Module 2: Monte Carlo Methods

Prof. Mike Giles

mike.giles@maths.ox.ac.uk

Oxford University Mathematical Institute

Greeks

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

$$V = \mathbb{E}[f(S(T))]$$

We also want to know a whole range of "Greeks" corresponding to first and second derivatives of V with respect to various parameters:

$$\Delta = \frac{\partial V}{\partial S_0}, \qquad \Gamma = \frac{\partial^2 V}{\partial S_0^2},$$

$$\rho = \frac{\partial V}{\partial r}, \qquad \text{Vega} = \frac{\partial V}{\partial \sigma}.$$

Greeks

The Greeks are needed for hedging and risk analysis.

Whereas prices can be obtained to some extent from market prices, simulation is the only way to determine the Greeks.

In this lecture we will explore 3 approaches:

- finite differences
- likelihood ratio method
- pathwise sensitivities

Finite difference sensitivities

If $V(\theta) = \mathbb{E}[f(S(T))]$ 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

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

(This approach is referred to as getting Greeks by "bumping" the input parameters.)

Finite difference sensitivities

The clear advantage of this approach is that it is very simple to implement (hence the most popular in practice?)

However, the disadvantages are:

- expensive (2 extra sets of calculations for central differences)
- significant bias error if $\Delta\theta$ too large
- machine roundoff errors is $\Delta\theta$ too small
- large variance if $f(S(T))$ 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 value for $\Delta\theta$ to minimise the bias due to the finite differencing.

Finite difference sensitivities

However, there are problems if $\Delta\theta$ is chosen to be extremely small.

In finite precision arithmetic,

$$X^{(i)}(\theta \pm \Delta\theta)$$

has an error which is approximately random with r.m.s. magnitude δ

- single precision $\delta \approx 10^{-6}|X|$
- double precision $\delta \approx 10^{-14}|X|$

Finite difference sensitivities

Consequently,

$$\frac{1}{2\Delta\theta}\left(X^{(i)}(\theta\!+\!\Delta\theta) - X^{(i)}(\theta\!-\!\Delta\theta)\right)$$

has an extra error term with approximate variance

$$\frac{\delta^2}{2(\Delta\theta)^2}$$

and therefore $\widehat{Y}$ has an extra error term with approximate variance

$$\frac{\delta^2}{2N(\Delta\theta)^2}.$$

Finite difference sensitivities

For double precision computations, if $\theta = O(1)$, then can probably use

$$\Delta\theta = 10^{-6}$$

without any problems, and even the $O(\Delta\theta)$ finite difference error from one-sided differencing will probably be small compared to the MC sampling error.

For single precision, better to use a larger perturbation, e.g.

$$\Delta\theta = 10^{-4}$$

and use the more expensive central differencing to minimise the discretisation error.

Finite difference sensitivities

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

In this case

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

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

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

Finite difference sensitivities

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

To determine the optimum choice we use the following result: if $\widehat{Y}$ is an estimator for $\mathbb{E}[Y]$ then

$$
\begin{aligned}
\mathbb{E}\left[\left(\widehat{Y}-\mathbb{E}[Y]\right)^2\right] &= \mathbb{E}\left[\left(\widehat{Y}-\mathbb{E}[\widehat{Y}] \;+\; \mathbb{E}[\widehat{Y}]-\mathbb{E}[Y]\right)^2\right] \\
&= \mathbb{E}\left[(\widehat{Y}-\mathbb{E}[\widehat{Y}])^2\right] + (\mathbb{E}[\widehat{Y}]-\mathbb{E}[Y])^2 \\
&= \mathbb{V}[\widehat{Y}] + \left(\mathbb{E}[\widehat{Y}]-\mathbb{E}[Y]\right)^2
\end{aligned}
$$

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

Finite difference sensitivities

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

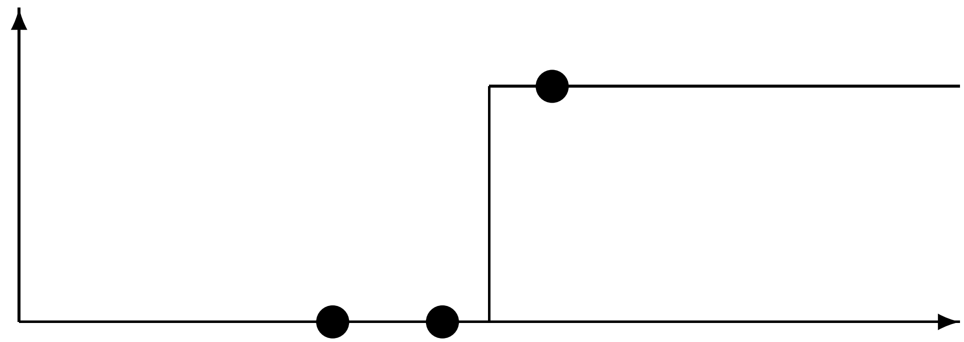
Second derivatives such as Γ can also be approximated by central differences:

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

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

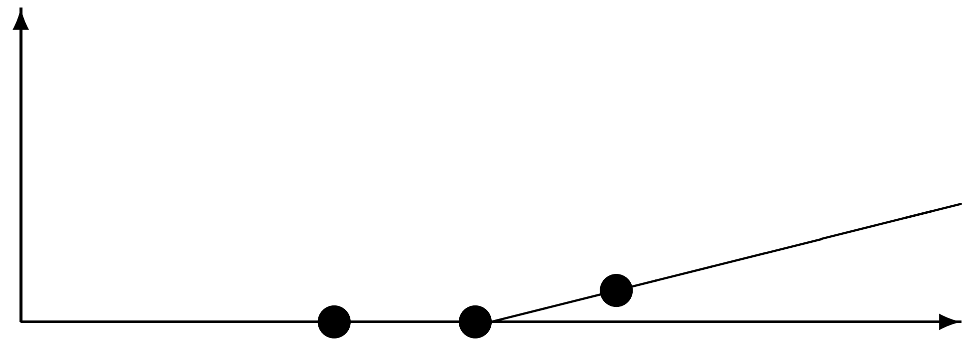
$$X^{(i)}(\theta{+}\Delta\theta) - 2X^{(i)}(\theta) + X^{(i)}(\theta{-}\Delta\theta) = O(1)$$

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

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

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

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

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 means that we can use the score function as a control variate – can be useful to reduce the variance, and is a handy check to make sure we have derived it 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

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

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

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

Likelihood ratio method

Hence

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

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

$$\Delta = \mathbb{E}\left[\frac{W(T)}{S_0\,\sigma\,T}\ f(S(T))\right]$$

– very easy to implement so you estimate Δ at the same time as estimating the price V

Likelihood ratio method

Similarly for vega we have

$$
\begin{aligned}
\frac{\partial \log p}{\partial \sigma} &= -\frac{1}{\sigma} - \sqrt{T}\left(\frac{\log(S/S_0) - (r-\frac{1}{2}\sigma^2)T}{S_0\sigma^2 T}\right) \\
&\quad + \frac{1}{\sigma}\left(\frac{\log(S/S_0) - (r-\frac{1}{2}\sigma^2)T}{S_0\sigma^2 T}\right)^2
\end{aligned}
$$

and hence

$$
\text{vega} = \mathbb{E}\left[\left(\frac{1}{\sigma}\left(\frac{W(T)^2}{T} - 1\right) - W(T)\right) f(S(T))\right]
$$

Likelihood ratio method

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

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

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

Pathwise sensitivities

Start instead with

$$V \equiv \mathbb{E}\left[f(S(T)) \right] = \int f(S(T)) \, p_W(W) \, \mathrm{d}W$$

and differentiate this to get

$$\frac{\partial V}{\partial \theta} = \int \frac{\partial f}{\partial S} \, \frac{\partial S(T)}{\partial \theta} \, p_W \, \mathrm{d}W = \mathbb{E}\left[\frac{\partial f}{\partial S} \, \frac{\partial S(T)}{\partial \theta} \right]$$

with $\partial S(T)/\partial \theta$ being evaluated at fixed W.

Note: this derivation needs $f(S)$ to be differentiable, but by considering the limit of a sequence of smoothed (regularised) functions can prove it's OK provided $f(S)$ is continuous and piecewise differentiable

Pathwise sensitivities

This leads to the estimator

$$\frac{1}{N}\sum_{i=1}^{N} \frac{\partial f}{\partial S}(S^{(i)})\ \frac{\partial S^{(i)}}{\partial \theta}$$

which is the derivative of the usual price estimator

$$\frac{1}{N}\sum_{i=1}^{N} f(S^{(i)})$$

Gives incorrect estimates when $f(S)$ is discontinuous.
e.g. for digital put $\dfrac{\partial f}{\partial S} = 0$ so estimated value of Greek is zero – clearly wrong.

Pathwise sensitivities

Extension to second derivatives is straightforward

$$
\begin{aligned}
\frac{\partial^2 V}{\partial \theta^2} &= \int \left\{ \frac{\partial^2 f}{\partial S^2} \left(\frac{\partial S(T)}{\partial \theta} \right)^2 + \frac{\partial f}{\partial S} \; \frac{\partial^2 S(T)}{\partial \theta^2} \right\} \, p_W \; \mathrm{d}W \\
&= \mathbb{E}\left[\frac{\partial^2 f}{\partial S^2} \left(\frac{\partial S(T)}{\partial \theta} \right)^2 + \frac{\partial f}{\partial S} \; \frac{\partial^2 S(T)}{\partial \theta^2} \right]
\end{aligned}
$$

with $\partial^2 S(T)/\partial\theta^2$ also being evaluated at fixed W.

