{\lambda \in \Gamma}$, where $\Gamma \subset \Lambda$ and $(\eta_\lambda)_{\lambda \in \Gamma}$.

Let $\beta = (\beta_\lambda \mathbf{1}_{|\hat{\beta}_\lambda| \geq \eta_\lambda})_{\lambda \in \Lambda}$.

Let $\epsilon > 0$ be fixed. If one finds $(F_\lambda)_{\lambda \in \Gamma}$ and $\kappa \in [0, 1[, \omega \in [0, 1], \zeta > 0$ st

(A1) For all $\lambda$ in $\Gamma$, $\mathbb{P} \left( |\hat{\beta}_\lambda - \beta_\lambda| > \kappa \eta_\lambda \right) \leq \omega$.

(A2) There exists $1 < a, b < \infty$ with $\frac{1}{a} + \frac{1}{b} = 1$ and $G > 0$ st $\lambda \in \Gamma$, $\left( \mathbb{E} \left[ |\hat{\beta}_\lambda - \beta_\lambda|^{2a} \right] \right)^{\frac{1}{a}} \leq G \max \left( F_\lambda, \frac{F_\lambda^{\frac{1}{a}} \epsilon^{\frac{1}{b}}}{\sqrt{\kappa}} \right)$. 
A general thresholding theorem

**Theorem (RB Rivoirard 2010)**

Let \( \beta = (\beta_\lambda)_{\lambda \in \Lambda} \) s.t. \( \|\beta\|_{\ell_2} < \infty \) be unknown. Let us observe \((\hat{\beta}_\lambda)_{\lambda \in \Gamma}\), where \( \Gamma \subset \Lambda \) and \((\eta_\lambda)_{\lambda \in \Gamma}\).

Let \( \tilde{\beta} = (\hat{\beta}_\lambda \mathbf{1}_{|\hat{\beta}_\lambda| \geq \eta_\lambda})_{\lambda \in \Lambda} \).

Let \( \epsilon > 0 \) be fixed. If one finds \((F_\lambda)_{\lambda \in \Gamma}\) and \( \kappa \in [0, 1[, \omega \in [0, 1], \zeta > 0 \) s.t.

(A1) For all \( \lambda \) in \( \Gamma \),
\[
P \left( |\hat{\beta}_\lambda - \beta_\lambda| > \kappa \eta_\lambda \right) \leq \omega.
\]

(A2) There exists \(1 < a, b < \infty\) with \(\frac{1}{a} + \frac{1}{b} = 1\) and \( G > 0 \) s.t.
\[
\lambda \in \Gamma, \quad \left( \mathbb{E} \left[ |\hat{\beta}_\lambda - \beta_\lambda|^{2a} \right] \right)^{\frac{1}{a}} \leq G \max \left( F_\lambda, F_\lambda^{\frac{1}{a}} \epsilon^{\frac{1}{b}} \right).
\]

(A3) there exists \( \tau \) s.t. for all \( \lambda \) in \( \Gamma \) / \( F_\lambda < \tau \epsilon \),
\[
P \left( |\hat{\beta}_\lambda - \beta_\lambda| > \kappa \eta_\lambda, |\hat{\beta}_\lambda| > \eta_\lambda \right) \leq F_\lambda \zeta.
\]
A general thresholding theorem (2)

**Theorem (RB Rivoirard 2010)**

*Then under (A1), (A2), (A3),*

\[
\mathbb{E}\|\tilde{\beta} - \beta\|^2_{\ell_2} \leq \\
\kappa \mathbb{E} \inf_{m \subset \Gamma} \left\{ \sum_{\lambda \notin m} \beta_\lambda^2 + \sum_{\lambda \in m} (\hat{\beta}_\lambda - \beta_\lambda)^2 + \sum_{\lambda \in m} \eta_\lambda^2 \right\} \\
+ \cdots \sum_{\lambda \in \Gamma} F_\lambda
\]

\[
\leq \kappa \mathbb{E} \inf_{m \subset \Gamma} [\|s - s_m\|^2 + \text{pen}(m)] + \text{reminder term}
\]
Bernstein and variance estimation

