Multilevel Stochastic Approximation of the Value-at-Risk and Expected Shortfall

Azar Louzi (joint work with Stéphane Crépey, Noufel Frikha, Gilles Pagès and Jonathan Spence)
LPSM, CNRS, Université Paris Cité.

Preprints:
arXiv:2304.01207 [q-fin.CP]
arXiv:2311.15333 [q-fin.RM]

ICCF 2024
April 5, 2024. Amsterdam, Netherlands

LABORATOIRE DE PROBABILITÉS STATISTIQUE & MODÉLISATION

- Consider **a loss** X defined at some time horizon τ by

$$X = \mathbb{E}[\psi(Y, Z)|Y] = \Psi(Y),$$

with Y the realization of risk factors up to τ, $\perp\!\!\!\perp$ Z the realization of risk factors beyond τ.
- **The VaR** $\xi_\star$ and **the ES** $\chi_\star$ of **the loss** X at **the confidence level** $\alpha \in (0, 1)$, are given by

$$\boxed{\mathbb{P}(X \leq \xi_\star) = \alpha, \quad \chi_\star = \mathbb{E}[X|X > \xi_\star].}$$

- A **stochastic approximation (SA)** point of view is adopted.
- If Ψ **is analytical**, then $X = \Psi(Y)$ can be **simulated exactly** (Bardou, Frikha, and Pagès, 2009).
- **Otherwise,** X can be **simulated by nested Monte Carlo** (Barrera et al., 2019).
- But **the nested Monte Carlo method is costly.**
- Crépey, Frikha, and Louzi, 2023 propose **a multilevel acceleration** of the nested approach.

According to Rockafellar and Uryasev, 2000, if F_X is increasing,

$$(\xi_\star, \chi_\star) = (\arg\min, \min) V, \quad \text{where} \quad V(\xi) = \xi + \frac{\mathbb{E}[(X-\xi)^+]}{1-\alpha}.$$

- **Robbins-Monro (SA) Algorithm.** We take **the step size (learning rate)**

$$\gamma_n = \frac{\gamma_1}{n^\beta}, \quad \gamma_1 > 0, \quad \beta \in (0,1].$$

We define the estimators $(\xi_n, \chi_n)_{n \geq 0}$ of $(\xi_\star, \chi_\star)$ by

$$\xi_{n+1} = \xi_n - \gamma_{n+1} H_1(\xi_n, X^{(n+1)}) \quad \text{and} \quad \chi_{n+1} = \chi_n - \frac{1}{n+1} H_2(\chi_n, \xi_n, X^{(n+1)}),$$

where $\xi_0 \perp\!\!\!\perp (X^{(n)})_{n \geq 1} \overset{\text{iid}}{\sim} X = \Psi(Y)$ s.t. $\mathbb{E}[|\xi_0|^2] < \infty$, $\chi_0 = 0$,

$$H_1(\xi, x) = 1 - \frac{\mathbf{1}_{x > \xi}}{1-\alpha} \quad \text{and} \quad H_2(\chi, \xi, x) = \chi - \Big(\xi + \frac{(x-\xi)^+}{1-\alpha}\Big).$$

As suggested by Bardou, Frikha, and Pagès, 2009, for **a prescribed accuracy $\varepsilon > 0$,** we have to take

$$n = \lceil \varepsilon^{-\frac{2}{\beta}} \rceil, \quad \text{i.e.} \quad \mathrm{Cost}^\beta_{\mathsf{SA}} = Cn = C\varepsilon^{-\frac{2}{\beta}} \underset{\beta \to 1}{\searrow} C\varepsilon^{-2}, \quad \text{if} \quad \gamma_1 > \lambda \quad \text{when} \quad \beta = 1,$$

where $\lambda > 0$ **is a constant that is explicit but tedious to compute.**

- **Averaged Stochastic Approximation (ASA) Algorithm.**

To overcome the constraint on γ_1, we average out the VaR estimators (Bardou, Frikha, and Pagès, 2009)

$$\overline{\xi}_n = \frac{1}{n}\sum_{k=1}^{n}\xi_k = \overline{\xi}_{n-1} - \frac{1}{n}\big(\overline{\xi}_{n-1} - \xi_n\big).$$

To achieve **a prescribed accuracy $\varepsilon > 0$,** we must take

$$n = \lceil \varepsilon^{-2} \rceil, \quad \text{i.e.} \quad \boxed{\mathrm{Cost}_{\mathsf{ASA}} = Cn = C\varepsilon^{-2}, \quad \beta \in \Big(\frac{1}{2}, 1\Big).}$$

However, these approaches are often not possible, since **one does not always have an analytical formula for Ψ to directly simulate $X = \Psi(Y)$.**

Outline

Following Barrera et al., 2019, we approximate $X = \mathbb{E}[\psi(Y,Z)|Y]$ by

$$\boxed{X_h = \frac{1}{K}\sum_{k=1}^{K}\psi(Y, Z^{(k)}),} \quad h = \frac{1}{K} \text{ a bias parameter } \in \mathcal{H} = \left\{\frac{1}{K'}, K' \in \mathbb{N}^*\right\},$$

where $(Z^{(k)})_{1\le k\le K} \overset{\text{iid}}{\sim} Z \perp\!\!\!\perp Y$.

We define the approximating optimization problem

$$\boxed{(\xi^h_\star, \chi^h_\star) = (\arg\min, \min) V_h, \quad \text{where} \quad V_h(\xi) = \xi + \frac{\mathbb{E}[(X_h - \xi)^+]}{1-\alpha}.}$$

- **Nested SA (NSA) Algorithm.** We devise the stochastic approximation scheme

$$\boxed{\xi^h_n = \xi^h_{n-1} - \gamma_n H_1(\xi^h_{n-1}, X^{(n)}_h), \quad \chi^h_n = \chi^h_{n-1} - \frac{1}{n} H_2(\chi^h_{n-1}, \xi^h_{n-1}, X^{(n)}_h),}$$

where $\xi^h_0 \perp\!\!\!\perp (X^{(n)}_h)_{n\ge1} \overset{\text{iid}}{\sim} X_h$ with $\mathbb{E}[|\xi^h_0|^2] < \infty$ and $\chi^h_0 = 0$.

