Stochastic Approximation Algorithms Applied to Optimal Execution

Sophie LARUELLE

LAMA-UPEC

April 3, 2026

Outline

1. Introduction to stochastic approximation
   - Motivations and framework
   - *A.s.* convergence
   - Rates of Convergence
   - Averaging principle
2. Optimal split of orders across liquidity pools
   - Modeling and mean execution cost
   - Design of the stochastic algorithm
   - Numerical experiments
3. Optimal posting price of limit orders: learning by trading
   - Execution process and intensity
   - Design of the stochastic algorithm
   - Cost function and its derivative as expectations
   - Numerical experiments

Outline

Outline

Motivations

In many fields, we often are confronted to optimization or zero search problems.

Motivations

In many fields, we often are confronted to optimization or zero search problems.

Examples

- In *Finance*: extraction of implicit parameters (volatility), calibration, optimization of exogenous parameter for variance reduction (regression, Importance sampling, etc);

Motivations

In many fields, we often are confronted to optimization or zero search problems.

Examples

- In *Finance*: extraction of implicit parameters (volatility), calibration, optimization of exogenous parameter for variance reduction (regression, Importance sampling, etc);
- in *Economics*: cost minimization, profit optimization;

Motivations

In many fields, we often are confronted to optimization or zero search problems.

Examples

- In *Finance*: extraction of implicit parameters (volatility), calibration, optimization of exogenous parameter for variance reduction (regression, Importance sampling, etc);
- in *Economics*: cost minimization, profit optimization;
- in *Physics*: search of equilibrium points of a system (potential minima);

Motivations

In many fields, we often are confronted to optimization or zero search problems.

Examples

- In *Finance*: extraction of implicit parameters (volatility), calibration, optimization of exogenous parameter for variance reduction (regression, Importance sampling, etc);
- in *Economics*: cost minimization, profit optimization;
- in *Physics*: search of equilibrium points of a system (potential minima);
- in *Statistics*: search of parameter estimator (maximum likehood, least squares, other contrast minimization);

Motivations

In many fields, we often are confronted to optimization or zero search problems.

Examples

- In *Finance*: extraction of implicit parameters (volatility), calibration, optimization of exogenous parameter for variance reduction (regression, Importance sampling, etc);
- in *Economics*: cost minimization, profit optimization;
- in *Physics*: search of equilibrium points of a system (potential minima);
- in *Statistics*: search of parameter estimator (maximum likehood, least squares, other contrast minimization);
- in *neural network*: search of the weights and the biases which minimize the cost function.

Motivations

In many fields, we often are confronted to optimization or zero search problems.

Examples

- In *Finance*: extraction of implicit parameters (volatility), calibration, optimization of exogenous parameter for variance reduction (regression, Importance sampling, etc);
- in *Economics*: cost minimization, profit optimization;
- in *Physics*: search of equilibrium points of a system (potential minima);
- in *Statistics*: search of parameter estimator (maximum likehood, least squares, other contrast minimization);
- in *neural network*: search of the weights and the biases which minimize the cost function.

Common point: all the functions have a representation as an expectation, *i.e.*

$$h(\theta) = \mathbb{E}\left[H(\theta, Y)\right], \quad \text{where } Y \text{ is a } q\text{-dimensional random vector.}$$

Motivations

In many fields, we often are confronted to optimization or zero search problems.

Examples

- In *Finance*: extraction of implicit parameters (volatility), calibration, optimization of exogenous parameter for variance reduction (regression, Importance sampling, etc);
- in *Economics*: cost minimization, profit optimization;
- in *Physics*: search of equilibrium points of a system (potential minima);
- in *Statistics*: search of parameter estimator (maximum likehood, least squares, other contrast minimization);
- in *neural network*: search of the weights and the biases which minimize the cost function.

Common point: all the functions have a representation as an expectation, *i.e.*

$$h(\theta) = \mathbb{E}\left[H(\theta, Y)\right], \quad \text{where } Y \text{ is a } q\text{-dimensional random vector.}$$

The stochastic approximation is a tool based on simulation to solve optimization or zeros search problems.

Framework

Consider the following recursive procedure defined on a filtered probabilistic space $(\Omega, \mathcal{A}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ having values in a convex set $C \subset \mathbb{R}^d$,

$$\forall\, n \geq 0, \quad \theta_{n+1} = \theta_n - \gamma_{n+1} h(\theta_n) + \gamma_{n+1} \left(\Delta M_{n+1} + r_{n+1}\right),$$

Framework

Consider the following recursive procedure defined on a filtered probabilistic space $(\Omega, \mathcal{A}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ having values in a convex set $C \subset \mathbb{R}^d$,

$$\forall\, n \geq 0, \quad \theta_{n+1} = \theta_n - \gamma_{n+1} h(\theta_n) + \gamma_{n+1} \left(\Delta M_{n+1} + r_{n+1}\right),$$

where

- $(\gamma_n)_{n\geq 1}$ is a $(0, \bar{\gamma}]$-valued step sequence for some $\bar{\gamma} > 0$,

Framework

Consider the following recursive procedure defined on a filtered probabilistic space $(\Omega, \mathcal{A}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ having values in a convex set $C \subset \mathbb{R}^d$,

$$\forall\, n \geq 0, \quad \theta_{n+1} = \theta_n - \gamma_{n+1} h(\theta_n) + \gamma_{n+1} \left(\Delta M_{n+1} + r_{n+1}\right),$$

where

- $(\gamma_n)_{n\geq 1}$ is a $(0, \bar{\gamma}]$-valued step sequence for some $\bar{\gamma} > 0$,
- $h : C \to \mathbb{R}^d$ is a continuous function with linear growth such that

$$(I_d - \gamma h)(C) \subset C \quad \text{for every} \quad \gamma \in (0, \bar{\gamma}],$$

Framework

Consider the following recursive procedure defined on a filtered probabilistic space $(\Omega, \mathcal{A}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ having values in a convex set $C \subset \mathbb{R}^d$,

$$\forall\, n \geq 0, \quad \theta_{n+1} = \theta_n - \gamma_{n+1} h(\theta_n) + \gamma_{n+1}\left(\Delta M_{n+1} + r_{n+1}\right),$$

where

- $(\gamma_n)_{n\geq 1}$ is a $(0, \bar{\gamma}]$-valued step sequence for some $\bar{\gamma} > 0$,
- $h : C \to \mathbb{R}^d$ is a continuous function with linear growth such that

$$(I_d - \gamma h)(C) \subset C \quad \text{for every} \quad \gamma \in (0, \bar{\gamma}],$$

- θ_0 is a $\mathcal{F}_0$-measurable finite random vector,

Framework

Consider the following recursive procedure defined on a filtered probabilistic space $(\Omega, \mathcal{A}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ having values in a convex set $C \subset \mathbb{R}^d$,

$$\forall\, n \geq 0, \quad \theta_{n+1} = \theta_n - \gamma_{n+1} h(\theta_n) + \gamma_{n+1} \left(\Delta M_{n+1} + r_{n+1}\right),$$

where

- $(\gamma_n)_{n\geq 1}$ is a $(0, \bar{\gamma}]$-valued step sequence for some $\bar{\gamma} > 0$,
- $h : C \to \mathbb{R}^d$ is a continuous function with linear growth such that

$$(I_d - \gamma h)(C) \subset C \quad \text{for every} \quad \gamma \in (0, \bar{\gamma}],$$

- θ_0 is a $\mathcal{F}_0$-measurable finite random vector,
- for every $n \geq 1$, ΔM_n is a $\mathcal{F}_n$-martingale increment,

Framework

Consider the following recursive procedure defined on a filtered probabilistic space $(\Omega, \mathcal{A}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ having values in a convex set $C \subset \mathbb{R}^d$,

$$\forall\, n \geq 0, \quad \theta_{n+1} = \theta_n - \gamma_{n+1} h(\theta_n) + \gamma_{n+1} \left(\Delta M_{n+1} + r_{n+1}\right),$$

where

- $(\gamma_n)_{n\geq 1}$ is a $(0, \bar{\gamma}]$-valued step sequence for some $\bar{\gamma} > 0$,
- $h : C \to \mathbb{R}^d$ is a continuous function with linear growth such that

$$(I_d - \gamma h)(C) \subset C \quad \text{for every} \quad \gamma \in (0, \bar{\gamma}],$$

- θ_0 is a $\mathcal{F}_0$-measurable finite random vector,
- for every $n \geq 1$, ΔM_n is a $\mathcal{F}_n$-martingale increment,
- r_n is a $\mathcal{F}_n$-adapted remainder term.

Outline

Convergence: The martingale approach

Assume there exists an essentially quadratic function $L : \mathbb{R}^d \to \mathbb{R}_+$ such that $\|h\|^2 \leq K(1+L)$ and

$$\forall \theta \neq \theta^*, \quad \langle \nabla L \mid h \rangle (\theta) > 0.$$

Convergence: The martingale approach

Assume there exists an essentially quadratic function $L : \mathbb{R}^d \to \mathbb{R}_+$ such that $\|h\|^2 \leq K(1+L)$ and

$$\forall \theta \neq \theta^*, \quad \langle \nabla L \mid h \rangle (\theta) > 0.$$

Assume that $\theta_0 \in \mathcal{F}_0$, $L(\theta_0) < +\infty$ *a.s.* and the nonnegative step sequence $(\gamma_n)_{n \geq 1}$ satisfies

$$\sum_{n \geq 1} \gamma_n = +\infty \quad \text{and} \quad \sum_{n \geq 1} \gamma_n^2 < +\infty.$$

Convergence: The martingale approach

Assume there exists an essentially quadratic function $L : \mathbb{R}^d \to \mathbb{R}_+$ such that $\|h\|^2 \leq K(1+L)$ and

$$\forall \theta \neq \theta^*, \quad \langle \nabla L \mid h \rangle (\theta) > 0.$$

Assume that $\theta_0 \in \mathcal{F}_0$, $L(\theta_0) < +\infty$ *a.s.* and the nonnegative step sequence $(\gamma_n)_{n\geq 1}$ satisfies

$$\sum_{n\geq 1} \gamma_n = +\infty \quad \text{and} \quad \sum_{n\geq 1} \gamma_n^2 < +\infty.$$

Furthermore, assume that

$$\mathbb{E}\left[\|\Delta M_{n+1}\|^2 \mid \mathcal{F}_n\right] \leq C\left(1 + L(\theta_n)\right) \quad \text{and} \quad \sum_{n\geq 1} \gamma_n \mathbb{E}\left[\|r_n\|^2 \mid \mathcal{F}_{n-1}\right] < +\infty.$$

Convergence: The martingale approach

Assume there exists an essentially quadratic function $L : \mathbb{R}^d \to \mathbb{R}_+$ such that $\|h\|^2 \leq K(1+L)$ and

$$\forall \theta \neq \theta^*, \quad \langle \nabla L \mid h \rangle (\theta) > 0.$$

Assume that $\theta_0 \in \mathcal{F}_0$, $L(\theta_0) < +\infty$ *a.s.* and the nonnegative step sequence $(\gamma_n)_{n\geq 1}$ satisfies

$$\sum_{n\geq 1} \gamma_n = +\infty \quad \text{and} \quad \sum_{n\geq 1} \gamma_n^2 < +\infty.$$

Furthermore, assume that

$$\mathbb{E}\left[\|\Delta M_{n+1}\|^2 \mid \mathcal{F}_n\right] \leq C\left(1 + L(\theta_n)\right) \quad \text{and} \quad \sum_{n\geq 1} \gamma_n \mathbb{E}\left[\|r_n\|^2 \mid \mathcal{F}_{n-1}\right] < +\infty.$$

Then,

$$\theta_n \xrightarrow[n\to\infty]{a.s.} \theta^*.$$

Convergence: The ODE method

We consider the differential system $ODE_h \equiv \dot{x} = -h(x)$ associated to the continuous *mean field* $h : C \to \mathbb{R}^d$.

Convergence: The ODE method

We consider the differential system $ODE_h \equiv \dot{x} = -h(x)$ associated to the continuous *mean field* $h : C \to \mathbb{R}^d$. We assume that this system has a C-valued flow $\Phi(t, \xi)_{t \in \mathbb{R}_+, \xi \in C}$.

Convergence: The ODE method

We consider the differential system $ODE_h \equiv \dot{x} = -h(x)$ associated to the continuous *mean field* $h : C \to \mathbb{R}^d$. We assume that this system has a C-valued flow $\Phi(t,\xi)_{t\in\mathbb{R}_+,\xi\in C}$.This flow exists as soon as h is locally Lipschitz with linear growth.

Convergence: The ODE method

We consider the differential system $ODE_h \equiv \dot{x} = -h(x)$ associated to the continuous *mean field* $h : C \to \mathbb{R}^d$. We assume that this system has a C-valued flow $\Phi(t,\xi)_{t\in\mathbb{R}_+,\xi\in C}$.This flow exists as soon as h is locally Lipschitz with linear growth.

Let K be a compact connected, flow invariant subset of C, *i.e.* such that $\Phi(t,K) \subset K$ for every $t \in \mathbb{R}_+$.

Convergence: The ODE method

We consider the differential system $ODE_h \equiv \dot{x} = -h(x)$ associated to the continuous *mean field* $h : C \to \mathbb{R}^d$. We assume that this system has a C-valued flow $\Phi(t,\xi)_{t\in\mathbb{R}_+,\xi\in C}$.This flow exists as soon as h is locally Lipschitz with linear growth.

Let K be a compact connected, flow invariant subset of C, *i.e.* such that $\Phi(t,K) \subset K$ for every $t \in \mathbb{R}_+$.

