Monte Carlo and Multilevel/Multifidelity Sampling Strategies for Forward UQ

SAND2017-11674PE

Gianluca Geraci

Sandia National Laboratories, Albuquerque

UQ for Computational Science and Engineering:
Monte Carlo and spectral stochastic methods

PART 1
25th October 2017

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energys National Nuclear Security Administration under contract DE-NA-0003525.

Monte Carlo

A brief of its history (1/2)

Halton (1970): *representing the solution of a problem as a parameter of a hypothetical population, and using a random sequence of numbers to construct a sample of the population, from which statistical estimates of the parameter can be obtained.*

- One of the first documented MC experiments is Buffon's needle experiment which Laplace (1812) suggested can be used to approximate π (Johansen and Evers, 2007)

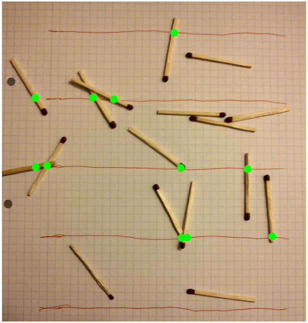


Figure: Buffon's needle experiment based on 17 throws. (Source: Wikipedia)

$$\pi \approx \frac{2Nl}{Pt},$$

where

- N: number of throws
- l: length of the needles
- P: number of needle crossing the lines
- t: distance between the lines

MONTE CARLO

A BRIEF OF ITS HISTORY (2/2) – *Los Alamos Science Special Issue 1987, by N. Metropolis*

Around 1940:

- ENIAC: first electronic computer at the University of Pennsylvania

[...] Stan's (Stanislaw Ulam) extensive mathematical background made him aware that statistical sampling techniques had fallen into desuetude because of the length and tediousness of the calculations. But with this miraculous development of the ENIAC, [...] it occurred to him that statistical techniques should be resuscitated, and he discussed this idea with von Neumann. Thus was triggered the spark that led to the Monte Carlo method.

- The name: *Ulam had a uncle who would borrow money from relatives because he "just had to go to Monte Carlo"*

Sampling Methods
Role in UQ

- There are several applications for the MC method
- In Uncertainty Quantification (UQ) we are often concerned with the computation of a the expected value of a function (or higher moments)

$$\mathbb{E}\left[f(\xi)\right] = \int_{\Xi} f(\xi)p(\xi)\mathrm{d}\xi$$

- Therefore one of the tasks to be performed in UQ is the quadrature in (very often) high-dimension ($\Xi \subset \mathbb{R}^d$)

UQ is a much richer area than 'just' numerical quadrature, but nevertheless this is an important task

Monte Carlo – Key elements

Main ingredients

From the random generator to the statistical estimator

Each **Monte Carlo** method is based upon three main steps:

- Pre-processing: generation of random numbers
- Evaluation step: Computation of the Quantity of Interest from the computational code
- Post-processing: Estimator and confidence interval evaluation

Pre-processing
Random Number Generator

- A random number generator is required for each Monte Carlo simulation
- Random number generation requires two main stages
  - Generation of independent random variables $\mathcal{U}(0, 1)$
  - Conversion of the RVs to desired distribution

Pre-processing
Random Number Generator

- A random number generator is required for each Monte Carlo simulation
- Random number generation requires two main stages
  - Generation of independent random variables $\mathcal{U}(0, 1)$
  - Conversion of the RVs to desired distribution

(Pseudo-) random generators use DETERMINISTIC algorithm to generate only APPARENTLY RANDOM numbers

Pre-processing
Random Number Generator

- A random number generator is required for each Monte Carlo simulation
- Random number generation requires two main stages
  - Generation of independent random variables $\mathcal{U}(0,1)$
  - Conversion of the RVs to desired distribution

(Pseudo-) random generators use DETERMINISTIC algorithm to generate only APPARENTLY RANDOM numbers

Properties for a good random generator

- Several statistical tests exist to measure randomness, therefore reliable software has been verified against them
- A long period is needed before the sequence repeats (at least 2^{40} is required)
- A control-based *seed* is provided to skip to an arbitrary point of the sequence (useful in parallel applications)

Pre-processing
Random Number Generator

- A random number generator is required for each Monte Carlo simulation
- Random number generation requires two main stages
  - Generation of independent random variables $\mathcal{U}(0, 1)$
  - Conversion of the RVs to desired distribution

(Pseudo-) random generators use DETERMINISTIC algorithm to generate only APPARENTLY RANDOM numbers

Properties for a good random generator

- Several statistical tests exist to measure randomness, therefore reliable software has been verified against them
- A long period is needed before the sequence repeats (at least 2^{40} is required)
- A control-based *seed* is provided to skip to an arbitrary point of the sequence (useful in parallel applications)

Bottom line...

