Multilevel methods for Nested Monte Carlo simulation

Longevity 11' Lyon

ISFA

V. Lemaire & G. Pagès

LPMA–UPMC & ANR LoLitA

7.09.2015

Motivation

- Illustrate the efficiency of recent advances in Monte Carlo simulation, especially
  - Multilevel methods initiated by M. Giles ($MLMC$, 2008)
  - weighted Multilevel methods developed in ($ML2R$, Lemaire-P., 2013).
- to speed up Nested Monte Carlo simulation...
- which is a key numerical tool in actuarial sciences, $e.g.$ to compute risk-based funding requirements, typically Solvency Capital Requirement (SCR), for Life insurance problems and contracts exposed to longevity risks.

A simulation benchmark devoted to multilevel methods is available at the url

https:simulations.lpma-paris.fr

A stylized toy model for SCR

(inspired by [Bauer et al., *Math. & Econ.* 2010])

- A risky asset process and a short rate Vasicek model

$$X_t = (A_t, r_t) \;\text{ with } \begin{cases} \mathrm{d}A_t = A_t\big(r_t dt + \sigma_A \mathrm{d}W_t\big), & A_0 > 0, \\ \mathrm{d}r_t = \kappa(\xi - \frac{\lambda\sigma_r}{\kappa} - r_t)\mathrm{d}t + \sigma_r \mathrm{d}B_t, & r_0 > 0, \end{cases}$$

(W, B) Brownian motion with $\langle W, B\rangle_t = \rho\, t$, $\rho \in (-1, 1)$ under risk neutral probability $\mathbf{Q}$.

A stylized toy model for SCR

(inspired by [Bauer et al., *Math. & Econ.* 2010])

- A risky asset process and a short rate Vasicek model

$$X_t = (A_t, r_t) \text{ with } \begin{cases} \mathrm{d}A_t = A_t\big(r_t dt + \sigma_A \mathrm{d}W_t\big), & A_0 > 0, \\ \mathrm{d}r_t = \kappa(\xi - \frac{\lambda\sigma_r}{\kappa} - r_t)\mathrm{d}t + \sigma_r \mathrm{d}B_t, & r_0 > 0, \end{cases}$$

  (W, B) Brownian motion with $\langle W, B\rangle_t = \rho\, t$, $\rho \in (-1, 1)$ under risk neutral probability $\mathbf{Q}$.

- History of the model $\equiv$ filtration

$$\mathcal{F}_t = \mathcal{F}_t^{(W,B)} = \mathcal{F}_t^{(A,r)},\ t \in [0, T].$$

A stylized toy model for SCR

(inspired by [Bauer et al., *Math. & Econ.* 2010])

- A risky asset process and a short rate Vasicek model

$$X_t = (A_t, r_t) \text{ with } \begin{cases} \mathrm{d}A_t = A_t\big(r_t dt + \sigma_A \mathrm{d}W_t\big), & A_0 > 0, \\ \mathrm{d}r_t = \kappa(\xi - \frac{\lambda\sigma_r}{\kappa} - r_t)\mathrm{d}t + \sigma_r \mathrm{d}B_t, & r_0 > 0, \end{cases}$$

  (W, B) Brownian motion with $\langle W, B\rangle_t = \rho\, t$, $\rho \in (-1, 1)$ under risk neutral probability $\mathbf{Q}$.

- History of the model $\equiv$ filtration

$$\mathcal{F}_t = \mathcal{F}_t^{(W,B)} = \mathcal{F}_t^{(A,r)},\ t \in [0, T].$$

- Yearly cash flows from the insurance company perspective over $[0, T]$

$$\pi_k(A_{1:k}),\ k = 1, \ldots, T, \quad \text{hence } \mathcal{F}_k\text{-adapted sequence.}$$

- Available capital at time $t = 0$

$$AC_0 = \mathsf{ANAV}_0 + \underbrace{\mathbf{E}_{\mathbf{Q}}\left[\sum_{k=1}^{T} e^{-\int_0^k r_s \mathsf{d}s} \pi_k(A_{1:k})\right]}_{=PVFP} \in \mathbf{R}.$$

- Available capital at time $t=0$

$$AC_0 = \mathsf{ANAV}_0 + \underbrace{\mathbf{E}_{\mathbf{Q}}\left[\sum_{k=1}^{T} e^{-\int_0^k r_s \mathsf{d}s}\pi_k(A_{1:k})\right]}_{=PVFP} \in \mathbf{R}.$$

- Available capital at time $t=1$

$$AC_1 = \underbrace{g(A_1)}_{=\mathsf{ANAV}_1} + \mathbf{E}_{\mathbf{Q}}\left[\sum_{k=1}^{T} e^{-\int_0^k r_u \mathsf{d}u}\pi_k(A_{1:k})\,|\,\mathcal{F}_1\right] = \mathbf{E}_{\mathbf{Q}}\left[\ldots\,|\,\mathcal{F}_1\right].$$

- Available capital at time $t = 0$

$$AC_0 = \text{ANAV}_0 + \underbrace{\mathbf{E}_{\mathbf{Q}}\left[\sum_{k=1}^{T} e^{-\int_0^k r_s \mathrm{d}s}\pi_k(A_{1:k})\right]}_{=PVFP} \in \mathbf{R}.$$

- Available capital at time $t = 1$

$$AC_1 = \underbrace{g(A_1)}_{=\text{ANAV}_1} + \mathbf{E}_{\mathbf{Q}}\left[\sum_{k=1}^{T} e^{-\int_0^k r_u \mathrm{d}u}\pi_k(A_{1:k}) \,|\, \mathcal{F}_1\right] = \mathbf{E}_{\mathbf{Q}}\left[\ldots \,|\, \mathcal{F}_1\right].$$

- One-year loss function L evaluated at $t = 0$

$$L = AC_0 - \frac{AC_1}{1 + s(0,1)}.$$

Solvency Capital Requirement as a quantile

- SCR at (confidence) level $\alpha \in [0,1]$ [Bâle II: $\alpha = 99.5\%$].

$$SCR = VaR_\alpha(L) = \sup_y \Big\{ \mathbf{Q}(L \leqslant y) \leqslant \alpha \Big\}.$$

- Computing $SCR \equiv$ compute α-quantile (α-Value-at-Risk).
  - Gaussian approximation based on (mean(L), variance(L)) estimation.