Global Error = Statistical Error + Bias Error

$$\xi_n^h - \xi_\star = (\xi_n^h - \xi_\star^h) + (\xi_\star^h - \xi_\star) \quad \text{and} \quad \chi_n^h - \chi_\star = (\chi_n^h - \chi_\star^h) + (\chi_\star^h - \chi_\star).$$

Assumption 1. $F_{X_h}(\xi) - F_X(\xi) = v(\xi)h + \mathrm{o}(h)$ as $\mathcal{H} \ni h \downarrow 0$.

Proposition 1 (Bias Error).

$$\xi_\star^h - \xi_\star = -\frac{v(\xi_\star)}{f_X(\xi_\star)}h + \mathrm{o}(h) \quad \text{and} \quad \chi_\star^h - \chi_\star = -h\int_{\xi_\star}^\infty \frac{v(\xi)}{1-\alpha}\mathrm{d}\xi + \mathrm{o}(h) \quad \text{as} \quad \mathcal{H} \ni h \downarrow 0.$$

Theorem 1 (Statistical Error). $\exists\lambda > 0$ s.t. **if** $\gamma_1 > \lambda$ **when** $\beta = 1$**,**

$$\mathbb{E}[(\xi_n^h - \xi_\star^h)^2] \leq C\gamma_n \quad \text{and} \quad \mathbb{E}[(\chi_n^h - \chi_\star^h)^2] \leq \frac{C}{n^{1\wedge 2\beta}}.$$

Eventually,

$$\mathbb{E}[(\xi_n^h - \xi_\star)^2] + \mathbb{E}[(\chi_n^h - \chi_\star)^2] \leq C(h^2 + \gamma_n).$$

For **a prescribed accuracy** $\varepsilon > 0$**,** we have to choose

$$h = \varepsilon, \quad n = \lceil \varepsilon^{-\frac{2}{\beta}} \rceil, \quad \text{i.e.} \quad \mathrm{Cost}_{\mathsf{NSA}}^\beta = C\frac{n}{h} = C\varepsilon^{-\frac{2}{\beta}-1} \underset{\beta\to 1}{\searrow} C\varepsilon^{-3}, \quad \text{if} \quad \gamma_1 > \lambda \quad \text{when} \quad \beta = 1,$$

for **some constant** $\lambda > 0$ **that is, although explicit, tedious to compute.**

Theorem 2 (Central Limit Theorem). Let $\beta \in (\frac{1}{2}, 1]$. $\exists \lambda > 0$ s.t. **if** $\gamma_1 > \lambda$ **when** $\beta = 1$,

$$\begin{pmatrix} h^{-\beta}(\xi^h_{\lceil h^{-2}\rceil} - \xi^h_\star) \\ h^{-1}(\chi^h_{\lceil h^{-2}\rceil} - \chi^h_\star) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma_\beta) \quad \text{as} \quad \mathcal{H} \ni h \downarrow 0,$$

where

$$\Sigma_\beta = \begin{pmatrix} \frac{\alpha\gamma_1}{2f_X(\xi_\star) - \gamma_1^{-1}(1-\alpha)\mathbf{1}_{\beta=1}} & \alpha\frac{\chi_\star - \xi_\star}{f_X(\xi_\star)}\mathbf{1}_{\beta=1} \\ \alpha\frac{\chi_\star - \xi_\star}{f_X(\xi_\star)}\mathbf{1}_{\beta=1} & \frac{\mathbb{V}\mathrm{ar}((X-\xi_\star)^+)}{(1-\alpha)^2} \end{pmatrix}.$$

Corollary 1. Under the context of Theorem 2,

$$\begin{pmatrix} h^{-\beta}(\xi^h_{\lceil h^{-2}\rceil} - \xi_\star) \\ h^{-1}(\chi^h_{\lceil h^{-2}\rceil} - \chi_\star) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix} -\frac{v(\xi_\star)}{f_X(\xi_\star)}\mathbf{1}_{\beta=1} \\ -\int_{\xi_\star}^\infty \frac{v(\xi)}{1-\alpha}\mathrm{d}\xi \end{pmatrix}, \Sigma_\beta\right) \quad \text{as} \quad \mathcal{H} \ni h \downarrow 0.$$

- **Averaged Nested Stochastic Approximation (ANSA) Algorithm.**

To circumvent the constraint on γ_1, we average out the VaR estimators (Crépey, Frikha, Louzi, and Pagès, 2023)

$$\boxed{\overline{\xi}^h_n = \frac{1}{n}\sum_{k=1}^n \xi^h_k = \overline{\xi}^h_{n-1} - \frac{1}{n}\big(\overline{\xi}^h_{n-1} - \xi^h_n\big).}$$

The global error can be decomposed into **a statistical error** and **a bias error**

$$\boxed{\overline{\xi}^h_n - \xi_\star = (\overline{\xi}^h_n - \xi^h_\star) + (\xi^h_\star - \xi_\star).}$$

Theorem 3 (Statistical Error). For $\beta \in \big(\frac{1}{2}, 1\big)$,

$$\boxed{\mathbb{E}[(\overline{\xi}^h_n - \xi^h_\star)^2] \le \frac{C}{n}.}$$

Finally, for $\beta \in \big(\frac{1}{2}, 1\big)$,

$$\mathbb{E}[(\overline{\xi}^h_n - \xi_\star)^2] + \mathbb{E}[(\chi^h_n - \chi_\star)^2] \le C\Big(h^2 + \frac{1}{n}\Big).$$

For **a prescribed accuracy $\varepsilon > 0$,** we must choose

$$h = \varepsilon, \quad n = \lceil \varepsilon^{-2} \rceil, \quad \text{i.e.} \quad \boxed{\mathrm{Cost}_{\mathsf{ANSA}} = C\frac{n}{h} = C\varepsilon^{-3}, \quad \beta \in \Big(\frac{1}{2}, 1\Big).}$$

