MATHEMATICS MAGAZINE

CONTENTS

VOLUME 36 • SEPTEMBER 1963 • NUMBER 4

MATHEMATICS MAGAZINE

ROBERT E. HORTON, *Editor*

EDITORIAL CORRESPONDENCE should be sent to the Editor, ROBERT E. HORTON, Department of Mathematics, Los Angeles City College, 855 North Vermont Avenue, Los Angeles 29, California. Articles should be typewritten and double-spaced on $8\frac{1}{2}$ by 11 paper. The greatest possible care should be taken in preparing the manuscript, and authors should keep a complete copy. Figures should be drawn on separate sheets in India ink and of a suitable size for photographing.

NOTICE OF CHANGE OF ADDRESS and other subscription correspondence should be sent to the Executive Director, H. M. GEHMAN, Mathematical Association of America, SUNY at Buffalo (University of Buffalo), Buffalo 14, New York.

ADVERTISING CORRESPONDENCE should be addressed to F. R. OLSON, Mathematical Association of America, SUNY at Buffalo (University of Buffalo), Buffalo 14, New York.

The MATHEMATICS MAGAZINE is published by the Mathematical Association of America at Buffalo, New York, bi-monthly except July-August. Ordinary subscriptions are: 1 year \$3.00; 2 years \$5.75; 3 years \$8.50; 4 years \$11.00; 5 years \$13.00. Members of the Mathematical Association of America may subscribe at the special rate of 2 years for \$5.00. Single copies are 65¢.

PUBLISHED WITH THE ASSISTANCE OF THE JACOB HOUCK MEMORIAL FUND OF THE MATHEMATICAL ASSOCIATION OF AMERICA

Second class postage paid at Buffalo, New York and additional mailing offices.

AN INTRODUCTION TO INFORMATION THEORY

MICHAEL MARCUS, The RAND Corporation, Santa Monica, California

I. Introduction. Information theory is concerned with developing a framework for the study of communication systems and their capacity to transmit messages in the presence of natural disturbances. The information function or entropy of a message source gives a quantitative measure of the usefulness of the source for transmitting messages. The capacity of the channel over which the message is being transmitted, gives a measure of the most efficient communication of information possible under given conditions. In the derivation of parameters of communication systems based on the information function and channel capacity it becomes clear that certain systems can be improved. Also, messages can be altered to provide more efficient transmission of information and to reduce the errors in a message which are caused by noise. The alterations are referred to as codes for the message. The proof of the existence of codes with desirable properties forms the basic result of information theory.

This subject can be developed in a purely mathematical fashion or by making reference to descriptive concepts in electrical engineering. We shall attempt to proceed along both paths, first using an axiomatic approach and then providing examples based on models of actual systems. The subject of information theory is rapidly expanding; this paper is a survey of some of its highlights. No attempt is made to prove the theorems which are stated.

Sections II and III of this paper are summarized from lecture notes by H. P. McKean in his course on probability at Massachusetts Institute of Technology in the fall of 1961. Section V is taken from a paper by the author and I. S. Reed [1], and the material in Section VI appeared first in Shannon's book [2].

II. Information Function. The information function is a numerical measure for the amount of information possessed by an ensemble of random events. The function is derived to satisfy four axioms. In understanding these axioms the nature of the information function should become clear.

Consider a probability sample space A which contains n independent events, $a_i, i=1, \cdots, n$. A probability distribution is defined on the sample space which represents the *a priori* probability of the occurrence of each event during a single "selection" from the sample space. The probability of the a_i-th event will be denoted by either $p(a_i)$ or p_i. By the above definitions

$$p(a_i \cup a_j) = p(a_i) + p(a_j), \qquad \sum_{i=1}^{n} p(a_i) = 1.$$

Suppose one event is selected from the sample space. We shall ask the question, how great is our uncertainty of what the event will be? Another way to phrase this is, how much information will knowledge of the event convey to us? Various degrees of uncertainty can be envisioned. If the *a priori* probabilities of the events are all zero except for one which is 1 (i.e., $p_i=0$, $i\neq j$, $i\leqq n$, $p_j=1$), then the event to be selected is certain. It would be reasonable to require that the information function be zero in this case. If two events have probability $\frac{1}{2}$

and the rest probability zero, then knowing which of these two events were selected constitutes information, but not as much, roughly speaking, as a case in which three events have probability $\frac{1}{3}$ and the rest probability zero. In the former case the event selected could be guessed more frequently than in the latter. If all the events were equally certain, each with probability $1/n$, then guessing which event was selected would be least likely; thus knowing of the selection, this information is greatest. (It should be noted that the examples of this paragraph apply to a sample space with n events.)

Formally, the information function for the probability sample space A will be denoted $H(A)$ or $H(p_1, \cdots, p_n)$, where $p_1, \cdots, p_n$ are the *a priori* probabilities of the events $a_1, \cdots, a_n$ contained in A. $H(A)$ is required to satisfy the following four axioms:

Axiom 1. *$H(p_1, \cdots, p_n)$ is a symmetric function of $p_1, \cdots, p_n$.* A symmetric function is a function with the same value regardless of the order in which the probabilities are taken. That $H(A)$ should be symmetric is clear since the information function is dependent only upon the relative probabilities of the events and not on the order in which the events are numbered.

Axiom 2. *$H(p_1, \cdots, p_n)=0$ if all the p_i equal zero except one which is equal to* 1. The justification of this axiom has already been given.

Axiom 3. $H(A\vee B)=H(A)+H(B|A)$. In this equation let B be a probability sample space of m events $b_1, \cdots, b_m$. By $A\vee B$ we indicate the probability sample space of nm events consisting of the pairs (a_i, b_j) with probabilities $p(a_i, b_j)=p_{ij}$. In general $p(a_i, b_j)=p(a_i)p(b_j|a_i)$, where $p(b_j|a_i)$ is the conditional probability of the event b_j given that the event a_i has occurred.

The function $H(B|A)$ is the conditional information function for the probability space B given knowledge of the probability space A. It is defined as:

$$H(B\mid A) = \sum_{i=1}^{n} p(a_i)H[p(b_1\mid a_i) \cdots p(b_m\mid a_i)]$$

that is, for each $a_i\in A$ we calculate the information function for the m probabilities $p(b_1|a_i), \cdots, p(b_m|a_i)$, and then take its expected value over A. Axiom 3 is often stated as: The information of the joint event $A\vee B$ is equal to the information of A plus the information of B when A is known. Intuitively one would want $H(B|A)$ to be equal to zero when B is completely determined by A and $H(B|A)$ to be equal to $H(B)$ when B is independent of A. We shall see later that this is the case.

Axiom 4. *If the number of events in A is two, with probabilities p_1 and p_2, then $H(p_1, p_2)$ is maximum when $p_1=p_2=\frac{1}{2}$.* We have discussed above how one would expect the information function defined on a space of n events to be greatest when the events are all equally probable. *Axiom* 4 is a special case of this general premise. It turns out that $H(1/n, \cdots, 1/n)$ is the maximum value for the information function on a space of n events. However, given Axioms 1 through 3, it is only necessary to require this condition for $n=2$.

Given these four axioms we obtain the following unusual and important result:

THEOREM 1. *The only function which satisfies Axioms* 1 *through* 4 *is*

$$H(p_1, \cdots, p_n) = -\text{ const.} \sum_{i=1}^{n} p_i \lg_2 p_i. \tag{1}$$

The proof of this theorem is far from trivial; it can be found in Refs. 2–5. The constant is arbitrary and is set equal to 1. The function $H(A)$ is often called the entropy of A because it resembles the entropy function of thermodynamics.

The remainder of this section will be devoted to stating some of the properties of the function H. We see that the conditional information function $H(B|A)$ is

$$H(B \mid A) = \sum_{i=1}^{n} p(a_i)\left(-\sum_{j=1}^{m} p(b_j \mid a_i) \lg_2 p(b_j \mid a_i)\right). \tag{2}$$

If A and B are independent, $p(b_j|a_i) = p(b_j)$ and

$$H(B \mid A) = \sum_{i=1}^{n} p(a_i) H(B) = H(B).$$

If B is completely determined by A then for each a_i, $p(b_j|a_i)$ is zero for all b_j except one and $H(B|A) = 0$.

An application of Jensen's inequality shows that $H(B|A) \leqq H(B)$. This yields the important inequality

$$H(A \vee B) \leqq H(A) + H(B), \tag{3}$$

with equality only when A and B are independent. Also, since

$$H = -\sum_{i=1}^{n} p_i \lg_2 p_i = \lg_2 \prod_{i=1}^{n} \left(\frac{1}{p_i}\right)^{p_i},$$

it can be shown using the relationship between the geometric and arithmetic mean that

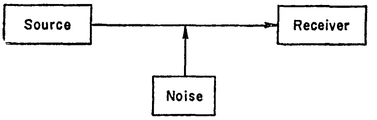
$$H(p_i, \cdots, p_n) \leqq \lg_2 n \tag{4}$$

with equality only when $p_1 = \cdots = p_n$.

Lastly we observe that H is a continuous function of $p_1, \cdots, p_n$. The function $\delta = -x \lg_2 x$ is found to be zero when $x = 0$. Thus if all $p_i = 0$ except one, $H(p_1, \cdots, p_n) = 0$.

In summation, we have postulated useful and sensible properties for a function which would evaluate the degree of "randomness" of an ensemble of random events and found that the function is unique. This function is used to analyze communication systems which by their nature are "random systems." The recognition of the random or probabilistic nature of communications systems is the basic idea of information theory. Norbert Wiener was one of the leaders in this area [6].

III. Communication System. A communication system consists of a source of a message and a channel over which the message is transmitted. The channel is composed of noise and a receiver, where the noise represents the possibility that the transmitted message is incorrectly received. The system is drawn schematically as follows:



Communication system

For our purposes the source will be taken to be a collection of messages of n symbols, each symbol being a letter from an alphabet of d letters. Let A represent the probability space of the source where $a \in A$ represents the individual messages; the source can produce d^n distinct messages, and $p(a)$ will denote the *a priori* probability that the source will produce the message a. Thus the source is a probability sample space of the type given in Section II and it is perfectly appropriate to compute $H(A)$. We note further that $H(A) \leqq n \lg_2 d$.

The received message is also taken to be of length n with each symbol being one of the d letters; however, due to the noise, the received message may differ from the one which was sent. The effect of the noise is characterized by the probabilities $p_a(b)$; this is the probability that if the message a was transmitted, the message b would be received. If we let B represent the probability space of the received messages, the probability of receiving the message b is

$$p(b) = \sum_{a \in A} p(a) p_a(b).$$

Thus the information function $H(B)$ is also well defined.

According to the development of this paper, the communication channel is completely determined by A, the probabilities $p(a)$, the noise characterization $p_a(b)$, and the rule that messages of length n from an alphabet of d letters go into messages of length n with an alphabet of d letters. Also, in the development thus far an application of the information function is seen; $H(A)$ is a measure of the information contained in a message from the source A.

We proceed to discuss the rate and capacity of a communication system and state a fundamental theorem on the capacity of a communication system. The theorem applies to systems which are more general than the one described here; however, by considering the implications of the theorem to our system considerable understanding of the problems of information theory can be gained.

The subscript n will be added to the information function to indicate that it is being applied to a communication system with a source that produces messages of length n. The information function of a communication system is defined as:

$$H_n \text{ (system)} = H_n \text{ (input)} - H_n \text{ (output} \mid \text{input)}. \tag{5}$$

H_n(input) is the information function of the source and is identical to $H_n(A)$. H_n(output | input) is the information function of the output when the input is known. It is equal to the expected value of the information function calculated for the probabilities $p_a(b)$ on the space A.

Equation (5) can be given a verbal interpretation: It states that the information content of a communication system is equal to the information content of the input minus the information content of the output once the input is known. It has been shown that if the input completely determines the output $H(\text{output}|\text{input})$ is zero, thus $H(\text{output}|\text{input})$ is a measure of the uncertainty in the output due to noise. Another way of saying this is that $H(\text{output}|\text{input})$ is a measure of the ambiguity in the output which cannot be resolved by the input. This is what Shannon [2] calls the equivocation.

In general, noise decreases the information of the system. For a noise-free system $H_n(\text{system}) = H_n(\text{input})$, as we would expect.

The n symbol mean rate of communication of the system is defined as:

$$R_n = \frac{1}{n} H_n\,(\text{system}) = \frac{1}{n}\left[H_n\,(\text{input}) - H_n\,(\text{output}\,|\,\text{input})\right]. \tag{6}$$

If the system is such that the following limit makes sense, we define the mean rate of communication to be

$$R = \lim_{n\to\infty} R_n = \lim_{n\to\infty} \frac{1}{n}\left[H_n\,(\text{input}) - H_n\,(\text{output}\,|\,\text{input})\right]. \tag{7}$$

So far we have not discussed the relationship between a source of messages of length n and one of length $n+1$ and, indeed, it has not been necessary. Now, however, we are forced to consider this relationship in order to make sense of the limit in (7). Unfortunately this problem lies beyond the scope of this paper. When the binary symmetric channel is discussed, the meaning of the limit in (7) will be clear. In general the limit will hold for stationary sources. It is also necessary to define the probability space for messages of infinite length; this can be done by standard procedures which will not be discussed in this paper.

The capacity C of a communication channel is defined as the supremum of the rate for all possible sources. We are assured that this exists if we restrict ourselves to stationary sources. An important distinction between the rate of a communication system and the capacity of a channel is that in considering the rate of the system the source is given, whereas the capacity of the channel is evaluated under the assumption that a source is matched to the channel which maximizes the rate of the communication system thus formed. The importance of the capacity of a channel is expressed by the following theorem:

THEOREM 2. *Given a communication channel with capacity C and an ϵ, $0<\epsilon<1$, and $C'<C$, the following will hold for sufficiently large n (dependent on ϵ and C'). A source A of messages of length n can be found from which N distinct messages $A_1, \cdots, A_N$ can be chosen. From the associated set, B, of received messages, N disjoint subsets $B_1, \cdots, B_N$ can be found such that $P_{A_i}(B_i)>1-\epsilon$ and $\lg_2 N/n > C'$.*

The probability $P_{A_i}(B_i)$, for fixed A_i is a probabilistic set function on B. $P_{A_i}(B_i)>1-\epsilon$ means that if the message A_i is transmitted, then with probability exceeding $1-\epsilon$ the received message will be a member of the set B_i. Since the B_i are disjoint subsets of B, this means that with probability greater than

$1-\epsilon$ the communication system will transmit a message correctly. (If a message in the set B_i is received, it is assumed that A_i was sent.) Furthermore, the theorem states that the system is capable of sending N different messages with this accuracy where $N > 2^{nC'}$ for all $C' < C$.

The function $\lg_2 N/n$ is called the empirical rate of the communication channel. Theorem 2 states that for a given ϵ and C', an n can be found such that $\lg_2 N/n > C'$. It is also true that $\lim_{n\to\infty} \lim_{\epsilon\to 0} \lg_2 N/n = C$. We shall use this result although we shall not discuss its derivation.

We shall now develop some inequalities and use these to show how Theorem 2 is used to demonstrate the existence of a "code" that will enable messages to be transmitted error-free over a noisy channel. Recall that $H_n(A) \leqq n \lg_2 d$ and that in general $H \geqq 0$. Applying these remarks to (6), we see that $R_n \leqq \lg d$ and, since this is true for all n, then $R \leqq \lg_2 d$. It follows directly that $C \leqq \lg_2 d$. If $\lim_{n\to\infty} 1/n\, H_n(\text{output}|\text{input}) \geq 0$, then C is strictly less than $\lg_2 d$.

Let us go back to sources of length n; in doing so we are on shaky ground since we shall try to relate results which hold for these sources with quantities which are valid when n goes to infinity. However, we know that the relations will be valid asymptotically. By Theorem 2 and the remarks of the preceding paragraph, $\lg_2 N/n < \lg_2 d$, thus $N < d^n$. Let m be the smallest integer such that $N \leqq d^m$; m is clearly less than or equal to n. The interesting cases occur when m is strictly less than n.

We emphasize that Theorem 2 states that asymptotically N of the possible d^n messages can be transmitted with error less than ϵ. There are two major interpretations concerning the existence of codes. The first is that since $N < d^n$ there should exist a subset of the messages of the source such that, with probability $1-\epsilon$, the disturbance of the message by noise still leaves it distinguishable from the other messages. A simple example is, if the source only transmitted the messages 111 and 000, and the noise caused a 1 to be received as a zero with probability 0.2, then with probability 0.96 these messages would be received correctly. Roughly speaking the N messages are spaced amongst the d^n possible messages so that errors do not cause a message to depart from its original position so far that it can be confused with a different message. This idea is discussed in a paper by Hamming [7].

The second interpretation of the coding problem is that, since $N \leqq d^m$, the first m symbols of the transmitted message can be used to carry the "information" while the remaining n-m symbols can be used for correction. These correction symbols are chosen after the m information symbols are given. They are chosen to work with the information symbols so that the receiver can figure out (decode) where the errors in the message have occurred. An example of this type of code will be given in Section VI.

It should be pointed out that the sender and receiver are working together to overcome the noise and that they both know the rules for coding and decoding.

In the following section the most common type of communication channel will be discussed and will be used to give examples of the meaning of the results of this section.

IV. Binary Symmetric Channel. The rate and capacity of a binary symmetric channel will be calculated and used to illustrate some of the concepts of information theory. The reader is probably familiar with the idea that complex messages can be transmitted in binary form so that this channel which is being described is not as restricted as it might seem.

We consider a binary source, that is, a source which at each stage transmits a 1 or a 0 with probabilities p_1 and $1-p_1$, the transmission being independent from stage to stage. (In reality the source would probably be constructed either to send or not to send an electrical impulse.) The channel is symmetric because the probability of receiving a 1 or a zero in error is the same. Thus

$$\begin{aligned} p_0(0) &= p_1(1) = 1 - p_2 \\ p_1(0) &= p_0(1) = p_2 \end{aligned} \tag{8}$$

where we continue the convention that $p_a(b)$ is the probability that the symbol b is received when the symbol a is transmitted.

The source is capable of transmitting 2^n different messages of length n. The probability of any message is simply $p_1^j(1-p_1)^{n-j}$ where j is the number of ones in the message. Furthermore all messages with j ones are equally likely. The information function of the source is

$$H_n\ (\text{source}) = \sum_{j=0}^{n} \frac{n!}{j!(n-j)!} p_1^j(1-p_1)^{n-j} \lg_2 p_1^j(1-p_1)^{n-j}$$

and by a simple manipulation common in dealing with binomial distributions we obtain

$$H_n\ (\text{source}) = n[-p_1 \lg_2 p_1 - (1-p_1) \lg_2 (1-p_1)]. \tag{9}$$

It is immediately evident that $H_n(\text{output}|\text{input})$ is the same as in (2) with p_1 replaced by p_2. From this we get the rate of the channel as

$$R = -p_1 \lg_2 p_1 - (1-p_1) \lg_2 (1-p_1) + p_2 \lg_2 p_2 + (1-p_2) \lg_2 p_2.$$

If $H_n(\text{output}|\text{input})$ is greater than $H_n(\text{input})$, the rate of the system is negative. In keeping with an axiomatic approach to information theory, we simply state that the rate of a system is only defined when it is equal to or greater than zero. However, to digress for a moment, it is easy to see that if $-p_2 \lg_2 p_2 - (1-p_2) \lg_2 p_2$ is greater than $-p_1 \lg_2 p_1 - (1-p_1) \lg_2 (1-p_1)$, the most accurate reception of a message is obtained when the receiver assumes either that all letters are 0 if $p_2 \geqq \frac{1}{2}$ or are 1 if $p_2 < \frac{1}{2}$. In other words the best method for receiving the message is independent of the source. We might say that the channel is too noisy for the source.

To calculate the capacity of the channel we need only note that the supremum of $(1/n)H_n$ (source) is 1, so that

$$C = 1 - p_2 \lg_2 p_2 - (1-p_2) \lg_2 (1-p_2).$$

To see how Theorem 2 relates to the binary symmetric channel, assume that the probability p_2 of the effects of noise on the message has been measured. (This can usually be accomplished in practice.) Assume also that we are required to transmit messages correctly with probability $1-\epsilon$, $\epsilon \ll p_2$, $p_2 < \frac{1}{2}$. It should be clear that, if a message is not "coded" in some way to compensate for errors due to the noise of the channel, the probability that a message of length k will be received correctly is $(1-p_2)^k$, a number which goes to zero very quickly as k increases. However, from Theorem 2 we see that for messages of some sufficiently long length n, the channel is capable of transmitting approximately 2^{nC} different messages with a probability of error of only ϵ. (We say approximately 2^{nC} to absorb the distinction between C' and C; C' approaches C from below for large n; or, to repeat, as $n \to \infty$, $C' \to C$ and $\epsilon \to 0$.) In other words, we can transmit messages with arbitrarily small error. The price that is paid is that only 2^{nC} messages of length n can be sent, $0 \leqq C \leqq 1$, whereas if the channel were noise-free 2^n different messages could be sent. As an example, suppose that $p_2 = 0.1$; then $C = 0.532$. Thus, for a probability of error per letter of only 0.1, almost half of the potential information-carrying letters of the source must be devoted to making the message error-free. The manner in which this is done has been discussed abstractly in Section III; an example will be given in Section VI.

One final remark; although it is useful to use "messages of length n" both to derive results and to aid in an intuitive understanding of the results of information theory, it is also important to be able to provide a qualitative measure of the usefulness of a system and channel without having to consider message length. The values for rate and channel capacity provide such a measure. Indeed this was the major initial contribution of information theory, whereby the means are provided for comparing communication systems and channels.

V. Compression of Finite Discrete Messages. So far we have spoken of coding messages to overcome noise; however, there is another kind of coding which we can do and which is simpler to understand. Consider the binary source described in Section IV which produces a 1 with probability p_1 and a zero with probability $(1-p_1)$. If we suppose that the source is being used to transmit messages of length n then H_n(source), the information function of the source, is $n[-p_1 \lg_2 p_1 - (1-p_1) \lg_2 (1-p_1)]$. Let m be the smallest integer greater than or equal to this function. A source which produces a 1 or a 0 with probability 1/2 would have $H_m(\text{source}) \geqq H_n(\text{source})$, where m could be considerably smaller than n. In other words, at least as much information could be obtained from the second source as from the first even though it is transmitting shorter messages. If we suppose a fixed cost of transmission per letter, the second source is preferable to the first and it is desirable to code the messages of the first source into messages which could be transmitted by the more efficient source. The process will be called compressing the message.

