High-frequency data and limit order books

—

Lecture 3 - March 11th, 2025

Ioane Muni Toke

Université Paris-Saclay, CentraleSupélec
MICS Laboratory, FiQuant Team

CentraleSupélec

U. Paris-Saclay CentraleSupelec cursus Ingénieur 3A Mathématiques et Data Science
IP Paris ENSAE M2 Statistics, Finance and Actuarial Science
U. Paris-Saclay Evry M2 Quantative Finance

Table of contents

Table of contents

Point processes

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be some probability space.

A point process on $\mathbb{R}_+ = [0, \infty)$ describes the occurrence (times) of random events.

Mathematically, a point process on $\mathbb{R}_+ = [0, \infty)$ can be seen as:

- a countable random subset of $[0, \infty)$;
- a sequence of nonnegative random variables ;
- a counting process ;
- a discrete random measure.

This lecture focuses on aspects useful for financial modeling. Possible references for further developments outside the scope of this course include [1, 2, 3].

Sequence of random times and counting process

- Let $0 < T_1 < T_2 < \ldots < T_n < \ldots$ be an increasing sequence of random times of $[0, \infty)$ (strictly, *simple process*) .
- We assume that $\lim_{n\to\infty} T_n = +\infty$ (*non-explosive process*).
- The counting process $(N_t)_{t\geq 0}$ associated with this sequence of random times is the stochastic process such that

$$N_t = n \quad \text{if } t \in [T_n, T_{n+1}),$$

or equivalently,

$$N_t = \sup\{n : T_n \leq t\} = \sum_{j\geq 1} \mathbf{1}_{\{T_j \leq t\}}.$$

- The random variables $S_n = T_n - T_{n-1}$ are called interarrival times, or durations.

Properties of a counting process

We thus have the following properties:

(i) $N_0 = 0$;

(ii) a counting process is non-negative integer-valued ;

(iii) a counting process is non-decreasing ;

(iv) a counting process is piecewise constant ;

(v) a counting process is càdlàg (right continuous with left limits) ;

(vi) all jumps of a counting process are of size 1.

The following relationships are often useful:

- $\{N_t \geq n\} = \{T_n \leq t\}$,
- $\{N_t = n\} = \{T_n \leq t < T_{n+1}\}$,
- $\{N_t < n \leq N_{t+s}\} = \{t < T_n \leq t + s\}$.

Random counting measure I

- Alternatively, for each realization $\omega \in \Omega$, we can define a measure m_ω on $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ by setting $m_\omega(A)$ equal to the number of T_i's in $A \in \mathcal{B}(\mathbb{R}_+)$:

$$m_\omega(A) = \sharp\{i \in \mathbb{N} : T_i(\omega) \in A\}.$$

- $m = \sum_{i \geq 0} \delta_{T_i}$ defines a random, discrete, counting measure.
- The number of events occurring in $(t, t+s]$ for any $s, t \geq 0$ is therefore :

$$m((t, t+s]) = N_{t+s} - N_t.$$

- We will use the notation N for both point of view, i.e. we will write $N((t, t+s]) = m((t, t+s]) = N_{t+s} - N_t$.

Random counting measure II

- For a given realization $\omega \in \Omega$, we may define an integral w.r.t the measure m_ω for a suitable function f:
$$\int_{(0,t]} f(u) m_\omega(du) = \sum_{i \in \mathbb{N}} f(T_i(\omega)),$$
and thus write in general
$$\int_{(0,t]} f(u) dN_u = \sum_{0 < T_i \leq t} f(T_i).$$

Table of contents

Construction with interarrival times I

Let $(S_i)_{i\in\mathbb{N}^*}$ be a sequence of independent and identically distributed random variables, with exponential distribution with parameter $\lambda \in \mathbb{R}_+^*$. Let $T_n = \sum_{i=1}^n S_i$, $n \geq 1$. This sequence of increasing random times defines a point process. Let $N_t = \sum_{j\geq 1} \mathbf{1}_{\{T_j \leq t\}}$ be the associated counting process.