Solvency Capital Requirement as a quantile

- SCR at (confidence) level $\alpha \in [0,1]$ [Bâle II: $\alpha = 99.5\%$].

$$SCR = VaR_\alpha(L) = \sup_y \Big\{ \mathbf{Q}(L \leqslant y) \leqslant \alpha \Big\}.$$

- Computing $SCR \equiv$ compute α-quantile (α-Value-at-Risk).
  - Gaussian approximation based on (mean(L), variance(L)) estimation.
  - If L can be simulated at a reasonable computational cost, simulation based methods
    - ★ Extreme values (Embrecht et al., 1997)
    - ★ Invert the empirical cumulative distribution function ([Egloff-Leipold, 2007])
    - ★ Robbins-Monro algorithm [Bardou et al., *MCMA*, 2011]

$+$ (for the last two cases)

Control variate & Importance sampling

([Glasserman], [Arouna, 2009], [Lemaire-P., *AAP*, 2011] , etc.)

Solvency Capital Requirement as a quantile

- SCR at (confidence) level $\alpha \in [0,1]$ [Bâle II: $\alpha = 99.5\%$].

$$SCR = VaR_\alpha(L) = \sup_y \Big\{ \mathbf{Q}(L \leqslant y) \leqslant \alpha \Big\}.$$

- Computing $SCR \equiv$ compute α-quantile (α-Value-at-Risk).
  - Gaussian approximation based on (mean(L), variance(L)) estimation.
  - If L can be simulated at a reasonable computational cost, simulation based methods
    - Extreme values (Embrecht et al., 1997)
    - Invert the empirical cumulative distribution function ([Egloff-Leipold, 2007])
    - Robbins-Monro algorithm [Bardou et al., *MCMA*, 2011]

$+$ (for the last two cases)

Control variate & Importance sampling
([Glasserman], [Arouna, 2009], [Lemaire-P., *AAP*, 2011] , etc.)

- First step toward SCR = Probability level computation $i.e.$

$$\mathbf{Q}(L \leqslant y) = \mathbf{E}_{\mathbf{Q}}\left[\varphi_y(L)\right] \quad \text{with} \quad \varphi_y(u) = \mathbf{1}_{[y,+\infty)}(u).$$

- Second step : recursive level earch

$$\theta^{N+1} = \theta^N - \frac{1}{N+1}\big(\mathbf{Q}(L \leqslant \theta^N) - \alpha\big)$$

$$\theta^N \to \theta_\alpha \quad \text{as } N \to +\infty$$

such that

$$\mathbf{Q}(L \leqslant \theta_\alpha) = \alpha \quad i.e. \quad \mathsf{SCR} = \theta_\alpha$$

Stochastic counterpart?

Nested Monte Carlo

[Devineau-Loisel, 2009], [Broadie et al., 2010].

- Here L cannot be simulated exactly because $L = \mathbf{E}_{\mathbf{Q}}\left[\,.\,|\,\mathcal{F}_1\right]$. But
- AC_0 can be computed by direct Monte Carlo methods and ...

$$L \;=\; \mathbf{E}_{\mathbf{Q}}\left[\Lambda\Big(e^{-\int_1^k r_s ds}, A_k,\; k=1:T\Big)\,|\,\mathcal{F}_1\right]$$

Markov property implies

$$\begin{aligned} L &= \mathbf{E}_{\mathbf{Q}}\left[F\Big(\underbrace{(W_{1+s}-W_1, B_{1+s}-B_1)_{s\in[0,T-1]}}_{Z\perp\!\!\!\perp X}, \underbrace{A_1, r_1}_{=:X\in\mathcal{F}_1}\Big)\,\Big|\, X\right] \\ &= \mathbf{E}_{\mathbf{Q}}\left[F(Z,X)\,|\,X\right] = \Big[\mathbf{E}_{\mathbf{Q}}\left[F(Z,\xi)\right]\Big]_{|\xi=X}, \\ &= \lambda(X) \quad \text{with} \qquad F(Z,X)\in L^2(\Omega,\mathcal{A},\mathbf{P}). \end{aligned}$$

- **Strong Law of Large numbers I:** (X_ℓ) i.i.d. copies of $X = (A_1, r_1)$. Then

$$L^\ell = \lambda(X_\ell),\, \ell \geqslant 1, \quad \text{are i.i.d. copies of } L$$

so that

$$\frac{1}{N}\sum_{\ell=1}^{N} \underbrace{\mathbf{1}_{\{L^\ell \leqslant y\}}}_{\text{outer simulations}} \xrightarrow{L^2\text{-}a.s.} \mathbf{Q}(L \leqslant y).$$

- **Strong Law of Large numbers I:** (X_ℓ) i.i.d. copies of $X = (A_1, r_1)$. Then

$$L^\ell = \lambda(X_\ell),\ \ell \geqslant 1, \qquad \text{are i.i.d. copies of } L$$

so that

$$\frac{1}{N}\sum_{\ell=1}^{N} \underbrace{\mathbf{1}_{\{L^\ell \leqslant y\}}}_{\text{outer simulations}} \xrightarrow{L^2\text{-}a.s.} \mathbf{Q}(L \leqslant y).$$

- **Strong Law of Large numbers II:** Let $(Z_k^\ell)_{k,\ell \geqslant 1}$ i.i.d. $\perp\!\!\!\perp (X_\ell)_\ell$.

