Mathematics: An Experimental Science

By Herbert S. Wilf

1 The Mathematician's Telescope

Albert Einstein once said "You can confirm a theory with experiment, but no path leads from experiment to theory." But that was before computers. In mathematical research now, there's a very clear path of that kind. It begins with wondering what a particular situation looks like in detail; it continues with some computer experiments to show the structure of that situation for a selection of small values of the parameters of the problem; and then comes the human part: the mathematician gazes at the computer output, attempting to see and to codify some patterns. If this seems fruitful, then the final step requires the mathematician to prove that the apparent pattern is really there, and is not a shimmering mirage above the desert sands.

A computer is used by a pure mathematician in much the same way that a telescope is used by a theoretical astronomer. It shows us "what's out there." Neither the computer nor the telescope can provide a theoretical explanation for what it sees, but both of them extend the reach of the mind by providing numerous examples that might otherwise be hidden, and from which one has some chance of perceiving, and then demonstrating, the existence of patterns, or universal laws.

In this article I would like to show you some examples of this process at work. Naturally the focus will be on examples in which some degree of success has been realized, rather than on the much more numerous cases where no pattern could be perceived, at least by my eyes. Since my work is mainly in combinatorics and discrete mathematics, the focus will also be on those areas of mathematics. It should not be inferred that experimental methods are not used in other areas; only that I don't know those applications well enough to write about them.

In one short article we cannot even begin do justice to the richly varied, broad, and deep achievements of experimental mathematics. For further reading, see the journal *Experimental Mathematics* and the books by Borwein and Bailey (2003) and Borwein, Bailey, and Girgensohn (2004).

In the following sections we give first a brief description of some of the useful tools in the armament of experimental mathematics, and then some successful examples of the method, if it is a method. The examples have been chosen subject to fairly severe restrictions:

(1) the use of computer exploration was vital to the success of the project; and

(2) the outcome of the effort was the discovery of a new theorem in pure mathematics.

I must apologize for including several examples from my own work, but those are the ones with which I am most familiar.

2 Some of the Tools in the Toolbox

2.1 Computer Algebra Systems

The mathematician who enjoys using computers will find an enormous number of programs and packages available, beginning with the two major computer algebra systems (CASs), Maple and Mathematica. These programs can provide so much assistance to a working mathematician that they must be regarded as essential pieces of one's professional armamentarium. They are extremely user-friendly and capable.

Typically one uses a CAS in interactive mode, meaning that you type in a one-line command and the program responds with its output, then you type in another line, etc. This modus operandi will suffice for many purposes but for best results one should learn the programming languages that are embedded in these packages. With a little knowledge of programming, one can ask the computer to look at larger and larger cases until something nice happens, then take the result and use another package to learn something else, and so forth. Many are the times when I have written little programs in Mathematica or Maple and then gone away for the weekend leaving the computer running and searching for interesting phenomena.

2.2 Neil Sloane's Database of Integer Sequences

Aside from a CAS, another indispensable tool for experimentally inclined mathematicians, particularly for combinatorialists, is Neil Sloane's "On-Line Encyclopedia of Integer Sequences," which

is on the web at ⟨www.research.att.com/~njas⟩. At present, this contains nearly $100\,000$ integer sequences and has full search capabilities. A great deal of information is given for each sequence.

Suppose that for each positive integer n you have an associated set of objects that you want to count. You might, for example, be trying to determine the number of sets of size n with some given property, or you might wish to know how many prime divisors n has (which is the same as counting the set of these prime divisors). Suppose further that you've found the answer for $n = 1, 2, 3, \ldots, 10$, say, but you haven't been able to find any simple formula for the general answer.

Here's a concrete example. Suppose you're working on such a problem, and the answers that you get for $n = 1, 2, \ldots, 10$ are 1, 1, 1, 1, 2, 3, 6, 11, 23, 47. The next step should be to look online to see if the human race has encountered your sequence before. You might find nothing at all, or you might find that the result that you'd been hoping for has long since been known, or you might find that your sequence is mysteriously the same as another sequence that arose in quite a different context. In the latter case, an example of which is described below in Section 3, something interesting will surely happen next. If you haven't tried this before, do look up the little example sequence above, and see what it represents.

2.3 Krattenthaler's package Rate

A very helpful Mathematica package for guessing the form of hypergeometric sequences has been written by Christian Krattenthaler and is available from his website. The name of the package is Rate (rot'-eh), which is the German word for "guess."

