HAL open science

Computation of VaR and CVaR using stochastic approximations and unconstrained importance sampling.

Olivier Aj Bardou, Noufel Frikha, G. Pagès

▶ To cite this version:

Olivier Aj Bardou, Noufel Frikha, G. Pagès. Computation of VaR and CVaR using stochastic approximations and unconstrained importance sampling.. 2008. hal-00348098v4

HAL Id: hal-00348098

https://hal.science/hal-00348098v4

Preprint submitted on 6 Nov 2009 (v4), last revised 3 Dec 2010 (v5)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Computing VaR and CVaR using Stochastic Approximation and Adaptive Unconstrained Importance Sampling

O. Bardou[1] , N. Frikha[2] , G. Pagès[3]

November 6, 2009

Abstract

Value-at-Risk (VaR) and Conditional-Value-at-Risk (CVaR) are two risk measures which are widely used in the practice of 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. Convergence rate of this algorithm to its target 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 [27]. Finally, to speed up the initialization phase of the IS algorithm, we replace the original confidence level of the VaR by a slowly 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 its efficiency is illustrated on several typical energy portfolios.

Keywords: *VaR, CVaR, Stochastic Approximation, Robbins-Monro algorithm, Importance Sampling, Girsanov.*

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 always be associated to efficient reduction variances techniques to encompass its slow convergence rate. In some specific cases, Gaussian approximations can lead to semi-closed form estimators. But, if these approximations can be of some interest when considering the *yield* of a portfolio, they turn out to be useless when estimating e.g. the VaR on the EBITDA (Earnings Before Interest, Taxes, Depreciation, and Amortization) of a huge portfolio as it is often the case in the energy sector.

In this article, we introduce an alternative estimation method to estmate both VaR and CVaR, relying on the use of recursive stochastic algorithms. By definition, the VaR at level $\alpha \in (0,1)$

[1]Laboratoire de Probabilités et Modèles aléatoires, UMR 7599, Université Pierre et Marie Curie and GDF Suez Research and Innovation Department, France, e-mail: olivier-aj.bardou@gdfsuez.com

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

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

(VaR_α) of a given portfolio is the lowest amount not exceeded by the loss with probability α (usually $\alpha \geq 95\%$). The Conditional Value-at-Risk at level α (CVaR_α) is the conditional expectation of the portfolio losses beyond the VaR_α level. 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. The historical simulation method usually assumes that the asset returns in the future are independent and identically distributed, having the same distribution as they had in the past. Over a time interval $[t, T]$, the loss is defined by $L := V(S_t, t) - V(S_t + \Delta S, T)$, where S_t denotes the market price vector observed at time t, $\Delta S = S_T - S_t$ the variation of S over the time interval $[t, T]$ -which can be calculated using historical data- and $V(S_t, t)$ the portfolio value at time t. The distribution of this loss L can be computed with the corresponding VaR at a given probability level by the inversion of the empirical function method. However, when the market price dynamics follow a general diffusion process solution of a stochastic differential equation (SDE), the assumption of asset returns independence is no longer available.

To circumvent this problem, Monte Carlo simulation tools are generally used. Another widely used method relies on a linear (Normal approximation) or quadratic expansion (Delta-Gamma approximation) and assume a joint normal (or log-normal) distribution for ΔS. The Normal approximation method gives L a normal distribution, thus the computation of the VaR_α is straightforward. However, when there is a non-linear dependence between the portfolio value and the prices of the underlying assets (think of a portfolio with options) such approximation is no longer acceptable. The Delta-Gamma approximation tries to capture some non linearity by adding a quadratic term in the loss expansion. Then, it is possible to find the distribution of the resulting approximation in order to obtain an approximation of the VaR. For more details about these methods, we refer to [7], [8], [15], [16] and [34]. Such approximations are no longer acceptable when considering portfolios with long maturity ($T - t = 1$ year up to 10 years) or when the loss is a functional of a general path-dependent SDE.

In the context of hedging or optimizing a portfolio of financial instruments to reduce the CVaR, it is shown in [33] 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 they address a different problem, this method can be used to compute both VaR and CVaR. The advantage of such a method is that it is possible to estimate both VaR and CVaR simultaneously without assuming that the market prices have a specified distribution (e.g. normal, 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 no limited by the number of generated sample paths used in the procedure.

The idea to compute both VaR and CVaR with one procedure comes from the fact that they are strongly linked as they appear as the solutions and the value of the same convex optimisation problem (see Proposition 2.1) as pointed out [33]. Moreover both the objective function of the minimization problem and its gradient read as an expectation. This leads us to define consistent and asymptotically normal estimators of both quantities as the limit of a global Robbins-Monro (RM) procedure. Consequently, we are no longer constrained by the number of samples paths used in the estimation.

A significant advantage of this recursive approach, especially in regard to the inversion of the empirical function method is that we only estimate the quantities of interest and not the whole inverse of the distribution function. Furthermore, we do not need to make approximations of the loss or of the convex optimization problem to be solved. Moreover, the implementation of the algorithm is straightforward. However to make it really efficient we need to modify it owing to

the fact that VaR and CVaR computation is closely related to the simulation of rare events. That is why as a necessary improvement, we introduce a (recursive and adaptive) variance reduction method based on an importance sampling (IS) paradigm.

Let us be a bit more specific. Basically in this kind of problem we are interested in events that are observed with a very low probability (usually less that 5%, 1% or even 0.1%) so that we obtain few significant replications to update our estimates. Actually, interesting losses are those that exceed the VaR, *i.e.* the ones that are "in the tail" of the loss distribution. Thus in order to compute more accurate estimates of both quantities of interest, it is necessary to generate more samples in the tail of L, the area of interest. A general tool used in this situation is IS.

The basic principle of IS is to modify the distribution of L 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 of IS is to find a right change of measure (among a parameterized family) that will induce a significant variance reduction. In [16] and [17], a change of measure based on a large deviation upper bound is proposed to estimate the loss probability $\mathbb{P}(L > x)$ for several values of x. Then, it is possible to estimate the VaR by interpolating between the estimated loss probabilities.

Although this approach provides an asymptotically optimal IS distribution, it is strongly based on the fact that the Delta-Gamma approximation holds exactly and relies on the assumption that, conditionally to the past data, market moves are normally distributed. Moreover, as shown in [18], importance sampling estimators based on a large deviations change of measure can have variance that increases with the rarity of the event, and even infinite variance. In [12], the VaR_α is estimated by using a quantile estimator based on the inversion of the empirical weighted function and combined with Robbins-Monro (RM) algorithm with repeated projection devised to produce the optimal measure change for IS purpose. This kind of IS algorithm is known to converge toward the optimal importance sampling parameter only after a (long) stabilization phase and provided that the compact sets have been appropriately specified. By contrast, our parameters are optimized by an adaptive unconstrained (*i.e.* without projections) RM algorithm naturally combined with our VaR-CVaR procedure.

One major issue that arises when combining the VaR-CVaR algorithm with the recursive IS procedure is to ensure that the IS 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 slowly increase from a low level (say 50%) to α by introducing a deterministic sequence $(\alpha_n)_{n\geq 0}$ of confidence level that converges toward α. This kind of incremental threshold increase has been proposed previously [22] in a different framework (use of cross entropy in rare event simulation). It speeds up the initialization phase of the IS algorithm and consequently improves the variance reduction. Thus, we can truly experiment asymptotic convergence results in practice.

The paper is organized as follows. In the next section, we present some theoretical results about VaR and CVaR. We introduce the VaR-CVaR stochastic algorithm in its first and naive version and study its convergence rate. We also introduce some background about IS using stochastic approximation algorithm. Section 3 is devoted to the design of an optimal procedure using an adaptive variance reduction procedure. We present how it modifies the asymptotic variance of our first CLT. In Section 4 we provide some extensions to the exponential change of measure and to deal with the case of infinite dimensional setting. Section 5 is dedicated to numerical examples. We propose several portfolios of options on several assets in order to challenge the algorithm and display variance reduction factors obtained using the IS procedure. To prevent the freezing of the algorithm during the first iterations of the IS procedure, we also consider a deterministic moving risk level α_n which replace α to speed up the initialization phase and improve the reduction of variance. We prove theoretically that modifying in this way the algorithm doesn't change the

previous CLT and fasten the convergence.

Notations: • $|.|$ will denote the canonical Euclidean norm on $\mathbb{R}^d$ and $\langle .,. \rangle$ will denote the canonical inner product.
• $\xrightarrow{\mathcal{L}}$ will denote the convergence in distribution and $\xrightarrow{a.s.}$ will denote the almost sure convergence.
• $x_+ := \max(0, x)$ will denote the positive part function.

2 VaR, CVaR using stochastic approximation and some background on recursive IS

It is rather natural to 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 $(\Omega, A, \mathbb{P})$ and $\varphi : \mathbb{R}^d \to \mathbb{R}$ is a Borel function. φ is the function representing the composition of the portfolio which remains fixed and X is a structural random vector used to model the market prices over the time interval; therefore we do not need to specify the dynamics of the market prices and only rely on the fact that it is possible to sample from the distribution of X. For instance, in a Black-Scholes framework, X is a Gaussian vector and φ can be a portfolio of vanilla options. In more sophisticated models or portfolio, X can be 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 $\varphi(X)$ *i.e.*:

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

Since $\lim_{\xi \to +\infty} \mathbb{P}(\varphi(X) \leq \xi) = 1$, we have $\{\xi \mid \mathbb{P}(\varphi(X) \leq \xi) \geq \alpha\} \neq \varnothing$. Moreover, we have $\lim_{\xi \to -\infty} \mathbb{P}(\varphi(X) \leq \xi) = 0$, which implies that $\{\xi \mid \mathbb{P}(\varphi(X) \leq \xi) \geq \alpha\}$ is bounded from below so that the VaR always exists. We assume that the distribution function of $\varphi(X)$ is continuous (*i.e.* without atoms) so that the VaR is the lowest solution of the equation:

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

Three values of α are commonly considered: 0.95, 0.99, 0.995 so that it is usually close to 1 and the tail of interest has probability $1 - \alpha$. If the distribution function is (strictly) increasing, the solution of the above equation is unique, otherwise, there may be more than one solution. In fact, in what follows, we will consider that *any* solution of the previous equation is the VaR. 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). As soon as $\varphi(X) \in L^1(\mathbb{P})$, it is defined by:

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

The CVaR of $\varphi(X)$ is simply the conditional expectation of $\varphi(X)$ given that it lies inside the critical risk area. To capture more information on the conditional distribution of $\varphi(X)$, it seems natural to consider more general risk measures like for example the conditional variance. In a more general framework we can be interested in estimating the Ψ-*Conditional Value at Risk* (Ψ-CVaR) (at level α) where $\Psi : \mathbb{R} \to \mathbb{R}$ is a continuous function. As soon as $\Psi(\varphi(X)) \in L^1(\mathbb{P})$, it is defined by:

$$\Psi\text{-}\mathrm{CVaR}_\alpha(\varphi(X)) := \mathbb{E}[\Psi(\varphi(X)) | \varphi(X) \geq VaR_\alpha(\varphi(X))]. \tag{1}$$

When $\Psi \equiv Id$ and $\varphi(X) \in L^1(\mathbb{P})$, (1) is the regular CVaR of $\varphi(X)$. When $\Psi \equiv x \mapsto x^2$, equation (1) is but the conditional quadratic norm of $\varphi(X)$.

2.1 Representation of VaR and Ψ-$CVaR$ as expectations

The idea to devise a stochastic approximation procedure to compute VaR and CVaR, and more generally the Ψ-CVaR, comes from the fact that these two quantities are solutions of a convex optimization problem whose value function can be represented as an expectation as pointed out by Rockafellar and Uryasev in [32].

Proposition 2.1. *Let V and V_Ψ be the functions defined by:*

$$V(\xi) = \mathbb{E}\left[v(\xi, X)\right] \quad \textit{and} \quad V_\Psi(\xi) = \mathbb{E}\left[w(\xi, X)\right] \tag{2}$$

where

$$v(\xi, x) := \xi + \frac{1}{1-\alpha}(\varphi(X) - \xi)_+ \quad \textit{and} \quad w(\xi, x) := \Psi(\xi) + \frac{1}{1-\alpha}(\Psi(\varphi(x)) - \Psi(\xi))\mathbf{1}_{\{\varphi(x)\geq\xi\}}. \tag{3}$$

Suppose that the distribution function of $\varphi(X)$ is continuous and that $\varphi(X) \in L^1(\mathbb{P})$. Then, the function V is convex, differentiable and the $\mathrm{VaR}_\alpha(\varphi(X))$ is any point of the set:

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

where V' is the derivative of V defined for every $\xi \in \mathbb{R}$ by

$$V'(\xi) = \mathbb{E}\left[\frac{\partial v}{\partial \xi}(\xi, X)\right]. \tag{4}$$

Furthermore,

$$\mathrm{CVaR}_\alpha(\varphi(X)) = \min_{\xi\in\mathbb{R}} V(\xi)$$

and, if Ψ is continuous and that $\Psi(\varphi(X)) \in L^1(\mathbb{P})$, for every $\xi^* \in \arg\min V$ (i.e., ξ^* is a $\mathrm{VaR}_\alpha(\varphi(X))$)

$$\Psi\text{-}\mathrm{CVaR}_\alpha(\varphi(X)) = V_\Psi(\xi^*).$$

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

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

Thanks to Lebesgue Dominated Convergence Theorem, one can interchange differentiation and expectation, so that V is differentiable with derivative $V'(\xi) = 1 - \frac{1}{1-\alpha}\mathbb{P}(\varphi(X) > \xi)$ 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, it is clear that:

$$\begin{aligned} V(\xi^*) &= \xi^* + \frac{\mathbb{E}[(\varphi(X) - \xi^*)_+]}{\mathbb{P}(\varphi(X) > \xi^*)} \\ &= \frac{\xi^*\mathbb{E}[\mathbf{1}_{\varphi(X)>\xi^*}] + \mathbb{E}[(\varphi(X) - \xi^*)_+]}{\mathbb{P}(\varphi(X) > \xi^*)} \\ &= \mathbb{E}\left[\varphi(X)|\varphi(X) > \xi^*\right] \end{aligned}$$

and, in the same way, $V_\Psi(\xi^*) = \Psi\text{-}\mathrm{CVaR}_\alpha(\varphi(X))$. This completes the proof. □

Remark: Actually, one could consider a more general framework by including any risk measure defined by an integral representation with respect to X:

$$\mathbb{E}[\Lambda(\xi^*, X)]$$

where Λ is a (computable) Borel function.