$$\bar{L}_h^\ell = \frac{1}{K}\sum_{k=1}^{K} \underbrace{F(Z_k^\ell, X_\ell)}_{\text{inner simulations}} \xrightarrow{L^2\text{-}a.s.} L^\ell \qquad \text{as } K \to +\infty.$$

where $h = \frac{1}{K}$. If $\mathbf{Q}(L = y) = 0$,

$$Y_h^\ell := \mathbf{1}_{\{\bar{L}_h^\ell \leqslant y\}} \xrightarrow{L^2} Y_0 := \mathbf{1}_{\{L \leqslant y\}} \qquad \text{as } K \to +\infty.$$

The Nested Monte Carlo estimator is defined by

$$I_h^N := \frac{1}{N}\sum_{\ell=1}^{N} Y_h^\ell$$

with

$$\boxed{Y_h^\ell = \mathbf{1}_{\{\bar{L}_h^\ell \leqslant y\}} = \mathbf{1}_{\{\frac{1}{K}\sum_{k=1}^{K} F(Z_k^\ell, X_\ell) \leqslant y\}}}$$

satisfies

$$\boxed{I_h^N \xrightarrow{L^2\text{-}a.s.} \mathbf{E_Q}[Y_0] = \mathbf{Q}(L \leqslant y) \text{ as } h = \frac{1}{K} \to 0,\ N \to +\infty}$$

Analysis of a crude Nested Monte Carlo I

In the spirit of [Broadie et al. 2010]

- The level estimator is biased since

$$\begin{aligned}\mathbf{E}_{\mathbf{Q}}[Y_h] &= \mathbf{Q}(\bar{L}_h \leqslant y)\\ &= \mathbf{Q}(L \leqslant y) + c_1 h + O(h^{\frac{3}{2}}) \qquad \text{[Gordy-Juneja, 2010].}\\ &= \mathbf{E}_{\mathbf{Q}}[Y_0] + c_1 h + O(h^{\frac{3}{2}})\end{aligned}$$

hence $h = \frac{1}{K}$ does appear as a bias parameter.

Analysis of a crude Nested Monte Carlo I

In the spirit of [Broadie et al. 2010]

- The level estimator is biased since

$$\begin{aligned} \mathbf{E}_{\mathbf{Q}}[Y_h] &= \mathbf{Q}(\bar{L}_h \leqslant y) \\ &= \mathbf{Q}(L \leqslant y) + c_1 h + O(h^{\frac{3}{2}}) \qquad \text{[Gordy-Juneja, 2010].} \\ &= \mathbf{E}_{\mathbf{Q}}[Y_0] + c_1 h + O(h^{\frac{3}{2}}) \end{aligned}$$

  hence $h = \frac{1}{K}$ does appear as a bias parameter.

- Compute the Mean Squared Error

$$\begin{aligned} MSE(I_h^N) = \|I_h^N - \mathbf{Q}(L \leqslant y)\|_2^2 &:= (\text{bias}(I_h^N))^2 + \text{var}(I_h^N) \\ &\approx c_1^2 h^2 + \frac{\text{var}(Y_h)}{N} \\ &\approx c_1^2 h^2 + \frac{\text{var}(Y_0)}{N} \end{aligned}$$

  (hence $\frac{1}{N}$ is a variance parameter).

Analysis of a crude Nested Monte Carlo II

- Complexity of the global simulation: $\mathrm{Cost}(h,N) = N \times K = \frac{N}{h}$.
- Optimization of the simulation: Let $\varepsilon > 0$ a target quadratic (L^2) accuracy

$$\min_{MSE(I_h^N) \leqslant \varepsilon^2} \mathrm{Cost}(h,N).$$

- Solution:

$$K^* = \frac{1}{h^*} = \underbrace{\frac{\sqrt{3}c_1}{\varepsilon}}_{\text{inner}} \quad \text{and} \quad N^* = \underbrace{\frac{3}{2}\frac{\mathrm{var}(Y_0)}{\varepsilon^2}}_{\text{outer}}$$

and a resulting complexity

$$\kappa^* = \kappa(h^*, N^*) = \frac{3^{\frac{3}{2}}}{2}\frac{c_1 \mathrm{var}(Y_0)}{\varepsilon^3} \dots$$

versus

$\frac{\mathrm{var}(Y_0)}{\varepsilon^2}$ for an *unbiased simulation* (if $Y_0 = \mathbf{1}_{\{L \leqslant y\}}$) were simulatable).

Can we improve that?

▷ Recent breakthroughs to speed up biased Monte Carlo simulations:

- Multistep Richardson-Romberg extrapolation (P. 2007, *MCMA*)
- Statistical Romberg (Kebaier, AAP, 2006, Ben Alaya-Kebaier, *Bernoulli*, 2014),
- Multilevel Monte Carlo (MLMC, Giles, 2008, *Oper. Research*) and many other co-authors or contributors (Belomestny, Dereich, Lemaire, Müller-Grunbach, P., Ritter, Sprutz, etc)
- Multilevel Richardson-Romberg (ML2R) simulation (Lemaire-P. 2013, to appear in *Bernoulli*, ArXiv 1401.1177v3).

▷ Other (recent) applications of similar ideas to Robbins-Monro stochastic zero search algorithms (Stochastic Approximation) by Frikha (2014), Lemaire-P. (2015, in progress), etc.

Regular Multilevel Monte Carlo framework(s)

- $Y_0 \in \mathbf{L}^2(\mathbf{P})$ not simulatable real-valued random variable.
- Aim: compute $\mathbf{E}\,[Y_0]$ with a given accuracy $\varepsilon > 0$.
- Let $Y_h \in \mathbf{L}^2(\mathbf{P})$, $h \in \mathcal{D} := \{\frac{\mathbf{h}}{n},\, n \geqslant 1\}$, be simulatable r.v. satisfying:
  - Higher order bias error expansion (weak error expansion):

$$(WE_{\alpha,R}) \equiv \mathbf{E}\,[Y_h] = \mathbf{E}\,[Y_0] + c_1 h^{\alpha} + c_2 h^{2\alpha} + \cdots + c_R h^{\alpha R} + o(h^{\alpha R}).$$

  - Strong approximation error assumption:

$$(SE_\beta) \equiv \forall h \in \mathcal{D}, \quad \left\|Y_h - Y_0\right\|_2^2 = \mathbf{E}\left[\left|Y_h - Y_0\right|^2\right] \leqslant V_1 h^{\beta}.$$

  - Unitary simulation complexity: $\kappa(h) = \dfrac{\kappa}{h}$.