However, this requires $f(S)$ to have a continuous first derivative – a problem in practice

Pathwise sensitivities

Extension to multivariate case is straightforward

$$S_k(T) = S_k(0)\ \exp\left((r - \tfrac{1}{2}\sigma_k^2)T + \sum_l L_{kl} X_l \right)$$

so

$$\log S_k(T) = \log S_k(0) + (r - \tfrac{1}{2}\sigma_k^2)T + \sum_l L_{kl} X_l$$

and hence

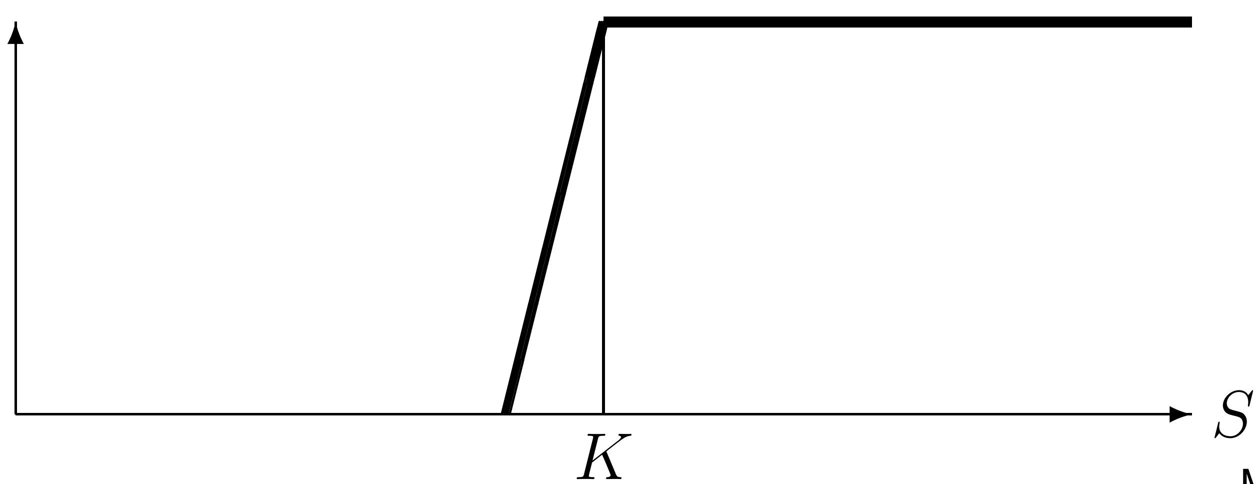
$$\frac{1}{S_k(T)}\ \frac{\partial S_k(T)}{\partial \theta} = \frac{1}{S_k(0)}\ \frac{\partial S_k(0)}{\partial \theta} - \sigma_k \frac{\partial \sigma_k}{\partial \theta} T + \sum_l \frac{\partial L_{kl}}{\partial \theta} X_l$$

Pathwise sensitivities

To handle payoffs which do not have the necessary continuity/smoothness one can modify the payoff

For digital options it is common to use a piecewise linear approximation to limit the magnitude of Δ near maturity – avoids large transaction costs

Bank selling the option will price it conservatively (i.e. over-estimate the price)



Pathwise sensitivities

The standard call option definition can be smoothed by integrating the smoothed Heaviside function

$$H_\varepsilon(S-K) = \Phi\left(\frac{S-K}{\varepsilon}\right)$$

with $\varepsilon \ll K$, to get

$$f(S) = (S-K)\,\Phi\left(\frac{S-K}{\varepsilon}\right) + \frac{\varepsilon}{\sqrt{2\pi}}\exp\left(-\,\frac{(S-K)^2}{2\,\varepsilon^2}\right)$$

This will allow the calculation of Γ and other second derivatives

Final Words

Estimating Greeks is an important task, often more important than estimating the prices

- finite differences:
  - simplest, but least accurate and most expensive
  - always use the same random numbers for both calcs
  - optimum “bump” size is a tradeoff between variance and bias
- LRM (Likelihood Ratio Method)
  - OK for discontinuous payoffs, but a little complicated
- pathwise sensitivity
  - my favourite – simple, lowest variance and least cost
  - needs continuous payoff for first derivatives, but smoothing can be used for discontinuous payoffs

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 $f'(x)$.

Performing a Taylor series expansion,

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

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

The problem with taking $\Delta x \ll 1$ is inaccuracy due to finite precision arithmetic.

Complex Variable Trick

This is a very useful "trick", which I learned about 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 331 citations according to Google Scholar.

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.

I typically 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 MATLAB functions, e.g. `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 code for a real function $f(x)$.