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-00348098v3

HAL Id: hal-00348098

https://hal.science/hal-00348098v3

Preprint submitted on 26 Feb 2009 (v3), 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.

Computation of VaR and CVaR using stochastic approximation and adaptive unconstrained importance sampling

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

December 2008

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 computing both VaR and CVaR using stochastic approximation (with decreasing steps): we propose a first Robbins-Monro procedure based on Rockaffelar-Uryasev's identity for the CVaR. The 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 importance sampling (I.S.) procedure which induces a significant variance reduction of both VaR and CVaR procedures. This idea, which goes back to the seminal paper [1], follows a new approach introduced in [22]. Finally, we consider a deterministic moving risk level to speed up the initialization phase of the algorithm. We prove that the convergence rate of the resulting procedure is ruled by a Central Limit Theorem with minimal variance and its efficiency is illustrated by considering several typical energy portfolios.

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

1 Introduction

Following financial institutions, energy companies are developping a risk management framework to face the new price and volatility risks associated to the growth of energy markets. VaR and 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 cases, Gaussian approximations can lead to semi-closed form estimators. But, if these approximations can be of some interest in considering the *yield* of a portfolio, they are useless when estimating the VaR on the EBITDA of a huge portfolio as it is often the case in the energy sector.

In this article, we introduce an alternative estimation method to compute both VaR and CVaR, relying on the use of stochastic algorithms. By definition, the VaR_α of a given portfolio at a specified level $\alpha \in (0,1)$ is the lowest amount not exceeded by the loss with probability α (usually $\alpha \geq 95\%$). The $CVaR_\alpha$ is the conditional expectation of the portfolio losses above the VaR_α. Several methods have been proposed to compute or approximate VaR and CVaR. To compute VaR, the simplest

[1]Research and Development Division, France, e-mail: olivier-aj.bardou@gdfsuez.com

[2]Research and Development Division, France, e-mail: noufel-externe.frikha@gdfsuez.com

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

method is the historical simulation. It assumes that 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 distribution of the loss $L := V(S_t,t) - V(S_t + \Delta S, T)$, where S_t denotes the vector of market prices at time t, ΔS the change in S over the time interval $T-t$ and $V(S_t,t)$ the portfolio value at time t, can be computed with the corresponding VaR at a given probability level by the inversion of the empirical function method. However, when market prices follow a general path-dependent SDE, the assumption of asset returns independence is no longer available. To circumvent this problem, Monte Carlo simulation tools are generally used. However, they require to store and sort loss samples in order to inverse the simulated empirical loss distribution function. Another method widely used relies on a linear (Normal approximation) or quadratic expansion (Delta Gamma approximation) and assume a joint normal (or log-normal) distribution of ΔS. The Normal approximation method gives L a normal distribution, thus the computation of the VaR_α is straightforward. However, when there is a nonlinear 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, one can refer to [6], [7], [12], [13] and [31]. Such approximations are no longer acceptable when considering portfolios with long maturity ($T-t \approx$1 year up to 10 years) or when the loss is a functional of a general path-dependent Stochastic Differential Equation (SDE).
In the context of hedging or optimizing a portfolio of financial instruments to reduce the CVaR, it is shown in [29] that it is possible to compute both VaR and CVaR (actually calculate VaR and optimize CVaR) by solving a linear programming problem which is an approximation of the original one. The advantage of such method is that it is possible to estimate both VaR and CVaR simulteanously and without assuming that the market prices have a special distribution like normal or log-normal. However, the linear programming problem is subject to a huge number of linear constraints (at least equal to the size of the generated loss samples) so that we are quickly limited in practice.
The idea to compute both with one procedure comes from the fact that the VaR and CVaR are strongly linked as they appear as the solutions and the value of the same convex optimisation problem (see Proposition 1.1) as demonstrated in [29]. This leads us to define consistent and asymptotically normal estimators of both quantities as the limit of a global Robbins-Monro (R.M.) procedure. So that we are no more constrained by the number of samples paths used in the estimation which in these approaches is the dimension of the resulting optimization problem. The great advantage of our point of vue, especially in regard to the inversion of the empirical function method is that we only estimate the quantities of interest. 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 very easy and efficient and as a necessary improvement, we also introduce a recursive variance reduction method based on an I.S. algorihm since both VaR and CVaR deal with rare events.
As a matter of fact, basically in this kind of problem we are interested by events that are observed with a very low probability (usually less that 5% or 1%) so that we obtain few significant replications on which to base our estimates. Actually, interesting losses are those that exceed the VaR, thus the ones that are in the tail of the loss distribution. So that, to compute more accurate estimates of both quantities of interest, it is crucial to generate more samples in the tail of L, the area of interest. A general tool used in this situation is importance sampling. The basic principle of importance sampling 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 importance sampling is to find a right change of measure (giving a parameterized family) that will

induce a significant variance reduction. Glasserman and al. in [13] and [14] proposed a change of measure based on a large deviation upper bound 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 importance sampling distribution it is strongly based on the fact that delta gamma approximation holds exactly and rely on the assumption that, conditional on past data, market moves are normally distributed. Moreover, as shown in [15], 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 [10], a quantile estimator based on the inversion of the empirical weighted function and as an improvement, they use a projected version (on convex compact set) of a Robbins-Monro I.S. algorithm to find the optimal change of measure. This kind of 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 specified properly. Our importance sampling procedure is based on a regular and unconstrained R.M. algorithm. One major issue that arises when combining the VaR-CVaR algorithm with the recursive importance sampling R.M. procedure is to ensure I.S. parameters do move appropriately toward the critical risk area. To circumvent this problem, we make the confidence level slowly move from $\alpha_0 = \frac{1}{2}$ to the final level α by introducing a deterministic confidence level sequence $(\alpha_n)_{n\geq 0}$. It speeds up the initialization phase and improves the variance reduction. Thus, we can truly experiment asymptotic convergence results in practice. Now, let us be more specific on the problem we are dealing with.

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.

Let $X : (\Omega, A, \mathbb{P}) \to \mathbb{R}^d$ be a random vector and $\varphi : \mathbb{R}^d \to \mathbb{R}$, $\Psi : \mathbb{R} \to \mathbb{R}$ be two Borel functions. Let $L = \varphi(X)$ be *the loss of a portfolio* over the considered time interval $T - t$. A positive loss is a loss sustained by the holder of the portfolio. Thus, φ is the function describing 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; thus 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 $\{\xi \mid \mathbb{P}(\varphi(X) \leq \xi) \geq \alpha\} \neq \emptyset$. 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 below so that the VaR always exists. We assume that the distribution function of $\varphi(X)$ is continuous (*i.e.* without atoms) so that it 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. 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:

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

The CVaR of $\varphi(X)$ is but 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 by computing the Ψ-*Conditional Value at Risk* (Ψ-CVaR) (at level α). As soon as $\Psi(\varphi(X)) \in L^1(\mathbb{P})$, it is defined by:

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

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

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 [28].

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

Proposition 1.1. *Suppose that φ is continuous. 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 $VaR_\alpha(\varphi(X))$ is any point of the set:

$$\arg\min V = \{\nabla V = 0\} = \{\xi \mid \xi\ , \mathbb{P}(\varphi(X) \leq \xi) = \alpha\}.$$

Furthermore,

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

*and, for every $\xi^* \in \arg\min V$ (*i.e. *a $VaR_\alpha(\varphi(X))$)*

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

Proof. Since the functions $\xi \mapsto (\varphi(x) - \xi)_+$ 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 differentiation Theorem, one can invert differentiation and expectation, 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, $V_\Psi(\xi^*) = \Psi\text{-}CVaR_\alpha(\varphi(X))$. □

Remark 1.1. Actually, one can 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.

