Sub-sampling and other considerations for efficient risk estimation in large portfolios

Mike Giles (Oxford)   Abdul-Lateef Haji-Ali (Heriot-Watt)

ESI Online Workshop
Multilevel and multifidelity sampling methods in UQ for PDEs

5 May 2020

Outline

- Risk applications and nested expectation.
- Problem setting and a Monte Carlo estimator.
- Estimating a sum.
- Multilevel Monte Carlo (MLMC) for nested expectations.
- Adaptive sampling.
- Results for a test problem.
- Other considerations.
- Concluding remarks.

Risk analysis

- Stochastic models are increasingly being adopted in real-life applications.
- An important question in such applications is assessing the risk of some extreme event:
  - in finance: risk of loss, default or ruin,
  - in industrial modelling: risk of component failure,
  - in crowd modelling: risk of stampede,
  - in workshop organization: risk of a pandemic,
  - ...
- Risk assessment is the first step to risk management.
- Computing **risk measures** is computationally difficult because
  - extreme events are extremely rare,
  - the risk measures are not smooth (either the event happened or not),
  - and the underlying stochastic models are difficult to evaluate (or expensive to approximate).
- In this talk, we address the last two points.

Nested expectation in risk applications

- The "losses" are modelled by P random variables $\{X_i\}_{i=1}^P$.
- $\{X_i\}_{i=1}^P$ depend on another (multi-dimensional) random variable Y, the risk factor.
- The **expected loss** for a given risk factor is

$$\Lambda = \mathbb{E}\left[\frac{1}{P} \sum_{i=1}^{P} X_i \,\middle|\, Y \right].$$

- We are interested in computing **probability of the expected loss** exceeding Λ_η as

$$\eta = \mathbb{P}[\, \Lambda > \Lambda_\eta \,] = \mathbb{E}\left[\mathrm{H}\left(\mathbb{E}\left[\frac{1}{P} \sum_{i=1}^{P} X_i - \Lambda_\eta \,\middle|\, Y \right] \right) \right]$$

where $\mathrm{H}(\cdot)$ is the Heaviside step function.

- **Key message**: the probability of a large expected loss involves a nested expectation.

Nested estimation using Monte Carlo

$$\mathbb{E}\left[\mathrm{H}\left(\mathbb{E}\left[\frac{1}{P} \sum_{i=1}^{P} X_i \,\middle|\, Y \right] \right) \right]$$

We use N inner samples of $\{X_i\}_{i=1}^{P}$ to estimate $\mathrm{H}(\mathbb{E}[\,X \mid Y\,]) \approx \mathrm{H}\big(\overline{X}_N(Y)\big)$ where

$$\overline{X}_N(Y) = N^{-1} P^{-1} \sum_{n=1}^{N} \sum_{i=1}^{P} X_i^{(n)}(Y)$$

This leads to a **bias** of $\mathcal{O}(\boldsymbol{P^{-1} N^{-1}})$. Using Monte Carlo for the outer expectation as well,

$$\mathbb{E}\big[\, \mathrm{H}(\mathbb{E}[\,X \mid Y\,]) \,\big] \approx \frac{1}{M} \sum_{m=1}^{M} \mathrm{H}\Big(\overline{X}_N(Y^{(m)})\Big)$$

leads to a **sampling error** of $\mathcal{O}(\boldsymbol{M^{-1/2}})$.

Nested estimation using Monte Carlo

To achieve a root mean-square error ε choose

$$N = \max\big(1, \mathcal{O}(P^{-1}\varepsilon^{-1})\big)$$
$$M = \mathcal{O}(\varepsilon^{-2})$$

Cost of nested Monte Carlo estimator is $M \times N \times P$.

Hence, complexity is $\mathcal{O}(\mathbf{max}\big(\boldsymbol{P\varepsilon^{-2}}, \boldsymbol{\varepsilon^{-3}}\big))$.

Ideally we would like the complexity to be $\mathcal{O}(\varepsilon^{-2})$, independently of P. Hence we will

- devise a strategy to sample the sum with a complexity that is independent of P so that the complexity is $\mathcal{O}(\varepsilon^{-3})$.
- use MLMC to reduce the complexity to almost $\mathcal{O}(\varepsilon^{-2})$.

