Karlsruhe Institute of Technology

Uncertainty Quantification

Multilevel Monte Carlo Methods for Parametric Expectations: Distribution and Robustness Measures

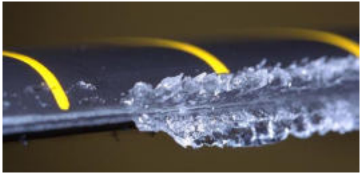
Sebastian Krumscheid (sebastian.krumscheid@kit.edu) | September 11, 2023

joint work with: Q. Ayoul-Guilmard, S. Ganesh, and F. Nobile

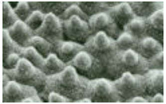
Uncertainty Quantification

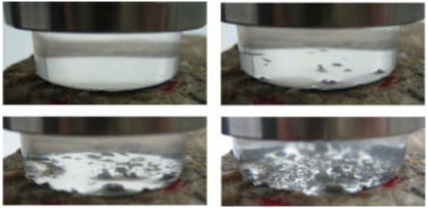
Mathematical models describing real-word phenomena *are affected by* uncertainties and incomplete knowledge.

- **Aerodynamic design**



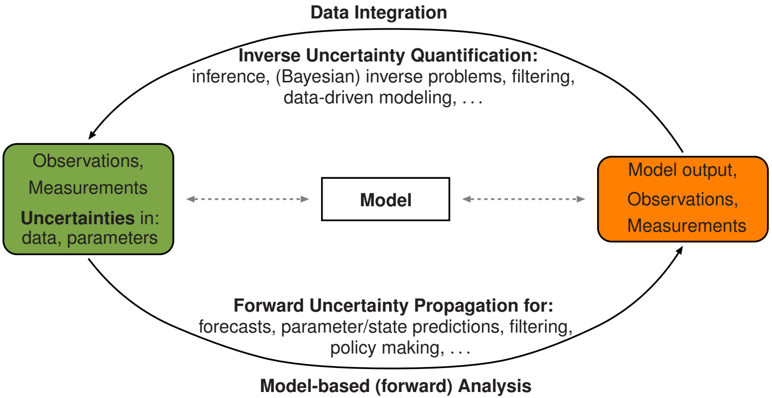
- **Rough contact mechanics** (Lotus effect, rubber on concrete, . . .)





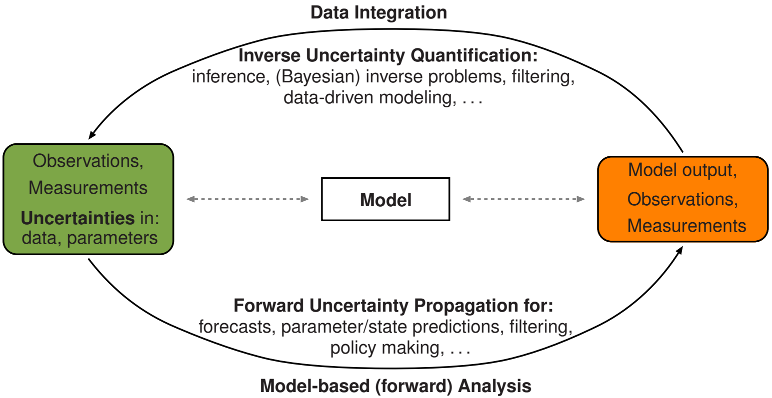
From models to data and back

Reliably using the models requires the study of the impact of all forms of error and uncertainty in models.



From models to data and back

Reliably using the models requires the study of the impact of all forms of error and uncertainty in models.

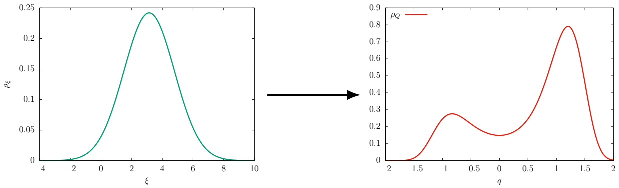


Today's focus: Forward Uncertainty Quantification — quantifying the effects of uncertainties on model predictions induced by model/input uncertainties

Forward Propagation of Uncertainties

random input: ξ $\longrightarrow$ Model $\longrightarrow$ quantity of interest: $Q \equiv Q(\xi)$

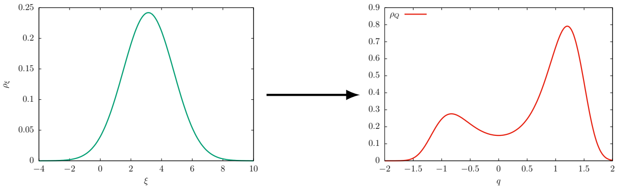
- Random input parameters ξ, possibly infinite dimensional
- **Model**: for ξ given, find u s.t. $\mathcal{M}_{\xi}(u) = 0 \rightsquigarrow$ **Quantity of interest:** $Q(\xi) = Q\big(u(\xi)\big)$



Forward Propagation of Uncertainties

random input: ξ $\longrightarrow$ Model $\longrightarrow$ quantity of interest: $Q \equiv Q(\xi)$

- Random input parameters ξ, possibly infinite dimensional
- **Model**: for ξ given, find u s.t. $\mathcal{M}_{\xi}(u) = 0 \rightsquigarrow$ **Quantity of interest:** $Q(\xi) = Q\big(u(\xi)\big)$



- **Goal:** compute statistical quantities of Q with minimal regularity assumptions.

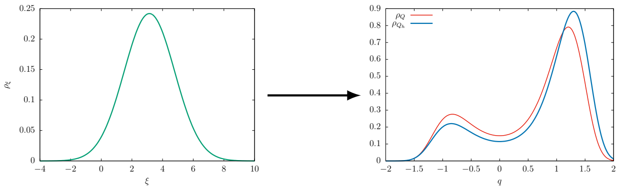
Forward Propagation of Uncertainties

random input: $\boldsymbol{\xi}$ → **Model** → quantity of interest: $Q \equiv Q(\boldsymbol{\xi})$

$\downarrow$ approximations

computational Model → $Q_h \equiv Q_h(\boldsymbol{\xi})$

- Random input parameters $\boldsymbol{\xi}$, possibly infinite dimensional
- **Model**: for $\boldsymbol{\xi}$ given, find u s.t. $\mathcal{M}_{\boldsymbol{\xi}}(u) = 0 \rightsquigarrow$ **Quantity of interest:** $Q(\boldsymbol{\xi}) = Q\big(u(\boldsymbol{\xi})\big)$



- **Goal:** compute statistical quantities of Q with minimal regularity assumptions.

Monte Carlo method

- General purpose method to approximate moments of Q_h for *given discretization h*:
  - generate sample of N i.i.d. realizations $\xi^{(1)}, \dots, \xi^{(N)} \sim \mathbb{P}_\xi$
  - compute corresponding model outputs $Q_h\big(\xi^{(i)}\big)$, $i = 1, \dots, N$
  - approximate expectation of Q_h by **sample average**:

$$\mathbb{E}(Q) \approx \mathbb{E}(Q_h) \approx E_N(Q_h) := \frac{1}{N}\sum_{i=1}^{N} Q_h^{(i)} \,, \quad Q_h^{(i)} \equiv Q_h\big(\xi^{(i)}\big) \text{ i.i.d. realizations}$$

Monte Carlo method

- General purpose method to approximate moments of Q_h for *given discretization h*:
  - generate sample of N i.i.d. realizations $\xi^{(1)}, \dots, \xi^{(N)} \sim \mathbb{P}_\xi$
  - compute corresponding model outputs $Q_h(\xi^{(i)})$, $i = 1, \dots, N$
  - approximate expectation of Q_h by **sample average**:

$$\mathbb{E}(Q) \approx \mathbb{E}(Q_h) \approx E_N(Q_h) := \frac{1}{N}\sum_{i=1}^{N} Q_h^{(i)} , \quad Q_h^{(i)} \equiv Q_h(\xi^{(i)}) \text{ i.i.d. realizations}$$

- **Question:** How should we select h and N to ensure a certain ***accuracy*** at *minimal* ***cost***?