- do not use your own generator, but use reputable sources
- For instance, Intel Math Kernel Library (MKL) are free for all

Pre-processing
Variable transformation

- Random generators produce uniform RV $\mathcal{U}(0,1)$, but usually we need other distributions
- Let's assume that the cumulative distribution function F_Ξ for a variable ξ is available

$$F_\Xi(\xi) = P(\Xi \leq \xi)$$

- The random generator produces $U \sim \mathcal{U}(0,1)$, *i.e.* $F_U(u) = u$
- We want to determine the function $g(U)$ which gives $\Xi = g(U)$ with cdf $F_\Xi(\xi)$
- We write the cdf for $F_\xi(\xi)$

$$F_\Xi(\xi) = \mathrm{P}(\Xi \leq \xi) = \mathrm{P}(g(U) \leq \xi)$$

- We also assume:
- The function g is invertible on its range
- The function g is strictly increasing (only for simplicity)

$$F_\Xi(\xi) = \mathrm{P}(g(U) \leq \xi) = \mathrm{P}(U \leq g^{-1}(\xi)) = F_U(g^{-1}(\xi)) = g^{-1}(\xi)$$

- Finally we can choose $g^{-1}(\xi) = F_\Xi(\xi)$, *i.e.* $\Xi = F_\Xi^{-1}(U)$ in order to get the desired distribution

Statistical Estimator
Evaluations step

Let consider a random variable Q:

$$\hat{Q}_N^{\mathrm{MC}} = \frac{1}{N} \sum_{i=1}^{N} Q^{(i)}$$

Statistical Estimator
Evaluations step

Let consider a random variable Q:

$$\hat{Q}_N^{\mathrm{MC}} = \frac{1}{N}\sum_{i=1}^{N} Q^{(i)}$$

Two main features

- Unbiased (for each choice of N!): $\mathbb{E}\left[\hat{Q}_N^{MC}\right] = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[Q^{(i)}\right] = \mathbb{E}\left[Q\right]$
- Convergent (Strong law of large numbers): $\lim_{N\to\infty}\hat{Q}_N^{\mathrm{MC}} = \mathbb{E}\left[Q\right]$ a.s.

Statistical Estimator
Evaluations step

Let consider a random variable Q:

$$\hat{Q}_N^{\mathrm{MC}} = \frac{1}{N} \sum_{i=1}^{N} Q^{(i)}$$

Two main features

- Unbiased (for each choice of N!): $\mathbb{E}\left[\hat{Q}_N^{MC}\right] = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}\left[Q^{(i)}\right] = \mathbb{E}\left[Q\right]$
- Convergent (Strong law of large numbers): $\lim_{N\to\infty} \hat{Q}_N^{\mathrm{MC}} = \mathbb{E}\left[Q\right]$ a.s.

Main mathematical tool used for the analysis is the Central Limit Theorem (CLT)

- Let's define the error $e_N = \mathbb{E}\left[Q\right] - \hat{Q}_N^{MC}$
- Let's assume $\mathbb{V}ar(Q)$ is finite, then for $N \to \infty$

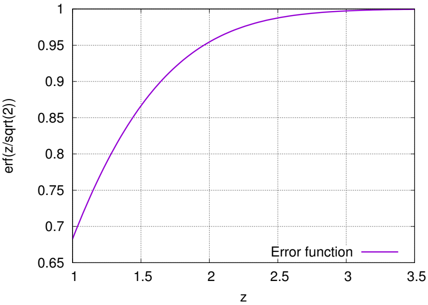
$$\frac{e_N}{\mathbb{V}ar\left(\hat{Q}_N^{\mathrm{MC}}\right)} = N^{1/2} \frac{e_N}{\mathbb{V}ar^{1/2}(Q)} \sim \mathcal{N}(0,1)$$

Central Limit Theorem
Confidence interval

CLT is the fundamental result that enable us to obtain a confidence interval for MC

- $\mathrm{P}\left(N^{1/2}\frac{e_N}{\mathbb{V}ar^{1/2}(Q)} \leq z\right) = F_Z(z)$, for $Z \sim \mathcal{N}(0,1)$
- $F_Z(z) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{z}{\sqrt{2}}\right)\right)$
- We want to control the probability of $\left|N^{1/2}\frac{e_N}{\mathbb{V}ar^{1/2}(Q)}\right|$, therefore

$$\mathrm{P}\left(\left|N^{1/2}\frac{e_N}{\mathbb{V}ar^{1/2}(Q)}\right| \leq z\right) = 1 - 2F_Z(z) = \mathrm{erf}\left(\frac{z}{\sqrt{2}}\right)$$



z	$1 - 2F_Z(z)$
1	0.683
2	0.954
3	0.997

Monte Carlo
Target accuracy

We can use the distribution of e_N to estimate the number of simulations required.

