Mathematical Entertainments

David Gale

For the general philosophy of this section see vol. 13, no. 1 (1991). *Contributors to this column who wish an acknowledgment of their contribution should enclose a self-addressed postcard.*

Somos Sequence Update

Some readers may recall that in the winter issue I reported on some sequences defined by a simple recursion that for unexplained reasons always seemed to yield integer terms. The sequences were originally introduced by Michael Somos and can be described as follows: Given an integer $k \geq 4$ a *Somos* (k) *sequence* is characterized by the recursion

$$a_n a_{n-k} = x_1 a_{n-1} a_{n-k+1} + x_2 a_{n-2} a_{n-k+2} + \ldots + x_r a_{n-r} a_{n-k+r} \tag{1}$$

where $r = \left[\frac{k}{2}\right]$ and the x_i are given integers.

Since in (1), a_n is defined in terms of the preceding k terms, one must choose *initial values* for $a_0, a_1, \ldots, a_{k-1}$.

When I simply refer to Somos (k), I will mean a k-sequence in which all the x_i and initial a_i are unity. It was first observed numerically and later proved that Somos 4, 5, 6, 7 always have integer terms whereas Somos 8, 9 and presumably all the rest do not. But there still remains a doubly-infinite family of Somos sequences that appear to have integer terms, although this has not been proved (see winter column for details).

Motivated by the Somos phenomena, Dana Scott discovered that sequences with initial values unity and the following recursions have integer terms:

$$a_n a_{n-k} = a_{n-1}^2 + a_{n-2}^2 + \ldots + a_{n-k+1}^2 \tag{2}$$

$$a_n a_{n-k} = a_{n-1} a_{n-2} + a_{n-3} a_{n-4} + \ldots + a_{n-k+2} a_{n-k+1} \tag{3}$$

for k odd.

The integer property also holds for

$$a_n a_{n-k} = a_{n-1} a_{n-2} + a_{n-2} a_{n-3} + \ldots + a_{n-k+2} a_{n-k+1} \tag{4}$$

Proofs of integrality of (2), (3) and (4) have now been found by Raphael Robinson with an assist at one point from Dean Hickerson. The proofs are quite elementary, so we will present the one for (2), the argument for (3) and (4) being similar. They involve finding rational functions that are *invariant* under the recursions. For (2) we define a new sequence (b_n) for $n \geq k$ by

$$b_n = \frac{a_n + a_{n-k}}{a_{n-1} a_{n-2} \ldots a_{n-k+1}} \tag{5}$$

and the claim is that the b_n are constants, that is, $b_{n+1} = b_n$. To see this note that

$$a_n(a_n + a_{n-k}) = (a_{n+1} + a_{n-k+1}) a_{n-k+1} \tag{6}$$

because from (2) both sides of (6) are equal to $a_n^2 + a_{n-1}^2 + \ldots + a_{n-k+1}^2$. Dividing both sides by $a_n a_{n-1} \ldots a_{n-k+1}$ gives $b_n = b_{n+1}$. But from the initial conditions $b_k = k$ hence $b_n = k$, so from (5) we have

$$a_n = k(a_{n-1} a_{n-2} \ldots a_{n-k+1}) - a_{n-k} \tag{7}$$

which gives a new recursion for the a_n where the right-hand side is a polynomial (rather than a rational function) of the a_i, so integrality follows, as does the fact that the sequence reduced mod m is *periodic* for any m.

But while some problems have now been solved, further numerical explorations by Robinson brought to light a host of new structural properties of Somos sequences, some of them number-theoretic, others analytic. Since these results will appear elsewhere, I will just mention a few of them. First, as mentioned at the end of the earlier column, all Somos sequences that give integers appear to be periodic when reduced mod m for any m. Robinson proved this for Somos 4 and 5 but not for 6 and 7. For 4 and 5 the period as a function of m seems unpredictable, but the following striking relation was observed: For all m except 2 the period mod m^k is equal to m^{k-1} times the period mod m. For 2 a somewhat more complicated pattern holds. (At this point I will stop qualifying every statement with "seems to," which is to be understood.) Robinson also investigated which primes divide the various terms of the sequences and found that for Somos 4 and 5 (but not for 6 and 7), the terms divisible by a given prime were equally spaced. Thus in Somos (4) every fifth

***Column editor's address:** Department of Mathematics, University of California, Berkeley, CA 94720 USA.

term is even, every seventeenth term is divisible by 11 and none is divisible by 5, while in Somos (5) every tenth term is divisible by 5 but none is divisible by 7.

