MATHEMATICAL SPECTRUM

A MAGAZINE FOR STUDENTS AT SCHOOLS COLLEGES AND UNIVERSITIES

Volume 15 1982/83 Number 1

Mathematical Spectrum is a magazine for the instruction and entertainment of student mathematicians in schools, colleges and universities, as well as the general reader interested in mathematics. It is published by the Applied Probability Trust, a non-profit making organisation established in 1963 with the support of the London Mathematical Society. The object of the Trust is the encouragement of study and research in the mathematical sciences.

Volume 15 of *Mathematical Spectrum* will consist of three issues, of which this is the first. The second will be published in January 1983 and the third in May 1983.

Articles published in *Mathematical Spectrum* deal with the entire range of mathematical disciplines (pure mathematics, applied mathematics, statistics, operational research, computing science, numerical analysis, biomathematics). Both expository and historical material may be included, as well as elementary research and information on educational opportunities and careers in mathematics. There is also a section devoted to problems.

Articles are normally commissioned by the Editors; the Editorial Committee also welcomes the submission of suitable material, including correspondence, queries and solutions to problems, for publication in *Mathematical Spectrum*. All correspondence about the contents should be sent to:

The Editor, Mathematical Spectrum,
Hicks Building, The University, Sheffield S3 7RH.

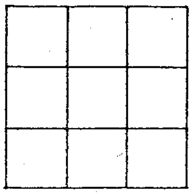
A Survey of Mathematical Puzzles II

KEITH AUSTIN, *University of Sheffield*

Keith Austin is a Lecturer in Pure Mathematics at the University of Sheffield. He has presented a monthly collection of mathematical puzzles on Sheffield's independent radio station and writes Brain-teasers for the *Sunday Times*. He is interested in the question of whether puzzles can be used to improve mental fitness in the same way that exercises are used to improve physical fitness.

In this second article in the series† we consider various types of finite mathematical puzzles. In this kind of puzzle we start with a finite list of possibilities and we have to find which one satisfies all the conditions of the puzzle.

Example. A magic square. Put the numbers $1, 2, 3, \ldots, 8, 9$ into the 3×3 square



so that the sum in each row, column and diagonal is the same. Also 9 must occur in the top row and 7 in the left-hand end column.

Now it is clear that the nine numbers can be put into the nine squares in only a finite number of ways. In fact, the number of ways is $9! = 9 \times 8 \times 7 \times \cdots \times 3 \times 2 \times 1$, but we do not usually have to worry about the actual number, just that it is finite. Thus one way of answering the puzzle is to write out all the possibilities and check each one in turn to see if it satisfies all the conditions given in the puzzle. This is sometimes a suitable way of solving the puzzle, but often the number of possibilities is too large and another way has to be found.

It is usual with these puzzles for there to be precisely one solution. If this is the case then you can use trial and error and you may hit on the solution by good luck. In other words, you can put the numbers into the square at random and check if they satisfy the conditions; if they do, then you have the solution. However, these puzzles are usually solved by a method which combines trial and error with a systematic search through the list. We often look for ideas which eliminate large numbers of the possibilities from the list at a stroke.

If you are constructing a finite puzzle then you have to check there is a solution and that there are no other solutions. You often start with your own solution and then you have to think up conditions which your solution satisfies but which together rule out all other possibilities.

There are many different types of finite puzzles, and we shall consider some of these. The answers to the puzzles will be given at the end of the article.

† The first part appeared in Volume 14 Number 2 of *Mathematical Spectrum*.

1. Cryptarithms

This title is one of many that are given to this type of puzzle; another title is alphametics. The puzzles combine arithmetic and coding.

Example.

$$\begin{array}{r} \mathrm{H\,O\,W} \\ \mathrm{I\,S} \\ +\mathrm{H\,I\,S} \\ \hline \mathrm{S\,I\,S} \\ \hline \end{array} \quad \text{(Note the 'O' is a letter.)}$$

Each letter stands for a different digit and SOW is a perfect square. A standard convention is that the left-hand end letter of each word does not stand for zero.

A popular puzzle is to take someone's name or a well-known saying and make it into an addition or subtraction sum. It is often necessary to add an extra condition in order to make the answer unique. A variation on this type of puzzle is to take a long multiplication or division and rub out most of the numbers but indicate where they were.

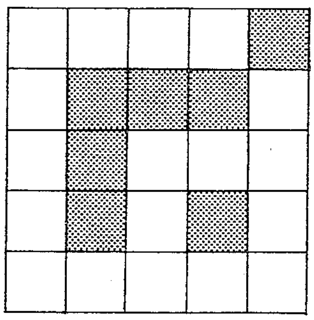
Example.

$$\begin{array}{r} * \; * \\ \times \quad 2 \; * \\ \hline * \; * \\ * \; 2 \;\;\, \\ \hline * \; * \; * \; * \\ \hline \end{array}$$

2. Fill-ins

Again there are alternative names such as jig-words.

Example.



Insert the following words into the diagram:

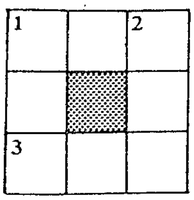
SHOTS, STEMS, SHOP, HOPS, FRO, FEE.

An alternative is to use numbers instead of words.

3. Crossnumbers

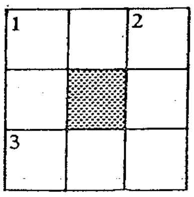
These are also known as crossfigures. They are similar to crosswords but numbers have to be inserted instead of words. As crosswords fall into two kinds, so do crossnumbers.

Example of a direct crossnumber.



Across	*Down*
1. 11^2	1. 5^3
3. 16^{th} prime $\times$ 10.	2. π radians $=$ — degrees.

Example of an indirect crossnumber.



Across	*Down*
1. 1 Down $-$ 4	1. a^b
3. $100a + 10b$	2. 1 Down $+$ 3 Across/10 $+$ 1

where a, b are positive integers less than 10.

There is a particular type of indirect crossnumber which is occasionally published in *The Listener*. One particular relationship between numbers is considered and various sets of the numbers which fit in the crossnumber and satisfy the relationship are given as the clues.

An example of this type of puzzle is given in the problems section.

4. Matching puzzles

These puzzles are not generally known by this name, and in fact they do not have a common name, although they often occur under the vague heading of 'puzzles in thought and logic'. They are very popular, probably because they offer considerable scope for varying the storyline.

Example. Mark, Paul and Boris are married to Anne, Mary and Susan, not necessarily respectively. Each couple has a pet and the pets are a cat, a guineapig and a pony. Use the following information to match up the husbands, wives and pets:

Paul's and Anne's pets were fighting.
Mary's husband's name has four letters.
Susan went round to feed Mark's pet when he was away.
Mark never goes to town.
Boris's pet is either the cat or the pony.
Susan's pet is not the cat.
The cat's male owner took him to the vet in town.
The guineapig hides when Anne visits its house.
Mary's pet sleeps in a shoe box.

In order to increase the variety of clues, one of the attributes is often numerical e.g. age or house number. Another variation is to introduce a second separate matching. For example, in the above example, let the pets also have names Mark, Paul and Boris. Then we have the new matching between each man and the pet with his name. This allows clues such as: Susan is married to the man whose name is the same as Mary's pet.

5. Logic puzzles

In these puzzles, we have a number of statements, and we have to decide for each one whether it is true or false from the information given.

Example. After a recent robbery, the police questioned five suspects. Each suspect made two statements of which one was true and one was false.

A said: B is guilty, C is innocent.
B said: A is guilty, D is innocent.
C said: A is guilty, B is guilty.
D said: C is guilty, E is innocent.
E said: We are all guilty, E is innocent.

Which of the suspects were guilty?

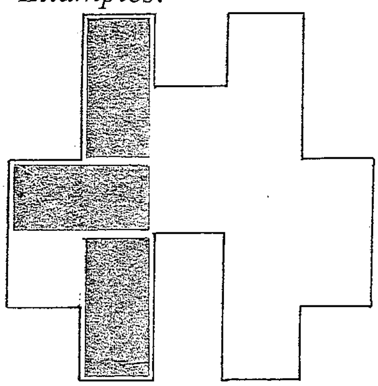
There are 32 possible answers for the above example, namely all combinations of true and false for each of the 5 statements, A is guilty, B is guilty, . . ., E is guilty. From this list of 32 possibilities you have to find the one which fits all the clues.

6. Moving puzzles

These include sliding blocks, pouring liquids, crossing rivers, stacking discs, moving chess pieces round the board and trains shunting trucks. Two problems that can be set are: to carry out the move and to do it in the smallest number of steps.

Examples.

1.

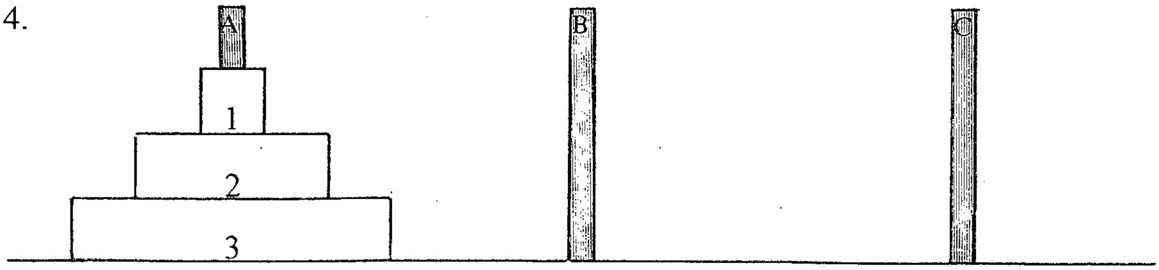


Slide the three blocks, without turning them, so that they occupy the right hand side of the region.

2. You have a large container of milk, a 5-pint jug and a 7-pint jug. By pouring back and forth how can you get exactly 6 pints of milk into the 7-pint jug?

