Mathematical Spectrum

1993/4
Volume 26
Number 3

A magazine for students and
teachers of mathematics in
schools, colleges and universities

Mathematical Spectrum is a magazine for students and teachers 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 26 of *Mathematical Spectrum* will consist of four issues, of which this is the third. The first was published in September 1993, the second in November 1993 and the fourth will appear in May 1994.

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. The copyright of all published material is vested in the Applied Probability Trust.

The Editorial Committee welcomes the submission of suitable material, including correspondence, queries and solutions to problems, for publication in *Mathematical Spectrum*. Students are encouraged to send in contributions. All correspondence about the contents should be sent to:

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

Fermat's Last Theorem— A Theorem At Last

ROGER COOK, *University of Sheffield*

The author is a professor and head of the pure mathematics section at the University of Sheffield. His mathematical interests are mainly in number theory and combinatorics; outside mathematics his interests include kite flying and supporting the Sheffield Steelers (ice) hockey team.

1. Introduction

At approximately 10.30 a.m. on Wednesday 23 June 1993 an announcement was made at the Isaac Newton Research Institute in Cambridge. It was widely reported by the media, with extensive coverage on the main TV news and three pages of the supplement to the *Guardian* newspaper on Thursday morning. What made all this so unusual was that the coverage was about mathematics.

The reason for such an abnormal interest is that Fermat's last theorem is a theorem at last, i.e. it now has a proof. To be more specific, it is now a corollary of Wiles. The proof was announced by Andrew Wiles, a British-born mathematician who now works in Princeton, USA. The title of his talk, 'Modular forms, elliptic curves and Galois representations', during a one-week workshop on 'p-adic Galois representations, Iwasawa theory and the Tamagawa numbers of motives' suggests that the proof may not be readily explained to the uninitiated. This is unfortunate, since Fermat's last theorem is one area of mathematics that has attracted the non-specialist. It is notorious as a result where amateurs will send in false proofs. One professor of mathematics had a standard reply printed. It read 'Your first mistake is on page ...'.

In this article we shall attempt to give readers something of the background to Fermat's last theorem and some idea of the mathematical theories that led Wiles to his historic proof.

2. History

Pierre de Fermat (1601–1665) was a French judge who lived in Toulouse. He was a keen amateur mathematician and although he is now chiefly remembered for his work in the theory of numbers he also worked in geometry and other areas. His most lasting work was on the solution of equations in integers. Such problems are called *Diophantine*, after Diophantus, a Greek mathematician who lived in Alexandria round about 250 AD. Diophantus wrote a 13-volume treatise, called *The Arithmetic*, of which 6 volumes survived. It is concerned not with arithmetic in the school sense but with the theory of numbers, i.e. the properties of the integers. The 6 volumes which survived were reprinted in 1621 by Claude Gaspard de Bachet and it was Fermat's copy of this edition that contained the original statement of his last theorem.

One section of *The Arithmetic* is concerned with Pythagorean triples: can we find positive integers x, y and z such that

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

(In fact there are infinitely many such triples.) Fermat developed a method called 'infinite descent' (essentially the well-ordering principle) to tackle similar problems. He was able to show that the equation

$$x^4+y^4 = z^4$$

has no solution in positive integers.

Fermat wrote (in Latin) in the margin of his copy of *The Arithmetic*:

> 'It is impossible to separate a cube into two cubes, or a fourth power into two fourth powers, or in general any power higher than the second into powers of like degree. I have discovered a truly remarkable proof which this margin is too narrow to contain.'

Fermat's copy of *The Arithmetic* has been lost but this comment appears in the 1670 edition of his collected works. Correspondence between Fermat and Mersenne suggests that the comment dates from 1637. In modern language, Fermat is asserting that, for any integer $n > 2$, the equation

$$x^n+y^n = z^n \qquad (1)$$

has no solution in positive integers x, y and z.

Leonhard Euler (1707–1783) provided an incomplete but essentially correct proof of the case $n = 3$ in his book on algebra in 1822. Another proof was given by Carl Friedrich Gauss (1777–1855) using the complex cube roots of unity

$$w = \tfrac{1}{2}(-1+i\sqrt{3}) = \cos\tfrac{2}{3}\pi+i\sin\tfrac{2}{3}\pi$$

and w^2. The case $n = 5$ was settled independently by Pierre Dirichlet (1805–1859) and Adrien-Marie Legendre (1752–1833) during the 1820s.

3. Cyclotomic fields

Clearly, if x, y and z satisfy (1) and d divides n, say $n = cd$, then

$$(x^c)^d+(y^c)^d = (z^c)^d.$$

Thus, if Fermat's last theorem is false for a particular exponent n then it is also false for all exponents d that divide n. Since the theorem is true for $n = 4$ it is sufficient to prove it whenever n is an odd prime, p say.

In 1847 Gabriel Lamé (1795–1870) announced a 'proof' of Fermat's last theorem to the Paris Academy. We can factorise

$$x^p+y^p = (x+y)(x+qy)(x+q^2y)\cdots(x+q^{p-1}y), \tag{2}$$

where

$$q = \cos\frac{2\pi}{p}+\mathrm{i}\sin\frac{2\pi}{p}$$

is a primitive pth root of unity. We have

$$0 = q^p-1 = (q-1)(q^{p-1}+q^{p-2}+\cdots+q+1)$$

and $q \neq 1$ so that

$$q^{p-1} = -q^{p-2}-\cdots-q-1.$$

Therefore each factor in (2) can be written in the form

$$a_0+a_1q+a_2q^2+\cdots+a_{p-2}q^{p-2}$$

where the a_i's are integers. Let

$$\mathbb{Z}[q] = \{a_0+a_1q+\cdots+a_{p-2}q^{p-2}: a_i \in \mathbb{Z}\}$$

then $\mathbb{Z}[q]$ forms a ring, called the ring of cyclotomic integers. (A ring is a set of objects which can be added, subtracted and multiplied according to the usual rules of elementary arithmetic.) Just as in the ring $\mathbb{Z}$ of integers we can talk about divisibility and factorisation, Lamé argued that each pair of the factors

$$(x+y), \qquad (x+qy), \qquad \ldots, \qquad (x+q^{p-1}y)$$

are coprime and their product is a pth power, namely z^p. He claimed that each factor must be a pth power, just as in factorisation in the integers. From this he was able to deduce Fermat's last theorem.

Joseph Liouville (1809–1882) pointed out a gap in the argument. Although we have divisibility and factorisation in the ring $\mathbb{Z}[q]$ it was not clear that factorisation was unique, whereas in $\mathbb{Z}$ integers have a unique factorisation into primes. Augustin-Louis Cauchy (1789–1857), who is now best remembered for laying the foundations of complex analysis, announced that he was also working on Fermat's last theorem and would be able to prove it soon. During 1847 he published 18 articles trying to plug the gaps in the argument. These efforts were doomed to failure. When we reach $p = 23$ unique factorisation no longer holds in $\mathbb{Z}[q]$. This had already been recognised by the German mathematician Ernst Kummer (1810–1893) in 1843. However, Kummer went on, stimulated by Fermat's last theorem, to discover properties of the cyclotomic fields $\mathbb{Q}(q)$ and to develop a theory of ideals in algebraic number fields.

Various criteria were established for Fermat's last theorem to be true for an exponent n, particularly when n was an odd prime. These calculations progressed further with the introduction of computers. In 1977 S. S. Wagstaff was able to show that Fermat's last theorem is true for each odd prime exponent $p < 125\,000$. It then followed that if the Fermat's last theorem was false for an exponent n then all the prime factors of n must be at least 125 000.

For general exponents m, the theorem is true for at least $\frac{1}{3}$ of all the exponents, since it is true for $n = 3$ and therefore true whenever 3 divides m. Similarly, the theorem is true if 4 divides m, so it is true for $\frac{1}{2} = 1-\frac{2}{3}\times\frac{3}{4}$ of all possible exponents. Using all the primes $p < 125\,000$ it follows from Wagstaff's result that Fermat's last theorem is true for a least a proportion

$$1 - \tfrac{3}{4} \prod_{2<p<125\,000} (1-p^{-1}) = 0.93\ldots$$

of all possible exponents.

4. Elliptic curves

The recent developments concerned with Fermat's last theorem are geometric in nature, and geometric properties are more easily stated in terms of homogeneous coordinates. Starting with a curve in the x-y plane, e.g.

$$y^2 = x^3-x, \tag{3}$$

we introduce a third coordinate z to make all the terms have the same degree, i.e.

$$y^2z = x^3-xz^2. \tag{4}$$

Now all the terms have degree 3 (where we add together the degrees of the different variables in a product).

Clearly, if (x,y,z) satisfies (4) then so does (tx,ty,tz) for any t and the point with $z = 1$ corresponds to a point on the curve (3). We call (3) and (4) the affine and projective versions of the curve, respectively. There is one important difference between them, which is that (4) allows the points $(x,y,0)$, where $z = 0$. This is called the *line at infinity* and including it has the advantage of making true sensible statements that would otherwise be false. For example, if we take a cubic curve and intersect it with a straight line then the intersection consists of three points. Whilst this remark clearly holds for the upper line of figure 1, it is difficult to see three points of intersection of the lower line and the curve. This is because two of them are at infinity! The other major difference is that, in the projective version, all the points (tx,ty,tz) as t varies are identified as a single point on the curve; if $z \neq 0$ we could put $t = 1/z$ and identify the projective point $(x/z,y/z,1) = (X,Y,1)$, say, with an affine point (X,Y).