To say what a hypergeometric sequence is let's first recall that a rational function of n is a quotient of two polynomials in n, like $(3n^2 + 1)/(n^3 + 4)$. A hypergeometric sequence $\{t_n\}_{n\geqslant 0}$ is one in which the ratio t_{n+1}/t_n is a rational function of the index n. For example, if $t_n = \binom{n}{7}$ then t_{n+1}/t_n works out to be $(n + 1)/(n - 6)$, which is a rational function of n, so $\{t_n\}_{n\geqslant 0}$ is a hypergeometric sequence. Other examples are

$$n!, \ (7n + 3)!, \ \binom{n}{7}t^n, \ \frac{(3n + 4)!(2n - 3)!}{4^n n!^4},$$

all of which are hypergeometric sequences.

If you input the first several members of the unknown sequence, Rate will look for a hypergeometric sequence that takes those values. It will also look for a hyper-hypergeometric sequence (i.e. one in which the ratio of consecutive terms is hypergeometric), and a hyper-hyper-hypergeometric sequence, etc.

For example, the line

```
Rate[1, 1/4, 1/4, 9/16, 9/4, 225/16]
```

elicits the (somewhat inscrutable) output

$$\{4^{1-i0}(-1 + i0)!^2\}.$$

Here $i0$ is the running index of Rate, so we would normally write that answer as, say,

$$\frac{(n - 1)!^2}{4^{n-1}} \quad (n = 1, 2, 3, 4, 5, 6),$$

which fits the input sequence perfectly. Rate is a part of the *Superseeker* front end to the Integer Sequences database, discussed in Section 2.2 above.

2.4 Identification of Numbers

Suppose that, in the course of your work, you encountered a number, let's call it β, which, as nearly as you could calculate it, was 1.218041583332573. It might be that β is related to other famous mathematical constants, like π, e, $\sqrt{2}$, and so forth, or it might not. But you'd like to know.

The general problem that is posed here is the following. We are given k numbers, $\alpha_1, \ldots, \alpha_k$ (the *basis*), and a target number α. We want to find integers $m, m_1, \ldots, m_k$ such that the linear combination

$$m\alpha + m_1\alpha_1 + m_2\alpha_2 + \cdots + m_k\alpha_k \qquad (2.1)$$

is an extremely close numerical approximation to 0. For, suppose we had a computer program that could find such integers, how would we use it to identify the mystery constant $\beta = 1.218041583332573$?

We would take the α_i to be a list of the logarithms of various well-known universal constants and prime numbers, and we would take $\alpha = \log \beta$. For example, we might use

$$\{\log \pi, 1, \log 2, \log 3\} \qquad (2.2)$$

as our basis. If we then find integers $m, m_1, \ldots, m_4$ such that

$$m \log \beta + m_1 \log \pi + m_2 + m_3 \log 2 + m_4 \log 3 \quad (2.3)$$

is extremely close to 0, then we will have found that our mystery number β is extremely close to

$$\beta = \pi^{-m_1/m} \mathrm{e}^{-m_2/m} 2^{-m_3/m} 3^{-m_4/m}. \quad (2.4)$$

At this point we will have a judgment to make. If the integers m_i seem rather large, then the presumed evaluation (2.4) is suspect. Indeed for any target α and basis $\{\alpha_i\}$ we can always find huge integers $\{m_i\}$ such that the linear combination (2.1) is *exactly* 0, to the limits of machine precision. The real trick is to find that the linear combination is extraordinarily close to 0, while using only "small" integers m, m_i, and that is a matter of judgment. If the judgment is that the relation found is real, rather than spurious, then there remains the little job of proving that the suspected evaluation of α is correct, but that task is beyond our scope here. For a nice survey of this subject see Bailey and Plouffe (1997).

There are two major tools that can be used to discover linear dependencies such as (2.1) among the members of a set of real numbers. They are the algorithms PSLQ, of Ferguson and Forcade (1979), and LLL, of Lenstra, Lenstra, and Lovász (1982), which uses their lattice basis reduction algorithm. For the working mathematician, the good news is that these tools are available in CASs. For example, Maple has a package, `IntegerRelations[LinearDependency]`, which places the PSLQ and the LLL algorithms at the immediate disposal of the user. Similarly there are Mathematica packages on the web that can be freely downloaded and which perform the same functions.

An application of these methods will be given below in Section 7. For a quick illustration, though, let us try to recognize the mystery number $\beta = 1.218041583332573$. We use as a basis the list in (2.2) above, and we put this list, augmented by $\log 1.218041583332573$, into the `IntegerRelations[LinearDependency]` package. The output is the integer vector $[2, -6, 0, 3, 4]$, which tells us that $\beta = \pi^3 \sqrt{2}/36$, to the number of decimal places carried.

2.5 Solving Partial Differential Equations

I had occasion recently to need the solution to a certain partial differential equation (PDE) that arose in connection with a research problem that was posed by Graham, Knuth, and Patashnik (1989). It was a first-order linear PDE, so in principle the method of characteristics gives the solution. As those who have tried that method know, it can be fraught with technical difficulties relating to the solution of the associated ordinary differential equations.

