The Distribution of Primes
and
The Twin Prime Conjecture

Bellevue College Math/Physics Colloquium
Larry Curnutt
Spring, 2026

- Definition

  A natural number bigger than 1 is prime, if its only divisors are 1 and itself.
  Otherwise it's called composite.

- Why are prime numbers important?

  Every natural number > 2 can be uniquely expressed as a product of primes.

Pf. (by math induction) Clearly, the first few natural numbers are either prime or easily factored into two or three primes. Assume that all natural numbers $k \le n$ can be expressed as products of primes. We must show that $n+1$ can be expressed as a product of primes. If $n+1$ is prime, then we're done. On the other hand, if $n+1$ is composite, then $n+1 = a \cdot b$, where $a, b \le n$. By the induction hypothesis, a and b can both be expressed as products of primes, and hence, so can $n+1$.

For uniqueness let's look at a simple case, and you'll see how the argument goes, in general.
Suppose $n = p_1 \cdot p_2 = q_1 \cdot q_2 \cdot q_3$. p_1 divides the righthand side, but none of the q's factor, so p_1 must equal one of the q's, say $p_1 = q_2$. Dividing both sides by $p_1 = q_2$, we have $p_2 = q_1 \cdot q_3$. Similarly, p_2 must equal q_1 or q_3, say $p_2 = q_3$. Dividing again, this time by $p_2 = q_3$, we get $1 = q_1$, which is impossible, so the number of p's and q's must be the same, and the factorizations are identical.

- How many primes are there?

The set of prime numbers is infinite. (Euclid)

Pf. (by contradiction) Assume the number of primes is finite: say $p_1, p_2, p_3, \ldots, p_n$ is all of them.
Let $P = p_1 \cdot p_2 \cdot p_3 \cdot \ldots \cdot p_n$ and define $Q = P + 1$. Since Q can be written as a product of primes,
Q is evenly divisible by one of the p_is: $Q = R \cdot p_k$.
By definition $P = S \cdot p_k$, and $1 = Q - P = R \cdot p_k - S \cdot p_k = (R - S) \cdot p_k$.
But $(R - S) \cdot p_k = 1$ is a clear contradiction. Therefore, our original assumption must have been wrong,
so the set of prime numbers must be infinite.

- Are the primes evenly distributed?

There are arbitrarily large gaps in the distribution of primes.

Here is a string of n consecutive composite numbers: for each $k \le n$, C_{k-1} is divisible by $(k + 1)$.

$$C_1 = n! + 2\,,\ C_2 = n! + 3\,,\ \ldots,\ C_{n-1} = n! + n\,,\ C_n = (n+1)! + (n+1)$$

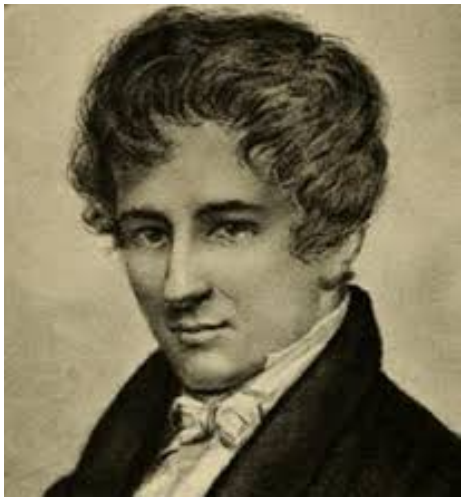
The average size of gaps between primes increases, but it doesn't appear to happen smoothly. You can start with $1{,}000{,}001! + 2$ and build list of 1,000,000 consecutive composites (as above), but primes tend to 'cluster' together here and there. For example, shortly beyond that gap of 1,00,000, there's a gap of length 2: 1,000,037 and 1,000,039 pop up. They're called 'twin primes'.

We'll have more to say about twin primes in a little bit.

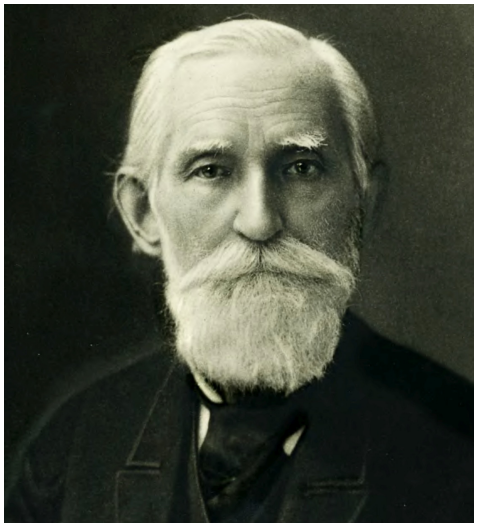
- Is there a formula for primes?

