High-frequency data and limit order books
Examen du jeudi 20 avril 2023

Durée : 2h00. Cet examen comporte 3 exercices énoncés sur 3 pages. Les exercices sont indépendants. Documents, calculatrices et ordinateurs sont interdits. Vous pouvez répondre au choix en français ou en anglais.

Exercise 1 - Time change theorems

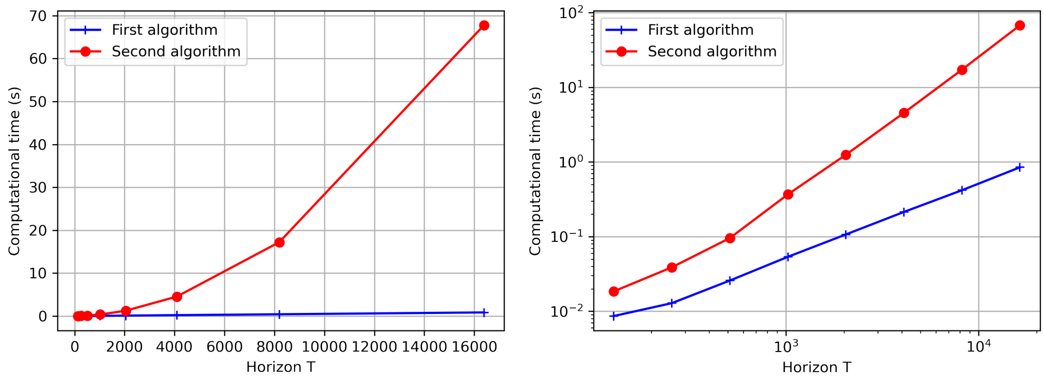
Let us consider the following result : "Let $(M_t)_{t\geq 0}$ be an homogeneous Poisson process with intensity $\mu = 1$. Let $f : \mathbb{R} \to \mathbb{R}$ be an increasing continuously differentiable function such that $f(0) = 0$. Then $N_t = M_{f(t)}$ defines a non-homogeneous Poisson process with intensity $\nu(t) = f'(t)$, $t \geq 0$."

1. Recall the definition of a non-homogeneous Poisson process.
2. Prove the above result (directly, without using the time change theorems seen in the course).
3. Give two examples of applications of time change theorems for point processes used in the course.

Exercise 2 - Miscellaneous questions

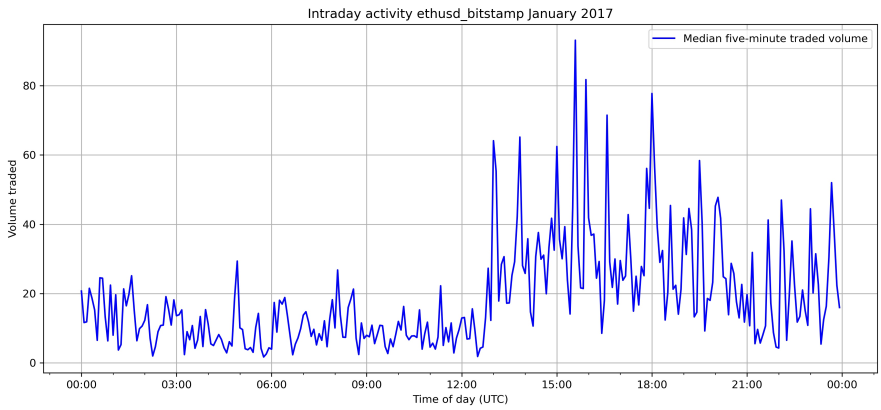
1. Let us consider some order flow data *at the best quotes* of a limit order book (LOB) of a large tick stock. Orders are splitted in 12 types denoted T_s^a, where $T \in \{L, M, C\}$ is the type of the order (L for limit, M for market or C for cancellation), $s \in \{B, A\}$ is the side of the book impacted by this order (B for bid, A for ask), and $a \in \{0, 1\}$ indicates if the order moves the price ($a = 1$) or not ($a = 0$). An empirical study computes the conditional probabilities $\mathbf{P}(x|y)$ that, given the observation of an order of type y, the following order is of type x. We consider :
   (a) $\mathbf{P}(L_B^1|M_B^1)$;
   (b) $\mathbf{P}(C_B^1|L_B^0)$.