2.2 Stochastic gradient and its adaptive companion procedure: a first naive approach

The above representation (4) naturally yields a stochastic gradient procedure derived from the convex Lyapunov function V which will (hopefully) converge toward $\xi^* := VaR_\alpha(\varphi(X))$. Then, a recursive companion procedure based on (2) can be easily devised having $C^* := \Psi\text{-CVaR}_\alpha(\varphi(X))$ as target. There is no reason to believe that this first version can do better than the empirical quantile estimate. But, it is a necessary phase in order to understand how our recursive IS algorithm (to be devised further on) can be combined with this first procedure.
First we set

$$H_1(\xi, x) := \frac{\partial v}{\partial \xi}(\xi, x) = 1 - \frac{1}{1-\alpha}\mathbf{1}_{\{\varphi(x)\geq\xi\}}, \tag{5}$$

so that,

$$V'(\xi) = \mathbb{E}\left[H_1\left(\xi, X\right)\right].$$

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

$$\xi_n = \xi_{n-1} - \gamma_n H_1(\xi_{n-1}, X_n), n \geq 1, \quad \xi_0 \in L^1(\mathbb{P}), \tag{6}$$

where $(X_n)_{n\geq1}$ 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 $(\gamma_n)_{n\geq1}$ is a deterministic step sequence (decreasing to 0) satisfying:

$$\sum_{n\geq1}\gamma_n = +\infty \quad \text{and} \quad \sum_{n\geq1}\gamma_n^2 < +\infty. \tag{A1}$$

In order to derive the *a.s.* convergence of (6) we introduce the following additional assumption on the distributions of $\varphi(X)$ and $\Psi(\varphi(X))$. Let $a > 0$,

$$\varphi(X) \text{ has a continuous distribution function and } \Psi(\varphi(X)) \in L^{2a}(\mathbb{P}). \tag{$A2)_a$}$$

Actually, Equation (6) can be seen either as a regular RM procedure with mean function V' since it is increasing (see e.g. [10] p.50 and p.66) or as a recursive gradient descent procedure derived from the Lyapunov function V. Both settings yield the *a.s.* convergence toward its target ξ^*. To establish the *a.s.* convergence of $(\xi_n)_{n\geq1}$ (and of our different RM algorithms), we will rely on the following theorem. For a proof of this slight extension of Robbins-Monro Theorem and of the *a.s.* convergence of $(\xi_n)_{n\geq1}$ (under assumptions $(A1)$ and $(A2)_1$), we refer to [13].

Theorem 2.2. *(Robbins-Monro Theorem (variant)). Let $H : \mathbb{R}^q \times \mathbb{R}^d \to \mathbb{R}^d$ be a Borel function and X be an $\mathbb{R}^d$-valued random vector such that $\mathbb{E}[|H(z, X)|] < \infty$ for every $z \in \mathbb{R}^d$. Then set*

$$\forall z \in \mathbb{R}^d, \quad h(z) = \mathbb{E}[H(z, X)].$$

Suppose that the function h is continuous and that $\mathcal{T}^* := \{h = 0\}$ satisfies

$$\forall z \in \mathbb{R}^d \setminus \mathcal{T}^*, \forall z^* \in \mathcal{T}^*, \quad \langle z - z^*, h(z)\rangle > 0. \tag{7}$$

Let $(\gamma_n)_{n\geq 1}$ be a deterministic step sequence satisfying condition (A1). Suppose that

$$\forall z \in \mathbb{R}^d, \quad \mathbb{E}[|H(z,X)|^2] \leq C(1+|z|^2) \tag{8}$$

(which implies that $|h(z)| \leq C'(1+|z|)$).
Let $(X_n)_{n\geq 1}$ be an i.i.d. sequence of random vectors having the distribution of X, let z_0 be a random vector independent of $(X_n)_{n\geq 1}$ satisfying $\mathbb{E}[|z_0|] < \infty$, all defined on the same probability space $(\Omega, A, \mathbb{P})$. Let $\mathcal{F}_n := \sigma(z_0, X_1, ..., X_n)$ and let $(r_n)_{n\geq 1}$ be an $\mathcal{F}_n$-measurable remainder sequence satisfying

$$\sum_n \gamma_n |r_n|^2 < \infty. \tag{9}$$

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

$$Z_n = Z_{n-1} - \gamma_n H(Z_{n-1}, X_n) + \gamma_n r_n,$$

satisfies:

$$\exists\, z_\infty, \quad \text{such that } \; Z_n \xrightarrow{a.s.} z_\infty \text{ and } z_\infty \in \mathcal{T}^* \text{ a.s.}$$

The convergence also holds in $L^p(\mathbb{P}), p \in (0,2)$, where $L^p(\mathbb{P})$ denotes the set of all random vectors defined on $(\Omega, A, \mathbb{P})$ such that $\mathbb{E}[|X|^p]^{\frac{1}{p}} < \infty$.

Remark: It is in fact a slight variant (see e.g. [13]) of the regular RM Theorem since Z_n converges to a random vector having its value in the set $\{h = 0\}$ even if $\{h = 0\}$ is not reduced to a singleton or a finite set. The remainder sequence in the above theorem plays a crucial role when we will (slightly) modify the first IS procedure to improve its efficiency.

The second step concerns procedure for the numerical computation of the Ψ-CVaR$_\alpha$. A naive idea is to compute the function V_Ψ at the point ξ^*:

$$\Psi\text{-CVaR}_\alpha = V_\Psi(\xi^*) = \mathbb{E}[w(\xi^*, X)]$$

using a regular Monte Carlo simulation,

$$\frac{1}{n}\sum_{k=0}^{n-1} w(\xi^*, X_{k+1}). \tag{10}$$

However, we first need to get from (6) a good approximate of ξ^* and subsequently to use another sample of the distribution X. A natural idea is to devise an adaptive *companion procedure* of the above quantile search algorithm by replacing ξ^* in (10) by its approximation at step k, namely

$$C_n = \frac{1}{n}\sum_{k=0}^{n-1} w(\xi_k, X_{k+1}), \; n \geq 1, \; C_0 = 0. \tag{11}$$

Hence, $(C_n)_{n\geq 0}$ is the sequence of empirical means of the non i.i.d. sequence $(w(\xi_k, X_{k+1}))_{k\geq 1}$, which can be written recursively:

$$C_n = C_{n-1} - \frac{1}{n} H_2\left(\xi_{n-1}, C_{n-1}, X_n\right), \; n \geq 1, \tag{12}$$

where $H_2(\xi, c, x) := c - w(\xi, x)$.

At this stage, we are facing two procedures (ξ_n, C_n) with different steps. This may appear not very consistent or at least natural. A second modification to the original Monte Carlo procedure (12)

consists in considering a general step β_n satisfying condition (A1) instead of $\frac{1}{n}$ (with in mind the possibility to set $\beta_n = \gamma_n$ eventually). This leads to:

$$C_n = C_{n-1} - \beta_n H_2\left(\xi_{n-1}, C_{n-1}, X_n\right), \ n \geq 1. \tag{13}$$

In order to prove the *a.s.* convergence of $(C_n)_{n\geq 1}$ toward C^*, we set for convenience $\beta_0 := \sup_{n\geq 1} \beta_n + 1$. Then, one defines recursively a sequence $(\Delta_n)_{n\geq 1}$ by

$$\Delta_{n+1} = \Delta_n \frac{\beta_{n+1}}{\beta_n} \frac{\beta_0}{\beta_0 - \beta_{n+1}}, \ n \geq 0, \ \Delta_0 = 1.$$

Elementary computations show by induction that

$$\beta_n = \beta_0 \frac{\Delta_n}{S_n}, \ n \geq 0, \ \text{with } S_n = \sum_{k=0}^{n} \Delta_k. \tag{14}$$

Furthermore, it follows from (14) that for every $n \geq 1$

$$\log(S_n) - \log(S_{n-1}) = -\log\left(1 - \frac{\Delta_n}{S_n}\right) \geq \frac{\Delta_n}{S_n} = \frac{\beta_n}{\beta_0}.$$

Consequently,

$$\log(S_n) \geq \frac{1}{\beta_0} \sum_{k=1}^{n} \beta_k$$

which implies that $\lim_n S_n = +\infty$.
Now using (13) and (14), one gets for every $n \geq 1$

$$S_n C_n = S_{n-1} C_{n-1} + \Delta_n \left(\Delta N_{n+1} + V_\Psi(\xi_n)\right)$$

where, $\Delta N_n := w(\xi_{n-1}, X_n) - V_\Psi(\xi_{n-1})$, $n \geq 1$, define a martingale increments sequence with respect to the natural filtration of the algorithm $\mathcal{F}_n := \sigma(\xi_0, X_1, \cdots, X_n)$, $n \geq 0$. Consequently,

$$C_n = \frac{1}{S_n} \left(\sum_{k=0}^{n-1} \Delta_{k+1} \Delta N_{k+1} + \sum_{k=0}^{n-1} \Delta_{k+1} V_\Psi(\xi_k) \right).$$

The second term in the right hand side of the above equality converges to $V_\Psi(\xi^*) = \Psi\text{-CVaR}_\alpha(\varphi(X))$ owing to the continuity of V_Ψ at ξ^* and Cesaro's Lemma.
The convergence to 0 of the first term will follow from the *a.s.* convergence of the series

$$N_n^\beta := \sum_{k=1}^{n} \beta_k \Delta N_k, \ n \geq 1$$

by the Kronecker Lemma since $\beta_n = \beta_0 \frac{\Delta_n}{S_n}$. The sequence $(N_n^\beta)_{n\geq 1}$ is an $\mathcal{F}_n$-martingale since the ΔN_k's are martingale increments and

$$\mathbb{E}\left[(\Delta N_n)^2 | \mathcal{F}_{n-1}\right] \leq \frac{1}{(1-\alpha)^2} \mathbb{E}\left[\left(\Psi\left(\varphi\left(X\right)\right) - \Psi\left(\xi\right)\right)^2\right]_{|\xi = \xi_{n-1}}.$$

The continuity of Ψ at ξ^*, assumption $(A2)_1$ and the *a.s.* convergence of ξ_k toward ξ^* imply that

$$\sup_{n\geq 1} \mathbb{E}[(\Delta N_n)^2 | \mathcal{F}_{n-1}] < \infty \quad a.s.$$

Consequently, assumption (A1) implies

$$\langle N^\beta \rangle_\infty = \sum_{n\geq 1} \beta_n^2 \mathbb{E}[(\Delta N_n)^2 | \mathcal{F}_{n-1}] < \infty$$

which in term yields the $a.s.$ convergence of $(N_n^\beta)_{n\geq 1}$, so that $C_n \xrightarrow{a.s.} \Psi\text{-}CVaR_\alpha(\varphi(X))$.
The resulting algorithm reads as for $n \geq 1$:

$$\begin{cases} \xi_n = \xi_{n-1} - \gamma_n H_1(\xi_{n-1}, X_n), \quad \xi_0 \in L^1(\mathbb{P}), \\ C_n = C_{n-1} - \beta_n H_2(\xi_{n-1}, C_{n-1}, X_n), \quad C_0 = 0, \end{cases} \tag{15}$$

and converges under $(A1)$ and $(A2)_1$.

The question of the joint weak convergence rate of (ξ_n, C_n) is not trivial owing to the coupling of the two procedures. The case of two different step scales refers to the general framework of two-time-scale stochastic approximation algorithms. Several results have been established by Borkar in [5], Konda and Tsitsiklis in [21] but the more relevant in our case are those of Mokkadem and Pelletier in [30]. The weak convergence rate of $(\xi_n)_{n\geq 1}$ is ruled by the CLT for "regular" (single-time scale) stochastic approximation algorithms (we refer to Kushner and Clark in [23], Métivier and Priouret in [4], Duflo in [10] among others). In order to achieve the best asymptotic rate of convergence, one ought to set $\gamma_n = \frac{\gamma_0}{n}$ where the choice of γ_0 depends on the value of the density $f_{\varphi(X)}$ of $\varphi(X)$ at ξ^*, which is unknown. To circumvent the difficulties induced by the specification of γ_0, which are classical in this field, we are led to modify again our algorithm by introducing the averaging principle independently introduced by Ruppert [35] and Polyak [19] and then widely investigated by several authors. It works both with two-time or single-time scale steps and leads to asymptotically efficient procedures, *i.e.*, satisfying a CLT at the optimal rate $\sqrt{n}$ and minimal variance (see also [30]). See also a variant based on a gliding window developed in [26]. Our numerical examples indicate that the averaged one-time-scale procedure provides less variance during the first iterations than the averaged procedure of the two-time-scale algorithm. Finally, we set $\gamma_n \equiv \beta_n$ in (15) so that, the VaR-CVaR algorithm can be written in a more synthetic way by setting $Z_n = (\xi_n, C_n)$ and for $n \geq 1$:

$$Z_n = Z_{n-1} - \gamma_n H(Z_{n-1}, X_n), \; Z_0 = (\xi_0, C_0), \; \xi_0 \in L^1(\mathbb{P}), \tag{16}$$

where $H(z, x) := (H_1(\xi, x), H_2(\xi, C, x))$. Throughout the rest of this section, we assume that the distribution $\varphi(X)$ has a positive probability density $f_{\varphi(X)}$ on its support. As a consequence the $VaR_\alpha(\varphi(X))$ is unique so that the procedure algorithm Z_n converges $a.s.$ to its single target $(\mathrm{VaR}_\alpha(\varphi(X)), \Psi\text{-}\mathrm{CVaR}_\alpha(\varphi(X)))$. Thus, the Cesaro mean of the procedure

$$\bar{Z}_n := \frac{Z_0 + \cdots + Z_{n-1}}{n}, \quad n \geq 1,$$

where Z_n is defined by (16), converges $a.s.$ to the same target. The Ruppert and Polyak's Averaging Principle says that an appropriate choice of the step yields for free the smallest possible asymptotic variance. We recall below this result (following a version established in [10], see [10] (p.169) for a proof).

