Statistics for processes

I likelihood

For a given model $X \sim \mathbb{P}_\theta$, $\theta \in \Theta$

then you can compute $g_\theta(x) = \mathbb{P}_\theta(X = x)$

The likelihood is $\theta \longmapsto L_\theta(X) = g_\theta(X)$

log likelihood is $\log L_\theta(X)$.

And the MLE is $\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} \, L_\theta(X) = \underset{\theta \in \Theta}{\operatorname{argmax}} \, \ell_\theta(X)$

⚠ you need θ to be the minimal possible set of parameters.

1/ Markov chain

$X_0, X_1, \dots, X_n$ Markov chain with value in $\{0,1\}$

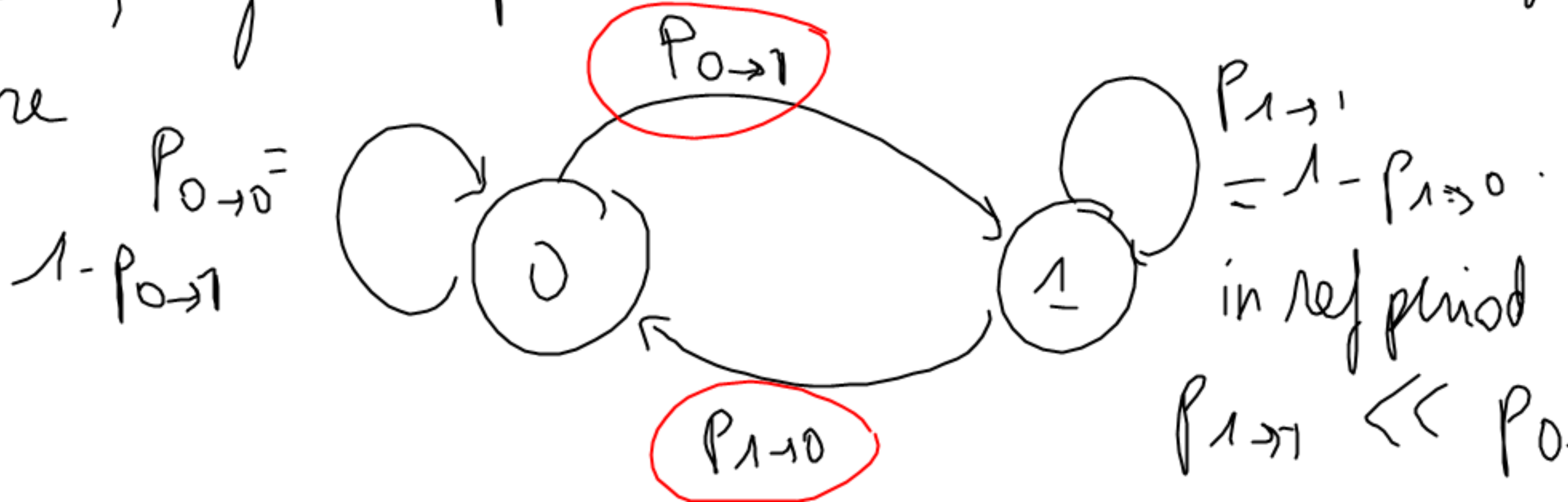
ex spike trains (binned)

0 0 0 1 0 0 1 1 0 0 0 . . .

there is at least one spike in this bin

no spike in this bin.

when refractory period happens, after a spike a neuron is less likely to produce another spike



in ref period

$P_{1\to 1} \ll P_{0\to 1}$

$P = \mathbb{P}(X_0 = 0) \quad +$

$$\theta = (p_{\cdot}, p_{1\to 0}, p_{0\to 1}) \in [0,1]^3$$

for a given realisation $x_0, \dots, x_n$

$$\mathbb{P}_\theta\big((X_0, \dots, X_n) = (x_0, \dots, x_n)\big)$$

$$= \mathbb{P}_\theta(X_0 = x_0 \;\cdots\; X_n = x_n)$$

$$= \mathbb{P}_\theta(X_0 = x_0, \dots, X_{n-1} = x_{n-1}) \underbrace{\mathbb{P}_\theta(X_n = x_n \mid \text{all the path from } 0 \text{ to } n-1)}_{\text{Markov property}}$$

$$= \mathbb{P}_\theta(X_n = x_n \mid X_{n-1} = x_{n-1})$$

$$= \mathbb{P}_\theta(X_0 = x_0)\,\mathbb{P}_\theta(X_1 = x_1 \mid X_0 = x_0) \times \dots \times \mathbb{P}_\theta(X_n = x_n \mid X_{n-1} = x_{n-1})$$

$$= \underbrace{p^{x_0} \cdot (1-p)^{1-x_0}}_{\mathbb{P}_\theta(X_0 = x_0)} \cdot \; p_{0\to1}^{n_{0\to1}} \cdot p_{1\to0}^{n_{1\to0}} \cdot p_{0\to0}^{n_{0\to0}} \cdot p_{1\to1}^{n_{1\to1}}$$

where $n_{i\to j}$ is the number of ij I have seen in the sequence $x_0 \dots x_n$.

0 0 0 1 0 0 1 0 0 1 1

$n_{1\to1} = 1$

$n_{1\to0} = 2$ ..

