Algorithme Stochastique avec Innovations Moyennisantes

Sophie Laruelle

LPMA - UPMC

$9^{\text{ème}}$ Colloque des Jeunes Probabilistes et Statisticiens

3-6 Mai 2010

Plan

1. Approximation récursive stochastique
2. Description du cadre et Motivations
3. Théorème de convergence
4. Applications
   - Innovations i.i.d.
   - Fonctionnelle d'un processus α-mélangeant
   - Chaîne de Markov homogène
   - Suites à discrépance faible
5. Exemples
   - Recherche de corrélation implicite

Plan

1. Approximation récursive stochastique
2. Description du cadre et Motivations
3. Théorème de convergence
4. Applications
   - Innovations i.i.d.
   - Fonctionnelle d'un processus α-mélangeant
   - Chaîne de Markov homogène
   - Suites à discrépance faible
5. Exemples
   - Recherche de corrélation implicite

Définition d'un algorithme stochastique

Il s'agit d'une **procédure récursive de recherche de zéro** d'une fonction dite *moyenne* car s'écrivant comme l'espérance d'une fonction dépendant du même paramètre et d'une v.a. appelée *innovation*, *i.e.*

$$\text{on cherche } \theta^* \text{ tel que } h(\theta^*) = 0 \text{ où } h(\theta) = \mathbb{E}\left[H(\theta, Y)\right].$$

Applications

- Recherche d'un paramètre pour un certain niveau de la fonction moyenne,
- Optimisation : annulation du gradient (*gradient stochastique*).

Exemple du dosage : Procédure de Robbins-Monro

Une dose θ d'un produit chimique crée un effet aléatoire mesuré par $F(\theta, Y)$, Y étant une v.a. et $F : \mathbb{R}^2 \mapsto \mathbb{R}$ une fonction inconnue. On suppose l'**effet moyen** $f(\theta) = \mathbb{E}\left[F(\theta, Y)\right]$ croissant. On cherche à déterminer la dose θ^* qui crée un effet moyen de niveau donné α, *i.e.* résoudre $f(\theta^*) = \alpha$.

La **procédure de Robbins-Monro** est la suivante

- choisir θ_0 arbitrairement et l'administrer à un sujet qui réagit avec l'effet $F(\theta_0, Y_1)$.
- *récurrence* : à l'instant n, choisir une dose θ_n administrée à un sujet indépendant des précédents, l'effet est $F(\theta_n, Y_{n+1})$.

(Y_n) est une suite de v.a. i.i.d. donc

$$f(\theta_n) = \mathbb{E}\left[F(\theta_n, Y_{n+1}) \mid F(\theta_0, Y_1), \ldots, F(\theta_{n-1}, Y_n)\right].$$

L'**algorithme de Robbins-Monro** de choix de θ_n est alors de la forme

$$\theta_{n+1} = \theta_n - \gamma_n \left(F(\theta_n, Y_{n+1}) - \alpha\right), \quad (\gamma_n) \text{ positive tendant vers } 0.$$

Plan

1. Approximation récursive stochastique
2. Description du cadre et Motivations
3. Théorème de convergence
4. Applications
   - Innovations i.i.d.
   - Fonctionnelle d'un processus α-mélangeant
   - Chaîne de Markov homogène
   - Suites à discrépance faible
5. Exemples
   - Recherche de corrélation implicite

Description du cadre

On s'intéresse à une procédure de la forme suivante

$$\theta_{n+1} = \theta_n - \gamma_{n+1}\left(H(\theta_n, Y_n) + \Delta M_{n+1}\right), \quad n \geq 0, \quad \theta_0 \in \mathbb{R}^d$$

où

- $H : \mathbb{R}^d \times \mathbb{R}^q \to \mathbb{R}^d$,
- $(Y_n)_{n\geq 0}$ est une suite à valeurs dans $\mathbb{R}^q$, $\mathcal{F}_n$-adaptée d' "innovations" des propriétés ν-**moyennisante**

$$\forall f \in \mathcal{V}, \quad \frac{1}{n}\sum_{k=1}^{n} f(Y_k) \xrightarrow[n\to\infty]{(?)} \int_{\mathbb{R}^q} f d\nu,$$

  pour une certaine classe de fonctions $\mathcal{V}$, mais $(Y_n)_{n\geq 0}$ pas nécessairement Markov !
- ΔM_n est un $\mathcal{F}_n$-accroissement de martingale.

