Predicting the Future

SANKARAN VISWANATH

Said Tweedledum to Tweedledee[1]
"Now looking back, I clearly see
That life has been indeed carefree.
But what the future holds in store,
Of happiness, less or more?
I do so wish I knew before!"

Said Tweedledee: "Oh, that's easy.
You'll soon admit, I guarantee
That here's the perfect recipe:
To estimate your future state,
To guess, but still be accurate,
Just simply ex-trapolate!"

We often find ourselves needing to predict the future value of something based on past trends. For instance, climate scientists have been trying to estimate the future rise in ocean temperatures based on (among other things) temperature data of the last 100 or more years. Or maybe you are a cricket enthusiast and want to predict how many centuries your favourite batsman will score this year based on his scoring statistics for past years. We all probably have our own ways of arriving at an estimate! In this article, let us explore one approach.

[1]Tweedledum and Tweedledee appear in Lewis Carroll's *Through the Looking Glass*. Carroll, whose real name was Charles Lutwidge Dodgson, was (among other things) a mathematician who taught at Oxford. The poem above is inspired by Carroll's *literary nonsense* style which is the hallmark of the Alice books.

Keywords: Patterns, polynomials, extrapolation, interpolation

First, imagine this: you are growing a bacterial culture in the lab, and would like to understand how the population of bacteria changes over time. So, at the end of every hour you measure the bacterial population, and write down the sequence of measurements as follows: $a_1, a_2, a_3, \ldots$, where a_n denotes the population after n hours. Now, suppose you find that the first five measurements are: $1, 4, 9, 16, 25$. Can you "predict" the population after 6 hours?

You probably guessed 36, right? If asked why you think this a good estimate, you would probably say "the pattern of the previous measurements suggests that the population is varying as the square of the natural numbers, i.e., $1^2,\ 2^2,\ 3^2, \ldots$; so at the end of 6 hours, it must be $6^2 = 36$."

This illustrates the following point: to predict a future value precisely, one has to presuppose that there is some pattern or orderly mechanism by which the values are being generated. If the values are just a collection of randomly generated numbers, then future values bear no relation to past ones and precise prediction is impossible.[2]

The simplest patterns are the sequences of powers of the natural numbers; for example, the sequences: (a) $1, 2, 3, 4, \ldots$ (b) $1, 4, 9, 16, \ldots$ (c) $1, 8, 27, 64, \ldots$. As before, letting a_n denote the n^{th} term of a sequence, the sequences above are given by the explicit formulas: (a) $a_n = n$, (b) $a_n = n^2$, and (c) $a_n = n^3$. Here n ranges over the natural numbers. Similarly one can look at the sequence of d^{th} powers, $a_n = n^d$ where d is some fixed natural number.

A somewhat more general sequence is obtained by taking combinations of powers of n; for example

$$a_n = n^3 - 4n^2 + 7n + 1. \tag{1}$$

The first few terms of the sequence (obtained for $n = 1, 2, 3, 4$) are: $5, 7, 13, 29$. An expression of the above form is called a *polynomial* in n, and the highest power of n that occurs (3 in the above example) is called the *degree* of the polynomial. The most general form of a polynomial of degree $\leq d$ is:

$$c_d\, n^d + c_{d-1}\, n^{d-1} + \cdots + c_1\, n + c_0 \tag{2}$$

Here $c_0, c_1, \ldots, c_d$ are real numbers. They are called the *coefficients* of the polynomial. The degree of this polynomial is the largest value of m for which $c_m \neq 0$. For instance, the degree 3 polynomial in equation (1) has coefficients $c_0 = 1, c_1 = 7, c_2 = -4, c_3 = 1$.

Let us now formulate a specific version of our original *prediction problem* for the bacterial populations. Recall that we have a sequence of measurements $a_1, a_2, a_3, \ldots$, with a new term added to this list every hour. Fix $d \geq 1$; suppose we are given the information that the bacterial population varies as a polynomial in n of degree $\leq d$, i.e., a_n is given by a formula of the form of Equation (2).

[2]There is a more nuanced version of this: if the values are randomly generated, but from a given *probability distribution*, i.e., when the probability of taking each value is known, then one can still make predictions which hold true with some probability. We won't deal with this situation here.

Problem 1. Determine the coefficients $c_0, c_1, \ldots, c_d$ of this polynomial, using only the first few population values.

