# Numerical methods for risk evaluation and control

Winter school : *Frontiers in Stochastic Modeling for Finance*

Nadia Oudjane
EDF Lab
Fime Centre

the 12th of February 2016, Padoue

# Numerical methods for risk evaluation and control

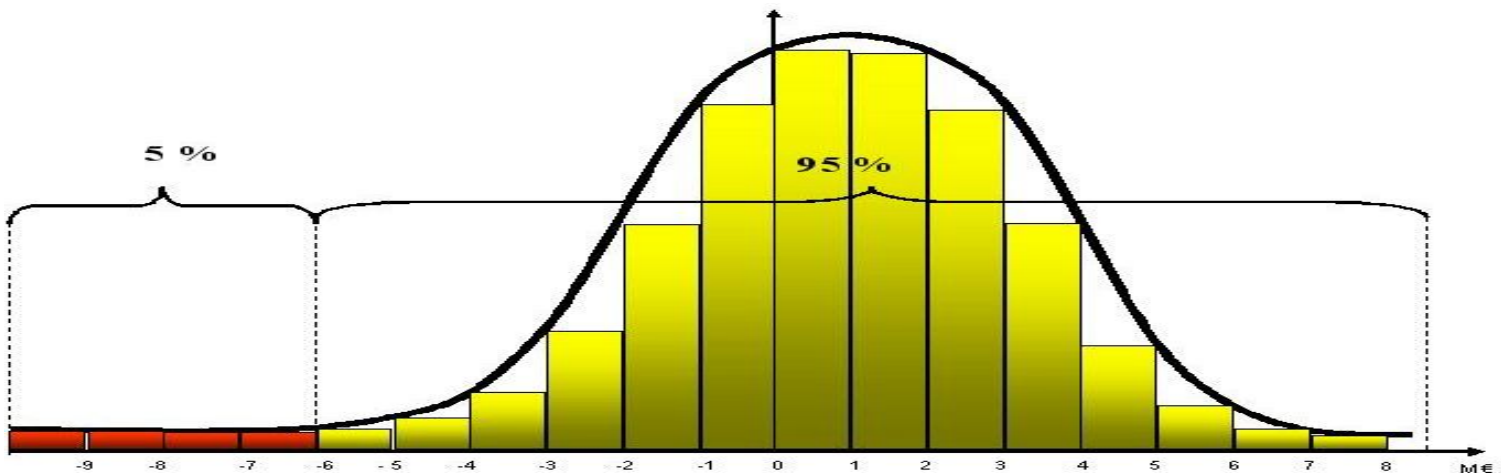
## Evaluating the risk of a portfolio

- We consider the variations of a portfolio value $V_t$ between $t = 0$ and $t = T$

$$R := \Phi(X) = V_T - V_0 + \int_0^T CF \ , \quad \text{with } X \sim \mu_0(dx) = p(x)dx \ ,$$

where $X \in \mathbb{R}^d$ is a random variable (r.v.) modeling the risk factors impacting the portfolio value between 0 and $T$.

- The portfolio value variations, $R$, is a real valued r.v. which can be characterized by its probability distribution.



## Coherent risk measure [ArtznerEtal99]

*This notion was introduced in mathematical finance to define the level of banks capital reserves, then it was extended to other domains. . .*

- Definition

$R$ is a real random variable (r.v) representing the variations of the portfolio value.
A risk measure $\rho : R \mapsto \rho(R) \in \mathbb{R}$ is said to be *coherent* if it verifies

- Positive homogeneity : $\rho(aR) = a\rho(R)$ for any $a \geq 0$
- Non penalization of non risky positions : $R \geq 0$ *a.s.* $\Rightarrow$ $\rho(R) \leq 0$.
- Risk compensation by capital reserve : $\rho(R + a) = \rho(R) - a$
- Sub-additivity : $\rho(R + Y) \leq \rho(R) + \rho(Y)$

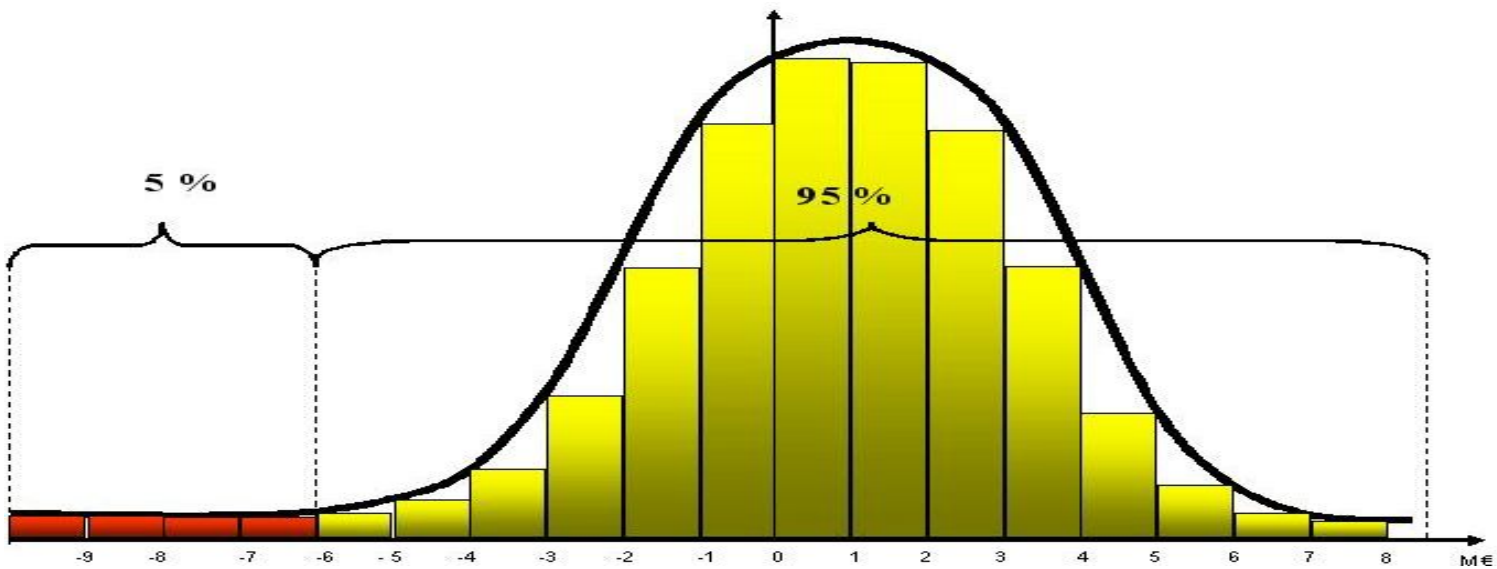
- Difficulty Build a risk measure that both :
  - characterizes properly the owner risk vision (situations to avoid);
  - easy to integrate as a constraint in an optimization problem
    ex : Homogeneity + Sub-additivity $\Rightarrow$ convexity : allowing to formulate convex optimization problems under risk constraints for which efficient resolution techniques have been developed[Bertsekas82, RockafellarWets98].

## Examples of usual risk measures

- Standard deviation is not a coherent risk measure symmetry w.r.t. losses and gains.
- VaR (Value at Risk) of level $(1-\alpha) \in (0,1)$

$$\begin{aligned} VaR_{1-\alpha} &= -\inf\{s \in \mathbb{R} \mid \mathbb{P}(R \leq s) \geq \alpha\} = -\xi_\alpha \quad \alpha\text{-quantile associated to } R \\ &= -F^-(\alpha) \ , \end{aligned}$$

where $F$ is the cumulative distribution of $R$ i.e. $F(s) = \mathbb{P}(R \leq s)$ for all $s \in \mathbb{R}$ .
VaR is not a coherent risk measure since it is not sub-additive.




- CVaR (Conditional Value at Risk) expected losses of the portfolio variations in the $\alpha$-proportion of worst cases is a coherent risk measure.

## Parametric quantile estimation

• Let $(R_1, \cdots, R_N)$ be $N$ i.i.d. realizations according to the cumulative distribution $F$, obtained by simulations or by historical data.

• Goal: To compute an estimation of $\xi_\alpha$, the $\alpha$-quantile associated to $F$, based on the sample $(R_1, \cdots, R_N)$.

• The real r.v. $R$ is supposed to belong to a parametric family parametrized by its mean $\mu$ and standard deviation $\sigma$ with cumulative distribution $F_{\mu,\sigma}$ i.e.

$$F_{\mu,\sigma}(x) := \mathbb{P}(R \leq x) \ , \quad \text{for any} \quad x \in \mathbb{R} \ .$$

• If $F$ is invertible, $\xi_\alpha$ is s.t. $\mathbb{P}(R \leq \xi_\alpha) = F_{\mu,\sigma}(\xi_\alpha) = \alpha$, which gives

$$\mathbb{P}\left(\frac{R-\mu}{\sigma} \leq \frac{\xi_\alpha - \mu}{\sigma}\right) = \alpha \ , \quad \text{i.e.} \quad F_{0,1}\left(\frac{\xi_\alpha - \mu}{\sigma}\right) = \alpha \ , \quad \text{since} \quad \frac{R-\mu}{\sigma} \sim F_{0,1} \ .$$

$\Rightarrow$ It results in estimating the mean and variance $(\mu, \sigma^2)$

$$\xi_\alpha = F_{0,1}^{-1}(\alpha)\sigma + \mu \approx \hat{\xi}_\alpha = F_{0,1}^{-1}(\alpha)\hat{\sigma} + \hat{\mu} \ .$$

where $\hat{\sigma}$ and $\hat{\mu}$ are the empirical standard deviation and mean based on $(R_1, \cdots, R_N)$.

