Calculus of Finite Differences

Fitting a line to n points

Given two arbitrary points (x_0, y_0) and (x_1, y_1) you can fit a straight line through them whose equation is:

$$y = mx + c \tag{1}$$

where
$$m = \frac{y_1 - y_0}{x_1 - x_0} \tag{2}$$

and
$$c = \frac{x_1 y_0 - x_0 y_1}{x_1 - x_0} \tag{3}$$

Likewise three points can be fitted uniquely using a parabola of the form:

$$y = ax^2 + bx + c \tag{4}$$

It is easy to see that $n + 1$ points can be fitted to a polynomial of the form

$$y = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n \tag{5}$$

because we always have $n + 1$ equations to solve for $n + 1$ variables.

Finding the right general equation when n is large is tedious and usually unnecessary. It is, however, possible to find the general equation in the simpler case when the given points have x coordinates equal to 0, 1, 2, …, n. For example, suppose we have two points (0,2) and (1,5). We shall write this as a list thus:

2 5 * * *

where the stars indicate that no points exist for the cases where $x = 2, 3, 4$ etc.

Now we shall calculate the differences between successive points:

2 5 * * *
3 * * *
* * *

The initial numbers of the two lines (in bold type) are **2** and **3** and it is no coincidence that the equation we desire is $y = 3x + 2$ (or rather $y = \mathbf{2} + \mathbf{3}x$) because if we substitute $x_0 = 0$ and $x_1 = 1$ into equations (2) and (3) we find that $m = y_1 - y_0$ (which is the initial number of the second line) and $c = y_0$. (which is the initial number of the first).

Now lets try a sequence of 3 numbers:

2 5 4 * *
3 -1 * *
-4 * *
* *

The addition of a new number adds an extra line and an extra number to our list of initial numbers which is now 2, 3, -4. We have seen how in the previous case, these numbers translate easily into the coefficients of the equation we are looking for but it is immediately obvious that the equation

$y = \mathbf{2} + \mathbf{3}x - \mathbf{4}x^2$ is not going to work because the x^2 term is going to mess up the second term as well as giving the wrong answer for the third!. Nevertheless, it would be nice if we could find an equation of the form:

$$y = \mathbf{2} + \mathbf{3}x - \mathbf{4}(f(x)) \tag{6}$$

when $f(x)$ is a second order function of x. But what could $f(x)$ be?

Now we know that $f(x)$ must be zero when $x = 0$ and $x = 1$. This suggests that we might try the function $f(x) = x(x - 1)$ which fits the bill and has the value 2 when $x = 2$.

Unfortunately, this still doesn't work because it generates the sequence 2, 5, 0. It appears that the $f(x)$ term is too large and we end up subtracting 8 when we should only be subtracting 4. This suggests that $f(x)$ should be $x(x - 1)/2$ not $x(x - 1)$.

Lets see if we can prove this generally by writing out the table of differences explicitly.

$$\begin{array}{ccccc} y_0 & y_1 & y_2 & * & * \\ & y_1\text{-}y_0 & y_2\text{-}y_1 & * & * \\ & & y_2\text{-}2y_1\text{+}y_0 & * & * \\ & & * & * & * \end{array}$$

This means that the general equation we are suggesting is

$$y = y_0 + (y_1 - y_0)x + (y_2 - 2y_1 + y_0)x(x-1)/2 \tag{7}$$

If this is correct, then putting $x = 0$, 1 and 2 into the equation should generate $y = y_0$, y_1 and y_2 respectively. We know it works for the first two cases. What about the third?

$$y = y_0 + (y_1 - y_0).2 + (y_2 - 2y_1 + y_0).2.(2-1)/2$$

$$y = y_0 + 2y_1 - 2y_0 + y_2 - 2y_1 + y_0$$

$$y = y_2$$

Hooray!!

We now know that the general equation of a parabola that fits three sequential integer points is

$$y = D_0 + D_1 x + D_2 x(x-1)/2 \tag{8}$$

where D_0, D_1 and D_2 are the initial numbers of the successive lines of the difference table. It would appear very probable that the equation for 4 points would be something like:

$$y = D_0 + D_1 x + D_2 x(x-1)/2 + D_3 x(x-1)(x-2)/k \tag{8}$$

The final term is a cubic (as it should be) and it is zero when $x = 0$, 1 or 2 but it is not immediately obvious what the constant k should be. On the other hand, the appearance of the descending sequence x, $(x$-1), $(x$-2) in the numerator strongly suggests that the denominator should be 1.2.3 and that k is simply $n!$ (factorial n). i.e.

$$\boxed{y = D_0 + D_1 x + D_2 \frac{x(x-1)}{2!} + \ldots + D_n \frac{x(x-1)(x-2)\ldots(x-(n-1))}{n!}} \tag{9}$$

This remarkable formula was discovered and proved by Newton and published in Principia in 1687. It is the discrete version of the Taylor series.

In order to prove that this is indeed the correct expression we have got to show that each time you 'differentiate' it (i.e. take successive differences) you get the sequence D_0, D_1, D_2 etc. when you put $x = 0$.

Equation (9) is the general equation of a polynomial of order n and it gives the value of y at any value of x. If we substitute $(x+1)$ for x we get the 'next' value of y in the sequence namely:

$$y' = D_0 + D_1(x+1) + D_2\frac{(x+1)x}{2!} + \dots + D_n\frac{(x+1)x(x-1)\dots(x-n)}{n!} \tag{10}$$

Now we subtract equation (9) from equation (10) to obtain an equation giving the first line of differences (which we can also call y without confusion).

$$y = D_1((x+1)-x) + D_2\frac{(x+1)x - x(x-1)}{2!} + D_3\frac{(x+1)x(x-1) - x(x-1)(x-2)}{3!} + \dots \tag{11}$$

$$y = D_1 + D_2\frac{2x}{2!} + D_3\frac{3x(x-1)}{3!} + \dots \tag{12}$$

$$y = D_1 + D_2 x + D_3\frac{x(x-1)}{2!} + \dots \tag{13}$$

Putting $x = 0$ gives $y = D_1$.

Now equation (9) has the identical form to equation (13) so when this equation is 'differentiated' the same thing will happen and the equation reduces to:

$$y = D_2 + D_3 x + \dots \tag{14}$$

and so on until we run out of terms.

How beautiful is that?!!!!

Carr Bank

August 2016