Multilevel Monte Carlo for VaR

Mike Giles, Abdul-Lateef Haji-Ali

Mathematical Institute, University of Oxford

(with acknowledgements to Wenhui Gou, Ralf Korn and Klaus Ritter)

Global Derivatives

Barcelona, May 10, 2017

Outline

- MLMC and randomised MLMC
- Value-at-Risk and other risk measures
- prior research on VaR
  - Gordy & Juneja (2010)
  - Broadie, Du & Moallemi (2011)
- portfolio sub-sampling
- estimating inner conditional expectation
- adding in SDE approximation

Multilevel Monte Carlo

MLMC is based on the telescoping sum

$$\mathbb{E}[P_L] \;=\; \mathbb{E}[P_0] + \sum_{\ell=1}^{L} \mathbb{E}[P_\ell - P_{\ell-1}] \;\equiv\; \sum_{\ell=0}^{L} \mathbb{E}[\Delta P_\ell]$$

where P_ℓ represents an approximation of some output P on level ℓ, and $\Delta P_\ell \equiv P_\ell - P_{\ell-1}$ with $P_{-1} \equiv 0$.

If the weak convergence is

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

and Y_ℓ is an unbiased estimator for $\mathbb{E}[P_\ell - P_{\ell-1}]$, with variance

$$\mathbb{V}[Y_\ell] = O(2^{-\beta\,\ell}),$$

and expected cost

$$\mathbb{E}[C_\ell] = O(2^{\gamma\,\ell}), \quad \ldots$$

Multilevel Monte Carlo

... then the finest level L and the number of samples N_ℓ on each level can be chosen to achieve an RMS error of ε at an expected cost

$$C = \begin{cases} O\left(\varepsilon^{-2}\right), & \beta > \gamma, \\ O\left(\varepsilon^{-2}(\log \varepsilon)^2\right), & \beta = \gamma, \\ O\left(\varepsilon^{-2-(\gamma-\beta)/\alpha}\right), & 0 < \beta < \gamma. \end{cases}$$

I always try to get $\beta > \gamma$, so the main cost comes from the coarsest levels – use of QMC can then give substantial additional benefits.

Original research in 2006 used 2^ℓ timesteps to approximate an SDE on level ℓ. Since then it has been used in a variety of applications, including SPDEs, stochastic reaction networks and nested simulation.

Randomised Multilevel Monte Carlo

Rhee & Glynn (2015) started from

$$\mathbb{E}[P] \;=\; \sum_{\ell=0}^{\infty} \mathbb{E}[\Delta P_\ell] \;=\; \sum_{\ell=0}^{\infty} p_\ell \; \mathbb{E}[\Delta P_\ell / p_\ell],$$

to develop an unbiased single-term estimator

$$Y = \Delta P_{\ell'} \,/\, p_{\ell'},$$

where ℓ' is a random index which takes value ℓ with probability p_ℓ.

$\beta > \gamma$ is required to simultaneously obtain finite variance and finite expected cost using

$$p_\ell \propto 2^{-(\beta+\gamma)\ell/2}.$$

The complexity is then $O(\varepsilon^{-2})$.

Value-at-Risk

Financial institutions (banks, pension companies, insurance companies) hold portfolios with a variety of financial assets:

- cash
- bonds
- stocks
- options

and also debts / obligations:

- pension payments
- insurance payments

Value-at-Risk

Collectively, the portfolio value can be expressed as a sum of risk-neutral expectations of discounted payoffs/cash-flows f_p:

$$V = \sum_{p=1}^{P} \mathbb{E}[f_p]$$

in which the individual expectations are obtained in a variety of ways:

- actual value (e.g. cash and stocks)
- analytically (e.g. Black-Scholes option prices)
- quasi-analytically (highly efficient FFT methods)
- simple Monte Carlo, with exact sampling of underlying
- complex Monte Carlo with SDE approximation for underlying
- finite difference approximation of PDE

Value-at-Risk

Institutions and regulators are concerned about the risk of a very large loss in a short time.

Given a risk horizon τ with a given distribution for risk factors R_τ at that time, the simplest question is

What is the probability of the portfolio loss L exceeding L_{max}?

