

HIGH-FREQUENCY DATA AND LIMIT ORDER BOOKS

EXAMEN DU VENDREDI 15 AVRIL 2022

Durée : 2h00. Cet examen comporte 5 exercices énoncés sur 3 pages. Les exercices sont indépendants. Documents, calculatrices et ordinateurs interdits.

Exercice 1 - Données *trades and quotes*

On considère le carnet d'ordres d'un actif financier. A t_0 , l'état des 5 meilleures limites *bid* et *ask* de ce carnet est le suivant :

side	Bid					Ask				
level	1	2	3	4	5	1	2	3	4	5
price	88.26	88.25	88.24	88.23	88.22	88.30	88.31	88.33	88.34	88.35
quantity	623	529	749	305	338	290	869	634	783	1147

On pourra noter $b_i(t)$ le prix de la i -ème limite *bid* au temps t , et $q_i^B(t)$ le volume associé. Symétriquement, on utilisera les notations $a_i(t)$ et $q_i^A(t)$ pour l'*ask*. On considère que la suite d'ordres suivante est envoyée sur le marché :

- au temps t_1 , un ordre de vente de l'actif de taille 100 au prix limite 88.31 ;
- au temps t_2 , un ordre d'achat de l'actif au prix de marché de taille 250 ;
- au temps t_3 , une annulation d'un ordre de taille 400 au prix limite 88.31 ;
- au temps t_4 , un ordre d'achat de l'actif au prix de marché de taille 50 ;
- au temps t_5 , un ordre d'achat de l'actif de taille 100 au prix limite 88.28 ;
- au temps t_6 , un ordre de vente de l'actif de taille 100 au prix limite 88.27 ;

1. Rappeler la définition du *ticksize* et donner sa valeur probable pour cet actif.
2. Rappeler la définition du *spread* et de l'*imbalance* et donner leur valeur au temps t_0 .
3. Rappeler la définition du *weighted mid price* et expliquer l'intérêt de cette notion.
4. Tracer l'évolution des *best bid price*, *best ask price* et *mid price* de t_0 à t_6 .
5. Ecrire le fichier *trades* issu de cette suite d'ordres au format utilisé en TP.
6. Ecrire le fichier *quotes* des 2 meilleures limites *bid* et *ask* issu de cette suite d'ordres au format utilisé en TP. On pourra utiliser le signe " pour signifier qu'une quantité est inchangée d'une ligne à l'autre.

Exercice 2 - Signature des transactions

On considère un fichier de données de transactions sur un actif financier. Les champs disponibles sont :

- **ts** : le *timestamp* de la transaction ;
- **price** : le prix de la transaction ;
- **qty** : le volume de la transaction.

1. Expliquer ce qu'est la signature des transactions.
2. Ecrire en pseudo-code détaillé (pas en Python) l'algorithme dit du *tick test*.

3. Donner le résultat attendu de votre algorithme appliqué au fichier exemple suivant :

ts	price	qty
2017-01-24 14 :54 :11	86.74	46
2017-01-24 14 :54 :11	86.74	59
2017-01-24 14 :54 :37	86.77	33
2017-01-24 14 :54 :37	86.78	43
2017-01-24 14 :54 :39	86.75	125
2017-01-24 14 :54 :39	86.75	25
2017-01-24 14 :54 :39	86.75	70
2017-01-24 14 :54 :39	86.75	70
2017-01-24 14 :54 :45	86.73	40
2017-01-24 14 :54 :57	86.72	58
2017-01-24 14 :54 :57	86.71	109
2017-01-24 14 :55 :16	86.72	169
2017-01-24 14 :55 :16	86.72	152

4. Cette signature est-elle exacte ? Avec quel type de données peut-on améliorer la signature des transactions ?

Exercice 3 - Signature plot d'un modèle de prix poissonien

Soit un carnet d'ordre et δ le *ticksize* de ce carnet. On note $(m_t)_{t \geq 0}$ le processus du prix *mid*, et sans perte de généralité on pose $m_0 = 0$. On suppose que les variations de ce prix *mid* sont à valeurs dans $\{-\frac{\delta}{2}, +\frac{\delta}{2}\}$. Les instants de hausse du prix *mid* sont modélisés par un processus de Poisson homogène $(N_1(t))_{t \geq 0}$ d'intensité $\lambda_1 > 0$ et les instants de baisse du prix *mid* par un processus de Poisson homogène $(N_2(t))_{t \geq 0}$ d'intensité $\lambda_2 > 0$. On suppose que N_1 et N_2 sont indépendants.