There have been many attempts to find a simple formulas for p_n = the *n*th prime and for $\pi(n)$ = the exact number of primes $\leq n$ — without success.

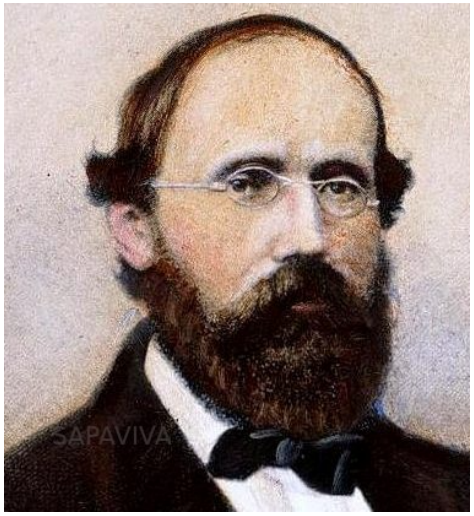
Here's a little history of what we did learn during the 19th century.



As a teenager, Gauss studied $\pi(x)$ by scouring tables of primes.
He guessed that $\pi(n)$ was approximately equal to $n/ln(n)$,
but was unable to prove it, even as an adult.



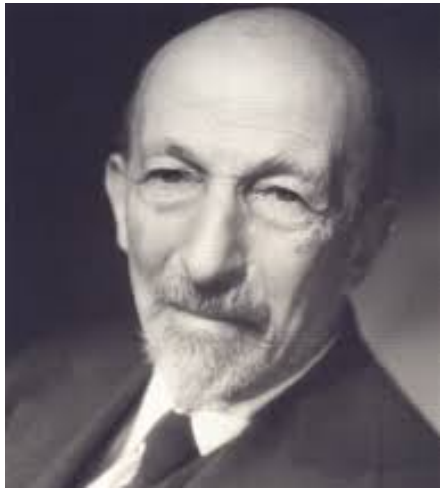
Around 1850 Pafnuty Chebyshev was able to bound $\frac{\pi(n)}{n/ln(n)}$ above and below by 9/8 and 7/8 for sufficiently large values of n.



In his famous 1859 manuscript Bernhard Riemann presented several brilliant results,

among them $\lim_{x\to\infty} \frac{\pi(x)}{x/ln(x)} = 1$.

But his arguments were so sketchy, or omitted completely, that in many cases his work was inconclusive.



It was not until 1896 that Jacques Hadamard and Charles Jean de la Poussin established the existence of this limit beyond a shadow of a doubt, and it became known as ‘The Prime Number Theorem’.

The Prime Number Theorem: $\lim_{x\to\infty} \dfrac{\pi(x)}{x/ln(x)} = 1$ or $\pi(x) \approx \dfrac{x}{ln(x)}$. **(very deep*)**

It takes a little simple, but clever, monkeying around with logarithms, limits, L'Hopital's Rule, and the PNT to see that $\lim_{n\to\infty} \dfrac{p_n}{n \cdot ln(n)} = 1$ or $p_n \approx n \cdot ln(n)$.

* What makes a result "**deep**"? I'd say 'one or more of the following characteristics':
Simple to state, but surprisingly profound, offering some fresh insight.
Long, complex proof, requiring unexpected methods far outside the scope of the statement.
Reveals or defines new fundamental underlying structures.
Connects disparate branches of mathematics.
Establishes, or at least suggests, other results.

[These approximations obviously improve, as x increases. They're useful theoretically, but as you can see in the tables on the next page, they're not very good for practical applications.]

