Parabola Volume 27, Issue 3 (1991)

IF IT FITS $\cdots$?

Peter Brown*

In experimental science in past eras data was collected from an experiment and some relationship between quantities in the form of equations was sought. In this article I wish to introduce to you three different methods of the technique known as curve fitting.

The first method, which is only approximate, is called $\cdots$

Line of Best Fit:

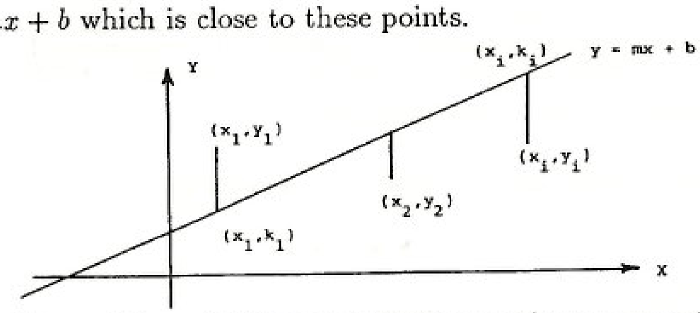
Suppose we look at the maximum amount of a certain substance that will dissolve in 1 litre of water at a certain temperature.

Temperature (°C)	Amount of substance (gm)
62	115
63	118
64	121
65	125
66	129
67	132
68	136

If we graph these numbers with amount of substance as our y-axis and temperature the x-axis, we see that the points lie approximately on a straight line.

Our problem is to find the 'best' straight line to describe the data.

First some theory. Suppose we have points (x_1, y_1), $(x_2, y_2) \cdots, (x_n, y_n)$ and try to find the line $y = mx + b$ which is close to these points.



Suppose (x_i, k_i) is a point on the line and also lies on the same vertical line as (x_i, y_i) one of our given points. It is not hard to see that the square of distance between (x_i, y_i)

* Peter is a Pure Mathematician at the University of New South Wales

and (x_i, k_i) is $(y_i - k_i)^2$. This gives some measure of the closeness of (x_i, y_i) to $y = mx + b$. In order to ensure that our line is the best one we look at the sum of **all** of these squared distances from each of our given points to the line, and calculate

$$\begin{aligned} S &= (y_1 - k_1)^2 + (y_2 - k_2)^2 + \cdots + (y_n - k_n)^2 \\ &= \sum_{i=1}^{n} (y_i - k_i)^2 \qquad (1) \end{aligned}$$

we now try to make this as **small** as possible. We know that (x_i, k_i) lies on the line $y = mx + b$, so

$$k_i = mx_i + b$$

Substituting into (1) we have

$$S = \sum_{i=1}^{n} (y_i - mx_i - b)$$

S is a function of two variables m and b. Using calculus (unfortunately university level calculus), we can show that S is smallest when

$$b = \frac{1}{n}\left[\sum_{i=1}^{n} y_i - m \sum_{i=1}^{n} x_i\right]$$

and

$$m = \frac{\sum_{i=1}^{n} x_i y_i - b \sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2}$$

These equations look formidable but are not too hard to apply.

Referring back to our table of temperatures and amounts, we can calculate the following table:

Temperature	Amount		
x_i	y_i	$x_i y_i$	x_i^2
62	115	7130	3844
63	118	7434	3969
64	121	7744	4096
65	125	8125	4225
66	129	8514	4356
67	132	8844	4489
68	136	9248	46 24
$\Sigma x_i = 455$	$\Sigma y_i = 876$	$\Sigma x_i y_i = 57039$	$\Sigma x_i^2 = 29603$

Here $n = 7$, so our equations become

$$b = \frac{1}{7}(876 - 455m)$$

and

$$m = \frac{57039 - 455b}{29603}$$

solving for m and b we get

$$m \simeq 3.54$$

$$b \simeq -104.68$$

So our line is approximately

$$y = 3.54x - 104.68$$

You can check by substituting (x_i, y_i) into the equation to see how close the line is.

If you found all the computation tedious, then take heart, since most modern calculators have all this built in.

If your calculator has buttons marked (x, y), a, b then putting the calculator into statistics mode we enter the x and y values by pushing 62 $\boxed{(x, y)}$ 115 $\boxed{\text{DATA}}$ etc. After all the scores are entered we push $\boxed{a}$ and then $\boxed{b}$ and the line is $y = bx + a$. The button marked $\boxed{r}$ gives some idea of how closely the points fit the line. If r is close to 1 then the points are close to the line.

