Event flows modeling

Crimes are random processes

Jean-Marc Vincent and Christine Plumejeaud

MESCAL-INRIA Project
Laboratoire d'Informatique de Grenoble
email Jean-Marc.Vincent@imag.fr

ANR GEOMEDIA

Outline

Event flow model

Continuous time modeling : occurrence of events

- traffic on a road, arrivals at a taxi station,
- birth and death in demography
- hit on web servers, messages on a link, phone calls
- crimes, delinquency,...
- ...

Basic model of a 2 time scale system

Randomness due to complexity of the environment
Superposition of many behaviors

$$\{N_t\}_{t\in\mathbb{R}}$$

N_t = number of events in $[0, t[$

Event flow model

Continuous time modeling : occurrence of events

- traffic on a road, arrivals at a taxi station,
- birth and death in demography
- hit on web servers, messages on a link, phone calls
- crimes, delinquency,...
- ...

Basic model of a 2 time scale system

Randomness due to complexity of the environment
Superposition of many behaviors

$$\{N_t\}_{t\in\mathbb{R}}$$

$N_t =$ number of events in $[0, t[$

Flow characteristics

Communication model : 2 counting processes
- emission/reception process

Throughput

$$\lambda = \lim_{t \to +\infty} \frac{1}{t} N_t.$$

Volume, Streaming
Link capacity...

Latency

$$\mathbb{E}(T_{n+1} - T_n)$$

Response time
Time constraints

Jitter

$$\mathbb{V}ar(T_{n+1} - T_n)$$

Variability of inter-arrivals
Periodic behavior

Loss rates

Communication reliability
Perturbed events

$$\lambda_{emission} - \lambda_{reception}$$

Justice management

RECHERCHES

SUR LA

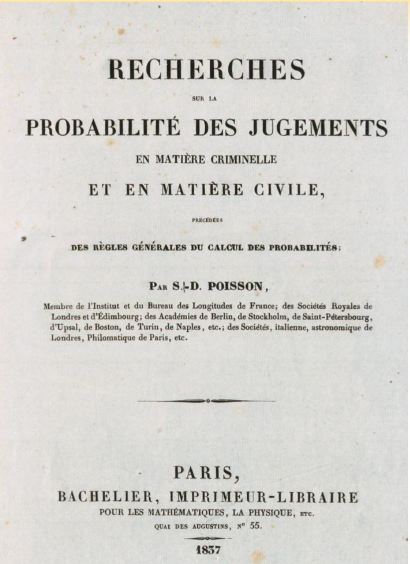
PROBABILITÉ DES JUGEMENTS

EN MATIÈRE CRIMINELLE

ET EN MATIÈRE CIVILE,

PRÉCÉDÉES

Justice management(2)


RECHERCHES

SUR LA

PROBABILITÉ DES JUGEMENTS

EN MATIÈRE CRIMINELLE

ET EN MATIÈRE CIVILE,

PRÉCÉDÉES

DES RÈGLES GÉNÉRALES DU CALCUL DES PROBABILITÉS;

PAR S.-D. POISSON,

Membre de l'Institut et du Bureau des Longitudes de France; des Sociétés Royales de Londres et d'Édimbourg; des Académies de Berlin, de Stockholm, de Saint-Pétersbourg, d'Upsal, de Boston, de Turin, de Naples, etc.; des Sociétés, italienne, astronomique de Londres, Philomatique de Paris, etc.

PARIS,

BACHELIER, IMPRIMEUR-LIBRAIRE

POUR LES MATHÉMATIQUES, LA PHYSIQUE, ETC.

QUAI DES AUGUSTINS, N° 55.

1837

Justice management

celui des accusés, a aussi augmenté d'une manière progressive (1). Voici des résultats extraits de ces documents, et que l'on pourra comparer à ce qui a lieu dans notre pays. Les nombres suivants se rapportent seulement à l'Angleterre et au pays de Galles. Ils répondent à trois périodes de chacune sept années, finissant en 1818, 1825, 1832.

