Mathematical Spectrum

Volume 11 1978/79 *Number* 1

A Magazine of *Contemporary Mathematics*

Published by the *Applied Probability Trust*

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 11 of *Mathematical Spectrum* will consist of three issues, the second of which will be published in January 1979, and the third in May 1979.

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.

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.

J. E. Littlewood (1885–1977)

H. BURKILL
University of Sheffield

The great C. F. Gauss (1777–1855) was aptly called 'the prince of mathematicians'. Today mathematics is too vast a subject for any one man to be pre-eminent in all its branches. However, J. E. Littlewood, who died a year ago at the age of 92, was, by common consent of his fellow practitioners, the prince of analysts. Although he was one of the most remarkable mathematicians of the century and certainly the most powerful analyst of modern times, his name is largely unknown to sixth-form mathematicians and to the educated public at large. There are several reasons for this. In the first place, the type of mathematics on which Littlewood worked is such that its character, let alone its substance, cannot be conveyed in a few short phrases. Secondly, Littlewood was interested in difficult individual *problems* rather than in *theories*, so that his name did not become associated with entire subjects, in the way that Newton's is linked with mechanics or Einstein's with relativity. Thirdly, though Littlewood was an enthusiastic teacher, he did not encapsulate his experience in undergraduate textbooks. Finally, men who are particularly distinguished in their profession are often drawn into public affairs, but Littlewood avoided commitments of this kind. It was symptomatic of the life he favoured that he spent practically all his adult years in Cambridge, the last 64 of them in the same set of rooms in Trinity College. All this might conjure up the image of a dry and retiring man. In fact, nothing could be further from the truth, for Littlewood had an exuberant personality, with a most astonishing vitality.

John Edensor Littlewood was born in Rochester on 9 June 1885. The Littlewoods had been north-country farmers for many generations (the family name of Edensor having been derived from an ancestress who came from the Derbyshire village of that name on the Duke of Devonshire's estate of Chatsworth). However, more recently the Littlewoods had turned to the professions, and JEL's father, a schoolmaster, was 9th Wrangler† in 1882. Many years later, in 1937, father and son were the joint authors of a mathematical paper. From 1892 to 1900 the family lived in South Africa, where the father was a headmaster. Before he left South Africa JEL had begun to attend Cape University (at the age of 14), but, when he returned to England, he became a pupil at St. Paul's School, London. There he was taught by F. S. Macaulay, a research mathematician who was eventually elected a Fellow of the Royal Society. In December 1902 JEL took the Entrance Scholarship Examination for Trinity College, Cambridge, and, amazingly enough, was awarded no more than a Minor Scholarship. However, he redeemed himself two terms after coming up to

† A student placed in the First Class of the Mathematical Tripos in Cambridge was, and still is, called a Wrangler. Until 1910 the list of Wranglers was arranged in order of merit; since then it has been alphabetical.

Trinity when he was top in the examination for Senior Scholarships. Littlewood recalls that, on this occasion, he was at the peak of his form: he had worked hard during the Michaelmas term, but had more or less given up mathematics for rowing during the Lent term. (It was Littlewood's abiding conviction that mathematicians ought not to work excessively long hours; in his opinion five hours a day produced the best results. This may well be correct for someone working at his intensity; lesser mortals should not be too sure that it applies to them as well.) At the end of his second year—a year earlier than normal—he took Part I of the Mathematical Tripos and was bracketed Senior Wrangler with a man who, though a very good mathematician, achieved nothing like Littlewood's eminence. A year later, in 1906, Littlewood took Part II of the Tripos, and he began his research immediately afterwards.

Littlewood's tutor and director of studies was E. W. Barnes, who later was appointed Bishop of Birmingham during the first Labour government's term of office (1924) and was to become known as the 'red bishop'. He suggested a problem in complex function theory, and, when that had been successfully disposed of, he proposed that Littlewood should prove a conjecture known as the 'Riemann Hypothesis'. This renowned problem is not too difficult to enunciate.

Ordinary undergraduate analysis shows that the series

$$\sum_{n=1}^{\infty} \frac{1}{n^z}$$

converges when Re z, the real part of the complex number z, is greater than 1.† An undergraduate course might also include the fact that there exists a unique function $\zeta(z)$, which is defined and very well behaved in the whole complex plane except at the point 1, such that

$$\zeta(z) = \sum_{n=1}^{\infty} \frac{1}{n^z}$$

when Re $z > 1$. This function ζ is called the 'Riemann zeta function'. The formula

$$\zeta(z) = \prod_p \left(1 - \frac{1}{p^z}\right)^{-1} \qquad (\text{Re } z > 1),$$

where the infinite product ranges over all the primes $p = 2, 3, 5, 7, 11, \ldots$, has been known for a long time and links the zeta function to the study of primes. The function has a certain amount of symmetry about the line Re $z = \frac{1}{2}$ and the famous (or notorious) Riemann hypothesis is that, if Re $z > 0$ and $\zeta(z) = 0$, then Re $z = \frac{1}{2}$. The hypothesis is important because its proof would immediately establish many very interesting results in number theory. One of these concerns $\pi(x)$, the number of primes less than or equal to x.

† When x and y are real, n^{x+iy} is defined as $e^{(x+iy)\log n} = n^x(\cos(y \log n) + i \sin(y \log n))$, so that $|n^{x+iy}| = n^x$.

In 1896 it was proved by J. S. Hadamard (1865–1963) and Ch. de la Vallée Poussin (1866–1962), working independently, that

$$\pi(x)\bigg/\frac{x}{\log x} \to 1 \quad \text{as} \quad x \to \infty,$$

i.e. that $\pi(x)$ is about $x/\log x$ for large values of x. (Here 'log' denotes the natural logarithm.) This is the celebrated 'Prime Number Theorem'. However, an even better approximation to $\pi(x)$ is the 'logarithmic integral'

$$\operatorname{Li} x = \int_2^x \frac{1}{\log t}\,dt,$$

and an important problem is to estimate the difference $E(x) = \pi(x) - \operatorname{Li} x$. It is, for example, known that, if m is any positive number, then

$$|E(x)| < \frac{x}{(\log x)^m}$$

for all sufficiently large values of x. But, if the Riemann hypothesis is true, then $|E(x)|$ can be proved not to exceed a constant multiple of the much smaller function $x^{1/2}\log x$.

There had been no dearth of continental mathematicians trying and failing to prove the Riemann hypothesis. But England was then mathematically a backwater and Barnes may not have realised what kind of suggestion he had made to his 21-year-old pupil. After a week's hard work Littlewood, however, came to appreciate the dimensions of the problem he had taken on. He did not succeed in proving the Riemann hypothesis, and no one else has done so up to now. Indeed, ultimately Littlewood inclined to the belief that the hypothesis might be false.

Although Littlewood's work on the Riemann zeta function did not yield the prize he first hoped for, it produced other results and it marked the beginning of his independent career as a mathematician. From then (1907) until 1972 hardly a year went by without the appearance of several papers of which he was sole or joint author. In the autumn of 1907 he began a three-year stint as a lecturer in Manchester. Littlewood recalled this as a term of genuine hard labour. The lecture load was more than twice as heavy as is now customary. It also fell to Littlewood to introduce a new course of rigorous analysis. Under present conditions this is a task that any competent mathematician can undertake, but at that time there were no suitable textbooks and there was no standard development, so that the lecturer could never be sure that he would not suddenly need a result that he ought to have proved weeks before. Littlewood's research output appeared to continue unabated, but it was largely the result of work done during the vacations. Nevertheless there seems to have been time for relaxation, for Littlewood used to say that he and Marie Stopes were the best dancers in Manchester. Walking and mountaineering were other recreations to which he was always devoted. Later in life he became an accomplished skier, but he did not take up this sport until he was 40.

In 1910 Littlewood returned to Cambridge and among the results he obtained during the next few years two are particularly famous. One deals with the function $E(x) = \pi(x) - \operatorname{Li} x$, in any discussion of which the Riemann hypothesis figures so prominently. All the numerical evidence available then (or now) suggests that $\operatorname{Li} x$ always exceeds $\pi(x)$. However Littlewood showed that, in fact, $\pi(x) - \operatorname{Li} x$ changes sign infinitely often; thus, in particular, $\pi(x) > \operatorname{Li} x$ for some x. The proof of this spectacular result did not provide a numerical estimate of any x for which $\pi(x) > \operatorname{Li} x$, but some 40 years later S. Skewes, a pupil of Littlewood's, showed that the least x does not exceed

$$10^{10^{10^{964}}}.$$

This number has now been whittled down to 10^{1166}. (By way of comparison: according to Eddington the universe contains 10^{79} protons.)

The second result concerns quite a different topic, namely a generalization of the notion of convergence. A number of 18th-century mathematicians, notably Euler, assigned 'sums' to divergent series. For instance the series

$$1 - 1 + 1 - 1 + \cdots \tag{1}$$

was given the 'sum' $\frac{1}{2}$. The argument was that, for $-1 < x < 1$, the power series

$$1 - x + x^2 - x^3 + \cdots \tag{2}$$

converges (in the usual sense) to the sum $1/(1 + x)$; and, as $x \to 1$,

$$\frac{1}{1+x} \to \frac{1}{2},$$

while each term of the series (2) tends to the corresponding term of the series (1). Similarly, since

$$1 - 2x + 3x^2 - 4x^3 + \cdots = \frac{1}{(1+x)^2} \quad \text{for} \quad -1 < x < 1$$

and

$$\frac{1}{(1+x)^2} \to \frac{1}{4} \quad \text{as} \quad x \to 1,$$

the series

$$1 - 2 + 3 - 4 + \cdots \tag{3}$$

was given the 'sum' $\frac{1}{4}$. Euler went on to use 'sums' of this kind for the derivation of various interesting formulae, including one that involves the zeta function $\zeta(z)$. During the early 19th century mathematical analysis was given a secure foundation, notably by A. L. Cauchy (1789–1857). To him the sum of a series meant the limit of the partial sums, and if the partial sums did not converge (as in (1) and (3)), then the series simply did not have a sum and was called 'divergent'. The attitude to divergent series of the new school of rigorous analysts was summed up rather forcibly by the great Swedish mathematician N. H. Abel (1802–1829): 'Divergent series are the invention of the devil, and it is shameful to base on them any demonstration whatsoever'. However, by the end of the 19th century it began to be realised that Euler's treatment of divergent series was perfectly admissible provided that one was clear what was meant by the term 'sum'; and the key to the new approach was the following theorem of Abel himself: *If the series* $\sum_{n=0}^{\infty} a_n$ *converges to sum* s, *then* $\sum_{n=0}^{\infty} a_n x^n$ *converges for* $-1 < x < 1$ *and*

$$\sum_{n=0}^{\infty} a_n x^n \to s \quad \textit{as } x \to 1 \textit{ from below.} \tag{4}$$

Suppose now that a series $\sum_{n=0}^{\infty} a_n$ (which need not converge) is such that $\sum_{n=0}^{\infty} a_n x^n$ converges for $-1 < x < 1$ and (4) holds. Then $\sum_{n=0}^{\infty} a_n$ is said to be *Abel summable*, and s is called its *Abel sum*. Thus the divergent series (1) and (3) have Abel sums $\frac{1}{2}$ and $\frac{1}{4}$, respectively. Moreover Abel's theorem shows that a series which converges to sum s is also Abel summable to s, so that Abel summability provides an extension of the notion of convergence.

