WHAT IS THE NEXT NUMBER IN THIS SEQUENCE?

PEE CHOON TOH, ENG GUAN TAY

Abstract. Given a sequence of numbers, or a sequence, in short, that is generated by an unknown polynomial or by a first order recurrence relation, we describe an easy way to write down the polynomial or the recurrence relation by computing successive differences of the terms of the sequence.

What is the next number in each of the following sequences?

$$S1 : 1, 3, 8, 19, 42, 89, \ldots$$
$$S2 : 23, 48, 84, 133, \ldots$$
$$S3 : 7, 27, 58, 102, \ldots.$$

One may immediately protest that it is impossible to uniquely determine a sequence based only on its first few terms. Richard Guy calls this the Strong Law Of Small Numbers [4]. Furthermore, it is well known that, given any n numbers, one can construct infinitely many polynomials $f(x)$ of degree n that evaluate to these n numbers when x takes on the value of the integers from 0 to $n-1$. As a result, Linderholm [5, p. 97] even suggests (with tongue firmly in cheek) that one should answer '19' to every "find the next term" question.

Yet such questions continue to appear in IQ tests[1] and, more importantly, also in the school syllabus for the purpose of teaching number patterns and algebra. We must acknowledge however that such number pattern questions have been useful as training in recognising patterns and appreciation of mathematics. What is needed in such texts is an insertion somewhere that such questions as they stand are 'nonsense' if there is no assumption about the sequence. One should also add that a sequence is only defined once *all* the terms are given, either explicitly (which in this case is no fun because then there would be no problem) or in some other way (for example, stating that the numbers in the sequence are the consecutive digits of the decimal representation of π).

Date: 26 July 2013.

[1]The sequences $S1$ to $S3$ are, respectively, puzzle 43 (from Test 1), 15 (Test 6) and 17 (Test 8), taken from [6], a book of IQ puzzles written by two UK Mensa puzzle editors.

One way to avoid ambiguity in such number pattern problems is to restrict ourselves to only sequences generated by polynomials of degree k, and when posing the question, to give at least $k+1$ terms. With that assumption, the standard technique to make sense of a complicated number pattern is the method of calculating successive differences. Take the sequence $S2$,

$$\begin{array}{lccccccccc} S2 & : & 23 & & 48 & & 84 & & 133 \\ \text{1st successive difference} & : & & 25 & & 36 & & 49 & \\ \text{2nd successive difference} & : & & & 11 & & 13 & & \\ \text{3rd successive difference} & : & & & & 2 & & & \end{array},$$

where the numbers in each new row is the difference of the two corresponding numbers of the previous row. Formally, if $a(n)$ denotes a sequence, then $a_i(n)$, the i-th successive difference of $a(n)$, is defined recursively by setting $a_0(n) = a(n)$ and

$$a_i(n) = a_{i-1}(n+1) - a_{i-1}(n) \text{ for integers } n \geq 0, i \geq 1,$$

Most people would be able to recognize that the first successive difference of $S2$ are square numbers starting from 5^2, and so the fifth term in $S2$ should be $133+8^2 = 197$, which is in agreement with the solution provided in [6].

$S3$ might provide more of a challenge,

$$\begin{array}{lccccccccc} S3 & : & 7 & & 27 & & 58 & & 102 \\ a_1(n) & : & & 20 & & 31 & & 44 & \\ a_2(n) & : & & & 11 & & 13 & & \\ a_3(n) & : & & & & 2 & & & \end{array},$$

since the patterns that emerge are slightly harder to decipher. [6] gives the next term as 161 with the explanation "the difference between the numbers are consecutive squares less 5". So, intrinsically, $S2$ and $S3$ are similar and one telling sign is that they share the same second and third successive differences. In fact both $S2$ and $S3$ are generated by polynomials of degree 3.

It is fairly well known that the k-th successive difference of a sequence generated by a polynomial of degree k would be a constant sequence. So if we are faced with a number pattern generated by a polynomial, we can simply compute successive differences until we arrive at a constant sequence which informs us of the value of k. After that, one can recover the $k+1$ coefficients of the unknown polynomial with some (possibly messy) algebra. What is perhaps not so well known is that there is a very easy way to write down the polynomial in question, if we allow ourselves some fancier notation.

