Financial risk estimation using nested MLMC and portfolio sub-sampling

Mike Giles[1]   Abdul-Lateef Haji-Ali[2]

[1]Oxford University

[2]Heriot-Watt University

ICIAM
19 July, 2019

Outline

- Risk applications and nested expectation.
- Problem setting and a Monte Carlo estimator.
- Estimating a sum.
- Multilevel Monte Carlo (MLMC) for nested expectations.
- MLMC + uniform inner sampling.
- MLMC + adaptive inner sampling.
- 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,
  - ...
- 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 work, we address the last two points.

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,
  - ...
- 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 work, 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 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,

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

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

Nested estimation using Monte Carlo

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

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,

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

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

Nested estimation using Monte Carlo

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

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,

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

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 $M\ N\ P$.

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

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

which is bounded for all P.

In nested expectation

Hence, we write

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

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.

Mixed sub-sampler

To illustrate the need for mixed sub-sampling, consider the simple example

$$\frac{1}{P}\sum_{i=1}^{P} X_i$$

where all X_i terms are deterministic. A mixed estimator for $0 \leq Q \leq N$ is

$$\begin{aligned}\frac{1}{P}\sum_{p=1}^{P} X_i &= \frac{1}{P}\sum_{p=1}^{Q} X_i + \frac{1}{P}\ \mathbb{E}\Big[\, X_j\ p_j^{-1}\,\Big] \\ &\approx \frac{1}{P}\sum_{p=1}^{Q} X_i + \frac{1}{P(N-Q)}\sum_{n=1}^{N-Q} X_{j^{(n)}}\ p_{j^{(n)}}^{-1}\end{aligned}$$

where j is a random integer with $\mathbb{P}[\,j=i\,] = p_i$ for $i \in \{Q+1, \ldots, P\}$. When $Q=0$ we have a fully random sub-sampler and when $Q=P$ we are computing the sum exactly.

Mixed sub-sampler

Using the previous choice for p_i the MSE is bounded by

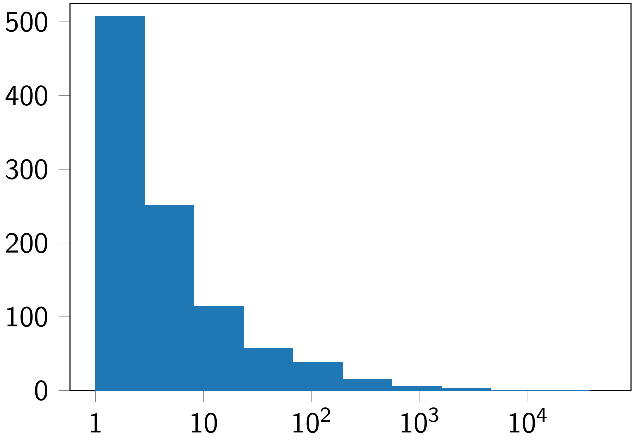
$$(\boldsymbol{N} - \boldsymbol{Q})^{-1} P^{-2} \left(\sum_{i=\boldsymbol{Q}+1}^{P} X_i \right)^2$$

since the error contribution is only due to the random sub-sampler. Hence, by sub-sampling the largest Q terms deterministically and optimizing with respect to Q, using a knapsack-type optimization, we can end up with a smaller MSE.

Numerical illustration

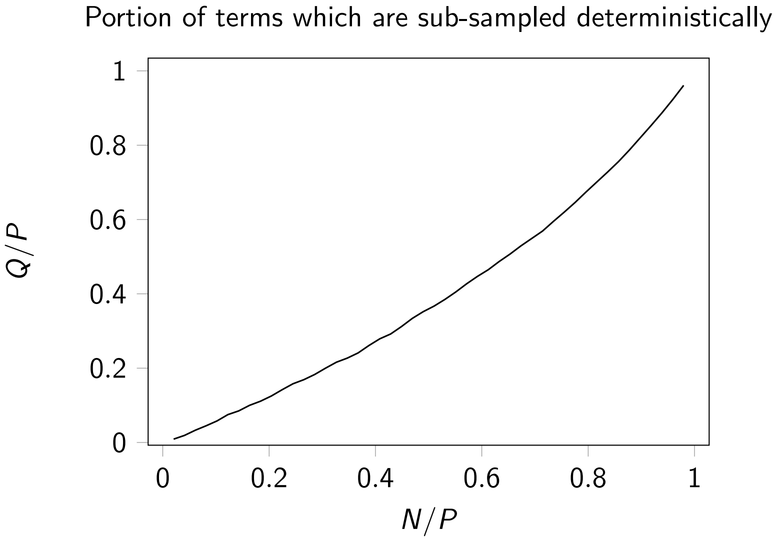
For $P = 1000$ and X_i^2 being i.i.d. samples from exponential distribution with rate 3.

