MATHEMATICS MAGAZINE

The Bieberbach Conjecture

Vol. 59 No. 3
June 1986

BRIDGE PROBABILITIES • PATROL PROBLEM
MODULAR ARITHMETIC • LARGEST UNIT BALL

MATHEMATICS MAGAZINE

Volume 59, Number 3, June 1986
(ISSN— 0025- 570X)

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

EDITORIAL POLICY

The aim of *Mathematics Magazine* is to provide lively and appealing mathematical exposition. This is not a research journal and, in general, the terse style appropriate for such a journal (lemma-theorem-proof-corollary) is not appropriate for an article for the *Magazine*. Articles should include examples, applications, historical background, and illustrations, where appropriate. They should be attractive and accessible to undergraduates and would, ideally, be helpful in supplementing undergraduate courses or in stimulating student investigations. Articles on pedagogy alone, unaccompanied by interesting mathematics, are not suitable. Neither are articles consisting mainly of computer programs unless these are essential to the presentation of some good mathematics. Manuscripts on history are especially welcome, as are those showing relationships between various branches of mathematics and between mathematics and other disciplines.

In addition to articles and notes the *Magazine* solicits proofs without words, mathematical verse, anecdotes, cartoons, and other such material consistent with the level and aims described above. Letters and comments are also welcome.

The full statement of editorial policy appears in this *Magazine*, Vol. 54, pp. 44–45, and is available from the Editor. Manuscripts to be submitted should not be concurrently submitted to, accepted for publication by, nor published by another journal or publisher. Send new manuscripts to: G. L. Alexanderson, Editor, Mathematics Magazine, University of Santa Clara, Santa Clara, CA 95053. Manuscripts should be typewritten and double spaced and prepared in a style consistent with the format of *Mathematics Magazine*. Authors should submit the original and one copy and keep one copy. Illustrations should be carefully prepared on separate sheets in black ink, the original without lettering and two copies with lettering added.

AUTHORS

Paul Zorn ("The Bieberbach Conjecture") received his Ph.D., in several complex variables, from the University of Washington. The BC story combines two of his interests—analysis and expository writing. He is also interested in using computer graphics and computer algebra systems to teach mathematical ideas, without teaching computer programming.

ILLUSTRATIONS

Photographs on pages 137 and 145 are from the George Pólya Picture Collection.

The *Mathematics Magazine* (ISSN 0025-570X) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, D.C. 20036 and Montpelier, VT, five times a year: January, March, May, September, and November.

The annual subscription price for the MATHEMATICS MAGAZINE to an individual member of the Association is $11 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $22. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The non-member/library subscription price is $28 per year. Bulk subscriptions (5 or more copies) are available to colleges and universities for classroom distribution to undergraduate students at a 41% discount ($6.50 per copy—minimum order $32.50).

Subscription correspondence and notice of change of address should be sent to the Membership/Subscriptions Department, 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 Company, Middletown, CT 06457. Microfilmed issues may be obtained from University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

Advertising correspondence should be addressed to Ms. Elaine Pedreira, Advertising Manager, The Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036.

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

Postmaster: Send address changes to Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036.

PRINTED IN THE UNITED STATES OF AMERICA

ARTICLES

The Bieberbach Conjecture

A famous unsolved problem and the story of de Branges' surprising proof.

PAUL ZORN
St. Olaf College
Northfield, Minnesota 55057

Introduction

The biggest mathematical news of 1984 was the proof of the 68-year-old Bieberbach conjecture. Louis de Branges of Purdue University had solved what Felix Browder of the University of Chicago described as "one of the most celebrated conjectures in classical analysis, one that has stood as a challenge to mathematicians for a very long time." Beginning in spring 1984, the news spread quickly through the mathematical world as de Branges lectured in Europe and as preprints and informal communications circulated. Even the New York Times (Sept. 4, 1984, p. C12) reported the story—incorrectly. What was all the fuss about?

The details of de Branges' ingenious proof are well beyond our scope. We will focus instead (mainly) on the conjecture itself. What does it say? Why would one conjecture it? What partial results are "obvious"? Why did so many mathematicians work so hard at the problem for so long? Who contributed to its solution? What comes next?

The Bieberbach conjecture is an attractive problem partly because it is easy to state—it says that under reasonable restrictions the coefficients of a power series are not too large. More precisely, let

$$f(z) = z + a_2 z^2 + a_3 z^3 + \cdots$$

be a power series in z, with complex coefficients $a_2, a_3, \ldots$. Assume that $f(z)$ converges for all complex numbers z $(= x + iy)$ with $|z|$ $(= (x^2 + y^2)^{1/2}) < 1$, and that the function $f(z)$ is one-to-one on the set of such z. Then

BIEBERBACH CONJECTURE. $|a_n| \leqslant n$, $n = 2, 3, \ldots$. *The inequality is strict for every n unless f is a "rotation" of the* **Koebe function**

$$k(z) = \sum_{n=1}^{\infty} nz^n = z + 2z^2 + 3z^3 + \cdots .$$

(The term "rotation" is defined at the end of the next section.)

The Bieberbach conjecture (hereafter, "BC") is at heart an assertion about extremality of the Koebe function. Therefore, understanding the Koebe function and why it is the natural candidate to be "biggest" in the sense of the conjecture is a recurring theme in this essay.

The BC first appeared in a footnote to a 1916 paper [4] of the German mathematician Ludwig Bieberbach, of which the principal result was the *second* coefficient theorem: $|a_2| \leqslant 2$; equality

940 Gesamtsitzung vom 20. Juli 1916. — Mitteilung vom 6. Juli

Über die Koeffizienten derjenigen Potenzreihen, welche eine schlichte Abbildung des Einheitskreises vermitteln.

Von Prof. Dr. Ludwig Bieberbach
in Frankfurt a. M.

(Vorgelegt von Hrn. Frobenius am 6. Juli 1916 [s. oben S. 775].)

The article

[1] Daß $k_n \geq n$ zeigt das Beispiel $\sum nz^n$. Vielleicht ist überhaupt $k_n = n$.

and the footnote that led to it all.

holds essentially only for the Koebe function. As we shall see, consequences of Bieberbach's theorem are at least as important as the theorem itself for understanding the BC, and in univalent function theory generally. Bieberbach himself proved no other analogous coefficient theorems. Before de Branges' general proof, $|a_n| \leqslant n$ was known only for $n \leqslant 6$.

The BC's words are familiar from elementary *real* calculus, but the meaning and interest of the conjecture are essentially rooted in *complex* analysis. This article aims to provide some mathematical and historical context for the BC. It is neither an exhaustive summary of the gigantic body of research in univalent function theory (which comprises thousands of papers; see, e.g., [2]) nor a careful presentation of de Branges' solution. It is an eclectic sample of background material, related results, and exercises related either to the conjecture itself or to standard ideas and techniques of the subject. We are less concerned with de Branges' dramatic achievement itself than with the stage setting—lighting, scenery, and backdrop—against which it is played.

Why write this essay when a wealth of clear and inviting expository books and articles already exists? (See, e.g., [3], [7], [8], [9], [11], [14], [15].) This article, which might have been titled "BC for Tourists" had not A. Baernstein [3] already used that phrase, is an invitation and a preamble to these sources, which assume more mastery of complex analysis. We assume that the reader is familiar with the rudiments of that subject (as developed in any undergraduate text), but *only* the rudiments; most of what is used is stated explicitly. Because all of the cited works contain extensive bibliographies, references for standard theorems discussed here are usually omitted.

What does the BC say?

The adjective "analytic" applied to a complex-valued function of a complex variable means "continuously differentiable in the complex sense"; i.e., f is a function of z, defined on a domain D in $\mathbb{C}$, and the limit

$$\lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0} = f'(z_0)$$

exists and is a continuous function of z_0. This generalization from real-variable calculus looks innocuous, but it is not. Analytic functions have pleasant properties not shared by their real-differentiable cousins. Most important, every analytic function can be expanded in a (complex) Taylor series

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n, \qquad a_n = \frac{f^{(n)}(z_0)}{n!}$$

about any point z_0 of its domain; the series converges to $f(z)$ in any circular neighborhood centered at z_0 that lies within the domain D. (Conversely, any series $\Sigma a_n z^n$ that converges for all z in D defines a function analytic on D.) None of this is true of an arbitrary real differentiable function: f may be no more than once differentiable, and so have no Taylor series at all; f's Taylor series, if it exists, may diverge; f's Taylor series, if it converges, may converge to a limit other than f.

Thus, an analytic function defined on any domain is, at least *locally*, a power series. An analytic function $f(z)$ that happens to be defined on the **unit disk** $\{z: |z| < 1\}$ in $\mathbb{C}$—hereafter referred to as D—is *globally* a (convergent) power series $\Sigma_{n=0}^{\infty} a_n z^n$. The sequence $\{a_0, a_1, a_2, \ldots\}$ of coefficients completely determines f's behavior. How, then, are analytic and geometric properties of f (e.g., univalence, boundedness, convexity) reflected in properties of f's coefficients (e.g., growth rate, individual bounds)? The BC (now de Branges' theorem) is the best-known "coefficient problem" of this kind.

The BC concerns functions which are analytic and also **univalent** on the unit disk. "Univalent" is the complex analyst's term for "one-to-one": $f(z_1) \neq f(z_2)$ unless $z_1 = z_2$. Synonyms for "analytic univalent" include the German **schlicht** (simple, unpretentious) and the Russian **odnolistni** (single-sheeted). These words emphasize a geometric property of a univalent function $w = f(z)$: it maps the disk D in the z-plane one-to-one and onto a domain $f(D)$ in the w-plane. (See FIGURE 1.) By contrast, the *p-valent* function $f(z) = z^p$ maps D onto D, but each image

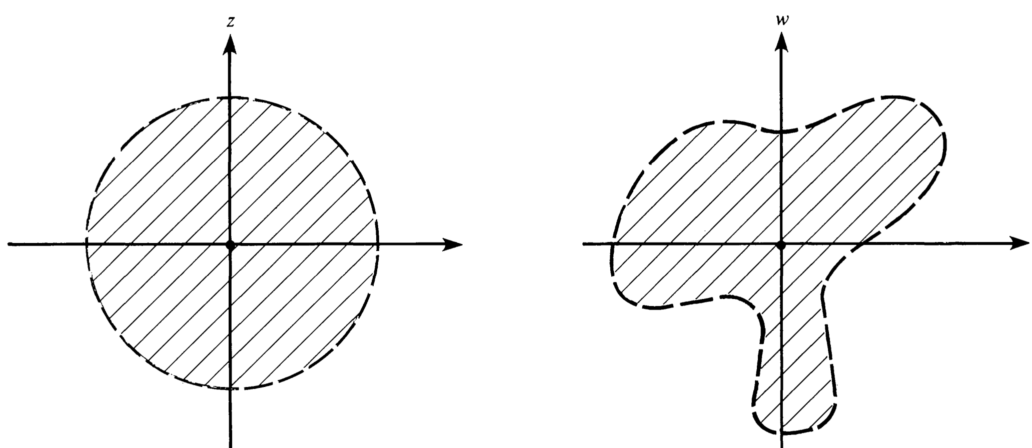


FIGURE 1

point (except $w=0$) has p different preimages. More picturesquely, $f(z)=z^p$ can be viewed as mapping D in the z-plane *univalently* onto a spiral-like surface with p layers ("sheets") hovering above D in the w-plane.

The most important property of univalent analytic functions is the famous **Riemann mapping theorem**, stated in 1851: every proper subdomain of the complex plane that is simply connected (without "holes") is the image of the unit disk under a univalent analytic mapping $f(z)$. The mapping function $f(z)$ is uniquely determined by the domain D, the image point $f(0)$ in D, and the requirement that $f'(0)$ be a positive real number. Thinking of univalent analytic functions as "Riemann mappings," one naturally wonders how analytic properties of f and geometric properties of the image domain $f(D)$ reflect each other. This is the viewpoint of *geometric function theory*; until de Branges' proof, the BC was its main problem.

The BC, as usually stated, is an assertion about a special family of analytic functions on D:

DEFINITION. The **normalized schlicht class**, denoted S, is the family of univalent analytic functions $f: D \to \mathbb{C}$ for which

1. $f(0)=0$
2. $f'(0)=1$.

Conditions 1 and 2 say that an S-function has power series

$$f(z)=z+a_2z^2+a_3z^3+\cdots=z+\sum_{n=2}^{\infty}a_nz^n.$$

The normalizing assumptions simplify the BC's appearance by eliminating irrelevant constants. Fortunately, the normalizations are harmless: if f is *any* univalent analytic function on D, then $g(z)=(f(z)-f(0))/f'(0)$ is in S, and properties of f are easily deduced from those of g. Geometrically, studying g rather than f corresponds to first *translating* the image domain by the vector $f(0)$, *dilating* by the factor $|f'(0)|$, and *rotating* through the angle $\arg(f'(0))$. All of these operations are reversible.

The identity $f(z)=z$ is in one sense a prototype for S: the normalization means that every S-function agrees with the identity up to order one at the origin. The most important nontrivial schlicht function is the Koebe function $k(z)=z/(1-z)^2$, named for the German mathematician P. Koebe, whose achievements include the first correct proof the Riemann mapping theorem. Using the fact that

$$k(z)=z\frac{d}{dz}\left[\frac{1}{1-z}\right],$$

it is easy to see that k can be written as the power series

$$k(z)=z+2z^2+3z^3+4z^4+\cdots,$$

which converges for every z in D. Bieberbach conjectured that k is at the other extreme from the identity function—*every* coefficient of k is as large as possible. Why did Bieberbach guess this almost 70 years before it was proved? As a first step toward understanding k, let's check explicitly that it is univalent on D and find its image. The identity

$$k(z)=\frac{1}{4}\left[\left(\frac{1+z}{1-z}\right)^2-1\right]$$

means that $k(z)$ is composed of the mappings

$$s=\frac{1+z}{1-z}; \qquad t=s^2; \qquad w=\frac{1}{4}(t-1);$$

in that order. The first, a *linear fractional transformation*, maps D univalently onto the right half of the s-plane. The mapping $t=s^2$ is one-to-one *when restricted to the right half-plane*; its image is the entire t-plane minus the nonpositive real axis. The last mapping is simpler yet: a translation

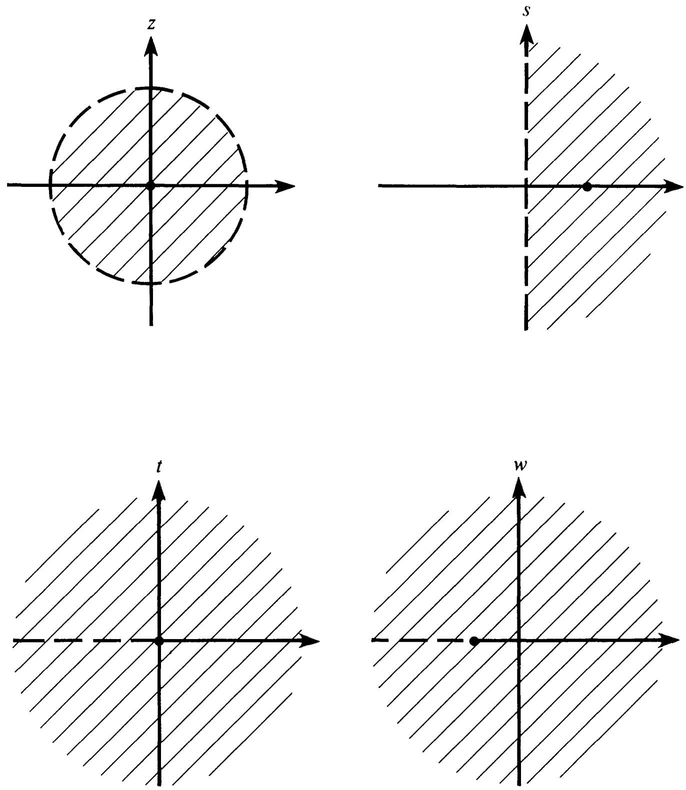


FIGURE 2

one unit to the left followed by a dilation with factor $1/4$. (See FIGURE 2.)

Could the Koebe function be unique among S-functions in having the largest possible coefficients? The answer is "essentially, but not quite": given one S-function, here is a way to construct infinitely many others with coefficients of the same modulus: if $f(z)$ is in S, and α is any real number, let

$$f_\alpha(z) = e^{-i\alpha} f(e^{i\alpha} z).$$

The f_α are called **rotations** of f, because the mapping $z \to e^{i\alpha} z$ is geometrically a counterclockwise rotation of $\mathbb{C}$, about $z = 0$, through α radians. Thus, f_α is formed by preceding and following f with opposite rotations through α radians. (The last operation keeps k_α in S.) In power series notation, if

$$f(z) = z + \sum_{n=2}^{\infty} a_n z^n,$$

then

$$f_\alpha(z) = z + \sum_{n=2}^{\infty} b_n z^n,$$

where $b_n = a_n e^{i(n-1)\alpha}$. Since $|e^{i\alpha}| = 1$, $|a_n| = |b_n|$. The BC, which we can now state precisely, says that the extremal function k is unique "up to rotations":

THE BIEBERBACH CONJECTURE. *If $f(z) = z + \sum_{n=2}^{\infty} a_n z^n$ is in S, then $|a_n| \leqslant n$ for every $n \geqslant 2$. If, for any n, $|a_n| = n$, then f is a rotation of the Koebe function.*

What does the BC mean? Why is it plausible?

Consider this infinite collection of optimization problems: for each $n \geqslant 2$, find in S a function $f(= z + \Sigma_{n=2}^{\infty} a_n z^n)$ for which $|a_n|$ is as large as possible. Two questions occur immediately:

1. Is there, for each n, an absolute bound on $|a_n|$, as f ranges over S?

2. If the nth coefficients of S-functions *are* bounded, is there an S-function f whose nth coefficient a_n *attains* that bound?

The answer to both questions is yes. Our first goal is to understand why (without, of course, assuming the BC).

Why are the nth coefficients of S-functions bounded? As a thought-experiment, consider what restriction, if any, the convergence of the power series $\Sigma a_n z^n$ of an S-function imposes on the a_n. The Cauchy-Hadamard formula for the radius of convergence of a power series implies that for every $R > 1$, $\lim_{n \to \infty} a_n / R^n = 0$. (Briefly, $|a_n| < R^n$ *asymptotically* as $n \to \infty$.) Although the Cauchy-Hadamard condition limits how fast the sequence $\{a_1, a_2, a_3, \dots\}$ can grow, it is no restriction at all on any *particular* coefficient. Every polynomial, for example, is a convergent power series, and the coefficients of polynomials can be any complex numbers at all. Our experiment is over, but we learn that analyticity alone does not explain the boundedness of coefficients of S-functions. We need to look at univalence, a stronger and more subtle property in the complex case than in the real.

To illustrate the difference between the real and complex settings, consider whether there is an interesting "real-variable BC." Let $f(x) = x + \Sigma_{n=2}^{\infty} a_n x^n$ be a one-to-one *real-analytic* function, defined for all x in the real interval $(-1, 1)$; the a_n are *real* constants. Is there any restriction on the size of, say, a_3? The answer is no: if a_3 is any positive number, the polynomial $x + a_3 x^3$ is univalent. In fact, every odd polynomial with positive coefficients is monotone increasing, so no bounds apply to any of the odd-indexed coefficients. *Conclusion*: the BC is essentially a complex result—it has no interesting real-variable analogue.

What, then, is the connection between univalence (of a complex function) and the size of the a_n? A first observation is that the complex monomial z^n is an n-to-one function on D. (Notice how the real monomial x^n on $(-1, 1)$ differs!) Intuitively speaking, the multivalent powers z^n threaten to swamp the univalent term z if the a_n are too large. Here is a proposition that illustrates the general idea.

PROPOSITION. *Let $f(z) = z + a_2 z^2 + a_3 z^3 + \cdots + a_n z^n$ be a polynomial of degree n. If f is univalent in D, then $|a_n| \leqslant 1/n$.*

Proof. Consider $f'(z) = 1 + 2a_2 z + \cdots + n a_n z^{n-1} = n a_n (\frac{1}{n a_n} + \cdots + z^{n-1})$. By the fundamental theorem of algebra, the polynomial within parentheses has $n - 1$ complex roots $c_1, \dots, c_{n-1}$ (some may be repeated). Hence, $f'(z)$ can be factored:

$$f'(z) = n a_n (z - c_1)(z - c_2) \cdots (z - c_{n-1}).$$

Since f is univalent on D, $f'(z)$ has no roots in D (compare the real case!) Therefore each c_i lies outside D; i.e., $|c_i| \geqslant 1$ for each i. Since f is in S,

$$1 = |f'(0)| = |n a_n| \, |c_1| \, |c_2| \cdots |c_{n-1}| \geqslant |n a_n|,$$

as claimed.

The converse is false, but:

Exercise. Let $f(z) = z + a_n z^n$. Show that $f(z)$ is univalent on D if and only if $|a_n| \leqslant 1/n$.

The key to questions 1 and 2 is the fact that S is a *compact* subset of the space of all analytic functions on D. The topology in question is that of **uniform convergence on compacta**: a sequence (f_n) of analytic functions on D converges to an analytic function f if (f_n) converges to f uniformly on every compact subset of D. The functional T that associates to an analytic function

Bieberbach (on the left) with Sierpiński at the Zürich International Congress in 1932.

on D its nth Maclaurin coefficient:

$$T(f) = T\left(\sum a_n z^n\right) = a_n,$$

is continuous in this topology. Therefore, T *attains* a maximum modulus somewhere on S.

We will not attempt a rigorous discussion of compactness in spaces of analytic functions. Roughly speaking, S is compact because it is closed and **locally bounded** (or **normal**). "Closed" means that the limit of a convergent sequence of S-functions is again an S-function. The nontrivial part of this property of S is a standard theorem of A. Hurwitz: the limit of a convergent sequence of schlicht functions is either schlicht or constant. (The normalizations in S rule out the latter possibility.) "Locally bounded" means that for every r in $(0,1)$, there is a positive number $M(r)$ so that $|f(z)| < M(r)$ whenever $|z| < r$ and f is in S. Curiously, the local boundedness of S follows from Bieberbach's second coefficient theorem—a uniform bound on the second coefficient leads to bounds on all the others. This is shown in the next section.

Granted that bounds on the coefficients of S-functions are attained, why should the Koebe function attain them? Here are several ways in which k is the "largest" member of S.

Consider first the image domain $k(D)$ $(=\mathbb{C} - \{x : x \leqslant -1/4\})$. It is as "big" as it can be: adding any open set to $k(D)$ would introduce some overlap, thereby destroying the schlicht property of k. Another extremal property of $k(D)$ has to do with the distance $1/4$ between 0 and the boundary of $k(D)$ in the w-plane. In 1907, Koebe showed that the image $f(D)$ under any mapping in S contains a disk $\{|w| < \rho\}$ of some fixed radius ρ, independent of f. Bieberbach deduced from his second coefficient theorem that (as Koebe conjectured) $1/4$ is the largest possible value of ρ, attained only by k and its rotations. This result, known as **Koebe's one-quarter theorem**, says that the distortion of the domain D after mapping by an S-function is not too severe: the boundary of the image domain cannot approach the origin too closely. Since the boundary of $k(D)$ misses the origin by precisely $1/4$, the Koebe function exhibits the maximum legal distortion. So do rotations $k_\alpha(z) = e^{-i\alpha}k(e^{i\alpha}z)$ of the Koebe function; the image $k_\alpha(D)$ is simply $k(D)$ rotated $-\alpha$ radians about the origin. The symmetry of $k(D)$ is further circumstantial evidence for its extremality. Like D itself (which is the image of D under the *other* extremal S-mapping $f(z) = z$) $k(D)$ exhibits the simplicity and regularity characteristic of extremal objects.

Analytic as well as geometric intuition suggests that we should pick on the Koebe function. For example, the Maclaurin coefficients of a function f are proportional to f's derivatives at the origin. At one extreme—$f(z) = z$—there is no distortion; all higher derivatives vanish. Since for $k(z) = \sum nz^n$ the distortion (in the sense of the one-quarter theorem) is maximal, it is natural to guess that k represents the other extreme.