The series (1) and (3), which are Abel summable but divergent, have relatively large terms. The question arises whether there are any Abel summable series with small terms which are nevertheless divergent. The answer is 'Not if the terms are small enough'. For it is fairly easy to show that, if $na_n \to 0$ as $n \to \infty$, and the series $\sum_{n=0}^{\infty} a_n$ is Abel summable, then it must be convergent. What Littlewood proved is that the weaker condition

$$|na_n| < K \quad \text{for all } n, \tag{5}$$

where K is an arbitrary constant, also ensures that Abel summability implies convergence. This result was very difficult to establish, but does not, at first sight, represent a very startling improvement on what had gone before. However, Littlewood had reached the ultimate truth; for he also showed that the constant K in (5) cannot be replaced by any function $\phi(n)$ tending to ∞ (however slowly) as $n \to \infty$.

The problem on Abel summability had been suggested to Littlewood by G. H. Hardy (1877–1947), another outstanding Trinity mathematician. The two men's mathematical interests were very similar, but their personalities were poles apart. Littlewood's ebullient but easy-going nature contrasted with Hardy's asceticism and devotion to passionately held beliefs. It was Hardy who was mainly responsible for the reform of the Mathematical Tripos at Cambridge. The old-fashioned syllabus with its emphasis on the high-speed solution of technically difficult problems was replaced by modern mathematics and an examination system intended to test understanding rather than manipulative skill. (Hardy also advocated the abolition of classes for the successful examinees, but he was able to secure only the suppression of the Wranglers' order of merit.) For the guidance of lecturers as well as students Hardy wrote a textbook (*A Course of Pure Mathematics*, Cambridge, 1908) whose twelfth edition is still in print. Its style, that of 'a missionary preaching to cannibals' (in Littlewood's phrase), will always make it a delight to read.

It is very natural that Hardy and Littlewood, living in the same college, should have collaborated in mathematical research. However, the extent of their cooperation is astonishing and without the remotest parallel in the history of mathematics or science. In 1913 five papers appeared under their joint names, and in all over a hundred were written during the course of a collaboration that lasted until Hardy's death in 1947. And all this time both Hardy and Littlewood were producing a great quantity of research on their own or with other mathematicians. Moreover, between 1919 and 1931 Hardy taught in Oxford and during these years he and Littlewood communicated largely by correspondence. Yet this situation was actually congenial to the two men, each of whom was anxious to preserve his independence and complete freedom of action. They evolved the following four principles as the basis of their relationship.

1. When one wrote to the other, he did not have to worry whether what he wrote was right or wrong.
2. When one received a letter from the other, he was under no obligation to read it, let alone answer it.
3. Although they could work on the same detail simultaneously, it was preferable that they should not do so.
4. It did not matter in what proportion they contributed to any particular paper.

Hardy and Littlewood worked on a great variety of topics. One of the principal areas for their investigations was the Theory of Numbers, and a memorable series of papers is entitled 'Some Problems of Partitio Numerorum'. It deals with a subject initiated by the minor 18th-century mathematician Edward Waring. In a book published in 1770 he stated without proof that every positive integer is the sum of at most 4 squares, 9 cubes, 19 fourth powers, 'and so on'. In other words, Waring claimed that, given the integer $k \geqslant 2$, there exists a least integer $g(k)$ such that every positive integer is the sum of at most $g(k)$ kth powers of positive integers; and that, moreover, $g(2) = 4$, $g(3) = 9$, $g(4) = 19$. J. L. Lagrange (1736–1813), in all probability unaware of Waring's book, proved the assertion for $k = 2$ in the year in which it was made. Though the proof was by no means easy, larger values of k present far greater obstacles. It was not, in fact, until 1909 that D. Hilbert (1862–1943) finally established the existence of $g(k)$ for all values of $k \geqslant 2$. Previously, in addition to Lagrange's result, there had been proofs of the existence of $g(k)$ for $k = 3, 4, 5, 6, 7, 8, 10$ and of the identity $g(3) = 9$. What still remained was the problem of determining the actual value of $g(k)$ for $k \geqslant 4$. However, of even greater interest is the number $G(k)$ defined as the least number s such that *all but a finite number* of integers are sums of at most s kth powers. Thus $G(k) \leqslant g(k)$ and, although $G(2) = g(2) = 4$, it is known that $4 \leqslant G(3) \leqslant 7 < g(3)$. (The determination of the lower bound is quite elementary, while that of the upper bound is a formidable affair. Numerical evidence suggests that only fifteen integers, the largest of them 8042, require 8 cubes; it has been proved that 23 and 239 are the sole integers requiring 9 cubes.) Hardy and Littlewood devised a powerful new method which yielded the general estimate

$$G(k) \leqslant (k-2)2^{k-1} + 5 \qquad (6)$$

as well as a number of particular inequalities better than this. The Russian mathematician I. M. Vinogradov later refined the Hardy–Littlewood method and showed that, for large k, $G(k)$ is of the order $k \log k$; this result is, of course, very much better than (6). Hardy and Littlewood were also interested in the *number* of representations of a given integer n as the sum of a given number r, say, of kth powers. Their results are remarkably precise, but rather too complicated to be given here.

A question that is often asked is whether the Hardy–Littlewood partnership contained a dominant member. When it is put in this extreme form the answer is certainly 'No'. But Hardy, with cool objectivity, has expressed his satisfaction at having been able to collaborate with Littlewood 'on something like equal terms'. In 1940 Hardy was awarded the (triennial) Sylvester medal of the Royal Society, and Littlewood was the next recipient, in 1943. On that occasion the President of the Royal Society summed up Littlewood's special qualities in words that cannot be bettered: 'He is the man most likely to storm and smash a really deep and formidable problem; there is no one else who can command such a combination of insight, technique and power'. In 1958 Littlewood also received the Copley Medal, the Royal Society's highest honour.

The First World War showed up some of the differences between Hardy and Littlewood. Hardy was an ardent pacifist and remained in Cambridge campaigning on behalf of like-minded men (such as Bertrand Russell) who had fallen foul of intolerant public opinion. Littlewood, on the other hand, though a supporter of 'liberal' causes, joined the Royal Artillery and from there found his way into the Ballistics Office. Incidentally, through cheerful indifference (as C. P. Snow has put it) he ended his military career as he had begun it: a Second Lieutenant. Throughout the war Hardy and Littlewood continued to collaborate, though the rate of their output was naturally reduced. However, yet another difference between the two men was that, whereas Hardy revelled in the practical uselessness of all his research, Littlewood was genuinely interested in ballistics as well as other aspects of applied mathematics and actually published several papers on these subjects.[†] In a fascinating little book, *A Mathematician's Miscellany*, which Littlewood published in 1953, an entire chapter is devoted to ballistics. A paragraph there throws new light on the battle of the Falkland Islands (December 1914). A squadron of German

[†] Some years ago Littlewood was asked if he would contribute to *Mathematical Spectrum*. He readily agreed, and decided that ballistics would be a suitable subject. Unfortunately the resulting two articles (Vol. 4, Nos. 1 and 2) turned out to be far too erudite. Clearly Littlewood had the same exalted ideas of sixth-formers' knowledge as Macaulay (who was convinced that 'every schoolboy knows who imprisoned Montezuma, and who strangled Atahualpa').

cruisers was attacked by a British force which was not only faster, but also had heavier guns. The British commander naturally fought the engagement at extreme range in order to minimise his casualties. Several hours, however, were required to sink the German ships. Historians have commented critically on this seemingly poor performance, but Littlewood provides the explanation. A naval officer who was present at the battle reported that, inexplicably, the British salvoes kept falling 100 yards to the left of their targets; and Littlewood makes the point that this distance is accounted for by the rotation of the earth. Apparently British gunsights *were* calibrated to take account of that factor, but the tacit assumption was that naval battles would be fought at a latitude of 50 °N. Since the Falklands are 50 ° south of the equator, a double error was introduced. In his *Miscellany* Littlewood also has a less serious item about his Ballistics Office days. He had written a report which ended with the words 'Thus σ should be made as small as possible'. This sentence did not appear in the printed minute. Instead there was a tiny speck which turned out to be the smallest σ the printers could lay their hands on.

In 1916, at the age of 30, Littlewood was elected a Fellow of the Royal Society; and in 1928 he became the first holder of the newly endowed Rouse Ball chair of Pure Mathematics in Cambridge. His principal mathematical activity during the inter-war years was his collaboration with Hardy, and a slight indication of its fruits has already been given. A beautiful theorem proved by him alone, and dating from 1930, is the following:

If the numbers a_{ij} are such that, for some constant K,

$$\left|\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} a_{ij}x_iy_j\right| \leqslant K \quad \textit{whenever} \quad |x_i| \leqslant 1 \quad \textit{and} \quad |y_j| \leqslant 1\ (i,j = 1, 2, \ldots), \qquad (7)$$

then the series $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} |a_{ij}|^{4/3}$ converges; moreover the result is best possible in the sense that, if, as $i,j, \to \infty$, $r_{ij} \to \infty$ (however slowly), then there exists a set of a_{ij} satisfying (7), *but such that $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} r_{ij}|a_{ij}|^{4/3}$ diverges.*

This theorem, which is complete in itself but also fits into a larger theory, has the quality of unexpectedness that is characteristic of so much of Littlewood's work.

