Multilevel Monte Carlo methods

Mike Giles

Mathematical Institute, University of Oxford

OMI Internal Seminar

April 30, 2014

Monte Carlo simulation

Interested in estimating the expected value of a function of a random variable $\mathbb{E}[P(\omega)]$.

The simplest estimator is an average of N independent samples

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

where $\widehat{P}$ is an approximation to P.

The Mean Square Error is

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

so greater accuracy requires more samples, and better accuracy for each sample — both drive up the cost.

Two-level Monte Carlo

If we want to estimate $\mathbb{E}[\widehat{P}_1]$ but it is much cheaper to simulate $\widehat{P}_0 \approx \widehat{P}_1$, then since

$$\mathbb{E}[\widehat{P}_1] = \mathbb{E}[\widehat{P}_0] + \mathbb{E}[\widehat{P}_1 - \widehat{P}_0]$$

we can use the estimator

$$N_0^{-1} \sum_{n=1}^{N_0} \widehat{P}_0^{(n)} \; + \; N_1^{-1} \sum_{n=1}^{N_1} \left(\widehat{P}_1^{(n)} - \widehat{P}_0^{(n)} \right)$$

If C_0, C_1 and V_0, V_1 are the cost and variance of $\widehat{P}_0$, $\widehat{P}_1 - \widehat{P}_0$ then

$$\text{total cost} = N_0\, C_0 + N_1\, C_1, \quad \text{total variance} = N_0^{-1}\, V_0 + N_1^{-1}\, V_1$$

so can optimise N_0/N_1 to minimise cost for given accuracy.

Multilevel Monte Carlo

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

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

so we can use the estimator

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

with independent estimation for each level

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 $\displaystyle\sum_{\ell=0}^{L} N_\ell^{-1} V_\ell$.

Using a Lagrange multiplier μ^2 to 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 Path Simulation

In 2006, I introduced the multilevel approach for infinite-dimensional integration arising from SDEs driven by Brownian diffusion.

Level ℓ corresponds to approximation using 2^ℓ timesteps, giving approximate payoff $\widehat{P}_\ell$.

Choice of finest level L depends on weak error (bias).

To make RMS error less than ε

- choose L so that $\left(\mathbb{E}[\widehat{P}_L] - \mathbb{E}[P]\right)^2 < \frac{1}{2}\,\varepsilon^2$
- choose $N_\ell \propto \sqrt{V_\ell / C_\ell}$ so total variance is less than $\frac{1}{2}\,\varepsilon^2$

MLMC Theorem

(Slight generalisation of original version)

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

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. When $\beta > \gamma$, the cost is optimal — $O(1)$ cost per sample on average.
- 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 Theorem

MLMC Theorem allows a lot of freedom in constructing the multilevel estimator. I sometimes use different approximations on the coarse and fine levels:

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

The telescoping sum still works provided

$$\mathbb{E}\left[\widehat{P}_\ell^f\right] \; = \; \mathbb{E}\left[\widehat{P}_\ell^c\right].$$

Given this constraint, can be creative to reduce the variance

$$\mathbb{V}\left[\widehat{P}_\ell^f - \widehat{P}_{\ell-1}^c\right].$$

Nested simulation

In some applications (especially in pensions / insurance industry for risk assessment?) interested in estimating quantities of the form

$$\mathbb{E}_Z \left[f\Big(\mathbb{E}[W|Z]\Big) \right]$$

- Z represents different risk *scenarios*
- $\mathbb{E}[W|Z]$ represents exposure, conditional on the scenario
- f might be an indicator function, to determine the percentage of scenarios under which the company has a loss in excess of its capital reserves
- alternatively, f might correspond to the expected loss in excess of the capital reserves.

Nested simulation

How is this simulated?

Nested Monte Carlo simulation:

- N *outer* samples $Z^{(n)}$
- M *inner* samples $W^{(m,n)}$, conditional on $Z^{(n)}$

$$\widehat{Y} = \frac{1}{N} \sum_{n=1}^{N} f\left(\frac{1}{M} \sum_{m=1}^{M} W^{(m,n)} \right)$$

Both M and N need to be increased to improve the accuracy of the estimate

Nested simulation

How can we use MLMC?

- level ℓ uses $M_\ell = 2^\ell$ inner samples
- (coarser levels could also use fewer, larger timesteps and perhaps small representative subsets of the full portfolio held by the company)

To estimate $\mathbb{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}]$ with a low variance we use an *antithetic* "trick":

$$\widehat{Y}_\ell \;=\; \frac{1}{N_\ell} \sum_{n=1}^{N_\ell} \left\{ f\left(\frac{1}{M_\ell} \sum_{m=1}^{M_\ell} W^{(m,n)} \right) \right. \\ \left. - \tfrac{1}{2} f\left(\frac{1}{M_{\ell-1}} \sum_{m=1}^{M_{\ell-1}} W^{(m,n)} \right) - \tfrac{1}{2} f\left(\frac{1}{M_{\ell-1}} \sum_{m=M_{\ell-1}+1}^{M_\ell} W^{(m,n)} \right) \right\}$$

which has the correct expectation.

