MLMC for Nested Expectations

Michael B. Giles

Dedicated to Ian H. Sloan on the occasion of his 80th birthday, with warm appreciation for the way in which he welcomed me into the MC/QMC community and introduced me to QMC methods when I switched research fields in 2006.

Abstract This paper discusses progress and future research possibilities in applying MLMC ideas to nested expectations of the form $\mathbb{E}[\,g(\mathbb{E}[f(X,Y)|X])\,]$, with an outer expectation with respect to one random variable X, and an inner conditional expectation with respect to a second random variable Y. The difficulty in treating such applications is shown to depend on whether the function g is i) smooth, ii) continuous and piecewise smooth, or iii) discontinuous.

1 Introduction

Considerable progress has been achieved over the past 10 years in the development, application and analysis of Multilevel Monte Carlo (MLMC) methods, applied to SDEs, SPDEs, continuous-time Markov processes, and a range of other stochastic models; see [10] and references therein.

This paper discusses an area of active research, the application of MLMC ideas to nested simulations, in which one is interested in estimating quantities of the form $\mathbb{E}[\,g(\mathbb{E}[f(X,Y)|X])\,]$ with an outer expectation with respect to one random variable X, and an inner conditional expectation with respect to a second random variable Y.

Such nested expectations arise in a number of applications; the two applications motivating the author's research are the evaluation of Expected Value of Partial Perfect Information (EVPPI) and Value-at-Risk (VaR).

Michael B. Giles (✉)
Mathematical Institute, University of Oxford, Oxford, UK
e-mail: mike.giles@maths.ox.ac.uk

EVPPI arises in fields such as medicine [1, 4] and the exploration and exploitation of oil and gas reservoirs [3, 24], the common element being decision making under a large degree of uncertainty. In the medical case, models of the effectiveness of different medical treatments are based on a number of uncertain parameters which we group into two independent sets X and Y. Given no knowledge of X and Y other than that they come from prescribed probability distributions, then given a finite set of possible treatments D, the optimal choice d_{opt} is the one which maximises $\mathbb{E}[f_d(X,Y)]$ where $f_d(X,Y)$ represents some measure of the patient outcome, such as QALY's (quality-adjusted life-year, see Wikipedia), with a larger value being better. Thus, with no knowledge, the expected optimal outcome is $\max_d \mathbb{E}[f_d(X,Y)]$. On the other hand, given perfect information on X,Y due to additional medical research, the best treatment maximises $f_d(X,Y)$, giving the overall expected outcome $\mathbb{E}[\max_d f_d(X,Y)]$. In the intermediate situation, if X is determined but not Y, then the best treatment has expected outcome value $\mathbb{E}[\max_d \mathbb{E}[f_d(X,Y)\,|\,X]]$. EVPI, the expected value of perfect information, is the difference

$$\text{EVPI} = \mathbb{E}[\max_d f_d(X,Y)] - \max_d \mathbb{E}[f_d(X,Y)],$$

and EVPPI, the expected value of partial perfect information, is the difference

$$\text{EVPPI} = \mathbb{E}[\max_d \mathbb{E}[f_d(X,Y)\,|\,X]] - \max_d \mathbb{E}[f_d(X,Y)].$$

EVPPI represents the benefit, on average, of knowing the value of X. This can be compared to the cost of the research required to determine X, to judge whether or not the research is cost-effective.

Value-at-Risk (VaR) is a financial risk measure used by investment banks [16, 17, 22, 23]. In this application, X represent a set of risk factors affecting the value of the bank's portfolio over some short risk horizon. For a given X, the loss in value of the portfolio is $L(X) \equiv \mathbb{E}[f(X,Y)|X]$ where the expectation corresponds to risk-neutral pricing, with Y representing the stochastic drivers for the behaviour of the underlying assets beyond the risk horizon. The objective with VaR is to compute the loss threshold L_α such that $\mathbb{P}(L(X) \geq L_\alpha) = \alpha$, for some small value of α. This defines L_α implicitly, but in this paper we will consider the simpler situation of a given threshold L^* and then computing $\mathbb{P}(L(X) \geq L^*) \equiv \mathbb{E}\left[\mathbf{1}_{\mathbb{E}[f(X,Y)|X] \geq L^*}\right]$. Hence in this case the function g is a discontinuous indicator function.

The paper begins with a quick review of MLMC and two important variants, the randomised unbiased MLMC method due to Rhee & Glynn [25], and the Multi-Index Monte Carlo (MIMC) method of Haji-Ali, Nobile & Tempone [21]. Based on material in [10], Section 3 addresses the case in which the function g is smooth, using an antithetic estimator to achieve a faster rate of multilevel variance convergence. Section 4 addresses the EVPPI problem; a similar antithetic estimator is used but the convergence is poorer due to the lack of smoothness when there is a switch in the optimal decision. Section 5 addresses the VaR problem, and the difficulty in dealing with the discontinuous indicator function, and the paper finishes with a few concluding comments.

2 MLMC and Two Important Variants

2.1 MLMC

The central idea behind MLMC is very simple: given a sequence $P_0, P_1, \ldots$ which approximates a random output variable P with increasing accuracy, but also increasing cost, we have the simple identity

$$\mathbb{E}[P_L] \;=\; \mathbb{E}[P_0] + \sum_{\ell=1}^{L} \mathbb{E}[P_\ell - P_{\ell-1}] \;=\; \sum_{\ell=0}^{L} \mathbb{E}[\Delta P_\ell], \tag{1}$$

if we define $\Delta P_\ell \equiv P_\ell - P_{\ell-1}$ and $P_{-1} \equiv 0$. Therefore, if Z_ℓ is an unbiased estimator for $\mathbb{E}[\Delta P_\ell]$ then $\sum_{\ell=0}^{L} Z_\ell$ is an estimator for $\mathbb{E}[P_L]$.