Observe that in our problem statement, we don't yet quantify the word "few"; we are allowed to use any finite number of initial values; this corresponds to the measurements that have been made until a given point of time. Knowing these, we would like to determine the coefficients of the polynomial. Once the coefficients are known, the polynomial is fully specified, and we can compute the value of a_n for any desired value of n. We are thus asking if we can just take finitely many measurements, and use those to determine what *all* future measurements will be. This process is usually termed *extrapolation*. [3]

Example 1. Suppose a_n is known to be a polynomial in n, of degree ≤ 1 (i.e., degree is 0 or 1). Given $a_1 = 3$ and $a_2 = 7$, find the formula for a_n.

To solve this, we start with the general form in Equation (2), with $d = 1$, i.e., $a_n = c_1\, n + c_0$. Here, the coefficients c_0, c_1 are unknown. Substituting $n = 1, 2$ we obtain two equations for these two unknowns:

$$c_1 + c_0 = 3,$$
$$2c_1 + c_0 = 7.$$

Solving, we obtain $c_1 = 4, c_0 = -1$; thus $a_n = 4n - 1$ is the desired formula.

This example suggests a general method; suppose d is any given natural number and we know that a_n is a polynomial of degree $\leq d$. This means a_n is given by Equation (2), but the $(d+1)$ coefficients $c_0, c_1, \ldots, c_d$ are unknown. If we also knew the values of $a_1, a_2, \ldots, a_{d+1}$, then we could substitute $n = 1, 2, \ldots, (d+1)$ in Equation (2) to obtain $d+1$ linear equations in the $d+1$ unknowns. Solving these equations would give us the values of $c_0, c_1, \ldots, c_d$. Why don't we try this in a slightly larger example, before proceeding ahead?

Exercise 1. You are given that a_n is a polynomial in n of degree ≤ 3 and that $a_1 = 1$, $a_2 = 2$, $a_3 = 9$, $a_4 = 28$. Using the above strategy, determine the coefficients c_0, c_1, c_2, c_3. You can use any method you like to solve the equations you get, but a simple elimination of variables (by subtracting successive pairs of equations) will work.

A second approach. Now, in this particular example, it turns out that another method of solution is possible. Observe that the given terms of the sequence are just one more than the cubes of the first four non-negative integers, i.e., $0^3 + 1, 1^3 + 1, 2^3 + 1, 3^3 + 1$. Thus, the first four terms are given by the simple formula: $(n-1)^3 + 1$ for $n = 1, 2, 3, 4$.

Let us define $b_n = (n-1)^3 + 1$; then we know the following: (i) Both a_n and b_n are polynomials in n of degree ≤ 3 (in fact, b_n has degree exactly 3 as is evident from the above formula, but we won't need this). (ii) $a_n = b_n$ for $n = 1, 2, 3, 4$.

We claim this implies that $a_n = b_n$ for all $n = 1, 2, \ldots$; in other words the formula for a_n is just $a_n = (n-1)^3 + 1 = n^3 - 3n^2 + 3n$; you should therefore have got $c_0 = 0$, $c_1 = 3$, $c_2 = -3$, $c_3 = 1$ in Exercise 1; did you? In fact, this is more generally true:

[3] A closely related term is *interpolation*. This refers to the process of computing the value at an intermediate time that lies between two measurement times. In our case, both these come down to the same problem, that of determining the coefficients of the polynomial.

Theorem 1. *Let $d \geq 1$ and suppose a_n, b_n are both polynomials in n of degree $\leq d$. If the first $d+1$ terms of both sequences match, then the sequences are identical, i.e., $a_n = b_n$ for $1 \leq n \leq (d+1)$ implies $a_n = b_n$ for all $n \geq 1$.*

Proof. To prove Theorem 1, it is better to enlarge our perspective a little, and work with *functions* instead of sequences. A *polynomial function* or *polynomial in x* of degree $\leq d$ is an expression of the form:

$$f(x) = c_d x^d + c_{d-1} x^{d-1} + \cdots + c_1 x + c_0$$

where c_i are real numbers for $i = 0, 1, \ldots, d$ and the variable x can take any real value.

The degree, denoted $\deg f(x)$, is as before the largest value of m for which $c_m \neq 0$. If all c_m are zero, i.e., $f(x)$ is the zero polynomial, then its degree is not defined. If $f(x)$ and $g(x)$ are nonzero polynomials, then so is their product and $\deg(f(x)g(x)) = \deg f(x) + \deg g(x)$. For instance, $f(x) = x^2 - x$ has degree 2, $g(x) = x^3 + x^2 + x + 1$ has degree 3 and their product $f(x)g(x) = x^5 - x$ has degree 5.