A non-empty subset $A \subset K$ is an internal attractor of K for ODE_h if

- $A \subsetneq K$,
- $\exists\, \varepsilon_0 > 0$ s.t. $\sup_{x\in K, \mathrm{dist}(x,A)\le\varepsilon_0} \mathrm{dist}\big(\Phi(t,x),A\big) \to 0$ as $t \to +\infty$.

Convergence: The ODE method

We consider the differential system $ODE_h \equiv \dot{x} = -h(x)$ associated to the continuous *mean field* $h : C \to \mathbb{R}^d$. We assume that this system has a C-valued flow $\Phi(t,\xi)_{t\in\mathbb{R}_+,\xi\in C}$.This flow exists as soon as h is locally Lipschitz with linear growth.

Let K be a compact connected, flow invariant subset of C, $i.e.$ such that $\Phi(t,K) \subset K$ for every $t \in \mathbb{R}_+$.

A non-empty subset $A \subset K$ is an internal attractor of K for ODE_h if

- $A \subsetneq K$,
- $\exists\, \varepsilon_0 > 0$ s.t. $\sup_{x\in K, \mathrm{dist}(x,A)\le\varepsilon_0} \mathrm{dist}\big(\Phi(t,x), A\big) \to 0$ as $t \to +\infty$.

A compact connected flow invariant set K is a minimal attractor for ODE_h if it contains no internal attractor. This terminology coming from dynamical systems may be misleading: Thus any equilibrium point of ODE_h (zero of h) is a minimal attractor by this definition, regardless of its stability.

Assume that ODE_h has a C-valued flow. Assume furthermore that

$$r_n \xrightarrow[n\to\infty]{a.s.} 0 \quad \text{and} \quad \sup_{n\geq n_0} \mathbb{E}\left[\|\Delta M_{n+1}\|^2 \mid \mathcal{F}_n\right] < +\infty \quad a.s.,$$

Assume that ODE_h has a C-valued flow. Assume furthermore that

$$r_n \xrightarrow[n\to\infty]{a.s.} 0 \quad \text{and} \quad \sup_{n\geq n_0} \mathbb{E}\left[\left\|\Delta M_{n+1}\right\|^2 \mid \mathcal{F}_n\right] < +\infty \quad a.s.,$$

and that $(\gamma_n)_{n\geq 1}$ is a positive sequence satisfying $(\gamma_n \in (0, \bar{\gamma}],\ n \geq 1)$ and

$$\sum_{n\geq 1} \gamma_n = +\infty \quad \text{and} \quad \sum_{n\geq 1} \gamma_n^2 < +\infty.$$

Assume that ODE_h has a C-valued flow. Assume furthermore that

$$r_n \xrightarrow[n\to\infty]{a.s.} 0 \quad \text{and} \quad \sup_{n\geq n_0} \mathbb{E}\left[\|\Delta M_{n+1}\|^2 \mid \mathcal{F}_n\right] < +\infty \quad a.s.,$$

and that $(\gamma_n)_{n\geq 1}$ is a positive sequence satisfying $(\gamma_n \in (0, \bar{\gamma}],\ n \geq 1)$ and

$$\sum_{n\geq 1} \gamma_n = +\infty \quad \text{and} \quad \sum_{n\geq 1} \gamma_n^2 < +\infty.$$

On the event $A_\infty = \left\{\omega : (h(\theta_n(\omega)))_{n\geq 0} \text{ is bounded}\right\}$, $\mathbb{P}(d\omega)$-$a.s.$, the set $\Theta^\infty(\omega)$ of the limiting values of $(\theta_n(\omega)_{n\geq 0})$ as $n \to +\infty$ is a compact connected flow invariant minimal attractor for ODE_h.

Assume that ODE_h has a C-valued flow. Assume furthermore that

$$r_n \xrightarrow[n\to\infty]{a.s.} 0 \quad \text{and} \quad \sup_{n\geq n_0} \mathbb{E}\left[\|\Delta M_{n+1}\|^2 \mid \mathcal{F}_n\right] < +\infty \quad a.s.,$$

and that $(\gamma_n)_{n\geq 1}$ is a positive sequence satisfying $(\gamma_n \in (0, \bar{\gamma}],\ n \geq 1)$ and

$$\sum_{n\geq 1} \gamma_n = +\infty \quad \text{and} \quad \sum_{n\geq 1} \gamma_n^2 < +\infty.$$

On the event $A_\infty = \{\omega : (h(\theta_n(\omega)))_{n\geq 0} \text{ is bounded}\}$, $\mathbb{P}(d\omega)$-$a.s.$, the set $\Theta^\infty(\omega)$ of the limiting values of $(\theta_n(\omega)_{n\geq 0})$ as $n \to +\infty$ is a compact connected flow invariant minimal attractor for ODE_h.

If $\{h = 0\} = \{\theta^*\}$ and $\Phi(\theta_0, t) \underset{t\to+\infty}{\longrightarrow} \theta^*$ locally uniformly in θ_0, then $\Theta^\infty(\omega) = \{\theta^*\}$ *i.e.* $\theta_n \xrightarrow[n\to+\infty]{a.s.} \theta^*$.

Outline

Rate(s) of convergence

We will say that h is η-differentiable ($\eta > 0$) at θ^* if

$$h(\theta) = h(\theta^*) + J_h(\theta^*)(\theta - \theta^*) + o(\|\theta - \theta^*\|^{1+\eta}) \quad \text{as} \quad \theta \to \theta^*.$$

Rate(s) of convergence

We will say that h is η-differentiable ($\eta > 0$) at θ^* if

$$h(\theta) = h(\theta^*) + J_h(\theta^*)(\theta - \theta^*) + o(\|\theta - \theta^*\|^{1+\eta}) \quad \text{as} \quad \theta \to \theta^*.$$

Let θ^* be an equilibrium point of $\{h = 0\}$ and $\{\theta_n \xrightarrow{a.s.} \theta^*\}$ the convergence event associated to θ^*. Set the gain parameter sequence $(\gamma_n)_{n\geq 1}$ as follows : $\forall n \geq 1$, $\gamma_n = \dfrac{1}{n}$.

Rate(s) of convergence

We will say that h is η-differentiable ($\eta > 0$) at θ^* if

$$h(\theta) = h(\theta^*) + J_h(\theta^*)(\theta - \theta^*) + o(\|\theta - \theta^*\|^{1+\eta}) \quad \text{as} \quad \theta \to \theta^*.$$

Let θ^* be an equilibrium point of $\{h = 0\}$ and $\{\theta_n \xrightarrow{a.s.} \theta^*\}$ the convergence event associated to θ^*. Set the gain parameter sequence $(\gamma_n)_{n\geq 1}$ as follows : $\forall n \geq 1$, $\gamma_n = \dfrac{1}{n}$.

Assume that the function h is differentiable at θ^* and all the eigenvalues of $J_h(\theta^*)$ have positive real parts.

Rate(s) of convergence

We will say that h is η-differentiable ($\eta > 0$) at θ^* if

$$h(\theta) = h(\theta^*) + J_h(\theta^*)(\theta - \theta^*) + o(\|\theta - \theta^*\|^{1+\eta}) \quad \text{as} \quad \theta \to \theta^*.$$

Let θ^* be an equilibrium point of $\{h = 0\}$ and $\{\theta_n \xrightarrow{a.s.} \theta^*\}$ the convergence event associated to θ^*. Set the gain parameter sequence $(\gamma_n)_{n\geq 1}$ as follows : $\forall n \geq 1$, $\gamma_n = \dfrac{1}{n}$.

Assume that the function h is differentiable at θ^* and all the eigenvalues of $J_h(\theta^*)$ have positive real parts. Assume that, for a real number $\delta > 0$,

$$\sup_{n\geq 0} \mathbb{E}\left[\|\Delta M_{n+1}\|^{2+\delta} \,|\, \mathcal{F}_n\right] < +\infty \ a.s., \quad \mathbb{E}\left[\Delta M_{n+1}\Delta M_{n+1}^t \,|\, \mathcal{F}_n\right] \xrightarrow[n\to+\infty]{a.s.} \Gamma^*,$$

on $\{\theta_n \xrightarrow{a.s.} \theta^*\}$, where $\Gamma^* \in \mathcal{S}^+(d, \mathbb{R})$ and for an $\varepsilon > 0$

Rate(s) of convergence

We will say that h is η-differentiable ($\eta > 0$) at θ^* if

$$h(\theta) = h(\theta^*) + J_h(\theta^*)(\theta - \theta^*) + o(\|\theta - \theta^*\|^{1+\eta}) \quad \text{as} \quad \theta \to \theta^*.$$

Let θ^* be an equilibrium point of $\{h = 0\}$ and $\{\theta_n \xrightarrow{a.s.} \theta^*\}$ the convergence event associated to θ^*. Set the gain parameter sequence $(\gamma_n)_{n \geq 1}$ as follows : $\forall n \geq 1$, $\gamma_n = \dfrac{1}{n}$.

Assume that the function h is differentiable at θ^* and all the eigenvalues of $J_h(\theta^*)$ have positive real parts. Assume that, for a real number $\delta > 0$,

$$\sup_{n \geq 0} \mathbb{E}\left[\|\Delta M_{n+1}\|^{2+\delta} \,|\, \mathcal{F}_n\right] < +\infty \; a.s., \quad \mathbb{E}\left[\Delta M_{n+1} \Delta M_{n+1}^t \,|\, \mathcal{F}_n\right] \xrightarrow[n \to +\infty]{a.s.} \Gamma^*,$$

on $\{\theta_n \xrightarrow{a.s.} \theta^*\}$, where $\Gamma^* \in \mathcal{S}^+(d, \mathbb{R})$ and for an $\varepsilon > 0$ and a positive sequence $(v_n)_{n \geq 1}$ (specified below),

$$n\, v_n \mathbb{E}\left[\|r_{n+1}\|^2 \mathbb{1}_{\{\|\theta_n - \theta^*\| \leq \varepsilon\}}\right] \xrightarrow[n \to +\infty]{} 0.$$

Let $\lambda_{\min}$ denote the eigenvalue of $J_h(\theta^*)$ with the lowest real part.

Let $\lambda_{\min}$ denote the eigenvalue of $J_h(\theta^*)$ with the lowest real part.

1. If $\Re e(\lambda_{\min}) > \frac{1}{2}$ and $v_n = 1$, $n \geq 1$, then, the weak convergence rate is ruled on $\{\theta_n \xrightarrow{a.s.} \theta^*\}$ by the following CLT

$$\sqrt{n}\left(\theta_n - \theta^*\right) \xrightarrow[n\to+\infty]{\mathcal{L}_{stably}} \mathcal{N}\left(0, \Sigma^*\right) \text{ with } \Sigma^* := \int_0^{+\infty} e^{-u\left(J_h(\theta^*)^t - \frac{I_d}{2}\right)} \Gamma^* e^{-u\left(J_h(\theta^*) - \frac{I_d}{2}\right)} du.$$

Let $\lambda_{\min}$ denote the eigenvalue of $J_h(\theta^*)$ with the lowest real part.

1. If $\Re e(\lambda_{\min}) > \frac{1}{2}$ and $v_n = 1$, $n \geq 1$, then, the weak convergence rate is ruled on $\{\theta_n \xrightarrow{a.s.} \theta^*\}$ by the following CLT

$$\sqrt{n}\left(\theta_n - \theta^*\right) \xrightarrow[n\to+\infty]{\mathcal{L}_{stably}} \mathcal{N}\left(0, \Sigma^*\right) \text{ with } \Sigma^* := \int_0^{+\infty} e^{-u\left(J_h(\theta^*)^t - \frac{I_d}{2}\right)} \Gamma^* e^{-u\left(J_h(\theta^*) - \frac{I_d}{2}\right)} du.$$

2. If $\Re e(\lambda_{\min}) = \frac{1}{2}$, $v_n = \log n$, $n \geq 1$, and h is η-differentiable at θ^*, then on $\{\theta_n \xrightarrow{a.s.} \theta^*\}$,

$$\sqrt{\frac{n}{\log n}}\left(\theta_n - \theta^*\right) \xrightarrow[n\to+\infty]{\mathcal{L}_{stably}} \mathcal{N}(0, \Sigma^*), \quad \Sigma^* := \lim_n \frac{1}{n} \int_0^n e^{-u\left(J_h(\theta^*)^t - \frac{I_d}{2}\right)} \Gamma^* e^{-u\left(J_h(\theta^*) - \frac{I_d}{2}\right)} du.$$

Let $\lambda_{\min}$ denote the eigenvalue of $J_h(\theta^*)$ with the lowest real part.

1. If $\Re e(\lambda_{\min}) > \frac{1}{2}$ and $v_n = 1$, $n \geq 1$, then, the weak convergence rate is ruled on $\{\theta_n \xrightarrow{a.s.} \theta^*\}$ by the following CLT

$$\sqrt{n}\,(\theta_n - \theta^*) \xrightarrow[n\to+\infty]{\mathcal{L}_{stably}} \mathcal{N}\,(0, \Sigma^*) \text{ with } \Sigma^* := \int_0^{+\infty} e^{-u\left(J_h(\theta^*)^t - \frac{I_d}{2}\right)} \Gamma^* e^{-u\left(J_h(\theta^*) - \frac{I_d}{2}\right)} du.$$

2. If $\Re e(\lambda_{\min}) = \frac{1}{2}$, $v_n = \log n$, $n \geq 1$, and h is η-differentiable at θ^*, then on $\{\theta_n \xrightarrow{a.s.} \theta^*\}$,

$$\sqrt{\frac{n}{\log n}}\,(\theta_n - \theta^*) \xrightarrow[n\to+\infty]{\mathcal{L}_{stably}} \mathcal{N}(0, \Sigma^*), \;\; \Sigma^* := \lim_n \frac{1}{n} \int_0^{n} e^{-u\left(J_h(\theta^*)^t - \frac{I_d}{2}\right)} \Gamma^* e^{-u\left(J_h(\theta^*) - \frac{I_d}{2}\right)} du.$$

3. If $\Lambda := \Re e(\lambda_{\min}) \in \left(0, \frac{1}{2}\right)$, $v_n = n^{2\Lambda-1+\epsilon}$, $n \geq 1$, for some $\epsilon > 0$, and h is η-differentiable at θ^*, then $n^{\Lambda}\,(\theta_n - \theta^*)$ is *a.s.* bounded on $\{\theta_n \xrightarrow{a.s.} \theta^*\}$ as $n \to +\infty$.

