A Strory On Anomaly Localisation

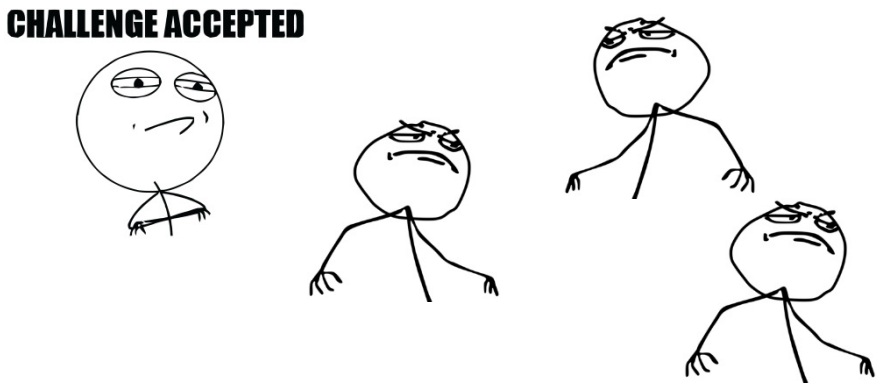
From Science to Data Science

romain.laby@criteo.com

A true story…

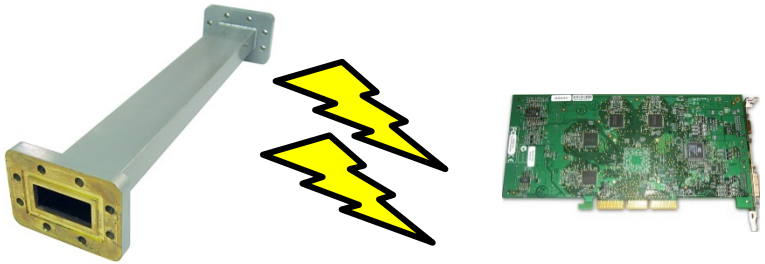
THALES

TELECOM ParisTech

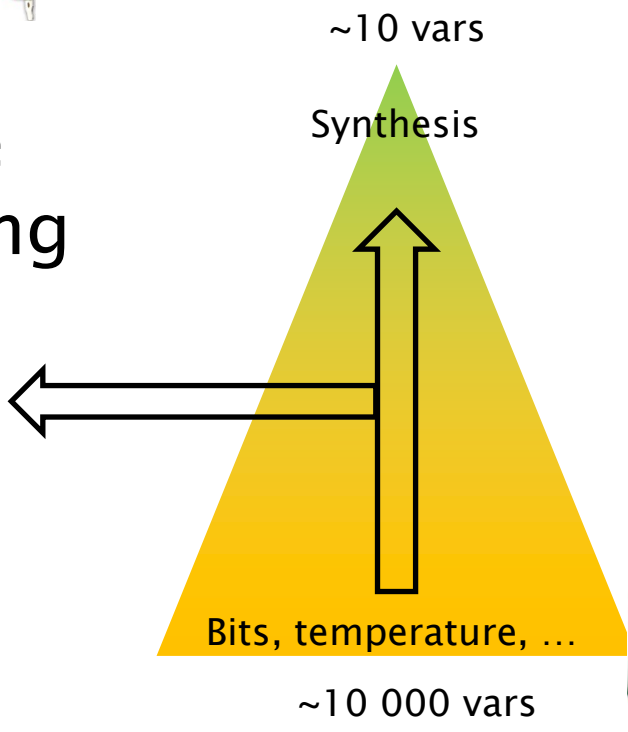


What was the problem ?

