Stochastic Simulation: Lecture 9a

Christoph Reisinger

Oxford University Mathematical Institute

Modified from earlier slides by Prof. Mike Giles.

Objectives

In presenting the multilevel Monte Carlo method, I want to emphasise:

- the simplicity of the idea
- its flexibility – it's not prescriptive, more an approach
- there are a variety of applications – there are lots of people around the world working on these

In this lecture I will focus on the fundamental ideas; in the next lecture we'll run through the implementation and an application.

Monte Carlo method

In stochastic models, we often have

$$\begin{array}{ccccc} \omega & \longrightarrow & S & \longrightarrow & P \\ \text{random input} & & \text{intermediate variables} & & \text{scalar output} \end{array}$$

The Monte Carlo estimate for $\mathbb{E}[P]$ is an average of N independent samples $\omega^{(n)}$:

$$Y = N^{-1} \sum_{n=1}^{N} P(\omega^{(n)}).$$

This is unbiased, $\mathbb{E}[Y] = \mathbb{E}[P]$, and as $N \to \infty$ the error becomes Normally distributed with variance $N^{-1} V$ where $V = \mathbb{V}[P]$.

RMS error of ε requires $N = \varepsilon^{-2} V$ samples, at a total cost of $\varepsilon^{-2} V\, C$, if C is the cost of a single sample.

Monte Carlo method

In many cases, this is modified to

$$\begin{array}{ccccc} \omega & \longrightarrow & \widehat{S} & \longrightarrow & \widehat{P} \\ \text{random input} & & \text{intermediate variables} & & \text{scalar output} \end{array}$$

where $\widehat{S}, \widehat{P}$ are approximations to S, P, in which case the MC estimate

$$\widehat{Y} = N^{-1} \sum_{n=1}^{N} \widehat{P}(\omega^{(n)})$$

is biased, and the Mean Square Error is

$$\mathbb{E}[\,(\widehat{Y} - \mathbb{E}[P])^2] \;=\; N^{-1}\, \mathbb{V}[\widehat{P}] + \left(\mathbb{E}[\widehat{P}] - \mathbb{E}[P]\right)^2$$

Greater accuracy requires both larger N and smaller weak error $\mathbb{E}[\widehat{P}] - \mathbb{E}[P]$.

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

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

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

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

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 $\left|\mathbb{E}[\widehat{P}_L] - \mathbb{E}[P]\right| < \varepsilon / \sqrt{2}$

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

MLMC Theorem

First version in: Mike Giles, "Multilevel Monte Carlo path simulation" *Operations Research*, 2008.
This slight generalisation from Giles & Reisinger, 2012.

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 generalisation

The theorem is for scalar outputs P, but it can be generalised to multi-dimensional (or infinite-dimensional) outputs with

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) $\mathbb{V}[\widehat{Y}_\ell] \equiv \mathbb{E}\left[\left\| \widehat{Y}_\ell - \mathbb{E}[\widehat{Y}_\ell] \right\|^2 \right] \leq c_2\, N_\ell^{-1} 2^{-\beta\,\ell}$

Original multilevel research by Heinrich in 1999 did this for parametric integration, estimating $g(\lambda) \equiv \mathbb{E}[f(x,\lambda)]$ for a finite-dimensional r.v. x.

Three MLMC extensions

- unbiased estimation – Rhee & Glynn (2015)
  - randomly selects the level for each sample
  - no bias, and finite expected cost and variance if $\beta > \gamma$
- Richardson-Romberg extrapolation – Lemaire & Pagès (2013)
  - reduces the weak error, and hence the number of levels required
  - particularly helpful when $\beta < \gamma$
- Multi-Index Monte Carlo – Haji-Ali, Nobile, Tempone (2015)
  - important extension to MLMC approach, combining MLMC with sparse grid methods

Randomised Multilevel Monte Carlo

Rhee & Glynn (2015) started from

$$\mathbb{E}[P] \;=\; \sum_{\ell=0}^{\infty} \mathbb{E}[\Delta P_\ell] \;=\; \sum_{\ell=0}^{\infty} p_\ell \; \mathbb{E}[\Delta P_\ell / p_\ell],$$

to develop an unbiased single-term estimator

$$Y = \Delta P_{\ell'} \,/\, p_{\ell'},$$

where ℓ' is a random index which takes value ℓ with probability p_ℓ.

$\beta > \gamma$ is required to simultaneously obtain finite variance and finite expected cost using

$$p_\ell \propto 2^{-(\beta+\gamma)\ell/2}.$$

The complexity is then $O(\varepsilon^{-2})$.

Multilevel Richardson-Romberg extrapolation

If the weak error on level ℓ satisfies

$$\mathbb{E}[Y_\ell - Y] = \sum_{j=1}^{L+1} c_j\, 2^{-\alpha j \ell} \;+\; r_{L,\ell}, \qquad |r_{L,\ell}| \leq C_{L+2}\, 2^{-\alpha(L+2)\ell}$$

then

$$\sum_{\ell=0}^{L} w_\ell\, \mathbb{E}[Y_\ell] = \left(\sum_{\ell=0}^{L} w_\ell \right) \mathbb{E}[Y] + \sum_{j=1}^{L+1} c_j \left(\sum_{\ell=0}^{L} w_\ell\, 2^{-\alpha j \ell} \right) \;+\; R_L,$$

with $|R_L| \leq C_{L+2} \sum_{\ell=0}^{L} (|w_\ell|\, 2^{-\alpha(L+2)\ell})$.