Let $\lambda_{\min}$ denote the eigenvalue of $J_h(\theta^*)$ with the lowest real part.

1. If $\Re e(\lambda_{\min}) > \frac{1}{2}$ and $v_n = 1$, $n \geq 1$, then, the weak convergence rate is ruled on $\{\theta_n \xrightarrow{a.s.} \theta^*\}$ by the following CLT

$$\sqrt{n}\left(\theta_n - \theta^*\right) \xrightarrow[n\to+\infty]{\mathcal{L}_{stably}} \mathcal{N}\left(0, \Sigma^*\right) \text{ with } \Sigma^* := \int_0^{+\infty} e^{-u\left(J_h(\theta^*)^t - \frac{I_d}{2}\right)} \Gamma^* e^{-u\left(J_h(\theta^*) - \frac{I_d}{2}\right)} du.$$

2. If $\Re e(\lambda_{\min}) = \frac{1}{2}$, $v_n = \log n$, $n \geq 1$, and h is η-differentiable at θ^*, then on $\{\theta_n \xrightarrow{a.s.} \theta^*\}$,

$$\sqrt{\frac{n}{\log n}}\left(\theta_n - \theta^*\right) \xrightarrow[n\to+\infty]{\mathcal{L}_{stably}} \mathcal{N}(0, \Sigma^*), \ \Sigma^* := \lim_n \frac{1}{n} \int_0^n e^{-u\left(J_h(\theta^*)^t - \frac{I_d}{2}\right)} \Gamma^* e^{-u\left(J_h(\theta^*) - \frac{I_d}{2}\right)} du.$$

3. If $\Lambda := \Re e(\lambda_{\min}) \in \left(0, \frac{1}{2}\right)$, $v_n = n^{2\Lambda - 1 + \epsilon}$, $n \geq 1$, for some $\epsilon > 0$, and h is η-differentiable at θ^*, then $n^{\Lambda}\left(\theta_n - \theta^*\right)$ is *a.s.* bounded on $\{\theta_n \xrightarrow{a.s.} \theta^*\}$ as $n \to +\infty$. If, moreover, $\Lambda = \lambda_{\min}$ ($\lambda_{\min}$ is real), then $n^{\Lambda}\left(\theta_n - \theta^*\right)$ *a.s.* converges as $n \to +\infty$ toward a finite random variable.

Outline

Averaging principle (Ruppert-Polyak)

The idea is to “smooth” the behavior of the converging stochastic algorithm by considering its arithmetic mean with scaling up the step sequence.

Averaging principle (Ruppert-Polyak)

The idea is to "smooth" the behavior of the converging stochastic algorithm by considering its arithmetic mean with scaling up the step sequence.

To be precise, let $(\gamma_n)_{n\geq 1}$ be e step sequence satisfying

$$\gamma_n \sim \left(\frac{c}{b+n}\right)^{\vartheta}, \qquad \vartheta \in (1/2, 1).$$

Averaging principle (Ruppert-Polyak)

The idea is to “smooth” the behavior of the converging stochastic algorithm by considering its arithmetic mean with scaling up the step sequence.

To be precise, let $(\gamma_n)_{n\geq 1}$ be e step sequence satisfying

$$\gamma_n \sim \left(\frac{c}{b+n}\right)^{\vartheta}, \qquad \vartheta \in (1/2, 1).$$

Then we implement the standard procedure

$$\forall n \geq 1, \quad \theta_{n+1} = \theta_n - \gamma_{n+1}(h(\theta_n) + \Delta M_{n+1})$$

and we set

$$\forall n \geq 1, \quad \bar{\theta}_n := \frac{\theta_0 + \cdots + \theta_{n-1}}{n}.$$

Assume that h is a Borel function, continuous at its unique zero θ^*, satisfying

$$\forall\, \theta \in \mathbb{R}^d, \quad h(\theta) = Dh(\theta^*)(\theta - \theta^*) + O(|\theta - \theta^*|^2)$$

where all the eigenvalues of $Dh(\theta^*)$ have positive real parts.

Assume that h is a Borel function, continuous at its unique zero θ^*, satisfying

$$\forall\,\theta\in\mathbb{R}^d,\quad h(\theta)=Dh(\theta^*)(\theta-\theta^*)+O(|\theta-\theta^*|^2)$$

where all the eigenvalues of $Dh(\theta^*)$ have positive real parts. Assume furthermore that for some constant $C>0$,

$$\mathbb{E}[\Delta M_{n+1}\,|\,\mathcal{F}_n]\mathbb{1}_{\{|\theta_n-\theta^*|\leq C\}}=0\quad a.s.$$

Assume that h is a Borel function, continuous at its unique zero θ^*, satisfying

$$\forall\, \theta \in \mathbb{R}^d, \quad h(\theta) = Dh(\theta^*)(\theta - \theta^*) + O(|\theta - \theta^*|^2)$$

where all the eigenvalues of $Dh(\theta^*)$ have positive real parts. Assume furthermore that for some constant $C > 0$,

$$\mathbb{E}[\Delta M_{n+1} \,|\, \mathcal{F}_n] \mathbb{1}_{\{|\theta_n - \theta^*| \leq C\}} = 0 \quad a.s.$$

and that there exists $\delta > 0$ such that

$$\mathbb{E}[\Delta M_{n+1} (\Delta M_{n+1})^t \,|\, \mathcal{F}_n] \xrightarrow{a.s.} \mathbb{1}_{\{|\theta_n - \theta^*| \leq C\}} \Gamma > 0 \,\in\, \mathcal{S}(d, \mathbb{R})$$

$$\sup_n \mathbb{E}[|\Delta M_{n+1}|^{2+\delta} \,|\, \mathcal{F}_n] \mathbb{1}_{\{|\theta_n - \theta^*| \leq C\}} < +\infty.$$

Assume that h is a Borel function, continuous at its unique zero θ^*, satisfying

$$\forall\, \theta\, \in \mathbb{R}^d, \quad h(\theta) = Dh(\theta^*)(\theta - \theta^*) + O(|\theta - \theta^*|^2)$$

where all the eigenvalues of $Dh(\theta^*)$ have positive real parts. Assume furthermore that for some constant $C > 0$,

$$\mathbb{E}[\Delta M_{n+1} \,|\, \mathcal{F}_n]\mathbb{1}_{\{|\theta_n - \theta^*| \le C\}} = 0 \quad a.s.$$

and that there exists $\delta > 0$ such that

$$\mathbb{E}[\Delta M_{n+1}(\Delta M_{n+1})^t \,|\, \mathcal{F}_n] \xrightarrow{a.s.} \mathbb{1}_{\{|\theta_n - \theta^*| \le C\}}\Gamma > 0 \,\in\, \mathcal{S}(d, \mathbb{R})$$

$$\sup_n \mathbb{E}[|\Delta M_{n+1}|^{2+\delta} \,|\, \mathcal{F}_n]\mathbb{1}_{\{|\theta_n - \theta^*| \le C\}} < +\infty.$$

Then, if $\gamma_n = \dfrac{c}{n^\alpha}$, $n \ge 1$, $1/2 < \alpha < 1$, the sequence of arithmetic mean $(\bar{\theta}_n)_{n\ge1}$ satisfies on $\{\theta_n \to \theta^*\}$, the *CLT* with the optimal variance

$$\sqrt{n}(\bar{\theta}_n - \theta^*) \xrightarrow{\mathcal{L}} \mathcal{N}(0; Dh(\theta^*)^{-1}\Gamma Dh(\theta^*)) \qquad \text{on } \{\theta_n \to \theta^*\}.$$

Outline

Outline

Modeling

The principle of a *Dark pool* is the following:

Modeling

The principle of a *Dark pool* is the following:

- It proposes a bid price with no guarantee of executed quantity at the occasion of an OTC transaction.

Modeling

The principle of a *Dark pool* is the following:

- It proposes a bid price with no guarantee of executed quantity at the occasion of an OTC transaction.
- Usually this price is lower than the bid price offered on the regular market (no market impact because no pre-trade transparency).

Modeling

The principle of a *Dark pool* is the following:

- It proposes a bid price with no guarantee of executed quantity at the occasion of an OTC transaction.
- Usually this price is lower than the bid price offered on the regular market (no market impact because no pre-trade transparency).

So one can model the impact of the existence of N dark pools ($N \geq 2$) on a given transaction as follows:

Modeling

The principle of a *Dark pool* is the following:

- It proposes a bid price with no guarantee of executed quantity at the occasion of an OTC transaction.
- Usually this price is lower than the bid price offered on the regular market (no market impact because no pre-trade transparency).

So one can model the impact of the existence of N dark pools ($N \geq 2$) on a given transaction as follows:

- Let $V > 0$ be the random volume to be executed,

Modeling

The principle of a *Dark pool* is the following:

- It proposes a bid price with no guarantee of executed quantity at the occasion of an OTC transaction.
- Usually this price is lower than the bid price offered on the regular market (no market impact because no pre-trade transparency).

So one can model the impact of the existence of N dark pools ($N \geq 2$) on a given transaction as follows:

- Let $V > 0$ be the random volume to be executed,
- Let $\theta_i \in (0, 1)$ be the *discount factor* proposed by the dark pool i.

Modeling

The principle of a *Dark pool* is the following:

- It proposes a bid price with no guarantee of executed quantity at the occasion of an OTC transaction.
- Usually this price is lower than the bid price offered on the regular market (no market impact because no pre-trade transparency).

So one can model the impact of the existence of N dark pools ($N \geq 2$) on a given transaction as follows:

- Let $V > 0$ be the random volume to be executed,
- Let $\theta_i \in (0, 1)$ be the *discount factor* proposed by the dark pool i.
- Let r_i denote the percentage of V sent to the dark pool i for execution.

Modeling

The principle of a *Dark pool* is the following:

- It proposes a bid price with no guarantee of executed quantity at the occasion of an OTC transaction.
- Usually this price is lower than the bid price offered on the regular market (no market impact because no pre-trade transparency).

So one can model the impact of the existence of N dark pools ($N \geq 2$) on a given transaction as follows:

- Let $V > 0$ be the random volume to be executed,
- Let $\theta_i \in (0, 1)$ be the *discount factor* proposed by the dark pool i.
- Let r_i denote the percentage of V sent to the dark pool i for execution.
- Let $D_i \geq 0$ be the quantity of securities that can be delivered (or mase available) by the dark pool i at price $\theta_i S$.

Cost of the executed order

The reminder of the order is to be executed on the regular market, at price S.

Cost of the executed order

The reminder of the order is to be executed on the regular market, at price S.

Then the **cost** C of the whole executed order is given by

Cost of the executed order

The reminder of the order is to be executed on the regular market, at price S.

Then the **cost** C of the whole executed order is given by

$$C = S\sum_{i=1}^{N}\theta_i \min\left(r_i V, D_i\right) + S\left(V - \sum_{i=1}^{N}\min\left(r_i V, D_i\right)\right)$$

Cost of the executed order

The reminder of the order is to be executed on the regular market, at price S.

Then the **cost** C of the whole executed order is given by

$$
\begin{aligned}
C &= S\sum_{i=1}^{N}\theta_i \min\left(r_i V, D_i\right) + S\left(V - \sum_{i=1}^{N}\min\left(r_i V, D_i\right)\right)\\
&= S\left(V - \sum_{i=1}^{N}\rho_i \min\left(r_i V, D_i\right)\right)
\end{aligned}
$$

where

$$\rho_i = 1 - \theta_i \in (0, 1), i = 1, \ldots, N.$$

Mean Execution Cost

Minimizing the mean execution cost, *given the price S*, amounts to solving the following **maximization problem**

Mean Execution Cost

Minimizing the mean execution cost, *given the price* S, amounts to solving the following **maximization problem**

$$\max\left\{\sum_{i=1}^{N}\rho_i\mathbb{E}\left(S\min\left(r_iV,D_i\right)\right),\, r\in\mathcal{P}_N\right\}$$

where $\mathcal{P}_N := \left\{r=(r_i)_{1\le i\le N}\in\mathbb{R}_+^N \mid \sum_{i=1}^N r_i = 1\right\}$.

Mean Execution Cost

Minimizing the mean execution cost, *given the price S*, amounts to solving the following **maximization problem**

$$\max\left\{\sum_{i=1}^{N}\rho_i\mathbb{E}\left(S\min\left(r_iV,D_i\right)\right), r\in\mathcal{P}_N\right\}$$

where $\mathcal{P}_N := \left\{r=(r_i)_{1\le i\le N}\in\mathbb{R}_+^N \mid \sum_{i=1}^N r_i = 1\right\}$.

It is then convenient to *include the price S into both random variables V and D_i* by considering

$$\widetilde{V} := V\,S \quad \text{and} \quad \widetilde{D}_i := D_iS$$

instead of V and D_i.