This section ends with a *caveat*: Informal evidence for the BC is easy to find. The evidence is valid, but the ease of finding it is misleading; rigorous proofs, as history shows, are much harder to find.

The second coefficient

The previous section says—informally—that for each $n \geqslant 2$, the problem of maximizing $|a_n|$ among S-functions has a solution, which is probably k. It is time to prove something. Bieberbach's second-coefficient theorem is the first concrete evidence for the general conjecture. Its corollaries (especially the *distortion theorem*, which implies that S is compact) are the basic tools for further study of the BC. The proof illustrates standard ideas and techniques of univalent function theory and shows how the Koebe functions arise as extrema.

BIEBERBACH'S THEOREM (1916). *If* $f = z + \sum_{n=2}^{\infty} a_n z^n$ *is in* S, *then* $|a_2| \leqslant 2$. *Equality holds if and only if* $f(z)$ *is a rotation of the Koebe function.*

A basic way to obtain Bieberbach-type inequalities is to relate power series coefficients to the (nonnegative) area of some region in the plane. The first such result is the **area theorem**, proved in

1914 by T. H. Gronwall. It refers not to S but to a related class of schlicht functions:

Definition. Let Σ denote the class of functions

$$g(z) = z + b_0 + b_1/z + \cdots = z + \sum_{n=0}^{\infty} b_n z^{-n}$$

that are analytic and univalent in $\Delta = \{z: |z| > 1\}$.

(Σ-functions are normalized to have a simple pole with residue one at infinity.) If g is in Σ, let $E = \mathbb{C} - g(\Delta)$ be the *complement* of the image domain. (See FIGURE 3.)

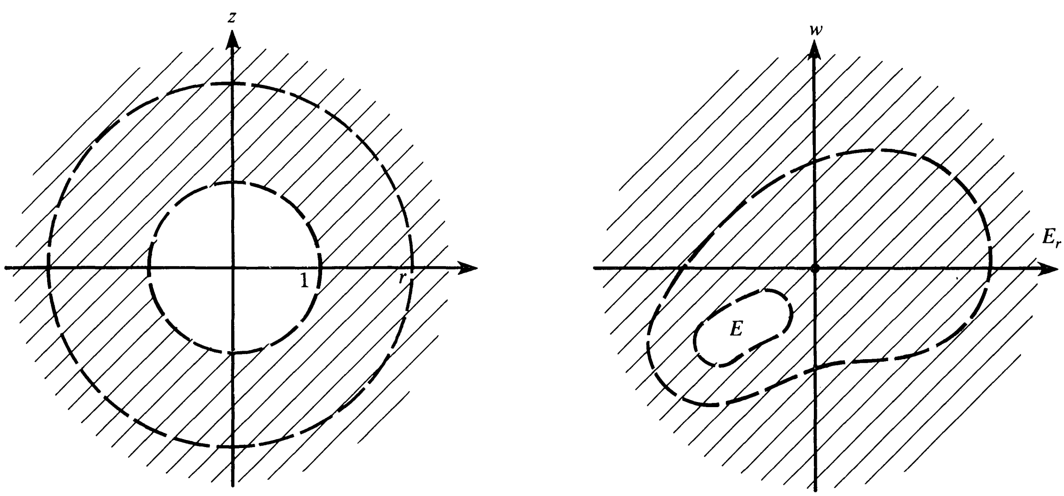


FIGURE 3

One hopes to calculate E's "area" in terms of the b_n. The quotes are necessary because E can be quite irregular. The solution is to approximate E from outside by nice domains $E(r) = \mathbb{C} - \{g(z): |z| > r\}$, and then *define*

$$\text{area } E = \lim_{r \to 1^+} \text{area } E(r).$$

Because the $E(r)$ have smooth (actually, analytic) boundary curves $\gamma(r)$, they have sensible areas which can be computed using Green's theorem in complex form. Explicitly, let $g(z) = w = u + iv$. Then

$$\begin{aligned} \frac{1}{\pi} \text{ area } E(r) &= \frac{1}{\pi} \iint_{E(r)} du\, dv \\ &= \frac{1}{2\pi i} \iint_{E(r)} d\bar{w}\, dw, \end{aligned}$$

since $du\, dv = (1/2i)\, d\bar{w}\, dw$. By Green's theorem,

$$\begin{aligned} \frac{1}{\pi} \text{ area } E(r) &= \frac{1}{2\pi i} \int_{\gamma(r)} \bar{w}\, dw \\ &= \frac{1}{2\pi i} \int_{|z|=r} \bar{g}(z)\, g'(z)\, dz \end{aligned}$$

$$= \frac{1}{2\pi} \int_0^{2\pi} re^{it} g'(re^{it}) \overline{g(re^{it})}\, dt,$$

where the changes of variable $w = g(z)$ and $z = re^{it}$ were made. (Note: the univalence of g was used here!) Writing g and g' as power series,

$$\frac{1}{\pi} \text{ area } E(r) = \frac{1}{2\pi} \int_0^{2\pi} \left(re^{it} + \sum_{n=1}^{\infty} nb_n r^{-n} e^{-int} \right) \left(re^{-it} + \sum_{n=0}^{\infty} \bar{b}_n r^{-n} e^{int} \right) dt$$

$$= r^2 - \sum_{n=1}^{\infty} r^{-2n} n |b_n|^2,$$

by the orthogonality of distinct powers of e^{it}. Since area $E(r) \geqslant 0$, the partial sum

$$\sum_{n=1}^{m} r^{-2n} n |b_n|^2 \leqslant r^2$$

for every $m > 0$. Letting $r \to 1^+$, we have

$$\sum_{n=1}^{m} n |b_n|^2 \leqslant 1, \qquad m = 1, 2, \dots .$$

We have proved:

AREA THEOREM. *If $g = z + \sum_{n=0}^{\infty} b_n z^{-n}$ is in Σ, then $\sum_{n=1}^{\infty} n |b_n|^2 \leqslant 1$.*

COROLLARY. *If $g(z)$ is in Σ, then $|b_1| \leqslant 1$. Equality holds if and only if $g(z) = z + b_0 + e^{i\alpha}/z$ (where $|e^{i\alpha}| = 1$).*

The area theorem is about Σ-functions but it leads indirectly to coefficient estimates for the class S. Beginning with an S-function, one applies algebraic transformations to concoct a Σ-function, keeping track of the coefficients. This is the idea of Bieberbach's proof.

Proof of Bieberbach's theorem. Given $f(z) = z + \sum_{n=2}^{\infty} a_n z^n$ in S, construct the auxiliary functions

$$g(z) = \left(f(z^2)\right)^{1/2} \quad \text{and} \quad h(z) = 1/g(1/z).$$

(The function $f(z^2)$ has a schlicht square root because f is univalent.) A routine calculation shows that $h(z)$ is in Σ, and has Laurent series

$$h(z) = z - \frac{a_2}{2z} + \cdots .$$

The corollary to the area theorem implies that $|a_2| \leqslant 2$; equality holds if and only if $h(z) = z + b/z$, where $|b| = 1$. Unravelling the definition of h in terms of f shows that

$$h(z) = z + \frac{b}{z} \quad \text{if and only if} \quad f(z) = \frac{z}{(1 + bz)^2}.$$

The function $f(z)$ is, as claimed, a rotation of the Koebe function.

The principle observed above—to start with an S- or Σ-function, carry out some algebraic transformation, and then apply a known coefficient theorem to the result—yields several interesting consequences of Bieberbach's theorem. Some of them give as much insight into the BC as the second coefficient theorem itself. A sampler of such results follows.

KOEBE'S ONE-QUARTER THEOREM. *Let $f = z + a_2 z^2 + \cdots$ be in S, and suppose that $f(D)$ omits the value w_0 (i.e., $f(z) \neq w_0$ if $|z| < 1$). Then $|w_0| \geqslant 1/4$; equality can hold if and only if f is a rotation of the Koebe function.*

Proof. The transformed function

$$g(z)=\frac{w_0 f(z)}{w_0-f(z)}=h(f(z)), \quad \text{where } h(w)=\frac{w_0 w}{w_0-w}$$

is univalent because h is (on $\mathbb{C}-\{w_0\}$). Computing g's derivatives at the origin yields

$$g(z)=z+\left(a_2+\frac{1}{w_0}\right)z^2+\cdots.$$

Thus, g is in S and by Bieberbach's theorem, $|a_2+1/w_0|\leqslant 2$. By the triangle inequality, $|w_0|\geqslant 1/4$; equality is possible only if $|a_2|=2$. In that case, f is a rotation of k.

Bieberbach's theorem says that for an S-function $f, |f''(0)|\leqslant 4$. This information can be transferred from the origin to any z_0 in D by composition with the linear fractional transformation

$$A(w)=\frac{w+z_0}{1+w\bar{z}_0}.$$

Because A is a schlicht mapping of D onto itself, with $A(0)=z_0$, the composite $f(A(w))$ is also schlicht, though not normalized. Setting

$$h(w)=\frac{f(A(w))-f(z_0)}{(1-|z_0|^2)f'(z_0)}$$

accomplishes the normalization; a messy (but explicit) calculation shows that

$$h(w)=w+\left[\frac{1}{2}(1-|z_0|^2)\frac{f''(z_0)}{f'(z_0)}-\bar{z}_0\right]w^2+\cdots.$$

From Bieberbach's theorem (dropping the subscript on z),

$$\left|z\frac{f''(z)}{f'(z)}-\frac{2|z|^2}{1-|z|^2}\right|\leqslant\frac{4|z|}{1-|z|^2}, \qquad |z|<1.$$

The last inequality can be "integrated" to give upper and lower bounds on $|f'(z)|$ and $|f(z)|$ in terms of $|z|$:

DISTORTION THEOREM. *If f is in S and $|z|<1$, then*

$$\frac{1-|z|}{(1+|z|)^3}\leqslant|f'(z)|\leqslant\frac{1+|z|}{(1-|z|)^3}; \tag{1}$$

$$\frac{|z|}{(1+|z|)^2}\leqslant|f(z)|\leqslant\frac{|z|}{(1-|z|)^2}. \tag{2}$$

("Distortion" in this context refers to the fact that the mapping f magnifies—distorts—small distances near z by the factor $|f'(z)|$.)

Proof. With $z=re^{i\alpha}$, the inequality before the theorem becomes

$$\frac{4}{1-r^2}\geqslant\left|e^{i\alpha}\frac{f''(re^{i\alpha})}{f'(re^{i\alpha})}-\frac{2r}{1-r^2}\right|=\left|\frac{\partial}{\partial r}\log[(1-r^2)f'(re^{i\alpha})]\right|;$$

the last equality is checked by direct calculation. Now we integrate along the ray from 0 to z; since $f'(0)=1$,

$$|\log[(1-r^2)f'(re^{i\alpha})]|=\left|\int_0^r\frac{\partial}{\partial\rho}[\log[(1-\rho^2)f'(\rho e^{i\alpha})]\,d\rho\right|$$
$$\leqslant\int_0^r\left|\frac{\partial}{\partial\rho}\log[(1-\rho^2)f'(\rho e^{i\alpha})]\right|d\rho$$

$$\leqslant \int_0^r \frac{4}{1-\rho^2}\,d\rho = 2\log\frac{1+r}{1-r}.$$

Taking the real part of $\log[(1-r^2)f'(re^{i\alpha})]$ gives

$$-2\log\frac{1+|z|}{1-|z|} \leqslant \log\left[(1-|z|^2)|f'(z)|\right] \leqslant 2\log\frac{1+|z|}{1-|z|},$$

which is equivalent to (1). The proof of (2) is similar.

The distortion theorems are basic to all further analysis of schlicht functions. The Koebe function plays its usual role: all of the inequalities are strict unless f is a rotation of k.

Now we can settle the *existence* question raised above of bounds for individual coefficients of S-functions. What remains undone from the previous section is to show that S is locally bounded. This is exactly what the right side of (2) means:

$$\text{if } |z| < r, \text{ then } |f(z)| \leqslant k(r) = \frac{r}{(1-r)^2}.$$

At last we know that coefficient bounds exist. One can also derive (admittedly unprepossessing) coefficient *estimates* from (2):

Exercise. Show that $|a_n| \leqslant n^2e^2/4$, $n = 2, 3, \ldots,$ for every f in S.

Hint. Use **Cauchy's inequalities**: if $f(z) = \Sigma a_n z^n$ is analytic and $|f(z)| \leqslant M$ for $|z| < r$, then $|a_n| \leqslant M/r^n$.

We will do much better for all coefficients in the next section. As a last corollary of Bieberbach's theorem, though, let's improve the a_3 estimate (the exercise gives $|a_3| \leqslant 16.63$).

PROPOSITION. *If f is in S, then $|a_3| \leqslant 5$.*

Proof. The function $g(z) = 1/f(1/z) = z - a_2 + (a_2^2 - a_3)z^{-1} + \cdots$ is in Σ. By the area theorem, $|a_2^2 - a_3| \leqslant 1$, so

$$|a_3| \leqslant |a_2^2 - a_3| + |a_2|^2 \leqslant 1 + 4 = 5.$$

Some history, two classical proofs, and de Branges' solution

The Riemann mapping theorem guarantees the *existence* of a schlicht mapping from D onto any given simply connected proper subdomain of the complex plane. Univalent function theory is the study of *concrete* properties (especially extremal properties) of "Riemann maps." The primeval fact is that S is *normal*; as discussed above, this assures that extremal problems (e.g., to maximize the nth coefficient) have solutions. This fundamental theorem was first proved in 1907 by Koebe, arguably the father of the discipline.

That the Koebe function plays some special role in S was evident before the BC appeared in 1916. For example, in 1914 Gronwall proved (among other properties of k) that if f is in S and $|z| < 1$, then

$$|k(-|z|)| \leqslant |f(z)| \leqslant |k(|z|)|,$$

which means that on every circle $|z| = r$, $|k(z)|$ attains both the largest maximum and the smallest minimum of any S-function. (This is also a corollary of the distortion theorem.) With so much evidence for the extremality of k in S, the conjecture that k also maximizes individual coefficients was certainly in the wind. Bieberbach's contribution is more in having proved the second coefficient theorem and its corollaries than in having issued the conjecture itself. (See also [**11**], ch. 2.)

Progress on the BC occurred in several directions. Here are three "genres" of partial results:

1. $|a_n| \leqslant n$ for specific n;
2. $|a_n| \leqslant n$ for subclasses of S;
3. $|a_n| \leqslant Cn$ for sufficiently large C.

(See [3] and [7] for progress in other directions.)

Results of the first type came slowly. The third coefficient theorem—$|a_3| \leqslant 3$—was proved in 1923 by the Czech-educated mathematician K. Loewner, who later emigrated to the United States. The proof is deep, delicate, and completely different from Bieberbach's second-coefficient proof. Loewner's partial differential equation method is notable both for having been found so early and because it figures in de Branges' proof of the general conjecture. No more "$|a_n| \leqslant n$" theorems were proved for more than 30 years, and then by still different methods. In 1955, P. Garabedian and M. Schiffer, who were then Loewner's colleagues at Stanford, used a specially developed calculus of variations in S to prove that $|a_4| \leqslant 4$. The sixth- and fifth-coefficient theorems followed in 1968 and 1972. By 1984, earlier complicated proofs had been revised and shortened, but the conjecture remained open for all n greater than six.

Two early "subclass" theorems, due to R. Nevanlinna (1920) and to W. Rogosinski and J. Dieudonné (independently, around 1930), respectively, assert that the BC holds for S-functions with either (i) starlike range ($f(D)$ contains the segment joining any of its points to the origin) or (ii) all real coefficients. (Loewner had proved a special case of (i) in 1917.) In a slightly different direction, certain subclasses of S were shown to satisfy more stringent coefficient growth estimates. For example, J. Clunie and Ch. Pommerenke showed in 1966 that the coefficients of *bounded* S-functions satisfy the order-of-growth estimate

$$|a_n| = O(n^{-1/2-\alpha})$$

for some positive number α (of unknown best value). Other recent work showed that $|a_n| \leqslant n$ for S-functions for which (iii) $|a_2| \leqslant 1.05$; (iv) f is "near" the Koebe function in an appropriate topology; or (v) f is "far" from k. As an example, we prove (ii).

PROPOSITION. *Let $f(z) = z + a_2 z^2 + a_3 z^3 + \cdots$ be in S, with each a_n real. Then $|a_n| \leqslant n$.*

Proof. Because the a_n are real, the identity

$$f(\bar{z}) = \sum a_n \bar{z}^n = \sum \overline{a_n z^n} = \overline{f(z)}$$

holds for all z in D. It follows that the image domain $f(D)$ is symmetric about the real axis; in particular, $f(z)$ is real if and (because f is univalent) only if z is real. Moreover, since $f'(0) = 1$, f maps the upper (resp. lower) half of D to the upper (resp. lower) half of $f(D)$. (See FIGURE 4.)

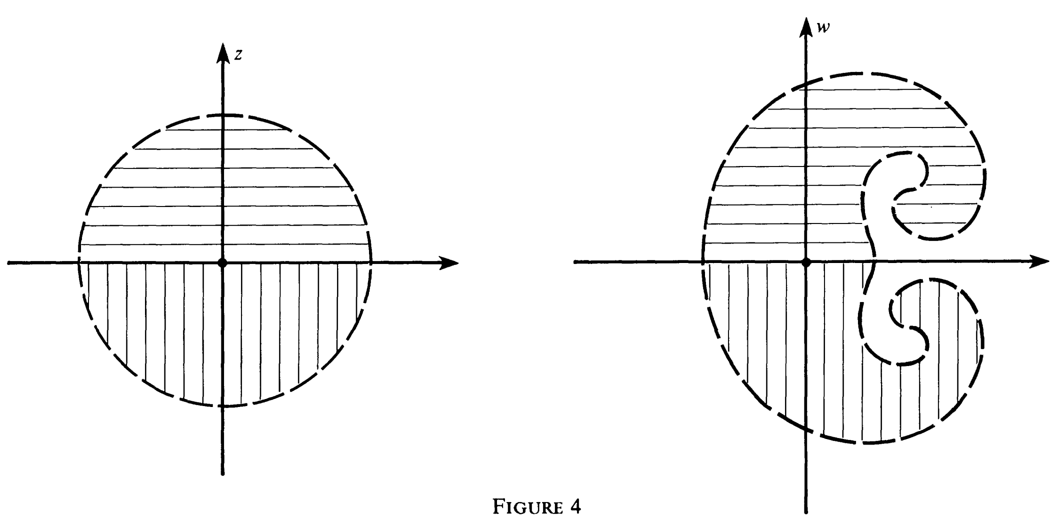


FIGURE 4

Writing f as a power series in polar coordinates (with $a_1 = 1$), we have

$$\begin{aligned} f(z) = f(re^{i\alpha}) &= \sum_{k=1}^{\infty} a_k r^k e^{ik\alpha} \\ &= \sum a_k r^k \cos(k\alpha) + i\sum a_k r^k \sin(k\alpha) \\ &= U(z) + iV(z), \end{aligned}$$

where U and V are *real-valued* functions of z. Recall from elementary calculus that if n and m are integers, functions of either form $\sin(n\alpha)$ or $\cos(m\alpha)$ are pairwise orthogonal with respect to integration in α over the interval $[-\pi, \pi]$, but that

$$\int_{-\pi}^{\pi} [\sin(n\alpha)]^2 \, d\alpha = \pi.$$

Hence, for each $r < 1$,

$$|a_n r^n| = \frac{1}{\pi}\left|\int_{-\pi}^{\pi} \sin(n\alpha) V(re^{i\alpha}) \, d\alpha\right|.$$

Because the integrand is an *even* function of α,

$$\begin{aligned} |a_n r^n| &= \frac{2}{\pi}\left|\int_0^{\pi} \sin(n\alpha) V(re^{i\alpha}) \, d\alpha\right| \\ &\leqslant \frac{2}{\pi}\int_0^{\pi} |\sin(n\alpha) V(re^{i\alpha})| \, d\alpha \\ &= \frac{2}{\pi}\int_0^{\pi} |\sin(n\alpha)| V(re^{i\alpha}) \, d\alpha; \end{aligned}$$

the last equality follows from the mapping property of f—$V(z)$ is nonnegative if α lies in $[0, \pi]$. Now we invoke a possibly unfamiliar but straightforward inequality of real analysis (see, e.g., [**12**, p. 356] for a proof):

$$|\sin(n\alpha)| \leqslant n \sin(\alpha), \qquad 0 \leqslant \alpha \leqslant \pi, \qquad n = 1, 2, \dots.$$

Given this,

$$\begin{aligned} |a_n r^n| &\leqslant \frac{2}{\pi} n \int_0^{\pi} \sin(\alpha) V(re^{i\alpha}) \, d\alpha \\ &= \frac{1}{\pi} n \int_0^{2\pi} \sin(\alpha) V(re^{i\alpha}) \, d\alpha \\ &= na_1 = n. \end{aligned}$$

Because $r < 1$ is arbitrary, we can let r tend to 1 from below to complete the proof.

The order of growth of the a_n was known early to be linear, as conjectured. The first good theorem of type $|a_n| \leqslant Cn$ appeared in 1925, when J. E. Littlewood proved that for f in S,

$$|a_n| \leqslant en.$$

The main ingredient is Littlewood's integral estimate (see [**8**, p. 38]): if $0 < r < 1$ and f is in S, then

$$\frac{1}{2\pi}\int_0^{2\pi} |f(re^{i\alpha})| \, d\alpha \leqslant \frac{r}{1-r}.$$

Assuming this, the proof is straightforward. Let $f(z) = \sum a_n z^n$; then

$$a_n = \frac{1}{2\pi i}\int_{|z|=r} \frac{f(z)}{z^{n+1}} \, dz.$$

M. M. Schiffer, S. Bergman, M. Protter, J. Herriot and C. Loewner at Stanford in the 1950's. Schiffer and Loewner played important roles in the history of the Bieberbach problem.

(This simplest version of the Cauchy integral theorem can be checked by direct computation.) Writing $z = re^{i\alpha}$ and $dz = ire^{i\alpha}\,d\alpha$,

$$|a_n| = \frac{1}{2\pi}\left|\int_0^{2\pi}\frac{f(re^{i\alpha})}{r^n e^{in\alpha}}\,d\alpha\right|$$

$$\leqslant \frac{1}{2\pi r^n}\int_0^{2\pi}|f(re^{i\alpha})|\,d\alpha \leqslant \frac{1}{r^{n-1}-r^n}.$$

This inequality holds for *every* r in $(0,1)$. Elementary calculus shows that the right side attains its minimum when $r = 1 - 1/n$. Substituting this value of r into the last inequality yields

$$|a_n| \leqslant \frac{1}{r^{n-1}-r^n} = n\left(1+\frac{1}{n-1}\right)^{n-1} < en.$$

In 1974, A. Baernstein improved Littlewood's integral inequality by showing that k has the largest possible integral mean of every order: for each f in S and each real number p,

$$\int_0^{2\pi}|f(re^{i\alpha})|^p\,d\alpha \leqslant \int_0^{2\pi}|k(re^{i\alpha})|^p\,d\alpha.$$

With $p = 1$, calculating the right side gives

$$\int_0^{2\pi}|f(re^{i\alpha})|\,d\alpha \leqslant \frac{r}{1-r^2}.$$

Exercise. Mimic the previous proof to show that

$$|a_n| \leqslant \frac{e}{2}n \approx 1.36\,n.$$

Given any S-function f, setting $g(z) = f(z^2)^{1/2} = z + \sum_{k=1}^{\infty} b_{2k+1}z^{2k+1}$ produces an *odd* univalent function. (We did this already in the proof of Bieberbach's theorem.) If $f = k$, then

$$k(z^2)^{1/2} = z + z^3 + z^5 + z^7 + \cdots,$$

i.e., $b_k = 1$ for all k. In 1932, J. Littlewood and R. Paley made the natural conjecture that $|b_k| \leqslant 1$ for every odd S-function. (They *proved* that $|b_k| < 14$.)

Exercise. By inverting the square-root transform, show that the Littlewood-Paley conjecture implies the BC.