Here's a list of the primes: $p_1 = 2,\ p_2 = 3,\ p_3 = 5,\ p_4 = 7,\ p_5 = 11, \ldots,\ p_n = ???, \ldots$

Here is a sample of $\pi(n)$-values: $\pi(2) = 1,\ \pi(3) = 2,\ \pi(4) = 2,\ \pi(20) = 7,\ \pi(50) = 15,\ \pi(p_n) = n$

x	$\pi(x)$	$x/\ln(x)$	$li(x)$
1,000	168	145	178
10,000	1,229	1,086	1,246
100,000	9,592	8,686	9,630
1,000,000	78,498	72,382	78,628
10,000,000	664,579	664,579	664,918

n	p_n	$n \cdot ln(n)$
10	29	23.02
100	541	460.5
1000	7919	6907.7
100,000	1,299,709	1,151,293
1,000,000	15,485,863	13,815,510

$\pi(x) \approx li(x) = \int_2^x \frac{dt}{ln(t)}$ *is much better; and* $n \cdot ln(n)$ *always under estimates* p_n *by as much as 15%, even for extremely large values of n . If* $n = 10^6$, *then* $p_{10^6} = 15{,}485{,}863$ *and* $10^6 \cdot ln(10^6) \approx 13{,}815{,}510$, *an error of about 10.8%. When* $n = 2 \cdot 10^{17}$, *the error is still about 6.4%.* $p_n \approx (ln(n) + ln(ln(n)) - 1)$ *is much better].*

- Nuances of 'countably infinite' revealed by primes.

The infinite sets $\{n\}$, $\{5n\}$, $\{n^2\}$, and $\{n!\}$ all contain the same number of terms. They can be put into one-to-one correspondence with one another. They're all countable.

Consider the infinite sums $\sum \frac{1}{n}$, $\sum \frac{1}{5n}$, $\sum \frac{1}{n^2}$, and $\sum \frac{1}{n!}$. The first two diverge (to ∞), while the second two both converge (to $\pi^2/6$ and e, respectively).

This suggests that there must be some subtle difference between the 'size' of $\{n\}$ & $\{5n\}$ and the 'size' of $\{n^2\}$ & $\{n!\}$. In some sense, there seem to be 'more' n's & $5n$'s than there are n^2's & $n!$'s. Here's something for you to chew on:

The size of the gaps between successive terms of $\{n\}$ & $\{5n\}$ is constant (1 and 5, respectively).

The size of the gaps between successive terms of $\{n^2\}$ & $\{n!\}$ increase dramatically, but predictably (2n +1 and $n \cdot n!$, respectively).

With these curiosities in mind, let's look at the set of primes, $\{p_n\}$, and ask about the series $\sum_{n=1}^{\infty} \frac{1}{p_n}$.

Does it converge or diverge?

Remember . . . the gaps in $\{p_n\}$ are very irregular. So what's your guess?

In 1737 Euler showed that $\sum \frac{1}{p_n}$ diverges.**

Here's his proof.

Step 1. $\displaystyle\prod_{k=1}^{n} \frac{1}{1-1/p_k} = \prod_{k=1}^{n} \left(1 + \frac{1}{p_k} + \frac{1}{p_k^2} + \frac{1}{p_k^3} + \frac{1}{p_k^4} + \cdots\right)$ [geometric series]

Multiplying all these series together forms a new series with all numerators = 1 and denominators eventually = all possible products of primes and powers of primes $\leq p_n$, which means all natural numbers with prime factors $\leq p_n$: $\displaystyle\sum \frac{1}{k}$.

Step 1. $$\prod_{k=1}^{n} \frac{1}{1-1/p_k} = \prod_{k=1}^{n} (1 + \frac{1}{p_k} + \frac{1}{p_k^2} + \frac{1}{p_k^3} + \frac{1}{p_k^4} + \cdots)$$ [geometric series]

Step 2. Bound the partial sums of the harmonic series:

$$\prod_{k=1}^{n} \frac{1}{1-1/p_k} \geq \sum_{k=1}^{p_n} \frac{1}{k} \geq \int_{1}^{p_n+1} \frac{dx}{x} = ln(p_n + 1) > ln(p_n) .$$

Step 1. $$\prod_{k=1}^{n} \frac{1}{1-1/p_k} = \prod_{k=1}^{n} (1 + \frac{1}{p_k} + \frac{1}{p_k^2} + \frac{1}{p_k^3} + \frac{1}{p_k^4} + \cdots)$$ [geometric series]

Step 2. Bound the partial sums of the harmonic series:

$$\prod_{k=1}^{n} \frac{1}{1-1/p_k} \geq \sum_{k=1}^{p_n} \frac{1}{k} \geq \int_1^{p_n+1} \frac{dx}{x} = ln(p_n + 1) > ln(p_n) \, .$$

Step 3. Taking reciprocals and then logarithms of the left and righthand sides yields:

$$\sum_{k=1}^{n} ln(1 - \frac{1}{p_k}) < - ln(ln(p_n))$$

Step 1. $\displaystyle\prod_{k=1}^{n}\frac{1}{1-1/p_k}=\prod_{k=1}^{n}\left(1+\frac{1}{p_k}+\frac{1}{p_k^2}+\frac{1}{p_k^3}+\frac{1}{p_k^4}+\cdots\right)$ [geometric series]