• However this estimation is very sensitive to the choice of the parametric family that characterizes the tails distribution

$$\text{Gaussian case} \quad F_{0,1}^{-1}(5\%) = -1.65 \ , \ \text{bi-exponential case} \quad F_{0,1}^{-1}(5\%) = -2.33$$

## Non parametric quantile estimation

• Let $(R_1, \cdots, R_N)$ be $N$ i.i.d. realizations according to the cumulative distribution $F$, obtained by simulations or by historical data.
• Goal: To compute an estimation of $\xi_\alpha$ based on the sample $(R_1, \cdots, R_N)$

• The order statistics denoted by $(R_{(1)}, \cdots, R_{(N)})$ is obtained by sorting the sample in the increasing order

$$R_{(1)} \leq \cdots \leq R_{(N)}$$

• The empirical quantile, $\hat{\xi}_\alpha$, associated to the sample $(R_1, \cdots, R_N)$ is defined by

$$\hat{\xi}_\alpha := R_{(\{N\alpha\})} \quad \text{where } \{N\alpha\} = \begin{cases} [N\alpha] + 1 & \text{if } [N\alpha] < N\alpha \\ N\alpha & \text{if } N\alpha = [N\alpha] \end{cases}$$

where for any $a$, $[a]$ denotes the integer part of $a \in \mathbb{R}^+$.

## Convergence and confidence interval for the empirical quantile

• Suppose $F$ is continuous and admits a unique solution $\xi_\alpha$ s.t. $F(\xi_\alpha) = \alpha$.
The empirical quantile $\hat{\xi}_\alpha$ converges almost surely (a.s.) to the quantile $\xi_\alpha$ when the size of the sample grows to infinity :

$$|\hat{\xi}_\alpha - \xi_\alpha| \xrightarrow[N\to\infty]{\text{a.s.}} 0 \ .$$

• When the density $f$ of $X$ is continuous with $f(\xi_\alpha) > 0$, $\hat{\xi}_\alpha$ satisfies a Central Limit Theorem

$$\sqrt{N}|\hat{\xi}_\alpha - \xi_\alpha| \xrightarrow[N\to\infty]{\text{Law}} \mathcal{N}(0, \sigma_\alpha) \ , \quad \text{where} \quad \sigma_\alpha = \frac{\sqrt{\alpha(1-\alpha)}}{f(\xi_\alpha)} \ . \tag{1}$$

We can derive the following confidence interval

$$\mathbb{P}(\xi_\alpha \in [\hat{\xi}_\alpha - s \, , \, \hat{\xi}_\alpha + s]) = x \ , \quad \text{with} \quad s = \frac{\sigma_\alpha}{\sqrt{N}} F^{-1}_{\mathcal{N}(0,1)}\left(\frac{x+1}{2}\right) \ .$$

where $F_{\mathcal{N}(0,1)}$ is the cumulative distribution function of the standard and centered Gaussian law.
$\sigma_\alpha$ can be estimated by density estimation techniques [Silverman86, Devroye87]

$$\hat{\sigma}_\alpha = \frac{\sqrt{\alpha(1-\alpha)}}{f_N^h(\hat{\xi}_\alpha)} \quad \text{where} \quad f_N^h \approx f \ . \tag{2}$$

## Non asymptotic confidence interval for the empirical quantile

- We look for a confidence interval of the form $\mathbb{P}(R_{(r)} \leq \xi_\alpha \leq R_{(s)}) = x \in (0,1)$ .
- We define the r.v. $B^\alpha_{1:N} := \sum_{i=1}^N \mathbb{I}_{R_i \leq \xi_\alpha} \sim$ Binomial$(\alpha, N)$ such that ,

$$\mathbb{P}(B^\alpha_{1:N} = k) = C_N^k \alpha^k (1-\alpha)^{N-k} \quad , \quad \text{for any } k \in \{0, 1, \cdots, N\} .$$

- For any $r \leq s$ in $\{1, \cdots, N\}$,

$$\begin{aligned}
\mathbb{P}(R_{(r)} \leq \xi_\alpha \leq R_{(s)}) &= \mathbb{P}(R_{(r)} \leq \xi_\alpha) - \mathbb{P}(R_{(s)} \leq \xi_\alpha) \\
&= \mathbb{P}(B^\alpha_{1:N} \geq r) - \mathbb{P}(B^\alpha_{1:N} \geq s) \\
&= \sum_{k=r}^N \mathbb{P}(B^\alpha_{1:N} = k) - \sum_{k=s}^N \mathbb{P}(B^\alpha_{1:N} = k) \\
&= \sum_{k=r}^{s-1} C_N^k \alpha^k (1-\alpha)^{N-k} .
\end{aligned}$$

=> This is the type of confidence interval which is mostly used in practice

## Idealized Multilevel splitting

- Suppose $X$ is a random vector in $\mathbb{R}^d$ with law $\mu$ that we can simulate
- Goal: To estimate $\alpha := \mathbb{P}(\Phi(X) > s)$ for a threshold $s$ lying in the tails of $\Phi(X)$, where the *score function* $\Phi : \mathbb{R}^d \to \mathbb{R}$ is a *black box*
- The crude Monte Carlo approach uses an i.i.d. sample $(X_1, \cdots, X_N)$

$$\hat{\alpha}^N := \frac{\sum_{i=1}^N \mathbb{I}_{\Phi(X_i)>s}}{N} \quad \text{with a relative error} \quad \frac{Var(\hat{\alpha}^N)}{\alpha^2} = \frac{1-\alpha}{\alpha}\frac{1}{N} .$$

- Consider a sequence of increasing levels $-\infty = L_0 < L_1 < \cdots < L_{n_0} = s$

$$\mathbb{P}(\Phi(X) > s) = \prod_{m=0}^{n_0-1} \mathbb{P}(\Phi(X) > L_{m+1} | \Phi(X) > L_m) .$$

- Idealized Multilevel Splitting
  - Estimate each probability $\alpha_m := \mathbb{P}(\Phi(X) > L_{m+1} | \Phi(X) > L_m)$ separately
  - Choose for all $m$, $\alpha_m = \alpha_0 = \alpha^{\frac{1}{n_0}}$ to minimize the variance.

Consider $\hat{\alpha}_0^N, \cdots, \hat{\alpha}_{n_0-1}^N$ some independent crude Monte Carlo estimators using $N$ runs each

$$\mathbb{E}[\hat{\alpha}_0^N \cdots \hat{\alpha}_{n_0-1}^N] = \alpha , \quad \text{and} \quad \frac{Var(\hat{\alpha}_0^N \cdots \hat{\alpha}_{n_0-1}^N)}{\alpha^2} \approx \frac{1-\alpha_0}{\alpha_0}\frac{n_0}{N} .$$

## Multilevel splitting

- The ideal setting is impractical because
  - it is difficult to *a priori* choose the levels $-\infty = L_0 < L_1 < \cdots < L_{n_0} = s$ so as to obtain for any $m = 0, \cdots, n_0 - 1$

$$\alpha_m := \mathbb{P}(\Phi(X) > L_{m+1} | \Phi(X) > L_m) = \alpha^{\frac{1}{n_0}}$$

  - it is difficult to build independent estimates $(\hat{\alpha}_0^N, \cdots, \hat{\alpha}_{n_0-1}^N)$
- Key points of this approach:
  1. Finding an efficient algorithm to compute the conditional probabilities
  2. Selecting properly the sequence of levels

## Reversible shaking kernels for splitting [GobetEtLiu15]

- Let $E := \{x \in \mathbb{R}^d \text{ s.t. } \Phi(x) > s\}$ and $E_k := \{x \in \mathbb{R}^d \text{ s.t. } \Phi(x) > L_k\}$ recall that

$$\mathbb{P}(X \in E) = \prod_{k=0}^{n_0-1} \mathbb{P}(X \in E_{k+1} | X \in E_k) = \prod_{k=0}^{n_0-1} \alpha_k \ .$$

- Consider the restriction of $\mu$ to $E_k$

$$\mu_k := \mathbb{P}(\cdot | X \in E_k) = \frac{\mathbb{I}_{x \in E_k} \mu}{\int_{x \in E_k}) \mu(dx)} = \frac{\mathbb{I}_{x \in E_k} \mu}{\mu(\mathbb{I}_{x \in E_k})} =: \mathbb{I}_{x \in E_k} \cdot \mu$$

- Find a kernel $Q_k$ which leaves invariant $\mu_k$:

$$\mu_k Q_k := \int_{\mathbb{R}^d} \mu_k(dx) Q_k(x, \cdot) = \mu_k$$

- Use Markov Chain Monte Carlo Methods (MCMC) to approximate $\alpha_k$:
  - Find $X_0^k \in E_k$, then for $i = 1, \cdots, N$ generate $X_i^k \sim Q_k(X_{i-1}^k, \cdot)$
  - Under ergodicity conditions on the transition kernel $Q_k$ for any $\varphi \in C_b(\mathbb{R}^d, \mathbb{R})$

$$\frac{1}{N} \sum_{i=1}^{N} \varphi(X_i^k) \xrightarrow[N \to \infty]{a.s.} \mu_k(\varphi) := \int_{\mathbb{R}^d} \mu_k(dx) \varphi(x) \quad \text{and} \quad \mathbb{E}[|\frac{1}{N} \sum_{i=1}^{N} \varphi(X_i^k) - \mu_k(\varphi)|^2] \leq \frac{C_k(\varphi)}{N} \ .$$

$$\Rightarrow \quad \hat{\alpha}_k^N := \frac{\sum_{i=1}^{N} \mathbb{I}_{X_i^k \in E_{k+1}}}{N} \xrightarrow[N \to \infty]{a.s.} \alpha_k \quad \text{and} \quad \mathbb{E}[|\hat{\alpha}_k^N - \alpha_k|^2] \leq \frac{C_k}{N} \ .$$

