Computational Methods in Uncertainty Quantification

Robert Scheichl

Department of Mathematical Sciences
University of Bath
Romberg Visiting Professor at University of Heidelberg

HGS MathComp Compact Course
IWR, Universität Heidelberg, June 11-17 2015

PART 2

Lecture 2 – Monte Carlo Methods

- Monte Carlo methods
- History
- Convergence analysis
- Variance reduction techniques
- Example: Predator-prey dynamical system
- Multilevel Monte Carlo methods

Monte Carlo Methods

Monte Carlo

- In 1777, George Louis Leclerc, Comte de Buffon (1707–1788), French naturalist and mathematician, posed the following problem:

> *Let a needle of length ℓ be thrown at random onto a horizontal plane ruled with parallel straight lines spaced by a distance $d > \ell$ from each other. What is the probability p that the needle will intersect one of these lines?*

- In 1777, George Louis Leclerc, Comte de Buffon (1707–1788), French naturalist and mathematician, posed the following problem:

  > *Let a needle of length ℓ be thrown at random onto a horizontal plane ruled with parallel straight lines spaced by a distance $d > \ell$ from each other. What is the probability p that the needle will intersect one of these lines?*

- Answer: $p = \frac{2\ell}{\pi d}$ (simple geometric arguments)

Monte Carlo Methods

The Buffon Needle Problem

- In 1777, George Louis Leclerc, Comte de Buffon (1707–1788), French naturalist and mathematician, posed the following problem:

  Let a needle of length ℓ be thrown at random onto a horizontal plane ruled with parallel straight lines spaced by a distance $d > \ell$ from each other. What is the probability p that the needle will intersect one of these lines?

- Answer: $p = \frac{2\ell}{\pi d}$ (simple geometric arguments)
- Laplace later used similar randomised experiment to approximate π.

Monte Carlo Methods

The Buffon Needle Problem

- In 1777, George Louis Leclerc, Comte de Buffon (1707–1788), French naturalist and mathematician, posed the following problem:

  > *Let a needle of length ℓ be thrown at random onto a horizontal plane ruled with parallel straight lines spaced by a distance $d > \ell$ from each other. What is the probability p that the needle will intersect one of these lines?*

- Answer: $p = \frac{2\ell}{\pi d}$ (simple geometric arguments)
- Laplace later used similar randomised experiment to approximate π.
- The term "Monte Carlo method" was coined by Ulam, von Neumann, Metropolis in the Manhattan project (Los Alamos, 1946).

THE ROYAL SOCIETY

Ants estimate area using Buffon's needle

Eamonn B. Mallon* **and Nigel R. Franks**

Centre for Mathematical Biology, and Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK

We show for the first time, to our knowledge, that ants can measure the size of potential nest sites. Nest size assessment is by individual scouts. Such scouts always make more than one visit to a potential nest before initiating an emigration of their nest mates and they deploy individual-specific trails within the potential new nest on their first visit. We test three alternative hypotheses for the way in which scouts might measure nests. Experiments indicated that individual scouts use the intersection frequency between their own paths to assess nest areas. These results are consistent with ants using a 'Buffon's needle algorithm' to assess nest areas.

Keywords: ants; colony emigration; individual-specific pheromones; *Leptothorax*; nest sites; rules of thumb

Proceedings of the Royal Society of London, 2000

- Let $\{H_k\}_{k\in\mathbb{N}}$ denote a sequence of i.i.d. binomial random variables s.t.
$$H_k(\omega) = \begin{cases} 1 & \text{if } k\text{-th needle intersects a line,} \\ 0 & \text{otherwise.} \end{cases}$$
- Their common distribution is that of a Bernoulli trial with success probability $p = 2\ell/\pi d$. In particular: $\mathbf{E}[H_k] = p \quad \forall k$.
- $S_N = H_1 + \cdots + H_N$ is the total number of hits after N throws.

- Let $\{H_k\}_{k\in\mathbb{N}}$ denote a sequence of i.i.d. binomial random variables s.t.

$$H_k(\omega) = \begin{cases} 1 & \text{if } k\text{-th needle intersects a line,} \\ 0 & \text{otherwise.} \end{cases}$$

- Their common distribution is that of a Bernoulli trial with success probability $p = 2\ell/\pi d$. In particular: $\mathbf{E}[H_k] = p \quad \forall k$.
- $S_N = H_1 + \cdots + H_N$ is the total number of hits after N throws.
- *Strong Law of Large Numbers:*

$$\frac{S_N}{N} \to p \quad \text{almost surely (a.s.)}$$

- Let $\{H_k\}_{k\in\mathbb{N}}$ denote a sequence of i.i.d. binomial random variables s.t.

$$H_k(\omega) = \begin{cases} 1 & \text{if } k\text{-th needle intersects a line,} \\ 0 & \text{otherwise.} \end{cases}$$

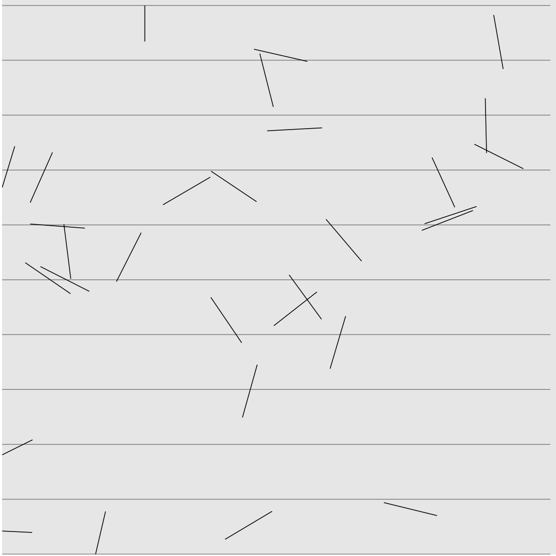
- Their common distribution is that of a Bernoulli trial with success probability $p = 2\ell/\pi d$. In particular: $\mathbf{E}[H_k] = p \quad \forall k.$
- $S_N = H_1 + \cdots + H_N$ is the total number of hits after N throws.
- *Strong Law of Large Numbers:*

$$\frac{S_N}{N} \to p \quad \text{almost surely (a.s.)}$$

- Compute realizations of H_k by sampling $X_k \sim \mathrm{U}[0, d/2]$ (distance of needle center to closest line) and $\Theta_k \sim \mathrm{U}[0, \pi/2]$ (acute angle of needle with lines) using a random number generator.

Monte Carlo Methods

Monte Carlo Simulation for the Buffon Needle Problem



- Setting $d = 2$, $\ell = 1$ gives $p = \frac{1}{\pi}$. We should get $N/S_N \stackrel{N\to\infty}{\longrightarrow} \pi$.

- Setting $d = 2$, $\ell = 1$ gives $p = \frac{1}{\pi}$. We should get $N/S_N \stackrel{N\to\infty}{\longrightarrow} \pi$.
- A Matlab experiment yields

N	S_N	N/S_N	rel. Error
10	3	3.333	6.10e-2
100	32	3.125	5.28e-3
1000	330	3.030	3.54e-2
10000	3188	3.137	1.54e-3

- Setting $d = 2$, $\ell = 1$ gives $p = \frac{1}{\pi}$. We should get $N/S_N \overset{N\to\infty}{\longrightarrow} \pi$.
- A Matlab experiment yields

N	S_N	N/S_N	rel. Error
10	3	3.333	6.10e-2
100	32	3.125	5.28e-3
1000	330	3.030	3.54e-2
10000	3188	3.137	1.54e-3

- Mario Lazzarini (1901) built machine that carries out repetitions of this random experiment. His needle was 2.5cm long and the lines 3.0cm apart. He claims to have observed 1808 intersections for 3408 throws, i.e

$$\pi \approx 2 \cdot \frac{2.5}{3} \cdot \frac{3408}{1808} = 3.14159292035398\ldots$$

  A relative error of $8.5 \cdot 10^{-8}$!

- Setting $d = 2$, $\ell = 1$ gives $p = \frac{1}{\pi}$. We should get $N/S_N \stackrel{N\to\infty}{\longrightarrow} \pi$.
- A Matlab experiment yields

N	S_N	N/S_N	rel. Error
10	3	3.333	6.10e-2
100	32	3.125	5.28e-3
1000	330	3.030	3.54e-2
10000	3188	3.137	1.54e-3

- Mario Lazzarini (1901) built machine that carries out repetitions of this random experiment. His needle was 2.5cm long and the lines 3.0cm apart. He claims to have observed 1808 intersections for 3408 throws, i.e
$$\pi \approx 2 \cdot \frac{2.5}{3} \cdot \frac{3408}{1808} = 3.14159292035398 3\ldots$$
A relative error of $8.5 \cdot 10^{-8}$!   Is this too good to be true?

