LE PROCESSUS DE POISSON [1]

Préparation à l'agrégation externe de Mathématiques de l'université Rennes 1
Année 2008/2009

1. LE PROCESSUS DE POISSON SIMPLE
[Réf. : toutes]

A titre d'exemple, considérons les phénomènes suivants : émission de particules radioactives, appels dans un central téléphonique, ou bien arrivées de clients devant un guichet. En terme de modélisation, ce qui caractérise ces phénomènes -considérés comme aléatoires-, c'est une répartition dans le temps d'instants aléatoires où se produisent certains événements spécifiques. Un premier modèle est fournit par la famille des *processus de comptage* :

Définition 1.1 *Soit $(X_t)_{t\geq 0}$ un processus stochastique à valeurs réelles. On dit que $(X_t)_{t\geq 0}$ est un processus de comptage si, pour $\mathbb{P}$-p.t. $\omega \in \Omega$, la trajectoire $t \mapsto X_t(\omega)$ est croissante par sauts d'amplitude 1, continue à droite et telle que $X_0(\omega) = 0$.*

Par exemple, X_t représente le nombre de clients arrivés devant un guichet donné dans l'intervalle de temps $[0, t]$. Une telle famille de modèles est en fait beaucoup trop générale pour pouvoir prétendre être étudiée. Dans les 3 exemples présentés ci-dessus, on peut imposer des hypothèses supplémentaires, qui restent compatibles avec une modélisation raisonnable, et qui permettront au modélisateur de fournir des réponses quantitatives.

Définition 1.2 *Un processus de comptage $(N_t)_{t\geq 0}$ est appelé processus de Poisson simple si :*
(i) pour tous $s, t \geq 0$, $N_{t+s} - N_s \perp\!\!\!\perp \sigma(N_u, u \leq s)$; [accroissements indépendants]
(ii) pour tous $s, t \geq 0$, $N_{t+s} - N_s \sim N_t$. [stationnarité]

Modèles.

- Guichet. Ici, N_t représente le nombre de clients qui sont arrivés au guichet avant l'instant t. L'hypothèse sur les sauts d'amplitude 1 exprime le fait que les clients arrivent un par un au guichet. En revanche, les hypothèses (i) et (ii), qui posent des conditions sur $N_{t+s} - N_s$ (le nombre de clients arrivés au guichet dans l'intervalle de temps $]s, t+s]$), sont plus discutables. Malgré cela, une telle modélisation est une approximation raisonnable de la réalité, qui a en plus la vertu de pouvoir donner des solutions quantitatives simples.
- Désintégration de l'uranium. Le processus de Poisson modélise de manière très convenable les émissions radioactives de l'uranium 235 : l'observation de son processus de désintégration -très lent- montre qu'il est stationnaire et à accroissements indépendants.

Existence du processus de Poisson. Soit $(D_n)_{n\geq 1}$ une suite de v.a.i.i.d. de loi $\mathcal{E}(\lambda)$. Le processus défini pour chaque $t \geq 0$ par

$$\sum_{n\geq 1} \mathbf{1}_{\{D_1+\cdots+D_n\leq t\}}, \quad t \geq 0. \qquad (\star)$$

est un processus de Poisson simple. C'est d'ailleurs une autre définition du processus de Poisson simple.

2. LOI D'UN PROCESSUS DE POISSON ET DE SES INTER-ARRIVEES
[Réf. : toutes]

Expliquons rapidemment de quelle manière on peut retrouver la loi marginale du processus de Poisson simple $(N_t)_{t\geq 0}$, lorsque N_t s'exprime par $(\star)$. Comme $S_n := D_1 + \cdots + D_n \sim \gamma(n, \lambda)$ et $\{S_n \leq t\} = \{N_t \geq n\}$, on a :

$$\mathbb{P}(N_t = n) = \mathbb{P}(S_n \leq t) - \mathbb{P}(S_{n+1} \leq t) = \exp(-\lambda t)\frac{(\lambda t)^n}{n!},$$

i.e. $N_t \sim \mathcal{P}(\lambda t)$. Avec la définition 1.2, on peut établir (plus difficile) :