Monte Carlo method

Karlsruhe Institute of Technology

- General purpose method to approximate moments of Q_h for *given discretization* h:
  - generate sample of N i.i.d. realizations $\xi^{(1)}, \ldots, \xi^{(N)} \sim \mathbb{P}_\xi$
  - compute corresponding model outputs $Q_h(\xi^{(i)})$, $i = 1, \ldots, N$
  - approximate expectation of Q_h by **sample average**:

$$\mathbb{E}(Q) \approx \mathbb{E}(Q_h) \approx E_N(Q_h) := \frac{1}{N} \sum_{i=1}^{N} Q_h^{(i)} \,, \quad Q_h^{(i)} \equiv Q_h(\xi^{(i)}) \text{ i.i.d. realizations}$$

- **Question:** How should we select h and N to ensure a certain ***accuracy*** at *minimal* ***cost***?
- **Mean squared error** (MSE) accuracy of Monte Carlo method:

$$\text{MSE} := \mathbb{E}\left(\left|\mathbb{E}(Q) - E_N(Q_h)\right|^2\right) = \underbrace{\left(\mathbb{E}(Q) - \mathbb{E}(Q_h)\right)^2}_{\text{squared bias}} + \underbrace{\frac{\text{Var}(Q_h)}{N}}_{\text{stat. error}}$$

Monte Carlo method

Karlsruhe Institute of Technology

- General purpose method to approximate moments of Q_h for *given discretization h*:
  - generate sample of N i.i.d. realizations $\xi^{(1)}, \ldots, \xi^{(N)} \sim \mathbb{P}_\xi$
  - compute corresponding model outputs $Q_h\big(\xi^{(i)}\big)$, $i = 1, \ldots, N$
  - approximate expectation of Q_h by **sample average**:

$$\mathbb{E}(Q) \approx \mathbb{E}(Q_h) \approx E_N(Q_h) := \frac{1}{N} \sum_{i=1}^{N} Q_h^{(i)} , \quad Q_h^{(i)} \equiv Q_h\big(\xi^{(i)}\big) \text{ i.i.d. realizations}$$

- **Question:** How should we select h and N to ensure a certain ***accuracy*** at *minimal* ***cost***?
- **Mean squared error** (MSE) accuracy of Monte Carlo method:

$$\text{MSE} := \mathbb{E}\Big(\big|\mathbb{E}(Q) - E_N(Q_h)\big|^2\Big) = \underbrace{\big(\mathbb{E}(Q) - \mathbb{E}(Q_h)\big)^2}_{\text{squared bias}} + \underbrace{\frac{\text{Var}(Q_h)}{N}}_{\text{stat. error}}$$

- That is, the parameters are implied by MSE tolerance with:
  - h dictated by the bias = discretization error of *approximate model*
  - sample size N given by the statistical error

Monte Carlo: cost vs. accuracy

Lemma: MSE complexity analysis MC (error vs. cost)

Suppose that:

- $\left|\mathbb{E}(Q - Q_h)\right| = \mathcal{O}(h^\alpha)$,
- $\mathrm{Var}(Q_h) = \mathcal{O}(1)$,
- $\mathrm{cost}\left({Q_h}^{(i)}\right) = \mathcal{O}(h^{-\gamma})$.

Achieving $\mathrm{MSE} \leq \varepsilon^2$ requires $N = \mathcal{O}(\varepsilon^{-2})$ and $h = \mathcal{O}(\varepsilon^{1/\alpha})$, resulting in

$$\text{comp. cost MC} = N \,\mathrm{cost}\left(Q_h\right) = \mathcal{O}\left(\varepsilon^{-(2+\gamma/\alpha)}\right) \,.$$

Monte Carlo: cost vs. accuracy

Lemma: MSE complexity analysis MC (error vs. cost)

Suppose that:
- $\big|\mathbb{E}(Q - Q_h)\big| = \mathcal{O}(h^\alpha)$,
- $\mathrm{Var}(Q_h) = \mathcal{O}(1)$,
- $\mathrm{cost}\big(Q_h{}^{(i)}\big) = \mathcal{O}(h^{-\gamma})$.

Achieving MSE $\leq \varepsilon^2$ requires $N = \mathcal{O}(\varepsilon^{-2})$ and $h = \mathcal{O}(\varepsilon^{1/\alpha})$, resulting in

$$\text{comp. cost MC} = N\,\mathrm{cost}\big(Q_h\big) = \mathcal{O}\big(\varepsilon^{-(2+\gamma/\alpha)}\big) .$$

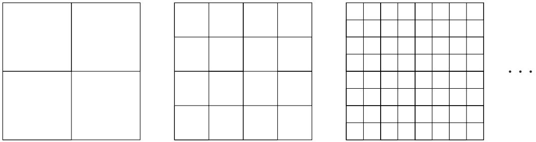
- Another error criterion is the **probability of failure**: find (h, N) such that

$$\mathbb{P}\big(|\mathbb{E}(Q) - E_N(Q_h)| \geq 2\varepsilon\big) \leq \tau , \quad \text{for } \varepsilon > 0 \text{ and } \tau \in (0, 1) .$$

- For $\varepsilon \ll 1$, the distribution of $E_N(Q_h)$ can be approximated in view of central limit theorem (CLT) as $\mathrm{Var}(Q_h)$ is bounded.

Multilevel Monte Carlo for expected values

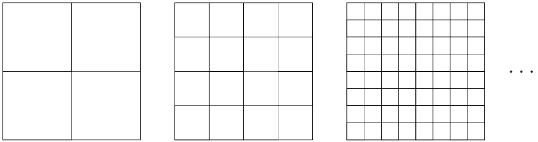
- **Main idea:** use a hierarchy of $L \in \mathbb{N}_0$ discretizations $h_0 > h_1 > \cdots > h_L$, instead of just one fine level:



$\ell = 0 \rightsquigarrow h_0$ $\quad$ $\ell = 1 \rightsquigarrow h_1$ $\quad$ $\ell = 2 \rightsquigarrow h_2$

Multilevel Monte Carlo for expected values

- **Main idea:** use a hierarchy of $L \in \mathbb{N}_0$ discretizations $h_0 > h_1 > \cdots > h_L$, instead of just one fine level:



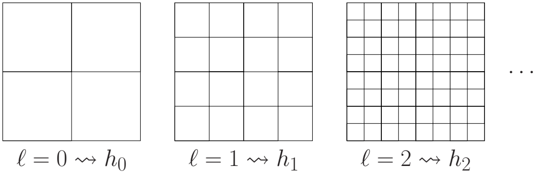
$\ell = 0 \rightsquigarrow h_0$ $\quad$ $\ell = 1 \rightsquigarrow h_1$ $\quad$ $\ell = 2 \rightsquigarrow h_2$

- The multilevel Monte Carlo (MLMC) estimator of $\mathbb{E}(Q)$ then is ($Q_\ell \equiv Q_{h_\ell}$):

$$\mathbb{E}(Q) \approx \mathbb{E}(Q_L) = \mathbb{E}(Q_0) + \sum_{\ell=1}^{L} \mathbb{E}(Q_\ell - Q_{\ell-1}) \approx E_{N_0}(Q_0) + \sum_{\ell=1}^{L} E_{N_\ell}(Q_\ell - Q_{\ell-1}) =: \hat{Q}_{N,L}$$

Multilevel Monte Carlo for expected values

- **Main idea:** use a hierarchy of $L \in \mathbb{N}_0$ discretizations $h_0 > h_1 > \cdots > h_L$, instead of just one fine level:



- The multilevel Monte Carlo (MLMC) estimator of $\mathbb{E}(Q)$ then is ($Q_\ell \equiv Q_{h_\ell}$):

$$\mathbb{E}(Q) \approx \mathbb{E}(Q_L) = \mathbb{E}(Q_0) + \sum_{\ell=1}^{L} \mathbb{E}(Q_\ell - Q_{\ell-1}) \approx E_{N_0}(Q_0) + \sum_{\ell=1}^{L} E_{N_\ell}(Q_\ell - Q_{\ell-1}) =: \hat{Q}_{N,L}$$

- MSE accuracy of multilevel Monte Carlo method is

$$\mathrm{MSE} \equiv \mathbb{E}\left(\left|\mathbb{E}(Q) - \hat{Q}_{N,L}\right|^2\right) = \underbrace{\left(\mathbb{E}(Q) - \mathbb{E}(Q_L)\right)^2}_{\text{squared bias}} + \underbrace{\sum_{\ell=0}^{L} \frac{\mathrm{Var}(Q_\ell - Q_{\ell-1})}{N_\ell}}_{\text{statistical error}}, \quad Q_{-1} \equiv 0$$

- Select parameters $L \in \mathbb{N}_0$ and $N \equiv (N_0, N_1, \ldots N_L)^T \in \mathbb{N}^{L+1}$ s.t. MSE criterion is met at minimal cost.
- Optimal parameters by **balancing errors:**
  - bias = discretization error $\rightsquigarrow L$,
  - **determine** N_ℓ by minimizing $\sum_{\ell=1}^{L} N_\ell C_\ell$ subject to MSE $\leq \varepsilon^2$,
  - where $C_\ell =$ cost for each $\left(Q_\ell^{(i,\ell)}, Q_{\ell-1}^{(i,\ell)}\right)$.

