High-frequency data and limit order books

—

Lecture 6 - April 8th, 2025

Ioane Muni Toke

MICS Laboratory and Chair of Quantitative Finance
CentraleSupélec – Université Paris-Saclay, France

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

Continuous-time Markov chains

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space. Let E be a countable state space. Let $(X_t)_{t\geq 0}$ be a stochastic process with values in E.

Definition (Continuous-time Markov chains)

X is a continuous-time Markov chain if for all $s, t \geq 0$, $k \in \mathbb{N}^*$, $(s_1, \dots, s_k) \in [0, s)^k$, $(j, i, i_1, \dots, i_k) \in E^{k+2}$,

$$\mathbf{P}(X_{s+t} = j | X_s = i, X_{s_1} = i_1, \dots, X_{s_k} = i_k) = \mathbf{P}(X_{s+t} = j | X_s = i).$$

- Homogeneous case : $\mathbf{P}(X_{s+t} = j | X_s = i)$ does not depend on s.
- Distribution vector of the variable X_t : $\mu(t) = (\mathbf{P}(X_t = i))_{i\in E}$.
- Few elements recalled here in the standard case (countable E, finite activity/regular jumps), excluding pathological cases. For further reading: [2].

Transition semi-group I

Definition (Transition semi-group)

Let $p_{ij}(t) = \mathbf{P}(X_{s+t} = j | X_s = i)$ and $P(t) = (p_{ij}(t))_{i,j \in E}$.
P is called the transition semi-group of the continuous-time homogeneous Markov chain X (*see property (b) below*).

Proposition

(a) $P(t)$ *is a stochastic matrix ;* $P(0)$ *is the identity matrix ;*

(b) $P(t+s) = P(t)P(s)$ *(Chapman-Kolmogorov equations) ;*

(c) $\mu^T(t) = \mu^T(0)\, P(t)$ *;*

(d) *For* $0 < t_1 < \ldots < t_k$, $\mathbf{P}\left(\bigcap_{j=1}^{k} \{X_{t_j} = i_j\}\right) = \sum_{i_0 \in E} \mathbf{P}(X_0 = i_0) \prod_{j=1}^{k} p_{i_{j-1} i_j}(t_j - t_{j-1})$.

Transition semi-group II

Sketch of proof.

(a) $\forall i \in E$, $\sum_{j\in E} p_{ij}(t) = \sum_{j\in E} \mathbf{P}(X_t = j | X_0 = i) = 1$;

(b) $\forall i, j \in E$, $p_{ij}(t+s) = \mathbf{P}(X_{s+t} = j | X_0 = i)$
$= \sum_{k\in E} \mathbf{P}(X_{s+t} = j | X_t = k, X_0 = i)\mathbf{P}(X_t = k | X_0 = i) = \sum_{k\in E} p_{ik}(t) p_{kj}(s)$;

(c) $\mu_i(t) = P(X_t = i) = \sum_{k\in E} \mathbf{P}(X_t = i | X_0 = k)\mathbf{P}(X_0 = k)$
$= \sum_{k\in E} p_{ki}(t)\mu_k(0)$;

(d) $\mathbf{P}\left(\bigcap_{j=1}^{k}\{X_{t_j} = i_j\}\right) = \sum_{i_0\in E} \mathbf{P}(X_0 = i_0) \prod_{j=1}^{k} \mathbf{P}(X_{t_j} = i_j | X_{t_{j-1}} = i_{j-1}, \dots, X_{t_0} = i_0)$
$= \sum_{i_0\in E} \mathbf{P}(X_0 = i_0) \prod_{j=1}^{k} \mathbf{P}(X_{t_j} = i_j | X_{t_{j-1}} = i_{j-1})$
$= \sum_{i_0\in E} \mathbf{P}(X_0 = i_0) \prod_{j=1}^{k} p_{i_{j-1}i_j}(t_j - t_{j-1})$.

Infinitesimal generator

Definition (Infinitesimal generator)

Let $A = \lim_{h\to 0} \frac{P(h)-P(0)}{h}$. A is called the infinitesimal generator of the continuous-time homogeneous Markov chain X.