Monte Carlo Methods

Basic Monte Carlo simulation – Convergence results

- Given a sequence $\{X_k\}$ of i.i.d. copies of a given random variable X, basic MC simulation uses the estimator

$$\mathbf{E}[X] \approx \frac{S_N}{N}, \qquad S_N = X_1 + \cdots + X_N.$$

- By the *Strong Law of Large Numbers*, $\dfrac{S_N}{N} \to \mathbf{E}[X]$ a.s.

Monte Carlo Methods

Basic Monte Carlo simulation – Convergence results

- Given a sequence $\{X_k\}$ of i.i.d. copies of a given random variable X, basic MC simulation uses the estimator

$$\mathbf{E}[X] \approx \frac{S_N}{N}, \qquad S_N = X_1 + \cdots + X_N.$$

- By the *Strong Law of Large Numbers*, $\dfrac{S_N}{N} \to \mathbf{E}[X]$ a.s.
- Also, for any measurable function f, $\dfrac{1}{N}\displaystyle\sum_{k=1}^{N} f(X_k) \to \mathbf{E}[f(X)]$ a.s.

- Given a sequence $\{X_k\}$ of i.i.d. copies of a given random variable X, basic MC simulation uses the estimator

$$\mathbf{E}[X] \approx \frac{S_N}{N}, \qquad S_N = X_1 + \cdots + X_N.$$

- By the *Strong Law of Large Numbers*, $\frac{S_N}{N} \to \mathbf{E}[X]$ a.s.
- Also, for any measurable function f, $\frac{1}{N}\sum_{k=1}^{N} f(X_k) \to \mathbf{E}[f(X)]$ a.s.
- If $\mathbf{E}[X] = \mu$ and $\mathbf{Var}[X] = \sigma^2$, then (via the *Central Limit Theorem*)

$$\mathbf{E}[S_N] = N\mu, \quad \mathbf{Var}[S_N] = N\sigma^2 \quad \text{and} \quad S_N^* = \frac{S_N - N\mu}{\sqrt{N}\sigma} \to \mathrm{N}(0,1),$$

i.e. the estimate is unbiased, the standard error is $\sigma N^{-1/2}$ and the distribution of the normalised RV S_N^* becomes Gaussian as $N \to \infty$.

1. Since
$$\mathbf{E}\left[\left(\frac{S_N}{N} - \mu\right)^2\right] = \mathbf{Var}\,\frac{S_N}{N} = \frac{\sigma^2}{N} \to 0,$$
we have mean square convergence of S_N/N to μ.

1. Since
$$\mathbf{E}\left[\left(\frac{S_N}{N}-\mu\right)^2\right] = \mathbf{Var}\,\frac{S_N}{N} = \frac{\sigma^2}{N} \to 0,$$
we have mean square convergence of S_N/N to μ.

2. Also *Chebyshev's Inequality* implies, for any $\epsilon > 0$,
$$\mathbf{P}\left\{\left|\frac{S_N}{N}-\mu\right| > N^{-1/2+\epsilon}\right\} \leq \frac{\sigma^2}{N^{2\epsilon}},$$
i.e. the probability of the error being larger than $N^{-1/2+\epsilon}$ converges to zero for any $\epsilon > 0$, as $N \to \infty$.

1. Since
$$\mathbf{E}\left[\left(\frac{S_N}{N}-\mu\right)^2\right] = \mathbf{Var}\,\frac{S_N}{N} = \frac{\sigma^2}{N} \to 0,$$
we have mean square convergence of S_N/N to μ.

2. Also *Chebyshev's Inequality* implies, for any $\epsilon > 0$,
$$\mathbf{P}\left\{\left|\frac{S_N}{N}-\mu\right| > N^{-1/2+\epsilon}\right\} \le \frac{\sigma^2}{N^{2\epsilon}},$$
i.e. the probability of the error being larger than $N^{-1/2+\epsilon}$ converges to zero for any $\epsilon > 0$, as $N \to \infty$.

3. If $\rho := \mathbf{E}\left[|X-\mu|^3\right] < \infty$, then the *Berry-Esseen Inequality* gives
$$|\mathbf{P}\{S_N^* \le x\} - \Phi(x)| \le \frac{\rho}{2\sigma^3\sqrt{N}},$$
where Φ denotes *cumulative density function (CDF)* of $\mathrm{N}(0,1)$.

Exercise 1

(a) Using the Berry-Esseen bound derive a confidence interval for the estimate S_N/N and (upper and lower) bounds on the probability that μ falls into this confidence interval.

(b) In the Buffon needle problem, we have

$$\mathbf{E}\left[H_k\right] = p, \ \mathbf{Var}[H_k] = p(1-p), \ \mathbf{E}\left[|H_k - p|^3\right] = p(1-p)(1-2p+2p^2).$$

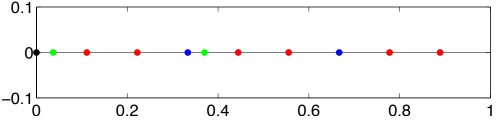
Calculate the confidence interval for this problem in the case $N = 3408$, $\ell = 2.5$, $d = 3$, and thus check how likely it is that Lazzarini's machine would produce 1808 intersections and a relative accuracy of π of $8.5 \cdot 10^{-8}$.

In quasi-Monte Carlo methods, the samples are not chosen randomly, but special (deterministic) number sequences, known as low-discrepancy sequences, are used instead. Discrepancy is a measure of equidistribution of a number sequence.

In quasi-Monte Carlo methods, the samples are not chosen randomly, but special (deterministic) number sequences, known as low-discrepancy sequences, are used instead. Discrepancy is a measure of equidistribution of a number sequence.

Example: The van der Corput sequence is such a low-discrepancy sequence for the unit interval. For base 3, it is given by $x_n = \frac{k}{3^j}$, where j increases monotonically and, for each j, k runs through all nonnegative integers such that $k/3^j$ is an irreducible fraction < 1. The ordering in k is obtained by representing k in base 3 and reversing the digits. The first 11 numbers are

$$\{x_n\}_{n=1}^{11} = \{0, \frac{1}{3}, \frac{2}{3}, \frac{1}{9}, \frac{4}{9}, \frac{7}{9}, \frac{2}{9}, \frac{5}{9}, \frac{8}{9}, \frac{1}{27}, \frac{10}{27}\}.$$



- Replacing i.i.d. random numbers sampled from $\mathrm{U}[0,1]$ in a standard Monte Carlo approximation of $\mathbf{E}[f(X)]$ for some $f \in C^\infty(0,1)$ and $X \sim \mathrm{U}[0,1]$, by the van der Corput sequence of length N, yields a quasi-Monte Carlo method.

- Replacing i.i.d. random numbers sampled from $\mathrm{U}[0,1]$ in a standard Monte Carlo approximation of $\mathbf{E}\left[f(X)\right]$ for some $f \in C^\infty(0,1)$ and $X \sim \mathrm{U}[0,1]$, by the van der Corput sequence of length N, yields a quasi-Monte Carlo method.

- The convergence rate is improved from $\mathscr{O}(N^{-1/2})$ to $\mathscr{O}(N^{-2})$.

- Replacing i.i.d. random numbers sampled from $\mathrm{U}[0,1]$ in a standard Monte Carlo approximation of $\mathbf{E}[f(X)]$ for some $f \in C^\infty(0,1)$ and $X \sim \mathrm{U}[0,1]$, by the van der Corput sequence of length N, yields a quasi-Monte Carlo method.
- The convergence rate is improved from $\mathscr{O}(N^{-1/2})$ to $\mathscr{O}(N^{-2})$.
- Although this improvement is impressive, the method does not generalise easily and the rate of convergence depends on the problem.

- Replacing i.i.d. random numbers sampled from $\mathrm{U}[0,1]$ in a standard Monte Carlo approximation of $\mathbf{E}[f(X)]$ for some $f \in C^\infty(0,1)$ and $X \sim \mathrm{U}[0,1]$, by the van der Corput sequence of length N, yields a quasi-Monte Carlo method.
- The convergence rate is improved from $\mathscr{O}(N^{-1/2})$ to $\mathscr{O}(N^{-2})$.
- Although this improvement is impressive, the method does not generalise easily and the rate of convergence depends on the problem.
- In particular, the rate of convergence for a quasi-Monte Carlo method generally does depend on the dimension.

- Replacing i.i.d. random numbers sampled from $\mathrm{U}[0,1]$ in a standard Monte Carlo approximation of $\mathbf{E}\left[f(X)\right]$ for some $f \in C^{\infty}(0,1)$ and $X \sim \mathrm{U}[0,1]$, by the van der Corput sequence of length N, yields a quasi-Monte Carlo method.

- The convergence rate is improved from $\mathscr{O}(N^{-1/2})$ to $\mathscr{O}(N^{-2})$.

- Although this improvement is impressive, the method does not generalise easily and the rate of convergence depends on the problem.