Motivations

- **Résoudre des problèmes de Probabilités Numériques**, *i.e.* avec des innovations simulées
  - $(Y_n)_{n\geq 0}$ i.i.d. de loi ν (Algorithme Stochastique (Regulier), Robbins-Monro [5], Duflo [5], etc),
  - $Y_n = \xi_n$, $n \geq 0$, suite à discrépance faible avec $\nu = \lambda_{[0,1]^q}$ et $\mathcal{F}_n = \{\emptyset, \Omega\}$ (Algorithme Quasi-Stochastique, Lapeyre-Pagès-Sab [2]).
- Les **innovations** sont **exogènes** (données de marché, output d'un système *e.g.* un schéma d'Euler) : $(Y_n)_{n\geq 0}$ est une $\mathcal{F}_n$-chaîne de Markov "ν-ergodique" (ou stationnaire)

$$\forall f \in L^p(\nu), \quad \frac{1}{n}\sum_{k=1}^{n} f(Y_k) \xrightarrow[n\to\infty]{} \int_{\mathbb{R}^q} f d\nu \quad \mathbb{P}_\nu\text{-}p.s. \text{ et } L^p(\mathbb{P}) \tag{1}$$

(ou bien même $\mathbb{P}_\mu$-$p.s.$ avec $\mu = \mathbb{P}_{Y_0} \neq \nu$).

Idées

- Hypothèse sur la vitesse dans (1) "$\mathbb{P}$-$p.s.$ & $L^p(\mathbb{P})$",
- Hypothèse (forte...) de Lyapunov sur H,
- Contrôle de $\mathbb{E}\left[(\Delta M_{n+1})^2 \,|\, \mathcal{F}_n\right]$.
- Condition sur la suite de pas $(\gamma_n)_{n\geq 1}$ [ni trop rapide, ni trop lente]

Alors, si $h(\theta) = \displaystyle\int_{\mathbb{R}^q} H(\theta, y)\nu(dy)$ est la *fonction moyenne*

$$\theta_n \underset{n\to\infty}{\longrightarrow} \theta^* = \{h = 0\}.$$

Plan

1. Approximation récursive stochastique
2. Description du cadre et Motivations
3. Théorème de convergence
4. Applications
   - Innovations i.i.d.
   - Fonctionnelle d'un processus α-mélangeant
   - Chaîne de Markov homogène
   - Suites à discrépance faible
5. Exemples
   - Recherche de corrélation implicite

Hypothèses

- **Hypothèse de vitesse de ν-ergodicité.** Soit $p \in [1, \infty)$ et $\varepsilon_n \to 0$. On définit la classe de fonctions suivante

$$\mathcal{V}_{(\varepsilon_n),p} := \left\{ f : \mathbb{R}^q \to \mathbb{R}^q \text{ t.q. } \frac{1}{n} \sum_{k=1}^{n} f(Y_k) - \int_{\mathbb{R}^q} f d\nu \overset{p.s.\&L^p}{=} O(\varepsilon_n) \right\}$$

  On suppose que $H(\theta^*, \cdot) \in \mathcal{V}_{(\varepsilon_n),p}$.

- **Hypothèse de Lyapunov sur H.** Soit $L : \mathbb{R}^d \to \mathbb{R}_+$ $\mathcal{C}^1$ satisfaisant
  - (i) ∇L est Lipschitz continue et $|\nabla L|^2 \le C(1 + L)$,
  - (ii) H satisfait l'hypothèse de *retour à la moyenne*

$$\begin{aligned} &\exists\, U \subset \mathbb{R}^q \text{ un ouvert avec } \nu(U) > 0 \text{ t.q. } \quad \forall \theta \in \mathbb{R}^d,\ \theta \neq \theta^*, \\ &\inf_{y \in U} \langle \nabla L(\theta) \mid H(\theta, y) - H(\theta^*, y) \rangle > 0 \text{ et s.c.i.} \end{aligned} \tag{2}$$

Hypothèses

- **Contrôle de H en l'infini (sous-linéarité).**

$$\forall \theta \in \mathbb{R}^d, \forall y \in \mathbb{R}^q, \quad |H(\theta, y)| \leq C_H \varphi(y)\sqrt{1 + L(\theta)}$$