- Existence of $q_{ij} = \lim_{h\to 0} \frac{p_{ij}(h)}{h} \in \mathbb{R}_+$, $i \neq j$ and $q_i = -q_{ii} = \lim_{h\to 0} \frac{1-p_{ii}(h)}{h} \in \overline{\mathbb{R}_+}$ by assuming continuity of P at the origin.
- In standard cases (E finite or finite activity jumps), $q_i = \sum_{j\in E, j\neq i} q_{ij} \in \mathbb{R}_+$.
- Local behaviour :

$$\mathbf{P}(X_{t+h} = j | X_t = i) = q_{ij} h + o(h)$$
$$\mathbf{P}(X_{t+h} = i | X_t = i) = 1 - q_i h + o(h)$$

- If $X_t = i$, the duration to the next jump is exponentially distributed with parameter $\sum_{j\neq i} q_{ij} = q_i$. The probability that the next transition is from i to j is $\frac{q_{ij}}{q_i}$.

Kolmogorov and balance equations

Proposition (Kolmogorov equations)

P and A satisfy the Kolmogorov backward (resp. forward) equation

$$\frac{d}{dt}P(t) = AP(t), \qquad \left(\textit{resp.} \quad \frac{d}{dt}P(t) = P(t)A \right).$$

Sketch of proof. One may write $\frac{P(t+h)-P(t)}{h} = \frac{P(h)P(t)-P(t)}{h} = \frac{P(h)-I}{h}P(t)$ or $\frac{P(t+h)-P(t)}{h} = \frac{P(t)P(h)-P(t)}{h} = P(t)\frac{P(h)-I}{h}$, so that Kolmogorov equations are obtained by taking the limit when $h \to 0$. In standard cases (e.g., E finite or finite activity jumps), the limit is well-defined, otherwise technical conditions apply.

- If E is finite, then a solution is available as an exponential of matrix : $P(t) = e^{tA}$.

Balance equations

Proposition (Balance equation)

μ and A satisfy the balance equations

$$\frac{d}{dt}\mu^T(t) = \mu^T(t)A.$$

Sketch of proof. We have $\mu_i(t+h) = \mu_i(t)p_{ii}(h) + \sum_{j\neq i} \mu_j(t)p_{ji}(h)$, hence:

$$\frac{\mu_i(t+h) - \mu_i(t)}{h} = -\mu_i(t)\frac{1 - p_{ii}(h)}{h} + \sum_{j\neq i} \mu_j(t)\frac{p_{ji}(h)}{h},$$

and the result is obtained if the limit when $h \to 0$ exist.

Stationary distributions

Definition (Stationary distribution)

A stationary distribution is any probability π on E such that for all $t \geq 0$

$$\pi^T P(t) = \pi^T.$$

Proposition (Characterization of stationary distributions)

A probability π on E is a stationary distribution if and only if $\pi^T A = 0$.

A simple birth-and-death process, or M/M/1 queue

A time-homogeneous Poisson process is a continuous time Markov chain.

Exercise. Let X be the size of a queue. The arrival process of customers is Poisson with intensity λ. The service time is exponential with parameter μ.

1. Write the infinitesimal generator of X.
2. Compute the stationary distribution of X.

Assume now that the intensities depend on the value of the process : $(\lambda_n)_{n\in\mathbb{N}}$, $(\mu_n)_{n\in\mathbb{N}}$.

1. Write the infinitesimal generator of X.
2. Compute the stationary distribution of X.

First passage times of a birth-and-death process I

Let us consider a birth-and-death process with intensities depending on the state of the process : $(\lambda_n)_{n\in\mathbb{N}}$, $(\mu_n)_{n\in\mathbb{N}}$. Let $\sigma_{i,i-1}$ be the first passage time to state $i-1$ given that i is the current state.

Theorem (Laplace transform of first passage time)

Let $f_{i,i-1}$ be the probability density of $\sigma_{i,i-1}$ and $\hat{f}_{i,i-1}$ its Laplace transform. Then

$$\hat{f}_{i,i-1}(s) = \frac{\mu_i}{\lambda_i + \mu_i + s - \lambda_i \hat{f}_{i+1,i}(s)} = -\frac{1}{\lambda_{i-1}} \mathbf{\Phi}_{k=i}^{\infty} \frac{\lambda_{k-1}\mu_k}{\lambda_k + \mu_k + s}$$

where the term on the right-hand side denotes the corresponding generalized continuous fraction.

- The density function of $\sigma_{i,i-1}$ can be found by applying the inverse Laplace transform.
- There exists efficient numerical methods to invert such continuous fractions.

