Processus de naissance et de Poisson

Olivier Scaillet

University of Geneva and Swiss Finance Institute

Outline

Introduction

Beaucoup de phénomènes naturels peuvent présenter des changements de valeurs à n'importe quel moment plutôt qu'à des dates fixes.

On aura besoin pour modéliser cela de processus en temps continu c.à.d. des processus $\{X_t : t \geq 0\}$ indicés par la demidroite positive $[0, +\infty)$.

Les processus que l'on analyse ici sont à valeurs entières.

Processus de Poisson

Le processus de Poisson sert à modéliser l'occurrence d'évènements successifs.

Chaque évènement est tel que dans un intervalle de temps $(t, t + \Delta t)$ avec Δt petit, sa probabilité d'occurrence est proportionnelle à Δt , et que la probabilité d'occurrence de deux évènements dans l'intervalle est négligeable.

Processus de Poisson (cont'd)

Un processus de Poisson d'intensité λ est un processus $N = N_t : t \geq 0$ à valeurs dans $0, 1, 2, 3, ...$ tel que

1. $N_0 = 0$ et $N_s \leq N_t$, si $s < t$
2. si $s < t$, le nombre $N_t - N_s$ d'évènements dans l'intervalle de temps $(s, t]$ est indépendant des temps d'arrivée et des arrivées dans l'intervalle$[0, s]$ (processus à accroissements indépendants)
3. $$P[N_{t+\Delta t} = n + i | N_t = n] = \begin{cases} \lambda \Delta t + o(\Delta t) & si \;\; i = 1 \\ o(\Delta t) & si \;\; i > 1 \\ 1 - \lambda \Delta t + o(\Delta t) & si \;\; i = 0 \end{cases}$$

avec la notation $f(\Delta t) = o(\Delta t)$ signifiant que $\lim_{\Delta t \to 0} \frac{f(\Delta t)}{\Delta t} = 0$.

Processus de Poisson (cont'd)

La fonction f est d'un plus petit ordre asymptotique que Δt, et est négligeable par rapport à Δt.

On parle de N_t comme le nombre d'arrivées ou d'occurrences ou d'évènements du processus jusqu'en t.

Le processus $\{N_t\}$ est un *processus de comptage*, et est un exemple simple de chaîne de Markov en temps continu.

L'intensité λ (ou taux de naissance) correspond à la probabilité d'occurrence d'un seul évènement par unité de temps.

Le processus N_t suit une loi de Poisson de paramètre λt, $P[N_t = i] = \frac{(\lambda t)^i}{i!} e^{-\lambda t}$, $i = 0, 1, 2, \ldots$. On a donc $E[N_t] = \lambda t$ en particulier.

Processus de Poisson (cont'd)

On peut également caractériser un processus de Poisson à l'aide du comportement des durées entre évènements, c.à.d. l'intervalle de temps s'étant écoulé entre deux évènements successifs.

Soit $T_0, T_1, \ldots$ donnés par $T_0 = 0$, $T_n = \inf\{t : N_t = n\}$. T_n est la date de la $n$ème arrivée.

Les durées ou intervalles de temps sont les variables aléatoires $d_n = T_n - T_{n-1}$. Ces variables $d_1, d_2, \ldots$ sont indépendantes de loi exponentielle de paramètre λ caractérisée par la fonction de densité $f(d) = \lambda e^{-\lambda d}$.

La fonction de répartition est $F(d) = 1 - e^{-\lambda d}$, et l'espérance des durées est $E[d_t] = \frac{1}{\lambda}$.

Processus de Poisson (cont'd)

Remarques:

- connaissant les durées d, on peut reconstruire N à l'aide de $T_n = \sum_{i=1}^{n} d_i$, $N_t = \max\{n : T_n \leq t\}$.
- l'estimation de lambda se fait aisément puisque l'estimateur du maximum de vraisemblance pour le paramètre d'une loi exponentielle correspond à l'inverse de la moyenne empirique: $\hat{\lambda} = \frac{1}{\bar{d}}$, où $\bar{d}$ est la moyenne empirique des durées observées.