This means estimating $\mathbb{P}[L > L_{max}] \equiv \mathbb{E}\left[\mathbf{1}(L > L_{max})\right]$ where

$$L(R_\tau) \;=\; \sum_{p=1}^{P} L_p(R_\tau) \;=\; \sum_{p=1}^{P} \mathbb{E}[f_p] - \mathbb{E}[f_p | R_\tau]$$

This is therefore a nested simulation problem, and the indicator function makes it even harder.

Value-at-Risk

In the current work we focus on this problem of estimating the probability of a large loss.

VaR is defined implicitly by $\mathbb{P}[L > L_\alpha] = \alpha$ for some specified small α.
A 1D root-finding procedure can be used to determine L_α.

CVaR (Expected Shortfall) is then defined as $\mathbb{E}\left[L \mid L > L_\alpha\right]$.

Value-at-Risk

What makes it expensive? Where is the potential for MLMC?

- large number of financial products in the portfolio (P)
- often needs lots of Monte Carlo samples for inner conditional expectation (M)
- sometimes needs lots of timesteps for SDE approximation (T)

P, M and T all offer possibilities for MLMC treatment

Prior research on VaR

Gordy & Juneja (2010) considered

$$\mathbb{P}\left[L > L_{max}\right] \equiv \mathbb{E}\left[\mathbf{1}\left(L > L_{max}\right)\right]$$

using N outer samples for R_τ, and M inner samples to estimate $L(R_\tau)$.

The variance for the estimator for $L(R_\tau)$ is $O(M^{-1})$, and Gordy & Juneja prove this produces a bias in the outer estimate of the same order.

Hence, for ε RMS accuracy require

- $M = O(\varepsilon^{-1})$
- $N = O(\varepsilon^{-2})$

and so the complexity is $O(M\,N\,P) = O(\varepsilon^{-3}P)$ since each inner sample has $O(P)$ cost.

Prior research on VaR

They also considered what happens as the number of products $P \to \infty$.

For this, they introduced a weighting $1/P$ for each product, so "total loss" is now "average loss".

In this case, the variance for the estimator for $L(R_\tau)$ is $O(M^{-1}P^{-1})$, if independent sampling is used for each product in the portfolio.

Hence, for ε RMS accuracy require

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

and so the complexity is $O(M\,N\,P) = O(\max(\varepsilon^{-2}P, \varepsilon^{-3}))$.

Prior research on VaR

Their analysis can be generalised if we need to approximate an SDE: if the inner conditional expectation estimate has bias μ and variance σ^2, then overall the bias in the outer expectation is

$$O(\mu + \sigma^2).$$

Interesting – standard Mean Square Error analysis for SDE approximations without nested simulation gives

$$\text{MSE} = \mu^2 + \sigma^2$$

and we usually balance these two terms so that $\mu \sim \sigma \sim \varepsilon$.

However, this nested simulation application needs $\mu \sim \sigma^2 \sim \varepsilon$
so $\mu \ll \sigma$ – ideally we'd like it to be unbiased.

Prior research on VaR

Broadie, Du & Moallemi (2011) improved on Gordy & Juneja by noting that we don't need many samples to determine whether $L > L_{max}$ unless $L - L_{max}$ is small.

Heuristic analysis: when using M inner samples, if

$$\sigma^2(R_\tau) = \mathbb{V}[\Delta f \mid R_\tau], \qquad d(R_\tau) = |L - L_{max}|$$

where Δf is a single sample of the conditional loss, then usual confidence interval is $\pm 3\,\sigma/\sqrt{M}$ so need roughly

$$M = 9\,\sigma^2(R_\tau)/d^2(R_\tau)$$

inner samples to be sure whether or not $L > L_{max}$.

Prior research on VaR

Remembering $\sigma^2 \sim P^{-1}$ in the large P asymptotic analysis, if we use

$$M = \lceil \min\left(c\,\varepsilon^{-1}P^{-1}, 9\,\sigma^2(R_\tau)/d^2(R_\tau)\right) \rceil$$

then the cross-over point is at $d = O(\varepsilon^{1/2})$ and the average number of inner samples is

$$\overline{M} = \max(1, O(\varepsilon^{-1/2}P^{-1})),$$

reducing the overall complexity to $O(\overline{M}\,N\,P) = O(\max(\varepsilon^{-2}P, \varepsilon^{-5/2})\,)$.

This is better, but still not the $O(\varepsilon^{-2})$ that we aim for.

Also, the issue of timestepping approximation hasn't been addressed yet.

