Recursive Computation of Value-at-Risk and Conditional Value-at-Risk using MC and QMC

Olivier Bardou, Noufel Frikha, and Gilles Pagès

Abstract Value-at-Risk (VaR) and Conditional-Value-at-Risk (CVaR) are two widely-used measures in risk management. This paper deals with the problem of estimating both VaR and CVaR using stochastic approximation (with decreasing steps): we propose a first Robbins-Monro (RM) procedure based on Rockafellar-Uryasev's identity for the CVaR. The estimator provided by the algorithm satisfies a Gaussian Central Limit Theorem. As a second step, in order to speed up the initial procedure, we propose a recursive and adaptive importance sampling (IS) procedure which induces a significant variance reduction of both VaR and CVaR procedures. This idea, which has been investigated by many authors, follows a new approach introduced in Lemaire and Pagès [20]. Finally, to speed up the initialization phase of the IS algorithm, we replace the original confidence level of the VaR by a deterministic moving risk level. We prove that the weak convergence rate of the resulting procedure is ruled by a Central Limit Theorem with minimal variance and we illustrate its efficiency by considering typical energy portfolios.

O. Bardou
Laboratoire de Probabilités et Modèles aléatoires France and GDF Suez, Research and Innovation Department
e-mail: olivier-aj.bardou@gdfsuez.com

N. Frikha
Laboratoire de Probabilités et Modèles aléatoires, Université Pierre et Marie Curie, UMR 7599 and GDF Suez, Research and Innovation Department, France
e-mail: noufel-externe.frikha@gdfsuez.com

G. Pagès
Laboratoire de Probabilités et Modèles aléatoires, UMR 7599, Université Pierre et Marie Curie, France
e-mail: gilles.pages@upmc.fr

P. L'Ecuyer, A.B. Owen (eds.), *Monte Carlo and Quasi-Monte Carlo Methods 2008*,
DOI 10.1007/978-3-642-04107-5_11,

1 Introduction

Following financial institutions, energy companies are developing a risk management framework to face the new price and volatility risks associated to the growth of energy markets. Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) are certainly the best known and the most common risk measures used in this context, especially for the evaluation of extreme losses potentially faced by traders. Naturally related to rare events, the estimation of these risk measures is a numerical challenge. The Monte Carlo method, which is often the only available numerical device in such a general framework, must preferably be associated to efficient variance reduction techniques to remedy its slow convergence rate.

By definition, the VaR_α of a given portfolio at a specified level $\alpha \in (0,1)$ is the lowest amount not exceeded by the loss with probability α. The CVaR_α is the conditional expectation of the portfolio's losses above the VaR_α. Compared to VaR, CVaR is known to have better properties. It is a coherent risk measure in the sense of Artzner, Delbaen, Eber and Heath, see [2]. The most commonly used method to compute VaR is the inversion of the simulated empirical loss distribution function using Monte Carlo or historical simulation tools. Another well-known method relies on linear or quadratic expansion of the distribution of the loss see e.g. [6], [7], [14], [15] and [24]. However, such approximations are no longer acceptable when considering portfolios over a long time interval as it is often the case in energy markets (1 year up to 10 years) or when the loss is a functional of a general path-dependent Stochastic Differential Equation (SDE).

In the context of hedging or optimizing a portfolio of financial instruments by reducing the Conditional Value-at-Risk, it is shown in [23] that it is possible to compute both VaR and CVaR (actually calculate VaR and optimize CVaR) by solving a convex optimization problem with a linear programming approach. It consists in generating loss scenarios and then in introducing constraints in the linear programming problem. Although a different problem is addressed in this paper, the method described in [23] can be used to compute both VaR and CVaR. The advantage of such an approach is that it is possible to estimate both VaR and CVaR simultaneously and without assuming that the market prices have a specific distribution like normal or log-normal. The main drawback is that the dimension (number of constraints) of the linear programming problem to be solved is equal to the number of simulated scenarios. In our approach, we are not constrained by the number of generated sample paths used in the estimation.

The idea to compute both VaR and CVaR with one procedure comes from the fact that they appear as the solutions and the value of the same convex optimisation problem (see Proposition 1) as demonstrated in [23]. Moreover, the convex objective function of the minimization problem reads as an expectation and its gradient too, so that a method to estimate both quantities is to devise a stochastic gradient algorithm and an averaging procedure. Thus, we derive a global recursive procedure which is an alternative method compared to the basic two step procedure which consists in first estimating the VaR using the inversion of the empirical function method and then estimating the CVaR by averaging. This optimization approach provides

a convex Lyapunov function (the gradient of the objective function) for the system which allows to derive the almost sure ($a.s.$) convergence of the VaR procedure. Moreover, the implementation of the algorithm is very easy. From a practical point of view, there is no reason to believe that this procedure behaves better than this alternative method. However, the proposed algorithm is just a building block of a recursive and adaptive IS procedure. As a matter of fact, basically in this kind of problem we are interested by events that are observed with a very small probability (usually less that 5%, 1% or even 0.1%) thus we obtain few significant replications to update our estimates. When α is close to 1 (otherwise it is not a numerical challenge), VaR and CVaR are fundamentally related to rare events thus as a necessary improvement, we also introduce a recursive variance reduction method. To compute more accurate estimates of both quantities of interest, it is necessary to generate more samples in the tail of the loss distribution, the area of interest. A general tool used in this situation is IS. The basic principle of IS is to modify the distribution of the loss by an equivalent change of measure to obtain more "interesting" samples that will lead to better estimates of the VaR and CVaR. The main issue is to find the right change of measure (among a parametrized family) that will induce a significant variance reduction. In [10], the VaR_α is estimated by using a quantile based on a weighted empirical distribution function and combined with a projected IS algorithm. This kind of algorithm is known to converge after a long stabilization phase and provided that the sequence of compact sets has been specified appropriately. Specifying adequately the sequence of compact sets is a significant challenge. Our IS parameters are optimized by an adaptive unconstrained (i.e., without projections) RM algorithm which is combined with our VaR-CVaR procedure. The fact that our estimates for both VaR and CVaR are recursive makes the algorithm well suited for the recursive IS procedure.

One major issue that arises when combining the VaR-CVaR algorithm with the recursive IS procedure is to ensure importance sampling parameters do move appropriately toward the critical risk area. They may remain "stuck" at the very beginning of the IS procedure. To circumvent this problem, we make the confidence level α slowly increase from a low level (say 50%) to the true value of α by introducing a deterministic sequence α_n that converges to α at a prespecified rate. This kind of incremental threshold increase has been proposed previously in [25] in a different framework (use of cross entropy). We finally discuss the possibility of plugging low-discrepancy sequences instead of pseudo-random numbers in the VaR-CVaR algorithm.