Estimating a sum

Recall that we have to compute $\frac{1}{P}\sum_{p=1}^{P} X_i$ for every sample of the risk factors, Y. Here, we focus on a single computation for a single risk scenario.

Using a random sub-sampler, we can approximate

$$\frac{1}{P}\sum_{p=1}^{P} X_i = \frac{1}{P}\mathbb{E}[\, X_j \; p_j^{-1}\,] \approx \frac{1}{PN}\sum_{n=1}^{N} X_{j^{(n)}}^{(n)} \; p_{j^{(n)}}^{-1}$$

where j is a random integer with $\mathbb{P}[\, j = i\,] = p_i$ for $i \in \{1, \ldots, P\}$.
The cost of this random sub-sampler is N while the MSE is bounded by

$$N^{-1} P^{-2} \sum_{i=1}^{P} \mathbb{E}[\, X_i^2\,] p_i^{-1}$$

Estimating a sum

Minimizing the MSE leads to the optimal expression for the probabilities

$$p_i = \widetilde{g}_i \Big/ \sum_{k=1}^{P} \widetilde{g}_k$$

for $\widetilde{g}_i^2 \approx \mathbb{E}[\, X_i^2 \,]$ and the optimal MSE

$$N^{-1} P^{-2} \left(\sum_{i=1}^{P} \frac{\mathbb{E}[\, X_i^2 \,]}{\widetilde{g}_i} \right) \left(\sum_{i=1}^{P} \widetilde{g}_i \right) \approx N^{-1} \left(P^{-1} \sum_{i=1}^{P} \widetilde{g}_i \right)^2 = \mathcal{O}(N^{-1})$$

which is bounded for all P.

In nested expectation

Hence, we write

$$\mathbb{E}\left[\mathrm{H}\left(\mathbb{E}\left[\frac{1}{P} \sum_{i=1}^{P} X_i \,\middle|\, Y \right] \right) \right] = \mathbb{E}[\, \mathrm{H}(\mathbb{E}[\, X \mid Y \,]) \,]$$

where

$$X = P^{-1} \; X_j \; p_j^{-1}$$

and

$$p_j = \widetilde{g}_j \Big/ \sum_{k=1}^{P} \widetilde{g}_k$$

for some sequence $\widetilde{g}_k$ independent of Y, e.g., $\widetilde{g}_k = \mathbb{E}[\, X_k^2 \,]$ so that the optimal probabilities have to computed only once.

Using the random sub-sampler, the computational complexity is independent of the number of terms P. Moreover, in some cases it can be reduced by a constant by using a “mixed” sub-sampler.

MLMC for nested expectation

Next, we want to apply MLMC to nested expectation to reduce the overall complexity from $\mathcal{O}(\varepsilon^{-3})$ to $\mathcal{O}(\varepsilon^{-2})$.
Building a hierarchy of $L+1$ estimators with N_ℓ inner samples for $\ell = 0, 1, \ldots, L$, the MLMC estimator is

$$\mathbb{E}[\,Q\,] = \sum_{\ell=0}^{\infty} \mathbb{E}[\,\Delta_\ell Q\,] \approx \sum_{\ell=0}^{L} \frac{1}{M_\ell} \sum_{m=1}^{M_\ell} \Delta_\ell Q^{(\ell,m)},$$

where

$$\begin{aligned} Q &= \mathrm{H}(\mathbb{E}[\,X \mid Y\,]), \\ Q_\ell &= \mathrm{H}\big(\overline{X}_{N_\ell}(Y)\big), \\ \Delta_\ell Q^{(\ell,m)} &= Q_\ell^{(\ell,m)} - Q_{\ell-1}^{(\ell,m)} \\ &= \mathrm{H}\Big(\overline{X}_{N_\ell}(Y^{(\ell,m)})\Big) - \mathrm{H}\Big(\overline{X}_{N_{\ell-1}}(Y^{(\ell,m)})\Big), \end{aligned}$$

and $Q_{-1} = 0$.