First passage times of a birth-and-death process II

Sketch of proof. Assume the process is in state i. Let $f_{i,i-1}$ the density of $\sigma_{i,i-1}$ (first passage time from state i to $i-1$) and $\hat{f}_{i,i-1}$ its Laplace transform ($\hat{f}_{i,i-1}(s) = \int_{\mathbb{R}} e^{-sx} f_{i,i-1}(x)\,dx = \mathbf{E}[e^{-s\sigma_{i,i-1}}]$). Let A the event "the next transition is to $i-1$". We have $\mathbf{P}(A) = \frac{\mu_i}{\lambda_i+\mu_i}$ and

$$\hat{f}_{i,i-1}(s) = \frac{\mu_i}{\lambda_i+\mu_i}\mathbf{E}[e^{-s\sigma_{i,i-1}} \mid A] + \frac{\lambda_i}{\lambda_i+\mu_i}\mathbf{E}[e^{-s\sigma_{i,i-1}} \mid A^c].$$

Since $\sigma_{i,i-1}|A$ has an exponential distribution $\mathcal{E}(\lambda_i+\mu_i)$ (next jump time), one can compute $\mathbf{E}[e^{-s\sigma_{i,i-1}} \mid A] = \frac{\lambda_i+\mu_i}{\lambda_i+\mu_i+s}$. Moreover,

$$\begin{aligned}
\mathbf{E}[e^{-s\sigma_{i,i-1}} \mid A^c] = \mathbf{E}[e^{-s(\sigma_{i,i+1}+\sigma_{i+1,i-1})} \mid A^c] &= \mathbf{E}[e^{-s\sigma_{i,i+1}} \mid A^c]\mathbf{E}[e^{-s\sigma_{i+1,i-1}} \mid A^c] \\
&= \frac{\lambda_i+\mu_i}{\lambda_i+\mu_i+s}\mathbf{E}[e^{-s\sigma_{i+1,i-1}}] = \frac{\lambda_i+\mu_i}{\lambda_i+\mu_i+s}\mathbf{E}[e^{-s\sigma_{i+1,i}}]\mathbf{E}[e^{-s\sigma_{i,i-1}}] \\
&= \frac{\lambda_i+\mu_i}{\lambda_i+\mu_i+s}\hat{f}_{i+1,i}(s)\hat{f}_{i,i-1}(s).
\end{aligned}$$

Finally, $\hat{f}_{i,i-1}(s) = \frac{1}{\lambda_i+\mu_i+s}(\mu_i + \lambda_i\hat{f}_{i+1,i}(s)\hat{f}_{i,i-1}(s))$, hence the result.

Table of contents

Table of contents

Stationary shape of a one-side LOB

[6] considers a one-side limit order book where 0 is the opposite reference price (e.g. bid price if we model the ask side), and all order flows are Poisson : intensity $\lambda_i \geq 0$ for limit orders at price i ; intensity $\mu > 0$ for market orders ; limit orders' lifetime is exponential with parameter $\theta > 0$.

Theorem (Average shape of the limit order book)

Let $L^k(t)$ denote the number of limit orders standing in the book at price k and time t. In this basic one-side Poisson model, the average shape of the LOB satisfies

$$\mathbf{E}[L^k] = \nu_{1\to k} - \nu_{1\to k-1} - \left(\frac{\Gamma_{\nu_{1\to k}}(1+\delta)}{\Gamma_{\nu_{1\to k}}(\delta)} - \frac{\Gamma_{\nu_{1\to k-1}}(1+\delta)}{\Gamma_{\nu_{1\to k-1}}(\delta)} \right),$$

with the normalized parameters $\nu_{1\to k} = \frac{\lambda_{1\to k}}{\theta}$ and $\delta = \frac{\mu}{\theta}$, and the lower incomplete version of the Euler gamma function $\Gamma_y : \mathbb{R}_+ \to \mathbb{R}, x \mapsto \int_0^y t^{x-1} e^{-t} dt$.

Extension with random sizes of limit orders

Extension: $(g_n^k)_{n\in\mathbb{N}^*}$: probability that a limit order at price k is of size n; falls into a family of queueing systems with bulk/batch arrival, which can still be studied with standard techniques.

