MLMC for Computing Probabilities

Abdul-Lateef Haji-Ali

Heriot-Watt University

Dundee — October 25, 2021

The problem: Computing probabilities

$$\mathbb{P}[\, Z \in \Omega \,] = \mathbb{E}[\, \mathbb{I}_{Z \in \Omega} \,]$$

where Z is a d-dimensional random variable and $\Omega \in \mathbb{R}^d$.

The problem: Computing probabilities

$$\mathbb{P}[\, Z \in \Omega \,] = \mathbb{E}[\, \mathbb{I}_{Z \in \Omega} \,]$$

where Z is a d-dimensional random variable and $\Omega \in \mathbb{R}^d$.

Two main reasons this problem can be challenging:

1. The event is rare,
2. and the complexity of sampling Z.

The problem: Computing probabilities

- Digital options $Z := S_1 - K$ where S_1 is an asset price at time 1 which satisfies an SDE and K is the strike price

$$\mathbb{E}[\, \mathbb{I}_{S_1 > K} \,]$$

- Financial risk assessment $Z := \mathbb{E}[\, Y \,|\, R \,] - \mathrm{MaxLoss}$

$$\mathbb{E}[\, \mathbb{I}_{\mathbb{E}[Y|R] > \mathrm{MaxLoss}} \,]$$

- Component failure: $Z := g(Y)$ where g depends on the solution of a PDE with random coefficients Y.

$$\mathbb{E}[\, \mathbb{I}_{g(Y)} \,]$$

The problem: Computing probabilities

- Digital options $Z := S_1 - K$ where S_1 is an asset price at time 1 which satisfies an SDE and K is the strike price

$$\mathbb{E}[\,\mathbb{I}_{S_1 > K}\,] \approx \mathbb{E}[\,\mathbb{I}_{\overline{S}_1^h > K}\,]$$

where $\overline{S}_\cdot^h$ is an Euler-Maruyama or Milstein approximations with step size h.

- Financial risk assessment $Z := \mathbb{E}[\,Y \,|\, R\,] - \mathrm{MaxLoss}$

$$\mathbb{E}[\,\mathbb{I}_{\mathbb{E}[Y|R] > \mathrm{MaxLoss}}\,] \approx \mathbb{E}\left[\,\mathbb{I}_{\frac{1}{N}\sum_{i=1}^N Y^{(i)}(R) > \mathrm{MaxLoss}}\,\right]$$

- Component failure: $Z := g(Y)$ where g depends on the solution of a PDE with random coefficients Y.

$$\mathbb{E}[\,\mathbb{I}_{g(Y)}\,] \approx \mathbb{E}[\,\mathbb{I}_{g^h(Y)}\,]$$

where g^h is a Finite Element approximation with grid size h.

The problem: Computing probabilities in 1D

The problem can be written in the form

$$\mathbb{P}[X > 0] = \mathbb{E}[\mathbb{I}_{X>0}]$$

for a one-dimensional random variable X which is the signed distance of Z to Ω.

Monte Carlo: A General Framework

Focus on

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

for some function f. For our setup, $f(X) := \mathbb{I}_{X>0}$. Assume we can approximate $X \approx X_\ell$ with $\ell \in \mathbb{N}$

Assumptions

- Work of X_ℓ is $\propto 2^{\gamma\ell}$ for $\gamma > 0$.
- Bias: $E_\ell := |\mathbb{E}[\, f(X_\ell) - f(X)\,]| \propto 2^{-\alpha\ell}$ for $\alpha > 0$.

When X depends on many random parameters, best option is to use Monte Carlo

$$\mathbb{E}[\, f(X)\,] \approx \frac{1}{M} \sum_{m=1}^{M} f(X_L^{(m)})$$

To approximate $\mathbb{E}[\, f(X)\,]$ with an error tolerance ε, need $M = \mathcal{O}(\varepsilon^{-2})$ and $L = \mathcal{O}(\frac{1}{\alpha}|\log \varepsilon|)$ hence complexity is $\mathcal{O}(\varepsilon^{-2-\gamma/\alpha})$.

Multilevel Monte Carlo: A General Framework

The MLMC estimator is based on

$$
\begin{aligned}
\mathbb{E}[\, f(X)\,] &= \mathbb{E}[\, f(X_0)\,] && + \sum_{\ell=1}^{\infty} \mathbb{E}[\, f(X_\ell) - f(X_{\ell-1})\,] \\
&\approx \mathbb{E}[\, f(X_0)\,] && + \sum_{\ell=1}^{L} \mathbb{E}[\, f(X_\ell) - f(X_{\ell-1})\,] \\
&\approx \frac{1}{M_0} \sum_{m=1}^{M_0} f(X_0^{0,m}) && + \sum_{\ell=1}^{L} \frac{1}{M_\ell} \sum_{m=1}^{M_\ell} f(X_\ell^{\ell,m}) - f(X_{\ell-1}^{\ell,m})
\end{aligned}
$$

Multilevel Monte Carlo: A General Framework

Assumptions

- Work of X_ℓ is $W_\ell \propto 2^{\gamma\ell}$ for $\gamma > 0$.
- Bias: $|\mathbb{E}[\, f(X_\ell) - f(X)\,]| \propto 2^{-\alpha\ell}$ for $\alpha > 0$.
- Strong Error: $\mathbb{E}[\,(X_\ell - X_{\ell-1})^2\,] \propto 2^{-\beta\ell}$ for $\beta > 0$.

