Quelques aspects de probabilistes / statistique non paramétrique des processus de comptage et des processus ponctuels

Introduction : ou les trucs que je trouve amusants.

Biblio : Daley Vere-Jones An introduction to the theory of point processes (Tome 1 et 2). dernière version

Andersen Borgan Gill Keiding Statistical models based on counting processes

Brémaud Point processes and Queues.

I/ Introduction (historique) et exemples

Sujet terriblement vaste qui représente un nombre de situations incalculables $\longrightarrow$ un aperçu de ce qui m'amuse plus particulièrement

1) Processus ponctuel

. N est un processus ponctuel si et seulement si c'est un ensemble aléatoire au plus dénombrable de point de $\mathbb{X}$ (espace mesurable).

. Si A est un ensemble mesurable on note N_A le nombre de points qui tombent dans A. $dN_t = \sum \delta_{T_i}$

2) En vrac :

Sans marques
- durées de vie / instants de mort dans une population
- dates de désintégration d'atomes dans matériau radioactif
- instants / position / durée d'appels téléphoniques (portable)
- date / position / magnitude de tremblements de terre
- date de découverte / position / taille champs de pétrole
- position / taille / âge arbres dans une forêt
- dates de panne de machine puis réparation
- entrées / sorties des abeilles d'une ruche
- instant où un potentiel d'action passe à un endroit précis sur un neurone

Avec marques (ie ont une info qualitative en plus)
- mort cancer ou autre chose
- plante OGM ou pas
- achat ou vente actifs — faillites banque / PME.
- plusieurs actifs
- plusieurs neurone

ADN . longue chaîne de nucléotides $\simeq 10^6$
- motifs ~~ex~~ = 4 à 6 lettres
ou position notable = début de gènes
⇝ points sur la chaîne d'ADN
(une des modélisations possibles)

⇝ On a noté tout ça car on soupçonne que ça donne de l'info sur le problème sous jacent

⟹ modélisation • qui puisse prendre en compte tout ce qu'on imagine
• qui représente un vrai objet probabiliste
• qui permette de quantifier les informations que l'on veut extraire

⇝ Ce n'est pas forcément (voire rarement) n variables iid

⇝ Origine stats / probas applis ?

3) Life tables (comme c'est l'origine, on va y passer un peu de temps)

On observe un groupe de mettons 1000 personnes et on note quand ils meurent. (en année x).

On a donc accès à :

l_x : nombre de personnes ayant atteint l'âge x

d_x : nombre de personnes mourant à l'âge x (+ chouille)

ie $l_x - l_{x+1}$

q_x : proportion de gens mourant juste après x parmi ceux qui ont atteint l'âge x ie $\frac{d_x}{l_x}$

Ces quantités représentent la version empirique et discrétisée de :

Si T est la durée de vie

$l_x \longrightarrow S(x) = P(T > x)$ (fonction de survie (queue

$q_x \longrightarrow q(x) = P(T \in]x, x+dx] \mid T > x)$

(taux de hasard (accident !)
(hazard rate

Rmq si T a pour densité f

$$S(x) = \int_x^{+\infty} f(y)\,dy \qquad f(x) = \frac{-dS}{dx}$$

$$q(x) = \frac{f(x)}{S(x)} = \frac{d}{dx} \log S(x) \quad , \quad S(x) = \exp\left(-\int_0^x q(y)\,d\right.$$

<u>1ère table de vie</u> : Graunt (1662) → Londres

Pascal (1654-1679). Euler proba.

Halley, Huyghens, de Moivre (1700...)
Laplace, Euler...

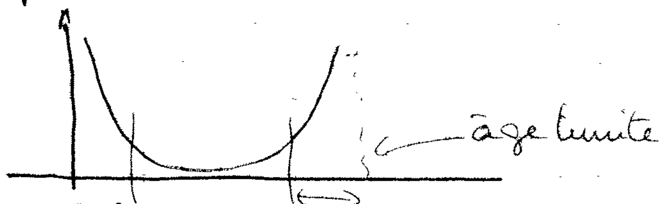
<u>La variable Exponentielle</u> $q(x) = \lambda = cte$

c'est la variable sans mémoire $\mathcal{E}(\lambda)$ / $S(x) = e^{-\lambda x}$

" l'âge n'influence pas la mort qui est aussi probable jeune que vieux ".

<u>Rmq</u> si $q(x) \searrow$ on est mieux si on est vieux
$q(x) \nearrow$ on est plus mal si on est vieux

→ Une Courbe typique



Pour trouver la courbe ? → $n = 1000$ répliques iid

oui mais à Londres Peste, accident, etc...
départ, arrivée...

⇝ Actuariat (population meme mais peut aussi croître)

4) <u>Renouvellement</u>

On regarde des durées entre évenements

La loi la plus connue $\tau_1, \dots, \tau_n$ iid $\mathcal{E}(\lambda)$

$$\overset{\tau_1}{\longleftrightarrow}\ \overset{\tau_2}{\longleftrightarrow} \quad \dots$$
$$0 \quad T_1 \quad T_2 \quad \dots \quad T_n$$

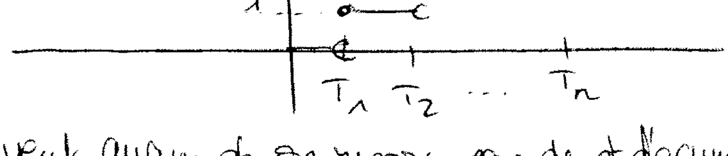
T_n suit la loi d'Erlang

(Erlang Danois 1920 qui modélise les appels téléphoniques)

⇝ ça remonte Newton (1728) Chronologies Hébreu/Juive ⇝ fonction des durées de règ

5) <u>Processus de comptage</u> (sur $\mathbb{R}$) (Car le temps c'est important !)