- Let's assume we want an estimator accurate at the 99.7% with error $e_N = \varepsilon$
- We need to select $z = 3$ (from the previous table)
- $N = 9\dfrac{\mathbb{V}ar(Q)}{\varepsilon^2}$

Few additional comments:

- The number of samples scales as ε^2, *i.e.* one order of increased accuracy is obtained with 100 times more samples
- Error is not a function of the dimension ($e_N \propto N^{-1/2}$)
- Error is not a function of the regularity of the quantity Q
- On the contrary the error for a composite Simpson's rule ($[0, 1]$) is bounded by

$$\frac{h^4}{180} \max_{x \in [0,1]} f^{(4)}(x), \quad \text{therefore} \quad e_N \propto N_{1D}^{-4} = N^{-4/d}$$

(MC integration is competitive for $d > 8$ w.r.t. Simpson's rule)

Monte Carlo

The role of the estimator variance

In summary we have seen so far:

- $e_N \sim \mathbb{V}ar\left(\hat{Q}_N^{\mathrm{MC}}\right) \mathcal{N}(0,1)$
- $e_N \propto N^{-1/2}$ and (numerical cost) is $\mathcal{C}^{\mathrm{MC}} \propto N$, therefore $\mathcal{C}^{\mathrm{MC}} \propto e_N^{-2}$

Monte Carlo

The role of the estimator variance

In summary we have seen so far:

- $e_N \sim \mathbb{V}ar\left(\hat{Q}_N^{\mathrm{MC}}\right)\, \mathcal{N}(0,1)$
- $e_N \propto N^{-1/2}$ and (numerical cost) is $\mathcal{C}^{\mathrm{MC}} \propto N$, therefore $\mathcal{C}^{\mathrm{MC}} \propto e_N^{-2}$

Variance of a MC estimator is

$$\begin{aligned}\mathbb{V}ar\left(\hat{Q}_N^{\mathrm{MC}}\right) &= \mathbb{V}ar\left(\frac{1}{N}\sum_{i=1}^{j} Q^{(i)}\right)\\ &= \frac{1}{N^2}\sum_{i=1}^{j}\mathbb{V}ar(Q)\\ &= \frac{1}{N}\mathbb{V}ar(Q)\end{aligned}$$

Monte Carlo
The role of the estimator variance

In summary we have seen so far:

- $e_N \sim \mathbb{V}ar\left(\hat{Q}_N^{\mathrm{MC}}\right) \mathcal{N}(0,1)$
- $e_N \propto N^{-1/2}$ and (numerical cost) is $\mathcal{C}^{\mathrm{MC}} \propto N$, therefore $\mathcal{C}^{\mathrm{MC}} \propto e_N^{-2}$

Variance of a MC estimator is

$$\begin{aligned}\mathbb{V}ar\left(\hat{Q}_N^{\mathrm{MC}}\right) &= \mathbb{V}ar\left(\frac{1}{N}\sum_{i=1}^{j} Q^{(i)}\right) \\ &= \frac{1}{N^2}\sum_{i=1}^{j} \mathbb{V}ar(Q) \\ &= \frac{1}{N}\mathbb{V}ar(Q)\end{aligned}$$

- So far we discussed what happens for increasing values of N, however the error can be reduced by manipulating $\mathbb{V}ar(Q)$
- This is the main idea behind the so-called *variance reduction* strategies

Variance reduction
An (incomplete list)

In the statistical literature several *variance reduction* techniques exist:

- Importance sampling
  - Very useful when the main contribution to $\mathbb{E}[Q]$ comes from rare events
- Stratified sampling
  - Very effective in 1D, not clear how to extend to multiple dimensions
- Latin hypercube
  - Effective if the function can be decomposed into a sum of 1D functions
- (Randomized) quasi-MC
  - Possibly provides better error than MC, but need to be randomized to get the confidence interval
- Control variate (more about it later...)

Monte Carlo – Extension to PDE with random input

Monte Carlo simulation

Introducing the spatial discretization

Problem statement: We are interested in the statistics of a functional (linear or non-linear) Q_M of the solution $\mathbf{u}_M$

$$Q_M = \mathcal{G}(\mathbf{u}_M) \quad \rightarrow \quad \mathbb{E}\left[Q_M\right]$$

- M is (related to) the number of spatial degrees of freedom
- $\mathbb{E}\left[Q_M\right] \xrightarrow{M \to \infty} \mathbb{E}\left[Q\right]$ for some RV $Q : \Omega \to \mathbb{R}$

$$\hat{Q}_{M,N}^{MC} \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^{N} Q_M^{(i)},$$

Looking at the Mean Square Error (MSE):

$$\begin{aligned}
\mathbb{E}\left[\left(\hat{Q}_{M,N}^{MC} - \mathbb{E}\left[Q\right]\right)^2\right] &= \mathbb{E}\left[\left(\hat{Q}_{M,N}^{MC} - \mathbb{E}\left[Q_M\right] + \mathbb{E}\left[Q_M\right] - \mathbb{E}\left[Q\right]\right)^2\right] \\
&= \mathbb{E}\left[\left(\hat{Q}_{M,N}^{MC} - \mathbb{E}\left[Q_M\right]\right)^2\right] + 2\mathbb{E}\left[\left(\hat{Q}_{M,N}^{MC} - \mathbb{E}\left[Q_M\right]\right)\left(\mathbb{E}\left[Q_M\right] - \mathbb{E}\left[Q\right]\right)\right] \\
&+ \mathbb{E}\left[\left(\mathbb{E}\left[Q_M\right] - \mathbb{E}\left[Q\right]\right)^2\right] \\
&= \mathbb{V}ar\left(\hat{Q}_{M,N}^{MC}\right) + \left(\mathbb{E}\left[Q_M - Q\right]\right)^2
\end{aligned}$$