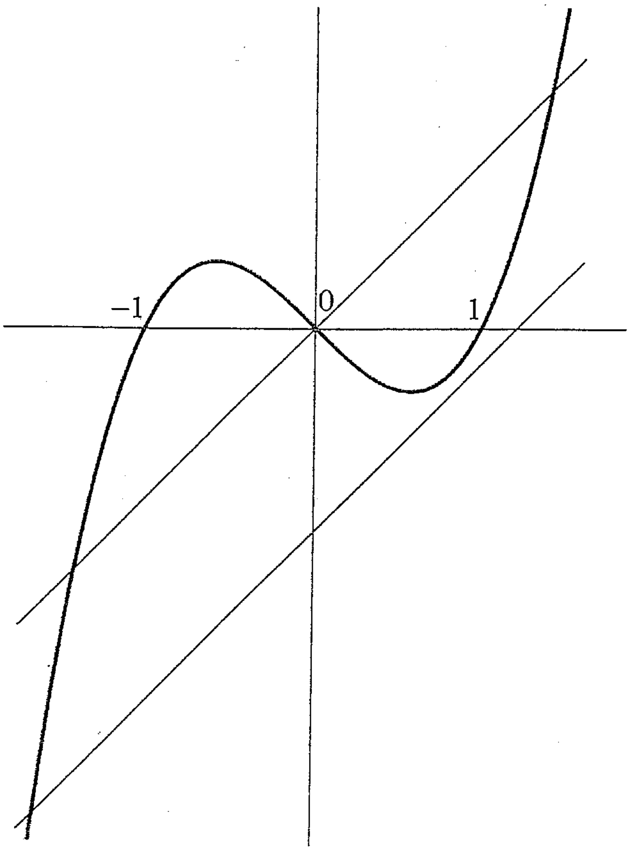


Figure 1

With the curve in projective form

$$C\colon\ F(x, y, z) = 0, \tag{5}$$

we can associate a topological parameter called the *genus*. Here F is a homogeneous polynomial of degree d and has integer coefficients. When the genus is 0 the degree d is 1 or 2 and the curve is a straight line or a conic. The arithmetic properties of such curves were developed by Legendre in the early nineteenth century. The curves of genus 1 are called elliptic curves and, with a suitable choice of coordinates, they can be written in the form

$$y^2 = f(x), \tag{6}$$

where $f(x)$ is a cubic polynomial with distinct roots.

This last condition has a geometrical significance. Writing the equation in the form

$$H(x, y) = f(x) - y^2 = 0, \tag{7}$$

we can find the tangent at a point by evaluating the partial derivatives

$$\frac{\partial H}{\partial x} \quad \text{and} \quad \frac{\partial H}{\partial y}.$$

The equation of the tangent then has the form

$$x\frac{\partial H}{\partial x}+y\frac{\partial H}{\partial y} = \text{constant}, \tag{8}$$

and this is well defined provided that the partial derivatives are not both zero. This can occur at a point (x_0, y_0) if and only if

$$f'(x_0) = 2y_0 = 0,$$

and then

$$f'(x_0) = f(x_0) = 0, \tag{9}$$

so that x_0 is a repeated root of $f(x)$. Thus the condition that $f(x)$ has distinct roots is precisely the condition that the curve has a tangent at each point (such curves are said to be *non-singular* or *smooth*). The elliptic curves of interest to us have equations with rational coefficients and then the roots of $f(x)$ are either three distinct real numbers or a single real number and a pair of complex conjugate roots. At this point you might find it helpful to draw the curves (3) and

$$y^2 = x(x^2+x+1), \tag{10}$$

which demonstrate typical elliptic curves (see figure 2).

The importance of elliptic curves in number theory is that if $f(x)$ has rational coefficients and the curve contains rational points, i.e. points with rational coordinates, then the points on the curve have an arithmetic structure.

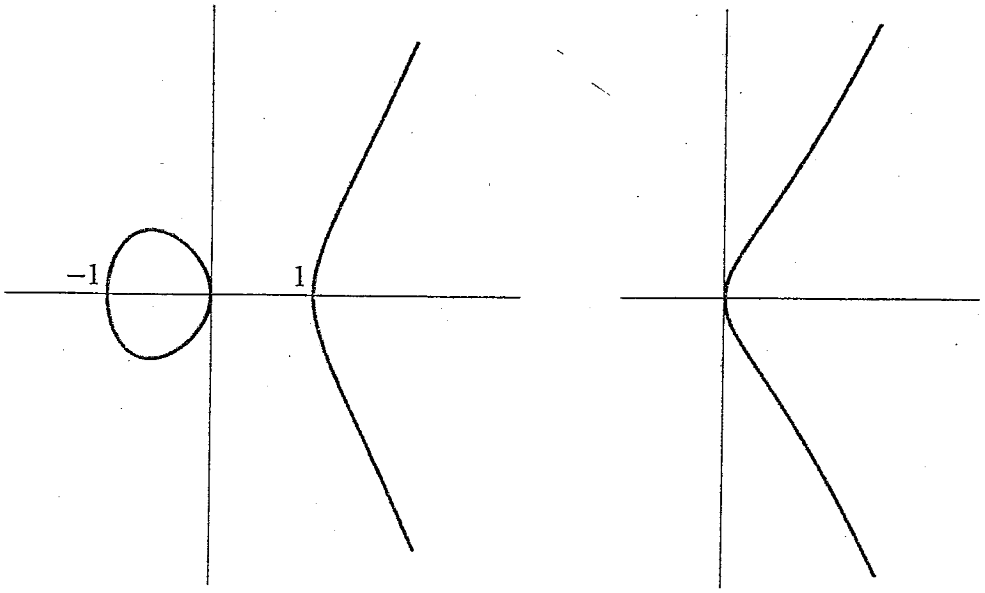
The degree d of F in (5) will be 3, so any straight line meets C in three points. If we take a point on the curve and draw the tangent at that point then the tangent will meet the curve in a third point; if we take two points on the curve and draw the chord through them it will meet the curve in a third point. This third point will also have rational coordinates. In this way, given a set of rational points on the curve C, we can construct other rational points on the curve, by what is called the *chord-tangent construction.*

In 1922 Louis Mordell (1888–1972), an American-born mathematician who worked at Cambridge University, showed that all the rational points on an elliptic curve can be constructed from a finite set of points using this chord-tangent construction.

The affine version of an elliptic curve can be written

$$y^2 = f(x) = x^3-Ax-B. \tag{11}$$

Here $f(x)$ is a cubic polynomial in x with distinct roots and, as indicated, the quadratic term can be taken to have coefficient 0 (after a suitable substitution $x' = x+a$). The condition for the curve to be non-singular is



$y^2 = x^3 - x$ $\qquad$ $y^2 = x(x^2+x+1)$

Figure 2

$$4A^3 - 27B^2 \neq 0, \tag{12}$$

and this is precisely the condition for the polynomial $f(x)$ to have distinct roots.

During the nineteenth century many mathematicians developed a theory of doubly periodic functions of a complex variable. These are called elliptic functions, and are analogous to the trigonometric functions $\sin x$ etc., which are singly periodic functions. One of the most important elliptic functions was Weierstrass's function $\wp(z)$ which satisfies a differential equation

$$\wp'(z)^2 = 4\wp^3(z) - 4A\wp(z) - 4B \tag{13}$$

for certain constants A and B. Thus the points $(\wp(z), \wp'(z))$ lie on the elliptic curve

$$(\tfrac{1}{2}y)^2 = x^3 - Ax - B \tag{14}$$

and so elliptic curves can be parametrised using elliptic functions, although our geometry now has complex coordinates.

5. Fermat's last theorem

In his 1922 paper, Mordell conjectured that a curve of genus greater than 1 contains only a finite number of rational points. This conjecture was proved in 1983 by Gerd Faltings, a young German mathematician who now works at Princeton. He received a Fields Medal for this work; the Fields Medal is roughly the mathematician's equivalent of a Nobel Prize. This was a major step towards Fermat's last theorem, since it showed that, for any fixed exponent $n > 2$ there can only be a finite number of positive integers x, y and z satisfying (1).

In 1985 Roger Heath-Brown showed that Faltings' result implies that Fermat's last theorem holds for almost all exponents, i.e. if $N(x)$ is the number of exponents $n < x$ for which Fermat's last theorem is false, then $N(x)/x \to 0$ as $x \to \infty$.

A connection between elliptic curves and Fermat's last theorem was discovered by Gerhard Frey in 1985. Suppose that there is an odd prime p for which there exist positive integers a, b and c satisfying

$$a^p + b^p = c^p.$$

Frey constructed the elliptic curve

$$E:\ y^2 = x(x - a^p)(x + b^p)$$

and was able to show that it had many interesting properties. In fact, it had so many interesting properties that it was unlikely to exist.

In 1955 Taniyama conjectured that all elliptic curves have a parametrisation in terms of 'modular curves'. This would tie them up with congruence properties modulo some integer N. In 1968 this conjecture was modified by Weil who specified that the integer N should be a geometric invariant of the curve, called the *conductor*. This conjecture was proved for a certain class of elliptic curves, those which have 'complex multiplication', by Goro Shimura in 1971. What Wiles has done is to prove the Taniyama–Weil conjecture for a further class, *semistable elliptic curves*. The exact definition of semistable need not concern us; essentially it means that the curve is non-singular not only when we look at it over the real numbers but also when we look at it '$\bmod p$', for each prime p. Of course we cannot define the 'tangent' $\bmod p$ by a limiting process but we can define the partial derivatives of the function $H(x, y)$ in (7) formally, and then take the tangent to be given by (8).