Optimal allocation of orders among N dark pools

We set for every $r = (r_1, \ldots, r_N) \in \mathcal{P}_N$,

$$\Phi(r_1, \ldots, r_N) := \sum_{i=1}^{N} \varphi_i(r_i).$$

Optimal allocation of orders among N dark pools

We set for every $r = (r_1, \dots, r_N) \in \mathcal{P}_N$,

$$\Phi(r_1, \dots, r_N) := \sum_{i=1}^{N} \varphi_i(r_i).$$

where for every $i \in I_N = \{1, \dots, N\}$,

$$\varphi_i(u) := \rho_i \mathbb{E}\left[\min\left(uV, D_i\right)\right], \quad u \in [0,1].$$

Optimal allocation of orders among N dark pools

We set for every $r = (r_1, \ldots, r_N) \in \mathcal{P}_N$,

$$\Phi(r_1, \ldots, r_N) := \sum_{i=1}^{N} \varphi_i(r_i).$$

where for every $i \in I_N = \{1, \ldots, N\}$,

$$\varphi_i(u) := \rho_i \mathbb{E}\left[\min\left(uV, D_i\right)\right], \quad u \in [0, 1].$$

We can formally extend Φ on the whole affine hyperplan spanned by $\mathcal{P}_N$ *i.e.*

$$\mathcal{H}_N := \left\{ r = (r_1, \ldots, r_N) \in \mathbb{R}^N \mid \sum_{i=1}^{N} r_i = 1 \right\}.$$

Outline

The Lagrangian Approach

We aim at solving the following maximization problem

$$\max_{r \in \mathcal{P}_N} \Phi(r).$$

The Lagrangian Approach

We aim at solving the following maximization problem

$$\max_{r \in \mathcal{P}_N} \Phi(r).$$

The Lagrangian associated to the sole affine constraint is

$$L(r, \lambda) = \Phi(r) - \lambda \left(\sum_{i=1}^{N} r_i - 1 \right)$$

The Lagrangian Approach

We aim at solving the following maximization problem

$$\max_{r \in \mathcal{P}_N} \Phi(r).$$

The Lagrangian associated to the sole affine constraint is

$$L(r, \lambda) = \Phi(r) - \lambda \left(\sum_{i=1}^{N} r_i - 1 \right)$$

Thus

$$\forall i \in I_N, \quad \frac{\partial L}{\partial r_i} = \varphi_i'(r_i) - \lambda.$$

The Lagrangian Approach

We aim at solving the following maximization problem

$$\max_{r \in \mathcal{P}_N} \Phi(r).$$

The Lagrangian associated to the sole affine constraint is

$$L(r, \lambda) = \Phi(r) - \lambda \left(\sum_{i=1}^{N} r_i - 1 \right)$$

Thus

$$\forall i \in I_N, \quad \frac{\partial L}{\partial r_i} = \varphi_i'(r_i) - \lambda.$$

This suggests that any $r^* \in \arg\max_{\mathcal{P}_N} \Phi$ iff $\varphi_i'(r_i^*)$ is constant when i runs over I_N

The Lagrangian Approach

We aim at solving the following maximization problem

$$\max_{r\in\mathcal{P}_N} \Phi(r).$$

The Lagrangian associated to the sole affine constraint is

$$L(r,\lambda) = \Phi(r) - \lambda\left(\sum_{i=1}^{N} r_i - 1\right)$$

Thus

$$\forall i \in I_N, \quad \frac{\partial L}{\partial r_i} = \varphi_i'(r_i) - \lambda.$$

This suggests that any $r^* \in \arg\max_{\mathcal{P}_N} \Phi$ iff $\varphi_i'(r_i^*)$ is constant when i runs over I_N or equivalently if

$$\forall i \in I_N, \quad \varphi_i'(r_i^*) = \frac{1}{N}\sum_{j=1}^{N} \varphi_j'(r_j^*).$$

Design of the stochastic algorithm

Then using that

$$\forall i \in I_N, \quad \varphi_i'(r) = \rho_i \mathbb{E}\left[\mathbb{1}_{\{r_i V < D_i\}} V\right],$$

Design of the stochastic algorithm

Then using that

$$\forall i \in I_N, \quad \varphi_i'(r) = \rho_i \mathbb{E}\left[\mathbb{1}_{\{r_i V < D_i\}} V\right],$$

this yields that, if $\arg\max_{\mathcal{H}_N} \Phi = \arg\max_{\mathcal{P}_N} \Phi \subset int(\mathcal{P}_N)$, then

$$r^* \in \arg\max_{\mathcal{P}_N} \Phi$$
$$\Updownarrow$$
$$\forall i \in I_N, \quad \mathbb{E}\left[V\left(\rho_i \mathbb{1}_{\{r_i^* V < D_i\}} - \frac{1}{N}\sum_{j=1}^{N} \rho_j \mathbb{1}_{\{r_j^* V < D_j\}}\right)\right] = 0.$$

Design of the stochastic algorithm

Then using that

$$\forall i \in I_N, \quad \varphi_i'(r) = \rho_i \mathbb{E}\left[\mathbb{1}_{\{r_i V < D_i\}} V\right],$$

this yields that, if $\arg\max_{\mathcal{H}_N} \Phi = \arg\max_{\mathcal{P}_N} \Phi \subset int(\mathcal{P}_N)$, then

$$r^* \in \arg\max_{\mathcal{P}_N} \Phi$$
$$\Updownarrow$$
$$\forall i \in I_N, \quad \mathbb{E}\left[V\left(\rho_i \mathbb{1}_{\{r_i^* V < D_i\}} - \frac{1}{N}\sum_{j=1}^{N} \rho_j \mathbb{1}_{\{r_j^* V < D_j\}}\right)\right] = 0.$$

Consequently, this leads to the following **recursive zero search procedure**

$$r_i^{n+1} = r_i^n + \gamma_{n+1} H_i(r^n, Y^{n+1}),\ r^0 \in \mathcal{P}_N,\ i \in I_N,$$

Design of the stochastic algorithm

where for $i \in I_N$, every $r \in \mathcal{P}_N$, every $V > 0$ and every $D_1, \ldots, D_N \geq 0$,

$$H_i(r, Y) = V \left(\rho_i \mathbb{1}_{\{r_i V < D_i\}} - \frac{1}{N} \sum_{j=1}^{N} \rho_j \mathbb{1}_{\{r_j V < D_j\}} \right)$$

where $(Y^n)_{n \geq 1}$ is a sequence of random vectors with non negative components such that, for every $n \geq 1$ and $i \in I_N$, $(V^n, D_i^n) \stackrel{d}{=} (V, D_i)$.

Design of the stochastic algorithm

where for $i \in I_N$, every $r \in \mathcal{P}_N$, every $V > 0$ and every $D_1, \ldots, D_N \geq 0$,

$$H_i(r, Y) = V \left(\rho_i \mathbb{1}_{\{r_i V < D_i\}} - \frac{1}{N} \sum_{j=1}^{N} \rho_j \mathbb{1}_{\{r_j V < D_j\}} \right)$$

where $(Y^n)_{n \geq 1}$ is a sequence of random vectors with non negative components such that, for every $n \geq 1$ and $i \in I_N$, $(V^n, D_i^n) \stackrel{d}{=} (V, D_i)$.

Comment: The underlying idea of the algorithm is to reward the dark pools which outperform the mean of the N dark pools by increasing the allocated volume sent at the next step (and conversely).

Constraint Problem: Practical recommendations

In this algorithm, we took into account the constraint

$$\sum_{i=1}^{N} r_i = 1,$$

but not

$$r_i > 0, \; \forall\, 1 \leq i \leq N.$$

Constraint Problem: Practical recommendations

In this algorithm, we took into account the constraint

$$\sum_{i=1}^{N} r_i = 1,$$

but not

$$r_i > 0, \ \forall\, 1 \le i \le N.$$

So the algorithm doesn't leave the simplex $\mathcal{P}_N$ stable. To overcome this problem, we have two possibilities

Constraint Problem: Practical recommendations

In this algorithm, we took into account the constraint

$$\sum_{i=1}^{N} r_i = 1,$$

but not

$$r_i > 0,\ \forall\, 1 \leq i \leq N.$$

So the algorithm doesn't leave the simplex $\mathcal{P}_N$ stable. To overcome this problem, we have two possibilities

- Use a Lyapunov function and an mean-reverting assumption out of $\mathcal{P}_N$: This solution is simpler from a mathematical point of view.

Constraint Problem: Practical recommendations

In this algorithm, we took into account the constraint

$$\sum_{i=1}^{N} r_i = 1,$$

but not

$$r_i > 0, \ \forall\, 1 \leq i \leq N.$$

So the algorithm doesn't leave the simplex $\mathcal{P}_N$ stable. To overcome this problem, we have two possibilities

- Use a Lyapunov function and an mean-reverting assumption out of $\mathcal{P}_N$: This solution is simpler from a mathematical point of view.
- Force the coefficients r_i to stay in $\mathcal{P}_N$ by a truncation-projection procedure: This solution is more efficient for users.

Outline

Numerical Tests

Two natural situations of interest can be considered *a priori*:

Numerical Tests

Two natural situations of interest can be considered *a priori*:

- *abundance*: $\mathbb{E}\,V \leq \sum_{i=1}^{N} \mathbb{E} D_i$,

Numerical Tests

Two natural situations of interest can be considered *a priori*:

- *abundance*: $\mathbb{E}\, V \leq \sum_{i=1}^{N} \mathbb{E} D_i$,
- *shortage*: $\mathbb{E}\, V > \sum_{i=1}^{N} \mathbb{E} D_i$.

Numerical Tests

Two natural situations of interest can be considered *a priori*:

- *abundance*: $\mathbb{E}\, V \leq \sum_{i=1}^{N} \mathbb{E} D_i$,
- *shortage*: $\mathbb{E}\, V > \sum_{i=1}^{N} \mathbb{E} D_i$.

We define a reference strategy named "**oracle**" devised by an insider who knows all the values V^n and D_i^n before making his/her optimal execution requests to the dark pools. The "oracle" strategy yields a cost reduction of the execution denoted by CR^{oracle}.

Numerical Tests

Two natural situations of interest can be considered *a priori*:

- *abundance*: $\mathbb{E}\,V \leq \sum_{i=1}^{N} \mathbb{E}D_i$,
- *shortage*: $\mathbb{E}\,V > \sum_{i=1}^{N} \mathbb{E}D_i$.

We define a reference strategy named "**oracle**" devised by an insider who knows all the values V^n and D_i^n before making his/her optimal execution requests to the dark pools. The "oracle" strategy yields a cost reduction of the execution denoted by CR^{oracle}.

We introduce indexes to measure the **performances** of our recursive allocation procedure as follows: the ratios between the relative cost reductions of our allocation algorithm and that of the oracle, *i.e.*

$$\frac{CR^{opti}}{CR^{oracle}}.$$

The (IID) setting

V and D_i, $i \in \mathcal{I}_N$, are log-normal variables and $N = 3$. The shortage setting is specified as follows:

$$\mathbb{E}\, V = \frac{3}{2} \sum_{i=1}^{N} \mathbb{E} D_i$$

The (IID) setting

V and D_i, $i \in \mathcal{I}_N$, are log-normal variables and $N = 3$. The shortage setting is specified as follows:

$$\mathbb{E}\,V = \frac{3}{2}\sum_{i=1}^{N}\mathbb{E}D_i$$

with

$$\mathbb{E}D_i = i,\ 1 \leq i \leq N, \quad \mathrm{Var}(V) = 1, \quad \mathrm{Var}(D_i) = 1,\ 1 \leq i \leq N$$

and

The (*IID*) setting

V and D_i, $i \in \mathcal{I}_N$, are log-normal variables and $N = 3$. The shortage setting is specified as follows:

$$\mathbb{E}\, V = \frac{3}{2} \sum_{i=1}^{N} \mathbb{E} D_i$$

with

$$\mathbb{E} D_i = i,\ 1 \le i \le N, \quad \mathrm{Var}(V) = 1, \quad \mathrm{Var}(D_i) = 1,\ 1 \le i \le N$$

and

$$\rho = \begin{pmatrix} 0.01 \\ 0.03 \\ 0.05 \end{pmatrix}.$$

The initial value for both algorithms is set at $r_i^0 = \frac{1}{N}$, $1 \le i \le N$.

Moving average on a window of 100 data for CR and performances

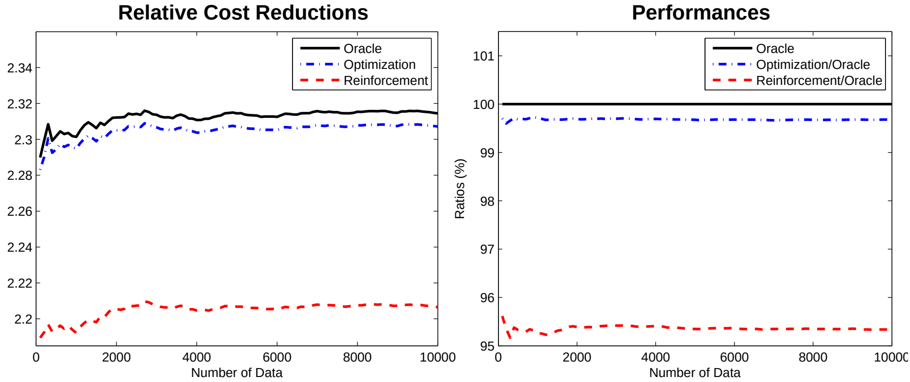

Figure: *Shortage setting*: Case $N = 3$, $m_V = \frac{3}{2}\sum_{i=1}^{N} m_{D_i}$, $m_{D_i} = i$, $\sigma_V = 1$, $\sigma_{D_i} = 1$, $1 \le i \le N$.