Monte Carlo simulation

Introducing the spatial discretization (MSE)

Two sources of error in the Mean Square Error:

$$\mathbb{E}\left[(\hat{Q}^{MC}_{M,N} - \mathbb{E}[Q])^2\right] = \mathbb{V}ar\left(\hat{Q}^{MC}_{M,N}\right) + \left(\mathbb{E}[Q_M - Q]\right)^2$$

- Sampling error: replacing the expected value by a (finite) sample average

$$\mathbb{V}ar\left(\hat{Q}^{MC}_{M,N}\right) = \frac{\mathbb{V}ar(Q)}{N}$$

- Spatial discretization: finite resolution implies $\mathbb{E}[Q_M - Q] = \mathcal{O}(M^{-\alpha})$

To get accuracy $\varepsilon^2 = \varepsilon^2/2 + \varepsilon^2/2$**:** $N = \mathcal{O}\left(\varepsilon^{-2}\right)$ **AND** $M = \mathcal{O}\left(\varepsilon^{-1/\alpha}\right)$

Accurate estimation ⇒ **Large number** of samples at high (spatial) resolution

Monte Carlo simulation

Introducing the spatial discretization (MSE)

Two sources of error in the Mean Square Error:

$$\mathbb{E}\left[(\hat{Q}^{MC}_{M,N} - \mathbb{E}[Q])^2\right] = \mathbb{V}ar\left(\hat{Q}^{MC}_{M,N}\right) + \left(\mathbb{E}[Q_M - Q]\right)^2$$

- Sampling error: replacing the expected value by a (finite) sample average

$$\mathbb{V}ar\left(\hat{Q}^{MC}_{M,N}\right) = \frac{\mathbb{V}ar(Q)}{N}$$

- Spatial discretization: finite resolution implies $\mathbb{E}[Q_M - Q] = \mathcal{O}(M^{-\alpha})$

To get accuracy $\varepsilon^2 = \varepsilon^2/2 + \varepsilon^2/2$**:** $N = \mathcal{O}\left(\varepsilon^{-2}\right)$ **AND** $M = \mathcal{O}\left(\varepsilon^{-1/\alpha}\right)$

Accurate estimation $\Rightarrow$ **Large number** of samples at high (spatial) resolution

In the remaining part of the lectures we will see how to reduce the MSE by using the resolution index M and reducing the variance $\mathbb{V}ar(Q)$

Multilevel/Multifidelity – Generalities

Multifidelity/Multilevel approaches

Computational cost reduction Vs Decreased Accuracy

- So far we have discussed only *single fidelity* sampling approaches (variance reduction strategies)
- However in engineering practice hierarchies of models are ubiquitous
- Very often low-fidelity models are order of magnitude cheaper than high-fidelity one (Potential for high numerical cost saving)

Few examples:

- Physical models (Laminar/Turbulent, Reacting/non-reacting, viscous/inviscid...)
- Numerical methods (high/low order, Euler/RANS/LES, *etc*...)
- Numerical discretization (fine/coarse mesh...)
- Quality of statistics (long/short time history for turbulent flow...)

Common features:

- Increasing the model level/fidelity the quality of the solution improves (numerical solution closer to the truth)
- Increasing the level/fidelity the numerical cost also increases

Multifidelity/Multilevel approaches
Few comments

- Low-fidelity models are potentially useful or their more efficient computational cost
- However, their accuracy is reduced w.r.t. high-fidelity models

$$\mathbb{E}\left[(\hat{Q}^{MC}_{M,N} - \mathbb{E}[Q])^2\right] = \mathbb{V}ar\left(\hat{Q}^{MC}_{M,N}\right) + \left(\mathbb{E}[Q_M - Q]\right)^2$$

Few additional comments:

- If we want to control the MSE it is important to consider the term $\mathbb{E}[Q_M - Q]$
- Therefore, we can distinguish two categories of low-fidelity models:
  - Multilevel: if $\mathbb{E}[Q_M - Q] \to 0$ for $M \to \infty$
  - Multifidelity: when (roughly speaking) the distance between the two models is not reduced increasing M

Examples:

- Level: spatial/time resolution, Fourier modes etc.
- Fidelity: LES/RANS, laminar/turbulent, viscous/inviscid etc.

