Adaptive MLMC for Computing Probabilities

Abdul-Lateef Haji-Ali

Jonathan Spence (HWU)
Aretha Teckentrup (University of Edinburgh)

Heriot-Watt University

BAMC — Decision Making Under Uncertainty
April 11, 2022

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 $\Omega \in \mathbb{R}^d$. This 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.

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 $\Omega \in \mathbb{R}^d$. This 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.

Two main reasons this problem can be challenging:

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

The problem: Computing probabilities

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

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

- Digital options $X := S(T) - K$ where S is an asset price satisfying an SDE and K is the strike price

$$\mathbb{E}[\, \mathbb{I}_{S(T) > K} \,]$$

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

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

The problem: Computing probabilities

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

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

- Digital options $X := S(T) - K$ where S is an asset price satisfying an SDE and K is the strike price

$$\mathbb{E}[\, \mathbb{I}_{S(T)>K}\,] \approx \mathbb{E}[\, \mathbb{I}_{S_h(T)>K}\,]$$

  where S_h is an Euler-Maruyama or Milstein approximations with step size h.

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

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

  where g_h is a Finite Element approximation with grid size h.

Monte Carlo: A General Framework

Focus on

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

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

Assumptions

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

When the dimensionality of X is high, best option is to use Monte Carlo

$$\mathbb{E}[\,\mathbb{I}_{X>0}\,] \approx \frac{1}{M}\sum_{m=1}^{M} \mathbb{I}_{X_L^{(m)}>0}$$

To approximate $\mathbb{P}[\, X > 0\,]$ 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 $W_\ell \propto 2^{\gamma\ell}$.
- Bias: $|\mathbb{E}[\, f(X_\ell) - f(X)\,]| \propto 2^{-\alpha\ell}$.
- Variance: $\mathrm{Var}[\, X_\ell - X_{\ell-1}\,] \propto 2^{-\beta\ell}$.

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

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

Multilevel Monte Carlo: A General Framework

Example

For a standard European call option we have $\mathbb{E}[\, f(X)\,]$ for $X = S(T) - K$ and $f(X) = \max(X, 0)$. Approximating $S(T)$ by Euler-Maruyama 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 $\mathbb{E}[\,\mathbb{I}_{X>0}\,]$ 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:

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

Previous research

- Giles et al., *Finance and Stochastics*, "Analysing Multi-Level Monte Carlo for options with non-globally Lipschitz payoff" (July 2009)

  Original analysis of classical MLMC for discontinuous payoffs in SDE example.
- Giles et al., *SIAM/ASA Journal on Uncertainty Quantification*, "Multilevel Monte Carlo Approximation of Distribution Functions and Densities" (Jan. 2015)

  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.
- Bayer et al., "Numerical smoothing and hierarchical approximations for efficient option pricing and density estimation" (2020)

  Same as before. Smoothes the discontinuity by intergrating using a high order method with respect to one of the dimensions. Needs to find point of discontinuity to achieve fast convergence.

Previous research (adaptivity)

- Broadie et al., *Management Science*, "Efficient Risk Estimation via Nested Sequential Simulation" (June 2011)

  Adaptive sampling for nested expectation with Monte Carlo methods.
- Giles and Haji-Ali, *SIAM/ASA Journal on Uncertainty Quantification*, "Multilevel Nested Simulation for Efficient Risk Estimation" (Jan. 2019)

  Adaptive sampling for MLMC applied to nested expectations only. Requires stronger conditions on the random variables than here.
- Elfverson et al., *SIAM/ASA Journal on Uncertainty Quantification*, "A Multilevel Monte Carlo Method for Computing Failure Probabilities" (Jan. 2016)

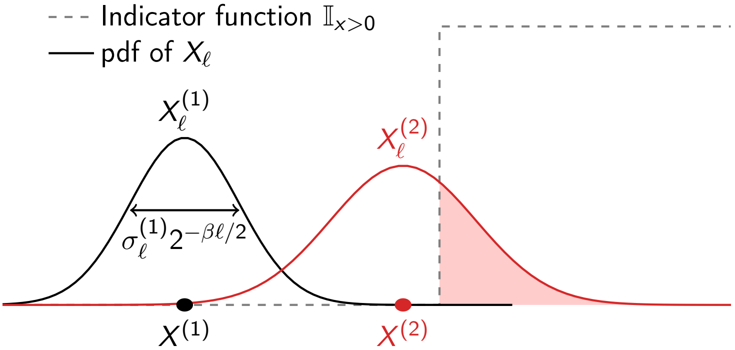
  Selective refinement of samples. Based on relaxing the condition. Assumes uniform almost sure error bounds (works well for PDEs with random coefficients but not stochastic models).