Theorem (Shape of the LOB with random sizes of limit orders)

In the discrete one-side LOB model with random size with distribution $(g_n^k)_{n\in\mathbb{N}^*}$, the average cumulative shape $L^{1\to k}(t) = \sum_{i=1}^k L^i(t)$ of the order book up to price k is:

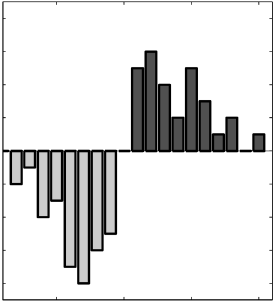
$$\mathbf{E}[L^{1\to k}] = \nu_{1\to k}\overline{g^{1\to k}} - \delta + \left(\int_0^1 v^{\delta-1} e^{\nu_{1\to k}\int_v^1 H^{1\to k}(u)\,du}\,dv\right)^{-1}.$$

- Proof: Infinitesimal generator of the process $L^{1\to k}$, balance equations, generating function.
- Explicit formulas for e.g., geometric distributions of the sizes of limit orders. Extension to the continuous case.

Table of contents

Model description

- [3] introduce an order book modeled as a Markov chain / queue
- Price constrained on a fixed grid $\{1, \ldots, n\}$
- State $X(t) = (X_1(t), \ldots, X_n(t))$, where $|X_i(t)|$ is the number of shares standing in the book at price i and time t.



- Limit orders at price p : $\lambda(p_A(t) - p) = \frac{k}{(p_A(t)-p)^\alpha}$ on the bid side ($p < p_A(t)$), and $\lambda(p - p_B(t)) = \frac{k}{(p-p_B(t))^\alpha}$ on the ask side ($p > p_B(t)$).
- Market orders : with intensity μ
- Cancellations at price p with intensity $\theta(\delta)|X_p(t)|$.

Source: [3]

Dynamics of the limit order book

- Notation : for $x \in \mathbb{Z}^n$, $x^{p\pm 1} = x \pm (0, \dots, 0, 1, 0, \dots, 0)$ where the element 1 is in p-th position.
- With the orders described above, X is a continuous-time homogeneous Markov chain with transition rates at time t and state x:

$$\begin{cases} x \to x^{p-1} & \text{with rate } \lambda(p_A(t) - p) \text{ if } p < p_A(t), \\ x \to x^{p+1} & \text{with rate } \lambda(p - p_B(t)) \text{ if } p > p_B(t), \\ x \to x^{p_B(t)+1} & \text{with rate } \mu, \\ x \to x^{p_A(t)-1} & \text{with rate } \mu, \\ x \to x^{p+1} & \text{with rate } \theta(p_A(t) - p)|x_p| \text{ if } p < p_A(t), \\ x \to x^{p-1} & \text{with rate } \theta(p - p_B(t))|x_p| \text{ if } p > p_B(t). \end{cases}$$

- Observe that X stays a well-defined limit order book after all transitions.

Some analytical results I

Theorem (Probability of the next price move)

Let us consider the LOB in the state $X_A(t) = a$ and $X_B(t) = b$. If the spread is equal to one tick, then the probability that the next mid-price movement will be an upward movement is given by the inverse Laplace transform of $\frac{1}{s}\hat{f}_a(s)\hat{f}_b(-s)$ evaluated at 0, where

$$\hat{f}_j(s) = \left(-\frac{1}{\lambda(1)}\right)^j \prod_{i=1}^{j} \Phi_{k=i}^{\infty} \frac{-\lambda(1)(\mu+k\theta(1))}{\lambda(1)+\mu+k\theta(1)+s}.$$

Proof: Analysis of the first passage times to 0 of the best quotes processes with Laplace transforms. Φ denotes a continuous fraction.

Some analytical results II

- Empirical (left) and analytical (right) probabilities of a mid-price increase when the spread is one:

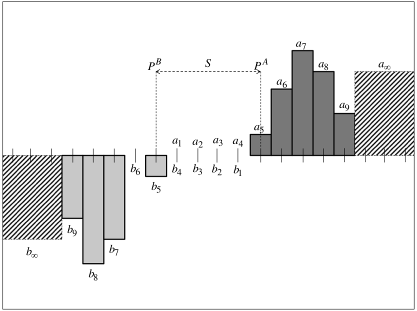
b \ a	1	2	3	4	5
1	.512	.304	.263	.242	.226
2	.691	.502	.444	.376	.359
3	.757	.601	.533	.472	.409
4	.806	.672	.580	.529	.484
5	.822	.731	.640	.714	.606

b \ a	1	2	3	4	5
1	.500	.336	.259	.216	.188
2	.664	.500	.407	.348	.307
3	.741	.593	.500	.437	.391
4	.784	.652	.563	.500	.452
5	.812	.693	.609	.548	.500

- Further analytical results can be derived, e.g.
  - Mid-price moves for any spread
  - Probability of execution before the next price move
  - Making the spread
  - ...
- Other analytical results using queueing theory: analytic expressions of the average shape of a Poisson LOB with non-deterministic order sizes [5].