3. A farmer has a fox, a goose and a cabbage and he wishes to take them across a river. There is a boat with room for the farmer and one of his possessions. He cannot leave the fox with the goose nor the goose with the cabbage. How does he get them across?

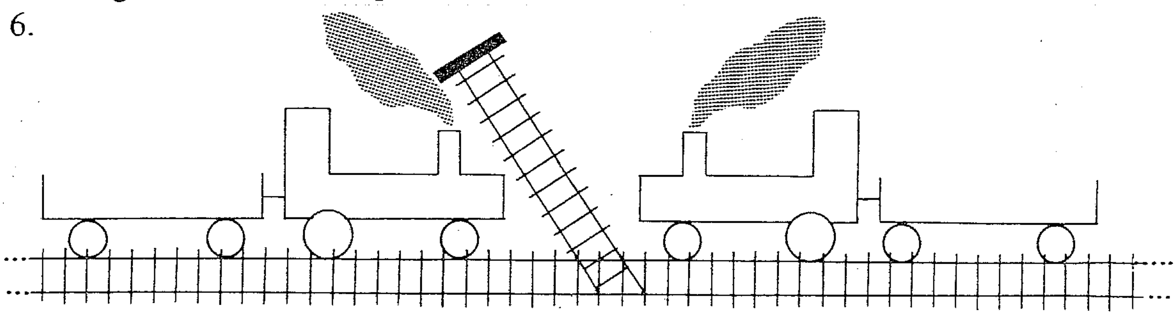
4.



There are 3 poles A, B, C. On A are stacked 3 discs 1, 2, 3. The aim is to get the 3 discs onto pole B. The discs may be stacked on A, B, and C. A move consists of taking the top disc from any pole and moving it to any other pole. However, no disc must ever be placed on a smaller disc.

5. Is it possible to place a knight on a 4×4 chess board and make 15 moves so that the knight visits all the squares on the board?

6.



The two trains wish to pass. There is room in the siding for one engine or one truck.

7. Puzzles involving arranging and counting

We shall give a typical example of each of these types of puzzles.

Examples.

1. Nine girls go out each day for a walk. They go in three rows with three girls in each row. How can the girls be arranged for four days so that no two girls are in the same row on more than one day?
2. There are seven girls and each day some of them go for a walk. The same group of girls may not go for a walk on more than one day. For how many days can the girls go for a walk?

(The final article in this series will be published in a forthcoming issue of *Mathematical Spectrum*. Ed.)

ANSWERS TO THE PUZZLES

2	9	4
7	5	3
6	1	8

Note. The conditions on 9 and 7 were necessary to guarantee a unique solution.

1. Cryptarithms

$$\begin{array}{r} 461 \\ 39 \\ 439 \\ \hline 939 \\ \hline \end{array} \qquad \begin{array}{r} 46 \\ \times\ 22 \\ \hline 92 \\ 92 \\ \hline 1012 \\ \hline \end{array}$$

2. Fill-ins

S	H	O	P	
H				H
O		F	R	O
T		E		P
S	T	E	M	S

3. Crossnumbers

1	2	1
2		8
5	3	0

2	1	2
1		8
6	3	0

4. Matching puzzles

Mark and Mary have the guinea pig. Paul and Susan have the pony. Boris and Anne have the cat.

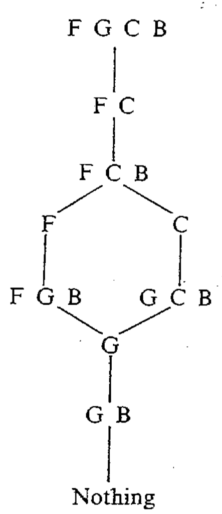
5. Logic puzzles

A and D were guilty.

6. Moving puzzles

1. See 3.
2. 5-pint : 0 5 0 2 2 5 0 4 4 5
   7-pint : 7 2 2 0 7 4 4 0 7 6

3. We shall use this example to illustrate a pictorial technique for solving moving puzzles. On a sheet of paper write out all the possible situations that can arise in the puzzle. Then draw a line from each situation to all the other situations we can move to directly from that situation. For puzzle 3 we get the following diagram, where each situation is denoted by those things on the starting bank. F = fox, G = goose, C = cabbage, B = boat and farmer.



It usually take some trial and error until we get a clear diagram. Such a diagram is called a *graph*. The problem is to find a path from the start to the finish. In this case the problem is more or less solved once we have the graph, as it is easy to spot the two solution paths.

Returning to puzzle 1, if we label the 3 blocks F, G, C, with the centre one G, then the block puzzle becomes almost the river puzzle and can be solved by one of the two solutions, except that there is no farmer and boat. Thus puzzles 1 and 3 are the same basic puzzle but dressed up in different words.

4.

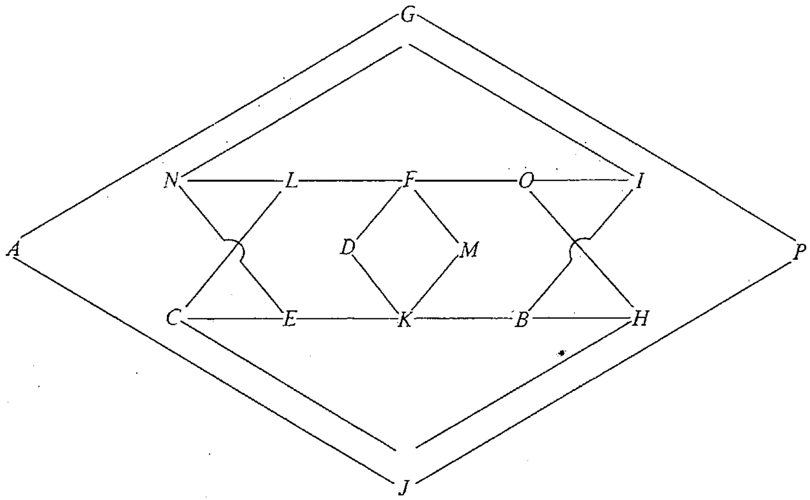
A	123	23	3	3		1	1	
B		1	1		3	3	23	123
C			2	12	12	2		

This puzzle is called the 'Tower of Hanoi'. The number of discs can be changed but we keep to 3 poles. The story goes that God placed 64 discs on pole A at the creation, and since then priests have been moving the discs one by one in order to transfer the whole pile to pole B. When the task is done the world will end. I will leave it to you to work out how long the priests will take.

5. No. We shall see why it is not possible.

A	*B*	*C*	*D*
E	*F*	*G*	*H*
I	*J*	*K*	*L*
M	*N*	*O*	*P*

We draw the graph of the puzzle. There are 16 situations, namely, the 16 squares where the knight can be.

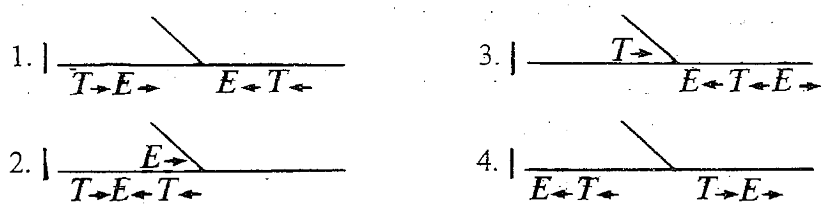


We have to trace a path through the graph visiting every letter exactly once. This is not possible. We can see this as follows.

We must start at *A* or *P*, complete the outer square and finish by completing the inner square and ending at *D* or *M*, or vice versa. However, in going from one square to the other we can only cover one end of the horizontal rectangle, *NLEC* or *OIHB*.

Note that the graph played a vital part in this solution, whereas puzzle 3 could be solved without it. However, it was used in puzzle 3 to illustrate the technique.

6.



7. Puzzles involving arranging and counting

Example 1.

	Day 1	*Day 2*	*Day 3*	*Day 4*
Row 1	1, 2, 3	1, 4, 7	1, 6, 8	1, 5, 9
Row 2	4, 5, 6	2, 5, 8	2, 4, 9	2, 6, 7
Row 3	7, 8, 9	3, 6, 9	3, 5, 7	3, 4, 8

A similar puzzle involving 15 girls and 7 days was proposed in 1847 by the Rev. T. P. Kirkman. (See 'The schoolgirls problem' by Norman L. Biggs, in Volume 12 Number 1 of *Mathematical Spectrum*.)

Example 2. $2^7 - 1$ days.

Roots of Polynomials

DAVID SHARPE, *University of Sheffield*

The author is a Lecturer in Pure Mathematics at the University of Sheffield, and is editor of *Mathematical Spectrum*.

1. A problem formulated

In Volume 14 Number 2 of *Mathematical Spectrum*, a letter appeared from Nigel McCann, a sixth-former at Immingham School, South Humberside. No doubt his sixth-form mathematics set had been dealing with roots of polynomials, and Nigel had been thinking about them on his own. First the standard fact. Suppose we have a polynomial

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

of positive degree n, so that $a_n \neq 0$, with real coefficients. It is a consequence of the fundamental theorem of algebra that this polynomial has n complex roots (where it is necessary to count each root an appropriate number of times). If we denote the roots of $f(x)$ by $\alpha_1, \ldots, \alpha_n$, we can then write

$$f(x) = a_n(x-\alpha_1)(x-\alpha_2)\cdots(x-\alpha_n).$$

Equating the coefficients of x^{n-1} in the two expressions for $f(x)$, we obtain

$$-a_n(\alpha_1+\alpha_2+\cdots+\alpha_n)=a_{n-1},$$

or

$$\alpha_1+\alpha_2+\cdots+\alpha_n=-\frac{a_{n-1}}{a_n},$$

a fact which is, or should be, well known to all advanced-level students, at least for a quadratic.

But Nigel noticed the following.

When $f(x)$ is differentiated $n-1$ times with respect to x, then

$$\frac{d^{n-1}f(x)}{dx^{n-1}}=a_n n!x+a_{n-1}(n-1)!,$$

and this polynomial has the one root $-(a_{n-1}/na_n)$. Putting these together, we see that *the root of the* $(n-1)$*th derivative of* $f(x)$ *is*

$$\frac{\alpha_1+\alpha_2+\cdots+\alpha_n}{n},$$

the arithmetic mean of the roots of $f(x)$.