The pseudo-real data setting

V is the traded volume of a very liquid security – namely the asset BNP – during an 11 day period.

The pseudo-real data setting

V is the traded volume of a very liquid security – namely the asset BNP – during an 11 day period.

Then we selected the N most correlated assets (in terms of traded volumes) with the original asset, denoted S_i, $i = 1, \ldots, N$.

The pseudo-real data setting

V is the traded volume of a very liquid security – namely the asset BNP – during an 11 day period.

Then we selected the N most correlated assets (in terms of traded volumes) with the original asset, denoted S_i, $i = 1, \ldots, N$.

The available volumes of each dark pool i have been modeled as follows

$$\forall 1 \leq i \leq N, \quad D_i := \beta_i \left((1 - \alpha_i) V + \alpha_i S_i \frac{\mathbb{E}\, V}{\mathbb{E} S_i} \right)$$

The pseudo-real data setting

V is the traded volume of a very liquid security – namely the asset BNP – during an 11 day period.

Then we selected the N most correlated assets (in terms of traded volumes) with the original asset, denoted S_i, $i = 1, \ldots, N$.

The available volumes of each dark pool i have been modeled as follows

$$\forall 1 \leq i \leq N, \quad D_i := \beta_i \left((1 - \alpha_i) V + \alpha_i S_i \frac{\mathbb{E}\, V}{\mathbb{E} S_i} \right)$$

where

- $\alpha_i \in (0, 1)$, $i = 1, \ldots, N$ are the recombining coefficients,
- β_i, $i = 1, \ldots, N$ are some scaling factors,
- $\mathbb{E}\, V$ and $\mathbb{E} S_i$ stand for the empirical mean of the data sets of V and S_i.

Parameters

The shortage situation corresponds to $\sum_{i=1}^{N} \beta_i < 1$ since it implies

$$\mathbb{E}\left[\sum_{i=1}^{N} D_i\right] < \mathbb{E}\, V.$$

Parameters

The shortage situation corresponds to $\sum_{i=1}^{N} \beta_i < 1$ since it implies

$$\mathbb{E}\left[\sum_{i=1}^{N} D_i\right] < \mathbb{E}\, V.$$

We will use the following parameters

$$N = 4, \quad \beta = (0.1, 0.2, 0.3, 0.2)' \quad \text{and} \quad \alpha = (0.4, 0.6, 0.8, 0.2)'$$

and the dark pool trading (rebate) parameters are set to

$$\rho = (0.01, 0.02, 0.04, 0.06)'.$$

Long-term optimization

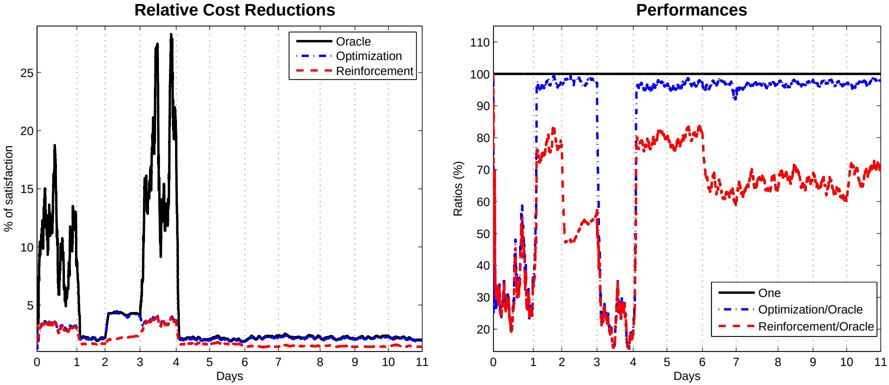

Figure: *Long term optimization*: Case $N = 4$, $\sum_{i=1}^{N} \beta_i < 1$, $0.2 \leq \alpha_i \leq 0.8$ and $r_i^0 = 1/N$, $1 \leq i \leq N$.

Daily resetting of the procedure

We reset the step γ_n at the beginning of each day and the satisfaction parameters and we keep the allocation coefficients of the preceding day.

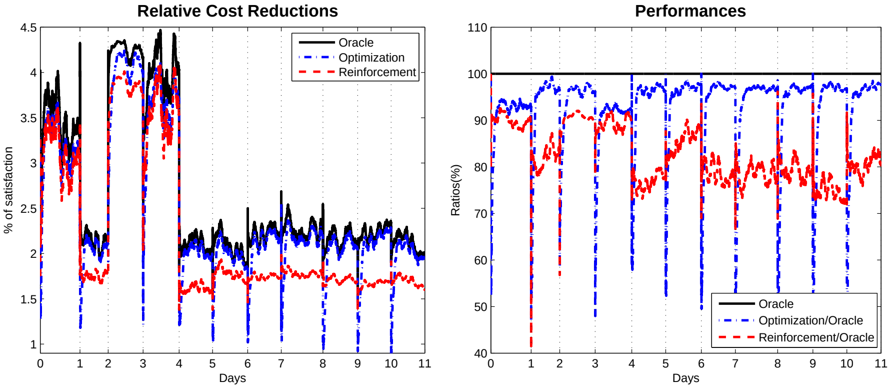

Figure: *Daily resetting of the algorithms parameters*: Case $N = 4$, $\sum_{i=1}^{N} \beta_i < 1$, $0.2 \leq \alpha_i \leq 0.8$ and $r_i^0 = 1/N$ $1 \leq i \leq N$.

Implementation

- Use the datasets on Paris Bourse: you have 4 assets (Bouygues, LVMH, Sanofi and Total). Select two weeks of data.
- Choose one of the asset for the volume you want to trade and the other to build the volumes on the dark pools.
- Fix an horizon T (in seconds) and compute the executed quantity on each period of T seconds executed for each assets (namely V and S_i).
- Choose the recombining coefficients α_i and the scaling factors β_i to be in the shortage setting and compute the volumes D_i.
- Choose ρ and implement both the stochastic algorithm and the oracle procedure. Print and comment your results.

Outline

Outline

Modeling the process of execution

We consider on a short period T a **Poisson process of "execution"** of buy orders

$$\left(N_t^{(\delta)}\right)_{0\leq t\leq T} \quad \text{with intensity} \quad \Lambda_T(\delta, S) := \int_0^T \lambda(S_t - (S_0 - \delta))dt$$

Modeling the process of execution

We consider on a short period T a **Poisson process of "execution"** of buy orders

$$\left(N_t^{(\delta)}\right)_{0\le t\le T} \quad \text{with intensity} \quad \Lambda_T(\delta, S) := \int_0^T \lambda(S_t - (S_0 - \delta))dt$$

where

- $0 \le \delta \le \delta_{\max}$ with $\delta_{\max} \in (0, S_0)$ denotes the *depth of the limit order book*,

Modeling the process of execution

We consider on a short period T a **Poisson process of "execution"** of buy orders

$$\left(N_t^{(\delta)}\right)_{0\leq t\leq T} \quad \text{with intensity} \quad \Lambda_T(\delta, S) := \int_0^T \lambda(S_t - (S_0 - \delta))dt$$

where

- $0 \leq \delta \leq \delta_{\max}$ with $\delta_{\max} \in (0, S_0)$ denotes the *depth of the limit order book*,
- $(S_t)_{t\geq 0}$ is a stochastic process modeling the dynamic of the *best ask price of a security stock*,

Modeling the process of execution

We consider on a short period T a **Poisson process of "execution"** of buy orders

$$\left(N_t^{(\delta)}\right)_{0\leq t\leq T} \quad \text{with intensity} \quad \Lambda_T(\delta, S) := \int_0^T \lambda(S_t - (S_0 - \delta))dt$$

where

- $0 \leq \delta \leq \delta_{\max}$ with $\delta_{\max} \in (0, S_0)$ denotes the *depth of the limit order book*,
- $(S_t)_{t\geq 0}$ is a stochastic process modeling the dynamic of the *best ask price of a security stock*,
- the function λ is defined on the whole real line as a finite non increasing convex function. We will use $\lambda(x) = Ae^{-ax}$.

Estimation of λ

Let $(\tau_n)_{n\geq 1}$ be the sequence of trade arrivals on $[0, T]$ assumed to be i.i.d. r.v. with exponential distribution of parameter λ.

Estimation of λ

Let $(\tau_n)_{n\geq 1}$ be the sequence of trade arrivals on $[0, T]$ assumed to be i.i.d. r.v. with exponential distribution of parameter λ. We estimate λ with the estimator $\hat{\lambda}_n$ defined by

$$\hat{\lambda}_n = \frac{\sum_{k=1}^{n} \mathbb{1}_{\{\tau_k < T\}}}{\sum_{k=1}^{n} \tau_k \wedge T}.$$

Estimation of λ

Let $(\tau_n)_{n\geq 1}$ be the sequence of trade arrivals on $[0, T]$ assumed to be i.i.d. r.v. with exponential distribution of parameter λ. We estimate λ with the estimator $\hat{\lambda}_n$ defined by

$$\hat{\lambda}_n = \frac{\sum_{k=1}^{n} \mathbb{1}_{\{\tau_k < T\}}}{\sum_{k=1}^{n} \tau_k \wedge T}.$$

This estimator converges *a.s.* toward λ and satisfies a CLT, namely

$$\hat{\lambda}_n \xrightarrow[n\to\infty]{a.s.} \lambda \quad \text{and} \quad \sqrt{n}\left(\hat{\lambda}_n - \lambda\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, \sigma).$$

Example on market data

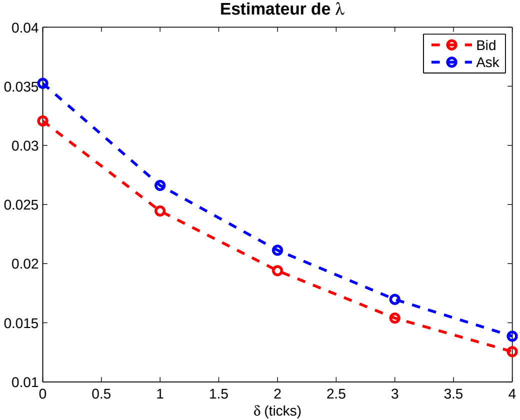

Figure: $\hat{\lambda}$ on the bid and ask sides for ACCOR on january 2012 for 5 limits and $T = 15$ trades.

Estimation of A and a

Assume now that the execution time depends on the posting distance with the following parametrization: $\hat{\lambda}_n(\delta) = Ae^{-a\delta}$. We look for estimating the parameters A and a from the estimator $\hat{\lambda}_n(\delta)$.

Estimation of A and a

Assume now that the execution time depends on the posting distance with the following parametrization: $\hat{\lambda}_n(\delta) = Ae^{-a\delta}$. We look for estimating the parameters A and a from the estimator $\hat{\lambda}_n(\delta)$. We will minimize the prediction error with the following linear regression

$$\log \hat{\lambda}_n(\delta) = c + \beta\delta + \epsilon.$$

Estimation of A and a

Assume now that the execution time depends on the posting distance with the following parametrization: $\hat{\lambda}_n(\delta) = Ae^{-a\delta}$. We look for estimating the parameters A and a from the estimator $\hat{\lambda}_n(\delta)$. We will minimize the prediction error with the following linear regression

$$\log \hat{\lambda}_n(\delta) = c + \beta\delta + \epsilon.$$

Then

$$\hat{A} = \exp(\hat{c}) \quad \text{and} \quad \hat{a} = -\hat{\beta}.$$

Application on market data

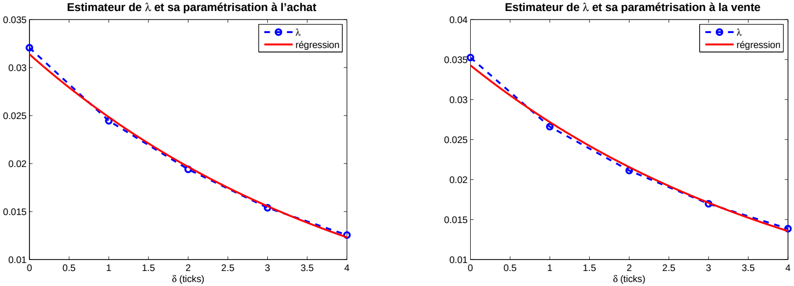

Figure: Comparison of parametrizations of estimators $\hat{\lambda}$ at the bid and the ask sides for ACCOR on january 2012 obtained by linear regression: $A_{buy} = 0.0314$, $a_{buy} = 0.2339$, $A_{sell} = 0.0343$, $a_{sell} = 0.2318$.

Outline

Design of the cost function

- Over the period $[0, T]$, we aim at executing a portfolio of size $Q_T \in \mathbb{N}$ invested in the asset S.

Design of the cost function

- Over the period $[0, T]$, we aim at executing a portfolio of size $Q_T \in \mathbb{N}$ invested in the asset S.
- At time $t = 0$, a buy order of size Q_T is posted at price $S_0 - \delta$.

Design of the cost function

- Over the period $[0, T]$, we aim at executing a portfolio of size $Q_T \in \mathbb{N}$ invested in the asset S.
- At time $t = 0$, a buy order of size Q_T is posted at price $S_0 - \delta$.
- On $[t, t + \Delta t]$, the *probability for such an order to be executed* is $\lambda(S_t - (S_0 - \delta))\Delta t = A \exp\left(-a\left(S_t - (S_0 - \delta)\right)\right)\Delta t$.

