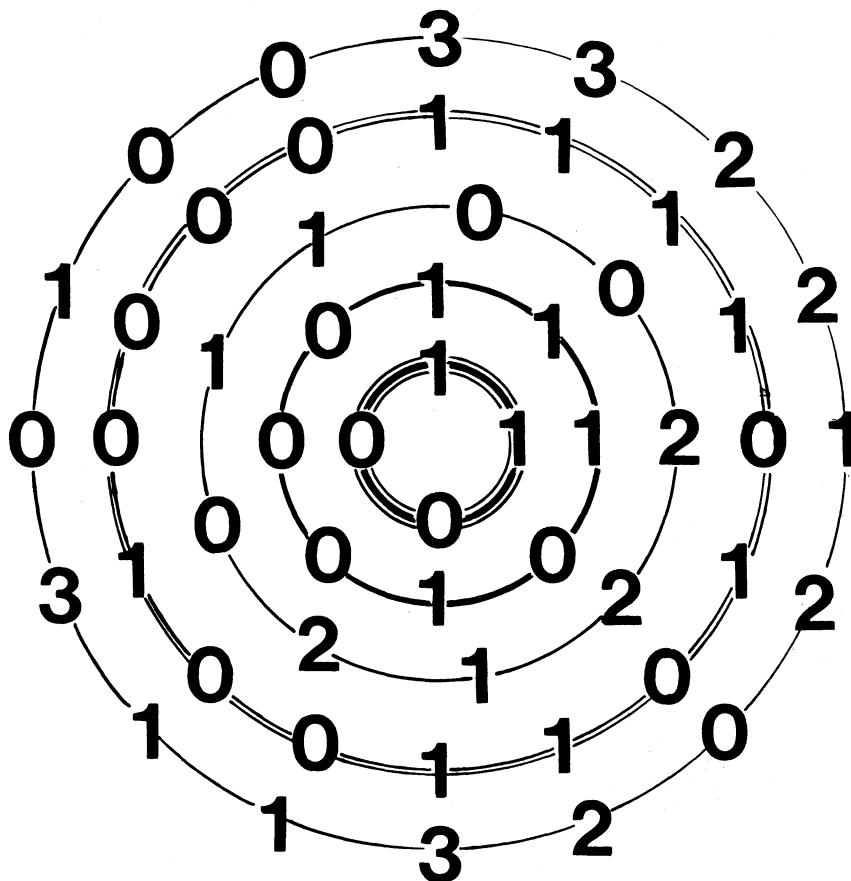


MATHEMATICS

Δ G Δ Z i N E



Vol. 55 No. 3
May, 1982

DE BRUIJN SEQUENCES • DE MORGAN SPOOF-PLAY
EQUATIONS OVER FINITE FIELDS • TIED RACES

SOME INFORMATIVE AND USEFUL BOOKS FROM THE MAA...

Browse through this list and see if your library is missing some of these important books published by the Association.

APPLICATIONS OF UNDERGRADUATE

MATHEMATICS IN ENGINEERING—written and edited by Ben Noble. A collection of articles prepared by engineers for the Committee on the Undergraduate Program in Mathematics. 364 pp. Hardbound. List: \$19.00. MAA Member: \$13.00.

ANNOTATED BIBLIOGRAPHY OF EXPOSITORY WRITING IN THE MATHEMATICAL SCIENCES,

prepared by M. P. Gaffney and L. A. Steen. An invaluable reference source of expository articles in mathematics. 282 pp. Paperbound. List: \$12.00. MAA Member: \$8.00.

AN ANNOTATED BIBLIOGRAPHY OF FILMS AND VIDEOTAPES FOR COLLEGE MATHEMATICS, by

David I. Schneider. An up-to-date listing of films and videotapes available for classroom use. 107 pp. Paperbound. List: \$9.00. MAA Members: \$6.00.

A BASIC LIBRARY LIST FOR TWO-YEAR COLLEGES,

prepared by the Committee on Basic Library Lists. A recommended library nucleus for two-year colleges. 66 pp. Paperbound. List: \$8.00. MAA Member: \$6.00.

A BASIC LIBRARY LIST FOR FOUR-YEAR

COLLEGES, prepared by CUPM. Presents listings of books and journals that should be in every college library. 106 pp. Paperbound. List: \$9.00. MAA Member: \$6.50

THE CHAUVENET PAPERS. *A Collection of Prize Winning Expository Papers in Mathematics*, edited by James C. Abbott. Two-volumes of the collected prize winning Chauvenet Papers. Vol. 1—312 pp. Hardbound. Vol. 2—282 pp. Hardbound. List: \$21.00 each. MAA Member \$16.00 each.

[Two volume sets
List: \$36.00. MAA Member: \$27.00.]

CRITICAL VARIABLES IN MATHEMATICS

EDUCATION: *Findings from a Survey of the Empirical Literature*, by E. G. Begle. A joint publication of the MAA and the National Council of Teachers of Mathematics. List: \$8.00. MAA Member: \$6.40.

A COMPENDIUM OF CUPM RECOMMENDATIONS.

Volumes I and II. A collection of the major recommendations of the Committee on the Undergraduate Program in Mathematics. Two volumes. 756 pp. Hardbound. List: \$16.50. MAA Member: \$12.00.

THE WILLIAM LOWELL, PUTNAM MATHEMATICAL COMPETITION: PROBLEMS AND SOLUTIONS—

1938-1964. Compiled the R. E. Greenwood, A. M. Gleason, and L. M. Kelly. Contains problems and solutions to the first 25 Putnam Exams. 652 pp. Hardbound. List: \$35.00. MAA Member: \$26.00.

THE MATHEMATICAL ASSOCIATION OF AMERICA:

Its First Fifty Years. An historical perspective of the Association. 170 pp. Hardbound. List: \$10.00. MAA Member: \$5.00.

FIFTY YEAR INDEX OF THE MATHEMATICS

MAGAZINE, edited by Lynn A. Steen, and J. Arthur Seebach. Cumulative index of volumes 1-50. 163 pp. paperbound. List: \$10.00. MAA Member: \$6.50.

INDEX OF THE AMERICAN MATHEMATICAL

MONTHLY. Contains the tables of contents for each issue of volumes 1-80 as well as subject and author indices. 269 pp. Hardbound. List: \$19.00. MAA Member: \$13.00.

PRIME-80. *Proceedings of a Conference on*

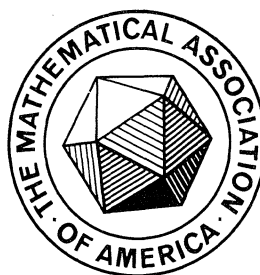
Prospects in Mathematics Education in the 1980's. Included are the background of the conference and the recommendations that resulted. 84 pp. Paperbound. \$3.50.

PROFESSIONAL OPPORTUNITIES IN THE

MATHEMATICAL SCIENCES, Tenth Edition. 1978. Designed for the student interested in a career in mathematics. 35 pp. Paperbound. \$1.50 each. 95¢ for orders of five or more.

RECOMMENDATIONS ON A GENERAL

MATHEMATICAL SCIENCES PROGRAM. Prepared by the Committee on the Undergraduate Program in Mathematics. (CUPM) 102 pp. Paperbound. \$3.50 each.



Order From:

THE MATHEMATICAL ASSOCIATION OF AMERICA

1529 Eighteenth Street, N.W.
Washington, D. C. 20036

EDITOR

Doris Schattschneider
Moravian College

ASSOCIATE EDITORS

Edward J. Barbeau, Jr.
University of Toronto

John Beidler
University of Scranton

Paul J. Campbell
Beloit College

Underwood Dudley
DePauw University

G. A. Edgar
Ohio State University

Joseph A. Gallian
Univ. of Minnesota, Duluth

Judith V. Grabiner
Calif. St. U., Dominguez Hills

Raoul Hailpern
SUNY at Buffalo

Joseph Malkevitch
York College of CUNY

Pierre J. Malraison, Jr.
MDSI, Ann Arbor

Leroy F. Meyers
Ohio State University

Jean J. Pedersen
University of Santa Clara

Gordon Raisbeck
Arthur D. Little, Inc.

Ian Richards
University of Minnesota

Eric S. Rosenthal
West Orange, NJ

David A. Smith
Duke University

EDITORIAL ASSISTANT

Dianne E. Chomko

ARTICLES

131 De Bruijn Sequences—A Model Example of the Interaction of Discrete Mathematics and Computer Science, *by Anthony Ralston.*

144 Why Study Equations over Finite Fields?, *by Neal Koblitz.*

NOTES

150 Historical Roots of Confusion Among Beginning Algebra Students: A Newly Discovered Manuscript, *by Helena M. Pycior.*

157 A Genealogy of 120° and 60° Natural Triangles, *by Alan Wayne.*

162 Why Your Classes Are Larger than “Average,” *by David Hemenway.*

164 Kirchhoff’s Third Law, *by Marlow Sholander.*

165 The Identity $(XY)^n = X^n Y^n$: Does It Buy Commutativity?, *by Howard E. Bell.*

170 Races with Ties, *by Elliott Mendelson.*

176 A Cross-Number Puzzle, *by Nick Franceschini III.*

PROBLEMS

177 Proposals Number 1144–1148.

178 Quickies Number 672, 673.

178 Solutions to Problems 1116–1120.

183 Answers to Quickies 672, 673.

REVIEWS

184 Reviews of recent books and expository articles.

NEWS AND LETTERS

187 Comments on recent issues; news; solutions to the 1981 Putnam Exam.

COVER: Five de Bruijn sequences. See p. 131 ff. and editor’s note p. 143. Design by the editor.

EDITORIAL POLICY

Mathematics Magazine is a journal of collegiate mathematics which aims to provide inviting, informal mathematical exposition of interest to undergraduate students. Manuscripts accepted for publication in the *Magazine* should be written in a clear and lively expository style and stocked with appropriate examples and graphics. Our advice to authors is: say something new in an appealing way or say something old in a refreshing way. The *Magazine* is not a research journal and so the style, quality, and level of articles submitted for publication should realistically permit their use to supplement undergraduate courses. The editor invites manuscripts that provide insight into the history and application of mathematics, that point out interrelationships between several branches of mathematics and that illustrate the fun of doing mathematics.

New manuscripts should be sent to: Doris Schattschneider, Editor, *Mathematics Magazine*, Moravian College, Bethlehem, Pennsylvania 18018. Manuscripts should be prepared in a style consistent with the format of *Mathematics Magazine*. They should be typewritten and double spaced on 8½ by 11 paper. Authors should submit the original and one copy and keep one copy as protection against possible loss. Illustrations should be carefully prepared on separate sheets of paper in black ink, the original without lettering and two copies with lettering added.

Authors planning to submit manuscripts should read the full statement of editorial policy which appears in the News and Letters section of this *Magazine*, Vol. 54, pp. 44–45. Additional copies of the policy are available from the Editor.

BUSINESS INFORMATION. *Mathematics Magazine* is published by the Mathematical Association of America at Washington, D.C., five times a year in January, March, May, September, and November. The annual subscription price for *Mathematics Magazine* to an individual member of the Association is \$8, included as part of the annual dues of \$40. Students receive a 50% discount. Bulk subscriptions (5 or more copies to a single address) are available to colleges and universities for distribution to undergraduate students at a 35% discount. The library subscription price is \$25.

Subscription correspondence and notice of change of address should be sent to A. B. Willcox, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N. W., Washington, D.C. 20036. Back issues may be purchased, when in print, from P. and H. Bliss Co., Middletown, Connecticut 06457.

Advertising correspondence should be addressed to Raoul Hailpern, Mathematical Association of America, SUNY at Buffalo, Buffalo, New York 14214.

Copyright © by The Mathematical Association of America (Incorporated), 1982, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. Reprint permission should be requested from Doris Schattschneider, Editor, Moravian College, Bethlehem, PA 18018.

General permission is granted to Institutional Members of the MAA for non-commercial reproduction in limited quantities of individual articles (in whole or in part), provided a complete reference is made to the source.

Second class postage paid at Washington, D.C., and additional mailing offices.

AUTHORS

Neal Koblitz ("Why Study Equations over Finite Fields?") is Associate Professor of Mathematics at the University of Washington. He received his Ph.D. from Princeton, where he studied algebraic geometry and number theory. From 1975 to 1979 he was a Benjamin Peirce Assistant Professor at Harvard. In 1974–75 and in spring 1978, Koblitz studied at Moscow University and the Steklov Institute, and in summer 1978 he gave a series of lectures at the Hanoi Mathematical Institute. He is the author of two books and several research articles on algebraic number theory and arithmetic algebraic geometry.

Anthony Ralston ("De Bruijn Sequences—A Model Example of the Interaction of Discrete Mathematics and Computer Science") is professor of computer science at SUNY at Buffalo. His degrees are all in mathematics and most of his early research was in numerical analysis. About ten years ago he turned to more strictly computer science pursuits (i.e., programming languages) but in recent years, has "returned" to mathematics where his current interests are in discrete mathematics and its place in undergraduate mathematics education.

ILLUSTRATIONS

Blaze Gallo sketched the scenes on pp. 162, 164.

Julia Mendelson sketched the photo finish on p. 171.

The sketch of **De Morgan** teaching (p. 151) is reproduced with the permission of University College London Library.

All other illustrations were provided by the authors.

De Bruijn Sequences—A Model Example of the Interaction of Discrete Mathematics and Computer Science

Combinatorics, graph theory, and abstract algebra can all be applied to the same algorithmic problem.

ANTHONY RALSTON

SUNY at Buffalo

Amherst, NY 14226

1. An exemplary problem

The mathematical tools needed in computer science are overwhelmingly those of discrete mathematics—combinatorics, graph theory, algebra and the other branches of mathematics which focus on discrete objects rather than continuous functions. It is these branches of mathematics which are crucial to the design and analysis of so many algorithms and it is algorithms which are the lifeblood of computer science. The problem we will discuss aptly illustrates the interaction of discrete mathematics and computer science. The problem is as follows.

PROBLEM. *Given $m + 1$ symbols (which, without loss of generality, we take to be $0, 1, 2, \dots, m$) and a positive integer n , find an algorithm to generate a sequence of the symbols having minimum length, which when arranged on a circle, contains as subsequences of consecutive symbols, every sequence of length n of the symbols.*

Since each of the $m + 1$ symbols may be repeated as often as desired in the sequences of length n , there are

$$(m + 1)^n \quad (1)$$

possible subsequences. We shall denote the sequence to be generated by any algorithm as

$$S = s_1 s_2 \dots s_L \quad 0 \leq s_i \leq m. \quad (2)$$

(We shall use S in this paper as a general-purpose symbol to represent sequences of the s_i .) A subsequence of consecutive symbols of length n or, an n -sequence, is just a string of the form

$$s_i s_{i+1} \dots s_{i+n-1}, \quad (3)$$

where we use the convention that the subscripts in (3) are to be taken modulo L . That is, we allow n -sequences of the form

$$s_{L-j} s_{L-j+1} \dots s_L s_1 s_2 \dots s_{n-j-1} \quad j = 0, 1, \dots, n-2 \quad (4)$$

which “wrap around” from the end of the sequence S to the beginning. Thus, any circular permutation of a sequence S which solves our problem is also a solution to the problem.

As an example, for $m = n = 2$, the sequence

$$221201100 \tag{5}$$

contains all possible sequences of length 2 of the symbols 0, 1, 2 (i.e., 22, 21, 20, 12, 11, 10, 02, 01, 00). Note that in (5), the subsequence 02 has the form (4) with $j = 0$.

Each s_i , $i = 1, \dots, L$ in S defines the beginning of an n -sequence because of our stipulation allowing wraparound. Therefore, the minimum possible value of L is given by (1) if S is to contain all possible n -sequences. A sequence S which solves our problem and for which L has the value (1) is called a **de Bruijn sequence**. (These have been called by various other names in the literature. Recently Fredricksen [9] has proposed that they be called *full cycles*.) The sequence (5) is such a sequence. As we shall see, de Bruijn sequences exist for all m and n .

In the next section we note some of the history and importance of the de Bruijn sequence problem, and then, in Section 3, discuss combinatorial, graph theoretic, and abstract algebraic approaches to the problem. In Section 4 we consider various algorithms for generating de Bruijn sequences. Finally, in Section 5 we shall discuss some implications of the growing importance of discrete mathematics for the mathematical education of computer science students and for the undergraduate mathematics curriculum.

2. History and applications

Like many problems which can be approached in a variety of different ways, the de Bruijn sequence problem has been “discovered” many times. For $m = 1$ the problem appears to have been first proposed in 1894 by A. de Rivière as just that—a problem—in the French problem journal, *l'Intermédiaire des Mathématiciens*. It was solved in that journal the same year by C. Flye Sainte-Marie [7] using a graphical method and three years later by W. Mantel [18] who, using an algebraic method, found a solution valid whenever $m + 1$ is prime. (See [5] and [9] for more of this history.) The earliest mention in the modern literature of de Bruijn sequences appears to have been in a paper by Martin [19] for whom the motivation was an unspecified “problem in dynamics.” He approached the problem combinatorially and proved the existence of de Bruijn sequences for all m and n by exhibiting an algorithm to construct such sequences. A decade after Martin, de Bruijn [4] and Good [13] independently “discovered” and solved the problem for the case $m = 1$ in graph theoretic terms. Their results are easily extended to any m . In a commentary on Good’s paper, Rees [26] approached the problem from the viewpoint of fields of prime characteristic and irreducible polynomials over such fields. Since the problem became generally known through de Bruijn’s 1946 paper, it has since been associated with his name.

Rees’ approach provides a general algorithm for the generation of de Bruijn sequences, but for $m > 1$ and, particularly, when $m + 1$ is not prime, it is a difficult algorithm to implement. For the case $m = 1$, however, this approach leads naturally to algorithms based on shift register techniques as described by Eldert et al. [6] (see also Golomb [12] and Fredricksen [9]). This approach is also discussed by Knuth [15] because of its relation to the problem of generating random sequences of binary digits.

The problem was discovered again, without any indicated motivation, by Roth [27] who used a combinatorial approach to develop an algorithm to generate de Bruijn sequences but his algorithm, like Martin’s, requires $(m + 1)^n$ units of memory (one for each digit of the sequence). Algorithms without this restriction for the case $m = 1$ have been found by Fredricksen [8] and Fredricksen and Kessler [10] and a related algorithm for general m was derived by Fredricksen and Maiorana [11]. Another such algorithm was found by Ralston [24]. Fredricksen’s paper [9] contains a general review of combinatorial algorithms for this problem.

While the graph theoretic approach leads to some elegant results about de Bruijn sequences, it does not lead to useful algorithms for the generation of these sequences. Good general discussions of de Bruijn sequences and their properties can be found in Hall [14] and van Lint [16].

As the preceding discussion implies, our problem is related to some applications of discrete mathematics, but the inherent mathematical interest of the problem rather than the applicability

of its solution has motivated most of the work on it. This certainly reflects our attitude here in presenting the problem as a paradigm of the application of discrete mathematics to algorithmics.

3. The existence of de Bruijn sequences

The combinatorial argument. The main idea of this approach is to generate a sequence one symbol at a time so that no n -sequence is ever repeated. A logical way to try to achieve this is to always add the *largest* symbol such that the resulting new n -sequence has not appeared previously. This is essentially the idea of Martin [19] and leads to the following algorithm:

Algorithm M

Step 1: Start with the sequence of n zeros.

Step 2: (Iterative Step) Append to the sequence S already generated the largest symbol possible such that the newly formed n -sequence (i.e., the last n symbols of S) has not already appeared.

Step 3: When Step 2 can no longer be repeated, remove the last $n - 1$ symbols of the sequence generated by Step 2.

Steps 1 and 2 of Algorithm M ignore the wraparound sequences discussed in Section 1, so when Step 2 terminates, the result cannot be a de Bruijn sequence. However, we shall show that Step 3 is sufficient to then produce a de Bruijn sequence. We first prove:

LEMMA 1. *Let S , the sequence generated at any stage, be denoted by*

$$S = s_1 s_2 \dots s_j.$$

Then, if at least one of the $n - 1$ symbols $s_{j-n+2}, s_{j-n+3}, \dots, s_{j-1}, s_j$ is not zero, Step 2 of Algorithm M may be applied to add s_{j+1} to S .

Proof. It is not possible to apply Step 2 only when the sequence of $n - 1$ symbols

$$s_{j-n+2} s_{j-n+3} \dots s_j \quad (6)$$

has appeared $m + 1$ times previously in S , each time followed by a different one of our $m + 1$ symbols. Thus, (6) would represent the $(m + 2)$ th appearance of this sequence. But one of the symbols in (6) is not zero, so this sequence cannot be the same as $s_1 s_2 \dots s_{n-1}$ since, by Step 1, these are all zeros. Thus, the $m + 2$ appearances of (6) must each have been *preceded* by a different one of our symbols. Since this is impossible, the lemma is proved.

Lemma 1 implies that Step 2 of Algorithm M can terminate when and only when the last $n - 1$ symbols in S are all zero. And this happens only on the $(m + 2)$ th occurrence of such a sequence.

Let us denote by

$$s_1^{j_1} s_2^{j_2} \dots s_i^{j_i}$$

the sequence of j_1 instances of s_1 followed by j_2 instances of s_2 , etc. Since we have shown that the sequence S generated by Algorithm M (before Step 3) has $m + 2$ occurrences of 0^{n-1} (the first of which has no predecessor), it follows that S contains all sequences of the form

$$s 0^{n-1} \quad s = 0, 1, 2, \dots, m.$$

Since, according to Step 2, $s 0^{n-2}$ is succeeded by 0 only when no larger digit can be appended, it follows that S also contains all sequences of the form

$$s 0^{n-2} t \quad s, t = 0, 1, 2, \dots, m. \quad (7)$$

We can now prove:

THEOREM 1. *When Step 2 of Algorithm M terminates, each possible n -sequence appears once and only once in the sequence S generated.*

Proof. That no sequence can occur more than once follows immediately from Step 2 of the algorithm.

Now let

$$T=t_1t_2\ldots t_n \tag{8}$$

be an arbitrary sequence of n symbols such that

$$t_2\ldots t_{n-1}\neq 0^{n-2}, \tag{9}$$

since we have shown in (7) that all sequences (8) with inequality replaced by equality in (9) do occur. To show that T appears in S , it is enough, by Step 2, to show that

$$U=t_1t_2\ldots t_{n-1}0 \tag{10}$$

appears. Suppose U does not appear so that $t_2\ldots t_{n-1}0$ occurs at most m times in S . But, therefore, again by Step 2,

$$t_2\ldots t_{n-1}0^2 \tag{11}$$

cannot be in S since the largest possible symbol is always chosen to follow $t_2\ldots t_{n-1}0$. Still under the assumption that U does not appear, it follows from (7) that, in $t_3\ldots t_{n-1}$, there must be a nonzero symbol since otherwise (11) would be in S . Applying the same argument as before with (11) now playing the role of (10), it follows that $t_3t_4\ldots t_{n-1}0^2$ appears, at most, m times in S and, therefore, $t_3t_4\ldots t_{n-1}0^3$ cannot occur in S . Continuing in this way we get eventually that $t_{n-1}0^{n-1}$ cannot appear, which contradicts the fact that all sequences of the form (7) do appear. This contradiction assures that U is in S and proves the theorem.

Finally, if we apply Step 3 of Algorithm M, the result is a de Bruijn sequence. This follows from Theorem 1 and the fact that deleting the final 0^{n-1} but allowing wraparound results in the original and new sequences having the same n -sequences. Therefore, we have proved the existence of de Bruijn sequences for all m and n .

For a given m and n , all de Bruijn sequences have the same length $L=(m+1)^n$, so that we can order this set of sequences lexicographically. (That is, if $S=s_1s_2\ldots s_L$ and $T=t_1t_2\ldots t_L$, we say $S\geq T$ if $s_i=t_i$, $0\leq i\leq j<L$ and $s_{j+1}>t_{j+1}$; $S=T$ only if $s_i=t_i$, $0\leq i\leq L$.)

We note that if the de Bruijn sequence generated by Algorithm M has its n initial zeros moved to the end, then the construction in Step 2 implies that the resulting sequence is greatest (with respect to the lexicographic ordering) among all de Bruijn sequences for a given m and n . We will later refer to this particular sequence as B_{mn} .

The graph theoretic argument. The idea behind a graphical approach to the PROBLEM (the approach used by de Bruijn [4]) is to define a digraph (directed graph) whose edges are each labeled with one of the $m+1$ symbols and which has a path such that the labels of successive

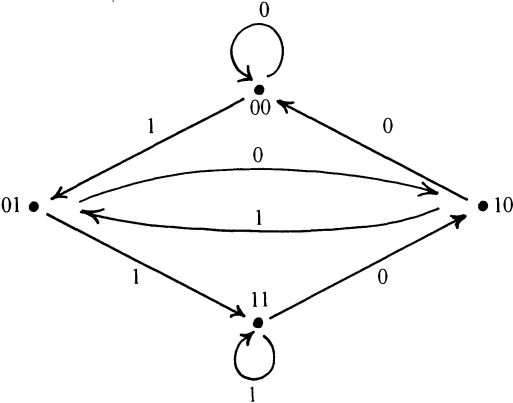


FIGURE 1. A de Bruijn graph for $m=1, n=3$.

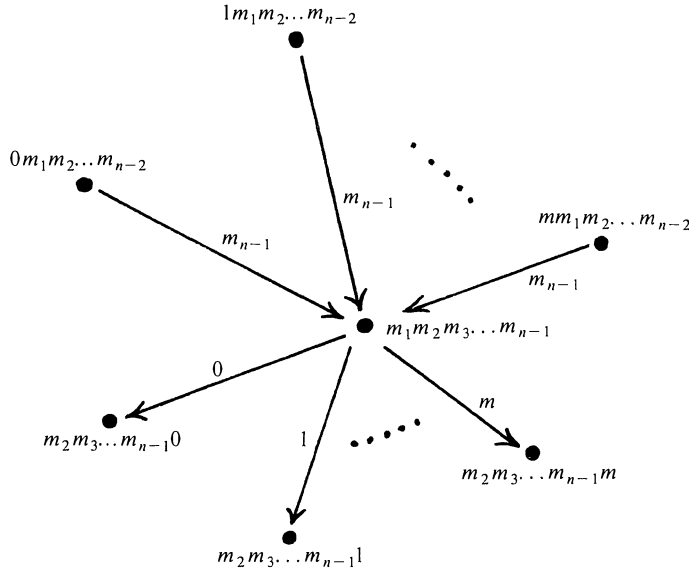


FIGURE 2. Part of a digraph for generating de Bruijn sequences.

edges in the path form the desired sequence. For $m = 1$ and $n = 3$ such a digraph is shown in FIGURE 1 with each node labeled by an $(n - 1)$ -sequence. Starting at the node labeled 00 and proceeding successively to the nodes labeled 01, 11, 11, 10, 01, 10, 00, 00, the corresponding edge labels form the sequence

11101000

which is a de Bruijn sequence.

Generalizing this idea, we can define a digraph G which will lead to a solution to our PROBLEM. Recall that the number of edges directed into (away from) a node is called the indegree (outdegree) of that node. G is defined by the following conditions:

- (a) G has $(m + 1)^{n-1}$ nodes each of which is labeled with a distinct $(n - 1)$ -sequence of our $m + 1$ symbols.
- (b) Each node $m_1 m_2 \dots m_{n-1}$ has indegree and outdegree $m + 1$ with the edges out to nodes $m_2 m_3 \dots m_{n-1} k$, $k = 0, 1, \dots, m$ labeled, respectively, $0, 1, 2, \dots, m$ and with the edges in from nodes $k m_1 m_2 \dots m_{n-2}$, $k = 0, 1, \dots, m$ all labeled m_{n-1} , the last symbol on the label of the node itself.

FIGURE 2 illustrates all of the edges which meet a single node of this digraph G .

An applicable, well-known theorem in graph theory (for instance, see [20]) is:

THEOREM 2. *A connected, directed graph has an Eulerian cycle (i.e., a path which traverses every edge once and only once and ends at the node at which it begins) if and only if each node has the same indegree and outdegree.*

Since our digraph G satisfies these conditions, it has an Eulerian cycle. Now suppose we have such an Eulerian cycle which starts at some node N . Suppose also that we construct a sequence S as follows:

Algorithm G

Step 1: Set S equal to the empty sequence.

Step 2: (Iterative Step) As each edge of the Eulerian path is traversed, append the label of the edge to S .

Since the graph G has $(m + 1)^{n-1}$ nodes and since each node has $m + 1$ edges emanating from it, G has $(m + 1)^n$ edges. Therefore, the sequence S generated by Algorithm G has $(m + 1)^n$ symbols. Since, additionally, the $m + 1$ edges leading from this node are each traversed once and only once in an Eulerian cycle, a unique n -sequence is formed each time a new edge is traversed. To see this note that the construction assures that, at any time, the last $n - 1$ symbols added to S are the unique label of the node just arrived at (using the wraparound for the first $n - 1$ nodes). Therefore, S is a de Bruijn sequence. The above is easily summarized by:

THEOREM 3. *Algorithm G generates a de Bruijn sequence for any m and n .*

If we choose our Eulerian cycle so that the initial node has the label 0^{n-1} and so that the first edge traversed is the loop at this node and the next edge chosen is always the one with the greatest label not yet traversed, then Algorithm G generates the same de Bruijn sequence (although a circularly permuted one) as Algorithm M. But what if we choose Eulerian cycles other than this one? For given m and n , how many distinct de Bruijn sequences are there up to a circular permutation? This combinatorial question is answered by the formula

$$(m + 1)^{-n} [(m + 1)!]^{(m+1)^{n-1}}. \tag{12}$$

For any n and $m = 1$, there are $2^{2^{n-1}-n}$ distinct de Bruijn sequences, which, even for $n = 5$, has a value of 2048. For $m = n = 2$, there are 23 de Bruijn sequences in addition to (5); all 24 sequences are listed in TABLE 1.