Table of contents

A Poisson LOB with a moving frame



Source:[1]

- [1] introduces a fixed-size sliding window instead of a large price grid: for a given $K \in \mathbb{N}^*$, the LOB at time t is represented by the vector

$$X(t) = (a_1(t), \ldots, a_K(t); b_1(t), \ldots, b_K(t)),$$

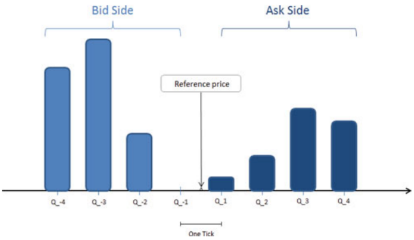
where $a_i(t)$ (resp. $b_i(t)$) is the number of shares available i ticks away from the opposite best bid (resp. ask) quote.

- Market orders, limit orders and cancellations are independent Poisson processes with constant intensities $\lambda_i^{M\pm}$, $\lambda_i^{L\pm}$ and $\lambda_i^{C\pm}$.

- Assuming cancellations are “sufficient”, ergodicity is proven. The ergodicity proof has been used in more complex settings since (including for Hawkes processes).

Table of contents

The Queue-Reactive model I



Source:[4]

- [4] propose several extensions to the previous pure Poisson LOB model.
- LOB is represented by the vector

$$X(t)) = (q_{-K}(t), \dots, q_{-1}(t), q_1(t), \dots, q_K(t)),$$

where q_i is the size of the i-th ask limit if $i > 0$, and the size of the i-th bid limit if $i < 0$. Limits are numbered around a reference price (e.g. mid-price).

- Now intensities of limit orders, markets orders and cancellations depend on the queue sizes: when $q_i = n$, these intensities are $\lambda^L_{\pm i}(n)$, $\lambda^M_{\pm i}(n)$ and $\lambda^C_{\pm i}(n)$.
- Dependence between the queues must be added to properly model market orders.

The Queue-Reactive model II

- Empirical measure of the intensities (see also a previous Lab).
- Monte Carlo simulation to derive applications on probability of execution, order placement or price impact.
- This line of modeling is still quite active in 2025.

Source:[4]

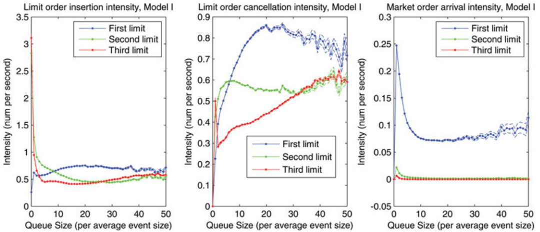


Figure 2. Intensities at $Q_{\pm i}, i = 1, 2, 3$, France Telecom.

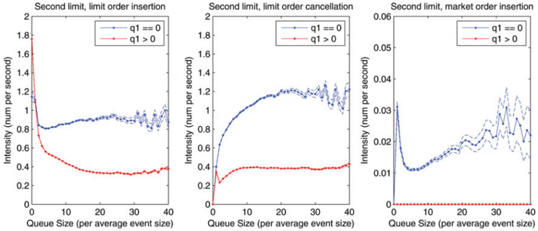


Figure 4. Intensities at Q_2 as functions of $1_{q_1>0}$ and q_2, France Telecom.

Table of contents

The "perfect market-making" case

- Prices in ticks (unconstrained, $\in \mathbb{Z}$); q: size of unit orders ; δ: unit step of the mid-price grid.
- Initial state: all limits above $p^A(0)$ and below $p^B(0)$ are filled with one limit order of unit size q; initial spread is equal to 1 tick.
- Market orders: two independent Poisson processes $N^A(t)$ (buy orders, affecting the ask side) and $N^B(t)$ (sell orders, affecting the bid side) with constant intensities λ^A and λ^B;
- Perfect market-making: one liquidity provider reacts *immediately* after a market order arrives so as to maintain the spread constantly equal to 1 tick. Liquidity is provided on the side affected by the market order (no mid-price change) with probability $1-u$ and on the opposite side with probability u (mid-price change).