Control variate
Pivotal role

A Control Variate MC estimator (function G with $\mathbb{E}[G]$ known)

$$\hat{Q}_N^{MCCV} = \hat{Q}_N^{MC} - \beta \left(\hat{G}_N^{MC} - \mathbb{E}[G] \right)$$

Properties:

- Unbiased, *i.e.* $\mathbb{E}\left[\hat{Q}_N^{MCCV}\right] = \mathbb{E}\left[\hat{Q}_N^{MC}\right]$
- $\underset{\beta}{\operatorname{argmin}} \, \mathbb{V}ar\left(\hat{Q}_N^{MCCV}\right) \rightarrow \beta = -\rho \dfrac{\mathbb{V}ar^{1/2}(Q)}{\mathbb{V}ar^{1/2}(G)}$
- Pearson's $\rho = \dfrac{\operatorname{Cov}(Q, G)}{\mathbb{V}ar^{1/2}(Q)\, \mathbb{V}ar^{1/2}(G)}$ where $|\rho| < 1$

$$\boxed{\mathbb{V}ar\left(\hat{Q}_N^{MCCV}\right) = \mathbb{V}ar\left(\hat{Q}_N^{MC}\right)\left(1 - \rho^2\right)}$$

The most difficult part is to find a well correlated function G

Multifidelity – Control variate with estimated LF control means

Multifidelity

Practical implications of unknown low-fidelity statistics

We have presented the CV idea, but...

- For complex applications guessing a function G is practically impossible
- Some information regarding the system under analysis need to be exploited

Multifidelity

Practical implications of unknown low-fidelity statistics

We have presented the CV idea, but...

- For complex applications guessing a function G is practically impossible
- Some information regarding the system under analysis need to be exploited

A viable alternative is to use a low-fidelity approximation of the system under analysis, however this is possible at a cost (Pasupathy *et. al* 2014)...

Multifidelity

Practical implications of unknown low-fidelity statistics

We have presented the CV idea, but...

- For complex applications guessing a function G is practically impossible
- Some information regarding the system under analysis need to be exploited

A viable alternative is to use a low-fidelity approximation of the system under analysis, however this is possible at a cost (Pasupathy *et. al* 2014)...

Let's modify the high-fidelity QoI, Q_M^{HF}, to decrease its variance

$$\hat{Q}_{M,N}^{\mathrm{HF},CV} = \hat{Q}_{M,N}^{\mathrm{HF}} + \alpha \left(\hat{Q}_{M,N}^{\mathrm{LF}} - \mathbb{E}\left[Q_M^{\mathrm{LF}} \right] \right).$$

Multifidelity

Practical implications of unknown low-fidelity statistics

We have presented the CV idea, but...

- For complex applications guessing a function G is practically impossible
- Some information regarding the system under analysis need to be exploited

A viable alternative is to use a low-fidelity approximation of the system under analysis, however this is possible at a cost (Pasupathy *et. al* 2014)...

Let's modify the high-fidelity QoI, Q_M^{HF}, to decrease its variance

$$\hat{Q}_{M,N}^{\mathrm{HF},CV} = \hat{Q}_{M,N}^{\mathrm{HF}} + \alpha \left(\hat{Q}_{M,N}^{\mathrm{LF}} - \mathbb{E}\left[Q_M^{\mathrm{LF}} \right] \right).$$

In practical situations

- the term $\mathbb{E}\left[Q_M^{\mathrm{LF}} \right]$ is unknown (low fidelity $\neq$ analytic function)
- we are using already N^{HF} simulations for evaluating $\hat{Q}_{M,N}^{\mathrm{HF}}$ and $\hat{Q}_{M,N}^{\mathrm{LF}}$
- in *Pasupathy et al.* an estimated control means approach has been proposed
- It uses an additional and independent set $\Delta^{\mathrm{LF}} = rN^{\mathrm{HF}}$ (total LF $N^{\mathrm{LF}} = (1+r)N^{\mathrm{HF}}$)

$$\mathbb{E}\left[Q_M^{\mathrm{LF}} \right] \simeq \frac{1}{(1+r)N^{\mathrm{HF}}} \sum_{i=1}^{(1+r)N^{\mathrm{HF}}} Q_M^{\mathrm{LF},(i)} = \frac{1}{(1+r)N^{\mathrm{HF}}} \sum_{i=1}^{N^{\mathrm{HF}}} Q_M^{\mathrm{LF},(i)} + \frac{1}{(1+r)N^{\mathrm{HF}}} \sum_{j=1}^{rN^{\mathrm{HF}}} Q_M^{\mathrm{LF},(j)}.$$