221201100	221100201	220211001	220021101
221200110	221100120	220210011	220021011
221120100	221020011	220121100	220012110
221120010	221011200	220112100	220011210
221102001	221002011	220110021	220011021
221101200	221001120	220100211	220010211

TABLE 1. The 24 de Bruijn sequences for $m = n = 2$.

The derivation of (12) is beyond us here. It involves the notion of *spanning trees* of a digraph, counting the number of spanning trees of graphs of the form G and then counting the number of Eulerian cycles of a graph as a function of the number of spanning trees. To do this is a major exercise in matrix algebra involving, in particular, the computation of minors and determinants of a matrix of order $(m + 1)^{n-1}$. I might note that the understanding of this derivation is a good test of a student's knowledge of matrix algebra. Van Lint [16] gives a rather terse version of the derivation.

The finite field approach. The use of recurrence relations is a well-known way of generating sequences. If the magnitude of the terms generated by a recurrence relation is bounded, then the *period* of the sequence generated must be finite. To achieve periods of maximum length using recurrence relations is one of the aims of random number generators used on computers. From our definition, a de Bruijn sequence is the sequence of maximum length with the property that there is no repeated n -sequence. These considerations lead to a recurrence relation approach to our problem and, from it, to the theory of finite fields.

For this discussion, we assume $m + 1 = p$, a prime. The basic idea is, given $s_0 \neq 0$, to generate a sequence $\{s_i\}$ by using a recurrence relation

$$s_i = a_1 s_{i-1} + a_2 s_{i-2} + \cdots + a_n s_{i-n} \bmod p, \quad i = 1, 2, \dots, \tag{13}$$

where $s_j = 0$ if $j < 0$, the coefficients a_i satisfy $0 \leq a_i \leq m$ for all i and all arithmetic is performed modulo p .

How long is the period of the sequence generated by (13)? We prove first

LEMMA 2. The first n -sequence $s_j s_{j+1} s_{j+n-1}$ to recur in the sequence $\{s_i\}$ generated by (13) is $s_0 s_1 \dots s_{n-1}$.

Proof. Suppose that

$$s_j s_{j+1} \dots s_{j+n-1}, \quad j \neq 0$$

recurs before $s_0 s_1 \dots s_{n-1}$ so that the sequence $\{s_i\}$ has the form $s_0 s_1 \dots s_j \dots s_{j+n-1} \dots s_j \dots s_{j+n-1} \dots$. But the digit preceding each s_j must be the same since arithmetic modulo a prime implies that (13) uniquely determines s_{j-1} from s_j, \dots, s_{j+n-1} . This contradicts the assumption about $s_j s_{j+1} \dots s_{j+n-1}$. Working back to s_0 , the lemma follows.

The sequence S generated by (13), therefore, has the form

$$s_0 s_1 \dots s_j s_0 s_1 \dots s_j s_0 s_1 \dots \quad (14)$$

with j the period of the "repeat." Now let

$$P(x) = 1 - a_1 x - a_2 x^2 - \dots - a_n x^n \quad (15)$$

with the a_i 's as in (13). Then, if we let $s_0 = 1$ and define

$$S(x) = 1 + \sum_{i=1}^{\infty} s_i x^i, \quad (16)$$

$S(x)$ is the formal power series expansion of $1/P(x)$ with arithmetic modulo p . This is easily shown by multiplying (15) and (16) and using (13). We may now prove

LEMMA 3. $P(x)$ divides $1 - x^k$ if and only if k is a multiple of the period of the sequence (14).

Proof.

$$\begin{aligned} \frac{(1 - x^k)}{P(x)} &= (1 - x^k)S(x) \\ &= s_0 + s_1 x + s_2 x^2 + \dots + s_k x^k + s_{k+1} x^{k+1} + \dots \\ &\quad - s_0 x^k - s_1 x^{k+1} - s_2 x^{k+2} + \dots \end{aligned} \quad (17)$$

But, if $P(x)$ divides $1 - x^k$, then the polynomial on the right-hand side of (22) must have degree $k - n$. Thus

$$s_i = s_{k+i} \quad i = 0, 1, \dots,$$

Since the truth of the above relation implies, from (17), that $P(x)$ divides $1 - x^k$, this proves the lemma. Note also that

$$s_{k-n+1} = s_{k-n+2} = \dots = s_{k-1} = 0$$

in order for the polynomial to have the correct degree.

We now need the following theorem from the theory of finite fields (Albert [2], Berlekamp [3]):

THEOREM 4. There exists an irreducible polynomial $P(x)$ of degree n with coefficients in a finite field of characteristic p which divides the polynomial $1 - x^{p^n-1}$ and such that no polynomial of the form $1 - x^j$ having degree less than $p^n - 1$ is a multiple of $P(x)$.

If we identify a polynomial $P(x)$ satisfying Theorem 4 with (15), then by Lemma 3, $p^n - 1$ must be the period of S . Of course, the period of S could not have been greater than $p^n - 1$, since there are only p^n possible n -sequences and the sequence 0^n cannot be in S (for otherwise, from (13), it would repeat immediately). Thus we have shown that the sequence generated by (13) contains all strings of length n of the digits $0, 1, \dots, p-1$ except the string of n zeros. Thus the first period of the sequence generated by (13) may be converted to a de Bruijn sequence by inserting a zero at some place where there is a sequence of $n-1$ zeros.

We have shown that equation (13) is effectively a recurrence relation for generating de Bruijn sequences. In algorithmic form we have

Algorithm F (assumes $m + 1 = p$, a prime)

Step 1: Set $s_0 = 1$.

Step 2: (Iterative Step) Use (13) to generate s_i , $i = 1, 2, \dots, p^n - 2$, where the a_i are coefficients of a polynomial $P(x)$ satisfying Theorem 4.

Step 3: Insert a zero in the sequence next to a subsequence of $n - 1$ consecutive zeros.

A point worth noting here is that, when $n = 1$, the recurrence relation (13) is of the same form as the relation used for the multiplicative congruential generation of a sequence $\{r_i\}$ of random numbers. This relation is

$$r_i \equiv ar_{i-1} \bmod \alpha,$$

where α is often chosen to be 2^β and β is the (binary) word length of the computer on which the random numbers are to be generated.

To see an example of the generation of de Bruijn sequences by Algorithm F, let $m = 1$ and $n = 4$. In this case, the polynomial $P(x) = 1 - x - x^4$ is irreducible and divides $1 - x^{15}$ but does not divide $1 - x^k$ for any $k < 15$. Now, using (13) with $s_0 = 1$, $a_1 = 1$, $a_2 = 0$, $a_3 = 0$, $a_4 = 1$, we obtain the sequence

$$1111010110010000 \quad (18)$$

with the final zero added to make a de Bruijn sequence. For the case $m = 2$ and $n = 2$, the polynomial $P(x) = 1 - x - x^2$ is irreducible and divides $1 - x^8$ but does not divide $1 - x^k$ for $k < 8$. Applying Algorithm F, we obtain the de Bruijn sequence

$$112002210 \quad (19)$$

in which the second of the two consecutive 0's has been added (per Step 3).

Two questions remain:

1. How do we find the coefficients a_i which satisfy the conditions of Theorem 4?
2. How can we extend Algorithm F to the case when $m + 1$ is not prime?

For an answer to the first question, see Knuth [15] or Alanen and Knuth [1]. For the second question, we need the following two lemmas due to Rees [26] which we present without proof.

LEMMA 4. *If p is prime and q is a power of p , we can construct a de Bruijn sequence with $m + 1 = p^q$ and any n as follows:*

- (a) *Use (13) with $m + 1 = p$ and $n' = qn$ to generate a sequence $S = s_0s_1s_2\dots$ and then*
- (b) *Construct $T = t_0t_1t_2\dots$ where*

$$t_i = s_{qi} + s_{qi+1}p + s_{qi+2}p^2 + \dots + s_{q(i+1)-1}p^{q-1} \quad (20)$$

so that $0 \leq t_i \leq p^q - 1$. Take the first $(p^q)^n - 1$ symbols of T and insert a zero next to a subsequence of $n - 1$ zeros.

If q is not a power of p , a modest modification of the procedure results in a de Bruijn sequence.

As an example, let $p = q = n = 2$ so that $n' = 4$ and $m = 1$. We use for S two periods of (18) (less its final zero) to get

$$1111 \ 0101 \ 1001 \ 0001 \ 1110 \ 1011 \ 0010 \ 00$$

and then use (20) with $t_i = s_{2i} + 2s_{2i+1}$, to obtain

$$332212023113010 \quad (21)$$

which contains all sequences of length $n = 2$ (except 00) of 0, 1, 2, and 3. To obtain a de Bruijn sequence we just insert a 0 at the end.

The next lemma indicates how to generate de Bruijn sequences for any n and $m + 1 = p_1p_2$ where p_1 and p_2 are primes or powers of primes.

LEMMA 5. Given positive integers n, p_1 and p_2 , with p_1 and p_2 relatively prime, we can construct a de Bruijn sequence with $m + 1 = p_1 p_2$ and n as follows.

Let the de Bruijn sequences generated by Algorithm F (or Lemma 4), corresponding to p_1 and p_2 , respectively, be $S_1 = s_{01}s_{11}s_{21}\dots$ and $S_2 = s_{02}s_{12}s_{22}\dots$. Then the desired sequence is given by

$$T = t_0 t_1 t_2 \dots$$

with

$$t_i = s_{i1}p_2 + s_{i2}. \quad (22)$$

To generate a de Bruijn sequence for any m and n we need only find the prime factorization of $m + 1$, use Lemma 4 to generate a de Bruijn sequence for each prime factor and n and then use Lemma 5 repeatedly.

As an example, let $m = 11$ and $n = 2$ so that $m + 1 = 2^2 \cdot 3$. For the 2^2 factor we first find the de Bruijn sequence with $m = 1$, $n = 4$ as given by (18) and then the sequence with $m = 3$ ($= 2^2 - 1$) and $n = 2$ as given by (21) with an additional zero appended. For the 3 factor, a de Bruijn sequence with $m = 2$ ($= 3 - 1$) and $n = 2$ is given by (19). Then, combining (19) and (21) and using A for 10 and B for 11, we apply (22) with $p_2 = 3$ to get the first digits of the 144 digit sequence for $m = 11$ and $n = 2$ as:

$$\begin{array}{cccccccc} 3322 & 1202 & 3113 & 0100 & 3322 & 1202 & 3113 & 0100 \\ 1120 & 0221 & 0112 & 0022 & 1011 & 2002 & 2101 & 1200 \\ & & & \downarrow & & & & \\ AA86 & 3827 & 944B & 0322 & A977 & 5608 & B43A & 1500 \end{array}$$

4. Combinatorial algorithms for generating de Bruijn sequences

Although all three approaches discussed in Section 3 produce a de Bruijn sequence for a given m and n , Algorithm F has the disadvantage of requiring a calculation of the coefficients a_1, \dots, a_n of a polynomial $P(x)$ satisfying Theorem 4. (For m not prime, several of these calculations may be required.) The combinatorial approach, represented by Algorithm M, poses fewer computational difficulties, and has been the one most commonly used in practice. In this section we focus on other combinatorial algorithms which generate de Bruijn sequences.

We begin by noting that Algorithm M may be extended in a simple and elegant way to generate *all* de Bruijn sequences for any m and n . This extension involves one of the most powerful ideas in algorithmics, namely, **backtracking**.

Algorithm B

Step 1: Start with the de Bruijn sequence B_{mn} generated by Algorithm M with the initial n zeros moved to the end; $B_{mn} = s_1 s_2 \dots s_L$ with $L = (m + 1)^n$. Find the largest $j < L$ such that $s_j \neq 0$. Let $S = s_1 s_2 \dots s_{j-1}$.

Step 2: Apply Step 2 of Algorithm M to S (except that at the first step look for the largest symbol *less than* s_j) until either

- (a) a new de Bruijn sequence has been generated. In this case, return to Step 1 above with the new sequence playing the role of B_{mn} .
- (b) Step 2 of Algorithm M fails with $S = s_1 s_2 \dots s_k$, $k < L$ (i.e., no symbol s_{k+1} can be added such that the new n -sequence has not already appeared). In this case go to Step 3 below.

Step 3: (Backtrack) Choose the largest $j \leq k$ such that $s_j \neq 0$ and return to Step 2 above with $S = s_1 s_2 \dots s_{j-1}$. The algorithm terminates when $j \leq n$.

The sequences in TABLE 1 for $m = n = 2$ were all generated using this algorithm. We illustrate in TABLE 2 the generation of the second and third sequences.

Why does Algorithm B succeed in generating all de Bruijn sequences? The answer is that the backtracking idea effectively tests *all possible sequences* in decreasing lexicographic order and, therefore, must “catch” each de Bruijn sequence. Algorithm B is described in Fredricksen [9] who ascribes it to Alberts.

s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	j	
2	2	1	2	0	1	<u>1</u>	0	0	7	[B_{22} from Algorithm M]
2	2	1	2	0	1	0	<u>2</u>	x	8	[x represents failure]
2	2	1	2	0	<u>1</u>	0	0	x	6	
2	2	1	2	0	0	<u>2</u>	x		7	
2	2	1	2	0	0	1	<u>1</u>	0	8	[Success]
2	2	1	2	0	0	<u>1</u>	0	x	7	
2	2	1	<u>2</u>	0	0	x			4	[Immediate failure at j]
2	2	1	1	2	0	<u>2</u>	x		7	
2	2	1	1	2	0	<u>1</u>	0	0	7	[Success]

TABLE 2

If we are only interested in generating a single de Bruijn sequence for given m and n , then Algorithm M poses a serious problem for all but small values of m and n . Unfortunately, it requires $(m+1)^n$ units of memory since, at every application of Step 2, we must be able to determine if the newly-formed n -sequence has appeared previously in the sequence. Algorithm M also requires an average of $(m+1)/2$ lookups in this table of $(m+1)^n$ entries every time a new symbol is added to the string.

It is most desirable to use a **memoryless algorithm**, that is, one which does not require the availability of the entire string already generated. We consider two such algorithms which are similar but have some distinctive differences. (For another such algorithm and a general discussion, see [8] and [9].) We begin with two definitions (see [11]).

DEFINITION. A **necklace** S of length n is an n -sequence with the property that $S \geq T$ for every n -sequence T which is a cyclic permutation of S . Thus, if $S = s_1 s_2 \dots s_n$ is a necklace

$$S \geq s_i s_{i+1} \dots s_n s_1 \dots s_{i-1} \quad (23)$$

for $2 \leq i \leq n$.

An interesting question is: How many necklaces $Z(m, n)$ are there of length n containing the symbols $0, 1, 2, \dots, m$? The answer, whose derivation is beyond us here (see Golomb [12]), is

$$Z(m, n) = \frac{1}{n} \sum_{d|n} \phi(d) (m+1)^{n/d} \quad (24)$$

where ϕ is Euler's totient function (i.e., $\phi(d)$ is the number of positive integers less than d which are relatively prime to d with $\phi(1) = 1$) and the summation is over all positive d which divide n . For example, when $m = 3$, $n = 6$, then d takes on the values 1, 2, 3, 6 and $\phi(1) = 1$, $\phi(2) = 1$, $\phi(3) = 2$, $\phi(6) = 2$ so that (24) becomes

$$Z(3, 6) = \frac{1}{6} [4^6 + 4^3 + 2 \cdot 4^2 + 2 \cdot 4] = 700.$$

DEFINITION. Let $S = s_1 s_2 \dots s_n$ and let $T = s_1 s_2 \dots s_j$, $j < n$. Denote by T^k the subsequence of k consecutive repetitions of T . If $S = T^k$ when $k > 1$, we say that S is **periodic** with repetition (or periodicity) k and T is its **periodic reduction**. If there is no T with $k > 1$ for which $S = T^k$, we say that S is **aperiodic**.

Given positive integers m and n , the following recursive algorithm, due to Fredricksen and Maiorana [11], produces a de Bruijn sequence by generating successive necklaces of length n .

In the algorithm we denote by $B_{mn}^{(FM)}$ the de Bruijn sequence to be generated, and define $B_{0n}^{(FM)} = 0$.

Algorithm FM

Step 1: Start with the empty string.

Step 2: (Iterative Step) Generate the necklaces of length n whose first symbol is m , in decreasing lexicographic order. Append to the string already generated each necklace if it is aperiodic or its periodic reduction otherwise.

Step 3: (Recursive Step) Append $B_{m-1,n}^{(FM)}$ to the string already generated.

Note that it is not necessary to state this algorithm in recursive form. Step 2 could have said just, "Generate the necklaces of length n in decreasing lexicographic order" and then Step 3 would have been unnecessary.

Clearly the crucial part of Algorithm FM is the generation of successive necklaces. Suppose $S = s_1 s_2 \dots s_n$ is a necklace and we wish to generate its successor necklace (i.e., the next smallest necklace lexicographically). Algorithm N below generates the successor of S ; in the formulation of Step 2, S_k denotes the string $s_1 s_2 \dots s_k$.

Algorithm N

Step 1: Find j such that $s_j \neq 0$ but $s_{j+1} = s_{j+2} = \dots = s_n = 0$.

Step 2: Generate $T = [S_{j-1}(s_j - 1)]^q S_{n-qj}$ where $0 \leq n - qj < j$. (That is, T consists of q repetitions of the first $j - 1$ symbols of S and the j th symbol reduced by 1 followed by as many of the first $j - 1$ symbols as needed to obtain n symbols.)

Step 3: Test whether T is a necklace [i.e., that it satisfies (23)]. If so, stop. If not, return to Step 1 with S replaced by T .

Fredricksen and Maiorana prove that this algorithm will produce the next necklace in at most $(1/2)(n + 1)$ steps. For example with $n = 12$ and $m = 2$

$$S = 20002 \ 00000 \ 00$$

is a necklace. From this Algorithm N generates successively

$$20001 \quad 20001 \quad 20 \quad [j = 5, \quad q = 2]$$

$$20001 \quad 20001 \quad 12 \quad [j = 11, \quad q = 1]$$

$$20001 \quad 20001 \quad 11 \quad [j = 12, \quad q = 1]$$

with the last string the necklace next smaller than S . Note that any time the last symbol is m , the result cannot be a necklace so that this may immediately be reduced to $m - 1$.

Fredricksen and Maiorana also prove that Algorithm FM indeed generates a de Bruijn sequence. From a computational point of view, Algorithm FM has the modest drawback that every necklace must be generated and, thus, Algorithm N must be executed many times. (Each necklace must also be tested for periodicity but this is trivial since we must only see if j in Algorithm N divides n .) An algorithm which ameliorates this problem and which we shall call Algorithm R has been developed by Ralston [24]. Although in some ways similar to Fredricksen and Maiorana's algorithm, Algorithm R has the following interesting features:

1. It is truly recursive in the sense that $B_{m,n}^{(R)}$, the de Bruijn sequence generated by Algorithm R for given m and n , consists of an initial segment followed by $B_{m-1,n}^{(R)}$.
2. The initial segment requires the generation only of those necklaces containing just m 's and $(m - 1)$'s. To each of these necklaces which is aperiodic is appended a sequence of n -sequences, each of which
 - (a) contains m 's only where the necklace contains m 's and
 - (b) is such that it has the greatest possible value less than the preceding n -sequence.

For example, when $m = 3$, $n = 5$, 33232 is a necklace and it is followed by

$$33231 \ 33230 \ 33132 \ 33131 \ 33130 \ 33032 \ 33031 \ 33030.$$

3. For periodic necklaces of m 's and $(m - 1)$'s let

t be the number of $(m - 1)$'s in the periodic reduction,

$$u = m^t - 1,$$

k be the length of the periodic reduction.

Then the periodic reduction of the necklace is followed by a sequence of k -sequences whose values are determined using $B_{u,n/k}^{(R)}$.

The details of how this is done are too complex to present here. But it is particularly worth

noting that, because of the use of $B_{u,n/k}^{(R)}$ in the generation of $B_{mn}^{(R)}$, Algorithm R contains that rather rare situation of a recursion within an already recursive algorithm.

The following instances of Algorithms FM and R for $n = 4, m = 2$ are taken, respectively, from Fredricksen and Maiorana [11] and Ralston [24].

$B_{24}^{(FM)} = 2$

2221 2220 2211 2210 2201 2200
21
2120 2111 2110 2101 2100
20
2011 2010 2001 2000
1
1110 1100
10
1000
0

$B_{24}^{(R)} = 2$

2221 2220
2211 2210 2201 2200
2121 2020
2111 2110 2101 2100 2011 2010 2001 2000
1
1110
1100
10
1000
0

For $n = 2$ and $n = 3$ the two algorithms give the same result. For $n \geq 4$, however, the results are always different. One noteworthy feature of Algorithm FM is that it generates exactly the same sequence as Algorithm M (after moving the n initial zeros to the end).

A few words are in order about the subject of the correctness of the algorithms we have discussed in this and the previous section. For those in Section 3 our discussion amounted to something close to “classical” (although very informal) mathematical proofs that they perform as claimed. For Algorithm B of this section we did no more than sketch a proof. And for Algorithms FM and R we referred only to the papers in which they were first presented and where standard mathematical proofs of their correctness are given. Of interest here is that a very active area of research in computer science concerns formal proofs of algorithms—and their implementations as computer programs—using the tools of mathematical logic. Although such proofs of even very simple algorithms tend to be quite difficult and involved, this is a most important and interesting area of research for both mathematicians and computer scientists. For a recent review of the status of research in this area see London [17].

5. Curriculum implications

The de Bruijn sequence problem is exceptional, yielding to successful solutions from several branches of discrete mathematics. Although few problems can be solved in such a variety of ways, the use of each of these branches—combinatorial analysis, graph theory, linear algebra, abstract algebra—is quite typical of the way these areas of mathematics impinge on computer science. The growing importance of discrete mathematics relative to more classical areas of mathematics and the symbiosis of discrete mathematics and computer science has clear and important educational implications for computer science and, although more speculative, there are also implications for the mathematics community.

It is not always fully recognized in the computer science community that mathematics should be an important component of any computer science curriculum. While there is room for

argument about just how much of the calculus-linear algebra sequence should be taught to computer science majors, to this author it is clear that discrete mathematics should play *at least an equal role* to the classical subject matter in the first two years. In addition, a one- or two-year sequence balanced between discrete and continuous mathematics would be of more professional value to students in the social, management, and behavioral sciences than the calculus sequence many such students are now required to take.

In my opinion, computer science will provide the largest source of problems for mathematicians for years to come. Thus, it is reasonable to predict that mathematical research will become increasingly oriented toward discrete mathematics. This suggests that an undergraduate major in mathematics better balanced between discrete and continuous mathematics should be developed. Undoubtedly for most readers, the points above raise more questions than they answer. Some of the author's answers to these questions can be found in Ralston [21, 22, 23, 25].

(Editor's note. The cover of this *Magazine* shows five de Bruijn sequences arranged on concentric circles. Beginning with the innermost circle and moving outwards, the sequences are for the following pairs (m, n) : (1, 2), (1, 3), (2, 2), (1, 4), (3, 2).)

References

- [1] J. Alanen and D. E. Knuth, Tables of finite fields, *Sankhyā* (Calcutta), 26 (1964) 305–328 (see also *Math. Rev.* 32, 4122).
- [2] A. A. Albert, *Fundamental Concepts of Higher Algebra*, Univ. of Chicago Press, 1956, pp. 128–131.
- [3] E. R. Berlekamp, *Algebraic Coding Theory*, McGraw-Hill, New York, 1968, chapt. 4.
- [4] N. G. de Bruijn, A combinatorial problem, *Nederl. Akad. Wetensch. Proc.*, 49 (1946) 758–764.
- [5] ———, Acknowledgment of priority to C. Flye Sainte-Marie on the counting of circular arrangements of 2^n zeros and ones that show each n -letter word exactly once, Report 75-WSK-06, Technical University Eindhoven, 1975.
- [6] C. Eldert, H. J. Gray, Jr., H. M. Gurk, and M. Rubinoff, Shifting counters, *AIIE Trans.*, 77 (1958) 70–74.
- [7] C. Flye Sainte-Marie, Solution to problem number 58, *l'Intermédiaire des Mathématiciens*, 1 (1894) 107–110.
- [8] H. Fredricksen, A class of nonlinear de Bruijn cycles, *J. Combin. Theory*, 19 (1975) 191–199.
- [9] ———, A survey of full cycle algorithms, *SIAM Review*, (to appear).
- [10] H. Fredricksen and I. J. Kessler, Lexicographic compositions and de Bruijn sequences, *J. Combin. Theory*, 22 (1977) 17–30.
- [11] H. Fredricksen and J. Maiorana, Necklaces of beads in k colors and k -ary de Bruijn sequences, *Discrete Math.*, 23 (1978) 207–210.
- [12] S. W. Golomb, *Shift Register Sequences*, Holden-Day, San Francisco, 1967, pp. 118–122.
- [13] I. J. Good, Normal recurring decimals, *J. London Math. Soc.*, 21 (1946) 167–169.
- [14] M. Hall, Jr., *Combinatorial Theory*, Blaisdell, Waltham, MA, 1967.
- [15] D. E. Knuth, *The Art of Computer Programming*, Addison-Wesley, Reading, MA, 1969, vol. 2, p. 27.
- [16] J. H. van Lint, *Combinatorial Theory Seminar*, vol. 382, *Lecture Notes in Mathematics*, Springer-Verlag, Berlin, 1974.
- [17] R. L. London, *Program Verification in Research Directions in Software Technology*, MIT Press, Cambridge, MA, 1979, pp. 302–315.
- [18] W. Mantel, Resten van wederkeerige reeksen, *Nieuw Arch. Wisk.*, Ser. 2, (1897) 172–184.
- [19] M. H. Martin, A problem in arrangements, *Bull. Amer. Math. Soc.*, 40 (1934) 859–864.
- [20] F. P. Preparata and R. T. Yeh, *Introduction to Discrete Structures*, Addison-Wesley, Reading, MA, 1973, p. 74.
- [21] A. Ralston, The twilight of the calculus, *Southeast Asian Bull. Math.*, 3 (1979) 49–56.
- [22] A. Ralston and M. Shaw, Curriculum 78—is computer science really that unmathematical?, *Comm. ACM*, 23 (1980) 67–70.
- [23] A. Ralston, Computer science, mathematics and the undergraduate curricula in both, Report No. 161, Dept. of Computer Science, SUNY at Buffalo, 1980.
- [24] ———, A new memoryless algorithm for de Bruijn sequences, *J. Algorithms*, 2 (1981) 50–62.
- [25] ———, Computer science, mathematics and the undergraduate curricula in both, *Amer. Math. Monthly*, 88 (1981) 472–484 (a shortened version of [23]).
- [26] D. Rees, Note on a paper by I. J. Good, *J. London Math. Soc.*, 21 (1946) 169–172.
- [27] E. Roth, Permutations arranged around a circle, *Amer. Math. Monthly*, 78 (1971) 990–992.

Why Study Equations over Finite Fields?

Finite field solutions to equations are related in a subtle and intriguing way to rational solutions and complex solutions.

NEAL KOBLITZ

University of Washington
Seattle, WA 98195

At least as far back as the Greeks, mathematicians have been interested in finding integer solutions to equations. Perhaps the best known example of such a “Diophantine equation” is

$$X^N + Y^N = Z^N, \quad (1)$$

where N is an integer greater than 2, and one looks for nonzero integers X , Y and Z for which the equation holds. The famous conjecture of Fermat is that there are none.

Fermat’s conjecture can be stated equivalently as follows. Let $x = X/Z$ and $y = Y/Z$. If we divide the equation (1) through by Z^N , we see that finding nonzero integer solutions (X, Y, Z) to equation (1) is equivalent to finding nonzero rational number solutions (x, y) to the equation

$$x^N + y^N = 1. \quad (1')$$

The equation (1') determines a curve in the xy -plane, and we’re interested in points (x, y) with (nonzero) rational coordinates which lie on the curve. Of course, we expect that there are none.

Such points with rational coordinates are called **Q-valued solutions** of (1'), or **Q-points** on the curve determined by (1'), where **Q** denotes the field of rational numbers. I’d like to discuss why one might want to take a finite field instead of **Q** and consider solutions to an equation such as (1'), where x and y are numbers in this other field.

Let me start by recalling the basic facts about finite fields. Let p be a prime number. For every power p^r of p , there is a field denoted F_{p^r} having exactly p^r elements; and there is essentially only one such field. The simplest case is when $r = 1$. Then F_p is the set of residues modulo p , i.e., the integers $0, 1, 2, \dots, p-1$ with the usual operations of addition and multiplication modulo p . (For example, $2 \times 5 = 3$ in F_7 .) One fundamental fact about a finite field F_{p^r} is that the set $F_{p^r}^*$ of nonzero elements has a generator g (generally the choice of g is not unique) such that all of the $p^r - 1$ numbers in $F_{p^r}^*$ are powers of g . For example, every number in F_7^* can be written as a power of 3; that is, $3, 3^2 = 2, 3^3 = 6, 3^4 = 4, 3^5 = 5, 3^6 = 1$ exhaust all of the nonzero elements $1, 2, \dots, 6$.