Si N est un processus ponctuel sur $\mathbb{R}^+$ on range les points dans l'ordre et



$N_t = N_{[0,t]}$ est le processus de comptage associé

Si on veut aucun pb on suppose pas de pt d'accumulation (fini)

N_t est un processus croissant càdlàg de sauts $= 1$

II Processus de Poisson (Le plus connu, le plus facile, défini en toute dim !) .

1) Définition

N est un processus de Poisson de mesure moyenne μ sur X si . $\forall n$ $\forall A_1, \dots, A_n$ mesurables disjoints $N_{A_1}, \dots, N_{A_n}$ sont des va $\perp\!\!\!\perp$

. $N_A \sim \mathcal{P}(\mu(A))$

$$\left(\text{ie } \forall k \quad P(N_A = k) = \frac{\mu(A)^k}{k!} e^{-\mu(A)} \right)$$

Rmq . En fait très peu d'hypothèses : principalement $\perp\!\!\!\perp$ et non accumulation. La loi de Poisson vient du passage à la limite ... (caractère ost. divisible)

. Un processus de Poisson est infiniment divisible :

Si $N^1 \dots N^k$ $PP(\frac{\mu}{k})$

alors $N^1 + \dots + N^k$ (au sens la réunion de tous les points)

est $PP(\mu)$.

. Généralement $X = \mathbb{R}^d$ et mesure de Lebesgue

→ Si $\mu \ll$ Lebesgue $\frac{d\mu}{dx} = \lambda(x)$ est l'intensité du Processus de Poiss
Not $PP(\mu)$ ou $PP(\lambda)$.

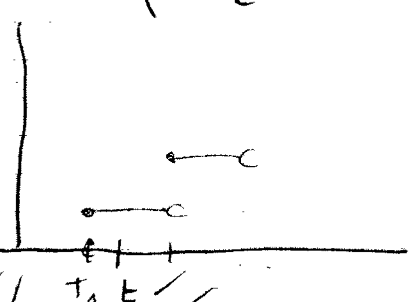
→ Un processus de Poisson est homogène si $\lambda(x) = cte = \lambda$
inhomogène sinon.

NE PAS OUBLIER → ici

2) Que peut-on dire d'un processus de Poisson vu en tant que processus de comptage sur $\mathbb{R}^+$?

$$P(N_t = 0) = P(T_1 > t) = e^{-\lambda t} \quad (\text{Rmq } \mu[0,t] = \lambda t)$$

donc $T_1 \sim \mathcal{E}(\lambda)$



On peut en fait montrer que ça correspond à un processus de renouvellement où les τ_i sont des va iid $\mathcal{E}(\lambda)$

(de plus $|N_X| = n$ la loi des pts sur X est une loi uniforme d'un éch. $(x_1, \dots, x_n)$)

(La 1ère modélisation des appels téléphoniques par Erlang revient à un processus de Poisson homogène mais pb de l'homogénéité ... etc ...

3) Exemple typique d'un processus de Poisson homogène est la <u>désintégration d'atomes radioactifs</u>.

<u>Pourquoi ?</u>

. La durée de vie d'un atome $\longrightarrow$ loi exponentielle (sans mémoire)
donc $P(T \in [t, t+dt] \mid T > t) = \lambda$ <u>constant</u> (hazard rate)

. λ est <u>très</u> faible $\sim 1{,}4 \cdot 10^{-11}\, s^{-1}$
(ie il y a ~~1,4~~ 1,4 chance sur 100 milliards qu'un atome précis de radium se désintègre dans l'intervalle d'1 s

. On observe M (très grand) ~~atomes~~ atomes
(0,1 µg de radium $= 2{,}66\ 10^{14}$ atomes)

On a $X_1, \dots, X_M$ leur durée de vie (cf table de vie)
(Au bout de X_i s ~~l'atome~~ radium n° i devient ras

$P(k$ désintégrations entre $[0,T])$
$= P(\# \{i \mid X_i \le t\} = k)$

a $\# \{i \mid X_i \le t\} \sim \mathcal{B}(M, \underbrace{1 - e^{-\lambda T}}_{\simeq \lambda T})$

$\simeq \lambda T$ très faible aux échelles de temps humaines ~~[illegible]~~ heure

$\Longrightarrow$ Approximation Binomiale Poisson

$\approx \mathcal{P}(M \lambda T)$ si $T = 1\,ms$ $\mathcal{P}(3{,}7)$

On voit en moyenne 3,7 désintégration par ms pour 0,1 µg de radium

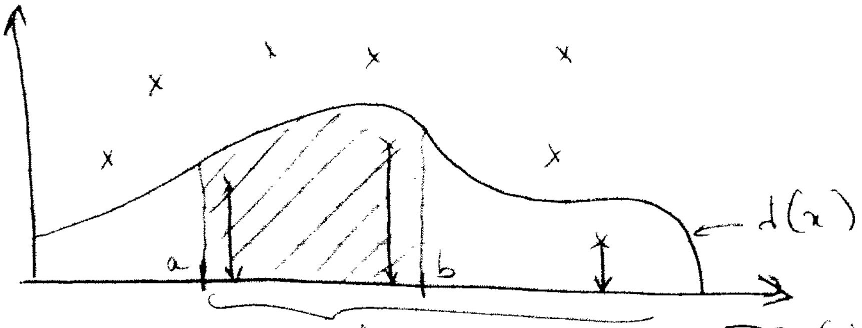
$\Longrightarrow$ On approche en fait le processus de désintégration par un processus de Poisson (cf lois mises en évidence par Bateman Rutherford et Geiger sur rayonnement α en 1910)

Bien sûr à très longue échéance le nb M d'atome de radium décroit exponentiellement lentement vers 0.