Processus de naissance

Le processus de poisson a un taux de naissance (ou intensité) constant. On peut introduire une dépendance de ce taux de naissance en fonction du nombre d'évènements déjà survenus.

Un *processus de naissance* d'intensités $\lambda_0, \lambda_1, \lambda_2, \ldots$ est un processus $N = \{N_t : t \geq 0\}$ à valeurs dans $\{0, 1, 2, 3, \ldots\}$ tel que

1. $N_0 = 0$ et $N_s \leq N_t$, si $s < t$
2. si $s < t$, conditionnellement à la valeur de N_s, le nombre $N_t - N_s$ d'événements dans l'intervalle de temps $(s, t]$ est indépendant de toutes les arrivées dans l'intervalle $[0, s]$
3. $P[N_{t+\Delta t} = n + i | N_t = n] = \begin{cases} \lambda_n \Delta t + o(\Delta t) & si \;\; i = 1 \\ o(\Delta t) & si \;\; i > 1 \\ 1 - \lambda_n \Delta t + o(\Delta t) & si \;\; i = 0 \end{cases}$

Processus de naissance (cont'd)

Cas particuliers:

- Processus de Poisson: $\lambda_n = \lambda, \forall n$
- Naissance simple: $\lambda_n = n\lambda$. Croissance d'une population dans laquelle chaque individu donne naissance indépendamment des autres, chacun donnant naissance à un nouvel individu avec probabilité $\lambda \Delta t + o(\Delta t)$ dans l'intervalle de temps $(t, t + \Delta t)$ et aucun individu ne pouvant mourir.
- Naissance avec immigration: $\lambda_n = n\lambda + v$. Processus de naissance où le taux de naissance dépend d'un taux d'immigration externe.

Processus de naissance et mort

Une modélisation plus réaliste qu'un processus de naissance pur dans le cadre de la croissance d'une population inclut la possibilité de décès.

Soit N_t le nombre d'individus en vie à la date t dans une certaine population. Ce nombre évoluera comme un *processus de naissance et de mort* si

1. N_t est une chaîne de Markov à valeurs dans $\{0, 1, 2, \dots\}$
2. $P[N_{t+\Delta t} = n + i | N_t = n] =$
$$\begin{cases} \lambda_n \Delta t + o(\Delta t) & si \;\; i = 1 \\ \mu_n \Delta t + o(\Delta t) & si \;\; i = -1 \\ o(\Delta t) & si \;\; |i| > 1 \\ 1 - \lambda_n \Delta t - \mu_n \Delta t + o(\Delta t) & si \;\; i = 0 \end{cases}$$
3. les taux de naissance $\lambda_0, \lambda_1, \lambda_2, \dots$ et de mort $\mu_0, \mu_1, \mu_2, \dots$ sont tels que $\lambda_i \geq 0$, $\mu_i \geq 0$, et $\mu_0 = 0$.

Processus de naissance et mort (cont'd)

Cas particuliers:

- Processus de naissance pour: $\mu_n = 0, \forall n$
- Mort simple avec immigration: $\lambda_n = \lambda$, $\mu_n = n\mu, \forall n$. Population ne se reproduisant pas mais bénéficiant d'un taux d'immigration correspondant à un processus de Poisson et dont chaque individu a une probabilité $\mu\Delta t + o(\Delta t)$ de décéder dans l'intervalle de temps $(t, t + \Delta t)$
- Naissance et mort simple: $\lambda_n = n\lambda$, $\mu_n = n\mu, \forall n$. Chaque individu en vie dans la population peut soit décéder dans l'intervalle de temps $(t, t + \Delta t)$ avec une probabilité $\mu\Delta t + o(\Delta t)$, soit se séparer en deux avec une probabilité $\lambda\Delta t + o(\Delta t)$.