We proceed to state these observations more rigorously. The quantity $(1/n)H_n(\text{source}) = -p_1 \lg_2 p_1 - (1-p_1) \lg_2 (1-p_1)$; following Khinchin [3] we define $\lim_{n\to\infty} (1/n)H_n(\text{source}) = k$, which we shall call the compression factor of the source. This quantity k is what Shannon denotes as H, the entropy per symbol of the source. Also notice that for a noiseless channel this is the same as R, the rate of the system. Theorem 3 paraphrases theorems of Shannon and

Khinchin, restricted to binary sources:

THEOREM 3. *Let the compression factor of a binary source be k; then given a δ and an ϵ for a sufficiently large n, messages of length n from the source can be divided into two classes. Class* 1 *consists of 2^m messages such that $m/n=k+\delta$ and such that, with probability exceeding* $1-\epsilon$, *a message produced by the source falls in this class. Class* 2 *contains the remaining messages, and the probability that the source produces a message in this class is less than ϵ. Furthermore as $n\to\infty$, $m/n\to k$ and* $\epsilon\to 0$. Thus asymptotically, for large n, a message of length n can be compressed by the factor k where $k=H$, the entropy per symbol of the source.

If this paper has failed to suggest to the reader the beauty and simplicity of the concepts of information theory, then it is only the author's failure. Still, there is something unsatisfying about theorems that continually require arbitrarily long messages for their realization, since in reality all messages have finite length. In 1960 I. S. Reed and the author [1] tried to approach the problem of coding sources for transmission over a noiseless channel as an engineer might who relied upon common sense rather than knowledge of information theory. We were fortunate since for this problem a best method readily presents itself. In the case of binary sources, besides obtaining the classical results by direct calculation, our approach shows the relation between the compressibility of a message and its length. The remainder of this Section is a summary of parts of our paper.

For a binary source which produces messages of some given length N (for all but very small N) we will show that the messages can be compressed from length N to length $N'<N$ with probability ϕ. For messages of length N we define the N-stage compression factor $k(N,\phi)=N'/N$. This is a function of both N and ϕ. As $N\to\infty$ and $\phi\to 1$, the compression factor will approach k, as before. However, before obtaining this result for binary sources a more general source will be studied.

Consider an alphabet of m letters A_m and a random source $\Omega_{N,m}$ that produces a sequence of N symbols called a message; each symbol of the message being one of the letters in A_m. An element $X\in\Omega_{N,m}$, $X=(x_1, x_2, \cdots, x_N)$, $x_i\in A_m$ represents a particular message from $\Omega_{N,m}$.

Let a probability density function $P(X)$ be defined on $\Omega_{N,m}$. Let S be a subset of $\Omega_{N,m}$. For each S in $\Omega_{N,m}$, $P(S)=\sum_{X\in S}P(X)$, $P(\Omega_{N,m})=1$. Let $v(S)$ be the cardinality of S, i.e., the number of messages in S. Let C_ϕ be the collection of sets S such that $P(S)\geq\phi$. The following lemma is obvious:

LEMMA. *Given ϕ, $0\leqq\phi\leqq 1$, there exists a set $S_\phi\in C_\phi$ such that $v(S_\phi)\leqq v(S)$ for all $S\in C_\phi$.*

Proof. Choose a set S in C_ϕ with minimum cardinality, call this set S_ϕ. Since C_ϕ is finite, this choice obtains without difficulty.

Thus, given a source which produces a random message of length N with a well-defined probability density function for the individual messages of the source, and given a probability ϕ, a set S_ϕ with the smallest number $v(S_\phi)$ of possible messages can be found such that with probability ϕ a message produced by the source will be contained in S_ϕ.

The messages in S_ϕ can be coded in the following simple way. Each message —there are $v(S_\phi)$ of them—is listed and assigned an m-nary number (a number in the base m) and this list is supplied in advance to the receiver of the message. If the source produces a message in this list, its number is sent to the receiver with the letters of the alphabet A_m used as terms of the number. The length of the transmitted message is clearly $\log_m v(S_\phi)$ and the N-stage compression factor of the message is $\log_m v(S_\phi)/N$. Evidently, for ϕ and N fixed, this method of encoding the source obtains optimal compression. In general, for a given source and probability ϕ the N-stage compression factor $k(N, \phi)$ is defined as:

$$k(N, \phi) = \frac{1}{N} \log_m v(S_\phi).$$

The N-stage compression factor will be obtained for the binary source described in Section IV, where the probability that a 1 is produced is p, and with the probability $(1-p)=q$ that a zero is produced. For the sake of the example we assume $p<q$. If $p=q=\frac{1}{2}$, no meaningful compression can take place.

For a given value of ϕ we want to find the subset S_ϕ of the 2^N possible messages that can be produced by the source and the number $v(S_\phi)$ of messages contained in S_ϕ. Clearly S_ϕ contains first the message consisting of all zeros, and then, all the messages containing a single 1, then all messages containing two ones, and so on up to the messages containing M ones. Here M is the smallest integer such that $\sum_{j=1}^{M} \binom{N}{j} q^{N-j}p^j \geqq \phi$. The number of messages in S_ϕ is $\sum_{j=1}^{M} \binom{N}{j}$. Note that the value of M is chosen so that $P(S_\phi)$ can exceed ϕ; however, for large N and for values of ϕ that occur in practice, the corresponding difference in $v(S_\phi)$ will not be significant.

For binary messages of length N, $v(S_\phi)=\sum_{j=1}^{M} \binom{N}{j}$ where M is chosen as before and $k(N, \phi)=1/N \log_2 v(S_\phi)$. This completes the problem. However, it is difficult to obtain $v(S_\phi)$ directly, just as it is difficult to find M as a function of ϕ. These calculations can be found in [1], so we shall simply state the result here.

$$k(N, \phi) = -p \lg_2 p - q \lg_2 q + \frac{(1.4)b\sqrt{pq} \lg \dfrac{q}{p}}{\sqrt{N}} - \frac{\lg_2 \left[\sqrt{2\pi} \left(\Delta \lg \dfrac{q}{p} - b \right) \right] + (1.4) \dfrac{b^2}{2}}{N} \tag{10}$$

where b is chosen so that $1/\sqrt{2\pi}\int_{\infty}^{b} \exp(-x^2/2)dx=\phi$, $\sigma=\sqrt{Npq}$, and we write lg for $\log_e$.

From (10) it is seen that as $N\to\infty$ we can let $\phi\to 1$ and $k(N, \phi)\to -p \lg_2 p -q \lg_2 q$, the latter being the compression factor k of the source. Notice though, that for ϕ to approach 1, b must go to infinity. This is possible but we must adjust the rate at which b and N go to infinity so that the second and third terms of (10) go to zero. Using (10), a sharper version of Theorem 3 can be given; one in which we obtain more information about the asymptotic behavior of the message compression.

THEOREM 3′. *Let the compression factor of a binary source be* k; *then given an* ϵ *for a sufficiently large* n, *messages of length* n *from the source can be divided into two classes. Class* 1 *consists of* 2^m *messages such that* $m = k \cdot n + O(\sqrt{n})$ [$O(\sqrt{n})$ *means that the term is bounded by a constant times* $\sqrt{n}$ *for sufficiently large* n] *and is such that, with probability exceeding* $1-\epsilon$, *a message produced by the source falls in this class. Class* 2 *contains the remaining messages, and the probability that the source produces a message in this class is less than* ϵ. *Furthermore as* $n \to \infty$, $m/n \to k$ *and* $\epsilon \to 0$.

Thus not only are the results of the axiomatic theory obtained, but also a method for the coding is given.

VI. Remark of Coding for Noisy Channels. Shannon's paper [2] which appeared in 1949 suggested the existence of codes for transmission over noisy channels; however, for many years only simple examples of such codes could be found. Recently, using concepts of algebra, in particular Galois field theory, many useful and deep codes have been developed. To describe them is beyond the scope of this paper; for bibliography see Ref. 8. Instead we will give the example of an efficient code which was given by Shannon. The example is somewhat artificial but is nevertheless interesting and easy to understand.

Suppose we have a binary source and a noisy channel which acts independently on each block of seven symbols from the source. The effect of the noise on the block of symbols is such that either the whole block is transmitted correctly or else one and only one error occurs in the block. All eight of these possibilities are equally likely. The channel capacity of this communication system is well defined only if we consider messages of lengths which are multiples of seven. In this case the maximum information function of the source is seven, whereas $H(\text{output}\,|\,\text{input})$ is $8 \cdot \frac{1}{8} \lg_2 \frac{1}{8}$. Since the system is independent in blocks of seven symbols, we have

$$C = \frac{1}{7}\left(7 + \frac{8}{8}\lg_2\frac{1}{8}\right) = \frac{4}{7}\cdot$$

A code which operates on each block of seven symbols permitting error-free transmission of four symbols, while using three symbols for purposes of coding, is an efficient code for this channel. By "efficient code" we mean a code for which the channel operates at its capacity. For the channel which has been described such a code can be constructed.

Let the seven symbols transmitted by the source be denoted by x_i, $i = 1, \cdots, 7$. The symbols x_1, x_2, x_3, x_4, comprise the message. The symbols x_5, x_6, x_7 are chosen as follows:

Choose x_5 so that $x_1 + x_3 + x_5 + x_7$ is even

Choose x_6 so that $x_2 + x_3 + x_6 + x_7$ is even

Choose x_7 so that $x_4 + x_5 + x_6 + x_7$ is even.

The message $x_1, x_2, x_3, x_4, x_5, x_6, x_7$ is transmitted and the receiver calculates

$$\alpha = x_1 + x_3 + x_5 + x_7$$
$$\beta = x_2 + x_3 + x_6 + x_7$$
$$\gamma = x_4 + x_5 + x_6 + x_7$$

in arithmetic module 2 (i.e., α is 0 or 1 according to whether $x_1+x_3+x_5+x_7$ is even or odd). Observe that the three-digit binary number $\alpha\beta\gamma$ gives the position of the error in the block of seven symbols; it is 000 if no error occurs.

References

1. Marcus, M. B. and I. S. Reed, *The Compression of Finite Discrete Messages*, The RAND Corporation, Paper P-2186, December 1960.
2. Shannon, C. E. and E. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, September 1949.
3. Khinchin, A., *Mathematical Foundations of Information Theory*, Dover Publications, New York, 1957.
4. Feinstein, A., *Foundations of Information Theory*, McGraw-Hill Book Co., Inc. New York, 1958.
5. Fano, R. M., *Transmission of Information; A Statistical Theory of Communication*, John Wiley and Son, New York, 1961.
6. Wiener, N., "What is Information Theory," *I.R.E. Transactions on Information Theory*, Vol. IT-2, June 1956, p. 48.
7. Hamming, R., "Error Detecting and Error Correcting Codes," *The Bell System Technical Journal*, April 1950, p. 147.
8. Peterson, W. W., *Error Correcting Codes*, M.I.T. Press, Cambridge, 1961.

A BASE SUGGESTION?

RODNEY T. HOOD, Franklin College of Indiana

Professor Dowling's "radical suggestion"[1] suggests the following comment. Since the square of any non-zero integer ends in an even number of zeros, it follows that if p and q are non-zero integers, $10p^2=q^2$ is an impossible equation, indicating an odd number of terminal zeros on the left and an even number on the right. This is true not only in the decimal system but also in any system in which the base (written β) is not a perfect square. This provides a simple proof that $\sqrt{n}$ is irrational so long as n is not a perfect square, for the assumption that $\sqrt{n}$ is rational leads to an equation of the form $\beta p^2=q^2$ (base n).

A similar argument shows that $\sqrt[m]{n}$ is irrational unless n is the mth power of an integer. For otherwise we should have the equation $\beta p^m=q^m$ (base n). This is impossible, since the number of terminal zeros on the right hand side of the equation is an exact multiple of m, while on the left hand side this is not the case.

---

[1] Dowling, Roy J., A Radical Suggestion, MATHEMATICS MAGAZINE, Volume 36, Number 1, Jan.–Feb. 1963, p. 59.

ON KHAZANOV'S FORMULAE

A. F. HORADAM, The University of New England, Australia

Introduction. As is well-known, the formulae

$$(1) \qquad x = p^2 - q^2, \qquad y = 2pq, \qquad z = p^2 + q^2$$

($p > q$; p, q mutually prime but not simultaneously odd in order to avoid repetitions) express all the integral solutions of the Pythagorean equation

$$(2) \qquad x^2 + y^2 = z^2.$$

Khazanov [1] has recently pointed out one more method for the derivation of these formulae. Using a generalised Fibonacci sequence, I have obtained ([2] and [3]) a formula by which all the Pythagorean number triples (i.e., Fibonacci number triples) may be found. The purpose of this note is to connect these two approaches.

Khazanov's Method. Briefly, Khazanov's technique is as follows. Let

$$(3) \qquad z = x + x_1, \qquad (x, x_1) = 1.$$

Put

$$(4) \qquad y = \frac{p}{q} x_1, \qquad (p, q) = 1, \quad p > q.$$

Then, from (2) and (3), we have the formulae

$$(5) \qquad x = \frac{p^2 - q^2}{2q^2} \cdot x_1, \qquad z = \frac{p^2 + q^2}{2q^2} \cdot x_1.$$

If p, q are such that one of them is even and the other odd, then $x_1 = 2q^2$ and (1) results. If p, q are both odd, then $x_1 = q^2$ and

$$(6) \qquad x = \frac{p^2 - q^2}{2}, \qquad y = pq, \qquad z = \frac{p^2 + q^2}{2}$$

which satisfy (2).

Use of Fibonacci Numbers. For our purposes, the only results required from [2] and [3] are, for any Fibonacci sequence H_{rs}:

(7) $H_{n+2} = H_{n+1} + H_n$ (recurrence law);

(8) H_n, H_{n+1} are always mutually prime provided r, s are mutually prime (otherwise repetitions of other sequences occur);

(9) $(H_n H_{n+3})^2 + (2H_{n+1}H_{n+2})^2 = (2H_{n+1}H_{n+2} + H_n)^2$ (Pythagorean theorem).

Now, from (2) and (9), we can associate x with either of the terms $H_n H_{n+3}$, $2H_{n+1}H_{n+2}$. Suppose $x = H_n H_{n+3}$. Then $z - x = 2H_{n+1}H_{n+2} + H_n^2 - H_n H_{n+3}$ $= 2H_{n+1}(H_{n+1} + H_n) + H_n^2 - H_n(2H_{n+1} + H_n) = 2H_{n+1}^2$ on using (7). That is,

$$x_1 = 2H_{n+1}^2. \tag{10}$$

Also,

$$\frac{H_nH_{n+3}}{2H_{n+1}^2} = \frac{H_n}{H_{n+1}}\left(1 + \frac{H_n}{2H_{n+1}}\right)$$

which, by (8), is in its lowest form.
Therefore

$$(x, x_1) = 1.$$

Next,

$$y = 2H_{n+1}H_{n+2} = \frac{p}{q}\cdot 2H_{n+1}^2$$

by (4) and (10). That is,

$$\frac{p}{q} = \frac{H_{n+2}}{H_{n+1}}$$

which is in its lowest form. Thus,

$$p = H_{n+2}, \qquad q = H_{n+1} \quad \text{with} \quad (p, q) = 1 \text{ by } (8) \tag{11}$$

and

$$x_1 = 2q^2 \text{ from (10), (11).}$$

Secondly, given $x=2H_{n+1}H_{n+2}$, we may likewise obtain $x_1=H_n^2$, $(x, x_1)=1$, $p=H_{n+3}$, $q=H_n$, $(p, q)=1$, $x_1=q^2$ (Notice that H_n, H_{n+3} are always both even, or both odd.) Hence, Fibonacci number triples satisfy all the conditions of Khazanov's method. Should H_n be even, it would of course be necessary to simplify in (9) by dividing by 2 in order to obtain the Pythagorean triples (for otherwise $(x, x_1)\neq 1$).

Example. Consider the Fibonacci sequence H_{31}:

H_1	H_2	H_3	H_4	H_5	H_6	$H_7 \cdots$
3	4	7	11	18	29	47 $\cdots$.

Putting $n=1$ in (9) we have $33^2+56^2=65^2$. If $x=33$, then $x_1=32$, $p=7$, $q=4$ while if $x=56$, then $x_1=9$, $p=11$, $q=3$. When $n=2$, we have, after simplification, $36^2+77^2=85^2$. If $x=77$, then $x_1=8$, $p=9$ $q=2$ while if $x=36$, then $x_1=49$, $p=11$, $q=7$.

References

1. M. B. Khazanov, "Formulas for Pythagorean numbers." Kabardin. Gos. Ped. Inst. Uč. Zap. **12** (1957), 14 (in Russian)

2. A. F. Horadam, "A Generalised Fibonacci Sequence." Amer. Math. Monthly **68** (5) (1961) 455–459.

3. ———, "Fibonacci Number Triples." Amer. Math. Monthly **68** (8) (1961) 751–753.

A NOTE ON THE SUM OF SQUARES

ALAN SUTCLIFFE, Knottingley, Yorkshire, England

Two theorems are proved which will be of interest to those introducing number theory at a fairly elementary level. They provide a convenient link between the simple theorem giving all solutions of $a^2+b^2=c^2$ and the more advanced theory of the solution of $a^2+b^2=M$. The significance of both theorems can conveniently be displayed in tabular form.

THEOREM I. *$A^2+B^2+C^2=D^2$ in positive integers, if and only if $A=2a$, $B=2b$, $C=x-y$ and $D=x+y$, where a and b are integers, and x and y are integers such that $xy=a^2+b^2$, and $x>y$.*

Proof. The remainder, on dividing a square by 4, is 0 or 1 according to whether the square is even or odd. Hence, considering $A^2+B^2+C^2=D^2$, if D is even, then A, B and C must be even, otherwise the two sides of the equation would not have the same remainder, while, if D is odd, two of the numbers A, B and C must be even, and the other odd.

Thus the equation may be re-written as $A^2+B^2=D^2-C^2$, where A and B are both even, and C and D are either both even or both odd. Let $A=2a$, $B=2b$, then

$$a^2+b^2=\left(\frac{D+C}{2}\right)\left(\frac{D-C}{2}\right).$$

Let

$$x=\frac{D+C}{2}, \qquad y=\frac{D-C}{2}\cdot$$

x and y are thus integers which satisfy $xy=a^2+b^2$, $x-y=C$, $x+y=D$, and $x>y$. So that all solutions are of the form stated.

If a and b are any integers, and x and y are integers such that $xy=a^2+b^2$, $x>y$, then

$$(x+y)^2-(x-y)^2=4xy=4(a^2+b^2),$$

hence $A=2a$, $B=2b$, $C=x-y$, $D=x+y$ is a solution. This concludes the proof.

Note that if we allow $x=y$, then $C=0$, and we get solutions of $A^2+B^2=D^2$, but only based on a, b and x chosen so that $a^2+b^2=x^2$.

By allowing a and b to increase through all possible values, and by taking each pair of unequal factors of a^2+b^2, all solutions of the equation can be tabulated. A partial tabulation appears below.

THEOREM II. *Any integer M is the sum of 2 squares, A^2+B^2, not both 0, if and only if it is of the form $2^n(4N+1)$, where N is the sum of two of the triangular numbers $\frac{1}{2}r(r+1)$, $(r=0, 1, 2, \cdots)$.*

If, of A and B, one is even and one is odd, then $n=0$; if both are odd, $n=1$; if both are even, $n>1$.

Proof. Let N be the sum of the triangular numbers $\frac{1}{2}a(a+1)$ and $\frac{1}{2}b(b+1)$.

a	b	a^2+b^2	x	y	$x-y$	$x+y$	Solution
1	1	2	2	1	1	3	$1^2+2^2+2^2=3^3$
1	2	5	5	1	4	6	$2^2+4^2+4^2=6^2$*
1	3	10	10	1	9	11	$2^2+6^2+9^2=11^2$
			5	2	3	7	$2^2+3^2+6^2=7^2$
2	2	8	8	1	7	9	$4^2+4^2+7^2=9^2$
			4	2	2	6	$2^2+4^2+4^2=6^2$*
1	4	17	17	1	16	18	$2^2+8^2+16^2=18^2$*
2	3	13	13	1	12	14	$4^2+6^2+12^2=14^2$*
1	5	26	26	1	25	27	$2^2+10^2+25^2=27^2$
			13	2	11	15	$2^2+10^2+11^2=15^2$
2	4	20	20	1	19	21	$4^2+8^2+19^2=21^2$
			10	2	8	12	$4^2+8^2+8^2=12^2$*
			5	4	1	9	$1^2+4^2+8^2=9^2$
3	3	18	18	1	17	19	$6^2+6^2+17^2=19^2$
			9	2	7	11	$6^2+6^2+7^2=11^2$
			6	3	3	9	$3^2+6^2+6^2=9^2$*

* By omitting solutions where x and y are both even or both odd, and those where a, b, x and y have a common factor, multiple and repeated solutions are avoided.

Then

$$2^{2m}(4N+1) = 2^{2m}(a-b)^2 + 2^{2m}(a+b+1)^2$$

and

$$2^{2m+1}(4N+1) = 2^{2m}(2a+1)^2 + 2^{2m}(2b+1)^2,$$

so that $2^n(4N+1)$ is the sum of two squares, with the stated restrictions for n.

For the converse, we shall first reduce the two cases where A and B are both even or both odd to the case of an odd and an even square.

(1) If $A=(2a+1)$, $B=(2b+1)$, then

$$\begin{aligned} A^2+B^2 &= (2a^2-4ab+2b^2)+(2a^2+4ab+2b^2+4a+4b+2) \\ &= 2[(a-b)^2+(a+b+1)^2], \end{aligned}$$

and of $a-b$ and $a+b+1$, one must be even and the other odd.