Quite rightly, this intrigued Nigel—hence his letter to *Mathematical Spectrum*. What it amounts to is this: if we mark the n complex numbers as points in the Argand diagram, then the root of the $(n-1)$th derivative of $f(x)$ lies at the centroid of the n roots of f. By a similar token, it will also lie at the centroid of the roots of $f'(x)$, so that the roots of $f'(x)$ have the same centroid as the roots of $f(x)$. Obviously now the centroid of the roots of $f''(x)$ will be the same point, and so on for all higher derivatives of $f(x)$ up to the $(n-1)$th.

Perhaps an example will help to clarify the situation. Let

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

Thus $f(x)$ has three roots 1, i, $-i$, which we have marked in the Argand diagram (Figure 1).

Now $f'(x)=3x^2-2x+1$, and $f'(x)$ has two roots, namely $(1\pm i\sqrt{2})/3$. These occur at points A, B of Figure 1. Lastly, $f''(x)=6x-2$ and $f''(x)$ has the one root $\frac{1}{3}$, which lies at the point C. This point is the midpoint of the line segment AB (and so the centroid of the points A, B), and also the centroid of of the points 1, i, $-i$, the roots of $f(x)$.

Nigel's observation prompts us to ask whether anything else can be said about the position of the roots of $f'(x)$ in the Argand diagram compared with the position of the roots of $f(x)$. From the example, we notice that the roots of $f'(x)$ lie inside the

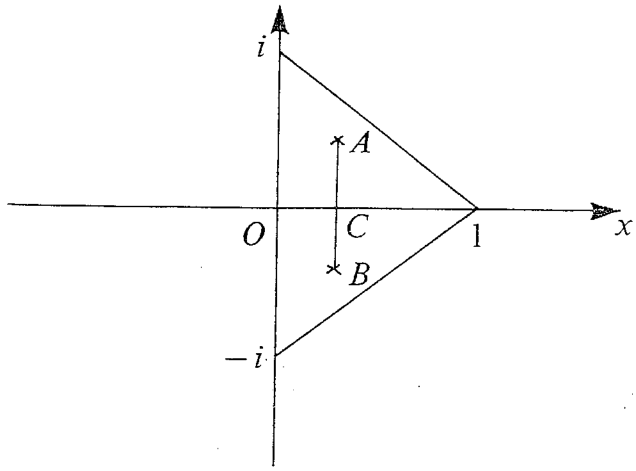


Figure 1

triangle formed by the roots of $f(x)$ and that the one root of $f''(x)$ lies on the line segment joining the roots of $f'(x)$.

Before we attempt an answer to this question, we introduce some simple mathematical machinery.

2. Mathematical machinery

A set S of points in the plane is said to be *convex* if, whenever A, B are points of S, then the whole straight-line segment AB lies in S (see Figure 2). Examples of convex sets are circular or elliptical discs, squares or rectangles. The *convex hull* of a set of points in the plane is the smallest convex set to contain that set. Intuitively, to form the convex hull of a set amounts to 'filling in its indentations' (see Figure 3). For example, consider a set consisting of two distinct points A, B. The convex hull of the two-point set $\{A, B\}$ is the line segment AB. The convex hull of the three-point set $\{A, B, C\}$ is the triangle ABC, its interior and its boundary. The convex hull of the four-point set A, B, C, D is the quadrilateral $ABCD$ provided that

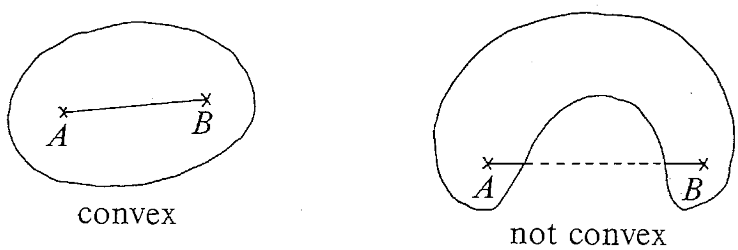


Figure 2

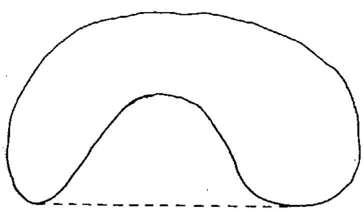
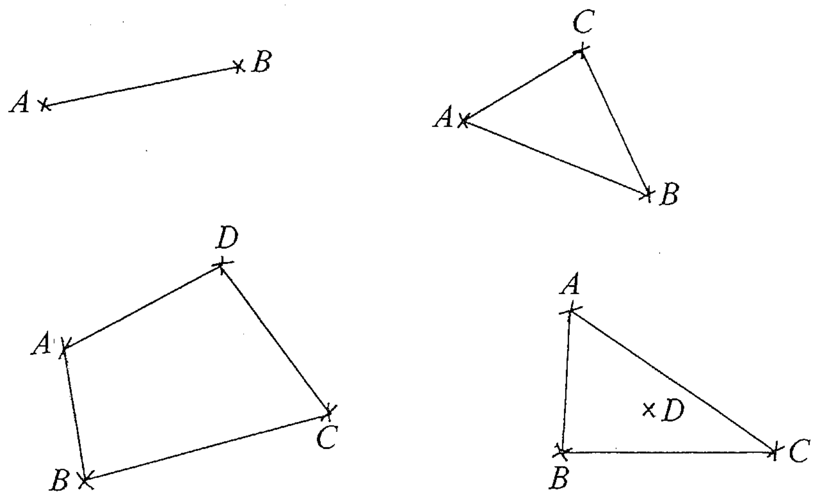


Figure 3



Convex hulls of finite sets of points

Figure 4

no point lies inside the triangle formed by the other three. If, for example, D lies inside the triangle ABC, then the convex hull of A, B, C, D is the triangle ABC. (See Figure 4.)

We need to describe algebraically the convex hull of a finite set of points. Consider two points A, B, with respective coordinates $\boldsymbol{a} = (x, y)$, $\boldsymbol{b} = (u, v)$. The coordinates of a typical point on the line segment AB are

$$\lambda \boldsymbol{a} + (1 - \lambda)\boldsymbol{b} = (\lambda x + (1 - \lambda)u, \lambda y + (1 - \lambda)v),$$

where λ is a real number and $0 \leq \lambda \leq 1$. For example, A is given by $\lambda = 1$, B is given by $\lambda = 0$, and the midpoint of AB is given by $\lambda = \frac{1}{2}$. Alternatively the points of the line segment AB, which is just the convex hull of the two-point set $\{A, B\}$, are those with coordinates

$$\lambda \boldsymbol{a} + \mu \boldsymbol{b}, \qquad \text{where } \lambda, \mu \geq 0 \text{ and } \lambda + \mu = 1.$$

If we now consider three points A, B, C, with coordinate vectors $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$, the points in the triangle ABC or on its boundary are given by

$$\lambda \boldsymbol{a} + \mu \boldsymbol{b} + \nu \boldsymbol{c}, \qquad \text{where } \lambda, \mu, \nu \geq 0 \text{ and } \lambda + \mu + \nu = 1.$$

This gives the convex hull of the three-point set $\{A, B, C\}$. Generally, if we have n points $A_1, A_2, \ldots, A_n$ with coordinate vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ respectively, then the convex hull of the set $\{A_1, A_2, \ldots, A_n\}$ consists of all points with coordinates

$$\lambda_1 \boldsymbol{a}_1 + \lambda_2 \boldsymbol{a}_2 + \cdots + \lambda_n \boldsymbol{a}_n,$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n \geq 0$ and $\lambda_1 + \lambda_2 + \cdots + \lambda_n = 1$.

We notice that the arithmetic mean of n complex numbers $\alpha_1, \ldots, \alpha_n$, namely

$$\frac{\alpha_1 + \cdots + \alpha_n}{n},$$

gives a point in the convex hull of $\alpha_1, \ldots, \alpha_n$ in the Argand diagram, because

$$\frac{1}{n}+\frac{1}{n}+\cdots+\frac{1}{n}=1,$$

where there are n terms in the sum.

3. The problem solved

We are now ready to solve the problem we formulated in Section 1. We have a polynomial $f(x)$ of positive degree n which we shall allow to have complex coefficients. We denote the complex roots of $f(x)$ by $\alpha_1, \ldots, \alpha_n$, which we can consider as points in the Argand diagram. The problem is: what can we say about the positions of the roots of the derivative $f'(x)$ compared with the positions of the roots of $f(x)$? The *Gauss–Lucas theorem* says that *the roots of $f'(x)$ all lie in the convex hull of the roots of $f(x)$*. Figure 5 illustrates this. The roots of $f(x)$ are denoted by α's and those of $f'(x)$ by β's, and the β's lie inside the polygon shown. We have already seen in Section 1 how this works in a particular example.

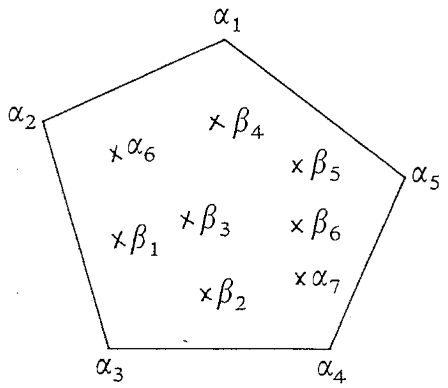


Figure 5

A consequence of this result has already made an appearance in *Mathematical Spectrum* as Problem 8.9. There, readers were asked to show that, if all the roots of a polynomial of positive degree have positive real parts, then the same is true of the roots of the derivative of the polynomial. You might like to think how this follows from the Gauss–Lucas theorem.