$R = 1 \rightsquigarrow MLMC$ (Giles, 08) and $R \geqslant 2 \rightsquigarrow ML2R$ (Lemaire-P., 13)

Typical setting: diffusions

We want to compute to compute by a Monte Carlo simulation.

$$I_0 = \mathbf{E}\left[Y_0\right].$$

(to compute AC_0!).

▷ Let $(X_t)_{t\in[0,T}$ be a Brownian diffusion. We want to compute by a Monte Carlo simulation and let

$$I_0 = \mathbf{E}\left[Y_0\right] \quad \text{with} \quad Y_0 = \varphi(X_{T}).$$

▷ $Y_h = \varphi(\bar{X}^h_T)$, $(\bar{X}^h_t)_{t\in[0,T]}$ scheme with step $h = \frac{T}{n}$ (Euler, Misltein).

- φ **smooth:** $\beta = 1$ (Euler) or $\beta = 2$ (Milsein) and $(WE_{\alpha,R})$, $\alpha = 1$ (Talay-Tubaro), [depending on drift b and diffusion σ regularity].
- φ **indicator function:** $\beta \precsim 1/2$ and $(WE_{\alpha,R})$, $\alpha = 1$ (Bally-Talay)

  [depending on regularity of b and σ and strong Hörmander ellipticity assumption].

▷ Extensions to Lévy driven diffusions ($R = 1$), fractional, $PDMP$, etc.

Back to Nested MC I: weak approximation rate

Let $R \geqslant 2$ be an integer. Let $\varphi : \mathbf{R} \to \mathbf{R}$.

$$h = \frac{1}{K},\ Y_0 = \varphi(L) \qquad Y_h = \varphi(\bar{L}_h) = \varphi\left(\frac{1}{K}\sum_{k=1}^{K} F(Z_k, X)\right).$$

Theorem (Smooth functions (Lemaire-P. '13):)

$\varphi : \mathbf{R} \to \mathbf{R}$ be a $2R+1$ times differentiable function with bounded derivatives $\varphi^{(k)}$, $k = R+1 : 2R+1$. Assume $X \in L^{2R+1}$.

$$\forall h \in \mathcal{D}, \quad \mathbf{E}_{\mathbf{Q}}\left[\varphi(\bar{L}_h)\right] = \mathbf{E}_{\mathbf{Q}}\left[\varphi(L)\right] + \sum_{r=1}^{R} c_r h^r + \mathcal{O}(h^{R+1/2}).$$

▷ Conclusion for $\mathbf{E}_{\mathbf{Q}}\left[\varphi(L)\right]$ computation, φ smooth:

$$(WE_{\alpha,R}) \quad \text{with} \quad \alpha = 1.$$

Theorem (Indicator functions $\varphi = \varphi_y = \mathbf{1}_{(-\infty,y]}$ (Lemaire-P. '15+):)

Assume $(L, \bar{L}_h)$ *has a density. If* λ *is monotonic,* λ, f_L *and* $f_{L|\bar{L}_h - L = \xi}$ *are* $\mathcal{C}^{2R}$, *and*

$$\lim_{|y|\to+\infty} |y|^{2R-k} f^{(k)}_{L|\bar{L}_h - L=\xi}(y) = 0,\ \forall\, \xi \in \mathbf{R},\ \forall\, k \in \{0,\dots,2R\},$$

then, uniformly in $y \in \mathbf{R}$,

$$\mathbf{Q}\big(\bar{L}_h \leqslant y\big) = \mathbf{Q}\big(L \leqslant y\big) + \sum_{r=1}^{R} \frac{\mathbf{E}\left[P_r(L)\mathbf{1}_{\{L\leqslant y\}}\right]}{r!}\, h^r + \mathcal{O}(h^{R+\frac{1}{2}}).$$

- Extends former results (for $R = 1$) by [Gordy & Juneja '10],
- Conclusion for $\mathbf{Q}(L \leqslant y) = \mathbf{E}_{\mathbf{Q}}\left[\varphi_y(L)\right]$ computation, φ_y indicator function: again

$$(WE_{\alpha,R}) \quad \text{with} \quad \alpha = 1.$$

How to “kill" this bias under $(WE_{\alpha,R})$, $R \geqslant 2$?

i.e.

How to take advantage of our higher order expansion to kill the bias faster?

Multistep Richardson-Romberg extrapolation [P. 2007]

▷ We introduce *refiners* an increasing R-tuple

$$\underline{n} = (n_1, n_2, \ldots, n_R) \in \mathbb{N}^R \quad \text{s.t.} \quad n_1 = 1 < n_2 < \cdots < n_R$$

and

the resulting refined approximations $Y_{\frac{h}{n_i}}$ (simulated with complexity $\frac{\kappa n_i}{h}$).