(2) If $A=2^p(2a+1)$, $B=2^q(2b+1)$, where $1 \leqq p \leqq q$, then, when $p<q$,

$$A^2 + B^2 = 2^{2p}[(2a+1)^2 + 2^{2(q-p)}(2b+1)^2],$$

and when $p=q$, by (1)

$$A^2 + B^2 = 2^{2p+1}[(a-b)^2 + (a+b+1)^2].$$

Thus in general the sum of two squares can be reduced to the sum of an odd and even square, multiplied by 2^n, where n satisfies the conditions of the theorem. It remains to prove that if

$$(2a)^2 + (2b+1)^2 = 4N + 1,$$

then N is the sum of 2 triangular numbers.

$$N = a^2 + b^2 + b,$$

so

$$N = \tfrac{1}{2}(a+b)(a+b+1) + \tfrac{1}{2}(a-b-1)(a-b),$$

and

$$N = \tfrac{1}{2}(a+b)(a+b+1) + \tfrac{1}{2}(b-a)(b-a+1),$$

which expresses N as the sum of two triangular numbers according as $a > b$ or $a \leqq b$, and the theorem is proved.

The relation between the sums of an odd and an even square, and the sums of two triangular numbers can be seen from the following arrays of such numbers:

0^2+1^2		0^2+3^2		0^2+5^2		0^2+7^2	
	1^2+2^2		1^2+4^2		1^2+6^2		1^2+8^2
		2^2+3^2		2^2+5^2		2^2+7^2	
			3^2+4^2		3^2+6^2		3^2+8^2
				4^2+5^2		4^2+7^2	
					5^2+6^2		

0+0		1+1		3+3		6+6	
	0+1		1+3		3+6		6+10
		0+3		1+6		3+10	
			0+6		1+10		3+15
				0+10		1+15	
					0+15		

ON THE EQUIVALENCE OF COMPLETENESS AND SEMI-COMPLETENESS FOR INTEGER SEQUENCES

J. L. BROWN, JR., The Pennsylvania State University

The notion of completeness for a sequence of positive integers was introduced by Hoggatt and King [1] in connection with a problem involving Fibonacci numbers and is defined as follows:

DEFINITION 1. *A sequence of positive integers,* $\{f_i\}_1^\infty$ *is said to be complete iff (if and only if) every positive integer n has a representation in the form*

$$n = \sum_{i=0}^{\infty} \alpha_i f_i, \tag{1}$$

where each α_i *is either zero or one; that is,* $0 \leqq \alpha_i \leqq 1$ *for* $i = 1, 2, 3, \cdots$.

Subsequently, it was shown ([2], Theorem 1, p. 558) that a nondecreasing sequence $\{f_i\}_1^\infty$ of positive integers with $f_1 = 1$ is complete iff

$$f_{n+1} \leqq 1 + \sum_{i=1}^{n} f_i \qquad \text{for } n = 1, 2, 3, \cdots. \tag{2}$$

More recently, Alder [3] has defined semi-completeness for such integer sequences:

DEFINITION 2. ([3], p. 147) *A sequence of positive integers* $\{f_i\}_1^\infty$ *is called semi-complete iff every positive integer n can be represented in the form*

$$n = \sum_{1}^{\infty} c_i f_i, \tag{3}$$

where the c_i *are nonnegative integers satisfying* $0 \leqq c_i \leqq k_i$ *for some initially prescribed set of positive integers,* $\{k_i\}_1^\infty$. (Strictly speaking, Alder restricts his definition to nondecreasing sequences, but for our purposes, the slightly broader definition is convenient.)

It is clear that completeness is subsumed under semi-completeness corresponding to the special selection, $k_i = 1$ for all $i = 1, 2, 3, \cdots$. As a result, semi-completeness is apparently more general than completeness, requiring separate methods for its investigation. The purpose of the present note is to show that the semi-completeness of a given integer sequence can be reduced to the consideration of completeness for an associated integer sequence which is essentially an augmented version of the original sequence arrived at by a certain repetition of terms. Once this equivalence is established, the characterization of semi-complete sequences obtained by Alder ([3], Theorem 1, p. 146) follows immediately from the known characterization for completeness ([2]).

In the following, $\{f_i\}_1^\infty$ is the given sequence of positive integers and $\{k_i\}_1^\infty$ is a fixed sequence of positive integers representing the coefficient bounds in Definition 2.

DEFINITION 3. *For a given pair of sequences,* $\{f_i\}_1^\infty$ *and* $\{k_i\}_1^\infty$, *the augmented sequence,* $\{g_i\}_1^\infty$, *is constructed as follows: For each positive integer* $i \geqq 1$, *there exists a unique positive integer,* $n_i \geqq 1$, *such that*

(4) $$1+\sum_{m=1}^{n_i-1} k_m \leqq i < 1+\sum_{m=1}^{n_i} k_m.$$

Define $g_i = f_{n_i}$ for $i=1, 2, 3, \cdots$.

In words, $\{g_i\}_1^\infty$ *is obtained from* $\{f_i\}_1^\infty$ *by repeating each* f_i, *in order,* k_i *times; that is, the g sequence has the form*

$$\left\{\underbrace{f_1, f_1, f_1}_{k_1 \text{ times}}, \underbrace{f_2, f_2, f_2, f_2}_{k_2 \text{ times}}, \cdots, \underbrace{f_i, f_i, f_i}_{k_i \text{ times}}, \cdots\right\}.$$

Our main result concerning the equivalence of the two completeness concepts is contained in the following theorem:

THEOREM. *The sequence* $\{f_i\}_1^\infty$ *is semi-complete iff the augmented sequence* $\{g_i\}_1^\infty$ *is complete.*

Proof. If $\{g_i\}_1^\infty$ is complete, then an arbitrary positive integer n can be represented in the form

(5) $$n = \sum_1^\infty \beta_i g_i = \sum_1^\infty \beta_i f_{n_i}$$

where each β_i is a nonnegative integer satisfying $0 \leqq \beta_i \leqq 1$. For $m=1, 2, \cdots$, define $A_m = \{i \mid n_i = m\}$. By the construction given in Definition 3, the set A_m contains exactly k_m members; furthermore, (5) may be rewritten

(6) $$n = \sum_{m=1}^\infty \left(\sum_{i\in A_m} \beta_i\right) f_m,$$

and the semi-completeness of $\{f_i\}_1^\infty$ is immediate.

Conversely, if $\{f_i\}_1^\infty$ is semi-complete, each positive integer n can be written in the form

(7) $$n = \sum_1^\infty \alpha_j f_j \quad \text{with} \quad 0 \leqq \alpha_j \leqq k_j.$$

Letting $i \to n_i$ denote the one-many mapping or correspondence specified in Definition 3, define

$$B_j = \{i \mid i \to j\} \qquad \text{for } j = 1, 2, 3, \cdots.$$

We may then rewrite (7) as

(8) $$n = \sum_{j=1}^\infty \sum_{i \in B_j} \gamma_i g_i,$$

where the γ_i are taken so that exactly α_j of the γ_i's are unity for $i \in B_j$ and the remaining $k_j - \alpha_j$ of the γ_i's are chosen equal to zero. Since $B_j \cap B_k = \varnothing$ for $j \neq k$, (8) establishes that the sequence $\{g_i\}_1^\infty$ is complete.

Theorem 1 of Alder's paper is an immediate corollary:

COROLLARY. *Let* $\{f_i\}_1^\infty$ *be a nondecreasing sequence of positive integers with* $f_1=1$ *and* $\{k_i\}_1^\infty$ *a fixed sequence of positive integers without order restrictions. Then* $\{f_i\}_1^\infty$ *is semi-complete iff*

$$f_{n+1} \leqq 1 + \sum_{i=1}^{n} k_i f_i \qquad \text{for } n = 1, 2, 3, \cdots. \tag{9}$$

Proof. By the preceding theorem, $\{f_i\}_1^\infty$ is semi-complete iff the augmented sequence $\{g_i\}_1^\infty$ is complete. Since $\{g_i\}_1^\infty$ is nondecreasing with $g_1=1$, Theorem 1 of [2] applies and $\{g_i\}_1^\infty$ is complete iff

$$g_{n+1} \leqq 1 + \sum_{i=1}^{n} g_i \qquad (n = 1, 2, \cdots). \tag{10}$$

By construction of the $\{g_i\}_1^\infty$, condition (10) obviously implies condition (9). Conversely, if (9) holds, then for $p \geqq k_1$,

$$g_p = f_{n_p} \leqq 1 + \sum_{1}^{n_p-1} k_i f_i = 1 + \sum^{p_0} g_i \leqq 1 + \sum^{p-1} g_i,$$

where

$$p_0 = \sum_{1}^{n_p-1} k_i.$$

Since $g_p=1$ for $p=1, 2, \cdots, k_1$, (10) is satisfied for all $n \geqq 1$.

Theorem 3 of [3] may also be proved by applying the corresponding theorem for complete sequences to the augmented sequence $\{g_i\}_1^\infty$.

In summary, semi-completeness and completeness for integer sequences have been shown to be equivalent in the sense that either may be regarded as a special case of the other. More particularly, we have proved that semi-completeness of a given sequence is equivalent to the completeness of a certain auxiliary sequence which is an augmented form of the original sequence. As a result, theorems involving the semi-completeness concept can be restated using only the apparently more restrictive concept of completeness.

Lastly, we observe as a matter of priority in connection with Theorem 2 of [3] that C. M. Sandwick [4] has stated the following theorem:

"*If* $f(n)$ *is a positive integer when* n *is a non-negative integer and if* $f_0=1$, *any nonnegative integer can be uniquely represented in the form*

$$\sum_{n=1}^{\infty} a_n \prod_{m=0}^{n-1} f(m), \qquad 0 \leqq a_n < f(n)."$$

References

1. V. E. Hoggatt and C. H. King, Problem E 1424, Am. Math. Monthly, **67**, No. 6 (1960), 593.

2. J. L. Brown, Jr., "Note on Complete Sequences of Integers," Am. Math. Monthly, **68**, No. 6 (1961), 557–560.

3. H. L. Alder, "The Number System in More General Scales," MATHEMATICS MAGAZINE, **35**, No. 2 (1962), 145–151.

4. C. M. Sandwick, Solution to Problem E 881, Am. Math. Monthly, **57**, No. 4 (1950), 262–264.

SOME PROBABILITY DISTRIBUTIONS AND THEIR ASSOCIATED STRUCTURES

NORMAN R. DILLEY, Huntington Beach High School, Huntington Beach, Calif.

PART II

In this section an application for the entries in the arrays which make up the extended binomial series will be presented after brief comment on some possible mappings for C or K.

The elements of C, the ordered positive integers, may be mapped upon a set of intersections generated by incident parallelograms or any other similar set of lattice points. One observes the usual metrical identity in this mapping. The elements of K, the extended binomial series, may map upon structures in a probability measure space. This type of space can achieve some metrical properties only by establishing isomorphisms. Schweizer and Sklar [4] designate this probability measure space as a Menger or Wald space. This space has some features which may be considered statistical metrical as well as topological.

Now to turn to an application for K in terms of distribution of discrete probability one may consider a process, to begin with, where the events are samplings without replacement from a sample space of two objects and where the process is repeated n times. In order to secure the unique discrete probabilities one may use the coefficients making up the binomial series. If samplings from a sample space of three objects are made without replacement, and the process is repeated n times, the probability distribution becomes the entries of K where $k=3$, and $n=1, 2, 3, \cdots, r$ (r would rarely be much greater than 10 because of the magnitude of the series at that point). If one extends the sample space to four objects then the probability distribution would utilize the entries from K where $k=4$, and $n=1, 2, 3, \cdots, r$. One understands that the denominator for the elements in each series will be $2^n, 3^n, 4^n, \cdots, r^n$ in order to determine the numerical value of the discrete probability.

At this point one may note the generalized probability proposition that may be stated as follows:

Given n consecutive possibility spaces (rows), over which $1, 2, \cdots$, or k samplings are performed (without replacement) on each possibility space for a random variable X_i, where X_i may be discovered on the 1st, 2nd, $\cdots$, but must be discovered on the kth sampling, then the system of sets K or $K_2, K_3, \cdots, K_m$ become the distributions of the discrete probability for locating n homogeneously distributed random variables $X_1, X_2, \cdots, X_n$ (one random variable being discovered on each of the n possibility spaces).

To prepare a simple demonstration of this proposition one may consider the basic process which involves the discrete probabilities. Suppose, now, one makes a listing of the samplings (choices, actions, etc.) performed, using the rule that 1's are written down for all samplings until the random variable X_i is discovered, then the 0's are written down as required to complete a row of k symbols. The row listings for $k=2$ comprise two possibilities: 1 0 (success on the first sampling), and 1 1 (success on the second sampling). For $k=3$, the row listings would be: 1 1 1, 1 1 0, or 1 0 0 (success on the third, the second, or the first sampling). For $k=4$, the row listings are: 1 1 1 1, 1 1 1 0, 1 1 0 0, 1 0 0 0. These individual row listings represent independent events with an equi-probability of $1/k$.

When the basic process is repeated n times, the consecutive row listings make up what may be considered to be matrices of 1's and 0's. For instance, if one allows $k=2$, 3, and 4 (and $n=4$ in the three cases), one may have, in the process considered, the following matrices:

$$H_2 = \begin{Vmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 0 \\ 1 & 1 \end{Vmatrix} \qquad H_3 = \begin{Vmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{Vmatrix} \qquad H_4 = \begin{Vmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{Vmatrix} \tag{12}$$

The H_2 matrix is the listing for the discovery of four random variables or $X_1, \cdots, X_4$ in exactly 6 samplings, the H_3 matrix is the listing for the discovery of $X_1, \cdots, X_4$ in exactly 8 samplings, and the H_4 matrix is the listing for the discovery of the four variables in exactly 10 samplings.

Now if one asks, what is the probability of discovering the four random variables $X_1, \cdots, X_4$ in H_2 taking exactly 6 samplings, in H_3 taking exactly 8 samplings, or in H_4 taking exactly 10 samplings, and in H_i taking exactly j samplings, one may make an observation that remains the basis for this demonstration of the stated proposition on probability. This observation is that the probability of the H_2 matrix becomes the number of the permutations of the inverse partitions of the integer 6 which generate by addition the positive number 6; that the probability of the H_3 matrix becomes the number of the permutations of the inverse partitions of the integer 8 which generate by addition the positive integer 8; that the probability of the H_4 matrix becomes the same, and that the probability of the H_i matrix becomes the number of the permutations of the inverse partitions made up of the n basis integers which generate by addition the positive integer j. One must note that one is restricted to partitions that involve only the basis integers of the specific set of integers.

The respective probabilities of the H_2, the H_3, and the H_4 matrix are then: 6/16 or 3/8, 12/81 or 4/27, and 24/256 or 3/32. Again one should note the use of the 2^4, 3^4, 4^4 or terms from the set n^k which gives the magnitude of the sample space or total number of events in the sample space.

There are several features of these probability distributions which are interesting and which serve to give models for various topics in mathematics. One may now, to introduce some of these features, note some of the characteristics of the algebra which may exist where the elements are all the possible matrices for $k=2, 3, \cdots, m$, and $n=1, 2, 3, \cdots, r$. This algebra will be designated as W and is characterized by the cardinal number, in terms of unique elements, of the sum: $2^n+3^n+4^n+ \cdots +r^n$. One may note, at this point, that this cardinal number is the number of unique search sequences connecting the X_i random variables. This algebra W would be a matrix algebra where the matrices are comprised of the symbols used in Boolean algebra which are the 1 and the 0. These symbols indicate only the vacuity and non-vacuity of a given point in a given set space. One does not associate a number with these matrices unless the measure operation is performed over the sequence. Since each matrix of the algebra W is a representation of a unique sequence that is a search for a random variable, then each matrix may also be considered a function in the probability measure space. If one allows the notion of the set operation of "intersection" to be analogous to the "zero" of a polynomial of the Cartesian plane, the reasonableness of the analogy between a matrix of 1's and 0's and a function in the set space is apparent. The algebra, W, has not been developed as far as it is known to the writer.

Since each sequence or space function intersects or converges (this would be convergence of the entire sequence towards either the optimum sequence or the maximum sequence) with all the other sequences for $X_1, X_2, \cdots, X_n$, one may note that an ordered intersection process must exist. Hence one may describe the algebra W in the following way:

$$W = W_2, W_3, \cdots, W_n \tag{13}$$

and,

$$\begin{aligned} W_2 &\rightarrow H_2 \cap H_3 \cap \cdots \cap H_r \\ W_3 &\rightarrow H_2 \cap H_3 \cap \cdots \cap H_r \\ &\cdots\cdots\cdots\cdots \\ W_n &\rightarrow H_2 \cap H_3 \cap \cdots \cap H_r \end{aligned} \tag{14}$$

where $H_2, H_3, \cdots, H_r$ are the sets of sequences associated with the sets of $K_2, K_3, \cdots, K_r$. One should recall that each K entry has that integral number of sequences and therefore the totality of K is the cardinal number of W. For instance, if the sequences were in the set of H_4 and $n=4$, then the algebra W_4 would be made up of 4^4 or 256 sequence functions.

The notion of a sequence is perhaps more suitable for a set space. Since in any sampling sequence one may require a minimum number of samplings, a maximum number of samplings, or some number intermediate between these two numbers, the sequences may be defined accordingly. The types of sequences become the H_m or maximum sequence, the H_o or optimum (least number of samplings to discover the random variables), and the student sequence or H_s. The obvious ordering of these sequences becomes:

$$H_o < H_{s1} < H_{s2} < \cdots < H_m. \tag{15}$$

One may note the order relation among the systems of sets introduced by the brief statement of:

$$T(C) \rightarrow K \rightarrow W(H). \tag{16}$$

The entries in K (where the familiar case would be the binomial coefficients) may also be generated by the usual expansion process discussed in elementary algebra texts. If one extends this expansion process, the set of functions generated may be designated as the set U where the coefficients become the entries of K. Let $U = U_2, U_3, \cdots, U_k$, then

$$\begin{array}{lll} U_2 = (x^0 + x^1)^2, & (x^0 + x^1)^3, \cdots, (x^0 + x^1)^n & \\ U_3 = (x^0 + x^1 + x^2)^2, & (x^0 + x^1 + x^2)^3, \cdots, (x^0 + x^1 + x^2)^n & \\ \cdots & \cdots & (17) \\ U_k = (x^0 + x^1 + \cdots + x^k)^2, & (x^0 + x^1 + \cdots + x^k)^3, \cdots, (x^0 + x^1 + \cdots + x^k)^n. & \end{array}$$

The set U provides a projection method to associate the entries of K as coefficients of functions.This may also be considered a type of transformation.

The fact that one begins with the ordered positive integers or C and by a transformation T changes them into entries of K, then represents them as matrices of 1's and 0's in W, leads one to search for the next transformation which will return the elements of W back to C. It appears that the general and elegant concept of the Haar integral [5] or the Haar measure of a given set or group (named after Alfred Haar, 1835–1933, a Hungarian mathematician) will provide this. If $I(W)$ is the Haar measure in which one may attach to each point, throughout the structure, the measure of one, then the following must exist,

$$I(W) \rightarrow I(W_2), \qquad I(W_3), \cdots, I(W_n) \rightarrow C. \tag{18}$$

This operation, while somewhat formal in appearance, is a counting of all the symbols of the 1 in all the matrices of W. One should note that Haar measure of the matrices or elements of W generates the ordered positive integers of C.

At this point it is desirable to attempt a brief summary of the discussion of this paper. Initially, one began with ordered sets of the positive integers which was called C. Next a transformation, which in detail was a permutation of the partitions of each integer of C into the basis integers, provided entries of the sets of K. These were shown to be distribution functions of the discrete probability for locating the X_i random variables. Next a proposition on the probability of an event where the event was an entire sampling sequence for X_i random variables was stated. The existence of the sampling sequence and its representation as a matrix of 1's and 0's was discussed. The sets of sequences was designated as H, and then a possible algebra or W using the matrices of 1's and 0's as elements was pointed out. An alternative realization of the entries of K as coefficients of functions was given by the sets of U. Finally the Haar measure was shown to provide the inverse transformation that projected the entries of K from the probability measure space back into the sets of the ordered positive integers of C.

Readers whose interest in the topics of this paper is genuine may be referred to Professor Loève's [6] masterful development of the subject of probability.

The writer would like to acknowledge the interest and assistance of Dr. Joseph W. Rigney, Director, Electronics Personnel Research, University of Southern California, during this investigation. Thanks are also due to Dr. Donald H. Schuster, Collins Radio Co., for helping the writer make contact with the problem of random variables. Special acknowledgment must be given to Dr. Dewey C. Duncan, and to Dr. D. V. Steed, Department of Mathematics, University of Southern California, for an introduction to some of the concepts of modern mathematics during a National Science Foundation Institute.

References

4. Schweizer, B., and Sklar, A., *Statistical Metric Spaces*, Pacific Journal of Mathematics, Vol. 10, No. 1, 1960, p. 2.

5. Loomis, Lynn H., *An Introduction to Abstract Harmonic Analysis*, D. Van Nostrand Co., New York, 1953, p. 107.

6. Loève, Michel, *Probability Theory*, D. Van Nostrand Co., Inc., New York, 1960. See Chapters II, III, IV, and V.

A SELF-DEFINING INFINITE SEQUENCE, WITH AN APPLICATION TO MARKOFF CHAINS AND PROBABILITY

ALEXANDER NAGEL, Bronx High School of Science

PART II

A completely generalized sequence S is defined as follows:

a) A term of the sequence is one of the numbers $T_1, T_2, T_3, \cdots, T_n$

b) S is composed of an infinite succession of finite segments $C_1, C_2, C_3, \cdots, C_i, \cdots$, so that

$$S = C_1C_2C_3 \cdots C_1 \cdots$$

c) The first two segments are given.

d) Each segment C_{i+1} for $i \geqq 2$ is generated from the segment C_i according to the following rule:

Replace each term T_k in C_i by the corresponding set of terms found in the table below:

TABLE I

Term in C_i	Set of terms generated (#T_1's	#T_2's	#T_3's	$\cdots$	#T_n's)
T_1	α_1	α_2	α_3		α_n
T_2	β_1	β_2	β_3		β_n
T_3	γ_1	γ_2	γ_3		γ_n
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
T_n	μ_1	μ_2	μ_3		μ_n

where the Greek letters with subscripts indicate the number of terms generated. We make the assumption that all these constants are greater than zero.

Let a_i be the number of T_1's in C_i.
Let b_i be the number of T_2's in C_i.
Let c_i be the number of T_3's in C_i.
. . . .
Let m_i be the number of T_n's in C_i.