Littlewood had a very large number of research students, but he also did his full share of undergraduate teaching. His single publication that resembles a textbook was, in fact, written as a set of lecture notes. It is called *The Elements of the Theory of Real Functions* (Heffer, 1926) and contains a mixture of topics from set theory. In the preface he commends the practice of providing lecture notes in advance of the course and remarks 'it is possible that the art of lecturing has not yet recognized the full importance of the younger invention of printing'. On the other hand, he warns prospective students not to skip his lectures since these help to convey the point of the subject matter by providing the necessary 'provisional nonsense' which 'would appear ridiculous if permanently enshrined in print'. However, 50 years after the *Elements*, printed (or duplicated) lecture notes are still the exception rather than the

J. E. Littlewood at the age of 80.
(Photograph reproduced by permission
of the London Mathematical Society.)

rule. University teachers who have experimented with them will probably agree that, in this respect at least, students are quite as conservative as they themselves are reputed to be.

Official retirement in 1950 only served to increase Littlewood's activity. Physically he was as strong as ever. He kept up the habit of the long daily walks on which, he often recalled, so many mathematical ideas came to him. He also continued to go for strenuous holidays in England and Switzerland. In addition he now accepted invitations to America, where his pungent personality was much appreciated. The quantity and quality of his mathematical research remained unabated until he was in his mid-eighties. In 1960, when he was 75, he was awarded a major prize by the London Mathematical Society for two papers he had published in the previous year. About the same time he submitted a paper to an electronics journal. When notifying him of the acceptance of the paper, the editor asked Littlewood his age in case he was eligible for a prize for authors under the age of 30.

Littlewood had never bothered to obtain a doctorate from his university. So, on his 80th birthday, the University of Cambridge conferred on him the honorary degree of Sc.D. His 90th birthday was celebrated with a day of mathematical lectures, most of which he attended, and a dinner in Trinity College, at which he made a witty and incisive speech. During these festivities a (much younger) guest said to him 'I hope to be present at your 100th birthday celebrations'. Littlewood looked at him quizzically and replied 'I don't see why you shouldn't; you look quite healthy to me'. In the winter of 1976 he spent his usual holiday in Switzerland. However, during the following summer he fell in his rooms and had to enter a nursing home. On the morning of 6 September a nurse went to his room to ask him what he wanted for breakfast. He ordered sausages, but, when the nurse returned with them, he was dead. It was an end that would have appealed to his sense of humour.

Probability as an Aid in Social Research: the Randomised Response Technique

J. R. ALEXANDER
University of Southampton

1. Introduction

One problem which plagues the social researcher is the difficult task of getting people to co-operate in answering questions on delicate personal matters. Another problem is to get them to answer such questions truthfully. These problems are particularly difficult when a survey is conducted by interview, rather than by a postal questionnaire. The respondent has to feel that he can trust the interviewer, a stranger, with highly personal information which may not only be embarrassing but may incriminate him legally or morally.

A number of techniques have been used in attempts to solve these problems. Barton (1958) (reference 1) gives an amusing and succinct account of some of the more commonplace ideas. Belson (1968) (reference 2) describes a study of stealing among schoolchildren, in which the researchers went to great lengths to win the confidence of the respondents.

In recent years a great deal of attention has been focussed on the randomised response technique. Starting with the notion that if you want to discover information about the population, you do not essentially need to have *full* information about every individual in your sample, Warner (1965) (reference 4) proposed a new procedure involving the use of probability. The modification outlined below is described by Horvitz et al. (1967) (reference 3) and is known as the Simmons technique.

2. The basis of randomised response

The central idea is that the respondent answers one of two questions, one totally innocuous, the other being the embarrassing item. The interviewer does not know to which question the stated answer refers. The respondent chooses which question to answer by carrying out a 'random trial', such as tossing a coin, throwing a die, cutting a deck of cards, etc., without letting the interviewer know the result. For example, the interviewer may ask the respondent to toss a coin. If the result is a head he is to answer the question

'Have you stolen from work?'

while if it is a tail, the question is

'Were you born in April?'

Probability theory then enables the researcher to estimate the proportion of people in the sample who have stolen from work; if the sample is random, this will give an estimate for the population from which it was drawn. To see how, we first need to write down the probability that a 'Yes' result occurs.

Let p be the probability that a respondent has stolen from work, and suppose that 1/12 of the population sampled were born in April. Putting λ as the probability of a 'Yes' we have

$$\begin{aligned}\lambda = P(Y) &= P(Y \cap H) + P(Y \cap \bar{H})\\ &= P(Y|H)P(H) + P(Y|\bar{H})P(\bar{H})\\ &= p \cdot \tfrac{1}{2} + (\tfrac{1}{12}) \cdot \tfrac{1}{2}\end{aligned}$$

where H is the event 'Head' and Y is the event of a response 'Yes', and $P(Y|H)$ denotes the probability that Y occurs, given that H occurs. Rearranging, we have

$$p = 2\lambda - \tfrac{1}{12}. \qquad (1)$$

We can estimate λ by the proportion of 'Yes' responses in the sample interviewed. From this p can be estimated using Equation (1).

3. A practical trial

We carried this out with a class of 112 young people. Two sensitive questions were asked, the coin being tossed independently for each question:

(1) For a head: 'I have consumed alcohol in a pub while under the legal age' True or False?

For a tail: 'My last birthday fell on a Sunday' True or False?

(2) For a head: 'I am not very interested in the opposite sex' True or False?

For a tail: 'My last birthday fell on a Sunday' True or False?

The results were: for Question 1, 59 True's; for Question 2, 11 True's. For these questions the equation corresponding to (1) is

$$p = 2\lambda - \tfrac{1}{7}.$$

The estimated values of λ were $59/112 = 0{\cdot}57$ and $11/112 = 0{\cdot}0982$ leading to estimates $\hat{p}_1 = 0{\cdot}91$ for Question 1 and $\hat{p}_2 = 0{\cdot}054$ for Question 2.

4. The precision of the estimates

These estimates of p are subject to sampling error. If we are only interested in the group of 112 people, then the error arises through the use of the randomising game; if we were using the values $\hat{p}_i$ $(i = 1, 2)$ to estimate the proportion of people in a population from which a sample of 112 had been drawn at random, then this randomisation would introduce further error. The way in which sampling error is usually expressed is in terms of the variance of the estimate; a knowledge of the binomial distribution enables us to calculate the required variances.

Let n_{YY} denote the number in the group whose answer to both questions would be 'Yes', n_{NY} the number whose answer is 'No' to the sensitive question and 'Yes' to the innocuous question, and so on. Then the variance in the total number of 'Yes' responses arises because of the uncertain replies of $(n_{NY} + n_{YN})$ respondents. Using the fact that the variance for a binomial distribution is $np(1 - p)$, we obtain the variance of the number of 'Yes' responses among the n_{NY} as

$$n_{NY}(\tfrac{1}{2})(\tfrac{1}{2}),$$

where the probability that the innocuous question is asked is $\frac{1}{2}$. A similar result holds for the n_{YN} respondents. The independence of each individual's responses leads to the variance of the total number of 'Yes' replies as

$$(\tfrac{1}{2})(\tfrac{1}{2})(n_{NY} + n_{YN})$$

and the variance of the estimator $\hat{p}$ as

$$\mathrm{Var}\,(\hat{p}) = \frac{4}{(112)^2}(n_{NY} + n_{YN})\left(\frac{1}{2}\right)\left(\frac{1}{2}\right).$$

Assuming independence of the sensitive and innocuous items we can estimate n_{YN} as $112 \times \hat{p} \times \frac{6}{7}$, etc., and thus calculate estimates of the variances, which leads to the following results.

$$\text{For} \quad \hat{p}_1 = 0{\cdot}91, \quad \mathrm{Var}\,(\hat{p}_1) = 0{\cdot}00708 \quad \text{and}$$

$$\text{for} \quad \hat{p}_2 = 0{\cdot}054, \quad \mathrm{Var}\,(\hat{p}_2) = 0{\cdot}00112.$$

If the 112 were a random sample (say from a population of people of a particular age and educational attainment) then the values of $\hat{p}_i$ $(i = 1, 2)$ would estimate the proportions in the population.

The variances would then be given by

$$\mathrm{Var}\,(\hat{p}) = \frac{4(1 - \lambda)}{112}.$$

For practical purposes we often convert the estimate and its variance into a confidence interval. A 95 % confidence interval for p is selected by a procedure which ensures that 95 % of such intervals will contain the true value of p. Using the normal approximation to the binomial distribution the confidence intervals are given by $p \pm 1{\cdot}96\sqrt{\mathrm{Var}\,p}$. Thus a 95 % confidence interval for the proportion *of the class* who had consumed alcohol illegally in a pub is (0·75, 1·07), while for the proportion who are not very interested in the opposite sex a 95% confidence interval is $(-0{\cdot}025, 0{\cdot}133)$. (Since the class on which the practical trials were carried out did not constitute a random sample, a population estimate is not appropriate.)

Looking at these intervals, however, we see immediately that they contain impossible values of p. Since the probability is 0·95 that such intervals contain the true value of p, the probability that such intervals, truncated at 0 and 1, contain p is also 0·95. So the final 95% confidence intervals are $(0{\cdot}75, 1)$ for p_1 and $(0, 0{\cdot}133)$ for p_2.

5. Concluding remarks

The use of the randomised response technique in social surveys can be adapted to classroom use to provide an interesting probability experiment which leads to consideration of certain aspects of inference. There are, however, some problems which can arise. One is that it is possible to find estimates of p outside the interval $[0, 1]$, or confidence intervals extending outside this. The logical procedure is to adopt 0 or 1 as the estimate in the former case and to truncate the interval in the latter. Another problem is that by injudicious choice of questions the answer can be 'almost' revealing in the sense that P('Yes' to sensitive item|'Yes' to randomised response question) is high. This can occur if the innocuous item is particularly rare.

References

1. A. H. Barton, Asking the embarrassing question, *Public Opinion Quarterly* **22** (1958), 67–68.
2. W. A. Belson, The extent of stealing by London boys and some of its origins, *Advancement of Science* **25** (1968), 71–184.
3. D. G. Horvitz, B. V. Shah, and W. R. Simmons, The unrelated question randomised response model, *Proceedings of the Social Statistics Section, American Statistical Association* (1967).
4. S. L. Warner, Randomised response: A survey technique for eliminating evasive answer bias. *J. Amer. Statist. Assoc.* **60** (1965) 63–69.