In a different direction, Robinson investigated analytic properties of the sequences for arbitrary k and initial values and found in all cases tested that there were (unique) constants C and D such that

$$a_n = C^{(n-D)^2}\,\phi(n) \tag{8}$$

where $\phi(n)$ has an oscillation with a well-defined period. The constants C and D depend on the initial values as well as on k but in an apparently unpredictable manner. Learning of Robinson's data, Clifford Gardner succeeded in finding explicit formulas for Somos 4 and 5 in terms of Jacobi elliptic functions, so in some sense the problem is starting to come full circle, since Somos originally discovered his sequences while studying properties of elliptic functions and was aware of some of the phenomena described here.

To a non-expert the analytic and number-theoretic properties of the Somos sequences seem unrelated, but perhaps algebraic-number theorists who are accustomed to such things will be able to make a connection. In any case, it is intriguing to see more and more properties of these sequences being revealed by numerical exploration.

Unconditionally Secure Protocols

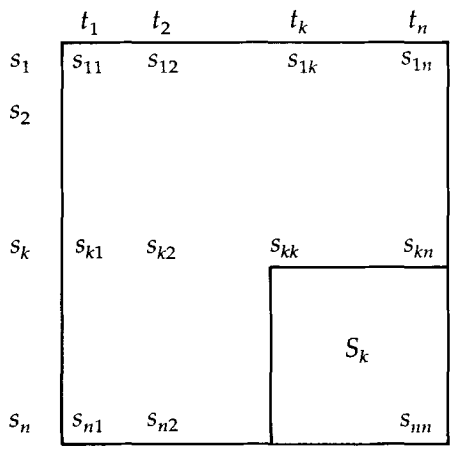
One of the exciting mathematical developments of the past decade was the discovery of so-called uncrackable public key codes. These are codes with the characteristic that everyone knows the method of encryption, but the amount of calculation required for an outsider to break the code is thought to be beyond present computational capabilities. In the best-known example, breaking the code was equivalent to being able to find the factors of, say, a 100-digit number, which was believed to be computationally infeasible.

A more recent but less well known development with somewhat the same flavor involves methods of conveying certain information that depends on other information that must remain secret. In these cases, however, it is literally impossible for anyone but a mind-reader to learn the secrets. Here is a simple example. Some people $P_1, \ldots, P_n$, say, the members of a mathematics department, are interested in learning their average salary but they are not willing to reveal their own salaries to anyone else. How can this be done? I put the problem to some of my colleagues and they were not able to come up with an answer. I also mentioned it at a social gathering and rather quickly a young woman who hadn't had a mathematics course since high school (and claimed she'd failed 9th grade algebra) proposed the following simple solution: P_1 chooses an arbitrary number x and tells it to P_2 who adds his salary and tells the total to P_3 who adds her

Some people $P_1, \ldots, P_n$, say, the members of a mathematics department, are interested in learning their average salary but they are not willing to reveal their own salaries to anyone else. How can this be done?

salary and tells it to P_4, and so on, until P_n adds his salary and tells it to P_1 who adds her salary, subtracts x, divides by n and announces the result. Clearly no one has learned anything about anyone else's salary except what can be inferred from knowing the average.

Now while it is true that in the above scheme no person acting alone can discover anything about the other people's salaries, the situation changes if people are allowed to collude. For example, if P_1 reveals x to P_3 then P_3 will know P_2's salary. Thus the scheme, or *protocol* as it is called, is said to be *1-private* but not *2-private*. One may then ask if there are any 2-private protocols for this problem. The answer is that, in fact, there is an n-private protocol which is also easy to describe. A protocol is called *n-private* if no subset of the people by colluding can learn anything about the complementary set except what can be inferred from their knowledge of the average. Here is how it works. Let s_i be the salary of P_i. Each P_i chooses n numbers s_{ij} subject only to the condition that they sum to s_i, and deals s_{ij} to P_j. Now each P_j announces the sum t_j of the numbers in his hand. The sum of the t_j is of course the sum of the s_i, as desired. The situation is represented by the matrix $S = (s_{ij})$ where the i^{th} row sum is s_i and the j^{th} column sum is t_j.



