Mathematical Modelling
Lecture 12 – Integration

Phil Hasnip
phil.hasnip@york.ac.uk

Overview of Course

- Model construction $\longrightarrow$ dimensional analysis
- Experimental input $\longrightarrow$ fitting
- Finding a ‘best’ answer $\longrightarrow$ optimisation
- Tools for constructing and manipulating models $\longrightarrow$ networks, differential equations, integration
- Tools for constructing and simulating models $\longrightarrow$ randomness
- Real world difficulties $\longrightarrow$ chaos and fractals

A First Course in Mathematical Modeling by Giordano, Weir & Fox, pub. Brooks/Cole. Today we're in chapter 5.

Integration

- Experimental data $\longrightarrow$ fit cubic splines
- General model $\longrightarrow$ discretise the independent variable

Euler again

- Last time we looked at integrals like

$$I = \int_0^x f(x')dx'$$

- $\longrightarrow$ we needed to extrapolate
- Today we're looking at

$$I = \int_a^b f(x')dx'$$

- $\longrightarrow$ we need to interpolate

Euler again

The simplest method is to use Euler again:

- Split integral into N intervals of width h
- Approximate function over each interval by its value at start

$$\int_a^b f(x)dx' \approx h \sum_{n=0}^{N-1} f(x_n) + O(h)$$

This is a first-order method since the error is $\sim h$.

Euler again

How well does Euler do? Consider

$$
\begin{aligned}
f(x) &= \frac{1}{1+x^2} \\
\Rightarrow I &= \int_0^1 f(x)dx \\
&= \tan^{-1}(1) \\
&= \frac{\pi}{4} \\
&\approx 0.785398
\end{aligned}
$$

Euler again

Using Euler:

h	N	I	Error
0.1	10	0.859981	9.54%
0.01	100	0.792894	0.954%
0.001	1000	0.786148	0.095%

Trapezium rule

Is Euler the best we can do? No!

- Replace ‘square’ Euler approximation by trapezia
- $f(x) \approx h\left\{\frac{1}{2}f(x_0) + f(x_1) + \ldots + f(x_{N-1}) + \frac{1}{2}f(x_N)\right\} + O(h^2)$
- Now have a second-order expression
- Only need one more evaluation: $f(x_N)$

Trapezium rule

h	N	I	Error
0.1	10	0.784981	-0.053%
0.01	100	0.785394	-0.00053%
0.001	1000	0.785398	-0.0000053%

Simpson's rule

Of course we can still do better:

- We could fit parabola in intervals $\longrightarrow$ Simpson's rule
- $f(x) \approx \frac{h}{3}\left\{f(x_0) + 4f(x_1) + 2f(x_2) + \ldots + 2f(x_{N-1}) + f(x_N)\right\}$
- Some error's cancel and get a fourth-order expression
- Need even no. points, but same cost as trapezium rule!

Beyond Simpson's rule

We can improve the answer by:

- Extend to yet higher-order
- Reduce h (the interval width)

However we can be slightly cleverer than this.

Romberg method

1. Choose an integration scheme
2. Compute integral I for a given h
3. Halve the width $h \leftarrow \frac{1}{2}h$
4. Repeat from step 2 for series of widths h, $\frac{1}{2}h, \frac{1}{4}h, \ldots$
5. Extrapolate to find limit of I as $h \to 0$

This is called the Romberg method.

Gaussian quadrature

We could use a non-uniform grid of points:

- Improved efficiency...
- ... but *much* more complex
- Only really worth it when function is expensive to calculate
- Gaussian quadrature – use special sampling points and weights

Higher dimensions

So far we've looked at functions of 1 variable. What happens for 2D? 3D?

- Suppose a given scheme and h leads to 30 sampling points in 1D
- For 2D we need 30^2 points
- For 3D we need 30^3 points

Higher dimensions

So far we've looked at functions of 1 variable. What happens for 2D? 3D?

- Suppose a given scheme and h leads to 30 sampling points in 1D
- For 2D we need 30^2 points
- For 3D we need 30^3 points
- For d dimensions we need 30^d points!

Higher dimensions

So far we've looked at functions of 1 variable. What happens for 2D? 3D?

- Suppose a given scheme and h leads to 30 sampling points in 1D
- For 2D we need 30^2 points
- For 3D we need 30^3 points
- For d dimensions we need 30^d points!
- This is rather bad news...

Higher dimensions

The error in our schemes for N total sampling points is:

- Euler is first-order, so error is $\sim N^{-\frac{1}{d}}$
- Trapezium is second-order, so error is $\sim N^{-\frac{2}{d}}$
- Simpson's rule is fourth-order, so error is $\sim N^{-\frac{4}{d}}$

The higher the dimensions, the more points we need to reduce the error. Help!