So the likelihood of a sequence $X_0, \dots X_n$ with counts:

$N_{i\to j}$ is the number of time a transition $i \to j$ occurs in the sequence

is

$$L_\theta(X) = p^{X_0}(1-p)^{1-X_0} \; p_{0\to1}^{N_{0\to1}} (1-p_{0\to1})^{N_{0\to0}} \; p_{1\to0}^{N_{1\to0}} (1-p_{1\to0})^{N_{1\to1}}$$

the log likelihood is

$$\ell_\theta(X) = X_0 \log(p) + (1-X_0)\log(1-p) + N_{1\to0}\log(p_{1\to0}) + N_{1\to1}\log(1-p_{1\to0})$$
$$+ N_{0\to0}\log(1-p_{0\to1}) + N_{0\to1}\log(p_{0\to1})$$

to find the MLE of the transitions

$$\frac{\partial \ell_\theta(x)}{\partial p_{1\to0}} = \frac{N_{1\to0}}{p_{1\to0}} - \frac{N_{1\to1}}{1-p_{1\to0}}.$$

		$\frac{N_{1\to0}}{N_{1\to1}+N_{1\to0}}$	
$\frac{\partial \ell}{\partial p_{1\to0}}$	$+$	0	$-$
$\ell_{p_{1\to0}}$	$\nearrow$		$\searrow$

the derivative vanishes when $$\frac{N_{1\to0}}{p_{1\to0}} = \frac{N_{1\to1}}{1-p_{1\to0}}$$

that is $$N_{1\to0} = (N_{1\to1}+N_{1\to0})\,p_{1\to0}$$ that is $$p_{1\to0} = \frac{N_{1\to0}}{N_{1\to1}+N_{1\to0}}$$

Hence

$$\hat{P}_{1\to 0} = \frac{N_{1\to 0}}{N_{1\to 1} + N_{1\to 0}} = \frac{\text{nb of times you saw "}1\to 0\text{"}}{\text{———— saw a "1"}}$$

In the same way

$$\hat{P}_{0\to 1} = \frac{N_{0\to 1}}{N_{0\to 1} + N_{0\to 0}}$$

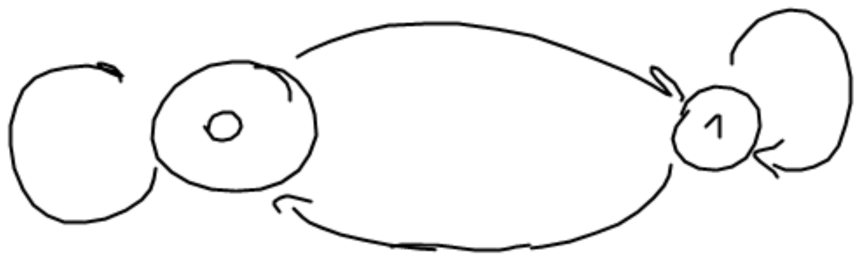
In general without further constraint the MLE of a probability for the transition $i\to j$

$$= \frac{N_{i\to j}}{\#\{X_k = i\}}$$

. This estimator is consistent if the chain is recurrent.

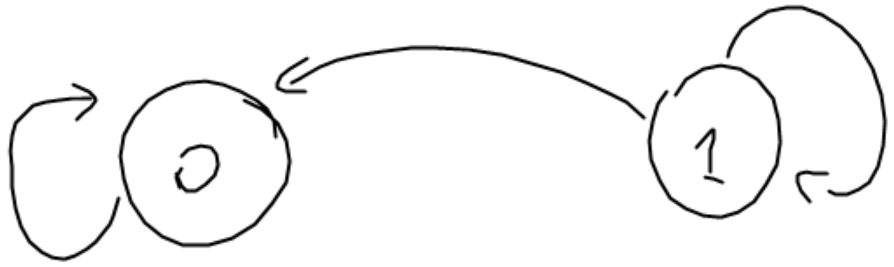
recurrence means that the chain will visit the transition an infinite number of time when $n \to +\infty$