For all $u > 0$,

$$P \left( |\hat{\beta}_\lambda - \beta_\lambda| \geq \sqrt{2uV_\lambda} + \frac{\| \varphi_\lambda \|_\infty u}{3L} \right) \leq 2e^{-u},$$

with $V_\lambda = \frac{1}{L} \int \varphi_\lambda^2(x)s(x)dx$
Bernstein and variance estimation

For all $u > 0$,

$$\mathbb{P} \left( |\hat{\beta}_\lambda - \beta_\lambda| \geq \sqrt{2u V_\lambda} + \frac{\|\varphi_\lambda\|_\infty u}{3L} \right) \leq 2e^{-u},$$

with $V_\lambda = \frac{1}{L} \int \varphi_\lambda^2(x)s(x)dx$

and also

$$\mathbb{P} \left( V_\lambda \geq \tilde{V}_\lambda(u) \right) \leq e^{-u},$$

with

$$\tilde{V}_\lambda(u) = \hat{V}_\lambda + \sqrt{2\hat{V}_\lambda \frac{\|\varphi_\lambda\|_\infty^2}{L^2}} u + 3\frac{\|\varphi_\lambda\|_\infty^2}{n^2} u,$$

where $\hat{V}_\lambda = \frac{1}{L^2} \int \varphi_\lambda^2(x)dN_x$. 
Bernstein and variance estimation

For all \( u > 0 \),

\[
\mathbb{P} \left( |\hat{\beta}_\lambda - \beta_\lambda| \geq \sqrt{2uV_\lambda} + \frac{\|\varphi\|_\infty u}{3L} \right) \leq 2e^{-u},
\]

with \( V_\lambda = \frac{1}{L} \int \varphi_\lambda^2(x)s(x)dx \)

and also

\[
\mathbb{P} \left( V_\lambda \geq \tilde{V}_\lambda(u) \right) \leq e^{-u}
\]

with

\[
\tilde{V}_\lambda(u) = \hat{V}_\lambda + \sqrt{2\hat{V}_\lambda \frac{\|\varphi\|_\infty^2}{L^2} u + 3\frac{\|\varphi\|_\infty^2}{n^2} u},
\]

where \( \hat{V}_\lambda = \frac{1}{L^2} \int \varphi_\lambda^2(x)dN_x \).

Hence

\[
\mathbb{P}(\hat{\beta}_\lambda - \beta_\lambda > \eta_\lambda(u)) \leq 3e^{-u}
\]

with \( \eta_\lambda(u) = \sqrt{2u\tilde{V}_\lambda(u) + \frac{\|\varphi\|_\infty u}{3L}} \).
Lasso for other counting processes

Reformulation of the least-square contrast:

\[
\gamma(f) = -\frac{2}{T} \int_0^T \psi_f(t) dN_t + \frac{1}{T} \int_0^T \psi_f(t)^2 dt.
\]
Lasso for other counting processes

Reformulation of the least-square contrast:

\[ \gamma(f) = -\frac{2}{T} \int_0^T \psi_f(t) dN_t + \frac{1}{T} \int_0^T \psi_f(t)^2 dt. \]

Let \( \Phi \) be a dictionary of \( \mathcal{H} \) and if \( a \in \mathbb{R}^\Phi \),

\[ f_a = \sum_{\varphi \in \Phi} a_\varphi \varphi. \]

Then

\[ \gamma(f) = -2b^*a + a^*Ga \]

where

- \( G \) is a random observable matrix.
- \( b \) is also a random observable vector.
Lasso criterion

\[ \hat{a} = \arg\min_{a \in \mathbb{R}} \{-2b^*a + a^*Ga + 2d^*|a|\} \]

- The vector $d^*$ is not constant: it is random and depends on the index, same role as the threshold $\eta$. 
**Lasso criterion**

\[ \hat{a} = \arg\min_{a \in \mathbb{R}} \Phi \{-2b^*a + a^*Ga + 2d^*|a|\} \]

- The vector \( d^* \) is not constant: it is random and depends on the index, same role as the threshold \( \eta \)
- \( \rightarrow \) data-driven penalty (see also Bertin, Le Pennec, Rivoirard (2011) in the density setting)
The vector $d^*$ is not constant: it is random and depends on the index, same role as the threshold $\eta$.