4) Comment construire un PP($\lambda(x)$) sur $\mathbb{R}_+$?

Il suffit de prendre un PP(1), CP sur $\underline{\underline{\mathbb{R}^2_+}}$.

(Si $\lambda(x)$ est L^1 loc)



on obtient ici un PP($\lambda(x)$) sur $\mathbb{R}_+$.

vérifions les 2 points.

- $\perp\!\!\!\perp$? OK car CP l'est (sur cylindres)
- $N_{[a,b]} = CP_{[////]} \sim \mathcal{P}(\text{aire } ////)$

$$= \mathcal{P}\left(\int_a^b \lambda(x)\,dx\right).$$

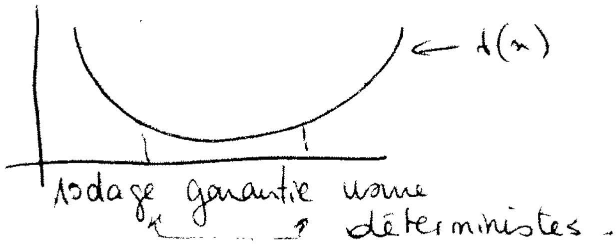
III Processus de Comptage généraux sur $\mathbb{R}_+$ et intensité

<u>Motivation</u> (ma motivation de statisticienne)

⇝ on veut de manière facile synthétiser les divers facteurs qui peuvent favoriser l'apparition d'un point à l'instant t.

→ Si c'est un processus de Poisson (inhomogène) c'est facile, il suffit de donner la fonction $\lambda(x)$ intensité (déterministe) du PP($\lambda(x)$)

<u>ex</u> Panne d'une machine on sait que



(forme (U baignoire))

Mais que se passe-t-il si d'autres facteurs (p ex aléatoire).

1) L'intensité d'un processus de comptage

Soit $(N_t)_{t\geq 0}$ un processus de comptage (sur $\mathbb{R}^+$)
(sans accumulation ps)

De manière informelle et lorsque cela existe, l'intensité $\lambda(t)$ de N_t est la limite des quotient par dt de ~~(probabilité~~ d'avoir un saut entre t et ~~[illegible]~~ l'espérance du nombre de saut t+dt sachant tout le passé jusqu'au temps t exclus.
(comme pas d'accumulation ça revient au m)

Pour l'instant on va essayer autant que possible de ne pas rentrer plus avant dans le formalisme (martingale etc filtration)

Rmq : Pour un PP $(\lambda(x))$ on retombe déjà sur ce qu'on veut.
Si on note $\mathcal{F}_{t^-}$ l'information (tribu...) du passé,

$$P\left(N_{[t,t+dt]} \geq 1 \mid \mathcal{F}_{t^-}\right) = P\left(N_{[t,t+dt]} \geq 1\right) \text{ par } \perp\!\!\!\perp$$
$$= 1 - e^{-\int_t^{t+dt}\lambda(u)\,du} \underset{t\to 0}{\approx} \lambda(t)\,dt.$$

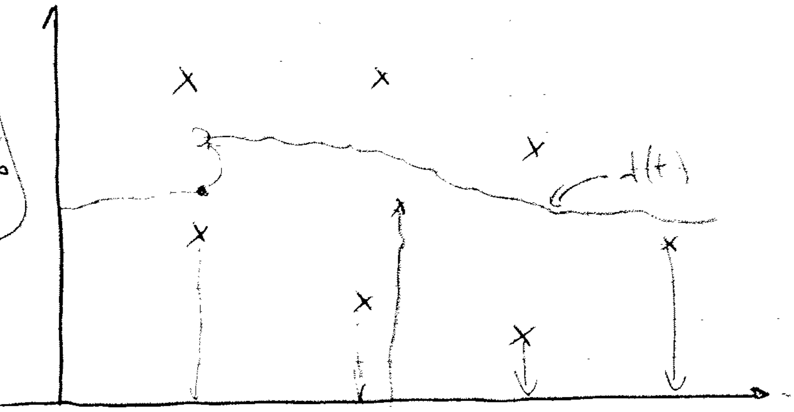
(autre déf) $E\left(N(t,t+dt) \mid \mathcal{F}_{t^-}\right) = \lambda(t)\,dt$)

2) Comment interpréter l'intensité $\lambda(t)$?

→ via le thinning (cf aussi les algorithmes de simulation qui en découlent).

Rmq du coup si intensité L'bc pas d'explosions

↑ NE PAS OUBLIER



cf PP(1) sur $\mathbb{R}_+^2$
traçons l'intensité de gauche à droite. Eventuellement les points découverts peuvent modifier la courbe mais on doit quand même pouvoir la tracer au fur et à mesure de gauche à droite.