Given a sequence a_n which is a polynomial in n, we can replace n by x to construct a polynomial function $f(x)$. For example, if $a_n = n^3 - 4n^2 + 7n + 1$, then $f(x) = x^3 - 4x^2 + 7x + 1$. To get the sequence back from the function, we note that $a_n = f(n)$ for $n = 1, 2, 3, \ldots$. Using this, it is now easy to see that the following theorem implies Theorem 1.

Theorem 2. *Let $d \geq 1$ and suppose $f(x), g(x)$ are polynomials of degree $\leq d$. Let $x_1, x_2, \ldots, x_{d+1}$ be any $d+1$ distinct real numbers. If $f(x_n) = g(x_n)$ for $n = 1, 2, \ldots, d+1$, then $f(x) = g(x)$ for all real x.*

Proof. To prove Theorem 2, we first set $h(x) = f(x) - g(x)$ and observe that (i) $h(x)$ has degree $\leq d$, and (ii) $h(x_n) = 0$ for $n = 1, 2, \ldots, d+1$. We now use the following important lemma:

Lemma 1. *Let $p(x)$ be a polynomial and suppose $p(a) = 0$ for some real number a. Then $p(x) = (x-a)q(x)$ for some polynomial $q(x)$.*

Proof. To prove the lemma, we use the *division algorithm* to write

$$p(x) = (x-a)q(x) + r(x),$$

where $r(x)$ is the remainder and $q(x)$ is the quotient. Here, since $x - a$ is of degree 1, the remainder $r(x)$ has degree < 1, i.e., it is just a constant (a polynomial of degree 0). Evaluating both sides at $x = a$ shows $r(x) = 0$. □

Now, back to the proof of Theorem 2.

Since $h(x_1) = 0$, the lemma implies $h(x) = (x - x_1)\, q_1(x)$ for some polynomial $q_1(x)$. Next, $h(x_2) = 0$ implies $q_1(x_2) = 0$ and by the lemma again, $q_1(x) = (x - x_2)\, q_2(x)$. Repeating this process, we obtain finally:

$$h(x) = (x - x_1)\,(x - x_2) \cdots (x - x_{d+1})\; q_{d+1}(x). \tag{3}$$

We now claim that $q_{d+1}(x)$ is the zero polynomial. If not, then the left side has degree $\leq d$, while the right side has degree $\geq d + 1$. This contradiction shows that $q_{d+1}(x) = 0$ and hence $h(x) = 0$. This proves Theorem 2, and thereby Theorem 1 as well. □

Thus, a polynomial function $f(x)$ of degree $\leq d$ is uniquely determined once its values at any $d+1$ distinct points are given. So, if you can somehow produce one polynomial that takes the prescribed values at those points (as in the second approach above, where our candidate polynomial $(x-1)^3 + 1$ had the required values at $x = 1, 2, 3, 4$) then you can be sure that that is indeed the solution.

A systematic method to find the polynomial

The next natural question is: does such a polynomial always exist, and if so, can we find it systematically, without having to make informed guesses?

Problem 2. Fix $d \geq 1$. Let $x_1, x_2, \ldots, x_{d+1}$ be *distinct* real numbers, and let $y_1, y_2, \ldots, y_{d+1}$ be any (not necessarily distinct) real numbers. Does there exist a polynomial $f(x)$ of degree $\leq d$ such that $f(x_n) = y_n$ for all $n = 1, 2, \ldots, d+1$?

Let us convince ourselves that the answer is 'Yes'. [4]

For $d = 1$, the polynomial is easy to produce by following the same procedure as in Example 1. It turns out to be (as you should check!):

$$f(x) = (y_2 - y_1)\,\frac{x - x_1}{x_2 - x_1} + y_1 \tag{4}$$

Next we show that such a polynomial exists in general, by mathematical induction on d. Let $d \geq 2$; our induction hypothesis is that the desired result holds for $d-1$. Considering only x_i and y_i for $1 \leq i \leq d$, the induction hypothesis ensures that there is a polynomial $g(x)$ of degree $\leq d-1$ satisfying $g(x_i) = y_i$ for $i = 1, 2, \ldots, d$. The polynomial $f(x)$ that we seek should have degree $\leq d$ and satisfy $f(x_i) = y_i$ for $i = 1, 2, \ldots, d+1$. This means in particular that $h(x) = f(x) - g(x)$ has degree $\leq d$ and vanishes at $x_1, x_2, \ldots, x_d$. By repeated application of Lemma 1, we obtain:

$$h(x) = C\,(x - x_1)\,(x - x_2) \cdots (x - x_d)\,,$$

where C is a constant. We can now solve for C by plugging in $x = x_{d+1}$. Carrying out this process, we obtain finally:

$$f(x) = (y_{d+1} - g(x_{d+1}))\;\frac{(x - x_1)\,(x - x_2) \cdots (x - x_d)}{(x_{d+1} - x_1)\,(x_{d+1} - x_2) \cdots (x_{d+1} - x_d)} + g(x), \tag{5}$$

a polynomial that satisfies all the required conditions. □

The above argument not only proves the existence of such a polynomial, it also allows us to construct it step-by-step by use of Equations (4) and (5). To see this, for each $n = 1, 2, \ldots, d$, let $f_n(x)$ denote the unique polynomial of degree $\leq n$ satisfying $f_n(x_i) = y_i$ for $i = 1, 2, \ldots, n+1$. The above argument shows that (i) $f_1(x)$ is given by the right hand side of Equation (4), and (ii) for all $2 \leq n \leq d$,

$$f_n(x) = (y_{n+1} - f_{n-1}(x_{n+1}))\;\prod_{i=1}^{n}\frac{x - x_i}{x_{n+1} - x_i} + f_{n-1}(x)$$

This recursion relation can be used repeatedly to find $f_d(x)$.

Exercise 2. Using the above procedure, find the polynomial $f(x)$ of degree ≤ 3 such that $f(1) = 1, f(2) = 2, f(3) = 9, f(4) = 28$. Check that your answer matches what you got for Exercise 1.

The Wikipedia articles given in the references are good starting points to get more information about the subject matter of this article.

[4] One way to do this is along the lines of Exercise 1, i.e., write out a system of linear equations for the unknown coefficients, and argue that this system always has a solution. This can be done, but requires some facts about matrices. In particular, the famous *Vandermonde matrix* makes an appearance here, and the fact that it has nonzero *determinant* implies the result. For more, see the references at the end.

Another systematic approach to interpolation

Let us begin by rewriting Equation (4) in a more "symmetrical" form as follows:

$$f(x) = y_1 \left(\frac{x - x_2}{x_1 - x_2}\right) + y_2 \left(\frac{x - x_1}{x_2 - x_1}\right).$$

This, we recall, is the unique polynomial of degree ≤ 1 satisfying $f(x_n) = y_n$ for $n = 1, 2$. Let us define

$$p_1(x) = \frac{x - x_2}{x_1 - x_2} \text{ and } p_2(x) = \frac{x - x_1}{x_2 - x_1}.$$

We then observe $f(x) = y_1\, p_1(x) + y_2\, p_2(x)$. The key properties of these polynomials are:

(1) $p_1(x)$ and $p_2(x)$ have degree ≤ 1.
(2) $p_1(x_1) = 1$, $p_1(x_2) = 0$; $p_2(x_1) = 0$, $p_2(x_2) = 1$.

This suggests the following idea to solve Problem 2 in general. Given x_n, y_n, $1 \leq n \leq d+1$ as in Problem 2, we wish to find a polynomial $f(x)$ of degree $\leq d$ such that $f(x_n) = y_n$ for $n = 1, 2, \ldots, d+1$. Let us first try to find polynomials $p_1(x), p_2(x), \ldots, p_{d+1}(x)$ satisfying the following properties:

(1) $p_n(x)$ has degree $\leq d$ for all $1 \leq n \leq d+1$.
(2) For all $1 \leq n, m \leq d+1$,

$$p_n(x_m) = \begin{cases} 1 & \text{if } n = m \\ 0 & \text{if } n \neq m \end{cases}$$

Suppose we can manage this, then

$$f(x) = y_1\, p_1(x) + y_2\, p_2(x) + \cdots + y_{d+1}\, p_{d+1}(x)$$

is a polynomial satisfying all the required conditions (check this!). Now, it only remains to find the $p_n(x)$. We know that $p_n(x) = 0$ for $x = x_m$, $m \neq n$. Using Lemma 1 repeatedly as before we obtain $p_n(x) = C(x - x_1)(x - x_2) \cdots \widehat{(x - x_n)} \cdots (x - x_{d+1})$, where C is a constant, and $\widehat{(x - x_n)}$ means that term is omitted from the product. We solve for C by plugging in $x = x_n$ and obtain:

$$p_n(x) = \frac{(x - x_1)(x - x_2) \cdots \widehat{(x - x_n)} \cdots (x - x_{d+1})}{(x_n - x_1)(x_n - x_2) \cdots \widehat{(x_n - x_n)} \cdots (x_n - x_{d+1})}.$$

Exercise 3. Redo Exercise 2 using the above method.