But now to prove the theorem. Consider any root α of $f'(x)$. We have to express α in the form

$$\alpha=\lambda_1\alpha_1+\lambda_2\alpha_2+\cdots+\lambda_n\alpha_n, \qquad \text{where } \lambda_1, \ldots, \lambda_n \geq 0 \text{ and } \lambda_1+\cdots+\lambda_n=1. \quad (*)$$

If α is one of the α_i's, this is not hard to do; after all,

$$\alpha_i=0\alpha_1+\cdots+1\alpha_i+\cdots+0\alpha_n.$$

We may thus suppose that $\alpha \neq \alpha_1, \ldots, \alpha_n$, i.e. α is not a root of $f(x)$. If we write

$$f(x)=a_n(x-\alpha_1)(x-\alpha_2)\cdots(x-\alpha_n),$$

then, using an extended form of the formula for the differentiation of a product, we have

$$
\begin{aligned}
f'(x) = a_n(x-\alpha_2)(x-\alpha_3)\cdots(x-\alpha_n) \\
+a_n(x-\alpha_1)(x-\alpha_3)\cdots(x-\alpha_n) \\
+-------------- \\
+a_n(x-\alpha_1)(x-\alpha_2)\cdots(x-\alpha_{n-1}).
\end{aligned}
$$

Now $f'(\alpha)=0$ but $f(\alpha)\neq 0$, so

$$0=\frac{f'(\alpha)}{f(\alpha)}=\frac{1}{\alpha-\alpha_1}+\frac{1}{\alpha-\alpha_2}+\cdots+\frac{1}{\alpha-\alpha_n}.$$

Experts on complex numbers will readily recall that, if $z=x+iy$ (x, y real), then $\bar{z}=x-iy$ and $z\bar{z}=x^2+y^2=|z|^2$. Thus, multiplying by conjugates on both top and bottom, we have

$$0=\frac{\overline{\alpha-\alpha_1}}{|\alpha-\alpha_1|^2}+\frac{\overline{\alpha-\alpha_2}}{|\alpha-\alpha_2|^2}+\cdots+\frac{\overline{\alpha-\alpha_n}}{|\alpha-\alpha_n|^2}.$$

If we take the conjugate of this, we have

$$0=\frac{\alpha-\alpha_1}{|\alpha-\alpha_1|^2}+\frac{\alpha-\alpha_2}{|\alpha-\alpha_2|^2}+\cdots+\frac{\alpha-\alpha_n}{|\alpha-\alpha_n|^2}.$$

We can now rearrange this expression to give

$$\left(\frac{1}{|\alpha-\alpha_1|^2}+\frac{1}{|\alpha-\alpha_2|^2}+\cdots+\frac{1}{|\alpha-\alpha_n|^2}\right)^{\alpha}=\frac{\alpha_1}{|\alpha-\alpha_1|^2}+\frac{\alpha_2}{|\alpha-\alpha_2|^2}+\cdots+\frac{\alpha_n}{|\alpha-\alpha_n|^2},$$

which leads to

$$\alpha=\frac{\dfrac{\alpha_1}{|\alpha-\alpha_1|^2}+\dfrac{\alpha_2}{|\alpha-\alpha_2|^2}+\cdots+\dfrac{\alpha_n}{|\alpha-\alpha_n|^2}}{\dfrac{1}{|\alpha-\alpha_1|^2}+\dfrac{1}{|\alpha-\alpha_2|^2}+\cdots+\dfrac{1}{|\alpha-\alpha_n|^2}}.$$

This is an expression of the form (*) (note that the coefficients of the α_i's are greater than or equal to 0 and add up to 1), which shows that α is in the convex hull of the set $\{\alpha_1, \ldots, \alpha_n\}$. Since α was a typical root of $f'(x)$, we have shown that the roots of $f'(x)$ lie in the convex hull of the roots of $f(x)$ in the Argand diagram. The proof of the Gauss–Lucas theorem is therefore complete.

4. A problem

A reader who has got this far might like to try to solve Problem 15.3 on p. 28. If you think you have been successful, send us your solution.

5. A question

We are all familiar with Gauss, but who was Lucas, and how do these two names come to be attached jointly to this beautiful result on roots of polynomials? Here our research has drawn a blank. Perhaps a reader can shed light on this.

Some Examples of Distribution-Free Statistical Tests

B. M. BROWN, *University of Tasmania*

Bruce Brown is a Senior Lecturer in the Department of Mathematics at the University of Tasmania, Hobart. His main research interests are in distribution-free statistical methods, the subject of the present article. Through his research, and his experience in statistical consulting, he is convinced that these methods deserve to be more widely known and used by practising statisticians.

1. Introduction

Statistical tests are very important in the theory and practice of statistical inference. Their most classical form is in the t- and F-tests of the analysis of variance, developed earlier this century by R. A. Fisher. These tests have formed the backbone of statistical methods for over half a century—details of them may be found in many elementary statistical text books—and their validity rests upon what are called 'standard assumptions' that observed data are equally accurate, and are distributed according to the normal probability distribution.

As the applications of statistics have extended into more and more diverse fields, it has been found that it is not always possible to obtain data conforming to these assumptions. in which case the application of classical statistical tests may lead to incorrect conclusions. In addition, many classical tests are rather susceptible to the presence of gross errors, or 'outliers' in data, as often occurs in practice.

Distribution-free tests were developed in response to these difficulties. The idea, roughly, is to replace observed numerical data by rankings, or orderings, or signs, and so on; so recasting the data into a form for which the probability distribution theory involves simple combinatorics, so that valid tests may be carried out rather simply. In the process, some information about the underlying situation under investigation may be lost, but this is often a small price to pay for the simplicity and validity of the resulting statistical tests.

In the remaining sections, two examples are presented of distribution-free tests in real-life situations where the classical tests are not strictly applicable. These examples are simple, and for their understanding need only elementary combinatorial probability theory, which means that the ideas behind distribution-free tests will be well within the grasp of many school students. It is hard to make a similar claim about the classical t- and F-tests, as a perusal of statistical textbooks will show.

It is probably necessary to state now the formal reasoning behind statistical tests; it must be admitted that though this reasoning, when broken into small steps, is fairly simple, it is common nevertheless for students to have some difficulty digesting all of it.

I. A vector X of random variables is to be observed; the probability distribution of X depends upon an unknown constant θ which is the subject of investigation.

II. A hypothesis about θ, called H_0, the *null-hypothesis*, is to be tested against an *alternative hypothesis* called H_1.

III. Some function of the observations, say $f(X)$, is found such that the probability distribution of $f(X)$ *when H_1 is true* is centred on higher numbers than is the case when H_0 is true. In other words, should H_1 be true we expect to observe larger values of $f(X)$ than if H_0 is true. The idea behind the test will then be to watch out for *large* values of $f(X)$.

IV. Observations are taken on X; suppose that $X = x$ is observed, so that $f(X)$ is observed to be $f(x)$. Then the question is asked: assuming H_0 true, how probable is it to obtain an outcome as extreme as the observed outcome, purely by chance?

In other words, before $X = x$ was observed, what would $\Pr\{f(X) \geq f(x)\}$ have been? Here, probability is calculated assuming H_0 true.

V. This probability is called the *significance level* of the observed $f(X) = f(x)$; if it is small, it means that the outcome observed could occur by chance only rarely, and if the significance level is too small, we may decide to reject H_0 in favour of H_1. In other words, the significance level might be too small for it to be believed that the observed outcome *did* occur purely by chance; in scientific work, it has become common to reject H_0 for significance levels $\leq 0{\cdot}05$; i.e. the level 1 in 20 is 'too small to believe'.

The steps I–V, or some equivalent sequence, are inherent in all statistical tests, and they may be identified in the following examples if one wishes to undergo a very thorough logical analysis. However, the examples will be presented with much less detail than this.

2. Student performance under repeated examinations

Many universities in Australia and other countries teach by the unit or component system, in which each unit is examined at the end of the term or semester in which it is taught. The nature of the learning process is much debated in educational circles, and there is a viewpoint that for many students, an examination at the end of the academic year would be more satisfactory, since the intervening time after the unit has been completed, plus end-of-the-year revision, can constitute a 'percolation' and consolidation period. The counterargument is that for other students, examination is better when the unit's material is freshest in the mind; i.e. immediately following the unit.

Because of this debate (and for other reasons), many universities offer alternative examinations at the end of the year, and the marks obtained by students taking both examinations allow investigation of the question as to whether students do in fact do better at a later examination. Here is some data for 13 students taking a third-year statistics course at La Trobe University in 1978.

Student	1	2	3	4	5	6	7	8	9	10	11	12	13
% Mark at term exam	64	55	58	70	69	38	43	40	73	82	36	44	80
% Mark at final exam	61	53	63	80	42	47	54	75	78	78	51	59	83

At first sight, it might appear that this data conforms to the standard assumptions which would enable the classical normal-theory test, in this case a 'paired t-test', to be carried out. But a closer inspection reveals that there may be some 'unreliable' marks in the above set, due possibly to a student being indisposed on the day of examination.

The improvements registered at the second examination are:

Student	1	2	3	4	5	6	7	8	9	10	11	12	13
Improvement	−3	−2	5	10	−27	9	11	35	5	4	15	15	3

Let the median improvement for all possible students (not just the 13 for whom data is available) be θ. To test $H_0 : \theta = 0$ (no median improvement) against $H_1 : \theta > 0$ (a positive median improvement), consider the signs of the above recorded improvements. If H_0 is true, these signs should be independent random variables, equalling + or − with probability $\frac{1}{2}$ each. Therefore, if S is the number of positive improvements, under H_0 S has a binomial probability distribution with parameters $n = 13$ and $p = \frac{1}{2}$. On the other hand, should H_1 be true, then S will tend to increase, so that the significance level of the observed $S = 10$ is (using the above binomial distribution)

$$\Pr(S \geq 10) = \sum_{x=10}^{13} \binom{13}{x} \left(\frac{1}{2}\right)^{13}$$

$$= \left(\frac{1}{2}\right)^{13} \left\{ \binom{13}{10} + \binom{13}{11} + \binom{13}{12} + \binom{13}{13} \right\}$$

$$= \left(\frac{1}{2}\right)^{13} \left\{ \frac{13.12.11}{1.2.3} + \frac{13.12}{1.2} + \frac{13}{1} + 1 \right\}$$

$$= \left(\frac{1}{2}\right)^{13} \{378\} = 0{\cdot}0461.$$

Since this is $< 0{\cdot}05$, the probability of as extreme an outcome as was observed occurring purely by chance is less than 1 in 20. Most statisticians would consider this sufficiently strong evidence against H_0: 'no median improvement' to reject it in favour of H_1: 'a positive median improvement'.