The Littlewood-Paley conjecture was disproved the following year by M. Fekete and G. Szegő, but it led M. Robertson to the slightly weaker conjecture that $|b_k| \leqslant 1$ in an average sense:

ROBERTSON CONJECTURE (1936). *If* $g(z) = z + b_3z^3 + b_5z^5 + \cdots$ *is in* S, *then* $1 + |b_3|^2 + |b_5|^2 + \cdots + |b_{2n-1}|^2 \leqslant n$.

The Robertson conjecture still implies the BC, even in a strong form (see, e.g., [**8**, p. 66]). It is what de Branges indirectly proved.

Iterating the square-root transform leads to looking at roots of higher order, and in a limiting sense, to the logarithm of a schlicht function. In 1939, H. Grunsky obtained a new class of coefficient inequalities that would prove, much later, to be an important step toward a general solution of the BC. Beginning with an S-function f, Grunsky studied the coefficients c_{jk} of the function

$$\log\frac{f(z)-f(w)}{z-w} = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} c_{jk}z^jw^k,$$

analytic in the *two* complex variables z and w. The **Grunsky inequalities** are an infinite family of estimates on the "logarithmic coefficients" c_{jk}. (Formally, they assert that the infinite matrices

(c_{jk}) associated to S-functions are bounded by one in a suitable norm.)

Given the Grunsky inequalities, a difficult question remains: how to use knowledge of the logarithmic coefficients c_{jk} to estimate the ordinary Taylor coefficients a_n of f, i.e., how to "exponentiate" the Grunsky inequalities. Different techniques were developed in the sixties and seventies by N. A. Lebedev and I. M. Milin in the USSR and by C. FitzGerald in the US. They were applied to improve earlier "$|a_n| \leqslant Cn$" theorems; by 1978, D. Horowitz had shown $|a_n| < 1.0657n$. (See [**13**] for a chronology of efforts to reduce C.)

Lebedev and Milin's work led to a conjecture of exceptional interest:

LEBEDEV-MILIN CONJECTURE (1967). *If f is in S and $\log f(z)/z = \sum_{k=1}^{\infty} c_k z^k$, then*

$$\sum_{m=1}^{n} \sum_{k=1}^{m} \left(k|c_k|^2 - \frac{4}{k} \right) \leqslant 0, \qquad n = 1,2,3,\dots.$$

Since for the Koebe function, $\log k(z)/z = \sum_{k=1}^{\infty}(2/k)z^k$, the obvious guess is that for every f in S, $|c_k| \leqslant 2/k$ for every k. This is too optimistic—it can be shown to imply the Littlewood-Paley conjecture. The Lebedev-Milin conjecture, like the Robertson conjecture, says that, nevertheless, the obvious guess is true in a weaker, average sense for the first n coefficients. Lebedev and Milin showed that their conjecture implies the Robertson conjecture, which is stronger than the BC. Thus some experts were surprised that the Lebedev-Milin conjecture is exactly what de Branges proved.

De Branges' proof has two main ingredients—a system of time-dependent weight functions, and the Loewner differential equation (part of the classical third coefficient proof), which also involves a time parameter. De Branges combines these to construct a time-dependent form of the Lebedev-Milin inequality which is clearly valid at time $t = \infty$ and which *equals* the Lebedev-Milin inequality at $t = 0$. The brilliant (and "miraculous," according to one expert) choice of the weighting functions assures that the inequality remains valid as t retreats to zero. The necessary property of the weight functions reduces to an inequality on generalized hypergeometric polynomials.

The computer played a role in de Branges' search for a solution, though not in his ultimate proof. At a late stage in his work, de Branges approached W. Gautschi of the computer science department at Purdue for help in testing his conjectured inequality numerically. The spectacular results—proofs of the BC for the first thirty coefficients—were encouraging evidence that de Branges' general method would work. They also led Gautschi to call R. Askey, at the University of Wisconsin, to inquire whether the inequality had been proven earlier. Surprisingly, it had, in a 1976 paper of Askey and G. Gasper [**1**]. By remarkable coincidence, a property of special functions of a *real* variable finished the proof of a *complex* result.

De Branges' originally lengthy proof was at first received skeptically. Incorrect proofs of the BC had been announced before. Nevertheless, the new proof was confirmed by the Leningrad Seminar in Geometric Function Theory (of which I. M. Milin is a member), in five marathon sessions in April and May, 1984. A preliminary version of the proof was issued in preprint form [**6**]. The proof was later revised and shortened; now, some of the individual coefficient proofs are longer and more difficult. De Branges' version was published in 1985 [**5**]. (See also [**15**], [**10**], and [**9**] for a slightly different version of the proof and for more detailed commentary.)

Mathematicians working on the BC created and advanced theories (including Schiffer's variational method and the theory of quadratic differentials) which, though not necessarily used in de Branges' solution, have found applications elsewhere. This is not to say that de Branges does not use earlier work. Loewner's method, the Lebedev-Milin conjecture (and, implicitly, the Grunsky inequalities), and Askey and Gasper's theorem are all essential—along with de Branges' unique contribution—to the ultimate solution.

Though its most famous problem has been solved, important questions in geometric function theory remain. For example, sharp coefficient estimates are unknown for normalized p-to-one analytic functions, for Σ-functions, and for S-functions with growth restrictions. De Branges' contribution was celebrated at an international symposium held at Purdue in March, 1985; many new problems and directions for research were proposed. At the University of Maryland, 1985–1986 was declared a special year in complex analysis. How and where de Branges' special techniques may apply remains to be seen, but his achievement seems certain to spur new work and interest in special functions, optimization theory, functional analysis, and complex analysis.

References

[1] R. Askey and G. Gasper, Positive Jacobi polynomial sums II, Amer. J. Math., 98 (1976) 709–737.

[2] S. Bernardi, Bibliography of Schlicht Functions, Mariner, Tampa, 1983.

[3] A. Baernstein, Bieberbach's conjecture for tourists, Lecture Notes in Math., vol. 908, Springer-Verlag, Berlin, 1982, 48–72.

[4] L. Bieberbach, Uber die Koeffizienten derjenigen Potenzreihen, welche eine schlichte Abbildung des Einheitskreises vermitteln, Sitzungsberichte Preussische Akademie der Wissenschaften, 1916, 940–955.

[5] L. de Branges, A proof of the Bieberbach conjecture, Acta Math., 154 (1985) 137–152.

[6] ______, A proof of the Bieberbach conjecture, Steklov Math. Inst., preprint E-5-84, Leningrad, 1984, 9–21.

[7] P. L. Duren, Coefficients of univalent functions, Bull. Amer. Math. Soc., 83 (1977) 891–909.

[8] ______, Univalent Functions, Springer-Verlag, New York, 1983.

[9] C. FitzGerald, The Bieberbach conjecture: retrospective, Notices Amer. Math. Soc., 32(1985) 2–6.

[10] ______, Ch. Pommerenke, The de Branges theorem on univalent functions, Trans. Amer. Math. Soc., 290 (1985) 683–690.

[11] A. W. Goodman, Univalent Functions, Vols. I, II, Mariner, Tampa, 1983.

[12] E. Hille, Analytic Function Theory, Vol. II, Chelsea, New York, 1962.

[13] D. Horowitz, A further refinement for coefficient estimates of univalent functions, Proc. Amer. Math. Soc., 71 (1978) 217–219.

[14] Ch. Pommerenke, Univalent Functions, Vandenhoeck and Ruprecht, Göttingen, 1975.

[15] ______, The Bieberbach conjecture, Math. Intelligencer, 7 (no. 2, 1985).

A Generalization of the Pythagorean Theorem Seen as a Problem of Equivalent Resistances

G. D. CHAKERIAN
University of California-Davis
Davis, CA 95616

M. S. KLAMKIN
University of Alberta
Edmonton, Alberta T6G 2G1 Canada

The following result (observed several years ago by the first author) may be viewed as a generalization of the Pythagorean Theorem. Let squares be erected outwardly on the sides of an acute-angled triangle ABC. The extended altitudes of the triangle cut the squares into rectangles, as in FIGURE 1. The theorem is that *rectangles with the same labels in the figure have the same areas*. The Pythagorean Theorem is obtained in the limiting case when A is a right angle, in which case the rectangles I degenerate to line segments.

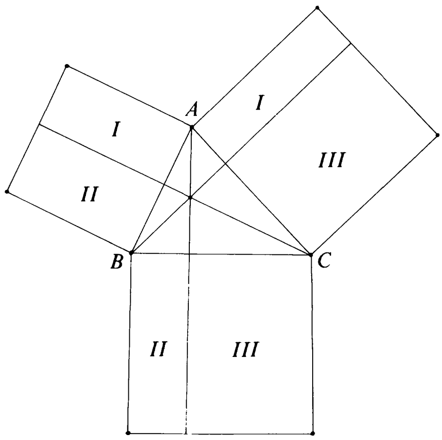


FIGURE 1

Let the triangle have sides a, b, c opposite the angles A, B, C, respectively. The theorem is immediate upon observing that, for example, the rectangle labelled I on side c has sides of lengths c and $b\cos A$, while the other rectangle labelled I has sides of lengths b and $c\cos A$. Thus (using the same symbol to represent both the region and its area) both rectangles have area

$$I = bc\cos A. \tag{1}$$

The result can also be proved synthetically exactly along the same lines as Euclid's proof of the Pythagorean Theorem.

One finds the theorem in the problem book of Larson [**3**, 8.3.17]. It is also implicit (though not explicitly stated) in a paper of Brown and Walter [**1**] (we are indebted to Lester Lange for this reference).

If Δ denotes the area of triangle ABC, then from $2\Delta = bc\sin A$ and (1) we obtain $I = 2\Delta\cot A$. This together with the corresponding relations for the other rectangles gives the following proportionality:

$$I : II : III = \cot A : \cot B : \cot C. \tag{2}$$

We will find later use for the quantities $\sigma_1, \sigma_2, \sigma_3$ defined as follows:

$$\sigma_1 = -a^2 + b^2 + c^2, \qquad \sigma_2 = a^2 - b^2 + c^2, \qquad \sigma_3 = a^2 + b^2 - c^2. \tag{3}$$

Under the assumption that ABC is acute-angled, $\sigma_1, \sigma_2, \sigma_3$ are positive real numbers. Indeed, applying the law of cosines in (1), we see that $2I = 2bc\cos A = -a^2 + b^2 + c^2 = \sigma_1$. Similarly $2II = \sigma_2$ and $2III = \sigma_3$. Thus we also have

$$I : II : III = \sigma_1 : \sigma_2 : \sigma_3. \tag{4}$$

In this note we ask whether the property in FIGURE 1 can be extended to other sets of concurrent cevians in the following way. Let AA_1, BB_1, CC_1 be cevians of the acute-angled triangle ABC concurrent in a point P. Let the squares on the sides of the triangle be cut into rectangles as indicated in FIGURE 2. For which points P is it true that $I = I'$, $II = II'$, and $III = III'$?

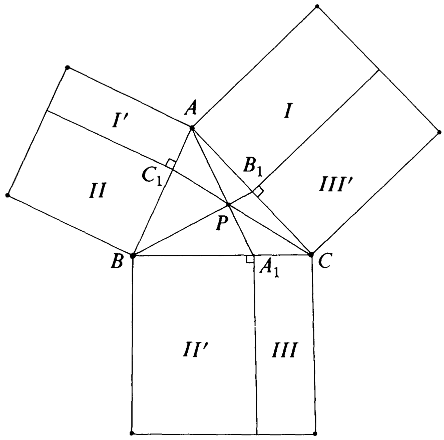


FIGURE 2

Our main result is the following theorem, which characterizes the orthocenter as the unique point having the required property.

THEOREM. *If the squares on the sides of an acute-angled triangle are cut into rectangles as depicted in* FIGURE 2, *then* $I = I'$, $II = II'$, *and* $III = III'$ *if and only if* P *is the orthocenter of the triangle.*

We give two proofs of the theorem. The first is a nice synthetic proof due to R. S. Luthar. The second proof, as we shall see, depends on the uniqueness of the solution of a certain set of three simultaneous nonlinear equations. Gilbert and Shepp [**2**] have dealt with the generalization of these same equations to n variables, in the context of a problem on equivalent resistances. They give an explicit solution in the case $n=3$. We shall derive essentially this solution in the course of the second proof.

Synthetic proof of theorem (R. S. Luthar). Let Q be the orthocenter of triangle ABC. If Q is interior to the triangle BPC_1, as it would be in FIGURE 2, then it is "below" the cevian CC_1. Hence the extended altitude from C will cut from the square on AB an upper rectangle larger than I'. Similarly Q is "above" the cevian BB_1. Hence the extended altitude from B will cut from the square on AC an upper rectangle smaller than I. Thus, since the upper rectangles cut off by the altitudes have equal area, we must have $I' < I$. Since Q cannot belong simultaneously to all the cevians determined by P if $P \neq Q$, one sees in a similar way that in every case one of the three conditions $I = I'$, $II = II'$, $III = III'$, does *not* hold if $P \neq Q$. In case $P = Q$, we have seen in the introduction that the three conditions hold, so the theorem is proved.

Analytic proof of theorem. With the situation as depicted in FIGURE 2, we represent $\mathbf{P}$ in terms of barycentric coordinates, $\mathbf{P} = x\mathbf{A} + y\mathbf{B} + z\mathbf{C}$, where $x, y, z \geqslant 0$ and $x + y + z = 1$. Then we have

$$\mathbf{A}_1 = \frac{y\mathbf{B} + z\mathbf{C}}{y+z}, \qquad \mathbf{B}_1 = \frac{x\mathbf{A} + z\mathbf{C}}{x+z}, \qquad \mathbf{C}_1 = \frac{x\mathbf{A} + y\mathbf{B}}{x+y}. \tag{5}$$

Since $AB_1 = zb/(x+z)$ and $AC_1 = yc/(x+y)$, we obtain $I = zb^2/(x+z)$ and $I' = yc^2/(x+y)$. The condition $I = I'$ is then equivalent to $zb^2/(x+z) = yc^2/(x+y)$, or $z(x+y)/c^2 = y(x+z)/b^2$. Dealing in the same way with the other rectangles, we see that the conditions $I = I'$, $II = II'$, and $III = III'$ are equivalent to

$$x(y+z) : y(x+z) : z(x+y) = a^2 : b^2 : c^2. \tag{6}$$

In terms of the quantities $\sigma_1, \sigma_2, \sigma_3$ defined by (3), this is equivalent to

$$x(y+z) : y(x+z) : z(x+y) = \sigma_2 + \sigma_3 : \sigma_1 + \sigma_3 : \sigma_1 + \sigma_2. \tag{7}$$

We now proceed to show that under the assumption $\sigma_1, \sigma_2, \sigma_3 > 0$, the system (7) is equivalent to

$$x : y : z = \frac{1}{\sigma_1} : \frac{1}{\sigma_2} : \frac{1}{\sigma_3}. \tag{8}$$

Assume first that x, y, z are positive numbers satisfying (8). Then there exists $\lambda > 0$ such that

$$\sigma_1 = \frac{\lambda}{x}, \qquad \sigma_2 = \frac{\lambda}{y}, \qquad \sigma_3 = \frac{\lambda}{z}. \tag{9}$$

This gives immediately

$$\sigma_1 + \sigma_2 = \left(\frac{\lambda}{xyz}\right) z(x+y), \qquad \sigma_1 + \sigma_3 = \left(\frac{\lambda}{xyz}\right) y(x+z), \qquad \sigma_2 + \sigma_3 = \left(\frac{\lambda}{xyz}\right) x(y+z), \tag{10}$$

which implies (7).

On the other hand, suppose x, y, z are positive numbers satisfying (7). Then for some $\mu > 0$ we have

$$x(y+z) = \mu(\sigma_2 + \sigma_3), \qquad y(x+z) = \mu(\sigma_1 + \sigma_3), \qquad z(x+y) = \mu(\sigma_1 + \sigma_2). \tag{11}$$

Subtracting the third equation from the sum of the first two equations in (11) gives $xy = \mu\sigma_3$, which can be rewritten as $z = xyz/\mu\sigma_3$. We obtain corresponding expressions for x and y in the same way, giving

$$x = \left(\frac{xyz}{\mu}\right)\frac{1}{\sigma_1}, \qquad y = \left(\frac{xyz}{\mu}\right)\frac{1}{\sigma_2}, \qquad z = \left(\frac{xyz}{\mu}\right)\frac{1}{\sigma_3}, \tag{12}$$

and so (8) follows.

Thus $I = I'$, $II = II'$, and $III = III'$ in FIGURE 2 if and only if P has barycentric coordinates x, y, z satisfying (8). With $\tau_i = 1/\sigma_i$, $i = 1,2,3$, this is equivalent to

$$x = \frac{\tau_1}{\tau_1 + \tau_2 + \tau_3}, \qquad y = \frac{\tau_2}{\tau_1 + \tau_2 + \tau_3}, \qquad z = \frac{\tau_3}{\tau_1 + \tau_2 + \tau_3}. \tag{13}$$

Thus there is at most one point P with the required property. This fact, together with our earlier observation that the property holds when P is the orthocenter, proves Theorem 1.

It is of course a consequence of our analytic proof of Theorem 1 that the barycentric coordinates x, y, z of the orthocenter of ABC satisfy (8), or equivalently (13). We can see this directly, starting with the well-known fact that for the orthocenter

$$x : y : z = \tan A : \tan B : \tan C. \tag{14}$$

Indeed (14) follows from $z/y = BA_1/CA_1 = \tan C/\tan B$ and the corresponding relations for the other ratios when P is the orthocenter. Comparing (14) with (2) we see that

$$x : y : z = \frac{1}{I} : \frac{1}{II} : \frac{1}{III}. \tag{15}$$

Then (15) and (4) give (8).

A problem of equivalent resistances

Consider three resistances $r_1, r_2, r_3 > 0$ connected in a triangular network as in FIGURE 3. Since

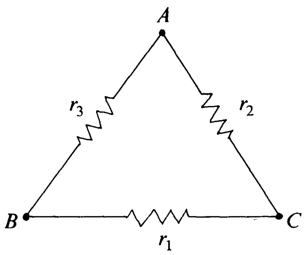


FIGURE 3

r_3 and r_2 are connected *in series* from B to C, they present a resistance to the flow of electric current from B to C equivalent to $r_3 + r_2$. Hence the given network presents a resistance to the flow of current from B to C equivalent to that of a resistor of magnitude $r_3 + r_2$ and a resistor of magnitude r_1 connected *in parallel* between B and C. From elementary physics we know this is equivalent to a resistance of magnitude

$$\left(\frac{1}{r_1} + \frac{1}{r_2 + r_3}\right)^{-1},$$

which we shall call the *equivalent resistance* between B and C.

Suppose now that the equivalent resistance between B and C is a^2, between C and A is b^2, and between A and B is c^2, where a, b, c are some nonzero real numbers. What we want to show is that r_1, r_2, r_3 *are uniquely determined by the equivalent resistances* a^2, b^2, c^2.

To see this, note that we are given

$$\frac{1}{r_1} + \frac{1}{r_2 + r_3} = \frac{1}{a^2}, \qquad \frac{1}{r_2} + \frac{1}{r_1 + r_3} = \frac{1}{b^2}, \qquad \frac{1}{r_3} + \frac{1}{r_1 + r_2} = \frac{1}{c^2}. \tag{16}$$

Simplifying the left-hand side of each equation leads to

$$r_1(r_2+r_3) : r_2(r_1+r_3) : r_3(r_1+r_2) = a^2 : b^2 : c^2. \tag{17}$$

But this is exactly the system (6) that we encountered in the geometric problem. To apply the analysis we applied to (6), we first verify that certain quantities are nonzero. Let $\sigma_1, \sigma_2, \sigma_3$ be defined as in (3). From (17) we have, for some $\lambda > 0$,

$$r_1(r_2+r_3) = \lambda a^2, \qquad r_2(r_1+r_3) = \lambda b^2, \qquad r_3(r_1+r_2) = \lambda c^2. \tag{18}$$

Subtracting each equation from the sum of the other two gives

$$2r_2r_3 = \lambda\sigma_1, \qquad 2r_1r_3 = \lambda\sigma_2, \qquad 2r_1r_2 = \lambda\sigma_3. \tag{19}$$

This shows that all the σ_i are positive, so the same analysis that led us from (6) to (8) can be applied to give, with $\tau_i = 1/\sigma_i$, $i = 1,2,3$,

$$r_1 : r_2 : r_3 = \tau_1 : \tau_2 : \tau_3. \tag{20}$$

Thus $r_i = \mu\tau_i$, $i = 1,2,3$, for some $\mu > 0$. The multiplier μ is determined by substituting into (16). We then obtain, with $T = \tau_1 + \tau_2 + \tau_3$,

$$r_1 = \frac{a^2T}{\tau_2+\tau_3}, \qquad r_2 = \frac{b^2T}{\tau_1+\tau_3}, \qquad r_3 = \frac{c^2T}{\tau_1+\tau_2}. \tag{21}$$

This shows that the r_i are uniquely determined by a^2, b^2, c^2, as claimed.

Gilbert and Shepp [**2**] showed that in the generalization to n resistances $r_1, \ldots, r_n$ on the sides of an n-gon, the r_i are uniquely determined by the equivalent resistances between consecutive vertices.

References

[1] S. I. Brown and M. I. Walter, What-if-not? An elaboration and second illustration, Mathematics Teaching, 51 (1970) 9–17.

[2] E. N. Gilbert and L. A. Shepp, A resistance problem, SIAM Review, 27 (1985) 449–452.

[3] L. C. Larson, Problem-Solving Through Problems, Springer-Verlag, New York, 1983.

Some Remarks About Bridge Probabilities

P. F. Zweifel

Virginia Polytechnic Institute and State University

Blacksburg, VA 24061

This paper is dedicated to the memory of Mark Kac, great mathematician and probabilist.

Probability theory is basic in many areas of physics and engineering, for example, quantum theory, statistical mechanics, and risk-benefit analysis. I have found it valuable in my teaching to go back to some very basic concepts and to discuss them in detail with the students, hoping that the proper building blocks will be available for constructing the more complicated situations necessary to describe, for instance, physical systems in statistical mechanics. Since my avocation

$$r_1(r_2+r_3) : r_2(r_1+r_3) : r_3(r_1+r_2) = a^2 : b^2 : c^2. \tag{17}$$

But this is exactly the system (6) that we encountered in the geometric problem. To apply the analysis we applied to (6), we first verify that certain quantities are nonzero. Let $\sigma_1, \sigma_2, \sigma_3$ be defined as in (3). From (17) we have, for some $\lambda > 0$,

$$r_1(r_2+r_3) = \lambda a^2, \qquad r_2(r_1+r_3) = \lambda b^2, \qquad r_3(r_1+r_2) = \lambda c^2. \tag{18}$$

Subtracting each equation from the sum of the other two gives

$$2r_2r_3 = \lambda\sigma_1, \qquad 2r_1r_3 = \lambda\sigma_2, \qquad 2r_1r_2 = \lambda\sigma_3. \tag{19}$$

This shows that all the σ_i are positive, so the same analysis that led us from (6) to (8) can be applied to give, with $\tau_i = 1/\sigma_i$, $i = 1,2,3$,

$$r_1 : r_2 : r_3 = \tau_1 : \tau_2 : \tau_3. \tag{20}$$

Thus $r_i = \mu\tau_i$, $i = 1,2,3$, for some $\mu > 0$. The multiplier μ is determined by substituting into (16). We then obtain, with $T = \tau_1 + \tau_2 + \tau_3$,

$$r_1 = \frac{a^2T}{\tau_2+\tau_3}, \qquad r_2 = \frac{b^2T}{\tau_1+\tau_3}, \qquad r_3 = \frac{c^2T}{\tau_1+\tau_2}. \tag{21}$$

This shows that the r_i are uniquely determined by a^2, b^2, c^2, as claimed.

Gilbert and Shepp **[2]** showed that in the generalization to n resistances $r_1, \ldots, r_n$ on the sides of an n-gon, the r_i are uniquely determined by the equivalent resistances between consecutive vertices.