Returning to our equation $x^N + y^N = 1$, suppose we ask about solutions x, y which are numbers in F_{p^r} . For a fixed p^r there can be only finitely many such pairs (x, y) , which we could find in principle simply by substituting all p^r possibilities for x and all p^r possibilities for y , and seeing which satisfy the equation. So, while it is hard to know how many **Q**-solutions our equation has, it is easy to find how many F_{p^r} -solutions it has: Just count ‘em!

Thus, the *first reason* for studying solutions to equations over finite fields rather than over **Q** is: *It’s easier!* This illustrates one basic principle of mathematical research: If you can’t solve the problem you want to solve, replace it by an easier problem.

Before tackling the Fermat equation, let’s count points on some simpler curves.

EXAMPLE 1. Any equation of the form $y = f(x)$ has p^r solutions (x, y) in F_{p^r} , because x can be given an arbitrary value in F_{p^r} , and then y is uniquely determined. Thus, if $N_{r,p}$ denotes the number of F_{p^r} -solutions to the equation $y = f(x)$, we have:

$$N_{r,p} = p^r. \quad (2)$$

EXAMPLE 2. The “unit circle” equation $x^2 + y^2 = 1$ can be shown to have either $p^r - 1$ or $p^r + 1$ solutions in F_{p^r} , depending on whether or not -1 is a square in F_{p^r} . Using the properties of finite fields, it is not a hard exercise to show that

$$N_{r,p} = \begin{cases} p^r - 1 & \text{if } p \text{ is of the form } 4k + 1, \\ p^r - (-1)^r & \text{if } p \text{ is of the form } 4k + 3, \end{cases} \quad (3)$$

where this time $N_{r,p}$ denotes the number of F_{p^r} -solutions to the equation $x^2 + y^2 = 1$.

Now let $N_{r,p}$ denote the number of F_{p^r} -solutions to the Fermat equation (1'). (Note: To be precise, we want to include the “trivial” solutions—those for which x or y is zero—and also the “points at infinity”. The points at infinity come from the solutions of $X^N + Y^N = Z^N$ for which $Z = 0$, solutions which we ignored when we wrote $x = X/Z$, $y = Y/Z$.)

From now on, I will assume for simplicity that N divides $p - 1$. Then there is a simple formula for $N_{r,p}$. Here and elsewhere, j and k will range over all integers from 1 to $N - 1$ whose sum is not N . (There are $(N - 1)(N - 2)$ such pairs of indices j, k .) Then

$$N_{r,p} = p^r + 1 - \sum_{j,k} J\left(\frac{j}{N}, \frac{k}{N}\right)^r. \quad (4)$$

I have to explain what $J(j/N, k/N)$ is. Choose a generator g of the group $F_p^* = \{1, 2, \dots, p - 1\}$; for example, we took 3 for F_7^* above. Define $\chi_{j/N}$ to be the map from F_p to the complex numbers which takes 0 to 0 and takes the m th power of g to the m th power of the complex N th root of unity $e^{2\pi i j/N}$; that is,

$$\chi_{j/N}(g^m) = e^{2\pi i m j/N}.$$

In the case of F_7 , for example, FIGURE 1 shows where $\chi_{1/6}$ takes the elements of $F_7 = \{0, 1, 2, 3, 4, 5, 6\}$. It is easy to see that $\chi_{j/N}$ is multiplicative, i.e., $\chi_{j/N}(xy) = \chi_{j/N}(x) \cdot \chi_{j/N}(y)$ for any x, y in F_p . The map $\chi_{j/N}$ is called a “multiplicative character”. Finally, $J(j/N, k/N)$ is defined to be

$$J\left(\frac{j}{N}, \frac{k}{N}\right) = - \sum_{x \in F_p} \chi_{j/N}(x) \chi_{k/N}(1 - x). \quad (5)$$

$J(j/N, k/N)$ is called a “Jacobi sum”. You might wish to compute $J(1/6, 2/6)$ for F_7 ; the answer is $-2 - i\sqrt{3}$.

In general, given an equation $F(x, y) = 0$ and given a prime p , we have a whole sequence of numbers $N_{r,p}$, $r = 1, 2, 3, \dots$, which tell us how many F_{p^r} -solutions our equation has. It turns out

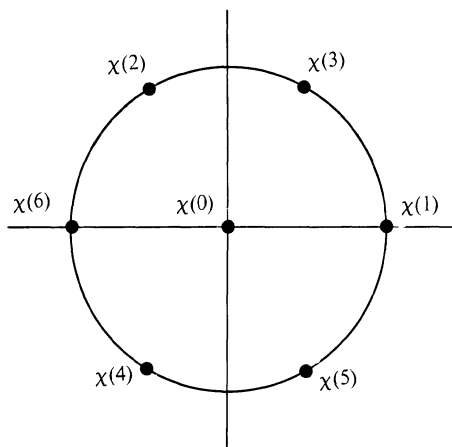


FIGURE 1

to be very useful to incorporate this information—this sequence of integers $N_{r,p}$ —into a generating series

$$Z_p(T) \stackrel{\text{def}}{=} e^{\sum N_{r,p} T^r / r}, \quad (6)$$

where the summation is over $r = 1, 2, 3, \dots$. This is the so-called “congruence zeta-function” introduced by E. Artin in the 1920’s.

The form chosen for $Z_p(T)$ may appear strange at first. In principle, we have a wide choice of power series that could be used to express the data $N_{r,p}$. However, the series (6) has two useful properties. First of all, if $N_{r,p}$ happens to be of the form α^r (as in Example 1 above, where $\alpha = p$), then $Z_p(T)$ takes a particularly simple form:

$$Z_p(T) = e^{\sum \alpha^r T^r / r} = e^{-\log(1-\alpha T)} = \frac{1}{1-\alpha T}. \quad (7)$$

In the second place, suppose that $N_{r,p}$ is a sum of two different functions of r :

$$N_{r,p} = N_{r,p}^* + N_{r,p}^{**}. \quad (8)$$

For example, if our equation $F(x, y) = 0$ is the product of two equations $F(x, y) = G(x, y)H(x, y)$ with no common solutions (i.e., $G(x, y) = H(x, y) = 0$ is impossible), then $N_{r,p}$ for $F(x, y) = 0$ is the sum of the $N_{r,p}$ for $G(x, y) = 0$ (call it $N_{r,p}^*$) and the $N_{r,p}$ for $H(x, y) = 0$ (call it $N_{r,p}^{**}$). Geometrically, our original curve is the disjoint union of the two curves given by G and H . In such a case the series $Z_p(T)$ for our equation is the *product* of the $Z_p(T)$ for the two parts:

$$Z_p(T) = e^{\sum N_{r,p} T^r / r} = e^{\sum N_{r,p}^* T^r / r} \cdot e^{\sum N_{r,p}^{**} T^r / r}. \quad (9)$$

Because of these two properties, it follows that whenever $N_{r,p}$ can be written in the form

$$N_{r,p} = \alpha_1^r + \alpha_2^r + \dots + \alpha_m^r - \beta_1^r - \beta_2^r - \dots - \beta_n^r, \quad (10)$$

where the α ’s and β ’s depend on p and on our equation but do *not* depend on r , then $Z_p(T)$ is a rational function. More precisely, if we substitute (10) in (6) and proceed as in (7) and (9), we obtain

$$Z_p(T) = \frac{(1 - \beta_1 T)(1 - \beta_2 T) \cdots (1 - \beta_n T)}{(1 - \alpha_1 T)(1 - \alpha_2 T) \cdots (1 - \alpha_m T)}.$$

Thus, $Z_p(T)$ is a rational function whose denominator has $1/\alpha_i$ for its roots and whose numerator has roots $1/\beta_j$.

Notice that in all of our examples we have a formula of the form (10). In (2), p is α_1 and there are no β ’s. In (3), p is α_1 , and either 1 or -1 is β_1 . In the case of the Fermat equation (1’), where $N_{r,p}$ is given by (4), we have $\alpha_1 = p$, $\alpha_2 = 1$, and the β ’s are the $J(j/N, k/N)$. Thus, for the Fermat equation we have

$$Z_p(T) = \frac{1}{(1-T)(1-pT)} \prod_{j,k} \left(1 - J\left(\frac{j}{N}, \frac{k}{N}\right) T \right),$$

which is a rational function whose numerator is a polynomial of degree $(N-1)(N-2)$.

In 1949, A. Weil noticed certain interesting properties of the function $Z_p(T)$ in the Fermat case and some other examples, which led him to a far-reaching conjecture concerning the congruence zeta-function. For example, he conjectured that $Z_p(T)$ is always a rational function. This is by no means obvious from the definition, but, as we saw, it is true whenever we have a formula of the form (10).

Moreover, in the case of curves (an equation with two variables, such as (1’)), the degree of the numerator is twice the so-called “Betti number” g of the corresponding complex Riemann surface. For our equation (1’) what this means is the following. Suppose we consider the set of *complex* points (x, y) which satisfy the equation and put in the points at infinity. What we obtain is a surface having $g = (N-1)(N-2)/2$ “handles”. (FIGURE 2 shows the case when $N = 4$, $g = 3$.)

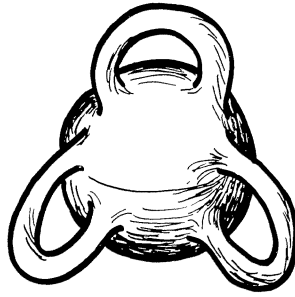


FIGURE 2

Then the numerator of the zeta-function $Z_p(T)$ —which, remember, came from counting the *finite-field* points (x, y) satisfying (1')—has degree precisely $2g$.

Thus there is an intriguing relationship between the number-theoretic properties of the equation considered modulo p and its “physical” properties (such as number of handles) when it is considered complex-analytically. Weil’s conjectures were the object of intensive research for a quarter of a century, culminating in P. Deligne’s proof in 1973 of the last and hardest part (the “Riemann hypothesis” for the congruence zeta-function).

Before leaving the example of the Fermat equation, I’d like to mention one further parallel between its finite-field theoretic and its complex analytic properties. In the classical theory of algebraic curves over the complex numbers, one associates to them certain definite integrals, called “periods”. In the case of the Fermat equation, these integrals are of the form

$$\int_0^1 x^{j/N} (1-x)^{k/N} \frac{dx}{x(1-x)}.$$

(Readers familiar with the beta-function will recognize this as $B(j/N, k/N)$.)

This expression should be compared to the definition of the Jacobi sum (5); it is closely analogous. Instead of summing over x in a finite field, we integrate over real numbers x . Instead of the multiplicative function $x \mapsto \chi_{j/N}(x)$ occurring in a summation (with x in a finite field), we have the multiplicative function $x \mapsto x^{j/N}$ occurring in an integral (with x a real number). So the *second reason* for studying equations over finite fields is that deep *analogies exist* between finite-field theoretic properties and complex-analytic properties of equations.

The *third and final reason* I offer for studying equations over finite fields is that sometimes such information can be pieced together to tell us something about the set of solutions in the field \mathbb{Q} of rational numbers. That is, *we learn something about Diophantine equations*.

The best example is the conjecture on elliptic curves formulated by Birch and Swinnerton-Dyer in the early 1960’s. To show the flavor of the Birch-Swinnerton-Dyer conjecture, let us suppose that we have an equation of the form

$$y^2 = f(x),$$

where $f(x)$ is a cubic polynomial with integer coefficients and distinct roots. This is called an “elliptic curve”. FIGURE 3 shows a typical example, $y^2 = x^3 - x$, graphed in the real plane.

Suppose that instead of pairs of real numbers (x, y) as in FIGURE 3, we take pairs of complex numbers (x, y) satisfying $y^2 = x^3 - x$. Then this set of pairs (x, y) can be identified in a natural way with the surface of a torus (“donut”; see FIGURE 4). This is true of any equation $y^2 = f(x)$ of this sort: the complex Riemann surface corresponding to this equation has one “handle”. That is, its Betti number g is equal to 1.

If we now consider this equation $y^2 = f(x)$ modulo a prime p and count all the solutions $x, y \in F_{p^r}$ for each r , then we can form the zeta-function

$$Z_p(T) = e^{\sum_{r=1}^{\infty} N_{r,p} T^r / r}, \quad (11)$$

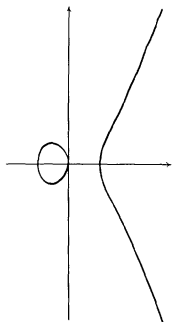


FIGURE 3. The real points of the elliptic curve $y^2 = x^3 - x$.

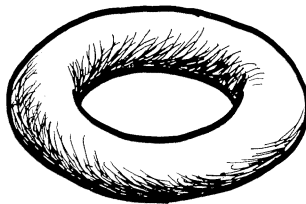


FIGURE 4

just as we did for the Fermat equation. Since $g = 1$, this rational function $Z_p(T)$ has as its numerator (which I denote $\text{num}_p(T)$) a polynomial of degree 2. More precisely, the Weil conjectures tell us that $Z_p(T)$ is of the form

$$Z_p(T) = \frac{1 - a_p T + pT^2}{(1 - T)(1 - pT)} = \frac{\text{num}_p(T)}{(1 - T)(1 - pT)}. \quad (12)$$

(Actually, for elliptic curves the Weil conjectures were already proved before the conjectures were formulated in full generality in 1949.) Note that the only part of this expression that varies from one elliptic curve $y^2 = f(x)$ to another is the linear coefficient a_p in $\text{num}_p(T)$.

If we take log of the two expressions (11) and (12) and equate coefficients of T in the two power series expansions, we immediately obtain the formula

$$a_p = p + 1 - N_{1,p}. \quad (13)$$

Thus, if we only know the number of F_p -solutions to the equation, we can substitute that number for $N_{1,p}$ in equation (13) to find a_p , and hence obtain the exact expression for $Z_p(T)$. Since $Z_p(T)$ determines all of the $N_{r,p}$ (just take its logarithm, expand in a power series, and compare with (11)), this means that once we know $N_{1,p}$, all of the $N_{r,p}$ are determined and can be found by a simple formula. In other words, once we find $N_{1,p}$ —which just means letting x and y take the values $0, 1, 2, \dots, p-1$ and counting the number of times $y^2 = f(x)$ modulo p —we can immediately determine the number of F_p -solutions for any r . This is what the Weil conjectures enable us to do in the case of an elliptic curve.

Once we know $Z_p(T)$ for each prime p , we can combine all of this information into a single function, which thereby incorporates all of the data $N_{r,p}$ for all r and all p . This function, called the “Hasse-Weil zeta-function”, is defined as follows for s a complex number:

$$Z(s) = \prod_p \frac{1}{\text{num}_p(p^{-s})} = \prod_p \frac{1}{1 - a_p p^{-s} + p^{1-2s}}. \quad (14)$$

(Actually, this definition is slightly off, but in a way which is irrelevant for our purposes.)

As an aside, let me mention that the Hasse-Weil zeta-function can be thought of as a generalization of the Riemann zeta-function. If we took the equation $y^2 = f(x) = x^3 + x^2$ (which is not really an elliptic curve, since this $f(x)$ has a double root), we would find that for this curve $\text{num}_p(T) = 1 - T$ (in other words, for this curve $Z_p(T)$ is simply $1/(1 - pT)$). Thus, for this curve our definition would give us $Z(s) = \prod_p 1/(1 - p^{-s})$, which is the “Euler product form” of the Riemann zeta-function.

What type of a function is the Hasse-Weil zeta-function $Z(s)$? It is a function of a complex variable s that is associated to our elliptic curve $y^2 = f(x)$. Using well-known estimates for a_p , it is easy to show that the infinite product converges when $\text{Re}(s) > 3/2$. Moreover, for a broad class of elliptic curves, $Z(s)$ is known to extend by analytic continuation to a meromorphic function on the entire complex s -plane; and this is conjectured to be the case for all elliptic curves.

Assuming that this extendibility conjecture is true, Birch and Swinnerton-Dyer make the following striking

CONJECTURE. $Z(1) = 0$ if and only if the equation $y^2 = f(x)$ has infinitely many rational solutions $x, y \in \mathbf{Q}$.

There is a heuristic argument—far from a proof—which shows why this conjecture might be true. Let us pretend that the infinite product for $Z(s)$ converges when $s = 1$. In that case, (14) and (13) imply that

$$Z(1) = \prod_p \frac{1}{1 - a_p p^{-1} + p^{-1}} = \prod_p \frac{p}{p - a_p + 1} = \prod_p \frac{p}{N_{1,p}}.$$

Now if there were infinitely many $x, y \in \mathbf{Q}$ which satisfied the equation $y^2 = f(x)$, then for each prime p , by reducing x and y modulo p (when neither x nor y has denominator divisible by p), we would expect to find a lot of distinct mod p solutions. In other words, the more \mathbf{Q} -solutions, the more F_p -solutions. This means that $N_{1,p}$ would be large, in fact, probably enough larger than p to cause the infinite product of $p/N_{1,p}$ to approach zero. This is a rough heuristic argument for the conjecture in the one direction: infinitely many \mathbf{Q} -solutions implies $Z(1) = 0$.

I should point out that the conjecture stated above is only a small part of the Birch-Swinnerton-Dyer conjecture, which relates the behavior of the function $Z(s)$ at $s = 1$ to many subtle arithmetic properties of the elliptic curve. The conjecture is very far from being proved, but there is now a fair amount of computational and theoretical evidence to support it. The most dramatic partial result so far has been the proof [4] by John Coates and Andrew Wiles about four years ago that for a large class of elliptic curves, an infinite number of \mathbf{Q} -solutions implies $Z(1) = 0$. (This is the implication we concluded from the heuristic argument above, but their proof bears no resemblance to those heuristics.)

As you see, the connections between information about finite field solutions to equations and information about \mathbf{Q} -solutions are very indirect, deep, and difficult. Detailed knowledge of F_p -solutions to an equation for all p and r does not in any simple way lead to answers to questions about its \mathbf{Q} -solutions. For example, even though we gave an exact formula for $N_{r,p}$ for the Fermat equation—and the properties of the Jacobi sums $J(j/N, k/N)$ in the formula are very well known—nevertheless, no one has any idea how to use all of this knowledge to prove Fermat's last theorem. The connections between F_p -solutions and \mathbf{Q} -solutions are subtle and not yet well understood. Intensive efforts by number theorists have been, and for a long time will continue to be, directed toward understanding such connections.

This paper is a revised version of the text of an M.A.A. invited address given in Portland, Oregon, on June 19, 1981.

This work is supported in part by N.S.F. grant #MCS80-02271.

References

Note: Of the following references, [7] and [9] are the most elementary, and parts of [2] and [3] are also easy to read; the others presuppose more technical background.

- [1] E. Artin, *Collected Papers*, S. Lang and J. Tate, eds., Addison-Wesley, Reading, MA, 1965.
- [2] B. Birch and P. Swinnerton-Dyer, Notes on elliptic curves II, *J. Reine Angew. Math.*, 218 (1965) 79–108.
- [3] J. W. S. Cassels, Arithmetic on an elliptic curve, *Proc. I. C. M. Stockholm*, (1962) 234–246.
- [4] J. Coates and A. Wiles, On the conjecture of Birch and Swinnerton-Dyer, *Invent. Math.*, 39 (1977) 223–251.
- [5] P. Deligne, La conjecture de Weil I, *Publ. Math. I. H. E. S.*, 43 (1974) 273–307.
- [6] H. Hasse, Bericht über neuere Untersuchungen und Probleme aus der Theorie der algebraischen Zahlkörper, Teil II, 2nd ed., Würzburg-Wien, Physica, 1965.
- [7] K. Ireland and M. Rosen, *Elements of Number Theory*, Including an Introduction to Equations over Finite Fields, Bogden and Quigley, Tarrytown, NY, 1972.
- [8] N. Katz, An overview of Deligne's proof of the Riemann hypothesis for varieties over finite fields, *Proc. Symp. in Pure Math.*, 28 (1976) 275–305.
- [9] A. Weil, Number of solutions of equations in finite fields, *Bull. Amer. Math. Soc.*, 55 (1949) 497–508.

Historical Roots of Confusion Among Beginning Algebra Students: A Newly Discovered Manuscript

HELENA M. PYCIOR

Department of History

University of Wisconsin-Milwaukee

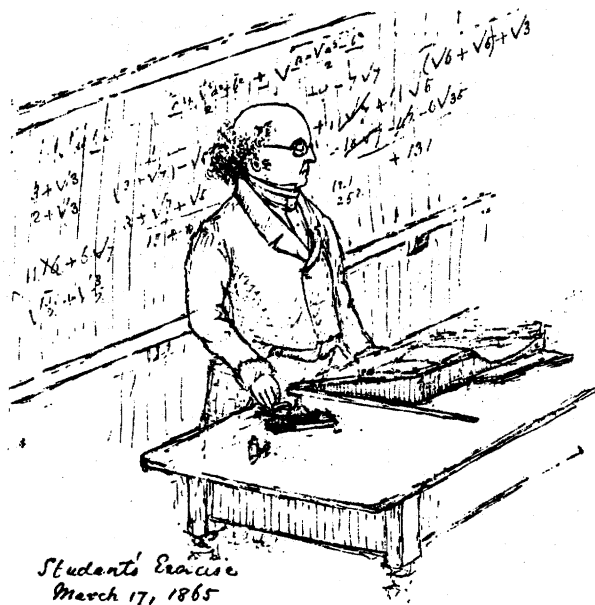
Milwaukee, WI 53201

Beginning students are usually told that modern algebra is the study of undefined entities which obey certain rules. The rules, the introductory textbooks explain, are rather arbitrarily chosen by mathematicians who keep in mind such desirable properties as consistency or freedom from contradictions. Most students require some time to get accustomed to algebraic arbitrariness and the related freedom of the algebraist.

The history of algebra can facilitate students' adjustment to the spirit and direction of not only abstract algebra but contemporary mathematics in general. The perplexed student might find comfort, for example, in the realization that arbitrariness and freedom are youthful additions to mathematics, born of the early nineteenth-century research on abstract algebra and non-Euclidean geometries, and widely accepted only later in the century. Around 1800 there was but one algebra. In this algebra, frequently called universal arithmetic, letters stood for numbers or quantities, and the laws of arithmetic, such as commutativity of addition and multiplication, prevailed. By the middle of the nineteenth century, however, mathematicians had created many different algebras, including the quaternions, whose multiplication is noncommutative. From this perspective, the confused modern algebra student may be a victim of the circumstance of having been born in the twentieth century. Additional solace might be derived from the revelation that bewilderment and occasionally even rejection greeted the original formulation of the symbolical approach to algebra. Thus the perplexed student has historical roots! A study of early objections to symbolical algebra can also shed light on and stimulate discussion of the oft-unvoiced concerns of present-day modern algebra students.

This paper offers a newly discovered nineteenth-century manuscript which documents confusion as an almost immediate response to the emergence of symbolical algebra in its limited form. (Like modern algebra, early nineteenth-century symbolical algebra admitted undefined entities; unlike modern algebra, symbolical algebra, in its original form, was governed by the laws of arithmetic.) The manuscript is a spoof-play on Augustus De Morgan's *Elements of Algebra* [5], one of the first undergraduate algebra textbooks to incorporate the symbolical (or limited modern) approach to algebra.

I found the play in November 1979 during a search through MS. Add. 163, labelled "Correspondence of Augustus De Morgan," in the D. M. S. Watson Library, University College London. In my prior research I had encountered but one example of early opposition to the symbolical approach—that of Sir William Rowan Hamilton, the famous Irish mathematician who in the 1830s rejected symbolical algebra in favor of his definition of algebra as the science of pure time (see [11]–[13]). Before me, then, was a new example of such early discontent and one possibly unknown to all other historians. It was proof of the existence of at least a second early critic of symbolical algebra.



Sketch of Augustus De Morgan, MS. Add. 7, D. M. S. Watson Library, University College London; artist unknown. Reproduced with permission of The Library, University College London.

Who was (were) the witty critic(s) of De Morgan's approach to algebra? Unfortunately, the manuscript is unsigned. When faced by such a problem, the historian turns sleuth. Internal analysis of the manuscript supplemented by general knowledge of nineteenth-century British mathematics finally led to the tentative conclusion that it was composed by either William Frend, an English mathematical renegade, or by Sophia Elizabeth Frend, William's eldest child and, from 1837 on, De Morgan's wife. At the outset there was circumstantial evidence to support this conclusion: the folder in which the manuscript was found contained some letters written by the Frends. Additionally, the manuscript's criticism of the negative numbers associated it with the tradition of opposition to these numbers and thereby, quite probably, with William Frend, who was during the early nineteenth century the leading British opponent of the negative numbers (see [6]). There soon came to light additional support for ascribing the play to either or both of the Frends: The hand in which it survives is that of Sophia Frend.

Given the historical sources to which I have had access, however, it appears impossible to decide between father and daughter as likely author. Although the manuscript's handwriting matches Sophia Frend's, we cannot, for example, rule out composition by her father, since the daughter occasionally served as her father's secretary [14, p. 301].

The exact date of the manuscript has also proven elusive. Although undated, it was clearly written after 1835, the year of publication of the first edition of De Morgan's *Elements of Algebra*. The play itself contains clues to the date of its composition. Scene Two, for instance, is set in University College London and introduces as a minor character a Mr. Kennel. J. Percival, the archivist of the Watson Library, has pointed out that in real life Kennel was the accountant for the college through 1838, when his embezzlement of £1500 was discovered. It seems unlikely—although it is possible—that a known embezzler would have been included in the play. If we accept the tentative conclusion that one or both of the Frends composed the manuscript, there is additional evidence in favor of its completion no later than the discovery of Kennel's crime. In 1838 William Frend suffered a stroke which left him "hardly able to speak or to move" [14, p. 306], and by that year Sophia Frend was De Morgan's wife.

History involves interpretation as well as occasional detective work. As can be seen from the secondary literature in this paper's bibliography, there already exist quite a few different explanations of the development of symbolical algebra. As essential background to the spoof-play,

I will sketch one of these interpretations: that symbolical algebra developed at least partially in response to the problem of the negative numbers (see [17] and [22]).

What is a negative number? Textbook definitions differ, but a naive reply could be: a number $-a$ such that when it is added to $+a$, the result is zero. But what really is a negative number? This question made sense to and became quite a burning issue among British mathematicians of the late eighteenth and early nineteenth centuries. Through the early nineteenth century mathematics was thought to be meaningful in a more than strictly logical way. Mathematical symbols, it was believed, stood for ideas which either were present *a priori* in the human mind or were derived from physical experience. Thus a negative number was supposed to mean something and therefore be susceptible to being defined. According to the period's most popular definitions, mathematics was the science of quantity; a negative number was (1) a quantity less than nothing or (2) a quantity obtained from the subtraction of a greater from a lesser.

Although Isaac Newton's powerful mind was comfortable with definition (1) of a negative number [18, p. 3], some later English mathematicians, including Francis Maseres and William Frend, charged that it was nonsense. Who, after all, had ever observed a quantity less than nothing or who could even form a clear idea of one? Maseres and Frend also dismissed definition (2), claiming that it was impossible to take more from less. Furthermore, (they correctly noted) the second definition depended on the operation of subtraction which had been defined exclusively for those cases in which the minuend was greater than or equal to the subtrahend. When no mathematician could answer their criticisms, Maseres and Frend called for total abandonment of the negatives and the restriction of algebra to universal arithmetic in the strictest sense—to the study of symbols standing only for nonnegative numbers and a subtraction operation covering only the cases in which the minuend was greater than or equal to the subtrahend (see [10] and [16]).

Reluctant to do algebra without the negatives (and to lose all the major algebraic results dependent on these numbers), some early nineteenth-century English mathematicians created symbolical algebra into which symbols were introduced without prior definition. In short, they developed the now-standard mathematical technique of reasoning on symbols whose meanings are unknown. Following his own maxim that meaning would follow and not precede, George Peacock, the original formulator of the approach, put the negatives into his system of algebra by assumption and without definition (see [20, pp. 194–196]). He thus did what modern mathematicians do: he simply stated that basic algebra would include the numbers $+a$ and $-a$ which obey certain rules.

In 1835 De Morgan, a student and follower of Peacock, published the elementary algebra textbook which is satirized below. In this work he supported a basically symbolical approach to algebra. Yet, perhaps to ease the beginning student's transition from arithmetic to algebra, he tried to justify the use of the negative numbers by “explaining” them as easily rectifiable mistakes. According to De Morgan, negatives arose from misinterpretation of the conditions of algebraic problems. A misinterpretation could, for example, introduce a negative number as the result of an impossible subtraction (where the subtrahend was greater than the minuend). The mathematician, however, could easily correct the mistake by inverting the terms of the impossible subtraction and reversing the “quality” of the original answer (see [5, p. 18]). An example will help. Suppose a mathematician is asked to calculate the profit or loss sustained when a person invests an amount a in a project and receives in return an amount b . Assuming that a loss is sustained, the mathematician calculates the loss by subtracting b from a . Now suppose that $a - b$ is a negative number. We know immediately that the mathematician's interpretation of this business enterprise was wrong. The investor made money, and the profit was $-(a - b)$, or $b - a$.

The following play pokes fun at De Morgan's use of inversion of the terms of subtraction and reversal of signs to explain the negative numbers, as well as at the negative numbers in general and the unrestricted subtraction operation. In the play, De Morgan's students cite the principle that the greater may be taken from the lesser as justification of their withdrawal from his classes. The overall impression derived from the play is that its author(s) felt that the symbolical approach

stripped algebra of all rhyme and reason. Algebra, they believed, had been reduced to the quite unreasonable whims of mathematicians, such as De Morgan.