3. An example in hydrology

The next example concerns hydrology, the study of the flow response of streams in a catchment area to precipitation in the form of rain or snow. The construction of good mathematical models in hydrology is very difficult because of the complexities of the physical processes involved, and this is particularly so for Australian catchments because of the notorious and erratic occurrence of flood and drought.

In the following example, it is not a question of a classical test being invalid; it is simply the case that no way of analyzing the data is known apart from the simple distribution-free method (or a variation of it) to be presented here.

The data consists of total annual streamflow for the years 1958–1972 for two adjacent catchments A and B. The catchments have different areas, but since they adjoin, their rainfalls will be very similar. A is a *control* catchment; B is an experimental catchment which in 1966 was logged. Because the exact time during the year at which logging took place is unknown, year 1966 is omitted from the data record. The hypothesis to be tested is H_0: logging does not affect streamflow; against H_1: logging tends to increase streamflow, at least in the six years immediately after logging takes place.

Annual streamflows

Year	58	59	60	61	62	63	64	65	67	68	69	70	71	72
A	477	66	100	281	314	130	166	0	121	319	716	88	707	560
B	274	15	27	117	173	70	84	0	92	327	570	102	374	380
$B \div A$	0·52	0·22	0·27	0·42	0·55	0·54	0·51	—	0·76	1·03	0·71	1·16	0·53	0·68

Omitting year 1965, rank the ratios (flow in B/flow in A) in order of magnitude from 1 to 13.

Year	58	59	60	61	62	63	64	67	68	69	70	71	72
Rank	9	13	12	11	6	7	10	3	2	4	1	8	5

The streamflows for the two catchments fluctuate a good deal from year to year, due mainly to varying rainfall. However, since the catchments experience similar rainfalls, the streamflow fluctuations should be very similar for the two catchments. In particular, if the null hypothesis is H_0: 'the 1966 logging of catchment B does not affect streamflow', and if H_0 is true, then fluctuations in streamflow ratios are due to chance alone, and the ranks from 1 to 13 should be spread randomly over both the pre-1965 and the post-1966 periods. On the other hand, if H_1: 'logging tends to *increase* streamflow' is true, then the post-1966 rankings would tend to be smaller than the pre-1965 rankings.

A reasonable way to test H_0 against H_1 is therefore to look at S, the sum of the pre-1965 rankings. We observe $S = 9 + 13 + 12 + 11 + 6 + 7 + 10 = 68$.

The significance level of this observation is $\Pr\{S \geq 68\}$, where probability is calculated assuming all subsets of 7 elements drawn from $\{1, 2, \ldots, 13\}$ are equally likely to constitute the pre-1965 set of rankings.

There are altogether

$$\binom{13}{7} = \frac{13.12.11.10.9.8}{1.2.3.4.5.6} = 1716$$

such subsets; those subsets leading to the highest values of S are as follows.

S	Subset
70	13, 12, 11, 10, 9, 8, 7
69	13, 12, 11, 10, 9, 8, 6
68	13, 12, 11, 10, 9, 8, 5
68	12, 12, 11, 10, 9, 7, 6

The significance level is $\Pr\{S \geq 68\}$, $= 4/1716 = 0{\cdot}0023$, the smallness of which easily provides strong enough evidence against H_0, that logging does not affect streamflow, for it to be rejected in favour of H_1, that logging tends to increase streamflow.

The criterion S is called *Wilcoxon's sum of ranks statistic.*

4. Notes

The following points regarding the preceding two examples should be mentioned; these points were omitted earlier in order to simplify the presentation.

(i) In Section 2 (students' exam performances) it is vital that the 13 students represented are selected at random from all students. The whole basis of validity for the given test would be destroyed if students were allowed to select themselves for the second exam, leading to an over-representation of those students who were conscious of having under-achieved, for whatever reason, on the first exam.

The 'classical' test in this situation is a paired t-test, which turns out to yield a similar result to the given distribution-free test. The reason for preferring the distribution-free test was the presence of 'unreliable' improvements 35 and -27; had these two numbers been larger in magnitude, the t-test would have been affected, while the distribution-free test would remain unaffected.

(ii) In regard to the hydrology example in Section 3, it is interesting to note that any alternative, more conventional approach would have involved setting up a hydrology model for the catchment's behaviour. Such models tend to be complicated, and quite unreliable unless good data is available for rainfall, and also for such difficult quantities to measure as evapotranspiration, the amount of moisture lost from evaporation, and transpiration through foliage of plants. The approach outlined is surprisingly simple and effective in detecting a small though consistent trend towards increased streamflow after logging.

(iii) When larger amounts of data are present than in the examples discussed, the computation of probabilities can become very tedious. In such cases, it is often convenient to use a normal approximation. For instance, in Section 2, the binomial probability distribution with parameters n, p is approximately normal with mean np and variance $np(1-p)$, so the distribution of S is approximately $N(6{\cdot}5,\ 3{\cdot}25)$.

In Section 3, the distribution of the sum of ranks of m objects among a total of $m+n$ objects is approximately normal with mean $\frac{1}{2}m(m+n+1)$ and variance $mn(m+n+1)/12$, so the distribution of S there is approximately $N(49, 49)$.

Chaos and Strange Attractors

G. ROWLANDS, *University of Warwick*

The author is Chairman of the School of Applied Mathematics at the University of Warwick. His interest in the topics discussed in this article arose from his research activities in the field of controlled thermonuclear reactions ('taming the hydrogen bomb').

Introduction

Almost by tradition the mathematical modelling of physical situations has been in terms of differential equations. For example, the decay of radioactive nuclei can be described by the simple first-order differential equation

$$\frac{dN}{dt} = -\lambda N, \tag{1}$$

where N is the number of nuclei at time t and λ the decay constant. The amplitude of oscillations of a simple pressure wave propagating down a tube satisfies the second-order differential equation

$$\frac{d^2A}{dx^2} + (\omega^2/c^2)A = 0, \tag{2}$$

where A is the pressure variation, ω the frequency of the wave, c is the velocity of the wave and x measures the distance down the tube.

These same equations are used to model systems other than those that arise in physics. Equation (1) describes the simplest ecological model of population growth if N is now identified as the number in the population and λ is allowed to take negative values. The equation then leads to the celebrated exponential growth of populations associated with the name of Malthus.

Such models are appropriate when, to a sufficient approximation, the quantities N and A may be considered to be continuous functions of t and x respectively. But situations exist, for example in a population of species where subsequent generations do not overlap (that is the parents die before their offspring reach maturity), where it is more appropriate to consider the number in the population per generation. Then the simplest model describing such a population takes the form of a first-order *difference* equation

$$N_{m+1} = \lambda N_m. \tag{3}$$

Here N_m is the number in the population after m generations and λ a growth constant. Similarly, if one considers the motion of a sound wave in a solid, which one may consider as a periodic array of atoms, then the equation analogous to (2) is

$$x_{n+1} - 2x_n + x_{n-1} + \frac{a^2\omega^2}{c^2} x_n = 0, \tag{4}$$

where x_n is the displacement of the nth atom from its equilibrium position and a the distance between the atoms.

The difference equations introduced above have one feature in common, namely, they are *linear*. That is, each term in the equation involves only x_m or N_m, there are no non-linear terms such as N_m^2. The solutions of linear difference equations are very similar to those of their analogous differential equation. The general solution of (2) may be written in the form

$$A = B\cos\left(\frac{\omega x}{c} + \phi\right)$$

where B and ϕ are constants. Similarly the solution of equation (4) is

$$x_n = D\cos(\alpha n + \beta)$$

where D and β are constant and

$$\sin\alpha = \sqrt{(b^2 - b^4/4)},$$

where $b = a\omega/c$. Thus both equations have the same sort of solution, namely a periodic variation along the length of the tube or along the array of atoms. Furthermore, for ω small we have that $\alpha = \omega a/c$ so, if we identify x with na, A and x_n show identical behaviour.

This and many more examples have led to the feeling that the solutions of difference equations are more or less the same as those of the analogous differential equations. Thus, because differential equations are more easily solved than difference equations, it has been the usual practice to model situations in terms of differential equations.

However, when one considers non-linear equations it is found that non-linear difference equations have a totally unexpected richness in the form of their solutions. This realization has only come about over the last few years, and the study of non-linear difference equations is an exciting new area of mathematics. I should like to give you a glimpse into this new field of research by considering two examples.

For the computer-minded it is worth emphasizing that difference equations are in a form suitable for direct numerical evaluation, whereas it is usually necessary first to replace differential equations by a set of approximate difference equations and then numerically solve this latter set. The difference equations discussed in this paper are readily solved using a calculator or mini-computer. The latter, with the simplest of graphical routines, is ideal.

Non-linear equations

The simplest population model described above has the disadvantage that if $\lambda > 1$ the population increases indefinitely with n. Obviously this is a shortcoming of the model as in practice there is always some limited resource, such as food, which curbs the growth rate λ. This effect itself may be accounted for by replacing equation (3) by the equation

$$N_{m+1} = r(1 - N_m)N_m, \tag{5}$$

where the fixed growth rate λ has been replaced by the effective growth rate $r(1 - N_m)$. This is called the logistic model and is a non-linear difference equation because of the term proportional to N_m^2 on the right-hand side.