However, some extremely intelligent packages are available for solving PDEs. I used the Maple command `pdsolve` to handle the equation

$$(1 - \alpha x - \alpha' y) \frac{\partial u(x, y)}{\partial x}$$
$$= y(\beta + \beta' y) \frac{\partial u(x, y)}{\partial y} + (\gamma + (\beta' + \gamma')y)u(x, y)$$

with $u(0, y) = 1$. `pdsolve` found that

$$u(x, y) = \frac{(1 - \alpha x)^{-\gamma/\alpha}}{(1 + (\beta'/\beta)y(1 - (1 - \alpha x)^{-\beta/\alpha}))^{1+\gamma'/\beta'}}$$

is the solution, and that enabled me to find explicit formulas for certain combinatorial quantities, with much less work and fewer errors than would otherwise have been possible.

3 Thinking rationally

The following problem appeared in the September/October 1997 issue of *Quantum* (and was chosen by Stan Wagon for the Problem of the Week archive).

How many ways can $90\,316$ be written as

$$a + 2b + 4c + 8d + 16e + 32f + \cdots,$$

where the coefficients can be any of 0, 1, or 2?

In standard combinatorial terminology, the question asks for the number of *partitions* of the integer $90\,316$ into powers of 2, where the multiplicity of each part is at most 2.

Let's define $b(n)$ to be the number of partitions of n, subject to the same restrictions. Thus $b(5) = 2$ and the two relevant partitions are $5 = 4 + 1$ and $5 = 2 + 2 + 1$. Then it is easy to see that

$b(n)$ satisfies the recurrences $b(2n+1) = b(n)$ and $b(2n+2) = b(n) + b(n+1)$, for $n = 0, 1, 2 \ldots$, with $b(0) = 1$.

It is now easy to calculate particular values of $b(n)$. This can be done directly from the recurrence, which is quite fast for computational purposes. Alternatively, it can be shown quite easily that our sequence $\{b(n)\}_0^\infty$ has the generating function

$$\sum_{n=0}^{\infty} b(n)x^n = \prod_{j=0}^{\infty} (1 + x^{2^j} + x^{2 \cdot 2^j}).$$

(For more information on generating functions, see ALGEBRAIC AND ENUMERATIVE COMBINATORICS , or see Wilf (1994).) This helps us to avoid much programming when working with the sequence, because we can use the built-in series-expansion instructions in Mathematica or Maple to show us a large number of terms in this series quite rapidly. Returning to the original question from *Quantum*, it is a simple matter to compute $b(90\,316) = 843$ from the recurrence. But let's try to learn more about the sequence $\{b(n)\}$ in general. To do that we open up our telescope, and calculate the first 95 members of the sequence, i.e. $\{b(n)\}_0^{94}$, which are shown in Table 1.1 below. The question is now, as it always is in the mathematics laboratory, what patterns do you see in these numbers?

Just for instance, one might notice that when n is 1 less than a power of 2, it seems that $b(n) = 1$. The reader who is fond of such puzzles is invited to cease reading here for the moment (without peeking at the next paragraph), and look at Figure 1.1 to spend some time finding whatever interesting patterns seem to be there. Computations up to $n = 94$ aren't as helpful for a quest like this as computations up to $n = 1000$ or so might be, so the reader is also invited to compute a much longer table of values of $b(n)$, using the above recurrence formulas, and to study it carefully for fruitful patterns.

OK, did you notice that if $n = 2^a$ then $b(n)$ appears to be $a + 1$? How about this one: in the block of values of n between 2^a and $2^{a+1} - 1$, inclusive, the largest value of $b(n)$ that seems to occur is the Fibonacci number F_{a+2}. There are many intriguing things going on in this sequence, but the one that was of crucial importance in understanding it was the observation that *consecutive values of $b(n)$ seem always to be relatively prime*.[1]

It was totally unexpected to find a property of the values of this sequence that involved the multiplicative structure of the positive integers, rather than their additive structure, which would have been quite natural. This is because the theory of partitions of integers belongs to the additive theory of numbers, and multiplicative properties of partitions are rare and always cherished.

Once this relative primality is noticed, the proof is easy. If m is the smallest n for which $b(n)$, $b(n+1)$ fail to be relatively prime, then suppose $p > 1$ divides both of them. If $m = 2k + 1$ is odd, then the recurrence implies that p divides $b(k)$ and $b(k+1)$, contradicting the minimality, whereas if $m = 2k$ is even, the recurrence again gives that result, finishing the proof.

Why was it so interesting that consecutive values appeared to be relatively prime? Well, at once that raised the question of whether every possible relatively prime pair (r, s) of positive integers occurs as a pair of consecutive values of this sequence, and if so, whether every such pair occurs once and only once. Both of those possibilities are supported by the table of values above, and upon further investigation both turned out to be true. See Calkin and Wilf (2000) for details.