Multilevel Monte Carlo: summary

For $Q \approx Q_\ell$ and $\Delta_\ell Q = Q_\ell - Q_{\ell-1}$ with $Q_{-1} = 0$, we have

$$\mathbb{E}[\,Q\,] = \sum_{\ell=0}^{\infty} \mathbb{E}[\,\Delta_\ell Q\,] \approx \sum_{\ell=0}^{L} \mathbb{E}[\,\Delta_\ell Q\,] \approx \sum_{\ell=0}^{L} \frac{1}{M_\ell} \sum_{m=1}^{M_\ell} \Delta_\ell Q^{(\ell,m)}$$

where $\Delta_\ell Q^{(\ell,m)}$ is the (ℓ, m)'th samples of $\Delta_\ell Q$. Assuming

$$\begin{aligned}
&\text{Bias:} && |\mathbb{E}[\,Q - Q_\ell\,]| = \mathcal{O}\left(2^{-\alpha\ell}\right),\\
&\text{Variance:} && V_\ell = \mathrm{Var}[\,\Delta_\ell Q\,] = \mathcal{O}\left(2^{-\beta\ell}\right),\\
&\text{Work:} && W_\ell = \mathcal{O}(2^{\gamma\ell}),
\end{aligned}$$

for $\alpha, \beta, \gamma \in \mathbb{R}_+$ and where the work to sample $\Delta_\ell Q$ is W_ℓ, then there are optimal choices of L and M_ℓ such that the MLMC estimator has complexity

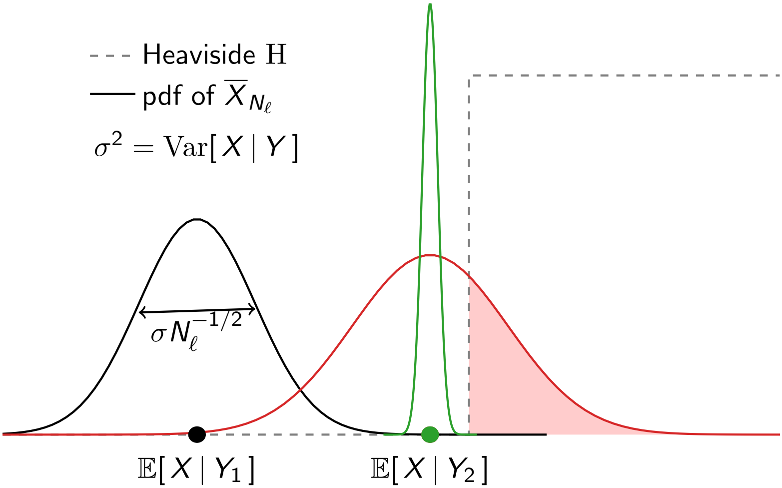
$$|\log \varepsilon|^p \varepsilon^{-2-\max\left(0, \frac{\gamma-\beta}{\alpha}\right)} \qquad \text{c.f. MC:} \quad \mathcal{O}(\varepsilon^{-2-\frac{\gamma}{\alpha}})$$

$p = 2$ if $\beta = \gamma$ and zero otherwise

Choice of N_ℓ: Need for adaptivity

$$Q := \mathrm{H}(\mathbb{E}[\, X \mid Y \,]) \approx \mathrm{H}\big(\overline{X}_{N_\ell}(Y)\big) =: Q_\ell$$

$$\text{adaptive } N_\ell \implies \quad \gamma = 1, \quad \alpha = 1, \quad \beta = 1$$



MLMC + adaptive inner sampling

Let

$$d = |\mathbb{E}[\,X \mid Y\,]|, \qquad \sigma^2 = \mathrm{Var}[\,X \mid Y\,], \qquad \delta = d/\sigma$$

We will instead use the following number of inner samples:

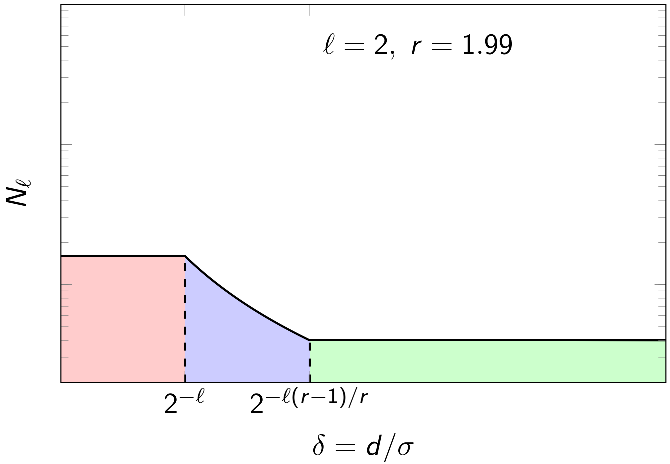
$$N_\ell = \max\Big(2^\ell, 4^\ell \min(1, (2^\ell \delta)^{-r})\Big), \quad 1 < r < 2,$$