Negative Chickens: An Exercise in Model Building

C. MAR MOLINERO
University of Southampton

Mathematical models

Many real-life situations require an understanding of a system before rational decisions can be taken. In general, we have some variables under our control—decision variables—and we want to know what the effect of changing them will be upon some characteristic of the system. Sometimes it is possible to experiment with a small-scale model—think for example of a small-scale aircraft in a wind tunnel—but very often experimentation is out of the question. We can envisage many such cases: no small-scale model can tell us what the effect of falling birth rates is going to be on the demand for schooling services in ten years' time; schools, however, have to be built now, and a replacement policy for teachers has to be envisaged. The normal approach in this case is to build a mathematical model.

A mathematical model is normally defined as 'an idealized representation of a real life system' (see reference 5). Such a model takes the form of a set of equations and, possibly, inequalities. The solution of this system will describe the dependent variable as a function of the decision variables, say

$$y = f(x_1, x_2, \ldots, x_n).$$

In our case y can be the number of new schools needed, and $x_1, x_2, \ldots, x_n$ the number of live births in the previous years. Notice that the function f is our representation of the real system. The more realistic a description of the actual process the function f is, the more reliable a solution obtained from the mathematical model will be.

Errors and the method of least squares

When we build a mathematical model we do not start from complete darkness, for normally we have a set of observations on the dependent variable and the decision variables. We also have some sort of hypothesis on the kind of relationship that is expected to hold. We can put it in equation form:

$$y_t = f(X_{1t}, X_{2t}, \ldots, X_{Nt}) + \varepsilon_t \qquad (t = 1, 2, \ldots, k), \tag{1}$$

where y_t is the value that the variable y takes at time t, X_{it} is the value that the variable X_i takes at time t, and ε_t is the error term at time t. We have data for k periods. The need to include an error term in equation (1) arises for many reasons. Variables are measured with error, perhaps the right variable is not measured at all and we have to use a related one, the exact form of the function f may not be known, or we may decide to include only the most important variables in our equation.

In equation (1) we have assumed that the function f can have any form. In practice, if f is unknown, it is normal to start with a linear function. Equation (1) then becomes

$$y_t = \beta_0 + \beta_1 X_{1t} + \cdots + \beta_N X_{Nt} + \varepsilon_t, \tag{2}$$

where $\beta_0, \beta_1, \ldots, \beta_N$ are parameters to be estimated. In general the parameters β_i have a certain meaning attached to them. Note that the set of errors ε_t is not known because the parameters β_i are not known. It is common practice to denote the estimates by $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_N$ and to compute these by the method of least squares, in which the sum of the squared errors is minimized. The sum is:

$$L(\beta_0, \beta_1, \ldots, \beta_N) = \sum_{t=1}^{k} \varepsilon_t^2 = \sum_{t=1}^{k} (y_t - \beta_0 - \beta_1 X_{1t} - \cdots - \beta_N X_{Nt})^2. \tag{3}$$

Since we look for a minimum of this function, we equate its derivatives to zero and so obtain

$$\frac{\partial L}{\partial \beta_i} = -2 \sum_{t=1}^{k} X_{it}(y_t - \beta_0 - \beta_1 X_{1t} - \cdots - \beta_N X_{Nt}) = 0 \qquad (i = 0, 1, \ldots, N), \tag{4}$$

where $X_{0t} = 1$.

The set of $N + 1$ equations (4) depends only on the $N + 1$ variables $\beta_0, \ldots, \beta_N$, and the system can in general be solved to produce optimal values $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_N$. By use of these computed parameter values estimates of the errors ε_t can be obtained. Goodness of fit of the equation (2) to the data is usually measured by a dimensionless value R^2 (see reference 4) defined by

$$R^2 = 1 - \frac{\sum_{t=1}^{k} \varepsilon_t^2}{\sum_{t=1}^{k} y_t^2}. \tag{5}$$

The value R^2 can be interpreted as the proportion of the variability of y which is explained by the set of variables $X_1, \ldots, X_N$.

Eggs and chickens

Egg production involves several steps. The first one, hatching, is normally carried out by specialised hatcheries. At the end of the hatching period a chick, known as a *day-old chick*, is born. Such a chick does not require any food for the first two days of its life, and this is the moment to move it out of the hatchery. From then on it is known as a *growing pullet*. Pullets have to grow over a period of five months before they start producing eggs. Here again there are specialised units that buy day-old chicks and rear them under optimal conditions of diet, heating, lighting, etc. At about 19 weeks of age pullets are moved into the laying flock, and are then classed as *hens*.

Data on the number of growing pullets that are alive are available on a quarterly basis from government census and surveys (see reference 2). The published statistics in this case, as in most other cases, are subject to error. For instance, some of the chickens placed by hatcheries will not be included in the census of growing pullets since they may be sold to 'statistically insignificant holdings' or to the general public. On the other hand, some growing pullets may have been bought abroad or sold by non-recorded hatchers, and they will not have been included in the figures of chicks placed by hatcheries.

We are interested in finding an equation to explain the number of growing pullets, given the number of chicks placed by hatcheries. Such an equation can be desirable for different reasons:

(i) We may be interested in estimating the proportion of day-old chicks sold to holdings not covered by the census.

(ii) We may want to estimate the population of growing pullets at times not covered by the census.

(iii) It may be desirable to forecast the number of growing pullets at a certain moment in the future. Such information can be of use when planning production of feeding stuff, for example.

A mathematical model

Let C_t be the number of growing pullets, as shown in the census taken at the beginning of month t. Let X_{it} be the number of chickens placed by hatcheries during month $t - i$. Not all of the X_{it} chickens will be alive at the beginning of the month t, and, if we take β_i to be the proportion of growing chickens that are included in the count C_t, we have the model

$$C_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \beta_5 X_{5t} + \varepsilon_t. \qquad (6)$$

We have not included terms beyond X_{5t} because all chickens of more than five months of age are recorded as hens and do not form part of the population measured by C_t.

What is the meaning of the coefficients β_i in equation (6)? The first coefficient, β_0, can be interpreted as the number of chickens that are included in the census but were not registered as being hatched. The sign of β_0 would indicate whether imports or exports predominated. Coefficients $\beta_1, \ldots, \beta_5$ give some sort of measure of survival rate, we expect them to be positive, and to satisfy the following relationship:

$$0 \leqslant \beta_5 \leqslant \beta_4 \leqslant \beta_3 \leqslant \beta_2 \leqslant \beta_1 \leqslant 1. \qquad (7)$$

The least squares method with data from England and Wales for the period 1969–76 leads to

$$C_t = 4{\cdot}25 + 0{\cdot}225X_{1t} + 1{\cdot}240X_{2t} + 0{\cdot}675X_{3t} - 0{\cdot}130X_{4t} + 0{\cdot}866X_{5t} + \varepsilon_t, \quad (8)$$

where C_t, X_{it} are in millions; here $R^2 = 0{\cdot}82$.

Judged by R^2, equation (8) can be considered as a good description of the process, since 82% of the variation in the census is accounted for by the expression obtained. But a closer look at the numerical values of the coefficients will make us think twice before being satisfied with it, for the following reasons.

(i) The constant term indicates that every quarter 4.25 million chickens are imported into England and Wales. Such a figure is excessive and does not match the data published in the Overseas Trade Statistics.

(ii) The inequalities (7) do not all hold for the coefficients $\beta_1, \ldots, \beta_5$. Indeed one of them, β_4 is negative. This means that chickens born four months ago have to be considered as being negative now. What is a negative chicken?

Perhaps it is not surprising that (7) does not hold, for the least squares method used did not make any allowance for it. We now incorporate (7) and formulate the problem as:

Minimise

$$L = \sum_{t=1}^{k} (y_t - \beta_0 - \beta_1 X_{1t} - \beta_2 X_{2t} - \beta_3 X_{3t} - \beta_4 X_{4t} - \beta_5 X_{5t})^2$$

subject to the conditions

$$0 \leqslant \beta_5 \leqslant \beta_4 \leqslant \beta_3 \leqslant \beta_2 \leqslant \beta_1.$$

By introducing the above set of constraints we are dropping the condition that the derivatives of the function L are to be zero at the minimum. We cannot make use of the previous technique to find the optimum values of the β_i. There are, however, other techniques available. The problem above has been solved by making use of a technique called quadratic programming (see reference 3), and the result obtained is

$$C_t = 4{\cdot}87 + 0{\cdot}85X_{1t} + 0{\cdot}58X_{2t} + 0{\cdot}58X_{3t} + 0{\cdot}58X_{4t} + 0{\cdot}46X_{5t} + \varepsilon_t.$$

Unfortunately the coefficients in the above equations do not decrease in a smooth way as was expected. So why not introduce this information in the set of constraints? If we write

$$\beta_{i+1} = \beta_i - \alpha \qquad (\mathrm{i} = 1, 2, 3)$$

and

$$\alpha \geqslant 0,$$

equation (6) now reads

$$C_t = \beta_0 + \beta_1 X_{1t} + (\beta_1 - \alpha)X_{2t} + (\beta_1 - 2\alpha)X_{3t} + (\beta_1 - 3\alpha)X_{4t} + \beta_5 X_{5t} + \varepsilon_t,$$

or

$$C_t = \beta_0 + \beta_1 y_{1t} - \alpha y_{2t} + \beta_5 X_{5t} + \varepsilon_t, \tag{9}$$

where

$$y_{1t} = X_{1t} + X_{2t} + X_{3t} + X_{4t},$$
$$y_{2t} = X_{2t} + 2X_{3t} + 3X_{4t}.$$

The meaning of the coefficients in equation (9) is as follows:

β_0 measures net imports or exports, according to its sign;

β_1 indicates the proportion of chickens registered in the placings that are also registered in the census;

α measures the mortality rate of growing pullets;

β_5 contains information about the average age at which growing pullets become hens for statistical purposes. We have said that this age is about five months, but the figure is only approximate, and we want to estimate its exact value from the data.