Théorème 2.1 *Soit $(N_t)_{t\geq 0}$ est un processus de Poisson simple. Il existe $\lambda \geq 0$ tel que pour chaque $t \geq 0$, $N_t \sim \mathcal{P}(\lambda t)$. Le paramètre λ, appelé intensité du processus de Poisson, le caractérise entièrement.*

[1] Benoît Cadre - ENS Cachan Bretagne

Remarques 2.1

(a) Le caractère "simple" de ce processus de Poisson tient essentiellement au fait qu'il est stationnaire, une propriété dont on a vu les limites en matière de modélisation. De manière plus générale, plutôt que l'hypothèse de stationnarité, on suppose que $N_{t+s} - N_s \sim \mathcal{P}\big(\int_s^{t+s} \lambda(u)du\big)$, où $\lambda(.)$ est localement intégrable et strictement positive. La fonction $m(t) = \int_0^t \lambda(u)du$ est alors inversible, et le processus $(N_{m^{-1}(t)})_{t\geq 0}$ est un processus de comptage, qui est de surcroît à accroissements indépendants et stationnaires : c'est donc un processus de Poisson simple (d'intensité 1).

(b) Le processus $(N_t)_{t\in\mathbb{N}}$ est une chaîne de Markov homogène d'espace d'états $\mathbb{N}$ et matrice de transition $\mathcal{P} = (p_{ij})_{i,j\in\mathbb{N}}$, où $p_{ij} = 0$ si $j < i$ et, dans le cas $j \geq i$:

$$p_{ij} = \exp(-\lambda)\frac{\lambda^{j-i}}{(j-i)!}.$$

De plus, $(N_t)_{t\geq 0}$ est une sous-martingale, et $(N_t - \lambda t)_{t\geq 0}$ est une martingale. En effet, pour tous $0 \leq s \leq t$:

$$\mathbb{E}(N_t - \lambda t | N_u, u \leq s) \quad = \quad \mathbb{E}(N_t - N_s | N_u, u \leq s) + N_s - \lambda t = \mathbb{E}(N_{t-s}) + N_s - \lambda t = N_s - \lambda s.$$

Si $0 = T_0 < T_1 < T_2 < ...$ sont les instants de sauts du processus de Poisson simple $(N_t)_{t\geq 0}$:

$$N_t = \sum_{n\geq 1} \mathbf{1}_{\{T_n \leq t\}}, \quad t \geq 0.$$

Théorème 2.2 *Les instants d'inter-arrivées $(T_n - T_{n-1})_{n\geq 1}$ du processus de Poisson simple d'intensité λ sont des v.a.r. indépendantes et de même loi $\mathcal{E}(\lambda)$. De plus, $(T_1, \cdots, T_n)$ possède une densité f_n définie par*

$$f_n(t_1, \cdots, t_n) = \begin{cases} \lambda^n \exp(-\lambda t_n) & \textit{si } 0 < t_1 < \cdots < t_n \,; \\ 0 & \textit{sinon.} \end{cases}$$

Preuve Cas $n = 2$. Soient $0 \leq s_1 < t_1 < s_2 < t_2$ et $A =]s_1, t_1] \times]s_2, t_2]$. En utilisant les propriétés de stationnarité et d'indépendance des accroissements d'un processus de Poisson, on obtient successivement :

$$\begin{aligned} \mathbb{P}((T_1, T_2) \in A) &= \mathbb{P}(N_{s_1} = 0, N_{t_1} - N_{s_1} = 1, N_{s_2} - N_{t_1} = 0, N_{t_2} - N_{s_2} \geq 1) \\ &= \mathbb{P}(N_{s_1} = 0)\mathbb{P}(N_{t_1 - s_1} = 1)\mathbb{P}(N_{s_2 - t_1} = 0)\mathbb{P}(N_{t_2 - s_2} \geq 1). \end{aligned}$$

Avec le théorème 2.1, cela donne finalement

$$\mathbb{P}((T_1, T_2) \in A) = \int_A \lambda^2 \exp(-\lambda x_2) dx_1 dx_2.$$