for instance



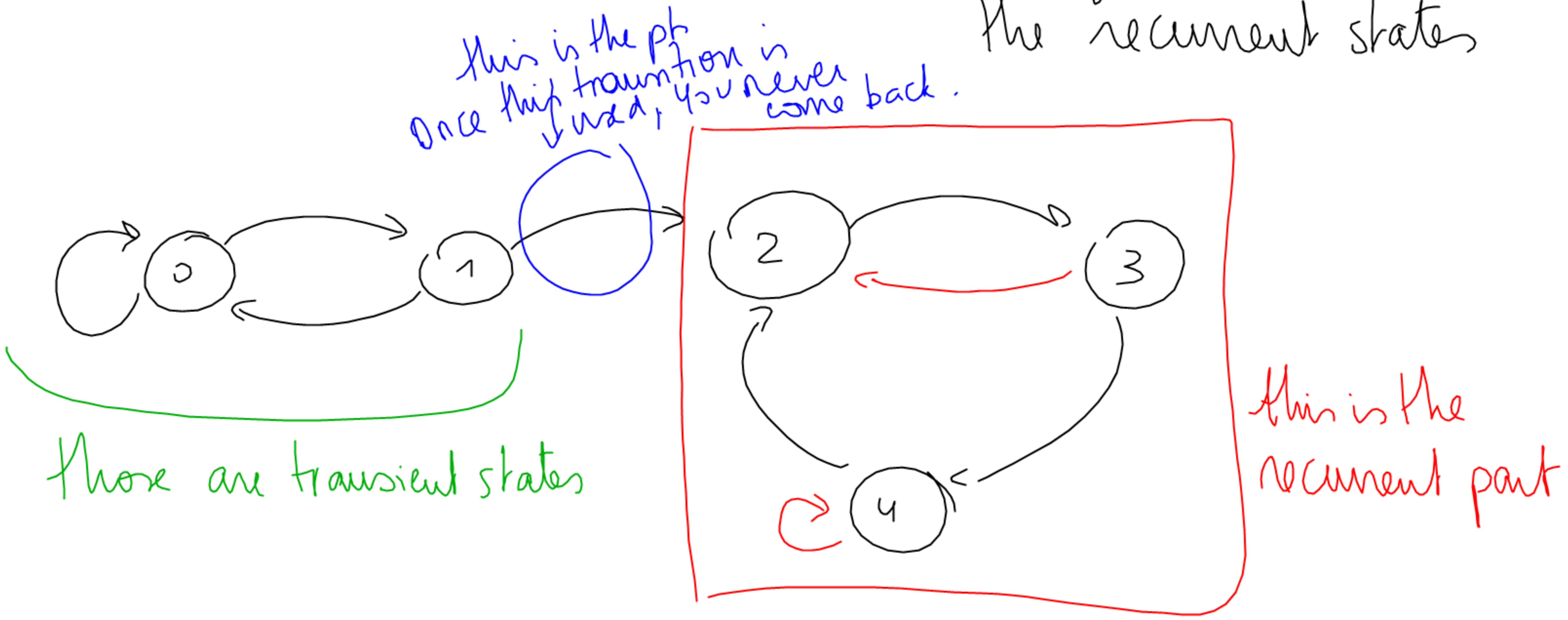
if all probabilities are non 0 this is recurrent.

whereas



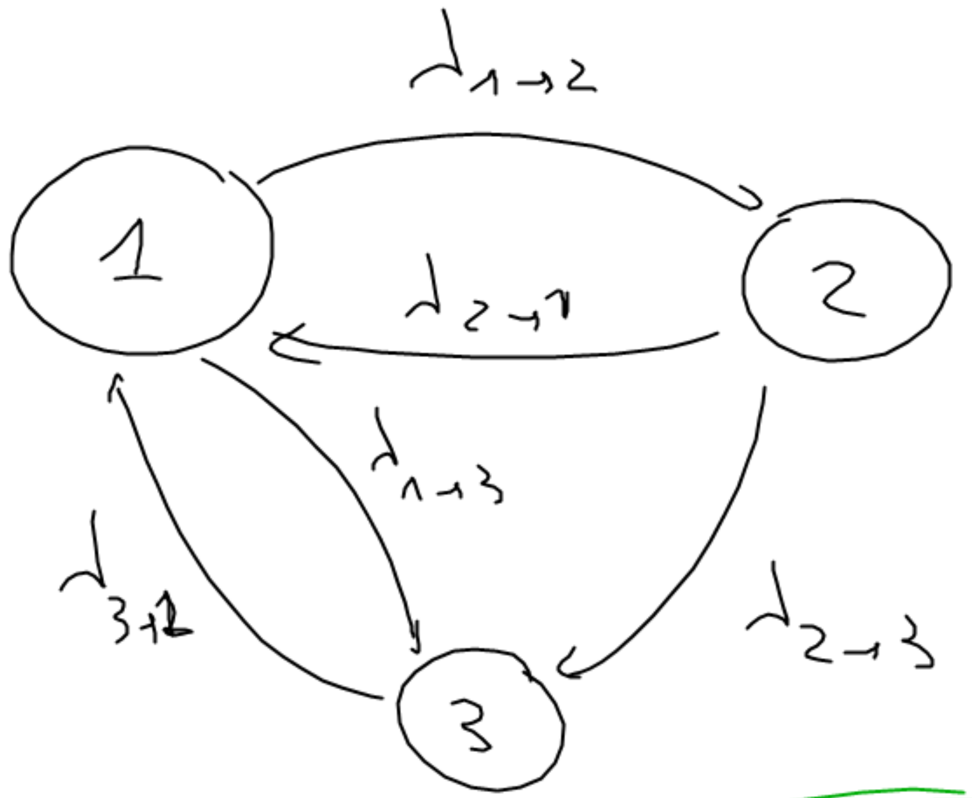
is not recurrent.