- Select parameters $L \in \mathbb{N}_0$ and $N \equiv (N_0, N_1, \dots N_L)^T \in \mathbb{N}^{L+1}$ s.t. MSE criterion is met at minimal cost.
- Optimal parameters by **balancing errors:**
  - bias = discretization error $\rightsquigarrow L$,
  - **determine** N_ℓ by minimizing $\sum_{\ell=1}^{L} N_\ell C_\ell$ subject to MSE $\leq \varepsilon^2$,
  - where $C_\ell =$ cost for each $\left(Q_\ell^{(i,\ell)}, Q_{\ell-1}^{(i,\ell)}\right)$.

Theorem: Complexity analysis MLMC for $h_{\ell-1}/h_\ell = s > 1$ [Giles, 2008; Cliffe et al., 2011]

Suppose that:
- $\left|\mathbb{E}(Q - Q_\ell)\right| = \mathcal{O}({h_\ell}^{\alpha})$,
- $\mathrm{Var}(Q_\ell - Q_{\ell-1}) = \mathcal{O}({h_\ell}^{\beta})$,
- $C_\ell = \mathcal{O}({h_\ell}^{-\gamma})$.

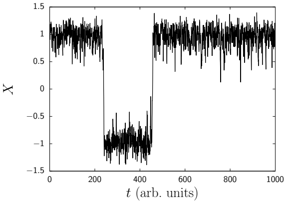
If $2\alpha \geq \min\{\beta, \gamma\}$, then there exists an MLMC estimator $\hat{Q}_{N,L}$ that satisfies MSE $\leq \varepsilon^2$ with

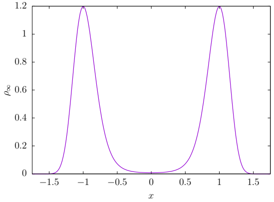
$$\text{comp. cost MLMC} = \sum_{\ell=1}^{L} N_\ell C_\ell \lesssim \begin{cases} \varepsilon^{-2}, & \beta > \gamma, \\ \varepsilon^{-2} \ln(\varepsilon)^2, & \beta = \gamma, \\ \varepsilon^{-\left(2 + \frac{\gamma-\beta}{\alpha}\right)}, & \beta < \gamma. \end{cases}$$

- Other error criteria are possible, e.g., tuning MLMC for *probability of failure* is possible thanks to CLT [Collier et al. 2015], [Hoel, K. 2019].

Assessing the distribution beyond moments

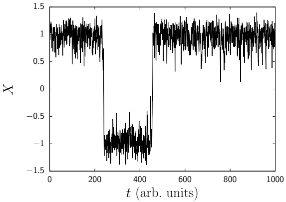
- Expectation is a **good starting point**, but we would like more: **assessing the distribution**

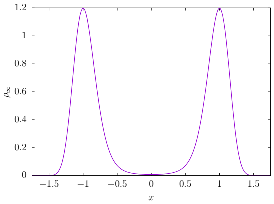




Assessing the distribution beyond moments

- Expectation is a **good starting point**, but we would like more: **assessing the distribution**





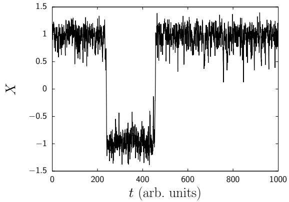
- One could estimate central moments of arbitrary order

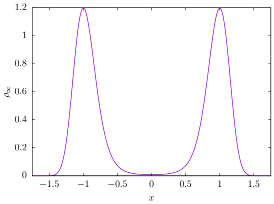
$$\mu_p(Q) := \mathbb{E}\left[\left(Q - \mathbb{E}(Q)\right)^p\right], \quad p \in \mathbb{N}$$

with MLMC [Bierig, Chernov. 2016], [K., Nobile, and Pisaroni. 2020]

Assessing the distribution beyond moments

- Expectation is a **good starting point**, but we would like more: **assessing the distribution**





- One could estimate central moments of arbitrary order

$$\mu_p(Q) := \mathbb{E}\left[\left(Q - \mathbb{E}(Q)\right)^p\right] , \quad p \in \mathbb{N}$$

  with MLMC [Bierig, Chernov. 2016], [K., Nobile, and Pisaroni. 2020]

- However, some applications require techniques that systematically go **beyond using a few moments**: characterization of entire distribution or applications in risk averse optimization, e.g., involving quantiles

$$\arg\min_{z\in\mathbb{R}^d} \mathcal{R}(z) , \quad \mathcal{R}(z) \equiv \mathcal{R}(z;\mathbb{P}_Q) : \text{"risk" of design parameter } z$$

Parametric Expectations

- Many such distributional assessments are related to **parametric expectations** of the form

$$\Phi(\vartheta) := \mathbb{E}\big(\phi(\vartheta, Q)\big) \ , \quad \vartheta \in \Theta$$

Parametric Expectations

- Many such distributional assessments are related to **parametric expectations** of the form

$$\Phi(\vartheta) := \mathbb{E}\big(\phi(\vartheta, Q)\big) \;, \quad \vartheta \in \Theta$$

Examples:

1. **Mean & variance**: $\phi(\vartheta, Q) = (Q - \vartheta)^2 \rightsquigarrow \mathrm{Var}\,(Q) = \min_\vartheta \Phi(\vartheta), \mathbb{E}(Q) = \arg\min_\vartheta \Phi(\vartheta)$
1. **Characteristic function** of Q: $\phi(\vartheta, Q) = e^{i\vartheta Q} = \cos(\vartheta Q) + i \sin(\vartheta Q)$
2. **Cumulative distribution function** of Q: $\phi(\vartheta, Q) = I(Q \leq \vartheta)$

Parametric Expectations

- Many such distributional assessments are related to **parametric expectations** of the form

$$\Phi(\vartheta) := \mathbb{E}\big(\phi(\vartheta, Q)\big)\ , \quad \vartheta \in \Theta$$

Examples:

1. **Mean & variance**: $\phi(\vartheta, Q) = (Q - \vartheta)^2 \rightsquigarrow \operatorname{Var}(Q) = \min_\vartheta \Phi(\vartheta)$, $\mathbb{E}(Q) = \arg\min_\vartheta \Phi(\vartheta)$
1. **Characteristic function** of Q: $\phi(\vartheta, Q) = e^{i\vartheta Q} = \cos(\vartheta Q) + i\sin(\vartheta Q)$
2. **Cumulative distribution function** of Q: $\phi(\vartheta, Q) = I(Q \leq \vartheta)$

Difficulty for CDF: $\phi(\vartheta, \cdot)$ is **discontinuous**! Consequently, the variance $\operatorname{Var}\big(\phi(\theta_j, Q_\ell) - \phi(\theta_j, Q_{\ell-1})\big)$ will decay slowly $\rightsquigarrow$ Not much gain in MLMC.

Parametric Expectations

- Many such distributional assessments are related to **parametric expectations** of the form

$$\Phi(\vartheta) := \mathbb{E}\big(\phi(\vartheta, Q)\big)\ , \quad \vartheta \in \Theta$$

Examples:

1. **Mean & variance**: $\phi(\vartheta, Q) = (Q - \vartheta)^2 \rightsquigarrow \operatorname{Var}(Q) = \min_\vartheta \Phi(\vartheta)$, $\mathbb{E}(Q) = \arg\min_\vartheta \Phi(\vartheta)$
1. **Characteristic function** of Q: $\phi(\vartheta, Q) = e^{i\vartheta Q} = \cos(\vartheta Q) + i \sin(\vartheta Q)$
2. **Cumulative distribution function** of Q: $\phi(\vartheta, Q) = I(Q \leq \vartheta)$

Difficulty for CDF: $\phi(\vartheta, \cdot)$ is **discontinuous**! Consequently, the variance $\operatorname{Var}\big(\phi(\theta_j, Q_\ell) - \phi(\theta_j, Q_{\ell-1})\big)$ will decay slowly $\rightsquigarrow$ Not much gain in MLMC.

