NUMBER THEORY_PRIMES

Jumping Champions

Ian Stewart leaps over the gaps between prime numbers

Mathematics is full of surprises. Who would have imagined, for instance, that something as straightforward as the natural numbers (1, 2, 3, 4,...) could give birth to anything so baffling as the prime numbers (2, 3, 5, 7, 11, ...)? The pattern of natural numbers is obvious: no matter which one you pick, it's easy to determine what the next one is. You can't say that for the primes. And yet it's a simple step from natural numbers to primes. Just take the natural numbers that have no proper divisors.

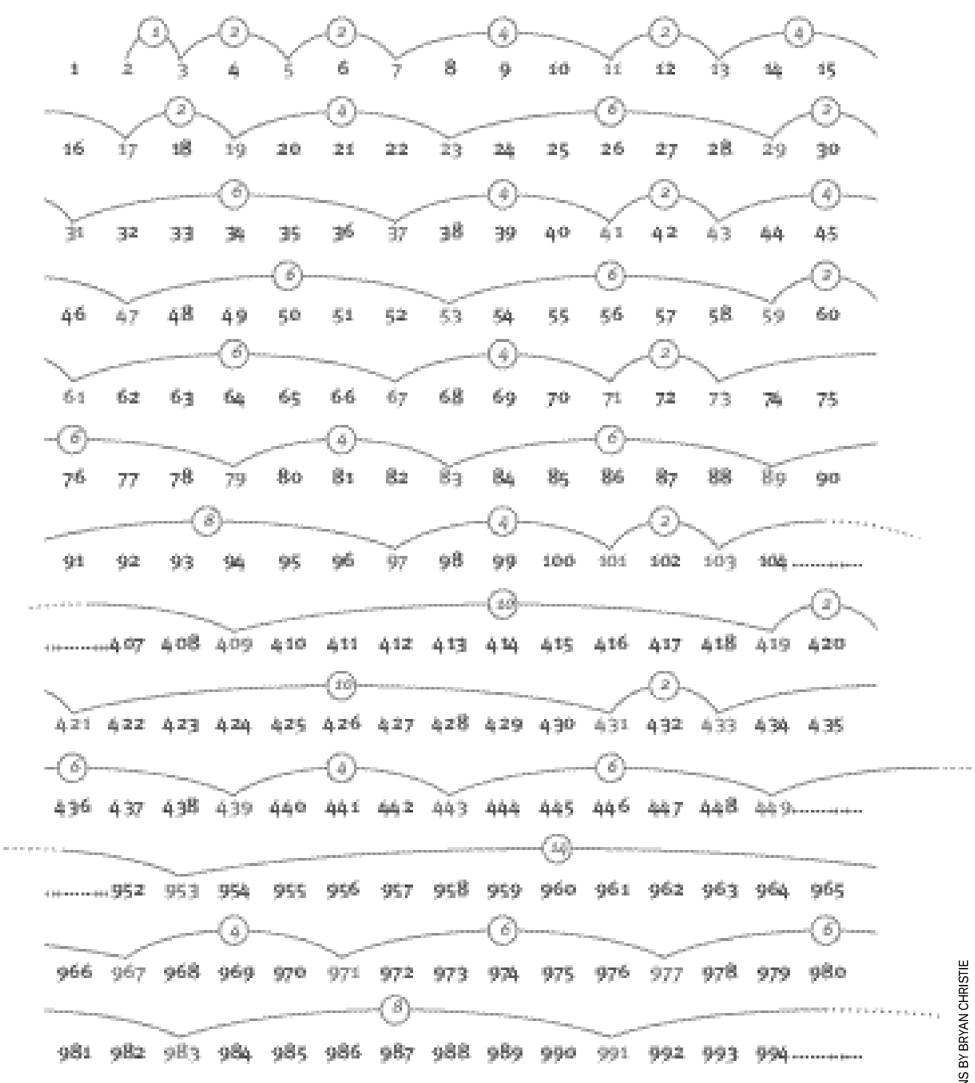
We know a lot about the primes, including some powerful formulas that provide good approximations when exact answers aren't forthcoming. The Prime Number Theorem states that the number of primes less than x is approximately $x / \log x$, where log denotes the natural logarithm. So, for instance, we know that there are roughly 4.3×10^{97} primes with less than 100 digits—but the exact number is a total mystery.

