Stochastic Simulation: Lecture 3

Christoph Reisinger

Oxford University Mathematical Institute

Modified from earlier slides by Prof. Mike Giles.

Sensitivity analysis

In many Monte Carlo applications we don't just want to know the expected value of some quantity

$$V = \mathbb{E}[f]$$

We also want to know a whole range of first (and possibly second) derivatives of V with respect to various input parameters.

In these two lectures we will explore 3 approaches:

- finite differences
- likelihood ratio method (LRM)
- IPA (= infinitesimal perturbation analysis) = pathwise sensitivities (next lecture)

Numerical differentiation

Suppose we have MATLAB code to compute $f(x)$ (with x and $f(x)$ both scalar) and we want to compute the derivative $f'(x)$.

What can we do? Performing a Taylor series expansion,

$$f(x+\Delta x) \approx f(x) + \Delta x\, f'(x) + \tfrac{1}{2}\Delta x^2\, f''(x) + \tfrac{1}{6}\Delta x^3\, f'''(x)$$

$$\begin{aligned}
\Longrightarrow \qquad \frac{f(x+\Delta x) - f(x)}{\Delta x} &\approx f'(x) + \tfrac{1}{2}\Delta x\, f''(x), \\
\frac{f(x+\Delta x) - f(x-\Delta x)}{2\Delta x} &\approx f'(x) + \tfrac{1}{6}\Delta x^2\, f'''(x), \\
\frac{f(x+\Delta x) - 2f(x) + f(x-\Delta x)}{\Delta x^2} &\approx f''(x) + \tfrac{1}{24}\Delta x^2\, f''''(x).
\end{aligned}$$

Numerical differentiation

These are finite difference approximations, and they are the basis for the finite difference method for approximating PDEs.

In Monte Carlo methods, we use similar ideas (often referred to as "bumping") for computing sensitivities (the "Greeks" in finance)

The problem with taking $\Delta x \ll 1$ is inaccuracy due to finite precision arithmetic, in which there is a relative rounding error of size 2^{-S} where S is the size of the mantissa.

Numerical differentiation

Error in computing $f(x+\Delta x) - f(x)$ is roughly of size $2^{-S} f(x)$, so error in computing one-sided difference estimate for $f'(x)$ is of order

$$\frac{2^{-S} f(x)}{2\Delta x}$$

while the finite difference error is $O(\Delta x)$.

To balance errors, want

$$\frac{2^{-S}}{\Delta x} \sim \Delta x \quad \Longrightarrow \quad \Delta x \sim 2^{-S/2}.$$

In single precision, this means taking $\Delta x \sim 10^{-3}$, and getting an error which is roughly of size 10^{-3}. This is not great, and making Δx smaller or bigger will make things worse.

This is why many people use double precision when doing "bumping" for sensitivity analysis.

Complex Variable Trick

This is a very useful "trick", from this very short article:

> *"Using Complex Variables to Estimate Derivatives of Real Functions", William Squire and George Trapp, SIAM Review, 40(1):110-112, 1998.*

which now has 650 citations.

Complex Variable Trick

Suppose $f(z)$ is a complex analytic function, and $f(x)$ is real when x is real.

Then

$$f(x+i\,\Delta x) \approx f(x) + i\,\Delta x\, f'(x) - \tfrac{1}{2}\Delta x^2\, f''(x) - i\,\tfrac{1}{6}\Delta x^3\, f'''(x)$$

and hence

$$\frac{\mathrm{Im} f(x+i\,\Delta x)}{\Delta x} \approx f'(x) - \tfrac{1}{6}\Delta x^2\, f'''(x)$$

Now, we can take $\Delta x \ll 1$, and there is no problem due to finite precision arithmetic.

Can use $\Delta x = 10^{-10}$!