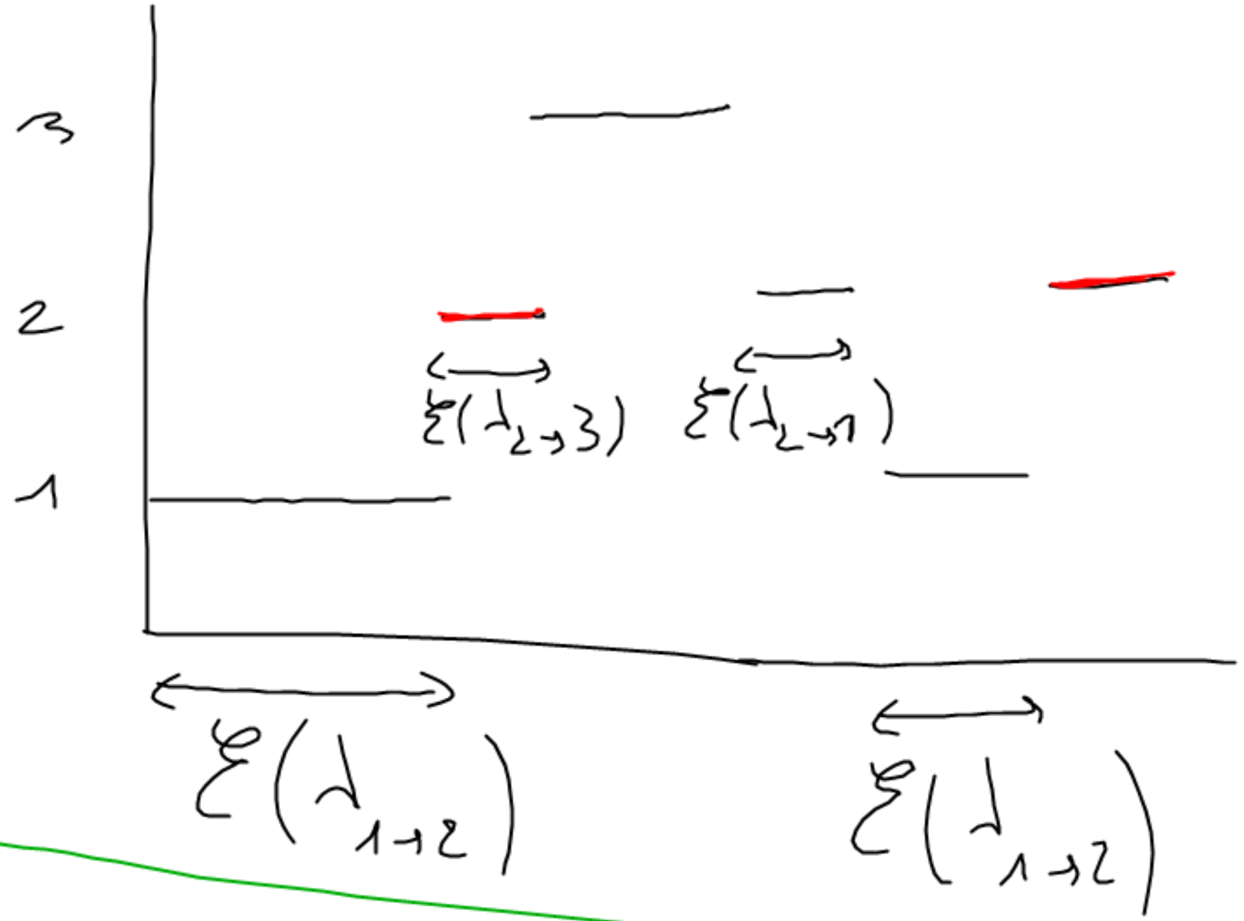
Usually when seeing a graph you can guess what are the recurrent states



Behind the fact that $\hat{p}_{i\to j} \xrightarrow[n\to+\infty]{} p_{i\to j}$ if i and j are recurrent, you have a thm called the ergodic theorem (a LLN for processes that are not ⫫)

About continuous Markov processes with discrete states





- The MLE is

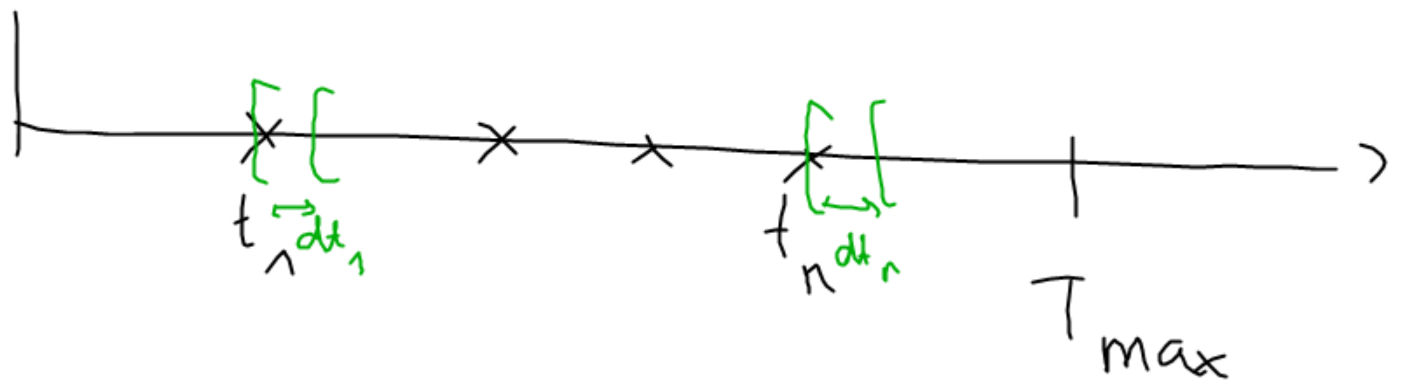
$$\hat{\lambda}_{i\to j} = \frac{1}{\text{mean duration of being in state } i \text{ and jumping to state } j}$$