According to definition d)

$$a_{i+1} = a_i\alpha_1 + b_i\beta_1 + c_i\gamma_1 + \cdots + m_i\mu_1 \qquad (11a)$$

$$b_{i+1} = a_i\alpha_2 + b_i\beta_2 + c_i\gamma_2 + \cdots + m_i\mu_2 \qquad (11b)$$

$$c_{i+1} = a_i\alpha_3 + b_i\beta_3 + c_i\gamma_3 + \cdots + m_i\mu_3 \tag{11c}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$m_{i+1} = a_i\alpha_n + b_i\beta_n + c_i\gamma_n + \cdots + m_i\mu_n \tag{11n}$$

Also,

Let A_i be the total number of T_1's in the partial sequence S_i.
Let B_i be the total number of T_2's in the partial sequence S_i.
Let C_i be the total number of T_3's in the partial sequence S_i.
. . . .
Let M_i be the total number of T_n's in the partial sequence S_i.

Accordingly

$$A_i = \sum_1^i a_k \tag{12a}$$

$$B_i = \sum_1^i b_k \tag{12b}$$

$$C_i = \sum_1^i c_k \tag{12c}$$

$$\cdots\cdots$$

$$M_i = \sum_1^i m_k \tag{12n}$$

We now prove that

$$A_{i+1} = A_i\alpha_1 + B_i\beta_1 + C_i\gamma_1 + \cdots + M_i\mu_1 \tag{13a}$$

$$B_{i+1} = B_i\alpha_2 + B_i\beta_2 + C_i\gamma_2 + \cdots + M_i\mu_2 \tag{13b}$$

$$C_{i+1} = A_i\alpha_3 + B_i\beta_3 + C_i\gamma_3 + \cdots + M_i\mu_3 \tag{13c}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$M_{i+1} = A_i\alpha_n + B_i\beta_n + C_i\gamma_n + \cdots + M_i\mu_n \tag{13n}$$

In doing this, we will show only (13a) to be true, but the method is general.

Proof. For $k \geqq 2$, we have from (11a)

$$a_{k+1} = a_k\alpha_1 + b_k\beta_1 + c_k\gamma_1 + \cdots + m_k\mu_1. \tag{i}$$

Therefore

$$\sum_2^i a_{k+1} = \alpha_1\sum_2^i a_k + \beta_1\sum_2^i b_k + \gamma_1\sum_2^i c_k + \cdots + \mu_1\sum_2^i m_k. \tag{ii}$$

But

$$\sum_2^i a_{k+1} = \sum_1^{i+1} a_k - (a_1 + a_2) = A_{i+1} - (a_1 + a_2) \tag{iii}$$

$$\left.\begin{aligned} \sum_2^i a_k &= \sum_1^i a_k - a_1 = A_i - a_1 \\ \sum_2^i b_k &= \sum_1^i b_k - b_1 = B_i - b_1 \\ &\cdots\cdots\cdots\cdots \\ \sum_2^i m_k &= \sum_1^i m_k - m_1 = M_i - m_1 \end{aligned}\right\} \tag{iv}$$

Therefore, from (ii), (iii) and (iv)

$$\begin{aligned} -(a_1 + a_2) + A_{i+1} = A_i\alpha_1 + B_i\beta_1 + C_i\gamma_1 + \cdots + M_i\mu_1 \\ - (a_1\alpha_1 + b_1\beta_1 + c_1\gamma_1 + \cdots + m_1\mu_1). \end{aligned} \tag{v}$$

We now define C_1 and C_2 so that

$$\left.\begin{aligned} a_1 + a_2 &= a_1\alpha_1 + b_1\beta_1 + c_1\gamma_1 + \cdots + m_1\mu_1 \\ b_1 + b_2 &= a_1\alpha_2 + b_1\beta_2 + c_1\gamma_2 + \cdots + m_1\mu_2 \\ c_1 + c_2 &= a_1\alpha_3 + b_1\beta_3 + c_1\gamma_3 + \cdots + m_1\mu_3 \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ m_1 + m_2 &= a_1\alpha_n + b_1\beta_n + c_1\gamma_n + \cdots + m_1\mu_n \end{aligned}\right\} \tag{vi}$$

Therefore, from (v) and (vi)

$$A_{i+1} = A_i\alpha_1 + B_i\beta_1 + C_i\gamma_1 + \cdots + M_i\mu_1. \qquad \text{Q.E.D.}$$

Now let N_i be the number of terms in S_i, and let t_k be the general term of the sequence. Again we have:

a) The total number of terms in S_i equals

$$N_i = A_i + B_i + C_i + \cdots + M_i \tag{14}$$

b) The sum of the N_i terms of S_i equals

$$\sum_1^{N_i} t_k = T_1A_i + T_2B_i + T_3C_i + \cdots + T_nM. \tag{15}$$

From (14) and (15), the average of the first N_i terms of the infinite sequence S equals:

$$\begin{aligned} \frac{\sum^{N_i} t_k}{N_i} &= \frac{T_1A_i + T_2B_i + T_3C_i + \cdots + T_nM_i}{A_i + B_i + C_i + \cdots + M_i} \\ &= \frac{T_1A_i}{A_i + B_i + \cdots + M_i} + \frac{T_2B_i}{A_i + B_i + \cdots + M_i} + \cdots \\ &\quad + \frac{T_nM_i}{A_i + B_i + \cdots + M_i} \end{aligned} \tag{16}$$

$$= \frac{T_1}{1 + \frac{B_i}{A_i} + \cdots + \frac{M_i}{A_i}} + \frac{T_2}{\frac{A_i}{B_i} + 1 + \cdots + \frac{M_i}{B_i}} + \cdots$$
$$+ \frac{T_n}{\frac{A_i}{M_i} + \frac{B_i}{M_i} + \cdots + 1}.$$

As in the special case discussed in Section I, we could now substitute from (13a) through (13n) for A_i through M_i. However, this would lead to an unmanageable expression. Instead, it is possible to find in a different way each of the ratios in the denominators of (16), in terms of the constant α's, β's, γ's, etc.

We first calculate the total number of these ratios. There are $(n-1)$ ratios whose denominators are A_i. If we do not count reciprocals of ratios already counted, there are $(n-2)$ ratios whose denominators are B_i, $(n-3)$ ratios whose denominators are C_i, etc. The total number of ratios under discussion therefore is

$$\sum_1^n (n-k) = n(n-1)/2.$$

If we take every possible pair of the equations (13a) through (13n) and divide one member of each pair by the other, we will have a new set of equations. The number of these new w equations will again be $n(n-1)/2$, since their number is ${}_nC_2$. The equations are:

$$\begin{aligned} \frac{A_{i+1}}{B_{i+1}} &= \frac{A_i\alpha_1 + B_i\beta_1 + C_i\gamma_1 + \cdots + M_i\mu_1}{A_i\alpha_2 + B_i\beta_2 + C_i\gamma_2 + \cdots + M_i\mu_2} \\ \frac{A_{i+1}}{C_{i+1}} &= \frac{A_i\alpha_1 + B_i\beta_1 + C_i\gamma_1 + \cdots + M_i\mu_1}{A_i\alpha_3 + B_i\beta_3 + C_i\gamma_3 + \cdots + M_i\mu_3} \\ &\cdots\cdots\cdots\cdots\cdots \end{aligned} \tag{17}$$

These equations can be rewritten as follows:

$$\begin{aligned} \frac{A_{i+1}}{B_{i+1}} &= \frac{\left(\frac{A_i}{B_i}\right)\alpha_1 + \beta_1 + \left(\frac{C_i}{B_i}\right)\gamma_1 + \cdots + \left(\frac{M_i}{B_i}\right)\mu_1}{\left(\frac{A_i}{B_i}\right)\alpha_2 + \beta_2 + \left(\frac{C_i}{B_i}\right)\gamma_2 + \cdots + \left(\frac{M_i}{B_i}\right)\mu_2} \\ \frac{A_{i+1}}{C_{i+1}} &= \frac{\left(\frac{A_i}{C_i}\right)\alpha_1 + \left(\frac{B_i}{C_i}\right)\beta_1 + \gamma_1 + \cdots + \left(\frac{M_i}{C_i}\right)\mu_1}{\left(\frac{A_i}{C_i}\right)\alpha_3 + \left(\frac{B_i}{C_i}\right)\beta_3 + \gamma_3 + \cdots + \left(\frac{M_i}{C_i}\right)\mu_3} \\ &\cdots\cdots\cdots\cdots\cdots \end{aligned} \tag{18}$$

Let us accept, for the moment without proof, that as i approaches infinity, each of these ratios is equal to some constant, though not necessarily the same for each ratio. We can then drop subscripts, and so obtain the following:

$$\frac{A}{B}=\frac{\left(\frac{A}{B}\right)\alpha_1+\beta_1+\left(\frac{C}{B}\right)\gamma_1+\cdots+\left(\frac{M}{B}\right)\mu_1}{\left(\frac{A}{B}\right)\alpha_2+\beta_2+\left(\frac{C}{B}\right)\gamma_2+\cdots+\left(\frac{M}{B}\right)\mu_2}$$

$$\frac{A}{C}=\frac{\left(\frac{A}{C}\right)\alpha_1+\left(\frac{B}{C}\right)\beta_1+\gamma_1+\cdots+\left(\frac{M}{C}\right)\mu_1}{\left(\frac{A}{C}\right)\alpha_3+\left(\frac{B}{C}\right)\beta_3+\gamma_3+\cdots+\left(\frac{M}{C}\right)\mu_3} \tag{19}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

If we regard each ratio as an unknown, we have a set of $n(n-1)/2$ simultaneous quadratic equations; and provided that the equations are not redundant or inconsistent, it is possible to solve for each of the unknowns in terms of the constant α's, β's, γ's, etc. However, these unknowns are simply the ratios in the denominators of (16) if subscripts are dropped. According to our assumption, dropping subscripts means that the number i of segments increases without limit. Therefore, if the values for these ratios in terms of the α's, β's, γ's, etc. are substituted in the right hand member of (16), the value of this member is the limit, as i approaches infinity, of the average of the first N_i terms of the sequence S, i.e., this value is $\lim_{i\to\infty}\sum_1^{N_i} t_k/N_i$. However, N_i is a monotonically increasing function of i and approaches infinity with i, so that the value just obtained equals $\lim_{N_i\to\infty}\sum_1^{N_i} t_k/N_i$.

We must now prove the assumption introduced above, namely, that as i approaches infinity, each quotient formed by dividing any right hand member of (13a) through (13n) by any other right hand member equals some constant, though not necessarily the same constant for each quotient. To establish this, we will perform an induction on the number n of distinct terms in the sequence S.

But before we proceed to do this, we must recall that if $n=2$ there are only two equations, each right member being the sum of two forms, as follows:

$$A_{i+1} = A_i\alpha_1 + B_i\beta_1$$
$$B_{i+1} = A_i\alpha_2 + B_i\beta_2.$$

Similarly, if $n=k$, there are k equations, each right member being the sum of k terms. We will first prove the assumption for the case of $n=2$. Next, we will consider a set of n equations (i.e., the case when there are n distinct terms in S), in which the right member of each is the sum of n products. From these equations we will construct a set of $(n-1)$ equations, each right member of which will be the sum of $(n-1)$ products, so that the $(n-1)$ equations correspond to the case when there are $(n-1)$ distinct terms in S. However, the $(n-1)$ equations will be such that, if the assumption holds for them it also holds for the n equations from which they were constructed. In other words, we will show that if the assumption holds for $(n-1)$ distinct terms in S, it holds for n distinct terms in S. We will thus prove by the principle of mathematical induction that the assumption holds for all $n \geq 2$.

We proceed to show that the assumption holds when $n=2$. Let the two equations corresponding to this case be:

$$A_{i+1} = A_i\alpha_1 + B_i\beta_1 + \gamma_1 \tag{20}$$

$$B_{i+1} = A_i\alpha_2 + B_i\beta_2 + \gamma_2 \tag{21}$$

where γ_1 and γ_2 are constants. (These constants, which may be zero, are introduced into these equations for the sake of the proof by mathematical induction, since each of the $(n-1)$ equations we will presently construct will also have a constant term in its right member. There is clearly no loss of generality from the introduction of these constants.)

Dividing (20) by (21) we obtain:

$$\frac{A_{i+1}}{B_{i+1}} = \frac{A_i\alpha_1 + B_i\beta_1 + \gamma_1}{A_i\alpha_2 + B_i\beta_2 + \gamma_2} \tag{22a}$$

$$= \frac{\dfrac{A_i}{B_i}\alpha_1 + \beta_1 + \dfrac{\gamma_1}{B_i}}{\dfrac{A_i}{B_i}\alpha_2 + \beta_2 + \dfrac{\gamma_2}{B_i}} \tag{22b}$$

$$= k_1 + \frac{k_2\dfrac{A_i}{B_i} + \dfrac{k_3}{B_i}}{\dfrac{A_i}{B_i}\alpha_2 + \beta_2 + \dfrac{\gamma_2}{B_i}}$$

$$= k_1 + \frac{k_2 + \dfrac{k_3}{A_i}}{\alpha_2 + \dfrac{\gamma_2}{A_i} + \beta_2 \bigg/ \dfrac{A_i}{B_i}}$$

where

$$k_1 = \beta_1/\beta_2$$

$$k_2 = (\alpha_1\beta_2 - \alpha_2\beta_1)/\beta_2$$

$$k_3 = (\gamma_1\beta_2 - \gamma_2\beta_1)/\beta_2.$$

The limit, as i approaches infinity, of A_{i+1}/B_{i+1} is now equal to:

$$\lim_{i\to\infty} \frac{A_{i+1}}{B_{i+1}} = k_1 + \frac{k_2}{\alpha_2 + \beta_2 \bigg/ \lim\limits_{i\to\infty} \dfrac{A_i}{B_i}}$$

$$= k_1 + \frac{k_2}{\alpha_2} + \frac{\beta_2}{\lim\limits_{i\to\infty} \dfrac{A_i}{B_i}}.$$

By continued resubstitution, the limit of A_{i+1}/B_{i+1} can now be expanded into the continued fraction:

$$\lim_{i\to\infty}\frac{A_{i+1}}{B_{i+1}} = k_1 + \frac{k_2}{\alpha_2+}\,\frac{\beta_2}{k_1+}\,\frac{k_2}{\alpha_2+}\,\frac{\beta_2}{k_1+\cdots}. \tag{23}$$

The only question left to be answered is under what conditions this fraction converges to a finite limit. A theorem in the general theory of continued fractions states [2] that, if a continued fraction is written as

$$\frac{b_1}{a_1+}\,\frac{b_2}{a_2+}\,\frac{b_3}{a_3+\cdots}$$

it will have a definite value if the limit of $a_n a_{n+1}/b_{n+1}$ is greater than zero when n approaches infinity. This theorem can now be applied to the continued fraction (23). Omitting the initial k_1 (the addition of a constant will not effect the convergence of the fraction) the value of $a_n a_{n+1}/b_{n+1}$ will be either $k_1\alpha_2/k_2$ or $k_1\alpha_2/\beta_2$, depending on whether n is odd or even. These two values are respectively $\beta_1\alpha_2/\alpha_1\beta_2-\alpha_2\beta_1$ and $\beta_1\alpha_2/\beta_2$.

If $\alpha_1\beta_2-\alpha_2\beta_1>0$, then clearly, since all the constants are assumed positive, the value of $a_n a_{n+1}/b_{n+1}$ will always be greater than zero, and so will its limit. If $\alpha_1\beta_2-\alpha_2\beta_1<0$ then $\alpha_2\beta_1-\alpha_1\beta_2>0$, then by dividing (21) by (20), and by using analogous arguments, it can be shown that the fraction obtained is still convergent, and that $\lim_{i\to\infty} B_{i+1}/A_{i+1}$ converges. Finally, if $\alpha_1\beta_2=\alpha_2\beta_1$, then, by taking the limit of (22b), we obtain:

$$\lim_{i\to\infty}\frac{A_{i+1}}{B_{i+1}} = \frac{\lim\dfrac{A_i}{B_i}\alpha_1+\beta_1}{\lim\dfrac{A_i}{B_i}\alpha_2+\beta_2}$$

$$= \frac{\beta_1}{\beta_2}\,\frac{\lim\dfrac{A_i}{B_i}\alpha_1\beta_2+\beta_1\beta_2}{\lim\dfrac{A_i}{B_i}\alpha_2\beta_1+\beta_1\beta_2}$$

$$= \frac{\beta_1}{\beta_2}\quad\text{(constant)}.$$

Thus we have proved that our assumption is true for the case when n equals two.

We next consider an infinite sequence S, each of whose terms is one of n distinct numbers. The n equations corresponding to this case, again with constants added, are:

$$A_{i+1} = A_i\alpha_1 + B_i\beta_1 + \cdots + L_i\lambda_1 + M_i\mu_1 + \nu_1 \tag{24a}$$

$$B_{i+1} = A_i\alpha_2 + B_i\beta_2 + \cdots + L_i\lambda_2 + M_i\mu_2 + \nu_2 \tag{24b}$$

. . . .

[2] See Hall and Knight, *Higher Algebra*, page 362.

. . . .

$$L_{i+1} = A_i\alpha_{n-1} + B_i\beta_{n-1} + \cdots + L_i\lambda_{n-1} + M_i\mu_{n-1} + \nu_{n-1} \qquad (24[\mathrm{n}-1])$$

$$M_{i+1} = A_i\alpha_n + B_i\beta_n + \cdots + L_i\lambda_n + M_i\mu_n + \nu_n. \qquad (24\mathrm{n})$$

Dividing (24a) through (24[n−1]) by M_i, and making the following substitutions:

$$\overline{A}_{i+1} = A_{i+1}/M_i \qquad \overline{A}_i = A_i/M_i$$

$$\overline{B}_{i+1} = B_{i+1}/M_i \qquad \overline{B}_i = B_i/M_i$$

. . . .

$$\overline{L}_{i+1} = L_{i+1}/M_i \qquad \overline{L}_i = L_i/M_i$$

we obtain:

$$\overline{A}_{i+1} = \overline{A}_i\alpha_1 + \overline{B}_i\beta_1 + \cdots + \overline{L}_i\lambda_1 + \mu_1 + \nu_1/M_i$$

$$\overline{B}_{i+1} = \overline{A}_i\alpha_2 + \overline{B}_i\beta_2 + \cdots + \overline{L}_i\lambda_2 + \mu_2 + \nu_2/M_i$$

. . . .

$$\overline{L}_{i+1} = \overline{A}_i\alpha_{n-1} + \overline{B}_i\beta_{n-1} + \cdots + \overline{L}_i\lambda_{n-1} + \mu_{n-1} + \nu_{n-1}/M_i.$$

Taking the limit, as i approaches infinity, of both sides, we obtain:

$$\lim_{i\to\infty} \overline{A}_{i+1} = \lim_{i\to\infty} (\overline{A}_i\alpha_1 + \overline{B}_i\beta_1 + \cdots + \overline{L}_i\lambda_1 + \mu_1) \qquad (25\mathrm{a})$$

$$\lim_{i\to\infty} \overline{B}_{i+1} = \lim_{i\to\infty} (\overline{A}_i\alpha_2 + \overline{B}_i\beta_2 + \cdots + \overline{L}_i\lambda_2 + \mu_2) \qquad (25\mathrm{b})$$

. . . .

$$\lim_{i\to\infty} \overline{L}_{i+1} = \lim_{i\to\infty} (\overline{A}_i\alpha_{n-1} + \overline{B}_i\beta_{n-1} + \cdots + \overline{L}_i\lambda_{n-1} + \mu_{n-1}). \qquad (25[n-1])$$

It should be noted that we now have the limit, as i approaches infinity, of $(n-1)$ equations that can be taken to refer to an infinite sequence S, each of whose terms is one of $(n-1)$ distinct numbers.

To complete the proof, suppose that our assumption (e.g. that $\lim_{i\to\infty} \overline{A}_{i+1}/\overline{B}_{i+1}$ is a constant) holds for the case when each term of S is any one of $(n-1)$ distinct numbers. Then each quotient formed by dividing any left member of (25a) through (25[$n-1$]) by any other left member is a constant, since, for example,

$$\lim_{i\to\infty} \frac{\overline{A}_{i+1}}{\overline{B}_{i+1}} = \frac{\lim \overline{A}_{i+1}}{\lim \overline{B}_{i+1}}.$$

But then

$$\frac{\lim \overline{A}_{i+1}}{\lim \overline{B}_{i+1}} = \frac{\lim A_{i+1}/M_i}{\lim B_{i+1}/M_i} = \frac{\lim A_{i+1}/\lim M_i}{\lim B_{i+1}/\lim M_i}.$$

$$= \frac{\lim A_{i+1}}{\lim B_{i+1}}$$

$$= \lim_{i\to\infty} \frac{A_{i+1}}{B_{i+1}}$$

so that lim (A_{i+1}/B_{i+1}) equals a constant. Similarly, each of the $\binom{n-1}{2}$ possible quotients obtained from (25) can be set equal to some constant, and in consequence, each of the $(n-1)(n-2)/2$ ratios involving $A_{i+1}, B_{i+1}, \cdots, L_{i+1}$ can be shown to be equal to a constant. However the total number of ratios obtainable from (13), (or (24)), which must be proved to be constant is $n(n-1)/2$, of which those we have just shown to be constant are only a proper subset. The difference between this total number of ratios and the number in this subset (i.e. between $n(n-1)/2$ and $(n-1)(n-2)/2$) is $(n-1)$. However, these remaining $(n-1)$ ratios which must be proved constant are precisely those involving M_{i+1} that were not obtained in the construction above because (24n) was ignored in the construction. But instead of ignoring this equation, we could ignore any one of the other $(n-1)$ equations in the set of n equations making up (24), and we could then show by an argument exactly like the one just used that the remaining $(n-1)$ ratios in question are also constants. In other words, all of the $\binom{n}{2}$ ratios are constants, so that the assumption holds for the case when each term in S is one of n distinct numbers, provided that it holds when each term in S is any one of $(n-1)$ distinct numbers.

This completes the proof of the statement, since we have shown that the assumption holds for $n=2$, and also that if it holds for $n-1$ it must hold for n.

Answers

A318. When two balls collide they will just exchange velocities. A simpler way of looking at this is to image the balls passing through each other. If we arrange the velocities in monotonic order, we will obtain $\binom{N}{2}$ collisions. That this is maximum follows by considering the worldliness of the balls (s vs. t). The maximum number of points of intersection of N straight lines is $\binom{N}{2}$. If we have an elastic wall at one point of the line, the maximum number of collisions will be doubled.