The paper is organized as follows. In Section 2 we briefly present the VaR-CVaR Robbins Monro algorithm in its first and naive version. It's a building block that is necessary in order to combine it with adaptive IS. Then, we introduce and study the adaptive variance reduction procedure and present how it modifies the asymptotic variance of our first CLT in Section 2.2. Numerical illustrations are given in Section 3.

2 Design of the VaR-CVaR Stochastic Approximation Algorithm

2.1 Devise of a VaR-CVaR Procedure (First Phase)

2.1.1 Definition of the VaR and the CVaR

We consider that the loss of the portfolio over the considered time horizon can be written as a function of a structural finite dimensional random vector, i.e., $L = \varphi(X)$ where X is a $\mathbb{R}^d$-valued random vector defined on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and $\varphi : \mathbb{R}^d \to \mathbb{R}$ is a Borel function. Thus, φ is the function describing the composition of the portfolio which remains fixed and X is a structural d-dimensional random vector used to model the market prices over a given time interval. We only rely on the fact it is possible to sample from the distribution of X. For instance, in a Black-Scholes framework, X is generally a vector of Brownian increments related to the Euler scheme of a diffusion. The VaR at level $\alpha \in (0,1)$ is the lowest α-quantile of the distribution $\varphi(X)$:

$$\mathrm{VaR}_\alpha(\varphi(X)) := \inf\{\xi \mid \mathbb{P}(\varphi(X) \le \xi) \ge \alpha\}.$$

Since $\lim_{\xi \to +\infty} \mathbb{P}(\varphi(X) \le \xi) = 1$ and $\lim_{\xi \to -\infty} \mathbb{P}(\varphi(X) \le \xi) = 0$, the VaR always exists. We assume that the distribution function of $\varphi(X)$ is continuous (i.e., without atoms) thus the VaR_α is the lowest solution of the equation:

$$\mathbb{P}(\varphi(X) \le \xi) = \alpha.$$

Another risk measure generally used to provide information about the tail of the distribution of $\varphi(X)$ is the *Conditional Value at Risk* (CVaR) (at level α). Assuming that $\varphi(X) \in L^1(\mathbb{P})$, the CVaR is defined by:

$$\mathrm{CVaR}_\alpha(\varphi(X)) := \mathbb{E}[\varphi(X) | \varphi(X) \ge \mathrm{VaR}_\alpha(\varphi(X))].$$

2.1.2 Stochastic Gradient and Averaging Procedure: A Naive Approach

In this paragraph, we present the framework and the first building block of the VaR-CVaR algorithm. Obviously, there is no reason to believe that this first and naive version can do better than others method, like the inversion of the empirical distribution function. However, our quantile estimate has the advantage to be recursive thus it can be combined later with a recursive IS algorithm in a suitable way.

Proposition 1. *Let V be the function defined on $\mathbb{R}$ by: $\xi \mapsto \xi + \frac{1}{1-\alpha}\mathbb{E}[(\varphi(X) - \xi)_+)]$. Suppose that the distribution function of $\varphi(X)$ is continuous. Then, the function V is convex, differentiable and the $\mathrm{VaR}_\alpha(\varphi(X))$ is any point of the set:*

$$\arg\min\ V = \left\{\xi \in \mathbb{R} \mid V'(\xi) = 0\right\} = \{\xi \mid \mathbb{P}(\varphi(X) \le \xi) = \alpha\}$$

where V' is the derivative defined of V. Moreover, for every $\xi \in \mathbb{R}$, $V'(\xi) = \mathbb{E}[H_1(\xi, X)]$ where

$$H_1(\xi, x) := 1 - \frac{1}{1-\alpha} \mathbf{1}_{\{\varphi(x) \geq \xi\}}.$$

Furthermore, $CVaR_\alpha(\varphi(X)) = \min_{\xi \in \mathbb{R}} V(\xi)$.

Proof. Since the function $\xi \mapsto (\varphi(x) - \xi)_+$, $x \in \mathbb{R}^d$, is convex, the function V is convex. $\mathbb{P}(dw)$-*a.s.*, $H_1(\xi, X(w))$ exists at every $\xi \in \mathbb{R}$ and

$$\mathbb{P}(dw)\text{-}a.s., \quad |H_1(\xi, X(w))| \leq 1 \vee \frac{\alpha}{1-\alpha}.$$

Thanks to Lebesgue Dominated Convergence Theorem, one can interchange differentiation and expectation, so that V is differentiable with derivative $V'(\xi) = \mathbb{E}[H_1(\xi, X)] = 1 - \frac{1}{1-\alpha}\mathbb{P}(\varphi(X) > \xi)$ and reaches its absolute minimum at any ξ^* satisfying $\mathbb{P}(\varphi(X) > \xi^*) = 1 - \alpha$, i.e., $\mathbb{P}(\varphi(X) \leq \xi^*) = \alpha$. Moreover,

$$V(\xi^*) = \frac{\xi^* \mathbb{E}[\mathbf{1}_{\varphi(X) > \xi^*}] + \mathbb{E}[(\varphi(X) - \xi^*)_+]}{\mathbb{P}(\varphi(X) > \xi^*)} = \mathbb{E}\big[\varphi(X) | \varphi(X) > \xi^*\big].$$

This completes the proof. □

Since we are looking for ξ for which $\mathbb{E}[H_1(\xi, X)] = 0$, we implement a stochastic gradient descent derived from the Lyapunov function V to approximate $\xi^* := \mathrm{VaR}_\alpha(\varphi(X))$, i.e., we use the RM algorithm:

$$\xi_n = \xi_{n-1} - \gamma_n H_1(\xi_{n-1}, X_n), \quad n \geq 1 \tag{1}$$

where $(X_n)_{n \geq 1}$ is an i.i.d. sequence of random variables with the same distribution as X, independent of ξ_0, with $\mathbb{E}[|\xi_0|] < \infty$ and $(\gamma_n)_{n \geq 1}$ is a positive deterministic step sequence (decreasing to 0) satisfying