Combining this with a geometric sequence of levels, and choosing the finest level L to control the magnitude of the weak error $\mathbb{E}[P_L - P]$, leads to the usual MLMC theorem in which we assume that there exist independent estimators Z_ℓ based on N_ℓ Monte Carlo samples, each with expected cost C_ℓ and variance V_ℓ, such that there are positive constants $\alpha, \beta, \gamma, c_1, c_2, c_3$ with $\alpha \geq \frac{1}{2}\min(\beta, \gamma)$ and

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

ii) $\mathbb{E}[Z_\ell] \;=\; \mathbb{E}[\Delta P_\ell]$

iii) $V_\ell \;\leq\; c_2\, 2^{-\beta \ell}$

iv) $C_\ell \;\leq\; c_3\, 2^{\gamma \ell}$,

and then conclude that there exists a positive constant c_4 such that for any desired root-mean-square accuracy $\varepsilon < e^{-1}$ there are values L and N_ℓ for which the multilevel estimator

$$Z = \sum_{\ell=0}^{L} Z_\ell,$$

has a mean-square-error with bound

$$MSE \equiv \mathbb{E}\left[\left(Z - \mathbb{E}[P]\right)^2\right] < \varepsilon^2$$

with a computational complexity C with bound

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

In each new application, the objective is to design an estimator so that $\beta > \gamma$ to achieve the best order of complexity.

2.2 Randomised MLMC for unbiased estimation

An important extension has been introduced by Rhee & Glynn in [25]. Rather than choosing the finest level of simulation L, based on the desired accuracy, and then using the optimal number of samples on each level based on an estimate of the variance, their "single term" estimator instead uses N samples in total, and for each sample they perform a level ℓ simulation with probability $p_\ell > 0$, with $\sum_{\ell=0}^{\infty} p_\ell = 1$.

The estimator is

$$Z = \frac{1}{N} \sum_{n=1}^{N} \Delta P^{(n)}_{\ell^{(n)}} / p_{\ell^{(n)}}$$

with the level $\ell^{(n)}$ for each sample being selected randomly with the relevant probability, so that

$$\mathbb{E}[Z] = \sum_{\ell} \mathbb{E}[\Delta P_\ell] = \mathbb{E}[P].$$

Hence, it is an unbiased estimator.

The choice of the probabilities p_ℓ is crucial. For both the variance and the expected cost to be finite, it is necessary that

$$\sum_{\ell=0}^{\infty} V_\ell / p_\ell \ < \ \infty, \quad \sum_{\ell=0}^{\infty} p_\ell C_\ell \ < \ \infty.$$

Under the conditions of the usual MLMC theorem, this is possible when $\beta > \gamma$ by choosing $p_\ell \propto 2^{-(\gamma+\beta)\ell/2}$, so that

$$V_\ell / p_\ell \propto 2^{-(\beta-\gamma)\ell/2}, \quad p_\ell C_\ell \propto 2^{-(\beta-\gamma)\ell/2}.$$

It is not possible when $\beta \le \gamma$, and for these cases the estimators in [25] have infinite expected cost.

2.3 Multi-Index Monte Carlo

In standard MLMC, there is a one-dimensional set of levels, with a scalar level index ℓ, although in some applications changing ℓ can change more than one aspect of the computation, such as both timestep and spatial discretisation in a parabolic SPDE application [14]. In [21], Haji-Ali, Nobile & Tempone generalised this, with the Multi-Index Monte Carlo (MIMC) method defining "levels" in multiple directions, so that the level index $\boldsymbol{\ell}$ is now a vector of integer indices. This is illustrated in Figure 1 for a 2D MIMC application.

Generalising (1) to D dimensions in [21], Haji-Ali, Nobile & Tempone first define a backward difference operator in one particular dimension, $\Delta_d P_{\boldsymbol{\ell}} \equiv P_{\boldsymbol{\ell}} - P_{\boldsymbol{\ell}-\boldsymbol{e}_d}$ where $\boldsymbol{e}_d$ is the unit vector in direction d, and then define the cross-difference

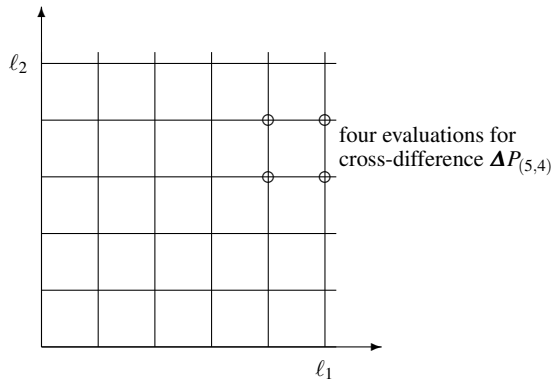

Fig. 1 "Levels" in 2D multi-index Monte Carlo application

$$\boldsymbol{\Delta} P_{\boldsymbol{\ell}} \equiv \left(\prod_{d=1}^{D} \Delta_d \right) P_{\boldsymbol{\ell}}$$

so that the telescoping sum becomes

$$\mathbb{E}[P] = \sum_{\boldsymbol{\ell} \geq 0} \mathbb{E}[\boldsymbol{\Delta} P_{\boldsymbol{\ell}}]. \tag{2}$$

As an example, Figure 1 marks the four locations at which $P_{\boldsymbol{\ell}}$ must be computed to determine the value of $\boldsymbol{\Delta} P_{(5,4)}$ in the 2D application.