Stochastic gradient and companion procedure

So in order to compute the VaR_α, we are interested in finding a zero of the non decreasing function V'. First note that V' admit the representation:

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

with

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

A classical procedure to compute such a zero is to consider the (recursive) R.M. algorithm defined by:

$$\xi_{n+1} = \xi_n - \gamma_{n+1}H(\xi_n, X_{n+1}), n \geq 0,\ \xi_0 \in L^1(\mathbb{P}), \tag{4}$$

where $(X_n)_{n\geq1}$ is an i.i.d. sequence of random variables with the same distribution as X, independent of ξ_0 and $(\gamma_n)_{n\geq1}$ is a step sequence (decreasing to 0) satisfying:

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

Actually, (4) can be seen as a regular R.M. procedure with mean function V' (see [9] p.50 and p.66) or as a recursive gradient descent procedure derived from the Lyapunov function V. Both settings yields to the *a.s.* convergence toward its target $\xi^* = VaR_\alpha(\varphi(X))$. We will temporarily assume this fact to carry on the design of the second step procedure: the computation of the $\Psi\text{-}CVaR_\alpha$. 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,

$$\frac{1}{n}\sum_{k=0}^{n-1} w(\xi^*, X_{k+1}). \tag{5}$$

However, we first need to get from (4) a good approximation of ξ^* and subsequently to use another sample of the distribution X. A natural idea is to devise a *companion procedure* of the above quantile search algorithm by replacing ξ^* in (5) by its approximation at step k, namely

$$C_n = \frac{1}{n}\sum_{k=0}^{n-1} w(\xi_k, X_{k+1}),\ n \geq 1,\ C_0 = 0. \tag{6}$$

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:

$$C_{n+1} = C_n - \frac{1}{n+1}\left(C_n - w(\xi_n, X_{n+1})\right),\ n \geq 0\ C_0 = 0. \tag{7}$$

In Proposition 2.2, it is shown that (7) *a.s.* converges toward $V_\Psi(\xi^*)$. 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 (7) consists in considering a general step β_n satisfying (A1) instead of $\frac{1}{n}$ (with in mind the possibility to set $\beta_n = \gamma_n$ eventually). This leads to:

$$C_{n+1} = C_n - \beta_{n+1}\left(C_n - w(\xi_n, X_{n+1})\right),\ n \geq 0\ C_0 = 0, \tag{8}$$

The resulting algorithm reads:

$$\begin{cases} \xi_{n+1} = \xi_n - \gamma_{n+1} H(\xi_n, X_{n+1}),\ \xi_0 \in L^1(\mathbb{P}),\ n \geq 0 \\ C_{n+1} = C_n - \beta_{n+1}\left(C_n - w(\xi_n, X_{n+1})\right),\ C_0 = 0,\ n \geq 0. \end{cases} \tag{9}$$

Our first aim in this paper is to study the almost sure convergence and the joint weak convergence rate of (ξ_n, C_n). The question is not trivial owing to the coupling of the two procedures. The case of two different step sizes $(\gamma_n) \equiv \gamma_0 n^{-b}$ and $(\beta_n) \equiv \beta_0 n^{-a}$ with $\gamma_0 > 0$, $\beta_0 > 0$ and $\frac{1}{2} < b < a \leq 1$ refers to the general framework of two-time-scale stochastic approximation algorithms. Several results have been established by Borkar in [4], Konda and Tsitsiklis [18] but the more relevant in our case are those of Mokkadem and Pelletier in [25]. It is possible to show that, if $\varphi(X)$ has a strictly positive density $f_{\varphi(X)}$ in a neighborhood of ξ^*, the couple satisifies a Gaussian CLT at the rate $\sqrt{\gamma_n^{-1}}$ for ξ_n and $\sqrt{\beta_n^{-1}}$ for C_n. The procedure for (ξ_n) is free of (C_n) so its convergence rate is ruled by the CLT for "regular" single-time scale stochastic approximation algorithms (see Kushner and Clark in [19], Métivier and Priouret in [3], Duflo in [9] and many others for more details) *i.e.* set $\gamma_n = \frac{\gamma_0}{n}$ where the choice of γ_0 depends on $f_{\varphi(X)}(\xi^*)$, which is unknown. The optimal choice for β_n is $\frac{1}{n}$. A posteriori, we verify that the resulting algorithm reduced to a one-time-scale procedure. 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 [32] and Polyak [16] and then widely investigated by several authors. It works in both two-time and single-time scale algorithm and leads to asymptotically efficient procedures *i.e.* satisfying a CLT at the optimal rate $\sqrt{n}$ and minimal variance. 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 and others procedure with different steps choice. So the final form of our procedure (without the variance reduction) is as follows:

set

$$\gamma_n = \beta_n = Cn^{-a},\ \text{with}\ \frac{1}{2} < a < 1, C > 0$$

and,

$$\overline{\xi}_n := \frac{1}{n}\sum_{k=1}^{n} \xi_k\ \text{ and }\ \overline{C}_n := \frac{1}{n}\sum_{k=1}^{n} C_k \tag{10}$$

where (ξ_k, C_k), $k \geq 0$ is defined by (9). Thanks to Ruppert and Polyak averaging principle, we obtain an asymptotically efficient procedure which satisfies the Gaussian CLT:

$$\sqrt{n}\begin{pmatrix} (\overline{\xi}_n - \xi^*) \\ (\overline{C}_n - C^*) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma) \tag{11}$$

where Σ is the asymptotic covariance matrix that will be explicited in Theorem 2.4. However, in pratice, the convergence of the algorithm will be very slow and especially chaotic for the VaR_α and so for the $CVaR_\alpha$. The bottleneck of this algorithm is the fact that it is only updated in 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 $\varphi(X)$. For instance, we can imagine large portfolio of complex derivative securities and options. Each evaluation may require a lot of computational time. So, for pratical implementation the above procedure must be combined with variance reduction techniques to achieve accurate results. An appropriate technique when we deal with rare events is importance sampling.

Unconstrained Recursive Importance Sampling

The second tool we want to introduce in this paper is a recursive importance sampling procedure which increases the probability of simulations for which $\varphi(X)$ exceeds ξ. 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 $\theta \in \mathbb{R}^d$,

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

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

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

If the following assumption

$$\forall \theta \in \mathbb{R}^d, \qquad \mathbb{E}\left[F^2(X)\frac{p(X)}{p(X-\theta)}\right] < +\infty \tag{12}$$

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

$$Q(\theta) = \mathbb{E}\left[F^2(X)\frac{p(X)}{p(X-\theta)}\right], \quad \theta \in \mathbb{R}^d. \tag{13}$$

Making few (natural) assumptions on p, one can show 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 [22]). Provided that ∇Q admits a representation as an expectation, then it is possible to devise a recursive Robbins-Monro procedure to approximate the optimal parameter θ^*. First investigated by Arouna (see [1]) in the Gaussian case, they used the natural representation of ∇Q as an expectation by formally differentiating (13) to design a stochastic gradient algorithm: $\nabla Q(\theta) = \mathbb{E}\left[K(\theta, X)\right]$. When $X = \mathcal{N}(0,1)$, one can show that $Q(\theta) = e^{\frac{|\theta|^2}{2}}\mathbb{E}[F^2(X)e^{-\theta X}]$ so that K is given by $K(x,\theta) = e^{\frac{|\theta|^2}{2}}F^2(x)e^{-\theta x}(\theta - x)$.

However, by this resulting form of K, the classical convergence results do not apply. Actually, the requested linear mean growth assumption is not fulfilled, *i.e.* $||K(\theta, X)||_2$ is not sub linear in θ as $|\theta|$ goes to infinity, inducing the explosion of the procedure at almost every implementation as pointed out in [1]. This lead the authors to introduce projected versions of the procedure based on repeated reinitializations when the algorithms exits from an increasing sequence of compact sets (while the step γ_n keeps going to 0). This approach is 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 [2], 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 [22] deeply revisited the original method to remove the constraints introduced by the previous algorithm. Thanks to *a third translation* of the variable θ, it is possible to plug back the parameter θ "into" $F(X)$, the function F having in common applications a known behaviour at infinity. We propose a regular and unconstrained importance sampling R.M. algorithm $(\theta_n(\xi_n), \mu_n(\xi_n))_{n\geq 1}$ as an application of [22] to the computation of both VaR and CVaR.

Our paper is now organized as follows. Section 2 is devoted to the exact design and to the study of its convergence rate. We firstly state our assumptions and give a result concerning the procedure without importance sampling; then, we introduce the adaptive variance reduction procedure and present how it modifies the asymptotic variance of our first CLT in Section 3. 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 I.S. procedure. To prevent the freezing of the algorithm during the first iterations of the importance sampling procedure, we also consider a deterministic moving risk level α_n which replace α in (4) 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 dramatically the convergence.

2 Study of the Var-CVaR stochastic approximation algorithm

In this section, we study the *a.s.* convergence and weak convergence rate of the one-time scale averaged procedure (10). The set $\arg\min V$, or more precisely, the random variables taking values in $\arg\min V$ will be the target of the first component of our (averaged) procedure $(\bar{\xi}_n)_{n\geq 0}$. If $\varphi(X)$ is strictly increasing then $\arg\min V = \{\xi^*\}$ is reduced to one single point. However, thanks to the next result, this will not be necessary to ensure the *a.s.* convergence of both $\bar{\xi}_n$ and $\bar{C}_n$. We will rely on this Theorem in order to prove that our different R.M. algorithms *a.s.* converge. For a proof of this slight extension of Robbins-Monro Theorem, we refer to [11]. The study of the CLT is postponed to the second subsection.