Synthetic portfolio construction

- 16 underlying assets modelled by correlated GBM
- a large number of single-asset puts and calls with varying weights, strikes and maturities, and 1-week risk horizon
- ratio of puts and calls is set to be Delta-neutral
- all holdings are short, not long, to ensure negative Gamma so small probability of a large loss
- portfolio can be valued in 3 different ways:
  - (BS) Black-Scholes prices
  - (Exact) MC estimate based on exact sample of underlying
  - (Unbiased) MLMC using Rhee & Glynn unbiased estimation

BS Portfolio

Key idea: conditional on risk factors R_τ, the total loss is

$$L = \sum_{p=1}^{P} L_p = P \; \mathbb{E}[L_p]$$

if p is uniformly distributed in $\{1, 2, \ldots, P\}$ in the r.h.s. expectation

Hence, it can be approximated by

$$\sum_{p=1}^{P} L_p \approx \frac{P}{M} \sum_{m=1}^{M} L_{p_m}$$

with M i.i.d. indices p_m.

(For simplicity, consider sampling with replacement, but sampling without replacement is also a possibility.)

BS Portfolio

In an MLMC framework, we can use $M_\ell = 2^\ell\, M_0$ inner samples on level ℓ.

If $\widehat{L}_\ell$ is an approximation on level ℓ, then by Central Limit Theorem

$$\widehat{L}_\ell - L \sim 2^{-\ell/2}, \;\; \widehat{L}_{\ell-1} - L \sim 2^{-\ell/2} \;\;\Longrightarrow\;\; \widehat{L}_\ell - \widehat{L}_{\ell-1} \sim 2^{-\ell/2}$$

If we are interested in $\mathbb{E}[\phi(L)]$ for some Lipschitz function $\phi(L)$, then our MLMC correction estimator

$$Y_\ell = \phi(\widehat{L}_\ell) - \phi(\widehat{L}_{\ell-1})$$

has mean $O(2^{-\ell})$, variance $O(2^{-\ell})$ and cost $O(2^\ell)$.

This gives $\alpha=1$, $\beta=1$, $\gamma=1$ in the MLMC theorem
$\Longrightarrow$ complexity is $O(\varepsilon^{-2}|\log\varepsilon|^2)$, independent of P.

BS Portfolio

The variance of the estimator can be improved by noting that

$$L_p \;\equiv\; \mathbb{E}[f_p] - \mathbb{E}[f_p | R_\tau] \;\approx\; -\,\Delta(S_\tau)_p \, \frac{\partial\, \mathbb{E}[f_p]}{\partial (S_0)_p}$$

when τ is small, and the overall loss is approximately $-\Delta S_\tau \cdot \Delta$ where Δ is the overall Delta vector for the portfolio, which is zero if it is Delta-neutral. Hence,

$$L = -\Delta S_\tau \cdot \Delta + \sum_{p=1}^{P} \left(L_p + \Delta(S_\tau)_p \, \frac{\partial\, \mathbb{E}[f_p]}{\partial (S_0)_p} \right)$$

so $(\Delta S_\tau)_p \, \partial \mathbb{E}[f_p] / \partial (S_0)_p$ is used as the control variate.

These ideas were investigated by Wenhui Gou in her 2016 MSc thesis, estimating moments of the loss to approximate its p.d.f.

BS Portfolio

Unfortunately, ϕ becomes an indicator function when estimating the large loss probability, so it is not Lipschitz.

The estimator

$$Y_\ell = \phi(\widehat{L}_\ell) - \phi(\widehat{L}_{\ell-1})$$

is only non-zero when $\widehat{L}_\ell$ and $\widehat{L}_{\ell-1}$ are on either side of L_{max}, and the probability of this is $O(2^{-\ell/2})$, so

$$\mathbb{V}\left[\phi(\widehat{L}_\ell) - \phi(\widehat{L}_{\ell-1})\right] = O(2^{-\ell/2}).$$

giving $\beta = 1/2$ and a complexity which is $O(\varepsilon^{-2.5})$.

BS Portfolio

To achieve a complexity which is nearly $O(\varepsilon^{-2})$, we add in the adaptive sampling idea of Broadie *et al*.