Complex Variable Trick

There are just a few catches, because $f(z)$ must be analytic:

- need analytic extensions for $\min(x,y)$, $\max(x,y)$ and $|x|$
- need analytic extensions to certain functions, e.g. MATLAB's `normcdf`
- in MATLAB, must be aware that A' is the Hermitian of A (complex conjugate transpose), so use $A.'$ for the simple transpose.

Using this, can very simply "differentiate" almost any MATLAB or C/C++ code for a real function $f(x)$.

Finite difference sensitivities

If $V(\theta) = \mathbb{E}[f\,|\,\theta]$ for an input parameter θ is sufficiently differentiable, then the sensitivity $\dfrac{\partial V}{\partial \theta}$ can be approximated by one-sided finite difference

$$\frac{\partial V}{\partial \theta} = \frac{V(\theta{+}\Delta\theta) - V(\theta)}{\Delta\theta} \; + \; O(\Delta\theta)$$

or by central finite difference

$$\frac{\partial V}{\partial \theta} = \frac{V(\theta{+}\Delta\theta) - V(\theta{-}\Delta\theta)}{2\Delta\theta} \; + \; O((\Delta\theta)^2)$$

(In the finance industry, the derivatives are known as the "Greeks" and this approach is referred to as "bumping".)

Finite difference sensitivities

The clear advantage of this approach is that it is very simple to implement.

However, the disadvantages are:

- expensive (2 extra sets of calculations for central differences)
- significant bias error if $\Delta\theta$ too large
- machine roundoff errors if $\Delta\theta$ too small
- large variance if $f(S_T)$ is discontinuous and $\Delta\theta$ small

Finite difference sensitivities

Let $X^{(i)}(\theta+\Delta\theta)$ and $X^{(i)}(\theta-\Delta\theta)$ be the values of $f(S_T)$ obtained for different MC samples, so the central difference estimate for $\dfrac{\partial V}{\partial \theta}$ is given by

$$\begin{aligned}
\widehat{Y} &= \frac{1}{2\Delta\theta}\left(N^{-1}\sum_{i=1}^{N} X^{(i)}(\theta+\Delta\theta) - N^{-1}\sum_{i=1}^{N} X^{(i)}(\theta-\Delta\theta) \right) \\
&= \frac{1}{2N\Delta\theta}\sum_{i=1}^{N}\left(X^{(i)}(\theta+\Delta\theta) - X^{(i)}(\theta-\Delta\theta) \right)
\end{aligned}$$

Finite difference sensitivities

If independent samples are taken for both $X^{(i)}(\theta+\Delta\theta)$ and $X^{(i)}(\theta-\Delta\theta)$ then

$$
\begin{aligned}
\mathbb{V}[\widehat{Y}] &\approx \left(\frac{1}{2N\Delta\theta}\right)^2 \sum_j \Big(\mathbb{V}[X(\theta+\Delta\theta)] + \mathbb{V}[X(\theta-\Delta\theta)]\Big) \\
&\approx \left(\frac{1}{2N\Delta\theta}\right)^2 2\,N\,\mathbb{V}[f] \\
&= \frac{\mathbb{V}[f]}{2N(\Delta\theta)^2}
\end{aligned}
$$

which is very large for $\Delta\theta \ll 1$.

Finite difference sensitivities

It is much better for $X^{(i)}(\theta{+}\Delta\theta)$ and $X^{(i)}(\theta{-}\Delta\theta)$ to use the same set of random inputs.

If $X^{(i)}(\theta)$ is differentiable with respect to θ, then

$$X^{(i)}(\theta{+}\Delta\theta) - X^{(i)}(\theta{-}\Delta\theta) \approx 2\,\Delta\theta\,\frac{\partial X^{(i)}}{\partial\theta}$$

and hence

$$\mathbb{V}[\widehat{Y}] \approx N^{-1}\ \mathbb{V}\left[\frac{\partial X}{\partial\theta}\right],$$

which behaves well for $\Delta\theta \ll 1$, so one should choose a small (but not ridiculously small) value for $\Delta\theta$ to minimise the bias due to the finite differencing.