Theorem

For Lipschitz f, the overall cost of Multilevel Monte Carlo for computing $\mathbb{E}[\, f(X)\,]$ to accuracy ε using optimal $L, \{M_\ell\}_{\ell=0}^{L}$ is

$$\begin{cases} \varepsilon^{-2} & \beta > \gamma \\ \varepsilon^{-2}(\log \varepsilon)^2 & \beta = \gamma \\ \varepsilon^{-2-\frac{\gamma-\beta}{\alpha}} & \beta < \gamma \end{cases}$$

Proof. $\mathrm{Var}[\,(f(X_\ell) - f(X_{\ell-1}))^2\,] \le C^2\, \mathbb{E}[\,(X_\ell - X_{\ell-1})^2\,] \propto 2^{-\beta\ell}$ $\square$

Multilevel Monte Carlo: A General Framework

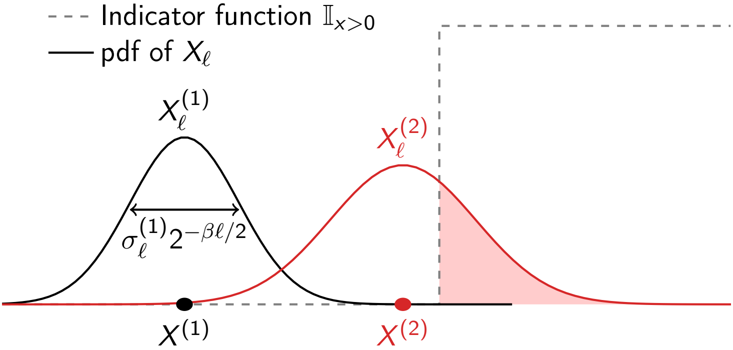
Example

For a standard European call option we have $\mathbb{E}[\, f(X)\,]$ for $X = S_1 - K$ and $f(X) = \max(X, 0)$. Approximating S_1 by Euler-Maruyama with 2^ℓ time-steps at level ℓ satisfies the previous assumptions with $\alpha = \beta = \gamma = 1$. The complexity is

- $\mathcal{O}\left(\varepsilon^{-3}\right)$ for Monte Carlo.
- $\mathcal{O}\left(\varepsilon^{-2}(\log \varepsilon)^2\right)$ using Multilevel Monte Carlo.

Discontinuous f: Key assumptions

Our quantity of interest is $f(x) = \mathbb{I}_{x>0}$ is discontinuous, need a different kind of analysis.



Discontinuous f: Key assumptions

Our quantity of interest is $f(x) = \mathbb{I}_{x>0}$ is discontinuous, need a different kind of analysis.

Assumptions

For all $\ell \in \mathbb{N}$ define

$$\delta_\ell := \frac{|X_\ell|}{\sigma_\ell} \geq 0,$$

for some random variable $\sigma_\ell > 0$. For all ℓ:

1. There is $\delta > 0$ such that for $x \leq \delta$ we have $\mathbb{P}[\,\delta_\ell \leq x\,] \lesssim x$.
2. There is $q > 2$ such that

$$\left(\mathbb{E}\left[\left(\frac{|X_\ell - X|}{\sigma_\ell}\right)^q\right]\right)^{1/q} \lesssim 2^{-\beta\ell/2}.$$

MLMC analysis

Lemma

$$\mathrm{Var}[\,\mathbb{I}_{X>0} - \mathbb{I}_{X_\ell>0}\,] \lesssim 2^{-\frac{q}{q+1}\ell\beta/2}$$

Proof. $|X - X_\ell| \approx \mathcal{O}(2^{-\ell\beta/2})$ □

Corollary

Computing $\mathbb{E}[\,\mathbb{I}_{X>0}\,]$ to accuracy ε using Multilevel Monte Carlo has cost:

$$\begin{cases} \varepsilon^{-2} & \beta > \frac{q+1}{q}\cdot 2\gamma \\ \varepsilon^{-2}(\log\varepsilon)^2 & \beta = \frac{q+1}{q}\cdot 2\gamma \\ \varepsilon^{-2-\left(\gamma-\frac{q}{q+1}\beta/2\right)/\alpha} & \beta < \frac{q+1}{q}\cdot 2\gamma \end{cases}$$

Another problem

- Note that for any even p

$$\mathbb{E}[\,(\mathbb{I}_{X_\ell>0} - \mathbb{I}_{X_{\ell-1}>0})^p\,] = \mathbb{E}[\,(\mathbb{I}_{X_\ell>0} - \mathbb{I}_{X_{\ell-1}>0})^2\,] = \mathcal{O}(2^{-\frac{q}{q+1}\beta\ell/2})$$

- Hence, the Kurtosis is $2^{\beta\ell/2}$.
- This means that estimating the variance becomes increasingly difficult as ℓ increases.
- This is caused by the indicator function.

Previous research

- M. B. Giles, D. J. Higham, and X. Mao. "Analysing multi-level Monte Carlo for options with non-globally Lipschitz payoff". In: *Finance and Stochastics* 13.3 (2009), pp. 403–413

  Original analysis of classical MLMC for discontinuous payoffs in SDE example.

- M. B. Giles, T. Nagapetyan, and K. Ritter. "Multilevel Monte Carlo approximation of distribution functions and densities". In: *SIAM/ASA journal on Uncertainty Quantification* 3.1 (2015), pp. 267–295

  Deals with similar problems in the generality of the current work. Uses different method based on smoothing the discontinuity. Assumes differentiability of PDF and requires further analysis to determine effect of smoothing parameter on bias/variance.

- C. Bayer, C. B. Hammouda, and R. Tempone. "Numerical smoothing and hierarchical approximations for efficient option pricing and density estimation". In: *arXiv preprint arXiv:2003.05708* (2020)

  Same as above. Smoothes the discontinuity by intergrating using a high order method with respect to one of the dimensions.