References

[1] S. I. Brown and M. I. Walter, What-if-not? An elaboration and second illustration, Mathematics Teaching, 51 (1970) 9–17.

[2] E. N. Gilbert and L. A. Shepp, A resistance problem, SIAM Review, 27 (1985) 449–452.

[3] L. C. Larson, Problem-Solving Through Problems, Springer-Verlag, New York, 1983.

Some Remarks About Bridge Probabilities

P. F. Zweifel

Virginia Polytechnic Institute and State University
Blacksburg, VA 24061

This paper is dedicated to the memory of Mark Kac, great mathematician and probabilist.

Probability theory is basic in many areas of physics and engineering, for example, quantum theory, statistical mechanics, and risk-benefit analysis. I have found it valuable in my teaching to go back to some very basic concepts and to discuss them in detail with the students, hoping that the proper building blocks will be available for constructing the more complicated situations necessary to describe, for instance, physical systems in statistical mechanics. Since my avocation has been, since my student days long ago, various games of chance (principally, backgammon and bridge), I find it useful to choose my examples from these areas. This also has the advantage of retaining the students' interest longer than, say, a laborious discussion of molecules confined in a cell of phase space. I felt it might be worthwhile to share some of these examples which, by and large, are not found in the standard elementary texts, with the readers of this journal.

A sample problem which arises in backgammon is: *given an arbitrary roll of two unbiased dice, what is the probability that a given number, say ⚀, will appear on one or both of the dice?* The correct answer of $11/36$ can be arrived at by brute force (i.e., enumerate all possible configurations) or by the formula

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B). \tag{1}$$

But the practical solution is simply to calculate the probability that a 1 does *not* appear. Almost anybody will come up with $(5/6)^2$ for that one, whereupon subtraction from 1 gives the correct answer.

In this trivial example, direct enumeration or application of the abstract formula may seem to be just about as easy as the *practical* solution, but let's complicate the problem a little by asking about n rolls. Then it is very hard to arrive at the result $1-(5/6)^n$ in any other way, *especially for arbitrary* n.

This method of calculation is actually based on De Morgan's Law $(A \vee B) = (A' \wedge B')$; since $P(A') = 1 - P(A)$ it follows that

$$P(A \vee B) = 1 - P(A')P(B') \tag{2}$$

because $P(A \wedge B) = P(A)P(B)$ always holds for independent events A and B. In fact, (2) holds *only* if A and B are independent. If A and B are *not* independent, $P(A \vee B) = 1 - P(A')P(B'|A')$ where $P(B'|A')$ is the conditional probability of B' given A'.

In my favorite pastime, bridge, a player is frequently called upon to choose among several lines of play; of course, the winning player will choose the line with the greatest probability of success. Suppose a given line will succeed if spades break 3–3 (each of the two *opponents* holds three cards of the spade suit) with probability p_1 or if West holds the king of hearts with probability p_2 or if the jack of spades lies in the hand with no trumps with probability p_3. Then the probability of success is approximately calculated as $1-(1-p_1)(1-p_2)(1-p_3)$; the calculation is approximate because the events may not be strictly independent.

I know that all bridge players (at least those who bother to calculate at all) have learned to calculate this way, because there is really no other way to do it. Now consider a slightly different situation: the line of play will succeed if spades break 4–2 (probability p_1) or 3–3 (probability p_2). These two events are evidently mutually exclusive so referring to (1) we see that the probability is $p_1 + p_2$. One can then go on to pose problems in which one must decide whether given events are independent or mutually exclusive or somewhere in between, and compute (or estimate) the probabilities in each case. All three situations arise in bridge; as a practical matter, only the two extremes are usually considered in actual play, during which only mental calculations are allowed by the rules of the game.

I would next like to illustrate the point that very often conditional probability is a more useful tool for carrying out practical computations than is absolute probability. In the course of this discussion, I shall demonstrate an effective and simple method for computing the probabilities of various bridge-suit breaks.

Suppose you have two red balls and two blue balls and place them at random, two each, in two boxes. What is the probability that both blue balls will appear in the same box? In this simple case, a direct enumeration is possible. Using an obvious notation, we have B_1B_2, B_1R_1, B_1R_2, B_2R_1, B_2R_2 and R_1R_2 as the equally likely possible contents of, say box one, from which we see that the desired probability is $1/3$. This is, of course, the absolute approach. The conditional approach is to assume that one of the blue balls, say B_1, is placed in one of the boxes, e.g., box number one. The remaining three balls are dealt at random. Clearly ball B_2 has twice as much chance of going into box number two, with two vacancies, as into box one which has only one vacancy remaining, so we again arrive at the correct answer 1/3. If the conditional computation seems to have little advantage over the absolute in this case, modify the problem just a bit to ask a bridge question. If thirteen cards are dealt to each of two players, East and West, and if among the twenty-six cards dealt are two spades, what is the probability that both spades are in the same hand, i.e., that spades break 2–0? Absolute computation now becomes lengthy while the conditional calculation is just as fast as in the two-ball case, easily giving the answer $12/25 = 48\%$.

Similarly, for three outstanding cards the probability of a 3–0 break is $(12/25) \cdot (11/24) = 22\%$ while a 2–1 break occurs in $3 \cdot (13/25) \cdot (12/24) = 78\%$ of the cases (the factor 3 here corresponds to the number of ways that a 2–1 break can occur, 6, times 1/2, since the first card can be placed in either hand). Following similar reasoning, we can calculate the probability of 4–0, 3–1 and 2–2 breaks with four cards outstanding as $11/115 = 9.6\%$; $286/575 = 49.7\%$; and $234/575 = 40.7\%$, respectively. A relatively simple rule for calculating the probability of an m–n break is the following. Calculate the number of ways the break can occur, multiply by 1/2, since either hand can hold the first card, and proceed to multiply by

$$\frac{12}{25} \cdot \frac{11}{24} \cdots \frac{13-(m-1)}{26-(m-1)} \cdot \frac{13}{26-m} \cdots \frac{13-(n-1)}{26-(m+n-1)}.$$

These are just the appropriate factors corresponding to "filling up" vacancies in the two hands. It is easy to verify that it makes no difference in which order one assumes the vacancies are filled. (Incidentally, the number of ways the m–n break can occur is given by $2C(m+n, n)$ if $n \neq m$ but by $C(m+n, m)$ if $m = n$, where $C(k, j)$ is the number of combinations of k objects taken j at a time.

It appears that this approach to computing suit division is not well known, even to the experts. Thus Hugh Kelsey, in chapter 6 of the otherwise outstanding book *Advanced Play at Bridge* [**5**] suggests that the probability of a 4–2 break be computed combinatorically as

$$\frac{2 \cdot C(6,2) \cdot C(20,11)}{C(26,13)} = 2\left[\frac{6!}{4!2!} \cdot \frac{20!}{11!9!}\right] \Big/ \frac{26!}{13!13!} = \frac{2519400}{10400600}.$$

This, of course gives the same answer (48.4%) as the formula I have suggested,

$$30 \cdot \frac{1}{2} \cdot \frac{12}{25} \cdot \frac{13}{24} \cdot \frac{12}{23} \cdot \frac{11}{22} \cdot \frac{10}{21},$$

except that Kelsey's approach seems to require somewhat more effort. Incidentally, a "quick and dirty" answer can be obtained by dividing the number of ways six cards can be distributed between two hands 4–2 ($2C(6,2) = 30$) by the total number of ways six cards can be distributed ($2^6 = 64$) to obtain 46.8% or, for "at table" mental calculations, $15/32 \approx 1/2$.

Probably the least intuitively understood aspect of probability theory involves applications of Bayes' Theorem. Recall that Bayes' Theorem tells us that if a probability measure space S is divided into n disjoint sets A_i, and if $B \subset S$ then

$$P(A_l|B) = \frac{P(B|A_l)P(A_l)}{\sum_i P(B|A_i)P(A_i)}, \qquad l = 1, 2, \ldots n, \tag{3}$$

where $P(A|B)$ represents the conditional probability of A given B. This relation is easy to derive from the two basic formulas

$$P(B) = \sum_{i=1}^{n} P(A_i)P(B|A_i)$$

and $P(A \wedge B) = P(B|A)P(A) = P(A|B)P(B)$.

A well-known nonbridge example of Bayes' Theorem is the so-called jailer's paradox. In this scenario, A, B, and C are in prison; one of the three is condemned to death while the other two will be freed, but only the jailer knows which. Now A wants to send a letter to his girl friend, and since he knows that with certainty either B or C is going free (remember, only one of the three will die) he asks the jailer to tell him the name of one of the other two who will be freed and thus would be able to deliver the letter (should A turn out to be the victim). The jailer demurs, explaining that as of now A has $1/3$ probability of dying but if he were to say, for example, that B would go free, then A's probability of dying would increase to $1/2$.

This seems intuitively suspect, since the jailer is providing no new information to A, so why should his probability of dying change? Here, Bayes' Theorem gives the correct answer (still $1/3$) easily: set A_1 = "A dies," A_2 = "C dies," and B = "Jailer says B goes free." Note: $P(B|A_1) = 1/2$ but $P(B|A_2) = 1$.

Aside from the formal application of Bayes' Theorem, one would like to understand this "paradox" from an intuitive point of view. The crucial point which can perhaps make the situation clear is the discrepancy between $P(B|A_1)$ and $P(B|A_2)$ noted above. Bridge players call this the "Principle of Restricted Choice" (more on this later). The probability of a restricted choice is obviously greater than that of a free choice, and a common error made by those who attempt to solve such problems intuitively is to overlook this point. In the case of the jailer's paradox, if the jailer says "B will go free" this is twice as likely to occur when C is scheduled to die (restricted choice; jailer *must* say "B) as when A is scheduled to die (free choice; jailer could say either "C" or "B").

The history of this principle in the game of bridge is extremely interesting. The prototype application occurs in the following situation. South, declarer, holding the ace, king and one or more other spades, plays the ace of spades. There are four spades outstanding between the defenders, East and West, the "2", "3", queen(Q) and jack(J). East plays the 2 and West the Q. South must now decide who has the J, so that he can play appropriately (finesse, or drop) in order to avoid losing a trick to the jack of spades. (The reader need know nothing about bridge in order to understand the probabilistic situation. The only point needed is that the queen and jack are equals; holding both, a (good) defender would play each one with probability $1/2$. Also, holding a low card $(3,2)$ in addition to a high card or cards (Q, J) a defender would invariably play the low card. Of course, a player is required to "follow suit" if able, i.e., when South plays a spade, the other players must play spades if they hold any.)

Let's look at a diagram of the situation. The possible holdings after the play of one round of spades are

	West	*East*
1)	J	3
2)	None	$J, 3$

Since there are only two cases, it seems obvious that the probability that West also holds the jack is $1/2$. Even very well trained mathematicians often cannot bring themselves to believe that the probability of case 1) above is not $1/2$, but $1/3$, without going through the detailed calculation. (To be precise, we have to take into account the probability of a 2–2 break vs. a 3–1 break as computed earlier. I will leave it as an exercise in the application of Bayes' Theorem to show that the precise probability of West's holding the Q, J is $6/17$ rather than $1/3$.) The point is, of course, that scenarios 1) and 2) above are *not* equally probable. Scenarios 1) and 2) are induced by the original holdings

	West	*East*
1′)	Q, J	$3, 2$
2′)	Q	$J, 3, 2$

In Case 1′, West would have played the jack $1/2$ of the time, whereas in Case 2′ his choice is *restricted* to playing the queen. Thus Case 2 is twice as probable as Case 1. From the point of view of Bayes' Theorem, $P(\text{West plays } Q|\text{West holds } Q, J) = 1/2$, whereas $P(\text{West plays } Q|\text{West holds only } Q) = 1$.

This is a subtle and easily overlooked point. Contract bridge has been played since 1929 and its predecessors (auction bridge, dummy bridge, and whist) several hundred years more. The same problem arises in all of these games, and yet, until approximately 1960, the two cases were treated as equally probable, even by expert players. Since expert bridge players always go with the odds (as they perceive them) this indicates how difficult the odds may be to figure when Bayes' Theorem is involved.

It is interesting that the correct odds for this, and similar bridge situations, were not calculated originally from Bayes' Theorem, but were estimated from the following ingenious heuristic argument. Since it is normal, in bridge, when not attempting to win a trick, to follow suit with the lowest possible card, a variance from the procedure is called a "false card." Now the argument runs, suppose West *never* plays a false card. Then his play of the Q the first time makes the probability of the finesse 100%, while if he *always* plays false cards, the finesse will still win 50% of the time. The true probability, then, must be between 100% and 50%, and assuming a random choice of false cards and true cards, the true probability for the finesse must be around 75%. A similar argument can be to analyze the jailer's paradox by considering the extreme situation that the jailer always chooses a name in alphabetical order or always chooses in antialphabetical order.

I should like to express my appreciation to Profs. Gerald Goodman, John van der Hoek, Stefano Pavari-Fontana, and Bernice Durand for stimulating discussions, and to Profs. I. J. Good, L. Barnett, John Perry, Henry Bauer, and Jeff Rubens for useful communications. Part of the work described here was carried out at Istituto Matematica "Ulissi Dini," Firenze, Italy, and the Aspen Center for Physics, whose hospitality I gratefully acknowledge.

Supported by U.S. Department of Energy Grant No. DE-AS05-80ER10711 and National Science Foundation Grant No. DMS8312451.

References

[1] Emile Borel and André Chéron, Théorie Mathématique du Bridge, Gauthier-Villars, Paris, 1940. (The classic treatise on bridge probabilities.)

[2] P. A. M. Dirac, The Principles of Quantum Mechanics, 3rd ed., Oxford, 1971, Chapter 1. (In quantum mechanics, probability is of course of fundamental importance. An excellent, qualitative discussion of probability in quantum mechanics.)

[3] Richard P. Feynman, Robert B. Leighton, and Matthew Sands, The Feynman Lectures on Physics, Addison-Wesley, Reading, Mass., 1963, Vol. I. Chapter 6. (A brief, readable account of probability concepts in physics.)

[4] Charles Goren, Go with the Odds—A Guide to Successful Gambling, Macmillan, New York, 1969.

[5] Mark Kac, Probability and Related Topics in Physical Sciences, Interscience, New York, 1959. (A somewhat advanced book which gives a really excellent summary of probabilistic methods in the physical sciences.)

[6] H. W. Kelsey, Advanced Play at Bridge, Hart, New York, 1968.

[7] Hugh Kelsey and Michael Glauert, Bridge Odds for Practical Players, Gallancz, London, 1980.

[8] Edward Packel, The Mathematics of Games and Gambling, Mathematical Association of America, 1981.

[9] Terence Reese and Roger Trézel, Master the Odds in Bridge, Gallancz, London, 1979.

[10] Sidney Silodor and John Tierny, Contract Bridge According to Silodor and Tierny, Stanley-Allen, 1961.

Two Paradoxes of Committee Elections

MIKE STARING
KLS College of Education
Sittard, The Netherlands

A well-known committee election procedure—a kind of limited voting—permits each voter to cast as many votes as there are vacancies to be filled in the committee [**1**]. In the Netherlands, for example, nearly all Catholic and Christian schools elect their school councils in this way [**2**].

This procedure, as well as a certain way of dealing with vacancies between two successive elections, may give rise to some strange phenomena. Let us consider the situation in which there are twelve voters (numbered 1 through 12), who will have to elect a committee out of nine candidates, named A through I. In case there are two vacancies to be filled, each voter votes for his/her two top-ranked candidates, and when the individual preferences of the voters for the candidates are as in FIGURE 1, the balloting will have the following results: candidates A and B both receive four votes, whereas H and I receive three votes each; the remaining candidates all receive two votes. Hence, A and B will fill the vacancies.

If, however, there are *three* vacancies instead of two, each voter has to cast three votes. This will result in C, D and E being chosen, because they will receive five votes each, whereas all the other candidates will acquire four or three votes. Similar calculations lead to the conclusion that no member of the *committee-of-two*, nor any member of the *committee-of-three*, will be elected in case there are *four* vacancies; in fact, F, G, H, and I will!

So, this first part of the analysis can be summarized in the *Increasing-Committee-Size Paradox*: a candidate can be elected in a committee of size N, without being sure that (s)he would also be elected in a committee of size $N+1$. In fact, the committee-of-N may be completely disjoint with the committee-of-$N+1$.

voter / preference order	1	2	3	4	5	6	7	8	9	10	11	12
high ↑	A	A	A	B	B	B	C	C	D	D	E	E
	F	F	G	G	H	H	H	I	I	A	B	I
	C	C	C	D	D	D	E	E	E	F	G	G
low ↓	H	H	H	I	F	F	F	G	G	I	I	A
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

FIGURE 1. **Scheme of individual preferences.**

A second phenomenon may occur when an (elected) member of the committee leaves between two successive elections. Often, no full-scale elections are held on such an occasion, but one simply appoints the candidate who at the last-held elections ended next to the elected-ones [**2**]. This seems reasonable, but look what can happen. Suppose there are twelve voters who have to elect a council of size two, out of five candidates. The individual preferences of the voters for the candidates are shown in FIGURE 2. If each voter must cast two votes, the result of the balloting will be that the committee will consist of A (acquiring 12 votes) and B (collecting 5 votes). Candidates C—with 3 votes—as well as D and E—both with 2 votes—will not be elected. If, after a few days, A leaves the committee and if all the individual preferences with respect to the other candidates have remained as they were, a new ballot would determine the winners to be D

voter(s) / preference order	1,2,3,4	5	6,7,8	9,10	11,12
high ↑	A	A	C	D	E
	B	B	A	A	A
	E	D	D	E	D
low ↓	C	C	B	B	B
	D	E	E	C	C

FIGURE 2. **Scheme of preferences.**

and E, with *eight* votes each. However, the procedure of appointing the next-to-the-winners candidate from the first election as the one who will replace the leaving member A, will result in the installation of candidate C. Hence the committee will consist of B and C, instead of D and E. Therefore, our conclusion may be formulated as the *Leaving-Member Paradox*: the procedure of appointing the next-to-the-winners candidate in case one of the elected members leaves the committee (who is, thus, no longer a candidate too) may result in establishing a committee completely disjoint with the one that would be chosen if the voters had the opportunity to vote again.

References

[1] Peter C. Fishburn, An analysis of simple voting systems for electing committees, SIAM J. Appl. Math., 41 (1981) 499–502.

[2] Handboek Medezeggenschapschap in de School, Samson, Alphen a/d Rijn, The Netherlands, 1981, Chapter 3.

A Patrol Problem

ALAN LEVINE
Franklin and Marshall College
Lancaster, PA 17604

Consider the following problem:

An officer of the Highway Patrol is assigned to assist motorists should they become involved in an accident or have a mechanical breakdown. He can use either of the following strategies:

(1) Position his patrol car at the midpoint of the highway and wait for an incident to occur.

(2) Continuously patrol the highway for incidents.

The objective is to determine which of these strategies is "better." (We will define what we mean by "better" later.)

This problem, which was adapted from Larson and Odoni [**2**], is a nice application of elementary probability theory suitable for use at a variety of different levels. Students at the junior high school or high school level, with no formal background in probability, can solve the problem in its simplest form by simulation with a hand calculator or computer. Students at the under-

preference order \ voter(s)	1,2,3,4	5	6,7,8	9,10	11,12
high ↑	A	A	C	D	E
	B	B	A	A	A
	E	D	D	E	D
low ↓	C	C	B	B	B
	D	E	E	C	C

FIGURE 2. **Scheme of preferences.**

and E, with *eight* votes each. However, the procedure of appointing the next-to-the-winners candidate from the first election as the one who will replace the leaving member A, will result in the installation of candidate C. Hence the committee will consist of B and C, instead of D and E. Therefore, our conclusion may be formulated as the *Leaving-Member Paradox*: the procedure of appointing the next-to-the-winners candidate in case one of the elected members leaves the committee (who is, thus, no longer a candidate too) may result in establishing a committee completely disjoint with the one that would be chosen if the voters had the opportunity to vote again.

References

[1] Peter C. Fishburn, An analysis of simple voting systems for electing committees, SIAM J. Appl. Math., 41 (1981) 499–502.

[2] Handboek Medezeggenschapschap in de School, Samson, Alphen a/d Rijn, The Netherlands, 1981, Chapter 3.

A Patrol Problem

ALAN LEVINE
Franklin and Marshall College
Lancaster, PA 17604

Consider the following problem:

An officer of the Highway Patrol is assigned to assist motorists should they become involved in an accident or have a mechanical breakdown. He can use either of the following strategies:

(1) Position his patrol car at the midpoint of the highway and wait for an incident to occur.

(2) Continuously patrol the highway for incidents.

The objective is to determine which of these strategies is "better." (We will define what we mean by "better" later.)

This problem, which was adapted from Larson and Odoni [**2**], is a nice application of elementary probability theory suitable for use at a variety of different levels. Students at the junior high school or high school level, with no formal background in probability, can solve the problem in its simplest form by simulation with a hand calculator or computer. Students at the undergraduate level can prove the results suggested by the simulation. At any level, the problem lends itself to an intuitive discussion which, in my experience, has not led to a consensus about the solution.

From a pedagogical point of view, the problem and its solution as we have presented it here illustrate the fact that mathematics is discovered in much the same way as any other science—by experimentation (here, simulation), followed by confirmation (proof). All too often, students think mathematics was created by divine inspiration since, by the time they see it in class, all the "dirty work" has been "cleaned up."

Let the length of the highway be $2L$ and introduce a coordinate system so that the midpoint of the highway has coordinate 0. (See FIGURE 1.) We note that, without loss of generality, we could

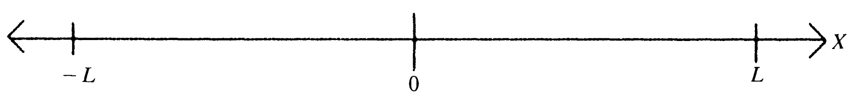


FIGURE 1

assume $L=1$; however, in order to insure that our conclusions are independent of L, we will not do so.

Let X be a continuous random variable representing the location of an incident, and let Y represent the location of the patrol car at the time it begins to respond to the incident. X and Y are presumed independent.

Under strategy (1), $Y \equiv 0$; that is, $Y=0$ with probability 1. Under strategy (2), we assume that the patrol car travels in such a way that Y is uniformly distributed over the interval $[-L, L]$.

Deciding on the distribution of X is a more difficult task. The simplest case is to assume that X is uniformly distributed on the interval $[-L, L]$. However, this may not be completely realistic. In reality, accidents may be more likely to occur at the ends of the highway (where there may be interchanges with other highways) or at some point in the middle (where there may be a dangerous curve). In this case, we assume that X has a distribution characterized by the probability density function $g(x)$ which vanishes outside $[-L, L]$.

We now consider the issue of defining what we mean by a "better" strategy. We first make several assumptions.

1. The patrol car travels at constant velocity v.

2. The highway is designed so that the patrol car can turn around at any point along the highway, in a negligible amount of time.

3. Incidents occur infrequently, thus allowing the patrol car (at least with high probability) to take care of one incident and, under strategy 1, return to the midpoint of the highway before another incident occurs.

Under these assumptions, the distances between the incident and the patrol car at the instant the patrol car responds for strategies (1) and (2) are given by $D_1 = |X|$ and $D_2 = |X-Y|$, respectively. In order to determine which strategy is better, we must define a way to measure the "performance" of a strategy. One way to do this is to consider the average time required for the patrol car to respond to the incident or, since the car travels at constant speed, the average distance between patrol car and incident. (We shall call this the "response distance.") Thus, for this criterion, strategy (1) will be better than strategy (2) if $E[D_1] < E[D_2]$. Another way to measure performance is to assume that there exists a critical time t^* (and corresponding critical distance $d^* = vt^*$) such that any injuries which may have occurred become more serious. It may therefore be desirable to choose a strategy which minimizes the likelihood of serious injury.

Hence, strategy (1) will be better with respect to this criterion if $\text{Prob}(D_1 > d^*) < \text{Prob}(D_2 > d^*)$ for fixed d^*. (We call these quantities "critical probabilities.")