Theorem 2.3. *(Ruppert and Polyak's Averaging Principle) Suppose that the $\mathbb{R}^d$-sequence $(Z_n)_{n\geq 0}$ is defined recursively by*

$$Z_n = Z_{n-1} - \gamma_n (h(Z_{n-1}) + \epsilon_n + r_n)$$

where h is a Borel function. Suppose that z^* is a zero of h satisfying

$$h(z) = M(z - z^*) + O(|z - z^*|^2) \tag{17}$$

where $M = Dh(z^*)$ is a uniformly repulsive matrix (all its eigenvalues have positive real parts), and $(\epsilon_n)_{n\geq 1}$ is a random sequence satisfying

$$\exists\ C > 0,\ \textit{such that a.s.}\ \begin{cases} (i)\ \mathbb{E}[\epsilon_{n+1}|\mathcal{F}_n]\mathbf{1}_{\{||Z_n - z^*||\leq C\}} = 0, \\ (ii)\ \exists b > 2,\ \sup_n \mathbb{E}[||\epsilon_{n+1}||^b|\mathcal{F}_n]\ \mathbf{1}_{\{||Z_n - z^*||\leq C\}} < +\infty, \\ (iii)\ \mathbb{E}\left[(\gamma_{n-1})^{-1}\left|r_n\right|^2\ \mathbf{1}_{\{||Z_n - z^*||\leq C\}}\right] \to 0, \\ (iv)\ \exists\ \Gamma \in \mathcal{S}^+(d,\mathbb{R})\ \textit{such that}\ \mathbb{E}\left[\epsilon_{n+1}\epsilon_{n+1}^T|\mathcal{F}_n\right] \xrightarrow{a.s.} \Gamma. \end{cases} \tag{18}$$

Set $\gamma_n = \frac{\gamma_1}{n^a}$ with $\frac{1}{2} < a < 1$, and

$$\bar{Z}_{n+1} := \frac{Z_0 + ... + Z_n}{n+1} = \bar{Z}_n - \frac{1}{n+1}(\bar{Z}_n - Z_n),\ n \geq 0.$$

Then, on the set of convergence $\{Z_n \to z^*\}$:

$$\sqrt{n}\left(\bar{Z}_n - z^*\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, M^{-1}\Gamma(M^{-1})^T\right) \qquad \textit{as } n \to +\infty,$$

where $(M^{-1})^T$ denotes the transpose of the matrix M^{-1}.

To apply this theorem to our framework we are led to compute the Cesaro means of both components, namely for $n \geq 1$

$$\begin{cases} \overline{\xi}_n := \frac{1}{n}\sum_{k=1}^n \xi_k = \overline{\xi}_{n-1} - \frac{1}{n}(\overline{\xi}_{n-1} - \xi_n), \\ \overline{C}_n := \frac{1}{n}\sum_{k=1}^n C_k = \overline{C}_{n-1} - \frac{1}{n}(\overline{C}_{n-1} - C_n), \end{cases} \tag{19}$$

where (ξ_k, C_k), $k \geq 0$ is defined by (16). In the following theorem, we provide the convergence rate of the couple $\bar{Z}_n := \left(\overline{\xi}_n, \overline{C}_n\right)$.

Theorem 2.4. *(Convergence rate of the VaR-CVaR procedure). Suppose $(A2)_a$ holds for some $a > 1$, Ψ is twice differentiable at ξ^* and the density function $f_{\varphi(X)}$ of $\varphi(X)$ is continuous, differentiable and strictly positive at ξ^*. If the step sequence is $\gamma_n = \frac{\gamma_1}{n^a}$ with $\frac{1}{2} < a < 1$ and $\gamma_1 > 0$ then*

$$\sqrt{n}\left(\bar{Z}_n - z^*\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \Sigma\right) \qquad \textit{as } n \to +\infty$$

where the asymptotic covariance matrix Σ is given by

$$\begin{pmatrix} \frac{\alpha(1-\alpha)}{f^2_{\varphi(X)}(\xi^*)} & \frac{\alpha}{(1-\alpha)f_{\varphi(X)}(\xi^*)}\mathbb{E}\left[\left(\Psi(\varphi(X)) - \Psi(\xi^*)\right)\mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\right] \\ \frac{\alpha}{(1-\alpha)f_{\varphi(X)}(\xi^*)}\mathbb{E}\left[\left(\Psi(\varphi(X)) - \Psi(\xi^*)\right)\mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\right] & \frac{1}{(1-\alpha)^2}\mathrm{Var}\left(\left(\Psi(\varphi(X)) - \Psi(\xi^*)\right)\mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\right) \end{pmatrix}. \tag{20}$$

Proof. First, the procedure (16) can be written as for $n \geq 1$

$$Z_n = Z_{n-1} - \gamma_n\left(h(Z_{n-1}) + \epsilon_n\right),\quad Z_0 = (\xi_0, C_0),\quad \xi_0 \in L^1(\mathbb{P}), \tag{21}$$

where $h(z) := \mathbb{E}[H(z,X)] = \left(1 - \frac{1}{1-\alpha}\mathbb{P}\left(\varphi(X) \geq \xi\right), C - \mathbb{E}[w(\xi,X)]\right)$ and $\epsilon_n := (\Delta M_n, \Delta N_n)$, $n \geq 1$, denotes the $\mathcal{F}_n$-adapted martingale increment sequence with

$$\Delta M_n := \frac{1}{1-\alpha}\left(\mathbb{P}\left(\varphi(X) \geq \xi\right)_{|\xi=\xi_{n-1}} - \ \mathbf{1}_{\{\varphi(X_n)\geq\xi_{n-1}\}}\right).$$

Owing to Assumption $(A2)_a$, the differentiability of Ψ at ξ^* and Lebesgue's differentiation Theorem, one can interchange expectation and derivation, so that the function h is differentiable at $z^* = (\xi^*, C^*)$ and

$$h'(z^*) = M := \begin{pmatrix} \frac{1}{1-\alpha} f_{\varphi(X)}(\xi^*) & 0 \\ \mathbb{E}\left[\left(\frac{\partial}{\partial \xi} w(\xi, X)\right)_{|\xi=\xi^*}\right] & 1 \end{pmatrix}. \tag{22}$$

Now, since Ψ is differentiable at ξ^*, $\mathbb{E}\left[\left(\frac{\partial}{\partial \xi} w(\xi, X)\right)_{|\xi=\xi^*}\right] = \left(1 - \frac{1}{1-\alpha}\mathbb{P}(\varphi(X) \geq \xi^*)\right)\Psi'(\xi^*) = 0$, so that, $M = \begin{pmatrix} \frac{1}{1-\alpha} f_{\varphi(X)}(\xi^*) & 0 \\ 0 & 1 \end{pmatrix}$ is diagonal. Since Ψ is twice differentiable at ξ^* and $f_{\varphi(X)}$ is differentiable at ξ^*, $D^2h(\xi^*)$ is bounded and (17) is satisfied.
To apply Theorem 2.3, we need to check assumptions (i)-(iv) of (18).
Let $A > 0$. First note that

$$\mathbb{E}\left[\Delta M_{n+1}^{2a} | \mathcal{F}_n\right] \mathbf{1}_{\{|Z_n - z^*| \leq A\}} \leq \left(\frac{1}{1-\alpha}\right)^{2a} 2^{2a} < +\infty.$$

Thanks to Assumption $(A2)_a$, there exists $C_{\alpha,\Psi} > 0$ such that

$$\mathbb{E}\left[\Delta N_{n+1}^{2a} | \mathcal{F}_n\right] \mathbf{1}_{\{||Z_n - z^*|| \leq A\}} \leq C_{\alpha,\Psi}\left(1 + \Psi(\xi_n)^{2a}\right) \mathbf{1}_{\{||Z_n - z^*|| \leq A\}} < C_{\alpha,\Psi}\left(1 + \sup_{|\xi - \xi^*| \leq A} |\Psi(\xi)|^{2a}\right) < +\infty.$$

Consequently, (ii) of (18) holds true with $b = 2a > 2$ since

$$\sup_{n \geq 0} \mathbb{E}\left[|\epsilon_{n+1}|^{2a} | \mathcal{F}_n\right] \mathbf{1}_{\{|Z_n - z^*| \leq A\}} < +\infty.$$

It remains to check (iv) for some positive definite symmetric matrix Γ. The dominated convergence theorem implies that

$$\begin{aligned}
\mathbb{E}\left[\left(\epsilon_{n+1}\epsilon_{n+1}^T\right)_{1,1} | \mathcal{F}_n\right] &= \left(\frac{1}{1-\alpha}\right)^2 \left(\mathbb{E}\left[\ \mathbf{1}_{\{\varphi(X) \geq \xi\}}\right]_{|\xi=\xi_n} - \mathbb{E}\left[\ \mathbf{1}_{\{\varphi(X) \geq \xi\}}\right]^2_{|\xi=\xi_n}\right) \\
&\xrightarrow{a.s.} \frac{\alpha}{1-\alpha}, \\
\mathbb{E}\left[\left(\epsilon_{n+1}\epsilon_{n+1}^T\right)_{1,2} | \mathcal{F}_n\right] &= \mathbb{E}\left[\left(\epsilon_{n+1}\epsilon_{n+1}^T\right)_{2,1} | \mathcal{F}_n\right] \\
&= \left(\frac{1}{1-\alpha}\right)^2 \mathbb{E}\left[\left(\Psi(\varphi(X)) - \Psi(\xi)\right) \mathbf{1}_{\{\varphi(X) \geq \xi\}}\right]_{|\xi=\xi_n} \\
&\qquad \times \left(1 - \mathbb{E}\left[\ \mathbf{1}_{\{\varphi(X) \geq \xi\}}\right]_{|\xi=\xi_n}\right) \\
&\xrightarrow{a.s.} \frac{\alpha}{(1-\alpha)^2} \mathbb{E}\left[\left(\Psi(\varphi(X)) - \Psi(\xi^*)\right) \mathbf{1}_{\{\varphi(X) \geq \xi^*\}}\right],
\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}\left[\left(\epsilon_{n+1}\epsilon_{n+1}^T\right)_{2,2}|\mathcal{F}_n\right] &= \mathbb{E}\left[(\Delta N_{n+1})^2|\mathcal{F}_n\right] \\
&= \frac{1}{(1-\alpha)^2}\Big(\mathbb{E}\left[(\Psi(\varphi(X_{n+1}))-\Psi(\xi))\ \mathbf{1}_{\{\varphi(X_{n+1})\geq\xi\}}|\mathcal{F}_n\right]_{|\xi=\xi_n} \\
&\qquad -\mathbb{E}\left[(\Psi(\varphi(X))-\Psi(\xi))\ \mathbf{1}_{\{\varphi(X)\geq\xi\}}\right]^2_{|\xi=\xi_n}\Big) \\
&\xrightarrow{a.s.} \frac{1}{(1-\alpha)^2}\Big(\mathbb{E}\left[(\Psi(\varphi(X))-\Psi(\xi^*))^2\ \mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\right] \\
&\qquad -\mathbb{E}\left[(\Psi(\varphi(X))-\Psi(\xi^*))\ \mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\right]^2\Big) \\
&= \frac{1}{(1-\alpha)^2}\mathrm{Var}\left((\Psi(\varphi(X))-\Psi(\xi^*))\ \mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\right).
\end{aligned}
$$

Using the continuity of both functions $\xi\mapsto\mathbb{E}\left[(\Psi(\varphi(X))-\Psi(\xi))\ \mathbf{1}_{\{\varphi(X)\geq\xi\}}\right]$ and $\xi\mapsto\mathbb{E}\left[(\Psi(\varphi(X))-\Psi(\xi))^2\ \mathbf{1}_{\{\varphi(X)\geq\xi\}}\right]$ at ξ^*, which follows from the continuity of Ψ and of the distribution function of $\varphi(X)$, finally yields the $a.s.$ convergence of $\mathbb{E}\left[\epsilon_{n+1}\epsilon_{n+1}^T|\mathcal{F}_n\right]$ toward

$$
\Gamma=\begin{pmatrix}\frac{\alpha}{1-\alpha} & \frac{\alpha}{(1-\alpha)^2}\mathbb{E}\left[(\Psi(\varphi(X))-\Psi(\xi^*))\ \mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\right] \\ \frac{\alpha}{(1-\alpha)^2}\mathbb{E}\left[(\Psi(\varphi(X))-\Psi(\xi^*))\ \mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\right] & \frac{1}{(1-\alpha)^2}\mathrm{Var}\left((\Psi(\varphi(X))-\Psi(\xi^*))\ \mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\right)\end{pmatrix}.
$$

If $\gamma_n=\frac{\gamma_1}{n^a}$ with $\gamma_1>0$ and $\frac{1}{2}<a<1$, Ruppert-Polyak's Theorem implies that

$$
\sqrt{n}\left(\bar{Z}_n-z^*\right)\xrightarrow{\mathcal{L}}\mathcal{N}(0,\Sigma)
$$

where $\Sigma=M^{-1}\Gamma\left(M^{-1}\right)^T$ is given by (20). This completes the proof. □

Remarks: • It is possible to replace $w(\xi,x)$ in (13) and (15) by $\tilde{w}(\xi,x)=\frac{1}{1-\alpha}\Psi(\varphi(x))\mathbf{1}_{\{\varphi(x)\geq\xi\}}$ since $C^*=\mathbb{E}\left[\tilde{w}\left(\xi^*,X\right)\right]$. Thus, we only have to change also the martingale increment sequence $(\Delta N_n)_{n\geq1}$ by $\left(\Delta\widetilde{N}_n\right)_{n\geq1}$ defined by

$$
\Delta\widetilde{N}_n:=\frac{1}{1-\alpha}\left(\mathbb{E}\left[\Psi(\varphi(X))\mathbf{1}_{\{\varphi(X)\geq\xi\}}\right]_{|\xi=\xi_{n-1}}-\Psi(\varphi(X_n))\mathbf{1}_{\{\varphi(X_n)\geq\xi_{n-1}\}}\right).
$$

This provides another procedure $\tilde{C}_n$ for the computation of the Ψ-CVaR$_\alpha$ which satisfies a Gaussian CLT with the same asymptotic covariance matrix.