	NOMBRE des accusés.	NOMBRE des condamnés.	RAPPORT du second nombre au 1er.	CONDAMNÉS à mort.	EXÉCUTÉS.	CONDAMNÉS à un emprisonnement de deux ans ou au-dessous.
1re période,	64538	41054	0,636...	5802	635	27168
2e	93718	63418	0,677...	7770	579	42713
3e	127910	90240	0,705...	9729	414	58757

Justice management

4755, 5081, 5018, 5552, 5582, 5296;

les nombres correspondants des condamnés, sous l'empire d'une même législation criminelle, se sont élevés à

882, 967, 948, 871, 834, 766,

pour les crimes de la première espèce, et à

3155, 3381, 3288, 3680, 3641, 3364,

pour ceux de la seconde. De là, on déduit

0,4649, 0,5071, 0,4961, 0,4723, 0,4657, 0,4598,

pour les rapports des nombres de condamnés à ceux des accusés de crimes contre les personnes, et

0,6635, 0,6654, 0,6552, 0,6628, 0,6523, 0,6352,

pour les rapports des nombres de condamnés à ceux des accusés de crimes contre les propriétés; où l'on voit que les uns et les autres n'ont pas beaucoup varié d'une année à une autre, mais que les derniers excèdent notablement les premiers.

En prenant pour μ et a_5 les sommes des nombres d'accusés et de condamnés dans le cas des crimes contre les personnes, et pour μ' et a'_5 leurs sommes dans le cas des crimes contre les propriétés, nous aurons

$$\mu = 11016, \quad a_5 = 5268, \quad \mu' = 31284, \quad a'_5 = 20509;$$

d'où il résulte ces deux rapports :

$$\frac{a_5}{\mu} = 0{,}4782, \qquad \frac{a'_5}{\mu'} = 0{,}6556,$$

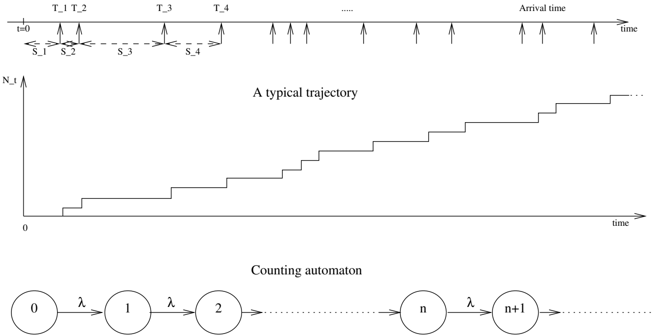
dont le second surpasse le premier d'un peu plus du tiers de celui-ci. Au moyen de ces nombres, on trouve

$$0{,}4782 \mp \alpha\,(0{,}00675)$$

pour les limites (a) de l'inconnue R_5, relative aux crimes contre les personnes, et

$$0{,}6556 \mp \alpha\,(0{,}00380),$$

Counting process



Macroscopic modeling

Definition (Macroscopic definition)

A continuous time stochastic process $\{N_t\}_{t\in\mathbb{R}^+}$ *is a counting Poisson process with intensity* λ *iff*

1. $N_0 = 0$
2. $\{N_t\}_{t\in\mathbb{R}^+}$ *have independent increments*
3. *The number of events occurring in a time interval* $]a, b]$ *is Poisson distributed with parameter* $\lambda(b-a)$*;*

$$\mathbb{P}(N_b - N_a = k) = e^{-\lambda(b-a)}\frac{(\lambda(b-a))^k}{k!}.$$

Properties

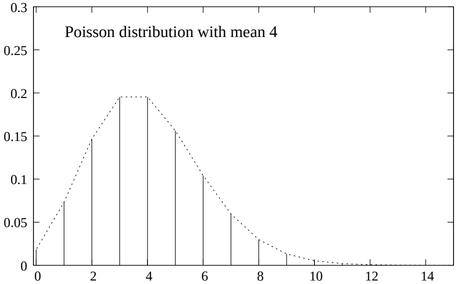
- Increments are stationary : homogeneous in time
- Linearity $\mathbb{E}(N_b - N_a) = \lambda(b-a)$
- $\lambda =$ intensity or throughput of the process
  number of events per unit of time

Poisson distribution $\mathcal{P}(\lambda)$

X random variable Poisson distributed with parameter λ

$$\mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

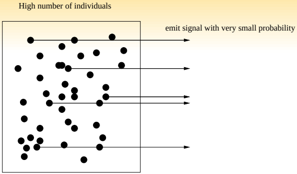
$$\mathbb{E}X = \lambda; \quad \mathbb{V}arX = \lambda.$$