- In particular, the rate of convergence for a quasi-Monte Carlo method generally does depend on the dimension.

Wed: Dimension-independent QMC results for the "fruit-fly"

Monte Carlo Methods

Variance reduction

The constant in the MC convergence rates is the variance σ^2 of the RV from which MC samples are being drawn. By designing an equivalent MC approximation with lower variance, we can expect faster convergence.

Monte Carlo Methods

Variance reduction

The constant in the MC convergence rates is the variance σ^2 of the RV from which MC samples are being drawn. By designing an equivalent MC approximation with lower variance, we can expect faster convergence.

- To approximate $\mathbf{E}[X]$ by standard MC, we draw independent samples $\{X_k\}_{k=1}^N$ of X and compute the sample average S_N/N.

The constant in the MC convergence rates is the variance σ^2 of the RV from which MC samples are being drawn. By designing an equivalent MC approximation with lower variance, we can expect faster convergence.

- To approximate $\mathbf{E}[X]$ by standard MC, we draw independent samples $\{X_k\}_{k=1}^N$ of X and compute the sample average S_N/N.
- Now assume a second set of samples $\{\tilde{X}_k\}_{k=1}^N$ of X is given with sample average $\tilde{S}_N/N$.

The constant in the MC convergence rates is the variance σ^2 of the RV from which MC samples are being drawn. By designing an equivalent MC approximation with lower variance, we can expect faster convergence.

- To approximate $\mathbf{E}[X]$ by standard MC, we draw independent samples $\{X_k\}_{k=1}^N$ of X and compute the sample average S_N/N.
- Now assume a second set of samples $\{\tilde{X}_k\}_{k=1}^N$ of X is given with sample average $\tilde{S}_N/N$.
- Since both sample averages converge to $\mathbf{E}[X]$, so does $\frac{1}{2}(S_N/N + \tilde{S}_N/N)$.
- When X_k and $\tilde{X}_k$ are negatively correlated they are called antithetic samples, and the approximation $\frac{1}{2N}(S_N + \tilde{S}_N)$ is a more reliable approximation of $\mathbf{E}[X]$ than $\frac{1}{2N}S_{2N}$.

Theorem

Let the two sequences of RVs $\{X_k\}$ and $\{\tilde{X}_k\}$ be identically distributed with

$$\mathbf{Cov}(X_j, X_k) = \mathbf{Cov}(\tilde{X}_j, \tilde{X}_k) = 0 \qquad \textit{for } j \neq k.$$

Then the sample averages S_N/N and $\tilde{S}_N/N$ satisfy

$$\mathbf{Var}\left[\frac{S_N + \tilde{S}_N}{2N}\right] = \mathbf{Var}\left[\frac{S_{2N}}{2N}\right] + \frac{1}{2}\,\mathbf{Cov}\left(\frac{S_N}{N}, \frac{\tilde{S}_N}{N}\right) \leq \mathbf{Var}\left[\frac{S_N}{N}\right].$$

Theorem

Let the two sequences of RVs $\{X_k\}$ *and* $\{\tilde{X}_k\}$ *be identically distributed with*

$$\mathbf{Cov}(X_j, X_k) = \mathbf{Cov}(\tilde{X}_j, \tilde{X}_k) = 0 \quad \text{for } j \neq k.$$

Then the sample averages S_N/N *and* $\tilde{S}_N/N$ *satisfy*

$$\mathbf{Var}\left[\frac{S_N + \tilde{S}_N}{2N}\right] = \mathbf{Var}\left[\frac{S_{2N}}{2N}\right] + \frac{1}{2}\,\mathbf{Cov}\left(\frac{S_N}{N}, \frac{\tilde{S}_N}{N}\right) \leq \mathbf{Var}\left[\frac{S_N}{N}\right].$$

- **Worst case:** Variance of average of N samples and N antithetic samples no better than variance of N independent samples.

Theorem

Let the two sequences of RVs $\{X_k\}$ *and* $\{\tilde{X}_k\}$ *be identically distributed with*

$$\mathbf{Cov}(X_j, X_k) = \mathbf{Cov}(\tilde{X}_j, \tilde{X}_k) = 0 \quad \text{for } j \neq k.$$

Then the sample averages S_N/N *and* $\tilde{S}_N/N$ *satisfy*

$$\mathbf{Var}\left[\frac{S_N + \tilde{S}_N}{2N}\right] = \mathbf{Var}\left[\frac{S_{2N}}{2N}\right] + \frac{1}{2}\mathbf{Cov}\left(\frac{S_N}{N}, \frac{\tilde{S}_N}{N}\right) \leq \mathbf{Var}\left[\frac{S_N}{N}\right].$$

- **Worst case:** Variance of average of N samples and N antithetic samples no better than variance of N independent samples.
- **Best case:** negatively correlated S_N/N and $\tilde{S}_N/N$, therefore variance of N samples and N antithetic samples less than variance of $2N$ indepependent samples.

Example: Predator-prey dynamical system

Explicit Euler discretisation

Consider the popular model of the dynamics of two interacting populations

$$\dot{\mathbf{u}} = \begin{bmatrix} \dot{u}_1 \\ \dot{u}_2 \end{bmatrix} = \begin{bmatrix} u_1(1-u_2) \\ u_2(u_1-1) \end{bmatrix}, \quad \mathbf{u}(0) = \mathbf{u}_0.$$

Assume the vector of initial conditions $\mathbf{u}_0$ is uncertain and that it is modeled as a (uniform) random vector $\mathbf{u}_0 \sim \mathrm{U}(\Gamma)$, where Γ denotes the square

$$\Gamma = \overline{\mathbf{u}}_0 + [-\epsilon, \epsilon]^2.$$

Example: Predator-prey dynamical system

Explicit Euler discretisation

Consider the popular model of the dynamics of two interacting populations

$$\dot{\mathbf{u}} = \begin{bmatrix} \dot{u}_1 \\ \dot{u}_2 \end{bmatrix} = \begin{bmatrix} u_1(1-u_2) \\ u_2(u_1-1) \end{bmatrix}, \quad \mathbf{u}(0) = \mathbf{u}_0.$$

Assume the vector of initial conditions $\mathbf{u}_0$ is uncertain and that it is modeled as a (uniform) random vector $\mathbf{u}_0 \sim \mathrm{U}(\Gamma)$, where Γ denotes the square

$$\Gamma = \overline{\mathbf{u}}_0 + [-\epsilon, \epsilon]^2.$$

- **Goal:** estimate $\mathbf{E}[u_1(T)]$ at time $T > 0$.
- Denote by $\mathbf{u}_M = \mathbf{u}_M(\omega)$ the explicit Euler approximation after M time steps of length $\Delta t = \frac{T}{M}$ starting with initial data $\mathbf{u}_0 = \mathbf{u}_0(\omega)$.
- Define the QoI $Q = \phi(\mathbf{u}(T)) = u_1(T)$ for $\mathbf{u} = [u_1, u_2]^\mathsf{T}$ and estimate $\mathbf{E}[Q_M]$ using the MC method just described, where $Q_M = \phi(\mathbf{u}_M)$.
- Expect better approximations for N large and Δt small.

Example: Predator-prey dynamical system

Monte Carlo Estimator

- Denote the Monte Carlo estimator for $\mathbf{E}[Q_M]$ by

$$\widehat{Q}_M := \widehat{Q}_{M,N} = \frac{1}{N}\sum_{k=1}^{N} Q_M^{(k)}$$

i.e the average over N samples $\{Q_M^{(k)}\}_{k=1}^{N}$ of Q_M.

Example: Predator-prey dynamical system

Monte Carlo Estimator

- Denote the Monte Carlo estimator for $\mathbf{E}[Q_M]$ by

$$\widehat{Q}_M := \widehat{Q}_{M,N} = \frac{1}{N}\sum_{k=1}^{N} Q_M^{(k)}$$

  i.e the average over N samples $\{Q_M^{(k)}\}_{k=1}^{N}$ of Q_M.

- Error with N samples and $M = T/\Delta t$ time steps:

$$e_{N,M} = |\mathbf{E}[Q] - \widehat{Q}_M| \le \underbrace{|\mathbf{E}[Q] - \mathbf{E}[Q_M]|}_{\text{discretisation error}} + \underbrace{|\mathbf{E}[Q_M] - \widehat{Q}_M|}_{\text{Monte Carlo error}}$$

Example: Predator-prey dynamical system

Monte Carlo Estimator

- Denote the Monte Carlo estimator for $\mathbf{E}[Q_M]$ by

$$\widehat{Q}_M := \widehat{Q}_{M,N} = \frac{1}{N}\sum_{k=1}^{N} Q_M^{(k)}$$

i.e the average over N samples $\{Q_M^{(k)}\}_{k=1}^N$ of Q_M.