The coefficients in (9) were estimated by use of quadratic programming. For that purpose imports, exports, and non-recorded transfers were assumed to be very small and so β_0 was set to zero. From knowledge of the industry it was obvious that the mortality rate has never been more than 2% per month, and therefore the constraint

$$0 \leqslant \alpha \leqslant 0{\cdot}02$$

was introduced. Finally, the coefficients β_1 and β_5 naturally satisfy the inequalities

$$0 \leqslant \beta_1 \leqslant 1, \qquad 0 \leqslant \beta_5 \leqslant 1.$$

The result is as follows:

$$C_t = 0{\cdot}910y_{1t} - 0{\cdot}02y_{2t} + 0{\cdot}779X_{5t} + \varepsilon_t, \qquad (10)$$
$$R^2 = 0{\cdot}64.$$

Conclusions

We have, finally, an equation with which we are satisfied. It provides very relevant information, given the limited amount of data and given its simplicity. In order to obtain it we had to change previous equations to fit our understanding of the process. We have tried to make the best possible use of all the information at hand. This is the usual approach in model-building.

We started our analysis with data on the number of growing pullets, and data on the number of placings. We did not know how good an indicator of changes in the growing pullet sector the series of placings was. Equation (10) tells us that we can use placings when building a model of the poultry industry. However, the analysis does not end here. Further analysis has shown that the residuals ε_t of equation (10) contain a lot of information, and that therefore (10) does not actually make the best possible use of the available data. In the light of the above results a more complete model has been developed. This model adds to our knowledge of how changing prices and varying demand affect the actions of chicken growers, and it can be used to forecast their future decisions.

References

1. Ministry of Agriculture, Fisheries and Food. Statistical Information: Agricultural Returns for England and Wales.
2. Ministry of Agriculture, Fisheries and Food. Statistical Information: Hatching Eggs and Placings by Hatcheries in the United Kingdom.
3. G. R. Walsh, *Methods of Optimisation* (Wiley, London, 1975).
4. H. T. Hayslett, *Statistics Made Simple* (W. H. Allen, London, 1967).
5. H. A. Taha, *Operations Research, An Introduction* (Macmillan, London, 1971).

Factorization and Random Numbers

H. J. GODWIN
Royal Holloway College, London

One of the simplest problems that one can state is—given a natural number N, what are its factors? Apart from the intrinsic pleasure of factorizing a difficult number, and the challenge of finding methods that will enable the hitherto unconquerable to be conquered, there are many problems in number theory that require factorization to be carried out. Recently, the invention of the 'trapdoor' method of coding messages (see reference 2) has given the problem some importance outside pure mathematics.

If N is small the solution of the problem is as simple as its statement; one divides N by 2 and if the division is exact one has the factor 2 and the smaller number $N/2$ for which to repeat the process. If 2 does not divide N exactly one tries 3, 5, and so on through prime numbers; if N has no factor less than $\sqrt{N}$ then N is itself prime.

If N is large this method becomes very time-consuming, and there is the additional disadvantage that one needs a large table of primes for possible use as divisors. In practice it is often easier to divide by all odd numbers, or all odd numbers excluding multiples of very small primes, than to have to discover if the divisor is itself prime. What one means by 'large N' depends on the equipment available; possibly $N > 10^3$ for mental arithmetic, $N > 10^5$ with a desk calculator, and $N > 10^{14}$ for a large computer, are reasonable values.

For large N other methods are available: some of these test whether N is prime and others find factors of N if it is known to be composite. Testing the primality of N is based on Fermat's theorem (see reference 1, p. 37) that if N is a prime, then 2^{N-1} leaves remainder 1 on division by N. Consequently if 2^{N-1} does not leave remainder 1 on division by N, then N is composite. Unfortunately the converse of Fermat's theorem is not true, the simplest example being $N = 341$, because $2^{340} - 1$ is divisible by 341, but 341 is not prime. However, there are ways of getting round this difficulty and it is a fairly simple matter to determine whether a number is prime or not, but far less easy to factorize it if it is composite. As long ago as 1909 it was proved that $2^{128} + 1$ and $2^{256} + 1$ were both composite, but the first was factorized only in 1970, and the second has so far resisted all efforts to find its factors.

The most common method for factorizing a large number N is to find numbers x and y such that $x^2 \equiv y^2 \pmod{N}$. (For those unfamiliar with this notation we recall that $a \equiv b \pmod{N}$ means that a and b leave the same remainder when divided by N. We shall make a lot of use of the fact that if $a \equiv b \pmod{N}$ and $c \equiv d \pmod{N}$, then $ac \equiv bd \pmod{N}$; it is also true that $a + c \equiv b + d \pmod{N}$, though we shall not use this.) If $x \equiv y \pmod{N}$ or $x \equiv -y \pmod{N}$ this does not help, but if neither of these two obtains, then N factorizes with one factor in common with $x + y$ and the other in common with $x - y$. The H.C.F. of two numbers can be found by using Euclid's algorithm (see reference 1, p. 42), without knowing the factorization of either.

To find x and y we first seek representations of N of the form $z_i^2 - k_i$, where each k_i is the product of small primes. We want to find a set of k_i that multiply to give a perfect square, and we can do this by pairing k's that have the same prime factor to an odd power, e.g. $2^2 . 5$ and $5^3 . 7^2$ multiply to give a perfect square. Since a square is positive we must also pair off -1's, and so for this purpose -1 is treated as a prime. If we have $r + 1$ k's formed from a set of r primes we can always find a product of some or all of them that is a perfect square (see Problem no. 11.1 on p. 28).

Suppose that, by renumbering, we have $k_1 k_2 \ldots k_{r+1} \equiv y^2 \pmod{N}$, then $y^2 \equiv z_1^2 z_2^2 \ldots z_{r+1}^2 = (z_1 z_2 \ldots z_{r+1})^2 \equiv x^2$, say, and we have found an x and y. To keep the numbers with which we are working as small as possible we can work throughout with remainders modulo N. If it turns out that $x \equiv \pm y(\text{mod } N)$ then we must try some more representations: there is an element of luck in the procedure.

To illustrate the method, consider $N = 53369$. To find k_i with small factors we try z_i near $\sqrt{N}$ and we obtain $z_1 = 231$, $k_1 = -2^3$, $z_2 = 229$, $k_2 = -2^5 . 29$, $z_3 = 227$, $k_3 = -2^4 . 5 . 23$, $z_4 = 225$, $k_4 = -2^3 . 7^3$, $z_5 = 223$, $k_5 = -2^3 . 5 . 7 . 13$. Another way of getting small factors is to arrange that some chosen small primes divide k several times: e.g. we can make 2^2 and 5^2 divide k by arranging that z ends in 69 as N does, and for this z must end in 13, 37, 63, or 87. By trial we find that $z_6 = 37$ gives $k_6 = -2^5 . 5^3 . 13$.

We now notice that k_1, k_4, k_5, and k_6 involve only 2, 5, 7, 13 (and -1): although we have not got enough values to guarantee forming a square it is worth seeing if it happens nevertheless. We find that $k_5 k_6$ is 7 times a perfect square, and $k_1 k_4$ is also 7 times a perfect square: hence $k_1 k_4 k_5 k_6$ is a perfect square, and on multiplying out and taking remainders modulo N we obtain $z_1 z_4 z_5 z_6 \equiv 25810 \text{ (mod } N)$ and $k_1 k_4 k_5 k_6 \equiv 10378^2 \text{ (mod } N)$.

Now the H.C.F. of N and $25810 + 10378$ is 83, and the H.C.F. of N and $25810 - 10378$ is 643, and so N is factorized as 83 times 643. There is, as has been said, an element of chance in the above method: it can be made more systematic by using the ideas of the theory of continued fractions (see reference 1, p. 49) to obtain small values of k, but even so success may come early or late. The ideas of the method are fairly old, but it is still in general the most powerful one, and it was by this method (adapted in a sophisticated way for use with a computer) that Morrison and Brillhart (reference 3) factorized $2^{128} + 1$.

Recently an English mathematician, J. M. Pollard (reference 4) devised a method that is entirely new and may sometimes be much quicker than the older methods. The use of the word 'sometimes' indicates that there is again an element of chance involved, and in fact Pollard's method is based on ideas from probability theory. For this reason we now consider what may appear at first sight to be a totally unrelated problem.

Suppose you are interviewing a queue of people and asking each one what day of the year his or her birthday falls. If you ask only two people it is highly unlikely that they will have the same birthday but if you ask 367 then it is certain that at least two will have the same birthday, though you cannot tell in advance what particular day it will be. Somewhere between these extremes there is an even chance of finding that two interviewees share a birthday, and it is rather remarkable that this is reached when there are no more than 23 people. A similar problem is that of collecting numbered cards, of the sort that are given away with some packets of tea. If there are 50 in the set then there is an approximately even chance that 9 cards will contain at least one duplicate. The birthdays and the card numbers are examples of random numbers, that is, numbers that are unknown before they are actually observed, but are such that there is a definite probability of any given number occurring at any given observation. In both the above examples it was assumed that all numbers occur with equal probability so that we had uniformly distributed random numbers. In practice not all days of the year are equally likely as birthdays (even leaving aside the obvious case of 29 February) and possibly cards are not uniformly distributed either: real life is seldom as simple as mathematical examples would like it to be.

If the numbers $1, 2 \ldots, N$ are all equally likely, then the probability that m numbers are all different is $(N - 1)(N - 2) \ldots (N - m + 1)/N^{m-1}$ and this is about one half for $m \doteqdot \sqrt{(2N \log_e 2)} \doteqdot \sqrt{(1{\cdot}38N)}$; to prove this result one approximates to $N!$ and $(N - m)!$ by means of Stirling's formula which states that $n!$ is approximately $\sqrt{2\pi} n^{n+1/2}e^{-n}$. Note that $\sqrt{(1{\cdot}38)(365)} \doteqdot 22{\cdot}4$ and $\sqrt{(1{\cdot}38)50} \doteqdot 8{\cdot}3$, which are in good agreement with the values 23 and 9 quoted above.