Adaptive Multilevel Monte Carlo: Algorithm

Refine samples of X_ℓ to $X_{\ell+\eta_\ell}$, where $0 \le \eta_\ell \le \lceil \theta\ell \rceil$ is the smallest integer for which

$$\delta_{\ell+\eta_\ell} \ge 2^{\frac{\gamma}{r}(\theta\ell(1-r)-\eta_\ell)}$$

for constants $r > 1$ and $0 \le \theta \le 1$. Recall that $\delta_\ell := \frac{|X_\ell|}{\sigma_\ell} \ge 0$.

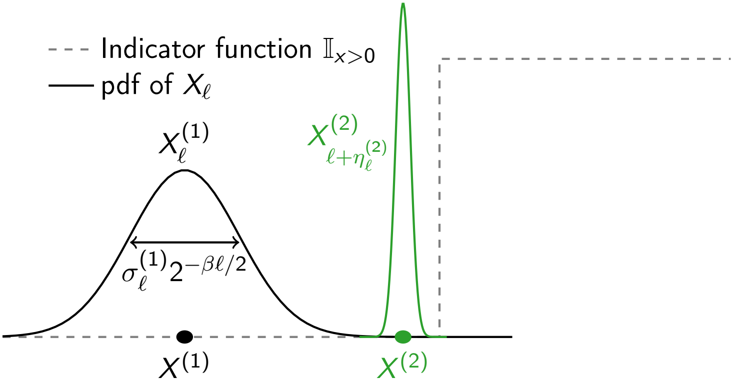


Adaptive Multilevel Monte Carlo: Algorithm

Refine samples of X_ℓ to $X_{\ell+\eta_\ell}$, where $0 \le \eta_\ell \le \lceil \theta\ell \rceil$ is the smallest integer for which

$$\delta_{\ell+\eta_\ell} \ge 2^{\frac{\gamma}{r}(\theta\ell(1-r)-\eta_\ell)}$$

for constants $r > 1$ and $0 \le \theta \le 1$. Recall that $\delta_\ell := \dfrac{|X_\ell|}{\sigma_\ell} \ge 0$.

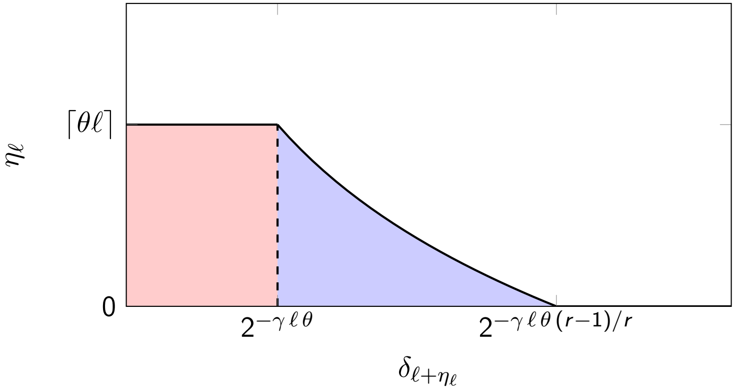


Adaptive Multilevel Monte Carlo: Algorithm

Refine samples of X_ℓ to $X_{\ell+\eta_\ell}$, where $0 \le \eta_\ell \le \lceil \theta\ell \rceil$ is the smallest integer for which

$$\delta_{\ell+\eta_\ell} \ge 2^{\frac{\gamma}{r}(\theta\ell(1-r)-\eta_\ell)}$$

for constants $r > 1$ and $0 \le \theta \le 1$. Recall that $\delta_\ell := \dfrac{|X_\ell|}{\sigma_\ell} \ge 0$.