A319. By symmetry, one factor must be $(a+b+c)$ and another factor must contain squared terms and terms of the form $-ab$ so that in the product, terms of the form a^2b will vanish, so $a^3+b^3+c^3=(a+b+c)(a^2+b^2+c^2-ab-bc-ca)$.

A320. Use $|u|=\sqrt{u^2}$ and assume that $|a|+|b|<|a+b|$. Then $\sqrt{a^2}+\sqrt{b^2}<\sqrt{(a+b)^2}$ or $a^2+2\sqrt{ab}+b^2<a^2+2ab+b^2$ which implies $|ab|<ab$, a contradiction. Therefore $|a|+|b|\geqq|a+b|$.

A321. The required condition is that n be composite. The partition obviously does not exist if n is prime, and if $n=ab$ where $a>1$ and $b>1$, the partition into a, b, and $ab-a-b$ ones does the job.

A322. Since $3ax^2+b=0$ gives $-b/3a$ as the product of the equal roots, and since the single root times $-b/3a$ equals $-c/a$, the single root is $-c/a\cdot 3a/-b=3c/b$.

(Quickies on page 270)

ISOTONE AND ANTITONE FRACTIONS

H. W. GOULD, West Virginia University

Garrett Birkhoff coined the words "isotone" and "antitone." If $a<b$ implies $f(a)<f(b)$ we call f an isotone function. If $a<b$ implies $f(a)>f(b)$ we call f an antitone function. We may apply this terminology to fractions by defining a function f by the relation

$$(1)\qquad f\left(\frac{p}{q}\right)=\frac{p+k}{q+k},\qquad k>0.$$

The example

$$\frac{1}{2}<\frac{3}{4}\quad\text{implies}\quad\frac{2}{3}<\frac{4}{5}$$

gives an illustration of isotone fractions. In fact in this example we may continue adding one's to the four elements of the fractions so obtained and always obtain the isotone relation. However it is not always true that fractions will have the isotone property. The example

$$\frac{1}{2}<\frac{5}{8},\qquad\frac{2}{3}=\frac{6}{9},\qquad\frac{3}{4}>\frac{7}{10}$$

shows how the sense of order may be reversed. The question arises then as to just when we may be confident that from a beginning inequality we may have the isotone property when using (1).

THEOREM 1. *Let $1<a<b$ and $1<k$. Then we have*

$$\frac{a-1}{b-1}<\frac{ka-1}{kb-1};$$

$$\frac{a}{b}=\frac{ka}{kb};$$

$$\frac{a+1}{b+1}>\frac{ka+1}{kb+1}.$$

Proof. The second relation is trivial. To establish the first inequality we form the difference between the two fractions and find that

$$\frac{ka-1}{kb-1}-\frac{a-1}{b-1}=\frac{(b-a)(k-1)}{(b-1)(kb-1)}.$$

Since $b-a>0$, $k-1>0$, $b-1>0$, and $kb-1>0$ the result follows. In the same way we have

$$\frac{a+1}{b+1}-\frac{ka+1}{kb+1}=\frac{(b-a)(k-1)}{(b+1)(kb+1)},$$

from which the third relation is evident.

Our first theorem allows us to manufacture any number of simple examples of fractions not possessing the isotone property. Thus we have:

$$\frac{2}{3} < \frac{5}{7}, \qquad \frac{3}{4} = \frac{6}{8}, \qquad \frac{4}{5} > \frac{7}{9}; \qquad (a = 3,\ b = 4,\ k = 2)$$

$$\frac{3}{4} < \frac{7}{9}, \qquad \frac{4}{5} = \frac{8}{10}, \qquad \frac{5}{6} > \frac{9}{11}; \qquad (a = 4,\ b = 5,\ k = 2)$$

$$\frac{e-1}{\pi-1} < \frac{e^2-1}{e\pi-1}, \qquad \frac{e}{\pi} = \frac{e^2}{e\pi}, \qquad \frac{e+1}{\pi+1} > \frac{e^2+1}{e\pi+1}; \qquad (a = e,\ b = \pi,\ k = e).$$

We may improve on the third relation in our theorem as follows.

THEOREM 2. *Let* $0<a<b$, $1<k$, $0<n$. *Then*

$$\frac{a+n}{b+n} > \frac{ka+n}{kb+n}.$$

Indeed we have the difference

$$\frac{a+n}{b+n} - \frac{ka+n}{kb+n} = \frac{(b-a)(k-1)n}{(b+n)(kb+n)} > 0$$

by our hypotheses. Note that we do not require a, b, k, n to necessarily be integers, although these are likely to be the most common cases. For an example, take $a=3$, $b=4$, $k=2$, and we get for successive values of n:

$$\frac{5}{6} > \frac{8}{10}, \qquad \frac{6}{7} > \frac{9}{11}, \qquad \frac{7}{8} > \frac{10}{12}, \qquad \frac{8}{9} > \frac{11}{13}, \qquad \text{et cetera.}$$

Let us now assume that $0<a<b$, $0<m<n$, and

$$\frac{a}{b} < \frac{m}{n} \tag{2}$$

and examine the relation of $(a+1)/(b+1)$ to $(m+1)/(n+1)$. We have

$$\frac{m+1}{n+1} - \frac{a+1}{b+1} = \frac{(bm-an)+(b-a)-(n-m)}{(b+1)(n+1)}$$

and we know that $b-a>0$, $n-m>0$, and $\delta=bm-an>0$. The denominator $(b+1)(n+1)$ is positive, so we have only to be concerned with the numerator. The fraction (2) will be isotone or antitone under the mapping defined by (1) according as the expression $\delta+(b-a)-(n-m)$ is greater or less than 0. A study of this relation leads us to offer the following theorem.

THEOREM 3. *Let* $a>0$, $m>0$, *and* $\epsilon>0$. *Then*

$$\frac{a}{a+\epsilon} < \frac{m}{m+\delta} \quad \text{and} \quad \frac{a+1}{a+\epsilon+1} < \frac{m+1}{m+\delta+1} \qquad (\mathit{ISOTONE}) \tag{3}$$

provided that

$$0 < \delta < \min\left(\frac{m}{a}\,\epsilon, \frac{m+1}{a+1}\,\epsilon\right);$$

(4) $$\frac{a}{a+\epsilon} < \frac{m}{m+\delta} \quad \textit{and} \quad \frac{a+1}{a+\epsilon+1} > \frac{m+1}{m+\delta+1} \qquad (\textit{ANTITONE})$$

provided that

$$0 < \frac{m+1}{a+1}\,\epsilon < \delta < \frac{m}{a}\,\epsilon;$$

(5) $$\frac{a}{a+\epsilon} < \frac{m}{m+\delta} \quad \textit{and} \quad \frac{a+1}{a+\epsilon+1} = \frac{m+1}{m+\delta+1} \qquad (\textit{EQUITONE})$$

provided that

$$(a+1)\delta = (m+1)\epsilon \quad \textit{and} \quad a\delta < m\epsilon.$$

Proof. We have for the difference between the fractions

$$\frac{a+1}{a+\epsilon+1} - \frac{m+1}{m+\delta+1} = \frac{\delta(a+1) - \epsilon(m+1)}{(a+\epsilon+1)(m+\delta+1)}\cdot$$

It is easy to see that the provisions in each case will make this properly negative, zero, or positive. The theorem allows us to generate any number of isotone or antitone fractions. For the case where $f(a) = f(b)$ when $a < b$ we have introduced the name "equitone" in order to have a name for this situation.

Let us take some examples which flow from Theorem 3. Let $a = 1$, $m = 5$, and $\epsilon = 1$. For an isotone property we have by (3) $0 < \delta < \min(5, 3) = 3$. So any positive number in this interval will suffice. Thus with $\delta = 2$ we get the example

$$\frac{1}{2} < \frac{5}{7} \quad \text{implies} \quad \frac{2}{3} < \frac{6}{8}\cdot$$

Or we might take $\delta = e$ so that we have

$$\frac{1}{2} < \frac{5}{5+e} \quad \text{implies} \quad \frac{2}{3} < \frac{6}{6+e}\cdot$$

With the same values of a and m we get the antitone property for any δ such that $3 < \delta < 5$. As an example take $\delta = 4$ and we have

$$\frac{1}{2} < \frac{5}{9} \quad \text{implies} \quad \frac{2}{3} > \frac{6}{10}\cdot$$

Finally in the same situation if we set $\delta = 3$ we get the equitone example stated just before Theorem 1:

$$\frac{1}{2} < \frac{5}{8} \quad \text{implies} \quad \frac{2}{3} = \frac{6}{9}\cdot$$

As a variation on Theorem 1 we next present

THEOREM 4. *Let* $0<a<b$ *and* $0<k$. *Then*

$$\frac{a}{b} < \frac{ka}{kb-1} \quad \textit{provided} \quad kb > 1;$$

$$\begin{cases} \dfrac{a+1}{b+1} < \dfrac{ka+1}{kb} & \textit{provided} \quad k < \dfrac{b+1}{b-a} \\ \dfrac{a+1}{b+1} = \dfrac{ka+1}{kb} & \textit{provided} \quad k = \dfrac{b+1}{b-a} \\ \dfrac{a+1}{b+1} > \dfrac{ka+1}{kb} & \textit{provided} \quad k > \dfrac{b+1}{b-a} \cdot \end{cases}$$

Proof. We have

$$\frac{a}{b} = \frac{ka}{kb} < \frac{ka}{kb-1} \qquad \text{for } kb > 1$$

so the first part is trivial. In the remaining three cases it is easily seen that the difference

$$\frac{ka+1}{kb} - \frac{a+1}{b+1} = \frac{b+1-k(b-a)}{kb(b+1)}$$

will be positive, zero, or negative according to whether the same is true for $b+1-k(b-a)$. This theorem also allows us to make up any number of simple isotone fractions.

As an example, let us take $a=4$ and $b=7$. For the isotone case we must choose $0<k<8/3$. With $k=2$ we have

$$\frac{4}{7} < \frac{8}{13} \quad \text{implies} \quad \frac{5}{8} < \frac{9}{14} \cdot$$

With $k>8/3$ we get the antitone case. Thus $k=3$ gives

$$\frac{4}{7} < \frac{12}{20} \quad \text{implies} \quad \frac{5}{8} > \frac{13}{21} \cdot$$

Let us turn now to the isotone situation

$$\frac{a}{b} < \frac{m}{n} \quad \text{implies} \quad \frac{a+1}{b+1} < \frac{m+1}{n+1} \cdot$$

If this is so, does it then follow that

$$\frac{a+k}{b+k} < \frac{m+k}{n+k}$$

and, if so, for what range of values of k? We offer the following theorem.

THEOREM 5. *Let* $0<a<b$, $0<m<n$, $t>0$. *If*

$$\frac{a}{b}<\frac{m}{n} \quad \text{and} \quad \frac{a+t}{b+t}<\frac{m+t}{n+t}$$

then

$$\frac{a+k}{b+k}<\frac{m+k}{n+k}$$

for any $k>0$, *provided that also* $k<t$.

Proof. Since $a/b<m/n$ we have $bm-an>0$. Now

$$\frac{m+t}{n+t}-\frac{a+t}{b+t}=\frac{(bm-an)+t[(b-a)-(n-m)]}{(b+t)(n+t)}>0$$

implies

$$bm-an>t[(n-m)-(b-a)]$$

and hence this would also hold for any value k such that $0<k<t$, which would be sufficient to give the result we want. Thus one could have a string of isotone fractions up to a certain step. If we supposed that $(b-a)-(n-m)\geq 0$ then the isotone case would hold for any positive k at all.

There is nothing really new or mysterious about the theorems we have given here. But they do seem to shed a little light on a subject which comes up every so often. The writer is indebted to E. E. Posey and C. H. Vehse who pointed out a passage in a well-known advanced calculus textbook where a published solution to a problem in infinite series seemed to imply that the unconscious assumption was being made that the function f defined by (1) above always preserves order. This motivates the idea of isotone and antitone fractions.

AN OPPENHEIM INEQUALITY

CHARLES W. TRIGG, Los Angeles City College

For P a point in a triangle whose distances are x, y, z from the vertices and p, q, r from the sides, Oppenheim [1] has proved that $xyz\geqq 8pqr$. If the joins of the vertices and P are extended to meet the sides, and the lengths of the segments between P and the sides are u, v, w, it has been shown [2] that $(x/u)(y/v)(z/w)\geqq 8$. Clearly, $uvw\geqq pqr$, so

$$xyz \geqq 8uvw \geqq 8pqr,$$

a stronger inequality and a proof of that of Oppenheim, with equality when the triangle is equilateral.

References

1. A. Oppenheim, The Erdos inequality and other inequalities for a triangle, The American Mathematical Monthly, Vol. 68, 1961, pp. 226–230.

2. O. J. Ramler and C. W. Trigg, E 1043, Property of three concurrent Cevians, The American Mathematical Monthly, Vol. 60, 1953, p. 421.

FORMULAS FOR A CURVED ROAD INTERSECTION

T. F. HICKERSON, University of North Carolina

It is required (see Fig. 1) to fit an access road extending from a known point A on "Curve A" to an unknown point C on "Curve B." Mathematically speaking, it is required to find the radius R of "Curve C" and the central angle Δ of the intercepted arc AC. Only the center-lines of the curved roadways in Fig. 1 are indicated, where the known data are: Points A and B of "Curves A and B," the corresponding radii R_1 and R_2, the angle Δ_1 and the tangents $T_1[T_1 = R_1 \tan \frac{1}{2}\Delta_1]$.

Assuming OA to be "North," the "Method of Coordinates" will be applied to the closed traverse $OAVO_2O$. The azimuths (bearings) will be reckoned clockwise from North. Thus, if L is *length* and ϕ is *azimuth*, *Departure* $= L \sin \phi$ and *Latitude* $= L \cos \phi$, where the same rules hold true for algebraic signs as in conventional trigonometry, provided the "initial line" is *North*.

Line	*Length*	*Azimuth*	*Latitude*	*Departure*
OA	R	0	R	0
AV	T_1	90	0	T_1
VO_2	(T_1+R_2)	$(90+\Delta_1)$	$-(T_1+R_2)\sin\Delta_1$	$(T_1+R_2)\cos\Delta_1$
O_2O	(R_2+R)	$(180+\Delta)$	$-(R_2+R)\cos\Delta$	$-(R_2+R)\sin\Delta$
			0	0

$$\Sigma \text{ Latitudes} = 0, \qquad \text{gives: } \cos\Delta = \frac{R - (T_1 + R_2)\sin\Delta_1}{R_2 + R} \tag{1}$$

$$\Sigma \text{ Departures} = 0, \qquad \text{gives: } \sin\Delta = \frac{T_1 + (T_1 + R_2)\cos\Delta_1}{R_2 + R}\cdot \tag{2}$$

The *cosine* formula, (1) is preferable to the *sine*, since the algebraic sign indicates whether to select Δ or $(180-\Delta)$.

Combining (1) and (2) gives

$$R = \frac{T_1 + (T_1 + R_2)\cos\Delta_1}{\sin\Delta} - R_2. \tag{3}$$

But (1) and (3) are not independent, hence the unknowns Δ and R cannot be found. Another (independent) expression for R will now be derived by a procedure involving analytic geometry. Thus, in Fig. 1,

$$AO = OC = OO_2 - CO_2; \text{ that is, } AO + CO_2 = OO_2; \text{ or}$$

$$(R + R_2) = OO_2; \qquad (R + R_2)^2 = (OO_2)^2. \tag{4}$$

Assuming AO to be the X-axis (origin at A), the coordinates (x_1, y_1) and (x_2, y_2) for points O and O_2 become: $[x_1 = R, y_1 = 0]$; and $[x_2 = (T_1+R_2)\sin\Delta_1, y_2 = T_1 + (T_1+R_2)\cos\Delta_1]$.

Substituting $(OO_2)^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2$ in (4),

$$(R + R_2)^2 = [(T_1 + R_2)\sin\Delta_1 - R]^2 + [T_1 + (T_1 + R_2)\cos\Delta_1 - O]^2. \tag{5}$$

Reducing, noting that $(T_1+R_2)^2(\sin^2\Delta_1 + \cos^2\Delta_1) = (T_1+R_2)^2$, factoring again, and solving, (5) becomes

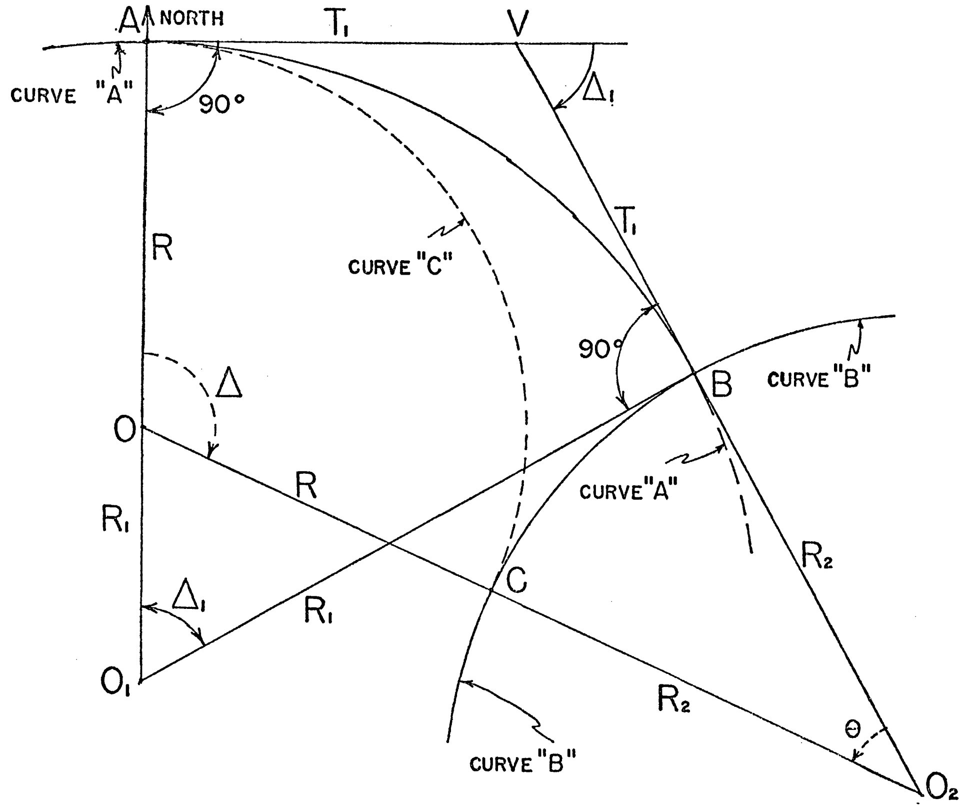


FIG. 1

$$R = \frac{(T_1)(T_1 + R_2)(1 + \cos \Delta_1)}{R_2 + (T_1 + R_2)(\sin \Delta_1)}\,. \tag{6}$$

By (6) and (1), the unknowns R and Δ may be found, while (3) serves merely as a check. Then θ (Fig. 1) $= 90 - (\Delta - \Delta_1)$. Knowing Δ, θ, and R, the arcs BC and AC may be calculated.

Example. Given $R_1 = 400$ ft, $R_2 = 300$ ft, $\Delta_1 = 60°$; to find R, Δ, θ, and the arcs BC and AC.

Solution. $T_1 = R_1 \tan \frac{1}{2}\Delta_1 = 230.94012$; by (6), $R = 242.0653$ ft; by (1), $\cos \Delta = -0.4016910$, then $(180 - \Delta) = 66°18'57.8''$, $\Delta = 113°41'12.2''$; $\theta = 90 - (\Delta - \Delta_1) = 36°18'47.8''$;

$$\text{arc } BC = \frac{R_2\theta}{57.29578} = \frac{(300)(36.3132778)}{57.29578} = 190.136 \text{ ft;}$$

$$\text{chord } BC = 2R_2 \sin \tfrac{1}{2}\theta = 2(300)(0.3116159) = 186.97 \text{ ft;}$$

$$\text{arc } AC = \frac{R\Delta}{57.29578} = \frac{(242.0653)(113.6867222)}{57.29578} = 480.308 \text{ ft;}$$

$$\text{chord } AC = 2R \sin \tfrac{1}{2}\Delta = 2(242.0653)(0.8371784) = 405.302 \text{ ft}$$

ON CERTAIN POLYNOMIALS ASSOCIATED WITH THE TRIANGLE

W. J. BLUNDON, Memorial University of Newfoundland

Many well-known properties of the triangle are expressed in terms of R, r, s, where these symbols represent respectively the circumradius, the inradius and the semiperimeter of the triangle. The aim of this paper is to set up identities from which necessary and sufficient conditions for certain properties of the triangle can be deduced immediately and simultaneously by inspection.

The central idea is to set up a correspondence between the truth of a certain property of the triangle and the vanishing of an appropriate polynomial. From these polynomials there is constructed a polynomial symmetrical in a, b, c, where these symbols represent the length of the sides of the triangle. The polynomial so constructed is now expressed as a polynomial function in R, r, s. It then remains only to examine the sign of the various factors of this function to establish the truth of the appropriate theorem.

The first step is to find expressions for the three elementary symmetrical polynomials in a, b, c. The fact that $a+b+c=2s$ follows immediately from the definition of s. Also, $abc=4R\Delta=4Rrs$, where Δ is the area of the triangle. To find $bc+ca+ab$, we note that

$$\begin{aligned} r^2s &= \Delta^2/s \\ &= (s-a)(s-b)(s-c) \\ &= s^3 - s^2\sum a + s\sum ab - abc \\ &= -s^3 + s\sum ab - 4Rrs, \end{aligned}$$

whence $\sum ab = s^2+4Rr+r^2$. To summarize

$$\begin{aligned} a+b+c &= 2s \\ bc+ca+ab &= s^2+4Rr+r^2 \\ abc &= 4Rrs. \end{aligned}$$

It is convenient to set down a few similar relations for later reference. Since $\sum a^2 = (\sum a)^2 - 2\sum ab$, $\sum a^2b = (\sum a)(\sum ab) - 3abc$, $\sum a^3 = (\sum a)^3 - 3(\sum a)(\sum ab)+3abc$, $16\Delta^2 = -\sum a^4 + 2\sum a^2b^2$, we have

$$\begin{aligned} \sum a^2 &= 2s^2 - 2r^2 - 8Rr \\ \sum (a^3 - a^2b) &= -8rs(R+r) \\ -\sum a^4 + 2\sum a^2b^2 &= 16r^2s^2. \end{aligned}$$

Consider first the problem of finding necessary and sufficient conditions on R, r, s that a triangle be right-angled. The angle A is a right angle if and only if $b^2+c^2-a^2=0$, with similar conditions for the angles B and C. Thus ABC is a right triangle if and only if $(b^2+c^2-a^2)(c^2+a^2-b^2)(a^2+b^2-c^2)=0$. Now the left side of this equation is equal to

$$-\sum a^6 + \sum a^4b^2 - 2a^2b^2c^2 = (\sum a^2)(-\sum a^4 + 2\sum a^2b^2) - 8a^2b^2c^2$$

$$= (2s^2 - 2r^2 - 8Rr)(16r^2s^2) - 128R^2r^2s^2$$
$$= 32r^2s^2(s + r + 2R)(s - r - 2R).$$

Every factor of this expression is necessarily positive with the exception of the last factor. Hence it follows that a necessary and sufficient condition for a triangle to be right-angled is that $s=r+2R$. Corresponding conditions for a triangle to be acute-angled or obtuse-angled are respectively $s>r+2R$, $s<r+2R$.