▷ *Example of Nested MC*: As $K = \frac{1}{h}$, the refiners expand the length of the inner simulations, namely

$$K \rightsquigarrow n_i K \rightsquigarrow Y_{\frac{h}{n_i}} = \varphi_y\Big(\frac{1}{n_i K}\sum_{k=1}^{n_i K} F(Z_k, X)\Big).$$

▷ Let $\mathbf{w} = (\mathbf{w}_1, \dots \mathbf{w}_R) \in \mathbf{R}$ and the *weighted sum*

$$\sum_{i=1}^{R} \mathbf{w}_i \, \mathbf{E}\left[Y_{\frac{h}{n_i}}\right] \overset{(WE_{\alpha,R})}{=} \underbrace{\left[\sum_{i=1}^{R} \mathbf{w}_i\right]}_{(\star)\ if\ =1} \mathbf{E}\left[Y_0\right] + \sum_{i=1}^{R} \mathbf{w}_i \left[\sum_{r=1}^{R-1} c_r \left(\frac{h}{n_i}\right)^{\alpha r}\right]$$

$$+ \widetilde{\mathbf{w}}_{R+1} c_R h^{\alpha R} + o\left(h^{\alpha R}\right)$$

$$\text{interchange } \Sigma \Rightarrow = \quad \mathbf{E}\left[Y_0\right] + \sum_{r=1}^{R-1} c_r h^{\alpha r} \underbrace{\left[\sum_{i=1}^{R} \mathbf{w}_i \, n_i^{-\alpha r}\right]}_{(\star\star)\ if\ =0}$$

$$+ \widetilde{\mathbf{w}}_{R+1} c_R h^{\alpha R} + o\left(h^{\alpha R}\right)$$

$$\overset{then}{=} \quad \mathbf{E}\left[Y_0\right] + \widetilde{\mathbf{w}}_{R+1} c_R h^{\alpha R} + o\left(h^{\alpha R}\right).$$

Equations $(\star)$ and $(\star\star)$ read as a *Vandermonde system*

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & n_2^{-\alpha} & \cdots & n_R^{-\alpha} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & n_2^{-\alpha(R-1)} & \cdots & n_R^{-\alpha(R-1)} \end{pmatrix} \mathbf{w} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

...whose solution $\mathbf{w}$ admits a closed form given by Cramer's rule:

$$\forall i \in \{1, \ldots, R\}, \quad \mathbf{w}_i = \frac{(-1)^{R-i} n_i^{\alpha(R-1)}}{\prod_{1 \leqslant j < i} (n_i^\alpha - n_j^\alpha) \prod_{i < j \leqslant R} (n_j^\alpha - n_i^\alpha)}.$$

Furthermore $$\widetilde{\mathbf{w}}_{R+1} := \sum_{i=1}^{R} \frac{\mathbf{w}_i}{n_i^{\alpha R}} = \frac{(-1)^{R-1}}{\underline{n}!^{\alpha}} \quad (\underline{n}! = \prod_{i=1}^{R} n_i.).$$

These weights $\mathbf{w}_i$ and $\widetilde{\mathbf{w}}_{R+1}$ only depend on α.

The origins: Multistep estimator

▷ Multistep Richardson-Romberg estimator

Definition (Multistep Richardson-Romberg estimator: $\pi = (h, R, \underline{n})$)

$$\bar{Y}^N_{h,\underline{n}} = \frac{1}{N}\sum_{\ell=1}^{N}\left(\mathbf{w}_1\, Y^{\ell}_{h} + \sum_{i=2}^{R} \mathbf{w}_i\, Y^{\ell}_{\frac{h}{n_i}}\right).$$

where $(Y^{\ell}_{\frac{h}{n_i}})_{1\leqslant i\leqslant R}$, $\ell \geqslant 1$, are i.i.d. copies of $(Y_{\frac{h}{n_i}})_{1\leqslant i\leqslant R}$

- When $Y_h = f(\bar{X}_{\frac{T}{nn_i}})$ this corresponds to Euler schemes computed on the same underlying B.M./Lévy process.
- Excellent bias killer... But the cost of the estimator grows linearly in R for a prescribed $MSE = \varepsilon^2$ (quadratic error)... even with optimized refiners $n_i = i$, $i = 1, \ldots, R$.

Introduction of the multilevel paradigm

▷ We go backward using an Abel transform (with $Y_{\frac{h}{n_0}}) \equiv 0$):

$$
\begin{aligned}
\mathbf{E}\left[Y_0\right] &= \sum_{i=1}^{R} \mathbf{w}_i\, \mathbf{E}\left[Y_{\frac{h}{n_i}}\right] - \widetilde{\mathbf{w}}_{R+1} c_R h^{\alpha R} - \dots \\
&= \sum_{i=1}^{R} \underbrace{(\mathbf{w}_i + \cdots + \mathbf{w}_R)}_{=\mathbf{W}_i} \mathbf{E}\left[Y_{\frac{h}{n_i}} - Y_{\frac{h}{n_{i-1}}}\right] - \widetilde{\mathbf{w}}_{R+1} c_R h^{\alpha R} \dots \\
&= \underbrace{\mathbf{E}\left[Y_h^{(1)}\right]}_{\text{``big''}} + \sum_{i=2}^{R} \mathbf{W}_i \underbrace{\mathbf{E}\left[Y_{\frac{h}{n_i}}^{(i)} - Y_{\frac{h}{n_{i-1}}}^{(i)}\right]}_{\text{``small''}} - \widetilde{\mathbf{w}}_{R+1} c_R h^{\alpha R} \dots
\end{aligned}
$$

where $(Y_{\frac{h}{n_i}}^{(i)})_{i=1,\dots,r}$ are i.i.d. copies of $(Y_{\frac{h}{n_i}})_{i=1,\dots,r}$.

▷ This suggests to introduce the multilevel paradigm introduced by M. Giles [Giles, Oper. research, 2008] ...

Multilevel Richardson-Romberg Estimator ($ML2R$)

▷ … $i.e.$ to perform a stratification of the coarse and n_i-refined levels by assigning

$$N_i = q_i N \text{ simulations to level } i.$$

Definition ($ML2R$ Estimator $\pi = (h, R, \underline{n}, q)$)

$$\bar{Y}^{N,q}_{h,\underline{n}} := \frac{1}{N}\sum_{i=1}^{R}\frac{\mathbf{W}_i}{q_i}\sum_{\ell=1}^{N_i}\left(Y^{(i),\ell}_{\frac{h}{n_i}} - Y^{(i),\ell}_{\frac{h}{n_{i-1}}}\right),\; N \geqslant 1,$$

still with $Y^{(i),\ell}_{\frac{h}{n_0}} = 0$ and $\left((Y^{(i),\ell}_{\frac{h}{n_i}})_i\right)_{\ell\geqslant 1}$ i.i.d. copies of $(Y^{(i)}_{\frac{h}{n_i}})_i$.

