CentraleSupélec, Université Paris-Saclay
3A – Dominante MDS – Mention MMF

Année universitaire 2023-2024
Ioane Muni Toke

High-frequency data and limit order books
Examen du lundi 8 avril 2024

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

Exercise 1 - A non-homogeneous Poisson process

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space. Let $(\alpha, \beta) \in (\mathbb{R}_+^*)^2$. Let $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$, $t \mapsto \alpha\beta e^{-\beta t}$. Let $(N_t)_{t\geq 0}$ a (non-homogeneous) Poisson process with intensity φ.

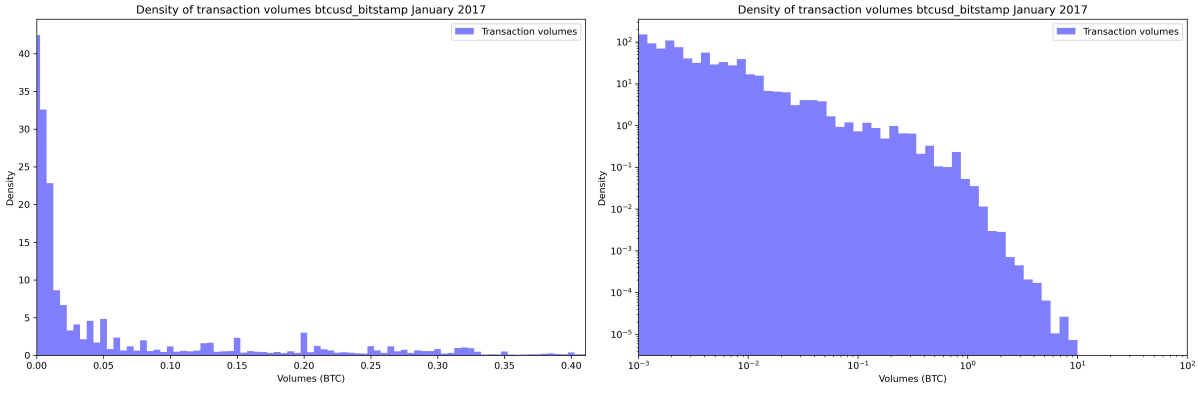
1. **Definition.**
   (a) Recall the definition of a non-homogeneous Poisson process.
   (b) Check that φ can indeed be used as the intensity of a (non-homogeneous) Poisson process.
2. **Distribution functions.** For $n \geq 0$, let T_n be the random variable equal to the time of the n-th jump of the process N. We set $T_n = +\infty$ if the process has strictly less than n jumps.
   (a) Let F_{T_1} be the cumulative distribution function of T_1. Show that
   $$F_{T_1}(t) = 1 - e^{-\alpha(1-e^{-\beta t})}, \quad t \in [0, \infty).$$
   (b) Compute $\mathbf{P}(T_1 = +\infty)$.
   (c) Compute $F_{T_n|T_{n-1}=t_{n-1}}$ the cumulative distribution function of T_n given $T_{n-1} = t_{n-1}$.
   (d) Compute $\mathbf{P}(T_n = +\infty \,|\, T_{n-1} = t_{n-1})$.
3. **Simulation.**
   (a) Explain how one can use the previous questions to simulate the process N.
   (b) Terminate the necessary computation so that the simulation method is complete.
   (c) Write the simulation routine in pseudo-code.
4. **Estimation.** Assume that the parameter β is known. Let $(s_1, \ldots, s_n)$ be a sample of the process N observed on $[0, T]$.
   (a) Compute the maximum likelihood estimator of the parameter α.
   (b) Show that the same estimator can be obtained by a method of moments.
   (c) What is the asymptotic distribution of this estimator as $T \to +\infty$?
5. **Goodness-of-fit.** Let $X_i = \alpha(e^{-\beta T_{i-1}} - e^{-\beta T_i})$.
   (a) Find the distribution of the random variable X_i.
   (b) Explain how a statistical goodness-of-fit test can be be built with the X_i's.

Exercise 2 - Miscellaneous questions

1. **On the ticksize.**
   (a) Recall the definition of the ticksize in a limit order book.
   (b) What is a large-tick stock ?
   (c) Why would a trading strategy be different on a large-tick stock and a small-tick stock ?
2. **On the imbalance.**
   (a) Recall the definition of the imbalance in a limit order book (the definition used in the course is expected, although alternative forms exist).
   (b) Draw (what you expect to be) the shape of the distribution of the imbalance in a limit order book. Justify the shape you have drawn.
   (c) Let X and Y be two i.i.d. random variable with exponential distribution with parameter λ and $V = \frac{X}{X+Y}$. Find the distribution of V by computing $\mathbf{E}[h(V)]$ for some measurable function h, using the change of variable

$$\begin{cases} u = x \\ v = \frac{x}{x+y} \end{cases} \Leftrightarrow \begin{cases} x = u \\ y = \frac{u}{v} - u. \end{cases}$$

   (d) What is the distribution of the imbalance in a limit order book model where the best bid and best ask queues are i.i.d. with exponential distribution ? Is this in accordance with your drawing above ?
3. **On the Epps effect.**
   (a) What is the Epps effect ?
   (b) What are the frequently given explanations of the Epps effect ?
4. **On trade volumes.** 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. **On several clocks.** Explain the concepts of calendar time, event time and volume time in the context of financial timeseries.

6. **On Poisson LOB models.** Let us consider a Poisson model of a limit order book. Let X_t be the number of shares at time t at some limit in the book far away from the best quotes, so that market orders can be neglected. Let $\lambda > 0$ be the intensity of limit orders arriving at this limit. Assume that all orders have size 1. Assume that the cancellation intensity is $n\theta$ when $X_t = n$, for some given $\theta > 0$. Compute the stationary distribution of the number of shares X_t.
7. **On price impact.**
   (a) Explain the concept of selective liquidity taking.
   (b) Explain the concept(s) of price impact.
   (c) Explain the concepts of metaorder and child order.
   (d) How is (empirically) the shape of the price impact of a metaorder with respect to the total metaorder size on a market with continuous trading ?

• • • Fin de l'énoncé. • • •