Stochastic Simulation: Lecture 13

Prof. Mike Giles

Oxford University Mathematical Institute

Nested expectations

In this lecture we are concerned with nested expectations.

In the simplest case, this is of the form $\mathbb{E}\left[f\left(\mathbb{E}[X|Y]\right)\right]$ where

- X and Y can be either multi-dimensional or scalar
- we will assume that $X|Y$ can be simulated exactly at $O(1)$ cost (on average)

Applications differ in the smoothness of f:

- C^2 – the nice case, e.g. x^2
- continuous, piecewise C^2 – the more typical case. e.g. $\max(x, 0)$
- discontinuous – the nasty case, e.g. Heaviside $H(x)$

Outline

As with previous topics we will discuss

- motivating applications
  - credit modelling with multiple firms
  - financial risk estimation
  - medical research decision making under uncertainty
- standard Monte Carlo method
- multilevel Monte Carlo

and will also outline the numerical analysis for both methods.

Financial credit modelling

First MLMC nested simulation – Bujok, Hambly, Reisinger (2013)

M individual firms each satisfying an SDE of the form:

$$\mathrm{d}x_t^{(i)} = a(x_t^{(i)})\,\mathrm{d}t + B(x_t^{(i)})\,\mathrm{d}W_t + b(x_t^{(i)})\,\mathrm{d}W_t^{(i)}$$

- $\{W_t\}_{0\leq t\leq T}$ is common to all firms (systemic risk factor)
- $\{W_t^{(i)}\}_{0\leq t\leq T}$ are independent Brownian motions (idiosyncratic risk)
- model views firms as bankrupt if $X_t^{(i)}$ drops below zero.

Quantity of interest is $\lim\limits_{M\to\infty} \mathbb{E}\left[f(X_M \mid Y)\right] = \mathbb{E}\left[f(\mathbb{E}[X \mid Y])\right]$

- Y is the systemic risk factor ($\{W_t\}_{0\leq t\leq T}$)
- X_M is the fraction of M firms not bankrupt at time T
- $X = 1_{\min_{[0,T]} x_t > 0}$ is corresponding indicator function
- f is continuous and piecewise linear

Financial risk measures

Value-at-risk (VaR) and Conditional Value-at-risk (CVaR)
– first MLMC paper by Giles, Haji-Ali (2019), key previous papers by Gordy & Juneja (2010) and Broadie, Du & Moallemi (2011)

Important financial risk question:

In a week, if the bank doesn't change its holdings of various stocks, bonds and other assets, can the reduction in the value of the combined portfolio exceed its cash reserves?

Current and future portfolio values

Current value of portfolio: $\mathbb{E}[P]$.

This average is over multiple driving Brownian motions which affect the value of various financial products within a large portfolio

$$P = \sum_i^{N_P} P^{(i)}$$

Future value of portfolio: $\mathbb{E}[P|Y]$.

Y here represents the Brownian motions for the first week, and the expectation is with respect to their subsequent evolution.

The loss is then $L = \mathbb{E}[P] - \mathbb{E}[P|Y]$.

Loss probability, VaR and CVaR

Given some expected loss L conditional on Y, the probability of a loss exceeding L_η is

$$\eta \; = \; \mathbb{P}[L > L_\eta]$$

VaR is defined implicitly by specifying η and computing L_η, and CVaR (Expected Shortfall) is defined as

$$\mathbb{E}\left[L \mid L > L_\eta\right] = \min_s \left\{ s + \frac{1}{\eta} \mathbb{E}\left[\max(L-s, 0)\right] \right\}$$

The important point is that the loss is a conditional expectation

$$L - L_\eta \equiv \mathbb{E}[X \mid Y]$$

so the loss probability is a nested expectation

$$\mathbb{P}[L > L_\eta] = \mathbb{E}\left[H\left(\mathbb{E}[X \mid Y]\right) \right]$$

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

Medical application – EVPI/EVPPI/EVSI

There are huge uncertainties in medicine:

- Individual patients vary hugely – in the future it is likely medicine will be more personalised, so that treatment is more customised to each individual person

  Is the cost of customisation justified?