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

A natural idea in order to estimate $C^* := \mathrm{CVaR}_\alpha$ is to devise an averaging procedure of the above quantile search algorithm, namely C_0 and for $n = 1, 2, \ldots$,

$$C_n = \frac{1}{n} \sum_{k=0}^{n-1} \xi_k + \frac{1}{1-\alpha} (\varphi(X_{k+1}) - \xi_k)_+ = C_{n-1} - \frac{1}{n} H_2(\xi_{n-1}, C_{n-1}, X_n). \tag{3}$$

where $H_2(\xi, c, x) := c - \xi - \frac{1}{1-\alpha} (\varphi(x) - \xi)_+$. The resulting algorithm reads for $n \geq 1$:

$$\begin{cases} \xi_n = \xi_{n-1} - \gamma_n H_1(\xi_{n-1}, X_n), & \xi_0 \in \mathbb{R} \\ C_n = C_{n-1} - \frac{1}{n} H_2(\xi_{n-1}, C_{n-1}, X_n), & C_0 = 0. \end{cases} \tag{4}$$

At this stage, we are facing a two-time scale algorithm in (4), i.e., with different steps γ_n and $1/n$. Since H_1 is free of C_n, we know that the stepsize sequence that provides

the best convergence rate (see e.g. [9]) is $\gamma_n = \gamma_1/n$ where γ_1 has to be chosen adequately. So we verify a posteriori that the resulting algorithm could theoretically be reduced to a one-time-scale procedure. Our numerical experiments indicate that the one-time scale procedure provides less variance during the first iterations than others procedures with different steps size. A slight modification consists in using both procedures with the same step size $(\gamma_n)_{n\geq 1}$ satisfying condition (2) (for more details about the different possible choice we refer to [3]). The resulting algorithm can be written as for $n \geq 1$

$$\begin{cases} \xi_n = \xi_{n-1} - \gamma_n H_1(\xi_{n-1}, X_n), \quad \xi_0 \in \mathbb{R}, \\ C_n = C_{n-1} - \gamma_n H_2(\xi_{n-1}, C_{n-1}, X_n), \quad C_0 = 0. \end{cases} \tag{5}$$

The recurrence for (ξ_n) does not involve (C_n) so its $a.s.$ convergence is ensured by the classical RM Theorem (see e.g. [9]). Then, it is possible to show that C_n converges $a.s.$ toward C^* provided that the distribution function of $\varphi(X)$ is continuous and increasing and that $\varphi(X)$ is square integrable (we refer to [3] for a proof). To achieve the best convergence rate, we are led to introduce the Ruppert and Polyak's averaging principle (see [17] and [26]). If we set $\gamma_n = cn^{-p}$, with $\frac{1}{2} < p < 1$, $c > 0$ in (5) and compute the Césaro means of both components

$$\begin{cases} \overline{\xi}_n := \frac{1}{n}\sum_{k=1}^n \xi_k = \overline{\xi}_n - \frac{1}{n+1}(\overline{\xi}_n - \xi_n) \\ \overline{C}_n := \frac{1}{n}\sum_{k=1}^n C_k = \overline{C}_n - \frac{1}{n+1}(\overline{C}_n - C_n) \end{cases} \tag{6}$$

where (ξ_k, C_k), $k \geq 0$ is defined by (5) then, provided that

$$\mathbb{E}\left[|\varphi(X)|^{2a}\right] < +\infty \text{ for some } a > 1, \tag{7}$$

and that the distribution of $\varphi(X)$ has a positive probability density $f_{\varphi(X)}$ on its support, we obtain asymptotically efficient estimators which satisfy the Gaussian CLT:

$$\sqrt{n}\begin{pmatrix} \overline{\xi}_n - \xi^* \\ \overline{C}_n - C^* \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma) \tag{8}$$

where the asymptotic covariance matrix Σ is given by

$$\Sigma = \begin{pmatrix} \frac{\alpha(1-\alpha)}{f_{\varphi(X)}(\xi^*)} & \frac{\alpha}{(1-\alpha) f_{\varphi(X)}(\xi^*)}\mathbb{E}\left[(\varphi(X) - \xi^*)_+\right] \\ \frac{\alpha}{(1-\alpha) f_{\varphi(X)}(\xi^*)}\mathbb{E}\left[(\varphi(X) - \xi^*)_+\right] & \frac{1}{(1-\alpha)^2}\mathrm{Var}\,(\varphi(X) - \xi^*)_+ \end{pmatrix}. \tag{9}$$

Remark 1. The result above is not surprising. The asymptotic variance of the quantile estimate based on the inversion of the empirical distribution function is the same than the one of our procedure $\overline{\xi}_n$, see for example [27], page 75.

The bottleneck in using the above algorithm lies in its very slow convergence owing to the fact that $\mathbb{P}(\varphi(X) > \xi^*) = 1 - \alpha$ is close to 0 in practical implementations, so that we observe few significant replications to update our estimates. Moreover,

in the bank and energy sectors, practitioners usually deal with huge portfolio composed by hundreds or thousands of risk factors and options. The evaluation step of $\varphi(X)$ may require a lot of computational time. Consequently, to achieve accurate estimates of both VaR_α and CVaR_α with reasonable computational effort, the above algorithm (6) needs to be speeded up by an IS procedure to recenter simulation "where things do happen", i.e., scenarios for which $\varphi(X)$ exceeds ξ.

2.2 Variance Reduction Using Adaptive Recursive Importance Sampling (Final Phase)

In this section, we investigate the IS by translation. We show how the IS algorithm investigated in [20] can be combined adaptively with our first algorithm. Consequently, every new sample is used to dynamically optimize the IS change of measure and the estimate of both VaR and CVaR.

2.2.1 Some Background on Recursive IS

Suppose that $F(X)$ is a square integrable random variable such that $\mathbb{P}(F(X) \neq 0) > 0$ and where X is a random vector with density function p over $\mathbb{R}^d$. The main idea of IS by translation, applied to the computation of $\mathbb{E}[F(X)]$, is to use the invariance of the Lebesgue measure by translation: it follows that for every $\theta \in \mathbb{R}^d$,

$$\mathbb{E}[F(X)] = \int_{\mathbb{R}^d} F(x)p(x)dx = \int_{\mathbb{R}^d} F(x+\theta)p(x+\theta)dx = \mathbb{E}\left[F(X+\theta)\frac{p(X+\theta)}{p(X)}\right]. \tag{10}$$

Among all these random vectors with the same expectation, we want to select the one with the lowest variance, i.e., the one with lowest quadratic norm

$$Q(\theta) := \mathbb{E}\left[F^2(X+\theta)\frac{p^2(X+\theta)}{p^2(X)}\right], \quad \theta \in \mathbb{R}^d.$$