The bottom line here is that *every positive rational number occurs once and only once, and in reduced form, among the members of the sequence* $\{b(n)/b(n+1)\}_0^\infty$. Hence the partition function $b(n)$ induces an enumeration of the rational numbers, a result which was found by gazing at a computer screen and looking for patterns.

Moral: be sure to spend many hours each day gazing at your computer screen and looking for patterns.

4 An Unexpected Factorization

One of the great strengths of computer algebra systems is that they are very good at factoring. They can factor very large integers and very complicated expressions. Whenever you run into some large expression as the answer to a problem that interests you, it is good practice to ask your CAS

---

[1]Two positive integers are *relatively prime* if they have no common factor.

Table 1.1. The first 95 values of $b(n)$.

| 0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1	2	1	3	2	3	1	4	3	5	2	5	3	4	1	5	4	7
19	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**
3	8	5	7	2	7	5	8	3	7	4	5	1	6	5	9	4	11	7
38	**39**	**40**	**41**	**42**	**43**	**44**	**45**	**46**	**47**	**48**	**49**	**50**	**51**	**52**	**53**	**54**	**55**	**56**
10	3	11	8	13	5	12	7	9	2	9	7	12	5	13	8	11	3	10
57	**58**	**59**	**60**	**61**	**62**	**63**	**64**	**65**	**66**	**67**	**68**	**69**	**70**	**71**	**72**	**73**	**74**	**75**
7	11	4	9	5	6	1	7	6	11	5	14	9	13	4	15	11	18	7
76	**77**	**78**	**79**	**80**	**81**	**82**	**83**	**84**	**85**	**86**	**87**	**88**	**89**	**90**	**91**	**92**	**93**	**94**
17	10	13	3	14	11	19	8	21	13	18	5	17	12	19	7	16	9	11
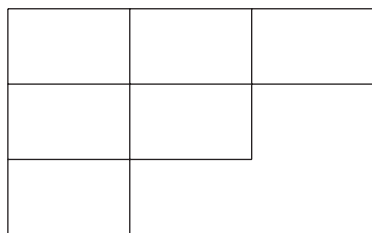
Figure 1.1. The Ferrers board.
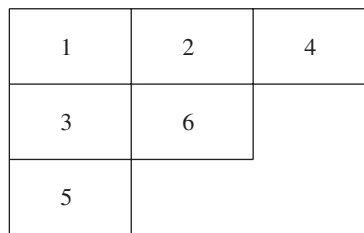
Figure 1.2. A Young tableau.

to factor it for you. Sometimes the results will surprise you. This is one such story.

The theory of *Young tableaux* forms an important part of modern combinatorics. To create a Young tableau we choose a positive integer n and a partition $n = a_1 + a_2 + \cdots + a_k$ of that integer. We'll use the integer $n = 6$ and the partition $6 = 3+2+1$ as an example. Next we draw the *Ferrers board* of the partition, which is a truncated chessboard that has a_1 squares in its first row, a_2 in its second row, etc., the rows being left-justified. In our example, the Ferrers board is as shown in Figure 1.1.

To make a tableau, we insert the labels $1, 2, \ldots, n$ into the n cells of the board in such a way that the labels increase from left to right across each row and increase from top to bottom down every column. With our example, one way to do this is as shown in Figure 1.2.

One of the important properties of tableaux is that there is a one-to-one correspondence, known as the Robinson–Schensted–Knuth (RSK) correspondence, which assigns to every permutation of n letters a pair of tableaux of the same shape. One use of the RSK correspondence is to find the length of the longest increasing subsequence in the vector of values of a given permutation. It turns out that this length is the same as the length of the first row of either of the tableaux to which the permutation corresponds under the RSK mapping. This fact gives us a good way, algorithmically speaking, of finding the length of the longest increasing subsequence of a given permutation.

Now suppose that $u_k(n)$ is the number of permutations of n letters that have no increasing subsequence of length greater than k. A spectacular theorem of Gessel (1990) states that

$$\sum_{n \geqslant 0} \frac{u_k(n)}{n!^2} x^{2n} = \det(I_{|i-j|}(2x))_{i,j=1,\ldots,k}, \quad (4.1)$$

in which $I_\nu(t)$ is (the modified Bessel function)

$$I_\nu(t) \stackrel{\text{def}}{=} \sum_{j=0}^{\infty} \frac{(\frac{1}{2}t)^{2j+\nu}}{j!(j+\nu)!}.$$

At any rate, it seems fairly "spectacular" to me that when you place various infinite series such as the above into a $k \times k$ determinant and then expand the determinant, you should find that the

coefficient of x^{2n}, when multiplied by $n!^2$, is exactly the number of permutations of n letters with no increasing subsequence longer than k.