Alors sur une portion $[t, t+dt]$,

$$E\left(N_{[t,t+dt]} \mid \mathcal{F}_{t^-}\right) = \lambda(t)\,dt.$$

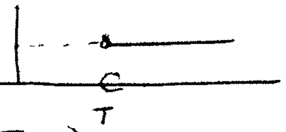
la courbe $\lambda(t) = \lambda(t^-)$ en aire
↳ on peut le sortir de l'intégrale

Alors (H) $\lambda(t)$ est continue à gauche positive ou nulle. (Prop 7.5 I Daley Vere Jones)
(Brémaud Massoulié 96)

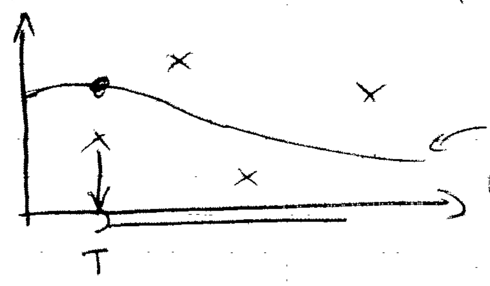
3) Calcul d'intensité, quelques exemples

Une durée de vie :

On observe un seul point T de densité f (c°), ie $N = \{T\}$

Donc on regarde $N_t = 1\!\!1_{T \leq t}$  càd de sau

$$P(T \in [t, t+dt] \mid \mathcal{F}_{t-})$$



$\hookrightarrow 0$ si $T < t$

qqch si $T \geq t$ $P(T \in [t,t+dt] \mid T \geq t) = q(t)\,dt$ ← taux de hasard

il n'y a que ça comme info dans le passé

Seulement une fois qu'on a projeté T on ne veut plus aucun autre poi

donc $\boxed{\lambda(t) = q(t)\, 1\!\!1_{T \geq t}}$ (OK càg).

n durées de vie :

On observe $T_1 .. T_n$ iid de densité f (c°).

On regarde N_t le processus agrégé $= \sum_{i=1}^{n} 1\!\!1_{T_i \leq t}$

Quelle est son intensité ?

$$E(N_{(t,t+dt)} \mid \mathcal{F}_{t-}) = \sum_{i=1}^{n} E(1\!\!1_{T_i \leq t} \mid \mathcal{F}_{t-})$$

↑ contient entre autre l'information sur T_i.

$$= \left(\sum_{i=1}^{n} 1\!\!1_{T_i \geq t} \right) q(t).$$

Une durée de vie censurée

Le patient a une durée de vie T de densité f c°.

Mais on n'observe pas toujours T car par exemple censure à droite (et mort accidentelle et pas cause)

$T \perp C$ et on observe $\begin{cases} X = T \wedge C \\ \delta = \begin{matrix} 0 & \text{si } X \neq T \\ 1 & \text{sinon} \end{matrix} \end{cases}$

pas forcément des fois C débile d'autres fois trucs + compliqués

$$= 1\!\!1_{T \leq C}$$

On regarde $N_t = \delta\, 1\!\!1_{X \leq t}$ (ie on ne compte que si pas censure N est soit ϕ soit réduit à 1 pt)

$$E\left(N_{(t,t+dt)} \mid \mathcal{F}_{t^-}\right) \begin{cases} \to 0 \text{ si } X<t. \\ \to \text{qqch si } X \geq t \to \text{evenement qui appartient à } \mathcal{F}_{t^-}. \end{cases}$$

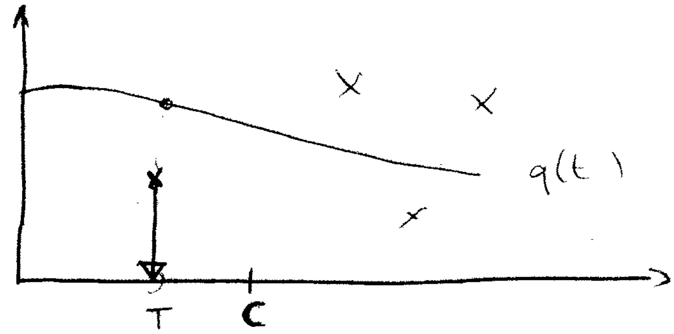
$= P\left(X \in [t, t+dt] \text{ et } \delta=1 \mid \text{ni } X \text{ ni } C \text{ ent}^- \right) \mathbb{1}_{X \geq t}$

$= P\left(T \in [t, t+dt] \text{ et } \underbrace{C \geq T}_{\text{pas d'info supp.}} \mid T>t \text{ et } C>t \right) \mathbb{1}_{X \geq t}$

$= P\left(T \in [t, t+dt] \mid T>t \underbrace{\text{ et } C>t}_{\text{ne sert plus à rien}} \right) \mathbb{1}_{X \geq t}$

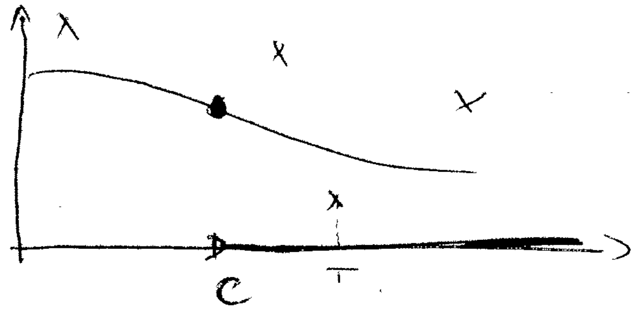
$= q(t) \mathbb{1}_{X \geq t}$

Pour le tracer :



cas $\delta = 1$.