• The quantile estimate based on the inversion of the empirical distribution function satisfies a Gaussian CLT with the same asymptotic covariance matrix than the one of the procedure $\overline{\xi}_n$, see for example [36] p.75. Obviously, there is no reason to believe that this first version can do better than the empirical quantile estimate. However, our quantile estimate has the advantage to be recursive: it naturally combines with a recursive IS algorithm in an adaptive way. In terms of computational complexity, once N loss samples have been generated, the behaviour of the inversion of the empirical distribution function method needs a sorting algorithm: good behaviour is $\mathcal{O}\left(N\log(N)\right)$ element comparisons to sort the list of loss samples. Whereas the behaviour of the recursive quantile algorithm is $\mathcal{O}\left(N\right)$.

• One shows that if we choose $\beta_n=\frac{1}{n}$, $n\geq1$ and $\gamma_n=\frac{1}{n^a}$ with $\frac{1}{2}<a<1$ in (15), the resulting two-time scale procedure satisfies a Gaussian CLT with the same asymptotic covariance matrix Γ (at rates $\sqrt{\gamma_n^{-1}}$ and $\sqrt{n}$). However, by averaging the first component ξ_n, the resulting procedure becomes asymptotically efficient (*i.e.* rate $\sqrt{n}$).

Proposition 2.5. *(Estimation of variance and confidence interval) For every $n \geq 1$, set*

$$\sigma_n^2 := \frac{1}{(1-\alpha)^2}\left(\frac{1}{n}\sum_{k=1}^{n}\left(\Psi(\varphi(X_k)) - \Psi(\xi_{k-1})\right)^2 \mathbf{1}_{\{\varphi(X_k)\geq\xi_{k-1}\}} - \left(\frac{1}{n}\sum_{k=1}^{n}\left(\Psi(\varphi(X_k)) - \Psi(\xi_{k-1})\right)\mathbf{1}_{\{\varphi(X_k)\geq\xi_{k-1}\}}\right)^2\right)$$

where $(\xi_n)_{n\geq 0}$ is the first component of (6). If $(A2)_a$ is satisfied for some $a \geq 2$, then

$$\sigma_n^2 \xrightarrow{a.s.} \frac{1}{(1-\alpha)^2}\ \mathrm{Var}\ \left(\left(\Psi(\varphi(X)) - \Psi(\xi^*)\right)\mathbf{1}_{\varphi(X)\geq\xi^*}\right)$$

and

$$\sqrt{n}\ \frac{C_n - C^*}{\sigma_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1). \tag{23}$$

Proof. The proof follows from standard arguments already used in the proof of the $a.s.$ convergence of the sequence $(C_n)_{n\geq 1}$ defined by (13). □

In practice, the convergence of the algorithm will be chaotic. The bottleneck of this algorithm is that it is only updated on rare events since it tries to measure the tail distribution of $\varphi(X)$: $\mathbb{P}(\varphi(X) > VaR_\alpha) = 1-\alpha \approx 0$. Another problem may be the simulation of $\varphi(X)$. In practice, we have to deal with large portfolios of complex derivative securities and options. Each evaluation may require a lot of computational efforts and takes a long time. So, for practical implementation it is necessary to combine the above procedure with variance reduction techniques to achieve accurate results at a reasonable cost. The most appropriate technique when dealing with rare events is IS.

2.3 Some background on IS using stochastic approximation algorithm

The second tool we want to introduce in this paper is a recursive IS procedure which increases the probability of simulations for which $\varphi(X)$ exceeds ξ. Our goal is to combine it adaptively with our first naive algorithm. Assume that X has an absolutely continuous distribution $\mathbb{P}_X(dx) = p(x)\lambda_d(dx)$ where λ_d denotes the Lebesgue measure on $(\mathbb{R}^d, \mathcal{B}or(\mathbb{R}^d))$. The main idea of importance sampling by translation applied to the computation of

$$\mathbb{E}[F(X)],$$

where $F \in L^2(\mathbb{P}_X)$ satisfies $\mathbb{P}(F(X) \neq 0) > 0$, is to use the invariance of the Lebesgue measure by translation, for every $\theta \in \mathbb{R}^d$,

$$\mathbb{E}[F(X)] = \mathbb{E}\left[F(X+\theta)\frac{p(X+\theta)}{p(X)}\right], \tag{24}$$

and 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

$$Q(\theta) := \mathbb{E}\left[F^2(X+\theta)\frac{p^2(X+\theta)}{p^2(X)}\right] \leq +\infty, \quad \theta \in \mathbb{R}^d. \tag{25}$$

If the following assumption

$$\forall \theta \in \mathbb{R}^d, \qquad \mathbb{E}\left[F^2(X)\frac{p(X)}{p(X-\theta)}\right] < +\infty \tag{B1}$$

holds true, then Q is everywhere finite and a reverse change of variable shows that:

$$Q(\theta) = \mathbb{E}\left[F^2(X)\frac{p(X)}{p(X-\theta)}\right], \quad \theta \in \mathbb{R}^d. \tag{26}$$

Now if p satisfies

$$\begin{cases} (i) \;\; \forall x \in \mathbb{R}^d,\; \theta \mapsto p(x-\theta) \text{ is log-}\textit{concave} \\ \\ (ii) \;\; \forall x \in \mathbb{R}^d,\; \lim_{|\theta|\to+\infty} p(x-\theta) = 0 \quad \text{or } \forall x \in \mathbb{R}^d,\; \lim_{|\theta|\to+\infty} \frac{p(x-\theta)}{p^2(x-\frac{\theta}{2})} = 0, \end{cases} \tag{B2}$$

one shows that Q is (strictly) finite, convex, goes to infinity at infinity so that $\arg\min Q = \{\nabla Q = 0\}$ is non empty (see [1] and [27]). Provided that ∇Q admits a representation as an expectation, then it is possible to devise a recursive RM procedure to approximate the optimal parameter θ^*. Recursive IS by stochastic approximation has been first investigated by Kushner and then by several authors, see e.g. [11] and [14] in order to "optimize" or "improve" the change of measure in IS using a stochastic gradient RM algorithm based on the representation of $\nabla Q(\theta)$. Recently, it has been brought back to light by Arouna (see [1]) in the Gaussian case, based on the natural representation of ∇Q obtained by formally differentiating (26). 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 (26) and not (25). We obtain $\nabla Q(\theta) = \mathbb{E}\left[K(\theta, X)\right]$.

When $X = \mathcal{N}(0,1)$, $Q(\theta) = e^{\frac{|\theta|^2}{2}}\mathbb{E}[F^2(X)e^{-\theta X}]$ so that $K(\theta, x) = e^{\frac{|\theta|^2}{2}}F^2(x)e^{-\theta x}(\theta - x)$. However, given this resulting form of K, the classical convergence results do not apply since $||K(\theta, X)||_2$ is not sub-linear in θ (see condition (8) of Theorem 2.2). This induces the explosion of the procedure at almost every implementation as pointed out in [1]. This leads the author to introduce a "constrained" variant of the regular procedure based on repeated reinitializations known as the projection "à la Chen". It forces the stability of the algorithm and prevents explosion. Let us also mention a first alternative approach investigated in [1] and [3], where Arouna and Bardou change the function to be minimized by introducing an entropy based criterion. Although it is only an approximation, it turns out to be often close to the original method.

Recently, Lemaire and Pagès in [27] revisited the original approach and provided a new representation of $\nabla Q(\theta)$ for which the resulting $K(\theta, X)$ has a linear growth in θ so that all assumptions of Theorem 2.2 are satisfied. Thanks to *a third translation* of the variable θ, it is possible to plug back the parameter θ "into" F, the function F having in common applications a known behaviour at infinity which makes possible to devise a "regular" and "unconstrained" stochastic algorithm. We will rely partially on this approach to devise our final procedure to compute both VaR and CVaR. To be more specific about the methodology proposed in [27], we introduce the following assumption on the probability density p of X

$$\exists b \in [1,2] \text{ such that } \begin{cases} (i) \;\; \frac{|\nabla p(x)|}{p(x)} = O(|x|^{b-1}) \quad \text{as} \quad |x| \to \infty \\ \\ (ii) \;\; \exists \rho > 0, \log\left(p(x)\right) + \rho|x|^b \text{ is convex,} \end{cases} \tag{B3}$$

and introduce the assumption on F :

$$\forall A > 0, \mathbb{E}\left[F(X)^2 e^{A|X|^{b-1}}\right] < +\infty. \tag{B4}$$

One shows that as soon as (B1), (B2), (B3) and (B4) are satisfied, Q_1 and Q_2 are both finite and differentiable on $\mathbb{R}^d$ with a gradient given by

$$\nabla Q(\theta) \quad := \quad \mathbb{E}\left[F(X-\theta)^2 \underbrace{\frac{p^2(X-\theta)}{p(X)p(X-2\theta)}\frac{\nabla p(X-2\theta)}{p(X-2\theta)}}_{W(\theta,X)}\right]. \tag{27}$$

This expression may look complicated at first glance but in fact the weight term $W(\theta, X)$ can be easily controlled by a deterministic function of θ since

$$|W(\theta, X)| \leq e^{2\rho|\theta|^b}(A|x|^{b-1} + A|\theta|^{b-1} + B) \tag{28}$$

for some real constants A and B. In the case of a normal distribution $X \stackrel{d}{=} \mathcal{N}(0;1)$,

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

So, if we have a control on the growth of the function F, typically for some positive constant c

$$\begin{cases} \forall x \in \mathbb{R}^d, |F(x)| \leq G(x) \quad \text{and} \quad G(x+y) \leq C(1+G(x))^c(1+G(y))^c \\ \mathbb{E}\left[|X|^{2(b-1)}G(X)^{4c}\right] < +\infty, \end{cases} \tag{B5}$$

then by setting

$$\widetilde{W}(\theta, X) := \frac{e^{-2\rho|\theta|^b}}{1+G(-\theta)^{2c}}W(\theta, X), \tag{29}$$

we can define K by

$$K(\theta, x) := F(x-\theta)^2\widetilde{W}(\theta, X) \tag{30}$$

so that it satisfies the linear growth assumption (8) of Theorem 2.2 and

$$\left\{\theta \in \mathbb{R}^d \mid \mathbb{E}\left[K(\theta, X)\right] = 0\right\} = \left\{\theta \in \mathbb{R}^d \mid \nabla Q(\theta) = 0\right\}.$$

Moreover, since Q is convex ∇Q satisfies (7). Now we are in position to derive a recursive unconstrained RM algorithm

$$\theta_n = \theta_{n-1} - \gamma_n K(\theta_{n-1}, X_n), \ \theta_0 \in \mathbb{R}^d, \tag{31}$$

that $a.s.$ converges to an $\arg\min Q$-valued (square integrable) random variable θ^*.

3 Design of a faster procedure: importance sampling and moving confidence level

3.1 Unconstrained adaptive importance sampling device

We noted previously that the bottleneck in using the above algorithm lies in its very slow and chaotic convergence owing to the fact that $\mathbb{P}(\varphi(X) > \xi^*) = 1 - \alpha$ is close to 0. This means that we observe fewer and fewer simulations for which $\varphi(X_k) > \xi_{k-1}$ as the algorithm evolves. Thus, it becomes more and more difficult to compute efficiently some estimates of VaR_α and CVaR_α when $\alpha \approx 1$. Moreover, in the bank and energy sectors, practitioners usually deal with huge portfolio made of hundreds or thousands of risk factors and options. The evaluation step of $\varphi(X)$ may be extremely time consuming. Consequently, to achieve accurate estimates of both VaR_α and CVaR_α

with reasonable computational effort, the above algorithm (16) drastically needs to be speeded up by an IS procedure to "recenter" the simulations where "things do happen", *i.e.* which generates scenarios for which $\varphi(X)$ exceeds ξ.