Theorem 4 (Central Limit Theorem). For $\beta \in \left(\frac{1}{2}, 1\right)$,

$$h^{-1}\begin{pmatrix} \overline{\xi}^h_{\lceil h^{-2}\rceil} - \xi^h_\star \\ \chi^h_{\lceil h^{-2}\rceil} - \chi^h_\star \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \overline{\Sigma}) \quad \text{as} \quad \mathcal{H} \ni h \downarrow 0,$$

where

$$\overline{\Sigma} = \begin{pmatrix} \frac{\alpha(1-\alpha)}{f_X(\xi_\star)^2} & \frac{\alpha}{1-\alpha}\frac{\mathbb{E}[(X-\xi_\star)^+]}{f_X(\xi_\star)} \\ \frac{\alpha}{1-\alpha}\frac{\mathbb{E}[(X-\xi_\star)^+]}{f_X(\xi_\star)} & \frac{\mathbb{V}\mathrm{ar}((X-\xi_\star)^+)}{(1-\alpha)^2}. \end{pmatrix}$$

Corollary 2. Within the framework of Theorem 4,

$$h^{-1}\begin{pmatrix} \overline{\xi}^h_{\lceil h^{-2}\rceil} - \xi_\star \\ \chi^h_{\lceil h^{-2}\rceil} - \chi_\star \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix} -\frac{v(\xi_\star)}{f_X(\xi_\star)} \\ -\int_{\xi_\star}^\infty \frac{v(\xi)}{1-\alpha}\mathrm{d}\xi \end{pmatrix}, \overline{\Sigma}\right) \quad \text{as} \quad \mathcal{H} \ni h \downarrow 0.$$

The implied nested Monte Carlo computational cost should be countered.

Let $h_0 = \frac{1}{K} \in \mathcal{H}$, an integer $M > 1$ and **a number of levels L.** Using telescopic summation,

$$\xi_\star^{h_L} = \xi_\star^{h_0} + \sum_{\ell=1}^{L} \xi_\star^{h_\ell} - \xi_\star^{h_{\ell-1}} \quad \text{and} \quad \chi_\star^{h_L} = \chi_\star^{h_0} + \sum_{\ell=1}^{L} \chi_\star^{h_\ell} - \chi_\star^{h_{\ell-1}}, \quad h_\ell = \frac{h_0}{M^\ell} = \frac{1}{KM^\ell} \in \mathcal{H}.$$

- **Multilevel SA (MLSA) Algorithm.** According to Heinrich, 2001, Giles, 2008, Dereich, 2011 and Frikha, 2016, we can take **iteration amounts** $N_0 \geq \cdots \geq N_L$ and define the multilevel estimators

$$\xi_{\mathsf{ML}}^{L} = \xi_{N_0}^{h_0} + \sum_{\ell=1}^{L} \xi_{N_\ell}^{h_\ell} - \xi_{N_\ell}^{h_{\ell-1}} \quad \text{and} \quad \chi_{\mathsf{ML}}^{L} = \chi_{N_0}^{h_0} + \sum_{\ell=1}^{L} \chi_{N_\ell}^{h_\ell} - \chi_{N_\ell}^{h_{\ell-1}}.$$

Each level $0 \leq \ell \leq L$ is simulated independently using the nested SA algorithm.

Within each level $1 \leq \ell \leq L$, X_{h_ℓ} and $X_{h_{\ell-1}}$ are perfectly correlated:

$$X_{h_{\ell-1}} = \frac{1}{KM^{\ell-1}} \sum_{k=1}^{KM^{\ell-1}} \psi(Y, Z^{(k)}) \quad \text{and} \quad X_{h_\ell} = \frac{1}{M} X_{h_{\ell-1}} + \frac{1}{KM^\ell} \sum_{k=KM^{\ell-1}+1}^{KM^\ell} \psi(Y, Z^{(k)}).$$

The global error is the sum of **the statistical error** and **the bias error**

$$\xi^L_{\mathsf{ML}} - \xi_\star = (\xi^L_{\mathsf{ML}} - \xi^{h_L}_\star) + (\xi^{h_L}_\star - \xi_\star) \quad \text{and} \quad \chi^L_{\mathsf{ML}} - \chi_\star = (\chi^L_{\mathsf{ML}} - \chi^{h_L}_\star) + (\chi^{h_L}_\star - \chi_\star).$$

Assumption 2. We consider 4 different frameworks:

1. $\exists p_* > 1$, $\mathbb{E}\big[\big|\psi(Y,Z) - \mathbb{E}[\psi(Y,Z)|Y]\big|^{p_*}\big] < \infty$.
2. $\exists C > 0$, $\mathbb{E}\big[\exp\big(\lambda\big(\psi(Y,Z) - \mathbb{E}[\psi(Y,Z)|Y]\big)\big)\big] \leq \exp(C\lambda^2)$, $\lambda \in \mathbb{R}$.
3. $G_\ell = h_\ell^{-\frac{1}{2}}(X_{h_\ell} - X_{h_{\ell-1}})$ satisfies $\sup_{\ell\geq 1} \mathbb{E}\big[[F_{X_{h_{\ell-1}}|G_\ell}]_{\mathrm{Lip}}\,|G_\ell|\big] < \infty$.
4. $\big(f_{X_{h_{\ell-1}}|G_\ell}\big)_{\ell\geq 1}$ is uniformly bounded.

Theorem 5 (Statistical Error). $\exists\lambda > 0$ s.t. **if** $\gamma_1 > \lambda$ **when** $\beta = 1$,

$$\mathbb{E}[(\xi^L_{\mathsf{ML}} - \xi^{h_L}_\star)^2] \leq C \sum_{\ell=0}^{L} \gamma_{N_\ell}\varphi(h_\ell) \quad \text{and} \quad \mathbb{E}[(\chi^L_{\mathsf{ML}} - \chi^{h_L}_\star)^2] \leq C \sum_{\ell=0}^{L} \frac{h_\ell}{N_\ell},$$

where

$$\varphi(h_\ell) = \sqrt{h_\ell}^{\frac{p_*}{1+p_*}} \text{ (Asp 2.1)}, \quad \sqrt{h_\ell\,|\ln h_\ell|} \text{ (Asp 2.2)}, \quad \sqrt{h_\ell} \text{ (Asp 2.3, 2.4)}.$$