MULTIFIDELITY

ESTIMATED CONTROL MEANS IMPACT (ESTIMATOR)

$$
\begin{aligned}
\hat{Q}_M^{\mathrm{HF,MF}} &= \hat{Q}_M^{\mathrm{HF}} + \alpha \left(\hat{Q}_M^{\mathrm{LF}} - \mathbb{E}\left[\hat{Q}_M^{\mathrm{LF}} \right] \right) \\
&= \frac{1}{N^{\mathrm{HF}}} \sum_{i=1}^{N^{\mathrm{HF}}} Q_M^{\mathrm{HF},(i)} \\
&+ \alpha \left(\frac{1}{N^{\mathrm{HF}}} \sum_{i=1}^{N^{\mathrm{HF}}} Q^{\mathrm{LF},(i)} - \frac{1}{N^{\mathrm{HF}} + \Delta^{\mathrm{LF}}} \sum_{i=1}^{N^{\mathrm{HF}}} Q^{\mathrm{LF},(i)} - \frac{1}{N^{\mathrm{HF}} + \Delta^{\mathrm{LF}}} \sum_{j=1}^{\Delta^{\mathrm{LF}}} Q^{\mathrm{LF},(j)} \right) \\
&= \frac{1}{N^{\mathrm{HF}}} \sum_{i=1}^{N^{\mathrm{HF}}} Q_M^{\mathrm{HF},(i)} + \frac{\alpha r}{N^{\mathrm{HF}}(1+r)} \sum_{i=1}^{N^{\mathrm{HF}}} Q^{\mathrm{LF},(i)} - \frac{\alpha}{N^{\mathrm{HF}}(1+r)} \sum_{j=1}^{\Delta^{\mathrm{LF}}} Q^{\mathrm{LF},(j)},
\end{aligned}
$$

NOTE: This is the form one uses to compute the estimator. Also i,j are used here for simplicity only to separate the independent sample sets

MULTIFIDELITY

ESTIMATED CONTROL MEANS IMPACT (VARIANCE)

The reason why we separated the i,j sets is evident when we write the variance of the estimator

$$
\begin{aligned}
\mathbb{V}ar\left(\hat{Q}_M^{\mathrm{HF,MF}}\right) &= \frac{1}{N^{\mathrm{HF}}}\mathbb{V}ar\left(Q_M^{\mathrm{HF}}\right) + \frac{\alpha^2 r^2}{N^{\mathrm{HF}}(1+r)^2}\mathbb{V}ar\left(Q_M^{\mathrm{LF}}\right) \\
&+ \frac{\alpha^2 r}{N^{\mathrm{HF}}(1+r)^2}\mathbb{V}ar\left(Q_M^{\mathrm{LF}}\right) + \frac{2\alpha r}{N^{\mathrm{HF}}(1+r)}\mathrm{Cov}\left(Q_M^{\mathrm{HF}}, Q_M^{\mathrm{LF}}\right) \\
&= \frac{1}{N^{\mathrm{HF}}}\mathbb{V}ar\left(Q_M^{\mathrm{HF}}\right) + \frac{\alpha^2 r}{N^{\mathrm{HF}}(1+r)}\mathbb{V}ar\left(Q_M^{\mathrm{LF}}\right) \\
&+ \frac{2\alpha r}{N^{\mathrm{HF}}(1+r)}\mathrm{Cov}\left(Q_M^{\mathrm{HF}}, Q_M^{\mathrm{LF}}\right),
\end{aligned}
$$

We need to address two points:

- α coefficient which minimizes the variance (equivalent to standard CV)
- r which minimizes the cost for a given variance

Multifidelity

Minimization of the variance (α)

$$\frac{\mathrm{d}\,\mathbb{V}ar\left(\hat{Q}_M^{\mathrm{HF,MF}}\right)}{\mathrm{d}\,\alpha} = 0 \quad \rightarrow \quad \alpha = -\rho \frac{\mathbb{V}ar^{1/2}(Q_M^{\mathrm{HF}})}{\mathbb{V}ar^{1/2}(Q_M^{\mathrm{LF}})}.$$

NOTE:

- $r/(r+1)$ multiply both terms containing α
- the optimal coefficient α is independent from r

Multifidelity

Optimal Variance

By using the optimal coefficient α

$$\alpha = -\rho \frac{\mathbb{V}ar^{1/2}(Q_M^{\text{HF}})}{\mathbb{V}ar^{1/2}(Q_M^{\text{LF}})}$$

It is possible to obtain the variance of the estimator

$$\boxed{\mathbb{V}ar\left(\hat{Q}_M^{\text{HF},CV}\right) = \mathbb{V}ar\left(\hat{Q}_M^{\text{HF}}\right)\left(1 - \frac{r}{1+r}\rho^2\right)}$$

NOTE:

- The result is similar to the standard CV
- The effect of the correlation is reduced by a factor $r/(r+1) \to 1$ for $r \to \infty$

Q: If $\dfrac{r}{r+1} \to 1$, why don't we use a very large r for the estimator? (Remember, $N^{\text{LF}} = (1+r)N^{\text{HF}}$)

Multifidelity

Minimization of the cost (r)

- Very often the LF models are very efficient in term of computational cost (this is why we use them), but they are not entirely free
- In order to build efficient estimators we need to include their computational cost