Notons $\mathcal{A}$ la famille des pavés du type $A =]s_1, t_1] \times]s_2, t_2]$ avec $t_1 < s_2$ et $G = \{(x_1, x_2) \,:\, 0 < x_1 < x_2\}$. Alors, $\mathcal{A}$ est un π-système et $\sigma(\mathcal{A}) = \mathcal{B}(G)$ donc, d'après le théorème de Dynkin, la formule ci-dessus est vraie pour tout $A \in \mathcal{B}(G)$. Les lois de T_1 et de $T_2 - T_1$ s'en déduisent aussitôt, de même que leur indépendance. •

File d'attente $M/M/1$ et processus de Poisson. Dans la cadre d'une file d'attente $M/M/1$, la loi des inter-arrivées est $\mathcal{E}(\lambda)$, et celle des temps de service est $\mathcal{E}(\mu)$. Le processus d'arrivée des clients au serveur est donc un processus de Poisson simple de paramètre λ. De plus, en régime stationnaire, le processus de sortie du système est aussi un processus de Poisson simple d'intensité λ (la loi stationnaire du processus décrivant l'évolution de la taille du système à l'instant t, qui existe lorsque $\rho := \lambda/\mu < 1$, est $(1-\rho)(1, \rho, \rho^2, \cdots)$).

Amnésie de la loi exponentielle. Dans le cadre d'une modélisation des arrivées de clients à un guichet, T_n représente l'instant d'arrivée du client n au guichet, et $(T_n - T_{n-1})$ représente le temps qui s'est écoulé entre les arrivées du $(n-1)$-ème et du n-ème client au guichet. La loi exponentielle des instants d'inter-arrivées des clients au guichet est héritée notamment de la propriété d'indépendance des accroissements du processus de Poisson. Rien d'étonnant à cela : l'indépendance des accroissements du processus de Poisson traduit un comportement "amnésique" des clients, et l'amnésie est précisémment ce qui caractérise la loi exponentielle. Rappelons en effet qu'une v.a.r. Z possédant une densité suit une loi exponentielle si, et seulement si, pour tous $x, y \geq 0$:

$$\mathbb{P}(Z > x + y | Z > x) = \mathbb{P}(Z > y).$$

Autrement dit -une fonction de répartition caractérisant la loi-, la loi exponentielle se caractérise par son absence de mémoire :

$$\forall x \geq 0 \,:\, \mathcal{L}(Z - x | Z > x) = \mathcal{L}(Z).$$

Cette propriété traduit bien le comportement du temps d'arrivée du prochain client dans une file d'attente (mais aussi de la durée de vie des ampoules, ...). Si $Z_1, \cdots, Z_n$ sont des v.a.r. indépendantes de lois exponentielles de paramètres $\lambda_1, \cdots, \lambda_n$, on montre que $\min(Z_1, \cdots, Z_n) \sim \mathcal{E}(\lambda_1 + \cdots + \lambda_n)$ et $I\!P(Z_i = \min(Z_1, \cdots, Z_n)) = \lambda_i/(\lambda_1 + \cdots + \lambda_n)$. Ainsi, dans le cadre d'une modélisation des arrivées des clients au guichet par un processus de Poisson simple d'intensité λ, la probabilité que le temps écoulé entre l'arrivée du $(i-1)$-ème et du i-ème client soit la plus petite parmis les inter-arrivées des n premiers clients est indépendante de λ, et vaut $1/n$. De plus, parmis les n premiers clients, le temps minimum entre l'arrivée de 2 clients consécutifs suit une loi $\mathcal{E}(n\lambda)$.

3. ESTIMATION DE L'INTENSITE
[Réf. : Foata et Fuchs]

Reprenons le modèle poissonnien des clients arrivant à une caisse. Soit $(N_t)_{t\geq 0}$ un processus de Poisson simple, dont l'intensité $\lambda > 0$ est donc le seul paramètre du modèle. En pratique, λ est inconnu et afin de connaître entièrement son modèle, le gérant du magasin doit donner une valeur pour λ. Dès lors, comment estimer λ ? Comment construire un intervalle de confiance pour λ ? Quel type de test statistique utiliser ?