Theorem 1. *If $f(x)$ is a polynomial of degree k and the sequence $a(n)$ is defined to be the value of $f(n)$, then the k-th successive difference $a_k(n)$ is a constant for all $n \geq 0$. Furthermore,*

$$f(n) = \sum_{i=0}^{k} a_i(0)\binom{n}{i},$$

where $a_i(0)$ is the first term of the i-th successive difference of $a(n)$.

We remark that $\binom{n}{i}$ is the binomial coefficients defined by

$$\binom{n}{i} = \frac{n!}{i!(n-i)!}$$

where we assume $\binom{n}{i} = 0$ if $i > n$. Furthermore $\binom{n}{i}$ can be viewed as a polynomial of degree i in n.

This theorem can be found in [7, p. 36-38], [1, p. 279] and also in standard numerical analysis textbooks disguised as a special case of Newton's divided difference formula. (See for example [2, p. 123].) However, the proof requires nothing more than the principle of mathematical induction and the following combinatorial identity [3, p. 79]:

$$\sum_{j=r}^{n} \binom{j}{r} = \binom{n+1}{r+1}.$$

We illustrate the theorem by writing down the generating polynomials for $S2$ and $S3$,

$$23\binom{n}{0} + 25\binom{n}{1} + 11\binom{n}{2} + 2\binom{n}{3} = \frac{2n^3 + 27n^2 + 121n + 138}{6}$$

and

$$7\binom{n}{0} + 20\binom{n}{1} + 11\binom{n}{2} + 2\binom{n}{3} = \frac{2n^3 + 27n^2 + 91n + 42}{6}.$$

Notationally, there is nothing superior about our usual way of writing polynomials that appear on the right side of the two equations above. In fact, from these expressions, it is not easy to see that these polynomials are always integer-valued when n is an integer.

We now prove Theorem 1 by induction on k, the degree of $f(x)$.

Proof. If $k = 1$, i.e. $f(n) = an + b$, then it is an arithmetic progression and the first successive difference would be the constant sequence consisting of all a, and

$$a_0(0)\binom{n}{0} + a_1(0)\binom{n}{1} = b + an = f(n).$$

Assume the theorem holds for some $k \geq 1$ and define

$$h(x) = f(x+1) - f(x).$$

Then

$$\begin{aligned}\sum_{j=0}^{n-1} h(j) &= \sum_{j=0}^{n-1} f(j+1) - f(j) \\ &= f(n) - f(0).\end{aligned}$$

On the other hand, by its definition, $h(x)$ is a polynomial of degree $k-1$. From the induction hypothesis, we have

$$h(n) = \sum_{\ell=0}^{k-1} b_\ell(0) \binom{n}{\ell}$$

where $b_\ell(0)$ is the first term of the ℓ-th successive difference of $h(n)$. This means that the $(k-1)$-th successive differences of $h(n)$ are constant, so the k-th successive differences of $f(n)$ are also constant. We also have $b_\ell(0) = a_{\ell+1}(0)$, the first term of the $(\ell+1)$-th successive difference of $f(n)$. Thus,

$$\begin{aligned}f(n) &= f(0) + \sum_{j=0}^{n-1} h(j) \\ &= f(0) + \sum_{j=0}^{n-1}\sum_{\ell=0}^{k-1} b_\ell(0)\binom{j}{\ell} \\ &= f(0) + \sum_{j=0}^{n-1}\sum_{\ell=0}^{k-1} a_{\ell+1}(0)\binom{j}{\ell} \\ &= f(0) + \sum_{\ell=0}^{k-1} a_{\ell+1}(0) \sum_{j=\ell}^{n-1}\binom{j}{\ell} \\ &= f(0) + \sum_{\ell=0}^{k-1} a_{\ell+1}(0)\binom{n-1+1}{\ell+1} \\ &= a_0(0)\binom{n}{0} + \sum_{i=1}^{k} a_i(0)\binom{n}{i},\end{aligned}$$

which proves the inductive step. □

What about the sequence $S1$? It turns out that this sequence is not generated by a polynomial, so Theorem 1 does not apply. But let us take a look at the successive differences anyway:

$$\begin{array}{llccccccccccc}
S1 & : & 1 & & 3 & & 8 & & 19 & & 42 & & 89 \\
a_1(n) & : & & 2 & & 5 & & 11 & & 23 & & 47 & \\
a_2(n) & : & & & 3 & & 6 & & 12 & & 24 & & \\
a_3(n) & : & & & & 3 & & 6 & & 12 & & & \\
a_4(n) & : & & & & & 3 & & 6 & & & & \\
a_5(n) & : & & & & & & 3 & & & & &
\end{array}.$$

Note that $a_2(n)$ and $a_3(n)$, the second and third successive differences *appear* to be identical. Furthermore, the successive ratios of $a_2(n)$ appear to be a constant, i.e. 6 is twice of 3, 12 is twice of 6 and 24 is twice of 12. There is some 'pattern' for further exploration.