If X and Y are independent random variable Poisson distributed with mean λ and μ then

$$X + Y \sim \mathcal{P}(\lambda + \mu).$$

Interpretation



N elements, each of them $p =$ probability of signal emission
X total number of emissions: binomial distribution $\mathcal{B}(N, p)$.
$\mathbb{E}X = Np \stackrel{def}{=} \lambda$ mean number of emissions.

$$\begin{aligned}
\mathbb{P}(X = k) &= \binom{N}{k} p^k (1-p)^{N-k}; \\
&= \underbrace{\frac{N(N-1)\cdots(N-k+1)}{N.N\cdots N}}_{\to 1} \underbrace{\frac{1}{(1-\frac{\lambda}{N})^k}}_{\to 1} \frac{\lambda^k}{k!} \underbrace{(1-\frac{\lambda}{N})^N}_{\to e^{-\lambda}}; \\
&\simeq e^{-\lambda}\frac{\lambda^k}{k!}.
\end{aligned}$$

for very large N, X is asymptotically Poisson distributed

Flow analysis

Traffic generated by a huge amount of individuals ⇒ Poisson process

- requests arrival on a web server
- arrivals of phone calls
- routed packets in a network
- cars on a road network
- ...

How to detect non-Poisson flows

- Time dependence or correlation (burstyness, periodicity,...)
- Mean < Variance : too much variability
- smoothers of the traffic (peack avoidance strategies)
- ...

Microscopic modeling

Definition

Microscopic definition A continuous time stochastic process $\{N_t\}_{t\in\mathbb{R}^+}$ is a counting Poisson process with intensity λ iff

1. $N_0 = 0$
2. $\{N_t\}_{t\in\mathbb{R}^+}$ *have independent and stationary increments*
3. *On a very small intervall* $]t, t+dt]$ *we have :*

$$\begin{array}{rcl} \mathbb{P}(N_{t+dt} - N_t = 1) & = & \lambda dt + o(dt) \\ \mathbb{P}(N_{t+dt} - N_t = 0) & = & 1 - \lambda dt + o(dt) \\ \mathbb{P}(N_{t+dt} - N_t \geqslant 2) & = & o(dt) \end{array}$$

Properties

- increments are stationary : homogeneous in time
- $\mathbb{E}(N_b - N_a) = \lambda(b - a)$
- $\lambda =$ intensity or throughput of the process
  number of events per unit of time

Differential system

$$p_n(t) = \mathbb{P}(N_t = n)$$

$$\begin{aligned}
p_n(t+dt) &= \mathbb{P}(N_{t+dt} = n) \\
&= \mathbb{P}(N_{t+dt} = n | N_t = n)\mathbb{P}(N_t = n) \text{ nothing happens} \\
&\quad +\mathbb{P}(N_{t+dt} = n | N_t = n-1)\mathbb{P}(N_t = n-1) \text{ one arrival} \\
&\quad +\mathbb{P}(N_{t+dt} = n | N_t < n-1)\mathbb{P}(N_t < n-1) \text{ more than one arrival} \\
&\quad \text{ independent increments} \\
&= \mathbb{P}(N_{t+dt} - N_t = 0)p_n(t) \text{ nothing happens} \\
&\quad +\mathbb{P}(N_{t+dt} - N_t = 1)p_{n-1}(t) \text{ one arrival} \\
&\quad +\mathbb{P}(N_{t+dt} - N_t \geqslant 2)\mathbb{P}(N_t < n-1) \text{ more than one arrival} \\
&= (1 - \lambda dt + o(dt))p_n(t) + (\lambda dt + o(dt))p_{n-1}(t) + o(dt) \\
&= p_n(t) + \lambda(p_{n-1}(t) - p_n(t))dt + o(dt)
\end{aligned}$$

recurrent differential equations

$$p_n'(t) = \lambda(p_{n-1}(t) - p_n(t)), \quad p_0(t) = \lambda p_0(t)$$

which is solved by recurrence (put $q_n(t) = e^{\lambda t} p_n(t)$) $\quad p_n(t) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}$