To see that the protocol is n-private, suppose, say, the first k players collude. Then they will know all entries of S except those in the lower $(n-k) \times (n-k)$ submatrix S_k and they will know as well $t_{k+1}, \ldots, t_n$, so they will know the column sums of S_k. But if one knows only the column sums c_j of a matrix then the row sums r_i can be any numbers subject only to the

condition $\Sigma r_i = \Sigma c_j$, so the colluding players will know only the sum of the other players' salaries, which they would know anyway from knowing the sum of all the salaries.

The sum protocol can be used to learn other things, for example, the distribution of salaries, that is, the number of people at each salary level, without revealing who they are. To find out how many people have salary x, do the sum protocol where P_i's secret number is 1 or 0 according as his salary is or is not equal to x, and repeat the protocol for all values of x. A more efficient method is for P_i to choose the secret number $(n + 1)^{s_i}$. When the sum is computed it is expressed in base $(n + 1)$ and the coefficient of $(n + 1)^x$ will be the number of people whose salary is x. The same trick can be used for a secret ballot. Suppose the candidates for some office are labeled 1 through m. A person who wants to vote for candidate k should use the secret number $(n + 1)^k$. The sum protocol then gives the vote count.

What about other functions? For example, the maximum rather than the sum of the salaries? If an upper bound $\bar{s}$ of the salaries is known the following procedure suggests itself. Use the sum protocol to find out how many people have salary $\bar{s}$. If the answer is zero try again with $\bar{s} - 1$ and so on until the sum is positive. The trouble is that one learns too much. One learns not only the maximum salary but also the number of people who earn the maximum. Is it possible to learn the maximum and nothing more? In the same spirit, is it possible to learn only the winning candidate(s) in an election but nothing about the distribution of votes? And a simple arithmatical question: the sum of n numbers can be computed n-privately—What about the product?

The answer to all of these questions is the same and is quite surprising. There exist t-private protocols for all of them if and only if t is less than $n/2$. Such protocols might be called *minority-private*. The existence of minority-private protocols was proved by Ben-Or, Goldwasser, and Wigderson [2] and independently by Chaum, Crépeau, and Damgard [3]. Given secret numbers $s_1, \ldots, s_n$ that may take on some finite set of values, it is shown that any function of the s_i can be computed minority-privately. It suffices to consider functions on a sufficiently large finite field. A minority-private protocol is given for multiplication, which is somewhat more complicated than that for addition. (Everything we have described up to now could probably be understood by a competent 7th grader. The multiplication protocol is about at the level of an undergraduate abstract algebra course.) Once one has addition and multiplication, one has polynomials and hence all functions on a finite field. It seems to be the case that by suitable encoding most problems of the sort one is interested in can be transformed into a problem of calculating a function from integers to integers, although this is not immediately obvious, for example, in the secret-ballot problem where one wants to know only which candidate won the election.

Perhaps even more striking than the sufficiency is the necessity of the condition $t < n/2$. This means there is no protocol, for example, for computing the product of n secret numbers that can maintain secrecy if half or more of the participants decide to collude. In fact, essentially the only functions that can be computed *majority-privately* are functions that can be obtained using only the sum protocol. This was first shown by Chor and Kushilevitz [4] for Boolean functions and then by Beaver [1] for general integer-valued functions. Notice that we have nowhere up to now said what a protocol actually is, but have simply exhibited examples. This is fine as long as one is proving existence theorems. By way of analogy, to show that there is a "formula" for the roots of third and fourth degree polynomials one simply displays them and checks that they work. On the other hand, in order to show *non-existence* of such expressions for higher degree polynomials, a strict formalization of the problem is necessary. In the same way, to prove non-existence of majority-private protocols, one must have precise definitions of protocols and privacy and then develop the necessary theory to deal with these concepts; and the arguments are considerably more involved than those for existence.

As a special case of Beaver's result we see that when there are only two people, essentially nothing can be learned privately, as, for example, whether they have the same secret number. On the other hand, from the existence theorem we know that if a third party P_3 enters the picture and is able to give and receive messages, then P_1 and P_2 can learn whether or not they have the same number 1-privately, and P_3 will not even know whether the answer is yes or no.

There is a good deal more to the theory than has been mentioned. For example if one does not require unconditional security but only "uncrackability" in the sense described in the first paragraph, then it has been shown that any function can be computed n-privately, including the situation where there are only two people. In the so-called "millionaires problem" of Yao [5], for example, P_1 and P_2 can learn which of them has the larger salary and nothing else.