Following the presentation in [10], the MIMC theorem formulated in [21] can be expressed in a form which matches quite closely the formulation of the MLMC theorem. If the level $\boldsymbol{\ell}$ MIMC estimator $Z_{\boldsymbol{\ell}}$, with variance $V_{\boldsymbol{\ell}}$ and cost $C_{\boldsymbol{\ell}}$, per sample, satisfies

i) $\left| \mathbb{E}[P_{\boldsymbol{\ell}} - P] \right| \longrightarrow 0$ as $\min_d \ell_d \longrightarrow \infty$

ii) $\mathbb{E}[Z_{\boldsymbol{\ell}}] = \mathbb{E}[\boldsymbol{\Delta} P_{\boldsymbol{\ell}}]$

iii) $\left| \mathbb{E}[Z_{\boldsymbol{\ell}}] \right| \leq c_1 2^{-\boldsymbol{\alpha} \cdot \boldsymbol{\ell}}$

iv) $V_{\boldsymbol{\ell}} \leq c_2 2^{-\boldsymbol{\beta} \cdot \boldsymbol{\ell}}$

v) $C_{\boldsymbol{\ell}} \leq c_3 2^{\boldsymbol{\gamma} \cdot \boldsymbol{\ell}}$,

then the complexity is $O(\varepsilon^{-2})$ provided $\beta_d > \gamma_d$ for all dimensions d, with additional $|\log \varepsilon|$ factors introduced if $\beta_d = \gamma_d$ for some d.

This complexity is achieved by truncating the set of increments in Eq. (2). It might seem natural that the summation region $\mathscr{L}$ should be rectangular, as illustrated on the left in Figure 2, so that

$$\sum_{\boldsymbol{\ell} \in \mathscr{L}} \mathbb{E}[Z_{\boldsymbol{\ell}}] = \mathbb{E}[P_{\boldsymbol{L}}]$$

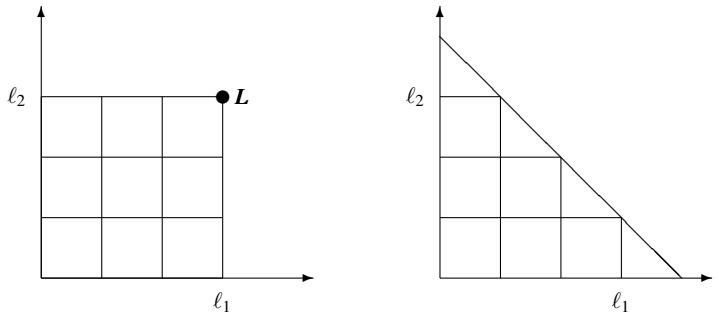

Fig. 2 Two choices of 2D MIMC summation region $\mathscr{L}$.

where $\boldsymbol{L}$ is the outermost point on the rectangle. However, [21] proves that in general this does not give the optimal order of complexity, and instead it is often best to use a region which in 2D is triangular, as illustrated on the right in Figure 2. This is very similar to the use of sparse grid methods in high-dimensional PDE approximations [7], and indeed MIMC can be viewed as a combination of sparse grid methods and Monte Carlo sampling.

3 The General Smooth Case

In this first section, we consider the case in which g is a smooth function. A particular case of interest is the VaR application which was discussed in the Introduction. If one can estimate moments of the loss function $L(X)$, then an approximation of the loss CDF can be generated using Maximum Entropy reconstruction [2, 19]. The critical loss value L_α can then be determined from this CDF approximation.

3.1 MLMC treatment

Following the presentation in [10], we are interested in estimating quantities of the form $\mathbb{E}\left[g\left(\mathbb{E}[f(X,Y)|X]\right)\right]$ where X is an outer random variable, and $\mathbb{E}[f(X,Y)|X]$ is a conditional expectation with respect to an independent inner random variable Y.

This can be simulated using nested Monte Carlo simulation with N outer samples $X^{(n)}$, M inner samples $Y^{(m,n)}$ and a standard Monte Carlo estimator:

$$Z = N^{-1}\sum_{n=1}^{N} g\left(M^{-1}\sum_{m=1}^{M} f(X^{(n)},Y^{(m,n)})\right)$$

Note that to improve the accuracy of the estimate we need to increase both M and N, and this will significantly increase the cost. In fact, it can be proved [18] that the root-mean-square error is $O(M^{-1}+N^{-1/2})$, so to achieve r.m.s. accuracy of ε it is best to choose $M=O(\varepsilon^{-1})$, $N=O(\varepsilon^{-2})$, giving a complexity which is $O(\varepsilon^{-3})$.

An MLMC implementation is straightforward; on level ℓ we can use $M_\ell = 2^\ell$ inner samples. To construct a low variance estimate for $\mathbb{E}[P_\ell - P_{\ell-1}]$ where

$$\mathbb{E}[P_\ell] \equiv \mathbb{E}\left[g\left(M_\ell^{-1} \sum_{m=1}^{M_\ell} f(X, Y^{(m)}) \right) \right],$$

we use an *antithetic* approach and split the M_ℓ samples for the "fine" value into two subsets of size $M_{\ell-1}$ for the "coarse" value:

$$\begin{aligned} Z_\ell = N_\ell^{-1} \sum_{n=1}^{N_\ell} \Bigg\{ & g\left(M_\ell^{-1} \sum_{m=1}^{M_\ell} f(X^{(n)}, Y^{(m,n)}) \right) \\ & - \tfrac{1}{2} g\left(M_{\ell-1}^{-1} \sum_{m=1}^{M_{\ell-1}} f(X^{(n)}, Y^{(m,n)}) \right) \\ & - \tfrac{1}{2} g\left(M_{\ell-1}^{-1} \sum_{m=M_{\ell-1}+1}^{M_\ell} f(X^{(n)}, Y^{(m,n)}) \right) \Bigg\} \end{aligned}$$

Note that this has the correct expectation, i.e. $\mathbb{E}[Z_\ell] = \mathbb{E}[P_\ell - P_{\ell-1}]$.

