Multilevel Monte Carlo methods for financial applications

Prof. Mike Giles

Oxford University Mathematical Institute

NAG webinar

Nov 17, 2021

Introduction

Outline:

- basic ideas
- three finance applications:
  - SDEs for option pricing
  - reduced precision computing and approximate random variables
  - nested simulations, such as CVaR

In doing this I hope to emphasise:

- the simplicity of the idea
- its flexibility – it's not prescriptive, more a generic approach
- sources of further information

Objective

To achieve a root-mean-square accuracy of ε, Monte Carlo simulation requires $O(\varepsilon^{-2})$ samples.

In many cases the cost of each sample also depends on ε, so the overall cost is often $O(\varepsilon^{-3})$ or worse.

The aim is to reduce the total cost to $O(\varepsilon^{-2})$, and reduce the cost even it is already $O(\varepsilon^{-2})$.

Control variates

Control variates are a well-established technique for reducing variance, and hence computational cost

If we want to estimate $\mathbb{E}[P(\omega)]$, and we know $\mathbb{E}[Q(\omega)]$ for some $Q(\omega)$ which is well correlated to $P(\omega)$, then

$$\mathbb{E}[P] = \underbrace{\lambda\,\mathbb{E}[Q]}_{\text{known}} + \mathbb{E}[P-\lambda Q]$$

so we can instead use Monte Carlo to estimate $\mathbb{E}[P-\lambda Q]$ and choose λ to minimise the variance, giving

$$\mathbb{V}[P-\lambda Q] = (1-\rho^2)\,\mathbb{V}[P]$$

where ρ is the correlation coefficient.

Two-level Monte Carlo

If we want to estimate $\mathbb{E}[P]$ but it is much cheaper to simulate $\widetilde{P} \approx P$, then since

$$\mathbb{E}[P] = \mathbb{E}[\widetilde{P}] + \mathbb{E}[P - \widetilde{P}]$$

we can use the estimator

$$N_0^{-1} \sum_{n=1}^{N_0} \widetilde{P}^{(0,n)} \; + \; N_1^{-1} \sum_{n=1}^{N_1} \left(P^{(1,n)} - \widetilde{P}^{(1,n)} \right)$$

Similar to a control variate except that

- we don't know analytic value of $\mathbb{E}[\widetilde{P}]$, so need to estimate it
- there is no multiplicative factor λ

Benefit: if $P - \widetilde{P}$ is small, its variance will be small, so won't need many samples to accurately estimate $\mathbb{E}[P - \widetilde{P}]$, so cost will be reduced greatly.

Two-level Monte Carlo

If we define

- C_0, V_0 cost and variance of one sample of $\widetilde{P}$
- C_1, V_1 cost and variance of one sample of $P - \widetilde{P}$

then the total cost and variance of this estimator is

$$C_{tot} = N_0\, C_0 + N_1\, C_1 \quad \Longrightarrow \quad V_{tot} = V_0/N_0 + V_1/N_1$$

Treating N_0, N_1 as real variables, using a Lagrange multiplier to minimise the cost subject to a fixed variance gives

$$\frac{\partial}{\partial N_\ell}\left(C_{tot} + \mu^2 V_{tot}\right) = 0, \quad N_\ell = \mu \sqrt{V_\ell / C_\ell}$$

Choosing μ s.t. $V_{tot} = \varepsilon^2$ gives

$$C_{tot} = \varepsilon^{-2}\left(\sqrt{V_0\, C_0} + \sqrt{V_1\, C_1}\right)^2$$

Multilevel Monte Carlo

Natural generalisation: given a sequence $\widehat{P}_0, \widehat{P}_1, \ldots, \widehat{P}_L$

$$\mathbb{E}[\widehat{P}_L] = \mathbb{E}[\widehat{P}_0] + \sum_{\ell=1}^{L} \mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}]$$

we can use the estimator

$$\widehat{Y} = N_0^{-1} \sum_{n=1}^{N_0} \widehat{P}_0^{(0,n)} \; + \; \sum_{\ell=1}^{L} \left\{ N_\ell^{-1} \sum_{n=1}^{N_\ell} \left(\widehat{P}_\ell^{(\ell,n)} - \widehat{P}_{\ell-1}^{(\ell,n)} \right) \right\}$$

with independent estimation for each level of correction

Multilevel Monte Carlo

If we define

- C_0, V_0 to be cost and variance of $\widehat{P}_0$
- C_ℓ, V_ℓ to be cost and variance of $\widehat{P}_\ell - \widehat{P}_{\ell-1}$

then the total cost is $\displaystyle\sum_{\ell=0}^{L} N_\ell \, C_\ell$ and the variance is $\displaystyle\sum_{\ell=0}^{L} N_\ell^{-1} V_\ell$.