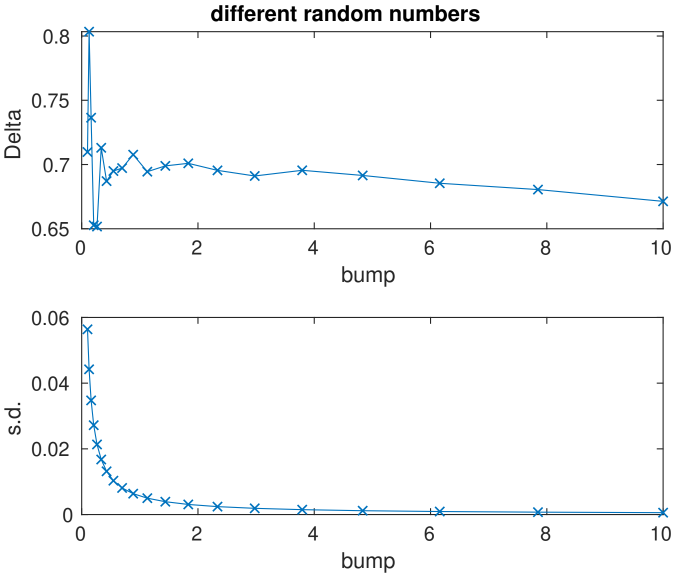
Basket call option

- 5 underlying assets starting at $S_0 = 100$, with call option on arithmetic mean with strike $K = 100$
- Geometric Brownian Motion model, $r = 0.05$, $T = 1$
- volatility $\sigma = 0.2$ and correlation matrix

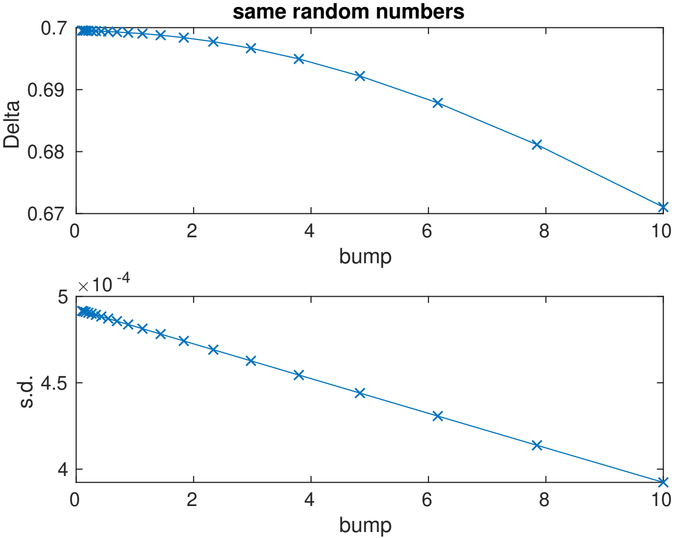
$$\Omega = \begin{pmatrix} 1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 1 \end{pmatrix}$$

The aim is to estimate $\dfrac{\partial}{\partial S_0}\mathbb{E}[\exp(-rT)\, f(S_T)]$ where $f(S_T)$ is the basket call option payoff, using central differences.

Basket call option



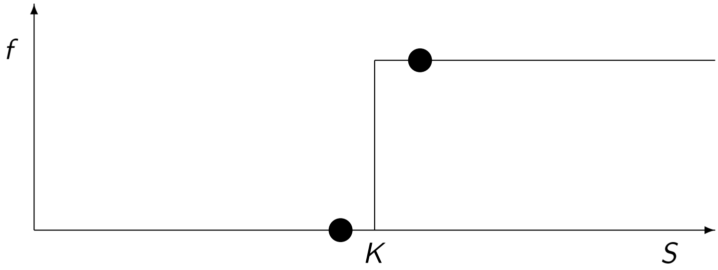
Basket call option



Finite difference sensitivities

Next, we analyse the variance of the finite difference estimator when the payoff is discontinuous.