If we define

$$\begin{aligned} M_{\ell-1}^{-1} \sum_{m=1}^{M_{\ell-1}} f(X^{(n)}, Y^{(m,n)}) &= \mathbb{E}[f(X^{(n)}, Y)] + \Delta f_1^{(n)}, \\ M_{\ell-1}^{-1} \sum_{m=M_{\ell-1}+1}^{M_\ell} f(X^{(n)}, Y^{(m,n)}) &= \mathbb{E}[f(X^{(n)}, Y)] + \Delta f_2^{(n)}, \end{aligned}$$

then if g is twice differentiable a Taylor series expansion gives

$$Z_\ell \approx -\frac{1}{4N_\ell} \sum_{n=1}^{N_\ell} g''\left(\mathbb{E}[f(X^{(n)}, Y)] \right) \left(\Delta f_1^{(n)} - \Delta f_2^{(n)} \right)^2.$$

By the Central Limit Theorem, $\Delta f_1^{(n)}, \Delta f_2^{(n)} = O(M_\ell^{-1/2})$ and therefore

$$g''\left(\mathbb{E}[f(X^{(n)}, Y)] \right) \left(\Delta f_1^{(n)} - \Delta f_2^{(n)} \right)^2 = O(M_\ell^{-1}).$$

It follows that $\mathbb{E}[Z_\ell] = O(M_\ell^{-1})$ and $V_\ell = O(M_\ell^{-2})$. For the MLMC theorem, this corresponds to $\alpha = 1$, $\beta = 2$, $\gamma = 1$, so the complexity is $O(\varepsilon^{-2})$.

This antithetic approach to nested simulation has been developed independently by several authors [6, 8, 20], and is related to an earlier use of an antithetic MLMC estimator for SDEs [15].

[20] used it in a mean field model for the motion of crowds, in which each person is modelled as a independent agent subject to random forcing and an additional force due to the collective influence of the crowd. This same approach is also relevant to mean field problems which arise in plasma physics [26].

[6] used multilevel nested simulation for a financial credit derivative application. In their case, the function g was piecewise linear, not twice differentiable, and so the rate of variance convergence was slightly lower, with $\beta = 1.5$. This will be discussed in Section 4, but it is still sufficient to achieve an overall $O(\varepsilon^{-2})$ complexity.

3.2 MIMC treatment

The previous analysis assumes we can compute $f(X,Y)$ with $O(1)$ cost, but suppose now that Y represents a complete Brownian path, and $f(X,Y)$ cannot be evaluated exactly; it can only be approximated using some finite number of timesteps. Using MLMC, on level ℓ we could use 2^ℓ timesteps and a Milstein discretisation (giving first order weak and strong convergence) which would still give $\alpha = 1$, $\beta = 2$. However, we would now have $\gamma = 2$, because on successive levels we would be using twice as many timesteps as well as twice as many inner samples. This then leads to an overall MLMC complexity which is $O(\varepsilon^{-2}(\log\varepsilon)^{-2})$.

Instead we can use MIMC to recover an optimal complexity of $O(\varepsilon^{-2})$. We now have a pair of level indices (l_1, l_2), with the number of inner samples equal to 2^{ℓ_1} and the number of timesteps proportional to 2^{ℓ_2}. If we use the natural extension of the MLMC estimator to the corresponding MIMC estimator, which means (for $l_1 > 0, l_2 > 0$) using

$$\begin{aligned} Z_\ell = N_\ell^{-1} \sum_{n=1}^{N_\ell} \Bigg\{ & g\left(2^{-\ell_1} \sum_{m=1}^{2^{\ell_1}} f_{\ell_2}(X^{(n)}, Y^{(m,n)})\right) - \tfrac{1}{2} g\left(2^{-\ell_1+1} \sum_{m=1}^{2^{\ell_1-1}} f_{\ell_2}(X^{(n)}, Y^{(m,n)})\right) \\ & - \tfrac{1}{2} g\left(2^{-\ell_1+1} \sum_{m=2^{\ell_1-1}+1}^{2^{\ell_1}} f_{\ell_2}(X^{(n)}, Y^{(m,n)})\right) \\ & - g\left(2^{-\ell_1} \sum_{m=1}^{2^{\ell_1}} f_{\ell_2-1}(X^{(n)}, Y^{(m,n)})\right) + \tfrac{1}{2} g\left(2^{-\ell_1+1} \sum_{m=1}^{2^{\ell_1-1}} f_{\ell_2-1}(X^{(n)}, Y^{(m,n)})\right) \\ & + \tfrac{1}{2} g\left(2^{-\ell_1+1} \sum_{m=2^{\ell_1-1}+1}^{2^{\ell_1}} f_{\ell_2-1}(X^{(n)}, Y^{(m,n)})\right) \Bigg\} \end{aligned}$$

The subscript on the f terms denotes the level of timestep approximation.

Carrying out the same analysis as before, performing the Taylor series expansion around $\mathbb{E}[f(X^{(n)}, Y)]$, we obtain

$$Z_\ell \approx -\frac{1}{4N_\ell} \sum_{n=1}^{N_\ell} g''\left(\mathbb{E}[f(X(n), Y)]\right) \left\{ \left(\Delta f_{1,\ell_2}^{(n)} - \Delta f_{2,\ell_2}^{(n)}\right)^2 - \left(\Delta f_{1,\ell_2-1}^{(n)} - \Delta f_{2,\ell_2-1}^{(n)}\right)^2 \right\}.$$

The difference of squares can be re-arranged as

$$\left(\Delta f^{(n)}_{1,\ell_2}-\Delta f^{(n)}_{2,\ell_2}\right)^2-\left(\Delta f^{(n)}_{1,\ell_2-1}-\Delta f^{(n)}_{2,\ell_2-1}\right)^2 = \left((\Delta f^{(n)}_{1,\ell_2}+\Delta f^{(n)}_{1,\ell_2-1})-(\Delta f^{(n)}_{2,\ell_2}+\Delta f^{(n)}_{2,\ell_2-1})\right)\times \left((\Delta f^{(n)}_{1,\ell_2}-\Delta f^{(n)}_{1,\ell_2-1})-(\Delta f^{(n)}_{2,\ell_2}-\Delta f^{(n)}_{2,\ell_2-1})\right)$$

