An inexistent series of railway crashes

by Gilles Dupin
RATP, engineer, risk assesment

PSAM9, may 2008

A derailment in Paris….

Black summer of 2006

On 3 July 2006, in Valencia (Spain)	a metro trainset derails killing 41 people
On 21 August, in Villada (Spain)	an intercity train derails killing 6 people
On 22 September, in Lathen (Germany)	a maglev train collides with a maintenance vehicle killing 21 people
On 11 October, in Zoufftgen (France)	an intercity train collides with a freight train killing 5 people
On 17 October, in Roma (Italy)	a metro trainset collides with a stationnary train killing 1 person

Was this succession of accidents the result of chance …

or should it be explained by other phenomena ?

- Was it due to an increase in rail traffic and more intensive operation of the networks ?
- Was it due to a relaxation of monitoring and maintenance ?
- Are drivers paying less attention to regulations ?
- Are the systems ageing and obsolete ?

etc

In short, are we observing an increase in accidents which reveals a reduction in the level of safety of the systems ?

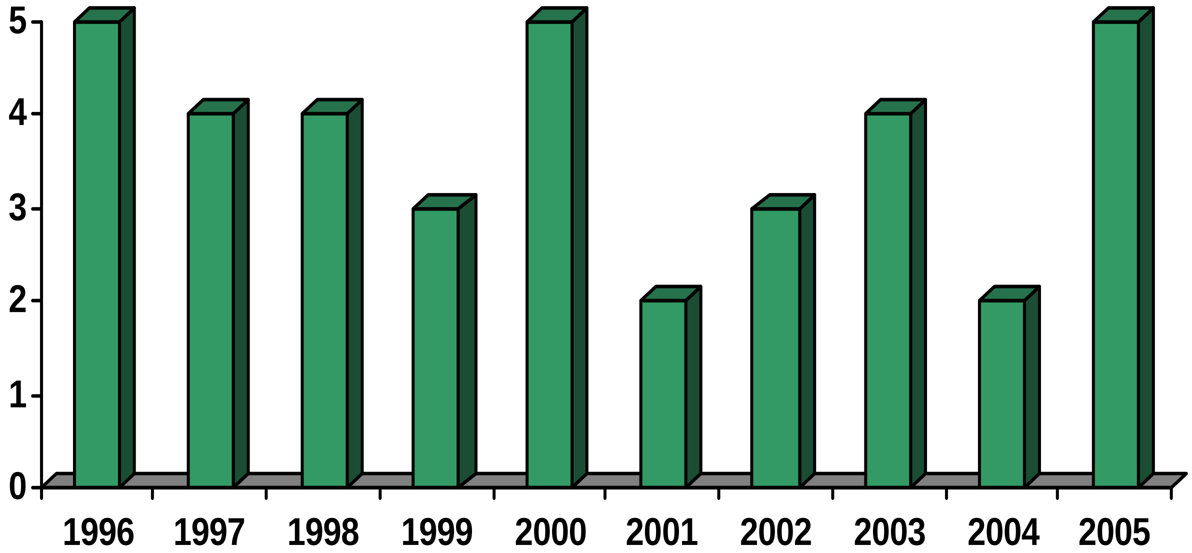
Statistical data

Between **1996 and 2005**, we have listed the railway accidents that received the most media attention,

That is to say **fatal accidents with a safety integrity level of 4, involving a collision or a derailment**.

Only the **most industrialized countries** are selected :
Western Europe, North America, Australia, Japan.

Statistical data



Number of annual railway accidents in the population concerned

Empirical mean : 3.7 accidents per year

Observed series : 5 accidents in 4 months = **density 4 times higher**

Probabilistic model

- The number of accidents that occur in a period **depends only on its duration**. The temporal density of accidents is therefore constant.

This hypothesis is equivalent that the appearance of the series can be explained by statistical variations.

- Accidents are **mutually independent**.

- Two accidents **cannot occur at the same time**.