1. Déterminer l'espérance et la variance du prix *mid* à la date t en fonction des paramètres du modèle.
2. Tracer l'allure du *signature plot* du prix *mid* dans ce modèle dans le cas $\lambda_1 \neq \lambda_2$ puis dans le cas $\lambda_1 = \lambda_2$.

Exercice 4 - Simulation de processus de Poisson non homogènes

Soit $(\Omega, \mathcal{F}, \mathbf{P})$ un espace de probabilité. Soit $N = (N_t)_{t \geq 0}$ un processus de Poisson non-homogène construit sur cet espace, d'intensité déterministe $\lambda : [0, +\infty[\rightarrow [0, +\infty[$. On notera $\Lambda : [0, +\infty[\rightarrow [0, +\infty[, t \mapsto \int_0^t \lambda(s) ds$ et $(T_i)_{i \geq 1}$ la suite ordonnée des temps de saut de N .

1. Soit $T > 0$ un horizon fixé. Soit $n \in \mathbb{N}^*$. Soit un n -uplet $(t_1, \dots, t_n) \in [0, +\infty[^n$. Soit $\epsilon > 0$ suffisamment petit pour que les intervalles $]t_i - \epsilon, t_i + \epsilon[$ soient disjoints. Calculer

$$\mathbf{P} \left(\bigcap_{i=1}^n \{T_i \in]t_i - \epsilon, t_i + \epsilon[\} \mid N_T = n, t_1 < t_2 < \dots < t_n \right).$$

2. En déduire en faisant $\epsilon \rightarrow 0$ la densité conditionnelle de (T_1, \dots, T_n) sachant $N_T = n$.
3. Interpréter ce résultat avec la fonction $F : [0, T] \rightarrow [0, 1], t \mapsto \Lambda(t)/\Lambda(T)$.
4. En déduire un algorithme de simulation du processus N sur $[0, T]$.
5. A quelle(s) condition(s) cet algorithme peut-il être performant ?

Exercice 5 - Un modèle de carnet d'ordres

On considère un modèle des K meilleures limites *bid* et *ask* d'un carnet d'ordre. On note $X_{-i}(t)$ la quantité disponible au i -ème meilleur *bid* et $X_i(t)$ la quantité disponible au i -ème meilleur *ask*, $i = 1, \dots, K$. On suppose que pour tout $k = -K, \dots, -1, 1, \dots, K$, le processus $(X_k(t))_{t \geq 0}$ est une chaîne de Markov à valeurs dans \mathbb{N} avec :

- $\lambda_k(n)$ l'intensité d'arrivée des *limit order* à la limite k dans l'état $X_k(t) = n$,
- $\mu_k(n)$ l'intensité d'arrivée des *market order* à la limite k dans l'état $X_k(t) = n$,
- et $\theta_k(n)$ l'intensité d'arrivée des annulations à la limite k dans l'état $X_k(t) = n$.

On suppose que tous les ordres sont de taille 1.

1. Ecrire le générateur infinitésimal de la chaîne de Markov $(X_k(t))_{t \geq 0}$.
2. Déterminer la distribution stationnaire de $(X_k(t))_{t \geq 0}$.
3. Proposer une méthode d'estimation des intensités $\lambda_k(n)$, $\mu_k(n)$ et $\theta_k(n)$ avec des données *trades and quotes*.
4. Tracer sur un même graphe pour $\lambda_1(n)$, $\mu_1(n)$, $\mu_2(n)$ et $\theta_1(n)$ une allure générale en fonction de n qui vous semble réaliste, et justifier votre proposition.
5. Commenter les hypothèses de ce modèle.

• • • Fin de l'énoncé. • • •