Files d'attente

La théorie des files d'attente est un grand domaine d'application des processus aléatoires.

Exemples pratiques d'utilisation:

- E-commerce: connexion à un serveur
- online banking: capacité d'un call center
- agence bancaire: nombre de guichets

Flux d'arrivée

Une *file d'attente* est constituée d'un *flux d'arrivée* qui représente les instants où arrivent les “clients” (terme générique représentant aussi bien des pièces de fabrication, des véhicules, des appels à la mémoire d'un ordinateur, sollicitations de l'organe central d'un réseau téléinformatique, . . .).

En première approximation on considère souvent que les délais entre les arrivées sont des variables aléatoires indépendantes équidistribuées.

Le flux est alors ce qu'on appelle un *processus de renouvellement stationnaire*. Le cas le plus simple est celui où la loi commune est la loi exponentielle, le flux d'arrivée est alors un processus de Poisson.

Service

Une file d'attente est également constituée d'un organe de service qui est caractérisé par

- le temps de service: un client qui commence à être servi sera immobilisé pendant un temps aléatoire dont on supposera connue la loi
- le nombre de guichets

Service (cont'd)

Une file d'attente est aussi caractérisée par une *règle de service* ou discipline qui indique comment fonctionne le système :

- Système avec attente ou sans attente (dans un système sans attente il n'y a pas de queue, un client non servi à son arrivée est perdu). Système de capacité d'attente limitée (au delà d'une certaine longueur de queue les nouveaux clients sont perdus)
- Ordre dans lequel les clients sont servis: "premier arrivé, premier servi", "dernier arrivé, premier servi". (Cas usuel entre deux postes d'usinage les pièces issues du premier poste sont empilées les unes sur les autres puis saisies en commençant par celle qui est sur le dessus).
- Plusieurs classes de clients chacune prioritaire par rapport aux classes suivantes.
- Le client à son arrivée, si la queue est trop longue, quitte le système avec une certaine probabilité dépendant de la longueur de la queue.

Notations de Kendall

Pour leur étude mathématique, on classe les files d'attente selon une notation standard A / B / C / D. Les différentes lettres sont dédiées:

- A: au flux d'arrivée
- B: au temps de service
- C: au nombre de guichets
- D: aux infos supplémentaires avec les conventions suivantes.

La lettre M (comme Markov) désigne, pour le flux d'arrivées, un flux poissonien, pour les temps de service, un temps exponentiel.

La lettre D (comme déterministe) désigne pour le flux d'arrivées, des arrivées à intervalles réguliers, pour les temps de service, un temps fixe pour tous les clients.

La lettre G désigne le cas général.

Exemple: File M/M/1

C'est l'exemple de file le plus simple:

- les arrivées constituent un processus de Poisson d'intensité λ, autrement dit les délais entres les arrivées sont des variables indépendantes de même loi exponentielle de paramètre λ.
- le temps de service d'un client suit une loi exponentielle de paramètre μ, les temps de service des divers clients sont indépendants.
- Il n'y a qu'un seul guichet, les clients sont servis dans l'ordre d'arrivée la capacité de la queue est infinie.

Remarque: la file $M/M/1$ peut se voir comme un processus de naissance et mort avec des taux de naissance et mort donnés par $\lambda_n = \lambda$, $\forall n$ et $\mu_0 = 0$, $\mu_n = \mu$, $\forall n > 0$.

Exemple: File M/M/1 (cont'd)

Si le guichet est utilisé en permanence le débit moyen de l'organe de service est égal à μ. Comme le flux d'arrivées est en moyenne égal à λ par unité de temps, on aura une impossibilité matérielle si $\lambda > \mu$.

L'intensité du trafic est ici mesurée par le rapport $\rho = \frac{\lambda}{\mu}$, et le comportement de la chaîne peut être étudié via ce rapport.

