Processus ponctuels

Olivier Scaillet

University of Geneva and Swiss Finance Institute

Outline

1. Processus ponctuels
2. Quotations

Introduction

On désire dans ce chapitre construire des modèles d'une distribution aléatoire de points sur un espace, typiquement un sous-ensemble de $[0, +\infty)$, $\mathbb{R}$, ou $\mathbb{R}^d$.

Beaucoup de phénomènes naturels peuvent être représentés ainsi:

- les temps de réparation d'un groupe de machines
- les positions et moments d'occurrence de tremblements de terre dans les 50 prochaines années
- la localisation des gisements de pétrole dans une zone donnée
- l'implantation des arbres dans une forêt
- la localisation des troupes sur un champ de bataille

Définition et notions de base

Soit E un sous-ensemble d'un espace Euclidien, ici pris comme étant un sous-ensemble de $[0, +\infty)$, $\mathbb{R}$, ou $\mathbb{R}^d$.

On souhaite distribuer les points aléatoirement sur E et avoir une notation simple de la fonction servant à compter le nombre aléatoire de points qui tombent dans l'ensemble fermé A. Si $E = \mathbb{R}^2$, A correspondra à un pavé (rectangle) de $\mathbb{R}^2$.

Supposons que $\{X_t : t \geq 0\}$ représente les points successifs dans l'espace E des états. Si on définit la mesure discrète $\epsilon_{X_t}(A) = \begin{cases} 1 & si\ X_t \in A \\ 0 & si\ X_t \notin A \end{cases}$ alors en sommant sur t, on obtient le nombre total de points X_t tombant dans A.

Définition et notions de base (cont'd)

Si on définit la mesure de comptage N par $N(\cdot) = \sum_t \epsilon_{X_t}(\cdot)$, alors $N(A) = \sum_t \epsilon_{X_t}(A)$ est le nombre aléatoire de points qui tombent dans l'ensemble A.

N est appelé un *processus ponctuel* et $\{X_t\}$ sont appelés les *points*.

L'*intensité* de N ou *mesure moyenne* de N est définie par $\mu(A) = E[N(A)]$, ce qui correspond au nombre espéré de points dans la région A.

Mesure de Poisson aléatoire

Soit N un processus ponctuel défini sur un espace d'états E dont on considère la σ-algèbre borélienne c.à.d la σ-algèbre Φ formée des ensembles ouverts de E, alors N est un processus de Poisson de mesure moyenne μ, ou encore une mesure aléatoire de Poisson si :

1. pour tout $A \in \Phi$,
$$P[N(A) = k] = \begin{cases} \frac{e^{-\mu(A)}\mu(A)^k}{k!} & si \ \mu(A) < \infty \\ 0 & si \ \mu(A) = \infty \end{cases}$$
2. si $A_1, A_2, \ldots, A_k$ sont des sous-ensembles disjoints de E dans Φ, alors $N(A_1), N(A_2), \ldots, N(A_k)$ sont des variables aléatoires indépendantes.

Mesure de Poisson aléatoire (cont'd)

Ainsi, N est poissonien si le nombre de points aléatoires dans un ensemble A est distribué suivant une loi de Poisson de paramètre $\mu(A)$ et le nombre de points dans des régions disjointes sont des variables aléatoires indépendantes.

La condition 2. lorsque $E = \mathbb{R}$ correspond à la propriété d'incréments indépendants puisque pour tout $t_1 < t_2 < \cdots < t_k$, $N((t_i, t_{i+1}])$ sont des variables indépendantes.

Exemple: localisation d'un compétiteur

Soit une agence bancaire se demandant à quelle distance moyenne un compétiteur sera localisé si le taux de concentration est de $\alpha = 3$ agences par kilomètre carré.

Soit R la distance au compétiteur le plus proche et soit $d(r)$ le disque de rayon r centré sur l'agence. On a ainsi $A = d(r)$ et $\mu(A) = \alpha\pi r^2$ puisque la surface d'un disque de rayon r vaut πr^2.

Cela donne $P[r < R] = P[N(A) = 0] = e^{-\mu(A)} = e^{-\alpha\pi r^2}$.