- **Contrôle de l'accroissement de martingale.** $\forall n \geq 0$,

$$\mathbb{E}\left[|\Delta M_{n+1}|^{2\vee\frac{p}{p-1}} \;|\mathcal{F}_n\right] \leq C_M \varphi(Y_n)^{2\vee\frac{p}{p-1}} \left(1 + L(\theta_n)\right)^{1\vee\frac{p}{2(p-1)}} \;\text{ si } p > 1$$

$$\operatorname{supess} \mathbb{E}\left[|\Delta M_{n+1}| \;|\mathcal{F}_n\right] \leq C_M \varphi(Y_n)\sqrt{1 + L(\theta_n)} \;\text{ si } p = 1$$

avec $\sup_{n\geq 1} \|\varphi(Y_n)\|_{\frac{p}{(p-1)}} < +\infty$.

Résultat principal

- **Conditions sur la suite de pas.**

$$\sum_{n\geq 1}\gamma_n = +\infty,\;\; n\varepsilon_n\gamma_n \longrightarrow 0 \text{ et } \sum_{k\geq 1} k\varepsilon_n \max\left(\gamma_k^2, |\Delta\gamma_{k+1}|\right) < +\infty.$$

Alors,

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

Plan

1 Approximation récursive stochastique

2 Description du cadre et Motivations

3 Théorème de convergence

4 Applications
- Innovations i.i.d.
- Fonctionnelle d'un processus α-mélangeant
- Chaîne de Markov homogène
- Suites à discrépance faible

5 Exemples
- Recherche de corrélation implicite

Cadre i.i.d.

On pose $\forall n \geq 0$, $Y_n = X_{n+1}$ où la suite $(X_n)_{n\geq 1}$ est i.i.d. de loi ν. Alors

$h(\theta) = \mathbb{E}\left[H(\theta, X_1)\right]$ et on pose $\Delta M_{n+1} = H(\theta_n, X_{n+1}) - h(\theta_n)$, $n \geq 0$

On a $\varphi \equiv 1$ donc l'hypothèse de **contrôle de** H devient

$$\forall \theta \in \mathbb{R}^d,\ \forall y \in \mathbb{R}^q, \quad |H(\theta, y)| \leq C_H \sqrt{1 + L(\theta)}.$$

Alors

$$\mathbb{E}\left|\Delta M_{n+1}\right|^2 \leq (1 + L(\theta_n)) \quad \text{et} \quad p = 2.$$

Par la loi du logartihme itéré, la **vitesse de convergence** est alors

$$\varepsilon_n = \sqrt{\frac{\log \log n}{n}}.$$

Fonctionnelle d'un processus α-mélangeant

On définit ici $Y_n = f(\cdots, X_{n-1}, X_n)$ avec $(X_n)_{n\in\mathbb{Z}}$ α-**mélangeant**, *i.e.* $(X_n)_{n\in\mathbb{Z}}$ est stationnaire (de loi ν) et $\alpha_n \underset{n\to\infty}{\longrightarrow} 0$ où $(\alpha_n)_n$ est la suite des coefficients d'α-mélange définis par

$$\alpha_n = \sup\left\{|\mathbb{P}(U \cup V) - \mathbb{P}(U \cap V)|, U \in \mathcal{F}^X_{0,k}, V \in \mathcal{F}^X_{k+n,\infty}, k \geq 1\right\} \quad (3)$$

avec $\mathcal{F}^X_{0,k} = \sigma(X_0, \ldots, X_k)$ et $\mathcal{F}^X_{k+n,\infty} = \sigma(X_{k+n}, \ldots)$.

On suppose aussi que $H(\theta^*, \cdot) \in L^{2+\delta}(\nu), \delta > 0$ et une condtion de sommabilité des coefficients d'α-mélange

$$\sum_{k=1}^{\infty} \frac{1}{\sqrt{k}} \alpha_k^{\frac{\delta}{2(2+\delta)}} < \infty. \quad (4)$$

Vitesse de ν-ergodicité

Proposition

Si $(X_n)_{n\geq 1}$ est stationnaire, α-mélangeant et satisfait (4), alors

$$\forall \delta > 0, \quad L^{2+\delta}(\nu) \subset \mathcal{V}_{\varepsilon_n,2}, \quad \varepsilon_n = n^\beta, \quad \textit{pour tout } \beta \in \left[0, \frac{1}{2}\right). \tag{5}$$

La preuve de cette proposition est basée sur

- l'inégalité de covariance pour les processus α-mélangeant
- le Théorème de Gàl-Koksma.