It is consistent if the chain is recurrent

2/ Point processes

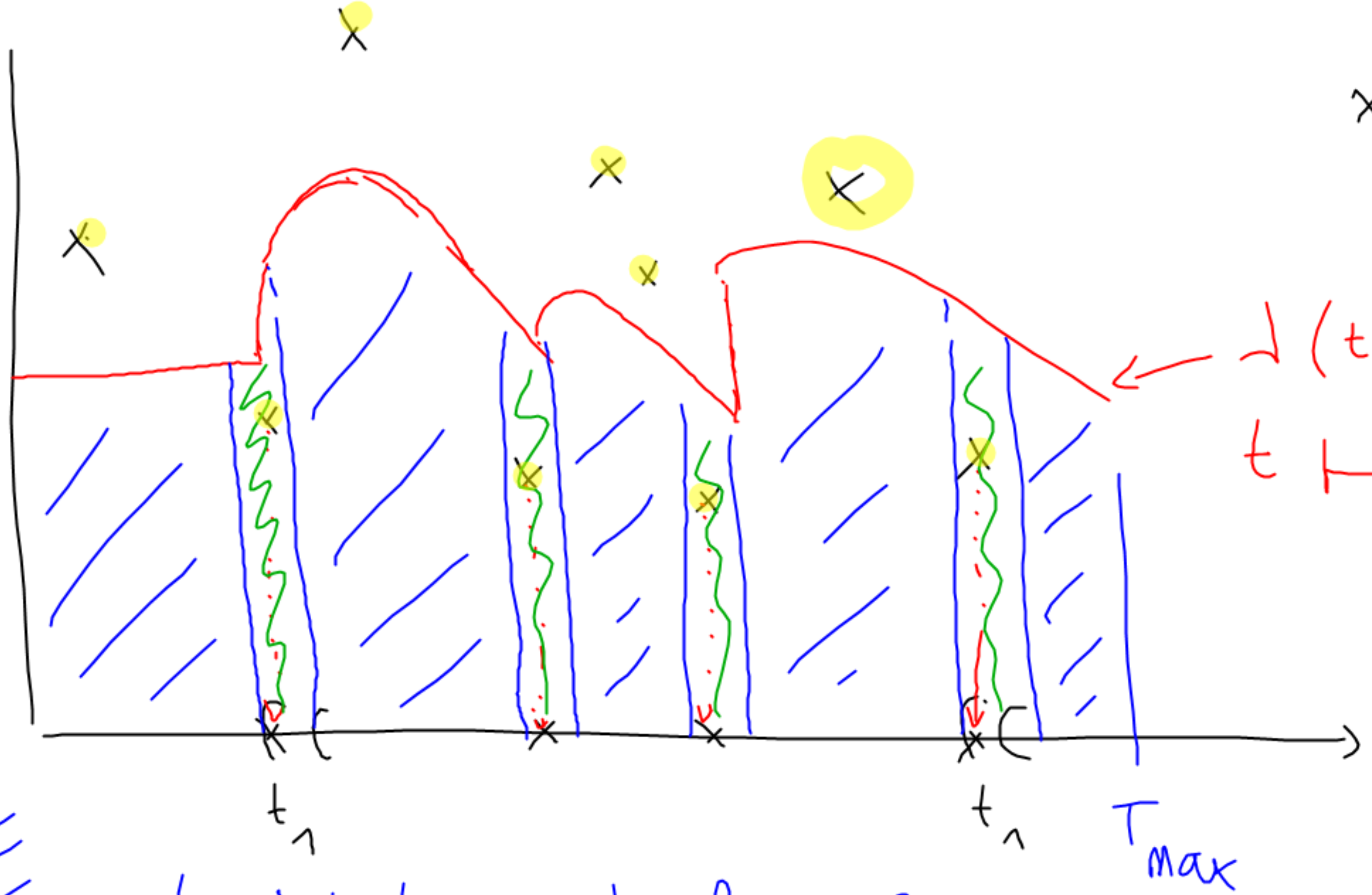
We observe a point process N with conditional intensity $\lambda(\,)$ that depends only on the previous point of N.

If we observe $N = \{t_1, \dots, t_n\}$ until T_{max}



$\mathbb{P}$(N has 1 point in $[t_1, t_1+dt_1[$, ..., 1 point in $[t_n, t_n+dt_n[$ and no points elsewhere)

Thanks to thinning you know that N can be generated as follows



$\times$: $\mathcal{N}$ Poisson process in $\mathbb{R}_+^2$.

$\lambda(t \mid \text{past})$

$t \longmapsto \lambda(t, \underbrace{t_1, \dots, t_n}_{})$

they are known → the red curve is fixed in this pb.

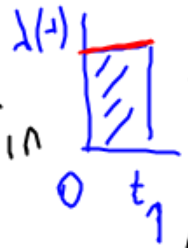
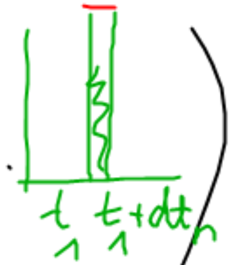
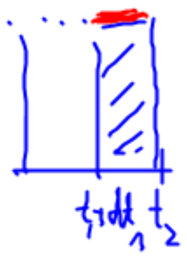
should be empty for $\mathcal{N}$

: I need exactly one point in each band for $\mathcal{N}$.