## Building a transition kernel that leaves invariant the Gaussian probability measure

- The notion of reversible kernel

Let $\mu$ be a probability measure on $\mathbb{R}^d$, and $Q$ a transition kernel, $Q$ is said to be $\mu$-reversible if

$$\mu(dx)Q(x,dx') = \mu(dx')Q(x',dx) \ , \quad \text{for any} \quad (x,x') \in \mathbb{R}^d \times \mathbb{R}^d \ . \tag{3}$$

- Any $\mu$-reversible kernel $Q$ leaves $\mu$ invariant Integrating w.r.t. $x'$

$$\int_{x' \in \mathbb{R}^d} \mu(dx)Q(x,dx') = \mu(dx) = \int_{x' \in \mathbb{R}^d} \mu(dx')Q(x',dx) = (\mu Q)(dx) \ , \quad \text{for any } x \in \mathbb{R}^d \ .$$

- The following kernel, $Q$, is $\mathcal{N}(0, I_m)$-reversible

$$Q(x,dx') := \frac{1}{\sqrt{2\pi(1-c^2)}} \exp\Big\{ -\frac{1}{2}\frac{(x'-cx)^2}{1-c^2} \Big\} dx' \ , \quad \text{with} \quad |c| < 1 \ . \tag{4}$$

- Indeed consider $W \sim \mathcal{N}(0, I_m)$ independent of $X$ then

$$X' := cX + \sqrt{1-c^2}\, W \ , \quad \text{is such that} \quad \mathbb{P}(X' \in dx' \,|\, X = x) = Q(x,dx') \tag{5}$$

If $X \sim \mathcal{N}(0, I_m)$ $\qquad (X, X') \sim \mathcal{N}\left(0, \begin{bmatrix} I_m & cI_m \\ cI_m & I_m \end{bmatrix}\right)$ is symmetric.

## From a $\mu$-reversible kernel to a kernel leaving $\mu_k$ invariant

• If $Q$ is a $\mu-$reversible kernel then $Q_k$ leaves invariant $\mu_k := \frac{\mathbb{I}_{x\in E_k}\mu}{\mu(\mathbb{I}_{E_k})} =: \mathbb{I}_{x\in E_k}\cdot\mu$,

where

$$Q_k(x,dx') := Q(x,dx')\mathbb{I}_{x'\in E_k} + \big[1 - Q(x,E_k)\big]\delta_x(dx') \ , \quad \text{for any } (x,x')\in E^2 \ .$$

Indeed, by integrating w.r.t. $\mu_k(dx)$ and using the $\mu-$reversibility of $Q$ yields
$\forall\, x'\in E$,

$$\begin{aligned}
&\int_{x\in E}\mu_k(dx)Q_k(x,dx') = \int_{x\in E}\frac{\mu(dx)}{\mu(E_k)}\mathbb{I}_{x\in E_k}Q(x,dx')\mathbb{I}_{x'\in E_k} + \int_{x\in E}\frac{\mu(dx)}{\mu(E_k)}\mathbb{I}_{x\in E_k}Q(x,E_k^c)\delta_x(dx')\\
&= \int_{x\in E}\frac{\mu(dx')}{\mu(E_k)}\mathbb{I}_{x\in E_k}Q(x',dx)\mathbb{I}_{x'\in E_k} + \frac{\mu(dx')}{\mu(E_k)}\mathbb{I}_{x'\in E_k}Q(x',E_k^c) \ ,\\
&= \frac{\mu(dx')}{\mu(E_k)}\mathbb{I}_{x'\in E_k}Q(x',E_k) + \frac{\mu(dx')}{\mu(E_k)}\mathbb{I}_{x'\in E_k}Q(x',E_k^c) = \mu_k(dx') \ .
\end{aligned}$$

• To simulate a r.v. $X'\sim Q_k(x,\cdot)$ notice that $Q_k$ rewrites

$$Q_k(x,dx') = Q(x,E_k)\frac{Q(x,dx')\mathbb{I}_{x'\in E_k}}{Q(x,E_k)} + \big[1-Q(x,E_k)\big]\delta_x(dx') \ , \quad \text{for any } (x,x')\in E^2 \ .$$

which provides the following procedure

1. generate $\tilde{X}'$ suivant $Q(x,\cdot)$ according to (5),
2. set $X' = \tilde{X}'$ if $\tilde{X}'\in E_k$ else set $X' = x$.

## Interacting particle methods [DelMoral04, DelMoralEtal06, CerouEtal06] Principle

- Consider the sequence of probability measures $(\mu_k)_{k=0,\cdots,n_0-1}$ s.t.

$$\mu_k(dx) := \mathbb{P}(X \in dx | X \in E_k) \ ,$$

and transition kernels $Q_k$ that leave $\mu_k$ invariant.

- Define the sequence of *Potential functions* $G_k$ on $\mathbb{R}^d$ s.t.

$$G_k(x) = \mathbb{I}_{x \in E_{k+1}} \ .$$

- $(\mu_k)_{k=0,\cdots,n_0-1}$ satisfies the following dynamics

$$\begin{aligned} \mu_k &= G_{k-1} \cdot \mu_{k-1} := \frac{G_{k-1}\mu_{k-1}}{\mu_{k-1}(G_{k-1})} \\ &= (G_{k-1} \cdot \mu_{k-1})Q_k \ , \quad \text{since} \quad \mu_k Q_k = \mu_k \ . \end{aligned}$$

- $(\mu_k)_{k=0,\cdots n_0}$ can be approximated by $(\mu_k^N)_{k=0,\cdots n_0}$ satisfying the dynamics

$$\left\{ \begin{array}{lcl} \mu_0^N &=& S^N(\mu_0) \\ \mu_k^N &=& S^N\Big((G_{k-1} \cdot \mu_{k-1}^N)Q_k\Big) \ , \end{array} \right.$$

where for any $\pi \in \mathcal{P}(E)$ defined on $\mathbb{R}^d$, $S^N(\pi)$ denotes a random discrete measure

$$S^N(\pi) = \frac{1}{N}\sum_{i=1}^N \delta_{\xi^i} \ , \quad \text{with} \quad (\xi^1, \cdots, \xi^N) \text{ i.i.d. } \sim \ \pi \ .$$

## Interacting particle methods [DelMoral04, DelMoralEtal06, CerouEtal06] Algorithm

- Initialization : Generate independently

$$(\xi_0^1, \cdots, \xi_0^N) \text{ i.i.d. } \sim \mu_0 := \mu \quad \text{then set} \quad \mu_0^N = S^N(\mu_0) := \frac{1}{N} \sum_{i=1}^N \delta_{\xi_0^i}$$

- For $k = 0, \cdots, n_0$
- Weighting : For each particle $i \in \{1, \cdots, N\}$, compute

$$\omega_k^i = \frac{G_k(\xi_k^i)}{\sum_{j=1}^N G_k(\xi_k^j)} \quad \text{then set} \quad \nu_k^N := \sum_{i=1}^N \omega_k^i \, \delta_{\xi_k^i}$$

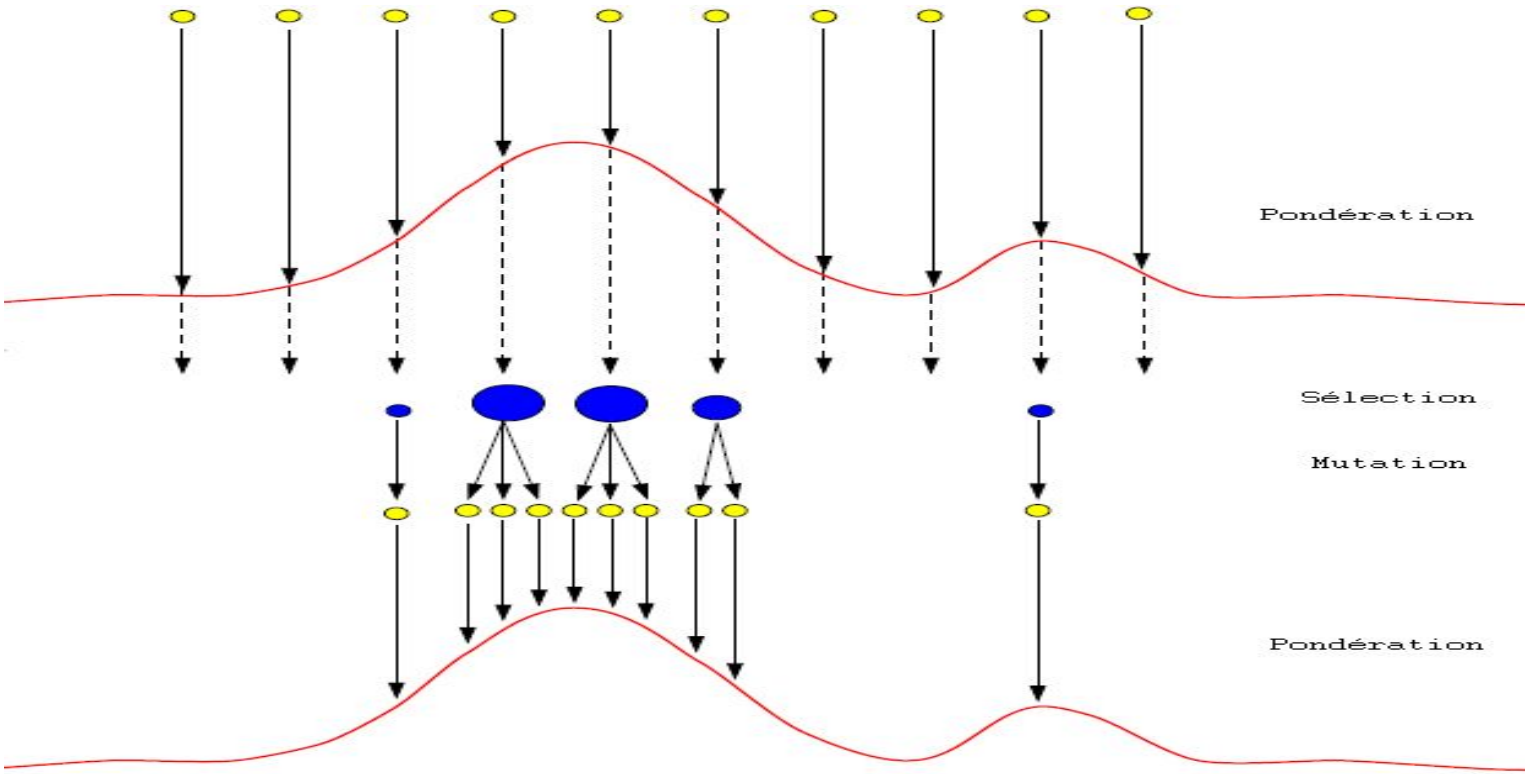
- Selection : Generate independently

$$(\tilde{\xi}_k^1, \cdots, \tilde{\xi}_k^N) \quad \text{i.i.d.} \quad \sim \quad \nu_k^N = \sum_{i=1}^N \omega_k^i \, \delta_{\xi_k^i}$$