As a second example, a necessary and sufficient condition that a triangle be equilateral is that the polynomial $(b-c)^2+(c-a)^2+(a-b)^2$ should vanish. This holds if and only if $s^2=3r^2+12Rr$, which is not a particularly interesting relation. However, we may replace this polynomial by any polynomial of the form $k_1(b-c)^2+k_2(c-a)^2+k_3(a-b)^2$, provided the k_i are positive and the resulting polynomial is symmetrical in a, b, c. Consider, for example, the polynomial

$$\begin{aligned}\sum (a + b - c)(a - b)^2 &= (a + b + c)\sum (a - b)^2 - 2\sum c(a - b)^2\\ &= (\sum a)(2\sum a^2 - 2\sum ab) - 2(\sum a^2b - 6abc)\\ &= 2(\sum a^3 - \sum a^2b) + 6abc\\ &= - 16rs(R + r) + 24Rrs\\ &= 8rs(R - 2r).\end{aligned}$$

Now the left side of this identity vanishes if and only if $a=b=c$, and the right side vanishes if and only if $R=2r$. This proved the familiar necessary and sufficient condition that a triangle be equilateral. Further, since the left side is never negative, we have for every triangle the relation $R\geqq 2r$.

We mention a few more examples. A necessary and sufficient condition that a triangle be isosceles is that the polynomial $(b-c)^2(c-a)^2(a-b)^2$ should vanish, and this polynomial is the discriminant of the monic cubic equation which has a, b, c as its roots. Consideration of the polynomial $(bc-a^2)(ca-b^2)(ab-c^2)$ and the polynomial $(b+c-2a)(c+a-2b)(a+b-2c)$ will give necessary and sufficient conditions that the sides of a triangle should be respectively in geometrical and arithmetical progression.

The reader is invited (1) to construct other symmetrical polynomials in a, b, c and thereby discover other properties of the triangle involving R, r, s; (2) to express in terms of R, r, s properties of the triangle readily expressible in terms of the ex-radii r_1, r_2, r_3, the preceding methods being applicable since it is easily verified that

$$\sum r_1 = 4R + r, \quad \sum r_1r_2 = s^2, \quad r_1r_2r_3 = rs^2.$$

References

1. H. S. M. Coxeter, *Introduction to Geometry*, Wiley, 1961.
2. E. W. Hobson, *A Treatise on Plane Trigonometry*, Cambridge University Press, 1925.
3. R. A. Johnson, *Modern Geometry*, Houghton Mifflin, 1929.

EDITED BY ROTHWELL STEPHENS, Knox College

This department is devoted to the teaching of mathematics. Thus, articles of methodology, exposition, curriculum, tests and measurements, and any other topic related to teaching, are invited. Papers on any subject in which you, as a teacher, are interested, or questions which you would like others to discuss, should be sent to Rothwell Stephens, Mathematics Department, Knox College, Galesburg, Illinois.

A LESSON IN GRAPHING

F. MAX STEIN, Colorado State University

The place was my office in the mathematics building. I answered a knock at my door.

"Good morning, Sir or Madam, as the case may be." (It was obvious that he had learned his salesmanship according to the book.) "I am Mr. Cross Hatch, your friendly graph paper salesman. I sell graph paper."

"No, thanks," said I. "I already have a sheet of graph paper."

"But, just a minute," said Cross, putting his foot in the door, "I am well stocked with a variety seldom found elsewhere."

"What kinds do you have?" I asked, playing the part of the perfect straight man.

"I have the usual rectangular and polar coordinate paper, with which everyone is acquainted, but I also have log and log-log." (He appeared to stutter.) "I have power paper of all orders, exponential paper, sine paper, and cosine paper as well as sine-sine and cosine-cosine paper." (He started his stuttering again.)

"What can I do with these various kinds of paper that I can't do with the sheet I already have?"

"Oh, this paper isn't for you," said C. H., now appealing to my vanity. "This paper is for your students. It will make graphing much simpler for them, albeit perhaps somewhat monotonous. Your students will have to know only how to draw a straight line. Also, think how much simpler your work will be when you check graphs."

Now C. H. was hitting home to me, since I appreciated lack of work with the best. I could see how $y=x$ or $y=2x$, or even $y=mx$, could be graphed as a straight line, but how could the graph of the parabola $y=x^2$, for instance, be represented by a straight line? The answer to my unasked question was not long in forthcoming.

"For $y=x$," said Mr. Hatch, "you have a line through the origin inclined at 45° to the x-axis on your graph paper. Here the intervals along each axis are taken to be equal, Fig. 1a. Suppose you let each interval along the x-axis be taken as two units but keep the units the same on the y-axis. The graph of $y=x$ is still a line through the origin, but it is now inclined at 63°26′ to the x-axis, Fig. 1b. I think that even you can see that $y=2x$, or $y=mx$, m finite, can be represented by a 45° line on the proper paper if you turn the argument around. You see that the problem is not one so much of graphing now as in choosing the proper graph paper."

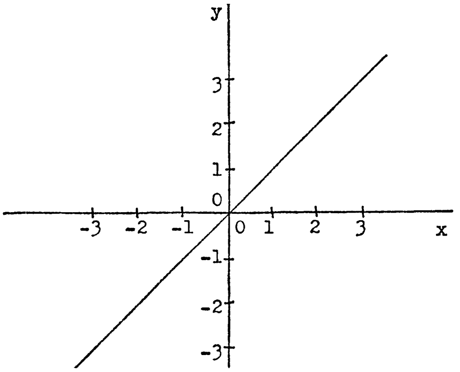


FIG. 1a

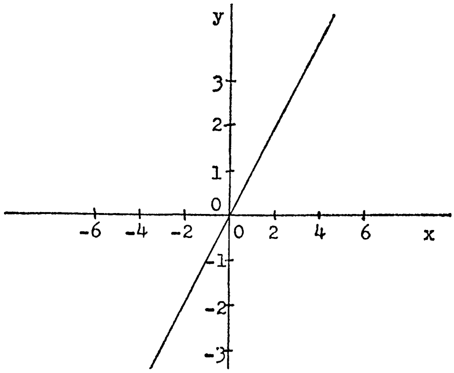


FIG. 1b

Without giving me time for a retort to his obvious dig at my mentality, he hurried on.

"Now to graph $y=x^2$, say, the student only has to pick the proper sheet of graph paper, draw a 45° line, and hand in his work, Fig. 2a. The paper he would

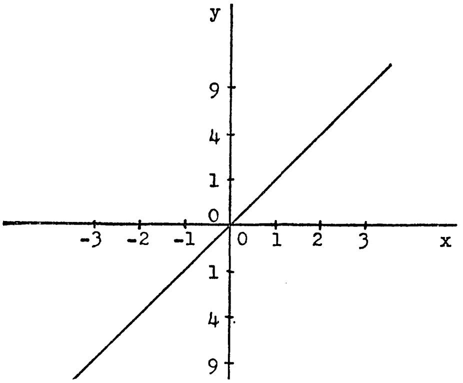


FIG. 2a

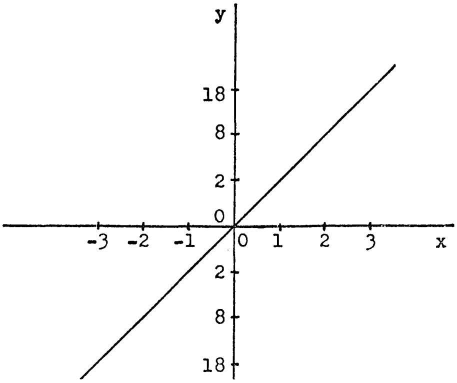


FIG. 2b

have to choose would have the intervals marked off as 1, 4, 9, $\cdots$ on the positive y-axis. Since all of our paper is square, the student would only need to turn his paper through 90° and make a few obvious changes and he could graph $y=\sqrt{x}$. I'm assuming your students are intelligent enough to handle the graph in other quadrants correctly in each case."

"Ah, ha!" thought I. "I can cut down on the number of kinds of graph paper my students will have to stock. All I need to do is to have them mark the intervals as 1^n, 2^n, 3^n, $\cdots$ along the positive y-axis for any positive integer n, and then they can graph $y=x^n$ on their usual paper as a straight line inclined at 45° to the positive x-axis."

"We also carry a little more expensive line of paper—multipurpose paper, we call it—so that students won't need to stock so many different kinds," said C. H., not bothering to comment on my unspoken thoughts. "This paper has a semi-elastic base so that it can be stretched without returning to its original size or it can be shrunk by dipping it in water for a short time. With this paper students can graph $y=2x^2$, or $y=mx^2$, in the same manner as they would graph $y=x^2$ by merely shrinking the paper and then halving the length of each subinterval along the y-axis, Fig. 2b. The intervals along the y-axis will then read 2, 8, 18, $\cdots$."

"Note," he continued, "that the paper for $y=x^2$, say, has two positive y-axes but no negative ones and that they both increase without bound both above and below $y=0$. However, our exponential paper, for graphing $y=e^x$, increases without bound above $y=1$ (no $y=0$ in this case), but it approaches zero as a limit below the x-axis, Fig. 3a. For our log paper this same phenomena is observed only with respect to the x-axis, Fig. 3b."

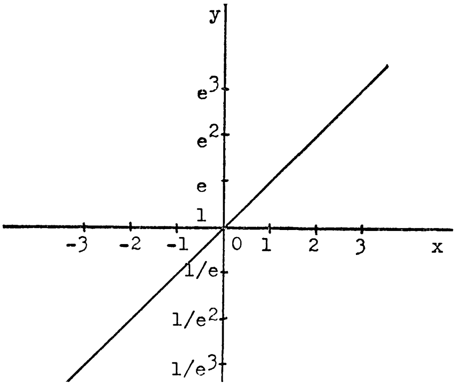


FIG. 3a

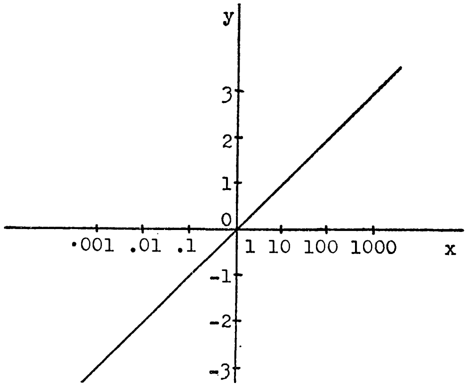


FIG. 3b

"This is a curious situation," I observed. "For all of your previous samples you had equal units marked off at equal subintervals on the x-axis, but now your equal units are on the y-axis. Elucidate," said I, displaying a multisyllable word.

"Actually," said C. H., "it is a matter of choice. My company has decided that for functions which are concave upward the equal units shall be on the x-axis and vice versa for functions concave downward."

"What advantage does this log-log paper have?" I asked after catching a glimpse of his next sample, Fig. 4. "It seems to me that, for $\log y=\log x$ and as far as we are concerned, we have $y=x$ and thus have the same units along

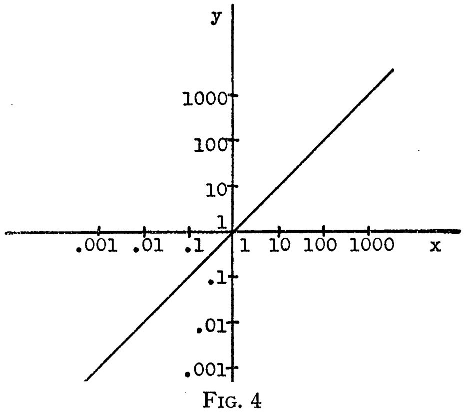


FIG. 4

each axis as on my paper."

Patiently C. H. led me along. "Suppose you have a set of points on the same graph with both coordinates near 10, another set with coordinates near 100, and a final set with coordinates near 1000, and you wish to plot these all on the same graph. Could you tell much from your graph?"

I was beginning to see the light. To get all of these points on the same graph I would need to crowd the first set to such an extent that they would be meaningless. I started to fill out an order blank.

"Notice how points whose coordinates are very small positive values can also be put on this graph, and the entire gamut will have meaning. Also observe that the axes cross at the point (1, 1) and not at (0, 0)." Hatch rambled on and on not appearing to notice that I was ripe for closing the sale.

Eventually, after the order for a generous supply of his entire stock had been signed and I was alone once more, I began to wonder—"Which kind of paper is best for graphing $y=\sqrt{25-x^2}$? Can I put 720° or only 180° on my polar graph paper and simplify my graphing in polar coordinates?" These questions led to others which in turn spawned others. I consoled myself; at least, I can now check a great portion of the homework papers involving graphing in much less time hereafter.

BOURBAKI. A FRENCH GENERAL—OR A MYSTERIOUS SOCIETY?

DAGMAR RENATE HENNEY, University of Maryland

"A revolution is going on in the field of mathematics and its leader is named Bourbaki."

An audience of more than 300, among them 75 midshipmen of the Naval Academy, were told that the usual subject matter studied in their Freshman classes, was totally obsolete. The speaker, calling trigonometry and analytic geometry subjects which are studied in a silly way, is recognized as one of the world's leading mathematicians, Professor Jean Dieudonné. All of trigonometry, according to the French mathematician, could be condensed into the single formula $e^{i(x+y)}=e^{ix}e^{iy}$. Dr. Dieudonné's lecture was the highpoint of a series of lectures presented by the Mathematics Honorary of the University of Maryland. The title of the lecture "Bourbaki" awoke the curiosity of many scientists and laymen who had come from as far as Baltimore and New York, to listen to the distinguished speaker. Almost everyone had heard the name of Bourbaki before and was aware of the mystery and numerous legends surrounding it.

The strangest fact about this famous mathematician is that he does not exist. A group of young French mathematicians has written numerous comprehensive volumes on mathematics using the collective pseudonym of Bourbaki. Nobody is quite certain why the name was chosen, maybe it was done in jest, or maybe to avoid writing all names of the authors that are participating on each article published by Bourbaki.

When the French army was disastrously defeated in 1871, little did its commander General Bourbaki realize that some day his name might be famous and talked about by every mathematician in the world. General Bourbaki's life had been a collection of whimsical mishaps. Though at one time well thought of (he was offered a chance to become king of Greece, but declined—because of his mistress' opposition)—his luck soon changed. He was imprisoned in Switzerland after being defeated in battle and tried to shoot himself. Apparently he did not succeed in this event either, since it is reported that at the ripe age of 80 years he challenged another retired officer to a duel.

The society, that now proudly bears his name was first formed in the early 1930's in Nancy, France. It is said that at this time students at the École Supérieure were exposed to a lecture by a distinguished visitor. This visitor, called Bourbaki, was an actor in fact and his lectures were a perfect piece of double-talk.

Their major publication "The Elements of Mathematics" appeared at the same time. The "Elements" constitute a complete and coherent survey of modern mathematics from a sophisticated point of view. Bourbaki reorganized all of mathematics in another giant work entitled "The Fundamental Structures of Analysis," which has already surpassed 25 volumes. The subdivisions of the first part might shock the classical mathematician or general scientist who thinks of mathematics in terms of algebra, trigonometry and analytic geometry. The first generation (or first part) instead includes topology, topological vector spaces, integration, set theory, functions of a real variable, and higher algebra. Now in progress is the second part (or second generation) consisting of Lie algebra, and commutative algebra. The third generation of Bourbaki plans the treatment of algebraic geometry, differential topology, algebraic number theory and partial differential equations. If any layman was ever of the opinion that mathematics is a dead science, he should but take a glance at any of these volumes to be convinced that this is not so. There exists even a topic in mathematics, namely, algebraic topology, which Bourbaki will not yet discuss since it changes rapidly from year to year.

There are young mathematicians in remote areas of the world whose serious mathematical education was obtained from the volumes of Bourbaki. Every good reference library is certain to include their books and the reader of Bourbaki, whether he studies these in the original French or in any other language (there exists a Russian translation without the permission of the authors) chuckles at the helpful notations such as "dangerous curve" ahead which indicates a "slippery" turn in the proof.

How is a treatise of this magnitude written? Much energy goes into the writing of each single volume, which often takes 10 to 20 years to complete. The society meets usually three times during the year. One person is elected to write the first draft. This draft is read during the general session and discussed at great length. Following this, a second member of the organization writes a second draft and it is read during the following year with much of the same results, i.e. it is examined again and usually will have to be rewritten. After 10, 12, 14 years everyone involved finally comes to an agreement and the book is turned over to the editor. The final approval of a draft is celebrated over a glass of wine. Since the Bourbaki books were a commercial success, the members have sufficient funds to pay for travel expenses and provide food and drinks to lubricate their endeavors.

It is an open secret that Professor Dieudonné is the driving force of the group. Legend has it, that Dieudonné once published a paper under the name Bourbaki, which later on was found out to contain an error. The mistake was consequently corrected in a paper entitled "On an Error of Bourbaki" and signed Jean Dieudonné.

Several mathematicians have come to regret having cast aspersions on the existence of Bourbaki. The society does not like to have its secrets told publicly. In these matters Bourbaki will retaliate swiftly. At one time an American mathematician, Boas, was asked to furnish a biography of Bourbaki, which was to be incorporated in a mathematics biography. The American returned a letter explaining that the name Bourbaki was but a pseudonym and that in fact Bourbaki did not exist. As soon as the Bourbaki group found out about this, they sent a telegram to the editor, giving in detail birthdate, birthplace and different Universities attended by Bourbaki and concluded that in fact the American mathematician Boas was but a collective pseudonym for a group of young Americans and that in fact Boas did not exist.

REMARKS ON THE EXPRESSION $\sum_{i=1}^{N}\left[\left(\frac{x_i}{a_i}\right)^2\right]^{n_i}$ = CONSTANT

JACQUES ALLARD, University of Sherbrooke

Several well-known geometrical figures can be obtained from the general expression above. In particular when $a_i=1$, $N=2$, and $n_i=n$, this expression reduces to:

$$x^{2n} + y^{2n} = r^{2n} \tag{1}$$

r being a constant and n a parameter. When $n\to\infty$, the expression represents in cartesian coordinates the locus of a square of side $2r$ as a limiting case. When $n\geqq 5$, a pseudo-square of side $2r$ is obtained as shown by Figure 1. When $n=2$,

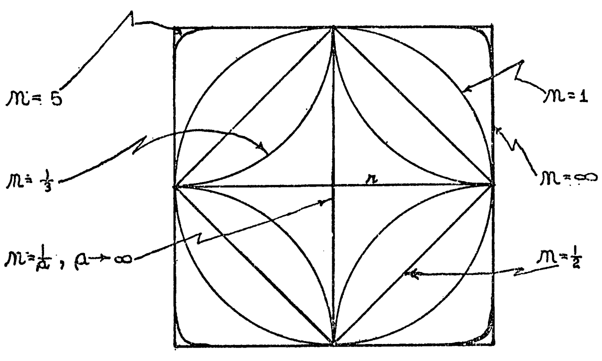


FIG. 1. Locus of $x^{2n}+y^{2n}=r^{2n}$ for $n=\infty$, 5, 1, $\frac{1}{2}$, $\frac{1}{3}$, 0.

the familiar quartic of Lamé is obtained. If $n=1$, (1) represents a circle of diameter $2r$. Putting $n=\frac{1}{2}$, gives:

$$(x^2)^{1/2} + (y^2)^{1/2} = (r^2)^{1/2}. \tag{2}$$

Equation (2) is not equivalent to $x+y=r$ as far as loci are concerned because it represents a square of diagonal $2r$ which is inscribed in the circle $x^2+y^2=r^2$. When $n=1/3$, the hypocycloid of four cusps is obtained and finally when $n=1/a$, $a\to\infty$, a cross of length $2r$ is obtained which is the hypocycloid of four cusps with a zero area. Some properties of these equations can be studied with gamma functions and can lead to interesting physical applications.[1]

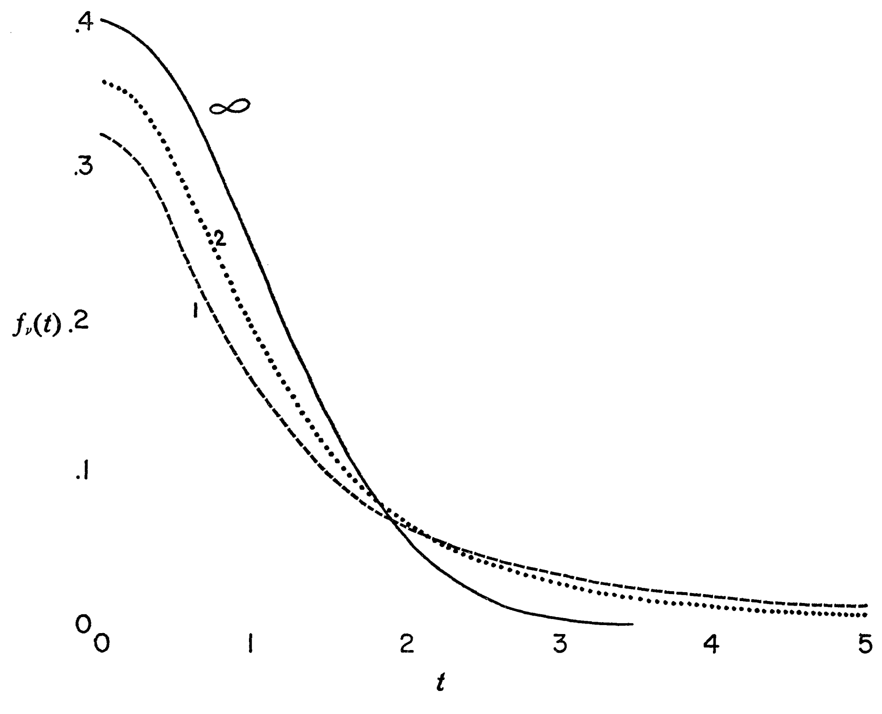
[1] J. Allard and J. Bazinet "Remarques Sur Quelques Figures Géométriques," Revue de l'Université de Sherbrooke, Vol. 2, No. 2, 1962.