(cas C déterminé)



Remarque d'où les pbs de censure informative ou pas cf Andersen etc

cas $\delta = 0$

Version agrégée

$T_1 \dots T_n$ iid $\qquad X_i = T_i \wedge C_i$

$C_1 \dots C_n$ iid $\qquad \delta_i = \mathbb{1}_{T_i \leq C_i}$

$$N_t = \sum_{i=1}^n \delta_i \mathbb{1}_{X_i \leq t}$$

$$\lambda(t) = \left(\sum_{i=1}^n \mathbb{1}_{X_i \geq t}\right) q(t). \qquad (*)$$

Rmq Cox Regression model

Pour chaque individu i, le taux de hasard dépend d'un vecteur de covariable z_i (ex pression sanguine fumeur ou non)

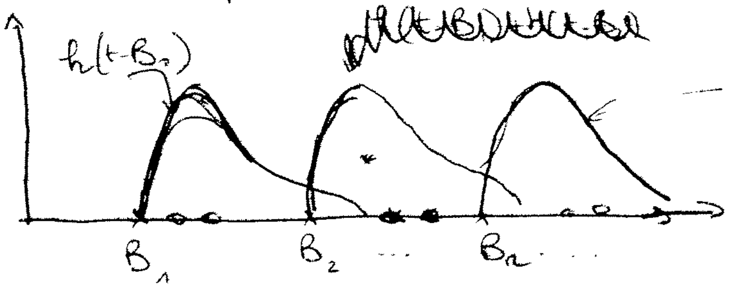
et $\quad q_i(t) = q_0(t) \exp(\beta^* z_i)$.

(les z_i dans le passé considérés comme connus)

$$\lambda(t) = \sum_{i=1}^n \mathbb{1}_{X_i \geq t} \, q_0(t) \exp(\beta^* z_i)$$

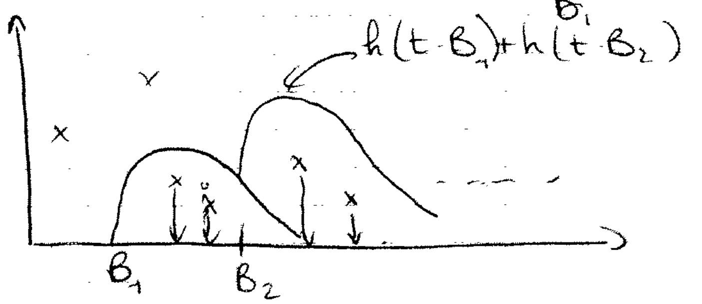
Contamination Poissonienne

Ex Faillite Banque ⇒ Faillite de PME selon un PP(h).

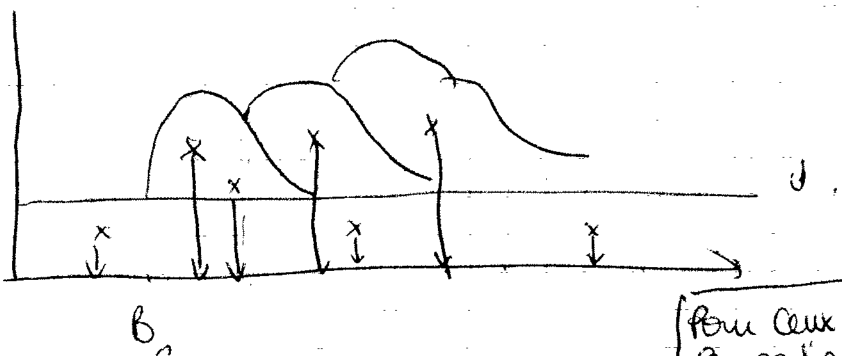


Plusieurs manières de le voir (aggrégation thinning etc).
à la fin l'intensité est $\lambda(t) = \sum_{B_i} h(t-B_i)$

Thinning



Et si autre cause → on peut rajouter un "bruit" Poissonien



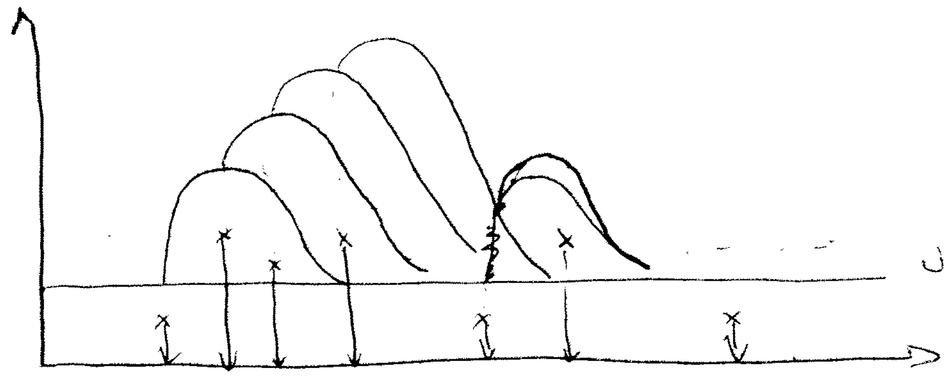
intensité $\lambda(t) = \nu + \sum_{B_i} h(t-B_i)$
= intensité conditionnellement aux B_i

Pour ceux qui connaissent
Cas particulier de Processus de Cox ou PP($\lambda_2(t)$) condit à 2 va (voire processus)
au sens Poissonnien

Processus de Hawkes (cas simple autoexcitant)

→ en fait 2 interprétations
- tremblements de terre → initiaux
  → répliques

et on peut être réplique de réplique.
Une fois que les tremblements de terre initiaux sont apparus $\sim$ PP(ν) un tb de terre donne naissance à un autre tb de terre selon un PP(h) partant du tb de terre parent.