Other criteria are possible; for instance, the "worst-case" scenario in which the goal is to minimize the maximum distance between car and incident. Here, strategy (1) will be better since the patrol car is never more than L units from the incident, while under strategy (2), he may be as much as $2L$ units away.

In the next section, we present a simple simulation program which can be done either on a hand calculator with random-number generator or on a computer. The simulation is presented initially for the case in which X is presumed to be uniformly distributed; we then show how to modify it for any arbitrary distribution of X.

A simulation approach

The first step is to generate a random number x, uniformly distributed on $[-L, L]$, representing the location of the incident. For strategy (1), the response distance is $d_1 = |x|$. For strategy (2), generate another random number, y, uniformly distributed on $[-L, L]$; the response distance is $d_2 = |x - y|$.

Repeat this process N times (see note below), thus obtaining N values of d_1 and d_2. If the criterion is to minimize expected response distance, compute the average of all the values of d_1 and the average of all the values of d_2 and determine which is smaller. If the criterion is to minimize critical probabilities, compute the relative frequency of values of d_1 (and d_2) which exceed d^*. These relative frequencies will serve as approximations to the desired probabilities. (Note: Experience has shown that values of N around 2500–5000 give answers which differ from the theoretical results by less than 1 percent.)

The results we obtained for 5000 trials with $L = 1$ are given in TABLE 1. We see in TABLE 1 that strategy (1) seems to result in both smaller expected response distance and critical probabilities.

	$E[D_1] = .4986$	$E[D_2] = .6671$
t	$\text{Prob}(D_1 > t)$	$\text{Prob}(D_2 > t)$
0	1.000	1.000
.1	.901	.906
.2	.794	.815
.3	.697	.724
.4	.594	.635
.5	.496	.557
.6	.400	.485
.7	.300	.421
.8	.203	.358
.9	.099	.301
1.0	0.000	.252
2.0	0.000	0.000

TABLE 1. **Uniform distribution.**

In order to modify the simulation to handle the case in which X is not uniformly distributed, we shall need the following lemma, which tells us how to generate random numbers from an arbitrary distribution. For a proof, see [**1**, p. 157].

LEMMA 1. *Let Z be uniformly distributed on* $[0,1]$ *and let X be any continuous random variable with cumulative distribution function* $G(x) = \text{Prob}(X < x)$. *Then the random variable $G^{-1}(Z)$ has the same distribution as X.*

Thus in order to generate a random number, x, from a distribution with cumulative distribution function G, first generate a random number z, uniformly distributed on $[0,1]$ and then let

$x = G^{-1}(z)$. This is the only modification needed in the simulation.

We considered three specific distributions. The first two are symmetric about $X = 0$ while the third is not. See FIGURE 2 for graphs of the density functions.

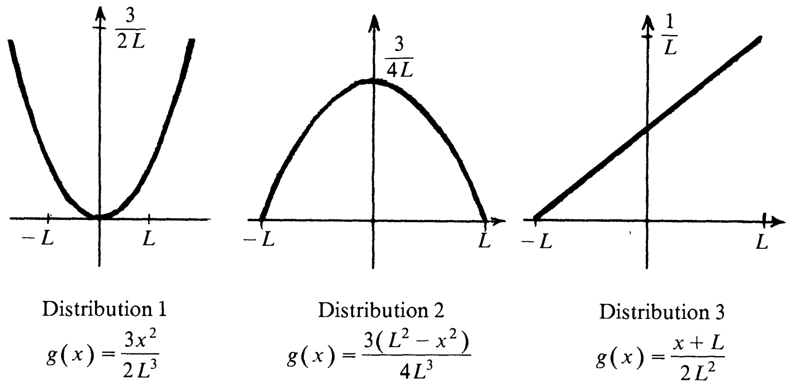


FIGURE 2

1. $g(x) = 3x^2/2L^3$; $G(x) = x^3/2L^3 + 1/2$. In this case, incidents are more likely to occur at the ends of the highway than in the middle.
2. $g(x) = 3(L^2 - x^2)/4L^3$; $G(x) = 3x/4L - x^3/4L^3 + 1/2$. This distribution describes the situation in which incidents are most likely to occur at the midpoint of the highway.
3. $g(x) = (x + L)/2L^2$; $G(x) = (x + L)^2/4L^2$. This corresponds to the case in which the incidents are most likely to occur at one end ($X = L$) of the highway and unlikely to occur at the other end ($X = -L$).

The results of the simulations (again for $L = 1$) are given in TABLE 2.

	Distribution 1		Distribution 2		Distribution 3	
	$E[D_1] = .749$ $E[D_2] = .797$		$E[D_1] = .374$ $E[D_2] = .601$		$E[D_1] = .499$ $E[D_2] = .667$	
t	$\Pr(D_1 > t)$	$\Pr(D_2 > t)$	$\Pr(D_1 > t)$	$\Pr(D_2 > t)$	$\Pr(D_1 > t)$	$\Pr(D_2 > t)$
0	1.000	1.000	1.000	1.000	1.000	1.000
.1	.999	.900	.850	.900	.905	.895
.2	.992	.829	.702	.803	.802	.799
.3	.974	.754	.557	.705	.699	.712
.4	.936	.696	.430	.609	.593	.635
.5	.875	.632	.314	.523	.500	.563
.6	.775	.573	.211	.446	.393	.498
.7	.651	.519	.122	.371	.299	.424
.8	.485	.468	.055	.307	.202	.362
.9	.273	.414	.017	.243	.105	.305
1.0	0.000	.368	0.000	.191	0.000	.258
2.0	0.000	0.000	0.000	0.000	0.000	0.000

TABLE 2. **Non-uniform distribution.**

As we can see from TABLE 2, strategy (1) yields smaller expected response distances in all cases. For distributions 2 and 3, strategy (1) results in smaller critical probabilities for all d^*. However, for distribution 1, strategy (2) has smaller critical probabilities for $d^* < .8$ (approximately). We shall prove these results in the next section. One curious fact is that the results for distribution 3 are precisely the same as for the uniform case. Note that D_1 is never bigger than L; hence, for $d^* \geq L$, strategy (1) will be better.

An interesting extension is to try the simulation for several values of L; statistical methods (regression) can then be used to find an empirical relationship between the expected response distances and critical probabilities and L.

There is a potential problem with this procedure. Finding G^{-1} in closed form may be difficult, as for distribution 2. Since G is monotonic increasing (thus guaranteeing the existence of G^{-1}), solving $x = G^{-1}(z)$ is equivalent to solving $G(x) = z$. We used an iterative bisection procedure to solve this equation numerically. Initially, set $x_0 = 0$. If $G(x_0) < z$, then the solution x^* must be greater than x_0; let $x_1 = x_0 + L/2$. Conversely, if $G(x_0) > z$, then x^* must be less than x_0; let $x_1 = x_0 - L/2$. In general, on the nth iteration, if $G(x_{n-1}) < z$, then let $x_n = x_{n-1} + L/2^n$. If $G(x_{n-1}) > z$, then $x_n = x_{n-1} - L/2^n$. Continue until some specified degree of accuracy is attained.

Analytical solution

In this section, we show that strategy (1) yields smaller expected response distance. As in the simulation, we first consider the case where X is uniformly distributed on $[-L, L]$; then we proceed to the general case.

Throughout the remainder of this note, we shall make use of the following lemmas, which we state without proof. (See [**3**, p. 39] for proofs.)

LEMMA 2. *Let X be a continuous random variable with probability density function $g(x)$ and let $U = Q(X)$ for some continuous function Q. Then*

$$E[U] = \int_{-\infty}^{\infty} Q(x)g(x)\,dx.$$

LEMMA 3. *Let X and Y be independent random variables with density functions $g(x)$ and $h(y)$, respectively, and let $U = Q(X, Y)$. Then*

$$E[U] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} Q(x, y)g(x)h(y)\,dy\,dx.$$

For the uniform case, $g(x) = 1/2L$, for $-L \leq x \leq L$. Thus by Lemma 2,

$$E[D_1] = \int_{-L}^{L} |x|g(x)\,dx = \frac{1}{2L}\cdot L^2 = \frac{L}{2}.$$

For strategy (2), Y is uniformly distributed on $[-L, L]$ so that $h(y) = 1/2L$. Thus, by Lemma 3, we have:

$$E[D_2] = \int_{-L}^{L}\int_{-L}^{L} |x-y|\frac{1}{4L^2}\,dy\,dx = \frac{1}{4L^2}\int_{-L}^{L}(x^2 + L^2)\,dx = \frac{2L}{3}.$$

We see then that $E[D_1] < E[D_2]$; hence, strategy (1) is better for this criterion as the simulation results suggested.

For the general case, let $g(x)$ be the probability density function of X, and let $G(x)$ be the corresponding cumulative distribution function. Our only assumption is that $g(x)$ vanishes outside $[-L, L]$.

Proceeding as before, we have

$$E[D_1] = \int_{-L}^{L} |x|g(x)\,dx,$$

$$E[D_2] = \int_{-L}^{L}\int_{-L}^{L}|x-y|g(x)\frac{1}{2L}\,dy\,dx = \frac{1}{2L}\int_{-L}^{L}(x^2+L^2)g(x)\,dx.$$

Thus, $E[D_1] < E[D_2]$ if and only if

$$\int_{-L}^{L}(x^2 - 2L|x| + L^2)g(x)\,dx > 0$$

or, equivalently,

$$\int_{-L}^{L}(|x|-L)^2 g(x)\,dx > 0. \tag{3}$$

Since $g(x) > 0$, the integrand in (3) is nonnegative and, hence, we have proven the following result:

THEOREM 1. *If X is any random variable with range $[-L, L]$ and finite mean, and if Y is uniformly distributed on $[-L, L]$, then $E(|X|) < E(|X-Y|)$.*

In other words, strategy (1) results in a smaller than expected response distance for every distribution of X.

For distribution 1 of the last section, $E[D_1] = 3L/4$, $E[D_2] = 4L/5$. For distribution 2, $E[D_1] = 3L/8$, $E[D_2] = 3L/5$. For distribution 3, $E[D_1] = L/2$, $E[D_2] = 2L/3$ (the same results as in the uniform case, as we noted before). These theoretical results agree with the simulation results presented earlier.

Note: The greatest difference in expected response distance for the two strategies occurs with distribution 2; the least difference is with distribution 1. An intuitive reason for this is that strategy (2) should do relatively well for distributions with large variance. Of the examples considered, distribution 1 has the largest variance; distribution 2 has the smallest. The maximum possible variance amongst distributions which vanish outside $[-L, L]$ occurs when a fraction p of the incidents occur at one end of the highway and the remainder occur at the other end. (No incidents occur anywhere else.) It is easily shown that for this distribution, $E[D_1] = E[D_2] = L$ for all $p \in [0,1]$.

Now we calculate the critical probabilities for the two strategies. Consider the event $\{D_1 > t\}$. This is equivalent to the event $\{X < -t \text{ or } X > t\}$ which is indicated by the cross-hatched portion of the number line in FIGURE 3. If X is uniformly distributed, then Prob($D_1 > t$) equals the ratio

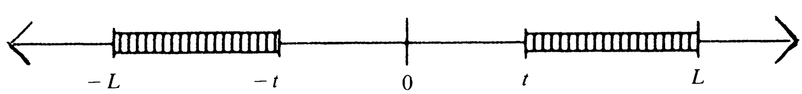


FIGURE 3

of the length of the shaded portion to the length of the highway, which is $2(L-t)/2L = 1 - t/L$. Thus,

$$P(D_1 > d^*) = 1 - d^*/L. \tag{4}$$

For strategy 2, we must consider the joint sample space of X and Y which is a square with vertices at (L, L), $(L, -L)$, $(-L, -L)$, $(-L, L)$. The event $\{D_2 > t\} = \{X - Y < -t$ or $X - Y > t\}$ is indicated by the shaded region in FIGURE 4.

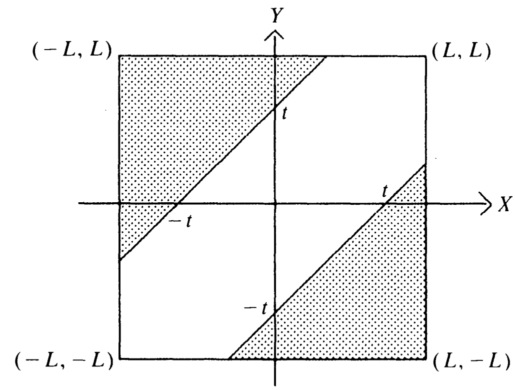


FIGURE 4

The probability of this event is the ratio of the area of the shaded region to the whole square, which is $(2L-t)^2/4L^2 = 1 - t/L + t^2/4L^2$. Thus,

$$\operatorname{Prob}(D_2 > d^*) = 1 - d^*/L + (d^*)^2/4L^2. \tag{5}$$

By comparing equations (4) and (5), we see $\operatorname{Prob}(D_1 > d^*) < \operatorname{Prob}(D_2 > d^*)$ for all values of d^*; thus strategy (1) is better with respect to this criterion as well.

Note: We could have found $E[D_1]$ and $E[D_2]$ by observing that equations (4) and (5) imply that the probability density functions of D_1 and D_2 are $f_1(t) = t/L$ and $f_2(t) = 1/L - t/2L^2$, whence

$$E[D_1] = \int_0^L tf_1(t)\,dt = \frac{L}{2} \quad \text{and} \quad E[D_2] = \int_0^{2L} tf_2(t)\,dt = \frac{2L}{3}.$$

Determining the critical probabilities is somewhat more difficult for arbitrary distributions since the geometrical derivation used above is no longer valid. For strategy (1), $\operatorname{Prob}(D_1 > t) = \operatorname{Prob}(X > t) + \operatorname{Prob}(X < -t) = 1 - G(t) + G(-t)$. For strategy (2), $\operatorname{Prob}(D_2 > t)$ equals the volume of the solid bounded above by the surface $z = g(x)h(y) = g(x)/2L$, below by the x-y plane over the shaded region in FIGURE 4. Thus,

$$\begin{aligned} P(D_2 > t) &= \int_{-L}^{L-t}\int_{x+t}^{L} g(x)/2L\,dy\,dx + \int_{-L+t}^{L}\int_{-L}^{x-t} g(x)/2L\,dy\,dx \\ &= \frac{1}{2L}\left(\int_{-L}^{L-t}(L-x-t)\,g(x)\,dx + \int_{-L+t}^{L}(x-t-L)\,g(x)\,dx\right). \end{aligned}$$

Although it does not seem possible to make general statements for all distributions, as we could in the case of expected response distance, it is possible to compute the critical probabilities for specific examples to determine the better strategy.

For distribution 1, we get $\operatorname{Prob}(D_1 > t) = 1 - t^3/L^3$ and $\operatorname{Prob}(D_2 > t) = 1 - t/L + 3t^2/4L^2 - t^3/2L^3 + t^4/8L^4$. Upon graphing the two functions (see FIGURE 5), we see that strategy (2) is better for $t < .8L$, while strategy (1) is better for $t > .8L$. This confirms the conclusions suggested by the simulation. We leave it as an exercise for the reader to work out the results for the other examples.

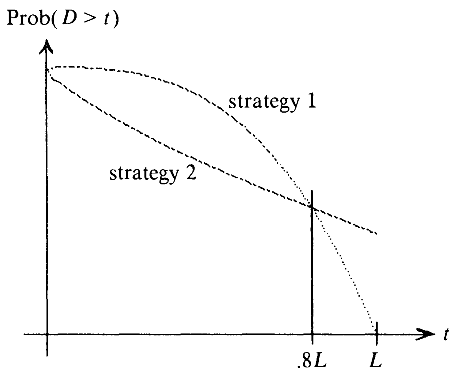


FIGURE 5

An interesting statistical exercise at this stage is to verify that the simulation results agree with the theoretical distributions of D_1 and D_2. This can be accomplished by using the cumulative distribution function to find the expected number of outcomes in each subinterval given in TABLE 2. Then, treating the simulation data as the "observed" values, compute an appropriate chi-square statistic.

Extensions

There are several directions in which these results can be extended. Other strategies can be considered; for example, the strategy in which the patrol car stays at one end of the highway. We could also complicate matters by relaxing the assumption that the car can turn around at any point on the highway; rather, we may assume that he has to go to one end of highway in order to turn around.

Another possibility is to consider what happens if there are two patrol cars. Some possible strategies include: allowing both cars to patrol the highway continuously and sending the one which is nearest to the incident; stationing both of them at the midpoint and sending whichever is available (if both are available, flip a coin); stationing them at coordinates $-L/3$ and $L/3$ and sending the nearest one.

We have only considered a one-dimensional (linear) highway. Suppose the patrol car is responsible for a rectangular network of highways. There are many possible strategies which could be compared. (Warning: You may encounter quadruple integrals in this version. Only the brave should proceed.)

Analytical results may be hard to come by for these cases; however, they can all be handled easily by simulation.

I would like to thank the members of my Stochastic Processes class that I taught at Franklin and Marshall College in the fall 1984 semester. In particular, John Allison helped with the computational aspects of this paper. Michelle Carter, David Garrison, Frank Narr, Douglas Reymann, and David Wiesenfeld all contributed to the results for the case in which X has an arbitrary distribution.

References

[1] R. Hogg and E. Tanis, Probability and Statistical Inference, Macmillan, 1983.
[2] R. Larson and A. Odoni, Urban Operations Research, Prentice-Hall, 1981.
[3] S. Ross, Introduction to Probability Models, Academic Press, 1980.

An Algorithm for Multiplication in Modular Arithmetic

WALTER GROSS
24 Eldridge Avenue
Johnson City, NY 13790

PETER HILTON
SUNY Binghamton
Binghamton, NY 13901

JEAN PEDERSEN
University of Santa Clara
Santa Clara, CA 95053

KIM YEW YAP
National University of Singapore
Kent Ridge, Singapore

Introduction

Modular arithmetic provides a good introduction to genuinely mathematical ideas in arithmetic and to the basic concepts of abstract algebra. In this note we offer an algorithm for multiplying in modular arithmetic. To be precise, let n and l be positive integers with $l < n$ and l prime to n. Then our algorithm will allow us to multiply by l modulo n without actually carrying out the multiplications in ordinary whole number arithmetic.

We will also discuss a refinement of the algorithm which enables us to perform multiplications modulo n by all such numbers l in a simple natural sequence.

The algorithm

Our algorithm naturally divides itself into two parts; we call these Algorithm A and Algorithm B. It is Algorithm B for which we claim originality, but we first describe Algorithm A.

Algorithm A. Find k, $1 \leqslant k < n$, such that $kl \equiv -1 (\text{mod } n)$ (note that such an integer k is unique). We may, for instance, proceed by using the Euclidean algorithm to find integers u, v, with $|u| < n$, $|v| < l$, and $ul + vn = 1$. Then

$$\begin{aligned} k &= -u \quad \text{if } u \text{ is negative,} \\ k &= n - u \quad \text{if } u \text{ is positive.} \end{aligned}$$

Algorithm B. Write out the residues modulo n, starting with 1 and in their natural order in a rectangular $(k \times l)$ array, which we call *NA* (natural array). That is, the residues increase from left to right, row by row. Then apply to this array the transformation T which consists of taking the *columns* of *NA* in *reverse* order and writing the entries again in a $(k \times l)$-array, row by row. Call the new array *TNA*. In a $(k \times l)$-array, we refer to the entry in the ith row, reading downwards, and the jth column, reading across, as the (i, j)-entry. Then *the* (i, j)*-entry of TNA is obtained from the* (i, j)*-entry of NA by multiplying by* l *modulo* n. Before proving this we give an example.

EXAMPLE 1. Let $n = 17$, $l = 3$. Algorithm A produces $k = 11$. Our (11×3)-array, modulo 17, reads

$$NA=\begin{Bmatrix} 1 & 2 & 3\\ 4 & 5 & 6\\ 7 & 8 & 9\\ 10 & 11 & 12\\ 13 & 14 & 15\\ 16 & 0 & 1\\ 2 & 3 & 4\\ 5 & 6 & 7\\ 8 & 9 & 10\\ 11 & 12 & 13\\ 14 & 15 & 16 \end{Bmatrix};\quad\text{and}\quad TNA=\begin{Bmatrix} 3 & 6 & 9\\ 12 & 15 & 1\\ 4 & 7 & 10\\ 13 & 16 & 2\\ 5 & 8 & 11\\ 14 & 0 & 3\\ 6 & 9 & 12\\ 15 & 1 & 4\\ 7 & 10 & 13\\ 16 & 2 & 5\\ 8 & 11 & 14 \end{Bmatrix}.$$

Then each entry in TNA is, modulo 17, 3 times the corresponding entry in NA. For example, the $(5,2)$-entry of NA is 14, the $(5,2)$-entry of TNA is 8, and $8\equiv 3\cdot 14 \pmod{17}$. Of course, the nonzero residues modulo 17 occur *twice* in NA because $11\times 3=2\times 17-1$.

We now justify Algorithm B. However it will be convenient for the sequel to prove a slightly generalized version. In this version we choose any positive integer b and we suppose that we have already written the residues modulo n, starting with b and then advancing by b modulo n, in their natural order in a rectangular $(k\times l)$-array, which we call $NA(b)$, so that $NA=NA(1)$. We then claim

$$TNA(b)=NA(lb). \tag{1}$$

Proof of (1). The $(1,l)$-entry in $NA(b)$ is lb, modulo n, and the $(r+1,s)$-entry is obtained from the (r,s)-entry by adding lb modulo n. It follows that $TNA(b)$ is the array starting with lb, modulo n, and proceeding by adding lb modulo n, that is, it is the array $NA(lb)$, except perhaps at the 'seams'. Precisely, it remains to prove that, in passing from the foot of one column in $NA(b)$ to the head of the preceding column, we advance by lb modulo n.

Now since the entry at the foot of the lth column of $NA(b)$ is congruent to $-b$ modulo n, it follows that the entry at the foot of the sth column is congruent to $-b-(l-s)b$ modulo n. If $s\geqslant 2$, the entry at the head of the $(s-1)$st column is congruent to $(s-1)b$. But

$$-b-(l-s)b+lb=(s-1)b,$$

so the proof of (1) is complete.

Of course Algorithm B is justified by taking $b=1$ in (1). Notice that the algorithm is especially nice if $l\mid n-1$. For then we simply find k by division and the array NA is of 'minimal area' $n-1$, with no repetition; we simply write out the nonzero residues mod n in a rectangular $(k\times l)$-array, with $kl=n-1$.

From (1) we immediately obtain by iteration

$$T^qNA=NA(l^q). \tag{2}$$

Thus we obtain the rule for multiplying by l^q modulo n, for any $q\geqslant 1$. This is especially satisfactory if l is a primitive residue (or root) modulo n, that is, if the powers of l run through *all* residues prime to n. *Such primitive residues exist if and only if* $n=2$, 4, p^m, $2p^m$, *where* p *is an odd prime and* $m\geqslant 1$ [**1**]. Thus in these cases, by judicious choice of l, we may quickly determine how to multiply modulo n by any number prime to n. In particular, if n is itself prime then, by suitably choosing l, we obtain the entire multiplication table of arithmetic modulo n. Let us give an example.

EXAMPLE 2. Let $n=13$. It is then easy to see that 2 is primitive modulo 13; this amounts to showing that the smallest power of 2 which is congruent to 1 is 2^{12}. But this must be so since $2^6=64\not\equiv 1$ and $2^4=16\not\equiv 1$. In fact the powers of 2 run successively through the residues $2,4,8,3,6,12,11,9,5,10,7,1$. Thus if we start with

$$NA = \begin{Bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \\ 7 & 8 \\ 9 & 10 \\ 11 & 12 \end{Bmatrix}$$

and successively carry out the transformations $T, T^2, T^3, \ldots, T^{11}$, we obtain the rules for multiplying by $2, 4, 8, 3, 6, 12, 11, 9, 5, 10, 7$, respectively, that is, the entire multiplication table. This example exhibits the favorable feature $l | n-1$ already referred to.

Some further points can be made relating to this example, but having a wider application.