2.1 The a.s convergence of the regular procedure

Theorem 2.1. *(Extended Robbins-Monro Theorem). 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

$$\forall z \in \mathbb{R}^d \setminus \mathcal{T}^*, \forall z^* \in \mathcal{T}^*, \quad \langle z - z^*, h(z)\rangle > 0. \tag{14}$$

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

$$\forall z \in \mathbb{R}^d, \mathbb{E}[|H(z,X)|^2] \leq C(1+|z|^2) \tag{15}$$

(which implies that $|h(z)|^2 \leq C(1+|z|^2)$).
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 $\mathbb{E}[|\theta_0|] < \infty$, all defined on the same probability space $(\Omega, A, \mathbb{P})$. Let $\mathcal{F}_n := \sigma(\theta_0, X_1, ..., X_n)$ and let $(r_n)_{n\geq 1}$ be the $\mathcal{F}_n$-measurable reminder sequence satisfying

$$\sum \gamma_{n-1}|r_n|^2 < \infty.$$

Then, the recursive procedure defined by

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

satisfies:

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

The convergence also holds in $L^p(\mathbb{P}), p \in (0,2)$.

In order to apply this theorem for our purpose, we write (9) ($\gamma_n = \beta_n$) in a more synthetic way by setting $Z_n = (\xi_n, C_n)$ and:

$$\begin{cases} Z_n = Z_{n-1} - \gamma_{n-1} H(Z_{n-1}, X_n), n \geq 0 \\ Z_0 = (\xi_0, 0)\,, \;\; \xi_0 \in L^1(\mathbb{P}) \end{cases} \tag{16}$$

where $H(z,x) := (H_1(\xi,x), H_2(\xi,C,x)) = \left(1 - \frac{1}{1-\alpha}\mathbf{1}_{\{\varphi(x)\geq\xi\}}, C - w(\xi,x)\right)$ is a Borel function from $\mathbb{R}^2 \times \mathbb{R}^d$ to $\mathbb{R}^2$.

The *a.s.* convergence of the algorithm resulting from this first approach is established in the proposition below. It is only a first step toward our final procedure in which we will add several devices to speed up the convergence: the Ruppert-Polyak's averaging principle on the one hand and some adaptive importance sampling procedure on the other hand. Moreover, we make the following additional assumption on the distributions of $\varphi(X)$ and $\Psi(\varphi(X))$

$$\varphi(X) \text{ is continuous and } \Psi(\varphi(X)) \in L^{2a}(\mathbb{P}) \text{ for } a > 0. \tag{$(A2)_a$}$$

Proposition 2.2. *(Convergence of the algorithm) Suppose that Ψ is continuous for every $\xi^* \in \arg\min V$ as well as (A1) and $(A2)_1$ are satisfied.*
Then, the recursive procedure defined by (16) a.s. converges toward $z^* := (\xi^*, C^*)$ where ξ^* is a square integrable VaR_α-valued random variable and $C^* = \Psi\text{-}CVaR_\alpha(\varphi(X))$.

Proof. For the convergence of the first component $(\xi_n)_{n\geq 1}$ we check that the function H_1 and its mean $h_1(\xi) := \mathbb{E}[H_1(\xi,X)] = 1 - \frac{1}{1-\alpha}\mathbb{P}(\varphi(X) \geq \xi)$ satisfy the assumptions of the Robbins-Monro Theorem without a reminder sequence:

- The function h_1 clearly satisfies (14) since it is non-decreasing.
- The function H_1 is uniformly bounded since

$$\forall \xi \in \mathbb{R}, \; |H_1(\xi,X)| \leq 1 \vee \frac{\alpha}{1-\alpha}$$

  so that the linear quadratic growth assumption (15) is clearly satisfied.

Consequently, $\xi_n \xrightarrow{a.s.} \xi^*$ and in every L^p, $p \in [1,2)$. Now let us deal with the second component. Set for convenience $\gamma_0 := \sup_{n\geq 1} \gamma_n + 1$. then, one defines recursively a sequence $(\Delta_n)_{n\geq 1}$ by

$$\Delta_{n+1} = \Delta_n \frac{\gamma_{n+1}}{\gamma_n} \frac{\gamma_0}{\gamma_0 - \gamma_{n+1}}, \ n \geq 0, \ \Delta_0 = 1.$$

Then elementary computations show by induction that

$$\gamma_n = \gamma_0 \frac{\Delta_n}{S_n}, \ n \geq 0, \ \text{with } S_n = \sum_{k=0}^{n} \Delta_k. \tag{17}$$

Furthermore, it follows from (17) that for every $n \geq 1$

$$\log(S_n) - \log(S_{n-1}) = -\log\left(1 - \frac{\Delta_n}{S_n}\right) \geq \frac{\Delta_n}{S_n} = \frac{\gamma_n}{\gamma_0}.$$

Consequently,

$$\log(S_n) \geq \frac{1}{\gamma_0} \sum_{k=1}^{n} \gamma_k$$

which implies that $\lim_{n\to+\infty} S_n = +\infty$.

Now using (8) and (17), 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

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

by the Kronecker lemma since $\gamma_n = \gamma_0 \frac{\Delta_n}{S_n}$. The sequence $(N_n^\gamma)_{n\geq 1}$ is an $\mathcal{F}_n$-martingale since the ΔN_k's are martingales increments and

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

The continuity of Ψ at ξ^* and the $a.s.$ convergence of ξ_k toward ξ^* imply that

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

Consequently, assumption (A1) implies

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

which in term yields the $a.s.$ convergence of $(N_n^\gamma)_{n\geq 1}$, so that $C_n \xrightarrow{a.s.} \Psi\text{-}CVaR_\alpha\left(\varphi(X)\right)$. □

Remark 2.1. • When the VaR_α is unique, the convergence is ensured by the regular Robbins-Monro Lemma.

• It is possible to replace $w(\xi, x)$ in (8) and (9) by $\tilde{w}(\xi, x) = \frac{1}{1-\alpha}\Psi(\varphi(x))\mathbf{1}_{\{\varphi(x)\geq\xi\}}$ since $\mathbb{E}\left[\tilde{w}\left(\xi^*, X\right)\right] = \Psi\text{-}CVaR_\alpha(\varphi(X))$. Thus, we only have to change also the martingale increments sequence $(\Delta N_n)_{n\geq 1}$ by $\Delta\tilde{N}_n := \frac{1}{1-\alpha}\left(\mathbb{E}\left[\Psi(\varphi(X))\mathbf{1}_{\{\varphi(X)\geq\xi\}}\right]_{|\xi=\xi_{n-1}} - \Psi(\varphi(X_n))\mathbf{1}_{\{\varphi(X_n)\geq\xi_{n-1}\}}\right)$. This provides another procedure $\tilde{C}_n$ for the computation of the $\Psi\text{-}CVaR_\alpha$.

2.2 Rate of convergence

Throughout this section, we assume that the distribution 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 to its single target $(VaR_\alpha(\varphi(X)), \Psi\text{-}CVaR_\alpha(\varphi(X))$.

As a consequence, the Césaro mean of the procedure $\bar{Z}_n := \frac{Z_0+\cdots+Z_{n-1}}{n}$, $n \geq 1$, 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 asymptotic variance. We recall below this result (following a version established in [9]).

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*

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

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

$$h(z) = M(z - z^*) + O(|z - z^*|^2)$$

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 such that

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

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 $\{Z_n \to z^*\}$:

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

See [9] (p.169) for a proof.

In fact the regular algorithm satisfies a Central Limit Theorem. By "optimal" we mean the algorithm for which the step $\gamma_n = \frac{\gamma_0}{n}$ (with γ_0 a $d \times d$ matrix) has been optimized to minimize the asymptotic variance of the normalized procedure $\gamma_n^{-\frac{1}{2}}(Z_n - z^*)$. As expected, unfortunately this optimal step choice depends on the unknown density $f_{\varphi(X)}(\xi^*)$. Ruppert and Polyak's averaging principle allows us to achieve the same optimal asymptotic variance by averaging our regular R.M. procedure. See also a variant based on a gliding window developed in [21]. Applying this theorem to our framework yields the following result for the rate of convergence of our procedure.

