Adaptive variance reduction for risk computing

Nadia Oudjane [(1)] joint work with François LeGland [(2)]
[(1)] Electricité de France (EDF) R&D
[(2)] INRIA Rennes

Plan

1

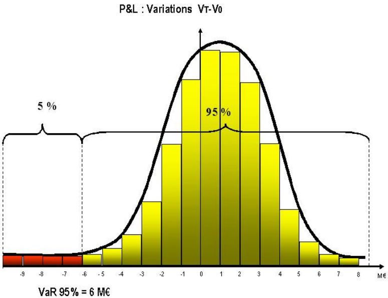
Computing Value at Risk

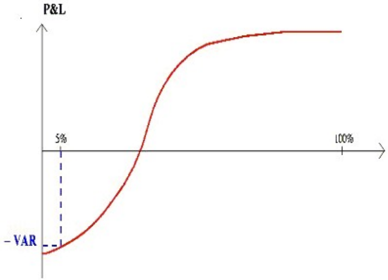
- We consider the variation of a portfolio value V_t between $t = 0$ and $t = T$

$$\Delta V(X) = V_T - V_0 + \int_0^T CF \ ,$$

where $X \in \mathbb{R}^d$ modellizes the risk factors between 0 and T.

- Value at Risk $\quad VaR_\alpha = \left| \inf\{s \in \mathbb{R} \mid \mathbb{P}(\Delta V \leq s) \geq 1-\alpha\} \right|$





$$VaR_\alpha = |F^{-}(\alpha)| \ , \quad \text{where } F(s) = \mathbb{P}(\Delta V \leq s) \ , \text{ for all } s \in \mathbb{R}$$

Monte Carlo method for VaR estimation

- The distribution function F can be viewed as an expectation

$$F(s) = \mathbb{E}[\mathbf{1}_{\Delta V(X) \leq s}] \ , \quad \text{for all} \quad s \in \mathbb{R}$$

- Traditional Monte Carlo Method for computing VaR

1. Monte Carlo simulations give an approximation of $F(s)$:

$$\hat{F}_N(s) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}(\Delta V(X_i) \leq s) \ , \quad \text{for all } s \in \mathbb{R}$$

$\Rightarrow$ Too many evaluations of ΔV for a given accuracy

2. Inversion of $\hat{F}^N$ and interpolation for approximating VaR

2

Importance Sampling

Importance Sampling

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

- Change of measure $p \longrightarrow q$ where q dominates φp

$$m = \mathbb{E}_p[\varphi(X)] = \mathbb{E}_q[\varphi(Y)\frac{p}{q}(Y)]\,, \quad \text{where} \quad X \sim p \quad \text{and} \quad Y \sim q$$

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

$$\hat{m}^q_M = \frac{1}{M}\sum_{i=1}^{M} \varphi(Y_i)\frac{p}{q}(Y_i) \xrightarrow[M\to\infty]{} \mathbb{E}_p[\varphi(X)]\,.$$

- Optimal change of measure $p \longrightarrow q^*$ (zero variance if $\varphi \geq 0$)

$$q^* = \frac{|\varphi|p}{\int |\varphi|(x)p(x)\,dx} = \frac{|\varphi|p}{\mathbb{E}_p[|\varphi|(X)]} = |\varphi|\cdot p$$

$\Rightarrow$ How to simulate and evaluate approximately q^* ?

Variance of the Importance Sampling estimate

• Let q be a (possibly random) importance probability density dominating q^*

$$Var(\hat{m}_M^q) = \mathbb{E}\Big[\, Var[\hat{m}_M^q \,|\, \mathcal{F}_q] \Big] + \underbrace{Var\Big[\mathbb{E}[\hat{m}_M^q \,|\, \mathcal{F}_q]\Big]}_{=0} \;,$$

where $\mathcal{F}_q$ denotes the sigma-algebra generated by q

• The variance of the IS estimate depends on the Chi-square distance between q and q^*