→ data-driven penalty (see also Bertin, Le Pennec, Rivoirard (2011) in the density setting).

Oracle inequality with "high" probability possible....

$Lasso criterion$

\[ \hat{a} = \arg\min_{a \in \mathbb{R}^p} \{-2b^*a + a^*Ga + 2d^*|a|\} \]
One of the main probabilistic ingredients

Bernstein type inequality for counting processes

Let \((H_s)_{s \geq 0}\) be a predictable process and

\[ M_t = \int_0^t H_s (dN_s - \lambda(s)ds). \]
One of the main probabilistic ingredients

Bernstein type inequality for counting processes

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 \(v > w > 0\).
One of the main probabilistic ingredients

Bernstein type inequality for counting processes

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 \(\phi(u) = \exp(u) - u - 1\).
One of the main probabilistic ingredients

Bernstein type inequality for counting processes

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

\[
\hat{V}_\mu^\tau = \frac{\mu}{\mu - \phi(\mu)} \int_0^\tau H_s^2 dN_s + \frac{b^2 x}{\mu - \phi(\mu)}, \quad \text{where } \phi(u) = \exp(u) - u - 1.
\]

Then for every stopping time \(\tau\) and every \(\varepsilon > 0\)

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

Bernstein type inequality for counting processes

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
\[
\hat{V}_\mu^\tau = \frac{\mu}{\mu - \phi(\mu)} \int_0^\tau H_s^2 dN_s + \frac{b^2 x}{\mu - \phi(\mu)}, \text{ where } \phi(u) = \exp(u) - u - 1.
\]

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

We apply it to \( \int_0^T \psi_\varphi(t)[dN_t - \lambda(t) dt] \). Then \( d \) is given by the right hand-side.
One of the main probabilistic ingredients

Bernstein type inequality for counting processes

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

We apply it to \(\int_0^T \Psi_\varphi(t)[dN_t - \lambda(t)dt]\). Then \(\mathbf{d}\) is given by the right hand-side.
For more details about the Lasso procedure, see V. Rivoirard’s talk.
Sketch of proof

\[ E_t = \exp(\xi \int_0^t H_s d(N-\Lambda)_s) - \int_0^t \phi(\xi H_s)\lambda(s)ds \] is a supermartingale.
Sketch of proof

- $E_t = \exp(\xi \int_0^t H_s d(N - \Lambda)_s - \int_0^t \phi(\xi H_s) \lambda(s) ds)$ is a supermartingale.

- For all $\xi \in (0, 3),
  \[ P\left( M_{\tau} \geq \frac{\xi}{2(1-\xi/3)} \int_0^\tau H_s^2 \lambda(s) ds + \xi^{-1}x \text{ and } \sup_{s \leq \tau} |H_s| \leq 1 \right) \leq e^{-x} \]
Sketch of proof

- \( E_t = \exp(\xi \int_0^t H_s d(N - \Lambda)_s - \int_0^t \phi(\xi H_s)\lambda(s)ds) \) is a supermartingale.

- For all \( \xi \in (0, 3) \),
  \[
  \mathbb{P} \left( M_\tau \geq \frac{\xi}{2(1-\xi/3)} \int_0^\tau H_s^2 \lambda(s)ds + \xi^{-1}x \text{ and } \sup_{s \leq \tau} |H_s| \leq 1 \right) 
  \leq e^{-x}
  \]

- \[
  \mathbb{P} \left( M_\tau \geq \frac{\xi}{2(1-\xi/3)} v + \xi^{-1}x \text{ and } \int_0^\tau H_s^2 \lambda(s)ds \leq v \text{ and } \sup_{s \leq \tau} |H_s| \leq 1 \right) 
  \leq e^{-x}.
  \]
Sketch of proof (2)