   For each of these probabilities, quickly draw three sketches of the LOB explaining the dynamics of these sequences when they occur (before y, after y, and after x). Indicate whether these probabilities are expected to be rather high or rather low (with respect to the unconditional probability $\mathbf{P}(x)$). Justify precisely each answer.
2. The following graphs plot the computational time used by two different algorithms to simulate a Hawkes process N with intensity $1.0 + \displaystyle\int_0^t \frac{e^{-\beta(t-s)}}{2}\, dN_s$ on $[0, T]$, as a function of the T, in natural scale and logarithmic scale. Both algorithm use a thinning method.



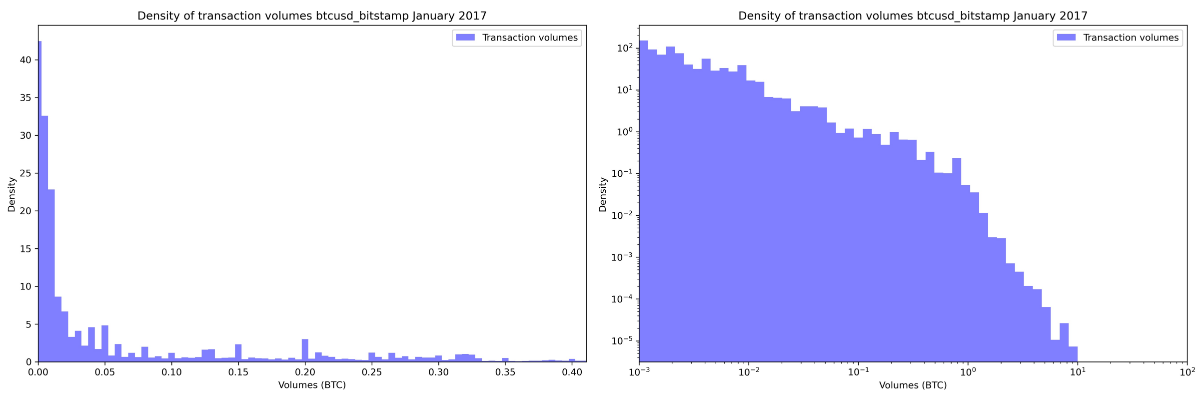
Estimate the complexity of each algorithm and explain the probable difference between the algorithm. Why is the second algorithm still worth implementing ?

3. The following graphs plot the median (over all trading days in January 2017) volume of the Ethereum cryptocurrency (ETH) traded every 5 minutes on the Bitstamp exchange (transactions in US dollars). Volume traded is expressed in ETH, time in UTC.



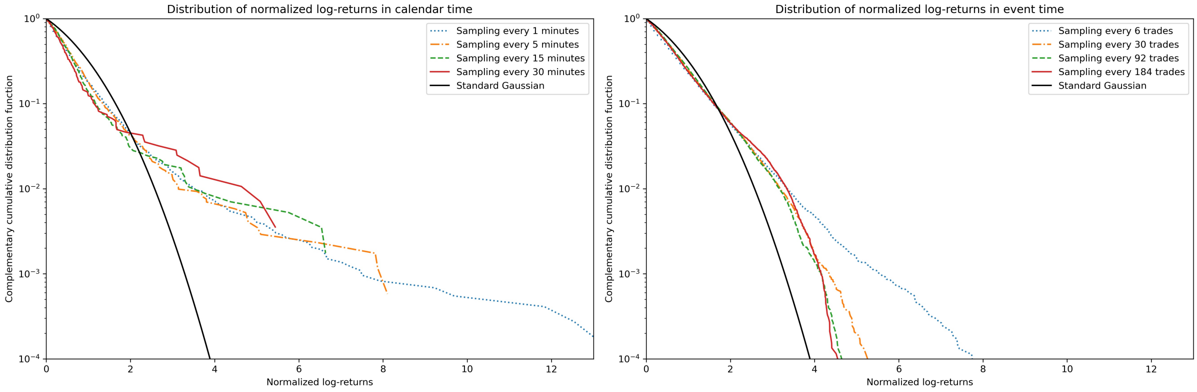
Comment the graph and highlight the similarities and differences with respect to the equity market data studied in the labs.