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

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

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

Some approaches to approximate importance distributions

- Large deviation approximation for rare events simulation [Bucklew04]
- Approximation of φ to obtain a simple expression for q^* ex : [Glasserman&al00] for computing VaR, Δ-Γ approximation of the portfolio value φ
- Cross-entropy [Homem-de-Mello&Rubinstein02] q^θ is chosen in a parametric family such as to minimize the entropy $K(q^\theta, q^*)$
- Mixture of kernels to approximate posterior distributions [West93] and [Raftery93]
- Progressive correction [Oudjane00]
- Review of different approaches [Evans&Schawrz95] and [Raftery93]

3

Particle approximation of the optimal importance distribution

Progressive correction [Musso&al01]

- We introduce a sequence of non negative functions $(G_k)_{0 \leq k \leq n}$

$$\begin{cases} G_0 = 1 \\ G_0 \cdots G_n = \varphi \\ G_k(x) = 0 \quad \text{implies} \quad G_{k+1}(x) = 0 \quad \text{for any} \quad x \in \mathbb{R}^d \ . \end{cases}$$

- For VaR computation $\varphi(x) = \mathbf{1}_{\Delta V(x) \leq s}$ then we choose

$$G_k(x) = \mathbf{1}_{\Delta V(x) \leq s_k} \ , \quad \text{with} \quad s = s_n \leq \cdots \leq s_0 = +\infty \ .$$

- Sequence of probability measures $(\nu_k)_{0 \leq k \leq n}$

$$\begin{cases} \nu_0 = p \, dx \\ \nu_k = \dfrac{G_k \nu_{k-1}}{\int_{\mathbb{R}^d} G_k(x) \nu_{k-1}(x) \, dx} = G_k \cdot \nu_{k-1} \ , \ \text{for all} \quad 1 \leq k \leq n \end{cases}$$

$\Rightarrow \nu_n = q^* \, dx$

- We introduce a sequence of Markov kernels $(Q_k)_{0\leq k\leq n}$ s. t.

$$\nu_k \approx \nu_k Q_k \quad \text{i.e.} \quad \nu_k(dx) \approx \int_{\mathbb{R}^d} \nu_k(du) Q_k(u, dx)\,, \quad \text{for all} \quad x \in \mathbb{R}^d$$

- In our case where $G_k(x) = \mathbf{1}_{\Delta V(x)\leq s_k}$, if p is Gaussian then Q_k is easily obtained from a Gaussian kernel Q reversible for p,

$$Q_k(x, dx') = Q(x, dx')\mathbf{1}_{\Delta V(x)\leq s_k}(x') + \left[1 - Q(x, \Delta V^{-}((-\infty, s_k]))\right]\delta_x(dx')$$

- Sequence of probability measures $(\nu_k)_{0\leq k\leq n}$

$$\begin{cases} \nu_0 = p\,dx \\ \nu_k = G_k \cdot (\nu_{k-1} Q_{k-1})\,, \quad \text{for all} \quad 1 \leq k \leq n \end{cases}$$

$\Rightarrow \nu_n = q^*\,dx$

Approximation of the dynamical system

Notation: empirical measure associated with μ

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

• The idea is to replace at each iteration k, $\nu_{k-1} Q_{k-1}$ by its N-empirical measure $S^N(\nu_{k-1} Q_{k-1})$ such that

$$S^N(\nu_{k-1} Q_{k-1}) = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_k^i} \quad \text{with} \quad (X_k^1, \cdots, X_k^N) \text{ i.i.d. } \sim \nu_{k-1} Q_{k-1}$$

• Sequence of dicrete probability measures $(\nu_k^N)_{0 \le k \le n}$

$$\begin{cases} \nu_0^N = S^N(\nu_0) \\ \nu_k^N = G_k \cdot S^N(\nu_{k-1}^N Q_{k-1}) \ , \quad \text{for all} \quad 1 \le k \le n \end{cases}$$

$\Rightarrow$ One can show [DelMoral04] that ν_n^N aproximates $q^* \, dx$ in the weak sense (when applied to tests functions).