To solve such difference equations one simply chooses a value for the constant r and an initial value N_0 and uses equation (5) to generate successive values for N_1, N_2, etc. As an exercise take a range of values of N_0 between 0 and 1 and values for r between 0 and 3. For $0 < r < 1$ you will find that, no matter what the value of N_0, N_m approaches 0 as m increases, whilst for $1 < r < 3$, N_m approaches a value $\bar{N}$ independent of m. This latter value $\bar{N}$ may be obtained analytically from (5) by replacing N_{m+1} and N_m by $\bar{N}$, that is, assume a value of N_m independent of m. In this manner one finds that

$$\bar{N} = 1 - 1/r. \tag{6}$$

You should check your numerical results with this analytic one. (This procedure illustrates the value of using a knowledge of the numerical solution to suggest possible analytic procedures. In this case the fact that the numerical solution shows that N_m for large m approaches a value independent of m suggests looking for the solution $\bar{N}$.) The variation of N_m with m can be compared with the solution of the analogous differential equation

$$\frac{dN}{dt} = \alpha N(1 - N), \tag{7}$$

which is

$$N(t) = 1/(1 + ae^{-\alpha t}).$$

Here a is a constant of integration such that $N_0 = N(t = 0) = 1/(1 + a)$. To compare with the value of N_m as a function of m one merely replaces t by m. For $0 < r < 3$ one finds good qualitative agreement between the differential and difference solutions by identifying α with $r - 1$. This is a further example of the fact that analogous difference and differential equations can have qualitatively similar solutions.

Bifurcations

However, for $r > 3$ the difference equation holds a few surprises. If one takes a value of r in the range between 3 and 3·449 and again solves equation (5) by successive iterations, one finds that, for large m, N_m settles down such that every other value of N_m is equal, i.e. $N_m = N_{m+2} = N_{m+4} \cdots = \tilde{N}$, but $N_m \neq N_{m+1}$. The original state $\bar{N}$ is said to *bifurcate*, that is, break into two. The important point is that the analogous differential equation, namely equation (7), shows no such behaviour, no matter what value of α is taken.

The equation for $\tilde{N}$ can be obtained by using equation (5) twice, first to express N_{m+2} in terms of N_{m+1} and then N_{m+1} in terms of N_m. Then N_{m+1} may be eliminated to express N_{m+2} in terms of N_m and finally both these quantities are put equal to $\tilde{N}$.

In this way one finds

$$\tilde{N} = \frac{(1+r) \pm \sqrt{(r-1)^2 - 4}}{2r}. \tag{8}$$

(Hint: One of the solutions of the cubic equation for $\tilde{N}$ is $\tilde{N} = (1 - (1/r))$.) For $r \geq 3$ one obtains two real solutions for $\tilde{N}$, so that $r = 3$ is the bifurcation point. Again one can test this analytic result against the solution obtained numerically.

As one increases r even further, it is found for $3{\cdot}449 < r < 3{\cdot}544$ that the solutions $\tilde{N}$ bifurcate again so that, for large enough m, $N_{m+4} = N_m$, and the asymptotic solution is four-fold periodic. This phenomenon is relatively easy to find numerically, and is illustrated in Figure 1.

Chaos

This process of bifurcation continues as one increases r, but the range of r over which a particular nth-fold periodic solution holds gets smaller and smaller and in fact this process reaches a limit at a critical value of $r = r_c$ ($= 3{\cdot}570$). For r values below r_c one has the possibility of bifurcating periodic solutions but above r_c the nature of the solution changes dramatically and one then has *chaos*. Chaos has a well-defined mathematical meaning which essentially is that even if sufficiently large values of m are taken, N_m never settles down to a periodic solution but seems to jump from one value to the next in a random fashion. (In practice it is difficult to distinguish

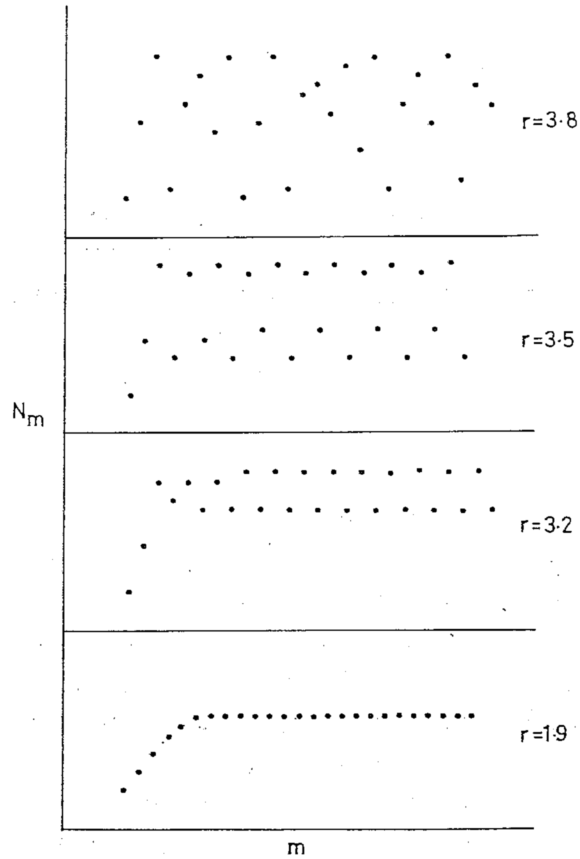


Figure 1. Solution of the logistic equation (5) for various values of the parameter r.

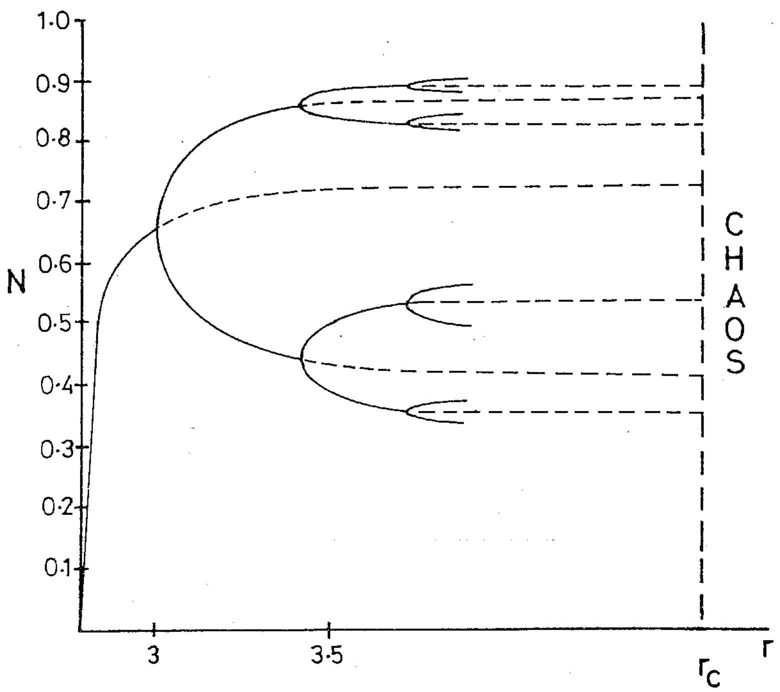


Figure 2. Schematic representation of the bifurcations of equation $N_{m+1} = rN_m(1 - N_m)$.

between a chaotic solution and one that is periodic but of high order, though conceptually these two types of solution are quite distinct.)

This whole phenomenon of continual bifurcation can be visualized by plotting the various asymptotic solutions (that is N_m as m gets large) as a function of r. For $r < 1 N = 0$, for $1 < r < 3$, N is given by (6) and for $3 < r < 3{\cdot}449$ it is given by (8). This behaviour is illustrated in Figure 2.

Try to estimate the critical values of r at which the bifurcations appear and also the critical value appropriate to the onset of chaos. For equation (5) these critical values for r are approximately $r = 3$, $3{\cdot}449$, $3{\cdot}544$, $3{\cdot}564$, and $r_c = 3{\cdot}570$. This same phenomenon occurs not for just the logistic equation but for a whole class of equations of the form $N_{m+1} = F(N_m)$, where F is a function with a general parabolic shape. As an example consider the equation

$$N_{m+1} = 1 - aN_m^2.$$

Remember that the range of values of a which separate the different orders of bifurcation is quite small.

We have seen that difference equations can give rise to much more interesting solutions than the corresponding differential equation. The major point now is whether this type of behaviour is found in real situations or is just a mathematical curiosity. There is some evidence that certain biological systems behave such that their population varies in a chaotic manner. This is encouraging, and suggests that the study of difference equations, particularly those showing chaotic behaviour, should be continued. In the ~~physical~~ sciences difference equations ~~arise, not~~ directly as in biological examples, but only after one has applied a substantial amount of mathematical analysis to the basic model equations. For this reason the relationship with experiment is indirect and hence much harder to substantiate.

Equations which relate N_{m+1} to N_m are called first-order difference equations, as they are similar to first-order differential equations. If one considers, by analogy with second-order differential equations, second-order difference equations in which N_{m+1} is related to N_m and N_{m-1}, then one expects even more interesting behaviour. Such equations occur quite naturally in the biological context when one considers, for example, the population dynamics of a species with two distinct age groups. If X_m and Y_m are the population of the two age groups, a model that illustrates this interaction is:

$$X_{m+1} = r(X_m + Y_m)\exp(-X_m - Y_m),$$
$$Y_{m+1} = X_m, \tag{9}$$

where r is a constant. Another set of equations, introduced by Hénon to model in an indirect manner some properties of a turbulent fluid, may be written in the form

$$N_{m+1} = 1 - aN_m^2 + bN_{m-1}, \tag{10}$$

where a and b are constants. One suspects that such equations will have chaotic solutions but it turns out they have other unexpected types of behaviour. This will be discussed in the next section.

