Multilevel Monte Carlo for discontinuous payoffs

Mike Giles

mike.giles@maths.ox.ac.uk

Oxford-Man Institute for Quantitative Finance

Mathematical Institute, University of Oxford

CTAC08, July 14-16

Generic Viewpoint

Want to estimate the value of an integral on an extremely high-dimensional hypercube:

$$I[f] = \int_{(0,1)^d} f(x)\, \mathrm{d}x$$

My particular motivation comes from finance, with each component of x being mapped to an independent Gaussian variable in the path simulation – d can be huge.

Same idea is also applicable in other contexts,
e.g. elliptic PDEs with stochastic diffusivity.

Generic Viewpoint

Suppose $f(x)$ has a decomposition:

$$f(x) = \sum_{l=0}^{L} f_l(x_l),$$

where

- x_l is first S_l elements of x, with S_l increasing (rapidly?) with l, and $S_L \equiv d$
- $\|f_l\|_2$ decreases (rapidly?) with l
- cost of evaluating f_l increases (rapidly?) with l

In my application, cost $\sim S_l \sim 2^l$ and $\|f_l\|_2 \sim 2^{-l}$.

Generic Viewpoint

It follows immediately that

$$I[f] = \sum_{l=0}^{L} I[f_l],$$

where $I[f_l]$ is a low-dimensional integral for small l.

The idea now is to estimate each of these integrals independently, using a (quasi-) Monte Carlo approach with N_l samples on level l:

$$I[f_l] \approx N_l^{-1} \sum_{n=1}^{N_l} f_l(x_l^{(n)})$$

Generic Viewpoint

Variance of combined estimator is

$$\sum_{l=0}^{L} N_l^{-1} V_l$$

and its computational cost is

$$\sum_{l=0}^{L} N_l \, C_l$$

so variance is minimised for fixed cost by choosing

$$N_l \propto \sqrt{V_l / C_l}$$

Generic Viewpoint

To make the r.m.s. error equal to ε requires

$$N_l = \varepsilon^{-2} \sqrt{V_l / C_l} \, \sum_{m=0}^{L} \sqrt{V_m \, C_m}$$

and hence the total cost is

$$\varepsilon^{-2} \left(\sum_{l=0}^{L} \sqrt{V_l \, C_l} \right)^2 .$$

If V_l decreases faster than C_l increases, then the total cost is approximately $\varepsilon^{-2} \, V_0 \, C_0$ compared to $\varepsilon^{-2} \, V[f] \, C_L$ for standard Monte Carlo.

Finance Application

Stochastic differential equation with general drift and volatility terms:

$$\mathrm{d}S(t) = a(S,t)\,\mathrm{d}t + b(S,t)\,\mathrm{d}W(t)$$

We want to compute the expected value of an option dependent on the final state: $P = f(S(T))$

The level l approximation $\widehat{P}_l$ uses 2^l timesteps, and we have the trivial identity

$$\mathbb{E}[\widehat{P}_L] = \mathbb{E}[\widehat{P}_0] + \sum_{l=1}^{L} \mathbb{E}[\widehat{P}_l - \widehat{P}_{l-1}]$$

Finance Application

- first paper used Euler discretisation: $O(\varepsilon^{-2}(\log \varepsilon)^2)$ complexity to achieve ε r.m.s. error for Lipschitz payoffs – worse for other payoffs.
- second paper used improved Milstein discretisation for $O(\varepsilon^{-2})$ complexity for a range of payoffs, including simple discontinuous payoffs
- collaboration with Ian Sloan, Frances Kuo & Ben Waterhouse combined multilevel with QMC for even greater savings (roughly $O(\varepsilon^{-1.5})$ for Lipschitz payoffs)
- collaboration with Des Higham and Xuerong Mao led to *a priori* numerical analysis of complexity for other payoffs in first paper

Discontinuous Payoffs

Why are discontinuous payoffs a problem?