Algorithm

- **Initialization** : Generate independently

$$(X_0^1, \cdots, X_0^N) \quad \text{i.i.d.} \quad \sim \quad p \quad \text{then set} \quad \nu_0^N = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_0^i}$$

- **Selection** : Generate independently

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

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

$$X_{k+1}^i \quad \sim \quad Q_k(\tilde{X}_k^i, \cdot)$$

- **Weighting** : For each particle $i \in \{1, \cdots, N\}$, compute

$$\omega_{k+1}^i = \frac{G_{k+1}(X_{k+1}^i)}{\sum_{j=1}^{N} G_{k+1}(X_{k+1}^j)} \quad \text{then set} \quad \nu_{k+1}^N = \sum_{i=1}^{N} \omega_{k+1}^i \, \delta_{X_{k+1}^i}$$

$\Rightarrow \nu_n^N = \sum_{i=1}^{N} \omega_i X_n^i$

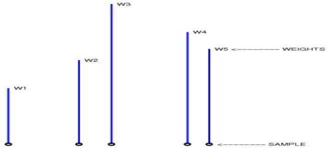
4

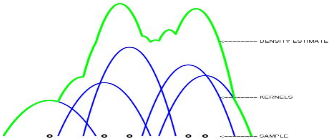
Smooth approximation of the optimal importance density

Weighted Density estimation

- Kernel K radially symetric density, with infinite support.
- Kernel of ordre 2 $\int \|x\|^2 K(x)dx < \infty$.
- Rescaled kernel K_h $\qquad K_h(x) = \frac{1}{h^d} K(\frac{x}{h})$
- Weighted Kernel $(X_1, \cdots, X_N)$ iid , $\omega_i \geq 0$ and $\sum \omega_i = 1$

$$\hat{\nu} = \sum \omega^i \, \delta_{X^i} \xrightarrow[K_h * \cdot]{\text{Weihted Density estimation}} \hat{q} = \sum \omega^i \, K_h(\cdot - X^i)$$





- For "non weighted" density estimation one proves that

$$\mathbb{E}\|\hat{q} - q\|_1 \leq \frac{C}{N^{\frac{2}{d+4}}} \, .$$

- No theoritical result on relative errors available

Theoritical result: variance bounds

• Let us consider $q_n^{N,h}$ the weighted density estimate obtained by convolution of $\nu_n^N = \sum_{i=1}^N \omega_i X_n^i$ by K_h

$$q_n^{N,h}(x) = K_h * \nu_n^N = \sum_{i=1}^N \omega_i K_h(x - X_n^i) \ .$$

• Assume that the target density q^* satisfies the following assumptions:

Assumption Q1: $q^* \propto Hp$ has a bounded support C in $\mathbb{R}^d$.

Assumption Q2: $q^* \propto Hp$ has second derivatives in $L^2(\mathbb{R}^d)$.

Assumption K2: The kernel K has "sufficiently heavy" tails.

• If the smoothing factor h is chosen s. t. $h \propto 1/N^{1/(d+4)}$, then

$$Var(\hat{m}_M^{q_n^{N,h}}) \leq \frac{C}{MN^{1/(d+4)}} \ .$$

5

Simulations results

Adaptive choice of the sequence $(G_k)_{0\leq k\leq n}$

[Musso&al01], [Hommem-de-Mello&Rubinstein02], [Cérou&al06]

• The performance of Interacting particle systems is known to deteriorate when the ratio $\frac{N \max G_k}{\sum_{i=1}^N G_k(X_i^k)}$ degenerate to infinity

The idea is to chose G_k such that $\frac{1}{N}\sum_{i=1}^N G_k(X_k^i)$ is not too small

• In our case where $G_k(x) = \mathbf{I}_{\Delta V(x)\leq s_k}$, the threshold s_k is chosen as a r.v. depending on the current particle system and on a parameter $\rho \in (0,1)$:

$$s_k = \inf\left\{ s \quad \text{such that} \quad \sum_{i=1}^N \mathbf{I}_{\Delta V(X^i)\leq s} \geq \rho N \right\}$$

• This choice of s_k is not prooved to guarantee that the algorithms ends in a finite number of iterations but this point does not seem to be a problem in our simulations

Some simulation results : Variance ratio

- Several test cases depending on the form of function $x \mapsto \Delta V(x)$ have been studied: results are all comparable
- X is a d dimensional Gaussian variable and $m = \mathbb{E}_p[\mathbb{1}_{\Delta V(X) \leq s}]$
- Particles $N = 500$; Iterations $n \approx 10$ to 60; Simulations $M = 10\,000$
- The performance of our approach has been compared to Interacting Particle Systems whithout I S [DelMoral&Garnier05]

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

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