Design of the cost function

- Over the period $[0, T]$, we aim at executing a portfolio of size $Q_T \in \mathbb{N}$ invested in the asset S.
- At time $t = 0$, a buy order of size Q_T is posted at price $S_0 - \delta$.
- On $[t, t + \Delta t]$, the *probability for such an order to be executed* is $\lambda(S_t - (S_0 - \delta))\Delta t = A \exp\left(-a\left(S_t - (S_0 - \delta)\right)\right)\Delta t$.
- The *execution cost* for a distance δ is $\mathbb{E}\left[(S_0 - \delta)\left(Q_T \wedge N_T^{(\delta)}\right)\right]$.

Design of the cost function

- Over the period $[0, T]$, we aim at executing a portfolio of size $Q_T \in \mathbb{N}$ invested in the asset S.
- At time $t = 0$, a buy order of size Q_T is posted at price $S_0 - \delta$.
- On $[t, t + \Delta t]$, the *probability for such an order to be executed* is $\lambda(S_t - (S_0 - \delta))\Delta t = A \exp\left(-a\left(S_t - (S_0 - \delta)\right)\right)\Delta t$.
- The *execution cost* for a distance δ is $\mathbb{E}\left[(S_0 - \delta)\left(Q_T \wedge N_T^{(\delta)}\right)\right]$.
- We add to this execution cost a **penalization** depending on the *remaining quantity* to execute $\left(Q_T - N_T^{(\delta)}\right)_+$, namely at the end of the period T, we want to have Q_T assets in the portfolio.

Optimization Problem

Then we introduce a **market impact penalization function** $\Phi : \mathbb{R} \mapsto \mathbb{R}_+$, nondecreasing and convex, with $\Phi(0) = 0$ to model the additional cost of the execution of the remaining quantity.

Optimization Problem

Then we introduce a **market impact penalization function** $\Phi : \mathbb{R} \mapsto \mathbb{R}_+$, nondecreasing and convex, with $\Phi(0) = 0$ to model the additional cost of the execution of the remaining quantity.

Then the *resulting cost of execution* on a period $[0, T]$ reads

$$C(\delta) := \mathbb{E}\left[(S_0 - \delta)\left(Q_T \wedge N_T^{(\delta)}\right) + \kappa S_T \Phi\left(\left(Q_T - N_T^{(\delta)}\right)_+\right)\right]$$

where $\kappa > 0$.

Optimization Problem

Then we introduce a **market impact penalization function** $\Phi : \mathbb{R} \mapsto \mathbb{R}_+$, nondecreasing and convex, with $\Phi(0) = 0$ to model the additional cost of the execution of the remaining quantity.

Then the *resulting cost of execution* on a period $[0, T]$ reads

$$C(\delta) := \mathbb{E}\left[(S_0 - \delta)\left(Q_T \wedge N_T^{(\delta)}\right) + \kappa S_T \Phi\left(\left(Q_T - N_T^{(\delta)}\right)_+\right)\right]$$

where $\kappa > 0$. Our aim is then to **minimize this cost**, namely to solve the following optimization problem

$$\min_{0 \le \delta \le \delta_{\max}} C(\delta).$$

Optimization Problem

Then we introduce a **market impact penalization function** $\Phi : \mathbb{R} \mapsto \mathbb{R}_+$, nondecreasing and convex, with $\Phi(0) = 0$ to model the additional cost of the execution of the remaining quantity.

Then the *resulting cost of execution* on a period $[0, T]$ reads

$$C(\delta) := \mathbb{E}\left[(S_0 - \delta)\left(Q_T \wedge N_T^{(\delta)}\right) + \kappa S_T \Phi\left(\left(Q_T - N_T^{(\delta)}\right)_+\right)\right]$$

where $\kappa > 0$. Our aim is then to **minimize this cost**, namely to solve the following optimization problem

$$\min_{0 \le \delta \le \delta_{\max}} C(\delta).$$

To solve this optimization problem, we will devise a **stochastic algorithm constrained** to stay in $[0, \delta_{\max}]$.

Strategy

We will devise a **stochastic algorithm constrained** to stay in $[0, \delta_{\max}]$.
To this end we have to

Strategy

We will devise a **stochastic algorithm constrained** to stay in $[0, \delta_{\max}]$.
To this end we have to

- prove that C, C' and C'' admit representations as expectations, *i.e.* in particular, there exists a Borel functional

$$H : [0, \delta_{\max}] \times \mathbb{D}([0,T], \mathbb{R}) \longrightarrow \mathbb{R} \quad \text{such that}$$

$$\forall \delta \in [0, \delta_{\max}], \quad C'(\delta) = \mathbb{E}\left[H\left(\delta, (S_t)_{t \in [0,T]}\right)\right]$$

Strategy

We will devise a **stochastic algorithm constrained** to stay in $[0, \delta_{\max}]$.
To this end we have to

- prove that C, C' and C'' admit representations as expectations, *i.e.* in particular, there exists a Borel functional

$$H : [0, \delta_{\max}] \times \mathbb{D}([0, T], \mathbb{R}) \longrightarrow \mathbb{R} \quad \text{such that}$$

$$\forall \delta \in [0, \delta_{\max}], \quad C'(\delta) = \mathbb{E}\left[H\left(\delta, (S_t)_{t\in[0,T]}\right)\right]$$

- find natural assumptions on Q_T and κ such that C is twice differentiable, strictly convex on $[0, \delta_{\max}]$ with $C'(0) < 0$.

Strategy

We will devise a **stochastic algorithm constrained** to stay in $[0, \delta_{\max}]$.
To this end we have to

- prove that C, C' and C'' admit representations as expectations, *i.e.* in particular, there exists a Borel functional

$$H : [0, \delta_{\max}] \times \mathbb{D}([0, T], \mathbb{R}) \longrightarrow \mathbb{R} \quad \text{such that}$$

$$\forall \delta \in [0, \delta_{\max}], \quad C'(\delta) = \mathbb{E}\left[H\left(\delta, (S_t)_{t\in[0,T]}\right)\right]$$

- find natural assumptions on Q_T and κ such that C is twice differentiable, strictly convex on $[0, \delta_{\max}]$ with $C'(0) < 0$.Consequently

$$\operatorname{argmin}_{\delta\in[0,\delta_{\max}]} C(\delta) = \{\delta^*\}, \quad \delta^* \in (0, \delta_{\max}]$$

and

$$\delta^* = \delta_{\max} \text{ iff } C \text{ is non increasing on } [0, \delta_{\max}].$$

Design of the algorithm

Once the two points are checked, we can devise the algorithm following the **standard stochastic approximation with projection**, namely

$$\delta_{n+1} = \mathrm{Proj}_{[0,\delta_{\max}]}\left(\delta_n - \gamma_{n+1} H\left(\delta_n, \left(\bar{S}_{t_i}^{(n+1)}\right)_{0\le i\le m}\right)\right), \ \delta_0 \in [0, \delta_{\max}],$$

Design of the algorithm

Once the two points are checked, we can devise the algorithm following the **standard stochastic approximation with projection**, namely

$$\delta_{n+1} = \text{Proj}_{[0,\delta_{\max}]}\left(\delta_n - \gamma_{n+1} H\left(\delta_n, \left(\bar{S}_{t_i}^{(n+1)}\right)_{0\le i\le m}\right)\right), \; \delta_0 \in [0, \delta_{\max}],$$

where

- $\text{Proj}_{[0,\delta_{\max}]}$ denotes the projection on $[0, \delta_{\max}]$,

Design of the algorithm

Once the two points are checked, we can devise the algorithm following the **standard stochastic approximation with projection**, namely

$$\delta_{n+1} = \mathrm{Proj}_{[0,\delta_{\max}]}\left(\delta_n - \gamma_{n+1} H\left(\delta_n, \left(\bar{S}_{t_i}^{(n+1)}\right)_{0\le i\le m}\right)\right),\ \delta_0 \in [0, \delta_{\max}],$$

where

- $\mathrm{Proj}_{[0,\delta_{\max}]}$ denotes the projection on $[0, \delta_{\max}]$,
- the positive step sequence $(\gamma_n)_{n\ge1}$ satisfies at least the minimal *decreasing step* assumption

$$\sum_{n\ge1} \gamma_n = +\infty \quad \text{and} \quad \gamma_n \to 0,$$

Design of the algorithm

Once the two points are checked, we can devise the algorithm following the **standard stochastic approximation with projection**, namely

$$\delta_{n+1} = \text{Proj}_{[0,\delta_{\max}]}\left(\delta_n - \gamma_{n+1} H\left(\delta_n, \left(\bar{S}_{t_i}^{(n+1)}\right)_{0\leq i\leq m}\right)\right), \ \delta_0 \in [0, \delta_{\max}],$$

where

- $\text{Proj}_{[0,\delta_{\max}]}$ denotes the projection on $[0, \delta_{\max}]$,
- the positive step sequence $(\gamma_n)_{n\geq 1}$ satisfies at least the minimal *decreasing step* assumption

$$\sum_{n\geq 1} \gamma_n = +\infty \quad \text{and} \quad \gamma_n \to 0,$$

- $\left\{\left(\bar{S}_{t_i}^{(n)}\right)_{0\leq i\leq m}, n \geq 0\right\}$ is either a sequence of i.i.d. copies of the true underlying dynamics of $(S_{t_i})_{0\leq i\leq m}$ or at least of its Euler scheme or a sequence sharing some averaging properties (*e.g*. stationary α-mixing).

Outline

1. Introduction to stochastic approximation
   - Motivations and framework
   - *A.s.* convergence
   - Rates of Convergence
   - Averaging principle
2. Optimal split of orders across liquidity pools
   - Modeling and mean execution cost
   - Design of the stochastic algorithm
   - Numerical experiments
3. Optimal posting price of limit orders: learning by trading
   - Execution process and intensity
   - Design of the stochastic algorithm
   - Cost function and its derivative as expectations
   - Numerical experiments

Basic facts on Poisson distributed variables

Let $(N^\mu)_{\mu>0}$ be a family of Poisson distributed random variables with parameter $\mu > 0$.

Basic facts on Poisson distributed variables

Let $(N^\mu)_{\mu>0}$ be a family of Poisson distributed random variables with parameter $\mu > 0$.

(*i*) For every function $f : \mathbb{N} \to \mathbb{R}_+$ such that $\log f(n) = O(n)$,

$$\frac{d}{d\mu}\mathbb{E}\left[f(N^\mu)\right] = \mathbb{E}\left[f(N^\mu+1) - f(N^\mu)\right] = \mathbb{E}\left[f(N^\mu)\left(\frac{N^\mu}{\mu} - 1\right)\right].$$

In particular, for any $k \in \mathbb{N}$, $\frac{d}{d\mu}\mathbb{P}\left(N^\mu \le k\right) = -\mathbb{P}\left(N^\mu = k\right)$.

Basic facts on Poisson distributed variables

Let $(N^\mu)_{\mu>0}$ be a family of Poisson distributed random variables with parameter $\mu > 0$.

(i) For every function $f : \mathbb{N} \to \mathbb{R}_+$ such that $\log f(n) = O(n)$,

$$\frac{d}{d\mu}\mathbb{E}\left[f(N^\mu)\right] = \mathbb{E}\left[f(N^\mu + 1) - f(N^\mu)\right] = \mathbb{E}\left[f(N^\mu)\left(\frac{N^\mu}{\mu} - 1\right)\right].$$

In particular, for any $k \in \mathbb{N}$, $\frac{d}{d\mu}\mathbb{P}\left(N^\mu \leq k\right) = -\mathbb{P}\left(N^\mu = k\right)$.

For any $k \in \mathbb{N}^*$,

(ii) $\mathbb{E}\left[k \wedge N^\mu\right] = k\mathbb{P}\left(N^\mu > k\right) + \mu\mathbb{P}\left(N^\mu \leq k - 1\right)$ and $\frac{d}{d\mu}\mathbb{E}\left[k \wedge N^\mu\right] = \mathbb{P}\left(N^\mu \leq k - 1\right)$,

Basic facts on Poisson distributed variables

Let $(N^\mu)_{\mu>0}$ be a family of Poisson distributed random variables with parameter $\mu > 0$.

(i) For every function $f : \mathbb{N} \to \mathbb{R}_+$ such that $\log f(n) = O(n)$,

$$\frac{d}{d\mu}\mathbb{E}\left[f(N^\mu)\right] = \mathbb{E}\left[f(N^\mu+1) - f(N^\mu)\right] = \mathbb{E}\left[f(N^\mu)\left(\frac{N^\mu}{\mu} - 1\right)\right].$$

In particular, for any $k \in \mathbb{N}$, $\frac{d}{d\mu}\mathbb{P}\left(N^\mu \le k\right) = -\mathbb{P}\left(N^\mu = k\right)$.

For any $k \in \mathbb{N}^*$,

(ii) $\mathbb{E}\left[k \wedge N^\mu\right] = k\mathbb{P}\left(N^\mu > k\right) + \mu\mathbb{P}\left(N^\mu \le k-1\right)$ and $\frac{d}{d\mu}\mathbb{E}\left[k \wedge N^\mu\right] = \mathbb{P}\left(N^\mu \le k-1\right)$,

(iii) $\mathbb{E}\left[(k - N^\mu)_+\right] = k\mathbb{P}\left(N^\mu \le k\right) - \mu\mathbb{P}\left(N^\mu \le k-1\right)$,

Basic facts on Poisson distributed variables

Let $(N^\mu)_{\mu>0}$ be a family of Poisson distributed random variables with parameter $\mu > 0$.

(i) For every function $f : \mathbb{N} \to \mathbb{R}_+$ such that $\log f(n) = O(n)$,

$$\frac{d}{d\mu}\mathbb{E}\left[f(N^\mu)\right] = \mathbb{E}\left[f(N^\mu+1) - f(N^\mu)\right] = \mathbb{E}\left[f(N^\mu)\left(\frac{N^\mu}{\mu} - 1\right)\right].$$

In particular, for any $k \in \mathbb{N}$, $\frac{d}{d\mu}\mathbb{P}\left(N^\mu \le k\right) = -\mathbb{P}\left(N^\mu = k\right)$.