- Error with N samples and $M = T/\Delta t$ time steps:

$$e_{N,M} = |\mathbf{E}[Q] - \widehat{Q}_M| \le \underbrace{|\mathbf{E}[Q] - \mathbf{E}[Q_M]|}_{\text{discretisation error}} + \underbrace{|\mathbf{E}[Q_M] - \widehat{Q}_M|}_{\text{Monte Carlo error}}$$

Exercise 2

Show that the mean square error can be expanded (with equality!)

$$\mathbf{E}\left[\left(\mathbf{E}[Q] - \widehat{Q}_M\right)^2\right] = \left(\mathbf{E}[Q - Q_M]\right)^2 + \frac{\mathbf{Var}[Q_M]}{N}$$

Hint: Note that $\mathbf{E}[Q]$ is constant and only $\widehat{Q}_M$ is actually random.

- Explicit Euler discretisation error (with some constant $K > 0$):

$$\|\mathbf{u}(T) - \mathbf{u}_M\| \le KM^{-1}.$$

Example: Predator-prey dynamical system

Discretisation Error – Bias

- Explicit Euler discretisation error (with some constant $K > 0$):

$$\|\mathbf{u}(T) - \mathbf{u}_M\| \leq KM^{-1}.$$

- ϕ Lipschitz-continuous with constant $L = 1$:

$$|\phi(\mathbf{u}(T)) - \phi(\mathbf{u}_M)| \leq KL\, M^{-1}.$$

Example: Predator-prey dynamical system

Discretisation Error – Bias

- Explicit Euler discretisation error (with some constant $K > 0$):

$$\|\mathbf{u}(T) - \mathbf{u}_M\| \leq K M^{-1}.$$

- ϕ Lipschitz-continuous with constant $L = 1$:

$$|\phi(\mathbf{u}(T)) - \phi(\mathbf{u}_M)| \leq KL\, M^{-1}.$$

- Therefore

$$|\mathbf{E}[Q] - \mathbf{E}[Q_M]| = |\mathbf{E}[Q - Q_M]| \leq KL\, M^{-1}.$$

Example: Predator-prey dynamical system

Balancing discretisation and MC error

- For the MC error, from **Exercise 1** with $\mathbf{Var}[Q_M] = \sigma_M^2$ we get

$$\mathbf{P}\left(\left| \mathbf{E}\left[Q_M\right] - \widehat{Q}_{M,N} \right| \le \frac{1.96\sigma_M}{\sqrt{N}} \right) > 0.95 + \mathscr{O}(N^{-1/2})$$

Example: Predator-prey dynamical system

Balancing discretisation and MC error

- For the MC error, from **Exercise 1** with $\mathbf{Var}[Q_M] = \sigma_M^2$ we get

$$\mathbf{P}\left(\left|\mathbf{E}\left[Q_M\right] - \widehat{Q}_{M,N}\right| \le \frac{1.96\sigma_M}{\sqrt{N}}\right) > 0.95 + \mathscr{O}(N^{-1/2})$$

- Combined with discretisation error:

$$\mathbf{P}\left(e_{N,M} \le \frac{KL}{M} + \frac{1.96\sigma_M}{\sqrt{N}}\right) > 0.95 + \mathscr{O}(N^{-1/2}).$$

Example: Predator-prey dynamical system

Balancing discretisation and MC error

- For the MC error, from **Exercise 1** with $\mathbf{Var}[Q_M] = \sigma_M^2$ we get

$$\mathbf{P}\left(\left|\mathbf{E}\left[Q_M\right] - \widehat{Q}_{M,N}\right| \le \frac{1.96\sigma_M}{\sqrt{N}}\right) > 0.95 + \mathscr{O}(N^{-1/2})$$

- Combined with discretisation error:

$$\mathbf{P}\left(e_{N,M} \le \frac{KL}{M} + \frac{1.96\sigma_M}{\sqrt{N}}\right) > 0.95 + \mathscr{O}(N^{-1/2}).$$

- Balance discretization and MC errors:

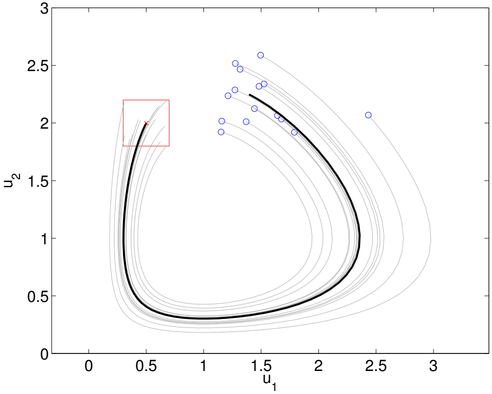
$$\frac{KL}{M} \approx \frac{\mathrm{TOL}}{2} \quad \text{and} \quad \frac{1.96\sigma_M}{\sqrt{N}} \approx \frac{\mathrm{TOL}}{2},$$

leads to

$$M \approx \frac{2KL}{\mathrm{TOL}}, \quad N \approx \frac{16\sigma_M^2}{\mathrm{TOL}^2} \quad \text{and so} \quad \text{Cost} = \mathscr{O}(\mathrm{TOL}^{-3})$$

Example: Predator-prey dynamical system

Sample trajectories



Population dynamics problem integrated over $[0, T = 6]$ with $\overline{\mathbf{u}}_0 = [0.5, 2]^{\mathsf{T}}$ and $\epsilon = 0.2$. Unperturbed trajectory (black) along with 15 perturbed trajectories. For the unperturbed trajectory $u_1(T) = 1.3942$.

Example: Predator-prey dynamical system

Antithetic sampling

We may introduce antithetic sampling to this problem by noting that, if $\mathbf{u}_0 \sim \mathrm{U}(\Gamma)$, then the same holds for the random vector

$$\tilde{\mathbf{u}}_0 := 2\overline{\mathbf{u}}_0 - \mathbf{u}_0.$$

Thus, the trajectories generated by the random initial data $\tilde{\mathbf{u}}_0$ have the same distribution as those generated by $\mathbf{u}_0$.

Example: Predator-prey dynamical system

Antithetic sampling

We may introduce antithetic sampling to this problem by noting that, if $\mathbf{u}_0 \sim \mathrm{U}(\Gamma)$, then the same holds for the random vector

$$\tilde{\mathbf{u}}_0 := 2\overline{\mathbf{u}}_0 - \mathbf{u}_0.$$

Thus, the trajectories generated by the random initial data $\tilde{\mathbf{u}}_0$ have the same distribution as those generated by $\mathbf{u}_0$.

- Let $Q_M = \phi(\mathbf{u}_M)$ be the basic samples and $\tilde{Q}_M = \phi(\tilde{\mathbf{u}}_M)$ the antithetic counterparts. Note that all pairs of samples are independent except each sample and its antithetic counterpart.

We may introduce antithetic sampling to this problem by noting that, if $\mathbf{u}_0 \sim \mathrm{U}(\Gamma)$, then the same holds for the random vector

$$\tilde{\mathbf{u}}_0 := 2\overline{\mathbf{u}}_0 - \mathbf{u}_0.$$

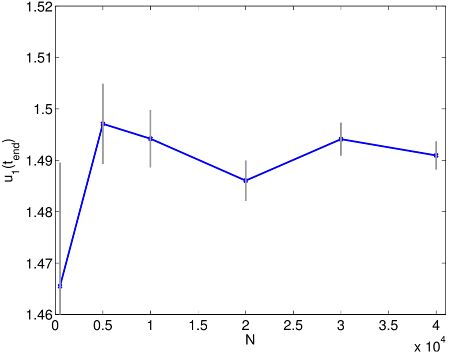
Thus, the trajectories generated by the random initial data $\tilde{\mathbf{u}}_0$ have the same distribution as those generated by $\mathbf{u}_0$.

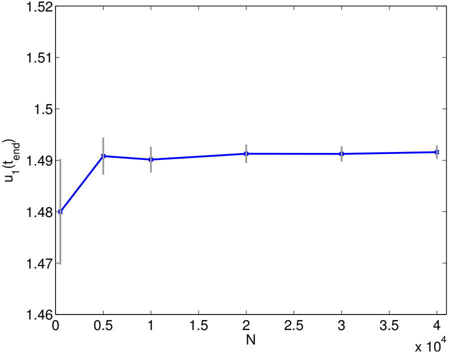
- Let $Q_M = \phi(\mathbf{u}_M)$ be the basic samples and $\tilde{Q}_M = \phi(\tilde{\mathbf{u}}_M)$ the antithetic counterparts. Note that all pairs of samples are independent except each sample and its antithetic counterpart.
- Then use $\frac{1}{2}(\widehat{Q}_{M,N} + \widehat{\tilde{Q}}_{M,N})$ instead of $\widehat{Q}_{M,2N}$ (same cost).
- To estimate $\mathbf{Var}[Q_M]$ and $\mathbf{Cov}(Q_M, \tilde{Q}_M)$ we use sample variance and covariance (resp.), i.e.

$$\frac{1}{N-1}\sum_{k=1}^{N}(Q_M^{(k)} - \widehat{Q}_{M,N})^2 \quad \text{and} \quad \frac{1}{N-1}\sum_{k=1}^{N}(Q_M^{(k)} - \widehat{Q}_{M,N})(\tilde{Q}_M^{(k)} - \widehat{\tilde{Q}}_{M,N})$$

Example: Predator-prey dynamical system

Numerical Experiment – Comparing standard and antithetic sampling





MC estimation of $\mathbf{E}\left[u_1(T)\right]$ using standard MC with N samples (left) vs. MC with antithetic sampling using $N/2$ samples of the initial data (right), showing the estimate along with 95% confidence intervals.