Minimise the cost for a fixed variance

$$\frac{\partial}{\partial N_\ell} \sum_{k=0}^{L} \left(N_k \, C_k + \mu^2 N_k^{-1} V_k \right) = 0$$

gives

$$N_\ell = \mu \sqrt{V_\ell / C_\ell} \quad \Longrightarrow \quad N_\ell \, C_\ell = \mu \sqrt{V_\ell \, C_\ell}$$

Multilevel Monte Carlo

Setting the total variance equal to ε^2 gives

$$\mu = \varepsilon^{-2} \left(\sum_{\ell=0}^{L} \sqrt{V_\ell \, C_\ell} \right)$$

and hence, the total cost is

$$\sum_{\ell=0}^{L} N_\ell \, C_\ell \; = \; \varepsilon^{-2} \left(\sum_{\ell=0}^{L} \sqrt{V_\ell C_\ell} \right)^2$$

in contrast to the standard cost which is approximately $\varepsilon^{-2} \, V_0 \, C_L$.

The MLMC cost savings are therefore approximately:

- V_L / V_0, if $\sqrt{V_\ell C_\ell}$ increases with level
- C_0 / C_L, if $\sqrt{V_\ell C_\ell}$ decreases with level

Multilevel Monte Carlo

If $\widehat{P}_0, \widehat{P}_1, \ldots \longrightarrow P$, then the Mean Square Error has the decomposition

$$\begin{aligned} \mathbb{E}\left[(\widehat{Y} - \mathbb{E}[P])^2\right] &= \mathbb{V}[\widehat{Y}] + \left(\mathbb{E}[\widehat{Y}] - \mathbb{E}[P]\right)^2 \\ &= \sum_{\ell=0}^{L} V_\ell / N_\ell + \left(\mathbb{E}[\widehat{P}_L] - \mathbb{E}[P]\right)^2 \end{aligned}$$

so can choose L so that $\quad \left|\mathbb{E}[\widehat{P}_L] - \mathbb{E}[P]\right| < \varepsilon/\sqrt{2}$

and then choose N_ℓ so that $\quad \displaystyle\sum_{\ell=0}^{L} V_\ell / N_\ell < \varepsilon^2/2$

MLMC Theorem

(Slight generalisation of version in my original 2008 *Operations Research* paper, "Multilevel Monte Carlo path simulation")

If there exist independent estimators $\widehat{Y}_\ell$ based on N_ℓ Monte Carlo samples, each costing C_ℓ, and positive constants $\alpha, \beta, \gamma, c_1, c_2, c_3$ such that $\alpha \geq \frac{1}{2} \min(\beta, \gamma)$ and

i) $\left| \mathbb{E}[\widehat{P}_\ell - P] \right| \leq c_1 \, 2^{-\alpha\,\ell}$

ii) $\mathbb{E}[\widehat{Y}_\ell] = \begin{cases} \mathbb{E}[\widehat{P}_0], & \ell = 0 \\ \mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}], & \ell > 0 \end{cases}$

iii) $\mathbb{V}[\widehat{Y}_\ell] \leq c_2 \, N_\ell^{-1} 2^{-\beta\,\ell}$

iv) $\mathbb{E}[C_\ell] \leq c_3 \, 2^{\gamma\,\ell}$

MLMC Theorem

then there exists a positive constant c_4 such that for any $\varepsilon\!<\!1$ there exist L and N_ℓ for which the multilevel estimator

$$\widehat{Y} = \sum_{\ell=0}^{L} \widehat{Y}_\ell,$$

has a mean-square-error with bound $\mathbb{E}\left[\left(\widehat{Y} - \mathbb{E}[P]\right)^2\right] < \varepsilon^2$

with an expected computational cost C with bound

$$C \leq \begin{cases} c_4\, \varepsilon^{-2}, & \beta > \gamma, \\ c_4\, \varepsilon^{-2} (\log \varepsilon)^2, & \beta = \gamma, \\ c_4\, \varepsilon^{-2-(\gamma-\beta)/\alpha}, & 0 < \beta < \gamma. \end{cases}$$