If coarse and fine paths are close

$$\left\| \widehat{S}_{l-1} - \widehat{S}_l \right\| = O(h_l),$$

then for Lipschitz payoffs this implies that

$$\widehat{P}_{l-1} - \widehat{P}_l = O(h_l)$$

whereas for discontinuous payoffs can have

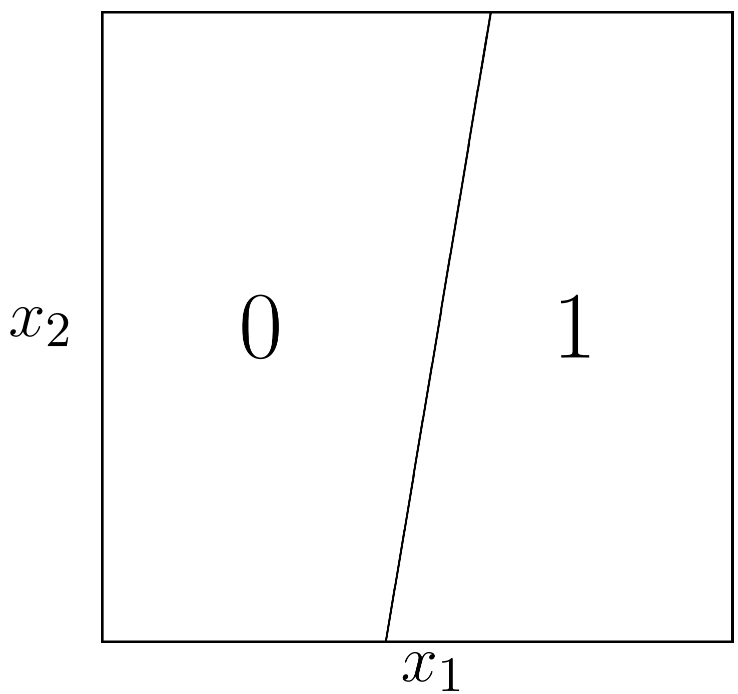
$$\widehat{P}_{l-1} - \widehat{P}_l = O(1)$$

Generic Viewpoint

Suppose we want to estimate

$$I[f] = \int_{(0,1)^2} f(x_1, x_2)\ \mathrm{d}x_1\ \mathrm{d}x_2$$

with $f(x_1,x_2) = H\left(x_1 + \delta\, g(x_1,x_2) - \alpha\right), \quad 0<\alpha<1, \quad \delta \ll 1$