Let's evaluate one of these determinants, say the one with $k = 2$. We find that

$$\det(I_{|i-j|}(2x))_{i,j=1,2} = I_0^2 - I_1^2,$$

which of course factors as $(I_0 + I_1)(I_0 - I_1)$. The arguments of the I_ν are all $2x$ and have been omitted.

When $k = 3$, no such factorization occurs. If you ask your CAS for this determinant when $k = 4$, it will show you

$$I_0^4 - 3I_0^2 I_1^2 + I_1^4 + 4I_0 I_1^2 I_2$$
$$- 2I_0^2 I_2^2 - 2I_1^2 I_2^2 + I_2^4 - 2I_1^3 I_3$$
$$+ 4I_0 I_1 I_2 I_3 - 2I_1 I_2^2 I_3 - I_0^2 I_3^2 + I_1^2 I_3^2,$$

where now we have abbreviated $I_\nu(2x)$ simply by I_ν. If we ask our CAS to factor this last expression, it (surprisingly) replies with

$$(I_0^2 - I_0 I_1 - I_1^2 + 2I_1 I_2 - I_2^2 - I_0 I_3 + I_1 I_3)$$
$$\times (I_0^2 + I_0 I_1 - I_1^2 - 2I_1 I_2 - I_2^2 + I_0 I_3 + I_1 I_3),$$

which is actually of the form $(A + B)(A - B)$, as a quick inspection will reveal.

We have now observed, experimentally, that for $k = 2$ and $k = 4$ Gessel's $k \times k$ determinant has a nontrivial factorization of the form $(A+B)(A-B)$, in which A and B are certain polynomials of degree $k/2$ in the Bessel functions. Such a factorization of a large expression in terms of formal Bessel functions simply cannot be ignored. It demands explanation. Does this factorization extend to all even values of k? It does. Can we say anything in general about what the factors mean? We can.

The key point, as it turns out, is that in Gessel's determinant (4.1), the matrix entries depend only on $|i - j|$ (such a matrix is called a *Toeplitz* matrix). The determinants of such matrices have a natural factorization, as follows. If $a_0, a_1, \dots$ is some sequence, and $a_{-i} = a_i$, then we have

$$\det(a_{i-j})_{i,j=1}^{2m}$$
$$= \det(a_{i-j} + a_{i+j-1})_{i,j=1}^{m} \det(a_{i-j} - a_{i+j-1})_{i,j=1}^{m}.$$

When we apply this fact to the present situation it correctly reproduces the above factorizations for $k = 2, 4$, and generalizes them to all even k, as follows.

Let $y_k(n)$ be the number of Young tableaux of n cells whose first row is of length at most k, and let

$$U_k(x) = \sum_{n \geqslant 0} \frac{u_k(n)}{n!^2} x^{2n},$$

$$Y_k(x) = \sum_{n \geqslant 0} \frac{y_k(n)}{n!} x^n.$$

In terms of these two generating functions, the general factorization theorem states that

$$U_k(x) = Y_k(x)Y_k(-x) \quad (k = 2, 4, 6, \dots).$$

Why is it useful to have such factorizations? For one thing we can equate the coefficients of like powers of x on both sides of this factorization (try it!). We then find an interesting explicit formula that relates the number of Young tableaux of n cells whose first row is of length at most k, on the one hand, and the number of permutations of n letters that have no increasing subsequence of length greater than k, on the other. No more direct proof of this relationship is known. For more details and some further consequences, see Wilf (1992).

Moral: cherchez les factorisations!

5   A Score for Sloane's Database

Here is a case study in which, as it happens, not only was Sloane's database utilized, but Sloane himself was one of the authors of the ensuing research paper.

Eric Weisstein, the creator of the invaluable web resource *MathWorld*, became interested in the enumeration of 0–1 matrices whose eigenvalues are all positive real numbers. If $f(n)$ is the number of $n \times n$ matrices whose entries are all 0s and 1s and whose eigenvalues are all real and positive, then by computation, Weisstein found for $f(n)$ the values

$$1, \ 3, \ 25, \ 543, \ 29\,281 \quad (\text{for } n = 1, 2, \dots, 5).$$

Upon looking up this sequence in Sloane's database, Weisstein found, interestingly, that this sequence is identical, as far as it goes, with sequence `A003024` in the database. The latter sequence counts vertex-labeled acyclic directed graphs ("digraphs") of n vertices, and so Weisstein's conjecture was born:

the number of vertex-labeled acyclic digraphs of n vertices is equal to the number of $n \times n$ 0–1 matrices whose eigenvalues are all real and positive.

This conjecture was proved in McKay et al. (2003). Enroute to the proof of the result, the following somewhat surprising fact was shown.