We want to estimate $\mathbb{E}[Y]$, so choose w_ℓ to satisfy

$$\sum_{\ell=0}^{L} w_\ell = 1, \qquad \sum_{\ell=0}^{L} w_\ell\, 2^{-\alpha j \ell} = 0, \quad j = 1, \ldots, L.$$

Multilevel Richardson-Romberg extrapolation

Given these weights, we then obtain

$$\sum_{\ell=0}^{L} w_\ell \, \mathbb{E}[Y_\ell] = \mathbb{E}[Y] + c_{L+1} \widetilde{w}_{L+1} + R_L,$$

where (see paper by Pagès and Lemaire)

$$\widetilde{w}_{L+1} = \sum_{\ell=0}^{L} w_\ell \, 2^{-\alpha(L+1)\ell} = (-1)^L \, 2^{-\alpha L(L+1)/2},$$

which is asymptotically much larger than $|R_L|$, but also very much smaller than the usual MLMC bias.

Multilevel Richardson-Romberg extrapolation

To complete the ML2R formulation we need to set

$$W_\ell \;=\; \sum_{\ell'=\ell}^{L} w_{\ell'} \;=\; 1 - \sum_{\ell'=0}^{\ell-1} w_{\ell'}\,.$$

$$\Longrightarrow \quad \sum_{\ell=0}^{L} w_\ell\, \mathbb{E}[Y_\ell] = W_0\, \mathbb{E}[Y_0] + \sum_{\ell=1}^{L} W_\ell\, \mathbb{E}[\Delta Y_\ell].$$

The big difference from MLMC is that now we need just

$$L_{\rm ML2R} \sim \sqrt{|\log_2 \varepsilon|/\alpha}$$

which is much better than the usual

$$L_{\rm MLMC} \sim |\log_2 \varepsilon|/\alpha$$

and can give good savings when $\beta \leq \gamma$.

Multi-Index Monte Carlo

Standard "1D" MLMC truncates the telescoping sum

$$\mathbb{E}[P] = \sum_{\ell=0}^{\infty} \mathbb{E}[\Delta \widehat{P}_\ell]$$

where $\Delta \widehat{P}_\ell \equiv \widehat{P}_\ell - \widehat{P}_{\ell-1}$, with $\widehat{P}_{-1} \equiv 0$.

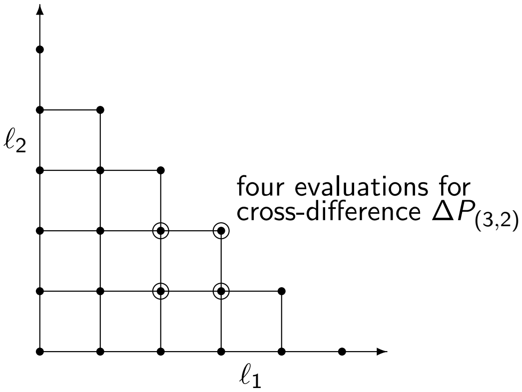
In "2D", MIMC truncates the telescoping sum

$$\mathbb{E}[P] = \sum_{\ell_1=0}^{\infty} \sum_{\ell_2=0}^{\infty} \mathbb{E}[\Delta \widehat{P}_{\ell_1,\ell_2}]$$

where $\Delta \widehat{P}_{\ell_1,\ell_2} \equiv (\widehat{P}_{\ell_1,\ell_2} - \widehat{P}_{\ell_1-1,\ell_2}) - (\widehat{P}_{\ell_1,\ell_2-1} - \widehat{P}_{\ell_1-1,\ell_2-1})$

Different aspects of the discretisation vary in each "dimension"

Multi-Index Monte Carlo



MIMC truncates the summation in a way which minimises the cost to achieve a target MSE – quite similar to sparse grids.

Can achieve $O(\varepsilon^{-2})$ complexity for a wider range of applications than plain MLMC.

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. 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

MLQMC

For further improvement in overall computational cost, can switch to QMC instead of MC for each level.

- use randomised QMC, with 32 random offsets/shifts
- define $V_{N_\ell,\ell}$ to be variance of average of 32 averages using N_ℓ QMC points within each average
- objective is therefore to achieve

$$\sum_{\ell=0}^{L} V_{N_\ell,\ell} \leq \varepsilon^2/2$$

- process to choose L is unchanged, but what about N_ℓ?

MLQMC

Numerical algorithm:

1. start with $L = 0$
2. get an initial estimate for $V_{1,L}$ using 32 random offsets and $N_L = 1$
3. while $\sum_{\ell=0}^{L} V_{N_\ell,\ell} > \varepsilon^2/2$, try to maximise variance reduction per unit cost by doubling N_ℓ on the level with largest value of $V_{N_\ell,\ell} \,/\, (N_\ell \, C_\ell)$
4. if $L < 2$ or the bias estimate is greater than $\varepsilon/\sqrt{2}$, set $L := L+1$ and go back to step 2

Final comments

- MLMC has become widely used in the past 10 years, and also MLQMC in some application areas (mainly PDEs)
- will cover a range of applications in this course
- most applications have a geometric structure as in the main MLMC theorem, but a few don't
- research worldwide is listed on a webpage: `people.maths.ox.ac.uk/gilesm/mlmc_community.html` along with links to all relevant papers