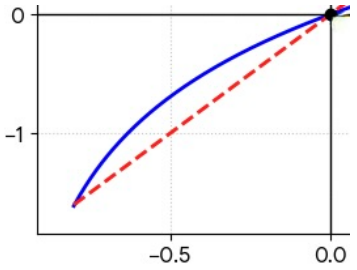
Step 2. Bound the partial sums of the harmonic series:

$$\prod_{k=1}^{n}\frac{1}{1-1/p_k}\geq\sum_{k=1}^{p_n}\frac{1}{k}\geq\int_{1}^{p_n+1}\frac{dx}{x}=ln(p_n+1)>ln(p_n)\,.$$

Step 3. Taking reciprocals and then logarithms of the left and righthand sides yields:

$$\sum_{k=1}^{n}ln(1-\frac{1}{p_k})<-ln(ln(p_n))$$

Step 4. Note that the line $y=2x$ lies below the graph of $y=ln(1+x)$ for $-\frac{1}{2}\leq x<0$. With $x=-\frac{1}{p_k}$, $-\frac{2}{p_k}<ln(1-\frac{1}{p_k})$.



Step 1. $\displaystyle\prod_{k=1}^{n}\frac{1}{1-1/p_k}=\prod_{k=1}^{n}(1+\frac{1}{p_k}+\frac{1}{p_k^2}+\frac{1}{p_k^3}+\frac{1}{p_k^4}+\cdots)$ [geometric series]

Step 2. Bound the partial sums of the harmonic series:

$$\prod_{k=1}^{n}\frac{1}{1-1/p_k}\geq\sum_{k=1}^{p_n}\frac{1}{k}\geq\int_1^{p_n+1}\frac{dx}{x}=ln(p_n+1)>ln(p_n)\ .$$

Step 3. Taking reciprocals and then logarithms of the left and righthand sides yields:

$$\sum_{k=1}^{n}ln(1-\frac{1}{p_k})<-ln(ln(p_n))$$

Step 4. Note that the line $y=2x$ lies below the graph of $y=ln(1+x)$ for $-\frac{1}{2}\leq x<0$. With $x=-\frac{1}{p_k}$, $-\frac{2}{p_k}<ln(1-\frac{1}{p_k})$.

Step 5. Take sums and apply Step 3:

$$-2\sum_{k=1}^{n}\frac{1}{p_k}<\sum_{k=1}^{n}ln(1-\frac{1}{p_k})<-ln(ln(p_k))\quad\text{or}\quad\sum_{k=1}^{n}\frac{1}{p_k}>\frac{ln(ln(p_k))}{2}\ .$$

Step 1. $\displaystyle\prod_{k=1}^{n}\frac{1}{1-1/p_k}=\prod_{k=1}^{n}(1+\frac{1}{p_k}+\frac{1}{p_k^2}+\frac{1}{p_k^3}+\frac{1}{p_k^4}+\cdots)$ [geometric series]

Step 2. Bound the partial sums of the harmonic series:

$$\prod_{k=1}^{n}\frac{1}{1-1/p_k}\geq\sum_{k=1}^{p_n}\frac{1}{k}\geq\int_1^{p_n+1}\frac{dx}{x}=ln(p_n+1)>ln(p_n)\ .$$

Step 3. Taking reciprocals and then logarithms of the left and righthand sides yields:

$$\sum_{k=1}^{n}ln(1-\frac{1}{p_k})<-ln(ln(p_n))$$

Step 4. Note that the line $y=2x$ lies below the graph of $y=ln(1+x)$ for $-\frac{1}{2}\leq x<0$. With $x=-\frac{1}{p_k}$, $-\frac{2}{p_k}<ln(1-\frac{1}{p_k})$.

Step 5. Take sums and apply Step 3:

$$-2\sum_{k=1}^{n}\frac{1}{p_k}<\sum_{k=1}^{n}ln(1-\frac{1}{p_k})<-ln(ln(p_k))\quad\text{or}\quad\sum_{k=1}^{n}\frac{1}{p_k}>\frac{ln(ln(p_k))}{2}\ .$$

Step 6. Conclusion: Since the righthand side $\to\infty$, so does the lefthand side. That is, $\displaystyle\sum_{k=1}^{\infty}\frac{1}{p_k}$ diverges. **QED**

Q.E.D. or Ta–da! or Voila! or Whew! or ■ or ∎ or The proof is complete!