The problem is that a small bump in the underlying S can produce a big bump in the output – not differentiable



Finite difference sensitivities

What is the probability that $S(\theta \pm \Delta\theta)$ will be on different sides of the discontinuity?

Separation of $S(\theta \pm \Delta\theta)$ is $O(\Delta\theta)$

$\mathbb{P}(\,|S(\theta) - K| < c\,\Delta\theta) = O(\Delta\theta)$

Hence, $O(\Delta\theta)$ probability of straddling the discontinuity.

Finite difference sensitivities

If we are interested in $\mathbb{E}[f(\omega)]$, and samples ω are either in set A, or its complement A^c, then

$$\begin{aligned}
\mathbb{E}[f(\omega)] &= \mathbb{E}[f(\omega) 1_A] + \mathbb{E}[f(\omega) 1_{A^c}] \\
&= \mathbb{P}(\omega \in A)\ \mathbb{E}[f(\omega) \mid \omega \in A] \\
&+ \mathbb{P}(\omega \notin A)\ \mathbb{E}[f(\omega) \mid \omega \notin A]
\end{aligned}$$

and similarly

$$\begin{aligned}
\mathbb{E}[f^2(\omega)] &= \mathbb{P}(\omega \in A)\ \mathbb{E}[f^2(\omega) \mid \omega \in A] \\
&+ \mathbb{P}(\omega \notin A)\ \mathbb{E}[f^2(\omega) \mid \omega \notin A]
\end{aligned}$$

Finite difference sensitivities

In this case of a discontinuous payoff

- For most samples, $X^{(i)}(\theta+\Delta\theta) - X^{(i)}(\theta-\Delta\theta) = O(\Delta\theta)$
- For an $O(\Delta\theta)$ fraction, $X^{(i)}(\theta+\Delta\theta) - X^{(i)}(\theta-\Delta\theta) = O(1)$

$$\implies \quad \mathbb{E}\left[\frac{X^{(i)}(\theta+\Delta\theta) - X^{(i)}(\theta-\Delta\theta)}{2\Delta\theta}\right] = O(1)$$

$$\mathbb{E}\left[\left(\frac{X^{(i)}(\theta+\Delta\theta) - X^{(i)}(\theta-\Delta\theta)}{2\Delta\theta}\right)^2\right] = O(\Delta\theta^{-1})$$

This gives $\mathbb{E}[\widehat{Y}] = O(1)$, but $\mathbb{V}[\widehat{Y}] = O(N^{-1}\Delta\theta^{-1})$.

So, small $\Delta\theta$ gives a large variance, while a large $\Delta\theta$ gives a large finite difference discretisation error.

Finite difference sensitivities

To determine the optimum choice we use the fact that

$$\text{Mean Square Error} = \text{variance} + (\text{bias})^2$$

In our case, the MSE (mean-square-error) is

$$\mathbb{V}[\widehat{Y}] + \text{bias}^2 \sim \frac{a}{N\,\Delta\theta} + b\,\Delta\theta^4.$$

This is minimised by choosing $\Delta\theta \propto N^{-1/5}$, giving

$$\sqrt{\text{MSE}} \;\propto\; N^{-2/5}$$

in contrast to the usual MC result in which

$$\sqrt{\text{MSE}} \;\propto\; N^{-1/2}$$

Finite difference sensitivities

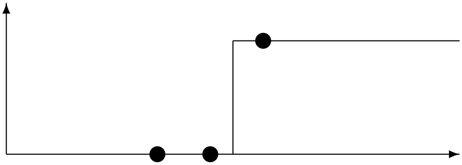
Second derivatives can also be approximated by central differences:

$$\frac{\partial^2 V}{\partial \theta^2} = \frac{V(\theta+\Delta\theta) - 2\,V(\theta) + V(\theta-\Delta\theta)}{\Delta\theta^2} \; + \; O(\Delta\theta^2)$$

This will again have a larger variance if either the payoff or its derivative is discontinuous.