Algorithm

Algorithm 1: Adaptive sampling at level ℓ

Input: $\ell, r, \theta, c > 0, \gamma, \beta$

Result: Adaptively refined sample $g_{\ell+\eta_\ell}$

Set $\eta_\ell = 0$;

Sample (g_ℓ, σ_ℓ);

Compute δ_ℓ given (g_ℓ, σ_ℓ);

while $|\delta_{\ell+\eta_\ell}| < c2^{\gamma(\theta\ell(1-r)-\eta_\ell)/r}$ *and* $\eta_\ell < \lceil \theta\ell \rceil$ **do**

  Refine $(g_{\ell+\eta_\ell}, \sigma_{\ell+\eta_\ell})$ to $(g_{\ell+\eta_\ell+1}, \sigma_{\ell+\eta_\ell+1})$;

  Compute $\delta_{\ell+\eta_\ell+1}$ given $(g_{\ell+\eta_\ell+1}, \sigma_{\ell+\eta_\ell+1})$;

  Set $\eta_\ell = \eta_\ell + 1$;

end

Output: $g_{\ell+\eta_\ell}$

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

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

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:

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

Adaptive Multilevel Monte Carlo: Complexity

Example

When approximating the price of a digital option

$$\mathbb{E}\left[\,\mathbb{I}_{S(T)>K}\,\right],$$

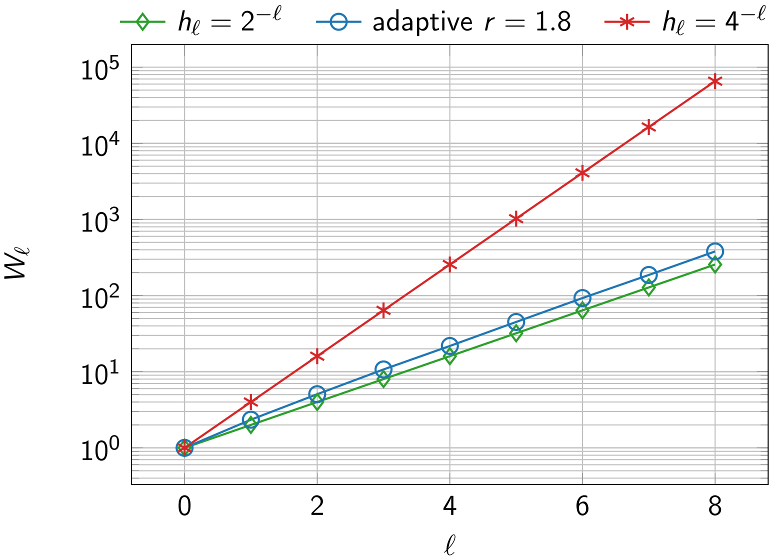
using Euler-Maruyama approximation of $S(T)$, or the risk estimation problem

$$\mathbb{E}\left[\,\mathbb{I}_{\mathbb{E}[Y|R]>0}\,\right],$$

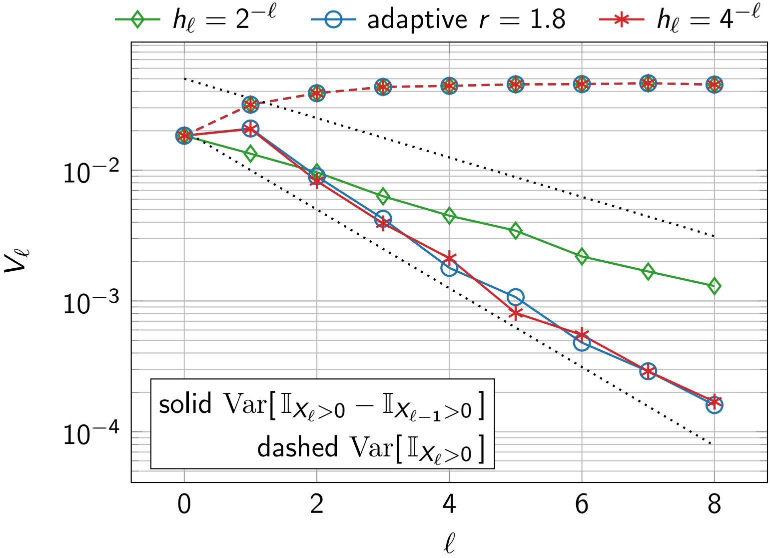
the assumptions hold for $\alpha = \beta = \gamma = 1$ and any $q < \infty$. The complexity is (for any $\nu > 0$)

- $\mathcal{O}\left(\varepsilon^{-3}\right)$ 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.