The importance of this is that Frey's elliptic curve E, if it exists, is known to be semistable; so Wiles has proved that it has a modular parametrisation. On the other hand, Ken Ribet had previously shown that, if the curve existed, then it would be a counterexample to the Taniyama–Weil conjecture; it could not have a modular parametrisation. Therefore such curves cannot exist, which means that there are no counterexamples to Fermat's last theorem!

References

Intermediate

1. R. B. J. T. Allenby and E. J. Redfern, *Introduction to Number Theory with Computing* (Edward Arnold, London, 1989).
2. E. T. Bell, *Men of Mathematics* (2 volumes) (Penguin, London).
3. C. B. Boyer, *A History of Mathematics* (Wiley, New York, 1968).

4. L. J. Mordell, *Diophantine Equations* (Academic Press, New York, 1969).

Advanced

5 H. M. Edwards, *Fermat's Last Theorem* (Springer-Verlag, New York, 1977).

6. D. Husemoller, *Elliptic Curves* (Springer-Verlag, New York, 1987).

7. K. Ireland and M. Rosen, *A Classical Introduction to Modern Number Theory*, 2nd edn. (Springer-Verlag, New York, 1990).

8. N. Koblitz, *Introduction to Elliptic Curves and Modular Forms* (Springer-Verlag, New York, 1984).

9. S. Lang, *Diophantine Geometry* (Wiley, New York, 1962).

10. S. Lang, *Cyclotomic Fields I & II* (Springer-Verlag, New York, 1990).

11. P. Ribenboim, *13 Lectures on Fermat's Last Theorem* (Springer-Verlag, New York, 1979).

12. J. H. Silverman, *The Arithmetic of Elliptic Curves* (Springer-Verlag, New York, 1986).

Articles

13. D. R. Heath-Brown, 'Fermat's last theorem for "almost all" exponents', *Bulletin of the London Mathematical Society*, **17** (1985), 15–16.

14. N. Koblitz, 'Why study equations over finite fields?', *Mathematics Magazine* **55** (1982), 144–149.

15. S. Lang, 'Review of L. J. Mordell's *Diophantine Equations*, *Bulletin of the American Mathematical Society*, **76** (1970), 1230–1234.

16. B. Mazur, 'Number theory as gadfly', *American Mathematical Monthly* **98** (1991), 593–610.

17. L. J. Mordell, 'Review of Lang's *Diophantine Geometry*', *Bulletin of the American Mathematical Society* **70** (1964), 491–498.

18. S. S. Wagstaff, 'The irregular primes to 125 000', *Mathematics of Computation* **32** (1977), 583–591.

Mordell's book on Diophantine equations contains a wealth of information about particular Diophantine equations, whereas a much more abstract approach is taken in Lang's book on Diophantine geometry. For a robust discussion of the merits of these approaches you should read Mordell's review of Lang's book and the subsequent review of Mordell's book by Lang.

The 1994 Puzzle

Readers would be disappointed not to have the annual challenge, to express a range of numbers in terms of the digits of the year in order, using only the operations $+$, $-$, $\times$, $\div$, $\sqrt{}$, $!$ and concatenation (putting 1 and 9 together to give 19, for instance). For example, $1 = -1+9-9+\sqrt{4}$. Since the numbers 1 to 100 seem easier for 1994, we invite readers to attempt the range 1 to 150.

An Introduction to Steiner Systems

MIKE GRANNELL AND TERRY GRIGGS, *University of Central Lancashire, Preston*

The authors are both graduates of the University of London, and have been working together in combinatorial design theory for 15 years. Construction of the Steiner system $S(5,6,108)$ described in this article was one of the first problems they tackled (unsuccessfully) many years ago, but a return to it in 1991 proved successful.

1. Introduction

Begin with a base set of nine elements, say the positive integers from 1 to 9 inclusive. Next consider the following subsets: 123, 456, 789, 147, 258, 369, 159, 267, 348, 168, 249 and 357. (Here, and elsewhere in this article, we will, when convenient, use the simpler notation *abc* for the set $\{a, b, c\}$.) Collectively the above subsets, which are more usually called blocks, have the property that every pair of elements of the base set is contained in precisely one of the blocks. Before proceeding further, readers should verify this fact for themselves. Such a collection of blocks is an example of a Steiner system and this particular configuration is usually denoted by $S(2,3,9)$. It is quite easy to see what the three integers 2, 3 and 9 describe. Considering them in reverse order, the 9 gives the number of elements in the base set, and these can be any nine elements; we named them as the positive integers 1 to 9 inclusive only for convenience. The 3 gives the number of elements in each block and the 2 tells us that every pair of elements of the base set, i.e. every subset consisting of two elements, is contained in precisely one block. More succinctly we have a covering of pairs of the base set, each precisely once, by a collection of triples. In this short article our aim will be to introduce some of the elementary theory of Steiner systems, which is a branch of combinatorial mathematics, and to try to give the flavour of what we feel is an intrinsically very interesting topic with many fascinating open problems to be solved.

In formal terms, a Steiner system $S(t,k,v)$ comprises a base set having v elements and a family of k-element subsets of this base set. These k-element subsets are called blocks. The blocks have the property that each t-element subset of the base set appears in precisely one block. The reason that these structures have the name of Steiner associated with them is that, in 1853, the Swiss geometer Jakob Steiner (1796–1863) proposed the problem of how to construct them. Six years later M. Reiss published a solution for the case $t = 2$ and $k = 3$. However, as we describe later, both Steiner and Reiss had been anticipated.

Steiner systems and other related structures have applications in statistics through the design of experiments and in coding theory. For example, consider the problem of comparing nine different breakfast cereals. A single person could not rank all nine with confidence; by the time he or she was on the ninth they would have forgotten the taste of the first. To be reasonable, we can only ask people to rank at most three each. If we asked people to rank two each then to get all $\frac{1}{2}\times9\times8 = 36$ comparisons we would need 36 people. However, if we use the $S(2,3,9)$ described above then we can reduce this to 12 people each testing three brands and be sure that every pair is compared precisely once.

As an example in coding theory, look at the 12 blocks of $S(2,3,9)$ above. Each block has built-in redundancy in the sense that if any two of its digits are correctly received then the block is uniquely defined. This property forms the basis for the construction of codes which can both detect and correct errors. Such codes are particularly important in telecommunications. Similar codes have been used in space missions. Further details can be found in the book by Ian Anderson (reference 1).

To conclude this introduction it is perhaps instructive to give a second example. Let the base set be $\{A,B,C,D,E,F,G,H\}$ and let the blocks be *ABCH*, *ADEH*, *AFGH*, *BDFH*, *BEGH*, *CDGH*, *CEFH*, *DEFG*, *BCFG*, *BCDE*, *ACEG*, *ACDF*, *ABEF* and *ABDG*. These form a Steiner system $S(3,4,8)$; there are eight elements in the base set, each block contains four elements and it is easily verified that every triple, i.e. every subset consisting of three elements, is contained in precisely one block.

2. Basic theory

Firstly we develop some elementary but important ideas. Suppose t, k and v are integers satisfying $0 < t < k < v$. Now it is impossible for a system $S(t,k,v)$ to be constructed for all values of t, k and v satisfying the inequality as will be quickly realized if the reader attempts to construct an $S(2,3,8)$. So we need to ascertain which values of t, k and v may allow us to construct Steiner systems. Simple necessary conditions on these paramenters can be deduced from the following two easy theorems.

Theorem 1. If there exists an $S(t,k,v)$ then there exists an $S(t-1,k-1,v-1)$.

Proof. Choose any element, say x, of the system. Remove all the blocks which do not contain x. Those which remain contain precisely once every t-element subset which also contains x. Thus if we remove the element x from these blocks, what is left is a base set of $v-1$ elements and blocks containing $k-1$ elements which collectively cover every $(t-1)$-element subset precisely once.

As an illustration of this theorem the reader may obtain a Steiner system $S(2,3,7)$ from the $S(3,4,8)$ given above. Simply choose x as any letter from A to H and apply the procedure described in the proof.

Theorem 2. If there exists an $S(t,k,v)$ then ${}^kC_t$ divides ${}^vC_t$.

Proof. ${}^kC_t$ is the number of t-element subsets in a block. If b is the total number of blocks in the system then the number of t-element subsets covered is $b \times {}^kC_t$. But every t-element subset appears precisely once and there are ${}^vC_t$ t-element subsets, so $b \times {}^kC_t = {}^vC_t$. Hence ${}^vC_t / {}^kC_t$ is an integer (b in fact).

Although the two theorems above are elementary, if we combine them we obtain a condition on the parameter set, the so-called admissibility condition, for the possible existence of a Steiner system $S(t,k,v)$.

Admissibility condition. If there exists an $S(t,k,v)$ then ${}^{k-i}C_{t-i}$ divides ${}^{v-i}C_{t-i}$ for each $i = 0,1,2,\ldots,t-1$.

Proof. By continued application of theorem 1, there exists an $S(t-i,k-i,v-i)$ for $i = 0,1,2,\ldots,t-1$. Then apply theorem 2.