Two mathematicians of the period mentioned in Scene One may not be familiar to all readers. Dionysius Lardner was an Irish mathematician and scientist who in the late 1820s received a gold medal for lectures on the steam engine. Charles Babbage was an English mathematician who is best known today for his formulation of the basic principles of modern computers. It should also be mentioned that the term *senior wrangler* refers to the top student of his class at Cambridge, determined primarily by performance on a mathematical test, called the *tripos*.

Today's beginning algebra students may profitably compare and contrast their attitudes towards modern algebra and its professors with those of the play's author(s) towards symbolical algebra and De Morgan. Discussion questions raised by the play include the following. Is modern algebra as arbitrary as implied in the play? Is it merely a product of the whims of mathematicians? What rules or guidelines do mathematicians follow in developing algebraic systems? Are the problems posed and solved in the play's last scene good algebra? What is the difference between good mathematics and "mathematical nonsense"?

**Illustrations of the Study of Mathematics
after the Manner of Miss Martineau**

"I consider that woman a demon—a fiend in human form! She delights in everything that is most degraded in society and when she cannot find anything bad to describe she invents it."

PROFESSOR DE MORGAN

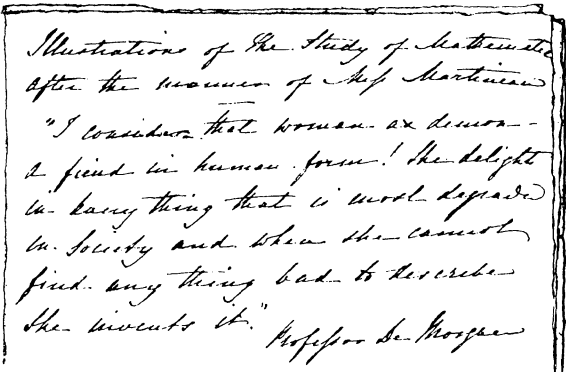
Dedication

To the Author of "Elements of Algebra" &c.

Sir-

The following pages compiled solely with a view to illustrate & describe the incalculable advantages of your system of teaching Mathematics are but a faint shadowing forth of all the excellencies of that inestimable method of generalization which enables the learner to adapt the operations & the symbols of Algebra to every possible elevation of sense & every impossible depth of nonsense for, as your invaluable production most lucidly sets forth, it is only necessary to place the negative sign before a word or an expression to alter its entire meaning. To whom can this ____ sense be so justly inscribed as to you Sir who are the Personification of Sense, & as such, command the overpowering *Respect* of —

THE AUTHOR



Scene 1. Trinity College Cambridge

1st student. Have you heard the news?

2nd student. No. What news?

1st. They are going to work Mathematics by steam in Gower St.

2nd. What do you mean?

1st. Upon my honor, I mean what I say. There is a Professor there of 1000 man power. I suppose it is a Steam Engine, or some of those infernal machines which Dr. Lardner and Mr. Babbage are always inventing to blow up the Church & the King.

3rd student. You have not got the right story. It's a man they have, who writes his name like this. (*Writes A De Morgan*¹⁰⁰⁰). He can do in every way 1000 times as much as any other man. He causes tremendous disturbances whenever he walks about, for he has 1000 times the specific gravity of any equal substance and it's thought, at least he says, that he has such immense power of attraction that he will pull the moon out of its orbit & alter the tides, & square the circle, & find the longitude, and,—(*Student stops, quite out of breath.*)

A freshman. But what is the meaning of 1000, written over his name?

The gentleman who expects to be senior wrangler (solemnly). That means that the gentleman has the power of 1000. He is raised to the thousandth power.

(*All the students debate together and at last they say they could learn Mathematics much cheaper and easier from Professor De Morgan*¹⁰⁰⁰ *than from any one else, which they settle in the following manner.*)

1st. Here are 20 of us. If that chap can teach us 1000 times as much as any one else in the same time, we can all of us learn as much mathematics in a day from him as we could in fifty days here. Besides, we need only take $1/1000$ part of the pains we are obliged to do here, & we shall get up our Mathematics much cheaper.

(*Students all come up to London in a body.*)

Scene 2. A room in University College, London.

Professor De Morgan reading The Way to Salvation, or, a Peep into Purgatory. (Enter 20 students.)

Mr. Kennel. Mr. De Morgan, my good Sir, here come twenty students from Cambridge, attracted by the fame of your scientific attainments!

Professor. (*Without stirring or looking up*) Ah! (*Goes on reading the Peep into Purgatory.*)

1st student. Sir, we have heard at Cambridge, that you have the power of 1000 men, and we are come to see if you can make us mathematicians in the twinkling of an eye.

P. We will try. (*Looks at his watch.*)

All. Then Sir, we will attend your course. But as we hear that you have 1000 times the power of any one else, certainly, you can teach us in a thousandth part of the time, so we suppose that the fee is one thousandth part of what any other professor ever had, £7/1000 or $1^d\ 17/25$. You see, Sir, I'm a dab at fractions but as we Cambridge men always do things liberally, we will make it two pence for the course.

P. Stay—Stay. A little too fast. We will turn to my book on Algebra, where you will find the anomaly you speak of, explained. (*P. takes up a book & after reading a great many p's & q's goes on.*) It is evident that we have made this equation wrong. All that we have to do, is, to suppose the case the direct reverse of what we have stated it.

The numbers will continue the same, but the sign of every term will be altered. Now gentlemen, to apply this. It is plain that you have misstated your problem. Suppose that instead of receiving $\frac{1^{th}}{1000}$ part of what other Professors have I were to receive 1000 times as much. The equation then becomes: Fee for one course = $\text{£}7 \times 1000$. This renders the problem rational for it is evident that a man who has the power of 1000 men must receive as much as 1000 others. *Vide* page 34 of my Elements of Algebra wherein I explain the meaning of half a horse, two men & three quarters &c.

Students. What a sell!

2nd student. Can't be helped. We must pay.

(Students pay £7000 each.)

Scene 3. Mathematical Lecture Room.

Professor. Before entering into this course of lectures, I must request that should any gentlemen find a difficulty in comprehending my meaning, he will state it either at the time or afterwards. I always set aside an hour for the express purpose of explaining difficulties. (*Looks at his watch, has not the remotest idea what the time is, & puts his watch into his pocket again.*) However, it may as well be understood at once, that if there are any persons who cannot understand what I say to the Senior Class, the Junior Class is the fit place for them, and if there are any who cannot understand what I say to the Junior Class, or in other words who cannot read fifty pages of my book in five minutes, all I can say is, such persons had better not learn Mathematics! (*Immense applause.*) Now having laid down these axioms, postulates, and common notions, we will begin. (*Professor takes a pinch of snuff.*) The first thing which occurs to the student of Algebra, is, that the subject is a hard one. (*Applause*) For instance—solve this problem. *A*, the lesser = 1; *B*, the greater = 20. What value of *A* will make it fifty times as great as *B*?

(All the students whisper to each other without answering, but stare & look very much bewildered.)

(In a few minutes the Professor says) There is only one case in which this can occur, & only three values which will satisfy the equation. I do not make an equation. It will suffice to tell you the values, & you must verify them yourselves.

A, (I by myself) 1, = 1000

B, my class, = 20

Here is an instance of the smaller number, one, equal to fifty times the greater, twenty. But as this is an exception, we must thence deduce the general rule.

A very brave student. I suppose, Sir, that this is French Mathematics.

P. Yes. We will now read my book. (*Reads sixty pages at the end of which he gives this*)

Problem—*A* may succeed in anything he likes. Only let him try, he'll do it. *A* tries to learn *X* of *B*. *B* always tells truth, and *B* tells *A* that if he cannot understand what he says, he (*A*) had better not try to learn *X*. Give the respective values of *A*, *B*, & *X*.

(Students scribble on little bits of paper, & at length they say that X is an impossible quantity.)

P. Stay. Stay. It appears to you that the result is not rational.

See how I state it

A = a student
 B = myself
 X = Algebra.

On verifying the problem we find nothing irrational, for X , Algebra, is the essence of rationality. B (myself), the personification of ditto, and you gentlemen can best appreciate the value of A .

(*Overpowering applause*)

Students. How fast we learn. We get on like bricks!

P. Now from these anomalies, some general rules may be deduced. Namely. That unless there is any reason to suppose the contrary, the greater number may always be taken from the lesser—2nd. That the values of terms are entirely arbitrary as well as every operation in Algebra.

Students. Then, Sir, as terms are entirely arbitrary, we think it very arbitrary in you to name such high terms for your course of lectures, and as the greater number may always be taken from the lesser I vote that we all take ourselves away to Cambridge.

(*Students all go back.*)

References

The best source for concise biographies of the individuals mentioned in this paper is the *Dictionary of Scientific Biography*, edited by Charles C. Gillispie.

- [1] Anonymous, Illustrations of the Study of Mathematics after the Manner of Miss Martineau, MS. Add. 163, Correspondence of Augustus De Morgan, D. M. S. Watson Library, University College London. (The author is indebted to The Library, University College London, for permission to reproduce this manuscript.)
- [2] Harvey W. Becher, Woodhouse, Babbage, Peacock, and modern algebra, *Historia Math.*, 7 (1980) 389–400.
- [3] Michael J. Crowe, A History of Vector Analysis: The Evolution of the Idea of a Vectorial System, Univ. of Notre Dame Press, Notre Dame, IN, 1967.
- [4] Augustus De Morgan, A Budget of Paradoxes, 2 vols., Open Court Publishing, Chicago, 1915.
- [5] ———, Elements of Algebra Preliminary to the Differential Calculus, 2nd ed., London, 1837. (In the Preface to this work De Morgan wrote: “This Second Edition differs from the first only in verbal amendments.”)
- [6] Augustus De Morgan, William Frend, Memoirs of the Royal Astronomical Society, 12 (1842) 458–468.
- [7] Sophia Elizabeth De Morgan, Memoir of Augustus De Morgan, Longmans, Green, and Company, London, 1882.
- [8] J. M. Dubbey, Babbage, Peacock and modern algebra, *Historia Math.*, 4 (1977) 295–302.
- [9] ———, The Mathematical Work of Charles Babbage, Cambridge Univ. Press, 1978.
- [10] William Frend, The Principles of Algebra, 2 vols., London, 1796–1799.
- [11] William Rowan Hamilton, Theory of conjugate functions, or algebraic couples; with a preliminary and elementary essay on algebra as the science of pure time, *Trans. Roy. Irish Acad.*, 17 (1837) 293–422.
- [12] Thomas L. Hankins, Algebra as pure time: William Rowan Hamilton and the foundations of algebra, in Motion and Time, Space and Matter: Interrelations in the History of Philosophy and Science, Peter K. Machamer and Robert G. Turnbull, eds., Ohio State Univ. Press, Columbus, 1976, pp. 327–359.
- [13] Thomas L. Hankins, Sir William Rowan Hamilton, Johns Hopkins Univ. Press, Baltimore, 1980.
- [14] Frida Knight, University Rebel: The Life of William Frend (1757–1841), Victor Gollancz, London, 1971.
- [15] Elaine Koppelman, The calculus of operations and the rise of abstract algebra, *Arch. Hist. Exact Sci.*, 8 (1971) 155–242.
- [16] Francis Maseres, A Dissertation on the Use of the Negative Sign in Algebra, London, 1758.
- [17] Ernest Nagel, “Impossible numbers”: a chapter in the history of modern logic, *Studies in the History of Ideas*, 3 (1935) 429–474.
- [18] Isaac Newton, Universal Arithmetick; or, A Treatise of Arithmetical Composition and Resolution, London, 1728, in The Mathematical Works of Isaac Newton, Derek T. Whiteside, ed., Johnson Reprint Corp., New York, 1967, vol. 2, pp. 3–134.
- [19] Luboš Nový, Origins of Modern Algebra, Jaroslav Tauer, translator, Academia, Prague, 1973.
- [20] George Peacock, Report on the recent progress and present state of certain branches of analysis, Report of the Third Meeting of the Brit. Assoc. for the Advancement of Science, 3 (1833) 185–352.
- [21] ———, A Treatise on Algebra, London, 1830.
- [22] Helena M. Pycior, George Peacock and the British origins of symbolical algebra, *Historia Math.*, 8 (1981) 23–45.

A Genealogy of 120° and 60° Natural Triangles

ALAN WAYNE

Pasco-Hernando Community College
New Port Richey, FL 33552

A fascinating part of elementary number theory deals with natural triangles, which are triangles having sides whose lengths are natural numbers. For example, the 90° natural triangles, more familiarly known as Pythagorean triangles, have been studied for more than three thousand years. Yet only recently was one of their most important properties discovered. In 1970 A. Hall proved that the most familiar 90° natural triangle of all, the (3,4,5) triangle, is the primordial ancestor of all other 90° natural triangles [1], [6]. Hall's genealogy is beautifully simple and shows that any 90° natural triangle is a branch on a tree generated by linear transformations acting on the (3,4,5) triangle.

We will first briefly describe Hall's work, then demonstrate a similar genealogy of the 120° natural triangles and also of the 60° natural triangles. Only primitive triangles need be considered—those in which the three side lengths have greatest common divisor 1.

An ordered triple of natural numbers (a, b, c) is Pythagorean if and only if $a^2 + b^2 = c^2$. Hall regards any Pythagorean triple as a row vector, a 1×3 matrix $[a \ b \ c]$. He then introduces the three matrices U, A, D , defined as follows:

$$U = \begin{bmatrix} 1 & 2 & 2 \\ -2 & -1 & -2 \\ 2 & 2 & 3 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 3 \end{bmatrix}, \quad \text{and} \quad D = \begin{bmatrix} -1 & -2 & -2 \\ 2 & 1 & 2 \\ 2 & 2 & 3 \end{bmatrix}.$$

These three matrices generate an infinite set S of 3×3 matrices which can be factored into finite products of U, A , and D . If $[a \ b \ c]$ is a primitive Pythagorean triple, and $M \in S$, then it is easy to show that $[a \ b \ c]M = [a' \ b' \ c']$ is also a primitive Pythagorean triple. Hall then demonstrates that to every primitive Pythagorean triple (other than [3 4 5]) there corresponds uniquely a matrix $M \in S$ which transforms [3 4 5] into the triple. His final result is that the set of all primitive Pythagorean triples may be exhibited as an infinite tree rooted in [3 4 5]. From each triple the branching is up (multiplication by U), or across (multiplication by A), or down (multiplication by D), as shown in FIGURE 1.

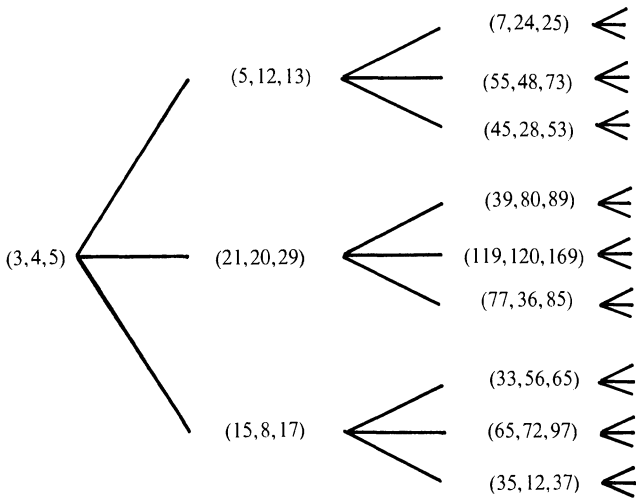


FIGURE 1. Hall's genealogy of primitive Pythagorean triangles.

We now consider the set of natural triangles (p, q, r) in which the angle opposite side r is 120° . By the Law of Cosines, (p, q, r) is in this set if and only if

$$r^2 = p^2 + q^2 + pq. \quad (1)$$

Those readers familiar with the traditional method of generating Pythagorean triples will recall that equations in two natural numbers m and n define the triples. Analogously, it is easily verified that given two natural numbers $m > n$, the following equations define a triple (p, q, r) in the set of 120° natural triangles:

$$p = m^2 - n^2, \quad q = 2mn + n^2, \quad \text{and} \quad r = m^2 + mn + n^2. \quad (2)$$

In 1887, Neuberg and Matthews [2] proved that *every* rational solution of (1) may be expressed in terms of two rational parameters $m > n$. If m and n have a common divisor d , equations (2) make clear that p , q , and r will have d^2 as a common divisor. Also, since these equations can be written as

$$p = (m - n)(m + n), \quad q = (m - n)^2 - m(m - n) + 3mn, \quad r = (m - n)^2 + 3mn, \quad (3)$$

it follows that if 3 divides $m - n$, then 3 is a common divisor of p , q , and r . Equations (3) also make it easy to show that each 120° triple (p, q, r) is determined by a unique pair of natural numbers $m > n$; solving for m and n in terms of p, q, r we have

$$m = [(p - q + 2r)/3]^{1/2} \quad \text{and} \quad n = [(-2p - q + 2r)/3]^{1/2}. \quad (4)$$

We define \mathbf{T} as the set of ordered pairs of natural numbers (u, v) which satisfy

- (i) $u > v$
 - (ii) u and v are relatively prime
 - (iii) $u - v \not\equiv 0 \pmod{3}$.
- (T)

Using our remarks above, it can be shown that equations (2) and (4), with m and n replaced, respectively, by u and v , with $(u, v) \in \mathbf{T}$, establish a one-to-one correspondence between the set \mathbf{T} and the set of primitive 120° natural triangles.

We wish to establish a 'genealogy' for the pairs $(u, v) \in \mathbf{T}$, and thereby establish an ancestral tree for 120° natural triangles. By definition of \mathbf{T} , if $(u, v) \in \mathbf{T}$, then $v \geq 1$, and conditions (i) and (iii) imply that the pairs $(2, 1)$ and $(3, 1)$ are the only ones with $v = 1$ and $u - v \leq 2$. Furthermore, if $(u, v) \in \mathbf{T}$ and $v > 1$, then u is not equal to $2v$, $(5/2)v$, $3v$, or $4v$, since in each of these cases either 3 divides $(u - v)$ or u and v are not relatively prime. Therefore \mathbf{T} can be partitioned into seven subsets (pairwise disjoint and whose union is \mathbf{T}):

$$\mathbf{F}_1 = \{(u, v) : u = 2v\} = \{(2, 1)\}$$

$$\mathbf{F}_2 = \{(u, v) : u = 3v\} = \{(3, 1)\}$$

$$\mathbf{T}_1 = \{(u, v) : v < u < 2v\}$$

$$\mathbf{T}_2 = \{(u, v) : 2v < u < (5/2)v\}$$

$$\mathbf{T}_3 = \{(u, v) : (5/2)v < u < 3v\}$$

$$\mathbf{T}_4 = \{(u, v) : 3v < u < 4v\}$$

$$\mathbf{T}_5 = \{(u, v) : 4v < u\}.$$

We will show that the two 'least' parameters, $(2, 1)$ and $(3, 1)$ in \mathbf{T} , will generate all others in \mathbf{T} . Define five nonsingular 2×2 matrices M_1, \dots, M_5 as follows:

$$M_1 = \begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix}, \quad M_2 = \begin{bmatrix} 2 & 1 \\ 3 & 1 \end{bmatrix}, \quad M_3 = \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix}, \quad M_4 = \begin{bmatrix} 3 & 1 \\ 1 & 0 \end{bmatrix}, \quad \text{and} \quad M_5 = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}.$$

LEMMA. (a) If $(u, v) \in T_i$, $i = 1, 2, 3, 4, 5$, and $[u v]M_i^{-1} = [u' v']$, then $(u', v') \in T$. Moreover, $u' < u$ and $v' < v$ for $1 \leq i \leq 4$, and for $i = 5$, $u' < u$ and $v' = v$.

(b) For every $(u, v) \in T$ other than $(2, 1)$ and $(3, 1)$, there is a unique product matrix M , all of whose factors are from the set $\{M_1, \dots, M_5\}$ such that $[u v] = [2 \ 1]M$ or $[u v] = [3 \ 1]M$.

Proof. (a) We sketch the proof for $i = 2$; similar verifications prove the other four cases. If $(u, v) \in T_2$, then $2v < u < (5/2)v$, and $[u' v'] = [u v]M_2^{-1}$ gives $u' = -u + 3v$ and $v' = u - 2v$. From these conditions we have

$$0 < u' = -u + 3v < -2v + 3v = v < u,$$

and

$$0 < v' = u - 2v < (5/2)v - 2v < v.$$

Also,

$$u' - v' = -2u + 5v = 5(v - u) + 3u,$$

so $u' > v'$ and $u' - v' \not\equiv 0 \pmod{3}$ (since $(u, v) \in T$). Since $\det M_2^{-1} = -1$, u' and v' must have the same greatest common divisor as u and v , thus $(u', v') \in T$.

(b) If $(u, v) \in T$ is not $(2, 1)$ or $(3, 1)$, then $(u, v) \in T_i$ for a unique i , $1 \leq i \leq 5$. Let $[u' v'] = [u v]M_i^{-1}$; from (a) we know that $(u', v') \in T$ and $u' < u$, $v' \leq v$. Call (u', v') the predecessor of (u, v) in T . If (u', v') is not $(2, 1)$ or $(3, 1)$, we repeat the procedure: $(u', v') \in T$ and we find the predecessor (u'', v'') of (u', v') in T , defined by $[u'' v''] = [u' v']M_j^{-1}$ for a unique j , $1 \leq j \leq 5$. Again, $u'' < u'$ and $v'' \leq v'$. Since we have a strictly decreasing sequence of natural numbers $u > u' > u'' \cdots > 0$ (and at each stage, $u' > v' > 0$), the repeated process must ultimately stop with the 'earliest' predecessor of (u, v) which is necessarily $(2, 1)$ or $(3, 1)$.

EXAMPLE. Let $(u, v) = (9, 5) \in T$. Since $(9, 5) \in T_1$, the predecessor of $(9, 5)$ in T is $[9 \ 5]M_1^{-1} = [5 \ 1]$. Since $(5, 1) \in T_5$, the predecessor of $(5, 1)$ in T is $[5 \ 1]M_5^{-1} = [2 \ 1]$. Thus $[9 \ 5] = [2 \ 1]M_5M_1$.

Using equations (2), the parameter pairs $(2, 1)$ and $(3, 1)$ produce, respectively, the 120° natural triples $(3, 5, 7)$ and $(8, 7, 13)$. By virtue of the Lemma and the one-to-one correspondence defined by (2) or (4), every other primitive 120° natural triple can be obtained by multiplying either the vector $[3 \ 5 \ 7]$, or the vector $[8 \ 7 \ 13]$ by a product matrix, all of whose factors are from a set of five 3×3 matrices $\{N_1, \dots, N_5\}$, corresponding to the 2×2 matrices M_i , $1 \leq i \leq 5$.

We illustrate a procedure for deriving the matrix N_i from the matrix M_i by indicating how this is done for the case $i = 1$. Recall that $M_1 = \begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix}$; let

$$N_1 = \begin{bmatrix} k_1 & k_4 & k_7 \\ k_2 & k_5 & k_8 \\ k_3 & k_6 & k_9 \end{bmatrix}.$$

For $(u, v) \in T$, we denote $[u v]M_1 = [u_1 v_1]$. The pair (u, v) defines a 120° triple (p, q, r) using equations (2):

$$p = u^2 - v^2, \quad q = 2uv + v^2, \quad r = u^2 + uv + v^2,$$

and similarly, the pair (u_1, v_1) defines a 120° triple (p_1, q_1, r_1) . The matrix N_1 we seek satisfies the equation $[p \ q \ r]N_1 = [p_1 \ q_1 \ r_1]$, that is, it makes the following diagram commute:

$$\begin{array}{ccc} [u \ v] & \xrightarrow{(2)} & [p \ q \ r] \\ \downarrow M_1 & & \downarrow N_1 \\ [u_1 \ v_1] & \xrightarrow{(2)} & [p_1 \ q_1 \ r_1] \end{array}$$

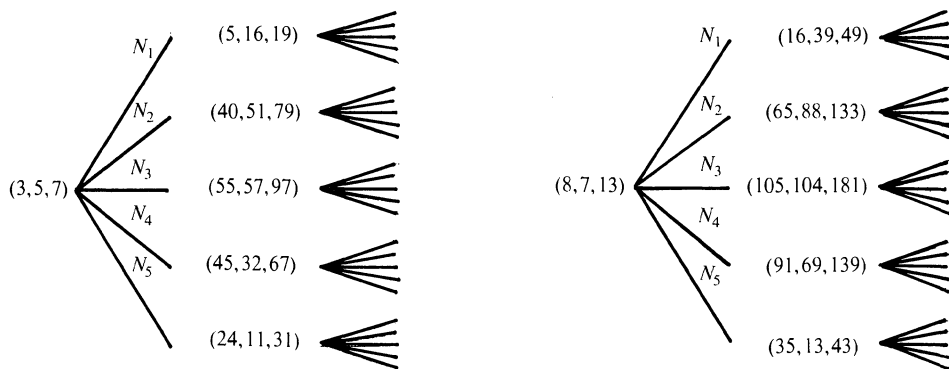


FIGURE 2. A genealogy of primitive 120° natural triangles.

Using this, the vectors $[p\ q\ r]$ and $[p_1\ q_1\ r_1]$ can be expressed in terms of the variables u and v to obtain nine linear equations in the k_i :

$$\begin{aligned} k_1 + k_3 &= 3, & -k_1 + k_2 + k_3 &= 1, & 2k_2 + k_3 &= -4, \\ k_4 + k_6 &= 5, & -k_4 + k_5 + k_6 &= 0, & 2k_5 + k_6 &= -2, \\ k_7 + k_9 &= 7, & -k_7 + k_8 + k_9 &= 1, & 2k_8 + k_9 &= -5. \end{aligned}$$

Now one can solve for the respective k_i , to obtain the matrix N_i . The five matrices N_i are:

$$\begin{aligned} N_1 &= \begin{bmatrix} -1 & 1 & 0 \\ -4 & -3 & -6 \\ 4 & 4 & 7 \end{bmatrix} & N_2 &= \begin{bmatrix} -1 & 1 & 0 \\ 3 & 4 & 6 \\ 4 & 4 & 7 \end{bmatrix} & N_3 &= \begin{bmatrix} 4 & 3 & 6 \\ 3 & 4 & 6 \\ 4 & 4 & 7 \end{bmatrix} \\ N_4 &= \begin{bmatrix} 4 & 3 & 6 \\ 1 & -1 & 0 \\ 4 & 4 & 7 \end{bmatrix} & N_5 &= \begin{bmatrix} -3 & -4 & -6 \\ 1 & -1 & 0 \\ 4 & 4 & 7 \end{bmatrix}. \end{aligned} \quad (5)$$

Note that $|\det N_i| = 1$ for all i ; thus the image $[p\ q\ r]N_i$ for any primitive 120° natural triple (p, q, r) is another such triple. The genealogy established in the Lemma for the set \mathbf{T} now gives rise to our main result.

THEOREM. *The complete set of primitive 120° natural triangles can be represented as a pair of infinite fivefold-branching trees, one rooted in $(3, 5, 7)$, and the other in $(8, 7, 13)$.*

FIGURE 2 shows the first branching of each of the trees of primitive 120° natural triangles.

A 60° natural triangle is represented by a triple of natural numbers (p, q, r) , where r is the side opposite the 60° angle. By the Law of Cosines, (p, q, r) is in this set if and only if $r^2 = p^2 + q^2 - pq$. Our results for the primitive 120° natural triangles lead at once to a genealogy for the set of primitive 60° natural triangles. FIGURE 3 makes clear the fact that to each primitive 120° natural

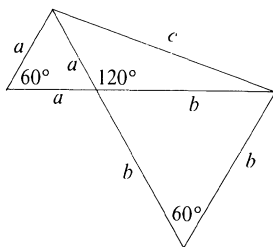


FIGURE 3. The 2-to-1 correspondence between 60° and 120° natural triangles. The 60° triangles $(a, a+b, c)$ and $(b, a+b, c)$ determine the 120° triangle (a, b, c) and vice-versa.