D'où la construction ① . On fait naître les ancêtres selon ν.

② On crée leurs enfants

③ . les enfants de leurs enfants

⋮

il y a extinction p.s. du nombre de descendants par ancêtres si $\int h < 1$ (Processus de Galton Watson dont la reproduction $\sim \mathcal{P}(\int h)$.)

Mais bien sûr on peut faire le dessin de la courbe de gauche à droite ($h(0)=0$ Pas d'enfants simultanés $\Rightarrow$ pas d'accumulation)

$\Longrightarrow$ l'intensité cf thinning est donnée par

$$\lambda(t) = \nu + \sum_{T<t} h(t-T)$$

$$\boxed{\lambda(t) = \nu + \int_{-\infty}^{t^-} h(t-u)\, dN_u .}$$

(cf Hawkes et Oakes (74))

$\mathcal{F}_s$ = toutes les infos jusqu'à la date s incluse.

4) <u>Martingales</u> (version douce) (M_t est une martingale si $\forall s<t$ $E(M_t|\mathcal{F}_s) = M_s$).

$$E(N[t,t+dt] \mid \mathcal{F}_{t^-}) = \lambda(t)\,dt$$

donc $\forall s<t$ $\quad E\left(\int_s^t [dN_u - \lambda(u)\,du] \,\middle|\, \mathcal{F}_s\right) = 0 \quad (*)$

Si on pose $\int_0^t \lambda(u)\,du = \Lambda_t$,

Λ_t est appelé le compensateur prévisible de N_t ie $N_t - \Lambda_t$ est une martingale (en effet de $(*)$ $\quad E(N_t - \Lambda_t | \mathcal{F}_s) = N_s - \Lambda_s$.

prévisible ? Pour simplifier ça veut dire "traçable" de gauche à droite, en particulier si le processus est continu à gauche et ne dépend que du passé tout va bien.
($\lambda(t)$ est prévisible ~~...~~ par hyp = cas classiques)

Soit (H_s) un processus prévisible, alors comme

$$E\left(H_t N_{[t,t+dt]} \mid \mathcal{F}_{t^-} \right) = H_t \, E\left(N_{[t,t+dt]} \mid \mathcal{F}_{t^-} \right)$$
$$= H_t \, \lambda(t)\, dt$$

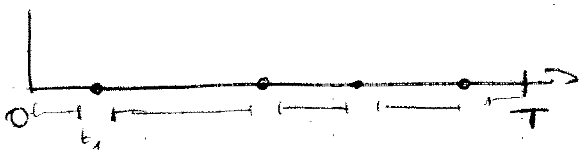
on a que $\int_0^t H_s (dN_s - \lambda(s)ds)$ est une martingale.

IV Vraisemblance d'un processus de comptage d'intensité $\lambda(t)$

1) Vraisemblance de quoi ?

Ici on vit avec pour variable aléatoire, un ensemble (supposé par hyp) fini en temps fini de points non ordonnés.

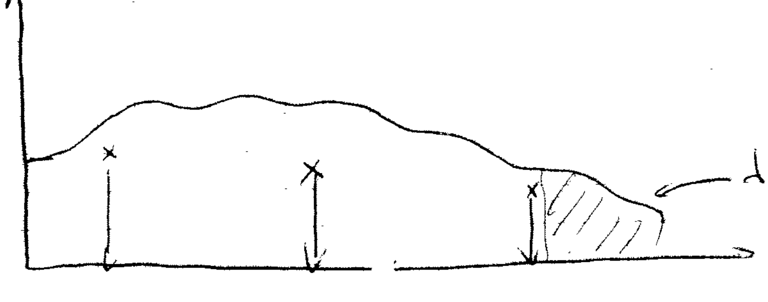
Donc on veut $\forall n \in \mathbb{N}$ quantifier la probabilité d'avoir



$P($ 0 point entre (0 et t_1), 1 point entre (t_1, t_1+dt_1), 0 point entre t_1+dt_1 et t_2 ..... 1 point entre (t_n, t_n+dt_n), 0 point entre t_n+dt_n et T)

$\Longrightarrow$ Thinning (cas le passé = le passé du processus) ∪ les points < t

cf PP(1) dans $\mathbb{R}_+^2$.



$$= E\left(P(\ldots \mid \mathcal{F}_{t_n+dt_n}) \right)$$

$\lambda(t)$ $$= E\left(E\left(\mathbb{1}_{0pt \ldots} \mathbb{1}_{0pt \text{ entre } t_n+dt_n} \mid \mathcal{F}_{t_n+dt_n} \right) \right)$$

$$= E\left(\mathbb{1}_{0pt} \ldots \mathbb{1}_{1pt \text{ entre } t_n \text{ et } t_n+dt_n} \; P(0pt \text{ entre } t_n+dt_n \text{ et } t \mid \right.$$

si pas de point je sais construire $\lambda(t)$ sur t_n+dt_n, t qui ne dépend que des $t_1 \ldots t_n$

et si pas de point c'est que personne pour le processus de Poisson en question

aire de

ie $P(0 \text{ pt entre } t_n+dt_n \text{ et } T \mid \mathcal{F}_{t_n+dt_n}) = e^{-\int_{t_n+dt_n}^{T} \lambda(u)\,du}$.