Proposition (Distributional properties)

(a) *For any $n \in \mathbb{N}^*$, T_n is Gamma distributed with shape parameter n and rate parameter λ, i.e. with density $x \mapsto \frac{(\lambda x)^{n-1}}{(n-1)!}\lambda e^{-\lambda x}\mathbf{1}_{\mathbb{R}_+}(x)$.*

(b) *For any $t \in \mathbb{R}_+$, N_t is Poisson distributed with parameter λt, i.e. $\mathbf{P}(N_t = k) = \frac{(\lambda t)^k}{k!}e^{-\lambda t}, k \in \mathbb{N}$.*

Construction with interarrival times II

Sketch of the proofs.

(a) $T_1 = S_1$, so $f_{T_1}(x) = \lambda e^{-\lambda x}\mathbf{1}_{\mathbb{R}_+}(x)$ and the property is valid for $n = 1$. $T_{n+1} = T_n + S_{n+1}$ and T_n and S_{n+1} are independent, so densities satisfy

$$f_{T_{n+1}}(x) = f_{T_n+S_{n+1}}(x) = (f_{T_n} * f_{S_{n+1}})(x) = \int_{\mathbb{R}} f_{T_n}(u) f_{S_{n+1}}(x-u)\, du.$$

If $f_{T_n}(x) = \frac{(\lambda x)^{n-1}}{(n-1)!}\lambda e^{-\lambda x}\mathbf{1}_{\mathbb{R}_+}(x)$, then

$$\begin{aligned} f_{T_{n+1}}(x) &= \int_{\mathbb{R}_+} \frac{(\lambda u)^{n-1}}{(n-1)!}\lambda e^{-\lambda u}\lambda e^{-\lambda(x-u)}\mathbf{1}_{\{x-u\geq 0\}}\, du = \lambda^{n+1} e^{-\lambda x}\mathbf{1}_{\mathbb{R}_+}(x)\int_0^x \frac{u^{n-1}}{(n-1)!}\, du \\ &= \lambda e^{-\lambda(x)}\mathbf{1}_{\mathbb{R}_+}(x)\frac{(\lambda x)^n}{n!}, \end{aligned}$$

which proves the result by induction.

Construction with interarrival times III

(b) $\mathbf{P}(N_t = n) = \mathbf{P}(T_n \leq t) - \mathbf{P}(T_{n+1} \leq t)$ and integrating by parts:

$$\begin{aligned}
\mathbf{P}(T_{n+1} \leq t) &= \int_0^t \frac{(\lambda x)^n}{n!} \lambda e^{-\lambda x} \, dx \\
&= \left[-\frac{(\lambda x)^n}{n!} e^{-\lambda x} \right]_0^t + \int_0^t \frac{(\lambda x)^{n-1}}{(n-1)!} \lambda e^{-\lambda x} \, dx \\
&= -e^{-\lambda t} \frac{(\lambda t)^n}{n!} + \mathbf{P}(T_n \leq t),
\end{aligned}$$

hence the result.

Definition following the construction

Definition (Time-homogeneous Poisson process)

A process $(N_t)_{t\geq 0}$ is called a time-homogeneous Poisson process with parameter $\lambda > 0$ if it is the counting process associated to a point process with i.i.d. interarrival times, exponentially distributed with parameter $\lambda > 0$.

- Remark : Construction can be generalized to i.i.d. interarrival times with any distribution (with finite mean). Such processes are called *renewal processes*. But without the memorylessness property of the exponential distribution, non-Poisson renewal processes do not have stationary and independent increments (see next slides), or satisfy the Markov property.

Increments of a Poisson process I

Proposition (Increments)

(a) *Increments of the process $(N_t)_{t\geq 0}$ are independent, i.e. for all $n \in \mathbb{N}^*$ and all strictly increasing finite sequence $t_0 < t_1 < \ldots < t_n$, the random variables $N_{t_1}, N_{t_2} - N_{t_1}, \ldots, N_{t_n} - N_{t_{n-1}}$ are mutually independent.*

(b) *Increments of the process $(N_t)_{t\geq 0}$ are stationary, i.e. for any $s < t$, $N_t - N_s$ and N_{t-s} have the same distribution, which is Poisson with parameter $\lambda(t-s)$.*