Altogether,

$$\mathbb{E}[(\xi^L_{\mathsf{ML}} - \xi_\star)^2] \leq C\left(h_L^2 + \sum_{\ell=0}^{L} \gamma_{N_\ell}\varphi(h_\ell)\right) \quad \text{and} \quad \mathbb{E}[(\chi^L_{\mathsf{ML}} - \chi_\star)^2] \leq C\left(h_L^2 + \sum_{\ell=0}^{L} \frac{h_\ell}{N_\ell}\right).$$

To achieve **a prescribed accuracy $\varepsilon > 0$,** one must choose

$$h_L = \frac{h_0}{M^L} = \varepsilon \quad \Leftrightarrow \quad \boxed{L = \left\lceil \frac{\ln \frac{h_0}{\varepsilon}}{\ln M} \right\rceil.}$$

The complexity of the MLSA algorithm is

$$\boxed{\mathrm{Cost}^{\beta}_{\mathsf{MLSA}} = C \sum_{\ell=0}^{L} \frac{N_\ell}{h_\ell}.}$$

- **VaR Focused Parametrization.** We optimize

$$\begin{array}{ll} \min_{N_0,\dots,N_L>0} & \sum_{\ell=0}^{L} \frac{N_\ell}{h_\ell}, \\ \text{subject to} & \sum_{\ell=0}^{L} \gamma_{N_\ell}\varphi(h_\ell) = \varepsilon^2. \end{array}$$

We infer the optimal iterations

$$N_\ell = \left\lceil \varepsilon^{-\frac{2}{\beta}} \left(\sum_{\ell'=0}^{L} h_{\ell'}^{-\frac{\beta}{1+\beta}} \varphi(h_{\ell'})^{\frac{1}{1+\beta}} \right)^{\frac{1}{\beta}} h_\ell^{\frac{1}{1+\beta}} \varphi(h_\ell)^{\frac{1}{1+\beta}} \right\rceil$$

i.e. $\boxed{\mathrm{Cost}^{\beta}_{\textsf{MLSA}} = C\sum_{\ell=0}^{L}\frac{N_\ell}{h_\ell} = C\varepsilon^{-\frac{2}{\beta}-1}\varphi(\varepsilon) \underset{\beta\to 1}{\searrow} C\varepsilon^{-3}\varphi(\varepsilon), \quad \text{if} \quad \gamma_1 > \lambda \quad \text{when} \quad \beta = 1,}$

or

$\boxed{\mathrm{Cost}^{\beta}_{\textsf{MLSA}} = C\varepsilon^{-\frac{5p_*+6}{2(1+p_*)}} \text{ (Asp 2.1)}, \quad C\varepsilon^{-\frac{5}{2}}|\ln\varepsilon|^{\frac{1}{2}} \text{ (Asp 2.2)}, \quad C\varepsilon^{-\frac{5}{2}} \text{ (Asp 2.3, 2.4)},}$

where $\lambda > 0$ **is an inaccessible albeit explicit constant.**

- **ES Focused Parametrization.** We solve

$$\begin{array}{ll} \min_{N_0,\dots,N_L>0} & \sum_{\ell=0}^{L} \frac{N_\ell}{h_\ell}, \\ \text{subject to} & \sum_{\ell=0}^{L} \frac{h_\ell}{N_\ell} = \varepsilon^2. \end{array}$$

On the condition $\beta = 1$, we obtain the amounts of iterations

$$N_\ell = \lceil \varepsilon^{-2} L h_\ell \rceil \quad \text{i.e.} \quad \boxed{\mathrm{Cost}^{\mathsf{ES}}_{\mathsf{MLSA}} = C \sum_{\ell=0}^{L} \frac{N_\ell}{h_\ell} = C\varepsilon^{-2} \left|\ln \varepsilon\right|^2 \quad \text{if} \quad \gamma_1 > \lambda,}$$

whereby λ **is an explicit yet inaccessible constant.**

This coincides with the optimal complexity of the multilevel MC algorithm derived by Giles, 2008.

Theorem 6 (Central Limit Theorem). Under **Assumption 2.4**, $G_\ell \xrightarrow{\mathcal{L}} G$ as $\ell \uparrow \infty$, and for $\beta \in \left(\frac{1}{2}, 1\right]$ and

$$N_\ell = \left\lceil h_L^{-\frac{2}{\beta}} \left(\sum_{\ell'=0}^{L} h_{\ell'}^{-\frac{2\beta-1}{2(1+\beta)}} \right)^{\frac{1}{\beta}} h_\ell^{\frac{3}{2(1+\beta)}} \right\rceil,$$

$\exists \lambda > 0$ s.t. **if** $\gamma_1 > \lambda$ **when** $\beta = 1$,

$$b \begin{pmatrix} h_L^{-1}(\xi_{\mathbf{N}}^{\mathrm{ML}} - \xi_\star^{h_L}) \\ h_L^{-\frac{1}{\beta}-\frac{2\beta-1}{4\beta(1+\beta)}}(\chi_{\mathbf{N}}^{\mathrm{ML}} - \chi_\star^{h_L}) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma_\beta^{\mathrm{ML}}) \quad \text{as} \quad L \uparrow \infty,$$

where $\Sigma_\beta^{\mathrm{ML}} =$

$$\begin{pmatrix} \frac{\gamma_1 \mathbb{E}[|G| f_G(\xi_\star)]}{(1-\alpha)(2 f_X(\xi_\star) - (1-\alpha)\gamma_1^{-1} \mathbf{1}_{\beta=1})} & 0 \\ 0 & \frac{h_0^{\frac{2\beta-1}{2(1+\beta)}} \left(M^{\frac{2\beta-1}{2(1+\beta)}} - 1\right)^{\frac{1}{\beta}}}{(1-\alpha)^2} \left(\frac{h_0^{-1} \mathbb{V}\mathrm{ar}((X_{h_0} - \xi_\star^{h_0})^+)}{M^{\frac{2\beta-1}{2\beta(1+\beta)}}} + \frac{\mathbb{V}\mathrm{ar}(\mathbf{1}_{X > \xi_\star} G)}{M^{\frac{2\beta-1}{2(1+\beta)}} - 1} \right) \end{pmatrix}.$$