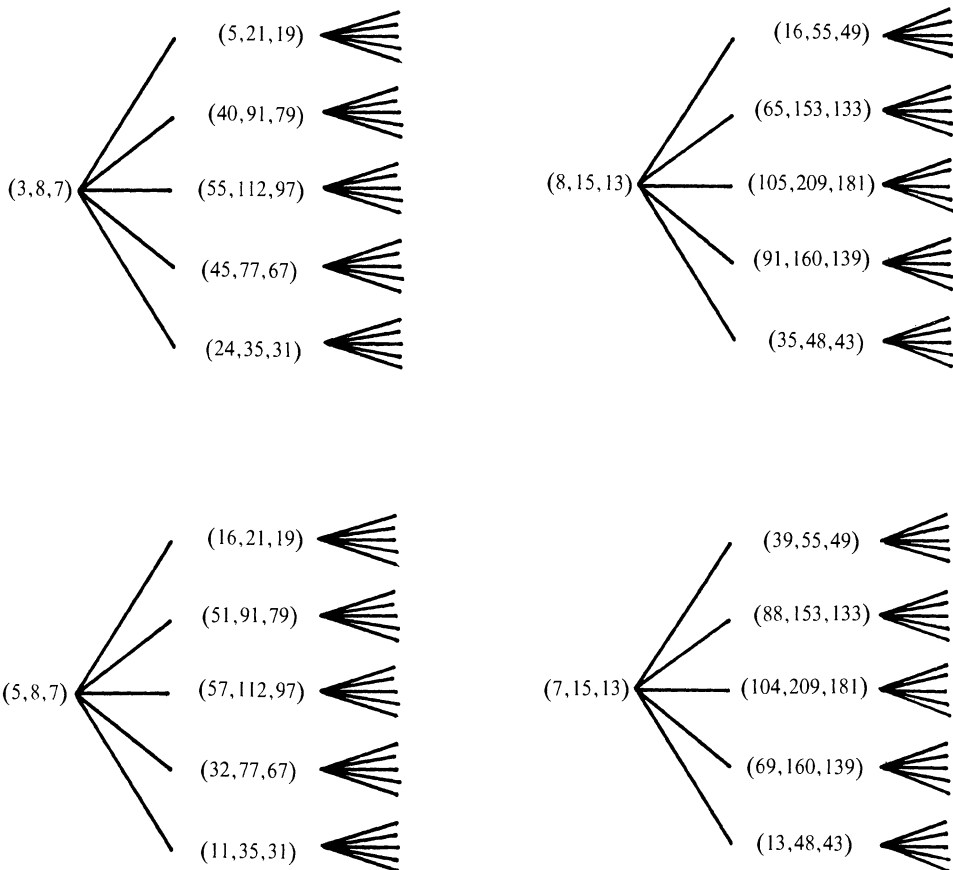


FIGURE 4. A genealogy of primitive 60° natural triangles.

triangle (a, b, c) with $a < c$ and $b < c$, there corresponds two distinct primitive 60° natural triangles, namely, $(a, a + b, c)$ and $(b, a + b, c)$, and conversely. (This correspondence was noted, though incorrectly quoted in my remark which appears in [5].) Consequently, *the totality of primitive 60° natural triangles may be represented by four infinite fivefold-branching trees, rooted respectively in $(3, 8, 7)$, $(5, 8, 7)$, $(8, 15, 13)$, and $(7, 15, 13)$.* From each triangle we proceed to five others by making use of the same five matrices N_1, \dots, N_5 . FIGURE 4 shows the first branches of these four trees.

Our results have been obtained mainly by exploiting the simple isomorphism of sets, $\{(u, v)\} \leftrightarrow \{(p, q, r)\}$, together with certain linear transformations in matrix notation, $M_i \leftrightarrow N_j$. Using a somewhat different approach, Pollina and Snover have arrived at the same results [4]. One way to summarize these results, together with those of Hall, is the following: for $k = 0, -1/2$, or $1/2$, the set of natural triangles in which all the triangles have a common angle $\theta = \arccos k$, can be generated by certain linear transformations from certain “root triangles.” For $k = 0, -1/2$, and $1/2$, there are, respectively, one, two, or four root triangles at the base of the tree containing all of the natural triangles having an angle $\theta = \arccos k$. We might ask if there are other values of k for which $\theta = \arccos k$ is rational in degrees, and for which a similar genealogy can be developed. Pollina and Snover point out that these are the *only* values of k which are possible. For if (a, b, c) is a natural triangle, then $\cos \theta$ is rational, and it is known [3] that if θ is also rational in degrees, then $\cos \theta = -1, -1/2, 0, 1/2$, or 1 . So this beautiful property is possessed only by the 60°, 120°, and 90° natural triangles.

References

- [1] A. Hall, Genealogy of Pythagorean triples, *Math. Gaz.*, 54 (1970) 377–379.
- [2] J. Neuberg and G. B. Matthews, *Math. Quest. Educ. Times*, 46 (1887) 97. This is cited in Leonard Eugene Dickson, *History of the Theory of Numbers*, vol. 2, G. E. Stechert & Co., New York, 1934, p. 406.
- [3] Ivan Niven, *Irrational Numbers*, The Carus Mathematical Monographs, no. 11, MAA, 1967, p. 41.
- [4] Benedict Pollina and Stephen Snover, 120° and 60° triples, *Pi Mu Epsilon Journal*, 7 (1981).
- [5] David P. Robbins and Kenneth L. Yocum, Solution to problem E 2566, *Amer. Math. Monthly*, 84 (1977) 220–221.
- [6] Joe Roberts, *Elementary Number Theory, A Problem Oriented Approach*, MIT Press, Cambridge, MA, 1977, pp. 84–85, 107S–109S.

Why Your Classes Are Larger than “Average”

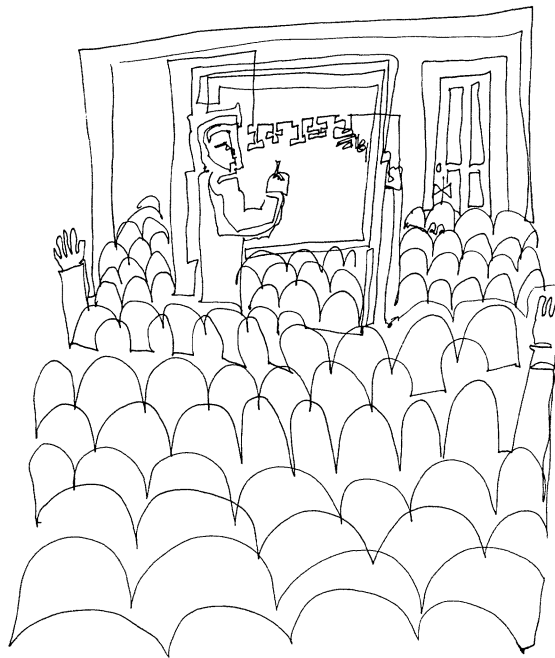
DAVID HEMENWAY

Harvard University School of Public Health

Boston, MA 02115

Most schools advertise their “average class size,” yet most students find themselves in larger classes most of the time. Here is a typical example.

In the first quarter of the 1980–81 academic year, 111 courses including tutorials, were given at Harvard School of Public Health. These ranged in size from one student to 229. The **average class size**, from the administration’s and professors’ perspective, was 14.5. The **expected class size** for a typical student was over 78! This huge discrepancy was due to the existence of a few very large classes. Indeed, only three courses had more than 78 students. One enrolled 105, another 171, and there were 229 in Epidemiology.



References

- [1] A. Hall, Genealogy of Pythagorean triples, *Math. Gaz.*, 54 (1970) 377–379.
- [2] J. Neuberg and G. B. Matthews, *Math. Quest. Educ. Times*, 46 (1887) 97. This is cited in Leonard Eugene Dickson, *History of the Theory of Numbers*, vol. 2, G. E. Stechert & Co., New York, 1934, p. 406.
- [3] Ivan Niven, *Irrational Numbers*, The Carus Mathematical Monographs, no. 11, MAA, 1967, p. 41.
- [4] Benedict Pollina and Stephen Snover, 120° and 60° triples, *Pi Mu Epsilon Journal*, 7 (1981).
- [5] David P. Robbins and Kenneth L. Yocum, Solution to problem E 2566, *Amer. Math. Monthly*, 84 (1977) 220–221.
- [6] Joe Roberts, *Elementary Number Theory, A Problem Oriented Approach*, MIT Press, Cambridge, MA, 1977, pp. 84–85, 107S–109S.

Why Your Classes Are Larger than “Average”

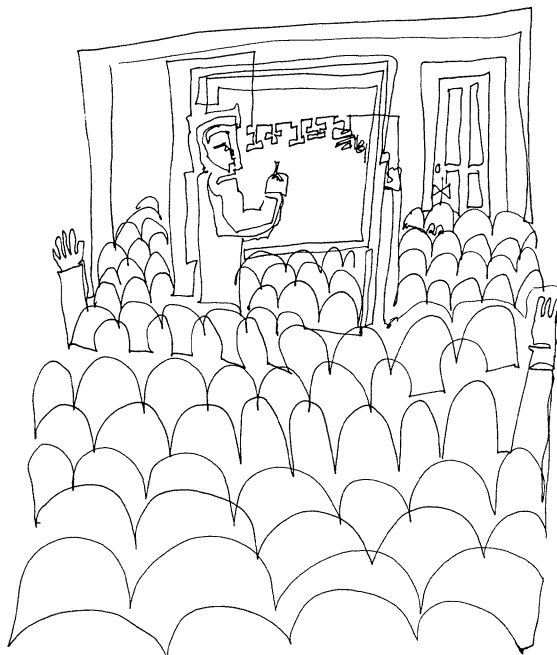
DAVID HEMENWAY

Harvard University School of Public Health

Boston, MA 02115

Most schools advertise their “average class size,” yet most students find themselves in larger classes most of the time. Here is a typical example.

In the first quarter of the 1980–81 academic year, 111 courses including tutorials, were given at Harvard School of Public Health. These ranged in size from one student to 229. The **average class size**, from the administration’s and professors’ perspective, was 14.5. The **expected class size** for a typical student was over 78! This huge discrepancy was due to the existence of a few very large classes. Indeed, only three courses had more than 78 students. One enrolled 105, another 171, and there were 229 in Epidemiology.



Given one class of the size of Epidemiology, an expected class size of approximately 78 for a typical student can be achieved in various ways. Four possible configurations for the rest of the classes are: (i) 450 individual tutorials, (ii) 50 courses of size 10, (iii) 25 courses of size 30, (iv) 25 courses of size 50. The administration's "average class size" for these four cases would be 1.5, 14.3 (close to the advertised figure), 38, and 57 respectively.

The discrepancy between average class size and expected class size for a typical student is explained by a simple computation. Suppose we have a population of M individuals divided into N groups, and we let X_i denote the size of the i th group, $1 \leq i \leq N$. Then the expected number of people in a randomly selected group ("average class size") is given by

$$\bar{X} = (\sum X_i)/N = M/N,$$

and the expected size of a group containing a randomly selected individual is given by

$$X^* = \sum (X_i/M) X_i = (\sum X_i^2)/M.$$

Hence

$$\begin{aligned} X^* - \bar{X} &= [N \sum X_i^2 - (\sum X_i)^2]/MN \\ &= \left[\frac{N \sum X_i^2 - (\sum X_i)^2}{N^2} \right] \frac{N}{M} \\ &= \sigma^2 / \bar{X}, \end{aligned}$$

where σ^2 is the variance in group sizes.

The difference between the two means \bar{X} and X^* is directly proportional to the variance in sizes of groups and inversely proportional to average group size. It follows that $X^* \geq \bar{X}$, with equality only when all the groups are the same size.

Here are additional examples from everyday life of the differences between \bar{X} and X^* .

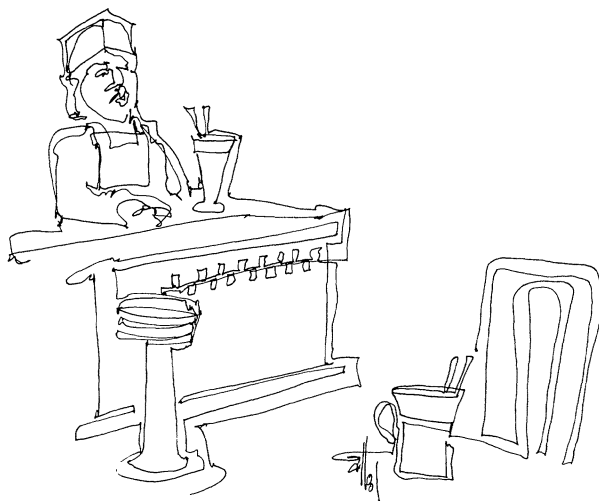
The Nationwide Personal Transportation Survey indicated that average car occupancy (\bar{X}) for "home-to-work" trips in metropolitan areas in 1969 was 1.4 people. The table below gives the data.

Number of Occupants	"Home-to-Work" Trips
1	73.5%
2	18.2
3	4.7
4	1.9
5	1.1
6	.5
7	.1

Calculating X^* from these statistics we find that the average number of occupants in the car of a typical commuter was 1.9.

To eliminate most congestion problems in U.S. cities would only require raising the average number of people per car (\bar{X}) to 2. This doesn't sound impossible. But suppose this were accomplished by inducing some drivers of single-occupant vehicles to join together in five-person car pools. The percentage of single-occupant cars would need to fall to 58.7%; five-occupant cars would rise to 15.9%. The percentage of people in single-occupant cars would fall below 30%. If X^* is calculated for this situation, one finds that the typical commuter would be in a car carrying more than three people.

I often buy dinner at a fast-food restaurant near my home. Although most customers order "to go," the place is almost always crowded, and I consider it quite a success. One evening about 6:30 I went in and there was no one in line. The manager was serving me, so I asked, "Where is everyone?" "It often gets quiet like this," he said, "even at dinnertime. The customers always seem to come in spurts. Wait fifteen minutes and it will be crowded again." I was surprised that I



had never before seen the restaurant so empty. But I probably shouldn't have been. If I am a typical customer, I am much more likely to be there during one of the spurts, so my estimate of the popularity of the restaurant (X^*) is likely to be much greater than its true popularity (\bar{X}).

The average number of people at the beach on a typical *day* will always be less than the average number of people the typical *beach-goer* finds there. This is because there are lots of people at the beach on a crowded day, but few people are ever there when the beach is practically deserted.

If the waiting time at a health clinic increases with the number of patients, the average waiting time for a typical *day* will always be less than the average waiting time for a typical *patient*. This is because there are more patients waiting on those days when the waiting time is especially long.

The expected size of a typical generation will be smaller than the expected number of contemporaries for a randomly chosen individual from one of those generations.

Figures for the population density of any region will understate the actual degree of crowding for the average inhabitant.

This Note distinguishes mathematically between two types of means. It does not report any original findings about human behavior. Yet it does indicate something about perceptions—especially my own. I was surprised at the restaurant. I was also surprised when the courses I took in college were larger than advertised. And I was surprised to realize how many commuters had to carpool to reach an average of even two people per car. If you are similarly surprised by any of these observations, your perceptions and perhaps even your behavior may be affected.

Helpful comments were received from Frederick Mosteller and an anonymous referee.

Kirchhoff's Third Law

The differential equations instructor, confronted with electrical circuit problems, may have trouble remembering names of components and corresponding units. The following mnemonic helps.

Remove the first, middle, and last letters from "Kirchhoff." This leaves triplets IRC and HOF. The former suggests "inductors, resistors, capacitors;" the latter "henries, ohms, farads."

—MARLOW SHOLANDER
Case Western Reserve University
Cleveland, OH 44106

The Identity $(XY)^n = X^nY^n$: Does It Buy Commutativity?

HOWARD E. BELL

Brock University

St. Catharines, Ontario, Canada L2S 3A1

Let $(S, *)$ be a set with an associative binary operation, which we shall think of as multiplication; denote the product $x * y$ by xy . If the operation $*$ is also commutative, then S satisfies the identity

$$(xy)^n = x^n y^n \quad (1)$$

for each positive integer n . Conversely, suppose that S satisfies (1) for one or more $n > 1$. Need $*$ be commutative? If not, under what additional hypotheses will $*$ be commutative? This problem is a natural one, and interesting answers can be obtained by using techniques covered in a first abstract algebra course. It is, therefore, somewhat surprising that the problem, at least in its ring-theory version, has only recently been investigated.

Let F be any field and consider the set of 3×3 matrices

$$M = \left\{ \begin{bmatrix} 0 & a & b \\ 0 & 0 & c \\ 0 & 0 & 0 \end{bmatrix} \mid a, b, c \in F \right\}.$$

Ordinary matrix multiplication is a binary operation on M , and $xyz = 0$ for all $x, y, z \in M$; hence M satisfies (1) for every $n \geq 2$. Since matrix multiplication on M is clearly noncommutative, it is already evident that something more than (1) must be assumed in order to prove commutativity.

Some results for groups

The positive result which sparked recent interest in the problem is an easy and probably long-known result for groups, which I first encountered as an exercise in Herstein's textbook on algebra [10].

THEOREM 1. *Let G be a group, and suppose that there exist three consecutive positive integers n for which G satisfies (1). Then G is a commutative group.*

Proof. Suppose G satisfies (1) for $n = k, k + 1, k + 2$. Making use of (1) for $n = k$ and $n = k + 1$, we obtain, for arbitrary $x, y \in G$,

$$x^{k+1}y^{k+1} = (xy)^{k+1} = (xy)^k(xy) = x^k y^k xy;$$

and cancelling x^k on the left and y on the right gives

$$xy^k = y^k x. \quad (2)$$

Repeating the argument with $k + 1$ and $k + 2$ gives

$$xy^{k+1} = y^{k+1}x; \quad (3)$$

substituting (2) into (3) we get

$$xy^{k+1} = yxy^k,$$

which implies $xy = yx$.

It is almost obvious that a group G satisfying (1) for $n = 2$ must be commutative; and while the analogous result for a single n greater than 2 does not hold, groups satisfying (1) for even one $n > 1$ are somewhat restricted in their behavior. (For details, see Alperin's paper [2], the major theorem of which is accessible to anyone with a little knowledge of free groups.)

A careful look at the proof of Theorem 1 yields information about groups satisfying (1) for two consecutive n : specifically, G is commutative if it satisfies (1) for $n = k$ and $n = k + 1$ and if every

element of G is of the form y^k for some $y \in G$. But satisfying (1) for two consecutive n does not by itself guarantee commutativity, as we see by considering the following example from [22].

EXAMPLE 1. Let Z_{10} denote the integers mod 10, with $+$ and \cdot denoting the usual operations, and let $G = \{(a, b, c) \mid a, b, c \in Z_{10}\}$. If the operation $*$ is defined on G by $(a, b, c) * (a', b', c') = (a + a', b + b', c + c' + 2a \cdot b')$, then G is a noncommutative group under $*$; however it is easily verified that G satisfies (1) for $n = 5$ and $n = 6$.

The problem for rings

Since the essential mechanism in the proof of Theorem 1 is cancellation, it is immediate that there is a version of Theorem 1 for rings without zero divisors. However, since these constitute a relatively small class of rings, it is reasonable to ask what conditions in addition to (1) will yield commutativity in more general rings. The first authors to consider this question were Johnsen, Outcalt, and Yaqub [13], who proved in 1968 that a ring having a multiplicative identity element must be commutative if it satisfies (1) with $n = 2$. Luh [17] in 1971 established commutativity of certain rings having a multiplicative identity element and satisfying (1) for three consecutive n . Several years later Anthony Richoux, at that time an undergraduate, and Steve Ligh, one of Richoux's professors, succeeded in proving the following ring analogue of Theorem 1 [16].

THEOREM 2. *Let R be a ring with a multiplicative identity element, and suppose R satisfies the identity (1) for three consecutive positive integers n . Then R is commutative.*

In the discussion that follows we shall use the symbol 1 to denote the multiplicative identity element of the ring R , and use expressions such as “ R has 1” or “a ring R with 1” to indicate that R has a multiplicative identity element. We shall also use the term “polynomial function of two variables on R ” to denote a function such as $f(x, y) = x^2yx + yxyx + y^4x^4$, where the x and y range over elements of R . (Note that since R is not assumed to be commutative, this function f is distinct from the function $g(x, y) = x^3y + x^2y^2 + x^4y^4$.) Among the important polynomial functions is the **commutator function** or **bracket function**, defined by $[x, y] = xy - yx$. Clearly, $[x, y] = 0$ if and only if the elements x and y commute. Moreover, $[,]$ is linear in each component; hence if R has 1, it follows that $[x + 1, y] = [x, y + 1] = [x, y]$ for all $x, y \in R$.

The Ligh-Richoux proof of Theorem 2, which is astonishingly simple, depends on a limited cancellation property in rings with 1.

LEMMA. *Let R be a ring with 1, and suppose f is any polynomial function of two variables on R with the property that $f(x + 1, y) = f(x, y)$ for all $x, y \in R$. If there exists a positive integer n such that $x^n f(x, y) = 0$ for all $x, y \in R$, then $f(x, y) = 0$ for all $x, y \in R$.*

Proof. Given that $x^n f(x, y) = 0$, replace x by $x + 1$, obtaining

$$0 = (x + 1)^n f(x + 1, y) = \left(x^n + nx^{n-1} + \binom{n}{2} x^{n-2} + \cdots + nx + 1 \right) f(x, y), \quad (4)$$

where the $\binom{n}{i}$ are the usual binomial coefficients. Left-multiplying (4) by x^{n-1} and using the fact that $x^n f(x, y) = 0$, we get $x^{n-1} f(x, y) = 0$; and simply repeating the argument finitely many times yields $f(x, y) = 0$.

Proof of Theorem 2. Suppose R satisfies (1) for $n = k, k + 1$, and $k + 2$. We begin as in the proof of Theorem 1, noting that the equation $x^{k+1}y^{k+1} = x^k y^k xy$ can be rewritten as $x^k [x, y^k] y = 0$. Repeat the argument, using $n = k + 1$ and $n = k + 2$ and apply the Lemma to obtain

$$[x, y^k] y = 0 \quad \text{and} \quad [x, y^{k+1}] y = 0 \quad \text{for all } x, y \in R. \quad (5)$$

Now left-multiply the first equation in (5) by y , obtaining $yxy^{k+1} = y^{k+1}xy$, and note that the second equation in (5) may be expressed as $xy^{k+2} = y^{k+1}xy$. Therefore, $xy^{k+2} = yxy^{k+1}$, which says

$$[x, y] y^{k+1} = 0 \quad \text{for all } x, y \in R. \quad (6)$$

A right-hand version of the Lemma now yields $[x, y] = 0$ for all $x, y \in R$.

As in the group case, we cannot get by in Theorem 2 with only two consecutive n .

EXAMPLE 2. Let R_1 be the set of all ordered 4-tuples with entries from the integers mod 10; define addition componentwise and define multiplication by

$$(a, b, c, d)(a', b', c', d') = (aa', ab' + ba', ac' + ca', ad' + da' + 2bc').$$

It is readily verified that R_1 is a noncommutative ring under these operations, and that the 4-tuple $(1, 0, 0, 0)$ is a multiplicative identity element, which we denote as usual by 1. Let W be the set of 4-tuples with first component 0; note that for all $w_1, w_2, w_3 \in W$, we have $w_1 w_2 w_3 = 0$ and $5w_1 w_2 = 0$. Observe also that every element of R_1 can be written as $k1 + w$ for some integer k and some $w \in W$. Now let $x = j1 + u$ and $y = k1 + v$ be arbitrary elements of R_1 , where j and k are integers and $u, v \in W$. Then $xy = jk1 + jv + ku + uv = jk1 + w_0$, where $w_0 = jv + ku + uv$ belongs to W . Since 1 commutes multiplicatively with w_0 , we can use the binomial theorem to obtain

$$(xy)^n = (jk)^n 1 + n(jk)^{n-1} w_0 + \frac{n(n-1)}{2} (jk)^{n-2} w_0^2$$

for any positive integer n . In particular, if $n(n-1)/2$ is divisible by 5, we have

$$(xy)^n = (jk)^n 1 + n(jk)^{n-1} w_0 = (jk)^n 1 + n(jk)^{n-1} (jv + ku + uv). \quad (7)$$

Subject to the same restriction on n , we have

$$x^n y^n = (j^n 1 + nj^{n-1} u)(k^n 1 + nk^{n-1} v) = (jk)^n 1 + n(jk)^{n-1} (jv + ku + nuv). \quad (8)$$

Since $n(n-1)/2$, and hence $n^2 - n$, was assumed to be divisible by 5, the right sides of (7) and (8) are equal; thus, R_1 satisfies the identity (1) for any n such that $n(n-1)/2$ is divisible by 5. In particular, for $n = 5$ and $n = 6$, identity (1) is satisfied by R_1 .

Despite the existence of examples such as this, we need not give up on the case of two consecutive n ; instead we can impose hypotheses which are incompatible with the "bad" behavior of R_1 . The theorem below is due to Harmanci [7]; the proof is based on the proof of Theorem 1 of [4].

THEOREM 3. Let R be a ring with 1. Suppose that R satisfies (1) for $n = k, k + 1$, and that R contains no nonzero elements x for which $k!x = 0$. Then R is commutative.

The basic strategy of the proof is to study a factor ring $\bar{R} = R/I$, where I is an ideal chosen so that \bar{R} is more tractable than R , and then to transfer information about \bar{R} back to R . In our case, we take I to be the set N of **nilpotent** elements of R , defined by

$$N = \{x \in R \mid x^j = 0 \text{ for some positive integer } j\}.$$

In our arguments, we shall also make use of the **center** C of R , defined by

$$C = \{x \in R \mid xy = yx \text{ for all } y \in R\},$$

and use the fact that the center of any ring is a subring. For arbitrary rings R , the set N need not be an ideal, nor even an additive subgroup; hence, the first step of the proof is to show that the hypotheses of Theorem 3 force the inclusion $N \subseteq C$. This inclusion implies that N is an ideal. Finally, we shall invoke yet another cancellation property, which we call Property C.

PROPERTY C. Let R be a ring with no nonzero nilpotent elements, and let f be a polynomial function in two variables on R such that every monomial term in $f(x, y)$ contains y . Then if R satisfies the identity $f(x, y)y = 0$, it also satisfies the identity $f(x, y) = 0$.

To establish property C, note first that if $ab = 0$, then $(ba)^2 = 0$, hence $ba = 0 = bax$ for every $x \in R$. Repeating the argument now yields $axb = 0$, so we have an **insertion-of-factors property** (IFP): in a ring with no nonzero nilpotent elements, if a product of finitely many elements is 0,

then all products obtained by inserting additional factors in any positions are also 0. (Of course, all commutative rings have IFP, but noncommutative rings in general do not.) Suppose now that R satisfies the identity $f(x, y)y = 0$, where $f(x, y) = \sum_{i=1}^n p_i(x, y)$ for monomials $p_i(x, y)$ having y as a factor. Then, because of IFP, R satisfies each of the identities $f(x, y)p_i(x, y) = 0$, $i = 1, \dots, n$; consequently R satisfies the identity $(f(x, y))^2 = 0$, which in the absence of nilpotent elements implies the identity $f(x, y) = 0$.

Proof of Theorem 3. Let R be any ring satisfying the hypotheses of the Theorem and C its center. Then, as in the proof of Theorem 2, we have $x^k[x, y^k]y = 0$, and applying the Lemma gives

$$[x, y^k]y = 0 \quad \text{for all } x, y \in R. \quad (9)$$

If y is not a zero divisor—in particular, if y is invertible (has a multiplicative inverse)—the obvious cancellation shows that $y^k \in C$. If $u \in N$, $u^n = 0$ implies that $(1+u)(1-u+u^2-\dots+(-1)^{n-1}u^{n-1}) = 1$, so that $1+u$ is invertible, and hence $(1+u)^k \in C$.

For arbitrary $u \in N$, let the **index** of u be the smallest n such that $u^n = 0$. We now proceed, by induction on the index of elements of N , to show that $N \subseteq C$. Expanding $(1+u)^k$ by the binomial theorem, we have

$$1 + ku + v \in C \quad (10)$$

for each $u \in N$, where

$$v = \binom{k}{2}u^2 + \binom{k}{3}u^3 + \dots$$

Thus, if u has index 2, $v = 0$ and $ku \in C$, so that $0 = [ku, x] = k[u, x] = k![u, x]$ for all $x \in R$. But recalling the hypotheses on R , we then get $[u, x] = 0$ for all $x \in R$, which says $u \in C$. Now suppose all nilpotent elements of index less than n are in C , and consider u of index n . It is easily seen that the corresponding v has index less than n , so (10) again yields $ku \in C$ and hence $u \in C$. Our induction is now complete.

Since $N \subseteq C$, the set N forms an ideal. (If $a, b \in N$, then $a^n = b^m = [a, b] = 0$; the fact that $a - b \in N$ follows by expanding $(a - b)^{n+m-1}$ by the binomial theorem and noting that each summand contains either a^n or b^m as a factor.) We consider the factor ring $\bar{R} = R/N$, which inherits all the original hypotheses and in addition has no nonzero nilpotent elements. Suppose, temporarily, that \bar{R} can be shown to be commutative. Then for every $x, y \in R$, $[x, y] = xy - yx \in N$, hence $[x, y] \in C$. It now follows by an easy induction that $[x, y^n] = ny^{n-1}[x, y]$ for all $x, y \in R$ and all positive integers n ; and recalling (9), we have $0 = [x, y^k]y = ky^k[x, y]$. It follows that $k!y^k[x, y] = 0$, and so, by hypothesis, $y^k[x, y] = 0$; hence, by the Lemma, R is commutative.

The proof of Theorem 3 is not yet complete; it is necessary to justify our temporary assumption concerning \bar{R} . We show, in fact, that any R satisfying the hypotheses of Theorem 3 and having $N = \{0\}$ must be commutative. Note first that Property C applied to (9) shows that $x^k \in C$ for all $x \in R$. Thus,

$$(1+x)^k - x^k - 1 = kx + \binom{k}{2}x^2 + \dots + kx^{k-1} \in C,$$

so that

$$\left[kx + \binom{k}{2}x^2 + \dots + kx^{k-1}, y \right] = 0 \quad \text{for all } x, y \in R. \quad (11)$$

Replacing x in (11) by $2x, 3x, \dots, (k-1)x$ in turn, we see that

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ 2 & 2^2 & \dots & 2^{k-1} \\ 3 & 3^2 & \dots & 3^{k-1} \\ \vdots & \vdots & \ddots & \vdots \\ k-1 & (k-1)^2 & \dots & (k-1)^{k-1} \end{bmatrix} \begin{bmatrix} [kx, y] \\ \left[\binom{k}{2}x^2, y \right] \\ \vdots \\ [kx^{k-1}, y] \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (12)$$

for all $x, y \in R$. Now the $(k-1) \times (k-1)$ matrix A on the left side of (12) is a Vandermonde matrix with determinant Δ equal to $\pm \Pi(i-j)$, where the factors $i-j$ range over all pairs with $i, j \in \{2, 3, \dots, k-1\}$ and $i < j$ (see [18, p. 15–16]); and since all these factors divide $k!$, it follows that Δ divides $(k!)^m$ for some positive integer m . Left-multiplying both sides of (12) by the matrix $\text{adj } A$ [15, p. 36] and recalling that $(\text{adj } A)A$ is equal to Δ times the $(k-1) \times (k-1)$ identity matrix, we get

$$\Delta[kx, y] = \Delta\left[\binom{k}{2}x^2, y\right] = \dots = \Delta[kx^{k-1}, y] = 0 \quad \text{for all } x, y \in R.$$

Consequently, $\Delta k[x, y] = 0 = \Delta k![x, y] = (k!)^{m+1}[x, y]$ for all $x, y \in R$. Repeatedly using the hypothesis that $k!z = 0$ implies $z = 0$, we obtain $[x, y] = 0$ for all $x, y \in R$; hence R is commutative.