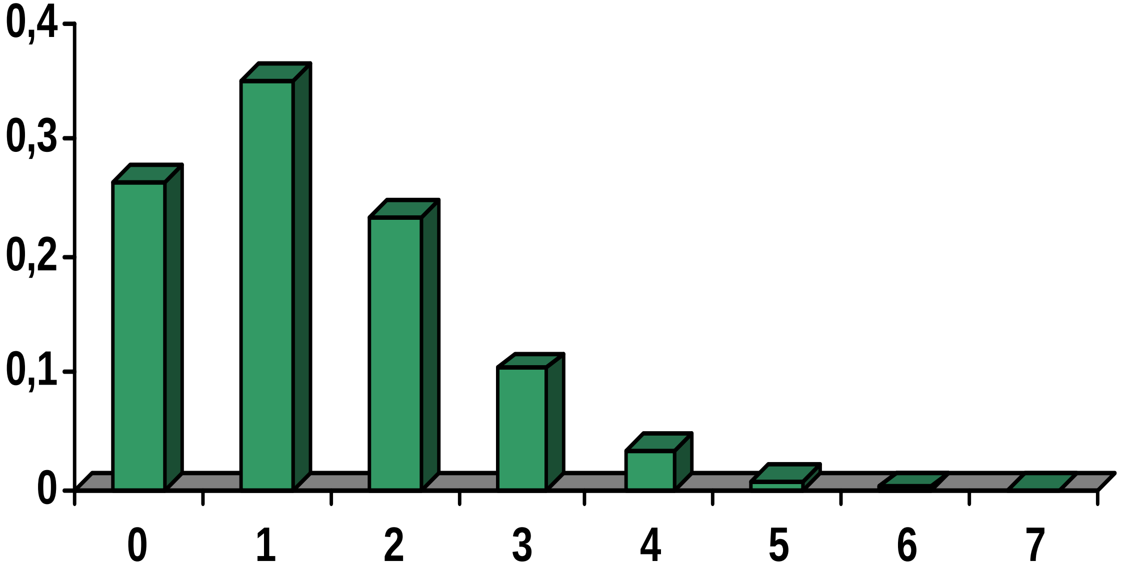
$X_{\Delta t}$: Number of accidents occurring during a period of duration Δt

$\mathbf{X_{\Delta t}} \rightarrow \boldsymbol{P}\ \mathbf{(0{,}3083.\Delta t\ )}$ with Δt in months

Simplistic calculation n°1
using one window

$X\Delta t \rightarrow$ *P* (0.3083.Δt)

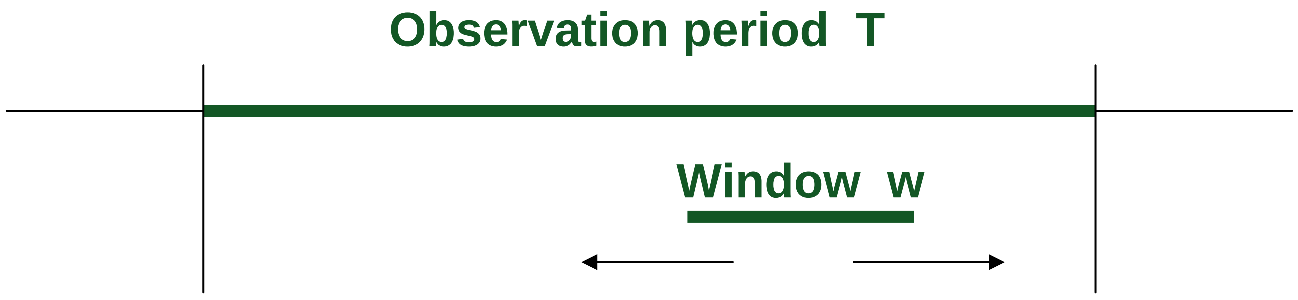
$X_4 \rightarrow$ *P* (1.233) for the period Δt = 4 months



distribution of Poisson with parameter ϕ = 1.233

P(series) $\approx$ P($X_4 \geqslant$ 5) $\approx$ **0.0086** **(1)**

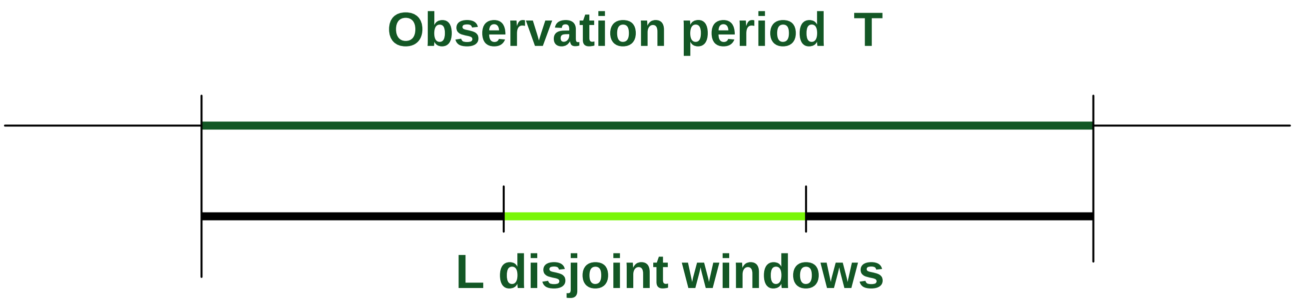
Sliding window



- **Window :** **w = 4 months**
- **Non-specialist observer :** **T = 1 year**
- **Professionnal observer :** **T = 10 years**