- Mutation : Generate independently for each $i \in \{1, \cdots, N\}$,

$$\xi_{k+1}^i \quad \sim \quad Q_{k+1}(\tilde{\xi}_k^i, \cdot) \quad \text{then set} \quad \mu_{k+1}^N = S^N(G_k \cdot \mu_k^N) := \frac{1}{N} \sum_{i=1}^N \xi_{k+1}^i$$

## Interacting Particle Systems



Adaptive choice of the levels $(L_k)_{k\geq 0}$
[MussoEtal01, Homem-de-MelloEtRubinstein02, CerouEtGuyader07, CerouEtal11, GobetEtLiu15, AgarwalEtal15]

• To reduce the variance and get closer to the idealized Multilevel splitting setting where $\alpha_1, \cdots \alpha_{n_0-1} = \alpha_0$, the threshold $L_k$ can be chosen as a r.v. depending on the current particle system

$$L_k = \inf \left\{ L \quad \text{such that} \quad \sum_{i=1}^{N} \mathbb{I}_{\Phi(\xi_k^i)\leq L} \geq (1-\alpha_0 N) \right\}$$

• The algorithm ends in a random number of iterations $K$ as soon as $L_K \geq s$ then we consider the estimate

$$\hat{\alpha} = \alpha_0^{K-1} \frac{\sum_{i=1}^{N} \mathbb{I}_{\Phi(\xi_K^i)\geq s}}{N} .$$

The asymptotic variance is similar to the idealized multilevel splitting approach with a small additional bias of order $1/N$.
• The same algorithm can be used to estimate quantiles by changing the stopping rule.

## Importance Sampling

**Goal** Computing $m = \mathbb{E}_f[\varphi(X)] = \int_{x\in\mathbb{R}^d} \varphi(x)\, f(x)\, dx$ .

- **Change of measure** $f \longrightarrow g$ where $g$ dominates $\varphi f$

$$m = \mathbb{E}_f[\varphi(X)] = \mathbb{E}_g[\varphi(Y)\frac{f}{g}(Y)] \ , \quad \text{where} \quad X \sim f \quad \text{and} \quad Y \sim g$$

- **Monte Carlo approximation** Generate $(Y_1, \cdots, Y_M)$ i.i.d. $\sim g$

$$\hat{m}_M^g = \frac{1}{M}\sum_{i=1}^{M} \varphi(Y_i)\frac{f}{g}(Y_i) \xrightarrow[M\to\infty]{a.s.\ L^2} \mathbb{E}_f[\varphi(X)] \qquad (\text{when} \quad Var\big(\varphi(Y)\frac{f}{g}(Y)\big)) \ .$$

- **Optimal change of measure** $f \longrightarrow g^*$ (zero variance for $\varphi \geq 0$)

$$g^* = \frac{|\varphi| f}{\int |\varphi|(x) f(x)\, dx} = \frac{|\varphi| f}{\mathbb{E}_f[|\varphi|(X)]} = |\varphi| \cdot f$$

$\Rightarrow$ How to simulate and evaluate (at least approximately) $g^*$ ?

## Some approaches to approximate importance distributions

- Large deviation approximation for rare events simulation [Bucklew04]
- Approximation of $\varphi$ to obtain a simple expression for $g^*$ ex : [GlassermanEtal00] for computing VaR, $\Delta$-$\Gamma$ approximation of the portfolio value $\varphi$
- Cross-entropy [Homem-de-MelloEtRubinstein02] $g^\theta$ is chosen in a parametric family such as to minimize the entropy $K(g^\theta, g^*)$
- Mixture of kernels to approximate posterior distributions [West93], [GivensEtRaftery96], [Zhang96]
- Progressive correction [MussoEtal01]
- Review of different approaches [EvansEtSwartz95] and [GivensEtRaftery96]
- ...

## Variance of adaptive Importance Sampling estimate

- Let $g$ be a (possibly random) importance probability density dominating $g^*$

$$Var(\hat{m}_M^g) = \mathbb{E}\big[\, Var[\hat{m}_M^g \,|\, \mathcal{F}_g]\,\big] + \underbrace{Var\big[\,\mathbb{E}[\hat{m}_M^g \,|\, \mathcal{F}_g]\,\big]}_{=0},$$

where $\mathcal{F}_g$ denotes the sigma-algebra generated by $g$

- The variance of the IS estimate is given by the Chi-square distance between $g$ and $g^*$

$$Var(\hat{m}_M^g) = \frac{m^2}{M}\mathbb{E}\left[\int [(g^* - g)\frac{g^*}{g}](x)dx\right].$$

- Idea: use a first set of $N$-simulations to approximate $g^*$ by $g^N$ to achieve variance reduction for $N$ and $M = M(N)$ sufficiently large

$$Var(\hat{m}_M^{g^N}) \leq \frac{C}{MN^\beta} \leq Var(\hat{m}_{M+N}^g) = \frac{C'}{M+N} \qquad \text{with} \quad 0 < \beta < 2/(d+4)$$

## Simulation results: Adaptive Importance Sampling vs Interacting Particle Method

- Several test cases depending on the form of function $x \mapsto \Phi(x)$ have been studied: results are all comparable
- $X$ is a $d$ dimensional Gaussian variable and $m = \mathbb{E}_f[\mathbb{I}_{\Phi(X)\geq s}]$
- Particles $N = 500$; Iterations $n \approx 10$ to $60$; Simulations $M = 10\,000$
- The performance of adaptive importance sampling (using IPS) has been compared to simple Interacting Particle Systems without IS [DelMoralEtal06]

| | $d=1$ | $d=2$ | $d=3$ | $d=4$ | $d=5$ |
|---|---|---|---|---|---|
| $m=10^{-2}$ | 150<br>$10^{-1}$ | 50 | 50 | 30 | 25 |
| $m=10^{-3}$ | 1000<br>2 | 300 | 300 | 200 | 140 |
| $m=10^{-6}$ | $2.10^5$<br>200 | $10^5$<br>400 | $10^5$<br>300 | $5.10^4$<br>460 | $2.10^4$<br>480 |

| | $d=6$ | $d=7$ | $d=8$ | $d=9$ | $\cdots$ | $d=30$ |
|---|---|---|---|---|---|---|
| $m=10^{-2}$ | 22 | 14 | 11 | 8 | $\cdots$ | $5.10^{-3}$ |
| $m=10^{-3}$ | 100 | 70 | 55 | 40 | $\cdots$ | $10^{-3}$ |
| $m=10^{-6}$ | $10^4$<br>250 | $2.10^3$<br>480 | $2.10^3$<br>300 | $4.10^3$<br>300 | $\cdots$ | 1<br>360 |

*Variance ratio of IS-IPS ans simple IPS w.r.t. crude Monte Carlo (indicates a variance reduction when greater than 1)*

# Numerical methods for risk evaluation and control

• Stochastic control problems are standard in energy management

Short-term: unit commitment problem minimizing production costs to satisfy a (stochastic) demand;

Long-term: investment decisions evaluating power plants flexibility (Gas turbines etc.), real options, when to invest ?

Increasing stochasticity due to demand, market prices (electricity, fuels, Co2), production (with the emergence of intermittent energies) ...

$$v_0(x) = \sup_{\nu} \mathbb{E}\left[\sum_{k=0}^{n} f_k(X_k^{\nu_k}, \nu_k)\right] .$$

• Ex: American option pricing is a specific stochastic control problem from financial mathematics, for which a great variety of numerical methods and variance reduction techniques have been proposed

$$u_0(x) = \sup_{\tau \in T_n} \mathbb{E}\left[\sum_{k=0}^{n} f(X_k)\mathbb{I}_{k=\tau}\right] .$$

- Control variate, Quasi Monte Carlo, Antithetic variables, Importance Sampling [LongstaffEtSchwartz01, BroadieEtGlasserman04, Arouna04, Moreni03, JunejaEtKalra09] ...