Note

$$2^\ell \le N_\ell \le 4^\ell.$$

MLMC + adaptive inner sampling

$$N_\ell = \max\Big(2^\ell, 4^\ell \min(1, (2^\ell \delta)^{-r})\Big), \quad 1 < r < 2,$$



Numerical analysis

Problem: In practice $\delta = d/\sigma$ is *unknown*, so the real adaptive algorithm has to use Monte Carlo estimates for $\widehat{d}$ and $\widehat{\sigma}$ to compute N_ℓ.

Algorithm: For a given outer sample Y, starting with the minimum, $N_\ell = 2^\ell$, keep doubling the number of inner samples, N_ℓ, until it is large enough based on current estimate $\widehat{\delta} = \widehat{d}/\widehat{\sigma}$, i.e.,

$$N_\ell \geq 4^\ell \; (2^\ell \widehat{\delta})^{-r},$$

or it reaches the maximum, 4^ℓ.

Concerns:

- If we use too many samples, the cost may be larger than we want.
- If we use too few samples, the variance may be larger than we want.

The main idea of the analysis is to prove that the probability of ending up with the "*wrong*" number of inner samples decays very rapidly as you move away from the "*right*" number, that we get if we use the exact δ.

MLMC + adaptive inner sampling

Theorem (main result on output of adaptive algorithm)

Provided

1. *the random variable* $\delta = d/\sigma$ *has bounded density near 0,*
2. *there exists* $q > 2$ *such that*

$$\sup_y \left\{ \mathbb{E}\left[\left(\frac{|X - \mathbb{E}[X|Y]|}{\sigma} \right)^q \,\middle|\, Y = y \right] \right\} < \infty,$$

3. *and for*

$$1 < r < 2 - \frac{\sqrt{4q+1}-1}{q}$$

then using the adaptive algorithm with this r *to compute* N_ℓ *we have*

$$\mathbb{E}[\, N_\ell \,] = \mathcal{O}(2^\ell) \qquad \textit{and} \qquad V_\ell := \mathrm{Var}[\, \Delta_\ell Q \,] = \mathcal{O}(2^{-\ell})$$

Numerical example

We consider a portfolio of 1000 heterogeneous options on 16 correlated assets which follow Geometric Brownian Motions.

The risk variable, Y, contains the values of the underlying assets at some short risk horizon τ compared to the option maturities.

The exact probability of large loss that we are trying to compute is around 1%.

Numerical example, cont.

The losses can be computed by one of three methods:

- 30%: exact computation of the expected loss using the analytic solution of the Black-Scholes PDE (zero variance, unit cost)
- 50% – 70%: exact simulation of the asset values by analytically solve the SDE (non-zero variance, unit cost)
- 0% – 20%: approximate simulation using the Milstein numerical scheme to estimate the asset values (non-zero variance, cost depending on approximation). An unbiased estimator is constructed using Unbiased MLMC.