Exercise 3

(a) Find an estimate for $\mathbf{Var}\left[\frac{1}{2}(\widehat{Q}_{M,N} + \widehat{\tilde{Q}}_{M,N})\right]$ based on the sample variances and covariances of $\{Q_M^{(k)}\}$ and $\{\tilde{Q}_M^{(k)}\}$ defined above.

(b) Implement the Monte Carlo method for the predator-prey system with $\overline{\mathbf{u}}_0 = [0.5, 2]^\mathsf{T}$, $\epsilon = 0.2$, $T = 6$, using explicit Euler discretisation, i.e.

$$\dot{\mathbf{u}} = \mathbf{f}(\mathbf{u}) \;\text{ and }\; \mathbf{u}(0) = \mathbf{u}_0 \quad \longrightarrow \quad \mathbf{u}_{j+1} = \mathbf{u}_j + \Delta t\, \mathbf{f}(\mathbf{u}_j).$$

Study the discretisation and MC errors and compute confidence intervals.

(c) Implement also the antithetic estimator and compare the variance of the two estimators. How much is the variance reduced? Does this reduction depend on the selected tolerance TOL.

Multilevel Monte Carlo Methods

History

- The **multilevel Monte Carlo** method is a powerful new variance reduction technique.
- First ideas for high-dimensional quadrature by Heinrich, 2000.
- Independently discovered and popularised by Giles, 2007 in the context of stochastic DEs in mathematical finance.

Multilevel Monte Carlo Methods

History

- The **multilevel Monte Carlo** method is a powerful new variance reduction technique.
- First ideas for high-dimensional quadrature by Heinrich, 2000.
- Independently discovered and popularised by Giles, 2007 in the context of stochastic DEs in mathematical finance.
- First papers in the context of UQ:
  - Cliffe, Giles, RS, Teckentrup, 2011
  - Barth, Schwab, Zollinger, 2011

Multilevel Monte Carlo Methods

History

- The **multilevel Monte Carlo** method is a powerful new variance reduction technique.
- First ideas for high-dimensional quadrature by Heinrich, 2000.
- Independently discovered and popularised by Giles, 2007 in the context of stochastic DEs in mathematical finance.
- First papers in the context of UQ:
  - Cliffe, Giles, RS, Teckentrup, 2011
  - Barth, Schwab, Zollinger, 2011
- Stochastic simulation of discrete state systems (biology, chemistry) by Anderson, Higham, 2012

Next week

- . . .

- To estimate the expectation $\mathbf{E}[Q]$ of a quantity of interest Q, assume only approximations $Q_M \approx Q$ are computable, where $M \in \mathbb{N}$ denotes a discretization parameter (#time steps, #grid points, ...) and

$$\lim_{M\to\infty} \mathbf{E}[Q_M] = \mathbf{E}[Q].$$

Multilevel Monte Carlo Methods

Mean-square error – Standard MC

- To estimate the expectation $\mathbf{E}[Q]$ of a quantity of interest Q, assume only approximations $Q_M \approx Q$ are computable, where $M \in \mathbb{N}$ denotes a discretization parameter (#time steps, #grid points, ...) and

$$\lim_{M\to\infty} \mathbf{E}[Q_M] = \mathbf{E}[Q].$$

- More precisely, we assume the error in mean to converge at a rate $-\alpha$, i.e.,

$$|\mathbf{E}[Q_M - Q]| \lesssim M^{-\alpha}, \qquad \text{as } M \to \infty, \qquad \alpha > 0.$$

(in the predator-prey case $\alpha = 1$)

Multilevel Monte Carlo Methods

Mean-square error – Standard MC

- To estimate the expectation $\mathbf{E}[Q]$ of a quantity of interest Q, assume only approximations $Q_M \approx Q$ are computable, where $M \in \mathbb{N}$ denotes a discretization parameter (#time steps, #grid points, ...) and

$$\lim_{M\to\infty} \mathbf{E}[Q_M] = \mathbf{E}[Q].$$

- More precisely, we assume the error in mean to converge at a rate $-\alpha$, i.e.,

$$|\mathbf{E}[Q_M - Q]| \lesssim M^{-\alpha}, \qquad \text{as } M \to \infty, \qquad \alpha > 0.$$

(in the predator-prey case $\alpha = 1$)

- From **Exercise 2** we know that the mean square error (MSE) is

$$\mathbf{E}\left[\left(\widehat{Q}_{M,N} - \mathbf{E}[Q]\right)^2\right] = \frac{\mathbf{Var}[Q_M]}{N} + \left(\mathbf{E}[Q_M - Q]\right)^2.$$

- Denote by $\mathscr{C}(Q_M^{(k)})$ cost associated with computing one sample $Q_M^{(k)}$ (e.g. in terms of the number of floating-point operations required)

Multilevel Monte Carlo Methods

Cost scaling – Standard MC

- Denote by $\mathscr{C}(Q_M^{(k)})$ cost associated with computing one sample $Q_M^{(k)}$ (e.g. in terms of the number of floating-point operations required)
- Cost typcially grows linearly or with some power $\gamma \geq 1$ with M. We assume
$$\mathscr{C}(Q_M^{(k)}) \lesssim M^\gamma, \qquad \gamma \geq 1.$$
so that $\mathscr{C}(\widehat{Q}_{M,N}) \lesssim NM^\gamma$ (in the predator-prey case $\gamma = 1$).

- Denote by $\mathscr{C}(Q_M^{(k)})$ cost associated with computing one sample $Q_M^{(k)}$ (e.g. in terms of the number of floating-point operations required)

- Cost typcially grows linearly or with some power $\gamma \geq 1$ with M. We assume
$$\mathscr{C}(Q_M^{(k)}) \lesssim M^\gamma, \qquad \gamma \geq 1.$$
so that $\mathscr{C}(\widehat{Q}_{M,N}) \lesssim NM^\gamma$ (in the predator-prey case $\gamma = 1$).

- To balance the two MSE components, assume each ist bounded by $\frac{\mathrm{TOL}^2}{2}$, resulting in a total bound of TOL^2 for the MSE.

Multilevel Monte Carlo Methods

Cost scaling – Standard MC

- Denote by $\mathscr{C}(Q_M^{(k)})$ cost associated with computing one sample $Q_M^{(k)}$ (e.g. in terms of the number of floating-point operations required)

- Cost typcially grows linearly or with some power $\gamma \geq 1$ with M. We assume
$$\mathscr{C}(Q_M^{(k)}) \lesssim M^\gamma, \qquad \gamma \geq 1.$$
so that $\mathscr{C}(\widehat{Q}_{M,N}) \lesssim NM^\gamma$ (in the predator-prey case $\gamma = 1$).

- To balance the two MSE components, assume each ist bounded by $\frac{\mathrm{TOL}^2}{2}$, resulting in a total bound of TOL^2 for the MSE.

- This yields (since $Q_M \to Q$, we have $\mathbf{Var}[Q_M] \approx \mathbf{Var}[Q] = \text{constant}$)
$$N \geq 2\,\mathbf{Var}[Q_M]\mathrm{TOL}^{-2} \qquad \text{and} \qquad M \gtrsim \mathrm{TOL}^{-1/\alpha}.$$

Multilevel Monte Carlo Methods

Cost scaling – Standard MC

- Denote by $\mathscr{C}(Q_M^{(k)})$ cost associated with computing one sample $Q_M^{(k)}$ (e.g. in terms of the number of floating-point operations required)
- Cost typcially grows linearly or with some power $\gamma \geq 1$ with M. We assume
$$\mathscr{C}(Q_M^{(k)}) \lesssim M^\gamma, \qquad \gamma \geq 1.$$
so that $\mathscr{C}(\widehat{Q}_{M,N}) \lesssim NM^\gamma$ (in the predator-prey case $\gamma = 1$).
- To balance the two MSE components, assume each ist bounded by $\frac{\text{TOL}^2}{2}$, resulting in a total bound of TOL^2 for the MSE.
- This yields (since $Q_M \to Q$, we have $\mathbf{Var}[Q_M] \approx \mathbf{Var}[Q] = \text{constant}$)
$$N \geq 2\,\mathbf{Var}[Q_M]\text{TOL}^{-2} \qquad \text{and} \qquad M \gtrsim \text{TOL}^{-1/\alpha}.$$
- So the total cost of achieving a MSE $< \text{TOL}^2$ using a standard MC estimator is
$$\mathscr{C}(\widehat{Q}_{M,N}) \lesssim \text{TOL}^{-2-\gamma/\alpha}$$

- **Key idea:** use realisations of Q_M on a hierarchy of different levels, i.e., for different values $M_0, \ldots, M_L$ of the discretization parameter, and decompose

$$\mathbf{E}\left[Q_{M_L}\right] = \mathbf{E}\left[Q_{M_0}\right] + \sum_{\ell=1}^{L} \mathbf{E}\left[Q_{M_\ell} - Q_{M_{\ell-1}}\right] =: \sum_{\ell=0}^{L} \mathbf{E}\left[Y_\ell\right],$$

where $M_0 \in \mathbb{N}$, $M_\ell = sM_{\ell-1}$, for $\ell = 1, \ldots, L$, and $s \in \mathbb{N} \setminus \{1\}$.