Due to the Central Limit Theorem, we have

$$\Delta f^{(n)}_{1,\ell_2}+\Delta f^{(n)}_{1,\ell_2-1}=O(2^{-\ell_1/2}),\quad \Delta f^{(n)}_{2,\ell_2}+\Delta f^{(n)}_{2,\ell_2-1}=O(2^{-\ell_1/2}),$$

and assuming first order strong convergence we also have

$$\Delta f^{(n)}_{1,\ell_2}-\Delta f^{(n)}_{1,\ell_2-1}=O(2^{-\ell_1/2-\ell_2}),\quad \Delta f^{(n)}_{2,\ell_2}-\Delta f^{(n)}_{2,\ell_2-1}=O(2^{-\ell_1/2-\ell_2}).$$

Combining these results we obtain

$$\left(\Delta f^{(n)}_{1,\ell_2}-\Delta f^{(n)}_{2,\ell_2}\right)^2-\left(\Delta f^{(n)}_{1,\ell_2-1}-\Delta f^{(n)}_{2,\ell_2-1}\right)^2=O(2^{-\ell_1-\ell_2})$$

and therefore $\mathbb{E}[Z_\ell]=O(2^{-\ell_1-\ell_2})$ and $V_\ell=O(2^{-2\ell_1-2\ell_2})$ with a cost per sample which is $O(2^{\ell_1+\ell_2})$. In the MIMC theorem this corresponds to $\alpha_1=\alpha_2=1$, $\beta_1=\beta_2=2$, and $\gamma_1=\gamma_2=1$, so the overall complexity is $O(\varepsilon^{-2})$.

3.3 Nested MLMC

MIMC is not the only way in which to generalise MLMC to multiple dimensions. Another option, which can sometimes be equivalent, but is often not, is to use nested MLMC, with an inner MLMC being used to generate samples within an outer MLMC computation.

The application in the previous section gives rise to a natural example of this. Ideally, we would like to generate exact samples of $f(X,Y)$ with $O(1)$ cost per sample. However, it is just as good to produce samples which have the correct expected value $\mathbb{E}[f(X,Y)|X]$, with an expected cost which is $O(1)$. This can be achieved by using the randomised MLMC discussed in Section 2.2, so that $f(X^{(n)},Y^{(m,n)})$ is replaced by

$$\left(f(X^{(n)},Y^{(m,n)}_\ell)-f(X^{(n)},Y^{(m,n)}_{\ell-1})\right)\ /\ p_\ell,$$

where the level ℓ which determines the number of timesteps is a random variable taking integer value $\ell'\geq 0$ with probability $p_{\ell'}>0$. The only requirement is that the variance for this inner randomised MLMC must decay faster with the number of timesteps than the increase in the computational cost, so that $p_{\ell'}$ can be specified appropriately to achieve both finite variance and finite expected cost.

4 EVPPI

For the estimation of the difference EVPI – EVPPI defined in the Introduction, we define a level ℓ approximation as

$$P_\ell = \overline{\max_d f_d}^\ell - \max_d \overline{f_d}^\ell$$

where $\overline{\max_d f_d}^\ell$ and $\overline{f_d}^\ell$ represent averages over 2^ℓ independent values of $Y^{(i)}$ for one particular value of X, so that

$$\text{EVPI} - \text{EVPPI} = \lim_{\ell\to\infty} \mathbb{E}[P_\ell].$$

Following the ideas in [6, 10, 20] we use the antithetic MLMC estimator

$$Z_\ell \;=\; \tfrac{1}{2}\left(\max_d \overline{f_d}^{(a)} + \max_d \overline{f_d}^{(b)}\right) \;-\; \max_d \overline{f_d}$$

where

- $\overline{f_d}^{(a)}$ is an average of $f_d(X,Y)$ over $2^{\ell-1}$ independent samples for Y;
- $\overline{f_d}^{(b)}$ is an average over a second independent set of $2^{\ell-1}$ samples;
- $\overline{f_d}$ is an average over the combined set of 2^ℓ inner samples.

The MLMC variance can be analysed by following the approach used by Giles & Szpruch for Theorem 5.2 in [15], which is also similar to the analysis by Bujok *et al* in [6]. Define

$$F_d(X) = \mathbb{E}_Y\left[f_d(X,Y)\right], \quad d_{opt}(X) = \arg\max_d F_d(X)$$

so the domain for X is divided into a number of regions in which the optimal decision $d_{opt}(X)$ is uniform, with a dividing lower-dimensional decision manifold K on which $d_{opt}(X)$ is not uniquely-defined.

Note that $\frac{1}{2}(\overline{f_d}^{(a)} + \overline{f_d}^{(b)}) - \overline{f_d} = 0$, and therefore $Z_\ell = 0$ if the same decision d maximises each of the terms in its definition. This is the key advantage of the antithetic estimator, compared to the alternative $\overline{f_d}^{(a)} - \overline{f_d}$. When ℓ is large and so there are many samples, $\overline{f_d}^{(a)}, \overline{f_d}^{(b)}, \overline{f_d}$ will all be close to $F_d(X)$, and therefore it is highly likely that $Z_\ell = 0$ unless X is very close to K at which there is more than one optimal decision. This idea leads to a theorem on the MLMC variance, but first we need to make three assumptions.

Assumption 1. $\mathbb{E}\left[\,|f_d(X,Y)|^p\right]$ *is finite for all* $p \geq 2$.

Comment: this enables us to bound the difference between $\overline{f_d}^{(a)}, \overline{f_d}^{(b)}, \overline{f_d}$ *and* $F_d(X)$.

Assumption 2. *There exists a constant* $c_0 > 0$ *such that for all* $0 < \varepsilon < 1$

$$\mathbb{P}\left(\min_{x\in K} \|X - x\| \leq \varepsilon\right) \leq c_0\,\varepsilon.$$

Comment: this bounds the probability of X being close to the decision manifold K.