Increments of a Poisson process II

Sketch of the proofs. Let $\mathcal{F}_t = \sigma\left(\{N_s : 0 \leq s \leq t\}\right)$. These properties are the consequence of the memorylessness property of the exponential distribution: if $X \sim \mathcal{E}(\lambda)$, then

$$\mathbf{P}(X > t + h \mid X > t) = \frac{\mathbf{P}(X > t + h)}{\mathbf{P}(X > t)} = \frac{e^{-\lambda(t+h)}}{e^{-\lambda t}} = \mathbf{P}(X > h).$$

Thus, at any time s, the distribution of the duration to the next jump is independent of $\mathcal{F}_s$, and the argument is valid for all subsequent jumps. Hence for all $t > s$, $N_t - N_s$ is independent of $\mathcal{F}_s$. (a) is then obtained by backward induction from t_{n-1} to t_1. The same argument shows that for all $t > s$, $N_t - N_s$ has the same distribution as N_{t-s}, which is (b).

Characterization by the increments

- Independent and stationary increments in fact characterize the Poisson process.

Lemma (Memoryless distribution)

A non-negative (and not identically zero) random variable has the memorylessness property if and only if it has an exponential distribution.

Proposition (Poisson characterization by the increments)

Let N be a (simple, locally integrable, and not identically zero) counting process. Then N is a Poisson process if and only if N has stationary and independent increments.

Another possible definition

- As a consequence, multiple equivalent definitions of the Poisson process can be proposed, depending on the point of view/emphasis of the presentation.

Definition (Time-homogeneous Poisson process)

A counting process N is called a time-homogeneous Poisson process if it satisfies the following conditions:

(i) N has independent increments ;

(ii) N has stationary increments ;

(iii) for any interval I with finite length, $N(I)$ is integrable ;

(iv) there exists an interval I such that $\mathbf{P}(N(I) > 0) > 0$.

(v) the process is *simple*: $\lim_{\epsilon\to 0} \frac{\mathbf{P}(N(\epsilon)\geq 2)}{\epsilon} = 0$.

Yet another (useful) definition

Definition (Time-homogeneous Poisson process)

A counting process $(N_t)_{t\geq 0}$ is called a time-homogeneous Poisson process with parameter $\lambda > 0$ if

(i) N has independent increments: for all $k \in \mathbb{N}^*$, for any collection $I_1, \ldots, I_k$ of pairwise disjoint intervals of $[0, \infty)$, the random variables $N(I_1), \ldots, N(I_k)$ are independent ;

(ii) the number of events in any interval with length s is Poisson distributed with parameter λs:

$$\forall s, t \geq 0,\ \forall n \in \mathbb{N},\ \ \mathbf{P}(N_{t+s} - N_t = n) = e^{-\lambda s}\frac{(\lambda s)^n}{n!}.$$

- This form is interesting since it can be generalized to non-homogeneous Poisson processes (see later) and Poisson processes in $\mathbb{R}^d$.

Asymptotic behavior of a Poisson process - Law of large numbers

Proposition (Law of large numbers)

If $(N_t)_{t\geq 0}$ is a (time-homogeneous) Poisson process with parameter $\lambda > 0$, then

$$\lim_{t\to\infty} \frac{N_t}{t} = \lambda, \quad \mathbf{P}\text{-}a.s. \text{ and in } L^2(\Omega, \mathcal{F}, \mathbf{P}).$$

Sketch of the proof. $\mathbf{E}\left[\frac{N_t}{t}\right] = \lambda$, hence $\mathbf{E}\left[\left|\frac{N_t}{t} - \lambda\right|^2\right] = \mathbf{V}\left[\frac{N_t}{t}\right] = \frac{\lambda}{t} \xrightarrow[n\to+\infty]{} 0$, which proves the L^2 convergence. Moreover, for any $n \in \mathbb{N}^*$,

$$\frac{N_n}{n} = \frac{1}{n}\sum_{i=1}^{n}(N_i - N_{i-1}) \xrightarrow[n\to+\infty]{} \mathbf{E}[(N_1 - N_0)] = \lambda \quad p.s.$$

by the strong law of large numbers. Finally, for any $t \in \mathbb{R}_+$, $\lfloor t \rfloor \leq t < \lfloor t \rfloor + 1$ yields

$$\frac{N_{\lfloor t \rfloor}}{\lfloor t \rfloor}\frac{\lfloor t \rfloor}{t} \leq \frac{N_t}{t} < \frac{N_{\lfloor t \rfloor + 1}}{\lfloor t \rfloor + 1}\frac{\lfloor t \rfloor + 1}{t},$$

hence the result.