Random numbers are needed in various mathematical applications and, although they can be generated by physical means such as tossing coins or by more sophisticated means such as ERNIE (the Electronic Random Number Indicating Equipment that controls the distribution of Premium Bond payments) it is frequently better to work instead with pseudo-random numbers, which behave in many ways like random numbers but are in fact completely determinate. (Their advantage is that work done with such numbers is completely reproducible, whereas truly random numbers cannot, by definition, be reproduced in the same sequence, to order.) There are various methods of producing pseudo-random numbers, but we consider only the one that we shall use later. We choose a number N and a starting value u_1 where $0 \leqslant u_1 < N$. Then u_2 is the remainder when $u_1^2 + 1$ is divided by N, u_3 is the remainder when $u_2^2 + 1$ is divided by N, and so on. For example, if $N = 221$, and $u_1 = 1$ then $u_2, u_3, \ldots, u_{14}$ are 2, 5, 26, 14, 197, 135, 104, 209, 145, 31, 78, 118, 2. Since each term is determined by the preceding one, we shall now repeat the cycle of terms from u_2 to u_{13} by having $u_{15} = u_3$, $u_{16} = u_4$ etc. The sequence thus has a loop of terms preceded by a 'tail' (in this case consisting of only one term), and all such sequences must be of this form, because there is only a finite number of possible values for the u_i, so that repetition must eventually occur.

If we now form a second sequence $v_1, v_2, \ldots$ by taking every other term of the first, so that we have in the above example $v_1 = 2$, $v_2 = 26$, $v_3 = 197$ etc., then both sequences will ultimately consist of terms in the loop, but the second sequence will go round the loop twice as quickly as the first one. If the length of the tail is t and the length of the loop is m, then we have $u_s = u_{t+k}$ if $s \equiv t + k \pmod{m}$, so long as $s \geqslant t + 1$. If a is an integer such that $am > t$, then we have $v_{am} = u_{2am} = u_{t+2am-t} = u_{t+am-t} = u_{am}$, so that a u_r and a v_r coincide. In the above example we have $t = 1$, $m = 12$, $a = 1$ and $v_{12} = u_{24} = u_{12}$.

Since the loop is first completed when two of the u_r are equal, the value of m is likely to be of the order of magnitude of $\sqrt{N}$, since, as in the birthday and card problems, we have first found a duplicate after getting $t + m + 1$ numbers out of N. By 'order of magnitude' we mean that m is unlikely to be a large multiple of $\sqrt{N}$ or a small fraction of $\sqrt{N}$: as an abbreviation we write 'm is $O(\sqrt{N})$'. (The symbol O is used in pure mathematics for order of magnitude in a rather less definite way than we are using it here.)

Now comes the crux of the method. Suppose that Q is a factor of N; then two numbers congruent modulo N are automatically congruent modulo Q. We expect to get two equal terms in the sequence, i.e. two terms congruent modulo N, in $O(\sqrt{N})$

terms, and so similarly we expect to get two numbers congruent modulo Q in $O(\sqrt{Q})$ terms. We shall thus find a u_r and v_r, where r is $O(\sqrt{Q})$, such that the H.C.F. of $v_r - u_r$ and N is Q. The Q that we find need not be prime, nor even if it is need it be the smallest factor of N. It cannot be N itself since both u_r and v_r lie between 0 and $N - 1$. The method may fail by u_r and v_r being equal and not merely congruent modulo Q: this happens for example with $N = 21$, $r = 3$, but is unlikely to happen if N is large enough for the method to be needed. If it does we can try a different starting value for u_1, or even a different formula for u_{r+1} in terms of u_r, e.g. $u_{r+1} \equiv u_r^2 - 1 \pmod{N}$.

In the above example we have $v_r - u_r$, and the H.C.F., as in Table 1 and so 13 is a factor of N.

TABLE 1

r	$v_r - u_r$	H.C.F
1	1	1
2	24	1
3	192	1
4	78	13

The numbers here were chosen small enough for the whole loop modulo N to be given without going to enormous length, but so small a value of N does not give a fair idea of the power of the method. If we take $N = 53369$, as in the previous example, we obtain Table 2.

TABLE 2

r	$v_r - u_r$	H.C.F.
1	1	1
2	24	1
3	31373	1
4	28899	1
5	24957	1
6	−16421	1
7	−5722	1
8	−5681	1
9	−8530	1
10	1909	83

We have seen that the number of terms of the sequence that we have to calculate is likely to be $O(\sqrt{Q})$; whereas, for the method of division by 2, 3, 5, . . .mentioned early on in this article, the number of trials is $O(Q/\log Q)$, since the number of primes less than Q is about $Q/\log Q$. Hence Pollard's method is likely to produce an answer much more quickly. The order of magnitude of the method of finding representations is not known, but experience suggests that it is smaller than $N^{1/4}$.

The moral of this story is that the combination of simple results from two apparently separate areas of mathematics can produce new and powerful methods—but it takes a genius to do it!

References

1. R. Courant and H. Robbins, *What is Mathematics?* (Oxford University Press, 1941).
2. M. Gardner, Mathematical Games, *Scientific American* **237** (August 1977), 120–124.
3. Michael A. Morrison and John Brillhart, A method of factorizing and the factorization of F_7, *Math. Comp.* **29** (1975), 183–205.
4. J. M. Pollard, A Monte Carlo method for factorization, *Nordisk Tidskr. Informationsbehandling* (*BIT*) **15** (1975), 331–334.

The Determination of Easter

H. V. SMITH
Leeds Polytechnic

Easter, the greatest of Church festivals, is celebrated on the Sunday following the first full moon occurring on or after 21 March, i.e. the vernal equinox. Rogosinski (reference 1) has discussed a method of calculating Easter Sunday based on a rule devised by Gauss. Here we discuss the following rule which appeared in Butcher's Ecclesiastical Calendar, 1876.

Divide	by	Quotient	Remainder
the year y	19		j
y	100	k	h
k	4	m	n
$k + 8$	25	p	
$k - p + 1$	3	q	
$19j + k - m - q + 15$	30		r
h	4	s	u
$32 + 2n + 2s - r - u$	7		v
$j + 11r + 22v$	451	w	
$r + v - 7w + 114$	31	x	z

Here x is the number of the month of the year y and $1 + z$ is the day of that month upon which Easter Sunday falls. It is now easy to write a computer program giving Easter's date for successive years. As an example a program written in BASIC language is as follows.

```
 80 PRINT     EASTER SUNDAY
 90 PRINT     YEAR , MONTH , DATE
100 FOR  Y = 1978 TO 2000 STEP 1
110 LET  J = Y-19*(INT(Y/19))
120 LET  K = INT(Y/100)
130 LET  L = Y-100*K
140 LET  M = INT(K/4)
150 LET  N = K-4*M
160 LET  P = INT((K+8)/25)
170 LET  Q = INT((K-P+1)/3)
180 LET  R = 19*J+K-M-Q+15-30*(INT((19*J+K-M-Q+15)/30))
190 LET  S = INT(L/4)
200 LET  U = L-4*S
210 LET  V=32+2*N+2*S-R-U-7*(INT((32+2*N+2*S-R-U)/7))
220 LET  W=INT((J+11*R+22*V)/451)
230 LET  X = INT((R+V - 7*W + 114)/31
240 LET  Z = R+V-7*W+115-31*X
250 PRINT Y,X,Z
260 NEXT Y
270 STOP
280 END
```

This program may be used to find the date of Easter Sunday for any particular year by omitting the FOR statement and setting Y equal to the desired year. It then gives the following dates for Easter Sunday up to the year 2000.

Year	Month	Day
1979	4	15
1980	4	6
1981	4	19
1982	4	11
1983	4	3
1984	4	22
1985	4	7
1986	3	30
1987	4	19
1988	4	3
1989	3	26
1990	4	15
1991	3	31
1992	4	19
1993	4	11
1994	4	3
1995	4	16
1996	4	7
1997	3	30
1998	4	12
1999	4	4
2000	4	23

I am grateful to Dr D. McNally for giving me the details of the method as cited in Butcher's Calendar.

Reference

1. H. P. Rogosinski, The perpetual calendar, *Math. Spectrum* **5** (1972/73), 42–46.

Letters to the Editor

Dear Editor,

Partitions of sets of integers

In *Mathematical Spectrum* (Volume 9, Number 1, p. 26) M. C. Sandford considered the problem of dividing the set of consecutive integers $\{1,2,\dots,k\}$ into n parts restricted by the condition that no part contains three integers in arithmetic progression. Given such a restricted partition of $\{1,2,\dots,k\}$, it may be possible to extend it to a partition of $\{1,2,\dots,k+1\}$ by adjoining $k+1$ to some existing part without violating the condition. However, if such an extension is impossible the partition of $\{1,2,\dots,k\}$ will be called a maximal restricted partition of length k. For example, the restricted partition of $\{1,2,\dots,6\}$ into 2 parts $\{1,3,6\}$, $\{2,4,5\}$ can be extended either by adjoining 7 to the first part, in which case it becomes maximal, or to the second part, in which case 8 can then be adjoined to the first part and the resulting partition of $\{1,2,\dots,8\}$ is maximal. Clearly every restricted partition of $\{1,2,\dots,k\}$ can be obtained from a list of all possible maximal restricted partitions. However, the same partition of $\{1,2,\dots,k\}$ may result from distinct maximal restricted partitions, so that counting the number of restricted partitions of $\{1,2,\dots,k\}$ is not simply related to counting the number of maximal restricted partitions of various lengths.

By systematically extending an initial state $\{1\}$, $\{\,\}$, $\dots$, $\{\,\}$ with 1 in the first part and nothing in the remaining $n-1$ parts, it is possible to enumerate all maximal restricted partitions into n parts in a 'dictionary' order. This has been done for $n = 3$ with the aid of a computer. There are 36923 distinct maximal restricted partitions and of these 8 have the minimum length 8 and the same number have the maximum length 26. The statistics of these partitions and a list of those of minimum length are given in Tables 1 and 2.

