On the twin prime conjecture

Alf van der Poorten and Teur Ah-Mah

Introductory Remarks

Just occasionally, the proposals of an amateur are sufficiently amusing to warrant that their dismissal be accompanied by some discussion with their perpetrator. Here I stage such a discussion in the form of a joint paper with Mr Teur, known to his friends as 'Bill'. *AJvdP*

Key words: prime pair, prime n-tuple.
MSC: Primary 11A41, 11N36.

1 Twin Primes

Schemata of Eratosthenes

The Sieve of Eratosthenes sequentially removes the n-th prime p_n, and all multiples of p_n, from the preceding array, S_n say, of integers. Here, for examples,

$$S_2 = \begin{matrix} 1 & 3 & 5 \\ 7 & 9 & 11 \\ 13 & 15 & 17 \\ \vdots & \vdots & \vdots \end{matrix} \qquad S_3 = \begin{matrix} 1 & 5 & 7 & 11 & 13 & 17 & 19 & 23 & 25 & 29 \\ 31 & 35 & 37 & 41 & 43 & 47 & 49 & 53 & 55 & 59 \\ 61 & 65 & 67 & 71 & 73 & 77 & 79 & 83 & 85 & 89 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{matrix}$$

where one finds it convenient to list S_n as an array in which successive elements in each column differ by $P_n = 2 \cdot 3 \cdot \ \cdots \ \cdot p_n$; thus by $2 \cdot 3 = 6$ in S_2, and by $2 \cdot 3 \cdot 5 = 30$ in S_3. Notice that the array S_n has $L_n = P_n(1 - 1/2)(1 - 1/3) \cdots (1 - 1/p_{n-1})$ columns. At the insistence of the first author we recall that 1 is not a prime, and at the suggestion of the second author we refer to each array S_n as a 'schema'.

It is now easy to describe just how the $(n+1)$-st schema S_{n+1} differs from the n-th schema S_n. Namely, to obtain S_{n+1}, remove from S_n those columns divisible by p_n — note that if any one element of a column of S_n is divisible by p_n then all are — and take what remains of the first p_{n+1} rows to form the first row of S_{n+1}. For example, to obtain S_4, we remove the columns headed by 5 and 25 from S_3, and take the remains of the first $p_4 = 7$ rows to make up the $7 \cdot (10 - 2) = L_4 = 56$ elements of the first row of S_4.

Pairs

We call a pair of integers $(k, k+2)$ — thus differing by two — a 'pair', and note that if a schema contains any one pair then it contains two columns differing by two (after a shift upwards if it is the last and first columns), so it contains infinitely many pairs. Indeed, it suffices to consider just the 'leading pairs', those essentially in the first row. For example, all pairs in S_3 are signalled by the leading pairs $(5, 7)$, $(11, 13)$, $(17, 19)$, $(23, 25)$ and — interfered with by that pesky 1 that one of us has insisted on keeping — $(29, 31)$. All other pairs in S_3 differ from these by translation by multiples of $P_3 = 30$.

It is mildly striking that all but one of these pairs is in fact a pair of primes, in brief, a 'prime pair'. Further, we can readily read off the leading $21 = (4-1) \cdot 7$ pairs in S_4 and learn that they are

$$(11,13),(17,19),(29,31),(41,43),(47,49),(59,61),(71,73),(77,79),(89,91),$$
$$(101,103),(107,109),(119,121),(131,133),(137,139),(149,151),(161,163),$$
$$(167,169),(179,181),(191,193),(197,199),(209,211)$$

once again, with a seemingly disproportionate number of prime pairs. In particular, the four pairs preceding the pair containing $49 = 7^2$ all are prime pairs, just as, in S_3, the three pairs preceding the pair containing $25 = 5^2$ all are prime pairs. By the way, note that 209 is blatantly divisible by 11.

Prime Pairs

Note, moreover, that it is not just happenstance that the pairs in S_n with both k and $k+2$ less than p_n^2 all are prime pairs. Indeed, p_n^2 is the *smallest* composite number in the schema S_n. Thus to show that there remains a prime pair in the n-th schema S_n it suffices to find a pair in S_n between p_n and p_n^2.