First, it is not necessary to compute the powers of 2 modulo 13 (or, more generally, of l modulo n) to know what we are multiplying by—*one simply looks at the leading entry in any $T^q NA$, and that entry indicates the multiplier.*

Second, suppose that $n = 4$, p^m, or $2p^m$, and that l is a primitive residue modulo n. Then, since the multiplicative group of residues prime to n is a cyclic group of even order $\varphi(n)$ generated by l, and since such a group has exactly one element of order 2, it follows that

$$l^{\frac{1}{2}\varphi(n)} \equiv -1 \pmod{n}, \tag{3}$$

where φ is Euler's totient function. Formula (3) means that, if T is executed $\frac{1}{2}\varphi(n)$ times, the result is our original NA multiplied by -1 modulo n. But this is equivalent to *reversing* NA, that is, writing the entries of NA in reverse order. One may verify that, with $n = 13$, $l = 2$, then $\varphi(n) = 12$ and

$$T^6 NA = \begin{Bmatrix} 12 & 11 \\ 10 & 9 \\ 8 & 7 \\ 6 & 5 \\ 4 & 3 \\ 2 & 1 \end{Bmatrix}.$$

Another good example is furnished by $n = 18$, (so that $\varphi(n) = 6$), $l = 5$, $k = 7$. Then

$$NA = \begin{Bmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \\ 16 & 17 & 0 & 1 & 2 \\ 3 & 4 & 5 & 6 & 7 \\ 8 & 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 & 17 \end{Bmatrix}, \quad T^3 NA = \begin{Bmatrix} 17 & 16 & 15 & 14 & 13 \\ 12 & 11 & 10 & 9 & 8 \\ 7 & 6 & 5 & 4 & 3 \\ 2 & 1 & 0 & 17 & 16 \\ 15 & 14 & 13 & 12 & 11 \\ 10 & 9 & 8 & 7 & 6 \\ 5 & 4 & 3 & 2 & 1 \end{Bmatrix}.$$

Third, since the transformation T enables us (in the general case) to multiply by l, it follows that the inverse transformation T^{-1} enables us to divide by l. Notice that division by l is defined since l is prime to the modulus n.

In a sense, *division by l is easier than multiplication by l*, since we may write out the array $T^{-1}NA$ without having to write out NA! For example, with $n = 7$ and $l = 2$ (so that $k = 3$), we simply write the numbers from 1 to 6 in two columns, starting on the right:

$$T^{-1}NA = \begin{Bmatrix} 4 & 1 \\ 5 & 2 \\ 6 & 3 \end{Bmatrix}.$$

Then we know that we get

$$NA = \begin{Bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{Bmatrix}$$

by doubling each entry in $T^{-1}NA$. We need to write out NA in order to know which residue has been halved to obtain the corresponding residue in $T^{-1}NA$, but not in order to be able to *produce* $T^{-1}NA$.

References

[1] Ivan Niven and H. S. Zuckerman, An Introduction to the Theory of Numbers, Fourth Edition, John Wiley and Sons, 1980, p. 62.

The Largest Unit Ball in Any Euclidean Space

JEFFREY NUNEMACHER

Oberlin College

Oberlin, OH 44074

In what dimensional Euclidean space does the unit ball have greatest volume? greatest surface area? The usual approach to this problem is to find explicit formulas for the volume and surface area and then to analyze their behavior. The standard derivations of these formulas are based on recurrence relations and typically involve some advanced calculus. For various approaches see, for example, [**1**, p. 411]; [**2**, p. 302]; [**3**, p. 220]; [**4**, p. 502]; [**5**, p. 324]. This note solves the problem by working directly from the recurrence relations. This approach is pleasingly simple and makes the argument accessible to a multivariable calculus class.

Let $B_n(r)$ denote the open ball of radius r in R^n, i.e.,

$$B_n(r) = \left\{ (x_1, x_2, \dots, x_n) \,\middle|\, \sum_{i=1}^{n} x_i^2 < r^2 \right\},$$

and let $V_n(r)$ denote its volume. Since balls are similar n-dimensional objects, it is not surprising that there are constants a_n so that $V_n(r) = a_n r^n$. This statement can be proved using the change of variables formula (see, e.g., [**4**, p. 500]). A more elementary argument can be carried out based on approximation by Riemann sums, using the basic observation that if all sides of an n-box are magnified by r, then the volume is magnified by r^n.

By definition we have

$$V_n(r) = \iint\limits_{B^n(r)} \cdots \int 1\, dx_1 \cdot dx_2 \cdots dx_n = \int_{-r}^{r} \left(\iint\limits_{\sum_{i=1}^{n-1} x_i^2 < r^2 - x_n^2} \cdots \int 1\, dx_1\, dx_2 \cdots dx_{n-1} \right) dx_n .$$

Since the value of the inner integral is $V_{n-1}\left(\sqrt{r^2 - x_n^2}\right)$, we find that

$$V_n(r) = \int_{-r}^{r} a_{n-1} \left(\sqrt{r^2 - x_n^2} \right)^{n-1} dx_n .$$

$$NA = \begin{Bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{Bmatrix}$$

by doubling each entry in $T^{-1}NA$. We need to write out NA in order to know which residue has been halved to obtain the corresponding residue in $T^{-1}NA$, but not in order to be able to *produce* $T^{-1}NA$.

References

[1] Ivan Niven and H. S. Zuckerman, An Introduction to the Theory of Numbers, Fourth Edition, John Wiley and Sons, 1980, p. 62.

The Largest Unit Ball in Any Euclidean Space

JEFFREY NUNEMACHER
Oberlin College
Oberlin, OH 44074

In what dimensional Euclidean space does the unit ball have greatest volume? greatest surface area? The usual approach to this problem is to find explicit formulas for the volume and surface area and then to analyze their behavior. The standard derivations of these formulas are based on recurrence relations and typically involve some advanced calculus. For various approaches see, for example, [**1**, p. 411]; [**2**, p. 302]; [**3**, p. 220]; [**4**, p. 502]; [**5**, p. 324]. This note solves the problem by working directly from the recurrence relations. This approach is pleasingly simple and makes the argument accessible to a multivariable calculus class.

Let $B_n(r)$ denote the open ball of radius r in R^n, i.e.,

$$B_n(r) = \left\{ (x_1, x_2, \dots, x_n) \,\middle|\, \sum_{i=1}^{n} x_i^2 < r^2 \right\},$$

and let $V_n(r)$ denote its volume. Since balls are similar n-dimensional objects, it is not surprising that there are constants a_n so that $V_n(r) = a_n r^n$. This statement can be proved using the change of variables formula (see, e.g., [**4**, p. 500]). A more elementary argument can be carried out based on approximation by Riemann sums, using the basic observation that if all sides of an n-box are magnified by r, then the volume is magnified by r^n.

By definition we have

$$V_n(r) = \iint_{B^n(r)} \cdots \int 1\, dx_1\, dx_2 \cdots dx_n = \int_{-r}^{r} \left(\iint_{\sum_{i=1}^{n-1} x_i^2 < r^2 - x_n^2} \cdots \int 1\, dx_1\, dx_2 \cdots dx_{n-1} \right) dx_n.$$

Since the value of the inner integral is $V_{n-1}\left(\sqrt{r^2 - x_n^2}\right)$, we find that

$$V_n(r) = \int_{-r}^{r} a_{n-1} \left(\sqrt{r^2 - x_n^2} \right)^{n-1} dx_n.$$

A standard trigonometric substitution now shows that

$$V_n(r) = 2a_{n-1}r^n\int_0^{\pi/2}\cos^n\theta\,d\theta.$$

Thus the volumes $V_n \equiv V_n(1)$ of the unit balls are related by the first-order recurrence relation

$$V_n = 2V_{n-1}\int_0^{\pi/2}\cos^n\theta\,d\theta,$$

where $V_1 = 2$, the length of the interval $(-1,1)$ in R^1.

For $n \geqslant 0$ let b_n denote the coefficient $2\int_0^{\pi/2}\cos^n\theta\,d\theta$. A standard integration by parts yields the second-order recurrence relation $b_n = ((n-1)/n)b_{n-2}$, which together with the initial conditions $b_0 = \pi$ and $b_1 = 2$ allows the calculation of any particular b_n. It is clear that $\{b_n\}$ is decreasing, and since for any $\varepsilon > 0$, $\lim_{n\to\infty}\cos^n\theta = 0$ uniformly for $\varepsilon \leqslant \theta \leqslant \pi/2$, we see that $\lim_{n\to\infty} b_n = 0$. We have $V_n = b_nV_{n-1}$, so that V_n increases until $b_n < 1$, then it decreases to 0. Calculation of the first several terms shows that $b_5 > 1$ but $b_6 < 1$; thus the greatest volume occurs in dimension $n = 5$.

A similar method can be used to show that S^{n-1}, the unit sphere of R^n, has greatest surface area when $n = 7$. Let $A_n(r)$ denote the surface area of S^{n-1}. The crucial ingredient is the believable but not-quite-obvious fact that $d/dr\,V_n(r) = A_n(r)$. This formula can be obtained by using the change of variables formula with generalized polar coordinates (see, e.g., [**3**, p. 340]). Here is an elementary proof.

Let $0 = r_0 < r_1 < \cdots < r_n = r$ be a partition of $[0,r]$ and let $\Delta r_n = r_n - r_{n-1}$. Then by elementary geometric considerations

$$\sum_{i=1}^{n} A_n(r_{i-1})\,\Delta r_i < V_n(r) < \sum_{i=1}^{n} A_n(r_i)\,\Delta r_i.$$

It follows that $V_n(r) = \int_0^r A_n(s)\,ds$, whence $(d/dr)\,V_n(r) = A_n(r)$.

Now since $V_n(r) = a_nr^n$, we have $A_n(r) = na_nr^{n-1}$, so that $A_n \equiv A_n(1) = nV_n$. Thus

$$A_n = nb_nV_{n-1} = \frac{n}{n-1}\,b_nA_{n-1},$$

which gives a one-step recurrence relation for A_n. Again the sequence of coefficients is decreasing towards 0, so that A_n increases until $(n/(n-1))\,b_n < 1$, then decreases towards 0. Since $(n/(n-1))\,b_n = b_{n-2}$ and b_5 is the last b_n to exceed 1, the surface area of the unit sphere is greatest in R^7.

It is an interesting consequence of this argument that both V_n and A_n approach 0 as n gets large.

References

[1] T. Apostol, Calculus, Vol. 2, 2nd ed., John Wiley & Sons, New York, 1969.

[2] R. Courant, Differential and Integral Calculus, Vol. 2, Wiley-Interscience, New York, 1936.

[3] W. Fleming, Functions of Several Variables, 2nd. ed., Springer-Verlag, New York, 1977.

[4] S. Lang, Undergraduate Analysis, Springer-Verlag, New York, 1983.

[5] J. Marsden, Elementary Classical Analysis, Freeman, San Francisco, 1974.

PROBLEMS

Loren C. Larson, Editor
Bruce Hanson, Associate Editor
St. Olaf College

Leroy F. Meyers, Past Editor
The Ohio State University

Proposals

To be considered for publication, solutions should be received by October 1, 1986.

1242. *Proposed by Ronald Graham, Bell Laboratories, Murray Hill, New Jersey.*

For each positive integer n, let $f(n)$ be the smallest integer r for which there is an increasing sequence of integers $n = a_1 < a_2 < \cdots < a_k = r$ such that the product $a_1 a_2 \cdots a_k$ is a perfect square. For example, $f(2) = 6, f(3) = 8, f(4) = 4, f(8) = 15$. Prove that f is a one-to-one function.

1243. *Proposed by Edwin Buchman, California State University, Fullerton.*

How many different numbers can be represented among the six expressions:

$$\begin{array}{ll} e_1\colon \lim\limits_{x\to a} g(f(x)) & e_4\colon g(f(a)) \\ e_2\colon g\left(\lim\limits_{x\to a} f(x)\right) & e_5\colon \lim\limits_{y\to f(a)} g(y) \\ e_3\colon \lim\limits_{x\to a}\ \lim\limits_{y\to f(x)} g(y) & e_6\colon \lim\limits_{y\to\left(\lim\limits_{x\to a} f(x)\right)} g(y) \end{array}$$

for any one choice of real functions f and g, and real number a? (Of course, the expressions are all equal if f and g are continuous.)

1244. *Submitted by the Problem Solving Class, University of California, Berkeley.*

Define a boomerang to be any nonconvex quadrilateral. Prove that it is impossible to tile any convex polygon with a finite number of (not necessarily congruent) boomerangs.

1245. *Proposed by Fouad Nakhli (student), American University of Beirut, Lebanon.*

For each number x in the open interval $(1, e)$ it is easy to show that there is a unique number y in (e, ∞) such that $(\ln y)/y = (\ln x)/x$. For such an x and y, show that $x + y > x \ln y + y \ln x$.

1246. *Proposed by Albert Wilansky, Lehigh University.*

Is this a test for convergence of $\sum_{n=1}^{\infty} a_n$? For each $\varepsilon > 0$ there exists a sequence $\{b_n\}_{n=1}^{\infty}$ with $|1 - b_n| < \varepsilon$ for all n such that $\sum_{n=1}^{\infty} a_n b_n$ is convergent.

ASSISTANT EDITORS: Clifton Corzatt and Theodore Vessey, *St. Olaf College. We invite readers to submit problems believed to be new and appealing to students and teachers of advanced undergraduate mathematics. Proposals should be accompanied by solutions, if at all possible, and by any other information that will assist the editors and referees. A problem submitted as a Quickie should have an unexpected, succinct solution. An asterisk (*) next to a problem number indicates that neither the proposer nor the editors supplied a solution.*

Solutions should be written in a style appropriate for Mathematics Magazine. Each solution should begin on a separate sheet containing the solver's name and full address.

Solutions and new proposals should be mailed in duplicate to Loren C. Larson, Department of Mathematics, St. Olaf College, Northfield, MN 55057.

Quickies

Answers appear on p. 180.

Q711. *Submitted by M. S. Klamkin, University of Alberta, Canada.*
. Determine the extreme values of the circumradii $R(\theta)$ of the set of triangles $T(\theta)$ whose sides are $\sin\theta, \cos\theta, \cos 2\theta$, for $0 < \theta < \pi/4$.

Q712. *Submitted by Barthel W. Huff, California State University, Fullerton.*
. Show that

$$\sum_{k=r}^{n} \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=r}^{n} \binom{k-1}{r-1} p^r (1-p)^{k-r}$$

for all positive integers $r \leqslant n$ and for all p.

Solutions

A Functional Differential Equation **May 1985**

***1216.** *Proposed by Stanley Rabinowitz, Merrimack, New Hampshire.*

Find all differentiable functions f that satisfy

$$f(x) = xf'\left(\frac{x}{\sqrt{3}}\right) \quad \text{for all real } x.$$

Editor's composite of partial solutions submitted by the solvers listed below.

Although repeated use of the functional relation shows that any solution function f must be infinitely differentiable on each of the open intervals $(-\infty, 0)$ and $(0, \infty)$, it is not guaranteed that f will be analytic on either of those intervals, nor even that $f''(0)$ will exist. However, if f is analytic at 0 then f will have a Maclaurin expansion

$$f(x) = \sum_{n=0}^{\infty} a_n x^n.$$

Substitution into the equation and comparison of coefficients yield

$$a_n = \frac{na_n}{\sqrt{3}^{\,n-1}} \quad \text{for} \quad n = 0, 1, 2, \ldots .$$

If $a_n \neq 0$, then $n^2 = 3^{n-1}$, and this occurs if and only if n is 1 or 3. (An easy induction shows that $3^{n-1} > n^2$ if $n > 3$.) Hence the only solutions analytic at 0 are of the form

$$f(x) = a_1 x + a_3 x^3 \quad \text{for all real } x,$$

where a_1 and a_3 are any real constants.

However, there are additional solutions. If g and h are any functions satisfying the equation on the intervals $(-\infty, 0]$ and $[0, \infty)$, respectively, then the function f defined so that $f(x)$ is $g(x)$

or $h(x)$ according as $x \leqslant 0$ or $x \geqslant 0$ will satisfy the equation on $(-\infty, \infty)$ provided that the one-sided derivatives $g'_{-}(0)$ and $h'_{+}(0)$ are equal. Such a function f is given by

$$f(x) = \begin{cases} ax + b_1 x^3 & \text{if } x \leqslant 0, \\ ax + b_2 x^3 & \text{if } x \geqslant 0, \end{cases}$$

where a, b_1, and b_2 are any constants. Note that the equation is linear, i.e., any linear combination of solutions is also a solution.

An interesting nonsolution, valid on the interval $(0, \infty)$ only, is given by $f(x) = x^r$, where r is the negative root, approximately $-.450727$, of $r^2 = 3^{r-1}$.

Solved partially by Michael H. Brill & Art Mansfield, Sheldon Degenhardt (student), Thomas E. Elsner, David Halprin (Australia), Douglas Henkin (student), R. K. Kittappa, William A. Newcomb, Bjorn Poonen (student), H. J. Seiffert (student, West Germany), Basil Skordinski, Nicholas Strauss, and the proposer.

Cumulating Rolls of a Die — May 1985

1217. *Proposed by David Callan, Lafayette College.*

A die is rolled repeatedly. Let p_n be the probability that the accumulated score is at some time equal to n. Find $\lim_{n \to \infty} p_n$.

Solution by Bjorn Poonen, student, Harvard College.

We shall consider the more general case in which the die has s sides numbered 1 through s, and is weighted so that the probability that the number i appears in a roll of the die is w_i, where $0 \leqslant w_i \leqslant 1$ and $\Sigma_{i=1}^{s} w_i = 1$. Note that w_i may be 0, but if the only i for which w_i is nonzero are multiples of an integer $j > 1$, then $\lim_{n \to \infty} p_n$ does not exist. This follows since $p_n = 0$ when $j \nmid n$, whereas $\Sigma_{i=1}^{s} p_{n+i} \geqslant 1$, since some number from $n+1$ through $n+s$ must be attained.

Now assume that the i for which w_i is nonzero have no common factor greater than 1. The coefficient of x^n in

$$\left(\sum_{i=1}^{s} w_i x^i \right)^d$$

gives the probability that the sum after d rolls of the die is exactly n. Thus

$$\sum_{n=0}^{\infty} p_n x^n = \sum_{d=0}^{\infty} \left(\sum_{i=1}^{s} w_i x^i \right)^d = \frac{1}{1 - \sum_{i=1}^{s} w_i x^i},$$

formally. If $|x| \leqslant 1$, then

$$\left| \sum_{i=1}^{s} w_i x^i \right| \leqslant \sum_{i=1}^{s} |w_i x^i| \leqslant \sum_{i=1}^{s} w_i = 1,$$

so that $\Sigma_{i=1}^{s} w_i x^i = 1$ if and only if $w_1 x, w_2 x^2, \ldots, w_s x^s$ are nonnegative real numbers, and $|x| = 1$. Because of the restriction on the nonzero w_i, this occurs only if $x = 1$. Thus the complex zeros of $Q(x) = 1 - \Sigma_{i=1}^{s} w_i x^i$ (other than 1), which we name r_i (with multiplicity s_i) for $i = 2, \ldots, m$, have absolute values greater than 1. The root 1 is simple, since $Q'(1) = [-\Sigma_{i=1}^{s}(i w_i x^{i-1})]_{x=1} = -\Sigma_{i=1}^{s}(i w_i) < 0$. Hence decomposition into partial fractions yields

$$\frac{1}{Q(x)} = \frac{1}{Q'(1)(x-1)} + \sum_{i=2}^{m} \frac{R_i(x)}{(x - r_i)^{s_i}},$$

where $\deg R_i < s_i$. Now $Q'(1)$, evaluated above, is simply $-E$, the negative of the expected value of a die roll. Thus, since the coefficient of x^n in the expansion of $1/[Q'(1)(x-1)]$ is $1/E$, we have

$$\lim_{n\to\infty} p_n = \lim_{n\to\infty}\left(\frac{1}{E} + \sum_{i=2}^{m} \frac{S_i(n)}{r_i^n}\right) = \frac{1}{E},$$

since $|r_i| > 1$. (Here $S_i(n)$ is the polynomial in n such that $S_i(n)/r_i^n$ is the coefficient of x^n in the expansion of $R_i(x)/(x-r_i)^{s_i}$.)

In the case that $s = 6$ and $w_i = 1/6$ for $1 \leqslant i \leqslant 6$ we have

$$\lim_{n\to\infty} p_n = \frac{1}{21/6} = \frac{2}{7}.$$

Also solved, or references provided, by Paul Abad & Jack Abad, Claude Belisle (student), Chris Bernhardt, Kenneth L. Bernstein, Robert E. Bernstein, J. Binz (Switzerland), Duane Broline, Roger Cuculière (France, two solutions), Robert Doucette & Bently Preece, Michael W. Ecker (two solutions), Michael V. Finn, Flinders Mathematics Endeavor Group (Australia, two solutions), Ram Prakash Gupta (Virgin Islands), Douglas Henkin (student), Víctor Hernández (Spain, two solutions), Bruce R. Johnson (Canada, three solutions), L. R. King & B. G. Klein, Kathleen Edwards Lewis, Kee-wai Lau (Hong Kong), David Mauro & Ralph Walde, Mike Molloy (student, Canada), William A. Newcomb, P. J. Pedler (Australia), Eric S. Rosenthal, Vincent P. Schielack, Jr., Allen J. Schwenk, Harry Sedinger, Bryan Shader (two solutions), Peter J. Slater (two solutions), John S. Sumner, Michael Vowe (Switzerland), William P. Wardlaw, Paul J. Zwier, and the proposer. There were three incorrect solutions.

Many solvers noted that this problem was problem 6 of the morning session of the 21st Putnam Competition (1960). (See Gleason, Greenwood, and Kelly, *The William Lowell Putnam Mathematical Competitions—Problems and Solutions: 1938–1964*, MAA, 1980, pp. 59, 522–526, 630, 635, where three solutions and further comments are provided.) It is also problem 5, pp. 136, 172, of Honsberger's *Mathematical Gems (I)*, 1973. The following intuitive argument (which can be made rigorous) was used by several solvers: if the expected value of a roll is E, then after many rolls we can expect approximately $1/E$ of the integers to be "hit." Other solvers used Markov processes or renewal theory. (See Feller, *An Introduction to Probability Theory and its Applications*, v. 1, pp. 285, 286, 291.)

Extending a Sequence — May 1985

1218. *Proposed by Vic Norton, Bowling Green State University.*

Begin with a list of n 1's. Adjoin the sum of the first two numbers in the list to the end of the list. Then adjoin the sum of the third and fourth numbers to the end of the list. Continue adjoining sums of pairs to the end of the list until no pair remains to be summed.

(a) How long is the final list?

(b) What is its last entry?

(c) What is the sum of the numbers in the final list?

I. *Solution by Jan van de Craats, Breda, The Netherlands.*

Suppose $2^{k-1} < n \leqslant 2^k$ and put $d = 2^k - n$. We extend the list to the left by starting with d pairs zero-one, followed by the initial $n - d$ ones of the original list, so that the new initial list has length $2d + (n-d) = n + d = 2^k$. The construction (summing in pairs) will then produce the remaining d ones, and from then on the rest of the list is generated as if we had started with n ones. But the extended list can be divided in a natural way into consecutive segments of lengths $2^k, 2^{k-1}, \ldots, 2, 1$, the first segment consisting of the new initial list. The sum of the terms in each segment is the same, and so the last segment (i.e., the last term of the list) has the same sum as the first segment, namely, $d + (n-d) = n$. Also, the total length of the extended list is $2^k + 2^{k-1} + \cdots + 2 + 1 = 2^{k+1} - 1$, and the sum of its terms is $(k+1)n$. Therefore, the original list (after construction) has length $2^{k+1} - 1 - 2d = 2n - 1$, and the sum of its terms is $(k+1)n - d = (k+2)n - 2^k$.

II. *Solution by D. M. Kilgour and Edward T. H. Wang, Wilfrid Laurier University, Canada.*

We solve the more general problem in which the initial n 1's are replaced by n arbitrary numbers $a_1, a_2, \ldots, a_n$. Let $f(n)$ denote the length of the final list, $l(n)$ its last entry, and $s(n)$ its sum.