Asymptotic behavior of a Poisson process - Central limit theorem I

Proposition (Central limit theorem)

If $(N_t)_{t\geq 0}$ is a (time-homogeneous) Poisson process with parameter $\lambda > 0$, then

$$\frac{N_t - \lambda t}{\sqrt{\lambda t}} \xrightarrow[t\to\infty]{\mathcal{L}} Z,$$

where Z is a standard Gaussian random variable.

Sketch of the proof. For any $n \in \mathbb{N}^*$,

$$\frac{N_n - n\lambda}{\sqrt{n\lambda}} = \frac{\sum_{i=1}^{n}(N_i - N_{i-1}) - n\lambda}{\sqrt{n\lambda}} \xrightarrow[n\to+\infty]{} Z \sim \mathcal{N}(0,1) \quad \text{in distribution}$$

by the central limit theorem. In the general case, for any $t \in \mathbb{R}_+$, we write

$$\frac{N_t - \lambda t}{\sqrt{\lambda t}} = \frac{N_t - N_{\lfloor t \rfloor}}{\sqrt{\lambda \lfloor t \rfloor}} \frac{\sqrt{\lambda \lfloor t \rfloor}}{\sqrt{\lambda t}} + \frac{N_{\lfloor t \rfloor} - \lambda \lfloor t \rfloor}{\sqrt{\lambda \lfloor t \rfloor}} \frac{\sqrt{\lambda \lfloor t \rfloor}}{\sqrt{\lambda t}} + \frac{\lambda \lfloor t \rfloor - \lambda t}{\sqrt{\lambda \lfloor t \rfloor}} \frac{\sqrt{\lambda \lfloor t \rfloor}}{\sqrt{\lambda t}}.$$

Asymptotic behavior of a Poisson process - Central limit theorem II

The second term converges in distribution to the expected limit and the third term converges to 0. For the first term, $N_t \leq N_{\lfloor t \rfloor + 1}$ implies

$$\frac{N_t - N_{\lfloor t \rfloor}}{\sqrt{\lambda \lfloor t \rfloor}} \leq \frac{N_{\lfloor t \rfloor + 1} - N_{\lfloor t \rfloor}}{\sqrt{\lambda \lfloor t \rfloor}} := \xi_{\lfloor t \rfloor}.$$

Then for $\epsilon > 0$,

$$\mathbf{P}(\xi_n > \epsilon) = \mathbf{P}(N_{n+1} - N_n > \epsilon\sqrt{\lambda n}) = \mathbf{P}(\mathcal{P}(1) > \epsilon\sqrt{\lambda n}) \xrightarrow[n \to +\infty]{} 0,$$

and the first term converges in probability towards 0. The result is thus obtained using Slutsky's theorem.

A functional Central Limit Theorem

- Let $(N_t)_{t\geq 0}$ be a (time-homogeneous) Poisson process with parameter $\lambda > 0$. For any $t, u \geq 0$, let us set
$$B_t^{(u)} = \frac{N_{tu} - \lambda tu}{\sqrt{\lambda u}}.$$
For any given $u \geq 0$, $(B_t^{(u)})_{t\geq 0}$ is a stochastic process with independent increments, and jump sizes $\frac{1}{\sqrt{\lambda u}}$.
- Given the previous result, for all $t \geq 0$,
$$B_t^{(u)} = \frac{N_{tu} - \lambda tu}{\sqrt{\lambda u}} \xrightarrow[u\to\infty]{\mathcal{L}} Z_t,$$
where Z_t is a Gaussian random variable with mean 0 and variance t.
- When $u \to +\infty$, we therefore obtain as a limit a Gaussian process, with independent increments, with continuous paths, with mean 0 and variance t, i.e. a Brownian motion.

