Playing dice with primes

Siddhartha Gadgil

Yitang Zhang, an unheralded mathematician in his 50s at the University of New Hampshire, USA, who previously had to work as an accountant and also at a Subway restaurant, has stunned the mathematical world by proving that primes come in pairs. This was a weaker form of a celebrated mathematical problem, the twin prime conjecture, which asks whether there are infinitely many numbers p so that both p and $p + 2$ are primes. Zhang could not find twins, merely distant cousins – he showed that there is a fixed number N, at most about 70 million, so that there are infinitely many pairs of primes p and q that differ by at most N.

At almost the same time, Harald Helfgott, a young mathematician working in Paris, showed that every odd number greater than five is the sum of exactly three primes (the so-called odd Goldbach conjecture), a weaker form of another celebrated problem, the Goldbach conjecture, which asks whether every even number greater than two is the sum of two primes. This improved a result from an year ago of Terence Tao, who had shown that odd numbers are sums of at most five primes.

On seeing such results, one may ask: (i) How can problems involving little more than basic arithmetic be so hard? (ii) Why should we care about such results anyway?

The first question first. It is a remarkable result of Matijasevic that every mathematical problem can be formulated as an arithmetic problem, specifically whether an equation with integer coefficients (e.g. $3x^2 + 12y^2 – 3214z^{12} + 32x = 0$) has solutions with all variables integers. For instance, we can write an equation of this form, so that it has an integral solution exactly if there is an algorithm to factorize integers in polynomial time (and another having solutions exactly if all global solutions to the Navier–Stokes equation are smooth). This result is the climax of a series of fundamental results in logic and computability, beginning with the work of Church, Gödel and Turing.

This does not mean, of course, that specific problems about numbers are interesting, whether or not they are considered so by mathematicians – interesting problems from elsewhere will usually have strange and complicated number theoretic formulations. However, what it does mean is that principles underlying nature often manifest themselves in the study of numbers, and the tools illuminating deep principles underlying numbers are likely to cast a wide light. So the solutions to number theoretic problems are of great value if they are consequences of a deep understanding of principles underlying the numbers, rather than some accidental feature of the specific problem.

Indeed Zhang's result is the consequence of such a principle, beautifully explicated by Tao, who, along with Ben Green, used it to prove another celebrated result in number theory – that primes contain arbitrarily long arithmetic progressions. Crudely, this principle suggests that we should think of primes like molecules in a gas.

Method from madness

A bottle full of gas has so many particles that trying to understand it by solving the governing equations will not get us very far. Yet this very disorder leads to order. While it is essentially impossible to predict the trajectory of a particle, we can predict that the time it spends in any part of the container is proportional to the volume of that part of the container – the particle is equidistributed. Such a view is the basis of statistical physics (whose successes include explaining previously empirical laws of thermodynamics), Einstein's theory of Brownian motion, and many other theories.

If we look at a molecule more closely, taking into account not just where it is but its velocity, which way it is pointing and so on, we get a path not in the physical container but in phase space. A given particle may not cover all of phase space – for instance its energy may be conserved, and perhaps its angular momentum as well. But remarkably, if we allow for such conservation laws, then the particle is (in an appropriate sense) equidistributed in phase space. This can be formulated precisely as ergodicity and mixing properties of the dynamics of the particle. Indeed, assuming a form of randomness, more precisely the so-called Markov properties, lies at the heart of Google's Translate.

Structure and randomness in mathematics

Primes clearly have order – for instance, all primes greater than two are odd. This order can be viewed in terms of mathematical structure. What one can postulate (as explicated by Tao) is that, just as a gas is random after taking into account conserved quantities, primes are random after taking into account appropriate structures. The randomness we seek is not that of traffic on Indian roads – leading to unpredictability, but that of a gas, with mixing properties giving global order from local disorder. Such mixing properties were earlier discovered in many other mathematical contexts.

So what does this say about twin primes? It is useful to view the twin prime conjecture as an instance of a general question: for which sets S of natural number are there infinitely many integers n so that both n and $n + 2$ are contained in S? This clearly depends on how big S is, more precisely the density of S – the fraction of elements between 1 and n that are contained in S (for n large). In the case of primes, the prime number theorem (from the 1800s) tells us that this proportion is approximately $1/\log(n)$.

For a set S with a lot of structure, such as the set of squares, it is straightforward to determine whether there are infinitely many twin pairs. What if we take a random set, with the same density as primes? Then the chance that a number k chosen at random between 1 and n is prime is about $1/\log(n)$, so the chance that both k and $(k + 2)$ are prime is about $(1/\log(n))^2$. But there are about n such pairs, so the expected number of twin primes is $n/(\log(n))^2$. As n becomes larger, this goes to infinity, so there are infinitely many twin primes. This means that a random set with the same density as primes will satisfy the twin prime conjecture.

On the other hand, by the same argument such a random set will contain inifinitely many pairs of adjacent numbers (n, $n + 1$). This is clearly not the case for primes, as one of n and $n + 1$ must be even and so not a prime unless $n = 2$. So we need a refinement, that primes are random enough, after understanding and then taking account the underlying structure. It is indeed such results that led Zhang, and earlier Tao and Green, to their deep results.

A little theory is a dangerous thing

So what are the lessons for science? I would postulate the following: to the extent that there is a genuinely useful theory, such as Newton's laws of motion, we should of course seek it. But beyond this, rather than building increasingly baroque models that are increasingly poor approximations to the observed, it is more fruitful to seek the kind of randomness that leads to order.

Siddhartha Gadgil is in the Department of Mathematics, Indian Institute of Science, Bangalore 560 012, India.
e-mail: gadgil@math.iisc.ernet.in