Quadratic Interpolation

Suppose I give you three points in the plane (not lying on a straight line), say $(1, 3)$, $(2, 9)$ and $(3, 17)$. Can you find a parabola of the form $y = ax^2 + bx + c$ on which these points lie? Obviously one way to procede is to substitute each of the points to get

$$3 = a + b + c$$

$$9 = 4a + 2b + c$$

$$17 = 9a + 3b + c$$

This system is not hard to solve but with more complicated points it is very tedious. Moreover we would like to get a 'formula' for the quadratic passing through 3 general points (x_1, y_1), (x_2, y_2), (x_3, y_3).

To do this we must 'cook up' a quadratic with the right property. Let $f(x)$ be a quadratic and write

$$f(x) = g_1(x) + g_2(x) + g_3(x)$$

where $g_i(x)$ is a quadratic.

We want $g_1(x)$ to have the property that

$$g_1(1) = 3 \text{ while } g_1(2) = 0 \text{ and } g_1(3) = 0$$

One guess is

$$g_1(x) = 3\frac{(x-2)(x-3)}{(1-2)(1-3)}$$

This gives $g_1(2) = g_1(3) = 0$ and $g_1(1) = 3$ since the factors cancel.

Similarly put

$$g_2(x) = 9\frac{(x-1)(x-3)}{(2-1)(2-3)}$$

this gives $g_2(1) = g_2(3) = 0$ and $g_2(2) = 9$.

Finally put

$$g_3(x) = 17\frac{(x-1)(x-2)}{(3-1)(3-2)} \text{ with } g_3(1) = g_3(2) = 0 \text{ and } g_3(3) = 17$$

Hence (after simplifying the demoninators)

$$f(x) = \frac{3}{2}(x-2)(x-3) - 9(x-1)(x-3) + \frac{17}{2}(x-1)(x-2)$$

You should check that $f(1) = 3$, $f(2) = 9$ and $f(3) = 17 \cdot f(x)$ can be further simplified to

$$f(x) = x^2 + 3x - 1.$$

Now for our general formula for the quadratic passing through $(x_1, y_1), (x_2, y_2), (x_3, y_3)$. Put

$$g_1(x) = y_1\frac{(x-x_2)(x-x_3)}{(x_1-x_2)(x_1-x_3)}$$

Check that $g_1(x_2) = g_1(x_3) = 0$ and $g_1(x_1) = y_1$ (again the factors will cancel)$\cdot$ $g_2(x)$ and $g_3(x)$ are constructed similarly, giving

$$f(x) = y_1\frac{(x-x_2)(x-x_3)}{(x_1-x_2)(x_1-x_3)} + y_2\frac{(x-x_1)(x-x_3)}{(x_2-x_1)(x_2-x_3)} + y_3\frac{(x-x_1)(x-x_2)}{(x_3-x_1)(x_3-x_2)}$$

This is certainly an example of 'creative cooking'.

The method (and formula) given above is called Lagrange interpolation and can be extended to higher degree equations. However as you see it gets messy. For higher degree problems we use instead ···

Newton's Formula: We need some new notation. Define $x^{(n)}$ (read "x falling factorial n") by

$$x^{(n)} = x(x-1)\cdots(x-n+1)$$

For example

$$x^{(2)} = x(x-1)$$

$$x^{(3)} = x(x-1)(x-2)$$

etc.

and define $\Delta f(x)$ by $\Delta f(x) = f(x+1) - f(x)$.

You can prove by induction that $\Delta x^{(n)} = nx^{(n-1)}$ for example

$$\begin{aligned}\Delta x^{(3)} &= (x+1)^{(3)} - x^{(3)} \\ &= (x+1)x(x-1) - x(x-1)(x-2) \\ &= 3x(x-1) \\ &= 3x^{(2)}.\end{aligned}$$

For those of you who know some calculus, you should compare this with $\frac{d}{dx}(x^n) = nx^{n-1}$.

We can extend to powers of Δ by defining, for example, $\Delta^2 f(x) = \Delta(\Delta f(x))$ and so on.

So $\Delta^2 x^{(3)} = \Delta 3x^{(2)} = 6x^{(1)}$

Now suppose we have a polynomial of degree 4 say.

$$f(x) = ax^4 + bx^3 + cx^2 + dx + e$$

This can be written in the form

$$f(x) = Ax^{(4)} + Bx^{(3)} + Cx^{(2)} + Dx^{(1)} + E$$

The problem is to find A, B, C, D and E in terms of $f(0)$, $\Delta f(0)$, etc.