- $$\mathbf{W}_i = \sum_{r=i}^{R}(-1)^{R-r}\frac{n_r^{\alpha(R-1)}}{\prod_{1\leqslant j<r}(n_r^\alpha - n_j^\alpha)\prod_{r<j\leqslant R}(n_j^\alpha - n_r^\alpha)},\; i = 1,\ldots,R.$$
- If $\mathbf{W}_i = 1$, $i = 1,\ldots,R \rightsquigarrow$ Giles' regular *MultiLevel MC* ($MLMC$).

What about Variance, complexity and effort?

▷ $$\mathrm{var}\left(\bar{Y}^{N,q}_{h,\underline{n}}\right) = \frac{1}{N}\sum_{i=1}^{R}\frac{\mathbf{W}_i^{\,2}}{q_i}\,\mathrm{var}\left(Y_{\frac{h}{n_i}} - Y_{\frac{h}{n_{i-1}}}\right)$$
$$\leqslant \frac{V_1 h^{\beta}}{N}\sum_{i=1}^{R}\frac{\mathbf{W}_i^{\,2}}{q_i}\left|\frac{1}{n_i}-\frac{1}{n_{i-1}}\right|^{\beta}.$$

▷ $$\text{Complexity}\left(\bar{Y}^{N,q}_{h,\underline{n}}\right) = \kappa\frac{N}{h}\sum_{i=1}^{R} q_i n_i$$

▷ $$\text{Effort}\left(\bar{Y}^{N,q}_{h,\underline{n}}\right) = \text{variance} \times \text{Complexity}$$ [Free of N!!]
$$= \frac{1}{\not{N}}\sum_{i=1}^{R}\frac{1}{q_i}\,\mathrm{var}\left(Y^{(i)}_{\frac{h}{n_i}} - Y^{(i)}_{\frac{h}{n_{i-1}}}\right) \times \frac{\not{N}}{h}\sum_{i=1}^{R} q_i n_i$$

What is the effort and "why" (Checklist!)

$$MSE(\bar{Y}_{h,\underline{n}}^{N,q}) := \|\bar{Y}_{h,\underline{n}}^{N,q} - \mathbf{E}\,[Y_0]\,\|_2^2 = \mathrm{Bias}\big(\bar{Y}_{h,\underline{n}}^{N,q}\big)^2 + \frac{\mathrm{Effort}(\pi)}{\mathrm{Complexity}(\pi, N)}$$

and our target problem is

$$\min_{\pi,\, MSE(\bar{Y}_{h,\underline{n}}^{N,q}) \leqslant \varepsilon^2} \left[\mathrm{Complexity}(\pi, N) = \frac{\mathrm{Effort}(\pi)}{\varepsilon^2 - \mathrm{Bias}\big(\bar{Y}_{h,\underline{n}}^{N,q}\big)^2}\right].$$

- $\varepsilon =$ prescribed $RMSE$ (L^2-error).
- The parameters of the simulation to be optimized are

$$\pi = (h, R, n_1, \ldots, n_R, q_1, \ldots, q_R).$$

(with $h = \frac{1}{K}$ for Nested MC)

Everything can be optimized explicitly in the parameter π.

From now on

$$n_i = M^{i-1}, \quad i = 1\ldots, R.$$

where M is an integer $M \geqslant 2$.

Rate optimal parameters for $ML2R$ (with $n_i = M^{i-1}$)

R	$\left\lceil \frac{1}{2} + \frac{\log(\widetilde{c}^{\frac{1}{\alpha}}\mathbf{h})}{\log(M)} + \sqrt{\left(\frac{1}{2} + \frac{\log(\widetilde{c}^{\frac{1}{\alpha}}\mathbf{h})}{\log(M)}\right)^2 + 2\frac{\log(1/\varepsilon)}{\alpha\log(M)}} \right\rceil$
h^{-1}	$\left\lceil (1+2\alpha R)^{\frac{1}{2\alpha R}} \varepsilon^{-\frac{1}{\alpha R}} M^{-\frac{R-1}{2}} \right\rceil_{\mathbf{h}}$
q	$q_1 = \mu^*(1+\theta h^{\frac{\beta}{2}})$ $q_j = \mu^*\theta h^{\frac{\beta}{2}} \left(\lvert\mathbf{W}_j(R,M)\rvert \frac{n_{j-1}^{-\frac{\beta}{2}} + n_j^{-\frac{\beta}{2}}}{\sqrt{n_{j-1}+n_j}} \right),$
N	$\dfrac{\mathrm{var}(Y_0)\left(1+\theta h^{\frac{\beta}{2}} \sum_{j=1}^{R} \lvert\mathbf{W}_j(R,M)\rvert \left(n_{j-1}^{-\frac{\beta}{2}} + n_j^{-\frac{\beta}{2}}\right)\sqrt{n_{j-1}+n_j}\right)^2}{\varepsilon^2\left(1+\frac{1}{2\alpha R}\right)^{-1} \sum_{j=1}^{R} q_j(n_{j-1}+n_j)}$

Theorem (Main (ML2R) result when $n_i = M^{i-1}$, $i = 1, \ldots, R$.)

Assume $\lim_{R\to+\infty} |c_R|^{1/R} = \widetilde{c} \in (0, +\infty)$. *The ML2R estimator satisfies*

$$\inf_{\substack{h\in\mathcal{D}, R\geqslant 2\\ \|\bar{Y}^{N,q}_{h,\underline{n}} - I_0\|_2^2 \leqslant \varepsilon^2}} \operatorname{Cost}\left(Y^{N,q}_{h,\underline{n}}\right) \lesssim K^{ML2R}_{\alpha,\beta,M} \begin{cases} \varepsilon^{-2} & \text{if } \beta > 1,\\ \varepsilon^{-2}(\log(1/\varepsilon)) & \text{if } \beta = 1,\\ \varepsilon^{-2} \underbrace{e^{\frac{1-\beta}{\sqrt{\alpha}}\sqrt{2\log(1/\varepsilon)\log(M)}}}_{=o(\varepsilon^{-\eta}),\forall\,\eta>0} & \text{if } \beta < 1, \end{cases}$$

as $\varepsilon \to 0$.