Differential system

$p_n(t) = \mathbb{P}(N_t = n)$

$$\begin{aligned}
p_n(t+dt) &= \mathbb{P}(N_{t+dt} = n) \\
&= \mathbb{P}(N_{t+dt} = n | N_t = n)\mathbb{P}(N_t = n) \text{ nothing happens} \\
&\quad + \mathbb{P}(N_{t+dt} = n | N_t = n-1)\mathbb{P}(N_t = n-1) \text{ one arrival} \\
&\quad + \mathbb{P}(N_{t+dt} = n | N_t < n-1)\mathbb{P}(N_t < n-1) \text{ more than one arrival} \\
&\quad \text{ independent increments} \\
&= \mathbb{P}(N_{t+dt} - N_t = 0)p_n(t) \text{ nothing happens} \\
&\quad + \mathbb{P}(N_{t+dt} - N_t = 1)p_{n-1}(t) \text{ one arrival} \\
&\quad + \mathbb{P}(N_{t+dt} - N_t \geqslant 2)\mathbb{P}(N_t < n-1) \text{ more than one arrival} \\
&= (1 - \lambda dt + o(dt))p_n(t) + (\lambda dt + o(dt))p_{n-1}(t) + o(dt) \\
&= p_n(t) + \lambda(p_{n-1}(t) - p_n(t))dt + o(dt)
\end{aligned}$$

recurrent differential equations

$$p_n'(t) = \lambda(p_{n-1}(t) - p_n(t)), \quad p_0(t) = \lambda p_0(t)$$

which is solved by recurrence (put $q_n(t) = e^{\lambda t}p_n(t)$) $\qquad p_n(t) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}$

Interarrivals

Let t be a fixed time and let T_t be the time to the next arrival after time t.

$$\mathbb{P}(T_t \geqslant s) = \mathbb{P}(N_{t+s} - N_t = 0) = e^{-\lambda s}.$$

T_t is exponentially distributed with rate λ

The inter-arrival process $\{A_n\}_{n\in\mathbb{N}}$ is a sequence of independent exponentially distributed random variable with rate λ

Exponential distribution

Density, rate λ :

$$f(x) = \lambda e^{-\lambda x}$$

Cumulative distribution function

$$F(x) = 1 - e^{-\lambda x}$$

Mean, Variance

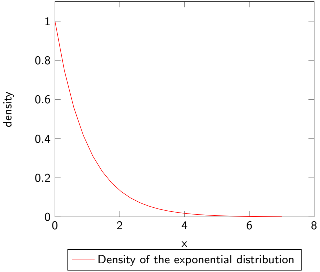
$$\mathbb{E}X = \frac{1}{\lambda}, \quad \mathrm{Var}X = \frac{1}{\lambda^2}$$

Hazard rate

$$h(x) = \lambda$$

Laplace transform

$$\mathcal{L}(t) = \mathbb{E}e^{-tX} = \frac{\lambda}{t+\lambda}$$



Memoryless property

$$\mathbb{P}(X > t + s | X > t) = \mathbb{P}(X > s)$$

Equivalence of definitions

Theorem (Global vision)

Macroscopic, **microscopic** *and* **independent exponentially distributed inter-arrivals** *are equivalent definitions of a Poisson process*

Proof : classical books

Maximum Entropy Process

Spread of Points

Let $[a, b]$ an interval, knowing $N_b - N_a = n$ the n points are distributed as the rearrangement of n points independents and uniformly distributed points on $[a, b]$

Theorem (Information Approach)

The **Poisson process** *is the model of process with a fixed intensity and* **minimal "a priori" information**

Maximum Entropy Process

Spread of Points

Let $[a, b]$ an interval, knowing $N_b - N_a = n$ the n points are distributed as the rearrangement of n points independents and uniformly distributed points on $[a, b]$

Theorem (Information Approach)

The **Poisson process** *is the model of process with a fixed intensity and* **minimal "a priori" information**

Scale Invariance

Theorem (Superposition)

Let $\{N_t^1\}$ and $\{N_t^2\}$ be two **independent** *Poisson processes then $\{(N^1 + N^2)_t\}$ is a Poisson process with rate $\lambda_1 + \lambda_2$*

Theorem (Extraction)

Probabilistic thinning of a Poisson process is a Poisson process.