Theorem 2.4. *(Convergence rate of the VaR-CVaR procedure) Suppose $(A2)_a$ (with $a > 1$) holds, Ψ is differentiable at ξ^* and that the density of $\varphi(X)$ is continuous and strictly positive at ξ^*. If the positive step sequence is $\gamma_n = \frac{\gamma_1}{n^a}$ with $\frac{1}{2} < a < 1$ and $\gamma_1 > 0$ then, the averaged procedure of the algorithm* (16) *satisfies the CLT:*

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

where Σ is given by

$$\begin{pmatrix} \frac{\alpha(1-\alpha)}{f^2_{\varphi(X)}(\xi^*)} & \frac{\alpha}{(1-\alpha)f_{\varphi(X)}(\xi^*)}\mathbb{E}\left[\left(\Psi(\varphi(X)) - \Psi(\xi^*)\right)\mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\right] \\ \frac{\alpha}{(1-\alpha)f_{\varphi(X)}(\xi^*)}\mathbb{E}\left[\left(\Psi(\varphi(X)) - \Psi(\xi^*)\right)\mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\right] & \frac{1}{(1-\alpha)^2}\,Var\left(\left(\Psi(\varphi(X)) - \Psi(\xi^*)\right)\mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\right) \end{pmatrix} \tag{19}$$

Proof. First, the procedure (16) can be written as

$$Z_n = Z_{n-1} - \gamma_n\left(h(Z_{n-1}) + \epsilon_n\right), n \geq 0,\ Z_0 = (\xi_0, 0),\ \xi_0 \in L^1(\mathbb{P}), \tag{20}$$

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 increments sequence with

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

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

$$h'(z^*) = M := \begin{pmatrix} \frac{1}{1-\alpha}f_{\varphi(X)}(\xi^*) & 0 \\ \mathbb{E}\left[\left(\frac{\partial}{\partial\xi}w(\xi,X)\right)_{|\xi=\xi^*}\right] & 1 \end{pmatrix} \tag{21}$$

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, $M = \begin{pmatrix} \frac{1}{1-\alpha}f_{\varphi(X)}(\xi^*) & 0 \\ 0 & 1 \end{pmatrix}$. To apply Theorem 2.3, we need to check assumptions (i)-(iv) (see (18)):

Let $A > 0$. First note that

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

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

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

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

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

It remains to check (iv) *i.e.* that $\mathbb{E}\left[\epsilon_{n+1}\epsilon_{n+1}^T|\mathcal{F}_n\right] \overset{a.s}{\to} \Gamma$, where Γ is a positive definite symmetric matrix. The dominated convergence theorem implies that

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

Using the continuity at ξ^* of the two functions: $\xi \mapsto \mathbb{E}\left[(\Psi(\varphi(X))-\Psi(\xi))\ \mathbf{1}_{\{\varphi(X)\geq\xi\}}\right]$, $\xi \mapsto \mathbb{E}\left[(\Psi(\varphi(X))-\Psi(\xi))^2\ \mathbf{1}_{\{\varphi(X)\geq\xi\}}\right]$ which follow from the continuity of Ψ and of the distribution function of $\varphi(X)$. Finally,

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

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

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

where $\Sigma = M^{-1}\Gamma\left(M^{-1}\right)^T$ is given by (19). □

Remark 2.2. • From a theoretical viewpoint the optimal rate of convergence of the procedure (20) is not impacted by the choice of γ_1 and a. In pratice, high or very small values of γ_1 will have as an effect to deteriorate the asymptotic variance at the usual range of simulations (say at most a few millions). Actually, if γ_1 is too high, the algorithm can freeze because it can take high values so that $\varphi(X)$ will not exceed ξ. Numerically, the "*CLT regime*" will take place later than with small steps once the algorithm is trapped. The algorithm spends most of the time "*exploring*" the state space before getting trapped.
• It is possible to show that if we choose $\beta_n = \frac{1}{n}$, $n \geq 1$ and $\gamma_n = \frac{1}{n^p}$ with $\frac{1}{2} < p < 1$ in (9), the resulting procedure satisfies a Gaussian CLT with the same asymptotic covariance matrix of the above theorem but not the same convergence rate. However, by averaging the first component ξ_n, the resulting procedure becomes asymptotically efficient.
• It is possible to show that if we replace C_n by the other procedure $\tilde{C}_n$ (see the remarks about Proposition 2.2) the procedure satisfies again a CLT with the same asymptotic covariance matrix.

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 (4). If (A2) is satisfied with $a = 2$, then on the convergence set $\{(Z_n) \to z^*\}$

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

and

$$\sqrt{n}\ \frac{C_n - C^*}{\sigma_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1). \tag{22}$$

Proof. The proof follows from standard arguments already used in the proof of Proposition 2.2 □

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 in practical implementations. 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 $CVaR_\alpha$ when $\alpha \approx 1$. Moreover, in the bank and energy sectors, practitioners usually deal with huge portfolio composed by hundreds or thousands of risk factors and options. The evaluation step of $\varphi(X)$ may require and consume a lot of computational time. Consequently, to achieve accurate estimates of both VaR_α and $CVaR_\alpha$ with reasonable computational effort, the above algorithm (16) drastically needs to be speeded up by an I.S. procedure to recenter simulation "where things do happen", *i.e.* scenarios for which $\varphi(X)$ exceeds ξ.

As we already said in the introduction, the main idea is to twist the distribution of X in order to minimize the asymptotic variance of the two components in the above CLT: the asymptotic variance of the VaR_α algorithm $\mathrm{Var}\left(\mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\right)$ and the asymptotic variance of $CVaR_\alpha$ algorithm $\mathrm{Var}\left(\left(\Psi(\varphi(X)) - \Psi(\xi^*)\right)\ \mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\right)$. It amounts to find the parameters θ_α^* and μ_α^* minimizing the two functionals:

$$Q_1(\theta,\xi^*) := \mathbb{E}\left[\mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\frac{p(X)}{p(X-\theta)}\right] \tag{23}$$

$$Q_2(\mu,\xi^*) := \mathbb{E}\left[\left(\Psi(\varphi(X)) - \Psi(\xi^*)\right)^2\mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\frac{p(X)}{p(X-\mu)}\right] \tag{24}$$

as soon as

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

Now if the density p function of X satisfies

$$\begin{cases} (i)\ \ \forall x \in \mathbb{R}^d,\ \theta \mapsto p(x-\theta) \text{ is log-}concave \\ (ii)\ \ \forall x \in \mathbb{R}^d,\ \lim_{|\theta|\to+\infty} p(x-\theta) = 0 \quad \text{or } \forall x \in \mathbb{R}^d,\ \lim_{|\theta|\to+\infty}\frac{p(x-\theta)}{p^2(x-\frac{\theta}{2})} = 0 \end{cases} \tag{B1}$$

and that,

$$Q_1(\theta,\xi) < +\infty,\ \forall\theta \in \mathbb{R}^d, \forall\xi \in \mathbb{R} \quad Q_2(\mu,\xi) < +\infty,\ \forall\mu \in \mathbb{R}^d, \forall\xi \in \mathbb{R}, \tag{B2}$$

then, one can show that the functions Q_1 and Q_2 are both finite, convex, goes to infinity at infinity. Consequently, the optimal parameters θ_α^* and μ_α^* exist since $\arg\min Q_1$ and $\arg\min Q_2$ are non empty (see [22] for more details). Recursive I.S. consists in devising a R.M. procedure to approximate the optimal parameters. Provided the gradients of the two target functions Q_1 and Q_2 admits a representation as an expectation, $\nabla Q_1(\theta,\xi^*) = \mathbb{E}[K_{Q_1}(\theta,\xi^*,X)]$ and $\nabla Q_2(\theta,\xi^*) = \mathbb{E}[K_{Q_2}(\mu,\xi^*,X)]$, one can easily devise two R.M. procedures by

$$\left\{\begin{array}{l} \theta_{n+1} = \theta_n - \gamma_{n+1} K_{Q_1}(\theta_n,\xi^*,X_{n+1}) \\ \mu_{n+1} = \mu_n - \gamma_{n+1} K_{Q_2}(\mu_n,\xi^*,X_{n+1}). \end{array}\right. \tag{25}$$

where K_{Q_1} and K_{Q_2} are naturally defined by the formal differentiation of Q_1 and Q_2

$$\left\{\begin{array}{l} \nabla Q_1(\theta,\xi^*) = \mathbb{E}\left[\mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\frac{p(X)}{p^2(X-\theta)}\nabla p(X-\theta)\right] \\ \nabla Q_2(\mu,\xi^*) = \mathbb{E}\left[(\Psi(\varphi(X))-\Psi(\xi^*))^2\,\mathbf{1}_{\{\varphi(X)\geq\xi^*\}}\frac{p(X)}{p^2(X-\mu)}\nabla p(X-\mu)\right]. \end{array}\right. \tag{26}$$

However, in our case, it is not possible to simulate or compute $K_{Q_1}(\mu_n,\xi^*,X)$ or $K_{Q_2}(\mu_n,\xi^*,X)$ by (25) since ξ^* is unknown. If one already has a good approximation of ξ^* and only wants to compute the optimal I.S. parameters (which seems unrealistic and uninteresting), it is possible to approximate θ_α^* and μ_α^* by devising directly the R.M. algorithm developed by Arouna (in the Gaussian framework) in its constrained form or the unconstrained R.M. procedure developed by Lemaire and Pagès in [22]. The main idea of our global procedure is to approximate both ξ^* and C^* and the optimal I.S. paramaters as the limit of an unconstrained I.S. algorithm coupled with the VaR_α procedure $(\xi_n)_{n\geq 0}$ using the *same innovation sequence* $(X_n)_{n\geq 1}$ for both procedure. The key idea of the unconstrained recursive I.S. algorithm developed in [22] is to introduce a third change of probability in order to control the annoying terms $\frac{p(x)}{p(x-\theta)}$ and $\frac{p(x)}{p(x-\mu)}$ by plugging back the parameters into $\mathbf{1}_{\{\varphi(X)\geq\xi\}}$ and $\Psi(\varphi(X))$ which has a controlled growth at infinity. To derive both new expressions for the gradients, we make the following assumption on the probability density p of X