These bounds are achieved (with $M = 2$ *if* $\beta < 1$*) and an order*

$$R^*(\varepsilon) = \left\lceil \frac{1}{2} + \frac{\log(\mathbf{h}\,\widetilde{c}^{\frac{1}{\alpha}})}{\log(M)} + \sqrt{\left(\frac{1}{2} + \frac{\log(\mathbf{h}\,\widetilde{c}^{\frac{1}{\alpha}})}{\log(M)}\right)^2 + 2\frac{\log(1/\varepsilon)}{\alpha\log(M)}}\right\rceil$$

and a closed from for the bias parameter $h^* = h^*(\varepsilon, R(\varepsilon)) \in (0, \mathbf{h}]$.

Theorem (Main result (MLMC) when $n_i = M^{i-1}$, $i = 1, \ldots, R$.)

The MLMC estimator satisfies

$$\inf_{\substack{h\in\mathcal{D},R\geqslant 2\\ \|\bar{Y}^{N,q}_{h,\underline{n}}-I_0\|_2^2\leqslant\varepsilon^2}} \text{Cost}\left(Y^{N,q}_{h,\underline{n}}\right) \leqslant K_{MLMC}(\alpha,\beta,M) \precsim \begin{cases} \varepsilon^{-2} & \text{if } \beta > 1,\\ \varepsilon^{-2}(\log(1/\varepsilon))^2 & \text{if } \beta = 1,\\ \varepsilon^{-2}\times\varepsilon^{-\frac{1-\beta}{\alpha}} & \text{if } \beta < 1.\end{cases}$$

These bounds are achieved with an order

$$R^*(\varepsilon) = \left\lfloor 1 + \frac{\log\left((1+2\alpha)^{\frac{1}{2\alpha}}|c_1|^{\frac{1}{\alpha}}\mathbf{h}\right)}{\log(M)} + \frac{\log(1/\varepsilon)}{\alpha\log(M)}\right\rfloor,$$

and a closed from for the bias parameter $h^* = h^*(\varepsilon, R(\varepsilon)) \in (0,\mathbf{h}]$.

Rate optimal parameters for MLMC (with $n_i = M^{i-1}$)

R	$\left\lfloor 1 + \dfrac{\log(\lvert c_1\rvert/\varepsilon)}{\alpha \log M} \right\rfloor$
h^{-1}	$\left\lceil (1+2\alpha)^{\frac{1}{2\alpha}} \left(\dfrac{\varepsilon}{\lvert c_1\rvert}\right)^{-\frac{1}{\alpha}} M^{-(R-1)} \right\rceil_{\mathbf{h}}$
q	$q_1 = \mu^*(1+\theta h^{\frac{\beta}{2}})$ $q_j = \mu^* \theta h^{\frac{\beta}{2}} \left(\dfrac{n_{j-1}^{-\frac{\beta}{2}} + n_j^{-\frac{\beta}{2}}}{\sqrt{n_{j-1}+n_j}}\right)$,
N	$\dfrac{\mathrm{var}(Y_0)\left(1+\theta h^{\frac{\beta}{2}} \sum_{j=1}^{R} \left(n_{j-1}^{-\frac{\beta}{2}} + n_j^{-\frac{\beta}{2}}\right)\sqrt{n_{j-1}+n_j}\right)^2}{\varepsilon^2 \left(1+\frac{1}{2\alpha}\right)^{-1} \sum_{j=1}^{R} q_j (n_{j-1}+n_j)}$

Back to Nested Monte Carlo II

- Weak approximation rate: $\alpha = 1$.
- Strong approximation rate: $X \in L^2$ and X, $F(Z,X)$ absolutely continuous.

$$h = \frac{1}{K},\ h' = \frac{1}{K'},\ L = \mathbf{E}_{\mathbf{Q}}\left[F(Z,X)\,|\,X\right],\ \bar{L}_h = \frac{1}{K}\sum_{k=1}^{K} F(Z_k, X)$$

$$\left\|\bar{L}_h - \bar{L}_{h'}\right\|_2^2 \leqslant \left(\|L\|_2^2 - \left\|\mathbf{E}(L|X)\right\|_2^2\right)|h-h'|$$

so that, with $Y_0 = \varphi(L)$ and $Y_h = \varphi(\bar{L}_h)$, $h > 0$

- If φ is Lipschitz then $\beta = 1$

$$\forall\, h, h' \geqslant 0, \quad \left\|Y_h - Y_{h'}\right\|_2^2 = \left\|\varphi(\bar{L}_h) - \varphi(\bar{L}_{h'})\right\|_2^2 \leqslant [\varphi]_{\rm Lip}^2 C_{L,X}|h-h'|.$$

- If $\varphi = \varphi_y = \mathbf{1}_{(-\infty,y]}$ and $L \in L^p$, $p \geqslant 2$, then $\beta \precsim \frac{1}{2} < 1$ since

$$\forall\, h, h' \geqslant 0, \left\|Y_h - Y_{h'}\right\|_2^2 \leqslant C_0(\|f_L\|_{\sup} + 1)^{\frac{2}{3}}|h-h'|^{\frac{p}{2(p+1)}}.$$

Numerical Applications

SCR computation in [Bauer et al]. model with confidence level $\alpha = 0.92$ but without variance reduction (I.S., etc).