"The symbol ∎ is definitely not my invention – it appeared in popular magazines (not mathematical ones) before I adopted it, but I seem to have introduced it into mathematics. It is used to indicate an end, usually the end of a proof. It is frequently called the 'tombstone', but at least one generous author referred to it as ***'the halmos'****."*

[This quote is from I Want To Be A Mathematician — an autopathography by Paul R. Halmos.]

Q.E.D. = *Quod erat demonstrandum* is a Latin phrase meaning "which was to be demonstrated"

** **Leonard Euler (1707 – 1783, Swiss)**

Legacy: Greatest mathematician of the 18th century ! ! !

Notation: π, e, i, Σ, $f(x)$, ln, $trig$

Polymath: geometry, graph theory, topology, analysis, logic, number theory, physics, astronomy, engineering, optics, demography, music theory, philosophy/religion

Prolific: ≈ one paper per week ≈ 900 works = volumes !

Education: studied under and worked with the Bernoulli's; tutored royalty (Russian, German)

Eyesight: totally blind in one eye by age 30; totally blind in other eye by age 60

Ranking: right up there with Gauss, Newton, and Archimedes

John von Neumann called Euler "the greatest virtuoso of the period."

Arago said "Euler calculated without any apparent effort, just as men breathe and as eagles sustain themselves in air."

Poincaré called Euler the "god of mathematics."

- One last hoo-rah! **Is the set of twin primes finite or infinite?**

Twin primes are pairs of prime numbers that differ by exactly 2. (e.g., 3 & 5, 11 & 13, 17 & 19, etc.).

Twin Prime Conjecture: The set of twin primes, $\{T_n\}$, is infinite. (. . . . maybe!)

Around 1850 Alphonse de Polignac hypothesized even more. He proposed that for every even number k there are infinitely many pairs of consecutive primes that differ by exactly k. Though many have tried, nobody has been able prove even the case $k = 2$; e.g., 175 years later, **the TPC is still unresolved**.

With all our modern computer power, there's lots of numerical evidence to suggest that the TPC is true.

$$3{,}756{,}801{,}695{,}685 \times 2^{666{,}669} - 1 \quad \text{and} \quad 3{,}756{,}801{,}695{,}685 \times 2^{666{,}669} + 1$$

are the largest twin primes discovered so far.

As you might guess, something might be learned by studying the sum of reciprocals of twin primes.

$$\sum \frac{1}{T_n} = \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \frac{1}{11} + \frac{1}{13} + \frac{1}{17} + \frac{1}{19} + \frac{1}{19} + \frac{1}{29} + \frac{1}{31} + \cdots$$

It turns out that that's almost as hard, but not impossible.

In 1919 a Norwegian number theorist, Viggo Brun, showed that $\sum \frac{1}{T_n} \approx 1.902160583...$, whether $\{T_n\}$ is finite or infinite. The twin primes, if infinite, must be extremely rare compared to the primes. In everyday language, the primes are pretty sparse among the natural numbers. Still, there are plenty of them — enough to make $\sum \frac{1}{p_n}$ diverge, while there are only enough twin primes to keep $\sum \frac{1}{T_n}$ smaller than 2 .

By 2013 it was known that "***no matter how far you go into the deserts of the truly gargantuan prime numbers**** — no matter how sparse the primes become — you will keep finding prime pairs that differ by less than 246)." Still a ways to go to settle TPC!

** *I stole, and then highlighted, this phrase from some article I read, because I liked the imagery so much. It's o.k. for mathematicians to care about such prissy things. It's not o.k. for mathematicians to omit references/citations, but I couldn't find this one, and I'm too lazy to search anymore.*

- To summarize, here are a few statistics/comments that illustrate some of what we've been saying.

Prime numbers are abundant at the beginning of the number line, but grow much sparser among large numbers. Of the first 20 numbers, for example, 40% are prime — 2, 3, 5, 7, 11, 13, 17, and 19 — but among 10-digit numbers, only about 4% are prime.

As we've seen, mathematicians have understood how on the average primes taper off (the PNT).
Among large numbers, the expected gap between primes is about 2.3 times the number of digits.
So, for example, among 100-digit numbers, the expected gap between primes is about 230.

How many ways can we say it:

"Infinity, even $\aleph_0$, is a very quirky place"?