Remedies in the context of MLMC (cf. [Giles. 2023]):

- Regularizing ϕ to yield $\Phi(\vartheta) = \mathbb{E}\big(\phi(\vartheta, Q)\big) \approx \mathbb{E}\big(\phi_\delta(\vartheta, Q)\big)$ [Giles, Nagapetyan, Ritter. 2015]
- Approximate CDF (PDF) based on MLMC estimates of moments [Bierig, Chernov. 2016]
- numerical smoothing via conditioning [Bayer, Ben Hammouda, Tempone. 2022]

Remedy for CDF: antiderivative/integration approach [K., Nobile. 2018]

For any $\tau \in (0,1)$ define

$$\Phi(\vartheta) = \mathbb{E}\big(\phi(\vartheta, Q)\big) \;, \quad \phi(\vartheta, Q) = \vartheta + \frac{1}{1-\tau}(Q-\vartheta)^+ \;.$$

Then

$$F(\vartheta) = (1-\tau)\Phi'(\vartheta) + \tau \;,$$

which is the **starting point for the MLMC estimator**:

$\rightsquigarrow$ requires effective estimator of Φ and its derivatives.

Remedy for CDF: antiderivative/integration approach [K., Nobile. 2018]

Karlsruhe Institute of Technology

For any $\tau \in (0,1)$ define

$$\Phi(\vartheta) = \mathbb{E}\big(\phi(\vartheta, Q)\big)\ , \quad \phi(\vartheta, Q) = \vartheta + \frac{1}{1-\tau}(Q-\vartheta)^+\ .$$

Then

$$F(\vartheta) = (1-\tau)\Phi'(\vartheta) + \tau\ ,$$

which is the **starting point for the MLMC estimator**:

$\rightsquigarrow$ requires effective estimator of Φ and its derivatives.

This approach may also offer approximations for:

- PDF: $f(\vartheta) = F'(\vartheta) = (1-\tau)\Phi''(\vartheta)$
- τ-quantile: $q_\tau = \inf\{\vartheta\colon F(\vartheta) \geq \tau\} = \arg\min_{\vartheta\in\mathbb{R}} \Phi(\vartheta)$
- Conditional Value at Risk: $\mathrm{CVaR}_\tau = \frac{1}{1-\tau}\int_{q_\tau}^{\infty} \vartheta\, dF(\vartheta) = \min_{\vartheta\in\mathbb{R}} \Phi(\vartheta)$

MLMC method for parametric expectations

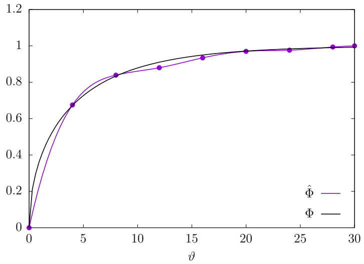
- **Goal:** given ϕ, approximate $\Phi(\vartheta) = \mathbb{E}\big(\phi(\vartheta, Q)\big)$ and its derivatives uniformly in Θ.

MLMC method for parametric expectations

- **Goal:** given ϕ, approximate $\Phi(\vartheta) = \mathbb{E}\big(\phi(\vartheta, Q)\big)$ and its derivatives uniformly in Θ.
- **Idea:** extend collection of pointwise estimators in nodes $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ to function:

MLMC method for parametric expectations

- **Goal:** given ϕ, approximate $\Phi(\vartheta) = \mathbb{E}\big(\phi(\vartheta, Q)\big)$ and its derivatives uniformly in Θ.
- **Idea:** extend collection of pointwise estimators in nodes $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ to function:

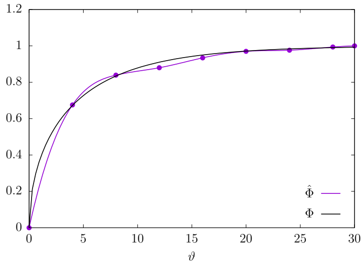


Interpolation approach (e.g., splines or polynomials):

- introduce grid $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n) \subset \Theta$
- compute MLMC estimate $\Phi_L^{\mathrm{MLMC}}(\theta_j)$ of $\Phi(\theta_j)$, $j = 1, \ldots, n$ (same samples of Q_ℓ for every θ_j)
- interpolate values $\Phi_L^{\mathrm{MLMC}}(\boldsymbol{\theta}) = \big(\Phi_L^{\mathrm{MLMC}}(\theta_j)\big)_{j=1}^n$:
  $\hat{\Phi}_L = \mathcal{I}_n\big(\Phi_L^{\mathrm{MLMC}}(\boldsymbol{\theta})\big)$

MLMC method for parametric expectations

- **Goal:** given ϕ, approximate $\Phi(\vartheta) = \mathbb{E}\big(\phi(\vartheta, Q)\big)$ and its derivatives uniformly in Θ.
- **Idea:** extend collection of pointwise estimators in nodes $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ to function:



Interpolation approach (e.g., splines or polynomials):

- introduce grid $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n) \subset \Theta$
- compute MLMC estimate $\Phi_L^{\text{MLMC}}(\theta_j)$ of $\Phi(\theta_j)$, $j = 1, \ldots, n$ (same samples of Q_ℓ for every θ_j)
- interpolate values $\Phi_L^{\text{MLMC}}(\boldsymbol{\theta}) = \big(\Phi_L^{\text{MLMC}}(\theta_j)\big)_{j=1}^n$:

  $\hat{\Phi}_L = \mathcal{I}_n\big(\Phi_L^{\text{MLMC}}(\boldsymbol{\theta})\big)$

Exemplary **a-priori properties** for Spline interpolation $\mathcal{I}_n$:

- $\Big\| f^{(m)} - \frac{d^m}{d\vartheta^m}\mathcal{I}_n\big(f(\boldsymbol{\theta})\big)\Big\|_{L^\infty} \leq c_1 n^{m-(k+1)}$, for any $f \in C^{k+1}(\bar{\Theta})$, $m \leq k$
- $\big\|\mathcal{I}_n(\boldsymbol{x})\big\|_{L^\infty} \leq c_2 \|\boldsymbol{x}\|_{\ell^\infty}$, for any $\boldsymbol{x} \in \mathbb{R}^n$
- $\text{cost}\big(\mathcal{I}_n(\boldsymbol{x})\big) \leq c_3 n$, $\boldsymbol{x} \in \mathbb{R}^n$
- $\Big\|\frac{d^m}{d\vartheta^m}\mathcal{I}_n(\boldsymbol{x})\Big\|_{L^\infty} \leq c_5 (n-1)^m \big\|\mathcal{I}_n(\boldsymbol{x})\big\|_{L^\infty}$, $\boldsymbol{x} \in \mathbb{R}^n$, $m \geq 1$ (inverse inequality)

Complexity Analysis

- Measure accuracy in terms of mean squared error:

$$\mathrm{MSE}\big(\hat{\Phi}_L^{(m)}\big) := \mathbb{E}\left(\left\|\hat{\Phi}_L^{(m)} - \Phi^{(m)}\right\|_{L^\infty}^2\right)\,, \quad m \in \mathbb{N}_0\,, \qquad L^\infty \equiv L^\infty(\Theta)\,.$$

Complexity Analysis

- Measure accuracy in terms of mean squared error:

$$\mathrm{MSE}\big(\hat{\Phi}_L^{(m)}\big) := \mathbb{E}\Big(\big\|\hat{\Phi}_L^{(m)} - \Phi^{(m)}\big\|_{L^\infty}^2\Big) \;, \qquad m \in \mathbb{N}_0 \;, \qquad L^\infty \equiv L^\infty(\Theta) \;.$$

- Using the notation $\Phi_L = \mathbb{E}\big(\phi(\cdot, Q_L)\big)$, the MSE can be decomposed as:

$$\mathrm{MSE}\big(\hat{\Phi}_L^{(m)}\big) \le 3\,\underbrace{\big\|\Phi^{(m)} - \mathcal{I}_n^{(m)}(\Phi)\big\|_{L^\infty}^2}_{\text{squared interpolation error}} + 3\,\underbrace{\big\|\mathcal{I}_n^{(m)}(\Phi - \Phi_L)\big\|_{L^\infty}^2}_{\text{squared bias}} + 3\,\underbrace{\mathbb{E}\Big(\big\|\mathcal{I}_n^{(m)}\big(\Phi_L - \Phi_L^{\mathtt{MLMC}}\big)\big\|_{L^\infty}^2\Big)}_{\text{stat. error}}$$