Previous research (adaptivity)

- D. Elfverson, F. Hellman, and A. Målqvist. "A multilevel Monte Carlo method for computing failure probabilities". In: *SIAM/ASA Journal on Uncertainty Quantification* 4.1 (2016), pp. 312–330

  Selective refinement of samples. Based on relaxing the condition. Assumes uniform almost sure error bounds (works well for PDEs with random coefficients but not stochastic models).

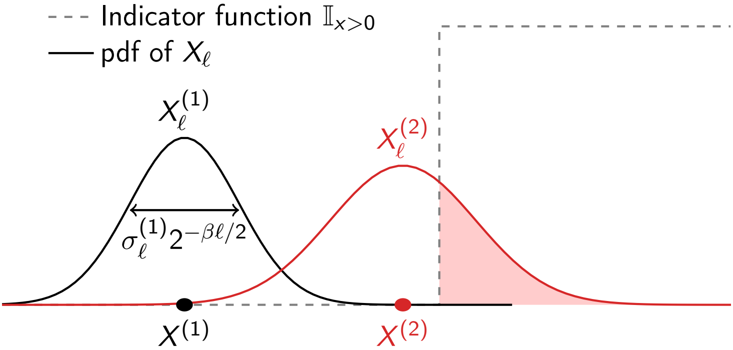
- M. Broadie, Y. Du, and C. C. Moallemi. "Efficient risk estimation via nested sequential simulation". In: *Management Science* 57.6 (2011), pp. 1172–1194

  Adaptive sampling for nested expectation with Monte Carlo methods.

- M. B. Giles and A.-L. Haji-Ali. "Multilevel nested simulation for efficient risk estimation". In: *SIAM/ASA Journal on Uncertainty Quantification* 7.2 (2019), pp. 497–525. DOI: 10.1137/18M1173186

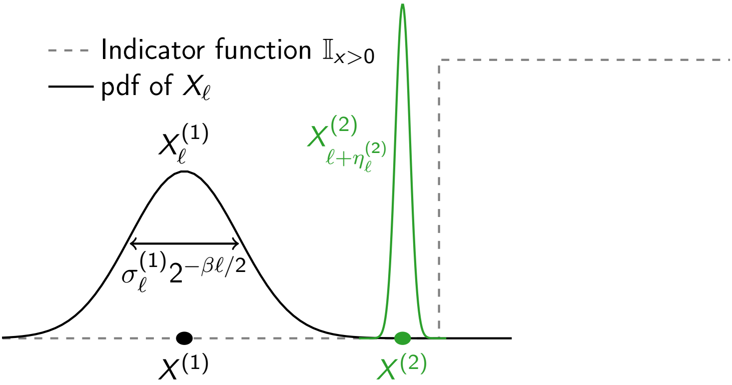
  Adaptive sampling for MLMC applied to nested expectations only. Requires stronger conditions on the random variables than here.

Adaptive refinement



Adaptive refinement

Refine samples of X_ℓ to $X_{\ell+\eta_\ell}$, where η_ℓ is an integer.



Adaptive refinement: Algorithm

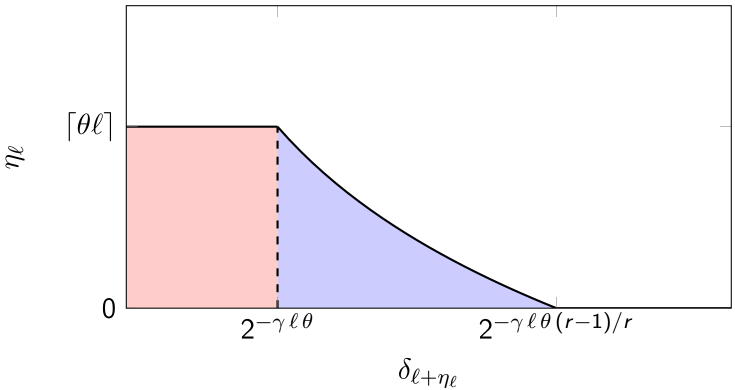
Recall that $\delta_\ell := \dfrac{|X_\ell|}{\sigma_\ell} \geq 0$.

Algorithm 1: Adaptive sampling at level ℓ

Input: $\ell, r, \theta, c > 0, \gamma, \beta$
Result: Adaptively refined sample $X_{\ell+\eta_\ell}$
Set $\eta_\ell = 0$;
Sample (X_ℓ, σ_ℓ);
Compute δ_ℓ given (X_ℓ, σ_ℓ);
while $|\delta_{\ell+\eta_\ell}| < c 2^{\gamma(\theta\ell(1-r)-\eta_\ell)/r}$ *and* $\eta_\ell < \lceil \theta\ell \rceil$ **do**
    Refine $(X_{\ell+\eta_\ell}, \sigma_{\ell+\eta_\ell})$ to $(X_{\ell+\eta_\ell+1}, \sigma_{\ell+\eta_\ell+1})$;
    Compute $\delta_{\ell+\eta_\ell+1}$ given $(X_{\ell+\eta_\ell+1}, \sigma_{\ell+\eta_\ell+1})$;
    Set $\eta_\ell = \eta_\ell + 1$;
end
Output: $X_{\ell+\eta_\ell}$

Adaptive refinement: Algorithm

Recall that $\delta_\ell := \dfrac{|X_\ell|}{\sigma_\ell} \geq 0$.