- When using the same treating for the whole population (with some particular ailment) there is uncertainty in how effective different treatments are

  If we knew more, could a better treatment be selected?

  Is it worth the cost of the required research?

  This second problem is the one we're focussed on here.

Medical application – EVPI

Suppose X and Y are independent r.v.'s representing unknown values of parameters in a cost-effectiveness model of a small set of possible medical treatments $d \in D$.

If we don't know either X or Y then the best treatment is the one which maximises:

$$\max_d \mathbb{E}[f_d(X, Y)]$$

but if we could somehow determine both of them then best treatment gives

$$\max_d f_d(X, Y)$$

Medical application – EVPI

On average this perfect information gives

$$\mathbb{E}[\max_d f_d(X,Y)]$$

and the difference is

$$\mathbb{E}[\max_d f_d(X,Y)] \;-\; \max_d \mathbb{E}[f_d(X,Y)]$$

which is the Expected Value of Perfect Information (EVPI), and can be computed using standard Monte Carlo with an $O(\varepsilon^{-2})$ cost.

Medical application – EVPPI

Totally eliminating uncertainty is rarely possible, so suppose instead the experiment will determine Y but not X.

In that case, we get the Expected Value of Partial Perfect Information (EVPPI)

$$\mathbb{E}\left[\max_d \mathbb{E}[f_d(X,Y)\,|\,Y]\right] \;-\; \max_d \mathbb{E}[f_d(X,Y)]$$

which is a nested simulation problem.

Similar problems arise in other fields – e.g. an oil company trying to decide whether to sink another exploratory well when mapping out a new oil reservoir.

Medical application – EVSI

Even identifying Y exactly is unlikely; more likely is to obtain some information Z which through Bayes Theorem updates a prior distribution for X and Y.

This leads to the Expected Value of Sample Information (EVSI) which has the form

$$\mathbb{E}_Z \left[\frac{\mathbb{E}[\max_d f_d(X,Y)\, \rho(X,Y|Z)]}{\mathbb{E}[\rho(X,Y|Z)]} \; - \; \frac{\max_d \mathbb{E}[f_d(X,Y)\, \rho(X,Y|Z)]}{\mathbb{E}[\rho(X,Y|Z)]} \right]$$

Monte Carlo for nested expectations

OK, let's return to generic problem of estimating

$$\mathbb{E}\left[f(\mathbb{E}[X \mid Y] \right]$$

The standard Monte Carlo estimator uses N outer samples, and M inner samples:

$$N^{-1} \sum_{n=1}^{N} f\left(\overline{X}(Y^{(n)}) \right)$$

where

$$\overline{X}(Y^{(n)}) = M^{-1} \sum_{m=1}^{M} X^{(m,n)} \mid Y^{(n)}$$

is an average of M samples of X conditional on $Y^{(n)}$.

How should M and N be chosen to achieve ε^2 Mean Square Error?

Nested expectations

If X is scalar, $f \in C^2$, and $|f''| < K$, then a Taylor series expansion gives

$$f\left(\overline{X}(Y^{(n)})\right) = f + f'\,\Delta X + \tfrac{1}{2} f''\,(\Delta X)^2$$

where f, f' are evaluated at $\mathbb{E}[X|\,Y^{(n)}]$ and f'' is evaluated elsewhere and

$$\Delta X \equiv \overline{X}(Y^{(n)}) - \mathbb{E}[X|\,Y^{(n)}].$$

Hence, for fixed $Y^{(n)}$,

$$\left|\mathbb{E}\left[f\left(\overline{X}(Y^{(n)})\right)\right] - f\left(\mathbb{E}[X|\,Y^{(n)}]\right)\right| \leq \frac{K}{2M}\mathbb{V}[X|\,Y^{(n)}]$$

Nested expectations

Hence overall

$$\left|\mathbb{E}\left[f\left(\overline{X}(Y)\right)\right] - \mathbb{E}\left[f\left(\mathbb{E}[X|Y]\right)\right]\right| \leq \frac{K}{2M}\,\mathbb{E}[\mathbb{V}[X|Y]]$$

so the bias is $O(M^{-1})$ and therefore we need $M = O(\varepsilon^{-1})$.