In this section we will focus on IS by mean translation. Our aim is to combine adaptively the IS (unconstrained) recursive procedure investigated in [27] with our first "naive" approach described in (16). Doing so every new sample is used to both optimize the IS change of measure and update VaR and CVaR procedures. We plan to minimize the asymptotic variance of both components of the algorithm (in its "averaged" form, as detailed in Theorem 2.4), namely

$$\frac{\alpha(1-\alpha)}{f_{\varphi(X)}(\xi^*)} = \frac{\mathrm{Var}(\mathbf{1}_{\varphi(X)\geq\xi^*})}{f_{\varphi(X)}(\xi^*)} \quad \text{for the VaR}_\alpha, \tag{32}$$

and,

$$\frac{\mathrm{Var}((\Psi(\varphi(X)) - \Psi(\xi^*))\mathbf{1}_{\varphi(\mathbf{X})\geq\xi^*})}{(1-\alpha)^2} \quad \text{for the CVaR}_\alpha, \tag{33}$$

provided the non-degeneracy assumption

$$\forall \xi \in \arg\min V,\ \mathbb{P}\left((\Psi(\varphi(X)) - \Psi(\xi))^2\, \mathbf{1}_{\{\varphi(X)\geq\xi\}} > 0\right) > 0, \tag{A3}$$

holds. Since the density $f_{\varphi(X)}(\xi^*)$ is an intrinsic constant (and comes in fact from the Jacobian matrix $Dh(\xi^*, C^*)$ of the mean function h of the algorithm) we are led to apply the IS paradigm described in Section 2.3 to

$$F_1^*(X) = \mathbf{1}_{\varphi(X)\geq\xi^*} \quad \text{and} \quad F_2^*(X) = (\Psi(\varphi(X)) - \Psi(\xi^*))\, \mathbf{1}_{\{\varphi(X)\geq\xi^*\}}.$$

Let us temporary forget that of course we do not know ξ^* at this stage. Those two functionals are related to the minimization of the two convex functions

$$Q_1(\theta, \xi^*) := \mathbb{E}\left[\mathbf{1}_{\{\varphi(X)\geq\xi^*\}} \frac{p(X)}{p(X-\theta)}\right] \tag{34}$$

$$Q_2(\mu, \xi^*) := \mathbb{E}\left[(\Psi(\varphi(X)) - \Psi(\xi^*))^2\, \mathbf{1}_{\{\varphi(X)\geq\xi^*\}} \frac{p(X)}{p(X-\mu)}\right]. \tag{35}$$

We can apply to these functions the minimizing procedure (31) described at section 2.3. Since

$$H_1(\xi^*, x) = 1 - \frac{1}{1-\alpha} F_1^*(x) \quad \text{and} \quad H_2(\xi^*, C^*, x) = C^* - \Psi(\xi^*) - \frac{1}{1-\alpha} F_2^*(x) \tag{36}$$

it is clear, owing to (24) that

$$\mathbb{E}\left[H_i(\xi^*, X)\right] = \mathbb{E}\left[H_i(\xi^*, X+\theta)\, \frac{p(X+\theta)}{p(X)}\right] \quad i = 1,\ 2.$$

Now, since we do not know either ξ^* and C^* (the VaR$_\alpha$ and the CVaR$_\alpha$) 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, for $n \geq 0$,

$$Z_n := (\xi_n, C_n, \theta_n, \mu_n),$$

where $\xi_n,\ C_n$ denotes the VaR$_\alpha$ and the CVaR$_\alpha$ approximate, $\theta_n,\ \mu_n$ denotes the variance reducers for the VaR and the CVaR procedures. We update this state variable recursively by

$$Z_n = Z_{n-1} - \gamma_n L(Z_{n-1}, X_n), \tag{37}$$

where $(X_n)_{n\geq 1}$ is an i.i.d. sequence with distributions X (and probability density p) and

$$\begin{aligned}
L_1(\xi,\theta,x) &:= e^{-\rho|\theta|^b}\left(1-\frac{1}{1-\alpha}\,\mathbf{1}_{\{\varphi(x+\theta)\geq\xi\}}\frac{p(x+\theta)}{p(x)}\right),\\
L_2(\xi,C,\mu,x) &:= C-\Psi(\xi)-\frac{1}{1-\alpha}(\Psi(\varphi(x+\mu))-\Psi(\xi))\,\mathbf{1}_{\{\varphi(x+\mu)\geq\xi\}}\frac{p(x+\mu)}{p(x)},\\
L_3(\xi,\theta,x) &:= e^{-2\rho|\theta|^b}\mathbf{1}_{\{\varphi(x-\theta)\geq\xi\}}\frac{p^2(x-\theta)}{p(x)p(x-2\theta)}\frac{\nabla p(x-2\theta)}{p(x-2\theta)}, \qquad (38)\\
L_4(\xi,\mu,x) &:= \frac{e^{-2\rho|\mu|^b}}{1+G(-\mu)^{2c}+\Psi(\xi)^2}\left(\Psi(\varphi(x-\mu))-\Psi(\xi)\right)^2\\
&\qquad\times\mathbf{1}_{\{\varphi(x-\mu)\geq\xi\}}\frac{p^2(x-\mu)}{p(x)p(x-2\mu)}\frac{\nabla p(x-2\mu)}{p(x-2\mu)}. \qquad (39)
\end{aligned}$$

The following proposition establishes the *a.s.* convergence of the procedure. For the sake of simplicity we will assume the uniqueness of the VaR_α of $\varphi(X)$.

Proposition 3.1. *(Efficient computation of VaR and CVaR). Suppose that $\Psi(\varphi(X))\in L^2(\mathbb{P})$, that the distribution function of $\varphi(X)$ is continuous and increasing (so that $VaR_\alpha(\varphi(X))$ is unique) and that (A3) holds. Assume that, for every $\xi\in\mathbb{R}$, $Q_i(.,\xi)$ (i=1,2) satisfies (B1), i.e.*

$$\forall\theta\in\mathbb{R}^d,\quad \mathbb{E}\left[\left(1+(\Psi(\varphi(X))-\Psi(\xi))^2\right)\mathbf{1}_{\varphi(X)\geq\xi}\frac{p(X)}{p(X-\theta)}\right]<+\infty. \qquad (40)$$

Suppose that p satisfies (B2) and (B3) and that

$$\forall A>0,\mathbb{E}\left[\left(\Psi(\varphi(X))^2+1\right)e^{A|X|^{b-1}}\right]<+\infty.$$

Assume that the step sequence $(\gamma_n)_{n\geq1}$ satisfies (A1). Then,

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

where $\xi^*=VaR_\alpha(\varphi(X))$, $C^*=\Psi\text{-}CVaR_\alpha(\varphi(X))$ and $(\theta^*_\alpha,\mu^*_\alpha)$ are the optimal variance reducers (to be precise some random vectors taking values in $\{\nabla Q_1(\xi^*,.)=0\}$ and $\{\nabla Q_2(\xi^*,.)=0\}$ respectively).

Proof. We first prove the *a.s.* convergence of the 3-tuple (ξ_n,θ_n,μ_n) that of $(C_n)_{n\geq1}$ will follow by the same arguments used in the proof in Section 2.2. The mean function l is defined by

$$l(\xi,\theta,\mu):=\left(1-\frac{1}{1-\alpha}\mathbb{P}\left(\varphi(X)\geq\xi\right),e^{-2\rho|\theta|^b}\nabla Q_1\left(\theta,\xi\right),\frac{e^{-2\rho|\mu|^b}}{1+G(-\mu)^c+\Psi(\xi)^2}\nabla Q_2(\mu,\xi)\right),$$

hence,

$$\mathcal{T}^*=\{l=0\}=\{\xi^*\}\times\{\nabla Q_1(\xi^*,.)=0\}\times\{\nabla Q_2(\xi^*,.)=0\}.$$

In order to apply the extended Robbins-Monro Theorem, we have to check the following facts:

- *Mean reversion*: One checks that $\forall\zeta=(\xi,\theta,\mu)\in\mathbb{R}\times\mathbb{R}^d\times\mathbb{R}^d\setminus\mathcal{T}^*$, $\forall\zeta^*\in\mathcal{T}^*$,

$$\begin{aligned}
\langle\zeta-\zeta^*,l(\zeta)\rangle &= (\xi-\xi^*)\frac{(\mathbb{P}(\varphi(X)\leq\xi)-\alpha)}{1-\alpha}+\frac{e^{-2\rho|\theta|^b}}{1-\alpha}\langle\theta-\theta^*_\alpha,\nabla Q_1(\theta,\xi)\rangle\\
&\quad+\frac{e^{-2\rho|\mu|^b}}{(1-\alpha)(1+F(-\mu)^{2c})}\langle\mu-\mu^*_\alpha,\nabla Q_2(\mu,\xi)\rangle>0,
\end{aligned}$$

owing to the convexity of $\theta \mapsto Q_1(\theta,\xi)$ and $\mu \mapsto Q_2(\mu,\xi)$, for every $\xi \in \mathbb{R}$.

• *Linear growth:* Let us first deal with L_1. First note that:

$$\mathbb{E}\left[L_1(\xi,\theta,X)^2\right] \leq C\left(1+\mathbb{E}\left[e^{-2\rho|\theta|^b}\mathbf{1}_{\{\varphi(X+\theta)\geq\xi\}}\frac{p^2(X+\theta)}{p^2(X)}\right]\right) \leq C\left(1+\mathbb{E}\left[e^{-2\rho|\theta|^b}\frac{p(X)}{p(X-\theta)}\right]\right).$$

Now, elementary computations show (see [27] for more details) that (B3)(ii) implies that

$$\frac{p^2(x)}{p(x-\theta)} \leq e^{2\rho|\theta|^b}p(x+\theta),$$

so that

$$\mathbb{E}\left[e^{-2\rho|\theta|^b}\frac{p(X)}{p(X-\theta)}\right] \leq \mathbb{E}\left[\frac{p(X+\theta)}{p(X)}\right] = 1.$$

L_3 and L_4 can be treated by a straightforward adaptation of the proofs in [27]. Then, one can apply Theorem 2.2 which yields the announced result for (ξ_n,θ_n,μ_n). The *a.s.* convergence of C_n toward C^* can be deduced from the *a.s.* convergence of the series

$$M_n^\gamma := \sum_{k=1}^n \gamma_k \Delta\widetilde{M}_k,\ n\geq 1,$$

where $\Delta\widetilde{M}_n$ are martingale increments defined by

$$\begin{aligned}\Delta\widetilde{M}_n &= \mathbb{E}[(\Psi(\varphi(X))-\Psi(\xi))\ \mathbf{1}_{\{\varphi(X)\geq\xi\}}]_{|\xi=\xi_{n-1}} \\ &\quad -(\Psi(\varphi(X_n+\mu_{n-1}))-\Psi(\xi_{n-1}))\ \mathbf{1}_{\{\varphi(X_n+\mu_{n-1})\geq\xi_{n-1}\}}\frac{p(X_n+\mu_{n-1})}{p(X_n)},\quad n\geq 1,\end{aligned}$$

satisfying

$$\mathbb{E}\left[\Delta\widetilde{M}_n^2|\mathcal{F}_{n-1}\right] \leq \mathbb{E}\left[(\Psi(\varphi(X+\mu))-\Psi(\xi))\ \mathbf{1}_{\{\varphi(X+\mu)\geq\xi\}}\frac{p(X+\mu)}{p(X)}\right]_{|\xi=\xi_{n-1},\theta=\theta_{n-1},\mu=\mu_{n-1}}.$$

We conclude by the same arguments used in the proof in Section 2.2. □

Now, we are interested by the rate of convergence of the procedure. It shows that the algorithm behaves as expected under quite standard assumptions: it satisfies a Gaussian CLT with optimal rate and minimal variances.

Theorem 3.2. *Suppose the assumptions of Proposition 3.1 hold true. Assume that $\Psi(\varphi(X)) \in L^{2a}(\mathbb{P})$ for some $a>1$ and that the step sequence is $\gamma_n = \frac{\gamma_1}{n^p}$ with $\frac{1}{2}<p<1$ and $\gamma_1>0$. Suppose furthermore that Ψ is twice differentiable at ξ^* and that the density $f_{\varphi(X)}$ is differentiable and strictly positive on its support. Let $(\overline{\xi}_n,\overline{C}_n)_{n\geq 1}$ be the sequence of Cesaro means defined by:*

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

This sequence satisfies the following CLT:

$$\sqrt{n}\begin{pmatrix}\overline{\xi}_n-\xi^* \\ \overline{C}_n-C^*\end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}(0,\Sigma^*) \qquad \text{as } n\to+\infty, \tag{41}$$

where

$$\begin{aligned}
\Sigma^*_{1,1} &= \frac{1}{f^2_{\varphi(X)}(\xi^*)}\mathrm{Var}\left(\mathbf{1}_{\{\varphi(X+\theta^*_\alpha)\geq\xi^*\}}\frac{p(X+\theta^*_\alpha)}{p(X)}\right),\\
\Sigma^*_{1,2} &= \Sigma^*_{2,1} = \frac{1}{(1-\alpha)f_{\varphi(X)}(\xi^*)}\mathrm{Cov}\left((\Psi(\varphi(X+\mu^*_\alpha))-\Psi(\xi^*))\,\mathbf{1}_{\{\varphi(X+\mu^*_\alpha)>\xi^*\}}\frac{p(X+\mu^*_\alpha)}{p(X)},\right.\\
&\qquad\qquad\left.\mathbf{1}_{\{\varphi(X+\theta^*_\alpha)\geq\xi^*\}}\frac{p(X+\theta^*_\alpha)}{p(X)}\right),\\
\Sigma^*_{2,2} &= \frac{1}{(1-\alpha)^2}\mathrm{Var}\left((\Psi(\varphi(X+\mu^*_\alpha))-\xi^*)\;\mathbf{1}_{\{\varphi(X+\mu^*_\alpha)\geq\xi^*\}}\frac{p(X+\mu^*_\alpha)}{p(X)}\right).
\end{aligned}$$

Proof. The proof is built like the one of Theorem 2.4. If we denote h the mean function of the global algorithm $h(z)=\mathbb{E}[L(z,X)]$, the algorithm (37) can be written as

$$Z_n = Z_{n-1}-\gamma_n\left(h(Z_{n-1})+\tilde{\epsilon}_n\right),\ n\geq 1,\quad Z_0=(\xi_0,0)\,,\ \xi_0\in L^1(\mathbb{P}), \tag{42}$$

where the first two components of h are the same function as the ones in the proof of Theorem 2.4 and $(\tilde{\epsilon}_n)_{n\geq1}$ denotes the $\mathcal{F}_n$-adapted martingale increments sequence where

$$\begin{aligned}
\tilde{\epsilon}_{1,n} &:= \frac{1}{1-\alpha}\left(\mathbb{P}\left(\varphi(X)\geq\xi\right)_{|\xi=\xi_n} - \mathbf{1}_{\{\varphi(X_{n+1}+\theta_n)\geq\xi_n\}}\frac{p(X_{n+1}+\theta_n)}{p(X_{n+1})}\right),\\
\tilde{\epsilon}_{2,n} &:= \frac{1}{1-\alpha}\left(\mathbb{E}[(\Psi(\varphi(X))-\Psi(\xi))\;\mathbf{1}_{\{\varphi(X)\geq\xi\}}]_{|\xi=\xi_n}\right.\\
&\qquad\left.-(\Psi(\varphi(X_{n+1}+\mu_n))-\Psi(\xi_n))\;\mathbf{1}_{\{\varphi(X_{n+1}+\mu_n)\geq\xi_n\}}\frac{p(X_{n+1}+\mu_n)}{p(X_{n+1})}\right).
\end{aligned}$$

One can check easily that the sequence $(\tilde{\epsilon}_n)_{n\geq1}$ satisfies $(i)-(iv)$ of (18). □

Remarks: • There exists a CLT for the whole sequence $(Z_n)_{n\geq1}$ and for its empirical mean $(\overline{Z}_n)_{n\geq1}$ according to Ruppert and Polyak averaging principle. We only stated the result for the two components of interest (the ones which converge to VaR and CVaR respectively) since we only need rough estimates for the other two (see below).