Chaîne de Markov homogène

Supposons que $(Y_n)_{n\geq 0}$ est une **chaîne de Markov homogène** à valeurs dans $\mathbb{R}^q$ de transition $(P(y,dx))_{y\in\mathbb{R}^d}$. La procédure récursive associée est alors

$$\theta_{n+1} = \theta_n - \gamma_{n+1} K(\theta_n, Y_{n+1}), \quad n \geq 0,$$

où

$$K(\theta, y) := \mathbb{E}\left(K(\theta, Y_1) \mid Y_0 = y\right)$$

$$\Delta M_{n+1} := K(\theta_n, Y_{n+1}) - \mathbb{E}\left(K(\theta_n, Y_{n+1}) \mid \mathcal{F}_n\right) = K(\theta_n, Y_{n+1}) - H(\theta_n, Y_n).$$

L'hypothèse de **contrôle sur** H devient

$$\forall \theta \in \mathbb{R}^d, \quad \forall y \in \mathbb{R}^q, \quad |K(\theta, y)| \leq C_K \widetilde{\phi}(y)\sqrt{1 + L(\theta)}$$

avec $\sup_n \left\|\widetilde{\phi}(Y_n)\right\|_{\frac{p}{p-1}} < +\infty$ et l'hypothèse de Lyapunov (2) sur H.

Suites à discrépance faible

Dans ce cadre, on a $\mathcal{F}_n = \{\emptyset, \Omega\}$, $Y_n = \xi_{n+1}$ où $(\xi_n)_{n\geq 1}$ est une **suite uniformément distribuée sur** $[0,1]^q$, *i.e.*

$$\frac{1}{n}\sum_{k=1}^{n}\delta_{\xi_k} \overset{(\mathbb{R}^q)}{\Longrightarrow} \lambda\big|_{[0,1]^q} \quad \text{quand} \quad n \to \infty.$$

On définit la *discrépance à l'origine* comme suit

$$D_n^*(\xi) := \sup_{y\in[0,1]^q}\left|\frac{1}{n}\sum_{k=1}^{n}\mathbb{1}_{[\![0,y]\!]} - \prod_{k=1}^{n} y^k\right|.$$

Théorème

Il existe des suites $(\xi_n)_{n\geq 1}$ (Halton, Kakutani, Sobol, ...) telles que

$$\max_{1\leq k\leq n} kD_k^* \leq (\log n)^q.$$

Exemple : Suite de Halton

Soient $p_1, \dots, p_q$ les q premiers nombres premiers. La suite de Halton q-dimensionelle est définie, pour tout $n \geq 1$, par

$$\xi_n = \left(\Phi_{p_1}(n), \dots, \Phi_{p_q}(n)\right)$$

où

$$\Phi_p(n) = \sum_{k=0}^{r} \frac{a_k}{p^{k+1}}$$

avec

$$n = a_0 + a_1 p + \dots + a_r p^r, \quad 0 \leq a_i \leq p-1, \quad a_r \neq 0,$$

désigne la décomposition p-adique de n. Alors, pour tout $n \geq 1$,

$$D_n^*(\xi) \leq O\left(\frac{(\log n)^q}{n}\right).$$

Vitesse de convergence

- **Cas à variation finie** Si f est à variation finie dans le sens de Hardy et Krause, alors par Koksma-Hlawka

$$\left|\frac{1}{n}\sum_{k=1}^{n} f(\xi_k) - \int_{[0,1]^q} f\,\lambda_q\right| \leq V_{HK}(f)\frac{D_n^*(\xi)}{n}.$$

Donc $\mathcal{V}_{HK} = \{f : [0,1]^q \to \mathbb{R} \text{ t.q. } V_{HK}(f) < +\infty\} \subset \mathcal{V}_{\epsilon_n}$ avec

$$\varepsilon_n = \frac{(\log n)^q}{n}.$$

- **Cas Lipschitz** Si f est Lipschitz continue, alors par Proinov

$$\left|\frac{1}{n}\sum_{k=1}^{n} f(\xi_k) - \int_{[0,1]^q} f(u)\lambda_q(du)\right| \leq C_d\,[f]_{\text{Lip}}\, D_n^*(\xi)^{\frac{1}{q}}.$$