Indeed, we now consider sequences $a(n)$ that are generated by a linear first order recurrence relation of the form

$$a(n) = ma(n-1) + f(n), \tag{1}$$

where m is a constant and $f(n)$ is a polynomial in n of degree at most k. If an initial value for $a(0)$ is given, such a relation can be shown [3, Ch. 6] to possess the unique solution $a(n) = p \cdot m^n + r(n)$ for some constant p and some polynomial $r(n)$. This gives rise to the following generalization.

Theorem 2. *Let $g(x) = p \cdot m^x$ and $f(x)$ be a polynomial of degree k. If we define the sequence $a(n)$ as the value of $g(n) + f(n)$, then the i-th successive difference of $a_i(n)$ is given by $p(m-1)^i m^n$ for all $i \geq k+1$ and $n \geq 0$. Furthermore,*

$$a(n) = \sum_{i=0}^{n} a_i(0) \binom{n}{i}.$$

Proof. We first note that if $m = 0$, then

$$a(n) = 0 + f(n) = \sum_{i=0}^{k} a_i(0) \binom{n}{i} = \sum_{i=0}^{n} a_i(0) \binom{n}{i},$$

since $a_i(0) = 0$ for all $i \geq k+1$.

Now let $g_i(n)$ and $f_i(n)$ be the i-th successive differences of the sequences $g(n)$ and $f(n)$ respectively. It is easy to show by induction that $g_i(n) = p(m-1)^i m^n$.

Since the operation of computing successive differences is linear, we have

$$a_i(n) = g_i(n) + f_i(n) = p(m-1)^i m^n,$$

for all $i \geq k+1$. Now

$$\begin{aligned}
\sum_{i=0}^{n} a_i(0)\binom{n}{i} &= \sum_{i=0}^{n} g_i(0)\binom{n}{i} + \sum_{i=0}^{k} f_i(0)\binom{n}{i} \\
&= \sum_{i=0}^{n} p(m-1)^i m^0 \binom{n}{i} + f(n) \quad \text{(from Theorem 1)} \\
&= p\sum_{i=0}^{n} (m-1)^i \binom{n}{i} + f(n) \\
&= p((m-1)+1)^n + f(n) \\
&= g(n) + f(n) \\
&= a(n). \qquad \square
\end{aligned}$$

Suppose we are told that the sequence $S1$ is generated by a recurrence relation of the form (1) where the degree of the polynomial $f(x)$ is not larger than 1. We can then assume that $a_i(0) = 3$ for all $i \geq 2$, so

$$\begin{aligned}
a(n) &= 1 + 2n + \sum_{i=2}^{n} 3\binom{n}{i} \\
&= -2 - n + 3\sum_{i=0}^{n} \binom{n}{i} \\
&= 3 \cdot 2^n - n - 2.
\end{aligned}$$

To answer our original question, the next term of the sequence is $a(6) = 3 \cdot 2^6 - 6 - 2 = 184$. We can also recover the recurrence relation by writing $a_i(0)$ as $3 \cdot (2-1)^i$ for all $i \geq 2$. Thus $m = 2$ and

$$\begin{aligned}
f(n) &= a(n) - 2a(n-1) \\
&= 3 \cdot 2^n - n - 2 - 2(3 \cdot 2^{n-1} - (n-1) - 2) \\
&= n.
\end{aligned}$$

In other words,

$$a(n) = 2a(n-1) + n,$$

agreeing with [6] which explained that "each of the numbers is doubled and 1, 2, 3, 4, 5, 6 is added in turn."