Extensions and related results

Our choice of theorems has been influenced by a desire to keep the proofs elementary and reasonably self-contained. Not surprisingly, by using more elaborate methods, one can obtain somewhat better results.

An examination of the proof of Theorem 3 shows that only in the final stages did we use the full force of the hypothesis that $k!x = 0$ implies $x = 0$; usually we employed only the weaker hypothesis that $kx = 0$ implies $x = 0$. In fact, the conclusion of Theorem 3 remains true if we assume only the weaker hypothesis [4]. In this case, we omit the last paragraph of the proof of Theorem 3 and show that commutators in R are nilpotent by appealing to a deep theorem of Herstein [3], [8], [9]: if the ring R (not necessarily with 1) satisfies (1) for some $n > 1$, then all commutators in R are nilpotent, and the ideal generated by the commutators consists entirely of nilpotent elements.

Clearly Herstein's result implies that a ring with no nonzero nilpotent elements is commutative if it satisfies (1) for even one $n > 1$. There are other one- n theorems available as well, for example, the following recent result due to Abu-Khuzam [1]:

THEOREM 4. *Let $n > 1$ be a positive integer, and let R be a ring with 1. If R satisfies (1) and contains no nonzero x for which $n(n-1)x = 0$, then R is commutative.*

Incidentally, in this theorem the hypothesis that $n(n-1)x = 0$ implies $x = 0$ cannot be weakened to the hypothesis that $nx = 0$ implies $x = 0$. (Consider the ring R_1 of Example 2 with $n = 21$.)

So far we have always assumed the existence of one, two, or three n such that every pair x, y of elements satisfies (1) for those n . One possibility of generalization is to assume that the n varies with x and y ; and it works—at least sometimes. Indeed, Richoux [21] has recently established commutativity of R with 1 under the hypothesis that for each $x, y \in R$ there exist three consecutive integers n , depending on x and y , for which (1) holds, and his result has been further generalized in [12], [19], and [20].

Other closely-related theorems assert commutativity of rings satisfying (1) together with other identities (see [5, Theorem 2] and [11]). We conclude with a sample of this kind of result.

THEOREM 5. *Let R be a ring with 1, and let n and m be relatively prime integers, greater than or equal to 2. If R satisfies the identities $(xy)^n = x^n y^n$, $(xy)^{n+1} = x^{n+1} y^{n+1}$, and $x^m y^m = y^m x^m$, then R is commutative.*

Supported by the Natural Sciences and Engineering Research Council of Canada, Grant No. A 3961.

This paper had its genesis in a talk given for undergraduates at Mount Holyoke College, South Hadley, Massachusetts, during the spring semester of 1979. The author acknowledges with gratitude the encouragement given by the members of the Mathematics Department at Mount Holyoke.

References

- [1] H. Abu-Khuzam, A commutativity theorem for rings, *Math. Japon.*, 25 (1980) 593–595.
- [2] J. Alperin, A classification of n -abelian groups, *Canad. J. Math.*, 21 (1969) 1238–1244.
- [3] H. E. Bell, On a commutativity theorem of Herstein, *Arch. Math.*, 21 (1970) 265–267.
- [4] ———, On the power map and ring commutativity, *Canad. Math. Bull.*, 21 (1978) 399–404.
- [5] ———, On rings with commuting powers, *Math. Japon.*, 24 (1979) 473–478.
- [6] L. O. Chung and J. Luh, Conditions for elements to be central in a ring, *Acta Math. Acad. Sci. Hungar.*, 34 (1979) 261–265.
- [7] A. Harmanci, Two elementary commutativity theorems for rings, *Acta Math. Acad. Sci. Hungar.*, 29 (1977) 23–29.
- [8] I. N. Herstein, Power maps in rings, *Michigan Math. J.*, 8 (1961) 29–32.
- [9] ———, A remark on rings and algebras, *Michigan Math. J.*, 10 (1963) 269–272.
- [10] ———, *Topics in Algebra*, Xerox College Publishing, Lexington, 1975.
- [11] Y. Hirano, M. Hongan, and H. Tominaga, Supplements to the previous paper “Some commutativity theorems for rings,” to appear.
- [12] M. Hongan and I. Mogami, A commutativity theorem for rings, *Math. Japon.*, 23 (1978) 131–132.
- [13] E. C. Johnsen, D. L. Outcalt, and A. Yaqub, An elementary commutativity theorem for rings, *Amer. Math. Monthly*, 75 (1968) 288–289.
- [14] A. Kaya, On a commutativity theorem of Luh, *Acta Math. Acad. Sci. Hungar.*, 28 (1976) 33–36.
- [15] P. Lancaster, *Theory of Matrices*, Academic Press, New York, 1969.
- [16] S. Ligh and A. Richoux, A commutativity theorem for rings, *Bull. Austral. Math. Soc.*, 16 (1977) 75–77.
- [17] J. Luh, A commutativity theorem for primary rings, *Acta Math. Acad. Sci. Hungar.*, 22 (1971) 211–213.
- [18] M. Marcus and H. Minc, *A Survey of Matrix Theory and Matrix Inequalities*, Allyn & Bacon, Boston, 1964.
- [19] I. Mogami and M. Hongan, Note on commutativity of rings, *Math. J. Okayama Univ.*, 20 (1978) 21–24.
- [20] I. Mogami, Note on commutativity of rings, II, *Math. J. Okayama Univ.*, 22 (1980) 51–54.
- [21] A. Richoux, On a commutativity theorem of Luh, *Acta Math. Acad. Sci. Hungar.*, 34 (1979) 23–25.
- [22] H. Trotter, Groups in which raising to a power is an automorphism, *Canad. Math. Bull.*, 8 (1965) 825–827.

Races with Ties

ELLIOTT MENDELSON

Queens College

Flushing, NY 11367

When there are n runners in a race, the number of possible outcomes is $n!$ if we assume that there are no ties. If any number of the runners are allowed to tie for arbitrarily many positions, calculation of the number J_n of outcomes becomes much more complicated. The number J_n has other interesting interpretations. It is the number of possible election ballots when there are n candidates and the voters are allowed to express equal preference among some of the candidates. It is also the number of preferential arrangements of n objects, allowing indifference among some of the objects.

The first few values of J_n are easy to calculate: $J_0 = 1$ and $J_1 = 1$, while $J_2 = 3$ (either (A, B) , (B, A) or a tie (AB)). When $n = 3$, we have the six standard permutations of A, B, C , plus (ABC) (all tied for first), (AB, C) , (AC, B) , (BC, A) (two tied for first), and (C, AB) , (B, AC) , (A, BC) (two tied for second). Thus, $J_3 = 13$. We shall derive recursion equations for J_n , several closed forms for J_n , and some other methods for calculating J_n .

Assume that there are $n + 1$ runners. If the number of runners who do not finish first is j , then those j runners can finish in 2nd, 3rd, ... places in J_j ways. Moreover, those j runners can be chosen from the $n + 1$ runners in $\binom{n+1}{j}$ ways. Hence, the number of possible outcomes is $\binom{n+1}{j} J_j$. Since j can be any number between 0 and n , the value of J_{n+1} is

References

- [1] H. Abu-Khuzam, A commutativity theorem for rings, *Math. Japon.*, 25 (1980) 593–595.
- [2] J. Alperin, A classification of n -abelian groups, *Canad. J. Math.*, 21 (1969) 1238–1244.
- [3] H. E. Bell, On a commutativity theorem of Herstein, *Arch. Math.*, 21 (1970) 265–267.
- [4] ———, On the power map and ring commutativity, *Canad. Math. Bull.*, 21 (1978) 399–404.
- [5] ———, On rings with commuting powers, *Math. Japon.*, 24 (1979) 473–478.
- [6] L. O. Chung and J. Luh, Conditions for elements to be central in a ring, *Acta Math. Acad. Sci. Hungar.*, 34 (1979) 261–265.
- [7] A. Harmanci, Two elementary commutativity theorems for rings, *Acta Math. Acad. Sci. Hungar.*, 29 (1977) 23–29.
- [8] I. N. Herstein, Power maps in rings, *Michigan Math. J.*, 8 (1961) 29–32.
- [9] ———, A remark on rings and algebras, *Michigan Math. J.*, 10 (1963) 269–272.
- [10] ———, *Topics in Algebra*, Xerox College Publishing, Lexington, 1975.
- [11] Y. Hirano, M. Hongan, and H. Tominaga, Supplements to the previous paper “Some commutativity theorems for rings,” to appear.
- [12] M. Hongan and I. Mogami, A commutativity theorem for rings, *Math. Japon.*, 23 (1978) 131–132.
- [13] E. C. Johnsen, D. L. Outcalt, and A. Yaqub, An elementary commutativity theorem for rings, *Amer. Math. Monthly*, 75 (1968) 288–289.
- [14] A. Kaya, On a commutativity theorem of Luh, *Acta Math. Acad. Sci. Hungar.*, 28 (1976) 33–36.
- [15] P. Lancaster, *Theory of Matrices*, Academic Press, New York, 1969.
- [16] S. Ligh and A. Richoux, A commutativity theorem for rings, *Bull. Austral. Math. Soc.*, 16 (1977) 75–77.
- [17] J. Luh, A commutativity theorem for primary rings, *Acta Math. Acad. Sci. Hungar.*, 22 (1971) 211–213.
- [18] M. Marcus and H. Minc, *A Survey of Matrix Theory and Matrix Inequalities*, Allyn & Bacon, Boston, 1964.
- [19] I. Mogami and M. Hongan, Note on commutativity of rings, *Math. J. Okayama Univ.*, 20 (1978) 21–24.
- [20] I. Mogami, Note on commutativity of rings, II, *Math. J. Okayama Univ.*, 22 (1980) 51–54.
- [21] A. Richoux, On a commutativity theorem of Luh, *Acta Math. Acad. Sci. Hungar.*, 34 (1979) 23–25.
- [22] H. Trotter, Groups in which raising to a power is an automorphism, *Canad. Math. Bull.*, 8 (1965) 825–827.

Races with Ties

ELLIOTT MENDELSON

Queens College

Flushing, NY 11367

When there are n runners in a race, the number of possible outcomes is $n!$ if we assume that there are no ties. If any number of the runners are allowed to tie for arbitrarily many positions, calculation of the number J_n of outcomes becomes much more complicated. The number J_n has other interesting interpretations. It is the number of possible election ballots when there are n candidates and the voters are allowed to express equal preference among some of the candidates. It is also the number of preferential arrangements of n objects, allowing indifference among some of the objects.

The first few values of J_n are easy to calculate: $J_0 = 1$ and $J_1 = 1$, while $J_2 = 3$ (either (A, B) , (B, A) or a tie (AB)). When $n = 3$, we have the six standard permutations of A, B, C , plus (ABC) (all tied for first), (AB, C) , (AC, B) , (BC, A) (two tied for first), and (C, AB) , (B, AC) , (A, BC) (two tied for second). Thus, $J_3 = 13$. We shall derive recursion equations for J_n , several closed forms for J_n , and some other methods for calculating J_n .

Assume that there are $n + 1$ runners. If the number of runners who do not finish first is j , then those j runners can finish in 2nd, 3rd, ... places in J_j ways. Moreover, those j runners can be chosen from the $n + 1$ runners in $\binom{n+1}{j}$ ways. Hence, the number of possible outcomes is $\binom{n+1}{j} J_j$. Since j can be any number between 0 and n , the value of J_{n+1} is

$$\binom{n+1}{0}J_0 + \binom{n+1}{1}J_1 + \dots + \binom{n+1}{n}J_n.$$

Thus, we obtain the recursion equations:

$$J_0 = 1$$

$$J_{n+1} = \sum_{j=0}^n \binom{n+1}{j} J_j. \quad (1)$$

Using these equations, we can calculate further values of J_n . For example,

$$J_4 = \sum_{j=0}^3 \binom{4}{j} J_j = J_0 + 4J_1 + 6J_2 + 4J_3 = 1 + 4 + 18 + 52 = 75.$$

TABLE 1 lists all values of J_n for $1 \leq n \leq 12$. Notice that the last digits occur in cycles of 1, 3, 3, 5. That this is always so has been proved by Gross [3], who showed that, for $n \geq 1$, $J_{n+4} - J_n$ is divisible by 10.

n	J_n
1	1
2	3
3	13
4	75
5	541
6	4, 683
7	47, 293
8	545, 835
9	7, 087, 261
10	102, 247, 563
11	1, 622, 632, 573
12	28, 091, 567, 595



TABLE 1. Values for J_n , $1 \leq n \leq 12$.

The sequence of values J_n , and their recursion equations, first appeared in an entirely different context in Cayley [2], where J_n was introduced as the number of different trees of a certain type. J_n reappeared, again in a graph-theoretic context, in MacMahon [5]. The first explicitly combinatorial treatment of J_n , in terms of preferential arrangements, seems to have appeared in Gross [3], although MacMahon was aware of their combinatorial significance. (He stated in [5] that the sequence of J_n 's is "identical with that of the compositions of certain multipartite numbers.")

The recursion equations (1) for J_n simplify its computation, but, in order to obtain more information about J_n and possibly also a more efficient algorithm for its computation, we should like to find an explicit formula for J_n . Let J_{nk} be the number of possible outcomes of a race among n runners in which only the first k places are occupied (i.e., the runner(s) who finish last, place k th). For example, $J_{31} = 1$, $J_{32} = 6$, and $J_{33} = 3! = 6$. Clearly, $J_n = \sum_{k=1}^n J_{nk}$. So, if we can find a formula for J_{nk} , a formula for J_n will result.

Let $W_k = \{1, 2, \dots, k\}$, and let \mathfrak{M}_{nk} be the set of all mappings from W_n onto W_k . Then J_{nk} is the number of mappings in \mathfrak{M}_{nk} . For, if we let a_1, a_2, \dots, a_n be the runners, then a particular outcome in which only the first k places are occupied corresponds to the mapping f from W_n onto W_k such that $f(i) = j$ if a_i finishes in the j th place. Thus, our problem reduces to finding the cardinality of \mathfrak{M}_{nk} . To do this, we shall need the following well-known result.

INCLUSION-EXCLUSION PRINCIPLE. If X_1, \dots, X_k are sets, then

$$|X_1 \cup \dots \cup X_k| = \sum_{i=1}^k |X_i| - \sum_{1 \leq i < j \leq k} |X_i X_j| + \sum_{1 \leq i < j < p \leq k} |X_i X_j X_p| - \dots + (-1)^{k+1} |X_1 X_2 \dots X_k|.$$

(Here, $|A|$ denotes the number of elements in set A , and AB designates the intersection of sets A and B .)

Proof. Assume y occurs in $X_1 \cup \dots \cup X_k$ and y is contained in exactly m of the sets X_1, \dots, X_k . Then, on the right side of the equation, y is added m times in the first sum, subtracted $\binom{m}{2}$ times in the second sum, added $\binom{m}{3}$ times in the third sum, etc. Hence, the number of times y is counted on the right is

$$\binom{m}{1} - \binom{m}{2} + \binom{m}{3} - \dots + (-1)^{m+1} \binom{m}{m}.$$

But, by the Binomial Theorem,

$$\begin{aligned} 0 &= (1-1)^m = 1 - \binom{m}{1} + \binom{m}{2} - \binom{m}{3} + \dots + (-1)^m \binom{m}{m} \\ &= 1 - \left[\binom{m}{1} - \binom{m}{2} + \binom{m}{3} - \dots + (-1)^{m+1} \binom{m}{m} \right]. \end{aligned}$$

Thus, y is counted exactly once on the right side.

For our purposes, we use the Inclusion-Exclusion Principle in the following manner. Let \mathcal{F}_{nk} be the set of all mappings from W_n into W_k . The cardinality of \mathcal{F}_{nk} is k^n . For $1 \leq i \leq k$, let X_i be the set of all mappings f in \mathcal{F}_{nk} such that i is not in the range of f . Now, $|X_i| = (k-1)^n$, $|X_i X_j| = (k-2)^n$ (if $i < j$), $|X_i X_j X_p| = (k-3)^n$ (if $i < j < p$), etc. Hence, by the Inclusion-Exclusion Principle,

$$\begin{aligned} |X_1 \cup X_2 \cup \dots \cup X_k| &= k(k-1)^n - \binom{k}{2}(k-2)^n + \binom{k}{3}(k-3)^n - \dots + (-1)^{k+1} \binom{k}{k}(k-k)^n \\ &= \sum_{j=1}^{k-1} (-1)^{j+1} \binom{k}{j} (k-j)^n = \sum_{i=1}^{k-1} (-1)^{k-i+1} \binom{k}{k-i} i^n \\ &= \sum_{i=1}^{k-1} (-1)^{k-i+1} \binom{k}{i} i^n. \end{aligned}$$

But, $X_1 \cup X_2 \cup \dots \cup X_k$ is the complement of \mathcal{N}_{nk} in \mathcal{F}_{nk} . Hence, $|X_1 \cup \dots \cup X_k| = k^n - |\mathcal{N}_{nk}|$. So,

$$\begin{aligned} |\mathcal{N}_{nk}| &= k^n - \sum_{i=1}^{k-1} (-1)^{k-i+1} \binom{k}{i} i^n = k^n + \sum_{i=1}^{k-1} (-1)^{k-i} \binom{k}{i} i^n \\ &= \sum_{i=1}^k (-1)^{k-i} \binom{k}{i} i^n. \end{aligned}$$

Since $J_{nk} = |\mathcal{N}_{nk}|$,

$$J_n = \sum_{k=1}^n \sum_{i=1}^k (-1)^{k-i} \binom{k}{i} i^n. \quad (2)$$

The argument employed here to count \mathcal{N}_{nk} is also found in Aigner [1], p. 164, Exercise 5, where it is used to find a formula for the so-called Stirling numbers of the second kind. A variant of formula (2) is obtained in a different way in Gross [3].

To see how (2) works, we compute J_4 .

$$J_4 = 1 + (-2 + 2^4) + (3 - 3 \cdot 2^4 + 3^4) + (-4 + 6 \cdot 2^4 - 4 \cdot 3^4 + 4^4) \\ = 1 + 14 + 36 + 24 = 75.$$

Another way to find an explicit formula for J_n is to use a generating function. This approach is found in much of the literature on the J_n 's (e.g., [2], [6]). The function

$$f(x) = \sum_{n=0}^{\infty} \frac{J_n}{n!} x^n \quad (3)$$

is called the exponential generating function for the J_n 's. Then

$$e^x f(x) = \left(\sum_{n=0}^{\infty} \frac{x^n}{n!} \right) \left(\sum_{n=0}^{\infty} \frac{J_n}{n!} x^n \right) = \sum_{n=0}^{\infty} c_n x^n.$$

Clearly, $c_0 = 1$, and, for $n \geq 1$,

$$c_n = \sum_{k=0}^n \frac{1}{(n-k)!} \frac{J_k}{k!} = \frac{1}{n!} \sum_{k=0}^n \frac{n!}{(n-k)!k!} J_k \\ = \frac{1}{n!} \sum_{k=0}^n \binom{n}{k} J_k = \frac{1}{n!} \left(\sum_{k=0}^{n-1} \binom{n}{k} J_k + J_n \right) \\ = \frac{1}{n!} (J_n + J_n) = \frac{2}{n!} J_n.$$

Hence,

$$e^x f(x) = 1 + 2 \sum_{n=1}^{\infty} \frac{J_n}{n!} x^n = 1 + 2(f(x) - 1) = 2f(x) - 1.$$

So,

$$f(x) = \frac{1}{2 - e^x}. \quad (4)$$

Equating (3) and (4), n -fold differentiation yields $J_n = f^{(n)}(0)$. Thus, a formula for $f^{(n)}(x)$ will produce a formula for J_n . Differentiating (4) gives

$$f'(x) = - \left[\frac{1}{2 - e^x} - \frac{2}{(2 - e^x)^2} \right].$$

Further differentiation suggests the following formula, which can be proved by means of a tedious induction:

$$f^{(n)}(x) = (-1)^n \sum_{j=1}^{n+1} \frac{c_j}{(2 - e^x)^j}, \quad (5)$$

where

$$c_j = (-1)^{j-1} 2^{j-1} \sum_{k=0}^{j-1} (-1)^k \binom{j-1}{k} (j-k)^n.$$

Hence,

$$J_n = f^{(n)}(0) = (-1)^n \sum_{j=1}^{n+1} (-1)^{j-1} 2^{j-1} \sum_{k=0}^{j-1} (-1)^k \binom{j-1}{k} (j-k)^n \\ = (-1)^n \sum_{i=0}^n (-1)^i 2^i \sum_{k=0}^i (-1)^k \binom{i}{k} (i+1-k)^n$$

							$\underline{J_n}$
							1
							3
							13
							75
							541
							4,683
1	126	1,806	8,400	16,800	15,120	5,040	47,293

TABLE 2. A Pascal-like triangle for computing J_n . The first entry in each row is 1, and the last entry in the n th row is $n!$. For $1 < k < n$, the k th entry in the n th row is J_{nk} , computed from the two entries above it, using (7). J_n is the sum of the entries in the n th row.

$$\begin{aligned}
 &= (-1)^n \sum_{i=0}^n (-1)^i 2^i \sum_{m=0}^i (-1)^{i-m} \binom{i}{i-m} (m+1)^n \\
 &= (-1)^n \sum_{i=0}^n 2^i \sum_{m=0}^i (-1)^m \binom{i}{m} (m+1)^n.
 \end{aligned} \tag{6}$$

We apply formula (6) to compute J_4 :

$$\begin{aligned}
 J_4 &= 1 + 2(1 - 2^4) + 4(1 - 2 \cdot 2^4 + 3^4) + 8(1 - 3 \cdot 2^4 + 3 \cdot 3^4 - 4^4) \\
 &\quad - 16(1 - 4 \cdot 2^4 + 6 \cdot 3^4 - 4 \cdot 4^4 + 5^4) \\
 &= 1 - 30 + 200 - 480 + 384 = 75.
 \end{aligned}$$

In contrast to the rather formidable formula (6), we exhibit two simple diagrammatic algorithms for computing J_n . Designate the runners in a race by A_1, A_2, \dots, A_n and consider the position of A_n in any outcome of a race with k positions. Either (i) A_n occupies a position alone or (ii) A_n is in a tie for one of the k positions. In case (i), the outcome can be thought of as arising from one of the $J_{n-1, k-1}$ outcomes of a race among A_1, A_2, \dots, A_{n-1} with $k-1$ positions. But there are k ways of inserting A_n into one such outcome in a position all by himself. Hence, case (i) can arise in $k \cdot J_{n-1, k-1}$ ways. In case (ii), the outcome can be thought of as arising from one of the $J_{n-1, k}$ outcomes of a race among A_1, A_2, \dots, A_{n-1} with k positions. Since A_n can be added to the runners in any one of the k positions, case (ii) occurs in $k \cdot J_{n-1, k}$ ways. Thus,

$$J_{nk} = k \cdot J_{n-1, k-1} + k \cdot J_{n-1, k} = k(J_{n-1, k-1} + J_{n-1, k}). \tag{7}$$

If we notice that $J_{n1} = 1$ and $J_{nn} = n!$, we can construct an analogue of Pascal's triangle as shown in TABLE 2. The upper right-hand side of the triangle is made up of the values of $n!$. For $1 < k < n$, the k th entry in the n th row is k times the sum of the two entries immediately above it in the $(n-1)$ th row. (For example, the third entry in the fourth row is 36, where $36 = 3(6 + 6)$.) Thus, by (7) and induction, the k th entry in the n th row is J_{nk} . The sum of the entries in the n th row is $\sum_{k=1}^n J_{nk} = J_n$, and this provides a quick tabular method for calculating the successive values of J_n .

TABLE 3 shows another Pascal-like triangle connected with J_n . Here, the upper right-hand side of the triangle consists of pairs of the same positive integer in ascending order. As in Pascal's triangle, each interior entry is the sum of the two entries immediately above it. The values of J_n can be computed in the following way. Let c_{nk} denote the k th entry in the n th row; then

					1															
				1		1														
			1		2		2													
		1		3		4		2												
			1	4		7		6	3											
				1	5		11		13	9	3									
					1	6		16		24	22	12	4							
							1	7		22		40		46		34		16		4

TABLE 3. Another Pascal-like triangle which can be used to compute J_n . The entries in the n th row are the coefficients $c_{n,j}$ in formula (8) for J_n .

$$\begin{aligned}
 J_n &= c_{n1}n^n - c_{n2}(n-1)^n + c_{n3}(n-2)^n + \cdots + (-1)^{n+1}c_{nn} \\
 &= \sum_{j=0}^{n-1} (-1)^j c_{n,j+1} (n-j)^n.
 \end{aligned} \tag{8}$$

For example, $J_3 = (1)3^3 - (2)2^3 + (2)1^3 = 27 - 16 + 2 = 13$, and $J_4 = (1)4^4 - (3)3^4 + (4)2^4 - (2)1^4 = 256 - 243 + 64 - 2 = 75$.

This computational procedure can be justified by transforming equation (2) as follows:

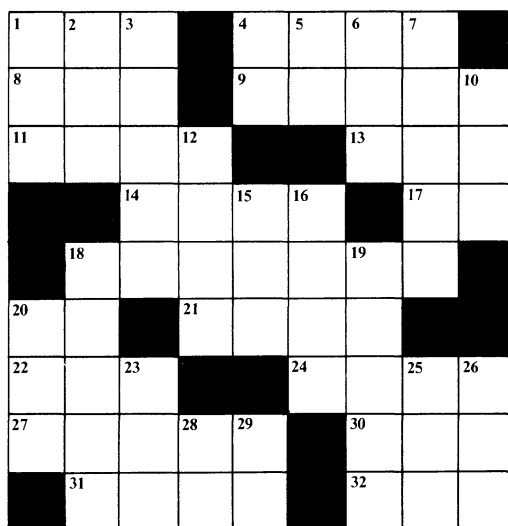
$$\begin{aligned}
 J_n &= \sum_{k=1}^n \sum_{i=1}^k (-1)^{k-i} \binom{k}{i} i^n = \sum_{i=1}^n (-1)^i \left[\sum_{k=i}^n (-1)^k \binom{k}{i} \right] i^n \\
 &= \sum_{j=0}^{n-1} (-1)^{n-j} \left[\sum_{k=n-j}^n (-1)^k \binom{k}{n-j} \right] (n-j)^n \\
 &= \sum_{j=0}^{n-1} \left[\sum_{p=0}^j (-1)^p \binom{p+n-j}{p} \right] (n-j)^n.
 \end{aligned}$$

Finally, an induction argument with respect to j shows that $\sum_{p=0}^j (-1)^p \binom{p+n-j}{p}$ represents $(-1)^j$ times the entry $c_{n,j+1}$ in the triangle in TABLE 3.

References

- [1] M. Aigner, *Combinatorial Theory*, Springer-Verlag, New York, 1979.
- [2] A. Cayley, On the analytic forms called trees, second part, *Philos. Mag.*, 18 (1859) 374–378. (Reprinted in *Collected Mathematical Papers of Arthur Cayley*, Cambridge, 1891, pp. 112–115.)
- [3] O. A. Gross, Preferential arrangements, *Amer. Math. Monthly*, 69 (1962) 4–8.
- [4] J. G. Kemeny, Mathematics without numbers, *Daedalus*, 88 (1959) 577–591. (For a more detailed presentation, see *Mathematical Models in the Social Sciences* by J. G. Kemeny and J. L. Snell, MIT Press, 1972, Chapter 2.)
- [5] P. A. MacMahon, Yoke-chains and multipartite compositions in connexion with the analytical forms called “trees”, *Proc. London Math. Soc.*, 22 (1891) 330–346. (Reprinted in *Collected Papers*, vol. 1, Combinatorics, MIT Press, 1978.)
- [6] T. S. Motzkin, Sorting numbers for cylinders and other classification numbers, *Proc. of Symposia in Pure Math.*, vol. 19, Combinatorics, Amer. Math. Soc., 1971, pp. 167–176.
- [7] N. J. A. Sloane, *A Handbook of Integer Sequences*, Academic Press, 1973 (Sequence 1191).