• In the first Central Limit Theorem (Theorem 2.4) for quantile estimation, the factor $\alpha(1-\alpha)$ is the variance of the indicator function of the event $\{\varphi(X)\geq\xi^*\}$. With our recursive IS procedure, it is replaced by the variance of the shifted indicator function modified by the measure change: $\mathrm{Var}\left(\mathbf{1}_{\{\varphi(X+\theta^*_\alpha)>\xi^*\}}\frac{p(X+\theta^*_\alpha)}{p(X)}\right)$. For further details on the rate of convergence of the unconstrained recursive importance sampling procedure, we refer to [27].

Now, let us point out an important issue. The algorithm (37) raises an important problem numerically speaking. Actually, we have two algorithm ξ_n and (θ_n,μ_n) that are in competitive conditions, *i.e.* on one hand, we added an IS procedure to $(\xi_n)_{n\geq1}$ to improve the convergence toward ξ^*, and on the other hand, the adjustment of the parameters (θ_n,μ_n) "need" some samples X_{n+1} satisfying $\varphi(X_{n+1}-\theta_n)>\xi_n$ and $\varphi(X_{n+1}-\mu_n)>\xi_n$ ($\Psi\equiv Id$) which tend to become rare events. Somehow, we postponed the problems resulting from rare events on the IS procedure itself which may "freeze". This in term suggests to break the link between the VaR-CVaR and the IS procedures by introducing a VaR *companion procedure* that will drive the IS parameters to the tail distribution. A solution to do this is to make the confidence level increase slowly from a lower value (say $\alpha_0=50\%$) up to the target level α. This kind of incremental threshold increase has been already proposed in [22] in a different framework. This idea is developed in the next section.

3.2 How to control the move towards the critical risk area: the final procedure

From a theoretical point of view, so far, we considered the purely adaptive approach where we approximate $(\xi^*, C^*, \theta^*_\alpha, \mu^*_\alpha)$ using the *same innovation sequences*. From a numerical point of view, we only need a rough estimate of the optimal IS parameters $(\theta^*_\alpha, \mu^*_\alpha)$. So that we are led to break the algorithm into two phases. Firstly, we compute a rough estimate of the optimal IS parameters (θ_M, μ_M) with a small number of iterations M and in a second time, estimate the VaR_α and the CVaR_α with those optimized parameters with N iterations ($M \ll N$ in practice).

Now, in order to circumvent the problem induced by the IS procedure, we propose to introduce companion VaR procedure (without IS, *i.e.*, based on H_1 from Section 2.2) that will lead the IS parameters into the critical risk area during a first phase of the simulation, say the first M iterations. An idea 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 $\varphi(X)$, is to move slowly toward the target critical risk area (at level α) in which $\varphi(X)$ exceeds ξ by introducing a non-decreasing sequence α_n slowly converging to α during the first phase. Since the algorithm for the CVaR component C_n is free of α, by doing so, we only modify the VaR procedure ξ_n. The function H_1 in (16) is replaced by its counterpart which depends on the moving confidence level α_n, namely

$$\hat{\xi}_n = \hat{\xi}_{n-1} - \gamma_n \hat{H}_1\left(\hat{\xi}_{n-1}, X_n, \alpha_n\right), \quad n \geq 1, \hat{\xi}_0 = \xi_0 \in L^1(\mathbb{P}). \tag{43}$$

where,

$$\forall\, \xi \in \mathbb{R},\ \forall x \in \mathbb{R}^d, \forall\, \hat{\alpha} \in]0,1[,\ \hat{H}_1\left(\xi, x, \hat{\alpha}\right) = 1 - \frac{1}{1-\hat{\alpha}}\ \mathbf{1}_{\{\varphi(x)\geq\xi\}}.$$

The sequence $\left(\hat{\xi}_n\right)_{n\geq 0}$ is only designed to drive "smoothly" the IS procedures toward the "critical area" at the beginning of the procedure, say during the first M iterations and in no case to approximate ξ^* or C^*. To be more precise, we define recursively the variance reducer sequence $(\hat{\theta}_n)_{n\geq 1}$, $(\hat{\mu}_n)_{n\geq 1}$ by plugging at each step n, $\hat{\xi}_{n-1}$ into $L_3(., \hat{\theta}_{n-1}, X_n)$ and $L_4(., \hat{\mu}_{n-1}, X_n)$ as defined in Section 3.1. This reads as follows, for $n \geq 1$,

$$\begin{cases} \hat{\xi}_n = \hat{\xi}_{n-1} - \gamma_n \hat{H}_1\left(\hat{\xi}_{n-1}, X_n, \alpha_n\right), \quad \hat{\xi}_0 \in L^1(\mathbb{P}), \\ \hat{\theta}_n = \hat{\theta}_{n-1} - \gamma_n L_3\left(\hat{\xi}_{n-1}, \hat{\theta}_{n-1}, X_n\right), \quad \theta_0 \in \mathbb{R}^d, \\ \hat{\mu}_n = \hat{\mu}_{n-1} - \gamma_n L_4\left(\hat{\xi}_{n-1}, \hat{\mu}_{n-1}, X_n\right), \quad \mu_0 \in \mathbb{R}^d. \end{cases} \tag{44}$$

Although, we are not really interested in the asymptotic of this procedure $(\hat{\xi}_n)$, its theoretical convergence follows from Theorem 2.2: as a matter of fact if we define a remainder term r_n by:

$$r_n := \hat{H}_1\left(\hat{\xi}_{n-1}, X_n, \alpha_n\right) - H_1\left(\hat{\xi}_{n-1}, X_n\right), \quad n \geq 1,$$

the procedure defined by (44) now reads

$$\hat{\xi}_n = \hat{\xi}_{n-1} - \gamma_n(H_1(\hat{\xi}_{n-1}, X_n) + r_n), \quad n \geq 1, \quad \hat{\xi}_0 \in L^1(\mathbb{P}). \tag{45}$$

One checks that

$$|r_n| \leq \frac{|\alpha_n - \alpha|}{(1-\alpha)^2},$$

so that Assumption (9) of Theorem 2.2 is satisfied as soon as

$$\sum_{n\geq 1} \gamma_n (\alpha - \alpha_n)^2 < +\infty.$$

3.3 A final procedure for practical implementation

In practice, we divided our procedure into two phases:
▷ Phase I is devoted to the estimation of the variance reducers $(\theta^*_\alpha, \mu^*_\alpha)$ using (44). The moving confidence level α has been settled as follows ($M \approx 15000$) :

$$\alpha_n = 50\% \text{ for } \ 1 \leq n \leq M_1 := M/3, \ \alpha_n = 80\% \text{ for } M_1 < n \leq 2M_1, \ \alpha_n = \alpha \text{ for } \ 2M_1 < n \leq M.$$

▷ Phase II produces some estimates for (ξ^*, C^*) based on the procedure defined by (37) and its Cesaro mean with N iterations. Note that during this phase, we keep on updating the IS parameters adaptively.

Now, we can summarize the two phase of the final procedure by the following pseudo-code:

Phase I: Estimation of $(\mu^*_\alpha, \theta^*_\alpha)$. $M \ll N$ (typically $M \approx N/100$).
for $n = 1$ to M **do**
  $\hat{\xi}_n \leftarrow \hat{\xi}_{n-1} - \gamma_n \hat{H}_1 \left(\hat{\xi}_{n-1}, X_n, \alpha_n\right),$
  $\hat{\theta}_n \leftarrow \hat{\theta}_{n-1} - \gamma_n L_3 \left(\hat{\xi}_{n-1}, \hat{\theta}_{n-1}, X_n\right),$
  $\hat{\mu}_n \leftarrow \hat{\mu}_{n-1} - \gamma_n L_4 \left(\hat{\xi}_{n-1}, \hat{\mu}_{n-1}, X_n\right).$
end for

Phase II: Estimation of (ξ^*, C^*). Set, for instance, $\xi_0 = \hat{\xi}_M$, $C_0 = 0$, $\theta_0 = \hat{\theta}_M$, and $\mu_0 = \hat{\mu}_M$.
for $n = 1$ to N **do**
  $\xi_n \leftarrow \xi_{n-1} - \gamma_n L_1 \left(\xi_{n-1}, \theta_{n-1}, X_n\right)$
  $C_n \leftarrow C_{n-1} - \gamma_n L_2 \left(\xi_{n-1}, C_{n-1}, \mu_{n-1}, X_n\right)$
  $\theta_n \leftarrow \theta_{n-1} - \gamma_n L_3 \left(\xi_{n-1}, \theta_{n-1}, X_n\right)$
  $\mu_n \leftarrow \mu_{n-1} - \gamma_n L_4 \left(\xi_{n-1}, \mu_{n-1}, X_n\right)$
  Compute the Cesaro means

$$\bar{\xi}_n \leftarrow \bar{\xi}_{n-1} - \frac{1}{n}\left(\bar{\xi}_{n-1} - \xi_n\right), \quad \bar{C}_n \leftarrow \bar{C}_{n-1} - \frac{1}{n}\left(\bar{C}_{n-1} - C_n\right)$$

end for
(ξ^*, C^*) is estimated by $(\bar{\xi}_N, \bar{C}_N)$.

An alternative, especially as concerns practical implementation, is to replace to Phase II by

Phase II' in which the variance reducers coming from Phase I are frozen at $\hat{\theta}_M$, $\hat{\mu}_M$. The only updated sequence is (ξ_n, C_n), as follows

$$\begin{aligned} \xi_n &= \xi_{n-1} - \gamma_n L_1 \left(\xi_{n-1}, \hat{\theta}_M, X_n\right), \\ C_n &= C_{n-1} - \gamma_n L_2 \left(\xi_{n-1}, C_{n-1}, \hat{\mu}_M, X_n\right). \end{aligned}$$

4 Towards some extensions

4.1 Extension to exponential change of measure: the Esscher transform

Considering an exponential change of measure (also called Esscher transform) instead of the mean translation is a rather natural idea that has already been investigated in [20] and [27] to extend the constrained IS stochastic approximation algorithm with repeated projections introduced in [1].

We briefly introduce the framework and give the main results without any proofs (for more details, see [27] and [13]). Let ψ denote the cumulant generating function (or log-Laplace) of X *i.e.* the function defined by $\psi(\theta) := \log \mathbb{E}[e^{\langle \theta, X\rangle}]$. We assume that $\psi(\theta) < +\infty$, which implies that ψ is an infinitely differentiable convex function and define

$$p_\theta(x) = e^{\langle \theta, x\rangle - \psi(\theta)} p(x), \quad x \in \mathbb{R}^d.$$

We denote by $X^{(\theta)}$ any random variable with distribution p_θ. We make the following assumption on the function ψ

$$\lim_{|\theta|} \psi(\theta) - 2\psi\left(\frac{\theta}{2}\right) = +\infty \quad \text{and} \quad \exists \delta > 0,\ \theta \mapsto \psi(\theta) - \delta|\theta|^2 \text{ is concave.} \tag{H_δ^{es}}$$

The two functionals to be minimized are

$$Q_1(\theta, \xi^*) := \mathbb{E}\left[\mathbf{1}_{\{\varphi(X) > \xi^*\}} e^{-\langle \theta, X\rangle + \psi(\theta)}\right] \tag{46}$$

$$Q_2(\mu, \xi^*) := \mathbb{E}\left[(\Psi(\varphi(X)) - \Psi(\xi^*))^2 \mathbf{1}_{\{\varphi(X) > \xi^*\}} e^{-\langle \mu, X\rangle + \psi(\mu)}\right]. \tag{47}$$

According to Proposition 3 in [27] as soon as ψ satisfies (H_δ^{es}) and that,

$$\forall \xi \in \mathbb{R}, \forall \theta \in \mathbb{R}^d, \quad \mathbb{E}[|X| \left(1 + \Psi(\varphi(X))^2\right) e^{\langle \theta, X\rangle}] < +\infty, \tag{48}$$

for every $\xi \in \mathbb{R}$, the functions $Q_1(., \xi)$ and $Q_2(., \xi)$ are finite, convex, differentiable on $\mathbb{R}^d$, go to infinity at infinity, so that $\arg\min Q_1(., \xi)$ and $\arg\min Q_2(., \xi)$ are non empty. Moreover, their gradients are given by

$$\nabla_\theta Q_1(\theta, \xi) = \mathbb{E}\left[(\nabla\psi(\theta) - X^{(-\theta)}) \mathbf{1}_{\left\{\varphi(X^{(-\theta)}) > \xi\right\}}\right] e^{\psi(\theta) - \psi(-\theta)} \tag{49}$$

$$\nabla_\mu Q_2(\mu, \xi) = \mathbb{E}\left[(\nabla\psi(\mu) - X^{(-\mu)})(\Psi(\varphi(X^{(-\mu)})) - \Psi(\xi))^2 \mathbf{1}_{\left\{\varphi(X^{(-\mu)}) > \xi\right\}}\right] e^{\psi(\mu) - \psi(-\mu)} \tag{50}$$

with $\nabla\psi(\theta) = \frac{\mathbb{E}[X e^{\langle \theta, X\rangle}]}{\mathbb{E}[e^{\langle \theta, X\rangle}]}$. Now, the main result of this section is the following theorem (for more details, we refer to [27] and [13]).