Theorem 1.1. *If a 0–1 matrix A has only real positive eigenvalues, then those eigenvalues are all equal to* 1.

To prove this, let $\{\lambda_i\}_{i=1}^n$ be the eigenvalues of A. Then

$$
\begin{aligned}
1 &\geqslant \frac{1}{n}\mathrm{trace}(A) \quad \text{(since all } A_{i,i} \leqslant 1) \\
&= \frac{1}{n}(\lambda_1 + \lambda_2 + \cdots + \lambda_n) \\
&\geqslant (\lambda_1 \lambda_2 \cdots \lambda_n)^{1/n} \\
&= (\det A)^{1/n} \\
&\geqslant 1,
\end{aligned}
$$

in which the third line uses the arithmetic–geometric mean inequality, and the last line uses the fact that $\det A$ is a positive integer. Since the arithmetic and geometric means of the eigenvalues are equal, the eigenvalues are all equal, and in fact all $\lambda_i(A) = 1$.

The proof of the conjecture itself works by finding an explicit bijection between the two sets that are being counted. Indeed, let A be an $n \times n$ matrix of 0s and 1s with positive eigenvalues only. Then those eigenvalues are all 1s, so the diagonal of A is all 1s, whence the matrix $A - I$ also has solely 0s and 1s as its entries. Regard $A - I$ as the vertex adjacency matrix of a digraph G. Then (it turns out that) G is acyclic.

Conversely, if G is such a digraph, let B be its vertex adjacency matrix. By renumbering the vertices of G, if necessary, B can be brought to triangular form with zero diagonal. Then $A = I + B$ is a 0–1 matrix with positive real eigenvalues only. But then the same must have been true for the matrix $I + B$ before simultaneously renumbering its rows and columns. For more details and more corollaries see McKay et al. (2003).

Moral: look for your sequence in the online encyclopedia!

6 The 21-Stage Rocket

Now we'll describe a successful attack that was carried out by Andrews (1998) on the problem of evaluating the Mills–Robbins–Rumsey determinant, which is the determinant of the $n \times n$ matrix

$$
M_n(\mu) = \left(\binom{i + j + \mu}{2j - i} \right)_{0 \leqslant i,j \leqslant n-1}. \tag{6.1}
$$

This problem arose (Mills, Robbins, and Rumsey 1987) in connection with the study of *plane partitions*. A plane partition of an integer n is an (infinite) array $n_{i,j}$ of nonnegative integers whose sum is n, subject to the restriction that the entries $n_{i,j}$ are nonincreasing across each row, and also down each column.

It turns out that $\det M_n(\mu)$ can be expressed neatly as a product, namely as

$$
\det M_n(\mu) = 2^{-n} \prod_{k=0}^{n-1} \Delta_k(2\mu), \tag{6.2}
$$

in which

$$
\Delta_{2j}(\mu) = \frac{(\mu + 2j + 2)_j (\frac{1}{2}\mu + 2j + \frac{3}{2})_{j-1}}{(j)_j (\frac{1}{2}\mu + j + \frac{3}{2})_{j-1}},
$$

and $(x)_j$ is the rising factorial $x(x+1)\cdots(x+j-1)$.

The strategy of Andrews's proof is elegant in conception and difficult in execution: we are going to find an upper triangular matrix $E_n(\mu)$, whose diagonal entries are all 1s, such that

$$
M_n(\mu) E_n(\mu) \stackrel{\text{def}}{=} L_n(\mu) \tag{6.3}
$$

is a lower triangular matrix, with the numbers $\{\frac{1}{2}\Delta_{2j}(2\mu)\}_{j=0}^{n-1}$ on its diagonal. Of course, if we can do this, then from (6.3), since $\det E_n(\mu) = 1$, we will have proved the theorem (6.2), since the determinant of the product of two matrices is the product of their determinants, and the determinant of a triangular matrix (i.e. of a matrix all of whose entries below the diagonal are 0s) is simply the product of its diagonal entries.

But how shall we find this matrix $E_n(\mu)$? By holding tightly to the hand of our computer and letting it guide us there. More precisely,

(1) we will look at the matrix $E_n(\mu)$ for various small values of n, and from those data we will conjecture the formula for the general (i, j) entry of the matrix; and then

(2) we will (well actually, "we" won't, but Andrews did) prove that the conjectured entries of the matrix are correct.

It was in step (2) above that an extraordinary 21-stage event occurred which was successfully managed by Andrews. What he did was to set up a system of 21 propositions, each of them a fairly technical hypergeometric identity. Next, he carried out a simultaneous induction on these 21 propositions. That is to say, he showed that if, say, the 13th proposition was true for a certain value of n, then so was the 14th, etc., and if they were all true for that value of n, then the first proposition was true for $n + 1$. The reader should be sure to look at Andrews (1998) to gain more of the flavor and substance of what was done than can be conveyed in this short summary.