A Cross-Number Puzzle



ACROSS

1. A perfect square
4. A palindromic integer
8. This many degrees Fahrenheit is 265 degrees Centigrade
9. With 3-down, a permutation of the digits 0 through 9
11. Second smallest prime of the form $n^2 + 2^n + 1$, $n > 0$
13. A gross number
14. 19-down minus 18-down plus 26-down
17. Number of 2-digit primes
18. The product of the digits of this number is 78,125
20. The sum of this number's positive divisors is 91
21. $33^2 + 3^2$
22. This number is the sum of the factorials of its digits
24. A power of 6
27. The sum of the 4th powers of 5 consecutive triangular numbers
30. A Mersenne prime
31. A power of 2
32. The number of the beast

DOWN

1. A multiple of 11
2. The product of the positive divisors of this number is $(202)^2$
3. See 9-across
4. A Fermat prime
5. Product of the first 3 primes
6. Colleague of 1-across
7. In base 2, this number is written 11010001110001
10. Yet another perfect square!
12. The first prime year after 1950
15. This many degrees is $25\pi/6$ radians
16. The 17th Fibonacci number
18. $210^2 + 111^2$
19. The least common multiple of 36 and 1631
20. The number of positive perfect squares less than 10^5
23. The number of positive integers less than 625 which are not divisible by 5
25. The sum of these digits is 15, and their product is 84
26. Palindromic square
28. The only even prime number
29. 20-across minus 28-down

—NICK FRANCESCHINE III
550 Dufranc Avenue
Sebastopol, CA 95472

Solution on p. 187

PROBLEMS

LEROY F. MEYERS, Editor
G. A. EDGAR, Associate Editor

The Ohio State University

Proposals

To be considered for publication, solutions should be mailed before October 1, 1982.

1144. Define the numbers $A_{r,s}^n$ by $A_{0,0}^n = 1$, $A_{r,-1}^n = A_{r,s}^{n-1} = A_{r,s}^{-1} = 0$, and

$$A_{r,s}^n = A_{r-1,s}^n + (n+r-s+1)A_{r,s-1}^n \text{ for } r+s > 0 \text{ and } n \geq 0.$$

(a) Show that $A_{r,s}^n = A_{r-1,s}^{n+1} + nA_{r,s-1}^{n-1}$ for $r+s > 0$ and $n \geq 0$.

(b) Show that $A_{r,s}^n = A_{r,s}^{n+1} - (r+s)A_{r,s-1}^n$ for $r, s, n \geq 0$.

(c) Find an explicit formula for $A_{r,s}^n$.

[H. L. Krall, *State College, Pennsylvania*.]

1145. Let $f(x) = ax^2 + bx + c$ be a quadratic polynomial with integral coefficients, where $a \neq 0$. Show that:

(a) if $f(x)$ is factorable into linear factors with integral coefficients, then there are integers d and e such that $d + e = b$ and $de = ac$; and

(b) if the integers d and e satisfy $d + e = b$ and $de = ac$, then

$$f(x) = \frac{ax+d}{(a,d)} \cdot \frac{ax+e}{a/(a,d)},$$

where each of the linear factors has integral coefficients. [Kenneth A. Brown, Jr., *Nova High School, Fort Lauderdale, Florida*.]

1146. Let $f(x) = 2^x$ and $g(x) = 3^x$ for all real x , and indicate iteration by superscripts. It is easy to check that $f(1) < g(1) < f^2(1) < f^3(1) < g^2(1) < f^4(1) < g^3(1) < f^5(1) < g^4(1)$. Is it true that $f^n(1) < g^{n-1}(1) < f^{n+1}(1)$ for all $n \geq 3$? [James Propp, *undergraduate, Harvard College*.]

1147. Marion Walter, in *Exploring a Rectangle Problem*, this MAGAZINE, 54 (1981) 131-134, discussed variations on the problem *If O is in the interior of the rectangle $ABCD$ and $|OA| = a$, $|OB| = b$, and $|OC| = c$, what is $|OD|$?* and suggested that it leads to other questions. Here is one such question. For a given triple (a, b, c) , what is the maximum area of the rectangle $ABCD$? [James S. Robertson, *Rochester, Minnesota*.]

ASSISTANT EDITORS: DANIEL B. SHAPIRO and WILLIAM A. MCWORTER, JR., *The Ohio State University*. We invite readers to submit problems believed to be new. Proposals should be accompanied by solutions, when available, and by any information that will assist the editors. Solutions to published problems should be submitted on separate, signed sheets. An asterisk (*) will be placed by a problem number to indicate that the proposer did not supply a solution. A problem submitted as a Quickie should be one that has an unexpected succinct solution. Readers desiring acknowledgment of their communications should include a self-addressed stamped card. Send all communications to this department to Leroy F. Meyers, *The Ohio State University, 231 W. 18th Ave., Columbus, Ohio 43210*.

1148. Tom Trotter from Toronto was a guest of his friend Paul Porter of Peoria at a Fourth-of-July bicentennial celebration. The following conversation took place:

Paul: When will you be back here again?

Tom: Not this year, but I'll be back before the Fourth of July eight years from now, and I'll look you up on the fourth day of the month.

Paul: If you tell me the day of the week for that day, and the year, will I be able to figure out the month when you'll be here?

Tom: No, but if I tell you just the day of the week for the thirtieth of that same month, you'll be able to figure it all out, even if I don't tell you the year!

Paul: Right you are! I'll have my swimming pool open.

When will Tom be visiting Paul? [*John M. H. Olmsted, Southern Illinois University.*]

Quickies

Solutions to Quickies appear at the conclusion of the Problems section.

Q672. A is a square matrix with complex number elements; p and q are different numbers; x and y are column vectors such that $y^T(A - pI)^s = 0$ and $(A - qI)^k x = 0$ for some integers s and k ; prove that $y^T x = 0$. [*H. Kestelman, University College London.*]

Q673. A man rowing upstream passes a log after a miles, then continues for b hours, and then rows downstream, meeting the log at his starting point. What is the rate of the stream? [*Charles F. Pinzka, University of Cincinnati.*]

Solutions

Redundancy is not always better

March 1981

1116. In a Hamming $(2^r - 1, 2^r - r - 1)$ single error correcting code, $2^r - r - 1$ information bits and r redundancy bits are transmitted. If at most one of the transmitted bits is in error, the receiver is able to correct the error and obtain the correct message. (If no redundancy bits were used, all $2^r - r - 1$ information bits would have to be correct in order to receive the correct message.) If p is the probability of error in each bit, for what values of p is the probability of the receiver obtaining the correct message greater with redundancy than without? [*G. A. Heuer, Concordia College and Tom Stratton, student, Massachusetts Institute of Technology.*]

Solution: Let $q = 1 - p$. Then for $r \geq 2$ the probability of obtaining the correct message with redundancy is

$$q^{2^r-1} + (2^r - 1)q^{2^r-2}(1 - q),$$

and that with the nonredundant system is

$$q^{2^r-r-1}.$$

If $p = 0$ or 1 , then both are equal. If $0 < p < 1$, then the redundant system is better than the

nonredundant system

$$\begin{aligned} \text{iff} \quad & q^r + (2^r - 1)q^{r-1}(1 - q) > 1 \\ \text{iff} \quad & (2^r - 2)q^r - (2^r - 1)q^{r-1} + 1 < 0 \\ \text{iff} \quad & (q - 1)(2q - 1) \sum_{i=1}^{r-1} (2^i - 1)q^{i-1} < 0. \end{aligned}$$

Since the first factor is strictly negative and the third factor is strictly positive, we need to have $2q - 1 > 0$. If $q = \frac{1}{2}$, then both the systems are equal. Hence the redundant system is better than (exactly as good as; worse than) the nonredundant system if and only if $0 < p < \frac{1}{2}$ ($p = 0, \frac{1}{2}, 1$; $\frac{1}{2} < p < 1$, respectively).

K. ASWATH RAO
George Washington University

Also solved by Steven B. Berger, Milton P. Eisner, Lee O. Hagglund, Peter M. Makus, Roger B. Nelsen, Scott Smith, and the proposers (two solutions).

Balls, Strikes, and Fouls

March 1981

1117. Suppose that every pitch in a baseball game has fixed probabilities p, q, r , respectively, of resulting in a ball, a strike that is not a foul, or a foul. (The probability for all other events is then $1 - p - q - r$; these constitute the events that end a player's time at bat in any way other than a walk or a strikeout.) Let P_K, P_B, P_E , respectively, denote the probabilities that a given batter strikes out, gets a walk, or does anything else. Find closed-form formulas for P_K, P_B , and P_E in terms of p, q, r . [*David A. Smith, Duke University.*]

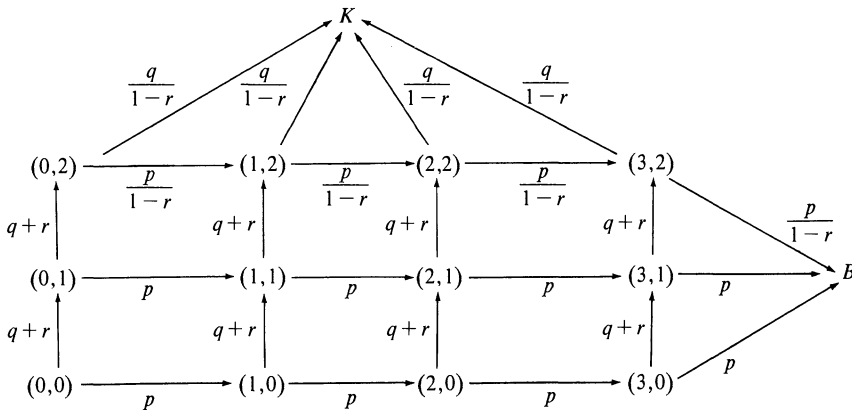
Solution: For any count $(x, y) = (\text{balls}, \text{strikes})$, the probability of “all other events,” i.e., a fair batted ball or a caught foul, is $s = 1 - p - q - r$. (We assume that r is the probability of an uncaught foul.) For count (x, y) , $y = 0$ or 1 , the count moves to

$(x + 1, y)$ with probability p , and
 $(x, y + 1)$ with probability $q + r$.

However, for count $(x, 2)$, the count moves to

$(x+1, 2)$ with probability $p(1+r+r^2+\dots)=p/(1-r)$, and
 $(x, 3) \in K$ with probability $q(1+r+r^2+\dots)=q/(1-r)$.

The diagram shows the conditional probabilities.



By enumerating possible paths through the diagram, we obtain:

$$P_B = p^4 + 4(q+r)p^4 + (q+r)^2 p^4 \sum_{n=1}^4 n(1-r)^{n-5},$$

$$P_K = q(q+r)^2 \sum_{m=1}^4 \sum_{n=1}^m np^{m-1}(1-r)^{n-m-1}, \text{ and}$$

$$P_E = 1 - P_B - P_K.$$

MILTON P. EISNER
Mount Vernon College

Also solved by Scott Smith and the proposer. There were two incorrect solutions.

A New Prime Conjecture

March 1981

1118*. Suppose P is a nonempty set of prime numbers such that for all p and q in P , all the prime divisors of $pq + 1$ are in P . Is P the set of all primes? [F. David Hammer, University of California, Davis.]

Editor's comment: No correct solutions, and two incorrect solutions, were received. J. L. Selfridge remarked, "It is easy to see that all primes up to 73 are in P , and chances seem excellent that $kp' = pq + 1$ will have enough solutions k, p, q so that a given p' cannot hold out indefinitely. But surely no one can prove it." On the other hand, Peter Ørno reported that if P_n is defined to be the smallest set containing the n th prime p_n and containing all prime divisors of $p_n q + 1$ whenever it contains q , then P_n is finite for $n \leq 22$ ($p_n \leq 79$).

Centroid and Circumcircle

March 1981

1119. Let the triangle ABC be inscribed in a circle and let point P be the centroid of the triangle. The line segments AP , BP , and CP are extended to meet the circle in points D , E , and F , respectively. Prove that

$$\frac{|AP|}{|PD|} + \frac{|BP|}{|PE|} + \frac{|CP|}{|PF|} = 3.$$

[K. R. S. Sastry, Addis Ababa, Ethiopia.]

Solution: Let M be the midpoint of side BC of $\triangle ABC$. (See figure on next page.) Since $\triangle ABP$ and $\triangle BPD$ have the same altitude, we have

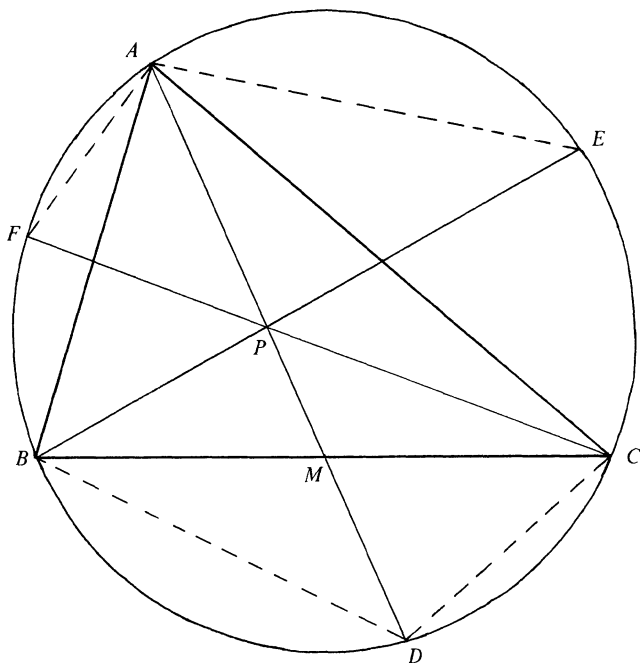
$$\frac{|AP|}{|PD|} = \frac{S_{ABP}}{S_{BPD}} = \frac{\frac{1}{3}S_{ABC}}{S_{BPD}},$$

where S_{ABC} denotes the area of $\triangle ABC$, etc. Similarly,

$$\frac{|BP|}{|PE|} = \frac{\frac{1}{3}S_{ABC}}{S_{APE}} \text{ and } \frac{|CP|}{|PF|} = \frac{\frac{1}{3}S_{ABC}}{S_{APF}}.$$

Since $\triangle APE \sim \triangle BPD$ and $\triangle APF \sim \triangle CPD$, we have

$$\frac{S_{APE}}{S_{BPD}} = \frac{|AP|^2}{|BP|^2} \text{ and } \frac{S_{APF}}{S_{BPD}} = \frac{S_{APF}}{S_{CPD}} = \frac{|AP|^2}{|CP|^2}.$$



Hence

$$\begin{aligned} \frac{|AP|}{|PD|} + \frac{|BP|}{|PE|} + \frac{|CP|}{|PF|} &= \frac{1}{3} \left(\frac{S_{ABC}}{S_{BPD}} + \frac{S_{ABC}}{S_{APE}} + \frac{S_{ABC}}{S_{APF}} \right) \\ &= \frac{1}{3} \frac{S_{ABC}}{S_{BPD}} \left(1 + \frac{|BP|^2}{|AP|^2} + \frac{|CP|^2}{|AP|^2} \right) \\ &= \frac{1}{3} \frac{S_{ABC}}{S_{BPD}} \left(\frac{|AP|^2 + |BP|^2 + |CP|^2}{|AP|^2} \right). \end{aligned}$$

But since $\triangle BMD \sim \triangle AMC$, we have

$$\frac{S_{BMD}}{\frac{1}{2} S_{ABC}} = \frac{S_{BMD}}{S_{AMC}} = \frac{|BM|^2}{|AM|^2} = \frac{1}{9} \frac{|BC|^2}{|AP|^2},$$

so that

$$\frac{S_{BPD}}{S_{ABC}} = \frac{S_{BPM}}{S_{ABC}} + \frac{S_{BMD}}{S_{ABC}} = \frac{1}{6} + \frac{1}{18} \frac{|BC|^2}{|AP|^2} = \frac{1}{18} \cdot \frac{3|AP|^2 + |BC|^2}{|AP|^2}.$$

From the formula for the length of the median of a triangle,

$$|AB|^2 + |CA|^2 = 2|AM|^2 + \frac{1}{2}|BC|^2 = \frac{9}{2}|AP|^2 + \frac{1}{2}|BC|^2,$$

we deduce

$$3|AP|^2 + |BC|^2 = \frac{2}{3}(|AB|^2 + |BC|^2 + |CA|^2),$$

and by addition of the formulas for all three medians we obtain

$$|AP|^2 + |BP|^2 + |CP|^2 = \frac{1}{3}(|AB|^2 + |BC|^2 + |CA|^2).$$

Hence

$$\frac{|AP|}{|PD|} + \frac{|BP|}{|PE|} + \frac{|CP|}{|PF|} = \frac{1}{3} \cdot \frac{18 \cdot |AP|^2}{3|AP|^2 + |BC|^2} \cdot \frac{|AP|^2 + |BP|^2 + |CP|^2}{|AP|^2} = 3.$$

THU PHAM, student
Tom C. Clark High School
San Antonio, Texas

Also solved by Anders Bager (Denmark), Jean-Marie Becker (France), Stavros A. Belbas, Walter Bluger (Canada), W. J. Blundon (Canada), Charles Chouteau, Clayton W. Dodge, Alan Edelman, Howard Eves, Jack Garfunkel, L. Kuipers (Switzerland), Moshe Lotan (Israel), Hubert J. Ludwig, John Oman, Jeremy D. Primer, St. Olaf College Problem Solving Group, J. M. Stark, Theodore Tollis (Greece), M. Vowe (Switzerland), Robert L. Young, Harry Zaremba, and the proposer, as well as (implicitly) all solvers of problem 1120 not mentioned above.

All solvers except Pham used methods which, except possibly for notation and terminology (vectors, complex numbers, power of a point with respect to a circle), are like the featured solutions of problem 1120, specialized to the conditions of this problem.

Centroid and Circumcircle, Generalized

March 1981

1120. Let the triangle ABC be inscribed in a circle and let P be a point in the interior of the circle. The line segments AP , BP , and CP are extended to meet the circle in points D , E , and F , respectively. Describe all such P for which

$$\frac{|AP|}{|PD|} + \frac{|BP|}{|PE|} + \frac{|CP|}{|PF|} \leq 3.$$

[Peter Ørno, The Ohio State University.]

Solution I: In the complex plane the following points are given: 0 as center of the unit circle C ; a_i ($i = 1, 2, \dots, n$) as n points on C , with $a_i \bar{a}_i = 1$; a point p in the interior of C , with $p\bar{p} < 1$; and b_i ($i = 1, 2, \dots, n$) as the n points in which the lines through a_i and p meet C , with $b_i \bar{b}_i = 1$.

Let $\lambda_i = (a_i - p)/(p - b_i)$ ($i = 1, 2, \dots, n$) and $\lambda = \sum_{i=1}^n \lambda_i$. Since λ is a positive real number, we have $b_i = (1 - p\bar{a}_i)/(\bar{p} - \bar{a}_i)$, $p - b_i = (1 - p\bar{p})/(\bar{a}_i - \bar{p})$, and

$$\lambda = \sum_{i=1}^n (a_i - p)(\bar{a}_i - \bar{p})/(1 - p\bar{p}).$$

Let s be the centroid of the n points a_i ($i = 1, 2, \dots, n$). Then the set of all points p yielding a given fixed λ is described by the equation

$$(\lambda + n)p\bar{p} - n\bar{s}p - ns\bar{p} + n - \lambda = 0. \quad (1)$$

The required set is a circle with

$$\text{center } m = \frac{ns}{n + \lambda} \text{ and radius } \rho = \frac{\sqrt{\lambda^2 - n^2(1 - s\bar{s})}}{n + \lambda},$$

if $\lambda \geq n\sqrt{1 - s\bar{s}}$.

For problem 1119 we are given that $n = 3$ and $\lambda = 3$, so that $m = s/2$, $\rho = |s|/2$, and s lies on the circle (1).

For problem 1120 we are given that $n = 3$ and $\lambda \leq 3$, so that the required set is the closed disc having the circle solving problem 1119 as boundary.

J. C. BINZ
Bern University, Switzerland

Solution II: The problem can be generalized to n points on a circle or sphere.

Let A_i ($i = 1, 2, \dots, n$) be n points on the surface of a sphere, and let P be an interior point. The segments A_iP are extended to meet the sphere at B_i . Describe the set of all such P for which

$$\sum_{i=1}^n \frac{|A_iP|}{|PB_i|} \leq n.$$

Let the points $A_i = (x_i, y_i, z_i)$ lie on the surface of the unit sphere with center at the origin, and let $P = (x, y, z)$ be an interior point. Then $|A_iP||PB_i| = 1 - x^2 - y^2 - z^2$, since P is an interior point of the great circle containing A_i and B_i . Hence

$$\sum_{i=1}^n \frac{|A_iP|}{|PB_i|} = \sum_{i=1}^n \frac{|A_iP|^2}{|A_iP||PB_i|} = \frac{\sum_{i=1}^n |A_iP|^2}{1 - x^2 - y^2 - z^2},$$

and so

$$\begin{aligned} \sum_{i=1}^n \frac{|A_iP|}{|PB_i|} \leq n &\Leftrightarrow \sum_{i=1}^n |A_iP|^2 \leq n(1 - x^2 - y^2 - z^2) \\ &\Leftrightarrow \sum_{i=1}^n ((x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2) \leq n(1 - x^2 - y^2 - z^2) \\ &\Leftrightarrow x^2 + y^2 + z^2 - \bar{x}x - \bar{y}y - \bar{z}z \leq 0, \end{aligned}$$

where $(\bar{x}, \bar{y}, \bar{z})$ is the centroid of the points A_i . The last inequality is satisfied just when P lies on or inside the sphere having a diameter whose endpoints are $(0, 0, 0)$ and $(\bar{x}, \bar{y}, \bar{z})$.

K. R. S. SASTRY
Addis Ababa, Ethiopia

Also solved by Anders Bager (Denmark), Jean-Marie Becker (France), Stavros A. Belbas, Walter Bluger (Canada), W. J. Blundon (Canada), Benny Cheng & Dinh Th  Hung, Howard Eves, Hans Kappus (Switzerland), L. Kuipers (Switzerland), Moshe Lotan (Israel), John Oman, Paul Zwier, and the proposer. Zwier also proved a generalization.

Answers

Solutions to the Quickies which appear near the beginning of the Problems section.

Q672. Since the g.c.d. of $(t-p)^s$ and $(t-q)^k$ is 1, there exist polynomials f and g such that $(t-p)^sf(t) + (t-q)^kg(t) = 1$ for all t ; hence $(A-pI)^sf(A) + (A-qI)^kg(A) = I$. This implies that $y^T(A-qI)^kg(A) = y^T$ and so $y^Tx = y^Tg(A)(A-qI)^kx = 0$.

Q673. Imagine the log as the front of a moving platform, which the rower mounts after a miles. After $2b$ more hours, he will have returned to the front of the platform, which will have moved a miles. Thus the rate of the stream is $a/(2b)$ miles per hour.

REVIEWS

PAUL J. CAMPBELL, Editor

Beloit College

PIERRE J. MALRAISON, Jr., Editor

MDSI, Ann Arbor

Assistant Editor: Eric S. Rosenthal, West Orange, NJ. Articles and books are selected for this section to call attention to interesting mathematical exposition that occurs outside the mainstream of the mathematics literature. Readers are invited to suggest items for review to the editors.

Begley, Sharon, *Life in two dimensions*, Newsweek (18 January 1982) 84-85.

The popularity of Rubik's cube may have opened the popular press wider to mathematical curiosities of all kinds. Here is an account of the "practicalities" of living in Flatland: two-dimensional laws of science, gadgets, and beings, as invented by Alexander Dewdney (U. Western Ontario) and furthered by more than 1000 others. Martin Gardner treated the topic 18 months earlier (*Scientific American*).

Abelson, Philip H., and Dorfman, Mary (eds.), *Computers and Electronics*, Science 215 (Special Issue), (12 February 1982) 749-873.

Collection of 21 articles focused on six themes. Of particular interest to mathematical scientists are articles on advances in computer graphics, software development, and the Unix operating system as a model for software design. Other articles survey computer prospects in scientific research, business, communications, and information retrieval.

Landau, Barbara, et al., *Spatial knowledge and geometric representation in a child blind from birth*, Science 213 (11 September 1981) 1275-1278.

"Our research indicates that the locomotion of the young blind child is guided by knowledge of the Euclidean properties of a spatial layout and by principles for making inferences based on those properties." More accurately, as the authors mention in a footnote, Euclidean, hyperbolic, and Riemannian geometries--but not topology or projective geometry--can all explain the observations.

Bühler, W.K., Gauss: A Biographical Study, Springer-Verlag, 1981; viii + 208 pp.

An attempt (in the words of Nicolaus von Fuss) "to describe the life of a great man who distinguishes his century by a considerable degree of enlightenment." Chapters of biography and description of Gauss's scientific work are interspersed with "interchapters" on sidelights such as the contemporary political and social situation and Gauss's style. The result is a rich and readable book. Appendices survey the volumes of Gauss's collected works, and the secondary literature, and give an index of Gauss's works.

Schmidt, Olaf, *On Plimpton 322. Pythagorean numbers in Babylonian mathematics*, Centaurus 24 (1980) 4-13.

New explanation of the numbers on the most famous Babylonian clay tablet, based on tables of reciprocals and the Babylonian method of solving certain quadratic equations.

Young, Laurence, Mathematicians and Their Times, North-Holland, 1981; x + 344 pp, \$36.50.

Lectures interpreting the history of mathematics up to World War II. Particularly valuable are personal glimpses of life at Cambridge and of mathematicians of the early decades of this century. The author was trained in mathematics at Cambridge, and knew most of the mathematicians of the time, in part through his parents, W.H. Young and G.C. Young.

Sondheimer, Ernst, and Rogerson, Alan, Numbers and Infinity: An Historical Account of Mathematical Concepts, Cambridge U Pr, 1981; x + 172 pp, \$15.95, \$7.95 (P).

Compact and insightful treatment of the concepts of number and infinity through history. Stimulating reading for calculus students, with suggestions for term paper topics.

Fox, Lynn H., *et al.* (eds.), Women and the Mathematical Mystique, Johns Hopkins U Pr, 1980; xi + 211 pp, \$5.95 (P).

"The thrust of this volume is to identify those sex differences with respect to mathematics that appear to be a result of social learning and to consider ways in which such behaviors and attitudes can be changed or prevented." The ten studies reported here defuse one stereotype after another and offer specific directions for research and for change. Essential reading for teachers of mathematics at all levels. Expanded version of 1976 AAAS Symposium entitled "Women and Mathematics".

Davis, Randall, and Lenat, Douglas B., Knowledge-Based Systems in Artificial Intelligence, McGraw-Hill, 1980; xxi + 490 pp.

Detailed descriptions of research using two computer programs. One, Teiresias, shows how a computer can acquire and incorporate knowledge from experts. The other, AM, deals with a paradigm for mathematical discovery and research. This program can start from a knowledge base of concepts, together with heuristic rules for exploring their connections; it tries to build "significant" new concepts from old ones. Starting from basic set theory, AM discovered cardinality, addition, multiplication, primality, powers, roots, factoring. AM cannot prove, only conjecture; but the possibility that it might inspire new mathematics is intriguing.

UMAP Modules, 1980: Tools for Teachings, Birkhäuser, 1981; xii + 690 pp.

Collection of 28 self-contained, lesson-length instructional units on applications of mathematics ranging from seismology to arms races, from the human cough to dietary management. First of annual sequels to UMAP Modules 1977-1979.

Elseth, G.D., and Baumgardner, K.D., Population Biology, Van Nostrand, 1981; xvi + 623 pp, \$21.95.

Emphasizes the theoretical basis of population biology; approaches it through mathematical model-building with testing against data. Calculus, matrix algebra, and statistics are assumed, and both example problems and exercises are included. Topics include genetics, ecology, and evolution. Perfect for a course in biology-oriented mathematical modeling.

Roberts, Fred S., Measurement Theory with Applications to Decision-making, Utility, and the Social Sciences, Addison-Wesley, 1979; xxiii + 420 pp, \$24.50.

Provides mathematical basis for measurement in the social and behavioral sciences. Features applications: measurement of utility, psycho-physical scaling and public policy decision-making. The language and presentation are akin to those of an abstract algebra book, so sophistication is demanded of the reader. Exercises are included.

Mason, Robert L., and Gunst, Richard F., Regression Analysis and Its Application: A Data-Oriented Approach, Dekker, 1980; xiv + 402 pp, \$39.75.

Joins theoretical coverage of regression analysis with a working knowledge of the broad range of regression modeling and estimation techniques. Data analysis is emphasized over classical parametric model formulation, and the emphasis is brought home by the analysis of real--and sometimes "dirty"--data sets. A prime consideration is what to do if the data don't perfectly fit the model, leading to consideration of care in model building, residual analysis, outlier detection, variable selection, and multicollinearity.

Greene, Daniel H., and Knuth, Donald E., Mathematics for the Analysis of Algorithms, Birkhäuser, 1981; 107 pp, \$10.

Collection of handouts for an advanced course at Stanford on analysis of algorithms, with much material drawn from Knuth's *The Art of Computer Programming*. Topics: binomial identities, recurrence relations, operator methods, and asymptotic analysis (most of the book). Includes midterm and final exams, with solutions. Not for computer students skittish about mathematics.

Koerner, James D., The New Liberal Arts: An Exchange of Views, Alfred P. Sloan Foundation, 1981; vii + 77 pp (P).

Examines the place of quantitative study in a liberal education, emphasizing analytic skills (mathematical modeling and manipulation) and "technological literacy" (including, but not limited to, computer literacy). An initial essay by Stephen White (Sloan Foundation) is followed by ten responses from educators, including Donald L. Kreider, Chairman of Mathematics at Dartmouth. Some respondents see growing emphasis on quantitative study as essential to the liberal arts enterprise, while others fear vocational and technical studies will expand to obliterate liberal education. Most believe the leadership of liberal education will implement a drive toward greater quantitative study: "The timing is right and the illiteracy is appalling."