Adaptive Multilevel Monte Carlo: Analysis

Theorem

There is $\bar{r} > 1$ such that for $1 < r < \bar{r}$:

- The expected work of sampling $\mathbb{I}_{X_{\ell+\eta_\ell}>0}$ is $W_\ell \propto 2^{\gamma\ell}$.
- The variance is

$$\mathrm{Var}[\,\mathbb{I}_{X_\ell>0} - \mathbb{I}_{X_{\ell+\eta_\ell}>0}\,] \propto 2^{-\frac{q}{q+1}\frac{1+\theta}{2}\beta\ell}$$

for

$$\theta = \begin{cases} \dfrac{1}{2^{\frac{q+1}{q}\frac{\gamma}{\beta}-1}} & \beta < \dfrac{q+1}{q}\gamma \\ 1 & \beta > \dfrac{q+1}{q}\gamma \end{cases}.$$

Adaptive Multilevel Monte Carlo: Complexity

Corollary

Computing $\mathbb{E}[\,\mathbb{I}_{X>0}\,]$ to accuracy ε using (non-)adaptive Multilevel Monte Carlo has cost:

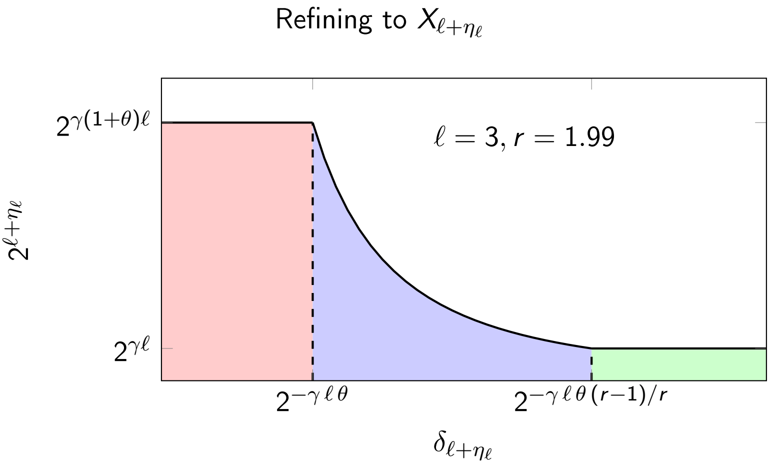
Non-Adaptive:

$$\begin{cases} \varepsilon^{-2} & \beta > \frac{q+1}{q} \cdot 2\gamma \\ \varepsilon^{-2}(\log \varepsilon)^2 & \beta = \frac{q+1}{q} \cdot 2\gamma \\ \varepsilon^{-2-\left(\gamma - \frac{q}{q+1}\beta/2\right)/\alpha} & \beta < \frac{q+1}{q} \cdot 2\gamma \end{cases}$$

Adaptive:

$$\begin{cases} \varepsilon^{-2} & \beta > \frac{q+1}{q} \cdot \gamma \\ \varepsilon^{-2}(\log \varepsilon)^2 & \beta = \frac{q+1}{q} \cdot \gamma \\ \varepsilon^{-2-\left(\gamma - \left(\frac{1}{2\frac{q+1}{q\beta} - \frac{1}{\gamma}}\right)\right)/\text{``}\alpha\text{''}} & \beta < \frac{q+1}{q} \cdot \gamma \end{cases}$$

Work/variance proof idea



Adaptive Multilevel Monte Carlo: Complexity

Example

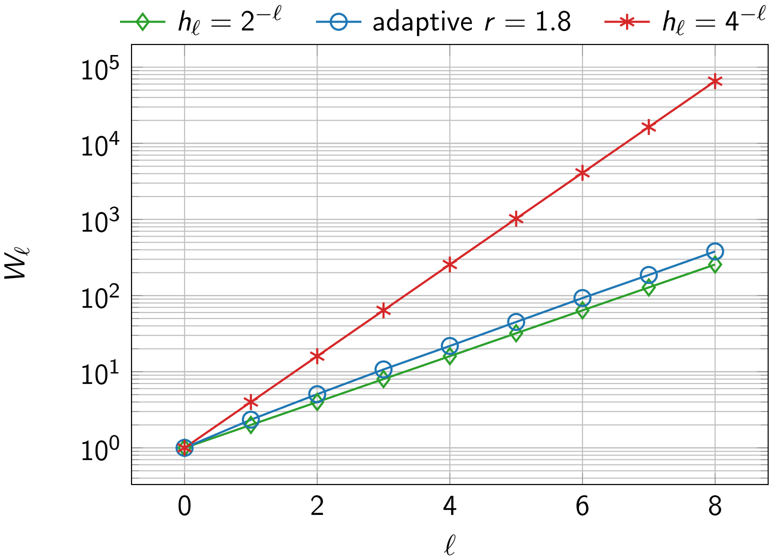
When approximating the price of a digital option

$$\mathbb{E}\left[\, \mathbb{I}_{S_1 > K} \,\right],$$

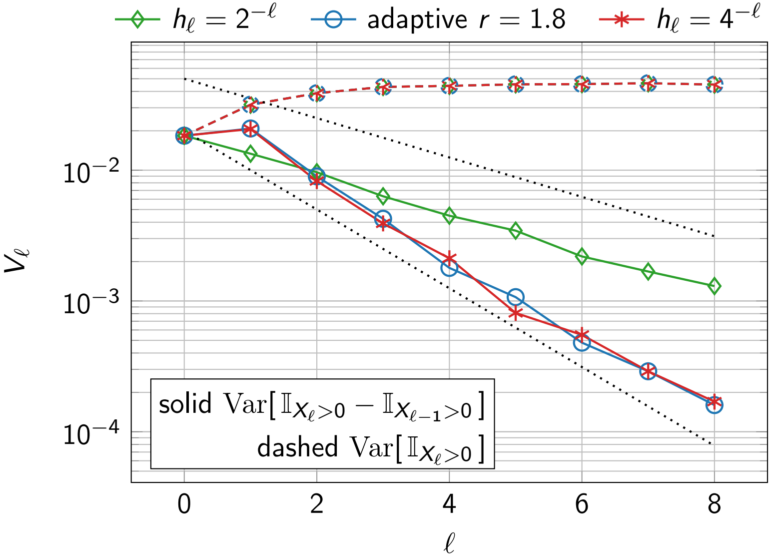
using Euler-Maruyama approximation of S_1 with 2^ℓ time-steps at level ℓ the assumptions hold for $\alpha = \beta = \gamma = 1$ and any $q < \infty$. The complexity is (for any $\nu > 0$)

- $\mathcal{O}\left(\varepsilon^{-3}\right)$ for Monte Carlo.
- $o\left(\varepsilon^{-2.5-\nu}\right)$ for non-adaptive Multilevel Monte Carlo.
- $o\left(\varepsilon^{-2-\nu}\right)$ for adaptive Multilevel Monte Carlo.