- A wave guide was exciting an electronic card

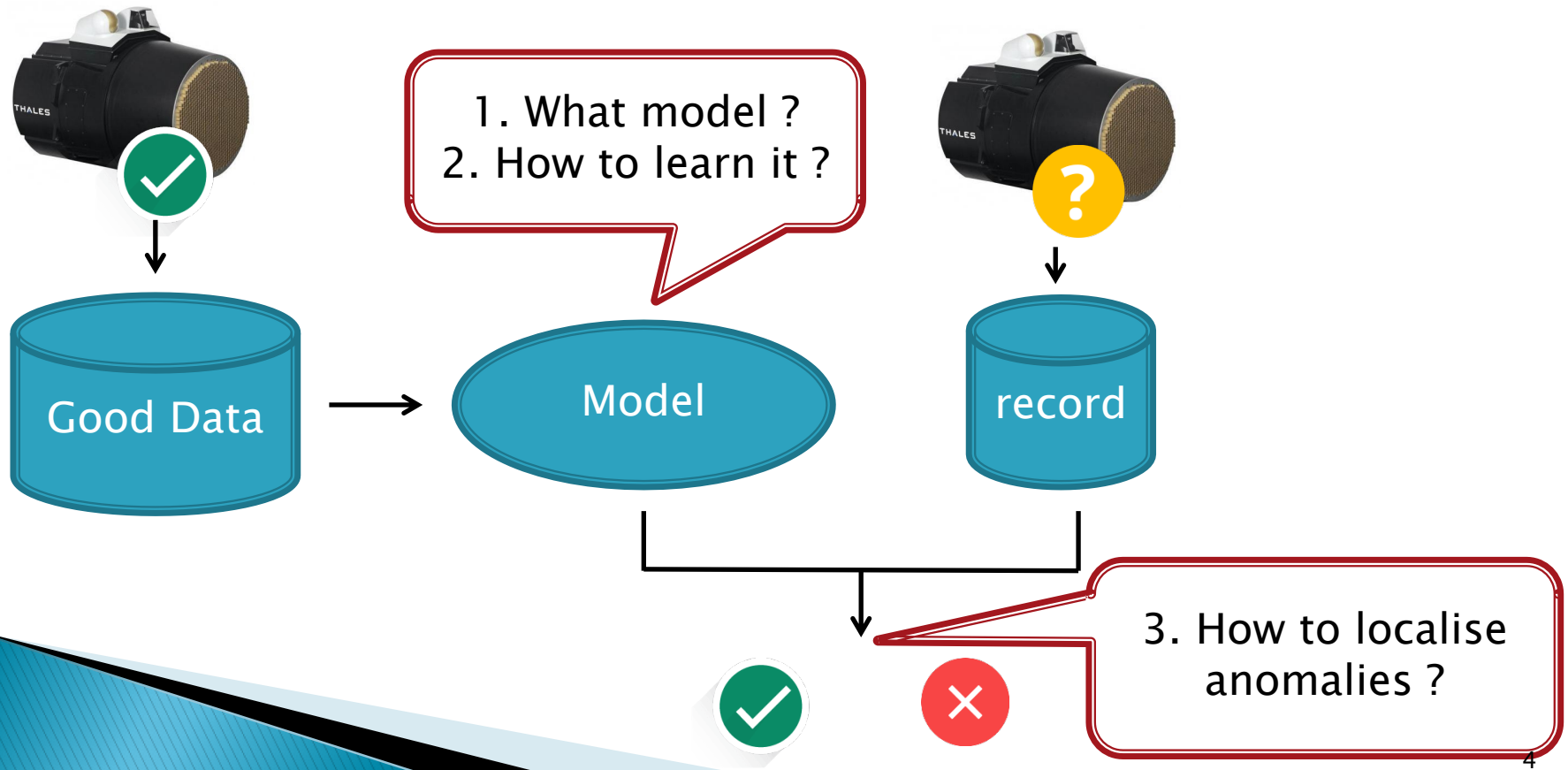


- Built-in tests: real-time monitoring and reporting



How to improve the built-in tests,

- It is impossible to list all breakdowns.
- ~ 10 000 variables !