On level ℓ we use

$$M_\ell \sim \begin{cases} 2^\ell, & |\widehat{L} - L_{max}| < 2^{-\ell/2} \\ 2^{\ell/2}, & |\widehat{L} - L_{max}| > 2^{-\ell/4} \\ |\widehat{L} - L_{max}|^{-2}, & \text{in between} \end{cases}$$

This leads to

$$\begin{aligned} \mathbb{E}[\phi(\widehat{L}_\ell) - \phi(\widehat{L}_{\ell-1})] &\sim 2^{-\ell} \\ \mathbb{V}[\phi(\widehat{L}_\ell) - \phi(\widehat{L}_{\ell-1})] &\sim 2^{-\ell/2} \\ \mathbb{E}[M_\ell] &\sim 2^{\ell/2} \end{aligned}$$

so that $\alpha \approx 1$, $\beta \approx 1/2$, $\gamma \approx 1/2$, and the complexity is approximately $O(\varepsilon^{-2} |\log \varepsilon|^2)$.

Portfolio with exact sampling

The previous MLMC estimator is easily extended to include Monte Carlo estimation of inner conditional expectations:

$$\sum_{p=1}^{P} L_p(R_\tau) = \sum_{p=1}^{P} \mathbb{E}[f_p] - \mathbb{E}[f_p | R_\tau] \approx \frac{P}{M} \sum_{m=1}^{M} \Big(f_{p_m}(R_m, W_m) - f_{p_m}(R_\tau, W_m) \Big)$$

where W_m represents all of the random inputs needed for the conditional expectation, and R_m is the extra random inputs for the time interval $[0, \tau]$ needed for the time 0 valuation.

This essentially combines the P and M issues into one, controlled by M.

(Again it is really important to use a Delta control variate to reduce the variance of the MLMC estimator.)

Portfolio with unbiased sampling

How do we add in time-stepping approximation of the SDEs?

For the inner conditional expectation what we want is an unbiased unit-cost estimator.

In many cases, can use Rhee & Glynn's unbiased estimator based on randomised MLMC – the analysis remains valid, since each inner sample has $O(1)$ expected cost.

(There are some issues with ensuring finite kurtosis so that we can estimate the variance accurately, but these are not insurmountable.)

Numerical analysis

Rigorous analysis of the variance and cost of the MLMC estimator is challenging due to the adaptive sampling:

- for each R_τ, need to estimate the conditional variance

$$\sigma^2(R_\tau) \;\equiv\; \mathbb{V}[f_p \,|\, R_\tau]$$

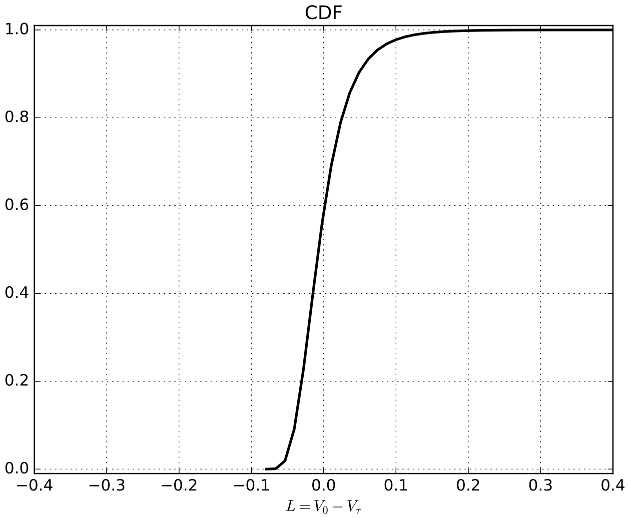
  and the distance from the indicator threshold

$$L(R_\tau) - L_{max} \;\approx\; \widehat{L} - L_{max}$$

- over-estimating $|L(R_\tau) - L_{max}|$ can lead to too few inner samples, and a higher likelihood of $\widehat{L}$ being on the wrong side of L_{max}
- under-estimating $|L(R_\tau) - L_{max}|$ can lead to too many inner samples, and an increased computational cost
- assuming a known upper bound for σ^2, we can bound these problems and prove that the overall complexity is $o(\varepsilon^{-2-\delta})$, for any small $\delta > 0$.