Assumption 3. *There exist constants $c_1, c_2 > 0$ such that if $X \notin K$, then*

$$\max_d F_d(X) - \max_{d \neq d_{opt}(X)} F_d(X) \;>\; \min\left(c_1,\; c_2 \min_{x \in K} \|X - x\|\right).$$

Comment: on K itself there are at least 2 decisions d_1, d_2 which yield the same optimal value $F_d(X)$; this assumption ensures at least a linear divergence between the values as X moves away from K.

Theorem 1. *If Assumptions 1-3 are satisfied, and Z_ℓ is as defined previously for level ℓ, then for any $\delta > 0$*

$$\mathbb{V}[Z_\ell] = o(2^{-(3/2-\delta)\ell}), \quad \mathbb{E}[Z_\ell] = o(2^{-(1-\delta)\ell}).$$

The proof of the theorem is given in [11], but a heuristic explanation is as follows:

- Because of Assumption 1, for any X, $\overline{f_d} - F_d(X) = O(2^{-\ell/2})$;
- Because of Assumption 2, there is an $O(2^{-\ell/2})$ probability of X being within $O(2^{-\ell/2})$ of the decision manifold K, in which case $Z_\ell = O(2^{-\ell/2})$;
- Because of Assumption 3, if it is further away from K then there is a clear separation between the different decision values, and hence $Z_\ell = 0$ with very high probability.
- This results in $\mathbb{E}[Z_\ell^2] = O(2^{-\ell/2}) \times \left(O(2^{-\ell/2})\right)^2 = O(2^{-3\ell/2})$.

The conclusion from the theorem is that the parameters for the MLMC theorem are $\beta \approx 3/2$, $\alpha \approx 1$, and $\gamma = 1$, giving the optimal complexity of $O(\varepsilon^{-2})$. Numerical results support this prediction.

A final comment is that sometimes the random variables in X or Y correspond to Bayesian posterior distributions, with samples generated by MCMC methods. In that case, it is possible to pre-generate a large set of MCMC samples, after the initial burn-in, and then MLMC can uniformly and randomly take samples from this dataset as required.

5 Value-at-Risk

The Value-at-Risk problem has been defined in the Introduction. In this section, we begin by introducing the idea of portfolio sub-sampling, and then proceed to discuss the difficulties in constructing efficient MLMC estimators for VaR because of the discontinuous nature of the indicator function.

5.1 Portfolio sub-sampling

In Sections 3.2 and 3.3, we considered Y to represent the driving Brownian motion, and the inner conditional expectation was with respect to this. However, in the context of the Value-at-Risk application, where we are considering estimation of moments of the loss for the purpose of Maximum Entropy reconstruction, there is an important second aspect to this conditional expectation.

The loss function $L(X)$ has contributions from a large number of financial options within a portfolio, so that it may be written as

$$L(X) = \sum_{i=1}^{N_o} L_i(X), \quad L_i(X) \equiv \mathbb{E}\left[f_i(X,Y)\,|X\right],$$

where $L_i(X)$ is the loss from the i^{th} option. In the existing literature, standard treatments evaluate each sample of the total loss by summing the contributions from all of the financial options, and the computational cost is inevitably proportional to N_o, the number of options. However, instead we can express the loss as

$$L(X) = \mathbb{E}\left[L_i(X)/p_i\right].$$

where the integer index i is randomly sampled from the set $\{1,2,\ldots,N_o\}$ with probability p_i.

Adding back in the expectation with respect to the Brownian motion we obtain the conditional expectation

$$L(X) = \mathbb{E}\left[f_i(X,Y)/p_i\,|X\right],$$

in which the expectation is now over both the Brownian motion and the index of the option being sampled. When 2^ℓ samples are generated to approximate the conditional expectation, they each can have a different option index as well as a different Brownian path sample. The overall benefit is to achieve a complexity, for a given accuracy ε expressed as a fraction of the total portfolio value, which no longer depends on N_o, the number of financial options in the portfolio.

This idea of sub-sampling a portfolio has been investigated by Wenhui Gou [19] whose research combined it with Maximum Entropy reconstruction of the loss distribution, but used an analytic expression for the conditional expectation with respect to the driving Brownian motion, and also used a control variate which substantially reduced the variance of the estimator.

5.2 Previous work on VaR

As explained in the Introduction, we are interested in determining

$$\mathbb{P}\left[L(X) \geq L^*\right] \equiv \mathbb{E}\left[\mathbf{1}\left(L(X) \geq L^*\right)\right]$$

This is again a nested simulation problem, but the indicator function makes it much harder than EVPPI, because small differences between the "coarse" and "fine" estimates for the conditional expectation in $L(X)$ can lead to a ± 1 change in the indicator value.

Gordy & Juneja [18] considered this problem, using a single level Monte Carlo method with M inner samples $Y^{(m,n)}$ for each of N outer samples $X^{(n)}$ to estimate

$$L(X) \equiv \mathbb{E}[f(X,Y)]$$

for each of N outer samples $X^{(n)}$, so that the overall estimate for the probability of exceeding the loss threshold is

$$\mathbb{P}\left[L(X) \geq L^*\right] \approx N^{-1} \sum_{n=1}^{N} \mathbf{1}\left(M^{-1} \sum_{m=1}^{M} f(X^{(n)}, Y^{(m,n)} \geq L^*)\right)$$

This problem setup assumes that it is possible to exactly simulate $f(X^{(n)},Y)$ at unit cost. Given this, they proved that the resulting RMS error is

$$O(M^{-1}+N^{-1/2}),$$

and hence, to achieve an ε RMS accuracy requires $M = O(\varepsilon^{-1})$, $N = O(\varepsilon^{-2})$ and so the complexity is $O(\varepsilon^{-3})$.