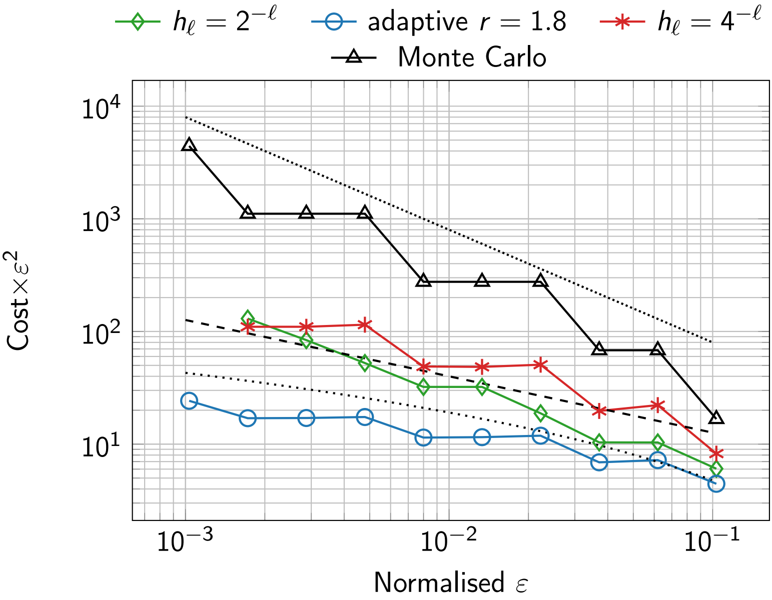
Numerical Test: Digital Options



Numerical Test: Digital Options



Numerical Test: Digital Options



Not included in this presentation

In https://arxiv.org/abs/2107.09148

- Nested expectation: Differs from previous work in that the same samples are used for computing the refinement, η_ℓ and for computing the estimate $X_{\ell+\eta_\ell}$. Leads to reduced cost and more relaxed assumptions.
- Discussion on choices of σ_ℓ in nested expectation.
- Motivation of previous assumptions in the case of nested expectation and SDEs.
- Analysis of weak error bounds and corresponding necessary assumptions.

Summary of Adaptive MLMC

- Accurate computation of probabilities by standard Monte Carlo techniques is expensive when the underlying observable must be approximated for each sample.
- Multilevel Monte Carlo is a great method to reduce this cost, but suffers for probabilities due to the intrinsic discontinuity.
- Adaptive sampling provides a general framework to improve Multilevel Monte Carlo performance for probabilities, in many cases to optimal $\mathcal{O}\left(\varepsilon^{-2}\right)$ cost.
- Other applications: Barrier options and computing sensitives.
- The adaptive method still suffers from increasing Kurtosis as ℓ increases. Might still need to smooth the discontinuity.

Multilevel Monte Carlo and Path Branching for Digital Options (briefly)

Abdul-Lateef Haji-Ali

Heriot-Watt University

Dundee — October 25, 2021

The problem: Pricing a Digital option

Let S_t be a d-dimensional stochastic process satisfying the SDE for $0 < t \leq 1$

$$\mathrm{d}S_t = a(S_t, t)\,\mathrm{d}t + \sigma(S_t, t)\,\mathrm{d}W_t.$$

Let $(\mathcal{F}_t)_{0 \leq t \leq 1}$ be the natural filtration of W_t.

The problem: Pricing a Digital option

Let S_t be a d-dimensional stochastic process satisfying the SDE for $0 < t \leq 1$

$$\mathrm{d}S_t = a(S_t, t)\,\mathrm{d}t + \sigma(S_t, t)\,\mathrm{d}W_t.$$

Let $(\mathcal{F}_t)_{0\leq t\leq 1}$ be the natural filtration of W_t.

We want to price a digital option of the form (dropping discounting)

$$\mathbb{P}[\, S_1 \in K \,] = \mathbb{E}[\, \mathbb{I}_{S_1 \in K} \,]$$

for some $K \subset \mathbb{R}^d$. Let $\{\overline{S}_t^\ell\}_{t=0}^1$ be an approximation of the path $\{S_t\}_{t=0}^1$ at level ℓ using $h_\ell^{-1} \equiv h_0^{-1} 2^\ell$ timesteps.

The problem: Pricing a Digital option

Let S_t be a d-dimensional stochastic process satisfying the SDE for $0 < t \leq 1$

$$\mathrm{d}S_t = a(S_t, t)\,\mathrm{d}t + \sigma(S_t, t)\,\mathrm{d}W_t.$$

Let $(\mathcal{F}_t)_{0 \leq t \leq 1}$ be the natural filtration of W_t.

We want to price a digital option of the form (dropping discounting)

$$\mathbb{P}[\, S_1 \in K \,] = \mathbb{E}[\, \mathbb{I}_{S_1 \in K} \,]$$

for some $K \subset \mathbb{R}^d$. Let $\{\overline{S}_t^{\ell}\}_{t=0}^{1}$ be an approximation of the path $\{S_t\}_{t=0}^{1}$ at level ℓ using $h_\ell^{-1} \equiv h_0^{-1} 2^{\ell}$ timesteps.