Soit n_t le nombre de clients dans le système à l'instant t (clients dans la queue ou en train d'être servis),

- la probabilité qu'un client que l'on est en train de servir à l'instant t ait fini d'être servi à l'instant $t + \Delta t$, ne dépend pas de t (services exponentiels).
- la probabilité qu'un nouveau client arrive dans l'intervalle de temps $(t, t + \Delta t)$ ne dépend pas de t (arrivées exponentielles)

Exemple: File M/M/1 (cont'd)

On peut montrer que si $\rho < 1$ (éq. $\lambda < \mu$), c.à.d. si le débit moyen est strictement supérieur au débit maximum

- Le nombre aléatoire n_t de personnes dans le système est un processus récurrent : il peut prendre une infinité de fois toute valeur arbitrairement grande et une infinité de fois la valeur zéro.
- Un régime stationnaire s'établir dont la loi est donnée par $P[n_t = n] \to \pi_n = (1-\rho)\rho^n$ lorsque $t \to \infty$.

On en déduit que le nombre moyen de personnes dans le système est $E^\pi[n_t] = \sum_n n\pi_n = \frac{\rho}{1-\rho}$.

Exemple: File M/M/1 (cont'd)

Si $\rho > 1$ (éq. $\lambda > \mu$), c.à.d. si le débit moyen est strictement inférieur au débit maximum, la longueur de la queue tend vers l'infini.

Si $\rho = 1$ (éq. $\lambda = \mu$), c.à.d. si le débit moyen est égal au débit maximum, le nombre aléatoire n_t de personnes dans le système est un processus récurrent mais aucun régime stationnaire ne s'établit (fortes oscillations du nombre de personnes en attente).

Exemple: File M/M/1 (cont'd)

Remarques:

- on peut également caractériser les lois et moments de la longueur de la queue c.à.d. le nombre de personnes qui attendent sans être servies, les temps nécessaires de fermeture des arrivées tout en maintenant le service pour résorber la queue, ...
- On peut aussi analyser les files $M/M/k$, $M/M/\infty$, ... imposer des bornes sur les temps d'arrivée et d'attente, travailler avec des lots d'arrivées (batch), ...

Processus de risque

En assurance on utilise des processus de risque afin de déterminer les montants à mettre en réserve pour faire face à des sinistres. Ces modèles de réserve de risque sont fondés sur l'utilisation de processus de Poisson pour les arrivées de sinistres.

Soit $N = \{N_t\}$ un processus de Poisson d'intensité constante λ et soit $U_1, U_2, \ldots$ des variables indépendantes de même loi.

Le processus $N^* = \{N_t^* = U_1 + U_2 + \cdots + U_{N_t}\}$ est appelé *processus de Poisson composé*. U_n est le changement de N_t^*.

Processus de risque (cont'd)

A la *n*ème arrivée du processus de Poisson, une horloge aléatoire prévient des arrivées du processus de Poisson et à la nème sonnerie rajoute une quantité U_n à N_t^*.

Les variables aléatoires $U_1, U_2, \ldots$ vont représenter les sinistres et le processus de Poisson décrit leurs temps d'arrivée. Les durées entre sinistres seront donc distribuées suivant une loi exponentielle de paramètre λ.

Modèle de réserve de risque

Le *modèle de réserve de risque* fondé sur le processus de Poisson composé est constitué

1. des instants aléatoires auxquels les sinistres sont enregistrés et décrits par un processus de Poisson $\{N_t\}$ d'intensité constante λ
2. les tailles des sinistres $U_1, U_2, \ldots$ i.i.d. prises par convention positives
3. une réserve de risque initiale u
4. les primes supposées collectées à un taux constant $\beta > 0$ de telle façon que le revenu soit une fonction linéaire du temps.

Le *processus de réserve de risque* $\{R_t : t \geq 0\}$ est alors donné par $R_t = u + \beta t - \sum_{i=1}^{N_t} U_i$.