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

Moreover, we need one more assumption on the function Ψ:

$$\forall C>0, \forall\xi\in\mathbb{R}, \mathbb{E}\left[(\Psi(\varphi(X))-\Psi(\xi))^2 e^{C|X|^{b-1}}\right] < +\infty. \tag{B4}$$

In [22], it is shown that as soon as (A3), (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

$$\nabla Q_1(\theta,\xi) := \mathbb{E}\left[\mathbf{1}_{\{\varphi(X-\theta)\geq\xi\}}\frac{p^2(X-\theta)}{p(X)p(X-2\theta)}\frac{\nabla p(X-2\theta)}{p(X-2\theta)}\right] \tag{27}$$

$$\nabla Q_2(\mu,\xi) := \mathbb{E}\left[(\Psi(\varphi(X-\mu))-\Psi(\xi))^2\;\mathbf{1}_{\{\varphi(X-\mu)\geq\xi\}}\frac{p^2(X-\mu)}{p(X)p(X-2\mu)}\frac{\nabla p(X-2\mu)}{p(X-2\mu)}\right]. \tag{28}$$

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

$$\frac{p^2(X-\theta)}{p(X)p(X-2\theta)}\frac{\nabla p(X-2\theta)}{p(X-2\theta)} = e^{\theta^2}(2\theta - X)$$

and more generally, if (B1) and (B3) hold, then there exist two constant A and B such that

$$\frac{p^2(x-\theta)}{p(x)p(x-2\theta)}\frac{|\nabla p(x-2\theta)|}{p(x-2\theta)} \leq e^{2\rho|\theta|^b}(A|x|^{b-1}+A|\theta|^{b-1}+B) \tag{29}$$

so that this weight can always be controlled by a deterministic function of θ (for more details, one can refer to [22]). Now, as soon as the growth of the function $x \mapsto \Psi(\varphi(x))$ at infinity can be explicitly controlled, it is possible to derive a regular Robbins-Monro procedure. For this, we introduce the following assumption where the function $\Psi o\varphi$ is bounded by a function F from $\mathbb{R}^d$ to $\mathbb{R}_+$ that satisfies a sub-multiplicative property and in which $c>0$ i a real constant

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

Now, we can devise our global procedure which consists in approximating both VaR_α and $CVaR_\alpha$ (in its I.S. version) with the optimal I.S. parameters. So, we are lead to devise the following algorithm, in which Z_n denotes the 4-tuple $(\xi_n, C_n, \theta_n, \mu_n)$,

$$Z_n = Z_{n-1} - \gamma_n L(Z_{n-1}, X_n) \tag{30}$$

where $(X_n)_{n\geq 1}$ is an i.i.d. sequence with the same distribution as X and

$$\begin{cases} L_1(\xi,\theta,x) := e^{-\rho|\theta|^b}\left(1-\frac{1}{1-\alpha}\,\mathbf{1}_{\{\varphi(x+\theta)\geq\xi\}}\frac{p(x+\theta)}{p(x)}\right) \\ L_2(\xi,C,\mu,x) := C - \bar{w}(\xi,\mu,x) \\ L_3(\xi,\theta,x) := e^{-2\rho|\theta|^b}\mathbf{1}_{\{\varphi(x-\theta)\geq\xi\}}\frac{p^2(x-\theta)}{p(x)p(x-2\theta)}\frac{\nabla p(x-2\theta)}{p(x-2\theta)} \\ L_4(\xi,\mu,x) := \frac{e^{-2\rho|\mu|^b}}{1+F(-\mu)^{2c}+\Psi(\xi)^2}\left(\Psi(\varphi(x-\mu))-\Psi(\xi)\right)^2\mathbf{1}_{\{\varphi(x-\mu)\geq\xi\}}\frac{p^2(x-\mu)}{p(x)p(x-2\mu)}\frac{\nabla p(x-2\mu)}{p(x-2\mu)} \end{cases} \tag{31}$$

with

$$\bar{w}(\xi,\mu,x) := \Psi(\xi) + \frac{1}{1-\alpha}(\Psi(\varphi(x+\mu))-\Psi(\xi))\,\mathbf{1}_{\{\varphi(x+\mu)\geq\xi\}}\frac{p(x+\mu)}{p(x)}. \tag{32}$$

Proposition 3.1. *(Efficient computation of VaR and CVaR) Suppose that $(A2)_a$ (with a=1), (A3), (B1), (B2), (B3), (B4) and (B5) are fulfilled and that the step sequence $(\gamma_n)_{n\geq 1}$ satisfies (A1). Then,*

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

where $(Z_n)_{n\geq 0}$ is the recursive algorithm defined by (30)*, where $C^* = \Psi\text{-}CVaR_\alpha$ and $(\xi^*, \theta^*_\alpha, \mu^*_\alpha) \in \mathcal{T}^* := \{(\xi,\theta,\mu)\ |\xi \in \arg\min V, (\theta,\xi) \in \{\nabla Q_1 = 0\}, (\mu,\xi) \in \{\nabla Q_2 = 0\}\}$.*

Proof. First we demonstrate that the 3-tuple (ξ_n, θ_n, μ_n) *a.s.* converges toward $(\xi^*, \theta^*_\alpha, \mu^*_\alpha)$ as $n \to +\infty$. Then, the convergence of $(C_n)_{n\geq 1}$ toward C^* will follow from standard arguments already used in the proof of Proposition 2.2. In order to apply the extended Robbins-Monro Theorem, we have to check the following facts:

• *Mean reversion*: the mean function of the three components procedure is defined by

$$l(z) := (\mathbb{E}\left[L_1(\xi,\theta,X)\right], \mathbb{E}\left[L_3(\xi,\theta,x)\right], \mathbb{E}\left[L_4(\xi,\mu,x)\right])$$

with,

$$\mathbb{E}\left[L_1(\xi,\theta,X)\right] = \left(1-\frac{1}{1-\alpha}\mathbb{P}\left(\varphi(X)\geq\xi\right)\right)e^{-\rho|\theta|^b},$$

$$\mathbb{E}\left[L_3(\xi,\theta,x)\right] = e^{-2\rho|\theta|^b}\nabla Q_1\left(\theta,\xi\right)$$

and

$$\mathbb{E}\left[L_4(\xi,\mu,x)\right] = \frac{e^{-2\rho|\mu|^b}}{1+F(-\mu)^{2c}+\Psi(\xi)^2}\nabla Q_2\left(\mu,\xi\right)$$

so that, $\{l=0\}=\mathcal{T}^*$ is such that $\forall z^*\in\mathcal{T}^*$ and $\forall z=(\xi,\ \theta,\ \mu)\in\mathbb{R}\times\mathbb{R}^d\times\mathbb{R}^d\setminus\mathcal{T}^*$

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

• *Linear growth:* Component by component it is easy to demonstrate that this condition is fulfilled:

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

Now, let f be the convex function defined on $\mathbb{R}^d$ by $f(x)=\log p(x)+\rho|x|^b$. Then, for every $x,\theta\in\mathbb{R}^d$,

$$\log\left(\frac{p^2(x)}{p(x-\theta)p(x+\theta)}\right)=2f(x)-f(x-\theta)-f(x+\theta)+\rho(|x+\theta|^a+|x-\theta|^a-2\,|x|^a).$$

Note that $x=\frac{(x-\theta)+(x+\theta)}{2}$, so that using the convexity of f and the following inequality (valid if $b\in(0,2]$)

$$\forall u,\ v\in\mathbb{R}^d,\quad |u|^b+|v|^b\le 2\left(\left|\frac{u+v}{2}\right|^b+\left|\frac{u-v}{2}\right|^b\right)$$

yields

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

so that

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

L_3 and L_4 are specially designed to fulfill this condition (for more details, one can refer to [22]). □

Now, we are interested by the rate of convergence of the two components procedure $(\bar{\xi}_n,\bar{C}_n)$. 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 that $(A2)_a$ with $a>1$, (A3), (B1), (B2), (B3), (B4) and (B5) are satisfied and that the positive step sequence is $\gamma_n=\frac{\gamma_1}{n^p}$ with $\frac{1}{2}<p<1$ and $\gamma_1>0$. Then, the empirical mean of the algorithm* (30) *satisfies the following local CLT:*