Histogram of $\{X_i\}_{i=1}^{P}$



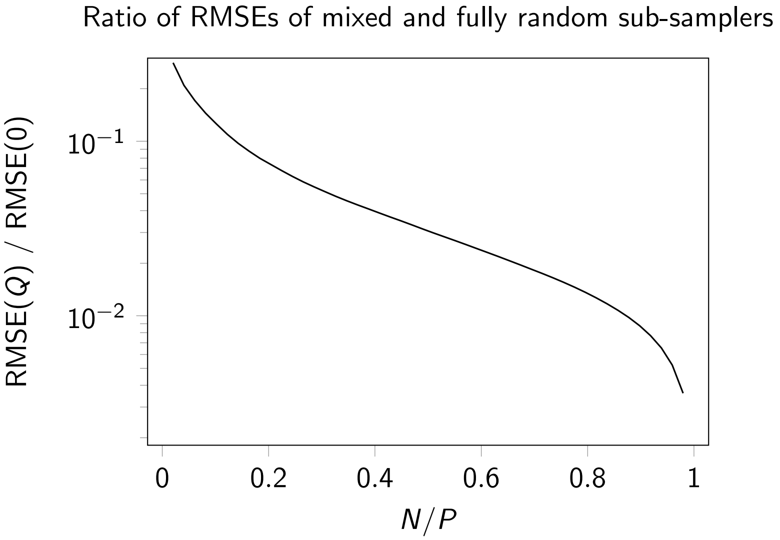
Numerical illustration

For $P = 1000$ and X_i^2 being i.i.d. samples from exponential distribution with rate 3.



Numerical illustration

For $P = 1000$ and X_i^2 being i.i.d. samples from exponential distribution with rate 3.



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

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

where

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

and $P_{-1} = 0$.

Multilevel Monte Carlo: summary

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

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

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

$$|\mathbb{E}[\,P - P_\ell\,]| = \mathcal{O}\left(2^{-\alpha\ell}\right),$$

$$V_\ell = \mathrm{Var}[\,\Delta_\ell P\,] = \mathcal{O}\left(2^{-\beta\ell}\right),$$

$$W_\ell = \mathcal{O}(2^{\gamma\ell}),$$

where the work to sample $\Delta_\ell P$ is W_ℓ, then there are optimal choices of L and M_ℓ so that the MLMC estimator has complexity

$$\begin{cases} \mathcal{O}\left(\varepsilon^{-2-\max\left(0,\frac{\gamma-\beta}{\alpha}\right)}\right), & \text{when } \gamma \neq \beta \\ \mathcal{O}\left(\varepsilon^{-2}|\log \varepsilon|^2\right) & \text{otherwise.} \end{cases}$$

Multilevel Monte Carlo: summary

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

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

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

$$|\mathbb{E}[\,P - P_\ell\,]| = \mathcal{O}\left(2^{-\alpha\ell}\right),$$
$$V_\ell = \mathrm{Var}[\,\Delta_\ell P\,] = \mathcal{O}\left(2^{-\beta\ell}\right),$$
$$W_\ell = \mathcal{O}(2^{\gamma\ell}),$$

where the work to sample $\Delta_\ell P$ is W_ℓ, then there are optimal choices of L and M_ℓ so that the MLMC estimator has complexity

$$\begin{cases} \mathcal{O}\left(\varepsilon^{-2-\max\left(0,\frac{\gamma-\beta}{\alpha}\right)}\right), & \text{when } \gamma \neq \beta \\ \mathcal{O}\left(\varepsilon^{-2}|\log\varepsilon|^2\right) & \text{otherwise.} \end{cases}$$