For a given finite time horizon $T > 0$, consider the stochastic control problem

$$\underbrace{V(t,x)}_{\text{Value function}} := \sup_{\nu \in \mathcal{U}} \underbrace{\mathbb{E}\left[\int_t^T f(s, X_s^{t,x,\nu}, \nu_s) ds + g(X_T^{t,x,\nu})\right]}_{J(t,x,\nu) = \text{Gain function}} \quad \text{for } (t,x) \in [0,T) \times \mathbb{R}^d \,. \tag{6}$$

where

- the stochastic *controlled process* $(X_s^{t,x,\nu})_{t \le s \le T}$ is the solution of the SDE

$$\begin{cases} dX_s^{t,x,\nu} &= b(s, X_s^{t,x,\nu}, \nu_s) ds + \sigma(s, X_s^{t,x,\nu}, \nu_s) dW_s \\ X_t^{t,x,\nu} &= x \,, \end{cases} \quad \text{where} \tag{7}$$

  ▲ the *control process* $\nu \in \mathcal{U}$, the set of all locally bounded predictable processes $\nu = \{\nu_t, t < T\}$ in $\mathbb{L}^2(\Omega \times [0,T))$ taking values in a subset $U \subset \mathbb{R}^k$;
  ▲ the coefficients $b : (t,x,u) \in \mathbb{R}_+ \times \mathbb{R}^d \times U \mapsto b(t,x,u) \in \mathbb{R}^d$ and $\sigma : (t,x,u) \in \mathbb{R}_+ \times \mathbb{R}^d \times U \mapsto \sigma(t,x,u) \in \mathbb{R}^{d \times d'}$ satisfy the usual Lipschitz and linear growth conditions.

- the *running reward function* $f : [0,T) \times \mathbb{R}^d \times U \to \mathbb{R}$ and the *terminal gain function* $g : \mathbb{R}^d \to \mathbb{R}$ satisfy the quadratic growth condition:

$$|f(t,x,u)| + |g(x)| \le C(1 + |x|^2) \,, \tag{8}$$

## Hamilton Jacobi Bellman equation

- Markovian control

There exists a measurable function $u : [0, T) \times \mathbb{R}^d \to U$ s.t. $\nu_t = u(t, X_t)$.

- Infinitesimal generator associated to the EDS (7) $\mathcal{L}^u$ s.t. for any $\varphi \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R})$

$$(\mathcal{L}^u)\varphi(t,x) := b(t,x,u)' D\varphi(x) + \frac{1}{2} Tr[\sigma\sigma'(t,x,u) D^2\varphi(x)] \ , \tag{9}$$

where $D$ and $D^2$ denote the gradient and the Hessian operators.

- Dynamic Programming Principle (DPP) gives that $V$ is a viscosity solution of the nonlinear PDE

$$\frac{\partial V}{\partial t}(t,x) + \sup_{u \in U} (\mathcal{L}^u V(t,x) + f(t,x,u)) = 0 \ , \tag{10}$$

- Consider the function $F$ on $[0, T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{S}^d$

$$F(t,x,\delta,\gamma) := \sup_{u \in U} \left( f(t,x,u) + b(t,x,u)'\delta(t,x) + \frac{1}{2} Tr[\sigma\sigma'(t,x,u)\gamma(t,x)] \right) \ , \tag{11}$$

($\mathcal{S}^d$ denoting the set of symmetric matrices in $\mathbb{R}^{d \times d}$), then the HJB equation can be rewritten as the non linear Partial Differential Equation (PDE):

$$\frac{\partial V}{\partial t}(t,x) + F(t,x,DV,D^2V) = 0 \ . \tag{12}$$

## Several approaches for the numerical solving of HJB equation

- Numerical analysis methods finite differences, finite elements...
  $V$ is interpreted as the viscosity solution of PDE (10). To prove the convergence of the numerical scheme, one can consider monotone approximation schemes in the sense of Barles and Souganidis [BarlesEtSouganidis91]
  (see [BonnansEtZidani03, ForsythEtVetzal12] etc).
- Discrete time and finite state space Dynamic Programming with the Markov chain approximation method [KushnerEtDupuis92].
  This approach is closely related to the finite differences approach since time-space discretization is obtained by renormalizing the finite differences scheme such as to obtain a proper Markov chain.
  The advantage is that the probabilistic setting helps to rely on probabilistic arguments to prove the convergence of the numerical scheme.
- Discrete time Dynamic Programming
  $V$ is interpreted as the solution to the control problem (6). Then one can try to solve a time discrete approximation of this control problem by using the time discrete DPP [BertsekasEtShreve78] and then optimizing backwardly in time the conditional expectation of the value function over the controls.
- When $V$ can be viewed as the solution of a semi-linear PDE, it can be related to the solution of a Forward Backward Stochastic differential Equation [PardouxEtPeng92]

## Finite differences scheme (in dimension 1): space discretization

- Consider the simplified case of a linear (and time homogeneous coefficients $b$ and $\sigma$) PDE

$$\begin{cases} \dfrac{\partial v}{\partial t}(t,x) + \dfrac{1}{2}\sigma^2(x)\dfrac{\partial^2}{\partial x^2}v(t,x) + b(x)\dfrac{\partial}{\partial x}v(t,x) + f(t,x) = 0 \\ v(T,x) = g(x)\ . \end{cases}$$

- Localization in space in $[-L, L]$ find judicious limit conditions
Ex. Dirichlet conditions: $v(t,-L) = \underline{v}(t)$; $v(t,-L) = \bar{v}(t)\ \forall\, t \in [0,T)$.
For simplicity, we suppose here that $\bar{v} = \underline{v} = 0$ allows to control the localization error.

- Space discretization $x_1 = -L < x_1 < \cdots < x_m = L\ ,\quad \text{s.t. } x_{i+1} - x_i = \frac{2L}{m-1} = \delta x\ .$

$$\frac{\partial v}{\partial x}(t,x_i) \approx \frac{v(t,x_{i+1}) - v(t,x_{i-1})}{2\delta x}\ ; \qquad \frac{\partial^2 v}{\partial x^2}(t,x_i) \approx \frac{v(t,x_{i+1}) - 2v(t,x_i) + v(t,x_{i-1})}{(\delta x)^2}\ .$$

- Finite differences approximation For $i = 2, \cdots, m-1$, set

$$\alpha_i = -\frac{\sigma^2(x_i)}{(\delta x)^2}\ ,\quad \beta_i = \frac{\sigma^2(x_i)}{2(\delta x)^2} + \frac{b(x_i)}{2\delta x}\ ,\quad \text{and}\quad \gamma_i = \frac{\sigma^2(x_i)}{2(\delta x)^2} - \frac{b(x_i)}{2\delta x}\ .$$

$$\Rightarrow \quad \frac{\partial v}{\partial t}(t,x_i) + \gamma_i v(t,x_{i-1}) + \alpha_i v(t,x_i) + \beta_i v(t,x_{i+1}) + f(t,x_i) \approx 0\ .$$

## Finite differences scheme in dimension 1: time discretization

- Consider the time mesh $t_1 = 0 < \cdots < t_j < \cdots < t_n = n\delta t = T$ .

$$\frac{v(t_{j+1}, x_i) - v(t_j, x_i)}{\delta t} + \gamma_i v(t_j, x_{i-1}) + \alpha_i v(t_j, x_i) + \beta_i v(t_j, x_{i+1}) + f(t_j, x_i) \approx 0 \ .$$

- Matrix representation We make the approximation $v(t_j, x_i) \approx v_{ij}$ , where $(v_{ij})_{1 \le i \le m, 1 \le j \le n}$ are the elements of a $m \times n$-matrix with column $v_{.j}$ s.t.

$$\begin{cases} v_{in} = g(x_i) \ , \quad \text{for} \quad i = 2, \cdots, m-1 & \text{Terminal condition} \\ v_{1j} = 0 = v_{mj} \ , \quad \text{for } j = 1, \cdots, n & \text{Localization condition} \\ \Big(I - \theta\delta t A\Big)\tilde{v}_{.j} = \Big(I + (1-\theta)\delta t A\Big)\tilde{v}_{.j+1} + \delta t \tilde{f}_{.j} \ , & \text{for } j = n-1, \cdots, 1 \ . \end{cases}$$

- $\tilde{v}_{.j}$ is the $(m-2)$ vector $\tilde{v}_{.j} = (v_{ij})_{i=2,\cdots,m-1}$;
- $\tilde{f}_{.j} = (f(t_j, x_i))_{i=2,\cdots,m-1}$;
- Parameter $\theta \in [0, 1]$;
- $A$ $(m-2) \times (m-2)$ tridiagonal matrix s.t.

$$\begin{bmatrix} \alpha_2 & \beta_2 & 0 & \cdots & \cdots & 0 \\ \gamma_3 & \alpha_3 & \beta_3 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \gamma_{m-2} & \alpha_{m-2} & \beta_{m-2} \\ 0 & \cdots & \cdots & 0 & \gamma_{m-1} & \alpha_{m-1} \end{bmatrix}$$

## Finite differences scheme in dimension 1: convergence conditions [Ciarlet88]

$$\begin{cases} v_{in} = g(x_i) \ , \quad \text{for} \quad i = 2, \cdots, m-1 & \text{Terminal condition} \\ v_{1j} = 0 = v_{mj} \ , \quad \text{for } j = 1, \cdots, n & \text{Localization condition} \\ \Big(I - \theta \delta t A\Big) \tilde{v}_{\cdot j} = \Big(I + (1-\theta)\delta t A\Big) \tilde{v}_{\cdot j+1} + \delta t \tilde{f}_{\cdot j} \ , & \text{for } j = n-1, \cdots, 1 \ , \end{cases}$$

Under regularity assumptions on $b$ $\sigma$ and $f$, $v_{ij}$ allows to build a stepwise (on the mesh $(x_i)_{i=1,\cdots,m}$ convergent approximation of $v$.

- $0 \leq \theta < 1/2$ convergence requires
  - $\delta t \to 0$,
  - $\delta x \to 0$
  - $(1-\theta)\sigma(x_i)\frac{\delta t}{(\delta x)^2} < \frac{1}{2}$ for any $x_i$ in the mesh (CFL condition).