MLMC Theorem

Two observations of optimality:

- MC simulation needs $O(\varepsilon^{-2})$ samples to achieve RMS accuracy ε, so when $\beta > \gamma$, the cost is optimal — $O(1)$ cost per sample on average.

  (Would need multilevel QMC to further reduce costs)

- When $\beta < \gamma$, another interesting case is when $\beta = 2\alpha$, which corresponds to $\mathbb{E}[\widehat{Y}_\ell]$ and $\sqrt{\mathbb{E}[\widehat{Y}_\ell^2]}$ being of the same order as $\ell \to \infty$.
  In this case, the total cost is $O(\varepsilon^{-\gamma/\alpha})$, which is the cost of a single sample on the finest level — again optimal.

MLMC

Numerical algorithm:

1. start with $L=0$
2. if $L < 2$, get an initial estimate for V_L using $N_L = 1000$ samples, otherwise extrapolate from earlier levels
3. for $\ell \leq L$, determine optimal N_ℓ to achieve $\displaystyle\sum_{\ell=0}^{L} V_\ell / N_\ell \leq \varepsilon^2/2$
4. perform extra calculations as needed, updating estimates of V_ℓ
5. if $L<2$ or the bias estimate is greater than $\varepsilon/\sqrt{2}$, set $L := L+1$ and go back to step 2

Application: SDEs

With SDEs, level ℓ corresponds to approximation using M^ℓ timesteps, giving approximate payoff $\widehat{P}_\ell$ at cost $C_\ell = O(M^\ell)$.

Usually choose M in the range $2-4$; often 4 for Euler-Maruyama and 2 for Milstein discretisation.

Simplest estimator for $\mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}]$ for $\ell > 0$ is

$$\widehat{Y}_\ell = N_\ell^{-1} \sum_{n=1}^{N_\ell} \left(\widehat{P}_\ell^{(n)} - \widehat{P}_{\ell-1}^{(n)} \right)$$

using same driving Brownian path for both levels.

The implementation is easy – for a coarse timestep of size $M\,h$ we simply sum the M fine path Brownian increments ΔW to get the coarse path Brownian increment.

Application: SDEs

Euler-Maruyama discretisation has $O(h^{1/2})$ strong convergence so

$$\mathbb{E}[\,(\widehat{S}_{\ell,T} - S_T)^2] = O(h_\ell) \quad\Longrightarrow\quad \mathbb{E}[(\widehat{S}_{\ell,T} - \widehat{S}_{\ell-1,T})^2] = O(h_\ell)$$

Hence for Lipschitz European payoff functions $P \equiv f(S_T)$,

$$\mathbb{V}[\widehat{P}_\ell - \widehat{P}_{\ell-1}] \le \mathbb{E}[(\widehat{P}_\ell - \widehat{P}_{\ell-1})^2] \le K^2\, \mathbb{E}[(\widehat{S}_{T,\ell} - \widehat{S}_{T,\ell-1})^2] = O(h_\ell)$$

In terms of the MLMC theorem, this means we have

$$\begin{aligned} C_\ell = O(M^\ell) &\quad\Longrightarrow\quad \gamma = \log_2 M, \\ V_\ell = O(h_\ell) = O(M^{-\ell}) &\quad\Longrightarrow\quad \beta = \log_2 M, \end{aligned}$$

so the overall cost to achieve ε RMS accuracy is $O(\varepsilon^{-2}|\log\varepsilon|^2)$.

Things are not so good for digital options – complexity is $O(\varepsilon^{-5/2})$.