- **Averaged Multilevel Stochastic Approximation (AMLSA) Algorithm.**

To avoid the limitation on γ_1, we average out the VaR estimators (Crépey, Frikha, Louzi, and Pagès, 2023)

$$\overline{\xi}^{L}_{\mathsf{ML}} = \overline{\xi}^{h_0}_{N_0} + \sum_{\ell=1}^{L} \overline{\xi}^{h_\ell}_{N_\ell} - \overline{\xi}^{h_{\ell-1}}_{N_\ell}.$$

The global error satisfies

$$\overline{\xi}^{L}_{\mathsf{ML}} - \xi_\star = (\overline{\xi}^{L}_{\mathsf{ML}} - \xi^{h_L}_\star) + (\xi^{h_L}_\star - \xi_\star).$$

Theorem 7 (Statistical Error). Under **Assumption 2.4,**

$$\mathbb{E}[(\overline{\xi}^{L}_{\mathsf{ML}} - \xi^{h_L}_\star)^2] \leq C \sum_{\ell=0}^{L} \frac{\varphi(h_\ell)}{N_\ell},$$

where

$$\varphi(h_\ell) = \sqrt{h_\ell}^{\,\frac{p_*}{1+p_*}} \text{ (Asp 2.1)}, \quad \sqrt{h_\ell\,|\ln h_\ell|} \text{ (Asp 2.2)}, \quad \sqrt{h_\ell} \text{ (Asp 2.3, 2.4)}.$$

In a nutshell,

$$\mathbb{E}[(\overline{\xi}^{L}_{\mathsf{ML}} - \xi_\star)^2] \leq C\left(h_L^2 + \sum_{\ell=0}^{L} \frac{\varphi(h_\ell)}{N_\ell}\right).$$

For **a prescribed accuracy $\varepsilon > 0$**, we parametrize $$L = \left\lceil \frac{\ln \frac{h_0}{\varepsilon}}{\ln M} \right\rceil.$$

The complexity of the AMLSA algorithm is

$$\mathrm{Cost}_{\mathsf{AMLSA}} = C \sum_{\ell=0}^{L} \frac{N_\ell}{h_\ell}.$$

We solve

$$\begin{array}{ll} \min_{N_0,\dots,N_L>0} & \sum_{\ell=0}^{L} \frac{N_\ell}{h_\ell}, \\ \text{subject to} & \sum_{\ell=0}^{L} \frac{\varphi(h_\ell)}{N_\ell} = \varepsilon^2. \end{array}$$

We find the iterations amounts

$$N_\ell = \left\lceil \varepsilon^{-2} \left(\sum_{\ell'=0}^{L} h_{\ell'}^{-\frac{1}{2}} \varphi(h_{\ell'})^{\frac{1}{2}} \right) h_\ell^{\frac{1}{2}} \varphi(h_\ell)^{\frac{1}{2}} \right\rceil$$ i.e. $$\mathrm{Cost}_{\mathsf{AMLSA}} = C \sum_{\ell=0}^{L} \frac{N_\ell}{h_\ell} = C\varepsilon^{-3}\varphi(\varepsilon), \quad \beta \in \left(\frac{8}{9}, 1\right),$$

which means

$$\mathrm{Cost}_{\mathsf{AMLSA}} = C\varepsilon^{-\frac{5p_*+6}{2(1+p_*)}} \text{ (Asp 2.1)}, \quad C\varepsilon^{-\frac{5}{2}} |\ln \varepsilon|^{\frac{1}{2}} \text{ (Asp 2.2)}, \quad C\varepsilon^{-\frac{5}{2}} \text{ (Asp 2.3, 2.4)}.$$

Theorem 8 (Central Limit Theorem). Under **Assumption 2.4**, $G_\ell \xrightarrow{\mathcal{L}} G$ as $\ell \uparrow \infty$, and for $\beta \in \left(\frac{8}{9}, 1\right)$ and

$$N_\ell = \left\lceil h_L^{-2} \left(\sum_{\ell'=0}^{L} h_{\ell'}^{-\frac{1}{4}} \right) h_\ell^{\frac{3}{4}} \right\rceil,$$

$$\begin{pmatrix} h_L^{-1} (\overline{\xi}_{\mathbf{N}}^{\mathrm{ML}} - \xi_\star^{h_L}) \\ h_L^{-\frac{9}{8}} (\chi_{\mathbf{N}}^{\mathrm{ML}} - \chi_\star^{h_L}) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \overline{\Sigma}^{\mathrm{ML}}) \quad \text{as} \quad L \uparrow \infty,$$

where

$$\overline{\Sigma}^{\mathrm{ML}} = \begin{pmatrix} \frac{\mathbb{E}[|G| f_G(\xi_\star)]}{(1-\alpha)^2 (1-M^{-\frac{1}{4}})} & 0 \\ 0 & \frac{h_0^{-\frac{3}{8}} (1-M^{-\frac{1}{4}})^{\frac{1}{2}} \mathbb{V}\mathrm{ar}((X_{h_0} - \xi_\star^{h_0})^+)}{(1-\alpha)^2} + \frac{h_0^{\frac{1}{4}} \mathbb{V}\mathrm{ar}(\mathbf{1}_{X > \xi_\star} G)}{(1-\alpha)^2 M_0^{\frac{1}{4}}} \end{pmatrix}.$$

Next step is to explore **adaptive refinement** to reduce the cost of MLSA/AMLSA to ε^{-2}.

