Approximation: An Introduction to Interpolation and Curve Fitting

We often know or record data at a set of points, and it is required that we estimate the value that the data would have taken at other points. This may be expressed as knowing the value of a function $f(x)$ at a set of n points in a range $[a,b]$ with $a = x_1 < x_2 < \cdots < x_n = b$ and we are required to determine an estimate of $f(x)$ at any value of $\in [a,b]$.

The problem seems simple, it may be considered to be equivalent to the task of plotting a graph and joining up the points. However, is this always a reasonable thing to do; for example, can we presume that $f(x)$ is smooth? Often the values $f(x_i)$ may be measured or are the result of numerical calculations and hence may include a margin of error; is it always reasonable to join up points that are in error, or to draw a much smoother line through the points?

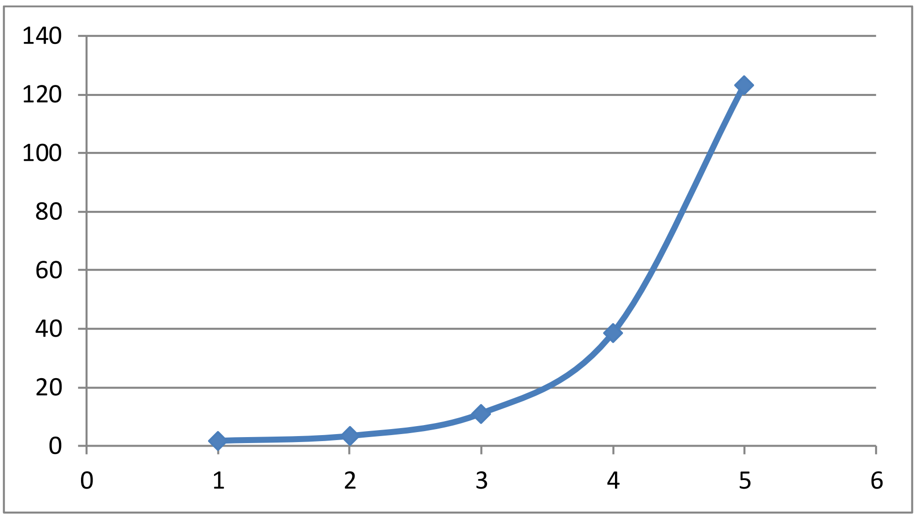
Interpolation is this process of joining up the points, curve fitting is the process of finding the best fitting line of a pre-determined form through the points . There is clearly not a unique solution and there is bias in the sense that we have to predetermine some properties of the function, for example that it should be smooth. The result of interpolation is termed the *interpolant.* Let us look a few illustrative examples.

Example 1

Consider the following data.

x_i	$f(x_i)$
1	1.718282
2	3.389056
3	11.08554
4	38.59815
5	123.4132

Typically an interpolation with a smooth function would give a result like the following.

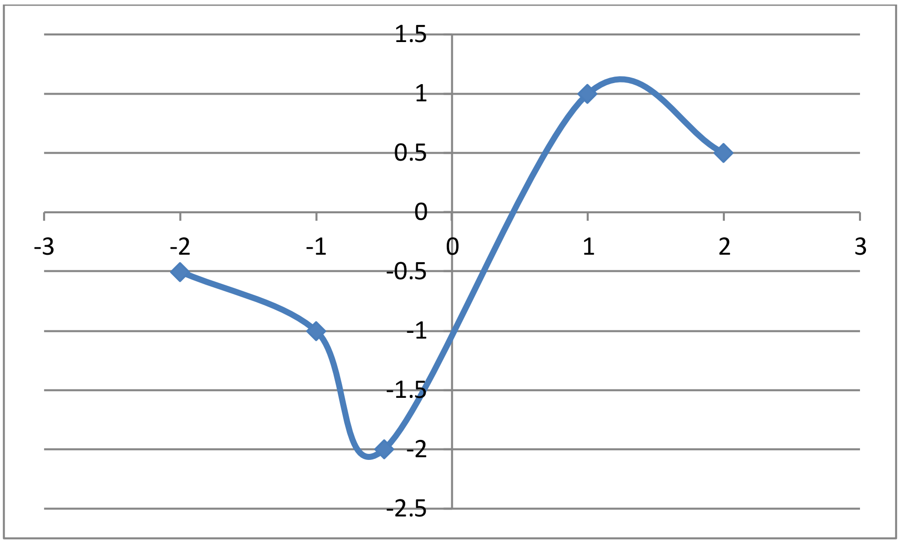


Example 2

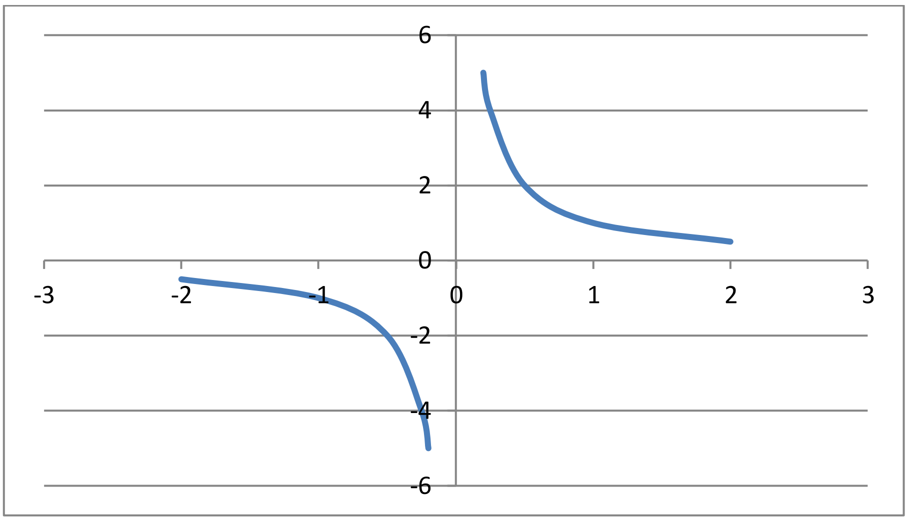
Consider the following data.

x_i	$f(x_i)$
-2	-0.5
-1	-1
-0.5	-2
1	1
2	0.5

Typically an interpolation with a smooth function would give a result like the following.