Plan

1. Approximation récursive stochastique
2. Description du cadre et Motivations
3. Théorème de convergence
4. Applications
   - Innovations i.i.d.
   - Fonctionnelle d'un processus α-mélangeant
   - Chaîne de Markov homogène
   - Suites à discrépance faible
5. Exemples
   - Recherche de corrélation implicite

Description du modèle

On considère un modèle de Black-Scholes 2-dimensionnel *i.e.* $X_t^0 = e^{rt}$ (actif sans risque) et

$$X_t^i = x_0^i e^{(r-\frac{\sigma_i^2}{2})t+\sigma_i W_t^i}, \quad x_0^i > 0, \quad i = 1, 2,$$

pour les deux actifs risqués où $\left\langle W^1, W^2 \right\rangle_t = \rho t$, $\rho \in [-1, 1]$ représente la corrélation entre W_1 et W_2 (qui est la corrélation entre les rendements des actifs risqués X_1 et X_2).

Dans ce marché, on considère une option best-of call caractérisée par le payoff

$$f(X_T^1, X_T^2) = \left(\max\left(X_T^1, X_T^2\right) - K\right)_+ .$$

On utilisera une procédure stochastique récursive pour résoudre le problème inverse en ρ

$$P_{BoC}(x_0^1, x_0^2, K, \sigma_1, \sigma_2, r, \rho, T) = P_0^{market}$$

où

$$P_{BoC}(x_0^1, x_0^2, K, \sigma_1, \sigma_2, r, \rho, T) := e^{-rT}\mathbb{E}\left[f(X_T^1, X_T^2)\right]$$
$$= e^{-rT}\mathbb{E}\left[\left(\max\left(x_0^1 e^{\mu_1 T + \sigma_1\sqrt{T}\zeta_1}, x_0^2 e^{\mu_2 T + \sigma_2\sqrt{T}\left(\rho\zeta_1 + \sqrt{1-\rho^2}\zeta_2\right)}\right) - K\right)_+\right]$$

où $\mu_i = r - \frac{\sigma_i^2}{2}$, $i = 1, 2$, et $\zeta = (\zeta_1, \zeta_2) \stackrel{d}{=} \mathcal{N}(0, I_2)$.

On suppose dorénavant que l'équation

$$P_{BoC}(x_0^1, x_0^2, K, \sigma_1, \sigma_2, r, \rho, T) - P_0^{market} = 0 \quad \text{(en } \rho\text{)}$$

a au moins une solution, notée ρ^*.

Définition de la fonction moyenne

Le moyen le plus commode pour éviter les effets de bords dûs au fait que $\rho \in [-1, 1]$ est d'utiliser une **paramétrisation trigonométrique de la corrélation** en posant

$$\rho = \cos\theta, \quad \theta \in \mathbb{R}.$$

Il est naturel de poser

$$\forall \theta \in \mathbb{R}, \forall \zeta = (\zeta^1, \zeta^2) \in \mathbb{R}^2, \quad H(\theta, \zeta) = P(\theta) - P_0^{market}$$

où $P(\theta) = P_{BoC}(x_0^1, x_0^2, K, \sigma_1, \sigma_2, r, \cos\theta, T)$ et de définir la procédure récursive

$$\theta_{n+1} = \theta_n - \gamma_{n+1} H(\theta_n, \zeta_{n+1}) \quad \text{où} \quad (\zeta_n)_{n\geq 1} \sim \mathcal{N}(0, I_2)$$

On a

- pour tout $\zeta \in \mathbb{R}^2$, $\theta \longmapsto H(\theta, \zeta)$ est **continue** et 2π-**périodique**,
- la **fonction moyenne** h est **continue** et 2π-**périodique**.

Fonction de Lyapunov

On définit la **fonction de Lyapunov** associée à ce problème par

$$L(\theta) = \int_0^\theta gh(u)du + C$$

où $C > 0$ telle que L soit positive et

$$g(\theta) := \begin{cases} \mathbb{1}_{\{h>0\}}(\theta) + \beta_+ \mathbb{1}_{\{h<0\}}(\theta) & si \quad \theta \geq \theta_0 \\ \mathbb{1}_{\{h>0\}}(\theta) + \beta_- \mathbb{1}_{\{h<0\}}(\theta) & si \quad \theta < \theta_0 \end{cases}$$

avec θ_0 solution de $h(\theta) = 0$ et $\beta_\pm$ tel que

$$0 < \beta_+ < \frac{\int_0^{2\pi} h_+(\theta)d\theta}{\int_0^{2\pi} h_-(\theta)d\theta} < \beta_-$$

Alors

- sa dérivée est donnée par $L^{'} = gh$ donc $L^{'}h = gh^2 \geq 0$,
- $\{L^{'}h = 0\} = \{h = 0\}$,
- $L^{'}$ est Lipschitz continue.