The idea of adaptive multilevel SA is to **refine the simulation X_{h_ℓ} to $X_{h_{\ell+\eta^n_\ell(\xi)}}$** to escape the discontinuity of the gradient $H_1(\xi,\cdot)$ at ξ **at level ℓ and iteration n.**

- **Sampling Refinement Algorithm.** Following Haji-Ali, Spence, and Teckentrup, 2022, we define

$$\boxed{\eta^n_\ell(\xi) = \lceil\theta\ell\rceil \wedge \min\left\{k \in [\![0, \lceil\theta\ell\rceil]\!] : |X_{h_{\ell+k}} - \xi| \geq C_0 \psi^n_{k,\ell}\right\}, \quad \ell \geq 0, \quad n \in \mathbb{N},}$$

with the convention $\min\varnothing = \infty$,

$$\psi^n_{k,\ell} = \begin{cases} u_n^{-\frac{1}{p_*}} h_{\theta\ell(r-1)+k}^{\frac{1}{r}} & \text{(Asp 2.1)}, \\ h_{\theta\ell(r-1)+k}^{\frac{1}{r}} \left(\ln \gamma_n^{-\frac{1}{2}} h_{\ell+k}^{-\frac{1}{2}(1+\theta)}\right)^{\frac{1}{2}} & \text{(Asp 2.2, 2.3)}, \end{cases}$$

$C_0, \theta, r > 0$,

$$h_s = \frac{h_0}{M^s}, \quad s \in \mathbb{R}.$$

and

$$u_n = \gamma_1 n^{-\delta}, \quad \delta \in (0,1].$$

- **Adaptive Multilevel Stochastic Approximation Algorithm.** Define

$$\widetilde{X}_{h_\ell}^{(n+1)} = X^{(n+1)}_{h_{\ell+\eta_\ell^{(n+1)}}}, \quad \eta_\ell^{(n+1)} = \eta_\ell^{n+1}(\widetilde{\xi}_n^{h_\ell}),$$

with

$$\widetilde{\xi}_{n+1}^{h_\ell} = \widetilde{\xi}_n^{h_\ell} - \gamma_{n+1} H_1(\widetilde{\xi}_n^{h_\ell}, \widetilde{X}_{h_\ell}^{(n+1)}).$$

For $N_0 \geq \cdots \geq N_L$, the adaptive multilevel SA estimator for the VaR is given by

$$\boxed{\widetilde{\xi}_{\mathsf{ML}}^L = \widetilde{\xi}_{N_0}^{h_0} + \sum_{\ell=1}^{L} \widetilde{\xi}_{N_\ell}^{h_\ell} - \widetilde{\xi}_{N_\ell}^{h_{\ell-1}}.}$$

The global error is

$$\boxed{\widetilde{\xi}_{\mathsf{ML}}^L - \xi_\star = (\widetilde{\xi}_{\mathsf{ML}}^L - \xi_\star^{h_{L+\lceil \theta L \rceil}}) + (\xi_\star^{h_{L+\lceil \theta L \rceil}} - \xi_\star),}$$

Assumption 3. We consider 3 different frameworks:

1. (Asp 2.1) + $p_* > 2$, $r < 2$, $\theta \leq \frac{p_*/2-1}{p_*/2+1}$.
2. (Asp 2.2) + $h_0 \geq (8C/C_0^2)^{\frac{1}{2/r-1}}$, $r \leq 2$, $\theta \leq 1$.
3. (Asp 2.3) + $\sup_{\ell \geq 0} \mathbb{E}[\exp(v_0 G_\ell^2)] < \infty$, $h_0 \geq (2/v_0 C_0^2)^{\frac{1}{2/r-1}}$, $r \leq 2$, $\theta \leq 1$.

Theorem 9 (Statistical Error). $\exists \lambda > 0$ s.t. **if** $\gamma_1 > \lambda$ **when** $\beta = 1$,

$$\mathbb{E}[(\widetilde{\xi}_{\mathsf{ML}}^{L} - \xi_\star^{h_{L+\lceil \theta L \rceil}})^2] \leq C \sum_{\ell=0}^{L} \Gamma_{N_\ell}^{\ell} \varphi(h_\ell)^{1+\theta},$$

where

$$\Gamma_{N_\ell}^{\ell} = \gamma_n \vee u_n h_\ell^{1+\theta} \text{ (Asp 3.1)}, \quad \gamma_n \text{ (Asp 3.2, 3.3)},$$

and

$$\varphi(h_\ell) = \sqrt{h_\ell}^{\frac{p_*}{1+p_*}} \text{ (Asp 3.1)}, \quad \sqrt{h_\ell \, |\ln h_\ell|} \text{ (Asp 3.2)}, \quad \sqrt{h_\ell} \text{ (Asp 3.3)}.$$

Overall,

$$\mathbb{E}[(\widetilde{\xi}^{L}_{\mathsf{ML}} - \xi_\star)^2] \leq C\left(h_L^{2(1+\theta)} + \sum_{\ell=0}^{L} \Gamma^{\ell}_{N_\ell}\varphi(h_\ell)^{1+\theta}\right).$$

To realize **a prescribed accuracy $\varepsilon > 0$**, we have to choose

$$h_{L+\lceil\theta L\rceil} = \frac{h_0}{M^{L+\lceil\theta L\rceil}} = \varepsilon \quad \Leftrightarrow \quad \boxed{L = \left\lceil \frac{\ln\frac{h_0}{\varepsilon}}{(1+\theta)\ln M}\right\rceil}.$$

The complexity of the adaptive multilevel SA algorithm is

$$\boxed{\mathrm{Cost}_{\mathsf{AdaMLSA}} \leq C \begin{cases} \sum_{\ell=0}^{L} \frac{N_\ell^{1+\frac{\delta}{p_*}}}{h_\ell} & \text{(Asp 3.1)}, \\ \sqrt{L}\sum_{\ell=0}^{L} \frac{N_\ell}{h_\ell} + \sum_{\ell=0}^{L} \frac{N_\ell\sqrt{\ln N_\ell}}{h_\ell} & \text{(Asp 3.2, 3.3)}. \end{cases}}$$

Under (Asp 3.1), the optimal complexity is achieved when $\delta < \beta$ for any $\theta \in \left(0, \frac{p_*/2-1}{p_*/2+1}\right]$ as long as $\delta \to \beta = 1$:

$$\boxed{\mathrm{Cost}_{\mathsf{AdaMLSA}} \leq C\varepsilon^{-2-\frac{2}{p_*}}.}$$