on recommence il sort

$$P(1 \text{ pt entre } t_n \text{ et } t_n+dt_n \mid \mathcal{F}_{t_n^-}) = \lambda(t_n)\,dt_n .$$

(cf def intensité)

A la fin on obtient que la vraisemblance de l'ensemble $\{t_1 \dots t_n\}$ est

$$\lambda(t_1) \dots \lambda(t_n)\, e^{-\int_0^T \lambda(u)\,du}$$

$$= \boxed{\exp\left[\int_0^T (\ln \lambda(u))\,dN_u - \int_0^T \lambda(u)\,du\right]}$$

Rmq Pb vraisemblance partielle relative à quelle filtration choisie...

2) Quelques calculs de vraisemblance

. Cas censure à droite : n va iid $X_1 \dots X_n = T_i \wedge C_i$ où C_i censure

$\delta_1 \dots \delta_n$ $\delta_i = 1_{T_i \le C_i}$.

. $N_t = \sum_{i=1}^{n} \delta_i 1_{X_i \le t}$

. $\lambda(t) = \left(\sum_{i=1}^{n} 1_{X_i \ge t}\right) q(t)$

$\uparrow$ Y_t processus prévisible observable

d'un pt de vue stat → Intensité mult. d'Aalen $\lambda(t) = Y_t\,\alpha(t)$ α déter, Y prév obs

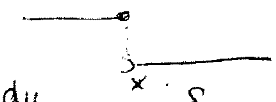
Rmq : Y_t le nb de personnes restant dans l'étude

la log vraisemblance :

$$\int_0^T \ln(\lambda(u))\,dN_u - \int_0^T \lambda(u)\,du$$

nb de personnes restant dans l'étude X_i compris au tps x.

$$= \sum_{i=1}^{n} \delta_i \ln(Y_{X_i}\, q(X_i)) - \int_0^T Y_u\, q(u)\,du .$$



Pour 1 seul pt, Y_{X_i} vaut toujours 1

donc la vraisemblance $q(x)^{\delta}\, e^{-\int_0^x q(u)\,du} = q(x)^{\delta}\, S(x)$

où $S(t) = P(X > t)$.

et on peut vérifier que c'est bien la vraisemblance qu'on imagine

car si $\delta=1$ on a vu T et on obtient $\frac{f(X)}{S(X)} S(X) = f(X)$ densité de la variable X.

si $\delta=0$ on sait que $T \geq X$ et on obtient comme vraisemblance $S(X)$ "Proba que $T > X$".

• <u>Cas Poisson</u> : $\lambda(u)$ est déterministe. Vérifions qqch sur cet ex.

Pourquoi moins la log vraisemblance serait elle un contraste ?

$$L(\lambda) = \int_0^T \ln(\lambda(u))\, dN_u - \int_0^T \lambda(u)\, du$$

$$E_{\lambda_0}(L(\lambda)) = \int_0^T \ln(\lambda(u))\, \lambda_0(u)\, du - \int_0^T \lambda(u)\, du$$

$$= \int_0^T \lambda(u) \left[\frac{\lambda_0(u)}{\lambda(u)} - \frac{\lambda_0}{\lambda} \ln\left(\frac{\lambda_0(u)}{\lambda(u)}\right) - 1 \right]$$

$$- \int_0^T \lambda_0(u)\, du + \int_0^T \ln(\lambda_0(u))\, \lambda_0(u)\, du.$$

or la fonction $x - x\ln x - 1$ est ≤ 0 sur $\mathbb{R}_+$ et vaut 0 ssi $x=1$ (dérivée $[1 - \ln x - 1] = -\ln x$)

donc ~~la log vrais~~ $E_{\lambda_0}(L(\lambda))$ est bien maximal quand $\lambda=\lambda_0$

et moins la log vrais. c'est donc un <u>contraste</u>.

<u>Remarque</u> :

$$K(PP(f), PP(s)) = E_f\left(\ln \frac{dP_f}{dP_s}\right) = E_f\left(L(f) - L(s)\right)$$

$$= E_f\left(\int_0^T \ln f\, dN_u - \int_0^T f\, du - \int_0^T \ln s\, dN_u + \int_0^T s\, du \right)$$

$$= \int_0^T \ln\left(\frac{f}{s}\right) f\, du - \int_0^T f\, du + \int_0^T s\, du)$$

$$= \int_0^T f \left[\frac{s}{f} - \ln\left(\frac{s}{f}\right) - 1 \right] = \int_0^T f\, \phi\left(\ln \frac{s}{f}\right)$$

$K(\ ) \geq 0$ et nul ssi

where $\phi(u) = e^u - u - 1$ minimal qd $u=0$ ie $s=f$ ps.

3) Et pour un processus de comptage général ?