Finite difference sensitivities

Discontinuous payoff:



For an $O(\Delta\theta)$ fraction of samples

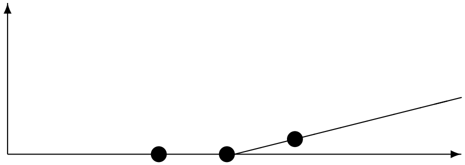
$$X^{(i)}(\theta+\Delta\theta) - 2X^{(i)}(\theta) + X^{(i)}(\theta-\Delta\theta) = O(1)$$

$$\Longrightarrow \quad \mathbb{E}\left[\left(\frac{X^{(i)}(\theta+\Delta\theta) - 2X^{(i)}(\theta) + X^{(i)}(\theta-\Delta\theta)}{\Delta\theta^2}\right)^2\right] = O(\Delta\theta^{-3})$$

This gives $\mathbb{V}[\widehat{Y}] = O(N^{-1}\Delta\theta^{-3})$.

Finite difference sensitivities

Discontinuous derivative:



For an $O(\Delta\theta)$ fraction of samples

$$X^{(i)}(\theta+\Delta\theta) - 2X^{(i)}(\theta) + X^{(i)}(\theta-\Delta\theta) = O(\Delta\theta)$$

$$\Longrightarrow \quad \mathbb{E}\left[\left(\frac{X^{(i)}(\theta+\Delta\theta) - 2X^{(i)}(\theta) + X^{(i)}(\theta-\Delta\theta)}{\Delta\theta^2}\right)^2\right] = O(\Delta\theta^{-1})$$

This gives $\mathbb{V}[\widehat{Y}] = O(N^{-1}\Delta\theta^{-1})$.

Finite difference sensitivities

Hence, for second derivatives the variance of the finite difference estimator is

- $O(N^{-1})$ if the payoff is twice differentiable
- $O(N^{-1}\Delta\theta^{-1})$ if the payoff has a discontinuous derivative
- $O(N^{-1}\Delta\theta^{-3})$ if the payoff is discontinuous

These can be used to determine the optimum $\Delta\theta$ in each case to minimise the Mean Square Error.

Likelihood ratio method

Defining $p(S)$ to the probability density function for the final state S_T, then

$$V = \mathbb{E}[f(S_T)] = \int f(S)\, p(S)\, \mathrm{d}S,$$

$$\implies \quad \frac{\partial V}{\partial \theta} = \int f\, \frac{\partial p}{\partial \theta}\, \mathrm{d}S = \int f\, \frac{\partial (\log p)}{\partial \theta}\, p\, \mathrm{d}S = \mathbb{E}\left[f\, \frac{\partial (\log p)}{\partial \theta} \right]$$

The quantity $\dfrac{\partial (\log p)}{\partial \theta}$ is sometimes called the “score function”.

Likelihood ratio method

Note that when $f = 1$, we get

$$\frac{\partial}{\partial\theta}\mathbb{E}[1] = 0$$

and therefore

$$\mathbb{E}\left[\frac{\partial(\log p)}{\partial\theta}\right] = 0$$

This is a handy check to make sure we have derived the score function correctly.

Likelihood ratio method

Example: GBM with arbitrary payoff $f(S_T)$.