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

where

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

Proof. The proof is built on the same way that the one of Theorem 2.4. Since, by the definition of translation by mean, $h_1(z) = \mathbb{E}\left[L_1(\xi,\theta,\mu,X)\right] = E\left[H_1(\xi,X)\right]$ and $h_2(z) = \mathbb{E}\left[L_2(\xi,C,\theta,\mu,X)\right] = E\left[H_1(\xi,C,X)\right]$, the algorithm (30) can be written as

$$\begin{cases} Z_{n+1} = Z_n - \gamma_{n+1}\left(l(Z_n) + \tilde{\epsilon}_n\right), n \geq 0 \\ Z_0 = (\xi_0, 0)\,, \;\; \xi_0 \in L^1(\mathbb{P}) \end{cases} \tag{33}$$

where the two first components of h are the same function as the ones in the proof of Theorem 2.4 and $(\tilde{\epsilon}_n)_{n\geq 1}$ denotes the $\mathcal{F}_n$-adapted martingale increments sequence with

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

One can without difficulties demonstrate that the sequence $(\tilde{\epsilon}_n)_{n\geq 1}$ satisfies $(i)-(iv)$ of (18). □

Remark 3.1. • Actually, only a rough estimate of θ^*_α and μ^*_α is needed to reduce the asymptotic variance of the VaR-CVaR algorithm. That's why we don't specify the asymptotic variance of the I.S. procedure. one can replace the unconstrained I.S. parameters (θ_n,μ_n) by any sequences $(\tilde{\theta}_n)_{n\geq 0}$ $(\tilde{\theta}_0 \in \mathbb{R}^d)$ and $(\tilde{\mu}_n)_{n\geq 0}$ $(\tilde{\mu}_0 \in \mathbb{R}^d)$ such that θ_n and μ_n are $\mathcal{F}_n$-adapted and a.s. converge toward θ^*_α and μ^*_α and obtain the same Gaussian CLT.

• 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 I.S. 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 [22].

• According to (B3), if there exists a positive real number ρ such that $x \mapsto \log(p(x)) + \rho|x|^b$ is concave, then $\exists\; C > 0$, such that $p(x) \leq Ce^{\rho|x|^b}(|x|+1)$ which implies that the function F in (B5) satisfies $F(x) \leq C' e^{\frac{\lambda}{4c_e}|x|^e}$ for some $e \in (0,b)$, some $\lambda > 0$ and with $c_e = 1$ if $e \in [0,1]$, $c_e = 2^{\frac{e}{2}}$ if $e > 1$. Consequently, for pratical implementation, to control the growth of the I.S. parameters (especially when φ has a non-isotropic behaviour near infinity), it may be more efficient to implement the renormalized procedure, namely:

$$L_3\left(\xi,\theta,X\right) := \frac{e^{-2\rho|\theta|^b}}{1+||\theta||_2}\;\mathbf{1}_{\{\varphi(X-\theta)\geq\xi\}}\frac{p^2(X-\theta)}{p(X)p(X-2\theta)}\frac{\nabla p(X-2\theta)}{p(X-2\theta)}$$

and

$$L_4(\xi,\mu,X) := \frac{e^{-2\rho|\mu|^b-\frac{\lambda}{2c_e}|\mu|^e}}{1+||\mu||_2}\left(\Psi(\varphi(X-\mu)) - \Psi(\xi)\right)^2\;\mathbf{1}_{\{\varphi(X-\mu)\geq\xi\}}\frac{p^2(X-\mu)}{p(X)p(X-2\mu)}\frac{\nabla p(X-2\mu)}{p(X-2\mu)}$$

to prevent the algorithm to take high values during the first steps, thus controling the growth of the importance sampling parameters.

Now, let us point out an important issue. For the sake of simplicity, we consider the VaR-CVaR algorithm so that $\Psi \equiv Id$. The algorithm (30) raises an important problem numerically talking. It is due to the fact that basically, we are dealing with rare events to update the VaR-CVaR and the importance sampling procedures. Somehow, we have two algorithm ξ_n and (θ_n,μ_n) that are in competitive conditions, *i.e.* on one hand, we added an I.S. procedure to $(\xi_n)_{n\geq 1}$ to improve the convergence toward ξ^*, and on the other hand, the adjustment of the parameters (θ_n,μ_n) are based

on samples X_{n+1} satisfying $\varphi(X_{n+1}-\theta_n) > \xi_n$ and $\varphi(X_{n+1}-\mu_n) > \xi_n$ which are rare events. For the sake of simplicity, at the beginning of the global algorithm, there are two basic situations:

$\triangleright$ If we don't have any idea how to start the VaR procedure ($\xi_0 = 0$ for instance), at the very begining of the algorithm, the I.S. procedure is going to be updated according to samples that do not correspond to the tail of distribution. As a consequence, it may twist the distribution to an area that is not the good one and the parameters may remain "stuck" at the very begining of the I.S. procedure. This may lead to poor performance in terms of variance reduction.

$\triangleright$ If we have a relatively good approximation ξ_0 of ξ^*, we may observe few replications in which to base the adjustment of θ_n and μ_n. This may result again in a poor performance in terms of both VaR and CVaR estimation and variance reduction.

To circumvent this problem, we propose to break the link between the VaR and the I.S. procedure by introducing a second artificial VaR procedure that will lead the I.S. parameters into the critical risk area.

3.2 How to control the move to the critical risk area?

As already noted, to obtain good estimates and efficient variance reduction it is crucial to sample in the good direction. An idea in order to control the growth of θ_n and μ_n at the beginning of the algorithm, since we have no idea on how to twist the distribution of $\varphi(X)$, is to move slowly toward the critical risk area at level α in which $\varphi(X)$ exceeds ξ by replacing α by a deterministic sequence α_n that converges to α in (30). Since the algorithm for the CVaR component C_n is free of α, by doing so, we only modify the VaR computation procedure ξ_n. The function H_1 of (16) now depends of the current step of the procedure, namely $H_{1,n-1}(\xi_{n-1}, X_n) = 1 - \frac{1}{1-\alpha_{n-1}}\,\mathbf{1}_{\{\varphi(X_n)\geq\xi_{n-1}\}}$ and the VaR_α algorithm ξ_n becomes

$$\hat{\xi}_n = \hat{\xi}_{n-1} - \gamma_n H_{1,n-1}(\hat{\xi}_{n-1}, X_n), \;\; n \geq 0, \;\; \hat{\xi}_0 \in L^1(\mathbb{P}). \tag{34}$$

The sequence $\hat{\xi}_n$ is only designed to drive smoothly the new I.S. procedures $\hat{\theta}_n$, $\hat{\mu}_n$ toward the critical risk area. Consequently, we are neither going to use it in order to approximate ξ^* nor to compute C^* but only to "help" the unconstrained recursive I.S. algorithm. This new procedure differs slightly from the original procedure, *i.e.* if we define the reminder sequence $(r_n)_{n\geq 1}$ by $r_n := H_{1,n-1}(\hat{\xi}_{n-1}, X_n) - H_1(\hat{\xi}_{n-1}, X_n)$, $n \geq 1$, the resulting procedure can be written as,

$$\hat{\xi}_n = \hat{\xi}_{n-1} - \gamma_n (H_1(\hat{\xi}_{n-1}, X_n) + r_n), \;\; n \geq 0, \;\; \hat{\xi}_0 \in L^1(\mathbb{P}). \tag{35}$$

To obtain the convergence of the VaR-CVaR algorithm, we only have to check that $\hat{\xi}_n$ still converges toward a VaR_α-valued random variable. In practice, we plug a deterministic sequence $(\alpha_n)_{n\geq 1}$ that converges toward α which is stepwise constant to let the importance sampling parameters move toward the good direction.

Proposition 3.3. *(Convergence of the new procedure) Suppose that the assumptions of Proposition 2 are satisfied and that $(\alpha_n)_{n\geq 0}$ is a deterministic sequence of real numbers taking its value in $]0,1[$ which converges to α and such that*

$$\sum \gamma_n (\alpha_n - \alpha)^2 < +\infty. \tag{36}$$

Then, the recursive procedure defined by (35) *a.s. converges toward ξ^*.*

Proof. The reminder sequence (r_n) is $\mathcal{F}_n$-adapted and $\exists C > 0$ such that

$$|r_n| \leq C\,(\alpha_{n-1} - \alpha)\,, \;\; n \geq 0$$

so that,

$$\sum \gamma_n |r_n|^2 < +\infty.$$

Consequently, according to Robbins-Monro Theorem with a reminder part, the procedure still converges toward a random variable taking its value in $\{V' = 0\}$. □

The aim of this new VaR procedure is to drive and control the two importance sampling recursive procedures. Indeed, in order to define a new unconstrained I.S. procedure $(\hat{\theta}_n, \hat{\mu}_n)$, we only need to replace ξ^* by $\hat{\xi}_n$ into the importance sampling algorithm (30)

$$\begin{cases} \hat{\theta}_{n+1} = \hat{\theta}_n - \gamma_{n+1} L_3(\hat{\xi}_n, \hat{\theta}_n, X_{n+1}) \;\; n \geq 0, \;\; \theta_0 \in \mathbb{R}^d \\ \hat{\mu}_{n+1} = \hat{\mu}_n - \gamma_{n+1} L_4(\hat{\xi}_n, \hat{\mu}_n, X_{n+1}), \;\; n \geq 0, \;\; \mu_0 \in \mathbb{R}^d. \end{cases} \tag{37}$$

To establish the convergence of this new procedure, we rely on the Extended Robbins-Monro Theorem with reminder part. In the same way as the proof of Proposition 3.1, it is easy to see that if the same assumptions are fullfiled and that the assumption on the sequence $(\alpha_n)_{n\geq 1}$ (36) holds then the mean reversion and the linear growth assumptions are satisfied. Finally, the procedure we use to estimate the couple (ξ^*, C^*) is the algorithm (30) with $(\hat{\theta}_n, \hat{\mu}_n)$ for the importance sampling parameters. Let us note that since the sequence $\left(\hat{\theta}_n, \hat{\mu}_n\right)$ is $\mathcal{F}_n$-adapted and converges towards $(\theta^*_\alpha, \mu^*_\alpha)$, the procedure (30) (with $\hat{\theta}_n$ and $\hat{\mu}_n$) satisfies the same CLT.