In particular,

- $\theta = 0$ is called *Explicit scheme*
- $\theta = 1/2$ is called *Crank-Nicholson scheme* (with a time discretization error of order 2 instead of 1 when $A$ is time homogeneous)
- $\theta = 1$ is called *fully implicit scheme*

## Finite differences scheme in dimension 1: HJB case

• Non-linear matrix equation $v(t_j, x_i) \approx v_{ij}$ , where $(v_{ij})_{1\leq i\leq m, 1\leq j\leq n}$ are the elements of a $m \times n$-matrix with column $v_{\cdot j}$ s.t.

$$\begin{cases} v_{in} = g(x_i) \ , & \text{for} \quad i = 2, \cdots, m-1 & \text{Terminal condition} \\ v_{1j} = 0 = v_{mj} \ , & \text{for } j = 1, \cdots, n & \text{Localization condition} \\ \tilde{v}_{\cdot j} = T_{\tilde{v}_{\cdot j+1}}(\tilde{v}_{\cdot j}) \ , & \text{for } j = n-1, \cdots, 1 \ , & \text{where} \quad \tilde{v}_{\cdot j} = (v_{ij})_{i=2,\cdots,m-1} \end{cases}$$

$T$ is defined for any $(v, w) \in \mathbb{R}^{m-2} \times \mathbb{R}^{m-2}$ by

$$\begin{aligned} T_w(v) &= \max_u T_w^u(v) \ , \quad \text{(supposed to exist)} \\ T_w^u(v) &= \theta \delta t A^u v + \left(I + (1-\theta)\delta t A^u\right) w + \delta t \tilde{f}_{\cdot j}^u \ , \end{aligned}$$

where the maximum is taken over functions $u : x_i \in \{x_2, \cdots, x_{m-1}\} \mapsto u(x_i) \in U$

One can prove that $T_w^u$ and $T_w$ are contractions (for $\delta t$ sufficiently small) so that one can rely on fixed point algorithms to solve $v = T_w(v)$

• Value iteration Choose $v^0$ (ex: $v_0 = \tilde{v}_{\cdot j+1}$) then for $\ell = 0, \cdots, L \quad v^{\ell+1} = T_w(v^\ell)$ .

• Policy iteration or Howard method Choose $u_0$ then for $\ell = 0, \cdots, L$

$$\begin{cases} v^\ell &= T_w^{u^\ell}(v^\ell) \\ u^{\ell+1} &= \operatorname{argmax}_{u\in U} T_w^u(v^\ell) \ . \end{cases}$$

## Discretizing the control problem in time

- Let $t_0 = 0, \cdots, t_n = T = n\delta t$ be a regular time discretization of $[0, T]$.

For a given sequence of controls $u_{k:p-1} := (u_k, \cdots u_{p-1}) \in \mathbb{U}^{p-k}$ ($\mathbb{U}$ being the set of $U$-valued measurable functions defined on $\mathbb{R}^d$) we consider the Euler scheme of $X$ i.e

$$\begin{cases} \bar{X}_{p+1}^{k,x,u_{k:p}} = \bar{X}_p^{k,x,u_{k:p-1}} + b_p \delta t + \sigma_p \sqrt{\delta t} \varepsilon_{p+1} & \text{for all } p \geq k, \\ \bar{X}_t^{t_k,x,u_{k:n-1}} = x \ , \end{cases} \quad \text{with}$$

$(\varepsilon_{p+1})_{p \geq k}$ being a sequence of i.i.d. $\sim \mathcal{N}(0, I_d)$ and

$$b_p = b\left(t_p, \bar{X}_{p-1}^{k,x,u_{k:p-1}}, u_p(\bar{X}_{p-1}^{k,x,u_{k:p-1}})\right) \quad \text{and} \quad \sigma_p = \sigma\left(t_p, \bar{X}_{p-1}^{k,x,u_{k:p-1}}, u_p(\bar{X}_{p-1}^{t_k,x,u_{k:p-1}})\right) .$$

- We consider the time discretized gain function $J_k : \mathbb{R}^d \times \mathbb{U}^{n-k} \to \mathbb{R}$ defined by

$$J_k(x, u_{k:n-1}) := \mathbb{E}\left[ g(\bar{X}_T^{k,x,u_{k:n-1}}) + \delta t \sum_{p=k}^{n-1} f(t_p, \bar{X}_p^{k,x,u_{k:p-1}}) \right] .$$

## Discrete time Dynamic Programming Principle

- Time disctretization of the value function $V_k$ For any $k \in \{0, \cdots, n\}$,

$$V_k(x) := \sup_{u_{k:n-1} \in \mathbb{U}^{n-k}} J_k(x, u_{k:n-1}) \qquad \text{for } x \in \mathbb{R}^d \ , \tag{13}$$

- By the discrete DPP, the value function, $V_k$, follows a backward recursion

$$\begin{cases} V_n(x) = g(x) \\ V_k(x) = f_k(x) + \sup_{u_k \in \mathbb{U}} \mathbb{E}[V_{k+1}(\bar{X}_{k+1}^{k,x,u_k})] \ , \quad \text{for any } k = 0, \cdots n-1 \ , \end{cases} \tag{14}$$

where $f_k(x) := f(t_k, x)$.

- Estimating the conditional expectations could be approximated by numerical methods such as PDE, Fourier, lattice or Monte Carlo methods,... Many numerical approximation schemes that have been developed for computing American option [BroadieEtGlasserman04, LongstaffEtSchwartz01, BallyEtal05, TsitsiklisEtVanRoy01, DelMoralEtal11, BouchardEtWarin12] prices or BSDE solutions [BouchardEtTouzi04, Zhang04, GobetEtal05] are then available.

- Maximizing the expectation over the controls at time step $k$ can be done via a parametrization of the control $x \mapsto u_k^\theta(x)$ so that parametric optimization methods such as the stochastic gradient algorithm could be applied to maximize the expectation over the parameter $\theta$ [KushnerEtYin97].

## Maximizing the expectation over the controls Back to the continuous time formulation

- Assume that for any $(t,x) \in [0,T) \times \mathbb{R}^d$, the following *argmin* reduces to one point in $U$ then on can define the mapping $G$ on $[0,T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{S}^d$ such that

$$G(t,x,\delta,\gamma) := \underset{u \in U}{\operatorname{argmin}} \left\{ f(t,x,u) + b(t,x,u)'\delta(t,x) + \frac{1}{2} Tr[\sigma\sigma'(t,x,u)\gamma(t,x)] \right\} ,$$

then, the HJB equation becomes

$$\begin{cases} \frac{\partial V}{\partial t}(t,x) + \mathcal{L}^u V(t,x) + f(t,x,u) = 0 & \text{Expectation} \\ u(t,x) = G(t,x,DV,D^2V) & \text{Maximization} , \end{cases} \tag{15}$$

- Time discretization *explicit scheme*

$$\begin{cases} V_k(x) &= f_k(x) + \mathbb{E}[V_{k+1}(\bar{X}_{k+1}^{k,x,u_k})] & \text{Expectation} \\ DV_k(x) &= f_k(x) + D\mathbb{E}[V_{k+1}(\bar{X}_{k+1}^{k,x,u_k})] & \text{Expectation derivative} \\ D^2V_k(x) &= f_k(x) + D^2\mathbb{E}[V_{k+1}(\bar{X}_{k+1}^{k,x,u_k})] & \text{Expectation derivative} \\ u_{k-1}(t,x) &= G(t,x,DV_k,D^2V_k) & \text{Simple evaluation} , \end{cases}$$

This type of approach was developed in [FahimEtal1], where $V$ is interpreted as a viscosity solution of (10) and in [Tan12], where $V$ is interpreted as the solution to the control problem (6).

Computing Derivatives of expectation can be achieved by

• Without any regularity assumption on $\varphi$ (likelihood ratio) Let $\varphi$ be a real valued function with sub-exponential growth then for any $t > 0$ and $x \in \mathbb{R}^d$

$$D\mathbb{E}[\varphi(x+W_t)] = \mathbb{E}[\varphi(x+W_t)\frac{W_t}{t}] \; ; \quad \text{and} \quad D^2\mathbb{E}[\varphi(x+W_t)] = \mathbb{E}[\varphi(x+W_t)\frac{W_t W_t^T - t\mathbb{I}_d}{t^2}] \; .$$

• Under regularity assumptions on $\varphi$

$$D\mathbb{E}[\varphi(x+W_t)] = \mathbb{E}[D\varphi(x+W_t)] \; ; \quad \text{and} \quad D^2\mathbb{E}[\varphi(x+W_t)] = \mathbb{E}[D^2\varphi(x+W_t)] \; .$$

• Those formula can be generalized to the case where $x + W_t$ is replaced by a diffusion process $X_t^{0,x}$ under suitable regularity assumptions on the SDE coefficients using

- Malliavin calculus (Elworthy's formula) [FournieEtal99];
- Tangent process approach [Kunita82] to compute $\nabla_x X_t^{0,x}$.

## FBSDE and semi-linear PDEs [PardouxEtPeng92, ElKarouiEtMazliak97]

• Definition Let $(W_t)_{0\leq t\leq T}$ be a $q$-dimensional Brownian.
Under Lipschitz and growth assumptions on the *driver* $f : [0,T]\times\mathbb{R}^d\times\mathbb{R}\times\mathbb{R}^q\to\mathbb{R}$ the terminal condition $g : \mathbb{R}^d\to\mathbb{R}$ and the coefficients $b : [0,T]\times\mathbb{R}^d\to\mathbb{R}^d$ and $\sigma : [0,T]\times\mathbb{R}^d\to\mathbb{R}^d\times\mathbb{R}^q$,

there exists a unique $\mathcal{F}_t^W$-adapted triplet on $[0,T]$, $(X,Y,Z)$, such that

$$\begin{cases} dX_t = b(t,X_t)dt + \sigma(t,X_t)dW_t\ , \quad \text{with} \quad X_0 = x \\ dY_t = -f(t,X_t,Y_t,Z_t)dt + Z_t dW_t\ , \quad \text{with} \quad Y_T = g(X_T)\ . \end{cases}$$

Moreover $Y_t = v(t,X_t)$.

