# Multilevel estimation of a CDF or probability density

Mike Giles, Tigran Nagapetyan, Klaus Ritter

University of Oxford, TU Kaiserslautern

March 29, 2012

# Objective

- lots of applications of multilevel approach to estimating a single expected value in a wide variety of contexts
- natural extension when there is a finite set of expected values to be estimated – main complication is optimal allocation of effort
- for finite-dimensional $X$, Stefan Heinrich's multilevel parametric integration method is very efficient for estimating

$$f(\lambda) = \mathbb{E}[g(\lambda, X)]$$

  when $g$ is a smooth function of $\lambda$
- we're interested in the case where $X$ is infinite-dimensional and we want to estimate the CDF for a scalar output $t(X)$:

$$C(t_0) = \mathbb{E}[H(t_0 - t(X))]$$

  or maybe a multi-dimensional probability density

# Motivating applications

- engineering flow device to separate two sizes of particle, as described this morning by Oleg Iliev
- deterministic PDE model for fluid
- stochastic model for particle transport – for time step $h$ there is an $O(h^{1/2})$ strong error due to reflected diffusion
- each "path" simulation gives a particle exit time $t$
- we want to estimate the CDF for each of the two particle sizes

- my own interest is in travel times in nuclear waste problem described by Oliver Ernst – collaboration with Andrew Cliffe and Rob Scheichl

# A diversion

When using multilevel to estimate $M$ different expectations, using $N_l$ samples on each level, the goal is to minimise the computational cost

$$\sum_{\ell=0}^{L} N_\ell \, C_\ell$$

subject to an acceptably small variance for each quantity

$$\sum_{\ell=0}^{L} N_\ell^{-1} \, V_{\ell,m} \leq \frac{1}{2} \, \varepsilon^2, \quad m = 1, \ldots, M$$

# A diversion

To perform the minimisation, introduce positive Lagrange multipliers $\lambda_m$ which are non-zero only for active constraints, and minimise

$$\sum_{\ell=0}^{L} \left( N_\ell \, C_\ell - N_\ell^{-1} \sum_m \lambda_m V_{\ell,m} \right)$$

which leads to

$$N_\ell^{-2} \sum_m \lambda_m V_{\ell,m} = C_\ell$$

with active $\lambda_m$ given by

$$\sum_{\ell=0}^{L} N_\ell^{-1} \, V_{\ell,m} = \frac{1}{2} \, \varepsilon^2$$

In most cases (e.g. when $V_{\ell,1} \geq V_{\ell,m}, \forall \ell$) probably only one active constraint, so it reduces to standard method – might be tricker if there's more than one.

# A diversion

Simpler to define

$$V_\ell = \max_m V_{\ell,m}$$

and make variance constraint

$$\sum_{\ell=0}^{L} N_\ell^{-1}\, V_\ell \leq \frac{1}{2}\,\varepsilon^2,$$

in which case it reduces to the standard method

# CDF interpolation

- consider a grid of points $t_0$ at which we estimate the CDF
- use cubic spline to interpolate to other times to approximate full CDF
- if we want $O(\varepsilon)$ accuracy, then need $O(\varepsilon^{1/4})$ spacing, so $O(\varepsilon^{-1/4})$ evaluation points, and at least $O(\varepsilon^{-1/4})$ cost per path
- could use higher order interpolation – CDF is probably $C^\infty$

- will interpolant be monotonic? may have to use a modified spline?
- if we care about the tail of the CDF, might need to use importance sampling?

# CDF evaluation

Suppose we use standard multilevel Monte Carlo approach to estimate $\mathbb{E}[H(t_0 - t)]$?