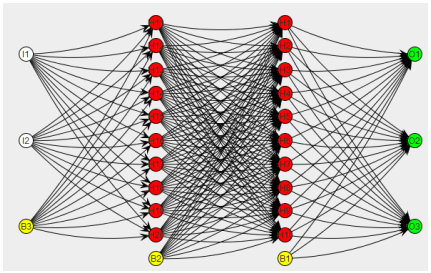


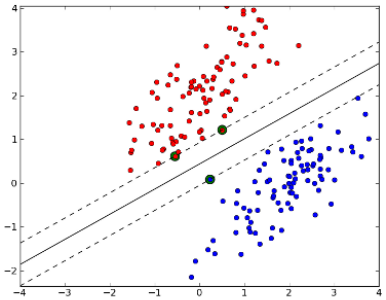
With some constraints…

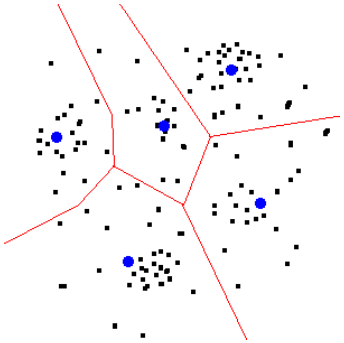
- Scalable method (>10 000 variables)
- Batch data file ~10 Go
- Online method
  - With no false alarms !
- Low computation power
  - ~ Gaming PC
- User friendly
- Heterogeneous data
  - Physical measures
  - Categorical variables

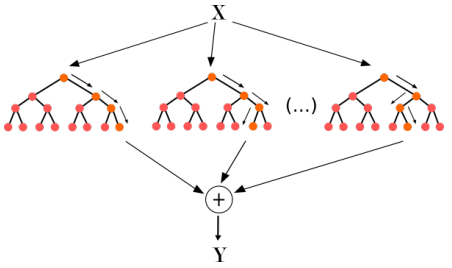
What ML tool for localisation ?

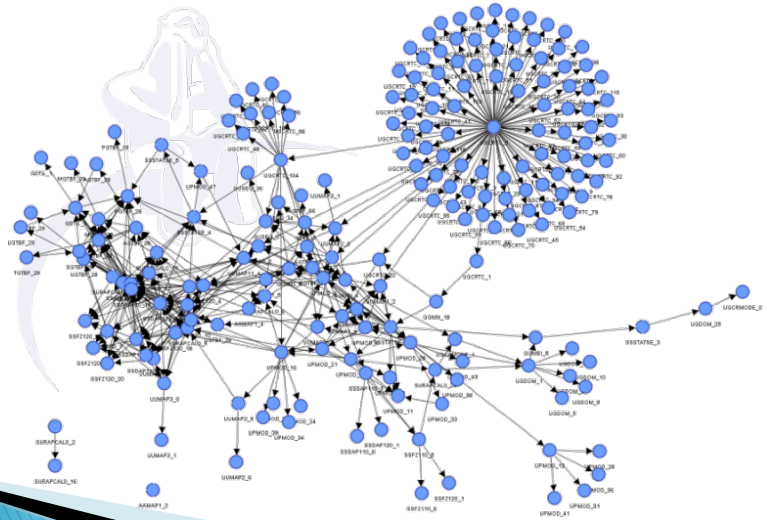
- Plenty of approaches to detect anomalies











- Most of them are *Blackbox…*

What is a graphical model ?

- Machine to compute probabilities
  - $P(Temp = 35°C \mid State = Finland) = ??$
- The density might be intractable
  - E.g. : n binary variables $\leftrightarrow 2^n - 1$ parameters
  - $2^{35} - 1 \approx 32.10^9$ parameters
- Association between a density and a graph.

$P(X_1, \dots, X_n)$ $\longleftrightarrow$ 

What is a graphical model ?