The polynomials obtained by the two procedures outlined above (Exercises 2 and 3) have very different forms, but check that they are equal. These are respectively called the *Newton form* and the *Lagrange form* of the *interpolating polynomial*.

References

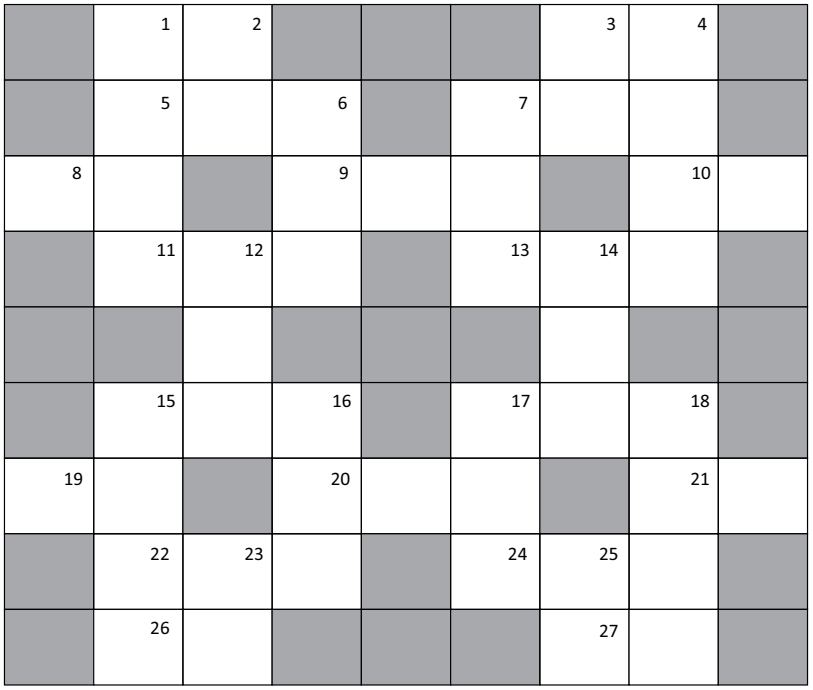
1. Newton Polynomial, https://en.wikipedia.org/wiki/Newton_polynomial
2. Lagrange Polynomial, https://en.wikipedia.org/wiki/Lagrange_polynomial
3. Vandermonde matrix, https://en.wikipedia.org/wiki/Vandermonde_matrix

SANKARAN VISWANATH is a faculty member at the Institute of Mathematical Sciences (IMSc), Chennai. He received his PhD in Mathematics from UC Berkeley (2004). His research interests lie in Algebraic Combinatorics and Lie Theory. He has been actively associated with mathematics outreach programmes at IMSc, and especially enjoys conveying the excitement of mathematics to school students and teachers. He may be contacted at svis@imsc.res.in.

NUMBER CROSSWORD

Solution on Page 29

D.D. Karopady & Sneha Titus



CLUES ACROSS

1 A prime number
3 2 more than 3 dozen
5 Product of the fourth power of 2 and 7
7 Product of a power of 2 and its reverse
8 Square root of 441
9 Perfect cube
10 A fortnight
11 Product of the 29th prime and 7
13 Multiple of 9
15 LCM of 3A and 10A
17 9 days short of a year
19 HCF of 2D and 19A is 13
20 14D divided by 7 x 37
21 2 score
22 Twin prime with 59 times 3
24 10 times 27 A
26 3.5 feet
27 20A divided by 25 and then digits reversed

CLUES DOWN

1 A hundred years hence
2 HCF of 2D and 19A is 13
6 Its negative signifies absolute zero
7 1 short of a millenium
12 Number remains a palidrome when divided by 2, 3 and 6
14 Quarter less than two centuries
15 Hundreds digit is the sum of the tens and units digit
17 Difference between middle digit and the end digits is the same