For the usual Geometric Brownian motion with constants r, σ, the final log-normal probability distribution is

$$p(S) = \frac{1}{S\sigma\sqrt{2\pi T}} \; \exp\left[-\tfrac{1}{2} \left(\frac{\log(S/S_0) - (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} \right)^2 \right]$$

$$\log p = -\log S - \log\sigma - \tfrac{1}{2}\log(2\pi T) - \tfrac{1}{2}\frac{\left(\log(S/S_0) - (r - \frac{1}{2}\sigma^2)T\right)^2}{\sigma^2 T}$$

$$\Longrightarrow \quad \frac{\partial \log p}{\partial S_0} = \frac{\log(S/S_0) - (r - \frac{1}{2}\sigma^2)T}{S_0 \sigma^2 T}$$

Likelihood ratio method

Hence

$$\frac{\partial V}{\partial S_0} = \mathbb{E}\left[\frac{\log(S/S_0) - (r - \frac{1}{2}\sigma^2)\,T}{S_0\,\sigma^2\,T} \;\; f(S_T) \right]$$

In the Monte Carlo simulation,

$$\log(S/S_0) - (r - \tfrac{1}{2}\sigma^2)\,T = \sigma\, W_T$$

so the expression can be simplified to

$$\frac{\partial V}{\partial S_0} = \mathbb{E}\left[\frac{W_T}{S_0\,\sigma\;T} \;\; f(S_T) \right]$$

– very easy to implement so you estimate $\partial V/\partial S_0$ at the same time as estimating V

Likelihood ratio method

Similarly,

$$\frac{\partial \log p}{\partial \sigma} = -\frac{1}{\sigma} - \frac{\log(S/S_0) - (r - \frac{1}{2}\sigma^2)T}{\sigma} + \frac{\left(\log(S/S_0) - (r - \frac{1}{2}\sigma^2)T\right)^2}{\sigma^3 T}$$

and hence

$$\frac{\partial V}{\partial \sigma} = \mathbb{E}\left[\left(\frac{1}{\sigma}\left(\frac{W_T^2}{T} - 1\right) - W_T\right) f(S_T)\right]$$

In both cases, the variance is very large when σ is small, and it is also large for Δ when T is small. More generally, LRM is usually the approach with the largest variance.

Likelihood ratio method

To get second derivatives, note that

$$\frac{\partial^2 \log p}{\partial \theta^2} \;=\; \frac{\partial}{\partial \theta}\left(\frac{1}{p}\,\frac{\partial p}{\partial \theta}\right) \;=\; \frac{1}{p}\,\frac{\partial^2 p}{\partial \theta^2} - \frac{1}{p^2}\left(\frac{\partial p}{\partial \theta}\right)^2$$

$$\Longrightarrow \quad \frac{1}{p}\,\frac{\partial^2 p}{\partial \theta^2} \;=\; \frac{\partial^2 \log p}{\partial \theta^2} + \left(\frac{\partial \log p}{\partial \theta}\right)^2$$

and hence

$$\frac{\partial^2 V}{\partial \theta^2} = \mathbb{E}\left[\left(\frac{\partial^2 \log p}{\partial \theta^2} + \left(\frac{\partial \log p}{\partial \theta}\right)^2\right) f(S_T)\right]$$

Likelihood ratio method

In the multivariate extension, $X = \log S_T$ can be written as

$$X = \mu + L\,Z$$

where μ is the mean vector, $\Sigma = L\,L^T$ is the covariance matrix and Z is a vector of uncorrelated Normals. The joint p.d.f. is

$$\log p = -\tfrac{1}{2}\log|\Sigma| - \tfrac{1}{2}(X-\mu)^T\Sigma^{-1}(X-\mu) - \tfrac{1}{2}d\,\log(2\pi).$$

and after a lot of algebra we obtain

$$\frac{\partial \log p}{\partial \mu} = L^{-T} Z,$$

$$\frac{\partial \log p}{\partial \Sigma} = \tfrac{1}{2}\,L^{-T}\left(ZZ^T - I\right)L^{-1}$$

Final Words

- estimating sensitivities is often an important task – in computational finance it is often more important than estimating the original expectations
- finite differences are simplest approach, but least accurate and most expensive
- always use the same random numbers for both sets of simulations
- in some cases, the optimum step size is a tradeoff between variance and bias (due to finite difference discretisation error)
- LRM (likelihood ratio method) usually has a higher variance, but can cope with discontinuous output functions