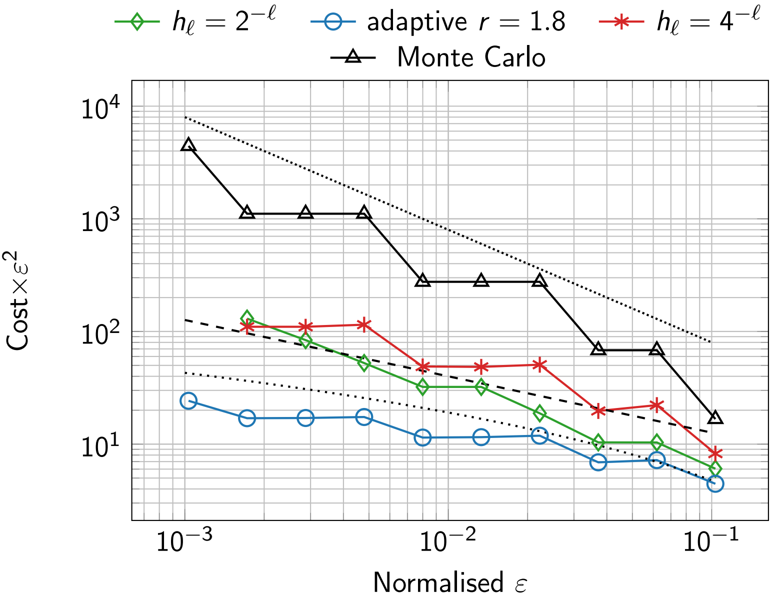
Numerical Test: Digital Options



Numerical Test: Digital Options



Numerical Test: Digital Options



Shortcomings

- The Kurtosis of the correcting grows with the level! This is a consequence of the 0-1 nature of the correction

$$\frac{\mathbb{E}[\,(\mathbb{I}_{X_\ell>0} - \mathbb{I}_{X_{\ell+\eta_\ell}>0})^4\,]}{\mathbb{E}[\,(\mathbb{I}_{X_\ell>0} - \mathbb{I}_{X_{\ell+\eta_\ell}>0})^2\,]^2} = \mathbb{E}\Big[\,|(\mathbb{I}_{X_\ell>0} - \mathbb{I}_{X_{\ell+\eta_\ell}>0})|\,\Big]^{-1}$$

- The better the variance convergence the worse the Kurtosis!
- High Kurtosis means that computing variance estimates is more difficult.
- Two methods to resolve this:
  - Smooth out the discontinuity. Not always natural to apply (for example in nested expectation).
  - More stable variance estimates (for example using a Bayesian model and previous levels). Upcoming work. See also Collier et al., *BIT Numerical Mathematics*, "A continuation multilevel Monte Carlo algorithm" (Sept. 2014).

Other points in the paper

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.

Conclusion

- 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.

Numerical Tests: Digital Options

For constant $\mu, \sigma, S(0)$ consider the asset

$$\mathrm{d}S(t) = \mu S(t)\,\mathrm{d}t + \sigma S(t)\,\mathrm{d}W(t).$$

Compute

$$\mathbb{E}[\,\mathbb{I}_{X>0}\,] \coloneqq \mathbb{E}[\,\mathbb{I}_{S(T)-K>0}\,]$$

for some strike price $K > 0$. We use Euler-Maruyama with a step size $h_\ell = 2^{-\ell}$ to approximate $S_{h_\ell}(\cdot) \approx S(\cdot)$ and set

$$X_\ell \coloneqq S_{h_\ell}(T) - K.$$

The assumptions are satisfied using constant $\sigma_\ell \equiv 1$ for $\alpha = \beta = \gamma = 1$ and any $q < \infty$ giving complexity $\mathcal{O}\big(\varepsilon^{-2.5-\nu}\big)$ for standard Multilevel Monte Carlo and $\mathcal{O}(\varepsilon^{-2-\nu})$ for any $\nu > 0$ using adaptive Multilevel Monte Carlo.

Numerical Tests: Digital Options

Consider the assets

$$\mathrm{d}S^{(i)}(t) = \mu^i S^{(i)}(t) + \sigma^i S^{(i)}(t)\mathrm{d}W^{(i)}(t)$$

where

$$W^i(t) = \rho W^{(i)}_{\mathrm{com}}(t) + \sqrt{1-\rho^2}\, W^{(i)}_{\mathrm{ind}}(t)$$

for $1 \le i \le 10$. Consider the digital option with payoff

$$\mathbb{I}_{\left(\frac{1}{10}\sum_{i=1}^{10} S^{(i)}(t)\right) > K}.$$

Thus, compute

$$\mathbb{E}\left[\, \mathbb{I}_{\left(\frac{1}{10}\sum_{i=1}^{10} S^{(i)}(t)\right) > K} \,\right].$$