For any $k \in \mathbb{N}^*$,

(ii) $\mathbb{E}\left[k \wedge N^\mu\right] = k\mathbb{P}\left(N^\mu > k\right) + \mu\mathbb{P}\left(N^\mu \le k-1\right)$ and $\frac{d}{d\mu}\mathbb{E}\left[k \wedge N^\mu\right] = \mathbb{P}\left(N^\mu \le k-1\right)$,

(iii) $\mathbb{E}\left[(k - N^\mu)_+\right] = k\mathbb{P}\left(N^\mu \le k\right) - \mu\mathbb{P}\left(N^\mu \le k-1\right)$,

(iv) $k\mathbb{P}\left(N^\mu = k\right) = \mu\mathbb{P}\left(N^\mu = k-1\right)$.

Representation as expectations of C and C'

To compute the cost function (or its gradient), it is convenient to use a pre-conditioning with respect to $\mathcal{F}_T^S := \sigma\left(S_t, 0 \leq t \leq T\right)$.

Representation as expectations of C and C'

To compute the cost function (or its gradient), it is convenient to use a pre-conditioning with respect to $\mathcal{F}_T^S := \sigma\left(S_t, 0 \leq t \leq T\right)$.We come down to compute the above quantity when $N^{(\delta)}$ is replaced by a standard Poisson random variable with parameter μ denoted N^{μ}.

Representation as expectations of C and C'

To compute the cost function (or its gradient), it is convenient to use a pre-conditioning with respect to $\mathcal{F}_T^S := \sigma\left(S_t, 0 \leq t \leq T\right)$.We come down to compute the above quantity when $N^{(\delta)}$ is replaced by a standard Poisson random variable with parameter μ denoted N^{μ}.

$$C(\delta) = \mathbb{E}\left[\widetilde{C}\left(\delta, \Lambda_T(\delta, S), (S_t)_{0\leq t\leq T}\right)\right],$$

where for every $x \in \mathcal{C}([0, T], \mathbb{R}_+)$ and every $\mu \in \mathbb{R}_+$,

$$\begin{aligned}\widetilde{C}\left(\delta, \mu, x\right) &= (x_0 - \delta)\left(Q_T \mathbb{P}\left(N^{\mu} > Q_T\right) + \mu \mathbb{P}\left(N^{\mu} \leq Q_T - 1\right)\right) \\ &\quad + \kappa x_T \mathbb{E}\left[\Phi\left(\left(Q_T - N^{\mu}\right)_+\right)\right].\end{aligned}$$

Representation as expectations of C and C'

To compute the cost function (or its gradient), it is convenient to use a pre-conditioning with respect to $\mathcal{F}^S_T := \sigma\left(S_t, 0 \le t \le T\right)$.We come down to compute the above quantity when $N^{(\delta)}$ is replaced by a standard Poisson random variable with parameter μ denoted N^μ.

$$C(\delta) = \mathbb{E}\left[\widetilde{C}\left(\delta, \Lambda_T(\delta, S), (S_t)_{0\le t\le T}\right)\right],$$

where for every $x \in \mathcal{C}([0, T], \mathbb{R}_+)$ and every $\mu \in \mathbb{R}_+$,

$$\begin{aligned}\widetilde{C}\left(\delta, \mu, x\right) &= (x_0 - \delta)\left(Q_T \mathbb{P}\left(N^\mu > Q_T\right) + \mu \mathbb{P}\left(N^\mu \le Q_T - 1\right)\right) \\ &\quad + \kappa x_T \mathbb{E}\left[\Phi\left(\left(Q_T - N^\mu\right)_+\right)\right].\end{aligned}$$

We introduce some notations for reading convenience: we set

$$\mathbb{P}^{(\delta)}\left(N^\mu > Q_T\right) = \mathbb{P}\left(N^\mu > Q_T\right)_{|\mu=\Lambda_T(\delta,S)}$$

and

$$\mathbb{E}^{(\delta)}\left[f(\mu)\right] = \mathbb{E}\left[f(\mu)\right]_{|\mu=\Lambda_T(\delta,S)}.$$

Representation as expectations of C and C'

- If $\Phi \neq \mathrm{id}$, then $C'(\delta) = \mathbb{E}\left[H(\delta, S)\right]$ with

$$\begin{aligned} H(\delta, S) \;&=\; -Q_T \mathbb{P}^{(\delta)}\left(N^\mu > Q_T\right) - \kappa S_T \frac{\partial}{\partial \delta} \Lambda_T(\delta, S) \varphi^{(\delta)}(\mu) \\ &\quad + \left(\frac{\partial}{\partial \delta} \Lambda_T(\delta, S)(S_0 - \delta) - \Lambda_T(\delta, S)\right) \mathbb{P}^{(\delta)}\left(N^\mu \leq Q_T - 1\right) \end{aligned}$$

where $\varphi^{(\delta)}(\mu) = \mathbb{E}^{(\delta)}\left[\left(\Phi\left(Q_T - N^\mu\right) - \Phi\left(Q_T - N^\mu - 1\right)\right) \mathbb{1}_{\{N^\mu \leq Q_T - 1\}}\right]$.

Representation as expectations of C and C'

- If $\Phi \neq \mathrm{id}$, then $C'(\delta) = \mathbb{E}\left[H(\delta, S)\right]$ with

$$
\begin{aligned}
H(\delta, S) \quad = \quad & -Q_T \mathbb{P}^{(\delta)}\left(N^\mu > Q_T\right) - \kappa S_T \frac{\partial}{\partial \delta} \Lambda_T(\delta, S) \varphi^{(\delta)}(\mu) \\
& + \left(\frac{\partial}{\partial \delta} \Lambda_T(\delta, S)(S_0 - \delta) - \Lambda_T(\delta, S)\right) \mathbb{P}^{(\delta)}\left(N^\mu \leq Q_T - 1\right)
\end{aligned}
$$

where $\varphi^{(\delta)}(\mu) = \mathbb{E}^{(\delta)}\left[\left(\Phi\left(Q_T - N^\mu\right) - \Phi\left(Q_T - N^\mu - 1\right)\right) \mathbb{1}_{\{N^\mu \leq Q_T - 1\}}\right]$.

- If $\Phi = \mathrm{id}$, then $C'(\delta) = \mathbb{E}\left[H(\delta, S)\right]$ with

$$
\begin{aligned}
H(\delta, S) \quad = \quad & -Q_T \mathbb{P}^{(\delta)}\left(N^\mu > Q_T\right) \\
+ \quad & \left((S_0 - \delta - \kappa S_T) \frac{\partial}{\partial \delta} \Lambda_T(\delta, S) - \Lambda_T(\delta, S)\right) \mathbb{P}^{(\delta)}\left(N^\mu \leq Q_T - 1\right).
\end{aligned}
$$

Criteria for the monotony at the origin and convexity

Assume that $(S_t)_{t\in[0,T]}$ satisfies a functional co-monotony principle. To simplify, set $\lambda(x) = Ae^{-ax}$, $A, a > 0$.

Criteria for the monotony at the origin and convexity

Assume that $(S_t)_{t\in[0,T]}$ satisfies a functional co-monotony principle. To simplify, set $\lambda(x) = Ae^{-ax}$, $A, a > 0$. (a) **Monotony at the origin:**

$C'(0) < 0$ as soon as

$$Q_T \geq 2T\lambda(-S_0) \quad \text{and} \quad \kappa \leq \frac{1 + aS_0}{a\mathbb{E}\left[S_T\right]\left(\Phi(Q_T) - \Phi(Q_T - 1)\right)}.$$

In particular, when $\Phi \equiv \mathrm{id}$, the condition reduces to

$$\kappa \leq \frac{1 + aS_0}{a\mathbb{E}\left[S_T\right]}.$$

Criteria for the monotony at the origin and convexity

Assume that $(S_t)_{t\in[0,T]}$ satisfies a functional co-monotony principle. To simplify, set $\lambda(x) = Ae^{-ax}$, $A, a > 0$. (a) **Monotony at the origin:**

$C'(0) < 0$ as soon as

$$Q_T \geq 2T\lambda(-S_0) \quad \text{and} \quad \kappa \leq \frac{1 + aS_0}{a\mathbb{E}\left[S_T\right]\left(\Phi(Q_T) - \Phi(Q_T - 1)\right)}.$$

In particular, when $\Phi \equiv \mathrm{id}$, the condition reduces to

$$\kappa \leq \frac{1 + aS_0}{a\mathbb{E}\left[S_T\right]}.$$

(b) **Convexity.** If $\Phi \neq \mathrm{id}$, assume that Φ satisfies

$$\forall x \in [1, Q_T - 1], \quad \Phi(x) - \Phi(x-1) \leq \rho_Q(\Phi(x+1) - \Phi(x)), \quad \rho_Q \in (0,1).$$

$C''(\delta) \geq 0$, $\delta \in [0, \delta_{\max}]$ (so that C is convex on $[0, \delta_{\max}]$) if

$$Q_T \geq 2T\lambda(-S_0) \quad \text{and} \quad \kappa \leq \frac{2}{a\mathbb{E}\left[S_T\right]\Phi'_\ell(Q_T)}.$$

Outline

1. Introduction to stochastic approximation
   - Motivations and framework
   - *A.s.* convergence
   - Rates of Convergence
   - Averaging principle
2. Optimal split of orders across liquidity pools
   - Modeling and mean execution cost
   - Design of the stochastic algorithm
   - Numerical experiments
3. Optimal posting price of limit orders: learning by trading
   - Execution process and intensity
   - Design of the stochastic algorithm
   - Cost function and its derivative as expectations
   - Numerical experiments

Numerical experiments: Simulated data

We assume that

$$dS_t = \sigma dW_t, \quad S_0 = s_0, \quad \text{and} \quad \Lambda_T(\delta, S) = A \int_0^T e^{-a(S_t - S_0 + \delta)} dt$$

where $(W_t)_{t \geq 0}$ is a standard Brownian motion and $\sigma, A, a > 0$ (this means that $\lambda(x) = Ae^{-ax}$).

Numerical experiments: Simulated data

We assume that

$$dS_t = \sigma dW_t, \quad S_0 = s_0, \quad \text{and} \quad \Lambda_T(\delta, S) = A \int_0^T e^{-a(S_t - S_0 + \delta)} dt$$

where $(W_t)_{t \geq 0}$ is a standard Brownian motion and $\sigma, A, a > 0$ (this means that $\lambda(x) = Ae^{-ax}$).We denote by $(\bar{S}_t)_{t \geq 0}$ the Euler scheme of $(S_t)_{t \geq 0}$ defined by

$$\bar{S}_{k+1} := \bar{S}_k + \sigma \sqrt{\frac{T}{m}} Z_{k+1}, \quad \bar{S}_0 = s_0, \quad Z_{k+1} \sim \mathcal{N}(0,1), \quad k \geq 0,$$

and we approximate $\Lambda_T(\delta, S)$ by

$$\bar{\Lambda}_T(\delta, S) = A \frac{T}{m} \sum_{k=1}^{m} e^{-a(\bar{S}_k - S_0 + \delta)}.$$

Numerical experiments: Simulated data

We assume that

$$dS_t = \sigma dW_t, \quad S_0 = s_0, \quad \text{and} \quad \Lambda_T(\delta, S) = A \int_0^T e^{-a(S_t - S_0 + \delta)} dt$$

where $(W_t)_{t \geq 0}$ is a standard Brownian motion and $\sigma, A, a > 0$ (this means that $\lambda(x) = Ae^{-ax}$).We denote by $(\bar{S}_t)_{t \geq 0}$ the Euler scheme of $(S_t)_{t \geq 0}$ defined by

$$\bar{S}_{k+1} := \bar{S}_k + \sigma \sqrt{\frac{T}{m}} Z_{k+1}, \quad \bar{S}_0 = s_0, \quad Z_{k+1} \sim \mathcal{N}(0,1), \quad k \geq 0,$$

and we approximate $\Lambda_T(\delta, S)$ by

$$\bar{\Lambda}_T(\delta, S) = A \frac{T}{m} \sum_{k=1}^{m} e^{-a(\bar{S}_k - S_0 + \delta)}.$$

The penalization function is of the following form

$$\Phi(x) = (1 + \eta(x))x \quad \text{with} \quad \eta(x) = A' e^{a' x}.$$

Simulated data

Now we present the cost function and its derivative for the following parameters:

- parameters of the asset dynamics: $s_0 = 100$ and $\sigma = 0.005$,
- parameters of the intensity of the execution process: $A = 5$ and $a = 1$,
- parameters of the execution: $T = 5$ and $Q = 10$,
- parameters of the penalization function: $A' = 1$ and $a' = 0.01$.

We use $m = 20$ for the Euler scheme and $M = 10\,000$ simulations Monte-Carlo.