Strange attractors

If in equation (10) we put b equal to 0 then we have an equation of the form mentioned in the last section. A critical value of a exists above which the solution goes chaotic. If b is not 0, then even to study the solution numerically is more complicated. Besides specifying the values of a and b it is necessary to specify two initial conditions, namely N_0 and N_1. Once these values are given, the solution N_m may be obtained by successive iteration of equation (10). Such a procedure leads to a whole set of numbers, but at first sight it is difficult to see any pattern emerging. However, a pattern does become apparent if one considers a 'phase plane plot'. For this one considers two adjacent values (N_m, N_{m+1}) to represent a point in a plane. Then by taking a range of values of m one generates a series of points in this plane. (For second-order differential equations, phase plane plots have recently been discussed in this journal by Jordan and Smith (reference 1).)

The first surprising thing is that, as long as the initial point (N_0, N_1) is not too far from the origin, after a few iterations the points (N_m, N_{m+1}) are found to lie on just a few curves in the plane (the general shape is shown schematically in Figure 3); that is, the points (N_m, N_{m+1}) are 'attracted' to these few curves, and for this reason these curves are said to form an *attractor*. No matter where the initial point lies (as long as it is not too far removed from the origin) the subsequent iterates of equation (10) end up on the attractor. Although it is found that the points (N_m, N_{m+1}) lie on just a few curves, the points jump from one curve to another and to other parts of the same curve in a chaotic manner. This perhaps is not too surprising since for $b = 0$ we have seen that the equation can give chaotic solutions. All this behaviour is found to exist

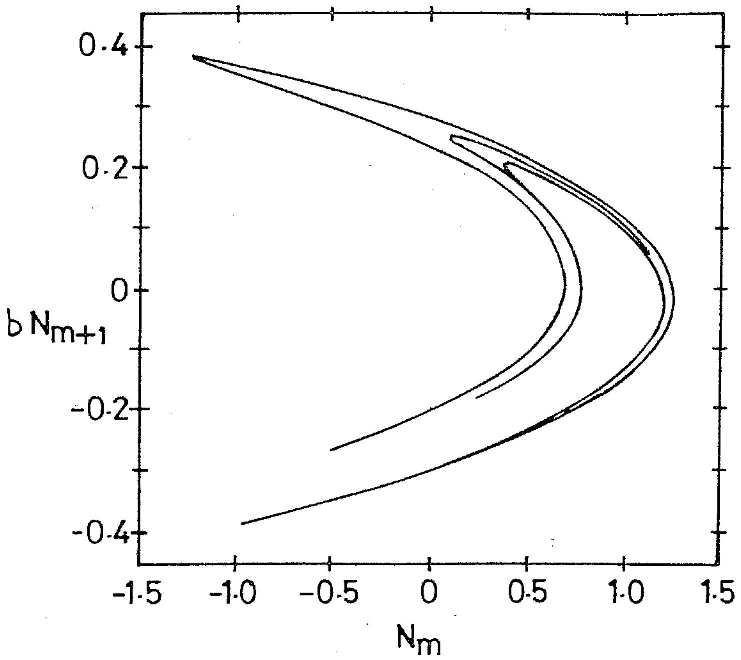


Figure 3. The form of the strange attractor for the Hénon map.

for a range of values of a and b but is best seen for the values originally chosen by Hénon, namely $a = 1{\cdot}4$ and $b = 0{\cdot}3$.

The second and perhaps bigger surprise is found if one looks at the curves of the attractor more closely, as though under a microscope. The number of curves making up the attractor depends on the accuracy of the numerical values of N_m and on the accuracy of their subsequent plotting. (The accuracy of the plotting is usually the limiting condition with modern calculators.) If one looks on a finer scale, that is, if one plots the points to more significant figures of accuracy, then what was one curve now splits into a number of similar curves displaced from each other by small distances. This process can be continued, and at each stage, as the magnifying power is increased, each curve splits up into a set of similar ones. To study this process even a few times unfortunately needs many iterations. In his original work Hénon iterated equation (10) something like ten million times. However, the first one or two splittings can be seen using far fewer iterations.

Based on his numerical results Hénon conjectured that this process of subdivision can be continued indefinitely and that an infinite set of curves is finally generated. This whole set is called a *strange attractor*.

For small b it is possible to obtain an approximation to equation (10) which illustrates many of the features of the attractor; namely (10) is replaced by

$$N_{m+1} = 1 - aN_m^2 \pm \beta\sqrt{1 - N_m \pm \beta}, \tag{11}$$

where $\beta = b/\sqrt{a}$. All possible combinations of the $+$ and $-$ are to be taken. We can write equation (11) in the form $N_{m+1} = F(N_m)$. This is the form of equation we considered in the section on chaos, and so we expect, for a certain range of values of a and b, to get a chaotic solution. This is what happens for $a = 1{\cdot}4$, $b = 0{\cdot}3$. Because of the $\pm$ signs, the function $F(N_m)$ is multivalued, and it is this which gives the first four major curves which constitute the strange attractor. The approximate method used to obtain equation (11) may be extended to generate the whole set of curves.

Equations (9) provide another example of equations whose solution is in the form of a strange attractor. In this case the attractor is even more complicated, being composed of three separate parts, each part having a similar structure to that of the Hénon model.

One cannot discuss strange attractors without mentioning the first one to be discovered, namely the Lorenz attractor. This arose in a study of the flow of fluids, and in particular the flow of air in the atmosphere. The amplitudes of the three most important types of disturbances in the atmosphere are governed by the equations

$$\frac{dx}{dt} = \sigma(x-y),$$

$$\frac{dy}{dt} = -xz + rx - y,$$

$$\frac{dz}{dt} = xy - bz,$$

where σ, b and r are constants. These are much more difficult to solve numerically than the difference equations discussed above, but this has been done, notably by Lorenz, and found to have the intriguing properties. Rather surprisingly these same equations, but with an entirely different interpretation of the symbols, have been shown to model the behaviour of a high-powered laser.

Finally, it should be mentioned that both fluid flows and high-powered lasers show behaviour which qualitatively at least, is explainable in terms of the presence of a strange attractor.

Summary

A few difference equations have been introduced which have immediate relevance to a range of phenomena in the sciences. Though these equations are simple to write down, this simplicity is largely illusory. Their solutions show an immensely complicated structure. We have given examples of equations which have chaotic solutions and others which combine this property with that of attraction to give rise to strange attractors.

Fortunately difference equations are very amenable to numerical solution and thus most of this exotic behaviour can be revealed with the help of relatively small calculators. They are ideal problems for solution on minicomputers with graphical capabilities.

Reference

1. D. W. Jordan and P. Smith, *Mathematical Spectrum*. Vol. 12, No. 3, p. 76, 1979/80.

Letters to the Editor

Dear Editor,

Parties and pies

In Volume 14, Number 2, I believe there is a mathematically significant slip in the penultimate line of the solution to problem 13·9 (page 62). This reads 'If D had not previously met E, then A, B, C, E would contradict the hypothesis.' I believe the points should be either A, B, D, E or A, C, D, E (either will do).

In the previous problem about fruit pies, the apparent paradox is easier to digest (!) once one realises that apple is still a better bet than blackberry even when cherry is considered, unless you are going for the best; which shows that better can be better than best!

Yours sincerely,

BRUCE ANDREWS
(Ipswich School, Henley Road, Ipswich, Suffolk IP1 3SG)

Dear Editor,

Close encounters of the Fermat kind

We all know that $3^2+4^2=5^2$. Indeed, the formula

$$(s^2-t^2)^2+(2st)^2=(s^2+t^2)^2$$

provides infinitely many triples of positive integers x, y, z such that

$$x^2+y^2=z^2.$$

It is also well known how Fermat wrote in the margin of his copy of Bachet's translation of Diophantus' *Arithmetica* by the side of this formula that he had discovered a 'truly wonderful proof' that the equation

$$x^n+y^n=z^n$$

has no solution in positive integers when $n>2$, but that the margin was too small to contain it. Whether or not Fermat's claim was justified, no such proof is known at the present time.

It is possible to obtain what might be termed 'close encounters of the Fermat kind'. Thus

$$5^3+6^3=7^3-2, \qquad 9^3+10^3=12^3+1.$$

The relative error in the latter is $1/12^3<0{\cdot}06\,\%$.

But suppose, for example, that we start with

$$3+4=7.$$

Then

$$(\sqrt[3]{3})^3+(\sqrt[3]{4})^3=(\sqrt[3]{7})^3$$

so, using a calculator, we have

$$(1{\cdot}4422496)^3+(1{\cdot}5874010)^3\simeq(1{\cdot}9129312)^3.$$

If we now multiply by 10^{21}, we obtain

$$(14422496)^3 + (15874010)^3 \simeq (19129312)^3,$$

which may perhaps be termed a 'very close encounter of the Fermat kind'!

Yours sincerely,
ALLAN J. MACLEAN
(Dunfermline High School)

Problems and Solutions

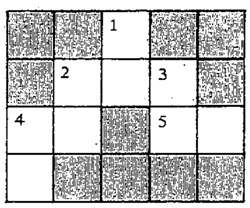
Sixth formers and students are invited to submit solutions to some or all of the problems below: the most attractive solutions will be published in subsequent issues. When writing to the Editorial Office, please state your full name and home address and also the postal address of your school, college or university.

Problems

15.1. (Submitted by M. G. Sykes, Huddersfield New College) Determine the exact value of $\sin 18°$.

15.2. (Submitted by A. K. Austin, University of Sheffield) Each square of the crossnumber is to be filled with one of the digits $0, 1, 2, \ldots, 9$ (no number beginning with 0) so that the following triples of numbers are in arithmetical progression with the terms in ascending order (so that the first plus the third equals twice the second):

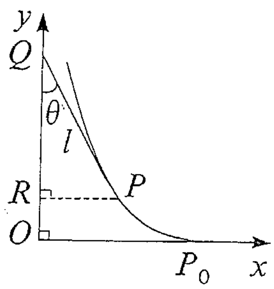
2 down, 4 across, 2 across
2 down, 3 down, 4 across
1 down, 3 down, 4 down,
4 down, 5 across, 2 across.



15.3. (See the article by David Sharpe in this issue.) The symbol f denotes a complex polynomial of positive degree, and z is a root of the derivative f' but not a root of f. If the maximum of the moduli of the roots of f is r, show that $|z| < r$.