**Lemma**

Let $a$, $b$ and $x$ be positive constants and let us consider on $(0, 1/b)$, $g(\xi) = \frac{a\xi}{(1-b\xi)} + \frac{x}{\xi}$. Then $\min_{\xi \in (0,1/b)} g(\xi) = 2\sqrt{ax} + bx$ and the minimum is achieved in $\xi(a, b, x) = \frac{xb - \sqrt{ax}}{xb^2 - a}$. 
Sketch of proof (2)

**Lemma**

Let $a$, $b$ and $x$ be positive constants and let us consider on $(0, 1/b)$, $g(\xi) = \frac{a\xi}{1-b\xi} + \frac{x}{\xi}$. Then $\min_{\xi \in (0, 1/b)} g(\xi) = 2\sqrt{ax} + bx$ and the minimum is achieved in $\xi(a, b, x) = \frac{xb - \sqrt{ax}}{xb^2 - a}$.

Then with $\xi(v/2, 1/3, x)$,

$$\mathbb{P}(M_\tau \geq \sqrt{2vx} + x/3 \text{ and } \int_0^\tau H_s^2 \lambda(s)ds \leq v \text{ and } \sup_{s \leq \tau} |H_s| \leq 1) \leq e^{-x}.$$
Sketch of proof (2)

Lemma

Let $a$, $b$ and $x$ be positive constants and let us consider on $(0, 1/b)$, \( g(\xi) = \frac{a\xi}{1-b\xi} + \frac{x}{\xi} \). Then \( \min_{\xi \in (0, 1/b)} g(\xi) = 2\sqrt{ax} + bx \) and the minimum is achieved in \( \xi(a, b, x) = \frac{xb - \sqrt{ax}}{xb^2 - a} \).

- Then with \( \xi(v/2, 1/3, x) \),
  \[
  \mathbb{P} \left( M_\tau \geq \sqrt{2vx} + x/3 \text{ and } \int_0^\tau H_s^2 \lambda(s) ds \leq v \text{ and } \sup_{s \leq \tau} |H_s| \leq 1 \right) \leq e^{-x}.
  \]
- But also
  \[
  \mathbb{P} \left( M_\tau \geq \sqrt{2(1 + \varepsilon) \int_0^\tau H_s^2 \lambda(s) ds} + x/3 \text{ and } \int_0^\tau H_s^2 \lambda(s) ds \leq v \text{ and } \sup_{s \leq \tau} |H_s| \leq 1 \right) \leq e^{-x}.
  \]
- Peeling + plug in ...
Conclusion

- If the concentration inequalities for the test statistics or the $\chi^2$ statistics are "tight" (dimension free) enough, possibility to aggregate / select in a large/complex family and hence be able to adapt to "ugly" situations.
Conclusion

- If the concentration inequalities for the test statistics or the $\chi^2$ statistics are "tight" (dimension free) enough, possibility to aggregate / select in a large/complex family and hence be able to adapt to "ugly" situations.

- For estimation, also need of
  - known, sharp constants
Conclusion

- If the concentration inequalities for the test statistics or the $\chi^2$ statistics are "tight" (dimension free) enough, possibility to aggregate / select in a large/complex family and hence be able to adapt to "ugly" situations.
- For estimation, also need of
  - known, sharp constants
  - observable quantities, eventually random ...
Conclusion

- If the concentration inequalities for the test statistics or the $\chi^2$ statistics are "tight" (dimension free) enough, possibility to aggregate / select in a large/complex family and hence be able to adapt to "ugly" situations.

- For estimation, also need of
  - known, sharp constants
  - observable quantities, eventually random ...
  - eventually change of method (threshold, Lasso)...
Conclusion

- If the concentration inequalities for the test statistics or the $\chi^2$ statistics are "tight" (dimension free) enough, possibility to aggregate / select in a large/complex family and hence be able to adapt to "ugly" situations.

- For estimation, also need of
  - known, sharp constants
  - observable quantities, eventually random ...
  - eventually change of method (threshold, Lasso)...

- Future work: multiple testing, group Lasso ???
References


Thank you !