ORDINATES FOR STUDENT'S DISTRIBUTION

JOHN M. HOWELL, Los Angeles City College

On rare occasions, one may wish to draw a picture of Student's Distribution. Many tables of the integral exist, but the table of ordinates is difficult to find or non-existent. This table was calculated on a computer using 12 digit accuracy and rounding off to four decimal places before type-out. The equation of Student's distribution is

$$f_\nu(t) = \frac{1}{\sqrt{\pi\nu}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2}.$$



ORDINATES FOR STUDENT'S DISTRIBUTION

$\nu\backslash t$	0	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
1	.3183	.2546	.1592	.0979	.0637	.0439	.0318	.0240	.0187	.0150	.0122
2	.3536	.2963	.1925	.1141	.0680	.0422	.0274	.0186	.0131	.0095	.0071
3	.3676	.3132	.2067	.1200	.0675	.0387	.0230	.0142	.0092	.0061	.0042
4	.3750	.3223	.2147	.1229	.0663	.0357	.0197	.0113	.0067	.0041	.0026
5	.3796	.3279	.2197	.1245	.0651	.0333	.0173	.0092	.0051	.0029	.0018
6	.3827	.3318	.2231	.1256	.0640	.0315	.0155	.0078	.0041	.0022	.0012
7	.3850	.3346	.2257	.1263	.0631	.0300	.0141	.0067	.0033	.0017	.0009
8	.3867	.3367	.2276	.1268	.0624	.0288	.0130	.0059	.0028	.0013	.0007
9	.3880	.3384	.2291	.1272	.0617	.0278	.0121	.0053	.0023	.0011	.0005
10	.3891	.3397	.2304	.1274	.0611	.0269	.0114	.0048	.0020	.0009	.0004
11	.3900	.3408	.2314	.1277	.0607	.0262	.0108	.0044	.0018	.0007	.0003
12	.3907	.3417	.2322	.1279	.0602	.0256	.0103	.0040	.0016	.0006	.0003
13	.3914	.3425	.2330	.1280	.0598	.0251	.0098	.0038	.0014	.0005	.0002
14	.3919	.3432	.2336	.1282	.0595	.0246	.0095	.0035	.0013	.0005	.0002
15	.3924	.3438	.2341	.1283	.0592	.0242	.0091	.0033	.0012	.0004	.0002
16	.3928	.3443	.2346	.1284	.0589	.0238	.0088	.0031	.0011	.0004	.0001
17	.3931	.3447	.2350	.1284	.0587	.0235	.0086	.0030	.0010	.0003	.0001
18	.3934	.3451	.2354	.1285	.0585	.0232	.0084	.0028	.0009	.0003	.0001
19	.3937	.3455	.2357	.1286	.0583	.0229	.0081	.0027	.0009	.0003	.0001
20	.3940	.3458	.2360	.1286	.0581	.0227	.0080	.0026	.0008	.0003	.0001
21	.3942	.3461	.2363	.1287	.0579	.0224	.0078	.0025	.0008	.0002	.0001
22	.3944	.3464	.2366	.1287	.0578	.0222	.0076	.0024	.0007	.0002	.0001
23	.3946	.3466	.2368	.1288	.0576	.0221	.0075	.0023	.0007	.0002	.0001
24	.3948	.3468	.2370	.1288	.0575	.0219	.0074	.0023	.0007	.0002	.0001
25	.3950	.3470	.2372	.1288	.0574	.0217	.0073	.0022	.0006	.0002	.0000
26	.3951	.3472	.2374	.1289	.0572	.0216	.0071	.0022	.0006	.0002	.0000
27	.3953	.3474	.2376	.1289	.0571	.0214	.0070	.0021	.0006	.0002	.0000
28	.3954	.3476	.2377	.1289	.0570	.0213	.0069	.0020	.0006	.0001	.0000
29	.3955	.3477	.2379	.1289	.0569	.0212	.0069	.0020	.0005	.0001	.0000
30	.3956	.3479	.2380	.1290	.0569	.0211	.0068	.0020	.0005	.0001	.0000
31	.3957	.3480	.2381	.1290	.0568	.0209	.0067	.0019	.0005	.0001	.0000
32	.3958	.3481	.2382	.1290	.0567	.0208	.0066	.0019	.0005	.0001	.0000
33	.3959	.3483	.2384	.1290	.0566	.0208	.0066	.0018	.0005	.0001	.0000
34	.3960	.3484	.2385	.1290	.0565	.0207	.0065	.0018	.0005	.0001	.0000
35	.3861	.3485	.2386	.1291	.0565	.0206	.0064	.0018	.0005	.0001	.0000
36	.3962	.3486	.2386	.1291	.0564	.0205	.0064	.0018	.0004	.0001	.0000
37	.3963	.3487	.2387	.1291	.0564	.0204	.0063	.0017	.0004	.0001	.0000
38	.3963	.3488	.2388	.1291	.0563	.0203	.0063	.0017	.0004	.0001	.0000
39	.3964	.3488	.2389	.1291	.0562	.0203	.0062	.0017	.0004	.0001	.0000
40	.3965	.3489	.2390	.1291	.0562	.0202	.0062	.0017	.0004	.0001	.0000
41	.3965	.3490	.2390	.1291	.0561	.0202	.0061	.0016	.0004	.0001	.0000
42	.3966	.3491	.2391	.1291	.0661	.0201	.0061	.0016	.0004	.0001	.0000
43	.3966	.3491	.2392	.1291	.0560	.0200	.0061	.0016	.0004	.0001	.0000
44	.3967	.3492	.2392	.1292	.0560	.0200	.0060	.0016	.0004	.0001	.0000
45	.3967	.3493	.2393	.1292	.0560	.0199	.0060	.0016	.0004	.0001	.0000
46	.3968	.3493	.2394	.1292	.0559	.0199	.0060	.0015	.0004	.0001	.0000
47	.3968	.3494	.2394	.1292	.0559	.0198	.0059	.0015	.0004	.0001	.0000
48	.3969	.3494	.2395	.1292	.0558	.0198	.0059	.0015	.0003	.0001	.0000
49	.3969	.3495	.2395	.1292	.0558	.0197	.0059	.0015	.0003	.0001	.0000
50	.3970	.3495	.2396	.1292	.0558	.0197	.0058	.0015	.0003	.0001	.0000
∞	.3989	.3521	.2420	.1295	.0540	.0175	.0044	.0009	.0001	.0000	.0000

COMMENTS ON PAPERS AND BOOKS

EDITED BY HOLBROOK M. MACNEILLE, Case Institute of Technology

This department will present comments on papers published in the MATHEMATICS MAGAZINE, lists of new books, and reviews.

In order that errors may be corrected, results extended, and interesting aspects further illuminated, comments on published papers in all departments are invited.

Communications intended for this department should be sent to Holbrook M. MacNeille, Department of Mathematics, Case Institute of Technology, Cleveland 6, Ohio.

A RELAXATION DIFFICULTY

WILLIAM R. RANSOM, Tufts University

In the relaxing method on page 295 of the November 1962 number of MATHEMATICS MAGAZINE, in step number 5, Δx is taken as 5, the exact value that reduces R_1 to zero being $16 \div 3$. A peculiar difficulty may arise from the use of approximate increments. For example, in solving:

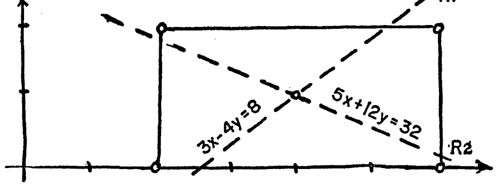
$$\left.\begin{aligned} 3x - 4y &= 8 \\ 5x - 12y &= 32 \end{aligned}\right\} \quad \text{if we take} \quad \left\{\begin{aligned} R_1 &= 8 - 3x + 4y \\ R_2 &= 32 - 5x - 12y \end{aligned}\right.$$

the first six steps will be

Step no.	x	y	R_1	R_2	exact	approximate
1	0	0	8	32	$32 \div 5$	$\Delta x = 6$
2	6	0	-10	2	$10 \div 4$	$\Delta y = 2$
3	6	2	-2	-22	$-22 \div 5$	$\Delta x = -4$
4	2	2	10	-2	$-10 \div 4$	$\Delta y = -2$
5	2	0	2	22	$22 \div 5$	$\Delta x = 4$
6	6	0	-10	2		

Now, every four steps there is a periodic repetition. Graphically this means that the points (x, y) are successive corners of a rectangle with the coordinates of its center as the solution. The true values of x and y are the means of their values at the corners: $x = (6+6+2+2) \div 4 = 4$, and $y = (2+2+0+0) \div 4 = 2$.

This same phenomenon occurs when the two lines have slopes which are negatives of each other.



COMMENTS ON A "TEACHING NOTE"

CHARLES T. SALKIND, Polytechnic Institute of Brooklyn

Herewith two comments on "A Teaching Note" by George Tyson, page 105 of the March–April 1963 issue of MATHEMATICS MAGAZINE: Comment 1. The solution to the equation $6t-1+t^2=1+t^2$ is given as $6t=2$. Since, however, t represents the tangent of an angle, it is proper, here, to include the infinite values of t as part of the solution. With this inclusion the substitution $\tan \frac{1}{2}x=t$ yields a complete solution of the given equation $3 \sin x-\cos x=1$.

Comment 2. If the procedure in Comment 1 is too risky for college freshmen, then the equally "standard" solution in terms of a phase angle may be used, to wit, $\sqrt{10}(3/\sqrt{10} \sin x - 1/\sqrt{10} \cos x) = 1$, $\sin (x-a) = 1/\sqrt{10}$, $x-a = \sin^{-1} 1/\sqrt{10}+2n\pi$, or $x-a=\pi-\sin^{-1} 1/\sqrt{10}+2n\pi$. The first of these values corresponds to the value $\tan \frac{1}{2}x=\frac{1}{3}$ and the second corresponds to $x=(2n+1)\pi$.

BOOK REVIEWS

"*Modern Mathematics: A Programed Textbook*," Developmental Edition. By Lewis D. Eigen, Jerome D. Kaplan, Ruth Emerson, and Harold M. Krouse, Science Research Associates, Chicago, 1961, 10 volumes, chapter tests, final examination and teacher's manual, \$13.75 for a complete set.

The content of *Modern Mathematics* reflects the modern approach to high school mathematics. Emphasis is placed on the ideas basic in understanding mathematics yet manipulative skills are not neglected. Although deduction is emphasized the rigor is not overwhelming. Compared with traditional approaches this one is most refreshing.

This is a vertical program text as each page is divided into two vertical columns. The right-hand column consists of statements to be completed or questions to be answered by the student. The left-hand column contains the answers and is supposed to be covered by a movable mask. The student writes his answer in the Student Answer Booklet and then slides the mask down to expose the correct answer in the text. After finding out whether he is right or wrong, he continues to the next item.

The assumption is made that the student can read at the seventh-grade level and is competent in arithmetic skills, including operations with common and decimal fractions.

It is claimed that the program has been used successfully as homework, as classwork, and for tutoring. As homework it has been used to reinforce material learned in class by assigning a related sequence, and also to introduce the new topic to be considered. As classwork, each student moves at his own pace with the teacher meeting small groups or individuals for discussion.

Among the topics that would normally be considered modern for this level are the following: numbers and numerals, absolute value, real numbers, sets and subsets, intersection of sets, ordered pairs and relations, functions, inequalities, functions and their converses, complements of sets, mappings, some order properties, binary operations, closure, identity elements, inverses, commutativity, associativity, distributivity, pronumerals, solution sets, and absolute value equations. All the traditional topics of graphing, solving equations, and word problems are included.

The quality of mathematics is extremely good for this level and when compared with traditional mathematics, it compares favorably with the better texts appearing on the market today. The programing, for the most part, is also good. The approach usually enables the student to become functional first with easy problems before considering definitions and theory. The good practice of beginning with easy and known material and leading gently into the new, is used. As for all programed texts, this permits each student to move at his own rate. The material may be rated as fairly interesting but not uniformly interesting.

Three conventions adopted by the writers might create some difficulties. There is an identification of identity with right identity and the definition $0^0=1$. Book 9, pg 71, frame 85. The convention of identifying fraction with rational number runs counter to many of the modern treatments.

The errors are very few.

Book 1:	pg 2 frame 7	—"When your teacher writes a number on the blackboard." (Should be numeral)
	pg 11 fr 50	—+3 instead of '+3'.
Book 2:	pg 69 fr 86	—The answer could be any superset of $\overline{A}$.
	pg 70 fr 87	—The answer here could be any subset of $\overline{A}$.
Book 4:	pg 18 fr 15	—Answer, "the set of real numbers."
	pg 26 fr 34	—D_* (subscript omitted). $R_* \subset N$ (and not $\in$).
	pg 28 fr 64	—Answer should be "usually" true.
	pg 40 fr 45	—Subscript on "R" omitted. Should be "R_+". Frames 45 and 46 should be interchanged.
Book 5:	pg 12 fr 35	—The solution to this equation is 1/2. Delete "set."
Book 8:	pg 75 fr 108	—One may quarrel with the implication that if the equation cannot be transformed into the form $ax+by+c=0$ then its graph is not a straight line. Note: $x^2-2xy+y^2=0$ and $x^2=0$.
Book 9:	pg 63 fr 15	—The answer is omitted. Should be "$-$."
	pg 101 fr 179	—Last expression should be "$2x^4+7$."

The type of programming may be classified as linear. With very few exceptions each frame must be taken in order. In about six places there is branching where the student is directed to proceed to a special frame depending upon the response he gave.

HARRY D. RUDERMAN
Hunter College High School

Inequalities. By P. P. Korovkin, Blaisdell, New York, 1961, v+60 pages, 95¢ (paper).

The Method of Mathematical Induction. By I. S. Sominskii, Blaisdell, New York, 1961, v+56 pages, 95¢ (paper).

Fibonacci Numbers. By N. N. Vorob'ev, Blaisdell, New York, 1961, v+66 pages, 95¢ (paper).

The lucid style characteristic of the Russian series *Popular Lectures in Mathematics* has survived the translation into English. These three little volumes, written a decade ago, deserve a place in the library of a beginning mathematics student (and his teacher).

Inequalities is a collection of 62 problems (with solutions)—a suitable companion to the recent books by Beckenbach, Bellman and Kazarinoff. Although oriented toward the classical inequalities usually encountered in calculus, the second problem, e.g., is to find the integral part of the sum of the reciprocals of the square roots of the first million natural numbers.

Mathematical Induction is a leisurely exposition which includes discussion of an intriguing conjecture (proposed in 1938 and solved in 1941) concerning the factorization of x^n-1, an ingenious proof of the AM-GM inequality, and such provocative problems as: $n+1$ numbers are picked at random from the $2n$ integers $1, 2, \cdots, 2n$. Prove that among the numbers picked we can find at least two, one of which is divisible by the other.

According to the author *Fibonacci Numbers*, "This book is designed to appeal basically to pupils 16 or 17 years of age in high school." (It is pointed out that the reader may omit sections involving trigonometry and the limit concept.) At first glance it might appear a bit too ambitious to expect even mature high school students to follow the development and application of the Euclidean algorithm and continued fractions. However, this reviewer feels that many of the number-theoretic topics discussed are more suitable for the development of mathematical sophistication than, say, the premature or misplaced rigor and abstraction which characterize so many of the "modern" texts.

Errors and misprints are few and minor. Occasionally the translation is unnecessarily literal (a whole n, a compound n, chain fraction, sine rule); but these books are ideal for enrichment and self-study. One only needs a magnifying glass to identify the microscopic exponents encountered in the first two volumes.

DMITRI THORO
San Jose State College

The Fibonacci Quarterly, The Official Journal of the Fibonacci Association, Vol. 1, No. 1, February, 1963.

That a quarterly journal could emerge devoted entirely to Fibonacci numbers and related developments may come as somewhat of a surprise to both mathematician and layman. That the venture shows promise of challenging and fascinating both the mathematician and the layman may come as an even greater surprise.

The Journal consists of two parts. Part I, for the more sophisticated mathematician, provides an outlet for the technical papers of researchers, including, in the initial issue, notions such as expansion of analytic functions, operational recurrences, semi-completeness, and primitive factors. Part II is for the beginner. Most of this material is well within the range of the competent high school student. Both Parts I and II include appropriate "Problems and Solutions" sections.

The editors state in the preface that "the journal is for active readers; the editors desire reader participation especially from mathematics teachers and students." The style, format, and content of the initial issue indicates that the editors have found a highly palatable combination of recreational and significant mathematics that should be of value to both teacher and student.

Those interested should inquire of Brother U. Alfred, St. Mary's College, California.

Ben K. Gold
Los Angeles City College

"*The Summation of Series.*" By Harold T. Davis, Principia Press of Trinity University, San Antonio, 1962, ix+140 pages, hard cover, $6.00.

The author has written this book to advance the reader's understanding of the problem of the summation of series in the real domain. The choice of subject matter and the level of presentation make this book a useful source of information for the student who has mastered the conventional treatment of the convergence and divergence of series, Taylor's Formula, Taylor's Series, and related topics usually included in a course in intermediate analysis.

In Chapter I the methods of the calculus of finite differences are developed, with due regard to analogous developments in the differential and integral calculus. The formulas for first and higher order differences are derived, a table of differences for particular functions is prepared, and the Gregory-Newton interpolation formula is constructed. The summation problem is defined and a table of sums is prepared for a collection of special functions. The Euler-Maclaurin Formula is applied to problems in numerical integration.

The Gamma and Psi functions are defined in Chapter II and their properties determined. Application of these properties is made to the problem of summing reciprocal polynomial functions and evaluating the sums of related series of constants. Chapter III contains the development of a number of summation methods, including Gregory's difference method, Lubbock's formula in both derivative and difference forms, and the Euler-Maclaurin Formula and the Bernoulli polynomials. Application is made to the calculation of Euler's constant γ, sums involving binomial coefficients, and the determination of the moments of the Bernoulli distribution in the theory of statistics.

The method of summation by tables in Chapter IV continues the method of Chapters I and II by systematically extending the tables of finite sums developed earlier and employing these sums in the evaluation of finite and infinite sums of series in a manner analogous to that in which a table of indefinite integrals aids in the evaluation of definite integrals. Chapter V begins with a review of the basic concepts and tests relating to the convergence of infinite series. The role of the Taylor's Series as a generator of sums is studied. The theory of entire functions in the complex domain provides a means for summing certain real series which appear as sums of powers of reciprocal roots. Poisson's Formula is applied to the summation problem, and its power in this role is compared with that of the Euler-Maclaurin Formula. The book concludes with a table of summation formulas and an adequate index.

The text contains numerous worked examples and provides frequent reference both to historical sources and related contemporary works. With each major topic a set of problems is provided for the development of the reader's mathematical strength. The book would serve well either as a text for an advanced undergraduate course or for individual study by anyone interested in the finite difference calculus and its applications.

PAUL E. GUENTHER
Case Institute of Technology

Introduction to Modern Algebra and Analysis. By Ralph Crouch and Elbert Walker. Holt, Rinehart and Winston, New York, 1962. viii+152 pp.

Topics in Modern Algebra. By C. P. Benner, Albert Newhouse, C. B. Rader, R. L. Yates. Harper, New York. x+144 pp.

The mathematics major who completes three semesters of calculus and perhaps one of differential equations normally takes in his junior year a two-semester course in algebra and a two-semester course in advanced calculus. Unless these courses are watered down, the mortality rate is rather high and for many of the students who pass both courses a good part of the material covered is only vaguely understood. A proposed solution which is gaining ground is to have the sophomore take an additional one or two semesters of mathematics which will better prepare him for these two junior courses. Each of the above mentioned texts is designed for a two-semester course with this avowed purpose, but they differ markedly in content, style and quality.

The book by Crouch and Walker is divided into two halves, the first devoted to algebra and the second to analysis. Part I first lays the foundation with a 27 page carefully written chapter on sets, relations, and functions and then continues with groups, rings, and fields. No effort is made to stuff the student with a large number of definitions and theorems; instead, the concepts are well illustrated and motivated, and theorems are proved with the proper amount of detail. Part II develops the real number system from the natural numbers using Peano's postulates and the Dedekind cut technique, and then with considerable care and precision discusses sequences of reals, limits of functions defined on intervals, continuity, differentiation, and integration. Part II serves as an excellent preparation for a good course in advanced calculus. For those students interested in only a one-semester course leading to advanced calculus, Part II can be covered without Part I with fairly minor adjustment. Part I, being independent of calculus, can be used with bright freshmen and was used successfully by the reviewer in an NSF summer program for high ability high school juniors.

The text by Benner, Newhouse, Rader and Yates is devoted entirely to algebra. An exceptionally large number of topics are covered as indicated by the chapter headings: foundations; vector spaces; linear transformations; matrices; systems of linear equations; linear programing; polynomials; groups; rings, integral domains, and fields; linear algebras; Boolean algebras; square matrices. There is very little discussion designed to orient or motivate the reader, while proofs of theorems are sometimes faulty and almost invariably too condensed for the student whom the authors profess to have in mind. For example, in the proof of a theorem on p. 19, the product of two non-zero elements of a field is assumed tacitly to be non-zero; on p. 20 it is claimed that every set of vectors generating a vector space contains a basis of that space; on p. 93 where necessary and sufficient conditions are given for a subset of a group to constitute a subgroup, the requirement that the subset be non-empty is omitted, and on p. 105 where the quotient field F of an integral domain D is obtained it is stated that the elements of F are ordered pairs of elements of D with second member non-zero. The product of two linear transformations is called a linear transformation on p. 28 without any justification, and on p. 30 the comment made relating a linear transformation to its associated matrix is utterly confusing.