So the $\mathbb{P}($ N has one point in each $[t_i, dt_i[$ and no point elsewhere $)$

$= \mathbb{P}(\mathcal{N}$ has exactly one point in each $\}$ and no point in each $)$

Poisson process for λ

$= \mathbb{P}(\mathcal{N}$ has no point in  $)\ \mathbb{P}(\mathcal{N}$ has 1 point in  $)\ \mathbb{P}(\mathcal{N}$ no point in  $)$

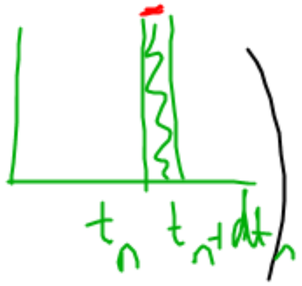
$\approx$ area $\approx \int_0^{t_1} \lambda(t)dt$

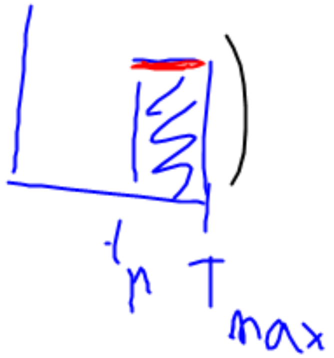
$\int_{t_1}^{t_1+dt_1} \lambda(t)dt =$ area $\approx \lambda(t_1)\, dt_1$ when $dt_1 \to 0$

$$= e^{-\int_0^{t_1} \lambda(t)dt} \cdot \lambda(t_1)\, dt_1 \cdot e^{-\int_{t_1}^{t_1+dt_1} \lambda(t)dt}\ e^{-\int_{t_1+dt_1}^{t_2} \lambda(t)dt} \cdots$$

$$\lambda(t_n)\, dt_n\ e^{-\int_{t_n}^{t_n+dt_n} \lambda(t)dt}$$

$$e^{-\int_{t_n+dt_n}^{T_{max}} \lambda(t)dt}$$

$\mathbb{P}(\mathcal{N}$ has 1 point in  $)$

$\mathbb{P}(\mathcal{N}$ — no point in  $)$

So the likelihood is.

$$\boxed{L_\lambda(N) = \prod_{\substack{T \in N \\ T \leq T_{max}}} \lambda(T) \; e^{-\int_0^{T_{max}} \lambda(t)\,dt}}$$

the log likelihood is

$$\ell_\lambda(N) = \sum_{\substack{T \in N \\ T \leq T_{max}}} \log(\lambda(T)) - \int_0^{T_{max}} \lambda(t)\,dt$$

$$\boxed{\ell_\lambda(N) = \int_0^{T_{max}} \log(\lambda(t))\,dN_t - \int_0^{T_{max}} \lambda(t)\,dt}$$

homogeneous Poisson process on $[0, T_{max}]$

We observe N a Poisson process with intensity (fixed) $\theta \in \mathbb{R}_+$

the conditional intensity is therefore $\lambda(t) = \theta$

$$\ell_\lambda(N) = \int_0^{T_{max}} \log(\lambda(t))\, dN_t - \int_0^{T_{max}} \lambda(t)\, dt$$

$$\ell_\theta(N) = \log(\theta)\, N_{[0,T_{max}]} - \theta\, T_{max}$$

nb of points between
0 and T_{max}

$$\frac{\partial l_\theta}{\partial \theta} = \frac{1}{\theta} N_{[0,T_{max}]} - T_{max}$$

$\longrightarrow$ (verify signs of the derivative)

$$\boxed{\hat{\theta} = \frac{N_{[0,T_{max}]}}{T_{max}}}$$

(classical estimator of the firing rate)

and it is consistant when $T_{max} \to +\infty$

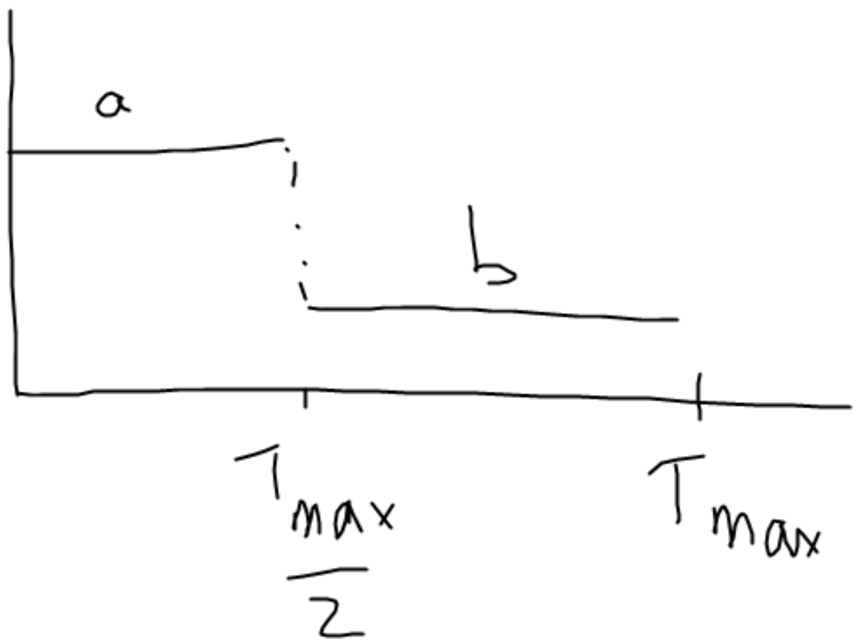
non homogeneous Poisson process

If we see $N_1, \dots, N_n$ n iid Poisson process with intensity $f(t)$

Then $N = N_1 \cup \dots \cup N_n$ is Poisson process with intensity $n f(t)$

The asymptotic here is when $n \to +\infty$ because you will have more and more points everywhere.