Complexity Analysis

- Measure accuracy in terms of mean squared error:

$$\mathrm{MSE}\big(\hat{\Phi}_L^{(m)}\big) := \mathbb{E}\left(\left\|\hat{\Phi}_L^{(m)} - \Phi^{(m)}\right\|_{L^\infty}^2\right), \quad m \in \mathbb{N}_0 , \qquad L^\infty \equiv L^\infty(\Theta) .$$

- Using the notation $\Phi_L = \mathbb{E}\big(\phi(\cdot, Q_L)\big)$, the MSE can be decomposed as:

$$\mathrm{MSE}\big(\hat{\Phi}_L^{(m)}\big) \leq 3 \underbrace{\left\|\Phi^{(m)} - \mathcal{I}_n^{(m)}(\Phi)\right\|_{L^\infty}^2}_{\text{squared interpolation error}} + 3 \underbrace{\left\|\mathcal{I}_n^{(m)}(\Phi - \Phi_L)\right\|_{L^\infty}^2}_{\text{squared bias}} + 3 \underbrace{\mathbb{E}\left(\left\|\mathcal{I}_n^{(m)}(\Phi_L - \Phi_L^{\mathtt{MLMC}})\right\|_{L^\infty}^2\right)}_{\text{stat. error}}$$

- **Parameter selection:**
  - interpolation error $\rightsquigarrow n$,
  - bias $\rightsquigarrow L$,
  - cost minimization s.t. **variance constraint** $\rightsquigarrow N_\ell$

Complexity Analysis

- Measure accuracy in terms of mean squared error:

$$\mathrm{MSE}\big(\hat{\Phi}_L^{(m)}\big) := \mathbb{E}\Big(\big\|\hat{\Phi}_L^{(m)} - \Phi^{(m)}\big\|_{L^\infty}^2\Big)\ , \quad m \in \mathbb{N}_0\ , \qquad L^\infty \equiv L^\infty(\Theta)\ .$$

- Using the notation $\Phi_L = \mathbb{E}\big(\phi(\cdot, Q_L)\big)$, the MSE can be decomposed as:

$$\mathrm{MSE}\big(\hat{\Phi}_L^{(m)}\big) \le 3\,\underbrace{\big\|\Phi^{(m)} - \mathcal{I}_n^{(m)}(\Phi)\big\|_{L^\infty}^2}_{\text{squared interpolation error}} + 3\,\underbrace{\big\|\mathcal{I}_n^{(m)}(\Phi - \Phi_L)\big\|_{L^\infty}^2}_{\text{squared bias}} + 3\,\underbrace{\mathbb{E}\Big(\big\|\mathcal{I}_n^{(m)}\big(\Phi_L - \Phi_L^{\mathtt{MLMC}}\big)\big\|_{L^\infty}^2\Big)}_{\text{stat. error}}$$

- **Parameter selection:**
  - interpolation error $\rightsquigarrow n$,
  - bias $\rightsquigarrow L$,
  - cost minimization s.t. **variance constraint** $\rightsquigarrow N_\ell$
- **Notation:** $\mathrm{Var}(\xi) = \mathbb{E}\big(\|\xi - \mathbb{E}(\xi)\|_{\ell^\infty}^2\big)$, for any r.v. ξ with values in $\mathbb{R}^n$
- **Useful technical result**: let $\big(\xi^{(1)}, \dots, \xi^{(N)}\big) \subset \mathbb{R}^n$ independent, then

$$\mathrm{Var}\Big(\sum_{i=1}^{N} \xi^{(i)}\Big) \le c\ln(n)\sum_{i=1}^{N}\mathrm{Var}\big(\xi^{(i)}\big)$$

Theorem: A-priori MSE complexity ($h_{\ell-1}/h_\ell = s > 1$) for Spline interpolation [K., Nobile. 2018]

Suppose that:

- $\sup_{\vartheta\in\Theta}\left|\Phi(\vartheta) - \mathbb{E}\left(\phi(\vartheta, Q_l)\right)\right| = \mathcal{O}(h_l{}^\alpha)$
- $\mathbb{E}\left(\left\|\phi(\cdot, Q_l) - \phi(\cdot, Q_{l-1})\right\|^2_{L^\infty(\Theta)}\right) = \mathcal{O}(h_l{}^\beta)$,
- cost for each $\left(Q_\ell^{(i,\ell)}, Q_{\ell-1}^{(i,\ell)}\right) = \mathcal{O}(h_\ell{}^{-\gamma})$

Let $m = 0$. If $\Phi \in C^{k+1}(\Theta)$ and $2\alpha \geq \min\{\beta,\gamma\}$, then there exists an MLMC estimator $\hat{\Phi}_L$ of Φ such that $\mathrm{MSE}\left(\hat{\Phi}_L\right) = \mathcal{O}(\varepsilon^2)$ with

$$\text{comp. cost} \lesssim \varepsilon^{-\left(2+\frac{1}{k+1}\right)}|\ln(\varepsilon)| + |\ln(\varepsilon)| \begin{cases} \varepsilon^{-2}\,, & \beta > \gamma\,, \\ \varepsilon^{-2}\ln(\varepsilon)^2\,, & \beta = \gamma\,, \\ \varepsilon^{-\left(2+\frac{\gamma-\beta}{\alpha}\right)}\,, & \beta < \gamma\,. \end{cases}$$

- **NB:** first term accounts for cost of computing interpolation: negligible for heavy computational models; systematically removable by $n = n_\ell$ (*different interpolation grid on each level*; Cor. 2.3 in [K., Nobile. 2018]).

Theorem: A-priori MSE complexity ($h_{\ell-1}/h_\ell = s > 1$) for Spline interpolation [K., Nobile. 2018]

Suppose that:

- $\sup_{\vartheta\in\Theta}\left|\Phi(\vartheta) - \mathbb{E}\left(\phi(\vartheta, Q_l)\right)\right| = \mathcal{O}(h_l{}^{\alpha})$
- $\mathbb{E}\left(\left\|\phi(\cdot, Q_l) - \phi(\cdot, Q_{l-1})\right\|^2_{L^\infty(\Theta)}\right) = \mathcal{O}(h_l{}^{\beta})$,
- cost for each $\left(Q_\ell^{(i,\ell)}, Q_{\ell-1}^{(i,\ell)}\right) = \mathcal{O}(h_\ell{}^{-\gamma})$

Let $m = 0$. If $\Phi \in C^{k+1}(\Theta)$ and $2\alpha \geq \min\{\beta, \gamma\}$, then there exists an MLMC estimator $\hat{\Phi}_L$ of Φ such that $\mathrm{MSE}\left(\hat{\Phi}_L\right) = \mathcal{O}(\varepsilon^2)$ with

$$\text{comp. cost} \lesssim \varepsilon^{-\left(2+\frac{1}{k+1}\right)}|\ln(\varepsilon)| + |\ln(\varepsilon)| \begin{cases} \varepsilon^{-2}, & \beta > \gamma, \\ \varepsilon^{-2}\ln(\varepsilon)^2, & \beta = \gamma, \\ \varepsilon^{-\left(2+\frac{\gamma-\beta}{\alpha}\right)}, & \beta < \gamma. \end{cases}$$

- **NB:** first term accounts for cost of computing interpolation: negligible for heavy computational models; systematically removable by $n = n_\ell$ (*different interpolation grid on each level*; Cor. 2.3 in [K., Nobile. 2018]).
- Neglecting first term: complexity is the same as for expectations, up to extra log factor.