At this point it is perhaps worth noting that, in general, each of the admissibility criteria ${}^kC_t$ divides ${}^vC_t$, ${}^{k-1}C_{t-1}$ divides ${}^{v-1}C_{t-1}$, etc., forces different conditions on the parameter set $\{t,k,v\}$. They are independent constraints and one does not necessarily imply another. A parameter set which satisfies the admissibility criteria is called an admissible set.

3. Steiner triple systems

We are now in a position to deduce, for given values of t and k, which integers v form an admissible parameter set $\{t,k,v\}$. When $t = 1$, the admissibility condition reduces to the statement that k must divide v, which is elementary anyway and it is equally elementary how to construct such systems. As an example, to construct an $S(1,5,15)$ let the base set be $\{a,b,c,d,e,f,g,h,i,j,k,l,m,n,o\}$; the blocks can then be chosen to be *abcde*, *fghij* and *klmno*. The first non-trivial case is when $t = 2$ and $k = 3$, the covering of pairs of elements by triples, and these systems are more often called Steiner triple systems. We begin by working out the admissibility condition on v. By theorem 1, if there exists $S(2,3,v)$ then there exists $S(1,2,v-1)$. Hence 2 divides $v-1$, i.e. v is odd and can be written in the form $v = 2s+1$, where s is a positive integer. Now applying theorem 2 to $S(2,3,v)$, we see that ${}^3C_2$ divides ${}^vC_2$, i.e. 3 divides $\frac{1}{2}v(v-1) = s(2s+1)$. Since 3 is prime then either 3 divides s or 3 divides $2s+1$, i.e. either s must be of the form $3r$ or of the form $3r+1$, where r is a positive integer. Hence $v = 6r+1$ or $6r+3$. Note here what we have done and, more importantly, what still remains to be done. We have not shown that there actually exist Steiner triple systems $S(2,3,v)$ when $v = 6r+1$ or $6r+3$. We have merely shown that these are the only possible values of v for which such systems can exist. The work of showing whether the necessary admissibility condition is also sufficient is still to come and is more difficult. What is required is a general construction or a number of constructions to produce Steiner triple systems. The first person to solve this problem was an Anglican clergyman living in the nineteenth century. The Reverend T. P. Kirkman (1806–1895) was Rector of Croft, near Warrington in what was then Lancashire. In 1847 (reference 4) he published a paper giving the complete solution to the problem of constructing Steiner triple systems. In the sense that he did not earn his living from mathematics, Kirkman belonged to the line of great amateurs whose contributions have so enriched and advanced the subject. By right the systems should be called Kirkman triple systems, but Kirkman's work was overlooked for many years. However, Kirkman's contributions to the development of combinatorial mathematics were eventually recognized and the name Kirkman triple system is now given to a special type of Steiner triple system with additional properties. Since Kirkman's time many other different constructions of Steiner triple systems have been discovered and we give below our favourite which occurs in the work of the American mathematician, R. M. Wilson (reference 8).

List all triples a,b,c (with a, b and c not necessarily distinct) such that $a+b+c \equiv 0 \pmod{v-2}$. It can be proved that the number of triples so obtained is precisely the required number of blocks for an $S(2,3,v)$. Because some triples contain repeated elements and there are only $v-2$ elements rather than v, we do not as yet have a Steiner triple system. But, as will be seen from the example which follows, such a system can easily be constructed with a little modification. We illustrate the method when $v=15$. There are three types of triples which sum to zero in arithmetic modulo 13.

Type A (all elements different)

0, 1, 12	0, 2, 11	0, 3, 10	0, 4, 9	0, 5, 8	0, 6, 7
1, 2, 10	1, 3, 9	1, 4, 8	1, 5, 7	2, 3, 8	2, 4, 7
2, 5, 6	3, 4, 6	3, 11, 12	4, 10, 12	5, 9, 12	5, 10, 11
6, 8, 12	6, 9, 11	7, 8, 11	7, 9, 10		

Type B (two elements equal)

1, 1, 11	11, 11, 4	4, 4, 5	5, 5, 3	3, 3, 7	7, 7, 12
12, 12, 2	2, 2, 9	9, 9, 8	8, 8, 10	10, 10, 6	6, 6, 1

Type C (all elements equal)

$0, 0, 0$

For all $v = 6r+1$ or $6r+3$, then $v-2 = 6r-1$ or $6r+1$ is not divisible by 3. Therefore $0, 0, 0$ will be the only type C triple. Repeated elements in the type B and type C triples must now be replaced by two further elements which we call X and Y. Firstly $0, 0, 0$ becomes $0, X, Y$. Next, considering type B triples, observe that these are listed to form a cycle with the non-repeated element of each triple being the repeated element of the next. The first triple $1, 1, 11$ becomes $X, 1, 11$; the second $11, 11, 4$ becomes $Y, 11, 4$; the next $4, 4, 5$ becomes $X, 4, 5$. Continuing in this way, replacing one of the repeated elements in the triples alternately with X and Y, we reach $Y, 6, 1$. It is easily verified that we have constructed an $S(2, 3, 15)$ on the base set $\{0, 1, 2, \ldots, 12, X, Y\}$.

The only problem with this method can occur when the type B triples do not form a single cycle as in the example used. This does not matter if all such cycles contain an even number of triples as the replacement by X and Y in each cycle can be handled independently. However, in certain cases, odd cycles occur. For example when $v-13$ the type B triples are $1, 1, 9$ $9, 9, 4$ $4, 4, 3$ $3, 3, 5$ $5, 5, 1$ and $10, 10, 2$ $2, 2, 7$ $7, 7, 8$ $8, 8, 6$ $6, 6, 10$. However, observe that these two cycles form a pair, each the negative of the other in arithmetic modulo 11. We replace these triples as follows. The first one becomes $X, 1, 9$ $Y, 9, 4$ $X, 4, 3$ $Y, 3, 5$ and $0, 5, 1$ since the latter cannot be either $X, 5, 1$ or $Y, 5, 1$. Similarly the second one becomes $X, 10, 2$ $Y, 2, 7$ $X, 7, 8$ $Y, 8, 6$ and $0, 6, 10$. Finally two of the type A triples $0, 1, 10$ and $0, 5, 6$ are amended to become $Y, 1, 10$ and $X, 5, 6$, respectively. If this procedure seems complicated then it is suggested that the reader tries out the cases $v = 19, 21, 25, 27$, etc., when it will be realized that this is an extremely simple method of constructing Steiner triple systems. In fact this method works for all $v = 6r+1$ or $6r+3$, thus showing that the necessary admissibility condition for a Steiner triple system is indeed sufficient.

4. Large Steiner systems

It is perhaps a surprise that after Kirkman's paper over a century passed until another case was completely solved. In 1960 H. Hanani proved that the necessary admissibility condition $v = 6r+2$ or $6r+4$ is also sufficient for Steiner systems $S(3, 4, v)$ (reference 3). Since then Hanani has also proved that the necessary admissibility condition $v = 12r+1$ or $12r+4$ and $v = 20r+1$ or $20r+5$ are also sufficient for Steiner systems $S(2, 4, v)$ and $S(2, 5, v)$, respectively. Today these are the only four pairs of values of t and k, i.e. $t = 2$, $k = 3, 4, 5$ and $t = 3$, $k = 4$

for which the problem of constructing Steiner systems $S(t,k,v)$ for all possible values of v is completely solved. However, some very recent work has resulted in the problem being almost completely solved when $t = 2$ and $k = 6, 7, 8, 9$. At this point it might also be worth noting that, unlike the cases which have been completely solved, in general it is known that the necessary admissibility condition is not always sufficient. For example $v = 36$ is an admissible value for $t = 2$ and $k = 6$, but there is no Steiner system $S(2,6,36)$. This problem, which is also known as the problem of the 36 officers and is related to ideas in finite geometry, goes back to Euler and was shown to be impossible by G. Tarry (reference 7).

However, if the state of knowledge concerning the existence of Steiner systems with $t = 2$ and $t = 3$ is patchy, results dealing with $t = 4$ and $t = 5$ are very scarce indeed and when $t \geqslant 6$ they are completely non-existent! Until 1975 the only known systems of these types were $S(5,6,12)$ and $S(5,8,24)$ together with the systems $S(4,5,11)$ and $S(4,7,23)$ obtained from them using theorem 1. The existence of all these systems is related to some deep results in group theory. In 1975, R. H. F. Denniston (reference 2) constructed further systems $S(5,6,v)$ for $v = 24$, 48 and 84 and $S(5,7,28)$. Two years later W. H. Mills (reference 5) added $S(5,6,72)$. There was then no further progress for over 10 years. Denniston's systems were constructed using hand calculations. Recently, using a computer, we have constructed $S(5,6,108)$. The reason for using a computer can be seen if the number of blocks in a Steiner system $S(5,6,108)$ is calculated: there are precisely 18 578 196 of them. Using sophisticated computer equipment at the University of Toronto and working with Professor Rudi Mathon of that university, we subsequently constructed $S(5,6,132)$, consisting of 51 553 216 blocks. Truly enormous systems! Our next target is $S(5,6,168)$, with no fewer than 175 036 708 blocks, though we appear to be on the limit both of mathematical reasoning concerning what the structure of such a system might be and of computer technology to effect the calculations involved.

To conclude, we give a simple construction of the Steiner system $S(5,6,12)$. The construction, which is purely combinatorial in nature, was first given by R. G. Stanton (reference 6). We start by constructing $S(4,5,11)$, which contains 66 blocks. Let the base set be $V = \{A,B,C,D,E,1,2,3,4,5,6\}$ and suppose that $ABCDE$ is the first block. It is then easy to calculate the numbers of other blocks which contain specified numbers of letters and numbers. Specifically there are 30 blocks of the form $LLLNN$, 20 blocks of the form $LLNNN$ and 15 blocks of the form $LNNNN$, where L is a letter and N a number.