TABLE 1. Number of maximal restricted partitions of k for $n = 3$

k	8	9	10	11	12	13	14	15	16	17
number	8	10	87	151	286	396	1436	1802	2947	3551
k	18	19	20	21	22	23	24	25	26	
number	6018	5576	5598	5063	2512	1009	420	45	8	

TABLE 2. The maximal restricted partitions of minimum length 8 for $n = 3$

$\{1,2,4,5\},\{3,6\},\{7,8\}$	$\{1,2,5\},\{3,4,6\},\{7,8\}$	$\{1,2,5\},\{3,6\},\{4,7,8\}$
$\{1,4,5\},\{2,3,6\},\{7,8\}$	$\{1,4,5\},\{2,7,8\},\{3,6\}$	$\{1,5\},\{2,3,6\},\{4,7,8\}$
$\{1,5\},\{2,4,7,8\},\{3,6\}$	$\{1,5\},\{2,7,8\},\{3,4,6\}$	

For $n = 4$, it is easily shown that there are 10 maximal restricted partitions with the minimum length of 10. Although the maximum length is not known, it is at least 60 since

$$\begin{array}{l}
\{\ 1,\ 2,\ 4,\ 5,10,11,13,23,31,32,37,38,40,47,50,55\},\\
\{\ 3,\ 6,\ 7,12,16,22,24,30,33,34,39,43,49,51,58\},\\
\{\ 8,\ 9,17,20,21,27,35,36,41,42,44,54,56,59,60\},\\
\{14,15,18,19,25,26,28,29,45,46,48,52,53,57\}
\end{array}$$

is a maximal restricted partition of $\{1, 2, \ldots, 60\}$. Eleven such examples are known and it is conjectured that even longer sets of consecutive integers can be partitioned into four parts, none of which contains three integers in arithmetic progression.

Yours sincerely,
J. K. MACKENZIE
(CSIRO Division of Chemical Physics,
Clayton, Victoria, Australia)

Dear Editor,

The circular and hyperbolic functions

With reference to the circle of unit radius, the magnitudes of the trigonometric ratios for a given angle θ (with $0 < \theta < \pi/2$) may be compared directly. Thus, for instance, in the notation of Figure 1,

$$\sin\theta = BE, \cos\theta = OE, \tan\theta = AD, \operatorname{cosec}\theta = OF, \sec\theta = OD, \cot\theta = \text{AF}.$$

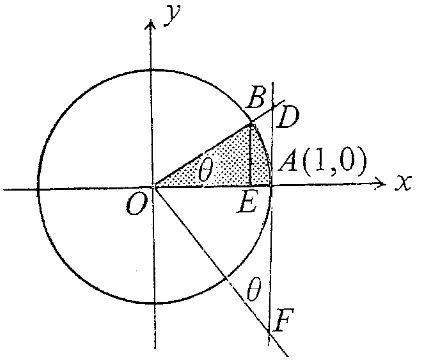

Figure 1

The angle θ can be interpreted as the length of the arc AB subtended by θ. It also represents twice the area $\mathscr{A}$ of the corresponding segment, since

$$\frac{\theta}{2\pi 1} = \frac{\mathscr{A}}{\pi 1^2} \quad \text{i.e.} \quad \theta = 2\mathscr{A}.$$

The equation of the circle, with its centre as the origin, is $x^2 + y^2 = 1$. In particular, $x = \cos\theta$, and so

$$2\mathscr{A} = \cos^{-1} x. \tag{1}$$

The hyperbolic functions $\sinh\phi$, $\cosh\phi$ are related to the hyperbola, defined by the equation $x^2 - y^2 = 1$, through the identity $\cosh^2\phi - \sinh^2\phi = 1$, in the same way that the trigonometric functions are related to the circle $x^2 + y^2 = 1$ via their identity $\cos^2\theta + \sin^2\theta = 1$. We construct an analogous diagram to that in Figure 1, in which the magnitudes of the hyperbolic functions may be compared with reference to a hyperbola and in which the parameter ϕ is interpreted as an area. The clue is provided by Equation (1), which suggests that ϕ (with $\phi > 0$) is equal to twice the area of the hyperbolic sector shown shaded in Figure 2, where $OR = \cosh\phi$, $PR = \sinh\phi$. To prove this, we refer the hyperbola to its asymptotes as axes by means of the co-ordinate transformation

$$x - y = \sqrt{2}\xi$$

and

$$x + y = \sqrt{2}\eta.$$

Then the equation of the hyperbola becomes

$$x^2 - y^2 = (x - y)(x + y) = 2\xi\eta = 1, \qquad \text{i.e.} \quad \xi\eta = \tfrac{1}{2}.$$

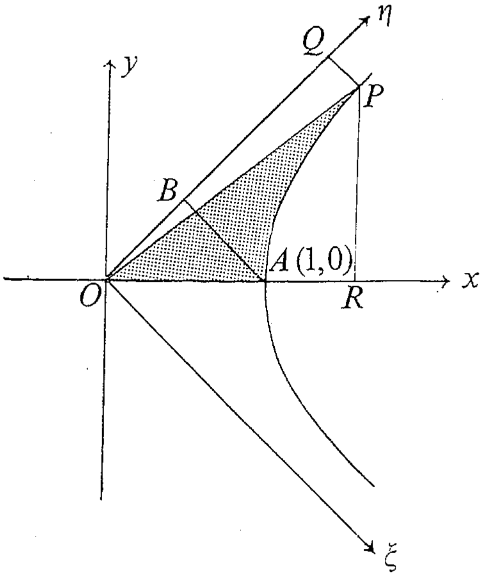

Figure 2

From this we deduce that the area in question is equal to the area of the figure $ABQP$, since the two right-angled triangles OPQ and OAB have the same area ($=\frac{1}{4}$ square units), according to the equation of the hyperbola. The two points A and P have the co-ordinates

$$\xi = 1/\sqrt{2}, \qquad \eta = 1/\sqrt{2}$$

and

$$\xi = \frac{x-y}{\sqrt{2}}, \qquad \eta = \frac{x+y}{\sqrt{2}}, \qquad \text{respectively;}$$

and for twice the area, $\mathscr{A}'$ say, of the figure $APQB$ we have

$$2\mathscr{A}' = 2\int_{1/\sqrt{2}}^{(x+y)/\sqrt{2}} \frac{1}{2\eta}\,d\eta = \log(x+y) = \log(x+\sqrt{(x^2-1)}) = \cosh^{-1}x = \phi;$$

which is the analogue of Equation (1).

We can further compare the magnitudes of the trigonometric functions with their hyperbolic counterparts if we construct a diagram which combines Figures 1 and 2 in such a way that the areas of the circular and hyperbolic segments are equal. The area $\mathscr{A}'$ of the hyperbolic segment will evidently be the same as the area $\mathscr{A}$ of the circular segment if

$$\int_{1/\sqrt{2}}^{\eta} \frac{d\eta}{2\eta} = \mathscr{A}, \qquad \text{i.e.}\quad \text{if } \log(\sqrt{2}\eta) = 2\mathscr{A},$$

or

$$\eta = (1/\sqrt{2})e^{2\mathscr{A}}.$$

Thus, given the angle θ, we measure a distance equal to $(1/\sqrt{2})e^{2\mathscr{A}} = (1/\sqrt{2})e^{\theta}$ along the η asymptote, and mark off the appropriate hyperbolic segment. We can then construct the various trigonometric and hyperbolic functions of θ so that their magnitudes are directly comparable. The reader may like to produce such a figure for himself.

Yours sincerely,
J. M. H. PETERS
(28 Belvidere Road, Liverpool 8)

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 the postal address of your school, college or university.

Problems

11.1. (Submitted by H. J. Godwin, Royal Holloway College.) The prime factorizations of $r + 1$ positive integers ($r \geqslant 1$) together involve only r primes. Prove that there is a subset of these integers whose product is a perfect square.

11.2. (Submitted by B. G. Eke, University of Sheffield.) The positive numbers x, y, z are such that

$$\frac{1}{x} + \frac{1}{y} + \frac{1}{z} = 1.$$

Show that $(x - 1)(y - 1)(z - 1) \geqslant 8$.

11.3. The rules for the card game of Clock Patience are as follows:

Shuffle the pack of cards and deal them into thirteen piles of four labelled A, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K. To play, take away the top card of the K pile (say it is 5), then the top card of the 5 pile (say it is J), then the top card of the J pile, and so on. The game proceeds until the fourth K is taken, and the game is said to 'come out' if, when the fourth K is taken, all the original piles are empty.

(a) What is the probability that a game comes out?

(b) Assume that, after dealing, the bottom cards on the piles form a rearrangement of A, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K. Show that this game comes out if and only if this rearrangement is a cyclic rearrangement.

Solutions to Problems in Volume 10, Number 2

10.4. The integers $a_1, a_2, \ldots, a_7$ are rearranged to give $b_1, b_2, \ldots, b_7$. Show that

$$(a_1 - b_1)(a_2 - b_2) \ldots (a_7 - b_7)$$

is even.

Solution by Glyn Normington (Leeds Grammar School)

Each odd integer present among $a_1, \ldots, a_7$ is also present among $b_1, \ldots, b_7$, so the number of odd integers among $a_1, \ldots, a_7, b_1, \ldots, b_7$ is even. Suppose the given product is odd. Then, for each i, $a_i - b_i$ is odd and exactly one of a_i, b_i is odd. This gives an odd number of odd integers among $a_1, \ldots, a_7, b_1, \ldots, b_7$ (namely 7). Contradiction!

Also solved by K. H. Yim (UMIST), John Savva (Winchester College), Bruce Westbury (Rugby School), John Ramsden (University of Aston).

A number of readers pointed out that 7 could be replaced by any odd positive integer.

10.5. The real polynomials $f_1(x), \ldots, f_{n-1}(x)$, $g(x)(n > 1)$ are such that

$$f_1(x^n) + xf_2(x^n) + \cdots + x^{n-2}f_{n-1}(x^n) = (1 + x + x^2 + \cdots + x^{n-1})g(x).$$

Show that $f_1(x), \ldots, f_{n-1}(x)$ all have $x - 1$ as a factor.

Solution by Bruce Westbury

Denote by $\omega_1, \ldots, \omega_{n-1}$ the complex nth roots of unity other than 1. Then, for $1 \leqslant i \leqslant n - 1$,

$$1 + \omega_i + \omega_i^2 + \cdots + \omega_i^{n-1} = 0.$$

and, if we evaluate the given identity at ω_i, we obtain

$$f_1(1) + \omega_i f_2(1) + \cdots + \omega_i^{n-2} f_{n-1}(1) = 0.$$

Thus the polynomial

$$f_1(1) + f_2(1)x + \cdots + f_{n-1}(1)x^{n-2}$$

has at least $n - 1$ distinct roots, yet its degree does not exceed $n - 2$. It follows that it must be the zero polynomial, so that

$$f_1(1) = f_2(1) = \cdots = f_{n-1}(1) = 0.$$

Hence $x - 1$ is a factor of the polynomials $f_1(x), f_2(x), \ldots, f_{n-1}(x)$.