In addition we need $N = O(\varepsilon^{-2})$ as usual, so the total cost is $O(\varepsilon^{-3})$.

As usual, we will aim to do better with MLMC.

Prior research on VaR

Gordy & Juneja (2010) considered

$$\mathbb{P}\left[\mathbb{E}[X|Y] > 0\right] \equiv \mathbb{E}\left[H\left(\mathbb{E}[X|Y]\right)\right]$$

using a uniform sampling approach with N outer samples for Y, and M inner samples to estimate $\mathbb{E}[X|Y]$.

The inner estimator has $O(M^{-1})$ variance, – they prove this again produces a bias in the outer estimate of the same order.

(The p.d.f. for $\overline{X}(Y)$ differs from p.d.f. for $\mathbb{E}[X|Y]$ by $O(M^{-1})$, even though $\overline{X}(Y)$ differs from $\mathbb{E}[X|Y]$ by $O(M^{-1/2})$; this is similar to weak and strong convergence with SDEs.)

Hence, ε RMS accuracy again requires $O(\varepsilon^{-3})$ work.

Prior research on VaR

Broadie, Du & Moallemi (2011) improved on Gordy & Juneja by noting that few samples are needed for $H(\mathbb{E}[X \mid Y])$ when $|\, \mathbb{E}[X \mid Y]\, |$ is large.

Their adaptive sampling algorithm used something like

$$M = \min(\varepsilon^{-1}, \varepsilon^{-1/2}\sigma/d)$$

where

$$\sigma^2 \equiv \mathbb{V}[X \mid Y], \qquad d = |\mathbb{E}[X \mid Y]|$$

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

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

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

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

MLMC for nested simulation

The MLMC treatment is quite natural:

- on level ℓ use $M_\ell = 2^\ell$ samples to compute $\overline{X}^{(f)} \approx \mathbb{E}[X|Y]$
- on level $\ell-1$ can use an independent set of $M_{\ell-1} = 2^{\ell-1}$ samples to compute $\overline{X}^{(c)} \approx \mathbb{E}[X|Y]$
- if f is globally Lipschitz then the level ℓ MLMC correction estimator is

$$Z_\ell \equiv f(\overline{X}^{(f)}) - f(\overline{X}^{(c)}) = O(\overline{X}^{(f)} - \overline{X}^{(c)}) = O(2^{-\ell/2})$$

  so the variance is $O(2^{-\ell})$ while the cost is $O(2^\ell)$
- hence $\beta=1, \gamma=1$ in MLMC theorem $\Longrightarrow O(\varepsilon^{-2}|\log\varepsilon|^2)$ complexity

MLMC for nested simulation

We can do even better if $f \in C^2$ and $|f''|$ is bounded:

- on level ℓ again use 2^ℓ samples to compute $\overline{X}^{(f)} \approx \mathbb{E}[X|Y]$
- on coupled level $\ell-1$, split these samples into two subsets of size $2^{\ell-1}$ and average each to get $\overline{X}^{(c1)}, \overline{X}^{(c2)}$
- MLMC "antithetic" correction estimator is then

$$Z = f(\overline{X}^{(f)}) - \tfrac{1}{2}\left(f(\overline{X}^{(c1)}) + f(\overline{X}^{(c2)}) \right)$$

- if $f(x)$ is linear then $Z=0$; with a bounded second derivative then

$$Z = O\left((\overline{X}^{(c1)} - \overline{X}^{(c2)})^2 \right) = O(2^{-\ell})$$

  so the variance is $O(2^{-2\ell})$ while the cost is $O(2^\ell)$
- hence $\beta=2, \gamma=1$ in MLMC theorem $\Longrightarrow O(\varepsilon^{-2})$ complexity

MLMC for nested simulation

What if $f(x) = \max(x, 0)$?