Begin with the *LNNNN* blocks. Consider the following scheme:

A	12	34	56
B	13	25	46
C	14	26	35
D	15	24	36
E	16	23	45

Note that each pair of digits occurs in the scheme precisely once and further that each digit occurs precisely once in each row. From each row form three blocks of the system by the letter and two of the three pairs of digits. Thus the first row generates blocks *A*1234, *A*1256 and *A*3456. This gives 15 blocks of the form *LNNNN*.

Considering next the blocks of the form *LLNNN*, observe that there are six numbers of which we require three in each block and that ${}^6C_3 = 20$, exactly the number required. Each triple of digits is contained in precisely one block and it is determined which the two letters must be. For example, consider 123. Blocks *A*1234, *B*1325 and *E*1623 already occur, so we must have 123*CD*.

Finally, when the 30 blocks of the form *LLLNN* are considered, everything is forced, as the reader will find if the construction is followed through. Obtaining the larger system $S(5,6,12)$ is easy. First a twelfth element, say ∞, is adjoined to all the blocks of the $S(4,5,11)$. Then 66 further blocks are created as the complements of the existing 66 blocks; the 132 blocks so formed are a Steiner system $S(5,6,12)$.

References

1. I. Anderson, *A First Course in Combinatorial Mathematics*, 2nd edn. (Clarendon Press, Oxford, 1989).

2. R. H. F. Denniston, Some new 5-designs, *Bull. London Math. Soc.* **8** (1976), 263–267.

3. H. Hanani, On quadruple systems, *Canad. J. Math.* **12** (1960), 145–157.

4. T. P. Kirkman, On a problem in combinations, *Cambridge and Dublin Math. J.* **2** (1847), 191–204.

5. W. H. Mills, A new 5-design, *Ars Combinatoria* **6** (1978), 193–195.

6. R. G. Stanton, A conjecture on quintuple systems, *Ars Combinatoria* **10** (1980), 187–192.

7. G. Tarry, Le problème de 36 officiers, *C.R. Assoc. France Avanc. Sci. Nat.* **1** (1900), 122–123.

8. R. M. Wilson, Some partitions of all triples into Steiner triple systems, *Springer Lecture Notes in Mathematics* **411** (1974), 267–277.

A Note on Wilson's Theorem

H. SAZEGAR, *Mashad, Iran*

The author works for a petrochemical company, and also finds time to pursue his interest in number theory.

Wilson's theorem (see reference 1, page 87, for example) states that for any prime p

$$(p-1)! \equiv -1 \pmod{p}.$$

Although the result is attributed to Wilson, the first proof was given by Lagrange in 1771.. However, there is evidence that Leibniz knew the result almost a century before.

The converse of Wilson's theorem is also true: for any integer $m > 1$ with $(m-1)! \equiv -1 \pmod{m}$, we have m prime.

Our purpose here is to give a proof of the following generalization.

Theorem. For any prime number p and any positive integer n,

$$\frac{(np-1)!}{(n-1)!\,p^{n-1}} \equiv (-1)^n \pmod{p}.$$

Wilson's theorem is just the case $n = 1$, and we use it as the first step in an induction argument.

Consider

$$\begin{aligned} g(x) &= (x+1)(x+2)\cdots(x+p-1) \\ &= x^{p-1}+a_{p-2}x^{p-2}+\cdots+a_1x+(p-1)!. \end{aligned}$$

Then

$$g(p) \equiv (p-1)! \equiv -1 \pmod{p}.$$

But

$$g(p) = \frac{(2p-1)!}{(p-1)!\,p},$$

and multiplying both sides by $(p-1)!$ we have

$$\frac{(2p-1)!}{p} \equiv (-1)(p-1)! \equiv (-1)^2 \pmod{p},$$

which is the case $n = 2$.

Now suppose that, for some positive integer k,

$$\frac{(kp-1)!}{(k-1)!\,p^{k-1}} \equiv (-1)^k \pmod{p}$$

(and that the expression on the left is an integer). Now

$$\begin{aligned} g(kp) &= (kp+1)\cdots[(k+1)p-1] \\ &\equiv (p-1)! \\ &\equiv -1 \pmod{p}, \end{aligned}$$

so

$$\begin{aligned} (-1)^{k+1} &\equiv g(kp)\frac{(kp-1)!}{(k-1)!\,p^{k-1}} \\ &\equiv [(k+1)p-1]\cdots(kp+1)\frac{(kp-1)!}{(k-1)!\,p^{k-1}} \\ &\equiv \frac{[(k+1)p-1]!}{k!\,p^k} \pmod{p}, \end{aligned}$$

which is just the case $k+1$, completing the inductive proof.

Further, the converse is true. Suppose that $m > 1$ is an integer such that

$$\frac{(nm-1)!}{(n-1)!\,m^{n-1}} \equiv (-1)^n \pmod{m}$$

for some positive integer n. Then m is prime.

If m is not prime we can write $m = cd$ for some divisors c and d with $1 < c, d < m$. Then $d\,|\,(m-1)!$ and

$$(nm-1)! = 1\times2\cdots(m-1)m\cdots2m\cdots(n-1)m\cdots(nm-1)$$

is divisible by $(m-1)!\,(n-1)!\,m^{n-1}$. Now

$$d\,|\,(m-1)! \quad \text{and} \quad (m-1)!\left|\,\frac{(nm-1)!}{(n-1)!\,m^{n-1}}\right.,$$

so

$$\frac{(nm-1)!}{(n-1)!\,m^{n-1}} \equiv 0 \pmod{d},$$

and this gives a contradiction.

References

1. G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, 4th edn. (Oxford University Press, 1965).

On a Theorem of Liouville

H. SAZEGAR, *Mashad, Iran*

A theorem of Liouville states that, for any prime number $p > 5$ and any positive integer m,

$$(p-1)!+1 \neq p^m.$$

(We know that $(p-1)!+1$ is a multiple of p by Wilson's theorem—see the previous article.) Here we prove a result which is a generalization of Liouville's theorem when m is odd.

Theorem. For any prime number p, any odd integer t in the range $-p+2 < t < p$ and any non-negative integer m,

$$(p+t)!-t^m \neq p^m.$$

Proof. Assume that $(p+t)!-t^m = p^m$ for some m.

(a) If $m = 0$ then $(p+t)! = 2$, contrary to $-p+2 < t$.

(b) If $m = 1$, then $(p+t)! = p+t$, so $p+t = 1$ or 2, and again we obtain a contradiction.

(c) For even $m > 1$ we have

$$(p+t)|(p+t)! = p^m+t^m$$

and

$$\begin{aligned} p^m &= [(p+t)-t]^m \\ &= (p+t)^m-m(p+t)^{m-1}t+\cdots+t^m \end{aligned}$$

and so $(p+t)|(p^m-t^m)$. Hence $(p+t)|2p^m$ $(2 < p+t < 2p)$. The only possibility is $p+t = p$. Since $t \neq 0$, this gives a contradiction.

(d) For odd $m > 1$ we use

$$\begin{aligned} p^m &= [(p+t)-t]^m \\ &= (p+t)^m-m(p+t)^{m-1}t+\cdots+m(p+t)t^{m-1}-t^m \end{aligned}$$

and hence we find that

$$(p+t)^2|[p^m+t^m-m(p+t)t^{m-1}].$$

If $p = 2$ then $t = 1$ and clearly $3!-1 = 5$ is not a power of 2. Otherwise $p > 2$ is odd, $p+t$ is even and so of the form $2s$ $(s > 1)$.

If $s = 2$ then $p+t = 4$ and we would have

$$24 = 4! = p^m+t^m = p^m-(p-4)^m,$$

as m is odd. Hence $24 \equiv 4mp^{m-1} \pmod{16}$, so $4 \equiv mp^{m-1} \pmod 4$. But m and p are odd, so this is not possible.

Now suppose that $p+t = 2s$, with $s > 2$. Then

$$(p+t-1)! = 1\times 2\cdots s\cdots(p+t-1)!$$

is divisible by $2s = p+t$. Hence

$$(p+t)^2 | (p+t)! = p^m + t^m.$$

Thus we now have

$$(p+t)^2 | m(p+t)t^{m-1}$$

and so $(p+t)|mt^{m-1}$. Let $d = \mathrm{hcf}(p+t, t^{m-1})$. Then

$$d | (p+t)! - t^m = p^m.$$

If $d \neq 1$, then d is a power of p. But $2 < p+t < 2p$ and $t \neq 0$, so we must have $d = 1$. Now $(p+t)|mt^{m-1}$ implies $(p+t)|m$. But $p+t$ is even and m is odd, so this contradiction completes the proof.

The Smarandache Function

I. For a positive integer n, $S(n)$ is defined as the smallest positive integer such that n divides $S(n)!$ (S is called the Smarandache function: see Volume 26 Number 2, pages 39–40 and Problem 26.5, and Problem 26.8 in this issue.) Thus

$$S(43) = 43, \quad S(46) = 23, \quad S(57) = 19, \quad S(68) = 17,$$
$$S(70) = 10, \quad S(72) = 6, \quad S(120) = 5,$$

so there is an increasing sequence of seven terms whose 'Smarandache values' are strictly decreasing. Can readers find a longer sequence with this property, and is there a bound on the length of such a sequence? Is there any way of extending the Smarandache function to rational or real or complex numbers?