Methods

In this example, the loss can be written as

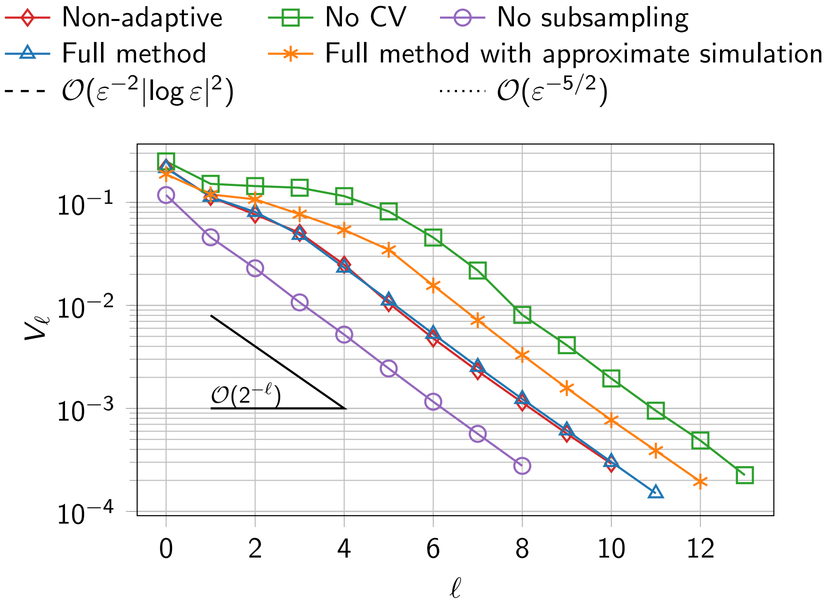
$$X_i = g_i(\widehat{S}_{i,T_i}; \tau, Y) - g_i(S_{i,T_i}; 0, S_0)$$

for some payoff function g_i.

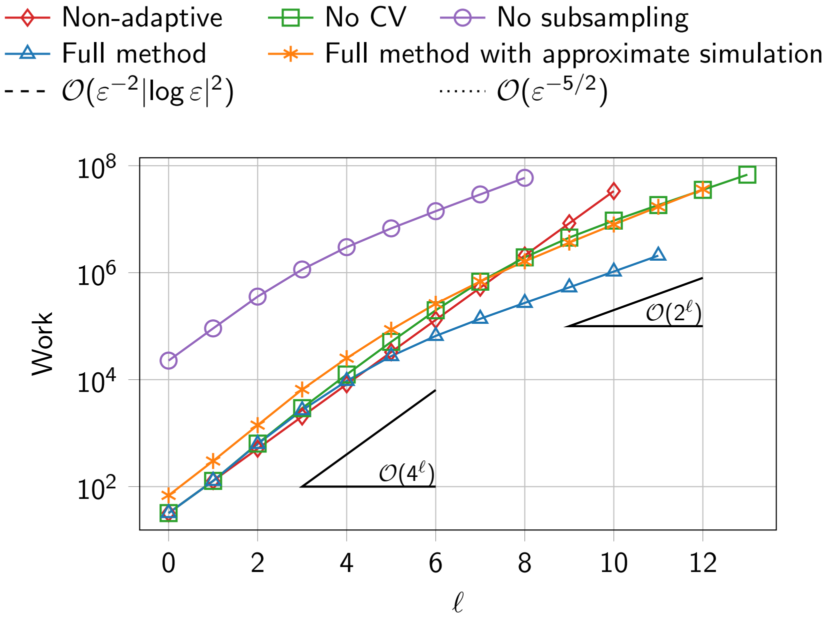
The method we consider uses MLMC with

- adaptive sampling,
- control variates that assume a short risk horizon τ
  - Delta: Linear model approximation with respect to risk parameter,
  - using the same Brownian path from risk horizon τ to maturity T_i,
  - and antithetic control variates based on the antithetic pairs of the Brownian Motion from 0 to risk horizon τ.
- random sub-sampling with optimal probabilities that depend on relative weight estimates,
- and an antithetic estimator for the MLMC differences based on the inner samples.