Le *processus de surplus de sinistres* $\{S_t : t \geq 0\}$ est donné par $S_t = \sum_{i=1}^{N_t} U_i - \beta t$.

Ruine

L'*instant de ruine* $\tau(u) = \min\{t : R_t < 0\} = \min\{t : S_t > u\}$ est le premier instant où le processus de réserve devient négatif ou de manière équivalente le processus de surplus excède le niveau u.

On s'intéresse en général au comportement des *probabilités de ruine*

1. à horizon fini $\psi(u; x) = P[\tau(u) \leq x]$
2. à horizon infini $\psi(u) = \lim_{x\to\infty} \psi(u; x) = P[\tau(u) \leq \infty]$

La probabilité $\psi(u)$ correspond à une *probabilité de ruine ultime*, et $\bar{\psi} = 1 - \psi(u)$ à une probabilité de survie.

Cat bond

Les cat bonds servent comme sources de financement alternatives en assurance.

Il s'agit de *titrisation* de risque catastrophe.

Pourquoi a-t-on besoin de sources alternatives par rapport au mécanisme usuel de réassurance ?

Cat bond (cont'd)

1. augmentation de la croissance de la population dans les zones à risque sujettes à des ouragans et tremblements de terre (croissance ces 30 dernières années en Californie est 3 fois plus importante que la moyenne aux US). Aux US passage de \$48.7 milliards entre 1950 et 1988 à \$98 milliards en pertes catastrophes entre 89 et 98, soit un accroissement de 101.2%.
2. consolidation parmi les gros réassureurs ce qui engendre des problèmes de capacité.

La tendance est à l'accroissement des grosses catastrophes. De 1970 à 1988: 11 événements en 19 ans. De 1989 à 1995: 19 événements en 7 ans. On estime qu'actuellement on doit s'attendre en moyenne à un sinistre supérieur à \$1 (\$3) milliard par an (2 ans et demi).

Fonctionnement des Cat Bonds

Dans un cat bond typique on crée un SPV (special purpose vehicle) qui va agir comme un réassureur en émettant de la dette sur les marchés financiers et en fournissant une police de réassurance à l'assureur cédant en cas de survenance d'un évènement causant des pertes supérieures à une certaine limite (point d'attachement).

Les flux servis aux détenteurs des obligations sont utilisés pour payer les dommages.

On suspend le paiement des intérêts et/ou le remboursement d'une fraction du capital et reverse ces montants à l'assureur devant payer les sinistres.

Utilisation des Cat bonds

Initialement (1996) les cat bonds ne concernaient qu'un seul type de sinistre dans une seule région (ouragan du bassin Atlantique) avec un horizon d'un an.

Aujourd'hui ils peuvent couvrir plusieurs types de sinistres à travers le monde sur plusieurs années.

On a eu 7 transactions en 97 ($1 milliard) et 8 transactions en 98 ($1 milliard). Le nombre et montant vont en croissant.

Les investisseurs considèrent les cat bonds comme une nouvelle classe d'actifs qui offrent des rendements non corrélés aux investissements traditionnels que sont les actions et obligations. L'inclusion dans un portefeuille d'actifs offre une opportunité de réduction de risque due à la diversification.

Exemple - Ouragan du bassin atlantique

Le facteur principal de sinistres est la vitesse du vent. On peut modéliser la survenance de vitesse très élevée par un processus de Poisson.

On calcule ensuite les périodes de retour c.à.d. le temps moyen d'attente d'arrivée d'un événement d'une ampleur donnée ou supérieure, et les probabilités de non survenance c.à.d. la probabilité qu'aucun événement d'une ampleur donnée ou supérieure survienne dans une période donnée, par exemple une année.

La période de retour est d'environ 3.6 années pour des vitesses de 125 mph ou supérieures et sa probabilité de non survenance est de 74% (il y donc 26% de chances qu'un tel événement survienne dans l'année). On modélise également les pressions atmosphériques puisque des pressions basses engendrent des tempêtes plus fortes.