Structure and randomness in the prime numbers

Terence Tao

University of California, Los Angeles

July 19 / IMO

The Distribution Of The Primes

- The prime numbers $2, 3, 5, 7, \ldots$ are one of the oldest topics studied in mathematics.
- We now have a lot of intuition as to how the primes should behave, and a great deal of confidence in our conjectures about the primes...
- ... but we still have a great deal of difficulty in proving many of these conjectures!
- Ultimately, this is because the primes are believed to not obey behave pseudorandomly in many ways, and not to follow any simple pattern.
- We have many ways of establishing that a pattern exists... but how does one demonstrate the absence of a pattern?

It is hard to find primes!

- **Euclid's theorem** ($\sim$ 300 BCE): There are infinitely many primes.
- In particular, given any k, there exists a prime with at least k digits.
- But there is no known quick and deterministic way to locate such a prime! (Here, "quick" means "computable in a time which is polynomial in k".)
- In particular, there is no known (deterministic) formula that can quickly generate large numbers that are guaranteed to be prime.
- The largest known prime is $2^{43,112,609} - 1$ (GIMPS 2008) - about 13 million digits long.

But primes can be found by random methods...

- Any k-digit number can be tested for primality quickly, either by probabilistic methods (Miller-Rabin 1980) or by deterministic methods (Agarwal-Kayal-Saxena 2002).
- **Prime number theorem** (Hadamard, de Vallée Poussin 1896): The number of primes less than n is $(1 + o(1))\frac{n}{\log n}$ as $n \to \infty$.
- In particular, the probability of a randomly selected k-digit number being prime is about $\frac{1}{k \log 10}$.
- So one can quickly find a k-digit prime with high probability by randomly selecting k-digit numbers and testing each of them for primality.

Is randomness really necessary?

To summarize:

- We do not know a quick way to find primes determinstically.
- However, we have quick ways to find primes randomly.
- Also, a major conjecture in complexity theory, $P = BPP$, asserts that any problem that can be solved quickly by probabilistic methods, can also be solved quickly by deterministic methods. (This conjecture is closely related to the more famous conjecture $P \neq NP$, which is a USD $ 1 million Clay Millennium prize problem.)
- Many other important probabilistic algorithms have been derandomised into deterministic ones, but this has not been done for the problem of finding primes.

Primes can be studied collectively rather than individually

- We've seen that it's hard to get a hold of any single large prime.
- But it is easier to study the set of primes collectively rather than one at a time.
- An analogy: it is difficult to locate and count all the grains of sand in a box, but one can get an estimate on this count by weighing the box, subtracting the weight of the empty box, and dividing by the average weight of a grain of sand.
- The point is that there is an easily measured statistic (the weight of the box) which is reflects the collective behaviour of the sand.

Euler's product formula

- For instance, from the fundamental theorem of arithmetic one can establish Euler's product formula

$$\sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p \text{ prime}} (1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \frac{1}{p^{3s}} + \ldots)$$

$$= \prod_{p \text{ prime}} (1 - \frac{1}{p^s})^{-1}$$

for any $s > 1$ (and also for other values of s, if one is careful enough).

The Riemann zeta function

- This formula links the collective behaviour of the primes to the behaviour of the Riemann zeta function $\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s}$:

$$\prod_{p \text{ prime}} (1 - \frac{1}{p^s}) = \frac{1}{\zeta(s)}$$

- One can then deduce information about the primes from information about the zeta function (and in particular, its zeroes).
- For instance, the fact that $\zeta(s)$ has no zeroes when $\mathrm{Re}(s) \geq 1$ implies (and is in fact equivalent to) the prime number theorem.

The Riemann hypothesis

- The famous Riemann hypothesis asserts that $\zeta(s)$ has no zeroes when $\mathrm{Re}(s) > 1/2$. It implies a much stronger version of the prime number theorem, and has many other consequences in number theory; it is another of the USD $ 1 million Clay Millennium prize problems.
- (A technical point: the sum $\sum_{n=1}^{\infty} \frac{1}{n^s}$ does not converge in the classical sense when $\mathrm{Re}(s) < 1$, so one has to interpret this sum in a fancier way, or else use a different definition of $\zeta(s)$ in this case; but I will not discuss these subtleties here.)