- $O(h^{1/2})$ difference between coarse and fine path exit times
- for given $t_0$, $O(h^{1/2})$ probability of coarse and fine exit times being on different sides of $t_0$
- hence variance of multilevel correction is $O(h^{1/2})$
- weak convergence is also $O(h^{1/2})$, so the overall complexity is $O(\varepsilon^{-3})$, compared to $O(\varepsilon^{-4})$ for standard Monte Carlo

- Using adaptive time-stepping may give $O(h)$ weak / strong error, and hence $O(\varepsilon^{-2.5})$ and $O(\varepsilon^{-3})$ complexity, respectively

# CDF smoothing

The poor variance convergence is due to the Heaviside discontinuity, so to improve variance we can use smoothing

$$\mathbb{E}[H(t_0-t)] \approx \mathbb{E}[H_\delta(t_0-t)]$$

where, for example,

$$H_\delta(\tau) = \Phi(\tau/\delta) \quad \text{Normal CDF}$$

What is the error due to smoothing?

$$\begin{aligned}
\mathbb{E}[H_\delta(t_0-t) - H(t_0-t)] &= \int p(t)\,(H_\delta(t_0-t) - H(t_0-t))\; dt \\
&= \int p(t_0+\delta\tau)\,(H_\delta(-\delta\tau) - H(-\tau))\;\delta\; d\tau
\end{aligned}$$

# CDF smoothing

If $H_\delta(\tau) - \frac{1}{2}$ is an odd function, then for even $n$

$$\int \tau^n \, (H_\delta(-\delta\tau) - H(-\tau)) \; d\tau = 0$$

and so a Taylor series expansion for $p(t_0 + \delta\,\tau)$ gives the error as an asymptotic expansion in even powers of $\delta$

$$\sum_{n=1}^{\infty} a_n \, \delta^{2n}$$

To get an improved order of error, can use Richardson extrapolation. For example

$$H_\delta(\tau) = \frac{4}{3}\,\Phi(\tau/\delta) - \frac{1}{3}\,\Phi(\tau/2\delta)$$

has leading error $O(\delta^4)$ provided $p \in C^3$, and then for accuracy $O(\varepsilon)$ can use $\delta = O(\varepsilon^{1/4})$

# CDF evaluation

Now how does the multilevel variance behave?

- for $h^{1/2} \gg \delta$, difference between coarse and fine paths is greater than the smoothing width $\delta$, so variance is still $O(h^{1/2})$
- for $h^{1/2} \ll \delta$, there is $O(\delta)$ probability of being in smoothing region, and for these paths the difference is $O(h^{1/2}/\delta)$ and hence the variance is $O(h/\delta)$
- consequently, the complexity to obtain $O(\varepsilon)$ accuracy is $O(\varepsilon^{-2}(\log\varepsilon)^2\,\delta^{-1})$
- very similar complexity to standard multilevel, just worse by factor $\delta^{-1}$ due to smoothing of Heaviside discontinuity

# Numerical results

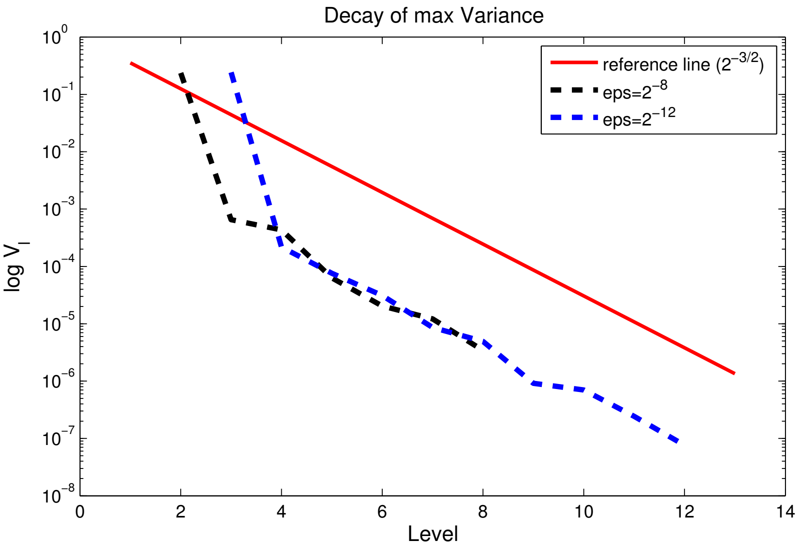
To check the initial ideas, we consider a different model problem