Brief calculation n°2

using disjoint windows



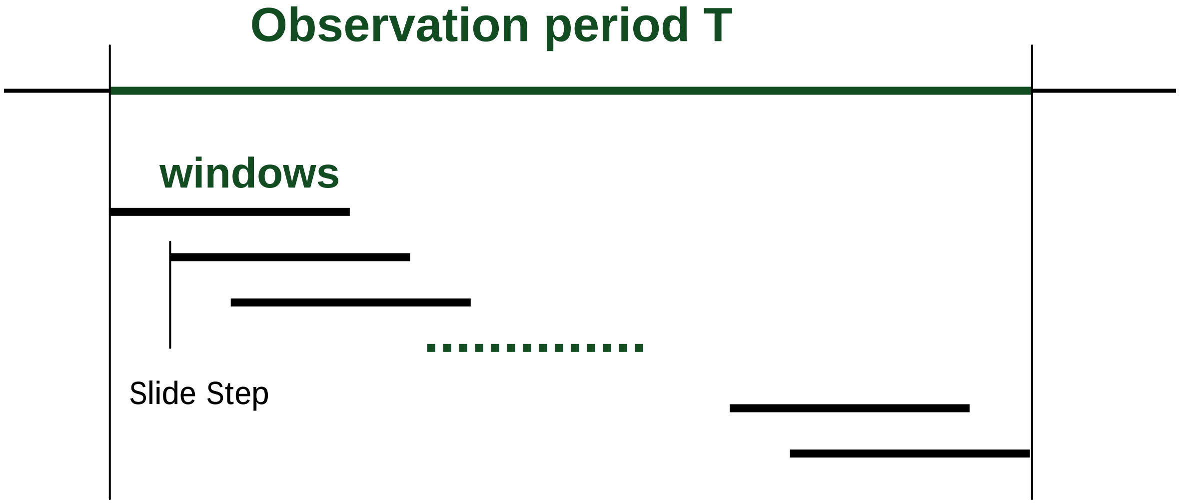
F_i : « the window i contains at least 5 accidents »

$$\mathbf{P\ (series)} = P\left(\bigcup_{i=1}^{L} F_i\right) = 1 - P\left(\bigcap_{i=1}^{L} \overline{F_i}\right) = 1 - P(\overline{F_i})^{L}$$

T = 1 year P (series) = $1-(1-0{,}0086)^{3}$ = **0.026 (2)**

T = 10 years P (series) = $1-(1-0{,}0086)^{30}$ = **0.228 (3)**

More precise calculation n°3
using Poincaré formula



For T = 1 year and a slide step of 1 month : **9 windows**

Certain windows overlap : **dependent events**

Poincaré's formula

Poincaré's formula :

$$\mathbf{P}\left(\bigcup_{i=1}^{9} F_i\right) = \sum_{p=1}^{9}\left[(-1)^{p+1} \cdot \sum_{1\le i1<i2<\ldots<ip\le 9} \mathbf{P}\left(F_{i1}\cap F_{i2}\cap\ldots\cap F_{ip}\right)\right]$$

When the events are **disjoint**, they are **independent** :

$$\mathbf{P}_{i\cap j=\varnothing}(E_i \cap E_j) = \mathbf{P}(E_i).\mathbf{P}(E_j)$$

$$X_{\Delta t} \rightarrow P\,(0.3083.\Delta t)$$

Numerical application for the Poincaré's formula

$$P\left(\bigcup_{i=1}^{9} F_i\right) = \sum_{p=1}^{9}\left[(-1)^{p+1} . \sum_{1 \le i1 < i2 < K < ip \le 9} P\left(F_{i1} \cap F_{i2} \cap K \cap F_{ip}\right)\right]$$

- **p = 1**

$$P(F_{i1}) = P(F_i) = P(X_4 \ge 5) = 0.0086$$

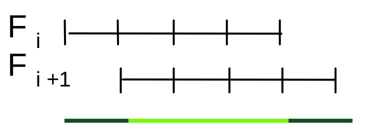
$$\sum_{1 \le i \le 9} P(F_i) \approx 9.0.0086 \approx 0.0774$$

$$P(F_{i1} \cap F_{i2})$$

- **p = 2**

$$P(F_i \cap F_{i+1})$$

$$1 \le i \le 8$$



(; ≥5 ;)

(≥1 ; 4 ; ≥1)