L. J. GREEN
Case Institute of Technology

THE THIRD ORDER MAGIC SQUARE

ROBERT H. SCOTT, Dunbar High School, West Virginia

Using 3 as a modulus, the nine positive digits can be partitioned into the residue classes 0, 1, and 2. The digits in these classes are (3, 6, 9), (1, 4, 7), and (2, 5, 8), respectively.

The nine residues may be arranged in a modular magic square, in which each row, column, and diagonal total 0 (mod 3), in three ways, namely:

$$\begin{matrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 2 & 2 & 2 \end{matrix} \qquad \begin{matrix} 1 & 2 & 0 \\ 0 & 1 & 2 \\ 2 & 0 & 1 \end{matrix} \qquad \begin{matrix} 0 & 1 & 2 \\ 1 & 2 & 0 \\ 2 & 0 & 1 \end{matrix}$$

If the nine digits are to be arranged in a magic square, each row, column and unbroken diagonal must total 15, since the sum of the nine digits is 45. It follows that the residues cannot be replaced by the members of their classes in the first and second arrays, since the sums of the residue classes 0 and 1 are 18 and 12, respectively.

In the third array the residues 2 may be replaced by some permutation of the members of the class. Each of the digits of the two other classes appears in a line with the central element. Hence 8 cannot be the central element, since $8+9>15$. Nor can 2 be the central element, since, choosing one element from each of the two other classes, only *two* pairs, $6+7$ and $9+4$, have a sum of 13.

Therefore 5 is the central element. If 2 is the upper right hand element, the last two pairs mentioned must occur with it in a column and a row with the extreme elements adding to 10. The "total of 15" requirement uniquely determines the positions of the two remaining digits in the fundamental third order magic square

$$\begin{matrix} 6 & 7 & 2 \\ 1 & 5 & 9 \\ 8 & 3 & 4 \end{matrix}$$

The eight "different looking" third order magic squares transform into this fundamental square by rotation or reflection or both.

PROBLEMS AND SOLUTIONS

EDITED BY ROBERT E. HORTON, Los Angeles City College

Readers of this department are invited to submit for solution problems believed to be new that may arise in study, in research, or in extra-academic situations. Proposals should be accompanied by solutions, when available, and by any information that will assist the editor. Ordinarily, problems in well-known textbooks should not be submitted. Solutions should be submitted on separate, signed sheets. Figures should be drawn in India ink and exactly the size desired for reproduction. Send all communications for this department to Robert E. Horton, Los Angeles City College, 855 North Vermont Avenue, Los Angeles 29, California.

PROPOSALS

523. *Proposed by Gilbert Labelle, Université de Montréal, Canada.*

It is well known that if, from a square of side A, we cut at each of its vertices a square of side $A/6$ and fold the resulting figure to form an open box, the box will have maximum volume. What must we do if the box of maximum volume is to be cut from a regular n-gon instead of a square?

524. *Proposed by L. Carlitz, Duke University.*

Let I denote the incenter, R the radius of the circumcircle, and r the radius of the inscribed circle of the triangle ABC. Show that $6r \leqq AI+BI+CI \leqq 3R$. Moreover, equality holds in either place if and only if ABC is equilateral.

525. *Proposed by Francis L. Miksa, Aurora, Illinois.*

It is known that the magic total, T, in a magic square of order n is $T=n(n^2+1)/2$. In how many ways can T be partitioned into n unequal parts, those parts to be chosen only from the consecutive integers from 1 to n^2?

526. *Proposed by C. W. Trigg, Los Angeles City College.*

Each of four strips, 4″ by 1″, and three strips, 3″ by 1″, has a different color of the spectrum. Into how many distinct square designs may they be arranged?

527. *Proposed by Sidney Kravitz, Dover, New Jersey.*

It is known that $f(n)=n^2-n+41$ yields prime numbers for $n=1, 2, \cdots, 40$. Find a sequence of 40 consecutive values of n for which $f(n)$ is composite.

528. *Proposed by Dewey C. Duncan, East Los Angeles College.*

In a right trihedral tetrahedron the square of the area of the face opposite the right trihedral angle is equal to the sum of the square of the areas of the other three faces. (A right trihedral angle has three right angles for its face angles.)

529. *Proposed by Huseyin Demir, Middle East Technical University, Ankara, Turkey.*

A cycloid (cardioid) rolls on a straight line without sliding. Prove that the locus of the center of curvature of the curve at the point of tangency is a circle (ellipse).

SOLUTIONS

Late Solutions

495. *Josef Andersson, Vaxholm, Sweden; Sister Marie Blanche, The Immaculata, Washington, D. C.; M. J. Pascual, Watervliet Arsenal, New York.*

496. *Josef Andersson, Vaxholm, Sweden; Sam Sesskin, Hempstead, New York.*

497. *Josef Andersson, Vaxholm, Sweden; J. L. Brown, Jr., Pennsylvania State University.*

498., 501. *Josef Andersson, Vaxholm, Sweden.*

499. *Josef Andersson, Vaxholm, Sweden; J. L. Brown, Jr., Pennsylvania State University; I. G. Patterson, Institute of Statistics, Raleigh, North Carolina; and James G. Wendel. Wendel referred to his paper, "A Problem in Geometric Probability," Math. Scand.* 11 (1962), 109–111.

500. *M. J. Pascual and Win Myint (Jointly), Watervliet Arsenal, New York.*

Errata

A number of solvers pointed out an error in Solution I to Problem 493 [September 1962 and March 1963] which invalidates that solution. When $n=5$, the units digit of n^4+4^n equals 9, not zero.

In Problem 488 [September 1962 and March 1963] the statement of the Problem should read, "For what values of n is the sum of the first n integral cubes a square?" An error occurred in translation.

A Classical Construction

502. [January 1963] *Proposed by Raphael T. Coffman, Richland, Washington.*

Using only plane geometry for construction and proof, show how to draw a tangent to an ellipse at any given point on the ellipse.

Solution by Hazel S. Wilson, Jacksonville University, Florida.

Let F_1 and F_2 be the foci and P a point on the ellipse. Bisect the angle F_1PF_2. This will be the normal to the curve at P. Construct a perpendicular at P to this angle bisector. This is the required tangent.

The proof depends on the property that the tangent to an ellipse makes equal angles with the focal radii. This is proved by elementary geometry in "Plane and Solid Analytic Geometry" by Osgood and Graustein, Macmillan, 1921, page 109. See also "Introduction to Modern Algebra" by John L. Kelley, Van Nostrand, 1960, pp. 198–199, ex 29.13 and 29.14.

Also solved by Maxey Brooke (Three solutions and references), Sweeny, Texas; Virginia Felder, University of Southern Mississippi; M. S. Klamkin, SUNY at Buffalo, New York; Aaron Lieberman, Merrick, New York; Brother Christopher Mark, St. Mary's College, California; Sam Sesskin, Hempstead, New York; Dale Woods, Northeast Missouri State Teachers College; and the proposer. Felder found the solution in "An Analytic Geometry" by E. W. Nichols, 1892, Leach, Shewell, and Sanborn, Boston. *Woods* located it in "Plane and Solid Geometry" by G. A. Wentworth, 1899. *Brooke* found a construction by straight edge only in "The Ruler in Geometrical Construction" by Smorgorzhevskii, Blaisdell, New York, 1961.

A Fibonacci Power Series

503. [January 1963] *Proposed by Brother U. Alfred, St. Mary's College, California.*

Consider the series a, a^2, a^3, a^5, a^8, $\cdots$ where every term beginning with the third is the product of the two previous terms. As the number of the terms approaches infinity, what is the limiting relationship that is approached by two successive terms?

Solution by J. L. Brown, Jr., Pennsylvania State University.

The nth term of the given series is $a^{F_{n+1}}$ where $F_1=1$, $F_2=1$, $F_{n+2}=F_{n+1}+F_n$ for $n\geq 1$, is the usual Fibonacci sequence. Since $a^{F_{n+2}}=(a^{F_{n+1}})F_{n+2}/F_{n+1}$ and $\lim_{n\to\infty} F_{n+2}/F_{n+1}=1+\sqrt{5}/2$, it follows that, asymptotically, each term is the $(1+\sqrt{5}/2)$th power of the preceding term.

Also solved by Josef Andersson, Vaxholm, Sweden; Daniel I. A. Cohen, Brooklyn, New York; Henry W. Gould, West Virginia University; John M. Howell, Los Angeles City College; Sidney Kravitz, Dover, New Jersey; Gilbert Labelle, Université de Montréal; I. G. Patterson, Institute of Statistics, Raleigh, North Carolina; Lawrence A. Ringenberg, Eastern Illinois University; David L. Silverman, Beverly Hills, California; Alan Sutcliffe, Knottingley, Yorkshire, England; and the proposer. Three solvers gave other interpretations to the statement of the problem.

An Eccentric Orbit

504. [January 1963] *Proposed by M. S. Demos, Drexel Institute of Technology.*

The orbit of the earth about the sun is an ellipse with the sun at the focus. Astronomy textbooks say that the mean distance of the sun from the earth is the semi-major axis a.

Show that the correct mean distance with respect to time is $(1+e^2/2)a$, where e is the eccentricity.

Solution by M. S. Klamkin, SUNY at Buffalo, New York.

For an elliptic orbit where

$$r = a(1 - e\cos E)$$

$$dE = \frac{k}{r}\frac{dt}{\sqrt{a}}$$

("Theoretical Mechanics," Vol. 1, Macmillan, p. 279)
Whence,

$$\bar{r} = \frac{\int r dt}{\int dt} = \frac{\int_0^{2\pi} r^2 dE}{\int_0^{2\pi} r dE} = a(1 + e^2/2),$$

by a simple integration. Also, it is easily shown that the average r with respect to polar angle is $a\sqrt{1-e^2}$. Both of these results are well known and in particular they are both posed as a problem in the aforementioned reference p. 304, prob. 25. Also, the problem as posed is given in "Theoretical Mechanics," C. J. Coe, p. 149, i.e., "In elliptic orbits the semi-major axis a of the ellipse is known in astronomy as the mean distance of the planet from the sun. Show that the actual average distance relative to the time is not a but $a(1+e^2/2)$."

Note: The arithmetic average of the perihelion distance and aphelion distance is a.

Also solved by Joseph Andersson, Vaxholm, Sweden; J. L. Brown, Jr., Pennsylvania State University; Jerome S. Shipman, Houston, Texas; and the proposer.

A Differential Sum

505. [January 1963] *Proposed by Leonard Carlitz, Duke University.*

Show that for

$$n \geqq 1 \qquad \left(x^2 \frac{d}{dx}\right)^n = \sum_{r=1}^{n} a_{nr} x^n \left(x \frac{d}{dx}\right)^r$$

where the a_{nr} are independent of x. Also identify the a_{nr}.

Solution by Henry W. Gould, West Virginia University.

We shall show that the numbers a_{nr} are essentially just the *Stirling numbers of the first kind.*

For simplicity we write

$$D = \frac{d}{dx} \cdot$$

Then clearly, for an arbitrary function y possessing required derivatives,

$$\begin{aligned}
(x^2D)y &= x(xD)y,\\
(x^2D)^2y &= x(xD)(x^2D)y = x^2(xD)y + x^2(xD)^2y,\\
(x^2D)^3y &= x(xD)(x^2D)^2y = 2x^3(xD)y + 3x^3(xD)^2y + x^3(xD)^3y.
\end{aligned}$$

In general, to proceed by induction, we suppose that there exist numbers a_{nr} such that

$$(x^2D)^n y = \sum_{r=1}^{n} a_{nr} x^n (xD)^r y;$$

and therefore

$$\begin{aligned}(x^2D)^{n+1}y &= x(xD)(x^2D)^ny\\ &= x(xD)\sum_{r=1}^{n} a_{nr}x^n(xD)^ry\\ &= x\sum_{r=1}^{n} a_{nr}(nx^n(xD)^ry + x^n(xD)^{r+1}y)\\ &= \sum_{r=1}^{n} na_{nr}x^{n+1}(xD)^ry + \sum_{r=2}^{n+1} a_{n,r-1}x^{n+1}(xD)^ry\\ &= \sum_{r=1}^{n+1} (na_{nr}x^{n+1}(xD)^ry + a_{n,r-1}x^{n+1}(xD)^ry)\end{aligned}$$

provided that we suppose $a_{nr}=0$ whenever $r>n$, and $a_{n,0}=0$ for $n\geqq 1$.

In order for the induction to hold we must therefore impose the recurrence relation

$$a_{n+1,r} = na_{n,r} + a_{n,r-1}.$$

Thus the coefficients are seen to be independent of x, and, moreover, recalling the first few values found explicitly above, we can see readily that these numbers are essentially "signless" Stirling numbers of the first kind. Indeed, following the notation of Riordan ("Combinatorial Analysis," New York, 1958, pp. 33, 48) we note that Stirling numbers of the first kind $s(n, r)$ satisfy the recurrence relation

$$s(n+1, r) = -ns(n, r) + s(n, r-1),$$

which along with the initial values implies that

$$a_{n,r} = (-1)^{n-r}s(n, r).$$

In contrast to the expansion given in this problem we may point out the related expansions

$$(xD)^ny = \sum_{r=1}^{n} S(n, r)x^rD^ry,$$

$$(x^2D)^ny = \sum_{r=0}^{n} \frac{r}{n}\binom{n}{r}\frac{n!}{r!}x^{r+n}D^ry, \qquad n\geq 0,$$

where in the former $S(n, r)$ is a *Stirling number of the second kind*, having recurrence relation $S(n+1, r) = S(n, r-1) + rS(n, r)$.

Another very interesting operational expansion which is easily established by induction is the formula

$$\{x(1+xD)\}^ny = \sum_{r=0}^{n}\binom{n}{r}\frac{n!}{r!}x^{r+n}D^ry\cdot n\geq 0.$$

Finally we remark that generalizations for $(x^pD)^n$ are readily obtainable.

Also solved by Josef Andersson, Vaxholm, Sweden; *J. L. Brown, Jr., Pennsylvania State University*; *Marvin L. Gray, Jr., Princeton, New Jersey*; *M. S. Klam-*

kin, SUNY at Buffalo, New York; Gilbert Labelle, Université de Montréal, Canada; and the proposer.

Intersecting N-Sectors

506. [January 1963] *Proposed by Leon Bankoff, Los Angeles, California.*

Let X, Y, Z be the intersections of the adjacent internal angle trisectors of a triangle ABC, forming the triangles ABZ, BCX, and CAY. If R, S, T are the feet of the internal angle bisectors XR, YS, and ZT, show that AR, BS, and CT are concurrent.

Solution by Josef Andersson, Vaxholm, Sweden.

Replace the word "trisectors" with "n-sectors" and designate the angles of the triangle by the same letters. Then

$$\frac{BR}{RC} = \frac{\sin C}{n} \Big/ \frac{\sin B}{n}, \; \cdots$$

and similarly for the other points. From the theorem of Ceva the proposition follows.

Also solved by Harry D. Ruderman, Hunter College High School, New York; and the proposer.

The Power of a Power

507. [January 1963] *Proposed by D. Rameswar Rao, Secundrabad, India.*

Prove that the power of a power always can be expressed as the geometric mean of two powers.

Solution by Alan Sutcliffe, Knottingley, Yorkshire, England.

This result is extended in the following general case.

An integer, N, is the geometric mean of two different powers greater than 1, if and only if N is neither a prime nor the square of a prime.

Proof. If $N=p$, a prime, then $N^2=p^2$ is the product of two factors in only two ways, namely $1\cdot p^2$ and $p\cdot p$. These are the only pairs of integers of which N is the geometric mean, so that N is not the geometric mean of two different powers greater than 1.

If $N=p^2$, the square of a prime, then $N^2=p^4$ is the product of two factors in only three ways, namely $1\cdot p^4$, $p\cdot p^3$, and $p^2\cdot p^2$. These are the only pairs of integers of which N is the geometric mean, so that N is not the geometric mean of two different powers greater than 1.

If N is neither a prime nor the square of a prime, we may write $N=r\cdot s$, where $r\neq s$, $r>1$, $s>1$. Since $N=\sqrt{r^2\cdot s^2}$, N is the geometric mean of two different powers greater than 1.

Note that we have also proved that any integer which is the geometric mean of two powers is the geometric mean of two squares.

Also solved by Josef Andersson, Vaxholm, Sweden; Monte Dernham, San Francisco, California; M. S. Klamkin, SUNY at Buffalo, New York; Aaron

Lieberman, Merrick, New York; Henry J. Ricardo, Fordham University; and the proposer.

Polyhedron Faces

508. [January 1963] *Proposed by David L. Silverman, Beverly Hills, California.*

Prove that every polyhedron has at least two faces with the same number of edges.

Solution by Viktors Linis, University of Ottawa, Canada.

Let e_i be the number of edges of the ith face, and let $e_i \leqq e_{i+1}$ for $i=1, 2, \cdots, F-1$ where F is the total number of faces. Assume no two faces have the same number of edges, then $e_{i+1}-e_i \geqq 1$ for all i in question. We obtain

$$e_F - e_1 = \sum_{i=1}^{F-1} (e_{i+1} - e_i) \geqq F - 1$$

and since $e_1 \geq 3$, it follows $e_F \geqq F+2$, i.e., the Fth face is adjacent to at least $F+2$ faces which is impossible.

Actually we obtain a stronger result: if the smallest number of edges of a face is k then the number of indices i satisfying $e_i = e_{i+1}$ is at least k.

Also solved by Josef Andersson, Vaxholm, Sweden; Robert Connelly, Carnegie Institute of Technology; Marvin L. Gray, Princeton, New Jersey; M. S. Klamkin, SUNY at Buffalo, New York; John W. Moon, University College, London, England; Harry D. Ruderman, Hunter College High School, New York; Alan Sutcliffe, Knottingley, Yorkshire, England; and the proposer.

QUICKIES

From time to time this department will publish problems which may be solved by laborious methods, but which with the proper insight may be disposed of with dispatch. Readers are urged to submit their favorite problems of this type, together with the elegant solution and the source, if known.

Q318. N perfectly elastic balls of equal mass are moving on the same straight line. What arrangement of velocities will produce the maximum number of collisions? [*Submitted by M. S. Klamkin*]

Q319. Factor $a^3+b^3+c^3-3abc$. [*Submitted by C. W. Trigg*]

Q320. Prove $|a|+|b| \geqq |a+b|$. [*Submitted by Elias Brettler*]

Q321. What is the necessary and sufficient condition on n that there exist a partition of n into two or more positive integers, the product of which is n? [*Submitted by D. L. Silverman*]

Q322. If the cubic $ax^3+bx+c=0$ has a double root, show that the single root is $3c/b$. [*Submitted by Harold B. Curtis*]

(Answers on page 239)

so that lim (A_{i+1}/B_{i+1}) equals a constant. Similarly, each of the $\binom{n-1}{2}$ possible quotients obtained from (25) can be set equal to some constant, and in consequence, each of the $(n-1)(n-2)/2$ ratios involving $A_{i+1}, B_{i+1}, \cdots, L_{i+1}$ can be shown to be equal to a constant. However the total number of ratios obtainable from (13), (or (24)), which must be proved to be constant is $n(n-1)/2$, of which those we have just shown to be constant are only a proper subset. The difference between this total number of ratios and the number in this subset (i.e. between $n(n-1)/2$ and $(n-1)(n-2)/2$) is $(n-1)$. However, these remaining $(n-1)$ ratios which must be proved constant are precisely those involving M_{i+1} that were not obtained in the construction above because (24n) was ignored in the construction. But instead of ignoring this equation, we could ignore any one of the other $(n-1)$ equations in the set of n equations making up (24), and we could then show by an argument exactly like the one just used that the remaining $(n-1)$ ratios in question are also constants. In other words, all of the $\binom{n}{2}$ ratios are constants, so that the assumption holds for the case when each term in S is one of n distinct numbers, provided that it holds when each term in S is any one of $(n-1)$ distinct numbers.

This completes the proof of the statement, since we have shown that the assumption holds for $n=2$, and also that if it holds for $n-1$ it must hold for n.

Answers

A318. When two balls collide they will just exchange velocities. A simpler way of looking at this is to image the balls passing through each other. If we arrange the velocities in monotonic order, we will obtain $\binom{N}{2}$ collisions. That this is maximum follows by considering the worldliness of the balls (s vs. t). The maximum number of points of intersection of N straight lines is $\binom{N}{2}$. If we have an elastic wall at one point of the line, the maximum number of collisions will be doubled.

A319. By symmetry, one factor must be $(a+b+c)$ and another factor must contain squared terms and terms of the form $-ab$ so that in the product, terms of the form a^2b will vanish, so $a^3+b^3+c^3=(a+b+c)(a^2+b^2+c^2-ab-bc-ca)$.

A320. Use $|u|=\sqrt{u^2}$ and assume that $|a|+|b|<|a+b|$. Then $\sqrt{a^2}+\sqrt{b^2} < \sqrt{(a+b)^2}$ or $a^2+2\sqrt{ab}+b^2<a^2+2ab+b^2$ which implies $|ab|<ab$, a contradiction. Therefore $|a|+|b| \geqq |a+b|$.

A321. The required condition is that n be composite. The partition obviously does not exist if n is prime, and if $n=ab$ where $a>1$ and $b>1$, the partition into a, b, and $ab-a-b$ ones does the job.

A322. Since $3ax^2+b=0$ gives $-b/3a$ as the product of the equal roots, and since the single root times $-b/3a$ equals $-c/a$, the single root is $-c/a\cdot 3a/-b =3c/b$.

(Quickies on page 270)

THE MATHEMATICAL ASSOCIATION OF AMERICA

The Association is a national organization of persons interested in mathematics at the college level. It was organized at Columbus, Ohio, in December 1915 with 1045 individual charter members and was incorporated in the State of Illinois on September 8, 1920. Its present membership is over 14,000, including more than 500 members residing in foreign countries.

Any person interested in the field of mathematics is eligible for election to membership. Annual dues of $5.00 include a subscription to the American Mathematical Monthly. Members are also entitled to reduced rates for purchases of the Carus Mathematical Monographs, MAA Studies, and for subscriptions to several journals.

Further information about the Association, its publications and its activities may be obtained by writing to:

HARRY M. GEHMAN, *Executive Director*
Mathematical Association of America
SUNY at Buffalo (University of Buffalo)
Buffalo 14, New York

MATHEMATICS MAGAZINE VOL. 36, NO. 4, SEPT.-OCT., 1963