MLMC SDE algorithm

Input: fine and coarse timesteps h^f, h^c, final time $T = N\, h^c$, refinement factor $M = h^c/h^f$, initial states $\widehat{S}^f = \widehat{S}^c = S_0$

for $n = 1, N$ **do**

$\Delta W^c := 0$

for $m = 1, M$ **do**

generate r.v. $\Delta W^f \sim N(0, h^f)$

$\Delta W^c := \Delta W^c + \Delta W^f$

$\widehat{S}^f := \widehat{S}^f + a(\widehat{S}^f)\, h^f + b(\widehat{S}^f)\, \Delta W^f$

end for

$\widehat{S}^c := \widehat{S}^c + a(\widehat{S}^c)\, h^c + b(\widehat{S}^c)\, \Delta W^c$

end for

$\widehat{P}_\ell - \widehat{P}_{\ell-1} := f(\widehat{S}^f) - f(\widehat{S}^c)$

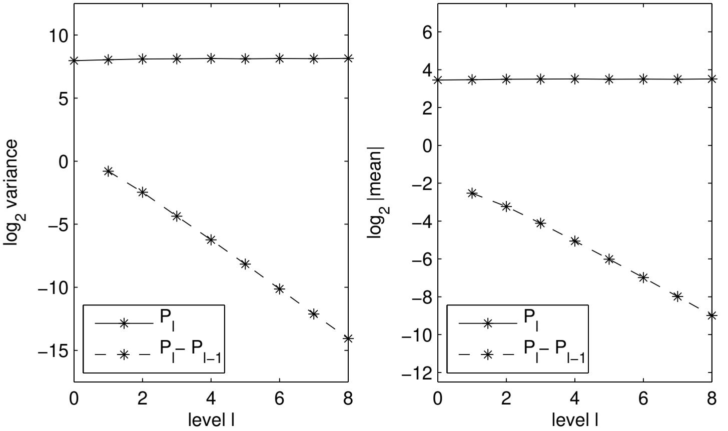
Application: SDEs

- Milstein discretisation – MBG (2008, 2012)
- MLQMC for SDEs – MBG, Waterhouse (2009)
- Greeks – Burgos (2011)
- American options – Belomestny & Schoenmakers (2011)
- jump-diffusion models – MBG, Xia (2012)
- Lévy-driven processes – Dereich (2010), Marxen (2010), Dereich & Heidenreich (2011), Kyprianou (2014)
- multi-dim Milstein without Lévy areas – MBG, Szpruch (2014)
- adaptive timesteps – Hoel, von Schwerin, Szepessy, Tempone (2012), MBG, Lester, Whittle (2014), Fang, MBG (2020)
- exponential Lévy processes – Xia (2017)

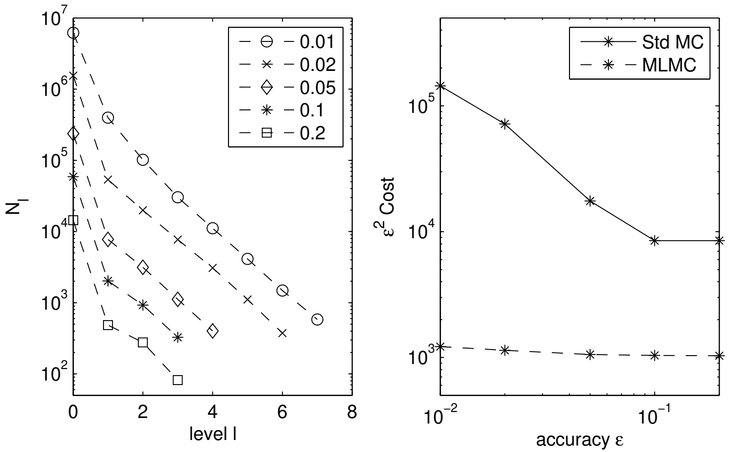
Basket call option

- 5 underlying assets, modelled by Geometric Brownian Motion with correlation between 5 driving Brownian motions
- Milstein numerical approximation
- standard call option based on average at final time T

Basket call option



Basket call option



Application: reduced precision

One simple use of two-level MLMC is with reduced precision floating point arithmetic:

- double precision on “fine” level
- single precision (or even half-precision?) on “coarse” level