4 Additional remarks

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 [17] and [22] to extend the constrained importance 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 [22] and [11]). 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

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

The two functionals to be minimized are

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

and

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

According to Proposition 3 in [22] as soon as ψ satisfies (H^{es}_δ) and that,

$$\forall \xi \in \mathbb{R}, \forall \theta \in \mathbb{R}^d, \quad \mathbb{E}[|X| e^{\langle \theta, X\rangle}] < +\infty \quad \text{and} \quad \mathbb{E}[|X| \Psi(\varphi(X))^2 e^{\langle \theta, X\rangle}] < +\infty \tag{38}$$

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,$\forall\xi \in \mathbb{R}$ their gradients are given by

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

and,

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

with $\nabla\psi(\theta) = \frac{\mathbb{E}[Xe^{\langle\theta,X\rangle}]}{\mathbb{E}[e^{\langle\theta,X\rangle}]}$. Now, the main result of this section is the following theorem (for more details, see [22] and [11]).

Theorem 4.1. *Suppose that ψ satisfies (H_δ^{es}) and that $(A2)_1$, (A3) hold. Assume that (38) is fullfiled and that*

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

Then, the recursive procedure defined by

$$Z_{n+1} = Z_n - \gamma_{n+1}L(Z_n, X_{n+1}), \quad n \geq 0, Z_0 = (\xi_0, C_0, \theta_0, \mu_0) \tag{39}$$

where $(\gamma_n)_{n\geq 1}$ satisfies the usual step assumption (A1), $Z_n := (\xi_n, C_n, \theta_n, \mu_n)$ and L can be written

$$L(Z_n, X_{n+1}) := \left(L_1\left(\xi_n, \theta_n, X_{n+1}^{(\theta_n)}\right), L_2\left(\xi_n, C_n, \mu_n, X_{n+1}^{(\mu_n)}\right), L_3\left(\xi_n, \theta_n, X^{(-\theta_n)}\right), L_4\left(\xi_n, \mu_n, X^{(-\mu_n)}\right)\right)$$

each component being defined by

$$\begin{cases} L_1(\xi,\theta,x) := 1 - \frac{1}{1-\alpha}\mathbf{1}_{\{\varphi(x)>\xi\}}\ e^{\psi(\theta)-\langle x,\theta\rangle}, \\ L_2(\xi,C,\mu,x) := C - \bar{w}(\xi,\mu,x), \\ L_3(\xi,\theta,x) := \mathbf{1}_{\{\varphi(x)\geq\xi\}}(\nabla\psi(\theta) - x), \\ L_4(\xi,\mu,x) := e^{-\frac{\lambda}{2}\sqrt{d}|\nabla\psi(-\mu)|}(\Psi(\varphi(x)) - \Psi(\xi))^2\mathbf{1}_{\{\varphi(x)>\xi\}}(\nabla\psi(\mu) - x), \end{cases}$$

and where $\bar{w}(\xi,\mu,x) := \Psi(\xi) + \frac{1}{1-\alpha}(\Psi(\varphi(x)) - \Psi(\xi))\mathbf{1}_{\{\varphi(x)>\xi\}}\ e^{\psi(\mu)-\langle\mu,x\rangle}$ a.s. converges toward $z^* := (\xi^*, C^*, \theta_\alpha^*, \mu_\alpha^*)$, ξ^* being a square integrable VaR_α-valued random variable, $C^* = \Psi\text{-}CVaR_\alpha(\varphi(X))$, θ_α^* being an $\arg\min Q_1(.,\xi^*)$-valued (square integrable) random vector and μ_α^* being an $\arg\min Q_2(.,\xi^*)$-valued (square integrable) 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 which were taken in the past. In Gaussian frameworks, X is a Gaussian vector. However in more sophisticated models or portfolio X can be a vector of Brownian increments related to the Euler scheme of a diffusion. 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 example, let's take 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 vectors of dimension d and we have to control the growth of each component. If one grows too quickly 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 importance sampling 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

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

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

$$dX_t = b(X_t)\ \mathrm{d}t + \sigma(X_t)\ \mathrm{d}W_t,\ X_0 = x \in \mathbb{R}^d, \tag{40}$$

$W = (W_t)_{t\in[0,T]}$ being a standard Brownian motion. A naive approach is to discretize (40) 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 to be poor for this kind of options. In this case, our I.S. parameters θ and μ are n-dimensional vectors which corresponds 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 I.S. procedure is no longer acceptable and tractable. To overcome this problem, the idea is to shift the entire distribution of X in (40) thanks to Girsanov transform. This last case is analyzed and investigated in [22]. It can be adapted to our framework (it is developed in [11]).

5 Numerical examples

For the sake of simplicity, we focus in this section on the finite dimensional setting and on the computation of the $CVaR_\alpha$ ($\Psi \equiv Id$). We consider first the usual Gaussian framework in which the exponential change of measure and translation by mean are the same, then we illustrate the algorithm (39) in a simple case.

5.1 Gaussian framework

In this setting, $X = \mathcal{N}(0, I_d)$ and its density is given on $\mathbb{R}^d$ 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

$$L_3(\xi, \theta, x) := \frac{1}{1 + ||\theta||_2} \mathbf{1}_{\{\varphi(X-\theta) \geq \xi\}} (2\theta - x)$$

and

$$L_4(\xi, \mu, x) := \frac{e^{-\frac{\lambda}{2}|\mu|^e}}{1 + \varphi^2(-\mu)} (\varphi(X - \mu) - \xi)_+^2 (2\mu - x).$$

Moreover, we use a stepwise constant sequence α_n that converges slowly toward α. For instance, if $\alpha = 0.95$, we firstly set $\alpha_n = 0.5$ during 5000-10000 replications to let $\hat{\theta}_n$ and $\hat{\mu}_n$ move towards the critical risk area, then we set $\alpha_n = 0.8$ for 5000-10000 replications and then finally set $\alpha_n = 0.95$. We consider three different portfolios of options (puts and calls) on 1 and 5 underlying assets except for the last case which is particular. 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 motion for the first two cases. In the third case, the assets (electricity and gaz day-ahead prices) are modeled as exponential of an Ornstein-Uhlenbeck process. This last derivative is priced using an approximation of Margrabe formulae. Each underlying assets are assumed uncorrelated except in the last case. We assume an annual risk free interest rate of 5%. In each case, we use three different values of loss probability $P(\varphi(X) \geq \xi^*) = 1 - \alpha = 5\%,\ 1\%,\ 0.5\%$. This probability is specified in the tables. We use the following test portfolios:

$\triangleright$ Short position in 1 put with strike $K = 110$ and maturity $T = 1$ year on a stock whose initial price is $S_0 = 100$ and volatility $\sigma = 20\%$.

$$\varphi^1(X) := (K - S_T)_+ - e^{rT} P_0$$

with

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

where $X = \mathcal{N}(0, 1)$ and P_0 is the initial price for which we sell the put option.

$\triangleright$ Short positions in 10 calls and 10 puts on each of 5 underlying assets, all options having the same maturity 0.25 years with several strikes, each underlying assets having a volatility of 20% and are assumed to be uncorrelated.

$\triangleright$ 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's and gas's initial spot prices are $S_0^e = 40\$/MWh$ and $S_0^g = 3\$/MMBtu$ with a Heat Rate equals $h_R = 10Btu/kWh$ and generation costs $C = 5\$/MWh$. The two spot prices have a correlation of 0.4. The payoff can be written

$$\varphi^3(X) = \sum_{k=1}^{30} \left(e^{r(T-t_k)} \left(S_{t_k}^e - h_R S_{t_k}^g - C\right)_+ - P_0^c e^{rT} \right) + \left(e^{rT} C_0 - e^{r(T-t_k)} \left(S_{t_k}^e - K\right)_+ \right)$$

This is a sum of spark spread options where we decide to exchange gas and electricity each day during one month.