Supposons que l'on observe un N Processus de comptage d'intensité $\lambda_0^{\theta_0}$. On propose une intensité candidate (donc prévisible) λ.

$$L(\lambda) = \int_0^T \ln(\lambda(t))\, dN_t - \int_0^T \lambda(t)\, dt .$$

paramétrée qd même par une déter

Alors $L(\lambda)$ a la même espérance que (passage au compensateur)

$$\int_0^T \ln(\lambda(t))\, \lambda_0(t)\, dt - \int_0^T \lambda(t)\, dt .$$

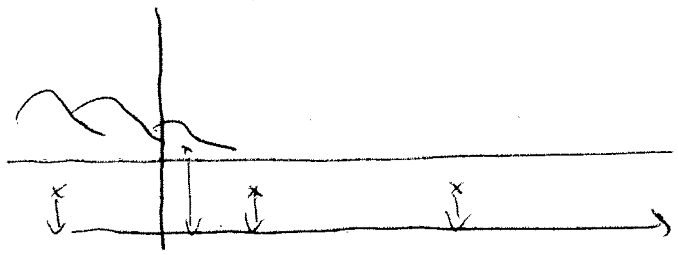
donc
$$E_{\theta_0}(L(\lambda)) = E_{\theta_0}\left[\int_0^T \ln(\lambda(t))\lambda_0(t)\,dt - \int_0^T \lambda(t)\,dt \right.$$
$$= E_{\theta_0}\left[\int_0^T \lambda \underbrace{\left[\frac{\lambda_0}{\lambda} \ln \frac{\lambda_0}{\lambda} - 1\right]}_{} \right]$$

≤ 0 et nul ssi $\lambda_0 = \lambda$ p.s (en ω et en t)

... On est donc caractérisé par l'intensité !.

Rmq de la même façon on définit la Kullback
$$K(N(\lambda), N(\lambda_0)) = E_\theta \left(L(\lambda) - L(\lambda_0) \right) \dashrightarrow \text{etc} \ldots$$

Attention tout de même : il faut pour que tout marche bien qu'on n'ait pas besoin de variables éventuellement non observables pour calculer l'intensité. Typiquement les ancêtres avant 0 du Hawkes :



la constructi
vrai ou l

ou
$$\lambda(t) = \nu + \sum_{T<t} h(t-T)$$

en 0 $\quad \lambda(0) = \nu + \sum_{T<0} h(t-T)$

soit pas d'ancêtres avant 0, et tout va bien

soit (par ex en régime stationnaire) et là il faut faire attention.

3) Et après qu'en fait-on ?

$\longrightarrow$ Calcul de l'estimateur par max de vraisemblance

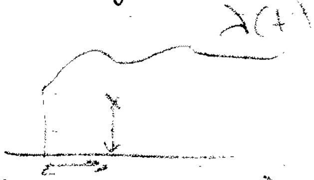
Propriétés : - ce qu'on attend CV, normalité asymptotique et

Le plus gros pb : c'est quand quoi tend vers ∞ ?

2 possibilités en général $\longrightarrow$ on a un échantillon iid de Proc (pex n patients) . $n \to +\infty$

$\longrightarrow$ on observe sur un temps long. $T \to +\infty$.

Comment savoir ? hypothèses (+/- atroces) à vérifier au cas par (Cf livre d'Andersen Borgan Gill et Keiding)



4) Goodness of fit ?

Basé +/- sur la réciproque du thinning $P(T_1 > t) = e^{-\int_0^t \lambda(u)du} = e^{-\Lambda(t)}$ $= P(\Lambda^{-1}(E_1) > t)$, $E_1 \sim \mathcal{E}(1)$, $P(E_1 > \Lambda(t)) = e^{-\Lambda(t)}$

Thm Soit N un processus de comptage.
Sous des hypothèses appropriées ($\lambda(t)$ intensité existe >0, prévisible, bornée ....)
Alors $N_{\Lambda^{-1}(t)}$ a pour loi un processus de Poisson homogène d'intensité 1

donc $\Lambda^{-1}(E_1) \sim T_1$ ie $E_1 \sim \Lambda(T_1)$

"Preuve" : $\Lambda(t) = \int_0^t \lambda(u)\,du$ f° strictement croissante donc inversible

soit $\mathcal{N}_t = N_{\Lambda^{-1}(t)}$ le passé de $\mathcal{N}$ à l'instant t

donc $E(d\mathcal{N}_t \mid \mathcal{G}_{t-})$ c'est toujours les mêmes points avec un changement de

$= E\left(dN_{\Lambda^{-1}(t)} \mid \mathcal{F}_{\Lambda^{-1}(t)}\right)$

$$= \lambda(\Lambda^{-1}(t))\, \underbrace{d\text{"}\Lambda^{-1}(t)\text{"}}_{\substack{\text{attention ici} \\ \text{l'élément de temps} \\ \text{passe plus ou moins vite} \\ \text{en fonction de } \Lambda^{-1}}} = \frac{\lambda(\Lambda^{-1}(t))}{\lambda(\Lambda^{-1}(t))}\, dt = dt$$

Remarque les sauts de $N_{\Lambda^{-1}(t)}$ sont les $\Lambda(T_i)$ où T_i sont les sauts/p de N.

Et ensuite il reste donc à tester que $\{\Lambda(T_i)\}$ forment un PP(1)
Pex (A la Kolmogorov Smirnov), on teste que
$\{\Lambda(T_1), \ldots \Lambda(T_{N_T})\}$ forment un N_T échantillon uniforme pour test de KS
(ou autre variante avec approx PP(1) - Brownien).
cf Daley Vere Jones.

Pb: (1) Λ est fixé donc on n'a pas le droit officiellement à $\Lambda_{\hat{\theta}}$ où $\hat{\theta}$ viendrait d'un max de vraisemblance pex
Ce qui n'empêche pas plein de gens de l'utiliser quand même dans ce cas même si il y a eu des contreex.

(2) C'est KS derrière $\rightarrow$ pas très puissant selon les alternatives.