Table of contents

Conditional intensity – Intuition I

- Conditional intensity provides a causal description of a point process (successive conditioning).
- Let $p_n(t|t_1, \ldots, t_{n-1})$, $t \geq t_{n-1}$, be the probability distribution function for T_n given $\{T_1 = t_1, \ldots, T_{n-1} = t_{n-1}\}$. The survival function is:

$$S_n(t) = \mathbf{P}(T_n > t|t_1, \ldots, t_{n-1}) = 1 - \int_{t_{n-1}}^{t} p_n(u|t_1, \ldots, t_{n-1})\, du.$$

- Let us set $h_n(t|t_1, \ldots, t_{n-1}) = \dfrac{p_n(t|t_1, \ldots, t_{n-1})}{S_n(t|t_1, \ldots, t_{n-1})}$. Then:

$$p_n(t|t_1, \ldots, t_{n-1}) = h_n(t|t_1, \ldots, t_{n-1}) e^{-\int_{t_{n-1}}^{t} h_n(u|t_1, \ldots, t_{n-1})\, du}.$$

- (Recall that in the time-homogeneous Poisson case $p_n(t|t_1, \ldots, t_{n-1}) = \lambda e^{-\lambda(t - t_{n-1})}$, $t \geq t_{n-1}$).

Conditional intensity – Intuition II

- Moreover,

$$\mathbf{E}[N(t_{n-1}+dt) - N(t_{n-1})|t_1,\dots,t_{n-1}] \approx h_n(t_{n-1}+|t_1,\dots,t_{n-1})dt.$$

- In summary, if we define a function λ^* successively on a sample path $(0 < t_1 < \dots < t_n < \dots)$ by

$$\lambda^*(t) = \begin{cases} h_1(t), & 0 < t \leq t_1, \\ h_n(t|t_1,\dots,t_{n-1}), & t_{n-1} < t \leq t_n, n \geq 2, \end{cases}$$

  then this function characterizes the probability structure of the point process (p_n, S_n expressed in terms of λ^*) and we may write the intuitive formula

$$\mathbf{E}[dN_t|\text{ history up to time } t] \approx \lambda^*(t)\,dt.$$

- λ^* is called the conditional intensity of the point process.

Conditional intensity of a counting process I

- General theory relies on the Doob-Meyer decomposition.
- Let $(N_t)_{t\geq 0}$ be an integrable counting process with generated filtration $(\mathcal{F}_t)_{t\geq 0}$. By the Doob-Meyer decomposition of a submartingale, there exists a unique increasing predictable process A, with $A_0 = 0$ a.s., such that $M_t = N_t - A_t$ is a martingale.
- If we assume further that $A_t = \int_0^t \lambda_s \, ds$ for some predictable process $(\lambda_s)_{s\geq 0}$, then $N_t - \int_0^t \lambda_s \, ds$ is a martingale.

Definition (Intensity of a counting process)

A (predictable) process $(\lambda_s)_{s\geq 0}$ is called the (predictable) intensity of a counting process $(N_t)_{t\geq 0}$ if $M_t = N_t - \int_0^t \lambda_s \, ds$ is a martingale.

- A is called a compensator of the counting process, and the martingale is called M the compensated counting process.

Conditional intensity of a counting process II

- For any $t, \epsilon \geq 0$, we thus have

$$\mathbf{E}[N_{t+\epsilon} - N_t \mid \mathcal{F}_t] = \mathbf{E}\left[\int_t^{t+\epsilon} \lambda_s \, ds \mid \mathcal{F}_t\right]$$

- Interpretation:

$$\lambda_{t_+}|\mathcal{F}_t = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \mathbf{E}[N_{t+\epsilon} - N_t | \mathcal{F}_t].$$

- Condensed notation:

$$\mathbf{P}[dN_t = 1 | \mathcal{F}_{t-}] = \mathbf{E}[dN_t | \mathcal{F}_{t-}] = \lambda_t \, dt.$$

Intensity of a Poisson process

Proposition (Intensity of a Poisson process)

Let N be a time-homogeneous Poisson process with parameter λ. N has a constant intensity process, equal to λ.

Sketch of the proof. $(N_t - \lambda t)_{t \geq 0}$ is a $\mathcal{F}_t$-martingale.

- Remark: we have
$$\begin{cases} \mathbf{P}(N_{t+h} - N_t = 0) = 1 - \lambda h + o(h), \\ \mathbf{P}(N_{t+h} - N_t = 1) = \lambda h + o(h), \\ \mathbf{P}(N_{t+h} - N_t \geq 2) = o(h). \end{cases}$$
(This could actually lead to yet another definition of the Poisson process.)