A reverse change of variable shows that:

$$Q(\theta) = \mathbb{E}\left[F^2(X)\frac{p(X)}{p(X-\theta)}\right], \quad \theta \in \mathbb{R}^d. \tag{11}$$

Now if the density function p of X satisfies

$$(i) \;\; p \text{ is log-}concave \quad \text{and} \quad (ii) \lim_{\|x\|\to+\infty} p(x) = 0 \tag{12}$$

where $\|.\|$ denotes the Euclidean norm, and

$$Q(\theta) < +\infty, \; \forall \theta \in \mathbb{R}^d \tag{13}$$

then, one shows that the function Q is finite, convex and goes to infinity at infinity, thus $\arg\min_\theta Q = \left\{\theta \in \mathbb{R}^d \mid \nabla Q(\theta) = 0\right\}$ is non empty, where ∇Q is the gradient of Q (for a proof, we refer to [20]). If ∇Q admits a representation as an expectation, then it is possible to devise a recursive RM procedure to approximate the optimal parameter θ^*, namely

$$\theta_n = \theta_{n-1} - \gamma_n K(\theta_{n-1}, X_n),\ n \geq 1 \tag{14}$$

where K is naturally defined by the formal differentiation of Q, for every $x \in \mathbb{R}^d$:

$$K(\theta, x) = F^2(x)\frac{p(x)}{p^2(x-\theta)}\nabla p(x-\theta). \tag{15}$$

Since we have no knowledge about the regularity of F and do not wish to have any, we differentiate the second representation of Q in (11) and not the first one. IS using stochastic approximation algorithms has been investigated by several authors, see e.g. [16], [11] and [13] in order to "optimize" or "improve" the change of measure in IS by a RM procedure. It has recently been studied in the Gaussian framework in [1] where (15) is used to design a stochastic gradient algorithm. However, the regular RM procedure (14) suffers from an instability issue coming from the fact that the classical sub-linear growth assumption in quadratic mean in the Robbins-Monro Theorem

$$\forall \theta \in \mathbb{R}^d,\ \ \mathbb{E}\left[K(\theta, X)^2\right]^{\frac{1}{2}} \leq C\,(1 + \|\theta\|) \tag{16}$$

is only fulfilled when F is constant, due to the behavior of the annoying term $p(x)/p(x-\theta)$ as θ goes to infinity. Consequently, θ_n can escape at infinity at almost every implementation as pointed out in [1]. To circumvent this problem, a "projected version" of the procedure based on repeated reinitializations when the algorithms exits from an increasing sequence of compact sets (while the step γ_n keeps going to 0) was used. This approach is known as the projection "à la Chen". It forces the stability of the algorithm and prevents explosion. Recently, IS using stochastic algorithm was deeply revisited in [20] to remove the constraints introduced by the original algorithm. Moreover, this construction is extended to a large class of probability distributions and to diffusion process. Thanks to another translation of the variable θ, it is possible to plug back the parameter θ "into" $F(X)$, the function F having in common applications a known behavior at infinity.

2.2.2 Unconstrained Adaptive Importance Sampling Algorithm Applied to the VaR-CVaR Procedure

Applied to the problem we are dealing with, the main idea is to twist (by translation) the distribution of X in order to minimize the asymptotic variance of the two components in the CLT (8): the asymptotic variances of the VaR_α and CVaR_α algorithm

$$\frac{\mathrm{Var}\left(\mathbf{1}_{\varphi(X)\geq\xi^*}\right)}{f_{\varphi(X)}(\xi^*)} = \frac{\alpha(1-\alpha)}{f_{\varphi(X)}(\xi^*)} \quad \text{and} \quad \frac{\mathrm{Var}\left((\varphi(X)-\xi^*)_+\right)}{(1-\alpha)^2}.$$

By importance sampling, it is not possible to modify the quantity $f_{\varphi(X)}(\xi^*)$ since it is an intrinsic constant which appears in the CLT (8) through the Jacobian matrix of h, where $h(\xi, C) := \mathbb{E}[H(\xi, C, X)]$ and $H(\xi, C, X) := (H_1(\xi, C, X), H_2(\xi, C, X))$. Consequently, we are led to find the parameters θ^* and μ^* minimizing the two functionals:

$$Q_1(\theta, \xi^*) := \mathbb{E}\left[\mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\frac{p(X)}{p(X-\theta)}\right], \quad Q_2(\mu, \xi^*) := \mathbb{E}\left[\left(\varphi(X)-\xi^*\right)_+^2\frac{p(X)}{p(X-\mu)}\right] \tag{17}$$

under the conditions that for every $(\xi, \theta) \in \mathbb{R}\times\mathbb{R}^d$,

$$\mathbb{E}\left[\left(1+(\varphi(X)-\xi)_+^2\right)\frac{p(X)}{p(X-\theta)}\right] < +\infty \tag{18}$$

and

$$\forall \xi \in \arg\min \; V, \;\; \mathbb{P}(\varphi(X) > \xi) > 0. \tag{19}$$

By differentiation we easily prove that

$$\nabla_\theta Q_1(\theta, \xi^*) = \mathbb{E}\left[\mathbf{1}_{\varphi(X)\geq\xi^*}\frac{p(X)}{p^2(X-\theta)}\nabla p(X-\theta)\right], \tag{20}$$

$$\nabla_\mu Q_2(\mu, \xi^*) = \mathbb{E}\left[\left(\varphi(X)-\xi^*\right)_+^2\frac{p(X)}{p^2(X-\mu)}\nabla p(X-\mu)\right]. \tag{21}$$

The key idea is to introduce a third change of probability in order to control the annoying terms $p(X)/p(X-\theta)$, $p(X)/p(X-\mu)$ by plugging back the parameters θ and μ into $\mathbf{1}_{\{\varphi(X)\geq\xi^*\}}$ and $\varphi(X)$ respectively. Now we follow [20] to design a regular unconstrained Robbins-Monro algorithm which converges $a.s.$ to the optimal parameters θ^* and μ^* (without risk of explosion) provided the growth of $x \mapsto \varphi(x)$ at infinity can be explicitly controlled. From now on, we assume that there exist two positive constants $a,\ C\ > 0$ such that

$$\forall x \in \mathbb{R}^d, \;\; |\varphi(x)|^2 \leq C\, e^{a\|x\|}. \tag{22}$$