This seems reasonable until it is revealed that the points are from the function $f(x) = \frac{1}{x}$ and therefore has a singularity at $x = 0$ and the graph is actually has the form illustrated in the graph below. This is an example of the bias required to produce the interpolant; the bias being for example the (mistaken) presumption that the interpolant is smooth.

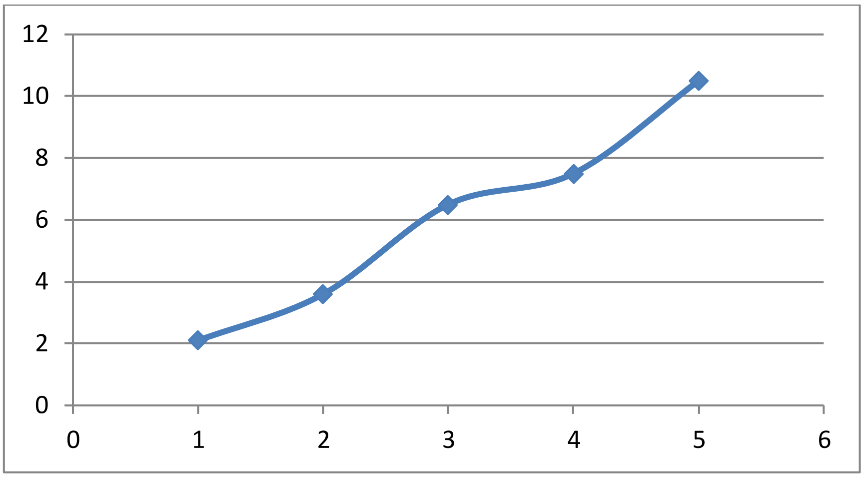


Example 3

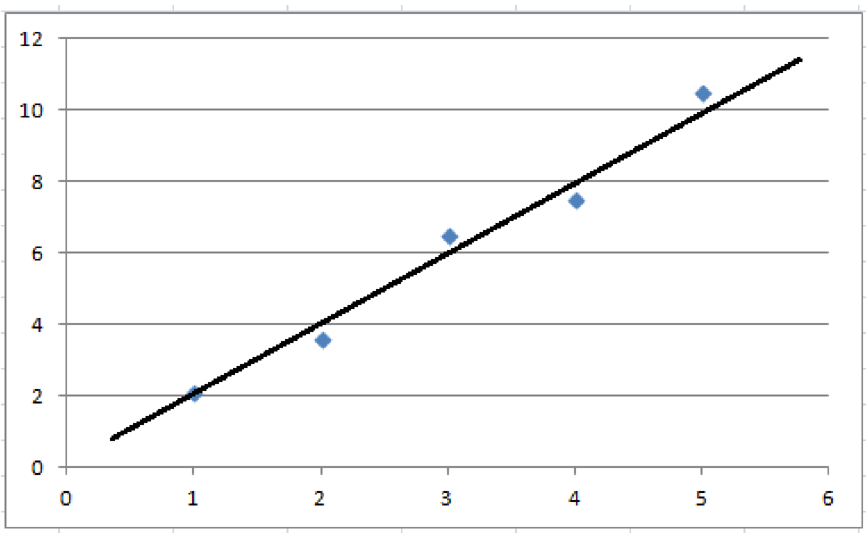
Consider the following data.

x_i	$f(x_i)$
1	2.1
2	3.6
3	6.5
4	7.5
5	10.5

Typically an interpolation with a smooth function would give a result like the following.



However, it may also be the case that we expect the points to fall on a straight line. For example the table could be the measured extension of a spring $f(x_i)$ under different forcings x_i and we expect the data to obey Hooke's Law[1] and hence that the quantities are proportional and that a straight line graph is expected. In this case it is appropriate to find the best fit straight line through the points as illustrated in the following graph.



[1] Hooke's Law

In this last example, the line does not normally pass through all the points. Some rationale is often attached to this such as that the difference between the point and the line is due to experimental or measurement error. The automation of the curve fitting process is termed *regression.*

Approximation is useful so that we find an approximation to the unknown function at the values of x between the data points, we can perform common operations such as differentiation[2] and integration[3] or draw other useful information from the data. A related topic is that of *extrapolation*. If we wish to estimate the function $f(x)$ outside of the set of data points then this is called extrapolation.

The functions that are most commonly used to form the basis of approximation include polynomials, trigonometric functions[4], exponential functions[5] and rational functions[6]. However, of these polynomial interpolation[7] is by far the most commonly used.

In general, the process of approximation – interpolation (extrapolation) or curve fitting – needs to be approached with some care. If we are given a set of data points alone then it is not necessarily reasonable to interpolate them or fit a line through them. In order to make progress we need further information about the data to make progress, such as the points lie on a smooth curve, the points should lie on a straight line, or the source of the data. This further information facilitates the choice of an appropriate approximation method.

[2] Differentiation
[3] Integration
[4] Trigonometric Functions
[5] Logarithm and Exponential Functions
[6] Rational Functions
[7] Polynomial Interpolation