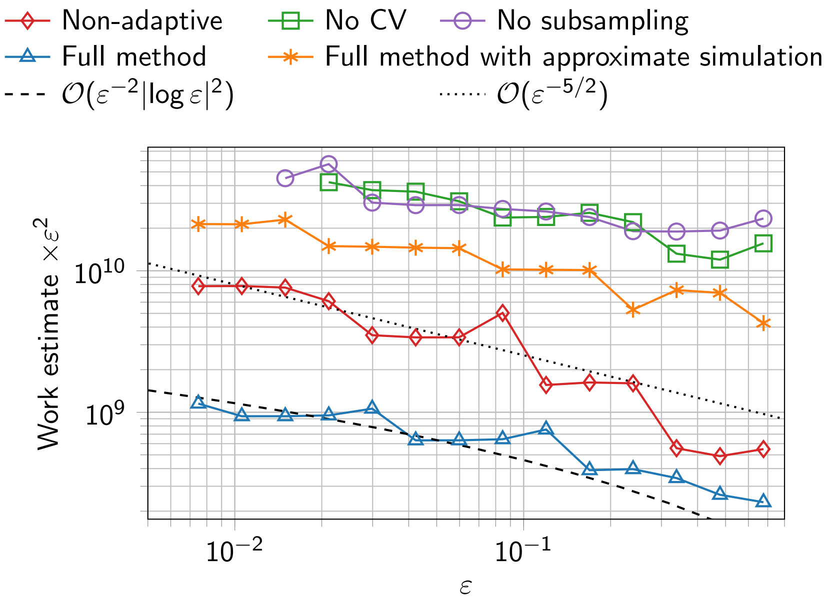
Results


Non-adaptive
No CV
No subsampling
Full method
Full method with approximate simulation
$\mathcal{O}(\varepsilon^{-2}|\log\varepsilon|^2)$
$\mathcal{O}(\varepsilon^{-5/2})$
10^{-1}
10^{-2}
10^{-3}
10^{-4}
V_ℓ
$\mathcal{O}(2^{-\ell})$
0
2
4
6
8
10
12
ℓ

Results

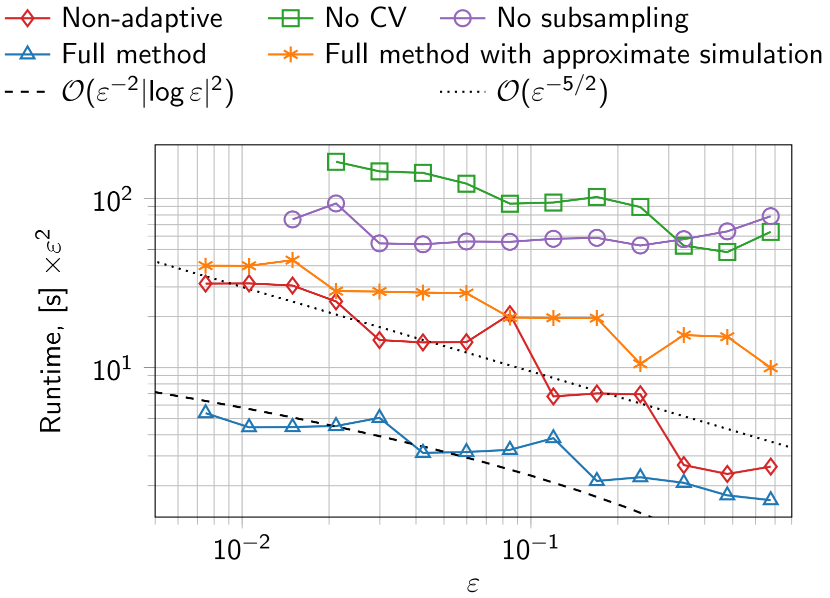

Non-adaptive
No CV
No subsampling
Full method
Full method with approximate simulation
$\mathcal{O}(\varepsilon^{-2}|\log\varepsilon|^2)$
$\mathcal{O}(\varepsilon^{-5/2})$
Work
10^8
10^6
10^4
10^2
$\mathcal{O}(2^\ell)$
$\mathcal{O}(4^\ell)$
0
2
4
6
8
10
12
ℓ

Results



Non-adaptive
No CV
No subsampling
Full method
Full method with approximate simulation
$\mathcal{O}(\varepsilon^{-2}|\log\varepsilon|^2)$
$\mathcal{O}(\varepsilon^{-5/2})$
Work estimate $\times\varepsilon^2$
10^{10}
10^{9}
10^{-2}
10^{-1}
ε

Results



Non-adaptive
No CV
No subsampling
Full method
Full method with approximate simulation
$\mathcal{O}(\varepsilon^{-2}|\log\varepsilon|^2)$
$\mathcal{O}(\varepsilon^{-5/2})$
10^2
10^1
Runtime, [s] $\times\varepsilon^2$
10^{-2}
10^{-1}
ε

Other risk measures: Value-at-Risk and Conditional VaR

- The Value-at-Risk (**VaR**), Λ_η, is defined implicitly by $\mathbb{P}[\Lambda > \Lambda_\eta] = \eta$.