Let's introduce the following notation

- Cost of one low-fidelity realization: $\mathcal{C}^{\mathrm{LF}}$
- Cost of one high-fidelity realization: $\mathcal{C}^{\mathrm{HF}}$
- The total cost of the estimator is
  - $\mathcal{C}^{\mathrm{tot}}_{\mathrm{MF}}\left(N^{\mathrm{HF}}, r\right) = N^{\mathrm{HF}}\mathcal{C}^{\mathrm{HF}} + N^{\mathrm{HF}}\left(1+r\right)\mathcal{C}^{\mathrm{LF}}$
  - Two free parameters, *i.e.* number of HF simulations (used also for the first LF term) and number of additional LF realizations

Remember...

$$\mathbb{E}\left[(\hat{Q}_M^{\mathrm{HF},CV} - \mathbb{E}\left[Q\right])^2\right] = \mathbb{V}ar\left(\hat{Q}_M^{\mathrm{HF},CV}\right) + \left(\mathbb{E}\left[Q_M - Q\right]\right)^2$$

Additional considerations:

- Let's assume someone is giving us the weak error $\mathbb{E}\left[Q_M - Q\right]$ committed on the resolution level M we are using
- Let's call $\mathbb{E}\left[Q_M - Q\right] = \varepsilon/2$ for simplicity

Multifidelity

Minimization of the computational cost (Problem definition)

We want to solve the following problem

- Minimization of the total computational cost:
  $\mathcal{C}_{\mathrm{MF}}^{\mathrm{tot}}\left(N^{\mathrm{HF}}, r\right) = N^{\mathrm{HF}}\mathcal{C}^{\mathrm{HF}} + N^{\mathrm{HF}}\left(1+r\right)\mathcal{C}^{\mathrm{LF}}$
- We want to reach a target MSE of ε, therefore $\mathbb{V}ar\left(\hat{Q}_M^{\mathrm{HF},CV}\right) = \varepsilon/2$

Multifidelity

Minimization of the computational cost (Problem definition)

We want to solve the following problem

- Minimization of the total computational cost:
  $\mathcal{C}_{\mathrm{MF}}^{\mathrm{tot}}\left(N^{\mathrm{HF}}, r\right) = N^{\mathrm{HF}}\mathcal{C}^{\mathrm{HF}} + N^{\mathrm{HF}}\left(1+r\right)\mathcal{C}^{\mathrm{LF}}$
- We want to reach a target MSE of ε, therefore $\mathbb{V}ar\left(\hat{Q}_M^{\mathrm{HF},CV}\right) = \varepsilon/2$

More formally, let's define our optimization problem (Lagrange constrain optimization)

$$\underset{N^{\mathrm{HF}},r,\lambda}{\operatorname{argmin}}(\mathcal{L}) \quad \mathcal{L} = \mathcal{C}_{\mathrm{MF}}^{\mathrm{tot}} - \lambda\left(\frac{1}{N^{\mathrm{HF}}}\mathbb{V}ar\left(Q_M^{\mathrm{HF}}\right)\Lambda(r) - \frac{\varepsilon^2}{2}\right)$$

$$\mathcal{C}_{\mathrm{MF}}^{\mathrm{tot}}\left(N^{\mathrm{HF}}, r\right) = N^{\mathrm{HF}}\mathcal{C}^{\mathrm{HF}} + N^{\mathrm{HF}}\left(1+r\right)\mathcal{C}^{\mathrm{LF}}$$

$$\Lambda(r) = 1 - \frac{r}{1+r}\rho^2.$$

Multifidelity

Minimization of the computational cost (Manipulations)

The three stationary conditions for the Lagrange function with respect to the variables $N^{\mathrm{HF}}, r, \lambda$ are

$$\frac{\partial \mathcal{L}}{\partial N^{\mathrm{HF}}} = \frac{\partial \mathcal{C}^{\mathrm{tot}}_{\mathrm{MF}}}{\partial N^{\mathrm{HF}}} - \lambda \frac{\mathbb{V}ar\left(Q_M^{\mathrm{HF}}\right) \Lambda(r)}{\left(N^{\mathrm{HF}}\right)^2} = 0$$

$$\frac{\partial \mathcal{L}}{\partial r} = \frac{\partial \mathcal{C}^{\mathrm{tot}}_{\mathrm{MF}}}{\partial r} + \lambda \frac{\mathbb{V}ar\left(Q_M^{\mathrm{HF}}\right)}{N^{\mathrm{HF}}} \frac{\partial \Lambda}{\partial r} = 0$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \frac{1}{N^{\mathrm{HF}}} \mathbb{V}ar\left(Q_M^{\mathrm{HF}}\right) \Lambda(r) - \frac{\varepsilon^2}{2} = 0,$$

where

$$\frac{\partial \Lambda}{\partial r} = -\frac{1}{(1+r)^2} \rho^2$$

$$\frac{\partial \mathcal{C}^{\mathrm{tot}}_{\mathrm{MF}}}{\partial N^{\mathrm{HF}}} = \mathcal{C}^{\mathrm{HF}} + (1+r)\mathcal{C}^{\mathrm{LF}} = \mathcal{C}^{\mathrm{eq}}(r)$$

$$\frac{\partial \mathcal{C}^{\mathrm{tot}}_{\mathrm{MF}}}{\partial r} = N^{\mathrm{HF}} \mathcal{C}^{\mathrm{LF}}.$$