(a) Each operation of adjoining the sum of two consecutive numbers to the end of the list uses two numbers and adjoins one new number. Therefore, the number of active numbers in the list (available for subsequent operations) is decreased by 1 for each operation. Hence, after $n-1$ operations, only one number remains active and no further operations are possible. Thus $f(n)=n+(n-1)=2n-1$.

(b) and (c). We consider first the case $n=2^k$. After the first 2^{k-1} operations, the original list is used and the 2^{k-1} numbers $a_1+a_2, a_3+a_4, \ldots, a_{n-1}+a_n$, whose sum is $\sum_{i=1}^n a_i$, are active. Similarly, the next 2^{k-2} operations yield 2^{k-2} numbers with the same sum, and so on, until the last, or $(n-1)$th, operation produces the single number $\sum_{i=1}^n a_i$. It follows that

$$l(n)=\sum_{i=1}^{n} a_i \quad \text{and} \quad s(n)=(k+1)\sum_{i=1}^{n} a_i \quad \text{if } n=2^k. \tag{1}$$

Now let $n=2^k+q$, where $1 \leqslant q < 2^k$. Since each operation decreases the number of active elements by 1, precisely 2^k active numbers remain after q operations. Since $2q<2^k+q$, these active numbers are $a_{2q+1}, a_{2q+2}, \ldots, a_n, a_1+a_2, \ldots, a_{2q-1}+a_{2q}$, whose sum is $\sum_{i=1}^n a_i$. Hence from (1), we infer that

$$l(n)=\sum_{i=1}^{n} a_i \quad \text{and} \quad s(n)=\sum_{i=1}^{2q} a_i+(k+1)\sum_{i=1}^{n} a_i \quad \text{if } n=2^k+q. \tag{2}$$

If $\sum_{i=1}^{2q} a_i$ is defined to be 0 when $q=0$, then (2) holds even if $n=2^k$.

In particular, if $a_1=a_2=\cdots=a_n=1$, then $f(n)=2n-1$, $l(n)=n$, and $s(n)=(k+1)n+2q$, where $n=2^k+q$ and $0 \leqslant q < 2^k$.

Also solved by Anders Bager (Denmark), Samuel F. Barger, Rich Bauer, Kenneth L. Bernstein, Michael H. Brill (part (b) only), Duane Broline, Rickie H. Chase, Samuel Chort, George Crofts, Charles R. Diminnie, Lorraine L. Foster (two solutions), John P. Georges & Ralph E. Walde, David Grabiner (student), Ralph P. Grimaldi, Jerrold W. Grossman, JoAnne S. Growney & Seymour Schwimmer, Hans Kappus (Switzerland), L. Kuipers (Switzerland), Kathleen Edwards Lewis, Mike Molloy (student, Canada), William A. Newcomb, P. J. Pedler (Australia), Bjorn Poonen (student), Ray Rosentrater, Allen J. Schwenk, John S. Sumner, Michael Vowe (Switzerland, parts (a) and (b)), William P. Wardlaw, Herbert Wills IV, Paul J. Zwier, and the proposer. There were two solutions without proof and one incorrect solution.

The more general result obtained in solution II can be obtained by the method of solution I if $a_{d+1}, 0, a_{d+2}, 0, \ldots, a_n, 0$ is prefixed to the original list. *Foster, Wardlaw, and Zwier* noted that a further generalization is possible: addition can be replaced by any commutative and associative binary operation. In fact, if only associativity is assumed, then the method of solution II can be applied to obtain the same result for $f(n)$ and $s(n)$, if $\sum_{i=1}^n a_i$ is interpreted as $a_1 * a_2 * \cdots * a_n$, where $*$ is the operation replacing $+$, and $(k+1)t$ is interpreted as $t*t*\cdots*t$ with k operations. However, $l(n)$ becomes $\sum_{i=2q+1}^n a_i + \sum_{i=1}^{2q} a_i$. (If $q=0$, the first "sum" is omitted.)

Two Series Involving Zeta Function Values — May 1985

1219. *Proposed by L. Matthew Christophe, Jr., Wilmington, Delaware.*

Sum the infinite series

$$\sum_{k=2}^{\infty}\left(\frac{(-1)^k}{k+1}(\zeta(k)-1)\right) \quad \text{and} \quad \sum_{k=2}^{\infty}\left(\frac{1}{k+1}(\zeta(k)-1)\right).$$

where $\zeta(k)=\sum_{n=1}^{\infty} n^{-k}$ (the Riemann zeta function).

I. *Solution by Michael V. Finn, Annandale, Virginia* (rearranged by the editor).

From the definition of the zeta function we have

$$\sum_{k=2}^{\infty}\frac{(\mp 1)^k}{k+1}(\zeta(k)-1)=\sum_{k=2}^{\infty}\left(\frac{(\mp 1)^k}{k+1}\sum_{n=2}^{\infty}\frac{1}{n^k}\right)=\sum_{n=2}^{\infty}\sum_{k=2}^{\infty}\frac{(\mp 1)^k}{(k+1)n^k},$$

where the interchange of order of summation is permitted since the double series is absolutely convergent. In fact, from

$$\sum_{n=2}^{\infty}\frac{1}{n^k}\leqslant\int_1^{\infty}x^{-k}\,dx=\frac{1}{k-1}\quad\text{for } k\geqslant 2$$

it follows that

$$\sum_{k=2}^{\infty}\left(\frac{1}{k+1}\sum_{n=2}^{\infty}\frac{1}{n^k}\right)\leqslant\sum_{k=2}^{\infty}\frac{1}{k+1}\cdot\frac{1}{k-1}=\sum_{k=2}^{\infty}\frac{1}{k^2-1},$$

the last series being obviously convergent.

By rearranging the Maclaurin series for $\ln(1+x)$ we find

$$\sum_{k=2}^{\infty}\frac{(-1)^k x^k}{k+1}=\frac{x}{2}-1+\frac{1}{x}\ln(1+x)\quad\text{for } x\in(-1,0)\cup(0,1],$$

and for $n\geqslant 2$ the substitution $x=\pm 1/n$ yields

$$\sum_{k=2}^{\infty}\frac{(\mp 1)^k}{(k+1)n^k}=\pm\frac{1}{2n}-1\pm\ln\left(1\pm\frac{1}{n}\right)^n.$$

For the upper signs, if M is a positive integer, we obtain

$$\begin{aligned}\sum_{n=2}^{M}\left(\frac{1}{2n}-1+\ln\left(1+\frac{1}{n}\right)^n\right)&=\frac{1}{2}\sum_{n=1}^{M}\frac{1}{n}-\frac{1}{2}-(M-1)+\ln\frac{(M+1)^M}{2(M!)}\\&=\frac{1}{2}\left(\sum_{n=1}^{M}\frac{1}{n}-\ln M\right)+\frac{1}{2}+\ln\left(1+\frac{1}{M}\right)^M\\&\quad-\frac{1}{2}\ln(8\pi)+\ln\frac{M^Me^{-M}\sqrt{2\pi M}}{M!}.\end{aligned}$$

Taking the limit as $M\to\infty$, using the definitions of Euler's constants γ and e together with Stirling's formula, we obtain

$$\sum_{k=2}^{\infty}\frac{(-1)^k}{k+1}(\zeta(k)-1)=\frac{1}{2}\gamma+\frac{1}{2}+\ln e-\frac{1}{2}\ln(8\pi)+\ln 1=\frac{1}{2}(\gamma+3-\ln(8\pi)).$$

For the lower signs, if M is a positive integer, we obtain

$$\begin{aligned}\sum_{n=2}^{M}\left(-\frac{1}{2n}-1-\ln\left(1-\frac{1}{n}\right)^n\right)&=-\frac{1}{2}\sum_{n=1}^{M}\frac{1}{n}+\frac{1}{2}-(M-1)+\ln\frac{M^{M+1}}{M!}\\&=-\frac{1}{2}\left(\sum_{n=1}^{M}\frac{1}{n}-\ln M\right)+\frac{3}{2}-\frac{1}{2}\ln(2\pi)+\ln\frac{M^Me^{-M}\sqrt{2\pi M}}{M!},\end{aligned}$$

and taking the limit as $M\to\infty$ yields

$$\sum_{k=2}^{\infty}\frac{1}{k+1}(\zeta(k)-1)=-\frac{1}{2}\gamma+\frac{3}{2}-\frac{1}{2}\ln(2\pi)+\ln 1=\frac{1}{2}(3-\gamma-\ln(2\pi)).$$

II. *Solution by John A. Crow, Hughes Aircraft Company.*

Let

$$S_1=\sum_{n=2}^{\infty}\frac{(-1)^n(\zeta(n)-1)}{n+1}\quad\text{and}\quad S_2=\sum_{n=2}^{\infty}\frac{\zeta(n)-1}{n+1}.$$

By using the similar series 6.1.33 in Abramowitz & Stegun, *Handbook of Mathematical Functions*, 1972, p. 256, we have

$$\log\Gamma(2+z)=\log(1+z)+\log\Gamma(1+z)=(1-\gamma)z+\sum_{n=2}^{\infty}\frac{(-1)^n(\zeta(n)-1)z^n}{n}\quad\text{for } |z|<2; \tag{1}$$

here γ is Euler's constant. Then simple integrations yield

$$\int_0^1\log\Gamma(2+z)\,dz=\frac{1}{2}(1-\gamma)+T_1\quad\text{and}\quad\int_{-1}^0\log\Gamma(2+z)\,dz=-\frac{1}{2}(1-\gamma)+T_2,$$

where

$$T_1=\sum_{n=2}^{\infty}\frac{(-1)^n(\zeta(n)-1)}{n(n+1)}\quad\text{and}\quad T_2=\sum_{n=2}^{\infty}\frac{\zeta(n)-1}{n(n+1)}.$$

From 6.441.1 of Gradshteyn & Ryzhik, *Table of Integrals, Series, and Products*, 1980, p. 661, we obtain

$$\int_0^1\log\Gamma(2+z)\,dz=\frac{1}{2}\log(2\pi)+2\log 2-2\quad\text{and}\quad\int_{-1}^0\log\Gamma(2+z)\,dz=\frac{1}{2}\log(2\pi)-1,$$

from which it follows that

$$T_1=\frac{1}{2}\log(2\pi)+2\log 2-2-\frac{1}{2}(1-\gamma)\quad\text{and}\quad T_2=\frac{1}{2}\log(2\pi)-1+\frac{1}{2}(1-\gamma).$$

From (1) we know that (with $z=\pm1$)

$$U_1=\sum_{n=2}^{\infty}\frac{(-1)^n(\zeta(n)-1)}{n}=\log 2-(1-\gamma)\quad\text{and}\quad U_2=\sum_{n=2}^{\infty}\frac{\zeta(n)-1}{n}=1-\gamma.$$

Finally, S_1 and S_2 are given by

$$S_1=U_1-T_1=\frac{1}{2}(3+\gamma-\log(8\pi))\quad\text{and}\quad S_2=U_2-T_2=\frac{1}{2}(3-\gamma-\log(2\pi)).$$

Also solved by Rich Bauer, Chico Problem Group, Douglas Henkin (student), Víctor Hernández (Spain), Hans Kappus (Switzerland), R. K. Kittappa, L. Kuipers (Switzerland, first sum only), Kee-wai Lau (Hong Kong), Syrous Marivani, Vania Mascioni (student, Switzerland), William A. Newcomb, P. J. Pedler (Australia), Bjorn Poonen (student), H. Roelants (Belgium), Volkhard Schindler (East Germany), Heinz-Jürgen Seiffert (student, West Germany), Robert E. Shafer, Michael Vowe (Switzerland), and the proposer. Late solution by H. M. Srivastava (Canada). There were two incorrect solutions.

The Chico Problem Group found the problem of evaluating U_2 (solution II above) to be problem E1736 in the *Monthly* (solution v. 72 (1965) 1023), and several other solvers evaluated related sums.

A Problem of Divisibility — May 1985

1220. *Proposed by Gregg Patruno, student, Princeton University.*

Prove that if the odd prime p divides a^b-1, where a and b are positive integers, then p appears to the same power in the prime factorization of $b(a^d-1)$, where d is the greatest common divisor of b and $p-1$.

Solution by Lorraine L. Foster, California State University, Northridge.

We use $p^n\|m$ to indicate that p^n is the highest power of the odd prime p which divides the integer m. Furthermore, we let a be any positive or negative integer different from ±1.

LEMMA. *Let $p^k\|(a^s-1)$, where $k\geqq 1$. If $p^r\|u$, then $p^{k+r}\|(a^{su}-1)$.*

Proof. Write $a^s=1+mp^k$, where $p\nmid m$.

(i) If t is any positive integer such that $p \nmid t$, then $a^{st} = (1 + mp^k)^t = 1 + mtp^k + vp^{2k} \equiv 1 + (mt)p^k \pmod{p^{k+1}}$ for some integer v. Since $p \nmid (mt)$, we have $p^k \| (a^{st} - 1)$.

(ii) $a^{sp} = (1 + mp^k)^p = 1 + mp^{k+1} + m^2(p-1)p^{2k+1}/2 + vp^{3k} \equiv 1 + mp^{k+1} \pmod{p^{k+2}}$ for some integer v. Hence $p^{k+1} \| (a^{sp} - 1)$.

(iii) Induction on r quickly settles the general case; the basic cases $r = 0$ and $r = 1$ are merely cases (i) and (ii). □

Now let $p^h \| (a^b - 1)$, $p^r \| b$, and $p^k \| (a^d - 1)$, where $h > 0$, $r \geqq 0$, and $k \geqq 0$. Since $d = \gcd(b, p-1) = bx + (p-1)y$ for some integers x and y, the Fermat-Euler theorem yields $a^d \equiv 1 \pmod{p}$, so that $k \geqq 1$. Now $p \nmid d$, since $d | (p-1)$. Since also $d | b$, we may write $b = ud$, where $p^r \| u$. Thus by the lemma we have $p^{k+r} \| (a^b - 1)$, so that $h = k + r$ and we are finished.

Also solved by Hugh M. Edgar, Syrous Marivani, Bjorn Poonen (student), Lawrence Somer, J.M. Stark, William P. Wardlaw, and the proposer. There were four incorrect or incomplete solutions.

The lemma in the above solution follows easily from Lemma 1 of Artin, "The orders of the linear groups," *Comm. Pure. Appl. Math.*, 8 (1955) 355–366.

Comments

936 (proposed March 1975; solution March 1976).

A. Oppenheim (formerly in Nigeria, now in England), *Paul Erdős* (Hungary), *H. S. M. Coxeter* (Canada), and *Clayton Dodge*, in private correspondence with *Leon Bankoff*, have pointed out a subtle error in the published solution. (The problem was to show that in triangle ABC we have, in the usual notation, $t_a + t_b + m_c \leqslant s\sqrt{3}$, with equality if and only if $a = b = c$.) The difficulty lies in the following argument: "Since $4bc \leqslant (b+c)^2$ with equality if and only if $b = c$ (with a similar argument connecting a and c), it is apparent that $m_a[4bc/(b+c)^2] + m_b[4ac/(a+c)^2] + m_c \leqslant m_a + m_b + m_c$, equality holding only when $a = b = c$, that is, when the triangle is equilateral and $m_a + m_b + m_c = s\sqrt{3}$. *Consequently* [editor's italics] the left side of the inequality attains its maximum value of $s\sqrt{3}$ only when $m_a = m_b = m_c$ and remains less than $s\sqrt{3}$ when the triangle is not equilateral." *Clayton Dodge* pointed out that the argument is similar to the argument that if $f(x) \leqslant g(x)$ for all x, with equality only when $x = a$, then $f(x) \leqslant g(a)$ [*sic*!] for all x; a counterexample is obtained by setting $f(x) = x^2(1 - x^2)$ and $g(x) = x^2$ for all x, and $a = 0$.

However, the conclusion of problem 936 is correct, and a proof using a combination of calculus and computer is given in G. S. Lessels and M. J. Pelling, "An inequality for the sum of two angle bisectors and a median," *Univ. Beograd Publ. Elecktrotehn. Fak.* ser. Mat. Fiz. no. 590 (1977) 59–62. A strengthened result,

$$h_a + t_b + m_c \leqslant t_a + t_b + m_c \leqslant \sqrt{s(s-a)} + \sqrt{s(s-b)} + m_c \leqslant s\sqrt{3},$$

with equality if and only if $a = b = c$, is proved in the same journal, no. 686 (1980) 45–47, "The triangle inequality of Lessells and Pelling," by B. E. Patuwo, R. S. D. Thomas, and Chung-Lie Wang. They noted that $t_a = 2\sqrt{bcs(s-a)}/(b+c) \leqslant \sqrt{s(s-a)}$ (equality only if $b = c$) by the harmonic and geometric inequality, and then, following Lessells and Pelling in setting $s = 1$, $x = 1 - a$, and $y = 1 - b$, they reduced the problem to showing that

$$g(x, y) = \sqrt{x} + \sqrt{y} + \sqrt{1 - x - y + \frac{1}{4}(x-y)^2} \leqslant \sqrt{3} \quad \text{for } x,\ y, \text{ and } x + y \text{ in } [0,1],$$

with equality if and only if $x = y = 1/3$. But the points (x, y) of interest are covered by the parabolic arcs $\sqrt{x} + \sqrt{y} = \sqrt{d}$ with $0 \leqslant d \leqslant 2$. Simple algebra now shows that if (x, y) is on such an arc, then

$$g(x, y) = \sqrt{d} + \sqrt{1 - \frac{d}{2} - \frac{2-d}{4d}(x-y)^2},$$

which for fixed d in $(0,2)$ shows that $g(x, y)$ is maximized when $x = y$. (There is only one point of interest if $d = 0$ or $d = 2$.) But elementary calculus shows that $g(t, t)$ for $0 \leqslant t \leqslant 1/2$ has a unique maximum at $t = 1/3$, which completes the proof.

Since $h_a \leqslant t_a$, this problem is a strengthening of MONTHLY problem E2504 (proposed v. 81 (1974) 1111; solution v. 83 (1976) 289–290). However, a further strengthening to $t_a + m_b + m_c \leqslant s\sqrt{3}$, or even to $h_a + m_b + m_c \leqslant s\sqrt{3}$, is false, as may be seen by using an isosceles triangle with $b = c = 1$ and base angle θ. If we set $f(\theta) = h_a + m_b + m_c - s\sqrt{3} = t_a + m_b + m_c - s\sqrt{3} = \sin\theta + \sqrt{1 + 8\cos^2\theta} - (1 + \cos\theta)\sqrt{3}$, then $f(0) = 3 - 2\sqrt{3} < 0$, $f(\pi/3) = 0$, and $f(\pi/2) = 2 - \sqrt{3} > 0$. Oppenheim and Wang (private communications) replied to the editor's query by showing that f is increasing on $[0, \pi/2]$.

1163 (proposed January 1983; solution March 1984).

An inquiry from a reader prompted a search of the solution file for a bijection f from R onto R which not only is continuous at some number b such that f^{-1} is discontinuous at $f(b)$, but also is differentiable at b. *Jerrold W. Grossman* (Oakland University) and *William A. Newcomb* (Lawrence Livermore National Laboratory) independently provided such an example, without stating so.

If $0 \leqq x < 1$ and x has the representation $\Sigma_{i=1}^{\infty}(a_i/r^i)$ in radix r, where each integer a_i is in $[0, r)$—choose either representation if x has two—, let $f(x) = \Sigma_{i=1}^{\infty}(a_i/r^{2i})$; if $-1 < x < 0$, let $f(x) = -f(-x)$; and complete f in any manner to a bijection from R onto R. Then $f(0) = 0$, f is continuous at 0, and f^{-1} is discontinuous at 0. Furthermore, for each positive integer j, if $r^{-j} \leqq x < r^{-j+1}$, then $0 < (f(x))/x \leqq r^{-j}$. Hence $f'(0) = 0$. By using $f(x) = \Sigma_{i=1}^{\infty}(a_i/r^{2ni})$ instead, we can achieve n-times differentiability at 0, and by using $f(x) = \Sigma_{i=1}^{\infty}(a_i/r^{i^2})$, we can achieve infinite differentiability at 0. Most solvers constructed functions f for which there is a positive real number δ such that $|f(x)| \leqq |x|$ whenever $|x| < \delta$. It is then easy to construct an infinitely differentiable example F by setting $F = g \circ f$, where $g(0) = 0$, $g(x) = \exp(-1/x^2)$ if $0 < x \leqq 1$, $g(x) = x/e$ if $x > 1$, and $g(x) = -g(-x)$ if $x < 0$.

Late solution by *Edilio Escalona F.* (*Venezuela*), buried in editor's files.

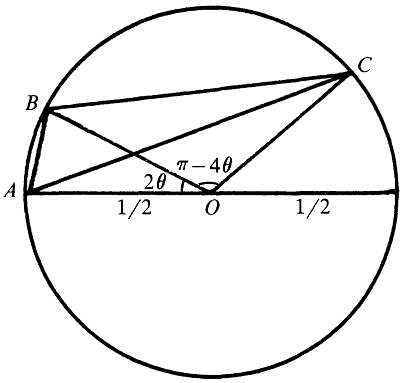
Answers

Solutions to the Quickies on p. 173.

A711. First one should verify that the triangles $T(\theta)$ actually exist for all θ in the given interval. This will follow from the elementary inequality $\cos 2\theta + \sin\theta > \cos\theta$. However, it will also follow from the subsequent geometry.

For $\theta = 0$, we get a degenerate triangle of sides $0, 1, 1$ whose circumradius is $R = 1/2$. For $\theta = \pi/4$, we get a degenerate triangle of sides $1/\sqrt{2}$, $1/\sqrt{2}$, 0 with $R = 1/2\sqrt{2}$. So it may appear that $1/2\sqrt{2} < R(\theta) < 1/2$. However, we will show that $R(\theta) = 1/2$ for all θ in the open interval $(0, \pi/4)$.

Consider a triangle ABC inscribed in a circle of radius $1/2$ as shown, where AB and BC subtend angles of 2θ and $\pi - 4\theta$, respectively, at the center O. Here, $AB = \sin\theta$, $BC = \cos 2\theta$, and $AC = \cos\theta$.



A712. Suppose $0 \leqslant p \leqslant 1$ and we have a sequence of independent Bernoulli trials with probability of success p. The left-hand expression gives the probability of at least r successes in n trials (binomial distribution) and the right-hand expression gives the probability that the rth success occurs in the rth through nth trials (negative binomial distribution). Obviously, the two events are the same and the probabilities are equal.

The polynomials are equal for all p since they agree on an interval.

REVIEWS

Paul J. Campbell, Editor
Beloit College

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.

Kac, Mark, *Enigmas of Chance*, Harper & Row, 1985; xxvii + 163 pp, $18.95.

Mark Kac (1914-1984) was a probabilist, one of the founders of the subject. He was also a Pole--a student of Hugo Steinhaus--who taught at Cornell and at Rockefeller. He wrote a little about those places, but the scenes that stick in one's mind are the descriptions of pre-War Poland and his mathematical education there. Virtually all his family and friends perished at the hands of the Nazis.

Crypton, Dr., How to win at Monopoly, *Science Digest* (September 1985) 66-71.

Oh-oh! The secret's out--the way to win at Monopoly is to play the percentages. And how are the latter calculated? By calculating the steady-state for a 123-state Markov chain. (See also "Monopoly as a Markov process," by Robert B. Ash and Richard L. Bishop, *Mathematics Magazine* 45 (1972) 26-29--i.e., you saw it here first!)

Alexanderson, G.L., L.F. Klosinski, and L.C. Larson, *The William Lowell Putnam Mathematical Competition Problems and Solutions: 1965-1984*, MAA, 1985; xii + 147 pp, $24.00.

At last the recent Putnam problems and solutions are all collected together. However, unlike the earlier MAA volume covering 1938-1964, which followed up on problems to compile better solutions, this book simply reprints material from the *Monthly* and *Mathematics Magazine*. But we're glad not to have to wait 20 more years!

Johnson, Deborah G., and Snapper, John W., *Ethical Issues in the Use of Computers*, Wadsworth, 1985; ix + 363 pp.