  This can be estimated by a stochastic root-finding algorithm, with the acceptable error ε being steadily reduced during the iteration.

  Complexity is $\mathcal{O}(\varepsilon^{-2}|\log \varepsilon|^2)$.

- Given a VaR estimate, $\widetilde{\Lambda}_\eta$, the Conditional VaR (**CVaR**) is then

$$
\begin{aligned}
\mathbb{E}[\,\Lambda \mid \Lambda > \Lambda_\eta\,] &= \min_x \left\{ x + \eta^{-1} \mathbb{E}[\max(0, \Lambda - x)] \right\} \\
&= \widetilde{\Lambda}_\eta + \eta^{-1} \mathbb{E}[\max(0, \Lambda - \widetilde{\Lambda}_\eta)] + \mathcal{O}(\widetilde{\Lambda}_\eta - \Lambda_\eta)^2 \\
&= \widetilde{\Lambda}_\eta + \eta^{-1} \mathbb{E}[\max(0, \mathbb{E}[X \mid Y])] + \mathcal{O}(\widetilde{\Lambda}_\eta - \Lambda_\eta)^2.
\end{aligned}
$$

  Complexity is $\mathcal{O}(\varepsilon^{-2})$.

Other considerations

- When sufficient regularity exists and/or Y is not too high-dimensional, we can build an accurate surrogate model

$$h(y) \approx \mathbb{E}[\, X \mid Y = y \,]$$

and then use it to estimate $\mathbb{E}[\,\mathrm{H}(\mathbb{E}[\, X \mid Y \,])\,] \approx \mathbb{E}[\,\mathrm{H}(h(Y))\,]$ or use it as a control variate. For example, using weighted least square polynomial approximation, c.f.

A.-L. Haji-Ali, F. Nobile, R. Tempone, and S. Wolfers. "Multilevel weighted least squares polynomial approximation". In: *ESAIM: Mathematical Modelling and Numerical Analysis* 54.2 (2020), pp. 649–677. DOI: 10.1051/m2an/2019045.

and other standard references therein.

Other considerations

- If sampling X given Y has more than one approximation parameter (e.g., approximating high dimensional PDEs, or a discretization particle system) then the multi-index methodology can be employed, c.f. talk by J. Dick for PDE application.

  For general complexity analysis of MIMC in arbitrary number of dimensions (under variance, bias and work assumptions), see

A.-L. Haji-Ali, F. Nobile, and R. Tempone. "Multi-index Monte Carlo: when sparsity meets sampling". In: *Numerische Mathematik* 132 (4 2016), pp. 767–806. DOI: 10.1007/s00211-015-0734-5.

Key messages

- Risk estimation (and nested expectations) is a great new application area for MLMC.
- Keys to performance:
  - MLMC approach with more inner samples on "finer" levels,
  - adaptive number of inner samples,
  - sub-sampling with optimal probabilities to obtain a cost that is independent of the number of options.
  - control variates that reduce the total variance.
- More complicated underlying assets, requiring time discretization, are also handled using (unbiased) MLMC/MIMC $\longrightarrow$ nested MLMC.
- Options with heterogeneous work can be easily handled using random sub-sampling.

Talk references

- Analysis of sub-sampling, Unbiased MLMC and more details on the used control variates in

M. B. Giles and A.-L. Haji-Ali. "Sub-sampling and other considerations for efficient risk estimation in large portfolios". In: arXiv:1912.05484 (2019).

C++ Code: https://github.com/haji-ali/nested-risk-mlmc

- More complete analysis of the adaptive method and VaR/CVaR algorithms

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

- All references are in www.randomoid.com

Other references

K. Bujok, B. Hambly, and C. Reisinger. "Multilevel simulation of functionals of Bernoulli random variables with application to basket credit derivatives". In: *Methodology and Computing in Applied Probability* 17.3 (2015), pp. 579–604.

M. B. Giles. "Multilevel Monte Carlo methods". In: *Acta Numerica* 24 (2015), pp. 259–328.

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

M. B. Gordy and S. Juneja. "Nested simulation in portfolio risk measurement". In: *Management Science* 56.10 (2010), pp. 1833–1848.