Cost function and its derivative

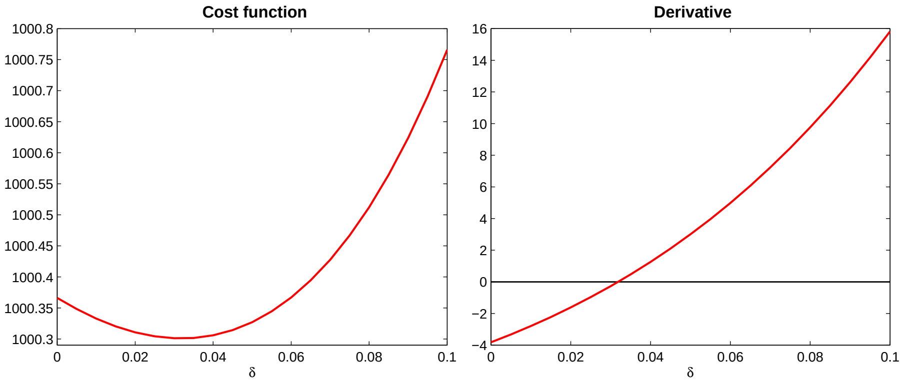

Figure: $\eta \neq 0$: $T = 5$, $A = 5$, $a = 1$, $s_0 = 100$, $\sigma = 0.005$, $Q = 10$, $\kappa = 6$, $A' = 1$, $a' = 0.01$, $m = 20$ and $M = 10000$.

δ obtained by SA

Now we present the result of the stochastic recursive procedure for the two cases with $n = 100$ and $\gamma_n = \frac{1}{200n}$.

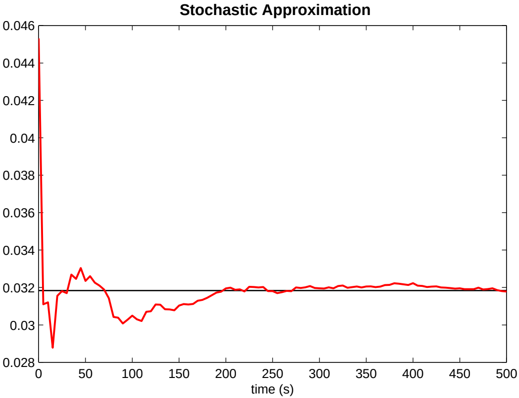

Figure: $\eta \not\equiv 0$: $T = 5$, $A = 5$, $a = 1$, $s_0 = 100$, $\sigma = 0.005$, $Q = 10$, $\kappa = 6$, $A' = 1$, $a' = 0.01$, $m = 20$ and $n = 100$.

Numerical experiments: Market data

As market data, we use the bid prices of Accor SA (ACCP.PA) of $11/11/2010$ for the fair price process $(S_t)_{t\in[0,T]}$.

Numerical experiments: Market data

As market data, we use the bid prices of Accor SA (ACCP.PA) of 11/11/2010 for the fair price process $(S_t)_{t\in[0,T]}$.We divide the day into periods of 15 trades which will denote steps of the stochastic procedure.

Numerical experiments: Market data

As market data, we use the bid prices of Accor SA (ACCP.PA) of 11/11/2010 for the fair price process $(S_t)_{t\in[0,T]}$.We divide the day into periods of 15 trades which will denote steps of the stochastic procedure.Let N_{cycles} be the number of these periods. For every $n \in \{1, \ldots, N_{\text{cycles}}\}$, we have a sequence of bid prices $(\bar{S}_{t_i}^{(n)})_{0\leq i\leq 14}$ and we approximate the jump intensity of the Poisson process $\Lambda_{T^n}(\delta, S)$, where $T^n = \sum_{i=0}^{14} t_i$, by

$$\forall n \in N_{\text{cycles}}, \quad \Lambda_{T^n}(\delta, S) = A \sum_{i=1}^{14} e^{-a(\bar{S}_{t_i}^{(n)} - \bar{S}_{t_0} + \delta)}(t_i - t_{i-1}).$$

Numerical experiments: Market data

As market data, we use the bid prices of Accor SA (ACCP.PA) of 11/11/2010 for the fair price process $(S_t)_{t\in[0,T]}$.We divide the day into periods of 15 trades which will denote steps of the stochastic procedure.Let N_{cycles} be the number of these periods. For every $n \in \{1,\ldots,N_{\text{cycles}}\}$, we have a sequence of bid prices $(\bar{S}^{(n)}_{t_i})_{0\le i\le 14}$ and we approximate the jump intensity of the Poisson process $\Lambda_{T^n}(\delta,S)$, where $T^n = \sum_{i=0}^{14} t_i$, by

$$\forall n \in N_{\text{cycles}}, \quad \Lambda_{T^n}(\delta,S) = A\sum_{i=1}^{14} e^{-a(\bar{S}^{(n)}_{t_i} - \bar{S}_{t_0}+\delta)}(t_i - t_{i-1}).$$

The empirical mean of the intensity function

$$\bar{\Lambda}(\delta,S) = \frac{1}{N_{\text{cycles}}}\sum_{n=1}^{N_{\text{cycles}}} \Lambda_{T^n}(\delta,S)$$

is plotted on the following figure.

Intensity function

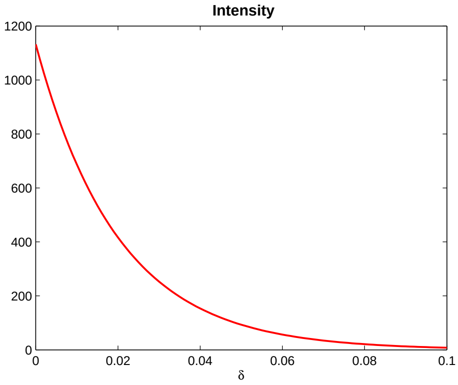

Figure: Fit of the exponential model on real data (Accor SA (ACCP.PA) 11/11/2010): $A = 1/50$, $a = 50$ and $N_{\mathrm{cycles}} = 220$.

The penalization function is of the following form

$$\Phi(x) = (1 + \eta(x))x \quad \text{with} \quad \eta(x) = A' e^{a' x}.$$

Cost function and its derivative

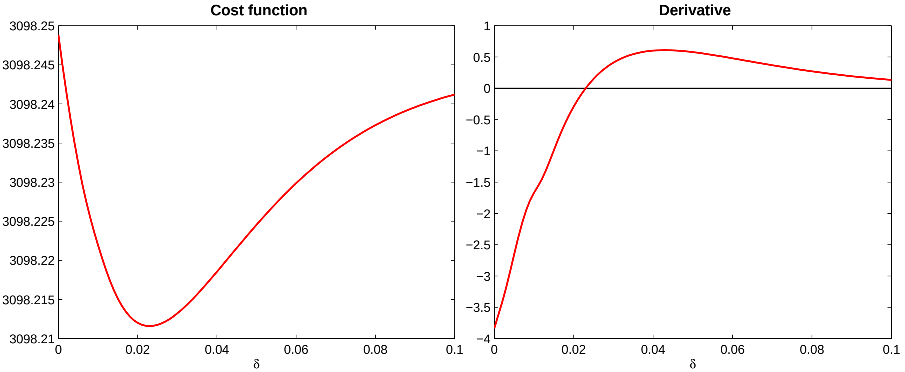

Figure: $\eta \not\equiv 0$: $A = 1/50$, $a = 50$, $Q = 100$, $\kappa = 1$, $A' = 0.001$, $a' = 0.0005$ and $N_{\text{cycles}} = 220$.

δ and posting price obtained by SA

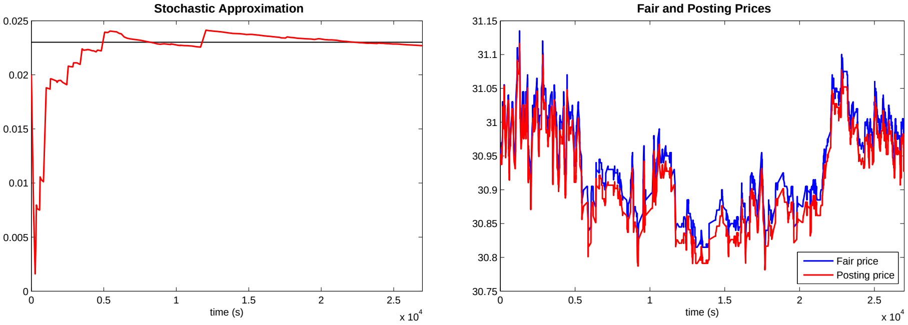

Figure: $\eta \neq 0$: $A = 1/50$, $a = 50$, $Q = 100$, $\kappa = 1$, $A' = 0.001$, $a' = 0.0005$ and $N_{\text{cycles}} = 220$. Crude algorithm with $\gamma_n = \frac{1}{550n}$.

Stochastic algorithm with RP averaging

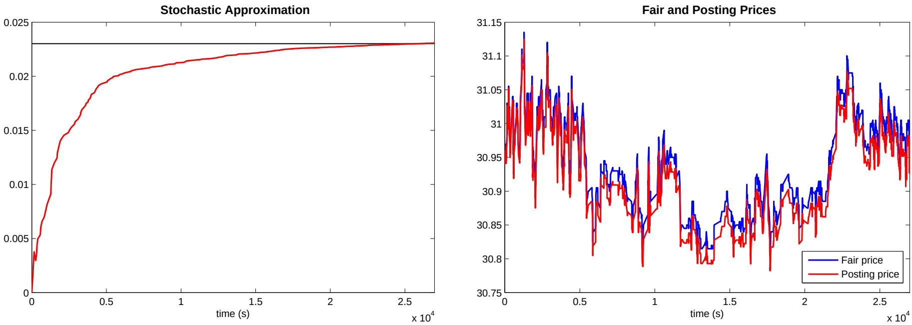

Figure: $\eta \not\equiv 0$: $A = 1/50$, $a = 50$, $Q = 100$, $\kappa = 1$, $A' = 0.001$, $a' = 0.0005$ and $N_{\mathrm{cycles}} = 220$. Averaging algorithm (Ruppert and Poliak) with $\gamma_n = \frac{1}{550 n^{0.95}}$.

Stochastic algorithm with RP averaging

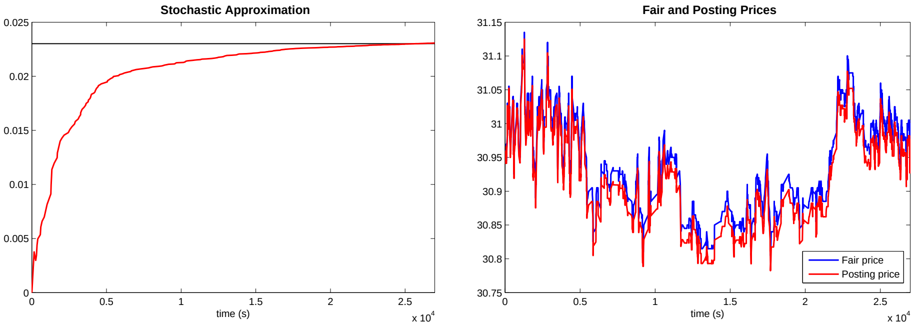

Figure: $\eta \not\equiv 0$: $A = 1/50$, $a = 50$, $Q = 100$, $\kappa = 1$, $A' = 0.001$, $a' = 0.0005$ and $N_{\text{cycles}} = 220$. Averaging algorithm (Ruppert and Poliak) with $\gamma_n = \frac{1}{550 n^{0.95}}$.

Reminder. Averaging algorithm (Ruppert and Poliak) $\bar{\delta}_n = \frac{1}{n+1} \sum_{k=0}^{n} \delta_k$, $n \geq 1$. The main asset of this averaging procedure is that, under appropriate assumptions, its weak rate of convergence is ruled by a *CLT* having a minimal asymptotic variance (among all recursive procedures).

Implementation

- Estimate on market data for one asset the intensity function λ depending on the first 5 limits.

- Estimate the coefficients A and a of the parametrization $\lambda(x) = Ae^{-ax}$.

- Fix T (in number of trades) and Q_T (approximately 1 ATS). Compute the cost function and its derivative using the previous parametrization (with the parameters given by the linear regression). You will have to tune κ to be sure that you have a convex cost.

- Implement the stochastic algorithm and check it converges toward the minimum (you could use averaging principle).

References

M. Benaïm, Dynamics of stochastic approximation algorithms, *Séminaire de Probabilités, XXXIII*, volume 1709 of *Lecture Notes in Math.*, p. 1-68. Springer, Berlin, 1999.

A. Benvéniste, M. Métivier, P. Priouret, *Adaptive algorithms and stochastic approximations*, volume 22 of *Applications of Mathematics (New York)*, Springer-Verlag, Berlin, 1990.

M. Duflo, *Algorihtmes Stochastiques*, volume 23 of Math. and Applications (Berlin), Springer-Verlag, Berlin, 1996.

M. Duflo, *Random iterative models*, volume 34 of *Applications of Mathematics (New York)*, Springer-Verlag, Berlin, 1997.

J.-C. Fort, G. Pagès, Convergence of stochastic algorithms: from the Kushner-Clark theorem to the Lyapounov functional method, *Adv. in Appl. Probab.*, 28(4):1072-1094,1996.

References

H. J. Kushner, D. S. Clark, *Stochastic approximation methods for constrained and unconstrained systems*, volume 26 of *Applied Mathematical Sciences*. Springer-Verlag, New York, second edition, 1978.

H. J. Kushner, G. G. Yin, *Stochastic approximation and recursive algorithms and applications*, volume 35 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 2003.

S. Laruelle, G. Pagès, *Stochastic Approximation with Averaging Innovation Applied to Finance*, Monte Carlo Methods and Applications, 18 (1) ,2012, pp. 1-51.

S. Laruelle, C.-A. Lehalle, G. Pagès, *Optimal split of orders across liquidity pools: a stochastic algorithm approach*, SIAM Journal of Financial Mathematics, Volume 2, pp.1042-1076, 2011.

S. Laruelle, C.-A. Lehalle, G. Pagès, *Optimal posting price of limit orders: learning by trading*, Mathematics and Financial Economics, Volume 7, Issue 3, pp359-403, 2013.

H. Robbins, S. Monro, *A stochastic approximation method*, Ann. Math. Stat.