Generic Viewpoint

We can split $f(x_1, x_2)$ into the sum of two parts

$$f(x_1, x_2) = f_1(x_1) + f_2(x_1, x_2)$$

where

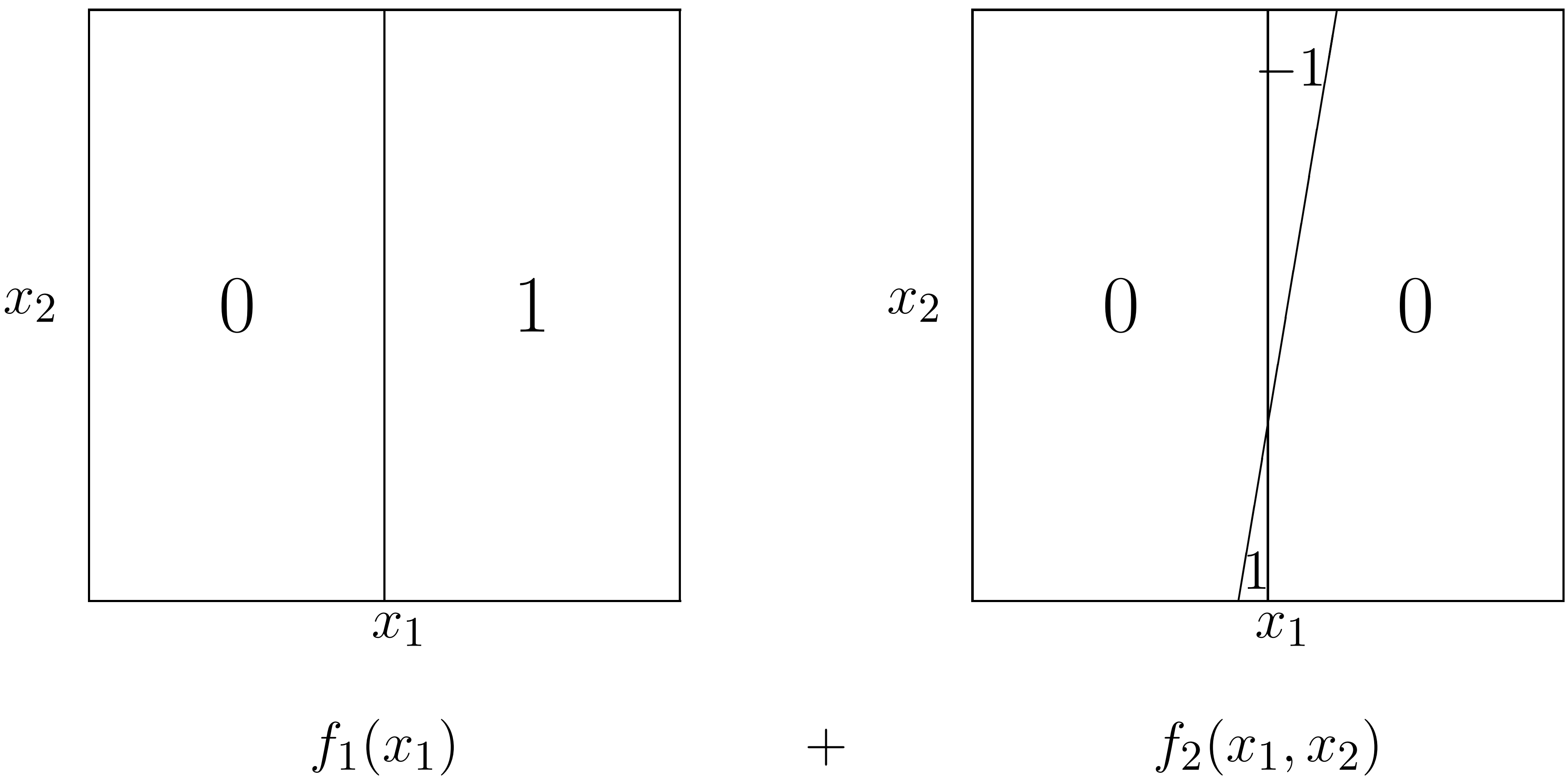
$$f_1(x_1) = H\left(x_1 - \alpha\right)$$

and

$$f_2(x_1, x_2) = H\left(x_1 + \delta\, g(x_1, x_2) - \alpha\right) - H\left(x_1 - \alpha\right)$$

The problem is that $f_2(x_1, x_2)$ is zero in most of the domain, and equals ± 1 in a narrow strip of width $O(\delta)$ around $x_1 = \alpha$.

Generic Viewpoint



Variance is $O(\delta)$, compared to the $O(\delta^2)$ one gets when $f_2(x_1, x_2) = O(\delta)$.

Generic Viewpoint

What can one do about this?

Use a change of variables

$$(x_1, x_2) \longrightarrow (\widetilde{x}_1, \widetilde{x}_2)$$

to map the unit square onto itself such that

$$H\left(x_1 + \delta\, g(x_1, x_2) - \alpha\right) = H\left(\widetilde{x}_1 - \alpha\right)$$

(I think this is similar to Lighthill's method of "strained coordinates")

Generic Viewpoint

We then get

$$
\begin{aligned}
&\int_{(0,1)^2} f_2(x_1, x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2 \\
&\quad = \int_{(0,1)^2} H\left(\widetilde{x}_1 - \alpha\right) - H\left(x_1 - \alpha\right)\ \mathrm{d}x_1\, \mathrm{d}x_2 \\
&\quad = \int_{(0,1)^2} H\left(\widetilde{x}_1 - \alpha\right) J(\widetilde{x}_1, \widetilde{x}_2)\, \mathrm{d}\widetilde{x}_1\, \mathrm{d}\widetilde{x}_2 - \int_{(0,1)^2} H\left(x_1 - \alpha\right)\ \mathrm{d}x_1\, \mathrm{d}x_2 \\
&\quad = \int_{(0,1)^2} \Big(J(x_1, x_2) - 1\Big)\ H\left(x_1 - \alpha\right)\ \mathrm{d}x_1\, \mathrm{d}x_2
\end{aligned}
$$

where $J(\widetilde{x}_1, \widetilde{x}_2) = \dfrac{\partial(x_1, x_2)}{\partial(\widetilde{x}_1, \widetilde{x}_2)}$ is the Jacobian of the mapping from $(\widetilde{x}_1, \widetilde{x}_2)$ to (x_1, x_2)

Generic Viewpoint

Since

$$J(x_1, x_2) - 1 = O(\delta)$$

we now get an integrand with an $O(\delta^2)$ variance, and it's also much more suitable for QMC integration.