Nested simulation

If we define

$$\frac{1}{M_{\ell-1}} \sum_{m=1}^{M_{\ell-1}} W^{(m,n)} = \mathbb{E}[W \,|\, Z^{(n)}] + \Delta W_1^{(n)}$$

$$\frac{1}{M_{\ell-1}} \sum_{m=M_{\ell-1}+1}^{M_\ell} W^{(m,n)} = \mathbb{E}[W \,|\, Z^{(n)}] + \Delta W_2^{(n)}$$

then if f is twice differentiable a Taylor series expansion gives

$$\widehat{Y} \approx -\frac{1}{4\,N_\ell} \sum_{n=1}^{N_\ell} f''(\mathbb{E}[W \,|\, Z^{(n)}]) \left(\Delta W_1^{(n)} - \Delta W_2^{(n)}\right)^2$$

$\Delta W_1^{(n)}, \Delta W_2^{(n)} = O(M_\ell^{-1/2})$ and hence $V_\ell = O(M_\ell^{-2})$. For the MLMC theorem, this corresponds to $\beta = 2, \gamma = 1$, so the complexity is $O(\varepsilon^{-2})$.

My MLMC research

- Lévy processes (Yuan Xia) and Greeks (Sylvestre Burgos)
- elliptic SPDEs (with Rob Scheichl, Bath)
- parabolic SPDEs (with Christoph Reisinger)
- multi-dimensional Milstein (with Lukas Szpruch, Edinburgh)
- approximating CDF (with Klaus Ritter, Kaiserslautern)
- continuous-time Markov procs (Ruth Baker, Kit Yates, Chris Lester)
- engineering SDE application (with Endre Süli)
- MLMC with reduced basis functions (with Jaime Peraire, MIT)
- stopped and reflected diffusions
- adaptive timestepping (with Raul Tempone, KAUST ?)
- nested simulation? mean field games?

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

MLMC Community

Abo Academi (Avikainen) – numerical analysis
Basel (Harbrecht) – elliptic SPDEs, sparse grid links
Bath (Kyprianou, Scheichl, Shardlow) – elliptic SPDEs, MCMC, Lévy-driven SDEs
Chalmers (Lang) – SPDEs
Christian-Albrechts University (Gnewuch) – multilevel QMC
Duisburg (Belomestny) – Bermudan and American options
Edinburgh (Davie, Szpruch) – SDEs, numerical analysis
ETH Zürich (Jenny, Jentzen, Schwab) – numerical analysis, SPDEs
Frankfurt (Gerstner, Kloeden) – numerical analysis, sparse grid links
Fraunhofer ITWM (Iliev) – SPDEs in engineering
Hong Kong (Chen) – Brownian meanders, nested simulation in finance
IIT Chicago (Hickernell) – SDEs, infinite-dimensional integration, complexity analysis
Kaiserslautern (Heinrich, Korn, Ritter) – finance, SDEs, complexity analysis, parametric integration
KAUST (Tempone) – adaptive time-stepping
Kiel (Gnewuch) – randomized multilevel QMC
Mannheim (Neuenkirch) – numerical analysis, fractional Brownian motion
Marburg (Dereich) – Lévy-driven SDEs
Munich (Hutzenthaler) – numerical analysis
Oxford (Giles, Hambly, Reisinger) – SDEs, jump-diffusion, SPDEs, numerical analysis
Passau (Müller-Gronbach) – infinite-dimensional integration, complexity analysis
Purdue (Gittelson) – SDPEs
Stanford (Glynn) – numerical analysis
Strathclyde (Higham, Mao) – numerical analysis, exit times, stochastic chemical modelling
Stuttgart (Barth) – SPDEs
Texas A&M (Efendiev) – SPDEs in engineering
UCLA (Caflisch) – Coulomb collisions in physics
UNSW (Dick, Kuo, Sloan) – multilevel QMC
WIAS (Schoenmakers) – Bermudan and American options
Wisconsin (Anderson) – numerical analysis, stochastic chemical modelling

Webpage: `people.maths.ox.ac.uk/gilesm/mlmc_community.html`

Conclusions

- MLMC is a very simple idea, but can provide very significant savings
- requires the construction of a hierarchy of approximation
- the savings are greatest when the coarsest approximations are much cheaper than the most accurate
- limited benefits ($10\times$ at most?) for pricing short-dated options?
- perhaps more significant opportunities with nested simulations?
- lots of MLMC research on a range of different applications
  – 100 journal articles in past 5 years