Exemple numérique

La suite de nombre quasi-aléatoire gaussien utilisée dans cet exemple est la suivante

$$(\zeta_n^1, \zeta_n^2) = \left(\sqrt{-2\log(\xi_n^1)} \sin(2\pi\xi_n^2), \sqrt{-2\log(\xi_n^1)} \cos(2\pi\xi_n^2) \right),$$

où $\xi_n = (\xi_n^1, \xi_n^2),\ n \geq 1$, est une **suite de Halton bidimensionnelle**.

Les valeurs des paramètres du modèle sont

$$x_0^1 = x_0^2 = 100, \quad r = 0.10, \quad \sigma_1 = \sigma_2 = 0.30, \quad \rho = -0.50$$

et les paramètres du payoff $T = 1$, $K = 100$.
Les paramètres de l'algorithme stochastique sont $\theta_0 = 0$, $n = 10^4$. Enfin on pose

$$\gamma_n = \frac{8}{n}.$$

Résultats graphiques

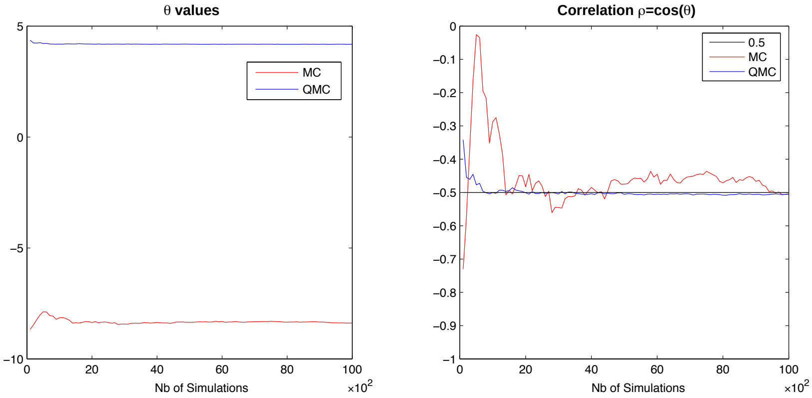

FIGURE: **B-S Best-of-Call option.**
$T = 1,\ r = 0.10,\ \sigma_1 = \sigma_2 = 0.30,\ x_0^1 = x_0^2 = 100,\ K = 100,\ n = 10^4.$
Gauche : convergence de θ_n vers un θ^*.
Droite : convergence de $\rho_n := \cos\theta_n$ vers -0.5.

Bibliographie

- M. Ben Alaya, G. Pagès, *Rate of Convergence for Computing Expectation of Stopping Functionals of an α-mixing Process*, Adv. Appl. Prob. **30**, 425-448 (1998).
- A. Benvéniste, M. Métivier, P. Priouret, *Algorithmes Adaptatifs et Approximations Stochatiques*, Masson, Paris, 1987.
- J. Dedecker, P. Doukhan, G. Lang, J. R. León R., S Louhichi, C. Prieur, *Weak Dependence : with Examples and Applications*, Lecture Notes in Statistics, Springer Science, 2007.
- P. Doukhan, *Mixing*, Lecture Notes in Statistics, Springer, 1995.
- M. Duflo, *Algorihtmes Stochastiques*, Math. and Applications, Springer, 1996.

Bibliographie

- KUSHNER, YIN, *Stochastic approximation and recursive algorithms and applications*, Springer, 2003.
- B. LAPEYRE, G. PAGÈS, K. SAB, *Sequences with Low Discrepancy, Generalisation and Application to Robbins-Monro Algotihm*, statistics 21 (1990) 2, 251-272.
- S. LARUELLE, C.-A. LEHALLE, G. PAGÈS, *Optimal split of orders across liquidity pools : a stochastic algorithm approach*, Pre-pub, 2009.
- G. PAGÈS, *Van der Corput sequences, Kakutani transforms and one-dimensional numerical integration*, J. of Comp. and Appl. Math. 44 (1992) 21-39.
- H. ROBBINS, S. MONRO, *A stochastic approximation method*, Ann. Math. Stat.