Broadie, Du & Moallemi [5] improved on this, by noting that unless $L(X) - L^*$ is small, we usually don't need many samples to determine whether $L(X) \geq L^*$. Their paper presents a rigorously analysed adaptive algorithm based on the theory of sequential sampling but here we give a simplified heuristic analysis. When using M inner samples, if

$$\sigma^2(X) = \mathbb{V}[f(X,Y)|X], \quad d(X) = \left|\mathbb{E}[f(X,Y)|X] - L^*\right|$$

then the usual CLT confidence interval for the estimate of $\mathbb{E}[f(X,Y)|X] - L^*$ has size $\pm 3\,\sigma/\sqrt{M}$. Hence, we need roughly

$$M = 9\,\sigma^2(X)/d^2(X)$$

inner samples to be sure whether or not $\mathbb{E}[f(X,Y)|X] \geq L^*$. If we now use

$$M = \min\left(c\,\varepsilon^{-1}, 9\,\sigma^2(X)/d^2(X)\right)$$

then the cross-over point between the two terms in the minimum is at $d = O(\varepsilon^{1/2})$, and it follows that the the average number of inner samples required is

$$\overline{M} = O(\varepsilon^{-1/2}),$$

reducing the overall complexity to $O(\varepsilon^{-5/2})$.

This is clearly a significant improvement on the complexity of the uniform sampling algorithm of Gordy & Juneja, but in both papers they are not using the subsampling introduced in Section 5.1 but are instead evaluating the full portfolio each time so the complexity is also proportional to the number of options in the portfolio. Furthermore, their analysis does not consider the additional cost which is incurred when one needs to approximate an SDE for the underlying assets.

5.3 Current research

Current research by the author and Abdul-Lateef Haji-Ali builds on the adaptive approach of Broadie *et al* [5] by incorporating MLMC ideas.

The first step is to extend Wenhui Gou's work to Monte Carlo estimation of the inner conditional expectations:

$$\sum_{i=1}^{N_o} \mathbb{E}[f_i(X,Y)|X] \approx M^{-1} \sum_{m=1}^{M} f_{i_m}(X,W_m) \;/\; p_{i_m}$$

where W_m represents the Brownian path and any additional random inputs needed for the conditional expectation. This essentially combines, or unifies, the Monte Carlo averaging over the portfolio samples with the averaging over the Brownian paths.

If we do this with the uniform inner sampling with $M_\ell = 4^\ell$ samples on level ℓ, assuming that $f_{p_m}(X,W_m)$ can be computed exactly at unit cost, then the error in the inner estimate is $O(M_\ell^{-1/2}) = O(2^{-\ell})$. There is an $O(2^{-\ell})$ probability of being within $O(2^{-\ell})$ of the indicator step, producing an $O(1)$ value for the MLMC estimator sample, so the MLMC variance is $V_\ell \sim 2^{-\ell}$. In addition we get bias $\sim M_\ell^{-1} \sim 4^{-\ell}$, $C_\ell \sim M_\ell \sim 4^\ell$, so $\alpha \approx 2$, $\beta \approx 1$, $\gamma \approx 2$ and therefore the complexity is $O(\varepsilon^{-5/2})$. The advantage over the previous method due to Broadie *et al* is that the complexity is independent of the value of N_o the number of options in the portfolio, but it still falls short of our target of $O(\varepsilon^{-2})$.

To further improve things, we add in the adaptive approach of Broadie *et al*, with the number of inner samples dependent on both X and the level ℓ, along the lines of

$$M_\ell(X) = \max\left(c_1\, 2^\ell, \min\left(c_2\, 4^\ell, 9\,\sigma^2(X)/d^2(X)\right)\right).$$

This gives approximately the same asymptotic behaviour in the variance and the bias, i.e. bias $\sim 4^{-\ell}$, $V_\ell \sim 2^{-\ell}$, but the cost is reduced to approximately $C_\ell \sim 2^\ell$. This

leads to $\alpha \approx 2, \beta \approx 1, \gamma \approx 1$ and hence the complexity is approximately $O(\varepsilon^{-2})$, independent of N_o.

The final challenge comes from the approximation of the underlying SDE. At first sight this looks very difficult, but the algorithm does not require the exact sampling of $f_{i_m}(X, W_m)$; it is sufficient to have an unbiased estimate with a unit expected cost. Following the ideas in Section 3.3, this is precisely what can be supplied in many cases by Rhee & Glynn's unbiased single-term estimator based on randomised MLMC. This requires the use of the Milstein time discretisation, because of the improved strong order of convergence and hence rate of MLMC variance convergence compared to an Euler-Maruyama discretisation. The complexity analysis is largely unchanged, and again we achieve an overall complexity of approximately $O(\varepsilon^{-2})$, to within log terms.

In practice, it is also very important to use an effective control variate, similar to the one used by Wenhui Gou [19], but the details are omitted here.

5.4 Future research

There are other aspects of the VaR problem to be investigated in the future.

One is associated with the fact that the different financial options within a portfolio vary greatly, both in their variance (in part due to differences in their financial magnitude) and in the computational cost involved in their simulation. Both of these factors need to be taken into account in optimising the probability p_i for sampling the option with index i. It might even be desirable to identify a few options which should always be sampled because of their large value, and apply the randomised sub-sampling to the remainder.

Secondly, the discussion so far has been about the simpler problem of determining

$$\mathbb{P}(L(X) \geq L^*) \equiv \mathbb{E}\left[\mathbf{1}_{L(X)\geq L^*}\right],$$

for some given loss value L^*. The research must be extended to the full VaR definition which requires some root-finding algorithm to determine L_α defined implicitly for some α by

$$\mathbb{P}(L(X) \geq L_\alpha) = \alpha.$$

We also need to consider other risk measures such as CVaR, or expected shortfall,

$$\text{CVaR} = \mathbb{E}[L(X) \mid L(X) \geq L_\alpha] = \alpha^{-1}\,\mathbb{E}\left[L(X)\,\mathbf{1}_{L(X)\geq L_\alpha}\right].$$

6 Conclusions

In this paper we have reviewed progress in applying MLMC ideas to problems with nested expectations. Such applications lead quite naturally to the use of the Multi-Index Monte Carlo method and other generalisations of MLMC such as nested MLMC. Randomised MLMC for the inner conditional expectation is particularly helpful as it is unbiased, which simplifies the treatment.