Scale Invariance

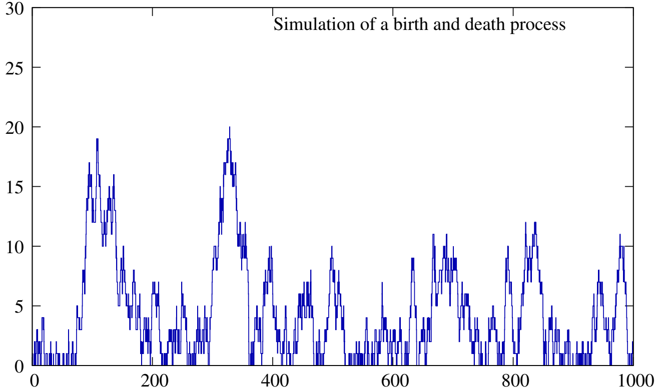
Theorem (Superposition)

Let $\{N_t^1\}$ *and* $\{N_t^2\}$ *be two* **independent** *Poisson processes then* $\{(N^1 + N^2)_t\}$ *is a Poisson process with rate* $\lambda_1 + \lambda_2$

Theorem (Extraction)

Probabilistic thinning of a Poisson process is a Poisson process.

Poisson Clumping heuristic



Non-homogeneity

Definition (Macroscopic definition)

A continuous time stochastic process $\{N_t\}_{t\in\mathbb{R}^+}$ is a non-homogeneous counting Poisson process with intensity $\lambda(t)$ iff

1. $N_0 = 0$
2. $\{N_t\}_{t\in\mathbb{R}^+}$ *have independent increments*
3. *The number of events occurring in a time interval $]a,b]$ is Poisson distributed with parameter $\int_a^b \lambda(t)dt = \Lambda(b) - \Lambda(a)$;*

$$\mathbb{P}(N_b - N_a = k) = e^{-(\Lambda(b)-\Lambda(a))}\frac{(\Lambda(b)-\Lambda(a))^k}{k!}.$$

- embedded periodicity
- exceptional period
- ...

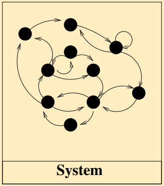
Doubly Stochastic

Randomness on the intensity

$$\{\lambda_t\}_{t \in real^+}$$

stationary process. Conditioned by λ_t, $\{N_t\}_{t \in \mathbb{R}^+}$ is a Poisson process.

Markov-modulated Poisson process



- several time scales
- algebra by composition of automata
- ON/OFF systems
- ...

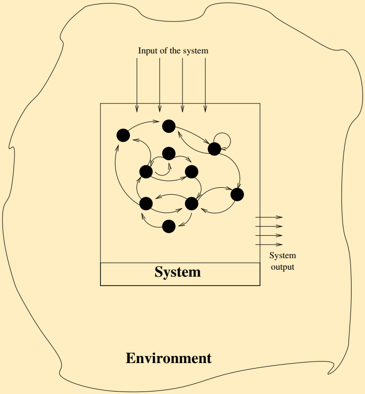
Doubly Stochastic

Randomness on the intensity

$$\{\lambda_t\}_{t \in real^+}$$

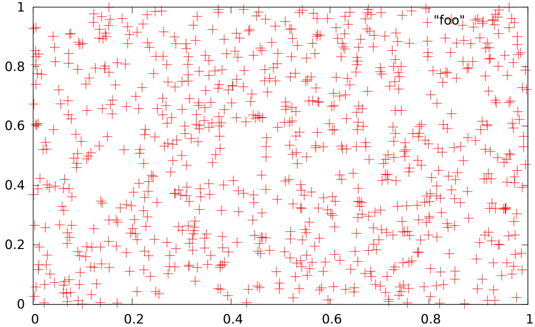
stationary process. Conditioned by λ_t, $\{N_t\}_{t\in\mathbb{R}^+}$ is a Poisson process.