3.1 Cas où le processus est observé jusqu'à un instant t

On peut écrire N_t comme une somme de v.a.r. indépendantes (et majoritairement de même loi) : $N_t = N_t - N_{[t]} + \sum_{i=0}^{[t]-1}(N_{i+1} - N_i)$. On peut alors montrer le résultat suivant :

Théorème 3.1 *Lorsque $t \to \infty$:*

$$\frac{N_t}{t} \overset{\text{p.s.}}{\to} \lambda \quad et \quad \sqrt{\frac{t}{\lambda}}\Big(\frac{N_t}{t} - \lambda\Big) \overset{\mathcal{L}}{\to} N(0,1).$$

Si le processus de Poisson a été observé jusqu'à l'instant t (suffisamment grand), l'estimateur naturel de λ est donc N_t/t, et il l'estime sans biais. En pratique, il suffira au gérant du magasin de compter le nombre de clients qui arrivent à la caisse avant un instant t suffisamment grand pour en déduire une estimation de λ. La construction de l'intervalle de confiance asymptotique (ou le test statistique) pour λ est basé sur la partie (ii) du théorème précédent. Notons $u_{1-\alpha/2}$ le quantile d'ordre $1-\alpha/2$ de la loi $N(0,1)$. Un intervalle de confiance (asymptotique) pour λ au niveau $1-\alpha$ est

$$\Big[- u_{1-\alpha/2}\frac{\sqrt{N_t}}{t} + \frac{N_t}{t},\, u_{1-\alpha/2}\frac{\sqrt{N_t}}{t} + \frac{N_t}{t}\Big].$$

Remarque 3.1 Cet estimateur de l'intensité est aussi l'estimateur du maximum de vraisemblance : ayant observé le processus jusqu'à l'instant t, on dispose d'une part du nombre de sauts n et d'autre part des instants de sauts $0 < t_1 < \cdots < t_n \leq t$ de la trajectoire de $(N_t)_{t\geq 0}$. La vraisemblance de ces observations est

$$L(n, t_1, \cdots, t_n; \lambda) = f(t_1, \cdots, t_n) I\!P(N_t = n),$$

où f désigne la densité de $\mathcal{L}(T_1, \cdots, T_n | N_t = n)$. On vérifie que $f(u_1, \cdots, u_n) = n!/t^n \mathbf{1}_{\{0<u_1<\cdots<u_n\leq t\}}$: en effet, pour $n = 2$, on a d'après le théorème 2.2, pour tous $0 < s_1 < u_1 < s_2 < u_2 < t$,

$$I\!P\Big(s_1 \leq T_1 \leq u_1, s_2 \leq T_2 \leq u_2 | N_t = 2\Big) = \frac{I\!P\Big(s_1 \leq T_1 \leq u_1, s_2 \leq T_2 \leq u_2, T_3 > t\Big)}{I\!P(N_t = 2)} = \frac{2!}{t^2}(u_1 - s_1)(u_2 - s_2).$$

On conclut, comme dans la preuve du théorème 2.2, en utilisant le théorème de Dynkin. La vraisemblance s'écrit donc :

$$L(n, t_1, \cdots, t_n; \lambda) = \exp(-\lambda t)\frac{(\lambda t)^n}{n!}\frac{n!}{t^n} = \lambda^n \exp(-\lambda t).$$

La valeur qui maximise cette expression est n/t, i.e. l'estimateur du maximum de vraisemblance de λ est N_t/t.

3.2 Cas où le processus est observé jusqu'à son n-ième saut

Notons $0 < t_1 < \cdots < t_n$ les instants de saut observés. La vraisemblance de ces observations est

$$L(t_1, \cdots, t_n; \lambda) = \lambda^n \exp(-\lambda t_n).$$

La valeur qui maximise cette expression est n/t_n, et l'estimateur du maximum de vraisemblance est donc n/T_n. En écrivant T_n sous la forme d'une somme de v.a.i.i.d. : $T_n = \sum_{i=1}^{n-1}(T_{i+1} - T_i)$, on démontre le résultat suivant, qui, tout comme le théorème précédent, permet de construire des intervalles de confiance et des tests statistiques pour la valeur de λ.