Diffusive limit: the generator view I

- The mid-price dynamics is

$$dP(t) = \delta\left(dN^+(t) - dN^-(t)\right),$$

where N^+ and N^- are independent Poisson processes with intensity $\lambda^+ = u\lambda^A$ and $\lambda^- = u\lambda^B$.

- Infinitesimal generator: for some function $f : \delta\mathbb{Z} \to \mathbb{R}$,

$$\mathcal{A}f(p) = \lim_{t\to 0} \frac{\mathbf{E}[f(P_t)|P_0 = p] - f(p)}{t} = \lambda^+\ \left(f(p+\delta) - f(p)\right) + \lambda^-\ \left(f(p-\delta) - f(p)\right).$$

- (Remark: Matrix form of the infinitesimal generator of the continuous-time Markov chain

$$\begin{pmatrix} \ddots & \ddots & \ddots & & \ddots & & \\ & 0 & \lambda^- & & -(\lambda^- + \lambda^+) & \lambda^+ & 0 \\ & & \ddots & & \ddots & \ddots & \ddots & \ddots \end{pmatrix})$$

Diffusive limit: the generator view II

- Observe that

$$\mathcal{A}f(p) = \frac{1}{2}\left(\lambda^+ + \lambda^-\right)\delta^2 \frac{f(p+\delta) - 2f(p) + f(p-\delta)}{\delta^2} + \left(\lambda^+ - \lambda^-\right)\delta \frac{f(p+\delta) - f(p-\delta)}{2\delta},$$

- Scaling:

$$\lim_{\delta\to 0} (\lambda^+ + \lambda^-)\delta^2 = \sigma^2 \in \mathbb{R}_+,$$

$$\lim_{\delta\to 0} (\lambda^+ - \lambda^-)\delta = \mu \in \mathbb{R}.$$

- Continuous limit

$$\mathcal{L}f(p) = \frac{\sigma^2}{2}\frac{\partial^2 f}{\partial p^2} + \mu \frac{\partial f}{\partial p}$$

Diffusive limit: the generator view III

- We obtain a diffusive limit for the mid-price at a large-scale: a Brownian motion with drift μ and volatility σ (cf. Bachelier model)
- Obviously, volatility increases with trading activity, and a symmetric Poisson model yields a model with a Brownian motion without drift.
- Local volatility models can be obtained similarly with price dependent intensities.

Reminder: Infinitesimal generator of an Ito diffusion I

- Let $(X_t^x)_{t\in[0,T]}$ be the (time-homogeneous) Itō diffusion satisfying

$$dX_t = a(X_t)\,dt + b(X_t)\,dB_t,$$

with initial state $X_0^x = x$, and where a and b satisfy the standard conditions.

Definition (Generator of an Itō diffusion)

The infinitesimal generator of a time-homogeneous Itō diffusion is the functional operator A defined by

$$\mathcal{A}f(x) = \lim_{t\to 0} \frac{\mathbf{E}[f(X_t^x)] - f(x)}{t},$$

where the limit exists for f and x.

Reminder: Infinitesimal generator of an Ito diffusion II

- Itō’s lemma applied to some (twice differentiable) function $f : \mathbb{R} \to \mathbb{R}$ yields

$$f(X_t^x) = f(x) + \int_0^t \Big[\frac{\partial f}{\partial x}(X_t^x)\left[a(X_t^x)\,dt + b(X_t^x)\,dB_t\right] + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(X_t^x) b^2(X_t^x)\,dt\Big],$$

- and by taking the expectation,

$$\mathbf{E}[f(X_t^x)] = f(x) + \mathbf{E}\left[\int_0^t \frac{\partial f}{\partial t}(X_t^x) b(X_t^x)\,dB_t\right] + \mathbf{E}\left[\int_0^t \left(a(X_t^x)\frac{\partial f}{\partial x}(X_t^x) + \frac{1}{2}b^2(X_t^x)\frac{\partial^2 f}{\partial x^2}(X_t^x)\right)\,dt\right].$$

Reminder: Infinitesimal generator of an Ito diffusion III

- If we assume that f is nice enough, e.g. that its derivatives are bounded, then $\int_0^t \frac{\partial f}{\partial t}(X_t^x) b(X_t^x)\, dB_t$ is a martingale, so that its expectation vanishes in the previous equation.
- In other words, we have (for e.g., f twice differentiable with bounded derivatives)

$$\mathcal{A}f(x) = a(x)\frac{\partial f}{\partial x}(x) + \frac{1}{2}b^2(x)\frac{\partial^2 f}{\partial x^2}(x).$$