One important nested expectation application is the estimation of EVPPI, the Expected Value of Partial Perfect Information. Substantial progress has been made on this topic, in both the construction and the analysis of efficient algorithms.

In the context of the financial Value-at-Risk application, we have pointed out the benefits to be achieved from sub-sampling the portfolio. Combining this with an adaptive MLMC estimator addresses the challenge due to the discontinuous indicator function in the outer expectation. This use of adaptive algorithms within MLMC fits well with other current research [9, 12, 13].

Acknowledgements

The author is very grateful to a number of people for discussions and collaborations: Ian Sloan and Frances Kuo on alternatives to the Multi-Index Monte Carlo method; Takashi Goda and Howard Thom on EVPPI estimation; Wenhui Gou, Abdul-Lateef Haji-Ali, Ralf Korn and Klaus Ritter on Value-at-Risk estimation.

References

1. Ades, A., Lu, G., Claxton, K.: Expected value of sample information calculations in medical decision modeling. Medical Decision Making **24**(2), 207–227 (2004)
2. Bierig, C., Chernov, A.: Approximation of probability density functions by the multilevel Monte Carlo Maximum Entropy method. Journal of Computational Physics **314**, 661–681 (2016)
3. Bratvold, R., Bickel, J., Lohne, H.: Value of information in the oil and gas industry: past, present, and future. SPE Reservoir Evaluation and Engineering **12**, 630–638 (2009)
4. Brennan, A., Kharroubi, S., O'Hagan, A., Chilcott, J.: Calculating partial expected value of perfect information via Monte Carlo sampling algorithms. Medical Decision Making **27**, 448–470 (2007)
5. Broadie, M., Du, Y., Moallemi, C.: Efficient risk estimation via nested sequential simulation. Management Science **57**(6), 1172–1194 (2011)
6. Bujok, K., Hambly, B., Reisinger, C.: Multilevel simulation of functionals of Bernoulli random variables with application to basket credit derivatives. Methodology and Computing in Applied Probability **17**(3), 579–604 (2015)
7. Bungartz, H.J., Griebel, M.: Sparse grids. Acta Numerica **13**, 147–269 (2004)
8. Chen, N., Liu, Y.: Estimating expectations of functionals of conditional expected via multilevel nested simulation. Presentation at conference on Monte Carlo and Quasi-Monte Carlo Methods, Sydney (2012)

9. Fang, W., Giles, M.: Adaptive Euler-Maruyama method for SDEs with non-globally Lipschitz drift: Part I, finite time interval. ArXiv preprint: 1609.08101 (2016)
10. Giles, M.: Multilevel Monte Carlo methods. Acta Numerica **24**, 259–328 (2015)
11. Giles, M., Goda, T.: Decision-making under uncertainty: using MLMC for efficient estimation of EVPPI. ArXiv preprint: 1708.05531 (2017)
12. Giles, M., Lester, C., Whittle, J.: Non-nested adaptive timesteps in multilevel Monte Carlo computations. In: R. Cools, D. Nuyens (eds.) Monte Carlo and Quasi-Monte Carlo Methods 2014. Springer (2016)
13. Giles, M., Ramanan, K.: MLMC with adaptive timestepping for reflected Brownian diffusions. In preparation (2017)
14. Giles, M., Reisinger, C.: Stochastic finite differences and multilevel Monte Carlo for a class of SPDEs in finance. SIAM Journal of Financial Mathematics **3**(1), 572–592 (2012)
15. Giles, M., Szpruch, L.: Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation. Annals of Applied Probability **24**(4), 1585–1620 (2014)
16. Glasserman, P., Heidelberger, P., Shahabuddin, P.: Variance reduction techniques for estimating Value-at-Risk. Management Science **46**, 1349–1364 (2000)
17. Glasserman, P., Heidelberger, P., Shahabuddin, P.: Portfolio Value-at-Risk with heavy-tailed risk factors. Mathematical Finance **12**, 239–269 (2002)
18. Gordy, M., Juneja, S.: Nested simulation in portfolio risk measurement. Management Science **56**(10), 1833–1848 (2010)
19. Gou, W.: Estimating Value-at-Risk using multilevel Monte Carlo Maximum Entropy method. MSc thesis, University of Oxford (2016).
20. Haji-Ali, A.L.: Pedestrian flow in the mean-field limit. MSc thesis, KAUST (2012).
21. Haji-Ali, A.L., Nobile, F., Tempone, R.: Multi Index Monte Carlo: when sparsity meets sampling. Numerische Mathematik **132**(4), 767–806 (2016)
22. Korn, R., Korn, E., Kronstadt, G.: Monte Carlo Methods and Models in Finance and Insurance. Chapman & Hall / CRC Financial Mathematics (2010)
23. Korn, R., Pupashenko, M.: A new variance reduction technique for estimating value-at-risk. Applied Mathematical Finance **22**(1), 83–98 (2015)
24. Nakayasu, M., Goda, T., Tanaka, K., Sato, K.: Evaluating the value of single-point data in heterogeneous reservoirs with the expectation maximization. SPE Economics and Management **8**, 1–10 (2016)
25. Rhee, C.H., Glynn, P.: Unbiased estimation with square root convergence for SDE models. Operations Research **63**(5), 1026–1043 (2015)
26. Rosin, M., Ricketson, L., Dimits, A., Caflisch, R., Cohen, B.: Multilevel Monte Carlo simulation of Coulomb collisions. Journal of Computational Physics **247**, 140–157 (2014)