Assuming

$$|\mathbb{E}[\, \mathbb{I}_{S_1 \in K} - \mathbb{I}_{\overline{S}_1^{\ell} \in K} \,]| \lesssim h_\ell^{\alpha}, \qquad \mathrm{Work}(\overline{S}_1^{\ell}) \lesssim h_\ell^{-1}$$

a Monte Carlo estimator of $\mathbb{E}[\, \mathbb{I}_{S_1 \in K} \,]$ has computational complexity $\varepsilon^{-2-\alpha}$ to achieve root MSE ε.

Multilevel Monte Carlo

Consider a hierarchy of corrections $\{\Delta P_\ell\}_{\ell=0}^{L}$ such that

$$\mathbb{E}[\, \Delta P_\ell \,] = \begin{cases} \mathbb{E}\left[\, \mathbb{I}_{\overline{S}_1^0 \in K} \,\right] & \ell = 0 \\ \mathbb{E}\left[\, \mathbb{I}_{\overline{S}_1^\ell \in K} - \mathbb{I}_{\overline{S}_1^{\ell-1} \in K} \,\right] & \text{otherwise.} \end{cases}$$

MLMC can be formulated as

$$\mathbb{E}\left[\, \mathbb{I}_{S_1 \in K} \,\right] = \sum_{\ell=0}^{\infty} \mathbb{E}[\, \Delta P_\ell \,] \approx \sum_{\ell=0}^{L} \frac{1}{M_\ell} \sum_{m=1}^{M_\ell} \Delta P_\ell^{(m)}$$

Multilevel Monte Carlo

Consider a hierarchy of corrections $\{\Delta P_\ell\}_{\ell=0}^{L}$ such that

$$\mathbb{E}[\,\Delta P_\ell\,] = \begin{cases} \mathbb{E}\left[\,\mathbb{I}_{\overline{S}_1^0 \in K}\,\right] & \ell = 0 \\ \mathbb{E}\left[\,\mathbb{I}_{\overline{S}_1^\ell \in K} - \mathbb{I}_{\overline{S}_1^{\ell-1} \in K}\,\right] & \text{otherwise.} \end{cases}$$

MLMC can be formulated as

$$\mathbb{E}\left[\,\mathbb{I}_{S_1 \in K}\,\right] = \sum_{\ell=0}^{\infty} \mathbb{E}[\,\Delta P_\ell\,] \approx \sum_{\ell=0}^{L} \frac{1}{M_\ell} \sum_{m=1}^{M_\ell} \Delta P_\ell^{(m)}$$

Assuming

$$\mathrm{Var}[\,\Delta P_\ell\,] \lessapprox h_\ell^{\beta/2}, \qquad |\mathbb{E}[\,\Delta P_\ell\,]| \lesssim h_\ell^{\alpha}, \qquad \mathrm{Work}(\Delta P_\ell) \lesssim h_\ell^{-1}$$

then to compute with root MSE ε the complexity of MLMC is $\mathcal{O}(\varepsilon^{-2+\max((\beta/2-1),0)/\alpha})$ when $\beta/2 \neq 1$ and $\mathcal{O}(\varepsilon^{-2}|\log \varepsilon|^2)$ otherwise.
E.g. Euler-Maruyama has $\alpha = \beta = 1$ and complexity is $\mathcal{O}(\varepsilon^{-5/2})$.

Conditional Expectation and Path Splitting

For some $0 < \tau < 1$, let

$$\Delta Q_\ell := \mathbb{E}[\, \Delta P_\ell \,|\, \mathcal{F}_{1-\tau} \,].$$

$$\text{Note} \qquad \mathbb{E}[\, \Delta Q_\ell \,] = \mathbb{E}[\, \Delta P_\ell \,].$$

Conditional Expectation and Path Splitting

For some $0 < \tau < 1$, let

$$\Delta Q_\ell := \mathbb{E}[\, \Delta P_\ell \,|\, \mathcal{F}_{1-\tau} \,].$$

$$\text{Note} \qquad \mathbb{E}[\, \Delta Q_\ell \,] = \mathbb{E}[\, \Delta P_\ell \,].$$

We can consider the MLMC estimator based on ΔQ_ℓ instead of ΔP_ℓ. The work and (hopefully improved) variance convergence of ΔQ_ℓ becomes relevant.

Conditional Expectation and Path Splitting

For some $0 < \tau < 1$, let

$$\Delta Q_\ell := \mathbb{E}[\, \Delta P_\ell \,|\, \mathcal{F}_{1-\tau} \,].$$

$$\text{Note} \qquad \mathbb{E}[\, \Delta Q_\ell \,] = \mathbb{E}[\, \Delta P_\ell \,].$$

We can consider the MLMC estimator based on ΔQ_ℓ instead of ΔP_ℓ. The work and (hopefully improved) variance convergence of ΔQ_ℓ becomes relevant.

Computing ΔQ_ℓ:

- In 1D, taking $\tau \equiv h_\ell$ and using Euler-Maruyama for the last step we know that the conditional distribution of ΔP_ℓ given $\mathcal{F}_{1-\tau}$ is Gaussian and we can compute ΔQ_ℓ exactly.
- More generally, for any method and any τ, we can use path splitting (Monte Carlo) with sufficient number of samples, leading to increased work.

See, e.g., Glasserman (2004) and Burgos & Giles (2012) for more information on this method (for computing options and sensitivities).