Table of contents

(Non homogeneous) Poisson processes I

Definition (Non-homogeneous Poisson process)

Let $\lambda : [0, \infty) \to \mathbb{R}_+$ be a locally integrable function. Let

$$\Lambda(t) = \int_0^t \lambda(s)\, ds, \quad t \geq 0.$$

A counting process N is called a Poisson process with intensity λ if

(i) N has independent increments ;

(ii) for all $0 \leq s \leq t$, the random variable $N_t - N_s$ is Poisson distributed with parameter $\Lambda(t) - \Lambda(s)$.

- Recall that we have assumed $N_0 = 0$ for all counting processes.
- λ is a deterministic function.

(Non homogeneous) Poisson processes II

- Increments are not stationary anymore (except of course in the constant λ case).
- Of course, $\mathbf{E}[N_t] = \Lambda(t)$ and $\mathbf{V}[N_t] = \Lambda(t)$ (Poisson distribution).
- If λ is continuous at t, then we have

$$\begin{cases} \mathbf{P}(N_{t+h} - N_t = 0) = 1 - \lambda(t)h + o(h), \\ \mathbf{P}(N_{t+h} - N_t = 1) = \lambda(t)h + o(h), \\ \mathbf{P}(N_{t+h} - N_t \geq 2) = o(h). \end{cases}$$

Sketch of the proof. $N_{t+h} - N_t$ is Poisson distributed with parameter $\int_t^{t+h} \lambda(s)\, ds$, so

$$\mathbf{P}(N_{t+h} - N_t = 0) = \exp\left(-\int_t^{t+h} \lambda(s)\, ds\right) = 1 - \lambda(t)h + o(h),$$

and similar computations for the other cases.

Time change for a Poisson process I

Proposition (Time change for a Poisson process)

Let $\lambda : \mathbb{R}_+ \to \mathbb{R}_+^*$ be a locally integrable function and for all $t \geq 0$, let $\Lambda(t) = \int_0^t \lambda(s)\, ds$. Let N be a time-homogeneous Poisson process with constant intensity 1. Then $\tilde{N}_t = N_{\Lambda(t)}$ is a non-homogeneous Poisson process with intensity $t \mapsto \lambda(t)$.

Sketch of the proof: Let $(T_n)_n$ be the jump times of process N, and let

$$S_n = \Lambda^{-1}(T_n) = \inf\{t \geq 0 : \Lambda(t) \geq T_n\}.$$

Then

$$\tilde{N}_t = N_{\Lambda(t)} = \sum_{i \geq 1} \mathbf{1}_{T_i \leq \Lambda(t)} = \sum_{i \geq 1} \mathbf{1}_{\Lambda^{-1}(T_i) \leq t} = \sum_{i \geq 1} \mathbf{1}_{S_i \leq t},$$

i.e. $\tilde{N}$ is the counting process associated to the event times $(S_i)_i$. Furthermore, for any $0 < t_1 < \ldots < t_n$, $\tilde{N}_{t_i} - \tilde{N}_{t_{i-1}} = N_{\Lambda(t_i)} - N_{\Lambda(t_{i-1})}$, hence increments are independent and Poisson distributed.

Time change for a Poisson process II

Proposition (Integrated intensity between event times)

Let $\tilde{N}$ be a non-homogeneous Poisson process with intensity $t \mapsto \lambda(t)$. Let $(S_i)_{i\geq 1}$ be the random times of counted events. Then the random variables $(\Lambda(S_i) - \Lambda(S_{i-1}))_{i\geq 2}$ are i.i.d. random variables, exponentially distributed with parameter 1.

Sketch of the proof: Using the above construction,

$$\Lambda(S_i) - \Lambda(S_{i-1}) = \Lambda(\Lambda^{-1}(T_i)) - \Lambda(\Lambda^{-1}(T_{i-1})) = T_i - T_{i-1} \sim \mathcal{E}(1).$$

- This can be used as a goodness-of-fit test of non-homogeneous Poisson processes to empirical data.

Table of contents

Time-homogeneous Poisson simulation

Simulation of a time-homogeneous Poisson process with parameter $\lambda > 0$ on $[0, T]$, for some horizon $T > 0$.