Welcome to Monte Carlo

We can get around the problem of dimensions with some statistical trickery. Recall:

$$P(x \leq x' < x + dx) = p(x)dx$$

and

$$\begin{aligned}
\int_{-\infty}^{\infty} p(x)dx &= 1 \\
\int_{-\infty}^{\infty} xp(x)dx &= < x > \\
\int_{-\infty}^{\infty} (x - < x >)^2 p(x)dx &= \sigma^2
\end{aligned}$$

Welcome to Monte Carlo

The average value of any function $f(x)$ is:

$$< f(x) > = \int_{-\infty}^{\infty} f(x)p(x)dx$$

Welcome to Monte Carlo

Using discrete variables and sampling at N random points:

$$
\begin{aligned}
< x > &= \frac{1}{N}\sum_{i=1}^{N} x_i \\
S^2 &= \frac{1}{N}\sum_{i=1}^{N}(x_i - < x >)^2
\end{aligned}
$$

The estimates for the *population* are:

$$
\begin{aligned}
\mu &= < x > \\
\sigma^2 &= \frac{N}{N-1} S^2
\end{aligned}
$$

The statistical error in this estimate of the mean is

$$\sim \frac{S}{\sqrt{N}}$$

Monte Carlo integration

What has this got to do with integration? Mean-value theorem:

$$\frac{1}{b-a}\int_a^b f(x)dx \quad = \quad < f >$$

Thus if we:

Monte Carlo integration

What has this got to do with integration? Mean-value theorem:

$$\frac{1}{b-a}\int_a^b f(x)dx \quad = \quad < f >$$

Thus if we:

- choose our N points randomly

Monte Carlo integration

What has this got to do with integration? Mean-value theorem:

$$\frac{1}{b-a}\int_a^b f(x)dx \quad = \quad < f >$$

Thus if we:

- choose our N points randomly
- calculate $f(x)$ at each point

Monte Carlo integration

What has this got to do with integration? Mean-value theorem:

$$\frac{1}{b-a}\int_a^b f(x)dx \quad = \quad < f >$$

Thus if we:

- choose our N points randomly
- calculate $f(x)$ at each point
- compute the average of f

Monte Carlo integration

What has this got to do with integration? Mean-value theorem:

$$\frac{1}{b-a}\int_a^b f(x)dx \quad = \quad < f >$$

Thus if we:

- choose our N points randomly
- calculate $f(x)$ at each point
- compute the average of f
- we get the integral!

Monte Carlo integration

We do have some statistical error (σ) in this value

$$\begin{aligned}\sigma &\sim \frac{b-a}{\sqrt{N}}\sqrt{<f^2> - <f>^2} \\ &= O(N^{-\frac{1}{2}})\end{aligned}$$

but this is independent of dimensions.

Monte Carlo integration

- Pretty useless in 1D

Monte Carlo integration

- Pretty useless in 1D
- Looks much better in higher dimensions

Monte Carlo integration

- Pretty useless in 1D
- Looks much better in higher dimensions
- Monte carlo error $\sim N^{-\frac{1}{2}}$

Monte Carlo integration

- Pretty useless in 1D
- Looks much better in higher dimensions
- Monte carlo error $\sim N^{-\frac{1}{2}}$
- Recall Trapezium rule error $\sim N^{-\frac{2}{d}}$

Monte Carlo integration

- Pretty useless in 1D
- Looks much better in higher dimensions
- Monte carlo error $\sim N^{-\frac{1}{2}}$
- Recall Trapezium rule error $\sim N^{-\frac{2}{d}}$
- $\longrightarrow$ Monte carlo better when $d > 4$

Monte Carlo integration

- Pretty useless in 1D
- Looks much better in higher dimensions
- Monte carlo error $\sim N^{-\frac{1}{2}}$
- Recall Trapezium rule error $\sim N^{-\frac{2}{d}}$
- $\longrightarrow$ Monte carlo better when $d > 4$
- Monte carlo always wins for large enough d

Monte Carlo integration

We need to ensure the sampling really is random:

- Can get tables of random numbers – use one, cross it off, move to next...
- Use a computer – pseudo-random numbers
- Both often give numbers between 0 and 1, so will often need to rescale $f(x)$:

$$\int_a^b f(x)dx = \int_0^1 f(x')dx'$$

Summary

- In 1D we can use a variety of fixed-interval methods
- Can extrapolate to zero-interval-width limit
- In higher dimensions interval sampling methods are expensive
- Can use Monte carlo techniques to solve efficiently