Theorem: A-priori MSE complexity ($h_{\ell-1}/h_\ell = s > 1$) for Spline interpolation [K., Nobile. 2018]

Suppose that:
- $\sup_{\vartheta\in\Theta}\left|\Phi(\vartheta) - \mathbb{E}(\phi(\vartheta, Q_l))\right| = \mathcal{O}({h_l}^{\alpha})$
- $\mathbb{E}\Big(\big\|\phi(\cdot, Q_l) - \phi(\cdot, Q_{l-1})\big\|^2_{L^\infty(\Theta)}\Big) = \mathcal{O}({h_l}^{\beta})$,
- cost for each $\left(Q_\ell^{(i,\ell)}, Q_{\ell-1}^{(i,\ell)}\right) = \mathcal{O}({h_\ell}^{-\gamma})$

Let $m = 0$. If $\Phi \in C^{k+1}(\Theta)$ and $2\alpha \geq \min\{\beta, \gamma\}$, then there exists an MLMC estimator $\hat{\Phi}_L$ of Φ such that $\mathrm{MSE}(\hat{\Phi}_L) = \mathcal{O}(\varepsilon^2)$ with

$$\text{comp. cost} \lesssim \varepsilon^{-\left(2+\frac{1}{k+1}\right)}|\ln(\varepsilon)| + |\ln(\varepsilon)| \begin{cases} \varepsilon^{-2}\,, & \beta > \gamma\,, \\ \varepsilon^{-2}\ln(\varepsilon)^2\,, & \beta = \gamma\,, \\ \varepsilon^{-\left(2+\frac{\gamma-\beta}{\alpha}\right)}\,, & \beta < \gamma\,. \end{cases}$$

- **NB:** first term accounts for cost of computing interpolation: negligible for heavy computational models; systematically removable by $n = n_\ell$ (*different interpolation grid on each level*; Cor. 2.3 in [K., Nobile. 2018]).
- Neglecting first term: complexity is the same as for expectations, up to extra log factor.
- Similar complexity for Φ analytic and using an interpolation in global polynomials:

$$\text{comp. cost} \lesssim \varepsilon^{-2}|\ln(\varepsilon)|^4 + |\ln(\varepsilon)|^3 \begin{cases} \varepsilon^{-2}\,, & \beta > \gamma\,, \\ \varepsilon^{-2}\ln(\varepsilon)^2\,, & \beta = \gamma\,, \\ \varepsilon^{-\left(2+\frac{\gamma-\beta}{\alpha}\right)}|\ln(\varepsilon)|^{\frac{\gamma-\beta}{\alpha}}\,, & \beta < \gamma\,. \end{cases}$$

Theorem (cont.)

If $\Phi \in C^{k+1}(\Theta)$ and $m \leq k$, then there exists $\hat{\Phi}_L$ such that $\mathrm{MSE}\left(\hat{\Phi}_L^{(m)}\right) = \mathcal{O}(\varepsilon^2)$:

$$\text{comp. cost (no interp. cost)} \lesssim |\ln(\varepsilon)| \begin{cases} \varepsilon^{-2\frac{k+1}{k+1-m}}, & \beta > \gamma\,, \\ \varepsilon^{-2\frac{k+1}{k+1-m}} \ln(\varepsilon)^2, & \beta = \gamma\,, \\ \varepsilon^{-\left(2+\frac{\gamma-\beta}{\alpha}\right)\frac{k+1}{k+1-m}}, & \beta < \gamma\,. \end{cases}$$

- Result applies to the approximation of CDF, quantiles and CVaR with $m = 1$, and with $m = 2$ for the PDF.

Theorem (cont.)

If $\Phi \in C^{k+1}(\Theta)$ and $m \leq k$, then there exists $\hat{\Phi}_L$ such that $\mathrm{MSE}\left(\hat{\Phi}_L^{(m)}\right) = \mathcal{O}(\varepsilon^2)$:

$$\text{comp. cost (no interp. cost)} \lesssim |\ln(\varepsilon)| \begin{cases} \varepsilon^{-2\frac{k+1}{k+1-m}} , & \beta > \gamma , \\ \varepsilon^{-2\frac{k+1}{k+1-m}} \ln(\varepsilon)^2 , & \beta = \gamma , \\ \varepsilon^{-\left(2+\frac{\gamma-\beta}{\alpha}\right)\frac{k+1}{k+1-m}} , & \beta < \gamma . \end{cases}$$

- Result applies to the approximation of CDF, quantiles and CVaR with $m = 1$, and with $m = 2$ for the PDF.
- Again, a similar complexity result is available for Φ analytic:

$$\text{comp. cost} \lesssim \varepsilon^{-2}|\ln(\varepsilon)|^{4(1+m)} + |\ln(\varepsilon)|^{3+4m} \begin{cases} \varepsilon^{-2} , & \beta > \gamma , \\ \varepsilon^{-2} \ln(\varepsilon)^2 , & \beta = \gamma , \\ \varepsilon^{-\left(2+\frac{\gamma-\beta}{\alpha}\right)} |\ln(\varepsilon)|^{\frac{\gamma-\beta}{\alpha}(1+2m)} , & \beta < \gamma . \end{cases}$$

MLMC for CDF, quantile, and CVaR: error control

- Recall that key quantities such as the CDF, quantile, and CVaR are all derived from the function

$$\Phi(\vartheta) = \mathbb{E}\big(\phi(\vartheta, Q)\big) \quad \text{for} \quad \phi(\vartheta, Q) = \vartheta + \frac{1}{1-\tau}(Q-\vartheta)^+ .$$

- However, complexity result with $m = 1$ applies directly only to Φ'. In fact, it “only” guarantees that $\mathrm{MSE}\big(\hat{\Phi}'\big) = \mathcal{O}(\varepsilon^2)$. What about these other derived quantities?

MLMC for CDF, quantile, and CVaR: error control

- Recall that key quantities such as the CDF, quantile, and CVaR are all derived from the function

$$\Phi(\vartheta) = \mathbb{E}\big(\phi(\vartheta, Q)\big) \quad \text{for} \quad \phi(\vartheta, Q) = \vartheta + \frac{1}{1-\tau}(Q-\vartheta)^+ .$$

- However, complexity result with $m = 1$ applies directly only to Φ'. In fact, it "only" guarantees that $\mathrm{MSE}\big(\hat{\Phi}'\big) = \mathcal{O}(\varepsilon^2)$. What about these other derived quantities?

Corollary

For $\tau \in (0, 1)$, let $\hat{\Phi}$ be the MLMC estimator of $\Phi \in C^{k+1}(\Theta)$, $k \geq 1 = m$, so that $\mathrm{MSE}\big(\hat{\Phi}'\big) = \mathcal{O}(\varepsilon^2)$. If the true τ-quantile is an interior point of Θ, then

$$\max\{\text{quantile MSE}, \text{CVaR MSE}, \text{unif. CDF MSE}\} = \mathcal{O}(\varepsilon^2) \ ,$$

at a cost dominated by $\mathrm{cost}\big(\hat{\Phi}'\big) = \mathrm{cost}(\text{CDF est.})$.

Corollary provides an all-at-once approach for the simultaneous approximation of CDF, quantiles, and CVaR.

Toy example: the characteristic function

- Let's consider the *toy model* to describe a European call option again, i.e., asset follows

$$dS = rS\,dt + \sigma S\,dW \ , \quad S(0) = S_0 \ ,$$

- Quantity of interest is the discounted "payoff": $Q := e^{-rT}\max\big(S(T) - K, 0\big)$

Toy example: the characteristic function

- Let's consider the *toy model* to describe a European call option again, i.e., asset follows

$$dS = rS\,dt + \sigma S\,dW\ , \quad S(0) = S_0\ ,$$

- Quantity of interest is the discounted "payoff": $Q := e^{-rT}\max\big(S(T) - K, 0\big)$
- **But**: Q has a *mixed distribution*, in the sense that $\mathbb{P}(Q = 0) > 0$.
- Consequently, its CDF $F_Q := \mathbb{E}\big(I(Q \leq \cdot)\big)$ has a jump discontinuity at the origin:

$\rightsquigarrow$ we **cannot** guarantee a uniformly accurate MLMC CDF-approximation, if $0 \in \Theta$.