Incidentally, the prime p_{n+1} of course lies between p_n and p_n^2. That's plain by Bertrand's postulate[1], on which a young Paul Erdős is said to have remarked:

> Chebychev said it, and I'll say it again,
> There's always a prime between n and $2n$.

The second of us now suggests as follows. It has long been known, at least since Euclid, that there are infinitely many primes. But, it is notorious that it is not known before now whether there are infinitely many prime *pairs*.

Suppose, however, that there are only finitely many prime pairs. Then there is an h so that there are no prime pairs exceeding p_h. Therefore, there certainly are no pairs between p_h and p_h^2 in S_h, and no pairs between p_{h+1} and p_{h+1}^2 in S_{h+1}, and no pairs between p_{h+2} and p_{h+2}^2 in S_{h+2}, and so on.

However, pairs in any schema S_{h+n} already occur in all the preceding schemata. Thus the observation just made shows that there are no pairs greater than p_h in any schema, and in particular no pairs at all in S_h nor in any succeeding schema.

But that's absurd, particularly (but not only) because the number of leading pairs in S_{h+1} always exceeds that in S_h.[2]

2 Afterwords (by AJvdP)

Prime n-tuples

But there's more. For example, it seems clear that there are infinitely many triples $(k, k+2, k+6)$, and triples $(k, k+4, k+6)$, of primes. In contrast, there is, plainly, only the one prime triple $(k, k+2, k+4)$, namely $(3,5,7)$, because obviously at least one of k, $k+2$, and $k+4$ must be divisible by 3. Just so, $(2,3)$ is the unique prime pair $(k, k+1)$, because at least one of k, $k+1$ must be even. Although the matter is less notorious than the twin prime conjecture, it is plain experimentally that every configuration[3] that could possibly consist of n primes infinitely many times does indeed

[1] See [1].

[2] There are $154 = (21-7) \cdot 11$ pairs in S_5, at least $(154-37) \cdot 13$ pairs in S_6, and so on.

[3] 'Configuration' is too cold a word. In discussion with my co-author, I used the word 'ménage', thinking of a ménage à n. I commend this terminology to the world.

occur infinitely often. I am happy to be able to add that the remarks of §1 deal just as certainly with these cases as they do with prime pairs.

Principles

What is one to think presented with the sort of arguments comprising Part 1? I had no doubt, even without reading the details. I wrote:

"When I get a letter purporting to make some great contribution to mathematics I first test the claim by applying several principles. The first is the *principle of the meatgrinder*[4].

That principle points out that there's no gain from busily turning the handle, no matter how energetically or painfully. To get chopped steak out of a meatgrinder you have to put some quality meat into it.

There's another important principle for which I don't as yet have a succinct catchy title. But it it boils down to this:

Mathematics ain't about getting it right; it's about not getting it wrong.

My thinking is this: There's no particular merit in getting an answer to a problem, even a correct answer. After all, applying some algorithm is better done by a machine, and in any case is no more than an organised technique for guessing an answer. An easier — and therefore better — way to guess is to look at the answers in the back of the book, or to lean across and look at what the person next to you has written (on the presumption she is either smarter or more energetic or both than you), or to phone a friend, or . . . . Contrary to what we tend to teach, finding an answer isn't mathematics at all; it's just guesswork, intelligent guesswork, maybe.

Mathematics begins when one checks and verifies that one's guess is indeed not necessarily wrong. So it's a very good idea to illustrate that one's argument fails when it should, that is, when it has no business in working. And one is doing meaningfully higher mathematics when one dissects one's argument — not the turgid computations and details, but the underlying logic — into immediately digestible pieces.

What's the point of this outburst? It's very difficult to prove that an argument is correct. Thus the principal obligation on a mathematician is to struggle to prove that her argument is false. If, and only if, she pursues that struggle but fails does she have any business purporting that her argument might possibly be correct."