Reference

1. Mike Mudge, 'The Smarandache function', *Personal Computer World* (1992) p. 420.

J. Rodriguez
(Sonora, Mexico)

II. Let $S(n)$ be defined as the smallest integer such that $S(n)!$ is divisible by n (the Smarandache function). For what triplets does this function satisfy the Fibonacci relationship? In other words, find n such that

$$S(n)+S(n+1) = S(n+2).$$

I have checked the first 1200 numbers and found just two triplets for which this holds:

$$S(9)+S(10) = S(11), \qquad \text{i.e. } 6+5 = 11$$

and

$$S(119)+S(120) = S(121), \qquad \text{i.e. } 17+5 = 22.$$

Reference

1. Mike Mudge, 'A return visit to the Florentin Smarandache function', *Personal Computer World* (February 1993), p. 403.

T. Yau
(Student,
Pima Community College,
Tucson, Arizona, USA)

III. Alphanumerics and solutions

Show that

$$
\begin{array}{rrl}
 & \texttt{SMARANDACHE} & * \\
 & \texttt{FUNCTION} & * \\
 & \texttt{IN} & \\ \cline{2-2}
= & \texttt{NUMBERTHEORY} &
\end{array}
$$

is impossible with * = +, −, × or ÷.

T. Yau

Palindromic numbers

What is the 140th palindromic number and how many palindromic numbers are there up to 14041?

Alex J-C. Chen
(University of California at Berkeley)

The Epi/hypocycloid

J. H. LITTLEWOOD

During the Second World War the author worked for the Mine Design Department of the Admiralty on countermeasures for the German acoustic mine. In 1946 he joined the Research Department of the London, Midland and Scottish Railway Company, where he was Head of Instrumentation. After the railway system was nationalised he became Superintendent of Field Trials at British Rail Research Department. He is now retired.

The *epicycloid* is defined as the locus of a point on the circumference of a circle which rolls round the outside of a fixed circle along the perimeter. If the rolling is on the inside, then the curve is a *hypocycloid.* If at the end of the first revolution of the rolling circle (say on the outside, being an epicycloid), the next revolution is on the inside (i.e. a hypocycloid), and so on, the curve is designated an *epi/hypocycloid.* With help from a computer, using GW-Basic, a family of these curves is presented here.

1. The equations

Figure 1 shows a quadrant of a large circle of radius R. Two smaller circles of radius r have rolled (one inside and the other outside) along its circumference. Originally the centres of the rolling circles were on the x-axis, and all three were in contact at the point K. The points $P(x_1, y_1)$ and $Q(x_2, y_2)$ represent the new positions when the circles have rolled through B radians at the origin and D radians at their own centres. Arcs JP, JQ and JK are all equal. Let T be the number of turns of the rolling circle, per circuit of the fixed circle. For the epi/hypocycloid, T can be any even number.

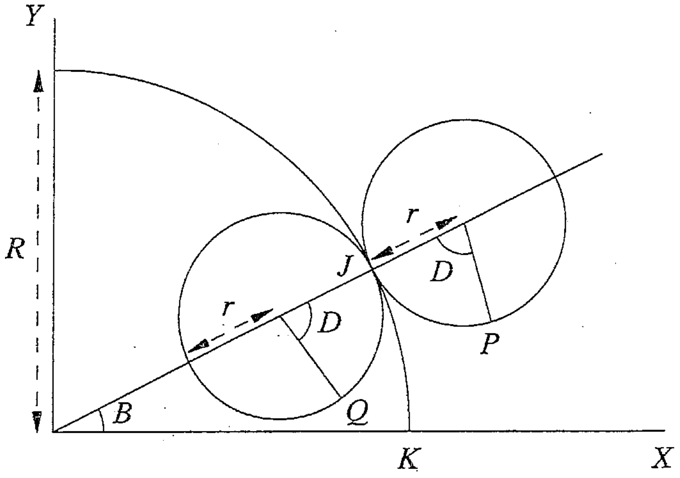


Figure 1. Rolling circles

Now $R/r = D/B = T$.
The coordinates of P are given by:

$$x_1 = (R+r)\cos B - r\cos(D+B),$$

$$y_1 = (R+r)\sin B - r\sin(D+B).$$

Similarly, the coordinates of Q are given by:

$$x_2 = (R-r)\cos B + r\cos(D-B),$$

$$y_2 = (R-r)\sin B - r\sin(D-B).$$

Inserting T as R/r and D/B gives:

$$x_1 = R\left(1+\frac{1}{T}\right)\cos B - \frac{R}{T}\cos B(T+1),$$

$$y_1 = R\left(1+\frac{1}{T}\right)\sin B - \frac{R}{T}\sin B(T+1),$$

$$x_2 = R\left(1-\frac{1}{T}\right)\cos B + \frac{R}{T}\cos B(T-1),$$

$$y_2 = R\left(1-\frac{1}{T}\right)\sin B - \frac{R}{T}\sin B(T-1).$$

Now introduce a variable H which can be switched from +1 to −1, and rewrite these equations thus:

$$x = R\left(1+\frac{H}{T}\right)\cos B - \frac{HR}{T}\cos B(H+T),$$

$$y = R\left(1+\frac{H}{T}\right)\sin B - \frac{R}{T}\sin B(H+T).$$

It will be seen that if $H = +1$ these are the coordinates of P and if $H = -1$ those of Q. If therefore H is switched at the end of each revolution of the rolling circle, it will alternate outside and inside the fixed circle.

Figure 2 presents a family of epi/hypocycloids in which $T = 24$.

2. The computer program

```
10 CLS
20 REM"EPI/HYPOCYCLOID":T = 24
30 PI = 22/7:SCREEN 2:T = 24
40 H =- 1
50 CIRCLE(300+2.4*65,100),0
60 FOR R = 65 TO 5 STEP -5
70 FOR A = 0 TO 36*T
80 B = PI*A/(18*T):C = B*(H+T)
```

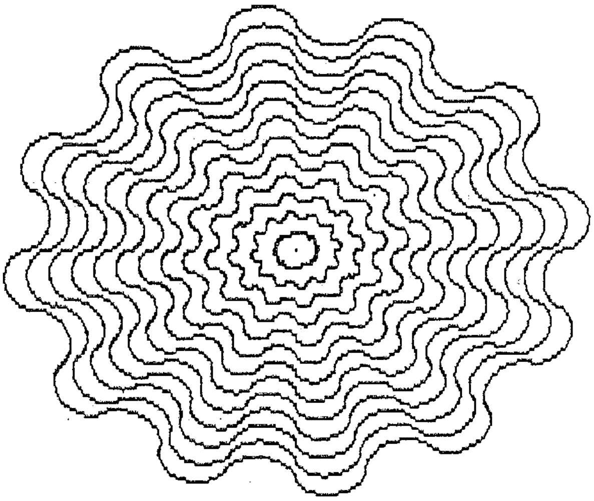


Figure 2. The epi/hypocycloid (T = 24)

```
90 X = 300 + R*(1+H/T)*COS(B) - H*R*COS(C)/T
100 X = (X-300)*2.4 + 300
110 Y = 100 + R*(1+H/T)*SIN(B) - R*SIN(C)/T
120 LINE - (X,Y)
130 IF((A MOD 36 = 0)) THEN H = H*(-1)
140 NEXT A:H = H*(-1)
150 CIRCLE(300+2.4*(R-5),100),0
160 NEXT R
```

Computer Column

MIKE PIFF

A Universal Turing Machine—the description

Our next program emulates a Turing machine. This is a machine with a finite number of states which reads from and writes to a doubly infinite tape. At some step in time, it is in state S and reads the symbol x. It changes that symbol to y, moves in the direction D and changes its state to T. The direction D can be one of L, R or S, standing for left, right or stay. A Turing machine is the most elementary model of general computation.

The program requests two files.

File 1: The machine. This file has the following format.

Line 1: Initial state.
Line 2,...,$n-1$: Transition quintuples of the form $SxyDT$.
Line n: Just a ~ by itself.

The lines in the machine file are conceptual only. The program ignores any spaces between quintuples, but the latter may be all on one line, and indeed may be contiguous. The only restriction is that *Sxy* must be contiguous; otherwise, the file is free-format. (*x* and/or *y* might be spaces, so no extraneous spaces are allowed here. But see below about invalid characters.)

File 2: The tape. A single line containing the symbols on the tape, terminated by a ~. The machine always starts on the first symbol of this tape. It is doubly infinite.

States are any ASCII symbols between ! and }, i.e., 33 to 125. Symbols are the same as states, but space `CHR(32)` is also allowed. Character 254 is interpreted as a visible space. Otherwise than this, invalid characters are simply ignored.

Note that if either filename is `CON` then input comes from the terminal. This is useful for tapes rather than machines!

As a Turing machine need never halt, nor is it possible to predict whether or not it will, interrupt the program by pressing `Q`.