Selected readings, including essays on responsiblity and liability, privacy and security, computers and power, and software as property (including texts of some U.S. court decisions), and the codes of conduct of computer societies. Eminently suitable as a text for a thought-provoking course on ethics of computer use, this book is an important resource for *all* courses in computer science.

Lawler, E.L., et al. (eds.), *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization*, Wiley, 1985; x + 465 pp.

Collection of 12 essays, each on a different aspect of the traveling salesman problem. A knowledge of graph theory and linear programming should suffice; exercises are included.

Kemp, Rainer, *Fundamentals of the Average Case Analysis of Particular Algorithms*, Wiley-Teubner, 1984; viii + 233 pp.

Excellent text for a course in analysis of algorithms (as opposed to a survey of algorithms) for students well-versed in undergraduate mathematics. Appendices summarize fundamental facts about specific mathematics needed (discrete math, plus Dirichlet series, some complex analysis, and the Euler summation formula). "...[T]he functions appearing in a detailed analysis of algorithms are sometimes very mysterious... They necessarily involve a knowledge of calculus or higher mathematics." Caution: Computer science students with only the minimum ACM-recommended mathematics will be woefully underprepared for this book.

Dudewicz, Edward J., and Karian, Zaven A., *Tutorial: Modern Design and Analysis of Discrete-Event Computer Simulations*, IEEE Computer Society Press, 1985; ix + 475 pp.

"Simulation has become the most widely used management science/operations research technique.... The object of this tutorial is to provide a working understanding of the design, implementation, and analysis of computer simulations." The tutorial consists of reprinted articles on random number generation and testing generators, transformation to other statistical distributions, and efficiency, and applications of simulations. Additional sections of original contributions cover fitting distributions to data and algorithms for sorting, plus an appendix reviewing statistics. (The 100-plus pages of the appendix, plus the section on sorting, could have been left out without loss, as sources for these are readily available.)

Gorenstein, Daniel, The enormous theorem, *Scientific American* 253:6 (December 1985) 104-115, 150.

"How could a single mathematical theorem require 15,000 pages to prove?" The unprecedented length of the classification theorem for finite simple groups can be traced to "the complexity of the subgroup structure of an arbitrary finite group compared with that of a simple group--just as a complex molecule has a far more tangled internal structure than a single atom does." What's ahead? A "second-generation" proof, slimmed down to a mere 3000 pages. And what is the impact of the result? Already there have been applications in various areas, as well as tantalizing connections to the theory of elliptic functions.

Dembart, Lee, Supercomputer comes up with whopping prime number, *Los Angeles Times* (17 September 1985) Part I, pp. 3, 19.

How else to test a new $10 million Cray X-MP supercomputer, but have it look for Mersenne primes? Scientists at Chevron thus discovered the 30th Mersenne prime and largest known prime, $2^{216,091}-1$.

Ifrah, Georges, *From One to Zero: A Universal History of Numbers*, Viking, 1985; xvi + 503 pp, $35.

Splendid history of numbers and numerals, chock-full of illustrations, and with even more detail than Karl Menninger's *Number Words and Number Symbols*. Unlike John N. Crossley in *The Emergence of Number*, Ifrah concentrates on number symbols. Despite some slips (e.g., many people *can* assess at a glance--without counting--how many objects they are shown, up to 40 or so), the omission of some intriguing theories (such as A. Seidenberg's on the ritual origin of counting), and the lack of discussion of number words, this is a fine work of lasting value. It is particularly strong on the history of Sumerian numerals. Valuable charts at the end of the volume show the evolution of both European and Eastern Arabic numerals.

Barnett, Arnold, Misapplications reviews: speed kills?, *Interfaces* 16:2 (March-April 1986) 63-68.

Is the 55 mph speed limit in the U.S. responsible for the 16% drop in road deaths? An article in *Road and Track* attributed the drop to three other causes; author Barnett here demolishes all three with very perceptive comments on the practical and statistical aspects involved. The column concludes with discussion of a different question, with a proposal that could reduce by half the rate of U.S. jetliner crashes.

Crypton, Dr., The limits of mathematical knowledge, *Science Digest* (March 1986) 72-75.

Can mathematics help Willy Loman after all? There is some encouraging news on the traveling-salesman problem and other NP-complete problems. Unfortunately for Willy, the news that is exciting mathematicians is bad news for Willy: there is mounting evidence that NP-complete problems are intractable. The new results show that logic circuits for certain computations grow exponentially large with problem size.

Kolata, Gina, New test finds "certified" primes in record time, *Science* 231 (31 January 1986) 452-453.

A new test for primality has been found. It is a probabilistic algorithm that is always correct, almost always runs in polynomial time, but is not yet fast enough to be practical. At worst, it *could* run as slowly as the algorithm of Adleman-Pomerance-Rumely, on a few primes. The proof of the new result of S. Goldwasser and J. Kilian (MIT) involves elliptic curves.

Kolata, Gina, Shakespeare's new poem: an ode to statistics, *Science* 231 (24 January 1986) 335-336; Letters, (21 March 1986) 1355.

Statistical analysis of word frequencies in a newly-discovered poem ascribed to Shakespeare concludes: "There is no convincing evidence for rejecting the hypothesis that Shakespeare wrote it." Put otherwise: the poem "fits Shakespeare as well as Shakespeare fits Shakespeare."

Schroeder, Manfred R., Auditory paradox based on fractal waveform, *Journal of the Acoustical Society of America* 7 (1986) 186-189.

Schroeder exploits the Weierstrass continuous nowhere-differentiable function to devise a curious auditory paradox: a waveform that, when played back on tape at double speed, sounds to the human ear a semitone lower (rather than an octave higher)! Such are the paradoxes of self-similarity of fractals.

Kolata, Gina, Solving knotty problems in math and biology, *Science* 231 (28 March 1986) 1506-1508.

Vaughan Jones (UC Berkeley) and Joan Birman (Columbia) have shown that the seemingly-unrelated theory of Von Neumann algebras provides a way to tell certain knots apart. Moreover, biologists are interested in the result, as it may help them understand knotting of DNA during replication and recombination. Jones has devised a new knot polynomial that--unlike the famous Alexander polynomial--can tell right-handed and left-handed overhand knots apart. More importantly for the biologist, the basic operations on DNA structure correspond exactly to the mathematical operations that Jones's polynomial uses.

Sharenow, Ira, Baseball strategy: an operations research approach, *OIRSA Student Communications* 4:2, 3-5.

Summary in dialogue of known results from operations research on baseball strategy. Is a sacrifice fly a good strategy? How about stealing a base? An intentional walk? Here's where to find out.

Kolata, Gina, What does it mean to be random?, *Science* 231 (7 March 1986) 1068-1070.

Persi Diaconis (Stanford) has tackled head-on the problem of what it means to be random. His theory of "randomness multipliers" explains why frequentists and subjectivists so often agree, and captures what is random about such physically determinate experiments as flipping a coin or spinning a roulette wheel. Diaconis uses the notion of a probability distribution having a "depth." His theory allows him to define chaos quantitatively. "In chaos, a little bit of uncertainty in initial conditions is quickly and enormously magnified.... To say a system is chaotic to a particular degree means that it is a specific distance from random after a specific number of iterations."

Shore, John, *The Sachertorte Algorithm and Other Antidotes to Computer Anxiety*, Viking Penguin, 1985; xvi + 269 pp, $6.95 (P).

Exceptionally well-written, wise, and insightful introduction to the nature of computers and their place in human affairs. This is a prime source to recommend to neighbors, students, and deans who have so far avoided learning about computers.

Peterson, Ivars, Acoustic residues: number theory, the paradigm of pure mathematics, helps change the sound of small rooms and concert halls, *Science News* 129 (4 January 1986) 12-13.

Concert-goers prefer to hear somewhat different signals at their two ears, and hence prefer long narrow halls over wide shallow ones. The latter can be improved by using a grating to redirect sound from the ceiling toward the walls. How deep the grating's notches should be comes from number theory--specifically, quadratic residues.

Peterson, Ivars, The troubled state of calculus: A push to revitalize college calculus teaching has begun, *Science News* 129 (5 April 1986) 220-221.

Discussions of discrete mathematics in the curriculum have had the fruitful side effect of provoking a reexamination of calculus courses. A Sloan Foundation conference earlier this year brought to the surface many exciting possibilities for change. Surprisingly, curricular inertia at colleges may be upstaged by faster change in high-school calculus, via changes in the syllabus for the Advanced Placement exam.

Strang, Gilbert, *Introduction to Applied Mathematics*, Wellesley-Cambridge Pr (Box 157, Wellesley, MA 02181), 1986; x + 758 pp, $39.00.

An earlier book by Gilbert Strang (*Linear Algebra and Its Applications*) offered a revolution in the teaching of linear algebra: a contemporary viewpoint, focus on the fundamentals of projection and orthogonality, emphasis on linear algebra (as opposed to matrix) applications, and abandonment of theorem-proof style for open prose exposition. His new book breathes a fresh spirit into applied mathematics. It combines the discrete with the continuous (e.g. Kirchoff's laws lead in the continuous case to the curl and the divergence), the algorithms (and numerical analysis) with the theory. Starting from linear systems, the book treats discrete and continuous equilibrium (including boundary-value problems), the Kalman filter, Fourier and complex methods, numerical methods (including finite element and fast Fourier transform), initial-value problems (with stability, chaos, solitons, and the Laplace transform), network flows and combinatorics, and optimization (including Karmarkar's algorithm). The author's synthesizing goal of showing a few ideas shared by a wide range of applications is magnificently achieved. Students will need a good background in linear algebra, and some background in differential equations would be helpful. The book is suitable for a year's course in applied mathematics, engineering mathematics, or "advanced calculus."

Walker, Jearl, *Roundabout: the Physics of Rotation in the Everyday World*, Freeman, 1985; viii + 70 pp, 10.95 (P).

Amusement park rides, racquetball, billiards, judo, ballet, tops, boomerangs and Frisbees: Walker has collected here his splendid *Scientific American* "Amateur Scientist" columns on rotating objects. His analysis is mostly qualitative (nary an equation appears) and entirely delightful.

Parker, Donn B., *Ethical Conflicts in Computer Science and Technology*, AFIPS Pr, 198 ; vi + 201 pp.

Forty-seven scenarios, with opinions of a panel on the ethics of practices involving computers. A summary is provided of the issues exposed, and appendices contain conduct codes of several computer societies. The panel's votes reveal great divisions of opinion; the voting itself leaves the unfortunate impression that ethics is a matter of majority vote. (But see "The evolution of ethics" by M. Ruse and E. O. Wilson, *New Scientist* (17 October 1985) 50-52.) However, this book, with its accompanying workbook, is a very valuable resource for the moral education and character formation of prospective computer professionals.

Problem solving: Evolve your answer, *The Economist* (23 November 1985) 92-93.

How about using the ideas of evolution to solve the traveling salesman problem? "Evolution is about finding better ways of doing things. First, random changes are made to the ways creatures solve problems. Second, natural selection kills off the less successful problem solvers. Third, the survivors pass on their tactics to the next generation." So one can try to find an optimum by perturbing a near-optimum. "But what really speeded up evolution was sex. Sex ... throws together new combinations, thus increasing the chances of escaping from a local optimum and hitting on a really successful innovation (or a catastrophic failure)." So Dr. [Robert] Brady [of Cambridge Univ.] has been trying to make his problem-solving programs sexual. The most favored solvers get to "shuffle their genes with others." The method seems to work, as it has produced better and faster solutions to the traveling salesman problem for 64 cities.

Kenner, Hugh, Neatness doesn't count after all, *Discover* (April 1986) 86-93.

A common rule of thumb in office filing is the 80-20 rule: 80% of the action involves 20% of the files. The rule extends to other realms of human endeavor, and author Kenner cites many; 40 different words make up 40% of all the words in Shakespeare. The rule is an instance of Zipf's law, which "says that *any* allocation of resources (people in cities, words in books, tools in a toolbox) will settle down to a harmonic arrangement"--"harmonic" as in "harmonic sequence," so that the nth largest city contains $1/n$ times the population of the largest.

Wagon, Stan, *The Banach-Tarski Paradox*, Cambridge U Pr, 1985; xvi + 251 pp, $37.50.

"The Banach-Tarski Paradox is a striking mathematical construction: it asserts that a solid ball may be taken apart into finitely many pieces that can be arranged using rigid motions to form a ball twice as large as the original." This is a detailed exposition of measure-theoretic paradoxes associated with the Axiom of Choice; it presumes that the reader is familiar with Lebesgue measure, Borel sets, the Axiom of Choice itself, and the like.

Fernie, J. Donald, Marginalia: Candid posterity and the Englishman, *American Scientist* 73:5 (October-November 1985) 471-473; II, 74:1 (January-February 1986) 55-58.

"His misfortune ... was that 'he had to shine in the same sky as that which was illuminated by the unparalleled genius of Newton.'" This Englishman? Edmond Halley. The first of these articles relates his scientific endeavors, the second his nonscientific activities.

NEWS & LETTERS

USCMI PRE-CONGRESS SESSION OF SURVEY TALKS

On August 2, 1986, the United States Commission on Mathematical Instruction will sponsor a series of invited survey talks aimed at enhancing understanding and appreciation of some of the major research-related work which will be discussed at the International Congress of Mathematicians meeting in Berkeley August 3-11. These survey talks will take place from 2 to 6 p.m. in Wheeler Auditorium on the campus of the University of California, Berkeley. There is no registration fee for this USCMI Pre-Congress Session.

The speakers and their topics follow:

2:00 p.m.	Robert Edwards, "Highlights of Low Dimensional Topology"
3:00 p.m.	Richard Karp, "The Polynomial-Time Frontier: Recent Developments in Computational Complexity Theory"
4:00 p.m.	Clifford Taubes, "The Physics and Geometry of Estimates in Nonlinear Partial Differential Equations"
5:00 p.m.	Andrew Ogg, "Modular Functions and Number Theory"

For further information, contact the session organizer:

Warren Page
New York City Technical College
300 Jay Street
Brooklyn, New York 11201

Telephone: (718) 643-3637 or
(718) 643-2470 .

DISCRETE MATHEMATICS CONFERENCE

The Fourteenth Annual Mathematics and Statistics Conference at Miami University, Oxford, Ohio, will be held October 3 and 4, 1986. The theme for this year's conference will be "Discrete Mathematics". Featured speakers will include Ron Graham, AT&T Bell Laboratories; Alan Tucker, SUNY, Stony Brook; and Herbert Wilf, University of Pennsylvania. There will be contributed paper sessions which should be suitable for a diverse audience of mathematicians, statisticians, and students. Abstracts should be sent by June 3, 1986 to

Professor Gary G. Gilbert
Department of Mathematics and Statistics
Miami University
Oxford, Ohio 45056

Information regarding preregistration and housing may also be obtained from Professor Gilbert.

STUDENT CONFERENCE

The Ohio Delta Chapter of Pi Mu Epsilon will hold its thirteenth annual Student Conference October 3, 4, 1986. Undergraduate mathematics and statistics students are invited to contribute papers, and should send abstracts by September 25, 1986 to

Professor Milton Cox
Department of Mathematics and Statistics
Miami University
Oxford, Ohio 45056 .

CUT HERE

JOIN MAA TODAY!

Send this application to: MAA Membership Department
1529 18th Street, N.W.
Washington, D.C. 20036

Name ______________________________

Mailing Address ______________________________

______________________________ Zip ______________

Employer/School ______________________________

Position (Rank) ______________________________

Employer's City/State ______________________________

Highest Earned Degree ______________ Year Degree Earned ______________

Institution Awarding Degree ______________________________

Month/Year of Birth ______________________________

Have you been a member of the MAA before? ☐ **Yes** ☐ **No**

Payment Enclosed $______________ *U.S. Funds Only* (see below)

Your membership may begin on July 1 or January 1. You may elect to receive any of the listed combinations of the three MAA journals:

THE AMERICAN MATHEMATICAL MONTHLY **(M)**
MATHEMATICS MAGAZINE **(G)**
THE COLLEGE MATHEMATICS JOURNAL **(J)**

Please circle the appropriate box corresponding to your initial membership period and the selection of journals you wish to receive. (Rates guaranteed for the indicated periods only.) Subscription prices are included with dues (see **).

	M	**G**	**J**	**M+G**	**M+J**	**G+J**	**M+G+J**
Student Membership							
1 year (Jan.-Dec. 1986)	$24	$18	$20	$30	$32	$26	$38
1½ year (July 1986-Dec. 1987)	$36	$27	$30	$45	$48	$39	$57
Regular Membership							
1 year (Jan.-Dec. 1986)	$40	$28	$31	$52	$55	$43	$67
1½ year (July 1986-Dec. 1987)	$60	$42	$46.50	$78	$82.50	$64.50	$100.50

*Student Membership available to high school and undergraduate students and to students regularly enrolled in graduate study at least half time. Enclose letter from school or department official confirming status. Student rates apply to unemployed persons under certain circumstances.

**Annual dues include subscription prices as follows: Reg. Memb. $24 (M), $15 (J), $12 (G), $1 (Focus). Std. Memb. $12 (M), $8 (J), $6 (G), $1 (Focus).

METHOD OF PAYMENT ☐ Check enclosed ☐ VISA ☐ MasterCard ☐ Bill me

Card Number ☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐ Expiration Date ☐☐☐☐ *(month/year)*

Interbank Number ☐☐☐☐ *(MasterCard only—located above name on card)*

SIGNATURE ______________________________

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 paperbound.

NUMBERS: RATIONAL AND IRRATIONAL by Ivan Niven. $8.75, $7.00* NML-01

WHAT IS CALCULUS ABOUT? by W. W. Sawyer. $8.75, $7.00* NML-02

AN INTRODUCTION TO INEQUALITIES, by E. F. Beckenbach, and R. Bellman. $8.75, $7.00* NML-03

GEOMETRIC INEQUALITIES, by N. D. Kazarinoff. $8.75, $7.00* NML-04

THE CONTEST PROBLEM BOOK. Problems from the Annual High School Mathematics Examinations 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. $8.75, $7.00* NML-05

THE LORE OF LARGE NUMBERS, by P. J. Davis. $10.00, $8.00* NML-06

USES OF INFINITY, by Leo Zippin. $8.75, $7.00* NML-07

GEOMETRIC TRANSFORMATIONS, by I. M. Yaglom, translated by Allen Shields. $8.75, $7.00* NML-08

CONTINUED FRACTIONS, by C. D. Olds. $10.00, $8.00* NML-09

GRAPHS AND THEIR USES, by Oystein Ore. $8.75, $7.00* NML-10

HUNGARIAN PROBLEM BOOKS I and II, based on the Eotvos Competitions 1894-1905 and 1906-1928. Translated by E. Rapaport. $8.75, $7.00* each NML-11 and NML-12

EPISODES FROM THE EARLY HISTORY OF MATHEMATICS, BY A. Aaboe. $8.75, $7.00* NML-13

GROUPS AND THEIR GRAPHS, by I. Grossman and W. Magnus. $10.00, $8.00* NML-14

THE MATHEMATICS OF CHOICE, by Ivan Niven. $10.00, $8.00* NML-15

FROM PYTHAGORAS TO EINSTEIN, by K. O. Friedrichs. $8.75, $7.00* NML-16

THE CONTEST PROBLEM BOOK II. A continuation of NML-05 containing problems and solutions from the Annual High School Mathematics Examinations for the period 1961-1965. $8.75, $7.00* NML-17

FIRST CONCEPTS OF TOPOLOGY, by W. G. Chinn and N. E. Steenrod. $10.00, $8.00* NML-18

GEOMETRY REVISITED, by H.S.M. Coxeter, and S. L. Greitzer. $10.00, $8.00* NML-19

INVITATION TO NUMBER THEORY, by Oystein Ore. $8.75, $7.00* NML-20

GEOMETRIC TRANSFORMATIONS II, by I. M. Yaglom, translated by Allen Shields. $10.00, $8.00* NML-21

ELEMENTARY CRYPTANALYSIS—A Mathematical Approach, by Abraham Sinkov. $10.00, $8.00* NML-22

INGENUITY IN MATHEMATICS, by Ross Honsberger. $10.00, $8.00 NML-23

GEOMETRIC TRANSFORMATIONS III, by I. M. Yaglom, translated by Abe Shenitzer. $10.00, $8.00* 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 Examinations for the period 1966-1972. $10.00, $8.00* NML-25

MATHEMATICAL METHODS IN SCIENCE, by George Polya. $10,00, $8.00* 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. $10.00, $8.00* NML-27

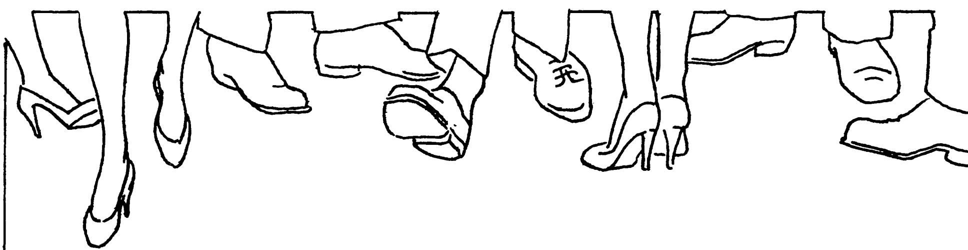
THE MATHEMATICS OF GAMES AND GAMBLING, by Edward W. Packel. $10.00, $8.00* NML-28

THE CONTEST PROGRAM BOOK IV, Annual High School Mathematics Examinations 1973-1982. Compiled and with solutions by R. A. Artino, A. M. Gaglione and Niel Shell. $11.50, $9.20* NML-29

THE ROLE OF MATHEMATICS IN SCIENCE, by M. M. Schiffer and Leon Bowden. $16.00, $12.50* NML-30

*Prices marked with an asterisk are for members of the MAA.

Send Orders to:
The Mathematical Association of America
1529 Eighteenth St., N.W.
Washington, D.C. 20036



MAA STUDIES IN MATHEMATICS

Studies in Numerical Analysis

MAA Studies in Mathematics #24
Gene H. Golub, Editor
415 pp. Hardbound.
List: $42.00   MAA Member: $31.00

This volume is a collection of papers describing the wide range of research activity in numerical analysis. The articles describe solutions to a variety of problems using many different kinds of computational tools. Some of the computations require nothing more than a hand held calculator: others require the most modern computer. While the papers do not cover all of the problems that arise in numerical analysis, they do offer an enticing and informative sample.

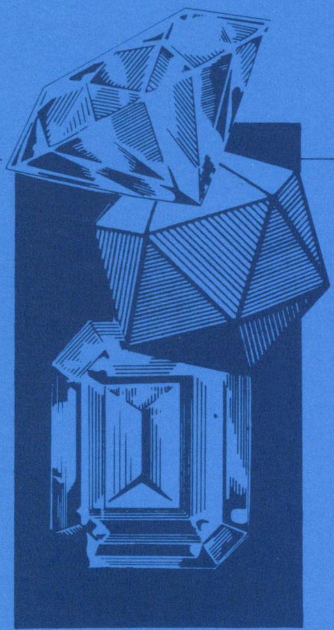
Numerical analysis has a long tradition within mathematics and science, beginning with the work of the early astronomers who needed numerical procedures to help them solve complex problems. The subject has grown and developed many branches, but it has not become compartmentalized. Solving problems using numerical techniques often requires an understanding of several of the branches. This fact is reflected in the papers in this collection.

Computational devices have expanded tremendously over the years, and the papers in this volume present the different techniques needed for and made possible by several of these computational devices.

Table of Contents

THE MATHEMATICAL ASSOCIATION OF AMERICA

Order From:
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036



Mathematical Gems III

by Ross Honsberger

260 pp. Hardbound, 1985.

ISBN-0-88385-313-2

MAA Member: $21.00

List: $27.00

This book is packed with gems from elementary combinatorics, geometry, number theory, and probability. It will bring the enjoyment of good mathematics to the student as well as the teacher.

Glance at the Table of Contents and order now!

Order from:
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, DC 20036

MATHEMATICS MAGAZINE VOL. 59, NO. 3, June 1986