- **Key idea:** use realisations of Q_M on a hierarchy of different levels, i.e., for different values $M_0, \dots, M_L$ of the discretization parameter, and decompose

$$\mathbf{E}\left[Q_{M_L}\right] = \mathbf{E}\left[Q_{M_0}\right] + \sum_{\ell=1}^{L} \mathbf{E}\left[Q_{M_\ell} - Q_{M_{\ell-1}}\right] =: \sum_{\ell=0}^{L} \mathbf{E}\left[Y_\ell\right],$$

where $M_0 \in \mathbb{N}$, $M_\ell = sM_{\ell-1}$, for $\ell = 1, \dots, L$, and $s \in \mathbb{N} \setminus \{1\}$.

- Given (unbiased) estimators $\{\widehat{Y}_\ell\}_{\ell=0}^{L}$ for $\mathbf{E}\left[Y_\ell\right]$, we refer to

$$\widehat{Q}_L^{\mathrm{ML}} := \sum_{\ell=0}^{L} \widehat{Y}_\ell$$

as a multilevel estimator for Q (today use standard MC on all levels).

- All expectations $\mathbf{E}\left[Y_\ell\right]$ sampled indep. $\Rightarrow$ **Var** $\widehat{Q}_L^{\mathrm{ML}} = \sum_{\ell=0}^{L}$ **Var** $\widehat{Y}_\ell$.

- If each $\widehat{Y}_\ell$ is itself a standard Monte Carlo estimator, i.e.,

$$\widehat{Y}_0 = \widehat{Y}_{0,N_0} := \frac{1}{N_0} \sum_{k=0}^{N_0} Q_{M_0}^{(k)}$$

and

$$\widehat{Y}_\ell = \widehat{Y}_{\ell,N_\ell} := \frac{1}{N_\ell} \sum_{k=0}^{N_\ell} \left(Q_{M_\ell}^{(k)} - Q_{M_{\ell-1}}^{(k)} \right), \qquad \ell = 1, \ldots, L,$$

one obtains a multilevel Monte Carlo estimator.

- The associated MSE then has the standard decomposition

$$\mathbf{E}\left[\left(\widehat{Q}_{L,\{N_\ell\}}^{\mathrm{ML}} - \mathbf{E}[Q] \right)^2 \right] = \sum_{\ell=0}^{L} \frac{\mathbf{Var}\, Y_\ell}{N_\ell} + \mathbf{E}\left[Q_{M_L} - Q \right]^2$$

into sample variance and bias (shown as for standard MC in Exerc. 2).

- Choose discretisation parameters and numbers of samples again to balance the terms in the MSE.
- The bias term is the same as for the standard MC estimator, leading again to a choice of $M_L = M \gtrsim \mathsf{TOL}^{-1/\alpha}$.

- Choose discretisation parameters and numbers of samples again to balance the terms in the MSE.
- The bias term is the same as for the standard MC estimator, leading again to a choice of $M_L = M \gtrsim \mathsf{TOL}^{-1/\alpha}$.
- **But why** do we get variance reduction – or rather lower cost for the same variance?

- Choose discretisation parameters and numbers of samples again to balance the terms in the MSE.
- The bias term is the same as for the standard MC estimator, leading again to a choice of $M_L = M \gtrsim \mathrm{TOL}^{-1/\alpha}$.
- **But why** do we get variance reduction – or rather lower cost for the same variance? Two reasons:

1. As we coarsen the problem, the cost per sample decays rapidly from level to level, with $\mathcal{O}(s^\gamma)$
2. Since $Q_M \to Q$, then $\mathbf{Var}[Y_\ell] = \mathbf{Var}[Q_{M_\ell} - Q_{M_{\ell-1}}] \to 0$ as $\ell \to \infty$, allowing for smaller and smaller sample sizes N_ℓ on finer and finer levels.

Multilevel Monte Carlo Methods

Optimal sample sizes

- The cost of the MLMC estimator is

$$\mathscr{C}(\widehat{Q}^{\mathrm{ML}}_{L,\{N_\ell\}}) = \sum_{\ell=0}^{L} N_\ell \mathscr{C}_\ell, \qquad \mathscr{C}_\ell := \mathscr{C}(Y_\ell^{(k)}).$$

Multilevel Monte Carlo Methods

Optimal sample sizes

- The cost of the MLMC estimator is

$$\mathscr{C}(\widehat{Q}^{\mathrm{ML}}_{L,\{N_\ell\}}) = \sum_{\ell=0}^{L} N_\ell \mathscr{C}_\ell, \qquad \mathscr{C}_\ell := \mathscr{C}(Y_\ell^{(k)}).$$

- Treating the N_ℓ as continuous variables, we can now minimise the cost of the MLMC estimator for a fixed variance

$$\sum_{\ell=0}^{L} \frac{\mathbf{Var}\, Y_\ell}{N_\ell} = \frac{\mathrm{TOL}^2}{2}$$

Multilevel Monte Carlo Methods

Optimal sample sizes

- The cost of the MLMC estimator is

$$\mathscr{C}(\widehat{Q}^{\mathrm{ML}}_{L,\{N_\ell\}}) = \sum_{\ell=0}^{L} N_\ell \mathscr{C}_\ell, \qquad \mathscr{C}_\ell := \mathscr{C}(Y_\ell^{(k)}).$$

- Treating the N_ℓ as continuous variables, we can now minimise the cost of the MLMC estimator for a fixed variance

$$\sum_{\ell=0}^{L} \frac{\mathbf{Var}\, Y_\ell}{N_\ell} = \frac{\mathrm{TOL}^2}{2}$$

- The solution to this constrained minimisation problem is

$$N_\ell \simeq \sqrt{\mathbf{Var}[Y_\ell]/\mathscr{C}_\ell}$$

with implied constant chosen such that the total variance is $\frac{\mathrm{TOL}^2}{2}$ (which leads to the constant $\frac{2}{\mathrm{TOL}^2} \sum_\ell \sqrt{\mathscr{C}_\ell \, \mathbf{Var}\, Y_\ell}$)

- This results in a total cost on level ℓ proportional to $\sqrt{\mathscr{C}_\ell \, \mathbf{Var}\, Y_\ell}$ and therefore

$$\mathscr{C}(\widehat{Q}^{\mathrm{ML}}_{L,\{N_\ell\}}) \le \frac{2}{\mathrm{TOL}^2} \left(\sum_{\ell=0}^{L} \sqrt{\mathscr{C}_\ell \, \mathbf{Var}\, Y_\ell} \right)^2$$

For comparison, the cost fo standard MC is $\mathscr{C}(\widehat{Q}_{M_L,N}) = \frac{2}{\mathrm{TOL}^2} \mathscr{C}_L \, \mathbf{Var}[Q_{M_L}]$.

- This results in a total cost on level ℓ proportional to $\sqrt{\mathscr{C}_\ell \, \mathbf{Var}\, Y_\ell}$ and therefore

$$\mathscr{C}(\widehat{Q}^{\mathrm{ML}}_{L,\{N_\ell\}}) \leq \frac{2}{\mathrm{TOL}^2} \left(\sum_{\ell=0}^{L} \sqrt{\mathscr{C}_\ell \, \mathbf{Var}\, Y_\ell} \right)^2$$

  For comparison, the cost fo standard MC is $\mathscr{C}(\widehat{Q}_{M_L,N}) = \frac{2}{\mathrm{TOL}^2} \mathscr{C}_L \, \mathbf{Var}[Q_{M_L}]$.