An example of an input machine file is the doubler. It takes a sequence `111` and converts it into `111111`:

```
1
111L2 211L3 2␣␣L3 3␣1S3 311L4 4␣1S4 411R5
511R5 5␣␣R6 611R6 6␣␣L7 71␣S7 7␣␣L8 8␣␣Lb
811L9 911L9 9␣␣La a11La a␣␣R2 b11Lb b␣␣Rc
~
```

Another example is the monic multiplier which converts `11*111=` into `11*111=111111`:

```
1
11xR2 1**L8 211R2 2**R3 31yR4 3==17 411R4
4==R5 511R5 5␣1L6 611L6 6==L6 611L6 6yyR3
6**R7 7y1L7 7**L7 711L7 7xxR1 7␣␣R8 8x1L8
~
```

It would be easy to write a description of a machine that did more complicated arithmetical operations than these. Indeed, any calculation whatsoever that can be performed by a computer can be performed by our Turing machine.

The listing of the Turing program will appear in the next issue of *Mathematical Spectrum.*

Letters to the Editor

Dear Editor,

On a formula of Ramanujan

It might interest your readers to know that the result

$$2 = \sqrt{1+\sqrt{1+2\sqrt{1+3\sqrt{1+4\sqrt{1+5\sqrt{1+6\sqrt{1+7\sqrt{1+\cdots}}}}}}}}$$

in the article by Alexander Abian and Sergei Sverchkov (see *Mathematical Spectrum* Volume 26 Number 1, pages 8–10) is essentially available in a number of places. For example, a slight variant of it was on the Putnam examination of 1966. The earliest reference I have been able to find is a question published by Ramanujan in the *Journal of the Indian Mathematical Society* as Question 289 (III, 90). See page 329 of the *Collected Papers of Srinivasa Ramanujan* (Chelsea, New York, 1962). It has the same version as the Putnam exam and another result of the same type:

$$4 = \sqrt{6+2\sqrt{7+3\sqrt{8+\cdots}}}$$

Also of interest are remarks attributed to T. Vijayaraghavan on page 348 of the *Collected Papers* regarding the convergence of such expressions.

Yours sincerely,
H. K. KRISHNAPRIYAN
(Math/CS, Drake University,
Des Moines, IA 50311, USA)

Dear Editor,

The Smarandache function—a previous existence

Looking at Volume 26 Number 2 of *Mathematical Spectrum*, I am intrigued by the Smarandache function. This function arose in my study of polynomial functions $(\operatorname{mod} m)$ when I was a graduate student about 1965—see definition 2 in the reference.

The paper is slightly technical in order to be careful, but I can give you a less formal version. The polynomial $s_k(x) = (x+1)(x+2)\cdots(x+k)$ is congruent to zero $(\operatorname{mod} m)$ if and only if m divides $k!$, i.e. $k \geqslant S(m)$. From this it follows that, if a monic polynomial is congruent to zero $(\operatorname{mod} m)$, then its degree is at least $S(m)$. Consequently the value $S(m)$ occurs in most of the later results in the paper.

Reference

1. David Singmaster, 'On polynomial functions $(\operatorname{mod} m)$', *Journal of Number Theory* **6** (1974), 345–352.

Yours sincerely,
DAVID SINGMASTER
(Computing, Information Systems and Mathematics,
South Bank University, London SE1 0AA, UK)

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, and are eligible for annual prizes. When writing to the Editorial Office, please state your full name and also the postal address of your school, college or university.

Problems

26.7 (Submitted by Kamal Rezaei, Tehran)
Curves $C_1,\dots,C_n$ in the plane have equations $r_1 = r_1(\theta)$, ..., $r_n = r_n(\theta)$ in polar form and

$$r_1(\theta)r_2(\theta)\cdots r_n(\theta) = \text{constant}.$$

A straight line l from the origin cuts the curves at points $P_1,\dots,P_n$, respectively, and the angle between l and the normal to the curve C_i at P_i is α_i, measured so that it is positive if the normal is on one side of l and negative if it is on the other side. Prove that

$$\sum_{i=1}^{n} \tan\alpha_i = 0.$$

26.8 (Submitted by J. Rodriguez, Sonora, Mexico)
For a positive integer n, $S(n)$ is defined as the smallest positive integer such that n divides $S(n)!$ (S is called the Smarandache function: see Volume 26 Number 2, pages 39–40 and Problem 26.5, page 56. Find an infinite strictly increasing sequence of positive integers $n_1, n_2, n_3, \dots$ such that no three terms in the sequence $S(n_1), S(n_2), S(n_3), \dots$ are increasing or decreasing. (See also page 84 of this issue.)

26.9 (Submitted by Anand Kumar, B. N. College, Patna, India)
Find all three-digit numbers which are equal to the sum of their hundreds digit, the square of their tens digit and the cube of their units digit.

Solutions to Problems in Volume 26 Number 1

26.1 A collection of 1993^n positive rational numbers has the property that, if any one of them is removed, the remaining ones can be divided into 1992 equal collections whose products are equal. Prove that all 1993^n numbers are the same.

Solution by V. Fawcett, Cambridge

Put $m = 1993^n$, consider a particular prime number p, and denote the exponents of p in the m given rational numbers by $a_1,\dots,a_m$; these can be positive or negative integers or zero. We need to show that they are all the same. If not, then we can reorder the numbers so that $a_2 \neq a_1$ and $a_2 - a_1$ is divisible by the smallest power of 1992 among all $a_i - a_1$, say by 1992^r. We can now write

$$a_1 = \lambda, \qquad a_2 = 1992^r b_2 + \lambda, \qquad \dots, \qquad a_m = 1992^r b_m + \lambda,$$

where $\lambda, b_2, \ldots, b_m \in \mathbb{Z}$ and $1992 \nmid b_2$. From the hypothesis, we can divide the $a_2, \ldots, a_m$ into 1992 equal collections whose sums are equal to A, say, so we can divide $b_2, \ldots, b_m$ into 1992 equal collections whose sums are equal to $B = (A - t\lambda)/1992^r$, where $t = (1993^n - 1)/1992$. Hence

$$b_2 + \cdots + b_m = 1992B.$$

Similarly we can divide $a_1, a_3, a_4, \ldots, a_m$ into 1992 equal collections whose sums are equal, from which we obtain

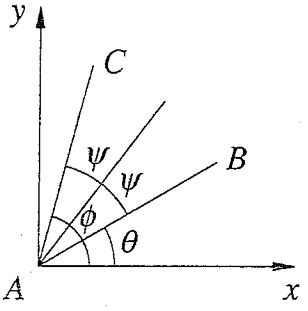
$$b_3 + \cdots + b_m = 1992C$$

for some integer C. If follows that b_2 is divisible by 1992, which is not so. Hence all the a_i's are equal.

26.2 Find an equation for the bisector of angle BAC, where A, B and C are the points representing the complex numbers a, b and c in the Argand diagram.

Solution by V. Fawcett

If AB and AC are at angles θ and ϕ to a fixed line, the x-axis, then the bisector of angle BAC is at an angle of $\theta + \psi$, where $\phi = \theta + 2\psi$. Hence the bisector is at an angle of $\frac{1}{2}(\theta + \phi)$. Thus $b = a + \lambda e^{i\theta}$ and $c = a + \mu e^{i\phi}$ for some real numbers λ and μ, and the bisector of angle BAC is given by



$$\begin{aligned} z &= a + \eta \exp \tfrac{1}{2}i(\theta + \phi) \\ &= a + \eta \exp \tfrac{1}{2}i\theta \exp \tfrac{1}{2}i\phi \\ &= a + \eta \sqrt{\left(\frac{b-a}{\lambda}\right)\left(\frac{c-a}{\mu}\right)} \\ &= a + \tau\sqrt{(b-a)(c-a)}. \end{aligned}$$

Also solved by Polly Shaw, Dame Allan's Girls' School, Newcastle upon Tyne.

26.3 Let X be a finite non-empty set for which there is a mapping $f: X \to X$ such that $f^{1993} = \mathrm{Id}_X$. Prove that the number of elements $x \in X$ such that $f(x) \neq x$ is a multiple of 1993.

Solution by V. Fawcett

If $f(x) \neq x$, consider the set $\{x, f(x), f^2(x), \ldots, f^{1992}(x)\}$. Suppose there exists an integer r between 1 and 1992 such that $f^r(x) = x$. Since 1993 is prime, r and 1993 are coprime, so there exist integers s and t such that $rs = 1993t + 1$. Then $x = f^{rs}(x) = f^{1993t+1}(x) = f(x)$, which is false. Similarly, if $f^r(x) = f^u(x)$ for some r and u ($1 \leqslant u < r \leqslant 1992$), then $f^{r-u}(x) = x$, contradicting what has just been proved. Hence the set has 1993 distinct numbers all of which clearly satisfy $f(y) \neq y$. We can thus partition X into subsets; firstly the subset consisting of all elements x such that $f(x) = x$, then subsets constructed as above. If there are n of these, then exactly $1993n$ members of x have the property that $f(x) \neq x$.

Note that 1993 can be replaced by any prime number.

Also solved by Sinefakopoulos Achilleas (University of Athens) and S. Pipinos (University of Bristol).

Reviews

Exploring Mathematics with Your Computer. By ARTHUR ENGEL. Mathematical Association of America, Washington DC, 1993. Pp. ix+301. Paperback $38.00 (ISBN 0-88385-636-0).

To quote from the preface: 'This is a mathematics not a programming book. It is intended for students, mathematics teachers and mathematicians who are just starting to explore mathematics on their own computers. In studying it, and especially in working through the exercises, readers will get to know many new, elementary topics and learn as much from the extensive exercises as from the examples.

We use Turbo Pascal ... easily picked up by readers as they work their way through the examples and exercises. ... Most of our programs are complete. Each serves as a reading exercise to help you learn Pascal; for this reason the programs are short and, for the most part, comprehensible without comment. Yet, many are not trivial!