Clearly $E = f(0)$. Applying Δ to both sides we get:

$$\Delta f(x) = 4Ax^{(3)} + 3Bx^{(2)} + 2Cx^{(1)} + D$$

So $\triangle f(0) = D$

$$\triangle^2 f(x) = 12Ax^{(2)} + 6Bx^{(1)} + 2C$$

So

$$\triangle^2 f(0) = 2C \text{ or } C = \frac{1}{2}\triangle^2 f(0)$$

We can continue the process to get

$$B = \frac{1}{3!}\triangle^3 f(0) \text{ and } A = \frac{1}{4!}\triangle^4 f(0)$$

Hence we have

$$f(x) = \frac{\triangle^4 f(0)}{4!}x^{(4)} + \frac{\triangle^3 f(0)}{3!}x^{(3)} + \frac{\triangle^2 f(0)}{2!}x^{(2)} + \frac{\triangle f(0)}{1!}x^{(1)} + f(0)$$

In general if $f(x)$ is a polynomial of degree n then, using induction, we can show that

$$f(x) = \frac{\triangle^n f(0)}{n!}x^{(n)} + \frac{\triangle^{n-1} f(0)}{(n-1)!}x^{(n-1)} + \cdots + \frac{\triangle f(0)}{1!}x^{(1)} + f(0)$$

We will now see how to use this formula to find polynomials to fit data.

Suppose we wish to find a polynomial to fit the points

$$(0,0),\ (1,2)\ (2,20),\ (3,90),\ (4,272),\ (5,650).$$

We construct a table of differences

x	$f(x)$	$\triangle f(x)$	$\triangle^2 f(x)$	$\triangle^3 f(x)$	$\triangle^4 f(x)$	$\triangle^5 f(x)$
0	0	2	16	36	24	0
1	2	18	52	60	24	
2	20	70	112	84		
3	90	182	196			
4	272	378				
5	650	*				

* each entry for $\triangle f(x)$ is given by $f(x+1) - f(x)$ e.g. 182 = 272 - 90.

The top row of the table gives the value of $f(0),\ \triangle f(0), \triangle^2 f(0)$ etc.

Hence our polynomial is

$$\begin{aligned} f(x) &= f(0) + \frac{\triangle f(0)}{1!}x + \frac{\triangle^2 f(0)}{2!}x^{(2)} + \frac{\triangle^3 f(0)}{3!}x^{(3)} + \frac{\triangle^4 f(0)}{4!}x^{(4)} \\ &= 0 + \frac{2x}{1!} + \frac{16x^{(2)}}{2!} + \frac{36x^{(3)}}{3!} + \frac{24x^{(4)}}{4!} \\ &= 2x + 8x(x-1) + 6x(x-1)(x-2) + x(x-1)(x-2)(x-3) \end{aligned}$$

which simplifies to $f(x) = x^4 + x^2$.

As a final example, we will find a formula for

$$1^2 + 2^2 + \cdots + n^2$$

Let $f(x) = 1^2 + 2^2 + \cdots + x^2$

$$f(0) = 0$$
$$f(1) = 1^2 = 1$$
$$f(2) = 1^2 + 2^2 = 5$$
$$f(3) = 1^2 + 2^2 + 3^2 = 14$$
$$f(4) = 1^2 + 2^2 + 3^2 + 4^2 = 30$$
$$f(5) = 1^2 + 2^2 + 3^2 + 4^2 + 5^2 = 55$$

We will **assume** that $f(x)$ can be written as a polynomial of degree 3. This assumption is not obvious but we will see that differences of order 4 and higher are 0, and so our assumption will be reasonable.

x	$f(x)$	$\Delta f(x)$	$\Delta^2 f(x)$	$\Delta^3 f(x)$	$\Delta^4 f(x)$	$\Delta^5 f(x)$
0	0	1	3	2	0	0
1	1	4	5	2	0	
2	5	9	7	2		
3	14	16	9			
4	30	25				
5	55					

So $f(x) = 0 + \frac{1}{1!}x + \frac{3}{2!}x^{(2)} + \frac{2}{3!}x^{(3)} = x + \frac{3}{2}x(x-1) + \frac{1}{3}x(x-1)(x-2)$ which simplifies to

$$f(x) = \frac{1}{6}x(x+1)(2x+1)$$

whence

$$1^2 + 2^2 + \cdots + n^2 = \frac{1}{6}n(n+1)(2n+1)$$

Functions which are **not** polynomials can be approximated by polynomials using this method.