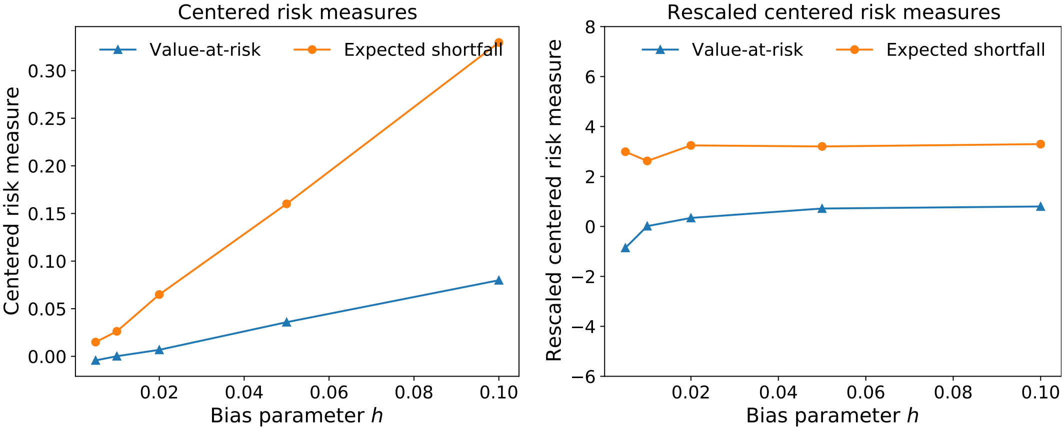
Under (Asp 3.2), the complexity is minimal when $\beta = \theta = 1$, in which case

$$\boxed{\mathrm{Cost}_{\mathsf{AdaMLSA}} \leq C\varepsilon^{-2} \left|\ln \varepsilon\right|^{\frac{7}{2}}.}$$

Under (Asp 3.3), the complexity is optimal when $\beta = \theta = 1$, where

$$\boxed{\mathrm{Cost}_{\mathsf{AdaMLSA}} \leq C\varepsilon^{-2} \left|\ln \varepsilon\right|^{\frac{5}{2}}.}$$

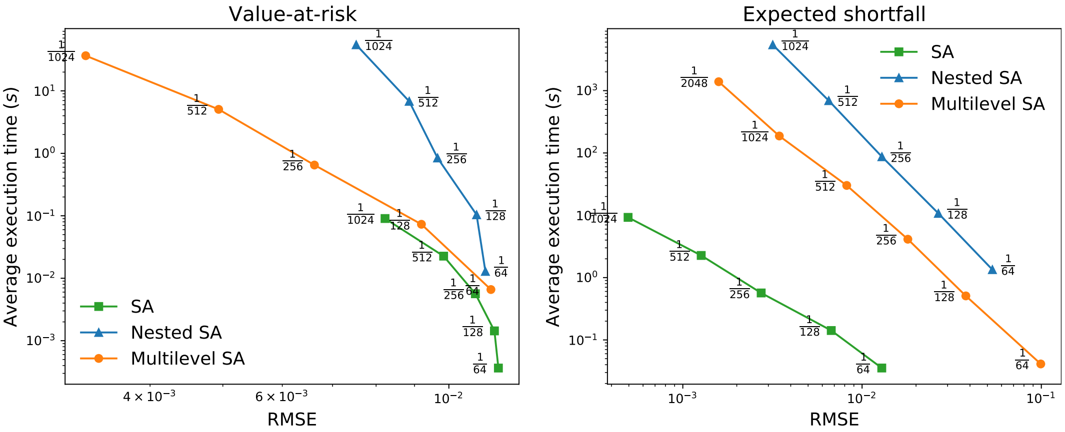
Consider **a European option with payoff $-W_1^2$ at $T = 1$.** Set $\tau = 0.5$ and $\alpha = 97.5\%$.



Bias errors $\xi_\star^h - \xi_\star$ and $\chi_\star^h - \chi_\star$ as $\mathcal{H} \ni h \downarrow 0$.

$\xi_\star^h - \xi_\star$ and $\chi_\star^h - \chi_\star$ **decrease linearly** as $\mathcal{H} \ni h \downarrow 0$.

Consider **a European option with payoff $-W_1^2$ at $T = 1$.** Set $\tau = 0.5$ and $\alpha = 97.5\%$.

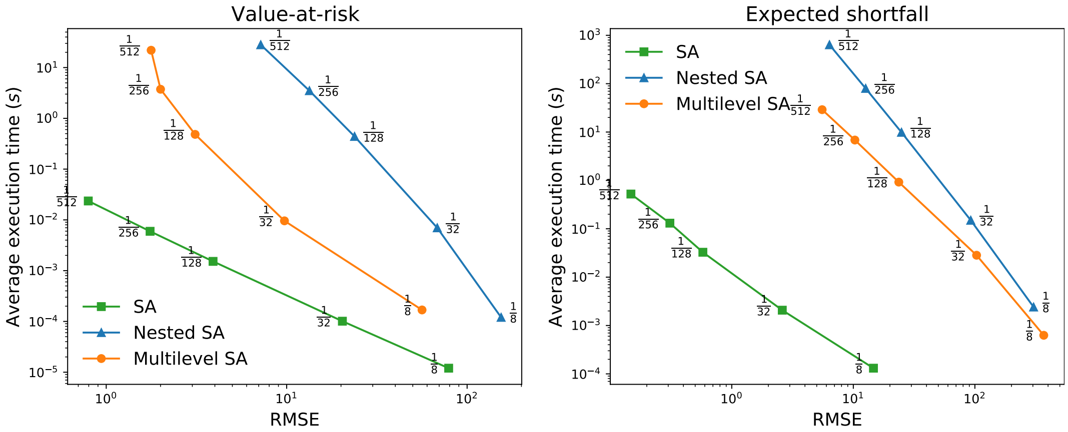


RMSE vs execution time as $\varepsilon \downarrow 0$. $\beta = 1$, $M = 2$, $p_* = 11$, 200 runs.

For a target RMSE $= 10^{-2}$,

- for the VaR, SA: 10^{-2}s, MLSA: 3.10^{-2}s and NSA: 5.10^{-1}s;
- for the ES, SA: 5.10^{-2}s, MLSA: 5.10s and NSA: 3.10^{2}s.

Consider **a short position on a swap issued at par on some Black-Scholes rate.** Set $\alpha = 85\%$.