Exemple: localisation d'un compétiteur (cont'd)

La distance moyenne au plus proche compétiteur vaut dès lors $E[R] = \int_0^\infty P(r < R)dr = \int_0^\infty e^{-\alpha\pi r^2} dr$.

En posant $\alpha\pi = \frac{1}{2\sigma^2}$, l'intégrale se réécrit $\int_0^\infty e^{-\frac{r^2}{2\sigma^2}} dr = \frac{1}{2}\int_{-\infty}^{\infty} e^{-\frac{r^2}{2\sigma^2}} dr = \frac{\sqrt{2\pi\sigma^2}}{2}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{r^2}{2\sigma^2}} dr = \frac{\sqrt{2\pi\sigma^2}}{2}$ en utilisant le fait que la densité d'une loi normale intègre à un.

On obtient finalement $E[R] = \frac{1}{2}\sqrt{\frac{1}{\alpha}}$, soit 0.2887 kilomètres pour $\alpha = 3$.

Processus ponctuels marqués

Etant donné un processus de Poisson initial on peut se demander dans quelles circonstances on peut élargir la dimension du point tout en gardant la structure poissonienne.

Soit $\{X_t\}$ un processus de Poisson de mesure moyenne μ défini sur l'espace état E_1.

Soit $\{J_t\}$ une suite de variables aléatoires i.i.d. sur un second espace E_2 de fonction de répartition F.

Processus ponctuels marqués (cont'd)

Si $\{X_t\}$ et $\{J_t\}$ sont définis sur le même espace de probabilité et sont indépendants, alors la mesure de comptage N, $N(\cdot) = \sum_t \epsilon_{(X_t, J_t)}(\cdot)$, définira un processus ponctuel correspondant à une mesure aléatoire de Poisson sur $E_1 \times E_2$ de mesure moyenne $\mu \times F$, au sens où si $A_i \subset E_i$, $i = 1, 2$, alors $(\mu \times F)(A_1 \times A_2) = \mu(A_1)F(A_2)$.

On dit alors que l'on rajoute au point X_t la *marque* J_t et on parle de manière générale de *Processus ponctuels marqués*.

Processus ponctuels marqués (cont'd)

- le processus de Poisson composé utilisé dans la modélisation des sinistres en assurance est un cas particulier de processus ponctuels marqués, les marques représentant les ampleurs des sinistres
- si on relâche l'hypothèse d'exponentialité des durées entre événements et/ou l'hypothèse d'indépendance entre arrivées et marques on obtient une classe beaucoup plus large de processus ponctuels marqués mais qui demande une étude séparée du comportement des arrivées et des marques (notion de compensateur).

Application aux cotations de prix

Les données de cotations à haute fréquence correspondent à des bases de données comprenant une description complète de l'évolution des prix, cotations achat-vente, prix échangé, quantité échangée, temps, ...

Leurs caractéristiques sont

- les intervalles de temps entre observations sont irréguliers.
- le comportement du marché est complètement observé.
- les données présentent des caractéristiques statistiques très particulières.
- la mesure de temps pour les observations est la seconde.

Application aux cotations de prix (cont'd)

La plupart des bourses d'actions ont leur propre système d'enregistrement des données (NYSE, NASDAQ, Bourse de Paris, Bourse de Madrid, ...) et distribuent leur base de données (TAQ database du NYSE).

Les mécanismes d'échange peuvent être

- dirigés par les prix
- dirigés par les ordres.

Pour les marches dirigés par les prix il existe un teneur de marché qui assure la liquidité du marché en achetant au prix bid et vendant au prix ask.

La rémunération du risque qu'il prend se fait à l'aide du bid-ask spread.

Application aux cotations de prix (cont'd)

Le teneur de marché possède une information privée à savoir les ordres limites.

Pour les marchés dirigés par les ordres il n'existe pas de teneur de marché.

Les ordres d'achat-vente et les ordres limites sont entrés directement dans le carnet d'ordres.

Une transaction s'effectue en faisant correspondre les ordres.

Il existe des règles de priorité en termes de prix et de temps.

Le spread est donné par la différence entre le meilleur prix de vente et le meilleur prix d'achat.