Opinions

Bill's arguments fail both tests, particularly the latter. In the event, I added:

"My immediate 'expert opinion' on your remarks is that there just isn't enough meat. Nonetheless, I'm not as critical of your remarks as I expected to be. I admit that your notion of the 'schematic sieve' at first did appear to me to contain some meat, indeed so much so that I was quickly convinced that that meat must be bad — that several critical allegations had to be wrong. On second look, I realised that there was much less wrong than I had guessed, and that to the contrary your opening claims were rather obvious. It's not that the meat is rotting, but that what had seemed to be meat is at best just eggplant."

[4] On my complaining to Kurt Mahler that hard work and apparent ingenuity was making no impact on a problem he had set me, Mahler responded: "Ach, Alf. If you want to get gehaktes Rindfleisch out of a meatgrinder then you must put some *steak* into the meatgrinder."

Facts

I suggest that Bill's schemata become somewhat less seductive when one looks at them for n large, say $n > 4$. For example, if p_n^2 is trivially small compared to P_n then there's a ton of room for lots of leading pairs in S_n without much pressure on those pairs to be forced to be prime pairs. The issue is to estimate quantities such as P_n. I remind myself how to do that by looking myself up [3], p75. First, I recall the Euler product formula

$$\sum_{n=1}^{\infty} \frac{1}{n^s} = \Big(1 - \frac{1}{2^s}\Big)^{-1}\Big(1 - \frac{1}{3^s}\Big)^{-1}\Big(1 - \frac{1}{5^s}\Big)^{-1}\Big(1 - \frac{1}{7^s}\Big)^{-1} \cdots = \prod_p \Big(1 - \frac{1}{p^s}\Big)^{-1}$$

for the Riemann ζ-function, a formula easily verified by multiplying both sides by $(1-1/2^s)$, and so on. Because $1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n} \approx \log n$ we find, after taking a logarithm, that $\sum_{p<x} 1/p \approx \log\log x$. Since $\log\log x = \int_e^x dt/t\log t$ we might deduce that the n-th prime $p_n \approx n\log n$ and that the number of primes less than x is $\approx x/\log x$. In this spirit, we estimate $\log P_n = \sum_{p\le n\log n} \log p$ by the integral $\int^{n\log n} \log t\,(dt/\log t)$, obtaining $P_n \approx n^n$. Compared to it, p_n^2 *is* trivial.

The largest *known* prime pair, not all that long ago, was $4\,650\,828 \cdot 1001 \cdot 10^{3429} \pm 1$; these primes are 3439 digits long. However, I've now seen $835\,335 \cdot 2^{39014} \pm 1$, (due to Ray Ballinger) which have a heftier 11751 digits.

Amazingly, the prime n-tuples conjecture alluded to in §2 is incompatible with a seemingly more obvious 'fact'. It is usual to denote the number of primes less than x by $\pi(x)$, and it is well known that $\pi(x) \approx x/\log x$; here the kicker is in the error term. Now, everyone 'knows' that there are more primes at first than later. Thus $\pi(100) = 25$ but there are only 21 primes in the next hundred, and only 16 in the next; $\pi(1000)$ is just 168. That is, for sizeable x, y, always $\pi(x) + \pi(y) > \pi(x+y)$. Not! [2]. The latter 'fact' is incompatible with the prime n-tuples conjecture, and anyone with feeling for these matters[5] votes for the conjecture.

References

[1] M. Aigner and G.M. Ziegler, *Proofs from The Book*, (Springer-Verlag Berlin 1999).

[2] D. Hensley and I. Richards, *On the incompatibility of two conjectures concerning primes*, in *Analytic Number Theory* Proc. Symp. Pure Math., **24**, (Amer. Math. Soc., Providence, R.I., 1973), 123–127; *Primes in intervals*, Acta Arith. **25** (1973/74), 375–391.

[3] A. van der Poorten, *Notes on Fermat's Last Theorem*, (Wiley-Interscience New York 1996).

ceNTRe for Number Theory Research, 1 Bimbil Place, Killara, NSW 2071
E-mail: alf@math.mq.edu.au

Received 19 September 2003, accepted 3 October 2003.

[5] Also, not! I recently had a totally incoherent conversation with Władysław Narkiewicz — my confusion being caused primarily by the fact that I just could not believe that Vlad was insisting that *everyone* (thus, he at least) is confident that $\pi(x) + \pi(y) > \pi(x+y)$ *does* hold generally.