The book contains 65 topics ... these are independent of each other. If you know the rudiments of Pascal you can tackle them in any order; if not, you are advised to cover them in sequence since new features of Pascal are introduced when needed.'

A 3.5-inch IBM-compatible disc containing the Pascal programs described in the book is packaged with this volume. I have never used Pascal before, but I managed to get the programs running (using Turbo Pascal Version 6.0) and also to write my own programs for many of the exercises in the text without difficulty and without having to refer to a Pascal Guide or any other source for help; the author is a master of didactics.

The author's style is pleasant, clear and precise, and his choice of material is superb. This includes: linear recursion, Euclid's algorithm, Gill's GCD and LCM algorithm, all representations of n in the form x^2+y^2, Pythagorean triples, Olympiad problems, partitions, the money changing problem, primes, empirical study of the Goldbach conjecture, representation of n as a sum of four squares, finding geometric probabilities by simulation, random sequence generation by cellular automata, the two sample problem of Wilcoxon, the binomial and hypergeometric distributions, sorting, the kth smallest element of an n-set, Archimedes' integration of the parabola, extrapolation to the limit and Romberg integration, one thousand decimals of e, coupled differential equations, continued fractions, self-avoiding random walks.

Professor Engel's examples and algorithms are real gems. Each is preceded by a detailed presentation of the underlying mathematical theory, leading logically to a complete program listing. The exercises are fascinating and the solutions are masterpieces. I would like to include several typical examples but I must restrict myself to one.

> Show that if a, b and q are positive integers with $a^2+b^2 = q(ab+1)$, then q is a perfect square.

This is a very difficult problem (set at the 29th International Mathematical Olympiad in Canberra, 1988): it defied the attempts of the four most eminent number theorists in Australia, but with the aid of a computer this problem becomes, as Professor Engel shows, comparatively simple.

I have spotted only five trivial misprints. On page 12: Ex. 3 (d), the answer should be 4.493 409 458; Ex. 4, one of the roots is given to be 3.472 963 5529 instead of 0.347 296 355 29. On page 52: in the given table $f(4) = 6$ and not 4. On page 252: Ex. 27, a left square bracket is missing from the x array. On page 270: first line, instead of s write sum.

This book is a gold mine and I will return to it many times to dig out more nuggets. Congratulations, Professor Engel, for one of the best texts I have seen; it is highly recommended.

Medical student, University of Newcastle upon Tyne GREGORY D. ECONOMIDES

The Wohascum County Problem Book. By GEORGE T. GILBERT, MARK I. KRUSEMEYER AND LOREN C. LARSON. Mathematical Association of America, 1993. Pp. ix+233. Paperback (ISBN 0-88385-316-7).

This book contains 130 problems intended to challenge undergraduate mathematicians. It contains the problems, solutions and an appendix listing prerequisites for the problems. The major part of the book is taken up by the solutions. The problems are arranged as far as possible in order of difficulty rather than according to the prerequisites required. Thus there are a few quite difficult problems accessible to sixth formers but several easier ones for which they would not have the necessary knowledge.

For problems like these to be successful, I feel that they should be attractive, in that a student can hardly resist attempting them. Rather than involve relatively straightforward applications of theorems or tedious calculations, the solutions should require some conceptual insight and encourage students to look at ideas from different points of view. This is, of course, not easy to achieve and I feel that this book is only partly successful.

Many of the problems are exercises in pure mathematics. Some of them did satisfy the above criteria but many did not. For example, Problem 40, which requires the solver to prove the existence of a positive number λ with

$$\int_0^\pi x^\lambda \sin x \, dx = 3,$$

is just an exercise in the intermediate value theorem together with a proof of continuity. On the other hand, Problem 16, which asks for the number of solutions to $e^x = x^n$ in terms of n, could provoke some discussion and lead to a greater understanding of the functions e^x and x^n.

The 'Wohascum County' context, when used, did not succeed in gripping my imagination. I have seen more entertaining scenarios in puzzles set in newspapers and magazines. Babe Ruth's batting average and slugging percentage, moreover, may be of interest to male Americans, but the problem about them was completely incomprehensible to me. I was sad to note the complete absence of female contributors.

The solutions are given in reasonable detail and should prove helpful to most readers although some might prove quite hard for less mathematically experienced students to understand. It is good that different solutions to the same problem are sometimes given. Occasionally an 'idea' which might serve as a hint precedes a solution.

On the whole I was a little disappointed in the book, essentially because I found too few of the problems really interesting.

University of Sheffield CAMILLA JORDAN

Groups. By MARK CARTWRIGHT. Macmillan, Basingstoke, 1993. Pp. 123. Hardback £8.50 (ISBN 0-333-54300-9).

This book is intended to provide an accessible introduction to group theory for undergraduates meeting the subject for the first time. It was the aim of the author to avoid abstraction as much as possible, since it was his view that abstraction is a stumbling-block for many people, and much of the book is devoted to examples and applications.

Unfortunately the word 'group' does not appear until page 65 of a book 123 pages long (despite the protestations of the index to the contrary), with the first 64 pages containing a long-winded, and frankly rather boring, derivation of the group axioms. The main problem is the author's insistence on providing examples for everything; for instance, in order to provide an application of addition, the author writes 'If I have three apples, and you give me two more apples, I end up with five apples' (page 1). While this is undeniably true, it is somewhat off-putting to find the first eight pages written in terms of apples, especially for someone reading mathematics at university.

Having survived the first three chapters, however, the rest of the book is of considerably better quality. It is a pity that closure is not specifically stated as a property of a group, but the various examples given show that this is required. Caley tables are introduced as analogous to the multiplication table—the fact that they are all Latin squares, which is one of the first results proved, is a useful one for the next part of the book. Isomorphism is well defined, with plenty of examples given.

The next section of the book contains various simple results—the definition of 'powers' and 'generators' leading naturally to the concept of a cyclic group. Much long-windedness could have been avoided by the use of the concept of the greatest common divisor, but this would involve referring to mathematics not necessarily known by the reader. Subgroups are introduced, although there is no mention of Lagrange's theorem, which is left to more advanced texts. This is followed by a chapter on group classification, which is in fact a listing of all possible groups up to order 5. The final chapter is a practical and useful application, using the symmetry group of an object to count the number of ways of colouring it with several colours. There are 2226 different six-colourings of a cube.

In all, the last four chapters of this book are a good introduction to the theory, but it is surprising to see no mention of rings or fields, which could have been introduced as further axioms very shortly after the group axioms are defined. Vector spaces are also conspicuous by their absence.

This is a reasonable textbook, but a very introductory one. Sawyer's *Concrete Approach to Abstract Algebra* remains a far better and wider-ranging introduction to the theory. It is not really a university-level book, but it would be a good introduction to group theory for those about to start a Further Mathematics course.

Sixth form Winchester College PAUL PHILLIPS AND THOMAS WOMACK

Other books received

Generalized Vector and Dyadic Analysis. By CHEN-TO TAI. IEEE Press, New York, 1992. Pp. x + 134. Hardback (ISBN 0-87942-288-2).

Algebra. By T. T. MOH. World Scientific Publishing Co, Singapore, 1992. Pp. viii + 350. Hardback £48 (ISBN 981-02-1195-3), Paperback £23 (ISBN 981-02-1196-1).

A comprehensive textbook covering pretty well all algebra that appears in an undergraduate mathematics course.

Advanced Level Mathematics. By R. C. SOLOMON. DP Publications Ltd, London, 1992, second edition. Pp. xvii + 599. Paperback £9.95 (ISBN 1-873981-24-4).

The second edition of an advanced-level textbook with around 2000 exercises and examination-style questions and answers, now with a companion revison course.

Advanced Level Mathematics Revision Course. By R. C. SOLOMON. DP Publications Ltd, London, 1993. Pp. vi + 165. Paperback £4.95 (ISBN 1-85805-041-3).

Business Mathematics and Statistics. By A. FRANCIS. DP Publications Ltd, London, 1993, third edition. Pp. vi + 516. Paperback £9.95 (ISBN 1-85805-003-0).

The third edition ofa manual for students who need a grounding in mathematics in finance. No previous knowledge is assumed.

Finite Difference Equations. By H. LEVY AND F. LESSMAN. Dover, New York, 1992. Pp. vii + 278. Paperback £7.95 (ISBN 0-486-67260-3).

This is a republication of a book first published in 1961 and is intended for advanced undergraduate or postgraduate readers.

Greek Mathematical Thought and the Origin of Algebra. By JACOB KLEIN. Dover, New York, 1992. Pp. xv + 360. Paperback £9.95 (ISBN 0-486-27289-3).

This is a republication of a book first published in German in the 1930s and first published in English in 1968. It is intended for serious readers and scholars in the history of mathematics.

CONTENTS

ISSN 0025–5653

PRICES *(postage included)*
Prices for Volume 26 (Issues Nos. 1, 2, 3 and 4)

£7.70 or US$12.30 or $A16.20

These prices apply to subscribers in all parts of the world. There is now only one set of sterling rates applicable to all countries.

A discount of 10% will be allowed on all orders of five or more copies of Volume 26 sent to the same address.

Details of reduced prices for two- and three-year subscriptions available on request.

Back issues

Most back issues are available; information concerning prices and a list of the articles published 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, UK.

Published by the Applied Probability Trust
Typeset by The Pi-squared Press, Nottingham, UK
Printed by Galliard (Printers) Ltd, Great Yarmouth, UK