(≥2 ; 3 ; ≥2)

etc

$$P(X_3 \ge 5) = 0.0026$$

$$P(X_1 \ge 1)^2 . P(X_3 = 4) = 0.2653^2 . 0.0121 = 0.0008$$

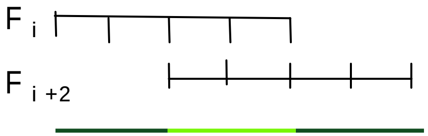
$$P(X_1 \ge 2)^2 . P(X_3 = 3) = 0.0388^2 . 0.0523$$

negligible

$$P(F_i \cap F_{i+1}) \approx 0.0035$$

Numerical application for the Poincaré’s formula

- **p = 2** $\quad \mathbf{P(F_i \cap F_{i+2})} \qquad 1 \le i \le 7$



(; ≥5 ;)

(≥1 ; 4 ; ≥1)

(≥2 ; 3 ; ≥2)

etc

$$\mathbf{P(X_2 \ge 5) = 0.0004}$$

$$\mathbf{P(X_2 \ge 1)^2.P(X_2 = 4) = 0.4602^2.0.0032 = 0.0007}$$

$$\mathbf{P(X_2 \ge 2)^2.P(X_2 = 3) = 0.1274^2.0.0211 = 0.0003}$$

negligible

$$\mathbf{P(F_i \cap F_{i+2}) \approx 0.0015}$$

$$\mathbf{P(F_i \cap F_{i+3})} \qquad 1 \le i \le 6$$

etc

$$\mathbf{P(F_i \cap F_{i+3}) \approx 0.0004}$$

Numerical application for the Poincaré's formula

- **p = 2** $\quad \mathbf{P(F_i \cap F_{i+p})} \qquad p \geq 4 \qquad 5 \leq i + p \leq 9$

$$\mathbf{P(F_i \cap F_{i+p}) = P(F_i).P(F_{i+p}) = P(F_1)^2 = 0.00007}$$

$$\sum_{\mathbf{1 \leq i1 < i2 \leq 9}} \mathbf{P(F_{i1} \cap F_{i2}) \approx 8.0.0035 + 7.0.0015 + 6.0.0004 + 15.0.00007 \approx 0.0420}$$

- **p = 3**

etc

- **p ≥ 4** $\quad \mathbf{P(F_{i1} \cap F_{i2} \cap ... \cap F_{ip})}$ negligible

Numerical application for the Poincaré's formula

therefore $\boxed{P(\bigcup_{i=1}^{9} F_i) \approx 0.043}$ **(4)**

This method has limits because the slide step should be shorter :

large number of windows

extremely large number of partitions

This calculus cannot give an accurate calculation

Scan Statistics

- Detection and analysis of a **succession of close events**, in time or space, to know if it occurs by coincidence or if it is an abnormal series

- American researchers since 1960's

- **Applications** :

  unusual clusters in cancer

  clusters of palindromes in DNA as a clue for virus replication

  capacity of phone centers

  clusters of defective objects within a production line

Powerful computation methods

- **by integration** (exact results)

- **by framing** (approximated results)

  Bonferroni bounds

- **by combinatorial analysis** (exact results)

  random walks and reflection principle

  ballot problems

  Karlin-MacGregor theorem (1959)

  Barton et Mallows corollary (1965)

  Kolmogorov-Smirnov statistic

- **by approximation** (approximated results)

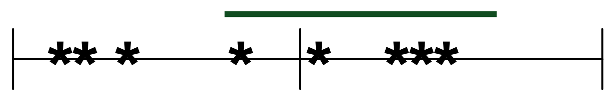
  product-type approximation

  Poisson approximation

Accurate calculation n°4 using a Scan Statistics formula

Naus' formula (1982)

$$P \approx 1 - Q_2 \left(\frac{Q_3}{Q_2}\right)^{L-2}$$



Wallenstein, Naus et Glaz approximation (1993)

$$Q_2 \approx 2\,F(k-1,\psi) - 1 - (k-1-\psi)\;p(k,\psi)$$

$$Q_3 \approx 2\,F(k-1,\psi) - 1 - (2k-1-2\psi)p(k,\psi)$$

$$L = \frac{T}{w} \qquad \phi = \lambda\, w \qquad F(k,\psi) = \sum_{i=0}^{k} p(i,\psi) \qquad p(i,\psi) = \exp(-\psi).\frac{\psi i}{i\,!}$$