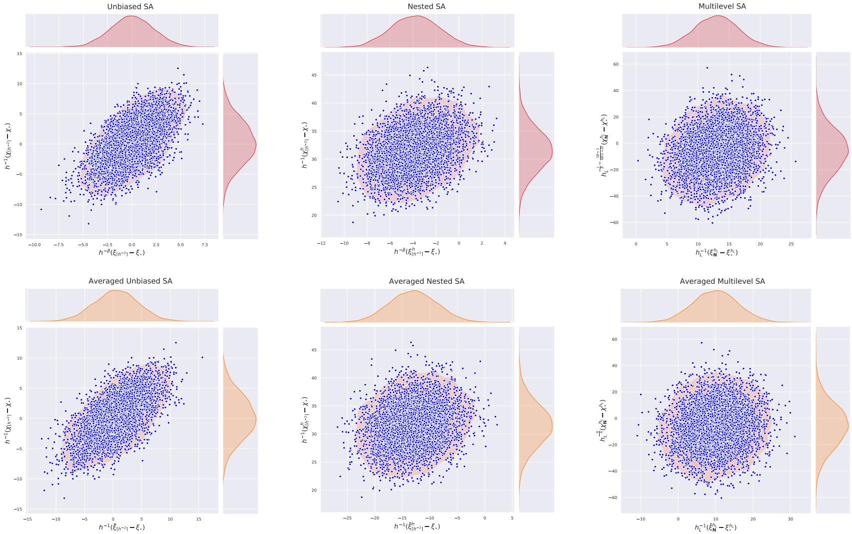


RMSE vs execution time as $\varepsilon \downarrow 0$. $\beta = 1$, $M = 2$, $p_* = 8$, 200 runs.

For a target RMSE $= 10$,

- for the VaR, SA: 3.10^{-4}s, MLSA: 10^{-2}s and NSA: 10s;
- for the ES, SA: 3.10^{-4}s, MLSA: 10s and NSA: 10^2s.

Consider **a short position on a swap issued at par on some Bachelier rate.** Set $\alpha = 85\%$.



CLTs. $\beta = 0.9$, $M = 2$, 5000 runs.

- **Key Takeaways.** Let **some prescribed accuracy $\varepsilon > 0$.** Taking $\beta = 1$,

$$\boxed{\mathrm{Cost}^1_{\mathsf{SA}} = C\varepsilon^{-2} \ll \begin{cases} \mathrm{Cost}_{\mathsf{AdaMLSA}} = C\varepsilon^{-2}\,|\ln\varepsilon|^{\frac{5}{2}} \\ \mathrm{Cost}^{\mathsf{ES}}_{\mathsf{MLSA}} = C\varepsilon^{-2}\,|\ln\varepsilon|^{2} \end{cases}}$$

$$\boxed{\ll \mathrm{Cost}^1_{\mathsf{MLSA}} = C\varepsilon^{-\frac{5}{2}} \ll \mathrm{Cost}^1_{\mathsf{NSA}} = C\varepsilon^{-3} \quad \text{if} \quad \gamma_1 > \lambda,}$$

whereby λ **is an explicit yet inaccessible constant,** and $\delta < 1$ depends on **the integrability degree** of the loss.

By averaging out the VaR estimators to overcome this limitation,

$$\boxed{\mathrm{Cost}_{\mathsf{ASA}} = C\varepsilon^{-2} \ll \mathrm{Cost}_{\mathsf{AMLSA}} = C\varepsilon^{-\frac{5}{2}} \ll \mathrm{Cost}_{\mathsf{ANSA}} = C\varepsilon^{-3}, \quad \beta \in \left(\frac{8}{9}, 1\right).}$$

- **Next steps.** Explore some numerical applications for **the adaptive refinement**.

Bardou, O., N. Frikha, and G. Pagès (2009). "Computing VaR and CVaR Using Stochastic Approximation and Adaptive Unconstrained Importance Sampling". In: *Monte Carlo Methods and Applications* 15.3, pp. 173–210. DOI: doi:10.1515/MCMA.2009.011.

Barrera, D., S. Crépey, B. Diallo, G. Fort, E. Gobet, and U. Stazhynski (2019). "Stochastic Approximation Schemes for Economic Capital and Risk Margin Computations". In: *ESAIM: Proceedings and Surveys* 65, pp. 182–218.

Crépey, S., N. Frikha, and A. Louzi (2023). *A Multilevel Stochastic Approximation Algorithm for Value-at-Risk and Expected Shortfall Estimation*. arXiv: 2304.01207 (q-fin.CP).

Crépey, S., N. Frikha, A. Louzi, and G. Pagès (2023). *Asymptotic Error Analysis of Multilevel Stochastic Approximations for the Value-at-Risk and Expected Shortfall*. arXiv: 2311.15333 (q-fin.RM).

Dereich, S. (July 15, 2011). "Multilevel Monte Carlo Algorithms For Lévy-Driven SDEs With Gaussian Correction". In: *The Annals of Applied Probability* 21.1, pp. 283–311. ISSN: 10505164.

Frikha, N. (2016). "Multi-level Stochastic Approximation Algorithms". In: *The Annals of Applied Probability* 26.2, pp. 933–985. DOI: 10.1214/15-AAP1109.

Giles, M. B. (2008). "Multilevel Monte Carlo Path Simulation". In: *Operations Research* 56.3, pp. 607–617. DOI: 10.1287/opre.1070.0496.

Haji-Ali, A.-L., J. Spence, and A. L. Teckentrup (2022). "Adaptive Multilevel Monte Carlo for Probabilities". In: *SIAM Journal on Numerical Analysis* 60.4, pp. 2125–2149. DOI: 10.1137/21M1447064.

Heinrich, S. (2001). "Multilevel Monte Carlo Methods". In: *Large-Scale Scientific Computing*. Springer, pp. 58–67. ISBN: 978-3-540-45346-8.

Rockafellar, R. T. and S. Uryasev (Apr. 1, 2000). "Optimization of Conditional Value-at-Risk". In: *Journal of Risk* 2.3, pp. 21–41. DOI: 10.21314/JOR.2000.038.