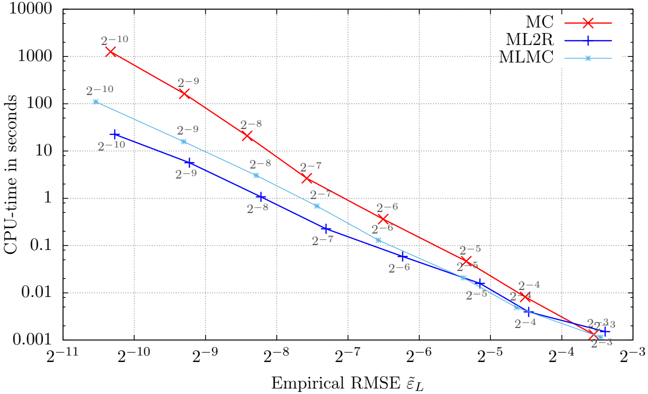

Figure: $ML2R$ — vs Nested Monte Carlo — (and vs $MLMC$ —)
$\varepsilon = 2^{-10} \approx 0.001$: $ML2R$ + (12 sec); Nested Monte Carlo × (1 200 sec)

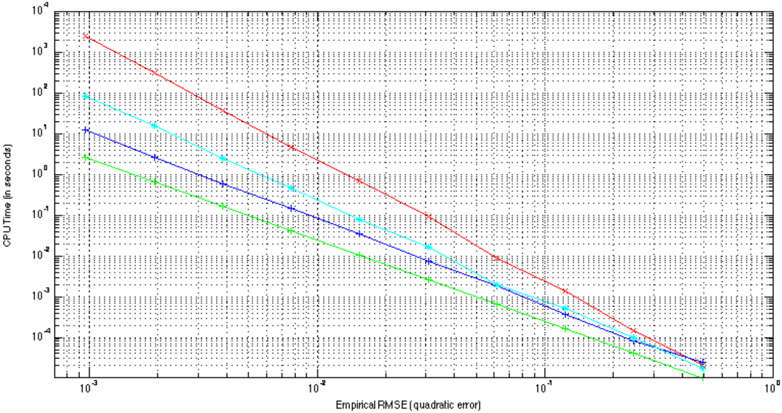

Figure: The same compared to —— Virtual unbiased simulation

▷ If $\varepsilon = 0.001$ then

ML2R Nested Monte Carlo for level probability and α-quantiles goes 100 times faster than an optimized crude Nested MC.

▷ HPC implementation: Processing the $ML2R$ simulation on a GPU (in CUDA or OpenCL) yields a second extra *multiplicative* 100 factor.

▷ Conclusion: Huge computation time contraction

$$3 \text{ hours} = 1 \text{ second.}$$

$$1 \text{ year} \approx 1 \text{ day}$$

Thank you for your attention

Theoretical application.

Theorem (Cornish-Fisher's expansion)

Quantile expansion : *Let* θ_{α}^{h} *an* a*-quantile of* $F_{\bar{L}_h}$, $\alpha \in (0,1)$. *Under the above assumptions of the above theorem*

$$\mathrm{SCR}_{\alpha}^{h} = SCR_{\alpha}^{0} + \sum_{r=1}^{R} \chi_r h^r + O(h^{R+\frac{1}{2}})$$

Quantile search by Stochastic Approximation

- Deterministic recursive zero search procedure: Let $\alpha \in (0,1)$.

$$\theta^{N+1} = \theta^N - \gamma_{N+1}\big(\mathbf{Q}(L \leqslant \theta^N) - \alpha\big)$$

with $\gamma_N = \frac{\gamma_0}{N^c}$, $\frac{1}{2} < c \leqslant 1$. Then $\theta^N \to \theta_\alpha$ as $N \to +\infty$.

- Stochastic zero search Robbins-Monto procedure: Assume $y \mapsto \mathbf{Q}(L \leqslant y)$ is unknown but L is simulatable. Let $(L^N)_{N\geqslant 1}$ be i.i.d. copies of L:

$$\begin{aligned}\theta^{N+1} &= \theta^N - \gamma_{N+1}\big(\mathbf{1}_{\{L^{N+1}\leqslant\theta^N\}} - \alpha\big)\\ &= \theta^N - \gamma_{N+1}\big(\mathbf{Q}(L \leqslant \theta^N) - \alpha\big) + \gamma_{N+1}\underbrace{\big(\mathbf{Q}(L \leqslant \theta^N) - \mathbf{1}_{\{L^{N+1}\leqslant\theta^N\}}\big)}_{=:\Delta M_{N+1}}\end{aligned}$$

with ΔM_{N+1} is a martingale increment perturbation since

$$\mathbf{Q}(L \leqslant y) = \mathbf{E}\left[\mathbf{1}_{\{L\leqslant y\}}\right].$$

As $\theta \mapsto \mathbf{Q}(L \leqslant \theta)$ is increasing one has

$$(\mathbf{Q}(L \leqslant \theta) - \alpha)(\theta - \theta_\alpha) > 0.$$

Robbins-Monro procedure

- Let $g : \mathbf{R}^d \to \mathbf{R}^d$ be a vector field having a representation

$$g(\theta) = \mathbf{E}\, G(\theta, Y_0)$$

with respect to a simulatable random vector Y_0.

- (i) *Step decrease:* $\gamma_N = \frac{\gamma_0}{N^c}$, $\frac{1}{2} < c \leqslant 1$
- (ii) *Mean reversion:* $\forall\, \theta \neq \theta^*$, $(g(\theta)\,|\,\theta - \theta^*) > 0$.
- (iii) *Linear growth (in L^2):* $\forall\, \theta \in \mathbf{R}^d$, $\big\|G(\theta, Y_0)\big\|_2 \leqslant C(1 + |\theta|)$.

then the Robbins-Monro procedure

$$\theta^{N+1} = \theta^N - \gamma_{N+1} G(\theta^N, Y_0^{(N+1)})$$

converges toward its “target”:

$$\theta^N \to \theta^* \quad a.s.$$

- When $g = \nabla V$, V (strictly) convex : stochastic gradient descent, etc.