To conclude let me return to the 7th grade level and describe a 1-private protocol that computes the maximum salary. For this we bring in an outsider P_0 who chooses some secret number x_0. The rules are then the following: if P_i's salary is $\bar{s}$ (the upper bound), she chooses some arbitrary positive number x_i. If not her secret number is 0. Now do the sum protocol. If the sum is not x_0, then P_0 announces that $\bar{s}$ is the maximum. If the sum is x_0, play again, replacing $\bar{s}$ by $\bar{s} - 1$, and so on until the maximum is found. Notice that it is necessary to bring in P_0 because if the others played the game without him and at some stage the sum

turned out to be x_i, then P_i would know that she was the only one getting the maximum. Similarly, the protocol with P_0 is only 1-private, because if P_0 gets together with a person earning the maximum salary, then the two of them will know whether or not anyone else is also earning this maximum.

I want to express my thanks to Donald Beaver of AT&T for much of the material I have presented and to Michael Hirsch of UC Berkeley for bringing this interesting subject to my attention. It seems there are more kinds of mathematics in heaven and on earth than are dreamed of in all your volumes of Bourbaki.

References

1. D. Beaver, Perfect privacy for two-party protocols, *Harvard Tec. Report* TR-11-89, Aiken Computer Laboratory.
2. M. Ben-Or, S. Goldwasser, and A. Wigderson, Completeness theorems for non-cryptographic fault tolerant distributed computations, *Proceedings 20th STOC* (1988), 1–10.
3. D. Chaum, C. Crépeau, and I. Damgard, Multiparty unconditionally secure protocols, *Proceedings 20th STOC* (1988), 11–19.
4. B. Chor and E. Kushilevitz, A zero-one law for Boolean privacy, *Proceedings 21st STOC* (1989), 62–72.
5. A. C. Yao, Protocols for secure computation, *Proceedings FOCS* (1982), 160–164.

A True Story

Once upon a time there was a little girl named Clara who was barely three years old and had just learned how to count. She could tell how many chairs there were in the living room and the number of steps down from the front porch. One day her father decided to test her. "Look" he said, "I've brought you these four lollipops," but he handed her only three. Clara took the lollipops and dutifully counted, "One, two, four." Then she looked up a bit puzzled and asked, "Where's the third one?"

Problems

Rational primes: Quickie 91-5 by W. Sierpiński (submitted by S. H. Weintraub).

Call a rational number a *prime rational* if it is the quotient of (integer) primes. Show that the set of prime rationals is dense in the positive reals.

Solutions

Derivatives eventually zero: Problem 91-1 by E. M. E. Wermuth (Jülich, Germany)

Let f be a C^∞-function defined on some open interval (a,b) such that for every x in (a,b) there is an integer $n(x)$ such that $f^{(n(x))}(x) = 0$. Show that f is a polynomial. (For multidimensional versions of the problem and its history see **MR**90e:26040.)

Solution by Proposer (slightly rephrased)

Call f *locally polynomial* at x if there is a polynomial P such that $f = P$ in some neighborhood of x. Now if f is locally polynomial at each point of some open subinterval (c,d), then it is given by a single polynomial in (c,d); for if P is the local polynomial at some point x, then the set of points x' where $f(x') = P(x')$ is both open and closed in (c,d) since f is C^∞. Let A be the closed set of all points x such that f is not locally polynomial at x. We must show that A is empty. First, A contains no isolated points; for if c were isolated in A, then f would be given by a polynomial in the component of $(a,b)\backslash A$ to the right of c and by a polynomial in the component to the left of c, and these two polynomials would have to be the same since $f \in C^\infty$.

Let $A_n = \{x \in A \mid f^{(n)}(x) = 0\}$. Then $A = \cup_{n=0}^{\infty} A_n$ by hypothesis and the result will follow from Baire's theorem if we can show that each closed subset A_n of A is nowhere dense in A. So suppose that for some n and for some open interval I, $A \cap I \subseteq A_n$. From the previous paragraph every x in $A \cap I$ is an accumulation point of A_n, hence by Rolle's theorem $f^{(k)}(x) = 0$ for $k \geq n$ for all x in $A \cap I$. But then this must also be the case for all x in I, since in each component of $I\backslash A$, f is a polynomial and the polynomials can have degree at most $n - 1$ (if the degree were $k \geq n$, $f^{(k)}$ would be a non-zero constant in this component, so we would have $f^{(k)}(x) \neq 0$ for some x in $A \cap I$, an endpoint of this component). But if $f^{(k)}(x) = 0$ for all x in I and for all $k \geq n$, then f is a polynomial on the interval I, so $A \cap I = \varnothing$. ■