- If $\mathbf{Var}\, Y_\ell$ decays faster than $\mathscr{C}_\ell$ increases, the cost on level $\ell = 0$ dominates. Since $\mathbf{Var}[Q_{M_0}] \approx \mathbf{Var}[Q_{M_L}]$, the cost ratio of MLMC to MC estimation is then approximately

$$\mathscr{C}_0 / \mathscr{C}_L \eqsim s^{-L\gamma}$$

- This results in a total cost on level ℓ proportional to $\sqrt{\mathscr{C}_\ell \, \textbf{Var}\, Y_\ell}$ and therefore
$$\mathscr{C}(\widehat{Q}^{\mathrm{ML}}_{L,\{N_\ell\}}) \le \frac{2}{\mathrm{TOL}^2} \left(\sum_{\ell=0}^{L} \sqrt{\mathscr{C}_\ell \, \textbf{Var}\, Y_\ell} \right)^2$$
For comparison, the cost fo standard MC is $\mathscr{C}(\widehat{Q}_{M_L,N}) = \frac{2}{\mathrm{TOL}^2} \mathscr{C}_L \, \textbf{Var}[Q_{M_L}]$.

- If $\textbf{Var}\, Y_\ell$ decays faster than $\mathscr{C}_\ell$ increases, the cost on level $\ell = 0$ dominates. Since $\textbf{Var}[Q_{M_0}] \approx \textbf{Var}[Q_{M_L}]$, the cost ratio of MLMC to MC estimation is then approximately
$$\mathscr{C}_0 / \mathscr{C}_L \eqsim s^{-L\gamma}$$

- If $\mathscr{C}_\ell$ increases faster than $\textbf{Var}\, Y_\ell$ decays, then the cost on level $\ell = L$ dominates, and then the cost ratio is approximately
$$\textbf{Var}[Y_L] / \textbf{Var}[Q_{M_L}] \;\eqsim\; \mathrm{TOL}^2$$
(provided $\textbf{E}\left[(Q - Q_L)^2\right] \eqsim (\textbf{E}\left[Q - Q_L\right])^2$, which is problem dependent).

Theorem

Let $TOL < \exp(-1)$ *and assume there are constants* $\alpha, \beta, \gamma > 0$ *such that* $\alpha \geq \min\{\beta, \gamma\}$ *and, for all* $\ell = 0, \ldots, L$,

(M1) $|\mathbf{E}[Q_{M_\ell}] - \mathbf{E}[Q]| \lesssim M_\ell^{-\alpha}$,

(M2) $\mathbf{Var}[\widehat{Y}_\ell] \lesssim N_\ell^{-1} M_\ell^{-\beta}$,

(M3) $\mathscr{C}(\widehat{Y}_\ell) \lesssim N_\ell M_\ell^{\gamma}$.

Then there are L *and* $\{N_\ell\}_{\ell=0}^L$ *s.t.* $\mathbf{E}\left[\left(\widehat{Q}^{ML}_{L,\{N_\ell\}} - \mathbf{E}[Q]\right)^2\right] \leq TOL^2$ *and*

$$\mathscr{C}(\widehat{Q}^{ML}_{L,\{N_\ell\}}) \lesssim \begin{cases} TOL^{-2}, & \text{if } \beta > \gamma, \\ TOL^{-2}\,|\log TOL|^2, & \text{if } \beta = \gamma, \\ TOL^{-2-(\gamma-\beta)/\alpha}, & \text{if } \beta < \gamma. \end{cases}$$

Exercise 4

(a) Solve the constrained minimisation problem on Slide 30 to find the otimal numbers of samples on each level. (*Hint:* Use a Lagrange multiplier approach to include the constraint and then consider the first-order optimality constraints to find the minimum.)

(b) Proof the complexity theorem.

- The following MLMC algorithm computes the optimal values of L and N_ℓ adaptively using (unbiased) sample averages ($\widehat{Y}_\ell$) and sample variances (s_ℓ^2) of Y_ℓ.
- The sample variances can be used directly in the MC error estimates.

- The following MLMC algorithm computes the optimal values of L and N_ℓ adaptively using (unbiased) sample averages ($\widehat{Y}_\ell$) and sample variances (s_ℓ^2) of Y_ℓ.
- The sample variances can be used directly in the MC error estimates.
- To bound the bias error, we assume there exists an $M^\star > 0$ such that the error decay in $\left|\mathbf{E}\left[Q_M - Q\right]\right|$ is monotonic for $M \geq M^\star$ and satisfies $\left|\mathbf{E}\left[Q_M - Q\right]\right| \eqsim M^{-\alpha}$.

- The following MLMC algorithm computes the optimal values of L and N_ℓ adaptively using (unbiased) sample averages ($\widehat{Y}_\ell$) and sample variances (s_ℓ^2) of Y_ℓ.
- The sample variances can be used directly in the MC error estimates.
- To bound the bias error, we assume there exists an $M^\star > 0$ such that the error decay in $|\mathbf{E}[Q_M - Q]|$ is monotonic for $M \geq M^\star$ and satisfies $|\mathbf{E}[Q_M - Q]| \eqsim M^{-\alpha}$.
- This ensures (via the inverse triangle inequality) that

$$|\mathbf{E}[Q_{M_\ell} - Q]| \leq \frac{1}{s^\alpha - 1} \widehat{Y}_\ell$$

  and gives a computable error estimator on level L to determine whether h_L is sufficiently small or whether L needs to be increased.

Adaptive MLMC Algorithm

1. Set TOL, $L = 1$ and $N_0 = N_1 = N_{\texttt{Init}}$.
2. For all levels $\ell = 0, \ldots, L$ do
   a. Compute new samples $Y_\ell^{(k)}$ until there are N_ℓ.
   b. Compute $\widehat{Y}_\ell$ and s_ℓ^2, and estimate $\mathscr{C}_\ell$.
3. Update estimates for N_ℓ using the formula on Slide 30 and if $\widehat{Y}_L > \frac{s^\alpha - 1}{\sqrt{2}}\mathrm{TOL}$, increase $L \to L+1$ and set $N_L = N_{\texttt{Init}}$.
4. If there is no change

   Go to 5.

   Else

   Return to 2.
5. Set $\widehat{Q}^{\mathrm{ML}}_{L,\{N_\ell\}} = \sum_{\ell=0}^{L} \widehat{Y}_\ell$.

Exercise 5

(a) Implement the multilevel MC method for the predator-prey problem. Choose M_0 not too small to avoid stability problems with the explicit Euler method. Compare the cost to achieve a certain tolerance TOL for the mean square error (in terms of floating point operations) against your other two implementations (standard MC and antithetic MC estimator). How big is the computational gain?

(b) Recall that $\alpha = \gamma = 1$ in that case. Verify this with your code. Compute $\mathbf{Var}[\widehat{Y}_\ell]$ and $\mathbf{Var}[\widehat{Q}_{M_\ell}]$ for a range of values of ℓ and M_0. What is the numerically observed rate β? Prove this theoretically.

(c) Can you think of any further enhancements of your code?

Exercise 6

Exercise 6

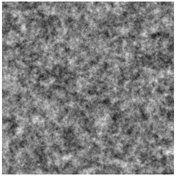
(a) Think of a UQ question in your field of research and try to formulate a simple model problem that encapsulates the essential question. What type of uncertainty is it? How could you model it within your problem? Can you formulate a Monte Carlo simulation to estimate the uncertainties in a derived quantity of interest from your model? Are any of the variance reduction techniques we discussed applicable? Is there a natural model hierarchy that could be exploited in a multilevel algorithm?

(b) Implement a simple Monte Carlo code to quantify the uncertainties. If your problem has natural model hierarchies and allows to couple them, try to estimate $\mathbf{Var}[\widehat{Y}_\ell]$ and $\mathbf{Var}[\widehat{Q}_{M_\ell}]$ in the same way as we did above to check whether multilevel Monte Carlo would be beneficial.

(c) Implement a multilevel MC method for your problem. Do you achieve the gains that were predicted in (b)?

$\boxed{\text{uncertain } k} \rightarrow \boxed{\begin{array}{lrcl} \textbf{Darcy's Law:} & \vec{q} + k\nabla p & = & f \\ \textbf{Incompressibility:} & \nabla \cdot \vec{q} & = & 0 \end{array}} \rightarrow \boxed{\text{uncertain } p,\ \vec{q}}$

Typical simplified model for k:



- $\log k(x,\omega)$ = isotropic, scalar **Gaussian** e.g. with exp. covariance ($\nu = \frac{1}{2}$): $R(x,y) := \sigma^2 \exp\left(-\frac{\|x-y\|}{\lambda}\right)$
- KL expansion: $\log k(x,\omega) \approx \sum_{j=1}^{J} \sqrt{\mu_j}\phi_j(x) Y_j(\omega)$
  with $Y_j(\omega)$ i.i.d. $N(0,1)$
- **FE discretisation**: $A(\omega)\,\mathbf{P}(\omega) = \mathbf{b}(\omega)$ $\boxed{\text{later}}$
- QoI $Q(\omega)$, e.g., particle travel time from repository to boundary

$D = (0,1)^2$, unconditioned KL expansion, $Q = \| - k \frac{\partial p}{\partial x_1} \|_{L^1(D)}$
using mixed FEs and the AMG solver `amg1r5` [Ruge, Stüben, 1992]

- Num. observed FE-error: $\approx \mathcal{O}(h^{-3/4}) \eqsim \mathcal{O}(M_h^{-3/8}) \;\Rightarrow\; \alpha \approx 3/8$
- Num. observed cost/sample: $\approx \mathcal{O}(h^{-d}) \eqsim \mathcal{O}(M_h) \;\Rightarrow\; \gamma \approx 1$
- **Total cost** to get RMSE $\mathcal{O}(\mathrm{TOL})$: $\approx \mathcal{O}(\mathrm{TOL}^{-14/3})$
  to get error reduction by a factor 2 $\rightarrow$ cost grows by a factor 25!