Solutions to Problems in Volume 14, Number 2

14.4. (*The Obstinate Dog Problem*) A man has his dog on a lead and walks in a straight line with the lead always taut. Initially the lead is perpendicular to the line along which the man walks. Describe the curve traced out by the dog.



Solution

We received no solutions to this problem from students, so we shall supply the solution given by Lloyd Taylor, who submitted the problem. Suppose the man starts at O and the dog at P_0. The man walks in the direction Oy. Consider the situation when the man has reached the point Q and the dog the point $P(x, y)$. If the lead is of length l, then $PQ = l$, and the slope of the curve traced out by the dog is given by

$$\frac{dy}{dx} = -\frac{QR}{RP} = \frac{-\sqrt{(l^2 - x^2)}}{x}.$$

Thus

$$\int dy = -\int \frac{\sqrt{(l^2 - x^2)}}{x}\, dx + c \ (c = \text{a constant}).$$

We now make the substitution $x = l \sin\theta$, so that

$$y = -\int \frac{l\cos\theta}{l\sin\theta}\, l\cos\theta\, d\theta + c$$

$$y = -l\int \frac{1 - \sin^2\theta}{\sin\theta}\, d\theta + c$$

$$y = l\int (\sin\theta - \operatorname{cosec}\theta)\, d\theta + c$$

$$y = -l\cos\theta + l\log(\operatorname{cosec}\theta + \cot\theta) + c$$

$$y = -\sqrt{(l^2 - x^2)} + l\log\left(\frac{l}{x} + \frac{\sqrt{(l^2 - x^2)}}{x}\right) + c.$$

When $y = 0$, $x = l$, so $c = 0$. Hence the equation of the curve traced out by the dog is

$$y = l\log\left(\frac{l + \sqrt{(l^2 - x^2)}}{x}\right) - \sqrt{(l^2 - x^2)}.$$

This curve is called a *tractrix*.

14.5. Solve the inequality

$$x(x+1)(x+2)(x+3) \geq \frac{9}{16}.$$

Solution (by M. G. Sykes of Huddersfield New College).
The inequality can be rewritten as

$$(x^2+3x)(x^2+3x+2) \geq \frac{9}{16},$$

$$\left(\left(x+\frac{3}{2}\right)^2-\frac{9}{4}\right)\left(\left(x+\frac{3}{2}\right)^2-\frac{1}{4}\right) \geq \frac{9}{16},$$

$$\left(x+\frac{3}{2}\right)^4 \geq \frac{5}{2}\left(x+\frac{3}{2}\right)^2,$$

and this is satisfied when either $x=-3/2$ or

$$\left(x+\frac{3}{2}\right)^2 \geq \frac{5}{2},$$

i.e. when

$$x=-\frac{3}{2} \quad \text{or} \quad x \geq -\frac{3}{2}+\sqrt{\left(\frac{5}{2}\right)} \quad \text{or} \quad x \leq -\frac{3}{2}-\sqrt{\left(\frac{5}{2}\right)}.$$

An alternative way of looking at the problem is to consider the curve with equation

$$y=x(x+1)(x+2)(x+3).$$

This is symmetrical about the line $x=-3/2$ as shown, and, when $x=-3/2$, $y=9/16$. If we put $y=9/16$, we obtain

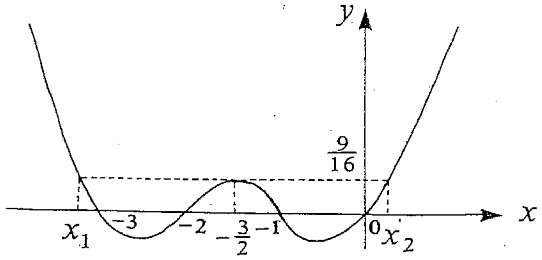
$$x^4+6x^3+11x^2+6x-\frac{9}{16}=0,$$

and this has $(x+3/2)^2$ as a factor, to give

$$\left(x+\frac{3}{2}\right)^2\left(x^2+3x-\frac{1}{4}\right)=0.$$

The values of x_1, x_2 on the diagram are now the roots of the quadratic equation

$$x^2+3x-\frac{1}{4}=0,$$



i.e.

$$x_1 = \frac{-3-\sqrt{10}}{2}, \qquad x_2 = \frac{-3+\sqrt{10}}{2}.$$

Thus, from the diagram, the inequality is satisfied when

$$x \leq \frac{-3-\sqrt{10}}{2} \quad \text{or} \quad x = -\frac{3}{2} \quad \text{or} \quad x \geq \frac{-3+\sqrt{10}}{2}.$$

You may agree that, by looking at the curve, you get a better idea of what is happening.
The problem was also solved by Paul Garcia of the Open University.

14.6. A polynomial in x has integer coefficients and leading coefficient 1, and it takes the value 1 for four distinct integer values of x. Show that there is no integer value of x for which the polynomial takes the value 24.

Solution

We first make a general observation.
Suppose that

$$g(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

is a polynomial with integer coefficients, and suppose that a is an integer root of $g(x)$. Then we can write

$$g(x) = (x-a)(b_{n-1}x^{n-1} + b_{n-2}x^{n-2} + \cdots + b_1 x + b_0),$$

and comparison of the coefficients will show that the b_i are also integers.

Now let $f(x)$ be the polynomial in question. Then $f(x) - 1$ has four distinct integer roots a, b, c, d. It follows from our general observation (applied four times) that we can write

$$f(x) - 1 = (x-a)(x-b)(x-c)(x-d)h(x),$$

where $h(x)$ is a polynomial with integer coefficients. Now suppose that there is an integer r such that $f(r) = 24$. Then

$$23 = (r-a)(r-b)(r-c)(r-d)h(r).$$

Since 23 is prime, at least three of the integers $r-a$, $r-b$, $r-c$, $r-d$ must take the values 1 or -1. Hence two must be the same, which is impossible. This proves the result.

You will notice that we have not needed $f(x)$ to have leading coefficient 1, which Mr R. Lyness pointed out to us. Mr Lyness also pointed out that, in the problem, 'four' cannot be replaced by 'three'. Perhaps you can come up with an example to show this, i.e. a polynomial in x with integer coefficients and leading coefficient 1 which takes the value 1 for *three* distinct integer values of x, and which also takes the value 24 for some integer value of x.

Book Review

An Introduction to BASIC. By M. R. EAGLE. Bell & Hyman, London, 1981 (2nd edition). Pp. vii + 147. £3·50.

30 Hour BASIC. By CLIVE PRIGMORE. National Extension College, Cambridge, 1981. Pp. 254. £5·50.

With the advent of the microcomputer, computing is going to be brought to more people in a more practical way than ever before. The programming language which is most widely available and probably the easiest to start on is the aptly named Beginners Allpurpose Symbolic Instruction Code, or BASIC. The aim of both books is to introduce the language and to generate an appreciation of computers and their potential uses. In my opinion, although it does not go as far as the first book, the second title is the more successful of the two in its aims.

M. R. Eagle's volume is the second edition of a work which was first published in 1976. It seems to have suffered by the passing of time. When first published, there were few microcomputers available in the U.K., so most applications of BASIC were on mainframes or minicomputers. This is no longer the case, as a great many schools now have their own Pets, Apples, and 380Z's and some even have the BBC computer. The book has not taken this enough into account. Lists longer than 255 items and multi-line function definitions will seldom be available to the user of a micro. Equally, the machines mentioned above all support more extensive variable names than those described in the book. None the less the main features of BASIC are all covered adequately, though I should have liked to have seen more worked examples in the text, especially in the use of the built-in functions and string handling. More emphasis could certainly have been placed on this latter topic. On the positive side the author is very good at pointing out the advantages of good use of flow diagrams to construct efficient programs. Overall, I had the impression that this might be a better book to teach from than to learn from, although the teacher will be somewhat frustrated by the fact that the many page cross-references in the text are wrong, all being 10 less than they should be!

30 Hour BASIC on the other hand is certainly a book from which the novice can learn. It has been produced by the National Extension College in cooperation with the BBC TV Computer Programme. It develops a fairly standard form of BASIC but has in mind the BBC computer and gives all the alterations to the programs contained in it which are required to make them run on the BBC machine. The book is intended for use as a home correspondence course with the NEC, and is structured with self-assessment questions, exercises and assignments to be assessed by a course tutor. Solutions to all but these last are contained in the book, so that it may very effectively be used as a self-teaching guide outside the formal NEC course. Cassettes are also available (for use on the BBC computer only) with the various programs in the book on them.

30 Hour BASIC does not describe all the features of the language (used defined functions are omitted, for example) but it is written so that the reader should not be able to move on until the particular point under discussion is properly understood. There are a few niggly misprints early on which might put a complete novice out of his stride, but in general it is well written and produced. It also contains a good chapter on the use of files, an important topic completely ignored by *An Introduction to BASIC*. This book certainly succeeds in its declared aim to 'help you to use a microcomputer with confidence'. After reading it, you will be looking forward to reading the second-stage BASIC course which is shortly to be available from NEC, if you can tear yourself away from using your machine, that is!

University of Durham — NIGEL MARTIN

CONTENTS

ISSN 0025–5653

PRICES (*postage included*)

Prices for Volume 15 (Issues Nos. 1, 2 and 3):

Subscribers in	*Price per volume*	
North, Central and South America	US\$11.50 or £6.00	(Note: These overseas prices apply even if the order is placed by an agent in Britain.)
Australia	\$A10.20 or £6.00	
Britain, Europe and all other countries	£2.75	

A discount of 10% will be allowed on all orders for five or more copies of Volume 15 sent to the same address

Back issues:

All back issues are available; information concerning prices may be obtained from the Editor.

Enquiries about rates, subscriptions and advertisements should be directed to:

Editor—Mathematical Spectrum,
Hicks Building,
The University,
Sheffield S3 7RH, England.

Published by the Applied Probability Trust
Printed in England by Galliard (Printers) Ltd, Great Yarmouth