• Link with PDEs By Ito's Lemma

$$\begin{aligned} Y_s - Y_t = v(s,X_s) - v(t,X_t) &= \int_t^s (\frac{\partial v}{\partial t} + \mathcal{L}v)(r,X_r)dr + \int_t^s Dv(r,X_r)\cdot\sigma(r,X_r)dW_r \\ &= \int_t^s -f(r,X_r,Y_r,Z_r)dr + \int_t^s Z_r dW_r\ . \end{aligned}$$

$$\Rightarrow \quad v \quad \text{satisfies} \quad \begin{cases} v(T,x) = g(x) \\ -\dfrac{\partial v}{\partial t} - \mathcal{L}v - f\big(t,x,v(t,x),Dv(t,x)\sigma(t,x)\big) = 0 \end{cases}$$

• Under some conditions, the semiliear PDE has a unique solution $v^*$ which can be represented as a FBSDE s.t. $Y_t = v^*(t,X_t)$ and $Z_t = Dv^*(t,X_t)\sigma(t,X_t)$

## Numerical schemes for FBSDE [BouchardEtTouzi04, Zhang04, GobetEtal05]

- Let $t_0 = 0, \cdots, t_n = T$ be a regular time discretization of $[0, T]$, s.t. $t_k = k\delta t$.

For any $k \in \{0, \cdots, n\}$,

$$Y_{t_{k-1}} = Y_{t_k} + \int_{t_{k-1}}^{t_k} f(t, X_t, Y_t, Z_t)dt - \int_{t_{k-1}}^{t_k} Z_t \cdot \sigma(t, X_t)dW_t \ , \quad \text{with} \quad Y_{t_n} = g(X_{t_n}) \ .$$

- Taking the conditional expectation w.r.t. $\mathcal{F}_{t_{k-1}}$ yields

$$Y_{t_{k-1}} = \mathbb{E}[Y_{t_k}|X_{t_{k-1}}] + \int_{t_{k-1}}^{t_k} \mathbb{E}[f(t, X_t, Y_t, Z_t)|X_{t_{k-1}}]dt \ , \quad \text{with} \quad Y_{t_n} = g(X_{t_n}) \ . \tag{16}$$

- Most numerical schemes consists of two steps
  1. Simulate (approximately) trajectories of the forward Markov chain $(X_{t_k})_{k\leq 0}$, typically using an Euler scheme $(\hat{X}_k)_{k\leq 0}$
  2. Approximate the backward dynamics (16), by discretizing the integral w.r.t. time

$$\left\{ \begin{array}{lcl} Y_n & = & g(X_n) \\ Y_{k-1} & = & \mathbb{E}[Y_k|X_{k-1}] + f(X_{k-1}, Y_{k-1}, Z_{k-1})\delta t \\ Z_{k-1} & = & \frac{1}{\delta t}(\sigma(t_{k-1}, X_{k-1}))^{-1}\mathbb{E}[Y_k(W_{t_k} - W_{t_{k-1}})|X_{k-1}] \ . \end{array} \right.$$

=> Again one has to rely on numerical approximations of conditional expectations.

## Several approach have been proposed

$(X, Y)$ a pair of square integrable r.v. on $\mathbb{R}^d \times \mathbb{R}$ observed via an $N$-sample

$$(X_i, Y_i)_{i=1,\cdots N} \quad \text{i.i.d.} \sim (X, Y)$$

- Likelihood ratio
- Kernel regression
- Least square regression
- Quantization method ...

## Likelihood ratio

$(X, Y)$ square integrable r.v. on $\mathbb{R}^d \times \mathbb{R}$ observed via an i.i.d. sample $(X_i, Y_i)_{i=1,\dots N}$

- Goal : Approximate the conditional expectation $\mathbb{E}(Y|X)$ on the base of the $N$-sample

$$\mathbb{E}(Y|X) = v^*(X) \ , \quad \text{where} \quad v^*(x) := \int_{y\in\mathbb{R}} y p_{Y|X=x}(y) dy \ .$$

where $p_{Y|X=x}(y)$ is the density of $Y$ conditioned on $X = x$ (assumed to exist).

- Assume that $p_{Y|X=x}(y)$ can be evaluated at any point $(x, y) \in \mathbb{R}^d \times \mathbb{R}$).
- Approximation of the conditional expectation

$$v^*(x) \approx \frac{1}{N} \sum_{i=1}^{N} Y_i p_{Y|X=X_i}(Y_i) \ .$$

## Kernel regression [GyorfiEtal02]

$(X, Y)$ square integrable r.v. on $\mathbb{R}^d \times \mathbb{R}$ observed via an i.i.d. sample $(X_i, Y_i)_{i=1,\cdots N}$

- Goal : Approximate the conditional expectation $\mathbb{E}(Y|X)$ on the base of the $N$-sample

$$\mathbb{E}(Y|X) = v^*(X) \ , \quad \text{where} \quad v^*(x) := \int_{y\in\mathbb{R}} y p_{Y|X=x}(y) dy \ .$$

where $p_{Y|X=x}(y)$ is the density of $Y$ conditioned on $X = x$ (assumed to exist).

- Bayes Formula $p_{Y|X=x}(y)dy = \dfrac{p_{X,Y}(x, dy)}{p_X(x)}$
- Kernel density estimation

$$p_X(x) \approx \frac{1}{Nh^d} \sum_{i=1}^{N} K(\frac{x - X_i}{h}) \ , \quad \text{and} \quad p_{X,Y}(x, dy) \approx \frac{1}{Nh^d} \sum_{i=1}^{N} K(\frac{x - X_i}{h}) \delta_{Y_i}(dy)$$

▲ $K$ is *regular* standard and centered probability density (e.g. $\mathcal{N}(0, I)$).

▲ $h > 0$ window bandwidth chosen s.t. $h \xrightarrow[N\to\infty]{} 0$ and $Nh^d \xrightarrow[N\to\infty]{} \infty$.

- Approximation of the conditional expectation

$$v^*(x) \approx \frac{\sum_{i=1}^{N} K(\frac{x-X_i}{h}) Y_i}{\sum_{i=1}^{N} K(\frac{x-X_i}{h})} \ .$$

## Least-square regression (i) [GyorfiEtal02]

$(X, Y)$ a pair of square integrable r.v. on $\mathbb{R}^d \times \mathbb{R}$ observed via an $N$-sample

$$(X_i, Y_i)_{i=1,\cdots N} \quad \text{i.i.d.} \sim (X, Y)$$

- Goal : Approximate the conditional expectation $\mathbb{E}(Y|X)$ on the base of the $N$-sample

$$\mathbb{E}(Y|X) = v^*(X) , \quad \text{where} \quad v^* := \arg \min_{v \in L^2(X)} \mathbb{E}\left[\left(Y - v(X)\right)^2\right] .$$

$v^*$ can be represented as a weighted sum of basis functions $v^* = \sum_{p=1}^{\infty} \alpha_p^* H_p$ with

$$(\alpha_p^*)_{p\geq 1} := \arg \min_{\alpha_p \in \mathbb{R}} \mathbb{E}\left[\left(\sum_{p=1}^{\infty} \alpha_p H_p(X) - Y\right)^2\right] .$$

where $H_p : \mathbb{R}^d \to \mathbb{R}$ are basis functions of $L^2(X)$.

- Truncation of the basis of $L^2(X)$ $v^* \approx \hat{v}^m := \sum_{p=1}^{m} \hat{\alpha}_p^m H_p$ with

$$(\hat{\alpha}_1^m, \cdots \hat{\alpha}_m^m) := \arg \min_{(\alpha_1,\cdots,\alpha_m)\in\mathbb{R}^m} \mathbb{E}\left[\left(\sum_{p=1}^{m} \alpha_p H_p(X) - Y\right)^2\right]$$

## Least-square regression (ii)

- Monte Carlo approximation of the error $\hat{v} \approx \hat{v}^{m,M} = \sum_{p=1}^{m} \hat{\alpha}_p^{m,M} H_p$ with

$$\hat{\alpha}^{m,M} := (\hat{\alpha}_1^{m,M}, \cdots, \hat{\alpha}_p^{m,M}) := \arg \min_{\alpha \in \mathbb{R}^m} \frac{1}{M} \sum_{i=1}^{M} \left( \sum_{p=1}^{m} \alpha_p H_p(X_i) - Y_i \right)^2 .$$

- Estimation of the coefficients vector $\hat{\alpha}^{m,M} := (\hat{\alpha}_1^{m,M}, \cdots, \hat{\alpha}_m^{m,M})$

$$\hat{\alpha}^{m,M} = \arg \min_{\alpha \in \mathbb{R}^m} \|\mathcal{H}\alpha - \mathcal{Y}\|_2$$

where

$$\mathcal{H} := \begin{bmatrix} H_1(X_1) & \cdots & H_p(X_1) & \cdots & H_m(X_1) \\ \cdot & & & & \\ \cdot & & & & \\ H_1(X_M) & \cdots & H_p(X_1) & \cdots & H_m(X_M) \end{bmatrix}, \quad \alpha := \begin{bmatrix} \alpha_1 \\ \cdot \\ \cdot \\ \alpha_m \end{bmatrix}, \quad \mathcal{Y} := \begin{bmatrix} Y_1 \\ \cdot \\ \cdot \\ Y_M \end{bmatrix}$$

- Approximation of the conditional expectation

$$\widehat{v}^{p,M} := \sum_{p=1}^{m} \hat{\alpha}_p^{m,M} H_p \xrightarrow[m,M \to \infty]{L^2} v$$

# References I

[Silverman86] Silverman, B. W., *Density Estimation for Statistics and Data Analysis*, Monographs on Statistics and Applied Probability, Chapman & Hall, Vol 26, London, 1986.