Cramér's random model for the primes

- A fruitful way to think about the set of primes is as a pseudorandom set - a set of numbers which is not actually random, but behaves like one.
- For instance, the prime number theorem asserts, roughly speaking, that a randomly chosen large integer n has a probability of about $1/\log n$ of being prime. One can then model the set of primes by replacing them with a random set of integers, in which each integer $n > 1$ is selected with an independent probability of $1/\log n$; this is Cramér's random model.

Refining the model

- This model is too crude, because it misses some obvious structure in the primes, such as the fact that most primes are odd. But one can improve the model to address this, by picking a model where odd integers n are selected with probability $2/\log n$ and even integers are selected with probability 0.
- One can also take into account other obvious structure in the primes, such as the fact that most primes are not divisible by 3, not divisible by 5, etc. This leads to fancier random models which we believe to accurately predict the asymptotic behaviour of primes.

Random models give heuristic support for conjectures

- One can use random models to give heuristic support for conjectures in number theory, such as
- **Twin prime conjecture**: There are infinitely many pairs $p, p+2$ which are both prime (twin prime pairs).
- Indeed, with Cramér's random model, if n is a large integer, then $n, n+2$ have probability $1/\log n$ and $1/\log(n+2)$ to be prime, so the probability that they are both prime is $\frac{1}{(\log n)(\log n+2)}$. Thus the total number of twin prime pairs less than N should be about $\sum_{n=2}^{N} \frac{1}{(\log n)(\log n+2)} \sim \int_2^N \frac{dx}{\log^2 x}$. This goes to infinity as $N \to \infty$, so this predicted count is consistent with the twin prime conjecture.

Random models cont.

- Using more refined models using the "obvious" structure in the primes, the total number of twin prime pairs less than N is predicted to be about $\Pi_2 \int_2^N \frac{dx}{\log^2 x}$, where Π_2 is the twin prime constant

$$\Pi_2 := 2 \prod_{p \geq 3 \text{ prime}} 1 - \frac{1}{(p-1)^2} \approx 1.32032\ldots.$$

  For $N = 10^{10}$, this prediction is accurate to four decimal places.
- Similar arguments based on random models give convincing heuristic support for many other conjectures in number theory, and are backed up by extensive numerical calculations.

How to convert probabilistic heuristics into rigorous arguments?

- Of course, the primes are a deterministic set of integers, not a random one, so the predictions given by random models are not rigorous. But can they be made so?
- There has been some progress in doing this. One approach is to try to classify all the possible ways in which a set could fail to be pseudorandom (i.e. it does something noticeably different from what a random set would do), and then show that the primes do not behave in any of these ways.

- For instance, consider the **odd Goldbach conjecture**: every odd integer larger than five is the sum of three primes. If, for instance, all large primes happened to have their last digit equal to one, then Goldbach's conjecture could well fail for odd integers whose last digit was different from three. Thus we see that the conjecture could fail if there was a sufficiently strange "conspiracy" among the primes.
- However, one can rule out this particular conspiracy by using the prime number theorem in arithmetic progressions, which tells us that (among other things) there are many primes whose last digit is 1. (The proof of this theorem is based on the proof of the classical prime number theorem.)

- Moreover, by using the techniques of Fourier analysis, we can show that all the conspiracies which could conceivably sink Goldbach's conjecture (for large integers, at least) are broadly of this type: an unexpected "bias" for the primes to prefer one remainder modulo 10 (or modulo another base, which need not be an integer), over another.
- Vinogradov eliminated each of these potential conspiracies, and established Vinogradov's theorem: every sufficiently large odd integer is the sum of three primes.
- Related methods were used to establish my theorem with Ben Green that the primes contained arbitrarily long arithmetic progressions.

Summary

- Individual primes are believed to behave randomly, but the collective behaviour of the primes is believed to be quite predictable.
- We believe that the primes do not observe any significant pattern beyond the obvious ones (e.g. mostly being odd), but we are still a long way from making this belief completely rigorous.
- Nevertheless, we can eliminate some types of patterns in the primes, and this is enough to give us some non-trivial results on the primes. But there is still a lot to do in this subject!