NOTE: An equivalent computational cost $\mathcal{C}^{\mathrm{eq}}(r) = \mathcal{C}^{\mathrm{HF}}\left[1 + (r+1)/w\right] = \mathcal{C}^{\mathrm{HF}}\Gamma(r)$ is introduced to measure the unit cost per HF simulation (given r)

Multifidelity

Minimization of the computational cost (Optimal solution)

The solution of the optimization problem is obtained as

$$r^{\star} = -1 + \sqrt{\frac{w\rho^2}{1-\rho^2}}$$

$$N^{\mathrm{HF},\star} = \frac{\mathbb{V}ar\left(Q_M^{\mathrm{HF}}\right)}{\varepsilon^2/2}\Lambda(r^{\star}),$$

Multifidelity

Minimization of the computational cost (Optimal solution)

The solution of the optimization problem is obtained as

$$r^\star = -1 + \sqrt{\frac{w\rho^2}{1-\rho^2}}$$
$$N^{\mathrm{HF},\star} = \frac{\mathbb{V}ar(Q_M^{\mathrm{HF}})}{\varepsilon^2/2}\Lambda(r^\star),$$

<u>How this compare to MC?</u>

- Total cost of MC: $\mathcal{C}_{tot}^{\mathrm{MC}} = N^{\mathrm{HF}}\mathcal{C}^{\mathrm{HF}} = \frac{\mathbb{V}ar(Q_M^{\mathrm{HF}})}{\varepsilon^2/2}\mathcal{C}^{\mathrm{HF}}$
- Total cost MF: $\mathcal{C}_{tot}^{\mathrm{MF}} = N^{\mathrm{HF},\star}\mathcal{C}^{\mathrm{eq}}(r^\star) = \frac{\mathbb{V}ar(Q_M^{\mathrm{HF}})}{\varepsilon^2/2}\mathcal{C}^{\mathrm{HF}}\Theta(w,\rho^2)$, where the function $\Theta(w,\rho^2)$

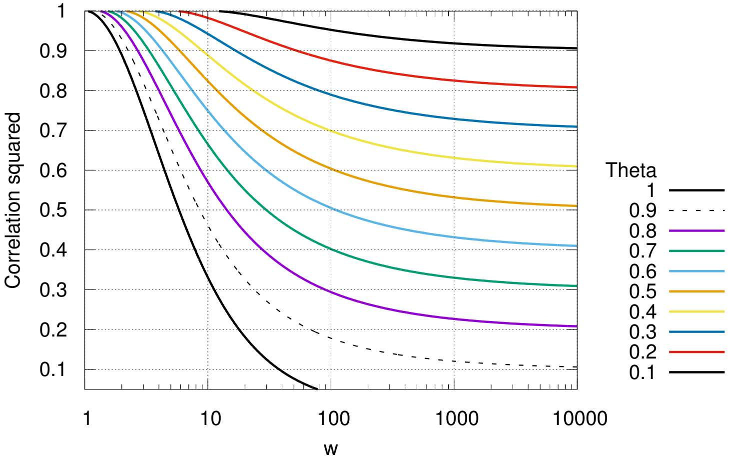
$$\Theta(w,\rho^2) \stackrel{\text{def}}{=} \Lambda(r^\star)\Gamma(r^\star)$$

measures the efficiency of the method (w.r.t. MC, *i.e.* we want $\Theta(w,\rho^2) < 1$)

Multifidelity

Efficiency compared to MC

$$\Theta(w,\rho^2) \stackrel{\text{def}}{=} \Lambda(r^\star)\Gamma(r^\star) = \left(1-\rho^2+\rho^2\sqrt{\frac{1-\rho^2}{w\rho^2}}\right)\left(1+\frac{\rho^2}{1-\rho^2}\frac{1}{w}\right)$$



Pause – In the next lecture

Next lecture
Plan

- Multilevel estimators
- Multifidelity + Multilevel = Multilevel-Multifidelity Estimators
- Numerical Examples

Few references:

- N. Metropolis, *The beginning of the Monte Carlo Method*, Los Alamos Science, Special Issue 1987.
- Mike Giles' website: `https://people.maths.ox.ac.uk/gilesm/` (I've borrowed some material from his lectures)
- *Monte Carlo Methods* by Johansen and Evers, Lecture note. University of Bristol
- Pasupathy et al, *Control-variate estimation using estimated control means,* IIE Transactions **44**(5), 381–385, 2014.
- Halton, J. H., *A retrospective and prospective survey of the Monte Carlo method*. SIAM Review, 12, 163, 1970.