From a theoretical point of view, 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, it is possible to break the algorithm into two parts. Firstly, we can compute a rough estimate of the

optimal I.S. parameters $\left(\hat{\theta}_M, \hat{\mu}_M\right)$ with a number of steps $M < N$ (compared with N, M can be small) and in a second time, compute the VaR_α and the $CVaR_\alpha$ with the resulting I.S. parameters estimate.
The results displayed in the following tables corresponds to Value-at-Risk, Conditional Value-at-Risk, variance reduction ratios for both VaR and CVaR procedure for three different values of α: $\alpha = 95\%$, $\alpha = 99\%$, $\alpha = 99.5\%$. Variance ratios corresponds to the ratio of variances using the averaged procedure of (16) and using (30). The results indicate that the I.S. procedure yields greater variance reduction, especially when α is closed to 1. The variance ratios in the table are estimated using different number of steps for the algorithm.

We observed that the unconstrained adaptive I.S. procedure converges very quickly. Only $10000 - 20000$ steps are needed to obtain a good estimate of θ^*_α and μ^*_α. Then, one can freeze those two parameters and let $(\bar{\xi}_n, \bar{C}_n)_{n\geq 1}$ converges.
The results indicate that the variance reduction achieved in estimating the *Conditional Value-at-Risk* is greater than the one achieved in estimating the *Value-at-Risk*. This can be explained by the presence of the factor $\frac{1}{f_{\varphi(X)}(\xi^*)}$ in the asymptotic variance of (19). This factor can be very large, especially when ξ^* is in the tail of the distribution, since the density may be close to zero whereas this factor doesn't appear in the asymptotic variance of the $CVaR$ procedure.

Table 1: Portfolio 1 Results

Number of steps	α	VaR	CVaR	RV_{VaR}	RV_{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%	34.18	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: Portfolio 2 Results

Number of steps	α	VaR	CVaR	RV_{VaR}	RV_{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

5.2 Esscher transform: the NIG distribution

Now, we consider a simple case of a portfolio composed by a long position on a Call option with strike $K = 0.6$ and maturity $T = 1$ *year*, where the underlying X_T ($X_0 = 0$) is a normal inverse

Table 3: Portfolio 3 Results

Number of steps	α	VaR	CVaR	RV_{VaR}	RV_{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

Gaussian variable $X_T \approx NIG(\alpha, \beta, \delta, \mu)$, $\alpha > 0$, $|\beta| \leq \alpha$, $\delta > 0$, $\mu \in \mathbb{R}$. 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 [22]). 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).

$\triangleright$ *Translation case.* The functions H_3 and H_4 of the I.S. procedure are defined by:

$$H_3(\xi, \theta, X) := \frac{e^{-2|\theta|}}{1 + |\theta|} \mathbf{1}_{\varphi(X-\theta)} \frac{p'(X - 2\theta)}{p(X)} \left(\frac{p(X - \theta)}{p(X - 2\theta)} \right)^2,$$

and

$$H_4(\xi, \mu, X) := \frac{e^{-2|\mu|}}{1 + |\mu|} (\varphi(X - \mu) - \xi)_+^2 \frac{p'(X - 2\mu)}{p(X)} \left(\frac{p(X - \mu)}{p(X - 2\mu)} \right)^2,$$

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

$\triangleright$ *Esscher Transform.* In this approach, the functions H_3 and H_4 are defined by

$$H_3(\xi, \theta, X) := \frac{e^{-|\theta|}}{1 + |\theta|} \mathbf{1}_{\varphi(X^{(-\theta)}) \geq \xi} (\nabla\psi(\theta) - X^{(-\theta)}),$$

and

$$H_4(\xi, \mu, X) := \frac{e^{-|\mu|}}{1 + |\mu|} (\varphi(X^{(-\mu)}) - \xi)_+^2 (\nabla\psi(\mu) - X^{(-\mu)}),$$

where $X^{(\pm\theta)} \approx NIG(\alpha, \beta \pm \theta, \delta, \mu)$.

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 $\alpha = 2.0$, $\beta = 0.2$, $\delta = 0.8$, $\mu = 0.04$. Table 4 compares

Table 4: Portfolio 4 Results

Number of steps	α	VaR	CVaR	RV^{tr}_{VaR}	RV^{tr}_{CVaR}	RV^{es}_{VaR}	RV^{es}_{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 variance ratios of the VaR_α and $CVaR_\alpha$ algorithms achieved by the translation of the mean (RV^{tr}_{VaR} and RV^{tr}_{CVaR}) and the one achieved by the Esscher Transform (RV^{es}_{VaR} and RV^{es}_{CVaR})

The I.S. procedure is very efficient when $\mathbb{P}(\varphi(X) \geq \xi^*) = 1 - \alpha$ is close to zero and becomes more and more efficient that α grows to 1. Even for the complex portfolio (3) where X is a Gaussian vector with $d = 60$, it is possible to provide good estimate with a great variance reduction for both VaR_α and $CVaR_\alpha$.

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 procedure. In our approach, for a given risk level α, the VaR_α and the $CVaR_\alpha$ are selected simultaneously by a regular Robbins-Monro algorithm. Ruppert and Polyak's averaging principle provides an asymptotically efficient procedure which satisfies a Gaussian CLT. However, due to the slow and chaotic convergence of the VaR component since we are interested by 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 recursive I.S. procedure. Then, at each step of the regular procedure, we plug the approximation obtained by the I.S. algorithm. The resulting procedure satisfies a CLT with minimal variances. To optimize the move to the critical risk area, the risk level α can be replaced by a deterministic sequence that converges slowly (stepwise constant in practice) to α. This produces a new VaR algorithm $(\hat{\xi}_n)_{n\geq 1}$ that controls the parameters (θ_n, μ_n). Numerically speaking, the resulting procedure converges efficiently and can greatly reduce variance. In current work, we extend the methods to portfolio whose losses depend of general path-dependent diffusion. This framework has particularly practical relevance. Numerical tests show that the VaR-CVaR procedure with I.S. is robust and is effective. Moreover, we are extending the method to low-discrepancy sequences instead of pseudo-random numbers into our procedure. This setting raises very interesting theoretical problems (for more details, see [11]).

References

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

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

[3] Benveniste A., Metivier M. and Priouret P., (1990). Adaptative Algorithms and Stochastic Approximations, **22**, *Applications of Mathematics*, transl. from French by S. Wilson, Springer-Verlag, Berlin.

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

[5] Bouton C., (1998). Approximation gaussienne d'algorithmes stochastiques, *The Annals of I.H.P.*, section B, **24**(1), p. 131-155.

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

[7] Duffie D. and Pan J., (2001). Analytical value-at-risk with jumps and credit risk, *Finance and Stochastics*, **5**(2), p. 155-180.

[8] Duflo M., (1997). Iterative random models, transl. from French, Springer-Verlag.

[9] Duflo M., (1996). Algorithmes Stochastiques, Springer, Berlin,

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

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

[12] Glasserman P. and Heidelberger P. and Shahabuddin P., (2002). Portfolio value-at-risk with heavy-tailed risk factors, *Mathematical Finance*, **9**, p.117-152.

[13] Glasserman P. and Heidelberger P. and Shahabuddin P., (1999). Variance reduction techniques for estimating value-at-risk, *Management Science*, **46**, p. 1349-1364.

[14] Glasserman P. and Heidelberger P. and Shahabuddin P., (2000). Importance sampling and Stratification for value-at-risk, *Computational Finance 1999*, MIT press.

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

[16] Juditsky A.B and Polyak B.T., (1992). Acceleration of stochastic approximation by averaging. *Control and optimization*, **30**, p. 838-855.

[17] Kawai R., (2008). Optimal importance sampling parameter search for Lévy Processes via stochastic approximation, pre-print.

[18] 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.

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

[20] Kushner H.J. and Yin G.G., (1997). Stochastic Approximation with averaging of the iterates: Optimal asymptotic rate of convergence for general processes, *Springer*, New York.

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

[22] Lemaire V. and Pagès G., (2008). Unconstrained Recursive Importance Sampling, *preprint*, LPMA-1231 ,UPMC (France).

[23] Lamberton D. and Pagès G. and Tarres P., (2004). When can the two-armed bandit algorithm be trusted?, *Ann. of Appl. Prob.*.**14**(3), 1424-1454.

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

[25] 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.

[26] 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.

[27] Revuz D. and Yor M., (1998). Continuous martingales and Brownian motion, *3rd edition, Springer*, Berlin, 1st edition, (1990).

[28] 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.

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

[30] Rogers L.C.G. and Williams D., (1986). Diffusions, Markov Processes and Martingales. *Cambridge Mathematical Library*, 2^{nd}edition.

[31] Rouvinez C., (1997). Going Greek with VAR, *Risk* **10** p. 57-65.

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

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