We make the following assumption on the probability density p of X

$$\exists b \in [1,2] \text{ such that } \begin{cases} (i) & \frac{\|\nabla p(x)\|}{p(x)} = O(||x||^{b-1}) \quad \text{as} \quad \|x\| \to \infty \\ (ii) & \exists\, \rho > 0 \text{ such that } \log(p(x)) + \rho\,\|x\|^b \;\; \text{is convex.} \end{cases} \tag{23}$$

Then, one shows that under the conditions (18), (19), (22), (23), Q_1 and Q_2 are both finite and differentiable on $\mathbb{R}^d$ with gradients given by

$$\nabla Q_1(\theta,\xi^*) := \mathbb{E}\left[\mathbf{1}_{\varphi(X-\theta)\geq\xi^*}\underbrace{\frac{p^2(X-\theta)}{p(X)p(X-2\theta)}\frac{\nabla p(X-2\theta)}{p(X-2\theta)}}_{W_1(\theta,X)}\right], \tag{24}$$

$$\nabla Q_2(\mu,\xi^*) := \mathbb{E}\left[\left(\varphi(X-\mu)-\xi^*\right)_+^2\underbrace{\frac{p^2(X-\mu)}{p(X)p(X-2\mu)}\frac{\nabla p(X-2\mu)}{p(X-2\mu)}}_{W_2(\mu,X)}\right]. \tag{25}$$

We refer to [20] for a proof. The two last expressions may look complicated at first glance but, in fact, the weight term of the expectation involving the probability density can be easily controlled by a deterministic function of θ. For instance, when $X \overset{d}{=} \mathcal{N}(0;1)$,

$$W_1(\theta,X) = e^{\theta^2}(2\theta - X)$$

and more generally, under conditions (18) and (23), there exist two constants A and B such that

$$|W_1(\theta,X)| \leq e^{2\rho|\theta|^b}\left(A\,\|x\|^{b-1} + A\,\|\theta\|^{b-1} + B\right) \tag{26}$$

so that this weight can always be controlled by a deterministic function of θ (for more details, one can refer to [20]). Then by setting,

$$\widetilde{W}_1(\theta,X) := e^{-2\rho|\theta|^b}W_1(\theta,X), \quad \widetilde{W}_2(\mu,X) := e^{-2\rho|\theta|^b - a(\|\mu\|^2+1)}W_2(\mu,X),$$

we can define K_1 and K_2 by

$$K_1(\xi^*,\theta,x) := \mathbf{1}_{\varphi(X-\theta)\geq\xi^*}\widetilde{W}_1(\theta,x), \quad K_2(\xi^*,\mu,x) := \left(\varphi(X-\mu)-\xi^*\right)_+^2\widetilde{W}_2(\mu,x)$$

so that they satisfy the linear growth assumptions (16) and

$$\left\{\theta\in\mathbb{R}^d \mid \mathbb{E}\left[K_1(\xi^*,\theta,X)\right]=0\right\} = \left\{\theta\in\mathbb{R}^d \mid \nabla Q_1(\theta,\xi^*)=0\right\},$$
$$\left\{\mu\in\mathbb{R}^d \mid \mathbb{E}\left[K_2(\xi^*,\mu,X)\right]=0\right\} = \left\{\mu\in\mathbb{R}^d \mid \nabla Q_2(\mu,\xi^*)=0\right\}.$$

Now, since we do not know either ξ^* and C^* (the VaR_α and the CVaR_α) respectively we make the whole procedure adaptive by replacing at step n, these unknown parameters by their running approximation at step $n-1$. This finally justifies to introduce the following global procedure. One defines the state variable $Z_n := (\xi_n, C_n, \theta_n, \mu_n)$ where ξ_n, C_n denote the VaR_α and the CVaR_α approximations, θ_n, μ_n denote the variance reducers for the VaR and the CVaR procedures. We update this state variable recursively by

$$Z_n = Z_{n-1} - \gamma_n L(Z_{n-1}, X_n), \quad n\geq 1 \tag{27}$$

where $(X_n)_{n\geq 1}$ is an i.i.d. sequence with the same distribution as X, and L is defined as follow $L(z,x) = (L_1(\xi,\theta,x),\ L_2(\xi,C,\mu,x),\ K_1(\xi,\theta,x),\ K_2(\xi,\mu,x))$ with

$$L_1(\xi,\theta,x) := e^{-\rho\|\theta\|^b}\left(1-\frac{1}{1-\alpha}\,\mathbf{1}_{\{\varphi(x+\theta)\geq\xi\}}\frac{p(x+\theta)}{p(x)}\right) \tag{28}$$

$$L_2(\xi,C,\mu,x) := C-\left(\xi+\frac{1}{1-\alpha}(\varphi(x+\mu)-\xi)_+\frac{p(x+\mu)}{p(x)}\right). \tag{29}$$

Theorem 1. *(a) Convergence. Suppose* $\mathbb{E}\left[\varphi(X)^2\right] < +\infty$. *Assume that conditions* (18), (19), (22), (23) *are fulfilled and that the step sequence* $(\gamma_n)_{n\geq 1}$ *satisfies* (2). *Then,*

$$Z_n \xrightarrow{a.s.} z^* := (\xi^*, C^*, \theta_\alpha^*, \mu_\alpha^*)$$

where $(Z_n)_{n\geq 0}$ *is the recursive sequence defined by* (27), *where* $C^* = CVaR_\alpha$ *and*

$$(\xi^*,\theta_\alpha^*,\mu_\alpha^*) \in \mathcal{T}^* := \Big\{(\xi,\theta,\mu)\in\mathbb{R}\times\mathbb{R}^d\times\mathbb{R}^d :\ \xi\in\arg\min V,\ \nabla Q_1(\theta,\xi)=\nabla Q_2(\mu,\xi)=0\Big\}.$$

(b) Ruppert and Polyak CLT. Suppose that the assumptions of (a) hold and that the density $f_{\varphi(X)}$ *of* $\varphi(X)$ *is positive on its support, differentiable, and that* (7) *hold. Let* $(\overline{\xi}_n,\overline{C}_n)_{n\geq 1}$ *be the sequence defined by:*

$$\overline{\xi}_n := \frac{\xi_0+\ldots+\xi_{n-1}}{n},\quad \overline{C}_n := \frac{C_0+\ldots+C_{n-1}}{n}. \tag{30}$$