$f(t)$



$\theta = (a, b)$.

intensity of N is

$$n a \mathbb{1}_{t \leq \frac{T_{max}}{2}} + n b \mathbb{1}_{\frac{T_{max}}{2} < t \leq T_{max}}$$

(ex: firing rate of one neuron after stimulus $i = 1 \dots n$ are the trials)

$$\ell(N) = \int_0^{T_{max}} \log(\lambda(t))\, dN_t - \int_0^{T_{max}} \lambda(t)\, dt$$

$$\ell_{a,b}(N) = \log(na)\, N_{[0, \frac{T_{max}}{2}]} + \log(nb)\, N_{]\frac{T_{max}}{2}, T_{max}]} - na\frac{T_{max}}{2} - nb\frac{T_{max}}{2}$$

$$\frac{\partial}{\partial a} \rightsquigarrow \boxed{\begin{aligned} \hat{a} &= \frac{N_{[0, \frac{T_{max}}{2}]}}{n\frac{T_{max}}{2}} \\ \hat{b} &= \frac{N_{]\frac{T_{max}}{2}, T_{max}]}}{n\frac{T_{max}}{2}} \end{aligned}}$$

You can also select models in general

- by using AIC : $m \in \Omega$ with d_m degrees of freedom

  then $\hat{m} = \underset{m \in \Omega}{\arg\min} \; -\ell_m(N) + d_m$

- or if it's possible find out iid trials or subjects, cross validation.

<u>Remark</u> : $-\log$ likelihood is not the only contrast for point processes

⟶ least square exists, this is

$$-2\int_0^{T_{max}} \lambda(t)\, dN_t + \int_0^{T_{max}} \lambda(t)^2\, dt$$

$\Longrightarrow$ more usable typically for Hawkes processes

II) Time rescaling theorem and the goodness-of-fit tests

1) Time-rescaling theorem (also time change theorem Wattanabe theorem)

If N is a point process of conditional intensity $\lambda(t)$ with compensator $\Lambda(t) = \int_0^t \lambda(u)\,du$ on $[0, T_{max}]$

then the process $N' = \{\Lambda(T) \text{ for } T \in N\}$ is a Poisson process of rate $\underline{1}$ on $[0, \Lambda(T_{max})]$

for the ones interested,

$t \mapsto N_t = N_{[0,t]}$ counting process. increasing

$t \mapsto \Lambda(t) = \int_0^t \lambda(u)\,du$

$\Lambda(t)$ is the compensator because $M_t = N_t - \Lambda(t)$ is a martingale.

This means that $\mathbb{E}(M_t \mid \underbrace{\mathcal{F}_s}_{\text{past until } s.}) = M_s$ for $s<t$.

$\Rightarrow \mathbb{E}(M_t \mid \mathcal{F}_0) = \mathbb{E}(M_t) = M_0 = N_0 - \Lambda(0) = 0$

So M_t is centered

$\rightsquigarrow$ The compensator will follow N_t more closely than its expectation

$\rightsquigarrow$ "super centered" noise for the difference $N_t - \Lambda(t)$.

Moreover

$$\mathbb{E}\left(N_{[0,t]}\right) = \mathbb{E}(N_t) = \underline{\mathbb{E}(\Lambda(t))}$$

$$\parallel$$

$$\underline{\mathbb{E}\left(N'_{[0,\Lambda(t)]}\right)}$$

mean nb of points in $[0, \Lambda(t)]$ is roughly $\Lambda(t)$

$\longrightarrow$ rate $\underline{1}$.

This explains why N' is of rate $\underline{1}$.

2) Ogata's tests

I observe a process N on $[0, T_{max}]$ with intensity λ

I want to test $H_0 : \lambda(t) = \lambda_0(t)$ vs this is not the case

→ you compute the candidate compensator $\Lambda_0(t) = \int_0^t \lambda_0(t)\,dt$

→ you compute $N' = \{\Lambda_0(T), T \in N\}$.

Under H_0, N' should be a Poisson process with rate 1

$\Longrightarrow$ a) – the points are uniformly distributed on $[0, \Lambda_0(T_{max})]$

→ Kolmogorov Smirnov test of uniformity on the points of T_{max} of N'

b) The delays between points of N' are exponential with parameter 1 — in R

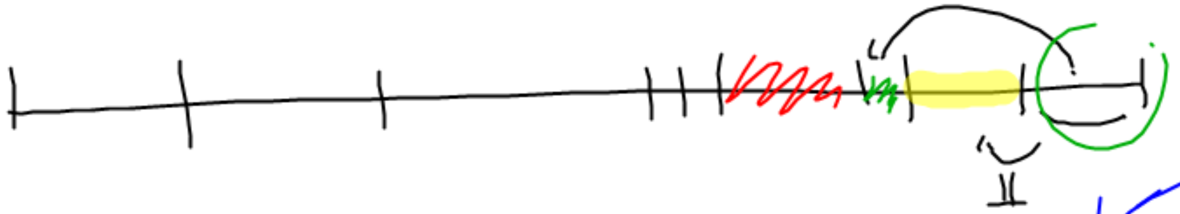
→ Kolmogorov Smirnov of exponentiality on the delays — ks.test

on N'

c) independance between the delays

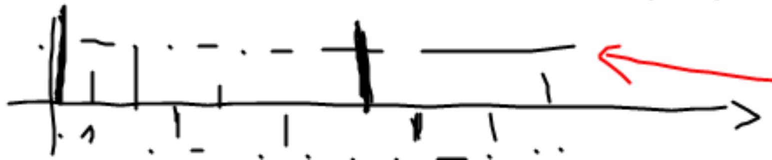
→ autocorrelation tests with different lags — on the delays of N' — $\text{acf}\left(\ , ci = 1 - \frac{0.05}{K} \right)$