4. The following graphs plot the distribution of the transaction volumes of the Bitcoin cryptocurrency (BTC) on the Bitstamp exchange (transactions in US dollars). Volume traded is expressed in BTC. Distribution is plotted in natural scale on the left, and log-log scale on the right.



Comment the graphs and highlight the similarities and differences with respect to the equity market data studied in the labs.

5. The following graphs plot the standardized log-returns of the transaction prices of the Ehtereum cryptocurrency (ETH) on the Bitstamp exchange (transactions in US dollars). On the left panel, log-returns are computed in calendar time. On the right panel, log-returns are computed in event time. On this data, there is approximately 6 trades per minute in average.



Explain the difference between calendar time and event time. Define (mathematically, with a formula) the log-returns in each case. Comment the graphs.

Exercise 3 - Marked Poisson processes and price models

Let $(T_n)_{n\geq 1}$ be a point process (an increasing sequence of random events). Let $(X_n)_{n\geq 1}$ be a sequence of i.i.d. random variables with Bernoulli distribution with parameter $p \in [0,1]$, independent of (T_n). The process $\{(T_n, X_n), n \geq 1\}$ is called a *marked point process*, since it can be seen as a process for which each point T_n is given a mark X_n.

Assume now that $(T_n)_{n\geq 1}$ is the sequence of events of a time-homogeneous Poisson process N with parameter $\lambda > 0$. Let $\mathcal{T}^{(i)} = \{T_n : X_n = i\}$, $i = 0, 1$, be the subset of times for each mark. Let $(N_t^{(0)})_{t\geq 0}$ and $(N_t^{(1)})_{t\geq 0}$ be the counting processes associated to these subset of points.

1. Compute $\mathbf{P}(N_t^{(1)} - N_s^{(1)} = k | N_t - N_s = n)$ for all $0 \leq s < t$, and all $k, n \in \mathbb{N}$, $k \leq n$.
2. Compute $\mathbf{P}(N_t^{(1)} - N_s^{(1)} = k)$ for all $0 \leq s < t$, and all $k \in \mathbb{N}$.
3. Prove that for all $0 \leq s < t \leq u < v$, $N_t^{(1)} - N_s^{(1)}$ and $N_v^{(1)} - N_u^{(1)}$ are independent.
4. What is the conclusion of the above questions ?
5. By analogy, write the equivalent result for the process $N^{(0)}$.
6. Prove that for all $0 \leq s < t$, $N_t^{(0)} - N_s^{(0)}$ and $N_t^{(1)} - N_s^{(1)}$ are independent.
7. Assume from now on that $(N_t)_{t\geq 0}$ is the counting process of the movements of the mid-price $(p_t)_{t\geq 0}$ in a high-frequency price model, and that $X_n = 1$ (resp. $X_n = 0$) indicates an upward (resp. downward) price movement at time T_n. For simplicity, let us set $p_0 = 0$ and assume that the absolute value of each price movement is $\frac{\delta}{2}$, $\delta > 0$.
   (a) Compute the mean and variance of p_t, $t \geq 0$.
   (b) Compute the theoretical signature plot of the realized variance in this model.

• • • Fin de l'énoncé. • • •