Multilevel Monte Carlo: summary

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

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

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

$$|\mathbb{E}[\,P - P_\ell\,]| = \mathcal{O}\left(2^{-\boldsymbol{\alpha}\ell}\right),$$
$$V_\ell = \mathrm{Var}[\,\Delta_\ell P\,] = \mathcal{O}\left(2^{-\boldsymbol{\beta}\ell}\right),$$
$$W_\ell = \mathcal{O}(2^{\boldsymbol{\gamma}\ell}),$$

where the work to sample $\Delta_\ell P$ is W_ℓ, then there are optimal choices of L and M_ℓ so that the MLMC estimator has complexity

$$\begin{cases} \mathcal{O}\left(\varepsilon^{-2-\max\left(0,\frac{\boldsymbol{\gamma}-\boldsymbol{\beta}}{\boldsymbol{\alpha}}\right)}\right), & \text{when } \boldsymbol{\gamma} \neq \boldsymbol{\beta} \\ \mathcal{O}\left(\varepsilon^{-2}|\log \varepsilon|^2\right) & \text{otherwise.} \end{cases}$$

Multilevel Monte Carlo: summary

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

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

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

$$\begin{aligned} |\mathbb{E}[\,P - P_\ell\,]| &= \mathcal{O}\left(2^{-\alpha\ell}\right), \\ V_\ell = \mathrm{Var}[\,\Delta_\ell P\,] &= \mathcal{O}\left(2^{-\beta\ell}\right), \\ W_\ell &= \mathcal{O}(2^{\gamma\ell}), \end{aligned}$$

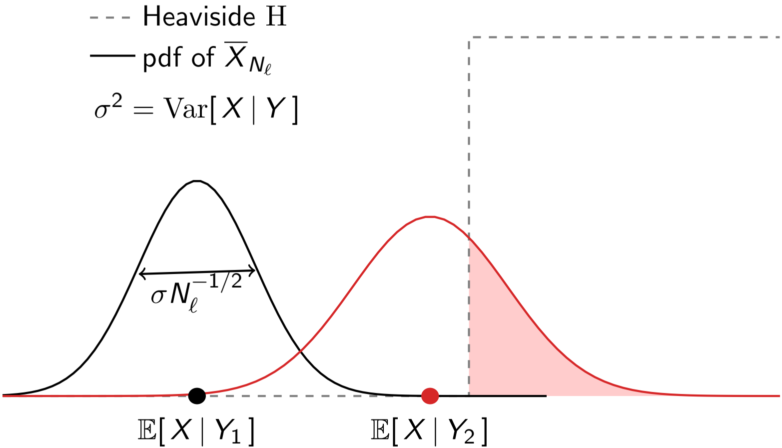
where the work to sample $\Delta_\ell P$ is W_ℓ, then there are optimal choices of L and M_ℓ so that the MLMC estimator has complexity

$$\begin{cases} \mathcal{O}\left(\varepsilon^{-2-\max\left(0,\frac{\gamma-\beta}{\alpha}\right)}\right), & \text{when } \gamma \neq \beta \\ \mathcal{O}\left(\varepsilon^{-2}|\log \varepsilon|^2\right) & \text{otherwise.} \end{cases}$$

c.f. MC: $\mathcal{O}(\varepsilon^{-2-\frac{\gamma}{\alpha}})$

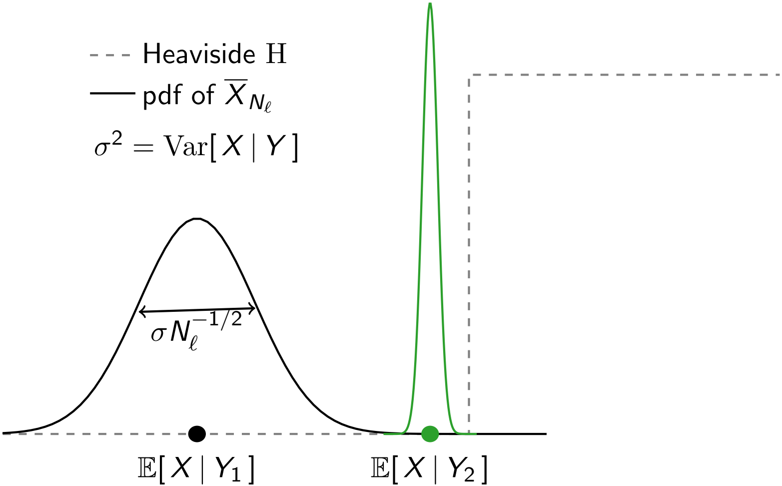
Choice of N_ℓ: Need for adaptivity

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



Choice of N_ℓ: Need for adaptivity

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



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:

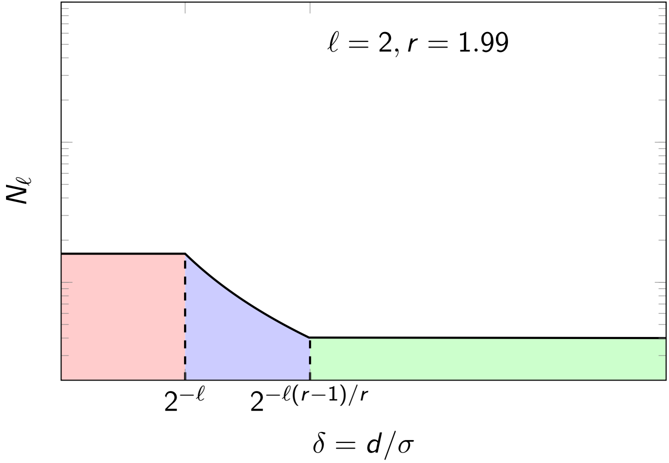
$$N_\ell = \max\Big(2^\ell, 4^\ell \min(1, (2^\ell \delta)^{-r})\Big), \quad 1 < r < 2,$$

Note

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

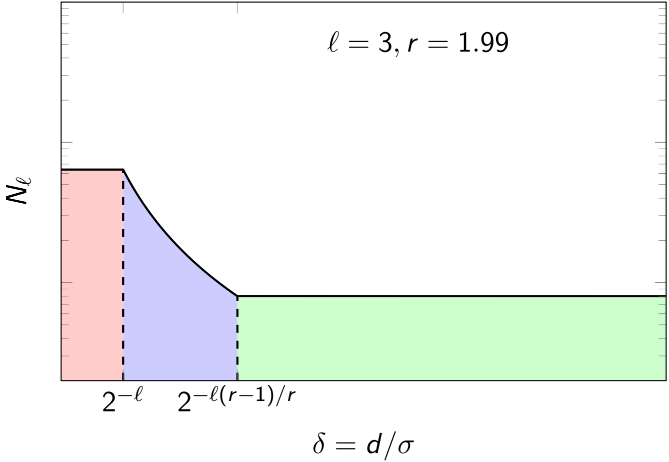
MLMC + adaptive inner sampling

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



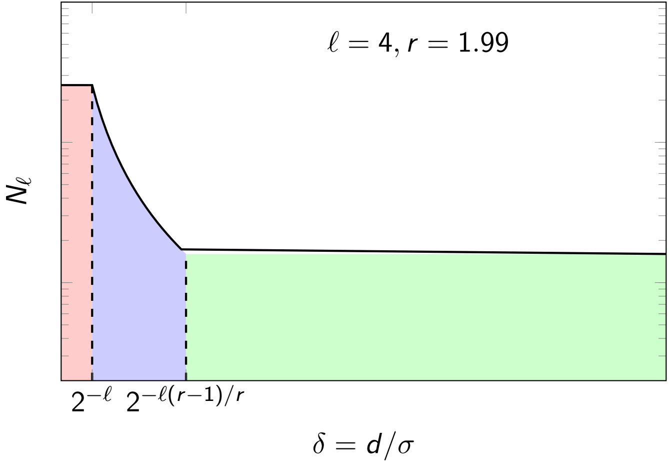
MLMC + adaptive inner sampling

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



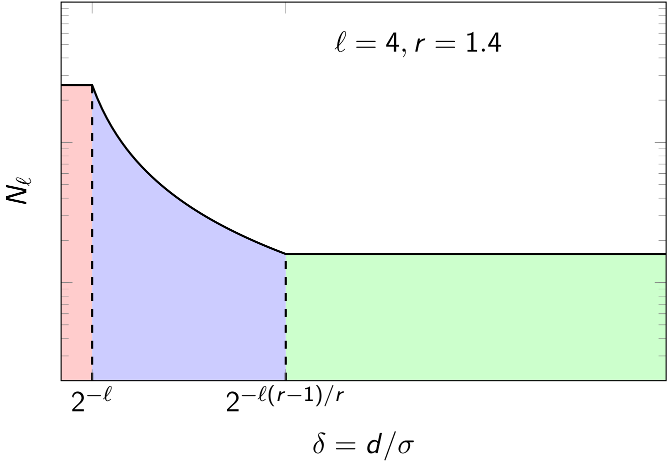
MLMC + adaptive inner sampling

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



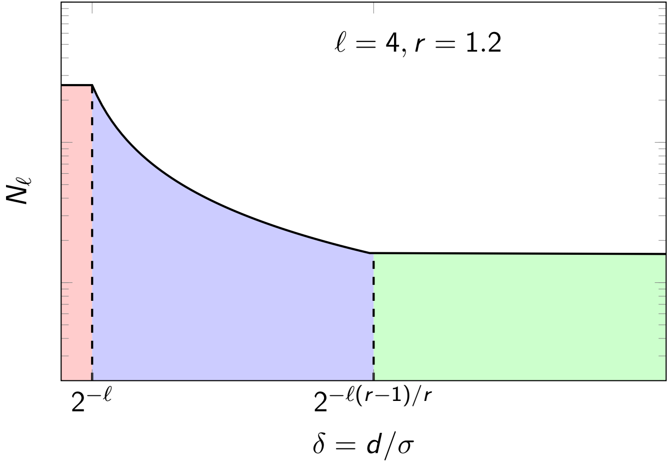
MLMC + adaptive inner sampling

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



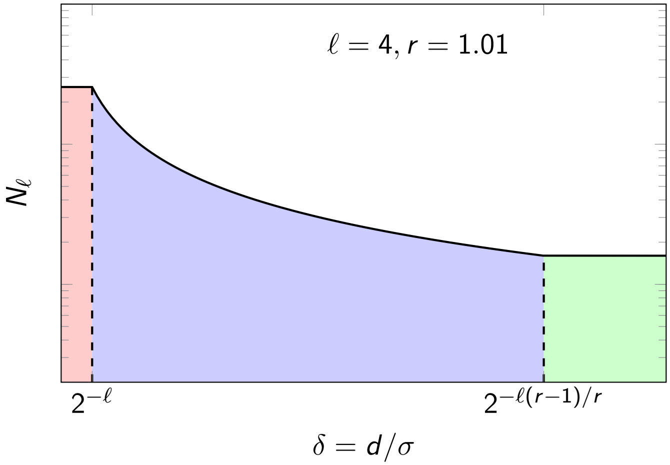
MLMC + adaptive inner sampling

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



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.

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.

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*

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

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

Epilogue: 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 to obtain a cost that is independent of the number of options.
- Using an antithetic estimator is possible and improves the computational complexity by a constant.
- The discussion can be easily extended to terms with heterogeneous work.
- More complicated underlying assets, requiring time discretization, are also handled using unbiased MLMC (leading to nested MLMC).