- Idea : simplify the density into a product of factors using independances in P

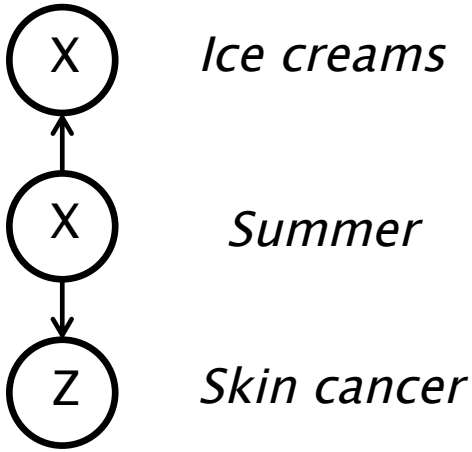
$$P(X,Y,Z) = P(X)P(Y|X)P(Z|X)$$

X	Y	Z	$P(X,Y,Z)$
0	0	0	0,14
0	0	1	0,06
0	1	0	0,02
0	1	1	0,18
1	0	0	0,084
1	0	1	0,036
1	1	0	0,048
1	1	1	0,432

X	$P(X)$
0	0,4
1	0,6

Y	X	$P(Y\|X)$
0	0	0,5
0	1	0,2
1	0	0,5
1	1	0,8

Z	X	$P(Z\|X)$
0	0	0,7
0	1	0,1
1	0	0,3
1	1	0,9



Pairwise Markov Models

- Pairwise Exponential Family:
  - Given a graph (V, E),

$$P_\theta(X) \propto \exp\left(\sum_{i \in V} \theta_i B_i(X_i) + \sum_{(i,j) \in E} \theta_{ij} B_i(X_i) B_j(X_j) + \sum_i C_i(X_i)\right)$$

- $\{B_i(\cdot)\}_i$ and $\{C_i(\cdot)\}_i$ depends on the choice of the family

Some Exponential families

- (zero-mean) Gaussian distribution:

$$B_i(X_i) = \frac{X_i}{\sigma_i}, \qquad C_i(X_i) = \frac{-X_i^2}{2\sigma_i^2}$$

- Poisson distribution:

$$B_i(X_i) = X_i, \qquad C_i(X_i) = -\log(X_i!)$$

- Ising/Potts distribution:

$$B_i(X_i) = X_i, \qquad C_i(X_i) = 0$$

- We can mix them too !

Mixed exponential families

- Gaussian + Ising
  - $\boldsymbol{X} \in \mathbb{R}^p,\ \boldsymbol{Y} \in \{0,1\}^q\ :\ p(\boldsymbol{X},\boldsymbol{Y})$ is valid!

- Poisson + Ising
  - $\boldsymbol{X} \in \mathbb{N}^p,\ \boldsymbol{Y} \in \{0,1\}^q\ :\ p(\boldsymbol{X},\boldsymbol{Y})$ is valid!
- Gaussian + Poisson
  - $\boldsymbol{X} \in \mathbb{R}^p,\ \boldsymbol{Y} \in \mathbb{N}^q\ :\ p(\boldsymbol{X},\boldsymbol{Y})$ is invalid…

Learning a model

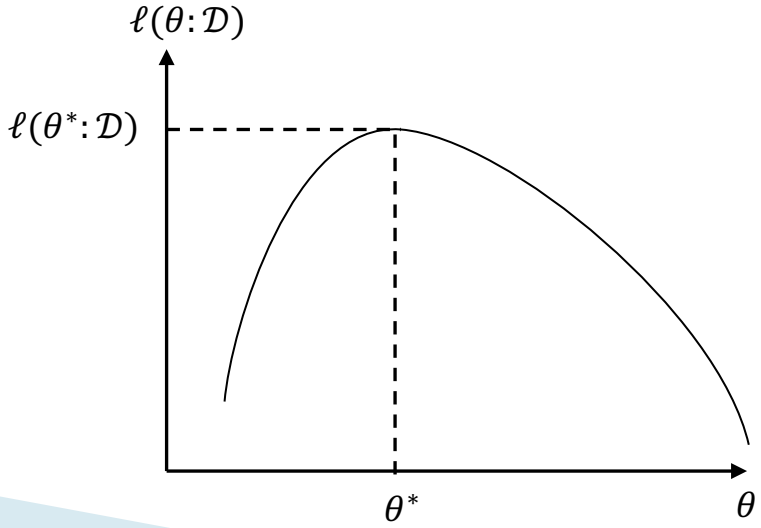
- Training set $\mathcal{D} = \left(X^{(t)}\right)_{t\in\mathcal{T}}$.
- We denotes $\theta = (\theta_i, i \in V) \cup (\theta_{ij}, ij \in E)$

$$P_\theta(X) = \frac{1}{Z_\theta}\exp\left(\sum_{i\in V}\theta_i B_i(X_i) + \sum_{(i,j)\in E}\theta_{ij}B_i(X_i)B_j(X_j) + \sum_i C_i(X_i)\right)$$