Markov-modulated Poisson process



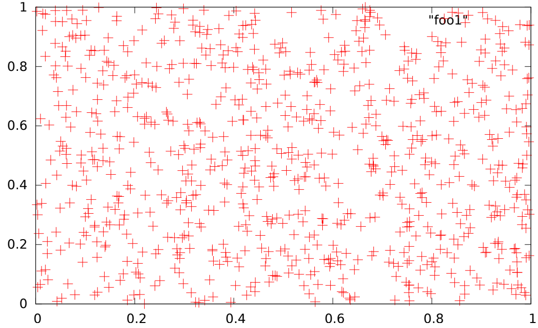
- several time scales
- algebra by composition of automata
- ON/OFF systems
- ...

Spatial Poisson Process



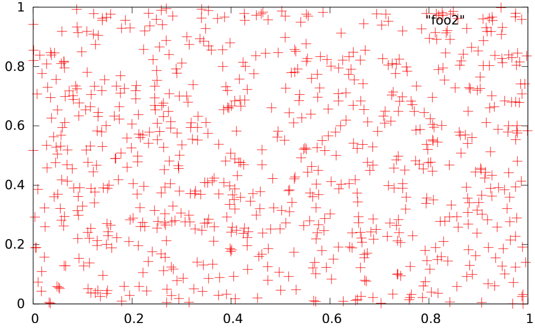
$$\mathbb{P}(N_A = k) = e^{-\mu(A)} \frac{\mu(A)^k}{k!}.$$

Spatial Poisson Process



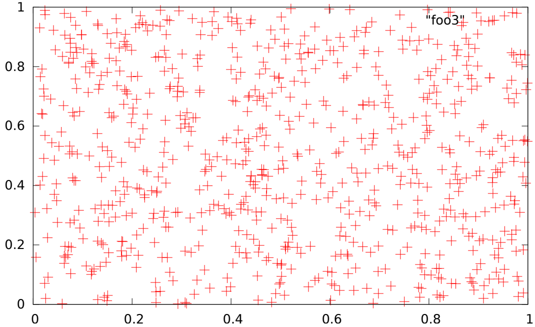
$$\mathbb{P}(N_A = k) = e^{-\mu(A)} \frac{\mu(A)^k}{k!}.$$

Spatial Poisson Process



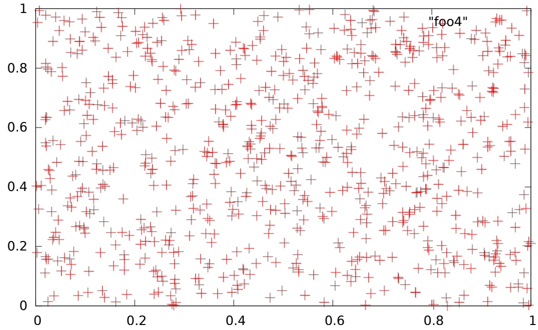
$$\mathbb{P}(N_A = k) = e^{-\mu(A)} \frac{\mu(A)^k}{k!}.$$

Spatial Poisson Process



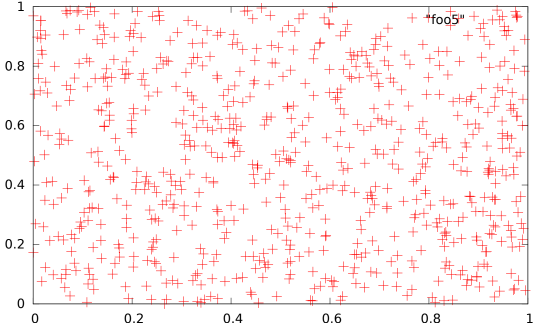
$$\mathbb{P}(N_A = k) = e^{-\mu(A)} \frac{\mu(A)^k}{k!}.$$

Spatial Poisson Process



$$\mathbb{P}(N_A = k) = e^{-\mu(A)} \frac{\mu(A)^k}{k!}.$$

Spatial Poisson Process



$$\mathbb{P}(N_A = k) = e^{-\mu(A)} \frac{\mu(A)^k}{k!}.$$

Synthesis

Base model

1. reference model $\Rightarrow$ deviation
2. refinement $\Rightarrow$ model extension
3. multi-scale analysis (algebra for superposition, composition,...)
4. statistical methods $\Rightarrow$ Poisson regression

Geomedia questions

1. Are RSS flow relevant of Poisson models ?
2. Scales of homogeneity ?
3. Development of the algorithms ?
4. ..