This sequence satisfies the following CLT:

$$\sqrt{n}\begin{pmatrix}\overline{\xi}_n-\xi^*\\ \overline{C}_n-C^*\end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}(0,\Sigma^*) \tag{31}$$

where the elements of Σ^* *are given by*

$$\Sigma^*_{1,1} = \frac{1}{f_{\varphi(X)}(\xi^*)}\ \mathrm{Var}\left(\mathbf{1}_{\{\varphi(X+\theta_\alpha^*)\geq\xi^*\}}\frac{p(X+\theta_\alpha^*)}{p(X)}\right),$$

$$\Sigma^*_{1,2} = \Sigma^*_{2,1} = \frac{\mathrm{Cov}\left((\varphi(X+\mu_\alpha^*)-\xi^*)_+\frac{p(X+\mu_\alpha^*)}{p(X)},\mathbf{1}_{\{\varphi(X+\theta_\alpha^*)\geq\xi^*\}}\frac{p(X+\theta_\alpha^*)}{p(X)}\right)}{(1-\alpha)\,f_{\varphi(X)}(\xi^*)},$$

$$\Sigma^*_{2,2} = \mathrm{Var}\left((\varphi(X+\mu_\alpha^*)-\xi^*)_+\frac{p(X+\mu_\alpha^*)}{p(X)}\right).$$

Proof. We give a summary of the arguments. (a) First one shows that the sequence (ξ_n,θ_n,μ_n) converges $a.s.$ using the classical Robbins-Monro Theorem (see e.g. [9] or [12]). Then, the $a.s.$ convergence of $(C_n)_{n\geq 1}$ follows. (b) One applies the Ruppert and Polyak's Averaging Principle following a version established in [9] (p.169). □

Remark 2. There exists a CLT for the sequence $(Z_n)_{n\geq 1}$ and for its empirical mean $(\overline{Z}_n)_{n\geq 1}$ thanks to the Ruppert and Polyak averaging principle. We only stated the result for the components of interest (the ones which converge to VaR and CVaR).

Now, let us point out an important issue. The IS procedure raises an important problem that can be noticed if we consider (27) and the definition of the two functions K_1 and K_2 in (29). It is due to the fact that basically, we are dealing with rare events to update the VaR and the IS procedures. Somehow, we have two RM procedures $(\xi_n)_{n\geq 1}$ and $(\theta_n, \mu_n)_{n\geq 1}$ that are in competitive conditions, i.e., on one hand, we added an IS procedure to $(\xi_n)_{n\geq 1}$ to improve the convergence toward ξ^*, and on the other hand, the adjustment of the parameters (θ_n, μ_n) are based on samples X_{n+1} satisfying $\varphi(X_{n+1}-\theta_n) > \xi_n$ and $\varphi(X_{n+1}-\mu_n) > \xi_n$ which tend to become rare events. Somehow, we "postponed" the problem resulting from rare events on the IS procedure itself which may "freeze". To circumvent this problem, we are led to slightly modify the IS procedure.

2.2.3 How to Control the Move to the Critical Risk Area

In order to control the growth of θ_n and μ_n at the beginning of the algorithm, since we have no idea on how to twist the distribution of $\varphi(X)$, we can move slowly toward the critical risk area at level α in which $\varphi(X)$ exceeds ξ by replacing α by a deterministic non decreasing sequence α_n that converges to α in (27) and (29). This kind of incremental threshold increase has been proposed previously in [25]. By doing so, we only modify the VaR computation procedure ξ_n. Our aim is to devise an artificial VaR *companion procedure* which will be dedicated to the optimization of the IS parameters and not to the computation of VaR-CVaR. This VaR algorithm will move slowly to the tail distribution of $\varphi(X)$ and hence will drive the IS parameters. In practice, we decide to plug a deterministic stepwise constant sequence, i.e., we set $\alpha_n = 50\%$ for the first 3000-5000 first iterations then we set $\alpha_n = 80\%$ for 3000-5000 first iterations and finally set $\alpha_n = \alpha$ when the sequence (θ_n, μ_n) has almost converged. Numerically speaking, this kind of stepwise growth leads to a stable IS procedure. The function L_1 in (27) now depends of the current step n of the procedure, namely $L_{1,n}(\xi_{n-1}, X_n) = 1 - \frac{1}{1-\alpha_n}\,\mathbf{1}_{\{\varphi(X_n)\geq \xi_{n-1}\}}$ and the VaR_α algorithm ξ_n becomes

$$\hat{\xi}_n = \hat{\xi}_{n-1} - \gamma_n L_{1,n}(\hat{\xi}_{n-1}, X_n), \quad n \geq 1, \quad \hat{\xi}_0 \in L^1(\mathbb{P}). \tag{32}$$

If we replace ξ_n by $\hat{\xi}_n$ into the procedure devised in (27), we obtain a new IS algorithm defined by for $n \geq 1$

$$\begin{cases} \hat{\theta}_n = \hat{\theta}_{n-1} - \gamma_n L_3\left(\hat{\xi}_{n-1}, \hat{\theta}_{n-1}, X_n\right), \quad \theta_0 \in \mathbb{R}^d, \\ \hat{\mu}_n = \hat{\mu}_{n-1} - \gamma_n L_4\left(\hat{\xi}_{n-1}, \hat{\mu}_{n-1}, X_n\right), \quad \mu_0 \in \mathbb{R}^d. \end{cases} \tag{33}$$

To establish the convergence of this new procedure, we rely on the Robbins-Monro Theorem with remainder sequence (see e.g. [12]). Finally, we use (27) to estimate the couple $(\xi^*,\ C^*)$ in which $(\theta_n,\ \mu_n)$ is replaced by $(\hat{\theta}_n,\ \hat{\mu}_n)$. One shows that this new sequence $(\xi_n,\ C_n,\ \hat{\theta}_n,\ \hat{\mu}_n)$ satisfies the same CLT.

Remark 3. When dealing with loss distribution depending on a path-dependent SDE X, one can also replace the IS based on mean translation in a finite dimensional setting by its equivalent based on a Girsanov transformation ([12] and [20]).