Also solved by M. J. Zahorodny (Imperial College).

10.6. Let (a, b) denote the highest common factor of a and b. For any positive integers a, b, m, n with $(a, b) = 1$, show that

$$(a^m - b^m, a^n - b^n) = a^{(m,n)} - b^{(m,n)}.$$

Solution

Put $d = (m, n)$. We use induction on $m + n$, the case $m + n = 2$ being obvious. We may suppose that $m \geqslant n$. Now

$$a^m - b^m = a^{m-n}(a^n - b^n) + (a^{m-n} - b^{m-n})b^n,$$
$$a^m - b^m = (a^n - b^n)b^{m-n} + a^n(a^{m-n} - b^{m-n}).$$

Thus $(a^m - b^m, a^n - b^n)$ divides $(a^{m-n} - b^{m-n})a^n$ and $(a^{m-n} - b^{m-n})b^n$. Since $(a^n, b^n) = 1$, it follows that $(a^m - b^m, a^n - b^n)$ divides $a^{m-n} - b^{m-n}$. It is now easy to see that

$$(a^m - b^m, a^n - b^n) = (a^{m-n} - b^{m-n}, a^n - b^n).$$

Since $(m - n, n) = d$ and $(m - n) + n < m + n$, the inductive hypothesis gives that

$$(a^m - b^m, a^n - b^n) = a^d - b^d,$$

as required.

Problem 10.3. The solution to this problem in the previous issue was inadequate, in that it solves the problem only when p, q, r are positive integers. A solution valid for all positive real numbers will be published in Volume 11, No. 2.

Book Reviews

The Foundations of Mathematics. By IAN STEWART and DAVID TALL. Oxford University Press, Oxford. Pp. xii + 263. Boards £7·50, paper covers £3·95.

When I first began to study mathematics at university I had no idea that the problem of defining irrational in terms of rational numbers ought to have been teasing my young mind for some time past; and therefore the first lectures in analysis on Dedekind cuts, while they brought illumination to more questing spirits, reduced me to a state of (not wholly unexpected) bewilderment. When I read now the notes I took then of those lectures I am astonished by their clarity, and am sure that our lecturer had first explained Dedekind's purpose with the utmost care; but so far was this purpose removed from what I had expected mathematics to be about that what he said had left no imprint whatever. Several years later, when I began to teach, I discovered that my state of youthful unpreparedness had been normal rather than exceptional and now, instead, a course in analysis started with a set of axioms summarising the basic properties of the real number system; the new hope was that these would reflect students' experience and intuition sufficiently to constitute an acceptable foundation on which to build. However, while this approach had the advantage of taking up less time, I doubt whether it was more successful. The difficulty now was that experience of number as gained at school was too chaotic and unselfconscious to be available for codifying; axiomatics are appropriate only when we are ready to abstract from well-digested experience—and readiness here means not only being willing to accept one set of axioms but also regarding the addition of further axioms as unacceptable. During the last twenty years reforms in the teaching of mathematics have made number concepts more explicit in the classroom; but as these reforms have gone hand in hand with a decline in actual number work, these concepts have rarely taken root. When I look now at the struggles of first-year students with their introductory analysis course I feel that in this context we have made little progress over the last thirty years. It is not that we have not tried to tackle the problem—it is much more that the problem of understanding number is genuinely difficult, not least because one meets it so early and in so unexpected a quarter.

The authors of *The Foundations of Mathematics* have set out to put matters right—and if the power of the printed word were what once it was, they might well succeed. With a relaxed and informal literary style they guide the reader from primitive number concepts and intuitive types of argument, through the meaning of proof and increasing awareness of the demands of logic, towards the formalisation of number systems and, in the very last section, of set theory. It seems to me that this is a book that aspiring students of mathematics might well start reading in the first year of the sixth form, discussing it (especially the exercises) among themselves and with their teachers; re-read during the summer before they enter University, have under the pillow during the first year at university, and still have by them for dipping into when preparing for finals. It is not a book to rush through; although the language is simple, many young people will be unaccustomed to see it harnessed to such ends and will need to be as patient in their reading as the authors have been in exposition. Moreover, as the use of logical symbols increases, young readers should be specially careful when assimilating it to their own use. All too often in undergraduate writings do logical symbols reveal only a lack of understanding. Indeed, this is one point where I wonder whether the authors would not have been wiser to restrict themselves more severely to the use of ordinary language, taking Halmos' 'Naive set theory' as a model in this respect.

The matter covered by this book evolved over more than two millenia, and the ideas basic to the most logical of subjects developed in an order that was anything but logical. Moreover, children's awareness of number tends to stumble through most of the historic pitfalls and confusions. This book, if studied early enough and well enough, could help students to a smoother entry into higher mathematics; and it should help teachers to understand why innumeracy is a disease that, up to now, has been so hard to cure.

University of Nottingham H. HALBERTSTAM

Mathematical Analysis. By M. D. HATTON. Hodder and Stoughton, Sevenoaks, Kent, 1977. Pp. 242. £4·75 hardback, £2·45 paperback.

The description of the book on the back of the cover claims

> 'This book provides an introduction to mathematical analysis for students who are familiar with elementary techniques of differential and integral calculus... the book should prove useful also to students to whom expert advice is not readily available'.

Students such as those cited above notoriously find analysis one of the most difficult subjects they have to study. They often start the subject without realising its purpose. Their elementary introduction to calculus has given them no hint of the nature of the problems about limiting processes which have demanded the construction of a rigorous deductive treatment; nor do they realise the lack of clarity and precision in their own concepts about limits. Calculus seems to students who have reached A-level largely to be concerned with properties of functions and their graphs; their A-level course has rightly given them no hint that it depends on subtle properties of the real numbers which were only put on a firm basis in the nineteenth century. It is the task of the 'Introduction to Mathematical Analysis', whether a book or a set of lectures, to bridge the gap between the student's ideas of calculus and the mathematician's awareness of the dependence of limiting processes on the real number system. Dr Hatton does not seem to appreciate the nature of this problem. He starts off with a chapter of 'Preliminaries' in which real numbers are defined as Dedekind sections of the rationals. This must come as rather a surprise to the student 'to whom expert advice is not readily available', when presented baldly and without explanation. Altogether the text is compressed and lacking in advice to the student as to whether a particular concept is important or not.

In fact, in this treatment Dedekind sections are important, as Dedekind's theorem on sections of the real numbers provides the basic completeness property of the reals. We are told 'this is one of the fundamental theorems of analysis', but there is no hint as to why it is important. It is only some pages later that it is used to prove the existence of the supremum of a function which is bounded above. And when the author comes to prove that a function continuous on a closed interval a, b is bounded, he assumes that a bounded *set* has a supremum although the supremum of a set has not previously been mentioned. How can the beginner be expected to recognize the essential equivalence of the two types of suprema?

Again, one of the commonest confusions among students is a lack of appreciation of the different roles of ε and δ in the definition of $\lim_{x\to a} f(x)$. But here the reader is given no guidance. Thus, the first worked example on the $\varepsilon - \delta$ method is set out as for a practised analyst with no commentary or explanation.

Any student who is so unfortunate as to need to study analysis when 'expert advice is not readily available' is unlikely to find this book's claim to be useful to him to be correct. Moreover, expert advice, if available, is likely to direct him to some more careful and sympathetic text.

Homerton College, Cambridge H. B. SHUARD

Notes on Contributors

H. Burkill is a Senior Lecturer in Pure Mathematics in the University of Sheffield. As an undergraduate in Cambridge he experienced J. E. Littlewood's inimitable style of lecturing. His principal mathematical interest is in classical analysis, but colleagues have also involved him in the younger subject of combinatorics.

John Alexander is a Lecturer at the University of Southampton and an Open University Tutor. He has previously taught in both a comprehensive school and a college of technology. He holds degrees in mathematics and social statistics, and his current interests lie in the field of medical statistics, with emphasis on applications in medical demography and the health service.

Cecilio Mar Molinero graduated from the University of Barcelona in mechanical and production engineering. Subsequently he obtained an M.Sc. degree in operational research from the University of Southampton. After several years' experience in management services, both in Spain and in the U.K., he is now a lecturer in the Department of Economics at Southampton University. His interests include statistical forecasting techniques, operational research and agricultural economics.

H. J. Godwin is a graduate of the University of Cambridge. His first university teaching post was in Swansea and from there, in 1968, he moved to Royal Holloway College (London University) to become Professor and Head of the joint Department of Statistics and Computer Science. Within pure mathematics, Professor Godwin's work has been largely in the theory of numbers and, in more recent years with the rigid development of high-speed computers, he has taken an especial interest in the numerous applications of computers to number theory.

Harry V. Smith took his first degree at Sir John Cass College, London, and obtained an M.Sc. by research into complex analysis from the University of Kent at Canterbury. After teaching at a grammar school he became a lecturer at a college of education. He is now a lecturer in the School of Mathematics and Computing at Leeds Polytechnic. Since his postgraduate student days his research interest has shifted to numerical analysis. For relaxation he indulges in judo and rambling.

Reprinting of Volume 1

After being unavailable for some time Volume 1 has now been reprinted. The price of the complete volume of two issues is £1·75 (US$3.20; $A2.80) and orders should be sent to

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

Please make remittances payable to 'Mathematical Spectrum'.

Contents

H. BURKILL 1 J. E. Littlewood (1885–1977)

J. R. ALEXANDER 10 Probability as an aid in social research: the randomised response technique

C. MAR MOLINERO 13 Negative chickens: an exercise in model building

J. J. GODWIN 18 Factorization and random numbers

H. V. SMITH 23 The determination of Easter

25 Letters to the Editor

28 Problems and Solutions

30 Book Reviews

32 Notes on Contributors

ISSN 0025–5653

PRICES (*postage included*)

Prices for Volume 11 (*Issues Nos.* 1, 2, *and* 3):

Subscribers in Britain and Europe: £1·30
Subscribers overseas: £2·60 (US$4.80; $A4.25)
(These prices apply even if the order is placed by an agent in Britain.)
A discount of 10% is allowed on all orders for five or more copies.

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.

Printed in England by Galliard (Printers) Ltd, Great Yarmouth