Here we will confine ourselves to a few comments about step (1) of the program above. So, let's look at the matrix $E_n(\mu)$ for some small values of n. The condition that $E_n(\mu)$ is upper triangular with 1s on the diagonal means that

$$\sum_{k=0}^{j-1} (M_n)_{i,k} e_{k,j} = -(M_n)_{i,j},$$

for $0 \leqslant i \leqslant j - 1$ and $1 \leqslant j \leqslant n - 1$. We can regard these as $\binom{n}{2}$ equations in the $\binom{n}{2}$ above-diagonal entries of $E_n(\mu)$ and we can ask our CAS to find those entries, for some small values of n. Here is $E_4(\mu)$:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -\dfrac{1}{\mu+2} & \dfrac{6(5+\mu)}{(\mu+2)(\mu+3)(2\mu+11)} \\ 0 & 0 & 1 & -\dfrac{6(\mu+5)}{(\mu+3)(2\mu+11)} \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

At this point the news is all good. While it is true that the matrix entries are fairly complicated, the fact that leaps off the page and warms the heart of the experimental mathematician is that all of the polynomials in μ factor into linear factors with pleasant-looking integer coefficients. So there is hope for conjecturing a general form of the E matrix. Will this benign situation persist when $n = 5$? A further computation reveals that $E_5(\mu)$ is as shown in Figure 1.3. Now it is a "certainty" that some nice formulas exist for the entries of the general matrix $E_n(\mu)$. The Rate package, described in

Section 2.3, would certainly facilitate the next step, which is to find general formulas for the entries of the E matrix. The final result is that the (i, j) entry of $E_n(\mu)$ is 0 if $i > j$ and

$$\frac{(-1)^{j-i}(i)_{2(j-i)}(2\mu + 2j + i + 2)_{j-i}}{4^{j-i}(j-i)!(\mu+i+1)_{j-i}(\mu+j+i+\frac{1}{2})_{j-i}}$$

otherwise.

After divining that the E matrix has the above form, Andrews now faced the task of proving that it works, i.e. that $M_n E_n(\mu)$ is lower triangular and has the diagonal entries specified above. It was in this part of the work that the 21-fold induction was unleashed. Another proof of the evaluation of the Mills–Robbins–Rumsey determinant is in Petkovšek and Wilf (1996). That proof begins with Andrews's discovery of the above form of the $E_n(\mu)$ matrix, and then uses the machinery of the so-called WZ method (Petkovšek, Wilf, and Zeilberger 1996), instead of a 21-stage induction, to prove that the matrix performs the desired triangulation (6.3).

Moral: never give up, even when defeat seems certain.

7   The computation of π

In 1997, a remarkable formula for π was found (Bailey, Borwein, and Plouffe 1997). This formula permits the computation of just a single hexadecimal digit of π, if desired, using minimal space and time. For example, we might compute the trillionth digit of π, without ever having to deal with any of the earlier ones, in a time that is faster than what we might attain if we had to calculate all of the first trillion digits. For example, Bailey, Borwein, and Plouffe found that in the hexadecimal expansion of π, the block of 14 digits in positions 10^{10} through $10^{10} + 13$ are 921C73C6838FB2. The formula is

$$\pi = \sum_{i=0}^{\infty} \frac{1}{16^i} \left(\frac{4}{8i+1} - \frac{2}{8i+4} - \frac{1}{8i+5} - \frac{1}{8i+6} \right). \tag{7.1}$$

In our discussion here we will limit ourselves to describing how we might have found the specific expansion (7.1) once we had decided that an inter-

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -\dfrac{1}{\mu+2} & \dfrac{6(5+\mu)}{(\mu+2)(\mu+3)(2\mu+11)} & -\dfrac{30(\mu+6)}{(\mu+2)(\mu+3)(\mu+4)(2\mu+15)} \\ 0 & 0 & 1 & -\dfrac{6(\mu+5)}{(\mu+3)(2\mu+11)} & \dfrac{30(\mu+6)}{(\mu+3)(\mu+4)(2\mu+15)} \\ 0 & 0 & 0 & 1 & -\dfrac{6(2\mu+13)}{(\mu+4)(2\mu+15)} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Figure 1.3. The upper triangular matrix $E_5(\mu)$.

esting expansion of the form

$$\pi = \sum_{i=0}^{\infty} \frac{1}{c^i} \sum_{k=1}^{b-1} \frac{a_k}{bi+k}. \qquad (7.2)$$

might exist. This, of course, leaves open the question of how the discovery of the form (7.2) was singled out in the first place.