Case 1: $\sigma^2 = 1$, $\lambda = 0.3$, $\nu = 0.5$

TOL	h^{-1}	N_h	Cost
0.01	129	1.4×10^4	21 min
0.002	1025	3.5×10^5	30 days

Case 2: $\sigma^2 = 3$, $\lambda = 0.1$, $\nu = 0.5$

TOL	h^{-1}	N_h	Cost
0.01	513	8.5×10^3	4 h
0.002	**Prohibitively large!!**		

(actual numbers & CPU times on a cluster of 2GHz Intel T7300 processors)

- Assuming optimal AMG solver (i.e. $\gamma \approx 1$) and $\beta \approx 2\alpha$. Then for $\alpha \approx 3/4d^{-1}$ (as in the example above) the **cost** in $\mathbb{R}^d$ is

d	**MC**	**MLMC**	per sample
1	$\mathcal{O}(\varepsilon^{-10/3})$	$\mathcal{O}(\varepsilon^{-2})$	$\mathcal{O}(\varepsilon^{-4/3})$
2	$\mathcal{O}(\varepsilon^{-14/3})$	$\mathcal{O}(\varepsilon^{-8/3})$	$\mathcal{O}(\varepsilon^{-8/3})$
3	$\mathcal{O}(\varepsilon^{-6})$	$\mathcal{O}(\varepsilon^{-4})$	$\mathcal{O}(\varepsilon^{-4})$

- Assuming optimal AMG solver (i.e. $\gamma \approx 1$) and $\beta \approx 2\alpha$. Then for $\alpha \approx 3/4d^{-1}$ (as in the example above) the **cost** in $\mathbb{R}^d$ is

d	**MC**	**MLMC**	per sample
1	$\mathcal{O}(\varepsilon^{-10/3})$	$\mathcal{O}(\varepsilon^{-2})$	$\mathcal{O}(\varepsilon^{-4/3})$
2	$\mathcal{O}(\varepsilon^{-14/3})$	$\mathcal{O}(\varepsilon^{-8/3})$	$\mathcal{O}(\varepsilon^{-8/3})$
3	$\mathcal{O}(\varepsilon^{-6})$	$\mathcal{O}(\varepsilon^{-4})$	$\mathcal{O}(\varepsilon^{-4})$

- Assuming optimal AMG solver (i.e. $\gamma \approx 1$) and $\beta \approx 2\alpha$. Then for $\alpha \approx 3/4d^{-1}$ (as in the example above) the **cost** in $\mathbb{R}^d$ is

d	**MC**	**MLMC**	per sample
1	$\mathcal{O}(\varepsilon^{-10/3})$	$\mathcal{O}(\varepsilon^{-2})$	$\mathcal{O}(\varepsilon^{-4/3})$
2	$\mathcal{O}(\varepsilon^{-14/3})$	$\mathcal{O}(\varepsilon^{-8/3})$	$\mathcal{O}(\varepsilon^{-8/3})$
3	$\mathcal{O}(\varepsilon^{-6})$	$\mathcal{O}(\varepsilon^{-4})$	$\mathcal{O}(\varepsilon^{-4})$

Optimality (for $\gamma > \beta = 2\alpha$)

MLMC cost is asymptotically the same as **one deterministic solve** to accuracy ε for $d > 1$, i.e. $\mathcal{O}(\varepsilon^{-\gamma/\alpha})$ **!!** (only true for rough problems!)

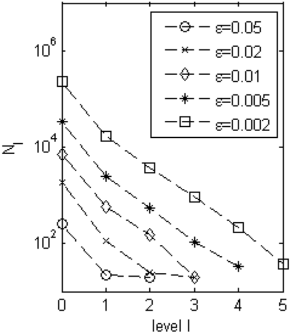
- Assuming optimal AMG solver (i.e. $\gamma \approx 1$) and $\beta \approx 2\alpha$. Then for $\alpha \approx 3/4d^{-1}$ (as in the example above) the **cost** in $\mathbb{R}^d$ is

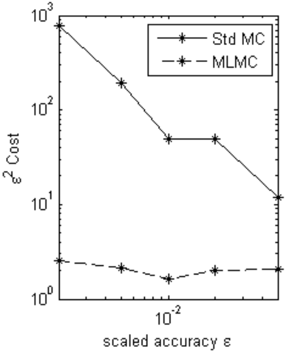
d	**MC**	**MLMC**	per sample
1	$\mathcal{O}(\varepsilon^{-10/3})$	$\mathcal{O}(\varepsilon^{-2})$	$\mathcal{O}(\varepsilon^{-4/3})$
2	$\mathcal{O}(\varepsilon^{-14/3})$	$\mathcal{O}(\varepsilon^{-8/3})$	$\mathcal{O}(\varepsilon^{-8/3})$
3	$\mathcal{O}(\varepsilon^{-6})$	$\mathcal{O}(\varepsilon^{-4})$	$\mathcal{O}(\varepsilon^{-4})$

Optimality (for $\gamma > \beta = 2\alpha$)

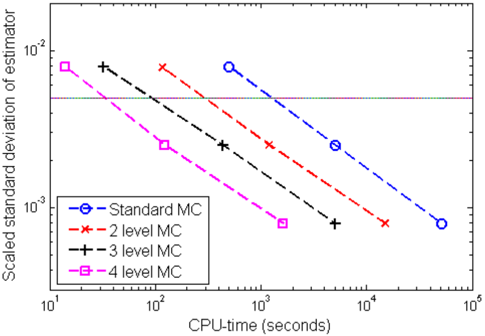
MLMC cost is asymptotically the same as **one deterministic solve** to accuracy ε for $d > 1$, i.e. $\mathcal{O}(\varepsilon^{-\gamma/\alpha})$ **!!** (only true for rough problems!)

Can we achieve such huge gains in practice?





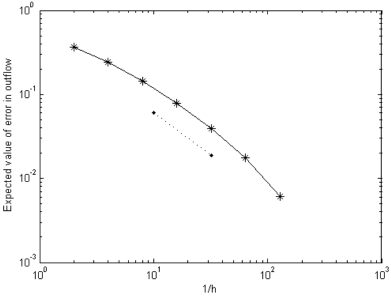
$$\nu = \frac{1}{2}, \quad \sigma^2 = 1, \quad \lambda = 0.3, \quad h_0 = \frac{1}{8}$$

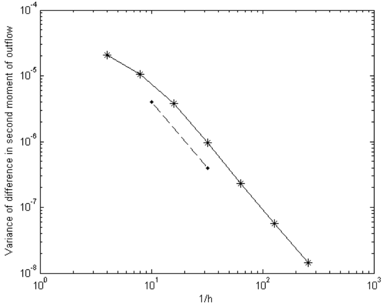


$h_L = 1/256$ (solid line is FE-error)

`Matlab` implementation on 3GHz Intel Core 2 Duo E8400 processor, 3.2GByte RAM, with **sparse direct solver**, i.e. $\gamma \approx 1.2$

Verifying Assumptions in MLMC Complexity Theorem: $\nu = 1/2$, $\sigma^2 = 1$, $\lambda = 0.3$





$$\left|\mathbb{E}\left[\mathscr{G}^{(1)}(p) - \mathscr{G}^{(1)}(p_h)\right]\right|$$

where $\mathscr{G}^{(1)}(p) := L_\omega(\Psi) - b_\omega(\Psi, v)$ given $\Psi(x) = x$ (outflow on right).

$$\mathbb{V}\left[\mathscr{G}^{(2)}(p_h) - \mathscr{G}^{(2)}(p_{2h})\right]$$

where $\mathscr{G}^{(2)}(p) := \left(\frac{1}{|D^*|}\int_{D^*} p(x)\,\mathrm{d}x\right)^2$ (i.e. 2nd moment of p over small patch)

$$\Longrightarrow \quad \alpha = 1 \text{ and } \beta = 2$$

Can be proved rigorously for the lognormal case! (more details Tue/Wed)