2.3 Quasi-Stochastic Approximation

It is rather natural to plug quasi-random numbers instead of pseudo-random numbers in a recursive stochastic approximation. In this framework, the loss distribution $\varphi(X)$ is replaced by $\Psi(U)$ where $U \stackrel{d}{=} U([0,1]^q)$ and $\varphi(X) \stackrel{d}{=} \Psi(U)$. Such an idea goes back to [19] in a one dimensional setting. We denote by F the loss distribution function. We make the following assumption on Ψ,

$$\Psi \text{ is continuous and } \forall\, u \in \mathbb{R},\ \mathbb{P}(\Psi(U) = u) = 0. \tag{34}$$

Let $x := (x_n)_{n\geq 1}$ be a uniformly distributed sequence over $[0,1]^q$ with low discrepancy, i.e., the star discrepancy of the first n terms (see [21]) is $O\left(n^{-1}(\log(n)^q)\right)$. From a theoretical point of view, the convergence of (ξ_n, C_n) can be derived from the weak convergence (see e.g. [5]) of $F_n^{\Psi} := \frac{1}{n}\sum_{k=1}^n \delta_{\Psi(x_k)}$ to F. For any subset A of $\mathbb{R}$, δ_x denotes the *unit mass at* x defined by $\delta_x(A) = \mathbf{1}_A(x)$. We write $D_n^*(x,\Psi)$ for the star discrepancy of the first n terms of x associated to the system and defined naturally by

$$D_n^*(x,\Psi) := \sup_{u\in\mathbb{R}} \left| \frac{1}{n}\sum_{k=1}^n \mathbf{1}_{\Psi(x_k)\leq u} - F(u) \right|.$$

Suppose that (34) is satisfied. Suppose moreover that Ψ is Lipschitz and that the probability density function f of $\Psi(U)$ is bounded. Using theoretical results about Jordan discrepancy (see [21], page 17), one can show that

$$l_n(x) := \max_{1\leq k\leq n} k D_k^*(x,\Psi) = o\left(n^{1-\frac{1}{q}+\epsilon}\right),\quad \forall\, \epsilon > 0.$$

Theorem 2. *Let* $H_1 : \mathbb{R}\times[0,1]^q \to \mathbb{R}$ *and* $H_2 : \mathbb{R}^2\times[0,1]^q \to \mathbb{R}$ *defined by:*

$$H_1(\xi,x) := 1 - \frac{1}{1-\alpha}\mathbf{1}_{\Psi(x)\geq\xi} \quad \text{and } H_2(\xi,C,x) := C - \left(\xi + \frac{1}{1-\alpha}(\Psi(x)-\xi)_+\right).$$

Suppose that condition (34) *is satisfied. Suppose that* Ψ *is Lipschitz and* f *is bounded. Let* $\gamma := (\gamma_n)_{n\geq 1}$ *be a non-increasing deterministic sequence of gain parameters satisfying*

$$\sum \gamma_n = +\infty,\ \ \gamma_n l_n \to 0 \ \text{ and } \ \sum_{n\geq 1} \max\left(\gamma_n - \gamma_{n+1}, \gamma_n^2\right) l_n < +\infty. \tag{35}$$

Then, the recursive procedure defined for every $n \geq 1$ *by*

$$\begin{cases} \xi_n = \xi_{n-1} - \gamma_n H_1(\xi_{n-1}, x_n), \\ C_n = C_{n-1} - \gamma_n H_2(\xi_{n-1}, C_{n-1}, x_n), \end{cases} \tag{36}$$

satisfies:

$$\xi_n \to \xi^* \quad \text{and} \quad C_n \to C^* \ \text{as}\ n \to \infty.$$

Proof. We give a summary of the arguments. Let L be the continuously differentiable Lyapunov function defined for every $x \in \mathbb{R}$ by $L(x) = \sqrt{1+(x-\xi^*)^2}$. A Taylor expansion of L at ξ_n leads to

$$L(\xi_{n+1}) \leq L(\xi_n) - \gamma_{n+1} L'(\xi_n) H_1(\xi^*, x) + C\gamma_{n+1}^2$$

for some positive constant C. Using (35) and successive Abel transforms, one deduces the convergence of $L(\xi_n)$ and then the convergence of ξ_n toward ξ^*. Then, the convergence of C_n can be deduced easily. □

Although, we do not have the rate of convergence of (36), in [19] some *a priori* error bounds (for some specific RM algorithms) show that using low-discrepancy sequences instead of pseudo-random numbers may significantly accelerate the convergence.

3 Numerical Illustrations

The assets are modeled as geometric Brownian Motions. We assume an annual risk free interest rate of $r = 5\%$ and volatility $\sigma = 20\%$.

Example 1. We consider a portfolio composed of a short position in a power plant that produces electricity from gas, day by day with a maturity of $T = 1$ month and 30 long positions in calls on electricity day-ahead prices all with the same strike $K = 60$. Electricity and gas initial spot prices are $S_0^e = 40$ \$/MWh and $S_0^g = 3$ \$/MMBtu (British Thermal Unit) with a Heat Rate equals $h_R = 10$ Btu/kWh and generation costs $C = 5$\$/MWh. The loss can be written

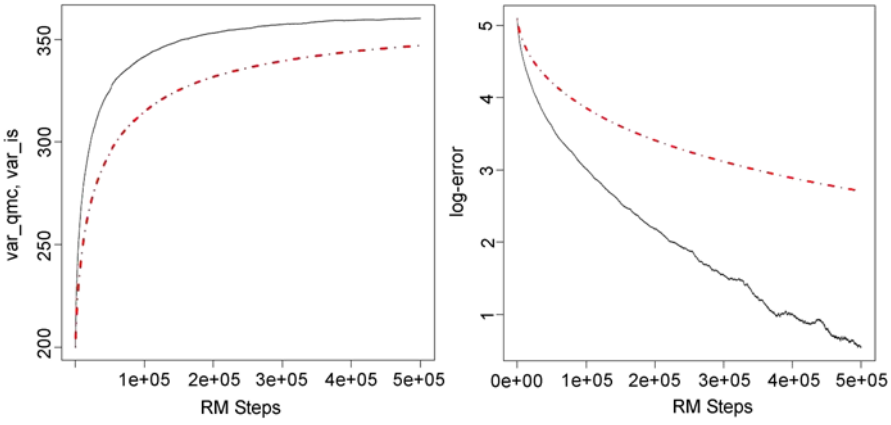
$$\sum_{k=1}^{30} e^{r(T-t_k)} \left(\left(S_{t_k}^e - h_R S_{t_k}^g - C \right)_+ - \left(S_{t_k}^e - K \right)_+ \right) + e^{rT}(C_0 - P_0^c)$$

where $P_0^c \approx 149.9$ is an estimate of the price of the option on the power plant (obtained by Monte Carlo using 500 000 samples) and C_0 is the price of the call options which is equal to 3.8. We use four different values of the confidence level $\alpha = 95\%,\ 99\%,\ 99.5\%$, and 99.9% for this example. In the RM procedure (27), we set the step sequence $\gamma_n = \frac{1}{n^p+100}$, where $p = 3/4$. Choosing the stepsize sequence is an important challenge for the analyst who wishes to apply this method. The numerical results are reported in Table 1. The first column denotes the number of steps used in the Robbins-Monro procedure in (27). The columns VaR and CVaR corresponds to the estimations of the VaR_α and the CVaR_α using (30). The columns VR_{VaR} and VR_{CVaR} denote variance reduction ratios estimations for both VaR and CVaR procedures. Variance Ratios (VR) corresponds to an estimate of the asymptotic variance using (6) divided by an estimate of the asymptotic variance using (27).