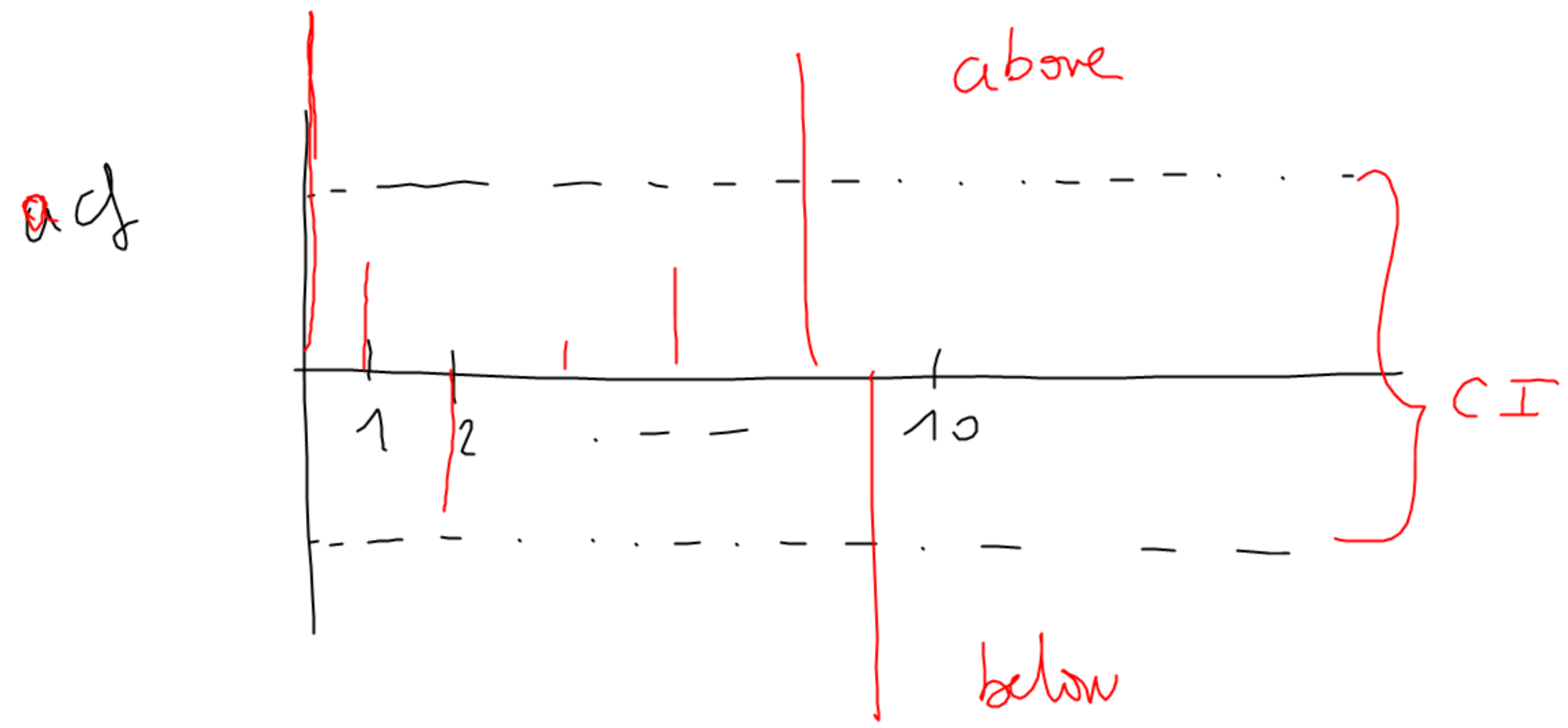
where K is the nb of tests



Be careful at the end you have $K = \overset{a,b}{2} + \overset{c}{L}$ tests (with L the nb of lags you tested)

→ Correct by Bonferroni the level / pvalues

We reject either if { pvalues in a or b are smaller than $0.05/K$.
or if  the CI — if one of the lag is above the CI or below


acf
above
1
2
10
CI
below

In practice

On simulation

- You have a new algorithm to simulate N with intensity λ.

To check that it works $\longrightarrow$ you compute $\Lambda(t) = \int_0^t \lambda(u)\,du$

$\longrightarrow$ $N' = \{ \Lambda(T), T \in N \}$.

$\longrightarrow$ Ogata's test on N'

and you want "large" pvalues (ie you don't reject)

On real data

You observe N.

you think the model λ_θ is good but you don't know θ and you want to verify it.

$\longrightarrow$ $\hat\theta$ (MLE or ...) please verify on simulation of the model that $\hat\theta$ is a good estimator for your model

$\longrightarrow$ $\hat\Lambda(t) = \int_0^t \lambda_{\hat\theta}(u)\,du$.

$\longrightarrow$ $\hat N' = \{\hat\Lambda(T),\ T \in N\}$.

$\longrightarrow$ Ogata's tests on $\hat N'$

Be extra careful this test is really conservative

$\longrightarrow$ if you reject : really there is a pb! with your model

$\longrightarrow$ if you accept, well you don't know much.