Theorem 4.1. *Suppose that ψ satisfies (H_δ^{es}) and that $(A2)_1$, (A3) hold. Assume that (48) is fulfilled and that*

$$\forall x \in \mathbb{R}^d, |\Psi(\varphi(x))| \le C e^{\frac{\lambda}{4}|x|} \quad \textit{and} \quad \mathbb{E}[|X|^2 e^{\lambda|X|}] < +\infty.$$

One considers the recursive procedure

$$Z_n = Z_{n-1} - \gamma_n L(Z_{n-1}, X_n),\ n \ge 1,\ Z_0 = (\xi_0, C_0, \theta_0, \mu_0) \tag{51}$$

where $(\gamma_n)_{n\ge1}$ satisfies the usual step assumption (A1), $Z_n := (\xi_n, C_n, \theta_n, \mu_n)$ and each component of L is defined by

$$\begin{cases} L_1\left(\xi_{n-1}, \theta_{n-1}, X_n^{(\theta_{n-1})}\right) := e^{-\frac{\psi(\theta_{n-1}) + \psi(-\theta_{n-1})}{2}} \left(1 - \frac{1}{1-\alpha} \mathbf{1}_{\left\{\varphi(X_n^{(\theta_{n-1})}) > \xi_{n-1}\right\}} e^{\psi(\theta_{n-1}) - \left\langle X_n^{(\theta_{n-1})}, \theta_{n-1}\right\rangle}\right), \\ L_2\left(\xi_{n-1}, C_{n-1}, \mu_{n-1}, X_n^{(\mu_{n-1})}\right) := C - \bar{w}(\xi_{n-1}, \mu_{n-1}, X_n^{(\mu_{n-1})}), \\ L_3\left(\xi_{n-1}, \theta_{n-1}, X_n^{(-\theta_{n-1})}\right) := \mathbf{1}_{\left\{\varphi(X^{(-\theta_{n-1})}) > \xi_{n-1}\right\}} (\nabla\psi(\theta_{n-1}) - X^{(-\theta_{n-1})}), \\ L_4\left(\xi_{n-1}, \mu_{n-1}, X_n^{(-\mu_{n-1})}\right) := \frac{e^{-\frac{\lambda}{2}\sqrt{d}|\nabla\psi(-\mu_{n-1})|}}{1 + \Psi(\xi_{n-1})^2} (\Psi(\varphi(X^{(-\mu_{n-1})})) - \Psi(\xi_{n-1}))^2 \mathbf{1}_{\left\{\varphi(X^{(-\mu_{n-1})}) > \xi_{n-1}\right\}} \\ \qquad\qquad \times (\nabla\psi(\mu_{n-1}) - X^{(-\mu_{n-1})}), \end{cases}$$

with $\bar{w}(\xi,\mu,x) := \Psi(\xi) + \dfrac{1}{1-\alpha}(\Psi(\varphi(x)) - \Psi(\xi))\mathbf{1}_{\{\varphi(x)>\xi\}}\ e^{\psi(\mu)-\langle\mu,x\rangle}$.

Then, Z_n *converges a.s. toward* $z^* := (\xi^*, C^*, \theta^*_\alpha, \mu^*_\alpha)$*, where* ξ^* *is a square integrable* VaR_α*-valued random variable,* $C^* = \Psi\text{-}CVaR_\alpha(\varphi(X))$*,* θ^*_α *is a (square integrable)* $\arg\min Q_1(.,\xi^*)$*-valued random vector and* μ^*_α *is a (square integrable)* $\arg\min Q_2(.,\xi^*)$*-valued random vector.*

4.2 Extension to infinite dimensional setting

In the above sections, we proposed our algorithm in a finite dimensional setting where the value of the loss $L = \varphi(X)$ is a function of a random vector having values in $\mathbb{R}^d$. This is due to the fact that generally the value of a portfolio may depend on a finite number of decisions taken in the past. Thus, the value of the loss at the horizon time $T-t$ may depend on a large number of dates in the past $t_0 = t < t_1 < t_2, ... < t_N = T-t$, with $N = 250$ for a portfolio with time interval $T-t = 1$ year. For instance, if we consider a simple portfolio composed of short positions on 250 calls with a maturity at each t_k and a strike K. The loss at time $t_N = 1$ year can be written:

$$L = \sum_{k=1}^{N} e^{r(t_N-t_k)}(S_{t_k} - K)_+ - e^{rt_N}C_0^k,$$

where C_0^i denotes the price of the call of maturity t_i and strike K, with

$$S_{t_{k+1}} = S_{t_k} e^{(r-\frac{\sigma^2}{2})(t_{k+1}-t_k)+\sigma\sqrt{(t_{k+1}-t_k)}Z_k}.$$

So that, $X = Z = (Z_1, ..., Z_{250})$ is a Gaussian vector with $d = 250$. Consequently, with our above procedure, θ_n and μ_n are two random vectors of dimension d and we have to control the growth of each component. If one grows too fast and take too high values, it may provides bad performance and bad estimates of both VaR and CVaR. To circumvent this problem, one can reduce the dimension of the problem by choosing the same shift parameters for several dates, *i.e.* for instance

$$\theta_n = (\underbrace{\theta_n^1, .., \theta_n^1}_{10 \text{ times}}, ..., \underbrace{\theta_n^{25}, .., \theta_n^{25}}_{10 \text{ times}}).$$

Now, we can run the IS algorithm for $\theta^1, ..., \theta^{25}$ so that, we have to deal with a procedure in dimension 25. It is sub-optimal with respect to the procedure in dimension 250 but it is more tractable. Another relevant example is a portfolio composed by only one barrier option, for instance a Down & In Call option

$$\varphi(X) = (X_T - K)_+ \mathbf{1}_{\{\min_{\{0\le t\le T\}} X_t \le L\}}$$

where the underlying X is a process solution of the path-dependent SDE

$$\mathrm{d}X_t = b(X_t)\ \mathrm{d}t + \sigma(X_t)\ \mathrm{d}W_t,\ X_0 = x \in \mathbb{R}^d, \tag{52}$$

$W = (W_t)_{t\in[0,T]}$ being a standard Brownian motion. A naive approach is to discretize (52) by an Euler-Maruyama scheme $\bar{X} = (\bar{X}_{t_k})_{k\in\{0,...,n\}}$

$$\bar{X}_{t_{k+1}} = \bar{X}_{t_k} + b(\bar{X}_{t_k})(t_{k+1} - t_k) + \sigma(\bar{X}_{t_k})(W_{t_{k+1}} - W_{t_k}),\ \ \bar{X}_0 = x_0 \in \mathbb{R}.$$

This kind of approximation is known to be poor for this kind of options. In this case, our IS parameters θ and μ are n-dimensional vectors which correspond to the number of steps in the Euler scheme. Now, if you consider a portfolio composed by several barrier options with different underlyings, the dimension can increase greatly and becomes an important issue, so that our first IS procedure is no longer acceptable and tractable. To overcome this problem, the idea is to shift the entire distribution of X in (52) thanks to a Girsanov transformation. This last case is analyzed and investigated in [27]. It can be adapted to our framework (see [13] for further developments).

5 Numerical examples

For the sake of simplicity, we focus in this section on the finite dimensional setting and on the computation of the regular CVaR_α ($\Psi \equiv Id$). We first consider the usual Gaussian framework in which the exponential change of measure coincide with the mean translation change of measure. Then we illustrate the algorithm (51) in a simple case.

5.1 Gaussian framework

In this setting, $X \sim \mathcal{N}(0, I_d)$ and p is given by

$$p(x) = (2\pi)^{-\frac{d}{2}} e^{-\frac{|x|^2}{2}}, \quad x \in \mathbb{R}^d,$$

so that (B3) and (B4) are satisfied with $\rho = \frac{1}{2}$ and $b = 2$. In this setting, we already noticed that

$$\begin{aligned} L_3(\xi, \theta, x) &:= \mathbf{1}_{\{\varphi(X-\theta)\geq\xi\}}(2\theta - x), \\ L_4(\xi, \mu, x) &:= \frac{1}{1 + G(-\mu) + \xi^2} (\varphi(X-\mu) - \xi)_+^2 (2\mu - x). \end{aligned}$$

Moreover, we use a stepwise constant sequence α_n that slowly converges toward α as proposed in Section 3.3. We consider three different portfolios of options (puts and calls) on 1 and 5 underlying assets (except for the last case). In the third case, we study the behaviour of a portfolio composed by a power plant that produces electricity from gas with short positions in calls on electricity. The assets are modeled as geometric Brownian motions for the first two examples. In the third example, the assets (electricity and gas day-ahead prices) are modeled as exponentials of an Ornstein-Uhlenbeck process. This last derivative is priced using an approximation of Margrabe formulae (see e.g. [29]). We assume an annual risk free interest rate of 5%. In each example, we use three different values of the confidence level $\alpha = 95\%,\ 99\%,\ 99.5\%$, which are specified in the Tables. We use the following test portfolios:

Example 1. *Short position in one put with strike $K = 110$ and maturity $T = 1$ year on a stock with initial value $S_0 = 100$ and volatility $\sigma = 20\%$. The loss is given by*

$$\varphi_1(X) := (K - S_T)_+ - e^{rT} P_0$$

with

$$S_T := S_0 e^{\left(\left(r - \frac{\sigma^2}{2}\right)T + \sigma\sqrt{T}X\right)}$$

where $X \sim \mathcal{N}(0,1)$ and P_0 is the initial price at which the put option was sold (it is approximately equal to 10.7). The dimension d of the structural vector X is equal to 1. The numerical results are reported in Table 1.

Example 2. *Short positions in 10 calls and 10 puts on each of the five underlying assets, all options having the same maturity 0.25 year. The strikes are set to 130 for calls, to 110 for puts and the initial spot prices to 120. The underlying assets have a volatility of 20% and are assumed to be uncorrelated. The dimension d of the structural vector X is equal to 5. The numerical results are reported in Table 2.*

Example 3. *Short position in a power plant that produces electricity day by day with a maturity of $T = 1$ month and 30 long positions in calls on electricity day-ahead price with the same strike $K = 60$. Electricity and gas initial spot prices are $S_0^e = 40$ \$/MWh and $S_0^g = 3$ \$/MMBTU*

*(*BTU*: British Thermal Unit) with a Heat Rate equals* $h_R = 10$ BTU/kWh *and generation costs* $C = 5$ \$/MWh*. The two spot prices have a correlation of 0.4. The payoff can be written*

$$\varphi_3(X) = \sum_{k=1}^{30} \left(e^{r(T-t_k)} \left(S^e_{t_k} - h_R S^g_{t_k} - C \right)_+ - P^c_0 e^{rT} \right) + \left(e^{rT} C_0 - e^{r(T-t_k)} \left(S^e_{t_k} - K \right)_+ \right)$$

where P^c_0 *is a proxy of the price of the option on the power plant and is equal to 149.9 and* C_0 *is the price of the call options which is equal to 3.8. This is a sum of spark spread options where we decide to exchange gas and electricity each day during one month. The dimension* d *of the structural vector* X *is equal to 60. The numerical results are reported in Table 3.*

The results displayed in the following tables correspond to the VaR, the CVaR and the variance reduction ratios estimations for both VaR and CVaR procedure using a number of steps specified in the first column, still for the same three levels of α. The variance ratios correspond to the ratio of an estimation of the asymptotic variance using the averaging procedure of (16) divided by an estimation of the asymptotic variance using the averaging procedure of (37): $\mathrm{VR}_{\mathrm{VaR}}$ corresponds to the variance reduction ratio of the VaR estimate and $\mathrm{VR}_{\mathrm{CVaR}}$ corresponds to the variance reduction ratio of the CVaR estimate. The results emphasize that the IS procedure yields a very significant, sometimes huge variance reduction especially when α is closed to 1.
In the three examples, we define the step sequence by $\gamma_n = \frac{1}{n^\beta + 100}$ where $\beta = \frac{3}{4}$.