[Devroye87] Devroye, L., *A Course on Density Estimation*, Birkhäuser, vol 14, Progress in Probability and Statistics, Boston 1987.

[ArtznerEtal99] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath, (1999), *Coherent Measures of Risk*. Mathematical Finance, 9: 203-228.

[Bertsekas82] D. Bertsekas *Constrained Optimization and Lagrange Multiplier methods.* 1982

[RockafellarWets98] Rockafellar, R.T. and Wets, R.J.-B. Variational analysis. Springer, Berlin, (1998).

[DelMoralEtal06] P. Del Moral, A. Doucet, and A. Jasra. *Sequential Monte Carlo samplers*, J. R. Stat. Soc. Ser. B Stat. Methodol., 68(3):411-436, 2006.

# References II

[MussoEtal01] Musso, C. and Oudjane, N. and Le Gland, F. , *Improving regularized particle filters*, in Sequential Monte Carlo Methods in Practice, A. Doucet N. de Freitas and N. Gordon editors, Statistics for Engineering and Information Science, 2001.

[CerouEtal06] Cérou, F. and Del Moral, P. and Le Gland, F. and Guyader, P. and Lezaud, H. Topart, *Some recent improvements to importance splitting*, Proceedings of the 6th International Workshop on Rare Event Simulation, Bamberg, October 9-10, 2006.

[CerouEtGuyader07] F. Cérou and A. Guyader, *Adaptive multilevel splitting for rare event analysis*. Stoch. Anal. Appl., 25(2):417-443, 2007.

[CerouEtal11] F. Cérou, P. Del Moral, T. Furon et A. Guyader, *Sequential Monte Carlo for Rare Event Estimation*, Statistics and Computing, Vol. 22 (3), 795-808, 2012.

[GobetEtLiu15] E. Gobet and G. Liu *Rare event simulation using reversible shaking transformations.* To appear in SIAM Journal on Scientific Computing, 2015.

# References III

[AgarwalEtal15] E. Gobet, G. Liu *Rare event simulation related to financial risks.* preprint, 2015.

[GlassermanEtal00] Glasserman, P. and Heidelberger, P. and Shahabuddin, P. *Reduction Variance techniques for estimating Value at Risk*, *Management Science, Vol. 46, No 10, 2000*.

[Homem-de-MelloEtRubinstein02] Homem-de-Mello, T. and Rubinstein, R.Y. *Estimation of rare event probabilities using cross-entropy*, Proceedings of the Winter Simulation Conference, 2002.

[Bucklew04] Bucklew, J. Introduction to rare event simulation, *Springer series in statistics*, Spr. Ver., 2004.

[West93] *Approximating posterior distributions by mixtures*, Journal of the Royal Statistical Society (Ser. B) 54:553-568, 1993.

[Zhang96] Zhang, P. *Nonparametric importance sampling*. Journal of the American Statistical Association, 91(434):1245-1253, September 1996.

# References IV

[EvansEtSwartz95] Evans M. and Swartz T. *Methods for Approximating Integrals in Statistics with Special Emphasis on Bayesian Integration Problems*. Stastical Science, Vol. 10 No. 3 254-272, 1995.

[GivensEtRaftery96] Givens, G.H. and Rfatery A.E. *Local Adaptive Importance Sampling for Multivariate Densities With Strong Nonlinear Relationships*. Journal of the american Staistical Association, Vol. 91, No433, 132-141, 1996.

[Arouna04] B. Arouna. *Adaptive Monte Carlo method, a variance reduction technique*. Monte Carlo Methods Appl., 10(1):1-24, 2004.

[BroadieEtGlasserman04] M. Broadie and P. Glasserman, A Stochastic Mesh Method for Pricing High-Dimensional American Options, *Journal of Comput. Finance*, vol. 7, 2004.

[DelMoral04] P. Del Moral, Feynman-Kac formula. Genealogical and interacting particle systems with applications, *Probability and its Applications*, Spr. Ver., 2004.

# References V

[DelMoralEtal11] P. Del Moral, P. Hu, N. Oudjane, B. Rémillard, *On the Robustness of the Snell Envelope*. SIFIN, 2011.

[DelMoralEta10] P. Del Moral, P. HU and N. Oudjane, *Snell envelope with small probability criteria*, preprint INRIA-00507794, 2010.

[JunejaEtKalra09] S. Juneja and H. Kalra, *Variance Reduction Techniques for Pricing American Options*, Journal of Computational Finance, 12 (3), 79-102, 2009.

[LongstaffEtSchwartz01] F. Longstaff and E. Schwartz, *Valuing American options by simulation: a simple least-square approach*, The Review of Financial Studies, 2001.

[Moreni03] N. Moreni, *Pricing American options : a variance reduction technique for the longstaff-schwartz algorithm*, Technical Report, CERMICS-ENPC, 2003.

[BarlesEtSouganidis91] G. Barles and P.E. Souganidis, *Convergence of approximation schemes for fully nonlinear second order equations*, Asymptotic Anal., 4(3):271-283, 1991.

# References VI

[BonnansEtZidani03] J.F. Bonnans, H. Zidani, *Consistency of Generalized Finite Difference Schemes for the Stochastic HJB Equation*. SIAM J. Numerical Analysis 41(3), 1008-1021. 2003.

[FahimEtal1] A. Fahim, N. Touzi, and X. Warin, *A probabilitic numerical method for fully nonlinear parabolic PDEs*, Ann. Appl. Probab., 21(4):1322-1364, 2011.

[ForsythEtVetzal12] P. A. Forsyth and K. R. Vetzal *Numerical Methods for Nonlinear PDEs in Finance* Chapter 22, pages 503-528 in Handbook of Computational Finance, Edited by J.C. Duan, J. Gentle, W. Hardle, Springer, 2012.

[Tan12] X. Tan, *Probabilistic numerical approximation for stochastic control problems*, preprint.

[KushnerEtDupuis92] H.J. Kushner and P. Dupuis, *Numerical methods for stochastic control problems in continuous time*, Vol 24 of Applications of Mathematics, Springer-Verlag, 1992.

# References VII

[BertsekasEtShreve78] D.P. Bertsekas, and S.E. Shreve, *Stochastic optimal control, the discrete time case, volume 139 of Mathematics in Science and Engineering*, Academic Press Inc. [Har- court Brace Jovanovich Publishers], New York, 1978.

[KushnerEtYin97] H.J. Kushner and G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, Stochastic Modelling and Applied Probability Vol. 35, Springer, 1997.

[BallyEtal05] V. Bally, G. Pagès and J. Printems, *A quantization tree method for pricing and hedging multidimensional American options*, Math. Finance, 15(1), pp. 119-168 (2005).

[TsitsiklisEtVanRoy01] J.N. Tsitsiklis and B. Van Roy, *Regression Methods for Pricing Complex American-Style Options*, IEEE Transactions on Neural Networks, 12 (4) pp. 694-703 (2001).

[Carriere96] J. Carrière, *Valuation of the Early-Exercise Price for Options using Simulations and Nonparametric Regression*, Insurance: Math. and Eco., 19 (1), pp. 19-30 (1996).

# References VIII

[BouchardEtWarin12] B. Bouchard and X. Warin, *Monte-Carlo valorisation of American options: facts and new algorithms to improve existing methods*, in *Numerical Methods in Finance*, ed. R. Carmona, P. Del Moral, P. Hu and N. Oudjane, 2012.

[DelMoralEtal11] P. Del Moral, P. Hu, N. Oudjane, B. Rémillard, *On the Robustness of the Snell Envelope* SIAM J. Finan. Math., Vol. 2, pp. 587-626 , 2011.

[Ciarlet88] P. G. Ciarlet, *Introduction à l'analyse numérique matricielle et à l'optimisation.*, Masson, 1988.

[FournieEtal99] , Fournié, E. and Lasry, J.M. and Lebuchoux, J. and Touzi N., *Some applications of Malliavin calculus to Monte Carlo methods in finance*, Finance and Stochastics, Vol 3, 391-412 1999,

[Kunita82] H. Kunita. *Stochastic differential equations and stochastic flows of diffeomorphisms*. Cours d'école d'été de Saint-Flour, LN-1097, Springer-Verlag, 1982.

# References IX

[PardouxEtPeng92] E. Pardoux and S. Peng, *Backward Stochastic Differential equations ans Quasi-linear Parabolic Partial Differential Equations*, Lecture Notes in CIS vol. 176, springer-verlag p.200-217 (1992).

[ElKarouiEtMazliak97] N. El Karoui and L. Mazliak, *Backward Stochastic Differential Equations* Pitman research notes in mathematics series, 364, 1997.

[BouchardEtTouzi04] B. Bouchard and N. Touzi, *Discrete-time approximation and Monte-Carlo sim- ulation of backward stochastic differential equations*, Stochastic Process. Appl., 111(2):175-206, 2004.

[GobetEtal05] E. Gobet, J.P. Lemor, and X. Warin, *A regression-based Monte-Carlo method to solve backward stochastic differential equations*, Ann. Appl. Probab., 15(3):2172- 2202, 2005.

[Zhang04] J. Zhang, *A numerical scheme for backward stochastic differential equations*, An- nals of Applied Probability, 14(1), 459-488, 2004.

[GyorfiEtal02] L. Gyorfi, M. Kohler, A. Krzyzak, and H. Walk. *A distribution-free theory of nonparametric regression* Springer Series in Statistics, 2002.