M. B. Giles. "Multilevel Monte Carlo Path Simulation". In: *Operations Research* 56.3 (2008), pp. 607–617.

Multi-index Monte Carlo: idea

We wish to approximate

$$\mathbb{E}[\, Q\,]$$

but can only sample, increasingly accurate, approximations of Q denoted by $\{Q_{\boldsymbol{\ell}}\}_{\boldsymbol{\ell}\in\mathbb{N}^d}$.

We write the telescoping sum

$$\mathbb{E}[\, Q\,] = \sum_{\boldsymbol{\ell}\in\mathbb{N}^d} \mathbb{E}[\, \Delta_{\boldsymbol{\ell}} Q\,]$$

where

$$\Delta_{\ell_i,\ldots,\ell_d} Q_{\ell_1,\ldots,\ell_{i-1}} = \Delta_{\ell_{i+1},\ldots,\ell_d} Q_{\ell_1,\ldots,\ell_i} - \Delta_{\ell_{i+1},\ldots,\ell_d} Q_{\ell_1,\ldots,\ell_i-1}$$

Then we make two approximations:

1. truncate the sum to some index-set $\mathcal{I} \subset \mathbb{N}^d$, resulting in a biased estimator.
2. and approximate each expectation with Monte Carlo where the number of samples depend on $\boldsymbol{\ell}$.

Multi-index Monte Carlo: summary

$$\mathbb{E}[\,Q\,] = \sum_{\boldsymbol{\ell}\in\mathbb{N}^d} \mathbb{E}[\,\Delta_{\boldsymbol{\ell}} Q\,] \approx \sum_{\boldsymbol{\ell}\in\mathcal{I}} \mathbb{E}[\,\Delta_{\ell} Q\,] \approx \sum_{\boldsymbol{\ell}\in\mathcal{I}} \frac{1}{M_{\boldsymbol{\ell}}} \sum_{m=1}^{M_{\boldsymbol{\ell}}} \Delta_{\boldsymbol{\ell}} Q^{(\boldsymbol{\ell},m)}$$

where $\Delta_{\boldsymbol{\ell}} Q^{(\boldsymbol{\ell},m)}$ is the $(\boldsymbol{\ell}, m)$'th samples of $\Delta_{\boldsymbol{\ell}} Q$. Assuming

$$\begin{aligned}
&\text{Bias:} & |\mathbb{E}[\,\Delta_{\boldsymbol{\ell}} Q\,]| &= \mathcal{O}\left(2^{-\boldsymbol{\alpha}.\boldsymbol{\ell}}\right), \\
&\text{Variance:} & V_{\boldsymbol{\ell}} = \mathrm{Var}[\,\Delta_{\boldsymbol{\ell}} Q\,] &= \mathcal{O}\left(2^{-\boldsymbol{\beta}\cdot\boldsymbol{\ell}}\right), \\
&\text{Work:} & W_{\boldsymbol{\ell}} &= \mathcal{O}(2^{\boldsymbol{\gamma}\cdot\boldsymbol{\ell}}),
\end{aligned}$$

for $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma} \in \mathbb{R}_+^d$ and where the work to sample $\Delta_{\boldsymbol{\ell}} Q$ is $W_{\boldsymbol{\ell}}$, then there are optimal choices of $\mathcal{I}$ and $M_{\boldsymbol{\ell}}$ such that the MIMC estimator has complexity

$|\log \varepsilon|^{p'} \varepsilon^{-2-\max\left(0, \max_i \frac{\gamma_i - \beta_i}{\alpha_i}\right)}$ c.f. MLMC: $|\log \varepsilon|^{p} \varepsilon^{-2-\max\left(0, \frac{|\gamma| - \min_i \beta_i}{\min_i \alpha_i}\right)}$

e.g. $p' = p = 0$ if $\beta_i \geq \gamma_i$ for all i.

Optimal choice of $\mathcal{I}$

$$\mathcal{I}_\varepsilon = \left\{ \boldsymbol{\ell} \in \mathbb{N}^d \ : \ \boldsymbol{w} \cdot \boldsymbol{\ell} \lesssim \varepsilon^{-1} \right\}$$

$$w_i = \alpha_i + (\gamma_i - \beta_i)/2$$