Toy example: the characteristic function

- Let's consider the *toy model* to describe a European call option again, i.e., asset follows

$$dS = rS\,dt + \sigma S\,dW\ , \quad S(0) = S_0\ ,$$

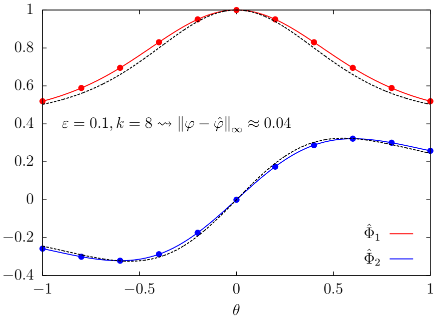
- Quantity of interest is the discounted "payoff": $Q := e^{-rT}\max\big(S(T) - K, 0\big)$
- **But**: Q has a *mixed distribution*, in the sense that $\mathbb{P}(Q = 0) > 0$.
- Consequently, its CDF $F_Q := \mathbb{E}\big(I(Q \leq \cdot)\big)$ has a jump discontinuity at the origin:

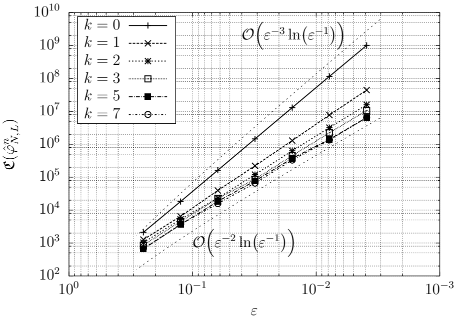
  $\rightsquigarrow$ we **cannot** guarantee a uniformly accurate MLMC CDF-approximation, if $0 \in \Theta$.
- But we **can** guarantee an accurate MLMC approximation of the **characteristic function** to characterize probability distribution of Q:

$$\varphi_Q(\vartheta) = \mathbb{E}\big(\underbrace{\cos(\vartheta Q)}_{=:\phi_1(\vartheta,Q)}\big) + \mathrm{i}\,\mathbb{E}\big(\underbrace{\sin(\vartheta Q)}_{=:\phi_2(\vartheta,Q)}\big) \equiv \Phi_1(\vartheta) + \mathrm{i}\,\Phi_2(\vartheta)\ ,$$

- **NB:** functions ϕ_i are smooth, no derivatives required (i.e., $m = 0$), moment approximations via post-processing possible [K., Nobile. 2018].

- Milstein scheme with $h_\ell = 2^{-\ell}T$; $\Theta = [-1, 1]$, $r = \frac{1}{20}$, $\sigma = \frac{1}{5}$, $T = 1$, $K = 10 = S_0$.





PDE toy example

Consider a simple Poisson equation

$$-\Delta u = f \,, \quad \text{in } D = (0,1)^2 \,,$$

with homogeneous Dirichlet boundary conditions and random forcing term f given by $f(x) = -72\xi({x_1}^2 + {x_2}^2 - x_1 - x_2)$, $\xi \sim \chi_1^2$. **Quantity of interest:** $Q := \int_D u \, dx = \xi$.

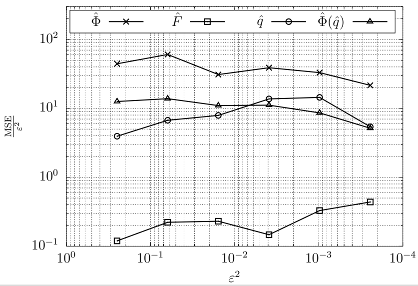
PDE toy example

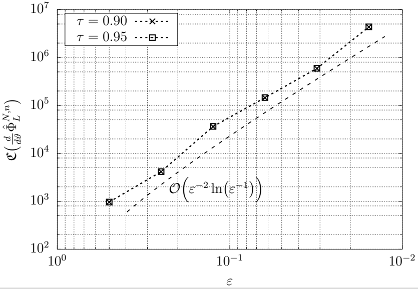
Consider a simple Poisson equation

$$-\Delta u = f \,, \quad \text{in } D = (0,1)^2 \,,$$

with homogeneous Dirichlet boundary conditions and random forcing term f given by $f(x) = -72\xi({x_1}^2 + {x_2}^2 - x_1 - x_2)$, $\xi \sim \chi_1^2$. **Quantity of interest:** $Q := \int_D u\,dx = \xi$.

We approximate $\Phi(\vartheta) = \vartheta + \frac{1}{1-\tau}\mathbb{E}\left[(Q-\vartheta)^+\right]$ with $\tau = 0.95$ on $\vartheta \in \Theta = [0,10]$ for $k = 3 \rightsquigarrow$ theory predicts cost $= \mathcal{O}\left(\varepsilon^{-2.3}\ln(\varepsilon^{-1})\right)$, appears to be conservative.





Practical bottleneck: a-priori bounds

Asymptotic complexity analysis is based on **a-priori upper bounds** for error estimates:

$$\frac{\mathrm{MSE}\big(\hat{\Phi}_L^{(m)}\big)}{3} \leq \|\Phi^{(m)} - \mathcal{I}_n^{(m)}(\Phi)\|_{L^\infty}^2 + \|\mathcal{I}_n^{(m)}(\Phi - \Phi_L)\|_{L^\infty}^2 + \mathbb{E}\big(\|\mathcal{I}_n^{(m)}(\Phi_L - \Phi_L^{\mathtt{MLMC}})\|_{L^\infty}^2\big)$$

$$\leq C_1(m)^2 n^{-2(k+1-m)} + C_2(m)^2(n-1)^{2m} {b_L}^2 + C_2(m)^2(n-1)^{2m}\log(n)\sum_{\ell=0}^{L}\frac{V_\ell}{N_\ell}$$

where $b_L = \|\Phi(\boldsymbol{\theta}) - \Phi_L(\boldsymbol{\theta})\|_{\ell^\infty}$ and $V_\ell = \mathrm{Var}\big(\phi(\boldsymbol{\theta}, Q_\ell) - \phi(\boldsymbol{\theta}, Q_{\ell-1})\big)$.

Practical bottleneck: a-priori bounds

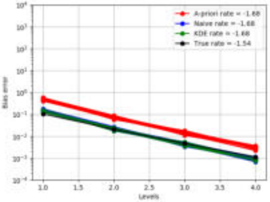
Asymptotic complexity analysis is based on **a-priori upper bounds** for error estimates:

$$\frac{\mathrm{MSE}\big(\hat{\Phi}_L^{(m)}\big)}{3} \leq \big\|\Phi^{(m)} - \mathcal{I}_n^{(m)}(\Phi)\big\|_{L^\infty}^2 + \big\|\mathcal{I}_n^{(m)}(\Phi - \Phi_L)\big\|_{L^\infty}^2 + \mathbb{E}\Big(\big\|\mathcal{I}_n^{(m)}(\Phi_L - \Phi_L^{\mathrm{MLMC}})\big\|_{L^\infty}^2\Big)$$

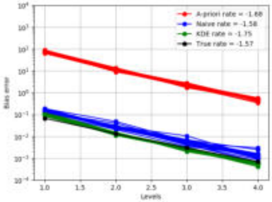
$$\leq C_1(m)^2 n^{-2(k+1-m)} + C_2(m)^2 (n-1)^{2m} b_L{}^2 + C_2(m)^2 (n-1)^{2m} \log(n) \sum_{\ell=0}^{L} \frac{V_\ell}{N_\ell}$$

where $b_L = \|\Phi(\theta) - \Phi_L(\theta)\|_{\ell^\infty}$ and $V_\ell = \mathrm{Var}\big(\phi(\theta, Q_\ell) - \phi(\theta, Q_{\ell-1})\big)$.

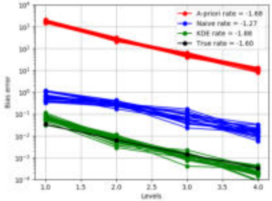
In particular, the *inverse inequality* makes error bound **severely conservative**. For example, the bias decay:



$m = 0$



$m = 1$



$m = 2$

Refined a-posterior error estimators I: bias

- **Starting point:** derive error estimators based on *first error splitting* directly:
- *Simplifications*: interpolation in cubic Splines

Refined a-posterior error estimators I: bias

- **Starting point:** derive error estimators based on *first error splitting* directly:
- *Simplifications*: interpolation in cubic Splines
- **Error estimator for bias via bivariate kernel smoothing** of empirical measure of (Q_L, Q_{L-1}):

$$\left\|\mathcal{I}_n^{(m)}(\Phi - \Phi_L)\right\|_{L^\infty}^2 \approx \frac{1}{s^\alpha - 1}\left\|\mathcal{I}_n^{(m)}(\mathbb{E}(\phi_L - \phi_{L-1}))\right\|_{L^\infty}$$

$$\left\|\mathcal{I}_n^{(m)}(\mathbb{E}(\phi_L - \phi_{L-1}))\right\|_{L^\infty} \approx \left\|\mathcal{I}_n^{(m)}\left(\mathbb{E}^{\mathrm{kde}}\left(\phi_L - \phi_{L-1}\right)\right)\right\|_{L^\infty}$$

where the *bivariate KDE approximates* the joint PDF of $(Q_\ell, Q_{\ell-1})$ using the N_ℓ correlated samples.