- 1D geometric Brownian motion
- interested in first exit time across a barrier
- Milstein approximation gives $O(h)$ strong convergence for path, but not for exit times
- best results obtained using Brownian Bridge interpolation within each time step, and then estimating the conditional expectation

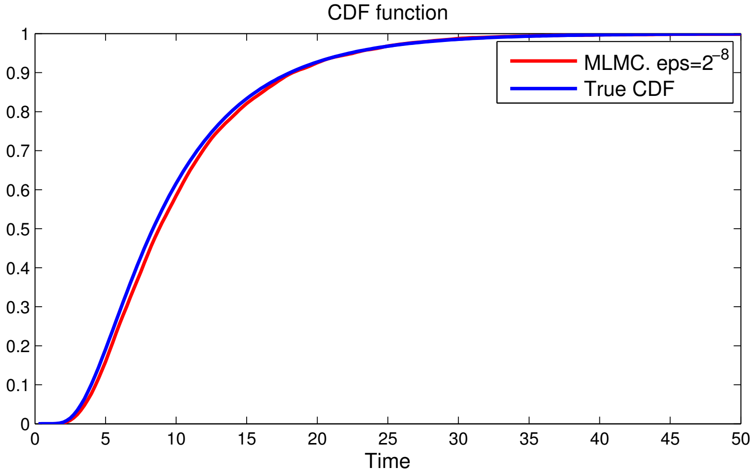
$$\mathbb{E}[\, H_\delta(t_0 - t) \mid \Delta W_n, n = 1, \ldots, N \,]$$

- coarse and fine levels have to be coupled carefully, as with financial barrier options in past papers
- based on analysis for financial barrier options, we expect $V_\ell \sim h_\ell^{3/2}$

# Numerical results



# Numerical results



# Numerical results

| $\varepsilon$ / $L_\infty$ error | $O(\delta^2)$ estimator ($\delta = \varepsilon^{0.5}$) | $O(\delta^4)$ estimator ($\delta = \varepsilon^{0.25}$) |
|---|---|---|
| $\varepsilon = 2^{-4}$ | $6.6 \cdot 10^{-2}$ | $2.1 \cdot 10^{-2}$ |
| $\varepsilon = 2^{-8}$ | $1.6 \cdot 10^{-2}$ | $6.6 \cdot 10^{-3}$ |

Table: Spline accuracy at discrete points.

# PDF approximation

This idea seems to be OK for scalar outputs, but what about $d$-dimensional outputs?

- in 1D, CDF $\Phi(t/\delta)$ corresponds to p.d.f. $\delta^{-1}\phi(t/\delta)$, so point measure has been replaced by narrow Gaussian
- in the multi-dimensional case, this suggests replacing output "particles" by $d$-dimensional Gaussian kernel of width $\delta$
- similar error analysis shows error is $O(\delta^2)$ using Gaussian kernel, but Richardson extrapolation can improve that to $O(\delta^s)$ if $p \in C^s$
- smoothed density can be used to define values on a regular grid and then use interpolation
- alternatively, could use a finite element projection of smoothed values

# PDF approximation

How does the multilevel variance behave?

- Gaussian has peak value $O(\delta^{-d})$
- for given output grid point, probability that a particle is within distance $O(\delta)$ is $O(\delta^{d})$
- if strong error $O(h^{\beta}) \ll \delta$ then difference between coarse and fine path contributions is $O(\delta^{-d}\, h^{\beta}\, \delta^{-1})$, and so multilevel variance is $O(\delta^{-d} h^{2\beta} \delta^{-2})$
- doesn't help to use a level coarser than $h = O(\delta^{1/\beta})$
- if $\beta > \frac{1}{2}$ then dominant cost is on the coarsest level and so the complexity is $O(\varepsilon^{-2}\, \delta^{-d-1/\beta})$
- if $\beta = \frac{1}{2}$ then the complexity is $O(\varepsilon^{-2} (\log \varepsilon)^2\, \delta^{-d-1/\beta})$

# Conclusions?

- very interesting engineering device
- multilevel seems to extend OK to handling output functions
- smoothing is important to get a good multilevel variance decay rate
- we'll have more supporting analysis and results next time