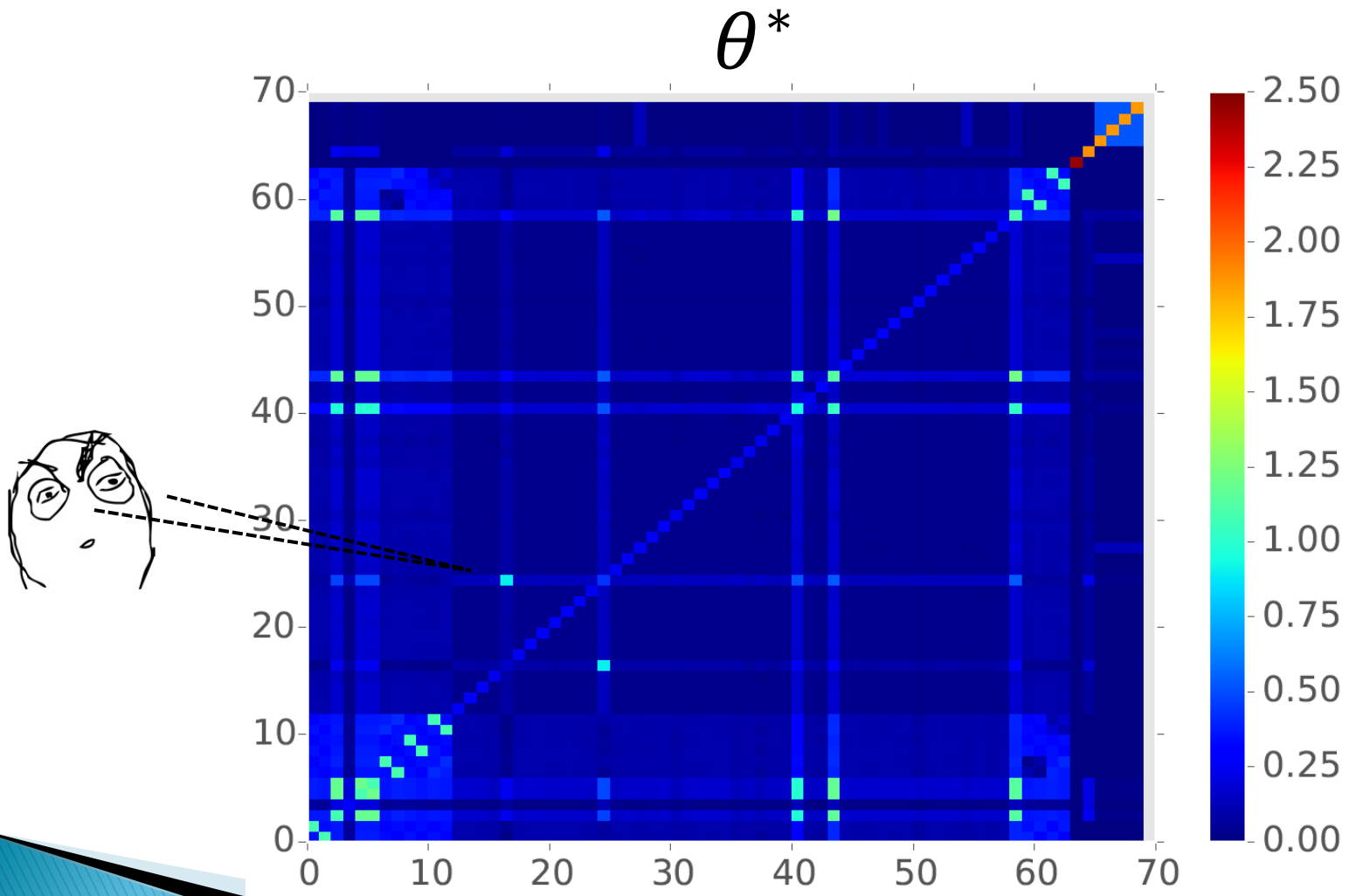
- Optimise the likelihood function

$$\ell(\theta|\mathcal{D}) = \sum_{t\in\mathcal{T}}\log P_\theta\left(X^{(t)}\right)$$

$$\theta^* = \underset{\theta\in\Theta}{\operatorname{argmax}}\,\ell(\theta|\mathcal{D})$$