- Simulation using interarrival times
  1. $t = 0, i = 1$
  2. Generate $s \sim \mathcal{E}(\lambda)$ (i.e. generate $u \sim \mathcal{U}([0,1])$ and set $s = -\frac{1}{\lambda}\log(u)$).
  3. If $t + s \leq T$:
     3.1 Set $t_i = t + s$, $t = t + s$, $i = i + 1$
     3.2 Go to 2.
  4. Else: Return $(t_i)_{i \geq 1}$
- Simulation using order statistics
  1. Generate $n \sim \mathcal{P}(\lambda T)$
  2. Generate $u_i \sim \mathcal{U}([0, T])$, $i = 1, \dots, n$, i.i.d.
  3. Return $(t_i)_{i=1,\dots,n}$ the ordered sequence of $\{u_i, i = 1, \dots, n\}$.

  (Proof: The joint density of $(T_1, \dots, T_n)|N_t = n$ is $f(t_1, \dots, t_n|n) = n!t^{-n}$.)

Non homogeneous Poisson simulation

Simulation of a non homogeneous Poisson process on $[0, T]$, for some horizon $T > 0$, with deterministic intensity function $\lambda : [0, T] \to \mathbb{R}_+$, and integrated intensity $\Lambda : t \mapsto \int_0^t \lambda(s)\, ds$.

- Simulation using interval times We mimic the time-homogeneous case, except interarrival times are not exponentially distributed, but should be generated from

$$
\begin{aligned}
F(s) &= \mathbf{P}(T_n \leq t_{n-1} + s | T_{n-1} = t_{n-1}) = 1 - \mathbf{P}(T_n > t_{n-1} + s | T_{n-1} = t_{n-1}) \\
&= 1 - \mathbf{P}(N_{t_{n-1}+s} - N_{t_{n-1}} = 0 | T_{n-1} = t_{n-1}) = 1 - \mathbf{P}(N_{t_{n-1}+s} - N_{t_{n-1}} = 0) \\
&= 1 - \exp\left(-(\Lambda(t_{n-1} + s) - \Lambda(t_{n-1})\right).
\end{aligned}
$$

(Need for an efficient inversion of F)

- Simulation using time change
  1. Generate $(t_i)_{i=1,\dots,n}$ sample path of a time-homogeneous Poisson process with intensity 1.
  2. Return $(\Lambda^{-1}(t_i))_{i=1,\dots,n}$.

(Need for an efficient inversion of Λ)

Thinning for non-homogeneous Poisson simulation I

The previous simulations often fail to be either numerically efficient or practical. Thinning is a flexible method to simulate non-homogeneous Poisson processes (and more, see later).

Theorem (Thinning for Poisson processes)

Let N^* be a (non-homogeneous) Poisson process with intensity $\lambda^*(t)$, $t \in [0, \infty)$. Let $\lambda(t)$ be a deterministic positive function such that $\lambda(t) \leq \lambda^*(t)$, $t \geq 0$. Let $T_1^*, \ldots, T_{N_T^*}^*$ be the sequence of random points of N^* on $[0, T]$. For $i = 1, \ldots, N_T^*$, delete the point T_i^* with probability $1 - \frac{\lambda(T_i^*)}{\lambda^*(T_i^*)}$. Then the remaining points form a (non-homogeneous) Poisson process with intensity $\lambda(t)$ on $[0, T]$.

- Introduced in [4].

Thinning for non-homogeneous Poisson simulation II

Sketch of the proof. For any $t \in (a, b]$,

$$\mathbf{P}(T_1^* > t | N_b^* - N_a^* = 1) = \frac{\mathbf{P}(N_t^* - N_a^* = 0, N_b^* - N_t^* = 1)}{\mathbf{P}(N_b^* - N_a^* = 1)} = \frac{\Lambda^*(b) - \Lambda^*(t)}{\Lambda^*(b) - \Lambda^*(a)},$$

hence the density $f_{T_1^* | N_b^* - N_a^* = 1}(t) = \frac{\lambda^*(t)}{\Lambda^*(b) - \Lambda^*(a)}$. The probability to keep a (single) point in $(a, b]$ after thinning is thus

$$\int_a^b \frac{\lambda^*(t)}{\Lambda^*(b) - \Lambda^*(a)} \frac{\lambda(t)}{\lambda^*(t)} \, dt = \frac{\Lambda(b) - \Lambda(a)}{\Lambda^*(b) - \Lambda^*(a)} := p.$$