Théorème 3.2 *Lorsque $n \to \infty$:*

$$\frac{n}{T_n} \xrightarrow{\text{p.s.}} \lambda \quad et \quad \sqrt{n}\Big(\lambda\frac{T_n}{n} - 1\Big) \xrightarrow{\mathcal{L}} N(0,1).$$

4. UN PROCESSUS DE NAISSANCE ET DE MORT
[Réf. : Dacunha-Castelle et Duflo, Foata et Fuchs, Grimmett et Stirzacker]

On veut décrire l'évolution, en fonction du temps, de la taille d'une population.

Modélisation.
- Chaque saut du processus de Poisson $(N_t)_{t\geq 0}$ d'intensité λ est interprété comme étant la naissance d'un individu ;
- Chaque individu a une durée de vie de fonction de répartition F, et les durées de vie des individus sont des v.a.i.i.d. et indépendantes du processus $(N_t)_{t\geq 0}$.

Soient T_k la date de naissance de l'individu k et X_k sa durée de vie. Le nombre d'individus en vie à l'instant $t \geq 0$, noté $Q(t)$, vaut alors (avec la convention habituelle " $\sum_{k=1}^{0} = 0$") :

$$Q(t) = \sum_{k=1}^{N_t} \mathbf{1}_{\{T_k + X_k \geq t\}}.$$

Loi de $Q(t)$. Pour tout $u \in I\!R$:

$$\varphi_{Q(t)}(u) := I\!E\Big(\exp(iuQ(t))\Big) = \sum_{n\geq 0} I\!P(N_t = n) I\!E\Big(\exp(iu\sum_{k=1}^{n}\mathbf{1}_{\{T_k+X_k\geq t\}})\Big|N_t = n\Big).$$

Or, la loi conditionnelle de $(T_1, \cdots, T_n)$ sachant $N_t = n$ a pour densité $n!/t^n \mathbf{1}_{\{0<t_1<\cdots<t_n<t\}}$ (cf. Section 3.1). Ainsi, $\mathcal{L}(T_1, \cdots, T_n|N_t = n) = \mathcal{L}(U_{(1)}, \cdots, U_{(n)})$, où $U_1, \cdots, U_n$ sont des v.a.i. de même loi $\mathcal{U}([0,t])$. De l'égalité en loi

$$\sum_{k=1}^{n} \mathbf{1}_{\{U_{(k)}+X_k\geq t\}} \sim \sum_{k=1}^{n} \mathbf{1}_{\{U_k+X_k\geq t\}},$$

on déduit que :

$$\varphi_{Q(t)}(u) = \sum_{n\geq 0} I\!P(N_t = n)\Big(I\!E\Big(\exp(iu\mathbf{1}_{\{U_1+X_1\geq t\}})\Big)\Big)^n.$$

De plus,

$$I\!P(U_1 + X_1 \geq t) = \frac{1}{t}\int_0^t I\!P(U_1 + X_1 \geq t|U_1 = s)ds = \frac{1}{t}\int_0^t (1 - F(u))du.$$

On conclut à l'aide de ces observations que $Q(t) \sim \mathcal{P}\big(\lambda \int_0^t (1 - F(u))du\big)$. Dans cette modélisation, on a ainsi, lorsque $t \to \infty$: $Q(t) \xrightarrow{\mathcal{L}} \mathcal{P}(\lambda I\!E(X))$. Cette conclusion est-elle conforme à la réalité ? Quelles simulations proposer ?

REFERENCES

- P. Billingsley, *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics, 1995.
- D. Dacunha-Castelle et M. Duflo, *Probabilités et statistiques - Tome 2 : Problèmes à temps mobile (Cours et Exercices)*, Masson, 1993.
- D. Foata et A. Fuchs, *Processus stochastiques - Processus de Poisson, chaînes de Markov et martingales*, Dunod, 2002.
- G.R. Grimmett et D. Stirzaker *Probability and Random Processes*. Oxford Science Publications, 1992.
- J.-Y. Ouvrard, *Probabilités 2 : maîtrise agrégation*, Cassini, 2001.