Theorem 2 now allows us to handle all sequences generated by recurrence relations of the form (1), as long as we are given at least $k+3$ terms of the sequence. The two extra terms are necessary for us to compute the constant m. For example, suppose that

$$\begin{array}{lcccccccccc} S4 & : & 1 & & 7 & & 19 & & 45 & & 109 \\ a_1(n) & : & & 6 & & 12 & & 26 & & 64 & \\ a_2(n) & : & & & 6 & & 14 & & 38 & & \\ a_3(n) & : & & & & 8 & & 24 & & & \\ a_4(n) & : & & & & & 16 & & & & \end{array}.$$

Working on the assumption that if we are given 5 terms, k must not exceed 2, we have $a_4(0) = p(m-1)^4$ and $a_3(0) = p(m-1)^3$. Thus $m - 1 = \frac{16}{8} = 2$ which implies $p = 1$ and $a_i(0) = 2^i$ for all $i \geq 3$.

$$\begin{aligned} a(n) &= 1 + 6n + 6\binom{n}{2} + \sum_{i=3}^{n} 2^i \binom{n}{i} \\ &= 4n + 2\binom{n}{2} + \sum_{i=0}^{n} 2^i \binom{n}{i} \\ &= 3^n + n^2 + 3n. \end{aligned}$$

A straightforward computation shows that the recurrence relation is

$$a(n) = 3a(n-1) - 2n^2 + 6.$$

Let us end with the following well known combinatorial problem: *Given a circle with $n+1$ arbitrary points on its circumference, if every possible pair of points is joined by a straight line, what is the maximal number of regions that the circle can be partitioned into?*

Working out the first 5 terms of this sequence and observing that they are respectively $1, 2, 4, 8$ and 16, one is tempted to jump to the conclusion that this sequence is generated by 2^n. We certainly should not do this. To be sure of the generating function, we need the additional knowledge that this sequence is generated by a polynomial of degree 4. With this fact in hand, plus the first five terms, we have

$$\begin{array}{lcccccccccc} a(n) & : & 1 & & 2 & & 4 & & 8 & & 16 \\ a_1(n) & : & & 1 & & 2 & & 4 & & 8 & \\ a_2(n) & : & & & 1 & & 2 & & 4 & & \\ a_3(n) & : & & & & 1 & & 2 & & & \\ a_4(n) & : & & & & & 1 & & & & \end{array}.$$

Theorem 1 then tells us that the generating function is

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \binom{n}{3} + \binom{n}{4}.$$

A geometric explanation for this sequence can be found in [4, Ex. 5].

Further reading

Ian Stewart [8] weaves an interesting tale about sequences that look like the Fibonacci numbers. Etta Mae Whitton [9] describes how one may extend a known sequence by adding a next term of any arbitrary value.

Bio

Pee Choon Toh is currently an Assistant Professor at the National Institute of Education, Nanyang Technological University. His main research interest is number theory, specifically the theory of partitions, modular forms and elliptic functions. He is also interested in problem solving and the teaching of mathematics at the undergraduate level.

Eng Guan Tay is an Associate Professor in the Mathematics and Mathematics Education Academic Group of the National Institute of Education at Nanyang Technological University, Singapore. Dr Tay specializes in graph theory and mathematics education and has had papers published in international scientific journals in both areas.

References

[1] R. A. Brualdi, *Introductory Combinatorics*, fifth edition, Prentice Hall, 2009.
[2] R. L. Burden and J. D. Faires, *Numerical Analysis*, eight edition, Thomson Brooks/Cole, 2005.
[3] C-C. Chen and K.-M. Koh, *Principles and Techniques in Combinatorics*, World Scientific, Singapore, 1992.
[4] R. K. Guy, *The Strong Law of Small Numbers*, Amer. Math. Monthly, **95** (1998) 697–712.
[5] C. E. Linderholm, *Mathematics Made Difficult,* World Publishing, New York, 1972.
[6] K. Russell and P. Carter, *Check Your IQ*, Foulsham, Berkshire, 1992.
[7] R. P. Stanley, *Enumerative Combinatorics*, Volume I, Cambridge University Press, Cambridge, 1997.
[8] I. Stewart, *Mathematical Recreations: Fibonacci Forgeries*, Scientific American, **272** (1995) 82–85.
[9] E. M. Whitton, *A General Formula for the Nth Term of a Sequence*, Two-Year College Math. J., **2** (1971) 96–98.

Mathematics & Mathematics Education, National Institute of Education, Nanyang Technological University, 1 Nanyang Walk, Singapore 637616

E-mail address: peechoon.toh@nie.edu.sg
E-mail address: engguan.tay@nie.edu.sg