- $Z = 0$ if $\overline{X}^{(c1)}, \overline{X}^{(c2)}$ have same sign
- if $\mathbb{E}[X|Y]$ has bounded density near 0, then $O(2^{-\ell/2})$ probability of $\overline{X}^{(c1)}, \overline{X}^{(c2)}$ having different signs, and in that case $Z = O(2^{-\ell/2})$
- MLMC correction variance is $O(2^{-3\ell/2})$ while the cost is $O(2^{\ell})$
- hence $\beta = 3/2, \gamma = 1$ in MLMC theorem $\Longrightarrow O(\varepsilon^{-2})$ complexity
- this is what Bujok, Hambly, Reisinger (2013) proved for their financial credit model with a piecewise linear f

MLMC for nested simulation

Finally, the nasty case $f(x) = H(x)$

- $Z = 0$ if $\overline{X}^{(c1)}, \overline{X}^{(c2)}$ have same sign
- if $\mathbb{E}[X|Y]$ has bounded density near 0, then $O(2^{-\ell/2})$ probability of $\overline{X}^{(c1)}, \overline{X}^{(c2)}$ having different signs, and in that case $Z = O(1)$
- MLMC correction variance is $O(2^{-\ell/2})$ while the cost is still $O(2^{\ell})$
- hence $\beta = 1/2, \gamma = 1$ in MLMC theorem – additional analysis gives $\alpha = 1$ so $O(\varepsilon^{-5/2})$ complexity
- This can be made rigorous under certain conditions, using

$$\mathbb{P}\left[\, |\overline{X} - \mathbb{E}[X|Y]| > c\,d \,\right] \;<\; (\sigma^2/M)\,/\,(c\,d)^2 \;\equiv\; c^{-2} M^{-1} \delta^{-2}$$

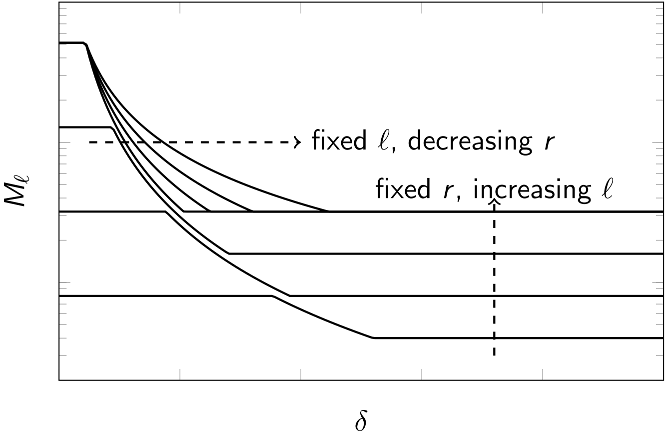
where $d \equiv |\mathbb{E}[X|Y]|$, $\sigma^2 \equiv \mathbb{V}[X|Y]$ and $\delta \equiv d/\sigma$.

MLMC + adaptive sampling

Instead of using $M_\ell = M_0\,2^\ell$ inner samples, we instead want to use

$$M_\ell = M_0\,4^\ell \max\left(2^{-\ell}, \min(1, (C^{-1}M_0^{1/2}2^\ell\delta)^{-r})\right), \quad 1 < r < 2,$$

with a minimum of $M_0\,2^\ell$ and maximum of $M_0\,4^\ell$.



MLMC + adaptive sampling

Heuristic analysis:

- $r > 1$ ensures the intermediate region is small enough that the average number of inner samples remains $O(2^\ell)$
- $r < 2$ ensures high probability of correct value for $H(\overline{X})$ in intermediate region
- $M_0\, 4^\ell$ in core region ensures $O(2^{-\ell})$ error in computed means, so $Z_\ell = O(1)$ for $O(2^{-\ell})$ fraction of outer samples
- hence, $V_\ell = \mathbb{V}[Z_\ell] = O(2^{-\ell})$ so now $\beta = 1$ and the overall complexity is $O(\varepsilon^{-2}|\log \varepsilon|^2)$

The heuristic analysis is fundamentally correct, but the rigorous analysis for the real algorithm took another year, and the upper bound on r had to be tightened.

Main challenge: real algorithm has to estimate δ.

Large portfolios

Another challenge with VaR and CVaR is the size of the portfolios:

$$P = \sum_i^{N_P} P^{(i)}$$

In real applications, can have $N_P = O(10^5)$ so cost of evaluating P is not really $O(1)$. What can we do?