(A functional CLT for martingales)

Theorem (Functional CLT for martingales)

Let $(M_n)_{n\geq 1}$ be a sequence of local martingales in the Skorokhod space $\mathcal{D}$. For all $n \geq 1$, $(M_n(t))_{t\geq 0}$ is a martingale with respect to the filtration $(\mathcal{F}_{n,t})_{t\geq 0}$ and satisfies $M_n(0) = 0$. Let $c \in \mathbb{R}_+^*$. If the following assumptions hold:

(i) *for any horizon, the expectation of the maximum jump of M_n is asymptotically negligible;*

(ii) $[M_n]_t \Rightarrow ct$ *in* $\mathbb{R}$ *as* $n \to +\infty$,

Then $M_n \Rightarrow \sqrt{c}W$ in $\mathcal{D}$ as $n \to +\infty$, where W is a standard Brownian motion.

Proof: See [7]. Alternative formulation for locally square-integrable martingales and predictable quadratic variations. Generalization in finite dimension with a covariance matrix C.

Diffusive limit: the CLT view I

- Mid-price dynamics: $P_t = \delta(N_t^+ - N_t^-)$.
- Standard computations

$$\mathbf{E}[P_t] = \delta(\lambda^+ - \lambda^-)t,$$

$$\mathbf{E}[P_t - \delta(\lambda^+ - \lambda^-)t|\mathcal{F}_s] = P_s - \delta(\lambda^+ - \lambda^-)s,$$

$$\mathbf{V}[P_t] = \delta^2 \left(\mathbf{V}[N_t^+] + \mathbf{V}[N_t^-]\right) = \delta^2(\lambda^+ + \lambda^-)t.$$

- $(P_t - \delta(\lambda^+ - \lambda^-)t)_{t\geq 0}$ is a square-integrable martingale.

Diffusive limit: the CLT view II

- Let $M_n(t) = \frac{P(nt)-\delta(\lambda^+-\lambda^-)nt}{\delta\sqrt{\lambda^++\lambda^-}\sqrt{n}}$.
- The sequence M_n satisfies the assumptions of the FCLT for martingales:
  - jump size is $\frac{\delta}{\sqrt{\delta^2(\lambda^++\lambda^-)n}} \to_{n\to+\infty} 0$;
  - and $[M_n]_t = \sum_{0<s\leq t}(\Delta M_n(s))^2 = \frac{1}{\delta^2(\lambda^++\lambda^-)n}\sum_{0<s\leq nt}\delta^2(\Delta N^+(s) + \Delta N^-(s))^2 = \frac{1}{(\lambda^++\lambda^-)n}\mathcal{P}((\lambda^+ + \lambda^-)nt) \to t$.
- Therefore, M_n converges weakly in $\mathcal{D}$ towards a standard Brownian motion.
- With the appropriate scaling, one retrieves a diffusive limit with drift $\mu = \delta\left(\lambda^+ - \lambda^-\right)$ and volatility $\sigma = \delta\sqrt{(\lambda^+ + \lambda^-)}$.
- Possible extension to local volatility models if the limits μ and σ depend on p and t (e.g. if the intensities are functions of p and t).

References I

[1] F. Abergel and A. Jedidi. "A mathematical approach to order book modeling". In: *International Journal of Theoretical and Applied Finance* 16.05 (2013), p. 1350025.

[2] P. Brémaud. *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*. Springer, 1999.

[3] R. Cont, S. Stoikov, and R. Talreja. "A stochastic model for order book dynamics". In: *Operations research* 58.3 (2010), pp. 549–563.

[4] W. Huang, C.-A. Lehalle, and M. Rosenbaum. "Simulating and analyzing order book data: The queue-reactive model". In: *Journal of the American Statistical Association* 110.509 (2015), pp. 107–122.

[5] I. Muni Toke. "Stationary distribution of the volume at the best quote in a Poisson order book model". In: *International Journal of Theoretical and Applied Finance* 20.06 (2017), p. 1750039.

References II

[6] I. Muni Toke. "The order book as a queueing system: average depth and influence of the size of limit orders". In: *Quantitative Finance* 15.5 (2015), pp. 795–808.

[7] W. Whitt et al. "Proofs of the martingale FCLT". In: *Probability Surveys* 4 (2007), pp. 268–302.