Recently Andrew Odlyzko of AT&T Labs, Michael Rubinstein of the University of Texas and Marek Wolf of the University of Wroclaw in Poland turned their attention to the gaps between successive primes. In an article in *Experimental Mathematics* (Vol. 8, No. 2, 1999), they addressed the following problem: What number is the most common gap between successive primes less than x? This question was posed in the late 1970s by Harry Nelson of Lawrence Livermore National Laboratory. Later on, John Horton Conway of Princeton University coined the phrase "jumping champions" to describe these numbers.

The primes up to 50 are 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43 and 47. The sequence of gaps—the differences between each prime and the next—is 1, 2, 2, 4, 2, 4, 2, 4, 6, 2, 6, 4, 2 and 4. The number 1 appears only once because all primes except for 2 are odd. The rest of the gaps are even numbers. In this sequence, 2 occurs six times, 4 occurs five times, and 6 occurs twice. So when $x = 50$, the most common gap is 2, and this number is therefore the jumping champion.

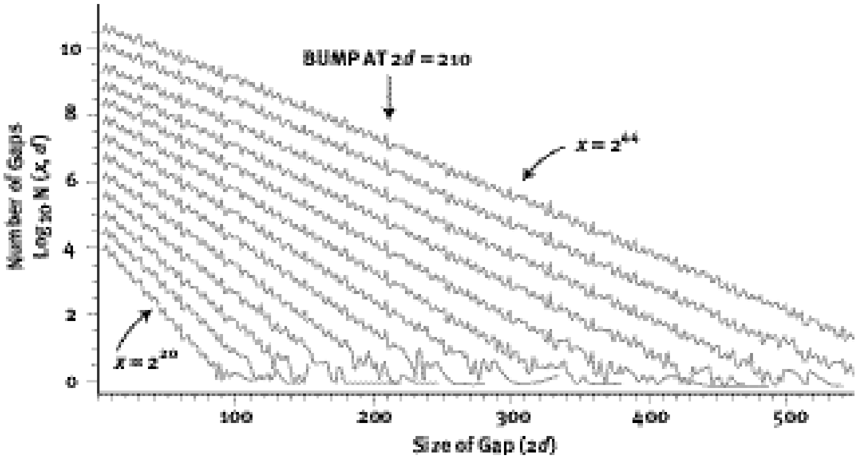
Sometimes several gaps are equally common. For instance, when $x = 5$ the gaps are 1 and 2, and each occurs once. For higher x, the sole jumping champion is 2 until we reach $x = 101$, when 2 and 4 are tied for the honor [*see illustration below*]. After that, the jumping champion is either 2 or 4, or both, until $x = 179$, when 2, 4 and 6 are involved in a three-way tie. At that point the challenge from 4 and 6 dies away, and 2 reigns supreme until $x = 379$, where 2 is tied with 6. Above $x = 389$ the jumping champion is mostly 6, occasionally tied with 2 or 4, or both. But when x ranges from 491 to 541, the jumping champion reverts to 4. From $x = 947$ onward the sole jumping champion is 6, and a computer search shows that this continues up to at least $x = 10^{12}$.

It seems reasonable to conclude that apart from some initial competition from 1, 2 and 4, the only long-term jumping champion is 6. But even a pattern that



MOST COMMON GAPS between successive prime numbers are 2, 4 and 6 for the sequences running up to 1,000. But no one knows the jumping champions for very long sequences of primes.

ALL ILLUSTRATIONS BY BRYAN CHRISTIE



LOGARITHMIC PLOT shows how the number of gaps between successive primes less than x varies with the size of the gap ($2d$). The plot suggests that 210 may be a jumping champion.

persists up to numbers in the trillions may well change as the numbers get still bigger. And that's where the surprise comes in. Odlyzko and his colleagues provide a persuasive argument that somewhere near $x = 1.7427 \times 10^{35}$ the jumping champion changes from 6 to 30. They also suggest that it changes again, to 210, near $x = 10^{425}$.