The strategy will be to use the linear dependency algorithm, described above in Section 2.4. More precisely, we want to find a nontrivial integer linear combination of π and the seven numbers

$$\alpha_k = \sum_{i=0}^{\infty} \frac{1}{(8i+k)16^i} \quad (k=1,\dots,7)$$

that sums to 0. As in equation (2.3), we now compute the seven numbers α_j and we look for a relation

$$m\pi + m_1\alpha_1 + m_2\alpha_2 + \cdots + m_7\alpha_7 = 0 \quad (m, m_i \in Z)$$

using, for example, the Maple `IntegerRelations` package. The output vector,

$$(m, m_1, m_2, \dots, m_7) = (1, -4, 0, 0, 2, 1, 1, 0),$$

yields the identity (7.1). You should do this calculation for yourself, then prove that the apparent identity is in fact true, and, finally, look for something similar that uses powers of 64 instead of 16. Good luck!

Moral: even as late as the year AD 1997, something new and interesting was said about the number π.

8   Conclusions

When computers first appeared in mathematicians' environments the almost universal reaction was that they would never be useful for proving theorems since a computer can never investigate infinitely many cases, no matter how fast it is. But computers are useful for proving theorems despite that handicap. We have seen several examples of how a mathematician can act in concert with a computer to explore a world within mathematics. From such explorations there can grow understanding, and conjectures, and roads to proofs, and phenomena that would not have been imaginable in the pre-computer era. This role of computation within pure mathematics seems destined only to expand over the coming years and to be imbued into our students along with EUCLID's axioms and other staples of mathematical education.

At the other end of the rainbow there may lie a more far-reaching role for computers. Perhaps one day we will be able to input some hypotheses and a desired conclusion, press the "Enter" key, and get a printout of a proof. There are a few fields of mathematics in which we can do such things, notably in the proofs of identities (Petkovšek, Wilf, and Zeilberger 1996; Greene and Wilf 2005), but in general the road to that brave new world remains long and uncharted.

Further Reading

Andrews, G. E. 1998. Pfaff's method (I): the Mills–Robbins–Rumsey determinant. *Discrete Mathematics* 193:43–60.

Bailey, D., P. Borwein, and S. Plouffe. 1997. On the rapid computation of various polylogarithmic constants. *Mathematics of Computation* 66:903–913.

Bailey, D. H. and S. Plouffe. 1997. Recognizing numerical constants. In *Proceedings of the Organic Mathematics Workshop, 12–14 December 1995, Simon Fraser University*. Conference Proceedings of the Canadian Mathematical Society, vol. 20.

Borwein, J. and D. Bailey. 2003. *Mathematics by Experiment: Plausible Reasoning in the 21st Century*. Wellesley, MA: AK Peters.

Borwein, J., D. Bailey, and R. Girgensohn. 2004. *Experimentation in Mathematics: Computational Paths to Discovery*. Wellesley, MA: AK Peters.

Calkin, N. and H. S. Wilf. 2000. Recounting the rationals. *American Mathematical Monthly* 107:363–363.

Ferguson, H. R. P. and R. W. Forcade. 1979. Generalization of the Euclidean algorithm for real numbers to all dimensions higher than two. *Bulletin of the American Mathematical Society* 1:912–914.

Gessel, I. 1990. Symmetric functions and P-recursiveness. *Journal of Combinatorial Theory* A 53:257–285.

Graham, R. L., D. E. Knuth, and O. Patashnik. 1989. *Concrete Mathematics*. Reading, MA: Addison-Wesley.

Greene, C. and Wilf, H. S. 2005. Closed form summation of C-finite sequences. *Transactions of the American Mathematical Society*. In press.

Lenstra, A. K., H. W. Lenstra Jr, and L. Lovász. 1982. Factoring polynomials with rational coefficients. *Mathematische Annalen* 261(4):515–534.

McKay, B. D., F. E. Oggier, G. F. Royle, N. J. A. Sloane, I. M. Wanless, and H. S. Wilf. 2004. Acyclic digraphs and eigenvalues of $(0, 1)$-matrices. *Journal of Integer Sequences* 7:04.3.3.

Mills, W. H., D. P. Robbins, and H. Rumsey Jr. 1987 Enumeration of a symmetry class of plane partitions. *Discrete Mathematics* 67:43–55.

Petkovšek, M. and H. S. Wilf. 1996. A high-tech proof of the Mills–Robbins–Rumsey determinant formula. *Electronic Journal of Combinatorics* 3:R19.

Petkovšek, M., H. S. Wilf, and D. Zeilberger. 1996. $A = B$. Wellesley, MA: AK Peters.

Wilf, H. S. 1992 Ascending subsequences and the shapes of Young tableaux. *Journal of Combinatorial Theory* A 60:155–157.

Wilf, H. S. 1994. *generatingfunctionology*, 2nd edn. New York: Academic Press. (This can also be downloaded at no charge from the author's web site.)