Refined a-posterior error estimators I: bias

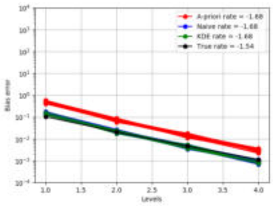
- **Starting point:** derive error estimators based on *first error splitting* directly:
- *Simplifications*: interpolation in cubic Splines
- **Error estimator for bias via bivariate kernel smoothing** of empirical measure of (Q_L, Q_{L-1}):

$$\left\|\mathcal{I}_n^{(m)}(\Phi - \Phi_L)\right\|_{L^\infty}^2 \approx \frac{1}{s^\alpha - 1}\left\|\mathcal{I}_n^{(m)}(\mathbb{E}(\phi_L - \phi_{L-1}))\right\|_{L^\infty}$$

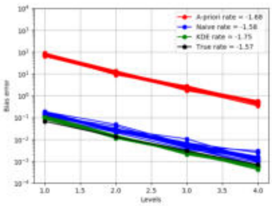
$$\left\|\mathcal{I}_n^{(m)}(\mathbb{E}(\phi_L - \phi_{L-1}))\right\|_{L^\infty} \approx \left\|\mathcal{I}_n^{(m)}\left(\mathbb{E}^{\text{kde}}\left(\phi_L - \phi_{L-1}\right)\right)\right\|_{L^\infty}$$

where the *bivariate KDE approximates* the joint PDF of $(Q_\ell, Q_{\ell-1})$ using the N_ℓ correlated samples.

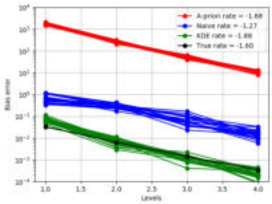
A-posteriori bias error estimation:



$m = 0$



$m = 1$



$m = 2$

Refined a-posterior error estimators II: statistical error

Idea: Statistical error estimator via bootstrapping MLMC estimators:

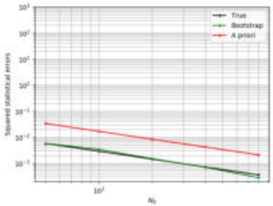
- "Observation": an MLMC estimator is defined through the hierarchy of samples

$$\left\{ \left\{ Q_\ell^{(i,\ell)}, Q_{\ell-1}^{(i,\ell)} \right\}_{i=1}^{N_\ell} \right\}_{\ell=0}^{L}$$

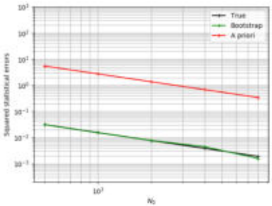
- Bootstrap principle: resample $N_{\mathrm{bs}} \gg 1$ "new" MLMC estimators Ψ_j

$$\mathbb{E}\left(\left\| \mathcal{I}_n^{(m)} (\Phi_L - \Phi_L^{\mathrm{MLMC}}) \right\|_{L^\infty}^2 \right) \approx \frac{1}{N_{\mathrm{bs}}} \sum_{j=1}^{N_{\mathrm{bs}}} \left\| \mathcal{I}_n^{(m)} \left(\Psi_j(\theta) - \bar{\Psi}(\theta) \right\|_{L^\infty}^2$$

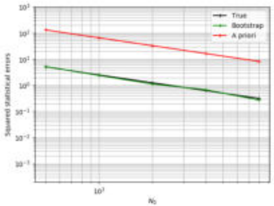
A-posteriori statistical error estimation:



$m = 0$



$m = 1$



$m = 2$

Implication for an (adaptive) MLMC implementation

- A-priori bounds yield error estimator that is convenient for an (adaptive) implementation:

$$\mathbb{E}\left(\left\|\mathcal{I}_n^{(m)}(\Phi_L - \Phi_L^{\mathtt{MLMC}})\right\|_{L^\infty}^2\right) \leq K(n,m)\sum_{\ell=0}^{L}\frac{V_\ell}{N_\ell} \approx K(n,m)\sum_{\ell=0}^{L}\frac{\hat{V}_\ell}{N_\ell}$$

- The a-posterior error bound does not provide level-wise errors

$$\mathbb{E}\left(\left\|\mathcal{I}_n^{(m)}(\Phi_L - \Phi_L^{\mathtt{MLMC}})\right\|_{L^\infty}^2\right) \approx \frac{1}{N_{\text{bs}}}\sum_{j=1}^{N_{\text{bs}}}\left\|\mathcal{I}_n^{(m)}\left(\Psi_j(\boldsymbol{\theta}) - \bar{\Psi}(\boldsymbol{\theta})\right)\right\|_{L^\infty}^2 =: \left(\hat{e}_{\text{a-post}}^{(m)}\right)^2$$

- Rescaling idea: we aim for redefined levelwise variances $\tilde{V}_\ell$ such that $\left(\hat{e}_{\text{a-post}}^{(m)}\right)^2 = \sum_{\ell=0}^{L}\frac{\tilde{V}_\ell}{N_\ell}$

Lemma

There exist positive constants $\mathfrak{K}_1(n)$ and $\mathfrak{K}_2(n)$: $\mathfrak{K}_1(n)\sum_{\ell=0}^{L}\frac{V_\ell}{N_\ell} \leq \sum_{m=0}^{2} k_m\left(\hat{e}_{\text{exact}}^{(m)}\right)^2 \leq \mathfrak{K}_2(n)\sum_{\ell=0}^{L}\frac{V_\ell}{N_\ell}$

- Rescaling heuristic: $\tilde{V}_\ell = r_e\hat{V}_\ell\ , \quad r_e = \dfrac{\sum_{m=0}^{2} k_m\left(\hat{e}_{\text{a-post}}^{(m)}\right)^2}{\sum_{k=0}^{L}\hat{V}_k/N_k}$

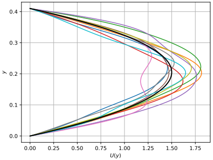
Computational Example

In [Ayoul-Guilmarda, Ganesh, K., Nobile. 2023] we combine these with an adaptive, *continuation MLMC framework* for an efficient implementation (Python library `XMC`) that can be tuned to target CDF, VaR, or CVaR.

Computational Example

In [Ayoul-Guilmarda, Ganesh, K., Nobile. 2023] we combine these with an adaptive, *continuation MLMC framework* for an efficient implementation (Python library `XMC`) that can be tuned to target CDF, VaR, or CVaR.

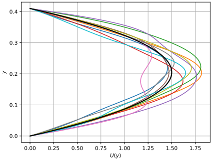
Example: (steady incompressible) Navier-Stokes Flow over a Cylinder in a Channel

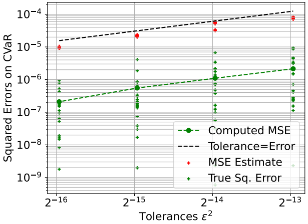


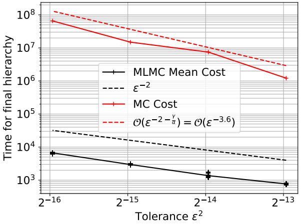
Computational Example

In [Ayoul-Guilmarda, Ganesh, K., Nobile. 2023] we combine these with an adaptive, *continuation MLMC framework* for an efficient implementation (Python library XMC) that can be tuned to target CDF, VaR, or CVaR.

Example: (steady incompressible) Navier-Stokes Flow over a Cylinder in a Channel







Take home message:

- we introduced a uniformly accurate MLMC estimator for a parametric expectation and its derivatives
- the approach enables approximating the characteristic function as well as CDF, PDF, VaR, and CVaR
- refined, a-posterior error estimates are required for computationally “heavy” problems when derivatives are required.

Take home message:

- we introduced a uniformly accurate MLMC estimator for a parametric expectation and its derivatives
- the approach enables approximating the characteristic function as well as CDF, PDF, VaR, and CVaR
- refined, a-posterior error estimates are required for computationally "heavy" problems when derivatives are required.

Thank you for your attention.

Further details:

SK and F. Nobile. Multilevel Monte Carlo Approximation of Functions. *SIAM/ASA J. Uncertain. Quantif.*, **3**(6):1256–1293, 2018.

Q. Ayoul-Guilmard, S. Ganesh, SK, and F. Nobile. Quantifying uncertain system outputs via the multi-level Monte Carlo method – distribution and robustness measures. *Int. J. Uncertain. Quantif.*, **13**(5):61–98, 2023.