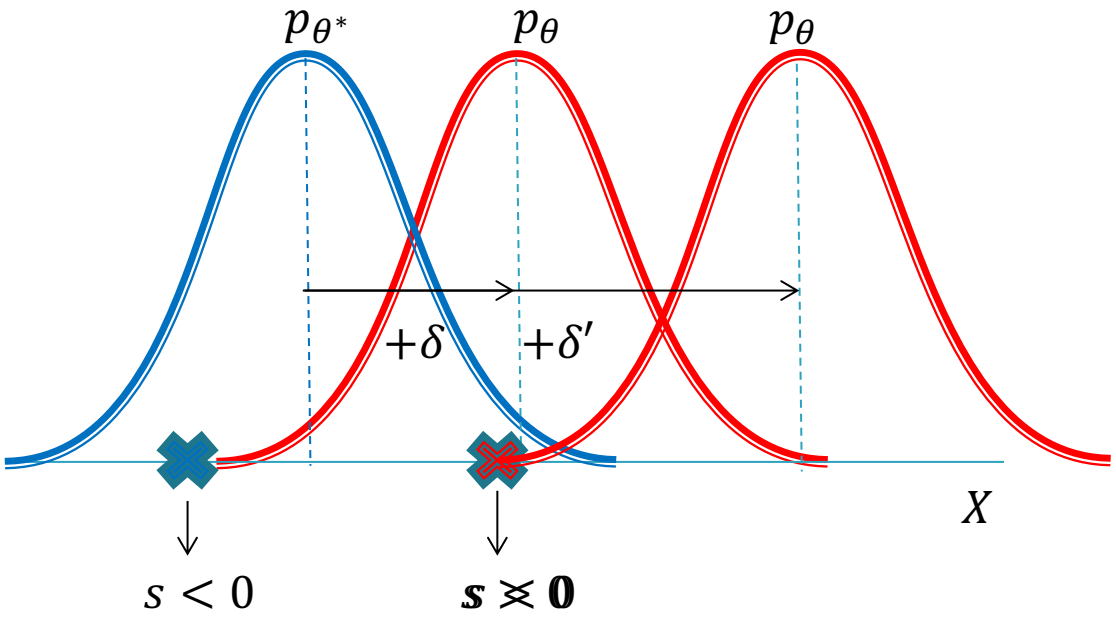


Learning a model



How to detect an anomaly ?

- We learned $p_{\theta^*} = \mathcal{N}(\mu, \sigma^2)$
- We define an anomalous density $p = \mathcal{N}(\mu + \delta, \sigma^2)$



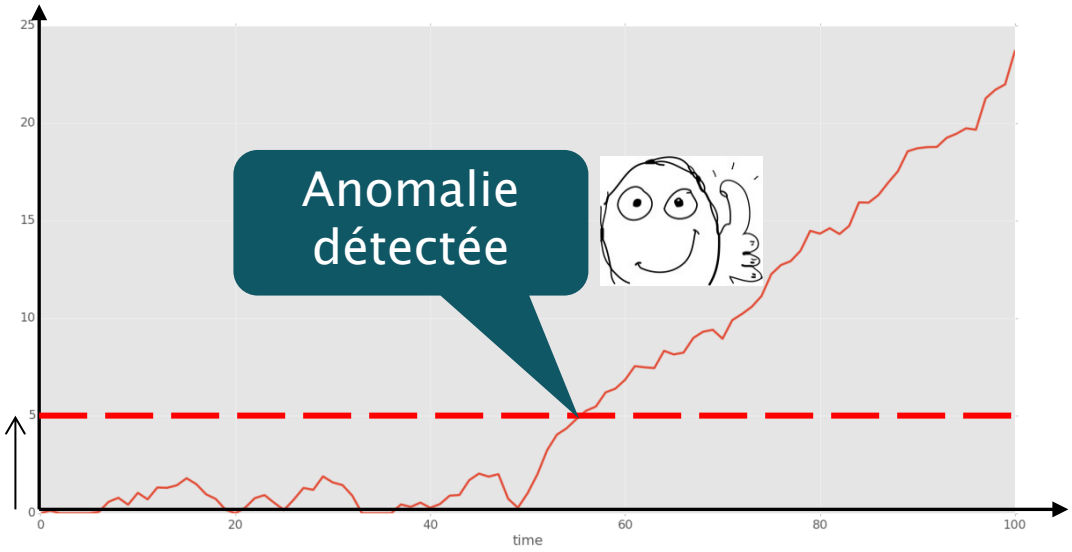
$$s(X) = \log \frac{p(X)}{p_{\theta^*}(X)}$$

δ controls the sensibility of the detection.