Except for 4, the conjectured jumping champions fit into an elegant pattern, which becomes obvious if we factor them into primes:

$$2 = 2$$
$$6 = 2 \times 3$$
$$30 = 2 \times 3 \times 5$$
$$210 = 2 \times 3 \times 5 \times 7$$

Each number is obtained by multiplying successive primes together. These numbers are called primorials—like factorials, but using primes—and the next few are 2,310, 30,030 and 510,510. In their article, Odlyzko and his co-authors propose the Jumping Champions Conjecture: the jumping champions are precisely the primorials, together with 4.

Here's a brief explanation of their analysis. Anyone who looks at the sequence of primes notices that every so often two consecutive odd numbers are prime: 5 and 7, 11 and 13, 17 and 19. The Twin Prime Conjecture states that there are infinitely many such pairs. It is based on the idea that primes occur "at random" among the odd numbers, with a probability based on the Prime Number Theorem. Of course, this sounds like nonsense—a number is either prime or not; there isn't any probability involved—but it is reasonable nonsense for this kind of problem. According to a calculation of probabilities, there is no chance that the list of twin primes is finite.

What about three consecutive odd numbers being prime? There is only one example: 3, 5, 7. Given any three consecutive odd numbers, one must be a multiple of 3, and that number is therefore not prime unless it happens to equal 3. Yet the patterns $p, p + 2, p + 6$ and $p, p + 4, p + 6$ cannot be ruled out by such arguments, and they seem to be quite common. For example, the first type of pattern occurs for 11, 13, 17 and again for 41, 43, 47. The second type of pattern occurs for 7, 11,13 and again for 37, 41, 43.

About 80 years ago English mathematicians Godfrey Harold Hardy and John Edensor Littlewood analyzed patterns of this kind involving larger numbers of primes. Using the same kind of probabilistic calculation that I described for the twin primes, they deduced a precise formula for the number of sequences of primes with a given pattern of gaps. The formula is complicated, so I won't show it here; see the article in *Experimental Mathematics* and the references therein.

From the Hardy-Littlewood work, Odlyzko and his colleagues extracted a formula for N (x, d), which is the number of gaps between consecutive primes when the gap is of size $2d$ and the primes are less than x. (We use $2d$ rather than d because the size of the gap has to be even.) The formula is expected to be valid only when $2d$ is large and x is much larger. The illustration at the left shows how log N (x, d) varies with $2d$ for 13 values of x ranging from 2^{20} to 2^{44} (in this graph, log denotes a base 10 logarithm). Each plot line is approximately straight but has lots of bumps. A particularly prominent bump occurs at $2d = 210$, the conjectured jumping champion for very large x. (It would look even more prominent if the logarithmic graphing didn't flatten it out.) This kind of information suggests that the N (x, d) formula is not too wide off the mark.

Now, if $2d$ is going to be a jumping champion, the value of N (x, d) has to be pretty big. The best way to achieve this is if $2d$ has many distinct prime factors. Also, $2d$ should be as small as possible subject to this condition, so the most plausible choices for $2d$ are the primorials. The known jumping champion 4 is presumably an exception. It occurs at a size where the N (x, d) formula isn't a good approximation anyway. The formula also lets us work out roughly when a given primorial takes over from the previous one as the new jumping champion.

What's left for recreational mathematicians to do? Prove the Jumping Champions Conjecture, of course—or disprove it. If you can't do either, try searching for other interesting properties of the gaps between primes. For example, what is the least common gap (that actually occurs) between consecutive primes less than x? And which gap occurs closest to the average number of times? As far as I know, these questions are wide open, even for relatively small values of x.

READER_FEEDBACK

In a recent column on logical paradoxes ["Paradox Lost," June], I argued that the Surprise Test paradox rests on an inconsistent interpretation of the word "surprise" and isn't really a paradox at all.

Several readers drew my attention to an article entitled "Surprise Maximization" in *American Mathematical Monthly* (Vol. 107, No. 6, June–July 2000). The authors define a measure of surprise and ask what strategy the teacher should follow to maximize the students' surprise. They conclude that in choosing the day of the week for the test, the teacher should use a probability distribution that remains roughly constant through the early part of the week but increases rapidly in the last few days. Under this strategy, Friday would be chosen most often. *—I.S.*