Certains marchés sont hybrides (NYSE).

Application aux cotations de prix (cont'd)

La théorie concernant la *microstructure des marchés* tente d'expliquer le comportement des agents et des prix dans un marché organisé selon certaines règles.

Cette théorie se concentre sur

- les systèmes de trading
- la liquidité
- l'asymétrie d'information entre agents intervenant sur le marché
- l'information transmise par les prix.
- ...

Application aux cotations de prix (cont'd)

On répond à des questions du type :

- Est-ce que le teneur de marché peut gérer optimalement son stock d'inventaire ?
- Est-ce que le teneur de marché prend en compte l'information potentielle privée des acheteurs et vendeurs ?
- Est-ce qu'il y a relation entre stock d'inventaire et cotations ?

Application aux cotations de prix (cont'd)

Le point important dans ces modèles est que les intervalles de temps entre transactions ne sont plus exogènes.

Cela implique en particulier que les études empiriques uniquement basées sur les données de prix peuvent mener à des conclusions erronées car on ignore l'information transmise par la longueur des intervalles de temps entre transactions.

Si les intervalles de temps sont longs cela veut dire que les agents, et surtout ceux susceptibles d'être informés, ont peu révisé leurs anticipations.

Application aux cotations de prix (cont'd)

Des intervalles courts révèlent qu'une information est détenue par certains agents et qu'un mécanisme d'ajustement des prix via l'augmentation des échanges se met en place.

De même l'importance des volumes échangés transmet l'information que certains agents sont informés.

L'analyse empirique des durées entre transactions est donc importante pour comprendre les mécanismes de fonctionnement et de révélation de l'information privée.

Application aux cotations de prix (cont'd)

On utilise

- les durées de transactions pour mesurer l'intensité de trading : temps passé jusqu'au prochain trade.
- les durées de prix pour mesurer la volatilité : temps passé jusqu'à un changement de prix cumulé donné.
- les durées de volumes pour mesurer la liquidité : temps passé jusqu'à un changement de volume cumulé donné.

Il s'agit à chaque fois de processus ponctuels éventuellement marqués si l'on rajoute des dimensions, par exemple durées entre transactions (durées entre événements) et changement de prix observés ou volumes échangés (marques).

Application aux cotations de prix (cont'd)

Les modèles utilisés sont dits à temps accéléré.

On modélise les durées $d_i = T_i - T_{i-1}$ entre événements comme $d_i = g(z_i; \theta)\epsilon_i$, avec g une fonction des variables exogènes qui dans une modélisation dynamique peuvent inclure les durées passées.

Le terme d'erreur ϵ_i représente une durée naïve qui suit une certaine distribution donnée.

Le facteur $g(z_i; \theta)$ conduit à une accélération ou décélération de la durée naïve en fonction des valeurs prises par les variables et par le paramètre θ.

Application aux cotations de prix (cont'd)

La densité conditionnelle de la durée en fonction de la densité $f_\epsilon(\cdot;\beta)$ de la durée naïve paramétrée par β est donnée par $f_d(y|z;\theta,\beta) = \frac{1}{g(z;\theta)} f_\epsilon(\frac{y}{g(z;\theta)};\beta)$.

Les modèles économétriques proposés de type ACD (autoregressive conditional duration) utilisent une forme linéaire faisant intervenir les valeurs passées des durées.

Ce type de modélisation permet de prendre en compte des phénomènes de clustering dans les durées c.à.d. que de petites, resp. grandes, durées ont tendance à être suivies par de petites, resp. grandes, durées.

Application aux cotations de prix (cont'd)

L'estimation de ces modèles se fait par maximum de vraisemblance et nécessite une désaisonnalisation des données.

Cette saisonnalité intrajournalière ou intrahebdomadaire est due a des typicités dans le comportement des intervenants (heures de repas, débouclage de position avant le WE, . . .).

L'estimation de ces modèles permet de vérifier si certaines prédictions de la théorie économique sont vérifiées empiriquement.

On les utilise également pour faire du trading intrajournalier ou dans le contrôle des risques liés à des positions intrajournalières.