Path splitting to estimate ΔQ_ℓ

- When $\tau \to 0$, i.e., splitting late,

$$\mathrm{Var}[\, \Delta Q_\ell \,] \le \mathbb{E}\Big[(\mathbb{E}[\, \Delta P_\ell \,|\, \mathcal{F}_{1-\tau} \,])^2 \Big] = \mathbb{E}\Big[(\Delta P_\ell)^2 \Big] = \mathcal{O}(h_\ell^{\beta/2})$$

- When $\tau \to 1$, i.e., splitting early,

$$\mathrm{Var}[\, \Delta Q_\ell \,] \le \mathbb{E}\Big[(\mathbb{E}[\, \Delta P_\ell \,|\, \mathcal{F}_{1-\tau} \,])^2 \Big] \le (\mathbb{E}[\, |\Delta P_\ell| \,])^2 = \mathcal{O}(h_\ell^{\beta})$$

Hence we want to take $\tau \to 1$, but the cost per inner sample increases; paths are approximated over $[1-\tau, 1]$ for every sample.

Path splitting to estimate ΔQ_ℓ

- When $\tau \to 0$, i.e., splitting late,

$$\mathrm{Var}[\,\Delta Q_\ell\,] \le \mathbb{E}\Big[\,(\mathbb{E}[\,\Delta P_\ell \,|\, \mathcal{F}_{1-\tau}\,])^2\,\Big] = \mathbb{E}\Big[\,(\Delta P_\ell)^2\,\Big] = \mathcal{O}(h_\ell^{\beta/2})$$

- When $\tau \to 1$, i.e., splitting early,

$$\mathrm{Var}[\,\Delta Q_\ell\,] \le \mathbb{E}\Big[\,(\mathbb{E}[\,\Delta P_\ell \,|\, \mathcal{F}_{1-\tau}\,])^2\,\Big] \le (\mathbb{E}[\,|\Delta P_\ell|\,])^2 = \mathcal{O}(h_\ell^{\beta})$$

Hence we want to take $\tau \to 1$, but the cost per inner sample increases; paths are approximated over $[1-\tau, 1]$ for every sample.

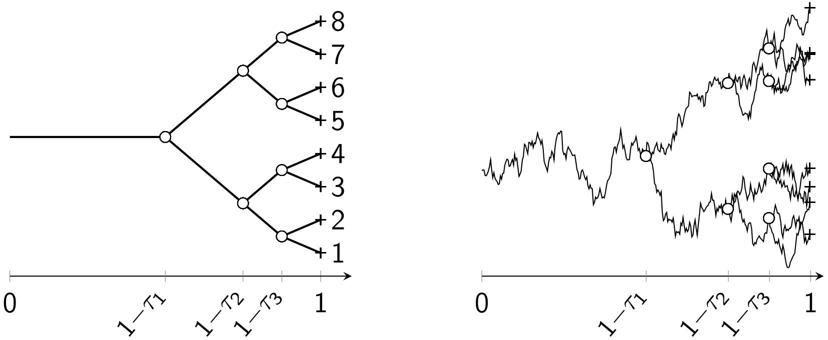
Solution: More splitting! For $\tau' > \tau$

$$\begin{aligned}\Delta Q'_\ell &= \mathbb{E}[\,\Delta Q_\ell \,|\, \mathcal{F}_{1-\tau'}\,] \\ &= \mathbb{E}[\,\mathbb{E}[\,\Delta P_\ell \,|\, \mathcal{F}_{1-\tau}\,] \,|\, \mathcal{F}_{1-\tau'}\,]\end{aligned}$$

Again $\qquad \mathbb{E}[\,\Delta Q'_\ell\,] = \mathbb{E}[\,\Delta P\,]$

Path Branching

- Let $1 - \tau_{\ell'} = 1 - 2^{-\ell'}$ for $\ell' \in \{1, \ldots, \ell\}$.
- For every ℓ', starting from $S_{1-\tau_{\ell'}}$ at time $1 - \tau_{\ell'}$, create two sample paths $\{S_t\}_{1-\tau_{\ell'} \le t \le 1-\tau_{\ell'+1}}$ which depend on two independent samples of the Brownian motion $\{W_t\}_{1-\tau_{\ell'} \le t \le 1-\tau_{\ell'+1}}$.
- Evaluate the payoff difference $\Delta P_\ell^{(i)}$ for every $S_1^{(i)}$ for $i \in \{1, \ldots, 2^\ell\}$
- Define the Monte Carlo average as $\Delta \mathcal{P}_\ell := 2^{-\ell} \sum_{i=1}^{2^\ell} \Delta P_\ell^{(i)}$



MLMC Complexity

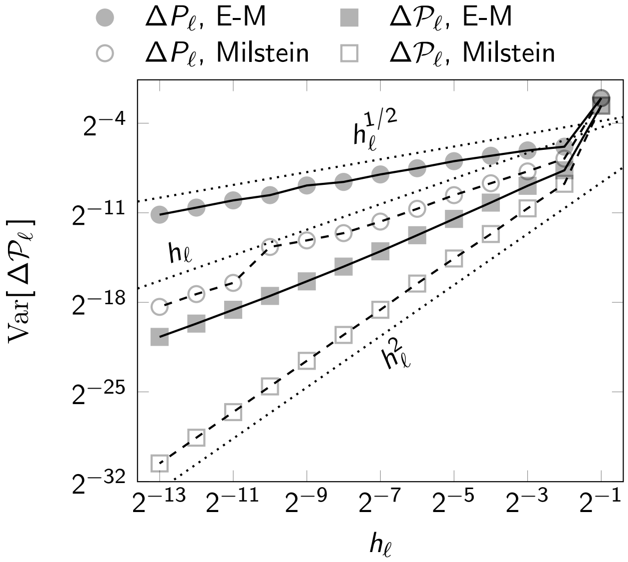
$$\text{Work}(\Delta\mathcal{P}_\ell) = \mathcal{O}(\ell h_\ell^{-1})$$