Table 1: Example 1 Results

Number of steps	α	VaR	CVaR	$\mathrm{VR}_{\mathrm{VaR}}$	$\mathrm{VR}_{\mathrm{CVaR}}$
10 000	95%	24.6	29.9	5.5	30.5
	99%	34.4	37.5	11.1	125.3
	99.5%	37.8	41.4	13.4	192.9
100 000	95%	24.6	30.4	6.6	32.2
	99%	34.2	37.9	11.5	127.9
	99.5%	37.3	40.7	15.1	185
500 000	95%	24.6	30.3	7.7	31.3
	99%	34.2	38	14.6	118.4
	99.5%	37.3	40.5	15.5	184

Table 2: Example 2 Results

Number of steps	α	VaR	CVaR	$\mathrm{VR}_{\mathrm{VaR}}$	$\mathrm{VR}_{\mathrm{CVaR}}$
10 000	95%	339	440.5	6.5	14.9
	99%	493.1	561.4	10.1	24.3
	99.5%	540.1	606.4	18.2	37.9
100 000	95%	349.8	439.7	6.7	17
	99%	495.7	563.8	11.3	28.6
	99.5%	544.8	607.8	18.9	40.3
500 000	95%	352.4	439.6	6.8	17.3
	99%	495.2	563	11.1	27.7
	99.5%	545.3	608.4	19.2	37

Table 3: Example 3 Results

Number of steps	α	VaR	CVaR	$\mathrm{VR}_{\mathrm{VaR}}$	$\mathrm{VR}_{\mathrm{CVaR}}$
10 000	95%	115.7	150.5	3.4	6.8
	99%	169.4	196	8.4	12.9
	99.5%	186.3	213.2	13.5	20.3
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
500 000	95%	119.2	150.4	5	9.2
	99%	169.8	195.7	13.1	18.6
	99.5%	188.7	212.8	17	29

5.2 Esscher transform: the NIG distribution

Now, we consider a simple case of portfolio composed by a long position on a Call option with strike $K = 0.6$ and maturity $T = 1$ year, where the underlying is e^{X_T} ($X_0 = 0$), where X_T is a Normal Inverse Gaussian (NIG) variable, $X_T \sim \mathrm{NIG}(\alpha, \beta, \delta, \mu)$, $\alpha > 0$, $|\beta| \leq \alpha$, $\delta > 0$, $\mu \in \mathbb{R}$. Its density is given by

$$p_{X_T}(x,\ \alpha,\ \beta,\ \delta,\ \mu) := \frac{\alpha\delta K_1(\alpha\sqrt{\delta^2 + (x-\mu)^2})}{\pi\sqrt{\delta^2 + (x-\mu)^2}} e^{\delta\gamma + \beta(x-\mu)},$$

where K_1 is a modified Bessel function of the second kind and $\gamma = \sqrt{\alpha^2 - \beta^2}$. Note that the generating function of the NIG distribution is given by

$$\psi(\theta) = \mu\theta + \delta(\gamma - \sqrt{\alpha^2 - (\beta+\theta)^2}),$$

and is not well defined for every $\theta \in \mathbb{R}$, so that we change the algorithm parametrization (see section 4.3 of [27]). The loss of the portfolio can be written $L = \varphi_4(X_T) = 50(e^{X_T} - K)_+ - e^{rT}C_0$. Note that the price C_0 is computed by a crude Monte Carlo and is approximately equal to 42. The parameters of the NIG random variable X_T are $\alpha = 2.0$, $\beta = 0.2$, $\delta = 0.8$, $\mu = 0.04$. We want to compare the variance reduction achieved by the translation of the mean (see section 3.1) and the one achieved by the Esscher Transform (see section 4.1). In the Robbins-Monro procedure, we define the step sequence by $\gamma_n = \frac{1}{n^\beta + 100}$ where $\beta = \frac{3}{4}$.

▷ *Translation case.* The functions L_3 and L_4 of the IS procedure are defined by:

$$\begin{aligned} L_3(\xi,\theta,X) &:= e^{-2|\theta|}\mathbf{1}_{\varphi(X-\theta)}\frac{p'(X-2\theta)}{p(X)}\left(\frac{p(X-\theta)}{p(X-2\theta)}\right)^2, \\ L_4(\xi,\mu,X) &:= \frac{e^{-2|\mu|}}{1+G(-\mu)+\xi^2}(\varphi(X-\mu)-\xi)_+^2\frac{p'(X-2\mu)}{p(X)}\left(\frac{p(X-\mu)}{p(X-2\mu)}\right)^2, \end{aligned}$$

where p' is easily obtained using the relation on the modified Bessel function $K_1'(x) = \frac{1}{x}K_1(x) - K_2(x)$.

▷ *Esscher Transform.* In this approach, the functions L_3 and L_4 are defined by

$$\begin{aligned} L_3(\xi,\theta,X) &:= \mathbf{1}_{\varphi(X^{(-\theta)})\geq\xi}(\nabla\psi(\theta) - X^{(-\theta)}), \\ L_4(\xi,\mu,X) &:= \frac{e^{-|\mu|}}{1+\xi^2}(\varphi(X^{(-\mu)})-\xi)_+^2(\nabla\psi(\mu) - X^{(-\mu)}), \end{aligned}$$

where $X^{(\pm\theta)} \sim \mathrm{NIG}(\alpha, \beta \pm \theta, \delta, \mu)$.
Table 4 compares the variance reduction ratios of the VaR_α and CVaR_α algorithms achieved by the translation of the mean (VR^{tr}_{VaR} and VR^{tr}_{CVaR}) and the one achieved by the Esscher Transform (VR^{es}_{VaR} and VR^{es}_{CVaR}).

Table 4: Example 4 Results

Number of steps	α	VaR	CVaR	$\mathrm{VR}^{tr}_{\mathrm{VaR}}$	$\mathrm{VR}^{tr}_{\mathrm{CVaR}}$	$\mathrm{VR}^{es}_{\mathrm{VaR}}$	$\mathrm{VR}^{es}_{\mathrm{CVaR}}$
10 000	95%	85.8	215.7	5	10	4.2	58.8
	99%	217	518	6	12	8	60
	99.5%	304	748	8	25	8.9	110
100 000	95%	87.2	215.1	5	12	4.5	60
	99%	218	521	5	12	8.2	70
	99.5%	303.5	747.8	7	30	12	100
500 000	95%	87.9	215.6	5	9	5	57
	99%	227	518.9	5.5	11.8	11.5	68
	99.5%	312.8	741.8	6	31	10	123

The IS procedure is very efficient when $\mathbb{P}(\varphi(X) \geq \xi^*) = 1 - \alpha$ is close to zero and becomes more and more efficient as α grows to 1. Even for the complex portfolio considered in Example 3, where X is a Gaussian vector with $d = 60$, it is possible to achieve a great variance reduction for both VaR_α and CVaR_α.

We observed that IS based on Esscher transform is well adapted to distributions with heavy tails (*i.e.* heavier tails than the normal distribution). It is therefore suitable when large values are more frequent than for the normal distribution, as it is the case when the vector X is a NIG random variable. Indeed, in this setting, the IS parameters modify the parameter β which controls the asymmetric shape of the NIG distribution. We think that the IS procedure by Esscher transform outperforms the IS procedure by mean translation when the IS parameter impacts on the symmetry of the distribution.

Concluding remarks.

In this article, we propose a recursive procedure to compute efficiently the Value-at-Risk and the Conditional Value-at-Risk using the same innovation for both procedures. In our approach, for a given risk level α, the VaR_α and the CVaR_α are estimated simultaneously by a regular RM algorithm. Ruppert and Polyak's averaging principle provides an asymptotically efficient procedure. The estimates satisfy a Gaussian CLT. However, due to the slow convergence of the global procedure since we are interested in rare events, the regular version of this algorithm cannot be used in practice. To speed-up and thus greatly reduce the number of scenarios, we devise an unconstrained adaptive IS procedure. The resulting procedure provides estimates that satisfy a CLT with minimal variances. To optimize the move to the critical risk area, the risk level α can be temporarily replaced by a slowly increasing level α_n (stepwise constant in practice) converging to α. This produces a VaR *companion procedure* $(\hat{\xi}_n)_{n\geq 1}$ that controls the IS change of measure parameters $(\hat{\theta}_n, \hat{\mu}_n)$. Numerically speaking, the resulting procedure converges efficiently and can drastically reduce the variance. It is possible to extend the methods to portfolio whose losses depend on a general diffusion process, using Girsanov transform to introduce a potentially infinite dimensional variance reducer. Finally, we aim at extending the method by implementing low-discrepancy sequences in our procedure instead of pseudo-random numbers. Preliminary numerical experiments showed a significant improvement of the convergence rate . This also raises interesting theoretical problems.

See [25] for some first theoretical results in that direction in a one-dimensional framework and [13] for further developments in higher dimensional setting.

References

[1] Arouna B. (2004). Adaptative Monte Carlo method, a variance reduction technique, *Monte Carlo Methods and Appl.*, **10**(1), p.1-24.

[2] Artzner P., Delbaen F., Eber J.-M. and Heath D. (1999). Coherent measures of risk, *Math. Finance*, **9**, p.203-228.

[3] Arouna B., Bardou O. (2004). Efficient variance reduction for functionals of diffusions by relative entropy, technical report, CERMICS-ENPC (France).

[4] Benveniste A., Métivier M. and Priouret P. (1987). *Algorithmes adaptatifs et approximations stochastiques : théorie et applications l'identification, au traitement du signal et la reconnaissance des formes*, Applications des Mathématiques, Masson, Paris, 367 p.

[5] Borkar V.S. (1997). Stochastic approximation with two time scales, *Systems Control Lett.*, **29**, p.291-294.

[6] Bouton C. (1998). Approximation gaussienne d'algorithmes stochastiques, *Annales de l'I.H.P.*, section B, **24**(1), p.131-155.

[7] Britten-Jones M. and Schaefer S.M. (1999). Non linear Value-at-Risk, *European Finance Review*, **2**, p.161-187.

[8] Duffie D. and Pan J. (2001). Analytical Value-At-Risk with Jumps and Credit Risk, *Finance and Stochastics*, **5**(2), p.155-180.

[9] Duflo M. (1997). *Iterative random models*, transl. from French, Springer-Verlag, 385p.

[10] Duflo M. (1996). *Algorithmes Stochastiques*, Springer, Berlin, 319p.

[11] Dufresne D. and Vázquez-Abad F.J. (1998). Accelerated simulation for pricing Asian options. *Proceedings of the 1998 Winter Simulation Conference*. Piscataway, NJ, IEEE Press, p.1493-1500.

[12] Egloff D. and Leippold M. (2007). Quantile estimation with adaptive importance sampling, *Electronic copy: http://ssrn.com/abstract=1002631.*

[13] Frikha N., PhD thesis, in progress.

[14] Fu M.C. and Su Y. (2000). Optimal importance sampling in securities pricing, *Journal of Computational Finance*, **5**(4), p.27-50.

[15] Glasserman P. and Heidelberger P. and Shahabuddin P. (2002). Portfolio Value-at-Risk with Heavy-Tailed Risk Factors, *Mathematical Finance*, **12**, p.239-270.

[16] Glasserman P. and Heidelberger P. and Shahabuddin P. (1999). Variance reduction techniques for estimating Value-at-Risk, *Management Science*, **47**, p.1349-1364.

[17] Glasserman P. and Heidelberger P. and Shahabuddin P. (1999). Importance Sampling and Stratification for Value-at-Risk, *Computational Finance 1999*, MIT press, p.7-24.

[18] Glasserman P. and Wang Y. (1997). Counterexamples in importance sampling for large deviation probabilities, *Annals of Applied Probability*, **7**(3), p.731-746.

[19] Juditsky A.B. and Polyak B.T. (1992). Acceleration of Stochastic Approximation by Averaging, *Journal on Control and optimization*, **30**(4), p.838-855.

[20] Kawai R. (2008). Optimal importance sampling parameter search for Lévy Processes via stochastic approximation, *SIAM Journal on Numerical Analysis*, **47**(1), p.293-307.

[21] Konda V.R. and Tsitsiklis J.N. (2004). Convergence rate of linear two-time-scale stochastic approximation, *Ann. Appl. Probab.*, **14**(2), p.796-819.

[22] Kroese D. P. and Rubinstein R. Y. (2004). *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte Carlo Simulation and Machine Learning*, Springer, New York, 300 p.

[23] Kushner H.J. and Clark D.S. (1978). *Stochastic Approximation Methods for Constrained and Unconstrained Systems*, Springer, New York, 276 p.

[24] Kushner H.J. and Yin G.G. (1993). *Stochastic Approximation with averaging of the iterates: Optimal asymptotic rate of convergence for general processes*, Springer, New York, **31**, p.1045-1062

[25] Lapeyre B., Pagès G., Sab K. (1990). Sequences with low discrepancy. Generalization and application to Robbins-Monro algorithm, *Statistics*, **21**(2), 251-272.

[26] Lelong J. (2007). Algorithmes stochastiques et Options parisiennes, *PhD thesis ENPC*, (France).

[27] Lemaire V. and Pagès G. (2008). Unconstrained Recursive Importance Sampling, To appear in Annals of Applied Probability.

[28] Ljung L. (1978). Strong convergence of a stochastic approximation algorithm, *Ann. Statist.*, **6**(3), p.680-696.

[29] Margrabe W. (1978). The Value of an Option to Exchange One Asset for Another. *Journal of Finance*, **33**(1), p.177-186.

[30] Mokkadem A. and Pelletier M. (2006). Convergence rate and averaging of non linear two-time-scale stochastic approximation algorithms, *The Annals of Applied Probability*, **16**(3), p.1671-1702.

[31] Pflug G.Ch. (2000). Some remarks on the value-at-risk and the conditional value-at-risk. In: Uryasev S. ed., *Probabilistic Constrained Optimization: Methodology and Applications*, Kluwer Academic Publishers, Dordrecht.

[32] Rockafellar R.T. and Uryasev S. (2002). Conditional Value-at-Risk for general loss distributions, *Journal of Banking and Finance*, **26**(7), p.1443-1471.

[33] Rockafellar R.T. and Uryasev S. (2000). Optimization of CVaR, *Journal of Risk*, **2**(3), p.21-41.

[34] Rouvinez C. (1997). Going Greek with VaR, *Risk*, **10**(2), p.57-65.

[35] Ruppert D. (1991). Stochastic Approximation. *Handbook of Sequential Analysis*, B. K. Ghosh and P.K. Sen, eds, p.503-529. Dekker, New York.

[36] Serfling R. J. (1980). *Approximation Theorems for Mathematical Statistics*, Wiley, New York.

[37] Uryasev S. (2000). Conditional Value-at-Risk: Optimization Algorithms and Applications, *Financial Engineering News*, **14**, p.1-5.