Note that if I is a random index which takes value i with probability p_i then

$$P \;=\; \sum_i^{N_P} P^{(i)} \;=\; \sum_i^{N_P} p_i (P^{(i)}/p_i) \;=\; \mathbb{E}[P^{(I)}/p_I]$$

The best thing is to choose p_i roughly proportional to the likely size of $P^{(i)}$. Now can replace P by $P^{(I)}/p_I$ in simulations.

EVPPI

We chose to calculate EVPI - EVPPI

$$\mathbb{E}\left[\max_d f_d(X,Y)\right] - \mathbb{E}\left[\max_d \mathbb{E}[f_d(X,Y)\,|\,Y]\right]$$

with antithetic estimator on level ℓ based on 2^ℓ inner samples

$$Z_\ell = \tfrac{1}{2}\left(\max_d \overline{f}_d^{(a)} + \max_d \overline{f}_d^{(a)}\right) - \max_d \overline{f}_d$$

- $\overline{f}_d^{(a)}$ is an average using $2^{\ell-1}$ inner samples
- $\overline{f}_d^{(b)}$ uses an independent set of $2^{\ell-1}$ inner samples
- $\overline{f}_d$ is the average using the combined set of 2^ℓ samples

EVPPI

Analysis proves $\mathbb{V}[Z_\ell] = O(2^{-3\ell/2})$ provided $\exists c_0, c_1, c_2 > 0$ s.t.

$$\mathbb{P}\left(\min_{y \in K} \| Y - y \| \leq \epsilon \right) \; \leq \; c_0 \, \epsilon$$

$$\max_d \mathbb{E}[f(X,Y)|Y] \; - \; \max_{d \neq d_{opt}} \mathbb{E}[f(X,Y)|Y] \; \geq \; \min\left(c_1, c_2 \min_{y \in K} \| Y - y \| \right)$$

where K is the set of Y for which the optimal d is not uniquely defined.

This gives $O(\varepsilon^{-2})$ complexity, as usual.

The same can be achieved for EVSI (Expected Value of Sample Information)

Future research?

We now have good algorithms and analysis for an increasing range of applications, but there's still lots to be done.

- Goda has new papers on Expected Information Gains, and Expected Value of Sample Information
- Haji-Ali has an arXiv paper on nested MLMC to calculate $X \mid Y$, when this requires the approximation of an SDE solution – also addresses the large portfolio problem
- Szpruch and others working on applications to McKean-Vlasov equations

Key references

M.B. Gordy, S. Juneja. “Nested simulation in portfolio risk measurement”. Management Science, 56(10):1833-1848, 2010.

M. Broadie, Y. Du, C.C. Moallemi. “Efficient risk estimation via nested sequential simulation”. Management Science, 57(6):1172-1194, 2011.

K. Bujok, B. Hambly, C. Reisinger. “Multilevel simulation of functionals of Bernoulli random variables with application to basket credit derivatives”. Methodology and Computing in Applied Probability, 17(3):579-604, 2015.

M.B. Giles, A.-L. Haji-Ali. “Multilevel nested simulation for efficient risk estimation”. SIAM/ASA Journal on Uncertainty Quantification, 7(2):497-525, 2019.

M.B. Giles, A.-L. Haji-Ali. “Sub-sampling and other considerations for efficient risk estimation in large portfolios”. arXiv pre-print, 1912.05484, 2019.

Key references

M.B. Giles. “MLMC for Nested Expectations”. pp.425-442 in Contemporary Computational Mathematics – A Celebration of the 80th Birthday of Ian Sloan, Springer, 2018.

M.B. Giles, T. Goda. “Decision-making under uncertainty: using MLMC for efficient estimation of EVPPI”. Statistics and Computing, 29(4):739-751, 2019.

T. Goda, T. Hironaka, T. Iwamoto. “Multilevel Monte Carlo estimation of expected information gains”. Stochastic Analysis and Applications, 38(4):581-600, 2020.

T. Hironaka, M.B. Giles, T. Goda, H. Thom. “Multilevel Monte Carlo estimation of the expected value of sample information”. SIAM/ASA Journal on Uncertainty Quantification, 8(3):1236-1259, 2020.