This can be combined with SDE treatment by using this two-level treatment for each of the expectations $\mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}]$,

$$\mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}] \;=\; \mathbb{E}[\widehat{P}_\ell^{\text{float}} - \widehat{P}_{\ell-1}^{\text{float}}] \;+\; \mathbb{E}[(\widehat{P}_\ell - \widehat{P}_{\ell-1}) - (\widehat{P}_\ell^{\text{float}} - \widehat{P}_{\ell-1}^{\text{float}})]$$

For the low-accuracy computations can also use imprecise conversion of uniform r.v.’s to Normal r.v.’s

Application: approximate r.v.'s

Continuing with that idea, one approach for the CIR model

$$\mathrm{d}r_t = a\,(b - r)\,\mathrm{d}t + \sigma\sqrt{r}\,\mathrm{d}W_t$$

is to use exact simulation which involves sampling from the non-central χ^2 distribution.

One method of sampling is to generate a $(0,1)$ uniform r.v. U and then apply the inverse of the non-central χ^2 CDF.

Doing this accurately is expensive, but one can construct a cheap fairly accurate piecewise bilinear approximation – linear in both U and the non-centrality parameter (for fixed degrees of freedom $d = 4ab/\sigma^2$).

Looks a good approach for interest rate models and Heston stochastic volatility.

Application: nested expectation

The general form of a nested expectation is $\mathbb{E}\left[f\left(\mathbb{E}[Y\,|\,X]\right)\right]$.

The standard approach uses N outer samples $X^{(n)}$, and for each M inner samples $Y^{(m,n)}$. To achieve ε RMS accuracy usually requires $N=O(\varepsilon^{-2})$, $M=O(\varepsilon^{-1})$, so the total cost is $O(\varepsilon^{-3})$.

The MLMC approach uses $M_\ell = 2^\ell\, M_0$ inner samples on level ℓ

- cost is $O(2^\ell)$ so $\gamma=1$
- an "antithetic" construction gives $\beta=2$ if f is smooth, and $\beta=3/2$ if f is continuous and piecewise smooth; both good enough for $O(\varepsilon^{-2})$ total cost
- if f is discontinuous $\beta=1/2$ and the cost is $O(\varepsilon^{-5/2})$; can be improved to $O(\varepsilon^{-2}|\log\varepsilon|^2)$ using adaptive sampling

Application: VaR and CVaR

VaR loss L_η is defined implicitly by $\mathbb{P}[L > L_\eta] = \eta$.

Given an estimate $\widetilde{L}_\eta$, Rockafellar & Uryasev (2000) show that CVaR is

$$\begin{aligned}\mathbb{E}[\,L\,|\,L > L_\eta] &= L_\eta + \eta^{-1}\mathbb{E}[\max(0, L - L_\eta)] \\ &= \min_x \left\{x + \eta^{-1}\mathbb{E}[\max(0, L - x)]\right\} \\ &= \widetilde{L}_\eta + \eta^{-1}\mathbb{E}[\max(0, L - \widetilde{L}_\eta)] + O(\widetilde{L}_\eta - L_\eta)^2\end{aligned}$$

For ε RMS error,

- first estimate $\widetilde{L}_\eta$ to accuracy $O(\varepsilon^{1/2})$ at cost $o(\varepsilon^{-2})$
- then estimate $\eta^{-1}\mathbb{E}[\max(0, L - \widetilde{L}_\eta)]$ to accuracy ε using MLMC; $\beta = 3/2$ so total cost is $O(\varepsilon^{-2})$

Can also use random sampling to reduce cost for portfolios with lots of products.

Final comments

- MLMC has become widely used in academia over the past 10 years, and also MLQMC in some areas (mainly PDEs)
- very large savings in some application areas (especially PDEs and stochastic modelling of chemical reactions)
- very limited uptake in the finance sector so far, but I think there are very good opportunities here
- research worldwide (inc. papers) is listed on a webpage:

  `people.maths.ox.ac.uk/gilesm/mlmc_community.html`
- MLMC software and examples available on another webpage:

  `people.maths.ox.ac.uk/gilesm/mlmc/`
- my papers are on:

  `people.maths.ox.ac.uk/gilesm/mlmc.html`

  best to start with review: 'Multilevel Monte Carlo methods'. Acta Numerica, 24:259-328, 2015.