Introduction to Computational Physics

Monte Carlo Methods - A glimpse on Multilevel Monte Carlo (MLMC)

Andreas Adelmann

Paul Scherrer Institut, Villigen
E-mail: andreas.adelmann@psi.ch

https://moodle-app2.let.ethz.ch/course/view.php?id=18025

7.8 Motivation Multilevel Monte Carlo (MLMC) I

This is based on several articles and presentations of Michael B. Giles, the inventor of MLMC. In stochastic models we often have

$$\underset{\text{random input}}{\omega} \longrightarrow \qquad \underset{\text{intermediate value}}{S} \longrightarrow \qquad \underset{\text{scalar output}}{P}$$

We search for the MC estimate for $\mathbb{E}[P]$ is an average of N samples $P(\omega^i)$

$$Y = \frac{1}{N}\sum_{i=1}^{N} P(\omega^i).$$

This is unbiased $\mathbb{E}[Y] = \mathbb{E}[P]$ and the CLT proves that as $N \to \infty$ the error becomes Normally distribute with variance $\frac{1}{N}\mathbb{V}\text{ar}[P] < \infty$ and hence on needs $N = \mathbb{O}(\varepsilon^{-2})$ samples to achieve ε RMS accuracy. In this particular case we are able to do the calculations exactly however, this is the exception.

7.8 Motivation Multilevel Monte Carlo (MLMC) II

In many cases we have

$$\underset{\text{random input}}{\omega \longrightarrow} \qquad \underset{\text{intermediate value}}{\hat{S} \longrightarrow} \qquad \underset{\text{scalar output}}{\hat{P}}$$

where $\hat{S}, \hat{P}$ are approximations (e.g. from discretisaton) to S, P in which case the MC estimate

$$\hat{Y} = \frac{1}{N} \sum_{i=1}^{N} \hat{P}(\omega^i).$$

is biased, and the Mean Square Error (MSE) is

$$\mathbb{E}[(\hat{Y} - \mathbb{E}(P)^2] = \underbrace{\frac{1}{N}\mathbb{V}\text{ar}[\hat{P}]}_{\text{as before}} + \underbrace{(\mathbb{E}(\hat{P}) - \mathbb{E}(P))^2}_{\textbf{weak error}},$$

i.e. the discretization error in expectation.

Example Stochastic Differential Equation (SDE) I

Lets look at a simple stochastic differential equation (SDE) for path simulation, appearing in finance physics etc. We look at the simplest scalar form

$$dS_t = a(S_t, t)dt + b(S_t, t)dE_t$$

with a we denote the deterministic drift term, b is the diffusion term and W is the increment of a Brownian motion – Normally distributed with variance dt.
This can be approximated by the Euler-Maruyama (standard Euler forward time integration) method

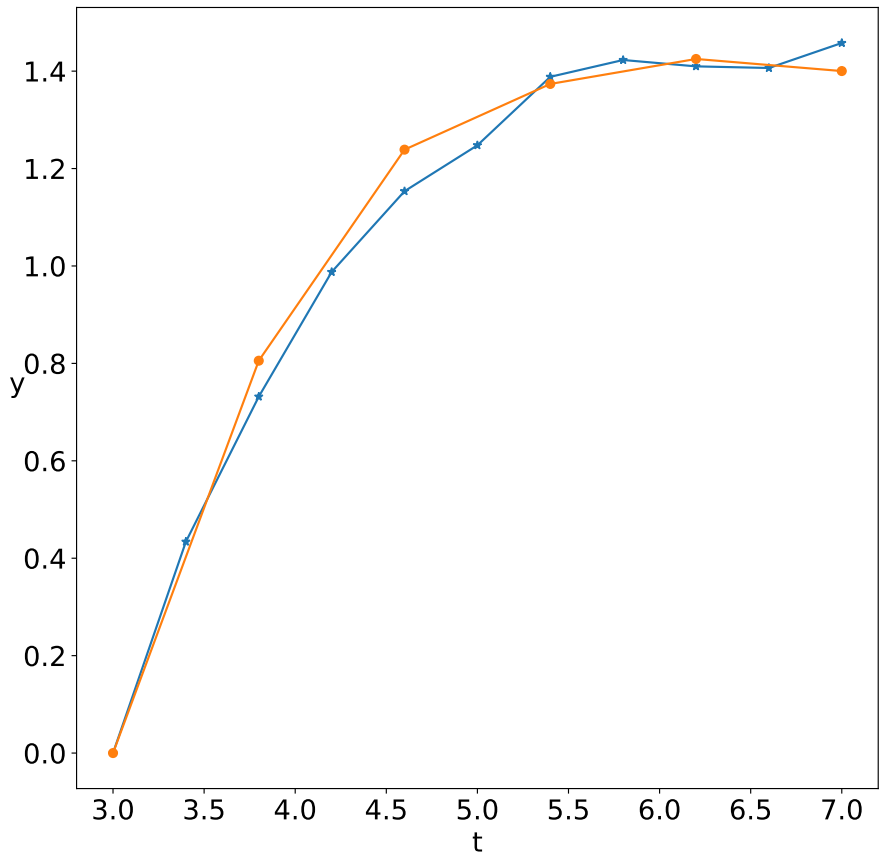
$$\hat{S}_{t_{n+1}} = \hat{S}_{t_n} + a(\hat{S}_{t_n} \cdot t_n)h + b(\hat{S}_{t_n} \cdot t_n)\Delta W_n$$

with uniform time step h and increments ΔW_n (change of Brownian path) with variance h. In simple applications, the output of interest is as function of the final value

$$\hat{P} \equiv f(\hat{S}_T).$$

Example Stochastic Differential Equation (SDE) II

Solving standard geometric Brownian Motion for 2 different timesteps:



Example Stochastic Differential Equation (SDE) III

Two kinds of discretization errors are involved here:

1. weak error:
$$\mathbb{E}[(\hat{P})] - \mathbb{E}[(P)] = \mathbb{O}(h)$$
2. strong error (in expectaion):
$$\left(\mathbb{E}\left[\sup_{[0,T]} (\hat{S}_t - S_t)^2 \right] \right)^{1/2} = \mathbb{O}(h^{1/2})$$

Better schemes exists for example Milstein discretization for which both weak and strong error are of $\mathbb{O}(h)$. The MSE is

Example Stochastic Differential Equation (SDE) IV

$$\frac{1}{N}\mathbb{V}\mathrm{ar}[\hat{P}] + (\mathbb{E}(\hat{P}) - \mathbb{E}(P))^2 = aN^{-1} + bh^2$$

If we want this to be ε^2, then we need

$$N = \mathbb{O}(\varepsilon^{-2}), \quad h = \mathbb{O}(\varepsilon)$$

hence, the total computational cost is

$$\mathbb{O}(\varepsilon^{-3})$$

this is the number of path $\times$ the number of time steps per path.
To improve we can work on two fronts:

- reduce N - variance reduction or Quasi-Monte Carlo methods
- reduce the cost of each path - MLMC

7.9 Control variates I

One of the classic approaches to Monte Carlo variance reduction is through the use of a control variate. Suppose we wish to estimate $\mathbb{E}[f]$, and there is a control variate g which is well correlated to f and has a known expectation $\mathbb{E}[g]$. In that case, we can use the following unbiased estimator for $\mathbb{E}[f]$:

$$\frac{1}{N}\sum_{n=1}^{N}\left\{f^{(n)}-\lambda\left(g^{(n)}-\mathbb{E}[g]\right)\right\}. \tag{1}$$

The optimal value for λ is $\rho\sqrt{\mathbb{V}\mathrm{ar}[f]\,/\,\mathbb{V}\mathrm{ar}[g]}$, where ρ is the correlation between f and g, and the variance of the control variate estimator is reduced by factor $1-\rho^2$ compared to the standard estimator.
A two-level version of MLMC (multilevel Monte Carlo) is very similar.

7.10 Two-level MC I

If we want to estimate $\mathbb{E}[P_1]$ but it is much cheaper to simulate $P_0 \approx P_1$, then since

$$\mathbb{E}[P_1] = \mathbb{E}[P_0] + \mathbb{E}[P_1 - P_0] \tag{2}$$

we can use the unbiased two-level estimator

$$N_0^{-1} \sum_{n=1}^{N_0} P_0^{(n)} \; + \; N_1^{-1} \sum_{n=1}^{N_1} \left(P_1^{(n)} - P_0^{(n)} \right). \tag{3}$$

Here $P_1^{(n)} - P_0^{(n)}$ represents the difference between P_1 and P_0 for the same underlying stochastic sample, so that $P_1^{(n)} - P_0^{(n)}$ is small and has a small variance; the precise construction depends on the application and various examples will be shown later. The two key differences from the control variate approach are that the value of $\mathbb{E}[P_0]$ is not known, so has to be estimated, and we use $\lambda = 1$.

7.10 Two-level MC II

If we define C_0 and C_1 to be the cost of computing a single sample of P_0 and $P_1 - P_0$, respectively, then the total cost is $N_0\, C_0 + N_1\, C_1$, and if V_0 and V_1 are the variance of P_0 and $P_1 - P_0$, then the overall variance is $N_0^{-1} V_0 + N_1^{-1} V_1$, assuming that $\sum_{n=1}^{N_0} P_0^{(n)}$ and $\sum_{n=1}^{N_1} \left(P_1^{(n)} - P_0^{(n)} \right)$ use independent samples.

Hence, treating the integers N_0, N_1 as real variables and performing a constrained minimisation using a Lagrange multiplier, the variance is minimised for a fixed cost by choosing $N_1 / N_0 = \sqrt{V_1 / C_1} \, / \sqrt{V_0 / C_0}$. For details see the script.

7.11 The MLMC I

Greater accuracy requires larger N and smaller weak error $\mathbb{E}(\hat{P}) - \mathbb{E}(P)$.
Given a sequence $\hat{P}_0, \hat{P}_1 \dots \hat{P}_L$

$$\mathbb{E}[\hat{P}_L] = \mathbb{E}[\hat{P}_0] + \sum_{l=1}^{L} \mathbb{E}[\hat{P}_l - \hat{P}_{l-1}]$$

and we can use again the estimator

$$N_0^{-1} \sum_{n=1}^{N_0} \hat{P}_0^n + \sum_{l=1}^{L} \left[N_l^{-1} \sum_{n=1}^{N_l} \hat{P}_l^n - \hat{P}_{l-1}^n \right]$$

If we define C_l the cost and V_l the variance of individual levels i.e.

$$C_0, V_0 \to \hat{P}_0, \text{ and } C_l, V_l \to (\hat{P}_l - \hat{P}_{l-1})$$

7.11 The MLMC II

then the total cost and variance becomes

$$C = \sum_{l=1}^{L} C_l N_l \quad \text{and} \quad \mathbb{V}ar = \sum_{l=1}^{L} V_l N_l^{-1}$$

Next we need to find the minimal cost for a given fixed variance. We use the Lagrange multiplyer to solve the constrained optimization problem.

$$\frac{d}{dN_l} = \sum_{l=1}^{L} \left(N_l C_l + \mu^2 N_l^{-1} V_l \right) = 0$$

$$\Rightarrow \quad N_l = \mu \sqrt{V_l / C_l}, \quad N_l C_l = \mu \sqrt{V_l C_l}$$

This is problem dependent not trivial to analyze. Some details can be found in the script.