This maybe seems a nice idea in principle, but impossible in practice.

Surprisingly (?) it is in fact possible for multilevel Monte Carlo path simulations.

Change of measure

The coordinate transformation corresponds to a change of measure in probability theory.

Given non-zero probability distributions $p_A(x), p_B(x)$ then

$$\begin{aligned} \mathbb{E}_A[f(x)] &= \int f(x)\, p_A(x)\, \mathrm{d}x \\ &= \int f(x)\, \frac{p_A(x)}{p_B(x)}\, p_B(x)\, \mathrm{d}x \; = \; \mathbb{E}_B[r(x) f(x)] \end{aligned}$$

where $r(x) = \dfrac{p_A(x)}{p_B(x)}$ is the Radon-Nikodym derivative due to the change in measure (similar to the Jacobian in the coordinate transformation).

Multilevel MC

In the multilevel MC approach we

- sample the underlying Brownian path up to one timestep before the end
- consider all possible Brownian increments for the final timestep
- obtain a narrow Normal distribution for the final path values

If (μ_A, σ_A) and (μ_B, σ_B) are the mean and std. deviation for the coarse and fine paths, then $\sigma_A, \sigma_B = O(\sqrt{h})$ and

$$\mu_A - \mu_B = O(h)$$

$$\sigma_A - \sigma_B = O(h)$$

Multilevel MC

The payoff difference is then

$$\mathbb{E}_A[f(S)] - \mathbb{E}_B[f(S)] = \mathbb{E}_B[(r(S)-1)\ f(S)]$$

where

$$r(S) = \frac{p_A(S)}{p_B(S)}$$

is the ratio of two Normal probability densities, and asymptotic analysis reveals that

$$r(S) - 1 = O(\sqrt{h})$$

Multilevel MC

A further refinement is that because

$$\mathbb{E}_B[r(S)] = \mathbb{E}_A[1] = 1 \quad \Longrightarrow \quad \mathbb{E}_B[r(S)-1] = 0,$$

we have

$$\mathbb{E}_B\left[(r(S)-1)\ f(S)\right] = \mathbb{E}_B\left[(r(S)-1)\ \left(f(S)-f(\mu_B)\right)\right]$$

This reduces the variance for paths which are not close to the discontinuity, so that overall the variance is $O(h^{3/2})$.

This final expectation is now estimated by Monte Carlo simulation using a few Z samples for each path.

Multilevel MC

The $O(h^{3/2})$ variance is sufficient to achieve an $O(\varepsilon^{-2})$ complexity for r.m.s. error ε using the multilevel scheme.

However, this relies on the stated assumption

$$\mu_A - \mu_B = O(h).$$

The Milstein discretisation gives this under certain conditions, but in general it is necessary to approximate Lévy areas and I think the complexity degrades slightly to $o(\varepsilon^{-2-\delta})$ for all $\delta > 0$.

Final words

The theory is developed – now need to implement it!

This is a multivariate generalisation of a technique I have used before, based on an idea due to Paul Glasserman, for simple cases in which $\mathbb{E}_A[f(S)], \mathbb{E}_B[f(S)]$ can be evaluated analytically.

Since that earlier technique worked well, I expect this "vibrato" generalisation to work as well.

Perhaps similar ideas can be used more generally for QMC integration of discontinuous functions?

Papers

M.B. Giles, “Multilevel Monte Carlo path simulation”, to appear in *Operations Research*, 2008.

M.B. Giles, “Improved multilevel convergence using the Milstein scheme”, Proceedings of *MCQMC06*, Springer-Verlag 2007.

M.B. Giles and B. Waterhouse, “Multilevel quasi-Monte Carlo path simulation”, Computational Methods in Finance conference, 2007.

`www.maths.ox.ac.uk/~gilesm/`

Email: `mike.giles@maths.ox.ac.uk`