Epilogue: extensions (in progress)

- 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 to obtain a cost that is independent of the number of options.
- Using an antithetic estimator is possible and improves the computational complexity by a constant.
- The discussion can be easily extended to terms with heterogeneous work.
- More complicated underlying assets, requiring time discretization, are also handled using unbiased MLMC (leading to nested MLMC).

References

- M.B. Giles and A. Haji-Ali. "Multilevel nested simulation for efficient risk estimation." SIAM/ASA Journal on Uncertainty Quantification, 7(2):497–525, 2019.
- M.B. Giles. "Multilevel Monte Carlo path simulation." *Operations Research*, 56(3):607–617, 2008.
- M.B. Gordy and S. Juneja. "Nested simulation in portfolio risk measurement." *Management Science*, 56(10):1833–1848, 2010.
- M. Broadie, Y. Du and C.C. Moallemi. "Efficient risk estimation via nested sequential simulation." *Management Science*, 57(6):1172–1194, 2011.
- K. Bujok, B.M. Hambly and C. Reisinger. "Multilevel simulation of functionals of Bernoulli random variables with application to basket credit derivatives." *Methodology and Computing in Applied Probability*, 17:579–604, 2015.
- M.B. Giles. "Multilevel Monte Carlo methods." *Acta Numerica*, 24:259–328, 2015.
- W. Gou. "Estimating Value-at-Risk using multilevel Monte Carlo Maximum Entropy method". Oxford University MSc thesis, 2016.