Anomaly localisation

- Do the same for every variable !

$s_i^{(t)}$



$$s_i(t) = \log \frac{p_{\theta_i}(X_i|X_{-i})}{p_{\theta^*}(X_i|X_{-i})}$$

$$S_i(0) = 0$$
$$S_i(t) = (S_i(t) + s_i(t))^+$$

$$p_\theta(X_i|X_{-i}) \propto \exp\left(B_i(X_i)\left(\theta_i + \sum_j \theta_{ij} B_j(X_j)\right) + C_i(X_i)\right)$$

What are we doing …

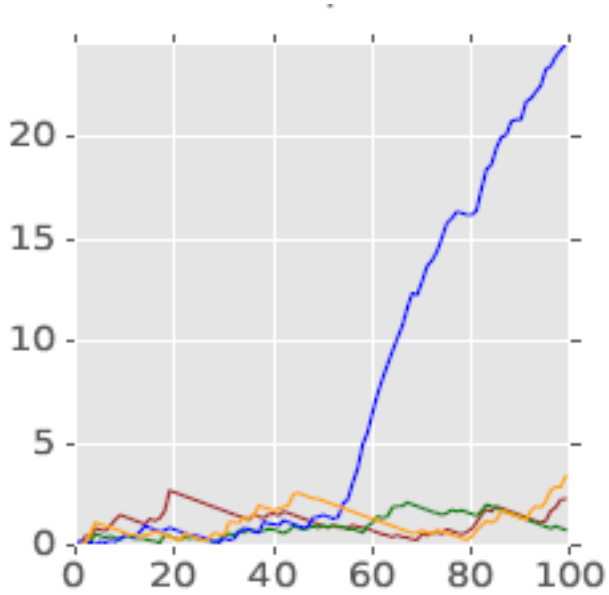
- We find the change time t_0:
  - $X^{(t)} \sim P_{\theta^*}$ for $t < t_0$
  - $X^{(t)} \sim P_{\theta}$ for $t > t_0$
- We have located a change in the parameters θ^*
- Localisation = finding a change in the parameters of the implicit distribution

Network traffic control

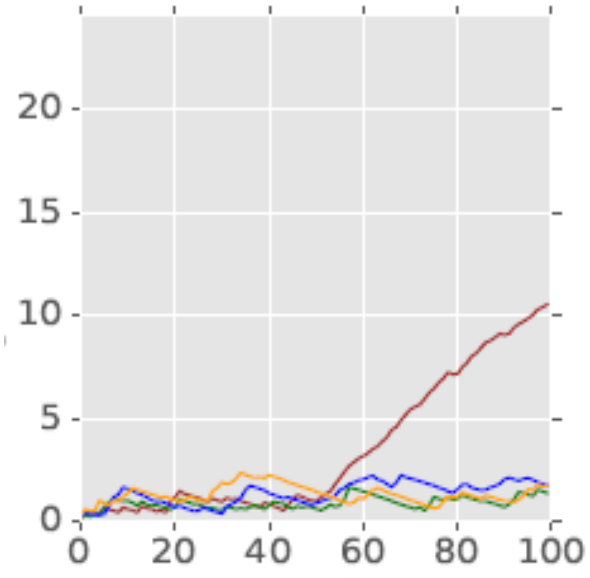
- N devices
- Each device i sends $X_i^{(t)}$ frames through WIFI during t and $t+1$
- Hypothesis : $X_i | X_{-i} \sim Poisson(\lambda_i)$
- Anomalous density : $Poisson(\lambda_i + 50)$

Network traffic control

- At $t = 50$, I started watching a Youtube video



$S_{youtube}(t)$



$S_{my\ phone}(t)$

Back to the company world…

- 40% : understand the data
- 30% : communicate
  - Proof of concept, updates, plannings, …
- 20% : research
  - The simplest works often fine !
- Remaining 10% : useless meetings

Thanks for your attention !