Nourse, James G., The Simple Solutions to Cubic Puzzles, Bantam, 1981; 64 pp, \$1.95 (P).

Solutions to new manipulative puzzles similar to Rubik's cube, including pyramid, barrel, octagon, and other slide-and-rotate puzzles. Most are easier than the famous cube. Also included are ideas for shapes for Rubik's new creation, the snake. Rubik himself might disapprove of suggestions for the latter, since he invented the snake to encourage creativity; unlike the cube, there's nothing you're *supposed* to do with it.

Runyon, Richard P., How Numbers Lie: A Consumer's Guide to the Fine Art of Numerical Deception, Lewis Publ. Co., 1981; viii + 182 pp, \$7.95 (P).

Essentially a revised version of the author's *Winning with Statistics* (1977), this easy-reading book offers new chapters on gambling, "graphic gullduggery," computer crime, and statistics and health. Suitable as a motivational supplement at the beginning of an elementary statistics course.

NEWS & LETTERS

MORE ON ELEVATORS

Readers of A Wuffle's "The Pure Theory of Elevators" (this *Magazine*, Jan. 1982, 30-37) may be interested in the applied theory of elevators, as treated by D.A. Field, "Investigating Mathematical Models," *Amer. Math. Monthly*, 85(1978) 196-197, which urges use of computer simulation; and B.A. Powell, "Mathematical modelling of elevator systems," in W.E. Boyce (ed.), *Case Studies in Mathematical Modeling*, Pitman, 1981, pp. 18-53, which includes a list of references on the subject.

Paul Campbell
Reviews Editor

(D.E. Knuth points out a fallacy in Gamow and Stern's reasoning for the case of more than one elevator in *J. Rec. Math.*, 2(1969) 131-137 -- editor.)

MIAMI UNIVERSITY CONFERENCE

The theme of the 10th annual conference, October 1-2, 1982 is "Mathematics and Computing." Speakers include Harold Abelson, MIT; William Buttelmann, Ohio State U.; James T. Fey, U. of Maryland; Stephen C. Kleene, U. of Wisconsin; Bruce Weide, Ohio State U. Contributed papers relating to the general theme, particularly those dealing with the use of computers in the teaching of mathematics, are invited. Abstracts should be sent by June 1, 1982, to Don Koehler or Fred Schuurmann, Dept. of Mathematics and Statistics, Miami University, Oxford, Ohio 45056.

S
O
L
U
T
I
O
N

(p. 176)

1	2	3	1		4	5	3	6	3	7	1	
8	5	0	9		7	0	6	3	10	8		
11	4	2	4	1				12	1	4	4	
				13	2	9	7	1		14	2	1
			15	5	5	5	5	16	5	5		
17	3	6		1	0	9	8					
18	1	4	19	5			20	7	7	15	16	6
21	6	2	0	10	3			21	1	2	7	
		22	1	0	2	4		23	6	6	6	

1981 PUTNAM SOLUTIONS

The following solutions to the 1981 Putnam Exam questions were prepared by Loren Larson of St. Olaf College, utilizing some of the results prepared by the Putnam Committee.

A-1. Let $E(n)$ denote the largest integer k such that 5^k is an integral divisor of the product $1!2!3!\dots n!$. Calculate

$$\lim_{n \rightarrow \infty} \frac{E(n)}{n^2}.$$

Sol. Let $T(m) = 1 + 2 + \dots + m = m(m+1)/2$, $[x]$ denote the greatest integer in x , $h = [\log_5 n]$, and $e_i = n/5^i - [n/5^i]$ for $1 \leq i \leq h$. Then $E(n) =$

$$\begin{aligned} & \sum_{i=1}^h 5^i T([n/5^i]) = \\ & \sum_{i=1}^h \frac{5^i}{2} ([n/5^i]^2 + [n/5^i]) = \\ & \sum_{i=1}^h \frac{5^i}{2} (n^2/5^{2i} - 2e_i n/5^i + \\ & e_i^2 + n/5^i - e_i) = \\ & \frac{n^2}{2} \left(\sum_{i=1}^h 1/5^i \right) - \frac{n}{2} \left(\sum_{i=1}^h e_i \right) + \frac{hn}{2} + \\ & \sum_{i=1}^h \frac{5^i}{2} (e_i^2 - e_i). \end{aligned}$$

It is now easy to show that $E(n)/n^2 \rightarrow 1/8$ as $n \rightarrow \infty$.

A-2. Two distinct squares of the 8 by 8 chessboard C are said to be adjacent if they have a vertex or side in common. Also, g is called a C -gap if for every numbering of the squares of C with all the integers $1, 2, \dots, 64$ there exist two adjacent squares whose numbers differ by at least g . Determine the largest C -gap g .

Sol. Two squares are adjacent if they are a (chess) King's move apart.

For any numbering, one can go from the square numbered 1 to the square numbered 64 in 7 or fewer successive King moves; thus a C -gap must necessarily be $\geq \frac{64-1}{7} = 9$. But no number greater than 9 can be a C -gap since 9 is attained in the labeling associated with the 8 by 8 matrix $A = (a_{ij})$, $a_{ij} = 8(i-1) + j$.

A-3. Find

$$\lim_{t \rightarrow \infty} [e^{-t} \int_0^t \int_0^t \frac{e^x - e^y}{x - y} dx dy]$$

or show that the limit does not exist.

Sol. Let $G(t)$ be the double integral. By L'Hopital's Rule, $\lim_{t \rightarrow \infty} G(t)/e^t = \lim_{t \rightarrow \infty} G'(t)/e^t$. One sees that $G'(t) = 2 \int_0^t \int_0^y \frac{e^x - e^y}{x - y} dx dy$, so $G'(t)/e^t = \frac{2}{e^t} \int_0^t \frac{e^x - e^t}{x - t} dx = 2 \int_0^t [1 + \frac{x-t}{2!} + \frac{(x-t)^2}{3!} + \dots] dx = 2[t + \frac{1}{2 \cdot 2!} t^2 + \dots] \rightarrow \infty$ as $t \rightarrow \infty$.

A-4. A point P moves inside a unit square in a straight line at unit speed. When it meets a corner it escapes. When it meets an edge its line of motion is reflected so that the angle of incidence equals the angle of reflection.

Let $N(T)$ be the number of starting directions from a fixed interior point P_0 for which P escapes within T units of time. Find the least constant α for which constants b and c exist such that

$$N(T) \leq \alpha T^2 + bT + c$$

for all $T > 0$ and all initial points P_0 .

Sol. Set up coordinates so that a vertex of the given unit square is $(0,0)$ and two sides of the square are on the axes. Using reflection properties, one can see that P escapes within T units of time if and only if the ray from P_0 , with the direction of the

first segment of the path, goes through a lattice point within T units of distance from P_0 . Thus $N(T)$ is at most the number $L(T)$ of lattice points in the circle with center at P_0 and radius T . Tiling the plane with unit squares having centers at the lattice points and considering areas, one sees that $N(T) \leq L(T) \leq \pi[T + \sqrt{2}/2]^2$. Hence there is an upper bound for $N(T)$ of the form $\pi T^2 + bT + c$, with b and c fixed. When just one coordinate of P_0 is irrational, $N(T) = L(T) \geq \pi[T - \sqrt{2}/2]^2$. This lower bound for $N(T)$ exceeds $\alpha T^2 + bT + c$ for sufficiently large T if $\alpha < \pi$; hence π is the desired α .

A-5. Let $P(x)$ be a polynomial with real coefficients and form the polynomial

$$Q(x) = (x^2 + 1)P(x)P'(x) + x([P(x)]^2 + [P'(x)]^2).$$

Given that the equation $P(x) = 0$ has n distinct real roots exceeding 1, prove or disprove that the equation $Q(x) = 0$ has at least $2n - 1$ distinct real roots.

Sol. We show that $Q(x)$ has at least $2n - 1$ real zeros. One finds that $Q(x) = F(x)G(x)$, where $F(x) = P'(x) + xP(x) = e^{-x^2/2} [e^{x^2/2} P(x)]'$, $G(x) = xP'(x) + P(x) = [xP(x)]'$. We can assume that $P(x)$ has exactly n zeros α_i exceeding 1 with $1 < \alpha_1 < \alpha_2 < \dots < \alpha_n$. It follows from Rolle's Theorem that $F(x)$ has $n - 1$ zeros b_i , $\alpha_i < b_i < \alpha_{i+1}$, and $G(x)$ has n zeros c_i with $0 < c_i < \alpha_i$, $\alpha_i < c_{i+1} < \alpha_{i+1}$, $i = 1, 2, \dots, n$. If $b_i \neq c_{i+1}$ for all i , we're done. So assume that $b_i = c_{i+1} = r$ for some i . Then $P'(r) + rP(r) = 0 = rP'(r) + P(r)$ and so $(r^2 - 1)P(r) = 0$. Since $r = b_i > 1$, $P(r) = 0$. Since $\alpha_i < r < \alpha_{i+1}$, this contradicts the fact that the α 's are all the zeros exceeding 1 of $P(x)$.

A-6. Suppose that each of the vertices of $\triangle ABC$ is a lattice point in the (x,y) -plane and that there is exactly one lattice point P in the interior of the triangle. The line AP is extended to meet BC at E . Determine

the largest possible value for the ratio of lengths of segments

$$|AP|/|PE|.$$

(A lattice point is a point whose coordinates x and y are integers.)

Sol. Each point X of the plane may represent the vector with initial point at A and final point at X . As vectors, $P = xB + yC$ where $0 < x$, $0 < y$, and $x + y < 1$. We may assume that $y \leq x$. Let $Q = 2P - B = (2x-1)B + 2yC$. Q is a lattice point and $Q \neq P$. As $2x-1 + 2y < 2-1 = 1$ and $0 < 2y$, Q would be inside $\triangle ABC$ if $0 < 2x-1$. Since P is the only lattice point inside $\triangle ABC$, $2x-1 \leq 0$ and so $x \leq 1/2$. Let $R = 3P - B - C = (3x-1)B + (3y-1)C$. As before, $(3x-1) + (3y-1) < 1$. This and $y \leq x$ imply $3y - 1 \leq 0$. Thus $y \leq 1/3$. Now $x + y \leq 1/2 + 1/3 = 5/6$. It follows that

$$\frac{|AP|}{|PE|} = \frac{(x+y)|AE|}{[1 - (x+y)]|AE|} \leq \frac{5/6}{1/6} = 5.$$

We see that this upper bound 5 is the maximum by considering the case with $A = (0,0)$, $B = (0,2)$, $C = (3,0)$, and $P = (1,1)$.

An alternate approach is given by Prof. Clifton Corzatt of St. Olaf College. Coordinatize the plane so that $A = (0,0)$. Let D be the lattice point on the line segment AB which is closest to, but different from, A (D may equal B). Suppose that $P = (a,b)$ and $D = (c,d)$. The matrix $\begin{bmatrix} -b & a \\ d-b & a-c \end{bmatrix}$ defines an area-preserving linear transformation which sends (a,b) to $(0,1)$ and (c,d) to $(1,1)$. [Note that $ad-bc = 1$ since, $(ad-bc)/2 = \text{Area } \triangle ADP$ (by vector analysis) $= 1/2$ (by Pick's Theorem).] Therefore, we may assume that $P = (0,1)$ and $D = (1,1)$. There are now only a small number of cases to check: the maximum ratio of 5 is attained when $B = (3,3)$ and $C = (-2,0)$.

B-1. Find

$$\lim_{n \rightarrow \infty} \left[\frac{1}{n^5} \sum_{h=1}^n \sum_{k=1}^n (5h^4 - 18h^2k^2 + 5k^4) \right].$$

Sol. Let $S_k(n) = 1^k + 2^k + \dots + n^k$.

One finds that $S_2(n) = n^3/3 + n^2/2 + an$ and $S_4(n) = n^5/5 + n^4/2 + bn^3 + cn^2 + dn$ with a, b, c, d constants. Then the

double sum is $10nS_4(n) - 18[S_2(n)]^2 = (2n^6 + 5n^5 + \dots) - (2n^6 + 6n^5 + \dots) = -n^5 + \dots$ and the desired limit is -1 .

B-2. Determine the minimum value of

$$(r-1)^2 + \left(\frac{s}{r}-1\right)^2 + \left(\frac{t}{s}-1\right)^2 + \left(\frac{4}{t}-1\right)^2$$

for all real numbers r, s, t with $1 \leq r \leq s \leq t \leq 4$.

Sol. First we let $0 < a < b$ and seek x that minimizes $f(x) = (x/a - 1)^2 + (b/x - 1)^2$ on $a \leq x \leq b$. Using standard methods of calculus one finds that the minimum occurs at $x = \sqrt{ab}$. Then the minimum for the given function of r, s, t occurs with $r = \sqrt{s}$, $t = \sqrt{4s} = 2\sqrt{s}$, and $s = \sqrt{rt} = r\sqrt{2}$. These imply $r = \sqrt{2}$, $s = 2$, $t = 2\sqrt{2}$. Thus the desired minimum value is $4(\sqrt{2} - 1)^2 = 12 - 8\sqrt{2}$.

B-3. Prove that there are infinitely many positive integers n with the property that if p is a prime divisor of $n^2 + 3$ then p is also a divisor of $k^2 + 3$ for some integer k with $k^2 < n$.

Sol. Let $f(x) = x^2 + 3$. Examination of $\{f(m)\} = \{3, 4, 7, 12, 17, 28, 39, 52, 67, 84, \dots\}$ leads one to conjecture that $f(x)f(x+1) = f(x(x+1)+3) = f(x^2+x+3)$. This is easily proved.

$$\text{Let } n = (m^2+m+2)(m^2+m+3) + 3.$$

Then one has

$$\begin{aligned} f(n) &= f(m^2+m+2)f(m^2+m+3) \\ &= f(m^2+m+2)f(m)f(m+1). \end{aligned}$$

Thus $p|f(n)$ with p prime implies $p|f(k)$ with k equal to m , $m+1$, or m^2+m+2 . Since each of these possibilities for k satisfies $k^2 < n$, the desired result follows.

B-4. Let V be a set of 5 by 7 matrices, with real entries and with the property that $rA + sB \in V$ whenever $A, B \in V$ and r and s are scalars (i.e., real numbers). Prove or disprove the following assertion: If V contains matrices of ranks 0, 1, 2, 4, and 5 then it also contains a matrix of rank 3. (The rank of a nonzero matrix M is the largest k such that the entries of some k rows and some k columns form a k by k matrix with a nonzero determinant.)

Sol. Let $M = M(a, b, c)$ denote the 5 by 7 matrix (a_{ij}) with $a_{11} = a$, $a_{22} = a_{33} = a_{44} = a_{55} = b$, $a_{16} = a_{27} = c$, and $a_{ij} = 0$ in all other cases. Then the set V of all such M (with a, b, c arbitrary real numbers) is closed under linear combinations, contains matrices of ranks 0, 1, 2, 4, 5, but does not contain a matrix of rank 3.

B-5. Let $B(n)$ be the number of ones in the base two expression for the positive integer n . For example, $B(6) = B(110_2) = 2$ and $B(15) = B(1111_2) = 4$. Determine whether or not

$$\exp \left(\sum_{n=1}^{\infty} \frac{B(n)}{n(n+1)} \right)$$

is a rational number. Here $\exp(x)$ denotes e^x .

Sol. If n has d digits in base two, $2^{d-1} \leq n$ and so $B(n) \leq d \leq 1 + \log_2 n$.

This implies that $\sum_{n=1}^{\infty} \frac{B(n)}{n(n+1)}$ converges to a real number S . It follows that

$$\begin{aligned} S &= \sum_{n=1}^{\infty} \frac{B(n)}{n(n+1)} \\ &= \sum_{m=0}^{\infty} \frac{B(2m+1)}{(2m+1)(2m+2)} + \sum_{m=1}^{\infty} \frac{B(2m)}{2m(2m+1)} \\ &= \sum_{m=0}^{\infty} \frac{1 + B(m)}{(2m+1)(2m+2)} + \sum_{m=1}^{\infty} \frac{B(m)}{2m(2m+1)} \\ &= \sum_{m=0}^{\infty} \frac{1}{(2m+1)(2m+2)} + \\ &\quad \sum_{m=1}^{\infty} B(m) \left[\frac{1}{2m(2m+1)} + \frac{1}{(2m+1)(2m+2)} \right] \\ &= \ln 2 + S/2. \end{aligned}$$

It follows that $S = 2 \ln 2$ and so $\exp(S) = 4$.

Alternately, if $n = \sum_{m=0}^{\infty} a_{n,m} 2^m$ is the binary representation of n ,

$$\begin{aligned} S &= \sum_{n=1}^{\infty} \frac{B(n)}{n(n+1)} = \sum_{n=1}^{\infty} \sum_{m=0}^{\infty} \frac{a_{n,m}}{n(n+1)} \\ &= \sum_{m=0}^{\infty} \sum_{n=1}^{\infty} \frac{a_{n,m}}{n(n+1)} = \dots = \sum_{m=0}^{\infty} \frac{1}{2^m} \ln 2 \\ &= 2 \ln 2. \end{aligned}$$

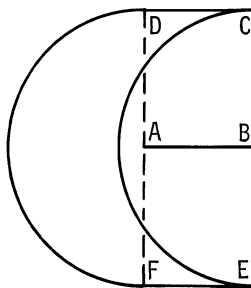
B-6. Let C be a fixed unit circle in the Cartesian plane. For any convex polygon P each of whose sides is tangent to C , let $N(P, h, k)$ be the number of points common to P and the unit circle with center at (h, k) . Let $H(P)$ be the region of all points (x, y) for which $N(P, x, y) \geq 1$ and $F(P)$ be the area of $H(P)$. Find the smallest number u with

$$I = \frac{1}{F(P)} \iint N(P, x, y) \, dx \, dy < u$$

for all polygons P , where the double integral is taken over $H(P)$.

Sol. Let $L = L(P)$ be the perimeter of P . One sees that $H(P)$ consists of the region bounded by P , the regions bounded by rectangles whose bases are the sides of P and whose altitudes equal 1, and sectors of unit circles which can be put together to form one unit circle. Hence $F(P) = L/2 + L + \pi = \pi + 3L/2$.

If A and B are two consecutive vertices of P , the contribution of side AB to the double integral I is double the area of the region (of the figure)



bounded by the unit semicircles with centers at A and B and segments CD and EF such that $ABCD$ and $ABEF$ are rectangles and $|AD| = 1 = |AF|$. One doubles this area because there is a symmetric region bounded by CD , EF , and the other halves of the unit circles centered at A and B . The overlap of the two regions counts twice. By Cavalieri's slicing principle, this contribution of side AB to I is four times the length of AB . Hence

$$\frac{I}{F(P)} = \frac{4L}{\pi + 3L/2} = \frac{8}{3 + 2\pi/L}.$$

One can make L arbitrarily large and therefore the desired least upper bound is $8/3$.

MAA STUDIES IN MATHEMATICS

MAA STUDIES IN MATHEMATICS present to the mathematical community expository papers in topics at the research frontiers written for a broad audience. Each volume is edited by a prominent mathematician active in the area of the study. There are twenty-one volumes currently available in this series:

		List Price	Member Price
Vol. 1.	Studies in Modern Analysis , R. C. Buck, editor	\$16.50	\$12.00
Vol. 2.	Studies in Modern Algebra , A. A. Albert, editor	16.50	12.00
Vol. 3.	Studies in Real and Complex Analysis , I. I. Hirschmann, Jr., editor	16.50	12.00
Vol. 4.	Studies in Global Geometry and Analysis , S. S. Chern, editor	16.50	12.00
Vol. 5.	Studies in Modern Topology , P. J. Hilton, editor	16.50	12.00
Vol. 6.	Studies in Number Theory , W. J. LeVeque, editor	16.50	12.00
Vol. 7.	Studies in Applied Mathematics , A. H. Taub, editor	16.50	12.00
Vol. 8.	Studies in Model Theory , M. D. Morley, editor	16.50	12.00
Vol. 9.	Studies in Algebraic Logic , Aubert Daigneault, editor	16.50	12.00
Vol. 10.	Studies in Optimization , George B. Dantzig and B. Curtis Eaves, editors	16.50	12.00
Vol. 11.	Studies in Graph Theory , Part I, D. R. Fulkerson, editor	16.50	12.00
Vol. 12.	Studies in Graph Theory , Part II, D. R. Fulkerson, editor	16.50	12.00
Vol. 13.	Studies in Harmonic Analysis , J. M. Ash, editor	21.00	16.00
Vol. 14.	Studies in Ordinary Differential Equations , Jack Hale, editor	21.00	16.00
Vol. 15.	Studies in Mathematical Biology , Part I, "Cellular Behavior and Development of Pattern," Simon Levin, editor	21.00	16.00
Vol. 16.	Studies in Mathematical Biology , Part II, "Populations and Communities," Simon Levin, editor	21.00	16.00
Vol. 17.	Studies in Combinatorics , Gian-Carlo Rota, editor	21.00	16.00
Vol. 18.	Studies in Probability , Murray Rosenblatt, editor	21.00	16.00
Vol. 19.	Studies in Statistics , R. V. Hogg, editor	18.00	13.50
Vol. 20.	Studies in Algebraic Geometry , Abraham Seidenberg, editor	16.00	12.00
Vol. 21.	Studies in Functional Analysis , Robert G. Bartle, editor	19.00	14.00

ORDER FROM: **THE MATHEMATICAL ASSOCIATION OF AMERICA**
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

Eminent Mathematicians and Mathematical Expositors speak to
STUDENTS and TEACHERS in . . .

The NEW MATHEMATICAL LIBRARY

An internationally acclaimed paperback series providing:

- stimulating excursions for students beyond traditional school mathematics.
- supplementary reading for school and college classrooms.
- valuable background reading for teachers.
- challenging problems for solvers of all ages from high school competitions in the US and abroad.

The New Mathematical Library is published by the MATHEMATICAL ASSOCIATION OF AMERICA. The volumes are paperback.

For information regarding the price of these publications, please contact The Mathematical Association of America at the address listed below.

NUMBERS: RATIONAL AND IRRATIONAL by Ivan Niven, NML-01

WHAT IS CALCULUS ABOUT? By W. W. Sawyer, NML-02

AN INTRODUCTION TO INEQUALITIES, by E. F. Beckenbach, and R. Bellman, NML-03

GEOMETRIC INEQUALITIES, by N. D. Kazarinoff, NML-04

THE CONTEST PROBLEM BOOK. Problems from the Annual High School Mathematics Contests sponsored by the MAA, NCTM, Mu Alpha Theta, The Society of Actuaries, and the Casualty Actuarial Society. Covers the period 1950-1960. Compiled and with solutions by C. T. Salkind. NML-05

THE LORE OF LARGE NUMBERS, by P. J. Davis, NML-06

USES OF INFINITY, by Leo Zippin, NML-07

GEOMETRIC TRANSFORMATIONS, by I. M. Yaglom, translated by Allen Shields, NML-08

CONTINUED FRACTIONS, by C. D. Olds, NML-09

GRAPHS AND THEIR USES, by Oystein Ore, NML-10

HUNGARIAN PROBLEM BOOKS I and II, based on the Eötvös Competitions 1894-1905 and 1906-1928. Translated by E. Rapaport, NML-11 and NML-12

EPISODES FROM THE EARLY HISTORY OF MATHEMATICS, by A. Aaboe, NML-13

GROUPS AND THEIR GRAPHS, by I. Grossman and W. Magnus, NML-14

THE MATHEMATICS OF CHOICE, by Ivan Niven, NML-15

FROM PYTHAGORAS TO EINSTEIN, by K. O. Friedrichs, NML-16

THE CONTEST PROBLEM BOOK II. A continuation of NML-05 containing problems and solutions from the Annual High School Mathematics Contests for the period 1961-1965. NML-17

FIRST CONCEPTS OF TOPOLOGY, by W. G. Chinn and N. E. Steenrod, NML-18

GEOMETRY REVISITED, by H. S. M. Coxeter, and S. L. Greitzer, NML-19

INVITATION TO NUMBER THEORY, by Oystein Ore, NML-20

GEOMETRIC TRANSFORMATIONS II, by I. M. Yaglom, translated by Allen Shields, NML-21

ELEMENTARY CRYPTANALYSIS — A Mathematical Approach, by Abraham Sinkov, NML-22

INGENUITY BY MATHEMATICS, by Ross Honsberger, NML-23

GEOMETRIC TRANSFORMATIONS III, by I. M. Yaglom, translated by Abe Shenitzer, NML-24

THE CONTEST PROBLEM BOOK III. A continuation of NML-05 and NML-17, containing problems and solutions from the Annual High School Mathematics Contests for the period 1966-1972. NML-25

MATHEMATICAL METHODS IN SCIENCE, by George Pólya, NML-26

INTERNATIONAL MATHEMATICAL OLYMPIADS, 1959-1977. Problems, with solutions, from the first nineteen International Mathematical Olympiads. Compiled and with solutions by S. L. Greitzer. NML-27

THE MATHEMATICS OF GAMES AND GAMBLING, by Edward W. Packel, NML-28



Send Orders to: **The Mathematical Association of America**
1529 Eighteenth St., N.W., Washington, D. C. 20036

The Mathematical Association of America

CARUS MATHEMATICAL MONOGRAPHS

Students and teachers of mathematics as well as nonspecialists and workers in other fields, will all appreciate the expository excellence of these monographs. A wide range of topics in pure and applied mathematics are covered, each one written by a renowned expositor in the field. A glance at the list of titles of the books and the list of authors will convince you of the quality of this respected series of monographs. Twenty titles are currently available:

		List price	MAA Member
No. 1.	Calculus of Variations , by G. A. Bliss	\$16.50	\$12.00
No. 2.	Analytic Functions of a Complex Variable , by D. R. Curtiss	16.50	12.00
No. 3.	Mathematical Statistics , by H. L. Rietz	16.50	12.00
No. 4.	Projective Geometry , by J. W. Young	16.50	12.00
No. 6.	Fourier Series and Orthogonal Polynomials , by Dunham Jackson	16.50	12.00
No. 8.	Rings and Ideals , by N. H. McCoy	16.50	12.00
No. 9.	The Theory of Algebraic Numbers (Second edition) by Harry Pollard, and H. G. Diamond	16.50	12.00
No. 10.	The Arithmetic Theory of Quadratic Forms , by B. W. Jones	16.50	12.00
No. 11.	Irrational Numbers , by Ivan Niven	16.50	12.00
No. 12.	Statistical Independence in Probability, Analysis and Number Theory , by Mark Kac	16.50	12.00
No. 13.	A Primer of Real Functions (Third edition), by R. P. Boas, Jr.	16.50	12.00 In Preparation
No. 14.	Combinatorial Mathematics , by H. J. Ryser	16.50	12.00
No. 15.	Noncommutative Rings , by I. N. Herstein	16.50	12.00
No. 16.	Dedekind Sums , Hans Rademacher and Emil Grosswald	16.50	12.00
No. 17.	The Schwarz Function and its Applications , by P. J. Davis	16.50	12.00
No. 18.	Celestial Mechanics , by Harry Pollard	16.50	12.00
No. 19.	Field Theory and its Classical Problems , by Charles Hadlock	21.00	16.00
No. 20.	The Generalized Riemann Integral , by R. M. McLeod	18.00	13.50

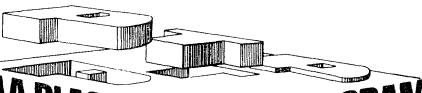


ORDER FROM:

THE MATHEMATICAL ASSOCIATION OF AMERICA

1529 Eighteenth Street, N.W.

Washington, D. C. 20036



MAA PLACEMENT TEST PROGRAM

EVERY STUDENT BELONGS

Let the MAA Placement Test Program help you match entering students with beginning mathematics courses according to *training* and *ability*, rather than transcripts and credentials. PTP tests are constructed by panels representing a broad spectrum of institutions and are carefully pretested. Information about pretesting scores and placement experience of a variety of participating institutions is published periodically in the PTP Newsletter.

Your institution can have unlimited use of annually updated MAA Placement Tests on

- Basic Mathematical Skills
- Basic Algebra
- Advanced Algebra
- Trigonometry/Elementary Functions
- Calculus Readiness

and also a subscription to the PTP Newsletter for one modest annual subscription fee.



For information write to:

The Mathematical Association of America

Department PTP

1529 Eighteenth Street, N.W.

Washington, D. C. 20036

OPEN COMPETITION IN GEOMETRY

\$2500 AWARD

FOR THE MOST COMPLETE SOLUTION TO THE PROBLEM:

What geometrical properties of an inversion locus correspond to (are the consequences of) the property of self-invertibility of a basis curve at angles other than 0° or 180° ?

EXAMPLE: A Cassinian monoval self-inverts about its center at angles of 90° and 270° . No dissimilar inversion locus of a Cassinian monoval self-inverts at either of these angles. In what manner is the property of self-inversion of the basis Cassinian in 90° and 270° modes manifested in its inversion loci?

Entries must be original contributions to the theory of plane curves that include direct geometrical interpretations. For competition details and further information send stamped self-addressed envelope to:

SCIENCE SOFTWARE SYSTEMS, INC.
11899 W. Pico. Blvd. Los Angeles, Calif. 90064

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, DC 20036

MATHEMATICS MAGAZINE VOL. 55, NO. 3, MAY 1982