Example 2. We consider a portfolio composed of short positions in 10 calls and 10 puts on each of 5 underlying assets, all options having the same maturity of 0.25 years. The strikes are set to 130 for calls, to 110 for puts and the initial spot prices to 120. The underlying assets are assumed to be uncorrelated. We only consider the confidence level $\alpha = 95\%$. We compare the performance of *quasi-stochastic approximation* using a Sobol sequence (see e.g. [21]) and stochastic approximation for the computation of the VaR_α with IS using pseudo-random numbers. The dimension d of the structural vector X is equal to 5. Note that Ψ is not continuous in this example. In the RM procedure, we merely set $\gamma_n = \frac{1}{n+100}$ in (36). The numerical results are summarized in the figure below. The graph on the left depicts the VaR estimations using the procedure (27) with pseudo-random numbers (dotted line) and using (36) (normal line) for several number of RM steps. A good approximation of VaR_α is 362. The graph on the right depicts the log-distance of the VaR estimates to this approximation using pseudo-random numbers (dotted line) and using a Sobol sequence (normal line).

Table 1 Example 1. VaR, CVaR using IS procedure. VR_{VaR}, VR_{CVaR}: variance reduction ratios.

Number of steps	α	VaR	CVaR	VR_{VaR}	VR_{CVaR}
10 000	95 %	115.7	150.5	3.4	6.8
	99 %	169.4	196.0	8.4	12.9
	99.5%	186.3	213.2	13.5	20.3
	99.9%	190.2	219.3	15.3	32.1
100 000	95 %	118.7	150.5	4.5	8.7
	99 %	169.4	195.4	12.6	17.5
	99.5%	188.8	212.9	15.6	29.5
	99.9%	197.4	217.4	21.3	35.5
500 000	95 %	119.2	150.4	5.0	9.2
	99 %	169.8	195.7	13.1	18.6
	99.5%	188.7	212.8	17.0	29.0
	99.9%	198.8	216.8	24.8	46.8



References

1. Arouna, B.: *Adaptive Monte Carlo method, a variance reduction technique*. Monte Carlo Methods and Applications, **10**(1), 1–24 (2004)
2. Artzner, P., Delbaen, F., : Eber, J-M., Heath, D.: *Coherent measures of risk*. Mathematical Finance, **9**(3), 203–228 (1999)
3. Bardou, O., Frikha, N., Pagès, G.: Computation of VaR and CVaR using Stochastic Approximations and Adaptive Importance Sampling. Preprint, LPMA-1263 (2008)
4. Benveniste, A., Metivier, M., Priouret, P.: Adaptative Algorithms and Stochastic Approximations. Springer, Berlin 1990
5. Billingsley, P.: Convergence of Probability Measures. Wiley-Interscience, 1999, p. 296
6. Britten-Jones, M., Schaefer, S.M.: *Non linear Value-at-Risk*. European Finance Review, **5**, 161–187 (1999)
7. Duffie, D., Pan, J.: *An Overview of Value-at-Risk*. Journal of Derivatives, **4**, 7–49 (1997)
8. Duflo, M.: Iterative Random Models. Springer-Verlag, 1997
9. Duflo, M.: Algorithmes Stochastiques. Springer, Berlin, 1996
10. Egloff, D., Leippold, M.: Quantile estimation with adaptive importance sampling, Electronic copy: http://ssrn.com/abstract=1002631, (2007)
11. Dufresne, D., Vázquez-Abad, F.J.: Accelerated simulation for pricing Asian options, Proceedings of the 1998 Winter Simulation Conference, 1493–1500 (1998)
12. Frikha, N., Phd Thesis, In progress
13. Fu, M.C., Su, Y.: *Optimal importance sampling in securities pricing*. Journal of Computational Finance, **5**(4), 27–50 (2000)
14. Glasserman, P., Heidelberger, P., Shahabuddin, P.: Portfolio Value-at-Risk with Heavy-Tailed Risk Factors, Mathematical Finance, **12**, 239–270 (2002)
15. Glasserman, P., Heidelberger, P., Shahabuddin, P.: Variance Reduction Techniques for Estimating Value-at-Risk, SIAM Journal on Numerical Analysis, **47**, 1349–1364 (2000)
16. Glynn, P.W., Iglehart, D.L.: *Importance sampling for stochastic simulations*. Management Science, **35**, 1367–1389 (1989)
17. Juditsky, A.B., Polyak, B.T.: Acceleration of stochastic approximation by averaging, SIAM Journal on Control and Optimization, **30**(4), 838–855 (1992)
18. Kawai, R.: *Optimal importance sampling parameter search for Lévy Processes via stochastic approximation*. SIAM Journal on Numerical Analysis, **47**(1), 293–307 (2008)
19. Lapeyre, B., Pagès, G., Sab, K.: Sequences with Low Discrepancy. Generalization and application to Robbins-Monro algorithm. Statistics, **21**(2) 251–272 (1990)
20. Lemaire, V., Pagès, G.: Unconstrained Recursive Importance Sampling. Preprint LPMA-1231, To appear in Annals of Applied Probability (2008)
21. Niederreiter, H.: Random Number Generation and Quasi-Monte Carlo Methods. SIAM CBMS-NSF Regional Conference Series in Applied Mathematics, **63**, SIAM, Philadelphia, PA, 1992
22. Pflug, G.Ch.: Some remarks on the Value-at-Risk and the Conditional Value-at-Risk. Kluwer Academic Publishers, Dordrecht, 2000
23. Rockafellar, R.T., Uryasev, S.: *Optimization of Conditional Value-At-Risk*. The Journal of Risk, **2**(3), 21–41 (2000)
24. Rouvinez, C.: *Going Greek with VaR*. Risk, **10**(2), 57–65 (1997)
25. Rubinstein, R.Y.: Simulation-based optimization with rare events. European Journal of Operations Reasearch, 89–112 (1997)
26. Ruppert, D.: Stochastic Approximation. B.K. Ghosh and P.K. Sen, eds, New York, 1991
27. Serfling, R.J.: Approximation Theorems for Mathematical Statistics, Wiley, New York, 1980