Generalizing to the case of keeping k points among n, one gets

$$\mathbf{P}(N_b - N_a = k | N_b^* - N_a^* = n) = \binom{n}{k} p^k (1-p)^{n-k},$$

and with a final computation

$$\mathbf{P}(N_b - N_a = k) = \sum_{k=n}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} e^{-(\Lambda^*(b) - \Lambda^*(a))} \frac{(\Lambda^*(b) - \Lambda^*(a))^n}{n!} = e^{-(\Lambda(b) - \Lambda(a))} \frac{(\Lambda(b) - \Lambda(a))^k}{k!}.$$

NHPP simulation by thinning

The previous theorem leads to new simulaton algorithms if for example λ is bounded by a constant $\lambda^* \in \mathbb{R}_+^*$ on $[0, T]$

- Global thinning
  1. Generate $(T_i^*)_{i=1,\ldots,n}$ a realization of a time-homogeneous Poisson process with parameter λ^*.
  2. Generate $u_i \sim \mathcal{U}([0,1])$, $i = 1, \ldots, n$.
  3. Return the set of points $\left\{ T_i^* : u_i \leq \frac{\lambda(T_i^*)}{\lambda^*} \right\}$.
- Thinning per interarrival time
  1. $t = 0, i = 1$
  2. Generate $s \sim \mathcal{E}(\lambda^*)$ (i.e. $s = -\frac{1}{\lambda^*} \log(u)$, $u \sim \mathcal{U}([0,1])$).
  3. If $t + s \leq T$:
     - 3.1 Generate $u \sim \mathcal{U}[0,1])$
     - 3.2 If $u \leq \frac{\lambda(t+s)}{\lambda^*}$: Set $t_i = t + s$, $i = i + 1$
     - 3.3 $t = t + s$
     - 3.4 Go to 2.
  4. Else: Return $(t_i)_{i \geq 1}$

Table of contents

Lab 2 - Questions I

1. **Non-homogeneous Poisson process and trade times.** Implement a thinning algorithm for the simulation of a non-homogeneous Poisson process on $[0, T]$ with a given intensity $\mu : [0, T] \to \mathbb{R}_+$. Make some tests to check that your simulation algorithm is correct. Propose/calibrate a deterministic intensity function f that represents the intensity of occurrence of trades during a trading day. Using simulations of a (non-homogeneous) Poisson process with intensity f, comment on the statistical properties of the fitted model and on the goodness-of-fit of this model.

Lab 2 - Questions II

2. **Brownian motions with Poisson sampling.** Consider two processes $p_i(t) = \sigma_i W_i(t)$, $i = 1, 2$, where W_1 and W_2 are two Brownian motions such that $\langle W_1, W_2 \rangle_t = \rho t$ for some constants $\sigma_1 > 0$, $\sigma_2 > 0$ and $\rho \in]-1, 1[$. Assume that for $i = 1, 2$ the process p_i is observed on $[0, T]$ at times $T_{i,k}$, $k \geq 0$, and that $(T_{i,k})_{k \geq 0}$ is a Poisson process with constant intensity $\mu_i > 0$. Such a framework could be used to model the trade times and log-trade prices for two correlated assets.
   Simulate multiple sample paths of the *observed* processes (trade times and prices) and on each path compute the standard covariance estimator of $[p_1, p_2]_t = \rho \sigma_1 \sigma_2 t$ at various sampling periods τ. Comment.

Lab 2 - Questions III

3. **Empirical intensities and LOB features.** The scientific literature on LOB suggests that intensities of order flows depend on observed LOB features. We will discuss this later in the course. On your dataset, compute empirical intensities of trades as a function, e.g., of the (rescaled) observed queue size or of the observed spread. Comment.

References

[1] P. Brémaud. *Point Processes and Queues, Martingale Dynamics*. Springer, 1981.

[2] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes: volume I: elementary theory and methods*. Springer, 2003.

[3] J. F. C. Kingman. *Poisson processes*. Clarendon Press, 1992.

[4] P. W. Lewis and G. S. Shedler. "Simulation of nonhomogeneous Poisson processes by thinning". In: *Naval research logistics quarterly* 26.3 (1979), pp. 403–413.