Assuming that the transition kernel is Gaussian or log-Gaussian and K is not pathological, then

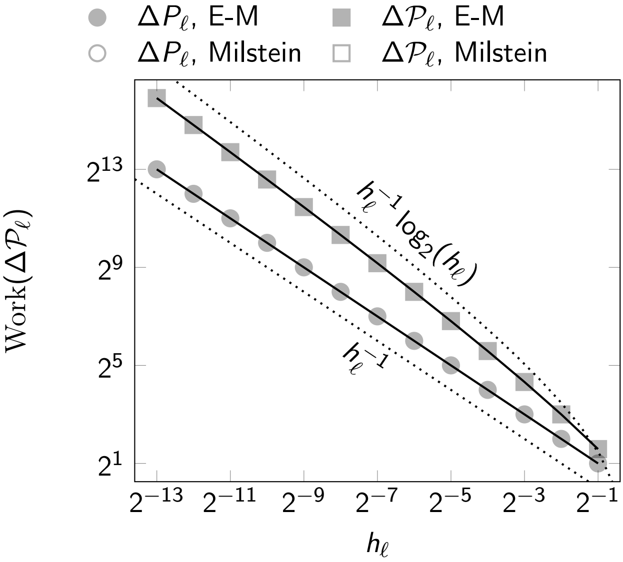
$$\mathrm{Var}[\,\Delta\mathcal{P}_\ell\,] \le \mathbb{E}\left[\left(2^{-\ell}\sum_{i=1}^{2^\ell}\Delta P_\ell^{(i)}\right)^2\right] \approx \mathcal{O}(h_\ell^{\min(1+\beta/2,\beta)})$$

- Using Euler-Maryama: $\beta = 1$ and the MLMC computational complexity is approximately $o\left(\varepsilon^{-2+\nu}\right)$ for any $\nu > 0$ and for MSE ε.
- Using Milstein: $\beta = 2$ and the complexity is $\mathcal{O}(\varepsilon^{-2})$.

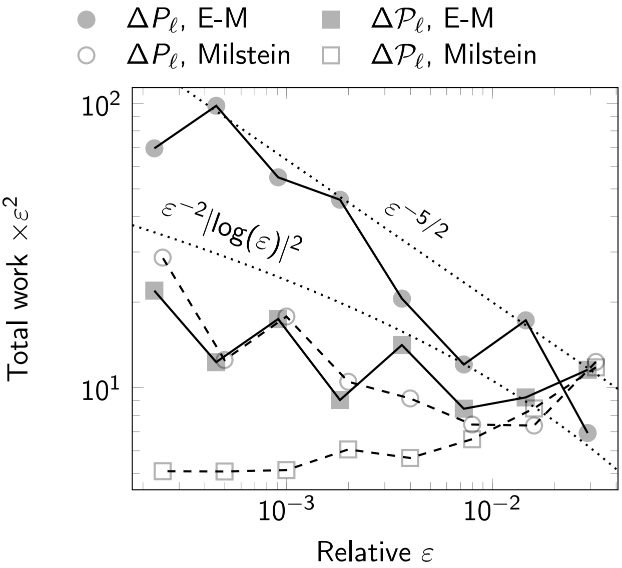
Numerical Results on GBM



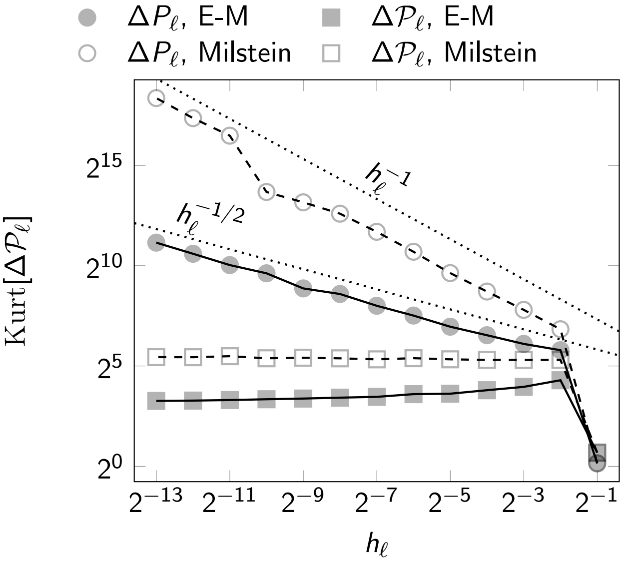
Numerical Results on GBM



Numerical Results on GBM



Numerical Results on GBM



Not included in this presentation

- We also consider a sequence $\tau_{\ell'} = 2^{-\eta \ell'}$ for some $\eta > 0$. For $\eta > 1$, this reduces the work of $\Delta \mathcal{P}_\ell$ to $\mathcal{O}(2^\ell)$.
- Rigorous analysis to show the previous rates.
- Application and analysis of antithetic estimators.
- Other applications to consider:
  - Pricing other options (Barrier).
  - Computing sensitivities: Using bumping, the variance increases as the bump distance decreases. Branching can help.
  - Particle systems.
  - Approximate CDFs. Need to tighten theory to deal with increasing number of discontinuities.
  - Parabolic SPDEs.

Final thoughts

- Adaptive MLMC can be applied to compute probabilities involving any models and their approximations, but it suffers from high Kurtosis.
- Path Branching, like conditional splitting, needs to be adapted/analysed for specific problems, but has better Kurtosis.
- These methods are well suited for non-rare events.
- MLMC can be combined with other rare event methods (like Sequential Monte Carlo, Subset simulation or adaptive importance sampling).

Q&A

Thank you!