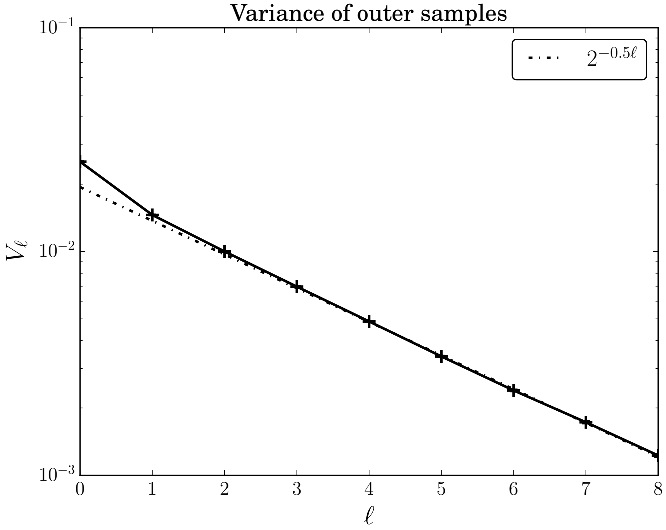
Numerical Results

1) CDF of loss for BS Portfolio (16 underlying assets)



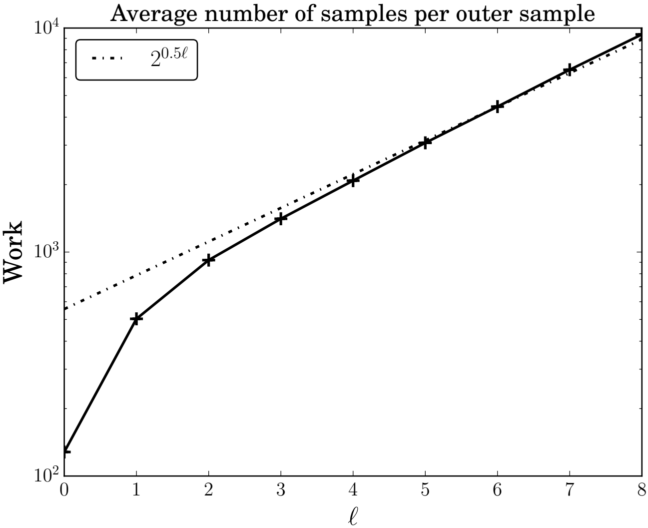
Numerical Results

2) variation of MLMC variance with level



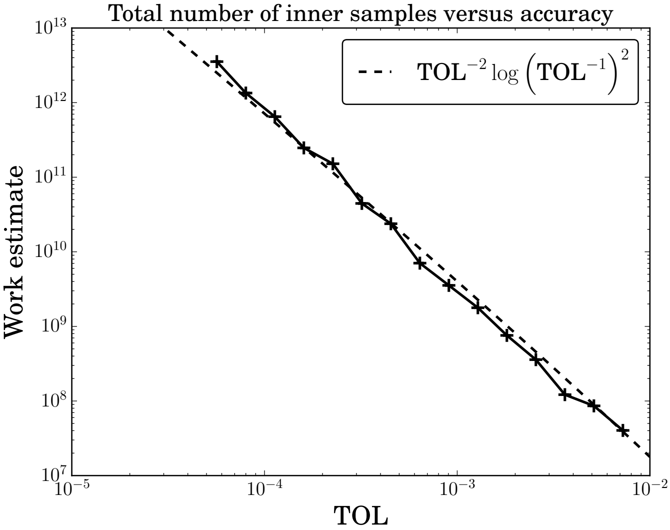
Numerical Results

3) variation of MLMC cost with level



Numerical Results

4) variation of total MLMC cost with accuracy



Future work

- root-finding for VaR, and then calculation of CVaR
- more realistic generation of risk factors (e.g. use of copula?)
- more realistic heterogeneous portfolio construction
- importance sampling of products in portfolio to optimise for varying variance/cost ratios
- importance sampling of risk factors, to get more extreme samples
- work with banks to test practical applications
- extend to pension and life insurance applications (longer maturities and risk horizon, added complications such as life expectancy for pension payments, but otherwise broadly similar)

Conclusions

- VaR is a great new application area for MLMC
- so far, banks haven't been very interested in MLMC, perhaps because the savings have been relatively modest – with VaR, I think the savings may be quite large
- three keys to performance:
  - sub-sampling the portfolio
  - adaptive number of inner samples (Broadie *et al*)
  - MLMC structure with more inner samples on "finer" levels

Webpages:

`http://people.maths.ox.ac.uk/gilesm/mlmc.html`

`http://people.maths.ox.ac.uk/gilesm/slides.html`

`http://people.maths.ox.ac.uk/gilesm/mlmc_community.html`

References

- 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.

- 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.