In our example : $k = 5$ $\phi = 1{,}233$

T = 1 year **$P \approx$ 0.062** **(5)**

T = 10 years **$P \approx$ 0.552** **(6)**

Synthesis of calculations

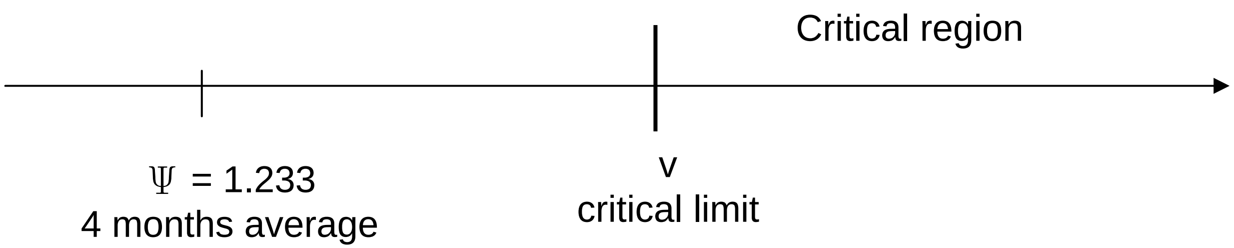
		T = w 4 months	T = 3w 1 year	T = 30w 10 years
Simple Approach	Disjoints windows	0.009 (1)	0.026 (2)	0.228 (3)
More precise Approach	Step of discretization of one month		0.043 (4)	
Accurate computation	Step of discretization of one day		0.062 (5)	0.553 (6)

Probability of appearance of the series with regard to the approach

Statistical test

$$\begin{cases} H_0 = \text{The temporal density of accidents is constant} \\ H_1 = \overline{H_0} \end{cases}$$

Statistic for the test : **Sw**



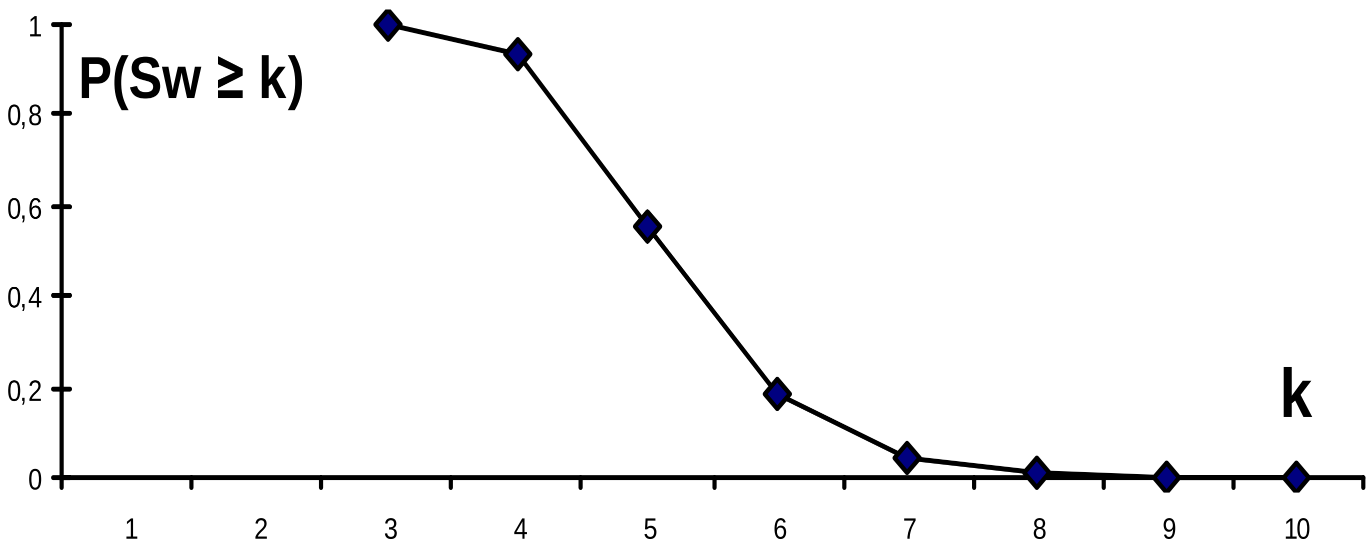
P(Sw ⩾ k) calculated with the Naus' formula

T = 10 years

Statistical test

k	0	1	2	3	4	5	6	7	8	9	10	…
P				≈ 1	0.935	0.552	0.182	0.041	0.008	0.001	≈ 0	

With a risk α = 0,05 **k = 7**



Conclusion

- **Some series of disasters may be explained by statistical variations**

  The hypothesis of the constant temporal density of accidents is not rejected

- **The phenomenon of cluster may be amplified**

  Media attention

  Situational or exogenous causes

- **Intuition is misleading**

  Unexpected results

  The coincidence does not share the events in a regular way